Stochastic Processes

Riley Wilson

last updated: 1/4/2021

These are notes compiled when I took math 171 at UCLA as taught by Moritz Voss. The textbook used for the course was Durrett's *Essentials of Stochastic Processes*,[1] so what follows is bound to be similar.

The document is arranged by week and lecture number, not mathematical content. Some weeks only have two lectures since there was a holiday or I decided to compress two classes' worth of material under one heading. Occasionally, I include material covered during discussion section. You'll find these notes titled "Jacob's section."

Any mathematical mistakes/typos are my own.

1   Week 0

First, we must understand what we mean by "stochastic". "Stochastic" in our context is very close in meaning to "random" For our purposes, imagine they are interchangeable.

Mathematically, we define something called a stochastic process.

Definition 1.1. Stochastic Process A stochastic process is a collection of random variables (X_t) $t \in I$, where I is the index set of the process. The random variables are defined on a common scenario space, Ω and take values in a common state space, S. Formally, we say

---

[1]The text is freely available online if your organization has access to Springer Books.

$$X_t : \Omega \to S, \qquad\qquad \omega \mapsto X_t(\omega) = x \in S$$

Remember, if $S \subseteq \mathbb{N}$, we say X_t is a discrete random variable process. Likewise, if $S \subseteq \mathbb{R}$, it would be a continuous one. We will get to continuous-time stochastic processes later in the notes.

Why are stochastic processes important? Let's consider a problem where they can be helpful.

Gambler's Ruin problem: Suppose we go to the casino with 10 dollars and need 80. We play european-style roulette, betting 1 dollar on red each time. If we win, we get our 1 dollar back, plus an extra dollar. If we lose, we lose the dollar. The probability of getting red on any given spin is $p = \frac{18}{37} \approx .486$. What is the probability we go bankrupt (are 'ruined') before winning 80 dollars?

Solution: First, we consult our intuition to get a sense of the problem. We know the game is "unfair", as the probability of winning a bet is below .5. But is it mathematically unfair? What's the expected value?

Let $W = P - c$ where P is a random variable representing our payout and c is our stake, or what we originally bet.

$$\mathbb{E}[W] = \mathbb{E}[P] - c = 2c(p) + 0(1 - p) - c$$
$$= c(2p - 1) = \frac{-c}{37} < 0$$

Which is not unfair.

To simplify things, we can model red with a random variable, X_i. $P(X_i = 1) = p$, and $P(X_i = -1) = 1 - p$.

We can define the gambler's total wealth as S_n, where n is the number of times she's made a bet. Hence, $S_0 = 10$, and

$$S_n = S_0 + X_1 + X_2 + \cdots + X_n = S_0 + \sum_{i=1}^{n} X_i$$

2

Where $n \geq 1$. Recall definition 1.1 and see S_n is a stochastic process. We introduce a new concept to sharpen our thinking about the problem.

Definition 1.2. Hitting times A "hitting time" is a random variable whose observed value is the time a stochastic process reaches a specific value.

For the gambler's ruin problem, we'll define hitting times as:

$$T_0 = \min\{n \geq 1 : S_n = 0\} \qquad T_{80} = \min\{n \geq: S_n = 80\}$$

Note T_0 and T_{80} are random variables that tell us when the gambler goes bankrupt or wins. In this vocabulary, we can describe the ruin probability as:

$$P(T_0 < T_{80})$$

We could immediately calculate this if we knew the joint probability distribution of T_0 and T_{80}. Unfortunately, we don't.

A different way relies on defining a "ruin function." Say

$$B_k = \{\text{gambler goes bankrupt with k initial capital before reaching K}\}$$

The ruin function is $R(k) = P(B_k)$. Notice $R(0) = 1$ and $R(K) = 0$.

We then introduce an event set.

$$W = \{\text{gambler wins first round}\}$$

Now, we can examine the ruin function a little closer.

$$
\begin{aligned}
R(k) &= P(B_k) \\
&= P(B_k \mid W)P(W) + P(B_k \mid W^c)P(W^c) \\
&\text{(by the law of total probability)} \\
&= P(B_{k+1})P(W) + P(B_{k-1})P(W^c)
\end{aligned}
$$

3

The last step is justified by the Markov property of a random walk. We will cover this later.

We can get a recursive formula for $R(k)$ this way.

$$R(k) = R(k+1)P(W) + R(k-1)P(W^c)$$
$$= R(k+1)(p) + R(k-1)(1-p)$$

We can solve this and get a closed form. The method is a little lengthy, and involves iterating the difference equation and a telescoping series. The calculations are not conceptually important. In the end, we get:

$$R(k) = \frac{(\frac{1-p}{p})^k - (\frac{1-p}{p})^K}{1 - (\frac{1-p}{p})^K}$$

for all $k = 1, \ldots, K$. Now, we can get the ruin probability for different starting capitals (k), target capitals (K), and probabilities of winning (p). For instance, if $k = 10$, $K = 80$, and $p = .486$, then

$$R(10) = \frac{(\frac{.514}{.486})^{10} - (\frac{.514}{.486})^{80}}{1 - (\frac{.514}{.486})^{80}} = .9914$$

In words, our gambler is going to have a bad time.

We've solved the problem, but we might want to see what happens if the gambler changes her strategy. For instance, she might bet 2 dollars each time instead of 1. What is the ruin probability then?

Our current closed form works perfectly for this. Imagine each "dollar" is now a chip, with x dollars. Now, the gambler bets x dollars, or 1 chip, per round. As before, she loses all the money if she loses the round, and wins it back at 1:1 odds if she wins. p doesn't need to be modified on the new interpretation. The new k and K's only need to be divided by x.

The ruin probability if the gambler bets 5 dollars per round is:

$$R(10/5) = \frac{\left(\frac{.514}{.486}\right)^{10/5} - \left(\frac{.514}{.486}\right)^{80/5}}{1 - \left(\frac{.514}{.486}\right)^{80/5}}$$

$$= \frac{\left(\frac{.514}{.486}\right)^2 - \left(\frac{.514}{.486}\right)^{16}}{1 - \left(\frac{.514}{.486}\right)^{16}}$$

$$= .9182$$

She does better. As it turns out, because this is a negative expected value game the optimal strategy is to minimize the number of bets you need to reach K. This means you should go all in every time.

The best ruin probability is then:

$$1 - P(\text{reaching } 80) = 1 - (.486)^3 = .8852.$$

1.1   bonus: computation

There's another way to compute the ruin probability given different initial starting capitals and targets. It's possible to simulate the gambler's game with some simple code. If you run the simulation many times, you can count how many times the gambler "won" and reached the target amount, and "lost," or was ruined. The ratio of these two numbers will be an approximation of the ruin probability.

A graph of 100 possible "paths" a gambler's wealth can take is shown below.[2] As in our problem, the gambler starts with 10 dollars and stops if she reaches 80 dollars.

---

[2]The graph was generated from code supplied by Moritz Voss. Notice how some paths are very long. This might give the gambler the impression she has a chance, even though her probability of reaching the target amount is actually very low. I imagine casinos like this property of the game.

Possible evolution of the gambler´s gain
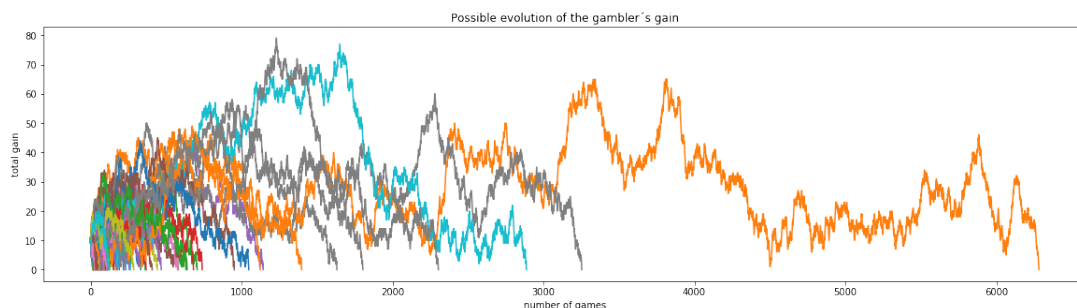
I will not reproduce the code. It only takes a little python to replicate.

2 Week 1

2.1 Lec 1 and 2

Definition 2.1. Markov Chain Let S be a discrete set. A Markov chain is a sequence of random variables, $X_0, X_1, X_2, \ldots$ takings values in S with the property:

$$P(X_{n+1} = j \mid X_0 = i_0, X_1 = i_1 \ldots X_n = i_n)$$
$$= P(X_{n+1} = j \mid X_n = i)$$

For all $i_0 \ldots i_{n-1}$ in S and $n \geq 0$. S is the state space of the Markov chain.

In words, this means the next random variable is influenced only by the previous one. If we are given a Markov chain and asked to compute the value of X_{100}, for instance, we only need to know the value of X_{99}.

A concrete example helps. Imagine a sequence of random variables represents my mood on a given day. If the sequence is a Markov chain, my mood on day 100 is only influenced by my mood on day 99.

Definition 2.2. Time Homogeneous Markov Chain A time homogeneous Markov chain is one where:

$$P(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i)$$

For all $i, j \in S$ and $n \geq 0$.

6

Definition 2.3. Transition Probability Matrix We define a transition probability matrix as $P = (P_{i,j})$, $i, j \in S$ where

$$P_{i,j} = P(X_1 = j \mid X_0 = i)$$

Example 2.1 (Weather model). Let (X_n), $n \geq 0$ be a time homogeneous Markov chain with state space $S = \{1, 2, 3\}$ where $1 = \text{rainy}$, $2 = \text{sunny}$, $3 = \text{cloudy}$. X_n corresponds to the weather on day n.

We define a transition probability matrix:

$$P = \begin{bmatrix} .2 & .6 & .2 \\ .1 & .8 & .1 \\ .1 & .6 & .3 \end{bmatrix}$$

The i, jth entry gives us the transition probability of moving to j given i. All the entries are condition probabilities. In light of this, the fact all the rows sum to 1 is unsurprising.

We can represent the transition probability matrix with a Markov chain diagram.
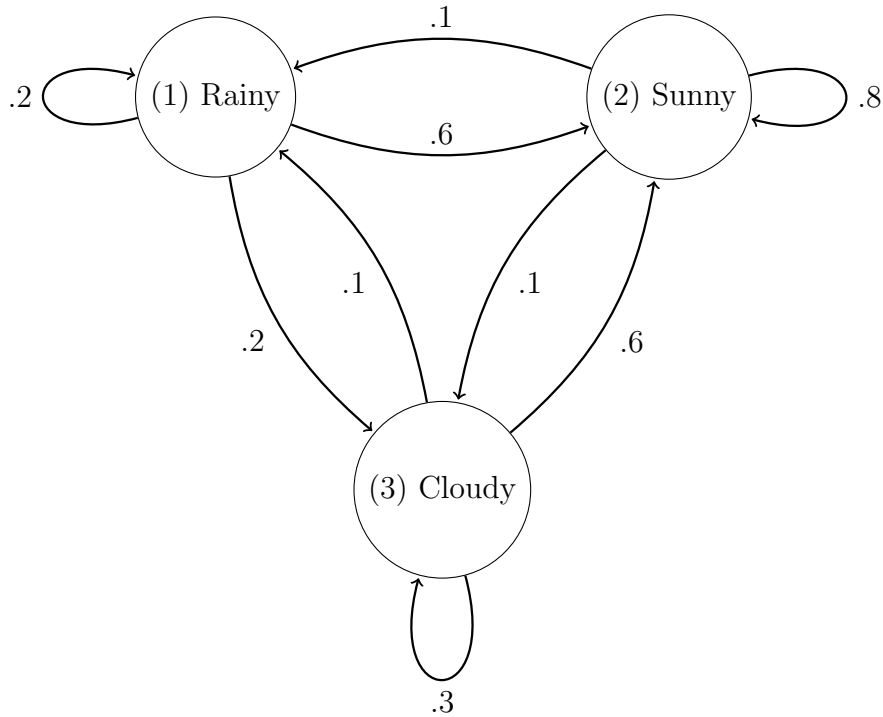
Each edge is labeled with the probability we transition from one state to another. Loops are labeled with the probability we stay in the same state.

Definition 2.4. Stochastic Matrix A stochastic matrix is a square matrix, $P = P_{i,j}$, $i, j \in S$ which satisfies:

1. $0 \le P_{i,j} \le 1$, $\forall i, j$.

2. $\sum_{j \in S} P_{i,j} = 1$, $\forall i$.

If we have a matrix that describes our transition probabilities, what about our initial distribution? How do we know what we are transitioning from?

Definition 2.5. Transition Vector Let (X_n), $n = 0, 1, \ldots$ be a Markov chain. The distribution of X_0 is the initial distribution of the entire chain. X_0's distribution is written as:

$$\alpha^T = (\alpha_1, \alpha_2, \alpha_3 \ldots)$$

Where $P(X_0 = i) = \alpha_i$. We might also call α a probability vector, since $0 \leq \alpha_i \leq 1 \;\forall i$ and $\sum_{i \in S} \alpha_i = 1$.

2.2   Lec 3

Example 2.2. Weather Model Computation What is $P(X_2 = 3 \mid X_0 = 2)$?

We rewrite in terms of one-step probabilities.

$$
\begin{aligned}
P(X_2 = 3 \mid X_0 = 2) &= \sum_{k \in S} P(X_2 = 3, X_1 = k \mid X_0 = 2) \text{(by LTP)} \\
&= \sum_{k \in S} \frac{P(X_2 = 3, X_1 = k, X_0 = 2)}{P(X_0 = 2)} \\
&= \sum_{k \in S} \frac{P(X_2 = 3, X_1 = k, X_0 = 2)}{P(X_0 = 2)} * \frac{P(X_1 = k, X_0 = 2)}{P(X_1 = k, X_0 = 2)} \\
&= \sum_{k \in S} \frac{P(X_2 = 3, X_1 = k, X_0 = 2)}{P(X_1 = k, X_0 = 2)} \frac{P(X_1 = k, X_0 = 2)}{P(X_0 = 2)} \\
&= \sum_{k \in S} P(X_2 = 3 \mid X_1 = k, X_0 = 2) P(X_1 = k \mid X_0 = 2) \\
&= \sum_{k \in S} P(X_2 = 3 \mid X_1 = k) P(X_1 = k \mid X_0 = 2) \\
&\quad \text{we ignore } X_0 = 2 \text{ by the Markov property} \\
&= \sum_{k \in S} P_{k,3} P_{2,k} = P_{2,3}^2
\end{aligned}
$$

Example 2.3. Joint Probabilities There's a multiplication rule for the probability of a certain sequence in a time-homogeneous Markov chain. Let $i, j, k \in S$.

$$
\begin{aligned}
P(X_1 = i, X_2 = j, X_3 = k) &= P(X_1 = i) P(X_2 = j \mid X_1 = i) P(X_3 = k \mid X_1 = i, X_2 = j) \\
&= P(X_1 = i) P_{i,j} P_{j,k}
\end{aligned}
$$

Note that $P(X_1 = i) = \sum_{m \in S} P(X_1 = i \mid X_0 = m) = (\alpha^T P)_i$. Substituting, we get:

$$= (\alpha^T P)_i P_{i,j} P_{j,k}$$

Proposition 2.1. Chapman-Kolmogorov equation Let (X_n) be a stochastic process and $i, j, k \in S$. Then,

$$P(X_{n+m} = j \mid X_0 = i) = \sum_{k \in S} P(X_m = k \mid X_0 = i) P(X_n = j \mid X_0 = k)$$

$$\iff P_{i,j}^{n+m} = \sum_{k \in S} P_{i,k}^m P_{k,j}^n$$

$$\iff P^{n+m} = P^n P^m$$

(We will explore the proof to this later. Also, take a look at the last line. Check how it relates to the second to last line.)

Definition 2.6. Marginal distribution Notice for all $n \geq 1$,

$$P_\alpha(X_n = j) = \sum_i P_\alpha(X_n = j \mid X_0 = i) P_\alpha(X_0 = i)$$

$$= \sum_i \alpha_i^T (P^n)_{i,j} = (\alpha^T P^n)_j$$

Then, we say the marginal distribution of X_n is

$$\mu_{X_n}^T = \alpha^T P^n \text{where} \qquad (\mu_{X_n})_j = P_\alpha(X_n = j)$$

Definition 2.7. Joint Distributions Let (X_n) be a time homogeneous Markov chain. For all $k \in \mathbb{N}$, $0 \leq n_1 \leq \cdots \leq n_k$, and all states $i_1, \cdots \in S$,

$$P_\alpha(X_{n_1} = i_1, X_{n_2} = i_2 \ldots X_{n_k} = i_k) = (\alpha^T P^{n_1})_{i_1} (P^{n_2 - n_1})_{i_1, i_2} \ldots (P^{n_k - n_{k-1}})_{i_{k-1}, i_k}$$

Hence, P and α completely determine a Markov chain.

3 Week 2

3.1 Lec 1

Theorem 3.1. Markov property (different statement): For all $m < n$,

$$P(X_{n+1} = j \mid X_0 = i_0, \ldots X_{n-m} = i)$$
$$= P(X_{n+1} = j \mid X_{n-m} = i)$$
$$= P(X_{m+1} = j \mid X_0 = i) = (P^{m+1})_{i,j}$$

for all $i, j, i_0 \ldots i_{n-m-1}$.

We want to classify the states of a Markov chain. First, we introduce some notation. Let,

$$P_x(X_n = j) = P(X_n = j \mid X_0 = x)$$

Also, recall an identity for conditional expectation:

$$\mathbb{E}[X_n \mid X_0 = x] = \sum_{k \in S} k P(X_n = k \mid X_0 = x)$$
$$= \sum_{k \in S} k P_{x,k}^n$$

Definition 3.1 (Return Times). Given a Markov chain $(X_n)_{n=0,1\ldots}$ let $T_y = \min\{n \geq 1 : X_n = y\}$ be the time of the first return to states $y \in s$. If $X_n \neq y$ for all $n \geq 1$, set $T_y = +\infty$.

We can do more with this notion of return times.

Let $\rho_{yy} = P_y(T_y < \infty)$ be the return probability. In words, it is the probability we return to state y after starting at y. Likewise, $\rho_{xy} = P_x(T_y < \infty)$ is the probability we go from x to y in a finite number of steps.

Definition 3.2 (**Stopping times**). We say a random variable T is a stopping time with respect to a Markov chain $(X_n)_{n=0,1\ldots}$ if for each $n \geq 0$ the occurrence of $\{T = n\}$ can be determined by the values of the process $X_0, X_1, \ldots X_n$ up to time n.

In words, we can't 'look into the future' to determine if our random variable takes on a value or not.

Example 3.1. Let $T_y = \min\{n \geq 1 : X_n = y\}$. T_y is a stopping time. This is because to determine whether $T_y = n$ we just need to check if $X_1 \neq y, X_2 \neq y, \ldots, X_n = y$.

A non-example of a stopping time is

$$U_y = \max\{n \geq 1 : X_n = y\}$$

If we want to determine if $U = n$, we need to look at all the values beyond U_n and ensure they're not equal to y. This disqualifies it from being a stopping time.

Theorem 3.2 (Strong Markov Property). Let $(X_n)_{n=0,1\ldots}$ be a markov chain with transition matrix P and let T be a stopping time.

Given $T = n$ and $X_T = y$, any other information about $X_0, \ldots, X_{T-1}$ is irrelevant for predicting the future. $(X_{T+k})_{k=0,1,\ldots}$ behaves like a Markov chain with initial state y.

A consequence is that:

$$P(X_{T+1} = z \mid X_T = y, T = n) = P_{yz}$$

Lemma 3.3. Let $T^1 = T_y$, and for $k \geq 2$ let

$$T_y^k = \min\{n > T_y^{k-1} : X_n = y\}$$

be the time of the kth return to y. Then,

$$P_y(T_y^k < \infty) = (P_y(T_y < \infty))^k = \rho_{yy}^k$$

12

The nice thing about the strong Markov property is we can 'split' Markov chains at random times (times set by random variables). With the original Markov property, we needed a deterministic time (like n = 5) in order to split the chain and disregard the past.

3.2   Lec 2

we want to classify the different states in space S of a Markov chain.

Definition 3.3 (Recurrence and Transience). .

1. States are **recurrent** if $\rho_{yy} = P_y(T_y < \infty) = 1$

2. States are **transient** if $\rho_{yy} = P_y(T_y < \infty) < 1$.

3. States are **positive recurrent** if $\mathbb{E}_y[T_y] < \infty$.

We can think of recurrent states as those we'd visit an infinite number of times, and transient states as those we would not. We formalize this in a lemma.

Lemma 3.4. .

1. If $x \in S$ is recurrent, then $P(X_n = x$ for infinitely many $n) = 1$

2. If $x \in S$ is transient, then $P(X_n = x$ for infinitely many $n) = 0$.

Here is another kind of space in S.

Definition 3.4 (Absorbing states). $x \in S$ is absorbing if $P_{i,i} = 1$. An entire chain is called absorbing if it has at least one absorbing state.

We also define some terms so we can talk about how different states are related.

Definition 3.5 (Communicating). .

1. $x \in S$ communicates with $y \in S$ if $\rho_{xy} = P_x(T_y < \infty) > 0$. For short, we write $x \to y$ to say x communicates with y.

2. We say x, y communicate if $x \to y, y \to x$.

Note $x \to y \iff \exists n > 0$ such that $P_{xy}^n > 0$.

Lemma 3.5. .

1. if $x \to y, y \to z$, then $x \to z$. (transitivity)

2. if $\rho_{xy} > 0$ and $\rho_{yx} < 1$, then x is transient. In particular, if $x \to y$ but $y \nrightarrow x$, then x is transient.

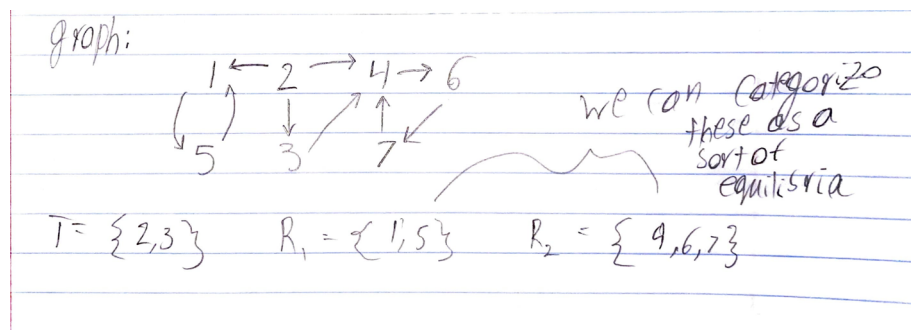3. if x is recurrent and $x \to y$, then y is recurrent.

3.3 Lec 3

Definition 3.6 (**Set Definitions**). .

1. $A \subset S$ is called closed if for all $i \in A$ and $j \in S$ it holds if $i \to j$ then $j \in A$.

2. $A \subset S$ is called irreducible if for all $i, j \in A$ it holds $i \to j$ and $j \to i$. In other words, i, j communicate.

Theorem 3.6. If $C \subset S$ is a finite, closed, and irreducible subset, then all of its states are recurrent.

Theorem 3.7 (State decomposition theorem). If the state space S is finite, then S can be written as a disjoint union $T \cup R_1 \cup \cdots \cup R_k$ where T is a set of transient states and $R_1 \dots R_k$ are closed irreducible sets of recurrent states.

Example 3.2. .

Note '2' seems to be a pivotal state.

Definition 3.7.
$$N(y) := |\ \{n \geq 1 : X_n = y\}\ |$$
This is the number of visits to y for all $n \geq 1$.

Lemma 3.8.

$$\mathbb{E}_x[N(y)] = \frac{\rho_{xy}}{1 - \rho_{yy}} = \sum_{n=1}^{\infty} P_{xy}^n$$

if this expectation is infinite, then y is recurrent.

Corollary 3.8.1. y is recurrent $\iff \sum_{n=1}^{\infty} P_{yy}^n = +\infty$

4 Week 3

4.1 Lec 1

There is another corollary of the previous lemma.

Corollary 4.0.1. If $y \in S$ is transient, then

$$\lim_{n \to \infty} P_{xy}^n = 0$$

$\forall x \in S$.

Definition 4.1 (Stationary distribution). Let there be a vector, π such that
$$\pi^T P = \pi^T$$
or in other words, $\pi_j = \sum_{i \in S} \pi_i P_{ij}$.

We also know that $\lim_{n \to \infty} P^n$, if it exists, is a stationary distribution.[3] Note π is just a left eigenvector with eigenvalue 1. We will explore this connection later in the notes.

---

[3]is there a proof for this? Why don't I have it as a separate lemma?

Example 4.1. Solving for the stationary distribution is solving for π in $\pi^T P = \pi^T$.. Suppose $\pi^T = [\pi_1, \pi_2, \pi_3]$. Then we'll get:

$$P_{1,1}\pi_1 + P_{2,1}\pi_2 + P_{3,1}\pi_3 = \pi_1$$
$$P_{1,2}\pi_1 + P_{2,2}\pi_2 + P_{3,2}\pi_3 = \pi_2$$
$$P_{1,3}\pi_1 + P_{2,3}\pi_2 + P_{3,3}\pi_3 = \pi_3$$

However, because we're solving for a probability vector, all the elements of π must add to 1. This means we introduce the condition $\pi_1 + \pi_2 + \pi_3 = 1$. Solving the four equations yields π.

Definition 4.2 (**Doubly stochastic matrix**). A matrix is called doubly stochastic if its columns also sum to 1. In other words, if $\sum_{i \in S} P_{ij} = 1$ for a fixed j.

Lemma 4.1. Let $(X_n)_{n=0,1\ldots}$ be a Markov chain on the state space $S = \{1, \ldots, N\}$ with doubly stochastic transition matrix P. Then the uniform distribution

$$\pi_i = \frac{1}{N} \qquad\qquad \forall i \in S$$

is a stationary distribution. [4]

Definition 4.3 (**Detailed balance condition**). Let P be the transition matrix of a Markov chain with state space S. We say that a probability vector λ satisfies the detailed balance condition if

$$\lambda_i P_{ij} = \lambda_j P_{ji} \qquad\qquad \forall i, j \in S$$

Proposition 4.1. If a probability distribution λ satisfies the detailed balance condition, then λ is a stationary distribution.

---

[4]the proof of this is the solution to problem 1.3.

Definition 4.4 (Birth and death chains). There are Markov chains with $S = \{l, l+1, \ldots, r-1, r\}$ and transition matrix P such that

$$P_{xy} = 0 \qquad \text{for } |x - y| > 1$$

and

$$
\begin{aligned}
P_{x,x+1} &= p_x & \text{for } x < r \\
P_{x,x-1} &= q_x & \text{for } x > l \\
P_{x,x} &= 1 - p_x - q_x & \text{for } l \leq x \leq r
\end{aligned}
$$

Example 4.2. Here is a 3x3 transition matrix for a birth and death chain.

$$
\begin{bmatrix}
1 - p_x & p_x & 0 \\
q_x & 1 - p_x - q_x & p_x \\
0 & q_x & 1 - q_x
\end{bmatrix}
$$

Proposition 4.2. Every birth and death Markov chain with $P_{xy} \neq 0$ for all $x, y \in S$ where $|x - y| = 1$ has a stationary distribution given by

$$\pi_{l+i} = \pi_l * \frac{\prod_{k=1}^{i} P_{l+k-1,l+k}}{\prod_{k=1}^{i} P_{l+k,l+k-1}} \qquad \forall i \in \{1, \ldots, r - l\}$$

5

4.2 Lec 2

Proposition 4.3. All Markov chains with a finite state space have at least one stationary distribution. The justification for this is the eigenvalue connection.[6]

---

[5]this is so strange. They use the products in the numerator and denominator but it's just p_x^i and q_x^i

[6]I'd like to investigate this. Perhaps I'll put the proof in.

Proposition 4.4. An irreducible Markov chain has at most one stationary distribution.

The conjunction of these two propositions gives us the following:

Corollary 4.1.1. A finite irreducible Markov chain has a unique, positive stationary distribution π.

We're on our way to understanding the limiting behavior of Markov chains. But first, we need the notion of periodicity.

Definition 4.5. The period of a state i is:

$$d(i) = \gcd\{n \geq 1 : P_{ii}^n > 0\}$$

A Markov chain (as a whole) is aperiodic if all states have period 1.

This definition is weird. The period of a state is *not* the length of the shortest cycle that includes it on a transition graph. Nor is it the average number of steps to traverse a cycle on the transition graph. Instead, it's greatest common devisor of the set of lengths of cycles you can make with the state.

A consequence of the greatest common devisor stipulation is that any state that can be revisited in a consecutive number of steps has period 1. For instance, if I can get from i to i in 4 steps and 5 steps, then the period of i is 1.

Lemma 4.2. Three mini-lemmas.

1. If $P_{ii} > 0$, the $d(i) = 1$.

2. If x and y communicate with each other, then $d(x) = d(y)$. (Lemma 1.17 in Durrett)

4.3 Lec 3 (limit theorems for Markov chains)

We're going to refer to these assumptions by the associated abbreviations.

- I: Markov chain is irreducible
- A: Markov chain is aperiodic
- R: All states are recurrent

- D: There exists a stationary distribution, π.

Theorem 4.3. (**Convergence theorem for Markov chains**) Say I, A, D are true. then

$$\lim_{n\to\infty} P_x[X_n = y] = \lim_{n\to\infty} P_{xy}^n = \pi_y$$

for all $x \in S$.

Denote the number of visits to y up to time n as

$$N_n(y) :=| \{k \in \{1, \ldots, n : X_k = y\} |$$

Theorem 4.4 (**Asymptotic frequency**). Suppose I is true. Then

$$\lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1}{E_y[T_y]}$$

for all $y \in S$.

Theorem 4.5. Suppose I and D are true. Then

$$\pi_y = \frac{1}{E_y[T_y]}$$

for all $y \in S$.

Theorem 4.6 (**Strong law of large numbers for Markov chains**). Suppose I and D are true. Let $f : \mathbb{R} \to \mathbb{R}$ be a function such that $\sum_{x \in S} | f(x) | \pi_x < \infty$. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(X_i) = \sum_{x \in S} f(x)\pi_x$$

Alternatively, let Z be a discrete random variable with values in S and distribution π. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(X_i) = E[f(Z)]$$

Compare this with the strong law of large numbers. This is the basis for the Markov Chain Monte Carlo method of estimating the expectation of a random variable.

The following theorem is helpful because it does not assume a finite state space.

Theorem 4.7. Let $(X_n)_{n=0,1...}$ be an irreducible Markov chain. Then, the following statements are equivalent.

1. There exists a unique stationary distribution

2. One state is positive recurrent

3. All states are positive recurrent

All states are automatically positive recurrent for a finite Markov chain.

5 Week 4

5.1 only substantive lecture in the week

We must review the exponential distribution. We say a random variable X is exponentially distributed by: $X \sim \text{Exp}(\lambda)$ with $\lambda > 0$.

we recall the density function of an exponential.

$$f_x(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The cumulative distribution function (CDF) is:

$$F_x(x) = P(X \leq x) = \int_{-\infty}^{x} f_x(y) dy$$
$$= 1 - e^{-\lambda x}$$
$$\forall x \geq 0$$

Additionally, it's helpful to recall the first and second moments.

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_x(x) dx = \frac{1}{\lambda}$$

and

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_x(x) dx = \frac{2}{\lambda^2}$$

This helps us compute the variance.

$$\mathrm{Var}(x) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

An important property of exponential distributions is that they are memoryless. Mathematically, we describe this as:

$$P(X > t + s \mid X > t) = P(X > s)$$

We can prove this.

$$P(X > t + s \mid X > t) = \frac{P(X > t + s, X > t)}{P(X > t)}$$

We note, as events, $X > t \subset X > t + s$. Hence we rewrite as:

$$= \frac{P(X > t + s)}{P(X > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda(t)}}$$
$$= e^{-\lambda s} = P(X > s)$$

Proposition 5.1. The exponential distribution is the only continuous memoryless distribution.

We now explore the minimum of exponential distributions.

Proposition 5.2. Suppose we have $X \sim \mathrm{Exp}(\lambda)$ and $Y \sim \mathrm{Exp}(\mu)$ and they are independent. Let $M = \min\{X, Y\}$.

Claim (i): $M \sim \mathrm{Exp}(\lambda + \mu)$

Claim (ii): $P(M = X) = \frac{\lambda}{\mu + \lambda}$

Proof. (i)

$$P(M > x) = P(\min\{X, Y\} > x) = P(X > x, Y > x) = P(X > x)P(Y > x)$$
$$= e^{-\lambda x}e^{-\mu x} = e^{-x(\lambda+\mu)}$$

which is 1 minus the cdf of $\text{Exp}(\lambda + \mu)$

(ii)

$$P(M = X) = P(\min\{X, Y\} = X) = P(Y \geq X)$$

We can use a generalized form of the law of total probability to calculate this.

$$= \int_0^\infty P(Y \geq X \mid X = x)P(X = x)dx$$
$$= \int_0^\infty P(Y \geq X \mid X = x)f_x(x)dx$$
$$= \int_0^\infty P(Y \geq x)\lambda e^{-\lambda x}dx$$
$$= \int_0^\infty e^{-\mu x}\lambda e^{-\lambda x}dx = \int_0^\infty \lambda e^{-x(\mu+\lambda)}$$
$$= \lambda[\frac{-1}{\lambda+\mu}e^{-x(\mu+\lambda)}]_0^\infty = \frac{\lambda}{\lambda+\mu}$$

$\square$

6   Week 5

6.1   Lec 1

Proposition 6.1. Suppose $X_1, \ldots, X_n$ are $\text{Exp}(\lambda_i)$ and $M = \min\{X_1, \ldots X_n\}$.

1. for $t > 0$, $P(M > t) = e^{-t(\lambda_1+,\ldots,+\lambda_n)}$. In words, M has an exponential distribution with parameter $\lambda_1 + \cdots + \lambda_n$.

2. for $k = 1, \ldots, n$, $P(M = X_k) = \frac{\lambda_k}{\lambda_1+\cdots+\lambda_n}$.

Also, the random variable M and the index of the random variable that is the minimum are independent.

Proposition 6.2. Let $X_1, X_2, \ldots X_n$ be independent $\text{Exp}(\lambda)$ random variables. Set $T_n = X_1 + \cdots + X_n$. Then T_n has a gamma distribution with parameters n and λ. The density function is:

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

$\forall t > 0$. The mean and variance of T_n are:

$$\mathbb{E}[T_n] = \frac{n}{\lambda} \qquad\qquad \text{Var}(T_n) = \frac{n}{\lambda^2}$$

We would say $T_n \sim \Gamma(n, \lambda)$.

With this background in place we can finally start to motivate the poisson process. We want to count instances of an event in continuous time. In this sense, we'll say a poisson process is a counting process.

Definition 6.1 (**Counting process**). A counting process $(N_t)_{t \geq 0}$ is a collection of non-negative integer-valued random variables such that if $0 \leq s \leq t$ then $N_s \leq N_t$.

Some notes about counting processes:

- Normally, $N_0 = 0$, since we're starting with zero instances of the event.

- N_k denotes the number of occurences by time k.

- for all $0 \leq s < t$, $N_t - N_s$ denotes the number of occurences in the time interval $(s, t]$.

Definition 6.2 (**Poisson process**). A poisson process with $\lambda > 0$ is a continuous time stochastic process with the following properties:

1. $N_0 = 0$

2. (independent increments)$\forall n \in \mathbb{N}$, $0 \leq t_1 < t_2 < \cdots < t_n$, the random variables $N_{t_2} - N_{t_1}, N_{t_3} - N_{t_2}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent.

3. (stationary increments) for all $0 \leq s < t$ the random variable $N_t - N_s$ is a poisson distribution with parameter $\lambda(t - s)$.

4. (redundant property) the function $t \mapsto N_t$ is a right-continuous step function with jump size 1.

property 4 just means the poisson process jumps by at most 1 (check this).

Several consequences flow from these properties.

- $N_t = N_t - 0 = N_t - N_0 \sim \text{Pois}(\lambda t) \ \forall t > 0$.

- $\mathbb{E}[N_t] = \lambda t$ (λ is called the arrival rate)

We must note that N_t is not independent of N_s. Only the difference is independent from other, non-overlapping differences.

6.2 Lec 2

Recall N_i's are poisson distributed with parameter λ_i where i is the time interval. We can see this by seeing:

$$N_t = N_t - N_s + N_s - N_0 \sim \text{Pois}(\lambda(t-s) + \lambda s) = \text{Pois}(\lambda t)$$

Example 6.1. Students come to the library at the rate 30 per hour starting at 6am. What is the probability more than 65 arrive between 9am and 11am?

<u>solution</u> Let N_t = the number of students who have arrived at the library by time t. $(N_t)_{t \geq 0}$ is a poisson process with $\lambda > 0$. Stipulate time is measured in hours. Then $\lambda = 30$ and $t_0 = $ 6am, $t_3 = $ 9am and $t_5 = $ 11am.

In these terms, our problem becomes calculating $P(N_5 - N_3 > 65)$. By stationary increments, we know $N_5 - N_3$ is poisson distributed with $\lambda = 2(30)$. Hence, we compute $P(\text{Pois}(60) > 65)$.

It's easy enough to plug this into a calculator, or wolfram, or whatever, but it's worth manipulating it a little to see what different forms this expression can take. We know $N_2 \sim \text{Pois}(60)$, so we can replace it.

$$P(\text{Pois}(60) > 65) = P(N_2 > 65) = 1 - P(N_2 \leq 65)$$
$$= 1 - \sum_{k \leq 65} P(N_2 = k)$$

we know the PMF of N_2 since it is $\text{Pois}(60)$

$$= 1 - \sum_{k \leq 65} e^{-60} \frac{60^k}{k!}$$

In the end, the probability comes out to be $\approx .235$.

Example 6.2. Visitors to a website start arriving at 10am every morning and come at a rate of 10 per hour according to a Poisson process.

1. What's the probability you will get exactly 18 visitors by noon and exactly 70 by 5pm?

2. Suppose you received exactly 18 visitors by noon. Find the probability you will get exactly 70 visitors by 5pm.

<u>Solution</u> The first question is asking for the probability of two events. We can express this as: $P(N_2 - N_0 = 18, N_7 - N_2 = 52)$. We set $N_7 - N_2 = 52$ because in order to get 70 visitors by 5, we need to get 18 from 10 to noon, and 52 from noon to 5. By independent increments, we split up the probability.

$$= P(N_2 - N_0 = 18)P(N_7 - N_2 = 52) = P(\text{Pois}(20) = 18)P(\text{Pois}(50) = 52)$$

The Poisson distributions replace the differences due to the stationary increments property

$$= e^{-20}\frac{20^{18}}{18!} * e^{50}\frac{50^{52}}{52}$$
$$\approx .084 * .053$$

The second question asks for a conditional probability. We want to find $P(N_7 = 70 \mid N_2 = 18)$. Approach this by using the definition of conditional probability.

$$= \frac{P(N_7 = 70, N_2 = 18)}{P(N_2 = 18)}$$
$$= P(N_7 - N_2 = 52)$$
$$= P(N_5 = 52) = P(\text{Pois}(50) = 52)$$
$$\approx .053$$

6.3 Bonus Section (indicator functions and expectation)

Indicator functions help us convert between events and random variables. From an event, A, we want to get a random variable 1_A that takes the value 1 if A obtains and 0 otherwise.

$$1_A = \begin{cases} 1 & A \\ 0 & \neg A \end{cases}$$

We note $\mathbb{E}[1_A] = P(A)$.

To go the other way around, from random variables to events, we notice for a random variable X,

$$X = \sum_x x * 1[X = x]$$

Think about what a random variable is. It takes on certain values with a given probability. Hence, we can split up all the possible values and pick the one that the random variable actually takes (this is accomplished by $1[X = x]$). We multiply all the other values by zero, add it to the value actually obtained, and we get the same result as a standard random variable.

Proposition 6.3. This allows us to put a conditional expectation into something more manageable.

$$\mathbb{E}[X \mid B] = \mathbb{E}[\sum_x x1[X = x] \mid B]$$

by linearity of expectation we can pull out the sum

$$= \sum_x x \mathbb{E}[1[X = x] \mid B]$$

$$= \sum_x x \frac{\mathbb{E}[1[X = x \cap B]]}{P(B)}$$

$$= \frac{1}{P(B)} \mathbb{E}[\sum_x x 1[X = x] 1[B]]$$

$$= \frac{\mathbb{E}[X 1[B]]}{P(B)}$$

Example 6.3. Let X and Y be two independent exponentially distributed random variables with parameters $\lambda > 0$ and $\mu > 0$, respectively. Show

$$\mathbb{E}[\min\{X, Y\} \mid X < Y] = \frac{1}{\lambda + \mu}$$

<u>Solution</u> From the prior proposition, we know

$$\mathbb{E}[\min\{X, Y\} \mid X < Y] = \frac{\mathbb{E}[\min\{X, Y\}] 1[X < Y]}{P(X < Y)}$$

We proceed by looking at the numerator.

$$\mathbb{E}[\min\{X,Y\}]1[X<Y]P(X<Y)] = \mathbb{E}[X1[X<Y]]$$

$$= \int\int_{\mathbb{R}} x1[X<Y]\mu\lambda e^{-x\lambda}e^{-y\mu}dydx$$

$$= \int_0^\infty \int_0^\infty x1[X<Y]\mu\lambda e^{-x\lambda}e^{-y\mu}dydx$$

$$= \int_0^\infty x\lambda e^{-x\lambda}dx \int_x^\infty \mu e^{-y\mu}dy$$

$$= \int_0^\infty x\lambda e^{-x\lambda} * e^{-x\mu}dx$$

$$= \int_0^\infty x\lambda e^{-x(\lambda+\mu)}dx$$

$$= \frac{\lambda}{\lambda+\mu}(\lambda + \mu e^{-x(\lambda+\mu)}dx$$

$$= P(X<Y)\mathbb{E}[Z]$$

where $Z \sim \mathrm{Exp}(\mu + \lambda)$. Now we can plug it into the original equation.

$$\mathbb{E}[\min\{X,Y\} \mid X < Y] = \frac{P(X<Y)\mathbb{E}[Z]}{P(X<Y)} = \mathbb{E}[Z] = \frac{1}{\lambda+\mu}$$

Add the correlation problem

6.4 Lecture 3

There is a different way to look at Poisson processes. We "decompose" them so to speak into the arrival times of events, and the times in between arrivals.

Proposition 6.4. Let $\tau_1, \tau_n \ldots$ be a sequence of iid exponential random variables with parameter $\lambda > 0$. Also, let $T_n = \tau_1 + \cdots + \tau_n$ and $T_0 = 0$ for $n \geq 1$. For $t \geq 0$ let

$$N_t = \max\{n \geq 0 : T_n \leq t\} = \sum_{k=1}^n 1_{(0,t]}(T_k)$$

Then N_t is a poisson process with parameter $\lambda > 0$.

T_i is the arrival time of the ith event in the poisson process. For instance, if the 8th patron enters a store according to a poisson process at time 40, then $T_8 = 40$. τ_i represents the time between T_i and T_{i-1}. If $T_7 = 30$ then $\tau_8 = 10$.

This leads us to the next proposition.

Proposition 6.5. Let N_t be a poisson process with parameter $\lambda > 0$. For some $t > 0$ let $N_t = n$ and have $T_1 \dots T_n$ be the corresponding arrival times.

Also, let $U_1 \dots U_n$ be iid uniform random variables on $[0, t]$. Then:

$$(T_1, \dots, T_n \mid N_t = n) = (U_{(1)}, \dots, U_{(n)})$$

In distribution, where $U_{(i)}$ is the ith order statistic of the iid uniform random variables.[7]

The proposition says something powerful. If we know the value of N_t at some time t, then we can say something useful about the joint distribution of arrival times. In particular, we know the joint distribution is identical to the joint distribution of n uniform random distributions on the interval $[0, t]$. We know that that is:

$$f(u_1, u_2, \dots, u_n) = \begin{cases} \frac{n!}{t^n} & \text{if } 0 \le t_1 \le t_2 \le \dots \le t_n < t \\ 0 & \text{otherwise} \end{cases}$$

Here are some examples of general problem-solving with poisson processes. Note they don't necessarily make us of the propositions we just covered.

Example 6.4. The times when goals are scored in hockey can be modeled as a Poisson process. Assume such a process where the average time between goals is 15 minutes.

1. In a 60-minute game, find the probability that a fourth goal occurs in the last 5 minutes of the game

---

[7]The statement of this proposition is altered a bit from Voss' class materials. He expresses the thought in words several times before givng a more concise mathematical formulation. This proposition is 2.10 in the slides (and probably in the Durrett book)

2. Assume at least three goals are scored in a game. What is the mean time of the third goal

<u>Solution</u> 1)

We're seeking $P(55 < T_4 < 60)$. Recall T_4 is gamma distributed.[8] This means we can evaluate the expression directly as:

$$\int_{55}^{60} (1/15)e^{-t/15}\frac{(t/15)^3}{3!}dt \approx .068$$

2) We're looking for: $\mathbb{E}[T_3 \mid T_3 < 60]$. We recall a formula for conditional expectation covered in section 6.3. Hence,

$$\mathbb{E}[T_3 \mid T_3 < 60] = \frac{1}{P(T_3 < 60)}\mathbb{E}[T_3 1_{T_3<60}]$$

We know $P(T_3 < 60)$ is evaluable because its gamma distributed. Looking closely at the expectation, we realize for all $T_3 > 60$ the indicator function will be 0. Hence, we can 'cut off' the expectation at 60.

$$= (\int_0^{60} \frac{1}{15}e^{-t/15}\frac{(t/15)^2}{2!}dt)^{-1} * \int_0^{60} t\frac{1}{15}e^{-t/15}\frac{(t/15)^2}{2!}dt$$

Going through the calculation, we see this comes out as:

$$\approx 33.46$$

Example 6.5. Starting at $t = 0$ patrons arrive at a restaurant according to a possion process with a rate of 20 customers per hour.

If 60 people arrive at the restaurant by time $t = 3$, find the probability that the 60th customer arrives in the interval $[2.9, 3]$.

<u>Solution</u> Formally, we're after $P(T_{60} \in [2.9, 3] \mid N_3 = 60)$. In words, this is the probability the 60th customer arrives between $t = 2.9$ and $t = 3$.

---

[8]This is because $T_n = \tau_1 + \tau_2 \ldots$ where τ is an exponential random variable. The sum of n iid exponential random variables is $\Gamma(n, \lambda)$

We use proposition 6.5 to say:

$$(T_{60} \in [2.9, 3]) = (U_{(60)} \in [2.9, 3])$$

in distribution. Hence,

$$P(T_{60} \in [2.9, 3] \mid N_3 = 60) = P(U_{(60)} \in [2.9, 3])$$
$$= P(U_{(60)} > 2.9)$$

The last step is valid since 3 is the end of the interval the U's are distributed on.

$$= 1 - P(U_{(60)} < 2.9)$$

Notice $U_{(60)}$ is an order statistc. This means that if $U_{(60)} < 2.9$, then all of the 60 iid U's must also be below 2.9.

$$= 1 - P(U_1 < 2.9, U_2 < 2.9, \ldots, U_60 < 2.9)$$
$$= 1 - (\int_0^{2.9} \frac{1}{3} dt)^{60}$$
$$= 1 - \frac{2.9^{60}}{3}$$
$$\approx .869$$

7 Week 6

The first lecture of week 6 was a continuation of the lecture 3 week 5. We did the same problems. The second lecture didn't happen because it was Veteran's day. I skip to Jacob's section.

7.1 Jacob's section

Example 7.1. Consider a series of payments of \$ 1000 at the random arrival times of a Poisson process with parameter $\lambda > 0$. Recall that if the continuously compounded interest rate is $r > 0$ the nthe present value at time 0 of

a payment of \$ 1000 at some future point in time $t > 0$ is given by \$ 1000 $*e^{-rt}$.

Demonstrate the expected present value at time 0 of the random series of payments until $T > 0$ is $1000\lambda(1 - e^{-rT})/r$.

<u>Solution</u> Let X be the random variable representing the present value of the future payments. For some T, there will be N_T payments of 1000 on the interval $[0, T]$. The value of X is just the sum of these payments discounted to the present. What's more, a payment arriving at T_k will be worth $1000e^{-rT_k}$ at time 0. Hence,

$$X = \sum_{k=1}^{N_T} 1000e^{rT_k}$$

Now, we use the law of total expectation to get the expected value of X. Recall the law is:[9]

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

Hence,

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} \mathbb{E}[\sum_{k=1}^{N_T} 1000e^{rT_k} \mid Y = n]P(Y = n)$$

The law of total expectation allows us to replace Y with any discrete random variable that takes positive natural values. Condition on $N_T = n$ since we're holding a time fixed and caluclating expected value from that.

$$= \sum_{n=0}^{\infty} \mathbb{E}[\sum_{k=1}^{n} 1000e^{rT_k} \mid N_T = n]P(N_T = n)$$

We know that N_T is a Poisson so we can replace.

$$= \sum_{n=0}^{\infty} \mathbb{E}[\sum_{k=1}^{n} 1000e^{rT_k} \mid N_T = n]\frac{(\lambda T)^n e^{-\lambda T}}{n!}$$

---

[9]There's a condition Y has to fulfill. Check this and include it.

Recall Proposition 6.5. We're going to use it to replace all the T_k's with uniform normals. Rigorously, we say:

let $\{A_1, A_2 \ldots A_n\} \sim \text{Uniform}(0, T)$.

Put them in order, like so: $A_{j_1} \le A_{j_2} \le \cdots \le A_{j_n}$

by proposition 6.5, we know $\{A_{j_k}\}_k = \{T_k \mid N_T = n\}$ in distribution. Now, we can make the swap.

$$\mathbb{E}[\sum_{k=1}^n 1000e^{rT_k} \mid N_T = n] = \mathbb{E}[\sum_{k=1}^n 1000e^{-rA_{jk}}] = \mathbb{E}[\sum_{k=1}^n 1000e^{-rA_j}](\text{by iid})$$

notice there is no dependence on k. We can pull out the relevant terms.

$$= n1000\mathbb{E}[e^{-rA_j}]$$
$$= 1000n\int_0^\infty e^{-ra}\text{PDF}(A_j = a)da$$

but remember, A_j is a normal on the interval $[0, T]$

$$= 1000n\int_0^T e^{-ra}\frac{1}{T}da$$
$$= 1000n\frac{1 - e^{-rT}}{rT}$$

Now we make the substitution

$$= \sum_{n=0}^\infty 1000n\frac{1 - e^{-rT}}{rT}\frac{(\lambda T)^n e^{-\lambda}}{n!} = 1000\frac{1 - e^{-rT}}{rT}\sum_{n=0}^\infty n\frac{(\lambda T)^n e^{-\lambda T}}{n!}$$

notice the sum is just the expectation of a Poisson with mean λT.

$$= 1000\frac{1 - e^{-rT}}{rT}\lambda T = 1000\lambda\frac{1 - e^{-rT}}{r}$$

as desired.

Example 7.2. Babies are born at a maternity ward according to a Poisson process with $\lambda > 0$.

The the babies' sexes are indpendent of each other (having a boy does not make a girl more likely). According to the UN, the global sex ration for boys to girls is $p = .519$.

How can the number of male and female births in the maternity ward be described?

<u>Solution</u> This is a strangely phrased question. Essentially, it's asking us to play around the with equation for a Poisson process while 'keeping track' of a single sex and to see what happens.

Let N_t be the the total number of babies born by time t and let M_t and F_t be the number of babies born by t with male and female sexes, respectively.

Let's compute the joint distribution of $P(M_t = m, N_t = m + f)$.

$$P(M_t = m, N_t = m + f) = P(M_t = m \mid N_t = m + f)P(N_t = m + f)$$

The last term is Poisson distributed so we hold off for now

We treat births like Bernoulli trials with $p = .519$. Hence, the probability of getting m successes in $m + f$ Bernoulli trials is

$$\binom{m + f}{m} P(\text{male})^m P(\text{female})^f$$

replace $P(M_t = m \mid N_t = m + f)$ with the previous expression.

$$= \binom{m + f}{m} P(\text{male})^m P(\text{female})^f P(N_t = m + f)$$

$$= \binom{m + f}{m} P(\text{male})^m P(\text{female})^f \frac{(\lambda t)^{m+f} e^{-\lambda t}}{(m + f)!}$$

$$= \frac{(m + f)!}{m! f!} (\lambda t)^m (\lambda t)^f \frac{1}{(m + f)!}$$

notice we can multiply the exponent of the exponential by 1, or $P(\text{male}) + P(\text{female})$.

$$= \frac{(t\lambda P(\text{male}))^m e^{-\lambda t P(\text{male})}}{m!} \frac{(t\lambda P(\text{female}))^f e^{-\lambda t P(\text{female})}}{f!}$$
$$= \text{Pois}(P(\text{male})\lambda) * \text{Pois}(P(\text{female})\lambda)$$

which is two Poisson distributions. This suggests that if we categorize different types of arrivals in a Poisson process, they are also Poisson distributed.

7.2 Lec 3

We will talk about thinning and superpositioning of Poisson processes

Proposition 7.1. Let $(N_t)_{t \geq 0}$ be a Poisson process with parameter $\lambda > 0$. Assume that each arrival, indpendent of other arrivals, is marked as a type k event with probability $p_k \in (0, 1)$, for $k = 1, 2, \ldots, n$, where $p_1 + p_2 + \cdots + p_n = 1$.

Let $(N^{(k)})_t$ be the number of type k events in $[0, t]$. Then, $(N_t^{(k)})_{t \geq 0}$ is a Poisson process with parameter $\lambda p_k > 0$.

Additionally, the processes,

$$(N_t^{(1)})_{t \geq 0}, \ldots, (N_t^{(n)})_{t \geq 0}$$

are independent. Each is called a thinned Poisson process. We also notice $N_t^{(1)} + N_t^{(2)} + \cdots + N_t^{(n)} = N_t$.

We can see a special case of this in birth example in the previous section.

Proposition 7.2. Assume that

$$(N_t^{(1)})_{t \geq 0}, \ldots, (N_t^{(n)})_{t \geq 0}$$

Are n indpendent Poisson processes with respective parameters $\lambda_1, \ldots, \lambda_n > 0$. Let

$$N_t = N_t^{(1)} + N_t^{(2)} + \cdots + N_t^{(n)}$$

Is a Poisson process with parameter $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

We call this the superpositioning process.

There are several ways to simulate a Poisson distribution.

<u>Method 1</u> This method relies on the idea a Poisson process is composed of stationary increments.

1. Split $[0, T]$ into n small subintervals of length $\Delta t = T/n$. Δt must be sufficiently small such that the probability of getting 2 occurrences from a Poisson random variable on it is very low.

2. create n iid Poisson random variables $\Delta N_1, \Delta N_2, \ldots, \Delta N_n$ with parameter $\lambda \Delta t$

3. Let $N_0 = 0$. For each $i = 0, \ldots, n-1$ let $N_{(i+1)\Delta t} = N_{i\Delta t} + \Delta N_{i+1}$

4. For all $i = 0, \ldots, n-1$ and each $t \in [i\Delta t, (i+1)\Delta t)$, set $N_t = N_{i\Delta t}$.

<u>Method 2</u> This method relies on the idea the interarrival times for a Poisson process are exponentially distributed.

1. Let $T_0 = 0$.

2. Generate iid exponential random variables $\tau_1, \tau_2, \ldots$, with parameter λ.

3. Let $T_n = \tau_1 + \cdots + \tau_n$ for $n = 1, 2, \ldots$.

4. For each $k = 0, 1, \ldots$ let $N_t = k$ for $T_k \leq t < \min\{T_{k+1}, T\}$.

<u>Method 3</u> This one relies on the fact the arrival times are normally distributed on the interval.

1. Simulate the total number of arrivals on $[0, T]$ from a Poisson distribution with parameters λT.

2. Generate $N_t = n$ iid random variables $U_1, \ldots, U_n$ uniformly distributed on $[0, T]$.

3. Sort the variables in increasing order (make order statistics) in order to get the arrival times.

4. For each $k = 0, 1 \ldots, n-1$ set $N_t = k$ for $T_k \leq t < T_{k+1}$ and set $N_t = n$ for $T_n \leq t < T$.

We're going to introduce a more realistic Poisson process. In this model, the λ changes as time varies. We might call this a nonhomogeneous Poisson process.

Definition 7.1 (Nonhomogeneous Poisson Process). A counting process is a nonhomogeneous Poisson process with intensity function $\lambda(t)$ if

1. $N_0 = 0$

2. (independent increments) For all $n \in \mathbb{N}, 0 \leq t_1 < t_2 < \cdots < t_n$ the random variables $N_{t_i} - N_{t_{i-1}}$ are independent.

3. For all $0 \leq s < t$ the random variable $N_t - N_s$ is Poisson distributed with parameter $\int_s^t \lambda(u) du$

We no longer have stationary distributions anymore. Furthermore, if λ is a constant we have a regular Poisson process.

Definition 7.2 (Compound Poisson Process). Let $(N_t)_{t \geq 0}$ be a Poisson process with $\lambda > 0$. Furthermore, let $Y_1, \ldots$ be a sequence of iid random variables with density function f_y, also independent of $(N_t)_{t \geq 0}$. Then,

$$C_t = \sum_{i=1}^{N_t} Y_i$$

for $t \geq 0$ is a compound Poisson process with jump-distribution f_y and jump intensity λ.

8 Week 7

8.1 Lecture 1

We will be finishing up Poisson processes and then moving onto renewal processes.

Recall what we've said about compound Poisson processes. We introduce and additional fact.

Fact If we have a compound Poisson process, C_t, then the following are true.

1. $C_0 = 0$

2. C_t has independent increments in the same way a regular Poisson process does.

3. The increments are stationary (though they are no long Poisson distributed)

Now, we have an example that makes use of these properties.

Example 8.1. Let $(C_t)_{t \geq 0}$ be a compound Poisson process. Show:

1. $\mathbb{E}[C_t] = \lambda t \mathbb{E}[Y_1]$

2. compute $\mathrm{Var}(C_t)$.

Solution First, in order to do this problem we must assume $\mathbb{E}[Y_1] < \infty, \mathbb{E}[Y_1^2] < \infty, \mathbb{E}[N] < \infty, \mathbb{E}[N^2] < \infty$.

1) We recall and apply the law of total expectation.

$$\mathbb{E}[C_t] = \mathbb{E}[\sum_{i=1}^{N_t} Y_i] = \sum_{k=0}^{\infty} \mathbb{E}[\sum_{i=1}^{N_t} Y_i \mid N_t = n] P(N_t = n)$$

We look at the conditional expectation and realize it fixes some n where the internal sum ends. Thus, we can replace a random sum with a deterministic sum.

$$= \sum_{k=0}^{\infty} \mathbb{E}[\sum_{i=1}^{n} Y_i] P(N_t = n)$$

We can bring the sum outside of the second expectation since the Y_i's are independent of n. The Y_i's are also iid, which allows us to replace them all with Y_1.

$$= \sum_{k=0}^{\infty} n\mathbb{E}[Y_1]P(N_t = n)$$

$$= \mathbb{E}[Y_1]\sum_{k=0}^{\infty} nP(N_t = n)$$

$$= \mathbb{E}[Y_1]\mathbb{E}[N] = \mathbb{E}[Y_1]\lambda t$$

As desired.

2) Computing the variance is a bit tougher. We recall that $\text{Var}(C_t) = \mathbb{E}[C_t^2] - \mathbb{E}[C_t]^2$. We already know $\mathbb{E}[C_t]^2$, so it remains to compute $\mathbb{E}[C_t^2]$. We use the law of total expectation again.

$$\mathbb{E}[C_t^2] = \sum_{k=0}^{\infty} \mathbb{E}[C_t^2 \mid N_t = k]P(N_t = k)$$

$$= \sum_{k=0}^{\infty} \mathbb{E}[(\sum_{i=1}^{N_t} Y_i)^2 \mid N_t = k]P(N_t = k)$$

from here, we use the law of conditional variance to calculate the conditional expectation. Recall for some random variable S,

$$\text{Var}(S \mid N_t = n) = \mathbb{E}[S^2 \mid N_t = n] - \mathbb{E}[S \mid N_t = n]^2$$

Now, we solve for the conditional expectation of the second moment substitute.

$$\mathbb{E}[C_t^2 \mid N_t = k] = \text{Var}(C_t \mid N_t = k) + \mathbb{E}[C_t \mid N_t = k]^2$$

$$= \text{Var}(\sum_{i=1}^{N_t} Y_i \mid N_t = k) + \mathbb{E}[\sum_{i=1}^{N_t} Y_i \mid N_t = k]^2$$

$$= \text{Var}(\sum_{i=1}^{k} Y_i) + \mathbb{E}[\sum_{i=1}^{k} Y_i]^2$$

$$= k\text{Var}(Y_1) + k^2\mathbb{E}[Y_i]^2$$

we plug in now.

$$\mathbb{E}[C_t^2] = \sum_{k=0}^{\infty} (k\text{Var}(Y_1) + k^2\mathbb{E}[Y_i]^2)P(N_t = k)$$
$$= \sum_{k=0}^{\infty} kP(N_t = k)\text{Var}(Y_1) + k^2P(N_t = k)\mathbb{E}[Y_1]^2$$

Bring the sums in

$$= \text{Var}(Y_1)\sum_{k=0}^{\infty} kP(N_t = k) + \mathbb{E}[Y_1]^2\sum_{k=0}^{\infty} k^2P(N_t = k)$$
$$= \text{Var}(Y_1)\mathbb{E}[N_t] + \mathbb{E}[Y_1]^2\mathbb{E}[N_t^2]$$

We now subtract $\mathbb{E}[C_t]^2$ to get the variance

$$\text{Var}(C_t) = \text{Var}(Y_1)\mathbb{E}[N_t] + \mathbb{E}[Y_1]^2\mathbb{E}[N_t^2] - \mathbb{E}[Y_1]^2\mathbb{E}[N_t]^2$$
$$= \text{Var}(Y_1)\mathbb{E}[N_t] + \mathbb{E}[Y_1]^2(\mathbb{E}[N_t^2] - \mathbb{E}[N_t]^2)$$
$$= \text{Var}(Y_1)\mathbb{E}[N_t] + \mathbb{E}[Y_1]^2\text{Var}(N_t)$$
$$= \lambda t\text{Var}(Y_1) + \mathbb{E}[Y_1]^2\lambda t$$
$$= \lambda t(\text{Var}(Y_1) + \mathbb{E}[Y_1]^2)$$
$$= \lambda t\mathbb{E}[Y_1^2]$$

Which finishes the problem.

Here is a proposition that allows us to pass from compound Poisson processes to regular ones. It is very similar to our proposition that details 'thinning' regular Poisson processes.

Proposition 8.1. Let $(C_t)_{t\geq0}$ be a compound Poisson process with jump intensity λ and jump distribution f_Y. Also, let $A_1, A_2, \ldots A_n$ be pairwise disjoint subsets of $\mathbb{R}$ and let

$$N_t^{(k)} = \sum_{i=1}^{N_t} 1_{\{Y_i \in A_k\}}$$

for all $k = 1, \ldots, n$ where $t \geq 0$. Then,

$$(N_t^{(1)})_{t\geq 0}, (N_t^{(2)})_{t\geq 0} \ldots (N_t^{(n)})_{t\geq 0}$$

are independent Poisson processes with rate $\lambda P(Y_1 \in A_k)$.

8.2 Lec 2

We have the basic components of a Poisson process down. We know arrivals over an interval follow a Poisson distribution with paratmeter λt and the interarrival times are exponentially distributed (which implies the arrival times ares gamma distributed). What happens if we relax one of these assumptions? For instance, we might want to model something where it doesn't make sense to say the interarrival times are exponentially distributed. It might be the case the longer you wait for an arrival, the more likely it is going to come in a relatively short time span.

In this case, we are dealing with a renewal process.

Definition 8.1 (Renewal process). Let $L_1, L_2, \ldots$ be a sequence of positive iid random variables. We call L_i the i-th holding time and $(L_i)_{i=1,2\ldots}$ the renewal sequence.

Let $T_0 = 0$ and define $T_n = \sum_{i=1}^n L_i$ for all $n \geq 1$. Each T_n is referred to as the n-th jump time, and $[T_n, T_{n+1}]$ as the renewal interval.

For $t \geq 0$ let

$$N_t = \max\{n \geq 0 : T_n \leq t\} = \sum_{k=1}^{\infty} 1_{(0,t]} T_k$$

Then N_t represents the number of jumps that have occurred by time $t > 0$ and the continuous-time stochastic process $(N_t)_{t\geq 0}$ is called a renewal process.

Notice the L_i's are comparable to a sequence of exponential random variables we might have had in a regular Poisson process. The difference here is that the L_i's can be distributed in any fashion, now. The fact the renewal sequence is not exponentially distributed means the jump times, or arrival times (T_i) are no longer gamma distributed.

Example 8.2. This is just meant to give you an idea of what renewal processes can model.

We might use such a process to model the breakdowns of parts in a machine.

- Let $L_1, L_2 \ldots$ be the lifetimes of the parameters

- T_n is the breakdown of the n-th part.

- N_t is the number of parts replaced in a machine by time t.

Since renewal processes are more general, it's more difficult to know things about them.[10] Consequently, we're left with asymptotic results.

Theorem 8.1. Let $(N_t)_{t \geq 0}$ be a renewal process. Then

$$\lim_{t \to +\infty} \frac{N_t}{t} = \frac{1}{\mathbb{E}[L_1]}$$

Proof. First, recall the strong law of large numbers. For $L_1, L_2, \ldots,$

$$\frac{1}{n} \sum_{i=1}^{\infty} L_i \to \mathbb{E}[L_1]$$

As $n \to \infty$. We also note that $N_t \to \infty$ as $t \to \infty$.

What follows is the crucial move of the proof. Observe that

$$T_{N_t} \leq t < T_{N_t + 1}$$

In words, the arrival time of the N_t-th arrival is going to be less than or equal to the time t. This is intuitive, since the last arrival happening at a time before t cannot occur at a time after t. Now, we realize that the arrival time of $N_t + 1$ must happen after t. If it happened at t or prior, it must count towards N_t. Now we manipulate the inequality.

$$\iff \frac{T_{N_t}}{N_t} \leq \frac{t}{N_t} < \frac{T_{N_t+1}(N_t + 1)}{N_t * (N_t + 1)}$$

---

[10]or so i've been told.

42

take the limit as $t \to \infty$. We use the fact $N_t \to \infty$ as $t \to \infty$ to invoke the law of large numbers.

$$\iff \mathbb{E}[L_1] \leq \frac{t}{N_t} < \lim_{t \to \infty} \frac{T_{N_t} + 1}{N_t + 1} * \frac{N_t + 1}{N_t}$$

$$\iff \mathbb{E}[L_1] \leq \frac{t}{N_t} < \mathbb{E}[L_1] * 1$$

$$\iff \lim_{t \to \infty} \frac{t}{N_t} = \mathbb{E}[L_1]$$

Take the reciprocal and we get

$$\lim_{t \to \infty} \frac{N_t}{t} = \frac{1}{\mathbb{E}[L_1]}$$

As desired

$\square$

Here is a quick application of the result, applied to our familiar Poisson process.

Example 8.3. Let N_t be a Poisson process with parameter $\lambda > 0$. Then

$$\lim_{t \to \infty} \frac{N_t}{t} = \frac{1}{\mathbb{E}[L_1]} = \frac{1}{1/\lambda} = \lambda$$

Now, we want to talk about coninuous time Markov chains. What's the difference between a continuous time chain and a discrete time one? First, we will note what stays the same.

- The state space S, of both the continuous time and discrete time chains are discrete.

- The Markov property continues to hold. The future only depends on the state we're in at the present. It does not depend on the past.

Here are some points of difference.

- Time evolves continuously. We can ask about the state of the chain at time t, where t is any real number, not just an integer.

- We can model the duration of time in each state with a continuous time Markov chain. Something analogous is possible in a discrete-time chain, but it's unwieldy.

- The Markov property manifests itself in a form of memorylessness. The time spent by the chain in each state in S is exponentially distributed.

Example 8.4. Recall our weather example from the beginning of the lecture notes. We have a state space, $S = \{1, 2, 3\}$ where each of the numbers correspond to rainy, sunny, and cloudy, respectively. We have a corresponding transition matrix:

$$\widetilde{P} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 3/4 & 0 & 1/4 \\ 1/4 & 3/4 & 0 \end{bmatrix}$$

To make this a continuous time Markov chain, we add the following information.

- rainfall lasts on average 3 hours at a time ($\lambda_r = 1/3$)

- sun shines on average 7 hours at a time ($\lambda_s = 1/7$)

- clouds stand for an average of 11 hours at a time ($\lambda_c = 1/11$)

Now, let X_t be the weather at time $t \geq 0$ (in hours). Then, $(X_t)_{t \geq 0}$ is a continuous time Markov chain. $\widetilde{P}$, the exponential parameters $(\lambda_r, \lambda_c, \lambda_s)$ completely specify the process.

Markov chains have the Markov property. We articulate this for the continuous case.

Definition 8.2 (Markov Property). A continuous time stochastic process $(X_t)_{t \geq 0}$ with discrete state space S is a continuous time Markov chain if

$$P(X_{t+s} = j \mid X_s = i, X_{s_n} = i_n, \ldots X_{s_0} = i_0) = P(X_{t+s} = j \mid X_s = i)$$

For all $n \in \mathbb{N}$ and $0 \leq s_0 < s_1 < \cdots < s_n < s < t$ and $i_0 \ldots i_n, i, j, \in S$.

The process is time-homogenous if the probability does not depend on what s is. That is,

$$P(X_{t+s} = j \mid X_s = i) = P(X_t = j \mid X_0 = i)$$

We need to have a function that tells us what the probability of moving from one state to another in t amount of time is. In the time-homogenous case we have the following definition to help.

Definition 8.3 (Transition function). for all $t \geq 0$ the tranition probabilities of a homogenous continuous time Markov chain can be arranged in a matrix function

$$P_{ij}(t) = P(X_t = j \mid X_0 = i)$$

The matrix $P(t) = (P_{ij}(t))_{ij \in S}$ is the transition function.

Proposition 8.2 (Chapman-Kolmogorov equations). For a continuous time Markov chain $(X_t)_{t \geq 0}$ with transition function $P(t)$, it holds that

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

For all $s, t, \geq 0$. That is,

$$P_{ij}(s + t) = [P(s)P(t)]_{ij} = \sum_{k \in S} P_{ik}(s)P_{kj}(t)$$

This is the same property as in the discrete time case.

9 Week 8

9.1 Lec 1

Now, we will introduce some more precise terminology to talk about continuous time Markov chains.

In general, let $(X_t)_{t \geq 0}$ be a continuous time Markov chain with a finite or countably infinite state space S.

Definition 9.1 (Holding times). T_i is the holding time for state $i \in S$. It is the length of time a Markov chain stays in state i before transitioning to a new state.

Because of time homogeneity and the Markov property, we can say the following:

Proposition 9.1. The holding times T_i for any state i are independent and exponentially distributed with holding time parameter $q_i \in [0, \infty)$.

Clearly, this means the mean time (expected duration) the Markov chain spends in state i is[11]

$$\mathbb{E}[T_i] = \frac{1}{q_i}$$

Definition 9.2 (Absorbing state). A state $i \in S$ is an absorbing state if $q_i = 0$. A Markov chain with at least one absorbing state is called an absorbing Markov chain.

When an absorbing state is visited the Markov chain never leaves that state. It's there forever.

It's also helpful to think about the 'chain within the chain,' so to speak.

Definition 9.3 (Embedded discrete time chain). Let $(X_t)_{t \geq 0}$ be a continuous time Markov chain and let $\widetilde{P} = p_{ij}$ where $i, j \in S$ describe the discrete transition probabilities from i to j.

Ignore time and just watch the state transitions of the continuous time Markov chain. Doing so, we get the sequence $Y_1, Y_2, \ldots, Y_n$ where Y_n is the n-th state visited by the continuous chain.

This embedded sequence, Y, is a discrete time Markov chain with transition probability matrix $\widetilde{P}$. We call it the embedded chain.

Note that $\widetilde{P}$ is a stochastic matrix whose diagonal entries are either 0 or 1.

---

[11]this is the expected time the chain spends in a state after it has just arrived. The total time a Markov chain spends in state i is likely to be infinite if the state is recurrent, and finite if the state is transient.

10 Week 9

10.1 Lec 1

We're going to specify a new way to talk about continuous time Markov chains. Recall that if we want to model a process with a continuous time chain, we need to gather a lot of information. We need to know the transition probabilities and the parameters of the exponential holding times. Finding the parameters can be accomplished from experimental data, but how exactly is one supposed to figure out the transition probabilities for the embedded chain?[12]

The solution is to conceive of a continuous time chain in a different way. Imagine we're in state $i \in S$. From state i, it's possible to visit $j, j_1, \ldots j_n$ other states (keeping in mind we cannot go back to i). For each of the n states we can hit from i, imagine an alarm clock that rings after an exponentially distributed amount of time is started. The parameter of the exponential alarm clock associated with the j-th state we can hit from i is q_{ij} and we call the random variable of the ring time T_{ij}. What's more, all the alarm clocks are independent.

The first alarm clock that rings determines the next state the chain visits. For example, if the alarm clock corresponding to the j_kth state rings first, we visit state j_k from i.

After we transition to state k, we begin the process again with an alarm clock for every state we can reach from k. Each of those alarm clocks will have exponentially distributed ring times with parameters $q_{k1}, q_{k2}, q_{k3}, \ldots$. From here, we follow the same procedure. This process models a continuous time Markov chain.

Now, we introduce more terminology to get a handle on the model of the chain we've introduced.

Definition 10.1. q_{ij} are called transition rates or instantaneous rates of the continuous time Markov chain.

Furthermore, we say $q_{ij} = 0$ if state j can't be hit from state i. What's more, we know $T_i = \min\{T_{i1}, T_{i2}, \ldots\}$, since the time spent in state i is the

---

[12]This is almost a philosophical problem.

minimum of the time to ring for all the exponential alarm clocks.

Knowledge about the exponential random variables gives the following facts:

Proposition 10.1. Recall, where $\forall i$, X_i is exponentially distributed with parameter λ_i, $P(\min\{X_1, X_2, \ldots, X_n\} = X_i) = \lambda_i/(\lambda_1 + \cdots + \lambda_n)$. Then,

- $q_i = \sum_{k=1}^{n_i} q_{ik}$ where n_i is the number of other states we can hit from i.

- The transition probabilities in the embedded chain $\widetilde{P}$ are: $p_{ij} = q_{ij}/\sum_{k=1}^{q_{ij}} q_{ik} = q_{ij}/q_i$

Definition 10.2 (generator matrix). The matrix Q with entries

$$Q_{ij} = \begin{cases} q_{ij} & \text{for } i \neq j \\ -q_i & \text{for } i = j \end{cases}$$

Is the generator matrix or infinitesimal generator matrix.

This is the most important matrix for a continuous time Markov chain. Also, note the entries of the matrix are not probabilities.

10.2   Lec 2

We're going to investigate some interesting properties of the generator matrix.

Proposition 10.2. The generator matrix, Q, satisfies

$$Q = P'(0)$$

Where $P(t)$ is the transition function.

What's important is to understand that Q is not a stochastic matrix. Even though we're differentiating a matrix of probabilities to get it, its entries are not probabilities.

Now, we introduce tools that will give us a principled way to solve for $P(t)$. In the end, we're going to be solving coupled systems of linear ODE's to get what we need.

Proposition 10.3 (Kolmogorov forward and backward equations)**.** A continuous time Markov chain with transition function $P(t)$ and generator Q satisfy the forward equation:

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

and the backward equation

$$P'(t) = QP(t)$$

With $P(0) = I$ being the identity matrix.

Equivalently, we might spell it out explicitly in sigma notation.

$$P'_{ij}(t) = \sum_k P_{ik}(t)q_{kj} = -P_{ij}(t)q_j + \sum_{k \neq j} P_{ik}(t)q_{ki}$$

or in the backwards form

$$P'_{ij}(t) = \sum_k q_{ik}P_{kj}(t) = -q_i P_{ij}(t) + \sum_{k \neq i} q_{ik}P_{kj}(t)$$

One might ask why we need two equations. The answer is that $P(t)$ and Q are matrices, so they don't necessarily commute. The forward and backward equations tell us that they do.

We want to solve this equation. It looks a lot like a simple differential equation of the form $f'(t) = qf(t)$ where $f(0) = 1$. In that case, we would recognize the solution is just $f(t) = e^{qt}$.

However, our differential equation deals with matrices, and we don't really know what it means to take the exponential of a matrix. That is, until we define it.

Definition 10.3. Let A be a kxk quadratic matrix. The matrix exponential e^A is the kxk matrix defined as:

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!}A^n = I + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \dots$$

Clearly, this takes after the definition of taking e to a regular power.

The definition is also nice because the resulting matrix exponential has some nice properties.

1. The matrix is well-defined since the series converges.

2. $e^0 = I$ with 0 being the zero matrix.

3. $e^A e^{-A} = I$

4. $e^{(s+t)A} = e^{sA} e^{tA}$.

5. if $AB = BA$ then $e^{A+B} = e^A e^B = e^B e^A$

6. $\frac{d}{dt} e^{tA} = A e^{tA} = e^{tA} A$

7. if $D \in \mathbb{R}^{n x n}$ is a diagonal matrix with entries $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ then e^D is a square matrix with only e^{λ_i} on the diagnonal. There are 0's everywhere else.

Given how we now understand what the form of the solution of the Kolmogorov forward and backward equations is, how do we solve it? Well, if you're in UCLA's math 171, they just give it to you.

Proposition 10.4. For a continuous-time Markov chain with transition function $P(t)$ and infinitesimal generator Q, it holds that:

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

In words, e^{tQ} is the solution to the Kolmogorov forward/backward equation.

Now, this isn't as great as we'd like. For one, the matrix exponential is really difficult to compute in the majority of cases. However, we can get lucky. If Q is diagonalizable, then the following holds.

Proposition 10.5. If the generator matrix Q is diagonalizable with $Q = SDS^{-1}$ then the transition function is given by

$$P(t) = e^{tQ} = Se^{tD}S^{-1}$$

The following refresher facts from linear algebra should also be helpful in conjuction with the above proposition.

Facts

1. If A is diagonalizable, then the entries of D are the eigenvalues of A and the columns of S are the corresponding eigenvectors.

2. if A has n distinct eigenvectors, then A is diagonalizable.

11 Week 10

11.1 Lec 1

Now we want to understand the limiting behavior of continuous-time Markov chains. To do so, we need to introduce additional concepts that are analogous to the discrete case.

Definition 11.1. A continuous-time Markov chain is irreducible if for any two states i and j it is possible to get from i to j in a finite number of jumps. We say there exists a finite number of states $k_0 = i, k_1, \ldots, k_{n-1}, k_n = j$ such that $q_{k_{m-1}, k_m} > 0$ for all $1 \leq m \leq n$.

A direct consequence of this:

Lemma 11.1. If $(X_t)_{t \geq 0}$ is an irreducible continuous time Markov chain and $s > 0$, then $P_{ij}(s) > 0$ for all states i, j. We can also say $P_{ij}(t) > 0$ for all $t > 0$.

Similar to the discrete case, we have a definition of a stationary distribution.

Definition 11.2 (Stationary distribution)**.** A probability distribution π is a stationary distribution if

$$\pi^T * P(t) = \pi^T$$

for all $t \geq 0$.

This is a more powerful notion than the discrete case. Here, we need to ensure π is stationary for all time.

There is a link to the generator matrix.

Proposition 11.1. A probability distribution π is a stationary distribution of a continuous time Markov chain with generator Q if and only if

$$\pi^T * Q = 0$$

in other words,

$$\sum_{i \in S} \pi_i * Q_{ij} = 0$$

for all $j \in S$.

Proof. we proof the "if" direction first.

Assume $\pi^T P(t) = \pi^T$. Differentiate with respect to t to yields

$$\pi^T P(t) = 0$$

For all $t \geq 0$. In particular, let $t = 0$. Then

$$\pi^T P'(0) = 0$$
$$\Longleftrightarrow \pi^T Q = 0$$

Since $P'(0) = Q$ by definition.

Now we prove the "only if" direction.

Assume $\pi^T Q = 0$.

$$\pi^T Q = 0$$
$$\Longleftrightarrow \pi^T Q P(t) = 0$$

Now, we use the Kolmogorov backward equation.

$$\Longleftrightarrow \pi^T P'(t) = 0$$

Since π^T is a non-zero vector, we can surmise that $P'(t)$ is the zero matrix. This means

$$\pi^T P(t) = k$$

for all t where k is some constant. We plug in $t = 0$ in order to solve for k.

$$\pi^T P(0) = \pi^T I_n = \pi^T$$

$k = \pi^T$. This means that

$$\pi^T P(t) = \pi^T$$

As desired. $\qquad\square$

Theorem 11.2 (Fundamental limit theorem). If a continuous time Markov chain $(X_t)_{t\geq 0}$ is irreducible and has a stationary distribution π, then for all j,

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

for all initial i.

We shouldn't neglect the embedded chain! Recall that we have a discrete-time Markov chain working 'under the hood' of our continuous-time one. This one also has a stationary distribution. Call it ψ.

Proposition 11.2. Let π, the stationary distribution of the 'parent' continuous-time Markov chain, be given. Then:

$$\psi_j = \frac{\pi_j q_j}{\sum_k \pi_k q_k}$$

for all j is the stationary distribution of the embedded chain.

Also, let ψ be given. Then,

$$\pi_j = \frac{\psi_j / q_j}{\sum_k \psi_k / q_k}$$

11.2   Lec 2

We introduce some notions for the continuous-time Markov chain that are nearly identical to the discrete-time case.

Definition 11.3. Let $(X_t)_{t \geq 0}$ be a continuous time Markov chain with generator Q. A probability distribution λ satisfies the detailed balance condition if

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

For all $i \neq j$.

There is a consequence of this.

Proposition 11.3. If a probability distribution satisfies the detailed balance condition, it is a stationary distribution.

This comes into play in proofs about birth and death processes, as defined below.

Definition 11.4. A birth and death process is a continuous time Markov chain with state space $S = \{0, 1, 2, \dots\}$ where transitions occur from state i to $i + 1$ at rate λ_i and from states i to $i - 1$ at rate μ_i.

The following proposition can tell us what the stationary distribution of a birth and death process will be like.

Proposition 11.4. Suppose we have a birth and death process with birth rates λ_i and death rates μ_i for $i = 0, 1, 2, \dots$. Assume that

$$\sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} < \infty$$

Then, the unique stationary distribution π is given by

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}$$

for all $k = 1, 2, \dots$.

and

$$(\pi_0)^{-1} = \sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}$$