stat333 Notes

Thaqib M

August 9, 2022

Linear Algebra

Matrix multiplication

If A is a $n \times m$ matrix and B is a $m \times k$ matrix then the matrix AB of dim $n \times k$ is defined by:

$$[AB]_{xy} = \sum_{\text{all}(z)} A_{xz} B_{zy}$$

Inner Product

The inner product (dot product) of 2 vectors $\vec{a}, \vec{b}$ in $\mathbb{R}^n$ is defined as

$$\vec{\mathbf{a}} \cdot \vec{\mathbf{b}} = \langle \vec{\mathbf{a}}, \vec{\mathbf{b}} \rangle = \sum_{k=1}^{n} a_k b_k$$

Eigenvalues and Eigenvectors

We can find eigenvalues by solving for the roots of the characteristic polynomial of the matrix $\mathbf{A}$.

$$\det(\mathbf{A} - tI_n) = 0$$

Where I_n is the $n \times n$ identity matrix. Then for each eigenvalue $t = c$ we can solve the system of linear equations

$$(\mathbf{A} - cI_n)\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$$

$\vec{\boldsymbol{x}}$ will be an eigenvector of $\mathbf{A}$.

Assignment Theorems

$$\mathbb{E}[Xf(Y)] = \mathbb{E}[f(Y) \cdot \mathbb{E}(X \mid Y)] \tag{hw1q6}$$

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k\mathbb{P}(X = k) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k) \tag{hw2q5}$$

If P is a **tridiagonal matrix** then the Markov chain satisfies the detail balance condition.

$$\pi_i P_{i,i+1} = \pi_{i+1} P_{i+1,i} \tag{hw2q7}$$

$$\{X_n\} \xrightarrow{a.s} X \Rightarrow \{X_n\} \xrightarrow{p} X \tag{hw3q5}$$

Stat

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] \tag{tower rule}$$

Walds Identity:
If $N \geq 0$ is a random variable and $X_i \sim X_1$ **i.i.d** then

$$\mathbb{E}\left[\sum_{n=1}^{N} X_n\right] = \mathbb{E}[N]\mathbb{E}[X_1] \tag{Walds identity}$$

Limit of Infinite sets Suppose $A = \bigcap_{i=1}^{\infty} A_i$ and $A_1 \supseteq A_2 \supseteq \cdots$ then

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$$

1 Week 1

1.1 Definition (Stochastic Process). Let $(X_t)_{t\in T}$ be a collection of random variables this is called a Stochastic Process. T is the *index set*.

1.2 Example (Simple Random Walk on $\mathbb{Z}$). Let $X_i \sim$ iid where $X_i \in \{-1,1\}$ with

$$P(X_i = 1) = \frac{1}{2}$$
$$P(X_i = -1) = \frac{1}{2}$$

now let

$$S_n = \sum_{i=0}^{n} X_i$$

Then $(S_i)_{k=0}^{\infty}$ is a stochastic process.

1.3 Definition (Transition Probability). Given $(X_s)_{s\leq t}$ we need the probability for X_{t+1}.

$$P(X_{(t+1)} = x_{t+1} | X_1 = x_1, X_2 = x_2, \ldots X_t = x_t)$$

1.4 Note. Conditional Probability Properties

$$P(A|B) = \frac{P(AB)}{P(B)}\ P(B) > 0$$
$$P(ABC) = P(A|BC) \cdot P(B|C) \cdot P(C)$$

1.5 Example. Transition Probabilities for SRW on $\mathbb{Z}^d$

$$P\left(\|X_{t+1} - X_t\| \mid (X_s)_{s\leq t}\right) = \frac{1}{2d}$$

1.1 Markov Chains

1.6 Definition (Markov Property)**.** A process has the Markov property if:

$$P(X_{t+1} = x_{t+1} \mid (X_s)_{s \le t}) = P(X_{t+1} = x_{t+1} \mid X_t = x_t)$$

(Next outcome only depends on the previous outcome)

1.7 Note (Markov Chain)**.** A stochastic process that satisfies the Markov property is called a Markov chain.

1.8 Definition (Time Homogeneous Markov Chain)**.** A Markov Chain is called time homogeneous if the following is true

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

1.9 Definition (Stochastic Matrix)**.** A matrix $\mathbf{P}$ is called stochastic if

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} & \cdots \\ p_{10} & p_{11} & \cdots \\ \vdots & \ddots & \end{pmatrix}$$

$$0 \le p_{ij} \le 1$$

$$\sum_{all(j)} p_{i_0 j} = 1 \text{ for fixed } i_0$$

1.10 Definition (Transition Matrix)**.** Let $\mathbf{P}$ be a Stochastic matrix and let p_{ij} = value in $i-$th row and $j-$th column. We define p_{ij} as

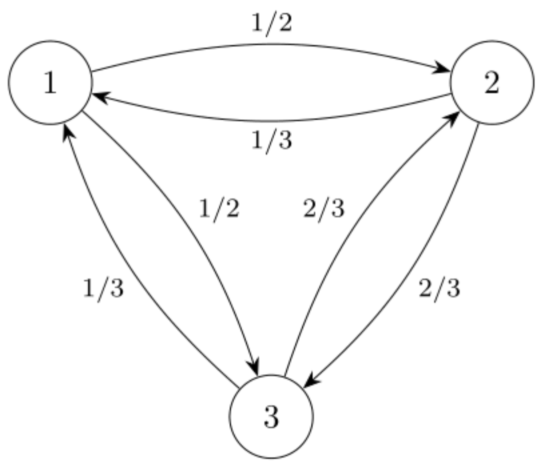
$$p_{ij} = P(X_t = j \mid X_{t-1} = i)$$

(probability of going from state i to state j in the chain).
This is called the transition matrix for $(X_t)_{t \in T}$.

1.11 Example. Transition Matrix Consider this transition matrix The transition matrix for this Markov Chain is

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

this can be visualized as:



1.1.1 Multistep Transition Probabilities

1.12 Definition.

$$[P(n, n+m)]_{xy} = P(X_{n+m} = y \mid X_n = x)$$

1.13 Theorem. Multistep Transition Probability Matrix Let $(X_t)_{t \in T}$ be a stochastic process satisfying the Markov property and be *time homogeneous* and let $\mathbf{P}$ be the transition matrix.

$$[P(n, n+m)]_{xy} = \mathbf{P}^m_{xy}$$

1.14 Lemma.

$$[P(n, m+1+n)]_{xy} = \sum_{\text{all}(z)} [P(n, m+n)]_{xz} P_{zy}$$

Proof. To go from state $x \to y$ we must add up all probabilities of going to an intermediate state $\mathbf{z}$, $x \to \mathbf{z} \to y$ we add possibilities of $\mathbf{z}$.

$[P(n, m+1+n)]_{xy} = P(X_{m+1+n} = y \mid X_n = x)$

$= \sum_{\text{all}(z)} P(X_{m+1+n} = y, X_{n+m} = z \mid X_n = x)$ Marginal probability function (stat240)

$= \sum_{\text{all}(z)} P(X_{m+1+n} = y \mid X_{n+m} = z, X_n = x) P(X_{n+m} = z \mid X_n = x)$ conditional probability

Since X_t satisfies the Markov property we get

$$= \sum_{\text{all}(z)} P(X_{m+1+n} = y \mid X_{n+m} = z) P(X_{n+m} = z \mid X_n = x)$$

By definition we have $P(X_{m+1+n} = y \mid X_{n+m} = z) = P_z y$ and $P(X_{n+m} = z \mid X_n = x) = [P(n, n+m)]_{xz}$. $\square$

Using Lemma 1.14 we can prove the Theorem 1.13.
Since 1.14's result is the definition of matrix multiplication we get

$$[P(n, m+1+n)]_{xy} = [P(n, m+n)P]_{xy}$$

by induction on m with base case $P(n, n+1) = P$ we get

$$[P(n, m+1+n)]_{xy} = \mathbf{P}^m$$

Since RHS does not depend on n we can write $P(n, n+m) = P(m)$ and time homogeneity applies for any m number of steps.

$$P(X_{n+m} = y \mid X_n = x) = P(X_m = y \mid X_0 = x)$$

2 Week 2

2.1 Initial Data

Let $(X_n)_{n\in I}$ be a time homogeneous Markov chain. We denote these by $0, 1, 2, \ldots |I| - 1$. We represent the state space as:

$$\{i_1, i_2, \ldots, i_{|I|}\} = \mathcal{X}$$

Let $\mathbf{P}$ be the transition matrix for this Markov chain.

2.1 Definition (Distribution Row Vector)**.**

$$\mu_j = P(X_0 = i_j)$$

Then the row vector $\vec{\mu}$ of $\dim = 1 \times |I|$ is defined as

$$\vec{\mu} = \left[\mu_1, \mu_2, \ldots, \mu_{|I|}\right]$$

$\vec{\mu}$ is called the distribution of X_0 denoted by $X_0 \sim m\vec{u}$.
The distribution vector for X_n is denoted by $\mu(n)$.

2.2 Theorem. Distribution of X_n The distribution row vector of X_n for a time homogeneous Markov chain is given by μP^n

Proof. Sketch.

$$P(X_n = i_k) = \sum_{j=1}^{|I|} P(X_n = i_k \mid X_0 = i_j)P(X_0 = i_j) = \sum \vec{\mu_j} P_{jk} = [\vec{\mu}P]_k$$

Implies $X_n \sim \vec{\mu}P^n$ □

2.2 Conditional Expectation

Given $f : \mathcal{X} \to \mathbb{R}$ what is the expected value of $f(X_m)$ given an initial distribution? The function f on a finite state space $\mathcal{X}$ is equivalent to a vector $\vec{f} \in \mathbb{R}^{|\mathcal{X}|}$

$$\vec{f} = \begin{pmatrix} f(1) \\ f(2) \\ \dots \\ f(n) \end{pmatrix}$$

The conditional expectation for $f(X_m)$ given $X_0 \sim \vec{\mu}$ is denoted by

$$\mathbb{E}(f(X_m) \mid X_0 \sim \vec{\mu})$$

By definition of conditional expectation we get

$$\begin{aligned} \mathbb{E}(f(X_m) \mid X_0 \sim \vec{\mu}) &= \sum_{k=1}^{|\mathcal{X}|} f(i_k) P(X_m = i_k \mid X_0 \sim \vec{\mu}) \\ &= \sum_{\text{all}(k)} f(i_k) \left[\vec{\boldsymbol{\mu}} \mathbf{P}^m\right]_k \\ &= \sum_{\text{all}(k)} \vec{\boldsymbol{f}}_k \left[\vec{\boldsymbol{\mu}} \mathbf{P}^m\right]_k \\ &= \langle \vec{\boldsymbol{\mu}} \mathbf{P}^m, \vec{\boldsymbol{f}} \rangle \end{aligned}$$

2.3 Stationary Distribution

Suppose $X_0 \sim \vec{\mu}$ then the distribution for $X_n \sim \vec{\mu}(n)$ then what is the limit of $\vec{\mu}(n)$ as $n \to \infty$. Suppose the limit $\lim_{n\to\infty} \vec{\mu}P^n = \vec{\pi}$ exists then we can write

$$\vec{\boldsymbol{\pi}} = \lim_{n\to\infty} \vec{\boldsymbol{\mu}}\mathbf{P}^n = \lim_{n\to\infty} \vec{\boldsymbol{\mu}}\mathbf{P}^{n-1}P = \lim_{n\to\infty} \vec{\boldsymbol{\mu}}(n-1)\mathbf{P} = \vec{\boldsymbol{\pi}}\mathbf{P}$$

So $\vec{\pi}$ is an `left eigenvector of` $\mathbf{P}$ with `eigenvalue 1`.

2.3 Definition (Stationary Distribution)**.** A probability vector $\vec{\boldsymbol{\pi}}$ is the Stationary Distribution for the stochastic matrix $\mathbf{P}$ if

$$\sum_k \vec{\boldsymbol{\pi}}_k = 1$$
$$\vec{\boldsymbol{\pi}}\mathbf{P} = \vec{\boldsymbol{\pi}}$$

2.4 Definition (Stationary Measure)**.** A measure $\vec{\nu}$ on $\mathcal{X}$ $\left(\vec{\nu} \in \mathbb{R}^{|\mathcal{X}|}\right)$ if

$$\vec{\nu}_i \geq 0$$
$$\sum \vec{\nu}_i > 0$$
$$\vec{\nu}\mathbf{P} = \vec{\nu}$$

2.5 Proposition (Stationary Distribution from Measure)**.** If $|X| < \infty$ and $\vec{\nu}$ is a stationary measure on $\mathbf{P}$

$$\vec{\boldsymbol{\pi}} = \frac{1}{\sum_i \vec{\nu}_i}\vec{\nu}$$

Then $\vec{\boldsymbol{\pi}}$ is a stationary distribution by definition.

2.6 Definition (Bi-stochastic Matrix)**.** A stochastic matrix is Bi-stochastic if

$$\sum_{\text{all}(i)} P_{ij_0} = 1 \qquad\qquad \text{for fixed } j_0$$

Sum of all rows = 1 and sum of all columns = 1.

2.7 Proposition (Stationary Distribution for Bi-stochastic Matrices)**.** If $\mathbf{P}$ is a **Bi-stochastic** transition matrix for Markov chain with finite state space $\mathcal{X}$ with $|\mathcal{X}| = N$ then the stationary distribution is given by

$$\vec{\boldsymbol{\pi}} = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \dots & \frac{1}{N} \end{pmatrix}$$

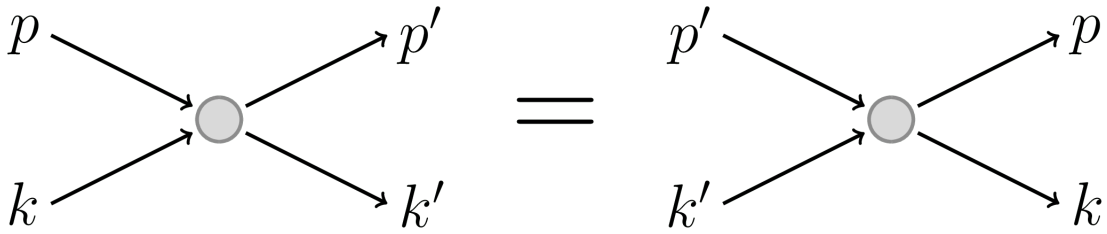
2.4 Detail Balance Condition

2.8 Definition (Detail Balance Condition)**.** $\vec{\pi}$ has the detail balance condition if:

$$\vec{\pi}_x \mathbf{P}_{xy} = \vec{\pi}_y \mathbf{P}_{yx}$$

2.9 Note. Detail balance condition means $\mathbb{P}(X_1 = x, X_0 = y) = \mathbb{P}(X_1 = y, X_0 = x)$.

2.10 Theorem (Detail Balance and Stationary Distribution)**.** If $X_0 \sim \vec{\pi}$ and $\vec{\pi}$ satisfies the detail balance condition then $X_n \sim \vec{\pi}$ for all $n \geq 1$



3 Week 3

3.1 Communicating States

3.1 Definition (communicating states)**.** A state x communicates with y if $\exists n \geq 1$ such that

$$[\mathbf{P}^n]_{xy} > 0$$

denoted by $x \to y$.

3.2 Note. $\mathbb{P}(A \mid X_{n-1} = x) = \mathbb{P}_x(A)$ and $\mathbb{E}(\cdot \mid X_n = x) = \mathbb{E}_x(\cdot)$

3.3 Definition (Time of the first return / first hitting time)**.**

$$\tau_x = \min\{n \mid X_n = x\}$$

$$\rho_{xy} = \mathbb{P}_x(\tau_y < \infty)$$

$\rho_{xy} = \mathbb{P}(X_n$ returns to y given it starts at $x)$.

3.4 Note.

$$1 - \rho_{xy} = \mathbb{P}_x(\tau_y = \infty)$$

3.5 Lemma (Communicating states and return probability)**.** $x \to y \iff \rho_{xy} > 0$.

3.6 Lemma (Transitivity)**.** $x \to y$ and $y \to z \Rightarrow x \to z$

3.7 Definition (Time of $k-th$ return)**.**

$$\tau_x^k = \min\{n > \tau_x^{k-1} \mid X_n = x\}$$

where $\tau_x^1 = \tau_x$.

3.2 Recurrent and Transient States

3.8 Definition (Recurrent and Transient States)**.** A state $x \in \mathcal{X}$ is called **recurrent** if

$$\rho_{xx} = 1$$

and **transient** if

$$\rho_{xx} < 1$$

3.9 Theorem (Escaping path)**.** If $x \to y$ and $\rho_{yx} < 1$ then x is transient.

3.10 Theorem (Corollary of Escaping Path theorem)**.** If $x \to y$ and x is recurrent then $\rho_{yx} = 1$.

3.3 Strong Markov Property

3.11 Definition (Stopping Time)**.** T is a stopping time if the occurrence (or non occurrence) of the event $\{T = n\}$ can be determined by $\{X_0, \ldots, X_n\}$.

3.12 Theorem (Strong Markov Property)**.** Suppose T is a stopping time. Given $T = n$ and $X_T = y$ the random variables $\{X_{T+k}\}_{k=0}^{\infty}$ behave like a Markov chain starting from initial state y. That is

$$\mathbb{P}(X_{T+1} = z \mid X_T = y, T = n) = \mathbb{P}(X_1 = z \mid X_0 = y) = \mathbf{P}_{yz}$$

3.13 Lemma ($k-$th return time and the strong Markov property)**.** Let τ_y^k be the $k-$th return time to y. Then the strong Markov property implies

$$\mathbb{P}_x(\tau_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1} \textbf{ or } \mathbb{P}_y(\tau_y^k < \infty) = \rho_{yy}^k \qquad \forall k \geq 1$$

3.14 Note. From the above **lemma** if we have $\rho_{yy} = 1$ (y is recurrent) then the chain returns to y for infinitely many k and it continually recurs in the Markov chain.
Otherwise if $\rho_{yy} < 1$ (y is transient) then $\rho_{yy}^k \to 0$ as $k \to \infty$ so after sometime y is never visited in the chain.

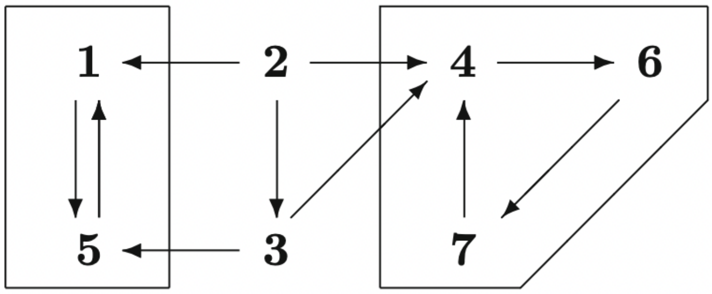
4 Week 4

4.1 Classification of States

4.1 Definition (Closed). A set A is **closed** if it is impossible to get out. Formally $c \in A$ and $y \notin A$ then $P_{xy} = 0$

4.2 Definition (irreducible). A set B is irreducible if every state is reachable from another in k steps or every state communicates with with all other states. Formally

$$x, y \in B \Rightarrow x \to y$$



4.3 Lemma (Commutating recurrent states). If x is recurrent and $x \to y$ then y is recurrent

4.4 Lemma (Existence of recurrent states in finite closed sets). If A is finite and closed then $\exists x \in A$ such that x is recurrent.

4.5 Theorem (Closed and irreducible sets are recurrent). If $C \subseteq \mathcal{X}$ is **finite**, closed and irreducible then all $x \in C$ are recurrent.

4.6 Theorem (Decomposition Theorem). If $\mathcal{X}$ is finite then

$$\mathcal{X} = T \cup R_1 \cup R_2 \cup \cdots \cup R_k$$

where T is the set of transient states and R_i for $1 \le i \le k$ are are closed irreducible sets of recurrent states.

4.7 Definition (Number of visits). $N(y)$ is the number of visits to y after initial time.

4.8 Lemma (Expected number of visits).

$$\mathbb{E}_x[N(y)] = \begin{cases} 0 & \rho_{xy} = 0 \\ \frac{\rho_{xy}}{1-\rho_{yy}} & \rho_{xy} > 0 \end{cases}$$

4.9 Lemma (Expected number of visits II).

$$\mathbb{E}_x[N(y)] = \sum_{n=1}^{\infty} [\mathbf{P}^n]_{xy}$$

4.10 Theorem (Equivalent condition for recurrence). y is recurrent if and only if

$$\sum_{n=1}^{\infty} [\mathbf{P}^n]_{yy} = \mathbb{E}_y[N(y)] = \infty$$

4.2 Existence of Stationary measure

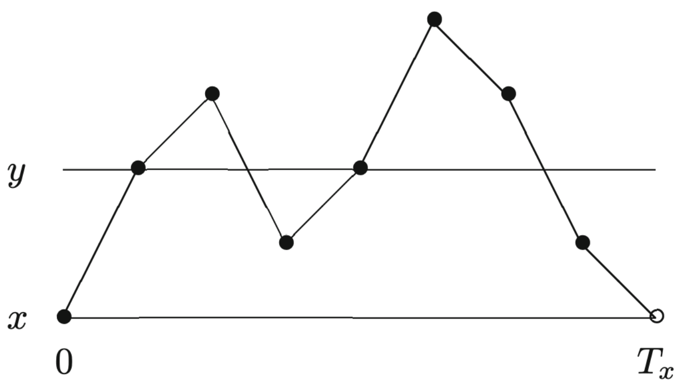
4.11 Theorem (Existence of Stationary measure). Suppose $\mathcal{X}$ is irreducible and recurrent there exists a stationary measure $\vec{\mu}$ with

$$0 < \mu_y < \infty \qquad\qquad y \in \mathcal{X}$$

Let $x \in \mathcal{X}$ be recurrent by Existence of recurrent states in finite closed sets. We define $\vec{\mu}^x$ as

$$\mu_y^x = \mathbb{E}_x[\# \text{ of visits to } y \text{ before } x] = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, \tau_x > n)$$

$\vec{\mu}^x$ is a stationary measure for $\mathbf{P}$.



5 Week 5

5.1 Definition (Ergodicity)**.** If $\vec{\pi}$ is a stationary distribution given $X_0 \sim \vec{\mu}$ if we have

$$\vec{\mu}\mathbf{P}^n \to \vec{\pi}$$

then $\mathbf{P}$ has Ergodicity.

5.2 Theorem (Ergodicity equivalent definition)**.**

$$\vec{\mu}\mathbf{P}^n \to \pi \iff [P^n]_{xy} \to \vec{\pi}_y \qquad\qquad \forall x \in \mathcal{X}$$

5.3 Remark. If y is transient then $\mathbb{E}_X[N(y)] < \infty$. Then

$$\mathbb{E}_X[N(y)] = \sum_{n=1}^{\infty}[\mathbf{P}^n]_{xy}$$

Then $[\mathbf{P}^n]_{xy} \to 0$. Meaning $\vec{\pi}_y = 0$, so all transient states have Ergodicity.

5.4 Definition (Periodicity)**.** A state x has a period d_x if

$$I_x = \{n \geq 0 \mid [P^n]_{xx} > 0\} \qquad\qquad \mathrm{d}_x = \gcd I_x$$

x is aperiodic if $\mathrm{d}_x = 1$ and periodic if $\mathrm{d}_x > 1$.

5.5 Definition (Class property)**.** A property $\mathcal{K}$ of a state x is called a **class property** if x has $\mathcal{K}$, $x \to y$ and $y \to x$ then y has $\mathcal{K}$.

5.6 Lemma. Periodicity is a class property If $x \to y$ and $y \to x$ then $\mathrm{d}_x = \mathrm{d}_y$.

5.7 Lemma. I_x is closed under addition

$$a, b \in I_x \Rightarrow (a + b) \in I_x$$

5.8 Lemma. If x is aperiodic then $\exists n_0$ such that for all $n \geq n_0$ $n \in I_x$.

5.9 Lemma. If $P_{xx} > 0$ then x is aperiodic.

6 Week 6

6.1 Convergence Theorems

6.1 Note.

$$\begin{aligned}&\mathbf{I} : \text{Irreducible}\\&\mathbf{A} : \text{Aperiodic}\\&\mathbf{R} : \text{Recurrent}\\&\mathbf{S} : \text{Stationary distribution } \pi \text{ exists}\end{aligned}$$

6.2 Remark. $I, S \Rightarrow R$

6.3 Definition (Convergence in Probability)**.** $\{X_n\} \xrightarrow{p} X$ if $\forall \varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

6.4 Definition (Almost surely convergence)**.** $\{X_n\} \xrightarrow{a.s} X$ if

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

6.5 Theorem (Weak Law of Large Numbers (WLLN))**.** Let $\{X_k\}_{k\in\mathbb{N}} \sim$ iid with $\mathbb{E}[X_k] = \mu$ then

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k \xrightarrow{p} \mu$$

6.6 Theorem (Strong Law of Large Numbers (SLLN))**.** Let $\{X_k\}_{k\in\mathbb{N}} \sim$ iid with $\mathbb{E}[X_k] = \mu$ then

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k \xrightarrow{a.s} \mu$$

6.7 Theorem (Convergence Theorem)**.** If I, A, S hold then

$$P_{xy}^n \to \pi_y \qquad\qquad n \to \infty$$

6.8 Theorem (Asymptotic Frequency)**.** Suppose I, R hold and let $N_n(y)$ be the number of visits to y upto time n then

$$\frac{N_n(y)}{n} \xrightarrow{a.s} \frac{1}{\mathbb{E}_y \tau_y}$$

Proof. If $R(k)$ is the $k-$th return time to y then by SLLN $\frac{R(k)}{k} \xrightarrow{a.s} \mathbb{E}_y[\tau_y]$ then we use squeeze theorem to get the result. $\square$

6.9 Lemma (Recurrent States)**.** If **S** holds and $\pi_y > 0 \Rightarrow y$ is recurrent.

6.10 Theorem (Stationary Distribution Uniqueness)**.** If $\mathbf{I}, \mathbf{S}$ hold then

$$\pi_y = \frac{1}{\mathbb{E}_y[\tau_y]}$$

6.11 Theorem. Consider $\vec{\mu}^x$ from this theorem

$$\vec{\mu}^x_y = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, \tau_x > n)$$

If $\mathbf{I}, (\mathbf{R}), \mathbf{S}$ hold then

$$\vec{\mu}^x_y = \frac{\vec{\pi}_y}{\vec{\pi}_x}$$

Proof. $\sum_{y \in \mathcal{X}} \vec{\mu}^x_y = \sum_{n=0}^{\infty} \mathbb{P}_x(\tau_x > n) = \mathbb{E}_x[\tau_x] = \frac{1}{\pi_y}$ □

6.12 Theorem (Expected value of function)**.** If $\mathbf{I}, \mathbf{S}$ hold and $\sum_x |f(x)|\vec{\pi}_x < \infty$ then

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \xrightarrow{a.s} \sum_{x \in \mathcal{X}} f(x)\vec{\pi}_x = \mathbb{E}_{\vec{\pi}}[f(X)]$$

6.13 Theorem (Stationary measures uniqueness)**.** If $\vec{\nu}$ is a stationary measure then $\vec{\nu} = \vec{\mu}^x$ for some x.

7 Summary of Convergence Theorems

Name	Result	Conditions
Existence of stationary measure $\vec{\mu}^x$	$\vec{\mu}^x_y = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, \tau_x > n)$	$\mathbf{I}, \mathbf{R}$
Convergence Theorem	$[\mathbf{P}^n]_{xy} \to \vec{\pi}_y$ as $n \to \infty$	$\mathbf{I}, \mathbf{A}, \mathbf{S}$
Asymptotic Frequency	$\frac{N_n(y)}{n} \xrightarrow{a.s} \frac{1}{\mathbb{E}_y[\tau_y]}$	$\mathbf{I}, \mathbf{R}$
Stationary Distribution Uniqueness	$\vec{\pi} = \frac{1}{\mathbb{E}_y[\tau_y]}$	$\mathbf{I}, \mathbf{S}$
Expected value of function	$\frac{1}{n} \sum_{n=0}^{\infty} \xrightarrow{a.s} \sum_{x \in \mathcal{X}} f(x)\vec{\pi}_x = \mathbb{E}_{\vec{\pi}}[f(X)]$	$\mathbf{I}, \mathbf{S}, \sum_{x \in \mathcal{X}} \lvert f(x) \rvert \vec{\pi}_x < \infty$
Expected number of visits before x	$\vec{\mu}^x_y = \frac{\vec{\pi}_y}{\vec{\pi}_x}$	$\mathbf{I}, \mathbf{S}$

7.1 Note.

$\mathbf{I}$: Irreducible
$\mathbf{A}$: Aperiodic
$\mathbf{R}$: Recurrent
$\mathbf{S}$: Stationary distribution π exists

7.2 Remark. $I, S \Rightarrow R$

8 Week 7

8.1 Exit Distributions

8.1 Definition (Visiting Time for Set)**.**

$$V_A = \min\{n \geq 1 : X_n \in A\}$$

8.2 Theorem (Exit Distribution)**.** For a Markov chain on state space $\mathcal{X}$. Let $A, B \subseteq \mathcal{X}$ and let $C = \mathcal{X} \setminus (A \cup B)$ if we have

$$\begin{aligned}
h(a) &= 1 && a \in A \\
h(b) &= 0 && b \in B \\
h(x) &= \sum_{y \in \mathcal{X}} [\mathbf{P}]_{xy} h(y) && x \in C
\end{aligned}$$

If $\mathbb{P}_x(V_A \wedge V_B < \infty) > 0$ for all $x \in C$ then $h(x) = P_x(V_A < V_B)$ (Chain visits A before B).

8.2 Exit Times

8.3 Theorem (Exit Time)**.** Let $T = \min\{n \geq 0 : X_n \in A\}$ be the time to exit. Suppose $C = \mathcal{X} \setminus A$ is finite and $\mathbb{P}_x(T < \infty) > 0$ for all $x \in C$. Then we define

$$\begin{aligned}
g(a) &= 0 && a \in A \\
g(x) &= 1 + \sum_{y \in \mathcal{X}} P_{xy} g(y)
\end{aligned}$$

Then $g(x) = \mathbb{E}_x[T]$

9 Infinite State Spaces

9.1 Laplace Matrix

9.1 Definition (Laplace Matrix)**.** The matrix $L = P - I$ is called the the Laplace matrix

For the exit distribution $h(x) = \mathbb{P}_x(V_A < V_B)$ it satisfies the Laplace equation:

$$\begin{cases} Lh = 0 & \text{on } C = \mathcal{X} \setminus (A \cup B) \\ h = 1 & \text{on } A \\ h = 0 & \text{on } B \end{cases}$$

For the exit time $g(x) = \mathbb{E}_x[T]$ it solves the *Poisson equation*

$$\begin{cases} Lg = -1 & \text{on } C \\ g = 0 & \text{on } \mathcal{X} \setminus C \end{cases}$$

9.2 Definition (Positive and Null Recurrent)**.**
A state is **positive recurrent** if $\mathbb{E}_x[\tau_x] < \infty$. A state x is null recurrent if it is recurrent but not null recurrent. ($\mathbb{P}_x(\tau_x < \infty) = 1$ but $\mathbb{E}_x[\tau_x] = \infty$)

9.3 Example. Reflecting Random Walk. The probabilities are defined as

$$\begin{aligned} P_{i,i+1} &= p \\ P_{i,i-1} &= 1 - p \qquad\qquad i \geq 1 \\ P_{0,0} &= 1 - p \end{aligned}$$

Using the detail balance condition since the matrix is tridiagonal we can say that

$$\begin{aligned} \pi_i P_{i,i+1} &= \pi_{i+1} P_{i+1,i} \\ \pi_i p &= \pi_{i+1}(1-p) \\ \pi_{i+1} &= \frac{p}{1-p}\pi_i \end{aligned}$$

Let $\pi_0 = c$ then we have the solution $\pi_i = c\left(\frac{p}{1-p}\right)^i$.
Case I $p < \frac{1}{2}$ then $\frac{p}{1-p} < 1$ then $\sum_i \pi_i < \infty$ then we can pick c to make π a stationary distribution. This gives the solution $\pi_0 = \frac{1-2p}{1-p}$. Since $\mathbb{E}_0[\tau_0] = \frac{1}{\pi_0} < \infty$ **0 is positive recurrent**.
Case II $p > \frac{1}{2}$ **all states are transient**
Case III $p = \frac{1}{2}$ then **0 is null recurrent**

9.4 Theorem (equivalent conditions)**.** For an **I**rreducible chain the following are equivalent:

(i) Some state is positive recurrent.

(ii) There is a stationary distribution π

(iii) All states are positive recurrent

9.2 Galton Watson Process

(1) Start with a single individual in generation 0

(2) This individual gives birth to $X \in \mathbb{N}$ random number of children with $\mathbb{P}(X = 0) > 0$ and $\mu = \mathbb{E}[X] < \infty$.

(3) The $r-$th individual in the $n-$th generation gives birth to $X_{r,n}$ which has the same distribution as X and independent of all other random variables.

Let Z_n be the size of the n-th generation then

$$\begin{cases} Z_{n+1} = X_{1,n} + X_{2,n} + \cdots + X_{Z_n,n} \\ Z_0 = 1 \end{cases}$$

$\{Z_n\}$ is a Markov chain. Using tower rule

$$\mathbb{E}[Z_{n+1}] = \mathbb{E}[\mathbb{E}[Z_{n+1}|Z_n]] = \sum_{j=1}^{\infty} \mathbb{E}[Z_{n+1}|Z_n = j]\mathbb{P}(Z_n = j) = \mu E[Z_n]$$

By induction we have $\mathbb{E}[Z_n] = \mu^n \mathbb{E}[Z_0]$
Survival and **Extinction**

$$\{\text{Survival}\} = \bigcap \{Z_n \neq 0\}, \qquad \{\text{Extinction}\} = \bigcup \{Z_n = 0\}$$
$$\mathbb{P}(\text{Extinction}) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0)$$

9.5 Example. Let $f_n = \mathbb{E}[\theta^{Z_n}]$ where Z_n is the size of the generation from Galton Watson Process. Then $f_1 = \mathbb{E}[\theta^{Z_1}] = \mathbb{E}[\theta^X]$ then we also have

$$f_{n+1}(\theta) = f_n(f_1(\theta)) = \underbrace{f_1 \circ f_1 \circ f_1 \cdots \circ f_1(\theta)}_{n-\text{times}} = f_1(f_n(\theta))$$

Assuming $\lim_{n\to\infty} f_n(0)$ exists we have $s = \lim_{n\to\infty} f_{n+1}(0) = f_1(\lim_{n\to\infty} f_n(0)) = f_1(s)$. So s is a **fixed-point** of f_1. Since $f_n(0) = \mathbb{P}(Z_n = 0)$ so

$$\mathbb{P}(\text{Extinction}) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) = s$$

Meaning $\mathbb{P}(\text{Extinction})$ is a fixed point of f_1.

9.6 Theorem (Fixed points and $\mathbb{P}$(Extinction))**.** If $\mathbb{E}[X] > 1$ then the extinction probability is the unique root of the equation $p = f_1(p)$ such that $p \in (0, 1)$. If $\mathbb{E}[X] \leq 1$ then $\mathbb{P}(\text{Extinction}) = 1$.

10 Generating Functions

10.1 Definition (Probability generating function (pgf))**.** The **pgf** of a non-negative integer valued random variable X,$G_X : [0,1] \to [0,1]$ is defined as:

$$\begin{aligned} G_X(\theta) = \mathbb{E}[\theta^X] &= \sum_{k=0}^{\infty} \theta^k \mathbb{P}(X = k) \\ &= \mathbb{P}(X=0) + \theta\mathbb{P}(X=1) + \theta^2\mathbb{P}(X=2) + \cdots + \theta^k\mathbb{P}(X=k) + \cdots \end{aligned}$$

10.2 Definition (Moment generating function (mgf))**.** The moment generating function M_X for a non-negative integer valued random variable X is the function $M_X : (-r, r) \to \mathbb{R}$ is defined as:

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \sum_{k=0}^{\infty} e^{tk} \mathbb{P}(X = k) \\ &= 1 + t\mathbb{E}[X] + \frac{t^2}{2!}\mathbb{E}[X^2] + \frac{t^3}{3!}\mathbb{E}[X^3] + \cdots + \frac{t^k}{k!}\mathbb{E}[X^k] \end{aligned}$$

and $r > 0$ such that the value exists on $(-r, r)$.

10.3 Theorem (Properties of generating functions)**.**

$$\begin{aligned} &G_X(0) = \mathbb{P}(X = 0) \\ &\mathbb{P}(X = n) = \frac{1}{n!} \left.\frac{\mathrm{d}^n G_X}{\mathrm{d}\theta^n}\right|_0 \\ &G'_X(\theta) = \mathbb{E}[X\theta^{X-1}] \qquad\qquad G'_X(1) = \mathbb{E}[X] \\ &\left.\frac{\mathrm{d}^n M_X}{\mathrm{d}t^n}\right|_0 = \mathbb{E}[X^n] \end{aligned}$$

10.4 Lemma (Generating functions are unique)**.** If 2 random variables have the same moment generating functions they have the same distribution.

11 Exponential Distribution

11.1 Definition (Exponential cdf)**.** If $T \sim EXP(\lambda)$ then

$$\mathbb{P}(T \leq t) = 1 - e^{-\lambda t}$$

Moreover

$$\begin{aligned}
\mathbb{E}[T] &= \frac{1}{\lambda} \\
\mathbb{E}[T^2] &= \frac{2}{\lambda^2} \\
\mathrm{Var}[T] &= \mathbb{E}[T^2] - (\mathbb{E}[T])^2 = \frac{1}{\lambda^2}
\end{aligned}$$

11.2 Theorem (Memoryless property)**.** Let $T \sim \mathrm{EXP}(\lambda)$ then $\forall t, s \leq 0$

$$\mathbb{P}(T > t + s | T > t) = \mathbb{P}(T > s)$$

11.3 Theorem (Exponential Races)**.** Let $V = \min\{T_1, \ldots, T_n\}$ such that $T_i = \mathrm{EXP}(\lambda_i)$ independently and let I be the index of the minimum T_i then

$$\begin{aligned}
V &= \mathrm{EXP}\left(\sum \lambda_i\right) \\
\mathbb{P}(I = i) &= \frac{\lambda_i}{\sum\limits_{k=1}^{n} \lambda_k}
\end{aligned}$$

and V, I are independent.

Proof. Sketch. $\mathbb{P}(\min\{T, S\} > t) = \mathbb{P}(T > t)\mathbb{P}(S > t)$ and induction. □

11.4 Theorem (Sum of exponential variables)**.** Let $\tau_1, \ldots, \tau_n$ be $\mathrm{EXP}(\lambda)$ then the sum $T_n = \tau_1 + \tau_2 + \cdots + \tau_n$ has a gamma(n, λ) distribution.

$$f_{T_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \qquad t \geq 0$$

12 Poisson Process

12.1 Definition (Poisson Distribution)**.** X has a Poisson distribution with mean λ if

$$P(X = n) = \frac{\lambda^n e^{-\lambda}}{n!} \qquad \text{for } n = 0, 1, 2 \dots$$

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

(2) $\text{Var}[X] = \lambda$

(3) **m.g.f** is $M_X(t) = e^{\lambda(e^t - 1)}$

12.2 Theorem (Limit of binomial is Poisson)**.** Let Y_n be binomial$(n, \frac{\lambda}{n})$ and $X =$ poisson(λ) then

$$\lim_{n\to\infty} \mathbb{P}(Y_n = k) = \mathbb{P}(X = k)$$

12.3 Theorem (Sum of poisson is poisson)**.** If $X_i \sim$ poisson(λ_i) for $1 \leq i \leq n$ independently then

$$S = \sum_{i=1}^{n} X_i \sim \text{poisson}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$$

12.4 Definition (Poisson Process)**.** Let $N(t)$ be the number of occurrences/arrivals in $[0, t]$. Then $\{N(t) : t \geq 0\}$ is a (homogeneous) poisson process if:

(1) $N(0) = 0$

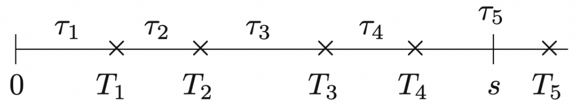
(2) $N(t + s) - N(s) = \text{poisson}(t\lambda)$

(3) $N(t)$ has **independent increments**

12.1 Constructing the Poisson Process

12.5 Definition. Let $\tau_1, \tau_2, \ldots, \tau_n$ be independent $\text{EXP}(\lambda)$ variables and let $T = \tau_1 + \tau_2 + \cdots + \tau_n$ and define $N(s) = \max\{n : T_n \leq s\}$

τ_i is the interval between arrivals implying T_n is the time for the n-th arrival and $N(s)$ is the number of arrivals by time s.



(i) $N(s)$ has a Poisson distribution with mean λs

(ii) $N(t+s) - N(s) = \text{poisson}(\lambda t)$ and is independent of $N(r), 0 \leq r \leq s$.

(iii) $N(t)$ has independent increments.

13 More complex models

13.1 Definition (Measure)**.** A **signed** measure $\mu : \Sigma \to \mathbb{R}$ on (Ω, Σ) such that $\Sigma \subseteq \text{Pow}(\Omega)$ such that

(1) $\mu(\varnothing) = 0$

(2) μ has countable additivity. If $\{E_i\}_{i=1}^{\infty}$ are mutually disjoint then $\mu(\cup E_i) = \sum_i \mu(E_i)$

(3) $\mu(\Omega) = 1$ (For distributions)

13.2 Definition (Metric)**.** A **distance** function is a metric if

(i) $d(x, y) \geq 0$ and $d(x, y) = 0 \iff x = y$ (Positive)

(ii) $d(x, y) = d(y, x)$ (Symmetry)

(iii) $d(x, y) \leq d(x, z) + d(z, y)$ (Triangle inequality)

13.3 Definition (Total variation Distance/Measure)**.** Suppose μ is a signed measure on (Σ, Ω) then

$$\|\mu\|_{TV} = \sup\{\mu(A) | A \in \Sigma\} - \inf\{\mu(A) | A \in \Sigma\}$$

If X, Y are **integer valued** random variables then

$$d_{TV}(X, Y) = \max_{A \subseteq \mathbb{Z}} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|$$

d_{TV} is a **metric**.

13.4 Remark. $d_{TV}(X, Y) = \|X - Y\|_{TV} = \frac{1}{2} \sum\limits_k |\mathbb{P}(X = k) - \mathbb{P}(Y = k)|$

13.5 Definition (Convergence in TV norm)**.** $X_n \to X$ in TV if

$$\lim_{n \to \infty} \|X_n - X\|_{TV} = 0$$

13.6 Theorem. Suppose $X_{n,m}$ ∼Bernoulli$(p_{n,m})$ are independent and let $S_n = \sum\limits_{m=1}^{n} X_{n,m}$ and $\lambda_n = \mathbb{E}[S_n] = \sum\limits_{m=1}^{n} p_{n,m}$ and $Y_n = \text{poisson}(\lambda_n)$ then

$$\|S_n - Y_n\|_{TV} \leq \sum_{m=1}^{n} p_{n,m}^2$$

If moreover, $\sup_m p_{n,m} \to 0$ and $\lambda_n \to \lambda$ as $n \to \infty$ and $Y = \text{poisson}(\lambda)$ then

$$\lim_{n\to\infty} \|S_n - Y\|_{TV} = 0$$

That is $S_n \to Y$ in TV-norm.

13.7 Example. Let $\lambda : [0,1] \to \mathbb{R}_+$ be continuous and let $p_{n,m} = \frac{1}{n}\lambda(\frac{m}{n})$ then

$$0 \leq \lim_{n\to\infty} p_{n,m} \leq \lim_{n\to\infty} \frac{\max(\lambda(t) : t \in [0,1])}{n} = 0$$

and $\lambda_n = \sum\limits_{m=1}^{n} \frac{1}{n}\lambda(\frac{m}{n})$, This is a Riemann sum and $\lim\limits_{n\to\infty} \lambda_n = \int_0^1 \lambda(t)\,\mathrm{d}t$ by **Theorem 13.6** we have

$$S_n \to \text{poisson}\left(\int_0^1 \lambda(t)\,\mathrm{d}t\right)$$

13.8 Definition (Non-homogeneous poisson process)**.** Let $N(t)$ be the total number of occurrences in $[0,t]$. $\{N(s) : s \geq 0\}$ is a Non-homogeneous poisson process if

(i) $N(0) = 0$

(i) $N(t+s) - N(s) = \text{poisson}\left(\int_s^{t+s} \lambda(t)\,\mathrm{d}r\right)$

(i) $N(t)$ has independent increments.

14 Compound Poisson Process

14.1 Theorem (Compound Expectation)**.** Let $Y_1, Y_2, \ldots$ be **i.i.d** and let N be an integer valued non-negative random variable. Such that $\mathbb{E}|Y_i|, \mathbb{E}[N] < \infty$ and we define $S = \sum_{i=1}^{N} Y_i$ then using

(i) $\mathbb{E}[S] = \mathbb{E}[N]\mathbb{E}[Y_i]$

(ii) If $\mathbb{E}[Y^2], \mathbb{E}[N^2] < \infty$ then

$$\mathrm{Var}[S] = \mathbb{E}[N]\mathrm{Var}[Y_i] + \mathrm{Var}[N]\left(\mathbb{E}[Y_i]\right)^2 \tag{ii}$$

(iii) If $N = poisson(\lambda)$ and $\mathbb{E}[Y^2] < \infty$ then

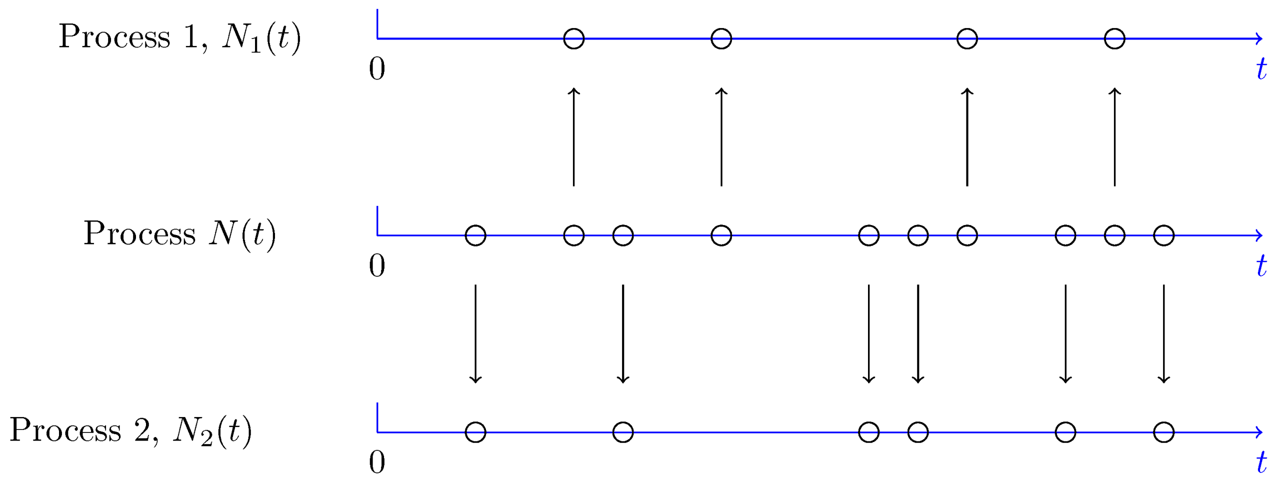
$$\mathrm{Var}[S] = \lambda\mathbb{E}[Y_i^2] \tag{iii}$$

14.1 Thinning

14.2 Theorem (Thinning)**.** Let $\{N(t) : t \geq 0\}$ be a poisson process with rate $\lambda(r)$ for each arrival i, Y_i is **i.i.d** $Y_i \sim Y$. Then let

$$N_j(t) = \#\{i \leq N(t) : Y_i = j\}$$

Then $N_j(t)$ are **independent** rate $\lambda(r)\mathbb{P}(Y = j)$ poisson processes.



14.2 M/G/∞ Queue

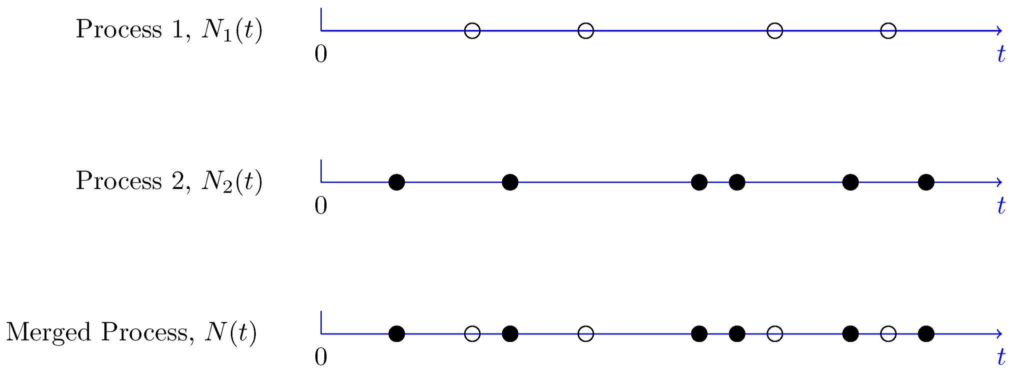
14.3 Theorem (M/G/∞)**.** Suppose the number of arrivals follow a poisson process with rate λ, duration of each arrival is **i.i.d** X with c.d.f G and $\mathbb{E}[X] = \mu$. Then the arrivals still in progress at time t is

$$\text{poisson}\left(\lambda \int_0^t [1 - G(r)]\, \mathrm{d}r\right)$$

14.4 Corollary (∞)**.** The number of arrivals in progress at $t = \infty$ is poisson $(\mu\lambda)$

14.3 Superposition

14.5 Theorem ($\bullet$)**.** Suppose $\{N_1(t)\}, \{N_2(t)\}, \cdots, \{N_k(t)\}$ are independent poisson processes then $\{N(t) = N_1(t) + N_2(t) + \cdots + N_k(t) : t \geq 0\}$ is a poisson process with rate $\lambda_1 + \lambda_2 + \cdots + \lambda_k$



14.4 Conditioning

Let $T_1, T_2, \ldots, T_n$ be the arrival time of a poisson process with rate λ and let $U_1, U_2, \ldots, U_n$ be uniformly distributed on $[0, t]$ and let $V_1 < V_2 < \cdots < V_n$ be the U_i in sorted order.

14.6 Theorem (Poisson Conditioning)**.** If we condition on $N(t) = n$ then the vector $(T_1, \ldots, T_n)$ has the same distribution as $(V_1, V_2, \ldots, V_n)$ and the set $\{T_1, T_2, \ldots, T_n\}$ has the same distribution as $\{U_1, U_2, \ldots, U_n\}$

14.7 Theorem (Conditioning Binomial Distribution)**.** If $0 \leq r < t$ and $0 \leq m \leq n$ then

$$\mathbb{P}(N(r) = m | N(t) = n) = \binom{n}{m} \left(\frac{r}{t}\right)^m ()^{n-m}$$

The distribution of $N(r)$ given $\{N(t) = n\}$ is binomial$(n, \frac{r}{t})$ and does not depend on λ.