STA447/2006 (Stochastic Processes) Lecture Notes, Winter 2017

by Jeffrey Rosenthal, University of Toronto

(Last updated: March 30, 2017)

Note: These lecture notes will be posted on the STA447/2006 course web page after the corresponding lecture material, for your convenience. However, they are just rough, point-form notes, with no guarantee of completeness or accuracy. They should in no way be regarded as a substitute for attending and actively learning from the course lectures.

Introduction:

- Course web page, outline, evaluation, etc. (www.probability.ca/sta447)
- Schedule: will take 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? 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).

Markov chains:

- EXAMPLE (Frog Example):
  - 20 lily pads arranged in a circle. (diagram)
  - Frog starts at pad #20.
  - Each minute, she jumps either straight up, or one pad clockwise, or one pad counter-clockwise, each with probability 1/3.
  - (see e.g. www.probability.ca/frogwalk)
- So, $\mathbf{P}$(at pad #1 after 1 step) $= 1/3$.
  - $\mathbf{P}$(at pad #20 after 1 step) $= 1/3$.
  - $\mathbf{P}$(at pad #19 after 1 step) $= 1/3$.
  - $\mathbf{P}$(at pad #2 after 2 steps) $= 1/9$.
  - $\mathbf{P}$(at pad #19 after 2 steps) $= 2/9$.
  - etc.
- What happens in the long run?
  - What is $\mathbf{P}$(frog at pad #14 after 987 steps)?
  - What is $\lim_{k\to\infty} \mathbf{P}$(frog at pad #14 after k steps)?

- Will the frog necessarily eventually return to pad #20?
- Will the frog necessarily eventually visit every pad?

- And what happens if we have:
  - different jump probabilities?
  - different arrangement of the pads?
  - more pads?
  - infinitely many pads?
  - etc.

- 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. (e.g. the 20 lily pads)
  - 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.)

- In the frog example, $S = \{1, 2, 3, \ldots, 20\}$, and

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

  and $\nu_{20} = 1$ (with $\nu_i = 0$ otherwise).

- Let X_n be the Markov chain's state at time n.
  - Then $\mathbf{P}(X_{n+1} = j \,|\, X_n = i) = p_{ij}$, $\forall i, j \in S$, $n = 0, 1, 2, \ldots$. (Doesn't depend on n: time-homogeneous.)
  - Also $\mathbf{P}(X_{n+1} = j \,|\, X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) = p_{i_n j}$. (Markov property.)
  - Also $\mathbf{P}(X_0 = i, X_1 = j, X_2 = k) = \nu_i p_{ij} p_{jk}$, etc.
  - More generally, $\mathbf{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \nu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}$.
  - The random sequence $\{X_n\}_{n=0}^{\infty}$ "is" the Markov chain.

- In the frog example:
  - $\mathbf{P}(X_0 = 20) = 1$, $\mathbf{P}(X_0 = 7) = 0$, etc.
  - $\mathbf{P}(X_1 = 1) = 1/3$, $\mathbf{P}(X_1 = 20) = 1/3$, $\mathbf{P}(X_2 = 2) = 1/9$, $\mathbf{P}(X_2 = 19) = 2/9$, etc.

More Examples of Markov Chains:

- Example: 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.

– Let X_n be net gain (in dollars) after n bets.
– Then $\{X_n\}$ is a Markov chain, with $S = \mathbf{Z}$, $\nu_0 = 1$, and

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

– What happens in the long run? Will you necessarily go broke? etc.
– (see e.g. www.probability.ca/randwalk)

- Example: Bernoulli process. (e.g. counting sunny days)
  – Let $0 < p < 1$. (e.g. $p = 1/2$)
  – Repeatedly flip a "p-coin" (i.e., a coin whose probability of heads is p).
  – 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

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

- Example: simple finite Markov chain.
  – Let $S = \{1, 2, 3\}$, and $X_0 = 3$, and

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

  – What happens in the long run? (diagram)
- Example: Ehrenfest's Urn
  – 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.
  – Then $\{X_n\}$ is Markov chain, with $S = \{0, 1, 2, \ldots, d\}$, and $p_{i,i-1} = i/d$, and $p_{i,i+1} = (d-i)/d$, with $p_{ij} = 0$ otherwise.
  – What happens in the long run? Does X_n become uniformly distributed? Does it stay close to X_0? to $d/2$?
- Example: human Markov chain?
  – Everyone take out a coin (or borrow one).
  – Then pick out two other students, one for "heads" and one for "tails".
  – Each time the frog comes to you, catch it, and flip your coin. Then toss the frog to either your heads or your tails student, depending on the result of the flip.

- What happens in the long run?

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.
- In frog example: $\mathbf{P}(X_0 = 20,\ X_1 = 19,\ X_2 = 20) = \nu_{20} p_{20,19} p_{19,20} = (1)(1/3)(1/3) = 1/9$, etc.
- Now, let $\mu_i^{(n)} = \mathbf{P}(X_n = i)$.
  - Then $\mu_i^{(0)} = \nu_i$.
- What is $\mu_j^{(1)}$ in terms of ν_i and p_{ij}?
  - $\mu_j^{(1)} = \mathbf{P}(X_1 = j) = \sum\limits_{i \in S} \mathbf{P}(X_0 = i,\ X_1 = j) = \sum\limits_{i \in S} \nu_i p_{ij}$.
  - ("Law of Total Probability", "additivity", "partition")
- In matrix form:
  - Write $\mu^{(n)} = (\mu_1^{(n)}, \mu_2^{(n)}, \mu_3^{(n)}, \ldots)$. [row vector]
  - And write $\mathbf{P} = (\,p_{ij}\,) = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & \vdots & \vdots & \ddots \end{pmatrix}$. [matrix]
  - And write $\nu = (\nu_1, \nu_2, \nu_3, \ldots)$. [row vector]
  - Then $\mu^{(1)} = \nu P = \mu^{(0)} P$. [matrix multiplication]
- e.g. if $S = \{1, 2, 3\}$, and $\mu^{(0)} = (1/7, 2/7, 4/7)$, and

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

  then $\mu_2^{(1)} = \mathbf{P}(X_1 = 2) = \mu_1^{(0)} p_{12} + \mu_2^{(0)} p_{22} + \mu_3^{(0)} p_{32} = (1/7)(1/2) + (2/7)(1/3) + (4/7)(1/4) = 13/42$.
- Similarly, $\mu_k^{(2)} = \sum_{i \in S} \sum_{j \in S} \nu_i p_{ij} p_{jk}$, etc.
  - Matrix form: $\mu^{(2)} = \mu^{(0)} P P = \mu^{(0)} P^2$.
  - By induction: $\mu^{(n)} = \mu^{(0)} P^n$, for $n = 1, 2, 3, \ldots$.
  - Convention: $P^0 = I$ (identity). Then true for $n = 0$ too.
  - e.g. in frog example, $\mu_{19}^{(2)} = \nu_{20} p_{20,19} p_{19,19} + \nu_{20} p_{20,20} p_{20,19} + 0 = (1)(1/3)(1/3) + (1)(1/3)(1/3) + 0 = 2/9$.

- n-step transitions: $p_{ij}^{(n)} = \mathbf{P}(X_{m+n} = j \mid X_m = i)$.
  - (Again, doesn't depend on m: time-homogeneous.)
  - $p_{ij}^{(1)} = p_{ij}$. (of course)
  - What about $p_{ij}^{(2)}$?
  - Well, $p_{ij}^{(2)} = \mathbf{P}(X_2 = j \mid X_0 = i) = \sum_{k \in S} \mathbf{P}(X_2 = j,\ X_1 = k \mid X_0 = i) = \sum_{k \in S} p_{ik} p_{kj}$.
  - Matrix form: $P^{(2)} = \left(p_{ij}^{(2)}\right) = P\,P = P^2$.
  - By induction: $P^{(n)} = P^n$, i.e. to compute probabilities of n-step jumps, you can take n^{th} powers of the transition matrix P.
  - Convention: $P^{(0)} = I$ = identity matrix, i.e. $p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & \text{otherwise} \end{cases}$
- Observation: $p_{ij}^{(m+n)} = \mathbf{P}(X_{m+n} = j \mid X_0 = i) = \sum_{k \in S} \mathbf{P}(X_{m+n} = j,\ X_m = k \mid X_0 = i) = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}$.
  - Matrix form: $P^{(m+n)} = P^{(m)} P^{(n)}$.
  - (Of course, since $P^{(m+n)} = P^{m+n} = P^m P^n$.)
  - "Chapman-Kolmogorov equations".
  - Follows that e.g. $p_{ij}^{(m+n)} \geq p_{ik}^{(m)} p_{kj}^{(n)}$ for any state k.

Classification of States:

- 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 the chain hits i. (Random variable; could be infinite.)
- Let $f_{ij} \ = \ \mathbf{P}_i(X_n = j \text{ for } \underline{\text{some}}\ n \geq 1)$.
- Defn: a state i of a Markov chain is <u>recurrent</u> (or, <u>persistent</u>) if $\mathbf{P}_i(X_n = i$ for some $n \geq 1) = 1$, i.e. if $f_{ii} = 1$. Otherwise, i is <u>transient</u>. (Previous examples? Frog? s.r.w.?)
- RECURRENCE THEOREM:
  - i recurrent <u>iff</u> $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ <u>iff</u> $\mathbf{P}_i(N(i) = \infty) = 1$.
  - And, i transient <u>iff</u> $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$ <u>iff</u> $\mathbf{P}_i(N(i) = \infty) = 0$.
- To prove this, let
- Also, $\mathbf{P}_i(N(i) \geq 1) = f_{ii}$, and $\mathbf{P}_i(N(i) \geq 2) = (f_{ii})^2$, etc.
  - In general, for $k = 0, 1, 2, \ldots$, $\mathbf{P}_i(N(i) \geq k) = (f_{ii})^k$.
- Also, recall that if Z is any non-negative-integer-valued random variable, then

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

- PROOF OF RECURRENCE THEOREM: First, by continuity of probabilities,

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

Second, using countable linearity (okay since non-negative),

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

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

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

END OF WEEK #1

- EXAMPLE: $S=\{1,2,3,4\}$, and $P=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}$.
  - Here $f_{11}=1$, $f_{22}=1/4$, $f_{33}=1$, and $f_{44}=1$.
  - So, states 1, 3, and 4 are recurrent, but state 2 is transient.
  - Also, $f_{12}=0=f_{13}=f_{14}=f_{32}=f_{31}$.
  - And, $f_{34}=1=f_{43}$.
  - And, $f_{21}=1/3$ [since e.g. $f_{21}=p_{21}+p_{22}f_{21}+p_{23}f_{31}+p_{24}f_{41}=(1/4)+(1/4)f_{21}+0+0$, so $f_{21}=(1/4)/(3/4)=1/3$; alternatively, in this special case only, $f_{21}=\mathbf{P}_2(X_1=1\,|\,X_1\neq 2)=(1/4)/[(1/4)+(1/2)]=1/3$].
  - And, $f_{23}=2/3$, and $f_{24}=2/3$, etc.
  - (Harder example to come on homework!)
- F-EXPANSION: $f_{ij}=p_{ij}+\sum_{k\neq j} p_{ik}f_{kj}$
- What about e.g. Frog Example? Harder. Later!
- EXAMPLE: Simple random walk (s.r.w.). ($S=\mathbf{Z}$, and $p_{i,i+1}=p$, and $p_{i,i-1}=1-p$.)
  - Is the state 0 recurrent?
  - Well, if n odd, then $p_{00}^{(n)}=0$.
  - If n even, then $p_{00}^{(n)}=\mathbf{P}(n/2$ heads and $n/2$ tails on first n tosses$)$ $=\binom{n}{n/2}p^{n/2}(1-p)^{n/2}=\frac{n!}{[(n/2)!]^2}p^{n/2}(1-p)^{n/2}$. [binomial distribution]
  - Stirling's approximation: if n large, then $n!\approx(n/e)^n\sqrt{2\pi n}$.
  - So, for n large and even,

$$p_{00}^{(n)} \;\approx\; \frac{(n/e)^n\sqrt{2\pi n}}{[(n/2e)^{n/2}\sqrt{2\pi n/2}]^2}\; p^{n/2}(1-p)^{n/2}$$

$$= [4p(1-p)]^{n/2}\sqrt{2/\pi n}\,.$$

- Now, if $p = 1/2$, then $4p(1-p) = 1$, so $\sum_{n=1}^{\infty} p_{00}^{(n)} \approx \sum_{n=2,4,6,\ldots} \sqrt{2/\pi n} = \infty$, so state 0 is <u>recurrent</u>.
- But if $p \neq 1/2$, then $4p(1-p) < 1$, so $\sum_{n=1}^{\infty} p_{00}^{(n)} \approx \sum_{n=2,4,6,\ldots} [4p(1-p)]^{n/2}\sqrt{2/\pi n} < \infty$, so state 0 is <u>transient</u>.
- (Similarly true for all other states besides 0, too.)

Communicating States:

- Say that state i <u>communicates</u> with state j, written $i \to j$, if $f_{ij} > 0$, i.e. if it is <u>possible</u> to get from i to j, i.e. if $\exists m \geq 1$ with $p_{ij}^{(m)} > 0$.
  - Write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$.
- Say a Markov chain is <u>irreducible</u> if $i \to j$ for all $i, j \in S$. (Previous examples?)
- 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 OF SUM LEMMA: Find $m, r \geq 1$ with $p_{ik}^{(m)} > 0$ and $p_{\ell j}^{(r)} > 0$. Note that $p_{ij}^{(m+s+r)} \geq p_{ik}^{(m)} p_{k\ell}^{(s)} p_{\ell j}^{(r)}$. Hence,

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

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

- SUM COROLLARY: if $i \leftrightarrow k$, then i is recurrent <u>iff</u> k is recurrent.
- PROOF: Combine Sum Lemma ($i = j$ and $k = \ell$) with Recurrence Theorem. *Q.E.D.*
- CASES THEOREM: For an <u>irreducible</u> Markov chain, either
  (a) $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ for all $i, j \in S$, so <u>all</u> states are recurrent.
  ("recurrent Markov chain")
  or (b) $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all $i, j \in S$, so <u>all</u> states are transient.
  ("transient Markov chain")
- EXAMPLE: simple random walk. Irreducible!
  - $p = 1/2$: case (a).
  - $p \neq 1/2$: case (b).
- What about Frog Example? Also irreducible, but which case?? Answer given by:
- FINITE SPACE THEOREM: an irreducible Markov chain on a <u>finite</u> 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 OF FINITE SPACE THEOREM:

- Choose any state $i \in S$. Then by exchanging the sums (okay since non-negative), we have
$$\sum_{j \in S} \sum_{n=1}^{\infty} p_{ij}^{(n)} = \sum_{n=1}^{\infty} \sum_{j \in S} p_{ij}^{(n)} = \sum_{n=1}^{\infty} 1 = \infty .$$
- Since S is finite, 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). *Q.E.D.*

- So, in Frog Example, $\mathbf{P}(\exists n \geq 1 \text{ with } X_n = 20 \,|\, X_0 = 20) = 1$.
  - But what about $\mathbf{P}(\exists n \geq 1 \text{ with } X_n = 14 \,|\, X_0 = 20)$??
- To continue, define $T_i = \min\{n \geq 1 : X_n = i\}$. ($T_i = \infty$ if never hit i.)
- HIT LEMMA: If $j \to i$ with $j \neq i$, then $\mathbf{P}_j(T_i < T_j) > 0$.
  - Intuitively obvious(?). But formal proof is:
  - 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 last time this path hits j.
  - 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, $\mathbf{P}_j(T_i < T_j) \geq \mathbf{P}_j(\text{this path}) = p_{x_S x_{S+1}} p_{x_{S+1} x_{S+2}} \ldots p_{x_{m-1} x_m} > 0$, *Q.E.D.*
- F-LEMMA: If $j \to i$ and $f_{jj} = 1$, then $f_{ij} = 1$.
- PROOF OF F-LEMMA:
  - Assume $i \neq j$ (otherwise trivial).
  - Since $j \to i$, $\mathbf{P}_j(T_i < T_j) > 0$ by Hit Lemma.
  - But $1 - f_{jj} = \mathbf{P}_j(T_j = \infty) \geq \mathbf{P}_j(T_i < T_j)\, \mathbf{P}_i(T_j = \infty)$
    $= \mathbf{P}_j(T_i < T_j)\,(1 - f_{ij})$.
  - If $f_{jj} = 1$, then $1 - f_{jj} = 0$, so must have $1 - f_{ij} = 0$, i.e. $f_{ij} = 1$. *Q.E.D.*
- Putting all of the above together, we obtain:
- STRONGER RECURRENCE THEOREM: If chain 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 $f_{ij} = 1$.
- PROOF:
  - (1) $\Rightarrow$ (2): Sum Lemma.
  - (2) $\Rightarrow$ (4): Recurrence Theorem (with $i = j$).
  - (4) $\Rightarrow$ (5): F-Lemma.

- (5) $\Rightarrow$ (3): Immediate.
- (3) $\Rightarrow$ (1): Recurrence Theorem (with $\ell = k$).
- *Q.E.D.*

- Frog Example: $\mathbf{P}(\exists n \geq 1 \text{ with } X_n = 14 \,|\, X_0 = 20) = f_{20,14} = 1$, etc.
- Simple random walk with $p = 1/2$: $\mathbf{P}(\exists n \geq 1 \text{ with } X_n = 1,000,000 \,|\, X_0 = 0) = 1$, etc. (And similarly for any conceivable pattern of values, i.e. the chain's values "fluctuate" arbitrarily.)
- Example: $S = \{1, 2, 3\}$, and $(p_{ij}) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.
  - Then $\sum_{n=1}^{\infty} p_{12}^{(n)} = \sum_{n=1}^{\infty} (1/2) = \infty$.
  - And $f_{22} = 1$. Recurrent!
  - But $f_{11} = 0 < 1$. Transient!
  - Also $f_{12} = 1/2 < 1$.
  - Not irreducible!
- Example: Simple random walk with $p > 1/2$.
  - Irreducible.
  - $f_{00} < 1$. (transient)
  - Claim: $f_{05} = 1$.
  - (Contradiction? No! Could still have $\sum_{n=1}^{\infty} p_{05}^{(n)} < \infty$.)
  - Indeed, let $Z_n = X_n - X_{n-1}$.
  - Then $\mathbf{P}(Z_n = +1) = p$, $\mathbf{P}(Z_n = -1) = 1 - p$, and $\{Z_n\}$ i.i.d.
  - So, by Strong Law of Large Numbers, 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, $\lim_{n\to\infty}(Z_1 + Z_2 + \ldots + Z_n) = +\infty$.
  - i.e., w.p. 1, $X_n - X_0 \to \infty$, so $X_n \to \infty$.
  - Follows that if $i < j$, then $f_{ij} = 1$ (since must pass j when going from i to ∞).
  - In particular, $f_{05} = 1$.
  - (Similarly, if $p < 1/2$ and $i > j$, then $f_{ij} = 1$.)
  - This is why the Stronger Recurrence Theorem does not include the equivalence "There are $k, \ell \in S$ with $f_{k\ell} = 1$".
- CLOSED SUBSET NOTE: if chain not irreducible, 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 can still apply the Stronger Recurrence Theorem etc to the chain restricted to C.

– e.g. in previous example with $P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}$, can take $C = \{3,4\}$, to prove that $f_{34} = 1$, $\sum_{n=1}^{\infty} p_{43}^{(n)} = \infty$, etc.

Stationary Distributions:

- What about a Markov chain's long-run probabilities?
  – Does $\lim_{n\to\infty} \mathbf{P}[X_n = i]$ exist?
  – What does it equal?
- Let π be a probability distribution on S, i.e. $\pi_i \geq 0$ for all $i \in S$, and $\sum_{i\in S} \pi_i = 1$.
- Defn: π is stationary for a Markov chain $P = (\,p_{ij}\,)$ if $\sum_{i\in S} \pi_i p_{ij} = \pi_j$ for all $j \in S$.
  – Matrix notation: $\pi\, P = \pi$.
  – Then by induction, $\pi\, P^n = \pi$ for $n = 0, 1, 2, \ldots$, i.e. $\sum_{i\in S} \pi_i p_{ij}^{(n)} = \pi_j$.
  – Intuition, if chain starts with probabilities $\{\pi_i\}$, then chain will keep the same probabilities one time unit later.
  – That is, 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. $\mu^{(n+1)}$ also equals π.
  – And then, by induction, $\mu^{(m)} = \pi$ for all $m \geq n$. ("stationary")
- Frog Example:
  – Let $\pi_i = \frac{1}{20}$ for all $i \in S$.

END OF WEEK #2

  – Then $\pi_i \geq 0$ and $\sum_i \pi_i = 1$.
  – Also, for all $j \in S$, $\sum_{i\in S} \pi_i p_{ij} = \frac{1}{20}(\frac{1}{3} + \frac{1}{3} + \frac{1}{3}) = \frac{1}{20} = \pi_j$.
  – So, $\{\pi_i\}$ is stationary distribution!
- (More generally, if chain is "doubly stochastic", i.e. $\sum_{i\in S} p_{ij} = 1$ for all $j \in S$, and if $|S| < \infty$, then if $\pi_i = 1/|S|$ for all $i \in S$, then $\{\pi_i\}$ is stationary [check].)
- Ehrenfest's Urn example: ($S = \{0, 1, 2, \ldots, d\}$, $p_{ij} = i/d$ for $j = i - 1$, $p_{ij} = (d-i)/d$ for $j = i + 1$)
  – Does $\pi_i = \frac{1}{d+1}$ for all i?
  – 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 not take $\pi_i = \frac{1}{d+1}$ for all i.
  – So, $\pi_i = ???$
- Defn: 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 $i, j \in S$.

- PROPOSITION: if chain is reversible w.r.t. $\{\pi_i\}$, then $\{\pi_i\}$ is a stationary distribution. (Converse false.)
  - PROOF: for $j \in S$, $\sum_{i\in S} \pi_i p_{ij} = \sum_{i\in S} \pi_j p_{ji} = \pi_j \sum_{i\in S} p_{ji} = \pi_j$. *Q.E.D.*

- Frog Example:
  - $\pi_i = 1/20$
  - 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 0.
  - So, reversible! (easier way to check stationarity)
- Example: $S = \{1,2,3\}$, $p_{12} = p_{23} = p_{31} = 1$, $\pi_1 = \pi_2 = \pi_3 = 1/3$. Then $\{\pi_i\}$ stationary (check!), but chain is not reversible w.r.t. $\{\pi_i\}$.
- Ehrenfest's Urn:
  - New idea: perhaps each ball is equally likely to be in either Urn.
  - That is, let $\pi_i = 2^{-d}\binom{d}{i} = 2^{-d}\frac{d!}{i!(d-i)!}$.
  - Then $\pi_i \ge 0$ and $\sum_i \pi_i = 1$.
  - Stationary? Need to check if $\sum\limits_{i\in S} \pi_i p_{ij} = \pi_j$ for all $j \in S$. Possible but messy. (Need the Pascal's Triangle identity that $\binom{d-1}{j-1} + \binom{d-1}{j} = \binom{d}{j}$.) Better way?
  - Use reversibility!
  - 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 $\pi_i p_{ij} = \pi_i p_{ij}$ [check! or just interchange i and j].
  - Otherwise both sides 0.
  - So, reversible!
  - So, $\{\pi_i\}$ is stationary distribution!
  - Intuitively, π_i is larger when i is close to $d/2$.
  - But does $\mu_i^{(n)} \to \pi_i$? We'll see!

Obstacles to Convergence:

- If chain has stationary distribution $\{\pi_i\}$, does $\lim_{n\to\infty} \mathbf{P}[X_n = i] = \pi_i$?
- Not necessarily!
- Example: $S = \{1, 2\}$, and $\nu_1 = 1$, and $(\, p_{ij}\,) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.
  - If $\pi_1 = \pi_2 = \frac{1}{2}$ (say), then $\{\pi_i\}$ stationary (check!).
  - But $\lim_{n\to\infty} \mathbf{P}[X_n = 1] = \lim_{n\to\infty} 1 = 1 \neq \frac{1}{2} = \pi_1$.
  - Not irreducible! ("reducible")
- Example: $S = \{1, 2\}$, and $\nu_1 = 1$, and $(\, p_{ij}\,) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.
  - Again, if $\pi_1 = \pi_2 = \frac{1}{2}$, then $\{\pi_i\}$ 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!
  - "periodic"
- Defn: the <u>period</u> of a state i is the greatest common divisor of the set $\{n \geq 1; p_{ii}^{(n)} > 0\}$.
  - e.g. if period of i is 2, this means that it is only possible to get from i to i in an <u>even</u> numbers of steps.
  - If period if each state is 1, say chain is "aperiodic".
- Example: $S = \{1, 2, 3\}$, and $p_{12} = p_{23} = p_{31} = 1$.
  - Then period of each state is 3.
- Example: $S = \{1, 2, 3\}$, and $(\, p_{ij}\,) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$.
  - Then period of state 1 is $gcd\{2, 3, \ldots\} = 1$.
  - Aperiodic!
- Observation: if $p_{ii} > 0$, then period of i is 1 (since $gcd\{1, \ldots\} = 1$).
  - (Converse false, as in previous example.)
  - Or, if there is some $n \geq 1$ with $p_{ii}^{(n)} > 0$ and $p_{ii}^{(n+1)} > 0$, then period of i is 1 (since $gcd\{n, n+1, \ldots\} = 1$).
- Frog Example: $p_{ii} > 0$, so chain aperiodic.
- Simple Random Walk: can only return after <u>even</u> number of steps, so period of each state is 2.
- Ehrenfest's Urn: again, can only return after <u>even</u> number of steps, so period of each state is 2.
- 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.
- Find $r, s \in \mathbf{N}$ with $p_{ij}^{(r)} > 0$ and $p_{ji}^{(s)} > 0$.
- Then $p_{ii}^{(r+s)} \geq p_{ij}^{(r)} p_{ji}^{(s)} > 0$, so t_i divides $r + s$.
- Also if $p_{jj}^{(n)} > 0$, then $p_{ii}^{(r+n+s)} \geq p_{ij}^{(r)} p_{jj}^{(n)} p_{ji}^{(s)} > 0$, so t_i divides $r + n + s$, hence t_i also divides n.
- So, t_i is a common divisor of $\{n \in \mathbf{N};\ p_{jj}^{(n)} > 0\}$.
- So, $t_j \geq t_i$ (since t_j is greatest common divisor).
- Similarly, $t_i \geq t_j$, so $t_i = t_j$.   *Q.E.D.*

- COR: if chain irreducible, then all states have same period.
- COR: if chain irreducible and $p_{ii} > 0$ for some state i, then chain is aperiodic.

Markov Chain Convergence Theorem:

- 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 some lemmas.
- STATIONARY RECURRENCE LEMMA: If chain irreducible, and has stationary dist, then it is recurrent.
- PROOF:
  - Suppose the chain is not recurrent.
  - Then by Stronger Recurrence Theorem, for all $i, j \in S$, $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$.
  - Hence, $\lim\limits_{n\to\infty} p_{ij}^{(n)} = 0$.
  - But $\pi_j = \sum_{i\in S} \pi_i p_{ij}^{(n)}$ for all n.
  - Since $\sum_{i\in S} \sup_n |\pi_i p_{ij}^{(n)}| \leq \sum_{i\in S} \pi_i = 1 < \infty$, and each term $\pi_i p_{ij}^{(n)} \to 0$, it follows from the Dominated Convergence Theorem (or "Weierstrass M-test"?) that as $n \to \infty$, $\sum_{i\in S} \pi_i p_{ij}^{(n)} \to 0$ as well.
  - This implies that $\pi_j = 0$ for all $j \in S$.
  - But we must have $\sum_{j\in S} \pi_j = 1$. Impossible!
  - So, the chain must be recurrent.   *Q.E.D.*

- NUMBER THEORY LEMMA: If a set A of positive integers is non-empty, and additive (i.e. $m + n \in A$ whenever $m \in A$ and $n \in A$), and aperiodic (i.e. $gcd(A) = 1$), then there is $n_0 \in \mathbf{N}$ such that $n \in A$ for all $n \geq n_0$.
- (Proof omited; see e.g. Durrett p. 51 / 2nd ed. p. 24, or Rosenthal p. 92.)
- COR: If a state i is aperiodic, and $f_{ii} > 0$, then there is $n_0(i)$ 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, A is additive since $p_{ii}^{(m+n)} \geq p_{ii}^{(m)} p_{ii}^{(n)}$.
- And, A is aperiodic by assumption.
- Hence, the result follows from the Number Theory Lemma. *Q.E.D.*

- COR: If a chain is irreducible and aperiodic, then for any states $i, j \in S$, there is $n_0(i, j)$ such that $p_{ij}^{(n)} > 0$ for all $n \geq n_0(i, j)$.
- PROOF:
  - Find $n_0(i)$ as above.
  - Find $m \in \mathbf{N}$ such that $p_{ij}^{(m)} > 0$.
  - 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 $p_{ij}^{(n)} \geq p_{ii}^{(n-m)} p_{ij}^{(m)} > 0$. *Q.E.D.*

- PROOF OF MARKOV CHAIN CONVERGENCE THEOREM (long!):
- Define a *new* Markov chain $\{(X_n, Y_n)\}_{n=0}^{\infty}$, with state space $\overline{S} = S \times S$, and transition probabilities $\overline{p}_{(ij),(k\ell)} = p_{ik} p_{j\ell}$.
  - Intuition: new chain has two coordinates, each of which is an independent copy of the original Markov chain. ("coupling")
- The new chain has stationary distribution $\overline{\pi}_{(ij)} = \pi_i \pi_j$ (because of independence).
- Furthermore, $\overline{p}_{(ij),(k\ell)}^{(n)} > 0$ whenever $n \geq \max\left[n_0(i, k),\, n_0(j, \ell)\right]$.
  - So, new chain is <u>irreducible</u> and <u>aperiodic</u>.
- So, by Stationary Recurrence Lemma, new chain is <u>recurrent</u>.
- Choose any $i_0 \in S$, and set $\tau = \inf\{n \geq 0; X_n = Y_n = i_0\}$.
- By Stronger Recurrence Theorem, $\overline{f}_{(ij),(i_0 i_0)} = 1$, i.e. $\mathbf{P}_{(ij)}(\tau < \infty) = 1$.
- Note also that if $n \geq m$, then

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

- Hence, for $i, j, k \in S$,

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

  which $\to 0$ as $n \to \infty$ since $\mathbf{P}_{(ij)}(\tau < \infty) = 1$.
  - (Above factor of "2" not really necessary, since both terms non-negative.)

END OF WEEK #3

- Hence, it follows that

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

which $\to 0$ as $n \to \infty$ since $|p_{ij}^{(n)} - p_{kj}^{(n)}| \to 0$ for all $k \in S$ (using the M-test).

- Finally, for any $\{\nu_i\}$ (again using the M-test),

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

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

Q.E.D. (phew!)

- So, for Frog Example, $\lim_{n\to\infty} \mathbf{P}(X_n = 14) = 1/20$, regardless of $\{\nu_i\}$.
- COR: If chain irreducible and aperiodic, then it has <u>at most one</u> stationary distribution.
  - PROOF: If it has at least one, then by the above, each one must be equal to $\lim_{n\to\infty} \mathbf{P}(X_n = j)$, so they're all equal. *Q.E.D.*

- Example: $S = \{1, 2, 3\}$, and $(\, p_{ij} \,) \;=\; \begin{pmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 2/3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
  - Stationary dist #1: $\pi_1 = \pi_2 = 1/2$ and $\pi_3 = 0$.
  - Stationary dist #2: $\pi_1 = \pi_2 = 0$ and $\pi_3 = 1$.
  - Stationary dist #3: $\pi_1 = \pi_2 = 1/8$ and $\pi_3 = 3/4$.
  - So, here, the stationary distribution is not unique!
  - But chain is not irreducible.

Periodic Convergence:

- What about periodic chains? (e.g. s.r.w., Ehrenfest)
- CYCLIC DECOMPOSITION LEMMA: Suppose chain irreducible, with period $b \ge 2$. 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 $0 \le r \le b-2$, then $\sum_{j\in S_{r+1}} p_{ij} = 1$, while if $i \in S_b$, then $\sum_{j\in S_0} p_{ij} = 1$. [picture] Furthermore, the chain $P^{(b)}$, when restricted to S_0, is irreducible and aperiodic.
- PROOF (outline): Fix $i_0 \in S$. Then for $r = 0, 1, 2, \ldots, b-1$, let $S_r = \{j \in S : p_{i_0 j}^{(bm+r)} > 0$ for some $m \in \mathbf{N}\}$. Then the stated properties are easily verified [CHECK!]. *Q.E.D.*

- PERIODIC CONVERGENCE THM: Suppose chain irreducible, with period $b \ge 2$, and stat dist $\{\pi_i\}$. Then $\forall i, j \in S$, $\lim_{n\to\infty} \frac{1}{b}\,[p_{ij}^{(n)} + \ldots + p_{ij}^{(n+b-1)}] = \pi_j$, i.e.

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

- PROOF (outline): Let $S_0, S_1, \ldots, S_{b-1}$ be as in Cyclic Decomposition Lemma.
  - Then must have $\pi(S_0) = \pi(S_1) = \ldots = \pi(S_{b-1}) = 1/b$ [check!].
  - Let $\hat{P}$ be the Markov chain corresponding to $P^{(b)}$ restricted to S_0.
  - Then let $\hat{\pi}_i = b\,\pi_i$ for all $i \in S_0$.
  - Then $\hat{\pi}$ is stationary for $\hat{P}$ [check!].
  - By usual Markov Chain Convergence Theorem, $\lim_{n\to\infty} \hat{p}_{ij}^{(n)} = \hat{\pi}_j$ for all $i, j \in S_0$, i.e. $\lim_{n\to\infty} p_{ij}^{(bn)} = b\,\pi_j$.
  - Then for $i \in S_0$ and $j \in S_r$, we have $\lim_{n\to\infty} p_{ij}^{(bn-r)} = \lim_{n\to\infty} \sum_{k\in S} p_{ik}^{(b-r)} p_{kj}^{(b(n-1))} = \sum_{k\in S} p_{ik}^{(b-r)} \lim_{n\to\infty} p_{kj}^{(b(n-1))} = \sum_{k\in S_0} p_{ik}^{(b-r)} \lim_{n\to\infty} p_{kj}^{(b(n-1))} = \sum_{k\in S} p_{ik}^{(b-r)} [b\,\pi_j] = b\,\pi_j$, again using the M-test.
  - Hence, for $i \in S_0$, $\lim_{n\to\infty} \frac{1}{b}[p_{ij}^{(bn)} + p_{ij}^{(bn+1)} + \ldots + p_{ij}^{(bn+b-1)}] = \frac{1}{b}[b\,\pi_j + 0] = \pi_j$ for any $j \in S$.
  - By relabeling, for any $i, j \in S$, $\lim_{n\to\infty} \frac{1}{b}[p_{ij}^{(bn)} + p_{ij}^{(bn+1)} + \ldots + p_{ij}^{(bn+b-1)}] = \pi_j$.
  - Follows (by reindexing) that $\lim_{n\to\infty} \frac{1}{b}\,[p_{ij}^{(n)} + \ldots + p_{ij}^{(n+b-1)}] = \pi_j$. *Q.E.D.*

- e.g. for Ehrenfest's Urn, if $i_0 = 0$, then $S_0 = \{\text{even } i \in S\}$, and $S_1 = \{\text{odd } i \in S\}$, and $\lim_{n\to\infty} \frac{1}{2}[p_{ij}^{(n)} + p_{ij}^{(n+1)}] = \pi_j = 2^{-d}\binom{d}{j}$.
- PERIODIC CONVERGENCE COR: 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: For aperiodic chains, this follows from the usual Markov Chain Convergence Theorem, together with "Cezàro sums". For chains with period $b \geq 2$, this follows from the Periodic Markov Chain Convergence Theorem, again by "Cezàro sums". *Q.E.D.*

- UNIQUE STATIONARY COROLLARY: If Markov chain P is irreducible (not necessarily aperiodic), then it has <u>at most one</u> stationary distribution (just like before).
- PROOF: If it has at least one, then by the above, each one 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. *Q.E.D.*

- Example: $S = \{1, 2, 3\}$, $p_{11} = 1$, $p_{22} = p_{23} = p_{32} = p_{33} = 1/2$. For <u>any</u> $a \in [0, 1]$, let $\pi_1 = a$, and $\pi_2 = \pi_3 = (1-a)/2$. Stationary! So, infinitely many stationary distributions! e.g. $\pi_1 = 1$ and $\pi_2 = \pi_3 = 0$, <u>or</u> $\pi_1 = 0$ and $\pi_2 = \pi_3 = 1/2$. Not unique! But not <u>irreducible</u>.
- What about simple random walk? Does it have a stationary dist?
  - (Would be reversible with respect to "uniform", but what is that??)
  - Know that $p_{ii}^{(n)} \approx [4p(1-p)]^{n/2}\sqrt{2/\pi n}$, so $p_{ii}^{(n)} \leq \sqrt{2/\pi n} \to 0$.
  - Then for any $i, j \in S$, find $m \in \mathbf{N}$ with $p_{ji}^{(m)} > 0$, then $p_{ii}^{(n+m)} \geq p_{ij}^{(n)} p_{ji}^{(m)}$, so we must have $p_{ij}^{(n)} \leq p_{ii}^{(n+m)} \,/\, p_{ji}^{(m)} \to 0$ as well.
  - Then, <u>if</u> had stat dist $\{\pi_i\}$, then $\forall j \in S$, $\pi_j = \sum_{i\in S} \pi_i p_{ij}^{(n)} \to 0$ (using M-test).

- [Or, alternatively, would have $\frac{1}{2}[p_{ij}^{(n)} + p_{ij}^{(n+1)}] \to \pi_j$ and also $\frac{1}{2}[p_{ij}^{(n)} + p_{ij}^{(n+1)}] \to 0$.]
- So, would have $\pi_j = 0$ for all j, so $\sum_j \pi_j = 0$. Impossible! No stationary distribution!
- [Aside: here $\sum_j p_{ij}^{(n)} = 1$ for all n, even though $\sum_j \lim_{n\to\infty} p_{ij}^{(n)} = 0$. So, M-test conditions are <u>not</u> satisfied for <u>that</u> limit.]

- If S is infinite, can there ever be a stationary distribution? Yes!
- Example: $S = \mathbf{N} = \{1, 2, 3, \ldots\}$, and for $i \geq 2$, $p_{i,i} = p_{i,i+1} = 1/4$ and $p_{i,i-1} = 1/2$, and $p_{1,1} = 3/4$ and $p_{1,2} = 1/4$.
  - Let $\pi_i = 2^{-i}$, 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 reversible! So, $\{\pi_i\}$ is stationary dist.
  - Also irreducible and aperiodic (easy).
  - So, $\lim_{n\to\infty} P(X_n = j) = \pi_j = 2^{-j}$ for all $j \in S$.

Application – Metropolis Algorithm (Markov Chain Monte Carlo) (MCMC):

- Let $S = \mathbf{Z}$, and let $\{\pi_i\}$ be <u>any</u> prob dist on S. Assume $\pi_i > 0$ for all i.
- Can we <u>create</u> Markov chain transitions $\{p_{ij}\}$ so that $\lim\limits_{n\to\infty} p_{ij}^{(n)} = \pi_j$.
- Yes! 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.
- Equivalent algorithmic version: Given X_{n-1}, let Y_n equal $X_{n-1} \pm 1$ (prob 1/2 each), and $U_n \sim \text{Uniform}[0, 1]$ (indep.), and

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

- Then $\pi_i p_{i,i+1} = \pi_i \, \frac{1}{2}\min[1, \frac{\pi_{i+1}}{\pi_i}] = \frac{1}{2}\min[\pi_i, \pi_{i+1}]$.
- Also $\pi_{i+1} p_{i+1,i} = \pi_{i+1} \, \frac{1}{2}\min[1, \frac{\pi_i}{\pi_{i+1}}] = \frac{1}{2}\min[\pi_{i+1}, \pi_i]$.
- So $\pi_i p_{ij} = \pi_j p_{ji}$ if $j = i + 1$, hence for all $i, j \in S$.
- So, chain is <u>reversible</u> w.r.t. $\{\pi_i\}$, so $\{\pi_i\}$ stationary.
- Also irreducible and aperiodic (easy).
- So, $\lim\limits_{n\to\infty} p_{ij}^{(n)} = \pi_j$, i.e. $\lim\limits_{n\to\infty} \mathbf{P}[X_n = j] = \pi_j$. *Q.E.D.*
- Widely used to <u>sample from</u> complicated distributions $\{\pi_i\}$, and thus estimate their probability / expected values / etc.
  - [** Animated version available at: www.probability.ca/met **]
- Also works on continuous state spaces, with π a <u>density function</u> (e.g. the Bayesian posterior density).

- "markov chain monte carlo" gives 2,130,000 hits in Google!

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) 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-loops ($w(u, u) > 0$), etc.
- Let $d(u) = \sum_{v \in V} w(u, v)$. ("degree" of vertex u)
- Define a Markov chain on $S = V$ by $p_{uv} = \frac{w(u,v)}{d(u)}$.
  - Check: $\sum_{v \in V} p_{uv} = \frac{\sum_{v \in V} w(u,v)}{\sum_{v \in V} w(u,v)} = 1$.
  - "(simple) random walk on the weighted undirected graph (V, w)"
- Usual case: 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 prob.

END OF WEEK #4

- Other examples: Irreducible? Aperiodic? Stationarity distribution?
- Example: $V = \{1, 2, 3, 4, 5\}$, 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. ("ring") (diagram) Irreducible? Aperiodic? Stationarity distribution?
- Example: $V = \{0, 1, 2, \ldots, K\}$, with $w(i, 0) = w(0, i) = 1$ for $i = 1, 2, 3$, with $w(i, j) = 0$ otherwise. ("star") (diagram)
- Example: $V = \{1, 2, \ldots, K\}$, with $w(i, i + 1) = w(i + 1, i) = 1$ for $1 \le i \le K - 1$, with $w(i, j) = 0$ otherwise. ("stick") (diagram)
- Example: $V = \mathbf{Z}$, with $w(i, i + 1) = w(i + 1, i) = 1$ for all $i \in V$, and $w(i, j) = 0$ otherwise.
  - Random walk on this graph corresponds to simple random walk with $p = 1/2$.
- Example: $V = \{1, 2, \ldots, 20\}$, with $w(i, i) = 1$ for $1 \le i \le 20$, and $w(i, i+1) = w(i+1, i) = 1$ for $1 \le i \le 19$, and $w(20, 1) = w(1, 20) = 1$, and $w(i, j) = 0$ otherwise.
  - Random walk on this graph corresponds to the Frog Example!
- Let $Z = \sum_{u \in V} d(u) = \sum_{u,v \in V} w(u, v)$.
  - In unweighted case, $Z = 2\times$(number of edges).
  - Assume that Z is finite (it might not be, if V is infinite).
  - And, assume that $d(u) > 0$ for all $u \in V$ (so any isolated point has a self-loop), to make $p_{uv} = \frac{w(u,v)}{d(u)}$ well-defined.
- Let $\pi_u = \frac{d(u)}{Z}$, so $\pi_u \ge 0$ and $\sum_u \pi_u = 1$.

- Then $\pi_u p_{uv} = \frac{d(u)}{Z} \frac{w(u,v)}{d(u)} = \frac{w(u,v)}{Z}$.
- And, $\pi_v p_{vu} = \frac{d(v)}{Z} \frac{w(v,u)}{d(v)} = \frac{w(v,u)}{Z} = \frac{w(u,v)}{Z}$. Same!
- So, chain is reversible w.r.t. $\{\pi_u\}$.
- So, $\{\pi_u\}$ is stationary dist.

- If graph is <u>connected</u>, then chain is irreducible.
- If graph is <u>bipartite</u> (i.e., can be divided into two subsets s.t. all links go from one to the other), then the chain has period 2.
  - Otherwise, the chain is aperiodic (since can return to u in 2 steps).
  - (i.e., 1 and 2 are the <u>only</u> possible periods)
- This proves: GRAPH CONVERGENCE THM: for random walk on a connected non-bipartite graph, if $Z < \infty$, then $\lim_{n\to\infty} p_{uv}^{(n)} = \pi_v = \frac{d(v)}{Z}$ for all $u, v \in V$.
  - i.e., $\lim_{n\to\infty} \mathbf{P}[X_n = v] = \frac{d(v)}{Z}$.
- What about bipartite graphs? Use Periodic Convergence Thm!
  - PERIODIC GRAPH CONVERGENCE THM: for random walk on any connected graph with $Z < \infty$ (whether bipartite or not), $\lim_{n\to\infty} \frac{1}{2}[p_{uv}^{(n)} + p_{uv}^{(n+1)}] = \frac{d(v)}{Z}$.
- Example: $V = \{1, 2, \ldots, K\}$, with $w(i, i+1) = w(i+1, i) = 1$ for $1 \le i \le K-1$, with $w(i,j) = 0$ otherwise. ("stick")
  - Connected, but bipartite.
  - $p_{12} = 1$, and $p_{K,K-1} = 1$, and $p_{i,i+1} = p_{i,i-1} = 1/2$ for $2 \le i \le K-1$.
  - $\pi_i = \frac{1}{2K-2}$ for $i = 1, K$, and $\pi_i = \frac{2}{2K-2}$ for $2 \le i \le K-1$.
  - Then, know that $\lim_{n\to\infty} \frac{1}{2}[p_{ij}^{(n)} + p_{ij}^{(n+1)}] = \pi_j$ for all $j \in V$.
- What about star? Or, star with an extra edge between 0 and 0?

Application – Gambler's Ruin:

- 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.
- At each bet, A wins \$1 with prob p, or loses \$1 with prob $1 - p$.
- Let X_n be the amount of money A has at time n.
  - So, $X_0 = a$.
- Let $T_i = \inf\{n \ge 0 : X_n = i\}$ be the first time A has i dollars.
- QUESTION: what is $\mathbf{P}_a(T_c < T_0)$, i.e. the prob that A reaches c dollars <u>before</u> reaching 0 (i.e., before losing all their money)?
  - (see e.g. www.probability.ca/randwalk)
- Example: What does it equal if $c = 10,000$, $a = 9,700$, and $p = 0.49$?
- Example: Is it 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 (good), but there's no limit to how long the game might take (bad).
  - So, how to solve it??
- Key: write $\mathbf{P}_a(T_c < T_0)$ as $s(a)$, and consider it to be a function of a.
  - Can we related the different unknown $s(a)$ to each other?
- Clearly $s(0) = 0$, and $s(c) = 1$.
- Furthermore, on the first bet, A either wins or loses \$1.
  - So, for $1 \le a \le c-1$,

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

- This gives $c-1$ equations for the $c-1$ unknowns.
  - Can solve 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 $x = s(1)$ (unknown).
  - 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 $1 \le a \le c-1$,

$$\begin{aligned} s(a) &= s(a) - s(0) \\ &= (s(a) - s(a-1)) + (s(a-1) - s(a-2)) + \ldots + (s(1) - s(0)) \\ &= \left(\left(\frac{1-p}{p}\right)^{a-1} + \left(\frac{1-p}{p}\right)^{a-2} + \ldots + \left(\frac{1-p}{p}\right) + 1 \right) x \\ &= \begin{cases} \left(\frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right) - 1} \right) x, & p \ne 1/2 \\ a x, & p = 1/2 \end{cases} \end{aligned}$$

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

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

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

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

- Example: If $c = 10,000$, $a = 9,700$, $p = 0.49$, then

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

- Example: If $c = 8$, $a = 6$, $p = 1/3$ ("born rich"),

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

- Example: If $c = 8$, $a = 2$, $p = 2/3$ ("born lucky"),

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

- So, it is better to be born lucky than rich!
- Check: is $s(a)$ continuous as a function of p, as $p \to 1/2$?

Eventual Ruin:

- What happens to Gambler's Ruin as $c \to \infty$?
- What is $\mathbf{P}(T_0 < \infty)$, the probability of eventual ruin?
- By continuity of probabilities, $\mathbf{P}(T_0 < \infty) = \lim_{c\to\infty} \mathbf{P}_a(T_0 < T_c)$.
  - So, let $r(a) = 1 - s(a) = \mathbf{P}_a(T_0 < T_c) = \mathbf{P}(\text{ruin})$.
  - Then $r(a)$ is like $s(a)$, but from the other player's perspective.
  - Hence, need to replace a by $c - a$, and replace p by $1 - p$.
  - Follows that

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

  - (Check: $r(a) + s(a) = 1$.)
- Then compute that

$$\mathbf{P}(T_0 < \infty) \;=\; \lim_{c\to\infty} \mathbf{P}_a(T_0 < T_c) \;=\; \begin{cases} \frac{0-1}{0-1} = 1, & p < 1/2 \\ \frac{1}{1} = 1, & p = 1/2 \\ \left(\frac{p}{1-p}\right)^{-a}, & p > 1/2 \end{cases}$$

- e.g. if $p = 2/3$ and $a = 2$, then $\mathbf{P}(T_0 < \infty) = \left(\frac{2/3}{1-(2/3)}\right)^{-2} = 2^{-2} = 1/4$.
  - Hence, $\mathbf{P}(T_0 = \infty) = 3/4$, i.e. have probability 3/4 of *never* going broke.

Expected Return Times:

- DEFN: the expected return time of a state i is $m_i = \mathbf{E}_i(\inf\{n \geq 1 : X_n - i\})$.
- DEFN: a state is positive recurrent if $m_i < \infty$.
  - (It is null recurrent if it is recurrent but $m_i = \infty$.)
  - Connection to stationary distributions??
- 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.
  - So, by the SLLN, $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k=i} = 1/m_i$.
  - Hence, by Bounded Convergence Theorem, $\lim_{n\to\infty} \mathbf{E}_i\left(\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k=i}\right) = 1/m_i$.
  - But if the chain converges to π, then $\lim_{n\to\infty} \mathbf{E}_i\left(\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k=i}\right) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbf{E}_i(\mathbf{1}_{X_k=i}) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbf{P}_i(X_k = i) = \lim_{n\to\infty} \mathbf{P}_i(X_n = i) = \lim_{n\to\infty} p_{ii}^{(n)} = \pi_i$.
  - (If the chain is periodic, then this still holds, by the Periodic Convergence Theorem.)
  - So, we must have $1/m_i = \pi_i$, i.e. $m_i = 1/\pi_i$!
  - This proves the RETURN TIME THEOREM: If an irreducible Markov chain on a discrete state space S has a stationary distribution π, then for any state $i \in S$, the mean return time satisfies $m_i \equiv \mathbf{E}_i(T_i) = 1/\pi_i$.
  - (For more details, see e.g. Rosenthal, Theorem 8.4.9.)
- By contrast, if $m_j = \infty$ for all j, then there is no stationary distribution.
  - So, for s.r.w., must have $m_j = \infty$ for all j.
  - However, if chain is irreducible and S is finite, then $m_j < \infty$ for all j, i.e. there is a stationary distribution.

Martingales:

- MOTIVATION: Gambler's ruin with $p = 1/2$. (So know that $s(a) = a/c$.)
  - Let $T = \inf\{n \geq 0 : X_n = 0 \text{ or } c\}$ = time when game ends.
  - 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".

END OF WEEK #5

- Makes sense since $\mathbf{E}(X_{n+1} \mid X_n = i) = (1/2)(i+1) + (1/2)(i-1) = i$.
- Reverse logic: If we *knew* that $\mathbf{E}(X_T) = \mathbf{E}(X_0) = a$, then could compute that $a = c\, s(a) + 0\,(1 - s(a))$, so must have $s(a) = a/c$. (Easier solution!)

- DEFN: A sequence $\{X_n\}_{n=0}^{\infty}$ of random variables is a martingale if $\mathbf{E}|X_n| < \infty$ for each n, and also $\mathbf{E}(X_{n+1}|X_0, ..., X_n) = X_n$ (i.e., it stays same on average).
- SPECIAL CASE: If $\{X_n\}$ is a Markov chain (with $\mathbf{E}|X_n| < \infty$), then $\mathbf{E}[X_{n+1}|X_0, ..., X_n] = \sum_j j\, P[X_{n+1} = j|X_0, ..., X_n] = \sum_j j\, p_{X_n,j}$, so martingale if $\sum_j j\, p_{ij} = i$ for all i.
- EXAMPLE: Let $\{X_n\}$ be simple random walk with $p = 1/2$ (i.e., "simple symmetric random walk", or s.s.r.w.).
  - Martingale, since $\sum_j j\, p_{ij} = (i+1)(1/2) + (i-1)(1/2) = i$.
- (Optional aside: in defn of martingale, suffices to check that $\mathbf{E}|X_n| < \infty$ for all n, and also $\mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_n$, where $\{\mathcal{F}_n\}$ is any nested filtration for $\{X_n\}$, i.e. $\mathcal{F}_n$ is a sub-σ-algebra with $\sigma(X_0, X_1, \ldots, X_n) \subseteq \mathcal{F}_n \subseteq \mathcal{F}_{n+1}$.)
- If $\{X_n\}$ martingale, then it follows from "double-expectation formula" that

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

  i.e. that $\mathbf{E}(X_n) = \mathbf{E}(X_0)$ for all n.
- But what about $\mathbf{E}(X_T)$ for a random time T?

Stopping Times:

- DEFN: 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., can't look into future before deciding to stop.
  - e.g. $T = \inf\{n \geq 0 : X_n = 5\}$ is a valid stopping time. ($= \infty$ if never hit 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).
- Do we always have $\mathbf{E}(X_T) = \mathbf{E}(X_0)$, if T is a stopping time?
  - At least if $P(T < \infty) = 1$?
- Not necessarily!
  - e.g. let $\{X_n\}$ be s.s.r.w. with $X_0 = 0$. Martingale!
  - Let $T = T_{-5} = \inf\{n \geq 0 : X_n = -5\}$. Stopping time!
  - And, $\mathbf{P}(T < \infty) = 1$ since s.s.r.w. is recurrent.
  - But $X_T = -5$, so $\mathbf{E}(X_T) = -5 \neq 0 = \mathbf{E}(X_0)$.
  - What went wrong? Need some boundedness conditions!

- OPTIONAL STOPPING LEMMA: If $\{X_n\}$ martingale, with stopping time T which is bounded (i.e., $\exists\, M < \infty$ with $\mathbf{P}(T \le M) = 1$), then $\mathbf{E}(X_T) = \mathbf{E}(X_0)$.
- PROOF: Using the double-expectation formula, and then the fact that "$1 - \mathbf{1}_{T \le k-1}$" is completely determined by $X_0, X_1, \ldots, X_{k-1}$ (and thus can be treated as a constant in the conditional expectation; this fact is optional), we have:

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

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

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

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

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

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

  - Question: How does this proof break down if $M = \infty$?
- Example: s.s.r.w., with $X_0 = 0$, and let $T = \min\big(10^{12},\ \inf\{n \ge 0 : X_n = -5\}\big)$.
  - Then $T \le 10^{12}$, so T bounded, so $\mathbf{E}(X_T) = \mathbf{E}(X_0) = \mathbf{E}(0) = 0$.
  - But nearly always have $X_T = -5$. Contradiction??
  - No, since by the Law of Total Expectation, $0 = \mathbf{E}(X_T) = \mathbf{P}(X_T = -5)\mathbf{E}(X_T \,|\, X_T = -5) + \mathbf{P}(X_T \ne 5)\mathbf{E}(X_T \,|\, X_T \ne 5)$, and $\mathbf{E}(X_T \,|\, X_T = -5) = -5$, and $\mathbf{P}(X_T = -5) \approx 1$, and $\mathbf{P}(X_T \ne 5) \approx 0$, but the equation still holds since $\mathbf{E}(X_T \,|\, X_T \ne 5)$ is huge.
- Can we apply this to the Gambler's Ruin problem?
  - No, since there T is not bounded!
  - Need something more general!
- OPTIONAL STOPPING THM: 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 $X_{S_m} = X_{\min(T,m)} = X_T\mathbf{1}_{T \le m} + X_m\mathbf{1}_{T>m} = X_T - X_T\mathbf{1}_{T>m} + X_m\mathbf{1}_{T>m}$.
  - So, $X_T = X_{S_m} + X_T\mathbf{1}_{T>m} - X_m\mathbf{1}_{T>m}$.

- So, $\mathbf{E}(X_T) = \mathbf{E}(X_{S_m}) + \mathbf{E}(X_T \mathbf{1}_{T>m}) - \mathbf{E}(X_m \mathbf{1}_{T>m})$. (three terms to consider)
- First term $= \mathbf{E}(X_0)$ from above.
- Second term $\to 0$ as $m \to \infty$ by Dominated Convergence Thm, since $\mathbf{E}|X_T| < \infty$ and $\mathbf{1}_{T>m} \to 0$ (since $\mathbf{P}(T < \infty) = 1$).
- Third term $\to 0$ as $m \to \infty$ by assumption.
- So, $\mathbf{E}(X_T) \to \mathbf{E}(X_0)$, i.e. $\mathbf{E}(X_T) = \mathbf{E}(X_0)$. *Q.E.D.*

- 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 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, result follows from Optional Stopping Theorem. *Q.E.D.*

- Example: Gambler's Ruin with $p = 1/2$, and $T = \inf\{n \ge 0 : X_n = 0 \text{ or } X_n = c\}$.
  - Then $\mathbf{P}(T < \infty) = 1$ (game must eventually end). [Formally: $\mathbf{P}(T > mc) \le (1 - p^c)^m \to 0$ as $m \to \infty$, since if win c times in a row then game over.]
  - Also, $|X_n| \mathbf{1}_{n \le T} \le c < \infty$ for all n.
  - So, by Optional Stopping Corollary, $\mathbf{E}(X_T) = \mathbf{E}(X_0) = a$.
  - Hence, as before, $a = c\, s(a) + 0\,(1 - s(a))$, so must have $s(a) = a/c$. (Easier solution!)
- What about Gambler's Ruin with $p \ne 1/2$?
  - Here $\{X_n\}$ is <u>not</u> a martingale: $\sum_j j\, p_{ij} = p(i+1) + (1-p)(i-1) = i + 2p - 1 \ne i$.
  - Trick: let $Y_n = \left(\frac{1-p}{p}\right)^{X_n}$. Then $\{Y_n\}$ is Markov chain too.
  - Then $\mathbf{E}(Y_{n+1} \,|\, Y_0, Y_1, \ldots, Y_n) = p\left[Y_n \left(\frac{1-p}{p}\right)\right] + (1-p)\left[Y_n \,/\, \left(\frac{1-p}{p}\right)\right] = Y_n(1-p) + Y_n(p) = Y_n$.
  - So, $\{Y_n\}$ is a martingale!
  - And, $\mathbf{P}(T < \infty) = 1$ as before (with the same T).
  - And, $|Y_n| \mathbf{1}_{n \le T} \le \max\left(\left(\frac{1-p}{p}\right)^0,\ \left(\frac{1-p}{p}\right)^c\right) < \infty$ for all n.
  - Hence, $\mathbf{E}(Y_T) = \mathbf{E}(Y_0) = \left(\frac{1-p}{p}\right)^a$.
  - But $Y_T = \left(\frac{1-p}{p}\right)^c$ if win, or $Y_T = \left(\frac{1-p}{p}\right)^0 = 1$ if lose.
  - 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, easier solution!)

- REMINDER: Midterm next week!
- REMINDER: Reading Week (no class) the week after!

END OF WEEK #6

- (Midterm.)

END OF WEEK #7

Wald's Theorem:

- WALD'S THM: Suppose $X_n = a + Z_1 + \ldots + Z_n$, where $\{Z_i\}$ are iid, with finite mean m. Let T be a stopping time for $\{X_n\}$ which has finite mean, i.e. $\mathbf{E}(T) < \infty$. Then $\mathbf{E}(X_T) = a + m\,\mathbf{E}(T)$.
- Special case: if $m = 0$, then $\{X_n\}$ is a martingale, and Wald's Thm says that $\mathbf{E}(X_T) = a = \mathbf{E}(X_0)$, as usual.
- Example: $\{X_n\}$ is s.s.r.w. with $X_0 = 0$, and $T = \inf\{n \geq 0 : X_n = -5\}$.
  - Then $\mathbf{E}(X_T) = -5 \neq 0 = \mathbf{E}(X_0)$.
  - Contradiction?? No; it turns out that here $\mathbf{E}(T) = \infty$.
- PROOF: We compute (using that Z_i indep of $\{T \geq i\} = \{T \leq i-1\}^C$) that

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

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

$$= \sum_{i=1}^{\infty} \mathbf{E}[Z_i]\,\mathbf{E}[\mathbf{1}_{T\geq i}] \;=\; m\sum_{i=1}^{\infty} \mathbf{P}[T \geq i] \;=\; m\,\mathbf{E}(T)\,, \quad Q.E.D.$$

  - (Note: the above calculation can be justified using the Dominated Convergence Thm with dominator $Y = \sum_{i=1}^{\infty} |Z_i|\,\mathbf{1}_{T\geq i}$, since $\mathbf{E}(Y) = \mathbf{E}[\sum_{i=1}^{\infty} |Z_i|\,\mathbf{1}_{T\geq i}] = \sum_{i=1}^{\infty} \mathbf{E}[|Z_i|\,\mathbf{1}_{T\geq i}] = \sum_{i=1}^{\infty} \mathbf{E}|Z_i|\,\mathbf{E}[\mathbf{1}_{T\geq i}] = \mathbf{E}|Z_1| \sum_{i=1}^{\infty} \mathbf{P}[T \geq i] = \mathbf{E}|Z_1|\,\mathbf{E}(T) < \infty$.)
- EXAMPLE: Gambler's Ruin with $p \neq 1/2$, and $T = \inf\{n \geq 0 : X_n = 0 \text{ or } c\}$. (see e.g. www.probability.ca/randwalk)
  - What is $\mathbf{E}(T)$ = expected number of bets in the game?
  - Well, here $m = \mathbf{E}(Z_i) = p(1) + (1-p)(-1) = 2p - 1$.
  - Also, $\mathbf{E}(X_T) = c\,s(a) + 0\,(1 - s(a)) = c\frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1}$.
  - And, $\mathbf{E}(T) < \infty$. [For example, this follows since $\mathbf{P}(T \geq cn) \leq (1-p^c)^n$ so $\mathbf{E}(T) = \sum_{i=1}^{\infty} \mathbf{P}(T \geq i) \leq \sum_{j=0}^{\infty} c\,\mathbf{P}(T \geq cj) \leq \sum_{j=0}^{\infty} c\,(1-p^c)^j = c/[1-(1-p^c)] = c/p^c < \infty$.]

– Hence, by Wald's Thm, $\mathbf{E}(X_T) = a + m\,\mathbf{E}(T)$.

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

– e.g. $p = 0.49$, $a = 9,700$, $c = 10,000$: $\mathbf{E}(T) = 484,997$. (large!)

- But what about $\mathbf{E}(T)$ when $p = 1/2$?
  – Then $m = 0$, so the above method does not work.

- LEMMA: Let $X_n = a + Z_1 + \ldots + Z_n$, where $\{Z_i\}$ 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:
  – Check: $\mathbf{E}|Y_n| \le \mathbf{Var}(X_n) + nv = 2nv < \infty$.
  – Also, since Z_{n+1} indep of $Z_1, \ldots, Z_n, Y_0, \ldots, Y_n$, we have

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

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

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

$$= \;Y_n + v - v + 2\,\mathbf{E}(Z_{n+1})\,\mathbf{E}\Big[Z_1 + \ldots + Z_n \;\Big|\; Y_0, Y_1, \ldots, Y_n\Big] \;=\; Y_n + 0\,, \quad Q.E.D.$$

- COR: 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 $v = 1$). Martingale (by Lemma)!
  – Choose $M > 0$, and let $S_M = \min(T, M)$. Stopping time! Bounded!
  – Hence, by 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 $M \to \infty$, $S_M \to T$ (obviously). This implies that $\mathbf{E}(S_M) \to \mathbf{E}(T)$ (by Monotone Convergence Thm), and $\mathbf{E}[(X_{S_M} - a)^2] \to \mathbf{E}[(X_T - a)^2]$ (by Bounded Convergence Thm, since for any n, $(X_{S_M} - a)^2 \le \max\left(a^2, (c-a)^2\right) < \infty$).
  – Hence, $\mathbf{E}(T) = \mathbf{E}[(X_T - a)^2] = \mathbf{Var}(X_T)$ (since $\mathbf{E}(X_T) = a$).
  – But $\mathbf{Var}(X_T) = (a/c)\,(c - a)^2 + (1 - a/c)a^2 = (a/c)(c^2 + a^2 - 2ac) + (a^2 - a^3/c) = ac + a^3/c - 2a^2 + a^2 - a^3/c = ac - a^2 = a(c - a)$, $\quad Q.E.D.$

- e.g. $c = 10,000$, $a = 9,700$, $p = 1/2$: $\mathbf{E}(T) = a(c - a) = 2,910,000$. (even larger!)

- e.g. "Double Until You Win":
  – Let $\{Z_i\}$ be iid ± 1 with probability $1/2$ each.
  – Suppose we bet 2^i dollars on the i^{th} bet.
  – Then our winnings after n bets equals $X_n = \sum_{i=1}^n 2^i Z_i$.
  – Let $T = \inf\{n \ge 1 : Z_n = +1\}$ be our first win.

- Then 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 $m = 0$ on each bet.
- Thus $\mathbf{E}(X_T) = 1 \neq a + m\,\mathbf{E}(T)$, even though $\mathbf{E}(T) = 2 < \infty$.
- This does not contradict Wald's Theorem, since $\{2^i Z_i\}$ is not iid.

Application – Sequence Waiting Times:

- Suppose we repeatedly flip a fair coin. Let τ be the first time the sequence "HTH" is completed. What is $\mathbf{E}(\tau)$?
  - And, is the answer the same for "THH"?
  - Try it out: file www.probability.ca/sta447/Rseqwait
- One solution: 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 HHTTHT, then $X_1 = 1$, $X_2 = 1$, $X_3 = 2$, $X_4 = 0$, $X_5 = 1$, and $X_6 = 2$.) Take $X_0 = 0$.
  - So, $S = \{0, 1, 2, 3\}$, with $X_0 = 0$, and $X_\tau = 3$.
  - Then $p_{01} = p_{12} = p_{23} = 1/2$. (Probability of continuing the sequence.)
  - Also $p_{00} = p_{20} = 1/2$. But instead of $p_{10} = 1/2$, have $p_{11} = 1/2$. Key!
  - (That is, if you fail to match the second flip, T, then you've already matched the first flip, H, for the next try.)
  - For completeness, assume we "start over" as soon as we win, so $p_{3j} = p_{0j}$ for all j, i.e. $p_{31} = p_{30} = 1/2$.
  - Thus, $P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix}$.
- Compute from equation $\pi P = \pi$ (check!) that the stationary distribution is $(0.3, 0.4, 0.2, 0.1)$.
  - So, by the Return Time Theorem, 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 10.
  - That is, mean waiting time for HTH is 10.
  - Solved it!
  - Try it out: file www.probability.ca/sta447/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 compute (check!) that the stationary distribution is $(1/8, 1/2, 1/4, 1/8)$.
- So, mean time to return to state 3 is $1/(1/8) = 8$. Smaller!
- Try it out: file www.probability.ca/sta447/Rseqwait

- ANOTHER APPROACH (to HTH), USING MARTINGALES:
- 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 or win three bets in a row.)
  - Let X_n be the total amount won by all the betters by time n.
  - Then since the bets were fair, $\{X_n\}$ is a martingale with stopping time τ.
  - Then have (check!) that $X_\tau = -(\tau - 3) + (-1) + (1) + (7) = -\tau + 10$.
  - It follows 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 Optional Stopping Lemma, and $\lim_{m\to\infty} \mathbf{E}(X_{T_m}) = \mathbf{E}(X_\tau)$ by Dominated Convergence Theorem with dominator $Y = 7\tau$ since $|X_n - X_{n-1}| \le 7$.)
  - Hence, $0 = \mathbf{E}(X_\tau) = -\mathbf{E}(\tau) + 10$, whence $\mathbf{E}(\tau) = 10$. Same as before!
- Similarly, for THH, get that $X_\tau = -(\tau - 3) + (-1) + (-1) + (7) = -\tau + 8$, whence $\mathbf{E}(\tau) = 8$. Same as before!

Martingale Convergence Theorem:

- 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 $i \in S$.
  - 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.
  - Hence, $X_n = 2^{Y_n} \to 2^{-\infty} = 0$ w.p. 1.
- EXAMPLE: Let $\{X_n\}$ be Gambler's Ruin with $p = 1/2$. Then $X_n \to X$ w.p. 1, where $\mathbf{P}(X = c) = a/c$ and $\mathbf{P}(X = 0) = 1 - a/c$.
- MARTINGALE CONVERGENCE THM: Any non-negative martingale $\{X_n\}$ converges w.p. 1 to some random variable X (e.g. $X \equiv 0$).
  - Intuition: since it's non-negative (i.e., bounded on one side), it can't "spread out" forever. And since it's a martingale, it can't "drift" anywhere. So eventually it has to stop somewhere.
  - Proof omitted; see e.g. Rosenthal, p. 169.
- Example: s.s.r.w. – martingale, but not non-negative, does not converge.

- Example: s.s.r.w. stopped at zero – martingale, non-negative, converges to zero.
- Example: symmetric random walk which makes smaller and smaller jumps when near zero – martingale, non-negative, converges to something.
- Example: s.r.w. with $p = 2/3$ stopped at zero – non-negative, does not converge (might increase to infinity), but not a martingale.
- COR: Any martingale $\{X_n\}$ with $X_n \geq -c$ for some $c \in \mathbf{R}$ converges w.p. 1 to some random variable X.
  - PROOF: Apply THM to $\{X_n + c\}$.
- COR: Any martingale $\{X_n\}$ with $X_n \leq c$ for some $c \in \mathbf{R}$ converges w.p. 1 to some random variable X.
  - PROOF: Apply THM to $\{-X_n + c\}$.

—————————— **END OF WEEK #8** ——————————

Application – Branching Processes:

- Let μ be any prob dist on $\{0, 1, 2, \ldots\}$. ("offspring distribution")
  - Have X_n individuals at time n. (e.g., people with colds)
  - 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)
  - (Just one "parent" per offspring, i.e. asexual reproduction.)
  - That is, $X_{n+1} = Z_{n,1} + Z_{n,2} + \ldots + Z_{n,X_n}$, where $\{Z_{n,i}\}_{i=1}^{X_n}$ are i.i.d. $\sim \mu$.
  - Then $\{X_n\}$ is Markov chain, on state space $S = \{0, 1, 2, \ldots\}$.
  - $p_{00} = 1$, with $p_{0j} = 0$ for all $j \geq 1$.
  - p_{ij} is more complicated; in fact (optional), $p_{ij} = (\mu * \mu * \ldots * \mu)(j)$, a <u>convolution</u> of i copies of μ.
- Will $X_n = 0$ for some n?
  - How can martingales help?
- Let $m = \sum_i i\, \mu\{i\}$ = mean of μ. ("reproductive number")
  - 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, also $\mathbf{P}(X_n \to \infty) > 0$. ("flourishing")
  - But assuming $\mu\{0\} > 0$, still have $\mathbf{P}(X_n \to \infty) < 1$, indeed $\mathbf{P}(X_n \to 0) > 0$ (e.g., if no one has any offspring at all on the first iteration: prob $= (\mu\{0\})^a > 0$).
  - So, have possible extinction, but also possible flourishing.
- What if $m = 1$?
  - Then $\mathbf{E}(X_n) = \mathbf{E}(X_0) = a$ for all n.
  - In fact, $\{X_n\}$ is a martingale, and non-negative.
  - So, by Martingale Convergence Thm, must have $X_n \to X$ w.p. 1, for some random variable X.
  - But how can $\{X_n\}$ converge w.p. 1? Either (a) $\mu\{1\} = 1$, or (b) $X = 0$.
  - (In all other cases, $\{X_n\}$ would continue to fluctuate, i.e. not converge w.p. 1.)
  - So, if non-degenerate (i.e., $\mu\{1\} < 1$), then $X \equiv 0$, i.e. $\{X_n\} \to 0$ w.p. 1.
  - Certain extinction, even when $m = 1$!

CONTINUOUS PROCESS: 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 speeded 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 $\{Y_t^{(M)}\}$.
- But $Y_t^{(M)} = \frac{1}{\sqrt{M}}(Z_1 + Z_2 + \ldots + Z_{tM})$ (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 $\mathbf{Var}(Y_t^{(M)}) = \frac{1}{M}(tM) = t$.
  - So, as $M \to \infty$, by the Central Limit Theorem, $Y_t^{(M)} \to \text{Normal}(0, t)$.
  - 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 it's independent 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 $\mathbf{Cov}(B_t, B_s) = \mathbf{E}(B_t B_s) = \mathbf{E}(B_t[B_s - B_t + B_t]) = \mathbf{E}(B_t[B_s - B_t]) + \mathbf{E}((B_t)^2) = \mathbf{E}(B_t)\,\mathbf{E}(B_s - B_t) + \mathbf{E}((B_t)^2) = (0)(0) + t = t$.
  - In general, $\mathbf{Cov}(B_t, B_s) = \min(t, s)$. ("covariance structure")
- DEFINITION: <u>Brownian motion</u> is a process $\{B_t\}_{t \ge 0}$ satisfying the above properties, and with continuous sample paths (i.e., the mapping $t \to B_t$ is continuous).
  - FACT: Such a process exists! (The above construction is <u>intuitive</u>, but a formal proof of existence requires measure theory.)
- Example: What is $\mathbf{E}[(B_2 + B_3 + 1)^2]$?
  - Well, $\mathbf{E}[(B_2+B_3+1)^2] = \mathbf{E}[(B_2)^2)]+\mathbf{E}[(B_3)^2]+1^2+2\mathbf{E}[B_2B_3]+2\mathbf{E}[B_2(1)]+2\mathbf{E}[B_3(1)] = 2 + 3 + 1 + 2(2) + 2(0) + 2(0) = 10$.
- Example: What is $\mathbf{Var}[B_3 + B_5 + 6]$?
  - Well, $\mathbf{Var}[B_3 + B_5 + 6] = \mathbf{E}[(B_3 + B_5)^2] = \mathbf{E}[(B_3)^2)] + \mathbf{E}[(B_5)^2] + 2\mathbf{E}[B_3B_5] = 3 + 5 + 2(3) = 14$.
- Aside: w.p. 1, the function $t \mapsto B_t$ is continuous everywhere, but differentiable nowhere.
- 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?
  - Similar to above, $B_s \,|\, B_t = B_t + (B_s - B_t) \mid B_t = B_t + \text{Normal}(0,\ s - t) \sim \text{Normal}(B_t,\ s - t)$. (i.e., given B_t, B_s is normal with mean B_t, 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 $p \equiv \mathbf{P}(B_\tau = b)$?
  - 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 continue, let $Y_t = B_t^2 - t$.

- Then for $0 < t < s$, $\mathbf{E}[Y_s \mid \{B_r\}_{r \le t}] = \mathbf{E}[B_s^2 - s \mid \{B_r\}_{r \le t}]$
  $= \mathbf{Var}[B_s \mid \{B_r\}_{r \le t}] + (\mathbf{E}[B_s \mid \{B_r\}_{r \le t}])^2 - s$
  $= (B_t)^2 + (s-t) - s = Y_t$.
- Then by double-expectation formula (see e.g. Rosenthal, Prop 13.2.7), we also have $\mathbf{E}[Y_s \mid \{Y_r\}_{r \le t}] = \mathbf{E}\Big[\mathbf{E}[Y_s \mid \{B_r\}_{r \le t}] \Big| \{Y_r\}_{r \le t}\Big] = \mathbf{E}[Y_t \mid \{Y_r\}_{r \le t}] = Y_t$.
- 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 $\mathbf{E}(Y_\tau) = 0$, solve to get $e = ab$. (like for discrete Gambler's Ruin)
  - But τ is not bounded ...

• To justify this argument, i.e. show 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.

• 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 $\mathbf{Var}(X_t) = 3^2\mathbf{Var}(B_t) = 9t$.
  - Follows that $X_t \sim \text{Normal}(2 + 5t,\, 9t)$.
  - Also for $0 < t < s$, $\mathbf{Cov}(X_t, X_s) = \mathbf{E}[(X_t - 5t - 2)(X_s - 5s - 2)] = \mathbf{E}[(3B_t)(3B_s)] = 9\,\mathbf{E}[B_t\, B_s] = 9t$.
  - Fancy notation: $dX_t = 5\,dt + 3\,dB_t$. ("diffusion")

END OF WEEK #9

• 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"; $\sigma \ge 0$)
  - Then $\mathbf{E}(X_t) = x_0 + \mu t$, and $\mathbf{Var}(X_t) = \sigma^2 t$, and $\mathbf{Cov}(X_t, X_s) = \sigma^2 \min(s,t)$.
  - 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.

Application – Financial Modeling:

- Common model for 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 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.
- Defn: A "European call option" is the option to buy the stock for some amount $\$K$ at some fixed future time $S > 0$?
  - At time S, this is worth $\max(0,\, X_S - K)$.
  - At time 0, it's worth only $e^{-rS} \max(0,\, X_S - K)$.
  - But at time 0, X_S is unknown (random).
- QUESTION: what is the "fair price" of this option?
  - This means the fair "no-arbitrage" price, i.e. a price such that you cannot make a guaranteed profit by buying or selling the option, combined with buying and selling the stock.
  - Note: 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!
- MOTIVATION – DISCRETE-TIME VERSION (cf. Durrett):
  - Suppose stock price equals 100 at time 0, and is either 80 or 130 at time S, and there is an option to buy the stock at time S for $K = 130$.
  - What is this option's fair price??
  - (Assume $r = 0$, i.e. no discounting.)
  - Suppose at time 0 you buy x stocks (for 100 each), and y options (for c each).
  - Then if the stock goes up to 130, your profit is $30x + (20 - c)y$. But if it goes down to 80, your profit is $-20x + (-c)y$.
  - If $y = -(5/2)x$, then these both equal $-(5/2)(8 - c)x$.
  - Hence, there is no arbitrage iff $c = 8$ (fair price).

- Here if we assign probabilities $\mathbf{P}(X_S = 80) = 3/5$ and $\mathbf{P}(X_S = 130) = 2/5$, then the stock price is a martingale since $(3/5)(80) + (2/5)(130) = 100$, and also the option is a martingale since $(3/5)(0) + (2/5)(130 - 110) = 8 = c$.
- Thus, fair price = martingale expected value = $(3/5)(0) + (2/5)(130 - 110) = 8$.

- BACK TO THE DIFFUSION CASE: What is the fair price?
- KEY: If $\mu = r - \frac{\sigma^2}{2}$, then $\{D_t\}$ becomes a martingale (HW#3).
- 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, 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.
  - (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 (HW#3), this fair price becomes:

$$x_0 \, \Phi\left(\frac{(r + \frac{\sigma^2}{2})S - \log(K/x_0)}{\sigma\sqrt{S}}\right) - e^{-rS} K \, \Phi\left(\frac{(r - \frac{\sigma^2}{2})S - \log(K/x_0)}{\sigma\sqrt{S}}\right),$$

    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!]
- 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.

Poisson Processes:

- MOTIVATING EXAMPLE:
  - Suppose an average of $\lambda = 2.5$ fires in some region (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$(n, \lambda/n)$.
  - So, as $n \to \infty$, $\mathbf{P}(\#\text{fires} = k) = \binom{n}{k} p^k (1-p)^{n-k}$
    $= \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} (\lambda/n)^k (1 - (\lambda/n))^{n-k} \approx \frac{n^k}{k!} (\lambda/n)^k (1 - (\lambda/n))^n \approx \frac{1}{k!} \lambda^k e^{-\lambda}$.

- So, if $\lambda = 2.5$, then $\mathbf{P}(\# \text{ fires} = k) \approx e^{-2.5}\frac{(2.5)^k}{k!}$, for $k = 0, 1, 2, 3, \ldots$
- i.e. # fires $\sim$ Poisson(λ) = Poisson(2.5).
- And, # fires today and tomorrow combined $\approx$ Poisson($2 * \lambda$) = Poisson(5), etc.
- Full distribution? $\mathbf{P}$(fire within next hour)? etc.

- POISSON PROCESS CONSTRUCTION:
- Let $\{Y_n\}_{n=1}^{\infty}$ be i.i.d. $\sim \text{Exp}(\lambda)$, 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. $\mathbf{P}(N(t) = m)$?
  - 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 $Y_n \sim \text{Exp}(\lambda) = \text{Gamma}(1, \lambda)$, so $T_m := Y_1 + Y_2 + \ldots + Y_m \sim \text{Gamma}(m, \lambda)$, 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,

$$\mathbf{P}(N(t) = m) \;=\; \mathbf{P}(T_m \leq t,\ T_{m+1} > t) \;=\; \mathbf{P}(\exists\, 0 \leq s \leq t :\ T_m = s,\ Y_{m+1} > t - s)$$

$$= \int_0^t f_{T_m}(s)\, \mathbf{P}(Y_{m+1} > t - s)\, ds \;=\; \int_0^t \frac{\lambda^m}{(m-1)!} s^{m-1} e^{-\lambda s}\, e^{-\lambda(t-s)}\, ds$$

$$= \frac{\lambda^m}{(m-1)!} e^{-\lambda t} \int_0^t s^{m-1}\, ds \;=\; \frac{\lambda^m}{(m-1)!} e^{-\lambda t} [\frac{t^m}{m}] \;=\; \frac{(\lambda t)^m}{m!} e^{-\lambda t}\,.$$

  - Hence, $N(t) \sim \text{Poisson}(\lambda t)$.
  - So, $\mathbf{E}(N(t)) = \lambda t$, and $\mathbf{Var}(N(t)) = \lambda t$.
- Now, recall the "memoryless" (or "forgetfulness") property of the Exp(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 $N(t + s) - N(s) \sim N(t) \sim \text{Poisson}(\lambda t)$.
  - 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 $0 \le t_1 \le s_1 \le t_2 \le s_2 \le \ldots \le t_k \le s_k$, then $N(s_i) - N(t_i) \sim \text{Poisson}(\lambda(s_i - t_i))$, 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 $\{N(t)\}_{t \ge 0}$ of random non-decreasing integer counts $N(t)$, satisfying: (a) $N(0) = 0$; (b) $N(t) \sim \text{Poisson}(\lambda t)$ for all $t \ge 0$; and (c) independent Poisson increments (as above).

- MOTIVATING EXAMPLE (cont'd):
  - Here, fires approximately follow a Poisson process with intensity 2.5.
  - So, $\mathbf{P}(\text{9 fires today and tomorrow combined}) \approx e^{-2*2.5}\frac{(2*2.5)^9}{9!} = e^{-5}(\frac{5^9}{9!}) \doteq 0.036$.
  - $\mathbf{P}(\text{at least one fire in next hour}) = 1 - \mathbf{P}(\text{no fires in next hour})$
    $= 1 - \mathbf{P}(N(1/24) = 0) = 1 - e^{-2.5/24}\frac{(2.5/24)^0}{0!} \doteq 1 - 0.90 = 0.10$.
  - $\mathbf{P}(\text{exactly 3 fires in next hour}) = e^{-2.5/24}\frac{(2.5/24)^3}{3!} \doteq 0.00017 \doteq 1/5891$, 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: Given $N(t)$, as $h \searrow 0$,
  - $\mathbf{P}(N(t+h) - N(t) = 1) = \lambda h + o(h)$.
  - $\mathbf{P}(N(t+h) - N(t) \geq 2) = o(h)$.
  - This (together with independent increments) is another way to characterise Poisson processes.
- NOTE: the $\{T_i\}$ tend to "clump up" in various patterns just by chance alone.
  - Doesn't "mean" anything at all: they're <u>independent</u>. ("Poisson clumping")
  - But it "seems" like it does have meaning!
  - See e.g. www.probability.ca/pois
- Illustration:
  - Suppose have PP on [0,100] with intensity $\lambda = 1$.
  - Then <u>expect</u> about one event every distance 1. But will get clumps! e.g.:

$$\mathbf{P}(\exists r \in [1, 100] : N(r) - N(r-1) = 4)$$

$$\geq \mathbf{P}(\exists m \in \{1, 2, \ldots, 100\} : N(m) - N(m-1) = 4)$$

$$= 1 - \mathbf{P}(\nexists m \in \{1, 2, \ldots, 100\} : N(m) - N(m-1) = 4)$$

$$= 1 - (\mathbf{P}(N(1) \neq 4))^{100} = 1 - (1 - P(N(1) = 4))^{100} = 1 - (1 - 1/24e)^{100} \doteq 0.787\,.$$

- 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 <u>two</u> in Feb 1 - Mar 15, 2010; <u>less</u> 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 <u>random</u>.

END OF WEEK #10

- 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 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)
  - If buses are completely random, then they form a Poisson process, so (by memoryless property) waiting time is $\sim \text{Exp}(\lambda)$, so mean $= \frac{1}{\lambda}$ 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 $\{N_1(t)\}_{t\ge 0}$ and $\{N_2(t)\}_{t\ge 0}$ are two independent Poisson processes, with rates λ_1 and λ_2 respectively. Let $N(t) = N_1(t) + N_2(t)$.
  - Then $\{N(t)\}_{t\ge 0}$ 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 $A \sim \text{Exp}(\lambda)$.
  - Hence, $\mathbf{E}(A) = 1/\lambda = 1/8$ hours, i.e. 7.5 minutes.

• THINNING: Let $\{N(t)\}_{t\ge 0}$ 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 independent Poisson processes, with rates λp_i.
- PROOF: "independent increments" is obvious.
- For the distribution, suppose for notational simplicity that there are just two types, with $p_1 + p_2 = 1$.
- Need to show: $\mathbf{P}(N_1(t) = j,\ N_2(t) = k)$
  $= \left(e^{-(\lambda p_1 t)}(\lambda p_1 t)^j / j!\right) \left(e^{-(\lambda p_2 t)}(\lambda p_2 t)^k / k!\right)$.
- 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$)$
  $= \left(e^{-\lambda t}(\lambda t)^{j+k}/(j+k)!\right) \times \binom{j+k}{j}(p_1)^j(p_2)^k$. 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 independent of # grads (and each follows a Poisson distribution).
- ASIDE: Can also have time-inhomogeneous Poisson processes, where $\lambda = \lambda(t)$, and $N(b) - N(a) \sim \text{Poisson}\left(\int_a^b \lambda(t)\, dt\right)$.
- ASIDE: Can also have Poisson processes on other regions, e.g. in two dimensions, etc., cf. www.probability.ca/pois

Continuous-Time, Discrete-Space Markov Processes:

- Recall: Markov chains $\{X_n\}_{n=0}^{\infty}$ defined in discrete (integer) time.
  - But Brownian motion $\{B_t\}_{t\geq 0}$, and Poisson processes $\{N(t)\}_{t\geq 0}$, both defined in continuous (real) time.
  - Can we define Markov processes in continuous time? Yes!
- DEFN: a continuous-time (time-homogeneous, non-explosive) Markov process, on a countable (discrete) state space S, is a collection $\{X(t)\}_{t\geq 0}$ of random variables such that

$$\mathbf{P}(X_0 = i_0, X_{t_1} = i_1, X_{t_2} = i_2, \ldots, X_{t_n} = i_n) \;=\; \nu_{i_0} p_{i_0 i_1}^{(t_1)} p_{i_1 i_2}^{(t_2 - t_1)} \ldots p_{i_{n-1} i_n}^{(t_n - t_{n-1})},$$

  for some initial distribution $\{\nu_i\}_{i\in S}$ (with $\nu_i \geq 0$, and $\sum_{i\in S} \nu_i = 1$), and transition probabilities $\{p_{ij}^{(t)}\}_{\substack{i,j\in S\\ t\geq 0}}$ (with $p_{ij}^{(t)} \geq 0$, and $\sum_{j\in S} p_{ij}^{(t)} = 1$).
  - Just like for discrete-time chains, except need to keep track of elapsed time t too.
  - As with discrete chains, $p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$
- Let $P^{(t)} = \left(p_{ij}^{(t)}\right)_{i,j\in S}$ = matrix version.
  - Then $P^{(0)} = I$ = identity matrix.
  - Also $p_{ij}^{(s+t)} = \sum_{k\in S} p_{ik}^{(s)} p_{kj}^{(t)}$, i.e. $P^{(s+t)} = P^{(s)}\, P^{(t)}$. ("Chapman-Kolmogorov equations", just like for discrete time)
  - If $\mu_i^{(t)} = \mathbf{P}(X(t) = i)$, and $\mu^{(t)} = \left(\mu_i^{(t)}\right)_{i\in S}$ = row vector, and $\nu = (\nu_i)_{i\in S}$ = row vector, then $\mu_j^{(t)} = \sum_{i\in S} \nu_i p_{ij}^{(t)}$, and $\mu^{(t)} = \nu\, P^{(t)}$, and $\mu^{(t)}\, P^{(s)} = \mu^{(t+s)}$, etc.

- Expect that $\lim_{t\searrow 0} p_{ij}^{(t)} = p_{ij}^{(0)} = \delta_{ij}$.
  - Assume this is true. ("standard" Markov process)
  - Follows that for small enough h, $p_{ii}^{(h)} > 0$. Hence, by Chapman-Kolmogorov, $p_{ii}^{(nh)} > 0$ for all small enough $h > 0$ and all $n \in \mathbf{N}$. Hence, $p_{ii}^{(t)} > 0$ for all $t \geq 0$.
- Then can compute the process's generator as $g_{ij} = \lim_{t\searrow 0} \frac{p_{ij}^{(t)} - \delta_{ij}}{t} = p'_{ij}(0)$. (right-handed derivative)
  - So, if $G = \left(g_{ij} \right)_{i,j \in S}$ = matrix, then $G = P'^{(0)} = \lim_{t\searrow 0} \frac{P^{(t)} - I}{t}$. (right-handed derivative)
  - Here $g_{ii} \leq 0$, while $g_{ij} \geq 0$ for $i \neq j$.
  - In fact, usually (e.g. if S is finite), have

$$\sum_{j \in S} g_{ij} \;=\; \sum_{j \in S} \lim_{t\searrow 0} \frac{p_{ij}(t) - \delta_{ij}}{t} \;=\; \lim_{t\searrow 0} \frac{\sum_{j \in S} p_{ij}(t) - \sum_{j \in S} \delta_{ij}}{t} \;=\; \lim_{t\searrow 0} \frac{1-1}{t} \;=\; 0\,.$$

  - Furthermore, if $t > 0$ is small, then $G \approx \frac{P^{(t)} - I}{t}$, so $P^{(t)} \approx I + tG$, i.e. $p_{ij}^{(t)} \approx \delta_{ij} + t g_{ij}$.
- RUNNING EXAMPLE: $S = \{1, 2\}$, and $G = \begin{pmatrix} -3 & 3 \\ 6 & -6 \end{pmatrix}$.
  - Then for small $t > 0$, $P^{(t)} \approx I + t\,G = \begin{pmatrix} 1-3t & 3t \\ 6t & 1-6t \end{pmatrix}$.
  - So $p_{11}^{(t)} \approx 1 - 3t$, $p_{12}^{(t)} \approx 3t$, etc.
  - e.g. if $t = 0.02$, then $p_{11}^{(0.02)} \doteq 1 - 3(0.02) = 0.94$, $p_{12}^{(0.02)} \doteq 3(0.02) = 0.06$, $p_{21}^{(0.02)} \doteq 6(0.02) = 0.12$, and $p_{22}^{(0.02)} \doteq 1 - 6(0.02) = 0.88$, i.e. $P^{(0.02)} \;\doteq\; \begin{pmatrix} 0.94 & 0.06 \\ 0.12 & 0.88 \end{pmatrix}$.
- What about for larger t?
  - Well, by Chapman-Kolmogorov eqn, for any $m \in \mathbf{N}$,

$$P^{(t)} \;=\; [P^{(t/m)}]^m \;=\; \lim_{n\to\infty} [P^{(t/n)}]^n \;=\; \lim_{n\to\infty} [I + (t/n)G]^n$$
$$=\; \exp(tG) \;:=\; I + tG + \frac{t^2 G^2}{2!} + \frac{t^3 G^3}{3!} + \ldots$$

    (matrix equation; similar to how $\lim_{n\to\infty}(1 + \frac{c}{n})^n = e^c$).
  - (Makes sense so that e.g. $P^{(s+t)} = \exp((s+t)G) = \exp(sG)\,\exp(tG) = P^{(s)}\,P^{(t)}$, etc.)
  - So, in principle, the generator G tells us $P^{(t)}$ for all $t \geq 0$.
- Can we actually compute $P^{(t)} = \exp(tG)$ this way? Yes!
- Method #1: Compute the infinite matrix sum on a computer, numerically and approximately.
- Method #2: Note that in above example, if $\lambda_1 = 0$ and $\lambda_2 = -9$, and $w_1 = (2, 1)$ and $w_2 = (1, -1)$, then $w_1 G = \lambda_1 w_1 = 0$, and $w_2 G = \lambda_2 w_2 = -9 w_2$. That is, $\{\lambda_i\}$ are the eigenvalues of G, with corresponding left-eigenvectors $\{w_i\}$.
  - Now, if w_i is a left-eigenvector with corresponding eigenvalue λ_i, then $w_i \exp(tG) = e^{t\lambda_i} w_i$. (Check.) Easy!

- So, if initial distribution is (say) $\nu = (1,0)$, then first compute that $\nu = \frac{1}{3}w_1 + \frac{1}{3}w_2$. Then,

$$\mu^{(t)} \;=\; \nu\, P^{(t)} \;=\; \nu\, \exp(tG) \;=\; (\frac{1}{3}w_1 + \frac{1}{3}w_2)\, \exp(tG)$$

$$=\; \frac{1}{3}e^{t\lambda_1}w_1 + \frac{1}{3}e^{t\lambda_2}w_2 \;=\; \frac{1}{3}e^{0\,t}(2,1) + \frac{1}{3}e^{-9t}(1,-1) \;=\; (\frac{2+e^{-9t}}{3},\; \frac{1-e^{-9t}}{3})\,.$$

- So, $\mathbf{P}[X_t = 1] = p_{11}^{(t)} = \frac{2+e^{-9t}}{3}$, and $\mathbf{P}[X_t = 2] = p_{12}^{(t)} = \frac{1-e^{-9t}}{3}$.
- Check: $p_{11}^{(0)} = 1$, $p_{12}^{(0)} = 0$, and $p_{11}^{(t)} + p_{12}^{(t)} = 1$. (Phew.)
- (Or, by instead choosing $\nu = (0,1)$, could compute $p_{21}^{(t)}$ and $p_{22}^{(t)}$.)

- Method #3: Note that

$$p'^{(t)}_{ij} \;=\; \lim_{h\searrow 0}\frac{p_{ij}^{(t+h)} - p_{ij}^{(t)}}{h} \;=\; \lim_{h\searrow 0}\frac{(\sum_{k\in S} p_{ik}^{(t)}p_{kj}^{(h)}) - p_{ij}^{(t)}}{h}$$

$$=\; \lim_{h\searrow 0}\frac{(\sum_{k\in S} p_{ik}^{(t)}[\delta_{kj} + hg_{kj}]) - p_{ij}^{(t)}}{h}$$

$$=\; \lim_{h\searrow 0}\frac{(p_{ij}^{(t)} + h\sum_{k\in S} p_{ik}^{(t)}g_{kj}) - p_{ij}^{(t)}}{h} \;=\; \sum_{k\in S} p_{ik}^{(t)}g_{kj}\,,$$

i.e. $P'^{(t)} = P^{(t)}\,G$. ("forward equations")

- (Makes sense since $P^{(t)} = \exp(tG)$, so $P'^{(t)} = \exp(tG)\,G = P^{(t)}\,G$.)
- So, in above example,

$$p'^{(t)}_{11} \;=\; p_{11}^{(t)}\,g_{11} + p_{12}^{(t)}\,g_{21} \;=\; (-3)p_{11}^{(t)} + (6)p_{12}^{(t)} \;=\; (-3)p_{11}^{(t)} + (6)(1 - p_{11}^{(t)})$$

$$=\; (-9)p_{11}^{(t)} + 6 \;=\; (-9)(p_{11}^{(t)} - \frac{2}{3})\,.$$

END OF WEEK #11

- But $p'^{(t)}_{ij} = \frac{d}{dt}(p_{11}^{(t)}) = \frac{d}{dt}(p_{11}^{(t)} - \frac{2}{3})$.
- So, $\frac{d}{dt}(p_{11}^{(t)} - \frac{2}{3}) = (-9)(p_{11}^{(t)} - \frac{2}{3})$.
- So, $p_{11}^{(t)} - \frac{2}{3} = Ke^{-9t}$, i.e. $p_{11}^{(t)} = \frac{2}{3} + Ke^{-9t}$.
- But $p_{11}^{(0)} = 1$, so $K = \frac{1}{3}$, so $p_{11}^{(t)} = \frac{2}{3} + \frac{1}{3}e^{-9t} = \frac{2+e^{-9t}}{3}$.
- And then $p_{12}^{(t)} = 1 - p_{11}^{(t)} = \frac{1-e^{-9t}}{3}$.
- Same answers as before. (Phew.)

- What about LIMITING PROBABILITIES?
- In above example, $\mu^{(t)} = (\frac{2+e^{-9t}}{3},\; \frac{1-e^{-9t}}{3})$, so $\lim_{t\to\infty}\mu^{(t)} = (\frac{2}{3},\; \frac{1}{3}) =: \pi$.
  - Note that $\sum_{i\in S}\pi_i g_{i1} = \frac{2}{3}(-3) + \frac{1}{3}(6) = 0$, and $\sum_{i\in S}\pi_i g_{i2} = \frac{2}{3}(3) + \frac{1}{3}(-6) = 0$.
  - i.e., $\sum_{i\in S}\pi_i g_{ij} = 0$ for all $j \in S$, i.e. $\pi G = 0$.

- Does this make sense?
  - Well, as in discrete case, $\{\pi_i\}$ should be stationary.
  - i.e. $\sum_{i \in S} \pi_i p_{ij}^{(t)} = \pi_j$ for all $j \in S$ and all $t \geq 0$.
  - In particular, for small $t > 0$,

$$\pi_j \;=\; \sum_{i \in S} \pi_i p_{ij}^{(t)} \;\approx\; \sum_{i \in S} \pi_i [\delta_{ij} + t g_{ij}] \;=\; \pi_j + t \sum_{i \in S} \pi_i g_{ij} \,.$$

  - So, need $\sum_{i \in S} \pi_i g_{ij} = 0$ for all $j \in S$.
  - So, can check if $\{\pi_i\}$ is stationary by checking if $\sum_{i \in S} \pi_i g_{ij} = 0$ for all $j \in S$.
- What about reversibility?
  - Well, if $\pi_i \, g_{ij} = \pi_j \, g_{ji}$ for all $i, j \in S$, then $\sum_i \pi_i \, g_{ij} = \sum_i \pi_j \, g_{ji} = \pi_j \sum_i g_{ji} = \pi_j \cdot 0 = 0$, so π is stationary.
  - So, again, reversibility (in the above sense) implies stationary!
  - In above example, $\pi_1 g_{12} = (2/3)(3) = 2$, while $\pi_2 g_{21} = (1/3)(6) = 2$, so it's reversible.
- Is convergence to $\{\pi_i\}$ guaranteed?
- CONTINUOUS-TIME MARKOV CONVERGENCE THEOREM: If a continuous-time M.C. is irreducible, and has a stationary distribution π, then $\lim_{t \to \infty} p_{ij}^{(t)} = \pi_j$ for all $i, j \in S$.
  - (Like discrete case, but don't need aperiodicity, since in continuous time we always have $p_{ii}^{(t)} > 0$, so it is automatically aperiodic.)
  - Proof: For any $h > 0$, $P^{(h)} \equiv \{p_{ij}^{(h)}\}$ is irreducible and aperiodic discrete-time chain with stationary distribution π, so $\lim_{n \to \infty} p_{ij}^{(hn)} = \pi_j$. Result follows (formally, using monotonicity of distance to π).
  - See e.g. Durrett, 2nd ed., Theorem 4.4, p. 128.
- CONNECTION TO DISCRETE-TIME MARKOV CHAINS:
  - Let $\{\hat{p}_{ij}\}_{i,j \in S}$ be the transition probabilities for a discrete-time Markov chain $\{\hat{X}_n\}_{n=0}^{\infty}$.
  - Let $\{N(t)\}_{t \geq 0}$ be a Poisson process with intensity $\lambda > 0$.
  - Then let $X_t = \hat{X}_{N(t)}$.
  - Then $\{X_t\}$ is just like $\{\hat{X}_n\}$ except that it jumps at Poisson process event times, not integer times. ("Exponential holding times")
  - In particular, $\{X_t\}$ is a continuous-time Markov process!
  - That is, we can "create" a continuous-time Markov process from a discrete-time Markov chain.
- What is the generator of this Markov process $\{X_t\}$?
  - Well, here $p_{ij}^{(t)} = \mathbf{P}_i[\hat{X}_{N(t)} = j] = \sum_{n=0}^{\infty} \mathbf{P}_i[N(t) = n, \; \hat{X}_n = j]$
    $= \sum_{n=0}^{\infty} \mathbf{P}[N(t) = n] \, \hat{p}_{ij}^{(n)} = \sum_{n=0}^{\infty} [e^{-\lambda t} \frac{(\lambda t)^n}{n!}] \, \hat{p}_{ij}^{(n)}$.

- So, as $t \searrow 0$,

$$
\begin{aligned}
p_{ij}^{(t)} &= \sum_{n=0}^{\infty} \mathbf{P}[N(t)=n]\,\hat{p}_{ij}^{(n)} \\
&= \mathbf{P}[N(t)=0]\,\hat{p}_{ij}^{(0)} + \mathbf{P}[N(t)=1]\,\hat{p}_{ij}^{(1)} + \mathbf{P}[N(t)=2]\,\hat{p}_{ij}^{(2)} + \ldots \\
&= \mathbf{P}[N(t)=0]\,(\delta_{ij}) + \mathbf{P}[N(t)=1]\,\hat{p}_{ij} + \mathbf{P}[N(t)=2]\,\hat{p}_{ij}^{(2)} + \ldots \\
&= [e^{-\lambda t}(\lambda t)^0/0!]\delta_{ij} + [e^{-\lambda t}(\lambda t)^1/1!]\,\hat{p}_{ij} + O(t^2) \\
&= (1-\lambda t)\delta_{ij} + (1-\lambda t)(\lambda t)\hat{p}_{ij} + O(t^2) \;=\; \delta_{ij} + t[\lambda(\hat{p}_{ij}-\delta_{ij})] + O(t^2)\,.
\end{aligned}
$$

- But as $t \searrow 0$, $p_{ij}^{(t)} = \delta_{ij} + t g_{ij} + O(t^2)$.
- Hence, $t g_{ij} = t[\lambda(\hat{p}_{ij} - \delta_{ij})]$, so $g_{ij} = \lambda\,(\hat{p}_{ij} - \delta_{ij})$.
- That is, $g_{ii} = \lambda\,(\hat{p}_{ii} - 1)$, and for $i \neq j$, so $g_{ij} = \lambda\,\hat{p}_{ij}$.
- Check: for $i \neq j$, $g_{ij} \geq 0$, and $g_{ii} \leq 0$. Good.
- Also, $\sum_{j \in S} g_{ij} = g_{ii} + \sum_{j \neq i} g_{ij} = \lambda(\hat{p}_{ii} - 1) + \sum_{j \neq i}(\lambda \hat{p}_{ij}) = -\lambda + \sum_{j \in S}(\lambda \hat{p}_{ij}) = -\lambda + \lambda \sum_{j \in S} \hat{p}_{ij} = -\lambda + \lambda(1) = 0$, as it must.

- SPECIAL CASE: if $\hat{X}_0 = 0$, and $\hat{p}_{i,i+1} = 1$ for all i, then $\hat{X}_n = n$ for all n, so $X_t = \hat{X}_{N(t)} = N(t) =$ Poisson process.
  - And, the Poisson process $\{N(t)\}_{t \geq 0}$ itself has generator (check):

$$
G = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \ldots \\ 0 & -\lambda & \lambda & 0 & \ldots \\ 0 & 0 & -\lambda & \lambda & \ldots \end{pmatrix}.
$$

Application – Queueing Theory:

- Consider a queue (i.e., a line of customers) with just one server.
  - Let T_n = time of arrival of n^{th} customer. (And set $T_0 = 0$.)
  - Let $Y_n = T_n - T_{n-1}$ = inter-arrival time between $(n-1)^{\text{st}}$ and n^{th} customers.
  - Let S_n = time it takes to serve the n^{th} customer.
  - Let $Q(t)$ = number of customers in the system (i.e., waiting in the queue or being served) at time $t \geq 0$. (Assume $Q(0) = 0$.)
- What happens as $t \to \infty$?
- M/M/1 QUEUE: $T_n - T_{n-1} \sim \text{Exp}(\lambda)$, and $S_n \sim \text{Exp}(\mu)$, all indep., $\lambda, \mu > 0$. (So $\{T_n\}$ are arrival times of a Poisson process with intensity λ.)
  - Then by memoryless property, $\{Q(t)\}$ is a Markov process!
- GENERATOR?
  - Well, let $r_{a,b,t} = \mathbf{P}[a$ arrivals and b served by time $t]$.
  - Then $r_{a,b,t} = \left[e^{-\lambda t}(\lambda t)^a/a!\right]\left[e^{-\mu t}(\mu t)^b/b!\right] = e^{-(\lambda+\mu)t}\lambda^a \mu^b t^{a+b}/a!b!$.

– Hence, if $a+b \geq 2$, then $r_{a,b,t} = O(t^2)$ as $t \searrow 0$.

– Also, $r_{1,0,t} = e^{-(\lambda+\mu)t}\lambda t$ and $r_{0,1,t} = e^{-(\lambda+\mu)t}\mu t$.

– Hence, for $n \geq 0$,

$$g_{n,n+1} \;=\; \lim_{t\searrow 0} \frac{\mathbf{P}[Q(t)=n+1 \,|\, Q(0)=n]}{t}$$

$$= \lim_{t\searrow 0} \frac{r_{1,0,t}+r_{2,1,t}+r_{3,2,t}+\ldots}{t} \;=\; \lim_{t\searrow 0} \frac{r_{1,0,t}+O(t^2)}{t} \;=\; \lim_{t\searrow 0} \frac{e^{-(\lambda+\mu)t}\lambda t+O(t^2)}{t} \;=\; \lambda\,.$$

– Similarly, for $n \geq 1$,

$$g_{n,n-1} \;=\; \lim_{t\searrow 0} \frac{\mathbf{P}[Q(t)=n-1 \,|\, Q(0)=n]}{t}$$

$$= \lim_{t\searrow 0} \frac{r_{0,1,t}+r_{1,2,t}+r_{2,3,t}+\ldots}{t} \;=\; \lim_{t\searrow 0} \frac{r_{0,1,t}+O(t^2)}{t} \;=\; \lim_{t\searrow 0} \frac{e^{-(\mu+\mu)t}\lambda t+O(t^2)}{t} \;=\; \mu\,.$$

– Also if $|n-m| \geq 2$ then $\mathbf{P}[Q(t)=m \,|\, Q(0)=n] = O(t^2)$, so $g_{n,m}=0$.

– But $\sum_{m=0}^{\infty} g_{n,m} = 0$, so the generator must be given by:

$$G \;=\; \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \ldots \\ \mu & -\lambda-\mu & \lambda & 0 & 0 & \ldots \\ 0 & \mu & -\lambda-\mu & \lambda & 0 & \ldots \\ 0 & 0 & \ddots & \ddots & \ddots & \end{pmatrix}$$

i.e. $g_{00} = -\lambda$ and $g_{nn} = -\lambda-\mu$ for $n \geq 1$.

– So we have solved for the queue generator matrix G.

- STATIONARY DISTRIBUTION $\{\pi_i\}$?

– Need $\sum_{i\in S} \pi_i g_{ij} = 0$ for all $j \in S$. (Or, can use <u>reversibility</u>: check.)

– $j=0$: $\pi_0(-\lambda)+\pi_1(\mu)=0$, so $\pi_1 = (\frac{\lambda}{\mu})\pi_0$.

– $j=1$: $\pi_0(\lambda)+\pi_1(-\lambda-\mu)+\pi_2(\mu)=0$,
so $\pi_2 = (\frac{\lambda}{-\mu})\pi_0 + (\frac{-\lambda-\mu}{-\mu})\pi_1 = (-\frac{\lambda}{\mu})\pi_0 + (1+\frac{\lambda}{\mu})(\frac{\lambda}{\mu})\pi_0 = (\frac{\lambda}{\mu})^2\pi_0$.

– Then by induction: $\pi_i = (\frac{\lambda}{\mu})^i\pi_0$, for $i=0,1,2,\ldots$

– So <u>if</u> $\lambda < \mu$, i.e. $\frac{1}{\mu} < \frac{1}{\lambda}$, i.e. $\mathbf{E}(S_n) < \mathbf{E}(T_n - T_{n-1})$, then since $\sum_i \pi_i = 1$,

$$\pi_0 \;=\; \frac{1}{\sum_{i=0}^{\infty}(\frac{\lambda}{\mu})^i} \;=\; 1-(\frac{\lambda}{\mu})\,,$$

and the stationary distribution is

$$\pi_i \;=\; (\frac{\lambda}{\mu})^i\,(1-\frac{\lambda}{\mu})\,, \qquad i=0,1,2,3,\ldots$$

(geometric distribution).

- Furthermore, since the process is clearly irreducible,

$$\lim_{n\to\infty} \mathbf{P}[Q(t)=i] \;=\; \pi_i \;=\; (\frac{\lambda}{\mu})^i\,(1-\frac{\lambda}{\mu})\,.$$

- By contrast, if $\lambda > \mu$, then $Q(t) \to \infty$ w.p. 1. (see below)
- If $\lambda = \mu$, then $Q(t) \to \infty$ in probability, but not w.p. 1. (see below)

General (G/G/1) Queue:

- What if we <u>don't</u> assume Exponential distributions, just that $\{T_n - T_{n-1}\}$ i.i.d., and $\{S_n\}$ i.i.d. (all indep.)?
- Then $Q(t)$ is not Markovian! Have to use "cruder" methods.
- Let D_n = time of <u>departure</u> of the n^{th} customer.
- Roughly speaking, $D_n = D_{n-1} + S_n$.
  - But no one served while queue is <u>empty</u>.
  - So, actually, $D_n = \max(T_n, D_{n-1}) + S_n$.
- Let $W_n = \max(0,\ D_{n-1} - T_n)$ = the amount of time that the n^{th} customer has to <u>wait</u>. (With $W_0 = 0$.)
- LINDLEY'S EQUATION: For $n \geq 1$, $W_n = \max(0,\ W_{n-1} + S_{n-1} - Y_n)$.
- PROOF:
  - The $(n-1)^{\text{st}}$ customer is in the system for a total time $W_{n-1} + S_{n-1}$.
  - But the n^{th} customer arrives a time Y_n <u>after</u> the $(n-1)^{\text{st}}$ customer.
  - If $W_{n-1} + S_{n-1} \leq Y_n$, then the n^{th} customer doesn't have to wait at all, so $W_n = 0$.
  - If $W_{n-1}+S_{n-1} \geq Y_n$, then W_n = [time the $(n-1)^{\text{st}}$ customer is in the system]$-$[amount of that time that the n^{th} customer was <u>not</u> present for] $= (W_{n-1} + S_{n-1}) - Y_n$, Q.E.D.
- LINDLEY'S COROLLARY: $W_n = \max_{0 \leq k \leq n} \sum_{i=k+1}^{n} (S_{i-1} - Y_i)$. (Here, if $k = n$, then the sum equals zero.)
  - PROOF #1. Write it out: $W_0 = 0$, $W_1 = \max(0, S_0 - Y_1)$,
    $W_2 = \max(0, W_1 + S_1 - Y_2) = \max(0, \max(0, S_0 - Y_1) + S_1 - Y_2) = \max(0,\ S_1 - Y_2,\ S_0 - Y_1 + S_1 - Y_2)$,
    $W_3 = \max(0, W_2 + S_2 - Y_3) = \max(0, \max(0,\ S_1 - Y_2,\ S_0 - Y_1 + S_1 - Y_2) + S_2 - Y_3) = \max(0, S_2 - Y_3,\ S_1 - Y_2 + S_2 - Y_3,\ S_0 - Y_1 + S_1 - Y_2 + S_2 - Y_3)$,
    etc., each corresponding to the claimed formula.
  - PROOF #2: Induction on n. When $n = 0$, both sides are zero. If n increases to $n+1$, then by Lindley's Equation, each possible value of the "max" gets $S_n - Y_{n+1}$ added to it. And the "max with zero" is covered by allowing for the possibility $k = n + 1$, Q.E.D.
- THEOREM: For a general (G/G/1) single-server queue:
  - (a) if $\mathbf{E}(Y_n) < \mathbf{E}(S_n)$, then $\lim_{n\to\infty} W_n = \infty$ w.p. 1.
    (Hence, also $\lim_{n\to\infty} W_n = \infty$ in probability, so for any $M < \infty$, $\lim_{n\to\infty} \mathbf{P}(W_n > M) = 1$.)
  - (b) if $\mathbf{E}(Y_n) > \mathbf{E}(S_n)$, then $\{W_n\}$ is "bounded in probability", i.e. for any $\epsilon > 0$ there is $M < \infty$ such that $\mathbf{P}(W_n > M) < \epsilon$ for all $n \in \mathbf{N}$.

- (c) if $\mathbf{E}(Y_n) = \mathbf{E}(S_n)$, and S_{n-1} and Y_n are not both constant (i.e., $\mathbf{Var}(S_{n-1} - Y_n) > 0$), then $W_n \to \infty$ in probability (but not w.p. 1). ("Borderline" case; similar to branching processes with $m = 1$, and to the absolute value of ssrw.)

- PROOF OF (a):
  - By Lindley's Equation, $W_{n+1} \geq W_n + S_n - Y_{n+1}$.
  - Here the sequence $\{S_n - Y_{n+1}\}$ is i.i.d., with mean > 0.
  - So, by the SLLN, $\liminf_{n\to\infty} \frac{W_n}{n} \geq \mathbf{E}(S_n - Y_{n+1}) > 0$, w.p. 1.
  - It follows that $\liminf_{n\to\infty} W_n \geq \infty$, w.p. 1, Q.E.D.
- PROOF OF (b):
  - By Lindley's Corollary, $\mathbf{P}(W_n > M) = \mathbf{P}\Big(\max_{0\leq k\leq n} \sum_{i=k+1}^{n} (S_{i-1} - Y_i) > M\Big)$.
  - But $\{S_{i-1} - Y_i\}$ are i.i.d., so this is equivalent to $\mathbf{P}(W_n > M) = \mathbf{P}\Big(\max_{0\leq k\leq n} \sum_{i=1}^{n-k} (S_{i-1} - Y_i) > M\Big)$.
  - This is the probability that i.i.d. partial sums with negative mean (since $\mathbf{E}(S_{i-1} - Y_i) < 0$) will ever be larger than M.
  - i.e., it is the probability that the maximum of a sequence of i.i.d. partial sums with negative mean will be larger than M.
  - But by the SLLN, i.i.d. partial sums with negative mean will eventually become negative, w.p. 1.
  - So, w.p. 1, only a finite number of the partial sums will have positive values.
  - So, w.p. 1, the maximum value of the partial sums will be finite.
  - So, as $M \to \infty$, the probability that the maximum value will be $> M$ must converge to zero.
  - So, for any $\epsilon > 0$, there is $M < \infty$ such that the probability that its maximum value is be $> M$ is $< \epsilon$, Q.E.D.
- PROOF OF (c):
  - Trickier! Omitted! For details see e.g. Grimmett & Stirzaker, 2nd ed., Theorem 11.5(4), pp. 432–435.

- REMINDER: FINAL EXAM Tues April 18, 9:00 am - 12:00 noon, in the Exam Centre (EX) room 300. BRING YOUR STUDENT CARD, and DO NOT SIT NEXT TO ANYONE THAT YOU KNOW.
- Good luck and best wishes! – J.R.

END OF WEEK #12