Stochastic Process 1

Rui Gong

April 22, 2024

Contents

1 Poisson Process

1.1 Poisson Approximation to Binomial

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

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

The probablity of a binomial random variable being k is

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

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

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

Proof.

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

□

With that, we consider three different binomial random variables:

$$\begin{aligned}
X_n &\sim \text{Binomial}(n, p_n), p_n \to 0, np_n \to \lambda > 0, \text{ as } n \to \infty. \\
Z_p &\sim \text{Binomial}(n(p), p), p \to 0, n(p)p \to \lambda > 0, \text{ as } p \to 0. \\
N_x &\sim \text{Binomial}(n(x), p(x)), p(x) \to 0, n(x) \to \lambda > 0, \text{ as } x.
\end{aligned}$$

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

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

1.2 Total Variance Distance

Let $X_1, \ldots, X_n$ 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 closedness?

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,

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

$\square$

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

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

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

And hence

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

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

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 probablity P, where

$$\begin{aligned}
&\Omega : \text{sample spaces - set of all outcomes} \\
&\mathcal{F} : \text{all events} \\
&P : \mathcal{F} \to [0, 1]
\end{aligned}$$

. Then we can write a random variable X_1 as:

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

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 $\Omega \in \mathcal{F}$

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

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

Definition 1.7. Probability Axioms:

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

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

P.3 $P(\Omega) = 1$.

Example 1.8. $X = (\Omega, \mathcal{F}) \to \mathbb{R}$, 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

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

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

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

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

For example, if we want to show $n \log(1 + \frac{\lambda_n}{n}) \to \lambda$. 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 $n \to \infty$, we have $h_n \to 0$, $nh_n \to \lambda$, $n \log(1 + \lambda_n / n) \to \lambda$.

Definition 1.9. Suppose X is nonnegative, integer-valued random variable $\Pr[X = k] = p_k$ for $k = 0, 1, 2, \ldots$, then the *probability-generating function* is defined as:

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

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

Then we have

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

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

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 $X_1, \ldots, X_n$ be i.i.d. Bernoulli random variable.

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

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

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

as $n \to \infty$.

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

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

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

$$F(t) := \Pr[X \le t], -\inf < t < \inf.$$

Given a Borel set A, we have

$$F(A) = \Pr[X \in A]$$

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

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 $X_n \overset{D}{\Longrightarrow} X$, or $X_n \to X$), if

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

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

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 $X_n \overset{D}{\Longrightarrow} X$.

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

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

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

Again X_n does not converge to X in variation. Let Q be the set of rational numbers. $\Pr[X \in Q] = 0$ because Q has measure zero, but $\Pr[X_n \in Q] = 1$. 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 $\Pr[T_n = k] = (1-p_n)^{k-1}p_n, k = 1, 2, \ldots$, $\Pr[T_n > k] = (1-p_n)^k$, 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

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

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 $F_n(t) \to F_X(T)$ for all $t \in C(X)$.

Proof.

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

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

1.5 Point Process

Consider $N \sim \text{Poisson}(\lambda)$ and let $X_1, \ldots, X_N$ be i.i.d. $\text{Bernoulli}(p)$. 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 valuesd r.v. $N(J) = N_J$, so that if $J_1, J_2, \ldots,$ 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 $J_1, J_2, \ldots,$ are nonoverlapping intervals, then $N(J_1), N(J_2), \ldots,$ are independent.

b) $N(J) \sim \text{Poisson}(\lambda|J|)$ where J is the length of the interval J.

Given a Poissson Point Process above, let $0 = T_0 < T_1 \leq T_2 \leq T_3 \leq \ldots$ be the time i^{th} customer arrives and $\tau_n = T_n - T_{n-1}$. Then $\tau_1, \tau_2, \ldots,$ 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 to 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 $\frac{1}{\lambda} = \frac{1}{5}\text{hr} = 12\text{min}$.

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

1.6 Bernoulli and Poisson

Let $X_1, X_2, \ldots$, 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 $X_1, X_2, \ldots$, 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 $0 = T_0 < T_1 = \tau_1 \leq T_2 = \tau_1 + \tau_2 \leq \ldots$ be the *occurence times* of a Poisson process which are the successive times $N(t)$ jumps. Let $\tau_1, \tau_2, \ldots$ be the *interoccurence time*, where $\tau_i := T_i - T_{i-1}$.

Theorem 1.20 (Interoccurence Time Theorem)**.**

(A) Interoccurence times $\tau_1, \tau_2, \ldots,$ of a Poisson process with rate λ are i.i.d. $\mathrm{Exp}(\lambda)$

(B) Let $Y_1, Y_2, \ldots,$ be i.i.d. $\mathrm{Exp}(\lambda)$.

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

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 $m\tau_1^m \sim \mathrm{Geo}(p_m)$. Let $T_2^m = \min\{n > T_1^m : X_n^m = 1\}$ and $\tau_2^m = T_2^m - T_1^m$, then $m\tau_2^m \sim \mathrm{Geo}(p_m)$ as well. Then with the occurrence time T_i, we have a counting process

$$N^m(t_1) \sim \mathrm{Binomial}(?\,, p_m)$$

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 $mp_m \to \lambda > 0$, 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:

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

Proof of Interoccurence Time Theorem.

a) We showed in the previous section that for a geometric r.v. with p_n with $np_n \to \lambda$. $T_n/n \xrightarrow{D} \mathrm{Exp}(\lambda)$. And we have seen that the interoccurrence times of Bernoulli $\{X_k^m\}_{k\in\mathbb{N}/n}$ are geometric, $\Delta_k^m = N^m(t_k) - N^m(t_{k-1}) \sim \mathrm{Binomial}(m(t_k - t_{k-1}) \pm 1, p_m)$ where $\pm$ considers the rounding of $m(t_k - t_{k-1})$. And this converges in distribution to $\Delta_k \sim \mathrm{Poisson}(\lambda(t_k - t_{k-1}))$. Thus the occurrence time of $N^m(t)$ converges to $N(t)$ in distribution. Thus, the interoccurence time of X_k^m, which is the interoccurence time of $N^m(t)$, converging to $\mathrm{Exp}(\lambda)$ implies that the interoccurence time of $N(t)$ converges to $\mathrm{Exp}(\lambda)$.

b) With a Poisson process with rate λ, and let τ_i be its interoccurence times, and we know $\tau_i \overset{iid}{\sim} \mathrm{Exp}(\lambda)$. Let Y_i be another sequence of i.i.d. expenentials with λ. Then since τ_i and Y_i have the same joint distribution, we also have

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

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

$\square$

Given B), now we can simulate Poisson with $U_i \overset{D}{\sim} \text{Uniform}([0,1])$ 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

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

Theorem 1.23 (Generalized Thinning Theorem)**.** Let $N \sim \text{Poisson}(\lambda)$, X_i be iid r.v. with $\Pr[X_i = k] = p_k, k = 1, \ldots, m$ 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 $N_1, \ldots, N_m$ 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

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

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

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

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

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

1. $J_1, J_2, \ldots$ disjoint Borel sets in $\mathbb{R}^k$; $N(J_1), N(J_2), \ldots$ 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 occurence 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 $\frac{\lambda}{\mu+\lambda}$ and $\frac{\mu}{\mu+\lambda}$ 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 occurence times and τ_i be the interoccurence times as before. Then $T_n = \tau_1 + \ldots + \tau_n$. 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 $T_1, T_2, \ldots,$ 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 and hence all $T_2, T_3, \ldots$ are know, so the process with T_i being the occurence 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

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

Theorem 1.28 (Dominant 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 $\lim_{n\to\infty} X_n(\omega)$ and such that for all $n \geq 1$

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

where Y is some integrable random variable. Then

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

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

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

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

where the last equality is by $\lim_{n\to\infty} \Pr[X_n > 0] = 0$ and $\lim_{n\to\infty} \Pr[X_n = 0] = 1$, so $X_n \to 0$ 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.m 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. $T \sim \text{Binomial}(n, p)$. $T = X_1 + \ldots + X_n$, where X_i are i.i.d. Bernoulli.
$$\begin{aligned} L(s) &= \mathbb{E}[e^{-sT}] \\ &= \int_{[0,\infty)} e^{-st} dF(t) \\ &= \mathbb{E}[e^{-s(X_1 + \ldots + X_n})] \\ &= \mathbb{E}[e^{-sX_1} \ldots e^{-sX_n}] \\ &= \mathbb{E}[e^{-sX_1}] \ldots \mathbb{E}[e^{-sX_n}] \\ &= (1 - p + pe^{-s})^n \end{aligned}$$
4. Let $X \sim \text{Exp}(\lambda)$, we have
$$\mathbb{E}[e^{-sX}] = \int_0^\infty e^{-st} \lambda e^{-\lambda t} dt = \frac{\lambda}{s + \lambda}. \qquad \text{(L.S. of Exp)}$$

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 equality 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. $\square$

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 $\Pr[N(T) = k] = \rho^k(1-\rho), k = 0, 1, \ldots$. Then

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

$$\begin{aligned} L(s) = \mathbb{E}[e^{-sT}] = G(1-s/\lambda) &= \frac{1-\rho}{1-\rho(1-s/\lambda)} \\ &= \frac{1-\rho}{1-\rho+\rho s/\lambda} = \frac{\frac{\lambda}{\rho}(1-\rho)}{\frac{\lambda}{\rho}(1-\rho)+s} \end{aligned}$$

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

2 Markov-Chain

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

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

1. Markov Property:

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

2. Time Homogeneity:

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

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

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

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

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

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

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

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

$$\begin{aligned} \Pr[Y_2 = 0 | Y_1 = 1, Y_0 = 0] = 0 \\ \Pr[Y_2 = 0 | Y_1 = 1, Y_0 = 2] = 1 - p \end{aligned}$$

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

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

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 interoccurence 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

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

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 just before $n+1$-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:

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

$M/M/1/3$ **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 argument applies 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

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

We have the transition plot of the above matrix,

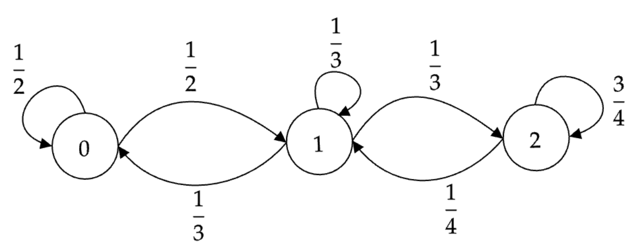


Figure 1: Transition Plot of P

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

Example 2.6. Given $X_0, X_1, \ldots$, we have

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

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 in (2.1), 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.

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

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, Recurence 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 j is *absorbing* i 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 by $P^{m+n}(i,k) \geq P^m(i,j)P^n(j,k)$, with $m+n$ steps, we might go from i to j.

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

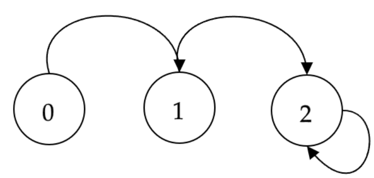


Figure 2: Self Commu States

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 divider of $\{n > 0 | P^n(i,i) > 0\}$. If $d(i) = 1$ for every state i, then the Markov Chain is *aperiodic*.

Example 2.19. Given the following graph:

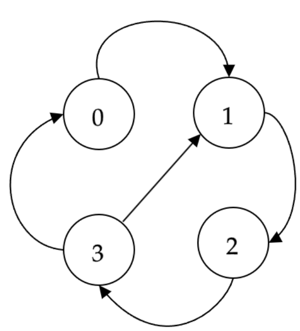


Figure 3: Period 1

Consider $i = 0$, then

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

For the following graph:

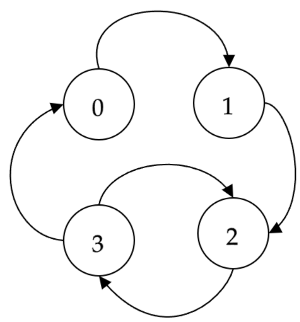


Figure 4: Period 2

Consider $i = 0$, then

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

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 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 if $\{T = n\}$ can be determined from $X_0, \ldots, X_n$, i.e.

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

for some function g_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. But $T_x = \inf\{n \geq 0 | X_{n+1} = x\}$ is not a stopping time, because $\{T = n\}$ requires knowing X_{n+1}.

Let T be a stopping time, then

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

and since T is a stopping time, $T = m$ is redundant by knowing $X_m, \dots, X_0$. This is called *Strong Markov Property*. That is, the Strong Markov Property says that if we know a stopping time $T = m$, then we can treat the Markov chain after T as one Markov chain Y with the same transition matrix P but starting with $Y_0 = X_m$.

Definition 2.23. Let $T_x^1 = T_x = \inf\{n \geq 1 | X_n = x\}$, $T_x^k = \inf\{n > T_x^{k-1} | X_n = x\}, k = 2, 3, \dots$, 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 recurrent* if $\mathbb{E}_x[T_x] < \infty$.
- State x is *null recurrent* 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. The number of times x will be revisited is represented as

$$N_x = \sum_{k=0}^{\infty} \mathbb{1}\{X_k = x\}.$$

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.

$$\begin{aligned}
\mathbb{E}_x[N_x] &= \mathbb{E}_x \sum_{k=0}^{\infty} \mathbb{1}[X_k = x] \\
&= \sum_{k=0}^{\infty} \mathbb{E}_x[\mathbb{1}[X_k = x]] \\
&= \sum_{k=0}^{\infty} \Pr_x[X_k = x] = \sum_{k=0}^{\infty} P^k(x, x)
\end{aligned}$$

$$N_x = 1 + \sum_{k=1}^{\infty} \mathbb{1}[T_x^k < \infty]$$

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

□

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 $P^\ell(y,x) > 0$.

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

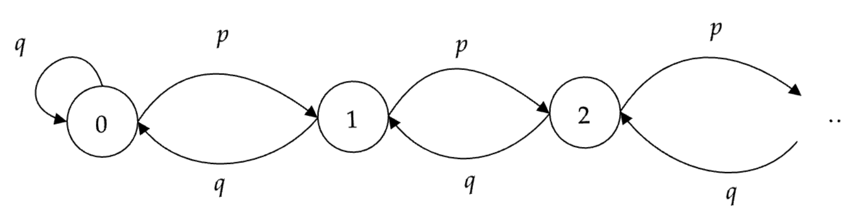
which implies that y is recurrent. □

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

$$
\begin{aligned}
\Phi(s) &= \sum_{k=0}^{\infty} p_k s^k \\
\Phi(s) &= qs + ps\Phi^2(s)
\end{aligned}
$$

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

$$\mathbb{E}_1[T_0] = \lim_{s \to 1} \Phi'(s).$$

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:

$$\frac{f(x_0) + \ldots + f(x_n)}{n+1} \to \pi f$$

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

Example 2.31 (x-excursion chain)**.** Let $X_0, X_1, \ldots$ be an irreducible Markov chain with stationary distribution π, transition matrix P and state space S. Let's 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

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

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

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

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

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

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

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, $\tau(i)=T_x$, $\tau(2)=T_x^2$ 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 workds $W_1,W_2,\dots$ are independent, and $W_2,W_3,\dots$ are identically distributed. Let $W_j=(X_{j,1},\dots,X_{j,m(j)})$, then

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

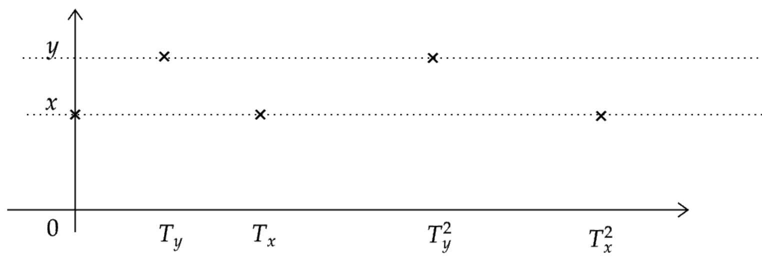
For P_y, $X_{1,1}=y$, all other $X_{j,1}$ remains at x, so $w_2,w_3,\dots$ are identically distributed.

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

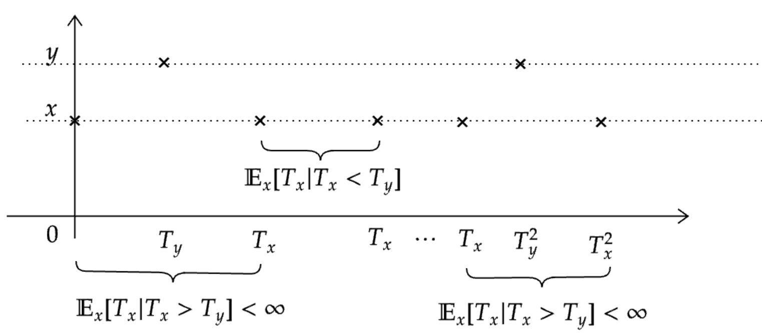
Proof.

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

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



Similar, we have $\mathbb{E}_y[T_y] < \infty$. $\square$

2.3 Stationary Distribution and Positive Reucurrence

Consider a random variable X, we can write it as $X = X^+ - X^-$, where $X^+ := \max(X, 0)$ and $X^- := \max(-X, 0)$. If both $\mathbb{E}[X^+], \mathbb{E}[X^-]$ are well-defined with value in $[0, \infty]$. Then

$$\mu := \mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$$

unless it is $\infty - \infty$. To avoid this, we can assume either X is nonnegative, or $|X|$ integrable ($\mathbb{E}[|X|] < \infty$), or $\mathbb{E}[X^-] < \infty$, then we have $\mu < \infty$ or μ is well-defined as ∞.

Theorem 2.35 (Strong Law of Large Number)**.** Consider $S_n = X_1 + \ldots + X_n$

1. If $X_1, X_2, \ldots$ are pairwise i.i.d. integrable with mean μ, then

2. Or if $X_1, X_2, \ldots$ are i.i.d. with $\mathbb{E}[X^+] = \infty$, $\mathbb{E}[X^-] < \infty$ with mean $\mu = \infty$, then

$$\frac{S_n}{n} \to \mu \text{ a.s. w.p. } 1$$

almost surely with probability 1.

When we say with almost surely with probability 1, we mean that the set

$$A = \left\{\omega \in \Omega : \frac{S_n(\omega)}{n} \to \mu\right\}$$

has a probability 1 when $n \to \infty$.

Example 2.36. Recall our "string" example, where $W_1 = (X_0, \ldots, X_{\tau(1)-1}), W_2 = (X_{\tau(0)}, \ldots, X_{\tau(2)-1}), \ldots$. Under P_x (start with $X_0 = x$), $W_1, W_2, \ldots$ are i.i.d., while under P_y, for $y \neq x$, $W_2, W_3, \ldots$ are i.i.d. and $W_1, W_2, \ldots$ are independent. Write $W_j = (X_{j,1}, \ldots, X_{j,m(j)})$, then

$$\Pr_x[W_1 = w_1, \ldots, W_k = w_k] = \prod_{j=1}^{k} \left(\prod_{\ell=1}^{m(j)-1} P(x_{j,\ell}, x_{j,\ell+1}) \right) P(x_{j,m(j)}, x).$$

Definition 2.37. Let $f : S \to \mathbb{R}_+$. The additive extension to the set of finite "words" with letters in S is the function f_+ where for $w = (x_1, \ldots, x_m)$,

$$f_+(w) = \sum_{i=1}^{m} f(x_i).$$

For any initial state $y \in S$ by the Strong Law of Large Number,

$$\lim_{k\to\infty} \frac{\sum_{i=1}^{k} f_+(w_i)}{k} = \mathbb{E}_x[f_+(w_1)] = \mathbb{E}_x[\sum_{j=0}^{\tau(1)-1} f(x_j)]$$

with P_y almost surely, because if $y \neq x$, then

$$\frac{f_+(w_1) + \ldots + f_+(w_k)}{k} = \frac{f_+(w_1)}{k} + \frac{f_+(w_2) + \ldots + f_+(w_k)}{k-1} \frac{k-1}{k} \to 0 + \mathbb{E}_x[f_+(w_2)] * 1.$$

In particular, if we set $f \equiv 1$, then

$$\lim_{k\to\infty} \tau(k)/k = \mathbb{E}_x[\tau(1)]$$

with P_y almost surely.

Let $N_n^x =$ the number of visits to state x up to time $n = \sum_{k=1}^{n} \mathbb{1}\{X_k = x\}$.

Theorem 2.38. Fix $x \in S$. If the Markov Chain is irreducible and positive recurrent, then $\exists!$ (there exists a unique) stationary distribution π and for all states x, y,

$$\lim_{n\to\infty} N_n^x/n = \pi(x), \ P_y\text{-a.s.}$$

If the chain is null recurrent, then there does not exist a stationary distribution and for all x, y,

$$\lim_{n\to\infty} N_n^x/n = 0, \ P_y\text{-a.s.}$$

Proof. First, we show $N_n^x/n \to 1/\mathbb{E}_x[T_x]$, P_y-a.s. Note, $N_n^x \le n$, and $N_n^x \to \infty$ P_y a.s.,

$$\frac{\tau(N_n^x)}{N_n^x} \le \frac{n}{N_n^x} < \frac{\tau(1+N_n^x)}{1+N_n^x}\frac{1+N_n^x}{N_n^x}.$$

where $n < \tau(1+N_n^x)$. And $\frac{\tau(N_n^x)}{N_n^x} \to \mathbb{E}_x[\tau(1)]$, $\frac{\tau(1+N_n^x)}{1+N_n^x} \to \mathbb{E}_x[\tau(1)]$, so $n/N_n^x \to \mathbb{E}_x[\tau(1)]$ with P_y-a.s..

Second, assume the Markov Chain has a stationary distribution π, then define $P_\pi(\cdot) = \sum_y \pi(y)P_y(\cdot)$,

$$N_n^x/n \to 1/\mathbb{E}_x[T_x],\ P_\pi\text{-a.s.}$$

by P_y-a.s and

$$\lim_{n\to\infty} \mathbb{E}_\pi[N_n^x/n] = \mathbb{E}_\pi \lim_{n\to\infty} N_n^x/n = \mathbb{E}_\pi[1/\mathbb{E}_x[T_x]] = 1/\mathbb{E}_x[T_x]$$

where the first equality is by $|N_n^x/n| \le 1$, $\mathbb{E}_\pi(1) = 1 < \infty$ by Dominant Convergence Theorem. The above equation is equivalent to

$$\lim_{n\to\infty} \mathbb{E}_\pi[N_n^x/n] = \lim_{n\to\infty} \mathbb{E}_\pi \frac{\sum_{j=1}^n \mathbb{1}[X_j = x]}{n} = \lim_{n\to\infty} \frac{n\pi(x)}{n} = \pi(x)$$

by π being stationary, $\mathbb{E}_x[\mathbb{1}[X_j = x]] = 1 * P_\pi(x) = \sum_y \pi(y)P_y(x) = \pi(x)$. Hence, for all state x,

$$\pi(x) = \frac{1}{\mathbb{E}_x[T_x]}.$$

For the positive recurrent case, π is uniquely defined as above. If it's the null recurrent case, then $\mathbb{E}_x[T_x] = \infty$, $\pi(x) = 0$, which is not even a distribution. $\square$

Lemma 2.39. If $X_0, X_1, \ldots$ is recurrent, then the invariant measure is unique up to multiplication by constants.

Proof. See Bremaul's book. $\square$

Combining the Lemma and Theorem, we know a recurrent Markov Chain's invariant measure sometimes does not give a stationary distribution because the sum of measure goes to infinity.

2.4 Period

2.4.1 Fundamental Theorem of Markov Chain

Let $a_1, a_2, \ldots$ be a sequence of integers. $d_k = g.c.d.(a_1, \ldots, a_k)$, if $1 \le d_k$ is nondecreasing and $d_k \to d$, then there exists k_0 such that $d_k = d$ for $k \ge k_0$.

Lemma 2.40. Let $S \subseteq \mathbb{Z}$ contain at least one non-zero element and be closed under addition and subtraction. Then S contains a smallest, positive integer a and $S = \{ka : k \in \mathbb{Z}\}$.

Proof. Let $c \in S$ with $c \neq 0$, then $0 = c - c \in S$ and $-c = 0 - c \in S$. Hence S contains at least one positive, one negative value. Then S contains a smallest positive element a. So

$$\begin{aligned} a, 2a, 3a, \ldots &\in S \\ -a, -2a, -3a, \ldots &\in S \end{aligned}$$

so $\{ka : k \in \mathbb{Z}\} \subseteq S$. Let $c \in S$, $c = ka + r$, $0 \leq r \leq a - 1, r \in \mathbb{Z}$. And $0 \leq r = c - ka \in S$ by subtraction, but $r < a$ and a is the smallest positive integer in S, so $r = 0$. □

Lemma 2.41. Let $a_1, a_2, \ldots, a_k$ be positive integer with g.c.d. d, there exist $n_1, n_2, \ldots, n_k \in \mathbb{Z}$ such that $d = \sum_{i=1}^{k} n_i a_i$.

Proof. The set $S = \{\sum_{i=1}^{k} n_i a_i : n_1, \ldots, n_k \in \mathbb{Z}\}$ is closed under additions and subtractions. So $S = \{ka : k \in \mathbb{Z}\}$ with $a = \sum_{i=1}^{k} n_i a_i$ being the smallest positive integer in S. Hence, d is a divisor of a by $a = \sum_{i=1}^{k} n_i a_i$. Then by $a_i = ka$, we know a is a divisor of a_i, so $a \leq g.c.d.(a_1, \ldots, a_k) = d$, so $a = d$. □

Theorem 2.42. $A = \{a_1, a_2, \ldots\}$ which is a set of positive integers. Let $d = g.c.d.(A)$, and A is closed under addition. Then A contains, all but a finite number of multiples of d.

Proof. WLOG, $d = 1$. For some k, we have $d = g.c.d.(a_1, \ldots, a_k)$. By Lemma (2.41).

$$1 = \sum_{i=1}^{k} n_i a_i, \text{ for some } n_1, \ldots, n_k \in \mathbb{Z}, 1 \quad = M - P, \text{ where } M \geq 0, P < 0, M, P \in A.$$

Let $n \in \mathbb{N}$, $n \geq P(P-1)$, $n = aP + r$, $0 \leq r \leq P - 1$, so $a \geq P - 1$ (If $a \leq P - 2, aP + r < P(P-1)$). By $1 = M - P$, we have

$$n = aP + r(M - P) = (a - r)P + rM$$

and $a - r \geq 0$ by $a \geq P - 1 \geq r$, which implies $n \in A$. Hence, $n \in A$ except for $n < P(P-1)$, $n \in \mathbb{N}$. □

Theorem 2.43 (Fundamental Theorem of Markov Chain)**.** For an irreducible positive recurrent aperiodic Markov chain X_i with the stationary distribution π and transition matrix P, we have

$$\lim_{n \to \infty} \Pr_i[X_n = j] = \lim_{n \to \infty} P^n(i, j) = \pi(j)$$

Proof. Consider two sequences of variables. Let $x = X_0, X_1, \ldots$ be the Markov chain starting with x, and $X_0^*, X_1^*, \ldots$ be a Markov chain where each $X_i^* \sim \pi$.

We have

$$\begin{aligned} |\Pr_x[X_n = y] - \Pr_\pi[X_n = y]| =& |\Pr_x[X_n = y] - \pi(y)| \\ =& |\Pr[X_n = y, X_n^* = y] + \Pr[X_n = y, X_n^* \neq y] \\ & - \Pr[X_n = y, X_n^* = y] - \Pr[X_n^* = y, X_n \neq y]| \\ \leq& \Pr[X_n \neq X_n^*] \end{aligned}$$

and we want to show $\Pr[X_n \neq X_n^*]$ goes to 0. Let $\tau := \min\{n \geq 0 : X_n = X_n^*\}$. And consider another independent Markov chain $X_0', X_1', \ldots$ which use the same transition matrix P. Consider a Markov chain V_n and its transition matrix Q:

$$V_n = (X_n, X_n'), \Pr[V_{n+1} = (y, y')|V_n = (x, x'), V_{n-1}, \ldots, V_0] = Q((x, x'), (y, y')) = P(x, y)P(x', y').$$

V_n has a stationary distribution where $\pi(x, x') = \pi(x)\pi(x')$ and

$$\begin{aligned}
\pi(y, y') &= \sum_x \sum_{x'} \pi(x, x')Q((x, x'), (y, y')) \\
&= \sum_x \sum_{x'} \pi(x)\pi(x')P(x, y)P(x', y') \\
&= \sum_x \pi(x)P(x, y) \sum_{x'} \pi(x')P(x', y') \\
&= \pi(y)\pi(y')
\end{aligned}$$

Consider

$$A_x = \{n \geq 1 : P^n(x, x) > 0\}$$

Then Theorem 2.42, there exists n_x such that $\forall n \geq n_x$, $P^n(x, x) > 0$ and there exists $k_{x,y}$ such that $P^{k_{x,y}}(x, y) > 0$, so $P^{n+k_{x,y}}(x, y) \geq P^{k_{x,y}}(x, y)P^n(x, x)$. Hence

$$P^n(x, y) > 0, \forall n \geq k_{x,y} + n_x,$$

similarly, we also have

$$P^n(x', y') > 0, \forall n \geq k_{x',y'} + n_{x'}.$$

Then for all $n \geq \max\{k_{x,y} + n_x, k_{x',y'} + n_{x'}\}$, we have

$$Q^n((x, x'), (y, y'))) > 0,$$

so V_n is irreducible and aperiodic (by letting $y, y' = x, x'$) and positive recurrent by having a stationary distribution.

Hence, all states are expected to be visited in finite time. $\tau' = \min\{n \geq 0 : X_n = X_n'\}$, $\tau' < \infty$ almost surely by considering arbitrary (x, x). Consider

$$\bar{X}_n = \begin{cases} X_n', & n \leq \tau' \\ X_n, & n > \tau' \end{cases}.$$

By the Strong Markov Property, the part of X_n' and X_n for $n \geq \tau'$ are i.i.d. Markov chain, so the $\bar{X}_n$ we construct follow the same distribution as X_n^* follows. That is,

$$\Pr[X_n \neq X_n^*] = \Pr[X_n \neq \bar{X}_n] = \Pr[\tau' > n] \to 0$$

by $\tau' < \infty$ almost surely. $\square$

2.5 Reversibility

Let P be a transition matrix for an irreducible Markov chain. Take a guess for stationary distribution π and a reverse transition matrix $\overset{\leftharpoonup}{P}$ with the same state space. If $\pi(j)\overset{\leftharpoonup}{P}(j,i) = \pi(i)P(i,j)$, then both guesses are right, we know this Markov chain is reversible with $\overset{\leftharpoonup}{P}$ and positive recurrent with π.

Consider $X_0, X_1, \ldots$ being stationary with the stationary distribution π, then

$$\begin{aligned}
\Pr[X_n = i, X_{n+1} = j] &= \Pr[X_n = i] \Pr[X_{n+1} = j | X_n = i] = \pi(i)P(i,j) \\
&= \Pr[X_{n+1} = j] \Pr[X_n = i | X_{n+1} = j] \\
&= \pi(j)\overset{\leftharpoonup}{P}(j,i)
\end{aligned}$$

Example 2.44. Consider a simple graph $G = (V, E)$ with vertices $0, \ldots, n$, then consider a Markov chain with states being the vertices with the transition matrix:

$$P(i,j) = \begin{cases} \frac{1}{d(i)}, & \text{if } ij \in E \\ 0, & \text{otherwise.} \end{cases}$$

Then $v = (d(0), d(1), \ldots)$ is an invariant measure. Consider

$$\begin{aligned}
& v(j)\overset{\leftharpoonup}{P}(j,i) = v(i)P(i,j) \\
\iff & d(j)\overset{\leftharpoonup}{P}(j,i) = d(i)\frac{1}{d(i)}
\end{aligned}$$

so $\overset{\leftharpoonup}{P}(j,i) = \frac{1}{d(j)}$, that is, $\overset{\leftharpoonup}{P} = P$.

2.6 Wald's (First) Lemma

Theorem 2.45 (Wald's (First) Lemma)**.** Let $X_1, X_2, \ldots,$ be i.i.d. integrable random variables with $\mathbb{E}[|X_1|] < \infty$, and let T be a stopping time with $\mathbb{E}[T] < \infty$, then

$$\mathbb{E}[\underbrace{X_1 + X_2 + \ldots + X_T}_{=:S_T}] = \mathbb{E}[X_1]\mathbb{E}[T]$$

where $S_0 := 0$.

Proof. We have

$$S_T = \sum_{n=1}^{T} X_n = \sum_{n=1}^{\infty} X_n \mathbb{1}\{n \leq T\} = \sum_{n=1}^{\infty} X_n \mathbb{1}\{T \leq n-1\}^c,$$

where the supscript c means the complement of the set. Also consider

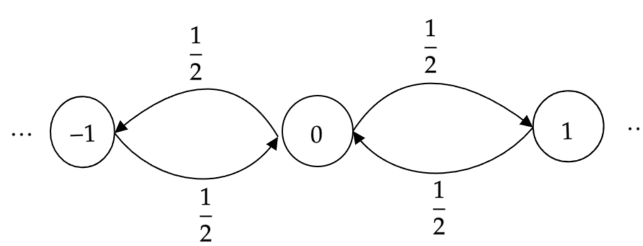
$$\begin{aligned}
\mathbb{E}[S_T] &\overset{?}{=} \sum_{n=1}^{\infty} \mathbb{E}[X_n]\mathbb{E}[\mathbb{1}\{T \leq n-1\}^c] \\
&= \sum_{n=1}^{\infty} \mathbb{E}[X_n] \Pr[T \geq n] \\
&= \mathbb{E}[X_1] \sum_{n=1}^{\infty} \Pr[T \geq n] \\
&= \mathbb{E}[X_1]\mathbb{E}[T].
\end{aligned}$$

For the "?" equation, note

$$\begin{aligned}
&\mathbb{E}[\sum_{n=1}^{\infty} X_n \mathbb{1}\{T \geq n\}] \\
\leq &\mathbb{E}\left[\sum_{n=1}^{\infty} \underbrace{|X_n|}_{\geq 0} \underbrace{\mathbb{1}\{T \geq n\}}_{\geq 0}\right] \\
= &\sum_{n=1}^{\infty} \mathbb{E}\left[|X_n| \mathbb{1}\{T \leq n-1\}^c\right] \\
= &\mathbb{E}[|X_1|] \sum_{n=1}^{\infty} \mathbb{E}\left[\mathbb{1}\{T \leq n-1\}^c\right] \\
= &\mathbb{E}[|X_1|] \sum_{n=1}^{\infty} \Pr[T \geq n] \\
= &\mathbb{E}[|X_1|]\mathbb{E}[T] < \infty.
\end{aligned}$$

then we can apply dominant expectation theorem and get "?". □

Example 2.46. We consider an example fails the assumptions. Consider

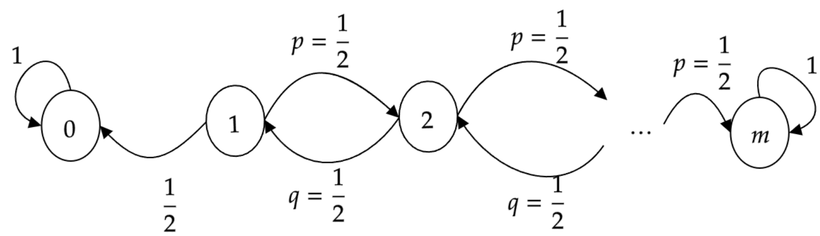


This Markov Chain is null recurrent. Let T be the first time reach 1 starting from 0.

$$X_i = \begin{cases} 1, \text{ with probability } \frac{1}{2}, \\ -1, \text{ with probability } \frac{1}{2}. \end{cases}$$

Then $\mathbb{E}[X_i] = 0$, $T = \inf\{n \geq 0 : S_n = 1 = X_1 + \ldots + X_n\}$ and $\mathbb{E}[T] = \infty$, so we cannot consider $\mathbb{E}[X_1]\mathbb{E}[T] = \mathbb{E}[S_T]$.

2.6.1 Gambler's Ruin



Suppose we start with i, then $S_0 = i$. Let $T = \inf\{n \geq 0 : X_n \in \{0, m\}\}$, $h_i = \Pr_i[S_T = m]$, $h_0 = 0, h_m = 1$. First, we show that $\mathbb{E}[T]$ is finite. Let $X_1, X_2, \ldots$ be i.i.d. r.v.s where $\Pr[X_i = 1] = p$ and $\Pr[X_i = -1] = q = 1 - p$ where $p \in (0, 1)$. Let $S_0 = i$, and $S_n = S_0 + X_1 + \ldots + X_n$. Now define

$$N = \inf\{k | X_{(k-1)m+1} = X_{(k-1)m+2} = ... = X_{km} = 1\}.$$

That is, we are looking at blocks of size m. First, we look at $X_1, \ldots, X_m$, then we look at $X_{m+1}, ..., X_{2m}$, and so on until we find a block that are all the X's are $+1$. Now, $T <= mN$ since if we have m successes in a row, then the gambler must have either m or 0 dollars. But, N is a geometric r.v. with parameter $p^m > 0$. Since $\mathbb{E}[N] = 1/p^m$, we have that $\mathbb{E}[T] <= \mathbb{E}[mN] = m/p^m < \infty$.

A way to interpret the setup is that we start with i dollars, for each play, we have p chances to win and q to lose. And we have to stop playing when we reach m or 0 dollar.

$\mathbb{E}[S_T | S_0 = i] = mh_i, \mathbb{E}[X_1 + \ldots + X_T] = (m-i)h_i + (-i)(1-h_i) = mh_i - i = 0$ by $\mathbb{E}[X_1] = 0$

where $\mathbb{E}[X_1 + \ldots + X_T] = (m-i)h_i + (-i)(1-h_i)$ is by the fact that $X_1 + \ldots + X_T$ is either $-i$ or $(m-i)$, and it is $m - i$ with probability h_i.

Consider $S'_n := S_n - i$, when $p = 1/2 = q$, as we have seen, $h_i = \frac{i}{m}$, then $\mathbb{E}_i[S_T] = mh_i + 0 = i$ or we can get it by $\mathbb{E}_i[S_T] = \mathbb{E}[S'_T] + i = 0 + i$, where $S'_T = X_1 + \ldots + X_T$ by the fact that $S_0 = i$ and $\mathbb{E}[S'_T] = \mathbb{E}[X_1 + \ldots + X_T] = \mathbb{E}[X_1]\mathbb{E}[T] = 0$.

If $p \neq q$, then note that $h_i = ph_{i+1} + qh_{i-1}$. Consider the probability matrix P of this chain, we know $h = Ph$. Consider

$$\begin{aligned}(p+q)h_i &= ph_{i+1} + qh_{i-1} \\ qh_i - qh_{i-1} &= ph_{i+1} - ph_i \\ \frac{q}{p}(h_i - h_{i-1}) &= h_{i+1} - h_i\end{aligned}$$

Notice we can write

$$0 = ah_{i+1} + bh_i + ch_{i-1}$$

which is in the form of a second order difference equation. The characterization equation of it is

$$0 = ar^2 + br + c$$

and once it is solved, we have two roots r_1, r_2 such that

$$h_i = c_1 r_1^i + c_2 r_2^i, \text{ when } r_1 \neq r_2$$
$$h_i = c_1 r_1^i + c_2 i r_2^i, \text{ when } r_1 = r_2$$

for some constants c_1, c_2.

For our case, we need to consider

$$0 = pr^2 - r + q \iff 0 = (r-1)(pr-q) \implies r_1 = 1, r_2 = q/p.$$

When $p = q$, $h_i = c_1 1^i + c_2 i 1^i = c_1 + c_2 i$. Note $h_0 = 0 = c_1 + c_2 * 0 = c_1 = 0$ and $h_m = 0 + c_2 m = 1$ so $c_2 = \frac{1}{m}$ and $h_i = \frac{i}{m}$ which agrees with what we had before. When $p \neq q$, $h_i = c_1 + c_2 (\frac{q}{p})^i$ and $h_0 = 0 \implies c_1 + c_2 = 0, c_2 = -c_1$. And $h_m = c_1 - c_1 (\frac{q}{p})^m = 1 \implies c_1 = \frac{1}{1-(\frac{q}{p})^m}$, so $h_i = \frac{1-(\frac{q}{p})^i}{1-(\frac{q}{p})^m}$.

Let $g(i) := \mathbb{E}_i[T]$, note $g(0) = 0 = g(m)$ and $g(i) = 1 + pg(i+1) + qg(i-1)$. For the same S_T defined before, consider $S'_T = S_T - i$ where $S'_n = S_n - i$. We know

$$\begin{aligned} \mathbb{E}_i[S'_T] &= \mathbb{E}_i[X_1] \underbrace{\mathbb{E}_i[T]}_{\leq \frac{m}{p^m} < \infty} \\ &= (p-q)\mathbb{E}_i[T]. \end{aligned}$$

When $p \neq q$, $\mathbb{E}_i[S'_T] = (m-i)h_i + (-i)(1-h_i) = mh_i - i$, so

$$\mathbb{E}_i[T] = \frac{mh_i - i}{p-q}.$$

When $p = q$, use second order difference equation, we have

$$\mathbb{E}_i[T] = g(i) = i(m-i)$$

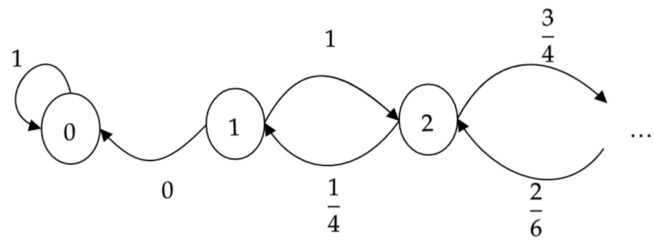
which can be verified that

$$i(m-i) = 1 + \frac{1}{2}(i+1)(m-i-1) + \frac{1}{2}(i-1)(m-i+1).$$

Now we consider the reverse probability, by supposing $h > 0$, then

$$\tilde{P}(i,j) = \frac{P(i,j)h_j}{h_i}$$
$$\sum_j \tilde{P}(i,j) = \frac{\sum_j P(i,j)h_j}{h_i} = \frac{h_i}{h_i} = 1.$$

When $p = q$, $\tilde{P}(i, i+1) = \frac{P(i,i+1)\frac{i+1}{m}}{\frac{i}{m}} = p\frac{i+1}{i} = \frac{i+1}{2i}$,


1
1
3/4
0
1
2
...
0
1/4
2/6

3 Renewal Theory

Definition 3.1. Let nonnegative random variables $S_0, S_1, \dots$ be a *renewal sequence* where $0 \leq S_0 < S_1 < \dots$. That is, consider S_i the time of i-th renewal. Let X_i be the time between $(i-1)$-th and i-th renewal, which is i.i.d. and nonnegative.

We call a renewal sequence an ordinary renewal process if

$$S_n = X_1 + \dots + X_n.$$

We call it a delayed renewal process if

$$S_n = S_0 + X_1 + \dots + X_n,$$

where S_0 is the delay.

We let $N(t) := \max\{n : S_n \leq t\}$ which the number of renewals until t.

If you wish, you can think about a renewal process as considering returning to a specific state of a Markov chain. Then it is recurrent if $\Pr[X_i < \infty] = 1$ and transient if $\Pr[X_i < \infty] < 1$.

Some examples of renewal processes are:

- **Arithmetic Bernoulli Process**: $\Pr[X_i \in h\mathbb{Z}] = 1$ with the largest possible h, assume, WLOG, $h = 1$.
- **Nonarithmetic Poisson Continuous Process**: For $X_i \sim \exp(\pi)$, then we have Poisson which is a renewal process.

3.1 Erdös Feller Pollard

The Erdos Feller Pollard (EFP) theorem we are discussing in this subsection is for arithmetic processes. For the non-arithmetic ones, one can check the Blackwell's theorem.

Let $f_k = \Pr[X_i = k]$, $\sum_{k=1}^{\infty} f_k = 1$, and $\mu = \sum_{k=1}^{\infty} k f_k < \infty$.

Definition 3.2. The *renewal measure* of a renewal process is defined as:

$$u(m) = \Pr[S_n = m \text{ for some } n \geq 0] = \sum_{n=0}^{\infty} \Pr[S_n = m],$$

which is the probablity that the renewal ever happens at time m.

Theorem 3.3 (Elementary Renewal Theorem)**.** For an ordinary, recurrent and arithmetic renewal process,

$$\frac{N(m)}{m} \to \frac{1}{\mu} \text{ a.s.}$$
$$\mathbb{E}\left[\frac{N(m)}{m}\right] \to \frac{1}{\mu}$$

Proof. Notice that

$$S_{N(m)} \le m \le S_{N(m)+1} \iff \frac{S_{N(m)}}{N(m)} \le \frac{m}{N(m)} \le \frac{S_{N(m)+1}}{N(m)}$$

where $\frac{S_{N(m)}}{N(m)} \to \mu$ a.s. by $\frac{S_n}{n} = \frac{X_1+\ldots+X_n}{n} \to \mu$ a.s., and we also have $\frac{S_{N(m)+1}}{N(m)} = \frac{S_{N(m)+1}}{N(m)+1}\frac{N(m)+1}{N(m)} \to \mu * 1$ a.s..

Also, since $N(m)/m \le 1$, by Dominant Convergence Theorem or Bounded Convergence Theorem, we have

$$\lim_{m\to\infty} \mathbb{E}[N(m)/m] = \mathbb{E}\lim_{m\to\infty} N(m)/m = 1/\mu.$$

□

Theorem 3.4 (Erdös Feller Pollard (EFP) Theorem)**.** Let $\{S_n, n \ge 0\}$ be an ordinary, recurrent arithmetic renewal process whose interoccurence times have distribution $f_k = \Pr[X_i = k]$, $\mu = \mathbb{E}[X_i] < \infty$ that is not supported by any proper subgroup of the integers (i.e., there does not exists an $m \ge 2$ such that $\Pr[X_i \in m\mathbb{Z}] = 1$). Then

$$\lim_{m\to\infty} u(m) = \frac{1}{\mu} = \frac{1}{\mathbb{E}[X_i]}.$$

Corollary 3.5. With the same conditions as EFP, except having a delayed renewal process

$$\lim_{m\to\infty} \Pr[S_n = m \text{ for some } n \ge 0] = \frac{1}{\mu}.$$

Proof. We can condition on S_0, then

$$\begin{aligned}
&\lim_{m\to\infty} \Pr[S_n = m, \text{ for some } n \ge 0] \\
=&\lim_{m\to\infty} \sum_{k=0}^{\infty} \Pr[S_0 = k] \Pr[S_n - S_0 = m-k, \text{ for some } n \ge 0] \\
=&\sum_{k=0}^{\infty} \Pr[S_0 = k]\frac{1}{\mu} \\
=&\frac{1}{\mu},
\end{aligned}$$

where the second equality is by DCT, so we can put $\lim$ inside. □

3.2 Life Time, Age, Remaining Time and Renewal Equation

Consider a point m, and the *age* $A(m)$ at m is the time from the last renewal to m, that is, $m - S_{N(m)}$ and the *remaining time* $R(m)$ at m is the time from m to the next renewal, that is, $S_{N(m)+1} - m$, and we also have the life time $L(m) = A(m) + R(m)$.

Example 3.6. Suppose we have a renewal process with $X_i \in \{1, \ldots, k\}$. Is $R(0), R(1), \ldots$ a Markov Chain?

Consider its probability matrix, $P(1,i) = f_i$, because when $R(t) = 1$, at $t+1$, a renewal happens, so the remaining time is actually, X_{t+1}, which follows the distribution f_i. For $j > 1$, $P(j, j-1) = 1$ and 0 otherwise. We can consider the stationary distribution, where

$$\begin{aligned}
\pi_1 &= \pi_1 f_1 + \pi_2 \\
\pi_2 &= \pi_1(1 - f_1) \\
&= f_2\pi_1 + \pi_3 \\
\pi_3 &= \pi_2 - f_2\pi_1 = \pi_1(1 - f_1 - f_2),
\end{aligned}$$

and

$$\sum_{i=1}^{\infty} \pi_i = \sum_{i=1}^{\infty} \pi_1 \Pr[X_1 \geq i] = \pi_1 \sum_{i=1}^{\infty} \Pr[X_1 \geq i] = \pi_1 \mathbb{E}[X_1] = 1.$$

so $\pi_1 = 1/\mathbb{E}[X_1]$.

Notice that $u(m) = \Pr[R(m-1) = 1]$, where $R(m-1)$ is an aperiodic and irreducible Markov Chain, so

$$\lim_{m\to\infty} \Pr[R(m-1) = 1] \to 1/\pi_1 = 1/\mu = 1/\mathbb{E}[X_1],$$

which agrees EFP.

3.2.1 Renewal Equation

Consider bounded sequences $\{z(m), m \geq 0\}$ and $\{b(m), m \geq 0\}$, where the renewal equation is

$$z(m) = b(m) + \sum_{k=1}^{\infty} f_k z(m-k), m \geq 0.$$

Let $z(m) = b(m) = 0$ for $m < 0$,

$$\begin{aligned}
z(m) &= b(m) + \sum_{k=1}^{\infty} f_k z(m-k) \\
&= b(m) + \mathbb{E}[z(m-k)]
\end{aligned}$$

1. Set up renewal equation.
2. Solution is $z(m) = \sum_{k=0}^{\infty} b(m-k)u(k) = \sum_{k=0}^{m} b(m-k)u(k) + 0$
3. Key Renewal Theorem: if $\sum_{k=0}^{\infty} |b(k)| < \infty$, then

$$\lim_{m\to\infty} z(m) = \sum_{k=0}^{\infty} b(k)/\mu.$$

Now we consider one renewal equation:

$$\begin{aligned}
z(m) &= \Pr[A(m) = r] \\
z(m) &= \sum_{k=m+1}^{\infty} f_k \Pr[A(m) = r | X_1 = k] + \sum_{k=1}^{m} f_k \Pr[A(m) = r | X_1 = k]. \\
&= \mathbb{1}\{m = r\} \sum_{k=m+1}^{\infty} f_k + \sum_{k=1}^{m} f_k z(m-k).
\end{aligned}$$

notice when $k > m$, $A(m) = m - 0 = m$, so $A(m) = r$ if and only if $m = r$. Let $\delta_{m,r} = \mathbb{1}\{m = r\}$. We then have

$$b(m) = \mathbb{1}\{m = r\} \sum_{k=m+1}^{\infty} f_k = \sum_{k=m+1}^{\infty} f_k \delta_{m,r}$$

from renewal equation.

Example 3.7. Consider a renewal process as a $\mathrm{Bernoulli}(p)$, that is, we have a renewal when the Bernoulli hits a success. Then $u(0) = 1$ by starting with a success and $u(k) = p$, and $X_i \sim \mathrm{Geo}(p)$, we have that, by EFP,

$$u(m) \to 1/p$$

Example 3.8. Consider a renewal process where $X_i \in \{1, 2\}$ and $f_1 = f_2 = 1/2$, from Lalley's notes,

$$\begin{aligned}
u(m) &= \frac{2}{3} + \frac{1}{3}(-1/2)^m \to 2/3 \\
u(1) &= 2/3 - 1/3(1/2)^1 = 2/3 - 1/6 = 1/2 \\
u(2) &= f_2 + f_1^2 = 3/4.
\end{aligned}$$

We can now start to try to solve the renewal equation.

Example 3.9. Back to the example above with Bernoulli process.

$$\begin{aligned}
z(m) = \Pr[A(m) = r] &= \sum_{k=0}^{\infty} b(m-k)u(k) \\
&= \sum_{k=0}^{m} b(m-k)u(k) + 0 \\
&= b(m) + p\sum_{k=1}^{m} b(m-k) \\
&= \sum_{k=m+1}^{\infty} f_k \delta_{m,r} + p\sum_{k=1}^{m} b(m-k)
\end{aligned}$$

say $m = 7$, $r = 5$, so $z(m) = \Pr[A(7) = 5]$, and $b(m-k) = \sum_{j=m-k+1}^{\infty} f_j \delta_{m-k,r}$, where $b(5) = \sum_{k=6}^{\infty} f_k \delta_{5,5} = \Pr[X_1 > 5] = (1-p)^5$. We have the above equation becoming:

$$\begin{aligned}
&= 0 + p\sum_{k=1}^{7} b(7-k) \\
&= 0 + pb(7-2) \\
&= p(1-p)^5
\end{aligned}$$

which makes sense because $A(7) = 5$ is saying the last renewal is at 2 and there is no renewal at $3, 4, 5, 6, 7$, and the probability for that is $p(1-p)^5$.

Another question we have is that for $z(m) = \Pr[A(m) = r]$, do we have

$$\sum_{k=0}^{\infty} |b(k)| < \infty?$$

The answer is yes.

$$\begin{aligned}
\sum_{m=0}^{\infty} |b(m)| = \sum_{m=0}^{\infty} b(m) &= \sum_{m=0}^{\infty} \sum_{k=m+1}^{\infty} f_k \delta_{m,r} \\
&= \sum_{k=1}^{\infty} f_k \sum_{m=0}^{k-1} \delta_{m,r} \\
&= \sum_{k=r+1}^{\infty} f_k = \Pr[X_1 > r] < \infty
\end{aligned}$$

Then by Key Renewal Theorem, we have

$$z(m) = \Pr[A(m) = r] \to \frac{\Pr[X_1 > r]}{\mu}.$$

For our Bernoulli example again, with $z(m) = \Pr[A(m) = r] = p(1-p)^r$, where

$$\frac{\Pr[X_1 > r]}{\mu} = \frac{(1-p)^r}{1/p} = p(1-p)^r$$

which agrees with what we have.

Now we consider another renewal equation, where $z(m) = \Pr[L(m) = r]$. Then

$$z(m) = \sum_{k=m+1}^{\infty} f_k \mathbb{1}\{k = r\} + \sum_{k=1}^{m} f_k z(m-k)$$

where for the first term, given $X_1 = k$, by $m \in \{0, 1, \dots, k\}$, we know $L(m) = k$, so $L(m) = r$ is equivalent to saying the process renewing at $k = r$.

3.3 $\mathbb{E}[N(m) - m/\mu]$

Recall that $\mathbb{E}[N(m)/m] = 1/\mu$ by the elementary renewal theorem, so we can approximate $\mathbb{E}[N(m)]$ by

$$\mathbb{E}[N(m)] \approx m/\mu.$$

Consider $z(m) = \mathbb{E}[N(m)] - m/\mu$, and then

$$z(m) = \sum_{k=m+1}^{\infty} f_k \mathbb{E}[N(m) - m/\mu | X_1 = k] + \sum_{k=1}^{m} f_k \mathbb{E}[N(m) - m/\mu | X_1 = k]$$

Note, when $k > m$, $\mathbb{E}[N(m)|X_1 = k] = 0$, so the first term is $\sum_{k=m+1}^{\infty} f_k(-m/\mu)$. For the second one, since $m \geq k$, $\mathbb{E}[N(m)|X_1 = k] = 1 + \mathbb{E}[N(m-k)]$, and so

$$\mathbb{E}[N(m) - m/\mu|X_1 = k] = 1 + \mathbb{E}[N(m-k)] - \frac{m-k}{\mu} + \frac{m-k}{\mu} - \frac{m}{\mu}$$

So the second term becomes

$$\begin{aligned}
&\sum_{k=1}^{m} f_k \left(1 + z(m-k) + \frac{m-k}{\mu} - \frac{m}{\mu}\right) \\
&= \sum_{k=1}^{m} f_k \left(1 + \frac{m-k}{\mu} - \frac{m}{\mu}\right) + \sum_{k=1}^{m} f_k z(m-k) \\
&= \sum_{k=1}^{m} f_k \left(1 - \frac{k}{\mu}\right) + \sum_{k=1}^{m} f_k z(m-k) \\
&= \left(\sum_{k=1}^{m} f_k\right) - 1 + \frac{\mu - \sum_{k=1}^{m} k f_k}{\mu} + \sum_{k=1}^{m} f_k z(m-k) \\
&= \sum_{k=m+1}^{\infty} (k/\mu - 1) f_k + \sum_{k=1}^{m} f_k z(m-k)
\end{aligned}$$

so

$$z(m) = \underbrace{\sum_{k=m+1}^{\infty} \left(\frac{k-m}{\mu} - 1\right) f_k}_{b(m)} + \sum_{k=1}^{m} f_k z(m-k).$$

Then we assume $\mathbb{E}[X_i^2]$ or $\mathrm{var}(X_i)$ is finite and apply Key Renewal Theorem. Then

$$\begin{aligned}
\sum_{m=0}^{\infty} |b(m)| &\leq \sum_{m=0}^{\infty} \sum_{k=m+1}^{\infty} \underbrace{f_k}_{\geq 0} \underbrace{\left(\frac{k-m}{\mu} + 1\right)}_{>0} \\
&= \sum_{k=1}^{\infty} f_k \left(\sum_{m=0}^{k-1} \frac{k-m}{\mu} + 1\right) \\
&= \sum_{k=1}^{\infty} f_k \left(\frac{k}{\mu} + \ldots + \frac{1}{\mu} + k\right) \\
&= \sum_{k=1}^{\infty} f_k \frac{k(k+1)}{2\mu} + \sum_{k=1}^{\infty} k f_k \\
&= \sum_{k=1}^{\infty} \frac{k^2 f_k}{2\mu} + \frac{\sum_{k=1}^{\infty} k f_k}{2\mu} + \mu \\
&= \frac{\sigma^2 + \mu^2}{2} + \frac{1}{2} + \mu < \infty,
\end{aligned}$$

then KRT says,

$$
\begin{aligned}
z(m) \to \frac{\sum_{k=0}^{\infty} b(k)}{\mu} = \sum_{m=0}^{\infty} \sum_{k=m+1}^{\infty} \frac{f_k}{\mu} \left(\frac{k-m}{\mu} - 1 \right) &= \sum_{k=1}^{\infty} \frac{f_k}{\mu} \sum_{m=0}^{k-1} \left(\frac{k-m}{\mu} - 1 \right) \\
&= \sum_{k=1}^{\infty} \frac{f_k}{\mu} \left(\frac{k(k+1)}{2\mu} - k \right) \\
&= \frac{\sigma^2 + \mu^2 + \mu}{2\mu^2} - \frac{\mu}{\mu} \\
&= \frac{\sigma^2 - \mu^2 + \mu}{2\mu}
\end{aligned}
$$

so $\mathbb{E}[N(m) - m/\mu] \to \frac{\sigma^2 - \mu^2 + \mu}{2\mu^2}$.

Note it "seems" like $\mathbb{E}[N(m)] = 1 * u_1 + 1 * u_2 + \ldots + 1 * u_m = mp = \frac{m}{\mu}$ as $u(m) \to \frac{1}{\mu}$, but is it true? Do we always have $\mathbb{E}[N(m) - m/\mu] \to 0$?

Example 3.10. Consider Bernoulli renewal, $\mathrm{Bernoulli}(P)$, where $X_1 \sim \mathrm{Geo}(p)$, $\mu = 1/p$ and $\sigma^2 = q/p^2$. Then

$$
\frac{\frac{q}{p^2} - \frac{1}{p^2} + \frac{p}{p^2}}{2/p^2} = 0.
$$

Example 3.11. $f_1 = f_2 = \frac{1}{2}$. Then recall we condition on the first renewal time,

$$
u(m) = \delta_{0,m} + \sum_{k=1}^{m} f_k u(m-k)
$$

by the fact that the first renewal is at $k \geq m$, then having a renewal at m if and only if $m = k$. Hence,

$$
\begin{aligned}
u(m) &= \frac{1}{2} u(m-1) + \frac{1}{2} u(m-2) \\
0 &= u(m) - \frac{1}{2} u(m) - \frac{1}{2} u(m-2) \\
0 &= r^2 - \frac{1}{2} r - \frac{1}{2} \\
0 &= (r-1)(r + \frac{1}{2}) \\
u(m) &= c_1 1^m + c_2 (-\frac{1}{2})^m = c_1 + c_2 (-1/2)^m \\
u(0) &= 1 = c_1 + c_2 \\
u(1) &= \frac{1}{2} = c_1 - \frac{1}{2} c_2 \\
\frac{1}{2} &= \frac{3}{2} c_2
\end{aligned}
$$

so $c_2 = \frac{1}{3}$, $c_1 = \frac{2}{3}$. Now $\mu = \frac{1}{2} * 1 + \frac{1}{2} * 2 = \frac{3}{2}$ and $\sigma^2 = \frac{1}{4}$ (same as $\mathrm{Bernoulli}(1/2)$ shifted by 1). Then

$$\mathbb{E}[N(m) - m/\mu] \to \frac{1/4 - 9/4 + 6/4}{2 * 9/4} = -1/9.$$

and

$$\begin{aligned}
\mathbb{E}[N(m) - m/\mu] &= \mathbb{E}[N(m)] - 2m/3 \\
&= \sum_{k=1}^{m} u(k) - 2m/3 \\
&= \sum_{k=1}^{m} (2/3 + (1/3)(-1/2)^k) - 2m/3 \\
&= 2m/3 + \frac{1}{3}\sum_{k=1}^{m} (-1/2)^k - 2m/3 \\
&= \frac{1}{3}\sum_{k=1}^{m} (-1/2)^k \\
&= \frac{1}{3}\frac{-1}{2}\sum_{k=0}^{m-1} (-1/2)^j \to \frac{1}{3}(-1/2)\frac{1}{1+\frac{1}{2}} = -\frac{1}{9}.
\end{aligned}$$

In fact, we can say

$$\mathbb{E}[N(m)] \to m/\mu + \frac{\sigma^2 - \mu^2 + \mu}{2\mu^2} + o(1).$$

Example 3.12. Consider the following example:

$z(0)$	$z(1)$	$z(2)$	$z(3)$	$z(4)$	$z(5)$
0	12	18	24	20	21

and $f_1 = 3/6$, $f_2 = 2/6$, $f_3 = 1/6$. This $z(m)$ is constructed as, $z(1), z(2), z(3)$ are given, $z(m) = f_1 z(m-1) + f_2 z(m-2) + f_3 z(m-3)$. Is this z a renewal equation? Define $z(0) = b(0) = 0$. We have $z(1) = b(1) + \sum_{k=1}^{1} f_k z(1-k)$.

$$\begin{aligned}
12 = z(1) &= b(1) + f_1 z(0) \implies b(1) = 12 \\
z(2) &= b(2) + f_1 z(1) + f_2 z(0) \\
z(3) &= b(3) + f_1 z(2) + f_2 z(1) + f_3 z(0) \\
z(4) &= b(4) + f_1 z(3) + f_2 z(2) + f_3 z(1) \\
&\implies b(4) = 0, b(5) = 0,
\end{aligned}$$

so $z(2) = b(2) + \frac{3}{6}12 + \frac{2}{6}0 \implies b(2) = 12$. Similarly, $b(3) = 11$.

$$\sum_{m=0}^{\infty} b(m) = 12 + 12 + 11 + 35$$

because by the way $z(m)$ being constructed, $b(m) = 0$ for all $m \geq 4$.

Proof of Key Renewal Theorem. Note $b(m) = z(m) = 0$ for $m \leq -1$. First we have that $\mathbb{E}[z(m - X_1)] = b(m - X_1) + \mathbb{E}[z(m - X_1 - X_2)]$. And hence

$$\begin{aligned}
z(m) &= b(m) + \mathbb{E}[z(m - X_1)] \\
&= b(m) + \mathbb{E}[b(m - X_1)] + \mathbb{E}[z(m - X_1 - X_2)] \\
&= b(m) + \mathbb{E}[b(m - X_1)] + \mathbb{E}[z(m - X_1 - X_2)] \\
&= b(m) + \mathbb{E}[b(m - X_1)] + \mathbb{E}[b(m - X_1 - X_2)] + \ldots + \mathbb{E}[b(m - X_1 - \ldots - X_m)]
\end{aligned}$$

because $m - X_1 - X_2 - \ldots - X_m - X_{m+1} < 0$. Now

$$\begin{aligned}
z(m) = \sum_{n=0}^{m} \mathbb{E}[b(m - S_n)] &= \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} [b(m-k)] \Pr[S_n = k] \\
&= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} b(m-k) \Pr[S_n = k] \\
z(m) &= \sum_{k=0}^{\infty} b(m-k) u(k) \\
&= \sum_{k=0}^{\infty} b(k) u(m-k).
\end{aligned}$$

With this,

$$\begin{aligned}
\lim_{m\to\infty} z(m) &= \lim_{m\to\infty} \sum_{k=0}^{\infty} b(k) u(m-k) \\
&= \sum_{k=0}^{\infty} b(k) \underbrace{\lim_{m\to\infty} u(m-k)}_{\to 1/\mu} \\
&= \sum_{k=0}^{\infty} b(k)/\mu
\end{aligned}$$

where the second equality is by $\sum_{k=0}^{\infty} b(k)u(m-k) \leq \sum_{k=0}^{\infty} |b(k)|$ since $u(m-k) \in [0,1]$ and dominant convergence theorem. ☐

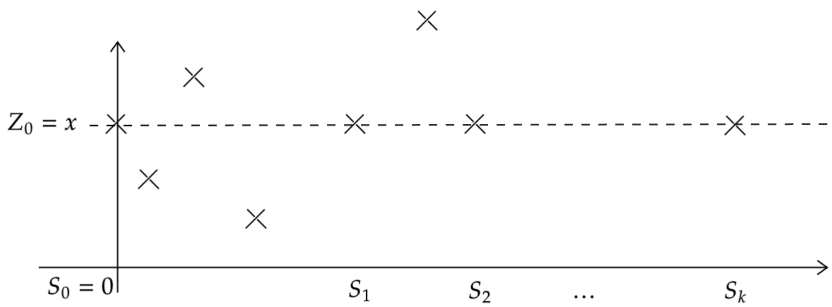
3.4 Regenerative Processes

Definition 3.13. Let $\{Z_m, m \geq 0\}$ be a stochastic process with arbitrary state space. Let $S_0, S_1, \ldots$ be a possibly delayed arithmetic renewal sequence with span 1. The process $Z_0, Z_1, \ldots$ is *regenerative* w.r.t. $S_0, S_1 \ldots$ if

1. $\{Z_{n+S_k}, n \geq 0\}$ is independent of $S_0, X_1, \ldots, X_k$ (stronger version requires everything before S_k).
2. $\{Z_{n+S_k}, n \geq 0\} \overset{D}{=} \{Z_{n+S_0}, n \geq 0\}$.

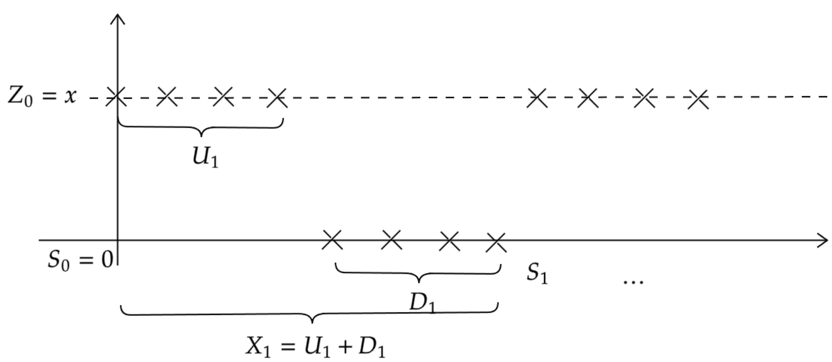
For the continuous case, we consider $\{Z_t, t \in \mathbb{R}_+\}$ and $\{Z_{t+S_k}, t \in \mathbb{R}_+\}$ instead.

Example 3.14. Suppose $Z_0, Z_1, \ldots$ is an irreducible recurrent Markov Chain. And let S_i be the i-th time the Markov Chain go back to x.



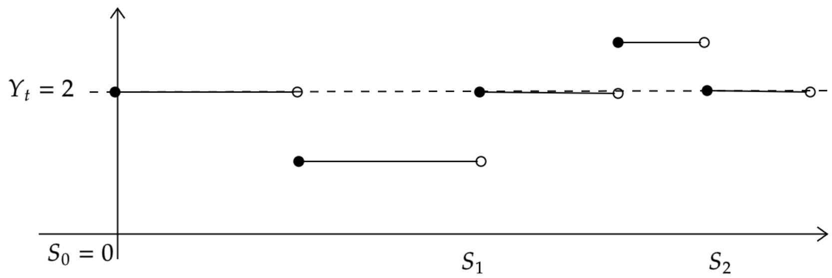
so this is a regenerative process by checking both condition 1) and 2) basically by the Strong Markov Theorem, that once we return to x, the process "restarts". We can also delay it by letting Markov Chain start with some other value, then S_0 is the first time it goes to x.

Example 3.15 (Reliability)**.** Consider a machine which has working (up) and not working (down) time, then we can consider the time it gets back to work as a regeneration.



where $U_1, D_1, \ldots$ be independent; $U_1, U_2, \ldots$ be i.i.d. and integer-valued and $D_1, D_2, \ldots$ be i.i.d. we also want $d = \max\{h : U_i + D_i \in \{h, 2h, \ldots\}\} = 1$. For the continuous case, just let U_i, D_i be a continuous time interval.

Example 3.16 ($M/M/1$ Queue)**.** Let Y_t be the number in the system at time t.



The behavior after S_k is probabilistic the same as the one after S_0.

Example 3.17 (Difference between independence of $S_0, X_1, \ldots, X_k$ and everything before S_k)**.** Consider independent copies of Markov Chains with probability matrix P. Denote the copies as:

$$\begin{gathered} x_{11} = 0, x_{12}, \ldots \\ x_{21} = 0, x_{22}, \ldots \\ \vdots \end{gathered}$$

Consider some independent generated renewal times which are arithmetic with span 1. Then starting from S_0 to $S_1 - 1$, we consider $Z_{S_0} = x_{11}, \ldots, Z_{S_1-1} = x_{1S_1}$, then for S_1, renew as x_{21} and then keep doing and get $Z_{S_2-1} = x_{2(S_2-S_1)}$. This is a regenerative process because the distribution after S_k is probabilistic same to the one after S_0 and independent of $S_0, X_1, \ldots, X_k$.

Modify it: let $r \in (0,1)$. Now at S_1, with probability r, we use $x_{11}, \ldots$, with probability $(1-r)$, use $x_{21}, \ldots$. That is, every time it renews, there is a probability r to reuse the previous one. But then, Z_{n+S_k} is not independent from everything before S_k but it is actually independent from $S_0, X_1, \ldots, X_k$. But 2) is satisfied, once we renew, it is just following a Markov Chain with probability matrix P.

Theorem 3.18 (Regenerative Theorem)**.** Let $Z_0, Z_1, \ldots$ be a non-delayed regenerative process, and let f be a real valued function such that

$$\mathbb{E}\left[\sum_{n=0}^{S_1-1} |f(Z_n)|\right] < \infty$$

and the renewal time is arithmetic with span 1, then

$$\lim_{n\to\infty} \mathbb{E}[f(Z_n)] = \frac{\mathbb{E}[\sum_{n=0}^{S_1-1} f(Z_n)]}{\mathbb{E}[S_1]}$$

Example 3.19. Consider a regenerative process which regenerates whenever an irreducible positive recurrent and aperiodic Markov Chain reaches x and starting with x. Let $f(Z_m) = \mathbb{1}\{Z_m = x\}$. Now from what we know from Markov Chain:

$$\mathbb{E}[f(Z_m)] = \Pr_x[Z_m = x] \to \pi_x$$

From the regeneraitive theorem, we know

$$\lim_{m\to\infty} \mathbb{E}[f(Z_m)] = \frac{\mathbb{E}[\sum_{m=0}^{S_1-1} f(Z_m)]}{\mathbb{E}[S_1]} = \frac{1}{1/\pi_x} = \pi_x$$

Example 3.20. For the reliability example. We let the "up" condition be 1, "down" be 0, then

$$\lim_{m\to\infty} \Pr[Z_m = 1] = \frac{\mathbb{E}[U_1]}{\mathbb{E}[U_1 + D_1]}$$

if $U_1 \sim \text{Geo}(p)$, then $D \sim \text{Geo}(1-p)$, so

$$\lim_{m\to\infty} \Pr[Z_m = 1] = \frac{1/p}{1/p + 1/(1-p)}.$$

Proof of Regenerative Theorem. For the arithmetic case:

$$\begin{aligned} z(m) &= b(m) + \mathbb{E}(z(m - X_1)) \\ &= b(m) + \sum_{k=1}^{m} f_k z(m-k) \end{aligned}$$

Let $z(m) = \mathbb{E}[f(z_m)]$, then

$$\begin{aligned} \mathbb{E}[f(z_m)] &= \mathbb{E}[f(z_m)\mathbb{1}\{m < X_1\}] + \mathbb{E}[f(z_m)\mathbb{1}\{X_1 \le m\}] \\ &= \underbrace{\mathbb{E}[f(z_m)\mathbb{1}\{m < X_1\}]}_{b(m)} + \sum_{k=1}^{m} f_k \mathbb{E}[f(z_{m-k})] \end{aligned}$$

For KRT, we need

$$\begin{aligned} &\sum_{m=0}^{\infty} |b(m)| \\ \le &\sum_{m=0}^{\infty} \mathbb{E}[|f(z_m)|\mathbb{1}\{m < X_1\}] \\ = &\mathbb{E}\left[\sum_{m=0}^{\infty} |f(z_m)|\mathbb{1}\{m < X_1\}\right] \\ = &\mathbb{E}\left[\sum_{m=0}^{X_1-1} f(z_m)\right] < \infty \end{aligned}$$

which is given, then KRT implies

$$\lim_{m\to\infty} \mathbb{E}[f(z_m)] = \frac{\mathbb{E}[\sum_{m=0}^{X_1-1} f(z_m)]}{\mathbb{E}[X_1]}$$

□

3.5 Delayed Regenerative Theorem

For a delayed regenerative process, we have

$$\mathbb{E}[f(z_m)] = \mathbb{E}[f(z_m)\mathbb{1}\{m \le S_0\}] + \mathbb{E}[f(z_m)\mathbb{1}\{S_0 < m\}].$$

To use KRT, we need $\mathbb{E}\left[\sum_{k=S_0}^{S_1-1}|f(z_k)|\right] < \infty$. We also need $|f(z_m)| \le b$ for some b, $\lim_{m\to\infty}\mathbb{E}[f(z_m)\mathbb{1}\{m \le S_0\}] = 0$. A sufficient condition is that f is a bounded function and we can apply DCT.

3.6 Renewal Equations: Non-arithmetic, non-lattice

When we say arithmetic, we mean $\Pr[X_1 \in \{0, h, 2h, \ldots\}] = 1$ for some positive integer h.

For non-arithmetic case, let X_1 be a continuous random variable defined by $F(t) = \Pr[X_1 \le t]$, $F(0) < 1$, $\Pr[X_1 \ge 0] = 1$. For the recurrent case, we also have $F(\infty) = 1$. Then the renewal equation is set up as: $z(t) = b(t) = 0$ if $t < 0$, and

$$z(t) = b(t) + \int_{[0,t]} dF(s)\, z(t-s).$$

If $b(t)$ is a bounded function, the unique solution to the renewal equation is

$$z(t) = \int_{[0,\infty)} du(s)\, b(t-s).$$

Let $N(t)$ be the number of renewals in $(0, t]$, then the renewal measure is defined as

$$u(t) = 1 + \mathbb{E}[N(t)]$$

For example, for a Poisson Process with rate $\lambda > 0$, we have the renewal measure

$$u(t) = 1 + \lambda t$$

And *Blackwell's Renewal Theorem* gives

$$\lim_{t\to\infty} u(t+h) - u(t) = h/\mu.$$

3.6.1 KRT non-arithmetic

Let X_1 be some continuous random variable. If $b(t)$ is directly Riemann Integrable (dRi), then

$$\lim_{t\to\infty} z(t) = \frac{\int_0^\infty b(t)dt}{\mu}.$$

dRi: $b(t) = 0$, for $t < 0$, and

$$\lim_{a\to 0^+}\sum_{n=1}^{\infty} a \inf_{(n-1)a\le t\le na} f(t) = \lim_{a\to 0^+}\sum_{n=1}^{\infty} a \sup_{(n-1)a\le t\le na} f(t) = \int_0^\infty b(t)dt$$

then b is *dRi*. The difference between this condition and Riemann integrable is that Riemann integral usually uses on finite interval then to infinity, this one is directly applied to infinity.

Proposition 3.21. If either one the following

1. $b(t)$ is nonnegative, nonincreasing and $\int_0^\infty b(t)dt < \infty$.
2. the support of $b(t)$ is compact and $b(t)$ is Riemann Integrable,

then $b(t)$ is dRi.

Lemma 3.22. For a continuous random variable,

$$X = \int_0^\infty (1 - \mathbb{1}\{X \le t\}dt = \int_0^\infty \mathbb{1}\{X > t\}dt = \int_0^X 1dt$$

and

$$\mathbb{E}[X] = \mathbb{E}\int_0^\infty \mathbb{1}\{X > t\}dt = \int_0^\infty \mathbb{E}[\mathbb{1}\{X > t\}]dt = \int_0^\infty \Pr[X > t]dt = \int_0^\infty (1 - F(t))dt$$

where the second equality is by Fubini's Theorem.

Example 3.23. Let $z(t) = \Pr[R(t) > y]$, where $R(t) = S_{N(t)+1} - t$ is the remaining time from t to the next renewal after t. Then if we condition on the first renewal after S_0, then we set up the renewal equation as:

$$z(t) = \Pr[X_1 > t + y] + \int_{[0,t]} dF(s)z(t-s) = \underbrace{1 - F(t+y)}_{b(t)} + \int_{[0,t]} dF(s)\, z(t-s)$$

Note $b(t) \ge 0$ and non-increasing, and by the above lemma, we have

$$\int_0^\infty 1 - F(t+y)dt \le \int_0^\infty (1 - F(t))dt = \mu < \infty$$

so it is dRi, then we have

$$\lim_{t\to\infty} z(t) = \lim_{t\to\infty} \Pr[R(t) > y] = \frac{\int_y^\infty (1 - F(t))dt}{\mu}$$

Corollary 3.24. Consider $\mathbb{E}[f(z_t)] = \underbrace{\mathbb{E}[f(z_t)\mathbb{1}\{t < X_1\}]}_{b(t)} + \int_{[0,t]} dF(s)\; z(t-s)$. If $b(t)$ is dRi, then

$$\lim_{t\to\infty} \mathbb{E}[f(z_t)] = \frac{\mathbb{E}\left[\int_0^{X_1} f(z_s)ds\right]}{\mu}$$

Example 3.25. Let $z(t) = \Pr[A(t) > x, R(t) > y]$, $\mathbb{E}[X_1] = \mu < \infty$. $\Pr[X_1 \le t] = F(t)$,

$$\begin{aligned} z(t) &= b(t) + \mathbb{E}[z(t - X_1)] = (1 - F(t+y))\mathbb{1}\{t \in (x,\infty)\} + \mathbb{E}[z(t - X_1)] \\ &= (1 - F(t+y))\mathbb{1}\{t \in (x,\infty)\} + \int_{[0,t]} dF(s)\; z(t-s) \end{aligned}$$

where the first term is when S_1 is greater than t, that means $A(t) = t$ and $R(t) > y$ is equivalent to $X_1 > t + y$. For the second term, the first renewal is before t, so the age becomes $t - X_1$ and remaining time is $R(t - X_1)$.

Then is $b(t)$ dRi?. It is zero before x and starts decreasing, so we know $b(t)$ is non-negative, "nonincreasing" (form x to ∞) and $\int_0^\infty b(t) < \infty$, then dRi, where

$$\int_0^\infty b(t)dt = \int_x^\infty 1 - F(t+y) \leq \int_0^\infty (1 - F(t))dt = \mu < \infty$$

so

$$\Pr[A(t) > x, R(t) > y] \to \frac{\int_{x+y}^\infty (1 - F(u))du}{\mu} \text{ as } t \to \infty$$

4 Some Other Examples

Example 4.1. For a $M/G/1$ queue, let V_t be the virtual waiting time at times t. How can we find regenerative processes?

The arrival to the empty system and departure leaving an empty system are regenerative processes. The reason is that we know the arrival to the empty system is always a regenerative one because the service time are i.i.d., and the departure is also one because of the memoryless property of a Markov Chain.

For a $G/G/1$ queue: arrivals to an empty system gives an regen process. If you consider the departure time leaving to an empty system, then suppose the service time is deterministic, we will know the arrival time once we know the departure time.

For $G/M/k$, you don't have memoryless property for interarrival time, but we can have departure left an emptys system as a regen process. For example, if it is $D/M/k$, we can know when the least, next arrival is, because the arrival time is deterministic.

For $M/G/2$, we don't want consider anything in the middle of service. So arrival to empty system and departure leaving empty system are two regen processes.

Example 4.2. Given two unknown unequal number x, y, we can see one of them x, I can tell if this one is larger with probability $> 1/2$.

Let $z \in N(0, 1)$, treat z as the hidden number, that is, if $z > x$, I say y is larger, $z < x$ then y is smaller. Hence, if $z > \max(x, y)$ or $z < \min(x, y)$, then we have $1/2$ chance to be correct depending on if revealed one is the smaller or larger one. If $z \in [x, y]$, then we are always correct.

Example 4.3. Suppose we are generating letters i.i.d following a uniform distribution (each letter with prob $1/26$). How long do we expect "miami" to be typed?

Consider a fair casino which stops once "miami" appear. At each time unit, one person came in with $\$1$, and they always bet all money they have, and always bet on "miami", that is, suppose a person enters the casino at 2, then they would bet m at time 2, i at time 3, a at time 4 and so on. If they got correct, they receive the amount they bet, otherwise, they lose everything and leave. Notice that for a fair casino, since every person has $1/26$ chance to get the correct letter, the return is 26 times the bet.

Let T be the first time "miami" appears, so $T - 4$ is m, $T - 3$ is i, $T - 2$ is a, $T - 1$ is m, and T is i. The one enters the casino at $T - 4$ got all correct, so they receive $\$26^5$, and the one enters at $T - 1$ got the first two letter correct, so receive $\$26^2$, and all other people lose their money.

Hence, the money the casino expect to pay is $26^5 + 26^2$ which is the expect amount it receives becausre it's a fair casino. Also, the amount it expects to receive is the $\$1$ every customer brings, so $\mathbb{E}[T] = 26^5 + 26^2$. is the expected time for "miami" to be typed.

In the sense of matingale, let $S_0 = 0$ and S_n be the change in casino bankroll. Then $S_n = X_1, \ldots + X_n$ with $\mathbb{E}[X_i] = 0$ because all bets are fair. Since $\mathbb{E}[S_{n+1}|S_n] = S_n$, it is a martingale. By $\mathbb{E}[T] < \infty$, $\mathbb{E}[S_T] = \mathbb{E}[S_0] = 0$.