MATH 170B LECTURE NOTE 6: ELEMENTARY STOCHASTIC PROCESSES

HANBAEK LYU

6. How can we use sequence of RVs to model real life situations?

Say we would like to model the USD price of bitcoin. We could observe the actual price at every hour and record it by a sequence of real numbers $x_1, x_2, \cdots$. However, it is more interesting to build a 'model' that could predict the price of bitcoin at time t, or at least give some meaningful insight on how the actual bitcoin price behaves over time. Since there are so many factors affecting its price at every time, it might be reasonable that its price at time t should be given by a certain RV, say X_t. Then our sequence of predictions would be a sequence of RVs, $(X_t)_{t\geq 0}$. This is an example of what is called a *stochastic process*. Here 'process' means that we are not interested in just a single RV, that their sequence as a whole: 'stochastic' means that the way the RVs evolve in time might be random.

In this section, we will be studying three elementary stochastic processes: 1) Bernoulli process, 2) Poisson process, and 3) discrete-time Markov chain.

6.1. The Bernoulli processes. Let $(X_t)_{t\geq 1}$ be a sequence of i.i.d. Bernoulli(p) variables. This is the *Bernoulli process* with parameter p, and that's it. Considering how simple it is conceptually, we can actually ask a lot of interesting questions about it.

First we envision this as a model of customers arriving at a register. Suppose a clerk rings a bell whenever she is done with her current customer or ready to take the next customer. Upon each bell ring, a customer arrives with probability p or no customer gets there with probability $1-p$, independently at each time. Then we can think of the meaning of X_t as

$$X_t = \mathbf{1}(\text{a customer arrives at the register after } t \text{ bell rings}). \tag{1}$$

To simplify terminology, let 'time' be measured by a nonnegative integer $t \in \mathbb{Z}_{\geq 0}$: time t means the time right after tth bell ring. Here are some of the *observables* for this process that we are interested in:

$$S_n = X_1 + \cdots + X_n = \#(\text{customers arriving at the register up to time } n) \tag{2}$$

$$T_i = \text{time that the } i\text{th customer arrives.} \tag{3}$$

$$\tau_i = T_i - T_{i-1} = \text{the inter-arrival time between the } i-1\text{st and } i\text{th customer.} \tag{4}$$

We also define $\tau_1 = T_1$. See Figure 1 for an illustration.

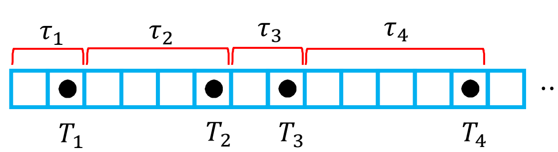

FIGURE 1. Illustration of Bernoulli process. First four customers arrive at times $T_1 = 2$, $T_2 = 6$, $T_3 = 8$, and $T_4 = 13$. The inter-arrival times are $\tau_1 = 2$, $\tau_2 = 4$, $\tau_3 = 2$, and $\tau_4 = 5$. There are $S_7 = 2$ customers up to time $t = 7$.

Exercise 6.1. Let $(X_t)_{t\geq 1}$ be a Bernoulli process with parameter p.

(i) Show that $S_n \sim \text{Binomial}(n, p)$.
(ii) Show that $T_1 \sim \text{Geom}(p)$.
(iii) Show that τ_i's are i.i.d. with distribution Geom(p).

Exercise 6.2 (Independence)**.** Let $(X_t)_{t\geq 1}$ be a Bernoulli process with parameter p. Show the following.

(i) Let U and V be the number of customers at times $t \in \{1, 2, \cdots, 5\}$ and $t \in \{6, 7, \cdots, 10\}$, respectively. Show that U and V are independent.

(ii) Let U and V be the first odd and even time that a customer arrives, respectively. Show that U and V are independent.

(iii) Let S_5 be the number of customers up to time $t = 5$ and let $\tau_3 = T_3 - T_2$ be the inter-arrival time between the second and third customers. Are S_5 and τ_3 independent?

Let $(X_t)_{t\geq 1}$ be a Bernoulli process with parameter p. If we discard the first 5 observations and start the process at time $t = 6$, then the new process $(X_t)_{t\geq 6}$ is still a Bernoulli process with parameter p. Moreover, the new process is independent on the past RVs $X_1, X_2, \cdots, X_5$. The following exercise generalizes this observation.

Exercise 6.3. Let $(X_t)_{t\geq 1}$ be a Bernoulli process with parameter p. Show the following.

(i) (Renewal property of Bernoulli RV) For any integer $k \geq 1$, $(X_t)_{t\geq k}$ is a Bernoulli process with parameter p and it is independent from $X_1, X_2, \cdots, X_{k-1}$.

(ii) (Memoryless property of Geometric RV) For any integer $k \geq 1$, let $\bar{T}$ be the first time that a customer arrives after time $t = k$. Show that $\bar{T} - k \sim \text{Geom}(p)$ and it is independent from $X_1, X_2, \cdots, X_k$. (hint: use part (i))

Exercise 6.4 (Alternative definition of Bernoulli process)**.** In this exercise, we show that the Bernoulli processes can be characterized in terms of their inter-arrival times. Let $(\tau_k)_{k\geq 0}$ be a sequence of i.i.d. Geom(p) variables. Define a sequence $(X_t)_{t\geq 0}$ of indicator RVs by

$$X_t = \mathbf{1}(\tau_1 + \cdots + \tau_k = t \text{ for some } k \geq 1). \tag{5}$$

(i) Show that $X_1 \sim \text{Bernoulli}(p)$.

(ii) Use the memoryless property of Geometric RVs to show show that $X_2 \sim \text{Bernoulli}(p)$ and that X_2 is independent of X_1.

(iii) Use the memoryless property of Geometric RVs to show show that $X_t \sim \text{Bernoulli}(p)$ and that X_t is independent of $X_1, \cdots, X_{t-1}$.

(iv) Conclude that $(X_t)_{t\geq 0}$ is a Bernoulli process with parameter p.

Example 6.5 (Renewal property at a random time)**.** Let $(X_t)_{t\geq 1}$ be a Bernoulli process with parameter p. Suppose N is the first time that we see two consecutive customers, that is,

$$N = \min\{k \geq 2 \,|\, X_{k-1} = X_k = 1\}. \tag{6}$$

Then what is the probability $\mathbb{P}(X_{N+1} = X_{N+2} = 0)$ that no customers arrive at times $t = N+1$ and $t = N+2$? Intuitively, what's happening after time $t = N$ should be independent from what happened up to time $t = N$, so we should have $\mathbb{P}(X_{N+1} = X_{N+2} = 0) = (1 - p^2)$. However, this is not entirely obvious since N is a random time.

Observe that the probability $\mathbb{P}(X_{N+1} = X_{N+2} = 0)$ depends on more than two source of randomness: N, X_{N+1}, and X_{N+2}. Our principle to handle this kind of situation was to use conditioning:

$$\mathbb{P}(X_{N+1} = X_{N+2} = 0) = \sum_{n=1}^{\infty} \mathbb{P}(X_{n+1} = X_{n+2} = 0 \,|\, N = n)\mathbb{P}(N = n) \tag{7}$$

$$= \sum_{n=1}^{\infty} \mathbb{P}(X_{n+1} = X_{n+2} = 0)\mathbb{P}(N = n) \tag{8}$$

$$= \sum_{n=1}^{\infty} (1-p)^2 \mathbb{P}(N = n) = (1-p)^2 \sum_{n=1}^{\infty} \mathbb{P}(N = n) = (1-p)^2. \tag{9}$$

Note that for the second equality we have used the renewal property of the Bernoulli process, namely, $(X_t)_{t\geq n+1}$ is a Bernoulli process with parameter p that is independent of $X_1,\cdots,X_n$, and the fact that the event $\{N=n\}$ is completely determined by the RVs $X_1,\cdots,X_n$.

Example 6.6 (Splitting and merging of Bernoulli processes)**.** Let $(X_t)_{t\geq 1}$ be a Bernoulli process with parameter p. Let us flip an independent probability $q\in[0,1]$ coin at every t, and define

$$Y_t = X_t\mathbf{1}(\text{coin lands heads}) \tag{10}$$

$$Z_t = X_t\mathbf{1}(\text{coin lands tails}). \tag{11}$$

Moreover, we have

$$X_t = Y_t + Z_t. \tag{12}$$

Note that $(Y_t)_{t\geq 1}$ and $(Z_t)_{t\geq 1}$ are also Bernoulli processes with parameters pq and $p(1-q)$, respectively. In other words, we splitted the Bernoulli process $(X_t)_{t\geq 1}$ with parameter p into two Bernoulli processes with parameters pq and $p(1-q)$. However, note that the processes Y_t and Z_t are not independent.

Conversely, let $(Y_t)_{t\geq 1}$ and $(Z_t)_{t\geq 1}$ be *independent* Bernoulli processes with parameters p and q, respectively. Is it possible to merge them into a single Bernoulli process? Indeed, we define

$$X_t = \mathbf{1}(Y_t = 1 \text{ or } Z_t = 1). \tag{13}$$

Then $\mathbb{P}(X_t=1) = 1-\mathbb{P}(Y_t=0)\mathbb{P}(Z_t=0) = 1-(1-p)(1-q) = p+q-pq$. By independence, X_t is a Bernoulli process with parameter $p+q-pq$.

Let $\tau_i \sim \text{Geom}(p)$ for $i\geq 0$ and let $N\sim\text{Geom}(q)$. Suppose all RVs are independent. Let $Y = \sum_{k=1}^{N}\tau_k$. In Exercise 4.24, we have shown that $Y\sim\text{Geom}(pq)$ using MGFs. In the following exercise, we show this by using splitting of Bernoulli processes.

Exercise 6.7 (Sum of geometric number of geometric RVs)**.** Let $(X_t)_{t\geq 0}$ be Bernoulli process of parameter p. Give each ball color Blue and Red independently with probability q and $1-q$, respectively. Let $X_t^B = \mathbf{1}(\text{there is a blue ball in box } t)$.

(i) Show that $(X_t^B)_{t\geq 1}$ is a Bernoulli process of parameter pq.
(ii) Let T_1^B be the location of first blue ball. Show that $T_1^B\sim\text{Geom}(pq)$.
(iii) Let N denote the total number of balls (blue or red) in the first T_1^B boxes. Show that $N\sim\text{Geom}(q)$.
(iv) Let T_k be the location of kth ball, and let $\tau_k = T_k - T_{k-1}$. Show that τ_k's are i.i.d. $\text{Geom}(p)$ RVs and they are independent of N. Lastly, show the identity

$$T_1^B = \sum_{k=1}^{N}\tau_k. \tag{14}$$

Example 6.8 (Applying limit theorems to BP)**.** Let $(X_t)_{t\geq 1}$ be a Bernoulli process with parameter p. Let T_k be the the smallest integer m such that $X_1+\cdots+X_m = k$, that is, the location of kth ball. Let $\tau_i = T_i - T_{i-1}$ for $i\geq 2$ and $\tau_0 = T_1$ be the inter-arrival times. Then

$$T_k = T_1 + (T_2-T_1)+(T_3-T_2)+\cdots+(T_k-T_{k-1}) \tag{15}$$

$$= \tau_1+\tau_2+\cdots+\tau_k. \tag{16}$$

Note that the τ_i's are i.i.d. $\text{Geom}(p)$ variables. Hence we can apply all limit theorems to T_k to bound/approximate probabilities associated to it.

To begin, recall that $\mathbb{E}(\tau_i) = 1/p$ and $\text{Var}(\tau_i) = (1-p)/p^2 < \infty$. Hence

$$\mathbb{E}(T_k) = k/p, \qquad \text{Var}(T_k) = \frac{(1-p)k}{p^2}. \tag{17}$$

If we apply SLLN to T_k, we conclude that

$$\mathbb{P}\left(\lim_{k\to\infty}\frac{T_k}{k}=\frac{1}{p}\right)=1. \tag{18}$$

So the line $y = x/p$ is the 'best fitting line' that explains the data points (k, T_k) (in the sense of linear regression). So we know that $1/p$ is a very good guess for T_k/k, which becomes more accurate as $k\to\infty$.

On the other hand, CLT describes how the sample mean T_k/k fluctuates around its mean $1/p$ as $k\to\infty$. The theorem says that as $k\to\infty$,

$$\frac{T_k - k/p}{\sqrt{k}\sqrt{(1-p)/p^2}} \Rightarrow Z \sim N(0,1). \tag{19}$$

What is this statement good for?

Lets take a concrete example by saying $p = 1/2$ and $k = 100$. Then $\mathbb{E}(T_{100}) = 200$ and $\mathrm{Var}(T_{100}) = 200$. Hence we expect the probability $\mathbb{P}(T_k \ge 250)$ to be very small. For this kind of tail probability estimation, we so far have three devices: Markov's and Chebyshef's inequality, and CLT itself.

First, Markov says

$$\mathbb{P}\left(T_{100}\ge 250\right)\le \frac{\mathbb{E}(T_{100})}{250}=\frac{200}{250}=\frac{4}{5}=0.8. \tag{20}$$

So this bound is not very useful here. Next, Chebyshef says

$$\mathbb{P}\left(|T_{100}-200|\ge 50\right)\le \frac{\mathrm{Var}(T_{100})}{50^2}=\frac{200}{2500}=0.08. \tag{21}$$

Moreover, an implication of CLT is that the distribution of T_k becomes more symmetric about its mean, so the probability on the left hand side is about twice of what we want.

$$\mathbb{P}(T_{100}\ge 250)\approx \frac{1}{2}\mathbb{P}\left(|T_{100}-200|\ge 50\right)\le 0.04. \tag{22}$$

So Chebyshef gives a much better bound.

But the truth is, the probability $\mathbb{P}\left(T_{100}\ge 250\right)$ in fact is extremely small. To see this, we apply CLT to get

$$\mathbb{P}\left(T_{100}\ge 250\right)=\mathbb{P}\left(\frac{T_{100}-200}{\sqrt{200}}\ge\frac{50}{10\sqrt{2}}\right)\approx \mathbb{P}(Z\ge 3.5355). \tag{23}$$

From the table for standard normal distribution, we know that $\mathbb{P}(Z \ge 1.96) = 0.025$ and $\mathbb{P}(Z \ge 2.58) = 0.005$. Hence The probability on the right hand side even smaller than these values.

6.2. **Poisson approximation of BP.** In this subsection, we will 'embed' the Bernoulli process into the real line and take a particular limit and get a preliminary form of the Poisson process. The basis is the following Exponential approximation of geometric distribution and Poisson approximation of Binomial distribution.

Exercise 6.9 (Exponential approximation of Geometric RV)**.** Let $\tau^{(n)} \sim \mathrm{Geom}(p)$. Let $p = \lambda/n$.

(i) For each real number $x \ge 0$, show that

$$\left(1-\frac{\lambda}{n}\right)^{n\lfloor x\rfloor}\le \mathbb{P}\left(\tau^{(n)}\ge nx\right)\le\left(1-\frac{\lambda}{n}\right)^{n\lceil x\rceil}, \tag{24}$$

where for each $x \ge 0$, $\lfloor x\rfloor$ (resp., $\lceil x\rceil$) denotes the largest (resp., smallest) integer that is at most (resp., least) x.

(ii) Show that $\tau^{(n)}/n \Rightarrow \mathrm{Exp}(\lambda)$.

Exercise 6.10 (Poisson approximation of Binomial RV)**.** In this exercise, we will show that Poisson distribution is obtained as a limit of the Binomial distribution as the number n of trials tend to infinity while the mean np is kept at constant λ. Recall that $X \sim \text{Poisson}(\lambda)$ if

$$\mathbb{P}(X=k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{25}$$

for all nonnegative integers $k \geq 0$. Let $Y_n \sim \text{Binomial}(n,p)$ with $np = \lambda$.

(i) Show that

$$\mathbb{P}(Y=k) = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k} \tag{26}$$

$$= \frac{n}{n}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\frac{(np)^k}{k!}(1-p)^{n-k} \tag{27}$$

$$= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k}. \tag{28}$$

(ii) Conclude that $Y_n \Rightarrow X$ as $n \to \infty$.

The punchline of this subsection is summarized in the following exercise.

Exercise 6.11. Let $(X_t)_{t\geq 0}$ be a Bernoulli process with parameter $p = \lambda/n$.

(i) Let $\tau_k^{(n)}$ denote one plus the number of boxes between $k-1$st and kth ball. Show that as $n \to \infty$,

$$\tilde{\tau}_k := \frac{\tau_k^{(n)}}{n} \Rightarrow \text{Exp}(\lambda). \tag{29}$$

(ii) Let S_k denote the number of balls in the first k boxes. Show that as $n \to \infty$,

$$\tilde{S}_n \Rightarrow \text{Poisson}(\lambda). \tag{30}$$

To give some more context, imagine starting from the first box at the origin and to the right, we examine each box for $1/n$ second. Let

$$\tilde{T}_k = \text{time that we discover } k\text{th ball} \tag{31}$$

$$\tilde{\tau}_k = \text{time to discover } k\text{th ball from } k-1\text{st ball}. \tag{32}$$

Since each box corresponds to $1/n$ second, we have

$$\tilde{T}_k = T_k/n \quad \text{and} \quad \tilde{\tau}_k = \tau_k/n. \tag{33}$$

Hence 6.11 (i) tells us that the time between consecutive balls converges in distribution to $\text{Exp}(\lambda)$ variable; (ii) tells us that the number of balls we discover in 1 second is asymptotically distributed as $\text{Poisson}(\lambda)$.

6.3. **The Poisson processes.** An *arrival* process is a sequence of strictly increasing RVs $0 < T_1 < T_2 < \cdots$. For each integer $k \geq 1$, its kth *inter-arrival time* is defined by $\tau_k = T_k - T_{k-1}\mathbf{1}(k \geq 2)$. For a given arrival process $(T_k)_{k\geq 1}$, the associated *counting process* $(N(t))_{t\geq 0}$ is defined by

$$N(t) = \sum_{k=1}^{\infty} \mathbf{1}(T_k \leq t) = \#(\text{arrivals up to time } t). \tag{34}$$

Note that these three processes (arrival times, inter-arrival times, and counting) determine each other:

$$(T_k)_{k\geq 1} \Longleftrightarrow (\tau_k)_{k\geq 1} \Longleftrightarrow (N(t))_{t\geq 0}. \tag{35}$$

Exercise 6.12. Let $(T_k)_{k\geq 1}$ be any arrival process and let $(N(t))_{t\geq 0}$ be its associated counting process. Show that these two processes determine each other by the following relation

$$\{T_n \leq t\} = \{N(t) \geq n\}. \tag{36}$$

In words, nth customer arrives by time t if and only if at least n customers arrive up to time t.

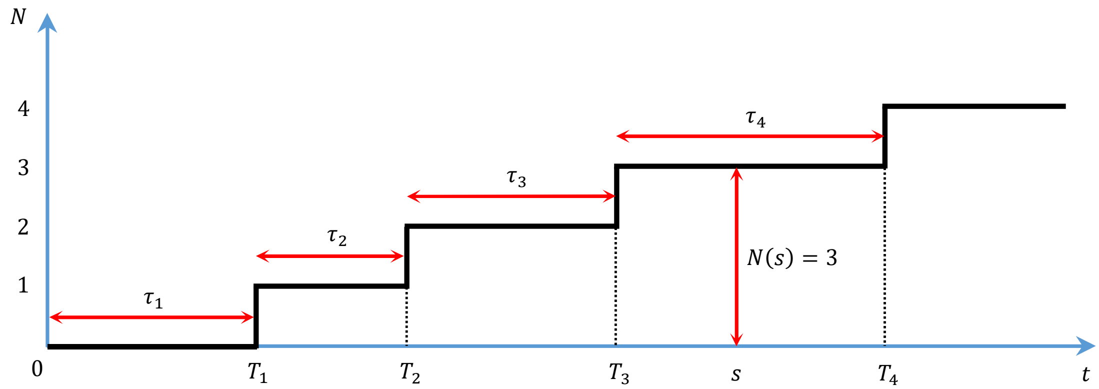

FIGURE 2. Illustration of a continuous-time arrival process $(T_k)_{k\geq 1}$ and its associated counting process $(N(t))_{t\geq 0}$. τ_k's denote inter-arrival times. $N(t) \equiv 3$ for $T_3 < t \leq T_4$.

Now we define Poisson process.

Definition 6.13 (Poisson process)**.** *An arrival process* $(T_k)_{k\geq 1}$ *is a Poisson process of rate* λ *if its inter-arrival times are i.i.d.* $\text{Exp}(\lambda)$ *RVs.*

Exercise 6.14. Let $(T_k)_{k\geq 1}$ be a Poisson process with rate λ. Show that $\mathbb{E}[T_k] = k/\lambda$ and $\text{Var}(T_k) = k/\lambda^2$. Furthermore, show that $T_k \sim \text{Erlang}(k, \lambda)$, that is,

$$f_{T_k}(z) = \frac{\lambda^k z^{k-1} e^{-\lambda z}}{(k-1)!}. \tag{37}$$

The following exercise explains what is 'Poisson' about the Poisson process.

Exercise 6.15. Let $(T_k)_{k\geq 1}$ be a Poisson process with rate λ and let $(N(t))_{t\geq 0}$ be the associated counting process. We will show that $N(t) \sim \text{Poisson}(\lambda t)$.

(i) Using the relation $\{T_n \leq t\} = \{N(t) \geq n\}$ and Exercise 6.14, show that

$$\mathbb{P}(N(t) \geq n) = \mathbb{P}(T_n \leq t) = \int_0^t \frac{\lambda^n z^{n-1} e^{-\lambda z}}{(n-1)!}\, dz. \tag{38}$$

(ii) Let $G(t) = \sum_{m=n}^{\infty} (\lambda t)^m e^{-\lambda t}/m! = \mathbb{P}(\text{Poisson}(\lambda) \geq n)$. Show that

$$\frac{d}{dt} G(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!} = \frac{d}{dt}\mathbb{P}(T_n \leq t). \tag{39}$$

Conclude that $G(t) = \mathbb{P}(T_n \leq t)$.

(iii) From (i) and (ii), conclude that $N(t) \sim \text{Poisson}(\lambda t)$.

The choice of exponential inter-arrival times is special due to the following 'memoryless property' of exponential RVs.

Exercise 6.16 (Memoryless property of exponential RV)**.** A continuous positive RV X is say to have *memoryless property* if

$$\mathbb{P}(X \geq t_1 + t_2) = \mathbb{P}(X \geq t_1)\mathbb{P}(X \geq t_2) \quad \forall x_1, x_2 \geq 0. \tag{40}$$

(i) Show that (40) is equivalent to

$$\mathbb{P}(X \geq t_1 + t_2 \,|\, X \geq t_2) = \mathbb{P}(X \geq t_1) \quad \forall x_1, x_2 \geq 0. \tag{41}$$

(ii) Show that exponential RVs have memoryless property.

(iii) Suppose X is continuous, positive, and memoryless. Let $g(t) = \log \mathbb{P}(X \geq t)$. Show that g is continuous at 0 and

$$g(x+y) = g(x) + g(y) \quad \text{for all } x, y \geq 0. \tag{42}$$

Using the following exercise, conclude that X must be an exponential RV.

Exercise 6.17. Let $g : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be a function with the property that $g(x+y) = g(x) + g(y)$ for all $x, y \geq 0$. Further assume that g is continuous at 0. In this exercise, we will show that $g(x) = cx$ for some constant c.

(i) Show that $g(0) = g(0+0) = g(0) + g(0)$. Deduce that $g(0) = 0$.

(ii) Show that for all integers $n \geq 1$, $g(n) = ng(1)$.

(iii) Show that for all integers $n, m \geq 1$,

$$ng(1) = g(n \cdot 1) = g(m(n/m)) = mg(n/m). \tag{43}$$

Deduce that for all nonnegative rational numbers r, we have $g(r) = rg(1)$.

(iv) Show that g is continuous.

(v) Let x be nonnegative real number. Let r_k be a sequence of rational numbers such that $r_k \to x$ as $k \to \infty$. By using (iii) and (iv), show that

$$g(x) = g\left(\lim_{k\to\infty} r_k\right) = \lim_{k\to\infty} g(r_k) = g(1) \lim_{k\to\infty} r_k = x \cdot g(1). \tag{44}$$

Given a Poisson process, we can restart it at any given time t. Then the first arrival time after t is simply the remaining inter-arrival time after time t. By memoryless property of exponential RVs, we see that this remaining time is also an exponential RV that is independent of what have happend so far. We will show this in the following proposition. The proof is essentially a Poisson version of Exercise 6.4.

Proposition 6.18 (Memoryless property of PP)**.** *Let $(T_k)_{k\geq 1}$ be a Poisson process of rate λ and let $(N(t))_{t\geq 0}$ be the associated counting process.*

(i) *For any $t \geq 0$, let $Z(t) = \inf\{s > t : N(s) > N(t)\}$ be the first arrival time after time t. Then $Z(t) \sim \text{Exp}(\lambda)$ and it is independent of the process up to time t.*

(ii) *For any $s \geq 0$, $(N(t+s) - N(t))_{t\geq 0}$ is the counting process of an independent Poisson process of rate λ, restarted at time s.*

Proof. (ii) follows immediately from (i). We show (i). Denote $Z = Z(t)$. We will show that Z is independent of $N(t)$ and it has distribution $\text{Exp}(\lambda)$. Since Z depends only on the current inter-arrival time (see Figure 5), this will show that Z is independent of the process up to time t.

For starter, first consider conditioning on the event that $N(t) = 0$, that is, no arrival occurred by time t. Then by Exercises 6.12 and 6.16,

$$\begin{aligned} \mathbb{P}(Z \geq x \,|\, N(t) = 0) &= \mathbb{P}(Z \geq x \,|\, T_1 > t) & (45)\\ &= \mathbb{P}(T_1 \geq x + t \,|\, T_1 > t) & (46)\\ &= \mathbb{P}(T_1 \geq x) = e^{-\lambda x}. & (47) \end{aligned}$$

Similarly, now suppose $N(t) = n$, that is, there have been n arrivals up to time t. Furthermore, we also assume that the last arrival is at time s, that is, $T_n = s \leq t$. Then

$$\begin{aligned} \mathbb{P}(Z \geq x \,|\, N(t) = n,\, T_n = s) &= \mathbb{P}(\tau_{n+1} \geq x + t - s \,|\, N(t) = n,\, T_n = s) & (48)\\ &= \mathbb{P}(\tau_{n+1} \geq x + t - s \,|\, \tau_{n+1} \geq t - s,\, T_n = s) & (49) \end{aligned}$$

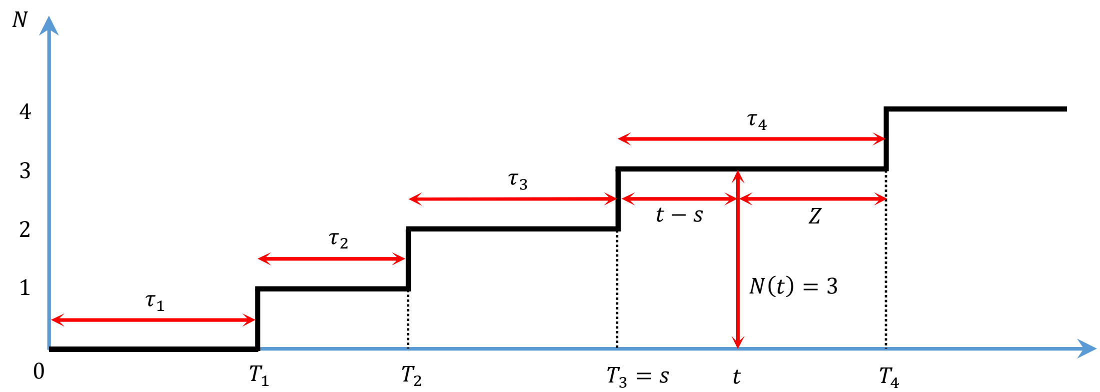

FIGURE 3. Assuming $N(t)=3$ and $T_3=s\le t$, we have $Z=\tau_4-(t-s)$. By memoryless property of exponential RV, Z follows $\mathrm{Exp}(\lambda)$ on this conditioning.

$$= \mathbb{P}(\tau_{n+1}\ge x+t-s\,|\,\tau_{n+1}\ge t-s) \tag{50}$$

$$= \mathbb{P}(\tau_{n+1}\ge x)=e^{-\lambda x}. \tag{51}$$

Hence by iterated expectation,

$$\mathbb{P}(Z\ge x\,|\,N(t)=n)=\mathbb{E}[\mathbb{P}(Z\ge x\,|\,N(t)=n,\,T_n)]=e^{-\lambda x}. \tag{52}$$

Since n is arbitrary, this shows that Z is independent of $N(t)$. By using another iterated expectation, this also shows that $Z\sim \mathrm{Exp}(\lambda)$. □

Exercise 6.19 (Sum of independent Poisson RV's is Poisson)**.** Let $(T_k)_{k\ge1}$ be a Poisson process with rate λ and let $(N(t))_{t\ge0}$ be the associated counting process. Fix $t,s\ge0$.

(i) Use memoryless property to show that $N(t)$ and $N(t+s)-N(t)$ are independent Poisson RVs of rates λt and λs.

(ii) Note that the total number of arrivals during $[0,t+s]$ can be divided into the number of arrivals during $[0,t]$ and $[t,t+s]$. Conclude that if $X\sim\mathrm{Poisson}(\lambda t)$ and $Y\sim\mathrm{Poisson}(\lambda s)$ and if they are independent, then $X+Y\in\mathrm{Poisson}(\lambda(t+s))$.

6.4. **Splitting and merging of Poisson process.** Recall the splitting of Bernoulli processes: If balls are given by $\mathrm{BP}(p)$ and we color each ball with blue and red independently with probability q and $1-q$, respectively, then the process restricted on blue and red balls are $\mathrm{BP}(pq)$ and $\mathrm{BP}(p(1-q))$, respectively. Considering blue balls process is sometimes called 'thinning' of the original BP. The same construction naturally works for Poisson processes as well. If customers arrive at a bank according to $\mathrm{PP}(\lambda)$ and if each one is male or female independently with probability q and $1-q$, then the 'thinned out' process of only male customers is a $\mathrm{PP}(q\lambda)$; the process of female customers is a $\mathrm{PP}((1-q)\lambda)$.

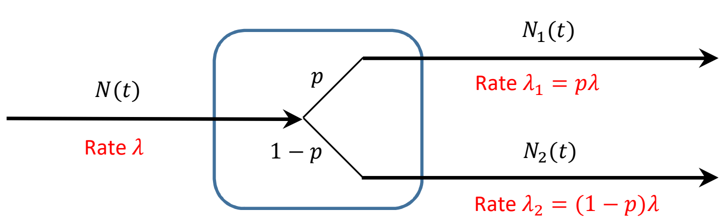

FIGURE 4. Splitting of Poisson process $N(t)$ of rate λ according to an independent Bernoulli process of parameter p.

The reverse operation of splitting a given PP into two complementary PPs is call the 'merging'. Namely, imagine customers arrive at a register through two doors A and B independently according to PPs of rates λ_A and λ_B, respectively. Then the combined arrival process of entire customers is again a PP of the added rate.

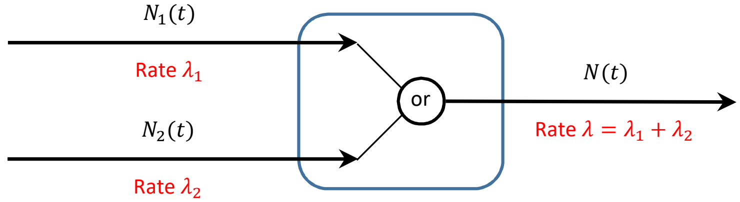

FIGURE 5. Merging two independent Poisson processes of rates λ_1 and λ_2 gives a new Poisson process of rate $\lambda_1 + \lambda_2$.

Exercise 6.20 (Excerpted from [BT02])**.** Transmitters A and B independently send messages to a single receiver according to Poisson processes with rates $\lambda_A = 3$ and $\lambda_B = 4$ (messages per min). Each message (regardless of the source) contains a random number of words with PMF

$$\mathbb{P}(1 \text{ word}) = 2/6, \quad \mathbb{P}(2 \text{ words}) = 3/6, \quad \mathbb{P}(3 \text{ words}) = 1/6, \tag{53}$$

which is independent of everything else.

(i) Find $\mathbb{P}$(total nine messages are recieved during $[0, t]$).
(ii) Let $M(t)$ be the total number of words received during $[0, t]$. Find $\mathbb{E}[M(t)]$.
(iii) Let T be the first time that the receiver receives exactly three messages consisting of three words from transmitter A. Find distribution of T.
(iv) Compute $\mathbb{P}$(exactly seven messages out of the first ten messages are from A).

Exercise 6.21 (Order statistics of i.i.d. Exp RVs)**.** One hundred light bulbs are simultaneously put on a life test. Suppose the lifetimes of the individual light bulbs are independent Exp(λ) RVs. Let T_k be the kth time that some light bulb fails. We will find the distribution of T_k using Poisson processes.

(i) Think of T_1 as the first arrival time among 100 independent PPs of rate λ. Show that $T_1 \sim \text{Exp}(100\lambda)$.
(ii) After time T_1, there are 99 remaining light bulbs. Using memoryless property, argue that $T_2 - T_1$ is the first arrival time of 99 independent PPs of rate λ. Show that $T_2 - T_1 \sim \text{Exp}(99\lambda)$ and that $T_2 - T_1$ is independent of T_1.
(iii) As in the coupon collector problem, we break up

$$T_k = \tau_1 + \tau_2 + \cdots + \tau_k, \tag{54}$$

where $\tau_i = T_i - T_{i-1}$ with $\tau_1 = T_1$. Note that τ_i is the waiting time between $i-1$st and ith failures. Using the ideas in (i) and (ii), show that τ_i's are independent and $\tau_i \sim \text{Exp}((100 - i)\lambda)$. Deduce that

$$\mathbb{E}[T_k] = \frac{1}{\lambda}\left(\frac{1}{100} + \frac{1}{99} + \cdots + \frac{1}{(100-k+1)}\right), \tag{55}$$

$$\text{Var}[T_k] = \frac{1}{\lambda^2}\left(\frac{1}{100^2} + \frac{1}{99^2} + \cdots + \frac{1}{(100-k+1)^2}\right). \tag{56}$$

(iv) Let $X_1, X_2, \cdots, X_{100}$ be i.i.d. Exp(λ) variables. Let $X_{(1)} < X_{(2)} < \cdots < X_{(100)}$ be their order statistics, that is, $X_{(k)}$ is the ith smallest among the X_i's. Show that $X_{(k)}$ has the same distribution as T_k, the kth time some light bulb fails. (So we know what it is from the previous parts.)

In the next two exercises, we rigorously justify splitting and merging of Poisson processes.

Exercise 6.22 (Splitting of PP)**.** Let $(N(t))_{t\geq 0}$ be the counting process of a PP(λ), and let $(X_k)_{k\geq 0}$ be an independent BP(p). We define two counting processes $(N_1(t))_{t\geq 0}$ and $(N_2(t))_{t\geq 0}$ by

$$N_1(t) = \sum_{k=1}^{\infty} \mathbf{1}(T_k \leq t)\mathbf{1}(X_k = 1) = \#(\text{arrivals with coin landing on heads up to time } t), \tag{57}$$

$$N_2(t) = \sum_{k=1}^{\infty} \mathbf{1}(T_k \leq t)\mathbf{1}(X_k = 0) = \#(\text{arrivals with coin landing on heads up to time } t). \tag{58}$$

In this exercise, we show that $(N_1(t))_{t\geq 0} \sim \text{PP}(p\lambda)$ and $(N_2(t))_{t\geq 0} \sim \text{PP}((1-p)\lambda)$.

(i) Let τ_k and $\tau_k^{(1)}$ be the kth inter-arrival times of the counting processes $(N(t))_{t\geq 0}$ and $(N_1(t))_{t\geq 0}$. Let Y_k be the location of kth ball for the BP $(X_t)_{t\geq 0}$. Show that

$$\tau_1^{(1)} = \sum_{i=1}^{Y_1} \tau_i. \tag{59}$$

(ii) Show that

$$\tau_2^{(1)} = \sum_{k=Y_1+1}^{Y_2} \tau_i. \tag{60}$$

(iii) Show that in general,

$$\tau_k^{(1)} = \sum_{k=Y_{k-1}+1}^{Y_k} \tau_i. \tag{61}$$

(iv) Recall that $Y_k - Y_{k-1}$'s are i.i.d. Geom(p) RVs. Use Exercise 4.23 and (iii) to deduce that $\tau_k^{(1)}$'s are i.i.d. Exp($p\lambda$) RVs. Conclude that $(N_1(t)) \sim \text{PP}(p\lambda)$. (The same argument shows $(N_2(t))_{t\geq 0} \sim \text{PP}((1-p)\lambda)$.)

Exercise 6.23 (Merging of independent PPs)**.** Let $(N_1(t))_{t\geq 0}$ and $(N_2(t))_{t\geq 0}$ be the counting processes of two independent PPs of rates λ_1 and λ_2, respectively. Define a new counting process $(N(t))_{t\geq 0}$ by

$$N(t) = N_1(t) + N_2(t). \tag{62}$$

In this exercise, we show that $(N(t))_{t\geq 0} \sim \text{PP}(p\lambda)$.

(i) Let $\tau_k^{(1)}, \tau_k^{(2)}$, and τ_k be the kth inter-arrival times of the counting processes $(N_1(t))_{t\geq 0}$, $(N_2(t))_{t\geq 0}$, and $(N(t))_{t\geq 0}$. Show that $\tau_1 = \min(\tau_1^{(1)}, \tau_1^{(2)})$. Conclude that $\tau_1 \sim \text{Exp}(\lambda_1 + \lambda_2)$.

(ii) Let T_k be the kth arrival time for the joint process $(N(t))_{t\geq 0}$. Use memoryless property of PP to deduce that N_1 and N_2 restarted from time T_k are independent PPs of rates λ_1 and λ_2, which are also independent from the past (before time t).

(iii) From (ii), show that

$$\tau_{k+1} = \min(\tilde{\tau}_1, \tilde{\tau}_2), \tag{63}$$

where $\tilde{\tau}_1$ is the waiting time for the first arrival after time T_k for N_1, and similarly for $\tilde{\tau}_2$. Deduce that $\tau_{k+1} \sim \text{Exp}(\lambda_1 + \lambda_2)$ and it is independent of $\tau_1, \cdots, \tau_k$. Conclude that $(N(t))_{t\geq 0} \sim \text{PP}(\lambda_1 + \lambda_2)$.

6.5. **Poisson process in terms of counting process*.** This subsection is optional. We have defined Poisson process in terms of the inter-arrival times of an arrival process. In this subsection, we look at alternative definitions based on its counting process.

Definition 6.24 (Def of PP:counting1)**.** *A counting process $(N(t))_{t\ge 0}$ is said to be a Poisson process with rate $\lambda > 0$ if it has the following properties*

(i) $N(0) = 0$;
(ii) *(Independent increment) For any $t, s \ge 0$, $N(t+s) - N(t)$ is independent of $(N(u))_{u\le t}$;*
(iii) *For any $t, s \ge 0$, $N(t+s) - N(t) \sim \text{Poisson}(\lambda s)$.*

Proposition 6.25. *The two definitions of Poisson process in Definitions 6.13 and 6.24 are equivalent.*

Proof. Let $(N(t))_{t\ge 0}$ be a counting process with the properties (i)-(iii) in Def 6.24. We want to show that the inter-arrival times are i.i.d. $\text{Exp}(\lambda)$ RVs. This is the content of Exercise 6.26.

Conversely, let $(T_k)_{k\ge 1}$ be an arrival process. Suppose its inter-arrival times are i.i.d. $\text{Exp}(\lambda)$ RVs. Let $(N(t))_{\ge 0}$ be its associated counting process. Clearly $N(0) = 0$ by definition so (i) holds. By the memoryless property (Proposition 6.18), $(N_u)_{u\ge t}$ is the counting process of a Poisson process of rate λ (in the sense of Def 6.13) that is independent of the past $(N(u))_{u\le t}$. In particular, the increment $N(t+s) - N(t)$ during time interval $[t, t+s]$ is independent of the past process $(N(u))_{u\le t}$, so (ii) holds. Lastly, the increment $N(t+s) - N(t)$ has the same distribution as $N(s) = N(s) - N(0)$ by the memoryless property. Since Exercise 6.15 shows that $N(t) \sim \text{Poisson}(\lambda t)$, we have (iii). □

Exercise 6.26. Let $(N(t))_{t\ge 0}$ be a counting process with the properties (i)-(iii) in Def 6.24. Let $T_k = \inf\{u \ge 0 \,|\, N(u) = k\}$ be the kth arrival time and let $\tau_k = T_k - T_{k-1}$ be the kth inter-arrival time.

(i) Use conditioning on T_k to show that for any $k \ge 1$ and $s \ge 0$, $N(T_k + s) - N(T_k)$ is independent of $(N(t))_{u\le T_k}$.
(ii) Let $Z(t) = \inf\{u \ge t \,|\, N(u) > N(t)\}$ be the first arrival time after time t. Show that $Z(t) \sim \text{Exp}(\lambda)$ for all $t \ge 0$.
(iii) Use (ii) and conditioning on T_{k-1} to show that $\tau_k \sim \text{Exp}(\lambda)$ for all $k \ge 1$.

Next, we give yet another definition of Poisson process in terms of the asymptotic properties of its counting process. For this, we need something called the 'small-o' notation. We say a function $f(t)$ is of order $o(t)$ or write $f(t) = o(t)$ if

$$\lim_{t\to 0} \frac{f(t)}{t} = 0. \tag{64}$$

Definition 6.27 (Def of PP:counting2)**.** *A counting process $(N(t))_{t\ge 0}$ is said to be a Poisson process with rate $\lambda > 0$ if it satisfies the following conditions:*

(i) $N(0) = 0$;
(ii) $\mathbb{P}(N(t) = 0) = 1 - \lambda t + o(t)$;
(iii) $\mathbb{P}(N(t) = 1) = \lambda t + o(t)$;
(iv) $\mathbb{P}(N(t) \ge 2) = o(t)$;
(v) *(Independent increment) For any $t, s \ge 0$, $N(t+s) - N(t)$ is independent of $(N(u))_{u\le t}$;*
(vi) *(Stationary increment) For any $t, s \ge 0$, the distribution of $N(t+s) - N(t)$ does not depend on t.*

It is easy to see that our usual definition of Poisson process in Definition 6.13 satisfies the properties (i)-(vi) above.

Proposition 6.24. *Let $(T_k)_{k\ge 1}$ be a Poisson process of rate λ (in the sense of Definition 6.13) and let $(N(t))_{t\ge 0}$ be its associated counting process. Then $(N(t))_{t\ge 0}$ is a Poisson process in the sense of Definition 6.24.*

Proof. There is no arrival at time $t=0$ so $N(0)=0$. We know that $N(t)\sim \text{Poisson}(\lambda t)$ for each $t>0$ from Exercise 6.15. Also note that

$$e^{-\lambda t} = 1-\lambda t + o(t) \tag{65}$$

for all $t>0$. Hence

$$\mathbb{P}(N(t)=n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!} \tag{66}$$

$$= (1-\lambda t+o(t))\frac{(\lambda t)^n}{n!}. \tag{67}$$

So plugging in $n=0$ and 1 gives (ii) and (iii). For (iv), we use (ii) and (iii) to get

$$\mathbb{P}(N(t)\ge 2) = 1-\mathbb{P}(N(t)\le 1) = 1-(1-\lambda t+o(t))-(\lambda t+o(t)) = o(t). \tag{68}$$

Lastly, (v) and (iv) follows from the memoryless property of Poisson process (Proposition 6.18). □

Next, we consider the converse implication. We will break this into several exercises.

Exercise 6.28. Let $(N(t))_{t\ge 0}$ is the Poisson process with rate $\lambda>0$ in the sense of Definition 6.24. In this exercise, we will show that $\mathbb{P}(N(t)=0)=e^{-\lambda t}$.

(i) Use independent/stationary increment properties to show that

$$\mathbb{P}(N(t+h)=0) = \mathbb{P}(N(t)=0, N(t+h)-N(t)=0) \tag{69}$$

$$= \mathbb{P}(N(t)=0)\mathbb{P}(N(t+h)-N(t)=0) \tag{70}$$

$$= \mathbb{P}(N(t)=0)(1-\lambda h+o(h)). \tag{71}$$

(ii) Denote $f_0(t)=\mathbb{P}(N(t)=0)$. Use (i) to show that

$$\frac{f_0(t+h)-f_0(h)}{h} = \left(-\lambda+\frac{o(h)}{h}\right)f_0(t). \tag{72}$$

By taking limit as $h\to 0$, show that $f(t)$ satisfies the following differential equation

$$\frac{df_0(t)}{dt} = -\lambda f_0(t). \tag{73}$$

(iii) Conclude that $\mathbb{P}(N(t)=0)=e^{-\lambda t}$.

Next, we generalize the ideas used in the previous exercise to compute the distribution of $N(t)$.

Exercise 6.29. Let $(N(t))_{t\ge 0}$ is the Poisson process with rate $\lambda>0$ in the sense of Definition 6.24. Denote $f_n(t)=\mathbb{P}(N(t)=n)$ for each $n\ge 0$.

(i) Show that

$$\mathbb{P}(N(t)\le n-2,\, N(t+h)=n) \le \mathbb{P}(N(t+h)-N(h)\ge 2). \tag{74}$$

Conclude that

$$\mathbb{P}(N(t)\le n-2,\, N(t+h)=n) = o(h). \tag{75}$$

(ii) Use (i) and independent/stationary increment properties to show that

$$f_n(t+h) = \mathbb{P}(N(t+h)=n) = \mathbb{P}(N(t)=n,\, N(t+h)-N(t)=0) \tag{76}$$

$$+\mathbb{P}(N(t)=n-1,\, N(t+h)-N(t)=1) \tag{77}$$

$$+\mathbb{P}(N(t)\le n-2,\, N(t+h)=n) \tag{78}$$

$$= f_n(t)(1-\lambda h+o(h)) + f_{n-1}(t)(\lambda h+o(h)) + o(h). \tag{79}$$

(iii) Use (ii) to show that the following differential equation holds:

$$\frac{df_n(t)}{dt} = -\lambda f_n(t)+\lambda f_{n-1}(t). \tag{80}$$

(iv) By multiplying the integrating factor $\mu(t) = e^{\lambda t}$ to (80), show that

$$(e^{\lambda t} f_n(t))' = \lambda e^{\lambda t} f_{n-1}(t). \tag{81}$$

Use the initial condition $f_n(0) = \mathbb{P}(N(0) = n) = 0$ to derive the recursive equation

$$f_n(t) = \lambda e^{-\lambda t} \int_0^t e\lambda s f_{n-1}(s)\, ds. \tag{82}$$

(v) Use induction to conclude that $f_n(t) = (\lambda t)^n e^{-\lambda t}/n!$.

(vi) Conclude that for all $t, s \geq 0$ and $n \geq 0$,

$$N(t+s) - N(s) \sim \text{Poisson}(\lambda t). \tag{83}$$

6.6. **Discrete-time Markov chains.** In this sebsection, we change our gear from arrival processes to *Markov processes*. Roughly speaking, Markov processes are used to model temporally changing systems where future state only depends on the current state. For instance, if the price of bitcoin tomorrow depends only on its price today, then bitcoin price can be modeled as a Markov process. (Of course, the entire history of price often affects decisions of buyers/sellers so it may not be a realistic assumption.)

Even through Markov processes can be defined in vast generality, we concentrate on the simplest setting where the state and time are both discrete. Let $\Omega = \{1, 2, \cdots, m\}$ be a finite set, which we call the *state space*. Consider a sequence $(X_t)_{t\geq 0}$ of Ω-valued RVs, which we call a *chain*. We call the value of X_t the *state* of the chain at time t. In order to narrow down the way the chain $(X_t)_{t\geq 0}$ behaves, we introduce the following properties:

(i) (Markov property) The distribution of X_{t+1} given the history $X_0, X_1, \cdots, X_t$ depends only on X_t. That is,

$$\mathbb{P}(X_{t+1} = k \,|\, X_t = j_t, X_{t-1} = j_{t-1}, \cdots, X_1 = j_1) = \mathbb{P}(X_{t+1} = k \,|\, X_t = j_t). \tag{84}$$

(ii) (Time-homogeneity) The transition probabilities

$$p_{ij} = \mathbb{P}(X_{t+1} = j \,|\, X_t = i) \qquad i, j \in \Omega \tag{85}$$

do not depend on t.

When the chain $(X_t)_{t\geq 0}$ satisfies the above two properties, we say it is a (discrete-time and time-homogeneous) *Markov chain*. Note that the Markov property (i) is a kind of a one-step complication of the memoryless property: We now forget all the past but we do remember the present. On the other hand, time-homogeneity (ii) states that the behavior of the chain does not depend on time. In this case, we define the *transition matrix* P to be the $m \times m$ matrix of transition probabilities:

$$P = (p_{ij})_{1\leq i,j\leq m} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}. \tag{86}$$

Finally, since the state X_t of the chain is a RV, we represent its PMF via a row vector

$$\mathbf{r}_t = [\mathbb{P}(X_t = 1), \mathbb{P}(X_t = 2), \cdots, \mathbb{P}(X_t = m)]. \tag{87}$$

Example 6.31. Let $\Omega = \{1, 2\}$ and let $(X_t)_{t\geq 0}$ be a Markov chain on Ω with the following transition matrix

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}. \tag{88}$$

We can also represent this Markov chain pictorially as in Figure 8, which is called the 'state space diagram' of the chain $(X_t)_{t\geq 0}$.

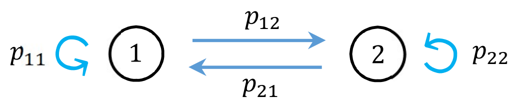

FIGURE 6. State space diagram of a 2-state Markov chain

For some concrete example, suppose

$$p_{11} = 0.2, \quad p_{12} = 0.8, \quad p_{21} = 0.6, \quad p_{22} = 0.4. \tag{89}$$

If the initial state of the chain X_0 is 1, then

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 1 \,|\, X_0 = 1)\mathbb{P}(X_0 = 1) + \mathbb{P}(X_1 = 1 \,|\, X_0 = 2)\mathbb{P}(X_0 = 2) \tag{90}$$

$$= \mathbb{P}(X_1 = 1 \,|\, X_0 = 1) = p_{11} = 0.2 \tag{91}$$

and similarly,

$$\mathbb{P}(X_1 = 2) = \mathbb{P}(X_1 = 2 \,|\, X_0 = 1)\mathbb{P}(X_0 = 1) + \mathbb{P}(X_1 = 2 \,|\, X_0 = 2)\mathbb{P}(X_0 = 2) \tag{92}$$

$$= \mathbb{P}(X_1 = 2 \,|\, X_0 = 1) = p_{12} = 0.8. \tag{93}$$

Also we can compute the distribution of X_2. For example,

$$\mathbb{P}(X_2 = 1) = \mathbb{P}(X_2 = 1 \,|\, X_1 = 1)\mathbb{P}(X_1 = 1) + \mathbb{P}(X_2 = 1 \,|\, X_1 = 2)\mathbb{P}(X_1 = 2) \tag{94}$$

$$= p_{11}\mathbb{P}(X_1 = 1) + p_{21}\mathbb{P}(X_1 = 2) \tag{95}$$

$$= 0.2 \cdot 0.2 + 0.6 \cdot 0.8 = 0.04 + 0.48 = 0.52. \tag{96}$$

In general, the distribution of X_{t+1} can be computed from that of X_t via a simple linear algebra. Note that for $i = 1, 2$,

$$\mathbb{P}(X_{t+1} = i) = \mathbb{P}(X_{t+1} = i \,|\, X_t = 1)\mathbb{P}(X_t = 1) + \mathbb{P}(X_{t+1} = i \,|\, X_t = 2)\mathbb{P}(X_t = 2) \tag{97}$$

$$= p_{1i}\mathbb{P}(X_t = 1) + p_{2i}\mathbb{P}(X_t = 2). \tag{98}$$

This can be written as

$$[\mathbb{P}(X_{t+1} = 2), \mathbb{P}(X_{t+1} = 2)] = [\mathbb{P}(X_{t+1} = 2), \mathbb{P}(X_{t+1} = 2)] \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}. \tag{99}$$

That is, if we represent the distribution of X_t as a row vector, then the distribution of X_{t+1} is given by multiplying the transition matrix P to the left.

We generalize this observation in the following exercise.

Exercise 6.32. Let $(X_t)_{t\geq 0}$ be a Markov chain on state space $\Omega = \{1, 2, \cdots, m\}$ with transition matrix $P = (p_{ij})_{1\leq i,j\leq m}$. Let $\mathbf{r}_t = [\mathbb{P}(X_t = 1), \cdots, \mathbb{P}(X_t = m)]$ denote the row vector of the distribution of X_t.

(i) Show that for each $i \in \Omega$,

$$\mathbb{P}(X_{t+1} = i) = \sum_{j=1}^{m} p_{ji}\mathbb{P}(X_t = j). \tag{100}$$

(ii) Show that for each $t \geq 0$,

$$\mathbf{r}_{t+1} = \mathbf{r}_t P. \tag{101}$$

(iii) Show by induction that for each $t \geq 0$,

$$\mathbf{r}_t = \mathbf{r}_0 P^t. \tag{102}$$

Exercise 6.33. Let $\Omega = \{1, 2\}$ and let $(X_t)_{t\geq 0}$ be a Markov chain on Ω with the following transition matrix

$$P = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix}. \tag{103}$$

(i) Show that P admits the following diagonalization

$$P = \begin{bmatrix} 1 & -4/3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -2/5 \end{bmatrix} \begin{bmatrix} 1 & -4/3 \\ 1 & 1 \end{bmatrix}^{-1}. \tag{104}$$

(ii) Show that P^t admits the following diagonalization

$$P^t = \begin{bmatrix} 1 & -4/3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (-2/5)^t \end{bmatrix} \begin{bmatrix} 1 & -4/3 \\ 1 & 1 \end{bmatrix}^{-1}. \tag{105}$$

(iii) Let $\mathbf{r}_t$ denote the row vector of distribution of X_t. Use Exercise 6.32 to deduce that

$$\mathbf{r}_t = \mathbf{r}_0 \begin{bmatrix} 1 & -4/3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (-2/5)^t \end{bmatrix} \begin{bmatrix} 1 & -4/3 \\ 1 & 1 \end{bmatrix}^{-1}. \tag{106}$$

Also show that

$$\lim_{t\to\infty} \mathbf{r}_t = \mathbf{r}_0 \begin{bmatrix} 3/7 & 4/7 \\ 3/7 & 4/7 \end{bmatrix} = [3/7, 4/7]. \tag{107}$$

Conclude that regardless of the initial distribution $\mathbf{r}_0$, the distribution of the Markov chain $(X_t)_{t\geq 0}$ converges to $[3/7, 4/7]$. This limiting distribution $\pi = [3/7, 4/7]$ is called the *stationary distribution* of the chain $(X_t)_{t\geq 0}$.

6.7. **Stationary distribution and examples.** Let $(X_t)_{t\geq 0}$ be a Markov chain on state space $\Omega = \{1, 2, \cdots, m\}$ with transition matrix $P = (p_{ij})_{1\leq i,j\leq m}$. If π is a distribution on Ω such that

$$\pi = \pi P, \tag{108}$$

then we say π is a *stationary distribution* of the Markov chain $(X_t)_{t\geq 0}$.

Example 6.34. In Exercise 6.33, we have seen that the distribution of the 2-state Markov chain $(X_t)_{t\geq 0}$ with transition matrix

$$P = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix}. \tag{109}$$

converges to $\pi = [3/7, 4/7]$. Since this is the limiting distribution, it should be invariant under left multiplication by P. Indeed, one can easily verify

$$[3/7, 4/7] = [3/7, 4/7] \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix}. \tag{110}$$

Hence π is a stationary distribution for the Markov chain $(X_t)_{t\geq 0}$. Furthermore, in Exercise 6.33 we also have shown the uniqueness of stationary distribution. However, this is not always the case.

Example 6.35. Let $(X_t)_{t\geq 0}$ be a 2-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{111}$$

Then any distribution $\pi = [p, 1-p]$ is a stationary distribution for the chain $(X_t)_{t\geq 0}$.

In Exercise 6.33, we used diagonalization of the transition matrix to compute the limiting distribution, which must be a stationary distribution. However, we can simply use the definition (108) to algebraically compute stationary distribution(s). Namely, by taking transpose,

$$\pi^T = P^T \pi^T. \tag{112}$$

Namely, the transpose of any stationary distribution is an eigenvector of P^T associated with eigenvalue 1. We record some properties of stationary distributions using some linear algebra stuff.

Exercise 6.36. Let $(X_t)_{t\geq 0}$ be a Markov chain on state space $\Omega = \{1, 2, \cdots, m\}$ with transition matrix $P = (p_{ij})_{1\leq i,j\leq m}$.

(i) Show that a distribution π on Ω is a stationary distribution for the chain $(X_t)_{t\geq 0}$ if and only if it is a left eigenvector of P associated with left eigenvalue 1.

(ii) Show that 1 is a right eigenvalue of P with right eigenvector $[1, 1, \cdots, 1]^T$.

(iii) Recall that a square matrix and its transpose have the same (right) eigenvalues and corresponding (right) eigenspaces have the same dimension. Show that the Markov chain $(X_t)_{t\geq 0}$ has a unique stationary distribution if any only if $[1, 1, \cdots, 1]^T$ spans the (right) eigenspace of P associated with (right) eigenvalue 1.

Now we look at some important examples.

Exercise 6.37 (Birth-Death chain)**.** Let $\Omega = \{0, 1, 2, \cdots, N\}$ be the state space. Let $(X_t)_{t\geq 0}$ be a Markov chain on Ω with transition probabilities

$$\begin{cases} \mathbb{P}(X_{t+1} = k+1 \,|\, X_t = k) = p & \forall 0 \leq k < N \\ \mathbb{P}(X_{t+1} = k-1 \,|\, X_t = k) = 1-p & \forall 1 \leq k \leq N \\ \mathbb{P}(X_{t+1} = 0 \,|\, X_t = 0) = 1-p \\ \mathbb{P}(X_{t+1} = N \,|\, X_t = N) = p. \end{cases} \tag{113}$$

This is called a Birth-Death chain. Its state space diagram is as below.

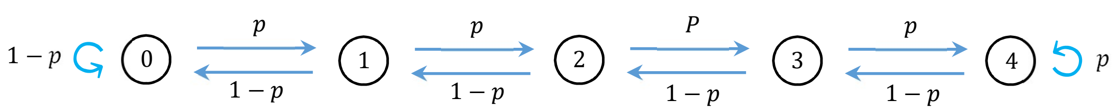

FIGURE 7. State space diagram of a 5-state Birth-Death chain

(i) Let $\pi = [\pi_0, \pi_1, \cdots, \pi_N]$ be a distribution on Ω. Show that π is a stationary distribution of the Birth-Death chain if and only if it satisfy the following 'balance equation'

$$p\pi_k = (1-p)\pi_{k+1} \qquad 0 \leq k < N. \tag{114}$$

(ii) Let $\rho = p/(1-p)$. From (ii), deduce that $\pi_k = \rho^k \pi_0$ for all $0 \leq k < N$.

(iii) Using the normalization condition $\pi_0 + \pi_1 + \cdots + \pi_N$, show that $\pi_0 = 1/(1 + \rho + \rho^2 + \cdots + \rho^N)$. Conclude that

$$\pi_k = \frac{\rho^k}{1 + \rho + \rho^2 + \cdots + \rho^N} = \rho^k \frac{1-\rho}{1-\rho^{N+1}} \qquad 0 \leq k \leq N. \tag{115}$$

Conclude that the Birth-Death chain has a unique stationary distribution given by (115).

In the following example, we will encounter a new concept of 'absorption' of Markov chains.

Exercise 6.38 (Gambler's ruin)**.** Suppose a gambler has fortune of k dolors initially and starts gambling. At each time he wins or loses 1 dolor independently with probability p and $1-p$, respectively. The game ends when his fortune reaches either 0 or N dolors. What is the probability that he wins N dolors and goes home happy?

We use Markov chains to model his fortune after betting t times. Namely, let $\Omega = \{0, 1, 2, \cdots, N\}$ be the state space. Let $(X_t)_{t\geq 0}$ be a sequence of RVs where X_t is the gambler's fortune after betting t times. Note that the transition probabilities are similar to that of the Birth-Death chain, except the 'absorbing boundary' at 0 and N. Namely,

$$\begin{cases} \mathbb{P}(X_{t+1} = k+1 \,|\, X_t = k) = p & \forall 1 \le k < N \\ \mathbb{P}(X_{t+1} = k \,|\, X_t = k-1) = 1-p & \forall 1 \le k < N \\ \mathbb{P}(X_{t+1} = 0 \,|\, X_t = 0) = 1 \\ \mathbb{P}(X_{t+1} = N \,|\, X_t = N) = 1. \end{cases} \tag{116}$$

Call the resulting Markov chain $(X_t)_{t\ge 0}$ the *gambler*'s chain. Its state space diagram is given below.

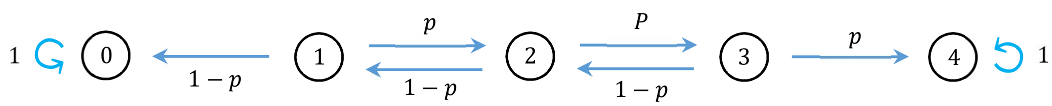

FIGURE 8. State space diagram of a 5-state gambler's chain

(i) Show that any distribution $\pi = [a, 0, 0, \cdots, 0, b]$ on Ω is stationary with respect to the gambler's chain. Also show that any stationary distribution of this chain should be of this form.

(ii) Clearly the gambler's chain eventually visits state 0 or N, and stays at that boundary state thereafter. This is called *absorbtion*. Let τ_i denote the time until absorbtion starting from state i:

$$\tau_i = \min\{t \ge 0 : X_t \in \{0, N\} \,|\, X_0 = i\}. \tag{117}$$

We are going to compute the 'winning probabilities': $q_i := \mathbb{P}(X_{\tau_i} = N)$.

By considering what happens in one step, show that they satisfy the following recursion

$$\begin{cases} q_i = p q_{i+1} + (1-p) q_{i-1} & \forall 1 \le i < N \\ q_0 = 0, \quad q_N = 1 \end{cases}. \tag{118}$$

(iii) Denote $\rho = (1-p)/p$. Show that

$$q_{i+1} - q_i = \rho(q_i - q_{i-1}) \qquad \forall 1 \le i < N. \tag{119}$$

Deduce that

$$q_{i+1} - q_i = \rho^i (q_1 - q_0) = \rho^i q_1 \qquad \forall 1 \le i < N, \tag{120}$$

and that

$$q_i = q_1 (1 + \rho + \cdots + \rho^{i-1}) \qquad \forall 1 \le i \le N. \tag{121}$$

(iv)* Conclude that

$$q_i = \begin{cases} \frac{1-\rho^i}{N - \rho(1-\rho^N)/(1-\rho)} & \text{if } p \ne 1/2 \\ \frac{2i}{N(N-1)} & \text{if } p = 1/2. \end{cases} \tag{122}$$

REFERENCES

[BT02] Dimitri P Bertsekas and John N Tsitsiklis, *Introduction to probability*, vol. 1, Athena Scientific Belmont, MA, 2002.