STA447/2006 (Stochastic Processes) Lecture Notes, Winter 2012

by Jeffrey S. Rosenthal, University of Toronto

(Last updated: April 5, 2012.)

Note: These lecture notes are available to STA447/2006 students for their 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 handout, evaluation, etc. (www.probability.ca/sta447)
- Schedule: will take 15-minute break <u>if</u> 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. (Do <u>not</u> require measure theory.)
- This class considers stochastic <u>processes</u>, i.e. randomness which <u>proceeds in time</u>.
  - Will develop their mathematical theory (with a <u>few</u> 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.
- 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 $\lim_{k\to\infty} \mathbf{P}$(frog at pad #428 after k steps)?
  - Will the frog necessarily <u>eventually</u> return to pad #1000?
  - Will the frog necessarily <u>eventually</u> 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) <u>Markov chain</u> is specified by three ingredients:
  - A <u>state space</u> S, any non-empty finite or countable set. (e.g. the 1000 lily pads)
  - <u>transition probabilities</u> $\{p_{ij}\}_{i,j\in S}$, 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.)
  - <u>initial probabilities</u> $\{\nu_i\}_{i\in S}$, 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.

- 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 <u>other</u> 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 $p_{ij} = \rho\big(j - i + \min(1, i)\big)$.

- 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}?
  - $\mu_j^{(1)} = \mathbf{P}(X_1 = j) = \sum_{i \in S} \mathbf{P}(X_0 = i,\ X_1 = j) = \sum_{i \in S} \nu_i p_{ij}$.
  - ("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, $p_{ij}^{(2)} = \mathbf{P}(X_2 = j \,|\, X_0 = i) = \sum\limits_{k \in S} \mathbf{P}(X_2 = j,\ X_1 = k \,|\, X_0 = i) = \sum\limits_{k \in S} p_{ik} p_{kj}$.
  - 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} <u>powers</u> of the transition matrix P.
  - Convention: $P^{(0)} = I =$ identity matrix, i.e. $p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & \text{otherwise} \end{cases}$

———————— **END OF WEEK #1** ————————

[Extra course handouts. Course web page / notes (www.probability.ca/sta447).]

[Best book: Durrett? E-mail/meet me any time!]

Summary of Previous Class:

∗ Markov chain has ingredients S, $\{\nu_i\}$, $\{p_{ij}\}$

∗ Chain $\{X_n\}$ satisfies $\mathbf{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \nu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}$.

∗ Lots of examples, questions.

∗ Matrix form: $\mu^{(n)} = \nu P^n$, and $P^{(n)} = \mathbf{P}^n$ (even if S is infinite).

- 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,

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

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

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

- 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. $f_{21} = p_{11} + p_{22} f_{21} + p_{23} f_{31} + p_{24} f_{41} = (1/4) + (1/4) f_{21} + 0 + 0$, 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 $p_{00}^{(n)} = \mathbf{P}(n/2 \text{ heads and } n/2 \text{ tails on first } n \text{ tosses})$
    $= \binom{n}{n/2} p^{n/2} (1-p)^{n/2} = \frac{n!}{[(n/2)!]^2} p^{n/2} (1-p)^{n/2}$. [binomial distribution]
  - Stirling's approximation: if n large, then $n! \approx (n/e)^n \sqrt{2\pi n}$.
  - So, for n large and even,

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

  - 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 transient.
- (Similarly true for all other states besides 0, too.)

Communicating States:

- Say that state i communicates with state j, written $i \to j$, if $f_{ij} > 0$, i.e. if it is possible to get from i to j, 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 irreducible if $i \to j$ for all $i, j \in S$. (Previous examples?)
- CASES THEOREM: For an irreducible Markov chain, either
  (a) $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ for all $i, j \in S$, so all states are recurrent.
  ("recurrent Markov chain")
  or (b) $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all $i, j \in S$, so all 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 finite state space always falls into case (a), i.e. $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ for all $i, j \in S$, and all states are recurrent.
- PROOF 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 $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$.

– 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 $\mathbf{P}(\exists n \geq 1 \text{ with } X_n = 428 \,|\, X_0 = 1000)$??
  – Next class!

——————————— **END OF WEEK #2** ———————————

[Homework #1 on course web page! Due Thurs Feb 9 at 6:10pm sharp. Start now!]

[Free Durrett version on-line. Raw TeX notes? E-mail/meet me any time!]

Summary of Previous Class:
∗ Recurrence Thm: equivalences to recurrence/transience
—— finite example
—— s.r.w.: rec iff $p = 1/2$
∗ Communicating states, irreducibility
∗ Case Thm ("Corollary Thm"): if irred, then cases (a) and (b)
—— Followed from Sum Lemma: if communicate, then sums both infinite or both finite
∗ Finite Space Thm: if S finite, then in case (a)
—— Frog Example: $\mathbf{P}(\exists n \geq 1 \text{ with } X_n = 1000 \,|\, X_0 = 1000) \;=\; 1$.

– But what about $\mathbf{P}(\exists n \geq 1 \text{ with } X_n = 428 \,|\, X_0 = 1000)$??

- F-LEMMA: if $j \to i$ and $f_{jj} = 1$, then $f_{ij} = 1$.
- PROOF OF F-LEMMA:
  – Assume $i \neq j$ (otherwise trivial).
  – Let $T_i = \min\{n \geq 1 : X_n = i\}$. ($T_i = \infty$ if never hit i.)
  – Since $j \to i$, $\mathbf{P}_j(T_i < \infty) > 0$.
  – In fact, must have $\mathbf{P}_j(T_i < T_j) > 0$. (Otherwise, would always return to j before hitting i, so would never reach i from j.)

- But $1 - f_{jj} = \mathbf{P}_j(T_j = \infty) \geq \mathbf{P}_j(T_i < T_j)\, \mathbf{P}_i(T_j = \infty) = \mathbf{P}_j(T_i < T_j)\,(1 - f_{ij})$.
- If $f_{jj} = 1$, then $1 - f_{jj} = 0$, so must have $1 - f_{ij} = 0$, i.e. $f_{ij} = 1$. *Q.E.D.*

- Putting all of this 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.)
- 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)
- But $f_{05} = 1$. Contradiction? No!

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 $\sum\limits_{i \in S} \pi_i = 1$.
- Defn: π is <u>stationary</u> for a Markov chain $P = (\, p_{ij} \,)$ if $\sum\limits_{i \in S} \pi_i p_{ij} = \pi_j$ 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 $j \in S$, $\sum\limits_{i \in S} \pi_i p_{ij} = \frac{1}{1000}(\frac{1}{3} + \frac{1}{3} + \frac{1}{3}) = \frac{1}{1000} = \pi_j$.
  - So, $\{\pi_i\}$ is stationary distribution!
- (More generally, if chain is "doubly stochastic", i.e. $\sum\limits_{i \in S} p_{ij} = 1$ 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 <u>not</u> 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 <u>messy</u>. (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 $\{\pi_i\}$, then does $\lim_{n\to\infty} \mathbf{P}[X_n = i] = \pi_i$?
- Not necessarily!
- Example: $S = \{1,2\}$, and $v_1 = 1$, and $(\, p_{ij} \,) \;=\; \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.
  - If $\pi_1 = \pi_2 = \frac{1}{2}$ (say), then $\{\pi_i\}$ stationary (check!).
  - But $\lim_{n\to\infty} \mathbf{P}[X_n = 1] = \lim_{n\to\infty} 1 = 1 \neq \frac{1}{2} = \pi_1$.
  - Not irreducible! ("reducible")
- Example: $S = \{1,2\}$, and $v_1 = 1$, 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"

—————————— **END OF WEEK #3** ——————————

[HW#1 due next week, Feb 9, 6:10 sharp.]

[Midterm Feb 16, 6:10, one hour, in U.C. room 266 (East Hall)]

[Office hours: next Wed Feb 8 at 10:30–11:00 and 3:30–5:00.]

Summary of Previous Class:
∗ F-Lemma: $f_{jj} = 1$ implies $f_{ij} = 1$ if $j \to i$
∗ Stronger Convergence Thm: five equivalences to case (a)
—— Frog Example, finite example, s.r.w.
∗ Stationary distribution: $\sum_{i\in S} \pi_i p_{ij} = \pi_j$
—— Frog Example, doubly stochastic, Ehrenfest's Urn
∗ Reversibility: implies stationarity
∗ Does $\lim_{n\to\infty} \mathbf{P}[X_n = i] = \pi_i$?
—— Not if reducible or "periodic".

- Aside: Consider s.r.w. $\{X_n\}$ with $p > 1/2$.
  - 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 $f_{ij} = 1$.)
- Defn: the <u>period</u> of a state i is the greatest common divisor of the set $\{n \geq 1; p_{ii}^{(n)} > 0\}$.
  - e.g. if period of i is 2, this means that it is only possible to get from i to i in an <u>even</u> numbers of steps.
  - If period if each state is 1, say chain is "aperiodic".
- Example: $S = \{1, 2, 3\}$, and $p_{12} = p_{23} = p_{31} = 1$.
  - Then period of each state is 3.
- Example: $S = \{1, 2, 3\}$, and $(\, p_{ij}\,) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$.
  - Then period of state 1 is $gcd\{2, 3, \ldots\} = 1$.
  - Aperiodic!
- Observation: if $p_{ii} > 0$, then period of i is 1 (since $gcd\{1, \ldots\} = 1$).
  - (Converse false, as in previous example.)
  - Or, if there is some $n \geq 1$ with $p_{ii}^{(n)} > 0$ and $p_{ii}^{(n+1)} > 0$, then period of i is 1 (since $gcd\{n, n+1, \ldots\} = 1$).
- Frog Example: $p_{ii} > 0$, so chain aperiodic.
- Simple Random Walk: can only return after <u>even</u> number of steps, so period of each state is 2.
- Ehrenfest's Urn: again, can only return after <u>even</u> number of steps, so period of each state is 2.

- EQUAL PERIODS LEMMA: if $i \leftrightarrow j$, then the periods of i and of j are equal.
- PROOF:
  - Let the periods of i and j be t_i and t_j.
  - Find $r, s \in \mathbf{N}$ with $p_{ij}^{(r)} > 0$ and $p_{ji}^{(s)} > 0$.
  - Then $p_{ii}^{(r+s)} \geq p_{ij}^{(r)} p_{ji}^{(s)} > 0$, so t_i divides $r + s$.
  - Also if $p_{jj}^{(n)} > 0$, then $p_{ii}^{(r+n+s)} \geq p_{ij}^{(r)} p_{jj}^{(n)} p_{ji}^{(s)} > 0$, so t_i divides $r + n + s$, hence t_i divides n.
  - So, t_i is <u>a common divisor</u> of $\{n \in \mathbf{N};\ p_{jj}^{(n)} > 0\}$.
  - So, $t_j \geq t_i$ (since t_j is <u>greatest</u> common divisor).
  - Similarly, $t_i \geq t_j$, so $t_i = t_j$. *Q.E.D.*
- COR: if chain <u>irreducible</u>, then all states have same period.
- COR: if chain irreducible and $p_{ii} > 0$ for <u>some</u> 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 <u>recurrent</u>.
- PROOF:
  - Suppose it's 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.
  - Let $n \to \infty$ (using "M-test"): $\pi_j = 0$ for all $j \in S$. Impossible!
  - So, it must be recurrent. *Q.E.D.*
- NUMBER THEORY LEMMA: If a set A of positive integers is non-empty, and additive (i.e. $m + n \in A$ whenever $m \in A$ and $n \in A$), and aperiodic (i.e. $gcd(A) = 1$), then there is $n_0 \in \mathbf{N}$ such that $n \in A$ for all $n \geq n_0$.

- (Proof omited; see e.g. Durrett p. 51 / 2nd ed. p. 35, or Rosenthal p. 92.)
- COR: If a state i is aperiodic, and $f_{ii} > 0$, then there is $n_0(i)$ such that $p_{ii}^{(n)} > 0$ for all $n \geq n_0(i)$.
- PROOF: Let $A = \{n \geq 1 : p_{ii}^{(n)} > 0\}$.
  - Then A is non-empty since $f_{ii} > 0$.
  - And, A is additive since $p_{ii}^{(m+n)} \geq p_{ii}^{(m)} p_{ii}^{(n)}$.
  - And, A is aperiodic by assumption.
  - Hence, the result follows from the Lemma. *Q.E.D.*
- COR: If a chain is irreducible and aperiodic, then for any states $i, j \in S$, there is $n_0(i,j)$ such that $p_{ij}^{(n)} > 0$ for all $n \geq n_0(i,j)$.
- PROOF:
  - Find $n_0(i)$ as above.
  - Find $m \in \mathbf{N}$ such that $p_{ij}^{(m)} > 0$.
  - Then let $n_0(i,j) = n_0(i) + m$.
  - Then if $n \geq n_0(i,j)$, then $n - m \geq n_0(i)$, so $p_{ij}^{(n)} \geq p_{ii}^{(n-m)} p_{ij}^{(m)} > 0$. *Q.E.D.*
- PROOF OF MARKOV CHAIN CONVERGENCE THEOREM (long!):
- Define a *new* Markov chain $\{(X_n, Y_n)\}_{n=0}^{\infty}$, with state space $\overline{S} = S \times S$, and transition probabilities $\overline{p}_{(ij),(k\ell)} = p_{ik} p_{j\ell}$.
  - Intuition: new chain has two coordinates, each of which is an independent copy of the original Markov chain. ("coupling")
- The new chain has stationary distribution $\overline{\pi}_{(ij)} = \pi_i \pi_j$ (because of independence).
- Furthermore, $\overline{p}_{(ij),(k\ell)}^{(n)} > 0$ whenever $n \geq \max\left[n_0(i,k),\, n_0(j,\ell)\right]$.
  - So, new chain is irreducible and aperiodic.
- So, by Stationary Recurrence Lemma, new chain is recurrent.
- Choose $i_0 \in S$, 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 $i, j, k \in S$,

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

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. (Since if it has one, then it must equal $\lim_{n\to\infty} \mathbf{P}(X_n = j)$.)

- 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, stationary distributions not unique!

– But chain is not irreducible.

- What about periodic chains? (e.g. s.r.w., Ehrenfest)

—————————— **END OF WEEK #4** ——————————

[Collect HW#1.]

[Midterm next week, Feb 16, 6:10, one hour, in U.C. room 266 (East Hall). Write NAME and STUDENT NUMBER and STA 447 / 2006 (CIRCLE ONE) at top of paper.]

[(Then return to our usual classroom for one-hour class 7:45 – 8:45.)]

[No class Feb 23 (Reading Week).]

[Office hours: next Wed Feb 15 at 10:30–11:00 and 3:30–4:30.]

Summary of Previous Class:
∗ period of states; examples; equal if they communicate
∗ Stationary Recurrence Lemma
∗ THM: if M.C. irreducible, aperiodic, and has stationary distribution π, then $p_{ij}^{(n)} \to \pi_j$ for all i, j. (Proof used "coupling": two copies of chain which eventually become equal, thus "forgetting" where they started.)
—— [Aside: can also use eigenvalues e.g. if S finite; see e.g. www.probability.ca/eigen.pdf]
—— Cor: for irreducible, aperiodic chain, stationary distribution UNIQUE.

- ASIDE: THE (WEIERSTRASS) M-TEST. (Optional.)
  - 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 $a_k \le \sup_n b_{nk}$, so $\sum_{k=1}^{\infty} \sup_n |b_{nk} - a_k| \le 2 \sum_{k=1}^{\infty} \sup_n |b_{nk}| < \infty$.
  - 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 $n \ge N$, $\left| \sum\limits_{k=1}^{\infty} b_{nk} - \sum\limits_{k=1}^{\infty} a_k \right| \le \sum\limits_{k=1}^{\infty} |b_{nk} - a_k| < K \frac{\epsilon}{2K} + \frac{\epsilon}{2} = \epsilon$. *Q.E.D.*
- What about periodic chains? (e.g. s.r.w., Ehrenfest)
- PERIODIC CONVERGENCE THM: Suppose chain irreducible, with period $b \ge 2$,

and stat dist $\{\pi_i\}$. Then $\forall i \in S$, $\lim_{n\to\infty} \frac{1}{b}\,[p_{ij}^{(n)} + ... + p_{ij}^{(n+b-1)}] = \pi_j$, 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):
- For $r = 0, 1, 2, \ldots, b-1$, let $S_r = \{j \in S : p_{ij}^{(bm+r)} > 0 \text{ for some } m \in \mathbf{N}\}$.
- 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.
- 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 $0 \le r \le b-1$.
- 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 $j \in S$.
- *Q.E.D.*
- (e.g. for Ehrenfest's Urn, if $i = 0$, then $S_0 = \{\text{even } i \in S\}$, and $S_1 = \{\text{odd } i \in S\}$.)

- 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 $p_{ii}^{(n)} \to 0$.
  - Then for any $i, j \in S$, find $m \in \mathbf{N}$ with $p_{ji}^{(m)} > 0$, then $p_{ii}^{(n+m)} \ge p_{ij}^{(n)} p_{ji}^{(m)}$, so we must have $p_{ij}^{(n)} \le p_{ii}^{(n+m)} \,/\, p_{ji}^{(m)} \to 0$ 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 $i \ge 2$, $p_{i,i} = p_{i,i+1} = 1/4$ 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

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

- 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 826,000 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)"

——————————— **END OF WEEK #5** ———————————

[Midterm today! At 6:10, one hour, in U.C. room 266 (East Hall). Write NAME and STUDENT NUMBER and STA 447 / 2006 (CIRCLE ONE) at top of paper.]

[(Then we'll return to our usual classroom for one-hour class 7:45 – 8:45.)]

[Midterm solutions posted on course web page.]

[No class Feb 23 (Reading Week).]

Summary of Previous Class:
∗ M-test (optional)
∗ Periodic Convergence Thm: $\frac{1}{b}[p_{ij}^{(n)} + \ldots + p_{ij}^{(n+b-1)}] \to \pi_j$ (proof optional)
∗ no stat dist for s.r.w.; yes for some other infinite chains
∗ Metropolis (MCMC) algorithm: propose new value Y_{n+1}, then accept or reject it. $p_{i,i+1} = (1/2) \min(1, \pi_{i+1}/\pi_i)$, etc.
—— Makes π stationary, and converges to it.
—— [Animated version available at: www.probability.ca/met]
∗ Random walk on weighted undirected graph (V, w)
—— $p_{uv} = \frac{w(u,v)}{d(u)}$

- 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 $1 \le i \le 1000$, and $w(i, i+1) = w(i+1, i) = 1$ for $1 \le i \le 999$, 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).
- Let $\pi_u = \frac{d(u)}{Z}$, so $\pi_u \ge 0$ and $\sum_u \pi_u = 1$.
  - Then $\pi_u p_{uv} = \frac{d(u)}{Z} \frac{w(u,v)}{d(u)} = \frac{w(u,v)}{Z}$.
  - And, $\pi_v p_{vu} = \frac{d(v)}{Z} \frac{w(v,u)}{d(v)} = \frac{w(v,u)}{Z} = \frac{w(u,v)}{Z}$. Same!
  - So, chain is reversible w.r.t. $\{\pi_u\}$.
  - So, $\{\pi_u\}$ is stationary dist.
- If graph is <u>connected</u>, then chain is irreducible.
- If graph is <u>bipartite</u> (i.e., can be divided into two subsets s.t. all links go from one to the other), then the chain has period 2.
  - Otherwise, the chain is aperiodic (since can return to u in 2 steps).
  - (i.e., 1 and 2 are the <u>only</u> possible periods)
- This proves: 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 (whether bipartite or not), $\lim_{n\to\infty} \frac{1}{2}[p_{uv}^{(n)} + p_{uv}^{(n+1)}] = \frac{d(v)}{Z}$.
- Example: $V = \{1, 2, \ldots, K\}$, with $w(i, i+1) = w(i+1, i) = 1$ for $1 \le i \le K-1$, with $w(i,j) = 0$ otherwise. (“stick”) (diagram)
  - Connected, but bipartite.
  - $p_{12} = 1$, and $p_{K,K-1} = 1$, and $p_{i,i+1} = p_{i,i-1} = 1/2$ for $2 \le i \le K-1$.
  - $\pi_i = \frac{1}{2K-2}$ for $i = 1, K$, and $\pi_i = \frac{2}{2K-2}$ for $2 \le i \le K-1$.
  - Then, know that $\lim_{n\to\infty} \frac{1}{2}[p_{ij}^{(n)} + p_{ij}^{(n+1)}] = \pi_j$ for all $j \in V$.

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)?
- 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 $1 \leq a \leq c - 1$,

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

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

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

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

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

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

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

- 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!

—————————— **END OF WEEK #6** ——————————

[Return HW#1 (mean 86.9%) and midterm (mean 89.6%).]

[HW#2 assigned, on web page, due at 6:10pm sharp on Thurs March 15.]

[HW#2, Q6, defn of $p_{(x,y),(z,w)}$: π should be f (4 times). (Corrected in web version.)]

[Final Exam: Thursday Apr 12, 7–10 p.m., Room 200, Brennan Hall, St. Michael's College, 81 St. Mary Street, 2nd floor.]

Summary of Previous Class:
$*$ Random walk on a graph: $p_{u,v} = w(u,v)/d(u)$, makes stationary distribution $\pi_u = d(u)/\sum_v d(v)$ (if sum finite).
$*$ Gambler's ruin: simple random walk with prob p, start at a, then $\mathbf{P}_a(T_c < T_0)$ is a/c if $p = 1/2$, otherwise $[((1-p)/p)^a - 1]/[((1-p)/p)^c - 1]$.

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 <u>martingale</u> if $\mathbf{E}|X_n| < \infty$ for all 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 $\mathbf{E}(X_T) = \mathbf{E}(X_0)$?

- 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 "$1 - \mathbf{1}_{T \leq k-1}$" 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:

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

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

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

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

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

- Example: s.s.r.w., with $X_0 = 0$, and $T = \min\Big(10^{12},\ \inf\{n \ge 0 : X_n = -5\}\Big)$.
  - 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 <u>huge</u>.
- 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 $\lim_{n\to\infty} \mathbf{E}\big(X_n\, \mathbf{1}_{T>n}\big) = 0$, 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 $T = \inf\{n \ge 0 : X_n = 0 \text{ or } X_n = c\}$.
  - Then $\mathbf{P}(T < \infty) = 1$ (game must eventually end). [Formally: $\mathbf{P}(T > mc) \le (1 - p^c)^m \to 0$ as $m \to \infty$, since if win c times in a row then game over.]
  - Also, $|X_n| \mathbf{1}_{n \le T} \le c < \infty$ for all n.
  - So, by Optional Stopping Corollary, $\mathbf{E}(X_T) = \mathbf{E}(X_0) = a$.
  - Hence, as before, $a = c\, s(a) + 0\, (1 - s(a))$, so must have $s(a) = a/c$. (Easier solution!)

- What about Gambler's Ruin with $p \ne 1/2$?
  - Here $\{X_n\}$ is <u>not</u> a martingale: $\sum_j j\, p_{ij} = p(i+1) + (1-p)(i-1) = i + 2p - 1 \ne i$.
  - Trick: let $Y_n = \left(\frac{1-p}{p}\right)^{X_n}$.
  - Then $\mathbf{E}(Y_{n+1} \,|\, Y_0, Y_1, \ldots, Y_n) = p\left[Y_n \left(\frac{1-p}{p}\right)\right] + (1-p)\left[Y_n \,/\, \left(\frac{1-p}{p}\right)\right] = Y_n(1-p) + Y_n(p) = Y_n$.
  - So, $\{Y_n\}$ is a martingale!
  - And, $\mathbf{P}(T < \infty) = 1$ as before (with the same T).
  - And, $|Y_n| \mathbf{1}_{T \le n} \le \max\left(\left(\frac{1-p}{p}\right)^0,\ \left(\frac{1-p}{p}\right)^c\right) < \infty$ 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!)

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

—————————— **END OF WEEK #7** ——————————

[Reminder: HW#2 assigned, due at 6:10pm sharp next class (Thurs March 15).]

[Question 3(e) was missing factors of 1/2 (ii), 1/3 (iii), 1/6 (iv). Corrected on web.]

[Office hours: next Wed, Mar 14, 1:30–3:30.]

[Reminder: Final Exam, Thursday Apr 12, 7–10 p.m., Room 200, Brennan Hall. (on web)]

Summary of Previous Class:
$*$ Martingales: defn, examples.
—— $\mathbf{E}(X_n) = \mathbf{E}(X_0)$ for all n.
$*$ If $\{X_n\}$ martingale, T stopping time, then $\mathbf{E}(X_T) = \mathbf{E}(X_0)$ provided that:
—— $\mathbf{P}(T \leq M) = 1$, or
—— $\mathbf{P}(T < \infty) = 1$, $\mathbf{E}|X_T| < \infty$, and $\lim_{n\to\infty} \mathbf{E}(X_n\,\mathbf{1}_{T>n}) = 0$, or
—— $\mathbf{P}(T < \infty) = 1$, and $|X_n|\mathbf{1}_{n\leq T} \leq M$.
$*$ Used this for alternate derivation of gambler's ruin probs.
$*$ Wald's Thm: if $X_n = a + Z_1 + \ldots + Z_n$, with $\{Z_i\}$ iid, finite mean m, and T stopping time for $\{X_n\}$ with $\mathbf{E}(T) < \infty$, then $\mathbf{E}(X_T) = a + m\,\mathbf{E}(T)$.

- PROOF: We compute (using that Z_i indep of $\{T \geq i\} = \{T \leq i-1\}^C$) that

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

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

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

  - (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 $\mathbf{E}(Y) = \mathbf{E}[\sum_{i=1}^{\infty} |Z_i|\, \mathbf{1}_{T \ge i}] = \sum_{i=1}^{\infty} \mathbf{E}[|Z_i|\, \mathbf{1}_{T \ge i}] = \sum_{i=1}^{\infty} \mathbf{E}|Z_i|\, \mathbf{E}[\mathbf{1}_{T \ge i}] = \mathbf{E}|Z_1| \sum_{i=1}^{\infty} \mathbf{E}[\mathbf{1}_{T \ge i}] = \mathbf{E}|Z_1|\, \mathbf{E}(T) < \infty$.)

- EXAMPLE: Gambler's Ruin with $p \neq 1/2$, and $T = \inf\{n \ge 0 : X_n = 0 \text{ or } c\}$.
  - 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, $\mathbf{E}(T) < \infty$ [easy, e.g. $\mathbf{P}(T \ge cn) \le (1 - p^c)^n$ so $\mathbf{E}(T) \le c/p^c < \infty$].
  - Hence, by Wald's Thm, $\mathbf{E}(X_T) = a + m\, \mathbf{E}(T)$.
  - So, $\mathbf{E}(T) = \frac{1}{m}\left(\mathbf{E}(X_T) - a\right) = \frac{1}{2p-1}\left(c \frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1} - a\right)$.
  - e.g. $p = 0.49$, $a = 9{,}700$, $c = 10{,}000$: $\mathbf{E}(T) = 484{,}997$. (large!)
- But what about $\mathbf{E}(T)$ when $p = 1/2$??
  - Then $m = 0$, so the above method does not work.
- LEMMA: Let $X_n = a + Z_1 + \ldots + Z_n$, where $\{Z_i\}$ i.i.d. with mean 0 and variance $v < \infty$. Let $Y_n = (X_n - a)^2 - nv = (Z_1 + \ldots + Z_n)^2 - nv$. Then $\{Y_n\}$ is a martingale.
- PROOF:
  - Check: $\mathbf{E}|Y_n| \le \mathbf{Var}(X_n) + nv = 2nv < \infty$.
  - Also, since Z_{n+1} indep of $Z_1, \ldots, Z_n, Y_0, \ldots, Y_n$, we have

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

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

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

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

- COR: If $\{X_n\}$ is Gambler's Ruin with $p = 1/2$, and $T = \inf\{n \geq 0 : X_n = 0 \text{ or } c\}$, 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 $M \to \infty$, $\mathbf{E}(S_M) \to \mathbf{E}(T)$ [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, $(X_{S_M} - a)^2 \leq \max\left(a^2, (c-a)^2\right) < \infty$].
  - Hence, $\mathbf{E}(T) = \mathbf{E}[(X_T - a)^2] = \mathbf{Var}(X_T)$ (since $\mathbf{E}(X_T) = a$).
  - Thus, $\mathbf{E}(T) = (a/c)\,(c-a)^2 + (1 - a/c)a^2 = (a/c)(c^2 + a^2 - 2ac) + (a^2 - a^3/c) = ac + a^3/c - 2a^2 + a^2 - a^3/c = ac - a^2 = a(c-a)$, *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!)
- 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.

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 $\mathbf{E}(X_{n+1} \,|\, X_0, \ldots, X_n) = \mathbf{E}(Z_{n+1,1} + Z_{n+1,2} + \ldots + Z_{n+1,X_n} \,|\, X_0, \ldots, X_n) = m\,X_n$.

• Let $Y_n = X_n / m^n$.

- Then since $Y_n \leftrightarrow X_n$ is one-to-one function,

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

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

- And, must have $\mathbf{E}|Y_n| < \infty$ (since $a < \infty$ and $m < \infty$).
- 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 $\mathbf{E}(X_n) = a\,m^n$.

• If $m < 1$, then $\mathbf{E}(X_n) = a\,m^n \to 0$.

- But $\mathbf{E}(X_n) = \sum_{k=0}^{\infty} k\,\mathbf{P}(X_n = k) \geq \sum_{k=1}^{\infty} \mathbf{P}(X_n = k) = \mathbf{P}(X_n \geq 1)$.
- Hence, $\mathbf{P}(X_n \geq 1) \leq \mathbf{E}(X_n) = a\,m^n \to 0$, i.e. $\mathbf{P}(X_n = 0) \to 1$.
- Certain extinction!

- If $m > 1$, then $\mathbf{E}(X_n) \to \infty$.
  - Indeed, in this case, $\mathbf{P}(X_n \to \infty) > 0$.
  - But still have $\mathbf{P}(X_n \to \infty) < 1$ (assuming $\mu\{0\} > 0$), i.e. could still have $X_n \to 0$ (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 $m = 1$!

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

[Collect HW#2! Note: will post HW#3 on web soon (due Apr 5).]

Summary of Previous Class:

∗ Proved Wald's Thm: $\mathbf{E}(X_T) = a + m\,\mathbf{E}(T)$.
—— Used this to compute $\mathbf{E}(T)$ for Gambler's Ruin, $p \neq 1/2$.
∗ Lemma: $\{(X_n - a)^2 - nv\}$ martingale.
—— Used this to compute $\mathbf{E}(T)$ for Gambler's Ruin, $p = 1/2$.
∗ Martingale Convergence Thm
∗ Branching processes:
—— $m < 1$: certain extinction
—— $m > 1$: possible extinction, possible flourishing
—— $m = 1$: certain extinction (!) by Mart Conv Thm

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.
- <u>Brownian motion</u> $\{B_t\}_{t\geq 0}$ 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 <u>independent</u> of $Y_t^{(M)}$.
  - CONCLUSION: $B_s - B_t \sim \text{Normal}(0, s-t)$, and it's <u>independent</u> 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 <u>independent</u>. ("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")
- Example: $\mathbf{E}[(B_2 + B_3 + 1)^2] = \mathbf{E}[(B_2)^2)] + \mathbf{E}[(B_3)^2] + 1^2 + 2\mathbf{E}[B_2 B_3] + 2\mathbf{E}[B_2(1)] + 2\mathbf{E}[B_3(1)] = 2 + 3 + 1 + 2(2) + 2(0) + 2(0) = 10$.
- Example: $\mathbf{Var}[B_3 + B_5 + 7] = \mathbf{E}[(B_3 + B_5)^2] = \mathbf{E}[(B_3)^2)] + \mathbf{E}[(B_5)^2] + 2\mathbf{E}[B_3 B_5] = 3 + 5 + 2(3) = 14$.

- 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 <u>given</u> B_r for $0 \le r \le t$, what is the <u>conditional</u> distribution of B_s?
  - Similar to above, $B_s \,|\, B_t \;=\; B_t + (B_s - B_t) \mid B_t \;=\; B_t + \text{Normal}(0,\, s-t) \;\sim\; \text{Normal}(B_t,\, s-t)$. (i.e., given B_t, B_s is normal with mean B_t, variance $s-t$.)
  - So, in particular, $\mathbf{E}[B_s \,|\, \{B_r\}_{0 \le r \le t}] = B_t$.
  - Hence, $\{B_t\}$ is a (continuous-time) <u>martingale</u>!
  - So, similar results apply just like for discrete-time martingales.
- Example: let $a, b > 0$, and let $\tau = \min\{t \ge 0 : B_t = -a \text{ or } b\}$.
  - What is $p \equiv \mathbf{P}(B_\tau = b)$?
  - Well, here $\{B_t\}$ is martingale, and τ is stopping time.
  - Furthermore, $\{B_t\}$ is bounded up to time τ, i.e. $|B_t| \mathbf{1}_{t \le \tau} \le \max(|a|, |b|)$.
  - So, just like for discrete martingales, must have $\mathbf{E}(B_\tau) = \mathbf{E}(B_0) = 0$.
  - Hence, $p(b) + (1-p)(-a) = 0$, so $p = \frac{a}{a+b}$. (as expected)
  - But what is $e \equiv \mathbf{E}(\tau)$?
- Let $Y_t = B_t^2 - t$.
  - Then for $0 < t < s$, $\mathbf{E}[Y_s \,|\, \{B_r\}_{r \le t}] = \mathbf{E}[B_s^2 - s \,|\, \{B_r\}_{r \le t}] = (\mathbf{E}[B_s \,|\, \{B_r\}_{r \le t}])^2 + \mathbf{Var}[B_s \,|\, \{B_r\}_{r \le t}] - s = (B_t)^2 + (s-t) - s = Y_t$.
  - Then, $\mathbf{E}[Y_s \,|\, \{Y_r\}_{r \le t}] = \mathbf{E}\Big[\mathbf{E}[Y_s \,|\, \{B_r\}_{r \le t}] \,\Big|\, \{Y_r\}_{r \le t}\Big] = \mathbf{E}\big[Y_t \,|\, \{Y_r\}_{r \le t}\big] = Y_t$.
  - So, $\{Y_t\}$ is also a martingale!
- Example (continued): What is $e \equiv \mathbf{E}(\tau)$?
  - Well, with $Y_t = B_t^2 - t$, have $\mathbf{E}(Y_\tau) = \mathbf{E}(B_\tau^2) - \mathbf{E}(\tau) = pb^2 + (1-p)(-a)^2 - e = \frac{a}{a+b} b^2 + \frac{b}{a+b} a^2 - e = ab - e$.
  - <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, let $\tau_M = \min(\tau, M)$, then $\mathbf{E}(Y_{\tau_M}) = 0$, so $\mathbf{E}(\tau_M) = \mathbf{E}(B_{\tau_M}^2)$. As $M \to \infty$, $LHS \to \mathbf{E}(\tau)$ by Bounded Convergence Thm, and $RHS \to \mathbf{E}(B_\tau^2)$ by Bounded Convergence Thm, so $\mathbf{E}(\tau) = \mathbf{E}(B_\tau)$, i.e. $\mathbf{E}(Y_\tau) = 0$ as above.

- Suppose $X_t = 2 + 5t + 3B_t$ for $t \geq 0$.
  - Then $\mathbf{E}(X_t) = 2 + 5t$, $\mathbf{Var}(X_t) = 3^2\mathbf{Var}(B_t) = 9t$, and $X_t \sim \text{Normal}(2 + 5t, 9t)$.
  - Also for $0 < t < s$, $\mathbf{Cov}(X_t, X_s) = \mathbf{E}[(3B_t)(3B_s)] = 9t$.
  - Fancy notation: $dX_t = 5\,dt + 3\,dB_t$. ("diffusion")
  - More generally, could have $X_t = X_0 + \mu t + \sigma B_t$.
  - Then $dX_t = \mu\,dt + \sigma\,dB_t$. ($\mu$ = "drift"; σ = "volatility"; $\sigma \geq 0$)
  - Then $\mathbf{E}(X_t) = \mathbf{E}(X_0) + \mu t$, 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.
- NEXT: stock price model: $X_t = X_0 \exp(\mu t + \sigma B_t)$.
  - what is the "fair price" to have the <u>option</u> to buy the stock for $\$K$ at some fixed time $S > 0$?

—————————— **END OF WEEK #9** ——————————

[Homework #3 posted on web; due April 5 at 6:10pm sharp.]

[Final Exam: Thursday Apr 12, 7–10 p.m., Room 200, Brennan Hall. Bring student card.]

Summary of Previous Class:

∗ Brownian motion:

—— limit of rescaled s.s.r.w. as $M \to \infty$

—— $B_0 = 0$, $B_s - B_t \sim \text{Normal}(0, s - t)$

—— independent increments, continuous sample paths

—— martingales $\{B_t\}$, $\{B_t^2 - t\}$

—— hitting probs & expectations ("continuous-time gambler's ruin")

—— diffusions: $dX_t = \mu\,dt + \sigma\,dB_t$

Application – Financial Modeling:

- Common model for stock price: $X_t = X_0 \exp(\mu t + \sigma B_t)$.
  - i.e. if $Y_t = \log(X_t)$, then $Y_t = Y_0 + \mu t + \sigma B_t$, i.e. $dY_t = \mu\, dt + \sigma\, dB_t$.
  - That is, changes occur proportional to total price (makes sense).
- 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

$$D_t \;\equiv\; e^{-rt} X_t \;=\; e^{-rt}\, X_0 \exp(\mu t + \sigma B_t) \;=\; X_0 \exp((\mu - r)t + \sigma B_t)\,.$$

- QUESTION: what is the fair ("no-arbitrage") price of a "European call option" to have the option to buy the stock for \$$K$ at some fixed 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).
  - So what is the fair price at time 0?
  - Is it simply $\mathbf{E}[e^{-rS} \max(0,\, X_S - K)]$? No! Arbitrage!
- FACT: the fair price for the option equals $\mathbf{E}[e^{-rS} \max(0,\, X_S - K)]$, but only after replacing μ by $r - \frac{\sigma^2}{2}$.
  - i.e., such that $X_S = X_0 \exp([r - \frac{\sigma^2}{2}]S + \sigma\, B_S)$, where $B_S \sim \text{Normal}(0, S)$.
  - (Note: this assumes the ability to buy/sell arbitrary amounts of stock at any time, infinitely often, including going negative, with no transaction fees.)
  - WHY?? Well, if $\mu = r - \frac{\sigma^2}{2}$, then $\{D_t\}$ becomes a martingale (HW#3), and this turns out to be a key fact. (finance/actuarial classes . . .)
- So, fair price is now just an integral (with respect to a normal density).
  - After some computation (HW#3), this fair price becomes:

$$X_0\, \Phi\left(\frac{(r + \frac{\sigma^2}{2})S - \log(K/X_0)}{\sigma\sqrt{S}}\right) - e^{-rS} K\, \Phi\left(\frac{(r - \frac{\sigma^2}{2})S - \log(K/X_0)}{\sigma\sqrt{S}}\right),$$

    where $\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}}\, e^{-v^2/2}\, dv$ is the cdf of a standard normal distribution. ["Black-Scholes formula". Do not have to memorise!]

- Note: this price does not depend on the drift ("appreciation rate") μ. [Surprising! Intuition: if μ large, then can make good money from stock, so don't need the option.]
- However, it is an increasing function of the volatility σ. [Makes sense.]

Poisson Processes:

- MOTIVATING EXAMPLE:
  - Suppose an average of $\lambda = 2.5$ fires in Toronto per day.
  - Intuitively, this is caused by a very large number n of buildings, each of which has a very small probability p of having a fire.
  - Then mean $= np = \lambda$, so $p = \lambda/n$.
  - Then # fires today is Binomial$(n, \lambda/n) \approx$ 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 $n \geq 1$. ("n^{th} arrival time")
  - [e.g. T_n = time of n^{th} fire.]
- Let $N(t) = \max\{n \geq 0 : T_n \leq t\} = \#\{n \geq 1 : T_n \leq t\} = \#$ arrivals up to time t.
  - "Counting process". (Counts number of arrivals.)
  - [e.g. $N(t) = \#$ fires between times 0 and t.]
  - "Poisson process with intensity λ"
- What is distribution of $N(t)$, i.e. $\mathbf{P}(N(t) = m)$?
  - Well, $N(t) = m$ iff both $T_m \leq t$ and $T_{m+1} > t$, which is iff there is $0 \leq s \leq t$ with $T_m = s$ and $T_{m+1} - T_m > t - s$.
  - Recall that $Y_n \sim$ Exp$(\lambda) = \Gamma(1, \lambda)$, so $T_m := Y_1 + Y_2 + \ldots + Y_m \sim \Gamma(m, \lambda)$, 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,

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

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

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

- 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 $\mathbf{Var}(N(t)) = \lambda t$.

• Now, recall the "memoryless" (or "forgetfulness") property of the Exp(λ) distribution: for $a, b > 0$, $\mathbf{P}(Y_n > b + a \,|\, Y_n > a) = \mathbf{P}(Y_n > b) = e^{-\lambda b}$.

  - This means the process $\{N(t)\}$ "starts over" in each new time interval.
  - It follows that $N(t+s) - N(s) \sim N(t) \sim \text{Poisson}(\lambda t)$.
  - Also follows that if $0 \le a < b \le c < d$, then $N(d) - N(c)$ indep. of $N(b) - N(a)$, and similarly for multiple non-overlapping time intervals. ("independent increments")
  - MORE GENERALLY: if $0 \le t_1 \le s_1 \le t_2 \le s_2 \le \ldots \le t_k \le s_k$, then $N(s_i) - N(t_i) \sim \text{Poisson}(\lambda(s_i - t_i))$, and $\{N(s_i) - N(t_i)\}_{i=1}^k$ are all independent. ("independent Poisson increments")

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

• MOTIVATING EXAMPLE (cont'd): 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.

- ALTERNATIVE APPROACH: Given $N(t)$, as $h \searrow 0$,
  - $\mathbf{P}\big(N(t+h) - N(t) = 1\big) \ = \ \lambda h + o(h)$.
  - $\mathbf{P}\big(N(t+h) - N(t) \geq 2\big) \ = \ o(h)$.
  - 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/ped1 and www.probability.ca/ped2
  - Later, just <u>two</u> in Feb 1 - Mar 15, 2010; <u>less</u> than expected (4), but no media!
- 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 <u>regular</u>, 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 <u>completely random</u>, 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.
- Related Approach:
  - Suppose have λ buses per hour, with completely random arrival times.
  - Model this as $T_1, T_2, \ldots, T_n \sim$ Uniform$[0, \frac{n}{\lambda}]$, 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!
- Aside: What about <u>streetcars</u>?
  - They can't <u>pass</u> each other, so they sometimes clump up even <u>more</u> than do (independent) buses. (e.g. Spadina streetcar; see my research paper!)

——————————— **END OF WEEK #10** ———————————

[Return HW#2]

[Reminder: HW#3 due next class (Apr 5) at 6:10pm sharp.]

[Office hours: Wed Apr 4, 2:30 – 4:00]

Summary of Previous Class:
∗ Application: Financial Modeling
—— Black-Scholes formula
∗ Poisson process:
—— $N(0) = 0$, $N(s) - N(t) \sim \text{Poisson}(\lambda(s-t))$, indep incr
—— Computations, cond prob
—— Poisson clumping, examples
—— Waiting times

- 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 $A \sim \text{Exp}(\lambda)$.
  - Hence, $\mathbf{E}(A) = 1/\lambda = 1/8 = 7.5$ 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: $\mathbf{P}(N_1(t) = j,\ N_2(t) = k) = \Big(e^{-(\lambda p_1 t)}(\lambda p_1 t)^j / j!\Big)\ \Big(e^{-(\lambda p_2 t)}(\lambda p_2 t)^k / k!\Big)$.
- But $\mathbf{P}(N_1(t) = j,\ N_2(t) = k) = \mathbf{P}(j + k$ arrivals up to time t, of which j of type 1 and k of type 2$) = \Big(e^{-\lambda t}(\lambda t)^{j+k}/(j+k)!\Big)\ \binom{j+k}{j}(p_1)^j(p_2)^k$. Equal!

- 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.
  - See e.g. www.probability.ca/pois

Continuous-Time Markov Processes:

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

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

  for some initial distribution $\{\nu_i\}_{i\in S}$ (with $\nu_i \geq 0$, and $\sum_{i\in S} \nu_i = 1$), and transition probabilities $\{p_{ij}^{(t)}\}_{\substack{i,j\in S\\ t\geq 0}}$ (with $p_{ij}^{(t)} \geq 0$, and $\sum_{j\in S} p_{ij}^{(t)} = 1$).
  - Just like for discrete-time chains, except need to keep track of the elapsed time (t) too.
  - As with discrete chains, $p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$
- Let $P^{(t)} = \big(p_{ij}^{(t)}\big)_{i,j\in S}$ = 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 $\mu^{(t)} = \left(\mu_i^{(t)}\right)_{i\in S}$ = 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.

- <u>Expect</u> that $\lim_{t\searrow 0} p_{ij}^{(t)} = p_{ij}^{(0)} = \delta_{ij}$.
  - <u>Assume</u> this is true. ("standard" Markov process)
- Then can compute the process's <u>generator</u> 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 $g_{ii} \le 0$, while $g_{ij} \ge 0$ for $i \ne j$.
  - 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 $p_{11}^{(0.02)} \doteq 1 - 3(0.02) = 0.94$, $p_{12}^{(t)} \doteq 3(0.02) = 0.06$, $p_{21}^{(t)} \doteq 6(0.02) = 0.12$, 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 $m \in \mathbf{N}$,

$$P^{(t)} \;=\; [P^{(t/m)}]^m \;=\; \lim_{n\to\infty} [P^{(t/n)}]^n \;=\; \lim_{n\to\infty} [I + (t/n)G]^n$$

$$=\; \exp(tG) \;:=\; I + tG + \frac{t^2G^2}{2!} + \frac{t^3G^3}{3!} + \ldots$$

(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 <u>all</u> $t \ge 0$.

- Can we actually <u>compute</u> $P^{(t)} = \exp(tG)$ this way? Yes!
- Method #1: Compute the infinite matrix sum on a computer (approximately).
- Method #2: Note that in above example, if $\lambda_1 = 0$ and $\lambda_2 = -9$, and $v_1 = (2,1)$ and $v_2 = (1,-1)$, then $v_1 G = \lambda_1 v_1 = 0$, and $v_2 G = \lambda_2 v_2 = -9 v_2$.
  - That is, $\{\lambda_i\}$ are the <u>eigenvalues</u> of G, with corresponding <u>left-eigenvectors</u> $\{v_i\}$.
  - So, if initial distribution is (say) $\nu = (1,0)$, then $\nu = \frac{1}{3}v_1 + \frac{1}{3}v_2$, so

$$\mu^{(t)} \;=\; \nu\,P^{(t)} \;=\; \nu\,\exp(tG) \;=\; (\frac{1}{3}v_1 + \frac{1}{3}v_2)\,\exp(tG)$$

$$=\; \frac{1}{3}e^{t\lambda_1}v_1 + \frac{1}{3}e^{t\lambda_2}v_2 \;=\; \frac{1}{3}e^{0\,t}(2,1) + \frac{1}{3}e^{-9t}(1,-1) \;=\; (\frac{2+e^{-9t}}{3},\;\frac{1-e^{-9t}}{3})\,.$$

  - 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 $p_{11}^{(t)} + p_{22}^{(t)} = 1$. (Phew.)

- 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,

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

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

  - 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 $p_{11}^{(t)} = \frac{2}{3} + \frac{1}{3}e^{-9t} = \frac{2+e^{-9t}}{3}$.

– Same answer as before. (Phew.)

- NOTE: you should learn at least one of Method #2 and Method #3.
- 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 again $\sum_{i\in S} \pi_i g_{ij} = 0$.
  - 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 $\sum_i \pi_i\, g_{ij} = \sum_i \pi_j\, g_{ji} = 0$, so π is stationary (similar to discrete case).

——————————— **END OF WEEK #11** ———————————

[Collect HW #3.]

[REMINDER: Final Exam, Thursday Apr 12, 7–10 p.m., Room 200, Brennan Hall. Bring your student card. No aids allowed.]

[Office hours: Wed Apr 11, 2:30 – 4:00?]

Summary of Previous Class:
∗ More on Poisson Processes:
—— Superposition, Thinning
∗ Continuous-Time Markov Processes:
—— $\{X_t\}_{t\geq 0}$, $p_{ij}^{(t)}$, etc.
—— Generator: $G = P'^{(0)} = \lim_{t\searrow 0} \frac{P^{(t)}-I}{t}$.
—— For small $t > 0$, $P^{(t)} \approx I + tG$.

—— For <u>all</u> $t \geq 0$, $P^{(t)} = \exp(tG)$.
—— Can compute this using eigenvectors, or diff eq's
—— $\{\pi_i\}$ is stationary dist. iff $\sum_{i \in S} \pi_i g_{ij} = 0$.
—— <u>Suffices</u> that $\pi_i g_{ij} = \pi_j g_{ji}$ for all $i, j \in S$. ("reversible")

- Is convergence to $\{\pi_i\}$ guaranteed?
  - THM: if continuous-time M.C. is irreducible and has stationary dist 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., automatically aperiodic).
- CONNECTION TO DISCRETE-TIME MARKOV CHAINS:
  - Let $\{\hat{p}_{ij}\}_{i,j \in S}$ be the transition probabilities for a <u>discrete-time</u> 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 (instead of at integer times). ("Exponential holding times")
  - In particular, $\{X_t\}$ is a continuous-time Markov process! So, can "create" a continuous-time Markov process from a discrete-time Markov chain.
  - (Special case: if $\hat{X}_0 = 0$, and $\hat{p}_{i,i+1} = 1$ for all i, then $X_t = N(t) =$ Poisson process.)
  - What is the <u>generator</u> of this Markov process?
  - Well, here $p_{ij}^{(t)} = \sum_{n=0}^{\infty} \mathbf{P}[N(t) = n] \, \hat{p}_{ij}^{(n)} = \sum_{n=0}^{\infty} [e^{-\lambda t} \frac{(\lambda t)^n}{n!}] \, \hat{p}_{ij}^{(n)}$.
  - So, for small $t > 0$ and $i \neq j$, $p_{ij}^{(t)} \approx \mathbf{P}[N(t) = 1] \, \hat{p}_{ij} \approx [t\lambda] \hat{p}_{ij}$, so $g_{ij} = \lambda \, \hat{p}_{ij}$.
  - Also $p_{ii}^{(t)} \approx \mathbf{P}[N(t) = 0] + \mathbf{P}[N(t) = 1] \, \hat{p}_{ii} \approx [1 - t\lambda] + [t\lambda] \hat{p}_{ii}$, so $g_{ii} = \lambda \, (\hat{p}_{ii} - 1) \leq 0$.
  - Hence, $g_{ii} = -\lambda \sum_{k \neq i} \hat{p}_{ik}$, so again $\sum_{j \in S} g_{ij} = \lambda \sum_{j \neq i} \hat{p}_{ij} - \lambda \sum_{k \neq i} \hat{p}_{ik} = 0$.
- Specific example: the Poisson process $\{N(t)\}_{t \geq 0}$ itself has generator:

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

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.
  - Let C_n = time taken to serve the n^{th} customer.
- For all $t \geq 0$, let $Q(t)$ = number of customers in the system (i.e., waiting in the queue or being served) at time t.
  - What happens to $Q(t)$ as $t \to \infty$?
- M/M/1 QUEUE: $T_n - T_{n-1} \sim \text{Exp}(\lambda)$, and $C_n \sim \text{Exp}(\mu)$. (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 $n \geq 0$, to first order as $t \searrow 0$,

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

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

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

  - Also, for $n \geq 1$,

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

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

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

    since $1 - e^{-\mu t} \approx 1 - [1 + (-\mu)t] = \mu t$.
  - Also if $|n - m| \geq 2$ then $\mathbf{P}[Q(t) = m \,|\, Q(0) = n] = O(t^2) = o(t)$, so $g_{n,m} = 0$.
  - But $\sum_{m=0}^{\infty} g_{n,m} = 0$, so generator must be given by:

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

i.e. $g_{00} = -\lambda$ and $g_{nn} = -\lambda - \mu$ for $n \geq 1$. (Zero arrivals and zero served.)

- STATIONARY DISTRIBUTION $\{\pi_i\}$?
  - Need $\sum_{i \in S} \pi_i g_{ij} = 0$ for all $j \in S$. (Or, can use reversibility: 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 $\pi_2 = (-\frac{\lambda}{\mu})\pi_0 + \pi_1 + \frac{\lambda}{\mu}\pi_1 = (\frac{\lambda}{\mu})^2\pi_0$.
  - 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. $\mathbf{E}(C_n) < \mathbf{E}(T_n - T_{n-1})$, then

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

    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 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.
  - (If $\lambda = \mu$, then $Q(t) \to \infty$ in probability, but not w.p. 1.)
- GENERAL (G/G/1) QUEUE: What if we don't assume Exponential distributions, just that $\{T_n - T_{n-1}\}$ i.i.d., and $\{C_n\}$ i.i.d. (all indep.)?
- In that case, we have the following FACTS:
  - if $\mathbf{E}(T_n - T_{n-1}) < \mathbf{E}(C_n)$, then $Q(t) \to \infty$ w.p. 1.
  - if $\mathbf{E}(T_n - T_{n-1}) > \mathbf{E}(C_n)$, then $\{Q(t)\}$ remains "bounded in probability", i.e. for any $\epsilon > 0$ there is $K < \infty$ such that $\mathbf{P}(Q(t) > K) < \epsilon$ for all $t \geq 0$.
  - if $\mathbf{E}(T_n - T_{n-1}) = \mathbf{E}(C_n)$, then $Q(t) \to \infty$ in probability (but not w.p. 1).

END OF WEEK #12

[REMINDER: Final Exam is Thursday Apr 12, 7–10 p.m., Room 200, Brennan Hall. Bring your student card. No aids allowed. Good luck and best wishes! – J.R.]