Stat 150 Notes

Leo Villani

April 2024

1 Basic Review

We define the conditional pmf of X given Y = y by $p_{X|Y}(x|y) = \frac{Pr(X=x,Y=y)}{Pr(Y=y)}$.
Note: $p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)}$.
Law of total probability says: $Pr(X = x) = \sum_{y=0}^{\infty} p_{X|Y}(x|y)p_Y(y)$
Let g be a function for which the expectation of g(X) is finite, then the conditional expected value of g(X) given Y = y is: $E[g(X)|Y = y] = \sum_x g(x)p_{X|Y}(x|y)$.
Law of total probability then says: $E[g(X)] = \sum_y E[g(X)|Y = y]p_Y(y)$

2 Markov Chains

2.1 Markov Chain Definition and Properties

"A Markov Process X_t is a stochastic process with the property that, given the value of X_t, the values of X_s for $s > t$ are not influenced by the values of X_u for $u < t$. In words, the probability of any particular future behavior of the process, when its current state is known exactly, is not altered by additional knowledge concerning its past behavior."
DTMC: (Markov chain with finite or countable state space)
Markov Property:
$Pr(X_{n+1} = j|X_0 = i_0, ..., X_n = i_n) = Pr(X_{n+1} = j|X_n = i_n)$
We denote the transition probabilities by a P matrix where
$P_{ij}^{n,n+1} = Pr(X_{n+1} = j|X_n = i)$.
This notation also considers the time at which transitions happen but when one-step transition probabilities are independent of the time variable n, we say that the Markov Chain has stationary transition probabilities. ($P_{ij}^{n,n+1} = P_{ij}$)

$$\mathbf{P} = \begin{Vmatrix} P_{00} & P_{01} & P_{02} & P_{03} & \cdots \\ P_{10} & P_{11} & P_{12} & P_{13} & \cdots \\ P_{20} & P_{21} & P_{22} & P_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & P_{i3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{Vmatrix}$$

The two classic conditions are:
1.$P_{ij} \geq 0$

2.$\sum_{j=0}^{\infty} P_{ij} = 1$

By Markov property computing joint probabilities is easy:
$Pr(X_0 = i_0, ..., X_n = i_n) = p_{i_0} P_{i_0,i_1} ... P_{i_{n-1},i_n}$
Moving onto n-step calculations we denote $P_{ij}^{(n)}$ as the probability that the process goes from state i to state j in n transitions.
Mathematically, $P_{ij}^{(n)} = Pr(X_{m+n} = j|X_m = i)$
It turns out $P_{ij}^{(n)} = \mathbf{P}^n$.

2.2 First Step Analysis

This method proceeds by analyzing, or breaking down, the possibilities that can arise at the end of the first transition, and then invoking the law of total probability coupled with the Markov property to establish a characterizing relationship among the unknown variables
Example:

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left\| \begin{array}{ccc} 1 & 0 & 0 \\ \alpha & \beta & \gamma \\ 0 & 0 & 1 \end{array} \right\| \end{array},$$

This Markov chain has two absorbing stares 0, and 2 but if started in state 1 it may stay in 1 or move to 0 or 2.
This may lead us to two different questions:
1. In which state, 0 or 2, is the process ultimately trapped?
2. How long, on average, does it take to reach one of these states?
Start by defining $T = min(n \geq 0; X_n = 0 \text{ or } X_n = 2)$, which is the time of absorption of the process.
We can answer the two questions by:
1. $u = Pr(X_T = 0|X_0 = 1)$
2. $v = E[T|X_0 = 1]$
For the first question consider at time 1 being at state 0,1 or 2:
$Pr(X_T = 0|X_1 = 0) = 1$
$Pr(X_T = 0|X_1 = 2) = 0$
$Pr(X_T = 0|X_1 = 1) = u$
Then by total probability:
$u = Pr(X_T = 0|X_0 = 1) = \sum_{k=0}^{2} Pr(X_T = 0|X_0 = 1, X_1 = k) * Pr(X_1 = k|X_0 = 1) = 1 * \alpha + u * \beta + 0 * \gamma$
So $u = \alpha + \beta u$ or $u = \frac{\alpha}{1-\beta} = \frac{\alpha}{\alpha+\gamma}$.
Observe that this quantity is the conditional probability of a transition to 0 given a transition to 0 or 2 occurred which makes sense.
Similar analysis for v we get $v = 1 + \alpha * 0 + \beta * v + \gamma * 0 = 1 + \beta v$.
So $v = \frac{1}{1-\beta}$
The intuition is that we need to wait to just stop returning to state 1 which is mean of a geometric R.V.
Transient states are the states that are non-absorbing.

Example:
If we have a 4 state Markov Chain with 0 and 3 as absorbing states we can define a more general u_i and v_i.
$u_i = Pr(X_T = 0|X_0 = i)$for $i = 1, 2$, and $v = E[T|X_0 = i]$ for $i = 1, 2$.
Its obvious also that $u_0 = 1, u_3 = 0$, and $v_0 = v_3 = 0$, where we seek to see state 0 before state 3.

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0.4 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

The first step analysis says:
$u_1 = 0.4 + 0.3u_1 + 0.2u_2$
$u_2 = 0.1 + 0.3u_1 + 0.3u_2$
Solving yields $u_1 = \frac{30}{43}, u_2 = \frac{19}{43}$, which says that beginning in state 1 or 2 the Markov Chain will end up in state 0 with probability u_1 or u_2 respectively.
Now for the v_i:
$v_1 = 1 + 0.3v_1 + 0.2v_2$
$v_2 = 1 + 0.3v_1 + 0.3v_2$
Solving yields $v_1 = \frac{90}{43}, v_2 = \frac{100}{43}$, which says that beginning in state 1 or 2 the Markov Chain will end up on average in state 0 after v_1 or v_2 steps respectively.

2.3 Generalization

This method generalizes to a finite-state $(0, ...N)$ Markov Chain with transient states $(0, ..., r-1)$ and absorbing states $(r, ...N)$.
The transition matrix has the form:

$$\mathbf{P} = \left\| \begin{array}{cc} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{array} \right\|$$

Where 0 is a 0-submatrix, and $Q_{ij} = P_{ij}$. In general, started at one of the transient states $X_0 = i$, where $0 \le i \le r-1$, such a process will remain in the transient states for some random duration, but ultimately the process gets trapped in one of the absorbing states $i = r, ..., N$.

2.4 Branching Processes

Suppose an organism at the end of its lifetime produces a random number ξ of offspring with probability distribution:
$Pr(\xi = k) = p_k$ for $k = 0, 1, 2, ...$
We assume all offspring act independently of each other and each have progeny according to the probability distribution.
The process (X_n), where X_n is the population size at the n-th generation is a Markov Chain of a special structure called a *branching process*.

Mathematically, in the n-th generation the X_n individuals independently give rise to $\xi_1^{(n)}, \xi_2^{(n)} ..., \xi_{X_n}^{(n)}$ offspring and hence $X_{n+1} = \xi_1^{(n)} + \xi_2^{(n)} ... + \xi_{X_n}^{(n)}$.
Let $\mu = E[\xi], \sigma^2 = Var[\xi]$. Let $M(n), V(n)$ be the mean and variance of X_n under the initial condition $X_0 = 1$.
Then $M(n+1) = \mu M(n)$ and $V(n+1) = \sigma^2 M(n) + \mu^2 V(n)$.
So $M(n) = \mu^n$ and the mean population size increases geometrically when $\mu > 1$, decreases geometrically when $\mu < 1$ and remains constant when $\mu = 1$.

2.4.1 Extinction Probabilities

Population extinction occurs when and if the population size is reduced to zero. The random time of extinction N is thus the first time n for which $X_n = 0$, and then, obviously, $X_k = 0$ for all $k \geq N$.
In Markov chain terminology, 0 is an absorbing state, and we may calculate the probability of extinction by invoking a first step analysis.
Let $u_n = Pr(N \leq n) = Pr(X_n = 0)$ be the probability of extinction at or prior to the n-th generation with $X_0 = 1$.
Suppose that the single parent represented by $X_0 = 1$ gives rise to $\xi_1^{(0)} = k$ offspring. In turn, each of these offspring will generate a population of its own descendants, and if the original population is to die out in n generations, then each of these k lines of descent must die out in $n-1$ generations.
As each of the k subpopulations have same properties as the original X_0 the probability of dying in $n-1$ generations is u_{n-1}. So for all k subpopulations to die is $(u_{n-1})^k$ by independence. Using law of total probability:
$u_n = \sum_{k=0}^{\infty} p_k (u_{n-1})^k$, $n = 1, 2, ...$
Clearly, $u_0 = 0$, and $u_1 = p_0$, where p_0 is probability of having 0 offspring.
Example:
Suppose a parent has no offspring with probability $\frac{1}{4}$ and two with probability $\frac{3}{4}$ then we get: $u_n = \frac{1}{4} + \frac{3}{4}(u_{n-1})^2$
Starting with $u_0 = 0$ you can successively compute all u_n and see it tends toward $\frac{1}{3}$.

2.4.2 Generating Functions and Extinction Probabilities

Considering ξ with $Pr(\xi = k) = p_k$ for $k = 0, 1, 2, ...$ the generating function associated with ξ is defined by:
$\phi(s) = E[s^\xi] = \sum_{k=0}^{\infty} p_k s^k$ for $0 \leq s \leq 1$.
Knowing a generating function is basically like knowing the values of the p_k as: $p_k = \frac{1}{k!} \frac{d^k \phi(s)}{ds^k}|_{s=0}$.
Generating functions are also useful for sums of random variables as if $X = \xi_1 + ... + \xi_n$ then $\phi_X(s) = \phi_1(s) ... \phi_n(s)$.
Finally they can be useful in finding moments i.e. $\frac{d\phi(s)}{ds} = E[\xi]$.
Shifting towards extinction probabilities:
Consider the normal branching process setup with generatin function $\phi(s)$ and

$u_n = Pr(X_n = 0)$, probability of extinction by stage n. Note: u_∞ is the probability that the population eventually becomes extinct so $u_\infty = Pr(X_n = 0)$ for some n.
Then, $u_n = \sum_{k=0}^{\infty} p_k (u_{n-1})^k = \phi(u_{n-1})$.
So if want to find probability of eventual distinction we are finding u_∞ so we just solve $u = \phi(u)$ and pick the minimum solution.
For the $\frac{1}{4}$ no offspring, $\frac{3}{4}$ chance of 2 offspring example from before we say $\phi(s) = \frac{1}{4} + \frac{3}{4}s^2$ so solving $u = \phi(u)$ yields $u = \frac{1}{3}, 1$. Thus the probability of eventual distinction is the minimum $\frac{1}{3}$.
As we will always have $u_\infty = 1$ other solutions are determined by the slope of $\phi'(1)$. If slope less than or equal to one then no crossing takes place of $\phi(s) = s$ but if bigger than one there was some crossing so extinction is not guaranteed.
Thus as $\phi'(1) = E[\xi]$, if the mean offspring size $E[\xi] \leq 1$, then $u_\infty = 1$ and extinction is certain. If $E[\xi] > 1$, then $u_\infty < 1$ and the population may grow unboundedly with positive probability. If $E[\xi] = 1$ mean population size is constant, but population is sure to die out eventually, so extinction also guaranteed.

3 The Long Run Behavior of Markov Chains

Suppose for some k, $\mathbf{P}^k$ has all strictly positive entries then the corresponding Markov Chain is called *regular*.
The most important fact concerning a regular Markov chain is the existence of a *limiting probability distribution* $\pi = (\pi_0, ..., \pi_N)$. Convergence means:
$\lim_{n\to\infty} Pr(X_n = j | X_0 = i) = \pi_j > 0$ for $j = 0, 1, ...N$
This convergence means that in the long run the probability of being in state j is approximately π_j no matter what the starting state.
Easy test for regularity:
1. There is a path from i to j, $\forall i, j$ (irreducible)
2. At least one state i for which $P_{ii} > 0$

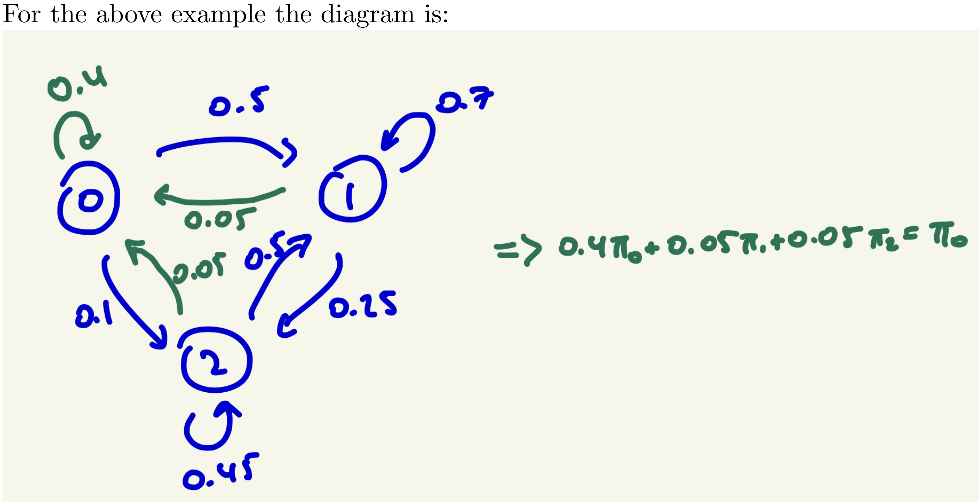
Theorem 1 *Let $\boldsymbol{P}$ be a regular transition probability matrix on the states 0,…,N. Then the limiting distribution π is the unique nonnegative solution of the equations:*
1. $\pi_j = \sum_{k=0}^{N} \pi_k P_{kj}$
2. $\sum_{k=0}^{N} \pi_k = 1$

Example:

$$\begin{array}{cc} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \mathbf{P} = \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \left\|\begin{array}{ccc} 0.40 & 0.50 & 0.10 \\ 0.05 & 0.70 & 0.25 \\ 0.05 & 0.50 & 0.45 \end{array}\right\| \end{array},$$

The equations determining the limiting distribution $(\pi_0, \pi_1, \pi_2) are$:
$0.4\pi_0 + 0.05\pi_1 + 0.05\pi_2 = \pi_0$
$0.5\pi_0 + 0.7\pi_1 + 0.5\pi_2 = \pi_1$

$0.1\pi_0 + 0.25\pi_1 + 0.45\pi_2 = \pi_2$
$\pi_0 + \pi_1 + \pi_2 = 1$
There is always one redundant equation of the non-normalization constraints so we just delete one and solve to get:
$\pi_0 = \frac{1}{13}, \pi_1 = \frac{5}{8}, \pi_2 = \frac{31}{104}$
Note: If given the diagram of the Markov Chain one way to set up these equations is just to analyze the flow in versus the flow out.
For the above example the diagram is:



3.0.1 Doubly Stochastic Matrices

Doubly stochastic matrices are just those that have both their columns and rows sum to 1.

Theorem 2 *Consider a doubly stochastic transition probability matrix on states* $0, ..., N-1$ *then if the matrix is regular the unique limiting distribution is the uniform distribution* $\pi = \frac{1}{N}(1, ..., 1)$.

3.0.2 Interpretations of Limiting Distribution

The interpretation of the limiting distribution we have been thinking out is that after a process has been in operation for a long time, the probability of finding the process in state j is π_j, regardless of the initial state.
Another interpretation is that the π_j also represents the *long run mean fraction* of time that the process (X_n) is in state j.
Thus if each visit to state j incurs a "cost" c_j then the long run mean cost per unit time associated with the Markov Chain is
Long run mean cost per unit time $= \sum_{j=0}^{N} \pi_j c_j$

Example:
Suppose that the weather on any day depends on the weather conditions for the previous 2 days. To be exact, we suppose that if it was sunny today and yesterday, then it will be sunny tomorrow with probability 0.8; if it was sunny today but cloudy yesterday, then it will be sunny tomorrow with probability 0.6; if it was cloudy today but sunny yesterday, then it will be sunny tomorrow with probability 0.4; if it was cloudy for the last 2 days, then it will be sunny tomorrow with probability 0.1.
Such a model can be transformed into a Markov chain, provided we say that the state at any time is determined by the weather conditions during both that day and the previous day.
We define states $(S,S),(S,C),(C,S),(C,C)$ if it was X yesterday but Y today.
Then $\mathbf{P}$ is:

$$
\begin{array}{cc|cccc|}
 & & \multicolumn{4}{c}{\text{Today's state}} \\
 & & (S,S) & (S,C) & (C,S) & (C,C) \\
\hline
 & (S,S) & 0.8 & 0.2 & & \\
\text{Yesterday's} & (S,C) & & & 0.4 & 0.6 \\
\text{state} & (C,S) & 0.6 & 0.4 & & \\
 & (C,C) & & & 0.1 & 0.9
\end{array}.
$$

With limiting distribution equations:
$0.8\pi_0 + 0.6\pi_2 = \pi_0$
$0.2\pi_0 + 0.4\pi_2 = \pi_1$
$0.4\pi_1 + 0.6\pi_3 = \pi_2$
$0.6\pi_1 + 0.9\pi_3 = \pi_3$
$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$
Solving yields: $\pi_0 = \frac{3}{11}, \pi_1 = \frac{1}{11}, \pi_2 = \frac{1}{11}, \pi_3 = \frac{6}{11}$.
If we wanted to know the long time fraction of it being sunny it would just be the sum of being sunny today and being sunny or cloudy tomorrow so the answer is $\pi_0 + \pi_1 = \frac{4}{11}$

3.1 The Classification of States

j is said to be *accessible* from state i if $P_{ij}^{(n)} > 0$ for some integer $n \geq 0$, in simple terms there is a probability that state j can be reached from state i in some finite number of transitions.
If two states i, j are accessible to each other than they *communicate*.
Communication is an equivalence relation.
Thus we can partition the states into equivalence classes that can communicate with each other.
We say that a Markov Chain is irreducible if there is only one communicating class, i.e. all states communicate with each other.
We define the *period* of state i, written $d(i)$ to be the g.c.d. of all integers $n \geq 1$ for which $P_{ii}^{(n)} > 0$.
A Markov Chain in which each state has period 1 is called aperiodic.

Period is a class property.

3.1.1 Recurrent and Transient States

Consider an arbitrary but fixed state i. We define for each integer $n \geq 1$
$f_{ii}^{(n)} = Pr(X_n = i, X_v \neq i, v = 1, 2, ..., n-1 | X_0 = i)$
In other words $f_{ii}^{(n)}$ is the probability that starting from state i the first return to state i occurs at the n-th transition.
Note: $f_{ii}^{(1)} = P_{ii}$ and $f_{ii}^{(n)}$ may be calculated recursively according to:
$P_{ii}^{(n)} = \sum_{k=0}^{n} f_{ii}^{(k)} P_{ii}^{(n-k)}$, $n \geq 1$
Also note $f_{ii}^{(0)} = 0 \ \forall i$.
When the process starts from state i, the probability that it returns to state i at some time is $f_{ii} = \sum_{n=0}^{\infty} f_{ii}^{(n)}$.
We say that a state i is *recurrent* if $f_{ii} = 1$. This definition says that a state i is recurrent if and only if, after the process starts from state i, the probability of its returning to state i after some finite length of time is one.
A nonrecurrent state is said to be transient.
If we let M be the number of times that the process returns to i then for a transient state $E[M|X_0 = i] = \frac{f_{ii}}{1-f_{ii}}$.

Theorem 3 *A state i is recurrent if and only if:* $\sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty$. *Equivalently state i is transient if and only if* $\sum_{n=1}^{\infty} P_{ii}^{(n)} < \infty$.

Recurrence is a class property.

3.2 The Basic Limit Theorem of Markov Chains

Consider a recurrent state i. Then,:
$f_{ii}^{(n)} = Pr(X_n = i, X_v \neq i, v = 1, 2, ..., n-1 | X_0 = i)$
is the probability distribution of the *first return time*
$R_i = min(n \geq 1; X_n = i)$
This is
$f_{ii}^{(n)} = Pr(R_i = n | X_0 = i)$, for $n = 1, 2, ...$

The mean duration between visits to state i is:
$m_i = E[R_i | X_0 = i] = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$
After starting in state i, then on average the process is in state i once every $m_i = E[R_i | X_0 = i]$ units of time.

Theorem 4 *a)Consider a recurrent, irreducible, aperiodic Markov Chain. Let* $P_{ii}^{(}n)$ *be the probability of entering state i at the n-th transition, given that* $X_0 = i$. *Let* $P_{ii}^{(0)} = 1$. *Let* $f_{ii}^{(n)}$ *be the probability of the first returning to state i at the nth transition, where* $f_{ii}^{(0)} = 0$. *Then*
$\lim_{n\to\infty} P_{ii}^{(n)} = \frac{1}{\sum_{n=0}^{\infty} n f_{ii}^{(n)}} = \frac{1}{m_i}$

Theorem 5 *In a positive recurrent aperiodic class then the "stationary distribution" is given by the set of equations*
$\sum_{n=0}^{\infty} \pi_i = 1$ *and* $\pi_j = \sum_{n=0}^{\infty} \pi_i P_{ij}$

Any set π is called a *stationary probability distribution.*
A limiting distribution when it exists is always a stationary distribution but the converse is not true.
The difference is that the stationary distribution needs the initial state for π_j to be j for the limit to converge while a general limiting distribution does not.
Contradicting Example:

$$\mathbf{P} = \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix},$$

No limiting distribution here but $\pi = (\frac{1}{2}, \frac{1}{2})$ is a stationary distribution.
A unique stationary distribution π exists for a positive recurrent, irreducible Markov Chain and the mean fraction of time in state i converges to π_i as the number of stages n grows to infinity. (does not need starting in state i here as we have made the constraints on the chain stricter)

4 Poisson Processes

Poisson distribution with parameter $\mu > 0$:
$p_k = \frac{e^{-\mu}\mu^k}{k!}$ for $k = 0, 1, ...$
If X is a Poisson R.V. with parameter μ then $E[X] = \mu$, $\sigma_X^2 = Var[X] = \mu$

Theorem 6 *Let X and Y be independent random variables having Poisson distributions with parameters μ and υ, respectively. Then the sum $X + Y$ has a Poisson distribution with parameter $\mu + \upsilon$*

A Poisson Process of rate $\lambda > 0$ is an integer valued stochastic process $(X(t); t \geq 0)$ for which:
1. for any time points $t_0 = 0 < t_1 < t_2 < ... < t_n$ the process increments $X(t_1) - X(t_0), ...X(t_n) - X(t_{n-1})$ are independent R.V.s.
2. for $x \geq 0$ and $t > 0$ the R.V. $X(t + s) - X(s)$ has Poisson distribution
$Pr(X(s + t) - X(s) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$ for $k = 0, 1, ...$
3. $X(0) = 0$
Note: $E[X(t)] = \lambda t, Var[X(t)] = \sigma_{X(t)}^2 = \lambda t$
Example:
Defects occur along an undersea cable according to $PP(0.1)$.
a) What is the probability that no defects appear in the first two miles?
b) Given that there are no defects in the first two miles what is the conditional probability of no defects between mile points 2 and 3?
a) $X(2)$ has poisson distribution with parameter $(0.1)(2) = 0.2$ so $Pr(X(2) =$

$0) = e^{-0.2}$.
b) By independence of $X(3) - X(2)$ and $X(2) - X(0)$ this question is just the same as $Pr(X(1) = 0 = e^{-1}$
A nonhomogeneous Poisson process is a Poisson process that has $\lambda = \lambda(t)$ which varies with time.
Note: Questions about nonhomogeneous Poisson processes can be transformed into corresponding questions about homogeneous processes.

4.1 The Law of Rare Events

The common occurrence of the Poisson distribution in nature is explained by the *law of rare events*. Informally, this law asserts that where a certain event may occur in any of a large number of possibilities, but where the probability that the event does occur in any given possibility is small, then the total number of events that do happen should follow, approximately, the Poisson distribution. Formally this has to do with the Poisson limit theorem (sending Binomial R.V. to limit).

Theorem 7 *Let $\epsilon_1, \epsilon_2, ...$ be independent Bernoulli random variables, where $Pr(\epsilon_i = 1) = p_i$ and $Pr(\epsilon_i = 0) = 1 - p_i$
and let $S_n = \epsilon_1 + ... + \epsilon_n$. The exact probabilities for S_n and Poisson probabilities $\mu = p_1 + ... + p_n$ differ by at most
$|Pr(S_n = k) - \frac{\mu^k e^{-\mu}}{k!}| \leq \sum_{i=1}^{n} p_i^2$*

4.1.1 The Law of Rare Events and the Poisson Process

Let $N((a, b])$ denote the number of events that occur during the interval $(a, b]$. That is, if $t_1 < t_2 < t_3 <$ denote the times (or locations, etc.) of successive events, then $N((a, b])$ is the number of values t_i for which $a < t_i \leq b$.
We make the following postulates:
1. The numbers of events happening in disjoint intervals are independent random variables. That is, for every integer $m = 2, 3, ...$ and time points $t_0 = 0 < t_1 < t_2 < \ < t_m$, the random variables $N((t_0, t_1]), N((t_1, t_2]), ..., N((t_{m-1}, t_m])$ are independent.
2. For any time t and positive number h the probability distribution of $N((t, t + h])$, the number of events occurring between time t and $t + h$ depends only on the interval length h and not on the time t.

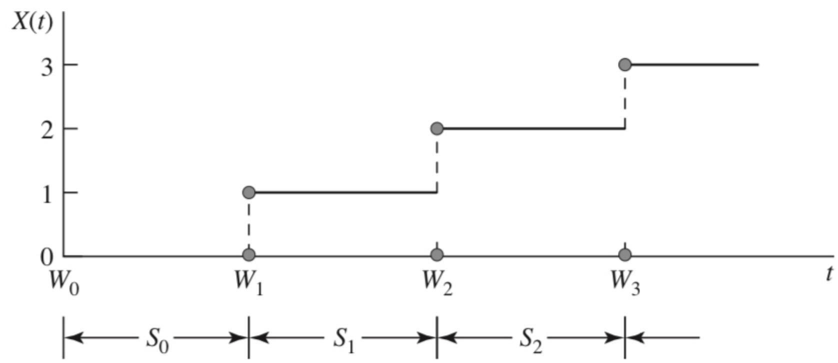
Let $N((s, t])$ be a random variable counting the number of events occurring in an interval $(s, t]$. Then $N((s, t])$ is a $PP(\lambda)$ if:
1. Postulate 1

2. for any times $s < t$ the random variable $N((s,t])$ has the Poisson distribution $Pr(N((s,t]) = k) = \frac{[\lambda(t-s)]^k e^{-\lambda(t-s)}}{k!}, k = 0, 1, ...$

4.2 Distributions Associated with the Poisson Process

A *Poisson point process* $N((s,t])$ counts the number of events occurring in an interval $(s,t]$. A *Poisson counting process*, or more simply a Poisson process $X(t)$, counts the number of events occurring up to time t. Formally, $X(t) = N((0,t])$.

If W_n is the time of occurrence of the n-th event, the so-called *waiting time*, with $W_0 = 0$, the differences $S_n = W_{n+1} - W_n$ are called *soujourn times*; S_n measures the duration the Poisson process sojourns in state n.



Theorem 8 *The waiting time W_n has the gamma distribution whose probability density function is*
$f_{W_n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}$, $n = 1, 2, ..., t \geq 0$
In particular W_1 the time to the first event is exponentially distributed
$f_{W_1}(t) = \lambda e^{-\lambda t}, t \geq 0$

Theorem 9 *The sojourn times $S_0, S_1, ..., S_{n-1}$ are independent random variables each having the exponential probability density function*
$f_{S_k}(s) = \lambda e^{-\lambda s}, s \geq 0$

Theorem 10 *Let $(X(t))$ be a Poisson process of rate $\lambda > 0$. Then for $0 < u < t$ and $0 \leq k \leq n$*
$Pr(X(u) = k | X(t) = n) = \frac{n!}{k!(n-k)!} (\frac{u}{t})^k (1 - \frac{u}{t})^{n-k}$

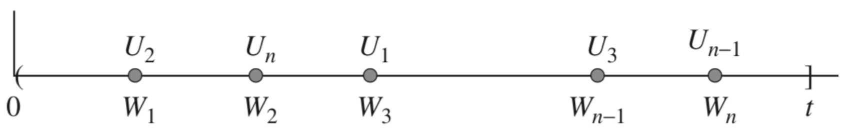
4.3 The Uniform Distribution and Poisson Processes

The main result in this section is that conditioned on a fixed total number of events in an interval, the locations of those events are uniformly distributed in a certain way.

First an introduction to order statistics:
We begin with a line segment t units long and a fixed number n of darts and throw darts at the line segment in such a way that each dart's position upon landing is uniformly distributed along the segment, independent of the location of the other darts. Let U1 be the position of the first dart thrown, U2 the position of the second, and so on up to Un. The probability density function is the uniform density

$$f_U(u) = \begin{cases} 1/\text{t}, 0 \leq u \leq t \\ 0, elsewhere \end{cases}$$

Now we define $W_1 \leq ... \leq W_n$ as the same positions sorted by order on the line. These have a different joint distribution
$f_{W_1,...,W_n}(w_1, ..., w_n) = n!t^{-n}$ for $0 < w_1 < ... < w_n \leq t$



Theorem 11 *Let $W_1, ..., W_n$ be the occurrence times in a Poisson process of rate $\lambda > 0$. Conditioned on $N(t) = n$ the random variables $W_1, ..., W_n$ have the joint probability density function*
$f_{W_1,...,W_n|X(t)=n}(w_1, ..., w_n) = n!t^{-n}$ *for* $0 < w_1 < ... < w_n \leq t$

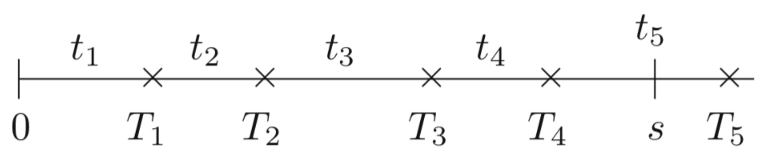
This theorem for various applications when dealing with symmetric functions.

5 Renewal Process

In the Poisson process the times between successive arrivals are independent and exponentially distributed. The lack of memory property of the exponential distribution is crucial for many of the special properties of the Poisson process however, in many situations the assumption of exponential interarrival times is not justified. Now we consider a generalization of Poisson processes called renewal processes in which the times $t_1, t_2, ...$ between events are independent and have distribution F.

In order to have a simple metaphor with which to discuss renewal processes, we will think of a single light bulb maintained by a very diligent janitor, who replaces the light bulb immediately after it burns out. Let t_i be the lifetime of the ith light bulb. We assume that the light bulbs are bought from one manufacturer, so we suppose

$P(t_i \leq t) = F(t)$
where F is a distribution function with $F(0) = 0$
If we start with a new bulb (numbered 1) at time 0 and each light bulb is replaced when it burns out, then $T_n = t_1 + ... + t_n$ gives the time that the nth bulb burns out, and
$N(t) = max(n : T_n \leq t)$
is the number of lightbulbs that have been replaced by time t.



Theorem 12 *Let $\mu = E[t_i]$ be mean interarrival time. If $P(t_i > 0) > 0$ then with probability one*
$\frac{N(t)}{t} \to \frac{1}{\mu}$ *as* $t \to \infty$

In words, this says that if our light bulb lasts μ years on the average then in t years we will use up about $\frac{t}{\mu}$ light bulbs.
Note: This holds for the Poisson process with $\frac{N(t)}{t} \to \lambda$ as exponential distribution has mean $\frac{1}{\lambda}$

Theorem 13 (Strong law of Large Numbers) *Let $x_1, x_2, x_3, ...$ be i.i.d. with $E[x_i] = \mu$ and let $S_n = x_1 + ... + x_n$. Then with probability one*
$\frac{S_n}{n} \to \mu$ *as* $n \to \infty$

5.0.1 Reward Process

We suppose that at the time of the ith renewal we earn a reward r_i. The reward r_i may depend on the ith interarrival time t_i, but we will assume that the pairs (r_i, t_i), $i = 1, 2, ...$ are independent and have the same distribution. Let
$R(t) = \sum_{i=1}^{N(t)} r_i$
be the total amount of reward by time t.

Theorem 14 *With probability one*
$\frac{R(t)}{t} \to \frac{E[r_i]}{E[t_i]}$

Intuitively this means that
$\frac{limitingreward}{time} = \frac{expectedreward/cycle}{expectedtime/cycle}$

6 CTMC

In discrete time we constructed the Markov Chain based on the property:
$P(X_{n+1} = j|X_n = i, ..., X_0 = i_0 = P(X_{n+1} = j|X_n = i)$
In continuous time it is technically difficult to define the conditional probability given all of the X_r for $r \leq s$ so we instead say that $X_t, t \geq 0$ is a Markov Chain if for any $0 \leq s_0 < s_1 < ... < s_n < s$ and possible state $i_0, ..., i_n, i, j$ we have
$P(X_{t+s} = j|X_s = i, ..., X_{s_0} = i_0 = P(X_t = j|X_0 = i)$
In words given the present state the rest of the past is irrelevant for predicting the future. Note that built into the definition is the fact that the probability of going from i at time s to j at time $s + t$ only depends on t the difference in the times.

Discrete time Markov Chains chains were described by giving their transition probabilities $\mathbf{P}(i, j)$ = the probability of jumping from i to j in one step. In continuous time there is no first time $t > 0$, so we introduce for each $t > 0$ a transition probability
$p_t(i, j) = P(X_t = j|X_0 = i)$
In continuous time, as in discrete time, the transition probability satisfies the following theorem:

Theorem 15 (Chapman Kolmogorov Equation) $\sum_k p_s(i, k)p_t(k, j) = p_{s+t}(i, j$

This makes sense as in order for the chain to go from i to j in time $s+t$ it must be in some intermediary state k at time s and the Markov property implies the independence of the rest of its trip.
$q(i, j) = \lim_{h\to 0} \frac{p_h(i,j)}{h}$ for $j \neq i$
If this limit exists we will call $q(i, j)$ the jump rate from i to j.
We can construct a "routing matrix" with:
$\lambda_i = \sum_{j\neq i} q(i, j)$, $r(i, j) = \frac{q(i,j)}{\lambda_i}$
Here the "routing matrix," is the probability the chain goes to j when it leaves i.

6.1 Computing the Transition Probability

How do you compute the transition probability pt from the jump rates q?
We introduce

$$Q(i, j) = \begin{cases} q(i, j), if j \neq i \\ -\lambda_i, if j = i \end{cases}$$

Definition 6.1 (Kolmogorovs backward Equation) $p'_t = Qp_t$

Where if we want to set $p_t = e^{Qt}$ to solve the differential equation we do matrix exponentiation via the taylor series.
$e^{Qt} = \sum_{n=0}^{\infty} Q^n \frac{t^n}{n!}$

Definition 6.2 (Kolmogorovs backward Equation) $p'_t = p_tQ$

6.2 Limiting Behavior

Having already worked hard to develop the convergence theory for discrete time chains, the results for the continuous time case follow easily. In fact the study of the limiting behavior of continuous time Markov chains is simpler than the theory for discrete time chains, since the randomness of the exponential holding times implies that we don't have to worry about aperiodicity.

Definition 6.3 *The Markov Chain X_t is* ***irreducible****, if for any two statesd i and j it is possible to get from i to j in a finite number of jumps.*

Lemma 16 *If X_t is irreducible and $t > 0$ then $p_t(i,j) > 0 \forall i, j$*

In discrete time for a stationary distribution we were solving $\pi\mathbf{P} = \pi$, now for the continuous time instead of using the p_ts we can use the rate matrix $Q(i,j)$

Lemma 17 *π is a stationary distribution if and only if $\pi Q = 0$*

This follows from the flow into a state needing to be equal to the flow out of the state for a stationary/unchanging setup.

Theorem 18 *If a continuous time Markov Chain X_t is irreducible and has a stationary distribution π then*
$\lim_{t\to\infty} p_t(i,j) = \pi(j)$

6.2.1 Detailed Balance Condition

Generalizing from discrete time we can formula the condition:
$\pi(k)q(k,j) = \pi(j)q(j,k)\forall j \neq k$
(*)

Theorem 19 *If (*) holds then π is a stationary distribution.*

The intuition is that this covers the need for flow in and flow out to be the same so it guarantees π to be a stationary distribution.

6.3 Exit Distributions and Exit Times

6.3.1 Exit Distribution

We will approach this first using the embedded jump chain with transition probability
$\lambda_i = \sum_{j\neq i} q(i,j)$, $r(i,j) = \frac{q(i,j)}{\lambda_i}$
Let $V_k = min(t \geq 0 : X_t = k)$ be the time of the first visit to k and let $T_k = min(t \geq 0 : X_t = k, X_s \neq k$ for some $s < t$ be the time of the first return. The second definition is made complicated as if $X_0 = k$ then the chain stays at

k for an amount of time that is exponential with rate λ_k.
We then would proceed similar to the discrete time case.

6.3.2 Exit times

Now for the continuous time analogue we define
$g(i) = \frac{1}{\lambda_i} + \sum_{j \neq i} \frac{q(i,j)}{\lambda_i} g(j)$
And again we proceed similar to discrete time its just now we are adding on a wait of $\frac{1}{\lambda_i}$ instead of 1 when time increases in steps.

7 Martingales

Given an event A we define its **indicator function**

$$1_A = \begin{cases} 1 x \in A \\ 0 x \in A^c \end{cases}$$

In words, 1_A is "1 on A" (and 0 otherwise). Given a random variable Y, we define the integral of Y over A to be $E[Y; A] = E[Y 1_A]$.
Note that multiplying Y by 1_A sets the product = 0 on A^c and leaves the values on A unchanged. Finally we define the **conditional expectation of Y given A** to be $E[Y|A] = \frac{E(Y;A)}{P(A)}$.

Lemma 20 *If X is a constant c on A, then $E(XY|A) = cE(Y|A)$.*

Lemma 21 *(Jensen's Inequality) If ϕ is convex, then $E(\phi(X)|A) \geq \phi(E(X|A))$*

Lemma 22 *If B is the disjoint union of $A_1, ..., A_k$, then*
$E(Y; B) = \sum_{j=1}^{k} E(Y; A_j)$

Lemma 23 *If B is the disjoint union of $A_1, ..., A_k$, then*
$E(Y|B) = \sum_{j=1}^{k} E(Y|A_j) * \frac{P(A_j)}{P(B)}$

In particular when $B = \Omega$ we just get the law of total expectation.

Definition 7.1 (Martingale) *Thinking of M_n as the amount of money at time n for a gambler betting on a fair game, and X_n as the outcomes of the gambling game we say that $M_0, M_1, ...$ is a **martingale** with respect to $X_0, X_1, ...$ if*
i) for any $n \geq 0$ we have $E|M_n| < \infty$
ii) the value of M_n can be determined from the values for $X_n, ..., X_0$ and M_0
iii) for any possible values $x_n, ..., x_0$
$E(M_{n+1} - M_n | X_n = x_n, ..., X_0 = x_0, M_0 = m_0) = 0$

The first condition is just needed for "integrability", a well defined conditional expectation.
The second condition just reflects the possibility of a loss of information that should be avoided.
The third condition says that conditional on the past up to time n, the average profit from the bet on the nth game is 0. ("fair game")
Note we define: $A_v = (X_n = x_n, ..., X_0 = x_0, M_0 = m_0)$

Definition 7.2 (Supermartingale) *In most cases, casino games are not fair but biased against the player. We say that M_n is a **supermartingale** with respect to X_n if a gambler's expected winnings on one play are negative:*
$E(M_{n+1} - M_n | A_v) \leq 0$

Definition 7.3 (Submartingale) *Similarly defined the **submartingale** with respect to X_n has:*
$E(M_{n+1} - M_n | A_v) \geq 0$

Theorem 24 *Let X_n be a Markov chain with $\boldsymbol{P}$ and let $f(x, n)$ be a function of the state x and the time n so that*
$f(x, n) \geq \sum_y \boldsymbol{P}(x, y) f(y, n+1)$
Then $M_n = f(X_n, n)$ is a super martingale with respect to X_n.

This result holds similarly based on the $\geq$, $\leq$, or $=$ sign for submartingales and martingales.

7.1 Gambling Strategies, Stopping Times

The first result should be intuitive if we think of supermartingale as betting on an unfavorable game: the expected value of our fortune will decline over time.

Theorem 25 *If M_n is a supermartingale and $m \leq n$ then $E[M_m] \geq E[M_n]$*

Theorem 26 *If M_n is a submartingale and $0 \leq m < n$ then $E[M_m] \leq E[M_n]$*

Note a process is a martingale if and only if it is both a supermartingale and a submartingale we can conclude:

Theorem 27 *If M_n is a martingale and $0 \leq m < n$ then $E[M_m] = E[M_n]$*

The most famous result of martingale theory is that "you can't beat an unfavorable game".

The amount of money we bet on the n-th game, H_n, clearly, cannot depend on the outcome of that game, nor is it sensible to allow it to depend on the outcomes of games that will be played later. We say that H_n is an admissible gambling strategy or **predictable process** if for each n the value of H_n can be determined from $X_{n-1}, X_{n-2}, ..., X_0, M_0$.

To motivate the next definition, which we give for a general M_m think of H_m as the amount of stock we hold between time $m-1$ and m, and M_m the price of a stock at time m. Then our wealth at time n is
$W_n = W_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1})$
since the change in our wealth from time $m-1$ to m is the amount we hold times the changer in the price of the stock $H_m(M_m - M_{m-1})$

Theorem 28 *Suppose that M_n is a supermartingale with respect to X_n, H_n is predictable, and $0 \leq H_n \leq c_n$ where c_n is a constant that may depend on n. Then*
$W_n = W_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1})$ *is a supermartingale*

Note: We need non-negative H_n as if we were betting a negative amount of money we would become the "house" and we need to be bounded just for conditional expectations to make sense.

Definition 7.4 *We say that T is a **stopping time** with respect to X_n if the occurrence of the event $T = n$ can be determined from the information known at time n, $X_n, ..., X_0, M_0$.*

Theorem 29 *If M_n is a supermartingale with respect to X_n and T is a stopping time, then the stopped process $M_{T \wedge n}$ is a supermartinglae with respect to X_n. In particular $E[M_{T \wedge n}] \leq E[M_0]$*

This result also holds similar for submartingales and martingales if we change the inequality.

Theorem 30 *Suppose M_n is a martingale and T a stopping time with $P(T < \infty) = 1$ and $|M_{T \wedge n}| \leq K$ for some constant K, Then $E[M_T] = E[M_0]$.*