

Stochastic Process 1

Rui Gong

February 26, 2024

Contents

1	Poisson Process	3
1.1	Poisson Approximation to Binomial	3
1.2	Total Variance Distance	3
1.3	Probability Axioms	5
1.4	Cumulative Distribution Function (c.d.f.)	7
1.4.1	Geometric Distribution to Exponential, the Memoryless variables	8
1.5	Point Process	9
1.6	Bernoulli and Poisson	10
2	Markov-Chain	17
2.1	Transition Matrix	19
2.2	Communication, Recurrence and Transience	21

1 Poisson Process

1.1 Poisson Approximation to Binomial

Given a Poisson random variable $Y \sim \text{Poisson}(\lambda)$ with pdf

$$\Pr[Y = k] = \frac{e^{-\lambda} \lambda^k}{k!}, \forall k \in N_0 = \{0, 1, \dots\}.$$

The probability of a binomial random variable being k is

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Theorem 1.1. Given $p \rightarrow 0$, $np \rightarrow \lambda$, we have that

$$\binom{n}{k} p^k (1-p)^{n-k} \rightarrow \frac{e^{-\lambda} \lambda^k}{k!}$$

Proof.

$$\begin{aligned} \binom{n}{k} p^k (1-p)^{n-k} &= \binom{n}{k} \left(\frac{np}{n}\right)^k \left(1 - \frac{np}{n}\right)^{n-k} \\ &= \frac{n!}{k! (n-k)!} \left(\frac{np}{n}\right)^k \left(1 - \frac{np}{n}\right)^n \left(1 - \frac{np}{n}\right)^{-k} \\ &= \frac{1}{k!} \underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\rightarrow 1} \underbrace{(np)^k}_{\rightarrow \lambda^k} \underbrace{\left(1 - \frac{np}{n}\right)^n}_{\rightarrow e^{-\lambda}} \underbrace{\left(1 - \frac{np}{n}\right)^{-k}}_{\rightarrow 1} \\ &\rightarrow \frac{1}{k!} \lambda^k e^{-\lambda} \end{aligned}$$

□

With that, we consider three different binomial random variables:

$$X_n \sim \text{Binomial}(n, p_n), p_n \rightarrow 0, np_n \rightarrow \lambda > 0, \text{ as } n \rightarrow \infty.$$

$$Z_p \sim \text{Binomial}(n(p), p), p \rightarrow 0, n(p)p \rightarrow \lambda > 0, \text{ as } p \rightarrow 0.$$

$$N_x \sim \text{Binomial}(n(x), p(x)), p(x) \rightarrow 0, n(x)p(x) \rightarrow \lambda > 0, \text{ as } x \rightarrow \infty.$$

For example, if $X_n \sim \text{Binomial}(n, 2/n)$, then we expect

$$\binom{n}{k} p^k (1-p)^{n-k} \rightarrow \frac{e^{-2} 2^k}{k!}$$

1.2 Total Variance Distance

Let X_1, \dots, X_n be n independent Bernoulli random variables, where $\mathbb{E}[X_i] = p_i$.

Given $S = \sum X_i$ and $T \sim \text{Poisson}(\lambda = \sum p_i)$, how close are these two distributions? Or, how to measure the closeness?

Definition 1.2. Given two random variables X, Y , (which shares the sample space), we have the *total variance distance* defined as

$$d_{TV}(X, Y) = \sup_A |\Pr[X \in A] - \Pr[Y \in A]|$$

where A is a Borel set defined with respect to the sample space σ -algebra.

Example 1.3. Given two distributions

	0	1	2	3
$\Pr[X = k]$	5/10	3/10	1/10	1/10
$\Pr[Y = k]$	2/10	1/10	1/10	6/10

Table 1: Discrete Distribution Distance

If $A = \{0, 1\}$, then

$$|\Pr[X \in A] - \Pr[Y \in A]| = 3/10 + 2/10 = 1/2.$$

If $A = \{3\}$,

$$|\Pr[X \in A] - \Pr[Y \in A]| = |1/10 - 6/10| = 1/2.$$

Lemma 1.4. If X, Y take values in a countable set E ,

$$\begin{aligned} d_{TV}(X, Y) &= \sum_{i \in E} (\Pr[X = i] - \Pr[Y = i])^+ \\ &= \sum_{i \in E} (\Pr[Y = i] - \Pr[X = i])^+ \\ &= \frac{1}{2} \sum_{i \in E} |\Pr[Y = i] - \Pr[X = i]| \end{aligned}$$

Proposition 1.5. Given two random variables, we have

$$d_{TV}(X, Y) \leq \Pr[X \neq Y]$$

Proof. For any A ,

$$\begin{aligned} &|\Pr[X \in A] - \Pr[Y \in A]| \\ &= |\Pr[X \in A, Y \in A] + \Pr[X \in A, Y \notin A] - \Pr[Y \in A, X \in A] - \Pr[Y \in A, X \notin A]| \\ &= |\Pr[X \in A, Y \notin A] - \Pr[Y \in A, X \notin A]| \\ &\leq \max\{\Pr[X \in A, Y \notin A], \Pr[Y \in A, X \notin A]\} \leq \Pr[X \neq Y]. \end{aligned}$$

□

Let X_1, \dots, X_n be independent Bernoulli random variables with $\mathbb{E}[X_i] = p_i$. Let $S := \sum X_i$, and $T \sim \text{Poisson}(\lambda := p_1 + \dots + p_n)$. Then $\mathbb{E}[S] = \mathbb{E}[\sum X_i] = \sum \mathbb{E}[X_i] = \sum p_i = \lambda$.

Let Y_1, \dots, Y_n be independent Poisson random variables with $\mathbb{E}[Y_i] = p_i$. Then $T = \sum Y_i \sim \text{Poisson}(\lambda)$. We also have

$$[S \neq T] \subseteq \underbrace{[X_1 \neq Y_1]}_{B_1} \cup [X_2 \neq Y_2] \cup \dots \cup [X_n \neq Y_n]$$

And hence

$$\begin{aligned} \Pr[S \neq T] &\leq \Pr[B_1 \cup \dots \cup B_n] \\ &\leq \Pr[B_1] + \dots + \Pr[B_n] \\ &\leq p_1^2 + \dots + p_n^2 \end{aligned}$$

where $\Pr[X_i = Y_i] = 1 - p + pe^{-p}$, $\Pr[X_i \neq Y_i] = p - pe^{-p} \leq p(1 - (1 - p + p^2/2! + \dots)) = p(p - p^2/2! + \dots) \leq p^2$.

Hence,

$$d_{TV}(S, T) \leq \Pr[S \neq T] \leq \sum_{i=1}^n p_i^2.$$

Consider $X_1 \sim \text{Bernoulli}(p_1 = 1/5)$, $X_2 \sim \text{Bernoulli}(p_2 = 1/6)$, $X_3 \sim \text{Bernoulli}(p_3 = 1/10)$, $S = X_1 + X_2 + X_3$ and $T \sim \text{Poisson}(\lambda = \frac{7}{15})$. Then if estimate T by S , for example,

$$\Pr[S \text{ is an odd number}] \approx \Pr[T \text{ is an odd number}]$$

the probability of getting an error is at most

$$(1/5)^2 + (1/6)^2 + (1/10)^2$$

by letting A be the set of odd numbers.

1.3 Probablity Axioms

Consider the sample space Ω , the set of events \mathcal{F} and the probability P , where

Ω : sample spaces - set of all outcomes

\mathcal{F} : all events

$P : \mathcal{F} \rightarrow [0, 1]$

. Then we can write a random variable X_1 as:

$$X_1 : \Omega \rightarrow \mathbb{R}$$

and an event as

$$B_1 = [X_1 \neq Y_1] = [w \in \Omega | X_1(w) \neq Y_1(w)].$$

Definition 1.6. Event Axioms:

E.1 $\Omega \in \mathcal{F}$

$$\text{E.2 } A \in \mathcal{F} \implies A^C \in \mathcal{F}$$

$$\text{E.3 } A_1, A_2, \dots \in \mathcal{F} \implies A_1 \cup A_2 \cup \dots \in \mathcal{F}$$

Definition 1.7. Probability Axioms:

$$\text{P.1 } A \in \mathcal{F} \implies P(A) \geq 0$$

$$\text{P.2 Countable additivity. } A_1, A_2, \dots \text{ being disjoint events, then } P(A_1 \cup A_2 \cup \dots) = \sum_{i=1}^{\infty} P(A_i).$$

$$\text{P.3 } P(\Omega) = 1.$$

Example 1.8. $X = (\Omega, \mathcal{F}) \rightarrow \mathbb{R}$, and let B be a Borel set. Then we can write $\{X = 3\} = \{w \in \Omega : X(w) = 3\} \in \mathcal{F}$. Similarly, $P(X \in B) \in \mathcal{F}$.

$\Omega = \{a, b, c\}$, $\mathcal{F} = \{\emptyset, \Omega, \{a\}, \{b, c\}\}$. Given $X(a) = 1, X(b) = 2, X(c) = 3$, we have

$$[X = 3] = [w \in \Omega : X(w) = 3] = [c]$$

which is not in the event, so X is not a random variable. If $X(b) = 3$, then X is a random variable.

Given $X \sim \text{Exp}(\lambda), Y \sim \text{Exp}(\mu)$ and $X \perp Y$, then

$$\begin{aligned} &P(X > s, Y - X > t | X < Y) \\ &= P(X > s | X < Y) P(Y - X > t | X < Y) \end{aligned}$$

$$\lambda_n \rightarrow \lambda \implies (1 + \frac{\lambda_n}{n})^n \rightarrow e^\lambda. f(h) = o(h) \implies f(h)/h \rightarrow 0 \text{ as } h \rightarrow 0.$$

Fix x , a function f is differentiable at x iff there exists a number $f'(x)$ such that

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + o(h) \\ \frac{f(x+h) - f(x)}{h} &= f'(x) + o(h)/h, h \rightarrow 0 \end{aligned}$$

For example, if we want to show $n \log(1 + \frac{\lambda_n}{n}) \rightarrow \lambda$. Take $h_n = \lambda_n/n, x = 1$.

$$n \log(1 + h_n) = nh_n + nO(h_n) = nh_n + \frac{\lambda_n}{h_n} O(h_n)$$

where $\log(1 + h) = \log(1) + h + o(h)$. Then as $n \rightarrow \infty$, we have $h_n \rightarrow 0, nh_n \rightarrow \lambda, n \log(1 + \lambda_n/n) \rightarrow \lambda$.

Definition 1.9. Suppose X is nonnegative, integer-valued random variable $P(X = k) = p_k$ for $k = 0, 1, 2, \dots$, then the *probability-generating function* is defined as:

$$G(s) := \mathbb{E}[s^X] = \sum_{k=0}^{\infty} p_k s^k = p_0 + p_1 s + p_2 s^2 + \dots$$

and $G(s) < \infty$ for $|s| < R$.

Then we have

$$\begin{aligned}
G'(s) &= \sum_{k=0}^{\infty} k p_k s^{k-1} = \mathbb{E}[X s^{X-1}] \\
G'(1) &= \mathbb{E}[X] \\
G''(s) &= \sum_{k=0}^{\infty} k(k-1) p_k s^{k-2} = \mathbb{E}[X(X-1) s^{X-2}] \\
G''(1) &= \mathbb{E}[X(X-1)] = \mathbb{E}[X^2 - X] = \mathbb{E}[X^2] - \mathbb{E}[X] \\
\mathbb{E}[X^2] &= G''(1) + G'(1) \\
\text{var}(X) &= G''(1) + G'(1) - [G'(1)]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2
\end{aligned}$$

$$G(0) = p_0, G'(0) = p_1, \frac{G''(0)}{2} = p_2.$$

Let X, Y be independent nonnegative, integer-value random variables.

$$\begin{aligned}
T &= X + Y \\
\mathbb{E}[s^T] &= \mathbb{E}[s^{X+Y}] = \mathbb{E}[s^X s^Y] = \mathbb{E}[s^X] \mathbb{E}[s^Y]
\end{aligned}$$

Example 1.10. Let X_1, \dots, X_n be i.i.d. Bernoulli random variable.

$$\begin{aligned}
T &= X_1 + \dots + X_n \\
\mathbb{E}[s^T] &= \mathbb{E}[s^{X_1 + \dots + X_n}] = (\mathbb{E}[s^{X_1}])^n = (1 - p + ps)^n \\
\mathbb{E}[s^{X_1}] &= s^0(1 - p) + sp
\end{aligned}$$

Let $X_n \sim \text{Binomial}(n, p_n)$, $p_n \rightarrow 0$, $np_n \rightarrow \lambda$, $n \rightarrow \infty$.

$$\begin{aligned}
G_n(s) &= \mathbb{E}[s^{X_n}] \\
&= (1 - p_n + p_n s)^n \\
&= \left(1 - \frac{np_n}{n} + \frac{np_n s}{n}\right)^n \\
&= \left(1 - \frac{np_n(1-s)}{n}\right)^n \rightarrow e^{-\lambda(1-s)}
\end{aligned}$$

as $n \rightarrow \infty$.

$X \sim \text{Poisson}(\lambda)$,

$$G(s) = \mathbb{E}(s^X) = \sum_{k=0}^{\infty} s^k P(X = k) = \sum_{k=0}^{\infty} s^k \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)}.$$

1.4 Cumulative Distribution Function (c.d.f.)

Definition 1.11. Given a random variable X , its *cumulative distribution function (c.d.f.)* is defined as

$$F(t) := P(X \leq t), -\infty < t < \infty.$$

Given a Borel set A , we have

$$F(A) = P(X \in A)$$

For example, $P(X \in (a, b]) = F(b) - F(a)$.

Definition 1.12 (Convergence in distribution). Let X_n be a sequence of random variables, X be a random variable. Let F_n be the cdf of X_n and F be the cdf of X . We can X_n converges to X in distribution (written as $X_n \xrightarrow{D} X$, or $X_n \rightarrow X$), if

$$\begin{aligned} F_n(t) &\rightarrow F(t), \forall t \in \mathcal{C}(F) \text{ (the continuous domain of } F) \text{] or} \\ \mathbb{E}[h(X_n)] &\rightarrow \mathbb{E}[h(X)], \forall \text{ bounded continuous function of } h \end{aligned}$$

Definition 1.13. We say X_n converges to X in (total) variation if $d_{TV}(X_n, X) \rightarrow 0$ as $n \rightarrow \infty$.

Example 1.14. Let X_n be constant random variable $1/n$ and $X = 0$. For every n , we have

$$d_{TV}(X_n, X) = \sup_A |F_n(A) - F_X(A)|$$

and $P(X_n = 0) = 0$, $P(X = 0) = 1$, so X_n does not converge to X in variation.

Given that $\mathcal{C}(F_X) = (-\infty, 0) \cup (0, \infty)$, we have for every $t \in \mathcal{C}(F_X)$, and for all large n ,

$$\begin{cases} F_n(t) = 1, & \text{if } t \in (0, \infty) \\ F_n(t) = 0, & \text{if } t \in (-\infty, 0) \end{cases}$$

Hence, $F_n(t)$ converges to $F_X(t)$ for every $t \in \mathcal{C}(F)$, so $X_n \xrightarrow{D} X$.

Example 1.15. If you have an n -sided die labelled $1/n, 2/n, \dots, n/n$. Then notice that

$$X_n \xrightarrow{D} U \sim \text{Uniform}(0, 1)$$

because if we consider any $t \in (0, 1)$, $F_U(t) = t$, and $F_n(t) = \frac{k}{n}$ where $(k-1)/n < t \leq k/n$. As $n \rightarrow \infty$, k/n converges to t .

Again X_n does not converge to X in variation. Let Q be the set of rational numbers. $P(X \in Q) = 0$ because Q has measure zero, but $P(X_n \in Q) = 1$. Hence $d_{TV}(X_n, X) = 1$ for every n .

1.4.1 Geometric Distribution to Exponential, the Memoryless variables

Let $T_n \sim \text{Geo}(p_n)$, then $\Pr[T_n = k] = (1 - p_n)^{k-1} p_n$, $k = 1, 2, \dots$, $\Pr[T_n > k] = (1 - p_n)^k$, and $\Pr[T_n > k + j | T_n > k] = \Pr[T_n > j]$. Also, $\mathbb{E}[T_n] = 1/p_n$.

And let $X \sim \text{Exp}(\lambda)$, $f_x(t) = \lambda e^{-\lambda t}$, $t \geq 0$, $\Pr[X > t] = e^{-\lambda t}$, $\Pr[X > t + s | X > t] = \Pr[X > s]$, $\mathbb{E}[X] = 1/\lambda$.

We will show

$$\frac{T_n}{n} \xrightarrow{D} X \sim \text{Exp}(\lambda).$$

First, let F_n be the c.d.f. of T_n and F_X be the c.d.f. of X . We need to show that $F_n(t) \rightarrow F_X(t)$ for all $t \in \mathcal{C}(X)$.

Proof.

$$\begin{aligned} 1 - F_n(t) &= \Pr\left[\frac{T_n}{n} > t\right] = \Pr[T_n > nt] = \Pr[T_n > \lfloor nt \rfloor] \\ &= \left(1 - \frac{np_n}{n}\right)^{\lfloor nt \rfloor} = \left(1 - \frac{\lambda_n}{n}\right)^{\lfloor nt \rfloor} \rightarrow e^{-\lambda t} \end{aligned}$$

where $\lambda_n := np_n \rightarrow \lambda$ as $n \rightarrow \infty$, and the convergence to $e^{-\lambda t}$ is by squeeze theorem. \square

1.5 Point Process

Consider $N \sim \text{Poisson}(\lambda)$ and let X_1, \dots, X_N be i.i.d. Bernoulli(p). Define $Y = \sum_{i=1}^N X_i$. If for each of its count of N , it has p chances to be 1 and $(1 - p)$ to be 0, then we can split N into two Poisson distributions

$$\begin{aligned} Y &\sim \text{Poisson}(\lambda p) \\ Z &\sim \text{Poisson}(\lambda(1 - p)) \end{aligned}$$

where $Z := N - Y$ and we have $Z \perp Y$ (seen that in homework 1).

Definition 1.16. A *point process* on $[0, \infty)$ is a mapping, assigning each Borel set $J \subseteq [0, \infty)$, a nonnegative extended integer valued r.v. $N(J) = N_J$, so that if J_1, J_2, \dots , are disjoint, then

$$N(\cup_i J_i) = \sum_i N(J_i)$$

A counting process associated with N (family of random variables), $N(t) = N_t$ for $t \geq 0$ where $N(t) = N((0, t])$ for $t > 0$. By convention, the sample paths are right continuous.

Definition 1.17. A *Poisson point process* with intensity $\lambda > 0$ is a point process with:

- a) If J_1, J_2, \dots , are nonoverlapping intervals, then $N(J_1), N(J_2), \dots$, are independent.
- b) $N(J) \sim \text{Poisson}(\lambda|J|)$ where J is the length of the interval J .

Given a Poisson Point Process above, let $0 = T_0 < T_1 \leq T_2 \leq T_3 \leq \dots$ be the time i^{th} customer arrives and $\tau_n = T_n - T_{n-1}$. Then τ_1, τ_2, \dots , are i.i.d. $\exp(\lambda)$.

Example 1.18. Let $N(t)$ be the number of customers arriving during $(0, t]$ and $N \sim \text{Poisson}(5)$. The probability of 0 arrivals up to time 2 is

$$\Pr[N(2) = 0] = e^{-5(2)} = e^{-10}$$

While the probability of k arrivals up up time 2 is

$$\Pr[N(2) = k] = \frac{e^{-10} 10^k}{k!}.$$

Consider

$$\begin{aligned}
& \{N(5) = 7 | N(2) = 1\} \\
& \{N((2, 5]) = 6 | N(2) = 1\} \\
& \Pr[N(5) - N(2) = 6 | N(2) = 1] \\
& = \Pr[N(5) - N(2) = 6] \\
& = \Pr[N((2, 5]) = 6] \\
& = \Pr[N(3) = 6]
\end{aligned}$$

We can also consider

$$\Pr[T_2 > 5.8 | T_1 = 3.7] = \Pr[\tau_2 > 2.1 | \tau_1 = 3.7] = e^{-\lambda(2.1)}$$

If you look at the store a 100 min, when will the next customer arrive?

We expect $\frac{1}{\lambda} = \frac{1}{5}\text{hr} = 12\text{min}$.

$$\begin{aligned}
\Pr[X_1 > t] &= \Pr[N(t) = 0] = e^{-\lambda t}, t \geq 0 \\
\Pr[X_2 > t | X_1 = s] &= \Pr[N((s, s + t]) = 0 | X_1 = s] \\
&= \Pr[N((s, s + t]) = 0] \\
&= e^{-\lambda t}
\end{aligned}$$

1.6 Bernoulli and Poisson

Let X_1, X_2, \dots , be Bernoulli Process with $p \in (0, 1)$.

Question:

- a) Is $\Pr[X_n = k | T = n]$ equal $\Pr[X_T = k | T = n]$? **Yes.**
Let $A = \{w \in \Omega : X_n(w) = k\}$, $B = \{w \in \Omega : T(w) = n\}$, $C = \{w \in \Omega : X_{T(w)}(w) = k\}$
and $A \cap B = \{w \in \Omega : X_n(w) = k, T(w) = n\}$, $C \cap B = \{w \in \Omega : X_{T(w)}(w) = k, T(w) = n\}$, which implies $\Pr[A \cap B] / \Pr[B] = \Pr[C \cap B] / \Pr[B]$
- b) Is $\Pr[X_n = k | T = n]$ equal to $\Pr[X_n = k]$? **No.** e.g. $T := \min\{n : X_n = 1\}$, and $\Pr[X_n = 1 | T = n] = 1$, $\Pr[X_n = 1] = p$.
e.g. $X_i \sim \text{Exp}(\lambda)$ where X_1, X_2, \dots , are event times.

$$\begin{aligned}
\Pr[X_2 > t | X_1 = s] &= \Pr[N((s, s + t]) = 0 | X_1 = s] \\
&= \Pr[N(s, s + t] = 0] \text{ by independent increment} \\
&= \Pr[N(X_1, X_1 + t] = 0 | X_1 = s]
\end{aligned}$$

But then let $T := \min\{r : N(r, r + t] = 10\}$. We have

$$\Pr[N(T, T + t) = 0 | T = 3.87] = 0, \Pr[N(3.87, 3.87 + t] = 0] = e^{-\lambda t}$$

Definition 1.19. Let $0 = T_0 < T_1 = \tau_1 \leq T_2 = \tau_1 + \tau_2 \leq \dots$ be the *occurrence times* of a Poisson process which are the successive times $N(t)$ jumps. Let τ_1, τ_2, \dots be the *interoccurrence time*, where $\tau_i := T_i - T_{i-1}$.

Theorem 1.20 (Interoccurrence Time Theorem).

- (A) Interoccurrence times τ_1, τ_2, \dots , of a Poisson process with rate λ are i.i.d. $\text{Exp}(\lambda)$
- (B) Let Y_1, Y_2, \dots , be i.i.d. $\text{Exp}(\lambda)$.

$$N(t) := \max\{n : \sum_{i=1}^n Y_i \leq t\} \implies \{N(t)\}_{t \geq 0} \text{ is a Poisson counting process with rate } \lambda > 0$$

Example 1.21. Consider Bernoulli processes $\{X_k^m\}_{k \in \frac{\mathbb{N}}{m}}$ with parameter $p_m \in (0, 1)$. Then $\tau_1^m = T_1^m = \min\{n \in \frac{\mathbb{N}}{m} : X_n^m = 1\}$. Then $m\tau_1^m \sim \text{Geo}(p_m)$. Let $T_2^m = \min\{n > T_1^m : X_n^m = 1\}$ and $\tau_2^m = T_2^m - T_1^m$, then $m\tau_2^m \sim \text{Geo}(p_m)$ as well. Then with the occurrence time T_i , we have a counting process

$$N^m(t_1) \sim \text{Binomial}(\cdot, p_m)$$

Useful later: $\{T_1 \geq t_1, T_2 \geq t_2\} \iff \{N(t_1) \geq 1, N(t_2) \geq 2\}$.

Theorem 1.22 (The law of small numbers for Bernoulli Process). Let $\{X_r^m\}_{r \in \mathbb{N}/m}$ be a Bernoulli Process with parameter p_m indexed by multipliers of \mathbb{N}/m . Let $N^m(t)$ be the corresponding counting process. If $mp_m \rightarrow \lambda > 0$, then the counting process N^m converges in distribution to the counting process of a Poisson process with rate $\lambda > 0$ in the following sense:

$$\forall n, 0 = t_0 < t_1 < \dots < t_n, (N^m(t_1), \dots, N^m(t_n)) \xrightarrow{D} (N(t_1), \dots, N(t_n))$$

Proof of Interoccurrence Time Theorem.

- a) We showed in the previous section that for a geometric r.v. with p_n with $np_n \rightarrow \lambda$. $T_n/n \xrightarrow{D} \text{Exp}(\lambda)$. And we have seen that the interoccurrence times of Bernoulli $\{X_k^m\}_{k \in \mathbb{N}/n}$ are geometric, $\Delta_k^m = N^m(t_k) - N^m(t_{k-1}) \sim \text{Binomial}(m(t_k - t_{k-1}) \pm 1, p_m)$ where \pm considers the rounding of $m(t_k - t_{k-1})$. And this converges in distribution to $\Delta_k \sim \text{Poisson}(\lambda(t_k - t_{k-1}))$. Thus the occurrence time of $N^m(t)$ converges to $N(t)$ in distribution. Thus, the interoccurrence time of X_k^m , which is the interoccurrence time of $N^m(t)$, converging to $\text{Exp}(\lambda)$ implies that the interoccurrence time of $N(t)$ converges to $\text{Exp}(\lambda)$.
- b) With a Poisson process with rate λ , and let τ_i be its interoccurrence times, and we know $\tau_i \stackrel{iid}{\sim} \text{Exp}(\lambda)$. Let Y_i be another sequence of i.i.d. exponentials with λ . Then since τ_i and Y_i have the same joint distribution, we also have

$$\left(\tau_1, \tau_1 + \tau_2, \dots, \sum_{i=1}^n \tau_i \right) \stackrel{D}{=} \left(Y_1, Y_1 + Y_2, \dots, \sum_{i=1}^n Y_i \right)$$

But $(Y_1, Y_1 + Y_2, \dots, \sum_{i=1}^n Y_i)$ determines the joint distribution of the occurrence time of $N(t)$. That is, the occurrence times of $N(t)$ are the occurrence times of a Poisson distribution. So $N(t)$ is Poisson.

□

Given B), now we can simulate Poisson with $U_i \stackrel{D}{\sim} \text{Uniform}([0, 1])$ and have $\tau_i = -\frac{1}{\lambda} \log(1 - U_i)$. However, if the actual $\lambda > \mu$ and we simulate with μ , then we have

$$\tilde{\tau}_i = -\frac{1}{\mu} \log(1 - U_i) \stackrel{D}{=} \frac{\lambda}{\mu} \tau_k$$

Theorem 1.23 (Generalized Thinning Theorem). Let $N \sim \text{Poisson}(\lambda)$, X_i be iid r.v. with $\Pr[X_i = k] = p_k, k = 1, \dots, m$ and $\sum_{i=1}^m p_k = 1$. And N is independent from X_i for all i . Let $N_k = \sum_{j=1}^N \mathbb{1}_{\{X_j=k\}}$.
e.g:

$$\begin{array}{cccccc} m = 3 & x_1 & x_2 & x_3 & x_4 & x_5 \\ N = 5 & 2 & 3 & 3 & 1 & 2 \end{array}$$

then $N_1 = 1, N_2 = 2, N_3 = 2, N_1 + N_2 + N_3 = N$.

We have that N_1, \dots, N_m are independent Poisson r.v. with $\mathbb{E}[N_k] = \lambda p_k$. (You can consider this as splitting a Poisson process into m different ones with probability p_k .)

And we have

$$\begin{aligned} \Pr[N_1 = j_1, N_2 = j_2, \dots, N_m = j_m] &= \Pr[N = j_1 + \dots + j_m, N_1 = j_1, \dots, N_m = j_m] \\ &= \underbrace{\Pr[N = j_1 + \dots + j_m]}_{\text{Poisson}} \underbrace{\Pr[N_1 = j_1, \dots, N_m = j_m | N = \sum_{i=1}^m j_i]}_{\text{multinomial}} \\ &= \frac{e^{-\lambda} \lambda^{j_1 + \dots + j_m}}{(j_1 + \dots + j_m)!} \binom{j_1 + \dots + j_m}{j_1, \dots, j_m} p_1^{j_1} \dots p_m^{j_m} \\ &= \prod_{i=1}^m \frac{e^{-\lambda p_i} (p_i \lambda)^{j_i}}{j_i!} \end{aligned}$$

Second Construction Let m_1, m_2, \dots be iid $\text{Poisson}(\lambda)$. Let U_1, U_2, \dots be iid $\text{Uniform}(0, 1)$ such that (m_1, m_2, \dots) independes (U_1, U_2, \dots) . Put points at U_1, \dots, U_{m_1} if $m_1 > 0$. Put points at $1 + U_{m_1+1}, \dots, 1 + U_{m_2}$ if $M_2 > 0$ and so on.

Claim 1.23.1. Above points form a Poisson point process (THM 7 of UChicago Notes).

Proof. $0 = t_1 < t_1 < \dots < t_n = 1, J_k = (t_{k-1}, t_k] \implies p_k = t_k - t_{k-1}. N(J_1), \dots, N(J_n)$ independent Poisson $\mathbb{E}[N(J_k)] = \lambda p_k = \lambda |J_k|$. \square

Definition 1.24. Poisson point process on \mathbb{R}^k with mean measure Λ is a point process on \mathbb{R}^k with

1. J_1, J_2, \dots disjoint Borel sets in \mathbb{R}^k ; $N(J_1), N(J_2), \dots$ are independent.
2. $N(J_k) \sim \text{Poisson}(\Lambda(J_k))$

Proposition 1.25. To show a point process is a Poisson point process, it suffices to verify the conditions above for rectangles J, J_i with sides parallel to the coordinate axes.

Example 1.26. Let T_i be the occurrence times of a Poisson process on $[0, \infty)$ with rate λ . Let S_j be the iid rv with CDF F . S_j, T_i are indep. Then we have $J = [t_1, t_2] \times [s_1, s_2]$. So $N(J) = \lambda(t_2 - t_1)(s_2 - s_1)$, where $J' \cap J = \emptyset$ implies $N(J)$ independent $N(J')$.

For a Poisson Point Process on \mathbb{R} with rate $\lambda > 0$, then given $t > 0$, we have

$$\begin{aligned}\Pr[N(0, t] = 0] &= e^{-\lambda t} \\ \Pr[N(-t, 0] = 0] &= e^{-\lambda t} \\ \Pr[N(-t, t] = 0] &= e^{-2\lambda t}\end{aligned}$$

Given 2 Poisson Processes on $[0, \infty)$ with $N \sim \text{Poisson}(\lambda)$, $M \sim \text{Poisson}(\mu)$, $\lambda > \mu$, how can we comply them so $N(J) \geq M(J)$ for every Borel set J ?

1. Superposition: Consider M as above and $L \sim \text{Poisson}(\lambda - \mu)$, which are independent, then take the superposition (a process made of all success of M, L) so we get another $\text{Poisson}(\lambda)$.
2. Decomposition: With the N above, for each success of N , split it to M with probability μ/λ , and L with $(1 - \mu/\lambda)$, then M and L are independent Poisson Processes and M is what's required.

Consider N, M with the distributions above, let T_1 be the time of first success of N , then what's the probability that $M(T_1) = k$? If we directly compute it, it will be

$$\Pr[M(T_1) = k] = \int_0^\infty \Pr[M(T_1) = k | T_1 = s] \underbrace{\lambda e^{-\lambda s}}_{\Pr[T_1=s]} ds$$

which is not that easy to compute. But we can consider $N + M \sim \text{Poisson}(\lambda + \mu)$. And split its success to N, M with probability $\frac{\lambda}{\lambda + \mu}$ and $\frac{\mu}{\lambda + \mu}$ respectively. Then T_1 is the time when a success is splitted to N the first time. That is, $M(T_1 = k)$ can be considered as a geometric process with k failure and one success, so

$$\Pr[M(T_1) = k] = \left(\frac{\mu}{\lambda + \mu}\right)^k \left(\frac{\lambda}{\lambda + \mu}\right)$$

Let $\{N(t)\}_{t \geq 0}$ be a counting process on $[0, \infty)$. Prove or disprove: If $N(t) \sim \text{Poisson}(\lambda t)$ for all $t > 0$, then N is a Poisson Process.

Let T_i be the occurrence times and τ_i be the interoccurrence times as before. Then $T_n = \tau_1 + \dots + \tau_n$. If τ_i are independent $\text{Exp}(\lambda)$, we know $T_n \sim \text{Erlang}(n, \lambda)$, so $\mathbb{E}[T_n] = n/\lambda$ and

$$F_n(t) = \Pr[T_n \leq t] = \Pr[N(t) \geq n] = \sum_{k=n}^{\infty} \frac{e^{-\lambda t} (\lambda t)^k}{k!} = 1 - \sum_{k=0}^{n-1} \frac{e^{-\lambda t} (\lambda t)^k}{k!}$$

so if T_1, T_2, \dots , have the "right" distribution, then $N(t)$ will be $\text{Poisson}(\lambda t)$. What if we don't have the independence? Consider $T_i := F_i^{-1}(U)$ where F_i is the cdf of $\text{Erlang}(i, \lambda)$ and $U \sim \text{Uniform}(0, 1)$. Then it's not hard to see that each $T_i \sim \text{Erlang}(i, \lambda)$, however, once T_1 is given, we can compute U_1 and hence all T_2, T_3, \dots are know, so the process with T_i being the occurrence time is not a Poisson.

limits of expectation and expectation of limits

Theorem 1.27 (Monotone Convergence Theorem). Let $\{X_n\}_{n \geq 1}$ be a sequence of random variables such that for all $n \geq 1$,

$$0 \leq X_n \leq X_{n+1}, \text{ Probably a.s.,}$$

then

$$\mathbb{E}[\lim_{n \rightarrow \infty} X_n] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n].$$

Theorem 1.28 (Dominated Convergence Theorem). Let $\{X_n\}_{n \geq 1}$ be a sequence of random variable such that for all ω outside a set \mathcal{N} of null probability there exists $\lim_{n \rightarrow \infty} X_n(\omega)$ and such that for all $n \geq 1$

$$|X_n| \leq Y, \text{ Probably a.s.,}$$

where Y is some integrable random variable. Then

$$\mathbb{E}[\lim_{n \rightarrow \infty} X_n] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n].$$

Example 1.29 ("Counter Example"). Suppose we are rolling a fair dice independently. Every time we get 6, we lose all the money, otherwise, we double the current amount. Starting with $X_0 = 100$, we have

$$X_n = \begin{cases} 100 * 2^n, & \text{with prob } (5/6)^n \\ 0, & \text{with prob } 1 - (5/6)^n \end{cases}$$

$$\begin{aligned} \mathbb{E}[X_n] &= 100 * (5/3)^n \\ \lim_{n \rightarrow \infty} \mathbb{E}[X_n] &= \infty \\ \mathbb{E}[\lim_{n \rightarrow \infty} X_n] &= 0 \end{aligned}$$

where the last inequality is by $\lim_{n \rightarrow \infty} \Pr[X_n > 0] = 0$ and $\lim_{n \rightarrow \infty} \Pr[X_n = 0] = 1$, so $X_n \rightarrow 0$ almost surely.

Let N be a Poisson on $[0, \infty)$ with rate λ . Let $T \geq 0$ be a r.v. such that N, T are independent. If we know the distribution of $N(T)$, can we determine the distribution of T ? First consider the *probability generating function* (p.g.f.) of a Poisson $X \sim \text{Poisson}(\lambda)$, we have

$$G(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} s^k \Pr[X = k] = \sum_{k=0}^{\infty} s^k \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)}$$

Or let x be nonnegative, integer-valued r.v. the *Laplace-Stieltjes Transformation* of X is

$$L(s) = \mathbb{E}[e^{-sX}] = \int_0^{\infty} e^{-st} dF(t) = \int_0^{\infty} e^{-st} F(dt)$$

note this formula prevent us from worrying about the continuity of X by $F(t)$.

Recall the moment generating function (m.g.f.) $m_X(\theta) = \mathbb{E}[e^{\theta X}]$. We give some examples,

Example 1.30.

1. When $\Pr[T = t] = 1$, we have $\mathbb{E}[e^{-sT}] = e^{-st}$.

2. When $T \sim \text{Bernoulli}(p)$,

$$L(s) = \mathbb{E}[e^{-sT}] = (1-p) * 1 + p * e^{-s} = \int_{[0,\infty)} e^{-st} dF(t)$$

3. $T \sim \text{Binomial}(n, p)$. $T = X_1 + \dots + X_n$, where X_i are i.i.d. Bernoulli.

$$\begin{aligned} L(S) &= \mathbb{E}[e^{-sT}] \\ &= \int_{[0,\infty)} e^{-st} dF(t) \\ &= \mathbb{E}[e^{-s(X_1 + \dots + X_n)}] \\ &= \mathbb{E}[e^{-sX_1} \dots e^{-sX_n}] \\ &= \mathbb{E}[e^{-sX_1}] \dots \mathbb{E}[e^{-sX_n}] \\ &= (1-p + pe^{-s})^n \end{aligned}$$

4. Let $X \sim \text{Exp}(\lambda)$, we have

$$\mathbb{E}[e^{-sX}] = \int_0^\infty e^{-st} \lambda e^{-\lambda t} dt = \frac{\lambda}{s + \lambda}. \quad (\text{L.S. of Exp})$$

Lemma 1.31. Given a $N(T) \sim \text{Poisson}(\lambda)$, and N being independent from T , we have $L_T(s) = G(1 - s/\lambda)$.

Proof.

$$\begin{aligned} G(z) &= \mathbb{E}[z^{N(T)}] \\ &= \mathbb{E}[\mathbb{E}[z^{N(T)} | T]] \\ &= \mathbb{E}[e^{-\lambda T(1-z)}] \\ &= L(\lambda(1-z)) \end{aligned}$$

where the second last inequality is by

$$G(z) = \mathbb{E}[z^{N(T)}] = \sum_{k=0}^{\infty} z^k \frac{e^{-\lambda T} (\lambda T)^k}{k!} = e^{-\lambda T(1-z)}.$$

And then let $s = \lambda(1 - z)$, we are done. □

Thus, when $N(T) \sim \text{Poisson}(\lambda T)$,

$$L(s) = G(1 - s/\lambda) = e^{-\lambda T(1 - (1 - s/\lambda))} = e^{-st}$$

so $\Pr[T = t] = 1$.

Theorem 1.32 (Not gonna prove). Like p.g.f. and m.g.f., $L(s)$ uniquely corresponds to a random distribution.

Example 1.33. Let $\Pr[N(T) = k] = \rho^k(1 - \rho)$, $k = 0, 1, \dots$. Then

$$\begin{aligned} G(z) &= \mathbb{E}[z^{N(T)}] = \sum_{k=0}^{\infty} z^k \rho^k (1 - \rho) = \frac{1 - \rho}{1 - \rho z}. \\ L(s) &= \mathbb{E}[e^{-sT}] = G(1 - s/\lambda) = \frac{1 - \rho}{1 - \rho(1 - s/\lambda)} \\ &= \frac{1 - \rho}{1 - \rho + \rho s/\lambda} = \frac{\frac{\lambda}{\rho}(1 - \rho)}{\frac{\lambda}{\rho}(1 - \rho) + s} \end{aligned}$$

which shows that $T \sim \text{Exp}(\frac{\lambda}{\rho}(1 - \rho))$ by **(L.S. of Exp)**.

2 Markov-Chain

Let X_0, X_1, \dots be discrete-time stochastic processes and let the state space be countable.

$$\Pr[X_0 = i_0, \dots, X_n = i_n], \forall n, i_0, \dots, i_n \in \text{state space.}$$

1. Markov Property:

$$\Pr[\underbrace{X_{n+1} = j}_{\text{future}} | \underbrace{X_n = i_n}_{\text{present}}, \underbrace{\dots, X_0 = i_0}_{\text{past}}] = \Pr[X_{n+1} = j | X_n = i_n]$$

2. Time Homogeneity:

$$\Pr[X_{n+1} = j | X_n = i] = \Pr[X_1 = j | X_0 = i] = \Pr(i, j)$$

Definition 2.1. X_0, X_1, \dots is a *discrete-time Markov chain (DTMC)* if X_0, X_1, \dots has the two properties above.

Example 2.2. Let X_0, X_1, \dots be an independent Bernoulli process with parameter p . Then the state space is $\{0, 1\}$.

$$\begin{aligned} \Pr[X_{n+1} = j | X_n = i_n, \dots, X_0 = i_0] &= \Pr[X_{n+1} = j] \\ \Pr[X_{n+1} = j | X_n = i_n] &= \Pr[X_{n+1} = j] \\ \Pr[X_{n+1} = j | X_n = j] &= \Pr(i, j). \end{aligned}$$

This forms a really special DTMC, basically every r.v. are i.i.d.. Its transition matrix looks like

$$P = \begin{bmatrix} 1-p & p & \dots \\ 1-p & p & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

where the rows represent the "from" and columns represent the "to". That is, $[P]_{ij} = \Pr(i, j)$.

Example 2.3. Let $X_0, X_1, \dots \sim \text{Bernoulli}(p), p \in (0, 1)$. $Y_n = X_n + X_{n+1} \in \{0, 1, 2\}$. Is Y_0, Y_1, \dots a Markov Chain? No.

$$\Pr[Y_2 = 0 | Y_1 = 1, Y_0 = 0] = 0$$

$$\Pr[Y_2 = 0 | Y_1 = 1, Y_0 = 2] = 1 - p$$

because $Y_0 = 0, Y_1 = 1$ implies that $X_2 = 1, X_0 = X_1 = 0$, first probability is the probability that $X_3 = -1$ and the second one is the probability that $X_3 = 0$.

What can we add to make it a DTMC?

Acquire more information. Let $Z_n = (X_n, Y_n)$, then we consider

$$\Pr[Z_{n+1} = (j_1, j_2) | Z_n = (i_1, i_2), Z_{n-1} = (k_{n-1}, \ell_{n-1}), \dots, Z_0 = (k_0, \ell_0)]$$

And the transition matrix is

	(0,0)	(0,1)	(1,0)	(1,1)
(0,0)	1-p	p	0	0
(0,1)	0	0	1-p	p
(1,0)	1-p	p	0	0
(1,1)	0	0	1-p	p

$M/M/1$ Queue Consider an $M/M/1$ queue, which is the queue with customers arriving according to $\text{Poisson}(\lambda)$, service time following i.i.d. $\exp(\mu)$ with 1 server. The model records the number of customers whenever a process (arrival or service) is done. Note that this process or a point from the Poisson process does not have to "happen". You can treat all events as a $\text{Poisson}(\lambda + \mu)$. For each point, there is a chance we have a service done, and another chance the we have an arrival. However, since this is an event, when there is 0 customer in the system, next point can still be a departure point, but the number of customers will stay at 0 instead of going to -1 . When there are at least one customer in the system, the server actually serves the customer and make the number of customers minus 1.

For example, if we have $X_0 = 0$ and the next event is finishing a service, $X_1 = 0$, if it's a customer arrival, $X_1 = 1$. This model is also called the birth and death model, basically we add one when we have a birth and minus one when we have a death. Since the moment starts, we can only have "deaths" (or departures) until the first arrival. That is, given $X_n = 0$, the probability that $X_{n+1} = 0$ is the probability that

$$\Pr[D < A] = \frac{\mu}{\lambda + \mu}$$

where $D \sim \exp(\mu)$ is the service time and $A \sim \exp(\lambda)$ is the interoccurrence time of $\text{Poisson}(\lambda)$ (i.e. the arrival time). Similarly, given $X_n = 0$, the probability that $X_{n+1} = 1$ is the probability that the customer arrives before the service time. So the transition matrix looks like

$$\begin{bmatrix} \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} & 0 & \dots & \dots \\ \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & 0 & \dots \\ 0 & \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & 0 & \dots \end{bmatrix}$$

where rows and columns are from 0 to infinity.

We can also consider $X_n :=$ number of customers in the system just before n -th arrival. For example, given $X_n = 0$, the probability $X_{n+1} = 0$ is $\frac{\mu}{\lambda+\mu}$, because $X_n = 0$, so between n -th and $n + 1$ th arrival, there is at most one customer in the system, and we have the probability $\frac{\mu}{\lambda+\mu}$ to finish the service before $n + 1$ -th arrival, otherwise, with probability $\frac{\lambda}{\mu+\lambda}$, we still have a customer in the system when $n + 1$ -th customer arrives.

Another way of considering this is treating the arrivals as a geometric distribution with $\frac{\lambda}{\lambda+\mu}$ success rate. For example, if $X_n = 1$. That means between n and $n + 1$ arrivals, there are 2 customers in the system, and we do the geometric experiment. The probability that there is no customer in the system when $n + 1$ th customer arrives is the probability we "fail" at least twice before the "success". Similarly, the probability that there is one customer in the system when $n + 1$ th customer arrives is the prob that we "fail" exactly once before the first success, and so on. So the transition matrix looks like:

$$\begin{bmatrix} \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} & 0 & \dots & \dots \\ \left(\frac{\mu}{\lambda+\mu}\right)^2 & \frac{\mu\lambda}{(\lambda+\mu)^2} & \frac{\lambda}{\lambda+\mu} & 0 & \dots \\ \left(\frac{\mu}{\lambda+\mu}\right)^3 & \left(\frac{\mu}{\lambda+\mu}\right)^2 \frac{\lambda}{(\lambda+\mu)} & \frac{\mu\lambda}{(\lambda+\mu)^2} & \frac{\lambda}{\mu+\lambda} & \dots \end{bmatrix}$$

$M/M/1/3$ Queue Consider the $M/M/1/3$ queue where the 3 means the capacity of the system. Let $Y_n :=$ number of customers in the system just after the n -th departure, so now the state space

is $\{0, 1, 2\}$. Then let's say $Y_n = 0$, then the probability $Y_{n+1} = 0$ is the probability that there is an arrival between n -th and $n + 1$ -th departures. In other words, for $n + 1$ -th departure to happen, there has to be an arrival, so the probability is actually the probability that the $(n + 1)$ -th departure happen before any arrivals except for the necessary one, which is $\frac{\mu}{\lambda + \mu}$, similar to other cases. So the transition matrix looks like:

$$\begin{bmatrix} \frac{\mu}{\mu + \lambda} & \frac{\lambda\mu}{(\mu + \lambda)^2} & \left(\frac{\lambda}{\lambda + \mu}\right)^2 \\ \frac{\mu}{\mu + \lambda} & \frac{\lambda\mu}{(\mu + \lambda)^2} & \left(\frac{\lambda}{\lambda + \mu}\right)^2 \\ 0 & \frac{\mu}{\mu + \lambda} & \frac{\lambda}{\mu + \lambda} \end{bmatrix}$$

2.1 Transition Matrix

Definition 2.4. A matrix P is a *stochastic matrix* if $P(i, j) \geq 0$, and $\sum_{j \in S} P(i, j) = 1$. It is called a *doubly stochastic matrix* if it is a stochastic matrix and $\sum_{i \in S} P(i, j) = 1$. It is called a *substochastic matrix* if $P(i, j) \geq 0$ and $\sum_{j \in S} P(i, j) \leq 1$.

Given $S = \{0, 1, 2\}$, and a transition matrix

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{bmatrix}. \quad (2.1)$$

We have the transition plot of the above matrix,

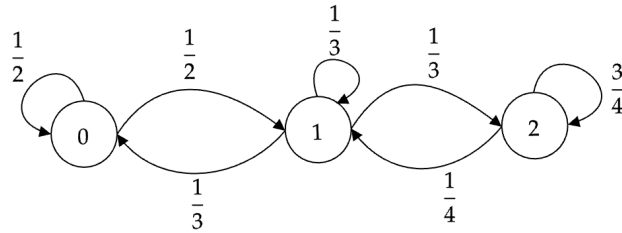


Figure 1: Transition Plot of P

Lemma 2.5. $\Pr[A, B, C, D] = \Pr[A] \Pr[B|A] \Pr[C|AB] \Pr[D|ABC]$

Example 2.6. Given X_0, X_1, \dots , we have

$$\begin{aligned} & \Pr[X_0 = i_0, \dots, X_n = i_n] \\ &= \Pr[X_0 = i_0] \Pr[X_1 = i_1 | X_0 = i_0] \dots \Pr[X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0] \\ &= \Pr[X_0 = i_0] P(i_0, i_1) P(i_1, i_2) \dots P(i_{n-1}, i_n) \end{aligned}$$

Definition 2.7. We use *measure distributions* on S that are functions from S to \mathbb{R} to describe a distribution of a random variable. We use α, β, μ, π to describe row vectors, and use f, g, h to describe column vectors. For example,

$$X_0 \sim \alpha = (1/3, 1/2, 1/6)$$

and a function

$$f = \begin{pmatrix} f(0) \\ f(1) \\ f(2) \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 3 \end{pmatrix},$$

then $\alpha f = \mathbb{E}[f(X_0)] \in \mathbb{R}$.

Example 2.8.

$$\begin{aligned} \Pr[X_2 = j | X_0 = i] &= \sum_{k \in S} \Pr[X_2 = j, X_1 = k | X_0 = i] \\ &= \sum_{k \in S} P(i, k) P(k, j) \\ &= P^2(i, j). \end{aligned}$$

For our P , we have $P^2(1, 1) = \frac{1}{6} + \frac{1}{9} + \frac{1}{12}$.

Lemma 2.9 (Chapman-Kolmogorov).

$$P^{m+n}(i, j) = \sum_{k \in S} P^m(i, k) P^n(k, j)$$

where $P^{m+n} = P^m P^n$.

Example 2.10. $\Pr[X_4 = 1, X_2 = 0, X_7 = 1 | X_1 = 2] = P(2, 0) P^2(0, 1) P^3(1, 1)$.

Lemma 2.11.

$$X_0 \sim \alpha \implies X_1 \sim \alpha P, \dots, X_n \sim \alpha P^n$$

And

$$\begin{aligned} \Pr[X_1 = j] &= \sum_i \Pr[X_1 = j | X_0 = i] \Pr[X_0 = i] \\ &= \sum_i \alpha(i) P(i, j) \end{aligned}$$

Example 2.12.

$$\Pr[X_4 = 1 | X_5 = 1] = \frac{\Pr[X_4 = 1, X_5 = 1]}{\Pr[X_5 = 1]} = \frac{\Pr[X_5 = 1 | X_4 = 1] \Pr[X_4 = 1]}{\Pr[X_5 = 1]} = \frac{\alpha P^4(1) P(1, 1)}{\alpha P^5(1)}$$

With the properties above, we can let f be a vector and have

$$\begin{aligned} [Pf]_i &= \mathbb{E}[f(X_1) | X_0 = i] \\ [P^n f]_i &= \mathbb{E}[f(X_n) | X_0 = i] \\ \alpha P^n f &= \mathbb{E}[f(X_n)] \end{aligned}$$

Definition 2.13. An *invariant measure* μ is a measure that $\mu = \mu P$. For our matrix P in (2.1), $\mu = (1, 3/2, 2)$ is an invariant measure.

A *stationary distribution* is an invariant measure that sums to 1. For our P in (2.1), $(2/9, 3/9, 4/9)$ is one.

2.2 Communication, Recurrence and Transience

Definition 2.14. We say j is *accessible* from i if $\exists n \geq 0$ such that $P^n(i, j) > 0$.

We say i and j *communicate* ($i \sim j$) if i is accessible from j and vice versa.

We say i is *absorbing* j if $P(i, j) = 1$.

Proposition 2.15. Communication is an equivalent relation being:

- reflective: $i \sim i$, which is always true by letting $n = 0$ and hence $P = I$.
- symmetric: $i \sim j \implies j \sim i$.
- transitive: $i \sim j, j \sim k \implies i \sim k$. (If there exists n with $P^n(i, j) > 0$ and m with $P^m(j, k) > 0$ then $m + n$ takes us from i to k).

Example 2.16. For the following plot, we see that for each state, they only communicate with themselves.

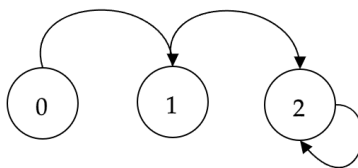


Figure 2: Self Commu States

Definition 2.17. If every two states communicate, then we say this Markov Chain is *irreducible*.

Definition 2.18. The *period* of state i is $d(i)$ defined as the greatest common divisor of $\{n > 0 \mid P^n(i, i) > 0\}$.

Example 2.19. Given the following graph:

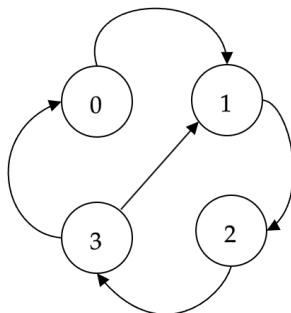


Figure 3: Period 1

Consider $i = 0$, then

$$\{n > 0 | P^n(0, 0) > 0\} = \{4, 7, 10, 13, \dots\} \implies d(0) = 1$$

For the following graph:

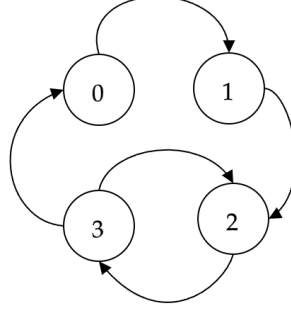


Figure 4: Period 2

Consider $i = 0$, then

$$\{n > 0 | P^n(0, 0) > 0\} = \{4, 6, 8, \dots\} \implies d(0) = 2$$

Proposition 2.20. If i and j communicate, $d(i) = d(j)$.

Proof. We know there exist m and n such that $P^m(i, j) > 0$ and $P^n(j, i) > 0$, so $P^{m+n}(i, i) > 0$, and $m + n$ is a multiplier of $d(i)$. Let ℓ be an integer such that $P^\ell(j, j) > 0$. Then

$$P^{m+n+\ell}(i, i) \geq P^m(i, j)P^\ell(j, j)P^n(j, i) > 0$$

so $m + n + \ell$ is a multiplier of $d(i)$. Hence, we know $m + n + \ell$ is a multiplier of $d(i)$, so ℓ is a multiplier of $d(i)$ which implies $d(j) \geq d(i)$. The argument for $d(i) \geq d(j)$ is similar, so $d(i) = d(j)$. \square

Definition 2.21. T is called a stopping time of $\{T = n\}$ can be determined from X_0, \dots, X_n , i.e.

$$\mathbb{1}_{T=n} = g_n(X_0, \dots, X_n).$$

Example 2.22. $T_x = \inf\{n \geq 0 | X_n = x\}$ is a stopping time. $T_x^k =$ time of k^{th} visit of x is also a stopping time.

Let T be a stopping time, then

$$\begin{aligned} & \Pr[X_{T+1} = i_{m+1}, X_{T+2} = i_{m+2}, \dots, X_{T+n} = i_{m+n} | T = m, X_m = i_m, X_{m-1} = i_{m-1}, \dots, X_0 = i_0] \\ &= P(i_m, i_{m+1}) \dots P(i_{m+n-1}, i_{m+n}) \end{aligned}$$

and since T is a stopping time, $T = m$ is redundant by knowing X_m, \dots, X_0 . This is called *Strong Markov Property*.

Definition 2.23. Let $T_x^1 = T_x = \inf\{n \geq 1 | X_n = x\}$, $T_x^k = \inf\{n \geq T_x^{k-1} | X_n = x\}$, $k = 2, 3, \dots$, and $\Pr[X_0 = x] = 1$.

- State x is *recurrent* if $\Pr_x[T_x < \infty] = 1$.
- State x is *transient* if $\Pr_x[T_x < \infty] < 1$.
- State x is *positive transient* if $\mathbb{E}_x[T_x] < \infty$.
- State x is *null* if x is recurrent and $\mathbb{E}_x[T_x] = \infty$.

Example 2.24. Let $\Pr[X = k] = \frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}$ for $k = 1, 2, \dots$. Then

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \frac{1}{k(k+1)} = \sum_{k=1}^{\infty} \frac{1}{k+1} = \infty$$

$$\Pr[X \leq n] = (1 - \frac{1}{2}) + (\frac{1}{2} - \frac{1}{3}) + \dots + (\frac{1}{n} - \frac{1}{n+1}) = 1 - \frac{1}{n+1}$$

Suppose x is recurrent. How many times will x be revisited is represented

$$N_x = \sum_{k=0}^{\infty} [X_k = x].$$

Suppose state x is transient, by strong Markov property,

$$\Pr[T_x^k < \infty] = \Pr_x[T_x < \infty]^k.$$

Assuming $X_0 = x$, $N_x \sim \text{Geo}(\Pr[T_x = \infty])$. That is, N_x stops (the number will not increase) once we fall into the case X_n never comes to x .

Proposition 2.25. State x is recurrent if and only if $\mathbb{E}_X[N_X] = \infty$.

Proof.

$$\begin{aligned} \mathbb{E}_X[N_X] &= \mathbb{E}_X \sum_{k=0}^{\infty} \mathbb{1}[X_k = x] \\ &= \sum_{k=0}^{\infty} \mathbb{E}_x[\mathbb{1}[X_k = x]] \\ &= \sum_{k=0}^{\infty} \Pr_X[X_k = x] = \sum_{k=0}^{\infty} P^k(x, x) \\ N_X &= 1 + \sum_{k=1}^{\infty} \mathbb{1}[T_x^k < \infty] \end{aligned}$$

$$\begin{aligned}
\mathbb{E}_X[N_X] &= 1 + \sum_{k=1}^{\infty} \mathbb{E}[\mathbb{1}[T_x^k < \infty]] \\
&= 1 + \sum_{k=1}^{\infty} \Pr[T_x^k < \infty] \\
&= 1 + \sum_{k=1}^{\infty} \Pr[T_x < \infty]^k \\
&= \begin{cases} \infty, & \text{if recurrent.} \\ \frac{1}{1 - \Pr[T_x < \infty]}, & \text{transient.} \end{cases}
\end{aligned}$$

□

Proposition 2.26. If x is recurrent and x, y communicate, then y is recurrent.

Proof. There exists k such that $P^k(x, y) > 0$, and there exists ℓ such that $P^\ell(y, x) > 0$.

$$\sum_{n=1}^{\infty} P^{k+\ell+n}(y, y) \geq \sum_{n=1}^{\infty} P^\ell(y, x) P^n(x, x) P^k(x, y) = \infty.$$

which implies that y is recurrent.

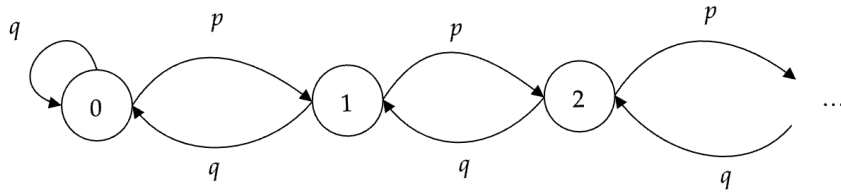
□

Example 2.27.

$$P = \begin{bmatrix} 0 & \frac{1}{6} & \frac{5}{6} \\ \frac{1}{6} & 0 & \frac{5}{6} \\ \frac{1}{6} & \frac{5}{6} & 0 \end{bmatrix}$$

and all states are recurrent.

Example 2.28. Consider the below Markov chain with $0 < p < 1$.



Consider the probability of starting at 1 and first time visit 0 at k ,

$$P_1[T_0 = k] = p_k,$$

and we have

$$\begin{aligned}
\Phi(s) &= \sum_{k=0}^{\infty} p_k s^k \\
\Phi(s) &= qs + ps\Phi(s)
\end{aligned}$$

where the second equality is by the fact that, $T_0 = 1$ when we go from 1 to 0 directly with probability q , otherwise, we go to 2 in the first step and then consider the steps required for us to go from 2 to 0, which is 2 to 1 then 1 to 0. In other words, we write

$$\begin{aligned}
\Phi(s) &= \sum_{k=0}^{\infty} p_k s^k \\
&= 0 * 1 + qs + \sum_{k=2} p_k s^k \\
&= qs + s \sum_{k=0} p_{k+1} s^k \\
&= qs + ps \sum_{k=0} P_2[T_0 = k] s^k \\
&= qs + ps \mathbb{E}[s^{X+Y}]
\end{aligned}$$

where $p_{k+1} = p * P_2[T_0 = k]$, and X is the random variable of number of steps from 0 to 1 and Y is from 2 to 1 which follow the same distribution as T_0 starting at 1 and are independent, so $\mathbb{E}[s^{X+Y}] = \mathbb{E}[s^X] \mathbb{E}[s^Y] = \Phi^2(s)$.

Then we can have that

$$\begin{aligned}
\Phi(s) &= \frac{1 - \sqrt{1 - 4pqs^2}}{2ps} \\
\Phi(1) &= \frac{1 - \sqrt{1 - 4pq}}{2p} = \begin{cases} 1, & \text{if } p \leq 1/2 \\ \frac{q}{p}, & \text{if } p > 1/2 \end{cases}
\end{aligned}$$

That is, when $p > 1/2$, there is a chance we never go to 0. Or we can find the expectation by

$$E_1[T_0] = \lim_{s \rightarrow 1} \Phi'(s).$$

Definition 2.29. We call π a *stationary distribution* for a Markov chain with transition matrix P , if

$$\pi = \pi P, \sum \pi(i) = 1.$$

Example 2.30.

$$(\pi(0), \pi(1), \pi(2)) \begin{bmatrix} 0 & \frac{1}{6} & \frac{5}{6} \\ \frac{1}{6} & 0 & \frac{5}{6} \\ \frac{1}{6} & \frac{5}{6} & 0 \end{bmatrix} = (\pi(0), \pi(1), \pi(2))$$

solve to get

$$(\pi(0), \pi(1), \pi(2)) = (11/77, 31/77, 35/77)$$

and then

$$\mathbb{E}_0[T_0] = \frac{77}{11} = \frac{1}{\pi(0)}$$

because now we can consider it as a geometric distribution with parameter $\pi(0)$, starting from $X_0 = 0$, you have 11/77 chance to get 0 at X_1 , similarly, if you get $X_1 \neq 0$, then you still have

11/77 for $X_2 = 0$ by π being stationary, and so on.

We can also consider the central limit theorem which gives:

$$\frac{f(x_0) + \dots + f(x_n)}{n+1} \rightarrow \pi f$$

for a function f valued on the states of the Markov chain X_i .

Example 2.31 (x -excursion chain). Let X_0, X_1, \dots be an irreducible Markov chain with stationary distribution π , transition matrix P and state space S . Let consider words (or strings if you prefer) that are finite, starting with x and containing only one x , call the set of all such words, S_y . Consider random variables Y_i with state space S_y , defined as

$$\begin{aligned} Y_0 &= x \\ Y_1 &= xX_1 \\ Y_2 &= xX_1X_2 \\ Y_3 &= xX_1X_2X_3 \\ &\vdots \end{aligned}$$

where we keep $X_0 = x$. So

$$\Pr[Y_3 = xy_1y_2y_3] = P(x, y_1)P(y_1, y_2)P(y_2, y_3).$$

and we can build the transition matrix Q for Y_i as

$$\begin{aligned} Q(xy_1 \dots y_k, xy_1 \dots y_k y_{k+1}) &= P(y_k, y_{k+1}) \\ Q(xy_1 \dots y_k, xy_1 \dots y_k x) &= P(y_k, x) \\ Q(x, xy) &= P(x, y) \\ Q(x, x) &= P(x, x). \end{aligned}$$

And we let $F : S_y \rightarrow S$ where $F(w)$ is the last letter of w .

Fact 2.32. If Y_0, Y_1, \dots is a Markov chain with transition matrix Q and state space S_y , then $F(Y_0), F(Y_1), \dots$ is a Markov chain with state space S and transition matrix P .

Now let's consider the stationary distribution for Y . Let ν be a stationary distribution of Y_i , then

$$\begin{aligned} \nu &= \nu Q \\ \nu(w) &= \sum_{w' \in S_y} \nu(w')Q(w', w), \quad \sum_{w \in S_y} \nu(w) = 1 \end{aligned}$$

Let $w = xy_1 \dots y_{k-1}y_k y_{k+1}$, we have

$$\begin{aligned} \nu(xy_1 \dots y_{k+1}) &= \nu(xy_1 \dots y_k)Q(y_k, y_{k+1}) \\ \nu(xy_1 \dots y_k) &= \nu(x)P(x, y_1)P(y_1, y_2) \dots P(y_{k-1}, y_k) \end{aligned}$$

Hence,

$$\begin{aligned}
\sum_{w \in S_y} \nu(w) &= \nu(x) + \sum_{k=1}^{\infty} \sum_{y_1 \dots y_k} \nu(x) P(x, y_1) P(y_1, y_2) \dots P(y_{k-1}, y_k) \\
&= \nu(x) + \nu(x) \sum_{k=1}^{\infty} \sum_{y_1 \dots y_k} P(x, y_1) P(y_1, y_2) \dots P(y_{k-1}, y_k) \\
&= \nu(x) P_x(T_x > 0) + \nu(x) \sum_{k=1}^{\infty} P_x(T_x > k) \\
&= \nu(x) \sum_{k=0}^{\infty} P_x(T_x > k) \\
&= \nu(x) \mathbb{E}_x[T_x] = 1
\end{aligned}$$

If state x is recurrent, then we have $\nu(x) = \frac{1}{\mathbb{E}_x[T_x]}$, otherwise, Q does not have a stationary distribution. Thus if X_0, X_1, \dots has a positive recurrent state x , then there exists at least one stationary distribution ν by the fact $\nu(w)$ can be defined by $\nu(x)$ and $P(x, y_1), \dots, P(y_{k-1}, y_k)$.

If $Y_0 \sim \nu$, and $Y_1, \dots \sim \nu$, let $\pi(z) = \sum_{w, F(w)=z} \nu(w)$, we have $\pi = \pi P$ and $\sum_{x \in S} \pi(x) = 1$.

Example 2.33. We consider a Markov chain X_0, X_1, \dots . For the case we start with $X_0 = x$, denote P_x , if we start with $X_0 = y$, denote P_y . Let $\tau(i)$ be the time we have the i -th x excluding X_0 , that is, $\tau(1) = T_x$, $\tau(2) = T_x^2$ and $\tau(0) = 0$. Define

$$\begin{aligned}
W_1 &= (X_0, X_1, \dots, X_{\tau(1)-1}) \\
W_2 &= (X_{\tau(1)}, \dots, X_{\tau(2)-1}) \\
&\vdots
\end{aligned}$$

Under P_x , the words W_1, W_2, \dots are i.i.d. Under P_y , $y \neq x$, the words W_1, W_2, \dots are independent, and W_2, W_3, \dots are identically distributed. Let $W_j = (X_{j,1}, \dots, X_{j,m(j)})$, then

$$\begin{aligned}
&P_x(W_1 = w_1, W_2 = w_2, \dots, W_k = w_k) \\
&= \prod_{j=1}^k \left(\prod_{\ell=1}^{m(j)-1} P(X_{j,\ell}, X_{j,\ell+1}) \right) P(X_{j,m(j)}, x) \\
&= \prod_{j=1}^k P(W_j = w_j)
\end{aligned}$$

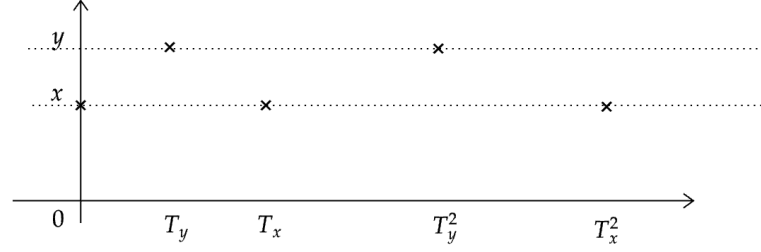
For P_y , $X_{1,1} = y$, all other $X_{j,1}$ remains at x , so w_2, w_3, \dots are identically distributed.

Proposition 2.34. WLOG, assume $x \neq y$, if x and y communicate, and x is positive recurrent, then y is positive recurrent.

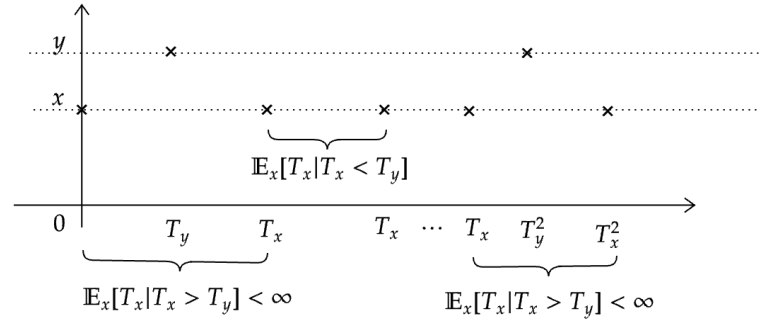
Proof.

$$\infty > \mathbb{E}_x[T_x] = \mathbb{E}_x[T_x | T_x > T_y] P_x[T_x > T_y] + \mathbb{E}_x[T_x | T_x < T_y] P_x[T_x < T_y]$$

If $P_x[T_x < T_y] = 0$, then $\mathbb{E}_y[T_y] \leq 2\mathbb{E}_x[T_x] < \infty$. The reason is that, we have $T_y \leq T_x$, then $\mathbb{E}_y[T_y]$ can be considered as $\mathbb{E}_x[T_y^2] - \mathbb{E}_x[T_y]$, but by $P_x[T_x < T_y] = 0$, we know for if we start at $X_0 = x$, then $T_y^2 \leq T_x^2$, see the plot below



If $P_x[T_x < T_y] > 0$, consider the plot



□