STA447/2006 (Stochastic Processes) Lecture Notes, Winter 2016

by Jeffrey Rosenthal, University of Toronto

(Last updated: April 7, 2016)

Note: These lecture notes will be posted on the STA447/2006 course web page for your convenience, and will be updated regularly. 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:

- Discuss course web page, outline, evaluation, etc. (www.probability.ca/sta447)
- Schedule: will take 15-minute break if you return promptly!
- Your background knowledge: STA347 last semester? previously? other?
- 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):
  - 1000 lily pads arranged in a circle. (diagram)
  - Frog starts at pad #1000.
  - 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 #1000 after 1 step) = 1/3.
  - $\mathbf{P}$(at pad #999 after 1 step) = 1/3.
  - $\mathbf{P}$(at pad #2 after 2 steps) = 1/9.
  - $\mathbf{P}$(at pad #999 after 2 steps) = 2/9.
  - etc.
- What happens in the long run?
  - What is $\mathbf{P}$(frog at pad #428 after 987 steps)?

– What is $\lim_{k\to\infty} \mathbf{P}(\text{frog at pad \#428 after } k \text{ steps})$?

– Will the frog necessarily eventually return to pad #1000?

– 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 1000 lily pads)

  – 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.)

  – 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 $\sum_i \nu_i = 1$.)

- In the frog example, $S = \{1, 2, 3, \ldots, 1000\}$, and

$$p_{ij} = \begin{cases} 1/3, & |j-i| \leq 1 \\ 1/3, & |j-i| = 999 \\ 0, & \text{otherwise} \end{cases}$$

  and $\nu_{1000} = 1$ (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:

  – $\mathbf{P}(X_0 = 1000) = 1$, $\mathbf{P}(X_0 = 972) = 0$, etc.

  – $\mathbf{P}(X_1 = 1) = 1/3$, $\mathbf{P}(X_1 = 1000) = 1/3$, $\mathbf{P}(X_2 = 2) = 1/9$, $\mathbf{P}(X_2 = 999) = 2/9$, 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/longrun)
- 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: Branching process. (e.g. amoebas, infected people)
  - Let ρ be any prob dist on $\{0, 1, 2, \ldots\}$, the "offspring distribution".
  - Let X_n be the size of a "population" at time n.
  - Each of the X_n items at time n has a random number of offspring which is i.i.d. $\sim \rho$. (diagram)
  - 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. $\sim \rho$.
  - Here $S = \{0, 1, 2, \ldots\}$.
  - p_{ij} is more complicated; in fact $p_{ij} = (\rho * \rho * \ldots * \rho)(j)$, a convolution of i copies of ρ. (In particular, $p_{00} = 1$.)
  - Will $X_n = 0$ for some n? etc.
- 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: simple discrete-time queue.
  - At each time n, one person (or internet packet or ...) gets "served", and Z_n new people arrive, where $\{Z_n\}$ are i.i.d. $\sim \rho$, with ρ an "arrival distribution" on $\{0, 1, 2, \ldots\}$.
  - Let $X_n = \#$ of people in the queue at time n.
  - Then $X_{n+1} = X_n - \min(1, X_n) + Z_n$.
  - Here $\{X_n\}$ is Markov chain, with $S = \{0, 1, 2, 3, \ldots\}$, and $p_{ij} = \rho(j - i + \min(1, i))$.
  - Important in many applications ...
- 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: $\mathbf{P}(X_0 = 1000,\ X_1 = 999,\ X_2 = 1000) = \nu_{1000} p_{1000,999} p_{999,1000} = (1)(1/3)(1/3) = 1/9$, 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, $\mu_{999}^{(2)} = \nu_{1000} p_{1000,999} p_{999,999} + \nu_{1000} p_{1000,1000} p_{1000,999} + 0 = (1)(1/3)(1/3) + (1)(1/3)(1/3) + 0 = 2/9$.
- n-step transitions: $p_{ij}^{(n)} = \mathbf{P}(X_{m+n} = j \,|\, X_m = i)$.
  - (Again, doesn't depend on m: time-homogeneous.)
  - $p_{ij}^{(1)} = p_{ij}$. (of course)
  - What about $p_{ij}^{(2)}$?
  - Well, $p_{ij}^{(2)} = \mathbf{P}(X_2 = j \,|\, X_0 = i) = \sum\limits_{k \in S} \mathbf{P}(X_2 = j,\ X_1 = k \,|\, X_0 = i) = \sum\limits_{k \in S} p_{ik} p_{kj}$.

END OF WEEK #1

- 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: $p_{ij}^{(m+n)} = \mathbf{P}(X_{m+n} = j \,|\, X_0 = i) = \sum_{k \in S} \mathbf{P}(X_{m+n} = j,\ X_m = k \,|\, X_0 = i) = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}$.
  - 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 $\mathbf{P}(\cdots \,|\, X_0 = i)$. And, write $\mathbf{E}_i(\cdots)$ for $\mathbf{E}(\cdots \,|\, X_0 = i)$.
- 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 $n \geq 1) = 1$. Otherwise, i is <u>transient</u>. (Previous examples? Frog? s.r.w.?)
- Let $N(i) = \#\{n \geq 1 : X_n = i) =$ total # times the chain hits i. (Random variable; could be infinite.)
- 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 $f_{ij} = \mathbf{P}_i(X_n = j$ for <u>some</u> $n \geq 1)$.
- Then i recurrent <u>iff</u> $f_{ii} = 1$.
  - And, i transient <u>iff</u> $f_{ii} < 1$.
- 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,

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

Second, using countable linearity,

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

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

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

- 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!)
- 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?)
- 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")
- This follows immediately from:
- 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,

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

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

- 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

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

- Since S is finite, there must be at least one $j \in S$ with $\sum\limits_{n=1}^{\infty} p_{ij}^{(n)} = \infty$.
- So, we must be in case (a). *Q.E.D.*

- So, in Frog Example, $\mathbf{P}(\exists n \geq 1 \text{ with } X_n = 1000 \,|\, X_0 = 1000) \;=\; 1$.
  - But what about $\mathbf{P}(\exists n \geq 1 \text{ with } X_n = 428 \,|\, X_0 = 1000)$??

- 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.*

END OF WEEK #2

- 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) $\Rightarrow$ (3): Recurrence Theorem (with $i = j = k$).

- (3) $\Rightarrow$ (1): Recurrence Theorem (with $\ell = k$).
- (2) $\Rightarrow$ (4): Recurrence Theorem (with $i = j$).
- (4) $\Rightarrow$ (5): F-Lemma.
- (5) $\Rightarrow$ (3): Immediate.
- *Q.E.D.*

- Frog Example: $\mathbf{P}(\exists n \geq 1 \text{ with } X_n = 428 \,|\, X_0 = 1000) = 1$, 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!
  - 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 $f_{ij} = 1$.)

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 <u>stationary</u> 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 $\pi_i = \frac{1}{1000}$ for all $i \in S$.
  - Then $\pi_i \geq 0$ and $\sum_i \pi_i = 1$.
  - Also, for all $j \in S$, $\sum_{i\in S} \pi_i p_{ij} = \frac{1}{1000}(\frac{1}{3} + \frac{1}{3} + \frac{1}{3}) = \frac{1}{1000} = \pi_j$.
  - 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 $\pi_i = 1/|S|$ 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 <u>not</u> take $\pi_i = \frac{1}{d+1}$ for all i.
  - So, $\pi_i = ???$
- Defn: a Markov chain is <u>reversible</u> (or <u>time reversible</u>, or satisfies <u>detailed balance</u>) 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:
  - $\pi_i = 1/1000$
  - If $|j-i| \le 1$ or $|j-i| = 999$, then $\pi_i p_{ij} = (1/1000)(1/3) = \pi_j p_{ji}$.
  - 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 $\lim_{n\to\infty} \mathbf{P}[X_n = 1] = \lim_{n\to\infty} 1 = 1 \ne \frac{1}{2} = \pi_1$.

- 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 $\{n \ge 1; p_{ii}^{(n)} > 0\}$.
  - 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 $n \ge 1$ 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 $p_{ii}^{(r+s)} \ge p_{ij}^{(r)} p_{ji}^{(s)} > 0$, so t_i divides $r + s$.
  - Also if $p_{jj}^{(n)} > 0$, then $p_{ii}^{(r+n+s)} \ge p_{ij}^{(r)} p_{jj}^{(n)} p_{ji}^{(s)} > 0$, so t_i divides $r + n + s$, hence t_i 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.*
- ASIDE: THE (WEIERSTRASS) M-TEST (our version, anyway; also follows from the Dominated Convergence Theorem).
  - THM: If $\lim\limits_{n\to\infty} b_{nk} = a_k\ \forall k$, and $\sum\limits_{k=1}^{\infty} \sup_n |b_{nk}| < \infty$, then $\lim\limits_{n\to\infty} \sum\limits_{k=1}^{\infty} b_{nk} = \sum\limits_{k=1}^{\infty} a_k$.
  - PROOF:
  - Let $\epsilon > 0$.
  - Note that $a_k \leq \sup_n b_{nk}$, so $\sum_{k=1}^{\infty} \sup_n |b_{nk} - a_k| \leq 2 \sum_{k=1}^{\infty} \sup_n |b_{nk}| < \infty$.
  - So, can find $K \in \mathbf{N}$ such that $\sum_{k=K+1}^{\infty} \sup_n |b_{nk} - a_k| < \frac{\epsilon}{2}$.

- Then for $1 \le k \le K$, find N_k with $|b_{nk} - a_k| < \frac{\epsilon}{2K}$ for all $n \ge N_k$.
- Let $N = \max(N_1, \ldots, N_K)$.
- Then for $n \ge N$, $\left|\sum_{k=1}^{\infty} b_{nk} - \sum_{k=1}^{\infty} a_k\right| \le \sum_{k=1}^{\infty} |b_{nk} - a_k| < K\frac{\epsilon}{2K} + \frac{\epsilon}{2} = \epsilon$. *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 $n \ge n_0$.

– (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 $n \ge n_0(i)$.
- PROOF: Let $A = \{n \ge 1 : p_{ii}^{(n)} > 0\}$.
  - Then A is non-empty since $f_{ii} > 0$.
  - And, A is additive since $p_{ii}^{(m+n)} \ge p_{ii}^{(m)} p_{ii}^{(n)}$.
  - 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 $n \ge n_0(i,j)$.
- 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 $n \ge n_0(i,j)$, then $n - m \ge n_0(i)$, so $p_{ij}^{(n)} \ge p_{ii}^{(n-m)} p_{ij}^{(m)} > 0$. *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")

END OF WEEK #3

- 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 $n \ge \max\left[n_0(i,k),\, n_0(j,\ell)\right]$.
  - 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.)
- Hence, it follows that

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

  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, $\lim_{n\to\infty} \mathbf{P}(X_n = 428) = 1/1000$, 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.
- 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.

- What about periodic chains? (e.g. s.r.w., Ehrenfest)
- PERIODIC CONVERGENCE THM: Suppose chain irreducible, with period $b \geq 2$, and stat dist $\{\pi_i\}$. Then $\forall i, j \in S$, $\lim_{n\to\infty} \frac{1}{b}\,[p_{ij}^{(n)} + ... + p_{ij}^{(n+b-1)}] = \pi_j$, and also

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

  - (Note: still have $\lim_{n\to\infty} \frac{1}{b}\,[p_{ij}^{(n)} + ... + p_{ij}^{(n+b-1)}] = \pi_j$ for aperiodic chains, too.)
  - e.g. Ehrenfest's Urn: $b = 2$, so $\lim\limits_{n\to\infty} \frac{1}{2}\,\mathbf{P}[X_n = j \text{ or } X_{n+1} = j] = 2^{-d}\binom{d}{j}$.
- PROOF (outline only; optional):
  - Fix $i \in S$.
  - For $r = 0, 1, 2, \ldots, b-1$, let $S_r = \{j \in S : p_{ij}^{(bm+r)} > 0$ for some $m \in \mathbf{N}\}$.
  - Then $S = S_0 \dot\cup S_1 \dot\cup \ldots \dot\cup S_{b-1}$. (disjoint) (partition)
  - Furthermore $P^{(b)}$ is irreducible and aperiodic when restricted to S_0.
  - Also $\pi(S_0) = \pi(S_1) = \ldots = \pi(S_{b-1}) = 1/b$.
  - And, $\{b\,\pi_i\}_{i\in S_0}$ is stationary for $P^{(b)}$ when restricted to S_0.
  - Follows that $\lim_{n\to\infty} p_{ij}^{(bn)} = b\,\pi_j$ for all $j \in S_0$.
  - Then follows that $\lim_{n\to\infty} p_{ij}^{(bn+r)} = b\,\pi_j$ for all $j \in S_r$, for $0 \leq r \leq b-1$.
  - Hence, $\lim_{n\to\infty} \frac{1}{b}[p_{ij}^{(n)} + p_{ij}^{(n+1)} + \ldots + p_{ij}^{(n+b-1)}] = \frac{1}{b}[b\,\pi_j + 0] = \pi_j$ for any $j \in S$.
  - Then the second statement follows from the first using the M-test, just like in the main proof. *Q.E.D.*
- (e.g. for Ehrenfest's Urn, if $i = 0$, then $S_0 = \{\text{even } i \in S\}$, and $S_1 = \{\text{odd } i \in S\}$, and $\lim_{n\to\infty} \frac{1}{2}[p_{ij}^{(n)} + p_{ij}^{(n+1)}] = \pi_j = 2^{-d}\binom{d}{j}$.)
- COROLLARY: If Markov chain P is irreducible (not necessarily aperiodic), then it has <u>at most one</u> stationary distribution (just like before).
- What about simple random walk? Does it have a stationary dist?
  - No!
  - 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, if 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!
- [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.]

- 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 635,000 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)"
- 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?
- [Reminder: HW#1 due next class at 6:10 sharp!]

END OF WEEK #4

- 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 $1 \leq i \leq K - 1$, 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: $V = \{1, 2, \ldots, 1000\}$, with $w(i, i) = 1$ for $1 \leq i \leq 1000$, and $w(i, i+1) = w(i+1, i) = 1$ for $1 \leq i \leq 999$, and $w(1000, 1) = w(1, 1000) = 1$, 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 <u>finite</u> (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 $\pi_u \geq 0$ 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: 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!
  - 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 $1 \leq i \leq K-1$, 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 $2 \leq i \leq K-1$.
  - $\pi_i = \frac{1}{2K-2}$ for $i = 1, K$, and $\pi_i = \frac{2}{2K-2}$ for $2 \leq i \leq K-1$.
  - 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 $T_i = \inf\{n \geq 0 : X_n = i\}$ 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/gamone)

- 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 <u>function</u> of a.
  – Can we related the different unknown $s(a)$ to each other?
- Clearly $s(0) = 0$, and $s(c) = 1$.
- Furthermore, on the <u>first</u> bet, A either wins or loses \$1.
  – So, for $1 \le a \le c-1$,

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

$$= \mathbf{P}_a(T_c < T_0,\ X_1 = X_0 + 1) + \mathbf{P}_a(T_c < T_0,\ X_1 = X_0 - 1)$$

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

- 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 $1 \le a \le c-1$,

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

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

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

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

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

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

- 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 $p \to 1/2$?

Martingales:

- MOTIVATION: Gambler's ruin with $p = 1/2$.
  - 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".
  - Makes sense since $\mathbf{E}(X_{n+1} \,|\, X_n = i) = (1/2)(i+1) + (1/2)(i-1) = i$.
  - 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
$$\mathbf{E}(X_{n+1}) \;=\; \mathbf{E}\Big[\mathbf{E}(X_{n+1}\,|\,X_0, X_1, \ldots, X_n)\Big] \;=\; \mathbf{E}(X_n)\,,$$
  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?
- 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 $\mathbf{P}(T \leq M) = 1$), 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:

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

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

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

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

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

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

- 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:
  - Let $S = \min(T, n)$. Stopping time! Bounded!
  - Then by Optional Stopping Lemma, $\mathbf{E}(X_S) = \mathbf{E}(X_0)$ (for any n).
  - But $X_S = X_{\min(T,n)} = X_T - X_T\mathbf{1}_{T>n} + X_n\mathbf{1}_{T>n}$.
  - So, $X_T = X_S + X_T\mathbf{1}_{T>n} - X_n\mathbf{1}_{T>n}$.

- So, $\mathbf{E}(X_T) = \mathbf{E}(X_S) + \mathbf{E}(X_T \mathbf{1}_{T>n}) - \mathbf{E}(X_n \mathbf{1}_{T>n})$. (three terms to consider)
- First term $= \mathbf{E}(X_0)$ from above.
- Second term $\to 0$ as $n \to \infty$ by Dominated Convergence Thm (optional), since $\mathbf{E}|X_T| < \infty$ and $\mathbf{1}_{T>n} \to 0$ (since $\mathbf{P}(T < \infty) = 1$).
- Third term $\to 0$ as $n \to \infty$ 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, $|\mathbf{E}(X_n \mathbf{1}_{T>n})| \le \mathbf{E}(|X_n| \mathbf{1}_{T>n}) \le \mathbf{E}(M \mathbf{1}_{T>n}) = M\,\mathbf{P}(T > n)$, 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 $T = \inf\{n \ge 0 : X_n = 0$ 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 not 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 $\mathbf{E}(Y_{n+1} \,|\, Y_0, Y_1, \ldots, Y_n) = p\left[Y_n \left(\frac{1-p}{p}\right)\right] + (1-p)\left[Y_n \,/\, \left(\frac{1-p}{p}\right)\right] = Y_n(1-p) + Y_n(p) = Y_n$.
  - 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!)

Reminders: No class Feb 18 (Reading Week). Midterm on Feb 25 in HA401 [last name A-P] and HA410 [last name Q-Z]) – BRING STUDENT CARD.

END OF WEEK #5

(Midterm Test)

END OF WEEK #6

- 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 $T = \inf\{n \ge 0 : X_n = -5\}$.
  – Then $\mathbf{E}(X_T) = -5 \ne 0 = \mathbf{E}(X_0)$.
  – Contradiction?? No; it turns out that here $\mathbf{E}(T) = \infty$.
- PROOF: We compute (using that Z_i indep of $\{T \ge i\} = \{T \le i-1\}^C$) that

$$\begin{aligned}\mathbf{E}(X_T) - a \;&=\; \mathbf{E}(X_T - a) \;=\; \mathbf{E}(Z_1 + \ldots + Z_T) \\ &=\; \mathbf{E}\left[\sum_{i=1}^{T} Z_i\right] \;=\; \mathbf{E}\left[\sum_{i=1}^{\infty} Z_i\,\mathbf{1}_{T \ge i}\right] \;=\; \sum_{i=1}^{\infty} \mathbf{E}\left[Z_i\,\mathbf{1}_{T \ge i}\right] \\ &=\; \sum_{i=1}^{\infty} \mathbf{E}[Z_i]\,\mathbf{E}[\mathbf{1}_{T \ge i}] \;=\; m \sum_{i=1}^{\infty} \mathbf{P}[T \ge i] \;=\; m\,\mathbf{E}(T)\,, \quad Q.E.D.\end{aligned}$$

  – (Optional aside: the above calculation uses the Dominated Convergence Thm; indeed, setting $Y = \sum_{i=1}^{\infty} |Z_i|\,\mathbf{1}_{T \ge i}$, we have $\mathbf{E}(Y) = \mathbf{E}[\sum_{i=1}^{\infty} |Z_i|\,\mathbf{1}_{T \ge i}] = \sum_{i=1}^{\infty} \mathbf{E}[|Z_i|\,\mathbf{1}_{T \ge i}] = \sum_{i=1}^{\infty} \mathbf{E}|Z_i|\,\mathbf{E}[\mathbf{1}_{T \ge i}]$ $= \mathbf{E}|Z_1| \sum_{i=1}^{\infty} \mathbf{P}[T \ge i] = \mathbf{E}|Z_1|\,\mathbf{E}(T) < \infty$.)
- EXAMPLE: Gambler's Ruin with $p \ne 1/2$, and $T = \inf\{n \ge 0 : X_n = 0$ or $c\}$. (see e.g. www.probability.ca/gamone)
  – 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 $\mathbf{P}(T \ge cn) \le (1-p^c)^n$ so $\mathbf{E}(T) = \sum_{i=1}^{\infty} \mathbf{P}(T \ge i) \le \sum_{j=0}^{\infty} c\,\mathbf{P}(T \ge cj) \le \sum_{j=0}^{\infty} c\,(1-p^c)^j = c/[1-(1-p^c)] = c/p^c < \infty$.]

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

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

– 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 (optional)

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

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

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

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

- 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 $\mathbf{E}(S_M) \to \mathbf{E}(T)$ [optional: by Monotone Convergence Thm], and $\mathbf{E}[(X_{S_M} - a)^2] \to \mathbf{E}[(X_T - a)^2]$ [optional: by Bounded Convergence Thm, since for any n, $(X_{S_M} - a)^2 \le \max(a^2, (c-a)^2) < \infty$].
  – Hence, $\mathbf{E}(T) = \mathbf{E}[(X_T - a)^2] = \mathbf{Var}(X_T)$ (since $\mathbf{E}(X_T) = a$).

- But $\mathbf{Var}(X_T) = (a/c)\,(c-a)^2 + (1-a/c)a^2 = (a/c)(c^2+a^2-2ac) + (a^2-a^3/c) = ac + a^3/c - 2a^2 + a^2 - a^3/c = ac - a^2 = a(c-a)$, *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!)

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 <u>not</u> non-negative, does <u>not</u> converge.
- Example: s.s.r.w. stopped at zero – martingale, non-negative, converges to zero.
- Example: s.r.w. with $p = 2/3$ stopped at zero – non-negative, does not converge (might increase to infinity), but not a martingale.

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)
- 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 $S = \{0, 1, 2, \ldots\}$.
- $p_{00} = 1$.
- p_{ij} is more complicated; in fact (optional), $p_{ij} = (\mu * \mu * \ldots * \mu)(j)$, a convolution of i copies of μ.

- Will $X_n = 0$ for some n?
  - How can martingales help?
- Let $m = \sum_i i\,\mu\{i\}$ = 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, $\mathbf{E}(X_n) < \infty$ for all n.
- Let $Y_n = X_n/m^n$.
  - Then since $Y_n \leftrightarrow X_n$ is one-to-one function,

$$\mathbf{E}(Y_{n+1}\,|\,Y_0,\ldots,Y_n) \;=\; \mathbf{E}(\frac{X_{n+1}}{m^{n+1}}\,|\,Y_0,\ldots,Y_n)$$
$$=\; \mathbf{E}(\frac{X_{n+1}}{m^{n+1}}\,|\,X_0,\ldots,X_n) \;=\; \frac{m\,X_n}{m^{n+1}} \;=\; \frac{X_n}{m^n} \;=\; Y_n\,.$$

  - And, must have each $\mathbf{E}|Y_n| < \infty$ (since $\mathbf{E}|X_n| < \infty$ and $m > 0$).
  - Hence, $\{Y_n\}$ is martingale.
- So, $\mathbf{E}(Y_n) = \mathbf{E}(Y_0) = a$ for all n, i.e. $\mathbf{E}(X_n/m^n) = a$, so $\mathbf{E}(X_n) = a\,m^n$.
  - (This also follows from the "induction" above.)
- If $m < 1$, then $\mathbf{E}(X_n) = a\,m^n \to 0$.
  - But $\mathbf{E}(X_n) = \sum_{k=0}^{\infty} k\,\mathbf{P}(X_n = k) \ge \sum_{k=1}^{\infty} \mathbf{P}(X_n = k) = \mathbf{P}(X_n \ge 1)$.
  - Hence, $\mathbf{P}(X_n \ge 1) \le \mathbf{E}(X_n) = a\,m^n \to 0$, i.e. $\mathbf{P}(X_n = 0) \to 1$.
  - Certain extinction!
- If $m > 1$, then $\mathbf{E}(X_n) \to \infty$.
  - In this case, it turns out that $\mathbf{P}(X_n \to \infty) > 0$. ("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.
- But 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 <u>some</u> 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. <u>not</u> 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 $m = 1$!

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)
- 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 get errors of order $O(1/\sqrt{M})$, which don'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 get $O(1/\sqrt{M})$ errors).
  - 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 $0 \leq t_1 \leq s_1 \leq t_2 \leq s_2 \leq \ldots \leq t_k \leq s_k$, 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 independent. ("independent normal increments")
- Finally, if $0 < t \leq s$, 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: Brownian motion is a process $\{B_t\}_{t\geq 0}$ 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 intuitive, but a formal proof of existence requires measure theory.)
- Example: What is $\mathbf{E}[(B_2 + B_3 + 1)^2]$?
  - Well, $\mathbf{E}[(B_2 + B_3 + 1)^2] = \mathbf{E}[(B_2)^2)] + \mathbf{E}[(B_3)^2] + 1^2 + 2\mathbf{E}[B_2 B_3] +$

$2\mathbf{E}[B_2(1)] + 2\mathbf{E}[B_3(1)] = 2 + 3 + 1 + 2(2) + 2(0) + 2(0) = 10.$

- Example: What is $\mathbf{Var}[B_3 + B_5 + 7]$?
  - Well, $\mathbf{Var}[B_3 + B_5 + 7] = \mathbf{E}[(B_3 + B_5)^2] = \mathbf{E}[(B_3)^2)] + \mathbf{E}[(B_5)^2] + 2\mathbf{E}[B_3 B_5] = 3 + 5 + 2(3) = 14.$
- 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 given B_r for $0 \le r \le t$, what is the conditional distribution of B_s?
  - Similar to above, $B_s \,|\, B_t \;=\; B_t + (B_s - B_t) \,|\, B_t \;=\; B_t + \text{Normal}(0,\, s - t) \;\sim\; \text{Normal}(B_t,\, s-t)$. (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) martingale!
  - 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 $e \equiv \mathbf{E}(\tau)$?

END OF WEEK #7

- To continue, let $Y_t = B_t^2 - t$.
  - Then for $0 < t < s$, $\mathbf{E}[Y_s \,|\, \{B_r\}_{r \le t}] = \mathbf{E}[B_s^2 - s \,|\, \{B_r\}_{r \le t}]$
    $= \mathbf{Var}[B_s \,|\, \{B_r\}_{r \le t}] + (\mathbf{E}[B_s \,|\, \{B_r\}_{r \le t}])^2 - s$
    $= (B_t)^2 + (s - t) - s = Y_t$.
  - By the law of iterated expectations (optional; e.g. Rosenthal, Prop 13.2.7), $\mathbf{E}[Y_s \,|\, \{Y_r\}_{r \le t}] = \mathbf{E}\Big[\mathbf{E}[Y_s \,|\, \{B_r\}_{r \le t}] \,\Big|\, \{Y_r\}_{r \le t}\Big] = \mathbf{E}[Y_t \,|\, \{Y_r\}_{r \le t}] = Y_t$.
  - So, $\{Y_t\}$ is also a martingale!
- Back to $\tau = \min\{t \ge 0 : B_t = -a \text{ or } b\}$. What is $e \equiv \mathbf{E}(\tau)$?

- Well, with $Y_t = B_t^2 - t$, have $\mathbf{E}(Y_\tau) = \mathbf{E}(B_\tau^2 - \tau) = \mathbf{E}(B_\tau^2) - \mathbf{E}(\tau) = pb^2 + (1-p)(-a)^2 - e = \frac{a}{a+b}b^2 + \frac{b}{a+b}a^2 - e = ab - e$.
- <u>Assuming</u> $\mathbf{E}(Y_\tau) = 0$, 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 $t \geq 0$.
  - 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")
- More generally, could have $X_t = x_0 + \mu t + \sigma B_t$. (file "Rbrownian")
  - Then $dX_t = \mu\,dt + \sigma\,dB_t$. ($\mu$ = "drift"; σ = "volatility"; $\sigma \geq 0$)
  - Then $\mathbf{E}(X_t) = x_0 + \mu t$, and $\mathbf{Var}(X_t) = \sigma^2 t$, and $\mathbf{Cov}(X_t, X_s) = \sigma^2 \min(s, t)$.
  - Optional: Even more generally, could have $dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dB_t$, where μ and σ are <u>functions</u>, 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 <u>proportional</u> to total price (makes sense).
  - So, $Y_t = \log(X_t)$ is a <u>diffusion</u>.
- 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

$$D_t \equiv e^{-rt} X_t = e^{-rt} x_0 \exp(\mu t + \sigma B_t) = x_0 \exp((\mu - r)t + \sigma B_t)\,.$$

- Defn: A "European call option" is the <u>option</u> 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: this assumes the ability to buy/sell arbitrary amounts of stock at any time, infinitely often, including going negative (i.e., "shorting" the stock), with <u>no</u> 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!
- FACT: the fair price for the option equals $\mathbf{E}[e^{-rS} \max(0, X_S - K)]$, but only after <u>replacing</u> 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 $B_S \sim \text{Normal}(0, S)$.
  - WHY?? Well, if $\mu = r-\frac{\sigma^2}{2}$, then $\{D_t\}$ becomes a martingale (HW#3), and this <u>turns out</u> to be a key fact. (finance/actuarial classes ...)
- So, fair price is now just an <u>integral</u> (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 <u>not</u> have to memorise!]
- Note: this price does <u>not</u> depend on the drift ("appreciation rate") μ. [Surprising! Intuition: if μ large, then can make good money from stock, so don't need the option.]
- However, it is an <u>increasing</u> function of the volatility σ. [Makes sense.]

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!

- Suppose an irreducible Markov chain on a (discrete) state space S has a stationary distribution π. Then if the chain starts at some $i \in S$, how long until it returns to i, on average? That is, what is $m_i \equiv \mathbf{E}_i(T_i)$?
  - Well, whatever m_i is, by the SLLN, over the long run, the chain will spend a fraction $1/m_i$ of iterations at i: $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k=i} = 1/m_i$.
  - Hence, in particular, by the Bounded Convergence Theorem, $\lim_{n\to\infty} \mathbf{E}\Big(\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k=i}\Big) = 1/m_i$.
  - But in the long run, the chain must spend an expected fraction π_i of iterations at i. That is, $\lim_{n\to\infty} \mathbf{E}\Big(\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k=i}\Big) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbf{E}_i(\mathbf{1}_{X_k=i}) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbf{P}_i(X_k = i) = \lim_{n\to\infty} \mathbf{P}_i(X_n = i) = \lim_{n\to\infty} p_{ii} = \pi_i$.
  - So, we must have $1/m_i = \pi_i$, i.e. $m_i = 1/\pi_i$!
  - This is 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.)
  - Let's apply this to sequence waiting times!
- Let X_n be the amount of the desired sequence (HTH) that the chain has "achieved so far". (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(!) that the stationary distribution is $(0.3, 0.4, 0.2, 0.1)$.
  - So, mean time to return from state 3 to state 3 is $1/0.1 = 10$.
  - 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 $P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix}$.
  - 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 S_n be the total amount won by <u>all</u> the betters by time n.
  - Then $\{S_n\}$ is a martingale with stopping time τ.
  - Then have(!) that $S_\tau = -(\tau - 3) + (-1) + (1) + (7) = -\tau + 10$.
  - It follows (optional: by Dominated Convergence Thm, since $|S_n - S_{n-1}| \le 7$, and $\mathbf{E}(\tau) < \infty$) that $\mathbf{E}(S_\tau) = \mathbf{E}(S_0) = 0$.
  - Hence, $0 = \mathbf{E}(S_\tau) = -\mathbf{E}(\tau) + 10$, whence $\mathbf{E}(\tau) = 10$. Same as before!
- Similarly, for THH, get that $S_\tau = -(\tau - 3) + (-1) + (-1) + (7) = -\tau + 8$, whence $\mathbf{E}(\tau) = 8$. Same as before!

Poisson Processes:

- MOTIVATING EXAMPLE:
  - Suppose an average of $\lambda = 2.5$ fires in Toronto per day.
  - Intuitively, this is caused by a very <u>large</u> number n of buildings, each of which has a very <u>small</u> probability p of having a fire.
  - Then mean $= np = \lambda$, so $p = \lambda/n$.
  - Then # fires today is Binomial$(n, \lambda/n) \approx$ Poisson(λ) = Poisson(2.5).
  - [That is, $\mathbf{P}(\#\text{ fires} = k) \approx e^{-2.5}\frac{(2.5)^k}{k!}$, for $k = 0, 1, 2, 3, \ldots$]
  - And, # fires today and tomorrow combined $\approx$ Poisson$(2 * \lambda)$ = Poisson(5), etc.
  - Full distribution? $\mathbf{P}$(fire within next hour)? etc.
- Let $\{Y_n\}_{n=1}^{\infty}$ be i.i.d. $\sim$ Exp(λ), for some $\lambda > 0$.
  - So, Y_n has density function $\lambda e^{-\lambda y}$ for $y > 0$.
  - And, $\mathbf{P}(Y_n > y) = e^{-\lambda y}$ for $y > 0$.
  - And, $\mathbf{E}(Y_n) = 1/\lambda$.
- Let $T_0 = 0$, and $T_n = Y_1 + Y_2 + \ldots + Y_n$ for $n \ge 1$. ("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, $N(t) \sim \text{Poisson}(\lambda t)$.
  - Thus, $\mathbf{P}(N(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$ for $k = 0, 1, 2, \ldots$
  - Hence also $\mathbf{E}(N(t)) = \lambda t$, and $\mathbf{Var}(N(t)) = \lambda t$.

END OF WEEK #8

- Now, recall the "memoryless" (or "forgetfulness") property of the $\text{Exp}(\lambda)$ 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 $0 \leq t_1 \leq s_1 \leq t_2 \leq s_2 \leq \ldots \leq t_k \leq s_k$, then $N(s_i) - N(t_i) \sim \text{Poisson}(\lambda(s_i - t_i))$, and $\{N(s_i) - N(t_i)\}_{i=1}^k$ are all <u>independent</u>. ("independent Poisson increments")

- DEFN: A <u>Poisson processes</u> with intensity $\lambda > 0$ is a collection $\{N(t)\}_{t \geq 0}$ of random non-decreasing integer counts $N(t)$, satisfying: (a) $N(0) = 0$; (b) $N(t) \sim \text{Poisson}(\lambda t)$ for all $t \geq 0$; and (c) independent Poisson increments (as above).

- MOTIVATING EXAMPLE (cont'd): average of $\lambda = 2.5$ fires per day.
  - Here, fires approximately follow a Poisson process with intensity 2.5.
  - So, $\mathbf{P}$(9 fires today and tomorrow combined) $\approx e^{-2*2.5}\frac{(2*2.5)^9}{9!} = e^{-5}(\frac{5^9}{9!}) \doteq 0.036$.
  - $\mathbf{P}$(at least one fire in next hour) $= 1 - \mathbf{P}$(no fires in next hour) $= 1 - \mathbf{P}(N(1/24) = 0) = 1 - e^{-2.5/24}\frac{(2.5/24)^0}{0!} \doteq 1 - 0.90 = 0.10$.
  - $\mathbf{P}$(exactly 3 fires in next hour) $= e^{-2.5/24}\frac{(2.5/24)^3}{3!} \doteq 0.00017 \doteq 1/5891$, etc.
- EXAMPLE: Let $\{N(t)\}$ be a Poisson process with intensity $\lambda = 2$. Then

$$\mathbf{P}[N(3) = 5,\ N(3.5) = 9] \;=\; \mathbf{P}[N(3) = 5,\ N(3.5) - N(3) = 4]$$

$$= \mathbf{P}[N(3) = 5]\ \mathbf{P}[N(3.5) - N(3) = 4]$$

$$= \left[e^{-\lambda 3}\frac{(\lambda\, 3)^5}{5!}\right]\ \left[e^{-\lambda\, 0.5}\frac{(\lambda\, 0.5)^4}{4!}\right]$$

$$= \left(e^{-6}\,\frac{6^5}{120}\right)\ \left(e^{-1}\,\frac{1^4}{24}\right) \;=\; e^{-7}(2.7) \;\doteq\; 0.0025 \;\doteq\; 1/400\,.$$

- EXAMPLE: Let $\{N(t)\}$ be a Poisson process with intensity λ.
  - Then for $0 < t < s$,

$$\mathbf{P}(N(t) = 1 \,|\, N(s) = 1) \;=\; \frac{\mathbf{P}(N(t) = 1,\ N(s) = 1)}{\mathbf{P}(N(s) = 1)}$$

$$= \frac{\mathbf{P}(N(t) = 1,\ N(s) - N(t) = 0)}{\mathbf{P}(N(s) = 1)}$$

$$= \frac{e^{-\lambda t}\frac{(\lambda t)^1}{1!}\ e^{-\lambda(s-t)}\frac{(\lambda(s-t))^0}{0!}}{e^{-\lambda s}\frac{(\lambda s)^1}{1!}} \;=\; t/s\,.$$

  - That is, conditional on $N(s) = 1$, the first event is <u>uniform</u> 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
- 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!
- RELATED APPROACH:
  - Suppose have λ buses per hour, i.e. about n buses every n/λ hours.
  - Suppose the arrival times are completely <u>random</u>.
  - 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 $\sim$ 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 $\sim$ Exp(λ), 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)

• SUPERPOSITION: Suppose $\{N_1(t)\}_{t\geq 0}$ and $\{N_2(t)\}_{t\geq 0}$ are two independent Poisson processes, with rates λ_1 and λ_2 respectively. Let $N(t) = N_1(t) + N_2(t)$.
  - Then $\{N(t)\}_{t\geq 0}$ 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 $A \sim$ Exp(λ).
  - Hence, $\mathbf{E}(A) = 1/\lambda = 1/8$ hours, i.e. 7.5 minutes.

• THINNING: Let $\{N(t)\}_{t\geq 0}$ be a Poisson process with rate λ.
  - Suppose each arrival is independently of "type i" with probability p_i, for $i = 1, 2, 3, \ldots$. (e.g. bus or streetcar, male or female, undergrad or grad, etc.)
  - Let $N_i(t)$ be number of arrivals of type i up to time t.
  - THM: The $\{N_i(t)\}$ are 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)
$= \Big(e^{-\lambda t}(\lambda t)^{j+k}/(j+k)!\Big) \times \binom{j+k}{j}(p_1)^j(p_2)^k$. Equal! (Check.)

- EXAMPLE: If students arrive for office hours according to a Poisson process, and each student is independently either undergrad (prob p_1) or grad (prob p_2), then # undergrads is independent of # grads (and each follows a Poisson distribution).
- ASIDE: Can also have time-inhomogeneous Poisson processes, where $\lambda = \lambda(t)$, and $N(b) - N(a) \sim \text{Poisson}\Big(\int_a^b \lambda(t)\,dt\Big)$.
- 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 $\{B_t\}_{t\ge 0}$, and Poisson processes $\{N(t)\}_{t\ge 0}$, 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 $\{X(t)\}_{t\ge 0}$ 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 $\nu_i \ge 0$, and $\sum_{i\in S} \nu_i = 1$), and transition probabilities $\{p_{ij}^{(t)}\}_{\substack{i,j\in S\\ t\ge 0}}$ (with $p_{ij}^{(t)} \ge 0$, and $\sum_{j\in S} p_{ij}^{(t)} = 1$).
  – Just like for discrete-time chains, except need to keep track of the elapsed time (t) too.
  – As with discrete chains, $p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \ne j \end{cases}$
- Let $P^{(t)} = \Big(p_{ij}^{(t)}\Big)_{i,j\in S}$ = 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 $\mu^{(t)} = \Big(\mu_i^{(t)}\Big)_{i\in S}$ = row vector, and $\nu = (\,\nu_i\,)_{i\in S}$ = row vector, then $\mu_j^{(t)} = \sum\limits_{i\in S} \nu_i p_{ij}^{(t)}$, 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)
- 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 $G = (\, g_{ij}\,)_{i,j \in S}$ = 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 $p_{22}^{(t)}$.)

END OF WEEK #9

- Method #3: Note that

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

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

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

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,

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

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

– 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, $\sum_{i \in S} \pi_i g_{ij} = 0$.
  - 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 (i.e., in continuous time, it is automatically aperiodic).
  - Proof omitted here; similar to discrete time.
  - 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 $p_{ij}^{(t)} = \mathbf{P}_i[\hat{X}_{N(t)} = j]$.

- So, $p_{ij}^{(t)} = \sum_{n=0}^{\infty} \mathbf{P}_i[N(t) = n,\ \hat{X}_n = j]$.
- So, $p_{ij}^{(t)} = \sum_{n=0}^{\infty} \mathbf{P}[N(t) = n]\, \hat{p}_{ij}^{(n)} = \sum_{n=0}^{\infty} [e^{-\lambda t} \frac{(\lambda t)^n}{n!}]\, \hat{p}_{ij}^{(n)}$.
- But for small $t > 0$, $\mathbf{P}[N(t) = n] = e^{-\lambda t}(\lambda t)^n / n! \approx t^n \lambda^n / n!$.
- So, for small $t > 0$ and $i \neq j$,

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

  to first order in t, as $t \searrow 0$.
- But $p_{ij}^{(t)} \approx \delta_{ij} + t g_{ij} = t g_{ij}$, so $t g_{ij} = [t\lambda]\hat{p}_{ij}$, so $g_{ij} = \lambda\, \hat{p}_{ij}$.
- Also $p_{ii}^{(t)} \approx \mathbf{P}[N(t) = 0] + \mathbf{P}[N(t) = 1]\, \hat{p}_{ii} + O(t^2) \approx [1 - t\lambda] + [t\lambda]\hat{p}_{ii}$.
- But $p_{ii}^{(t)} \approx \delta_{ii} + t g_{ii} = 1 + t g_{ii}$, so $[1 - t\lambda] + [t\lambda]\hat{p}_{ii} = 1 + t g_{ii}$, so $g_{ii} = \lambda\, (\hat{p}_{ii} - 1)$.
- 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, for $n \geq 0$, $\mathbf{P}[Q(t) = n+1 \,|\, Q(0) = n]$
    $= \mathbf{P}$[one arrival and zero served by time t] $+ \mathbf{P}$[two arrivals and one served by time t] $+ \ldots$
    $= \mathbf{P}$[one arrival and zero served by time t] $+ O(t^2)$
    $\approx \mathbf{P}$[one arrival and zero served by time t].
  - Hence,, for $n \geq 0$, to first order as $t \searrow 0$,

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

$$= \lim_{t\searrow 0} \frac{\mathbf{P}[\text{one arrival and zero served by time } t]}{t}$$

$$= \lim_{t\searrow 0} \frac{[e^{-\lambda t}\frac{(\lambda t)^1}{1!}]\,[e^{-\mu t}\frac{(\mu t)^0}{0!}]}{t} = \lim_{t\searrow 0} e^{-\lambda t}\lambda e^{-\mu t} = \lambda\,.$$

  - Similarly, for $n \geq 1$,

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

$$= \lim_{t\searrow 0} \frac{\mathbf{P}[\text{zero arrivals and one served by time } t]}{t}$$

$$= \lim_{t\searrow 0} \frac{[e^{-\lambda t}\frac{(\lambda t)^0}{0!}]\,[e^{-\mu t}\frac{(\mu t)^1}{1!}]}{t} = \mu\,.$$

  - 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 $n \geq 1$. (This corresponds to zero arrivals and zero served by time t; check.)
  - 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 if $\lambda < \mu$, 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 don't 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 departure of the n^{th} customer.
- Roughly speaking, $D_n = D_{n-1} + S_n$.
  – But no one served while queue is empty.
  – 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 wait. (With $W_0 = 0$.)
- LINDLEY'S EQUATION: For $n \ge 1$, $W_n = \max(0,\; W_{n-1} + S_{n-1} - Y_n)$.
- 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 after the $(n-1)^{\text{st}}$ customer.
  – If $W_{n-1} + S_{n-1} \le Y_n$, then the n^{th} customer doesn't have to wait at all, so $W_n = 0$.
  – If $W_{n-1} + S_{n-1} \ge Y_n$, then W_n = [time the $(n-1)^{\text{st}}$ customer is in the system] − [amount of that time that the n^{th} customer was not present for] $= (W_{n-1} + S_{n-1}) - Y_n$, Q.E.D.
- LINDLEY'S COROLLARY: $W_n \;=\; \max_{0 \le k \le n} \sum_{i=k+1}^{n} (S_{i-1} - Y_i)$. (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 <u>not</u> w.p. 1). ("Borderline" case; similar to branching processes with $m = 1$.)

- PROOF OF (a):
  - By Lindley's Equation, $W_{n+1} \ge W_n + S_n - Y_{n+1}$.
  - Here the sequence $\{S_n - Y_{n+1}\}$ is i.i.d., with mean > 0.
  - So, by the SLLN, $\liminf_{n\to\infty} \frac{W_n}{n} \ge \mathbf{E}(S_n - Y_{n+1}) > 0$, w.p. 1.
  - It follows that $\liminf_{n\to\infty} W_n \ge \infty$, w.p. 1, Q.E.D.

- PROOF OF (b):
  - By Lindley's Corollary,
    $\mathbf{P}(W_n > M) = \mathbf{P}\Big(\max_{0\le k\le n} \sum_{i=k+1}^{n} (S_{i-1} - Y_i) > M\Big)$.
  - But $\{S_{i-1} - Y_i\}$ are i.i.d., so this is equivalent to
    $\mathbf{P}(W_n > M) = \mathbf{P}\Big(\max_{0\le k\le n} \sum_{i=1}^{n-k} (S_{i-1} - Y_i) > M\Big)$.
  - 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 <u>maximum</u> 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 <u>negative</u>, w.p. 1.
  - So, w.p. 1, only a <u>finite</u> number of the partial sums will have positive values.
  - So, w.p. 1, the maximum value of the partial sums will be <u>finite</u>.
  - 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.

Application: Quantum Mechanics (simplified view!):

- According to quantum mechanics, the universe on a fundamental level behaves according to probabilities(!).
  - More specifically, it has a complex-valued wave function Ψ which evolves according to various rules, and whose squared absolute value $|\Psi|^2$ gives the probability of observing a given state.
- Here is a very simplified example ...
- Suppose an elementary particle has (just) two "eigenstates", E_1 and E_2, with "energies" $\lambda_1 \neq \lambda_2$, and an "observable state" $B = \frac{1}{\sqrt{2}}E_1 + \frac{1}{\sqrt{2}}E_2$.
  - State of system is described by a complex-valued "wave function" Ψ_t.
  - Then, the projection of the wave function onto each of the two eigenstates E_1 and E_2 evolves by the formulas: $\langle \Psi_t \,|\, E_1 \rangle = \langle \Psi_0 \,|\, E_1 \rangle \, e^{i\lambda_1 t}$, and $\langle \Psi_t \,|\, E_2 \rangle = \langle \Psi_0 \,|\, E_2 \rangle \, e^{i\lambda_2 t}$, where $i = \sqrt{-1}$.
  - (This is a special case of the "Schrödinger Wave Equation".)
  - Then, by linearity, the projection of the wave function onto the observable state B is given by: $\langle \Psi_t \,|\, B \rangle = \frac{1}{\sqrt{2}}\langle \Psi_t \,|\, E_1 \rangle + \frac{1}{\sqrt{2}}\langle \Psi_t \,|\, E_2 \rangle$. ("superposition")
- Now, suppose we "observe" the system at some time t.
  - Then, $\mathbf{P}$(observe state B at time t)
    $= |\langle \Psi_t \,|\, B \rangle|^2 = |\frac{1}{\sqrt{2}}\langle \Psi_t \,|\, E_1 \rangle + \frac{1}{\sqrt{2}}\langle \Psi_t \,|\, E_2 \rangle|^2$.
- QUESTION: Suppose the system is in state B at time 0, so $\langle \Psi_0 \,|\, E_1 \rangle = \langle \Psi_0 \,|\, E_2 \rangle = \frac{1}{\sqrt{2}}$. Suppose we then observe the system at some later time $t > 0$. Then what is $\mathbf{P}$(observe state B at time t)?

END OF WEEK #10

- SOLUTION:
  - We know that $\langle \Psi_0 \,|\, E_1 \rangle = \langle \Psi_0 \,|\, E_2 \rangle = \frac{1}{\sqrt{2}}$.
  - So, at time t, $\langle \Psi_t \,|\, E_j \rangle = \langle \Psi_0 \,|\, E_j \rangle \, e^{i\lambda_j t} = \frac{1}{\sqrt{2}} \, e^{i\lambda_j t}$.
  - $\mathbf{P}$(observe state B at time t) $= |\langle \Psi_t \,|\, B \rangle|^2 = |\frac{1}{\sqrt{2}}\langle \Psi_t \,|\, E_1 \rangle + \frac{1}{\sqrt{2}}\langle \Psi_t \,|\, E_2 \rangle|^2$.
  - This is the absolute square of a complex number.
  - We can compute it, using rules of complex numbers. e.g., for $a, b \in \mathbf{R}$, $|a + ib|^2 = a^2 + b^2$, and $e^{ia} = \cos(a) + i\sin(a)$, and $|e^{ia}| = 1$.

- Compute that $\mathbf{P}(\text{observe state } B \text{ at time } t) = |\frac{1}{2}\, e^{i\lambda_1 t} + \frac{1}{2}\, e^{i\lambda_2 t}|^2 = \frac{1}{4}|e^{i\lambda_1 t}|\, |1+e^{i(\lambda_2-\lambda_1)t}|^2 = \frac{1}{4}(1)\, |1+\cos((\lambda_2-\lambda_1)t)+i\sin((\lambda_2-\lambda_1)t)|^2 = \frac{1}{4}\, [(1+\cos((\lambda_2-\lambda_1)t))^2 + (\sin((\lambda_2-\lambda_1)t))^2]$
  $= \frac{1}{4}\, [1+1+2\cos((\lambda_2-\lambda_1)t))] = \frac{1}{2}[1+\cos((\lambda_2-\lambda_1)t))]$.

- That is, the probability fluctuates between 1 (at times 0, $2\pi/(\lambda_2-\lambda_1)$, etc.), and 0 (at times $\pi/(\lambda_2-\lambda_1)$, etc.). (Probability "phase".)
  - This is, apparently, actually observed in physics labs!

- <u>Aside:</u> If we observe the system in state B at time s, then the system immediately <u>resets</u> to B, so that $\langle \Psi_s \,|\, E_1\rangle = \langle \Psi_s \,|\, E_2\rangle = \frac{1}{\sqrt{2}}$.

- Note: Actual quantum mechanics is more complicated: the wave function $\Psi(x,t)$ can depend on all spatial points x, not just on two different eigenstates corresponding to one single observable state.
  - Then the more general Schrödinger Wave Equation is the partial differential equation: $i\frac{h}{2\pi}\frac{\partial}{\partial t}\Psi(x,t) = \hat{H}\Psi(x,t)$, where $i = \sqrt{-1}$, h is Planck's constant (a very small constant), and $\hat{H}$ is a "Hamiltonian" energy operator (complicated).
  - The above example is the very special case where x takes on just two values (for the two states), and $\hat{H}$ is a 2×2 diagonal matrix with diagonal entries λ_1 and λ_2.

Renewal Theory:

- Like for Poisson Processes, have "arrival times" $\{T_n\}$.
- Here $T_0 = 0$, and $T_n = Y_1 + Y_2 + \ldots + Y_n$, where $\{Y_n\}_{n=1}^{\infty}$ are <u>independent</u> interarrival times, with $\{Y_n\}_{n=2}^{\infty}$ i.i.d.
- Then $N(t) = \max\{n \ge 0; T_n \le t\} = \#\{n \ge 1; T_n \le t\}$ is a "renewal process", on the state space $S = \{0,1,2,\ldots\}$.
  - If $\{Y_n\}_{n=1}^{\infty}$ are <u>all</u> i.i.d., then the process is <u>zero-delayed</u> (or <u>pure</u> or <u>ordinary</u>).
  - If $\{Y_n\}$ are i.i.d. $\sim \text{Exp}(\lambda)$, then $\{N(t)\}$ is Markovian (because of the "memoryless property"), and in fact $\{N(t)\}$ is a Poisson process (already studied).
  - But for other distributions of the Y_i, usually $\{N(t)\}$ is <u>not</u> Markovian, e.g. perhaps $\mathbf{P}[N(11) = 4 \,|\, N(10) = 3] \ne \mathbf{P}[N(11) = 4 \,|\, N(10) = 3,\ N(5) = 3]$.
  - In fact, Poisson processes are the <u>only</u> Markovian renewal processes; see Grimmett & Stirzaker, 2nd ed., Theorem 8.3(5).
- EXAMPLE: Suppose we replace a light-bulb whenever it burns out.
  - Let T_n be the n^{th} time we replace it. ($T_0 = 0$)
  - Then $Y_n = T_n - T_{n-1}$ is the <u>lifetime</u> of the n^{th} bulb.

- If the bulbs are identical, then $\{Y_n\}_{n=2}^{\infty}$ are i.i.d.
- Let $N(t)$ be the number of bulbs replaced by time t.
- Then $\{N(t)\}$ is a renewal process.
- If we started with a fresh bulb, then $\{Y_n\}_{n=1}^{\infty}$ are all i.i.d., so $\{N(t)\}$ is a "zero-delayed" renewal process. Otherwise probably not.

- Similarly for fixing a machine, etc.
- EXAMPLE: Let $\{Q(t)\}$ be a single-server queue.
  - Let $T_0 = 0$, and let T_n be the n^{th} time the queue empties. (i.e., the n^{th} time s such that $Q(s) = 0$ but $\lim_{t \nearrow s} Q(t) > 0$)
  - Let $Y_n = T_n - T_{n-1}$ be the time between queue emptyings.
  - Let $N(t) = \#\{n \geq 1; T_n \leq t\}$ be the number of queue emptyings by time t.
  - Then $\{N(t)\}$ is a renewal process.
  - (Probably not zero-delayed ... unless start at an emptying time ... or start empty and have exponential interarrival times ...)
- EXAMPLE: Let $\{X_n\}_{n=0}^{\infty}$ be an irreducible recurrent discrete-time Markov chain on a discrete state space S.
  - Let $i, j \in S$, and assume $X_0 = j$.
  - Let $T_0 = 0$, and for $n \geq 2$, let $T_n = \min\{t > T_{n-1} : X(t) = i\}$ be the time of the n^{th} visit to the state i. (integer-valued)
  - Let $Y_n = T_n - T_{n-1}$.
  - Then for $y = 1, 2, \ldots$, we have $\mathbf{P}(Y_1 = y) = f_{ji}^{(y)}$, where

    $$f_{ji}^{(n)} \;:=\; \mathbf{P}_j(X_n = i,\ \text{but}\ X_m \neq i \text{ for } 1 \leq m \leq n-1)\,,$$

    so $\sum_{y=1}^{\infty} f_{ji}^{(y)} = f_{ji} = 1$.
  - Similarly for $n \geq 2$, $\mathbf{P}(Y_n = y) = f_{ii}^{(y)}$.
  - So, if $N(t) = \#\{n \geq 1; T_n \leq t\}$ is the number of visits to i by time t, then $\{N(t)\}$ is a renewal process.
  - If $j = i$, then the process is zero-delayed, otherwise probably not.
- EXAMPLE: Let $\{X_n\}_{n=0}^{\infty}$ be an irreducible recurrent continuous-time Markov process on a discrete state space S.
  - Let $i, j \in S$, and assume $X(0) = j$.
  - Let $T_0 = 0$, and for $n \geq 2$, let $T_n = \min\{t > T_{n-1} : X(t) = i\}$ be the time of the n^{th} arrival at the state i, i.e. the n^{th} time s such that $Q(s) = i$ but there is a sequence of times $t_1, t_2, \ldots \nearrow s$ with $Q(t_m) \neq i$.
  - Let $Y_n = T_n - T_{n-1}$ be the time between arrivals at i.
  - Let $N(t) = \#\{n \geq 1; T_n \leq t\}$ be the number of arrivals to i by time t.

- Then $\{N(t)\}$ is a renewal process.
- If $j = i$, then the process is zero-delayed, otherwise probably not.

- ELEMENTARY RENEWAL THEOREM: For a renewal process as above, if $\mathbf{P}(Y_1 < \infty) = 1$, and if the "mean interarrival time" $\mu := \mathbf{E}(Y_2) < \infty$, then (a) $\lim_{t\to\infty} \frac{N(t)}{t} = 1/\mu$ w.p. 1,

  and (b) $\lim_{t\to\infty} \frac{\mathbf{E}[N(t)]}{t} = 1/\mu$.

- PROOF of (a):
  - By the Strong Law of Large Numbers (SLLN), $\lim_{n\to\infty} \frac{1}{n} T_n = \lim_{n\to\infty} \frac{1}{n}(Y_1 + Y_2 + \ldots + Y_n) = \mu$ w.p. 1.
  - Also, $Y_n < \infty$ w.p. 1 (for $n = 1$ by assumption, $n \geq 2$ since $\mu < \infty$).
  - Therefore, $\lim_{t\to\infty} N(t) = \infty$ w.p. 1.
  - Hence, $\lim_{t\to\infty} \frac{1}{N(t)} T_{N(t)} = \lim_{n\to\infty} \frac{1}{n} T_n = \mu$ w.p. 1.
  - But $N(t) = \max\{n \geq 0;\ T_n \leq t\}$. So, $T_{N(t)} \leq t < T_{N(t)+1}$. Hence,

$$\frac{T_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{T_{N(t)+1}}{N(t)} = \frac{T_{N(t)+1}}{N(t)+1} \, \frac{N(t)+1}{N(t)} .$$

  - As $t \to \infty$, w.p. 1, $\frac{T_{N(t)}}{N(t)} \to \mu$, and $\frac{T_{N(t)+1}}{N(t)+1} \to \mu$, and $\frac{N(t)+1}{N(t)} \to 1$.
  - So, $\frac{t}{N(t)} \to \mu$ w.p. 1 (by "sandwich theorem").
  - So, $\frac{N(t)}{t} \to 1/\mu$ w.p. 1, Q.E.D.

- PROOF of (b):
  - This is similar to part (a), but it's a bit more subtle (since $N(t)/t$ is neither bounded nor monotone).
  - So, we won't prove it here. Instead see e.g. Grimmett & Stirzaker, 2nd ed., p. 397, or Resnick, p. 191.

- EXAMPLE: Let $\{X_n\}_{n=0}^{\infty}$ be an irreducible recurrent discrete-time Markov chain on a discrete state space S.
  - Again let $N(t)$ be the number of visits to the state i by time t.
  - Then $\{N(t)\}$ is a renewal process.
  - Here $\mu = m_i =$ the mean return time for the state i.
  - Hence, by the Elementary Renewal Theorem part (a), $N(t)/t \to 1/m_i$.
  - But here if $t = n$, then $N(n) = \#\{k \leq n : X_k = i\} = \sum_{k=1}^{n} \mathbf{1}_{X_k = i}$.
  - Hence, $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{X_k=i} = 1/m_i$.
  - Same as we already showed previously!
  - (And, since $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{X_k=i} = \pi_i$, this showed that $\pi_i = 1/m_i$.)
  - Furthermore, by the Elementary Renewal Theorem part (b), or by the Bounded Convergence Theorem since in this case $N(t) \leq t$ so

$N(t)/t \leq 1$, we also have $\mathbf{E}[N(t)]/t \to 1/m_i$.

- BLACKWELL RENEWAL THEOREM: For a renewal process as above, suppose $\mathbf{P}(Y_1 < \infty)$, and $\mu := \mathbf{E}(Y_2) < \infty$, and Y_2 is "not arithmetic", i.e. there is <u>no</u> $\lambda > 0$ such that $\mathbf{P}(Y_2 = k\lambda \text{ for some } k \in \mathbf{Z}) = 1$. (Similar to "aperiodicity".) Then for any fixed $h > 0$,

$$\lim_{t\to\infty} \mathbf{E}[N(t+h) - N(t)] = h/\mu\,.$$

  - This is a "more refined" theorem, since it doesn't just consider the overall average $N(t)/t$, but rather considers the specific number of renewals between times t and $t+h$.
  - For a PROOF, see e.g. Grimmett & Stirzaker, 2nd ed., pp. 408–409, or Resnick, Section 3.10.3.

- PRACTICE PROBLEM: Let $Y_1, Y_2, \ldots$ be i.i.d. $\sim$ Uniform$[0,10]$. Let $T_0 = 0$, and $T_n = Y_1 + Y_2 + \ldots + Y_n$ for $n \geq 1$. Let $N(t) = \max\{n \geq 0 : T_n < t\}$. **(a)** Compute (with explanation) $\lim_{t\to\infty} N(t)/t$.
(b) Approximate (with explanation) $\mathbf{E}(\#\{n \geq 1 : \ 1234 < T_n < 1236\})$.

Renewal Reward Processes:

- Consider a renewal process as above, with $P(Y_1 < \infty) = 1$.
- Suppose the mean interarrival time $\mu := \mathbf{E}(Y_2) < \infty$.
- Suppose at the k^{th} renewal time T_k, you receive an "reward" (or cost) R_k.
  - Assume the $\{R_k\}$ are i.i.d.
- Let $R(t) = \sum_{k=1}^{N(t)} R_k$ be the total reward received by time t.
- RENEWAL REWARD THEOREM: $\lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbf{E}[R_1]}{\mu}$ w.p. 1.
  - PROOF: Here

$$\frac{R(t)}{t} = \frac{\sum_{k=1}^{N(t)} R_k}{t} = \frac{\sum_{k=1}^{N(t)} R_k}{N(t)} \frac{N(t)}{t}\,.$$

  - As $t \to \infty$, w.p. 1, $\frac{\sum_{k=1}^{N(t)} R_k}{N(t)} \to \mathbf{E}[R_1]$ by SLLN, and $\frac{N(t)}{t} \to 1/\mu$ by the Elementary Renewal Theorem part (a).
  - Hence, $\frac{R(t)}{t} \to \mathbf{E}[R_1] \times (1/\mu) = \mathbf{E}[R_1]/\mu$, Q.E.D.

Application – Car Purchases:

- Suppose each new car's lifetime L has distribution Uniform[0,10] (in years), after which it breaks down.
- Suppose your strategy is to buy a new car as soon as your old car breaks down, <u>or</u> after S years if the car hasn't broken down by then. $(0 \leq S \leq 10)$
- Suppose a new car costs 30 (in thousands of dollars).

- Suppose further that if your car breaks down before you sell it, then that costs you an extra 5 (in thousands of dollars).
- QUESTION: On average, how many cars will you buy each year?
- SOLUTION:
  - Here $\mu = \mathbf{E}(Y_2) = \mathbf{E}[\min(L, S)]$.
  - So, $\mu = S\,\mathbf{P}(L > S) + \mathbf{E}[L \,|\, L < S]\,\mathbf{P}(L < S)$
    $= S[(10 - S)/10] + (S/2)(S/10) = S - (S^2/20)$.
  - So, by the Elementary Renewal Theorem part (a), $\lim_{t\to\infty} N(t)/t = 1/\mu = 1/[S - (S^2/20)]$.
  - So, you will buy an average of $1/[S - (S^2/20)]$ cars per year, i.e. one car every $[S - (S^2/20)]$ years.
  - e.g. if $S = 10$, then you will buy an average of one car every 5 years. (Of course.)
  - e.g. if $S = 9$, then you will buy an average of one car every $99/20 = 4.95$ years.
  - Or, if $S = 1$, then you will buy an average of one car every $19/20 = 0.95$ years.
- QUESTION: About how many cars will you buy between times 562 and 566 (in years)?
- SOLUTION:
  - To apply the Blackwell Renewal Theorem, we want $t = 562$, and $t + h = 566$ so $h = 4$.
  - Hence, by the Blackwell Renewal Theorem, since $t = 562$ is reasonably large, the expected number of purchases between times 562 and 566 is approximately $h/\mu = 4/[S - (S^2/20)]$.
  - e.g. if $S = 9$, then this equals about 0.808.
- QUESTION: What is your long-run average car cost per year?
  - (And, what choice of S minimises this?)
- SOLUTION:
  - Let T_k be the time (in years) of the purchase of your k^{th} car. ($T_0 = 0$)
  - Let $Y_k = T_k - T_{k-1}$ be the k^{th} interarrival time.
  - Let $R(t)$ be your total car cost by time t.
  - Then this is a renewal reward process!
  - Here $E[R_1] = 30 + 5\,\mathbf{P}(L \le S) = 30 + 5\,(S/10) = 30 + S/2$.
  - Also $\mu = \mathbf{E}[Y_2] = S - (S^2/20)$ as above.

- Hence, w.p. 1, by the Renewal Reward Theorem,

$$\lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbf{E}[R_1]}{\mu} = \frac{30 + S/2}{S - (S^2/20)}\,.$$

- If $S = 10$ (i.e., never sell early), this equals $\frac{35}{5} = 7$.
- If $S = 9$, this equals $\frac{34.5}{99/20} \doteq 6.970$. Less!
- Minimised when $S \doteq 9.282$, then this equals about 6.964.
- Plot: www.probability.ca/sta447/Rcar

- CONCLUSION: Your best policy is to buy a new car as soon as your old car breaks, or is 9.282 years old, whichever comes first.

Discrete-time Markov Chains on Continuous State Spaces:

- Suppose instead of a discrete set S, our state space $\mathcal{X}$ is now <u>any</u> non-empty (perhaps uncountable) set, e.g. $\mathbf{R}$.
- The (one-step) transition probabilities are then given by $P(x, A)$, for each $x \in \mathcal{X}$ and each (measurable) $A \subseteq \mathcal{X}$.
  - So, for each fixed state $x \in \mathcal{X}$, $P(x, \cdot)$ is a <u>probability measure</u>.
  - (Also, for each fixed measurable $A \subseteq \mathcal{X}$, $P(x, A)$ is a measurable function of $x \in \mathcal{X}$.)
  - Here $P(x, A)$ is the probability, if the chain is at a point x, that it will jump to <u>somewhere</u> in the subset A at the next step.
- If $\mathcal{X}$ is countable, then $P(x, \{i\})$ corresponds to the transition probability p_{xi} of the discrete Markov chains as before.
  - But on a general state space, might have $P(x, \{i\}) = 0$ for all $i \in \mathcal{X}$.
- We also require an initial distribution ν, which is any probability distribution on $(\mathcal{X}, \mathcal{F})$.
- We then have a (discrete-time, general state space, time-homogeneous) *Markov chain* $X_0, X_1, X_2, \ldots$, where

$$\begin{aligned}
&\mathbf{P}(X_0 \in A_0,\ X_1 \in A_1, \ldots, X_n \in A_n) \\
&\quad = \int_{x_0 \in A_0} \nu(dx_0) \int_{x_1 \in A_1} P(x_0, dx_1) \ldots \\
&\qquad \ldots \int_{x_{n-1} \in A_{n-1}} P(x_{n-2}, dx_{n-1}) \int_{x_n \in A_n} P(x_{n-1}, dx_n)\,.
\end{aligned}$$

  - If we want, we can simplify $\int_{x_n \in A_n} P(x_{n-1}, dx_n) = P(x_{n-1}, A_n)$.
  - But <u>cannot</u> further simplify e.g. $\int_{x_{n-1} \in A_{n-1}} P(x_{n-2}, dx_{n-1})\, P(x_{n-1}, A_n)$.

END OF WEEK #11

- EXAMPLE: Consider the Markov chain on the real line (i.e. with $\mathcal{X} = \mathbf{R}$), where $P(x, \cdot) = N(\frac{x}{2}, \frac{3}{4})$ for each $x \in \mathcal{X}$.
  - Simulation: www.probability.ca/sta447/RconMCex
  - Equivalently, $X_{n+1} = \frac{1}{2}X_n + Z_{n+1}$, where $\{Z_n\}$ are i.i.d. with $Z_n \sim N(0, \frac{3}{4})$.
  - Stationary distribution? Convergence? (Soon!)
- Similar to before, write $\mathbf{P}_x(\cdots)$ for the probability of an event conditional on $X_0 = x$, i.e. under the assumption that the initial distribution ν is a point-mass at the single state x.
- And, define higher-order transition probabilities inductively by $P^1(x, A) = P(x, A)$, and $P^{n+1}(x, A) = \int_{z \in \mathcal{X}} P(x, dz) P^n(z, A)$ for $n \geq 1$.
- A stationary distribution for such a Markov chain is a probability measure $\pi(\cdot)$ on $\mathcal{X}$, such that $\pi(A) = \int_{\mathcal{X}} \pi(dx) P(x, A)$ for all $A \subseteq \mathcal{X}$.
  - (This generalises our earlier definition $\pi_j = \sum_{i \in S} \pi_i \, p_{ij}$.)
  - Like in the discrete case, Markov chains on general state spaces may or may not have stationary distributions.
- How to define a concept like "irreducible"?
- Problem: often (e.g. in the above example), $p_{ij}^{(n)} = 0$ for all $i, j \in \mathcal{X}$ and all $n \geq 1$. What to do?
- Let $\tau_A = \inf\{n \geq 0;\ X_n \in A\}$ be the first hitting time of the subset A.
  - Thus, $\tau_A < \infty$ iff the chain eventually hits the subset A.
- DEFN: a Markov chain on a general state space $\mathcal{X}$ is ϕ-irreducible if there is a non-zero (σ-finite) measure ψ on $\mathcal{X}$ such that if $\psi(A) > 0$, then $\mathbf{P}_x(\tau_A < \infty) > 0$ for all $x \in \mathcal{X}$.
  - That is, the chain has positive probability of eventually hitting any subset A of positive ψ measure.
  - Common choice: ψ = Lebesgue (length) measure on $\mathbf{R}$.
- And, how to define "period"? "aperiodicity"? Can't use "gcd" defn!
- DEFN: the period of a general-state-space Markov chain is the largest (finite) positive integer d such that there are non-empty disjoint subsets $\mathcal{X}_1, \ldots, \mathcal{X}_d \subseteq \mathcal{X}$, with $P(x, \mathcal{X}_{i+1}) = 1$ for all $x \in \mathcal{X}_i$ ($1 \leq i \leq d-1$) and $P(x, \mathcal{X}_1) = 1$ for all $x \in \mathcal{X}_d$. ("forced cycle") (diagram)
- DEFN: The chain is aperiodic if its period (as above) equals 1.
- GENERAL STATE SPACE MARKOV CONVERGENCE THEOREM: If a discrete-time Markov chain on a general state space is ϕ-irreducible and aperiodic, and has a stationary distribution $\pi(\cdot)$, then for π-almost every $x \in \mathcal{X}$,

$$\lim_{n \to \infty} \sup_{A \in \mathcal{F}} |P^n(x, A) - \pi(A)| \ = \ 0\,.$$

- That is, the Markov chain converges to its stationary distribution in probability (and in "total variation distance").
- For a proof see e.g. my review paper, or the Meyn and Tweedie book.

Application – Convergence of the Normal Example:

- Consider again the Markov chain from the above Example, with $\mathcal{X} = \mathbf{R}$, and with $P(x,\cdot) = N(\frac{x}{2}, \frac{3}{4})$ for each $x \in \mathcal{X}$. Let $\pi(\cdot) = N(0,1)$ be the standard normal distribution.
- CLAIM: $\pi(\cdot)$ is a stationary distribution for this chain.
  - PROOF: One way to think of the chain is, to get from X_{n-1} to X_n, we first <u>divide</u> X_{n-1} by 2, and then <u>add</u> an independent $N(0, \frac{3}{4})$ random variable.
  - That is, $X_n = \frac{X_{n-1}}{2} + Z_n$, where $\{Z_n\}$ are i.i.d. $\sim N(0, \frac{3}{4})$.
  - Now, if $X_{n-1} \sim \pi = N(0,1)$, then $\frac{X_{n-1}}{2} \sim N(0, \frac{1}{4})$, so $X_n = \frac{X_{n-1}}{2} + Z_n \sim N(0, \frac{1}{4} + \frac{3}{4}) = N(0,1) = \pi$.
  - Hence, π is stationary for this chain, Q.E.D.
- CLAIM: This Markov chain is ϕ-irreducible, where ψ is the usual Lebesgue (length) measure on $\mathbf{R}$.
  - PROOF: Indeed, if $x \in \mathbf{R}$, and if $A \subseteq \mathbf{R}$ has positive Lebesgue measure (essentially, has positive length), then $P(x, A) > 0$.
  - Hence, chain is ϕ-irreducible [in fact, with $\mathbf{P}(\tau_A = 1) > 0$], Q.E.D.
- CLAIM: This Markov chain is aperiodic.
  - PROOF (outline): If instead it had period $d \geq 2$, then for $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, we would have $P(x_1, \mathcal{X}_2) = 1$ while $P(x_2, \mathcal{X}_2) = 0$.
  - However, if a subset $A \subset \mathbf{R}$ has positive Lebesgue measure (i.e., positive length), then as above, $P(x, A) > 0$ for <u>all</u> $x \in \mathbf{R}$.
  - Or, if a subset $A \subset \mathbf{R}$ has <u>zero</u> Lebesgue measure (i.e., zero length), then $P(x, A) = 0$ for <u>all</u> $x \in \mathbf{R}$.
  - So, there is <u>no</u> subset $A \subset \mathbf{R}$ with $P(x_1, A) > 0$ but $P(x_2, A) = 0$.
  - Contradiction! So, period=1, i.e. it is aperiodic, Q.E.D.
- HENCE, by the General State Space Markov Chain Convergence Theorem, for π-almost every $x \in \mathcal{X}$, $\lim_{n\to\infty} \sup_{A\in\mathcal{F}} |P^n(x, A) - \pi(A)| = 0$.
  - In particular, $P^n(x, A) \to \pi(A)$.
  - So, e.g., $\lim_{n\to\infty} \mathbf{P}(X_n < 2) = \pi\{(-\infty, 2)\} = \Phi(2)$, where Φ is the standard normal c.d.f.

- PRACTICE PROBLEM: Consider a discrete-time Markov chain with state space $\mathcal{X} = \mathbf{R}$, and with transition probabilities such that $P(x, \cdot)$ is uniform on the interval $[x-1,\, x+1]$. Determine whether or not this chain is ϕ-irreducible.

- PRACTICE PROBLEM:
  (a) Prove that a Markov chain on a countable state space $\mathcal{X}$ is ϕ-irreducible if and only if there is $j \in \mathcal{X}$ such that $\mathbf{P}_i(\tau_j < \infty) > 0$ for all $i \in \mathcal{X}$, i.e. such that j can be reached from any state i.
  (b) Give an example of a Markov chain on a countable state space which is ϕ-irreducible as above, but which is not irreducible according to our previous (discrete state space) definition.

Convergence From Where?:

- QUESTION: Why does Theorem say "π-almost every" $x \in \mathcal{X}$?
- EXAMPLE:
  - State space $\mathcal{X} = \{1, 2, 3, \ldots\}$ (actually discrete). (diagram)
  - Transitions $P(1, \{1\}) = 1$, and for $x \geq 2$, $P(x, \{1\}) = 1/x^2$ and $P(x, \{x+1\}) = 1 - (1/x^2)$.
  - Stationary distribution? $\pi\{1\} = 1$ (of course).
  - ϕ-irreducible? Yes! If $\pi(S) > 0$, then $1 \in S$, so $P(x, S) \geq P(x, \{1\}) > 0$ for all $x \in \mathcal{X}$. So, can take $\psi = \pi$.
  - Aperiodic? Yes, since e.g. $P(1, \{1\}) > 0$.
  - So, by Theorem, for π-a.e. $x \in \mathcal{X}$, have $\lim_{n\to\infty} P(x, S) = \pi(S)$.
  - That is, $\lim_{n\to\infty} \mathbf{P}_x(X_n = 1) = 1$.
  - But if $X_0 = x \geq 2$, then $\mathbf{P}_x[X_n = x + n \text{ for all } n] = \prod_{j=x}^{\infty}(1 - (1/j^2))$.
  - This infinite product is positive, since $\sum_{j=x}^{\infty}(1/j^2) < \infty$. (optional)
  - This means that $\lim_{n\to\infty} \mathbf{P}_x(X_n = 1) \neq 1$. Contradiction??
  - No! Here, convergence holds if $x = 1$, which is π-a.e. since $\pi(1) = 1$, but not if $x \geq 2$.
- So, the convergence can be subtle! But it usually holds from any $x \in \mathcal{X}$ ("Harris recurrent"); see e.g. my Harris recurrence paper.

END OF WEEK #12

- REMINDER: FINAL EXAM! (And office hours.)
- Good luck and best wishes! – J.R.