STA447/2006 (Stochastic Processes) Lecture Notes, Winter 2019

by Professor Jeffrey Rosenthal, University of Toronto

(Last updated: March 19, 2019)

Note: These lecture notes will be posted on the STA447/2006 course web page after the corresponding lecture material. However, they are just rough notes, with no guarantee of completeness or accuracy. They should not be regarded as a substitute for attending and actively learning from the course lectures and supplementary readings and practice problems.

Introduction:

- Course web page, outline, evaluation, etc. (www.probability.ca/sta447)
- Schedule: will take a 15-minute break if you return promptly!
- Background: STA347 prerequisite – required for undergrads! (last semester? previously?) (includes various math and stat second-year prerequisites)
- Your status: undergrad? grad? special? STA specialist? major? Act Sci? Comp Sci? Math? Physics / Chem? Econ / Finance / Management? Engineering? other?
- You should already know basic probability theory: probability spaces, random variables, expected value, independence, conditional probability, discrete and continuous distributions, etc. (You do not need to know measure theory.)
- This class considers stochastic processes, i.e. randomness which proceeds in time.
  - Will develop their mathematical theory (with a few applications).

1. Markov Chain Probabilities

- In these notes we will explore the mathematical foundations of *stochastic processes*, i.e. processes that proceed in time in a random way.
  - (For background and notation, see Appendix A.)
- One of the most natural and common and interesting types of stochastic processes are *Markov chains*, and we begin with them.

1.1. First Example: the Frog Walk

- Imagine 20 lily pads arranged in a circle.
  - Suppose a frog starts at pad #20.
  - Each second, it either stays where it is, or jumps one pad in the clockwise ↻ direction, or jumps one pad in the counter-clockwise ↺ direction, each with probability 1/3. (See Figure 1.)
  - (For an animated Java illustration, see: www.probability.ca/frogwalk.)
- This immediately leads to some direct computations. For example:
  - $\mathbf{P}$(at pad #1 after 1 step) $= 1/3$.

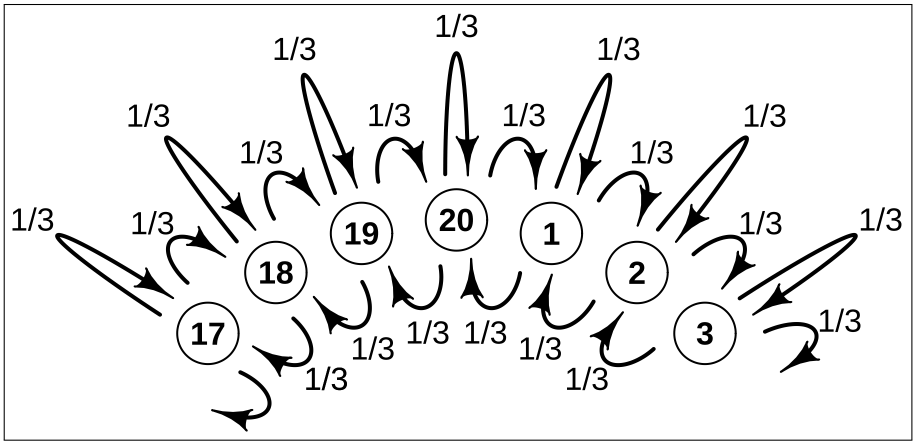


Figure 1: A diagram of part of the Frog Walk example.

- $\mathbf{P}$(at pad #20 after 1 step) = 1/3.
- $\mathbf{P}$(at pad #19 after 1 step) = 1/3.
- But $\mathbf{P}$(at pad #18 after 1 step) = 0, etc.

- Proceeding for two steps, we see that e.g. $\mathbf{P}$(at pad #2 after 2 steps) = $(1/3)(1/3) = 1/9$, since starting at pad #20, there is a 1/3 chance of jumping to pad #1, and then from pad #1 there is a further 1/3 chance of jumping to pad #2.
  - And, $\mathbf{P}$(at pad #19 after 2 steps) = $(1/3)(1/3) + (1/3)(1/3) = 2/9$, since the frog could *either* first jump from pad #20 to pad #19 and then stay there on the second jump with probability $(1/3)(1/3)$, *or* it could first stay at pad #20 and then jump from there to pad #19 again with probability $(1/3)(1/3)$.
  - And so on.
- Markov chain theory concerns itself largely with the question of what happens in the long run.
  - For example, what is $\mathbf{P}$(frog at pad #14 after 27 steps)?
  - Farther ahead, what is $\lim_{k\to\infty} \mathbf{P}$(frog at pad #14 after k steps)?
  - And, will the frog necessarily *eventually* return to pad #20?
  - And, will the frog necessarily *eventually* visit every pad?
- These questions will all be answered soon!

1.2. Markov Chain Basics

(1.2.1) Definition. A (discrete time, discrete space, time homogeneous) *Markov chain* is specified by three ingredients:

- A *state space* S, any non-empty finite or countable set.
- *initial probabilities* $\{\nu_i\}_{i\in S}$, where ν_i is the probability of starting at i (at time 0). (So, $\nu_i \geq 0$, and $\sum_i \nu_i = 1$.)
- *transition probabilities* $\{p_{ij}\}_{i,j\in S}$, where p_{ij} is the probability of jumping to j if you start at i. (So, $p_{ij} \geq 0$, and $\sum_j p_{ij} = 1$ for all i.)

- We will make frequent use of Definition (1.2.1).
- In the frog example:
  - State space: $S = \{1, 2, 3, \ldots, 20\}$. (All possible lily pads.)

- Initial distribution: $\nu_{20} = 1$, and $\nu_i = 0$ for all $i \neq 20$. (It starts at pad #20.)
- Transition probabilities:

$$p_{ij} \;=\; \begin{cases} 1/3, & |j-i| \leq 1 \\ 1/3, & |j-i| = 19 \\ 0, & \text{otherwise} \end{cases}$$

- For any Markov chain, let X_n be the Markov chain's state at time n.
- Then $\forall i \in S$, at time 0 we have $\mathbf{P}(X_0 = i) = \nu_i$.
- In terms of conditional probabilities (A.5.1), assuming $\mathbf{P}(X_n = i) > 0$, then $\forall i, j \in S$, and all $n = 0, 1, 2, \ldots$, $\mathbf{P}(X_{n+1} = j \,|\, X_n = i) = p_{ij}$. (This doesn't depend on n, because the chain is *time homogeneous*.)
- Also, $\mathbf{P}(X_{n+1} = j \,|\, X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) = \mathbf{P}(X_{n+1} = j \,|\, X_n = i_n) = p_{i_n j}$, i.e. the probabilities at time $n+1$ depend only on the state at time n (this is called the *Markov property*).
- Then for all $i, j \in S$, $\mathbf{P}(X_0 = i, X_1 = j) = \mathbf{P}(X_0 = i)\,\mathbf{P}(X_1 = j \,|\, X_0 = i) = \nu_i p_{ij}$.
- Similarly, $\mathbf{P}(X_0 = i, X_1 = j, X_2 = k) = \mathbf{P}(X_0 = i)\,\mathbf{P}(X_1 = j \,|\, X_0 = i)\,\mathbf{P}(X_2 = k \,|\, X_1 = j) = \nu_i p_{ij} p_{jk}$, etc.
- More generally,

$$\textbf{(1.2.2)} \qquad \mathbf{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) \;=\; \nu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}\,.$$

  - This completely defines the probabilities of the sequence $\{X_n\}_{n=0}^{\infty}$.
  - The random sequence $\{X_n\}_{n=0}^{\infty}$ "is" the Markov chain.
- In the frog example:
  - At time 0, $\mathbf{P}(X_0 = 20) = \nu_{20} = 1$, and $\mathbf{P}(X_0 = 7) = \nu_7 = 0$, etc.
  - Also, at time 1, $\mathbf{P}(X_1 = 1) = 1/3$, and $\mathbf{P}(X_1 = 20) = 1/3$, etc.
  - And, at time 2, our above results say that $\mathbf{P}(X_2 = 2) = (1/3)(1/3) = 1/9$, and $\mathbf{P}(X_2 = 19) = (1/3)(1/3) + (1/3)(1/3) = 2/9$, etc.
  - Using the formula (1.2.2), we have $\mathbf{P}(X_0 = 20,\ X_1 = 19,\ X_2 = 18) = \nu_{20} p_{20,19} p_{19,18} = (1)(1/3)(1/3) = 1/9$, etc.

(1.2.3) Problem. For the frog example:

(a) Compute $\mathbf{P}(X_0 = 20,\ X_1 = 19)$. **[sol]** (**Note:** Problems marked with **[sol]** have solutions provided at the end of these notes.)

(b) Compute $\mathbf{P}(X_0 = 20,\ X_1 = 19,\ X_2 = 20)$. **[sol]**

(c) Compute $\mathbf{P}(X_0 = 20,\ X_1 = 1,\ X_2 = 2,\ X_3 = 2)$.

(d) Compute $\mathbf{P}(X_0 = 20,\ X_1 = 20,\ X_2 = 19,\ X_3 = 20)$.

(1.2.4) Problem. Repeat Problem (1.2.3) for a modified frog example in which each second the frog jumps one pad clockwise ↻ with probability 1/2, or one pad counter-clockwise ↺ with probability 1/3, or stays where it is with probability 1/6.

1.3. More Examples of Markov Chains

- There are lots of different Markov chains. Here we present some more examples.

(1.3.1) Example. (Simple finite Markov chain)

– Let $S = \{1, 2, 3\}$, and $\nu = (1/7, 2/7, 4/7)$, and (see Figure 2)

$$(p_{ij}) = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}.$$

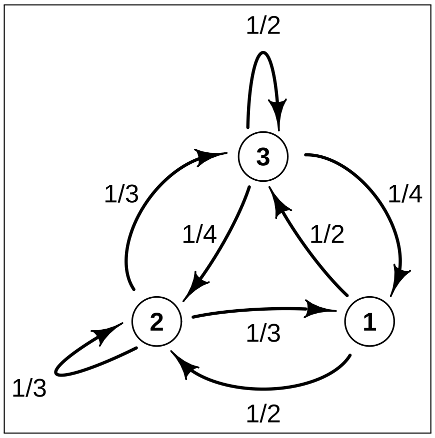


Figure 2: A diagram of Example (1.3.1).

– Then $\mathbf{P}(X_0 = 3,\ X_1 = 2) = \nu_3 p_{32} = (4/7)(1/4) = 1/7$, and $\mathbf{P}(X_0 = 2,\ X_1 = 1,\ X_2 = 3) = \nu_2 p_{21} p_{13} = (2/7)(1/3)(1/2) = 1/21$, etc.

– What happens in the long run, as $n \to \infty$?

– Is every state visited for sure? What is $\lim_{n\to\infty} \mathbf{P}(X_n = 3)$? etc.

(1.3.2) Problem. For Example (1.3.1):

(a) Compute $\mathbf{P}(X_0 = 1,\ X_1 = 1)$.

(b) Compute $\mathbf{P}(X_0 = 1,\ X_1 = 2)$.

(c) Compute $\mathbf{P}(X_0 = 1,\ X_1 = 1,\ X_2 = 1)$.

(d) Compute $\mathbf{P}(X_0 = 1,\ X_1 = 1,\ X_2 = 1,\ X_3 = 2)$.

(1.3.3) Problem. Consider a Markov chain with $S = \{1, 2, 3\}$, and $\nu = (1/3, 2/3, 0)$, and

$$(p_{ij}) = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/3 & 0 & 2/3 \\ 0 & 1/4 & 3/4 \end{pmatrix}.$$

(a) Draw a diagram of this Markov chain. [**sol**]

(b) Compute $\mathbf{P}(X_0 = 1)$.

(c) Compute $\mathbf{P}(X_0 = 1,\ X_1 = 1)$.

(d) Compute $\mathbf{P}(X_0 = 1,\ X_1 = 2)$.

(e) Compute $\mathbf{P}(X_0 = 1,\ X_1 = 1,\ X_2 = 1)$.

(f) Compute $\mathbf{P}(X_0 = 1,\ X_1 = 1,\ X_2 = 1,\ X_3 = 2)$.

(1.3.4) Bernoulli process. (Like counting sunny days.)

- Let $0 < p < 1$. (e.g. $p = 1/2$)
- Repeatedly flip a "p-coin" (i.e., a coin with probability of heads $= p$), at times $1, 2, 3, \ldots$
- Let $X_n = \#$ of heads on first n flips.
- Then $\{X_n\}$ is Markov chain, with $S = \{0, 1, 2, \ldots\}$, $X_0 = 0$ (i.e. $\nu_0 = 1$, and $\nu_i = 0$ for all $i \neq 0$), and

$$p_{ij} = \begin{cases} p, & j = i+1 \\ 1-p, & j = i \\ 0, & \text{otherwise} \end{cases}$$

(1.3.5) Problem. For the Bernoulli process with $p = 1/4$:

(a) Compute $\mathbf{P}(X_0 = 0,\ X_1 = 1)$.

(b) Compute $\mathbf{P}(X_0 = 0,\ X_1 = 1,\ X_2 = 1)$.

(c) Compute $\mathbf{P}(X_0 = 0,\ X_1 = 1,\ X_2 = 1,\ X_3 = 2)$.

(1.3.6) Simple random walk (s.r.w.).

- Let $0 < p < 1$. (e.g. $p = 1/2$)
- Suppose you repeatedly bet \$1.
- Each time, you have probability p of winning \$1, and probability $1-p$ of losing \$1.
- Animated Java illustration: www.probability.ca/randwalk
- Let X_n be net gain (in dollars) after n bets.
- Then $\{X_n\}$ is a Markov chain, with $S = \mathbf{Z}$, and (see Figure 3)

$$p_{ij} = \begin{cases} p, & j = i+1 \\ 1-p, & j = i-1 \\ 0, & \text{otherwise} \end{cases}$$

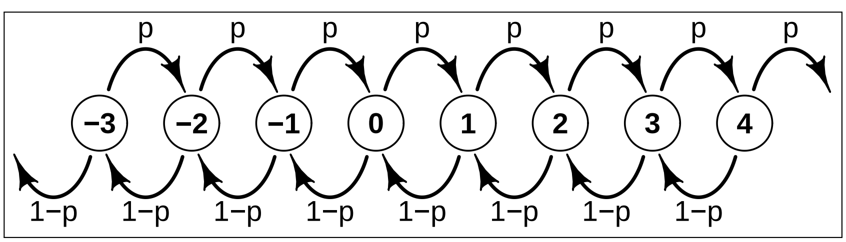


Figure 3: A diagram of part of Simple random walk (s.r.w.).

- For the initial value, usually we take $X_0 = 0$ so $\nu_0 = 1$. But sometimes we will instead take $X_0 = a$ for some other $a \in \mathbf{Z}$, so $\nu_a = 1$.
- What happens in the long run? Will you eventually lose \$1,000? etc.
- An important special case is when $p = 1/2$, called *simple symmetric random walk (s.s.r.w.)* since then $p = 1 - p$.

(1.3.7) Problem. For simple random walk with $p = 2/3$ and $X_0 = 0$:

(a) Compute $\mathbf{P}(X_0 = 0,\ X_1 = 0)$.

(b) Compute $\mathbf{P}(X_0 = 0,\ X_1 = 1)$.

(c) Compute $\mathbf{P}(X_0 = 0,\ X_1 = 1,\ X_2 = 0)$.

(d) Compute $\mathbf{P}(X_0 = 0,\ X_1 = 1,\ X_2 = 0,\ X_3 = -1)$.

(1.3.8) Ehrenfest's Urn.

- We have d balls in total, divided into two urns.
- At each time, we choose one of the d balls uniformly at random, and move it to the other urn.
- Let $X_n = \#$ balls in Urn 1 at time n (see Figure 4).

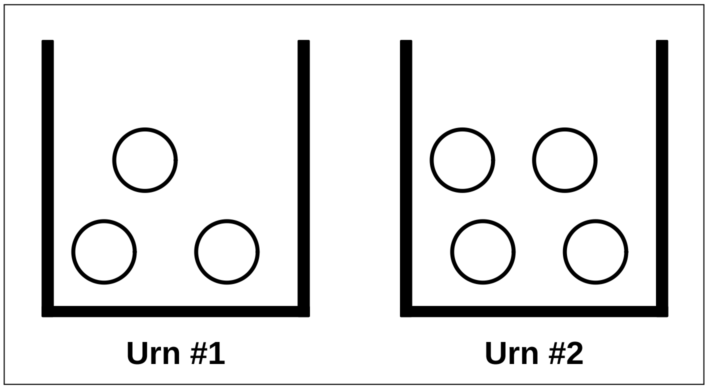


Figure 4: Ehrenfest's Urn with $d = 7$ balls, when $X_n = 3$.

- Then $\{X_n\}$ is a Markov chain, with $S = \{0, 1, 2, \ldots, d\}$, and transitions $p_{i,i-1} = i/d$, and $p_{i,i+1} = (d-i)/d$, with $p_{ij} = 0$ otherwise.
- (For an animated Java illustration, see: www.probability.ca/ehrenfest.)
- What happens in the long run? Does X_n become uniformly distributed? Does it stay close to X_0? to $d/2$?

(1.3.9) Problem. For Ehrenfest's Urn with $d = 5$ balls and $X_0 = 2$:

(a) Compute $\mathbf{P}(X_0 = 2,\ X_1 = 2)$.

(b) Compute $\mathbf{P}(X_0 = 2,\ X_1 = 1)$.

(c) Compute $\mathbf{P}(X_0 = 2,\ X_1 = 1,\ X_2 = 0)$.

(d) Compute $\mathbf{P}(X_0 = 2,\ X_1 = 1,\ X_2 = 0,\ X_3 = 1)$.

(1.3.10) Human Markov chain.

- Suppose each student takes out a coin (or borrows one).
- And, each student selects two other students, one for "heads" and one for "tails".
- To begin, the frog is tossed randomly to one student.
- Every time the frog comes to a student, that student catches it, and flips their coin, and tosses the frog to their student corresponding to the coin flip result.
- What happens in the long run?
- Will every student eventually get the frog?
- What about every student who is selected by someone else?
- What are the long-run probabilities? etc.

1.4. Elementary Computations

- Let $\{X_n\}$ be a Markov chain, with state space S, and transition probabilities p_{ij}, and initial probabilities ν_i.
- Recall that:
  - $\mathbf{P}(X_0 = i_0) = \nu_{i_0}$.
  - $\mathbf{P}(X_0 = i_0,\ X_1 = i_1) = \nu_{i_0} p_{i_0 i_1}$.
  - $\mathbf{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \nu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}$.
  - etc.
- Now, let $\mu_i^{(n)} = \mathbf{P}(X_n = i)$.
  - Then at time 0, $\mu_i^{(0)} = \nu_i$.
- At time 1, what is $\mu_j^{(1)}$ in terms of ν_i and p_{ij}?
  - Well, by the *Law of Total Probability* (A.2.8), $\mu_j^{(1)} = \mathbf{P}(X_1 = j) = \sum_{i \in S} \mathbf{P}(X_0 = i,\ X_1 = j) = \sum_{i \in S} \nu_i p_{ij} = \sum_{i \in S} \mu_i^{(0)} p_{ij}$.
- These equations can be written nicely with *matrix multiplication* (A.11.1).
  - Let $m = |S|$ be the number of elements in S (could be infinity).
  - Write $\nu = (\nu_1, \nu_2, \nu_3, \ldots)$ as a $1 \times m$ row vector.
  - And, write $\mu^{(n)} = (\mu_1^{(n)}, \mu_2^{(n)}, \mu_3^{(n)}, \ldots)$ as a $1 \times m$ row vector.
  - And, write $\mathbf{P} = (\, p_{ij}\,) = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & \vdots & \vdots & \ddots \end{pmatrix}$ as an $m \times m$ matrix.
  - Then in terms of matrix multiplication, $\mu^{(1)} = \nu P = \mu^{(0)} P$.
- Similarly, $\mu_k^{(2)} = \sum_{i \in S} \sum_{j \in S} \nu_i p_{ij} p_{jk}$, etc.
  - In matrix form: $\mu^{(2)} = \nu P P = \nu P^2 = \mu^{(0)} P^2$.
  - Then, by induction: $\mu^{(n)} = \nu P^n = \mu^{(0)} P^n$, for $n = 1, 2, 3, \ldots$.
  - By convention, let $P^0 = I$ be the identity matrix.
  - Then the formula $\mu^{(n)} = \nu P^n$ holds for $n = 0$, too.
- For Example (1.3.1), with $S = \{1, 2, 3\}$, and $\nu = (1/7, 2/7, 4/7)$, and

$$(\, p_{ij}\,) = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{pmatrix},$$

  we compute that $\mathbf{P}(X_1 = 2) = \mu_2^{(1)} =$
  $\sum_{i \in S} \nu_i p_{i2} =$
  $\nu_1 p_{12} + \nu_2 p_{22} + \nu_3 p_{32} =$
  $(1/7)(1/2) + (2/7)(1/3) + (4/7)(1/4) = 13/42$, etc.

(1.4.1) Problem. For Example (1.3.1):

(a) Compute $\mathbf{P}(X_1 = 3)$. **[sol]**

(b) Compute $\mathbf{P}(X_1 = 1)$.

(c) Compute $\mathbf{P}(X_2 = 1)$.

(d) Compute $\mathbf{P}(X_2 = 2)$.

(e) Compute $\mathbf{P}(X_2 = 3)$.

- Applying these formulas to the frog example, we have e.g. $\mathbf{P}(X_2 = 19) =$

$\mu_{19}^{(2)} =$
$\sum_{i \in S} \sum_{j \in S} \nu_i p_{ij} p_{j,19} =$
$\nu_{20} p_{20,19} p_{19,19} + \nu_{20} p_{20,20} p_{20,19} + 0 =$
$(1)(1/3)(1/3) + (1)(1/3)(1/3) + 0 = 2/9$, just like before.

(1.4.2) Problem. For the frog example:

(a) Compute $\mathbf{P}(X_1 = 1)$.

(b) Compute $\mathbf{P}(X_1 = 20)$.

(c) Compute $\mathbf{P}(X_2 = 20)$.

(d) Compute $\mathbf{P}(X_2 = 1)$.

(e) Compute $\mathbf{P}(X_2 = 2)$.

(f) Compute $\mathbf{P}(X_2 = 3)$.

(1.4.3) Problem. For the frog example, show that $\mathbf{P}(X_3 = 1) = 6/27$.

- Another way to track the probabilities of a Markov chain is with *n-step transitions*: $p_{ij}^{(n)} = \mathbf{P}(X_n = j \,|\, X_0 = i)$.
- (Since the chain is *time-homogeneous*, this must be the same as $p_{ij}^{(n)} = \mathbf{P}(X_{m+n} = j \,|\, X_m = i)$ for any $m \in \mathbf{N}$.)
- We must again have $p_{ij}^{(n)} \geq 0$, and $\sum_{j \in S} p_{ij}^{(n)} = \sum_{j \in S} \mathbf{P}_i(X_n = j) = \mathbf{P}_i(X_n \in S) = 1$.
- Here $p_{ij}^{(1)} = p_{ij}$. (of course)
- But what about $p_{ij}^{(2)}$?
  - Well, $p_{ij}^{(2)} = \mathbf{P}(X_2 = j \,|\, X_0 = i) =$
    $\sum_{k \in S} \mathbf{P}(X_2 = j,\ X_1 = k \,|\, X_0 = i) =$
    $\sum_{k \in S} p_{ik} p_{kj}$.
  - In matrix form: $P^{(2)} = \left(p_{ij}^{(2)} \right) = P\,P = P^2$.
- Similarly, $p_{ij}^{(3)} =$
  $\sum_{k \in S} \sum_{\ell \in S} p_{ik} p_{k\ell} p_{\ell j}$, i.e. $P^{(3)} = P^3$.
- By induction, $P^{(n)} = P^n$ for all $n \in \mathbf{N}$. That is, to compute probabilities of n-step jumps, you can take the n^{th} power of the transition matrix P.
- Convention: $P^{(0)} = I =$ identity matrix, i.e. $p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & \text{otherwise} \end{cases}$
  - Then $P^{(0)} = P^0$, too.
- Next, observe that by the *Law of Total Probability* (A.2.8) again, $p_{ij}^{(m+n)} =$
  $\mathbf{P}(X_{m+n} = j \,|\, X_0 = i) =$
  $\sum_{k \in S} \mathbf{P}(X_{m+n} = j,\ X_m = k \,|\, X_0 = i) =$
  $\sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}$.
- Similarly, $p_{ij}^{(m+s+n)} = \mathbf{P}(X_{m+s+n} = j \,|\, X_0 = i) =$
  $\sum_{k \in S} \sum_{\ell \in S} \mathbf{P}(X_{m+s+n} = j,\ X_m = k,\ X_{m+s} = \ell \,|\, X_0 = i) =$
  $\sum_{k \in S} \sum_{\ell \in S} p_{ik}^{(m)} p_{k\ell}^{(s)} p_{\ell j}^{(n)}$, etc. That is:

(1.4.4) Chapman-Kolmogorov equations. $p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}$, and $p_{ij}^{(m+s+n)} = \sum_{k \in S} \sum_{\ell \in S} p_{ik}^{(m)} p_{k\ell}^{(s)} p_{\ell j}^{(n)}$, etc.

- In matrix form, this says: $P^{(m+n)} = P^{(m)}P^{(n)}$, $P^{(m+s+n)} = P^{(m)}P^{(s)}P^{(n)}$, etc.
  - (Which is obvious, since e.g. $P^{(m+n)} = P^{m+n} = P^m P^n = P^{(m)}P^{(n)}$.)
- Then, since all probabilities are ≥ 0, it immediately follows that:

(1.4.5) Chapman-Kolmogorov Inequality. $p_{ij}^{(m+n)} \geq p_{ik}^{(m)} p_{kj}^{(n)}$ for any one fixed state $k \in S$, and $p_{ij}^{(m+s+n)} \geq p_{ik}^{(m)} p_{k\ell}^{(s)} p_{\ell j}^{(n)}$ for fixed $k, \ell \in S$, etc.

(1.4.6) Problem. Consider a Markov chain with state space $S = \{1, 2\}$, and transition probabilities $p_{11} = 2/3$, $p_{12} = 1/3$, $p_{21} = 1/4$, and $p_{22} = 3/4$.

(a) Draw a diagram of this Markov chain. **[sol]**

(b) Compute $p_{12}^{(2)}$. **[sol]**

(c) Compute $p_{12}^{(3)}$.

(1.4.7) Problem. Suppose a fair six-sided die is repeatedly rolled, at times $0, 1, 2, 3, \ldots$. (So, each roll is independently equally likely to be 1, 2, 3, 4, 5, or 6.) Let $X_0 = 0$, and for $n \geq 1$ let X_n be the *largest* value that appears among all of the rolls up to time n.

(a) Find (with justification) a state space S, initial probabilities $\{\nu_i\}$, and transition probabilities $\{p_{ij}\}$, for which $\{X_n\}$ is Markov chain.

(b) Compute the two-step transition probabilities $\{p_{35}^{(2)}\}$ and $\{p_{15}^{(2)}\}$.

(c) Compute the three-step transition probability $\{p_{15}^{(3)}\}$.

1.5. Recurrence and Transience

- Shorthand: write $\mathbf{P}_i(\cdots)$ for $\mathbf{P}(\cdots \mid X_0 = i)$. And, write $\mathbf{E}_i(\cdots)$ for $\mathbf{E}(\cdots \mid X_0 = i)$.
- Let $N(i) = \#\{n \geq 1 : X_n = i) =$ total # times that the chain hits i (not counting time 0).
  - (Note: $N(i)$ is a random variable, which could be infinite, cf. Section A.4.)
- Let f_{ij} be the *return probability* from i to j.
  - Formally, f_{ij} is defined as $f_{ij} := \mathbf{P}_i(X_n = j \text{ for } \underline{\text{some}}\ n \geq 1)$.
  - That is, f_{ij} is the probability, starting from i, that the chain will eventually visit j at least once.
  - Equivalently, $f_{ij} = \mathbf{P}_i(N(j) \geq 1)$.
- By stringing different paths together, it follows that e.g. $\mathbf{P}_i$(chain will eventually visit j, and then eventually visit k) $= f_{ij} f_{jk}$, etc.
  - Hence, in particular, $\mathbf{P}_i(N(i) \geq 2) = (f_{ii})^2$, and $\mathbf{P}_i(N(i) \geq 3) = (f_{ii})^3$, etc.
  - In general, for $k = 0, 1, 2, \ldots$,

$$\text{(1.5.1)} \qquad \mathbf{P}_i(N(i) \geq k) \;=\; (f_{ii})^k .$$

  - And similarly,

$$\text{(1.5.2)} \qquad \mathbf{P}_i(N(j) \geq k) \;=\; f_{ij}(f_{jj})^{k-1} .$$

– It also follows that $f_{ik} \geq f_{ij} f_{jk}$, etc.

Considering the complementary probabilities gives:

(1.5.3) $1 - f_{ij} \;=\; \mathbf{P}_i(X_n \neq j \text{ for } \underline{\text{all}}\ n \geq 1)$.

A very important definition is:

(1.5.4) Definition. A state i of a Markov chain is *recurrent* (or, *persistent*) if $\mathbf{P}_i(X_n = i \text{ for some } n \geq 1) = 1$, i.e. if $f_{ii} = 1$.

– Otherwise, if $f_{ii} < 1$, then i is *transient.*

- (Is the Frog Example recurrent? Is s.r.w.? Other previous examples?)
- The following result relates recurrence of a state i, to the question of whether the chain will return to i infinitely often, and also to whether a certain sum is finite or infinite (for background on infinite series, see Subsection A.3).

(1.5.5) Recurrent State Theorem.

– State i is recurrent <u>iff</u> $\mathbf{P}_i(N(i) = \infty) = 1$ <u>iff</u> $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.

– And, i is transient <u>iff</u> $\mathbf{P}_i(N(i) = \infty) = 0$ <u>iff</u> $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$.

Proof. By continuity of probabilities (A.7.3) and then (1.5.1),

$$\mathbf{P}_i(N(i) = \infty) \;=\; \lim_{k\to\infty} \mathbf{P}_i(N(i) \geq k) \;=\; \lim_{k\to\infty} (f_{ii})^k \;=\; \begin{cases} 1, & f_{ii} = 1 \\ 0, & f_{ii} < 1 \end{cases}$$

Then, using countable linearity for non-negative random variables (A.8.2), and then the trick (A.2.6) that if Z is non-negative-integer-valued then

$$\mathbf{E}(Z) \;=\; \sum_{k=1}^{\infty} \mathbf{P}(Z \geq k)\,,$$

we have

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} \;=\; \sum_{n=1}^{\infty} \mathbf{P}_i(X_n = i) \;=\; \sum_{n=1}^{\infty} \mathbf{E}_i(\mathbf{1}_{X_n = i})$$

$$= \mathbf{E}_i\left(\sum_{n=1}^{\infty} \mathbf{1}_{X_n = i}\right) \;=\; \mathbf{E}_i\left(N(i)\right) \;=\; \sum_{k=1}^{\infty} \mathbf{P}_i\left(N(i) \geq k\right)$$

$$= \sum_{k=1}^{\infty} (f_{ii})^k \;=\; \begin{cases} \infty, & f_{ii} = 1 \\ \frac{f_{ii}}{1 - f_{ii}} < \infty, & f_{ii} < 1 \end{cases}$$

∎

- Note: The above proof shows that, even though f_{ii} is a very different concept from $p_{ii}^{(n)}$, the surprising identity $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \sum_{k=1}^{\infty} (f_{ii})^k$ holds.
- What about simple random walk (1.3.6)?
  – Is the state 0 recurrent in that case?
  – Use the Recurrent State Theorem! Check if $\sum_{n=1}^{\infty} p_{00}^{(n)} = \infty$, or not.
  – Well, if n is odd, then $p_{00}^{(n)} = 0$.
  – If n is even, $p_{00}^{(n)} = \mathbf{P}(n/2 \text{ heads and } n/2 \text{ tails on first } n \text{ tosses})$.

– This is the Binomial(n,p) distribution (A.2.2), so

$$p_{00}^{(n)} = \binom{n}{n/2} p^{n/2}(1-p)^{n/2} = \frac{n!}{[(n/2)!]^2} p^{n/2}(1-p)^{n/2}.$$

– But Stirling's approximation (A.12.3) says that if n is large, then $n! \approx (n/e)^n\sqrt{2\pi n}$. Hence, also $(n/2)! \approx (n/2e)^{n/2}\sqrt{2\pi n/2}$.

– So, for n large and even,

(1.5.6)
$$p_{00}^{(n)} \approx \frac{(n/e)^n\sqrt{2\pi n}}{[(n/2e)^{n/2}\sqrt{2\pi n/2}]^2}\, p^{n/2}(1-p)^{n/2}$$
$$= [4p(1-p)]^{n/2}\sqrt{2/\pi n}\,.$$

– Now, if $p = 1/2$, then $4p(1-p) = 1$, so $\sum_{n=1}^{\infty} p_{00}^{(n)} \approx \sum_{n=2,4,6,\ldots} \sqrt{2/\pi n}$, which is infinite by (A.3.2) with $a = 1/2$, so state 0 is <u>recurrent</u>, and the chain will return to state 0 infinitely often with probability 1.

– But if $p \neq 1/2$, then $4p(1-p) < 1$, so

$$\sum_{n=1}^{\infty} p_{00}^{(n)} \approx \sum_{n=2,4,6,\ldots} [4p(1-p)]^{n/2}\sqrt{2/\pi n}$$
$$< \sum_{n=2,4,6,\ldots} [4p(1-p)]^{n/2} = \frac{4p(1-p)}{1-4p(1-p)}$$

since it is a geometric series (A.3.1) with initial term $c = 4p(1-p)$ and common ratio $r = [4p(1-p)]^{1/2} < 1$, so state 0 is <u>transient</u>, and the chain will *not* return to state 0 infinitely often.

– Conclusion: the state 0 is recurrent (with $f_{00} = 1$) if $p = 1/2$, otherwise state 0 is transient (with $f_{00} < 1$).

– The same exact calculation applies to any other state i, so:

(1.5.7) For simple random walk, if $p = 1/2$ then $f_{ii} = 1\ \forall i$, but if $p \neq 1/2$ then $f_{ii} < 1\ \forall i$.

- What about f_{ij} for e.g. the Frog Example?
  Coming soon!

END OF WEEK #1

Recall: Markov chains, S, ν_i, p_{ij}, P, examples (Frog, Bernoulli, finite matrix, srw, Ehrenfest), $\mu_i^{(n)}$, $P^{(n)}$, Chapman-Kolmogorov, f_{ij}, recurrence/transience, Recurrent State Theorem.

- Computing actual values of the return probabilities f_{ij} can be challenging.

(1.5.8) Example. Let $S = \{1,2,3,4\}$, and (see Figure 5)

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 2/5 & 3/5 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}.$$

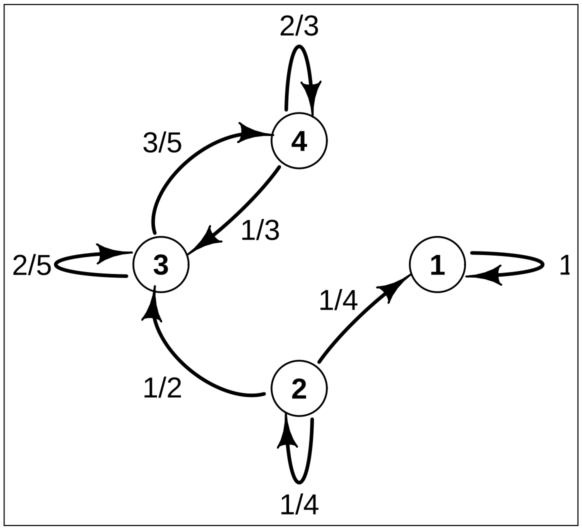


Figure 5: A diagram of Example (1.5.8).

- What are the various f_{ij} values?
- Here $f_{11} =$
  1 (of course).
- Also, $f_{22} =$
  $1/4$, since after leaving f it can never return.
- Also $f_{33} =$
  1, since e.g. $\mathbf{P}_3$(don't return to 3 in first n steps) $= (3/5)(2/3)^{n-1}$ which $\to 0$ as $n \to \infty$. (Alternatively, this can be shown by an infinite sum similar to Solution #2 below.)
- By similar reasoning, $f_{44} =$
  1.
- So, states 1, 3, and 4 are recurrent, but state 2 is transient.
- Also, clearly $f_{12} = f_{13} = f_{14} = f_{32} = f_{31} =$
  0, since e.g. from state 1 it can never get to state 2, etc.
- And, $f_{34} = f_{43} =$
  1 (similar to f_{33} above).
- What about f_{21}? Harder?

- To compute the return probabilities f_{ij}, the following result sometimes helps:

(1.5.9) f-Expansion. $f_{ij} = p_{ij} + \sum_{\substack{k \in S \\ k \neq j}} p_{ik} f_{kj}$

- The f-Expansion shows that $f_{ij} \geq p_{ij}$ (of course).
- It essentially follows from logical reasoning: from i, to get to j eventually, we have to either jump to j immediately (with probability p_{ij}), or jump to some other state k (with probability p_{ik}) and *then* get to j eventually (with probability f_{kj}).
- More formally, it follows by:

$$f_{ij} = \mathbf{P}_i(\exists n \geq 1 : X_n = j) = \sum_{k \in S} \mathbf{P}_i(X_1 = k,\ \exists n \geq 1 : X_n = j)$$

$$= \mathbf{P}_i(X_1 = j,\ \exists n \geq 1 : X_n = j) + \sum_{k \neq j} \mathbf{P}_i(X_1 = k,\ \exists n \geq 1 : X_n = j)$$

$$= p_{ij}(1) + \sum_{k \neq j} p_{ik}(f_{kj}) .$$

- (Note: If $i \neq j$, the sum *does* include the term where $k = i$.)

- Returning to Example (1.5.8), what is f_{21}?
  - Solution #1: By the f-Expansion (1.5.9), $f_{21} = p_{21} + p_{22} f_{21} + p_{23} f_{31} + p_{24} f_{41} = (1/4) + (1/4) f_{21} + (1/2)(0) + (0)(0)$. So, $f_{21} = (1/4) + (1/4) f_{21}$, so $(3/4) f_{21} = (1/4)$, so $f_{21} = (1/4) \,/\, (3/4) = 1/3$.
  - Solution #2: Let τ be the first time we hit 1. Then we compute that $f_{21} = \mathbf{P}_2[\tau < \infty] = \sum_{m=1}^{\infty} \mathbf{P}_2[\tau = m] = \sum_{m=1}^{\infty} (1/4)^{m-1}(1/4) = \sum_{m=1}^{\infty} (1/4)^m = (1/4)/[1 - (1/4)] = 1/3$.
  - Solution #3: In this special case only, $f_{21} = \mathbf{P}_2(X_1 = 1 \,|\, X_1 \neq 2) = \mathbf{P}_2(X_1 = 1) \,/\, \mathbf{P}_2(X_1 \neq 2) = (1/4)/[(1/4) + (1/2)] = 1/3$.

(1.5.10) Problem. In Example (1.5.8):

(a) Show that $f_{23} = 2/3$, in three different ways.

(b) Show that $f_{24} = 2/3$.

(1.5.11) Problem. Consider a (discrete-time) Markov chain $\{X_n\}$ on the state space $S = \{1, 2, 3, 4\}$, with transition probabilities

$$(\, p_{ij} \,) \;=\; \begin{pmatrix} 1/3 & 1/6 & 1/2 & 0 \\ 1/4 & 0 & 3/4 & 0 \\ 2/5 & 3/5 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix} .$$

(a) Compute the value of $p_{43}^{(3)} \equiv \mathbf{P}(X_3 = 3 \,|\, X_0 = 4)$.

(b) Compute f_{23}.

(c) Compute f_{21}.

(1.5.12) Problem. For each of the following sets of conditions, either provide (with explanation) an example of Markov chain transition probabilities $\{p_{ij}\}$ on some state space S such that the conditions are satisfied, or prove that no such Markov chain exists.

(a) $3/4 < p_{12}^{(n)} < 1$ for all $n \geq 1$.

(b) $p_{11} > 1/2$, and the state 1 is transient.

(c) $p_{12} = 0$ and $p_{12}^{(3)} = 0$, but $0 < p_{12}^{(2)} < 1$.

(d) $f_{12} = 1/2$, and $f_{13} = 2/3$.

(e) $p_{12}^{(n)} \geq 1/4$ and $p_{21}^{(n)} \geq 1/4$ for all $n \geq 1$, and the state 1 is transient.

(f) S is finite, and $\lim_{n \to \infty} p_{ij}^{(n)} = 0$ for all $i, j \in S$.

(g) S is infinite, and $\lim_{n \to \infty} p_{ij}^{(n)} = 0$ for all $i, j \in S$.

1.6. Communicating States

(1.6.1) Definition. Say that state i *communicates* with state j, written $i \to j$, if $f_{ij} > 0$, i.e. if it is possible to get from i to j.

- An alternative formulation, which follows easily from countable additivity of probabilities (A.2.7), is:

(1.6.2) $f_{ij} > 0$ iff $\exists\, m \geq 1$ with $p_{ij}^{(m)} > 0$, i.e. there is some time m for which it is possible to get from i to j in m steps.

- We will write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$.

(1.6.3) Definition. A Markov chain is *irreducible* if $i \to j$ for all $i, j \in S$, i.e. if $f_{ij} > 0$ for all $i, j \in S$. Otherwise, the chain is *reducible*.

- (Check: Are our previous examples irreducible? Yes for e.g. the Frog Example, and Simple Random Walk, and Ehrenfest's Urn.)
- Next, recall the condition $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ from the Recurrent State Theorem (1.5.5). To check it, the following lemma sometimes helps.

(1.6.4) Sum Lemma. If $i \to k$, and $\ell \to j$, and $\sum_{n=1}^{\infty} p_{k\ell}^{(n)} = \infty$, then $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$.

Proof. By (1.6.2), find $m, r \geq 1$ with $p_{ik}^{(m)} > 0$ and $p_{\ell j}^{(r)} > 0$.

- By the Chapman-Kolmogorov inequality (1.4.5), $p_{ij}^{(m+s+r)} \geq p_{ik}^{(m)} p_{k\ell}^{(s)} p_{\ell j}^{(r)}$.
- Hence, since each $p_{ij}^{(n)} \geq 0$,

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} \;\geq\; \sum_{n=m+r+1}^{\infty} p_{ij}^{(n)} \;=\; \sum_{s=1}^{\infty} p_{ij}^{(m+s+r)} \;\geq\; \sum_{s=1}^{\infty} p_{ik}^{(m)} p_{k\ell}^{(s)} p_{\ell j}^{(r)}$$

$$= \; p_{ik}^{(m)} p_{\ell j}^{(r)} \sum_{s=1}^{\infty} p_{k\ell}^{(s)} \;=\; (positive)(positive)(\infty) \;=\; \infty\,. \qquad \blacksquare$$

- Setting $j = i$ and $\ell = k$ in the Sum Lemma says that if $i \leftrightarrow k$, then $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ <u>iff</u> $\sum_{n=1}^{\infty} p_{kk}^{(n)} = \infty$.
  - Combining that fact with the Recurrent State Theorem (1.5.5) says:

(1.6.5) Sum Corollary. If $i \leftrightarrow k$, then i is recurrent <u>iff</u> k is recurrent.

- Then, applying this Sum Corollary to an *irreducible* chain (where all states communicate) says:

(1.6.6) Cases Theorem. For an <u>irreducible</u> Markov chain, either
(a) $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ for all $i, j \in S$, and <u>all</u> states are recurrent.
("*recurrent Markov chain*")
or (b) $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all $i, j \in S$, and <u>all</u> states are transient.
("*transient Markov chain*")

- So what about simple random walk (1.3.6)?
  - Is it irreducible?
    Yes!
  - So, the Cases Theorem applies. But which case?
  - We have already seen (1.5.7) that if $p = 1/2$ it's in case (a), so all states are recurrent, and $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ for all $i, j \in S$.

- Or, if $p \neq 1/2$, then it's in case (b), so all states are transient, and $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all $i, j \in S$.

- What about Frog Example?
  - Is it irreducible?
    Yes!
  - So, again the Cases Theorem applies. But which case?
  - The answer is given by:

(1.6.7) Finite Space Theorem. An irreducible Markov chain on a finite state space always falls into case (a), i.e. $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ for all $i, j \in S$, and all states are recurrent.

Proof.

- Choose any state $i \in S$. Then by exchanging the sums (which is valid by (A.8.3) since $p_{ij}^{(n)} \geq 0$), and recalling that $\sum_{j \in S} p_{ij}^{(n)} = 1$, we have

$$\sum_{j \in S} \sum_{n=1}^{\infty} p_{ij}^{(n)} = \sum_{n=1}^{\infty} \sum_{j \in S} p_{ij}^{(n)} = \sum_{n=1}^{\infty} 1 = \infty .$$

- This equation holds whether or not S is finite.
- But if S is finite, then it follows that there must be at least one $j \in S$ with $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$.
- So, we must be in case (a). ∎

- Note: In the above proof, the equation $\sum_{j \in S} \sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ always holds, whether S is finite or infinite. But it is only informative when S is finite.

- The Frog Example has $|S| = 20$, which is finite.
  - So, the Frog Example is in case (a)!
  - So, all states are recurrent.
  - So, in the Frog Example, $\mathbf{P}_{20}(\exists n \geq 1 \text{ with } X_n = 20) = 1$, etc.
  - But what about e.g. $f_{20,14} = \mathbf{P}_{20}(\exists n \geq 1 \text{ with } X_n = 14)$? Unclear!
- To solve this, for $i \neq j$ let H_{ij} be the event that the chain hits the state i before returning to j, i.e.

$$H_{ij} = \{\exists n \in \mathbf{N} : X_n = i, \text{ but } X_m \neq j \text{ for } 1 \leq m \leq n-1\} .$$

(1.6.8) Hit Lemma. If $j \to i$ with $j \neq i$, then $\mathbf{P}_j(H_{ij}) > 0$. (That is: If it is possible to get from j to i at all, then it is possible to get from j to i without first returning to j.)

- The Hit Lemma is intuitively obvious: If there is some path from j to i, then the final part of the path (starting with the last time it visits i) is a possible path from j to i which does not return to i.
- For a more formal proof (optional):
  - Since $j \to i$, there is some possible path from j to i, i.e. there is $m \in \mathbf{N}$ and $x_0, x_1, \ldots, x_m$ with $x_0 = j$ and $x_m = i$ and $p_{x_r x_{r+1}} > 0$ for all $0 \leq r \leq m-1$.

- Let $S = \max\{r : x_r = j\}$ be the <u>last</u> time this path hits j. (So, we might have $S = 0$, but might have $S > 0$.)
- Then $x_S, x_{S+1}, \ldots, x_m$ is a possible path which goes from j to i without first returning to j.
- So, $\mathbf{P}_j(H_{ij}) \geq \mathbf{P}_j(\text{this path}) = p_{x_S x_{S+1}} p_{x_{S+1} x_{S+2}} \ldots p_{x_{m-1} x_m} > 0$, ∎

- Using the Hit Lemma and (1.5.3), we can prove the important:

(1.6.9) f-Lemma. If $j \to i$ and $f_{jj} = 1$, then $f_{ij} = 1$.

Proof.

- If $i = j$ it's trivial, so assume $i \neq j$.
- Since $j \to i$, we have $\mathbf{P}_j(H_{ij}) > 0$ by the Hit Lemma.
- But one way to never return to j, is to first hit i, and then from i never return to j.
- That is, $\mathbf{P}_j(\text{never return to } j) \geq \mathbf{P}_j(H_{ij})\, \mathbf{P}_i(\text{never return to } j)$.
- By (1.5.3), this means $1 - f_{jj} \geq \mathbf{P}_j(H_{ij})\,(1 - f_{ij})$.
- But if $f_{jj} = 1$, then $1 - f_{jj} = 0$, so $\mathbf{P}_j(H_{ij})\,(1 - f_{ij}) = 0$.
- Since $\mathbf{P}_j(H_{ij}) > 0$, we must have $1 - f_{ij} = 0$, i.e. $f_{ij} = 1$. ∎

- This also allows us to prove:

(1.6.10) Infinite Returns Lemma. For an irreducible Markov chain, if it is recurrent then $\mathbf{P}_i(N(j) = \infty) = 1$ for all $i, j \in S$, but if it is transient then $\mathbf{P}_i(N(j) = \infty) = 0$ for all $i, j \in S$.

Proof. If the chain is recurrent, then $f_{ij} = f_{jj} = 1$ by the f-Lemma.

- Hence, similar to the proof of the Recurrent State Theorem (1.5.5), again using continuity of probabilities (A.7.3) and then the formula (1.5.2) for $\mathbf{P}_i(N(j) \geq k)$, we have

$$\begin{aligned} \mathbf{P}_i(N(j) = \infty) &= \lim_{k\to\infty} \mathbf{P}_i(N(j) \geq k) \\ &= \lim_{k\to\infty} f_{ij}(f_{jj})^{k-1} \;=\; \lim_{k\to\infty} (1)(1)^{k-1} \;=\; 1\,. \end{aligned}$$

- Or, if instead the chain is transient, then $f_{jj} < 1$, so instead $\mathbf{P}_i(N(j) = \infty) = \lim_{k\to\infty} f_{ij}(f_{jj})^{k-1} = 0$. ∎

- Putting all of the above together, we obtain:

(1.6.11) Recurrence Equivalences Theorem. If a chain is irreducible, then the following are equivalent (and all correspond to "case (a)"):

(1) There are $k, \ell \in S$ with $\sum_{n=1}^{\infty} p_{k\ell}^{(n)} = \infty$.
(2) For all $i, j \in S$, we have $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$.
(3) There is $k \in S$ with $f_{kk} = 1$, i.e. with k recurrent.
(4) For all $j \in S$, we have $f_{jj} = 1$, i.e. all states are recurrent.
(5) For all $i, j \in S$, we have $f_{ij} = 1$.
(6) There are $k, \ell \in S$ with $\mathbf{P}_k(N(\ell) = \infty) = 1$.
(7) For all $i, j \in S$, we have $\mathbf{P}_i(N(j) = \infty) = 1$.

Proof. All of the necessary implications follow from results that we have already proven:

- (1) ⇒ (2): Sum Lemma.
- (2) ⇒ (4): Recurrent State Theorem (with $i = j$).
- (4) ⇒ (5): f-Lemma.
- (5) ⇒ (3): Immediate.
- (3) ⇒ (1): Recurrent State Theorem (with $\ell = k$).
- (4) ⇒ (7): Infinite Returns Lemma.
- (7) ⇒ (6): Immediate.
- (6) ⇒ (3): Infinite Returns Lemma. ∎

- Or, considering the opposites (and (1.6.10)), we obtain:

(1.6.12) Transience Equivalences Theorem. If chain irreducible, then the following are equivalent (and all correspond to "case (b)"):

(1) For all $k, \ell \in S$, $\sum_{n=1}^{\infty} p_{k\ell}^{(n)} < \infty$.
(2) There is $i, j \in S$ with $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$.
(3) For all $k \in S$, $f_{kk} < 1$, i.e. k is transient.
(4) There is $j \in S$ with $f_{jj} < 1$, i.e. some state is transient.
(5) There are $i, j \in S$ with $f_{ij} < 1$.
(6) For all $k, \ell \in S$, $\mathbf{P}_k(N(\ell) = \infty) = 0$.
(7) There are $i, j \in S$ with $\mathbf{P}_i(N(j) = \infty) = 0$.

- What about the Frog Example?
  - The Finite Space Theorem (1.6.7) says it's in case (a).
  - Then, the Recurrence Equivalences Theorem (1.6.11) says $f_{ij} = 1$ for all $i, j \in S$.
  - So, $\mathbf{P}(\exists\, n \geq 1 \text{ with } X_n = 14 \,|\, X_0 = 20) = f_{20,14} = 1$, etc.
- What about simple symmetric random walk (i.e. with $p = 1/2$)?
  - We showed (1.5.7) that it's in case (a).
  - So, the Recurrence Equivalences Theorem (1.6.11) says $f_{ij} = 1$ for all $i, j \in S$.
  - Hence, $\mathbf{P}(\exists n \geq 1 \text{ with } X_n = 1,000,000 \,|\, X_0 = 0) = 1$, etc.
  - More generally, for any sequence $i_1, i_2, i_3, \ldots \in S$,

(1.6.13)
$$f_{i_1 i_2} f_{i_2 i_3} f_{i_3 i_4} \ldots \;=\; 1\,.$$

  - This means that for any conceivable pattern of values, the chain will certainly eventually hit each of them, in sequence.
  - That is, the chain values have infinite *fluctuations*.

- Example: Let $S = \{1, 2, 3\}$, and $(\,p_{ij}\,) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.
  - Then of course $f_{22} = 1$ (since $p_{22} = 1$), so state 2 is recurrent.
  - Also $f_{11} =$

0 < 1, so state 1 is transient.

- Also, for all $n \in \mathbf{N}$, $p_{12}^{(n)} = 1/2$.
- So, $\sum_{n=1}^{\infty} p_{12}^{(n)} = \sum_{n=1}^{\infty}(1/2) = \infty$.
- Also, for all $n \in \mathbf{N}$, $p_{21}^{(n)} = 0$.
- So, $\sum_{n=1}^{\infty} p_{12}^{(n)} = \sum_{n=1}^{\infty} 0 = 0 < \infty$.
- Also, $f_{12} = 1/2 < 1$.
- Summary: Some states are recurrent, and some transient.
- And, sometimes $f_{ij} = 1$, but sometimes $f_{ij} < 1$.
- And, sometimes $\sum_n p_{ij}^{(n)} = \infty$, but sometimes $\sum_n p_{ij}^{(n)} < \infty$.
- Contradictions to the Recurrence Equivalence Theorem?
- No, since the chain is reducible (not irreducible)!

- One observation which sometimes helps with reducible chains is:

(1.6.14) Closed Subset Note. Suppose a chain is reducible, but it has a *closed subset* $C \subseteq S$ (i.e., $p_{ij} = 0$ for $i \in C$ and $j \notin C$) on which it is irreducible (i.e., $i \to j$ for all $i, j \in C$). Then, the Recurrence Equivalences Theorem, and other results about irreducible chains, still apply to the chain when *restricted* to C.

- To illustrate this, consider again Example (1.5.8) with

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 2/5 & 3/5 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}.$$

  - Is this chain irreducible? No!
  - But it *is* irreducible on the closed subset $C = \{3, 4\}$.
  - Since C is finite, the Finite Space Theorem (1.6.7) says that the chain restricted to C is in case (a).
  - This immediately proves that states 3 and 4 are recurrent, and that $f_{34} = f_{43} = 1$, and $\sum_{n=1}^{\infty} p_{43}^{(n)} = \infty$, etc.
  - Easier than before!

- Now, consider simple random walk with $p > 1/2$ (e.g. $p = 0.51$).
  - Irreducible? Yes!
  - Is $f_{00} = 1$? No, $f_{00} < 1$. (transient) Similarly $f_{55} < 1$, etc.
  - And, by the Cases Theorem (1.6.6), $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all $i, j \in S$.
  - On the other hand, surprisingly, we still have e.g. $f_{05} = 1$:

(1.6.15) Proposition. For simple random walk with $p > 1/2$, $f_{ij} = 1$ whenever $j > i$. (Similarly, if $p < 1/2$ and $j < i$, then $f_{ij} = 1$.)

Proof. Let $X_0 = 0$, and $Z_n = X_n - X_{n-1}$ for $n = 1, 2, \ldots$.

- Then, by construction, $X_n = \sum_{i=1}^n Z_i$.
- The $\{Z_n\}$ are i.i.d., with $\mathbf{P}(Z_n = +1) = p$ and $\mathbf{P}(Z_n = -1) = 1 - p$.
- So, by the *Law of Large Numbers* (A.6.2), w.p. 1, $\lim_{n\to\infty} \frac{1}{n}(Z_1 + Z_2 + \ldots + Z_n) = \mathbf{E}(Z_1) = p(1) + (1-p)(-1) = 2p - 1 > 0$.
- So, w.p. 1, $\lim_{n\to\infty}(Z_1 + Z_2 + \ldots + Z_n) = +\infty$.
- So, w.p. 1, $X_n - X_0 \to \infty$, i.e. $X_n \to \infty$.
- So, starting from i, the chain will converge to ∞,
- But if $i < j$, then to go from i to ∞, the chain must pass through j.
- This will happen with probability 1, so $f_{ij} = 1$. ∎

- For example, if $p = 0.51$ and $i = 0$ and $j = 5$, then $f_{05} = 1$.
  - Does this contradict the Recurrence Equivalences Theorem?
  - No! Could still have $\sum_{n=1}^\infty p_{05}^{(n)} < \infty$, etc.
  - This is why the Recurrence Equivalences Theorem does not include the equivalence "There are $k, \ell \in S$ with $f_{k\ell} = 1$".

(1.6.16) Problem. Consider the Markov chain from Problem (1.4.6), with state space $S = \{1, 2\}$, and transition probabilities $p_{11} = 2/3$, $p_{12} = 1/3$, $p_{21} = 1/4$, and $p_{22} = 3/4$.

(a) Determine whether or not this chain is irreducible. [sol]

(b) Determine whether or not $f_{11} = 1$. [sol]

(c) Determine whether or not $f_{21} = 1$. [sol]

(d) Determine whether or not $\sum_{n=1}^\infty p_{11}^{(n)} = \infty$. [sol]

(e) Determine whether or not $\sum_{n=1}^\infty p_{21}^{(n)} = \infty$. [sol]

(1.6.17) Problem. Let $S = \{1, 2, 3\}$, with transition probabilities defined by $p_{11} = p_{12} = p_{22} = p_{23} = p_{32} = p_{33} = 1/2$, with $p_{ij} = 0$ otherwise.

(a) Is this chain irreducible? [sol]

(b) Compute f_{ii} for each $i \in S$. [sol]

(c) Specify which states are recurrent, and which states are transient. [sol]

(d) Compute the value of f_{13}. [sol]

(1.6.18) Problem. For each of the following sets of conditions, either provide (with explanation) an example of a state space S and Markov chain transition probabilities $\{p_{ij}\}_{i,j\in S}$ such that the conditions are satisfied, or prove that no such Markov chain exists.

(a) The chain is irreducible and transient, and there are $k, \ell \in S$ with $p_{k\ell}^{(n)} \geq 1/3$ for all $n \in \mathbf{N}$. [sol]

(b) The chain is irreducible, and there are distinct states $i, j, k, \ell \in S$ such that $f_{ij} < 1$, and $\sum_{n=1}^\infty p_{k\ell}^{(n)} = \infty$. [sol]

(c) There are distinct states $k, \ell \in S$ such that if the chain is started at k, then there is a positive probability that the chain will visit ℓ exactly five times (and then never again). **[sol]**

(d) The chain is irreducible and transient, and there are $k, \ell \in S$ with $f_{k\ell} = 1$. **[sol]**

(1.6.19) Problem. Consider the chain from Problem (1.5.11), with $S = \{1, 2, 3, 4\}$ and

$$(p_{ij}) = \begin{pmatrix} 1/3 & 1/6 & 1/2 & 0 \\ 1/4 & 0 & 3/4 & 0 \\ 2/5 & 3/5 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix}.$$

Use (1.6.14) to compute f_{21} and f_{23} more easily.

(1.6.20) Problem. Consider a Markov chain with state space $S = \{1, 2, 3, 4\}$, and transition probabilities $p_{11} = p_{12} = 1/2$, $p_{21} = 1/3$, $p_{22} = 2/3$, $p_{32} = 1/7$, $p_{33} = 2/7$, $p_{34} = 4/7$, $p_{44} = 1$, with $p_{ij} = 0$ otherwise.

(a) Draw a diagram of this Markov chain.

(b) Compute $p_{32}^{(2)}$. **[sol]**

(c) Determine whether or not $\sum_{n=1}^{\infty} p_{12}^{(n)} = \infty$. [Hint: perhaps let $C = \{1, 2\}$.] **[sol]**

(d) Compute f_{32}, in three different ways. **[sol]**

(1.6.21) Problem. Consider a Markov chain with state space $S = \{1, 2, 3\}$, and transition probabilities $p_{11} = 1/6$, $p_{12} = 1/3$, $p_{13} = 1/2$, $p_{22} = p_{33} = 1$, and $p_{ij} = 0$ otherwise.

(a) Draw a diagram of this Markov chain.

(b) Compute (with explanation) f_{12}.

(c) Prove that $p_{12}^{(n)} \geq 1/3$, for all positive integers n.

(d) Compute $\sum_{n=1}^{\infty} p_{12}^{(n)}$.

(e) Why do the answers in parts (d) and (b) not contradict the implication (1) ⇒(5) in the Recurrence Equivalences Theorem?

(1.6.22) Problem. Consider a Markov chain with $S = \{1, 2, 3, 4, 5, 6, 7\}$, and transition probabilities

$$(p_{ij}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/5 & 4/5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/5 & 2/5 & 2/5 & 0 & 0 \\ 1/10 & 0 & 0 & 0 & 7/10 & 0 & 1/5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

(a) Draw a diagram of this Markov chain.

(b) Determine which states are recurrent, and which are transient.

(c) Compute f_{i1} for each $i \in S$. (Hint: leave f_{41} until last.)

(1.6.23) Problem. For each of the following sets of conditions, either provide (with explanation) an example of a state space S and Markov chain transition probabilities $\{p_{ij}\}_{i,j\in S}$ such that the conditions are satisfied, or prove that no such Markov chain exists.

(a) There are states $i, j \in S$ with $0 < f_{ij} < 1$, and $p_{ij}^{(n)} = 0$ for all $n \geq 3$.

(b) There are distinct states $i, j \in S$ with $f_{ij} > 0$ and $f_{ji} > 0$, and i is transient.

(c) The chain is irreducible, and there are distinct states $i, j, k, \ell \in S$ such that $f_{ij} = 1$, and $\sum_{n=1}^{\infty} p_{k\ell}^{(n)} < \infty$.

(d) There are distinct states $i, j, k \in S$ with $i \leftrightarrow j$ and $k \to i$, and $\sum_{n=1}^{\infty} p_{kj}^{(n)} < \infty$.

(e) The state space $S = \{1, 2, 3\}$, and $1 \leftrightarrow 2$, and $3 \to 1$, and $\sum_{n=1}^{\infty} p_{32}^{(n)} < \infty$.

END OF WEEK #2

Recall: Markov chains, f_{ij}, f-Expansion, $i \to j$, irreducibility, Cases Theorem, Finite Space Theorem, Recurrence Equivalences Theorem; application to Frog Walk, srw, etc.

2. Markov Chain Convergence

- Now that we know the basics about Markov chains, we turn to questions about its long-run probabilities:
  - What happens to $\mathbf{P}(X_n = j)$ for large n?
  - Does $\lim_{n\to\infty} \mathbf{P}(X_n = j)$ exist?
  - What does it equal?
- One clue is given by the following.
  - Suppose $q_j := \lim_{n\to\infty} \mathbf{P}(X_n = j)$ exists for all $j \in S$.
  - That is, $\lim_{n\to\infty} \mu_j^{(n)} = q_j$.
  - Then also $\lim_{n\to\infty} \mu_j^{(n+1)} = q_j$.
  - So, $\lim_{n\to\infty} \sum_{i\in S} \mu_i^{(n)} p_{ij} = q_j$.
  - If we could exchange the limit and sum, this would mean that $\sum_{i\in S} \left(\lim_{n\to\infty} \mu_i^{(n)}\right) p_{ij} = q_j$, i.e. $\sum_{i\in S} (q_i)\, p_{ij} = q_j$, i.e. $\sum_{i\in S} q_i p_{ij} = q_j$.
  - That means that the limits $\{q_j\}$, if they exist, must satisfy a certain "stationary" property, as we now discuss.

2.1. Stationary Distributions

A key concept is the following.

(2.1.1) Definition. If π is a *probability distribution* on S (i.e., $\pi_i \geq 0$ for all $i \in S$, and $\sum_{i\in S} \pi_i = 1$), then π is *stationary* for a Markov chain with transition probabilities $(\, p_{ij}\,)$ if $\sum_{i\in S} \pi_i p_{ij} = \pi_j$ for all $j \in S$.

- In matrix notation: $\pi\, P = \pi$.
- (Equivalently: π is a *left eigenvector* for the matrix P with *eigenvalue* 1, cf. (A.11.2).)

- Intuitively, π being stationary means if the chain starts with probabilities $\{\pi_i\}$, then it will keep the same probabilities one time unit later.
  - That is, if $\mu^{(0)} = \pi$, i.e. $\mathbf{P}(X_0 = i) = \pi_i$ for all i, then also $\mu^{(1)} = \nu P = \pi P = \pi$, i.e. $\mu^{(1)} = \mu^{(0)}$.
  - Similarly, if $\mu^{(n)} = \pi$, i.e. $\mathbf{P}(X_n = i) = \pi_i$ for all i, then $\mu^{(n+1)} = \mu^{(n)} P = \pi P = \pi$, i.e. $\mu^{(n+1)} \mu^{(n)}$.
  - Then, by induction, $\mu^{(m)} = \pi$ for all $m \geq n$, too. ("stationary")
  - In particular, if $\nu = \pi$, then $\mu^{(n)} = \pi$ for all $n \in \mathbf{N}$.
  - Hence, also $\pi \, P^n = \pi$ for $n = 0, 1, 2, \ldots$, i.e. $\sum_{i \in S} \pi_i p_{ij}^{(n)} = \pi_j$.
  - (This also follows since e.g. $\pi \, P^{(2)} = (\pi \, P) P = \pi \, P = \pi$.)
- In the Frog Example:
  - Let π be the *uniform distribution* (A.2.1) on S, i.e. $\pi_i = \frac{1}{20}$ for all $i \in S$.
  - Then $\pi_i \geq 0$ and $\sum_i \pi_i = 1$.
  - And, if e.g. $j = 8$, then $\sum\limits_{i \in S} \pi_i p_{i8} = \pi_7 p_{78} + \pi_8 p_{88} + \pi_9 p_{98} = \frac{1}{20}(\frac{1}{3}) + \frac{1}{20}(\frac{1}{3}) + \frac{1}{20}(\frac{1}{3}) = \frac{1}{20} = \pi_j$.
  - Similarly, for all $j \in S$, $\sum\limits_{i \in S} \pi_i p_{ij} = \frac{1}{20}(\frac{1}{3}) + \frac{1}{20}(\frac{1}{3}) + \frac{1}{20}(\frac{1}{3}) = \frac{1}{20} = \pi_j$.
  - So, $\{\pi_i\}$ is a stationary distribution!
  - Hence, if $\nu_i = 1/20$ for all i, then $\mathbf{P}(X_1 = i) = 1/20$ for all i, and indeed $\mathbf{P}(X_n = i) = 1/20$ for all i and all $n \geq 1$, too.
- More generally, suppose that $|S| < \infty$, and we have a chain which is *doubly stochastic*, i.e. $\sum_{i \in S} p_{ij} = 1$ for all $j \in S$ (in addition to the usual condition that $\sum_{j \in S} p_{ij} = 1$ for all $i \in S$).
  - (This holds for the Frog Example.)
  - Let π be the *uniform distribution* (A.2.1) on S, i.e. $\pi_i = \frac{1}{|S|}$ for all $i \in S$.
  - Then for all $j \in S$, $\sum\limits_{i \in S} \pi_i p_{ij} = \frac{1}{|S|} \sum\limits_{i \in S} p_{ij} = \frac{1}{|S|}(1) = \frac{1}{|S|} = \pi_j$.
  - So, $\{\pi_i\}$ is stationary.
- Ehrenfest's Urn example: ($S = \{0, 1, 2, \ldots, d\}$, $p_{ij} = i/d$ for $j = i - 1$, $p_{ij} = (d - i)/d$ for $j = i + 1$)
  - Is $\pi_i = \frac{1}{d+1}$ for all i a stationary distribution?
  - Well, if e.g. $j = 1$, then $\sum_{i \in S} \pi_i p_{ij} = \frac{1}{d+1}(p_{01} + p_{21}) = \frac{1}{d+1}(1 + \frac{2}{d}) \neq \frac{1}{d+1} = \pi_j$.
  - So, should <u>not</u> take $\pi_i = \frac{1}{d+1}$ for all i.
  - So, how should we choose π?

2.2. Searching for Stationarity

- How can we find stationary distributions, or decide if they even exist?
- One trick to help find stationary distributions is:

(2.2.1) Definition. A Markov chain is *reversible* (or *time reversible*, or satisfies *detailed balance*) with respect to a probability distribution $\{\pi_i\}$ if $\pi_i p_{ij} = \pi_j p_{ji}$ for all $i, j \in S$.

- The usefulness of reversibility is:

(2.2.2) Proposition. If a chain is reversible with respect to π, then π is a stationary distribution.

Proof. Reversibility means $\pi_i p_{ij} = \pi_j p_{ji}$, so then for $j \in S$, $\sum_{i \in S} \pi_i p_{ij} = \sum_{i \in S} \pi_j p_{ji} = \pi_j \sum_{i \in S} p_{ji} = \pi_j(1) = \pi_j$. ∎

- However, the converse to Proposition (2.2.2) is false; it is possible to have a stationary distribution even if the chain is not reversible:

(2.2.3) Problem. Let $S = \{1, 2, 3\}$, with transition matrix (p_{ij}) specified by $p_{12} = p_{23} = p_{31} = 1$, and let $\pi_1 = \pi_2 = \pi_3 = 1/3$.
(a) Show that π is a stationary distribution for P. [Hint: Either check directly, or use the doubly stochastic property, or find the left eigenvector with eigenvalue 1.]
(b) Show that the chain is not reversible w.r.t. π.

- In the Frog Example, let $\pi_i = 1/20$ for all i.
  - If $|j - i| \le 1$ or $|j - i| = 19$, then $\pi_i p_{ij} = (1/20)(1/3) = \pi_j p_{ji}$.
  - Otherwise, both sides equal 0.
  - So, the chain is reversible with respect to π!
  - (This provides an easier way to check stationarity.)
- What about Ehrenfest's Urn? What π?
  - New idea: perhaps each ball is equally likely to be in either Urn.
  - Then, the number of balls in Urn #1 follows a Binomial$(d, 1/2)$ distribution (A.2.2).
  - That is, let $\pi_i = \binom{d}{i}(1/2)^i(1/2)^{d-i} = 2^{-d}\binom{d}{i} = 2^{-d}\frac{d!}{i!(d-i)!}$.
  - Then $\pi_i \ge 0$ and $\sum_i \pi_i = 1$, so π is a distribution.
  - Is π stationary? Need to check if $\sum\limits_{i \in S} \pi_i p_{ij} = \pi_j$ for all $j \in S$.
  - This can be checked directly, but the calculations are pretty messy; see Problem (2.2.4).
- What to do instead?
  Use reversibility!
  - Need to check if $\pi_i p_{ij} = \pi_j p_{ji}$ for all $i, j \in S$.
  - Clearly, both sides are 0 unless $j = i + 1$ or $j = i - 1$.
  - If $j = i + 1$, then

$$\pi_i p_{ij} = 2^{-d}\binom{d}{i}\frac{d-i}{d} = 2^{-d}\frac{d!}{i!(d-i)!}\,\frac{d-i}{d} = 2^{-d}\frac{(d-1)!}{i!(d-i-1)!}\,.$$

  Also

$$\pi_j p_{ji} = 2^{-d}\binom{d}{j}\frac{j}{d} = 2^{-d}\frac{d!}{j!(d-j)!}\,\frac{j}{d}$$

$$= 2^{-d}\frac{(d-1)!}{(j-1)!(d-j)!} = 2^{-d}\frac{(d-1)!}{i!(d-i-1)!} = \pi_i p_{ij}\,.$$

  - If $j = i - 1$, then again $\pi_i p_{ij} = \pi_j p_{ji}$ (by a similar calculation, or simply by exchanging i and j in the above).

- So, it is reversible!
- So, π is stationary distribution!
- Intuitively, π_i is larger when i is close to $d/2$, which makes sense.
- But does $\mu_i^{(n)} \to \pi_i$? We'll see!

(2.2.4) Problem. Prove by direct calculation that the above π is stationary for Ehrenfest's Urn. [Hint: Don't forget the *Pascal's Triangle* identity that $\binom{d-1}{j-1} + \binom{d-1}{j} = \binom{d}{j}$.]

- What about simple random walk? Does it have a stationary distribution?
  - If $p = 1/2$, then s.s.r.w. is symmetric, so it would be reversible with respect to a "uniform" distribution on $\mathbf{Z}$, but what is that?
  - Or perhaps it doesn't have a stationary distribution?
  - But how could we prove that?
  - One method is given by:

(2.2.5) Vanishing Probabilities Proposition. If a Markov chain's transition probabilities satisfy that $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ for all $i, j \in S$, then the chain does <u>not</u> have a stationary distribution.

Proof. If there were a stationary distribution π, then we would have $\pi_j = \sum_{i\in S} \pi_i p_{ij}^{(n)}$ for any n, so also

$$\pi_j \;=\; \lim_{n\to\infty} \pi_j \;=\; \lim_{n\to\infty} \sum_{i\in S} \pi_i p_{ij}^{(n)} .$$

- If we could exchange the sum and limit, then this would imply that $\pi_j = \sum_{i\in S} \lim_{n\to\infty} \pi_i p_{ij}^{(n)} = \sum_{i\in S}(0) = 0$.
- So, we would have $\pi_j = 0$ for all j.
- But this means that $\sum_j \pi_j = 0$, which is a contraction.
- Therefore, there is no stationary distribution.
- Here, exchanging the limit and sum is justified by the *M-test* (A.10.1), since if $x_{ni} = \pi_i p_{ij}^{(n)}$, then

$$\sum_{i=1}^{\infty} \sup_n |x_{ni}| \;=\; \sum_{i=1}^{\infty} \sup_n |\pi_i p_{ij}^{(n)}| \;\le\; \sum_{i=1}^{\infty} |\pi_i| \;=\; 1 \;<\; \infty . \qquad \blacksquare$$

- So what about simple random walk (s.r.w.)?
  - Well, $p_{00}^{(n)} = 0$ for n odd.
  - And, we showed (1.5.6) that $p_{00}^{(n)} \approx [4p(1-p)]^{n/2}\sqrt{2/\pi n}$ for n large and even.
  - But we always have $[4p(1-p)]^{n/2} \le 1$.
  - So, for all large n, $p_{00}^{(n)} \le \sqrt{2/\pi n}$.
  - In particular, for s.r.w. with any value of p:

(2.2.6)
$$\lim_{n\to\infty} p_{00}^{(n)} \;=\; 0 .$$

  - Does this mean that $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ for <u>all</u> $i, j \in S$?

– Actually, yes it does!

(2.2.7) Vanishing Lemma. If a Markov chain has <u>some</u> $k, \ell \in S$ with $\lim_{n\to\infty} p_{k\ell}^{(n)} = 0$, then for <u>any</u> $i, j \in S$ with $k \to i$ and $j \to \ell$, $\lim_{n\to\infty} p_{ij}^{(n)} = 0$.

Proof. By (1.6.2), we can find $r, s \in \mathbf{N}$ with $p_{ki}^{(r)} > 0$ and $p_{j\ell}^{(s)} > 0$.

– By the Chapman-Kolmogorov Inequality (1.4.5), $p_{k\ell}^{(r+n+s)} \geq p_{ki}^{(r)} p_{ij}^{(n)} p_{j\ell}^{(s)}$.

– Hence, $p_{ij}^{(n)} \leq p_{k\ell}^{(r+n+s)} \big/ (p_{ki}^{(r)} p_{j\ell}^{(s)})$.

– But the assumptions imply that $\lim_{n\to\infty} \left[p_{k\ell}^{(r+n+s)} \big/ (p_{ki}^{(r)} p_{j\ell}^{(s)}) \right] = 0$.

– Also, we always have $p_{ij}^{(n)} \geq 0$.

– So, $p_{ij}^{(n)}$ is "sandwiched" between 0 and a sequence converging to 0.

– Hence, by the *Sandwich Theorem* (A.3.8), we have $\lim_{n\to\infty} p_{ij}^{(n)} = 0$. ∎

- This immediately implies:

(2.2.8) Vanishing Together Corollary. For an <u>irreducible</u> Markov chain, either (i) $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ for all $i, j \in S$, or (ii) $\lim_{n\to\infty} p_{ij}^{(n)} \neq 0$ for all $i, j \in S$.

- Then, combining this Vanishing Together Corollary with the Vanishing Probabilities Proposition gives:

(2.2.9) Vanishing Probabilities Corollary. If an irreducible Markov chain's transition probabilities satisfy that $\lim_{n\to\infty} p_{k\ell}^{(n)} = 0$ for <u>some</u> $k, \ell \in S$, then the chain does <u>not</u> have a stationary distribution.

- Returning to simple random walk:
  – Simple random walk is irreducible.
  – It may be recurrent or transient, depending on p.
  – However , we know that $\lim_{n\to\infty} p_{00}^{(n)} = 0$.
  – Hence, by the Vanishing Together Corollary, $p_{ij}^{(n)} \to 0$ for <u>all</u> $i, j \in S$.
  – And, by the Vanishing Probabilities Corollary, simple random walk does <u>not</u> have a stationary distribution.
- In particular, simple <u>symmetric</u> random walk (with $p = 1/2$) is in case (a) of the Cases Theorem, but in case (i) of the Vanishing Together Corollary. Hence, for all $i, j \in S$, $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ even though $\lim_{n\to\infty} p_{ij}^{(n)} = 0$. This is not a contradiction, as noted in (A.3.4).
- Aside: We always have $\sum_j p_{ij}^{(n)} = 1$ for all n, so $\lim_{n\to\infty} \sum_j p_{ij}^{(n)} = 1$, even though for simple random walk $\sum_j \lim_{n\to\infty} p_{ij}^{(n)} = 0$. How can this be? It must be that we cannot interchange the sum and limit. In particular, the M-test (A.10.1) does not apply, so the M-test conditions are <u>not</u> satisfied, i.e. we must have $\sum_j \sup_n p_{ij}^{(n)} = \infty$.
- It also follows that:

(2.2.10) Transient Not Stationary Corollary. A Markov chain which is irreducible and <u>transient</u> cannot have a stationary distribution.

Proof. If a chain is irreducible and transient, then by the Transience Equivalences Theorem (1.6.12), $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all $i, j \in S$.

– Hence, by (A.3.3), also $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ for all $i, j \in S$.

– Thus, by the Vanishing Probabilities Proposition (2.2.5), there is no stationary distribution. ∎

- If S is infinite, can there ever be a stationary distribution?
Yes!

(2.2.11) Example. Let $S = \mathbf{N} = \{1, 2, 3, \ldots\}$, with $p_{1,1} = 3/4$ and $p_{1,2} = 1/4$, and for $i \geq 2$, $p_{i,i} = p_{i,i+1} = 1/4$ and $p_{i,i-1} = 1/2$.

– This chain is clearly irreducible.

– But does it have a stationary distribution?

– Let $\pi_i = 2^{-i}$ for all $i \in S$, so $\pi_i \geq 0$ and $\sum_i \pi_i = 1$.

– Then for any $i \in S$, $\pi_i p_{i,i+1} = 2^{-i}(1/4) = 2^{-i-2}$.

– Also, $\pi_{i+1} p_{i+1,i} = 2^{-(i+1)}(1/2) = 2^{-i-2}$. Equal!

– And, $\pi_i p_{i,j} = 0$ if $|j - i| \geq 2$.

– So, the chain is reversible with respect to π.

– So, π is a stationary distribution.

– Does this mean that $\lim_{n\to\infty} P(X_n = j) = \pi_j = 2^{-j}$ for all $j \in S$?

2.3. Obstacles to Convergence

- If chain has stationary distribution $\{\pi_i\}$, does $\lim_{n\to\infty} \mathbf{P}[X_n = i] = \pi_i$?
- Not necessarily!

(2.3.1) Example. Let $S = \{1, 2\}$, and $\nu_1 = 1$, and $(p_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (see Figure 6).

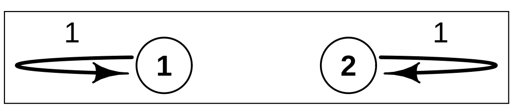


Figure 6: A diagram of Example (2.3.1).

– Let $\pi_1 = \pi_2 = \frac{1}{2}$ Then $\{\pi_i\}$ is stationary. (Check this either directly, or by the doubly stochastic property, or by reversibility.)

– But $\lim_{n\to\infty} \mathbf{P}(X_n = 1) = \lim_{n\to\infty}$
$1 = 1 \neq \frac{1}{2} = \pi_1$.

– However, this chain is not
irreducible (i.e. it's "reducible").

(2.3.2) Example. Let $S = \{1, 2\}$, and $\nu_1 = 1$, and $(p_{ij}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ (see Figure 7).

– This example is irreducible.

– Again, if $\pi_1 = \pi_2 = \frac{1}{2}$, then π is stationary (check!).

– But $\mathbf{P}(X_n = 1) = \begin{cases} 1, & n \text{ even} \\ 0, & n \text{ odd} \end{cases}$

– So, $\lim_{n\to\infty} \mathbf{P}[X_n = 1]$ does not even exist!

– In fact, this chain is "periodic", as we now discuss.

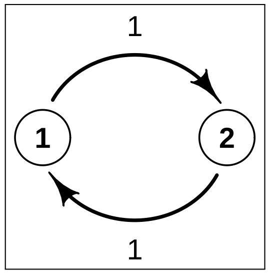


Figure 7: A diagram of Example (2.3.2).

(2.3.3) Definition. The *period* of a state i is the *greatest common divisor (gcd)* of the set $\{n \geq 1; p_{ii}^{(n)} > 0\}$, i.e. the largest number m such that all the values of n with $p_{ii}^{(n)} > 0$ are all integer multiples of m.

- (We assume that $f_{ii} > 0$, so that $\{n \geq 1; p_{ii}^{(n)} > 0\}$ is non-empty.)
- e.g. if the period of i is 2, this means that it is only possible to get from i to i in an even numbers of steps. (Like Example (2.3.2) above.)
- If the period of each state is 1, we say the chain is *aperiodic*.
- Otherwise we say the chain is *periodic*.

- It follows from the definition of "common divisor" that:

(2.3.4) If state i has period t, and $p_{ii}^{(m)} > 0$, then m is an integer multiple of t, i.e. t divides m.

(2.3.5) Example. Let $S = \{1, 2, 3\}$, and $p_{12} = p_{23} = p_{31} = 1$.

- Then (check!) the period of each state is 3.

(2.3.6) Example. Let $S = \{1, 2, 3\}$, and $(p_{ij}) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$ (see Figure 8).

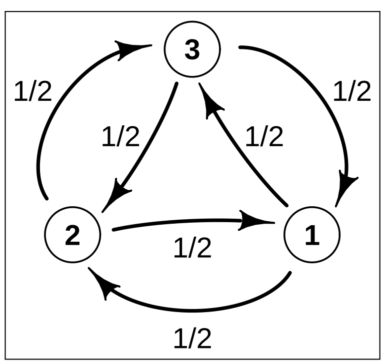


Figure 8: A diagram of Example (2.3.6).

- Then the period of state 1 is $gcd\{2, 3, \ldots\} = 1$.
- In fact, the chain is aperiodic (i.e., all states have period 1).

- The gcd of any collection of positive integers which includes 1 must equal 1, i.e. $gcd\{1, \ldots\} = 1$. Hence:

(2.3.7) If $p_{ii} > 0$, then the period of state i is 1.

- (The converse to (2.3.7) is false, as shown by Example (2.3.6).)
- More generally, since also $gcd\{n, n+1, \ldots\} = 1$, we have:

(2.3.8) If $p_{ii}^{(n)} > 0$ and $p_{ii}^{(n+1)} > 0$, then the period of state i is 1.

- What about the Frog Example?
  - There $p_{ii} = 1/3 > 0$, so each state has period 1, i.e. the chain is aperiodic.
- What about Simple Random Walk?
  - It can only return to i after an <u>even</u> number of steps, so the period of each state is 2.
- What about Ehrenfest's Urn?
  - Again, it can only return to a state after an <u>even</u> number of steps, so the period of each state is 2.

(2.3.9) Problem. Consider the example from Problem (1.6.17), where $S = \{1, 2, 3\}$, and $p_{11} = p_{12} = p_{22} = p_{23} = p_{32} = p_{33} = 1/2$, with $p_{ij} = 0$ otherwise. Is this chain aperiodic? **[sol]**

- Finally, we present a few helpful properties of periodicity.

(2.3.10) Equal Periods Lemma. If $i \leftrightarrow j$, then the periods of i and of j are equal.

Proof.

- Let the periods of i and j be t_i and t_j.
- By (1.6.2), find $r, s \in \mathbf{N}$ with $p_{ij}^{(r)} > 0$ and $p_{ji}^{(s)} > 0$.
- Then by (1.4.5), $p_{ii}^{(r+s)} \geq p_{ij}^{(r)} p_{ji}^{(s)} > 0$.
- So, by (2.3.4), t_i divides $r + s$.
- Suppose now that $p_{jj}^{(n)} > 0$.
- Then $p_{ii}^{(r+n+s)} \geq p_{ij}^{(r)} p_{jj}^{(n)} p_{ji}^{(s)} > 0$, so t_i divides $r + n + s$.
- Since t_i divides $r + n + s$ and also divides $r + s$, therefore t_i must divide n.
- Since this is true for any n with $p_{jj}^{(n)} > 0$, it follows that t_i is a common divisor of $\{n \in \mathbf{N};\ p_{jj}^{(n)} > 0\}$.
- But t_j is the <u>greatest</u> such common divisor.
- So, $t_j \geq t_i$.
- Similarly, $t_i \geq t_j$, so $t_i = t_j$. ∎

(2.3.11) Equal Periods Corollary. If a chain is <u>irreducible</u>, then all states have the same period.

- (So, we can say that e.g. the whole <u>chain</u> has period 3, or the whole <u>chain</u> is aperiodic, etc.)

- Combining this fact with (2.3.7) shows:

(2.3.12) Corollary. If a chain is irreducible, and $p_{ii} > 0$ for <u>some</u> state i, then the chain is aperiodic.

(2.3.13) Problem. Consider a Markov chain made up of two intersecting cycles of length 4 and 5 respectively, with state space $S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, and $p_{01} = p_{05} = 1/2$, and $p_{12} = p_{23} = p_{34} = p_{40} = 1$, and $p_{56} = p_{67} = p_{78} = p_{89} = p_{90} = 1$.

(a) Draw a diagram of this Markov chain.

(b) Determine whether or not this chain is irreducible.

(c) Compute the period of the state 0.

(d) Compute the period of each state i.

(e) Determine if this chain is periodic or aperiodic.

2.4. Convergence Theorem

- The good news is that the above represent *all* of the obstacles to stationarity.
  - That is, once we have found a stationary distribution, and ruled out the obstacles of reducibility or periodicity, then the probabilities must converge:

(2.4.1) Markov Chain Convergence Theorem. If a Markov chain is irreducible, and aperiodic, and has a stationary distribution $\{\pi_i\}$, then $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ for all $i, j \in S$, and $\lim_{n\to\infty} \mathbf{P}(X_n = j) = \pi_j$ for any initial probabilities $\{\nu_i\}$.

- To prove this (big) theorem, we need two preliminary results.
- The first one connects stationary distributions to recurrence.

(2.4.2) Stationary Recurrence Theorem. If chain irreducible, and has a stationary distribution, then it is <u>recurrent</u>.

Proof. The Transient Not Stationary Corollary (2.2.10) says that a chain cannot be irreducible and transient and have a stationary distribution.

- So, if a chain is irreducible and has a stationary distribution, then it cannot be transient, i.e. it must be recurrent. ■

- The second one uses a result from number theory (A.12.4) to provide a more concrete way to use aperiodicity.

(2.4.3) Proposition. If a state i has $f_{ii} > 0$ and is *aperiodic*, then there is $n_0(i) \in \mathbf{N}$ such that $p_{ii}^{(n)} > 0$ for all $n \geq n_0(i)$.

Proof. Let $A = \{n \geq 1 : p_{ii}^{(n)} > 0\}$.

- Then A is non-empty since $f_{ii} > 0$.
- And, if $m \in A$ and $n \in A$, then $p_{ii}^{(m)} > 0$ and $p_{ii}^{(n)} > 0$, so by the Chapman-Kolmogorov Inequality (1.4.5), $p_{ii}^{(m+n)} \geq p_{ii}^{(m)} p_{ii}^{(n)} > 0$, so $m + n \in A$, which shows that A satisfies *additivity*.
- Also, $gcd(A) = 1$ since the state i is aperiodic.
- Hence, from the Number Theory Lemma (A.12.4), there is $n_0 \in \mathbf{N}$ such that for all $n \geq n_0$ we have $n \in A$, i.e. $p_{ii}^{(n)} > 0$. ■

(2.4.4) Corollary. If a chain is irreducible and aperiodic, then for any states $i, j \in S$, there is $n_0(i,j) \in \mathbf{N}$ such that $p_{ij}^{(n)} > 0$ for all $n \geq n_0(i,j)$.

Proof. Find $n_0(i)$ as above, and find $m \in \mathbf{N}$ with $p_{ij}^{(m)} > 0$.

- Then let $n_0(i,j) = n_0(i) + m$.
- Then if $n \geq n_0(i,j)$, then $n - m \geq n_0(i)$, so $p_{ij}^{(n)} \geq p_{ii}^{(n-m)} p_{ij}^{(m)} > 0$. ∎

- This next lemma is the key to the proof.

(2.4.5) Markov Forgetting Lemma. If a Markov chain is irreducible and aperiodic, and has stationary distribution $\{\pi_i\}$, then for all $i, j, k \in S$,

$$\lim_{n\to\infty} \left| p_{ik}^{(n)} - p_{jk}^{(n)} \right| \;=\; 0\,.$$

(Intuitively, after a long time n, the chain "forgets" whether it started from state i or from state j.)

Proof. The proof uses the *coupling* technique of defining two different copies of the chain.

- Specifically, define a new Markov chain $\{(X_n, Y_n)\}_{n=0}^{\infty}$, with state space $\overline{S} = S \times S$, and transition probabilities $\overline{p}_{(ij),(k\ell)} = p_{ik} p_{j\ell}$.
  - Intuitively, the new chain has two coordinates, each of which is an independent copy of the original Markov chain.
- Since the two copies are independent, this new chain has a joint stationary distribution given by $\overline{\pi}_{(ij)} = \pi_i \pi_j$ for $i, j \in S$.
- Furthermore, whenever $n \geq \max\left[n_0(i,k),\, n_0(j,\ell)\right]$, we have $\overline{p}_{(ij),(k\ell)}^{(n)} = p_{ij}^{(n)} p_{k\ell}^{(n)} > 0$.
  - It follows that the new chain is irreducible (since it eventually has positive transition probabilities from anywhere to anywhere), and is aperiodic by (2.3.8).
- So, by the Stationary Recurrence Theorem (2.4.2), the new chain is recurrent.

END OF WEEK #3

Reminder: Midterm #1 next Thurs, Feb 7, at 6:10 pm, in MS 2170 [surnames A-K] and MS 3154 [surnames L-Z]. It will cover all lecture material up to that time. No aids allowed. Bring your student card. Do not sit next to anyone you know. Explain your answers well.

Reminder: TA office hours in SS 2119, Wednesdays & Fridays 3:10–5:00, plus Tues Feb 5 from 3:10–7:00 in IN 313. Instructor office hours Tuesday Feb 5 from 12:30 to 2:30 and Wednesday Feb 6 from 11:30 to 1:30.

Recall: stationary distributions, reversibility, Vanishing Probabilities Prop, periods, application to examples, Equals Periods Lemma, Markov chain Convergence Theorem, Markov Forgetting Lemma

Proof of the Markov Forgetting Lemma, continued:

- Next, choose any $i_0 \in S$, and let $\tau = \inf\{n \geq 0; X_n = Y_n = i_0\}$ be the first time that the new chain hits (i_0, i_0), i.e. that both copies of the chain equal i_0 at the same time.
- Since the new chain is irreducible and recurrent, it follows from the Recurrence Equivalences Theorem (1.6.11) that $\overline{f}_{(ij),(i_0 i_0)} = 1$.
  - This means that, starting from (i, j), the new chain must eventually hit (i_0, i_0), so $\mathbf{P}_{(ij)}(\tau < \infty) = 1$.
- Then, using the *Law of Total Probability* (A.2.8),

$$p_{ik}^{(n)} \;=\; \mathbf{P}_{(ij)}(X_n = k) \;=\; \sum_{m=1}^{\infty} \mathbf{P}_{(ij)}(X_n = k,\ \tau = m)$$

$$= \sum_{m=1}^{n} \mathbf{P}_{(ij)}(X_n = k,\ \tau = m) + \mathbf{P}_{(ij)}(X_n = k,\ \tau > n)\,.$$

  - Similarly,

$$p_{jk}^{(n)} \;=\; \sum_{m=1}^{n} \mathbf{P}_{(ij)}(Y_n = k,\ \tau = m) + \mathbf{P}_{(ij)}(Y_n = k,\ \tau > n)\,.$$

- On the other hand, if $n \geq m$, then by conditioning (A.5.2),

$$\mathbf{P}_{(ij)}(X_n = k, \tau = m) \;=\; \mathbf{P}_{(ij)}(\tau = m)\ \mathbf{P}_{(ij)}(X_n = k \,|\, X_m = Y_m = i_0)$$

$$= \mathbf{P}_{(ij)}(\tau = m)\ \mathbf{P}(X_n = k \,|\, X_m = i_0) \;=\; \mathbf{P}_{(ij)}(\tau = m)\ p_{i_0,k}^{(n-m)}\,.$$

  - Similarly,

$$\mathbf{P}_{(ij)}(Y_n = k, \tau = m) \;=\; \mathbf{P}_{(ij)}(\tau = m)\ p_{i_0,k}^{(n-m)}\,.$$

  - So, $\mathbf{P}_{(ij)}(X_n = k, \tau = m) \;=\; \mathbf{P}_{(ij)}(Y_n = k, \tau = m)$. This is key!
  - (Intuitively, this makes sense, since once $X_m = Y_m = i_0$, then X_n and Y_n have the same probabilities after that.)
- Hence, for all $i, j, k \in S$, using the triangle inequality (A.12.1) and monotonicity of probabilities (A.2.9),

$$\begin{aligned}
\left|p_{ik}^{(n)} - p_{jk}^{(n)}\right| &= \left|\mathbf{P}_{(ij)}(X_n = k) - \mathbf{P}_{(ij)}(Y_n = k)\right| \\
&= \Big|\textstyle\sum_{m=1}^{n} \mathbf{P}_{(ij)}(X_n = k,\ \tau = m) + \mathbf{P}_{(ij)}(X_n = k,\ \tau > n) \\
&\qquad - \textstyle\sum_{m=1}^{n} \mathbf{P}_{(ij)}(Y_n = k,\ \tau = m) - \mathbf{P}_{(ij)}(Y_n = k,\ \tau > n)\Big| \\
&= \left|\mathbf{P}_{(ij)}(X_n = k,\ \tau > n) - \mathbf{P}_{(ij)}(Y_n = k,\ \tau > n)\right| \\
&\leq |\mathbf{P}_{(ij)}(X_n = k,\ \tau > n)| + |\mathbf{P}_{(ij)}(Y_n = k,\ \tau > n)| \\
&\leq \mathbf{P}_{(ij)}(\tau > n) + \mathbf{P}_{(ij)}(\tau > n) \;=\; 2\,\mathbf{P}_{(ij)}(\tau > n)\,.
\end{aligned}$$

- But using (A.7.3), $\lim_{n\to\infty} 2\,\mathbf{P}_{(ij)}(\tau > n) = 2\,\mathbf{P}_{(ij)}(\tau = \infty) = 0$, since $\mathbf{P}_{(ij)}(\tau < \infty) = 1$. ∎

- (Aside: the above factor of "2" isn't really necessary, since $\mathbf{P}_{(ij)}(X_n = k,\ \tau > n)$ and $\mathbf{P}_{(ij)}(Y_n = k,\ \tau > n)$ are each between 0 and $\mathbf{P}_{(ij)}(\tau > n)$.)

Proof of Markov Chain Convergence Theorem.

- Intuition: The Markov Forgetting Lemma says the intial probabilities don't matter in the long run. So, we might as well have started in the stationary distribution. But then we would always remain in the stationary distribution, too!

- More precisely, using the triangle inequality (A.12.1),

$$\left|p_{ij}^{(n)} - \pi_j\right| \;=\; \left|\sum_{k\in S} \pi_k \left(p_{ij}^{(n)} - p_{kj}^{(n)}\right)\right| \;\le\; \sum_{k\in S} \pi_k \left|p_{ij}^{(n)} - p_{kj}^{(n)}\right| .$$

- By the Markov Forgetting Lemma (2.4.5), $\lim_{n\to\infty} \pi_k \, |p_{ij}^{(n)} - p_{kj}^{(n)}| = 0$ for all $k \in S$.
  - But what about the limit of the sum? Can we use the M-test?
  - Well, $\sup_n |p_{ij}^{(n)} - p_{kj}^{(n)}| \le 2$ by the triangle inequality (A.12.1), so we have $\sum_{k\in S} \sup_n \pi_k |p_{ij}^{(n)} - p_{kj}^{(n)}| \le \sum_{k\in S} 2\,\pi_k = 2 < \infty$.
  - Hence, by the M-test (A.10.1), we can exchange the limit and sum:

$$\lim_{n\to\infty} \left|p_{ij}^{(n)} - \pi_j\right| \;\le\; \lim_{n\to\infty} \sum_{k\in S} \pi_k \left|p_{ij}^{(n)} - p_{kj}^{(n)}\right|$$

$$=\; \sum_{k\in S} \lim_{n\to\infty} \pi_k \left|p_{ij}^{(n)} - p_{kj}^{(n)}\right| \;=\; \sum_{k\in S} (0) \;=\; 0 .$$

  - That is, $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ for all $i, j \in S$, as claimed.
- Finally, for <u>any</u> $\{\nu_i\}$, we have (again using the M-test):

$$\lim_{n\to\infty} \mathbf{P}(X_n = j) \;=\; \lim_{n\to\infty} \sum_{i\in S} \mathbf{P}(X_0 = i,\ X_n = j) \;=\; \lim_{n\to\infty} \sum_{i\in S} \nu_i \, p_{ij}^{(n)}$$

$$=\; \sum_{i\in S} \nu_i \lim_{n\to\infty} p_{ij}^{(n)} \;=\; \sum_{i\in S} \nu_i \, \pi_j \;=\; (1)\pi_j \;=\; \pi_j . \qquad \blacksquare$$

- So, for the Frog Example, regardless of the choice of $\{\nu_i\}$, we have $\lim_{n\to\infty} \mathbf{P}(X_n = 14) = 1/20$, etc.
- Similarly all other examples which are irreducible and aperiodic and have a stationary distribution.
- For example, we showed that Example (2.2.11), with $S = \mathbf{N} = \{1, 2, 3, \ldots\}$ and $p_{1,1} = 3/4$ and $p_{1,2} = 1/4$, and for $i \ge 2$, $p_{i,i} = p_{i,i+1} = 1/4$ and $p_{i,i-1} = 1/2$, is irreducible and aperiodic with stationary distribution $\pi_i = 2^{-i}$.
  - Hence, by the Markov Chain Convergence Theorem (2.4.1), in Example (2.2.11) we have $\lim_{n\to\infty} \mathbf{P}(X_n = j) = \pi_j = 2^{-j}$ for all $j \in S$.
- Theorem (2.4.1) says nothing about *quantitative convergence rates*, i.e. how <u>large</u> n has to be to make $|\mathbf{P}(X_n = j) - \pi_j| < \epsilon$ for given $\epsilon > 0$. This is an active area of research*, but it is difficult since it depends on all of the details of all of the Markov chain transitions, so we do not pursue it further here.
- Another application of (2.4.1) is:

(2.4.6) Corollary. If a chain is irreducible and aperiodic, then it has at most <u>one</u> stationary distribution.

---

*For an introduction, see e.g. J.S. Rosenthal, "Convergence rates of Markov chains", *SIAM Review* **37** (1995), 387–405.

Proof. By Theorem (2.4.1), any stationary distribution that it <u>does</u> have must be equal to $\lim_{n\to\infty} \mathbf{P}(X_n = j)$, so they're all equal. ∎

- Corollary (2.4.6) is generalised in (2.5.3) below.

(2.4.7) Example. Let $S = \{1, 2, 3\}$, and $(\, p_{ij}\,) = \begin{pmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 2/3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

  - Stationary dist #1: $\pi_1 = \pi_2 = \pi_3 = 1/3$ (since doubly stochastic).
  - Stationary dist #2: $\pi_1 = \pi_2 = 1/2$ and $\pi_3 = 0$.
  - Stationary dist #3: $\pi_1 = \pi_2 = 0$ and $\pi_3 = 1$.
  - Stationary dist #4: $\pi_1 = \pi_2 = 1/8$ and $\pi_3 = 3/4$.
  - So, in this example the stationary distribution is not unique.
  - But the chain is not <u>irreducible</u>.

(2.4.8) Example. Let $S = \{1, 2, 3\}$, and

$$(\, p_{ij}\,) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}.$$

  - Is it irreducible?
    Yes! e.g. $f_{11} \geq p_{12}\, p_{21} = (1/2)(1/3) > 0$, etc.
  - Is it aperiodic?
    Yes! e.g. $p_{33} = 1/2 > 0$, and irreducible.
  - Does it have a stationary distribution? If yes, what is it?
  - Well, we need $\pi P = \pi$, i.e. $\sum_{i\in S} \pi_i p_{ij} = \pi_j$ for all $j \in S$.
  - $j = 1$: $\pi_1(0) + \pi_2(1/3) + \pi_3(1/4) = \pi_1$.
  - $j = 2$: $\pi_1(1/2) + \pi_2(1/3) + \pi_3(1/4) = \pi_2$.
  - $j = 3$: $\pi_1(1/2) + \pi_2(1/3) + \pi_3(1/2) = \pi_3$.
  - Subtract j=2 line from j=3 line: $\pi_3[(1/2) - (1/4)] = \pi_3 - \pi_2$, i.e. $\pi_2 = \pi_3(3/4)$.
  - Substitute into j=1 line: $\pi_1 = \pi_2(1/3) + \pi_3(1/4) = \pi_3(3/4)(1/3) + \pi_3(1/4) = \pi_3/2$.
  - Then, we need $\pi_1 + \pi_2 + \pi_3 = 1$, i.e. $\pi_3/2 + \pi_3(3/4) + \pi_3 = 1$, i.e. $\pi_3(9/4) = 1$, so we must have $\pi_3 = 4/9$.
  - Thus, $\pi_2 = \pi_3(3/4) = (4/9)(3/4) = 1/3$.
  - And, $\pi_1 = \pi_3/2 = (4/9)/2 = 2/9$.
  - Hence, a stationary distribution is $\pi = (2/9, 1/3, 4/9)$.
  - It can then be checked directly that $\pi P = \pi$. (Check it!)
  - And, this stationariy distribution is unique by (2.4.6).
  - Finally, does $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ for all $i, j \in S$?
  - Yes, by the Markov Chain Convergence Theorem (2.4.1)!

- Finally, we consider convergence of simple random walk (s.r.w.).
  - We know by (2.2.9) that s.r.w. has no stationary distribution, so it does not converge in the sense of (2.4.1).

- But we know (1.6.15) that with probability 1 (w.p. 1), if $p > 1/2$ then s.r.w. converges to $+\infty$, or if $p < 1/2$, then s.r.w. converges to $-\infty$.
- But what about simple symmetric random walk (s.s.r.w.), with $p = 1/2$?
- Well, we know (1.5.7) that s.s.r.w. is recurrent, so $f_{ij} = 1$ for all $i, j \in \mathbf{N}$, and it has infinite *fluctuations* (1.6.13), i.e. w.p. 1 it eventually hits every possible collection of integers, in sequence, without end.
- So, s.s.r.w. certainly doesn't converge to anything w.p. 1.
- Its absolute values $|X_n|$ still have infinite fluctuations on the non-negative integers, so they don't converge to anything w.p. 1 either.
- However, the absolute values still converge in some sense:

(2.4.9) Proposition. If $\{X_n\}$ is simple symmetric random walk, then the absolute values $|X_n|$ converge weakly to positive infinity. (So, $|X_n|$ converges weakly but not strongly; cf. Section A.6.)

Proof. We can write (similar to the proof of (1.6.15)) that $X_n = \sum_{i=1}^n Z_i$, where $\{Z_i\}$ are i.i.d. ± 1 with probability 1/2 each, and hence common mean $m = 0$ and variance $v = 1$.

- Hence, the Central Limit Theorem (A.6.3) says that

$$\lim_{n\to\infty} \mathbf{P}\Big(-b < \frac{X_n}{\sqrt{n}} < b\Big) \;=\; \int_{-b}^{b} \phi(x)\,dx\,,$$

  where ϕ is the density function of the standard normal distribution.
- Hence, $\lim_{b\searrow 0} \lim_{n\to\infty} \mathbf{P}(-b < \frac{X_n}{\sqrt{n}} < b) = 0$.
- But $\mathbf{P}(|X_n| < K) = \mathbf{P}(-K < X_n < K) = \mathbf{P}(\frac{-K}{\sqrt{n}} < \frac{X_n}{\sqrt{n}} < \frac{K}{\sqrt{n}})$.
- So, by monotonicity (A.2.9), for any finite K and any $b > 0$, $\mathbf{P}(|X_n| < K) < \mathbf{P}(-b < \frac{X_n}{\sqrt{n}} < b)$ for all sufficiently large n.
- It follows that $\lim_{n\to\infty} \mathbf{P}(|X_n| < K) = 0$ for any finite K.
- Thus, as in (A.6.1), $|X_n|$ converges weakly to positive infinity. ∎

(2.4.10) Problem. Consider a Markov chain on the state space $S = \{1, 2, 3, 4\}$ with the following transition matrix:

$$P \;=\; \begin{pmatrix} 0.1 & 0.2 & 0.5 & 0.2 \\ 0.4 & 0.3 & 0.2 & 0.1 \\ 0.3 & 0.2 & 0.1 & 0.4 \\ 0.2 & 0.3 & 0.2 & 0.3 \end{pmatrix}$$

Let π be the uniform distribution on S, so $\pi_i = 1/4$ for all $i \in S$.

(a) Compute $p_{14}^{(2)}$. [**sol**]

(b) Is this Markov chain reversible with respect to π? [**sol**]

(c) Is π a stationary distribution for this Markov chain? [**sol**]

(d) Does $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ for all $i, j \in S$? Why or why not? [**sol**]

(2.4.11) Problem. Consider the Markov chain from Problem (1.4.6), with state space $S = \{1, 2\}$, and transition probabilities $p_{11} = 2/3$, $p_{12} = 1/3$,

$p_{21} = 1/4$, and $p_{22} = 3/4$, which was shown to be irreducible in Problem (1.6.16).

(a) Determine whether or not this chain is aperiodic. [**sol**]

(b) Find a probability distribution $\{\pi_i\}$ which is stationary for this chain. [**sol**]

(c) Determine whether or not $\lim_{n\to\infty} p_{12}^{(n)} = \pi_2$. [**sol**]

(2.4.12) Problem. Suppose there are 10 lily pads arranged in a circle, numbered consecutively clockwise from 1 to 10. A frog begins on lily pad #1. Each second, the frog jumps one pad clockwise with probability 1/4, or two pads clockwise with probability 3/4.

(a) Specify a state space S, initial probabilities $\{\nu_i\}$, and transition probabilities $\{p_{ij}\}$, with respect to which this process is a Markov chain. [**sol**]

(b) Determine if this Markov chain is irreducible. [**sol**]

(c) Determine if this Markov chain is aperiodic, or if not then what its period equals. [**sol**]

(d) Determine whether or not $\sum_{n=1}^{\infty} p_{15}^{(n)} = \infty$. [**sol**]

(e) Either find a stationarity distribution $\{\pi_i\}$ for this chain, or prove that no stationary distribution exists. [**sol**]

(f) Determine whether or not $\lim_{n\to\infty} p_{15}^{(n)}$ exists, and if so what it equals. [**sol**]

(2.4.13) Problem. For each of the following sets of conditions, either provide (with explanation) an example of a state space S and Markov chain transition probabilities $\{p_{ij}\}_{i,j\in S}$ such that the conditions are satisfied, or prove that no such Markov chain exists.

(a) The chain is irreducible and periodic (i.e., not aperiodic), and has a stationary probability distribution. [**sol**]

(b) The chain is irreducible, and there are states $k \in S$ having period 2, and $\ell \in S$ having period 4. [**sol**]

(c) The chain is irreducible and transient, and is reversible with respect to some probability distribution π. [**sol**]

(d) The chain is irreducible and has a stationary probability distribution π, and $p_{ij} < 1$ for all $i, j \in S$, but the chain is not reversible with respect to π. [**sol**]

(e) There are distinct states $i, j, k \in S$ with $p_{ij} > 0$, $p_{jk}^{(2)} > 0$, and $p_{ki}^{(3)} > 0$, and the state i is periodic (i.e., has period > 1). [**sol**]

(f) $p_{11} > 1/2$, and the state 1 has period 2.

(g) $p_{11} > 1/2$, and the period of state 2 equals 3, and the chain is irreducible.

(2.4.14) Problem. Let $S = \mathbf{Z}$ (the set of all integers), and let $h : S \to (0, 1)$ with $h(i) > 0$ for all $i \in S$, and $\sum_{i\in S} h(i) = 1$. Consider the transition probabilities on S given by $p_{ij} = (1/4)\min(1, h(j)/h(i))$ if $j = i-2, i-1, i+1$, or $i+2$, and $p_{ii} = 1 - p_{i,i-2} - p_{i,i-1} - p_{i,i+1} - p_{i,i+2}$, and $p_{ij} = 0$ whenever $|j-i| \geq 3$. Prove that $\lim_{n\to\infty} p_{ij}^{(n)} = h(j)$ for all $i, j \in S$. (Hint: reversibility might help.)

(2.4.15) Problem. For each of the following sets of conditions, either provide (with explanation) an example of a state space S and Markov chain transition probabilities $\{p_{ij}\}_{i,j\in S}$ such that the conditions are satisfied, or prove that no such a Markov chain exists.

(a) The chain is irreducible, with period 3, and has a stationary distribution. [**sol**]

(b) There is $k \in S$ having period 2, and $\ell \in S$ having period 4. [**sol**]

(c) The chain has a stationary distribution π, and $0 < p_{ij} < 1$ for all $i, j \in S$, but the chain is not reversible with respect to π. [**sol**]

(d) The chain is irreducible, and there are distinct states $i, j, k, \ell \in S$ such that $f_{ij} < 1$, and $\sum_{n=1}^{\infty} p_{k\ell}^{(n)} = \infty$. [**sol**]

(e) The chain is irreducible, and there are are distinct states $i, j, k \in S$ with $p_{ij} > 0$, $p_{jk}^{(2)} > 0$, and $p_{ki}^{(3)} > 0$, and state i is periodic with period equal to an odd number. [**sol**]

(f) There are states $i, j \in S$ with $0 < f_{ij} < 1$, and $p_{ij}^{(n)} = 0$ for all $n \in \mathbf{N}$. [**sol**]

(g) There are states $i, j \in S$ with $0 < f_{ij} < 1$, and $p_{ij}^{(n)} = 0$ for all $n \geq 2$. [**sol**]

(h) There are distinct states $i, j \in S$ with $f_{ij} > 0$ and $f_{ji} > 0$, and i is transient. [**sol**]

(i) There are distinct states $i, j, k \in S$ with $f_{ij} = 1/2$, $f_{jk} = 1/3$, and $f_{ik} = 1/10$. [**sol**]

(2.4.16) Problem. Consider a Markov chain with state space $S = \{1, 2, 3\}$, and transition probabilities $p_{12} = 1/2$, $p_{13} = 1/2$, $p_{21} = 1/3$, $p_{23} = 2/3$, $p_{31} = 1/4$, $p_{32} = 3/4$.

(a) Compute $p_{11}^{(2)}$.

(b) Compute $p_{13}^{(3)}$.

(c) Find a probability distribution π which is stationary for this chain.

(d) Determine (with explanation) whether or not $\lim_{n\to\infty} p_{13}^{(n)} = \pi_3$.

(e) Determine (with explanation) whether or not $f_{13} = 1$.

(f) Determine (with explanation) whether or not $\sum_{n=1}^{\infty} p_{13}^{(n)} = \infty$.

(2.4.17) Problem. Consider a Markov chain on the state space $S = \{1, 2, 3\}$ with transition matrix:

$$P = \begin{pmatrix} 1/3 & 0 & 2/3 \\ 1/6 & 1/3 & 1/2 \\ 4/5 & 0 & 1/5 \end{pmatrix}$$

(a) Compute $p_{13}^{(2)}$.

(b) Specify (with explanation) which states are recurrent, and which states are transient.

(c) Compute f_{23}.

(d) Find a stationarity distribution $\{\pi_i\}$ for the chain.

(e) Determine if the chain reversible with respect to π.

(f) Determine whether or not $\lim_{n\to\infty} p_{1j}^{(n)} = \pi_j$ for all $j \in S$. [Hint: Don't forget the Closed Subset Note (1.6.14).]

(2.4.18) Problem. Consider a Markov chain with state space $S = \{1, 2, 3\}$, and transition probabilities $p_{12} = 1/2$, $p_{13} = 1/2$, $p_{21} = 1/3$, $p_{23} = 2/3$, $p_{31} = 2/5$, $p_{32} = 3/5$, otherwise $p_{ij} = 0$.

(a) Compute $p_{11}^{(2)}$. [**sol**]

(b) Find a probability distribution π which is stationary for this chain. [**sol**]

(c) Determine whether or not $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ for all $i, j \in S$. [**sol**]

(2.4.19) Problem. Suppose there are 10 lily pads arranged in a circle, numbered consecutively clockwise from 1 to 10. A frog begins on lily pad #1. Each second, the frog jumps one pad clockwise with probability 1/4, or two pads clockwise with probability 3/4.

(a) Specify a state space S, initial probabilities $\{\nu_i\}$, and transition probabilities $\{p_{ij}\}$, with respect to which this process is a Markov chain. [**sol**]

(b) Determine if this Markov chain is irreducible. [**sol**]

(c) Determine if this Markov chain is aperiodic, or if not then what its period equals. [**sol**]

(d) Determine whether or not $\sum_{n=1}^{\infty} p_{23}^{(n)} = \infty$. [**sol**]

(e) Either find a stationarity distribution $\{\pi_i\}$ for this chain, or prove that no stationary distribution exists. [**sol**]

(f) Determine whether or not $\lim_{n\to\infty} p_{23}^{(n)}$ exists, and if so what it equals. [**sol**]

(g) Determine whether or not $\lim_{n\to\infty} \frac{1}{2}[p_{23}^{(n)} + p_{23}^{(n+1)}]$ exists, and if so what it equals. [**sol**]

2.5. Periodic Convergence

- What about periodic chains? (e.g. s.r.w., Ehrenfest)
- Recall Example (2.3.2): $S = \{1, 2\}$, and $\nu_1 = 1$, and $(p_{ij}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.
  - This example is irreducible, and has period $b = 2$, and the uniform distribution $\pi_1 = \pi_2 = \frac{1}{2}$ is stationary, but
    $\mathbf{P}(X_n = 1) = \begin{cases} 1, & n \text{ even} \\ 0, & n \text{ odd} \end{cases}$
- Do these transition probabilities converge to a stationary distribution?
  - No! They oscillate between zero and positive values (e.g. at even versus odd times).
  - However, if we average over different times (e.g. both odd and even times), then the average transition probabilities do converge to π.
  - In this example, $\lim_{n\to\infty} \frac{1}{2}[p_{ij}^{(n)} + p_{ij}^{(n+1)}] = \pi_j = 1/2$.
  - Similar convergence holds in general:

(2.5.1) Periodic Convergence Theorem. Suppose a Markov chain is irreducible, with period $b \geq 2$, and stationary distribution $\{\pi_i\}$. Then for all

$i, j \in S$,

$$\lim_{n\to\infty} \frac{1}{b} \, [p_{ij}^{(n)} + ... + p_{ij}^{(n+b-1)}] \;=\; \pi_j \,,$$

and

$$\lim_{n\to\infty} \frac{1}{b} \Big(\mathbf{P}[X_n = j] + \mathbf{P}[X_{n+1} = j] + \ldots + \mathbf{P}[X_{n+b-1} = j] \Big) \;=\; \pi_j \,,$$

and also

$$\lim_{n\to\infty} \frac{1}{b} \, \mathbf{P} \, [X_n = j \text{ or } X_{n+1} = j \text{ or } \ldots \text{ or } X_{n+b-1} = j] \;=\; \pi_j \,.$$

- For example, if $b = 2$, then $\lim_{n\to\infty} \frac{1}{2} \, [p_{ij}^{(n)} + p_{ij}^{(n+1)}] = \pi_j$ for all $i, j \in S$.
  - Of course, in the aperiodic case, by (2.4.1) we also have $\lim_{n\to\infty} \frac{1}{2} \, [p_{ij}^{(n)} + p_{ij}^{(n+1)}] = \pi_j$ for all $i, j \in S$.
- The proof of the Periodic Convergence Theorem is not overly difficult, but it requires a surprisingly large number of steps, so it is deferred to Problems (2.5.6) and (2.5.7) below.
- In particular, Problem (2.5.6) first proves the *Cyclic Decomposition Lemma*, that there is a disjoint partition $S = S_0 \dot\cup S_1 \dot\cup \ldots \dot\cup S_{b-1}$ such that with probability 1, the chain always moves from S_0 to S_1, and then to S_2, and so on to S_{b-1}, and then back to S_0.
- For example, *Ehrenfest's Urn* has period $b = 2$, and oscillates between the two subsets $S_0 = \{\text{even } i \in S\}$ and $S_1 = \{\text{odd } i \in S\}$.
  - And, it satisfies the periodic convergence property that

$$\lim_{n\to\infty} \frac{1}{2} [p_{ij}^{(n)} + p_{ij}^{(n+1)}] \;=\; \pi_j \;=\; 2^{-d} \binom{d}{j} \,.$$

- One type of convergence holds for all irreducible chains with stationary distributions, whether periodic or not:

(2.5.2) Average Probability Convergence. If a Markov chain is irreducible with stationary distribution $\{\pi_i\}$ (whether periodic or not), then for all $i, j \in S$, $\lim_{n\to\infty} \frac{1}{n} \, [p_{ij}^{(1)} + p_{ij}^{(2)} + \ldots + p_{ij}^{(n)}] = \pi_j$, i.e. $\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} p_{ij}^{(\ell)} = \pi_j$.

Proof. This follows from either the usual Markov Chain Convergence Theorem (for aperiodic chains), or the Periodic Markov Chain Convergence Theorem (for chains with period $b \geq 2$), by the *Cesàro sum* principle (A.3.7): if a sequence converges, then its partial averages also converge to the same value. ∎

- One use of this result is to generalise (2.4.6) to periodic chains:

(2.5.3) Unique Stationary Corollary. If Markov chain P is irreducible (whether periodic or not), then it has at most one stationary distribution.

Proof. By Average Probability Convergence (2.5.2), any stationary distribution which exists must be equal to $\lim_{n\to\infty} \frac{1}{n} \, [p_{ij}^{(1)} + p_{ij}^{(2)} + \ldots + p_{ij}^{(n)}]$, so they're all equal. ∎

(2.5.4) Problem. Let $S = \{1, 2, 3\}$, $p_{11} = 1$, and $p_{22} = p_{23} = p_{32} = p_{33} = 1/2$. For $a \in [0, 1]$, let π^a be the probability distribution on S defined by $\pi^a_1 = a$, and $\pi^a_2 = \pi^a_3 = (1 - a)/2$.

(a) Prove that for any $a \in [0, 1]$, π^a is a stationary distribution.

(b) How many stationary distributions does this chain have?

(c) Why does this result not contradict the Unique Stationary Corollary?

(2.5.5) Problem. Consider the Markov chain with state space $S = \{1, 2, 3\}$, and transition probabilities $p_{12} = p_{32} = 1$, $p_{21} = 1/4$, and $p_{23} = 3/4$, with $p_{ij} = 0$ otherwise. Let $\pi_1 = 1/8$, $\pi_2 = 1/2$, and $\pi_3 = 3/8$.

(a) Verify that the chain is reversible with respect to π. **[sol]**

(b) Determine (with explanation) which of the following statements are true and which are false: (i) $\lim_{n\to\infty} p_{11}^{(n)} = 1/8$. (ii) $\lim_{n\to\infty} \frac{1}{2}[p_{11}^{(n)} + p_{11}^{(n+1)}] = 1/8$. (iii) $\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} p_{11}^{(\ell)} = 1/8$. **[sol]**

(2.5.6) Problem. The *Cyclic Decomposition Lemma* says that for an irreducible Markov chain with period $b \geq 2$, then there is a "cyclic" disjoint partition $S = S_0 \dot\cup S_1 \dot\cup \ldots \dot\cup S_{b-1}$ such that if $i \in S_r$ for some $0 \leq r \leq b-2$, then $\sum_{j\in S_{r+1}} p_{ij} = 1$, while if $i \in S_{b-1}$, then $\sum_{j\in S_0} p_{ij} = 1$. (That is, the chain is forced to repeatedly move from S_0 to S_1 to S_2 to $\ldots$ to S_{b-1} and then back to S_0.) Furthermore, the b-step chain $P^{(b)}$, when restricted to S_0, is irreducible and aperiodic. Prove this lemma, using the following steps.

(a) Fix $i_0 \in S$, and let $S_r = \{j \in S : p_{i_0 j}^{(bm+r)} > 0$ for some $m \in \mathbf{N}\}$. Show that $\bigcup_{r=0}^{b-1} S_r = S$. [Hint: use irreducibility.]

(b) Show that if $0 \leq r < t \leq b-1$, then S_r and S_t are disjoint, i.e. $S_r \cap S_t = \emptyset$. [Hint: Suppose $j \in S_r \cap S_t$. Find $m \in \mathbf{N}$ with $p_{ji}^{(m)} > 0$. What can you conclude about $\gcd\{n \geq 1 : p_{ii}^{(n)} > 0\}$?]

(c) Let $i \in S_r$ for some $0 \leq r \leq b-2$, and $p_{ij} > 0$. Prove that $j \in S_{r+1}$.

(d) Suppose $i \in S_{b-1}$, and $p_{ij} > 0$. Prove that $j \in S_0$.

(e) Let $\widehat{P} = P^{(b)}|_{S_0}$, i.e. $\widehat{p}_{ij} = p_{ij}^{(b)}$ corresponding to b steps of the original chain, except restricted to just $i, j \in S_0$. Prove that $\widehat{P}$ is irreducible on S_0. [Hint: Suppose there are $i, j \in S_0$ with $\widehat{p}_{ij}^{(n)} = 0$ for all $n \geq 1$. What does this say about f_{ij} for the original chain?]

(f) Prove that $\widehat{P}$ is aperiodic. [Hint: Suppose there is $i \in S_0$ with $gcd\{n \geq 1 : \widehat{p}_{ii}^{(n)} > 0\} = m \geq 2$. What does this say about the period of i in the original chain?]

(2.5.7) Problem. Complete the proof of the *Periodic Convergence Theorem*, by the following steps. Assume $\{p_{ij}\}$ are the transition probabilities for an irreducible Markov chain on a state space S with period $b \geq 2$, and stationary distribution $\{\pi_i\}$. Let $S_0, S_1, \ldots, S_{b-1}$ and $\widehat{P}$ be as in *Cyclic Decomposition Lemma* in the previous problem. For any subset $A \subseteq S$, write $\pi(A) = \sum_{i\in A} \pi_i$ for the total probability of A according to π.

(a) Prove that $\pi(S_0) = \pi(S_1) = \ldots = \pi(S_{b-1}) = 1/b$. [Hint: What is the relationship between $\mathbf{P}[X_0 \in S_0]$ and $\mathbf{P}[X_1 \in S_1]$? On the other hand, if we begin in stationarity, then how does $\mathbf{P}[X_n \in S_0]$ change with n?]

(b) Let $\widehat{\pi}_i = b\,\pi_i$ for all $i \in S_0$. Show that $\widehat{\pi}$ is a stationary distribution for $\widehat{P}$.

(c) Conclude that $\lim_{m\to\infty} \widehat{p}_{ij}^{(m)} = \widehat{\pi}_j$ for all $i, j \in S_0$, i.e. $\lim_{m\to\infty} p_{ij}^{(bm)} = b\,\pi_j$.

(d) Prove that similarly $\lim_{m\to\infty} p_{ij}^{(bm)} = b\,\pi_j$ for all $i, j \in S_r$, for any $1 \le r \le b-1$.

(e) Show that for $i \in S_0$ and $j \in S_r$ for any $1 \le r \le b-1$, we have $\lim_{m\to\infty} p_{ij}^{(bm+r)} = b\,\pi_j$. [Hint: $p_{ij}^{(bm+r)} = \sum_{k\in S} p_{ik}^{(r)} p_{kj}^{(bm)}$.]

(f) Conclude that for $i \in S_0$, $\lim_{m\to\infty} \frac{1}{b}[p_{ij}^{(bm)} + p_{ij}^{(bm+1)} + \ldots + p_{ij}^{(bm+b-1)}] = \pi_j$ for any $j \in S$.

(g) Show that the previous statement holds for any $i \in S$ (not just $i \in S_0$).

(h) Show that $\lim_{n\to\infty} \frac{1}{b}\,[p_{ij}^{(n)} + \ldots + p_{ij}^{(n+b-1)}] = \pi_j$, i.e. that we can take the limit over all n not just $n = bm$. [Hint: for any n, let $m = \lfloor n/b \rfloor$ be the *floor* of n/b, i.e. the greatest integer not exceding n/b.]

2.6. Application – MCMC Algorithms

- Let $S = \mathbf{Z}$, or more generally let S be any continguous subset of $\mathbf{Z}$, e.g. $S = \{1,2,3\}$, or $S = \{-5,-4,\ldots,17\}$, or $S = \mathbf{N}$, etc.
  - Let $\{\pi_i\}$ be any probability distribution on S.
  - Assume for simplicity that $\pi_i > 0$ for all $i \in S$.
- Suppose we want to *sample* from π, i.e. create a random variable X with $\mathbf{P}(X = i) \approx \pi_i$ for all $i \in S$. (This method is called "*Monte Carlo*".)
  - One method: Create a Markov chain $X_0, X_1, X_2, \ldots$ such that $\lim_{n\to\infty} \mathbf{P}(X_n = i) = \pi_i$ for all $i \in S$. ("*Markov chain Monte Carlo (MCMC)*")
  - Can we do this? Can we create Markov chain transitions $\{p_{ij}\}$ so that $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$?
- Yes! One way is the *Metropolis algorithm*†.
- Let $p_{i,i+1} = \frac{1}{2}\min[1,\ \frac{\pi_{i+1}}{\pi_i}]$, $p_{i,i-1} = \frac{1}{2}\min[1,\ \frac{\pi_{i-1}}{\pi_i}]$, and $p_{i,i} = 1 - p_{i,i+1} - p_{i,i-1}$, with $p_{ij} = 0$ otherwise.
  - (We take $\pi_j = 0$ if $j \notin S$.)
  - An equivalent algorithmic version is: Given X_{n-1}, let Y_n equal $X_{n-1} \pm 1$ (probability $1/2$ each), and let $U_n \sim \text{Uniform}[0,1]$ as in (A.2.5), with the $\{U_n\}$ chosen i.i.d., and then let

$$X_n = \begin{cases} Y_n, & U_n < \frac{\pi_{Y_n}}{\pi_{X_{n-1}}} \quad (\text{"accept"}) \\ X_{n-1}, & \text{otherwise} \quad (\text{"reject"}) \end{cases}$$

  - An animated illustration is at: www.probability.ca/met
- Does this chain have $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$?
  - Here $\pi_i p_{i,i+1} = \pi_i\,\frac{1}{2}\min[1,\ \frac{\pi_{i+1}}{\pi_i}] = \frac{1}{2}\min[\pi_i,\ \pi_{i+1}]$.
  - Also $\pi_{i+1} p_{i+1,i} = \pi_{i+1}\,\frac{1}{2}\min[1,\ \frac{\pi_i}{\pi_{i+1}}] = \frac{1}{2}\min[\pi_{i+1},\ \pi_i]$.

---

†First introduced by N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equations of state calculations by fast computing machines", J. Chem. Phys. **21**, 1087–1091, way back in 1953 (!). For an overview of later research in this area, see e.g. S. Brooks, A. Gelman, G.L. Jones, and X.-L. Meng, eds., "Handbook of Markov chain Monte Carlo", Chapman & Hall / CRC Press, 2011.

- So $\pi_i p_{ij} = \pi_j p_{ji}$ if $j = i+1$, hence for all $i, j \in S$ (since otherwise both sides are 0).
- So, the chain is reversible w.r.t. $\{\pi_i\}$.
- So, $\{\pi_i\}$ is stationary.
- Also, the chain is easily checked to be irreducible and aperiodic.
- So, $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$, and $\lim_{n\to\infty} \mathbf{P}[X_n = j] = \pi_j$, for all i and j and ν.
- Hence, for "large enough" n, X_n is approximately a *sample* from π.

- Such Markov chains are very widely used to sample from complicated distributions $\{\pi_i\}$, to estimate their probabilities and expected values.
  - Indeed, "markov chain monte carlo" gives over a million hits in Google!
  - Some of the applications are on continuous state spaces, with π a *density function* (e.g. a *Bayesian posterior density*).
  - And some of the applications use more general versions of the Metropolis algorithm (see Problem (2.6.3)), or related MCMC algorithms such as the *Gibbs sampler* (see Problem (2.6.4)).

(2.6.1) Problem. Let $S = \{1, 2, 3\}$, with $\pi_1 = 1/2$ and $\pi_2 = 1/3$ and $\pi_3 = 1/6$. Find (with proof) irreducible transition probabilities $\{p_{ij}\}_{i,j\in S}$ such that π is a stationarity distribution. **[sol]**

(2.6.2) Problem. Let $S = \mathbf{N}$, with $\pi_i = 2/3^i$ for each $i \in S$. Find (with proof) irreducible transition probabilities $\{p_{ij}\}_{i,j\in S}$ such that π is a stationarity distribution.

(2.6.3) Problem. [*General Metropolis algorithm.*] Let S be a contiguous subset of $\mathbf{Z}$, let π be a probability distribution on S with $\pi_i > 0$ for all $i \in S$, and let $q : S \times S \to \mathbf{R}^+$ be a *proposal distribution* with $\sum_j q(i, j) = 1$ for all $i \in S$. Assume q is *symmetric*, i.e. $q(i, j) = q(j, i)$ for all $i, j \in S$. For $i \neq j$, let $p_{ij} = q(i, j) \min[1, \frac{\pi_j}{\pi_i}]$, and let $p_{ii} = 1 - \sum_{j\neq i} p_{ij}$ be the leftover probabilities. (As before, we take $\pi_j = 0$ if $j \notin S$.)

(a) Prove that $(\, p_{ij}\,)$ is a valid Markov chain transition matrix.

(b) Prove that this Markov chain is reversible with respect to π.

(c) Specify conditions which guarantee that $\lim_{n\to\infty} \mathbf{P}(X_n = j) = \pi_j$ for all $j \in S$. [Hint: Don't forget (2.4.1).]

(2.6.4) Problem. [*Metropolis-Hastings algorithm.*] Let S, π, and q be as in Problem (2.6.3), except do not assume that q is symmetric, merely that it is symmetrically positive, i.e. $q(i, j) > 0$ iff $q(j, i) > 0$. For $i \neq j$, let $p_{ij} = q(i, j) \min[1, \frac{\pi_j q(j,i)}{\pi_i q(i,j)}]$, and again let $p_{ii} = 1 - \sum_{j\neq i} p_{ij}$.

(a) Prove that $(\, p_{ij}\,)$ is a valid Markov chain transition matrix.

(b) Prove that this Markov chain is reversible with respect to π.

(c) Specify conditions which guarantee that $\lim_{n\to\infty} \mathbf{P}(X_n = j) = \pi_j$ for all $j \in S$.

(2.6.5) Problem. [*Gibbs sampler.*] Let $S = \mathbf{Z} \times \mathbf{Z}$, and let $f : S \to (0, \infty)$ be some function from S to the positive real numbers. Let $K = \sum_{(x,y)\in S} f(x, y)$, and assume that $K < \infty$. For $x, y \in \mathbf{Z}$, let $C(x) = \sum_{w\in\mathbf{Z}} f(x, w)$, and $R(y) = \sum_{z\in\mathbf{Z}} f(z, y)$. Consider an algorithm which proceeds at each time n as follows. Given a pair $(X_{n-1}, Y_{n-1}) \in S$ at time $n-1$, it chooses either the "horizontal" or "vertical" option, with probability 1/2 each. If it chooses horizontal, then it sets $Y_n = Y_{n-1}$, and chooses X_n randomly to equal x with probability $f(x, Y_{n-1}) / R(Y_{n-1})$ for each $x \in \mathbf{Z}$. If it chooses vertical, then it sets $X_n = X_{n-1}$, and chooses Y_n randomly to equal y with probability $f(X_{n-1}, y) / C(X_{n-1})$ for each $y \in \mathbf{Z}$.

(a) Verify that the resulting sequence $\{(X_n, Y_n)\}_{n \in \mathbf{N}}$ has transition probabilities given by

$$p_{(x,y),(z,w)} \;=\; \begin{cases} \frac{f(z,w)}{2\,C(x)} + \frac{f(z,w)}{2\,R(y)}\,, & x = z \text{ and } y = w \\ \frac{f(z,w)}{2\,C(x)}\,, & x = z \text{ and } y \neq w \\ \frac{f(z,w)}{2\,R(y)}\,, & x \neq z \text{ and } y = w \\ 0\,, & \text{otherwise} \end{cases}$$

(b) Verify directly that $\sum_{(z,w) \in S} p_{(x,y),(z,w)} = 1$ for all $(x, y) \in S$.

(c) Show that the chain is reversible with respect to $\pi_{(x,y)} = \frac{f(x,y)}{K}$.

(d) Compute $\lim_{n \to \infty} p^{(n)}_{(x,y),(z,w)}$ for all $x, y, z, w \in \mathbf{Z}$ (with justification).

2.7. Application – Random Walks on Graphs

- Let V be a non-empty finite or countable set.
- Let $w : V \times V \to [0, \infty)$ be a symmetric weight function (i.e. $w(u, v) = w(v, u)$).
  - Usual (unweighted graph) case: $w(u, v) = 1$ if there is an edge between u and v, otherwise $w(u, v) = 0$. (diagram)
  - Or can have other weights, multiple edges, self-edges ($w(u, u) > 0$), etc.
- Let $d(u) = \sum_{v \in V} w(u, v)$ be the *degree* of the vertex u.
  - Assume that $d(u) > 0$ for all $u \in V$ (for example, by giving any isolated point a self-edge.
  - Then we can define:

(2.7.1) Definition. Given a vertex set V with symmetric weights w, the *(simple) random walk on the (undirected) graph* (V, w) is the Markov chain with state space $S = V$ and transition probabilities $p_{uv} = \frac{w(u,v)}{d(u)}$ for all $u, v \in V$.

- It follows (check!) that $\sum_{v \in V} p_{uv} = \frac{\sum_{v \in V} w(u,v)}{\sum_{v \in V} w(u,v)} = 1$, as it must.

- The most common case is where each $w(u, v) = 0$ or 1, so from u, the chain moves to one of the $d(u)$ vertices connected to u with equal probability.

(2.7.2) Ring Graph. Suppose $V = \{1, 2, 3, 4, 5\}$, with $w(i, i + 1) = w(i + 1, i) = 1$ for $i = 1, 2, 3, 4$, and $w(5, 1) = w(1, 5) = 1$, with $w(i, j) = 0$ otherwise (see Figure 9).

- Is random walk on this graph irreducible? Is it aperiodic? Does it have a stationarity distribution? Do its transition probabilities converge?

(2.7.3) Stick Graph. Suppose $V = \{1, 2, \ldots, K\}$, with $w(i, i+1) = w(i + 1, i) = 1$ for $1 \le i \le K - 1$, with $w(i, j) = 0$ otherwise (see Figure 10).

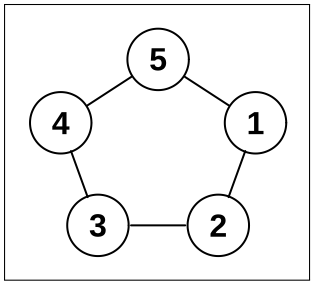


Figure 9: A diagram of the Ring Graph.

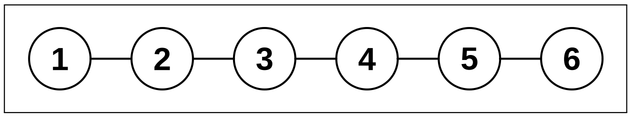


Figure 10: A diagram of the Stick Graph, with $K = 6$.

– Does it have a stationary distribution? etc.

(2.7.4) Star Graph. Suppose $V = \{0, 1, 2, \ldots, K\}$, with $w(i, 0) = w(0, i) = 1$ for $i = 1, 2, 3$, with $w(i, j) = 0$ otherwise. (see Figure 11).

– Does it have a stationary distribution? Will it converge?

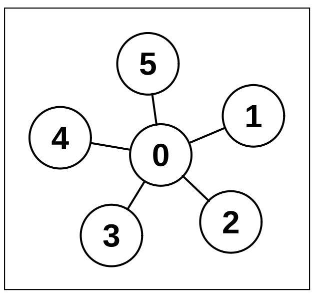


Figure 11: A diagram of the Star Graph, with $K = 5$.

(2.7.5) Infinite Graph. Suppose $V = \mathbf{Z}$, with $w(i, i+1) = w(i+1, i) = 1$ for all $i \in V$, and $w(i, j) = 0$ otherwise. (diagram)

– Random walk on this graph corresponds exactly to simple symmetric random walk.

(2.7.6) Frog Graph. Suppose $V = \{1, 2, \ldots, K\}$, with $w(i, i) = 1$ for $1 \le i \le K$, and $w(i, i+1) = w(i+1, i) = 1$ for $1 \le i \le K - 1$, and $w(K, 1) = w(1, K) = 1$, and $w(i, j) = 0$ otherwise (see Figure 13).

– In fact, when $K = 20$, random walk on this graph corresponds exactly to the original Frog Example from Section 1.1.

- Do these walks have stationary distributions?
  Usually yes!
  – Let $Z = \sum_{u \in V} d(u) = \sum_{u,v \in V} w(u, v)$.
  – In the unweighted case, Z equals two times the number of edges.
  – (Except that self-edges are only counted once, not twice.)
  – Here Z might be infinite, if V is infinite.
  – But if $Z < \infty$, then we have a precise formula for a stationary distribution:

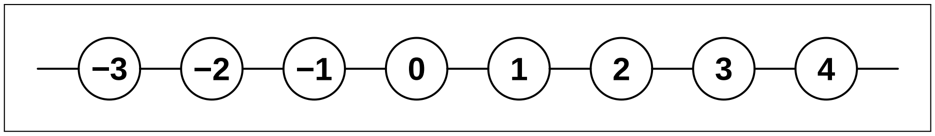


Figure 12: A diagram of part of the Infinite Graph (2.7.5).

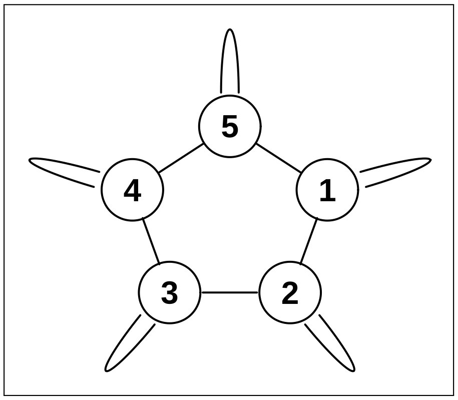


Figure 13: A diagram of the Frog Graph with $K = 5$.

(2.7.7) Graph Stationary Distribution. Consider a random walk on a graph V with degrees $d(u)$. Assume that Z is finite. Then if $\pi_u = \frac{d(u)}{Z}$, then π is a stationary distribution for this walk.

Reminder: (again) Midterm #1 next Thurs, Feb 7, at 6:10 pm, in MS 2170 [surnames A-K] and MS 3154 [surnames L-Z].

END OF WEEK #4

(Midterm #1.)

END OF WEEK #5

Proof. It is easily checked that $\pi_u \geq 0$, and $\sum_u \pi_u = 1$, so π is a probability distribution on V.

- Also, $\pi_u p_{uv} = \frac{d(u)}{Z} \frac{w(u,v)}{d(u)} = \frac{w(u,v)}{Z}$.
- And, $\pi_v p_{vu} = \frac{d(v)}{Z} \frac{w(v,u)}{d(v)} = \frac{w(v,u)}{Z} = \frac{w(u,v)}{Z}$. Equal!
- So, the chain is reversible w.r.t. π.
- So, by Proposition (2.2.2), π is a stationary distribution. ■

- Thus, (2.7.7) gives a precise formula for the stationary distribution of almost any (simple) random walk on any (undirected) graph. This is quite striking, since no such formula exists for most other Markov chains.
  - (However, in e.g. the Infinite Graph (2.7.5) corresponding to s.s.r.w., there is no stationary distribution, and (2.7.7) does not apply, because $Z = \infty$.)
- What about irreducibility?
  - Well, if the graph is *connected* (i.e. it is possible to move from any vertex to any other vertex through edges of positive weight), then the chain is clearly irreducible.

- What about periodicity?
  - Well, the walk can always return to any vertex u in 2 steps, by moving to any other vertex and then back along the same edge.
  - So, 1 and 2 are the only possible periods.
  - If the graph is a *bipartite graph* (i.e., it can be divided into two subsets such that all the links go from one subset to the other one), then the chain has period 2.
  - But if the chain is not bipartite, then it is aperiodic.
  - Furthermore, if there is any cycle of odd length, then the graph cannot be bipartite, so it must be aperiodic.
- We conclude:

(2.7.8) Graph Convergence Theorem. For a random walk on a connected non-bipartite graph, if $Z < \infty$, then $\lim_{n\to\infty} p_{uv}^{(n)} = \frac{d(v)}{Z}$ for all $u, v \in V$, and $\lim_{n\to\infty} \mathbf{P}[X_n = v] = \frac{d(v)}{Z}$ (for any initial probabilities).

Proof. By the above, if it is connected then it is irreducible, and if it is non-bipartite then it is aperiodic.

  - Also, if $Z < \infty$ then $\pi_v = \frac{d(v)}{Z}$ is a stationary distribution by (2.7.7).
- Hence, by the Markov Chain Convergence Theorem (2.4.1), for all $u, v \in V$, $\lim_{n\to\infty} p_{uv}^{(n)} = \pi_v = \frac{d(v)}{Z}$. ■

- What about graphs which might not be bipartite?
  - Since the only possible periods are 1 and 2, the Periodic Convergence Theorem (2.5.1) and Average Probability Convergence (2.5.2) give:

(2.7.9) Graph Average Convergence. For a random walk on any connected graph with $Z < \infty$ (whether bipartite or not), for all $u, v \in V$, $\lim_{n\to\infty} \frac{1}{2}[p_{uv}^{(n)} + p_{uv}^{(n+1)}] = \frac{d(v)}{Z}$, and $\lim_{n\to\infty} \frac{1}{n}\sum_{\ell=1}^{n} p_{uv}^{(\ell)} = \frac{d(v)}{Z}$.

- Consider again the Stick Graph (2.7.3), with $V = \{1, 2, \ldots, K\}$, and $w(i, i+1) = w(i+1, i) = 1$ for $1 \le i \le K-1$, and $w(i, j) = 0$ otherwise.
  - This graph is clearly connected, but is bipartite (since every edge goes between and even vertex and an odd vertex).
  - Also $d(u) = 1$ for $i = 1$ or K, otherwise $d(u) = 2$.
  - Hence, $Z = 1 + 2 + 2 + \ldots + 2 + 1 = 1 + 2(K-2) + 1 = 2K - 2$.
  - Then if $\pi_i = \frac{d(i)}{Z} = \frac{1}{2K-2}$ for $i = 1$ or K, and $\pi_i = \frac{2}{2K-2}$ for $2 \le i \le K-1$, then π is a stationary distribution by (2.7.7).
  - Then, we must have $\lim_{n\to\infty} \frac{1}{2}[p_{ij}^{(n)} + p_{ij}^{(n+1)}] = \pi_j$ for all $j \in V$.

(2.7.10) Problem. Consider again the Star Graph (2.7.4).

(a) Is this random walk irreducible?

(b) What periods do the states of this random walk have?

(c) Does $\lim_{n\to\infty} p_{uv}^{(n)}$ exist, and if so then what does it equal?

(d) Does $\lim_{n\to\infty} \frac{1}{2}[p_{uv}^{(n)} + p_{uv}^{(n+1)}]$ exist, and if so then what does it equal?

(2.7.11) Problem. Repeat Problem (2.7.10) except with an extra edge between 0 and itself, i.e. with $w(0, 0) = 1$ instead of 0.

(2.7.12) Problem. Consider the undirected graph with vertex set $V = \{1, 2, 3, 4\}$, and an undirected edge (of weight 1) between each of the following four pairs of edges (and no other edges): (1,2), (2,3), (3,4), and (2,4). Let $\{p_{uv}\}_{u,v \in V}$ be the transition probabilities for random walk on this graph. Compute (with full explanation) $\lim_{n\to\infty} p_{21}^{(n)}$, or prove that this limit does not exist. [**sol**]

(2.7.13) Problem. Consider the undirected graph on the vertices $V = \{1, 2, 3, 4, 5\}$, with weights given by $w(1, 2) = w(2, 1) = w(2, 3) = w(3, 2) = w(1, 3) = w(3, 1) = w(3, 4) = w(4, 3) = w(3, 5) = w(5, 3) = 1$, and $w(u, v) = 0$ otherwise.

(a) Draw a picture of this graph.

(b) Compute (with full explanation) $\lim_{n\to\infty} \mathbf{P}[X_n = 3]$, where $\{X_n\}$ is the usual (simple) random walk on this graph.

2.8. Application – Gambler's Ruin

- Consider the following gambling game.
  - Let $0 < a < c$ be integers, and let $0 < p < 1$.
  - Suppose player A starts with a dollars, and player B starts with $c - a$ dollars, and they repeatedly bet.
  - At each bet, A wins \$1 from B with probability p, or B wins \$1 from A with probability $1 - p$.
- If X_n is the amount of money that A has at time n, then clearly $X_0 = a$, and $\{X_n\}$ follows a simple random walk (1.3.6).
- Let $T_i = \inf\{n \geq 0 : X_n = i\}$ be the first time A has i dollars.
- The *Gambler's Ruin* question is: what is $\mathbf{P}_a(T_c < T_0)$, i.e. what is the probability that A reaches c dollars <u>before</u> reaching 0 (i.e., before losing all their money)?
  - For an animated version, see: www.probability.ca/gambler
- Some specific examples are:
  - What is the probability if $c = 10,000$, $a = 9,700$, and $p = 0.5$?
  - What if instead $c = 10,000$, $a = 9,700$, and $p = 0.49$?
  - Is the probability higher if $c = 8$, $a = 6$, $p = 1/3$ ("born rich"), or if $c = 8$, $a = 2$, $p = 2/3$ ("born lucky")?
- Here $\{X_n\}$ is a Markov chain.
  - However, there is no limit to how long the game might take.
  - So, how can we find this probability?
- But what is the player's win probability $\mathbf{P}_a(T_c < T_0)$?
- Key: Write $\mathbf{P}_a(T_c < T_0)$ as $s(a)$, and consider it to be a <u>function</u> of the player's initial fortune a.
  - Can we related the different unknown $s(a)$ to each other? Yes!
- Clearly $s(0) = 0$, and $s(c) = 1$.
- Furthermore, on the <u>first</u> bet, A either wins or loses \$1.
  - So, for $1 \leq a \leq c - 1$, by conditioning (A.5.4),

$$s(a) = \mathbf{P}_a(T_c < T_0)$$

$$= \mathbf{P}_a(T_c < T_0,\ X_1 = X_0 + 1) + \mathbf{P}_a(T_c < T_0,\ X_1 = X_0 - 1)$$
$$= \mathbf{P}(X_1 = X_0 + 1)\ \mathbf{P}_a(T_c < T_0 \,|\, X_1 = X_0 + 1)$$
$$+ \mathbf{P}(X_1 = X_0 - 1)\ \mathbf{P}_a(T_c < T_0 \,|\, X_1 = X_0 - 1)$$
$$= p\, s(a+1) + (1-p)\, s(a-1)\,.$$

- This gives $c - 1$ equations for the $c - 1$ unknowns.
  - We can solve this using simple algebra!
- Re-arranging, $p\, s(a) + (1-p)\, s(a) = p\, s(a+1) + (1-p)\, s(a-1)$.
  - Hence, $s(a+1) - s(a) = \frac{1-p}{p}\,[s(a) - s(a-1)]$.
  - Let $x = s(1)$ (an unknown quantity).
  - Then $s(1) - s(0) = x$, and $s(2) - s(1) = \frac{1-p}{p}[s(1) - s(0)] = \frac{1-p}{p}x$.
  - Then $s(3) - s(2) = \frac{1-p}{p}[s(2) - s(1)] = \left(\frac{1-p}{p}\right)^2 x$.
  - In general, for $1 \le a \le c-1$, $s(a+1) - s(a) = \left(\frac{1-p}{p}\right)^a x$.
  - Hence, for $1 \le a \le c-1$,

$$s(a) = s(a) - s(0)$$

$$= [s(a) - s(a-1)] + [s(a-1) - s(a-2)] + \ldots + [s(1) - s(0)]$$

$$= \left[\left(\frac{1-p}{p}\right)^{a-1} + \left(\frac{1-p}{p}\right)^{a-2} + \ldots + \left(\frac{1-p}{p}\right) + 1\right] x$$

$$= \begin{cases} \left[\dfrac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right) - 1}\right] x, & p \neq 1/2 \\ ax, & p = 1/2 \end{cases}$$

  - But $s(c) = 1$, so we can solve for x:

$$x \;=\; \begin{cases} \dfrac{\left(\frac{1-p}{p}\right) - 1}{\left(\frac{1-p}{p}\right)^c - 1}, & p \neq 1/2 \\ 1/c, & p = 1/2 \end{cases}$$

- We then obtain our final *Gambler's Ruin formula*:

(2.8.1)
$$s(a) \;=\; \begin{cases} \dfrac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1}, & p \neq 1/2 \\ a/c, & p = 1/2 \end{cases}$$

thus finally solving the Gambler's Ruin problem as desired.

- Some specific examples are:
  - If $c = 10,000$, $a = 9,700$, and $p = 0.5$, then

$$s(a) \;=\; a/c \;=\; 0.97\,.$$

  - Or, if $c = 10,000$ and $a = 9,700$, but $p = 0.49$, then

$$s(a) \;=\; \frac{\left(\frac{0.51}{0.49}\right)^{9,700} - 1}{\left(\frac{0.51}{0.49}\right)^{10,000} - 1} \;\doteq\; 0.000006134 \;\doteq\; \frac{1}{163,000}\,.$$

- So, changing p from 0.5 to 0.49 makes a huge difference!

- Also, if $c = 8$, $a = 6$, and $p = 1/3$ ("born rich"), then

$$s(a) \;=\; \frac{\left(\frac{2/3}{1/3}\right)^6 - 1}{\left(\frac{2/3}{1/3}\right)^8 - 1} \;=\; \frac{63}{255} \;\doteq\; 0.247\,,$$

but if $c = 8$, $a = 2$, and $p = 2/3$ ("born lucky"), then

$$s(a) \;=\; \frac{\left(\frac{1/3}{2/3}\right)^2 - 1}{\left(\frac{1/3}{2/3}\right)^8 - 1} \;=\; \frac{3/4}{255/256} \;\doteq\; 0.753\,.$$

So, it is better to be born lucky than rich!

- We will sometimes write $s(a)$ as $s_{c,p}(a)$, to show the explicit dependence on c and p.

(2.8.2) Problem. [Calculus Challenge] Is $s(a)$ continuous as a function of p, as $p \to 1/2$? That is, is it true that $\lim_{p\to 1/2} s_{c,p}(a) = s_{c,1/2}(a)$? [Hint: Don't forget *L'Hôpital's Rule*.]

- We can also consider $r_{c,p}(a) = \mathbf{P}_a(T_0 < T_c) = \mathbf{P}_a(\text{ruin})$.
  - Then $r_{c,p}(a)$ is like $s(a)$, but from the other player's perspective.
  - Hence, we need to replace a by $c - a$, and replace p by $1 - p$.
- It follows that:

$$\text{(2.8.3)} \qquad r_{c,p}(a) \;=\; s_{c,1-p}(c-a) \;=\; \begin{cases} \dfrac{\left(\frac{p}{1-p}\right)^{c-a} - 1}{\left(\frac{p}{1-p}\right)^{c} - 1}, & p \neq 1/2 \\ (c-a)/c, & p = 1/2 \end{cases}$$

(2.8.4) Problem. Check directly that $r_{c,p}(a) + s_{c,p}(a) = 1$, thus showing that Gambler's Ruin must eventually end.

- Next we consider $\mathbf{P}_a(T_0 < \infty)$, the probability of <u>eventual</u> ruin.
  - Clearly $\lim_{c\to\infty} T_c = \infty$ (since e.g. $T_c \geq c - a$).
  - Hence, by continuity of probabilities (A.7.3) and then (2.8.3), we can compute the probability of eventual ruin as:

$$\mathbf{P}_a(T_0 < \infty) \;=\; \lim_{K\to\infty} \mathbf{P}_a(T_0 < K)$$

$$=\; \lim_{c\to\infty} \mathbf{P}_a(T_0 < T_c) \;=\; \lim_{c\to\infty} r_{c,p}(a)$$

$$\text{(2.8.5)} \qquad =\; \begin{cases} \frac{0-1}{0-1} = 1, & p < 1/2 \\ \frac{1}{1} = 1, & p = 1/2 \\ \left(\frac{p}{1-p}\right)^{-a}, & p > 1/2 \end{cases}$$

- So, eventual ruin is certain if $p \leq 1/2$, but not if $p > 1/2$.
- For example, suppose $p = 2/3$ and $a = 2$.
  - Then $\mathbf{P}(T_0 < \infty) = \left(\frac{2/3}{1-(2/3)}\right)^{-2} = 2^{-2} = 1/4$.

– Hence, $\mathbf{P}(T_0 = \infty) = 1 - \mathbf{P}(T_0 < \infty) = 3/4$.
– That is, if we start with \$2, and have probability 2/3 of winning each bet, then we have probability 3/4 of never going broke.

- Finally, we consider the time $T = \min(T_0, T_c)$ when the Gambler's Ruin game ends.
  – It follows from Problem (2.8.4) that the game must end eventually, with probability 1.
  – But a crude bound on the tail probabilities of T proves more directly that T is finite and has finite expectation:

(2.8.6) Proposition. Let $T = \min(T_0, T_c)$ be the time when the Gambler's Ruin game ends. Then $\mathbf{P}(T > mc) \le (1 - p^c)^m$, and $\mathbf{P}(T = \infty) = 0$, and $\mathbf{E}(T) < \infty$.

Proof. If the player ever wins c bets in a row, then the game must be over (since if they haven't already lost, then they will have reached c).

– So, if $T > mc$, then the player has failed to win c bets in a row, despite having m independent attempts to do so.
– But the probability of winning c bets in a row is p^c.
– So, the probability of failing to win c bets in a row is $1 - p^c$.
– So, the probability of failing on m independent attemps is $(1 - p^c)^m$.
– Hence, $\mathbf{P}(T > mc) \le (1 - p^c)^m$, as claimed.

- Then, by continuity of probabilities (A.7.3),

$$\mathbf{P}(T = \infty) \;=\; \lim_{m\to\infty} \mathbf{P}(T > mc) \;\le\; \lim_{m\to\infty} (1 - p^c)^m \;=\; 0\,.$$

- Finally, using the trick (A.2.6), and a geometric series (A.3.1),

$$\mathbf{E}(T) \;=\; \sum_{i=1}^{\infty} \mathbf{P}(T \ge i) \;\le\; \sum_{i=0}^{\infty} \mathbf{P}(T \ge i)$$

$$= \;\mathbf{P}(T \ge 0) + \mathbf{P}(T \ge 1) + \mathbf{P}(T \ge 2) + \mathbf{P}(T \ge 3) + \mathbf{P}(T \ge 4) + \ldots$$

$$\le \;\mathbf{P}(T \ge 0) + \mathbf{P}(T \ge 0) + \ldots + \mathbf{P}(T \ge 0) + \mathbf{P}(T \ge c) + \ldots$$

$$= \;\sum_{j=0}^{\infty} c\,\mathbf{P}(T \ge cj) \;\le\; \sum_{j=0}^{\infty} c\,(1 - p^c)^j$$

$$= \;\frac{c}{1 - (1 - p^c)} \;=\; \frac{c}{p^c} \;<\; \infty\,. \qquad \blacksquare$$

- That is, with probability 1 the Gambler's Ruin game must eventually end, and the time it takes to end has finite expected value.

(2.8.7) Remark. A *betting strategy* (or, *gambling strategy*) involves wagering different amounts on different bets. One common example is the *double 'til you win* strategy (also sometimes called a *martingale* strategy, though it is different from the martingales discussed in the next chapter): First you wager \$1, then if you lose you wager \$2, then if you lose again you wager \$4, etc. As soon as you win any one bet you stop, with total net gain always equal to \$1. For any $p > 0$, with probability 1 you will eventually win a bet, so this appears to guarantee a \$1 profit. (And then by repeating the strategy, or scaling up the wagers, you can guarantee larger profits too.) What's the catch? Well, you might have a run of very bad luck, and reach your credit limit, and have to stop with a very negative net gain. (For example, if you happen to lose your first 30 bets, then you will be down over a billion dollars, with little hope of recovery!) Indeed, it can be proven that provided there is some finite limit to how much you can lose, then if $p \leq 1/2$ then your expected net gain can never be positive; see e.g. Section 7.3 of Rosenthal (2006), or Dubins and Savage (2014). See also Example (3.3.7) below.

(2.8.8) Problem. Suppose a gambler starts with \$7, and repeatedly bets \$1, with (independent) probability 2/5 of winning each bet.

(a) Compute the probability that the gambler will reach \$10 before losing all their money.

(b) Compute the probability that the gambler will reach \$20 before losing all their money.

(c) Compute the probability that the gambler will never lose all their money.

(2.8.9) Problem. Suppose $p = 0.49$ and $c = 10,000$. Find the smallest value of a such that $s_{c,p}(a) \geq 1/2$. Interpret your answer in terms of a gambler at a casino.

(2.8.10) Problem. Suppose $a = 9,700$ and $c = 10,000$. Find the smallest value of p such that $s_{c,p}(a) \geq 1/2$. Interpret your answer in terms of a gambler at a casino.

(2.8.11) Problem. Consider the Markov chain with state space $S = \{1, 2, 3, 4\}$, $X_0 = 3$, and transition probabilities specified by $p_{11} = p_{22} = 1$, $p_{31} = p_{32} = p_{34} = 1/3$, $p_{43} = 3/4$, $p_{41} = 1/6$, and $p_{42} = 1/12$. Compute $\mathbf{P}(T_1 < T_2)$. [Hint: You may wish to set $s(a) = \mathbf{P}_a(T_1 < T_2)$.]

(2.8.12) Problem. Consider the Markov chain with state space $S = \{1, 2, 3, 4\}$, $X_0 = 2$, and transition probabilities specified by $p_{11} = p_{44} = 1$, $p_{21} = 2/3$, $p_{23} = p_{24} = 1/6$, and $p_{32} = p_{34} = 1/2$. Compute $\mathbf{P}(T_1 < T_4)$.

2.9. Mean Recurrence Times

- Another important quantity is the following:

(2.9.1) Definition. The *mean recurrence time* of a state i is $m_i = \mathbf{E}_i(\inf\{n \geq 1 : X_n = i\}) = \mathbf{E}_i(\tau_i)$ where $\tau_i = \inf\{n \geq 1 : X_n = i\}$.

- That is, m_i is the expected value of the time to return from i back to i.
- Now, if the chain never returns to i, then $\tau_i = \infty$. If this has positive probability (which holds iff i is transient), then $m_i = \infty$, cf. (A.4.1).
  - So, if i is transient, then $m_i = \infty$.
  - It follows that if $m_i < \infty$, then i must be recurrent.
  - However, even if i is recurrent, i.e. $\mathbf{P}_i(\tau_i < \infty) = 1$, it is still possible that $\mathbf{E}_i(\tau_i) = \infty$; cf. Section A.4.
  - This leads to the following definitions:

(2.9.2) Definition. A state is *positive recurrent* if $m_i < \infty$.

- Or, it is *null recurrent* if it is recurrent but $m_i = \infty$.

- Is there any connection to stationary distributions?
  Yes!
- Suppose a chain starts at i, and $m_i < \infty$.
- Over the long run, what fraction of time will it spend at i?
  - Well, on average, the chain returns to i once every m_i steps.
  - That is, for large r, by the *Law of Large Numbers* (A.6.2), it takes about rm_i steps for the chain to return to i a total of r times.
  - So, in a large number $n \approx rm_i$ of steps, it will return to i about r times, i.e. about n/m_i times, i.e. about $1/m_i$ of the time.
  - Hence, the limiting fraction of time it spends at i is equal to $1/m_i$.
  - That is:

(2.9.3) $$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{X_k=i} \;=\; 1/m_i, \;\; w.p.\ 1\,.$$

- It then follows from the Bounded Convergence Theorem (A.9.1) that

(2.9.4) $$\lim_{n\to\infty} \mathbf{E}_i\Big(\frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{X_k=i}\Big) \;=\; 1/m_i\,.$$

- However, if the chain is irreducible, with stationary distribution π, then by finite linearity and Average Probability Convergence (2.5.2),

$$\lim_{n\to\infty} \mathbf{E}_i\Big(\frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{X_k=i}\Big) \;=\; \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{E}_i(\mathbf{1}_{X_k=i})$$

$$=\; \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{P}_i(X_k=i) \;=\; \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ii}^{(k)} \;=\; \pi_i\,.$$

- So, surprisingly, we must have $1/m_i = \pi_i$, i.e. $m_i = 1/\pi_i$.
  - It then follows that $\sum_{i\in S}(1/m_i) = 1$, which seems surprising too.
  - (This explains the names "positive recurrent" and "null recurrent": they come from considering $1/m_i$, not m_i.)
- If $m_i = \infty$, then it follows as above that we must have $\pi_i = 1/m_i = 1/\infty = 0$.
  - Hence, if $m_j = \infty$ for all j, then we must have $\pi_j = 0$ for all j, which contradicts that $\sum_{j\in S} \pi_j = 1$.
  - So, in that case, there is no stationary distribution.
- In turns out that, similar to (2.2.8), if a chain is irreducible, and $m_j < \infty$ for some $j \in S$, then $m_i < \infty$ for all $i \in S$, and $\{1/m_i\}$ is stationary; see e.g. Section 8.4 of Rosenthal (2006).
- Putting this all together, we conclude:

(2.9.5) Recurrence Time Theorem. For an irreducible Markov chain, either (I) $m_i < \infty$ for all $i \in S$, and there is a unique stationary distribution given by $\pi_i = 1/m_i$; or (II) $m_i = \infty$ for all $i \in S$, and there is no stationary distribution.

- Furthermore, we have the following fact similar to the Finite Space Theorem (1.6.7) (for a proof see e.g. Rosenthal, 2006, Proposition 8.4.10):

(2.9.6) Proposition. An irreducible Markov chain on a finite state space S always falls into case (I) above, i.e. has $m_i < \infty$ for all $i \in S$, and a unique stationary distribution given by $\pi_i = 1/m_i$.

- The converse to Proposition (2.9.6) is false.
  - For example, Example (2.2.11) had infinite state space $S = \mathbf{N}$, but still had a stationary distribution, so it falls into case (I).
- What about simple random walk (s.r.w.)?
  - It is irreducible, but has no stationary distribution by (2.2.5).
  - This means that it must be in case (II).
  - So, we must have $m_i = \infty$ for all i.
  - That is, on average s.r.w. (including s.s.r.w. when $p = 1/2$) takes an infinite amount of time to return to where it started.
  - This is surprising, since s.s.r.w. is recurrent, and usually returns quickly. But the small chance of taking a very long time to return is sufficient to make the expected return time infinite even though the actual return time is always finite; cf. Section A.4.
- In summary, simple symmetric random walk (s.s.r.w.) is recurrent and hence in case (a) of the Cases Theorem (1.6.6) with $f_{ij} = 1$ and $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ for all $i, j \in S$; but is in case (i) of the Vanishing Together Corollary (2.2.8) with $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ for all $i, j \in S$ and no stationary distribution; and is in case (II) of the Recurrence Time Theorem (2.9.5) with expected return times $m_i = \infty$ for all $i \in S$.
- Finally, for s.r.w. we consider $\mathbf{E}_i(\tau_{i+1})$, i.e. the expected time starting from i to first hit the state $i+1$.
  - Well, at time 0, s.r.w. moves from i to either $i+1$ with probability p, or $i-1$ with probability $1-p$, which takes one step. Hence,

$$m_i \;:=\; \mathbf{E}_i(\tau_i) \;=\; 1 + p\,\mathbf{E}_{i+1}(\tau_i) + (1-p)\,\mathbf{E}_{i-1}(\tau_i)\,.$$

  - Then, by shift-invariance, this is the same as

$$m_i \;=\; 1 + p\,\mathbf{E}_i(\tau_{i-1}) + (1-p)\,\mathbf{E}_i(\tau_{i+1})\,.$$

  - But we know that $m_i = \infty$, so

$$\infty \;=\; 1 + p\,\mathbf{E}_i(\tau_{i-1}) + (1-p)\,\mathbf{E}_i(\tau_{i+1})\,.$$

  - Now, if $p = 1/2$, then by symmetry $\mathbf{E}_i(\tau_{i+1}) = \mathbf{E}_i(\tau_{i-1})$, so we must have $\mathbf{E}_i(\tau_{i+1}) = \mathbf{E}_i(\tau_{i-1}) = \infty$, i.e. on average it takes an infinite amount of time to progress just one step.
  - By contrast, for s.r.w. with $p > 1/2$, $\mathbf{E}_i(\tau_{i+1}) < \infty$, but in that case $f_{i,i-1} < 1$ by (2.8.5), so $\mathbf{P}_i(\tau_{i-1} = \infty) > 0$, so $\mathbf{E}_i(\tau_{i-1}) = \infty$ by (A.4.1).

2.10. Application – Sequence Waiting Times

- Suppose we repeatedly flip a fair coin, and get Heads (H) or Tails (T) independently each time with probability 1/2 each.
  - Let τ be the first time the sequence "HTH" is completed.

- What is $\mathbf{E}(\tau)$?
- And, is the answer the same for "THH"?

- To find $\mathbf{E}(\tau)$, one solution is: use Markov chains!
- Let X_n be the partial amount of the desired sequence (HTH) that the chain has "achieved so far" after n flips.
  - (For example, if the flips begin with HHTTHTH, then $X_1 = 1$, $X_2 = 1$, $X_3 = 2$, $X_4 = 0$, $X_5 = 1$, $X_6 = 2$, and $X_7 = 3$, and $\tau = 7$.)
  - We always have $X_\tau = 3$, since we "win" upon reaching state 3.
  - Assume we "start over" right after we win, i.e. that $X_{\tau+1} = 1$ if flip $(\tau + 1)$ is Heads, otherwise $X_{\tau+1} = 0$.
  - We take $X_0 = 0$, i.e. at the beginning we have not achieved any of the sequence.
  - This is equivalent to taking $X_0 = 3$, since after that we start over anyway.
- The mean waiting time of "HTH" is thus equal to the mean recurrence time of the state 3!
- Here $\{X_n\}$ is a Markov chain with state space is $S = \{0, 1, 2, 3\}$.
  - What are the transition probabilities?
  - Well, $p_{01} = 1/2 = p_{12} = p_{23}$ (the probability of continuing the sequence).
  - Also $p_{00} = p_{20} = 1/2$ (the probability of a Tail ruining the sequence).
  - But is $p_{10} = 1/2$? No!
  - If the sequence is ruined with a Head, then we've already taken the first step towards a fresh sequence beginning with H. (That is, if you fail to match the second flip, T, then you've already matched the first flip, H, for the next try. This is key!)
  - Hence, $p_{11} = 1/2$, while $p_{10} = 0$.
  - Also, since we "start over" right after we win, therefore $p_{3j} = p_{0j}$ for all j, i.e. $p_{31} = p_{30} = 1/2$.
  - Thus, the transitions are $P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix}$.
- Using the equation $\pi P = \pi$, it can then be computed (check!) that the stationary distribution is $(0.3, 0.4, 0.2, 0.1)$.
  - So, by the Recurrence Time Theorem (2.9.5), the mean time to return from state 3 to state 3 is $1/\pi_3 = 1/0.1 = 10$.
  - But returning from state 3 to state 3 has the same probabilities as going from state 0 to state 3.
  - Hence, the mean time to go from state 0 to state 3 is also 10.
  - That is, the mean waiting time for HTH is 10. We've solved it!
  - Is it true?

  - Try it out, with the R program at: www.probability.ca/Rseqwait

- What about THH? Is it the same?
  - Here we compute similarly (check!) that $P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix}$.
  - Then (check!) the stationary distribution is $(1/8, 1/2, 1/4, 1/8)$.
  - So, the expected waiting time for "THH" is equal to the mean recurrence time of state 3, which is $1/\pi_3 = 1/(1/8) = 8$.
  - This is smaller than the previous answer 10. They are not the same!
  - Is it true? Try it, with www.probability.ca/Rseqwait

(2.10.1) Problem. Suppose we repeatedly flip a fair coin. Let τ be the number of flips until we first see two Heads in a row, i.e. see the pattern "HH". Compute the expected value $\mathbf{E}(\tau)$. [Answer: 6.]

(2.10.2) Problem. Repeat the previous question where the coin is now <u>biased</u> (not fair), and instead shows H with probability 2/3 or T with probability 1/3 (independently on each flip). [Answer: 15/4.]

(2.10.3) Problem. Suppose we repeatedly flip a fair coin. Let τ be the number of flips until we first see three Heads in a row, i.e. see the pattern "HHH". Compute $\mathbf{E}(\tau)$.

(2.10.4) Problem. Suppose we repeatedly flip a coin which is <u>biased</u> (not fair), and shows H with probability 3/4 or T with probability 1/4 (independently on each flip). Let τ be the number of flips until we first see three Heads in a row, i.e. see the pattern "HHH". Compute $\mathbf{E}(\tau)$.

(2.10.5) Problem. Suppose we repeatedly flip a fair coin.

(a) Compute the expected value of the number of flips until we first see the pattern "HTHT".

(b) Repeat part (a) for the pattern "HTTH".

(2.10.6) Problem. Suppose we repeatedly roll a fair six-sided die (which is equally likely to show 1, 2, 3, 4, 5, or 6). Let τ be the number of rolls until the pattern "232" first appears. Compute $\mathbf{E}[\tau]$. [Answer: 222.]

(2.10.7) Problem. Suppose we repeatedly roll a fair six-sided die (which is equally likely to show 1, 2, 3, 4, 5, or 6). Let τ be the number of rolls until the pattern "223" first appears. Compute $\mathbf{E}[\tau]$. [Answer: 216.]

Reminder: No class on Feb. 21 (Reading Week).

END OF WEEK #6

3. Martingales

- Roughly speaking, martingales[‡] are stochastic processes which "stay the same on average".
- For motivation, consider again the Gambler's Ruin problem of Section 2.8, in the case where $p = 1/2$.
  - Let $T = \inf\{n \geq 0 : X_n = 0 \text{ or } c\}$ = time when game ends.
  - We know from (2.8.1) that $\mathbf{P}(X_T = c) = s(a) = a/c$, and $\mathbf{P}(X_T = 0) = r(a) = 1 - s(a) = 1 - (a/c)$.
  - Then $\mathbf{E}(X_T) = c\,\mathbf{P}(X_T = c) + 0\,\mathbf{P}(X_T = 0) = c\,s(a) + 0\,(1 - s(a)) = c\,(a/c) + 0\,(1 - (a/c)) = a$.
  - So, $\mathbf{E}(X_T) = \mathbf{E}(X_0)$, i.e. "on average it stays the same".
  - This makes sense since if $p = 1/2$ then $\mathbf{E}(X_{n+1} \,|\, X_n = i) = (1/2)(i + 1) + (1/2)(i - 1) = i$, i.e. each step stays the same on average.
- We can also apply the reverse logic:
  - Suppose we *knew* that the process stayed the same on average, so that $\mathbf{E}(X_T) = \mathbf{E}(X_0) = a$, but we did not know $s(a)$.
  - We still must have $\mathbf{E}(X_T) = c\,s(a) + 0\,(1 - s(a))$.
  - It follows that $c\,s(a) + 0\,(1 - s(a)) = \mathbf{E}(X_0) = a$.
  - Re-arranging, we would solve that $s(a) = a/c$.
  - This provides a much easier solution to the Gambler's Ruin problem (at least when $p = 1/2$).
  - But can we be sure that $\mathbf{E}(X_T) = \mathbf{E}(X_0)$?
    We'll see!

3.1. Martingale Definitions

- For a formal definition, let $\{X_n\}_{n=0}^{\infty}$ be a sequence of random variables.
- We assume throughout that the random variables X_n have *finite expectation* (or, are *integrable*), i.e. that $\mathbf{E}|X_n| < \infty$ for all n.
  - (This allows us to take conditional expectations below.)

(3.1.1) Definition. A sequence $\{X_n\}_{n=0}^{\infty}$ is a *martingale* if for all n,

$$\mathbf{E}(X_{n+1} \mid X_0, ..., X_n) \;=\; X_n\,.$$

- Definition (3.1.1) says[§] that no matter what has happened so far, the average of the next value will be equal to the most recent value.
- For discrete random variables, the definition means that

(3.1.2)
$$\mathbf{E}[X_{n+1} \mid X_0 = i_0, ..., X_n = i_n] \;=\; i_n$$

[‡] The name "martingale" appears to come from the betting strategies discussed in Remark (2.8.7); see e.g. "The origins of the word martingale" by R. Mansuy, Electronic Journal for History of Probability and Statistics **5(1)**, June 2009.

[§] Technical Remark: Martingales are often defined instead by the condition that $\mathbf{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$ for some nested "filtration" $\{\mathcal{F}_n\}$, i.e. sub-σ-algebras $\{\mathcal{F}_n\}$ with $\sigma(X_0, X_1, \ldots, X_n) \subseteq \mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. But it follows from a version of the double-expectation formula that this condition is actually equivalent to Definition (3.1.1); see e.g. Remark 14.0.1 of Rosenthal (2006).

for all $i_0, i_1, \ldots, i_n$ (see Section A.5).

- An important case is where the sequence $\{X_n\}$ is a *Markov chain.* Then

$$\mathbf{E}[X_{n+1} \mid X_0 = i_0, ..., X_n = i_n] = \sum_{j\in S} j\, P[X_{n+1} = j | X_0 = i_0, ..., X_n = i_n]$$

$$= \sum_j j\, P[X_{n+1} = j | X_n = i_n] = \sum_j j\, p_{i_n,j}\,.$$

  - To be a martingale, this must equal i_n.
  - That is, a Markov chain (with $\mathbf{E}|X_n| < \infty$) is a martingale if

(3.1.3) $$\sum_{j\in S} j\, p_{ij} = i \quad \text{for all } i \in S\,.$$

- As our first example, let $\{X_n\}$ be s.s.r.w. (i.e., simple random walk with $p = 1/2$).
  - Is it a martingale?
  - Well, we always have $|X_n| \leq n$, so $\mathbf{E}|X_n| \leq n < \infty$, so there is no problem with finite expectations. (Indeed, we will almost always have $\mathbf{E}|X_n| < \infty$, so sometimes we won't even mention it.)
  - More importantly, for all $i \in S$, we compute that $\sum_{j\in S} j\, p_{ij} = (i+1)(1/2) + (i-1)(1/2) = i$.
  - This satisfies (3.1.3), so s.s.r.w. is indeed a martingale.
  - (In fact, it's my favourite one!)
  - (Aside: for s.s.r.w., the $\mathbf{E}|X_n|$ are not <u>uniformly</u> bounded, i.e. there is no constant $M < \infty$ such that $\mathbf{E}|X_n| < M$ for <u>all</u> n at once. But this is not required for a martingale. All we require is that that $\mathbf{E}|X_n| < \infty$ for each <u>individual</u> n, a much weaker condition.)
- If $\{X_n\}$ is a martingale, then by the *Double-expectation formula* (A.5.5),

$$\mathbf{E}(X_{n+1}) = \mathbf{E}\Big[\mathbf{E}(X_{n+1} \,|\, X_0, X_1, \ldots, X_n)\Big] = \mathbf{E}(X_n)\,,$$

i.e.

(3.1.4) $$\mathbf{E}(X_n) = \mathbf{E}(X_0) \quad \text{for all } n\,.$$

  - This is not surprising, since martingales stay the same on average.
  - However, (3.1.4) is <u>not</u> a sufficient condition for $\{X_n\}$ to be a martingale, as the following problem shows:

(3.1.5) Problem. Let $\{X_n\}$ be a Markov chain with state space $S = \mathbf{Z}$, and $X_0 = 0$, and $p_{0,1} = p_{0,-1} = 1/2$, and $p_{i,i+1} = 1$ for all $i \geq 1$, and $p_{i,i-1} = 1$ for all $i \leq -1$.

(a) Draw a diagram of this Markov chain.

(b) Compute $\mathbf{P}(X_n = i)$ for each $i \in S$.

(c) Use this to compute $\mathbf{E}(X_n)$ for all $n \in \mathbf{N}$.

(d) Verify that (3.1.4) is satisfied, i.e. that $\mathbf{E}(X_n) = \mathbf{E}(X_0)$ for all n.

(e) Show that (3.1.3) is <u>not</u> satisfied in this case, e.g. when $i = 1$.

(f) Compute $\mathbf{P}(X_2 = i \,|\, X_0 = 0,\ X_1 = 1)$ for all $i \in S$.

(g) Use this to compute $\mathbf{E}(X_2 \mid X_0 = 0,\ X_1 = 1)$.

(h) Show that (3.1.2) is not satisfied in this case, i.e. that $\{X_n\}$ is not a martingale.

(3.1.6) Problem. Consider a Markov chain $\{X_n\}$ with state space $S = \mathbf{Z}$, with $X_0 = 0$, and with transition probabilities given by $p_{00} = 0$, $p_{0j} = c/|j|^3$ for all $j \neq 0$, and $p_{i,i+i^2} = p_{i,i-i^2} = 1/2$ for all $i \neq 0$, with $p_{ij} = 0$ otherwise, where $c > 0$ is chosen so that $\sum_{j \in S} p_{0j} = 1$. For each of the following statements, determine (with proof) whether it is true or false. [Hint: Don't forget (A.3.2) that $\sum_{k=1}^{\infty}(1/k^a)$ is finite for $a > 1$, but infinite for $a \leq 1$.]

(a) $\sum_{j \in S} j\, p_{ij} = i$ for all $i \in S$.

(b) $\mathbf{E}|X_1| < \infty$.

(c) $\mathbf{E}|X_2| < \infty$.

(d) $\{X_n\}$ is a martingale.

3.2. Stopping Times

- For martingales, we know from (3.1.4) that $\mathbf{E}(X_n) = \mathbf{E}(X_0)$ for each fixed time n.
  - However, we will often want to consider $\mathbf{E}(X_T)$ for a random time T.
  - Will we still have $\mathbf{E}(X_T) = \mathbf{E}(X_0)$, even if T is random?
  - One issue is, we need to prevent the random time T from looking into the future of the process, before deciding whether to stop.
  - This prompts the following definition.

(3.2.1) Definition. A non-negative-integer-valued random variable T is a *stopping time* for $\{X_n\}$ if the event $\{T = n\}$ is determined by $X_0, X_1, \ldots, X_n$, i.e. if the indicator function $\mathbf{1}_{T=n}$ is a function of $X_0, X_1, \ldots, X_n$.

- Intuitively, Definition (3.2.1) says that a stopping time T must decide whether to stop at time n based solely on what has happened up to time n, i.e. without first looking into the future.
- Some examples can help to clarify this:
  - e.g. $T = 5$
    is a valid stopping time.
  - e.g. $T = \inf\{n \geq 0 : X_n = 5\}$
    is a valid stopping time.
    (Note: here $T = \infty$ if $\{X_n\}$ never hits 5.)
  - e.g. $T = \inf\{n \geq 0 : X_n = 0 \text{ or } X_n = c\}$
    is a valid stopping time.
  - e.g. $T = \inf\{n \geq 2 : X_{n-2} = 5\}$
    is a valid stopping time.
  - e.g. $T = \inf\{n \geq 2 : X_{n-1} = 5, X_n = 6\}$
    is a valid stopping time.
  - e.g. $T = \inf\{n \geq 0 : X_{n+1} = 5\}$
    is not a valid stopping time (since it looks into the future).
- Another issue is, what if $\mathbf{P}(T = \infty) > 0$?
  - Then perhaps X_T isn't even defined!

- So, suppose T is a stopping time, with $P(T < \infty) = 1$.
  - Then do we always have $\mathbf{E}(X_T) = \mathbf{E}(X_0)$?
  - No, not necessarily!
- For example, let $\{X_n\}$ be s.s.r.w. with $X_0 = 0$.
  - We already know that this is a martingale.
  - Now, let $T = T_{-5} = \inf\{n \geq 0 : X_n = -5\}$.
  - Is T a stopping time?
    Yes indeed.
  - And, $\mathbf{P}(T < \infty) = 1$ since s.s.r.w. is recurrent.
  - However, in this case, $X_T =$
    -5.
  - So, $\mathbf{E}(X_T) = -5 \neq 0 = \mathbf{E}(X_0)$.
  - Why not? What went wrong?
  - It turns out that we need various additional conditions.
  - One possibility is a boundedness condition:

(3.2.2) Optional Stopping Lemma. If $\{X_n\}$ is a martingale, and T is a stopping time which is bounded (i.e., $\exists\, M < \infty$ with $\mathbf{P}(T \leq M) = 1$), then $\mathbf{E}(X_T) = \mathbf{E}(X_0)$.

Proof. Using a telescoping sum, and then indicator functions, and then finite linearity of expectation (A.8.1),

$$\mathbf{E}(X_T) - \mathbf{E}(X_0) \;=\; \mathbf{E}(X_T - X_0) \;=\; \mathbf{E}\left[\sum_{k=1}^{T}(X_k - X_{k-1})\right]$$

$$= \mathbf{E}\left[\sum_{k=1}^{M}(X_k - X_{k-1})\mathbf{1}_{k\leq T}\right] \;=\; \sum_{k=1}^{M}\mathbf{E}\Big[(X_k - X_{k-1})\mathbf{1}_{k\leq T}\Big]$$

$$= \sum_{k=1}^{M}\mathbf{E}\Big[(X_k - X_{k-1})\,(1 - \mathbf{1}_{T\leq k-1})\Big]\,.$$

Next, using the *Double-expectation formula* (A.5.5), and then the fact that "$1 - \mathbf{1}_{T\leq k-1}$" is completely determined by $X_0, X_1, \ldots, X_{k-1}$ and thus can be conditionally factored out as in (A.5.6), it follows that

$$\mathbf{E}(X_T) - \mathbf{E}(X_0) \;=\; \sum_{k=1}^{M}\mathbf{E}\Big(\mathbf{E}\Big[(X_k - X_{k-1})\,(1 - \mathbf{1}_{T\leq k-1})\;\Big|\; X_0, X_1, \ldots, X_{k-1}\Big]\Big)$$

$$= \sum_{k=1}^{M}\mathbf{E}\Big((1 - \mathbf{1}_{T\leq k-1})\;\mathbf{E}\Big[(X_k - X_{k-1})\;\Big|\; X_0, X_1, \ldots, X_{k-1}\Big]\Big)\,.$$

But $\mathbf{E}\Big[X_k \;\Big|\; X_0, X_1, \ldots, X_{k-1}\Big] = X_{k-1}$ since $\{X_n\}$ is a martingale, and $\mathbf{E}\Big[X_{k-1} \;\Big|\; X_0, X_1, \ldots, X_{k-1}\Big] = X_{k-1}$ by (A.5.7) since X_{k-1} is a function of $X_0, X_1, \ldots, X_{k-1}$. Hence,

$$\mathbf{E}(X_T) - \mathbf{E}(X_0) \;=\; \sum_{k=1}^{M}\mathbf{E}\Big((1 - \mathbf{1}_{T\leq k-1})\,(X_{k-1} - X_{k-1})\Big)$$

$$= \sum_{k=1}^{M}\mathbf{E}\Big((1 - \mathbf{1}_{T\leq k-1})\,(0)\Big) \;=\; 0\,. \qquad \blacksquare$$

(3.2.3) Problem. Identify where the above proof breaks down if $M = \infty$.

(3.2.4) Example. Consider s.s.r.w. with $X_0 = 0$, and let

$$T \;=\; \min\left(10^{12},\ \inf\{n \geq 0 : X_n = -5\}\right).$$

- Then T is a stopping time.
- Also, $T \leq 10^{12}$, so T is bounded.
- Hence, by the Optional Stopping Lemma (3.2.2), $\mathbf{E}(X_T) = \mathbf{E}(X_0) = \mathbf{E}(0) = 0$.
- But nearly always, we will have $X_T = -5$. How can this be?
- Well, by the *Law of Total Expectation* (A.5.3),

$$0 \;=\; \mathbf{E}(X_T)$$

$$= \mathbf{P}(X_T = -5)\,\mathbf{E}(X_T \,|\, X_T = -5) + \mathbf{P}(X_T \neq -5)\,\mathbf{E}(X_T \,|\, X_T \neq -5)\,.$$

- Here $\mathbf{E}(X_T \,|\, X_T = -5) = -5$.
- Also $\mathbf{P}(X_T = -5) \approx 1$, and $\mathbf{P}(X_T \neq 5) \approx 0$.
- So how is this possible?
- Well, $\mathbf{E}(X_T \,|\, X_T \neq 5)$ must be huge!

• Can we apply Optional Stopping Lemma to the Gambler's Ruin problem?
  - No, since there T is not bounded (though still finite, cf. Section A.4).
  - We need a more general result:

(3.2.5) Optional Stopping Theorem. If $\{X_n\}$ is martingale with stopping time T, and $\mathbf{P}(T < \infty) = 1$, and $\mathbf{E}|X_T| < \infty$, and $\lim_{n\to\infty} \mathbf{E}(X_n\, \mathbf{1}_{T>n}) = 0$, then $\mathbf{E}(X_T) = \mathbf{E}(X_0)$.

Proof.

- For each $m \in \mathbf{N}$, let $S_m = \min(T, m)$. Stopping time! Bounded!
- Then by Optional Stopping Lemma, $\mathbf{E}(X_{S_m}) = \mathbf{E}(X_0)$ (for any m).
- But $X_{S_m} = X_{\min(T,m)} = X_T \mathbf{1}_{T \leq m} + X_m \mathbf{1}_{T>m}$.
- i.e., $X_{S_m} = X_T(1 - \mathbf{1}_{T>m}) + X_m \mathbf{1}_{T>m} = X_T - X_T \mathbf{1}_{T>m} + X_m \mathbf{1}_{T>m}$.
- So, $X_T = X_{S_m} + X_T \mathbf{1}_{T>m} - X_m \mathbf{1}_{T>m}$.
- So, $\mathbf{E}(X_T) = \mathbf{E}(X_{S_m}) + \mathbf{E}(X_T \mathbf{1}_{T>m}) - \mathbf{E}(X_m \mathbf{1}_{T>m})$.
- So, from above, for <u>any</u> m, $\mathbf{E}(X_T) = \mathbf{E}(X_0) + \mathbf{E}(X_T \mathbf{1}_{T>m}) - \mathbf{E}(X_m \mathbf{1}_{T>m})$.
- Then take $m \to \infty$.
- Here $\lim_{m\to\infty} \mathbf{E}(X_T \mathbf{1}_{T>m}) = 0$ by the Dominated Convergence Theorem (A.9.3), since $\mathbf{E}|X_T| < \infty$ and $\mathbf{1}_{T>m} \to 0$ (since $\mathbf{P}(T < \infty) = 1$).
- Also, $\lim_{m\to\infty} \mathbf{E}(X_m \mathbf{1}_{T>m}) = 0$ by assumption.
- So, $\mathbf{E}(X_T) \to \mathbf{E}(X_0)$, i.e. $\mathbf{E}(X_T) = \mathbf{E}(X_0)$. ∎

• Can we apply <u>this</u> to the Gambler's Ruin problem?
  - Almost ... we need one more corollary first.

(3.2.6) Optional Stopping Corollary. If $\{X_n\}$ is martingale with stopping time T, which is "bounded up to time T" (i.e., $\exists\, M < \infty$ with $\mathbf{P}(|X_n| \mathbf{1}_{n\le T} \le M) = 1$ for all n), and $\mathbf{P}(T < \infty) = 1$, then $\mathbf{E}(X_T) = \mathbf{E}(X_0)$.

Proof.

- It follows, clearly, that $\mathbf{P}(|X_T| \le M) = 1$.
  [Formally, this holds since $\mathbf{P}(|X_T| > M) = \sum_n \mathbf{P}(T = n,\ |X_T| > M) = \sum_n \mathbf{P}(T = n,\ |X_n| \mathbf{1}_{n\le T} > M) \le \sum_n \mathbf{P}(|X_n| \mathbf{1}_{n\le T} > M) = \sum_n (0) = 0$.]
- Hence, $\mathbf{E}|X_T| \le M < \infty$.
- Also, $|\mathbf{E}(X_n \mathbf{1}_{T>n})| \le \mathbf{E}(|X_n| \mathbf{1}_{T>n}) = \mathbf{E}(|X_n| \mathbf{1}_{n\le T} \mathbf{1}_{T>n}) \le \mathbf{E}(M \mathbf{1}_{T>n}) = M\, \mathbf{P}(T > n)$, which $\to 0$ as $n \to \infty$ since $\mathbf{P}(T < \infty) = 1$.
- Hence, the result follows from the Optional Stopping Theorem. ∎

- Now, consider again the Gambler's Ruin problem of Section 2.8, with $p = 1/2$.
  - Let $T = \inf\{n \ge 0 : X_n = 0 \text{ or } X_n = c\}$ be the time when the game ends.
  - Then $\mathbf{P}(T < \infty) = 1$ by Proposition (2.8.6).
  - Also, if the game has <u>not</u> yet ended, i.e. $n \le T$, then X_n must be between 0 and c. Hence, $|X_n| \mathbf{1}_{n\le T} \le c < \infty$ for all n.
  - So, by the Optional Stopping Corollary, $\mathbf{E}(X_T) = \mathbf{E}(X_0) = a$.
  - Hence, as before, $a = c\, s(a) + 0\,(1 - s(a))$, so we must have $s(a) = a/c$.
  - (This is a much easier solution than in Section 2.8!)
- What about Gambler's Ruin with $p \ne 1/2$?
  - Then is $\{X_n\}$ a martingale?
  - <u>No</u>, since $\sum_j j\, p_{ij} = p(i+1) + (1-p)(i-1) = i + 2p - 1 \ne i$.
  - Instead, we use a trick: Let $Y_n = \left(\frac{1-p}{p}\right)^{X_n}$.
  - Then $\{Y_n\}$ is also a Markov chain. And,

$$\mathbf{E}(Y_{n+1} \,|\, Y_0, Y_1, \ldots, Y_n)$$
$$= p\left(\frac{1-p}{p}\right)^{X_n+1} + (1-p)\left(\frac{1-p}{p}\right)^{X_n-1}$$
$$= p\left[Y_n\left(\frac{1-p}{p}\right)\right] + (1-p)\left[Y_n \Big/ \left(\frac{1-p}{p}\right)\right]$$
$$= Y_n(1-p) + Y_n(p) \;=\; Y_n\,.$$

  - So, $\{Y_n\}$ is a martingale!
  - And, again $\mathbf{P}(T < \infty) = 1$ by Proposition (2.8.6).
  - And, $|Y_n| \mathbf{1}_{n\le T} \le \max\left(\left(\frac{1-p}{p}\right)^0, \left(\frac{1-p}{p}\right)^c\right) =: M < \infty$ for all n.
  - Hence, by the Optional Stopping Corollary, $\mathbf{E}(Y_T) = \mathbf{E}(Y_0) = \left(\frac{1-p}{p}\right)^a$.
  - But $Y_T = \left(\frac{1-p}{p}\right)^c$ if the game is won, or $Y_T = \left(\frac{1-p}{p}\right)^0 = 1$ if it is lost.
  - Hence, $\left(\frac{1-p}{p}\right)^a = \mathbf{E}(Y_T) = s(a)\left(\frac{1-p}{p}\right)^c + [1 - s(a)](1) = 1 + s(a)\left[\left(\frac{1-p}{p}\right)^c - 1\right]$.

– Solving, $s(a) = \frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1}$. (Again, a much easier solution!)

(3.2.7) Problem. Let $\{X_n\}$ be a Markov chain on the state space $S = \{1,2,3,4\}$, with $X_0 = 3$, and with transition probabilities $p_{11} = p_{44} = 1$, $p_{21} = 1/3$, $p_{34} = 1/4$, and $p_{23} = p_{31} = p_{12} = p_{13} = p_{14} = p_{41} = p_{42} = p_{43} = 0$. Let $T = \inf\{n \geq 0 : X_n = 1 \text{ or } 4\}$.

(a) Find non-negative values of p_{22}, p_{24}, p_{32}, and p_{33}, such that $\sum_{j \in S} p_{ij} = 1$ for all $i \in S$ (as it must), and also $\{X_n\}$ is a martingale. **[sol]**

(b) For the values found in part (a), compute $\mathbf{E}(X_T)$. **[sol]**

(c) For the values found in part (a), compute $\mathbf{P}(X_T = 1)$. **[sol]**

(3.2.8) Problem. Let $\{X_n\}$ be a Markov chain on the state space $S = \{1,2,3,4\}$, with $X_0 = 2$, and with transition probabilities $p_{11} = p_{44} = 1$, $p_{21} = 1/4$, $p_{34} = 1/5$, and $p_{23} = p_{31} = p_{12} = p_{13} = p_{14} = p_{41} = p_{42} = p_{43} = 0$. Let $T = \inf\{n \geq 0 : X_n = 1 \text{ or } 4\}$.

(a) Find non-negative values of p_{22}, p_{24}, p_{32}, and p_{33}, such that $\sum_{j \in S} p_{ij} = 1$ for all $i \in S$ (as it must), and also $\{X_n\}$ is a martingale. **[sol]**

(b) For the values found in part (a), compute $\mathbf{E}(X_T)$. **[sol]**

(c) For the values found in part (a), compute $\mathbf{P}(X_T = 1)$. **[sol]**

(3.2.9) Problem. Let $\{X_n\}$ be a Markov chain on the state space $S = \{1,2,3,\ldots\}$ of all positive integers, which is also a martingale. Assume $X_0 = 5$, and that there is $c > 0$ such that $p_{i,i-1} = c$ and $p_{i,i+2} = 1 - c$ for all $i \geq 2$. Let $T = \inf\{n \geq 0 : X_n = 1 \text{ or } X_n \geq 10\}$.

(a) Determine (with explanation) what c must equal. [Hint: remember that $\{X_n\}$ is a martingale.] **[sol]**

(b) Determine (with explanation) what p_{11} must equal. [Hint: again, remember that $\{X_n\}$ is a martingale.] **[sol]**

(c) Determine (with explanation) the value of $\mathbf{E}(X_3)$. **[sol]**

(d) Determine (with explanation) the value of $\mathbf{E}(X_T)$. **[sol]**

(e) Prove or disprove that $\sum_{n=1}^{\infty} p_{55}^{(n)} = \infty$. **[sol]**

(3.2.10) Problem. Let $\{Z_i\}_{i=1}^{\infty}$ be an i.i.d. collection of random variables with $\mathbf{P}[Z_i = -1] = 3/4$ and $\mathbf{P}[Z_i = C] = 1/4$, for some $C > 0$. Let $X_0 = 5$, and $X_n = 5 + Z_1 + Z_2 + \ldots + Z_n$ for $n \geq 1$. Finally, let $T = \inf\{n \geq 1 : X_n = 0 \text{ or } Z_n > 0\}$.

(a) Find (with explanation) a value of C such that $\{X_n\}$ is a martingale.

(b) For this value of C, compute (with explanation) $\mathbf{E}(X_9)$.

(c) For this value of C, compute (with explanation) $\mathbf{E}(X_T)$. [Hint: is T bounded?]

3.3. Wald's Theorem

We begin with a warm-up example:

(3.3.1) Example. Suppose we repeatedly roll a fair six-sided die (which is equally likely to show 1, 2, 3, 4, 5, or 6).

– Let Z_n be the result of the n'th roll.

– Let $R = \inf\{n \geq 1 : Z_n = 5\}$ be the first time we roll 5.

- Let A be the sum of all the numbers rolled up to time R, i.e. $A = \sum_{n=1}^{R} Z_n$.
- What is $\mathbf{E}(A)$?
- Also, let $S = \inf\{n \geq 1 : Z_n = 3\}$, and $B = \sum_{n=1}^{S} Z_n$.
- Which is larger, $\mathbf{E}(A)$ or $\mathbf{E}(B)$?
- Finally, let $A' = \sum_{n=1}^{R-1} Z_n$ and $B' = \sum_{n=1}^{S-1} Z_n$ be the sums not counting the final roll.
- Then which is larger, $\mathbf{E}(A')$ or $\mathbf{E}(B')$??

To answer questions like this, we shall use:

(3.3.2) Wald's Theorem. Suppose $X_n = a + Z_1 + \ldots + Z_n$, where $\{Z_i\}$ are i.i.d., with finite mean m. Let T be a stopping time for $\{X_n\}$ which has finite mean, i.e. $\mathbf{E}(T) < \infty$. Then $\mathbf{E}(X_T) = a + m\,\mathbf{E}(T)$.

- A special case is when $m = 0$. Then, $\{X_n\}$ is a martingale, and Wald's Theorem says that $\mathbf{E}(X_T) = a = \mathbf{E}(X_0)$, as usual, but under different assumptions (an i.i.d. sum, with $\mathbf{E}(T) < \infty$).
- Suppose $\{X_n\}$ is s.s.r.w. with $X_0 = 0$, and $T = \inf\{n \geq 0 : X_n = -5\}$.
  - Then $a = 0$, and $m = 0$, and $\mathbf{P}(T < \infty) = 1$.
  - But $\mathbf{E}(X_T) = -5 \neq 0 = \mathbf{E}(X_0)$.
  - Does this contradict Wald's Theorem?
  - No, since here $\mathbf{E}(T) = \infty$.
- Let's apply Wald's Theorem to Example (3.3.1).
  - In this case, the $\{Z_i\}$ are i.i.d. with mean $m = 3.5$.
  - And, clearly R and S are stopping times.
  - Also, R and S have Geometric(1/6) distributions (A.2.4), so $\mathbf{E}(R) = \mathbf{E}(S) = 6$.
  - In the notation of Wald's Theorem, $a = 0$, and $A = X_R$, and $B = X_S$.
  - So, $\mathbf{E}(A) = \mathbf{E}(X_R) = a + m\,\mathbf{E}(R) = 3.5(6) = 21$.
  - Similarly, $\mathbf{E}(B) = \mathbf{E}(X_S) = a + m\,\mathbf{E}(S) = 3.5(6) = 21$. Equal!
- What about A' and B'?
  - Well, $A' = X_{R-1}$ and $B' = X_{S-1}$.
  - However, $R-1$ and $S-1$ are not stopping times, so we cannot apply Wald's Theorem.
  - On the other hand, $A' = A - 5$, and $B' = B - 3$.
  - So, $\mathbf{E}(A') = \mathbf{E}(A) - 5 = 21 - 5 = 16$.
  - And, $\mathbf{E}(B') = \mathbf{E}(B) - 3 = 21 - 3 = 18$. Larger!

Proof of Wald's Theorem. We compute that

$$\mathbf{E}(X_T) - a \;=\; \mathbf{E}(X_T - a) \;=\; \mathbf{E}(Z_1 + \ldots + Z_T)$$

$$= \mathbf{E}\left[\sum_{i=1}^{T} Z_i\right] \;=\; \mathbf{E}\left[\sum_{i=1}^{\infty} Z_i\,\mathbf{1}_{T \geq i}\right] \;=\; \mathbf{E}\left[\lim_{N\to\infty} \sum_{i=1}^{N} Z_i\,\mathbf{1}_{T \geq i}\right].$$

Assuming we can interchange the expectation and limit, this gives

$$\mathbf{E}(X_T) - a \;=\; \lim_{N\to\infty} \mathbf{E}\left[\sum_{i=1}^{N} Z_i\,\mathbf{1}_{T \geq i}\right]$$

$$= \lim_{N\to\infty} \sum_{i=1}^{N} \mathbf{E}\left[Z_i \, \mathbf{1}_{T\ge i}\right] \;=\; \sum_{i=1}^{\infty} \mathbf{E}\left[Z_i \, \mathbf{1}_{T\ge i}\right].$$

Now, we use the fact that $\{T \ge i\} = \{T \le i-1\}^C$, so since T is a stopping time, Z_i is independent of the event $\{T \ge i\}$, and hence

$$\mathbf{E}(X_T) - a \;=\; \sum_{i=1}^{\infty} \mathbf{E}[Z_i]\,\mathbf{E}[\mathbf{1}_{T\ge i}] \;=\; \sum_{i=1}^{\infty} m\,\mathbf{P}[T \ge i] \;=\; m \sum_{i=1}^{\infty} \mathbf{P}[T \ge i]\,.$$

And this last expression equals $m\,\mathbf{E}(T)$, by the trick (A.2.6).

It remains to justify interchanging the above expectation and limit. This follows (optional) from the Dominated Convergence Theorem (A.9.3) with sequence $X_N = \sum_{i=1}^{N} Z_i \, \mathbf{1}_{T\ge i}$ and limit $X = \sum_{i=1}^{\infty} Z_i \, \mathbf{1}_{T\ge i}$ and dominator $Y = \sum_{i=1}^{\infty} |Z_i| \, \mathbf{1}_{T\ge i}$ since then $|X_N| \le Y$ for all N (by the triangle inequality), while (by non-negative countable linearity and the trick (A.2.6)) $\mathbf{E}(Y) = \mathbf{E}[\sum_{i=1}^{\infty} |Z_i| \, \mathbf{1}_{T\ge i}] = \sum_{i=1}^{\infty} \mathbf{E}[|Z_i| \, \mathbf{1}_{T\ge i}] = \sum_{i=1}^{\infty} \mathbf{E}|Z_i| \, \mathbf{E}[\mathbf{1}_{T\ge i}] = \mathbf{E}|Z_1| \sum_{i=1}^{\infty} \mathbf{P}[T \ge i] = \mathbf{E}|Z_1| \, \mathbf{E}(T) < \infty$. ∎

- Consider again the Gambler's Ruin problem of Section 2.8, with $p \ne 1/2$.
  - Again let $T = \inf\{n \ge 0 : X_n = 0 \text{ or } c\}$.
  - What is $\mathbf{E}(T)$ = expected number of bets in the game?
  - We have the following:

(3.3.3) Corollary. If $\{X_n\}$ is Gambler's Ruin with $p \ne 1/2$, and $T = \inf\{n \ge 0 : X_n = 0 \text{ or } c\}$, then

$$\mathbf{E}(T) \;=\; \frac{1}{2p-1}\left(c\,\frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1} - a\right).$$

Proof. We again apply Wald's Theorem.

- Here $Z_i = +1$ if you win the i^{th} bet, otherwise $Z_i = -1$.
- So, $m = \mathbf{E}(Z_i) = p(1) + (1-p)(-1) = 2p - 1$.
- And, $\mathbf{E}(T) < \infty$ by Proposition (2.8.6).
- Hence, by Wald's Theorem, $\mathbf{E}(X_T) = a + m\,\mathbf{E}(T)$.
- But $\mathbf{E}(X_T) = c\,s(a) + 0\,(1 - s(a)) = c\frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1}$.
- So, $\mathbf{E}(T) = \frac{1}{m}\left(\mathbf{E}(X_T) - a\right) = \frac{1}{2p-1}\left(c\frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1} - a\right)$, as claimed. ∎

- For example, if $p = 0.49$, $a = 9{,}700$, and $c = 10{,}000$, then $\mathbf{E}(T) = 484{,}997$. (large!)
- What about $\mathbf{E}(T)$ when $p = 1/2$?
  - If $p = 1/2$, then $m = 0$, so the above method does not work.
  - Instead, we need to use a "second order" martingale:

(3.3.4) Lemma. Let $X_n = a + Z_1 + \ldots + Z_n$, where $\{Z_i\}$ are i.i.d. with mean 0 and variance $v < \infty$. Let $Y_n = (X_n - a)^2 - nv = (Z_1 + \ldots + Z_n)^2 - nv$. Then $\{Y_n\}$ is a martingale.

Proof. We first check that $\mathbf{E}|Y_n| \le \mathbf{Var}(X_n) + nv = 2nv < \infty$.

– Then, since Z_{n+1} is independent of $Z_1, \ldots, Z_n, Y_0, \ldots, Y_n$, with $\mathbf{E}[Z_{n+1}] = 0$ and $\mathbf{E}[Z_{n+1})^2] = v$, we have

$$\mathbf{E}[Y_{n+1} \,|\, Y_0, Y_1, \ldots, Y_n] \;=\; \mathbf{E}\Big[(Z_1+\ldots+Z_n+Z_{n+1})^2 - (n+1)v \;\Big|\; Y_0, Y_1, \ldots, Y_n\Big]$$

$$= \mathbf{E}\Big[(Z_1+\ldots+Z_n)^2 + (Z_{n+1})^2 + 2\,Z_{n+1}(Z_1+\ldots+Z_n) - nv - v \;\Big|\; Y_0, Y_1, \ldots, Y_n\Big]$$

$$= \mathbf{E}\Big[Y_n + (Z_{n+1})^2 - v + 2\,Z_{n+1}(Z_1 + \ldots + Z_n) \;\Big|\; Y_0, Y_1, \ldots, Y_n\Big]$$

$$= Y_n + v - v + 2\,\mathbf{E}(Z_{n+1})\,\mathbf{E}\Big[Z_1 + \ldots + Z_n \;\Big|\; Y_0, Y_1, \ldots, Y_n\Big]$$

$$= Y_n + v - v + 0 \;=\; Y_n\,. \qquad \blacksquare$$

(3.3.5) Corollary. If $\{X_n\}$ is Gambler's Ruin with $p = 1/2$, and $T = \inf\{n \ge 0 : X_n = 0 \text{ or } c\}$, then $\mathbf{E}(T) = \mathbf{Var}(X_T) = a(c-a)$.

Proof. Let $Y_n = (X_n - a)^2 - n$ (since here $v = 1$).

– Then $\{Y_n\}$ is a martingale by the above Lemma.

– Choose $M > 0$, and let $S_M = \min(T, M)$.

– Then S_m is a stopping time, which is bounded by M.

– Hence, by the Optional Stopping Lemma, $\mathbf{E}[Y_{S_M}] = \mathbf{E}[Y_0] = (a - a)^2 - 0 = 0$.

– But $Y_{S_M} = (X_{S_M} - a)^2 - S_M$, so $\mathbf{E}(S_M) = \mathbf{E}[(X_{S_M} - a)^2]$.

– As $M \to \infty$, S_M increases monotonically to T, so $\mathbf{E}(S_M) \to \mathbf{E}(T)$ by the Monotone Convergence Theorem (A.9.2).

– Also, $\mathbf{E}[(X_{S_M} - a)^2] \to \mathbf{E}[(X_T - a)^2]$ by the Bounded Convergence Theorem (A.9.1), since for any n, $(X_{S_M} - a)^2 \le \max\left(a^2, (c-a)^2\right) < \infty$.

– Hence, letting $M \to \infty$ gives $\mathbf{E}(T) = \mathbf{E}[(X_T - a)^2]$.

– And, since $\mathbf{E}(X_T) = a$, $\mathbf{E}[(X_T - a)^2] = \mathbf{Var}(X_T)$.

– Finally, we compute that $\mathbf{Var}(X_T) = (a/c)\,(c-a)^2 + (1 - a/c)a^2 = (a/c)(c^2 + a^2 - 2ac) + (a^2 - a^3/c) = ac + a^3/c - 2a^2 + a^2 - a^3/c = ac - a^2 = a(c-a)$. $\blacksquare$

- For example, if $c = 10{,}000$, $a = 9{,}700$, and $p = 1/2$, then $\mathbf{E}(T) = a(c-a) = 2{,}910{,}000$. (even larger!)

(3.3.6) Problem. Prove that the formula for $\mathbf{E}(T)$ for Gambler's Ruin is continuous as $p \to 1/2$. **[sol]**

(3.3.7) Example. Consider again the "double 'til you win" betting strategy discussed in Remark (2.8.7).

– Specifically, let $p \le 1/2$ be the probability of winning each bet.

– Let $\{Z_i\}$ be i.i.d., with $\mathbf{P}[Z_i = +1] = p$ and $\mathbf{P}[Z_i = -1] = 1 - p$.

– Then $m := \mathbf{E}(Z_i) = p - (1-p) = 2p - 1 \le 0$.

– Suppose we bet 2^{i-1} dollars on the i^{th} bet.

– Then our winnings after n bets equals $X_n = \sum_{i=1}^{n} 2^{i-1} Z_i$.

– Let $T = \inf\{n \ge 1 : Z_n = +1\}$ be time of our first win, and suppose we <u>stop</u> betting at time T.

- Then T has a *geometric distribution* (A.2.4), with $\mathbf{E}(T) = 1/p < \infty$.
- Also, w.p. 1, $T < \infty$ and $X_T = -1 - 2 - 4 - \ldots - 2^{T-1} + 2^T = +1$.
- That is, we are guaranteed to be up \$1 at time T, even though we start at $a = 0$ and have average gain $m \leq 0$ on each bet.
- So, if X_n is the total amount won by time n, then $\mathbf{E}(X_T) = +1$.
- However, $a + m\,\mathbf{E}(T) = (2p-1)/p \leq 0$.
- Does this contradict Wald's Theorem?
  No, since even though $\{Z_i\}$ is i.i.d., the sequence $\{2^{i-1}Z_i\}$ is not i.i.d.
- (In any case, as discussed in Remark (2.8.7), this strategy is <u>not</u> a guaranteed money-maker, due to the possibility of reaching your credit limit and being forced to stop betting with a huge loss.)

3.4. Application – Sequence Waiting Times

- As in Section 2.10, suppose we repeatedly flip a fair coin, and let τ be the number of flips until we first see the pattern HTH.
  - Again we can ask, what is the expected time $\mathbf{E}(\tau)$?
  - But this time, we'll use a martingale approach!
- Suppose that at each time n, a new "player" appears, and bets \$1 on heads, then if they win they bet \$2 on tails, then if they win again they bet \$4 on heads.
  - Each player stops betting as soon as they either lose once (and hence are down a total of \$1), or win three bets in a row (and hence are up a total of \$7).
- Let X_n be the total amount won by <u>all</u> the betters by time n.
- Then since the bets were fair, $\{X_n\}$ is a <u>martingale</u> with <u>stopping time</u> τ.
  - Here each of the first $\tau - 3$ players (i.e., all but the last three) each has a net loss of \$1 (after either one or two or three bets).
  - Player $\tau - 2$ wins all three bets, for a gain of \$7.
  - Player $\tau - 1$ bets on "H" and loses, for a loss of \$1.
  - Player τ bets on "H" and wins, for a gain of \$1.
  - Hence, $X_\tau = (\tau - 3)(-1) + (7) + (-1) + (1) = -\tau + 10$.
- And, it follows here that $\mathbf{E}(X_\tau) = \mathbf{E}(X_0) = 0$.
  - Indeed, if $T_m = \min(\tau, m)$, then $\mathbf{E}(X_{T_m}) = 0$ by the Optional Stopping Lemma (3.2.2), and $\lim_{m\to\infty} \mathbf{E}(X_{T_m}) = \mathbf{E}(X_\tau)$ by the Dominated Convergence Theorem (A.9.3) with dominator $Y = 7\tau$ since $|X_n - X_{n-1}| \leq 7$ and $\mathbf{E}(\tau) < \infty$.
- Hence, $0 = \mathbf{E}(X_\tau) = -\mathbf{E}(\tau) + 10$, whence $\mathbf{E}(\tau) = 10$.
  - Same as before! But easier: no stationary distribution to compute.
- Similarly, for THH, get that $X_\tau = -(\tau - 3) + (7) + (-1) + (-1) = -\tau + 8$ [check], whence $\mathbf{E}(\tau) = 8$. Same as before!
- And, for TTT, we get [check] that $X_\tau = -(\tau - 3) + 7 + 3 + 1 = -\tau + 14$, so $\mathbf{E}(\tau) = 14$.

(3.4.1) Problem. Suppose we repeatedly flip a fair coin. Use the martingale method of this section to compute $\mathbf{E}(\tau)$, where τ is the number of flips until we first see the pattern:

(a) HH.

(b) HHH.

(c) HTHT.

(d) HTTH.

——————— **END OF WEEK #7** ———————

3.5. Martingale Convergence Theorem

- Suppose $\{X_n\}$ is a martingale.
- Then $\{X_n\}$ could have infinite fluctuations in both directions, as we have seen for s.s.r.w. (1.6.13).
- Or, $\{X_n\}$ could converge w.p. 1 to a fixed (perhaps random) value, as in the following two examples.

(3.5.1) Example. Let $\{X_n\}$ be Gambler's Ruin with $p = 1/2$, where we stop as soon as we either win or lose. Then $X_n \to X$ w.p. 1, where $\mathbf{P}(X = c) = a/c$ and $\mathbf{P}(X = 0) = 1 - a/c$.

(3.5.2) Example. Let $\{X_n\}$ be a Markov chain on $S = \{2^m : m \in \mathbf{Z}\}$, with $X_0 = 1$, and $p_{i,2i} = 1/3$ and $p_{i,i/2} = 2/3$ for $i \in S$.

- This is a martingale, since $\sum_j j\, p_{ij} = (2i)(1/3) + (i/2)(2/3) = i$.
- What happens in the long run?
- Trick: let $Y_n = \log_2 X_n$. Then $Y_0 = 0$, and $\{Y_n\}$ is s.r.w. with $p = 1/3$, so $Y_n \to -\infty$ w.p. 1 by the Law of Large Numbers (A.6.2).
- Hence, $X_n = 2^{Y_n} \to 2^{-\infty} = 0$ w.p. 1.

- It turns out that these are the only two possibilities.
  - That is, if a martingale does not have infinite fluctuations in both directions, then it must converge to a specific (perhaps random) value.
  - More precisely, for non-negative martingales, we have:

(3.5.3) Martingale Convergence Theorem. Any martingale $\{X_n\}$ which is bounded below (i.e. $X_n \ge c$ for all n, for some finite number c), or is bounded above (i.e. $X_n \le c$ for all n, for some finite number c), converges w.p. 1 to some random variable X.

- The intuition behind this theorem is:
  - Since the martingale is bounded on one side, it cannot "spread out" forever.
  - And, since it is a martingale, it cannot "drift" in a positive or negative direction.
  - So, it has nowhere to go, and eventually has to stop somewhere.
- We do not prove this theorem here; for a proof see e.g. Section 14.2 of Rosenthal (2006), or Williams (1991), or Billingsley (1995), or many other advanced probability books.
- Note: The Martingale Convergence Theorem is usually stated for non-negative martingales, i.e. assuming $X_n \ge 0$, i.e. taking $c = 0$.

  - However, our more general version follows from this, since if $X_n \geq c$ then $\{X_n - c\}$ is a non-negative martingale, or if $X_n \leq c$ then $\{-X_n + c\}$ is a non-negative martingale, and in either case the non-negative martingale converges iff $\{X_n\}$ converges.

- Consider some examples:
  - s.s.r.w. is a martingale, but it does not converge. However, it is not bounded above or below.
  - If we modify s.s.r.w. to stop whenever it reaches 0 (after starting at a positive number), then it is still a martingale, and now it is non-negative and hence bounded below. But since $f_{i0} = 1$ for all i, it will eventually reach 0, and hence it converges to $X = 0$.
  - Or, if we instead modify s.s.r.w. so that from 0 it always moves to 1, then it is again non-negative, but it does not converge. However, this modification is not a martingale.
  - Or, if we instead modify s.s.r.w. to make smaller and smaller (but still symmetric) jumps when it is near 0, to avoid ever going negative, then it is still a martingale, and is non-negative. But then it will converge to some value X (perhaps random, or perhaps 0, depending on how we have modified the jumps).
  - Or, if we consider s.r.w. with $p = 2/3$ stopped at 0, then it is again non-negative, and if $X_0 = i > 0$ then since $f_{i0} < 1$ it might never reach 0, so it might fail to converge. However, it is not a martingale.
- In short, if you try to modify a martingale to make it bounded above or below, then the modification will either converge, or will no longer be a martingale.

(3.5.4) Problem. Let $X_1, X_2, \ldots$ be independent random variables with

$$X_n = \begin{cases} 1 & \text{with probability } (2n)^{-1} \\ 0 & \text{with probability } 1 - n^{-1} \\ -1 & \text{with probability } (2n)^{-1} . \end{cases}$$

Let $Y_1 = X_1$, and for $n \geq 2$,

$$Y_n = \begin{cases} X_n & \text{if } Y_{n-1} = 0 \\ nY_{n-1}|X_n| & \text{if } Y_{n-1} \neq 0 . \end{cases}$$

(a) Determine whether or not $\{Y_n\}$ is a martingale.

(b) Determine whether or not $\{Y_n\}$ converges almost surely, i.e. whether or not there is a random variable Y such that $\mathbf{P}(\lim_{n\to\infty} Y_n = Y) = 1$.

(c) Relate your answers in parts (a) and (b) to the Martingale Convergence Theorem.

(3.5.5) Problem. (*Pòlya's Urn*) A bag contains Red and Blue balls, with initially $r > 0$ Red balls and $b > 0$ Blue balls. Each minute, a ball is drawn from the bag, its colour noted, and then it is returned to the bag together with a new ball of the same colour. So, after n minutes, there are a total of $r + b + n$ balls in the bag, of which some random number R_n are Red, and the remaining $B_n = r + b + n - R_n$ are Blue. (Here $R_0 = r$ and $B_0 = b$.) Let $Y_n = R_n/(n + r + b)$ be the fraction of Red balls after n minutes.

(a) Prove that $\{Y_n\}$ is a martingale.

(b) Prove there is a random variable Y such that $\mathbf{P}(\lim_{n\to\infty} Y_n = Y) = 1$.

3.6. Application – Branching Processes

- Let μ be any prob dist on $\{0, 1, 2, \ldots\}$, the *offspring distribution*.
  - Let X_n be the number of individuals at time n.
  - (e.g., people with colds, or bacteria, or ...)
  - Start with $X_0 = a$ individuals. Assume $0 < a < \infty$.
  - Each of the X_n individuals at time n has a random number of offspring which is i.i.d. $\sim \mu$, i.e. has i children with probability $\mu\{i\}$. (diagram)
  - (There is just one "parent" per offspring, i.e. asexual reproduction.)
  - Hence, $X_{n+1} = Z_{n,1} + Z_{n,2} + \ldots + Z_{n,X_n}$, where $\{Z_{n,i}\}_{i=1}^{X_n}$ are i.i.d. $\sim \mu$.
  - Here $\{X_n\}$ is Markov chain, on the state space $S = \{0, 1, 2, \ldots\}$.
  - What about the transition probabilities?
  - Well, $p_{00} = 1$, and $p_{0j} = 0$ for all $j \geq 1$. That is, if X_n ever reaches 0, then it stays there forever (this is called *extinction*).
  - But p_{ij} for other i is more complicated; in fact (optional), $p_{ij} = (\mu * \mu * \ldots * \mu)(j)$, a *convolution* of i copies of μ.
- Will $X_n = 0$ for some n?
  - How can martingales help?
- Let $m = \sum_i i\, \mu\{i\}$ = the mean of μ, called the *reproductive number*.
  - What relevance does this number have?
  - Well, assume $0 < m < \infty$.
  - Then $\mathbf{E}(X_{n+1} \,|\, X_0, \ldots, X_n) = \mathbf{E}(Z_{n,1} + Z_{n,2} + \ldots + Z_{n,X_n} \,|\, X_0, \ldots, X_n) = m\, X_n$.
  - So, by induction, $\mathbf{E}(X_n) = m^n \mathbf{E}(X_0) = m^n a < \infty$.
- If $m < 1$, then $\mathbf{E}(X_n) = a\, m^n \to 0$.
  - But $\mathbf{E}(X_n) = \sum_{k=0}^{\infty} k\, \mathbf{P}(X_n = k) \geq \sum_{k=1}^{\infty} \mathbf{P}(X_n = k) = \mathbf{P}(X_n \geq 1)$.
  - Hence, $\mathbf{P}(X_n \geq 1) \leq \mathbf{E}(X_n) = a\, m^n \to 0$, i.e. $\mathbf{P}(X_n = 0) \to 1$.
  - Certain extinction!
- If $m > 1$, then $\mathbf{E}(X_n) = a\, m^n \to \infty$.
  - FACT: In this case, $\mathbf{P}(X_n \to \infty) > 0$, i.e. there is a positive probability that X_n converges to infinity (called *flourishing*).
  - But assuming $\mu\{0\} > 0$, we still also have $\mathbf{P}(X_n \to 0) > 0$ (for example, this will happen if no one has any offspring at all on the first iteration, which has probability $(\mu\{0\})^a > 0$).
  - So, $\mathbf{P}(X_n \to \infty) \leq 1 - \mathbf{P}(X_n \to 0) < 1$.
  - So, if $m > 1$, then we have possible extinction, but also possible flourishing.
- What about the case where $m = 1$?
  - Then $\mathbf{E}(X_n) = \mathbf{E}(X_0) = a$ for all n.
  - Here $\mathbf{E}(X_{n+1} | X_0, \ldots, X_n) = mX_n = X_n$, i.e. $\{X_n\}$ is a martingale.
  - Also, it is clearly non-negative. Hence, by the Martingale Convergence Theorem, we must have $X_n \to X$ w.p. 1, for some random variable X.

- But how can $\{X_n\}$ converge w.p. 1?
- After reaching X, the process $\{X_n\}$ would still continue to fluctuate, i.e. would not converge w.p. 1, unless either
  (a) $\mu\{1\} = 1$ (i.e. the branching process is *degenerate*), or
  (b) $X = 0$.
- Conclusion: If μ is non-degenerate (i.e., $\mu\{1\} < 1$), then $X \equiv 0$, i.e. $\{X_n\} \to 0$ w.p. 1.
- So, there is certain extinction, even when $m = 1$!

- In summary: Assuming $\mu\{0\} > 0$ (so $\mu\{1\} < 1$ too), then extinction is certain if $m \leq 1$, but both flourishing and extinction are possible if $m > 1$.

3.7. Application – Stock Options (Discrete)

- In mathematical finance, it is common to model the price of one share of some *stock* as a random process.
  - For now, we work in discrete time, and suppose that X_n is the price of one share of the stock at each date n.
  - If you buy the stock, then the situation is clear: if X_n increases then you will make a profit, but if X_n decreases then you will suffer a loss.
  - A more interesting situation arises when considering stock *options*:

(3.7.1) Definition. A *(European call) stock option* is the option to buy one share of the stock for some fixed *strike price* K at some fixed future *strike date (time)* $S > 0$.

- Such stock options are commonly traded in actual markets; indeed, millions of dollars are spent on them each day.
- Now, if at the strike time S, the stock price X_S is less than the strike price K, then the option would not be exercised, and would thus be worth exactly zero.
  - But the stock price X_S is more than K, then the option would be exercised to obtain a stock worth X_S for a price of just K, for a net profit of $X_S - K$.
  - Hence, at time S, the stock option is worth $\max(0, X_S - K)$.
  - But at time 0 (when the option is to be purchased), X_S is an unknown (random) quantity.
- This leads to the question: at time 0, what is the *fair price* of the stock option?
  - Is it simply the expected value of its worth at time S, i.e. is it $\mathbf{E}[\max(0, X_S - K)]$?
  - It turns out that this is not the fair price.
  - Instead, the fair price is defined in terms of the concept of *arbitrage*, i.e. a combination of investments which guarantees you a profit no matter how the stock performs.
  - Specifically, the fair price of a stock option is defined to be the *no-arbitrage price*, i.e. the price for the option which makes it impossible to make a guaranteed profit through any combination of buying or selling the option, and buying and selling the stock.

  - It is not immediately clear that such a price exists, nor that it is unique. But it turns out that, under appropriate conditions, it is.

- In what follows, we assume that you have the ability to buy or sell arbitrary amounts of stock and/or the option at any (discrete) time, without incurring any transaction fees.
  - We shall see how to determine the no-arbitrage price.
- We begin with an example.

(3.7.2) Example. Suppose a stock price X_0 at time 0 is equal to 100, and at time S the stock price X_S is random with $\mathbf{P}(X_S = 80) = 9/10$ and $\mathbf{P}(X_S = 130) = 1/10$.

  - Suppose there is an option to buy the stock at time S for $K = 110$.
  - What is the fair (no-arbitrage) price of this option?
  - Well, at time S, the option is worth $\max(0, X_S - K)$, which is equal to 0 if $X_S = 80$, or 20 if $X_S = 130$.
  - Hence, option's expected value at time S is $\mathbf{E}(\text{option}) = \mathbf{E}[\min(0, X_S - K)] = (9/10)(0) + (1/10)(20) = 2$.
  - Is this option's fair price? Is it more? less? Does it even exist?

- Suppose that at time 0, you buy x stock shares (for \$100 each), and y option shares (for \$$c$ each).
  - Note: We allow for the possibility that x or y is negative, corresponding to selling (i.e. *shorting*) the stock or option instead of buying positive amounts.
  - Then if the stock goes up to \$130, then you make $\$130 - \$100 = \$30$ on each stock share, and make $\$20 - \$c = \$(20 - c)$ on each option share, for a total profit of $(130 - 100)x + (20 - c)y = 30x + (20 - c)y$.
  - But if the stock instead goes down to 80, you lose \$20 on each stock share, and lose \$$c$ on each option share, for a total profit of $(80 - 100)x + (0 - c)y = -20x - cy$.
  - To attempt to make a guaranteed profit, regardless of how the stock performs, we could attempt to make these two different total profit amounts equal to each other.
- Indeed, if $y = -(5/2)x$, then these profits both equal $(5/2)(c - 8)x$, i.e. your net profit is no longer random.
  - So, if $c > 8$, then if you buy $x > 0$ stock shares and $y = -(5/2)x < 0$ option shares, then you will make a guaranteed profit of $(5/2)(c - 8)x > 0$.
  - Or, if $c < 8$, then if you buy $x < 0$ stock shares and $y = -(5/2)x > 0$ option shares, then you will make a guaranteed profit of $(5/2)(8 - c)(-x) > 0$.
  - Either way, there is arbitrage, i.e. a guaranteed profit.
  - But if $c = 8$, then these both equal 0, so there is no arbitrage.
  - More generally, if $c = 8$, there is no choice of x and y which makes both possible profits positive.
- In summary, for Example (3.7.2), there is no arbitrage iff $c = 8$.
  - This means that $c = \$8$ is the unique fair (no-arbitrage) price.

- (Not $\mathbf{E}(X_S) = 85$, nor $\mathbf{E}(X_S - K) = -15$, nor even $\mathbf{E}(\text{option}) = \mathbf{E}[\min(0, X_S - K)] = (9/10)(0) + (1/10)(20) = 2$.)

- What is the connection to martingales?
  - Well, suppose we assign the new probabilities that $\mathbf{P}(X_S = 80) = 3/5$ and $\mathbf{P}(X_S = 130) = 2/5$, instead of the actual probabilities 9/10 and 1/10.
  - (These are called the *Martingale probabilities*.)
  - Then, for these probabilities, the stock price is a martingale since $(3/5)(80) + (2/5)(130) = 100 =$ initial price, and also the option is a martingale since $(3/5)(0) + (2/5)(130 - 110) = 8 = c =$ initial price.
  - Then, the fair price = the martingale expected value (i.e. the expected value with respect to the martingale probabilities), i.e. it equals $(3/5)(0) + (2/5)(130 - 110) = 8$.
  - (In fact, the original probabilities 9/10 and 1/10 are irrelevant and do not affect the option's at all value!)
- We summarise this connection as:

(3.7.3) Martingale Pricing Principle. The fair price of an option is equal to its expected value under the martingale probabilities.

(3.7.4) Problem. Suppose a stock costs \$20 today, and will cost either \$10 or \$50 tomorrow. Compute the fair (no-arbitrage) price of a (European call) option to buy the stock tomorrow for \$30, in two ways:

(a) By directly computing the potential profits, as above. [**sol**]

(b) By using the Martingale Pricing Principle. [**sol**]

- By similar reasoning, we now solve a more general case:

(3.7.5) Proposition. Suppose a stock price at time 0 equals $X_0 = a$, and at time $S > 0$ equals either $X_S = d$ (down) or $X_S = u$ (up), where $d < a < u$. Then if $d < K < u$, then at time 0, the fair (no-arbitrage) price of an option to buy the stock at time S for K is equal to: $(a - d)(u - K)/(u - d)$.

Proof #1: Profit Computation. Suppose you buy x shares of the stock, plus y shares of the option.

- Then if the stock goes down to $X_S = d$, your profit is $x(d-a)+y(-c)$. If the stock goes up to $X_S = u$, your profit is $x(u-a)+y(u-K-c)$.
- These are equal if $x(d - u) = y(u - K)$, i.e. $y = x(d - u)/(u - K) = -x(u - d)/(u - K)$, in which case your guaranteed profit is $x(d - a) - yc = x(d - a) + x(u - d)/(u - K)c$.
- If there is no arbitrage, then this equals zero, which implies that $c = (a - d)(u - K)/(u - d)$. ■

Proof #1: Martingale Pricing Principle.. We need to find martingale probabilities $q_1 = \mathbf{P}(X_S = d)$ and $q_2 = \mathbf{P}(X_S = u)$ to make the stock price a martingale.

- So, we need $dq_1+uq_2 = a$, i.e. $dq_1+u(1-q_1) = a$, i.e. $(d-u)q_1+u = a$, i.e. $q_1 = (u - a)/(u - d)$.
- Then $q_2 = 1 - q_1 = (a - d)/(u - d)$.

  - Then, by the Martingale Pricing Principle, the fair price is the martingale expectation of the option's worth, which equals $q_1(0)+q_2(u-K) = [(a-d)/(u-d)] * (u-K)$, the same as before. ∎

(3.7.6) Problem. For the setup of Problem (3.7.4), use the above general formulas to verify the answer obtained previously. **[sol]**

- Similar (but messier) calculations work in the multi-step discrete case too, and the Martingale Pricing Principle still holds.
  - We do not pursue those calculations here, but see e.g. Durrett (2011).
- And, we shall see later that the Martingale Pricing Principle works in continuous time, too.

4. Continuous Processes

- So far, we have mostly considered discrete processes, where the time is indexed by non-negative integers, and the process takes on a finite or countable number of different values.
  - We now consider various generalisations of this to continuous time and/or space.
  - We begin with a continuous generalisation of s.s.r.w., called Brownian motion.

4.1. Brownian Motion

- Let $\{X_n\}_{n=0}^{\infty}$ be s.s.r.w., with $X_0 = 0$.
- Represent this as $X_n = Z_1 + Z_2 + \ldots + Z_n$, where $\{Z_i\}$ are i.i.d. with $\mathbf{P}(Z_i = +1) = \mathbf{P}(Z_i = -1) = 1/2$.
  - That is, $X_0 = 0$, and $X_{n+1} = X_n + Z_{n+1}$.
  - Here $\mathbf{E}(Z_i) = 0$ and $\mathbf{Var}(Z_i) = 1$.
- Let M be a large integer, and let $\{Y_t^{(M)}\}$ be like $\{X_n\}$, except with time sped up by a factor of M, and space shrunk down by a factor of $\sqrt{M}$.
  - That is, $Y_0^{(M)} = 0$, and $Y_{\frac{i+1}{M}}^{(M)} = Y_{\frac{i}{M}}^{(M)} + \frac{1}{\sqrt{M}} Z_{i+1}$. (diagram)
  - Fill in $\{Y_t^{(M)}\}_{t \geq 0}$ by linear interpolation. (file www.probability.ca/sta447/Rbrownian; some images at www.probability.ca/sta447/brownian/)
- Brownian motion $\{B_t\}_{t\geq 0}$ is (intuitively) the limit as $M \to \infty$ of $\{Y_t^{(M)}\}$.
- e.g. since $Y_0^{(M)} = 0$ for all M, also $B_0 = 0$.
- Also, note that $Y_t^{(M)} = \frac{1}{\sqrt{M}}(Z_1 + Z_2 + \ldots + Z_{tM})$ (at least, if $tM \in \mathbf{Z}$; otherwise within $O(1/\sqrt{M})$, which doesn't matter when $M \to \infty$).
  - Thus, $\mathbf{E}(Y_t^{(M)}) = 0$, and $\mathbf{Var}(Y_t^{(M)}) = (\frac{1}{\sqrt{M}})^2 (tM) = t$.
  - As $M \to \infty$, by the Central Limit Theorem (A.6.3), $Y_t^{(M)} \to \text{Normal}(0, t)$.
  - CONCLUSION: $B_t \sim \text{Normal}(0, t)$. ("normally distributed")
- Also, if $0 < t < s$, then $Y_s^{(M)} - Y_t^{(M)} = \frac{1}{\sqrt{M}}(Z_{tM+1} + Z_{tM+2} + \ldots + Z_{sM})$ (at least, if $tM, sM \in \mathbf{Z}$; otherwise within $O(1/\sqrt{M})$).
  - So, $Y_s^{(M)} - Y_t^{(M)} \to \text{Normal}(0, s-t)$, and it is independent of $Y_t^{(M)}$.

- CONCLUSION: $B_s - B_t \sim \text{Normal}(0, s-t)$, and <u>independent</u> of B_t.

- MORE GENERALLY: if $0 \le t_1 \le s_1 \le t_2 \le s_2 \le \ldots \le t_k \le s_k$, then $B_{s_i} - B_{t_i} \sim \text{Normal}(0, s_i - t_i)$, and $\{B_{s_i} - B_{t_i}\}_{i=1}^k$ are all <u>independent</u>. ("independent normal increments")

- Finally, if $0 < t \le s$, then

$$\begin{aligned}\mathbf{Cov}(B_t,\, B_s) \;&=\; \mathbf{E}(B_t B_s) \;=\; \mathbf{E}(B_t[B_s - B_t + B_t]) \\ &=\; \mathbf{E}(B_t[B_s - B_t]) + \mathbf{E}((B_t)^2) \\ &=\; \mathbf{E}(B_t)\,\mathbf{E}(B_s - B_t) + \mathbf{E}((B_t)^2) \\ &=\; (0)(0) + t \;=\; t\,.\end{aligned}$$

  - In general, $\mathbf{Cov}(B_t,\, B_s) = \min(t, s)$. ("covariance structure")

- DEFINITION: <u>Brownian motion</u> is a process $\{B_t\}_{t\ge 0}$ with $B_0 = 0$ satisfying the above three properties [1. $B_t \sim N(0,t)$; 2. Independent normal increments; 3. $\mathbf{Cov}(B_s, B_t) = \min(s,t)$], and also with continuous sample paths (i.e., the random function $t \to B_t$ is always continuous).

  - FACT: Such a process exists! (The above construction is <u>intuitive</u>; a formal proof of existence requires measure theory; see e.g. Billingsley, 1995, or many other advanced probability and stochastic processes books.)

- Example: What is $\mathbf{E}[(B_2 + B_3 + 1)^2]$?

  - Well, $\mathbf{E}[(B_2 + B_3 + 1)^2] = \mathbf{E}[(B_2)^2)] + \mathbf{E}[(B_3)^2] + 1^2 + 2\mathbf{E}[B_2 B_3] + 2\mathbf{E}[B_2(1)] + 2\mathbf{E}[B_3(1)] = 2 + 3 + 1 + 2(2) + 2(0) + 2(0) = 10$.

- Example: What is $\mathbf{Var}[B_3 + B_5 + 6]$?

  - Well, $\mathbf{Var}[B_3 + B_5 + 6] = \mathbf{E}[((B_3 + B_5 + 6) - E(B_3 + B_5 + 6))^2] = \mathbf{E}[((B_3 + B_5 + 6) - 6))^2] = \mathbf{E}[(B_3 + B_5)^2]$
    $= \mathbf{E}[(B_3)^2)] + \mathbf{E}[(B_5)^2] + 2\mathbf{E}[B_3 B_5] = 3 + 5 + 2(3) = 14$.

- Example: Let $\alpha > 0$, and let $W_t = \alpha\, B_{t/\alpha^2}$.

  - Then $W_t \sim \text{Normal}(0,\ \alpha^2(t/\alpha^2)) = \text{Normal}(0, t)$. (same as for B_t)
  - Also for $0 < t < s$, $\mathbf{E}(W_t W_s) = \alpha^2 \mathbf{E}(B_{t/\alpha^2} B_{s/\alpha^2}) = \alpha^2(t/\alpha^2) = t$.
  - In fact, $\{W_t\}$ has all the same properties as $\{B_t\}$.
  - That is, $\{W_t\}$ "is" Brownian motion, too. ("transformation")

- If $0 < t < s$, then <u>given</u> B_r for $0 \le r \le t$, what is the <u>conditional</u> distribution of B_s?

  - Well, $B_s \,|\, B_t \;=\; B_t + (B_s - B_t) \;|\; B_t \;=\; B_t + \text{Normal}(0,\, s-t) \;\sim\; \text{Normal}(B_t,\, s-t)$.
  - That is, given B_t, B_s is normal with mean B_t and variance $s - t$.
  - So, in particular, $\mathbf{E}[B_s \,|\, \{B_r\}_{0 \le r \le t}] = B_t$.
  - Hence, $\{B_t\}$ is a (continuous-time) <u>martingale</u>!
  - So, similar results apply just like for discrete-time martingales.

- Example: let $a, b > 0$, and let $\tau = \min\{t \ge 0 : B_t = -a \text{ or } b\}$.

  - What is $p \equiv \mathbf{P}(B_\tau = b)$?

END OF WEEK #8

- Well, here $\{B_t\}$ is martingale, and τ is stopping time.
- Furthermore, $\{B_t\}$ is bounded up to time τ, i.e. $|B_t| \mathbf{1}_{t \le \tau} \le \max(|a|, |b|)$.
- So, just like for discrete martingales, must have $\mathbf{E}(B_\tau) = \mathbf{E}(B_0) = 0$.
- Hence, $p(b) + (1-p)(-a) = 0$, so $p = \frac{a}{a+b}$. (as expected)

- But what is $e \equiv \mathbf{E}(\tau)$?
- To solve for e, let $Y_t = B_t^2 - t$.
  - Then for $0 < t < s$, $\mathbf{E}[Y_s \,|\, \{B_r\}_{r \le t}] = \mathbf{E}[B_s^2 - s \,|\, \{B_r\}_{r \le t}]$
    $= \mathbf{Var}[B_s \,|\, \{B_r\}_{r \le t}] + (\mathbf{E}[B_s \,|\, \{B_r\}_{r \le t}])^2 - s$
    $= (s-t) + (B_t)^2 - s = (B_t)^2 - t = Y_t$.
  - Then by the Double-expectation formula (A.5.5),

$$\mathbf{E}[Y_s \,|\, \{Y_r\}_{r \le t}] \;=\; \mathbf{E}\Big[\mathbf{E}[Y_s \,|\, \{B_r\}_{r \le t}] \,\Big|\, \{Y_r\}_{r \le t}\Big] = \mathbf{E}[Y_t \,|\, \{Y_r\}_{r \le t}] = Y_t\,.$$

  - So, $\{Y_t\}$ is also a martingale!
- Back to $\tau = \min\{t \ge 0 : B_t = -a \text{ or } b\}$. What is $e \equiv \mathbf{E}(\tau)$?
  - Well, with $Y_t = B_t^2 - t$, have $\mathbf{E}(Y_\tau) = \mathbf{E}(B_\tau^2 - \tau) = \mathbf{E}(B_\tau^2) - \mathbf{E}(\tau) = pb^2 + (1-p)(-a)^2 - e = \frac{a}{a+b} b^2 + \frac{b}{a+b} a^2 - e = ab - e$.
  - Assuming $\mathbf{E}(Y_\tau) = 0$, we can solve this to get that $e = ab$. (just like discrete Gambler's Ruin)
  - But τ is not bounded. So, how to show that $\mathbf{E}(Y_\tau) = 0$?
- To justify that $\mathbf{E}(Y_\tau) = 0$:
  - Let $\tau_M = \min(\tau, M)$.
  - Then τ_M is bounded, so $\mathbf{E}(Y_{\tau_M}) = 0$.
  - But $Y_{\tau_M} = B_{\tau_M}^2 - \tau_M$, so $\mathbf{E}(\tau_M) = \mathbf{E}(B_{\tau_M}^2)$.
  - As $M \to \infty$, $\mathbf{E}(\tau_M) \to \mathbf{E}(\tau)$ by the Monotone Convergence Thm, and $\mathbf{E}(B_{\tau_M}^2) \to \mathbf{E}(B_\tau^2)$ by the Bounded Convergence Thm.
  - Therefore, $\mathbf{E}(\tau) = \mathbf{E}(B_\tau^2)$, i.e. $\mathbf{E}(Y_\tau) = 0$ as above.
- Fact: Although the function $t \mapsto B_t$ is always continuous everywhere, with probability 1 it is differentiable nowhere!
  - Intuitive reason #1: $\{Y_t^{(M)}\}$ has non-differentiable "spikes" at $t = i/M$ for all $i \in \mathbf{Z}$, and as $M \to \infty$ there are more and more spikes.
  - Intuitive reason #2: If it were differentiable, then its derivative at t would be $\lim_{h \to \infty} \frac{1}{h}(B_{t+h} - B_t)$. But $B_{t+h} - B_t \sim N(0, h)$, so $\frac{1}{h}(B_{t+h} - B_t) \sim N(0, \frac{1}{h})$, with variance $\to \infty$ as $h \to 0$.
  - Intuitive reason #3: Since $B_{t+h} - B_t \sim N(0, h)$, therefore $\mathbf{E}\left([\frac{1}{h}(B_{t+h} - B_t)]^2\right) = \frac{1}{h^2}(h) = \frac{1}{h}$ which $\to \infty$ as $h \to 0$.
  - Intuitive reason #4: Since $B_{t+h} - B_t \sim N(0, h)$, therefore $B_{t+h} - B_t \approx O(\sqrt{h})$, so $\frac{1}{h}(B_{t+h} - B_t) \approx O(\sqrt{1/h})$, which $\to \infty$ as $h \to 0$.
- Example: Suppose $X_t = 2 + 5t + 3B_t$ for $t \ge 0$.
  - What are $\mathbf{E}(X_t)$ and $\mathbf{Var}(X_t)$ and $\mathbf{Cov}(X_t, X_s)$?
  - Well, $\mathbf{E}(X_t) = 2 + 5t$, and $\mathbf{Var}(X_t) = 3^2 \mathbf{Var}(B_t) = 9t$.
  - Also for $0 < t < s$, $\mathbf{Cov}(X_t, X_s) = \mathbf{E}[(X_t - 5t - 2)(X_s - 5s - 2)] = \mathbf{E}[(3B_t)(3B_s)] = 9\,\mathbf{E}[B_t\, B_s] = 9t$.
  - Note: Also follows that $X_t \sim \text{Normal}(2 + 5t, 9t)$.
  - Fancy notation: $dX_t = 5\,dt + 3\,dB_t$. ("diffusion")

- More generally, could have $X_t = x_0 + \mu t + \sigma B_t$. (file "Rbrownian")
  - Then $dX_t = \mu\, dt + \sigma\, dB_t$. ($\mu$ = "drift"; σ = "volatility"; $\sigma \geq 0$)
  - Then $\mathbf{E}(X_t) = x_0 + \mu t$, and $\mathbf{Var}(X_t) = \sigma^2 t$, and $\mathbf{Cov}(X_t, X_s) = \sigma^2 \min(s,t)$.
  - Here $X_t \sim N(x_0 + \mu t,\, \sigma^2 t)$.
  - Optional: Even more generally, could have $dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dB_t$, where μ and σ are functions, i.e. non-constant drift and volatility.

(4.1.1) Problem. Let $\{B_t\}_{t\geq 0}$ be Brownian motion. Compute $\mathbf{Var}(B_5 B_8)$, the variance of $B_5 B_8$. [Hint: You may use without proof that if $Z \sim$ Normal$(0,1)$, then $\mathbf{E}(Z) = \mathbf{E}(Z^3) = 0$, $\mathbf{E}(Z^2) = 1$, and $\mathbf{E}(Z^4) = 3$. And don't forget that $B_8 = B_5 + (B_8 - B_5)$.]

(4.1.2) Problem. Let $\{B_t\}_{t\geq 0}$ be Brownian motion. Let $\theta \in \mathbf{R}$, and let $Y_t = \exp(\theta B_t - \theta^2 t/2)$. Prove that $\{Y_t\}_{t\geq 0}$ is a martingale. [Hint: You may use without proof that if $Z \sim$ Normal$(0,1)$, and $a \in \mathbf{R}$, then $\mathbf{E}[e^{aZ}] = e^{a^2/2}$.]

4.2. Application – Stock Options (Continuous)

- Assume now that the stock price X_t is equal to some (random) positive value, for each time $t \geq 0$.
- Common model for the stock price: $X_t = x_0 \exp(\mu t + \sigma B_t)$.
  - i.e. if $Y_t = \log(X_t)$, then $Y_t = y_0 + \mu t + \sigma B_t$, i.e. $dY_t = \mu\, dt + \sigma\, dB_t$.
  - That is, changes occur proportional to total price (makes sense).
  - So, $Y_t = \log(X_t)$ is a diffusion.
- Also assume a risk-free interest rate r, so that \$1 today is worth \$$e^{rt}$ a time t years later.
  - Equivalently, \$1 at a future time $t > 0$ is worth \$$e^{-rt}$ at time 0 (i.e. "today").
  - So, "discounted" stock price (in "today's dollars") is

$$D_t \;\equiv\; e^{-rt} X_t \;=\; e^{-rt} x_0 \exp(\mu t + \sigma B_t) \;=\; x_0 \exp((\mu - r)t + \sigma B_t)\,.$$

  - Special case: if $r = 0$ then no discounting.
  - (Discounting might not seem important, but it turns out to provide a key step below.)
- Defn: A "European call option" is the option to buy the stock for some amount \$$K$ at some fixed future time $S > 0$ years later.
  - At time S, this is worth $\max(0,\, X_S - K)$.
  - So, at time 0, it's worth the "discounted" value $e^{-rS} \max(0,\, X_S - K)$.
  - But at time 0, X_S is unknown (random), so this (discounted) value is a random variable.
- QUESTION: What is the "fair price" of this option?
  - As before, this means the "no-arbitrage" price, i.e. a price such that you cannot make a guaranteed profit by combinations of buying or selling the option, and buying and selling the stock.
  - Note: Again we assume the ability to buy/sell arbitrary amounts of stock and/or at any time, infinitely often, including going negative (i.e., "shorting" the stock or option), with no transaction fees.

- So, what is the fair price at time 0?
- Is it simply the expected value, $\mathbf{E}[e^{-rS} \max(0, X_S - K)]$?
- No! This would allow for arbitrage!

- So, what is the fair price?
- Use the Martingale Pricing Principle again!
- KEY: If $\mu = r - \frac{\sigma^2}{2}$, then it can be computed (Problem (4.2.2)) that $\{D_t\}$ becomes a martingale.
- FACT: If $\{D_t\}$ is a martingale, then similar to the discrete case, there is a (no-arbitrage) continuous-time fair price process $\{C_t\}$ which makes the option value into a martingale too! (finance/actuarial classes ...)
  - Hence, the Martingale Pricing Principle still holds: Under the martingale probabilities, the initial value (fair price) of the option at time zero is the same as the expected value of the option at time S.
  - (For a proof, see e.g. Theorem 1.2.1 of I. Karatzas (1997), *Lectures on the Mathematics of Finance*, CRM Monograph Series **8**, American Mathematical Society.)
- CONCLUSION: The fair price for the option equals $\mathbf{E}[e^{-rS} \max(0, X_S - K)]$, but only after replacing μ by $r - \frac{\sigma^2}{2}$.
  - i.e., such that $X_S = x_0 \exp([r-\frac{\sigma^2}{2}]S+\sigma B_S)$, where $B_S \sim \text{Normal}(0, S)$.
- So, the fair price can be computed by an integral (with respect to a normal density). After some computation (Problem (4.2.3)), this fair price becomes:

  (4.2.1)
  $$x_0 \, \Phi\left(\frac{(r + \frac{\sigma^2}{2})S - \log(K/x_0)}{\sigma\sqrt{S}}\right) - e^{-rS} K \, \Phi\left(\frac{(r - \frac{\sigma^2}{2})S - \log(K/x_0)}{\sigma\sqrt{S}}\right),$$

  where $\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-v^2/2} \, dv$ is the cdf of a standard normal distribution. ["Black-Scholes formula". Do not have to memorise!]
- Can compute with e.g. www.probability.ca/sta447/Rblackscholes or `https://www.mystockoptions.com/black-scholes.cfm`.
  - e.g. Suppose $x_0 = 100$ (dollars), $K = 110$ (dollars), $S = 1$ (years), $r = 0.05$ (i.e. 5% per year), $\sigma = 0.3$ (i.e. 30% per year). Then fair option price is:
    \$10.02.
  - Or, if $x_0 = 200$, $K = 250$, $S = 2$, $r = 0.1$, $\sigma = 0.5$, then price is: \$53.60.
  - Or, if $x_0 = 200$, $K = 250$, $S = 2$, $r = 0.1$, but $\sigma = 0.8$, then price is: \$84.36 (more).
- Note: This price does not depend on the drift ("appreciation rate") μ, since we first have to replace μ by $r - \frac{\sigma^2}{2}$. This seems surprising!
  - Intuition: if μ large, then can make good money from the stock itself, so the option doesn't add much value ...
- However, the price is an increasing function of the volatility σ.
  - This makes sense, since the option "protects" you against large drops in the stock price.

(4.2.2) Problem. Let $\{B_t\}_{t\geq 0}$ be Brownian motion, let $X_t = x_0 \exp(\mu t +$

$\sigma B_t)$ be the stock price model (where $\sigma > 0$), and let $D_t = e^{-rt} X_t$ be the discounted stock price. Show that if $\mu = r - \frac{\sigma^2}{2}$, then $\{D_t\}$ is a martingale. [Hint: Don't forget Problem (4.1.2).]

(4.2.3) Problem. Let $\{B_t\}_{t\geq 0}$ be Brownian motion, and let $X_t = x_0 \exp(\mu t + \sigma B_t)$ be the stock price model (where $\sigma > 0$). Show that if $\mu = r - \frac{\sigma^2}{2}$, then $\mathbf{E}\Big[e^{-rS} \max(0, X_S - K)\Big]$ is equal to the formula (4.2.1). [Hint: Write the expectation as an integral with respect to the density function for B_S. Then, break up the integral into the part where $X_S - K \geq 0$ and the part where $X_S - K < 0$.]

(4.2.4) Problem. Consider the price formula (4.2.1), with r, σ, S, and x_0 fixed positive quantities.

(a) What happens to the price (4.2.1) as $K \searrow 0$? Does this result make intuitive sense?

(b) What happens to the price (4.2.1) as $K \to \infty$? Does this result make intuitive sense?

(4.2.5) Problem. Consider the price formula (4.2.1), with r, σ, x_0, and K fixed positive quantities.

(a) What happens to the price (4.2.1) as $S \searrow 0$? [Hint: Consider separately the cases $K > x_0$, $K = x_0$, and $K < x_0$.] Does this result make intuitive sense?

(b) What happens to the price (4.2.1) as $T \to \infty$? Does this result make intuitive sense?

4.3. Poisson Processes

- MOTIVATING EXAMPLE:
  - Suppose an average of $\lambda = 2.5$ fires in Toronto per day.
  - Intuitively, this is caused by a very large number n of buildings, each of which has a very small probability p of having a fire.
  - Then mean $= np = \lambda$, so $p = \lambda/n$.
  - Then # fires today is Binomial(n, p) = Binomial$(n, \lambda/n)$.
  - So, as $n \to \infty$, using the binomial formula (A.2.2),

$$\mathbf{P}(\#\text{fires} = k) \;=\; \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} (\lambda/n)^k (1-(\lambda/n))^{n-k}$$

$$\approx \frac{n^k}{k!} (\lambda/n)^k (e^{-\lambda/n})^n$$

$$= \frac{1}{k!} \lambda^k e^{-\lambda}.$$

  - i.e. # fires $\sim$ Poisson(λ) = Poisson(2.5).
  - So, if $\lambda = 2.5$, then $\mathbf{P}(\#\text{ fires} = k) \approx e^{-2.5} \frac{(2.5)^k}{k!}$, for $k = 0, 1, 2, 3, \ldots$
  - And, # fires today and tomorrow combined $\approx$ Poisson$(2 * \lambda)$ = Poisson(5), etc.

- POISSON PROCESS CONSTRUCTION:
- Let $\{Y_n\}_{n=1}^\infty$ be i.i.d. $\sim$ Exponential(λ), for some $\lambda > 0$.
  - So, Y_n has density function $\lambda e^{-\lambda y}$ for $y > 0$.
  - And, $\mathbf{P}(Y_n > y) = e^{-\lambda y}$ for $y > 0$.
  - And, $\mathbf{E}(Y_n) = 1/\lambda$.
- Let $T_0 = 0$, and $T_n = Y_1 + Y_2 + \ldots + Y_n$ for $n \geq 1$. ("n^{th} arrival time")
  - [e.g. T_n = time of n^{th} fire.]
- Let $N(t) = \max\{n \geq 0 : T_n \leq t\} = \#\{n \geq 1 : T_n \leq t\} = \#$ arrivals up to time t.
  - "Counting process". (Counts number of arrivals.)
  - [e.g. $N(t) = \#$ fires between times 0 and t.]
  - "Poisson process with intensity λ"
- What is distribution of $N(t)$, i.e. $\mathbf{P}(N(t) = m)$?
- PROPOSITION: $\mathbf{P}(N(t) = m) \;=\; \frac{(\lambda t)^m}{m!} e^{-\lambda t}$, for $m = 0, 1, 2, 3, \ldots$.
- PROOF:
  - Well, $N(t) = m$ iff both $T_m \leq t$ and $T_{m+1} > t$, which is iff there is $0 \leq s \leq t$ with $T_m = s$ and $T_{m+1} - T_m > t - s$.
  - Recall that $Y_n \sim$ Exponential(λ) = Gamma($1, \lambda$), so $T_m := Y_1 + Y_2 + \ldots + Y_m \sim$ Gamma(m, λ), with density function $f_{T_m}(s) = \frac{\lambda^m}{\Gamma(m)} s^{m-1} e^{-\lambda s} = \frac{\lambda^m}{(m-1)!} s^{m-1} e^{-\lambda s}$.
  - Also $\mathbf{P}(T_{m+1} - T_m > t - s) = \mathbf{P}(Y_{m+1} > t - s) = e^{-\lambda(t-s)}$. So,

$$\mathbf{P}(N(t) = m) \;=\; \mathbf{P}(T_m \leq t,\, T_{m+1} > t) \;=\; \mathbf{P}(\exists\, 0 \leq s \leq t:\; T_m = s,\; Y_{m+1} > t - s)$$

$$= \int_0^t f_{T_m}(s)\, \mathbf{P}(Y_{m+1} > t - s)\, ds \;=\; \int_0^t \frac{\lambda^m}{(m-1)!} s^{m-1} e^{-\lambda s}\, e^{-\lambda(t-s)}\, ds$$

$$= \frac{\lambda^m}{(m-1)!} e^{-\lambda t} \int_0^t s^{m-1}\, ds \;=\; \frac{\lambda^m}{(m-1)!} e^{-\lambda t} [\frac{t^m}{m}] \;=\; \frac{(\lambda t)^m}{m!} e^{-\lambda t}, \quad Q.E.D.$$

- Hence, $N(t) \sim$ Poisson(λt).
- So, $\mathbf{E}(N(t)) = \lambda t$, and $\mathbf{Var}(N(t)) = \lambda t$.
- Now, recall the "memoryless" (or "forgetfulness") property of the Exponential(λ) distribution: for $a, b > 0$, $\mathbf{P}(Y_n > b + a \,|\, Y_n > a) = \mathbf{P}(Y_n > b) = e^{-\lambda b}$.
  - This means the process $\{N(t)\}$ "starts over" in each new time interval.
  - It follows that $N(t + s) - N(s) \sim N(t) \sim$ Poisson(λt).
  - Also follows that if $0 \leq a < b \leq c < d$, then $N(d) - N(c)$ indep. of $N(b) - N(a)$, and similarly for multiple non-overlapping time intervals. ("independent increments")
- MORE GENERALLY: if $0 \leq t_1 \leq s_1 \leq t_2 \leq s_2 \leq \ldots \leq t_k \leq s_k$, then $N(s_i) - N(t_i) \sim$ Poisson($\lambda(s_i - t_i)$), and $\{N(s_i) - N(t_i)\}_{i=1}^k$ are all independent. ("independent Poisson increments")
- DEFN: A Poisson processes with intensity $\lambda > 0$ is a collection $\{N(t)\}_{t \geq 0}$ of random non-decreasing integer counts $N(t)$, satisfying: 1. $N(0) = 0$; 2. $N(t) \sim$ Poisson(λt) for all $t \geq 0$; and 3. Independent Poisson increments.
- MOTIVATING EXAMPLE (cont'd):
  - Here, fires approximately follow a Poisson process with intensity 2.5.

- So, $\mathbf{P}$(9 fires today and tomorrow combined) $\approx e^{-2*2.5}\frac{(2*2.5)^9}{9!} = e^{-5}(\frac{5^9}{9!}) \doteq$ 0.036.
- $\mathbf{P}$(at least one fire in next hour)
  $= 1 - \mathbf{P}$(no fires in next hour)
  $= 1 - \mathbf{P}(N(1/24) = 0)$
  $= 1 - e^{-2.5/24}\frac{(2.5/24)^0}{0!} \doteq 1 - 0.90 = 0.10$.
- $\mathbf{P}$(exactly 3 fires in next hour)
  $= e^{-2.5/24}\frac{(2.5/24)^3}{3!} \doteq 0.00017 \doteq 1/5891$, etc.

- EXAMPLE: Let $\{N(t)\}$ be a Poisson process with intensity $\lambda = 2$. Then

$$\mathbf{P}[N(3) = 5,\ N(3.5) = 9]$$

$$= \mathbf{P}[N(3) = 5,\ N(3.5) - N(3) = 4]$$

$$= \mathbf{P}[N(3) = 5]\ \mathbf{P}[N(3.5) - N(3) = 4]$$

$$= \left[e^{-\lambda(3)}\frac{(\lambda(3))^5}{5!}\right]\ \left[e^{-\lambda(0.5)}\frac{(\lambda(0.5))^4}{4!}\right]$$

$$= \left(e^{-6}\frac{6^5}{120}\right)\ \left(e^{-1}\frac{1^4}{24}\right) = e^{-7}(2.7) \doteq 0.0025 \doteq 1/400\,.$$

- EXAMPLE: Let $\{N(t)\}$ be a Poisson process with intensity λ.
  - Then for $0 < t < s$,

$$\mathbf{P}(N(t) = 1 \,|\, N(s) = 1)$$

$$= \frac{\mathbf{P}(N(t) = 1,\ N(s) = 1)}{\mathbf{P}(N(s) = 1)}$$

$$= \frac{\mathbf{P}(N(t) = 1,\ N(s) - N(t) = 0)}{\mathbf{P}(N(s) = 1)}$$

$$= \frac{e^{-\lambda t}\frac{(\lambda t)^1}{1!}\, e^{-\lambda(s-t)}\frac{(\lambda(s-t))^0}{0!}}{e^{-\lambda s}\frac{(\lambda s)^1}{1!}}$$

$$= t/s\,.$$

  - That is, conditional on $N(s) = 1$, the first event is uniform over $[0, s]$. (Distribution does not depend on λ.)

- Also, e.g.

$$\mathbf{P}(N(4) = 1 \,|\, N(5) = 3) = \frac{\mathbf{P}(N(4) = 1,\ N(5) = 3)}{\mathbf{P}(N(5) = 3)}$$

$$= \frac{\mathbf{P}(N(4) = 1,\ N(5) - N(4) = 2)}{\mathbf{P}(N(5) = 3)}$$

$$= \frac{(e^{-4\lambda}(4\lambda)^1/1!)(e^{-\lambda}\lambda^2/2!)}{e^{-5\lambda}(5\lambda)^3/3!} = \frac{(4)^1/1!)(1/2!)}{(5)^3/3!}$$

$$= \frac{4/2}{125/6} = 24/250 = 12/125\,.$$

  - This also does not depend on λ. [And equals $\binom{3}{1}(4/5)^1(1/5)^2$. Why?]

- ALTERNATIVE APPROACH:

(4.3.1) Proposition. If $N(t)$ is a Poisson process with rate $\lambda > 0$, then as $h \searrow 0$:
(a) $\mathbf{P}(N(t+h) - N(t) = 1) \ = \ \lambda h + o(h)$.
(b) $\mathbf{P}(N(t+h) - N(t) \geq 2) \ = \ o(h)$.

Proof. Here $\mathbf{P}(N(t+h) - N(t) = 1) = \mathbf{P}(N(h) = 1) = e^{\lambda h}(\lambda h)^1/1! = [1 + \lambda h + O(h^2)](\lambda h) = \lambda h + \lambda^2 h^2 + O(h^3) = \lambda h + O(h^2) = \lambda h + o(h)$.

Also, $\mathbf{P}(N(t+h) - N(t) \geq 2) = \mathbf{P}(N(h) \geq 2) = 1 - \mathbf{P}(N(h) = 0) - \mathbf{P}(N(h) = 1) = 1 - e^{\lambda h}(\lambda h)^0/0! - e^{\lambda h}(\lambda h)^1/1! = 1 - [1+\lambda h + O(h^2)](1) - [1+\lambda h + O(h^2)](\lambda h) = 1 - 1 + \lambda h - \lambda h + \lambda^2 h^2 + O(h^2) + O(h^3) = O(h^2) = o(h)$. ∎

- FACT: Proposition (4.3.1) has a converse: If $X(t)$ is a stochastic process with independent increments which satisfies properties (a) and (b), then $X(t)$ must be a Poisson process with rate λ.
  - Hence, this provides another way to define and characterise Poisson processes.
  - Also, Poisson processes are the only integer-valued processes with independent increments such that $\mathbf{E}[N(s+t) - N(s)] = \lambda t \ \forall s, t > 0$.
- What about "clumping"?
  - Suppose we have a Poisson Process on [0,100] with intensity $\lambda = 1$.
  - Then expect about one event every time distance 1.
  - What is $\mathbf{P}(\exists r \in [1, 100] : N(r) - N(r-1) = 4)$?
  - Well,

$$\begin{aligned}
&\mathbf{P}(\exists r \in [1, 100] : N(r) - N(r-1) = 4) \\
&\geq \mathbf{P}(\exists m \in \{1, 2, \ldots, 100\} : N(m) - N(m-1) = 4) \\
&= 1 - \mathbf{P}(\nexists m \in \{1, 2, \ldots, 100\} : N(m) - N(m-1) = 4) \\
&= 1 - (\mathbf{P}(N(1) \neq 4))^{100} = 1 - (1 - P(N(1) = 4))^{100} \\
&= 1 - (1 - e^{-1}(1^4/4!))^{100} = 1 - (1 - 1/24e)^{100} \doteq 0.787\,.
\end{aligned}$$

  - That is, there will probably be some clumps!
  - This is an illustration of:
- POISSON CLUMPING: the $\{T_i\}$ tend to "clump up" in various patterns just by chance alone.
  - Doesn't "mean" anything at all: they're independent.
  - But it "seems" like it does have meaning!
  - See e.g. www.probability.ca/pois
- APPLICATION: pedestrian deaths example (true story).
  - 7 pedestrian deaths in Toronto (14 in GTA) in January 2010.
  - Media hype, friends concerned, etc.
  - Facts: Toronto averages about 31.9 per year, i.e. $\lambda = 2.66$ per month.
  - So, $\mathbf{P}(7 \text{ or more}) = \sum_{j=7}^{\infty} e^{-2.66} \frac{(2.66)^j}{j!} \doteq 1.9\%$, about once per 52 months, i.e. about once per 4.4 years.
  - Not so rare! doesn't "mean" anything! (Though tragic.) "Poisson clumping"
  - See e.g. www.probability.ca/ped

  - Later, just two in Feb 1 - Mar 15, 2010; less than expected (4), but no media!

- APPLICATION – BUS MODEL:
  - Suppose have λ buses per hour, i.e. about n buses every n/λ hours.
  - Suppose the arrival times are completely random.
  - Model this as $T_1, T_2, \ldots, T_n \sim \text{Uniform}[0,\ n/\lambda]$, i.i.d.
  - Then for $0 < a < b$, as $n \to \infty$,

$$\#\{i : T_i \in [a,b]\} \ \sim\ \text{Binomial}(n, \frac{b-a}{n/\lambda})$$

$$=\ \text{Binomial}(n, \frac{\lambda(b-a)}{n})\ \approx\ \text{Poisson}(\lambda(b-a))\,.$$

  - Like a Poisson process!

- APPLICATION: WAITING TIME PARADOX.
  - Suppose there are an average of λ buses per hour. (e.g. $\lambda = 5$)
  - You arrive at the bus stop at a random time.
  - What is your expected waiting time until the next bus?
  - If buses are completely regular, then there is one bus every $1/\lambda$ minutes. (e.g. $\lambda = 5$: one bus every 12 minutes)
  - Then the waiting time is $\sim \text{Uniform}[0, \frac{1}{\lambda}]$, so mean $= \frac{1}{2\lambda}$ hours. (e.g. $\lambda = 5$, mean $= \frac{1}{10}$ hours $= 6$ minutes)
  - However, if the buses are instead completely random, then they form a Poisson process with rate λ.
  - So, by the memoryless property, the waiting time is $\sim \text{Exponential}(\lambda)$.
  - So, the mean waiting time is $\frac{1}{\lambda}$ hours. Twice as long! (e.g. $\lambda = 5$, mean $= \frac{1}{5}$ hours $= 12$ minutes)
  - But same number of buses! Contradiction??
  - No: you're more likely to arrive during a longer gap.

- Aside: What about streetcars? They can't pass each other, so they sometimes clump up even more than do (independent) buses. (e.g. Spadina streetcar; see this paper)

- SUPERPOSITION: Suppose $\{N_1(t)\}_{t\ge0}$ and $\{N_2(t)\}_{t\ge0}$ are two independent Poisson processes, with rates λ_1 and λ_2 respectively. Let $N(t) = N_1(t) + N_2(t)$.
  - Then $\{N(t)\}_{t\ge0}$ is a Poisson process with rate $\lambda_1 + \lambda_2$.
  - Proof? Sum of two independent Poissons is Poisson!

- EXAMPLE:
  - Suppose undergrads arrive for office hours according to a Poisson process with intensity $\lambda_1 = 5$ (i.e. one every 12 minutes on average).
  - And, grads arrive independently according to their own Poisson process with intensity $\lambda_2 = 3$ (i.e. one every 20 minutes on average).
  - Then, what is expected number of minutes until first student arrives?
  - Well, total # arrivals $N(t)$ is Poisson process with $\lambda = \lambda_1 + \lambda_2 = 5 + 3 = 8$.
  - Let A = time of first arrival.

- Then, $\mathbf{P}(A > t) = \mathbf{P}(N(t) = 0) = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t}$; so $A \sim$ Exponential(λ).
- Hence, $\mathbf{E}(A) = 1/\lambda = 1/8$ hours, i.e. 7.5 minutes.

- THINNING: Let $\{N(t)\}_{t\geq 0}$ be a Poisson process with rate λ.
  - Suppose each arrival is independently of "type i" with probability p_i, for $i = 1, 2, 3, \ldots$. (e.g. bus or streetcar, male or female, undergrad or grad, etc.)
  - Let $N_i(t)$ be number of arrivals of type i up to time t.
  - THM: The $\{N_i(t)\}$ are <u>independent</u> Poisson processes, with rates λp_i.
  - PROOF: "independent increments" is obvious.
  - For the distribution, suppose for notational simplicity that there are just <u>two</u> types, with $p_1 + p_2 = 1$.
  - Need to show: $\mathbf{P}(N_1(t) = j,\ N_2(t) = k)$ $= \Big(e^{-(\lambda p_1 t)}(\lambda p_1 t)^j/j!\Big)\ \Big(e^{-(\lambda p_2 t)}(\lambda p_2 t)^k/k!\Big)$.
  - But $\mathbf{P}(N_1(t) = j,\ N_2(t) = k)$ $= \mathbf{P}(j+k$ arrivals up to time t, of which j of type 1 and k of type 2) $= \Big(e^{-\lambda t}(\lambda t)^{j+k}/(j+k)!\Big) \times \binom{j+k}{j}(p_1)^j(p_2)^k$. Equal! (Check.)
- EXAMPLE: If students arrive for office hours according to a Poisson process, and each student is independently either undergrad (prob p_1) or grad (prob p_2), then # undergrads is <u>independent</u> of # grads (and each follows a Poisson distribution).
- NOTE: Can also have <u>time-inhomogeneous</u> Poisson processes, where $\lambda = \lambda(t)$, and $N(b) - N(a) \sim \text{Poisson}\Big(\int_a^b \lambda(t)\,dt\Big)$.
- NOTE: Can also have Poisson processes on other <u>regions</u>, e.g. in two dimensions, etc., cf. www.probability.ca/pois

NOTE: I will try to add some Poisson Process practice problems later. For now, there are lots of problems beginning on page 94 of the Durrett book available at: `http://www.math.duke.edu/~rtd/EOSP/EOSP2E.pdf`

END OF WEEK #9

Reminder: Midterm #2 next week during class time; surname A-L in Bahen (BA), 40 St. George St, room 1160; surname M-Z in Haultain (HA), 170 College St (rear), room 403.

A. Appendix: Background

Before reading these notes, you should already have a solid background in undergraduate-level probability theory, including discrete and continuous probability distributions, random variables, joint distributions, expected values, inequalities, conditional probability, limit theorems, etc. (but not measure theory).

You should also know basic calculus concepts such as derivatives and integrals, and basic real analysis including limits and infinite series, and such linear algebra notions as matrix multiplication and eigenvectors.

This Appendix provides background material and certain supplementary results that will be required at various points in these notes.

A.1. Notation

The following notation will be used throughout the book.

$\mathbf{Z}$ are the integers, $\mathbf{N}$ are the positive integers, and $\mathbf{R}$ are the real numbers.

X, Y, etc. usually denote *random variables*, which take on different values with different probabilities.

$\mathbf{P}(\cdots)$ is the *probability* of an event.

$\mathbf{E}(\cdots)$ is the *expected value* of a random variable.

$\{X_n\}_{n=0}^{\infty}$ is an infinite sequence of random variables. It will often denote a *stochastic process*, such as a *Markov chain* or a *martingale*.

$\mathbf{P}_i(\cdots)$ and $\mathbf{E}_i(\cdots)$ are probability and expectation (respectively) of events *conditional* on $X_0 = i$, i.e. *assuming* that the stochastic process $\{X_n\}$ begins in the state i.

$\mathbf{1}_{\ldots}$ is the *indicator function* of an event, e.g. $\mathbf{1}_{X>5}$ equals 1 if $X > 5$, or equals 0 if $X \le 5$. It follows that $\mathbf{E}(\mathbf{1}_A) = \mathbf{P}(A)$.

iff means "if and only if", i.e. the statements before and after are always either both true or both false.

$\exists$ means "there exists", and $\forall$ means "for all".

$\emptyset$ is the *empty set*, i.e. the set containing no elements.

inf is the *infimum* of a set (a generalisation of *minimum*), with the convention that $\inf(\emptyset) = \infty$.

sup is the *supremum* of a set (a generalisation of *maximum*), with the convention that $\sup(\emptyset) = -\infty$.

A.2. Basic Probability

The reader is assumed to already know undergraduate-level probability theory well. Here we just mention a few very basic facts.

If S is a non-empty finite set, the *(discrete) uniform distribution* on S gives equal probability to each point in S. Hence, if $X \sim \text{Uniform}(S)$, then

(A.2.1) $$\mathbf{P}(X = s) \;=\; \frac{1}{|S|} \quad \text{for each } s \in S\,,$$

where $|S|$ is the number of elements in S.

The *binomial distribution* $\text{Binomial}(n, p)$ represents the probabilities for the number of successes from n independent trials each with probability p of succeeding. Hence, if $X \sim \text{Binomial}(n,p)$, then for $k = 0, 1, \ldots, n$,

$$\textbf{(A.2.2)} \qquad \mathbf{P}(X = k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \;=\; \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}\,.$$

The *geometric distribution* Geometric(p) represents the probabilities for the number of trials before the first success from independent trials each with probability p of succeeding. Hence, if $X \sim \text{Geometric}(p)$, then

$$\textbf{(A.2.3)} \qquad \mathbf{P}(X = k) \;=\; p^k\,(1-p)\,, \quad \text{for } k = 0, 1, 2, \ldots\,,$$

and it is computed that $\mathbf{E}(X) = \frac{1}{p} - 1$. Alternatively, the geometric distribution is sometimes defined as one more than this (i.e., counting the first success in addition to all the failures), in which case

$$\textbf{(A.2.4)} \qquad \mathbf{P}(X = k) \;=\; p^{k-1}\,(1-p)\,, \quad \text{for } k = 1, 2, 3, \ldots\,,$$

and $\mathbf{E}(X) = \frac{1}{p}$.

If X is *discrete* (i.e. takes on only a finite or countable number of different values), then its *expected value* is $\mathbf{E}(X) = \sum_\ell \ell\, \mathbf{P}(X = \ell)$.

So, if X is non-negative-integer-valued, then $\mathbf{E}(X) = \sum_{\ell=1}^\infty \ell\, \mathbf{P}(X = \ell)$.

If X is *absolutely continuous* with *density function* f, then for any $a < b$, $\mathbf{P}(a < X < b) = \int_a^b f(x)\,dx$, and $\mathbf{E}(X) = \int_{-\infty}^\infty x\, f(x)\,dx$.

If $\mathbf{E}(X) = m$, then the *variance* of X is $v = \mathbf{E}[(X-m)^2] = \mathbf{E}(X^2) - m^2$.

The *(continuous) uniform distribution* on an interval $[L, R]$ (where $L < R$) gives probability $(b-a)/(R-L)$ to $[a, b]$ whenever $L \le a \le b \le R$, e.g.

$$\textbf{(A.2.5)} \qquad \text{If } X \sim \text{Uniform}[0,1] \quad \text{then } \mathbf{P}(\frac{1}{2} \le X \le \frac{2}{3}) = \frac{1}{2} - \frac{2}{3} = \frac{1}{6}\,, \text{etc}\,.$$

The *normal distribution* $N(m, v)$ with mean m and variance v has density $f(x) = \frac{1}{\sqrt{2\pi v}}\, e^{-(x-m)^2/2v}$. In particular, the *standard normal distribution* $N(0,1)$ has density $\phi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$.

One trick for computing expectations is:

(A.2.6) Proposition. If Z is non-negative-integer-valued, then

$$\mathbf{E}(Z) \;=\; \sum_{k=1}^\infty \mathbf{P}(Z \ge k)\,.$$

Proof. This follows since

$$\begin{aligned}
\sum_{k=1}^\infty \mathbf{P}(Z \ge k) \;&=\; \sum_{k=1}^\infty [\mathbf{P}(Z = k) + \mathbf{P}(Z = k+1) + \mathbf{P}(Z = k = 2) + \ldots] \\
&= [\mathbf{P}(Z = 1) + \mathbf{P}(Z = 2) + \mathbf{P}(Z = 3) + \mathbf{P}(Z = 4) + \ldots] \\
&\quad + [\mathbf{P}(Z = 2) + \mathbf{P}(Z = 3) + \mathbf{P}(Z = 4) + \ldots] \\
&\quad + [\mathbf{P}(Z = 3) + \mathbf{P}(Z = 4) + \mathbf{P}(Z = 5) + \ldots] \\
&= \mathbf{P}(Z = 1) + 2\,\mathbf{P}(Z = 2) + 3\,\mathbf{P}(Z = 3) + \ldots \\
&= \sum_{\ell=1}^\infty \ell\, \mathbf{P}(Z = \ell) \;=\; \mathbf{E}(Z)\,.
\end{aligned}$$

∎

Probabilities satisfy *additivity* over countable disjoint subsets.

That is, if A and B are *disjoint* (i.e. $A \cap B = \emptyset$), then $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.

And more generally, if $A_1, A_2, \ldots$ are a disjoint sequence, then

(A.2.7) $$\mathbf{P}\left(\bigcup_n A_n\right) \;=\; \sum_n \mathbf{P}(A_n)\,.$$

For example, if X and Y are discrete random variables, then the collection $\{X = i,\ Y = j\}$ are all disjoint, and $\bigcup_j \{X = i,\ Y = j\} = \{X = i\}$, so

(A.2.8) $$\mathbf{P}(X = i) \;=\; \sum_j \mathbf{P}(X = i,\ Y = j)\,,$$

which is one version of the *Law of Total Probability*.

It also follows that probabilities are *monotone*, i.e.:

(A.2.9) $$\text{If } A \subseteq B, \text{ then } \mathbf{P}(A) \leq \mathbf{P}(B)\,.$$

A.3. Infinite Series and Limits

An infinite sum like $\sum_{n=1}^\infty x_n$ is a shorthand way of writing $\lim_{N\to\infty} \sum_{n=1}^N x_n$.

If the x_n are non-negative, then $\sum_{n=1}^\infty x_n$ either (a) converges to some finite value, in which case we write $\sum_{n=1}^\infty x_n < \infty$, or (b) diverges to infinity, in which case we write $\sum_{n=1}^\infty x_n = \infty$.

For example, for $c \in \mathbf{R}$ and $|r| < 1$, the *geometric series* satisfies:

(A.3.1) $$\sum_{n=0}^{\infty} c\, r^n \;=\; c + cr + cr^2 + cr^3 + \ldots \;=\; \frac{c}{1-r}\,.$$

Also, by showing that $\int_1^\infty t^{-a} dt < \sum_{n=2}^\infty n^{-a} < \int_2^\infty t^{-a} dt$, it follows that:

(A.3.2) $$\sum_{n=1}^{\infty} (1/n^a) \;=\; \infty \qquad \underline{\text{iff}} \qquad a \leq 1\,.$$

If $\sum_{n=1}^\infty x_n$ equals some finite number a, then $\lim_{n\to\infty} x_n \;=\; \lim_{n\to\infty} \Big[(\sum_{i=1}^n x_i) - (\sum_{i=1}^{n-1} x_i)\Big] \;=\; \lim_{n\to\infty} (\sum_{i=1}^n x_i) - \lim_{n\to\infty} (\sum_{i=1}^{n-1} x_i) \;=\; a - a \;=\; 0$, hence:

(A.3.3) If the x_n are non-negative, and $\sum_{n=1}^\infty x_n < \infty$, then $\lim_{n\to\infty} x_n \;=\; 0$.

The converse to (A.3.3) is false. For example, if $x_n = 1/n$, then we still have $\lim_{n\to\infty} x_n = 0$, but in this case $\sum_{n=1}^\infty x_n = \infty$ by (A.3.2). That is:

(A.3.4) It is possible that $\sum_{n=1}^\infty x_n = \infty$ even if $\lim_{n\to\infty} x_n = 0$.

(A.3.5) Problem. Prove that $\sum_{n=2}^\infty (1 \,/\, [n \log(n)]) = \infty$.
[Hint: First show that $\sum_{n=2}^\infty (1 \,/\, [n \log(n)]) \geq \int_2^\infty (dx \,/\, [x \, \log(x)])$.]

(A.3.6) Problem. Prove that $\sum_{n=3}^\infty (1 \,/\, [n \log(n) \log\log(n)]) \;=\; \infty$, but $\sum_{n=3}^\infty \Big(1 \,/\, [n \log(n) [\log\log(n)]^2]\Big) < \infty$.

Next, we consider the *Cesàro sum*, i.e. the limit of partial averages $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n x_i$, and its relationship to $\lim_{n\to\infty} x_n$.

For example, suppose $x_n = 1$ for n odd, and $x_n = 0$ for n even. Then $\lim_{n\to\infty} x_n$ does not exist. However, $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n x_i$, does exist (and equals 1/2).

On the other hand, the reverse implication is always true, i.e. if the limit exists then the Cesàro sum also exists and equals the same value. That is:

(A.3.7) $$\text{If } \lim_{n\to\infty} x_n = r, \quad \text{then } \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n x_i = r\,.$$

Finally, one simple but helpful fact about limits is the *Sandwich Theorem* (or, *Squeeze Theorem*):

(A.3.8) If $a_n \le b_n \le c_n$, and $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = L$, then $\lim_{n\to\infty} b_n = L$.

A.4. Bounded, Finite, and Infinite

X is a *bounded random variable* if there is $M < \infty$ with $\mathbf{P}(|X| \le M) = 1$, i.e. if it is always in some interval $[-M, M]$ for some finite number M.

X is a *finite random variable* if $\mathbf{P}(|X| < \infty) = 1$, i.e. if $\mathbf{P}(|X| = \infty) = 0$, i.e. if it always takes on finite values.

A random variable X has *finite expectation* if $\mathbf{E}|X| < \infty$; this is also sometimes called being *integrable*.

If a random variable is bounded then it must have finite expectation; indeed in that case $\mathbf{E}|X| \le M$.

But the converse is false, e.g. if $\mathbf{P}(X = 2^{k/2}) = 2^{-k}$ for $k = 1, 2, 3, \ldots$, then $\mathbf{P}(|X| > M) > 0$ for any $M < \infty$, but by (A.3.1) we have $\mathbf{E}|X| = \sum_{k=1}^\infty 2^{k/2}(2^{-k}) = \sum_{k=1}^\infty 2^{-k/2} = 2^{-1/2}/(1 - 2^{-1/2}) \doteq 2.414 < \infty$.

If a random variable has finite expectation then it must be finite. Indeed, $\mathbf{E}|X| = \sum_\ell \ell\, \mathbf{P}(|X| = \ell) \ge \infty\, \mathbf{P}(|X| = \infty)$, which shows:

(A.4.1) If $\mathbf{P}(|X| = \infty) > 0$, then $\mathbf{E}|X| = \infty$.

Hence, also:

(A.4.2) If $\mathbf{E}|X| < \infty$, then $\mathbf{P}(|X| = \infty) = 0$, i.e. $\mathbf{P}(|X| < \infty) = 1$.

But the converse is false, e.g. if $\mathbf{P}(X = 2^k) = 2^{-k}$ for $k = 1, 2, 3, \ldots$, then $\mathbf{P}(|X| < \infty) = 1$ so X is finite, but $\mathbf{E}|X| = \sum_{k=1}^\infty 2^k(2^{-k}) = \sum_{k=1}^\infty (1) = \infty$.

A.5. Conditioning

If $\mathbf{P}(B) > 0$, then the *conditional probability* of A given B is

(A.5.1) $$\mathbf{P}(A \,|\, B) \;=\; \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}\,,$$

from which it follows that

(A.5.2) $$\mathbf{P}(A \cap B) \;=\; \mathbf{P}(B)\, \mathbf{P}(A \,|\, B)\,.$$

So, if Y is a discrete random variable, and $\mathbf{P}(Y = y) > 0$, then

$$\mathbf{P}(A \,|\, Y = y) \;=\; \frac{\mathbf{P}(A,\; Y = y)}{\mathbf{P}(Y = y)}\,.$$

Hence, if X is also discrete, with finite expectation, then

$$\mathbf{E}(X \,|\, Y = y) \;=\; \sum_x x\,\mathbf{P}(X = x \,|\, Y = y) \;=\; \sum_x x\,\frac{\mathbf{P}(X = x,\; Y = y)}{\mathbf{P}(Y = y)}\,.$$

(A.5.3) Law of Total Expectation. If X and Y are discrete random variables, then

$$\mathbf{E}(X) \;=\; \sum_y \mathbf{P}(Y = y)\;\mathbf{E}(X \,|\, Y = y)\,,$$

i.e. we can compute $\mathbf{E}(X)$ by averaging conditional expectations.

Proof. We compute that

$$\begin{aligned}
&\sum_y \mathbf{P}(Y = y)\,\mathbf{E}(X \,|\, Y = y) \\
&= \sum_y \mathbf{P}(Y = y)\,\sum_x x\,\mathbf{P}(X = x \,|\, Y = y) \\
&= \sum_y \mathbf{P}(Y = y)\,\sum_x x\,\frac{\mathbf{P}(X = x,\; Y = y)}{\mathbf{P}(Y = y)} \\
&= \sum_x x\,\sum_y \mathbf{P}(X = x,\; Y = y) \\
&= \sum_x x\,\mathbf{P}(X = x) \;=\; \mathbf{E}(X)
\end{aligned}$$

where the last line uses the Law of Total Probability (A.2.8). ∎

In particular, if $X = \mathbf{1}_A$ is an indicator function, then (A.5.3) becomes:

(A.5.4) $$\mathbf{P}(A) \;=\; \sum_y \mathbf{P}(Y = y)\,\mathbf{P}(A \,|\, Y = y)\,.$$

We sometimes consider $\mathbf{E}(X \,|\, Y)$ to be another random variable, the *conditional expectation* of X given Y, whose value when $Y = y$ is equal to $\mathbf{E}(X \,|\, Y = y)$ as above. We then have:

(A.5.5) Double-expectation formula. $\mathbf{E}\big[\mathbf{E}(X \,|\, Y)\big] \;=\; \mathbf{E}(X)$, i.e. the random variable $\mathbf{E}(X \,|\, Y)$ equals X on average.

Proof. Since $\mathbf{E}(X \,|\, Y)$ is equal to $\mathbf{E}(X \,|\, Y = y)$ with probability $Y = y$, we compute that

$$\mathbf{E}\big[\mathbf{E}(X \,|\, Y)\big] \;=\; \sum_y \mathbf{P}(Y = y)\;\mathbf{E}(X \,|\, Y = y)\,,$$

whence the result follows from (A.5.3). ∎

Another useful fact is:

(A.5.6) Conditional Factoring. If $h(Y)$ is some function of Y, then

$$\mathbf{E}[h(Y)\,X \,|\, Y] \;=\; h(Y)\,\mathbf{E}[X \,|\, Y]\,,$$

i.e. when conditioning on Y, we can treat any function of Y as a *constant* and factor it out.

Proof. This follows since

$$\begin{aligned}\mathbf{E}[h(Y)\,X \,|\, Y=y] \;&=\; \sum_r r\,\mathbf{P}[h(Y)\,X = r \,|\, Y=y] \\ &=\; \sum_x [h(y)\,x]\,\mathbf{P}[X=x \,|\, Y=y] \\ &=\; h(y)\,\sum_x x\,\mathbf{P}[X=x \,|\, Y=y] \\ &=\; h(y)\,\mathbf{E}[X|\,Y=y]\,. \end{aligned}$$

∎

In particular, setting $X=1$, we obtain that

(A.5.7) $$\mathbf{E}[h(Y)\,|\,Y] \;=\; h(Y)\,,$$

i.e. conditioning on Y has no effect on a function of Y.

If X and Y are general (not just discrete) random variables, then a formal definition of $\mathbf{E}(X\,|\,Y)$ is more technical, involving σ-algebras, and we do not describe it here (see e.g. Section 13 of Rosenthal, 2006, or many other measure-theoretic probability books). Intuitively, $\mathbf{E}(X\,|\,Y)$ is still the average value of X conditional on a particular value of Y. And, the Double-expectation formula (A.5.5) and conditional factoring (A.5.6) still hold.

A.6. Convergence of Random Variables

There are various senses in which a sequence $X_1, X_2, X_3, \ldots$ of random variables could converge to another random variable X.

Convergence in distribution means that for any $a<b$, $\lim_{n\to\infty}\mathbf{P}(a < X_n < b) = \mathbf{P}(a<X<b)$.

Weak convergence means that for any $\epsilon>0$, $\lim_{n\to\infty}\mathbf{P}(|X_n - X| \ge \epsilon) = 0$.

Strong convergence (or, convergence *with probability 1* or *w.p. 1*) means $\mathbf{P}(\lim_{n\to\infty} X_n = X) = 1$, i.e. the random sequence of values always converges.

The limiting random variable X is often just a constant, e.g. $X=0$.

Strong convergence implies weak convergence, and weak convergence implies convergence in distribution; for proofs see e.g. Propositions 5.2.3 and 10.2.1 of Rosenthal (2006) or many other advanced probability books.

However, the converse is false. For example, suppose the $\{X_n\}$ are mostly equal to 0, except that one of $X_1,\ldots,X_9$ (chosen uniformly at random) equals 1, and one of $X_{10},\ldots,X_{99}$ equals 1, and one of $X_{100},\ldots,X_{999}$ equals 1, etc. This sequence has an infinite number of 1's, so $\{X_n\}$ does not converge strongly to 0. But $\mathbf{P}(X_n\neq 0)\to 0$, so $\{X_n\}$ does converge weakly to 0.

Convergence to infinity requires slightly modifying the weak convergence definition. A sequence $\{X_n\}$ of random variables *converges weakly to positive infinity* if for all finite K,

(A.6.1) $$\lim_{n\to\infty}\mathbf{P}(X_n<K) \;=\; 0\,.$$

(For negative infinity, replace $X_n<K$ by $X_n>K$.)

The *Law of Large Numbers* (LLN) says that if the sequence $\{X_n\}$ is *i.i.d.* (i.e., both *independent* and *identically distributed*), with common mean m, then the sequence $\frac{1}{n}\sum_{i=1}^n X_i$ converges to m (both weakly and strongly), i.e.

(A.6.2) $$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_i \;=\; m \quad \textit{w.p. } 1\,.$$

The *Central Limit Theorem* (CLT) says that if the sequence $\{X_n\}$ is *i.i.d.*, with common mean m and common variance v, then the sequence $\frac{1}{\sqrt{nv}}\sum_{i=1}^{n}(X_i - m)$ converges in distribution to the standard normal distribution $N(0,1)$, so for any $a < b$,

$$\text{(A.6.3)} \qquad \lim_{n\to\infty} \mathbf{P}\Big(a < \frac{1}{\sqrt{nv}}\sum_{i=1}^{n}(X_i - m) < b\Big) \;=\; \int_a^b \phi(x)\,dx\,,$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}$ is again the $N(0,1)$ density.

A.7. Continuity of Probabilities

Suppose $A_1, A_2, \ldots$ are a sequence of events which "increase" to an event A, meaning that $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$, and $\bigcup_n A_n = A$.

Then we can write $\{A_n\} \nearrow A$, and think of A as a sort of "limit" of the events $\{A_n\}$, i.e. $\lim_{n\to\infty} A_n = A$.

But here $A = \bigcup_n (A_n \setminus A_{n-1})$ (taking $A_0 = \emptyset$), which is a disjoint union, so $\mathbf{P}(A) = \sum_n \big(\mathbf{P}(A_n) - \mathbf{P}(A_{n-1})\big) = \lim_{n\to\infty}\big(\mathbf{P}(A_n) - \mathbf{P}(A_0)\big) = \lim_{n\to\infty}\mathbf{P}(A_n)$. Hence, probabilities respect these limits:

$$\text{(A.7.1)} \qquad \text{If } \{A_n\} \nearrow A, \text{ then } \lim_{n\to\infty} \mathbf{P}(A_n) \;=\; \mathbf{P}(A)\,.$$

Similarly, if $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$, and $\bigcap_n A_n = A$, then we can write $\{A_n\} \searrow A$. Taking complements, $\{A_n^C\} \nearrow A^C$, so by (A.7.1) we have $\lim_{n\to\infty} \mathbf{P}(A_n^C) = \mathbf{P}(A^C)$, so $\lim_{n\to\infty}\big(1 - \mathbf{P}(A_n^C)\big) = 1 - \mathbf{P}(A^C)$, and hence:

$$\text{(A.7.2)} \qquad \text{If } \{A_n\} \searrow A, \text{ then } \lim_{n\to\infty} \mathbf{P}(A_n) \;=\; \mathbf{P}(A)\,.$$

These facts are called *continuity of probabilities*; see e.g. Section 3.3 of Rosenthal (2006), or many other advanced probability books.

If X is a random variable, and A_n is the event $\{X \geq n\}$, and A is the event $\{X = \infty\}$, then $\{A_n\} \searrow A$, so $\lim_{n\to\infty}\mathbf{P}(A_n) = \mathbf{P}(A)$, i.e.

$$\text{(A.7.3)} \qquad \lim_{n\to\infty} \mathbf{P}(X \geq n) \;=\; \mathbf{P}(X = \infty)\,.$$

A.8. Exchanging Sums and Expectations

For a *finite* collection of random variables, the expectation of a sum always equals the sum of the expectations, e.g. $\mathbf{E}(X+Y) = \mathbf{E}(X)+\mathbf{E}(Y)$, and more generally

$$\text{(A.8.1)} \qquad \mathbf{E}\Big(\sum_{i=1}^{M} X_i\Big) \;=\; \sum_{i=1}^{M} \mathbf{E}(X_i)\,.$$

That is, expectation satisfies finite *linearity*.

But for an infinite collection, this might not be true. For example, suppose Z is a random variable with $\mathbf{P}(Z = n) = 2^{-n}$ for all $n \in \mathbf{N}$. Then let

$$Y_n \;=\; \begin{cases} 2^n, & Z = n \\ -2^{n+1}, & Z = n+1 \\ 0, & \text{otherwise} \end{cases}$$

i.e. $Y_n = 2^n\,\mathbf{1}_{Z=n} - 2^{n+1}\,\mathbf{1}_{Z=n+1}$ for each $n \in \mathbf{N}$.

Then $\mathbf{E}(Y_n) = 2^n\,\mathbf{P}(Z = n) + (-2^{n+1})\,\mathbf{P}(Z = n+1) = 2^n\,(2^{-n}) - 2^{n+1}(2^{-(n+1)}) = 1 - 1 = 0$. So, also $\sum_{n=1}^{\infty} \mathbf{E}(Y_n) = 0$. However,

$$\sum_{n=1}^{\infty} Y_n = \sum_{n=1}^{\infty} (2^n\,\mathbf{1}_{Z=n} - 2^{n+1}\,\mathbf{1}_{Z=n+1})$$

$$= (2^1\,\mathbf{1}_{Z=1} - 2^2\,\mathbf{1}_{Z=2}) + (2^2\,\mathbf{1}_{Z=2} - 2^3\,\mathbf{1}_{Z=3}) + (2^3\,\mathbf{1}_{Z=3} - 2^4\,\mathbf{1}_{Z=4}) + \ldots$$

$$= 2\,\mathbf{1}_{Z=1}$$

since the rest of the sum all cancels out. Hence, $\mathbf{E}[\sum_{n=1}^{\infty} Y_n] = \mathbf{E}[2\,\mathbf{1}_{Z=1}] = 2\,\mathbf{P}(Z = 1) = 2\,(1/2) = 1$. So, in this case, $\sum_{n=1}^{\infty} \mathbf{E}(Y_n) = 0 \neq 1 = \mathbf{E}[\sum_{n=1}^{\infty} Y_n]$, i.e. we cannot interchange the sum and expected value.

What goes wrong in such examples? Roughly speaking, it is because these expressions really equal $\infty - \infty$. That is, both the positive and the negative parts of these expressions are infinite, so the calculations involve an infinite amount of cancellation, and the final result depends on the order in which the cancelling is done.

However, if the Y_n are *non-negative* random variables, then there is no cancellation, so such problems do not arise:

(A.8.2) Countable Linearity. If $\{Y_n\}$ is a sequence of non-negative random variables, then $\sum_{n=1}^{\infty} \mathbf{E}(Y_n) = \mathbf{E}[\sum_{n=1}^{\infty} Y_n]$, i.e. we can exchange the order of the sum and expected values.

Indeed, this follows from the Monotone Convergence Theorem (below) upon setting $X_n = \sum_{k=1}^{n} Y_k$.

Similarly, upon replacing expected values by sums, it follows that if x_{nk} are non-negative real numbers, then

(A.8.3)
$$\sum_n \sum_k x_{nk} = \sum_k \sum_n x_{nk}\,.$$

A.9. Exchanging Expectations and Limits

Next, we consider exchanging expected values and limits.

If a sequence of random variables $\{X_n\}$ converges to X with probability 1 (cf. Section A.6), this does not necessarily imply that $\lim_{n\to\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$.

For example, if $U \sim \text{Uniform}(0,1)$, and $X_n = n\,\mathbf{1}_{0<U<1/n}$, then $X_n \to 0$ (since for any value of U, we have $X_n = 0$ for all $n > 1/U$), but $\mathbf{E}(X_n) = 1$ for all n. So, in this case, $\mathbf{E}[\lim_{n\to\infty} X_n] = \mathbf{E}[0] = 0$, but $\lim_{n\to\infty} \mathbf{E}[X_n] = \mathbf{E}[1] = 1$, i.e. we cannot interchange the limit and expected value.

However, under some *additional* conditions, this does imply that $\lim_{n\to\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$. Indeed, the following results are well-known (for proofs see e.g. Section 9.1 of Rosenthal, 2006, or many other advanced probability books).

(A.9.1) Bounded Convergence Theorem. If $\lim_{n\to\infty} X_n = X$, and the $\{X_n\}$ are uniformly bounded (i.e., there is $M < \infty$ with $|X_n| \leq M$ for all n), then $\lim_{n\to\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$.

(A.9.2) Monotone Convergence Theorem. If $\lim_{n\to\infty} X_n = X$, and $0 \leq X_1 \leq X_2 \leq X_3 \leq \ldots$, then $\lim_{n\to\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$.

(A.9.3) Dominated Convergence Theorem. If $\lim_{n\to\infty} X_n = X$, and there is some random variable Y with $\mathbf{E}|Y| < \infty$ and $|X_n| \leq Y$ for all n, then $\lim_{n\to\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$.

A.10. Exchanging Limits and Sums

The next result concerns the question of when we can exchange the order of limits and sums, i.e. when we can be sure that $\lim_{n\to\infty} \sum_{k=1}^\infty x_{nk} = \sum_{k=1}^\infty \lim_{n\to\infty} x_{nk}$.

If there are only a *finite* number of k in the sum, then this is always true, since (from first-year calculus) the limit of finite sums always equals the sum of the corresponding limits.

But if there are an *infinite* number of different k, then this might fail. For example, suppose $x_{nn} = 1$ but $x_{nk} = 0$ for $k \neq n$. Then $\lim_{n\to\infty} x_{nk} = 0$ for each fixed k (since $x_{nk} = 0$ whenever $n > k$), so $\sum_{k=1}^\infty \lim_{n\to\infty} x_{nk} = \sum_{k=1}^\infty 0 = 0$. On the other hand, $\sum_k x_{nk} = 1$ for each fixed n, so $\lim_{n\to\infty} \sum_{k=1}^\infty x_{nk} = \lim_{n\to\infty} 1 = 1$. Hence, in this case, $\lim_{n\to\infty} \sum_{k=1}^\infty x_{nk} = 1 \neq 0 = \sum_{k=1}^\infty \lim_{n\to\infty} x_{nk}$.

Still, the following result gives a useful condition under which the limit and sum can be exchanged. It may be regarded as a special case of the *Weierstrasse M-test*, or of the above *Dominated Convergence Theorem* (with expected values replaced by sums); here we call it simply the *M-test*.

(A.10.1) M-test. Let $\{x_{nk}\}_{n,k\in\mathbf{N}}$ be a collection of real numbers. Suppose that $\lim_{n\to\infty} x_{nk}$ exists for each fixed $k \in \mathbf{N}$. Suppose further that $\sum_{k=1}^\infty \sup_n |x_{nk}| < \infty$. Then $\lim_{n\to\infty} \sum_{k=1}^\infty x_{nk} = \sum_{k=1}^\infty \lim_{n\to\infty} x_{nk}$.

Proof. Let $a_k = \lim_{n\to\infty} x_{nk}$. The assumption implies that $\sum_{k=1}^\infty |a_k| < \infty$. Hence, by replacing x_{nk} by $x_{nk} - a_k$, it suffices to assume for simplicity that $a_k = 0$ for all k.

Fix $\epsilon > 0$. Since $\sum_{k=1}^\infty \sup_n x_{nk} < \infty$, we can find $K \in \mathbf{N}$ such that $\sum_{k=K+1}^\infty \sup_n x_{nk} < \epsilon/2$. Since $\lim_{n\to\infty} x_{nk} = 0$, we can find (for $k = 1, 2, \ldots, K$) numbers N_k with $x_{nk} < \epsilon/2K$ for all $n \geq N_k$. Let $N = \max(N_1, \ldots, N_K)$. Then for $n \geq N$, we have $\sum_{k=1}^\infty x_{nk} < K\frac{\epsilon}{2K} + \frac{\epsilon}{2} = \epsilon$. Since this is true for all $\epsilon > 0$, we must actually have $\lim_{n\to\infty} \sum_{k=1}^\infty x_{nk} \leq 0 = \sum_{k=1}^\infty a_k$.

Similarly, $\lim_{n\to\infty} \sum_{k=1}^\infty x_{nk} \geq \sum_{k=1}^\infty a_k$. The result follows. ∎

If $\lim_{n\to\infty} x_{nk} = a_k$ for each fixed $k \in \mathbf{N}$, with $x_{nk} \geq 0$, but if we do *not* know that $\sum_{k=1}^\infty \sup_n x_{nk} < \infty$, then we still have

(A.10.2)
$$\lim_{n\to\infty} \sum_{k=1}^\infty x_{nk} \;\geq\; \sum_{k=1}^\infty a_k\,,$$

assuming this limit exists. Indeed, if not then we could find some finite $K \in \mathbf{N}$ with $\lim_{n\to\infty} \sum_{k=1}^\infty x_{nk} < \sum_{k=1}^K a_k$, contradicting the fact that $\lim_{n\to\infty} \sum_{k=1}^\infty x_{nk} \geq \lim_{n\to\infty} \sum_{k=1}^K x_{nk} = \sum_{k=1}^K a_k$.

A.11. Linear Algebra

Linear algebra is not used very much in these notes, but a few aspects are helpful at times.

We first recall *matrix multiplication*. If A is an $r \times m$ matrix, and B is an $m \times n$ matrix, then the product AB is a $r \times n$ matrix whose ij entry is:

(A.11.1)
$$(AB)_{ij} \;=\; \sum_{k=1}^m A_{ik}\,B_{kj}\,, \quad i = 1, 2, \ldots, r, \quad j = 1, 2, \ldots, n\,.$$

For example,

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix} \begin{pmatrix} 13 & 14 \\ 15 & 16 \\ 17 & 18 \\ 19 & 20 \end{pmatrix} = \begin{pmatrix} 170 & 180 \\ 426 & 452 \\ 682 & 724 \end{pmatrix}$$

since e.g. $170 = 1 \times 13 + 2 \times 15 + 3 \times 17 + 4 \times 19$, and $452 = 5 \times 14 + 6 \times 16 + 7 \times 18 + 8 \times 20$, etc.

These formulas still hold for infinite matrices, where $m = \infty$, in which case the sum (A.11.1) becomes an infinite sum.

Of particular use in these notes is the case $r = 1$ and $m = n$, e.g.

$$\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 48 & 54 & 60 \end{pmatrix}.$$

In that case, if v is a $1 \times m$ vector, and A is an $m \times m$ matrix, then v is called a *left eigenvector* for A with *eigenvalue* λ if

(A.11.2)
$$v\,A = \lambda\,v\,,$$

i.e. if multiplying v by the matrix A is equivalent to just multiplying it by the scalar constant λ. (Note that in probability theory we usually put the vector v to the *left* of the matrix, though in pure linear algebra it is usually put to the right, and $v\,A$ is not the same as $A\,v$.) In particular, if v is a left eigenvector with eigenvalue $\lambda = 1$, then $v\,A = v$ (which turns out to be the definition of a *stationary distribution*).

A.12. Miscellaneous Math

Finally, we present a few additional mathematical facts.

First, the well-known *triangle inequality* says that $|a + b| \le |a| + |b|$, and more generally

(A.12.1)
$$\Big|\sum_i a_i\Big| \le \sum_i |a_i|\,.$$

In terms of expected values, this implies that

(A.12.2)
$$\Big|\mathbf{E}(X)\Big| \le \mathbf{E}|X|\,.$$

Second, when dealing with binomial distributions (A.2.2), it is challenging to work with the factorials in the choose formula. So, the following approximation from Real Analysis is very useful:

(A.12.3) Stirling's approximation. If n is large, then $n! \approx (n/e)^n \sqrt{2\pi n}$. More precisely,

$$\lim_{n\to\infty} \frac{n!}{(n/e)^n \sqrt{2\pi n}} = 1\,.$$

Finally, to deal with aperiodicity, we need the following result from Number Theory about a property of the *greatest common divisor (gcd)*. For a proof, see e.g. Durrett (2011) p. 24, or Rosenthal (2006) p. 92.

(A.12.4) Number Theory Lemma. If a set A of positive integers is non-empty, and satisfies *additivity* (i.e. $m + n \in A$ whenever $m \in A$ and $n \in A$), and $gcd(A) = 1$, then there is some $n_0 \in \mathbf{N}$ such that for all $n \ge n_0$ we have $n \in A$, i.e. the set A includes all of the integers n_0, $n_0 + 1$, $n_0 + 2$, ...

B. Bibliography

The following books provide additional details and alternative perspectives about stochastic processes, and might be useful for further reading:

P. Billingsley (1995), *Probability and measure*, 3rd ed. John Wiley & Sons, New York. (Advanced.)

J.K. Blitzstein and J. Hwang (2019), *Introduction to Probability*, 2nd ed. Taylor & Francis Group. https://t.co/2iwKDNe0ag (Elementary; see especially Chapters 11 and 12.)

R. Douc, E. Moulines, P. Priouret, and P. Soulier (2018), *Markov chains*. Springer, New York. (Advanced.)

L.E. Dubins and L.J. Savage (2014), *How to gamble if you must: inequalities for stochastic processes*. Dover Publications, New York, reprint edition.

R. Durrett (2011), *Essentials of stochastic processes*, 2nd ed. Springer, New York. http://www.math.duke.edu/~rtd/EOSP/

G.R. Grimmett and D.R. Stirzaker (2001), *Probability and random processes*, 3rd ed. Oxford University Press.

O. Häggström (2002), *Finite Markov chains and algorithmic applications*. Cambridge University Press.

G.E. Lawler (2006), *Introduction to stochastic processes*, 2nd ed. Chapman & Hall.

S.P. Meyn and R.L. Tweedie (1993), *Markov chains and stochastic stability*. Springer-Verlag, London. http://probability.ca/MT/ (Advanced.)

S. Resnick (1992), *Adventures in stochastic processes*. Birkhäuser, Boston.

J.S. Rosenthal (2006), *A first look at rigorous probability theory*, 2nd ed. World Scientific Publishing Company, Singapore. (Especially chapters 7, 8, 14, and 15.)

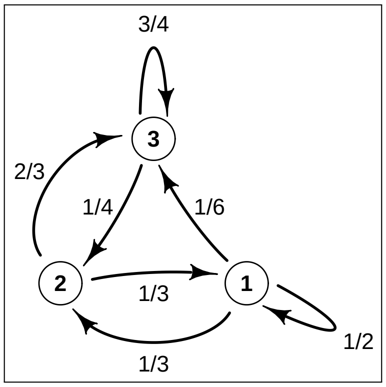
D. Williams (1991), *Probability with Martingales*. Cambridge University Press. (Advanced.)

C. Solutions to Problems Marked [sol]

1.2.3(a): $\mathbf{P}(X_0 = 20,\ X_1 = 19) = \nu_{20} p_{20,19} = (1)(1/3) = 1/3$.

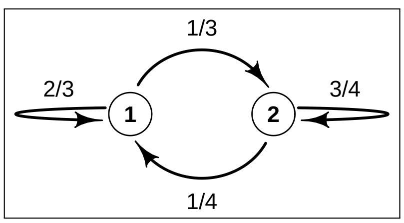
1.2.3(b): $\mathbf{P}(X_0 = 20,\ X_1 = 19,\ X_2 = 20) = \nu_{20} p_{20,19} p_{19,20} = (1)(1/3)(1/3) = 1/9$.

1.3.3(a):



1.4.1(a): $\mathbf{P}(X_1 = 3) = \mu_3^{(1)} = \sum_{i \in S} \nu_i p_{i3} = \nu_1 p_{13} + \nu_2 p_{23} + \nu_3 p_{33} = (1/7)(1/2) + (2/7)(1/3) + (4/7)(1/2) = 19/42$.

1.4.6(a):



1.4.6(b): $p_{12}^{(2)} = \sum_{k \in S} p_{1k} p_{k2} = p_{11} p_{12} + p_{12} p_{22} = (2/3)(1/3) + (1/3)(3/4) = 2/9 + 1/4 = 17/36$.

1.6.16(a): *Yes, it's irreducible: since* $p_{ij} > 0$ *for all* $i, j \in S$, *therefore* $i \to j$ *for all* $i, j \in S$.

1.6.16(b): *Yes. Since the chain is irreducible, and has a stationary distribution, it is recurrent (by the Stationary Recurrence Lemma). Hence,* $f_{ii} = 1$ *for all* $i \in S$. *In particular,* $f_{11} = 1$. *(Or, use the Finite Space Theorem. Or compute it directly.)*

1.6.16(c): *Yes. Since* $f_{11} = 1$ *and* $1 \to 2$ *(by irreducibility), it follows from the f-Lemma (with* $i = 2$ *and* $j = 1$*) that* $f_{21} = 1$. *(Or compute it directly.)*

1.6.16(d): *Yes it is. By the Recurrent State Theorem, since* $f_{11} = 1$, *therefore* $\sum_{n=1}^{\infty} p_{11}^{(n)} = \infty$. *(Or, use the Finite Space Theorem.)*

1.6.16(e): *Yes it is. By the Recurrence Equivalences Theorem, since the chain is irreducible and* $f_{11} = 1$, *therefore* $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ *for all* $i, j \in S$, *and in particular* $\sum_{n=1}^{\infty} p_{21}^{(n)} = \infty$.

1.6.17(a): *No, it isn't: since* $p_{21} = 0$ *and* $p_{31} = 0$, *it is impossible to ever get from state 2 to state 1, so the chain is* <u>not</u> *irreducible.*

1.6.17(b): $f_{11} = p_{11} = 1/2$, *since once we leave state 1 then we can never return to it. But* $f_{22} = 1$, *since from state 2 we will either return to state 2 immediately, or go to state 3 but from there eventually return to* *state 2. Similarly, $f_{33} = 1$, since from state 3 we will either return to state 3 immediately, or go to state 2 but from there eventually return to state 3.*

1.6.17(c): *Since $f_{11} = 1/2 < 1$, therefore state 1 is transient. But since $f_{22} = f_{33} = 1$, therefore states 2 and 3 are recurrent.*

1.6.17(d): *From state 1, the chain might stay at state 1 for some number of steps, but with probability 1 will eventually move to state 2. Then, from state 2, the chain might stay at state 2 for some number of steps, but with probability 1 will eventually move to state 3. So, $f_{13} = 1$.*

1.6.18(a): *Does not exist. We know from the Cases Theorem that for any irreducible transient Markov chain, $\sum_{n=1}^{\infty} p_{k\ell}^{(n)} < \infty$ for all $k, \ell \in S$. In particular, we must have $\lim_{n\to\infty} p_{k\ell}^{(n)} = 0$. So, it is impossible that $p_{k\ell}^{(n)} \geq 1/3$ for all $n \in \mathbf{N}$.*

1.6.18(b): *Does not exist. We know from the Recurrence Equivalences Theorem that for any irreducible Markov chain, if $\sum_{n=1}^{\infty} p_{k\ell}^{(n)} = \infty$ for any one pair $k, \ell \in S$, then the chain is recurrent, and $f_{ij} = 1$ for all $i, j \in S$.*

1.6.18(c): *Exists. For example, let $S = \{1, 2, 3\}$, with $p_{12} = 1$, $p_{22} = 1/3$, $p_{23} = 2/3$, and $p_{33} = 1$ (with $p_{ij} = 0$ otherwise). Then if the chain is started at $k = 1$, then it will initially follow the path $1 \to 2 \to 2 \to 2 \to 2 \to 2 \to 3$ with probability $(1)(1/3)(1/3)(1/3)(1/3)(2/3) > 0$, after which it will remain in the state 3 forever.*

1.6.18(d): *Exists. For example, consider simple random walk with $p = 3/4$, so $S = \mathbf{Z}$ and $p_{i,i+1} = 3/4$ and $p_{i,i-1} = 1/4$ for all $i \in S$ (with $p_{ij} = 0$ otherwise). Let $k = 0$ and $\ell = 5$. Then as shown in class, $f_{05} = 1$, and the chain is irreducible and transient. (Of course, S is infinite here; if S is finite then all irreducible chains are recurrent.)*

1.6.20(b): $p_{32}^{(2)} = \sum_{k\in S} p_{3k}p_{k2} = p_{31}p_{12} + p_{32}p_{22} + p_{33}p_{32} + p_{34}p_{42} = (0)(1/2) + (1/7)(2/3) + (2/7)(1/7) + (4/7)(0) = 2/21 + 2/49 = 20/147$.

1.6.20(c): *The subset $C = \{1, 2\}$ is <u>closed</u> since $p_{ij} = 0$ for $i \in C$ and $j \notin C$. Furthermore, the Markov chain restricted to C is irreducible (since it's possible to go $1 \to 2 \to 1$), and C is finite. Hence, by the Finite Space Theorem, we must have $\sum_{n=1}^{\infty} p_{12}^{(n)} = \infty$.*

1.6.20(d): *Here $f_{32} = \sum_{n=1}^{\infty} \mathbf{P}_3[\text{first hit 2 at time } n] = \sum_{n=1}^{\infty} (2/7)^{n-1}(1/7) = (1/7)/(1-(2/7)) = (1/7)/(5/7) = 1/5$. Or, alternatively, $f_{32} = \mathbf{P}_3[\text{hit 2 when we first leave 3}] = \mathbf{P}_3[\text{hit 2} \mid \text{leave 3}] = (1/7)/((1/7)+(4/7)) = 1/5$. Or, alternatively, by the f-Expansion, $f_{32} = p_{32} + p_{31}\,f_{12} + p_{33}\,f_{32} + p_{34}\,f_{42} = (1/7) + 0 + (2/7)f_{32} + 0$, so $(5/7)f_{32} = 1/7$, so $f_{32} = (1/7)/(5/7) = 1/5$.*

2.3.9: *Yes, it's aperiodic: since $p_{ii} > 0$ for all $i \in S$, therefore every state i has period 1.*

2.4.10(a): $p_{14}^{(2)} = \sum_{k\in S} p_{1k}p_{k4} = p_{11}p_{14} + p_{12}p_{24} + p_{13}p_{34} + p_{14}p_{44} = (0.1)(0.2) + (0.2)(0.1) + (0.5)(0.4) + (0.2)(0.3) = 0.02 + 0.02 + 0.20 + 0.06 = 0.30 = 0.3$.

2.4.10(b): *No. For example, if $i = 1$ and $j = 2$, then $\pi_i p_{ij} = (1/4)(0.2) = 1/20$, while $\pi_j p_{ji} = (1/4)(0.4) = 1/10$, so $\pi_i p_{ij} \neq \pi_j p_{ji}$.*

2.4.10(c): *Yes. The matrix P has each column-sum equal to 1 (as well as, of course, having each row-sum equal to 1), i.e. it is doubly-stochastic. Hence,* *for any $j \in S$, we have that $\sum_{i\in S} \pi_i p_{ij} = \sum_{i\in S}(1/4)p_{ij} = (1/4)\sum_{i\in S} p_{ij} = (1/4)(1) = 1/4 = \pi_j$. So, π is a stationarity distribution.*

2.4.10(d): *Yes. P is irreducible (since $p_{ij} > 0$ for all $i, j \in S$), and aperiodic (since $p_{ii} > 0$ for some, in fact all, $i \in S$), and π is stationary (by part c above), so by the Markov Chain Convergence Theorem, we have $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ for all $i, j \in S$.*

2.4.11(a): *Yes, it's aperiodic: since $p_{ii} > 0$ for all $i \in S$ and it is irreducible, therefore every state i has period 1.*

2.4.11(b): *We need $\pi_1 + \pi_2 = 1$, and also $\pi_1 p_{11} + \pi_2 p_{21} = \pi_1$ and $\pi_1 p_{12} + \pi_2 p_{22} = \pi_2$. The second equation gives $\pi_1(2/3) + \pi_2(1/4) = \pi_1$, i.e. $\pi_2(1/4) = \pi_1(1/3)$, i.e. $\pi_2 = (4/3)\pi_1$. Then since $\pi_1+\pi_2 = 1$, $(4/3)\pi_1+\pi_1 = 1$, so $(7/3)\pi_1 = 1$, so $\pi_1 = 3/7$, and then $\pi_2 = 4/7$. Then all the equations are satisfied.*

2.4.11(c): *Yes it does. The chain is irreducible, aperiodic, and has a stationary distribution $\{\pi_i\}$, so by the Markov Chain Convergence Theorem, $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ for all $i, j \in S$. Setting $i = 1$ and $j = 2$ gives the result.*

2.4.12(a): *Here $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, and $\nu_1 = 1$ (with $\nu_i = 0$ for all other i). Also, for $1 \le i \le 9$, $p_{i,i+1} = 1/4$, and for $1 \le i \le 8$, $p_{i,i+2} = 3/4$, and $p_{10,1} = 1/4$, and $p_{9,1} = p_{10,2} = 3/4$, with $p_{ij} = 0$ otherwise.*

2.4.12(b): *Yes, since it is always possible to move one space clockwise, and hence eventually get to every other state with positive probability.*

2.4.12(c): *From any state i, it is possible to return in 10 seconds by moving one pad clockwise at each jump, or to return in 9 seconds by moving two pads clockwise on the first jump and then one pad clockwise for 8 additional jumps. Since $gcd(10, 9) = 1$, the chain is aperiodic.*

2.4.12(d): *Since the chain is irreducible, and the state space is finite, by the Finite Space Theorem, $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ for all $i, j \in S$, so $\sum_{n=1}^{\infty} p_{15}^{(n)} = \infty$.*

2.4.12(e): *For every state $j \in S$, $\sum_{i\in S} p_{ij} = (1/4) + (3/4) = 1$. Hence, the chain is <u>doubly stochastic</u>. So, since $|S| < \infty$, the <u>uniform</u> distribution on S is a stationary distribution. Hence, we can take $\pi_1 = \pi_2 = \ldots = \pi_{10} = 1/10$.*

2.4.12(f): *Yes, since the chain is irreducible and aperiodic with stationary distribution $\{\pi_i\}$, therefore by the Markov Chain Convergence Theorem, $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j = 1/10$ for all $i, j \in S$, so $\lim_{n\to\infty} p_{15}^{(n)} = \pi_5 = 1/10$ exists.*

2.4.13(a): *Exists. For example, let $S = \{1, 2\}$, with $p_{12} = p_{21} = 1$ (and $p_{11} = p_{22} = 0$). Then the chain is irreducible (since it can get from each i to $3 - i$ in one step, and from i to i in two steps), and periodic with period 2 (since it only returns to each i in even numbers of steps). Furthermore, if $\pi_1 = \pi_2 = 1/2$, then for $i \ne j$, we have $\pi_i p_{ij} = (1/2)(1) = \pi_j p_{ji}$. Hence, the chain is reversible with respect to π, so π is a stationarity distribution.*

2.4.13(b): *Does not exist. By the Equal Periods Lemma, since the chain is irreducible, all states must have the same period.*

2.4.13(c): *Does not exist. If the chain is reversible with respect to π, then π is a stationarity distribution. Then if it is also irreducible, then by the Stationarity Recurrence Lemma, it is recurrent, i.e. it is not transient.*

2.4.13(d): *Exists. For example, let* $S = \{1,2,3\}$, *with* $p_{12} = p_{23} = p_{31} = 1/3$, *and* $p_{21} = p_{32} = p_{13} = 2/3$ *(with* $p_{ij} = 0$ *otherwise). And let* $\pi_1 = \pi_2 = \pi_3 = 1/3$, *so* π *is a probability distribution on* S. *Then* $\pi_1 p_{12} = (1/3)(1/3) \neq (1/3)(2/3) = \pi_2 p_{21}$, *so the chain is not reversible with respect to* π. *On the other hand, for any* $j \in S$, *we have* $\sum_i \pi_i p_{ij} = (1/3)(1/3 + 2/3) = 1/3 = \pi_j$, *so* π *is a stationary distribution.*

2.4.13(e): *Exists. For example, let* $S = \{1,2,3,4,5,6\}$, *with* $p_{12} = p_{23} = p_{34} = p_{56} = p_{61} = 1$ *(with* $p_{ij} = 0$ *otherwise). Let* $i = 1$, *and* $j = 2$, *and* $k = 4$. *Then* $p_{ij} = p_{12} = 1 > 0$, *and* $p_{jk}^{(2)} = p_{23}p_{34} = 1 > 0$, *and* $p_{ki}^{(3)} = p_{45}p_{56}p_{61} = 1 > 0$, *but state* i *has period 6 since it is only possible to return from* i *to* i *in multiples of six steps.*

2.4.15(a): *Possible. For example, let* $S = \{1,2,3\}$, *with* $p_{12} = p_{23} = p_{31} = 1$ *(and* $p_{ij} = 0$ *otherwise). Then the chain is irreducible (since it can get from* $1 \to 2 \to 3 \to 1$*), and periodic with period 3 (since it only returns to each* i *in multiples of three steps). Furthermore the chain is doubly stochastic, so if* $\pi_1 = \pi_2 = \pi_3 = 1/3$, *then* π *is a stationarity distribution.*

2.4.15(b): *Possible. For example, let* $S = \{1,2,3,4,5,6\}$, *with* $p_{12} = p_{21} = 1$, *and with* $p_{34} = p_{45} = p_{56} = p_{63} = 1$. *Then state* $k = 1$ *has period 2 since it only returns in multiples of 2 steps, and state* $\ell = 3$ *has period 4 since it only returns in multiples of 4 steps. (Of course, this chain is not irreducible; for irreducible chains, all states must have the same period.)*

2.4.15(c): *Possible. For example, let* $S = \{1,2,3\}$, *with* $p_{12} = p_{23} = p_{31} = 1/3$, *and* $p_{21} = p_{32} = p_{13} = 1/2$, *and* $p_{11} = p_{22} = p_{33} = 1/6$. *Then* $0 < p_{ij} < 1$ *for all* $i, j \in S$ *(yes, even when* $i = j$*). Next, let* $\pi_1 = \pi_2 = \pi_3 = 1/3$, *so* π *is a probability distribution on* S. *Then* $\pi_1 p_{12} = (1/3)(1/3) \neq (1/3)(1/2) = \pi_2 p_{21}$, *so the chain is not reversible with respect to* π. *On the other hand, for any* $j \in S$, *we have* $\sum_i \pi_i p_{ij} = (1/3)(1/3 + 1/2 + 1/6) = 1/3 = \pi_j$. *(Or, alternatively,* $\sum_i p_{ij} = 1/3 + 1/2 + 1/6 = 1$, *so the chain is doubly stochastic.) Hence,* π *is a stationary distribution.*

2.4.15(d): *Not possible. If the chain is irreducible, and* $\sum_{n=1}^{\infty} p_{k\ell}^{(n)} = \infty$, *then by the Recurrence Equivalences Theorem, we must have* $f_{ij} = 1$ *for all* i *and* j.

2.4.15(e): *Possible. For example, let* $S = \{1,2,3,4,5,6\}$, *with* $p_{12} = p_{15} = 1/2$, *and* $p_{23} = p_{34} = p_{45} = p_{56} = p_{61} = 1$, *with* $p_{ij} = 0$ *o.w. Let* $i = 1$, *and* $j = 2$, *and* $k = 4$. *Then* $p_{ij} = p_{12} = 1/2 > 0$, *and* $p_{jk}^{(2)} = p_{23}p_{34} = 1(1) = 1 > 0$, *and* $p_{ki}^{(3)} = p_{45}p_{56}p_{61} = 1(1)(1) = 1 > 0$, *but state* i *has period 3 (which is odd) since from* i *the chain can return to* i *in three steps (*$1 \to 5 \to 6 \to 1$*) or six steps (*$1 \to 2 \to 3 \to 4 \to 5 \to 6 \to 1$*), and* $gcd(3,6) = 3$.

2.4.15(f): *Not possible. If* $f_{ij} > 0$, *then there is some* $n \in \mathbf{N}$ *with* $p_{ij}^{(n)} > 0$.

2.4.15(g): *Possible. For example, let* $S = \{1,2,3,4\}$, *and* $i = 1$ *and* $j = 2$, *and* $p_{12} = p_{13} = 1/2$, *and* $p_{24} = p_{34} = p_{44} = 1$. *Then* $p_{ij}^{(1)} = 1/2$, *but* $p_{ij}^{(n)} = 0$ *for all* $n \geq 2$, *so* $f_{ij} = 1/2$.

2.4.15(h): *Possible. For example, simple random walk with* $p = 3/4$ *has all states transient, but also* $f_{ij} > 0$ *and* $f_{ji} > 0$ *for all states* i *and* j *by irreducibility.*

2.4.15(i): *Not possible. One way to eventually get from* i *to* k, *is to* first eventually get from i to j, and then eventually get from j to k. This means we must have $f_{ik} \geq f_{ij} f_{jk} = (1/2)(1/3) = 1/6$, so we cannot have $f_{ik} = 1/10$.

2.4.18(a): $p_{11}^{(2)} = \sum_{j \in S} p_{1j} p_{j1} = p_{11} p_{11} + p_{12} p_{21} + p_{13} p_{31} = (0)(0) + (1/2)(1/3) + (1/2)(2/5) = (1/6) + (1/5) = 11/30$.

2.4.18(b): *We need* $\pi P = \pi$, *i.e.* $\sum_{i \in S} \pi_i p_{ij} = \pi_j$ *for all* $j \in S$. *When* $j = 1$ *this gives* $\pi_2(1/3) + \pi_3(2/5) = \pi_1$. *When* $j = 2$ *this gives* $\pi_1(1/2) + \pi_3(3/5) = \pi_2$, *so* $\pi_1(1/2) = \pi_2 - \pi_3(3/5)$, *so* $\pi_1 = 2\pi_2 - (6/5)\pi_3$. *Combining the two equations,* $\pi_2(1/3) + \pi_3(2/5) = 2\pi_2 - (6/5)\pi_3$, *so* $\pi_3(8/5) = \pi_2(2 - (1/3)) = (5/3)\pi_2$, *and* $\pi_3 = (5/8)(5/3)\pi_2 = (25/24)\pi_2$. *Then* $\pi_1 = 2\pi_2 - (6/5)\pi_3 = 2\pi_2 - (6/5)(25/24)\pi_2 = 2\pi_2 - (5/4)\pi_2 = (3/4)\pi_2$. *We need* $\pi_1 + \pi_2 + \pi_3 = 1$, *i.e.* $(3/4)\pi_2 + \pi_2 + (25/24)\pi_2 = 1$, *i.e.* $(67/24)\pi_2 = 1$. *So,* $\pi_2 = 24/67$. *Then* $\pi_1 = (3/4)\pi_2 = (3/4)(24/67) = 18/67$, *and* $\pi_3 = (25/24)\pi_2 = (25/24)(24/67) = 25/67$. *So, the stationary distribution is* $\pi = (18/67, 24/67, 25/67)$.

2.4.18(c): *Yes. Here* π *is stationary by part (b), and the chain is irreducible since e.g. it can go* $1 \to 2 \to 3 \to 2 \to 1$, *and the chain is aperiodic since e.g. it can get from 1 to 1 in two steps* $(1 \to 2 \to 1)$ *or three steps* $(1 \to 2 \to 3 \to 1)$ *and* $gcd(2, 3) = 1$. *Hence, by the Markov chain Convergence Theorem,* $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ *for all* $i, j \in S$.

2.4.19(a): *Here* $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, *and* $\nu_1 = 1$ *(with* $\nu_i = 0$ *for all other* i*). Also, for* $1 \leq i \leq 9$, $p_{i,i+1} = 1/4$, *and for* $1 \leq i \leq 8$, $p_{i,i+2} = 3/4$, *and* $p_{10,1} = 1/4$, *and* $p_{9,1} = p_{10,2} = 3/4$, *with* $p_{ij} = 0$ *otherwise.*

2.4.19(b): *Yes, since it is always possible to move one space clockwise, and hence eventually get to every other state with positive probability.*

2.4.19(c): *From any state* i, *it is possible to return in 10 seconds by moving one pad clockwise at each jump, or to return in 9 seconds by moving two pads clockwise on the first jump and then one pad clockwise for 8 additional jumps. Since* $gcd(10, 9) = 1$, *the chain is aperiodic.*

2.4.19(d): *Since the chain is irreducible, and the state space is finite, by the Finite Space Theorem we have* $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ *for all* $i, j \in S$, *so in particular* $\sum_{n=1}^{\infty} p_{23}^{(n)} = \infty$.

2.4.19(e): *For every state* $j \in S$, $\sum_{i \in S} p_{ij} = (1/4) + (3/4) = 1$. *Hence, the chain is* <u>*doubly stochastic*</u>. *So, since* $|S| < \infty$, *the* <u>*uniform*</u> *distribution on* S *is a stationary distribution. Hence, we can take* $\pi_1 = \pi_2 = \ldots = \pi_{15} = 1/10$.

2.4.19(f): *Yes, since the chain is irreducible and aperiodic with stationary distribution* $\{\pi_i\}$, *therefore* $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j = 1/15$ *for all* $i, j \in S$, *and in particular* $\lim_{n\to\infty} p_{23}^{(n)} = \pi_3 = 1/10$.

2.4.19(g): *Since* $\lim_{n\to\infty} p_{15}^{(n)} = 1/10$, *therefore also* $\lim_{n\to\infty} p_{15}^{(n+1)} = 1/10$, *and hence also* $\lim_{n\to\infty} \frac{1}{2}[p_{15}^{(n)} + p_{15}^{(n+1)}] = 1/10$.

2.5.5(a): *Here* $\pi_1 p_{12} = (1/8)(1) = (1/2)(1/4) = \pi_2 p_{21}$, *and* $\pi_1 p_{13} = (1/8)(0) = (3/8)(0) = \pi_3 p_{31}$, *and* $\pi_3 p_{32} = (3/8)(1) = (1/2)(3/4) = \pi_2 p_{23}$, *so* $\pi_i p_{ij} = \pi_j p_{ji}$ *for all* $i, j \in S$, *so the chain is reversible with respect to* π.

2.5.5(b): *Here* π *is stationary by part (a), and the chain is irreducible since it can go* $1 \to 2 \to 3 \to 2 \to 1$, *but the chain has period 2 since it always moves from odd to even or from even to odd. Hence,* $p_{11}^{(n)} = 0$ *whenever* n

is odd, so we do not have $\lim_{n\to\infty} p_{11}^{(n)} = 1/8$*. But by the Periodic Convergence Theorem (2.5.1), we do still have* $\lim_{n\to\infty} \frac{1}{2}[p_{11}^{(n)} + p_{11}^{(n+1)}] = \pi_1 = 1/8$*, and by Average Probability Convergence (2.5.2) we also have* $\lim_{n\to\infty} \frac{1}{n}\sum_{\ell=1}^{n} p_{11}^{(\ell)} = \pi_1 = 1/8$*. So, in summary, (i) does not hold, but (ii) and (iii) do hold.*

2.6.1: *The Metropolis algorithm says that for* $i \neq j$ *we want* $p_{ij} = (1/2)\min(1, \pi_j/\pi_i)$*. So, we set* $p_{21} = p_{32} = (1/2)(1) = 1/2$*, and* $p_{12} = (1/2)(2/3) = 1/3$ *and* $p_{23} = (1/2)(3/6) = 1/4$*. Then to make* $\sum_j p_{ij} = 1$ *for all* $i \in S$*, we set* $p_{11} = 2/3$*, and* $p_{22} = 1/4$*, and* $p_{33} = 1/2$*. Then by construction,* $\pi_i p_{ij} = \pi_j p_{ji}$ *for all* $i, j \in S$*. Hence, the chain is reversible with respect to* π*. Hence,* π *is a stationary distribution.*

2.7.12: *The graph is connected (since we can get from* $1 \to 2 \to 3 \to 4$ *and back), so the walk is irreducible. Also, the walk is aperiodic since e.g. we can get from 2 to 2 in 2 steps by* $2 \to 3 \to 2$*, or in 3 steps by* $2 \to 3 \to 4 \to 2$*, and* $gcd(2,3) = 1$*. And, if* $\pi_u = d(u)/Z = d(u)/2|E| = d(u)/8$*, then* π *is a stationary distribution by (2.7.7). Hence,* $\lim_{n\to\infty} p_{21}^{(n)} = \pi_1 = d(1)/8 = 1/8$*, since* $d(1) = 1$ *because there is only one edge originating from the vertex 1.*

3.2.7(a): *Intuitively, from 2, the chain has probability 1/3 of decreasing by 1, so to preserve expectation it must have probability 1/6 of increasing by 2, so* $p_{24} = 1/6$*. Similarly, from 3, the chain has probability 1/4 of increasing by 1, so to preserve expectation it must have probability 1/4 of decreasing by 1, so* $p_{32} = 1/4$*. More formally, for a Markov chain to be a martingale, we need that* $\sum_{j\in S} j\, p_{ij} = i$ *for all* $i \in S$*. With* $i = 2$*, we need that* $(1/3)(1) + p_{22}(2) + p_{24}(4) = 2$*, so* $p_{22}(2) + p_{24}(4) = 5/3$*, so* $p_{22} + 2p_{24} = 5/6$*. But also* $p_{21} + p_{22} + p_{24} = 1$*, whence* $p_{22} + p_{24} = 2/3$*. Subtracting this equation from the previous one gives* $p_{24} = (5/6) - (2/3) = 1/6$*, hence* $p_{22} = (2/3) - (1/6) = 1/2$*. Similarly, with* $i = 3$*, we need* $p_{32}(2) + p_{33}(3) + (1/4)(4) = 3$*, so* $2p_{32} + 3p_{33} = 2$*. But also* $p_{32} + p_{33} + p_{34} = 1$*, whence* $p_{32} + p_{33} = 3/4$*. Subtracting twice this equation from the previous one gives* $p_{33} = 2 - 2(3/4) = 1/2$*, hence* $p_{32} = 1/4$*. In summary, if* $p_{24} = 1/6$*,* $p_{22} = 1/2$*,* $p_{32} = 1/4$*, and* $p_{33} = 1/2$*, then we have valid Markov chain transitions which make it a martingale.*

3.2.7(b): *Clearly the chain is bounded up to time* T*, indeed we always have* $|X_n| \le 4$*. Hence, by the Optional Stopping Corollary,* $\mathbf{E}(X_T) = \mathbf{E}(X_0) = 3$*.*

3.2.7(c): *Let* $p = \mathbf{P}(X_T = 1)$*. Then since we must have* $X_T = 1$ *or 4, therefore* $\mathbf{P}(X_T = 4) = 1 - p$*, and* $\mathbf{E}(X_T) = p(1) + (1-p)(4) = 4 - 3p$*. Solving and using part (b), we must have that* $3 = 4 - 3p$*, so* $3p = 4 - 3 = 1$*, whence* $p = 1/3$*.*

3.2.8(a): *For a Markov chain to be a martingale, we need that* $\sum_{j\in S} j\, p_{ij} = i$ *for all* $i \in S$*. With* $i = 2$*, we need that* $(1/4)(1) + p_{22}(2) + p_{24}(4) = 2$*. But we must have* $\sum_{j\in S} p_{ij} = 1$*, i.e.* $(1/4) + p_{22} + p_{24} = 1$*, i.e.* $p_{22} = 3/4 - p_{24}$*, so we must have* $(1/4)(1) + (3/4 - p_{24})(2) + p_{24}(4) = 2$*, or* $p_{24}(4-2) = 2 - 1/4 - 3/2 = 1/4$*, so* $p_{24} = (1/4)/2 = 1/8$*. (Or, more simply, from 2 the chain has probability 1/4 of decreasing by 1, so to preserve expectations it must have probability 1/8 of increasing by 2.) Then* $p_{22} = 3/4 - p_{24} = 3/4 - 1/8 = 5/8$*. Similarly, with* $i = 3$*, we need* $p_{32}(2) + p_{33}(3) + (1/5)(4) = 3$*. For simplicity, since we must have* $p_{32} + p_{33} + (1/5) = 1$*, we can subtract 3 from each term, to get that* $p_{32}(-1) + p_{33}(0) + (1/5)(1) = 0$*, so* $p_{32} = 1/5$*, and then* $p_{33} = 1 - 1/5 - p_{32} = 3/5$*. In summary, if* $p_{24} = 1/8$*,* $p_{22} = 5/8$*,* $p_{32} = 1/5$*,* *and $p_{33} = 3/5$, then we have valid Markov chain transitions which make it a martingale.*

3.2.8(b): *Clearly the chain is bounded up to time T, indeed we always have $|X_n| \le 4$. Hence, by the Optional Stopping Corollary, $\mathbf{E}(X_T) = \mathbf{E}(X_0) = 2$.*

3.2.8(c): *Let $p = \mathbf{P}(X_T = 1)$. Then since we must have $X_T = 1$ or 4, therefore $\mathbf{P}(X_T = 4) = 1 - p$, and $\mathbf{E}(X_T) = p(1) + (1-p)(4) = 4 - 3p$. Solving and using part (b), we must have that $2 = 4 - 3p$, so $3p = 4 - 2 = 2$, whence $p = 2/3$.*

3.2.9(a): *For $i \ge 2$, we need $\sum_j j p_{ij} = i$, so $(i-1)c + (i+2)(1-c) = i$, so $i + 2 - 3c = i$, so $2 = 3c$, so $c = 2/3$.*

3.2.9(b): *We need $\sum_{j \in S} j p_{1j} = 1$. But $\sum_{j \in S} j p_{1j} = \sum_{j=1}^\infty j p_{1j} = p_{11} + \sum_{j=2}^\infty j p_{ij} \ge p_{11} + \sum_{j=2}^\infty 2 p_{ij} = p_{11} + 2(1 - p_{11}) = 2 - p_{11}$. For this to equal 1, we need $p_{11} = 1$.*

3.2.9(c): *Since $\{X_n\}$ is a martingale, $\mathbf{E}(X_n) = \mathbf{E}(X_0) = 5$ for all n, so in particular $\mathbf{E}(X_3) = 5$.*

3.2.9(d): *Clearly the chain is bounded up to time T, indeed we always have $|X_n| \, \mathbf{1}_{n \le T} \le 11$. Hence, by the Optional Stopping Corollary, $\mathbf{E}(X_T) = \mathbf{E}(X_0) = 5$.*

3.2.9(e): *Starting at 5, the chain has positive probability of immediately going $5 \to 4 \to 3 \to 2 \to 1$ and then getting stuck at 1 forever and never returning to 5. Hence, the state 5 is <u>transient</u>. It thus follows from the Recurrence Theorem that $\sum_{n=1}^\infty p_{55}^{(n)} < \infty$, i.e. $\sum_{n=1}^\infty p_{55}^{(n)} \ne \infty$. (Note that this chain is <u>not</u> irreducible, since e.g. $f_{12} = 0$, so the Cases Theorem etc do <u>not</u> apply.)*

3.3.6: *Let $r = (1-p)/p$, so $p = 1/(r+1)$, and $2p - 1 = 2/(r+1) - 1 = (1-r)/(1+r)$. Then the formula for $\mathbf{E}(T)$ when $p \ne 1/2$ can be written as*

$$\frac{c\frac{r^a - 1}{r^c - 1} - a}{(1-r)/(1+r)} \, .$$

As $p \to 1/2$, we have $r \to 1$. So, we compute (using L'Hôpital's Rule twice, and the fact that as $r \to 1$, $1 + r \to 2$) that

$$\lim_{r \to 1} \frac{c\frac{r^a - 1}{r^c - 1} - a}{(1-r)/(1+r)} = 2 \lim_{r \to 1} \frac{c\frac{r^a - 1}{r^c - 1} - a}{1 - r}$$

$$= 2 \lim_{r \to 1} \frac{c(r^a - 1) - a(r^c - 1)}{(r^c - 1)(1 - r)} = 2 \lim_{r \to 1} \frac{car^{a-1} - acr^{c-1}}{cr^{c-1}(1-r) - (r^c - 1)}$$

$$= 2ac \lim_{r \to 1} \frac{(a-1)r^{a-2} - (c-1)r^{c-2}}{c(c-1)r^{c-2}(1-r) - cr^{c-1} - cr^{c-1}}$$

$$= 2ac \frac{(a-1) - (c-1)}{c(c-1)1^{c-2}(1-1) - c1^{c-1} - c1^{c-1}}$$

$$= 2ac \frac{(a-c)}{0 - c - c} = 2ac \frac{(a-c)}{(-2c)} = a(c-a) \, .$$

3.7.4(a): *Here if you buy x stock and y options, then your profit if the stock goes down is $x(-10) + y(-c)$, or if the stock goes up is $x(30) + y(20 - c)$.*

These are equal if $y = -2x$, *in which case they both equal* $-10x + 2xc = 2x(c-5)$. *So, there is no arbitrage* iff $c = 5$, *i.e. the fair price is \$5.*

3.7.4(b): *The stock price is a martingale if the stock price goes down with probability 3/4, or up with probability 1/4. And, under those martingale probabilities, the expected value of the option tomorrow is* $(3/4)(0) + (1/4)(20) = 5 = c$. *So, again, the fair price for the option is \$5. (And, once again the* true *probabilities that stock tomorrow equals \$10 or \$50 are irrelevant.)*

3.7.6: *In this case,* $a = 20$, $d = 10$, $u = 50$, *and* $K = 30$. *So, the fair price* $= c = (20-10)(50-30)/(50-10) = 5$. *Alternatively, the martingale probabilities are* $q_1 = (50-20)/(50-10) = 3/4$ *and* $q_2 = 1 - q_1 = 1/4$, *so the fair price = martingale expected value* $= q_1(0) + q_2(u-K) = (1/4)(50-30) = 5$. *Either way, it gives the same answer as before.*

Index

(References to figures are underlined.)