Lecture notes of Stochastic Process

lectured by prof. Hsueh-I Lu

pishen

AlgoLab, CSIE, NTU

June 17, 2012

Thank list

LeoSW, windker

Stochastic Process

Definition

A Stochastic process is a set of random variables $\{X(t)|t \in T\}$ where T is a index (time) set.
State Space: possible value of X(t) for each t, which is defined as subset of R.

Markov Chain

Definition

A Stochastic Process $\mathbb{X}$ with state space S is a Markov Chain if $\exists 0 \leq p_{ij} \leq 1 \quad \forall i,j \in S$ such that

$$
\begin{aligned}
(a) &\quad \sum_{j \in S} p_{ij} = 1 \quad \forall i \in S \\
(b) &\quad P(X(t+1) = j | X(0) = i_0, X(1) = i_1, ..., X(t) = i) = p_{ij} \\
&\qquad \forall t, i_0, i_1, ..., i_{t-1}
\end{aligned}
$$

$\mathbb{P}$ denotes the matrix form of p_{ij} with sum of any row is 1.

Lemma: $P(X(n) = j | X(0) = i) = \mathbb{P}^n[i,j]$

Proof of lemma

We know statement is true for $(m+n)=0$. For $(m+n)>0$:

$$
\begin{aligned}
&P(X(m+n)=j|X(0)=i) \\
&\quad = \sum_{k\in S} P(X(m+n)=j \text{ and } X(m)=k|X(0)=i) \\
&\quad = \sum_{k\in S} P(X(m+n)=j|X(m)=k \text{ and } X(0)=i)\cdot \\
&\qquad P(X(m)=k|X(0)=i) \\
&\quad = \sum_{k\in S} P(X(m+n)=j|X(m)=k)\cdot P(X(m)=k|X(0)=i) \\
&\quad = \sum_{k\in S} P^n[k,j]\cdot P^m[i,k] \\
&\quad = \sum_{k\in S} P^m[i,k]\cdot P^n[k,j] \\
&\quad = \mathbb{P}^n[i,j]
\end{aligned}
$$

Proof of lemma(cont)

$=$

: conditional on $X(m)$

$=$: definition of conditional probability

$=$: (see next page)

$=$: inductive hypothesis

Proof of lemma(cont)

$$
\begin{aligned}
&P(X(m+n)=j|X(m)=k \text{ and } X(0)=i)\\
&\quad = \sum_{r\in S} P(X(m+n)=j|\\
&\qquad X(m+n-1)=r \text{ and } X(m)=k \text{ and } X(0)=i)\cdot\\
&\qquad P(X(m+n-1)=r|X(m)=k \text{ and } X(0)=i)\\
&\quad = \sum_{r\in S} P(X(m+n)=j|X(m+n-1)=r)\cdot\\
&\qquad P(X(m+n-1)=r|X(m)=k)\\
&\quad = P(X(m+n)=j|X(m)=k)
\end{aligned}
$$

$=$: conditional on $X(m+n-1)$
$=$: first part by definition of Markov chain and second part by inductive hypothesis

Absorbing State

Let $\mathbb{A}$ be a set of accepting states. We would like to know the probability that $\mathbb{X}$ has ever entered some state in $\mathbb{A}$. Technique: merge all state of $\mathbb{A}$ into a new absorbing state a. Set matrix of $\mathbb{X}$ by once enter a, then probability of a goes to a is 1.

Recurrent & transient

Definition

The *recurrent probability* of state i of Markov chain $\mathbb{X}$ is

$$f_i = P(\text{there exists an index } t \geq 1 \text{ with } X(t) = i | X(0) = i)$$

- State i of $\mathbb{X}$ is *recurrent* if $f_i = 1$.
- State i of $\mathbb{X}$ is *transient* if $f_i < 1$.

Recurrent & transient (cont.)

- If state i is recurrent, by the property of Markov chain, once it re-enter the state i, we can take it as starting from $X(0)$ again. Hence we know that it will keep re-entering the state i again and again in the process.
- If state i is transient, in each period it start going from i, it may have probability $1 - f_i$ that it won't come back anymore.
  Hence the probability that the process will be in state i for exactly n periods equals $f_i^{n-1}(1 - f_i)$, $n \geq 1$, which is a geometric distribution.

Recurrent & transient (cont.)

- From the preceding page, it follows that state i is recurrent if and only if, starting in state i, the expected number of steps that the process is in state i is infinite.
- We can also derive that, if the Markov chain has finite states, at least one state is recurrent.

Expected number of visits

Let

$$I(n) = \begin{cases} 1 & \text{if } X(n) = i \\ 0 & \text{if } X(n) \neq i \end{cases}$$

we have $\sum_{n=0}^{\infty} I(n)$ represents the number of steps that the process is in state i, and

$$\begin{aligned} E\left[\sum_{n=0}^{\infty} I(n)|X(0) = i\right] &= \sum_{n=0}^{\infty} E[I(n)|X(0) = i] \\ &= \sum_{n=0}^{\infty} 1 \cdot P(X(n) = i|X(0) = i) \\ &= \sum_{n=0}^{\infty} P_{ii}^{n} \end{aligned}$$

We set $T = \sum_{n=0}^{\infty} I(n)$

Lemma 1

From the above statements, we prove the following

Lemma

State i is

$$\textit{recurrent} \iff \sum_{n=0}^{\infty} P_{ii}^{n} = \infty,$$

$$\textit{transient} \iff \sum_{n=0}^{\infty} P_{ii}^{n} < \infty$$

Proof of Lemma 1

($\Rightarrow$:)
($\Leftarrow$:)
Suppose state i is transient($f_i < 1$), consider $P(T=k) = f_i^{k-1} \cdot (1-f_i)$.
Since T is a geometric distribution, we have

$$E[T] = \sum_{k=0}^{\infty} k \cdot f_i^{k-1} \cdot f_i$$
$$= \frac{1}{1-f_i} < \infty$$

Communicated states

Definition

State i and j *communicate*, denoted $i \leftrightarrow j$, if there exist integers $m \geq 0$ and $n \geq 0$ such that

$$P_{ij}^{m} > 0 \text{ and } P_{ji}^{n} > 0$$

We say a Markov chain $\mathbb{X}$ is irreducible if $i \leftrightarrow j \quad \forall i, j \in S$

Lemma 2

Lemma

If $i \leftrightarrow j$, then the following statements hold.

- *State i is recurrent if and only if state j is recurrent.*
- *State i is transient if and only if state j is transient.*

Corollary: $\mathbb{X}$ is finite and irreducible $\implies$ all states are recurrent.

- *$\mathbb{X}$ is finite $\implies$ $\exists i \in S$ is recurrent (proof later)*
- *By Lemma 2, all states are recurrent*

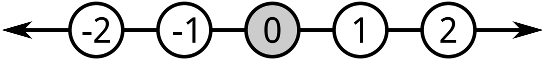
Proof

Let m and n be nonnegative integers with $P_{ij}^m, P_{ji}^n > 0$. Suppose that state j is recurrent, i.e., $\sum_{t=0}^{\infty} P_{jj}^t = \infty$. We have

$$\begin{aligned}
\sum_{t=0}^{\infty} P_{ii}^t &\geq \sum_{t=0}^{\infty} P_{ii}^{m+t+n} \\
&\geq \sum_{t=0}^{\infty} P_{ij}^m \cdot P_{jj}^t \cdot P_{ji}^n \\
&= P_{ij}^m \cdot P_{ji}^n \cdot \sum_{t=0}^{\infty} P_{jj}^t = \infty
\end{aligned}$$

Thus, state i is also recurrent.

Infinite drunken man problem



Let the state space consist of all integers. Let $X(0) = 0$ (i.e. at time 0 the drunken man is in state 0). The transition probabilities are such that

$$P_{i,(i+1)} = P_{i,(i-1)} = 0.5$$

holds for all states i of $\mathbb{X}$.

Gambler's ruin

Outline

1. Limiting probabilities
2. Stationary distribution
3. Long-run proportion
4. (Inverse of) Expected return time

Limiting Probabilities

Definition

Number π_j is the *limiting probability* of j if

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

holds for all states $i \in S$ ($S \subseteq \mathbb{N}$ is the state space).

- π_j is independent of i.
- $\lim_{n\to\infty} P^n = \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix}$, where $\pi = (\pi_1, \pi_2, \ldots)$

Stationary Probability Distribution

Definition

Non-negative row vector $\pi = (\pi_1, \pi_2, \ldots)$ is a *stationary probability distribution* of $\mathbb{X}$ if $\pi \times P = \pi$ holds and $\sum_{i \in S} \pi_i = 1$

- π is a normalized left eigenvector with eigenvalue $= 1$.
- If $X(0)$ has distribution π, then $X(t)$ has the same distribution π for all $t \geq 1$. π is also called as *steady-state distribution*.
- It doesn't mean that each $X(t)$ become independent. π only means the distribution of $X(t)$ when the previous random variable's value is unknown.

Theorem 1

Theorem

Let $\mathbb{X}$ be an irreducible, aperiodic, positive recurrent Markov chain, then

- *The limiting probability π_j of each state j exists.*
- *$\pi = (\pi_1, \pi_2, \ldots)$ is the unique stationary probability distribution.*

- The proof will be stated at page 38.

Expected return time

Definition

The *expected return time* of state $i \in S$ is

$$\mu_i = \sum_{n \geq 1} n \cdot f_i^{(n)}$$

where

$$f_i^{(n)} = P(\min\{t : X(t) = i, t \geq 1\} = n | X(0) = i)$$

- $f_i = \sum_{n \geq 1} f_i^{(n)}$

Positive recurrent & null recurrent

Definition

State i is *positive recurrent* if $\mu_i < \infty$

Definition

State i is *null recurrent* if $\mu_i = \infty$

- Both are recurrent states, and are *class properties*, which means that if state i and j communicate, they will share this property.
- If $\mathbb{X}$ is finite, then each recurrent state of $\mathbb{X}$ is positive recurrent. Proof stated at page 63.

Example of null recurrent

Example

For a Markov chain with n states $(1, \ldots, n)$, if

$$P(X(t+1) = i+1 | X(t) = i) = 1 - 1/n$$

and

$$P(X(t+1) = 1 | X(t) = i) = 1/n$$

According to geometric distribution (taking $p = 1/n$), the expectation value of "steps taken for state 1 to come back" will be $1/p = n$, hence $\lim_{n\to\infty} n = \infty$.

Period of a chain

Definition

The *period* of state i is d if d is the largest integer such that

$$P_{ii}^{n} = 0$$

holds for all n which is not divisible by d.

Definition

If each state of $\mathbb{X}$ has period 1, then $\mathbb{X}$ is called *aperiodic*.

- If $P_{ii} > 0$ for all $i \in S$, then $\mathbb{X}$ is aperiodic.
- Period can be seen as the gcd of all n that have $P_{ii}^{n} > 0$, note that $P_{ii}^{\mathrm{gcd}} > 0$ is not necessary.
- The period of drunken man problem is 2.

Lemma 1

Lemma

If state j has period 1 and is positive recurrent, then

$$\pi_{ij} \equiv \lim_{n\to\infty} P_{ij}^n$$

exists and is positive for all states $i \in S$.

- This can be proved by the Blackwell theorem in Renewal theory.
- It doesn't promise that $\pi_{ij} = \pi_{i'j}$ for any $i, i' \in S$. But they will be the same if we add the irreducible property ($i \leftrightarrow i'$).

Property of lim

- The position of $\lim$ may not be switched arbitrarily in an equation.

Example

$$1 = \lim_{n\to\infty} \lim_{m\to\infty} \frac{m}{m+n} \neq \lim_{m\to\infty} \lim_{n\to\infty} \frac{m}{m+n} = 0$$

- $\lim$ would not influence the inequality.

Example

If $f(n) \geq g(n)$, then $\lim_{n\to\infty} f(n) \geq \lim_{n\to\infty} g(n)$

Property of lim (cont.)

- $\lim$ is linear operator under finite number of functions.

Example

For $m < \infty$,

$$\sum_{i=1}^{m} \lim_{n\to\infty} f_i(n) = \lim_{n\to\infty} \sum_{i=1}^{m} f_i(n)$$

need an example of $m = \infty$

Inequality 1

Inequality

$$\sum_{j \in S} \pi_{ij} \leq 1 \quad \forall i \in S$$

Proof

$$\begin{aligned}
\lim_{m\to\infty}\sum_{j=1}^{m}\pi_{ij} &= \lim_{m\to\infty}\sum_{j=1}^{m}\lim_{n\to\infty}P_{ij}^{n} \\
&= \lim_{m\to\infty}\lim_{n\to\infty}\sum_{j=1}^{m}P_{ij}^{n} \\
&\leq \lim_{m\to\infty}\lim_{n\to\infty}\sum_{j\in S}P_{ij}^{n} = 1
\end{aligned}$$

- The last equation works since $\sum_{j\in S}P_{ij}^{n}=1$.

Inequality 2

Inequality

For state $j \in S$, we have

$$\pi_{ij} \geq \sum_{k \in S} \pi_{ik} P_{kj}$$

Proof

For $m \geq 1$ and $n \geq 1$,

$$P_{ij}^{n+1} = \sum_{k \in S} P_{ik}^{n} P_{kj} \geq \sum_{k=1}^{m} P_{ik}^{n} P_{kj}$$

then

$$\pi_{ij} = \lim_{n \to \infty} P_{ij}^{n+1} \geq \lim_{n \to \infty} \sum_{k=1}^{m} P_{ik}^{n} P_{kj} = \sum_{k=1}^{m} \lim_{n \to \infty} P_{ik}^{n} P_{kj} = \sum_{k=1}^{m} \pi_{ik} P_{kj}$$

hence, we know

$$\lim_{m \to \infty} \pi_{ij} = \pi_{ij} \geq \lim_{m \to \infty} \sum_{k=1}^{m} \pi_{ik} P_{kj} = \sum_{k \in S} \pi_{ik} P_{kj}$$

Equality 1

Equality

$$\pi_{ij} = \sum_{k \in S} \pi_{ik} P_{kj}$$

Proof

Assume for contradiction $\pi_{ij} > \sum_{k \in S} \pi_{ik} P_{kj}$, then

$$
\begin{aligned}
\lim_{m\to\infty} \sum_{j=1}^{m} &> \lim_{m\to\infty} \sum_{j=1}^{m} \lim_{p\to\infty} \sum_{k=1}^{p} \pi_{ik} P_{kj} \\
&= \lim_{m\to\infty} \lim_{p\to\infty} \sum_{j=1}^{m} \sum_{k=1}^{p} \pi_{ik} P_{kj} \\
&= \lim_{m\to\infty} \lim_{p\to\infty} \sum_{k=1}^{p} \pi_{ik} \sum_{j=1}^{m} P_{kj} \\
&= \lim_{p\to\infty} \sum_{k=1}^{p} \pi_{ik} \lim_{m\to\infty} \sum_{j=1}^{m} P_{kj} \\
&= \lim_{p\to\infty} \sum_{k=1}^{p} \pi_{ik} \cdot 1 = \lim_{p\to\infty} \sum_{k=1}^{p} \pi_{ik}
\end{aligned}
$$

Proof (cont.)

- Since a value cannot be greater than itself, we got contradiction.
- In the 4th line, two $\lim$ can be switched because the value can only get larger when applying $\lim$ on it. not sure

Proof of theorem 1

- **Step 0**: existence of limiting probability.
- **Step 1**: existence of stationary probability distribution.
- **Step 2**: uniqueness.

0. Existence of limiting probability

Proof.

By lemma 1, we know that there exists a π_j for row i. Since the Markov chain is irreducible and all the states are positive recurrent, for any state i' other than i, we know that i' surely will visit i in finite steps. Therefore, the π_j value at row i' will equal to the π_j value at row i, which means that all the π_j for column j are the same, and is the limiting probability. □

still not clear enough

1. Existence of stationary probability distribution

We want to prove that

Target

There's a vector $s = (s_1, s_2, \ldots)$ *such that*

1. $\sum_{i \in S} s_i = 1$
2. $s \times P = s$

Proof.

By lemma 1, we know that there exists a $\pi = (\pi_1, \pi_2, \ldots)$.
And by equality 1, we know that

$$(\pi_1, \pi_2, \ldots) \times P = (\pi_1, \pi_2, \ldots)$$

Hence π can satisfy the 2nd part of our target.
Then, we take $k = \sum_{i \in S} \pi_i$. By inequality 1, we know that $k < \infty$, and can get

$$(\frac{\pi_1}{k}, \frac{\pi_2}{k}, \ldots) \times P = (\frac{\pi_1}{k}, \frac{\pi_2}{k}, \ldots)$$

where $\sum_{i \in S} \frac{\pi_i}{k} = 1$ also satisfy the 1st part of our target.
Therefore, this vector can be s, which means that it exists. □

2. Uniqueness

Target

If $s = (s_1, s_2, \dots)$ is a stationary distribution of $\mathbb{X}$, then $s = \pi$.

- We'll prove this by inequality 3 & 4.

Inequality 3

Inequality

$$s_j \geq \pi_j, \forall j \in S$$

Proof.

Let the distribution of $X(0)$ be s, by the property of stationary distribution, we have

$$\begin{aligned}
s_j = P(X(n) = j) &= \sum_{i \in S} P(X(n) = j | X(0) = i) P(X(0) = i) \\
&= \sum_{i \in S} P_{ij}^n \cdot s_i \\
&\geq \sum_{i=1}^{m} P_{ij}^n \cdot s_i \\
\Rightarrow s_j &= \lim_{m \to \infty} \lim_{n \to \infty} s_j \\
&\geq \lim_{m \to \infty} \lim_{n \to \infty} \sum_{i=1}^{m} P_{ij}^n \cdot s_i = \lim_{m \to \infty} \sum_{i=1}^{m} \pi_j \cdot s_i = \pi_j
\end{aligned}$$

$\square$

Inequality 4

Inequality

$$s_j \leq \pi_j, \forall j \in S$$

Proof.

Similar in the proof above, $\forall m, n \geq 1$, we have

$$\begin{aligned} s_j &= \sum_{i \in S} P_{ij}^n \cdot s_i \\ &\leq \sum_{i=1}^{m} P_{ij}^n \cdot s_i + \sum_{i=m+1}^{\infty} s_i \\ \Rightarrow s_j &= \lim_{m \to \infty} \lim_{n \to \infty} s_j \\ &\leq \lim_{m \to \infty} \lim_{n \to \infty} \left(\sum_{i=1}^{m} P_{ij}^n \cdot s_i + \sum_{i=m+1}^{\infty} s_i \right) \\ &= \pi_j \end{aligned}$$

□

An example Markov chain

Example

$$P = \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix}, 0 < \alpha, \beta < 1$$

$$\pi = \left(\frac{\beta}{1+\beta-\alpha}, \frac{1-\alpha}{1+\beta-\alpha} \right)$$

Real world example: Hardy-Weinberg Law

Example

There're two kinds of allele:

- dominant: **A**
- recessive: **a**

And three kinds of senotype with population proportion as follow:

- AA: p
- aa: q
- Aa: $r = 1 - (p + q)$

Example (cont.)

$$P = \begin{array}{c c} & \begin{array}{ccc} AA & aa & Aa \end{array} \\ \begin{array}{c} AA \\ aa \\ Aa \end{array} & \begin{pmatrix} p+\frac{r}{2} & 0 & q+\frac{r}{2} \\ 0 & q+\frac{r}{2} & p+\frac{r}{2} \\ \frac{p}{2}+\frac{r}{4} & \frac{p}{2}+\frac{r}{4} & \frac{p+q+r}{2} \end{pmatrix} \end{array}$$

we get $\pi = (p, q, r)$ when

- $p = \left(p + \frac{r}{2}\right)^2$
- $q = \left(q + \frac{r}{2}\right)^2$
- $r = 2\left(p + \frac{r}{2}\right)\left(q + \frac{r}{2}\right)$

Long-run proportion

Definition

We say that r_j is the *long-run proportion* of state $j \in S$ if

$$r_j = \lim_{n \to \infty} \frac{1}{n} \sum_{1 \le t \le n} P_{ij}^t$$

holds for each state $i \in S$.

- It represents the average appearance times of state j in the whole process.
- We will show that (in theorem 3) if $\mathbb{X}$ is irreducible, then the long-run proportion of all states exist.

Theorem 2

Theorem (type 1)

If r_j *exists for each* $j \in S$ *and* $\sum_{j \in S} r_j > 0$, *then* $r = (r_1, r_2, \ldots)$ *is the unique stationary distribution of* $\mathbb{X}$.

or

Theorem (type 2)

If r_j *exists for each* $j \in S$ *and* **a stationary distribution exists**, *then* $r = (r_1, r_2, \ldots)$ *is the unique stationary distribution of* $\mathbb{X}$.

Proof

Existence of stationary distribution in type 1:
Let

$$R = \begin{pmatrix} r \\ r \\ \vdots \end{pmatrix} = \lim_{n\to\infty} \frac{1}{n} \sum_{1\le t\le n} P^t$$

then

$$\begin{aligned} R \times P &= \lim_{n\to\infty} \frac{1}{n} \sum_{1\le t\le n} P^{t+1} \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{1\le t\le n} P^t + \lim_{n\to\infty} \frac{1}{n}(P^{n+1} - P) \\ &= R \end{aligned}$$

As stated later, $\sum_{j\in S} r_j \le 1$, hence by normalizing r, we prove that stationary distribution exist.

- $(\lim f(n)) \cdot g(n) = \lim f(n) \cdot g(n)$?
- can replace the proof on page 40?

Proof (cont.)

Uniqueness:

Let π be an arbitrary stationary distribution, then

$$
\begin{aligned}
r &= \pi \times R \\
&= \pi \times \lim_{n\to\infty} \frac{1}{n} \sum_{1\le t\le n} P^t \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{1\le t\le n} \pi \times P^t \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{1\le t\le n} \pi \\
&= \pi
\end{aligned}
$$

can replace the proof for page 42?

Proof (cont.)

Prove that $\sum_{j \in S} r_j \leq 1$:

$$\begin{aligned}
\sum_{j \in S} r_j &= \lim_{m \to \infty} \sum_{j=1}^{m} \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} P_{ij}^{t} \\
&= \lim_{m \to \infty} \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{m} P_{ij}^{t} \\
&\leq \lim_{m \to \infty} \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{j \in S} P_{ij}^{t} \\
&= \lim_{m \to \infty} \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} 1 = 1
\end{aligned}$$

Example 1

On a highway, if we know the probability that

- A truck is followed by a truck: $1/4$
- A truck is followed by a car: $3/4$
- A car is followed by a truck: $1/5$
- A car is followed by a car: $4/5$

We can construct a matrix

$$\begin{array}{cc} & \begin{array}{cc} T & C \end{array} \\ \begin{array}{c} T \\ C \end{array} & \begin{pmatrix} 1/4 & 3/4 \\ 1/5 & 4/5 \end{pmatrix} \end{array}$$

and get the portion of trucks and cars on the whole highway as the eigenvector $(4/19, 15/19)$ (we will know that long-run proportion exists by Theorem 3).

Example 2

For a system which has several good and bad states, we have a matrix P:

$$
\begin{array}{c} \\ g_1 \\ g_2 \\ \vdots \\ b_1 \\ b_2 \\ \vdots \end{array}
\begin{array}{c}
\begin{array}{cccccc} g_1 & g_2 & \cdots & b_1 & b_2 & \cdots \end{array} \\
\left(\begin{array}{cccccc} \; & \; & \; & \; & \; & \; \\ \\ \\ \\ \\ \\ \end{array}\right)
\end{array}
$$

Example 2 (cont.)

Q1: Breakdown rate (breakdown times / total time)
The long-run frequency of going to a bad state from a good state is

$$\sum_{i \in g} \sum_{j \in b} r_i P_{ij}$$

Example 2 (cont.)

Q2: The expected time μ_G (resp. μ_B) of staying in good (resp. bad) states once we reach a good (resp. bad) state?

Ans:

For each $t = 1, 2, \ldots$, let G_t (resp. B_t) be the length of the t-th good (resp. bad) phase of consecutive good (resp. bad) states. By the strong law of large numbers,

$$P\left(\lim_{t\to\infty} \frac{G_1 + B_1 + G_2 + B_2 + \cdots + G_t + B_t}{t} = \mu_G + \mu_B\right) = 1$$

Since the reciprocal of above is the breakdown rate, we get equation (1):

$$P\left(\sum_{i\in G}\sum_{j\in B} \pi_i P_{ij} = \frac{1}{\mu_G + \mu_B}\right) = 1$$

Example 2 (cont.)

Also, with probability 1, we get equation (2):

$$P\left(\sum_{i\in G} r_i = \lim_{t\to\infty} \frac{G_1+G_2+\cdots+G_t}{G_1+B_1+\cdots+G_t+B_t} = \frac{\mu_G}{\mu_G+\mu_B}\right) = 1$$

Then, by $(2)/(1)$, we get that

$$P\left(\mu_G = \frac{\sum_{i\in G} r_i}{\sum_{i\in G}\sum_{j\in B} r_i P_{ij}}\right) = 1$$

- $\lim \frac{f(n)}{g(n)} = \frac{\lim f(n)}{\lim g(n)}$?

Theorem 3

Theorem

If $\mathbb{X}$ is irreducible, then the long-run proportion r_i exists with probability 1, moreover,

1. *If state i is positive recurrent (i.e. $0 < \mu_i < \infty$), then $P(r_i = \frac{1}{\mu_i}) = 1$.*
2. *If state i is null recurrent (i.e. $\mu_i = \infty$) or transient, then $P(r_i = 0) = 1$.*

where μ_i is the expected return time of state i

Proof

Part 1:
Suppose $X(0) = i$, T_k is the number of steps required for the k-th i goes to $(k+1)$-st i, then by the strong law of large number,

$$P\left(\lim_{k\to\infty} \frac{T_1 + T_2 + \cdots + T_k}{k} = \mu_i\right) = 1$$
$$\Rightarrow P\left(r_i = \lim_{k\to\infty} \frac{k}{T_1 + T_2 + \cdots + T_k} = \frac{1}{\mu_i}\right) = 1$$

- $\lim(A/B) = \frac{1}{\lim(B/A)}$?

Proof (cont.)

Part 2:

1. If i is transient, i will only appear finite times in the long-run, hence

$$r_i = \frac{finite}{\infty} = 0$$

2. If i is null recurrent, μ_i is ∞, then

$$P\left(\lim_{k\to\infty} \frac{T_1 + T_2 + \cdots + T_k}{k} = \infty\right) = 1$$

$$P\left(r_i = \lim_{k\to\infty} \frac{k}{T_1 + T_2 + \cdots + T_k} = 0\right) = 1$$

(The first equation is not promised by the strong law of large number. But if it's not ∞, we can say that μ_i is not ∞, which is a contradiction.)

Example 1

Example (type 1)

If $\mathbb{X}$ is **irreducible** and finite, then $\mathbb{X}$ has no null recurrent states.

Example (type 2)

If $\mathbb{X}$ is finite, then $\mathbb{X}$ has no null recurrent states.

- Finite irreducible imply positive recurrent.

Proof

- **Type 1:**
  If there's a state which is null recurrent, by irreducible, all the states will be null recurrent. Then, all states have $P(r_i = 0) = 1$. By changing the proof in page 54 into finite states version, we know that $\sum r_i = 1$. So it's impossible for finite r_i, which are all close to 0, to sum up to 1.
- **Type 2:**
  If it's not irreducible, the finite set of communicated null recurrent states still form an irreducible and finite Markov chain, which can fit the requirement of type 1.

Example 2

Example

In the drunken man problem with infinite states, no state will be positive recurrent.

- Infinite drunken man imply no positive recurrent. Note that it doesn't mean all infinite irreducible Markov chain has no positive recurrent state.

Proof

If all the states are positive recurrent, then by theorem 3, we know that all the $r_i > 0$ and is a finite value. Since each state of drunken man problem has the same structure, all the r_i has same value. We then set $r = \epsilon \cdot r_i$ $(0 < \epsilon < 1)$ such that $r_i > r > 0, \forall i$. And get

$$\sum_{i \in S} r_i > \sum_{i \in S} r = \infty > 1$$

which is contradiction to page 54.

Example 3: Poisson Hotel

Example

There's a hotel, with N representing the number of newly occupied rooms each day (N is a poisson distribution with parameter λ). And the number of consecutive check-in days of each room is a geometric distribution with probability p (p is the probability of check-out). $X(t)$ is the number of occupied rooms in day t.

Q1: $P_{ij} =?$

We set R_i as a binomial distribution with parameter $(i, 1-p)$, which represents the number of rooms which will remain occupied in the next day, then

$$\begin{aligned}
P_{ij} &= P(R_i + N = j) \\
&= \sum_{k\geq 0} P(R_i + N = j | R_i = k) P(R_i = k) \\
&= \sum_{k\geq 0} P(N = j-k) P(R_i = k) \\
&= \sum_{0\leq k\leq \min(i,j)} \frac{e^{-\lambda}\cdot\lambda^{j-k}}{(j-k)!}\binom{i}{k}(1-p)^k p^{1-k}
\end{aligned}$$

Q2: $r_i =$?

We guess (by a dream?) there's a stationary distribution which is a poisson distribution with parameter λ_0. Setting $X(0)$ with this distribution. And let R as the number of rooms in $X(0)$ which remain check-in in the next day (R is a poisson distribution with parameter $\lambda_0(1-p)$). $X(1)$ will have distribution $R+N$, which is a poisson distribution with parameter $\lambda_0(1-p)+\lambda$. Then since $X(0)$ is a stationary distribution, it will have the same distribution with $X(1)$, which means that $\lambda_0 = \lambda_0(1-p)+\lambda$, and we get $\lambda_0 = \lambda/p$.
After getting r_i, we get that with probability 1,

$$\mu_i = \frac{1}{P(X(0)=i)} = \frac{i!}{e^{-\lambda/p} \cdot (\lambda/p)^i}$$

not clear enough

Corollary of theorem 2 & 3

Corollary

If $\mathbb{X}$ is irreducible, then

$$\mathbb{X} \textit{ is positive recurrent } \iff \mathbb{X} \textit{ admits a stationary distribution.}$$

Moving to transient states

For transient states i and j, we define the following:

1. Expected steps in a transient state:

Definition

E is a matrix where E_{ij} is the expected number of steps t with $X(t) = j$ when $X(0) = i$.

2. Probability of reaching a transient state:

Definition

F is a matrix where

$$F_{ij} = P(X(t) = j \text{ for some } t \geq 1 | X(0) = i)$$

Computing E & F

Theorem

For a Markov chain $\mathbb{X}$ consisting finite transient states,

$$E = (I - T)^{-1}$$

where I is an identity matrix, T is the induced matrix of P by all the transient states in P. Moreover,

$$F_{ij} = \frac{E_{ij} - \delta_{ij}}{E_{jj}} \text{ ,where } \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Proof

Conditioned on $X(1)$, we have

$$E_{ij} = \underbrace{\delta_{ij}}_{\text{step}=0} + \underbrace{\sum_k P_{ik} \cdot E_{kj}}_{\text{step}\geq 1} = \delta_{ij} + \sum_k T_{ik} \cdot E_{kj}$$

The 2nd equation works since the process will not go back to transient state once it enter a recurrent state. Then, we have

$$\begin{aligned} I \times E = E &= I + T \times E \\ \Longrightarrow (I - T) \times E &= I \\ \Longrightarrow E &= (I - T)^{-1} \end{aligned}$$

Proof (cont.)

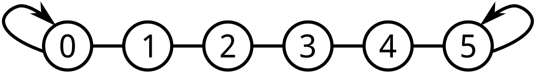
Conditioned on whether or not $X(t) = j$ holds for some $t \geq 1$, we have

$$E_{ij} = \underbrace{\delta_{ij}}_{\text{step}=0} + \underbrace{F_{ij} \cdot E_{jj}}_{\text{steps} \geq \text{the first } j}$$

therefore,

$$F_{ij} = \frac{E_{ij} - \delta_{ij}}{E_{jj}}$$

Example: Gambler's ruin



$$T = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0 \end{pmatrix} \end{array} \quad E = \begin{pmatrix} 1.6 & 1.2 & 0.8 & 0.4 \\ 1.2 & 2.4 & 1.6 & 0.8 \\ 0.8 & 1.6 & 2.4 & 1.2 \\ 0.4 & 0.8 & 1.2 & 1.6 \end{pmatrix}$$

$$F = \begin{pmatrix} 0.375 & 0.5 & 1/3 & 0.25 \\ 0.75 & 1.75/3 & 1.9/3 & 0.5 \\ 0.5 & 1.9/3 & 1.75/3 & 0.75 \\ 0.25 & 1/3 & 0.5 & 0.375 \end{pmatrix}$$

Branching process

In the beginning, there're $X(0)$ life forms, each life form has probability p_i of becoming i life forms in the next step.

- state 0 is recurrent (absorbing).
- if $p_0 > 0$, all other states $(1, 2, \ldots)$ are transient since $P(X(t+1) = 0 | X(t) = i) = {p_0}^i > 0$

We'll show that

$$E[X(n)] = \mu^n \cdot X(0)$$

where

$$\mu = \sum_{j \geq 1} j \cdot p_j = E[Z_k]$$

and Z_k is the number of offspring of the k-th life form, all Z_k are i.i.d.

Proof

$$
\begin{aligned}
E[X(n)] &= E[E[X(n)|X(n-1)]] \\
&= E\left[E\left[\sum_{k=1}^{X(n-1)} Z_k|X(n-1)\right]\right] \\
&= E[X(n-1)\cdot\mu] \\
&= \mu\cdot E[X(n-1)] \\
&= \mu^n\cdot X(0)
\end{aligned}
$$

Probability of extinction

Definition

e_i is the probability of extinction when $X(0) = i$.

Case 1: $\mu < 1$

$$
\begin{aligned}
1 - e_i &= \lim_{n\to\infty} P(X(n) \geq 1 | X(0) = i) \\
&= \lim_{n\to\infty} \sum_{j\geq 1} P(X(n) = j | X(0) = i) \\
&\leq \lim_{n\to\infty} \sum_{j\geq 1} j \cdot P(X(n) = j | X(0) = i) \\
&= \lim_{n\to\infty} E[X(n) | X(0) = i] \\
&= \lim_{n\to\infty} \mu^n \cdot i = 0
\end{aligned}
$$

Probability of extinction (cont.)

Case 2: $\mu \geq 1$

$$e_2 = {e_1}^2, \quad e_3 = e_2 \cdot e_1, \quad \ldots$$

$$\begin{aligned} e_1 &= P(\text{extinct} | X(0) = 1) \\ &= \sum_{j \geq 0} P(\text{extinct} | X(1) = j) \cdot P_{1j} \\ &= \sum_{j \geq 0} e_j \cdot p_j \\ &= \sum_{j \geq 0} {e_1}^j \cdot p_j \end{aligned}$$

We then solve the above equation to get e_1.

Example

$$
\begin{aligned}
& p_0 = p_1 = 0.25, \quad p_2 = 0.5 \\
\Longrightarrow\ & \mu = 1 \cdot 0.25 + 2 \cdot 0.5 > 1 \\
\Longrightarrow\ & e_1 = {e_1}^0 \cdot 0.25 + {e_1}^1 \cdot 0.25 + {e_1}^2 \cdot 0.5 \\
\Longrightarrow\ & e_1 = \{1/2,\ 1\}
\end{aligned}
$$

Since $\mu > 1$, we know $\lim_{n\to\infty} E[X(n)] = \infty$.
But if $e_1 = 1$, we have $\lim_{n\to\infty} P(X(n) = 0) = 1$, which would not make $\lim_{n\to\infty} E[X(n)] = \infty$, hence $e_1 \neq 1$.

Reversed Markov chain

Definition

Let $\mathbb{X}$ (resp. $\mathbb{Y}$) be a Markov chain with matrix P (resp. Q).
We say that $\mathbb{Y}$ is the *reversed chain* of $\mathbb{X}$ if there exists a stationary distribution π of $\mathbb{X}$ such that

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

holds for all states $i, j \in S$.

Observation 1

Observation

The reversed sequence $\mathbb{Y}$ of $\mathbb{X}$ is a Markov chain.

Proof

$$
\begin{aligned}
&P(Y(n) = i_0 | Y(n-1) = i_1, Y(n-2) = i_2, \ldots, Y(n-k) = i_k) \\
=&P(X(n) = i_0 | X(n+1) = i_1, X(n+2) = i_2, \ldots, X(n+k) = i_k) \\
=&\frac{P(X(n) = i_0, X(n+1) = i_1, \ldots, X(n+k) = i_k)}{P(X(n+1) = i_1, \ldots, X(n+k) = i_k)} \\
=&\frac{P(X(n) = i_0) \cdot P(X(n+1) = i_1 | X(n) = i_0) \cdot P_{i_1 i_2} \cdots P_{i_{k-1} i_k}}{P(X(n+1) = i_1) \cdot P_{i_1 i_2} \cdots P_{i_{k-1} i_k}} \\
=&\frac{P(X(n) = i_0, X(n+1) = i_1)}{P(X(n+1) = i_1)} \\
=&P(X(n) = i_0 | X(n+1) = i_1) \\
=&P(Y(n) = i_0 | Y(n-1) = i_1)
\end{aligned}
$$

Observation 2

Observation

If $\mathbb{Y}$ is the reversed sequence of Markov chain $\mathbb{X}$ and π is a stationary distribution of $\mathbb{X}$, then

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

holds for all $i, j \in S$, where Q is the transition matrix of $\mathbb{Y}$.

Proof

Let $\mathbb{X}$ and $\mathbb{Y}$ have distribution π

$$\begin{aligned}
\pi_i \cdot Q_{ij} &= P(Y(n-1) = i) \cdot P(Y(n) = j | Y(n-1) = i) \\
&= P(Y(n-1) = i, Y(n) = j) \\
&= P(Y(n-1) = i | Y(n) = j) \cdot P(Y(n) = j) \\
&= P(X(n+1) = i | X(n) = j) \cdot P(X(n) = j) = \pi_j \cdot P_{ji}
\end{aligned}$$

Observation 3

Observation

Let P (resp. Q) be the transition matrix of $\mathbb{X}$ (resp. $\mathbb{Y}$), if vector π satisfy the following

- $\sum_{i \in S} \pi_i = 1$
- $\pi_i \geq 0 \quad \forall i \in S$
- $\pi_i \cdot Q_{ij} = \pi_j \cdot P_{ji} \quad \forall i, j \in S$

then $\mathbb{Y}$ is the reversed sequence of $\mathbb{X}$.

- The long-run proportion of $i \to j$ in the sequence of $\mathbb{Y}$ is equal to the long-run proportion of $j \to i$ in the sequence of $\mathbb{X}$.
- Reversed Markov chain is the reversed sequence.

Proof

From the third property, we have

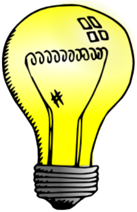
$$\sum_{j \in S} \pi_i \cdot Q_{ij} = \pi_i = \sum_{j \in S} \pi_j \cdot P_{ji} \quad \forall i \in S$$

From the 2nd equation, we know that $\pi \times P = \pi$, hence π is a stationary distribution of $\mathbb{X}$.
Then by observation 2, we know that for any π, there's a reversed sequence $\mathbb{Y}'$, whose transition matrix Q' satisfy

$$\pi_i \cdot Q'_{ij} = \pi_j \cdot P_{ji} \quad \forall i, j \in S$$

hence $\mathbb{Y} = \mathbb{Y}'$, which is a reversed sequence of $\mathbb{X}$.

Example: Bulb's life



There's a room which need to be lighted by one bulb, when the bulb in use fails, it will be replaced by a new one on next day.

- $X(n) = i$ if the bulb in use on day n is in its ith day of use.
- L is a random variable representing the lifetime of a bulb.

We want to know the stationary probability π_i of state i.

Example: Bulb's life (cont.)

$\mathbb{X}$ is a irreducible, positive recurrent, aperiodic Markov chain which has the sequence like this:

$$1, 2, 3, 1, 2, 3, 4, 5, 1, 1, 2, 1, 2, 3, 4, \ldots$$

We know that

$$P_{i1} = P(\text{buld, on its } i\text{th day of use, fails}) = \frac{P(L = i)}{P(L \geq i)} = 1 - P_{i(i+1)}$$

And the expected return time of state 1 is $E[L]$, which means that the long-run proportion of state 1 is $1/E[L]$ by page 60.

Example: Bulb's life (cont.)

Take $\mathbb{Y}$ (with matrix Q) as the reversed chain of $\mathbb{X}$, we know that for all $i \in S$,

- $Q_{(i+1)i} = 1$
- $Q_{1i} = P(L = i)$
- $\pi_1 \cdot Q_{1i} = \pi_i \cdot P_{i1}$

Hence,

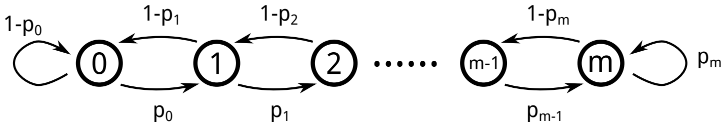
$$\pi_i = \frac{\pi_1 \cdot Q_{1i}}{P_{i1}} = \frac{P(L = i) \cdot P(L \geq i)}{E[L] \cdot P(L = i)} = \frac{P(L \geq i)}{E[L]}$$

Time-reversible

Definition

$\mathbb{X}$ is *time-reversible* if $\mathbb{X}$ is the reversed chain of $\mathbb{X}$.

Example: Reversed drunken man



- $0 < p_0 \leq 1$
- $0 \leq p_m < 1$
- $0 < p_i < 1 \quad \forall i = 1, \ldots, m-1$

The long-run proportion of transition $i \to i+1$ and $i+1 \to i$ are the same, since one must go back to i from $i+1$ in order to go to $i+1$ from i. Hence the drunken man problem is time-reversible.

Example: Reversed drunken man (cont.)

$$
\begin{aligned}
\pi_0 \cdot p_0 &= \pi_1 \cdot (1 - p_1) \\
\pi_1 \cdot p_1 &= \pi_2 \cdot (1 - p_2) \\
&\vdots \\
\pi_{m-1} \cdot p_{m-1} &= \pi_m \cdot (1 - p_m)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\pi_1 &= \pi_0 \cdot p_0 / (1 - p_1) \\
\pi_2 &= \pi_1 \cdot p_1 / (1 - p_2) \\
&\vdots \\
\pi_m &= \pi_{m-1} \cdot p_{m-1} / (1 - p_m)
\end{aligned}
$$

Example: Reversed drunken man (cont.)

$$\pi_i = \underbrace{\frac{\prod_{j=0}^{i-1} p_j}{\prod_{j=1}^{i}(1-p_j)}}_{q_i} \cdot \pi_0 \quad \forall i = 1, \ldots m$$

$$\Longrightarrow \pi_0 + \sum_{i=1}^{m} \pi_i = 1 = \pi_0 + \sum_{i=1}^{m} q_i \cdot \pi_0$$

$$\Longrightarrow \pi_0 = \frac{1}{1 + \sum_{i=1}^{m} q_i}$$

$$\Longrightarrow \pi_k = \frac{q_k}{1 + \sum_{i=1}^{m} q_i} \quad \forall k = 0, 1, \ldots m$$

Example: Two bukkits of balls

There're two bukkits contain total m balls.
In each step, we randomly choose one ball and put it in another bukkit.
Let $X(n)$ represent the number of balls in the first bukkit, it's the Markov chain of previous example with

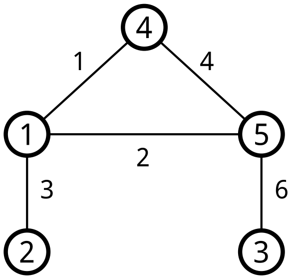
$$p_0 = 1,\ p_m = 0,\ p_i = \frac{m-i}{m} \quad \forall i = 1, \dots, m-1$$

We can get that

$$q_i = \frac{\prod_{j=0}^{i-1} \frac{m-j}{m}}{\prod_{j=1}^{i} \frac{j}{m}} = \frac{\prod_{j=0}^{i-1} m-j}{\prod_{j=1}^{i} j} = \binom{m}{i} \quad \forall i = 1, \dots m$$

$$\implies \pi_0 = \frac{1}{1 + \sum_{i=1}^{m} \binom{m}{i}} = \frac{1}{2^m} \implies \pi_k = \frac{\binom{m}{k}}{2^m} \quad \forall k = 0, 1, \dots m$$

Example: A random walk



$$P_{ij} = \frac{w(i,j)}{\sum_k w(i,k)}$$

where $w(a,b)$ is the weight of edge (a,b).
To make it as a time-reversible chain, we let

$$\pi_i = \frac{\sum_k w(i,k)}{\sum_\ell \sum_k w(\ell,k)} \quad \forall i$$

We can see that

$$\pi_i \cdot P_{ij} = \pi_j \cdot P_{ji} \quad \forall i,j$$

Hastings-Metropolis sampling algorithm

Design an irreducible Markov chain $\mathbb{X}$ such that the unique stationary distribution of $\mathbb{X}$ is the distribution of random variable Y.
Since the long-run proportion of state i is $P(Y = i)$,

$$\lim_{n\to\infty} \frac{X(1) + X(2) + \ldots + X(n)}{n} = \sum_{i\in S} i \cdot P(Y = i) = E[Y] = \mu$$

While computing μ by the law of large number is difficult (hard to sample on Y), we use this alternative method to compute μ by generating a sequence of $\mathbb{X}$, which is sometime easier.

Hastings-Metropolis sampling algorithm (cont.)

There's a random variable Y such that

$$P(Y = i) = \frac{b_i}{C}$$

for some unknown (or intractable) $C = \sum_{i \in S} b_i$.
We then design a Markov chain $\mathbb{X}$ that

- $P_{ii} = Q_{ii} + \sum_{k \in S, k \neq i} Q_{ik} \cdot (1 - q_{ik})$
- $P_{ij} = Q_{ij} \cdot q_{ij} \quad \forall j \neq i$

where

- Q is the transition matrix of an arbitrary irreducible Markov chain $\mathbb{X}$ which has the same state space as Y.
- q is a matrix to be determined later.

Hastings-Metropolis sampling algorithm (cont.)

For $n = 0, 1, \ldots$,

1. If $X(n) = i$, set Z such that $P(Z = j) = Q_{ij} \quad \forall j \in S$.
2. If $Z = j$, set $X(n+1)$ such that
   - $P(X(n+1) = j) = q_{ij}$
   - $P(X(n+1) = i) = 1 - q_{ij}$

One can see that this satisfies the requirement on previous page.

Hastings-Metropolis sampling algorithm (cont.)

Then, we let

$$
\begin{aligned}
q_{ij} &= \min\left(\frac{b_j \cdot Q_{ji}}{b_i \cdot Q_{ij}}, 1\right) \\
\Longrightarrow b_i \cdot Q_{ij} \cdot q_{ij} &= b_j \cdot Q_{ji} \cdot q_{ji} \\
\Longrightarrow \frac{b_i}{C} \cdot P_{ij} &= \frac{b_j}{C} \cdot P_{ji}
\end{aligned}
$$

By observation 3 on page 86, we know that $(b_1/C, b_2/C, \ldots)$ is the stationary distribution of $\mathbb{X}$.

Example: Space of permutations

Example

Let S consist of all the permutations $(x_1, x_2, \ldots, x_n)$ of $\{1, 2, \ldots, n\}$ that

$$\sum_{k=1}^{n} k \cdot x_k \geq \frac{n^3}{4}$$

- This is same as Y in page 98 with $C = |S|$ and $b_i = 1\ \forall i$.
- S is hard to compute.
- We need to design a matrix Q such that when given a permutation x, it's efficient to compute the value of $Q_{xy}\ \forall y \in S$.

Example: Space of permutations (cont.)

We let

$$Q_{xy} = \frac{1}{N(x)} \quad \text{, if } y \text{ can be obtained from } x \text{ by one swap}$$

where $N(x)$ is the number of permutations that can be obtained from x by one swap. For example:

$$\underbrace{(1,2,3,4,5)}_{y} \leftrightarrow \underbrace{(1,3,2,4,5)}_{x} \leftrightarrow \underbrace{(1,3,4,2,5)}_{y}$$

This chain is irreducible since each $x \in S$ can go to $(x_1, x_2, \ldots, x_n)$, where $x_1 \leq x_2 \leq \ldots \leq x_n$, by several swaps.
Also, given a x, finding all the obtainable y can be done efficiently.

Counting process

Definition

A collection $\mathbb{N}$ of random variables is a *counting process* if $N(t)$ denotes the total number of events that occur by time t.

- $N(t)$ is a nonnegative integer.
- The value of $N(t)$ is increasing as t increase.
- $N(t) - N(s)$ is the number of events that occur between time index s and t, where $t > s$.

Two properties

Independent increments:

Definition

A counting process is *independent increments* if the number of events in two non-overlapping time intervals are independent.

- For example, $N(s) - N(0)$ and $N(s+t) - N(s)$ are independent.

Stationary increments:

Definition

A counting process is *stationary increments* if the number of events in any time interval depends only on the length of the interval.

- For example, $P(N(s_1+t) - N(s_1) = k) = P(N(s_2+t) - N(s_2) = k)$.

Poisson process

Definition

A *Poisson process* with rate λ is a counting process with independent increments and stationary increments such that

$$P(N(s+t) - N(s) = n) = \frac{e^{-\lambda t} \cdot (\lambda t)^n}{n!}$$

holds for all nonnegative integers.

- $N(s+t) - N(s)$ is Poisson distributed with parameter λt.
- The average number of events that occur in an unit time interval $(t = 1)$ is λ (since the expectation value of Poisson distribution with parameter λ is λ.)

An operational definition

Theorem

Let $\mathbb{N}$ be a counting process with independent increments and stationary increments. Then $\mathbb{N}$ is a Poisson process if and only if the following two conditions hold:

1. $P(N(t) = 1) = \lambda \cdot t + o(t)$
2. $P(N(t) \geq 2) = o(t)$

- We say that $f(t) = o(t)$ if

$$\lim_{t \to 0} \frac{f(t)}{t} = 0$$

Proof

($\Longrightarrow$):
Since $N(t)$ is Poisson distributed with parameter λt,

$$
\begin{aligned}
P(N(t)=1) = \frac{(\lambda t)\cdot e^{-\lambda t}}{1!} &= \lambda t \cdot \left(1 - \frac{\lambda t}{1!} + \frac{(\lambda t)^2}{2!} - \cdots\right) \\
&= \lambda t - \lambda^2 t^2 + \cdots \\
&= \lambda t + o(t)
\end{aligned}
$$

$$
\begin{aligned}
P(N(t)=2) = \frac{(\lambda t)^2\cdot e^{-\lambda t}}{2!} &= \frac{(\lambda t)^2}{2!} \cdot \left(1 - \frac{\lambda t}{1!} + \frac{(\lambda t)^2}{2!} - \cdots\right) \\
&= o(t)
\end{aligned}
$$

One can prove that $P(N(t)=k)=o(t)$ for all $k \geq 2$,
hence $P(N(t)\geq 2)=o(t)$.

Proof (cont.)

($\Longleftarrow$):
The Laplace transform of a random variable X is

$$\phi(u) = E[e^{-u \cdot X}]$$

We say that two random variables have the same distribution if their Laplace transform are the same.
And if X is Poisson distributed with parameter λt, then

$$E[e^{-u \cdot X}] = e^{(e^{-u}-1) \cdot \lambda t}$$

Proof (cont.)

We define $\phi_u(t) = E[e^{-u \cdot N(t)}]$, then we know that

$$\begin{aligned}
\phi_u(s+t) &= E[e^{-u \cdot N(s+t)}] \\
&= E[e^{-u \cdot (N(s) - N(0))} e^{-u \cdot (N(s+t) - N(s))}] \\
&= E[e^{-u \cdot N(s)}] \cdot E[e^{-u \cdot (N(s+t) - N(s))}] \\
&= E[e^{-u \cdot N(s)}] \cdot E[e^{-u \cdot N(t)}] \\
&= \phi_u(s) \cdot \phi_u(t)
\end{aligned}$$

The 3rd equation is because two independent random variables X and Y will make

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

Proof (cont.)

By the two conditions in page 106, we know

$$P(N(t) = 0) = 1 - \lambda t + o(t)$$

Therefore,

$$\begin{aligned}
\phi_u(t) &= E[e^{-u \cdot N(t)}] \\
&= e^{-u \cdot 0} \cdot (1 - \lambda t + o(t)) + e^{-u \cdot 1} \cdot (\lambda t + o(t)) \\
&\qquad + (e^{-u \cdot 2} + e^{-u \cdot 3} + \cdots) \cdot o(t) \\
&= 1 - \lambda t + e^{-u} \cdot \lambda t + o(t) \\
&= 1 + (e^{-u} - 1) \cdot \lambda t + o(t)
\end{aligned}$$

And

$$\phi_u(s+t) = \phi_u(s) \cdot \phi_u(t) = \phi_u(s) \cdot (1 + (e^{-u} - 1) \cdot \lambda t + o(t))$$

Proof (cont.)

Differentiate on $\phi_u(s)$, we can get

$$\begin{aligned}\phi_u{}'(s) &= \lim_{t\to 0}\frac{\phi_u(s+t)-\phi_u(s)}{t} = \lim_{t\to 0}(\phi_u(s)\cdot(e^{-u}-1)\cdot\lambda + o(t))\\ &= \phi_u(s)\cdot(e^{-u}-1)\cdot\lambda\end{aligned}$$

By $\frac{\phi_u{}'(s)}{\phi_u(s)} = (e^{-u}-1)\cdot\lambda$, we have

$$\ln\phi_u(s) = \int (e^{-u}-1)\cdot\lambda\, ds = (e^{-u}-1)\cdot\lambda s + C$$

By $\phi_u(0) = 1$ and $\ln 1 = 0$, we know $C = 0$, hence

$$\phi_u(s) = e^{(e^{-u}-1)\cdot\lambda s} \quad \forall s, u$$

which means that $N(s)$ is Poisson distributed for all s.

Inter-arrival time

Definition

The kth inter-arrival time T_k of $\mathbb{N}$ is the time interval between the $(k+1)$st and kth events.
$\mathbb{T} = T_1, T_2, \ldots$ is the sequence of inter-arrival times of $\mathbb{N}$.

- 0th event arrives at time 0.

Observation 1: Independent & exponential distributed

Observation

If $\mathbb{N}$ is a Poisson process with rate λ, then each T_k is an independent exponential distribution with parameter λ.

Proof:
The cumulative distribution function of T_1 is

$$\begin{aligned} F_1(s) &= P(T_1 \le s) \\ &= 1 - P(T_1 > s) \\ &= 1 - P(N(s) = 0) \\ &= 1 - e^{-\lambda s} \end{aligned}$$

The 3rd equation is because $T_1 > s \iff N(s) = 0$.
We can observe that T_1 is exponential distributed.

Proof (cont.)

$$\begin{aligned} P(T_2 > t | T_1 = s) &= P(N(T_1 + t) - N(T_1) = 0 | T_1 = s) \\ &= P(N(T_1 + t) - N(T_1) = 0) \\ &= P(N(t) = 0) \\ &= e^{-\lambda t} \end{aligned}$$

The equations are derived by stationary increments.

Thus, T_2 is also exponential distributed with parameter λ.
And T_1, T_2 are independent.
One can prove for T_k with $k \geq 3$ by the same approach.

Observation 2: Waiting time is gamma distributed

Observation

The waiting time $S_k = T_1 + T_2 + \cdots + T_k$ of the kth event is gamma distributed with parameter (k, λ). check gamma distribution

- The probability density function is

$$f(t) = \lambda e^{-\lambda t} \cdot \frac{(\lambda t)^{k-1}}{(k-1)!}$$

- It's also called Erlang distribution since $k \in \mathbb{Z}^+$.

Verification

$$P(S_k \leq t) = P(N(t) \geq k) = \sum_{i \geq k} P(N(t) = i) = \sum_{i \geq k} \frac{(\lambda t)^i \cdot e^{-\lambda t}}{i!}$$

So,

$$\begin{aligned} \frac{dP(S_k \leq t)}{dt} &= \sum_{i \geq k} \frac{\lambda \cdot (\lambda t)^{i-1} \cdot e^{-\lambda t}}{(i-1)!} - \sum_{i \geq k} \frac{(\lambda t)^i \cdot e^{-\lambda t} \cdot \lambda}{i!} \\ &= \frac{\lambda \cdot (\lambda t)^{k-1} \cdot e^{-\lambda t}}{(k-1)!} \end{aligned}$$

Property 1 (Different types of events)

Property

Let $\mathbb{N}$ be a Poisson process with rate λ, each event is classified as type 1 with probability p or type 2 with probability $1-p$.
Then the arrival of type 1 and type 2 events are both Poisson processes with rate $p \cdot \lambda$ and $(1-p) \cdot \lambda$. And the two processes are independent.

- Let $\mathbb{N}_1$ be the process of type 1 event, $N_1(k)$ is the number of type 1 events that occur by time k. (same for $\mathbb{N}_2$)
- $\mathbb{N}_1$ and $\mathbb{N}_2$ are said to be independent if $N_1(s_1+t_1) - N_1(s_1)$ and $N_2(s_2+t_2) - N(s_2)$ are independent for all s_1, t_1, s_2, t_2.

Proof

Here we prove that $\mathbb{N}_1$ is a Poisson process with rate λp.

Stationary increments:

$$P(N_1(s+t) - N_1(s) = k_1 | N(s+t) - N(s) = k) = \binom{k}{k_1} \cdot p^{k_1} \cdot (1-p)^{k-k_1}$$

Therefore,

$$P(N_1(s+t) - N_1(s) = k_1) = \sum_{k \geq 0} \binom{k}{k_1} \cdot p^{k_1} \cdot (1-p)^{k-k_1} \cdot \frac{(\lambda t)^k \cdot e^{\lambda t}}{k!}$$

which has nothing to do with s.

Proof (cont.)

Independent increments:
Let $(s, s+t)$ and $(u, u+v)$ be two non-overlapping time intervals,

$$
\begin{aligned}
&P(N_1(s+t) - N_1(s) = k_1, N_1(u+v) - N_1(u) = \ell_1) \\
&= \sum_{k \geq 0} \sum_{\ell \geq 0} P(N_1(s+t) - N_1(s) = k_1, N_1(u+v) - N_1(u) = \ell_1 \\
&\qquad |N(s+t) - N(s) = k, N(u+v) - N(u) = \ell) \\
&\qquad \cdot P(N(s+t) - N(s) = k, N(u+v) - N(u) = \ell) \\
&= \sum_{k \geq 0} \sum_{\ell \geq 0} P(N_1(s+t) - N_1(s) = k_1 | N(s+t) - N(s) = k) \\
&\qquad \cdot P(N_1(u+v) - N_1(u) = \ell_1 | N(u+v) - N(u) = \ell) \\
&\qquad \cdot P(N(s+t) - N(s) = k) \cdot P(N(u+v) - N(u) = \ell) \\
&= P(N_1(s+t) - N_1(s) = k_1) \cdot P(N_1(u+v) - N_1(u) = \ell_1)
\end{aligned}
$$

Proof (cont.)

Conditions on page 106:

1. $P(N_1(t) \geq 2) \leq P(N(t) \geq 2) = o(t)$
2. $$\begin{aligned} P(N_1(t) = 1) &= P(N_1(t) = 1 | N(t) = 1) \cdot P(N(t) = 1) \\ &\quad + P(N_1(t) = 1 | N(t) \geq 2) \cdot P(N(t) \geq 2) \\ &= p \cdot (\lambda t + o(t)) + o(t) \\ &= p\lambda t + o(t) \end{aligned}$$

Hence we know that $\mathbb{N}_1$ is a Poisson process with rate λp.
Seems like we can also derive this from the result of next page and omit this page's proof? (By Example 3.23 on textbook?)

Proof (cont.)

$\mathbb{N}_1$ **and** $\mathbb{N}_2$ **are independent**:

$$
\begin{aligned}
&P(N_1(t) = i, N_2(t) = j) \\
&= P(N_1(t) = i, N_2(t) = j | N(t) = i + j) \cdot P(N(t) = i + j) \\
&= \binom{i+j}{i} \cdot p^i \cdot (1-p)^j \cdot \frac{e^{-\lambda t} \cdot (\lambda t)^{i+j}}{(i+j)!} \\
&= \frac{e^{-\lambda p t} \cdot (\lambda p t)^i}{i!} \cdot \frac{e^{-\lambda(1-p)t} \cdot (\lambda(1-p)t)^j}{j!} \\
&= P(N_1(t) = i) \cdot P(N_2(t) = j)
\end{aligned}
$$

We only prove for two intervals that have the same length here.
This also prove that $N_1(t)$ and $N_2(t)$ are Poisson distributed over t (Example 3.23 on textbook).

Example 1

$\mathbb{N}$ has rate 10 and $p = 1/12$,

$$P(N_1(4) = 0) = \frac{e^{-\frac{40}{12}} \cdot (\frac{40}{12})^0}{0!} = e^{-\frac{10}{3}}$$

Example 2: Type transitions

There're r classes of particles.

- $Y_i(k)$ is the number of class i particles at time k.
- The time is discrete in this case.
- $Y_i(0)$ is Poisson distributed with parameter λ_i.
- P_{ij} is the transition probability for a class i particle to class j.

We prove that $Y_j(n)$ is Poisson distributed with parameter $\sum_{i=1}^{r} P_{ij}^{n} \cdot \lambda_i$.

Proof

Take class i for example, we consider a Poisson process $\mathbb{N}$ with rate λ_i, where each event is classified as type k with probability P^n_{ik}.

For an arbitrary unit time interval, the number of events that occur in this interval is Poisson distributed with parameter λ_i. We take this Poisson distributed number as the value of $Y_i(0)$.

By property 1, we know that the number of type k events in this interval is Poisson distributed with parameter $P^n_{ik} \cdot \lambda_i$, which also means that the number of class i particles that eventually become class k at time n, which is denoted as C^n_{ik}, is also Poisson distributed with parameter $P^n_{ik} \cdot \lambda_i$.

$Y_j(n) = \sum_i C^n_{ij}$, which is Poisson distributed with parameter $\sum_{i=1}^r P^n_{ij} \cdot \lambda_i$. (since the Poisson parameter can be summed up.)

Example 3: Selling a product

Consider a Poisson process with rate λ, where each event is an offer that has density function $f(x)$.
A product is sold if an offer with value higher than the price y comes.
Assume the accepted offer comes at time t, then the storage cost is $c \cdot t$, where c is a constant decided by the product.

We want to know the expected profit, which is $E[f(x) - ct | f(x) \geq y]$.

Solution

The probability for each offer being accepted is

$$p(y) = P(X \geq y) = \int_y^\infty f(x)\,dx$$

The expectation of storage time t is $1/(\lambda \cdot p(y))$.
Hence,

$$\begin{aligned}
&E[f(x)|f(x) \geq y] - E[ct|f(x) \geq y] \\
&= \int_0^\infty x \cdot f_{X|X\geq y}(x)\,dx - \frac{c}{\lambda \cdot p(y)} \\
&= \int_y^\infty x \cdot \frac{f_X(x)}{P(X \geq y)}\,dx - \frac{c}{\lambda \cdot p(y)} \\
&= \frac{1}{p(y)}\left(\int_y^\infty x \cdot f(x)\,dx - \frac{c}{\lambda}\right)
\end{aligned}$$

There are r types of coupons, and p_i is the probability for a collected coupon being type i.

We want to know the expectation of N, where N is the number of collected coupons so that all r types of coupons are collected.

Solution: First attempt

Let N_i be the number of coupons collected to receive the first type i coupon. We know that

$$E[N] = E[\max(N_1, N_2, \ldots, N_r)]$$

And

$$P(N \leq n) = P(N_1 \leq n, N_2 \leq n, \cdots, N_r \leq n)$$

But since each N_i are not independent, we can't go even further from here. For example, given that $N_1 = 1$, $P(N_2 = 1 | N_1 = 1) = 0$.

Solution: Second attempt

Without loss of generality, we assume that the coupons arrive as a Poisson process $\mathbb{N}$ with rate 1.
$\mathbb{N}_i$ is the process of type i coupons, which has rate p_i.
X_i is the time that the first type i coupon appears, and

$$X = \max(X_1, X_2, \ldots, X_r)$$

$X_1, X_2, \ldots, X_r$ are independent since $\mathbb{N}_1, \mathbb{N}_2, \ldots, \mathbb{N}_r$ are independent.

Solution: Second attempt (cont.)

We can see that

$$\begin{aligned} P(X \leq t) &= P(X_1 \leq t, X_2 \leq t, \cdots, X_r \leq t) \\ &= \prod_{i=1}^{r} P(X_i \leq t) = \prod_{i=1}^{r} (1 - e^{-t \cdot p_i}) \end{aligned}$$

And

$$E[X] = \int_0^\infty P(X > t)\, dt = \int_0^\infty 1 - \prod_{i=1}^{r} (1 - e^{-t \cdot p_i})\, dt$$

The first equation is from the property of probability.
Surprisingly, $E[X] = E[N]$ as explained below.

Solution: Second attempt (cont.)

$$X = T_1 + T_2 + \cdots + T_N$$

where T_j is the jth inter-arrival time of $\mathbb{N}$.
Each T_j are i.i.d. and are exponential distributed with rate 1.
Also, N is independent with each T_j.

$$E[X|N = n] = E[T_1 + T_2 + \cdots + T_n] = n \cdot E[T_1] = n$$

$$E[X|N] = N \cdot E[T_1] = N$$

Hence,

$$E[X] = E[E[X|N]] = E[N]$$

Property 2-simple (Distribution of one event)

Property

Given that exactly one event of a Poisson process arrives in the interval $[0, t]$*, this arrival time is uniformly distributed over* $[0, t]$*.*

Proof:

$$\begin{aligned} P(T_1 \leq s | N(t) = 1) &= \frac{P(T_1 \leq s, N(t) = 1)}{P(N(t) = 1)} \\ &= \frac{P(N(s) = 1) \cdot P(N(t) - N(s) = 0)}{e^{-\lambda t} \cdot (\lambda t)^1 / 1!} \\ &= \frac{e^{-\lambda s} \cdot \lambda s \cdot e^{-\lambda(t-s)}}{e^{-\lambda t} \cdot \lambda t} = \frac{s}{t} \end{aligned}$$

Property 2-advanced (Distribution of several events)

Property

Given that exactly n events of a Poisson process arrive in the interval $[0, t]$, each with arrival time $X_1, X_2, \ldots, X_n$.
The order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ of these random variables have the joint density function

$$
\begin{aligned}
& f_{X_{(1)}, X_{(2)}, \ldots, X_{(n)}}(x_1, x_2, \ldots, x_n | N(t) = n) \\
&= \begin{cases} \frac{n!}{t^n} & \text{if } 0 < x_1 < x_2 < \cdots < x_n < t \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

- This implies that $X_1, X_2, \ldots, X_n$ are i.i.d. and each is uniformly distributed over $[0, t]$.

Proof

For any $0 < x_1 < x_2 < \ldots < x_n < t$,

$$
\begin{aligned}
& f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots, x_n | N(t) = n) \\
&= \frac{f_{X_{(1)},X_{(2)},\ldots,X_{(n)},N(t)}(x_1, x_2, \ldots, x_n, n)}{P(N(t) = n)} \\
&= \frac{f_{T_1,T_2,\ldots,T_n}(x_1, x_2 - x_1, \ldots, x_n - x_{n-1}) \cdot P(T_{n+1} > t - x_n)}{P(N(t) = n)} \\
&= \frac{\lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda(x_2 - x_1)} \ldots \lambda e^{-\lambda(x_n - x_{n-1})} \cdot e^{-\lambda(t - x_n)}}{\frac{e^{-\lambda t}(\lambda t)^n}{n!}} \\
&= \frac{n!}{t^n}
\end{aligned}
$$

T_i are the inter-arrival times, which are exponential distributed.

Corollary of property 2 (simple version)

Corollary

Consider a Poisson process with rate λ.
Each event is classified into a type i, where there are r types of event.
Suppose that $p_i(\cdot)$ is a sampling function over interval $[0, t]$ such that each arrived event at time x has probability $p_i(x)$ to be classified as type i.
Then the number $N_i(t)$ of type i events in $[0, t]$ is Poisson distributed with parameter

$$\lambda \int_0^t p_i(x)\, dx = \lambda \cdot t \cdot R_i \quad \text{,where } R_i = \frac{1}{t} \int_0^t p_i(x)\, dx$$

And each $N_i(t)$ for all i are independent.

- $N_i(\cdot)$ does not form a Poisson process here.
  It doesn't satisfy stationary distribution because of $p_i(\cdot)$.

Proof

Assume that $N(t) = n$.
Let $n = n_1 + n_2 + \cdots + n_r$, where $n_i \geq 0 \ \forall i = 1, \ldots, r$.
These n events arrive independent and uniformly at random over $[0, t]$.
If an event arrives at time $x \in [0, t]$, then with probability $p_i(x)$ it becomes type i. Therefore, each event is of type i with probability

$$\begin{aligned}
&\int_0^t P(\text{type } i | \text{arrives at time } x) \cdot \frac{1}{t} \, dx \\
&= \frac{1}{t} \int_0^t p_i(x) \, dx \\
&= R_i
\end{aligned}$$

which can be seen as the average of $p_i(\cdot)$ over $[0, t]$.

Proof (cont.)

$$
\begin{aligned}
P(\bigwedge_{1\le i\le r} N_i(t) = n_i) &= P(\bigwedge_{1\le i\le r} N_i(t) = n_i | N(t) = n) \cdot P(N(t) = n) \\
&= \left(\frac{n!}{n_1! n_2! \cdots n_r!} \cdot \prod_{1\le i\le r} R_i^{n_i} \right) \cdot \frac{e^{-\lambda t} \cdot (\lambda t)^n}{n!} \\
&= \prod_{1\le i\le r} \frac{e^{-\lambda t R_i} (\lambda t R_i)^{n_i}}{n_i!} \\
&= \prod_{1\le i\le r} P(N_i(t) = n_i)
\end{aligned}
$$

Hence we know that each $N_i(t)$ is Poisson distributed with parameter $\lambda t R_i$ and are independent. (Check Example 3.23 on textbook.)

Corollary of property 2 (full version)

Corollary

Consider a Poisson process with rate λ.
Each event is classified into a type i, where there are r types of event.
Suppose that $p_i(\cdot)$ is a sampling function over interval $[s, s+t]$ such that each arrived event at time x has probability $p_i(x)$ to be classified as type i.
Then the number $N_i(s+t) - N_i(s)$ of type i events in $[s, s+t]$ is Poisson distributed with parameter

$$\lambda \int_s^{s+t} p_i(x)\, dx$$

And each $N_i(s+t) - N_i(s)$ for all i are independent.

Proof

Regarding $[s, s+t]$ as $[0, t]$, the sampling function becomes

$$p'_i(x) = p_i(x+s)$$

From the simple version, we know that $N_i(s+t) - N_i(s)$ is Poisson distributed with parameter

$$\lambda \int_0^t p'_i(x)\, dx = \lambda \int_0^t p_i(x+s)\, dx = \lambda \int_s^{s+t} p_i(x)\, dx$$

Example 1: Infinite server queue

Suppose that jobs arrive at a Poisson rate λ, and we have infinite number of servers. The running time of each job is independent and distributed with function $T(\cdot)$. We want to know

1. The distribution of the number $X(t)$ of completed jobs by time t.
2. The distribution of the number $Y(t)$ of running jobs by time t.
3. The joint distribtuion of $Y(t_1)$ and $Y(t_2)$, where $t_1 < t_2$.

Solution of question 1 & 2

We classify the jobs into two types:

- **type 1**: completed by time t.
- **type 2**: not completed by time t.

If a job arrives at time x, then the sampling function is

- $p_1(x) = T(t-x)$
- $p_2(x) = 1 - T(t-x)$

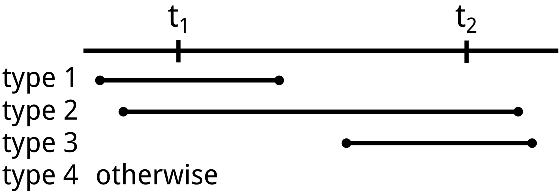
Thus, $X(t) = N_1(t)$ is Poisson distributed with parameter

$$\lambda \int_0^t T(t-x)\, dx$$

And $Y(t) = N_2(t)$ is Poisson distributed with parameter

$$\lambda \int_0^t (1 - T(t-x))\, dx = \lambda t - \lambda \int_0^t T(t-x)\, dx$$

Solution of question 3



$$p_1(x) = T(t_2 - x) - T(t_1 - x) \quad \text{if } 0 < x < t_1$$
$$p_2(x) = 1 - T(t_2 - x) \quad \text{if } 0 < x < t_1$$
$$p_3(x) = 1 - T(t_2 - x) \quad \text{if } t_1 < x < t_2$$

Solution of question 3 (cont.)

We know that $Y(t_1) = N_1(t_2) + N_2(t_2)$ and $Y(t_2) = N_2(t_2) + N_3(t_2)$, hence $Y(t_1)$ and $Y(t_2)$ are not independent. Therefore, we use the following method:

$$
\begin{aligned}
& P(Y(t_1) = m_1, Y(t_2) = m_2) \\
&= \sum_{n_2=0}^{\infty} P(N_1(t_2) = m_1 - n_2, N_2(t_2) = n_2, N_3(t_2) = m_2 - n_2) \\
&= \sum_{n_2=0}^{\infty} P(N_1(t_2) = m_1 - n_2) \cdot P(N_2(t_2) = n_2) \cdot P(N_3(t_2) = m_2 - n_2) \\
&= \cdots
\end{aligned}
$$

The ∞ in $\sum$ can be replaced by $\min(m_1, m_2)$.

Example 2: Encounters on a highway

Cars enter a distance-d highway in a Poisson rate λ.
The fixed speed of each car is i.i.d. with function $F_S(\cdot)$.
Suppose our car enters the highway and moves at a fixed speed s, what's the distribution of the number of encountering with other cars?

Solution

Suppose we enter at time t_1 and leave at $t_2 = t_1 + d/s$.
Each car choose a fixed speed S according to F_S, its travel time $T = d/S$.
The distribution function of T is

$$F_T(t) = P(T \leq t) = P(S \geq \frac{d}{t}) = 1 - F_S(\frac{d}{t})$$

We classify the cars into three types:

- **type a** (overtaken by us): $0 < t < t_1$, $t + T > t_2$.
- **type b** (overtake us): $t_1 < t < t_2$, $t + T < t_2$.
- **type c**: otherwise

Solution (cont.)

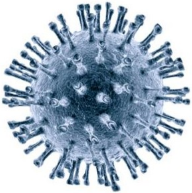
$$p_a(t) = P(T > t_2 - t) = 1 - F_T(t_2 - t) \quad \text{if } 0 < t < t_1$$
$$p_b(t) = P(T < t_2 - t) = F_T(t_2 - t) \quad \text{if } t_1 < t < t_2$$

Since the Poisson parameters can be summed up, $N_a(t_2) + N_b(t_2)$ is Poisson distributed with parameter

$$\lambda \int_0^{t_1} (1 - F_T(t_2 - t))\, dt + \lambda \int_{t_1}^{t_2} F_T(t_2 - t)\, dt$$

which is the distribution we want.

Example 3: HIV infection



People are infected with an unknown Poisson rate λ.
The incubation time for each infected person has distribution function F.
At time t, we know the number n_1 of people who already have the AIDS symptoms (finished incubations).
We want to estimate the value of λ, and the number n_2 of the incubating people at time t.

Solution

We classify the infected people into two types:

- **type 1**: have symptoms appear by time t.
- **type 2**: still in incubation by time t.

Then

$$E[N_1(t)] = \lambda \int_0^t F(t-x)\,dx = \lambda \int_0^t F(x)\,dx$$

$$E[N_2(t)] = \lambda \int_0^t (1 - F(t-x))\,dx = \lambda \int_0^t (1 - F(x))\,dx$$

$$n_1 \approx E[N_1(t)] \implies \hat{\lambda} = \frac{n_1}{\int_0^t F(x)\,dx}$$

$$n_2 \approx \hat{\lambda} \int_0^t (1 - F(x))\,dx = \frac{n_1 \int_0^t (1 - F(x))\,dx}{\int_0^t F(x)\,dx}$$

Example 4: Hidden bugs

Suppose the errors come as a Poisson process.
Each error belongs to one of the m bugs in the program.
For all $i = 1, \ldots, m$, $\mathbb{N}_i$ denotes the Poisson process of errors caused by bug i, which has an unknown rate λ_i.
At time t, we know the value of $M_j(t)$, which is the number of bugs causing exactly j errors.
And a bug is still *hidden* if it hasn't cause any error.
We want to know the expected error rate of hidden bugs, which is the expectation of the summation of all hidden bugs' λ_i.

Solution

Let

$$H_i(t) = \begin{cases} 1 & \text{if the } i\text{th bug is still hidden by time } t \\ 0 & \text{otherwise} \end{cases}$$

$$\Lambda(t) = \sum_{i=1}^{m} H_i(t) \cdot \lambda_i$$

Then

$$P(H_i(t) = 1) = P(N_i(t) = 0) = \frac{(\lambda_i t)^0 \cdot e^{-\lambda_i t}}{0!} = e^{-\lambda_i t}$$

$$E[\Lambda(t)] = \sum_{i=1}^{m} \lambda_i \cdot E[H_i(t)] = \sum_{i=1}^{m} \lambda_i \cdot P(H_i(t) = 1) = \sum_{i=1}^{m} \lambda_i \cdot e^{-\lambda_i t}$$

$$= \frac{1}{t} \sum_{i=1}^{m} \lambda_i \cdot t \cdot e^{-\lambda_i t} = \frac{1}{t} \sum_{i=1}^{m} P(N_i(t) = 1) = \frac{1}{t} E[M_1(t)]$$

Non-homogeneous Poisson process

Definition

$\mathbb{N}$ is a non-homogeneous Poisson process with intensity function $\lambda(\cdot)$ if $\mathbb{N}$ is a counting process such that the following four conditions hold:

1. $N(0) = 0$
2. $\mathbb{N}$ satisfies independent increments.
3. $P(N(t+h) - N(t) \geq 2) = o(h)$
4. $P(N(t+h) - N(t) = 1) = \lambda(t) \cdot h + o(h)$

- It's a Poisson process without stationary increments.
- If $\lambda(t) = \lambda$, it becomes the homogeneous (normal) Poisson process.

Proposition 1

Part 1:
For a Poisson process $\mathbb{N}$ with rate λ, suppose that each arrived event has sampling function $p_i(t)$, then the counting process $\mathbb{N}_i$ describing the arrival of type-i events is a non-homogeneous Poisson process with intensity function

$$\lambda_i(t) = \lambda \cdot p_i(t)$$

Part 2:
All non-homogeneous Poisson process with bounded $\lambda(\cdot)$ can be obtained in the above way.
Proof of part 2: Since there's a λ that $\lambda(t) \leq \lambda$ for all t, we just let $p_i(t) = \lambda(t)/\lambda$ for a Poisson process with rate λ.

Proof of part 1

Condition 1: trivial.
Condition 3:

$$P(N_i(t+h) - N_i(t) \geq 2) \leq P(N(t+h) - N(t) \geq 2) = o(h)$$

Condition 4:

$$\begin{aligned}
&P(N_i(t+h) - N_i(t) = 1) \\
=& P(N_i(t+h) - N_i(t) = 1 | N(t+h) - N(t) = 1) \\
&\quad \times P(N(t+h) - N(t) = 1) \\
&\quad + P(N_i(t+h) - N_i(t) = 1 | N(t+h) - N(t) \geq 2) \\
&\quad \times P(N(t+h) - N(t) \geq 2) \\
=& (p_i(t) + o(h)) \cdot (\lambda h + o(h)) + o(h) \\
=& p_i(t) \cdot \lambda h + o(h)
\end{aligned}$$

Proof of part 1 (cont.)

The yellow part holds because

$$\lim_{h \to 0} P(N_i(t+h) - N_i(t) = 1 | N(t+h) - N(t) = 1) = p_i(t)$$
$$\Longrightarrow P(N_i(t+h) - N_i(t) = 1 | N(t+h) - N(t) = 1) = p_i(t) + o(h)$$

Condition 2:
Consider non-overlapped intervals $[s, s+t]$ and $[u, u+v]$, the distribution in each interval is decided by Poisson parameter $\lambda \int_s^{s+t} p_i(x)\, dx$ and $\lambda \int_u^{u+v} p_i(x)\, dx$, which doesn't influence each other.

Distribution of non-homogeneous Poisson process

According to part 2 of proposition 1, for a non-homogeneous Poisson process with intensity function $\lambda(\cdot)$, we can observe that its distribution in $[s, s+t]$ is Poisson distributed with parameter

$$\lambda \int_s^{s+t} p_i(x)\, dx = \int_s^{s+t} \lambda(x)\, dx$$

Example 1: Poisson tea shop

Suppose that customers come to tea shop as a non-homogeneous Poisson process with $\lambda(t)$ as follow:

$$\lambda(t) = \begin{cases} 0 & 0 \le t \le 8 \\ 5 + 5 \cdot (t - 8) & 8 < t \le 11 \\ 20 & 11 < t < 13 \\ 20 - 2 \cdot (t - 13) & 13 < t \le 17 \\ 0 & 17 < t < 24 \end{cases}$$

We want to know the probability that no one comes in $[8.5, 9.5]$.

Solution

The distribution in $[8.5, 9.5]$ is a Poisson with parameter

$$\int_{8.5}^{9.5} (5 + 5 \cdot (t - 8))\, dt = \int_{0.5}^{1.5} (5 + 5t)\, dt = 10$$

Hence the probability is e^{-10}.

Example 2: Infinite server queue again

Suppose that jobs arrive at a Poisson rate λ, and we have infinite number of servers. The running time of each job is independent and distributed with function $F(\cdot)$.

We want to prove that the process of jobs departing (finishing) is a non-homogeneous Poisson process.

Proof

Condition 1: trivial.

Condition 2:
Consider two non-overlapping intervals $[s, s+t]$ and $[u, u+v]$, where $u > s$. Let the jobs finishing in $[s, s+t]$ be type 1, and those finishing in $[u, u+v]$ be type 2.
According to corollary of property 2, we know that the distribution of type 1 and type 2 jobs are independent in $[0, u+v]$, which means that the number of jobs finishing in $[s, s+t]$ and $[u, u+v]$ are independent.

Proof (cont.)

Condition 3 & 4: Let $f(\cdot)$ be the density function of $F(\cdot)$. The number of jobs departing in $[t, t+h]$ is Poisson distributed with parameter

$$\begin{aligned}
&\lambda \int_0^{t+h} (F(t+h-x) - F(t-x))\, dx \\
=&\lambda \int_0^{t+h} (f(t+h-x) \cdot h + o(h))\, dx \\
=&\left(\lambda h \int_0^{t+h} f(x)\, dx\right) + o(h) \\
=&\lambda h \cdot F(t+h) + o(h) \\
=&\lambda h \cdot F(t) + o(h)
\end{aligned}$$

Proof (cont.)

The first equation is because

$$\lim_{h\to 0}\frac{F(t+h-x)-F(t-x)}{h}=f(t-x)=\lim_{h\to 0}f(t-x+h)$$

Then we know that

$$\begin{aligned}
&P(N(t+h)-N(t)=1)\\
=&(\lambda h\cdot F(t)+o(h))\cdot e^{-\lambda h\cdot F(t)+o(h)}\\
=&(\lambda h\cdot F(t)+o(h))\cdot(1-\lambda h\cdot F(t)+o(h))\\
=&\lambda h\cdot F(t)+o(h)\\
\text{and}\qquad &P(N(t+h)-N(t)\geq 2)\\
=&\sum_{k\geq 2}\frac{(\lambda h\cdot F(t)+o(h))^k\cdot e^{-\lambda h\cdot F(t)+o(h)}}{k!}\\
=&o(h)
\end{aligned}$$

Proposition 2

If $\mathbb{N}_1$ and $\mathbb{N}_2$ are independent non-homogeneous Poisson process with intensity function $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$,
then $\mathbb{N} = \mathbb{N}_1 + \mathbb{N}_2$ is also a non-homogeneous Poisson process with intensity function $\lambda(t) = \lambda_1(t) + \lambda_2(t)$.

Proof is left as exercise.

Continuous Markov chain

Definition

The collection $\mathbb{X} = \{X(t) | t \geq 0\}$ of nonnegative integral random variables is a *continuous-time Markov chain* if for all function $x(\cdot)$, nonnegative real numbers s and t, and integer i and j,

$$\begin{aligned}
&P(X(s+t) = j | X(s) = i, X(r) = x(r) \quad \forall r \in [0, s)) \\
=&P(X(s+t) = j | X(s) = i)
\end{aligned}$$

holds.

- The condition above is called *Markovian property*.
- Non-homogeneous Poisson process is a continuous Markov chain, where we take the values of $N(t)$ as the states.

Homogeneous transition probabilities

Definition

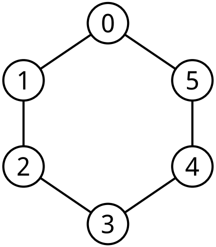
We say a continuous Markov chain has homogeneous transition probabilities if

$$P(X(s+t) = j | X(s) = i)$$

is independent of s.

- Our discussion of continuous Markov chain assumes this property.

State graph



The state graph of continuous Markov chain can be taken as the state graph of Markov chain without self-links. The process will stay on a state for a certain amount of time, and go into a possible next state.

Observation 1

Observation

The time for $\mathbb{X}$ to stay in state i is exponentially distributed.

- Suppose that $X(r) = i$ for some time $r \geq 0$. Let T_i be the waiting time for $\mathbb{X}$ to transit to a state other than i. We want to prove that T_i is memoryless, that is

$$P(T_i \geq s + t | T_i \geq s) = P(T_i \geq t)$$

  hence T_i is exponential distributed.
- This also means that for a Poisson process, no matter when the last event happened before, start measuring from now, the waiting time for the next event is still exponential distributed with parameter λ.

Proof

$$
\begin{aligned}
& P(T_i \geq s+t | T_i \geq s) \\
=& P(X(v) = i, \forall r+s \leq v \leq r+s+t | X(u) = i, \forall r \leq u \leq r+s) \\
=& P(X(v) = i, \forall r+s \leq v \leq r+s+t | X(r+s) = i) \\
=& P(X(v) = i, \forall r \leq v \leq r+t | X(r) = i) \\
=& P(T_i \geq t)
\end{aligned}
$$

$=$ holds because of homogeneous transition probability.

$=$ holds because of Markovian property (more details at next page.)

Proof (cont.)

$\boxed{=}$ holds because we can use something like

$$\begin{aligned}
&P(A, B, C, \ldots, Y|Z) \\
=&P(B, C, \ldots, Y|A, Z) \cdot P(A|Z) \\
=&P(C, \ldots, Y|A, B, Z) \cdot P(B|A, Z) \cdot P(A|Z) \\
=&\cdots
\end{aligned}$$

and rearrange the condition parts?

Observation 2

Observation

Given that $X(s) = i$, the probability that the next state other than i is j is a value $\mathbb{P}[i,j]$ that is independent of s.

Proof: For any state j other than i, let T_i be the waiting time for the first transition from state i to a state other than i. We have

$$\mathbb{P}[i,j] = P(X(T_i + s) = j | X(s) = i) = P(X(T_i) = j | X(0) = i)$$

where the last equation is by homogeneous transition probability.

Interpretation

- Starting with $X(0) = i$, the time staying in i is exponential distributed with parameter λ_i, then $\mathbb{X}$ transits to a different state j with probability $\mathbb{P}[i,j]$.
- $\mathbb{X}$ can be characterized by

$$\lambda_i \quad \forall i \in S \qquad \text{and} \qquad \mathbb{P}[i,j] \quad \forall i,j \in S$$

Example: Hair cut

- State 0: no customer
- State 1: cut hair
- State 2: wash hair

Assume customers appear as a Poisson process with rate λ'. When a customer appear, if the service is in state 1 or 2, then the customer just leave.

- Staying time in state 1 is exponential distributed with parameter $\lambda_1{}'$.
- Staying time in state 2 is exponential distributed with parameter $\lambda_2{}'$.

Example: Hair cut (cont.)

Characterized as follow:

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

$$\lambda_0 = \lambda', \qquad \lambda_1 = {\lambda_1}', \qquad \lambda_2 = {\lambda_2}'$$

Birth and death process

Definition

A *birth and death process* with birth rate $\{\beta_n\}, \forall n \geq 0$ and death rate $\{\delta_n\}, \forall n \geq 0$ is a Markov chain that when the state is n,

- The waiting time B_n for it to go to state $n+1$ is exponential distributed with parameter β_n.
- The waiting time D_n for it to go to state $n-1$ is exponential distributed with parameter δ_n.

Birth and death process: characterize

- $\lambda_0 = \beta_0$
- $\lambda_n = \beta_n + \delta_n \quad \forall n > 0$

The last equation can be seen as a Poisson process with rate λ_n, where there're two types of events, one is birth and one is death. By page 117, we get the above result and

- $\mathbb{P}[n, n+1] = \frac{\beta_n}{\beta_n + \delta_n} \quad \forall n > 0$
- $\mathbb{P}[n, n-1] = \frac{\delta_n}{\beta_n + \delta_n} \quad \forall n > 0$

Example: M/M/s Servers

Given s servers, where

- Jobs arrive in a Poisson process with rate λ.
- Processing time for each job is exponential distributed with parameter μ.

The number of jobs currently waiting or processing is a B & D process with

- $\beta_n = \lambda$
- $\delta_n = \begin{cases} \mu \cdot n & \forall n \leq s \\ \mu \cdot s & \forall n > s \end{cases}$

Example: Linear growth with immigration

- $\beta_n = \beta \cdot n + \theta$
- $\delta_n = \delta \cdot n$

We want to know the value of $E[X(t)]$.

Solution:

$$P(X(t+h) = n+1 | X(t) = n) = (\beta n + \theta) \cdot h + o(h) \quad \forall n \geq 0$$
$$P(X(t+h) = n-1 | X(t) = n) = \delta n \cdot h + o(h) \quad \forall n \geq 1$$
$$P(X(t+h) = n | X(t) = n) = 1 - (\beta n + \delta n + \theta) \cdot h + o(h) \quad \forall n \geq 0$$

Since we make h close to 0, we ignore other probability that are all $o(h)$.

Solution (cont.)

$$
\begin{aligned}
E[X(t+h)|X(t)=n] &= (n+1)\cdot((\beta n+\theta)\cdot h+o(h)) \\
&\quad + (n-1)\cdot(\delta n\cdot h+o(h)) \\
&\quad + n\cdot(1-(\beta n+\delta n+\theta)\cdot h+o(h)) \\
&\quad + o(h) \\
&\vdots \\
&= n+((\beta-\delta)n+\theta)\cdot h+o(h)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
E[X(t+h)] &= E[E[X(t+h)|X(t)]] \\
&= (1+(\beta-\delta)\cdot h)\cdot E[X(t)]+\delta h+o(h)
\end{aligned}
$$

Solution (cont.)

$$\frac{E[X(t+h)] - E[X(t)]}{h} = (\beta - \delta) \cdot E[X(t)] + \theta + \frac{o(h)}{h}$$
$$\frac{d}{dt}E[X(t)] = (\beta - \delta) \cdot E[X(t)] + \theta$$

If $\beta = \theta$,

$$\frac{dE[X(t)]}{dt} = \theta$$
$$\implies E[X(t)] = X(0) + \theta t$$

Solution (cont.)

If $\beta \neq \theta$, define $g(t) = (\beta - \delta) \cdot E[X(t)] + \theta$,

$$g'(t) = (\beta - \delta) \cdot \frac{d}{dt} E[X(t)] = (\beta - \delta) \cdot g(t)$$

$$\vdots$$

$$\Longrightarrow E[X(t)] = X(0) \cdot e^{(\beta-\delta)t} + \frac{\theta}{\beta - \delta} \cdot (e^{(\beta-\delta)t} - 1)$$

Note that if $\delta > \beta$,

$$\lim_{t \to \infty} E[t] = \frac{\theta}{\delta - \beta}$$

which is dominated by the immigration.

Increment time

Definition

Given a birth and death process, if $X(0) = i$, the increment time is $T_i = \min\{t | X(t) = i + 1\}$.

- We want to know the value of $E[T_i]$.

Expectation of increment time

- If $i = 0$, $E[T_0] = \dfrac{1}{\beta_0}$
- If $i > 0$,

$$\begin{aligned}
E[T_i] = {} & \text{waiting time for the first event} \\
& + 0 \cdot P(\text{the first event is birth}) \\
& + (E[T_{i-1}] + E[T_i]) \cdot P(\text{the first event is death}) \\
= {} & \frac{1}{\beta_i + \delta_i} + 0 + (E[T_{i-1}] + E[T_i]) \cdot \frac{\delta_i}{\beta_i + \delta_i} \\
\Longrightarrow E[T_i] = {} & \frac{1}{\beta_i} \cdot (\delta_i \cdot E[T_{i-1}] + 1)
\end{aligned}$$

Transition probability function

Definition

The *transition probability function* $P_{ij}(t) = P(X(t) = j | X(0) = i)$.

- Note that this is different from the $\mathbb{P}[i,j]$ mentioned before.
- There's also value for $P_{ii}(t)$.

Transition probability function of pure birth process

Theorem

If $\mathbb{X}$ is a pure birth process, where $\delta_n = 0 \quad \forall n \geq 0$, and $\beta_i \neq \beta_j \quad \forall i \neq j$, then

$$P_{ij}(t) = \begin{cases} e^{-\beta_i t} & \text{if } i = j \\ F(i,j) - F(i,j-1) & \text{if } i < j \end{cases}$$

where

$$F(i,j) = P(X(t) \leq j | X(0) = i) = \sum_{i \leq k \leq j} e^{-\beta_k t} \prod_{i \leq r \leq j,\ r \neq k} \frac{\beta_r}{\beta_r - \beta_k}$$

Proof

When $i < j$,

$$\begin{aligned} F(i,j) &= P(X(t) < j+1 | X(0) = i) \\ &= P(T_i + T_{i+1} + \cdots + T_j > t) \\ &\boxed{=} \sum_{i \le k \le j} e^{-\beta_k t} \prod_{i \le r \le j,\ r \ne k} \frac{\beta_r}{\beta_r - \beta_k} \end{aligned}$$

$\boxed{=}$ is from equation 5.6 on textbook.

$$P(T_i + T_{i+1} + \cdots + T_j > t) = 1 - P(T_i + T_{i+1} + \cdots + T_j \le t)$$

where $P(T_i + T_{i+1} + \cdots + T_j \le t)$ is the distribution of sum of $j - i + 1$ exponential distributions where each one must has different rate.

Yule process

Definition

Yule process is a pure birth process with linear rate, where $\delta_n = 0\ \forall n \geq 0$ and $\beta_n = \beta \cdot n\ \forall n \geq 0$.

According to page 183,

$$P_{ij}(t) = \cdots = \binom{j-1}{i-1} (e^{-\beta t})^i \cdot (1 - e^{-\beta t})^{j-i} \quad \forall 1 \leq i \leq j$$

which is a negative binomial distribution with parameter $(i, e^{-\beta t})$.

Instantaneous transition rate

Definition

Given a continuous Markov chain with each state i having the exponential parameter λ_i, the *instantaneous transition rate* is

$$q_{ij} = \lambda_i \cdot \mathbb{P}[i,j]$$

we let $q_{ii} = 0$ for the ease of later explanation.

- Given all the q_{ij}, one can compute λ_i by $\sum_{j \in S} q_{ij}$, and then compute all the value of $\mathbb{P}[i,j]$. Hence q_{ij} also characterize the continuous Markov chain.

Komolgorov forward equation

Theorem

For most of the continuous Markov chain (ex. birth and death process), the following holds:

$$P_{ij}'(t) = \sum_{k \in S} P_{ik}(t) \cdot q_{kj} - \lambda_j \cdot P_{ij}(t)$$

Komolgorov backward equation

Theorem

For all the continuous Markov chain, the following holds:

$$P_{ij}{}'(t) = \sum_{k \in S} q_{ik} \cdot P_{kj}(t) - \lambda_i \cdot P_{ij}(t)$$

- We only prove this equation (at page 194).

Lemma 1

Lemma

1. $\lim_{h \to 0} \frac{1 - P_{ii}(h)}{h} = \lambda_i$
2. $\lim_{h \to 0} \frac{P_{ij}(h)}{h} = q_{ij} \quad \textit{if } i \neq j$

We prove this lemma by proving the equivalence that

1. $1 - P_{ii}(h) = \lambda_i \cdot h + o(h)$
2. $P_{ij}(h) = q_{ij} \cdot h + o(h)$

Proof

Equation 1: In the interval of h,

$$\begin{aligned}
P_{ii}(h) &= P(\text{no transition happen}) \\
&\quad + P(\text{more than one transitions and come back to state } i) \\
\Longrightarrow\ 1 - P_{ii}(h) &= P(\text{one transition happens}) \\
&\quad + P(\text{more than one transitions and doesn't come back}) \\
&= \lambda_i \cdot h + o(h)
\end{aligned}$$

$=$ is according to operational definition on page 106.

Proof (cont.)

Equation 2: In the interval of h,

$$\begin{aligned}
P_{ij}(h) &= P(\text{one } i \to j \text{ transition happens}) \\
&\quad + P(\text{more than one transitions and the final state is } j) \\
&= (\lambda_i \cdot h + o(h)) \cdot \mathbb{P}[i,j] + o(h) \\
&= \lambda_i \cdot \mathbb{P}[i,j] \cdot h + o(h) \\
&= q_{ij} \cdot h + o(h)
\end{aligned}$$

Lemma 2

Lemma

$$P_{ij}(t+s) = \sum_{k \in S} P_{ik}(t) \cdot P_{kj}(s) \quad \forall s \geq 0, t \geq 0$$

Proof

$$
\begin{aligned}
P_{ij}(t+s) &= P(X(t+s)=j|X(0)=i) \\
&= \sum_{k\in S} P(X(t+s)=j, X(t)=k|X(0)=i) \\
&= \sum_{k\in S} P(X(t+s)=j|X(t)=k, X(0)=i)\cdot P(X(t)=k|X(0)=i) \\
&= \sum_{k\in S} P(X(t+s)=j|X(t)=k)\cdot P(X(t)=k|X(0)=i) \\
&= \sum_{k\in S} P(X(s)=j|X(0)=k)\cdot P(X(t)=k|X(0)=i) \\
&= \sum_{k\in S} P_{ik}(t)\cdot P_{kj}(s)
\end{aligned}
$$

= is by Markovian property.

= is by homogeneous transition probability.

Proof of backward equation

$$
\begin{aligned}
P_{ij}(h+t) - P_{ij}(t) &\colorbox{yellow}{$=$} \left(\sum_{k\geq 0} P_{ik}(h) \cdot P_{kj}(t) \right) - P_{ij}(t) \\
&= \left(\sum_{k\geq 0, k\neq i} P_{ik}(h) \cdot P_{kj}(t) \right) - (1 - P_{ii}(h)) \cdot P_{ij}(t) \\
\implies P_{ij}{}'(t) &= \lim_{h\to 0} \left(\sum_{k\geq 0, k\neq i} \frac{P_{ik}(h)}{h} \cdot P_{kj}(t) \right) - \frac{1 - P_{ii}(h)}{h} \cdot P_{ij}(t) \\
&\colorbox{cyan}{$=$} \sum_{k\geq 0} q_{ik} \cdot P_{kj}(t) - \lambda_i \cdot P_{ij}(t)
\end{aligned}
$$

$\colorbox{yellow}{$=$}$ is by lemma 2.

$\colorbox{cyan}{$=$}$ is by lemma 1 and that lim can be swap (photo P1070502).

Example: Pure birth process

$$\lambda_i = \beta_i \text{ and } \mathbb{P}[i, i+1] = 1$$

Hence

$$P_{ij}{}'(t) = \beta_i \cdot P_{i+1,j}(t) - \beta_i \cdot P_{ij}(t)$$

Example: Birth and death process

If $i = 0$,

$$P_{0j}{}'(t) = \beta_0 \cdot P_{1j}(t) - \beta_0 \cdot P_{0j}(t)$$

If $i > 0$,

$$\begin{aligned} P_{ij}{}'(t) = (\beta_i + \delta_i) \cdot \frac{\beta_i}{\beta_i + \delta_i} \cdot P_{i+1,j}(t) \\ + (\beta_i + \delta_i) \cdot \frac{\delta_i}{\beta_i + \delta_i} \cdot P_{i-1,j}(t) \\ - (\beta_i + \delta_i) \cdot P_{ij}(t) \end{aligned}$$

Example: A simple case

There're only two states 0 and 1, where 0 has rate β and 1 has rate δ.

$$\left.\begin{array}{l} {P_{00}}'(t) = \beta(P_{10}(t) - P_{00}(t)) \\ {P_{10}}'(t) = \delta(P_{00}(t) - P_{10}(t)) \end{array}\right\} \delta \cdot {P_{00}}'(t) + \beta \cdot {P_{10}}'(t) = 0$$

$$\Longrightarrow \delta \cdot P_{00}(t) + \beta \cdot P_{10}(t) = C$$

$$\Longrightarrow \text{since } P_{00}(0) = 1 \text{ and } P_{10}(0) = 0, \text{ we know } C = \delta$$

$$\Longrightarrow \beta \cdot P_{10}(t) = \delta(1 - P_{00}(t))$$

$$\Longrightarrow {P_{00}}'(t) = \delta - (\delta + \beta) \cdot P_{00}(t)$$

$$\Longrightarrow P_{00}(t) = \frac{\beta}{\beta + \delta} e^{-(\beta+\delta)t} + \frac{\delta}{\beta + \delta} \text{ and}$$

$$P_{10}(t) = \cdots$$

Limiting probability

Definition

If $\lim_{t \to \infty} P_{ij}(t)$ exists and is independent of i, then

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

is the limiting probability of state j.

Some facts of the limiting probability

If $\mathbb{X}$ is irreducible and positive recurrent, then

1. $\pi = (\pi_1, \pi_2, \dots)$ exists.
2. $\sum_{i \in S} \pi_i = 1$
3. π is the unique stationary distribution of $\mathbb{X}$.
4. π_j is the long-run proportion of state j.

Balance equation

Consider on the birth and death process and finite states continuous Markov chain. We can prove that above two processes have limiting probability, which means that

$$\lim_{t\to\infty} P_{ij}{}'(t) = 0$$

then from the forward equation, we know

$$\begin{aligned}\lim_{t\to\infty} P_{ij}{}'(t) &= \lim_{t\to\infty} \sum_{k\geq 0} P_{ik}(t) q_{kj} - \lambda_j P_{ij}(t) \\ &= \sum_{k\geq 0} \pi_k \cdot q_{kj} - \lambda_j \cdot \pi_j = 0\end{aligned}$$

The $\lim$ at $=$ can be swap inside.

Balance equation (cont.)

Hence, we know that

$$\lambda_j \pi_j = \sum_{k \geq 0} q_{kj} \pi_k \quad \forall j \geq 0$$

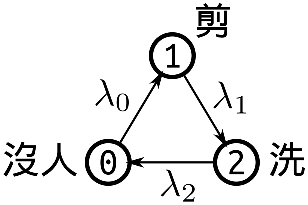
which is called the *balance equation*.

If we take each transition in the process as an event, the events of "leaving from state j" forms a Poisson process with rate $\lambda_j \pi_j$, and the events of "entering state j" forms a Poisson process with rate $\sum_{k \geq 0} q_{kj} \pi_k$.

Interpretation of balance equation

- Take the process as infinite transitions (include the type as $i \to i$).
- Now consider the transitions that can form a Poisson process with rate λ_j.
- One can observe that only π_j of these transitions can match as a transition leaving from state j.
- Hence we know that this type of transitions is also a Poisson process with rate $\lambda_j \pi_j$.
- if still not clear, prove it formally :p

Example: Hair cut



According to balance equation:

$$\lambda_0 \pi_0 = \lambda_2 \pi_2$$
$$\lambda_1 \pi_1 = \lambda_0 \pi_0$$
$$\lambda_2 \pi_2 = \lambda_1 \pi_1$$

And by $\pi_0 + \pi_1 + \pi_2 = 1$, we can represent π_0, π_1, π_2 by λ_0, λ_1, and λ_2.

Example: Birth and death process

According to balance equation:

$$
\begin{aligned}
\beta_0 \pi_0 &= \delta_1 \pi_1 \\
(\beta_1 + \delta_1)\pi_1 &= \beta_0 \pi_0 + \delta_2 \pi_2 \\
&\vdots \\
(\beta_n + \delta_n)\pi_n &= \beta_{n-1}\pi_{n-1} + \delta_{n+1}\pi_{n+1}
\end{aligned}
$$

which implies

$$
\begin{aligned}
\beta_0 \pi_0 &= \delta_1 \pi_1 \\
\beta_1 \pi_1 &= \delta_2 \pi_2 \\
&\vdots \\
\beta_n \pi_n &= \delta_{n+1}\pi_{n+1}
\end{aligned}
$$

Example: Birth and death process (cont.)

$$\pi_1 = \frac{\beta_0}{\delta_1}\pi_0$$

$$\pi_2 = \frac{\beta_1}{\delta_2}\pi_1 = \frac{\beta_1\beta_0}{\delta_2\delta_1}\pi_0$$

$$\vdots$$

$$\pi_n = \frac{\beta_0\beta_1\cdots\beta_{n-1}}{\delta_1\delta_2\cdots\delta_n}\pi_0$$

By $\sum_{i\geq 0}\pi_i = 1$, we have

$$\pi_0 = \frac{1}{1+\sum_{k\geq 1}\frac{\beta_0\beta_1\cdots\beta_{k-1}}{\delta_1\delta_2\cdots\delta_k}} \text{ and } \pi_n = \frac{\beta_0\beta_1\cdots\beta_{n-1}}{\delta_1\delta_2\cdots\delta_n(1+\sum_{k\geq 1}\frac{\beta_0\beta_1\cdots\beta_{k-1}}{\delta_1\delta_2\cdots\delta_k})}$$

Example: Birth and death process (cont.)

As a matter of fact,

$$\sum_{k\geq 1} \frac{\beta_0\beta_1\cdots\beta_{k-1}}{\delta_1\delta_2\cdots\delta_k} < \infty \iff \sum_{i\geq 0} \pi_i = 1$$

Ohterwise, each π_i would have value 0.

Example: M/M/s servers

$$\beta_n = \beta, \qquad \delta_n = \begin{cases} n \cdot \delta & \text{if } n < s \\ s \cdot \delta & \text{if } n \geq s \end{cases}$$

To make $\sum_{i \geq 0} \pi_n = 1$,

$$\sum_{k \geq 1} \frac{\beta_0 \beta_1 \cdots \beta_{k-1}}{\delta_1 \delta_2 \cdots \delta_k} < \infty \iff \sum_{k \geq s} \frac{\beta^k}{(s \cdot \delta)^k} < \infty$$

$$\iff \frac{\beta}{s \cdot \delta} < 1 \iff \frac{\beta}{\delta} < s$$

Example: Server crash

There are s servers. Each server has parameter β of crashing. Only one person is fixing the servers, with parameter δ. We want to know:

1. The time proportion that no server is in crash.
2. Expectation number of crashing servers.

Solution:
Let state n represents that there are n crashed servers.

$$\delta_n = \begin{cases} \delta & \forall 1 \leq n \leq s \\ 0 & \text{otherwise} \end{cases} \qquad \beta_n = \begin{cases} (s-n)\beta & \forall 0 \leq n \leq s \\ 0 & \text{otherwise} \end{cases}$$

Solution to question 1

$$
\begin{aligned}
\pi_0 &= \frac{1}{1 + \sum_{k=1}^{s} \frac{\beta_0 \beta_1 \cdots \beta_{k-1}}{\delta_1 \delta_2 \cdots \delta_k}} \\
&= \frac{1}{1 + \sum_{k=1}^{s} \frac{s\beta \cdot (s-1)\beta \cdots (s-k+1)\beta}{\delta^k}} \\
&= \frac{1}{1 + \sum_{k=1}^{s} \left(\frac{\beta}{\delta}\right)^k \frac{s!}{(s-k)!}}
\end{aligned}
$$

Solution to question 2

From previous page, we know

$$\pi_n = \frac{\left(\frac{\beta}{\delta}\right)^n \frac{s!}{(s-n)!}}{1 + \sum_{k=1}^{s} \left(\frac{\beta}{\delta}\right)^k \frac{s!}{(s-k)!}}$$

Hence, the expectation is

$$\sum_{n=0}^{s} n \cdot \pi_n$$

Example: Server crash advanced

Given n servers. The working time parameter for the ith server is s_i. And the fixing time parameter is r_i. There's only one TA, who will fix the most recently crashed server in the first priority.

We use $(i_1, i_2, \ldots, i_k)$ to represent the state that k servers crash, i_j is the jth recently crashed server. And use ϕ as the state that all the servers are working. Hence, we have $\sum_{k=0}^{n} \binom{n}{k} k!$ different states.

Example: Server crash advanced (cont.)

According to balance equation:

$$\sum_{i=1}^{n} s_i \cdot \pi(\phi) = \sum_{i=1}^{n} r_i \cdot \pi(i)$$

$$\left(r_{i_1} + \sum_{i \notin \{i_1, i_2, \ldots, i_k\}} s_i\right) \cdot \pi(i_1, i_2, \ldots, i_k)$$
$$= s_{i_1} \cdot \pi(i_2, \ldots, i_k) + \sum_{i \notin \{i_1, i_2, \ldots, i_k\}} r_i \cdot \pi(i, i_1, i_2, \ldots, i_k)$$

Example: Server crash advanced (cont.)

$$\pi(\phi) = \frac{1}{1 + \sum_{(i_1, i_2, \ldots, i_k)} \frac{s_{i_1} s_{i_2} \cdots s_{i_k}}{r_{i_1} r_{i_2} \cdots r_{i_k}}}$$

$$\pi(i_1, i_2, \ldots, i_k) = \frac{s_{i_1} s_{i_2} \cdots s_{i_k}}{r_{i_1} r_{i_2} \cdots r_{i_k}} \cdot \pi(\phi)$$

Time reversibility

- Assume $\mathbb{X}$ is irreducible and positive recurrent. We say $\mathbb{X} \to \mathbb{X}^*$ is a discretization that changes each inter-arrival time of $\mathbb{X}$ to 1.
- $\mathbb{X}^*$ can now be seen as a discrete Markov chain, which still satisfies irreducible and positive recurrent.
- Assume $\mathbb{X}^*$ is aperiodic, we can take the matrix $\mathbb{P}$ of $\mathbb{X}$ as the transition matrix of $\mathbb{X}^*$ in the discrete Markov chain.
- We know that $\mathbb{X}^*$ has the limiting probability vector π^*, and

$$\pi^* \times \mathbb{P} = \pi^*, \quad \sum_{i \in S} \pi_i^* = 1$$

Time reversibility (cont.)

According to balance equation,

$$\lambda_j \cdot \pi_j = \sum_{k \in S} \lambda_k \cdot \pi_k \cdot \mathbb{P}[k, j]$$

If we let $\hat{\pi} = (\lambda_1 \pi_1, \lambda_2 \pi_2, \ldots)$, we know that $\hat{\pi} \times \mathbb{P} = \hat{\pi}$ and $\frac{\hat{\pi}}{\sum_{i \in S} \hat{\pi}_i}$ is the unique vector that is π^*. Hence,

$$\pi_i^* = \frac{\pi_i \cdot \lambda_i}{\sum_{j \in S} \pi_j \cdot \lambda_j} \text{ and } \pi_i = \frac{\pi_i^* / \lambda_i}{\sum_{j \in S} \pi_j^* / \lambda_j}$$

Time reversibility (cont.)

According to the time reversibility of discrete Markov chain, we know

- The reversed chain $\mathbb{Y}^*$ of $\mathbb{X}^*$, which is also a Markov chain, has transition matrix $\mathbb{Q}$ such that

$$\pi_i^* \cdot \mathbb{Q}[i,j] = \pi_j^* \cdot \mathbb{P}[j,i] \quad \forall i,j \in S$$

- If

$$\pi_i^* \cdot \mathbb{P}[i,j] = \pi_j^* \cdot \mathbb{P}[j,i] \quad \forall i,j \in S$$

  then $\mathbb{X}^*$ is time-reversible.
- If we can find any vector $\hat{\pi}$ such that

$$\hat{\pi}_i \cdot \mathbb{P}[i,j] = \hat{\pi}_j \cdot \mathbb{P}[j,i] \quad \forall i,j \in S \text{ and } \sum_{i \in S} \hat{\pi}_i = 1$$

  then $\hat{\pi} = \pi^*$.

Time reversibility (cont.)

- If $\mathbb{X}$ is in steady state, which means that $P(X(t) = i) = \pi_i \; \forall i \in S$, we can prove that the reversed chain $\mathbb{Y}$ of $\mathbb{X}$ is also a continuous Markov chain.
- Since the discretization of $\mathbb{Y}$ is $\mathbb{Y}^*$, $\mathbb{Q}$ would also be the matrix for $\mathbb{Y}$.
- To characterize $\mathbb{Y}$, we use $\mathbb{Q}[i,j] = \pi_j^* \cdot \mathbb{P}[j,i] / \pi_i^*$ and each λ_i are the same as $\mathbb{X}$.

Time reversibility (cont.)

If $\mathbb{X}$ is time reversible, then $\mathbb{Q} = \mathbb{P}$, which means that $\mathbb{X}^*$ is also time reversible. And we know that

$$\begin{aligned} & \pi_i^* \cdot \mathbb{P}[i,j] = \pi_j^* \cdot \mathbb{P}[j,i] \\ \Longleftrightarrow\ & \lambda_i \cdot \pi_i \cdot \mathbb{P}[i,j] = \lambda_j \cdot \pi_j \cdot \mathbb{P}[j,i] \\ \Longleftrightarrow\ & \pi_i \cdot q_{ij} = \pi_j \cdot q_{ji} \end{aligned}$$

- LHS is the Poisson rate of $i \to j$ transitions.
- RHS is the Poisson rate of $j \to i$ transitions.

Example: Birth and death process

Since $\pi_i \cdot q_{i,(i+1)}$ is the Poisson rate (i.e. λ) of $i \to i+1$ transitions,

$$E[\text{number of } i \to i+1 \text{ in } [s, s+t]] = \pi_i \cdot q_{i,(i+1)} \cdot t$$

Hence

$$\begin{aligned} \pi_i \cdot q_{i,(i+1)} &= \frac{E[\text{number of } i \to i+1 \text{ in } [s, s+t]]}{t} \\ &= \frac{E[\text{number of } i+1 \to i \text{ in } [s, s+t]]}{t} = \pi_{i+1} \cdot q_{(i+1),i} \end{aligned}$$

- Then we know that B & D process is time reversible.
- $=$ is because one must go back from $i+1$ to i to go from i to $i+1$ again.

Example: M/M/s servers

M/M/s is a birth and death process, it's time-reversible when in steady state ($\frac{\beta}{\delta} < s$). We want to prove that the overall process of job departing is a Poisson process with rate β.

Proof: Since it's time reversible, the rate of $i \to i-1$ is equal to the rate of $i-1 \to i$, which is β.

Example: M/M/1 server

In the M/M/1 system, if a job J stays for a time interval t, what is the distribution of the number of jobs already in the system when J enter this system?

Solution: Assume J arrives at s and depart at $s+t$, we know that all the jobs queuing in system at time s will depart in $[s, s+t]$, and the distribution of job departing in $[s, s+t]$ is Poisson with parameter $\beta \cdot t$ from the previous example.

Theorem on the other direction

Theorem

If there exists a nonnegative vector π *with*

$$\sum_{i \in S} \pi_i = 1 \text{ and } \pi_i \cdot q_{ij} = \pi_j \cdot q_{ji} \quad \forall i, j \in S$$

then $\mathbb{X}$ *is a time-reversible chain and* π *is the limiting probability of* $\mathbb{X}$.

Example: Server crash advanced

Given n servers. The working time parameter for the ith server is β_i. And the fixing time parameter is δ_i/k if there are k servers currently crashed. We want to know if this process is time reversible?

We use a set $D \subseteq \{1, 2, \ldots, n\}$ to represent the state of crashed servers. There are 2^n different states. If $i \notin D$ and $|D| = k - 1$ with $k \geq 1$, then

$$q_{D, D \cup \{i\}} = \beta_i \text{ and } q_{D \cup \{i\}, D} = \frac{\delta_i}{k}$$

Example: Server crash advanced (cont.)

We prove the time reversibility by guessing a π to satisfy

$$\pi_D \cdot \beta_i = \pi_{D \cup \{i\}} \cdot \frac{\delta_i}{k}$$

that is

$$\pi_\phi = \frac{1}{1 + \sum_{\phi \neq D \subseteq \{1,2,\ldots,n\}} |D|! \cdot \prod_{i \in D} \frac{\beta_i}{\delta_i}}$$
$$\pi_D = (|D|! \cdot \prod_{i \in D} \frac{\beta_i}{\delta_i}) \cdot \pi_\phi$$

Truncated Markov chain

Definition

Given a set A of states of $\mathbb{X}$, $\mathbb{X}^A$ is the *truncated chain* of $\mathbb{X}$ w.r.t. A, where the matrix of q_{ij} is induced by $i, j \in A$.

- We can see that if $\mathbb{X}$ is time reversible, so is $\mathbb{X}^A$. Since we can find a π^A that $\pi_i^A = \pi_i / \sum_{k \in A} \pi_k$.

Uniformization

Definition

Uniformization is a transformation that makes $\mathbb{X}$ into $\mathbb{Y}_\lambda$, where $\lambda_i = \lambda \quad \forall i \in S$.

- All the transitions of $\mathbb{Y}_\lambda$ forms a Poisson process with rate λ.

Theorem: Transition matrix of uniformized chain

Theorem

If $\mathbb{Q}$ is the transition matrix of $\mathbb{Y}_\lambda$, then

$$Q_{ij}(t) = \sum_{n=0}^{\infty} \mathbb{Q}^n[i,j] \cdot e^{-\lambda t} \cdot \frac{(\lambda t)^n}{n!}$$

- $Q_{ij}(t)$ is the transition probability function.
- Note that $\mathbb{Q}^n[i,j] \neq (\mathbb{Q}[i,j])^n$.

Proof

Let $N(t)$ represents the number of transitions in $[0, t]$.

$$\begin{aligned}
\text{LHS} &= \sum_{n=0}^{\infty} P(Y_\lambda(t) = j | Y_\lambda(0) = i, N(t) = n) \cdot P(N(t) = n | Y_\lambda(0) = i) \\
&= \sum_{n=0}^{\infty} \mathbb{Q}^n[i,j] \cdot \frac{(\lambda t)^n e^{-\lambda t}}{n!} \\
&= \text{RHS}
\end{aligned}$$

Application of uniformization

- Let $\mathbb{X}$ be a continuous Markov chain with $\lambda_i \leq \lambda \quad \forall i$.
- We can think as the transition rate for all states i are λ, but when the transition happens, it has probability λ_i/λ that it will really transit to other state, and with probability $1-(\lambda_i/\lambda)$ it will keep in state i.
- We then have $\mathbb{Y}_\lambda$ with

$$\mathbb{Q}[i,j] = \begin{cases} 1-\frac{\lambda_i}{\lambda} & i=j \\ \frac{\lambda_i}{\lambda}\cdot \mathbb{P}[i,j] & i\neq j \end{cases}$$

Example

For an example with only two states 0 and 1, where $\lambda_0 = \beta$ and $\lambda_1 = \delta$, we do the uniformization with $\lambda = \beta + \delta$. Then

$$\mathbb{Q} = \begin{pmatrix} \frac{\delta}{\beta+\delta} & \frac{\beta}{\beta+\delta} \\ \frac{\delta}{\beta+\delta} & \frac{\beta}{\beta+\delta} \end{pmatrix}$$

We can observe that $\mathbb{Q}^n = \mathbb{Q} \quad \forall n \geq 1$. And we can obtain the value of each $Q_{ij}(t)$.