STAT 171 NOTES

GREG TAM

Contents

1. January 27th, 2014

Stochastic Processes:

$$\{X_0, X_1, X_2, \dots, X_n\}$$

Some examples include X_n as

(1) Stock Prices
(2) Molecular Dynamics
(3) Weather Patterns
(4) Evolution
(5) Movement of a tea particle in a cup of hot water
(6) Speech
(7) Sway
(8) Customers in a Restaurant
(9) Score in a Sports Game

The first topic we study is Markov Chains, which can be split into discrete time or continuous time. Discrete time typically is used to model long run behaviour. Continuous time is used in Poisson processes and birth-death chains.

2. January 29th, 2014

More review:

- $X \sim \text{Bern}(p)$. Then
$$X = \begin{cases} 0 & 1-p \\ 1 & p \end{cases}$$
$$\mathbb{E}[X^m] = p \quad \forall m$$
- If $X \sim \mathcal{N}(0,1)$,
$$\mathbb{E}\left[X^{99}\right] = 1$$
- $X \sim \text{Bin}(n,p)$
- $X \sim \text{Pois}(\lambda)$ if
$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$
One approximation is that
$$\text{Bin}(n,p) \approx \text{Pois}(c)$$
with n large, p small and $np \approx c$.
- $X \sim \text{Geom}(p)$. $A_1, A_2, \dots,$ are independent events with $\mathbb{P}(A_i) = p$ for all i. Then
$$\mathbb{P}(X = k) = (1-p)^k p$$
To verify this, we sum over all k
$$\begin{aligned}\sum_{k=0}^{\infty} \mathbb{P}(X = k) &= \sum_{k=0}^{\infty} p(1-p)^k \\ &= p\left(1 + (1-p) + \cdots\right) \\ &= p\frac{1}{1-(1-p)} \\ &= 1\end{aligned}$$
With this, we have
$$\mathbb{E}[X] = \frac{1-p}{p}$$
and
$$\text{Var}(X) = \frac{1-p}{p^2}$$
- Let Z be the number of failures before r successes. Then this is called the negative binomial. When $r = 1$, this is the geometric.
$$\mathbb{P}(Z = k) = \binom{k+r-1}{k} p^r (1-p)^k$$
We have
$$\mathbb{E}[Z] = \frac{r(1-p)}{p}$$

- Multinomial: We have n independent trials and r possible outcomes with probabilities $p_1, p_2, \dots, p_r$ with $p_1 + \cdots + p_r = 1$. A draw $Z \sim \text{Mult}(n, p_1, \dots, p_r)$ is such that

$$Z = (k_1, \dots, k_r) \qquad k_1 + \cdots + k_r = n$$

has probability

$$\frac{n!}{k_1! k_2! \cdots k_r!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$$

- Exponential has a parameter λ. The probability density function is

$$f_T(t) = \lambda e^{-\lambda t}, t > 0$$

where $T \sim \text{Expo}(\lambda)$. We further have

$$F_T(t) = 1 - e^{-\lambda t}$$

and

$$\mathbb{E}[X] = \frac{1}{\lambda}$$

"Memoryless" Property. Let us calculate

$$\mathbb{P}(T > 10 | T > 5)$$

First,

$$\mathbb{P}(T > 5) = e^{-\lambda 5}$$

We know that

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

So,

$$\begin{aligned}
\mathbb{P}(T > 10 | T > 5) &= \frac{\mathbb{P}(T > 10 \cap T > 5)}{\mathbb{P}(T > 5)} \\
&= \frac{\mathbb{P}(T > 10)}{\mathbb{P}(T > 5)} \\
&= \frac{e^{-\lambda 10}}{e^{-\lambda 5}} \\
&= e^{-\lambda 5}
\end{aligned}$$

Let us calculate

$$\mathbb{P}(S \in [s, s + \mathrm{d}s] | S \geq s) = \frac{g(s)\mathrm{d}s}{1 - G(s)} = r(s)\mathrm{d}s$$

where $r(s)$ is called the hazard function.

- $X \sim \text{Uniform}[0, 1]$ if the density is constant across $[0, 1]$. $\mathbb{E}[X] = \frac{1}{2}$. $\text{Var}(X) = \frac{1}{12}$.

Suppose $X = 0, 1, 2, \dots$. Then

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

Claim.

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} \mathbb{P}(X > k)$$

Proof.

$$\begin{aligned}
\sum_{k=0}^{\infty} \mathbb{P}(X > k) &= \mathbb{P}(X \geq 1) + \mathbb{P}(X \geq 2) + \cdots \\
&= \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) \\
&\quad + \qquad\qquad \mathbb{P}(X = 2) + \mathbb{P}(X = 3) \\
&\quad + \cdots \\
&= \sum_{k=0}^{\infty} k \mathbb{P}(X = k)
\end{aligned}$$

$\square$

Suppose a jar has n chips. We pick with replacement until we pick the same chip twice.

$$\begin{aligned}
&\mathbb{P}(X=1)=0\\
&\mathbb{P}(X>2)=\frac{n}{n}\cdot\frac{n-1}{n}=1\left(1-\frac{1}{n}\right)\\
&\mathbb{P}(X>k)=1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\\
&\mathbb{P}(X=k)=\mathbb{P}(X>k-1)-\mathbb{P}(X>k)
\end{aligned}$$

More importantly,

$$\mathbb{E}[X]=\sum_{k=1}^{n+1}\mathbb{P}(X>k)$$

For a random variable X, $\phi(t)$ is called the characteristic function.

$$\phi_X(t)=\mathbb{E}\big[e^{itX}\big]\qquad i=\sqrt{-1}$$

which by definition is equal to

$$\int_{-\infty}^{\infty}e^{itx}f_X(x)\,\mathrm{d}x$$

For the normal, $X\sim\mathcal{N}(0,1)$,

$$\begin{aligned}
\phi(t)&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{itx}e^{-x^2/2}\,\mathrm{d}x\\
&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}[(x-it)^2-(it)^2]}\,\mathrm{d}x\\
&=e^{\frac{1}{2}(it)^2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-it)^2}\,\mathrm{d}x\\
&=e^{-\frac{t^2}{2}}
\end{aligned}$$

Two functions are nearly the same if their characteristic functions are nearly the same.

Let $X_1,\dots,X_n\overset{iid}{\sim}\mathcal{P}$ with $\mathbb{E}[X_i]=0,\mathrm{Var}(X_i)=1$, that is they are standardized and come from any distribution.

$$\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$$

The Central Limit Theorem says that

$$\sqrt{n}(\bar{X}_n-\mathbb{E}\big[\bar{X}_n\big])\overset{d}{\rightsquigarrow}\mathcal{N}(0,1)$$

3. February 3rd, 2014

From last lecture, we have that the Central Limit Theorem says if X_i are iid and $\mathbb{E}[X_i]=0$, $\mathbb{E}\big[X_i^2\big]=1$, then

$$\sqrt{n}\bar{X}_n\overset{d}{\rightsquigarrow}\mathcal{N}(0,1)$$

where

$$\sqrt{n}\bar{X}_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i$$

Now, if $Z\sim\mathcal{N}(0,1)$, $S=\sqrt{n}\bar{X}_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i$. The characteristic function is

$$\mathbb{E}\big[e^{itZ}\big]=e^{-t^2/2}$$

Knowing this, we have

$$\begin{aligned}
\mathbb{E}\big[e^{itS}\big]&=\mathbb{E}\Big[e^{i\frac{t}{\sqrt{n}}\sum_{j=1}^{n}X_j}\Big]\\
&=\prod_{j=1}^{n}\mathbb{E}\Big[e^{i\frac{t}{\sqrt{n}}X_j}\Big]
\end{aligned}$$

Now we Taylor expand.

$$e^{i\frac{t}{\sqrt{n}}X_j}\approx 1+i\frac{t}{\sqrt{n}}X_j+\frac{1}{2}\left(\frac{it}{\sqrt{n}}\right)^2X_j^2+O\left(\frac{1}{n^{3/2}}\right)$$

Taking the expectation, we have

$$\mathbb{E}\Big[e^{i\frac{t}{\sqrt{n}}X_j}\Big]\approx 1+\frac{1}{2}\left(\frac{it}{\sqrt{n}}\right)^2=1-\frac{t^2}{2n}$$

This gives us that

$$\mathbb{E}\left[e^{itS}\right] \approx \left(1 - \frac{t^2}{2n}\right)^n$$
$$\approx e^{-t^2/2}$$

The last step follows from the fact that

$$\left(1 - \frac{t^2}{2n}\right)^n = e^{n \log(1 - t^2/n)}$$
$$\to e^{-\frac{nt^2}{2n}}$$
$$= e^{-t^2/2}$$

since $e^{\log x} = x$, so $\log(1+x) \approx x$.

Suppose we take a poll and ask people what their favourite number is? How certain are we? Suppose $n = 7, \bar{x} = 5, \sigma^2_{pop} = 1$. Then our confidence interval is

$$5 \pm \frac{1}{\sqrt{n}} \cdot C_\alpha$$

3.1. **Random Walk.**

Definition 1 (Random Walk). *Suppose $\{X_n\}_{n\geq 0} = X_0, X_1, X_2, \ldots$ is a stochastic process. $\{X_n\}_{n\geq 0}$ is a random walk if*

$$X_{n+1} = \begin{cases} X_n + 1 & \textit{prob. } \frac{1}{2} \\ X_n - 1 & \textit{prob. } \frac{1}{2} \end{cases}$$

This is the fundamental building block of stock prices.

3.2. **Markov Chains.**

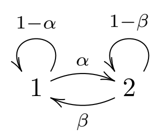
Definition 2 (Markov Chain).

(1) $\mathbb{P}(X_0 = i_0) = p_{i_0}$

(2) $\mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) = \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1})$

These two properties are what is called the Markovian property.

Example 1.

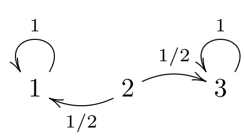


The transition matrix here is

$$\mathcal{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$$

Example 2. *Suppose we have transition matrix*

$$\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}$$



It is interesting to note that $p_{ij} \geq 0$ and $\sum_{j=1}^{\infty} p_{ij} = 1$. The fact that if you are at states 1 or 3 and will stay there is called ***absorption****.*

How would you compute

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n)$$

We do this by multiplying conditional probabilities. Start with $n = 2$.

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2) = \mathbb{P}(X_2 = i_2 | X_1 = i_1, X_0 = i_0)\,\mathbb{P}(X_1 = i_1, X_0 = i_0)$$
$$= \mathbb{P}(X_2 = i_2 | X_1 = i_1)\,\mathbb{P}(X_1 = i_1 | X_0 = i_0)\,\mathbb{P}(X_0 = i_0)$$
$$= p_{i_1 i_2} p_{i_0 i_1} p_{i_0}$$

What is an example of something that is not a Markov Chain?

$$X_{n+1} = \begin{cases} \max(X_n, X_{n-1}) + 1 & \frac{1}{2} \\ \max(X_n, X_{n-1}) - 1 & \frac{1}{2} \end{cases}$$

Example 3 (Ehrenfest Model (Urn Models))**.** *Suppose we have two urns, urn A and urn B with a permeable membrane between the two. Suppose there are $2s$ balls, k of which are in urn A and the remaining $2a-k$ in urn B.*

Let Y_n be the number of balls in urn A at the nth step. Let $X_n = Y_n - a$. The range of X_n is

$$X_n \in \{-a, -a+1, \ldots, -1, 0, 1, \ldots, a\}$$

Our transition matrix here is a $2a+1 \times 2a+1$ matrix with

$$p_{ij} = \begin{cases} \frac{a-i}{2a} & j = i+1 \\ \frac{a+i}{2a} & j = i-1 \\ 0 & \textit{otherwise} \end{cases}$$

3.3. **Simulation Assignment.**

(1) Understand/Verify CLT
(2) Simulate Discrete Markov Chains (2 state and 3 state)

Example 4 (Wright-model)**.**

- *In generation 1, there are two types of genes, type a and type A. There are $2N$ people in the population. Suppose j have type a and $2N-j$ have type A.*
- *In generation 2, we have $2N$ Bernoulli trials for the genes with*

$$p_j = \frac{j}{2N} \qquad q_j = 1 - p_j$$

Suppose X_n is the number of people with type a gene. The absorbing states are 0 and $2N$, in which case there would be no new people of the respective gene types.

What is $\mathbb{P}(X_{n+1} = j | X_n = i)$? $p_{ij} = \frac{i}{2N}$ and we have $2N$ independent Bernoulli trials.

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \binom{2N}{j} p_i^j (1-p_i)^{2N-j}$$

Suppose that we alter this so we have a mutation, where

$$\begin{cases} a \to A & \alpha \\ A \to a & \beta \end{cases}$$

Here, we can instead look at

$$\frac{\mathbb{E}[\textit{type a gene}]}{2N} \to \frac{i(1-\alpha) + (2N-i)\beta}{2N}$$

4. February 5th, 2014

Example 5 (Discrete Queue Markov Chain)**.** *Suppose there is a queue of people and at each time, only one person can be serviced. Let ξ_n be a random variable which tells you the distribution over the number of customers serviced over the nth period.*

$$\mathbb{P}(\xi_n = k) = a_k$$

- *Let X_n be the number of customers waiting in line after time n*
- $$j = \begin{cases} i - 1 + \xi_n & i \geq 1 \\ \xi_n & i = 0 \end{cases}$$
- $$p_{ij} = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ 0 & a_0 & a_1 & \cdots \\ 0 & 0 & a_0 & \cdots \end{bmatrix}$$

  and you see that this matrix is upper triangular.

  $$\lim_{n\to\infty} \mathbb{P}(X_n = 0 | X_1 = 10) = \pi_0$$

  Consider

  $$\pi_k = \lim_{n\to\infty} \mathbb{P}(X_n = k | X_1 = 10)$$

Then the average is

$$\sum_{k=0}^{\infty} k\pi_k$$

Example 6 (AR-1: Autoregressive Process)**.**

$$X_n = \rho X_{n-1} + \varepsilon_n$$

where

$$\varepsilon_n \sim \mathcal{N}(0,1) \qquad \rho \in \mathbb{R}$$

and $X_0 = x_0 \in \mathbb{R}$*.*

$$\begin{aligned} X_n &= \rho X_{n-2} + \varepsilon_n \\ &= \rho(\rho X_{n-2} + \varepsilon_{n-1}) + \varepsilon_n \\ &= \rho^2 X_{n-2} + \rho\varepsilon_{n-1} + \varepsilon_n \end{aligned}$$

We know

$$X_n | X_{n-1} \sim \mathcal{N}(\rho X_{n-1}, 1)$$

By recursion,

$$\begin{aligned} X_n | X_{n-2} &\sim \mathcal{N}(\rho^2 X_{n-2}, 1 + \rho^2) \\ X_n | X_{n-3} &\sim \mathcal{N}(\rho^3 X_{n-3}, 1 + \rho^2 + \rho^4) \\ &\sim \mathcal{N}(\rho^n X_0, 1 + \rho^2 + \rho^4 + \rho^6 + \cdots) \quad \textit{as } n \to \infty \end{aligned}$$

Now, if $|\rho| < 1$*, then*

$$1 + \rho^2 + \rho^4 + \rho^6 + \cdots = \frac{1}{1-\rho^2}$$

For AR-1, the "stationary distribution" is $\mathcal{N}(0, \frac{1}{1-\rho^2})$*.*

Question: What if we have $X_0 \sim \mathcal{N}(0, \frac{1}{1-\rho^2})$?
Activity: Simulate AR-1. Find an M to see how quickly it converges to this stationary distribution. What happens when $\rho \approx 1$.

$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \end{array}$$

$$\begin{tikzcd} 1 \arrow[loop above, "1-\alpha"] \arrow[r, bend left, "\alpha"] & 2 \arrow[loop above, "1-\beta"] \arrow[l, bend left, "\beta"] \end{tikzcd}$$

What is $\mathbb{P}(X_3 = 1 | X_1 = 1)$? The "first step analysis" is to find the in-between step.

$$\mathbb{P}(X_3 = 1 | X_1 = 1) = \mathbb{P}(X_3 = 1, X_2 = 1 | X_1 = 1) + \mathbb{P}(X_3 = 1, X_2 = 2 | X_1 = 1)$$

which is an identity from the law of total probability,

$$\sum_{k=1}^{2} \mathbb{P}(X_3 = 1, X_2 = k | X_1 = 1) = \mathbb{P}(X_3 = 1 | X_1 = 1)$$

Now look at the first term.

$$\begin{aligned} \mathbb{P}(X_3 = 1, X_2 = 1 | X_1 = 1) &= \mathbb{P}(X_3 = 1 | X_2 = 1, X_1 = 1)\, \mathbb{P}(X_2 = 1 | X_1 = 1) \\ &= \mathbb{P}(X_3 = 1 | X_2 = 1)\, \mathbb{P}(X_2 = 1 | X_1 = 1) \\ &= (1-\alpha)^2 \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{P}(X_3 = 1, X_2 = 2 | X_1 = 1) &= \mathbb{P}(X_3 = 1 | X_2 = 2)\, \mathbb{P}(X_2 = 2 | X_1 = 1) \\ &= \beta\alpha \end{aligned}$$

We can also get this from the transition matrix

$$P^2 = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$$

so

$$(P^2)_{11} = \mathbb{P}(X_3 = 1 | X_1 = 1)$$

(note that this is the $(1,1)$ entry of the matrix P^2, not P_{11} squared)

This means that

$$\mathbb{P}(X_{n+1} = j | X_1 = i) = (P^n)_{ij}$$

so,

$$(P^2)_{ij} = \sum_{k=1}^{n} P_{ik} P_{kj}$$

Now, consider

$$P = \begin{array}{c c} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ \alpha & \beta & \gamma \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

Let $T = \min\{n \geq 0 : X_n = 0 \text{ or } X_n = 2\}$. Two valid quantities of interest are

(1) $u = \mathbb{P}(X_T = 0 | X_0 = 1)$

(2) $v = \mathbb{E}[T | X_0 = 1]$.

$$\begin{aligned} u &= \mathbb{P}(X_T = 0 | X_0 = 1) \\ &= \mathbb{P}(X_T = 0 | X_1 = 0, X_0 = 1)\, \mathbb{P}(X_1 = 0 | X_0 = 1) \\ &\quad + \mathbb{P}(X_T = 0 | X_1 = 1, X_0 = 1)\, \mathbb{P}(X_1 = 1 | X_0 = 1) \\ &\quad + \mathbb{P}(X_T = 0 | X_1 = 2, X_0 = 1)\, \mathbb{P}(X_1 = 2 | X_0 = 1) \\ &= 1\alpha + \beta u \end{aligned}$$

so

$$u = \frac{\alpha}{1-\beta} = \frac{\alpha}{\alpha+\gamma}$$

Now we compute the expectation. We have two methods of doing this

(1)

$$\mathbb{E}[T | X_0 = 1] = \sum_{k=1}^{\infty} k \mathbb{P}(T = k)$$

We have

$$\begin{aligned} \mathbb{P}(T = 1) &= 1 - \beta \\ \mathbb{P}(T = 2) &= \beta(1-\beta) \\ \mathbb{P}(T = k) &= \beta^{k-1}(1-\beta) \end{aligned}$$

Then

$$\begin{aligned} \mathbb{E}[T | X_0 = 1] &= \sum_{k=1}^{\infty} k(1-\beta)\beta^{k-1} \\ &= (1-\beta)(1 + 2\beta + 3\beta^2 + 4\beta^3 + \cdots) \\ &= (1-\beta)\frac{\mathrm{d}}{\mathrm{d}\beta}(\beta + \beta^2 + \beta^3 + \cdots) \\ &= (1-\beta)\frac{1}{(1-\beta)^2} \\ &= \frac{1}{1-\beta} \end{aligned}$$

(2) We have $\mathbb{P}(T > k) = \beta^k$. Recall that

$$\begin{aligned} \mathbb{E}[T | X_0 = 1] &= \sum_{k \geq 1} \mathbb{P}(T > k | X_0 = 1) \\ &= \sum_{k \geq 1} \beta^k \\ &= \frac{1}{1-\beta} \end{aligned}$$

5. February 10th, 2014

First Step Analysis:

$$\begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & 1 & 0 & 0 \\ 1 & \alpha & \beta & \gamma \\ 2 & 0 & 0 & 1 \end{array}$$

(1) $u = \mathbb{P}(X_T = 0 | X_0 = 1)$
(2) $v = \mathbb{E}[T | X_0 = 1]$
We solved this by

$$\begin{aligned} u &= \mathbb{P}(X_T = 0|X_0 = 1) \\ &= \mathbb{P}(X_T = 0|X_1 = 0)\,\mathbb{P}(X_1 = 0|X_0 = 1) \\ &\quad + \mathbb{P}(X_T = 0|X_1 = 1)\,\mathbb{P}(X_1 = 1|X_0 = 1) \\ &\quad + \mathbb{P}(X_T = 0|X_1 = 2)\,\mathbb{P}(X_1 = 2|X_0 = 1) \\ &= \alpha + u\beta \end{aligned}$$

which implies $u = \frac{\alpha}{1-\beta} = \frac{\alpha}{\alpha+\gamma}$.

$$v = 1 + \alpha \times 0 + \beta v \Rightarrow v = \frac{1}{1-\beta}$$

$$\begin{array}{c|cccc} & 0 & 1 & 2 & 3 \\ \hline 0 & 1 & 0 & 0 & 0 \\ 1 & p_{10} & p_{11} & p_{12} & p_{13} \\ 2 & p_{20} & p_{21} & p_{22} & p_{23} \\ 3 & 0 & 0 & 0 & 1 \end{array}$$

We have

$$T = \min\{n \geq 0, X_n = 0 \text{ or } 3\}$$

What is $\mathbb{P}(X_T = 0|X_0 = 1)$? Let's look at the quantity $u_i = \mathbb{P}(X_T = 0|X_0 = i)$ for $i = 1, 2$.

$$\begin{aligned} u_1 &= p_{10} + p_{11}u_1 + p_{12}u_2 \\ u_2 &= p_{20} + p_{21}u_1 + p_{22}u_2 \end{aligned}$$

Exercise: How would we find $\mathbb{E}[T|X_0 = 1]$?

Let T be the number of tosses needed until we get 2 consecutive tosses. What is $\mathbb{E}[T]$? Suppose X_n is the number of running heads.

$$\begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 & \frac{1}{2} \\ 2 & 0 & 0 & 1 \end{array}$$

It is interesting that this problem can be setup using a Markov Chain. We have that the expected time for T to happen is actually the expected absorption time (when we are in state 2).

$$\begin{aligned} u &= \mathbb{E}[T|X_0 = 0] \\ v &= \mathbb{E}[T|X_0 = 1] \end{aligned}$$

Then the first-step analysis gives that

$$\begin{aligned} u &= 1 + \frac{1}{2}u + \frac{1}{2}v \\ v &= 1 + \frac{1}{2}u \end{aligned}$$

Solving this gives $u = 6, v = 4$. Since we actually start at $X_0 = 0$, the answer here is 6.

Suppose now T is the time needed for n consecutive heads. Let

$$E_n = \mathbb{E}[T]$$

What happens when you have E_{n-1}? You need another head in which case you're done or you get a tail, but then you would need another E_n heads.

$$E_n = \frac{1}{2}(E_{n-1} - 1) + \frac{1}{2}(E_{n-1} + 1 + E_n) \Rightarrow E_n = 2E_{n-1} + 2$$

Solving this gives

$$E_n = 2^{n+1} - 2$$

5.1. **Fibonacci Sequence.**

$$0, 1, 1, 2, 3, 5, 8, 21, 34, 55, 89$$

which follows the property

$$F_n = F_{n-1} + F_{n-2}$$

One interesting property is

$$F_n^2 = F_{n-1}F_n \pm 1$$

$\{F_n\}$ itself relies on F_{n-1} and F_{n-2} so it is not Markov. Consider $Z_n = (F_n, F_{n-1})$.

$$\begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}}_{A} \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}$$

Now if we let $Z_{n+1} = AZ_n$, then this is a Markov chain. By recursion, we get

$$Z_n = A^n Z_1$$

To calculate this easily, we have to get a handle on A. Suppose A is diagonalizable, i.e. $A = u\Lambda u^{-1}$ is u is orthogonal and Λ is orthogonal. Why is this useful? Consider A^2.

$$\begin{aligned} A^2 &= AA \\ &= u\Lambda u^{-1} u\Lambda u^{-1} \\ &= u\Lambda^n u^{-1} \end{aligned}$$

So

$$A^n = u\Lambda^n u^{-1}$$

To get the determinant, we have $|A - \lambda I| = 0$.

$$\left\| \begin{matrix} 1-\lambda & 1 \\ 1 & -\lambda \end{matrix} \right\| = \lambda(\lambda - 1) - 1 = 0$$

Equivalently,

$$\lambda^2 - \lambda - 1 = 0 \Rightarrow \lambda = \frac{1 \pm \sqrt{5}}{2}$$

This is the golden ratio.

$$\varphi = \frac{1 + \sqrt{5}}{2}$$

$$A^n = u \begin{bmatrix} \varphi^n & 0 \\ 0 & (1-\varphi)^n \end{bmatrix} u^{-1}$$

$A_{ij}^n = c_1 u_1^n + c_2 u_2^n$ so

$$F_n = c_1 \varphi^n + c_2 (1 - \varphi)^n$$

We have

$$\begin{aligned} F_0 &= 0 \\ F_1 &= 1 \\ c_1\varphi + c_2(1-\varphi) &= 1 \\ c_1 + c_2 &= 0 \end{aligned}$$

which gives

$$c_1 = \frac{1}{\sqrt{5}} \qquad c_2 = -\frac{1}{\sqrt{5}}$$

$$F_n = \frac{1}{\sqrt{5}}\varphi^n - \frac{1}{\sqrt{5}}(1-\varphi)^n$$

$$u = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}}$$

5.2. **Two-state Markov Chain.**

$$\begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 1-\alpha & \alpha \\ 2 & \beta & 1-\beta \end{array}$$

$$\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \end{array}$$

Suppose

$$\pi_1 = \lim_{n\to\infty} \mathbb{P}(X_n = 1|X_1 = 1)$$

$P = 1, 1-\alpha-\beta$.

$$\begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

So this gives us our eigenvector and eigenvalues.

$$(P^n)_{11} = A + B(1-\alpha-\beta)^n$$

We know

$$P_{11}^{(0)} = A + B = 1$$
$$P_{11}^{(1)} = 1-\alpha$$

so we can solve for A and B. Computing this gives us

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}$$

This is the stationary distribution.

A transition matrix T is regular if there exists some K such that T^K has all positive entries. For regularity, it is suffiient to check

1) For every i, j, there is a path $k_1, k_2, \ldots, k_r$

$$p_{ik_1} p_{k_1k_2} p_{k_2k_3} \cdots p_{k_nj} > 0$$

2) There exists at least one i such that $p_{ii} > 0$.

Example 7.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is not regular since they get stuck in their states.

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{2}{3} \\ \frac{1}{4} & \frac{3}{4} & 0 \end{pmatrix} \end{array}$$

You cannot go from state 1 to 2 in one step, but you can go in more than one step, so this is regular.

Theorem 1. *Suppose P is a regular transition matrix on $N+1$ states, $j = 0, 1, 2, \ldots, N$. Then there exists a unique probability vectors*

$$(\pi_0, \pi_1, \ldots, \pi_N)$$

such that $\sum_{j=0}^N \pi_j = 1, \pi_i > 0$, and $\pi_K = \lim_{n\to\infty} \mathbb{P}(X_n = K|X_0 = i)$ for all i.

For any K,

$$\pi_K = \sum_{j=0}^N P_{jK} \pi_j$$

$\pi = \pi P$.

6. February 12th, 2014

Theorem 2. *Suppose P is a regular matrix with state $\{0, 1, 2, \ldots, N\}$.*

(1) Then there exists a probability vector $\pi = (\pi_0, \pi_1, \ldots, \pi_N)$ with $\sum_{k=0}^{N} \pi_k = 1$ and $\pi_k > 0$.
(2) Then there is a unique solution to the set of equations

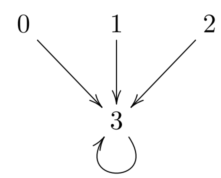
$$\pi_j = \sum_{k=0}^{N} \pi_k P_{kj}$$

*Then π is called the **stationary** or **invariant distribution***

Why exactly does

$$\pi_j = \lim_{n\to\infty} \underbrace{\mathbb{P}(X_n = j | X_0 = i)}_{(P^n)_{ij}}$$

for any i. This is by the Perron-Frobenius theorem which we will not get into.



$$\pi_3 = \pi_0 P_{03} + \pi_1 P_{13} + \pi_2 P_{23} + \pi_3 P_{33}$$

Proof.

$$P^n = P^{n-1} P$$

Therefore,

$$(P^n)_{ij} = \sum_{k=0}^{N} (P_{ik}^{n-1}) P_{kj}$$

Taking limits, we get

$$\lim_{n\to\infty} (P^n)_{ij} = \sum_{k=0}^{N} \lim_{n\to\infty} (P^{n-1})_{ik} P_{kj}$$
$$\pi_j = \sum_{k=0}^{N} \pi_k P_{kj}$$

We must now prove uniqueness. If $x = (x_0, x_1, \ldots, x_N)$ with $\sum_{k=0}^{N} x_k = 1$, we require that

$$x_j = \sum_{k=0}^{N} x_k P_{kj} \Rightarrow x_j = \pi_j$$

We have by definition

$$x_l = \sum_{j=0}^{N} x_j P_{jl} \qquad x_j = \sum_{k=0}^{N} P_{kj} x_k$$

This gives

$$x_l = \sum_{j=0}^{N} \left(\sum_{k=0}^{N} P_{kj} x_k \right) P_{jl}$$
$$= \sum_{k=0}^{N} x_k \underbrace{\left(\sum_{j=0}^{N} P_{kj} P_{jl} \right)}_{(P^2)_{kl}}$$

This implies

$$
\begin{aligned}
x_l &= \sum_{k=0}^{N} x_k P_{kl} \\
x_l &= \sum_{k=0}^{N} x_k (P^2)_{kl} \\
&\vdots \\
x_l &= \sum_{k=0}^{N} x_k (P^n)_{kl} \\
x_l &= \sum_{k=0}^{N} x_k \pi_l \\
&= \pi_l \underbrace{\sum_{k=0}^{N} x_k}_{=1}
\end{aligned}
$$

So $x_l = \pi_l$ $\square$

From a probabilistic perspective, this means $\pi = \pi P$. This just means π is the left eigenvector. So $P'\pi' = \pi'$. For

$$P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$$

What this says is

$$\begin{bmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{bmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}$$

$$
\begin{aligned}
(1-\alpha)\pi_1 + \beta\pi_2 &= \pi_1 \\
\pi_1 + \pi_2 &= 1
\end{aligned}
$$

so

$$\pi_1 = \frac{\beta}{\alpha + \beta}$$

6.1. **Doubly Stochastic Matrices.**

$$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

This is an example of a doubly stochastic matrix, where the columns sum to 1 in addition to the rows summing to 1.

$$
\begin{aligned}
\pi_j &= \sum_{j=0}^{N} P_{kj}\pi_k \\
&= \frac{1}{N+1} \sum_{k=0}^{N} P_{kj} \\
&= 1
\end{aligned}
$$

Example 8. *Suppose we have X_n and $S_n = \sum_{m=1}^{n} X_m$. Let $Y_n = S_n \mod 7$ Then what do we expect the probability $Y_n = 0$ in the long term? This is $\frac{1}{7}$. From a Markov Chain point of view, if you're in state i, the probability of going to any other state is $\frac{1}{6}$ and is 0 to go back to itself. Then this matrix is a doubly stochastic process.*

Suppose we can sample from a uniform distribution. How would we sample from $\mathcal{N}(0,1)$?

6.2. **Inverse CDF Method.** Algorithm:
(1) Get $U \sim \mathcal{U}[0,1]$
(2) Let $Z = F_X^{-1}(U)$
We have

$$F_X : \mathbb{R} \mapsto [0,1]$$
$$F_X^{-1} : [0,1] \mapsto \mathbb{R}$$

And the lemma is that $Z \sim F_X$.

Example 9. *Suppose we have a lattice. $x_i \in \{-1, 1\}$, $x = (x_1, \dots, x_n)$, with*

$$P(x) \propto e^{\beta \sum_i x_i x_{i+1}}$$

In this case, it is not possible to generate data from here.

Example 10. *Suppose we have*

$$(X,Y) \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$$

How would we sample from this assuming we have can sample from a univariate normal and there is no software to sample from the bivariate case?

$$X|Y = y \sim \mathcal{N}(\rho y, 1-\rho^2)$$
$$Y|X = x \sim \mathcal{N}(\rho x, 1-\rho^2)$$

We sample by doing the following
(1) $X_0 = 0$
(2) $Y_1 \sim f_{Y|X=0}$
(3) $X_1 \sim f_{X|Y=y_1}$
and so forth to get $X_0, Y_1, X_1, Y_2, X_2, \dots$.

When we simulate it, the time it takes to get to convergence depends on ρ. We start with $X_0 = x_0$, then $Y_1 \sim \mathcal{N}(\rho X_0, 1-\rho^2)$. Alternatively, this is the same as

$$Y_1 = \rho X_0 + \sqrt{1-\rho^2} Z_1$$

Similarly, $X_1 \sim \mathcal{N}(\rho Y_1, 1-\rho^2)$

$$X_1 = \rho Y_1 + \sqrt{1-\rho^2} Z_2$$

so

$$X_1 = \rho^2 X_0 + \rho\sqrt{1-\rho^2} Z_1 + \rho\sqrt{1-\rho^2} Z_2$$

7. February 19th, 2014

If P is a transition matrix on N states, a probability vector $\pi = (\pi_0, \dots, \pi_{N-1})$ is a **stationary distribution(measure)** if $\pi P = \pi$.

Theorem 3. *If P is regular,*
(1) Then $(P^n)_{ij} \to \pi_j$
(2) $\pi P = \pi$, *so* π *is unique.*

One surprising fact is that we can have $\pi P = \pi$ even if P is not regular. Take for example

$$P = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

Then any vector $\pi = (\pi_1, \pi_2)$ satisfies this.

$$\begin{pmatrix}\pi_1 & \pi_2\end{pmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} = \begin{pmatrix}\pi_1 & \pi_2\end{pmatrix}$$

Next, consider

$$P = \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$$

here, $\pi = (\frac{1}{2}, \frac{1}{2})$.

$$\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\end{pmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} = \begin{pmatrix}\frac{1}{2} & \frac{1}{2}\end{pmatrix}$$

Definition 3. *We say that two states i and j communicate, if*

$$P_{ij}^{(n)} > 0$$
$$P_{ji}^{(n)} > 0$$

This is denoted $i \leftrightarrow j$. This is in fact an equivalent class since
(1) $i \leftrightarrow i$
(2) $i \leftrightarrow j$, $j \leftrightarrow k$ implies $i \leftrightarrow k$.

This implies that a matrix is irreducible if all states communicate to each other.

7.1. **Period of a state i.**

$$\mathrm{GCD}\{n : p_{ii}^{(n)} > 0\}$$

$d(i)$ is denoted as the period of i. The state i is aperiodic if its period is 1. P is aperiodic if all states have period 1.

If we go back to our example where

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

then $P_{11}^{(1)} = 0$, $P_{11}^{(2)} = 1 = P_{11}^{(4)} = \cdots$.

If state i is aperiodic, then, $P_{ii}^{(n)} > 0$ for all sufficiently large n.

Example 11. *Suppose $i \leftrightarrow j$. Then $d(i) = d(j)$.*

If $P_{ij}^{(r)} > 0$ and $P_{ji}^{(s)} > 0$, then we can go from $i \to j \to i$ in $r + s$ steps with positive probability. This implies that $P_{ii}^{(r+s)} > 0$. This says $d(i)$ divides $r + s$. Suppose t is such that $P_{jj}^{(t)} > 0$. Then

$$\begin{tikzcd} i \arrow[r, "r"] & j \arrow[loop above, "t"] \arrow[r, "s"] & i \end{tikzcd}$$

which implies $P_{ii}^{(r+t+s)} > 0$, so $d(i)$ divides $r + t + s$.

Definition 4 (Recurrent)**.** *A state i is **recurrent** if we go from state i back to i in some finite time. Otherwise it is called **transient**. Let $X_0 = 1$ and*

$$M = \inf\{n \geq 1 : X_n = i\}$$

Then state i is recurrent if $M < \infty$. In fact, this is true if and only if

$$\sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty$$

We have

$$\lim_{n \to \infty} P_{ij}^{(n)} = \pi_j$$

How do we know this? Suppose $X_0 = i$.

$$a = \frac{1}{M} \sum_{k=1}^{M} \mathbb{1}\{X_k = j\}$$

Then

$$\mathbb{E}[a] = \frac{1}{M} \sum_{k=1}^{M} \mathbb{E}[\mathbb{1}\{X_k = j\}]$$
$$= \frac{1}{M} \sum_{k=1}^{M} P_{ij}^{(k)}$$

So if we have $\{a_n\}$ such that $\lim_{n \to \infty} a_n = a$, then

$$\lim_{M \to \infty} \frac{1}{M} \sum_{k=1}^{M} a_k = a$$

Applying both of these, we get

$$\lim_{M \to \infty} \mathbb{E}[a] = \pi_j$$

7.2. **Gibb's Sampler.** Suppose we have a bivariate distribution $f_{X,Y}(x,y)$, and know the conditional distributions $f_{X|Y}(x)$ and $f_{Y|X}(y)$. Also suppose that the joint distribution is hard to sample from, but the conditional distribution is easy to sample from. Then we can set $X_0 = x$ and iterate, sampling

(1) $Y_1 \sim f_{Y|X_0=x}$
(2) $X_1 \sim f_{X|Y_1=y_1}$

Suppose instead that we have a multivariate distribution $f_{X_1,X_2,\ldots,X_n}$ and we iterate $f_{X_1|X_2,\ldots,X_n}$, $f_{X_2|X_1,X_3,\ldots,X_n}$, and so forth. However, we must note that

(0) The Markov chain algorithm has stationary distribution $f_{X,Y}$.
(1) No iid samples.

7.3. **Metropolis-Hastings Algorithm.** Suppose $\pi(x)$ is a density you wish to sample from, (where we know $\pi(x)$ up to a normalizing constant). We start with $X_0 = x_0$.

(1) Choose a point $y = x_0 + Z$ where $Z \sim \mathcal{N}(0,1)$.
(2) Let

$$X_1 = \begin{cases} y & \min\left(1, \frac{\pi(y)}{\pi(x)}\right) \\ x_0 & 1 - \min\left(1, \frac{\pi(y)}{\pi(x)}\right) \end{cases}$$

So we simply take a uniform, $U \sim \mathcal{U}(0,1)$ and set $X_1 = y$ if $U < \min\left(1, \frac{\pi(y)}{\pi(x)}\right)$.

Suppose $y \sim q(y|x)$. Then the acceptance probability is given by

$$\min\left(1, \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}\right)$$

Suppose that $y \sim g$. Then we have

$$\min\left(1, \frac{\pi(y)}{\pi(x)}\frac{g(x)}{g(y)}\right)$$

Example 12. *Suppose we have a circle with points $x = (x_1, \ldots, x_n) \in \{-1,1\}^n$ equidistant around the edge of the circle. We have*

$$\pi(x) \propto \exp\left(\beta \sum_{\substack{i=1 \\ i\sim j}}^{n} x_i x_j\right)$$

where $i \sim j$ means i is adjacent to j and $\beta > 0$.

The algorithm for this is to have

$$X^0 = (x_1^0, x_2^0, \ldots, x_n^0)$$

Pick a random i and flip $x_i \leftrightarrow -x_i$.

8. February 24th, 2014

8.0.1. *Metropolis Hastings Re-explanation.* Suppose the goal is to sample from $\pi(x)$. We will create $\{X_n\}_{n\geq 1}$ with $X_0 = 0$ which is a Markov Chain such that the stationary distribution is π.

(1) We have a proposal distribution $y \sim q(y|X_0)$.
(2) Then the acceptance probability is

$$a = \min\left(1, \frac{\pi(y)}{\pi(X_0)}\frac{q(X_0|y)}{q(y|X_0)}\right)$$

(3) Simulate a $U \sim \mathcal{U}[0,1]$. If $U < a$, then $X_1 = y$. Otherwise, $X_1 = X_0$.

In the basic version, $y = X_0 + Z$, where $Z \sim \mathcal{N}(0,1)$. Then

$$q(y|X_0) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(X_0-y)^2} = q(X_0|y)$$

When doing a Metropolis-Hastings algorithm, the variance of the sample is usually higher than the target variance because of the autocorrelation. They are only equal when they are iid.

What can we vary? If σ^2 is small, then $Y = X_0 + \sigma Z$. Then $\pi(y) \approx \pi(x_0)$, so $\frac{\pi(y)}{\pi(x_0)} \approx 1$, which means we accept often. The disadvantage to this, is that we would never get anywhere. On the other hand, if σ^2 is large, then we can move far, but then $\frac{\pi(y)}{\pi(x_0)} \approx 0$ so the acceptance probability is low, so we will not move very much.

8.1. **Time Reversibility.** Let $\{X_n\}_{0\le n\le N}$ be a Markov chain and let $Y_n = X_{N-n}$. Think of this as running the Markov chain backwards. Let (π, P) be the stationary distribution of $\{X_n\}$

Theorem 4. *If we have $\{Y_n\}$, π, $\hat{P}$ then $\pi_j \hat{P}_{ji} = \pi_j P_{ij}$.*

Let's check this. For $\hat{P}$ to be a transition matrix, we require that $\sum_i \hat{P}_{ji} = 1$.

$$\begin{aligned}\sum_i \hat{P}_{ji} &= \sum_i \pi_i \frac{P_{ij}}{\pi_j} \\ &= \frac{1}{\pi_j}\pi_j \\ &= 1\end{aligned}$$

We also have

$$\begin{aligned}\sum_j \pi_j \hat{P}_{ji} &= \sum_j \pi_i P_{ij} \\ &= \pi_i \sum_j P_{ij} \\ &= \pi_i\end{aligned}$$

The only thing left to verify is that the transition matrix of $\{Y_n\}$ is indeed $\hat{P}$.

$$\begin{aligned}\mathbb{P}(Y_0 = i_0, Y_1 = i_1, \ldots, Y_N = i_N) &= \mathbb{P}(X_0 = i_N, X_1 = i_{N-1}, \ldots, X_n = i_0) \\ &= \pi_{i_{N_0}} P_{i_N, i_{N-1}} P_{i_{N-1}, i_{N-2}} \cdots \\ &= \pi_{i_{N-1}} \hat{P}_{i_{N-1}, i_N} P_{i_{N-1}, i_{N-2}} \cdots \\ &= \pi_{i_0} \hat{P}_{i_0, i_1} \hat{P}_{i_1, i_2} \cdots \hat{P}_{i_{N-1}, i_N}\end{aligned}$$

A Markov chain (π, P) is said to be reversible (time reversible) if

$$\pi_i P_{ij} = \pi_j P_{ji}$$

which implies $P = \hat{P}$.

Why is this true? Say for arbitrary λ_i

$$\lambda_i P_{ij} = \lambda_j P_{ji}$$

Then our conclusion in this case would be that λ is stationary for P. This says that

$$\sum_i \lambda_i P_{ij} = \sum_i \lambda_j P_{ji} = \lambda_j \sum_i P_{ji} = \lambda_j$$

so that

$$\lambda P = P$$

and so λ would be the stationary distribution.

Example 13.

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

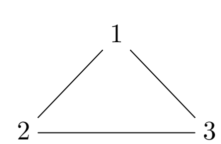
The claim here is that $\left(\frac{4}{7}, \frac{3}{7}\right)$ is the stationary distribution of this. We have $\pi_1 P_{12} = \frac{4}{7}\frac{1}{2} = \frac{2}{7}$ and $\pi_2 P_{21}\frac{3}{7}\frac{2}{3} = \frac{2}{7}$.

Example 14. *Take our homework example where we have states $0, \ldots, N$ and the probability of going from i to $i-1$ is q_i and the probability of going from i to $i+1$ is p_i, unless at states 0 and N. Then if*

$$\pi_{i-1} p_{i-1} = \pi_i q_i$$

then

$$\pi_i = \frac{p_{i-1}}{q_i}\pi_{i-1} = \frac{p_{i-1}}{q_i}\frac{p_{i-2}}{q_{i-1}}\pi_{i-2} = \cdots = \frac{p_{i-1}p_{i-2}\cdots p_{i_0}}{q_i q_{i-1}\cdots q_i}\pi_0$$



Suppose

$$P = \begin{pmatrix} 0 & a & 1-a \\ 1-a & 0 & a \\ a & 1-a & 0 \end{pmatrix}$$

Then

$$\begin{cases} \pi - 1 = \pi_2(1-a) \\ \pi_2 a = \pi_3(1-a) \\ \pi_3 a = \pi_1(1-a) \\ a = \frac{1}{2} \end{cases} \Rightarrow \pi_1 = \pi_2 = \pi_3$$

This is only true if $a = \frac{1}{2}$.

9. February 26th, 2014

Definition 5 (Detailed Balance)**.** *(π, P) satisfies detailed balance if*

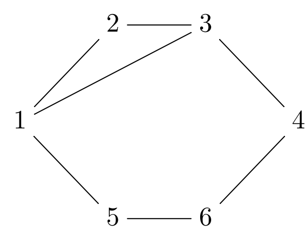
$$\pi_i P_{ij} = \pi_j P_{ji}$$

Reversibility states that if we have $\{X_n\}_{0 \le n \le N}$ with $\{\pi, P\}$ and $Y_n = X_{N-n}$ with $\{\pi, \hat{P}\}$, then we have

$$\pi_j \hat{P}_{ji} = \pi_i P_{ij}$$

Then if we have detailed balance, $\pi_i P_{ij} = \pi_j P_{ji}$.

Example 15.



Let v_i be the degree of a node. Then we assume equal probability between any nodes that share an edge, so $P_{ij} = \frac{1}{v_i}$. This implies that $v_i P_{ij} = 1$ so

$$v_i P_{ij} = v_j P_{ji}$$

Then this says that for any graph with structure, we have a random walk on the graph and we can do this.

Example 16 (Knight on a chessboard)**.** *Given a knight in a starting position, what is the expected amount of time for the knight to return? Number the squares 1 through 64. Then, this will form a graph, of which we have formed a random walk on.*

In the corners of the grid, there are two possible positions, so

$$\pi_1 = \frac{2}{2v_i} = \frac{2}{2 \times 168} = \frac{1}{168}$$

We know that the mean time is 1 over the stationary distribution, so the mean time to return to state 1 (bottom left corner) is 168.

Example 17 (Metropolis-Hastings)**.** *How do we know that the Metropolis-Hastings algorithm converges?*

We want to verify the detailed balance condition,

$$\pi(x)\mathbb{P}(X \to Y) = \pi(y)\mathbb{P}(Y \to X)$$

Note that

$$\begin{aligned} \pi(x)\mathbb{P}(X \to Y) &= \pi(x)\min\left(1, \frac{\pi(y)}{\pi(x)}\right) \\ &= \min(\pi(x), \pi(y)) \end{aligned}$$

and this is symmetric in x and y, so they are interchangeable and the detailed balance condition clearly holds.

Example 18 (Gibbs Sampler)**.** *We had that the acceptance probability in the Metropolis-Hastings is*

$$\min\left(1, \frac{\pi(y)}{\pi(x)} \frac{q(x|y)}{q(y|x)}\right)$$

This is in fact the detailed balance condition.

Suppose we have $f_{X_1,X_2}(x_1, x_2)$, $X_2' \sim f_{X_2|X_1}(x_2|x_1)$, i.e. one step of the Gibbs sampler. Then the acceptance probability is

$$\min\left(1, \frac{f_{X_1,X_2}(x_1, x_2')}{f_{X_1,X_2}(x_1, x_2)} \frac{f_{X_2|X_1}(x_2|x_1)}{f_{X_2|X_1}(x_2'|x_1)}\right)$$

Now note

$$\frac{f_{X_1,X_2}(x_1, x_2')}{f_{X_1,X_2}(x_1, x_2)} = \frac{f_{X_2|X_1}(x_2')}{f_{X_2|X_1}(x_2')} \frac{f_{X_1}(x_1)}{f_{X_2}(x_2)}$$

and so

$$\frac{f_{X_1,X_2}(x_1,x_2')}{f_{X_1,X_2}(x_1,x_2)}\frac{f_{X_2|X_1}(x_2|x_1)}{f_{X_2|X_1}(x_2'|x_1)} = 1$$

and we have that the acceptance probability is simply $\min(1,1) = 1$. *Thus, for the Gibbs sampler is a special case of Metropolis-Hastings where we* ***always*** *accept.*

Example 19. *Let $X_0 = i$. Define*

$$f_{ii}^{(n)} = \mathbb{P}(X_n = i, X_v \neq i, 1 \leq v \leq n-1 | X_0 = i)$$

that is this is the probability that the first return time to state i is n steps. We have

$$f_{ii}^{(n)} = \sum_{k=0}^{n} f_{ii}^{(k)} P_{ii}^{(n-k)}$$

The probability that we return to state i is

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)}$$

A state is recurrent if $f_{ii} = 1$ and transient if $f_{ii} < 1$.

Theorem 5. *A state i is recurrent if and only if*

$$\sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty$$

9.1. **Random walk on $\mathbb{Z}$.** This is defined as

$$X_{k+1} = \begin{cases} X_k + 1 & p \\ X_k - 1 & q \end{cases}$$

where $p + q = 1$.

Proof of above theorem. First note that $P_{00}^{(1)} = 0$ since two steps are required to return to itself. This also implies $P_{00}^{(2n+1)} = 0$. We also have

$$P_{00}^{(2n)} = \binom{2n}{n} p^n q^n$$

so our sum is given by

$$\sum_{n=1}^{\infty} \frac{(2n)!}{n!n!}(pq)^n$$

Stirling's approximation tells us that

$$\begin{aligned} n! &\approx n^{n+\frac{1}{2}} e^{-n}\sqrt{2\pi} \\ &\approx e^{\left(n+\frac{1}{2}\right)\log n} e^{-n} e^{\log(\sqrt{2\pi})} \end{aligned}$$

We use the inequality that $pq = p(1-p) \leq \frac{1}{4}$ and $p - q = \frac{1}{2}$. So we have

$$\sum_{n=1}^{\infty} \frac{(4pq)^n}{\sqrt{\pi n}} = \infty$$

only for $p = q$ since this is close to $\sum \frac{1}{\sqrt{n}}$. ∎

Suppose we have a random walk on 2d lattice. Recurrence also holds by almost the exact same calculation. However, if we go to 3 dimensions, then it is actually transient.

In finite state space, irreducibility and aperiodicity imply the Markov chain converges to a stationary distribution and it is unique. Suppose we have an altered random walk

$$X_{k+1} = \begin{cases} X_k - 1 & \frac{1}{3} \\ X_k & \frac{1}{3} \\ X_k + 1 & \frac{1}{3} \end{cases}$$

In this case, a stationary distribution would say $\pi P = \pi$, that is

$$\pi_k = \frac{1}{3}\pi_{k-1} + \frac{1}{3}\pi_k + \frac{1}{3}\pi_{k-1} \Rightarrow \pi_{k+1} = 2\pi_k - \pi_{k-1}$$

This says that

$$\pi_2 = \pi_1 - \pi_0$$
$$\pi_3 = 2\pi_1 + \pi_0$$
$$\vdots$$
$$\pi_k = k(\pi_1 - \pi_0) + \pi_0$$

This means that $\pi_k \to \infty$, therefore we cannot have $\sum_{k=0}^{\infty} \pi_k = 1$ and so there exists no π that is stationary. This is example of a Markov chain that is aperiodic and irreducible, but is null recurrent, which means that the expected return time is ∞. Positive recurrent means that the expectation of return is finite. For a countable state space, in addition to the irreducibility and aperiodicity conditions, we also require it to be positive recurrent for it to imply that it converges to a unique, stationary distribution.

Everything up until the discussion before recurrence is on exam.

10. March 5th, 2014

Second midterm will be April 14th. In subsequent tests, he will test for definitions.

Theorem 6. *In finite state spaces, Markov chains are aperiodic, irreducible, which implies*

(1) The stationary distribution is unique.
(2) The Markov chain converges to its stationary distribution.

What happens if we have infinite state spaces? In addition to aperiodicity and irreducibility, we need positive recurrence. Recall that in a finite state spaces, the mean return time is $\frac{1}{\pi_i}$, where $\pi_i > 0$, so $\frac{1}{\pi_i} < \infty$. Consider the random walk we can go from X_n to X_{n-1}, X_n, or X_{n+1} each with probability $\frac{1}{3}$.

Let

$$f_{ii}^{(n)} = \mathbb{P}(X_n = i, X_\nu \neq i, 1 \leq \nu \leq n-1 | X_0 = 1)$$

and

$$E_k = \{X_n = i, \text{ the first transition from } i \leftrightarrow i \text{ happens at } k\}$$

Then we have that

$$\mathbb{P}(E_k) = \underbrace{\mathbb{P}(X_k = i, X_\nu \neq i, 1 \leq \nu \leq k-1 | X_0 = i)}_{f_{ii}^{(k)}} P_{ii}^{(n-k)}$$

As this is the probability of going from i to i in K steps, and then also going from i to i in $n-k$ steps. Note that we also have

$$P_{ii}^{(n)} = \sum_{k=1}^{n} \mathbb{P}(E_k) = \sum_{k=1}^{n} P_{ii}^{(n-k)} f_{ii}^{(k)}$$

The importance of this calculation is to condition on the first time we go somewhere, i.e. $f_{ii}^{(k)}$.

What does $f_{11}^{(0)} + f_{11}^{(1)} + f_{11}^{(2)}$? This is the probability that we come back in at most 2 steps. Then this means

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)}$$

is the probability that we come back in finite time. Recurrence implies that $f_{ii} = 1$.

Let M be the number of return times. Then

$$\mathbb{P}(M \geq 1 | X_0 = i) = f_{ii}$$

What about if we have $\mathbb{P}(M \geq k | X_0 = i) = f_{ii}$? We claim that

$$\mathbb{P}(M \geq k | X_0 = i) = (f_{ii})^k$$

Then this says that

$$\mathbb{E}[M] = \frac{f_{ii}}{1 - f_{ii}}$$

10.1. **Coupling.** Joint distributions for two Markov chains. Consider

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

Run two Markov chains independently, where Y is stationary, i.e. $\left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$.

- $X_1 = 1, X_2, \ldots, X_n = 2$
- $Y_0, Y_1, \ldots, Y_n = 2$

As soon as $X_n = Y_n$, then we can run the chain together, that is just run one single chain. Since $\{Y_n\}$ was at stationarity, we know at n, $\{X_n\}$ is also at stationarity.

11. March 10th, 2014

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)}$$

which is the probability that the Markov chain returns to itself in finite time. The chain is recurrent if $f_{ii} = 1$ and transient if $f_{ii} < 1$.

Let $m_i = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$ be the expected return time to i. The chain is positive recurrent if $m_i < \infty$ and null recurrent if $m_i = \infty$. In infinite state spaces, irreducibility, aperiodicty, and positive recurrence implies

(1) There exists π such that $\pi P = \pi$.
(2) For any i, $\lim_{n\to\infty} P_{ji}^{(n)} = \frac{1}{m_i} = \pi_i$.

Let M be the number of times the Markov chain returns to itself. How do we find $\mathbb{P}(M \geq k)$? We know that $\mathbb{P}(M \geq 1) = f_{11}$. The probability $\mathbb{P}(M \geq 2) = f_{ii}^2$ by the Markovian property. Therefore,

$$\mathbb{P}(P \geq k) = f_{ii}^k$$

This gives

$$\mathbb{E}[M] = \frac{f_{ii}}{1 - f_{ii}}$$

Note that if this is recurrent, then $f_{ii} = 1$, and $\mathbb{E}[M] = \infty$. We can rewrite

$$M = \sum_{n=0}^{\infty} \mathbb{1}\{X_n = i\}$$

Suppose $X_0 = i$. Then

$$\mathbb{E}[M] = \sum_{n=1}^{\infty} \mathbb{P}(X_n = i | X_0 = i) = \sum_{n=1}^{\infty} P_{ii}^{(n)}$$

11.1. **A Return to Coupling (Not Testable).** Suppose

$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \end{array}$$

We have

$$\pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$$

$$X_0 = 1, X_1, \ldots X_\tau$$

$$X_{\tau+1}, X_{\tau+2}$$

$$Y_0 \sim \pi, Y_1, \ldots$$

We know

$$\lim_{n\to\infty} P^{(n)}(i,j) = \pi_j$$

How do we quantitatively verify that the two distributions have converged. To do this we look at something called total variation distance.

$$d_{TV}(P^n(i,\cdot), \pi) \leq \mathbb{P}(\tau > n) = \mathbb{P}(X_n = Y_n)$$

11.1.1. *Coupling from the past.* If we start the chain at $t = 0$, then at $t = \infty$, we should have achieved stationarity. We can just shift back and start the chain from $t = -\infty$ so that at $t = 0$, we have achieved stationarity. We can start the chain from $t = -1, -2, \ldots$ and hope that they will become coupled by $t = 0$, by which point we will have achieved stationarity.

11.2. **Branching Processes.** Suppose we start with one parent. It has probability $\frac{3}{4}$ of not having children and probability $\frac{1}{4}$ of having 2 children. What is the probability that it will die out?

Let ξ be such that

$$\mathbb{P}(\xi = k) = a_k, \quad k \geq 0$$

and $\sum_k a_k = 1$. Let X_n be the number of children at the nth step.

$$X_{n+1} = \xi_1^{(n)} + \xi_2^{(n)} + \cdots + \xi_{X_n}^{(n)}$$

11.2.1. *Random Sums.* Let N be a random variable such that $\mathbb{E}[N] = \nu$. Let

$$Z = \xi_1 + \xi_2 + \cdots + \xi_N$$

We condition on N to get this value.

$$\mathbb{E}[Z] = \mathbb{E}[\mathbb{E}[Z|N]] = \mathbb{E}[N\mu] = \mu\mathbb{E}[N] = \mu\nu$$

Let $\mathbb{E}[X_n] = M_N$. Then we know $\mathbb{E}[X_{n+1}] = M_{n+1} = M_n \cdot \mu$ so

$$M_n = \mu^n$$

$\mu < 1$ is called a subcritical case, $\mu > 1$ is called a supercritical case and $\mu = 1$ is called a critical case.

If $\text{Var}(N) = \tau^2$ and $\text{Var}(\xi) = \sigma^2$, then

$$\text{Var}(Z) = \nu\sigma^2 + k^2\tau^2$$

In our branching process case, $\text{Var}(X_n) = V_n$, where

$$V_{n+1} = \sigma^2 k^n + \mu^2 V_n$$

and

$$V_n = \begin{cases} \sigma^2 k^{n-1} n & k = 1 \\ \sigma^2 k^{n-1} \frac{1-k^n}{1-k} & k \neq 1 \end{cases}$$

If N is the random time the population dies out, then

$$u_n = \mathbb{P}(N \leq n) = \mathbb{P}(X_n = 0)$$

Doing a first step analysis, we have

$$u_n = \sum_{k=0}^{\infty} a_k (u_{n-1})^k$$

12. March 12th, 2014

12.1. **Gambler's Ruin.** Suppose we have a game where two players flip coins successively. with probability p and q, player 1 will give or receive a dollar to player 2 respectively. There are N total dollars. The game ends when player 1 has 0 or player 2 has 0 (in which case player 1 has N). Let these flips be $R_1, R_2, \ldots$ where $R_i = \pm 1$. Let $\{X_n\}$ be the number of dollars player 1 has, with $X_0 = i, X_1 + i + R_1, \ldots$. By first step analysis, we have

$$u_i = qu_{i-1} + pu_{i+1}$$

The probability of player 1 going broke is

$$\frac{N-i}{N}$$

if $p = \frac{1}{2} = q$. For general p and q, theis value is

$$\frac{\left(\frac{q}{p}\right)^i - \left(\frac{q}{n}\right)^N}{1 - \left(\frac{q}{p}\right)^N}$$

12.2. **Branching Processes.** We have bacteria that have $\frac{3}{4}$ probability of not producing offspring and $\frac{1}{4}$ probability of producing 2 offspring.

$$u_n = \mathbb{P}(\text{population is extinct in } n \text{ steps})$$

recalling that $\mathbb{P}(\xi = k) = a_k$ and

$$u_n = \sum_{k=0}^{\infty} (u_{n-1})^k a_k$$

12.2.1. *Generating Function.* $\mathbb{P}(\xi = k) = a_k$ is called the generating function. Then

$$\phi_\xi(s) = \mathbb{E}\big[s^\xi\big] \qquad \text{for } 0 \le s \le 1$$

We can write this as

$$\phi_\xi(s) = \sum_{k=0}^{\infty} a_k s^k = a_0 + a_1 s + a_2 s^2 + \cdots$$

Note that $\phi_\xi(0) = a_0$, which is the probability that there are no children. If we differentiate, we get

$$\frac{\mathrm{d}}{\mathrm{d}s}\phi_\xi(s) = a_1 + a_2 s + 3a_3 s^2$$

and we note that $\phi'_\xi(0) = a_1$. Doing this iteratively, we get

$$a_k = \left.\frac{1}{k!}\frac{\mathrm{d}^k}{\mathrm{d}s^k}\phi_\xi(s)\right|_{s=0}$$

Notice that

$$\begin{aligned}
\phi_\xi(s) &= a_0 + a_1 s + a_2 s^2 + \cdots \\
\phi'_\xi(s) &= a_1 + 2a_2 s + \cdots \\
\phi'_\xi(1) &= a_1 + 2a_2 + 3a_3 \\
&= \sum_{k=0}^{\infty} k a_k \\
&= \mathbb{E}[\xi]
\end{aligned}$$

and we can see why this would be called the generating function.

$$\begin{aligned}
\phi''_\xi(1) &= 2a_2 + 3 \cdot 2a_3 + \cdots \\
&= \sum_{k=0}^{\infty} k(k-1)a_k \\
&= \sum_{k=0}^{\infty} k^2 a_k - \sum_{k=0}^{\infty} k a_k \\
&= \mathbb{E}[\xi(\xi - 1)]
\end{aligned}$$

12.2.2. *Independence.* If ξ, θ are independent, then

$$\begin{aligned}
\phi_{\xi+\theta}(s) &= \mathbb{E}\big[s^{\xi+\theta}\big] \\
&= \mathbb{E}\big[s^\xi\big]\,\mathbb{E}\big[s^\theta\big] \\
&= \phi_\xi(s)\phi_\theta(s)
\end{aligned}$$

Recall that

$$\phi_\xi(s) = \sum_{k=0}^{\infty} a_k s^k$$

and

$$u_n = \sum_{k=0}^{\infty} (u_{n-1})^k a_k$$

From this, we see that

$$u_n = \phi_\xi(u_{n-1})$$

Suppose we take limits (denoted by u_∞). Then

$$u_\infty = \phi_\xi(u_\infty)$$

We essentially want to solve $s = \phi_\xi(s)$. Recall that we have 0 children with $\frac{3}{4}$ probability and 2 with $\frac{1}{4}$ probability. So then

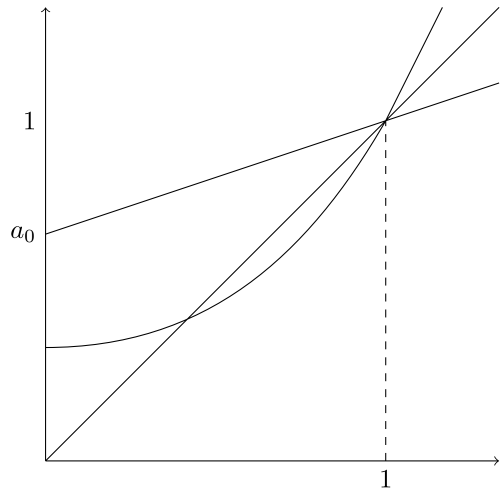
$$\phi_B(s) = \mathbb{E}\big[s^B\big] = \frac{3}{4} + \frac{1}{4}s^2$$

This tells us

$$\begin{aligned}
s &= \frac{3}{4} + \frac{1}{4}s^2 \\
s^2 - 4s + 3 &= 0 \\
s &= \{1, 3\}
\end{aligned}$$

This implies 1 is a fixed point.

Suppose now we switch the probabilities so that we have 0 children with $\frac{1}{4}$ probability and 2 children with $\frac{3}{4}$ probability. Doing the identical calculations, we have $s = \{1, \frac{1}{3}\}$.



Suppose we now have

$$g_N(s) = \sum_{k=0}^{\infty} s^k \mathbb{P}(N = k)$$

and

$$Q = \xi_1 + \xi_2 + \cdots + \xi_N$$

Then

$$\begin{aligned}
\sum_{k=0}^{\infty} s^k \mathbb{P}(Q = k) &= \sum_{k=0}^{\infty} s^k \left(\sum_{n=0}^{\infty} \mathbb{P}(Q = k | N = n)\, \mathbb{P}(N = n) \right) \\
&= \sum_{n=0}^{\infty} \sum_{n=0}^{\infty} s^k \mathbb{P}(Q = k | N = n)\, \mathbb{P}(N = n) \\
&= \sum_{n=0}^{\infty} \left(\sum_{k=0}^{\infty} s^k \mathbb{P}(\xi_1 + \cdots + \xi_n = k) \right) \mathbb{P}(N = n)
\end{aligned}$$

Note that we can interchange the summations since all terms are positive. We also have

$$\{\phi_\xi(s)\}^n = \sum_{k=0}^{\infty} s^k \mathbb{P}(\xi_1 + \cdots + \xi_n = k)$$

and so

$$\begin{aligned}
\sum_{k=0}^{\infty} s^k \mathbb{P}(Q = k) &= \sum_{n=0}^{\infty} \{\phi_\xi(s)\}^n \, \mathbb{P}(N = n) \\
&= g_N(\phi_\xi(s))
\end{aligned}$$

Now, if $X_{n+1} = \xi_1^{(n)} + \xi_2^{(n)} + \cdots + \xi_{X_n}^{(n)}$, then $g_N \equiv \phi_n$ and

$$\phi_{n+1}(s) + \phi_n(\phi(s))$$

12.3. **"Pure-Death" Process.** Suppose $\phi(s) = q + ps$ with $q + p = 1$. Then

$$\begin{aligned}
\phi_2(s) &= \phi(\phi(s)) \\
&= q + p(q + ps) \\
&= q + pq + p^2 s \\
&= (1 + p)(1 - p) + p^2 s \\
&= 1 - p^2 + p^2 s
\end{aligned}$$

Let us let T be a random time, which is when the population dies out. Let $X_0 = 1$.

$$\begin{aligned}
\mathbb{P}(T = n | X_0 = 1) &= \mathbb{P}(X_n = 0 | X_0 = 1) - \mathbb{P}(X_{n-1} = 0 | X_0 = 1) \\
&= \phi_n(0) - \phi_{n-1}(0)
\end{aligned}$$

If instead, $X_0 = k$, then

$$\mathbb{P}(T = n | X_0 = k) = \phi_n^{(k)}(0) - \phi_{n-1}^{(k)}(0)$$

13. March 24th, 2014

13.1. **Poisson Processes.** Recall that the probability mass function for a Poisson distribution is

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad k = 0, 1, 2, \ldots$$

If $X \sim \text{Pois}(\mu)$ and $Y \sim \text{Pois}(\lambda)$, with X and Y iid, then

$$X + Y \sim \text{Pois}(\mu + \lambda)$$

We can show this by using convolutions.

$$\begin{aligned}
\mathbb{P}(X + Y = n) &= \sum_{k=0}^{n} \mathbb{P}(X = k, Y = n - k) \\
&= \sum_{k=0}^{n} \mathbb{P}(X = k)\,\mathbb{P}(Y = n - k) \\
&= \sum_{k=0}^{n} \frac{\mu^k}{k!} e^{-\mu} \frac{\lambda^{n-k}}{(n-k)!} \\
&= \frac{e^{-(\mu+\lambda)}}{n!} \sum_{k=0}^{n} \frac{\mu^k \lambda^{n-k}}{k!(n-k)!} n! \\
&= \frac{e^{-(\mu+\lambda)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \mu^k \lambda^{n-k} \\
&= \frac{e^{-(\mu+\lambda)}}{n!} (\mu + \lambda)^n
\end{aligned}$$

This family is special in that it belongs to the family of infinitely divisible distributions, in that if we have n iid $\text{Pois}(\lambda)$ random variable, then we can obtain a $\text{Pois}(n\lambda)$ random variable. Similarly,

$$\text{Pois}(\lambda) \stackrel{d}{=} Y_1 + Y_2 + \cdots + Y_n$$

where $Y_i \sim \text{Pois}(\lambda/n)$.

Example 20. *Suppose $N \sim \text{Pois}(\lambda)$ and $M|N \sim \text{Bin}(N, p)$. Then what is the marginal distribution of M?*

$$M \sim \text{Pois}(\lambda p)$$

If we are told that M is Poisson, then we can immediately find its parameter since

$$\mathbb{E}[M] = \mathbb{E}[\mathbb{E}[M|N]] = \mathbb{E}[Np] = \lambda p$$

Definition 6 (Poisson Process)**.** *Let $N(t)$ (or $X(t)$) with intensity (or rate) $\lambda > 0$ be such that*

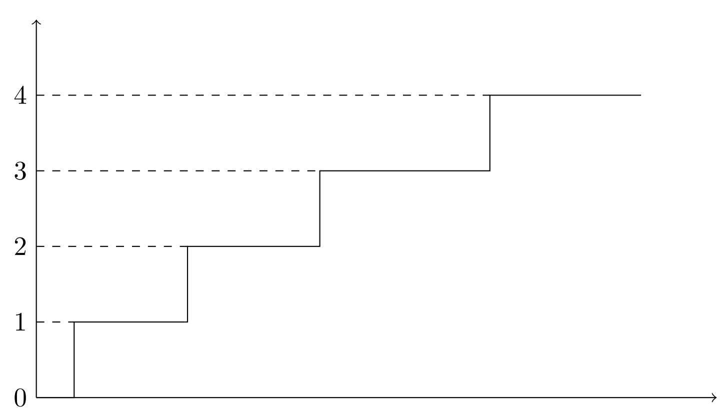
(1) $N(0) = 0$

(2) For any finite set of time intervals, $t_0, t_1, t_2, \ldots, t_n$,

$$N(t_1) - N(t_0), N(t_2) - N(t_1), \ldots, N(t_n) - N(t_{n-1})$$

are independent.

(3) For any $t, s > 0$, $N(t + s) - N(s) \sim \text{Pois}(\lambda t)$.



Note that we have

$$N(t+2) - N(t+1) \sim \text{Pois}(\lambda)$$
$$N(t+100) - N(t+99) \sim \text{Pois}(\lambda)$$

so we have stationary, independent increments.

How do we know the increments go up by 1? Note that

$$N(t+\mathrm{d}t) - N(t) \sim \text{Pois}(\lambda \mathrm{d}t)$$

A heuristic way to look at this is

$$\mathbb{P}(N(t+\mathrm{d}t) - N(t) \geq 2) \approx \lim_{\mathrm{d}t \to 0} \frac{1}{\mathrm{d}t} e^{-\lambda \mathrm{d}t} \left[\frac{(\lambda \mathrm{d}t)^2}{2!} + \frac{(\lambda \mathrm{d}t)^3}{3!} + \cdots \right] = 0$$

Typically in probability and mathematics, units are not important as they are in physics. However, they do have importance with Poisson processes. Here, λ is the rate. Suppose λ is 4 per hour. Suppose we start at $9:00$. At $9:30$, there has been 1 customer. At $11:30$, there have been 5 customers. Then we have

$$\mathbb{P}\left(N\left(\frac{1}{2}\right) = 1, N\left(\frac{5}{2}\right) - N\left(\frac{1}{2}\right) = 4\right)$$

We know $N\left(\frac{1}{2}\right) \sim \text{Pois}(2)$ and $N\left(\frac{5}{2}\right) - N\left(\frac{1}{2}\right) \sim \text{Pois}(8)$. By independence, we know that the probability above, has distribution

$$\text{Pois}(2) \cdot \text{Pois}(8)$$

13.1.1. *Law of Rare Events.* Recall that if we have n, p such that $n \to \infty$, $p \to 0$ with $np = c$, then

$$\text{Bin}(n, p) \approx \text{Pois}(c)$$

A good example to illustrate this is considering a highway with a large number of cars passing by. Each one have them has a low probability of getting into an accident.

Theorem 7. *Suppose $\xi_i \in \{0, 1\}$ and $\mathbb{P}(\xi_i = 1) = p_i$ where ξ_i are independent. Let $S_n = \xi_1 + \xi_2 + \cdots + \xi_n$. Then*

$$\left| \mathbb{P}(S_n = k) - \frac{\mu^k e^{-\mu}}{k!} \right| \leq \sum_{i=1}^{n} p_i^2$$

where $\mu = p_1 + p_2 + \cdots + p_n$.

Suppose $p_i = \frac{p}{n}$. Then

$$\sum_{i=1}^{n} p_i^2 \approx \frac{p^2}{n} \to 0$$

as $n \to \infty$. Using this as a heuristic, we can see why the densities are closely approximated.

Lemma (Coupling Lemma)**.** *For random variables X and Y (not necessarily iid), and Borel set B,*

$$|\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)| \leq \mathbb{P}(X \neq Y)$$

Proof. Note that

$$\{X \in B\} = \{X \in B, Y \in B\} \cup \{X \in B, Y \notin B\}$$

so

$$\mathbb{P}(X \in B) = \mathbb{P}(X \in B, Y \in B) + \mathbb{P}(X \in B, Y \notin B)$$
$$\mathbb{P}(Y \in B) = \mathbb{P}(Y \in B, X \in B) + \mathbb{P}(Y \in B, X \notin B)$$

Subtracting these, we have

$$|\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)| \leq \mathbb{P}(X \neq Y)$$

□

For the above theorem, we have $B = \{k\}$.

Proof of Theorem.

$$S_n = \xi(p_1) + \xi(p_2) + \cdots + \xi(p_n)$$
$$N = N(p_1) + N(p_2) + \cdots + N(p_n)$$

By the coupling lemma,

$$\begin{aligned}|\mathbb{P}(S_n = k) - \mathbb{P}(N = k)| &\le \mathbb{P}(S_n \neq N) \\ &\le \sum_{i=1}^{n} \mathbb{P}(\xi(p_i) \neq N(p_i))\end{aligned}$$

If we can show that for each i, we have

$$\mathbb{P}(\xi(p_i) \neq N(p_i)) \le p_i^2$$

then we are done.

Let $U \sim \mathcal{U}[0,1]$. For any p, we construct $\xi(p)$ such that

$$\mathbb{P}(\xi(p) \neq N(p)) \le p^2$$

Set $\xi(p) = 1$ if $U < p$. Otherwise, $\xi(p) = 0$. Set $N(p) = k$ if

$$\sum_{i=0}^{k-1} \frac{p^i e^{-p}}{i!} < U \le \sum_{i=0}^{k+1} \frac{p^i e^{-p}}{i!}$$

that is, we set

$$N(p) = \begin{cases} 0 & U \le e^{-p} \\ 1 & e^{-p} < U \le e^{-p} + pe^{-p} \\ 2 & e^{-p} + pe^{-p} < U \le e^{-p} + pe^{-p} + \frac{p^2}{2!}e^{-p} \\ \vdots \end{cases}$$

□

14. March 26th, 2014

NOTE: There is a typo in the book. In this lecture, $\varepsilon(p)$ and $X(p)$ are used instead of $\xi(p)$ and $N(p)$

To show that

$$\mathbb{P}(\varepsilon(p) \neq X(p)) \le p^2$$

we can alternatively show

$$\mathbb{P}(\varepsilon(p) = X(p)) \ge 1 - p^2$$

When do we have $\varepsilon(p) = X(p) = 0$ and $\varepsilon(p) = X(p) = 1$?. We know $\varepsilon(p) = 0$ if $0 < U < 1-p$ and $X(p) = 0$ if $p < U \le e^{-p}$. Using the fact that $1 - p \le e^{-p}$, we have

$$\mathbb{P}(\varepsilon(p) = X(p) = 0) = 1 - p$$

$$\mathbb{P}(\varepsilon(p) = X(p) = 1) = pe^{-p}$$

Then we have

$$\mathbb{P}(\varepsilon(p) = X(p)) = 1 + p + pe^{-p} = 1 - p(1 - e^{-p})$$

Rearranging the $1 - p \le e^{-p}$ again, we get $1 - e^{-p} \le p$, so that

$$\mathbb{P}(\varepsilon(p) = X(p)) = 1 - p(1 - e^{-p}) \ge 1 - p^2$$

14.1. Poisson Processes.

$X(t)$ is a Poisson process if

(1) $X(0) = 0$
(2) $X(t)$ has stationary, independent increments

and

$$X(t) - X(s) \sim \text{Pois}(\lambda(t-s))$$

Let $N[s,t)$ be the number of events in $[s,t)$.

(1) $N[t_1,t_2)$ and $N[t_3,t_4)$ are independent if $[t_1,t_2)$ is disjoint from $[t_3,t_4)$.
(2) The distribution of $N[t,t+h)$ only depends on h.
(3) There exists $\lambda > 0$ such that $\mathbb{P}(N[t,t+h) = 1) = \lambda h + o(h)$. The notation $o(h)$ means that if $s_h = o(h)$, then

$$\lim_{h \to 0} \frac{s_h}{h} = 0$$

(4) $\mathbb{P}(N[t,t+h) \ge 2) = o(h)$

Theorem 8. *$X(t) = N[0,t)$ is a Poisson process with $\lambda > 0$*

$$\varepsilon_i = \begin{cases} 1 & \text{if there is at least one event in the } i\text{th interval } \left[(i-1)\frac{t}{n}, \frac{it}{n}\right) \\ 0 & \text{otherwise} \end{cases}$$

Let

$$S_n = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n$$

We have

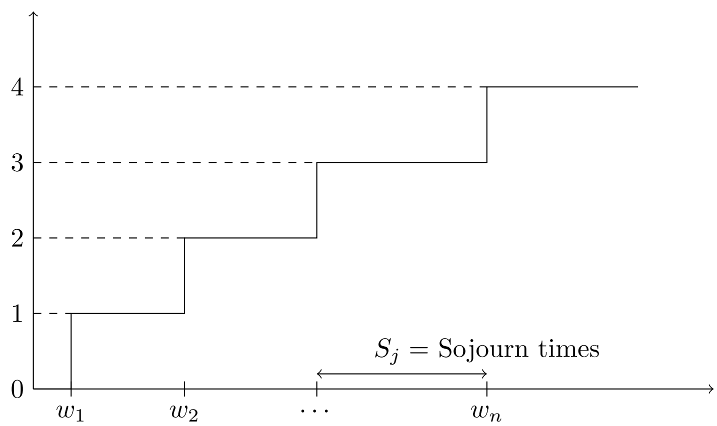
$$\mathbb{P}(\varepsilon_i = 1) = \lambda \frac{t}{n} + o\left(\frac{t}{n}\right)$$

By the law of rare events,

$$|\mathbb{P}(S_n = k) - \mathbb{P}(X(t) = k)| \le \sum \left(\lambda \frac{t}{n} + o\left(\frac{t}{n}\right)\right)^2 \longrightarrow 0 \text{ as } n \to \infty$$

The event $\{S_n \neq N[0,t)\}$ occurs if we have at least one interval in which we have 2 or more events. That is, we have

$$\begin{aligned} \mathbb{P}(S_n \neq N[0,t)) &= \sum_{i=1}^{n} \mathbb{P}\left(N\left[(i-1)\frac{t}{n}, \frac{it}{n}\right) \ge 2\right) \\ &= o\left(\frac{t}{n}\right) n \longrightarrow 0 \text{ as } n \to \infty \end{aligned}$$



(1) $f_{W_n}(w) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}$
(2) S_j are iid exponential, $f_{S_n}(s) = \lambda e^{-\lambda s}$.
(3) For $u < t$, $X(u) | X(t) = n \sim \text{Bin}\left(n, \frac{u}{t}\right)$
(4) $(W_1, W_2, \dots, W_n) | N(t) = n = \frac{t^n}{n!}$.

14.1.1. *"Semi-rigourous" proof.*

$$\begin{aligned} f_{W_n}(t) &= \mathbb{P}(W_n \in [t_n, t_n + \Delta t]) \\ &= \mathbb{P}(N[0,t) = n-1)\,\mathbb{P}(N[t, t+\Delta t) = 1) \\ &= \frac{1}{\Delta t} (\lambda t)^{n-1} \frac{e^{-\lambda t}}{(n-1)!} \lambda \Delta t \\ &= \lambda^n t^{n-1} \frac{e^{-\lambda t}}{(n-1)!} \end{aligned}$$

15. March 31st, 2014

15.1. **Poisson Processes: Distributions.**

15.1.1. *Distribution 1.*

$$f_{W_n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}$$

How would we link the waiting time to the Poisson process?

$$\mathbb{P}(W_n \le t) = \mathbb{P}(X(t) \ge n)$$

To show this, we have

$$
\begin{aligned}
\mathbb{P}(W_n \le t) &= 1 - \sum_{k=0}^{n-1} \mathbb{P}(X(t) = k) \\
&= 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!}
\end{aligned}
$$

Differentiate both sides to get

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}(W_n \le t) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!}\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left\{e^{-\lambda t}\left(1 + \lambda t + \frac{(\lambda t)^2}{2!} + \frac{(\lambda t)^3}{3!} + \cdots + \frac{(\lambda t)^{n-1}}{(n-1)!}\right)\right\} \\
&= e^{-\lambda t}\left(\lambda + 2\lambda\frac{\lambda t}{2!} + 3\lambda\frac{(\lambda t)^2}{3!} + \cdots + (n-1)\lambda\frac{(\lambda t)^{n-2}}{(n-1)!}\right) \\
&\quad - \lambda e^{-\lambda t}\left(1 + \lambda t + \frac{(\lambda t)^2}{2!} + \cdots + \frac{(\lambda t)^{n-1}}{(n-1)!}\right) \\
&= \frac{\lambda(\lambda t)^{n-1}}{(n-1)!}e^{-\lambda t}
\end{aligned}
$$

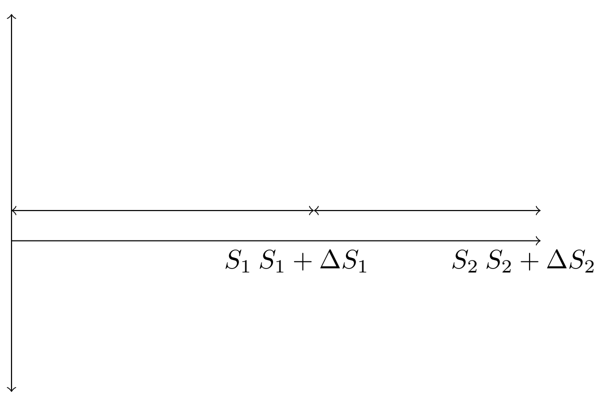
15.1.2. *Distribution 2.* Now if we look at the Sojourn times $S_1, S_2, \ldots, S_n$, we have

$$f_{S_n}(s_n) = \lambda e^{-\lambda s_n}$$

so that the joint distribution is

$$f_{S_1, S_2, \ldots, S_n}(s_1, s_2, \ldots, s_n) = (\lambda e^{-\lambda s_1})(\lambda e^{-\lambda s_2}) \cdots (\lambda e^{-\lambda s_n})$$

Proof. Suppose $n = 2$.



Then we can consider the intervals

$$
\begin{aligned}
&\underbrace{[0, S_1)}_{=0}\underbrace{[S_1, S_1 + \Delta S_1)}_{=1}\underbrace{[S_1 + \Delta S_1, S_1 + \Delta S_1 + S)}_{=0}\underbrace{[S_1 + \Delta S_1 + S_2, S_1 + \Delta S_1 + S_2 + \Delta S_2)}_{=1} \\
&= e^{-\lambda s_1} \cdot \lambda \Delta s_1 e^{-\lambda \Delta s_1} \cdot e^{-\lambda s_2} \cdot \lambda \Delta s_2 e^{-\lambda \Delta s_2}
\end{aligned}
$$

So that

$$
\begin{aligned}
f_{S_1, S_2}(s_1, s_2) &= \lim_{\Delta s_1 \to 0, \Delta s_2 \to 0} (\lambda e^{-\lambda s_1})(\lambda e^{-\lambda s_2}) \Delta s_1 \Delta s_2 e^{-\lambda s_1} e^{-\lambda s_2} \\
&= \lim_{\Delta s_1 \to 0, \Delta s_2 \to 0} \Delta s_1 \Delta s_2 (1 + \lambda \Delta s_1 \Delta s_2 + \cdots) \\
&= (\lambda e^{-\lambda s_2})(\lambda e^{-\lambda s_2})
\end{aligned}
$$

$\square$

15.1.3. *Distribution 3.* The third distribution is

$$\mathbb{P}(X(u) = k | X(t) = n)$$

where $u < t$. We have

$$\begin{aligned}
\mathbb{P}(X(u) = k | X(t) = n) &= \frac{\mathbb{P}(X(u) = k, X(t) = n)}{\mathbb{P}(X(t) = n)} \\
&= \frac{\mathbb{P}(X(u) = k)\,\mathbb{P}(X(t) - X(u) = n - k)}{\mathbb{P}(X(t) = n)} \\
&= \frac{(\lambda u)^k e^{-\lambda u} \left\{\lambda(t-u)\right\}^k e^{-\lambda(t-u)}}{k! e^{-\lambda t} \frac{(\lambda t)^n}{n!} (n-k)!} \\
&= \frac{n!}{(n-k)!k!} \left(\frac{u}{t}\right)^k \left(1 - \frac{u}{t}\right)^{n-k} \\
&\sim \text{Bin}\left(n, \frac{u}{t}\right)
\end{aligned}$$

15.2. **Order Statistics.** Suppose $U_1, U_2 \overset{iid}{\sim} \mathcal{U}[0,1]$. Sort these in increasing order to get (W_1, W_2). We know

$$f_{U_1, U_2, \ldots, U_n}(u_1, u_2, \ldots, u_n) = \frac{1}{t^n}$$

Then

$$f_{W_1, \ldots, W_n}(w_1, \ldots, w_n) = \frac{n!}{t^n}$$

Let us illustrate this with the $n = 2$ case. We know

$$\mathbb{P}(w_1 \le W_1 \le w_2, w_2 \le W_2 \le 1) = \mathbb{P}(w_1 \le U_1 \le w_2, w_2 \le U_2 \le 1) + \mathbb{P}(w_1 \le U_2 \le w_2, w_2 \le U_1 \le 1)$$

since U_1 and U_2 are exchangeable because of their iid nature.

15.2.1. *Distribution 4.*

$$f_{W_1, \ldots, W_n}(w_1, \ldots, w_n) = \frac{n!}{t^n}$$

By the "random variable perspective",

$$\begin{aligned}
&\underbrace{[0, w_1)}_{=0} \underbrace{[w_1, w_1 + \Delta w_1)}_{=1} \underbrace{[w_1 + \Delta w_1, w_1 + \Delta w_1 + w_2)}_{=0} \\
&= e^{-\lambda w_1} \cdot \lambda e^{-\lambda \Delta w_1} \cdot e^{-\lambda w_2} \cdots \\
&= \frac{e^{-\lambda t}}{\mathbb{P}(X(t) = n)} (\lambda \Delta w_1)(\lambda \Delta w_2) \cdots (\lambda \Delta w_n) \\
&= \frac{e^{-\lambda t}}{e^{-\lambda t} \frac{(\lambda t)^n}{n!}} \lambda^n \\
&= \frac{n!}{t^n}
\end{aligned}$$

How would we calculate

$$\mathbb{E}\left[e^{-\beta \sum_{k=1}^{X(t)} W_k}\right]$$

We can condition on the upper sum. Then we can use

$$\mathbb{E}\left[e^{-\beta \sum_{k=1}^{n} W_k} \middle| X(t) = n\right] = \mathbb{E}\left[e^{-\beta \sum_{k=1}^{n} U_k}\right]$$

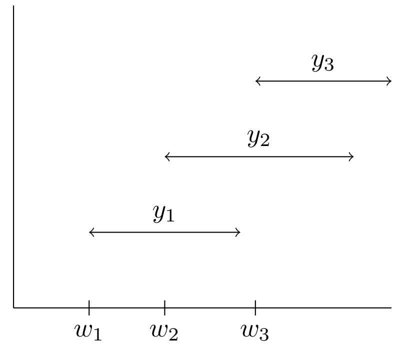
where $U_k \sim \mathcal{N}[0,1]$. We can solve this using

$$\int_0^1 e^{-\beta u}\, \mathrm{d}u$$

Then solve the entire thing using

$$\mathbb{E}\left[e^{-\beta \sum_{k=1}^{X(t)} W_k}\right] = \sum_{n=0}^{\infty} \mathbb{E}\left[e^{-\beta \sum_{k=1}^{n} W_k} \middle| X(t) = n\right] \mathbb{P}(X(t) = n)$$

Example 21. *Suppose we have particles which have lifetime determined by $G(t) = \mathbb{P}(Y_k \le y)$.*



Let $M(t)$ be the number of particles alive at t conditional on $X(t) = n$. We have $M(t) \leq X(t)$. The kth particle is dead is $\{w_k + y_k \geq t\}$. Then

$$\begin{aligned}
\mathbb{P}(M(t) = m | X(t) = n) &= \mathbb{P}\left(\sum_{k=1}^{n} \mathbb{1}_{\{W_k + Y_k \geq y\}} = m \middle| X(t) = n \right) \\
&= \mathbb{P}\left(\sum_{k=1}^{n} \mathbb{1}_{\{U_k + Y_k \geq y\}} = m \right) \\
&= \binom{n}{m} p^m (1-p)^{n-m}
\end{aligned}$$

We also have

$$\begin{aligned}
\mathbb{P}(Y_k + U_k \geq t) &= \frac{1}{t} \int_0^t \mathbb{P}(Y_k \geq t - u) \, \mathrm{d}u \\
&= \frac{1}{t} \int_0^t \{1 - G(t-u)\} \, \mathrm{d}u \\
&=: p
\end{aligned}$$

Now, we have

$$\begin{aligned}
\mathbb{P}(M(t) = m) &= \sum_{n=m}^{\infty} \mathbb{P}(M(t) = m | X(t) = n) \, \mathbb{P}(X(t) = n) \\
&= \sum_{n=m}^{\infty} \frac{n!}{(n-m)!m!} p^m (1-p)^{n-m} (\lambda t)^n \frac{e^{-\lambda t}}{n!} \\
&= e^{-\lambda p t} \frac{(\lambda p t)^m}{m!}
\end{aligned}$$

Check that this last step works. Note that since

$$p = \frac{1}{t} \int_0^t \{1 - G(t-u)\} \, \mathrm{d}u$$

then

$$\begin{aligned}
\lambda p t &= \lambda \int_0^t \{1 - G(t-u)\} \, \mathrm{d}u \\
&= \lambda \int_0^t \{1 - G(u)\} \, \mathrm{d}u \qquad \textit{by change of variables}
\end{aligned}$$

So, if we take $t \to \infty$,

$$M(t) \sim \mathrm{Pois}\left(\lambda \int_0^t \{1 - G(u)\} \, \mathrm{d}u \right)$$

and we know that

$$\mathbb{E}[Y_k] = \int_0^\infty \mathbb{P}(Y_k \geq u) \, \mathrm{d}u$$

16. April 2nd, 2014

Charles Stein proved that if

$$\mathbb{E}[\mathcal{A} f(W)] = 0$$

then W is Poisson. He also showed that if $W \approx \mathrm{Pois}$, then

$$\mathbb{E}[\mathcal{A} f(W)] \approx 0$$

For Normality, we had a similar result. If $X \sim \mathcal{N}(0,1)$, then

$$\mathbb{E}[f'(X)] = \mathbb{E}[Xf(X)]$$

so that

$$\mathbb{E}[f'(X) - Xf(X)] = 0$$

Here, we would set

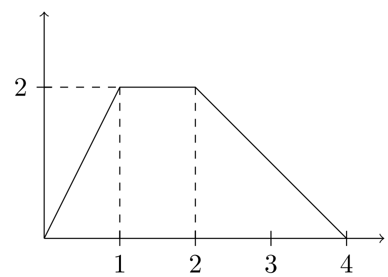
$$\mathcal{A}f = f'(x) - xf(x)$$

16.1. **Time-inhomoegeneous Poisson Processes.** In a time-inhomogeneous Poisson process, the rate λ is not constant as it was in the time-homogeneous case. Let

$$\lambda(t) = \begin{cases} 2t & 0 \le t < 1 \\ 2 & 1 \le t < 2 \\ 4-t & 2 < t \le 4 \end{cases}$$

In a time-homoegeneous case, we have

$$\mathbb{P}(X(t+h) - X(t) = 1) \approx \lambda h + o(h) \approx \lambda h$$



Then

$$\mathbb{P}(X(2) = 2, X(4) = 4) = \mathbb{P}(X(2) = 2)\,\mathbb{P}(X(4) - X(2) = 2)$$

and so

$$m_1 = \int_0^2 \lambda(u)\,\mathrm{d}u = \int_0^1 2t\,\mathrm{d}t + \int_1^2 2\,\mathrm{d}u = 3$$

which implies

$$\mathbb{P}(X(2) = 2) = \frac{3^2 e^{-3}}{2!}$$

We also have

$$m_2 = \int_2^4 (4-t)\,\mathrm{d}t = 2$$

so

$$\mathbb{P}(X(4) - X(2) = 2) = \frac{2^2 e^{-2}}{2!}$$

We integrate here to determine the rate since in the homogeneous case, the parameter of the Poisson process is

$$\lambda s = \int_t^{t+s} \lambda\,\mathrm{d}u$$

so it is natural to use

$$\int_t^{t+s} \lambda(u)\,\mathrm{d}u$$

How would we simulate a Poisson process? The easiest way would be to simulate iid Exponential random variables with the desired λ as the parameter, which would then become the waiting times. Why do we choose the same λ in the waiting times as in the Poisson process? This is easy to see if $\lambda = 2$. Then for a Poisson process with this rate, we expect to see 2 observations in the interval $[0, 1)$. This is precisely what we would get if the waiting times had expectation $\frac{1}{2}$, which is when $\lambda = 2$.

Example 22. *Suppose $Y \sim \mathrm{Pois}(\lambda(u))$. Let*

$$S = \Lambda(t) = \int_0^t \lambda(u)\,\mathrm{d}u$$

Theorem 9.

$$Z(\Lambda(t)) = X(t) \Rightarrow Z(t) = X(\Lambda^{-1}(t))$$

16.2. **Yule Process.** Let $\beta > 0$ be the rate. We call that we have

$$\mathbb{P}(X(t+h) - X(t) = 1) \approx \beta h + o(h)$$

Then

$$\mathbb{P}(X(t+h) - X(t) = 1 | X(t) = n) = \binom{n}{1}(\beta h + o(h))(1 - \beta(h) - o(h))^{n-1} = n\beta h + o(h)$$

16.3. **Birth-Death Process.** Assume $X(0) = 0$. Then we have

(1) $\mathbb{P}(X(t+h) - X(t) = 1 | X(t) = k) = \lambda_k h + o(h)$
(2) $\mathbb{P}(X(t+h) - X(t) = 0 | X(t) = k) = 1 - \lambda_k h + o(h)$
(3) $\mathbb{P}(X(t+h) - X(t) < 0 | X(t) = k) = 0$

Let $P_n(t) = \mathbb{P}(X(t) = n)$.

Lemma. *We have*

$$\begin{cases} P_0'(t) = -\lambda_0 P_0(t) \\ P_n'(t) = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t) \end{cases}$$

with boundary conditions

$$\begin{aligned} P_0(0) &= \mathbb{P}(X(0) = 0) = 1 \\ P_n(0) &= \mathbb{P}(X(0) = n) = 0 \qquad \textit{for } n \geq 1 \end{aligned}$$

Proof.

$$P_n'(t) = \lim_{h \to 0} \frac{P_n(t+h) - P_n(t)}{h}$$

We have

$$\begin{aligned} P_n(t+h) &= \mathbb{P}(X(t+h) = n) \\ &= \sum_{k=0}^{n} \underbrace{\mathbb{P}(X(t+h) = n | X(t) = k)}_{\mathbb{P}(X(t+h) - X(t) = n-k)} \underbrace{\mathbb{P}(X(t) = k)}_{P_k(t)} \end{aligned}$$

For $k = 0, 1, 2, \dots, n-1$, $P_k(t) = o(h)$

$$P_n(t+h) = (1 - \lambda_n h) P_n(t) + \lambda_{n-1} h P_{n-1}(t) + o(h)$$

which implies

$$\frac{P_n(t+h) - P_n(t)}{h} = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t)$$

□

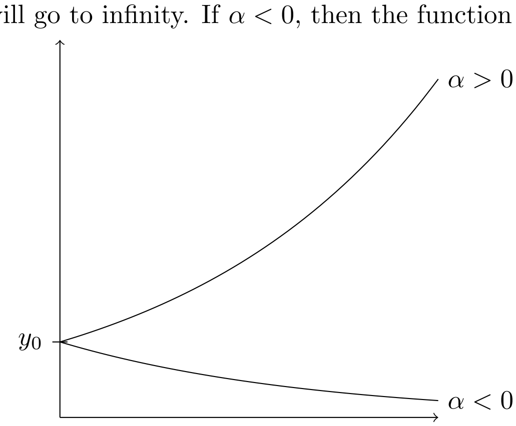
Suppose we have a differential equation of the form

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = \alpha y$$

Then we have

$$y(t) = y_0 e^{\alpha t}$$

Note that if $\alpha > 0$, then the function will go to infinity. If $\alpha < 0$, then the function will decay to 0.



Suppose $P_n'(t) = -\beta n P_n(t) + \beta(n-1)P_{n-1}(t)$. Then

$$P_n(t) = e^{-\beta t}(1 - e^{-\beta t})^{n-1}$$

for $n \geq 1$. so

$$\mathbb{P}(X(t) = n) = e^{-\beta t}$$

17. April 7th, 2014

17.1. Pure Birth Process.

(1) $\mathbb{P}(X(t+h) = k+1 | X(t) = k) = \lambda_k h + o(h)$
(2) $\mathbb{P}(X(t+h) = k | X(t) = k) = 1 - \lambda_k(h) + o(h)$
(3) $\mathbb{P}(X(t+h) < X(t) | X(t) = k)$
(4) $X(0) = 0$ (This is not always an assumption) This implies $X(t) = N[0, t)$.

$$\begin{tikzcd}
 & \text{Continuous Time Markov Chains} \\
\text{Poisson Processes} \arrow[ur] \arrow[dr] & \\
 & \text{Renewal Processes}
\end{tikzcd}$$

Last time, we had that

$$\begin{aligned}
P_n(t) &= \mathbb{P}(X(t) = n) \\
P_n'(t) &= -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t) \\
P_0'(t) &= -\lambda_0 P_0(t)
\end{aligned}$$

If we wish to impose the condition $X(0) = 0$, then we set $P_0(0) = 1$.

For the Yule process case, if you start with $P_0(1) = 1$, and $\lambda_k = \beta k$, then

$$P_n(t) = e^{-\beta t}(1 - e^{-\beta} t)^{n-1}$$

Define

$$Q_n(t) = e^{\lambda_n t} P_n(t)$$

where $P_0(0) = 1$. Differentiating, we get

$$Q_n'(t) = \lambda_n e^{\lambda_n t} P_n(t) + e^{\lambda_n t}(-\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t)$$

and so

$$Q_n'(t) = \lambda_{n-1} e^{\lambda_n t} P_{n-1}(t)$$

which implies

$$Q_n(t) = \int_0^t \lambda_n e^{-\lambda_n x} P_{n-1}(x) \, \mathrm{d}x$$

Suppose we have the differential equation

$$\mathrm{d}y(t) = y^2(t)$$

where $y(0) = 1$. The solution to this is

$$y(t) = \frac{1}{1-t}$$

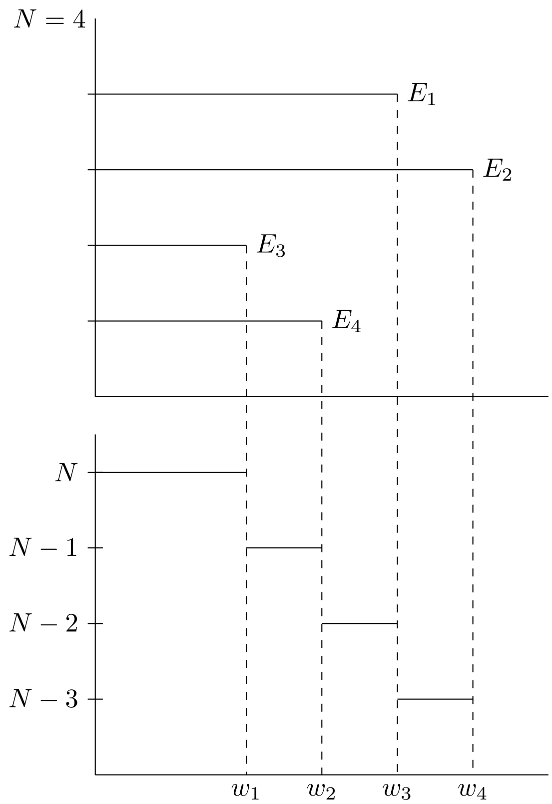
We verify that

$$y'(t) = \frac{1}{1-t^2}$$

Note that this blows up at $t = 1$. What is the analog in the pure birth process case? $\frac{1}{\lambda_k}$ is the expected amount of time for the kth particle to come. Then

$$\sum_{k=1}^{\infty} \frac{1}{\lambda_k}$$

is the expected amount of time for all particles to come. If this is less than ∞, then the expected amount of time you need to wait for all the particles to come is finite.



Here, the w_i's are all exponentially distributed due to the memoryless property. This is **pure-death process**. Now we have
(1) $\mathbb{P}(X(t+h) - X(t) = 1 | X(t) = k) = -\lambda_k h$
Suppose X and Y are independent Exponential random variables with rates μ and λ respectively. Let $Z = \min(X, Y)$. Then

$$\begin{aligned}\mathbb{P}(Z > u) &= \mathbb{P}(X > u, Y > u)\\ &= \mathbb{P}(X > u)\,\mathbb{P}(Y > u)\\ &= e^{-\mu u}e^{-\lambda u}\\ &= e^{-(\lambda+\mu)u}\end{aligned}$$

so Z is Exponential with rate parameter $\lambda + \mu$. So we have

$$\mathbb{P}(X(t+h) - X(t) = 1 | X(t) = k) = -\lambda_k h = -k\lambda h$$

What is the link between pure-death processes and pure-birth processes? Recall the transition matrix

$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \end{array}$$

We should define a continuous time finite state-space Markov chain is through a matrix Q, which is the generator or infinitesimal matrix. This represents the infinitesimal probability of going from one state to another.

$$Q = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix} \end{array}$$

Suppose you are in state 1. Then you would stay in state 1 for an Exponential amount of time with rate α. If you are in state 2, you would stay in state 2 for an Exponential amount of time with rate β. Then

$$\mathbb{P}(X(t+h) = 2 | X(t) = 1) \approx \alpha e^{-\alpha h}$$

Suppose we have states $1, 2, \ldots, N$. In the discrete time Markov chain, we had $P_{ij}^{(n)}$, the probability of going from state i to j in n steps. In a continuous Markov chain, we would have $P_{ij}(t)$, the probability of going from i to j in time t. Note that

$$P_{ij}(t+s) = \sum_{k=1}^{N} P_{ik}(t) P_{kj}(s)$$

Then, we have that $P(t)$ is a matrix for any t. Note that from the above equation,

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

This is known as the Kolmogorov-Chapman equation.

Suppose

$$Q = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} -5 & 2 & 3 \\ 3 & -7 & 4 \\ 1 & 2 & -3 \end{pmatrix} \end{array}$$

In the case of a 2×2 matrix,

$$P_{11}(0) = P_{22}(0) = 1$$
$$P_{12}(0) = P_{21}(0) = 0$$

so $P(0) = I$. The fact that the rows of Q sum to 0 is given by the fact that

$$P(h) - I = Qh + o(h)$$

When h is small, then

$$P(h) - P(0) \approx Qh$$

which gives

$$\lim_{h \to \infty} \frac{P(h) - P(0)}{h} = Q$$

Now, we have that

$$\begin{aligned} P(t+h) - P(t) &= P(t)P(h) - P(t) \\ &= P(t)(P(h) - I) \end{aligned}$$

If we divide both sides by h and take the limit as $h \to 0$, we have

$$\lim_{h \to 0} \frac{P(t+h) - P(h)}{h} = \lim_{h \to 0} P(t) \frac{P(h) - I}{h}$$

This gives us that

$$P'(t) = P(t)Q$$

and so

$$P(t) = e^{Qt}$$

Now what exactly does this mean? (since Q is a matrix). Recall

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \cdots$$

so for a matrix A, we have

$$\begin{aligned} \exp(A) &= I + A + \frac{A^2}{2!} + \cdots \\ &= \sum_{k=0}^{\infty} \frac{A^k}{k!} \end{aligned}$$

Now, if

$$D = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}$$

then

$$\exp(D) = \begin{bmatrix} e^{-d_1} & 0 \\ 0 & e^{-d_2} \end{bmatrix}$$

But if A is symmetric, then $A = UDU'$ which implies

$$A^K = UD^K U'$$

Putting this into the Taylor Series, we get

$$\exp(A) = U \exp(D) U'$$

Suppose we have a rate matrix

$$Q = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}$$

The square of this is

$$Q^2 = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} = \begin{bmatrix} (\alpha+\beta)\alpha & -(\alpha+\beta)\alpha \\ (\alpha+\beta)(-\beta) & (\alpha+\beta)\beta \end{bmatrix}$$

and so

$$Q^2 = -(\alpha + \beta) Q$$

By induction,

$$Q^n = -(\alpha+\beta)^{n-1} Q$$

Using this, we have

$$\exp(Qt) = I + \sum_{k=1}^{\infty} \frac{(Qt)^k}{k!}$$

where the sum is

$$\begin{aligned}\sum_{k=1}^{\infty} \frac{(-(\alpha+\beta))^k}{\alpha+\beta} \frac{t^k}{k!} Q &= -\frac{Q}{\alpha+\beta} \sum_{k=1}^{\infty} \frac{(-(\alpha+\beta)t)^k}{k!} \\ &= -\frac{Q}{\alpha+\beta}\left(e^{-(\alpha+\beta)t} - 1\right)\end{aligned}$$

and so

$$P(t) = \exp(Qt) = I + \frac{Q}{\alpha+\beta} - \frac{Q}{\alpha+\beta} e^{-(\alpha+\beta)t}$$

Note that

$$\lim_{t\to\infty} = I + \frac{Q}{\alpha+\beta}$$

Note that

$$I + \frac{Q}{\alpha+\beta} = \begin{bmatrix} 1 + \frac{\alpha}{\alpha+\beta} & -\frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & 1 - \frac{\alpha}{\alpha+\beta} \end{bmatrix}$$

Hence we can solve the relation

$$\lim_{t\to\infty} P'(t) = \lim_{t\to\infty} P(t) Q$$

which means

$$0 = \pi Q$$

18. April 9th, 2014

If $X \sim \text{Expo}(\mu)$ and $Y \sim \text{Expo}(\lambda)$, then what is $\mathbb{P}(X < Y)$?

$$\begin{aligned}\mathbb{P}(X<Y) &= \int_0^{\infty} \mathbb{P}(X>y)\, f_Y(y)\,\mathrm{d}y \\ &= \int_0^{\infty} e^{-\mu y} \lambda e^{-\lambda y}\,\mathrm{d}y \\ &= \frac{\lambda}{\lambda+\mu}\end{aligned}$$

This is important in determining what state a continuous Poisson process switches to.

18.1. **Renewal Process.** Let $X_i \overset{iis}{\sim} f$, $\mathbb{P}(X \le x) = F(x)$ with $X_i > 0$. As before, let

$$W_n = \sum_{i=1}^{n} X_i$$

be the waiting time. The renewal process

$$N(t) = N[0,t)$$

is the number of events before time t. Also, we have

$$\mathbb{E}[N(t)] = M(t)$$

which is the expected number of events. In a Poisson process, this is equal to λt. How would we find

$$\mathbb{P}(W_n \le x) = F_n(x)$$

We can write this recursively,

$$\begin{aligned}F_n(x) &= \int_0^{\infty} F_{n-1}(x-y)\,\mathrm{d}F(y) \\ &= \int_0^{x} F_{n-1}(x-y)\,\mathrm{d}F(y)\end{aligned}$$

Fundamental Relationship:

$$N(t) \ge k \text{ if and only if } W_k \le t$$

This tell us that

$$\begin{aligned}\mathbb{P}(N(t) \geq k) &= \mathbb{P}(W_k \leq t) \\ &= F_k(t)\end{aligned}$$

where $t \geq 0$ and $k = 1, 2, \ldots$. Immediately, we get

$$\begin{aligned}\mathbb{P}(N(t) = k) &= \mathbb{P}(N(t) \geq k) - \mathbb{P}(N(t) \geq k+1) \\ &= F_k(t) - F_{k+1}(t)\end{aligned}$$

Then we have

$$\begin{aligned}\mathbb{E}[N(t)] &= \sum_{k=1}^{\infty} \mathbb{P}(N(t) \geq k) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(W_k \leq t) \\ &= \sum_{k=1}^{\infty} F_k(t)\end{aligned}$$

Example 23. *Suppose* $S = \sum_{i=1}^{N} X_i$. *Then*

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}[N]\,\mathbb{E}[X_1] \\ &= \mu\mathbb{E}[X_1]\end{aligned}$$

Recall that $W_k = \sum_{i=1}^{k} X_i$

Theorem 10.

$$\begin{aligned}\mathbb{E}\big[W_{N(t)+1}\big] &= \mathbb{E}[X_1]\,\mathbb{E}[N(t)+1] \\ &= \mathbb{E}[X_1]\,(M(t)+1)\end{aligned}$$

What is so surprising about this result is that we do not have independence here, but the result still holds

Proof. We know that

$$N(t) \geq j-1 \Leftrightarrow X_1 + X_2 + \ldots X_{j-1} \leq t$$

We can leverage the fact that the X_i are independent.

$$\begin{aligned}\mathbb{E}[X_j \mathbb{1}\{X_1 + X_2 + \cdots + X_{j-1} \leq t\}] &= \mu\mathbb{P}(X_1 + X_2 + \cdots + X_{j-1} \leq t) \\ &= \mu\mathbb{P}(X_{j-1} \leq t) \\ &= \mu F_{j-1}(t)\end{aligned}$$

This gives

$$\begin{aligned}\mathbb{E}\big[W_{N(t)+1}\big] &= \mathbb{E}\left[\sum_{i=1}^{N(t)+1} X_i\right] \\ &= \mathbb{E}[X_1] + \mathbb{E}\left[\sum_{i=2}^{N(t)+1} X_i\right]\end{aligned}$$

Now, we can write

$$\begin{aligned}\mathbb{E}\left[\sum_{i=2}^{N(t)+1} X_i\right] &= \mathbb{E}\left[\sum_{i=2}^{\infty} X_i \cdot \mathbb{1}\{N(t)+1 \geq i\}\right] \\ &= \mathbb{E}\left[\sum_{i=2}^{\infty} X_i \cdot \mathbb{1}\{N(t) \geq i-1\}\right] \\ &= \mu \sum_{i=2}^{\infty} F_{i-1}(t)\end{aligned}$$

Hence

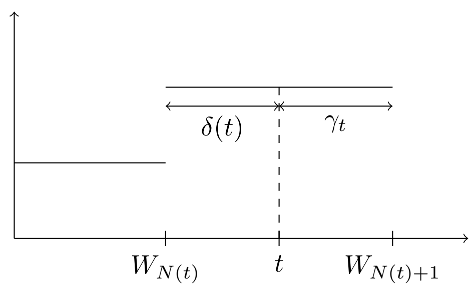
$$\begin{aligned}\mathbb{E}\big[W_{N(t)+1}\big] &= \mu + \mu\sum_{i=2}^{\infty} F_{i-1}(t)\\ &= \mu + \mu N(t)\\ &= \mu(N(t)+1)\end{aligned}$$

□

Suppose you are running a Metropolis-Hastings algorithm with target distribution $\mathcal{N}(0,1)$. Suppose your samples are $X_1, \ldots, X_n$. Then what are the estimates for the population mean and variance?

$$\begin{aligned}\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} X_i\\ \hat{\sigma}^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\end{aligned}$$

You will see that $\hat{\mu}$ will be close to 0, but $\hat{\sigma}^2$ will be slightly larger than 1. We would expect that the Markov chain here would converge to the proper distribution. However, the issue we have is that the Metropolis-Hastings algorithm will give correlated values.



where δ_t is called the **current age** and γ_t is called the **excess life**.

Suppose this is a Poisson process. Then what is the distribution of γ_t?

$$\begin{aligned}\mathbb{P}(\gamma_t > x) &= \mathbb{P}(N(t+x) - N(t) = 0)\\ &= e^{-\lambda x}\end{aligned}$$

Similarly,

$$\mathbb{P}(\delta_t < x) = \begin{cases} 1 - e^{-\lambda x} & x < t\\ 1 & x \geq t\end{cases}$$

Now, define

$$B_t = \delta_t + \gamma_t$$

which is the total life.

$$\begin{aligned}\mathbb{E}[B_T] &= \mathbb{E}[\delta_t] + \mathbb{E}[\gamma_t]\\ &= \int_0^t \mathbb{P}(\delta_t > x) + \frac{1}{\lambda}\\ &= \frac{1}{\lambda} + \frac{1}{\lambda}(1 - e^{-\lambda t})\end{aligned}$$

However, this does not make sense since B_t should be exponential. This is known as the length-based sampling paradox.

Example 24. *Suppose you have patients with pancreatic cancer in a hospital and you wish to find an estimate of the amount of time they have left. If you ask all the patients how long they have each been in the hospital and average that number, the result will be biased since it already takes into account the fact that the patients have survived up until that point. It is unlikely to find patients who only survived for two days.*

Example 25 (Cancer)**.** *Suppose there are two kinds of tumours: One that is fast growing and another that is slow growing, but does not show symptoms until the very end.*

If you screen people for cancer, you are then more likely to find people with a slow-growing tumour as opposed to ones with the fast-growing tumour.

18.2. **Sum-Quota Sampling.** Suppose you an engineer and you want to estimate how long a certain part lives. The first one lasts 9 months, the second 5 months, the third 7 months, and the fourth one is still lasting. A naive estimate of the lifetime here would be

$$\frac{9+5+7}{3} = 7 \text{ months}$$

that is our estimate is

$$\frac{W_{N(t)}}{N(t)}$$

We have

$$\mathbb{E}\left[\frac{W_{N(t)}}{N(t)}\middle|N(t)>0\right] = \sum_{n=1}^{\infty}\mathbb{E}\left[\frac{W_n}{n}\middle|N(t)=n\right]\mathbb{P}(N(t)=n)$$

Now, recall

$$\mathbb{E}[W_n|N(t)=n] = \mathbb{E}[\max\{U_1,\dots,U_n\}] = \frac{nt}{n+1}$$

Putting these together, we get

$$\mathbb{E}\left[\frac{W_{N(t)}}{N(t)}\middle|N(t)>0\right] = \frac{1}{\lambda}\left(1+\frac{\lambda t}{e^{-\lambda t}-1}\right)$$

Note that if t goes to infinity, the bias disappears (i.e. it gives us $\frac{1}{\lambda}$.)

19. April 16th, 2014

19.1. **Sum-Quota Sampling.** Find failure times

$$X_1,\dots,X_{n_1}$$

for $n \le T$. Instead of taking the sample average to estimate the mean, we have

$$\bar{X} = \frac{X_{N(t)}}{N(t)}$$

Recall that

$$\mathbb{E}[\bar{X}] = \frac{1}{\lambda}\left(1-\frac{\lambda T}{e^{\lambda T}-1}\right)$$

From last time, we have that

$$M(t) = \mathbb{E}[N(t)]$$

where $M(t) = \lambda t$ for Poisson processes. This gives us

$$\frac{M(t)}{t} = \frac{1}{\lambda} = \frac{1}{\mathbb{E}[X_1]}$$

Hence we have

$$\lim_{t\to\infty}\frac{M(t)}{t} = \frac{1}{\mu} \qquad \mu = \mathbb{E}[X_1]$$

Theorem 11. *For any renewal process,*

$$\lim_{t\to\infty}\frac{M(t)}{t} = \frac{1}{\mu}$$

$M(t)$ is the expected number of points in $(0,t]$. If you divide this by t, then this gives the expected amount of time for one to come, so this result makes sense. It turns out that this result is very hard to prove, so we will just accept it to be true.

19.1.1. *"Naive Renewal".* Suppose you have $\sum_k X_k$, where the X_k are the inter-arrival times. Suppose $X_k = 1$ for all k, that is there is no randomness. Then

$$N(t) = \begin{cases} 0 & 0 \le t < 1 \\ 1 & 1 \le t < 2 \end{cases}$$

In particular,

$$M(t) = N(t) = \lfloor t \rfloor$$

The renewal theorem says

$$\lim_{t\to\infty}\frac{M(t)}{t} = 1$$

that is

$$\lim_{t\to\infty}\frac{\lfloor t\rfloor}{t} = 1$$

An interesting question is what is $M(t)-t$? This is equal to $\lfloor t\rfloor - t$, which is between -1 and 0. Hence, we see that $M(t)$ is certainly not linear in t.

Theorem 12. *For any $h > 0$,*

$$\frac{M(t+h) - M(t)}{h} \longrightarrow \frac{1}{\mu}$$

as $t \to \infty$.

We just saw that

$$\lim_{t\to\infty} \frac{\mathbb{E}[N(t)]}{t} = \frac{1}{\mu}$$

If σ^2 is the variance of the renewal arrival times, then

$$\lim_{t\to\infty} \frac{\mathrm{Var}(N(t))}{t} = \frac{\sigma^2}{\mu^3}$$

We get an interesting result if we standardize this:

Theorem 13.

$$\frac{N(t) - \frac{t}{\mu}}{\sqrt{t\frac{\sigma^2}{\mu^3}}} \xrightarrow{D} \mathcal{N}(0,1)$$

This is interesting since a Normal random variables comes out of nowhere. It can also be shown that

$$\lim_{t\to\infty} M(t) - \frac{t}{\mu} = \frac{\sigma^2 - \mu^2}{2\mu^2}$$

19.2. **Major Themes of the Course.**

- Markov Processes
  - Branching Processes
  - Continuous Time Markov Chains
    - Birth-death Processes
    - Yule Processes
  - Poisson Processes
    - Renewal Processes
- Martingales

19.3. **Conditional Expectations.** $\mathbb{E}[X]$ can be thought of as an average. What happens when we condition, i.e. $\mathbb{E}[X|Y]$? Suppose $Y = X^2$ and

$$X = \begin{cases} -2 & \frac{1}{3} \\ 0 & \frac{1}{3} \\ 2 & \frac{1}{3} \end{cases}$$

Then

$$\begin{aligned} \mathbb{E}[X|Y=0] &= 0\mathbb{1}_{Y=0} \\ \mathbb{E}[X|Y=4] &= 0 \\ \mathbb{E}\left[X^2|Y=4\right] &= 4 \end{aligned}$$

19.4. **Martingales.** $\{M_n\}_{n\geq 1}$ is a martingale if

$$\mathbb{E}[M_n|M_{n-1}] = M_{n-1}$$

and $\mathbb{E}[|M_n|] < \infty$. $\mathcal{F}_n$ is the information contained in $\{M_1, \ldots, M_n\}$. Clearly, $\mathcal{F}_{n-1} \subset \mathcal{F}_n$. This is called filtration. Another way to write the definition of a martingale is

$$\mathbb{E}[M_n|\mathcal{F}_{n-1}] = M_{n-1}$$

Example 26. *Let $X_i \overset{iid}{\sim} f$ with $\mathbb{E}[X_i] = 0$. Let*

$$M_n = \sum_{i=1}^{n} X_i, \quad \mathcal{F}_n = \{X_1, \ldots, X_n\}$$

We have

$$\begin{aligned} \mathbb{E}[M_n|M_{n-1}] &= \mathbb{E}[M_{n-1} + X_n|M_{n-1}] \\ &= M_{n-1} + \mathbb{E}[X_n|M_{n-1}] \\ &= M_{n-1} + \mathbb{E}[X_n] \\ &= M_{n-1} \end{aligned}$$

Example 27. *Let U_i be iid and*

$$M_n = \prod_{i=1}^{n} U_i, \quad \mathbb{E}[U_i] = 1$$

Then we have

$$\begin{aligned}
\mathbb{E}[M_n | M_{n-1}] &= \mathbb{E}[M_{n-1} U_n | M_{n-1}] \\
&= M_{n-1} \mathbb{E}[U_n] \\
&= M_{n-1}
\end{aligned}$$

Example 28. *Let X_i be iid, with $\mathbb{E}[X_i] = 0$, $\mathrm{Var}(X_i) = \sigma^2$ and*

$$S_n = \sum_{i=1}^{n} X_i$$

Let $M_n = S_n^2 - n\sigma^2$. Now, since $S_n = S_{n-1} + X_n$, so

$$\begin{aligned}
\mathbb{E}[M_n | M_{n-1}] &= \mathbb{E}\left[S_{n-1}^2 + 2S_{n-1}X_n + X_n^2 - n\sigma^2 | M_n\right] \\
&= \mathbb{E}\left[\left(S_{n-1}^2 - (n-1)\sigma^2\right) + 2S_{n-1}X_n + X_n^2 - n\sigma^2 | M_{n-1}\right] \\
&= M_{n-1} + \underbrace{\mathbb{E}[2S_{n-1}X_n | M_{n-1}]}_{=0} + \underbrace{\mathbb{E}\left[X_n^2 - \sigma^2 | M_{n-1}\right]}_{=0} \\
&= M_{n-1}
\end{aligned}$$

The first expectation is 0 since $X_n \perp\!\!\!\perp M_{n-1}$. The second one is zero since the variance of X_n is σ^2.

Example 29. *Let $\phi(\lambda) = \mathbb{E}\left[e^{\lambda X}\right]$ and X_i be iid. Let*

$$Y_i = \frac{e^{\lambda X_i}}{\phi(\lambda)}$$

We have

$$M_n = \prod_{i=1}^{n} Y_i$$

is a martingale. Note that

$$\mathbb{E}[Y_i] = \frac{\mathbb{E}\left[e^{\lambda X_i}\right]}{\phi(\lambda} = 1$$

If $\exists \lambda_0$ such that $\phi(\lambda_0) = 1$, then

$$M_n = e^{\lambda_0 \sum_{i=1}^{n} X_i}$$

is a martingale.

Example 30. *Suppose $X_i \overset{iid}{\sim} f$ with f unknown. What do we do if we wish to test for f, that is*

$$\begin{cases} H_0 : f = f_0 \\ H_1 : f = f_1 \end{cases}$$

Neyman and Pearson proposed to look at

$$T = \prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)}$$

and to reject H_0 if $T > c_\alpha$ for some c_α. In fact,

$$M_n = \prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)}$$

is a martingale if H_0 is true.

$$\begin{aligned}
\mathbb{E}\left[\frac{f_1(X_i)}{f_0(X_i)}\right] &= \int \frac{f_1(x_i)}{f_0(x_i)} f_0(x_i) \,\mathrm{d}x_i \\
&= 1
\end{aligned}$$

Example 31. *Suppose H_0 is $f_0 \sim \mathcal{N}(0,1)$ and H_1 is $f_1 \sim \mathcal{N}(\lambda, 1)$. Then*

$$\begin{aligned}
\frac{f_1(x)}{f_0(x)} &= \frac{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\lambda)^2\right\}}{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}} \\
&= e^{\lambda x - \frac{\lambda^2}{2}}
\end{aligned}$$

Now we see that since

$$\begin{aligned}\mathbb{E}\big[e^{\lambda}x\big] &= e^{\lambda^2/2} \\ &= \phi(\lambda)\end{aligned}$$

this is immediately a martingale. However, this is also connected to the previous example where we have a ratio of two densities.

Example 32. *Let H_0 be $f = \mathrm{Pois}(1)$ and H_1 be $f = \mathrm{Pois}(\lambda)$. Then*

$$\begin{aligned}\frac{f_1(x)}{f_0(x)} &= \frac{\lambda^{x_i}e^{-\lambda}}{1^{x_i}e^{-1}} \\ &= e^{x_i \log\lambda - (\lambda-1)}\end{aligned}$$

and so our martingale is

$$M_n = \prod_{i=1}^{n} e^{x_i \log\lambda - (\lambda-1)}$$

This follows as a result of the fact that if $X \sim \mathrm{Pois}(1)$,

$$\mathbb{E}\big[e^{X\log\lambda}\big] = e^{\lambda-1}$$

Example 33. *Let X_n be a branching process and μ be the progeny distribution. Then*

$$M_n = \frac{X_n}{\mu^n}$$

is a martingale. We can verify this easily:

$$\begin{aligned}\mathbb{E}[M_{n+1}|X_n] &= \mathbb{E}\left[\frac{X_{n+1}}{\mu^{n+1}}\middle| X_n\right] \\ &= \frac{1}{\mu^{n+1}}\mathbb{E}[X_{n+1}|X_n] \\ &= \frac{\mu^{n+1}}{\mu X_n} \\ &= \frac{X_n}{\mu^n}\end{aligned}$$

20. April 21st, 2014

20.1. Some Solutions to Midterm-Exam.

4) Minimum of Exponentials.

5b)

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} -\lambda & \lambda & 0 & 0 \\ \mu & -(\lambda+\mu) & \lambda & 0 \\ 0 & \mu & -(\lambda+\mu) & \lambda \\ 0 & 0 & \mu & -\mu \end{pmatrix}\end{array}$$

Then we solve $\pi Q = 0$. Note that this is not the same as solving $Q\pi = 0$.

5c)

$$\left(\frac{n}{t}\right)\left(1-\frac{x}{t}\right)^{n-1} = \frac{\mathbb{P}(N(t-x)=n-1)}{\mathbb{P}(N(t)=n)}f(x)$$

Let $u_n = \mathbb{P}(N(t)=n)$. Plugging in $x=0$, we get

$$\frac{n}{t} = \frac{u_{n-1}}{u_n}f(0)$$

20.2. Martingales.

$\{M_n, \mathcal{F}_n\}$ is a martingale. We can write the martingale property as

$$\mathbb{E}[M_n|\mathcal{F}_n] = M_{n-1}$$

or

$$\mathbb{E}[M_n|M_{n-1},\ldots,M_1] = M_{n-1}$$

Let $N(t) \sim \text{Pois}(\lambda t)$. Then $M_t = N(t) - \lambda t$ is a martingale. For any $s < t$,

$$\begin{aligned}
\mathbb{E}[M_t|M_s] &= \mathbb{E}[N(t) - \lambda t|M_s] \\
&= \mathbb{E}[N(t) - N(s) + N(s) - \lambda t|M_s] \\
&= \mathbb{E}[N(t) - N(s)] + \mathbb{E}[N(s) - \lambda t|M_s] \\
&= \lambda t - \lambda s + N(s) - \lambda t \\
&= N(s) - \lambda s \\
&= M_s
\end{aligned}$$

Note that this is not the martingale property we had earlier, i.e. $\mathbb{E}[M_n|M_{n-1}] = M_{n-1}$. However, we have

Theorem 14. *If $\mathbb{E}[M_n|M_{n-1}] = M_{n-1}$, then for $m < n$,*

$$\mathbb{E}[M_n|M_m] = M_m$$

Proof.

$$\begin{aligned}
\mathbb{E}[M_n|M_{n-2}] &= \mathbb{E}[\mathbb{E}[M_n|M_{n-1}, M_{n-2}]|M_{n-2}] \\
&= \mathbb{E}[M_{n-1}|M_{n-2}] \\
&= M_{n-2}
\end{aligned}$$

and we can continue this by induction. □

Theorem 15. *Suppose you have $\{M_n, \mathcal{F}_n\}$ with*

$$\sup_n \mathbb{E}\left[M_n^2\right] \le C < \infty$$

Then $M_n \longrightarrow M_\infty$ for some random variable M_∞.

Let $X_i = \pm 1$, each with a probability $\frac{1}{2}$. Let

$$M_n = \sum_{k=1}^{n} \frac{1}{k} X_i$$

Note that we have our condition

$$\begin{aligned}
\sup_n \mathbb{E}\left[M_n^2\right] &= \sum_{k=1}^{n} \frac{1}{k^2} \\
&\le \sum_{k=1}^{\infty} \frac{1}{k^2} \\
&= \frac{\pi^2}{6}
\end{aligned}$$

Corollary. *This result holds if*

$$\sup_n \mathbb{E}[M_n^p] \le C < \infty$$

for any $p \ge 2$. This is, in fact, not true for $p = 1$.

Definition 7 (Stopping Time)**.** *An integer τ is a stopping time if when $\{\tau = k\}$, we can answer our question by looking at $X_1, \ldots, X_k$ and have no information about the future.*

$\tau = C$ is a stopping time, where C is a constant.

Theorem 16. *Suppose we have $\{M_n, \mathcal{F}_n\}$ and τ be a stopping a time with $\tau \le N$ with probability 1. Then*

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$$

Why is this result so fascinating?

$$\mathbb{E}[M_n|M_{n-1}] = M_{n-1}$$

By applying Adam's law,

$$\mathbb{E}[\mathbb{E}[M_n|M_{n-1}]] = \mathbb{E}[M_{n-1}]$$

and so

$$\mathbb{E}[M_n] = \mathbb{E}[M_{n-1}]$$

By induction, we get

$$\mathbb{E}[M_n] = \mathbb{E}[M_0]$$

However, in the above theorem, note that τ is random.

Definition 8 (Submartingale)**.** *$\{M_n\}$ is a submartingale if*

$$\mathbb{E}[M_n|M_{n-1}] \ge M_{n-1}$$

Theorem 17. *If X_n is a martingale and $M_n = U(X_n)$, where U is any convex function, then M_n is a submartingale.*

21. April 23rd, 2014

Recall that if M_n is a martingale, then

$$\mathbb{E}[M_n] = \mathbb{E}[M_m] = \mathbb{E}[M_0]$$

for $m \leq n$.

Theorem 18 (Doob's Optional Stopping Theorem)**.** *If τ is a bounded stopping time,*

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$$

It is important to note that bounded is not the same as being finite with probability 1. For example, a Normal random variable is bounded, but not finite.

Proof of Doob's Optional Stopping Theorem. Suppose $\tau \leq N$, since τ is bounded. Then

$$\begin{aligned} M_\tau &= \sum_{n=0}^{\infty} M_n \mathbb{1}_{\{\tau=n\}} \\ &= \sum_{n=0}^{N} M_n \mathbb{1}_{\{\tau=n\}} \end{aligned}$$

Taking expectations, we get

$$\begin{aligned} \mathbb{E}[M_\tau] &= \mathbb{E}\left[\sum_{n=0}^{N} M_n \mathbb{1}_{\{\tau=n\}}\right] \\ &= \mathbb{E}\left[\sum_{n=0}^{N} \mathbb{E}[M_n|\mathcal{F}_n]\, \mathbb{1}_{\{\tau=n\}}\right] \end{aligned}$$

By the martingale property,

$$\mathbb{E}[M_n|M_m] = M_m$$

so we have

$$\begin{aligned} \mathbb{E}[M_\tau] &= \mathbb{E}\left[\sum_{n=0}^{N} \mathbb{E}\left[M_n \mathbb{1}_{\{\tau=n\}}|\mathcal{F}_n\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\sum_{n=0}^{N} \mathbb{E}\left[M_N \mathbb{1}_{\{\tau=n\}}|\mathcal{F}_n\right]\right]\right] \\ &= \mathbb{E}\left[\sum_{n=0}^{N} M_N \mathbb{1}_{\{\tau=n\}}\right] \\ &= \mathbb{E}\left[M_N \sum_{n=0}^{N} \mathbb{1}_{\{\tau=n\}}\right] \\ &= \mathbb{E}[M_N] \end{aligned}$$

and so

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_N] = \mathbb{E}[M_0]$$

□

Theorem 19 (Wald's Theorem)**.** *Let $S_n = \sum_{i=1}^{n} X_i$. If τ is a stopping, then*

$$\mathbb{E}[S_\tau] = \mathbb{E}[\tau]\, \mathbb{E}[X_i]$$

Proof. First,

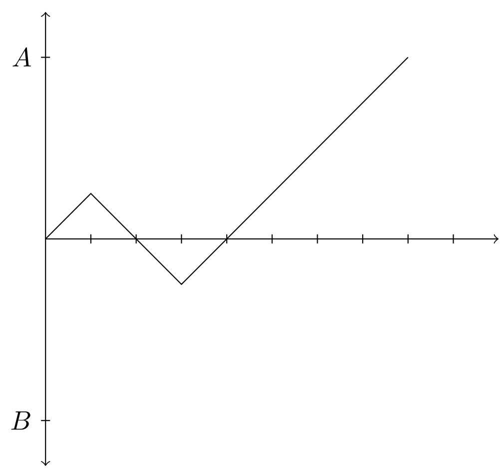
$$\begin{aligned} S_\tau &= \sum_{k=1}^{\tau} X_k \\ &= \sum_{k=1}^{\infty} X_k \mathbb{1}_{\{\tau>k\}} \end{aligned}$$

Taking expectations, we get

$$\begin{aligned}
\mathbb{E}[S_\tau] &= \sum_{k=1}^{\infty} \mathbb{E}\big[X_k \mathbb{1}_{\{\tau > k-1\}}\big] \\
&= \sum_{k=1}^{\infty} \mathbb{E}[\mathbb{E}[X_k \mathbb{1}_{\tau > k-1} | \mathcal{F}_{k-1}]] \\
&= \sum_{k=1}^{\infty} \mathbb{E}[X_k]\, \mathbb{E}[\mathbb{1}_{\tau > k-1} | \mathcal{F}_{k-1}] \\
&= \mathbb{E}[X_1] \sum_{k=1}^{\infty} \mathbb{P}(\tau > k-1) \\
&= \mathbb{E}[X_1]\, \mathbb{E}[\tau]
\end{aligned}$$

□

Suppose we have a regular random walk and two integer points $A > 0$ and $B < 0$.



Let

$$\tau = \min\{n \geq 1, S_n = A \text{ or } B\}$$

be a stopping time. What is $\mathbb{P}(S_\tau = A)$? That, is what is the probability we hit A before B.

We cannot say that

$$\mathbb{E}[S_0] = \mathbb{E}[S_\tau]$$

since τ is not bounded. Instead, let

$$\tau \wedge n := \min(\tau, n)$$

in which case we would have

$$0 = \mathbb{E}[S_0] = \mathbb{E}[S_{\tau \wedge n}]$$

We also have

$$0 = \mathbb{E}[S_\tau] = A\mathbb{P}(S_\tau = A) - B\, \underbrace{\mathbb{P}(S_\tau = B)}_{1-\mathbb{P}(S_\tau = A)}$$

which implies

$$\mathbb{P}(S_\tau = A) = \frac{|B|}{A + |B|}$$

If we take $n \to \infty$, then $\lim_{n\to\infty} \tau \wedge n = \tau$ and

$$\lim_{n\to\infty} S_{\tau\wedge n} = S_\tau$$

Claim.

$$\lim_{n\to\infty} \mathbb{E}[S_{\tau\wedge n}] = \mathbb{E}\Big[\lim_{n\to\infty} S_{\tau\wedge n}\Big] = \mathbb{E}[S_\tau]$$

This result will not always hold:

Example 34. *Suppose we have a random variable*

$$X_n = \begin{cases} 2^n & \frac{1}{2^n} \\ 0 & 1 - \frac{1}{2^n} \end{cases}$$

Then

$$\mathbb{P}(X_n \neq 0) = \frac{1}{2^n}$$

so

$$\lim_{n\to\infty} \mathbb{P}(X_n \neq 0) = 0$$

This implies

$$\lim_{n\to\infty} X_n = 0$$

Note that

$$\mathbb{E}[X_n] = 1$$

and so

$$\underbrace{\lim_{n\to\infty} \mathbb{E}[X_n]}_{=1} \neq \underbrace{\mathbb{E}\Big[\lim_{n\to\infty} X_n\Big]}_{=0}$$

The answer to the above problem is in fact

$$\mathbb{E}[\tau] = AB$$

Proof. Let $M_n = S_n^2 - n$, which is our martingale. Then

$$0 = \mathbb{E}\big[S_\tau^2 - \tau\big] \implies \mathbb{E}[\tau] = \mathbb{E}\big[S_\tau^2\big]$$

We know

$$0 = \mathbb{E}\big[S_{\tau\wedge n}^2 - \tau \wedge n\big]$$

and

$$\lim_{n\to\infty} \tau \wedge n = \tau \qquad \lim_{n\to\infty} S_{\tau\wedge n} = S_\tau$$

By monotonicity, we have

$$\begin{aligned} 0 &= \lim_{n\to\infty} \mathbb{E}\big[S_{\tau\wedge n}^2 - \tau \wedge n\big] \\ &= \mathbb{E}\Big[\lim_{n\to\infty} S_{\tau\wedge n}^2 - \tau \wedge n\Big] \\ &= \mathbb{E}\big[S_\tau^2 - \tau\big] \end{aligned}$$

There will be a question on the exam that has this on it. □

Recall that $\{M_n, \mathcal{F}\}$ is called a submartingale if

$$\mathbb{E}[M_n|\mathcal{F}_{n-1}] \geq M_{n-1}$$

Similarly, we have something called a supermartingale if

$$\mathbb{E}[M_n|\mathcal{F}_{n-1}] \leq M_{n-1}$$

How would we construct these?

Lemma. *If M_n is a martingale and φ is a convex function, then $X_n = \phi(M_n)$ is a submartingale. E.g. ($\varphi(x) = x^2, e^x, \ldots$)*

Proof.

$$\begin{aligned} \mathbb{E}[X_n|\mathcal{F}_{n-1}] &= \mathbb{E}[\phi(M_n)|\mathcal{F}_{n-1}] \\ &\geq \varphi(\mathbb{E}[M_n|\mathcal{F}_{n-1}]) \qquad \text{by Jensen's inequality} \\ &= \varphi(M_{n-1}) \\ &= X_{n-1} \end{aligned}$$

□

The idea behind this is essentially that this is stochastically non-decreasing. Another way to see this is to take expectations of both sides,

$$\mathbb{E}[\mathbb{E}[M_n|\mathcal{F}_{n-1}]] \geq \mathbb{E}[M_{n-1}] \implies \mathbb{E}[M_n] \geq \mathbb{E}[M_{n-1}]$$

Let $\{X_n, \mathcal{F}_n\}$ be a martingale and $\{A_n\}$ be a sequence of random variables that only depend on $\mathcal{F}_{n-1}$, that is

$$\mathbb{E}[A_n|\mathcal{F}_{n-1}] = A_n$$

Let

$$Z_n = \sum_{k=1}^{n} A_k(M_k - M_{k-1})$$

Claim. *$\{Z_n, \mathcal{F}_n\}$ is a martingale.*

Proof.

$$\begin{aligned}\mathbb{E}[Z_n|\mathcal{F}_{n-1}] &= Z_{n-1} + \mathbb{E}[A_n(M_n - M_{n-1})|\mathcal{F}_{n-1}]\\ &= Z_{n-1} + A_n\mathbb{E}[M_n - M_{n-1}|\mathcal{F}_{n-1}]\\ &= Z_{n-1}\end{aligned}$$

□

Recall that the Markov inequality states that if $Y > 0$ is a random variable, then

$$\mathbb{P}(Y > \lambda) \leq \frac{\mathbb{E}[Y]}{\lambda}$$

Let $M_n^* = \max_{0\leq m\leq n} M_m$. Then

Theorem 20 (Doob's Maximal Inequality)**.** *If M_n is a non-negative sub-martingale,*

$$\mathbb{P}(M_n^* > \lambda) \leq \frac{\mathbb{E}[M_n]}{\lambda}$$

Note that in the right hand side we have M_n, not M_n^*.

22. April 28th, 2014

22.1. **Doob's Maximal Inequality.** Recall that Markov's inequality states that if $X \geq 0$, then

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}[X]}{\lambda}$$

In fact, for any $p \geq 1$,

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}[X^p]}{\lambda^p}$$

and

$$\mathbb{P}\left(e^{tX} \geq e^{\lambda t}\right) \leq \frac{\mathbb{E}\left[e^{tX}\right]}{e^{\lambda t}}$$

Let $\{M_n\}$ be a non-negative submartingale. Define $M_n^* = \max_{0\leq m\leq n} M_m$. Then recall that Doob's Maximal Inequality states that

$$\mathbb{P}(M_n^* > \lambda) \leq \frac{\mathbb{E}[M_n]}{\lambda}$$

Proof. For $\lambda > 0$, let $\tau = \min\{m \geq 0, M_n \geq \lambda\}$. Then

$$\mathbb{P}(M_n^* > \lambda) = \mathbb{P}(\tau \leq n)$$

We also know

$$\begin{aligned}\lambda\mathbb{1}\{\tau \leq n\} &\leq M_\tau\mathbb{1}\{\tau \leq n\}\\ &= \sum_{k=0}^{n} M_k\mathbb{1}\{\tau = k\}\end{aligned}$$

Taking expectations on both sides, we get

$$\begin{aligned}\lambda\mathbb{P}(\tau \leq n) &\leq \mathbb{E}\left[\sum_{k=0}^{n} M_k\mathbb{1}\{\tau = k\}\right]\\ &\leq \mathbb{E}\left[\sum_{k=0}^{n} M_n\mathbb{1}\{\tau = k\}\right] && \text{since } M_k \leq \mathbb{E}[M_n|M_k] \text{ by sub-martingale property}\\ &= \mathbb{E}\left[M_n\sum_{k=0}^{n}\mathbb{1}\{\tau = k\}\right]\\ &= \mathbb{E}[M_n\mathbb{1}\{0 \leq \tau \leq n\}]\\ &= \mathbb{E}[M_n]\end{aligned}$$

□

22.2. **Doob's Optional Stopping Theorem.** Recall that if $\{M_n\}$ is a martingale and $\tau < C$, then $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$. However, the condition that $\tau < C$ is one of three conditions that are necessary for this to hold. These three are

(1) $\tau < C$.
(2) $\tau_n < \infty, |M_n| < C$.
(3) $\mathbb{E}[\tau] < \infty, |M_n - M_{n-1}| \leq C$. This is often the most realistic condition.

22.3. **Brownian Motion.**

Definition 9 (Brownian Motion)**.** *$\{B_t\}$ follows Brownian motion if*

(1) $B_0 = 0$.
(2) For any n, $B_{t_1} - B_{t_2}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$ *are independent. (Independence)*
(3) $B_{t+s} - B_t \sim \mathcal{N}(0,1)$. *(Stationarity)*
(4) $\{B_t\}$ *is a continuous function.*

The fact that $\{B_t\}$ is continuous is completely non-trivial. It is also known that if continuity holds, (2) and (3) are the only properties that can hold.

Theorem 21 (Scale Invariance)**.** *Let $c > 0$. Define $X_t = \frac{1}{\sqrt{c}} B_{ct}$. Then X_t is a Brownian motion.*

Note that

$$X_t \sim \frac{1}{\sqrt{c}} \mathcal{N}(0, ct) \sim \mathcal{N}(0,t)$$

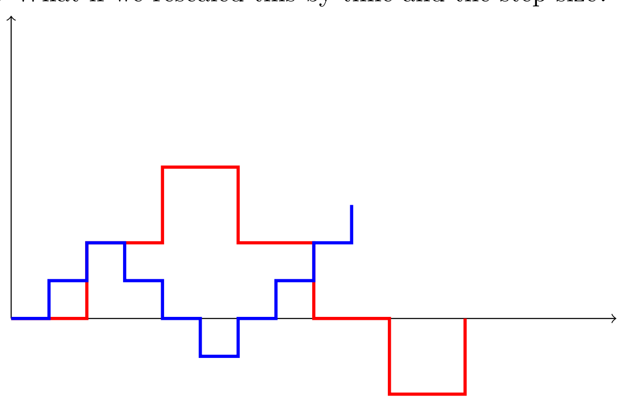
so this makes sense. The theorem essentially says that if we take a Brownian motion and stretch or compress it, it is still a Brownian motion. Is this a Markov process? Note that $B_t - B_t \perp\!\!\!\perp B_t$. To find the distribution of B_T, we need to know the distribution of B_t and so this is a Markov process.

Consider $\{B_t, \mathcal{F}_t\}$. Then

$$\begin{aligned}\mathbb{E}[B_t | \mathcal{F}_s] &= \mathbb{E}[B_t - B_s + B_s | \mathcal{F}_s] \\ &= \mathbb{E}[B_t - B_s] + B_s \\ &= B_s\end{aligned}$$

What is remarkable here is that this also says that Brownian motion is a martingale. What else is similar to this? A Poisson process comes close. We have, in fact, that $N_t - \lambda t$ satisfies both these properties.

Consider a typical random walk. What if we rescaled this by time and the step size?



This in fact converges to Brownian motion (provided we control the rate at which we scale time and step size). Why is this true? Central limit theorem results state that given $\{X_k\}$,

$$S_n = \sum_{k=1}^{\lfloor nt \rfloor} \frac{1}{\sqrt{nt}} X_k \xrightarrow{d} \mathcal{N}(0,t) = B_t$$

Another interesting fact is that the total variation of Brownian motion is infinite. That is, if we were to trace the curve of the Brownian motion and project if onto the y-axis, the distance the projection would travel is infinite.

22.3.1. *Sampling from a Brownian Motion.* If we have a set of points $t_1, t_2, \dots, t_n$, which are points where we want to sample, we can simply use the fact that $B_{t_1} - H_0 \perp\!\!\!\perp B_0$ and sample form a $\mathcal{N}(0, t_1)$ random variable. Then, using the fact that $B_{t_2} - B_{t_1} \perp\!\!\!\perp B_{t_1}$, where $B_{t_2} - B_{t_1} \sim \mathcal{N}(0, t_2 - t_1)$. Then we have

$$B_{t_2} = B_{t_1} + (B_{t_2} - B_{t_1})$$

and we keep doing this.

How do we find the covariance?

$$\begin{aligned}\mathrm{Cov}(B_t, B_s) &= \mathbb{E}[B_t B_s] - \mathbb{E}[B_t]\,\mathbb{E}[B_s] \\ &= \mathbb{E}[(B_t - B_s + B_s) B_s] \\ &= \underbrace{\mathbb{E}[(B_t - B_s) B_s]}_{=0} + \underbrace{\mathbb{E}\left[B_s^2\right]}_{=s} \\ &= s\end{aligned}$$

assume that $t > s$. Hence we have

$$\mathrm{Cov}(B_t, B_s) = \min(t, s)$$

Let $\{B_t\}$ be a Brownian motion. Let

$$M_t = e^{\lambda B_t - \frac{\lambda^2 t}{2}}$$

Theorem 22. *$\{M_t, \mathcal{F}_t\}$ is a martingale.*

Proof.

$$\begin{aligned}
\mathbb{E}[M_t | M_s] &= \mathbb{E}\left[e^{\lambda(B_t - B_s) + \lambda B_s - \frac{\lambda^2 t}{s}} \middle| M_s\right] \\
&= e^{\lambda B_s} \mathbb{E}\left[e^{\lambda(B_t - B_s)}\right] e^{-\frac{\lambda^2 t}{2}} \\
&= e^{\lambda B_s} e^{\frac{\lambda^2}{2}(t-s) - \frac{\lambda^2 t}{2}} \\
&= e^{\lambda B_s - \frac{\lambda^2 s}{2}} \\
&= M_s
\end{aligned}$$

$\square$

Example 35. *Suppose we want to find the expected time it takes for a Brownian motion to hit a given value a. We can find this by using Doob's Optional Stopping Theorem.*

$$\tau = \inf\{t \geq 0, B_t = a\}$$

We know

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$$

where $M_0 = e^{\lambda B_0 - \frac{\lambda^2 \cdot 0}{2}} = 1$. Then we have

$$\mathbb{E}\left[e^{\lambda B_\tau - \frac{\lambda^2}{2}\tau}\right] = \mathbb{E}\left[e^{\lambda a - \frac{\lambda^2 \tau}{2}}\right] = 1 \Rightarrow \mathbb{E}\left[e^{-\frac{\lambda^2}{2}\tau}\right] = e^{-\lambda a}$$

So this gives

$$\frac{\lambda^2}{2} = \mathbb{E}\left[e^{t\tau}\right] = e^{2ta}$$

23. April 30th, 2014

Suppose we have

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

We have, marginally, that

$$\begin{aligned}
X &\sim \mathcal{N}(0, 1) \\
Y &\sim \mathcal{N}(0, 1)
\end{aligned}$$

How do we simulate from this distribution? If $Z \sim \mathcal{N}(0, \sigma^2)$, how do we simulate from this? We have that

$$Z(0, \sigma^2) \stackrel{d}{=} \sigma \mathcal{N}(0, 1)$$

so we sample from a standard Normal and then scale it. A similar property holds for the multivariate case. If

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

then we have

$$\begin{pmatrix} X \\ Y \end{pmatrix} \stackrel{d}{=} \Sigma^{1/2} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

where we obtain $\Sigma^{1/2}$ by using the Cholesky decomposition.

In the case of Brownian motion, $\mathrm{Cov}(B_s, B_t) = \min(s, t)$.'

The length of the path of a Brownian motion is

$$\begin{aligned}
\int |\mathrm{d}f| &= \int \frac{|\mathrm{d}f|}{\mathrm{d}x} \, \mathrm{d}x \\
&= \lim_{n \to \infty} \sum \left| f\left(\frac{i}{n}\right) - f\left(\frac{i-1}{n}\right) \right|
\end{aligned}$$

Quadratic Variation. If we instead consider

$$\lim_{n\to\infty} \sum \left|B_{iT/n} - B_{(i-1)T/n}\right|^2$$

Theorem 23.

$$\lim_{n\to\infty} \sum \left|B_{iT/n} - B_{(i-1)T/n}\right|^2 = T$$

Proof. Let

$$Z = \sum \left|B_{iT/n} - B_{(i-1)T/n}\right|^2$$

Then we have

$$\begin{aligned}\mathbb{E}[Z] &= \sum \frac{T}{n} \\ &= T\end{aligned}$$

We also have

$$\begin{aligned}\operatorname{Var}(Z) &= \sum \left(\left|B_{iT/n} - B_{(i-1)T/n}\right|^2 - \frac{T}{n}\right)^2 \\ &= \frac{T}{n}\end{aligned}$$

□

23.1. **Stochastic Differential Equations (Not Examinable).** If we have for a given $y(t)$,

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \alpha y(t)$$

The solution to this is

$$y(t) = y_0 e^{\alpha t}$$

Suppose we want to compensate for error. Then we have

$$\mathrm{d}y = \alpha y(t)\mathrm{d}t + \mathrm{d}B_t$$

where

$$\mathrm{d}B_t \sim \mathcal{N}(0, \mathrm{d}t)$$

This changes our notion of calculus. Recall that the fundamental theorem of calculus states that

$$f(x) - f(0) = \int_0^x f'(y)\,\mathrm{d}y$$

where

$$\int_0^x g(y)\,\mathrm{d}y = \lim_{h\to\infty} \sum_i g\left(\frac{i-1}{n}\right)\Delta n$$

Motivated by this, we have

$$\lim_{n\to\infty} \sum_i g(B_{iT/n})(B_{it/n} - B_{(i-1)T/n})$$

This turns out to satisfy

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}s$$

$$\begin{aligned}f(B_t) - f(B_0) &= \sum_i f(B_{it/n}) - f(B_{(i-1)t/n}) \\ &= \sum f'(B_{it/n})(B_{it/n} - B_{(i-1)t/n}) + \frac{1}{2} f(B_{(i-1)t/n})(B_{it/n} - B_{(i-1)t/n}) \\ &= \int_0^t f'(B_s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}B_s\end{aligned}$$

Suppose we want to discretize our formula

$$\mathrm{d}y = \alpha y(t)\mathrm{d}t + \mathrm{d}B_t$$

Then we can write it as

$$y_{i+1} - y_i = \alpha y_i(t_{i+1} - t_i) + \sqrt{t_{i+1} - t_i}\mathcal{N}(0,1)$$

Rearranging gives us

$$y_{i+1} = (1 + \alpha(t_{i+1} - t_i))y_i + Z_i$$

and we note that this is an AR(1) process.