Stat 150 Stochastic Processes **Spring 2009**

Lecture 1: To be determined

Lecturer: Jim Pitman

- **Stochastic Process**: Random process, evolution over time/space. Especially models for sequences of random variables:
  - $X_0, X_1, X_2, \ldots$
  - Time $= \{0, 1, 2, \ldots\}$
  - Random vector $(X_0, X_1, \ldots, X_N)$
  - $(X_t, t \in \{0, 1, \ldots, N\})$
  - $(X_t, 0) \quad$ Time $= [0, 1]$
- In background always have
  - Probability space: $(\Omega, \mathcal{F}, \mathbb{P}) \sim$ TRIPLE
  - Set $\Omega =$ "set of all possible outcomes"
  - Collection $\mathcal{F}$ of "events" = subsets of Ω
  - Assume $\mathcal{F}$ is closed under $\bigcup, \bigcap$, complement, and countable set operations.
- If $A_1, A_2, \ldots$ is a sequence of events, we
  - consider $A_1 \bigcup A_2 \bigcup A_3 \ldots$ is an event;
  - **Axiom:** If the A_i's are disjoint, $A_i \bigcap A_j = \emptyset, i \neq j$, then $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.
- **Measurability.** $\mathbb{P}\{\omega : X(\omega) \leq x\} = F(x)$. F is called the cumulative distribution function of X. $\mathbb{P}(X \leq x) = F(x)$.
- Discrete value: List of values $x_0, x_1, x_2 \ldots$, often $\{0, 1, 2, 3, \ldots\}$.

  $\mathbb{P}(X = x_i) = p_i$. If $\sum_{i=0}^{\infty} \mathbb{P}_i = 1$, we say X has discrete distribution with probabilities (p_i) at values (x_i).

  $\mathbb{E}(X) = \sum_i x_i p_i$. Technical issue: $\sum_i x_i p_i := \lim_{N \to \infty} \sum_{i=0}^{N} x_i p_i$, we need to assume the limit exists, i.e., $\sum_{i=0}^{\infty} |x_i| p_i < \infty$.

Other cases:

$$F(x) = \mathbb{P}(X \le x)$$

Density case: $$F(x) = \int_{-\infty}^{x} f(y)dy$$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} g(x)f(x)dx \qquad \text{provided } \int |g(x)|f(x)dx = \mathbb{E}|g(X)| < \infty$$

- We start with naïve ideas of $\mathbb{E}(X)$ defined by $\sum$ and $\int$.
  We often encounter situations where

$$0 \le X_n \uparrow X$$
$$0 \le X_1 \le X_2 \le \ldots \uparrow X$$
$$0 \le \mathbb{E}(X_n) \uparrow \mathbb{E}(X)$$

Claim: $X \le Y \Longrightarrow \mathbb{E}(X) \le \mathbb{E}(Y)$.
Proof: $Y = X + Y - X$. $\mathbb{E}(Y) = \mathbb{E}(X) + \mathbb{E}(Y - X) \ge \mathbb{E}(X)$.

Monotone Convergence Theorem: If $0 \le X_n \le\uparrow X$, then $0 \le \mathbb{E}(X_n) \uparrow \mathbb{E}(X)$.

- **Conditional Probability**

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)} \qquad \text{provided } \mathbb{P}(B) = 0$$

Easily for a discrete random variable X we can define

$$\mathbb{E}(Y|X = x) \begin{cases} = \sum_i y_i d\mathbb{P}(Y = y_i|X = x) & \text{discrete} \\ = \int_y y d\mathbb{P}(Y = y|X = x) & \text{density} \end{cases}$$

Suppose you know values of $\mathbb{E}(Y|X = x)$ for all x of X, how to find $\mathbb{E}(Y)$?

$$\mathbb{E}(Y) = \sum_x \underbrace{\mathbb{E}(Y|X = x)}_{\varphi(x)} \cdot \mathbb{P}(X = x)$$
$$= \mathbb{E}(\varphi(x))$$

For a discrete X, $\mathbb{E}(Y|X)$ is a random variable, whereas $\mathbb{E}(Y|X=x)$ is a value. So

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X))$$

- Exercises with **random sums**

$X_1, X_2 \ldots$ is a sequence of independent and identically distributed random variables. $S_n := X_1 + \cdots + X_n$, N is a random variable. For simplicity, $N \perp X_1, X_2, \ldots$ Want a formula for $\mathbb{E}(X_N) = \mathbb{E}(X_1 + \cdots + X_N) = \mathbb{E}(\mathbb{E}(S_N|N))$.

$$\mathbb{E}(S_N|N=n) = \mathbb{E}(X_1 + \cdots X_n) = n\mathbb{E}(X_1)$$

$$\implies \mathbb{E}(\mathbb{E}(S_N|N)) = \mathbb{E}(N\mathbb{E}(X_1)) = \mathbb{E}(N)\mathbb{E}(X_1)$$

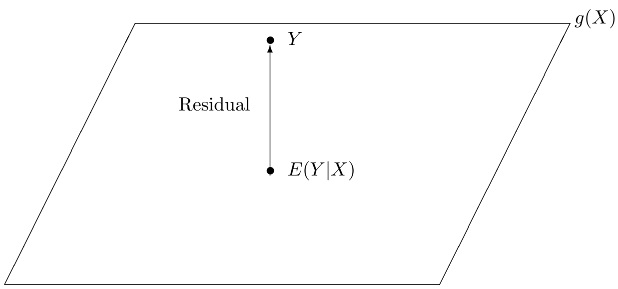
<u>Stopping time</u>: Rule for deciding when to quit.

Stat 150 Stochastic Processes **Spring 2009**

Lecture 2: Conditional Expectation

Lecturer: Jim Pitman

- **Some useful facts** (assume all random variables here have finite mean square):
  • $\mathbb{E}(Yg(X)|X) = g(X)\mathbb{E}(Y|X)$
  • $Y - \mathbb{E}(Y|X)$ is orthogonal to $\mathbb{E}(Y|X)$, and orthogonal also to $g(X)$ for every measurable function g.
  • Since $\mathbb{E}(Y|X)$ is a measurable function of X, this characterizes $\mathbb{E}(Y|X)$ as the *orthogonal projection* of Y onto the linear space of all square-integrable random variables of the form $g(X)$ for some measurable function g.
  • Put another way $g(X) = \mathbb{E}(Y|X)$ minimizes the mean square prediction error $\mathbb{E}[(Y - g(X))^2]$ over all measurable functions g.



These facts can all be checked by computations as follows: Check orthogonality:

$$\begin{aligned}
\mathbb{E}[(Y - \mathbb{E}(Y|X))g(X)] &= \mathbb{E}(g(X)Y - g(X)\mathbb{E}(Y|X)) \\
&= \mathbb{E}(g(X)Y) - \mathbb{E}(g(X)\mathbb{E}(Y|X)) \\
&= \mathbb{E}(\mathbb{E}(g(X)Y|X)) - \mathbb{E}(g(X)\mathbb{E}(Y|X)) \\
&= \mathbb{E}(g(X)\mathbb{E}(Y|X)) - \mathbb{E}(g(X)\mathbb{E}(Y|X)) \\
&= 0
\end{aligned}$$

- Recall: $\mathbf{V}ar(Y) = \mathbb{E}(Y - \mathbb{E}(Y))^2$ and $\mathbf{V}ar(Y|X) = \mathbb{E}([Y - \mathbb{E}(Y|X)]^2|X)$.
  Claim: $\mathbf{V}ar(Y) = \mathbf{V}ar(\mathbb{E}(Y|X)) + \mathbb{E}(\mathbf{V}ar(Y|X))$

Proof:

$$
\begin{aligned}
Y &= \mathbb{E}(Y|X) + Y - \mathbb{E}(Y|X) \\
\mathbb{E}(Y^2) &= \mathbb{E}(\mathbb{E}(Y|X)^2) + \mathbb{E}([Y - \mathbb{E}(Y|X)]^2) + 0 \\
&= \mathbb{E}(\mathbb{E}(Y|X)^2) + \mathbb{E}(\mathbf{V}ar(Y|X)) \\
\mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 &= \mathbb{E}(\mathbb{E}(Y|X)^2) - (\mathbb{E}(Y))^2 + \mathbb{E}(\mathbf{V}ar(Y|X)) \\
&= \mathbb{E}(\mathbb{E}(Y|X)^2) - (\mathbb{E}(\mathbb{E}(Y|X)))^2 + \mathbb{E}(\mathbf{V}ar(Y|X))
\end{aligned}
$$

$$\Longrightarrow \mathbf{V}ar(Y) = \mathbf{V}ar(\mathbb{E}(Y|X)) + \mathbb{E}(\mathbf{V}ar(Y|X))$$

- Exercise P. 84, 4.3

  T is uniform on [0,1]. Given T, U is uniform on $[0, T]$. What is $\mathbb{P}(U \geq 1/2)$?

$$
\begin{aligned}
\mathbb{P}(U \geq 1/2) &= \mathbb{E}(\mathbb{E}[\mathbf{1}(U \geq 1/2)|T]) \\
&= \mathbb{E}[\mathbb{P}(U \geq 1/2|T)] \\
&= \mathbb{E}\left[\frac{T - 1/2}{T}\mathbf{1}(T \geq 1/2)\right] \\
&= \int_{1/2}^{1} \frac{t - 1/2}{t} dt
\end{aligned}
$$

- **Random Sums:** Random time T. $S_n = X_1 + \cdots + X_n$. Wants a formula for $\mathbb{E}(S_T)$ which allows that T might not be independent of $X_1, X_2, \ldots$.

  Condition: For all $n = 1, 2, \ldots$ the event $(T = n)$ is determined by $X_1, X_2, \ldots, X_n$.
  Equivalently: $(T \leq n)$ is determined by $X_1, X_2, \ldots, X_n$.
  Equivalently: $(T > n)$ is determined by $X_1, X_2, \ldots, X_n$.
  Equivalently: $(T \geq n)$ is determined by $X_1, X_2, \ldots, X_{n-1}$.
  Call such a T a *stopping time* relative to the sequence $X_1, X_2, \ldots$
  Example: The first n (if any) such that $S_n \leq 0$ or $S_n \geq b$. Then $(T = n) = (S_1 \in (0, b), S_2 \in (0, b), \ldots, S_{n-1} \in (0, b), S_n \notin (0, b))$ is a function of $S_1, \ldots, S_n$.

- **Wald's identity:** If T is a stopping time relative to $X_1, X_2, \ldots$, which are i.i.d. and $S_n := X_1 + \cdots + X_n$, then $\mathbb{E}(S_T) = \mathbb{E}(T)\mathbb{E}(X_1)$, provided $\mathbb{E}(T) < \infty$.

Sketch of proof:

$$\begin{aligned} S_T &= X_1 + \cdots + X_T \\ &= X_1 \mathbf{1}(T \geq 1) + X_2 \mathbf{1}(T \geq 2) + \cdots \end{aligned}$$

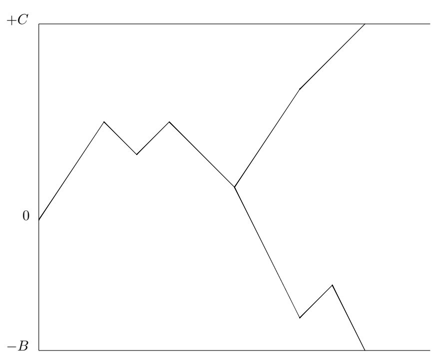
$$\begin{aligned} \mathbb{E}(S_T) &= \mathbb{E}(X_1 \mathbf{1}(T \geq 1) + \mathbb{E}(X_2 \mathbf{1}(T \geq 2)) + \cdots \\ &= \mathbb{E}(X_1) + \mathbb{E}(X_2)\mathbb{P}(T \geq 2) + \mathbb{E}(X_3)\mathbb{P}(T \geq 3) + \cdots \\ &= \mathbb{E}(X_1) \sum_{n=1}^{\infty} \mathbb{P}(T \geq n) \\ &= \mathbb{E}(X_1)\mathbb{E}(T) \end{aligned}$$

Key point is that for each n the event $(T \geq n)$ is determined by $X_1, X_2, \ldots, X_{n-1}$, hence is independent of X_n. This justifies the factorization

$$\mathbb{E}(X_n 1(T \geq n) = \mathbb{E}(X_n)\mathbb{E}(\mathbf{1}(T \geq n)) = \mathbb{E}(X_1)\mathbb{P}(T \geq n).$$

It is also necessary to justify the swap of $\mathbb{E}$ and Σ. This is where $\mathbb{E}(T) < \infty$ must be used in a more careful argument. Note that if $X_i \geq 0$ the swap is justified by monotone convergence.

- Example. Hitting probabilities for simple symmetric random walk



$S_n = X_1 + \cdots X_n$, $X_i \sim \pm 1$ with probability 1/2,1/2. T= first n s.t. $S_n = +C$ or $-B$.

It is easy to see that $\mathbb{E}(T) < \infty$. Just consider successive blocks if indices of length $B + C$. $\underbrace{\qquad}_{B+C}\underbrace{\qquad}_{B+C}\underbrace{\qquad}_{B+C}\underbrace{\qquad}_{\ldots}$ Wait until a block of length $B + C$ with $X_i = 1$ for all i in the block. Geometric distribution of this upper bound on $T \Longrightarrow \mathbb{E}(T) < \infty$.

Let $p+ = \mathbb{P}(S_T = +C)$ and $p_- := \mathbb{P}(S_T = -C)$. Then

$$\begin{aligned}
\mathbb{E}(S_T) &= \mathbb{E}(T)\mathbb{E}(X_1) = \mathbb{E}(T) \cdot 0 = 0 \\
0 &= p_+ C - p_- B \\
1 &= p_+ + p_- \\
\implies p_+ &= \frac{B}{B+C} \qquad p_- = \frac{C}{B+C}
\end{aligned}$$

Stat 150 Stochastic Processes **Spring 2009**

Lecture 3: Martingales and hitting probabilities for random walk

Lecturer: Jim Pitman

- A sequence of random variables $S_n, n \in \{0, 1, 2, \dots\}$, is a *martingale* if

$$1)\ \mathbb{E}|S_n| < \infty \qquad \text{for each } n = 0, 1, 2, \dots$$
$$2)\ \mathbb{E}(S_{n+1}|(S_0, S_1, \dots, S_n)) = S_n$$

E.g., $S_n = S_0 + X_1 + \cdots + X_n$, where X_i are differences, $X_i = S_i - S_{i-1}$. Observe that (S_n) is a martingale if and only $\mathbb{E}(X_{n+1}|S_0, S_1, \dots, S_n) = 0$.

Then say (X_i) is a martingale difference sequence.

Easy example: let X_i be independent random variables (of each other, and of S_0), where $\mathbb{E}(X_i) \equiv 0 \Longrightarrow X_i$ are martingale differences $\Longrightarrow S_n$ is a martingale.

- Some general observations:

If S_n is a martingale, then $\mathbb{E}(S_n) \equiv \mathbb{E}(S_0)$ (Constant in n).
Proof: By induction, suppose true for n, then

$$\mathbb{E}(S_{n+1}) = \mathbb{E}(\mathbb{E}(S_{n+1}|(S_0, S_1, \dots, S_n))) = \mathbb{E}(S_n) = \mathbb{E}(S_0)$$

- Illustration: Fair coin tossing walk with absorption at barriers 0 and b. Process: Start at a. Run until the sum of ± 1's hits 0 or b, then freeze the value.

  - By construction, $0 \le S_n \le b \Longrightarrow \mathbb{E}|S_n| \le b < \infty$.
  - Given $S_0, S_1, \dots, S_n$, with $0 < S_i < b$ for $0 \le i \le n$, we assume that

$$S_{n+1} = \begin{cases} S_n + 1 & \text{with probability } 1/2 \\ S_n - 1 & \text{with probability } 1/2 \end{cases}$$

So

$$\begin{aligned} &\mathbb{E}(S_{n+1}|S_0, S_1, \dots, S_n \text{ with } 0 < S_i < b, 0 \le i \le n) \\ =&\frac{1}{2}(S_n + 1) + \frac{1}{2}(S_n - 1) = S_n \end{aligned}$$

whereas given $S_0, S_1, \dots, S_n$, with either $S_n = 0$ or $S_n = b$, then $S_{n+1} = S_n$. So also

$$\mathbb{E}(S_{n+1}|S_0, S_1, \dots, S_n \text{ with } S_i = 0 \text{ or } b) = S_n$$

Since the only possible values of S_n are in the range from 0 to b, no matter what the value of S_n,

$$\mathbb{E}(S_{n+1}|S_0, S_1, \dots, S_n) = S_n$$

So the fair random walk with absorbtion at barriers 0 and b is a martingale.

- Revisit the Gambler's ruin problem for a fair coin. Observe

$$(S_n = b) \Longrightarrow (S_{n+1} = b)$$
$$\mathbb{P}(S_n = b) \le \mathbb{P}(S_{n+1} = b)$$

$\mathbb{P}(S_n = b)$ is an increasing sequence which is bounded above by 1.
<u>Proof</u>: We know $\mathbb{E}(S_n) \equiv a$. But

$$a = \mathbb{E}(S_n) = 0 \cdot \mathbb{E}(S_n = 0) + b \cdot \mathbb{P}(S_n = b) + \sum_{i=1}^{b-1} i \cdot \mathbb{P}(S_n = i)$$

Observe that $\sum_{i=1}^{b-1} \mathbb{P}(S_n = i) = \mathbb{P}(T > n) \to 0$ as $n \to \infty$, where T is the time to hit the barriers. Then

$$\mathbb{P}(S_n = b) = \frac{a}{b} - \frac{\epsilon_n}{b} \to \frac{a}{b} \text{ as } n \to \infty$$

- <u>Gambler's ruin for an unfair coin $p \uparrow q \downarrow,\ p + q = 1.$</u>

Idea: Choose an increasing function g so that we make the game fair; that is, we will make an $M_n = g(S_n)$ so that (M_n) is a MG.

Look at r^{S_n} instead of S_n for some ratio r to be determined. Suppose you are given $r^{S_0}, \dots r^{S_n}$, with $0 < S_n < b$, then

$$r^{S_{n+1}} = \begin{cases} r^{S_n} \times r & \text{with prob } p \\ r^{S_n} \times \frac{1}{r} & \text{with prob } q \end{cases}$$

Compute

$$\mathbb{E}(r^{S_{n+1}}|r^{S_1}, \dots r^{S_n}) = r^{S_n}(rp + \frac{1}{r}q) = r^{S_n}$$
$$\Longleftrightarrow rp + \frac{1}{r}q = 1 \Longleftrightarrow r = \frac{q}{p} \text{ or } r = 1.$$

So $r = q/p$ implies r^{S_n} is a martingale. Again, the walk is stopped if it reaches the barriers at 0 or b, and given such values the martingale property for the next step is obvious. The martingale property gives: $\mathbb{E}r^{S_n} \equiv r^a$ if $S_0 = a$.

$$(\frac{q}{p})^a = (\frac{q}{p})^0\mathbb{P}(S_n = 0) + (\frac{q}{p})^b\mathbb{P}(S_n = b) + \sum_{i=1}^{b-1}(\frac{q}{p})^i\mathbb{P}(S_n = i)$$

Let $P_{\text{up}} = \lim_{n\to\infty} \mathbb{P}(S_n = b),\ 1 - P_{\text{up}} = \lim_{n\to\infty} \mathbb{P}(S_n = 0)$,

$$(\frac{q}{p})^a = (\frac{q}{p})^0(1 - P_{\text{up}}) + (\frac{q}{p})^b P_{\text{up}}$$
$$\implies P_{\text{up}} = \frac{(q/p)^a - 1}{(q/p)^b - 1}$$

Case $p > q$, fix a and let $b \to \infty$,

$$\lim_{b\to\infty} P_{\text{up}} = 1 - (\frac{q}{p})^a$$
$$\lim_{b\to\infty} P_{\text{down}} = (\frac{q}{p})^a$$

Stat 150 Stochastic Processes **Spring 2009**

Lecture 4: Conditional Independence and Markov Chain

Lecturer: Jim Pitman

1 Conditional Independence

- Q: If X and Y were conditionally independent given λ, are X and Y independent? (Typically no.)

- Write $X \overset{Y}{\perp} Z$ to indicate X and Z are conditionally independent given Y.
  Assume for simplicity all variables are discrete, then $X \overset{Y}{\perp} Z$ means

  $$\mathbb{P}(X = x, Z = z|Y = y) = \mathbb{P}(X = x|Y = y)\mathbb{P}(Z = z|Y = y)$$

  for all possible values of x, y, and z.

- Suppose $X \overset{Y}{\perp} Z$, **claim:** $\mathbb{P}(Z = z|X = x, Y = y) = P(Z = z|Y = y)$.
  (Key point: The value of X is irrelevant.)

  Proof: From $\mathbb{P}(X = x, Z = z|Y = y) = \mathbb{P}(X = x|Y = y)\mathbb{P}(Z = z|Y = y)$, we get

  $$\begin{aligned}\mathbb{P}(Z = z|Y = y) &= \frac{\mathbb{P}(X = x, Z = z|Y = y)}{\mathbb{P}(X = x|Y = y)} \\ &= \frac{\mathbb{P}(X = x, Y = y, Z = z)/\mathbb{P}(Y = y)}{\mathbb{P}(X = x, Y = y)/P(Y = y)}\end{aligned}$$

  Also $\mathbb{P}(Z = z|X = x, Y = y) = \dfrac{\mathbb{P}(X = x, Y = y, Z = z)}{\mathbb{P}(X = x, Y = y)}$ ■

- $Z \overset{Y}{\perp} X$ is the same as $X \overset{Y}{\perp} Z$. $\mathbb{P}(X = x|Z = z, Y = y) = P(X = x|Y = y)$, $\forall x, y, z$ is a further extension.

- Suppose $X \overset{Y}{\perp} Z$, then $\mathbb{E}[g(Z)|X, Y] = \mathbb{E}[g(Z)|Y]$.
  Proof is by application of $\mathbb{P}(Z = z|X = x, Y = y) = P(Z = z|Y = y)$

2 Markov Chain

- **Definition:** A sequence of r.v.'s $(X_0, X_1, \dots, X_n)$ is called a Markov chain if for every $1 \le m < n$, $(X_0, X_1, \dots, X_{m-1})_{\text{(past)}} \overset{X_{m\,\text{(present)}}}{\perp} X_{m+1\,\text{(future)}}$.

  In other words, at each stage, to describe the conditional distribution of X_{m+1} given $(X_0, X_1, \dots, X_m)$, you just need to know X_m.

  Often it is assumed further that the conditional distribution of X_{m+1} given $X_m = x$ depends only on x, not on m. These are **homogeneous transition probabilities**.

- MC's with homogenous transition probabilities are specified by a matrix P. Various notations for $P : P = P_{ij} = P(i, j)$.

  Assuming discrete values for $X_0, X_1, \dots,$ then

  $$\mathbb{P}(X_{m+1}|X_m = i, X_0 = i_0, X_1 = i_1, \dots, X_{m-1} = i_{m-1}) = P_{ij} = P(i, j)$$

  Write the unconditional probability:

  $$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_0 = i_0)P(i_0, i_1) \cdots P(i_{n-1}, i_n)$$

- Consider the matrix $P(i, j)$. Fix i and sum on j, and get $\sum_j P(i, j) = 1$.

- Trivial case: IID $\leftrightarrow$ P with all rows identical.

- Notice that we need a state space of values S — one row and one column for each stat $i \in S$.

- The textbook uses $S = \{0, 1, 2, \dots\}$. However, sometimes it is unnatural to have a state space of nonnegative integers. e.g. shuffling a deck of cards. State space = 52! orderings of the deck.

- **Transition diagram** v.s. **transition matrix**

- Examples of Markov chains: random walks of graphs, webpage rankings.

- **n-step transition probability of a MC**

  <u>Problem</u>: Know $P(i, j) = \mathbb{P}(X_{m+1} = j|X_m = i), \forall i, j$, want to compute $\mathbb{P}(X_n = j|X_0 = i) = \mathbb{P}(X_{m+n} = j|X_m = i)$.

<u>Method</u>: For n = 2,

$$\begin{aligned}\mathbb{P}(X_2 = j|X_0 = i) &= \frac{\mathbb{P}(X_0 = i, X_2 = j)}{\mathbb{P}(X_o = i)} \\ &= \frac{\sum_k \mathbb{P}(X_0 = i, X_1 = k, X_2 = j)}{\mathbb{P}(X_0 = i)} \\ &= \frac{\sum_k \mathbb{P}(X_0 = i)P(i,k)P(k,j)}{P(X_0 = i)} \\ &= \sum_k P(i,k)P(k,j)\end{aligned}$$

Notice that $\mathbb{P}(X_2 = j|X_0 = i) = P^2(i, j)$.

Claim: $\mathbb{P}(X_n = j|X_0 = i) = P^n(i, j)$. **Proof:** The case for n = 2 has been proved. Suppose true for n, then

$$\begin{aligned}\mathbb{P}(X_{n+1} = j|X_0 = i) &= \frac{\mathbb{P}(X_0 = i, X_{n+1} = j)}{\mathbb{P}(X_o = i)} \\ &= \frac{\sum_k \mathbb{P}(X_0 = i, X_1 = k, X_{n+1} = j)}{\mathbb{P}(X_0 = i)} \\ &= \frac{\sum_k \mathbb{P}(X_0 = i)P(i,k)P^n(k,j)}{P(X_0 = i)} \\ &= \sum_k P(i,k)P^n(k,j) = P^{n+1}(i,j)\end{aligned}$$

■

- **Action of matrix P on row vectors**

  Suppose X_0 has distribution λ on state space S. $\mathbb{P}(X_0 = i) = \lambda_i, i \in S$, $P = P(i, j), i, j \in S$, then

$$\mathbb{P}(X_1 = j) = \sum_{i\in S} \mathbb{P}(X_0 = i, X_1 = j) = \sum_{i\in S} \lambda_i P(i,j) = (\lambda P)_j$$

  Repeating this process, we get distribution of X_n is λP^n.

- **Action of matrix P on column vectors**

  Given a function f on the state space S, then

$$\mathbb{E}(f(X_1)|X_0 = i) = \sum_j f(j)\mathbb{P}(X_1 = j|X_0 = i) = \sum_j P(i,j)f(j) = (Pf)_i$$

Stat 150 Stochastic Processes **Spring 2009**

Lecture 5: Markov Chains and First Step Analysis

Lecturer: Jim Pitman

1 Further Analysis of Markov Chains

- Q: Suppose $\lambda =$ row vector, $f =$ column vector, and P is a probability transition matrix, then what is the meaning of $\lambda P^n f$? Here $\lambda P^n f$ is the common value of $(\lambda P^n)f = \lambda(P^n f)$ for usual operations involving vectors and matrices. The above equality is elementary for finite state space. For countable state space it should be assumed e.g. that both λ and f are non-negative. Assume that λ is a probability distribution, i.e. $\lambda_i >= 0$ and $\sum_i \lambda_i = 1$. Suppose then that $X_0, X_1, \dots, X_n$ is a Markov chain and the distribution of X_0 is λ: $\mathbb{P}(X_0 = i) = \lambda_i$. Then $(\lambda P^n)_j = \mathbb{P}(X_n = j)$. So

$$\begin{aligned} \lambda P^n f &= \sum_j (\lambda P^n)_j f(j) \\ &= \sum_j \mathbb{P}(X_n = j) f(j) \\ &= \mathbb{E} f(X_n) \end{aligned}$$

  So the answer is: $\lambda P^n f = \mathbb{E} f(X_n)$ for a Markov chain with initial distribution λ and transition probability matrix P.

- Notice that the same conclusion is obtained by first recognizing that $P^n f(i) = \mathbb{E}[f(X_n)|X_0 = i]$:

$$\begin{aligned} \lambda P^n f &= \lambda(P^n f) \\ &= \sum_i \lambda_i (P^n f)_i \\ &= \sum_i \lambda_i \mathbb{E}[f(X_n)|X_0 = i] \\ &= \mathbb{E}[\mathbb{E}[f(X_n)|X_0]] \\ &= \mathbb{E} f(X_n) \end{aligned}$$

- Q: What sort of functions of a Markov chain give rise to Martingales?

  Say we have a MC $X_0, X_1, \ldots$ with transition probability matrix P. Seek a function h such that $h(X_0), h(X_1), \ldots$ is a Martingale.

  Key computation:

$$\begin{aligned}\mathbb{E}[h(X_{n+1})|X_0, X_1, \ldots, X_n] =& \mathbb{E}[h(X_{n+1})|X_n] \\ =& (Ph)(X_n)\end{aligned}$$

  where

$$\begin{aligned}Ph(i) &= \sum_j P(i,j)h(j) \\ &= \mathbb{E}[h(X_1)|X_0 = i] \\ &= \mathbb{E}[h(X_{n+1})|X_n = i]\end{aligned}$$

  Notice that if $Ph = h$, then

$$\begin{aligned}\mathbb{E}[h(X_{n+1})|h(X_0), \ldots, h(X_n)] &= \mathbb{E}[\mathbb{E}[h(X_{n+1})|X_0, \ldots, X_n]|h(X_0), \ldots, h(X_n)] \\ &= \mathbb{E}[h(X_n)|h(X_0), \ldots, h(X_n)] \\ &= h(X_n)\end{aligned}$$

  Definition: A function h such that $h = Ph$ is called a harmonic function associated with the Markov matrix P, or a P-harmonic function for short.

- **Conclusion:** If h is a P-harmonic function and (X_n) is a Markov chain with transition matrix P, then $(h(X_n))$ is a martingale. You can also check the converse: if $(h(X_n))$ is a martingale, no matter what the initial distribution of X_0, then h is P-harmonic.

- **Note:** A constant function $h(i) \equiv c$ is harmonic for every transition matrix P, because every row of P sums to 1.

- Example: A simple random walk with $\pm 1, (1/2, 1/2)$ on $\{0, \ldots, b\}$, and 0 and b absorbing: $P(0,0) = 1, P(b,b) = 1, P(i, i \pm 1) = 1/2,\ 0 < i < b$.

  Now describe the harmonic function solutions $h = (h_0, h_1, \ldots, h_b)$ of $h = Ph$. For $0 < i < b$, $h(i) = \sum_j P(i,j)h(j) = 0.5h(i-1) + 0.5h(i+1)$, $h(0) = h(0), h(b) = h(b)$.

  Observe that $h = Ph \Rightarrow i \to h(i)$ is a straight line.
  So if $h(0)$ and $h(b)$ are specified, get $h(i) = h(0) + i\dfrac{h(b) - h(0)}{b}$. So there is just

a two-dimensional subspace of harmonic functions. And a harmonic function is determined by its boundary values.

Exercise: Extend the discussion to a $p \uparrow, q \downarrow$ walk with absorbtion at 0 and b for $p \neq q$. (Hint: Consider $(q/p)^i$ as in discussion of the Gambler's ruin problem for $p \neq q$)

2 First Step Analysis

First step analysis is the general approach to computing probabilities or expectations of a functional of a Markov chain $X_0, X_1, \ldots$ by conditioning the first step X_1. The key fact underlying first step analysis is that if $X_0, X_1, \ldots$ is a MC with some initial state $X_0 = i$ and transition matrix P, then conditionally on $X_1 = j$, the sequence $X_1, X_2, X_3, \ldots$ (with indices shifted by 1) is a Markov chain with $X_1 = j$ and transition matrix P. You can prove this by checking from first principles e.g.

$$\mathbb{P}(X_2 = j_2, X_3 = j_3, X_4 = j_4 | X_1 = j) = P(j, j_2)P(j_2, j_3)P(j_3, j_4)$$

- Back to gambler's ruin problem: Consider random walk $X_0, X_1, \ldots$ with $\pm 1, (1/2, 1/2)$ on $\{0, \ldots, b\}$. We want to calculate $\mathbb{P}(\text{walk is absorbed at } b \mid \text{starts at } X_0 = a)$.

- New approach: let $h_b(i) := \mathbb{P}(\text{reach b eventually} \mid X_0 = i)$.

  Now $h_b(0) = 0, h_b(b) = 1$, and for $0 < i < b$,

$$\begin{aligned} h_b(i) =& P(i, i+1)\mathbb{P}(\text{reach b eventually} \mid X_1 = i+1) \\ &+ P(i, i-1)\mathbb{P}(\text{reach b eventually} \mid X_1 = i-1) \\ =& 0.5\ h_b(i+1) + 0.5\ h_b(i-1) \end{aligned}$$

  Observe that $h = h_b$ solves $h = Ph$. So from the results of harmonic functions, $h_b(i) = i/b$.

- Example: Random walk on a grid inside square boundary.

  For the symmetric nearest neighbor random walk, with all four steps to nearest neighbors equally likely, and absorbtion on the boundary, a function is harmonic if its value at each state (i, j) in the interior of the grid is the average of its values at the four neighbors $(i \pm 1, j \pm 1)$. This is a discrete analog of the notion of a harmonic function in classical analysis, which is a function in a two dimensional domain whose value at (x, y) is the average of its values around every circle centered at (x, y) which is inside the domain.

Lecture 6 : Markov Chains

STAT 150 Spring 2006 *Lecturer: Jim Pitman* Scribe: *Alex Michalka* <>

Markov Chains

- Discrete time
- Discrete (finite or countable) state space S
- Process $\{X_n\}$
- Homogenous transition probabilities
- matrix $P = \{P(i,j);\ \ i,j \in S\}$

$P(i,j)$, the $(i,j)^{th}$ entry of the matrix P, represents the probability of moving to state j given that the chain is currently in state i.

Markov Property:
$\mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n, \ldots, X_0 = i_0) = P(i_n, i_{n+1})$
This means that the states of $X_{n-1} \ldots X_0$ don't matter. The transition probabilities only depend on the current state of the process. So,

$$\begin{aligned}
\mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n, \ldots, X_0 = i_0) &= \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n) \\
&= P(i_n, i_{n+1})
\end{aligned}$$

To calculate the probability of a path, multiply the desired transition probabilities:

$\mathbf{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) = P(i_0, i_1) \cdot P(i_1, i_2) \cdot \ldots \cdot P(i_{n-1}, i_n)$

Example: iid Sequence $\{X_n\}$, $\mathbf{P}(X_n = j) = p(j)$, and $\sum_j p(j) = 1$.

$P(i,j) = j$.

Example: Random Walk. $S = \mathbf{Z}$ (integers), $X_n = i_0 + D_1 + D_2 + \ldots + D_n$, where D_i are iid, and $\mathbf{P}(D_i = j) = p(j)$.

$P(i,j) = p(j - i)$.

Example: Same random walk, but stops at 0.

$$P(i,j) = \begin{cases} p(j-i) & \text{if } i \neq 0; \\ 1 & \text{if } i = j = 0; \\ 0 & \text{if } i = 0, j \neq 0. \end{cases}$$

Example: $Y_0, Y_1, Y_2 \ldots$ iid. $\mathbf{P}(Y = j) = p(j)$. $X_n = max(Y_0, \ldots, Y_n)$.

$$P(i,j) = \begin{cases} p(j) & \text{if } i \leq j; \\ \sum\limits_{k \leq i} p(k) & \text{if } i = j; \\ 0 & \text{if } i > j. \end{cases}$$

Matrix Properties:
$P(i, \cdot) = [P(i,j);\ \ j \in S]$, the i^{th} row of the matrix P, is a row vector. Each row of P is a probability distribution on the state-space S, representing the probabilities of transitions out of state i. Each row should sum to 1. ($\sum\limits_{j \in S} P(i,j) = 1\ \ \forall i$)

The column vectors $[P(\cdot, j); j \in S]$ represent the probability of moving into state j.

Use the notation $\mathbf{P}_i$ = probability for a Markov Chain that started in state i ($X_0 = i$). In our transition matrix notation, $\mathbf{P}_i(X_1 = j) = P(i,j)$.

***n*-step transition probabilities:**

Want to find $\mathbf{P}_i(X_n = j)$ for $n = 1, 2, ...$
This is the probability that the Markov Chain is in state j after n steps, given that it started in state i. First, let $n = 2$.

$$\begin{aligned} \mathbf{P}_i(X_2 = k) &= \sum_{j \in S} \mathbf{P}_i(X_i = j, X_2 = k) \\ &= \sum_{j \in S} P(i,j)P(j,k). \end{aligned}$$

This is simply matrix multiplication.

$$\text{So, } \mathbf{P}_i(X_2 = k) = P^2(i,k)$$

We can generalize this fact for n-step probabilities, to get:

$$\mathbf{P}_i(X_n = k) = P^n(i,k)$$

Where $P^n(i,k)$ is the $(i,k)^{th}$ entry of P^n, the transition matrix multiplied by itself n times. This is a handy formula, but as n gets large, P^n gets increasingly difficult and time-consuming to compute. This motivates theory for large values of n.

Suppose we have $X_0 \sim \mu$, where μ is a probability distribution on S, so $\mathbf{P}(X_0 = i) =$

$\mu(i)$ for $i \in S$. We can find n^{th} step probabilities by conditioning on X_0.

$$\begin{aligned}
\mathbf{P}(X_n = j) &= \sum_{i \in S} \mathbf{P}(X_0 = i) \cdot \mathbf{P}(X_n = j | X_0 = i) \\
&= \sum_{i \in S} \mu(i) P^n(i,j) \\
&= (\mu P^n)_j \\
&= \text{the } j^{th} \text{ entry of } \mu P^n.
\end{aligned}$$

Here, we are regarding the probability distribution μ on S as a vector indexed by $i \in S$.

So, we have $X_1 \sim \mu P, X_2 \sim \mu P^2, \ldots, X_n \sim \mu P^n$.

Note: To compute the distribution of X_n for a particular μ, it is not necessary to find $P^n(i,j)$ for all i, j, n. In fact, there are very few examples where $P^n(i,j)$ can be computed explicitly. Often, P has certain special initial distributions μ so that computing μP^n is fairly simple.

Example: Death and Immigration Process

$X_n =$ the number of individuals in the population at time (or generation) n.

$S = \{0, 1, 2, \ldots\}$

Idea: between times n and $n+1$, each of the X_n individuals dies with probability p, and the survivors contribute to generation X_{n+1}. Also, immigrants arrive each generation, following a Poisson(λ) distribution.

Theory: What is the transition matrix, P, for this process? To find it, we condition on the number of survivors to obtain:

$$P(i,j) = \sum_{k=0}^{i \wedge j} \left(\begin{array}{c} i \\ j \end{array} \right) (1-p)^k (p)^{i-k} \cdot \frac{e^{-\lambda} \lambda^{j-k}}{(j-k)!}$$

Here, $i \wedge j = min\{i, j\}$. This formula cannot be simplified in any useful way, but we can analyze the behavior for large n using knowledge of the Poisson/Binomial relationship. We start by considering a special initial distribution for X_0.

Let $X_0 \sim$ Poisson(λ_0). Then the survivors of X_0 have a Poisson($\lambda_0 q$) distribution, where $q = 1-p$. Since the number of immigrants each generation follows a Poisson(λ) distribution, independent of the number of people currently in the population, we have $X_1 \sim$ Poisson($\lambda_0 q + \lambda$). We can repeat this logic to find:

$$X_2 \sim \text{Poisson}((\lambda_0 q + \lambda) q + \lambda) = \text{Poisson}(\lambda_0 q^2 + \lambda q + \lambda), \text{ and}$$

$$X_3 \sim \text{Poisson}((\lambda_0 q^2 + \lambda q + \lambda)q + \lambda) = \text{Poisson}(\lambda_0 q^3 + \lambda q^2 + \lambda q + \lambda).$$

In general,

$$X_n \sim \text{Poisson}(\lambda_0 q^n + \lambda \sum_{k=0}^{n-1} q^k).$$

In this formula, $\lambda_0 q^n$ represents the survivors from the initial population, λq^{n-1} represents the survivors from the first immigration, and so on, until λq represents the survivors from the previous immigration, and λ represents the immigrants in the current generation.

Now we'll look at what happens as n gets large.

As we let $n \to \infty$: $\quad \lambda_0 q^n \to 0$, and $\quad \lambda_0 q^n + \lambda \sum_{k=0}^{n-1} q^k \to \frac{\lambda}{1-q} = \frac{\lambda}{p}$.

So, no matter what λ_0, if X_0 has Poisson distribution with mean λ_0,

$$\lim_{n\to\infty} \mathbf{P}(X_n = k) = \frac{e^{-\nu}\nu^k}{k!}, \text{ where } \nu = \frac{\lambda}{1-q} = \frac{\lambda}{p}.$$

It is easy enough to show that this is true no matter what the distribution of X_0. The particular choice of initial distribution for X_0 that is Poisson with mean $\lambda_0 = \nu$ gives an invariant (also called stationary, or equilibrium, or steady-state) distribution of X_0. With $\lambda_0 = \nu$, we find that each X_n will follow a Poisson(ν) distribution. This initial distribution $\mu(i) = \frac{e^{-\nu}\nu^i}{i!}$ is special because it has the property that

$$\sum_{i \in S} \mu(i) P(i,j) = \mu(j) \text{ for all } j \in S.$$

Or, in matrix form, $\mu P = \mu$. It can be shown that this μ is the *unique* stationary probability distribution for this Markov Chain.

Stat 150 Stochastic Processes **Spring 2009**

Lecture 6: Markov Chains and First Step Analysis II

Lecturer: Jim Pitman

1 Further Analysis of Markov Chain

- In class last time: We found that if $h = Ph$, then

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

  $\Rightarrow$ (2) $\mathbb{E}[h(X_{n+1}) \mid h(X_0), h(X_1), \ldots, h(X_n)] = h(X_n) \Rightarrow h(X_n)$ is a MC.

  Why? This illustrates a conditional form of $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$, which can be written more generically as

$$(\beta) \qquad \mathbb{E}[Y|Z] = \mathbb{E}[\mathbb{E}[Y|X,Z]|Z]$$

  To obtain (2), take $Y = h(X_{n+1})$, $Z = (h(X_0), h(X_1), \ldots, h(X_n))$, $X = (X_0, X_1, \ldots, X_n)$. Notice that $(X, Z) = (X_0, X_1, \ldots, X_n, h(X_0), h(X_1), \ldots, h(X_n))$ is the same information as $(X_0, X_1, \ldots, X_n)$ since Z is a function of X.

  So

$$\begin{aligned}
&\mathbb{E}[h(X_{n+1}) \mid h(X_0), h(X_1), \ldots, h(X_n)] \\
=&\mathbb{E}[\mathbb{E}[h(X_{n+1}) \mid X_0, X_1, \ldots, X_n, h(X_0), h(X_1), \ldots, h(X_n)] \mid h(X_0), h(X_1), \ldots, h(X_n)] \\
=&\mathbb{E}[\mathbb{E}[h(X_{n+1})|X_0, X_1, \ldots, X_n] \mid h(X_0), h(X_1), \ldots, h(X_n)] \\
=&\mathbb{E}[h(X_n) \mid h(X_0), h(X_1), \ldots, h(X_n)] \\
=&h(X_n)
\end{aligned}$$

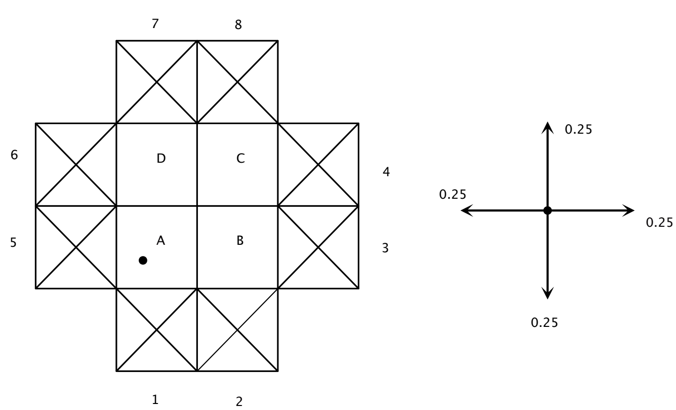
- Let us derive (β) from the definition of $\mathbb{E}[Y|Z]$:

$$\begin{aligned}
\mathbb{E}[Y|Z=z] &= \frac{\mathbb{E}[Y\mathbf{1}(Z=z)]}{\mathbb{P}(Z=z)} \\
&= \frac{\mathbb{E}[\sum_x Y\mathbf{1}(X=x, Z=z)]}{\mathbb{P}(Z=z)} \\
&= \frac{\sum_x \mathbb{E}[Y\mathbf{1}(X=x, Z=z)]}{\mathbb{P}(Z=z)} \\
&= \frac{\sum_x \mathbb{E}[Y|X=x, Z=z]\mathbb{P}(X=x|Z=z)\mathbb{P}(Z=z)}{\mathbb{P}(Z=z)} \\
&= \sum_x \mathbb{E}[Y|X=x, Z=z]\mathbb{P}(X=x|Z=z) \\
&= \mathbb{E}[\mathbb{E}[Y|X, Z=z]|Z=z]
\end{aligned}$$

Therefore, $\mathbb{E}[Y|Z] = \mathbb{E}[\mathbb{E}[Y|X,Z]|Z]$. Note that the above argument uses $\mathbf{1} = \sum_x \mathbf{1}(X=x)$ which implies $Y = \sum_x Y\mathbf{1}(X=x)$. The switch of sum and $\mathbb{E}$ can be justified provided either $Y \geq 0$ or $\mathbb{E}(|Y|) < \infty$).

2 First Step Analysis — An Example

- Consider a symmetric nearest neighbour random walk in a 2-D grid:



Suppose squares 1 to 8 form an absorbing boundary. Start the walk in Square A, what is the distribution of the exit point on the boundary?

Let f_j denote the probability of hitting Square j. Obvious symmetries: $f_1 = f_5$, $f_2 = f_6$, $f_3 = f_7$, $f_4 = f_8$.

Less obvious: $f_2 = f_3$.

- Idea: Let T_B denote the time of first hitting state B (if ever), T = time of first hitting some state $i \in \{1, 2, \ldots, 8\}$. Observe that if $X_T \in \{2, 3\}$, then we must get there via B, which implies $T_B < \infty$. That $f_2 = f_3$ now follows from:

- General fact (Strong Markov Property): For a MC $X_0, X_1, \ldots,$ with transition matrix P, a state B,
  if T_B= first hitting time of B:

$$T_B = \begin{cases} \text{first } n : X_n = B & \text{if any} \\ \infty & \text{if none} \end{cases}$$

  Then $(X_{T_B}, X_{T_B+1}, \ldots)$ given $T_B < \infty$ is a MC with transition matrix P and initial state B.

- Let $f_{ij} = \mathbb{P}_i(X_T = j)$ denote the probability, starting at state i, of exiting in square j. So $f_j = p_{Aj}$. From the above symmetries, and the obvious fact that $\mathbb{P}_A(T < \infty) = 1$, we see

$$f_1 + 2f_2 + f_4 = \frac{1}{2}.$$

  By conditioning on X_1, called *first step analysis*,

$$\begin{aligned} f_{A1} &= \frac{1}{4}0 + \frac{1}{4}1 + \frac{1}{4}f_{B1} + \frac{1}{4}f_{D1} \\ f_1 &= \frac{1}{4} + \frac{1}{4}f_2 + \frac{1}{4}f_2 \qquad \text{(By symmetry)} \\ f_{A2} &= \frac{1}{4}0 + \frac{1}{4}0 + \frac{1}{4}f_{B2} + \frac{1}{4}f_{D2} \\ f_2 &= \frac{1}{4}f_1 + \frac{1}{4}f_4 \end{aligned}$$

  Solving

$$\begin{aligned} f_1 + 2f_2 + f_4 &= \frac{1}{2} \\ f_1 &= \frac{1}{4} + \frac{1}{4}f_2 + \frac{1}{4}f_2 \\ f_2 &= \frac{1}{4}f_1 + \frac{1}{4}f_4 \end{aligned}$$

  yields $f_1 = 7/24, f_2 = 1/12, f_4 = 1/24$. And we can obtain the distribution of other exit points by symmetry.

Stat 150 Stochastic Processes **Spring 2009**

Lecture 7: Limits of Random Variables

Lecturer: Jim Pitman

- Simplest case: pointwise limits.
- Recall that formally a random variable X is a function of outcomes $\omega \in \Omega$:

$$\begin{aligned} X : \Omega &\to \mathbb{R} \\ \omega &\to X(\omega) \end{aligned}$$

  A sequence of random variables (functions) $X_n(\omega)$ *converges pointwise* to a limit $X(\omega)$ means $\lim_{n\to\infty} X_n(\omega) = X(\omega)$, $\forall \omega \in \Omega$.

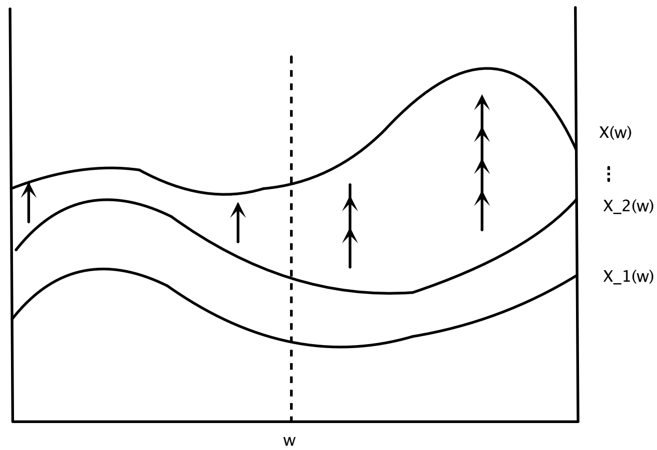
- Important example:

  $X_n(\omega)$ is $\uparrow$ as $n \uparrow$, that is, $X_n(\omega) \le X_{n+1}(\omega) \le \dots$

  For example: $X_n(\omega) = \sum_{k=1}^{n} Y_k(\omega)$ for $Y_k(\omega) \ge 0$. (e.g. if $Y_k = I_{A_k}$ is the indicator of some event A_k; then X_n is the number of events that occur among $A_1, \dots, A_n$.) Since $X_n(\omega)$ is a sequence of nondecreasing random variables, by properties of real numbers, for every ω there is a pointwise limit $X(\omega) = \lim_{n\to\infty} X_n(\omega) \in \mathbb{R} \bigcup \{+\infty\}$. In the example with indicators, $X(\omega) = \sum_{k=1}^{\infty} I_{A_k}(\omega)$. This is the total number of events which occur in the infinite sequence of events $A_1, A_2, \dots$. This number $X(\omega)$ might be finite or infinite, depending on ω.

- If $0 \le X_n \uparrow X$ as above, then the **Monotone Convergence Theorem** gives $0 \le \mathbb{E}[X_n] \uparrow \mathbb{E}[X]$. Applying to $X_n = \sum_{k=1}^{n} Y_k$ for $Y_k \ge 0$ this justifies exchange of infinite sums and $\mathbb{E}$ for non-negative sequences of random variables:

$$\mathbb{E}\left(\sum_{k=1}^{\infty} Y_k\right) = \sum_{k=1}^{\infty} \mathbb{E}(Y_k).$$



1 Law of Large Numbers

- Suppose $X, X_1, X_2, \ldots$ are i.i.d. with $\mathbb{E}|X| < \infty, \mathbb{E}X^2 < \infty$, $S_n := X_1 + \ldots X_n$, $S_n/n =$ the average of $X_1, \ldots, X_n$, then $\mathbb{E}[S_n/n] = \mathbb{E}[X]$, $Var(S_n/n) = Var(X)/n \to 0$ as $n \to \infty$.

- **Convergence in the Mean Square:**

$$\mathbb{E}\left[(S_n/n - \mathbb{E}[X])^2\right] = Var(X)/n \to 0 \quad \text{as } n \to \infty$$

  This is a more subtle sense of convergence of S_n/n to limit $\mathbb{E}[X]$.

- **Almost Sure Convergence:** Suppose $X, X_1, X_2, \ldots$ are i.i.d., if $\mathbb{E}|X| < \infty$, then

$$\mathbb{P}(\omega : \lim_{n\to\infty} S_n(\omega)/n = \mathbb{E}(X)) = 1$$

  Note that there may be an exceptional set of points that do not satisfy $\lim_{n\to\infty} S_n(\omega)/n = \mathbb{E}(X)$, but that set has probability 0 under the current probability measure.

- **The Strong Law of Large Numbers (Kolmogorov):**

$$S_n/n \xrightarrow{a.s.} \mathbb{E}[X] \quad \text{as } n \to \infty$$

- **The Weak Law of Large Numbers:**

$$SLLN \Rightarrow \forall \epsilon > 0, \lim_{n\to\infty} \mathbb{P}(|S_n/n - \mathbb{E}[X]| > \epsilon) = 0$$

A quick proof: By Chevbyshev's inequality,

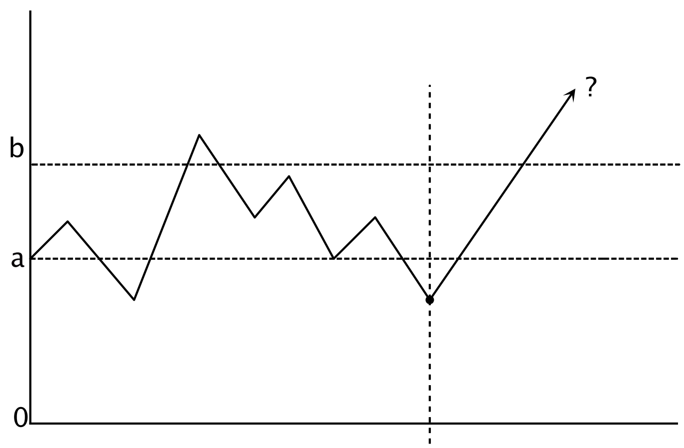
$$\mathbb{P}(|S_n/n - \mathbb{E}[X]| > \epsilon) \leq \frac{Var(X)}{n\epsilon^2} \to 0.$$

- Two results about convergence of MGs (MG convergence theorem):

  (1) If $(X_n, n \geq 0)$ is a nonnegative MG, then X_n converges a.s. to a limit X.

  (2) If $(X_n, n \geq 0)$ is a MG which has $\sup_n \mathbb{E}[X_n^2] < \infty$, then X_n converges both a.s. and in the mean square to some X.

  Hint of proof of (1): From the text, if $X_0 \leq a$, then $\mathbb{P}(\sup_n X_n \geq b) \leq a/b$.



$$\mathbb{P}(\text{\# of upcrossings of [a,b]} \geq 1) \leq a/b$$
$$\mathbb{P}(\text{\# of upcrossings of [a,b]} \geq 2) \leq (a/b)^2$$
$$\mathbb{P}(\text{\# of upcrossings of [a,b]} \geq k) \leq (a/b)^k$$

Let $U_{a,b}$:= total # of upcorssings of [a,b] by path of $X_0, X_1, \ldots,$

$$\mathbb{P}(U_{a,b} \geq k) \leq (a/b)^k \quad \Longrightarrow \quad \mathbb{P}(U_{a,b} < \infty) = 1$$

So with probability one, for every $0 < a < b < \infty$, the sequence X_n must eventually stop crossing the interval $[a, b]$. This leads to almost sure convergence. (Reference: see any graduate level textbook in probability, e.g. Durrett).

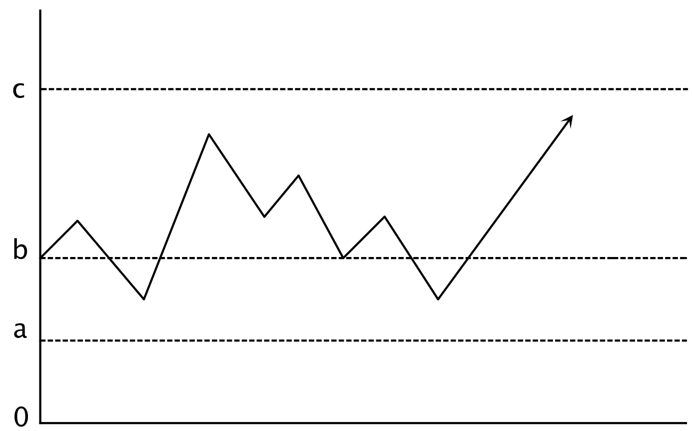
- Example:

  For fair coin tossing, given $X_0 = b, T_a =$ first $n : X_n = a$, and $0 < a < b$, prove that $\mathbb{P}_b(T_a < \infty) = 1$.

Proof:

$$\mathbb{P}_b(T_a < T_c) = \frac{c-b}{c-a} \to 1 \quad \text{as } c \to \infty \qquad \text{(Recall gambler's ruin problem)}$$

but $1 \geq \mathbb{P}(T_a \leq \infty) \geq \mathbb{P}_b(T_a < T_c) \uparrow 1$, therefore, $\mathbb{P}_b(T_a < \infty) = 1$.



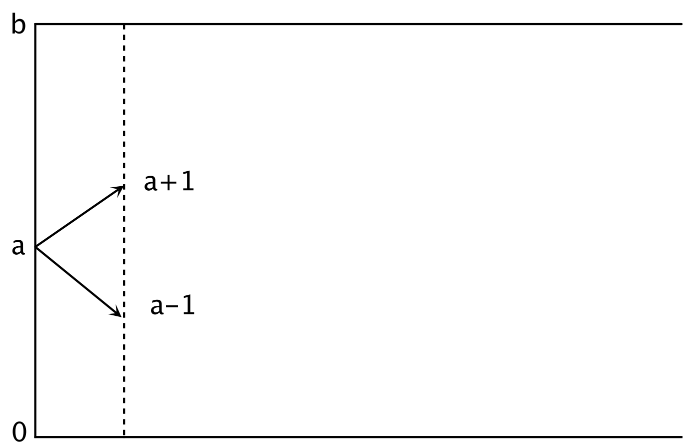
2 Mean first passage times by first step analysis

- Consider a fair coin tossing walk that starts at a and runs to until first reaching either 0 or b. Problem: Find the expected time to reach 0 or b.

$$\begin{aligned} S_0 &= a \\ S_n &= a + X_1 + X_2 + \cdots + X_n \\ T_{0,b} &: = \text{first } n : S_n = 0 \text{ or } b \\ \mathbb{E}_a[T_{0,b}] &: = \mathbb{E}[T_{0,b}|S_0 = a] \end{aligned}$$

Method: Define $m_{a,b} = \mathbb{E}_a[T_{0,b}]$. Fix b and vary a. Condition on the first step:

$$\begin{aligned} \mathbb{E}_a(T_{0,b}|S_1 = s) &= 1 + \mathbb{E}_s(T_{0,b}) \\ m_{a,b} &= 1/2(1 + m_{a+1,b}) + 1/2(1 + m_{a-1,b}) \\ m_{b,b} &= 0 \\ m_{0,b} &= 0 \end{aligned}$$



Simplify the main equation to

$$(m_{a+1,b} + m_{a-1,b})/2 - m_{a,b} = -1$$

which shows that $a \to m_{a,b}$ is a concave function. The simplest concave function is a parabola. The function must vanish at $a = 0$ and $a = b$, which suggests the solution

$$m_{a,b} = ca(b - a)$$

for some $c > 0$. Simple algebra shows this works for $c = 1$. The solution is obviously unique subject to the boundary conditions, hence

$$m_{a,b} = a(b - a).$$

Note that the maximum is attained for $a = b/2$ if b is even and for $a = (b \pm 1)/2$ if b is odd.

Stat 150 Stochastic Processes	**Spring 2009**

Lecture 8: First passage and occupation times for random walk

Lecturer: Jim Pitman

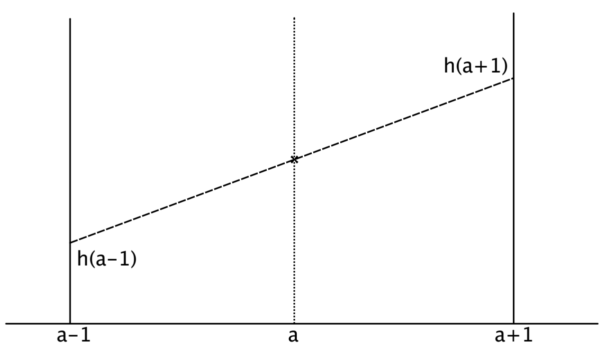
1 First passage and occupation times for random walk

- Gambler's ruin problem on $\{0, 1, \ldots, b\}$ with $P(a, a-1) = P(a, a+1) = \frac{1}{2}$ and 0 and b absorbing. As defined in last class,

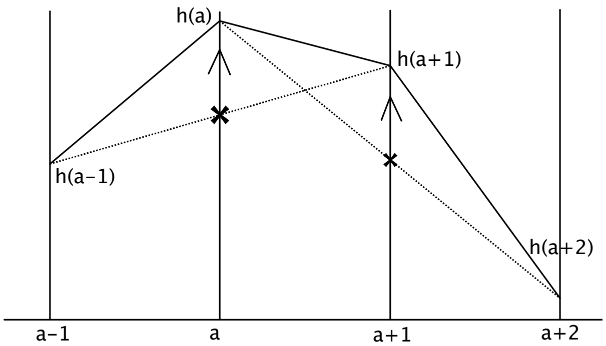
$$\begin{aligned} T_{0,b} &= \text{first } n : X_n \in \{0, b\} \\ m_{a,b} &:= \mathbb{E}_a(T_{0,b}) \\ m_{0,b} &= m_{b,b} = 0 \end{aligned}$$

- Solve this system of equations $m_{a,b} = 1 + \frac{1}{2}(m_{a+1,b} + m_{a-1,b})$.

Recall harmonic equation $h(a) = \frac{1}{2}h(a+1) + \frac{1}{2}h(a-1)$:



Now graphically $h(a) = 1 + \frac{1}{2}h(a+1) + \frac{1}{2}h(a-1)$ looks like this: a concave function whose value at a exceeds by 1 the average of values at $a \pm 1$



Simplify the main equation to

$$(m_{a+1,b} + m_{a-1,b})/2 - m_{a,b} = -1$$

which shows that $a \to m_{a,b}$ is a concave function. The simplest concave function is a parabola, that is a quadratic function of a. The function must vanish at $a = 0$ and $a = b$, which suggests the solution

$$m_{a,b} = ca(b - a)$$

for some $c > 0$. Simple algebra shows this works for $c = 1$, meaning that $m_{a,b} := a(b - a)$ solves the equations. The solution is obviously unique subject to the boundary conditions, hence

$$\mathbb{E}_a(T_{0,b}) = a(b - a).$$

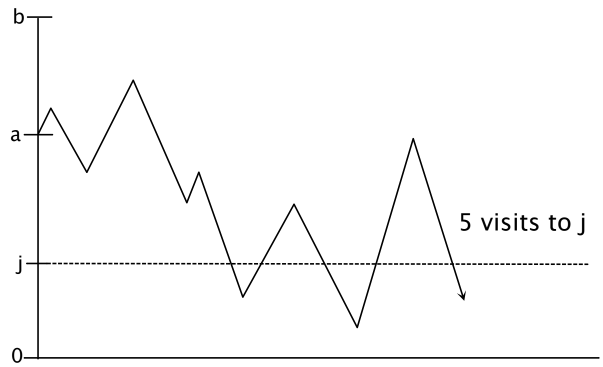
Note that the maximum is attained for $a = b/2$ if b is even and for $a = (b \pm 1)/2$ if b is odd.

- **Exercise:** Do this by a martingale argument. For $S_n = a + X_1 + \cdots + X_n$, with no barrier at 0 and b, and $X_n = \pm 1$ $(1/2, 1/2)$, check that $(S_n^2 - n)$ is a MG. Now argue that this process stopped at time $T_{0,b}$ is still a martingale, and deduce the result.

- **Occupation times**
In the previous example, take a state j with $1 \leq j \leq b$ and calculate the expected number of visits to j before $T_{0,b}$. Denote $N_j := \#$ times hitting j.

Observe that $\displaystyle\sum_{j=1}^{b-1} = T_{0,b}$. To make this true, we must agree that for $j = a$, we count the visit at time 0. (e.g. $0 < a = 1 < b = 2$, must hit 0 or b at time 1)

Formally, $N_j := \sum_{n=0}^{\infty} \mathbf{1}(S_n = j)$



- Discuss the case $j = a$. Key observations:
  - At every return to a, start a fresh Markov chain from a.
  - If we let $p = \mathbb{P}_a(T_{0,b}$ before returning to $a) = P(a, b)$, we see

$$\mathbb{P}_a(N_a = 1) = p$$
$$\mathbb{P}_a(N_a = 2) = (1-p)p$$
$$\mathbb{P}_a(N_a = k) = (1-p)^{k-1}p$$

  Then $\mathbb{E}_a(N_a) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = 1/p.$

- Now find $p = P(a, b)$. Method: condition on the first step.
  With probability $1/2$ of going up, starting at $a + 1$:

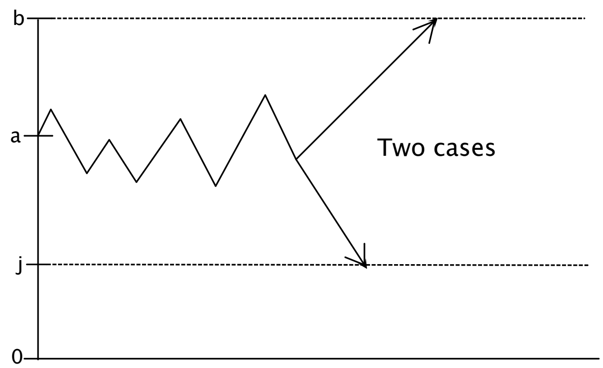
$$\mathbb{P}_{a+1}(\text{hit } b \text{ before } a) = \mathbb{P}_1(\text{hit } (b-a) \text{ before } 0) = \frac{1}{b-a}$$

  Similarly,

$$\mathbb{P}_{a-1}(\text{hit } 0 \text{ before } a) = 1 - \mathbb{P}_{a-1}(\text{hit } a \text{ before } 0) = 1 - \frac{a-1}{a} = \frac{1}{a}$$

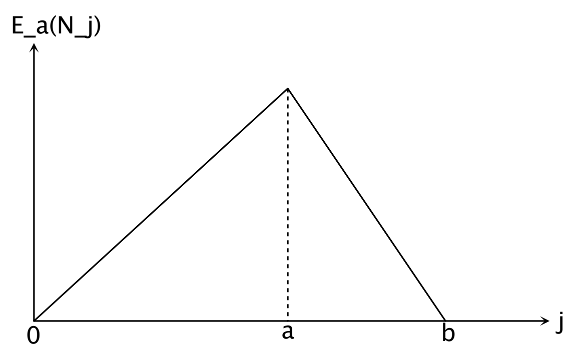
  Thus $P(a, b) = \frac{1}{2}(\frac{1}{a} + \frac{1}{b-a}) = \frac{b}{2a(b-a)}$. So $\mathbb{E}_a(N_a) = \frac{2a(b-a)}{b}$.

- Now consider the case $0 < j < a$.



$$
\begin{aligned}
\mathbb{E}_a(N_j) =& \mathbb{P}_a(\text{hit } j \text{ before } b) \cdot \mathbb{E}_j(N_j) + \mathbb{P}_a(\text{hit } b \text{ before } j) \cdot \mathbb{E}_a(N_j | \text{hit } b \text{ before } j) \\
=& \frac{b-a}{b-j} \cdot \frac{2j(b-j)}{b}, \qquad \text{since } (\mathbb{E}_a(N_j | \text{hit } b \text{ before } j) = 0) \\
=& 2j \frac{(b-a)}{b}, \quad \text{for } 1 \le j \le a
\end{aligned}
$$

The graph below shows that for $a \le j \le b$ we must get $\mathbb{E}_a(N_j) = \frac{2a}{b}(b-j)$.



- <u>Final Check</u>: $\sum_{j=1}^{b-1} \mathbb{E}_a(N_j) = \mathbb{E}_a(T_{0,b}) = a(b-a)$

$$
\begin{aligned}
\sum_{j=1}^{b-1} \mathbb{E}_a(N_j) =& \sum_{j=1}^{a-1} \mathbb{E}_a(N_j) + \sum_{j=a}^{a} \mathbb{E}_a(N_j) + \sum_{j=a+1}^{b-1} \mathbb{E}_a(N_j) \\
=& \sum_{j=1}^{a-1} 2j \frac{b-a}{b} + 2a \frac{b-a}{b} + \sum_{j=a+1}^{b-1} 2(b-j) \frac{a}{b} \\
=& \frac{2(b-a)}{b} \cdot \frac{(a-1)(1+a-1)}{2} + \frac{2a}{b}(b-a) + \frac{2a}{b} \cdot \frac{(b-a-1)(b-a-1+1)}{2} \\
=& a(b-a).
\end{aligned}
$$

Stat 150 Stochastic Processes **Spring 2009**

Lecture 9: Waiting for patterns

Lecturer: Jim Pitman

1 Waiting for patterns

- **Expected waiting time for patterns in Bernoulli trials**

Suppose $X_1, X_2, \ldots$ are independent coin tosses with $\mathbb{P}(X_i = H) = p$, $\mathbb{P}(X_i = T) = 1-p = q$. Take a particular pattern of some finite length K, say $\underbrace{HH\ldots H}_{K}$ or $\underbrace{HH\ldots H}_{K-1}T$. Let

$$T_{\text{pat}} := \text{ first } n \text{ s.t. } (\underbrace{X_{n-K+1}, X_{n-K+2}, \ldots, X_n}_{K}) = \text{pattern}$$

You can try to find the distribution of T_{pat}, but you will find it very difficult.

We can compute $\mathbb{E}[T_{\text{pat}}]$ with our tools. Start with the case of pat $= \underbrace{HH\ldots H}_{K}$.

- For $K = 1$,

$$\mathbb{P}(T_H = n) = q^{n-1}p, \qquad n = 1, 2, 3 \ldots$$
$$\mathbb{E}(T_H) = \sum_n nq^{n-1}p = \frac{1}{p}$$

For a general K, notice that when the first T comes (e.g. $HHHT$, considering some $K \geq 4$), we start the counting process again with no advantage.

Let m_K be the expected number of steps to get K heads in a row.
Let T_T be the time of the first tail.

<u>Observe</u>:
If $T_T > K$, then $T_{HH\ldots H} = K$;
if $T_T = j \leq K$, then $T_{HH\ldots H} = j +$ (fresh copy of) $T_{HH\ldots H}$.

Therefore $\mathbb{E}(T_{HH\ldots H}|T_T > K) = K$ and $\mathbb{E}(T_{HH\ldots H}|T_T = j) = j + m_K$. So condition m_K on T_T:

$$
\begin{aligned}
m_K &= \sum_{j=1}^{K} \mathbb{P}(T_T = j)(j + m_K) + \mathbb{P}(T_T > K)K \\
&= \sum_{j=1}^{K} p^{j-1} q (j + m_k) + p^K K \\
&= (\sum_{j=1}^{K} p^{j-1} q) m_K + (\sum_{j=1}^{K} p^{j-1} q j) + p^K K \\
&= (1 - p^K) m_K + \mathbb{E}(T_T \mathbf{1}(T_T \leq K)) + p^K K
\end{aligned}
$$

Recall that $\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(X|A)\mathbb{P}(A)$, and notice that $\mathbb{E}(T_T) = 1/q$, $\mathbb{E}(T_T|T_T > K) = K + 1/q$, so

$$
\begin{aligned}
\mathbb{E}(T_T\mathbf{1}(T_T \leq K)) &= \mathbb{E}(T_T) - \mathbb{E}(T_T\mathbf{1}(T_T > K)) \\
&= \frac{1}{q} - (K + \frac{1}{q}) p^K
\end{aligned}
$$

Therefore

$$
\begin{aligned}
m_K &= (1 - p^K) m_K + \frac{1}{q} - (K + \frac{1}{q}) p^K + p^K K \\
&= \frac{1 - p^K}{1 - p} \, \frac{1}{p^K} \\
&= \frac{1 + p + p^2 + \cdots + p^{K-1}}{p^K}
\end{aligned}
$$

or

$$m_K p^K = 1 + p + p^2 + \cdots + p^{K-1}.$$

- Try to understand equation $m_K p^K = 1 + p + p^2 + \cdots + p^{K-1}$.

  Idea: Imagine you are gambling and observe the sequence of H and T evolving. Each time you bet that the next K steps will be $\underbrace{HH\ldots H}_{K}$. Scale to get \$1 if you win.

  Accounting: Suppose that K = 3.

Return: $\quad \$1 \quad p\$1 \quad p^2\$1$

$H \quad T\ H\ H\ T\ T\ T\ H\ H \quad H \quad ? \quad ?$

Cost: $p^3 + p^3 + \cdots + p^3$

Expected cost $= \$\ p^K m_K$
Expected return $= \$1 \cdot (1 + p + \cdots + p^{K-1})$
So by "Conservation of Fairness", $m_K p^K = 1 + p + p^2 + \cdots + p^{K-1}$. Rigorously, this is a martingale argument, but details are beyond scope of this course. See "A Martingale Approach to the Study of Occurrence of Sequence Patterns in Repeated Experiments" by Shuo-Yen Robert Li, Ann. Probab. Volume 8, Number 6 (1980), 1171-1176.

- Try pattern $= HTHT$

Return: $\quad \$1 \quad pq\$1$

$| \cdots \quad H\ T\ H\ T\ ?\ ?\ ?$

Cost: $p^2q^2 + \cdots + p^2q^2$

Again, make fairness argument:

$$p^2q^2 m_{\text{HTHT}} = 1 + pq \Rightarrow m_{\text{HTHT}} = \frac{1+pq}{p^2q^2}$$

- Try

$$\begin{array}{llllllllllllllll}
H & H & T & T & H & H & T & T & & & & & & & & \longrightarrow \$1 \\
 & H & H & T & T & H & H & T & T & & & & & & & \longrightarrow 0 \\
 & & H & H & T & T & H & H & T & T & & & & & & \longrightarrow 0 \\
 & & & H & H & T & T & H & H & T & T & & & & & \longrightarrow 0 \\
 & & & & H & H & T & T & H & H & T & T & & & & \longrightarrow p^2q^2\$1 \\
 & & & & & H & H & T & T & H & H & T & T & & & \longrightarrow 0 \\
 & & & & & & H & H & T & T & H & H & T & T & & \longrightarrow 0 \\
 & & & & & & & H & H & T & T & H & H & T & T & \longrightarrow 0
\end{array}$$

Then $m_{\text{HHTTHHTT}} = \dfrac{1 + p^2q^2}{p^4q^4}$.

- To verify the method described above actually works for all patterns of length $K = 2$, make a M.C. with states {HH, TT, TH, TT} by tracking the last two states.

	HH	HT	TH	TT
HH	p	q	0	0
HT	0	0	p	q
TH	p	q	0	0
TT	0	0	p	q

We have a chain with states $i \in \{HH, HT, TH, TT\}$, and we want the mean first passage time to HT. Say we start at $i = TT$, and define

$$m_{ij} = \text{mean first passage time to } (j = HT)$$

In general, for a M.C. with states $i, j, \ldots,$

$$m_{ij} = \mathbb{E}_i(\text{ time to reach j })$$

If $i \neq j$,

$$\begin{aligned} m_{ij} &= \sum_{k \neq j} P(i,k)(1 + m_{kj}) + \sum_{k=j} 1 \cdot P(i,j) \\ &= 1 + \sum_{k \neq j} P(i,k) m_{kj} \end{aligned}$$

Remark: The derivation shows that j need not be an absorbing state.

So we want

$$\begin{aligned} m_{\text{TT,HT}} &= 1 + p\ m_{\text{TH,HT}} + q\ m_{\text{TT,HT}} \\ m_{\text{TH,HT}} &= 1 + p\ m_{\text{HH,HT}} \\ m_{\text{HH,HT}} &= 1 + p\ m_{\text{HH,HT}} \end{aligned}$$

Solve the system of equations and verify that $m_{\text{TT,HT}} = \dfrac{1}{pq}$. In previous notation, this is just m_{HT}, and the general formula is seen to be working. You can check similarly that the formula works for patterns of length 3 by considering a chain with 8 patterns as its states, and so on (in principle) for patterns of length K with 2^K states.

Stat 150 Stochastic Processes **Spring 2009**

Lecture 10: The fundamental matrix (Green function)

Lecturer: Jim Pitman

1 The fundamental matrix (Green function)

- Formulate for Markov chains with an absorbing boundary. (Applications to other chains can be made by suitable sets of states to make them absorbing.)

  Suppose the state space $S = I \cup B$ satisfies the following regularity condition:

  - Finite number total number of *internal states* I.
  - Finite or countably infinite set of *boundary states* B: each $b \in B$ is *absorbing*, meaning that $P(b,b) = 1$.
  - Starting at any $i \in I$, there is some path to the boundary, say $i \to j_1 \to j_2 \to \cdots \to j_n \in B$ with $P(i,j_1)P(j_1,j_2)\cdots P(j_{n-1},j_n) > 0$.

  It's easy to see that $\mathbb{P}_i(T_B < \infty) = 1$ for all $i \in I$: say $|I| = N < \infty$, no matter what the starting state i, within $N+1$ steps, by avoiding any repeat, there must be a path as above from i to some $j_n \in B$ in $n \leq N$ steps.
  So $\mathbb{P}_i(T_B \leq N) > 0$ for each $i \in I$. Because I is finite, this implies

  $$\epsilon := \min_{i \in I} \mathbb{P}_i(T_B \leq N) > 0.$$

  Then

  $$\begin{aligned}
  \mathbb{P}_i(T_B > N) &\leq 1 - \epsilon \\
  \mathbb{P}_i(T_B > 2N) &\leq (1-\epsilon)^2 \qquad \text{(Condition on } X_N) \\
  &\vdots \\
  \mathbb{P}_i(T_B > kN) &\leq (1-\epsilon)^k \quad \downarrow 0 \text{ as } k \uparrow \infty \\
  \implies \mathbb{P}_i(T_B < \infty) &= 1
  \end{aligned}$$

  That is, if $T_B = \infty$, then $T_B > KN$ for arbitrary K, so

  $$0 \leq \mathbb{P}_i(T_B = \infty) \leq \mathbb{P}_i(T_B > KN) \leq (1-\epsilon)^K \to 0 \text{ as } K \to \infty.$$

  Therefore $\mathbb{P}_i(T_B = \infty) = 0$. Also, T_B has finite expectation:

  $$\mathbb{E}_i(T_B) \leq N \sum_{K=0}^{\infty} \mathbb{P}_i(T_B > KN) \leq \sum_{K=0}^{\infty} (1-\epsilon)^K = \epsilon^{-1}$$

- (Text Page 169) Assume the structure of the transition matrix P is as follows:

$$P = \begin{array}{ccccc} I & \{| & \overbrace{Q}^{I} & \overbrace{R}^{B} & | \\ B & \{| & 0 & I & | \end{array}$$

  $Q_{ij} = P_{ij}$ for $i, j \in I$ and $R_{ij} = P_{ij}$ for $i \in I, j \in B$. Note that Q is a square matrix, and I denotes an identity matrix, in this diagram indexed by $b \in B$, and later indexed by $i \in I$. Matrix R indexed by $I \times B$ is typically not square.

- Key observation: Look at the mean number of hits on j before hitting B starting at some different $i \in I$. (Always count a hit at time 0)
  First: case $j \in I$

$$\mathbb{E}_i[\sum_{n=0}^{\infty} \mathbf{1}(X_n = j)] = \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = j) = \sum_{n=0}^{\infty} P^n(i, j)$$

  Observe that for $i, j \in I$, $W_{ij} := \mathbb{E}_i N_j = \sum_{n=0}^{\infty} P^n(i, j) < \infty$. Also:

$$\sum_{j \in I} W_i j = \mathbb{E}_i[\sum_{n=0}^{\infty} \sum_{j \in I} \mathbf{1}(X_n = j)] = \mathbb{E}_i[\sum_{n=0}^{\infty} \mathbb{P}_i(T_B > n)] = \mathbb{E}_i(T_B) \leq \epsilon^{-1} < \infty$$

  so $W := (W_{ij}, i, j \in I)$ is an $I \times I$ matrix with all entries finite.

  Claim: $\mathbb{P}^n(i, j) = Q^n(i, j)$ for all $i, j \in I$.

  $W_{ij} = \sum_{n=0}^{\infty} Q_{ij}^n$ for $i, j \in I$ is a finite matrix. Consider the following analogy:

  Real numbers: $w = \sum_{n=0}^{\infty} q^n = 1 + q + q^2 + \cdots = (1 - q)^{-1}$
  Matrices: $W = \sum_{n=0}^{\infty} Q^n = I + Q + Q^2 + \cdots = (I - Q)^{-1}$

  How do we know $I - Q$ is invertible? Let us prove that W really is an inverse of $I - Q$. Need to show: $W(I - Q) = I$, that is, $W - WQ = I$ or $W = I + WQ$. Check:

$$\begin{aligned} W &= \sum_{n=0}^{\infty} Q^n = I + Q + Q^2 + \cdots \\ WQ &= \sum_{n=1}^{\infty} Q^n = Q + Q^2 + Q^3 + \cdots \\ &\Longrightarrow I + WQ = W \end{aligned}$$

- Summary: In this setting, matrix W is called the **Fundamental Matrix**, also the **Green Matrix(or Function)**. We find that $W = (I - Q)^{-1}$. The meaning of entries of W is that W_{ij} is the mean number of hits of state $j \in I$, starting from state $i \in I$.

- Final points:

  (1) W and R encode all the hitting probabilities:

$$
\begin{aligned}
\mathbb{P}_i(X_{T_B} = b) &= \sum_{n=1}^{\infty} \mathbb{P}_i(T_B = n, X_n = b) \\
&= \sum_{n=1}^{\infty} \sum_{j \in I} \mathbb{P}_i(X_{n-1} = j, X_n = b) \\
&= \sum_{n=1}^{\infty} \sum_{j \in I} Q_{ij}^{n-1} P_{jb} \\
&= \sum_{j \in I} (\sum_{n=1}^{\infty} Q_{ij}^{n-1}) P_{jb} \\
&= \sum_{j \in I} W_{ij} P_{jb} \\
&= (WR)_{ib} \qquad \text{since } R = P|_{I \times B}
\end{aligned}
$$

  (2) From Q^n and R you pick up distribution of T_B:

$$
\begin{aligned}
\mathbb{P}_i(T_B = n) &= \sum_{b \in B} \mathbb{P}_i(T_B = n, X_n = b) \\
&= \sum_{b \in B} \sum_{j \in I} Q_{ij}^{n-1} R_{jb} \\
&= (Q^{n-1} R \ 1)_i
\end{aligned}
$$

  (3) Exercise: Q^n and R give you the joint distribution of T_B and X_{T_B}.

- (Compare page 241) Brief discussion of infinite state spaces with no boundary.

  Example: Coin tossing walk, $p \uparrow q \downarrow$ on $\mathbb{Z}$, $N_0 := \sum_{n=0}^{\infty} \mathbf{1}(X_n = 0)$. Compute:

$$
\begin{aligned}
\mathbb{E}_0 N_0 = \mathbb{E}_0 \sum_{n=0}^{\infty} \mathbf{1}(X_n = 0) & \qquad (= \text{ expected number of returns to } 0) \\
= \sum_{n=0}^{\infty} P^n(0,0) & \\
= \sum_{m=0}^{\infty} P^{2m}(0,0) & \\
= \sum_{m=0}^{\infty} \binom{2m}{m} p^m q^m & \\
= \sum_{m=0}^{\infty} \binom{2m}{m} (1/2)^{2m} (4pq)^m &
\end{aligned}
$$

Case $p = q = q/2$: recall Stirling's formula $n! \sim (n/e)^n \sqrt{2\pi n}$. Then

$$\binom{2m}{m}(1/2)^{2m} \approx \frac{c}{\sqrt{m}}$$

$$4pq = 1$$

Then $\displaystyle\sum_{n=0}^{\infty} \frac{c}{\sqrt{m}} = \infty$.

Case $p \neq q$:

$$\sum_{n=0}^{\infty} \underbrace{\binom{2m}{m}(1/2)^{2m}}_{\leq 1} \underbrace{(4pq)^m}_{<1} \leq \frac{1}{1-4pq} < \infty$$

by comparison with a geometric series with ratio $4pq < 1$ for $p \neq q$.

Thus we have $\mathbb{E}_0 N_0 = \infty$ for $p = q = 1/2$ (recurrent case) and $\mathbb{E}_0 N_0 < \infty$ for $p \neq q$ (transient case). In the transient case, $\mathbb{E}_0 N_0 < \infty$ implies $\mathbb{P}_0(N_0 < \infty) = 1$, meaning that the random walk visits state 0 only a finite number of times. The same argument shows that in the transient case, for each fixed state j, $\mathbb{E}_0 N_j < \infty$ and hence $\mathbb{P}_0(N_j < \infty) = 1$, meaning that the random walk visits state j only finite number of times with probability one. It follows that in the transient case, with probability one the random walk eventually leaves each finite set of states, and hence that with probability one $|X_n| \to \infty$ as $n \to \infty$. In the recurrent case, the reason why $\mathbb{E}_0 N_0 = \infty$ is that $\mathbb{P}_0(N_0 = \infty) = 1$. This is not fully obvious yet, but argued in next lecture. So in the recurrent case, the random walk keeps coming back to its initial state, infinitely often, with probability one.

Stat 150 Stochastic Processes **Spring 2009**

Lecture 11: Return times for random walk

Lecturer: Jim Pitman

1 Recurrence/Transience

- From last time: look at the simple random walk on $\mathbb{Z}$ with $p \uparrow q \downarrow$.
  $S_n := a + X_1 + \cdots + X_n$. Compute:

$$\begin{aligned} u_{2n} &= \mathbb{P}_0(S_{2n} = 0) \\ f_{2n} &= \mathbb{P}_0(T_0 = 2n) \\ T_0 &= \text{first return time to } 0 \\ &= \inf\{n : S_n = 0\} \quad (= \infty \text{ if no such } n \text{ exists}) \end{aligned}$$

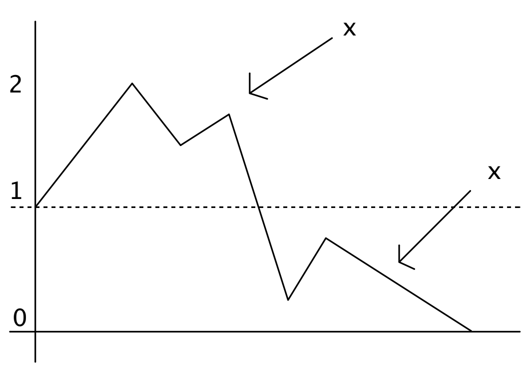
- We already know that

$$\mathbb{P}_0(T_0 < \infty) = \begin{cases} 1 & \text{if } p = q = \frac{1}{2} \\ p \cdot \frac{q}{p} + q \cdot 1 = 2q & \text{if } p > q \\ q \cdot \frac{p}{q} + p \cdot 1 = 2p & \text{if } p < q \end{cases}$$

  To see this,

$$\mathbb{P}_1(T_0 < \infty) = \begin{cases} \frac{q}{p} & \text{if } p \geq q \\ 1 & \text{if } p < q \end{cases}$$

  let $\mathbb{P}_1(x < \infty) = x$, then $x = q \cdot 1 + p \cdot x^2 \implies x = 1$ or q/p.

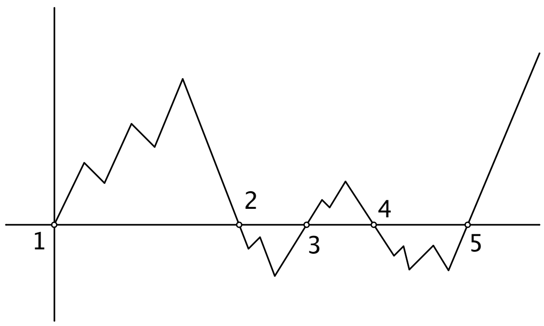


- **Recurrence/Transience**
  We know
  $$u_{2n} = \binom{2n}{n}(pq)^n$$
  What is f_{2n}?

- Let $N_0 = \sum_{n=0}^{\infty} \mathbf{1}(S_n = 0) = \#$ of visits of sums to 0. From last class,
  $$\mathbb{E}_0 N_0 = \sum_{n=0}^{\infty} \binom{2n}{n}(1/2)^{2n}(4pq)^n$$
  $$\binom{2n}{n}(1/2)^{2n} \sim \frac{c}{\sqrt{n}} \qquad\qquad (4pq)^n = \begin{cases} 1 & \text{if } p = q \\ < 1 & \text{if } p \neq q \end{cases}$$
  then
  $$\mathbb{E}_0 N_0 = \begin{cases} \infty & \text{if } p = q \\ < \infty & \text{if } p \neq q \end{cases}$$
  In fact, $\mathbb{P}_0(N_0 = \infty) = \begin{cases} 1 & \text{for } p = q \\ 0 & \text{for } p \neq q \end{cases}$.

  Obviously, $\mathbb{E}_0 N_0 < \infty \implies \mathbb{P}_0(N_0 < \infty) = 1$, but $\mathbb{E}_0 N_0 = \infty$ does not so obviously imply $\mathbb{P}_0(N_0 = \infty) = 1$.

- <u>Idea</u>: Look at the first return probability.
  $$\begin{aligned} f_* := \sum_{n=1}^{\infty} f_{2n} = \sum_{n=1}^{\infty} \mathbb{P}_0(T_0 = 2n) &= \mathbb{P}_0(T_0 < \infty) \\ &= \text{probability of ever returning to } 0 \end{aligned}$$
  Notice that $\mathbb{P}_0(N_0 = 5) = f_*^4(1 - f_*)$. It's easy to see that
  $$\mathbb{P}_0(N_0 = k) = f_*^{k-1}(1 - f_*) \sim \text{Geometric}(1 - f_*)$$
  $$\mathbb{E}_0 N_0 = \frac{1}{1 - f_*} \begin{cases} < \infty & \text{iff } f_* < 1 \\ = \infty & \text{iff } f_* = 1 \end{cases}$$



- **Summary:**

  If $\mathbb{P}_0(T_0 < \infty) < 1$, then $\mathbb{E}_0 N_0 < \infty$.

  If $\mathbb{P}_0(T_0 < \infty) = 1$, then $\mathbb{E}_0 N_0 = \infty$.

- How to find a formula for $f_{2n} = P_0(T_0 = 2n)$?
  Idea: Develop a relation between sequences f_{2n}(unknown) and u_{2n}(known).

  For every $n = 1, 2, 3, \ldots,$

$$\begin{aligned} u_{2n} &= \mathbb{P}_0(S_{2n} = 0) \\ &= \sum_{m=1}^{n} \mathbb{P}_0(T_0 = 2m, S_{2n} = 0) \\ &= \sum_{m=1}^{n} f_{2m} u_{2n-2m} \end{aligned}$$

  Idea: Handle a sequence by its generating function.

  Try the generating function derived from u_{2n}:

$$U(z) = \sum_{n=0}^{\infty} u_{2n} z^{2n}.$$

  Also

$$F(z) = \sum_{n=1}^{\infty} f_{2n} z^{2n}$$

  See text for examples of probability generating functions.

Look at

$$\begin{aligned} F(z)U(z) &= (\sum_{m=1}^{\infty} f_{2m} z^{2m})(\sum_{k=0}^{\infty} u_{2k} z^{2k}) \\ &= \sum_{n=1}^{\infty} (\sum_{m=1}^{n} f_{2m} u_{2n-2m}) z^{2n} \\ &= \sum_{n=1}^{\infty} u_{2n} z^{2n} \\ &= U(z) - 1 \end{aligned}$$

- **Progress:** We have turned the nasty looking convolution formula into a nicer relation of generating functions

$$F(z)U(z) = U(z) - 1 \qquad \text{or} \qquad F(z) = \frac{U(z)-1}{U(z)} = 1 - \frac{1}{U(z)}.$$

$$\begin{aligned} U(1) = \sum_{n=0}^{\infty} u_{2n} &= \mathbb{E}_0 N_0 \\ &= \frac{1}{1 - \mathbb{P}_0(T_0 < \infty)} \\ &= \frac{1}{1 - F(1)} \end{aligned}$$

- **Algebra:** It is a key observation that for $p = q = 1/2$, $u_{2k} = (-1)^k \binom{-\frac{1}{2}}{k}$. See e.g. Feller's book, "An introduction to probability theory and its applications" Vol I, Page 96.

  Recall that

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \qquad \text{for any positive integer } n$$

  generalizes to

$$(1+x)^a = \sum_{k=0}^{\infty} \binom{a}{k} x^k \qquad \text{for } |x| < 1$$

  for all real numbers a, where

$$\binom{a}{k} := \frac{a(a-1)\cdots(a-k+1)}{k!}.$$

Claim: $\binom{2n}{n} = 2^{2n}(-1)^n \binom{-\frac{1}{2}}{n}$.

<u>Proof</u>:

$$\begin{aligned}
\binom{2n}{n} &= \frac{1}{n!}\frac{2n(2n-1)\cdots(n+1)(n)(n-1)\cdots}{n(n-1)\cdots 1} \\
&= \frac{1}{n!}2(2n-1)2(2n-3)2\cdots(5)2(3)2(1) \\
&= \frac{1}{n!}2^n \cdot 1 \cdot 3 \cdots (2n-1) \\
&= \frac{1}{n!}2^{2n}(\frac{1}{2})(\frac{1}{2}+1)(\frac{1}{2}+2)\cdots(\frac{1}{2}+n-1) \\
&= \frac{1}{n!}2^{2n}(-1)^n(-\frac{1}{2})(-\frac{1}{2}-1)\cdots(-\frac{1}{2}-n+1) \\
&= 2^{2n}(-1)^n\binom{-\frac{1}{2}}{n}
\end{aligned}$$

Now apply the binomial expansion for the power $-1/2$

$$\begin{aligned}
U(z) := \sum_{n=0}^{\infty} u_{2n} z^{2n} &= \sum_{n=0}^{\infty} \binom{2n}{n} 2^{-2n} z^{2n} \\
&= \sum_{n=0}^{\infty} \binom{-\frac{1}{2}}{n} (-z^2)^n \\
&= (1-z^2)^{-\frac{1}{2}}
\end{aligned}$$

$$F(z) = \sum_{n=1}^{\infty} f_{2n} z^{2n}, \qquad f_{2n} = \mathbb{P}_0(T_0 = 2n)$$

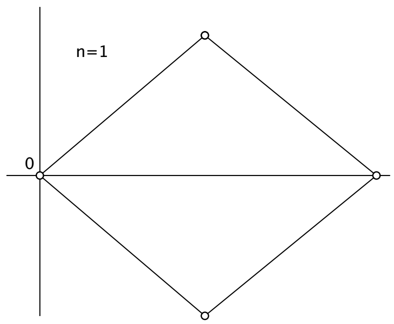
$$\begin{aligned}
F(z) &= 1 - U(z)^{-1} \\
&= 1 - (1-z^2)^{\frac{1}{2}} \\
&= 1 - \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n} (-z^2)^n \\
&= \sum_{n=1}^{\infty} \binom{\frac{1}{2}}{n} (-1)^{n-1} z^{2n}
\end{aligned}$$

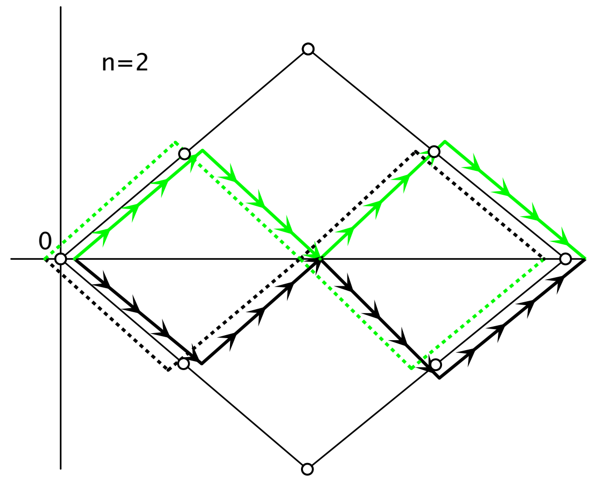
- Finally, compare coefficients to deduce that

$$f_{2n} = (-1)^{n-1}\binom{\frac{1}{2}}{n} = \frac{u_{2n}}{2n-1}$$

Check:

For $n=1$, $f_2/u_2 = \frac{2/4}{2/4} = \frac{1}{1}$ For $n=2$, $\dfrac{f_4}{u_4} = \dfrac{2/8}{6/8} = \dfrac{1}{3}$





For a biased coin,

$$u_{2n}(p) = \binom{2n}{n}(pq)^n = u_{2n}(4pq)^n$$

where $u_{2n} = u_{2n}(1/2)$ as before. So

$$\begin{aligned} U_p(z) &= \sum_{n=0}^{\infty} \binom{-\frac{1}{2}}{n} (4pq)^n (-z^2)^n \\ &= (1-4pqz^2)^{-\frac{1}{2}} \\ F_p(z) &= 1-(1-4pqz^2)^{\frac{1}{2}} \end{aligned}$$

In particular,

$$\begin{aligned} \mathbb{P}_0^{(p)}(T_0 < \infty) = F_p(1) &= 1-(1-4pq)^{\frac{1}{2}} \\ &= 1-|1-2p| \end{aligned}$$

Check: For $p \geq q$,

$$\begin{aligned} \mathbb{P}_0^{(p)}(T_0<\infty) &= p\mathbb{P}_1^{(p)}(T_0<\infty) + q\mathbb{P}_{-1}^{(p)}(T_0<\infty) \\ &= p \cdot q/p + q \cdot 1 \\ &= 2q = 2-2p. \end{aligned}$$

And similarly for $p < q$.

Stat 150 Stochastic Processes **Spring 2009**

Lecture 12: Probability Generating Functions

Lecturer: Jim Pitman

(See text, around page 185.) Previous lecture showed the value of identifying a sequence with a corresponding generating function defined by a power series, then recognizing that various operations on sequences correspond to familar algebraic and analytic operations on the generating functions. For instance, if sequences (f_n) and (g_n) have generating functions

$$F(z) := \sum_{n=0}^{\infty} f_n z^n \text{ and } G(z) := \sum_{n=0}^{\infty} g_n z^n$$

then the *convolution*

$$h_n = \sum_{m=0}^{n} f_m g_{n-m}$$

has generating function

$$H(z) := \sum_{n=0}^{\infty} h_n z^n = F(z)G(z)$$

- Discrete random variable X with values $0, 1, 2, 3, \dots$
  Probability $\mathbb{P}(X = n) = p_n, n = 0, 1, 2, \dots$
  Porbability GF (of X, or of $p_0, p_1, \dots$)

$$\phi(s) := \phi_X(s) := \sum_{n=0}^{\infty} p_n s^n, \quad |s| \le 1$$

The generic notation is $\phi(s)$. The subscript X is just used to indicate what random variable X the generating function is derived from.

- **Basic Properties:**

$$
\begin{aligned}
\phi(1) &= 1 \qquad [\text{Assuming } \mathbb{P}(0 \leq X < \infty) = 1] \\
\phi(0) &= p_0 = \mathbb{P}(X = 0) \\
\phi'(s) &= \sum_{n=0}^{\infty} n p_n s^{n-1}, \quad |s| < 1 \\
\phi'(0) &= p_1 \\
\phi'(1) &= \sum_{n=0}^{\infty} n p_n = \mathbb{E}X \\
\phi''(s) &= \sum_{n=0}^{\infty} n(n-1) p_n s^{n-1}, \quad |s| < 1 \\
\phi''(0) &= 2p_2 \\
\phi''(1) &= \sum_{n=0}^{\infty} n(n-1) p_n = \mathbb{E}(X(X-1))
\end{aligned}
$$

and so on for higher derivatives: evaluating the kth derivative at 0 gives $k!p_k$, and evaluating at 1 gives $E[X(X-1)\cdots(X-k+1)]$. In case any of these moments are infinite, so is the derivative evaluated as a limit as $s \uparrow 1$. In particular, the first two derivatives of ϕ at 1 give the mean and variance: So

$$
\begin{aligned}
E(X) &= 1\phi'(1) \\
E(X^2) &= 1\phi''(1) + \phi'(1) \\
Var(X) &= E(X^2) - [E(X)]^2 = \phi''(1) + \phi'(1) - (\phi'(1))^2
\end{aligned}
$$

- **Uniqueness:**

For X with values $\{0, 1, 2, \dots\}$, the function $s \longrightarrow \phi_X(s)$ for $|s| \leq 1$, or even for $|s| < \epsilon$ for any $\epsilon > 0$, determines the distribution of X uniquely.

Proof: $\mathbb{P}(X = n) = \phi_X^{(n)}(0)/n!$.

- **Sums of independent r.v.'s**

Write $\phi_X(s)$ for the GF of X, $\phi_Y(s)$ for the GF of Y. Assume X and Y are independent. Then

$$\phi_{X+Y}(s) = \phi_X(s)\phi_Y(s)$$

because

$$\begin{aligned}\phi_X(s)\phi_Y(s) &= \left(\sum_{k=0}^{\infty} \mathbb{P}(X=k)s^k\right)\left(\sum_{m=0}^{\infty} \mathbb{P}(Y=m)s^m\right) \\ &= \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} \mathbb{P}(X=k)\mathbb{P}(Y=n-k)\right)s^n \\ &= \sum_{n=0}^{\infty} \mathbb{P}(X+Y=n)s^n\end{aligned}$$

<u>Alternative proof</u>:

$$\phi_X(s) = \sum_{n=0}^{\infty} \mathbb{P}(X=n)s^n = \mathbb{E}[s^X]$$

$$\phi_{X+Y}(s) = \mathbb{E}[s^{X+Y}] = \mathbb{E}[s^X \cdot s^Y] = \mathbb{E}[s^X]\mathbb{E}[s^Y]$$

because independence of X and Y implies independence of s^X and s^Y, and using the rule for expectation of a product of independent random variables.

- Exercise: Use probability GF to confirm that the sum of independent Poisson's is Poisson.

  1) Compute GF of $Poi(\lambda)$, where $p_n = e^{-\lambda}\lambda^n/n!$:

  $$\phi(s) = \sum_{n=0}^{\infty} e^{-\lambda}\frac{\lambda^n}{n!}s^n = e^{\lambda s-\lambda} = e^{\lambda(s-1)}.$$

  2)Look at product of $Poi(\lambda)$ and $Poi(\mu)$'s GF:

  $$e^{\lambda(s-1)}e^{\mu(s-1)} = e^{(\lambda+\mu)(s-1)}$$

  is the GF of $Poi(\lambda+\mu)$. That is,

  $$X \sim Poi(\lambda),\ Y \sim Poi(\mu),\ X \perp Y \Longrightarrow X+Y \sim Poi(\lambda+\mu)$$

- **Random sums**

  Suppose $X_1, X_2, \dots$ are i.i.d. on $\{0, 1, 2, \dots\}$. N is a random index independent of $X_1, X_2, \dots$

  Problem: Find the distribution of $X_1 + \cdots + X_N = S_N$. (Note: $S_0 = 0$ by convention)

Solution: Apply GF. Let $\phi_X(s)$ be the GF of the $X_i's$.

$$\phi_X(s) = \mathbb{E}[s^X] = \sum_{n=0}^{\infty} \mathbb{P}(X = n)s^n, \qquad X = X_i, \text{for any } i$$

Compute by conditioning on N:

$$\begin{aligned}
\phi_X(s) = \mathbb{E}[s^X] &= \sum_{n=0}^{\infty} \mathbb{P}(N = n)\mathbb{E}[s^{S_N}|N = n] \\
&= \sum_{n=0}^{\infty} \mathbb{P}(N = n)\mathbb{E}[s^{X_1+\cdots+X_n}] \\
&= \sum_{n=0}^{\infty} \mathbb{P}(N = n)[\phi_X(s)]^n \\
&= \phi_N(\phi_X(s))
\end{aligned}$$

From the GF of $S_N, \phi_N(\phi_X(s))$, we get formulas for means and variances. Compare with text, first chapter.

- **Example: Poisson Thinning** This is a Stat 134 exercise, much simplified by use of GF. Let $X_1, X_2, \ldots$ be independent 0/1 Bernoulli(p) trials.
  Let N be $Poi(\lambda)$, independent of the $X_i's$.
  Let $S_N = X_1 + \cdots + X_N = \#$ of successes in $Poi(\lambda)$ # of trials.
  Then $S_N \sim Poi(\lambda p)$. This can be checked directly, most easily by showing also that S_N and $N - S_N$ are independent and $S_N \sim Poi(\lambda(1-p))$. But the GF computation is very quick:

$$\begin{aligned}
\phi_{S_N}(s) &= \phi_N(\phi_X(s)) \\
&= \phi_N(q + ps) \\
&= e^{\lambda((q+ps)-1)} \\
&= e^{\lambda p(s-1)}
\end{aligned}$$

  This is the GF of $Poi(\lambda p)$, hence the conclusion, by uniqueness of the GF.

- **Compare: Moment GF**

  Usually the Moment GF is defined for a real valued X as

$$M_X(t) := \mathbb{E}[e^{tX}] = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbb{E}[X^n]$$

  provided the series converges in some neighbourhood of $t = 0$. For discrete X with values in $\{0, 1, 2, \ldots\}$ the change of variables: $e^t = s$ shows that

$$M_X(t) = \phi_X(e^t) \text{ and } \phi_X(s) = M_X(\log s)$$

- **Application of GF to recover a discrete distribution from its factorial moments**.

  Consider a discrete distribution of X on $\{0, 1, 2, \ldots, n\}$. Let us check that such a distribution is determined by its binomial moments:

$$\mathbb{E}\binom{X}{k} = \frac{\mathbb{E}[(X)_k]}{k!}$$

  where $(X)_k := X(X-1)\cdots(X-k+1)$ is a falling factorial function of X, and $\mathbb{E}[(X)_k]$ is the kth factorial moment of X. Note that the kth binomial moment is just some linear combination of the first k moments of X. Now

$$\begin{aligned}
\phi_X(s) &= \mathbb{E}[s^X] \\
&= \phi_X(1 + (s-1)) \\
&= \sum_{k=0}^{n} \phi_X^{(k)}(1) \frac{(s-1)^k}{k!} \\
&= \sum_{k=0}^{n} \mathbb{E}\binom{X}{k}(s-1)^k
\end{aligned}$$

  Hence for $0 \leq j \leq n$

$$\begin{aligned}
P(X = j) &= \text{ Coefficient of } s^j \text{ in } \phi_X(s) \\
&= \sum_{k=j}^{n} (-1)^{k-j} \binom{k}{j} \mathbb{E}\binom{X}{k}
\end{aligned}$$

  In particular, if X is the number of events that occur in some sequence of events $A_1, \ldots, A_n$, in terms of indicators $X_i = 1_{A_i}$

$$X := \sum_{i=1}^{n} X_i$$

  and then

$$\binom{X}{k} = \sum_{1 \leq i_1 < i_2 < \cdots < i_k} \prod_{j=1}^{k} X_{i_j}$$

  is the number of ways to choose k distinct events A_i among those which happen to occur. So

$$\mathbb{E}\binom{X}{k} = \sum_{1 \leq i_1 < i_2 < \cdots < i_k} \mathbb{P}(A_{i_1} A_{i_2} \cdots A_{i_k})$$

is the usual sum appearing in the inclusion exclusion formula for the probability of a union of n events, which can be written in present notation as

$$P(\cup_{i=1}^n A_i) = P(X \geq 1) = \sum_{k=1}^n (-1)^{k-1} \mathbb{E}\binom{X}{k}$$

Since $P(X \geq 1) = 1 - P(X = 0)$ and $\binom{X}{0} = 1$, this agrees with the previous formula for $P(X = j)$ in the case $j = 0$, which is

$$P(X = 0) = \sum_{k=0}^n (-1)^k \mathbb{E}\binom{X}{k} = 1 + \sum_{k=1}^n (-1)^k \mathbb{E}\binom{X}{k}$$

- **Application to the matching problem**

  Let M_n be the number of matches, that is i with $i = \pi_n(i)$, where π_n is a uniformly distributed random permutation of $\{1, \ldots, n\}$. Apply the previous discussion with $X = M_n$ and A_i the event $i = \pi_n(i)$ to see that

$$\mathbb{E}\binom{M_n}{k} = \sum_{1 \leq i_1 < i_2 < \cdots < i_k} \mathbb{P}(A_{i_1} A_{i_2} \cdots A_{i_k}) = \binom{n}{k} \frac{1}{(n)_k} = \frac{1}{k!}$$

  and hence

$$\mathbb{P}(M_n = j) = \sum_{k=j}^n (-1)^{k-j} \binom{k}{j} \frac{1}{k!} = \frac{1}{j!} \sum_{k=j}^n (-1)^{k-j} \frac{1}{(k-j)!}$$

  which converges as $n \to \infty$ to the limit

$$\frac{1}{j!} \sum_{k=j}^\infty \frac{(-1)^{k-j}}{(k-j)!} = \frac{e^{-1}}{j!} = P(M_\infty = j)$$

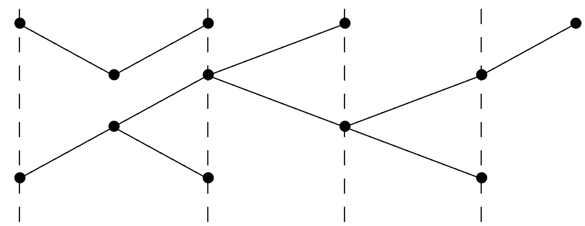
  for a random variable M_∞ with Poisson(1) distribution. Thus the limit distribution of the number of matches in a large random permutation is Poisson(1).

Stat 150 Stochastic Processes **Spring 2009**

Lecture 13: Branching Processes

Lecturer: Jim Pitman

- Consider the following branching process:



Model:
Intuitively: a random genealogical tree. Or a forest of trees, e.g. 2 trees above starting from $Z_0 = 2$ individuals at time 0.
Formalism captures $Z_n :=$ number of individuals at n^{th} generation.
Note that $Z_n = 0 \implies$ population extinct by time n.

$$\begin{aligned}\text{Extinction time } &= \text{first } n \text{ (if any) such that } (Z_n = 0)\\ &= \infty \text{ if no such } n \text{ exists}\end{aligned}$$

$Z_0 =$ initial number of individuals.
$(Z_0, Z_1, Z_2, \dots)$ is the discrete time branching process.

Mechanism: There is a fixed distribution of probabilities $p_0, p_1, p_2, \dots$ for the number of offspring of each individual.
Generally, let X denote the number of offspring of a generic individual, so $\mathbb{P}(X = k) = p_k$.

Assumptions: Informally, each individual present at time n has some number of offspring, distributed like X, independently of all past events and all other individuals present at time n. So for each $n = 0, 1, 2, \dots$ given $Z_0, Z_1, \dots, Z_n = z$ say, Z_{n+1} is the sum of z independent random variables with the same distribution as X:

$$\begin{aligned}(Z_1 | Z_0 = 1) &\stackrel{d}{=} X\\ (Z_1 | Z_0 = 2) &\stackrel{d}{=} X_1 + X_2, \quad \text{two independent copies of } X\\ (Z_{n+1} | Z_0, \dots, Z_{n-1}, Z_n = k) &\stackrel{d}{=} X_1 + \cdots + X_k\end{aligned}$$

Clearly, (Z_n) is a M.C. with state space $\{0, 1, 2, \dots\}$ and state 0 is absorbing. The kth row $P(k, .)$ of the transition matrix is the distribution of $X_1 + \dots + X_k$, which is k-fold convolution of the offspring distribution with itself.

- Basic problem: Describe general features of the B.P. and how they depend on the offspring distribution $p_0, p_1, p_2, \dots$ and Z_0.

  Avoid trivial case: $p_1 = 1, p_i = 0$ for $i \geq 2$, $Z_n = Z_0$ for all n. There is nothing to discuss in this case.

  Main Theorem:
  Let $\mu := \sum_n np_n = \mathbb{E}[X]$,
  Ignoring the trivial case $p_1 = 1$, either $\mu \leq 1$, in which case $\mathbb{P}(\underbrace{Z_n = 0 \text{ eventually}}_{\text{extinction}}) = 1$,
  or $\mu > 1$, in which case $\mathbb{P}(Z_n = 0 \text{ eventually}) < 1$.
  If the extinction event does not occur, then $Z_n \to \infty$ as $n \to \infty$. Note that Z_n is roughly like μ^n. Check: Z_n/μ^n is a martingale, hence $\mathbb{E}(Z_n|Z_0 = 1) = \mu^n$,

$$\begin{aligned}
\mu < 1 &\sim \text{ subcritical} \\
\mu = 1 &\sim \text{ critical} \\
\mu > 1 &\sim \text{ super critical}
\end{aligned}$$

- **Method:** Use theory of P.G.F.

$$\phi(s) = \phi_X(s) := \mathbb{E}[s^X] = p_0 + p_1 s + p_2 s^2 + \dots = \sum_{k=0}^{\infty} p_k s^k$$

  Recall:

$$\begin{aligned}
\phi_{Z_{n+1}}(s) = \mathbb{E}[s^{Z_{n+1}}] &= \mathbb{E}[\mathbb{E}[s^{Z_{n+1}}|Z_n]] \\
&= \sum_k \mathbb{P}(Z_n = k)\mathbb{E}[s^{Z_{n+1}}|Z_n = k] \\
&= \sum_k \mathbb{P}(Z_n = k)[\phi(s)]^k \\
&= \phi_{Z_n}(\phi(s))
\end{aligned}$$

  For simplicity, take $Z_0 = 1$.

$$\begin{aligned}
\phi_{Z_1}(s) &= \phi_X(s) = \phi(s) \\
\phi_{Z_2}(s) &= \phi(\phi(s)) \\
\phi_{Z_3}(s) &= \phi(\phi(\phi(s)))
\end{aligned}$$

and so on. $\phi_{Z_n}(s) =$ n-fold composition of ϕ with itself.

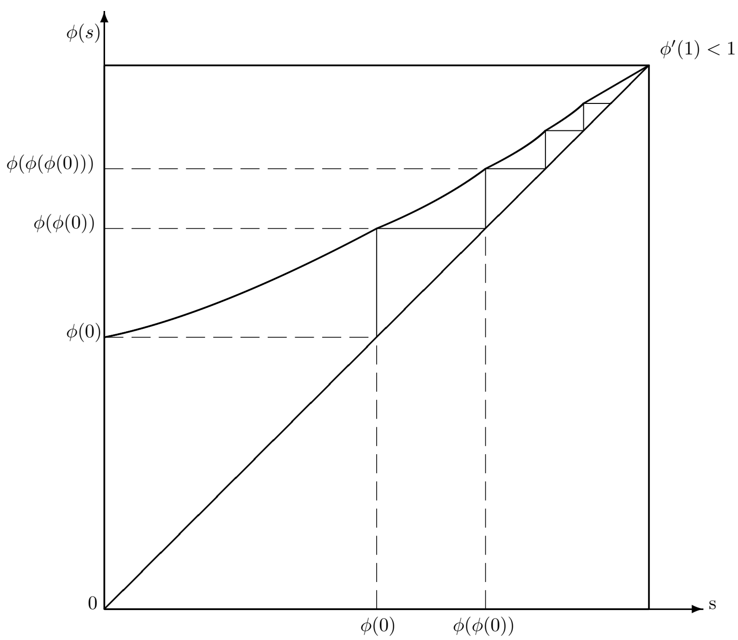
Interest is in

$$\begin{aligned}\mathbb{P}(Z_n = 0) = \mathbb{P}_1(Z_0 = 0) &= \underbrace{\phi(\phi\cdots(\phi}_{n \text{ times}}(0))\cdots) \\ &= \phi_{Z_n}(0) \\ &= \text{coefficient of } s^0 \text{ in this composition}\end{aligned}$$

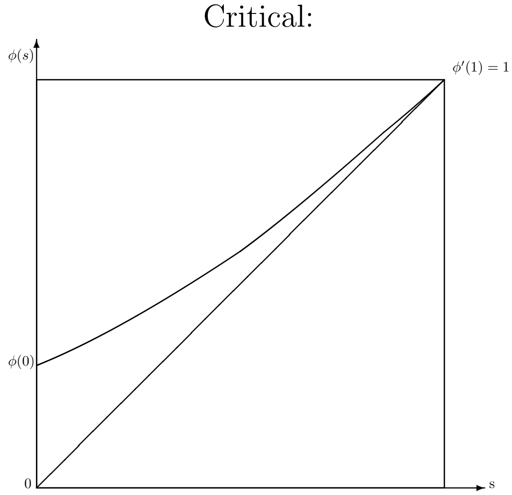
Graphical method for P.G.F.'s:

Recall $\phi'(1) = \sum_k kp_k 1^k = \mathbb{E}[X]$.

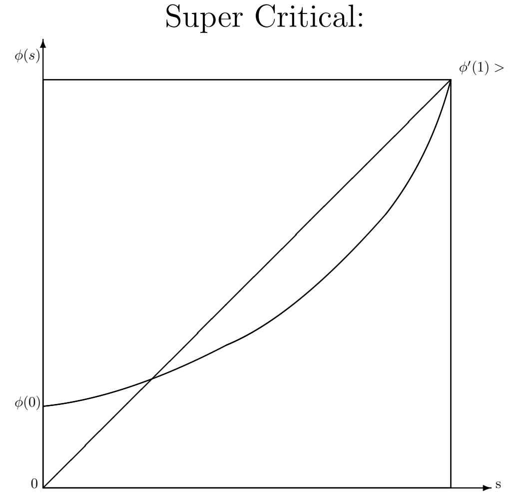
Subcritical:



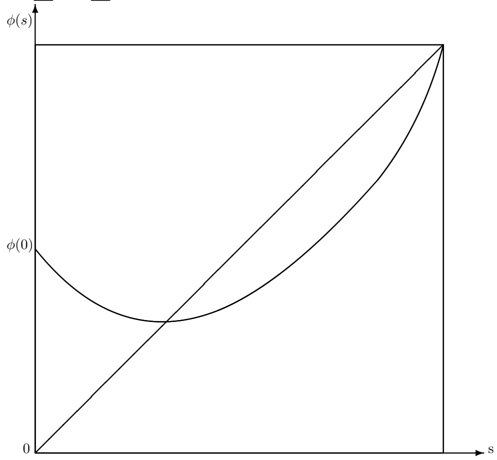
Critical:



Super Critical:



However, the graph below is not a PGF since $\phi'(s) = p_1 + 2p_2 s + \cdots \geq 0$ for $0 \leq s \leq 1$.



Observe that $\mathbb{P}_1(Z_n = 0) \uparrow$ as $n \uparrow$ and it is $\leq 1 \Longrightarrow$ converges to a real limit in $[0, 1]$. This is the extinction probability $\mathbb{P}_1(Z_n = 0) = \underbrace{\phi(\phi \cdots (\phi}_{n \text{ times}}(0))))$.

Let $u_n = \mathbb{P}_1(Z_n = 0)$, $u_1 = \phi(0)$, $u_{n+1} = \phi(u_n) \longrightarrow u = \phi(u)$ because ϕ is continuous.

- **Claim:** If $\mu \leq 1$, then u (= extinction probability) is the **unique root** of $u = \phi(u)$, which is $u = 1$ by calculus.

  Case: $\mu \leq 1 : \phi'(1) \leq 1$
  ϕ is convex $\Longrightarrow \phi(s) \geq s$ for all $s \in [0, 1]$. Suppose $\phi(s) = s$ for some $s_0 < 1$, then $\phi(s) = s$, for all $s_0 \leq s \leq 1$ by convexity.

- All cases:
  $u =$ least root s of $s = \phi(s)$ by graphical argument as indicated before.
  • $\mu \leq 1 \Longrightarrow$ this root is 1.
  • $\mu > 1 \Longrightarrow$ this root is < 1 and it is the only root < 1.

- Example: $X \sim \text{Geometric}(p)$ on $\{0, 1, 2, \dots\}$, $p_n = q^n p, \mu = \mathbb{E}[X] = q/p$.

$$p \geq 1/2 \Longrightarrow \mu \leq 1 \Longrightarrow \text{subcritical/critical} \Longrightarrow u = 1$$
$$p < 1/2 \Longrightarrow \mu > 1 \Longrightarrow \text{super critical} \Longrightarrow u = p/q < 1$$

Check:

$$\phi(s) = \sum_{n=0}^{\infty} s^n q^n p = \frac{p}{1 - sq}$$
$$s = \phi(s) = \frac{p}{1 - sq}$$

$$s - s^2 q = p \Longrightarrow s = \frac{1 \pm \sqrt{1 - 4pq}}{2q} \Longrightarrow s = 1 \text{ or } p/q$$

Stat 150 Stochastic Processes **Spring 2009**

Lecture 14: Branching Processes and Random Walks

Lecturer: Jim Pitman

Common setting: $p_0, p_1, p_2, \ldots$ probability distribution of X on $\{0, 1, 2, \ldots\}$.

- **Two surprisingly related problems:**

  Problem 1: Use X as offspring variable of a branching process $Z_0, Z_1, Z_2, \ldots$. Consider the random variable $\Sigma := \sum_{n=0}^{\infty} Z_n$, the *total progeny* of the branching process. Describe the distribution of Σ given $Z_0 = k$.

  Problem 2: Use $X - 1$ as the increment variable of a random walk $S_n^{(-)}$, $n = 0, 1, 2, \ldots$ on integer states. Let $T_0 := \inf\{n : S_n^{(-)}\} = 0$. Describe the distribution of T_0 given $S_0^{(-)} = k$.

- **Theorem:**
  • These two problems have the same solution.
  • Moreover, the solution is

$$\begin{aligned}\mathbb{P}(\Sigma = n | Z_0 = k) &= \mathbb{P}(T_0 = n | S_0^{(-)} = k) \\ &= \frac{k}{n}\mathbb{P}(S_n = n - k)\end{aligned}$$

  where $S_n := X_1 + \cdots + X_n$ is the sum of n independent copies $X_i \stackrel{d}{=} X$.

- Notes:
  • $\mathbb{P}(\Sigma < \infty | Z_0 = k) = \mathbb{P}(T_0 < \infty | S_0^{(-)} = k)$;
  Here $(\Sigma < \infty)$ is the event of extinction of the branching process, which given $k = 1$ has probability which is the least root $s \in [0, 1]$ of $s = \phi(s)$, and for general k is the kth power of the probability for $k = 1$. •. The formula is very explicit for any distribution of X for which there is a simple for the distribution of S_n. e.g. for $X \stackrel{d}{=} \text{Poisson}(\lambda)$, $S_n \stackrel{d}{=} \text{Poisson}(n\lambda)$, so

$$\mathbb{P}(S_n = n - k) = e^{-n\lambda}\frac{(n\lambda)^{n-k}}{(n-k)!}$$

- Approach both problems with the technique of probability generating functions.. This will show both solutions are the same. Then we can pick which one to work with to establish the formula.

Branching process problem: Introduce G.F. $G(s) := \sum_{n=0}^{\infty} s^n \mathbb{P}(\Sigma = n | Z_0 = 1)$. Notice that Σ given $Z_0 = k$ is the sum of k independent copies of Σ given $Z_0 = 1$,

$$[G(s)]^k = \sum_{n=0}^{\infty} s^n \mathbb{P}(\Sigma = n | Z_0 = k).$$

First step analysis: Start from $Z_0 = 1$, condition on $Z_1 = k$ for $k = 0, 1, 2, \ldots$ to see that

$$(\Sigma | Z_0 = 1, Z_1 = k) \stackrel{d}{=} (1 + \Sigma | Z_0 = k)$$

which gives

$$\begin{aligned} G(s) &= \sum_{k=0}^{\infty} p_k s [G(s)]^k \quad \text{where } s[G(s)]^k \text{ is the GF of } (1 + \Sigma | Z_0 = k) \\ &= s\phi(G(s)) \end{aligned}$$

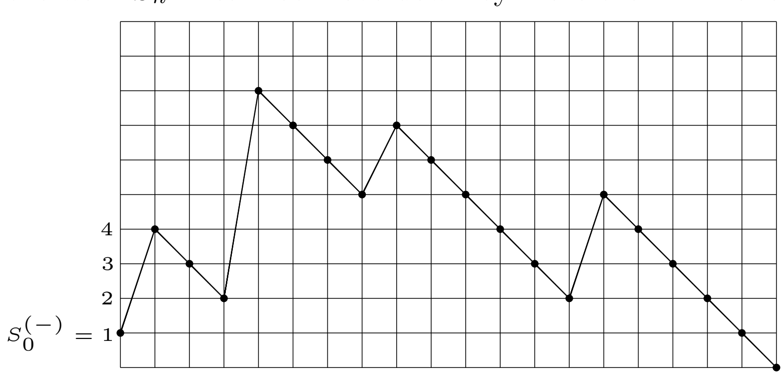
$$\Longrightarrow \frac{G(s)}{\phi(G(s))} = s$$

Define $F(z) := \dfrac{z}{\phi(z)}$, then $F(G(s)) = s$. So G is the functional inverse of $F(z)$. With care, you can argue this equation specifies $G(s)$ uniquely.

Random Walk Problem: Random walk $S_n^{(-)} := S_0^- + S_n - n$ for $n = 1, 2, \ldots$. Look at T_0 the first passage time to 0 starting at $S_0^- = k$.

Key Idea:

Given that $S_0^{(-)} = 1$ and $X_1 = k$, so $S_1^{(-)} = 1 + k - 1 = k$, the walk must pass down through each level $k-1, k-2, \ldots, 1$ before 0. This exploits the fact that the walk $S_n^{(-1)}$ cannot move down by more than 1 in one step.



So according to the Strong Markov Property, given $S_0^{(-1)} = 1$ and $S_1^{(-)} = k$, the distribution of T_0 becomes like 1+ the sum of k indep copies of T_0 given $S_0^{(-1)} = 1$, one copy for each of the passages from k to $k-1$ to $\ldots$ to 1 to 0.

<u>First step analysis</u>:

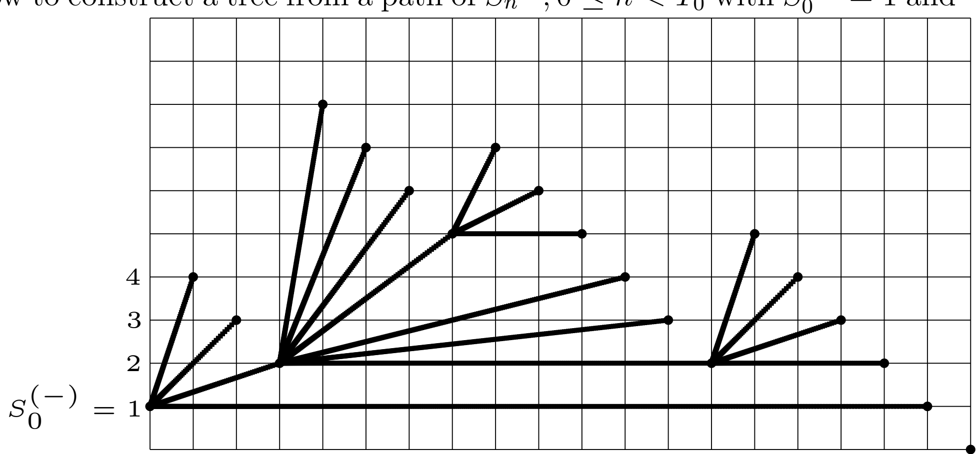
$$(T_0 | S_0^{(-1)} = 1, S_1^{(-)} = k) \stackrel{d}{=} (1 + T_0 | S_0^{(-)} = k)$$

Let $H(s)$ = G.F. of T_0 given $S_0^{(-)} = 1$. Condition on value of X_1, so $S_1^{(-)} = S_0^{(-)} + X_1 - 1$. Given $X_1 = k$, see 1 plus k independent copies of T_0. Hence

$$H(s) = s\phi(H(s)).$$

Same as the previous equation for $G(s)$, the generating function of Σ given $Z_0 = 1$. Can argue uniqueness of the solution. This explains analytically why the two problems have the same solution.

Where is the branching process in the random walk? The following diagram shows how to construct a tree from a path of $S_n^{(-)}, 0 \leq n < T_0$ with $S_0^{(-)} = 1$ and



$T_0 < \infty$:
In this graph, the walk $S_n^{(-)}, 0 \leq n < T_0$ with $S_0^{(-)} = 1$ is

$$1, 4, 3, 2, 8, 7, 6, 5, 7, 6, 5, 4, 3, 2, 5, 4, 3, 2, 1, 0 \text{ with } T_0 = 19$$

The increments of this walk the $X_n - 1$ for the following sequence of X-values:

$$4, 0, 0, 7, 0, 0, 0, 3, 0, 0, 0, 0, 0, 4, 0, 0, 0, 0, 0 \text{ with sum } S_19 = 18.$$

The associated tree has one root node and 18 child nodes. The numbers Z_n of nodes in successive levels of the tree are

$$1, 4, 7, 7, 0, 0, \ldots \text{ with sum } \sum_{n=0}^{\infty} Z_n = 1 + 18 = 19.$$

In general associated with every walk path starting at $S_0^{(-)} = 1$ with $T_0 < \infty$, the tree can be constructed as follows:

- The nodes of the tree are the pairs $(n, S_n^{(-1)})$ with $0 \le n < T_0$, so the tree has tree has root $(0,1)$, and total number of nodes of the tree is T_0.
- For each $0 \le m < T_0$ such that $S_{m+1}^{(-)} - S_m^{(-)} = j - 1$, meaning that $X_{m+1} = j$, the node $(m, S_m^{(-)})$ has exactly j children, which for $j \ge 1$ are $(m+1, S_{m+1}^{(-)})$ and the next $j-1$ *descending ladder points* of the path $(r, S_r^{(-)}), r = m+2, m+3, \ldots$, that is to say points $(\ell, S_\ell^{(-)})$ such that $S_\ell^{(-)} < \min_{m+1 \le r < \ell} S_r^{(-)}$, the last of these points occurring when the walk path first returns to the level $S_m^{(-)}$.

In general, it can be checked that the random tree derived this way from the path of $S_n^{(-)}, 0 \le n < T_0$ with $S_0^{(-)} = 1$ and $T_0 < \infty$ is distributed on the event $T_0 < \infty$ exactly as if it were the random tree associated with a branching process with offspring numbers distributed like X, on the event that the branching process becomes extinct. This explains why the distribution of T_0 given $S_0^{(-)} = 1$ is identical to the distribution of $\Sigma := \sum_n Z_n$ given $Z_0 = 1$, without appealing to the theory of generating functions. The argument can be carried further to explain why the distribution of T_0 given $S_0^{(-)} = k$ is identical to the distribution of Σ given $Z_0 = k$. For in each problem it is clear that the random variable of interest with initial condition k is the sum of k independent cpoies of the random variable with initial condition $k = 1$.

<u>Next</u>: Having established the identity of solutions of the two problems, it remains to verify the explicit formula. Put in a slightly different but obviously equivalent way, the formula reads:

$$\mathbb{P}(T_{(-k)} = n) = \frac{k}{n}\mathbb{P}(S_n - n = k),$$

where $T_{(-k)} := \inf\{n : S_n - n = -k\}$ with $S_n := X_1 + \cdots + X_n$ for independent and identically distributed non-negative integer random variables X_i with arbitrary distribution $p_0, p_1, p_2, \ldots$.

Left side: $\mathbb{P}(T_{(-k)} = n) = \mathbb{P}(\text{the walk first reaches } -k \text{ at time } n)$
Right side: $\mathbb{P}(S_n - n = -k) = \mathbb{P}(\text{the walk is at } -k \text{ at time } n)$

So the formula says that given the walk is at walk is at $-k$ at time n, the chance that it first reached that level at time n is k/n. Why such a simple relation? This relation is implied by a combinatorial fact about sequences with values in $\{-1, 0, 1, 2, \ldots\}$.

- Take any sequence of integers in $\{-1, 0, 1, 2, \dots\}$ of length n, say $y_1, y_2, \dots, y_n$. E.g. $-1, 2, 3, -1, 3, -1, \dots$, with $\Sigma := \sum_{i=1}^n y_i = -k < 0$. Look at the n cyclic shifts of the sequence:

$$\begin{aligned} y_1, y_2, \dots, y_{n-1}, y_n \qquad & \text{with } \Sigma = -k \\ y_2, y_3, \dots, y_n, y_1 \qquad & \text{with } \Sigma = -k \\ \vdots \qquad & \\ y_n, y_1, \dots, y_{n-2}, y_{n-1} \qquad & \text{with } \Sigma = -k \end{aligned}$$

  Among the n cyclic shifts, count how many of them have partial sums which reach $-k$ for the first time at the nth step.

  The Cycle Lemma. No matter what the sequence $y_1, y_2, \dots, y_n$ with values in $\{-1, 0, 1, 2, \dots\}$ with $\sum_{i=1}^n y_i = -k < 0$, exactly k of the cyclic shifts have the property that their partial sums which reach $-k$ for the first time at the nth step.

  Sketch of Proof There is always at least one such cyclic shift, obtained by starting the sequence just after the first time the partial sums attain their minimum. So there is no loss of generality in supposing that the original sequence has the property. Now $k-1$ more cyclic shifts with the property are obtained from the $k-1$ descending ladder indices when the partial sums first reach level $-j$ for $1 \le j < k$, and it can be checked that there are no more shifts with this property. This is clear from a picture, where a copy of the original walk path with n steps is attached to the end of the original walk. The walk paths with cyclic shifts are derived from overlapping stretches of length n within this path of length $2n$.

- <u>Now conclude</u>: Let $B_{nk} := \{T_{(-k)} = n\}$ for $X_1, X_2, \dots$ with $\sum_{i=1}^n (X_i - 1) = -k < 0$, and notice that B_{nk} is determined by the first n variables $X_1, X_2, \dots, X_n$. Let $B_{nk}^{(i)}$ denote the same event with $X_1, X_2, \dots, X_n$ replaced by the ith cyclic shift of these variables. Since the distribution of the sequence $X_1, X_2, \dots, X_n$ is unchanged by a permutation of its indices

$$P(B_{nk}^{(i)}) = \mathbb{P}(B_{nk}) \qquad \text{for all } 1 \le i \le n, 1 \le k \le n.$$

  On the other hand, by the cycle lemma applied to $Y_i = X_i - 1, 1 \le i \le n$ whose sum is $S_n - n$, no matter what the values of Y_i, the number of events $B_{nk}^{(i)}$ which occur is k if $S_n - n = -k$, and obviously zero otherwise. In terms of indicator random variables:

$$\sum_{i=1}^n \mathbf{1}_{B_{nk}}^{(i)} = k\mathbf{1}(S_n - n = -k).$$

Take expections of both sides to deduce that

$$n\mathbb{P}(B_{nk}) = k\mathbb{P}(S_n - n - k)$$

and the conclusion follows.

References: For more about the functional equation for generating functions

$$G(z) = z\phi(G(z))$$

appearing here, and its analytic, combinatorial and probabilistic interpretations, see

- Richard P. Stanley. *Enumerative Combinatorics, Volume 2* (CUP 1999). Section 5.4, The Lagrange Inversion Formula. `http://openlibrary.org/b/OL7754023M`
- Jim Pitman. *Enumerations of trees and forests related to branching processes and random walks*. In *Microsurveys in Discrete Probability*, D. Aldous and J. Propp editors, DIMACS Ser. Discrete Math. Theoret. Comp. Sci No. 41, 163-180, Amer. Math. Soc., Providence RI (1998). `http://www.stat.berkeley.edu/tech-reports/482.pdf`

Stat 150 Stochastic Processes **Spring 2009**

Lecture 17: Limit distributions for Markov Chains

Lecturer: Jim Pitman

- Finite state space Markov chain $X_0, X_1, \dots$ with state space S having N elements, $N < \infty$. Matrix P, $\mathbb{P}_i(X_n = j) = P^n(i,j)$.

  Problem: Suppose you know the initial distribution λ of $X_0, \lambda_j = \mathbb{P}(X_0 = j)$. Want to evaluate $\lim_{n\to\infty} \mathbb{P}_\lambda(X_n = j)$.

  Notation: $\mathbb{P}_\lambda(\cdot) := \sum_i \lambda_i \mathbb{P}_i(\cdot)$

  Questions:

  - When does limit F exist?
  - If F exists, how to evaluate it?

- We say P is **regular** if $\exists n : P^n(i,j) > 0$ for all $i, j \in S$.

  Equivalently, there exists for every i and j some sequence $i_0 = i, i_1, i_2, \dots, i_n = j$ with $P(i_k, i_{k+1}) > 0$ for all $1 \le k \le n$ some path from i to j in exactly n steps.

  Obviously $P^n(i,j) > 0, \forall i,j \implies P^m(i,j) > 0, \forall i,j, m \ge n$:
  Let $m = n + k$,
  $$P^{n+k}(i,j) = \sum_l P^k(i,l) P^n(l,j) > 0$$
  because $P^n(l,j) > 0$ for all n and $\sum_l P^k(i,l) = 1 \Rightarrow P^k(i,l) > 0$ for some l.

- **Theorem:** If P is a regular transition matrix on a finite set, then there exists a unique probability distribution π on S such that $\pi P = \pi$ (π is called the stationary, equilibrium, invariant, or steady state distribution) and $\lim_{n\to\infty} \mathbb{P}_\lambda(X_n = j) = \pi_j$.

  Remarks:

  - $\pi_j > 0$, for all j.
  - Let $T_j :=$ first $n \ge 1 : X_n = j$, then $\pi_j = \dfrac{1}{\mathbb{E}_j T_j}$.
  - $\dfrac{\pi_j}{\pi_i} = \mathbb{E}_i(\#$ of hits on j before $T_i)$. This can be proved by using previous techniques.

– Take $\lambda_j = \mathbf{1}(j = i),\ \lambda P^n(j) = P^n(i,j) \to \pi_j$. Start in state $i,\ X_0 = i$.

Discussion: If such limits $\lim_{n\to\infty} \lambda P^n(j)$ exist for all j, say $\lim_{n\to\infty} \lambda P^n(j) = \pi_j$, then $\pi P = \pi$.

Proof:

$$\begin{aligned}
\pi_j &= \lim_{n\to\infty} \lambda P^n(j) \\
&= \lim_{n\to\infty} \sum_{i\in S} (\lambda P^n)_i P(i,j) \\
&= \sum_i \pi_i P(i,j) \\
&= \pi P(j)
\end{aligned}$$

$\left[\pi \text{ is a probability distribution }, \pi P = \pi \right]$ is a system of $N+1$ equations with N unknowns:

$$\sum_{i=1}^N \pi_i = 1$$

$$\sum_{i=1}^N \pi_i P(i,j) = \pi_j,\ \text{for } 1 \le j \le N$$

There is one too many equation because the P matrix has rows sum to 1.

- Trick/Technique: **Reversible equilibrium**

  π is an equilibrium distribution if given $X_0 \stackrel{d}{=} \pi$, then

  $$\pi \stackrel{d}{=} X_0 \stackrel{d}{=} X_1 \stackrel{d}{=} X_2 \stackrel{d}{=} \cdots \stackrel{d}{=} X_n$$

  Sometimes $(X_0, X_1) \stackrel{d}{=} (X_1, X_0)$; that is

  $$\begin{aligned}
  \mathbb{P}(X_0 = i, X_1 = j) &= \mathbb{P}(X_1 = i, X_0 = j) \\
  &= \mathbb{P}(X_0 = j, X_1 = i)
  \end{aligned}$$

  With $\pi_i = \mathbb{P}(X_0 = i)$, this becomes

  $$\pi_i P(i,j) = \pi_j P(j,i)$$

  This system of equations $\boxed{\pi_i P(i,j) = \pi_j P(j,i),\ \ i,j \in S}$ is called **system of reversible equilibrium equations**, also called *detailed balance* equations. Notice that this system has $\binom{N}{2}$ equations with N unknowns, and there is still the additional constraint $\sum_i \pi_i = 1$. Observe: If π solves REV EQ EQNS, then π solves the usual EQ EQNS:

$$\text{Suppose} \qquad \pi_i P(i,j) = \pi_j P(j,i), \text{ for all } i,j \in S$$
$$\text{sum over } i: \qquad (\pi P)_j = \sum_i \pi_i P(i,j) = \pi_j \sum_i P(j,i) = \pi_j \implies \pi P = \pi$$

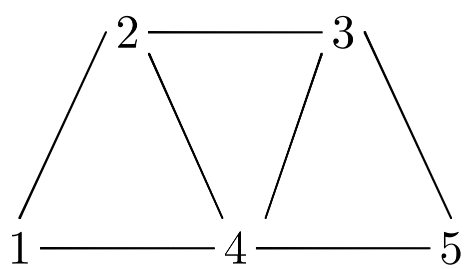
It is usually easy to tell if the REV EQ EQNS have a solution, and if so to evaluate it. Starting from the value π_i any particular state i, the value of π_j for every state j with $P(i,j) > 0$ is immediately determined. Continuing this way determines π_k for every state k with can be reached from i in two steps, say via j with $P(i,j)P(j,k) > 0$, and so on. The only issue with this procedure is if two different paths from i to k lead force two different values for π_k, in which case the REV EQ EQNS have no solution.

- **Equilibrium Mass Transfer** Imagine mass π_i located at state i. At each step, proportion $P(i,j)$ of the mass at i is moved to j. Equilibrium means that after the transfer, the mass distribution is the same. Reversible equilibrium means this is achieved because for each pair of states i and j, the mass transferred from i to j equals the mass transferred from j to i.

- Example Nearest neighbour R.W. on graph.

  This is really a meta-example, with large numbers of instances.

  Graph G on set S, G is a collection of edges $\{i,j\}$ for $i,j \in S$.
  $P(i,j) = \frac{1}{N(i)} \mathbf{1}(i \text{ neighbour of } j)$
  $N(i) = \#$ of neighbours of i



  Easy fact: The measure $(N(i), i \in S)$ is always a reversible equilibrium for RW on graph. $N(i)P(i,j) = N(j)P(j,i)$, for all $i,j \in S$ because these equations read either $0 = 0$ if i is not a neighbour of j or $1 = 1$ if i is a neighbour of j.

  For instance

$$(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5) = (2/14, 3/14, 3/14, 4/14, 2/14)$$

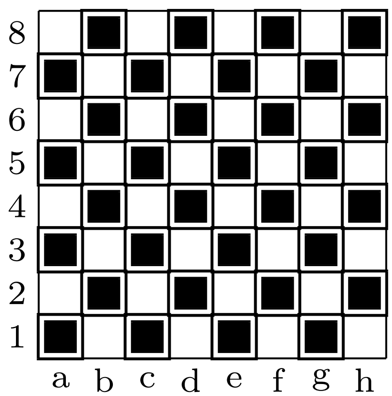
  in the example above.

Stat 150 Stochastic Processes **Spring 2009**

Lecture 18: Markov Chains: Examples

Lecturer: Jim Pitman

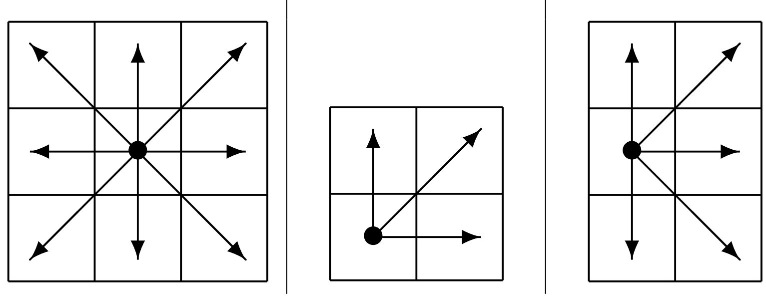
- A nice collection of random walks on graphs is derived from random movement of a chess piece on a chess board. The state space of each walk is the set of $8 \times 8 = 64$ squares on the board:



Each kind of chess piece at a state i on an otherwise empty chess board has some set of all states j to which it can allowably move. These states j are the *neighbours of* i in a graph whose vertices are the 64 squares of the board. Ignoring pawns, for each of king, queen, rook, bishop and knight, if j can be reached in one step from i, this move can be reversed to reach i from j. Note the pattern of black and white squares, which is important in the following discussion.

The King

interior	corner	edge
8 possible moves	3 possible moves	5 possible moves



In the graph, $i \longleftrightarrow j$ means that j is a king's move from i, and i is a king's move from j. Observe that $N(i) := \#(\text{ possible moves from } i) \in \{3, 5, 8\}$. From general discussion of random walk on a graph in previous lecture, the reversible equilibrium distribution is $\pi_i = \dfrac{N(i)}{\Sigma}$ where

$$\Sigma := \sum_j N(j) = (6 \times 6) \times 8 + (4 \times 6) \times 5 + 4 \times 3$$

Question: Is this walk regular? Is it true that $\exists m : P^m(i, j) > 0$ for all i, j? Yes. You can easily check this is so for $m = 7$.

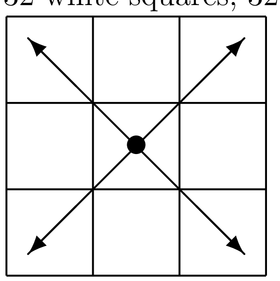
Rook and Queen

Very similar treatment. Just change the values of $N(i)$.

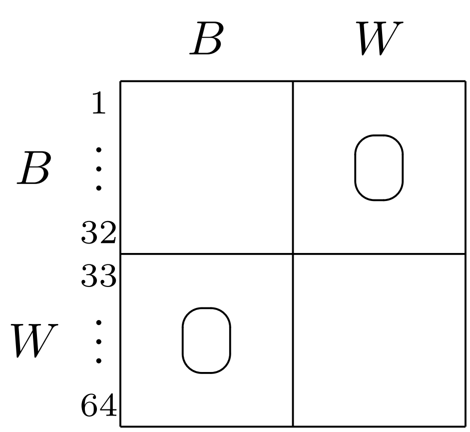
Both walks are regular because P^2 is strictly positive: even the rook can get from any square to any other square in two moves.

Bishop

32 white squares, 32 black squares. Bishop stays on squares of one colour.

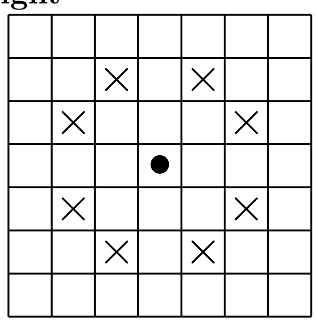


Is this chain regular? No, because the bishop's walk is **reducible**: there are two disjoint sets of states (B and W). For each $i, j \in B$, $\exists N; P^N(i, j) > 0$. In fact, $N = 3$. Similarly for each $i, j \in W$, $P^3(i, j) > 0$, hence $P^N(i, j) > 0$ for all $N \geq 3$. But B and W do not communicate.



The matrix is decomposed into two: $B \longleftrightarrow B, W \longleftrightarrow W$. If the bishop starts on a black square, it moves as if a Markov chain with state space B, and the limit distribution is the equilibrium distribution for the random walk on B, with $\pi_i = \dfrac{N(i)}{\sum}$ as before, but now $\sum = \sum_{j \in B} N(j)$, with 32 terms, rather than a sum over all 64 squares. This is typical of a Markov chain with two disjoint communicating classes of states.

Knight



Is it regular? (No)
After an even number of steps n, $P^n(i,j) > 0$ only if j is of the same color as i. No matrix power of P is strictly > 0 at all entries $\Longrightarrow$ not regular.

- **Periodicity** Fix a state i. Look at the set of $n : P^n(i,i) > 0$. For the knight's move $\{n : P^n(i,i) > 0\} = \{2,4,6,8,\dots\}$. You cannot return in an odd number of steps because knight's squares change colour $B \to W \to B$ at each step.

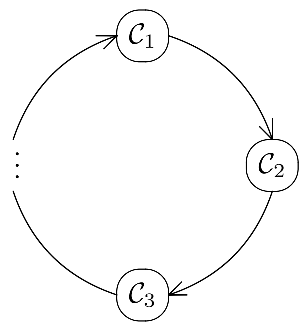
  Definition: Say P is **irreducible** if for all states i, j, $\exists n : P^n(i,j) > 0$.
  This condition is weaker than regular, which requires the n to work for all i, j.

  Deal with periodic chains: Suppose P is irreducible, say state i has period $d \in \{1,2,3,\dots\}$ if the greatest common divisor of $\{n : P^n(i,i) > 0\} = d$. E.g., $d = 2$ for each state i for the knight's walk.

  Fact (Proof see Feller Vol.1): For an irreducible chain, every state i has the same period.

  - $d = 1$: called *aperiodic*
  - $d = 2, 3, \dots$: called *periodic with period d*

  In periodic case, $\exists$ d disjoint sets of states, $\mathcal{C}_1, \dots, \mathcal{C}_d$ (Cyclically moving subclasses): $P(i,j) > 0$ only if $i \in \mathcal{C}_m, j \in \mathcal{C}_{m+1}$ $(m+1 \bmod d)$.



<u>Note</u> If $\exists i : P(i,i) > 0$, then i has period 1. Then, assuming P is irreducible, all states have period 1, and the chain is aperiodic.

<u>Fact</u> An irreducible, aperiodic chain on finite S is regular.

- **Death and immigration chain**

  Population story. State space $S = \{0, 1, 2, \dots\}$. Let $X_n \in S$ represent the number of individuals in some population at time n.

  <u>Dynamics</u>: Between times n and $n+1$,

  - each individual present at time n dies with probability p and remains with probability $q := 1 - p$.
  - add an independent Poisson(λ) number of immigrants.

  <u>Problem</u> Describe limit behavior of X_n as $n \to \infty$. Natural first step: write down the transition matrix.
  For $i \geq 0, j \geq 0$, condition on number of survivors:

$$P(i,j) = \sum_{k=0}^{i} \binom{i}{k} q^k p^{i-k} e^{-\lambda} \frac{\lambda^{j-k}}{(j-k)!} \mathbf{1}(k \leq j)$$

  Now try to solve the equations $\pi P = \pi$. Very difficult!

- <u>Idea</u>: Suppose we start at state $i = 0, X_0 = 0$,

$$\begin{aligned}
&X_1 \sim \text{Poi}(\lambda) \\
&X_2 \sim \text{Poi}(\lambda q + \lambda) \qquad \text{by thinning and “+” rule for Poisson} \\
&X_3 \sim \text{Poi}((\lambda q + \lambda) q + \lambda) \\
&\quad \vdots \\
&X_n \sim \text{Poi}(\lambda(1 + q + \cdots + q^{n-1})) = \text{Poi}\left(\lambda \frac{1 - q^n}{1 - q}\right) \to \text{Poi}(\lambda/p)
\end{aligned}$$

So given $X_0 = 0$, we see $X_n \xrightarrow{d} \text{Poi}(\lambda/p)$.

If we start with $X_n > 0$, it is clear from the dynamics that we can write

$$X_n = X_n^* + Y_n$$

$$\text{where} \qquad \begin{aligned} X_n^* &= \text{ survivors of the initial population} \\ Y_n &= \text{ copy of the process given } (X_n = 0) \end{aligned}$$

Note that

$$X_n^* \to 0, \qquad Y_n \to \text{Poi}(\lambda/p)$$

Conclusion: $X_n \xrightarrow{d} \text{Poi}(\lambda/p)$ no matter what X_0 is.

- **Ad hoc analysis** — special properties of Poisson.

  General idea: What do you expect in the limit? Answer: stationary distribution $\pi P = \pi$. Here π is $\text{Poi}(\lambda/p)$.

  Does $\pi P = \pi$? That is, does $X_0 \sim \text{Poi}(\lambda/p)$ imply $X_1 \sim \text{Poi}(\lambda/p) \ldots$?

  Check: Say $X_0 \sim \text{Poi}(\mu)$, then by thinning and addition rules for Poisson, as before $X_1 \sim \text{Poi}(\mu q + \lambda)$. So

  $$X_1 \stackrel{d}{=} X_0 \iff \mu = \mu q + \lambda \iff \mu = \frac{\lambda}{1-q} = \frac{\lambda}{p}$$

  Therefore, the unique value of μ which makes $\text{Poi}(\mu)$ invariant for this chain is $\mu = \lambda/p$. In fact, this is the unique stationary distribution for the chain. This follows from the previous result that no matter what the distribution of X_0, the distribution of X_n converges to $\text{Poi}(\lambda/p)$. Because if π was some other stationary distribution, if we started the chain with $X_0 \sim \pi$, then $X_n \sim \pi$ for every n, and the only way this can converge to $\text{Poi}(\lambda/p)$ is if in fact $\pi = \text{Poi}(\lambda/p)$.

Stat 150 Stochastic Processes **Spring 2009**

Lecture 19: Stationary Markov Chains

Lecturer: Jim Pitman

- Symmetry Ideas: General idea of symmetry: make a transformation, and something stays the same. In probability theory, the transformation may be conditioning, or some rearrangement of variables. What stays the same is the distribution of something. For a sequence of random variables $X_0, X_1, X_2, \ldots$, various notions of symmetry:

  (1) Independence. X_0 and X_1 are independent. Distribution of X_1 given X_0 does not involve X_0; that is, $(X_1 | X_0 \in A) \stackrel{d}{=} X_1$.

  (2) Identical distribution. The X_n are identically distributed: $X_0 \stackrel{d}{=} X_n$ for every n. Of course IID $\Longrightarrow$ LLN / CLT.

- **Stationary:** $(X_1, X_2, \ldots) \stackrel{d}{=} (X_0, X_1, \ldots)$ (shifting time by 1)
  By measure theory, this is the same as

$$(X_1, X_2, \ldots, X_n) \stackrel{d}{=} (X_0, X_1, \ldots, X_{n-1}), \text{ for all } n.$$

  Obviously IID $\Longrightarrow$ Stationary. But the converse is not necessarily true. E.g., $X_n = X_0$ for all n, for any non-constant random variable X_0.

  Another example: Stationary MC, i.e., start a MC with transition matrix P with initial distribution π. Then easily, the following are equivalent:

$$\pi P = \pi$$

$$X_1 \stackrel{d}{=} X_0$$

$$(X_1, X_2, \ldots) \stackrel{d}{=} (X_0, X_1, \ldots)$$

  . Otherwise put, (X_n) is stationary means $(X_1, X_2, \ldots)$ is a Markov chain with exactly the same distribution as $(X_0, X_1, \ldots)$.

- Other symmetries:

  - Cyclic symmetry: $(\underbrace{X_1, X_2, \ldots, X_n}_{n}) \stackrel{d}{=} (\underbrace{X_2, X_3, \ldots, X_n, X_1}_{n})$
  - Reversible: $(X_n, X_{n-1}, \ldots, X_1) \stackrel{d}{=} (X_1, X_2, \ldots, X_n)$

- Exchangeable: $(X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)}) \stackrel{d}{=} (X_1, X_2, \ldots, X_n)$ for every permutation π of $(1, \ldots, n)$.

- Application of stationary idea to MC's.

  Think about a sequence of RVs $X_0, X_1, \ldots$ which is stationary. Take a set A in the state space. Let $T_A = \begin{cases} \text{least } n \geq 1 (\text{if any}) \text{ s.t. } X_n \in A \\ \infty \text{ if no such } n \end{cases}$; that is, $T_A :=$ least $n \geq 1 : \mathbf{1}_A(X_n) = 1$. Consider the $\mathbf{1}_A(X_n)$ process.

  E.g., 0 0 0 0 1 0 0 1 1 0 0 0 . . .
  To provide an informal notation:

$$\mathbb{P}(T_A = n) = \mathbb{P}(?\ \underbrace{0\,0\ldots0\,0}_{n-1 \text{ zeros}}\ 1)$$

$$\mathbb{P}(T_A \geq n) = \mathbb{P}(?\ \underbrace{0\,0\ldots0\,0}_{n-1 \text{ zeros}}\ ?)$$

$$\mathbb{P}(X_0 \in A, T_A \geq n) = \mathbb{P}(1\ \underbrace{0\,0\ldots0\,0}_{n-1 \text{ zeros}}\ ?)$$

  Lemma: For every stationary sequence $X_0, X_1, \ldots,$

$$\mathbb{P}(T_A = n) = \mathbb{P}(X_0 \in A, T_A \geq n) \qquad \text{for } n = 1, 2, 3, \ldots$$

  With the informal notation, the claim is:

$$\mathbb{P}(?\ \underbrace{0\,0\ldots0\,0}_{n-1 \text{ zeros}}\ 1) = \mathbb{P}(1\ \underbrace{0\,0\ldots0\,0}_{n-1 \text{ zeros}}\ ?)$$

  <u>Notice.</u> This looks like reversibility, but reversibility of the sequence $X_0, X_1, \ldots, X_n$ is *not* being assumed. Only stationarity is required!

  <u>Proof</u>

$$\begin{array}{ccc}
P(?\ \underbrace{0\,0\ldots0\,0}_{n-1 \text{ zeros}}\ 1) & & P(1\ \underbrace{0\,0\ldots0\,0}_{n-1 \text{ zeros}}\ ?) \\
+ & & + \\
P(?\ \underbrace{0\,0\ldots0\,0}_{n-1 \text{ zeros}}\ 0) & \stackrel{stationarity}{=} & P(0\ \underbrace{0\,0\ldots0\,0}_{n-1 \text{ zeros}}\ ?) \\
\| & & \| \\
P(?\ \underbrace{0\,0\ldots0\,0}_{n-1 \text{ zeros}}\ ?) & = & P(?\ \underbrace{0\,0\ldots0\,0}_{n-1 \text{ zeros}}\ ?)
\end{array}$$

Take the identity above, sum over $n = 1, 2, 3, \ldots,$

$$\begin{aligned}
\mathbb{P}(T_A < \infty) &= \sum_{n=1}^{\infty} \mathbb{P}_A(T = n) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(X_0 \in A, T_A \geq n) \\
&= \mathbb{E} \sum_{n=1}^{\infty} \mathbf{1}(X_0 \in A, T_A \geq n) \\
&= \mathbb{E}(T_A \mathbf{1}(X_0 \in A))
\end{aligned}$$

This is *Marc Kac's identity*: For every stationary sequence (X_n), and every measurable subset A of the state space of (X_n):

$$\mathbb{P}(T_A < \infty) = \mathbb{E}(T_A \mathbf{1}(X_0 \in A))$$

- Application to recurrence of Markov chains: Suppose that an irreducible transition matrix P on a countable space S has a stationary probability measure π, that is $\pi P = \pi$, with $\pi_i > 0$ for some state i. Then
  - π is the unique stationary distribution for P
  - $\pi_j > 0, \forall j$.
  - $\mathbb{E}_j T_j < \infty, \forall j$.
  - $\mathbb{E}_j T_j = \frac{1}{\pi_j}, \forall j$.

  Proof:

  Apply Kac's identity to $A = \{i\}$, with $\mathbb{P} = \mathbb{P}_\pi$ governing (X_n) as a MC with transition matrix P started with $X_0 \stackrel{d}{=} \pi$.

$$1 \geq \mathbb{P}_\pi(T_i < \infty) \stackrel{Kac}{=} \mathbb{E}_\pi[T_i \mathbf{1}(X_0 = i)] = \pi_i \mathbb{E}_i T_i$$

  Since $\pi_i > 0$ this implies first $\mathbb{E}_i T_i < \infty$, and hence $\mathbb{P}_i(T_i < \infty) = 1$. Next, need to argue that $P_j(T_i < \infty) = 1$ for all states j. This uses irreducibility, which gives us an n such that $P^n(i,j) > 0$, hence easily an $m \leq n$ such that it is possible to get from i to j in m steps without revisiting i on the way, i.e.

$$\mathbb{P}_i(T_i > m, X_m = j) > 0$$

  But using the Markov property this makes

$$\mathbb{P}_j(T_i < \infty) = \mathbb{P}_i(T_i < \infty \,|\, T_i > m, X_m = j) = 1$$

Finally
$$\mathbb{P}_\pi(T_i < \infty) = \sum_j \pi_j \mathbb{P}_j(T_i < \infty) = \sum_j \pi_j = 1$$
and the Kac formula becomes $1 = \pi_i \mathbb{E}_i T_i$ as claimed. Lastly, it is easy that $\pi_j > 0$ and hence $1 = \pi_j \mathbb{E}_j T_j$ because $\pi = \pi P$ implies $\pi = \pi P^n$ and so
$$\pi_j = \sum_k \pi_k P^n(k,j) \geq \pi_i P^n(i,j) > 0$$
for some n by irreducibility of P.

Stat 150 Stochastic Processes **Spring 2009**

Lecture 20: Markov Chains: Examples

Lecturer: Jim Pitman

- A nice formula:

$$\mathbb{E}_i(\text{number of hits on } j \text{ before } T_i) = \frac{\pi_j}{\pi_i}$$

Domain of truth:

  - P is irreducible.
  - π is an invariant measure: $\pi P = \pi$, $\pi_j \geq 0$ for all j, and $\sum \pi_j > 0$
  - Either $\sum \pi_j < \infty$ or (weaker) $\mathbb{P}_i(T_i < \infty) = 1$ (recurrent).

Positive recurrent case: if $\mathbb{E}_i(T_i) < \infty$, then π can be a probability measure. Null recurrent case: if $\mathbb{P}_i(T_i < \infty) = 1$ but $\mathbb{E}_i(T_i) = \infty$, then π cannot be a probability measure.

- The formula above is a refinement of the formula $\mathbb{E}_i(T_i) = 1/\pi_i$ for a positive recurrent chain with invariant π with $\sum_j \pi_j = 1$. To see this, observe that

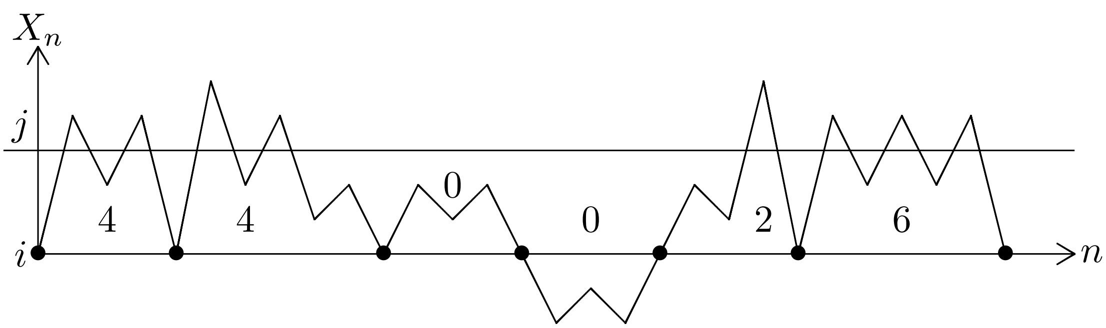
$$N_{ij} := \text{number of hits on } j \text{ before } T_i$$
$$N_{ij} = \sum_{n=0}^{\infty} \mathbf{1}(X_n = j, n < T_i)$$

$$\begin{aligned} T_i &= \sum_{n=0}^{\infty} \mathbf{1}(n < T_i) \\ &= \sum_{n=0}^{\infty} \underbrace{\sum_{j \in S} \mathbf{1}(X_n = j)}_{=1} \mathbf{1}(n < T_i) \\ &= \sum_{j \in S} \sum_{n=0}^{\infty} \mathbf{1}(X_n = j, n < T_i) \\ &= \sum_{j \in S} N_{ij} \end{aligned}$$

Then

$$\mathbb{E}_i T_i = \mathbb{E}_i \sum_j N_{ij} = \sum_j \mathbb{E}_i N_{ij} = \sum_j \frac{\pi_j}{\pi_i} = \frac{1}{\pi_i} \quad \text{with } \sum_j \pi_j = 1$$

- Strong Markov property : Chain refreshes at each visit to $i \Longrightarrow$ numbers of visits to j in successive i-blocks are iid.



Sequence $N_{ij}^{(1)}, N_{ij}^{(2)}, N_{ij}^{(3)}, \ldots$ Each $N_{ij}^{(k)}$ is an independent copy of $N_{ij} = N_{ij}^{(1)}$. So $\mathbb{E}_i N_{ij}$ = expected # of j's in an i-block, and according to the law of large numbers:

$$\underbrace{\frac{1}{m} \sum_{k=1}^{m} N_{ij}^{(k)}}_{\text{average \# of visits to } j \text{ in the first } m \text{ i-blocks}} \qquad \longrightarrow \mathbb{E}_i N_{ij}$$

Push this a bit further:

$$\sum_{n=1}^{N} \mathbf{1}(X_n = j) = \text{ \# of } j\text{'s in the first } N \text{ steps.}$$
$$\sum_{n=1}^{N} \mathbf{1}(X_n = i) = \text{ \# of } i\text{'s in the first } N \text{ steps.}$$

Previous discussion: In the long run, the # of j's / (visits to i) $\longrightarrow \mathbb{E}_i N_{ij}$:

$$\frac{\sum_{n=1}^{N} \mathbf{1}(X_n = j)}{\sum_{n=1}^{N} \mathbf{1}(X_n = i)} = \frac{\text{\# of } j\text{'s in the first } N \text{ steps}}{\text{\# of } i\text{'s in the first } N \text{ steps}} \longrightarrow \mathbb{E}_i N_{ij}$$

But also in positive recurrent case, for simplicity, assume aperiodic/regular:

$$\mathbb{P}_i(X_n = j) \longrightarrow \pi_j$$
$$\mathbb{E}_i\left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{1}(X_n = j)\right) = \frac{1}{N} \sum_{n=1}^{N} \underset{\substack{\downarrow \\ \pi_j}}{P^n(i,j)} \longrightarrow \pi_j$$

because averages of a convergent sequence tend to the same limit.

What's happening of course is

$$\left.\begin{array}{l} \frac{1}{N}\sum_{n=1}^{N}\mathbf{1}(X_n=j)\longrightarrow\pi_j \\[2ex] \frac{1}{N}\sum_{n=1}^{N}\mathbf{1}(X_n=i)\longrightarrow\pi_i \end{array}\right\}\Longrightarrow \frac{\sum_{n=1}^{N}\mathbf{1}(X_n=j)}{\sum_{n=1}^{N}\mathbf{1}(X_n=i)}\longrightarrow\frac{\pi_j}{\pi_i}$$

Hence $\mathbb{E}_i N_{ij} = \pi_j/\pi_i$.

- Example

  $p\uparrow q\downarrow$ walk on $\{0,1,2,\ldots\}$ with partial reflection at 0.

  $0<p<1$

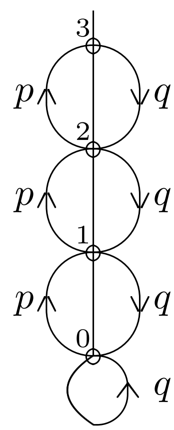
  $P(0,1) = p$

  $P(0,0) = q$

  for $i=1,2,\ldots$

  $P(i,i+1) = p$

  $P(i,i-1) = q$



Analysis:

1) $p>q$: walk is transient (goes to infinity)
2) $p=q$: walk is null recurrent (keep returning to 0 with infinite mean return time)
3) $p<q$: walk is positive recurrent, $\lim\mathbb{P}_i(X_n=j)=\pi_j$ for (π_i) a probability measure

We will focus on 3) and the determination of π.

Appeal to result from previous class. If $\exists$ a stationary π with $\sum_j \pi_j = 1$, then P is positive recurrent and $\mathbb{E}_i(T_i)=1/\pi_i$.

Advice: always look for a reversible equilibrium. Say the equilibrium distribution is μ_j.

$$\text{REV EQ} \quad \mu_i P_{ij} = \mu_j P_{ji}$$

$$\begin{array}{l} i/j \\ 0/1 : \mu_0 p = \mu_1 q \\ 1/2 : \mu_1 p = \mu_2 q \\ 2/3 : \mu_2 p = \mu_3 q \\ \vdots \end{array} \quad \Longrightarrow \quad \begin{array}{l} \mu_1 = \mu_0(p/q) \\ \mu_2 = \mu_1(p/q) = \mu_0(p/q)^2 \\ \quad \vdots \\ \mu_n = \mu_0(p/q)^n \end{array}$$

So there is a unique solution to REQ with $\mu_0 = 1$ namely $\mu_n = (p/q)^n$. If $p > q$ then $\sum_n \mu_n = \infty$ which shows there is no invariant probability distribution for this chain. If $p < q$ then $\sum_n \mu_n = 1/(1 - p/q) < \infty$ which gives the unique invariant probability distribution

$$\pi_n = \frac{(p/q)^n}{\sum_{n=0}^{\infty}(p/q)^n} = (p/q)^n(1 - p/q) \sim \text{Geometric}(1 - p/q) \text{ on } \{0, 1, 2, \dots\}$$

From previous theory, existence of this invariant probability distribution proves that the chain is positive recurrent for $p < q$.

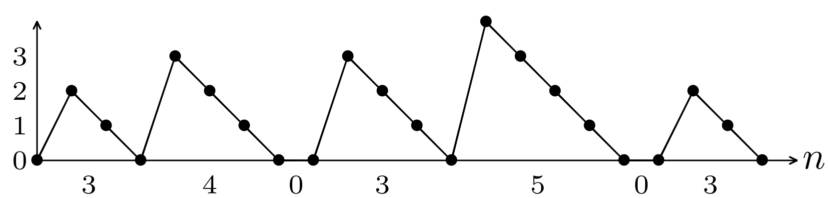
- Example: Compute $\mathbb{E}_1 T_0$ by two methods and compare.

  Method 1: First step analysis. Let $m_{ij} = \mathbb{E}_i T_j$. From 0,

$$m_{00} = 1 + p m_{10} \Longrightarrow m_{10} = \frac{m_{00} - 1}{p} = \frac{\frac{1}{1-p/q} - 1}{p} = \frac{1}{q - p}$$

  Method 2: We can use Wald's identity for the coin-tossing walk, without concern about the boundary condition at 0, to argue that $m_{10} = \frac{1}{q-p}$.

- Example: For any particular probability distribution $p_1, p_2, \dots$ on $\{1, 2, \dots\}$, there is a MC and a state 0 whose recurrent time T_0 has $\mathbb{P}_0(T_0 = n) = p_n$.



Start with $T_0^{(1)}, T_0^{(2)}, T_0^{(3)}, \ldots$ iid with distribution probability $(p_1, p_2, \ldots)$.

$$\begin{aligned}
P(0,j) &= p_{j+1}, \text{ for } j = 0,1,2,\ldots \\
P(1,j) &= \mathbf{1}(j=0) \\
P(2,j) &= \mathbf{1}(j=1) \\
&\vdots \\
P(i,j) &= \mathbf{1}(j=i-1), \text{ for } i > 0
\end{aligned}$$

<u>Describe the stationary distribution</u>: Solve $\pi P = \pi$. Not reversible. Use

$$\frac{\pi_j}{\pi_0} = \mathbb{E}_0(\text{ \# of visits to } j \text{ before } T_0 \text{ })$$

Notice chain dynamics: # of hits on state $j \geq 1$ is either 1 or 0. It is 1 if and only if initial jump from 0 is to j or greater. Probability of that is $p_{j+1} + p_{j+2} + \cdots$ $\Longrightarrow \dfrac{\pi_j}{\pi_0} = p_{j+1} + p_{j+2} + \cdots$.

Check $\dfrac{1}{\pi_0} = \mathbb{E}_0(T_0)$:

$$\frac{1}{\pi_0} = \sum_j \frac{\pi_j}{\pi_0} = \sum_{j=1}^{\infty} \mathbb{P}_0(T_0 > j) = \mathbb{E}_0(T_0)$$

Stat 150 Stochastic Processes **Spring 2009**

Lecture 21: Poisson Processes

Lecturer: Jim Pitman

- **Poisson processes**

  General idea: Want to model a random scatter of points in some space.

  - # of points is random
  - locations are random
  - want model to be simple, flexible and serve as a building block

  Variety of interpretations:

  - Natural phenomena such as weather, earthquakes, meteorite impacts
  - points represent discrete "events", "occurrences", "arrivals" in some continous space of possibilities
  - Continuous aspect: location/attributes of points in physical space/time
  - Discrete aspect: counts of numbers of points in various regions of space/time give values in $\longrightarrow \{0, 1, 2, 3, \dots\}$

  Abstractly: A general, abstract space S. Suitable subsets B of S for which we discuss counts, e.g. intervals, regions for which area can be defined, volumes in space, ...

  Generically $N(B)$ should be interpreted as the number of occurrence/arrivals/events with attributes in B. Then by definition, if $B_1, B_2, \dots, B_n$ are disjoint subsets of S, then

  $$N(B_1 \cup B_2 \cup \cdots \cup B_n) = N(B_1) + \cdots + N(B_n)$$

  Simple model assumptions that lead to Poisson processes:

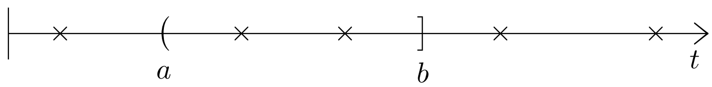
  (1) Independence: If $B_1, B_2, \dots, B_n$ are disjoint subsets of S, then the counts $N(B_i)$ are independent.

  (2) No multiple occurrences. Example: we model #'s of accidents, not #'s of people killed in accidents, or numbers of vehicles involved in accidents.

It can be shown in great generality that these assumptions imply the Poisson model: $N(B) \sim \text{Poisson}(\mu(B))$ for some measure μ on subsets B of S. If μ is λ times some notion of length/area/volume on S, then the process is called *homogeneous* and λ is called the rate. In general, $\mu(B)$ is the expected number of points in B. In the homogeneous case, the rate λ is the expected number of points per unit of length/area/volume.

- **Specific**

Most basic study is for a Poisson arrival process on a time line $[0, \infty)$ with arrival rate λ.



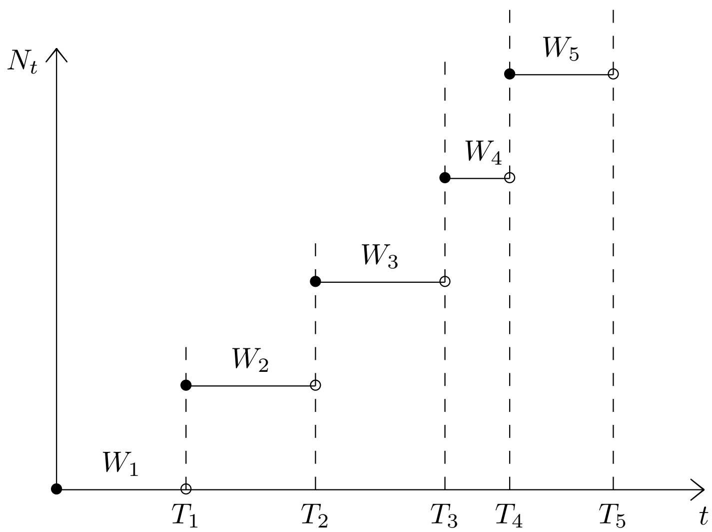
$$(a, b] = \text{ interval of time}$$
$$(a, b] \cup (b, c] = (a, c]$$
$$N(a, b] := \text{ \# of arrivals in } (a, b] \sim \text{ Poisson } \underbrace{(\lambda(\overbrace{b-a}^{\text{Length}}))}_{\mathbb{E}[\text{\# of arrivals}]}$$

$N_t := N(0, t]$. $(N_t, t \geq 0)$ continuous time counting process:

- only jumps by 1
- stationary independent increments
- for $0 \leq s \leq t$, $N_t - N_s \stackrel{d}{=} N_{t-s} \stackrel{d}{=} \text{ Poisson}(\lambda(t-s))$

Define

$$T_r := W_1 + W_2 + \cdots + W_r = \inf\{t : N_t = r\}, \qquad 0 = T_0 < T_1 < T_2 \cdots$$
$$W_i := T_{i+1} - T_i$$



- **Fact/Theorem**

  $(N_t, t \leq 0)$ is a Poisson process with rate $\lambda \iff W_1, W_2, W_3, \ldots$ is a sequence of iid $\text{Exp}(\lambda)$.

$$(W_1 > t) \iff (N_t = 0)$$
$$\mathbb{P}(W_1 > t) = \mathbb{P}(N_t = 0) = e^{-\lambda t}$$

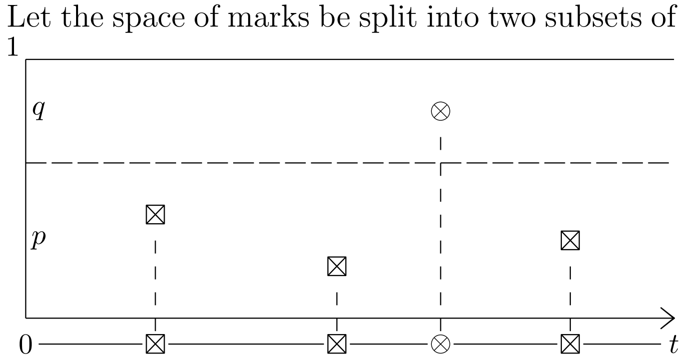
- **Fundamental constructions: Marked Poisson point process**

  Assume space of marks is [0,1] at first.

  ◦ Points arrive according to a PP(λ)

  ◦ Each point is assigned an independent uniform mark in [0,1]

  Let the space of marks be split into two subsets of length p and q with $p+q=1$.



Each point is a $\boxtimes$ with probability p and $\otimes$ with probability q. Let

$$N_{\boxtimes}(t): \ \# \text{ of } \boxtimes \text{ up to } t$$
$$N_{\otimes}(t): \ \# \text{ of } \otimes \text{ up to } t$$

Then

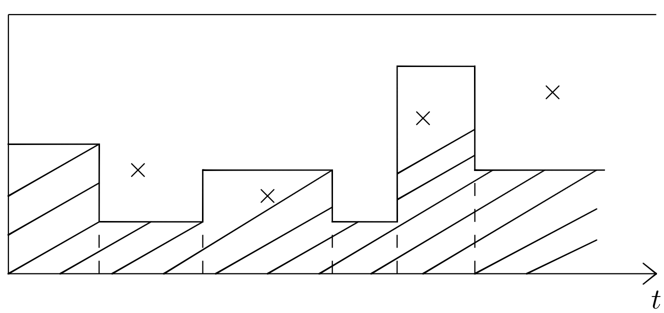
$$\begin{aligned} N_{\boxtimes}(t) &\sim \text{Poisson}(\lambda pt) \\ N_{\otimes}(t) &\sim \text{Poisson}(\lambda qt) \\ N(t) &= N_{\boxtimes}(t) + N_{\otimes}(t) \end{aligned}$$

These relations hold because of the *Poisson thinning property*: if you have Poisson number of individuals with mean μ, then

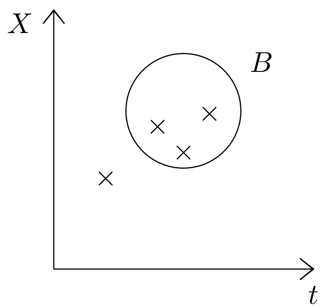
- keep each individual with probability p
- discard each individual with probability q

Then # kept and # discarded are independent Poissons with means μp and μq, respectively. Pushing this further, $(N_{\boxtimes}(t), t \geq 0)$ and $(N_{\otimes}(t), t \geq 0)$ are two independent homogeneous Poisson processes with rates λp and λq respectively.

Take a fixed region (t, u) space, count the number of (T_i, U_i) with $(T_i, U_i) \in R$. In the graph below, there are two. Since the sum of independent Poissons is Poisson, the number of (T_i, U_i) has to be Poisson as well.



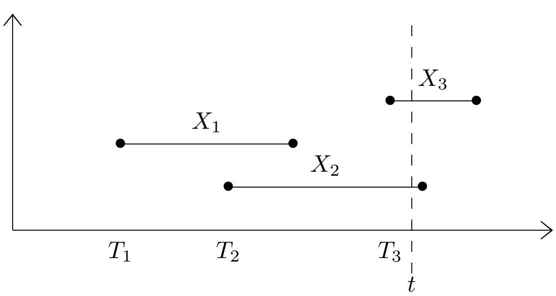
- ◦ Here we simply suppose the marks are uniform
- ◦ Generalize to marks which are independent identical with density $f(x)$
- ◦ $T_1, T_2, \dots$ are time arrivals of PP(λ)
- ◦ $X_1, X_2, \dots$ are independent identical with $\mathbb{P}(X_i \in dx) = f(x)dx$.
- ◦ Then $(T_1, X_1), (T_2, X_2), \dots$ are the points of a Poisson process in $[0, \infty) \times \mathbb{R}$ with measure density $\lambda dt f(x) dx$

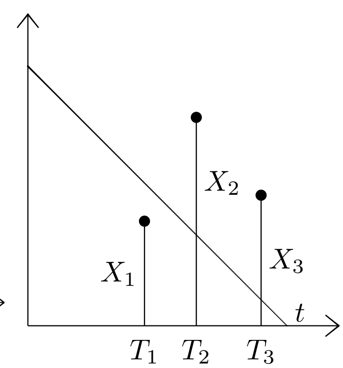


Here, $N_B \sim$ Poisson with mean $\iint_B \lambda dt f(x)dx$

- Practical example (Queueing model)
  - Customers arrive according to a PP(λ)
  - A customer arrives at time t and stays in the system for a random time interval with density $f(x)$.

  Find a formula for the distribution of # of customers in system at time t.





  Mark each arrival by time spent in the system.

$$\mu = \int_0^t \lambda ds \int_{t-s}^{\infty} f(x)dx$$
$$= \int_0^t \lambda ds(1 - F(s)), \qquad \text{where } F(s) = \int_0^s f(x)dx$$

  Actual number is Poisson with mean μ calculated above.

- Connection between PP and uniform order statistics

  Construction of a PP with rate λ on [0,1]

  Step 1: Let $N_1 \sim$ Poisson(λ). This will be the total number of points in [0,1]

Step 2: Given $N_1 = n$, let $U_1, \dots, U_n$ be independent and uniform on [0,1]. Let the points of PP be at $U_1, U_2, \dots, U_n$. Let

$$T_1 = \min_{1 \le i \le n} U_i$$
$$T_k = \min\{U_i : U_i > T_{k-1}, 1 \le i \le n\} \text{ for } k > 1$$

We can only define $T_1, T_2, \dots, T_n$ = ordered values of $U_1, U_2, \dots, U_n$. Pick all the T_r's with $T_r \le 1$ this way.

Create $(N_t, 0 \le t \le 1) : N_t = \#\{i : U_i \le t\} = \sum_{i=1}^{\infty} \mathbf{1}(N_1 \ge i, U_i \le t)$

Claim: $(N_t, 0 \le t \le 1)$ is the usual PP(λ)
This means

- $N_t \sim \text{Poisson}(\lambda t)$
- $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}, N_1 - N_{t_n}$ are independent for $0 \le t_1 \le \cdots \le t_n \le 1$
- $N_t - N_s \sim \text{Poisson}(\lambda(t-s))$

Why?
$N_t \sim \text{Poisson}(\lambda t)$ by thinning.
Independence also by thinning: I can assign Poisson(N_1) counts into categories $(0, t_1], (t_1, t_2], \dots, (t_n, 1]$ with probabilities $t_1, t_2 - t_1, \dots, 1 - t_n$.

Poissonization of multinomials:

$$\mathbb{P}(N_{t_1} = n_1, N_{t_2} - N_{t_1} = n_2, \dots, N_1 - N_{t_{k-1}} = n_k | n_1 + n_2 + \cdots + n_k = n)$$

$$= \mathbb{P}(N_1 = n_1 + \cdots + n_k)\mathbb{P}(N_{t_1} = n_1, N_{t_2} - N_{t_1} = n_2, \dots, N_1 - N_{t_{k-1}} = n_k | N_1 = n_1 + \cdots + n_k)$$

$$= e^{-\lambda} \frac{\lambda^{n_1 + \cdots + n_k}}{(n_1 + \cdots n_k)!} \binom{n_1 + \cdots + n_k}{n_1, n_2, \dots, n_k} t_1^{n_1} (t_2 - t_1)^{n_2} \cdots (1 - t_{k-1})^{n_k}$$

$$= e^{-\lambda t_1} \frac{(\lambda t_1)^{n_1}}{n_1!} \; e^{-\lambda(t_2 - t_1)} \frac{(\lambda(t_2 - t_1))^{n_2}}{n_2!} \; \cdots \; e^{-\lambda(1 - t_{k-1})} \frac{(\lambda(1 - t_{k-1}))^{n_k}}{n_k!}$$

Therefore $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}, N_1 - N_{t_n}$ are independent Poisson as claimed.

Stat 150 Stochastic Processes **Spring 2009**

Lecture 22: Continuous Time Markov Chains

Lecturer: Jim Pitman

- **Continuous Time Markov Chain**

  Idea: Want a model for a process $(X_t, t \geq 0)$ with continuous time parameter t.

  - State space a finite/countable set S
  - Markov property (time homogenous transition mechanism)

  Strong parallels with discrete time
  Expression of Markov: conditionally given that $X_t = j$, the past process $(X_s, 0 \leq s \leq t)$ and the future $(X_{t+v}, v \geq 0)$ are conditionally independent, and the future is a distinct copy of $(X_v, v \geq 0)$ given $X_0 = j$.

  Notations
  Recall in discrete $P^n = n$ step transition matrix, we need now $P^t =$ time t transition matrix.

  Conceptual challenge: raising a matrix to a real t power such as $t = 1/2, 1/3$.

  Use notation $P(t)$ instead of P^t for the transition matrix over time t. Idea: $P(t)$ is something like a tth power:

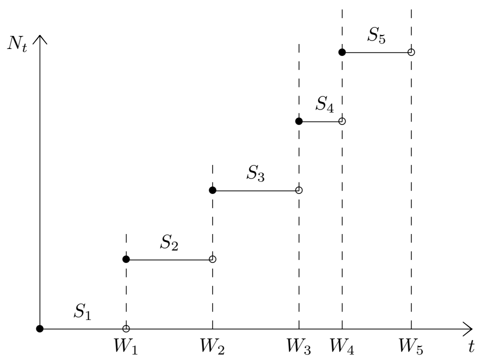
$$P(s)P(t) = P(s+t).$$

  This *semigroup property* of transition probability matrices follows easily from the Markov property:

$$\begin{aligned} P_{ij}(s+t) :=& \mathbb{P}_i(X_{s+t} = j) \\ =& \sum_k \mathbb{P}_i(X_s = k, X_{s+t} = j) \\ =& \sum_k \mathbb{P}_i(X_s = k)\mathbb{P}_k(X_t = j) \end{aligned}$$

- **Basic examples:** The Poisson process as a Markov chain.

  Let $(N_t, t \geq 0)$ be a Posson process with $N_0 = 0$ and rate λ.



$S_1, S_2, \dots$ independent $\exp(\lambda)$.

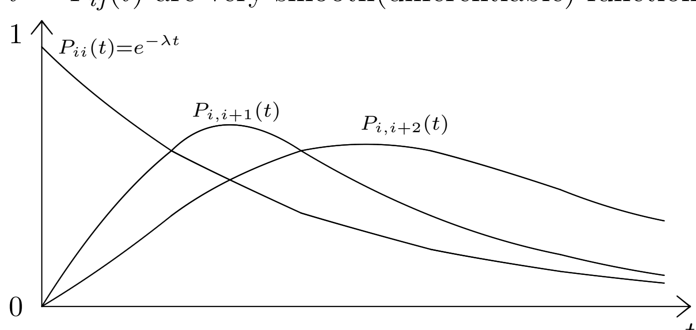
Obserbe $(N_t, t \geq 0)$ is a MC with transition matrices

$$\begin{aligned}
P_{ij}(t) &= \mathbb{P}_0(N_{s+t} = j | N_s = i) \\
&= \mathbb{P}_0(N_{s+t} - N_s = j - i | N_s = i) \\
&= \mathbb{P}_0(N_{s+t} - N_s = j - i) \qquad \text{since } N_{s+t} - N_s \perp N_s \\
&= \mathbb{P}_0(N_t = j - i) \\
&= e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} \mathbf{1}(j \geq i)
\end{aligned}$$

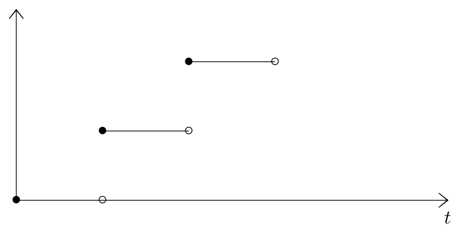
Independence property of PP $\Longrightarrow$ $(N_t, t \geq 0)$ is a MC in state space $\{0, 1, \dots\}$ with transition matrix $\mathbb{P}_{ij}(t) = e^{-\lambda t} \dfrac{(\lambda t)^{j-i}}{(j-i)!} \mathbf{1}(j \geq i)$.

Distinguish carefully two features of this process:

(1) The semi-group $(P(s+t) = P(s)P(t))$ of transition $P(t), t \geq 0$: the entries $t \to P_{ij}(t)$ are very smooth(differentiable) functions of t.



(2) The path of the process $t \to N_t$ is a step function.



Recall:

$$\begin{aligned} \text{Discrete time :} & P^{n+1} = P^n P = P P^n \\ \text{Continuous time :} & P(n+1) = P(n)P(1) = P(1)P(n) \\ & P(t+h) = P(t)P(h) = P(h)P(t) \qquad \text{with } h \text{ small} \end{aligned}$$

We should evaluate $\frac{d}{dt}P(t) := \lim_{t\to 0} \frac{P(t+h) - P(t)}{h}$. We write this as a convenient abbreviation for $\frac{d}{dt}P_{ij}(t) = \lim_{t\to 0} \frac{P_{ij}(t+h) - P_{ij}(t)}{h}, \forall i, j \in S$.

$$\begin{aligned} \frac{d}{dt}P(t) &= \lim_{t\to 0} \frac{P(t+h) - P(t)}{h} \\ &= \lim_{h\to 0} P(t)\frac{P(h) - I}{h} \\ &= P(t) \lim_{h\to 0} \frac{P(h) - I}{h} \\ &= P(t)A, \qquad \text{where } A := \lim_{h\to 0} \frac{P(h) - I}{h} = \frac{d}{dt}P(t)\bigg|_{t=0} \end{aligned}$$

Note that $\frac{d}{dt}P(t) = P(t)A = AP(t)$. These are **Kolmogorov forward and backward difference equations for a Markov semi-group**. The matrix A is called the *infinitesimal generator* of the semi-group. It should be clear that A is rather important. If you know A, you can figure out $P(t)$ by solving Kolmogorov's equations.

- Compute A for the PP:

$$\begin{aligned} A_{i,i} &= -\lambda \\ A_{i,i+1} &= \lambda \\ A_{i,i+2} &= 0 \\ A_{i,i+3} &= 0 \\ &\vdots \end{aligned}$$

$$A_{i,j} = \begin{cases} -\lambda & \text{if } j = i \\ \lambda & \text{if } j = i+1 \\ 0 & \text{else} \end{cases}$$

Note that

$$\sum_j P_{ij}(t) = 1 \Longrightarrow \frac{d}{dt}\sum_j P_{ij}(t) = 0 \Longrightarrow \sum_j \frac{d}{dt} P_{ij}(t) = 0$$

Take $t = 0$ in this, $\left.\frac{d}{dt}P_{ij}(t)\right|_{t=0} = A_{ij} \Longrightarrow \sum_j A_{ij} = 0.$

Notice $A_{ii} \le 0$ for all i, all transition matrix $P(t)$.

$$\left.\begin{matrix} P_{ii}(0) = 1 \\ P_{ii}(t) \le 1 \end{matrix}\right\} \Longrightarrow \left.\frac{d}{dt}P_{ii}(t)\right|_{t=0} \le 0$$

Customary notations: (see Pg. 396 of text)

$$\begin{aligned} q_i &:= -A_{ii} \\ q_{ij} &:= A_{ij} \ge 0 \qquad \text{for all } j \ne i \\ q_i &= \sum_{j \ne i} q_{ij} \end{aligned}$$

- **Key idea:** Start a continuous time chain in state i,
  - it will hold there for an exponentially distributed sojourn time with $\exp(q_i)$, mean $1/q_i$.
  - independent of length of this sojourn time, when it leaves for the first time, it jumps to j with probability q_{ij}/q_i.

$$A = \begin{bmatrix} -q_1 & q_{12} & q_{13} & \cdots \\ q_{21} & -q_2 & q_{23} & \cdots \\ \vdots & & -q_3 & \\ \vdots & & & \ddots \end{bmatrix}$$

Now sum $= 0$, diagonal $D_{ij} = -q_i \le 0$, off diagonal $d_{ij} = q_{ij} \ge 0$, $\sum_{j \ne i} q_{ij} = q_i$, $\sum_{j \ne i} q_{ij}/q_i = 1$.

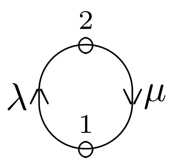
Example: A Poisson process that starts in state i

- hold in state i for $\exp(\lambda)$, $\lambda = q_i = -A_{ii}$
- only way to leave state i is by jumping to $i+1$, $q_{i,i+1}/q_i = \lambda/\lambda = 1$

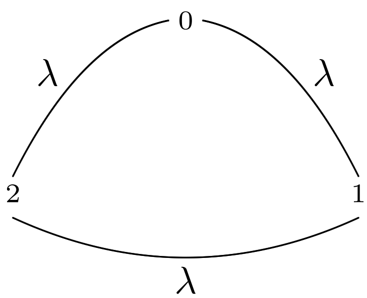
   – for other states j, $q_{ij}/q_i = 0/\lambda = 0$

- **Transition rate diagram:** Mark off diagonal entries of A.

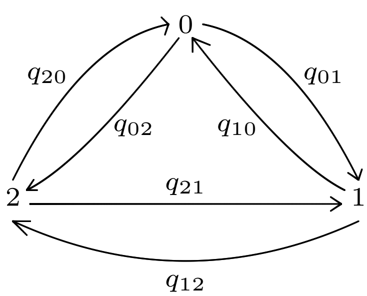
  Two state MC: $A = \begin{array}{c} 1 \\ 2 \end{array} \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$



  Three state MC on a circle: $N_t^{(3)} = N_t \bmod 3$, $N_t = \mathrm{PP}(\lambda)$.



  General 3 state chain:



Lecture 21 : Continuous Time Markov Chains

STAT 150 Spring 2006 *Lecturer: Jim Pitman* Scribe: *Stephen Bianchi* <>

(*These notes also include material from the subsequent guest lecture given by Ani Adhikari.*)

Consider a continuous time stochastic process $(X_t, t \geq 0)$ taking on values in the finite state space $S = \{0, 1, 2, \ldots, N\}$.

Recall that in discrete time, given a transition matrix $P = p(i,j)$ with $i, j \in S$, the n-step transition matrix is simply $P^n = p^n(i,j)$, and that the following relationship holds:

$$P^n P^m = P^{n+m}.$$

Where $p^n(i,j)$ is the probability that the process moves from state i to state j in n transitions. That is,

$$\begin{aligned}
p^n(i,j) &= \mathbb{P}_i(X_n = j), \\
p^n(i,j) &= \mathbb{P}(X_n = j | X_0 = i), \\
p^n(i,j) &= \mathbb{P}(X_{m+n} = j | X_m = i).
\end{aligned}$$

Now moving to continuous time, we say that the process $(X_t, t \geq 0)$ is a *continuous time markov chain* if the following properties hold for all $i, j \in S$, $t, s \geq 0$:

- $P_t(i,j) \geq 0$,
- $\sum_{j=0}^{N} P_t(i,j) = 1$,
- $\mathbb{P}(X_{t+s} = j | X_0 = i) = \sum_{k=0}^{N} \mathbb{P}(X_{t+s} = j | X_s = k)\mathbb{P}(X_s = k | X_0 = i)$, or $\mathbb{P}(X_{t+s} = j | X_0 = i) = \sum_{k=0}^{N} P_s(i,k) P_t(k,j)$.

This last property can be written in matrix form as

$$P_{s+t} = P_s P_t.$$

This is known as the *Chapman-Kolmogorov* equation (the semi-group property).

Example 21.1 (Poissonize a Discrete Time Markov Chain) *Consider a poisson process $(N_t, t \geq 0)$ with rate λ, and a jump chain with transition matrix $\hat{P}$, which* *makes jumps at times* $(N_t, t \geq 0)$*. Let* $X_t = Y_{N_t}$ *(*Y_{N_t} *is the value of the jump chain at time* N_t*), where* X_t *takes values in the discrete state space* S*. Assume* $(Y_0, Y_1, \ldots)$ *and* N_t *are independent. Find* P_t *for the process* X_t*.*

By definition

$$P_t(i,j) = \mathbb{P}(X_t = j | X_0 = i).$$

Condition on N_t *to give,*

$$\begin{aligned} P_t(i,j) &= \sum_{n=0}^{\infty} \mathbb{P}(N_t = n) \hat{P}^n(i,j) \\ &= \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \hat{P}^n(i,j) \\ &= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t \hat{P})^n(i,j)}{n!} \\ &= e^{-\lambda t} e^{-\lambda t \hat{P}}. \end{aligned}$$

Dropping indices then gives,

$$P_t = e^{-\lambda t(\hat{P} - I)}.$$

P_t *is a paradigm example of a continuous time semi-group of transition matrices. It is easy to check that* $P_t P_s = P_{t+s}$*, since* $e^{A+B} = e^A e^B$*.*

21.1 Generalization

Let A be a matrix of transition *rates*, where $q_{ij} = A(i,j)$ is the rate of transitions from state i to state j $(i \neq j)$, $q_{ij} \geq 0$, and $q_i = A(i,i)$ is the total rate of transitions out of state i. Hence, $q_i = -\sum_{j=0, j\neq i}^{N} q_{ij}$ and each row of A sums to 0. Also, since P_t is continuous and differentiable we have

$$\begin{aligned} q_{ij} &= \lim_{h \to 0} \frac{P_h(i,j)}{h} \\ q_i &= \lim_{h \to 0} \frac{1 - P_h(i,i)}{h}. \end{aligned}$$

In words, q_{ij} is the probability per unit time of transitioning from state i to state j in a small time increment (h).

The limit relations above can be expressed in matrix form,

$$\frac{d}{dt} P_t = \lim_{h \to 0+} \frac{P_{t+h} - P_t}{h} = \lim_{h \to 0+} \frac{P_t P_h - P_t}{h} = \lim_{h \to 0+} \frac{P_t [P_h - I]}{h}$$

hence

$$\frac{d}{dt}P_t = P_t \lim_{h\to 0+} \frac{P_h - I}{h} = P_t \frac{d}{dt}P_0 = P_t A = AP_t,$$

where $A = \frac{d}{dt}P_0$. Given the initial condition $P_0 = I$, the solution to the equation

$$\frac{d}{dt}P_t = P_t A = AP_t$$

is $P_t = e^{At}$. $\frac{d}{dt}P_t = P_t A$ is known as the *Kolmogorov forward equations* and $\frac{d}{dt}P_t = AP_t$ is known as the *Kolmogorov backward equations*. We claim that $(P_t, t \geq 0)$ is a collection of transition matrices of some continuous time markov chain.

In the example given above we have,

$$\begin{aligned} A &= \lambda(\hat{P} - I) \\ A(i,i) &= \lambda(\hat{P}(i,i) - 1) \\ A(i,j) &= \lambda \hat{P}(i,j). \end{aligned}$$

Example 21.2 *Consider the two state markov chain with states $\{0,1\}$ and*

$$A = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

The sojourn times in state 0 and 1 are independent and exponentially distributed with parameters λ and μ, respectively. Find P_t for this chain.

First note that

$$\frac{d}{dt}P_t = AP_t = \begin{pmatrix} -\lambda P_t(0,0) + \lambda P_t(1,0) & -\lambda P_t(0,1) + \lambda P_t(1,1) \\ \mu P_t(0,0) - \mu P_t(1,0) & \mu P_t(0,1) - \mu P_t(1,1) \end{pmatrix}.$$

Then

$$\frac{d}{dt}(P_t(0,0) - P_t(1,0)) = -(\lambda + \mu)(P_t(0,0) - P_t(1,0)).$$

This equation has the form $f'(t) = cf(t)$, and a solution is

$$P_t(0,0) - P_t(1,0) = ce^{-(\lambda+\mu)t}.$$

Since c evaluated at $t = 0$ is one (i.e., $P_0(0,0) - P_0(1,0) = 1$)), we have

$$P_t(0,0) - P_t(1,0) = e^{-(\lambda+\mu)t},$$

which implies

$$\frac{d}{dt}P_t(0,0) = -\lambda e^{-(\lambda+\mu)t}.$$

Then

$$\begin{aligned} P_t(0,0) &= P_0(0,0) + \int_0^t -\lambda e^{-(\lambda+\mu)s} ds \\ P_t(0,0) &= 1 + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} - \frac{\lambda}{\lambda+\mu} \\ P_t(0,0) &= \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t}. \end{aligned}$$

Likewise

$$\begin{aligned} P_t(1,0) &= \frac{\mu}{\lambda+\mu} - \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t} \\ P_t(0,1) &= \frac{\mu}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} \\ P_t(1,1) &= \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t}. \end{aligned}$$

Notice that as $t \longrightarrow \infty$

$$\begin{aligned} P_t(0,0) &= P_t(1,0) = \frac{\mu}{\lambda+\mu} \\ P_t(0,1) &= P_t(1,1) = \frac{\lambda}{\lambda+\mu}. \end{aligned}$$

These limits are independent of the initial state. The stationary distribution of the chain is

$$\begin{aligned} \pi_0 &= \frac{\mu}{\lambda+\mu} \\ \pi_1 &= \frac{\lambda}{\lambda+\mu}. \end{aligned}$$

21.2 Jump Hold Description

Let $(X_t, t \geq 0)$ be a markov chain with transition rate matrix A, such that $A(i,j) \geq 0$ for $i \neq j$ and $A(i,i) = -\sum_{i \neq j} A(i,j)$. Let H_i be the holding time in state i (i.e., the process "holds" in state i for an amount of time H_i and then jumps to another state). The markov property implies

$$\mathbb{P}_i(H_i > s + t | H_i > s) = \mathbb{P}_i(H_i > t).$$

We also recognize this as the memoryless property of the exponential distribution (fact: the only continuous distribution with the memoryless property is the exponential).

By the idea of *competing risks* we will see that $\mathbb{P}_i(H_i > t) = e^{A(i,i)t}$. Imagine that each state $j \neq i$ has a random variable $\xi_j \sim \exp(A(i,i))$. Think of a bell that rings in state j at time ξ_j, and from state i the process moves to the state where the bell rings first. Construct $H_i = \min_{i \neq j} \xi_j$, then

$$\begin{aligned}
\mathbb{P}_i(H_i > t) &= \mathbb{P}(\bigcap_{i \neq j} \xi_j > t) \\
&= \Pi_{i \neq j} e^{-A(i,j)t} \\
&= e^{-(\sum_{i \neq j} A(i,j))t} \\
&= e^{A(i,i)t},
\end{aligned}$$

where $A(i,i) < 0$. What is the probability is that the first transition out of state i is to state k? This is given by

$$\begin{aligned}
\mathbb{P}(\min_{i \neq j} \xi_j = \xi_k) &= \int_0^\infty \mathbb{P}(\min \in dt,\ \min = \xi_k) dt \\
&= \int_0^\infty \mathbb{P}(\xi_k \in dt, \text{ all the others are } > t) dt \\
&= \int_0^\infty A(i,k) e^{-A(i,k)t} e^{-(\sum_{i \neq k,j} A(i,k))t} dt \\
&= \int_0^\infty A(i,k) e^{-(\sum_{i \neq j} A(i,j))t} dt \\
&= \frac{A(i,k)}{\sum_{i \neq j} A(i,j)} = \frac{A(i,k)}{-A(i,i)}.
\end{aligned}$$

Observe from this calculation that the following mechanisms are equivalent:

1. (a) competing random variables $\xi_j \sim \exp(A(i,j))$ for state $j \neq i$
   (b) $H_i = \min_{j \neq i} \xi_j$
   (c) jump to the state j which attains the minimum
2. (a) $H_i \sim \exp(-A(i,i) = \sum_{i \neq j} A(i,j))$
   (b) at time H_i jump to k with probability $\frac{A(i,k)}{-A(i,i)}$

Example 21.3 (Hold Rates for a Poissonized Chain) *Consider a general transition matrix $\hat{P}$, where some of the diagonal entries maybe greater than zero. So we have*

$$\begin{aligned}
P_t &= e^{-\lambda t(\hat{P} - I)} \\
A &= \lambda(\hat{P} - I) \\
A(i,i) &= \lambda(\hat{P}(i,i) - 1) = -\lambda(1 - \hat{P}(i,i)) \\
A(i,j) &= \lambda(\hat{P}(i,j)).
\end{aligned}$$

Hence $A(i,i) = \lambda(1 - \hat{P}(i,i)) = \lambda \sum_{i \neq j} \hat{P}(i,j) = \sum_{i \neq j} A(i,j)$ *is the total rate of transitions out of state* i*, and* $A(i,j) = \lambda(\hat{P}(i,j))$ *is the rate of* i *to* j *transitions. Let*

$$H_i = S_1 + S_2 + \cdots + S_G,$$

where $S_1, S_2, \ldots \sim \exp(\lambda)$ *are independent. Define* G *to be the number of steps of the* $\hat{P}$ *chain until the process leaves state* i*. Then*

$$\mathbb{P}(G = n) = \hat{P}(i,i)^{n-1}(1 - \hat{P}(i,i)).$$

So G *has a geometric distribution with parameter* $1 - \hat{P}(i,i)$ *and* H_i *has an exponential distribution with parameter* $\lambda(1 - \hat{P}(i,i))$*. It is easy to check the expectations, since* $\mathbb{E}[S_i] = 1/\lambda$, $\mathbb{E}[G] = 1/(1 - \hat{P}(i,i))$*, and* $\mathbb{E}[H_i] = 1/\lambda(1 - \hat{P}(i,i))$*. Then by Wald's identity*

$$\mathbb{E}[H_i] = \mathbb{E}[S_i]\mathbb{E}[G] = \frac{1}{\lambda}\left(\frac{1}{1 - \hat{P}(i,i)}\right) = \frac{1}{\lambda(1 - \hat{P}(i,i))}.$$

21.3 Theorem

Theorem 21.4 (for finite state irreducible chain) *There is a unique probability distribution* π *so that*

1. $\lim_{t \to \infty} P_t(i,j) = \pi_j$ *for all* i*.*
2. π *solves the "balance equations"* $\pi A = 0$ *(equivalently,* $\pi P_t = \pi$ *for all* t*).*
3. π_j *is the long-run proportion of time spent in state* j
4. *the expected return time is*
$$\mathbb{E}_j[T_j] = \frac{1}{q_j \pi_j}.$$

Proof: [for part (2)] Assume (1) is true.

$$P_{s+t}(i,j) = \sum_k P_s(i,k) P_t(k,j)$$

Letting $s \to \infty$ we get

$$\pi_j = \sum_k \pi_k P_t(k,j)$$

Therefore, $\pi = \pi P_t$ for all t. Differentiating with respect to t then gives

$$0 = \pi A.$$

■

Look at the jth element of πA (which we get by multiplying π by the jth column of A). We see that

$$\sum_{i \neq j} \pi_i q_{ij} - \pi_j q_j = 0$$

which gives

$$\sum_{i \neq j} \pi_i q_{ij} = \pi_j q_j.$$

This is the reason these equations are referred to as *balance equations*. Computations can often be simplified if there is also *detailed balance*, which simply means that for all pairs i, j ($i \neq j$) we have

$$\pi_i q_{ij} = \pi_j q_{ji}.$$

This is the condition of *reversibility* of the markov chain.

Example 21.5 (Birth-Death Process) *Consider a process that, from state i, can only move to state $i+1$ or state $i-1$. That is*

$$\begin{aligned} q_{i,i+1} &= \lambda_i \\ q_{i,i-1} &= \mu_i \\ q_{i,j} &= 0, \forall j \notin \{i+1, i-1\}. \end{aligned}$$

So λ_i is the birth rate and μ_i is the death rate. We try to find π so that

$$\begin{aligned} \pi_0 q_{0,1} &= \pi_1 q_{1,0} \\ \pi_1 q_{1,2} &= \pi_2 q_{2,1} \\ &\vdots \\ \pi_{i-1} q_{i-1,i} &= \pi_i q_{i,i-1}. \end{aligned}$$

Then

$$\begin{aligned} \pi_0 q_{0,1} &= \pi_1 q_{1,0} \\ \pi_0 \lambda_0 &= \pi_1 \mu_1 \\ \pi_1 &= \frac{\lambda_0}{\mu_1} \pi_0. \end{aligned}$$

And

$$\begin{aligned} \pi_1 q_{1,2} &= \pi_2 q_{2,1} \\ \pi_1 \lambda_1 &= \pi_2 \mu_2 \\ \pi_2 &= \frac{\lambda_1}{\mu_2} \pi_1 = \frac{\lambda_1 \lambda_0}{\mu_1 \mu_2} \pi_0. \end{aligned}$$

Hence

$$\pi_k = \frac{\lambda_0\lambda_1\cdots\lambda_{k-1}}{\mu_1\mu_2\cdots\mu_k}\pi_0,$$

and

$$\pi_0 = \frac{1}{1+\sum_{k=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{k-1}}{\mu_1\mu_2\cdots\mu_k}}.$$

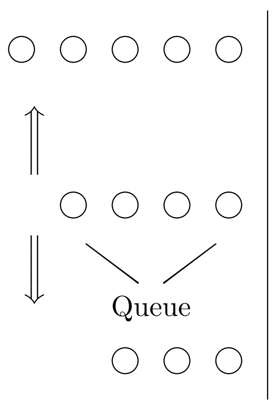
Stat 150 Stochastic Processes **Spring 2009**

Lecture 24: Queuing Models

Lecturer: Jim Pitman

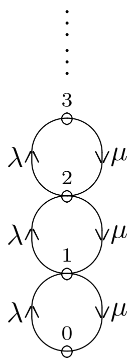
- **Queueing Models**

  The simplest setup is a single server with some service rules takes a random to process then the customer departs.



  Generically, $X(t) = \#$ of customers in system(queue) at time t, and model for evolution of X.

- **M/M/1**
  - M = Markov
  - First M is for input stream
  - Second M is for service process
  - $1 = \#$ of servers
  - Poisson arrival at rate λ
  - Independent exponential service time with rate μ
  - Each customer that is served takes $\exp(\mu)$ time
  - The service times $T_1, T_2, \ldots$ are i.i.d. $\exp(\mu)$ random variables independent of the arrival stream



$$
\begin{aligned}
q_{ij} &= \text{rate of transition from } i \text{ to } j \\
&= \text{off diagonal elements of the } A \text{ matrix} \\
q_0 &= \lambda = -A_{00} \\
q_{01} &= \lambda \\
q_{0j} &= 0, \qquad \text{for all } j \neq 0, j \neq 1 \\
q_1 &= \lambda + \mu = A_{11} \\
q_{10} &= \mu \\
q_{12} &= \lambda \\
q_{1j} &= 0, \qquad \text{for all } j \neq 0, j \neq 1, j \neq 2
\end{aligned}
$$

- Condition for stability of the queue : $\lambda < \mu$
  $\lambda = \mu$ is exceptional (null-recurrent chain).
  $\lambda > \mu$ is linear growth (like a RW with upward drift: transient chain).
  Stability means $\lim_{t\to\infty} \mathbb{P}(X_t = j) = \pi_j > 0$ with $\sum_j \pi_j = 1$.
  We can calculate π. How?

  $$\pi P_t = \pi, \qquad \text{for all } t$$

  Take $\frac{d}{dt}$ and evaluate at $0 \Longrightarrow \pi A = 0$. (Note: we did not try to solve differential equations for P_t. In the M/M/1 queue $P_t(i,j)$ is a nasty power series (involving Bessel functions))

  $$\text{Use} \qquad q_i = -A_{ii}, \quad q_{ij} = A_{ij}, i \neq j, \quad \text{and} \quad \sum_i \pi_i A_{ij} = 0, \text{ for all } j$$

  We get

  $$
  \begin{aligned}
  \forall j, \qquad & \sum_{i \neq j} \mu_i q_{ij} - \pi_j q_j = 0 \\
  \Longrightarrow \forall j, \qquad & \sum_{i \neq j} \mu_i q_{ij} = \pi_j q_j
  \end{aligned}
  $$

That is , at equilibrium,

$$\text{Rate in to } j = \text{ Rate out of } j$$

Fact: For birth/death chain on integers, the only possible equilibrium is reversible. So always look for solution of Rev EQ.

$$\begin{bmatrix} \pi_0\lambda & = & \pi_1\mu \\ \pi_1\lambda & = & \pi_2\mu \\ & \vdots & \end{bmatrix} \Longrightarrow \frac{\pi_n}{\pi_{n-1}} = \frac{\lambda}{\mu} \Longrightarrow \pi_n = \pi_0(\frac{\lambda}{\mu})^n$$

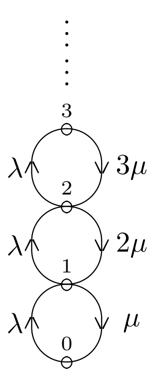
Use $\sum_{n=0}^{\infty} \pi_n = 1$, get

$$\pi_n = (\frac{\lambda}{\mu})^n(1-\frac{\lambda}{\mu}), \qquad n = 0, 1, 2, \dots$$

Check P549(2.9)

In the limit, mean queue length $= 1/(1-\frac{\lambda}{\mu}) = \frac{\mu}{\mu-\lambda}$. If λ is close to μ, this is large. Fix λ, as $\mu \downarrow \lambda$, $\frac{\mu}{\mu-\lambda} \uparrow \infty$.

- **M/M/∞**
  - Input rate λ
  - Unlimited number of servers, each operate at rate μ on one customer at a time



Use the fact that mean of $\exp(\mu_1), \exp(\mu_2), \dots, \exp(\mu_n)$ is $\exp(\mu_1+\mu_2+\cdots+\mu_n)$. Condition for stability? No matter what $\lambda > 0, \mu > 0$, the system is stable.To see this, establish a Rev EQ probablilty distribution:

$$\begin{bmatrix} \pi_0\lambda & = & \pi_1\mu \\ \pi_1\lambda & = & \pi_2 2\mu \\ \pi_2\lambda & = & \pi_3 3\mu \\ & \vdots & \end{bmatrix} \Longrightarrow \frac{\pi_n}{\pi_{n-1}} = \frac{\lambda}{n\mu} \Longrightarrow \pi_n = \pi_0(\frac{\lambda}{\mu})^n \frac{1}{n!}$$

So $\pi_0 = e^{-\lambda/\mu}$. The equilibrium probability distribution is Poisson(λ/μ) and the mean queue length is λ/μ.

Stat 150 Stochastic Processes **Spring 2009**

Lecture 25: Renewal Theory

Lecturer: Jim Pitman

- **Renewal Process**
  - Model for points on the real line.
  - Typically interpreted as times that something happens, e.g.
    * arrival in a queue
    * departure in a queue
    * end of a busy cycle of a queue
    * time that a component is replaced

- Many examples arise from more complex models where there is an increasing sequence of random times of interest, say $0 < W_1 < W_2 < \cdots$, with $W_n = X_1 + X_2 + \cdots + X_n$, where
  - the X_i are strictly positive random variables (perhaps discrete, perhaps with a density)
  - the X_i are independent and identically distributed copies of $X := X_1$.

  Language: Events/Arrivals/... are called *Renewals*.

  Generic image: Lightbulb replacements:

  - X_n = lifetime of n^{th} bulb
  - Install new bulb at time 0
  - Leave it until it burns out; replace with a new bulb immediately
  - W_n = time of replacement of the nth bulb.

  Notation ready to roll from PP setup. Introduce the counting process:

$$\begin{aligned} N(t) &:= \# \text{ of renewals up to and including time } t \\ &:= \text{largest } n : 0 < W_n \leq t \end{aligned}$$

  Note $N(0) := 0$. The path $t \mapsto N(t)$ is by construction a right continuous non-decreasing step function, which jumps up by 1 at each of the times W_n for

$n \geq 1$. The sequence of renewal times (W_n) and the counting process $N(t)$ are inverses of each other. So for instance:

$$(N(t) < n) = (W_n > t)$$

$$(N(t) \geq n) = (W_n \leq t)$$

which allows the distribution of $N(t)$ for each t to be derived from knowledge of the distribution of W_n for each n, and vice versa. In principle this generalizes the relation between Poisson counts and exponential inter-arrival times. In practice, these distributions are not computationally tractable except in a small number of cases. But limiting results are available in great generality.

- **Scope of theory**

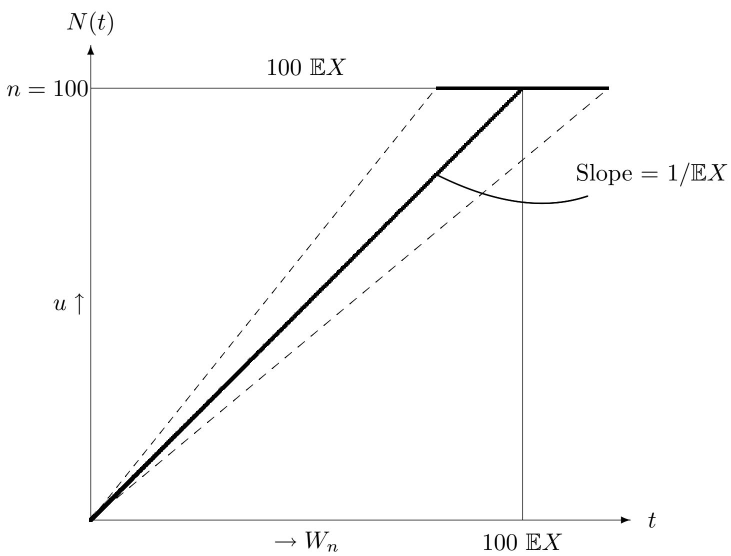
  Interested in limit behaviours of various quantities/random variables.

  $M(t) := \mathbb{E}[N(t)]$. How does $M(t)$ behave for large t?

  Example: If $\mathbb{P}(X > t) = e^{-\lambda t}$ for $\lambda > 0$, then $N(t) \sim \text{Poisson}(\lambda t)$. $M(t) = \mathbb{E}[N(t)] = \lambda t$, linear growth with t. Notice that $\lambda =$ expected rate of renewals per unit time, which is realized also as a long run rate of arrivals per unit time. Recall $E[X] = 1/\lambda$.

  $$\text{one renewal per time } 1/\lambda \longleftrightarrow \text{ expect } \lambda \text{ renewals per unit time}$$

  In general, we have **LLN**: After a large number n of replacements, the time used up will be around $\mathbb{E}(W_n) = n\mathbb{E}(X)$.



  By LLN, $\mathbb{E}N(t) \sim \dfrac{t}{\mathbb{E}X}$, where $g(t) \sim f(t)$ means $\lim_{t \to 0} \dfrac{f(t)}{g(t)} = 1$. This is called *asymptotic equivalence.*

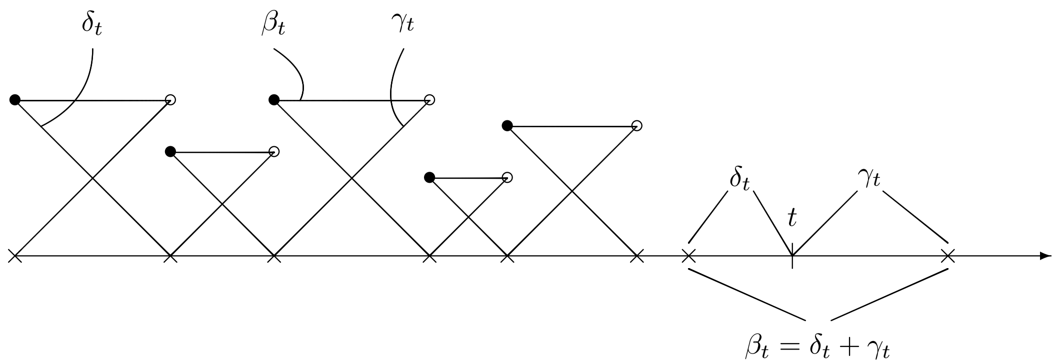
Theorem: If $\mathbb{E}X < \infty$, then $\lim_{t\to\infty} \frac{\mathbb{E}N(t)}{t/\mathbb{E}X} = 1$.

Usual notation: $\mathbb{E}N(t) \sim t/\mathbb{E}X$.

$\mathbb{E}[N(t)/t] \to 1/\mathbb{E}X$. This form is also true if $\mathbb{E}X = \infty$ with $1/\infty = 0$

Finer result: If you avoid discrete(lattice) case, meaning there is some $\Delta > 0$: $\mathbb{P}($ X is a multiple of $\Delta) = 1$ (Discrete renewal theory), then $M(t) := \mathbb{E}N(t)$ satisfies $M(t+h) - M(t) \to h/\mathbb{E}X$, for each fixed $h > 0$. This is D. Blackwell's renewal theorem.

- Now look at residual life and age process.
  - residual life/excess life = γ_t
  - age/current life = δ_t
  - $\beta_t := \gamma_t + \delta_t$ = total life of component in use at time t



- Example: Poisson process
  - Excess life $\gamma_t \sim exp(\lambda)$ for every t by the memoryless property of exponential variables.
  - Age δ_t. Notice $0 \le \delta_t \le t$, also $(\delta_t > s) \equiv (\underbrace{N(t) - N(t-s)}_{\text{Poisson}(\lambda s)} = 0)$, provided $0 < s < t$.

    So $\mathbb{P}(\delta_t > s) = e^{-\lambda s}$ for $0 \le s < t$, and $\mathbb{P}(\delta_t > t) = 0$.

    Clever representation:

$$\delta_t \overset{d}{=} \min\,(X, t), \text{where } X \sim \exp(\lambda)$$
$$\xrightarrow{d} X \sim \exp(\lambda) \text{ as } t \to \infty$$

Notice From Poisson assumptions, γ_t and δ_t are independent.

$$\gamma_t \overset{d}{=} X, \delta_t \xrightarrow{d} X \qquad \text{and} \qquad \gamma_t, \delta_t \text{ independent}$$
$$\Longrightarrow \beta_t := \gamma_t + \delta_t \xrightarrow{d} X_1 + X_2 = W_2$$

where we know $W_2 \sim \text{Gamma}(2, \lambda)$ has the $\text{Gamma}(2, \lambda)$ density

$$f_{W_2}(x) = \frac{1}{\Gamma(2)}\lambda^2 x e^{-\lambda x} = \frac{x f_X(x)}{E(X)}$$

- Remarkably, this structure of size-biasing the lifetime density $f_X(x)$ be a factor of x is completely general.

  <u>Theorem</u>: Whatever the distribution of X with a density $f_X(x)$, the limit distribution of β_t as $t \to \infty$ has density $\dfrac{x f_X(x)}{\mathbb{E}X}$. This is a density because:

$$\int_0^\infty \frac{x f_X(x)}{\mathbb{E}X} = \frac{\mathbb{E}X}{\mathbb{E}X} = 1$$

  <u>More</u>

$$(\gamma_t, \delta_t) \xrightarrow{d} (UY, (1-U)Y)$$

  where $\mathbb{P}(Y \in dx) = \dfrac{x f_X(x)}{\mathbb{E}X}$ and U is uniform [0,1], independent of Y.

  <u>Exercise</u>: Show this implies limit density of γ_t at x is $\frac{\mathbb{P}(X>x)}{\mathbb{E}X}$, and that the same is true for δ_t. Note that this function of x is a probability density because

$$\int_0^\infty \mathbb{P}(X > x)dx = \int_0^\infty \mathbb{E}[1(X > x)]dx = \mathbb{E}\int_0^\infty 1(X > x)dx = \mathbb{E}X.$$

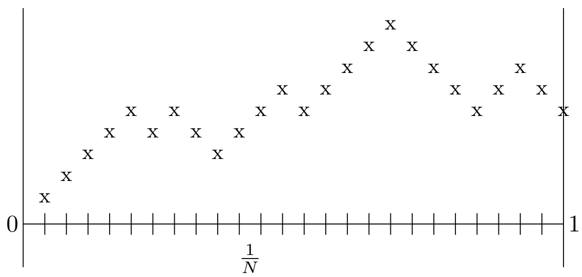
Stat 150 Stochastic Processes **Spring 2009**

Lecture 26: Brownian Motion

Lecturer: Jim Pitman

- **Brownian Motion**

  Idea: Some kind of continuous limit of random walk.



Consider a random walk with independent $\pm\Delta$ steps for some Δ with $\frac{1}{2} \uparrow \frac{1}{2} \downarrow$, and N steps per unit time. Select $\Delta = \Delta_N$, so the value of the walk at time 1 has a limit distribution as $N \to \infty$.

Background: Coin tossing walk with independent ± 1 steps with $\frac{1}{2} \uparrow \frac{1}{2} \downarrow$.

$$S_N = X_1 + \cdots + X_N$$
$$\mathbb{E}S_N = 0$$
$$\mathbb{E}S_N^2 = N$$

If instead we add $\pm\Delta_N$, get $S_N \Delta_N$ and

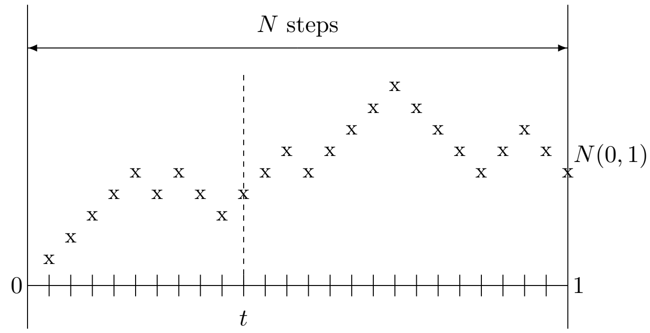
$$\mathbb{E}(S_N \Delta_N)^2 = \mathbb{E}(S_N^2)\Delta_N^2 = N\Delta_N^2 \equiv 1$$

if we take $\Delta_N = 1/\sqrt{N}$.

Value of our process at time 1 with scaling is then

$$S_N/\sqrt{N} \xrightarrow{d} \text{Normal}(0,1) \qquad \text{as } N \to \infty$$

according the the normal approximation to the binomial distribution. The same is true for any distribution of X_i with mean 0 and variance 1, by the Central



Limit Theorem.

Scaled process at time t has value

$$\frac{S_{\lfloor tN \rfloor}}{\sqrt{N}} = \frac{S_{\lfloor tN \rfloor}}{\sqrt{\lfloor tN \rfloor}} \sqrt{\frac{\lfloor tN \rfloor}{N}}$$

Fix $0 < t < 1$, $N \to \infty$,

$$\frac{S_{\lfloor tN \rfloor}}{\sqrt{\lfloor tN \rfloor}} \to N(0,1) \quad \text{and} \quad \sqrt{\frac{\lfloor tN \rfloor}{N}} \to \sqrt{t}$$

So

$$\frac{S_{\lfloor tN \rfloor}}{\sqrt{N}} \xrightarrow{d} \text{Normal}(0,t)$$

the normal distribution with mean 0, variance t, and standard deviation $\sqrt{t}$. In the distributional limit as $N \to \infty$, we pick up a process $(B_t, t \geq 0)$ with some nice properties:

- $B_t \sim \text{Normal}(0,t)$, $\mathbb{E}B_t = 0$, $\mathbb{E}B_t^2 = t$.
- For $0 < s < t$, B_s and $B_t - B_s$ are independent.

$$\begin{aligned} B_t &= B_s + (B_t - B_s) \\ \implies B_t - B_s &\sim \text{Normal}(0, t-s) \\ (B_t | B_s) &\sim \text{Normal}(B_s, t-s) \end{aligned}$$

For times $0 < t_1 < t_2 < \cdots < t_n$, $B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent Normal$(0, t_i - t_{i-1})$ random variables for $1 \leq i \leq n, t_0 = 0$. This defines the finite dimensional distributions of a stochastic process $(B_t, t \geq 0)$ called a *Standard Brownian Motion* or a *Wiener Process*.

- **Theorem:** (Norbert Wiener) It is possible to construct such a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is a countably additive probability measure, and the path $t \to B_t(w)$ is continuous for all $w \in \Omega$.

  More formally, we can take $\Omega = \mathbb{C}[0,\infty)$ = continuous functions from $[0,\infty) \to \mathbb{R}$. Then for a function (also called a *path*) $w = (w(t), t \geq 0) \in \Omega$, the value of

$B_t(w)$ is simply $w(t)$. This is the *canonical path space* viewpoint. This way $\mathbb{P}$ is a probability measure on the space Ω of continuous functions. This $\mathbb{P}$ is called *Wiener measure*. Note. It is customary to use the term Brownian motion only if the paths of B are continuous with probability one.

- Context: Brownian motion is an example of a **Gaussian process** (Gaussian = Normal). For an arbitrary index set I, a stochastic process $(X_t, t \in I)$ is called Gaussian if
  - for every selection of $t_1, t_2, \ldots, t_n \in I$, the joint distribution of $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ is multivariate normal.
  - Here multivariate normal
    $\iff \sum_i a_i X_{t_i}$ has a one-dimensional normal distribution for whatever choice of $a_1, \ldots, a_n$.
    $\iff X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ can be constructed as n (typically different) linear combinations of n independent normal variables.

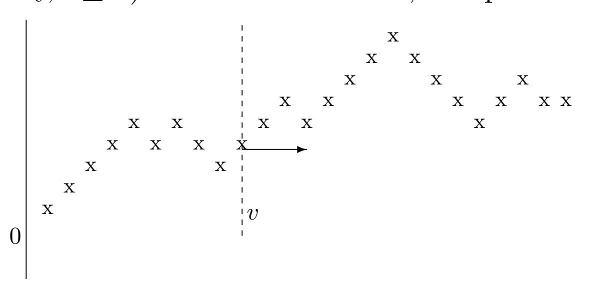
Easy fact: The FDD's of a Gaussian process are completely determined by its mean and covariance functions

$$\mu(t) := \mathbb{E}(X_t) \qquad \sigma(s,t) = \mathbb{E}(X_s - \mu(s))(X_t - \mu(t))$$

Here the function σ is *non-negative definite* meaning it satisfies the system of inequalities implied by $Var(\sum_i a_i X_{t_i}) \geq 0$ whatever the choice of real numbers a_i and indices t_i.

Key consequences: If we have a process $\hat{B}$ with the same mean and covariance function as a BM B, then the finite-dimensional distributions of $\hat{B}$ and B are identical. If the paths of $\hat{B}$ are continuous, then $\hat{B}$ a BM.

Example: Start with $(B_t, t \geq 0) =$ BM, fix a number $v \geq 0$, look at $(B_{v+t} - B_v, t \geq 0)$. This is a new BM, independent of $(B_s, 0 \leq s < v)$.



- Time inversion

  Let $\hat{B}_t := \begin{cases} tB(1/t) & \text{for } t > 0 \\ 0 & \text{for } t = 0 \end{cases}$. Check that $\hat{B}$ is a BM. Note first that $\hat{B}$ is also a Gaussian process.

(1) Continuity of paths
Away from $t = 0$, this is clear. At $t = 0$, does $tB(1/t) \to 0$ as $t \to 0$? Compute $Var(tB(1/t)) = t^2 Var(B(1/t)) = t \to 0$. With some more care, it is possible to establish path continuity at 0 (convergence with probability one).

(2) Mean and covariances

$$\mathbb{E}(tB(1/t)) = t\mathbb{E}(B(1/t)) = 0 = \mathbb{E}(B_t)$$

Covariances: for $0 < s < t$,

$$\begin{aligned}\mathbb{E}(B_s B_t) &= \mathbb{E}(B_s(B_s + B_t - B_s)) \\ &= \mathbb{E}(B_s^2) + \mathbb{E}(B_s(B_t - B_s)) \\ &= s + 0 \\ &= s\end{aligned}$$

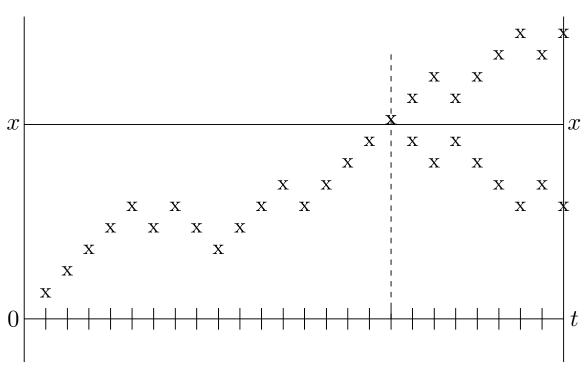
$$\begin{aligned}\mathbb{E}(\hat{B}_s \hat{B}_t) &= \mathbb{E}(sB(1/s)tB(1/t)) \\ &= st\mathbb{E}(B(1/s)B(1/t)) \\ &= st \cdot \frac{1}{t} \qquad \text{since } \frac{1}{t} < \frac{1}{s} \\ &= s\end{aligned}$$

- Example:

  Distribution of maximum on $[0, t]$, $M_t := \sup_{0 \le s \le t} B_s$

  Fact: $M_t \overset{d}{=} |B_t|$

$$\mathbb{P}(M_t) = 2\mathbb{P}(B_t > x)$$
$$\mathbb{P}(M_t \in dx) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$



Sketch a proof: We have $B \stackrel{d}{=} -B$.

$$\mathbb{P}(M_t > x, B_t < x) = \mathbb{P}(M_t > x, B_t > x) = \mathbb{P}(B_t > x)$$

by a reflection argument. This uses $B \stackrel{d}{=} -B$ applied to $(B(T_x + v) - x, v \geq 0)$ instead of B (strong Markov property) where T_x is the first hitting time of x by B.. Add and use $\mathbb{P}(B_t = x) = 0$, to get $\mathbb{P}(M_t > x) = 2\mathbb{P}(B_t > x)$.

Stat 150 Stochastic Processes **Spring 2009**

Lecture 27: Hitting Probabilities for Brownian Motion

Lecturer: Jim Pitman

- The French mathematician Paul Lévy first showed that several random variables have the same distribution for a Brownian path:

  (1) $g_1 :=$ Time of last zero before 1 (See Problem 2.5 on Page 497 for $g_1 = \tau_0$ in text)

  (2) $A^+(1) = \int_0^1 \mathbf{1}(B_t > 0)dt$, function of time spent > 0 up to time 1

  (3) Time when the maximum of B on [0,1] is attained

  (4) Time when the minimum of B on [0,1] is attained

  All of these variables have the Beta$(1/2, 1/2)$ distribution, also called the *Arcsin Law* on [0,1] with density at $u \in [0, 1]$

  $$\frac{1}{\pi\sqrt{u}\sqrt{1-u}}$$

  Surprising (at least for g_1, but not for the others):

  – symmetry about 1/2

  – U shape

- Hitting probability for BM

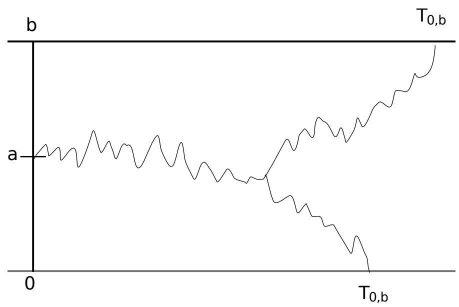
  First discuss BM started at x. This is $x + B_t,\ t \geq 0$ for (B_t) the standard BM as before.

  Let P_x governs BM started at x, then $P_x,\ x \in \mathbb{R}$ fit together as the distributions of a Markov process indexed by its starting state x.

  Problem: Look at BM started at $a,\ 0 < a < b$, run until $T_{0,b}$ when B first hit 0 or b. Then

  $$\mathbb{P}_a(B(T_{0,b}) = b) = \mathbb{P}_a(\text{B hits } b \text{ before } 0) = a/b.$$

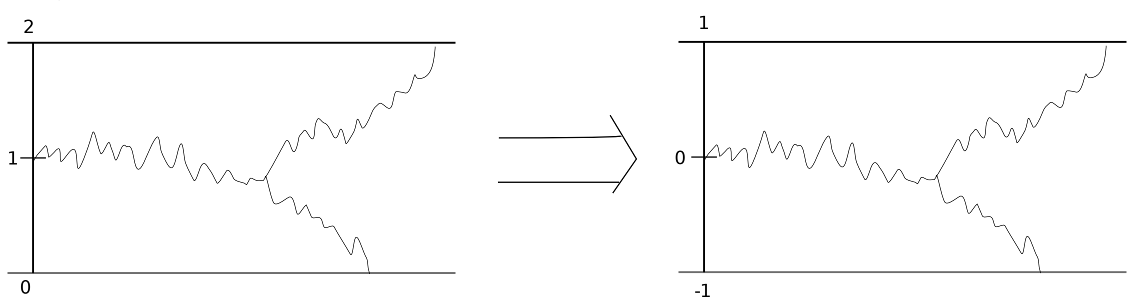
  .



Idea: BM is the continuous limit of RW. We know this formula for $\pm 1, \frac{1}{2} \uparrow \frac{1}{2} \downarrow$ RW, a, b integers, $0 < a < b$.

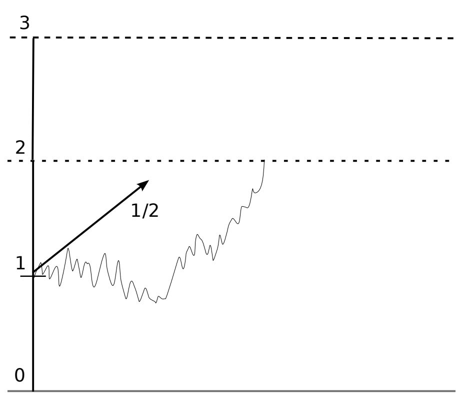
Look at the BM case for a and b integers.

Easy: $a = 1, b = 2$



Recall hitting +1 before -1 has the same probability as hitting -1 before +1.

Easy analysis shows $\mathbb{P}_1(B(T_{0,3}) = 3) = 1/3$



Key Idea: After first reach 2, the BM starts afresh as if it starts at 2.

- There is a coin tossing walk embedded in the BM.
- Exploit this to derive the a/b formula for BM.

Next, suppose a, b are rational numbers. Work on multiples of c where a/c and b/c are integers. Then $\dfrac{a/c}{b/c} = a/b$

Use rational numbers are dense in reals and obvious monotonicity in a, b to get result for real a, b.

- **Martingales** Definition: $(M_t, t \geq 0)$ is called a Martingale if

$$\mathbb{E}(M_t | M_s, 0 \leq s \leq u) = M_u \qquad \text{for } 0 < u < t$$

$$\mathbb{E}[M_t g(M_s, 0 \leq s \leq u)] = \mathbb{E}[M_u g(M_s, 0 \leq s \leq u)]$$

for $0 < u < t$, all bounded measurable function g.

Notice: $(B_t, t \geq 0)$ is a MG. Given $B_s, 0 \leq s \leq u$,

$$B_t \sim N(B_u, t - u)$$
$$\implies \mathbb{E}(B_t | B_s, 0 \leq s \leq u) = B_u \qquad \operatorname{Var}(B_t | B_s, 0 \leq s \leq u) = t - u$$

Key fact: Define idea of stopping time for (M_t), $(\tau \leq t)$ is determined by $(M_s, 0 \leq s \leq t)$ for all $t \geq 0$.

$$\text{Stopped process} \quad M_{t \wedge \tau} := \begin{cases} M_t & \text{if } t \leq \tau \\ M_\tau & \text{if } t > \tau \end{cases}$$

Theorem: If M is a MG with continuous path and τ is a stopping time for M, then $(M_{t \wedge \tau}, t \geq 0)$ is also a MG.

- Application: Take $M_t = a + B_t$ for B a standard BM.

$$T := T_{0,b} = \text{ first } t : M_t = 0 \text{ or } b$$
$$\mathbb{E}[M_{t \wedge \tau}] \equiv a$$

Then

$$\begin{aligned} a = \mathbb{E}[M_{t \wedge \tau}] =& 0 \cdot \mathbb{P}(M_T = 0, T \leq t) \\ & + b \cdot \mathbb{P}(M_T = b, T \leq t) \qquad (\to b\mathbb{P}(M_T = b)) \\ & + \mathbb{E}(M_t \mathbf{1}(T > t)) \qquad (\to 0) \end{aligned}$$

$\implies a = b \cdot \mathbb{P}(M_T = b) \implies \mathbb{P}(M_T = b) = a/b$ as claimed.

- Introduce a drift

Look at D_t a drifting BM started at $a > 0$.

$$D_t = a + \delta t + B_t$$

Find $\mathbb{P}_a(\text{D hits } b \text{ before } 0)$.

Parallel to $p \uparrow q \downarrow$ walk S_n. Found a suitable MG: $M_n := (q/p)^{S_n}$

Key idea: Look at r^{S_n}

$$M_n = \frac{r^{S_n}}{\mathbb{E}(r^{S_n})}$$

works for any r. Good choice $r = q/p$ makes $\mathbb{E}(r^{S_n}) = 1$ for $S_0 = 0$.

For continuous time/space, let $r = e^\theta$, then $M_t = \dfrac{e^{\theta B_t}}{\mathbb{E}(e^{\theta B_t})}$. Look for continuous analog with B instead of S_n.

First

$$\mathbb{E}(e^{\theta B_t}) = \int_{-\infty}^{\infty} e^{\theta x} \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}\frac{x^2}{t}} dx = e^{\frac{1}{2}\theta^2 t}$$

Then

$$\begin{aligned}
\mathbb{E}[M_{t+s}^{(\theta)} | B_u, 0 \leq u \leq t] &= \frac{\mathbb{E}(e^{\theta B_{t+s}} | B_u, 0 \leq u \leq t)}{\mathbb{E}(e^{\theta B_{t+s}})} \\
&= \frac{\mathbb{E}(e^{\theta(B_{t+s} - B_t)} e^{\theta B_t} | B_u, 0 \leq u \leq t)}{e^{\frac{1}{2}\theta^2 (t+s)}} \\
&= \frac{e^{\frac{1}{2}\theta^2 s} e^{\theta B_t}}{e^{\frac{1}{2}\theta^2 t} e^{\frac{1}{2}\theta^2 s}} \\
&= M_t^{(\theta)}
\end{aligned}$$

We have this for all θ!

Return to $D_t = a + \delta t + B_t$. Look at $e^{\theta D_t} = e^{\theta a} e^{\theta \delta t} e^{\theta B_t}$. Want $\theta\delta = -\frac{1}{2}\theta^2 \Leftrightarrow \theta = -2\delta$, then $e^{\theta D_t} = e^{\theta a} M_t^{(\theta)}$

Conclusion: If D_t is a drifting BM, then $(e^{-2\delta D_t}, t \geq 0)$ is a MG. Analog of S_n is $p \uparrow q \downarrow$ walk and $(q/p)^{S_n}$ is a MG.

Exercise: Use this MG to solve the problem of P_a(D hits b before 0).

Stat 150 Stochastic Processes **Spring 2009**

Lecture 28: Brownian Bridge

Lecturer: Jim Pitman

- From last time:

  Problem: Find hitting probability for BM with drift δ, $D_t := a + \delta t + B_t$, $\mathbb{P}_a^{(\delta)}(D_t \text{ hits } b \text{ before } 0)$.

  Idea: Find a suitable MG. Found $e^{-2\delta D_t}$ is a MG.

  Use this: Under $\mathbb{P}_a$ start at a. $\mathbb{E}_a e^{-2\delta D_0} = e^{-2\delta a}$. Let T = first time D_t hits 0 or b. Stop the MG at time T, $M_t = e^{-2\delta D_t}$. Look at $(M_{t\wedge T}, t \geq 0)$.

  For simplicity, take $\delta > 0$, then

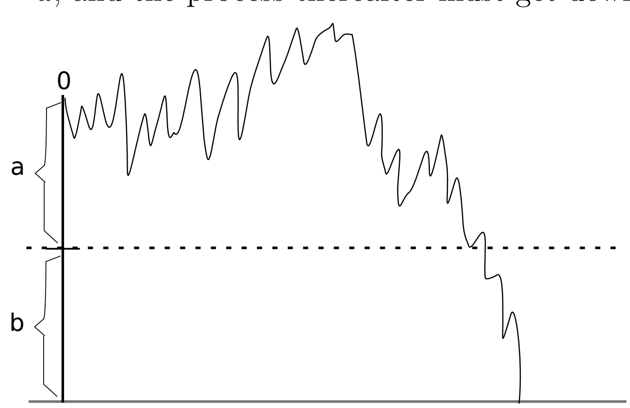
  $$1 \geq e^{-2\delta D_t} > e^{-2\delta a} > e^{-2\delta b} > 0.$$

  For $0 \leq t \leq T$, get a MG with values in [0,1] when we stop at T. Final value of MG $= 1 \cdot \mathbf{1}(\text{hit } 0 \text{ before } b) + e^{-2\delta b} \cdot \mathbf{1}(\text{hit } b \text{ before } 0)$

  $$\left.\begin{array}{rcl} 1 & = & \mathbb{P}(\text{hit } b) + \mathbb{P}(\text{hit } 0) \\ e^{-2\delta a} & = & 1 \cdot \mathbb{P}(\text{hit } 0) + e^{-2\delta b} \cdot \mathbb{P}(\text{hit } b) \end{array}\right\} \Longrightarrow \mathbb{P}(\text{hit } b) = \frac{1 - e^{-2\delta a}}{1 - e^{-2\delta b}}$$

  Equivalently, start at 0, $\mathbb{P}(B_t + \delta t \text{ ever reaches } -a) = e^{-2\delta a}$.

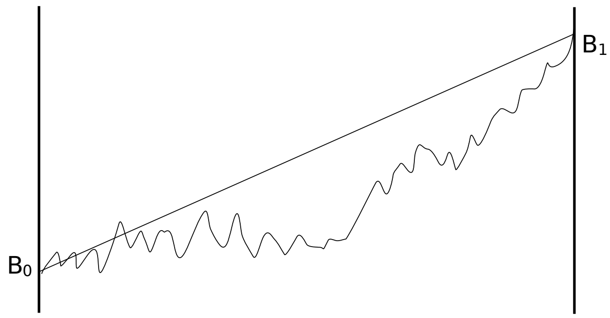
  Let $M := -\min_t(B_t + \delta t)$, then $\mathbb{P}(M \geq a) = e^{-2\delta a} \Longrightarrow M \sim \text{Exp}(2\delta)$. Note the *memoryless property*: $\mathbb{P}(M > a + b) = \mathbb{P}(M > a)\mathbb{P}(M > b)$ This can be understood in terms of the strong Markov property of the drifting BM at its first hitting time of $-a$: to get below $-(a+b)$ the process must get down to $-a$, and the process thereafter must get down a further amount $-b$.



- **Brownian Bridge**

  This is a process obtained by conditioning on the value of B at a fixed time, say time=1. Look at $(B_t, 0 \le t \le 1 | B_1)$, where $B_0 = 0$.

  Key observation: $\mathbb{E}(B_t|B_1) = tB_1$

  To see this, write $B_t = tB_1 + (B_t - tB_1)$, and observe that tB_1 and $B_t - tB_1$ are independent. This independence is due to the fact that B is a Gaussian process, so two linear combinations of values of B at fixed times are independent if and only if their covariance is zero, and we can compute:

  $$\mathbb{E}(B_1(B_t - tB_1)) = \mathbb{E}(B_tB_1 - tB_1^2) = t - t \cdot 1 = 0$$

  since $\mathbb{E}(B_sB_t) = s$ for $s < t$.

  Continuing in this vein, let us argue that $\underbrace{(B_t - tB_1, 0 \le t \le 1)}_{\text{process}}$ and $\underbrace{B_1}_{r.v.}$ are independent.

  Proof:

  Let $\hat{B}(t_i) := B_{t_i} - t_iB_1$, take any linear combination:

  $$\mathbb{E}((\sum_i \alpha_i \hat{B}(t_i))B_1) = \sum_i \alpha_i \mathbb{E}\hat{B}(t_i)B_1 = \sum_i \alpha_i \cdot 0 = 0$$

  This shows that every linear combination of values of $\hat{B}$ at fixed times $t_i \in [0,1]$ is independent of B_1. It follows by measure theory (proof beyond scope of this course) that B_1 is independent of every finite-dimensional vector $(\hat{B}(t_i), 0 \le t_1 < \cdots < t_n \le 1)$, and hence that B_1 is independent of the whole process $(\hat{B}(t), 0 \le t \le 1)$. This process $\hat{B}(t) := (B_t - tB_1, 0 \le t \le 1)$ is called **Brownian bridge**.

  Notice: $\hat{B}(0) = \hat{B}(1) = 0$. Intuitively, $\hat{B}$ is B conditioned on $B_1 = 0$. Moreover, B conditioned on $B_1 = b$ is $(\hat{B}(t) + tb, 0 \le t \le 1)$. These assertions about conditioning on events of probability can be made rigorous either in terms of limiting statements about conditioning on e.g. $B_1 \in= (b, b+\Delta)$ and letting Δ tend to 0, or by integration, similar to the interpretation of conditional distributions and densities.

- Properties of the bridge $\hat{B}$.

  (1) Continuous paths.

  (2)

$$\begin{aligned}\sum_i \alpha_i \hat{B}(t_i) &= \sum_i \alpha_i (B(t_i) - t_i B(1)) \\ &= \sum_i \alpha_i B(t_i) - (\sum_i \alpha_i t_i) B(1) \\ &= \text{some linear combination of } B(\cdot)\end{aligned}$$

  $\Longrightarrow \hat{B}$ is a **Gaussian process**. As such it is determined by its mean and covariance functions: $\mathbb{E}(\hat{B}(t)) = 0$. For $0 \le s < t \le 1$,

$$\begin{aligned}\mathbb{E}(\hat{B}(s)\hat{B}(t)) &= \mathbb{E}[(B_s - sB_1)(B_t - tB_1)] \\ &= \mathbb{E}(B_s B_t) - s\mathbb{E}(B_t B_1) - t\mathbb{E}(B_s B_1) + st\mathbb{E}(B_1^2) \\ &= s - st - st + st \\ &= s(1-t)\end{aligned}$$

  Note symmetry between s and $1-t \Longrightarrow (\hat{B}(1-t), 0 \le t \le 1) \stackrel{d}{=} (\hat{B}(t), 0 \le t \le 1)$

  (3) $\hat{B}$ is an inhomogeneous Markov Process. (Exercise: describe its transition rules, i.e. the conditional distribution of $\hat{B}_t$ given $\hat{B}_s$ for $0 < s < t$.)

- Application to statistics

  Idea: How to test if variables $X_1, X_2, \ldots$ are like an independent sample from some continuous cumulative distribution function$F$?

  Empirical distribution $F_n(x) :=$ fraction of $X_i, 1 \le i \le n$, with values $\le x$. $F_n(x)$ is a random function of x which increases by $1/n$ at each data point. If the X_i's are IID(F), then $F_n(x) \to F(x)$ as $n \to \infty$ by LLN. In fact, $\sup_x |F_n(x) - F(x)| \to 0$. (Glivenko-Cantelli theorem).

  Now $\mathrm{Var}(F_n(x)) = F(x)(1 - F(x)/n$. So consider $D_n := \sqrt{n} \sup_x |F_n(x) - F(x)|$ as a test statistic for data distributed as F.

  Key observation: Let $U_i := F(X_i) \sim U_{[0,1]}$. The distribution of D_n does not depend on F at all. To compute it, may as well assume that $F(x) = x, 0 \le x \le 1$; that is, sampling from $U_{[0,1]}$. Then

$$\begin{aligned}D_n :=& \sqrt{n} \sup_n |F_n(u) - u| \\ F_n(t) =& \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(U_i \le u)\end{aligned}$$

Idea: Look at the process $X_n(t) := (\sqrt{n}(F_n(t) - t), 0 \leq t \leq 1)$. You can easily check for $0 \leq s \leq t \leq 1$:

$$\mathbb{E}(X_n(t)) = 0$$
$$\mathrm{Var}(X_n(t)) = t(1-t)$$
$$\mathbb{E}[X_n(s)X_n(t)] = s(1-t)$$

We see $(X_n(t), 0 \leq t \leq 1) \xrightarrow{d} (\hat{B}(t), 0 \leq t \leq 1)$ in the sense of convergence in distribution of all finite-dimensional distributions: In fact, with care from linear combination,

$$\underbrace{(X_n(t_1), \dots, X_n(t_k))}_{Centered and Scaled Discrete} \xrightarrow{d} \underbrace{(\hat{B}(t_1), \dots, \hat{B}(t_k))}_{Gaussian}$$

Easy: For any choice of α_i and t_i, $1 \leq i \leq m$,
$\sum_{i=1}^{m} \alpha_i X_n(t_i)$ = centered sum of n IID random variables $\xrightarrow{d}$ Normal by usual CLT. And it can be shown that also:
$D_n := \sup_t |X_n(t)| \xrightarrow{d} \sup_t |\hat{B}(t)|$
Formula for the distribution function is a theta function series which has been widely tabulated. The convergence in distribution to an explicit limit distribution is due to Kolmogorov and Smirnov in the 1930s. Identification of the limit in terms of Brownian bridge was made by Doob in the 1940s. Rigorous derivation of a family of such limit results for various functionals of random walks converging to Brownian motion is due to Donsker in the 1950's (Donsker's theorem).

Theorem: (Dvoretsky-Erdös-Kakutani) With probability 1 the path of a BM is nowhere differentiable.