IOE 515 Stochastic Process♠

Fall 2000

Qin Lei
Department of Finance
University of Michigan

♠ These notes are based without restraint on the class notes provided by Professor Shane Henderson.

CHAPTER 1 INTRODUCTION TO PROBABILITY

1.1 Probability Spaces

Definition 1.1 An *experiment* is a specific set of actions the results of which cannot be predicted with certainty.

Definition 1.2 Each possible result of the experiment defines a *sample point* or *outcome*, ω.

Definition 1.3 The set of all outcomes of an experiment is called the *sample space* and is denoted by Ω.

Definition 1.4 An *event* is a subset of the sample space, i.e., it is a group of outcomes. We let $\mathcal{F}$ be the set of all "reasonable" events (the set of all subsets of Ω if Ω is countable.)

Definition 1.5 We assign a probability $\Pr$ to each event in $\mathcal{F}$. We refer to $(\Omega, \mathcal{F}, \Pr)$ as a *probability space*. Such a $\Pr$ must satisfies:
(1) $0 \le \Pr(A) \le 1, \forall A \in \mathcal{F}$;
(2) $\Pr(\Omega) = 1$;
(3) *Countable Additivity*: If $A_1, A_2, A_3, \ldots$, is a finite or countably infinite sequence of mutually exclusive events in F then $\Pr(\bigcup_i A_i) = \sum_i \Pr(A_i)$.

Lemma 1.1 Some consequences of the above three axioms in definition 1.5 are that
(1) $\Pr(\varnothing) = 0$;
(2) If $A \subseteq B$ then $\Pr(A) \le \Pr(B)$;
(3) $\Pr(A^c) = 1 - \Pr(A)$;
(4) (*Boole's Inequality*) $\Pr(\bigcup_{i=1}^{\infty} A_i) \le \sum_{i=1}^{\infty} \Pr(A_i)$.

1.2 Random Variables

Definition 1.6 A *random variable* (r.v.) is a function from Ω (the sample space) to the real line, i.e., a random variable assigns a number to every outcome of the sample space. Random variables are usually denoted by capital letters, e.g., X, Y, and Z. We sometimes write $X(\omega)$ to make the dependence on the outcome ω explicit. When we write $\Pr(X \in A)$ we are in effect writing $\Pr\{\omega : X(\omega) \in A\}$.

Definition 1.7 The *cumulative distribution function* (c.d.f.) or *distribution function* of X is the function $F(x)$ defined by $F(x) = \Pr(X \le x)$. All proper distribution functions $F(x)$ satisfy:
(1) F is non-decreasing;
(2) $\lim_{x \to -\infty} F(x) = 0$;

(3) $\lim_{x\to+\infty} F(x)=1$;
(4) F is continuous from the right.

Definition 1.8 A *discrete* random variable is a r.v. whose range (set of possible values) is finite or countable. If X is discrete, then $F(x)=\Pr(X\le x)=\sum_{y\le x}\Pr(X=y)$.

Definition 1.9 An (absolutely) *continuous* variable, X, is a random variable whose distribution function F is given by $F(x)=\int_{-\infty}^{x}\mathrm{f}(y)dy$ for some function f. The function f is called the *probability density function* (p.d.f.).

Definition 1.10 *Stieltjes Integral*: We define the integral

$$\int_A g(x)dF(x)=\begin{cases}\sum_{x\in A} g(x)\Pr(X=x) & X \text{ discrete}\\ \int_A g(x)\,\mathrm{f}(x)dx & X \text{ continuous}\end{cases}$$

Definition 1.11 The *expected value* or *mean* of X is defined to be

$$\mathrm{E}\,X=\int_{-\infty}^{+\infty} x dF(x)=\begin{cases}\sum_x x\Pr(X=x) & X \text{ discrete}\\ \int_{-\infty}^{+\infty} x\,\mathrm{f}(x)dx & X \text{ continuous}\end{cases}$$

We also define

$$\mathrm{E}[g(x)]=\int_{-\infty}^{+\infty} g(x)dF(x).$$

Definition 1.12 The *variance* of X is defined to be $\mathrm{Var}(X)=\mathrm{E}[(X-\mathrm{E}\,X)^2]=\mathrm{E}(X^2)-[\mathrm{E}(X)]^2$.

Example 1.1 Let the random variable X be uniformly distributed between $[1,3]$. We have the p.d.f. as

$$\mathrm{f}(x)=\begin{cases}\frac{1}{2} & 1\le x\le 3\\ 0 & \text{else.}\end{cases}$$

The cumulative density function is:

$$F(x)=\begin{cases}0 & x<1\\ \frac{1}{2}(x-1) & 1\le x\le 3\\ 1 & x>3.\end{cases}$$

We then have the following results:

$$\mathrm{E}(x)=\int_{-\infty}^{+\infty} x\,\mathrm{f}(x)dx=\int_1^3 x\cdot\tfrac{1}{2}\,dx=2,$$
$$\mathrm{E}(x^2)=\int_{-\infty}^{+\infty} x^2\,\mathrm{f}(x)dx=\int_1^3 x^2\cdot\tfrac{1}{2}\,dx=\tfrac{13}{3},$$
$$\mathrm{Var}(x)=\mathrm{E}(x^2)-\mathrm{E}(x)^2=\tfrac{13}{3}-4=\tfrac{1}{3}.$$

Comment 1.1 If X is a continuous r.v. with c.d.f. F, then $F(X)$ is uniformly distributed over $(0,1)$.

Proof: Let $F(X) = U$. Apparently, $U \in [0,1]$ satisfies the domain requirement for uniform distribution.

$$\Pr(U \le u) = \Pr[F(X) \le u] = \Pr[X \le F^{-1}(u)] \qquad <F(\cdot) \text{ is non-decreasing}>$$
$$= F[F^{-1}(u)] = u$$

Therefore, U is uniformly distributed over $(0,1)$. QED

Comment 1.2 If U is a uniform r.v. over $(0,1)$, then $F^{-1}(U)$ has distribution F , where $F^{-1}(x)$ is the value of y such that $F(y) = x$.

Proof: Following the property of uniform distribution, we have $\Pr(U \le u) = u$. Let $F^{-1}(U) = Y$.

$$\Pr(Y \le y) = \Pr[F^{-1}(U) \le y] = \Pr[U \le F(y)] \qquad <F(\cdot) \text{ is non-decreasing, so is } F^{-1}(\cdot)>$$
$$= F(y) \qquad <\Pr(U \le u) = u>$$

Therefore, $F^{-1}(U)$ has distribution F . QED

Comment 1.3 The property mentioned in comment 1.2 is often used to generate continuous r.v.s. in simulation. As long as we can define $F^{-1}(\cdot)$, we can easily generate uniform $U(0,1)$ and thus $F^{-1}(U)$ for distribution F .

Definition 1.13 The *joint distribution function* (c.d.f.) F of two random variables X and Y is defined to be $F(x,y) = \Pr(X \le x, Y \le y)$. Given the joint c.d.f. of X and Y , $F_X(x) = \lim_{y\to\infty} F(x,y)$ is the c.d.f. of X and $F_Y(y) = \lim_{x\to\infty} F(x,y)$ is the c.d.f. of Y .

Example 1.2 Let's toss a fair coin twice. Let X be the number of heads shown up and $Y = 1$ if the results of the two tosses are the same and $Y = 0$ otherwise.

We find that $F(x,y) = \Pr(X \le x, Y \le y) = \Pr\{\omega : X(\omega) \le x, Y(\omega) \le y\}$. If X, Y are discrete r.v. then $F(x,y) = \sum_{x' \le x} \sum_{y' \le y} \Pr(x', y')$. If X, Y are jointly continuous r.v.s., then we have $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x', y') dx' dy'$.

Definition 1.14 Two random variables X and Y are *independent* if $F(x,y) = F_X(x) F_Y(y)$ for all x and y , where F is the joint c.d.f. of X and Y ; $F_X(x)$ and $F_Y(y)$ are the marginal c.d.f.'s of X and Y respectively.

Lemma 1.2 If X and Y are independent then $\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\,\mathrm{E}[h(Y)]$ for all functions g and h for which these expectations are well defined. (But the converse may not be true.)

Definition 1.15 The *covariance* of X and Y is defined as

$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X - \mathrm{E}\,X)(Y - \mathrm{E}\,Y)] = \mathrm{E}[XY] - \mathrm{E}(X)\,\mathrm{E}(Y).$$

Example 1.3 Does zero covariance imply independence? Not necessarily. Consider the following example. All points on a unit circle are equally likely distributed and we have

$\text{Cov}(X,Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y) = 0 - 0 \cdot 0 = 0$.
However, we also have $\Pr(X = 0 \mid Y = 1) = 0$ but $\Pr(X = 0) \neq 0$, which implies that X and Y are not independent.

1.3 Moment Generating, Characteristic Functions, and Laplace Transforms

Definition 1.16 The *moment generating function* (m.g.f.) of a r.v. X, written as $\psi_X(t)$, is given by

$$\psi_X(t) = \text{E}(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} dF(x),$$

for all real t. When a moment generating function exists in a neighborhood of 0, it uniquely determines the distribution of X.

Example 1.4 Let $X \sim \text{Exp}(\lambda)$. We have

$$\text{f}(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The moment generating function is then

$\psi_X(t) = \text{E}(e^{tX}) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \int_0^{\infty} \lambda e^{(t-\lambda)x} dx = \frac{\lambda}{t-\lambda} e^{(t-\lambda)x} \mid_0^{\infty} = \frac{\lambda}{\lambda - t}$ (for $t < \lambda$).

$\psi_X\text{'}(t) = \frac{d}{dt}\text{E}(e^{tX}) = \text{E}(\frac{d}{dt} e^{tX})$ (if interchangability is allowed)

$= \text{E}(Xe^{tX})$.

So $\psi_X\text{'}(0) = \text{E}(Xe^{0 \cdot X}) = \text{E}(X)$. Similarly, $\psi_X\text{''}(t) = \text{E}(X^2 e^{tX})$ and thus $\psi_X\text{''}(0) = \text{E}(X^2)$. Hence $\text{Var}(X) = \text{E}(X^2) - [\text{E}(X)]^2 = \psi_X\text{''}(0) - [\psi_X\text{'}(0)]^2$. In this particular example, $\psi_X\text{'}(t) = \frac{\lambda}{(\lambda - t)^2}$, $\psi_X\text{''}(t) = \frac{2\lambda}{(\lambda - t)^3}$, $\text{E}(X) = \frac{1}{\lambda}$ and $\text{Var}(X) = \frac{1}{\lambda^2}$.

Comment 1.4 The m.g.f. may not be found. For example, the m.g.f. of a Cauchy r.v. can not be found. Can you find the m.g.f. for a distribution that has p.d.f. of $\text{f}(x) = \frac{1}{\pi(1+x^2)} \forall x$? Sometimes, it is more convenient to define characteristic function, which always exists, rather than moment generating function. Note also that there are many cases where we can write down the m.g.f. but there is no explicit distribution, such as the usual cases in queuing applications.

Example 1.5 Let $X \sim \text{Poisson}(\lambda)$, then we have the following m.g.f.:

$$\psi_X(t) = \text{E}[\exp(tX)] = \sum_{x=0}^{\infty} \frac{e^{tx} e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} \qquad \left\langle \sum_{x=0}^{\infty} \frac{\alpha^x}{x!} = e^{\alpha} \right\rangle$$

$$= e^{\lambda(e^t - 1)}.$$

Definition 1.17 The *characteristic function* of a r.v. X, written $\phi_X(u)$, is given by

$$\phi_X(u) = \text{E}(e^{iuX}) = \int_{-\infty}^{+\infty} e^{iux} dF(x),$$

for all real u (where $i = \sqrt{-1}$). Note that ϕ uniquely defines F and vice versa.

Definition 1.18 For non-negative r.v.s. the *Laplace transform* of the r.v. X with c.d.f. F, written $\tilde{F}(s)$, is given by

$$\tilde{F}(s) = \mathrm{E}(e^{-sX}) = \int_0^{+\infty} e^{-sx}\, dF(x),$$

for all real $s \geq 0$.

Example 1.6 Suppose X, Y are independent r.v. with known distribution function. What is the distribution of $X+Y$?

Let $\phi_{X+Y}, \phi_X, \phi_Y$ be the characteristic functions and ϕ_X, ϕ_Y are known.

$$\begin{aligned}
\phi_{X+Y}(u) &= \mathrm{E}\{\exp[iu(X+Y)]\} = \mathrm{E}[\exp(iuX)\exp(iuY)] \\
&= \mathrm{E}[\exp(iuX)] \cdot \mathrm{E}[\exp(iuY)] \qquad \langle \text{By independence of } X \text{ and } Y \rangle \\
&= \phi_X(u) \cdot \phi_Y(u).
\end{aligned}$$

Comment 1.5 Some common distributions and their moments as well as moment generating functions.

(1) Bernoulli Distribution

One experiment with two possible outcomes x, success or failure. $x(\text{success}) = 1$ and $x(\text{failure}) = 0$. The p.d.f. for x is $\mathrm{f}(x) = p^x(1-p)^{1-x}$. $\mu = p$ and $\sigma^2 = p(1-p)$. The m.g.f. is $\psi_X(t) = 1 - p + pe^t$.

(2) Binomial Distribution

x numbers of success within n Bernoulli experiments. The p.d.f. is $\mathrm{f}(x) = \binom{n}{x} p^x (1-p)^{n-x}$. $\mu = np$ and $\sigma^2 = np(1-p)$. The m.g.f. is $\psi_X(t) = (1-p+pe^t)^n$.

(3) Negative Binomial Distribution

x numbers of failures before the k^{th} success within $x+k$ Bernoulli experiments. The p.d.f. is $\mathrm{f}(x,k) = \binom{x+k-1}{k-1} p^k (1-p)^x$.

(4) Geometric Distribution (Two versions with different starting values)

x numbers of failures before the first success within $x+1$ Bernoulli experiments. The p.d.f. is $\mathrm{f}(x) = p(1-p)^x, x \geq 0$. $\mu = \frac{1-p}{p}$ and $\sigma^2 = \frac{1-p}{p^2}$. The m.g.f. is $\psi_X(t) = p[1 - e^t(1-p)]^{-1}$.

$x-1$ numbers of failures before the first success within x Bernoulli experiments. The p.d.f. is $\mathrm{f}(x) = p(1-p)^{x-1}, x \geq 1$. $\mu = \frac{1}{p}$ and $\sigma^2 = \frac{1-p}{p^2}$. The m.g.f is $\psi_X(t) = pe^t / [1-(1-p)e^t]$.

Note that we have different means $\mu = \mathrm{E}(x)$ for two versions because the expectations were calculated based upon different lower bounds of the domain for x.

(5) Hyper-geometric Distribution

There are N balls with R red ones. Get x red balls within n draws without replacement. The p.d.f. is $\mathrm{f}(x) = \binom{R}{x}\binom{N-R}{n-x} / \binom{N}{n}$.

(6) Pareto Distribution

The p.d.f. is $\mathrm{f}(x,\theta) = \theta \cdot x_0^\theta \cdot x^{(\theta+1)}$ for $x \geq x_0$. $\mu = \frac{x_0\theta}{\theta-1}$.

(7) Trinomial Distribution

One experiment with there possible outcomes. Repeat the experiment n times. The p.d.f. is $\mathrm{f}(x,y) = \frac{n!}{x!y!(n-x-y)!} p_1^x \cdot p_2^y \cdot p_3^{n-x-y}$ with $p_1 + p_2 + p_3 = 1$. The m.g.f. is $\psi_{X,Y}(t_1,t_2) = (p_1 e^{t_1} + p_2 e^{t_2} + p_3)^n$.

(8) Multinomial Distribution

One experiment with k possible outcomes. Repeat the experiment n times. The p.d.f. is $\mathrm{f}(x_1,\ldots,x_{k-1}) = \frac{n!}{x_1!x_2!\ldots x_k!} p_1^{x_1} \cdot \ldots \cdot p_k^{x_k}$ with $\sum_k p_k = 1$ and $\sum_k x_k = n$. The m.g.f. is $\psi_{x_1,\ldots,x_{k-1}}(t_1,\ldots t_{k-1}) = (p_1 e^{t_1} + \ldots + p_{k-1}e^{t_{k-1}} + p_k)^n$.

(9) Poisson Distribution

The p.d.f. is $\mathrm{f}(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0,1,2,\ldots$. We can get the Poisson distribution from $\mathrm{Bin}(n,p)$ if we let $n \to \infty$ and $np = \lambda$. $\mu = \sigma^2 = \lambda$. The m.g.f. is $\psi_X(t) = e^{\lambda(e^t-1)}$.

(10) Gamma Distribution

The p.d.f. is $\mathrm{f}(x,\alpha,\beta) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}$ for $x \geq 0, \alpha, \beta > 0$. When $\beta = 1$, we have $\mathrm{f}(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}$, hence $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$. $\mu = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$. The m.g.f. is $\psi_X(t) = (1-\beta t)^{-\alpha}$ for $t < \frac{1}{\beta}$.

(11) Chi-square Distribution

The p.d.f. is $\mathrm{f}(x) = \frac{1}{\Gamma(r/2)\sqrt{2}} x^{\frac{r}{2}-1}e^{-\frac{x}{2}}$. This is the special case of Gamma distribution where $\alpha = \frac{r}{2}$ and $\beta = 2$. The m.g.f. is $\psi_X(t) = (1-2t)^{-\frac{r}{2}}$ for $t < \frac{1}{2}$.

(12) Exponential Distribution

The p.d.f. is $\mathrm{f}(x) = \lambda e^{-\lambda x}$ for $x > 0$. This is the special case of Gamma distribution where $\alpha = 1$ and $\beta = \frac{1}{\lambda}$. $\mu = \frac{1}{\lambda}$ and $\sigma^2 = \frac{1}{\lambda^2}$. The m.g.f. is $\psi_X(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$. The c.d.f. is $F(x) = 1 - e^{-\lambda x}$ for $x > 0$.

(13) Normal Distribution

The p.d.f. is $\mathrm{f}(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$ and the m.g.f. is $\psi_X(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2})$.

(14) Uniform Distribution

The p.d.f. is $\mathrm{f}(x) = \frac{1}{b-a}$ for $a < x < b$. $\mu = \frac{a+b}{2}$ and $\sigma^2 = \frac{(b-a)^2}{12}$. The m.g.f. is $\psi_X(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}$.

(15) Beta Distribution

The p.d.f. is $\mathrm{f}(x,\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{b-1}$ for $0 < x < 1$. $\mu = \frac{\alpha}{\alpha+\beta}$ and $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

1.4. Some Useful Tools

Theorem 1.1 *Fubini's Theorem*

(1) If $X_n \geq 0$ almost surely, then $\mathrm{E}\sum_{n=1}^\infty X_n = \sum_{n=1}^\infty \mathrm{E}\, X_n$.

(2) If $\mathrm{E}\sum_{n=1}^\infty |X_n| < \infty$, then $\mathrm{E}\sum_{n=1}^\infty X_n = \sum_{n=1}^\infty \mathrm{E}\, X_n$.

(3) If $X(t) \geq 0$, then $\mathrm{E}\int_{-\infty}^{+\infty} X(t)dt = \int_{-\infty}^{+\infty} \mathrm{E}\, X(t)dt$.

(4) If $\mathrm{E}\int_{-\infty}^{+\infty} |X(t)|dt < \infty$, then $\mathrm{E}\int_{-\infty}^{+\infty} X(t)dt = \int_{-\infty}^{+\infty} \mathrm{E}\, X(t)dt$.

Note that an expectation is just a sum or integral so the above results hold if the expectations are replaced by sums or integrals. Indeed, Fubini's theorem must be used whenever interchanging infinite sums or integrals.

Comment 1.6 When we calculate the sum or integral, we can insert an identity in the form of sum or integral, and then use *Fubini's Theorem* to simplify the calculation. For example, we can use this trick to show that if X is nonnegative with distribution F, then

$$\mathrm{E}(X) = \int_0^\infty [1 - F(x)]dx \text{ and } \mathrm{E}(X^n) = \int_0^\infty nx^{n-1}[1 - F(x)]dx.$$

Here is how:

$$\begin{aligned}\mathrm{E}(X) &= \int_0^\infty x dF(x) = \int_0^\infty \int_0^x dy dF(x) && \left\langle x = \int_0^x dy \right\rangle \\ &= \int_0^\infty \int_y^\infty dF(x) dy && \langle 0 \le y \le x < \infty;\ \text{Fubini's Theorem} \rangle \\ &= \int_0^\infty [1 - F(y)]dy.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\mathrm{E}(X^n) &= \int_0^\infty x^n dF(x) = \int_0^\infty \int_0^x ny^{n-1} dy dF(x) && \left\langle x^n = \int_0^x ny^{n-1} dy \right\rangle \\ &= \int_0^\infty \int_y^\infty n dF(x) y^{n-1} dy && \langle 0 \le y \le x < \infty;\ \text{Fubini's Theorem} \rangle \\ &= \int_0^\infty n[1 - F(y)]y^{n-1} dy.\end{aligned}$$

Comment 1.7 When using *Fubini's Theorem* to switch sums or integrals, take extra caution when determining the domain of two indexing variables. In the example above, we use two identities, $x = \int_0^x ydy$ and $x^n = \int_0^x ny^{n-1}dy$. Note the fact that $0 \le y \le x < \infty$ when determining the new ranges for the integrals after switching.

Definition 1.19 Define the *indicator variable* of the event A to be $I(A) = 1$ if event A occurs and $I(A) = 0$ otherwise. Then $\mathrm{E}[I(A)] = \Pr(A)$. A useful tool is to multiply an expression by $I(A) + I(A^c)$ (which is identically equal to one) and then calculate the terms individually.

Comment 1.8 Let N be a positive integer valued r.v. For an infinite series $\{X_1, X_2, \ldots X_n, \ldots\}$, where $n = 1, 2, \ldots, \infty$, note that $I(N = n)X_n$ would select the N^{th} element of the original series and $I(N \ge n)X_n$ would select the first N elements of the original series. We now want to prove the very useful result $\mathrm{E}(N) = \sum_{i=1}^\infty \Pr(N \ge i) = \sum_{k=0}^\infty \Pr(N > k)$, a natural counterpart to the result in **Comment 1.6** for positive integer r.v. N.

$$\begin{aligned}\mathrm{E}(N) &= \sum_{n=1}^\infty n \Pr(N = n) && \langle \text{By definition of expectation} \rangle \\ &= \sum_{n=1}^\infty \sum_{i=1}^n \Pr(N = n) && \langle \text{By creating identity} \rangle \\ &= \sum_{i=1}^\infty \sum_{n=i}^\infty \Pr(N = n) && \langle \text{By Fubini's theorem}; 1 \le i \le n < \infty \rangle \\ &= \sum_{i=1}^\infty \Pr(N \ge i) && \langle \text{By countable additivity} \rangle\end{aligned}$$

Lemma 1.3 Suppose $\{X_i\}$ are i.i.d. non-negative r.v.s., independent of N, a positive integer valued r.v., then $\mathrm{E}\left[\sum_{i=1}^N X_i\right] = \mathrm{E}[N]\mathrm{E}[X_i]$.

Proof (Version 1): Use the selector $I(N=n)\sum_{i=1}^{n} X_n$ to get the N^{th} element from the series $\left\{\sum_{i=1}^{1} X_i, \sum_{i=1}^{2} X_i, \ldots \sum_{i=1}^{n} X_i, \ldots\right\}$, where $n = 1,2,\ldots,\infty$. Hence $\sum_{i=1}^{N} X_i = \sum_{n=1}^{\infty}\left[I(N=n)\sum_{i=1}^{n} X_i\right]$.

Now we have

$$
\begin{aligned}
\mathrm{E}\sum_{i=1}^{N} X_i &= \mathrm{E}\sum_{n=1}^{\infty}\left[I(N=n)\sum_{i=1}^{n} X_i\right] \\
&= \sum_{n=1}^{\infty}\mathrm{E}\left[I(N=n)\sum_{i=1}^{n} X_i\right] && \text{<By Fubini's theorem; non-negativity of } X_i\text{>} \\
&= \sum_{n=1}^{\infty}\mathrm{E}[I(N=n)]\,\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] && \text{<By independence of } X_i \text{ and } N\text{>} \\
&= \mathrm{E}(X_i)\sum_{n=1}^{\infty} n\Pr[N=n] \\
&= \mathrm{E}(N)\,\mathrm{E}(X_i).\ \text{QED}
\end{aligned}
$$

Proof (Version 2): Use the selector $I(N \ge n)X_n$ to get the first N elements from the series $\{X_1, X_2, \ldots, X_n, \ldots\}$, where $n = 1,2,\ldots,\infty$. Hence we have the following proof:

$$
\begin{aligned}
\mathrm{E}\sum_{i=1}^{N} X_i &= \mathrm{E}\left(\sum_{n=1}^{\infty} I(N \ge n)X_n\right) && \text{<By definition of selector>} \\
&= \sum_{n=1}^{\infty}\mathrm{E}[I(N \ge n)X_n] && \text{<By Fubini's Theorem; non-negativity of } X_n\text{>} \\
&= \sum_{n=1}^{\infty}\mathrm{E}[I(N \ge n)]\,\mathrm{E}(X_n) && \text{<By independence of } N \text{ and } X_n\text{>} \\
&= \mathrm{E}(X_i)\sum_{n=1}^{\infty}\Pr(N \ge n) && \text{<By definition of indicator variable>} \\
&= \mathrm{E}(N)\,\mathrm{E}(X_i).\ \text{QED} && \text{<By \textbf{Comment 1.8} proved above>}
\end{aligned}
$$

Comment 1.9 Here are three useful identities regarding exponential function.

(1) $e^{\alpha} = \lim_{n\to\infty}(1+\frac{\alpha}{n})^n$

(2) $e^{\alpha} = \sum_{n=0}^{\infty}\frac{\alpha^n}{n!}$

(3) $\int_{-\infty}^{+\infty} e^{-at^2}dt = \sqrt{\frac{\pi}{a}}$

Proof for (1):

$$
\begin{aligned}
\lim_{n\to\infty}(1+\tfrac{\alpha}{n})^n &= \lim_{n\to\infty}\{\exp[n\ln(1+\tfrac{\alpha}{n})]\} = \lim_{n\to\infty}\left\{\exp\left[\frac{\ln(1+\frac{\alpha}{n})}{\frac{1}{n}}\right]\right\} \\
&= \exp\left\{\lim_{n\to\infty}\frac{\frac{1}{1+\frac{\alpha}{n}}\left(-\frac{\alpha}{n^2}\right)}{-\frac{1}{n^2}}\right\} = \exp\left\{\lim_{n\to\infty}\left[\frac{\alpha}{1+\frac{\alpha}{n}}\right]\right\} = e^{\alpha}.\ \text{QED}
\end{aligned}
$$

Proof for (2): Using Taylor series expansion to expand e^{α} around $\alpha = 0$. QED

Proof for (3): $\int_{-\infty}^{+\infty} e^{-at^2}dt = \sqrt{\frac{\pi}{a}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}(\frac{1}{\sqrt{2a}})}\exp\left\{-\frac{1}{2}\left(\frac{t}{\frac{1}{\sqrt{2a}}}\right)^2\right\}dt = \sqrt{\frac{\pi}{a}}$. QED

Comment 1.10 Here are two useful formula for calculating the derivative of integration:

(1) $\frac{d}{dx}\int_a^x \mathrm{f}(t)dt = \mathrm{f}(x)$ and $\frac{d}{dx}\int_x^b \mathrm{f}(t)dt = -\mathrm{f}(x)$

(2) $\frac{d}{dx}\int_a^b \mathrm{f}(x,t)dt = \int_a^b \mathrm{f}'_x(x,t)dt$ and $\frac{d}{dx}\int_c^{\infty} \mathrm{f}(x,t)dt = \int_c^{\infty} \mathrm{f}'_x(x,t)dt$

(3) $\frac{d}{dx}\int_{u(x)}^{v(x)} \mathrm{f}(x,t)dt = \mathrm{f}[x,v(x)]v'(x) - \mathrm{f}[x,u(x)]u'(x) + \int_{u(x)}^{v(x)} \mathrm{f}'_x(x,t)dt$

The property (3) is also called *Leibniz's Rule*, and the properties (1) and (2) are special cases of *Leibniz's Rule*.

1.5 Conditional Probability and Expectation

Definition 1.20 Let X and Y be discrete r.v.s., then we have $\Pr(X = x \mid Y = y) = \frac{\Pr(X=x,Y=y)}{\Pr(Y=y)}$, for all y such that $\Pr(Y = y) > 0$.

Definition 1.21 $\mathrm{E}(X \mid Y = y) = \sum_x x \Pr(X = x \mid Y = y)$. If X and Y are independent then $\mathrm{E}(X \mid Y = y) = \mathrm{E}(X)$.

Definition 1.22 Define the expectation of X given Y , $\mathrm{E}(X \mid Y)$, to be the r.v. that takes on the value $\mathrm{E}(X \mid Y = y)$ whenever $Y = y$.

Theorem 1.2 $\mathrm{E}(X) = \mathrm{E}[\mathrm{E}(X \mid Y)]$ if $\mathrm{E}(|X|)$ is finite or $X \geq 0$ almost surely.

Proof in the discrete case:

$$\begin{aligned}\mathrm{E}[\mathrm{E}(X|Y)] &= \sum_y \mathrm{E}(X \mid Y = y)\Pr(Y = y) = \sum_y \sum_x x \Pr(X = x \mid Y = y)\Pr(Y = y) \\ &= \sum_y \sum_x x \Pr(X = x, Y = y) = \sum_x x \sum_y \Pr(X = x, Y = y) \\ &= \sum_x x \Pr(X = x) = \mathrm{E}(X). \text{ QED}\end{aligned}$$

Example 1.7 Suppose a manufacture produces a lot of n items. Each item is defective with probability P , where P is a r.v. with c.d.f. F and mean p_0 . What is the expected number of defective items in the lot?

Let N be the number of defective items in the lot. Apparently the conditional expectation is $\mathrm{E}(N \mid P = p) = np$, which is a result similar to Binomial distribution, but here we don't need the independence for each examination. Then the unconditional mean becomes

$$\mathrm{E}(N) = \int_{-\infty}^{+\infty} \mathrm{E}(N \mid P) dF(P) = \int_{-\infty}^{+\infty} nP dF(P) = np_0.$$

Alternatively, we have $\mathrm{E}(N \mid P) = nP$ since $\mathrm{E}(N \mid P = p) = np$. Thus

$$\mathrm{E}(N) = \mathrm{E}[\mathrm{E}(N \mid P)] = \mathrm{E}(nP) = n\,\mathrm{E}(P) = np_0 .$$

Armed with the **Theorem 1.2**, we now try to prove **Lemma 1.3** in a third way.

By the Theorem 1.2, we have $\mathrm{E}\left[\sum_{i=1}^N X_i\right] = \mathrm{E}\left\{\mathrm{E}\left[\sum_{i=1}^N X_i \mid N\right]\right\}$. Since

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^N X_i \mid N = n\right] &= \mathrm{E}\left[\sum_{i=1}^n X_i\right] \qquad \text{<By independence of } X_i \text{ and } N\text{>} \\ &= n\,\mathrm{E}(X_i),\end{aligned}$$

we have $\mathrm{E}\left[\sum_{i=1}^N X_i\right] = \mathrm{E}\left\{\mathrm{E}\left[\sum_{i=1}^N X_i \mid N\right]\right\} = \mathrm{E}\left[N\,\mathrm{E}(X_i)\right] = \mathrm{E}(X_i)\mathrm{E}(N)$. QED

Example 1.8 Suppose that $X_1, X_2, \ldots$ is a sequence of independent Bernoulli random variables (trials) with $\Pr(X_i = 1) = p = 1 - \Pr(X_i = 0)$. Let N be the number of trials until the first failure ($X_i = 0$). What is $E(N)$ and $E(N^2)$?

By definition, we have $E(N) = \sum_{n=1}^{\infty} n \Pr(N = n) = \sum_{n=1}^{\infty} n(1-p)p^{n-1}$, but this is not a good way to proceed. Using the Theorem 1.2 again, we have $E(N) = E\left[E(N \mid X_1)\right]$, where X_1 is the result of the first trial. If $X_1 = 0$, then the experiment ends with $\{N \mid X_1 = 0\} = 1$; otherwise, we get a recursive game with $\{N \mid X_1 = 1\} = 1 + N'$, where N' is identically distributed as N. Therefore, we have the following results

$$E(N \mid X_1 = 0) = 1 \text{ and } E(N^2 \mid X_1 = 0) = 1;$$

$$E(N \mid X_1 = 1) = 1 + E(N) \text{ and } E(N^2 \mid X_1 = 1) = E[(1+N)^2] = 1 + 2E(N) + E(N^2).$$

Therefore, we have

$$E(N) = 1 \cdot (1-p) + [1 + E(N)] \cdot p \Rightarrow E(N) = \tfrac{1}{1-p},$$

and

$$E(N^2) = 1 \cdot (1-p) + \{1 + 2E(N) + E(N^2)\} \cdot p \Rightarrow E(N^2) = \tfrac{1+p}{(1-p)^2}.$$

Example 1.9 "Thief of Baghdad" Suppose the thief is in a dungeon that has three doors. One door leads to freedom immediately, and the second door leads back to the dungeon after one day, and the third door leads back to the dungeon after three days. It is assumed that the thief would forget which door he previously took once he gets back to the dungeon, and further that the probability of getting to each of the three doors is equal. What is the expected number of days before the thief can get out of the dungeon?

Let N be the number of days until freedom and X_1 be the consequence of the door chosen at the very beginning. The theorem 1.2 tells us that $E(N) = E\left\{E\left[N \mid X_1\right]\right\}$. We know that $E(N \mid X_1 = 0) = 0$, $E(N \mid X_1 = 1) = 1 + E(N)$, and $E(N \mid X_1 = 3) = 3 + E(N)$. Hence we have

$$E(N) = E(N \mid X_1 = 0) \cdot \tfrac{1}{3} + E(N \mid X_1 = 1) \cdot \tfrac{1}{3} + E(N \mid X_1 = 3) \cdot \tfrac{1}{3} = [4 + 2E(N)]/3.$$

Finally, we have $E(N) = 4$.

Comment 1.10 It is a good habit to define clearly the events before tackling a problem in probability. It is very important to define an event X such that the game conditional upon the occurrence of X will evolve into a recursive game.

Theorem 1.3 *Law of Total Probability* Let A be some event and Y some r.v. with c.d.f. F, then $\Pr(A) = \int_y \Pr(A \mid Y = y) dF(y)$.

Proof: First of all, $\Pr(A) = E\left[I(A)\right] = E\left\{E\left[I(A) \mid Y\right]\right\}$. Then $E\left[I(A) \mid Y = y\right] = \Pr(A \mid Y = y)$, thus we have $E\left\{E\left[I(A) \mid Y\right]\right\} = \int_y \Pr(A \mid Y = y) dF(y)$. QED

Example 1.10 "Monte Hall's Problem" There is one prize behind one of three doors. You are asked to choose one door, and suppose you pick door 1. Before door 1 is opened, Monte Hall would open one of two doors left, say, door 3, showing you that there is nothing behind. Should you switch your choice to door 2 or stick with door 1?

For $i = 1,2,3$, let $C(i) = 1$ be the event that you choose the i^{th} door; here we have $C(1) = 1$. Let $D(i) = 1$ be the event that the prize is behind the i^{th} door. Let $X(i) = 1$ be the event that Monte Hall opens the i^{th} door; here we have $X(3) = 1$.

Before Monte Hall opens the third door, we have $\Pr[C(i) = 1, D(j) = 1] = \frac{1}{3}, \forall i, j$. After Monte Hall opened the third door, however, we need update our believes according to Bayes' rule.

Given the choice of door 1, the probability that Monte Hall opens the third door is:

$$\Pr[X(3) = 1 \mid C(1) = 1] = \Pr[X(3) = 1 \mid C(1) = 1, D(1) = 1] \cdot \Pr[C(1) = 1, D(1) = 1]$$
$$+ \Pr[X(3) = 1 \mid C(1) = 1, D(2) = 1] \cdot \Pr[C(1) = 1, D(2) = 1] = \tfrac{1}{2} \cdot \tfrac{1}{3} + 1 \cdot \tfrac{1}{3} = \tfrac{1}{2}.$$

Note that here we use the discrete version of *Law of Total Probability*.

Given the choice of door 1 and the fact that nothing is behind door 3, the probability that the prize is behind door 1 is given by

$$\Pr\{D(1) = 1 \mid [X(3) = 1 \mid C(1) = 1]\} = \frac{\Pr\{D(1) = 1, [X(3) = 1 \mid C(1) = 1]\}}{\Pr[X(3) = 1 \mid C(1) = 1]}$$
$$= \frac{\Pr[X(3) = 1 \mid C(1) = 1, D(1) = 1] \cdot \Pr[C(1) = 1, D(1) = 1]}{\Pr[X(3) = 1 \mid C(1) = 1]} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \tfrac{1}{3}.$$

Similarly, we have

$$\Pr\{D(2) = 1 \mid [X(3) = 1 \mid C(1) = 1]\} = \frac{\Pr[X(3) = 1 \mid C(1) = 1, D(2) = 1] \cdot \Pr[C(1) = 1, D(2) = 1]}{\Pr[X(3) = 1 \mid C(1) = 1]} = \tfrac{2}{3}.$$

Therefore, you should always switch the door upon such a choice is given.

Example 1.11 Suppose that X and Y are independent, and $X \sim F, Y \sim G$. Find $\Pr(X \le Y)$.

$$\Pr(X \le Y) = \int_{-\infty}^{+\infty} \Pr(X \le Y \mid Y = y) dG(y)$$
$$= \int_{-\infty}^{+\infty} \Pr(X \le y \mid Y = y) dG(y) \qquad \text{<By Theorem 1.2>}$$
$$= \int_{-\infty}^{+\infty} \Pr(X \le y) dG(y) \qquad \text{<By independence>}$$
$$= \int_{-\infty}^{+\infty} F(y) dG(y).$$

Example 1.12 Suppose that X and Y are independent, and $X \sim F, Y \sim G$. Find the distribution of $X + Y$.

$$\Pr(X + Y \le z) = \int_{-\infty}^{+\infty} \Pr(X + Y \le z \mid Y = y) dG(y) = \int_{-\infty}^{+\infty} \Pr(X \le z - y) dG(y) = \int_{-\infty}^{+\infty} F(z - y) dG(y).$$

This result is actually so useful that we give it a special name, called *convolution*.

Definition 1.23 If F and G are two distribution functions, then the *convolution* of F and G , written $F \otimes G$, is given by $F \otimes G(z) = \int_{-\infty}^{+\infty} F(z - y) dG(y)$. Note that if X has c.d.f. F and Y has c.d.f. G and X and Y are independent then $X + Y$ has c.d.f. $F \otimes G$.

Example 1.13 If X_1, X_2 are independent exponential r.v.s. with mean λ, what is the distribution of $X+Y$?

We have proved previously for two independent r.v.s. $\psi_{X+Y}(t) = \psi_X(t) \cdot \psi_Y(t)$, hence we have $\psi_{X+Y}(t) = \frac{\lambda}{\lambda-t} \cdot \frac{\lambda}{\lambda-t} = \frac{\lambda^2}{(\lambda-t)^2}$. It turns out that $X+Y$ is distributed as a $\text{Gamma}(2,\lambda)$ r.v. The result for the generalized case is given in the following lemma.

Lemma 1.4 If $X_1, X_2, X_3, \ldots$ is a sequence of independent $\text{Exp}(\lambda)$ r.v.s., then $\sum_{i=1}^n X_i$ is distributed as an $\text{Gamma}(n,\lambda)$ r.v.

Lemma 1.5 If X and Y are r.v.s. with $\mathrm{E}\left[|XY|\right] < \infty$ then $\mathrm{E}\left[X\,\mathrm{E}\left[Y \mid X\right]\right] = \mathrm{E}\left[\mathrm{E}\left[XY \mid X\right]\right]$.

Proof:

$$\begin{aligned}\mathrm{E}\left[X\,\mathrm{E}(Y \mid X)\right] &= \sum_x x\,\mathrm{E}(Y \mid X=x)\Pr(X=x) = \sum_x \mathrm{E}(xY \mid X=x)\Pr(X=x) \\ &= \sum_x \mathrm{E}(XY \mid X=x)\Pr(X=x) = \mathrm{E}(XY). \text{ QED}\end{aligned}$$

This lemma says that when finding $\mathrm{E}(XY \mid X)$, X can be factored out of the expectation operator as if it were a constant.

1.6 Probability Inequalities

Lemma 1.6 *Markov's Inequality* If X is a nonnegative random variable, then for any $x > 0$, $\Pr(X \ge x) \le \mathrm{E}(X)/x$.

Proof of version 1:

$$\begin{aligned}X &= X \cdot I(X<x) + X \cdot I(X \ge x) && \langle\text{By definition of indicator variable}\rangle \\ &\ge X \cdot I(X \ge x) && \langle\text{By non-negativity of } X\rangle \\ &\ge x \cdot I(X \ge x) && \langle\text{By the fact that } X \ge x \text{ by now}\rangle\end{aligned}$$

Hence $\mathrm{E}(X) \ge x \cdot \Pr(X \ge x) \Rightarrow \Pr(X \ge x) \le \mathrm{E}(X)/x$. QED

Proof of version 2:

$$\begin{aligned}\mathrm{E}(X) &= X \cdot \Pr(X \ge x) + X \cdot \Pr(X < x) \ge X \cdot \Pr(X \ge x) && \langle\Pr(X \ge x) \ge 0, X \ge 0\rangle \\ &\ge x\Pr(X \ge x). && \langle X \ge x\rangle \text{ QED}\end{aligned}$$

Lemma 1.7 *Chebychev's Inequality* If X is a random variable and both c and ε are constants, where $\varepsilon > 0$, then we have $\Pr(|X-c| > \varepsilon) \le \mathrm{E}(X-c)^2/\varepsilon^2$.

Proof: $\Pr(|X-c| > \varepsilon) = \Pr(|X-c|^2 > \varepsilon^2) \le \mathrm{E}(X-c)^2/\varepsilon^2$, by *Markov's Inequality*. QED

Lemma 1.8 *Chernoff Bounds* If X is a random variable with moment generating function $\psi_X(t) = \mathrm{E}(e^{tX})$, then for $x > 0$, $\Pr(X \ge x) \le e^{-tx}\psi_X(t), \forall t > 0$, or $\Pr(X \le x) \le e^{-tx}\psi_X(t), \forall t < 0$.

Proof:

$$\Pr(X \ge x) = \Pr(e^{tX} \ge e^{tx}) \qquad \text{<increasing exponential function>}$$
$$\le \mathrm{E}(e^{tX}) / e^{tx} \qquad \text{<by Markov's Inequality>}$$
$$= e^{-tx}\psi_X(t). \text{ QED}$$

Comment 1.11 The *Markov's Inequality* is very useful when we would like to find out the tail probabilities yet the only thing we know about the distribution is the first moment. Note that *Chernoff Bounds* is less useful in practice in the sense that *Markov's Inequality* requires only first moment yet the former requires more information.

Lemma 1.9 *Jensen's Inequality* If f is a convex function and $\mathrm{E}(X), \mathrm{E}[\mathrm{f}(X)] < \infty$, then $\mathrm{E}[\mathrm{f}(X)] \ge \mathrm{f}[\mathrm{E}(X)]$.
Proof: We assume that $\mathrm{f}(x)$ is twice differentiable and thus use the Tyler Series Expansion to expand $\mathrm{f}(x)$ around the mean $\mu = \mathrm{E}(X)$.

$$\mathrm{f}(x) \cong \mathrm{f}(\mu) + \mathrm{f}'(\mu)(x-\mu) + \tfrac{1}{2}\mathrm{f}''(\mu)(x-\mu)^2 \ge \mathrm{f}(\mu) + \mathrm{f}'(\mu)(x-\mu) \qquad \text{<By convexity>}$$

This implies that $\mathrm{f}(X) \ge \mathrm{f}(\mu) + \mathrm{f}'(\mu)(X-\mu)$.
Taking expectation on both sides leads to $\mathrm{E}[\mathrm{f}(X)] \ge \mathrm{f}[\mathrm{E}(X)]$. QED

Example 1.14 In general, we have $\mathrm{f}[\mathrm{E}(X)] \ne \mathrm{E}[\mathrm{f}(X)]$. An extreme case of this inequality is given in the example below. Let X stand for the position of a drunk person wandering uniformly over the width of the highway and f stand for the result of either the drunk person is alive or dead. Apparently, we have $\mathrm{E}(X) = \text{center line}$, and $\mathrm{f}[\mathrm{E}(X)] = \text{alive}$. But we also have $\mathrm{E}[\mathrm{f}(X)] = \text{dead}$. How different are $\mathrm{f}[\mathrm{E}(X)]$ and $\mathrm{E}[\mathrm{f}(X)]$? Life and death.

1.7 Types of Convergence

Definition 1.24 A sequence of random variables $X_n : n \ge 1$ is said to *converge almost surely* or *with probability 1* to X if $\Pr(X_n \to X) = 1$. We denote *convergence almost surely* by $X_n \to X$.

Example 1.15 Let's spin a spinner once. The final angle will be within $[0, 360)$, and the gambler can choose one of the following game. In game n, the gambler gets $\$2^n$ if the final angle is within $[0, \frac{360}{n}]$. Let X_n denote the event that the gambler wins in game n. If we define $X = 0$ then $X_n \to X$ almost surely. Why? If the final angle is in $(0, 360)$, then we have $X_n = 0$ for large enough n. (Note that if the final angle is 0, then $X_n \ne 0$.) Furthermore, $\Pr(\text{The final angle is between (0,360)}) = 1$.
In the same time, we have $\mathrm{E}(X_n) = 2^n \cdot \Pr(X_n = \mathit{win}) + 0 \cdot \Pr(X_n = \mathit{lose}) = 2^n(\frac{360}{n} / 360) = \frac{2^n}{n}$. Note that $\lim_{n\to\infty} \mathrm{E}(X_n) \to \infty$.

Definition 1.25 A sequence of random variables $X_n : n \ge 1$ is said to *converge in probability* to X ($X_n \xrightarrow{P} X$) if for all $\varepsilon > 0$, $\lim_{n\to\infty} \Pr(|X_n - X| \ge \varepsilon) = 0$.

Definition 1.26 A sequence of random variables $X_n : n \geq 1$ is said to *converge in distribution* to X ($X_n \Rightarrow X$) if for all x such that $\Pr(X = x) = 0$, $\lim_{n\to\infty} \Pr(X_n \leq x) = \Pr(X \leq x)$.

Theorem 1.4 If $X_n \to X$ almost surely then $X_n \xrightarrow{P} X$. If $X_n \xrightarrow{P} X$ then $X_n \Rightarrow X$.

Example 1.16 Toss a sequence of fair coins. $X_n = 1$ if the n^{th} toss is a head and $X_n = 0$ if it is a tail. Hence $\Pr(X_n = 1) = \frac{1}{2} = \Pr(X_n = 0)$ and $X_n \Rightarrow X$, where X has the distribution function of:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & x \geq 1. \end{cases}$$

1.8 Probability Limit Theorem

Theorem 1.5 *The Strong Law of Large Numbers*

Let $X_1, X_2, \ldots$ be i.i.d. r.v.s with mean $\mu < \infty$. Let $S_n = \sum_{i=1}^n X_i$, then $\lim_{n\to\infty} \frac{S_n}{n} = \mu$ almost surely.

Example 1.17 Let $X_i = 1$ if the i^{th} toss is a head and $X_n = 0$ if it is a tail. We then have $\frac{S_n}{n} = \frac{X_1 + \ldots + X_n}{n} \Rightarrow \mathrm{E}(X_1) = \frac{1}{2}$.

Theorem 1.6 *The Central Limit Theorem*

Let $X_1, X_2, \ldots$ be i.i.d. r.v.s with mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Let $S_n = \sum_{i=1}^n X_i$, then

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow N(0,1).$$

Comment 1.12 The intuition behind the Central Limit Theorem is that although S_n is $\sqrt{n}$ times more variable than X_1, but the mean of S_n grows even faster. That is, the variance of S_n is increasing as n becomes larger, but not as fast as the mean does.

CHAPTER 2 STOCHASTIC PROCESSES AND BROWNIAN MOTION

Definition 2.1 A *stochastic process* is an indexed collection of random variables, i.e., $X = \{X(t): t \in T\}$, where T is some index set. We often interpret t as time. Some common sets T are $T = \{0,1,2,...\}$ (discrete time) and $T = [0,\infty]$ (continuous time). A realization of X is referred to as a *sample path.*

Example 2.1 Suppose that $Z \sim N(0,1)$ and $Y(t) = \sqrt{t}Z, \forall t \geq 0$. Then we have a stochastic process $\{Y(t): t \geq 0\}$ and $Y(t) \sim N(0,t)$.

Theorem 2.1 (*Kolmogrov's Extension Theorem*) The distribution of a stochastic process $X = \{X(t): t \in T\}$ is *uniquely determined* by the finite dimensional distributions $F(x_1, x_2, ..., x_n) = \Pr(X(t_1) \leq x_1, X(t_2) \leq x_2, ..., X(t_n) \leq x_n)$, for all $t_1 < t_2 < ... < t_n$ and all n and $x_1, x_2, ..., x_n$. (Note that although the distribution is uniquely determined, the sample paths are not. The proof of this theorem can be found in Billingsley.)

Definition 2.2 A continuous time stochastic process $X = \{X(t): t \in T\}$ is said to have *independent increments* if for all n and $t_0 < t_1 < ... < t_n$, the random variables $X(t_1) - X(t_0)$, $X(t_2) - X(t_1)$, …, $X(t_n) - X(t_{n-1})$ are independent. In other words, the magnitude of the change of the process over a time interval is independent of the magnitudes of the changes in the process over other *non-overlapping* time intervals.

Definition 2.3 A continuous time stochastic process $X = \{X(t): t \geq 0\}$ is said to have *stationary increments* if $X(t+s) - X(s)$ has the same distribution as $X(t) - X(0)$ for all $s, t \geq 0$. In other words, the distribution of the change in the process over any time interval depends only on the duration of that interval.

Comment 2.1 If a process has stationary and independent increments, then for $\Delta t > 0$ we have $\{X_n : n \geq 0\}$ as an i.i.d. sequence, where $X_n = X[n \cdot \Delta t] - X[(n-1) \cdot \Delta t]$.

Theorem 2.2 If the continuous time stochastic process $X = \{X(t): t \geq 0\}$ has stationary and independent increments with $X(0) = 0$, then the distribution of the r.v.s. $X(t), \forall t > 0$, uniquely specifies the distribution of the stochastic process $X = \{X(t): t \geq 0\}$.

Example 2.2 Suppose that $X(t) \sim \text{Poisson}(\lambda \cdot t)$. We have $\Pr[X(t) = n] = \frac{1}{n!} e^{-\lambda t} (\lambda t)^n, n = 0,1,....$ For $t_1 < t_2 < t_3$, we have

$\Pr[X(t_1) = 1, X(t_2) = 3, X(t_3) = 6]$

$= \Pr[X(t_1) - X(0) = 1, X(t_2) - X(t_1) = 2, X(t_3) - X(t_2) = 3]$ <define $X(0) = 0$>

$= \Pr[X(t_1) - X(0) = 1] \cdot \Pr[X(t_2) - X(t_1) = 2] \cdot \Pr[X(t_3) - X(t_2) = 3]$ <by independent increments>

$= \Pr[X(t_1)=1]\cdot\Pr[X(t_2-t_1)=2]\cdot\Pr[X(t_3-t_2)=3]$ <by stationary increments>

$= \frac{e^{-\lambda t_1}\lambda t_1}{1!}\cdot\frac{e^{-\lambda(t_2-t_1)}[\lambda(t_2-t_1)]^2}{2!}\cdot\frac{e^{-\lambda(t_3-t_2)}[\lambda(t_3-t_2)]^3}{3!}.$

Definition 2.4 A stochastic process $X=\{X(t):t\ge 0\}$ is said to be a *Brownian motion process* with drift $\mu\in\mathcal{R}$ and variance $\sigma^2>0$ if the following three conditions hold:
(1) $X(t)\sim N(\mu t,\sigma^2 t),\forall t\ge 0$ (so we have $X(0)=0$)
(2) $X=\{X(t):t\ge 0\}$ has stationary independent increments
(3) $X=\{X(t):t\ge 0\}$ has continuous sample paths.

Comment 2.2 (1) and (2) can be shown to imply (3) almost surely but we will assume (3) everywhere. If $\mu=0$ and $\sigma^2=1$ then $X=\{X(t):t\ge 0\}$ is *standard Brownian motion*, which is often also called the *Weiner process*. Also note that financial time series is not geometric Brownian motion, but Geometric Brownian motion is a good approximation of financial time series. Brownian motion is infinitely bumpy, without any flat portions, everywhere continuous but nowhere differentiable.

Example 2.3 Let $\{Y_i:i\ge 1\}$ be an i.i.d. sequence with $\Pr(Y_i=1)=\Pr(Y_i=-1)=\frac{1}{2}$. We have $\mathrm{E}(Y_i)=0$ and $\mathrm{Var}(Y_i)=1$. Let $S_0=0$ and $S_n=Y_1+...+Y_n$. Let's define further $X_n^1(t)=S_{\lfloor nt\rfloor}$, where $\lfloor nt\rfloor$ is a floor function taking the smallest integer of nt. We also define $X_n^2(t)=S_{\lfloor nt\rfloor}/n$.

For given t, we have
$\lim_{n\to\infty}X_n^2(t)=\lim_{n\to\infty}t\cdot\frac{S_{\lfloor nt\rfloor}}{nt}=t\cdot\lim_{n\to\infty}\frac{S_{\lfloor nt\rfloor}}{nt}=t\,\mathrm{E}(Y_1)=0$ <by Strong Law of Large Numbers almost surely>, which is not very interesting.

Redefine $X_n(t)=S_{\lfloor nt\rfloor}/\sqrt{n}$. We get that $\mathrm{E}\left[\frac{S_{\lfloor nt\rfloor}}{\sqrt{nt}}\right]=\frac{\lfloor nt\rfloor\mathrm{E}(Y_i)}{\sqrt{nt}}=0$ and $\mathrm{Var}\left[\frac{S_{\lfloor nt\rfloor}}{\sqrt{nt}}\right]=\frac{\lfloor nt\rfloor\cdot 1}{nt}$. By the Central Limit Theorem, we have $\frac{S_{\lfloor nt\rfloor}}{\sqrt{nt}}\Rightarrow N(0,1)$ as $n\to\infty$, which implies that $X_n(t)\Rightarrow N(0,t)$ as $n\to\infty$. Now we claim that $\{X_n(t):t\ge 0\}$ has approximately stationary independent increments. Why?

The increments are given by $X_n(t_2)-X_n(t_1)=\frac{1}{\sqrt{n}}\left(S_{\lfloor nt_2\rfloor}-S_{\lfloor nt_1\rfloor}\right)=\frac{1}{\sqrt{n}}\left(Y_{\lfloor nt_1\rfloor+1}+\cdots+Y_{\lfloor nt_2\rfloor}\right)$. Clearly, the independence of Y_i implies the independence of increments. Since the term in bracket depends on only the number of terms, which is a function of (t_2-t_1), we have stationary increments. Therefore, we conclude that $\{X_n(t):t\ge 0\}$ is a Brownian motion.

Proposition 2.1 Parts (1) and (2) in **Definition 2.4** are equivalent to
(1) $X(t)-X(s)\sim N[\mu(t-s),\sigma^2(t-s)]\ \forall s,t\ge 0, s<t$;
(2) $\{X(t);t\ge 0\}$ has independent increments; and
(3) $X(0)=0$.

Proof: It is easy to show that the two parts in definition 2.4 imply the proposition 2.1. If part (1) in the proposition 2.1 holds, then we know the increments are stationary, and thus we have definition 2.4.

Proposition 2.2 Parts (2) and (3) of **Definition 2.4** together with $X(0) = 0$ imply part (1) of **Definition 2.4**. (The proof is hard and omitted here.)

Proposition 2.3 Suppose $\{W(t); t \geq 0\}$ is a Weiner process. For all $t \geq 0$, let $X(t) = \mu t + \sigma W(t)$, then $\{X(t); t \geq 0\}$ is Brownian motion with drift $\mu \in \mathcal{R}$ and variance $\sigma^2 > 0$.
Proof: Let's check the three qualifications for a Brownian motion. The continuous sample path follows from the linear transformation $X(t) = \mu t + \sigma W(t)$. Furthermore, we have

$$\begin{aligned} X(t) - X(s) &= \mu(t-s) + \sigma[W(t) - W(s)] \\ &= \mu(t-s) + \sigma[W(t-s)] \quad \text{<by stationary increments of Weiner process>} \\ &= X(t-s), \end{aligned}$$

which implies that $X(t) - X(s)$ are stationary increments. Increments $X(t) - X(s)$ are also independent because of the independence of increments in Weiner Process. QED

Example 2.4 Suppose X is a (μ, σ)-Brownian motion. Let $a > 0$ and $W(t) = \frac{1}{a} X(a^2 t), \forall t$. (a) Show that W is also a Brownian motion and give its parameters. (b) How would the formula for W need to be adjusted for W to be a (μ, σ)-Brownian motion?

(a) Clearly, we have $W(0) = 0$ and W has a continuous sample path following the continuity of the sample path of Brownian motion X.
$W(t) - W(s) = \frac{1}{a}[X(a^2 t) - X(a^2 s)] = \frac{1}{a} X[a^2(t-s)]$ implies stationary increments of W. W also has independent increments, following independent increments of X.
$X(t) \sim N(\mu t, \sigma^2 t) \Rightarrow W(t) = \frac{1}{a} X(a^2 t) \sim \frac{1}{a} N(\mu a^2 t, \sigma^2 a^2 t) = N(\mu a t, \sigma^2 t)$ implies that W is a $(\mu a, \sigma)$-Brownian motion.
(b) Clearly $\mathrm{E}[X(t) - W(t)] = (1-a)\mu t$. We need adjust $W(t)$ in the following way
$W(t) = \frac{1}{a} X(a^2 t) + (1-a)\mu t$ to achieve a (μ, σ)-Brownian motion.

Chapter 3 Poisson Processes

Definition 3.1 A *counting process* $\{N(t): t \geq 0\}$ is a stochastic process where $N(0) = 0$, and every sample path consists of integer-valued, non-decreasing random variables. Here $N(t)$ represents the number of events that have occurred by time $t, t \geq 0$.

Definition 3.2 A *Poisson process*, $\{N(t): t \geq 0\}$, is a counting process with stationary independent increments where the r.v. $N(t)$ is Poisson distributed with mean λt for all $t \geq 0$, i.e.,
$\Pr[N(t) = n] = \frac{1}{n!} e^{-\lambda t} (\lambda t)^n$, for $n = 0, 1, 2, \ldots$.

Comment 3.1 λ is called the rate of the process. We claim that the "stationary" condition in the definition is redundant. Proof?
Using m.g.f. approach, we can prove that the sum (or difference) of two independent Poisson r.v.s. is a Poisson r.v. with parameter that equals to the sum (or difference) of the two original parameters. We also have $N(t+s) = [N(t+s) - N(t)] + [N(t) - N(0)]$ for the counting process. If the increments are independent, then we have $\psi_{N(t+s)}(u) = \psi_{N(t+s)-N(t)}(u) \cdot \psi_{N(t)-N(0)}(u)$, where $\psi(\cdot)$ is the moment generating function. Hence we have

$$\psi_{N(t+s)-N(t)}(u) = \frac{\psi_{N(t+s)}(u)}{\psi_{N(t)}(u)} = \frac{e^{\lambda(t+s)(e^u - 1)}}{e^{\lambda(t)(e^u - 1)}} = e^{\lambda s(e^u - 1)},$$

which doesn't depend upon t. Therefore, the increments of Poisson process are stationary. QED

Definition 3.3 The function f is said to be $o(t)$, read as "of order t", if $\lim_{t \to 0} \mathrm{f}(t)/t = 0$.

Comment 3.2 The function f is of order t, or $\mathrm{f}(t) = o(t)$, meaning that $\mathrm{f}(t)$ goes to zero faster than t does. For example, $t^2 = o(t)$, but $t \neq o(t)$.

Theorem 3.1 A counting process is a Poisson process if and only if it satisfies the following three assumptions:
(1) The process has stationary and independent increments;
(2) The probability of one event occurring in an interval of length t is $\lambda t + o(t)$, i.e., is approximately proportional to the length of that interval. The proportionality constant λ is constant over time (stationary) and may be considered as the rate at which events occur.
(3) The probability of more than one event occurring in an interval of length t is $o(t)$.
Proof: (a) The definition of a Poisson process implies the theorem above. (Note that we need expand $e^{-\lambda t}$ around $t = 0$ to reach the goal.)
(b) A rigorous proof for the other direction is given in the book. Here we provide an intuitive version. Let's split the interval of length t into n pieces so that $\Delta t = \frac{t}{n}$. Let $X_i^n = 1$ if one event occurs in an interval $[(i-1)\Delta t, i\Delta t)$ and $X_i^n = 0$ otherwise, where $i = 1, 2, \ldots, n$. It is true that $N(t) \geq \sum_{i=1}^n X_i^n$ since

maybe more than two events occur in each time interval Δt and an event may occur at time t, the last boundary point.

The property (3) of the theorem implies that $N(t) \approx \sum_i^n X_i^n$. Then the property of independent increments implies that X_i^n are independent for $i = 1,...,n$. Apparently, X_i^n is distributed as a Bernoulli r.v. The property (2) of the theorem implies that $\Pr(X_i^n = 1) \cong \lambda\Delta t = \lambda \cdot \frac{t}{n}$. Hence $N(t) = \sum_{i=1}^n X_i^n$ is distributed as a Binomial with parameters $(n, \frac{\lambda t}{n})$ since X_i^n are independent Bernoulli r.v.s. Let $n \to \infty$, we have $\text{Bin}(n, \frac{\lambda t}{n}) \to \text{Poisson}(\lambda t)$. Why? The moment generating function of $\text{Bin}(n, \frac{\lambda t}{n})$ is $(\frac{\lambda t}{n} e^u + 1 - \frac{\lambda t}{n})^n = [1 + \frac{\lambda t}{n}(e^u - 1)]^n \to e^{\lambda t(e^u - 1)}$ as $n \to 0$, because we have $\lim_{n\to 0}(1 + \frac{\alpha}{n})^n = e^\alpha$.

Theorem 3.2 Let X be a non-negative continuous r.v. Then X has the memoryless property if and only if X is exponentially distributed, i.e., $\Pr(X > t + s \mid X > s) = \Pr(X > t)$ for all $s, t \geq 0$ if and only if $\Pr(X \leq t) = 1 - e^{-\lambda t}$ for some $\lambda > 0$.

Proof: Here we discuss the result that an exponential distribution has a memoryless property.

$$\Pr(X > t + s \mid X > s) = \frac{\Pr(X > t + s, X > s)}{\Pr(X > s)} = \frac{\Pr(X > t + s)}{\Pr(X > s)}$$
$$= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = \Pr(X > t). \text{ QED}$$

(The proof for the other direction is on page 37 of the text.)

Theorem 3.3 Let $\{N(t) : t \geq 0\}$ be a counting process. Let X_n be the time between the $(n-1)^{th}$ and n^{th} event (n^{th} inter-event time) for $n = 1, 2, \ldots$. Then $X_i, i = 1, 2, \ldots$ are i.i.d. exponential r.v.s. with parameter λ and mean $\frac{1}{\lambda}$ if and only if $\{N(t) : t \geq 0\}$ is a Poisson process with rate λ.

Proof: We prove that a Poisson process $\{N(t) : t \geq 0\}$ implies exponential $X_i, i = 1, 2, \ldots$.

$\Pr(X_1 > s) = \Pr(1^{st}$ event occurs after $s) = \Pr[N(s) = 0] = e^{-\lambda s}(\lambda s)^0 = e^{-\lambda s}$, i.e., X_1 is exponential with parameter λ.

$$\begin{aligned}
\Pr(X_2 > s \mid X_1 = t) &= \Pr\{0 \text{ events in } (t, t+s] \mid X_1 = t\} \\
&= \Pr\{0 \text{ events in } (t, t+s]\} && \langle\text{by independent increments}\rangle \\
&= \Pr\{0 \text{ events in } (0, s]\} && \langle\text{by stationary increments}\rangle \\
&= \Pr[N(s) = 0] = e^{-\lambda s}.
\end{aligned}$$

Hence $\Pr(X_2 > s) = \int_0^\infty e^{-\lambda s} \, \text{f}_{X_1}(t) dt = e^{-\lambda s}$, i.e., X_2 is exponential with parameter λ.

We can use the method of induction to prove the claim. QED

Example 3.1 For a Poisson process show, for $s < t$, that

$$\Pr[N(s) = k \mid N(t) = n] = \binom{n}{k}\left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}, k = 0, 1, \ldots, n.$$

Proof: For $s < t, k = 0, 1, \ldots, n,$ we can write

$$\Pr[N(s) = k \mid N(t) = n] = \frac{\Pr[N(s) = k, N(t) = n]}{\Pr[N(t) = n]} = \frac{\Pr[N(s) = k, N(t) - N(s) = n - k]}{\Pr[N(t) = n]}$$

$$= \frac{\Pr[N(s)=k]\cdot\Pr[N(t)-N(s)=n-k]}{\Pr[N(t)=n]} = \frac{\frac{e^{-\lambda s}(\lambda s)^k}{k!}\frac{e^{-\lambda(t-s)}[\lambda(t-s)]^{n-k}}{(n-k)!}}{\frac{e^{-\lambda t}(\lambda t)^n}{n!}} = \binom{n}{k}\left(\tfrac{s}{t}\right)^k\left(1-\tfrac{s}{t}\right)^{n-k}.\ \text{QED}$$

Example 3.2 Let $\{N(t): t \ge 0\}$ be a Poisson process with rate λ. Calculate $\mathrm{E}[N(t)\cdot N(t+s)]$.

$$\mathrm{E}[N(t)\cdot N(t+s)] = \mathrm{E}\{N(t)\cdot[N(t+s)-N(t)+N(t)]\}$$
$$= \mathrm{E}\{N(t)\cdot[N(t+s)-N(t)]\} + \mathrm{E}[N(t)]^2 = \mathrm{E}[N(t)]\cdot\mathrm{E}[N(t+s)-N(t)] + \mathrm{E}[N(t)]^2$$
$$= \mathrm{E}[N(t)]\cdot\mathrm{E}[N(s)] + \mathrm{Var}[N(t)] + \{\mathrm{E}[N(t)]\}^2 = \lambda t\cdot\lambda s + \lambda t + (\lambda t)^2 = \lambda t[1+\lambda(s+t)].$$

Theorem 3.4 The time until the n^{th} event in a Poisson process is a $\mathrm{Gamma}(n,\lambda)$ r.v. The density of a $\mathrm{Gamma}(n,\lambda)$, where n is a positive integer, is $\dfrac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}$.

Proof: We know that the time until n^{th} event is equal to $\sum_{i=1}^{n} X_i$, where X_i is defined in **Theorem 3.3**, and that they are independent $\mathrm{Exp}(\lambda)$ r.v.s. We have known from **Lemma 1.4** that the sum of n $\mathrm{Exp}(\lambda)$ is a $\mathrm{Gamma}(n,\lambda)$. QED

Lemma 3.1 Let $\{N(t): t \ge 0\}$ be a Poisson process with inter-event times $X_1, X_2, \ldots$. Conditioned on $N(t)=1$, the r.v. X_1 is uniform on $[0,t]$.

Proof: Let $0 < s < t$ and suppose the first event occurs before s. Conditioned on $N(t)=1$, we have

$$\Pr(X_1 \le s \mid N(t)=1) = \frac{\Pr(X_1 \le s, N(t)=1)}{\Pr[N(t)=1]} = \frac{\Pr[N(s)=1, N(t)-N(s)=0]}{\Pr[N(t)=1]}$$

$$= \frac{\Pr[N(s)=1, N(t-s)=0]}{\Pr[N(t)=1]} \qquad \langle\text{Stationarity of Increments}\rangle$$

$$= \frac{\Pr[N(s)=1]\cdot\Pr[N(t-s)=0]}{\Pr[N(t)=1]} \qquad \langle\text{Independence of Increments}\rangle$$

$$= \frac{\frac{e^{-\lambda s}(\lambda s)}{1!}\cdot\frac{e^{-\lambda(t-s)}[\lambda(t-s)]^0}{0!}}{\frac{e^{-\lambda t}(\lambda t)}{1!}} = \frac{s}{t}.\ \text{QED}$$

Comment 3.3 A natural extension to **Lemma 3.1** is that conditioned on $N(t)=n$, the inter-event times $X_1, \ldots X_n$ should be uniformly distributed with the restriction that $X_1 < \ldots < X_n$. A formal result follows.

Definition 3.4 Consider any n r.v.s. $Y_1, Y_2, \ldots Y_n$. The i^{th} order statistic, $Y_{(i)}$ of these r.v.s. is the i^{th} smallest among them, $i = 1, 2, \ldots, n$.

Lemma 3.2 Suppose that $Y_1, Y_2, \ldots, Y_n$ are i.i.d. continuous r.v.s. with common p.d.f. $\mathrm{f}(y)$. Then the joint p.d.f. of their order statistics $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ is given by

$$g_{Y_{(1)},Y_{(2)},\ldots,Y_{(n)}}(y_1, y_2, \ldots, y_n) = \begin{cases} n!\,\mathrm{f}(y_1)\,\mathrm{f}(y_2)\cdots\mathrm{f}(y_n) & \text{if } y_1 < y_2 < \ldots < y_n \\ 0 & \text{otherwise.} \end{cases}$$

Proof: We have $\left\{Y_{(1)},Y_{(2)},...Y_{(n)}\right\} = \{y_1, y_2,..., y_n\}$ if and only if $\left\{Y_{(1)},Y_{(2)},...Y_{(n)}\right\}$ is some permutation of $\{y_1, y_2,..., y_n\}$. Each permutation is equally likely with probability $\mathrm{f}(y_1)\mathrm{f}(y_2)\cdots\mathrm{f}(y_n)$ and there are $n!$ such permutations. So the result follows. QED

Comment 3.4 If we integrate the joint density $g(y_1, y_2,..., y_n)$, taking care of the range condition, $y_1 < y_2 < ... < y_n$, we can get the marginal density for the i^{th} smallest r.v. y_i. That is:

$$g_i(y_i) = \tfrac{n!}{(i-1)!(n-i)!}[F(y_i)]^{i-1}[1-F(y_i)]^{n-i}\,\mathrm{f}(y_i).$$

Theorem 3.5 Let $\{N(t): t \ge 0\}$ be a Poisson process and let S_n be the time to the n^{th} event, $n = 1,2,...$, then given that $N(t) = n$, the n event times $S_1, S_2,...S_n$ have the same distribution as the order statistics corresponding to n independent r.v.s. uniformly distributed on the interval $[0,t]$.

Proof: Suppose $0 < t_1 < t_2 < ... < t_n < t_{n+1} = t$. Define $h > 0$ such that all n events fall into n intervals with length h, i.e., $(t_1, t_1 + h], (t_2, t_2 + h],..., (t_n, t_n + h]$. (Of course we can let the event occurring intervals be centered around $t_1,....t_n$, but it will only complicate our procedure but not altering the result.) To ensure that these time intervals won't overlap, we impose $t_i + h < t_{i+1}, \forall i$.

Using the definition of p.d.f. $\mathrm{f}(x)$ of a continuous r.v. X, we have

$\mathrm{f}(x) = F'(x) = \lim_{h\to 0} \frac{1}{h}[F(x+h) - F(x)] = \lim_{h\to 0} \frac{1}{h}\Pr[X \in (x, x+h)]$.

Here, we have

$$\begin{aligned}
\mathrm{f}_{S_1,S_2,...,S_n|N(t)=n}(t_1,...,t_n) &= \lim_{h\to 0} \tfrac{1}{h^n}\Pr\{S_1 \in (t_1,t_1+h], S_2 \in (t_2,t_2+h],...S_n \in (t_n,t_n+h] \mid N(t)=n\} \\
&= \lim_{h\to 0} \frac{\Pr\{S_1 \in (t_1,t_1+h], S_2 \in (t_2,t_2+h],...,S_n \in (t_n,t_n+h], N(t)=n\}}{h^n \cdot \Pr[N(t)=n]} \\
&= \lim_{h\to 0} \frac{\Pr\{\text{one event in } (t_1,t_1+h],...,\text{one event in } (t_n,t_n+h], \text{ zero event elsewhere}\}}{h^n \cdot \Pr[N(t)=n]} \\
&= \lim_{h\to 0} \frac{\Pr\{\text{one event in } (t_1,t_1+h]\}\cdots\Pr\{\text{one event in } (t_n,t_n+h]\}\cdot\Pr\{\text{zero event elsewhere}\}}{h^n \cdot \Pr[N(t)=n]} \\
&= \lim_{h\to 0} \frac{\frac{e^{-\lambda h}(\lambda h)}{1!}\cdots\frac{e^{-\lambda h}(\lambda h)}{1!}\cdot\frac{e^{-\lambda(t-nh)}(t-nh)^0}{0!}}{h^n \cdot \frac{e^{-\lambda t}(\lambda t)^h}{n!}} \\
&= \lim_{h\to 0} \frac{\left(\frac{h}{t}\right)^n \cdot n!}{h^n} = \tfrac{n!}{t^n}.
\end{aligned}$$

It is apparent that the density of $U(0,t)$ is $\frac{1}{t}$, and the order statistic of $U(0,t)$ is $\frac{n!}{t^n}$, by **Lemma 3.2**. Therefore $S_1, S_2,...S_n$ have the same distribution as the order statistics corresponding to n independent r.v.s. uniformly distributed on the interval $[0,t]$. QED

Theorem 3.6 *Decomposition of Poisson Processes* Suppose that events are being generated by a Poisson process of rate λ. Whenever an event occurs it is assigned to one of k streams with the i^{th} stream $(1 \le i \le k)$ being chosen with probability $p_i(\sum_{i=1}^{k} p_i = 1)$, independently of any previous assignments.

Then the i^{th} stream is itself a Poisson process of rate λp_i. Furthermore, the streams are independent of each other.

Proof for the case of two sub-streams: Let $\{N(t): t \geq 0\}$ be a Poisson process with rate λ. Let $\{N_1(t): t \geq 0\}$ and $\{N_2(t): t \geq 0\}$ be the two sub-streams of the original Poisson process and the probabilities are p and $1-p$. We need to show that both sub-streams are Poisson process.

(a) Independent increments of sub-processes follow immediately from the independent increments of the original process, given that p is constant.

(b) p.d.f. of the sub-streams:

$$
\begin{aligned}
&\Pr[N_1(t) = n, N_2(t) = m] \\
&\quad = \sum_{k=0}^{\infty} \Pr\{N_1(t) = n, N_2(t) = m \mid N(t) = k\} \cdot \Pr[N(t) = k] \\
&\quad = \Pr\{N_1(t) = n, N_2(t) = m \mid N(t) = n+m\} \cdot \Pr[N(t) = n+m] \\
&\quad = \binom{n+m}{n} p^n (1-p)^m \frac{e^{-\lambda t}(\lambda t)^{n+m}}{(n+m)!} \qquad \text{<See explanation below>} \\
&\quad = \frac{e^{-\lambda p t}(\lambda p t)^n}{n!} \cdot \frac{e^{-\lambda(1-p)t}[\lambda(1-p)t]^m}{m!} \\
&\quad = \text{Poisson}(\lambda p t) \cdot \text{Poisson}[\lambda(1-p)t] \\
&\quad = \Pr[N_1(t) = n] \cdot \Pr[N_2(t) = m] \qquad \text{<Independence of } N_1(t) \text{ and } N_2(t)\text{>}
\end{aligned}
$$

Given that there are $n+m$ events occurring between $[0,t]$ in the original Poisson process, the probability of getting n numbers of outcome 1 and m numbers of outcome 2 is the density of a Binomial distribution. We make the rough conclusion here that both sub-streams are Poisson processes, with parameter λp and $\lambda(1-p)$, respectively. Strictly speaking, we should derive the marginal density from the joint density and confirm that the each of the marginal density is indeed the p.d.f. of Poisson r.v. We also skipped the proof for the independence of two sub-streams, which is kind of cumbersome, without adding new insights.

For the case of more sub-streams, we use the same technique except that we'll consider a multi-nomial distribution instead.

Example 3.3 Customers arrive at a restaurant according to a Poisson process with rate 60 per hour. Each customer is either male or female with equal probability, independent of everything else. Suppose that at a certain hour 50 men arrived at the restaurant. What is the expected number of women arrivals?

Because the two sub-streams (male or female customers) are independent from each other, the female customers' arrival follows a Poisson process with parameter $60 \cdot \frac{1}{2} = 30$ per hour. Therefore, the expected female customer arrivals will be 30, despite the fact that there are substantially more male customers arrived an hour ago.

Example 3.4 Cars pass a point on the highway at a Poisson rate of one per minute. If 2% of the cars on the road are Mercury, then (a) what is the probability that at least two Mercury pass by during an hour?

(b) given that five Mercury have passed by in a given hour, what is the expected number of cars to have passed by in that same hour? (c) if 25 cars have passed by in an hour, what is the probability that 2 of them were Mercury?

Let $N(t)$ be the number of cars through time t (in minutes), then $\{N(t) : t \geq 0\}$ is a Poisson process with rate one per minute. Let $A = 1$ if the car passing by is a Mercury and $A = 0$ otherwise. We also know that $\Pr(A = 1) = 0.02$. Assume that the event $A = 1$ is independent of the event $A = 0$.
(a) Let $N_1(t)$ be the number of Mercury through time t, then $N_1(t) \sim \text{Poisson}(0.02t)$ and $\Pr[N_1(60) \geq 2] = 1 - \Pr[N_1(60) = 0] - \Pr[N_1(60) = 1] = 0.3374$.
(b) Let $N_2(t)$ be the number of non-Mercury through time t, then $N_2(t) \sim \text{Poisson}(0.98t)$. And thus we have $\text{E}[N(60) \mid N_1(60) = 5] = \text{E}[N_1(60) + N_2(60) \mid N_1(60) = 5] = \text{E}(5 + 0.98 \times 60) = 63.8 \approx 64$.
(c) $\Pr[N_1(60) = 2 \mid N(60) = 25] = \binom{25}{2}(0.02)^2(0.98)^{23} = 0.0756$.

Lemma 3.3 Suppose that events are being generated by a Poisson process of rate λ. Whenever an event occurs at time s, it is assigned to one of 2 streams with the first stream being chosen with probability $p(s)$, independent of any previous assignments. Let $N_i(t)$ be the number of events classified type i through time t for $i = 1,2$. Then $N_1(t)$ and $N_2(t)$ are independent Poisson r.v.s. with mean λpt and $\lambda(1-p)t$, respectively, where $p = \frac{1}{t}\int_0^t p(s)ds$.
Proof: Conditioned on $N(t) = n$, let's consider one of the unordered events. Let S be the time of occurrence for this particular event, conditioned on $N(t) = n$. Let A be either 1 or 2, depending on which stream the event is assigned to. We know that the arrival time S has a uniform $[0,t]$ distribution, and we also have

$$\Pr(A = 1) = \text{E}[\Pr(A = 1 \mid S)] = \int_0^t \Pr(A = 1 \mid S = s)\tfrac{1}{t}ds = \tfrac{1}{t}\int_0^t p(s)ds \equiv p.$$

Hence, we have

$$\begin{aligned}\Pr[N_1(t) = n, N_2(t) = m] &= \Pr[N_1(t) = n, N_2(t) = m \mid N(t) = n+m] \cdot \Pr[N(t) = n+m] \\ &= \tbinom{n+m}{n} p^n(1-p)^m \tfrac{e^{-\lambda t}(\lambda t)^{n+m}}{(n+m)!} = \text{Poisson}(\lambda pt) \cdot \text{Poisson}[\lambda(1-p)t]. \text{ QED}\end{aligned}$$

Example 3.5 Suppose that calls arrived at an phone exchange center according to a Poisson process with rate λ per minute. We are interested in the calls made between 11:00pm and 11:10pm in the example, and let's assume that at time $11:00\text{pm} + t$, a fraction, $0.1 - \frac{t}{100}$, of the calls is for Peter, where $0 \leq t \leq 10$. What's the distribution of the calls for Peter during this particular "crazy" ten minutes?

Clearly, the calls for Peter during this ten minutes period follow a Poisson process with parameter $\lambda p \cdot 10$, where p is defined as $p = \frac{1}{t}\int_0^t p(s)ds = \frac{1}{10}\int_0^{10}(0.1 - \frac{s}{100})ds = 0.05$. Hence the distribution of the calls during this ten minute period is a $\text{Poisson}(\frac{\lambda}{2})$.

Theorem 3.7 If $X_1, X_2, ..., X_n$ are independent exponential r.v.s. with parameters $\lambda_1, \lambda_2, ..., \lambda_n$, respectively, then $M = \min(X_1, X_2, ..., X_n)$ is exponential with parameter $\lambda = \lambda_1 + ... + \lambda_n$. We also have $\Pr(M = X_j)$, i.e., the probability that X_j is the minimum of the X_i's, equals $\frac{\lambda_j}{\lambda}, 1 \le j \le n$.

Proof: (a) To prove $\Pr(\min_{i=1}^n X_i > x) = e^{-\lambda x}$, we have

$$\Pr(M > x) = \Pr(X_i > x, \forall i) = \prod_i \Pr(X_i > x) = \prod_i e^{-\lambda_i x} = e^{-x \sum \lambda_i} = e^{-\lambda x}.$$

(b) To prove $\Pr(X_j < X_i, \forall i \ne j) = \frac{\lambda_j}{\lambda}$, we have

$$\Pr(M = X_j) = \Pr(X_j < X_i, \forall i \ne j) = \mathrm{E}[\Pr(X_j < X_i, \forall i \ne j \mid X_j)].$$

$$\Pr(X_j < X_i, \forall i \ne j \mid X_j = x) = \prod_{i \ne j} \Pr(X_i > x) = \prod_{i \ne j} e^{-\lambda_j x} = e^{-(\lambda - \lambda_j)x}.$$

$$\mathrm{E}[\Pr(X_j < X_i, \forall i \ne j \mid X_j)] = \int_0^\infty e^{-(\lambda - \lambda_j)x} \cdot \lambda_j e^{-\lambda_j x} dx = \frac{\lambda_j}{\lambda}. \text{ QED}$$

Theorem 3.8 *Recomposition of Poisson Processes* Suppose we have k streams of events each generated independently according to a Poisson process with the i^{th} stream $(1 \le i \le k)$ having rate λ_i. Whenever an event in one of these streams occurs it is assigned to a new combined stream. This new stream is itself a Poisson process with events being generated at rate $\lambda = \sum_{i=1}^k \lambda_i$. The probability of an event in the combined stream coming from stream i is $\frac{\lambda_i}{\lambda}, 1 \le i \le k$.

Proof: The independent increments of $N_i(t), \forall i$ imply the independent increments of the combined stream $N(t)$. To show that $N(t) \sim \text{Poisson}(\lambda t)$, let $\psi_N(u)$ be the m.g.f. of $N(t)$. Hence, we have

$$\begin{aligned}
\psi_{N(t)}(u) &= \mathrm{E}\{\exp[uN(t)]\} = \mathrm{E}\{\exp[u\textstyle\sum_{i=1}^k N_i(t)]\} \\
&= \textstyle\prod_{i=1}^k \mathrm{E}\{\exp[uN_i(t)]\} && \langle\text{Independence of } N_i(t), \forall i\rangle \\
&= \textstyle\prod_{i=1}^k \exp[\lambda_i t(e^u - 1)] && \langle\text{m.g.f. of } N_i(t), \forall i\rangle \\
&= \exp[t(e^u - 1)\textstyle\sum_i^k \lambda_i] \\
&= \exp[t(e^u - 1)\lambda] \\
&= \psi_{\text{Poisson}(\lambda t)}(u).
\end{aligned}$$

The part about probability $\frac{\lambda_i}{\lambda}$ follows immediately from **Theorem 3.7**.

Definition 3.5 A stochastic process $\{X(t) : t \ge 0\}$ is said to be a *compounded Poisson process* if and only if it can be represented as $X(t) = \sum_{i=1}^{N(t)} Y_i$, where $N(t)$ is a Poisson process and $Y_i, i = 1, 2, ...$ are i.i.d. r.v.s. which are all independent of $N(t)$.

Lemma 3.4 If X is a compound Poisson process and Y_i has characteristic function $\phi_Y(u)$ then $X(t)$ has characteristic function $\phi_{X(t)}(u) = \exp\{\lambda t[\phi_Y(u) - 1]\}$, where λ is the rate of the underlying Poisson process. Furthermore, $\mathrm{E}[X(t)] = \lambda t \,\mathrm{E}(Y_1)$ and $\mathrm{Var}[X(t)] = \lambda t \,\mathrm{E}(Y_1^2)$.

Proof:

$$\phi_{X(t)}(u) = \mathrm{E}\{\exp[iuX(t)]\} = \mathrm{E}\{\exp[iu\textstyle\sum_{j=1}^{N(t)} Y_j]\} = \mathrm{E}\Big[\mathrm{E}\{\exp[iu\textstyle\sum_{j=1}^{N(t)} Y_j] \mid N(t)\}\Big].$$

We also have

$$\begin{aligned}
&\mathrm{E}\{\exp[iu\textstyle\sum_{j=1}^{N(t)} Y_j] \mid N(t)=n\} \\
&= \mathrm{E}\{\exp[iu\textstyle\sum_{j=1}^{n} Y_j]\} && \text{<Independence of } Y_j \text{ and } N(t)\text{>} \\
&= \textstyle\prod_{j=1}^{n} \mathrm{E}[\exp(iuY_j)] && \text{<Independence of } Y_j, \forall j\text{>} \\
&= [\phi_Y(u)]^n,
\end{aligned}$$

and thus

$$\begin{aligned}
\phi_{X(t)}(u) &= \mathrm{E}\{[\phi_Y(u)]^{N(t)}\} = \textstyle\sum_{n=0}^{\infty} [\phi_Y(u)]^n \frac{e^{-\lambda t}(\lambda t)^n}{n!} = e^{-\lambda t} e^{\lambda t \phi_Y(u)} \qquad <\textstyle\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x> \\
&= \exp\{\lambda t[\phi_Y(u)-1]\}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\phi'_{X(t)}(u) &= e^{\lambda t[\phi_Y(u)-1]} \lambda t \phi'_Y(u), \text{ and} \\
\phi''_{X(t)}(u) &= e^{\lambda t[\phi_Y(u)-1]} \lambda t \phi''_Y(u) + e^{\lambda t[\phi_Y(u)-1]} \lambda t \phi'_Y(u) \lambda t \phi'_Y(u) \\
&= \phi'_{X(t)}(u) \tfrac{\phi''_Y(u)}{\phi'_Y(u)} + \phi'_{X(t)}(u) \lambda t \phi'_Y(u).
\end{aligned}$$

If we evaluate both of them at $u = 0$, we get the first and second moments of $X(t)$ as follows:

$$\begin{aligned}
\phi'_{X(t)}(0) &= e^{\lambda t(1-1)} \lambda t\, \mathrm{E}(Y_1) \qquad <\phi_Y(0) = \mathrm{E}(e^0) = 1> \\
\phi''_{X(t)}(0) &= \lambda t\, \mathrm{E}(Y_1) \tfrac{\mathrm{E}(Y_1^2)}{\mathrm{E}(Y_1)} + \lambda t\, \mathrm{E}(Y_1) \lambda t\, \mathrm{E}(Y_1) \\
&= \lambda t\, \mathrm{E}(Y_1^2) + (\lambda t)^2 [\mathrm{E}(Y_1)]^2
\end{aligned}$$

and thus

$$\mathrm{E}[X(t)] = \lambda t\, \mathrm{E}(Y_1) \text{ and } \mathrm{Var}[X(t)] = \phi''_{X(t)}(0) - [\phi'_{X(t)}(0)]^2 = \lambda t\, \mathrm{E}(Y_1^2). \text{ QED}$$

Comment 3.5 The compound Poisson process is widely used in the insurance industry. Let Y_i be the insurance claim, we can use the *Markov's Inequality* or *Chernoff Bound* to get the tail probabilities.

Example 3.6 An insurance company pays out claims on its rental insurance policies in accordance with a Poisson process having rate $\lambda = 6$ per week. If the amount of money paid on each policy is exponentially distributed with mean \$5,000, what is the mean and variance of the amount of money paid by the insurance company in a four week span? Give a bound on the probability that the money paid out on policies in a four week span is more than \$200,000.

Let $N(t)$ be the number of claims, then we have $N(t) \sim \text{Poisson}(6t)$. Let Y_i be the amount of money paid on each cliam, and $Y_i \sim \exp(\frac{1}{5000})$. When $t = 4$, we have

$$\mathrm{E}\left[\textstyle\sum_{i=1}^{N(t)} Y_i\right] = \mathrm{E}[N(t)] \cdot \mathrm{E}(Y_i) = 6 \times 4 \times 5000 = \$120,000,$$

and

$$\begin{aligned}
\mathrm{Var}\left[\textstyle\sum_{i=1}^{N(t)} Y_i\right] &= \mathrm{E}[N(t)] \cdot \mathrm{E}(Y_i^2) = \mathrm{E}[N(t)] \cdot \{\mathrm{Var}(Y_i) + [\mathrm{E}(Y_i)]^2\} \\
&= 6 \times 4 \times (5000^2 + 5000^2) = \$1,200,000,000.
\end{aligned}$$

By Markov's Inequality we have

$$\Pr\left[\textstyle\sum_{i=1}^{N(t)} Y_i \geq 200000\right] \leq \tfrac{1}{200000} \cdot \mathrm{E}\left[\textstyle\sum_{i=1}^{N(t)} Y_i\right] = 0.6.$$

Definition 3.6 A *non-homogenous* Poisson process, $\{N(t): t \geq 0\}$, is a counting process with independent increments where the r.v. $N(t)$ is Poisson distributed with mean $m(t)$, for all $t \geq 0$, i.e., $\Pr[N(t) = n] = \frac{e^{-m(t)}[m(t)]^n}{n!}$, for $n = 0,1,2,....$ The function $m(t)$ is called the mean value function of the process and is non-decreasing with $m(0) = 0$. Furthermore, we can write $m(t) = \int_0^t \lambda(s)ds$, where $\lambda(s) \geq 0$ is called the rate of the process at time s. If $\lambda(\cdot)$ is continuous at t, then $\frac{dm(t)}{dt} = \lambda(t)$.

Comment 3.6 In regular Poisson process, we have flat rate which implies stationary increments. But it doesn't make sense for some problems with peak-time and off-peak-time. If $\{N(t): t \geq 0\}$ is a non-homogenous Poisson process, then we have $N(t+s) - N(t) \sim \text{Poisson}[m(t+s) - m(t)]$. Note that $\text{Poisson}[m(t+s) - m(t)] \neq \text{Poisson}[m(s)]$, which implies that a non-homogenous Poisson process doesn't have stationary increments. Following the same argument using m.g.f. approach in **Comment 3.1**, we can prove this result.

Theorem 3.9 A counting process is a non-homogenous Poisson process if and only if its satisfies the following three assumptions:
(1) The process has independent increments with $N(0) = 0$;
(2) $\Pr[N(t+h) - N(t) = 1] = \lambda(t)h + o(h)$;
(3) $\Pr[N(t+h) - N(t) \geq 2] = o(h)$.
Proof: We can use the similar technique used in proving **Theorem 3.1**. Skipped here.

Theorem 3.10 Let $\{N(t): t \geq 0\}$ be a homogenous Poisson process with rate λ^*, and let $\{\lambda(t): t \geq 0\}$ be such that $0 \leq \lambda(t) \leq \lambda^*$. If an event in N occurring at time t is accepted with probability $\frac{\lambda(t)}{\lambda^*}$ independent of everything else, then the process of accepted events is a non-homogenous Poisson process with rate function $\{\lambda(t): t \geq 0\}$.
Proof: Denote as $M(t)$ the number of events accepted, which is often called the "thinned" process. We need prove that $\{M(t): t \geq 0\}$ is a non-homogenous Poisson process with rate function $\lambda(t)$.
(1) $\{M(t): t \geq 0\}$ has independent increments, following from independent increments of the original process $N(t)$. We also have $M(0) = 0$ since $N(0) = 0$.
(2) Let $A = 1$ stand for the case where the event is accepted.

$$\begin{aligned} &\Pr[M(t+h) - M(t) = 1] = \Pr[N(t+h) - N(t) = 1, A = 1] \\ &= \Pr[N(t+h) - N(t) = 1] \cdot \Pr(A = 1) && \text{<By independence of } A \text{ and } N(t)\text{>} \\ &= [\lambda^* h + o(h)]\tfrac{\lambda(t)}{\lambda^*} && \text{<By \textbf{Theorem 3.1.(2)}>} \\ &= \lambda(t)h + o(h) \end{aligned}$$

(3) Using the same condition, $A = 1$, we have

$$\Pr[M(t+h) - M(t) \geq 2] \leq \Pr[N(t+h) - N(t) \geq 2] = o(h) \quad \text{<By \textbf{Theorem 3.1.(3)}>}$$

Therefore, we have $\{M(t): t \geq 0\}$ as a non-homogenous Poisson process with rate function $\lambda(t)$ by **Theorem 3.9**. QED

Comment 3.7 Theorem 3.10 is very useful in practice in that it helps us generate non-homogenous processes from homogenous ones. We can also use the idea of thinned process to patchwork the crowd scenes when making films.

Example 3.7 Let $T_1, T_2, \ldots$ denote the inter-arrival times of events of a non-homogenous Poisson process having intensity function $\lambda(t)$. Find out the distribution of T_1 and T_2. Determine whether T_i are i.i.d.
Let $m(t) = \int_0^t \lambda(v)dv$. $\Pr(T_1 > s) = \Pr[N(s) = 0] = \frac{e^{-m(s)}[m(s)]^0}{0!} = e^{-m(s)}$. Hence $F_{T_1}(s) = 1 - e^{-m(s)}$ and $\mathrm{f}_{T_1}(s) = e^{-m(s)} \cdot \lambda(s)$.
We also have

$$\begin{aligned}\Pr(T_2 > s \mid T_1 = t) &= \Pr[N(t+s) - N(t) = 0 \mid N(t) = 1] \\ &= \Pr[N(t+s) - N(t) = 0] \qquad \langle\text{Independent Increments}\rangle \\ &= e^{-[m(t+s)-m(t)]} \qquad \langle N(t+s) - N(t) \sim \text{Poisson}[m(t+s) - m(t)]\rangle\end{aligned}$$

Therefore, we can write

$$\Pr(T_2 > s) = \int_0^\infty e^{-[m(t+s)-m(t)]}\, \mathrm{f}_{T_1}(t)dt = \int_0^\infty e^{-[m(t+s)-m(t)]} e^{-m(t)}\lambda(t)dt = \int_0^\infty e^{-m(t+s)}\lambda(t)dt,$$

which implies that

$$F_{T_2}(s) = \int_{-\infty}^0 e^{-m(t+s)}\lambda(t)dt.$$

It is now easy to reach the conclusion that T_i are not i.i.d.

Theorem 3.11 Given a non-homogenous Poisson process, $\{N(t) : t \geq 0\}$, let's set $N'(t') = N(m^{-1}(t'))$ for all $t' = m(t)$ and $t' \geq 0$. Then $\{N'(t') : t' \geq 0\}$ is a homogenous Poisson process with rate $\lambda' = 1$. (Note that the prime ' stands for prime time, t', not first order derivative.)
Proof: It is clear that $N'(t')$ has independent increments. We also have $N'(0) = N[m^{-1}(0)] = N(0) = 0$. Since

$$\begin{aligned}\Pr[N'(t') = n] &= \Pr[N[m^{-1}(t')] = n] = \Pr[N(t) = n, \text{ where } t = m^{-1}(t')] \\ &= \{\tfrac{e^{-m(t)}[m(t)]^n}{n!}, \text{ where } t = m^{-1}(t')\} = \tfrac{1}{n!} e^{-m(m^{-1}(t'))}[m(m^{-1}(t'))]^n = \tfrac{1}{n!} e^{-t'}(t')^n,\end{aligned}$$

we know $N'(t') \sim \text{Poisson}(t')$, i.e., $\{N'(t') : t' \geq 0\}$ is a homogenous Poisson process with rate 1. QED

Comment 3.8 This theorem is very useful when proving stuff. The intuition behind the theorem is that a process with slow rate $\lambda(t)$ in certain portion of the regular time would be mirrored, through the mean value function $m(t)$, into a process with much faster rate in the corresponding portion of the prime time, and vice versa. Hence a non-homogenous Poisson process with varying rates will be smoothed out into a homogenous Poisson process with rate 1.

Example 3.8 Let $\{N(t) : t \geq 0\}$ be a non-homogenous Poisson process with mean value function $m(t)$. Given $N(t) = n$, show that the unordered set of arrival times has the same distribution as n i.i.d. r.v.s. with c.d.f. being $F(x) = \frac{m(x)}{m(t)}, \forall x \leq t$ and $F(x) = 1, \forall x > t$.

$\{N(t): t \geq 0\}$ is a non-homogenous Poisson process with mean value function $m(t)$. Let's set up the prime-time t' as $m^{-1}(t') = t$. Then we have $\{N'(t') = N[m^{-1}(t')] : t' \geq 0\}$ as a homogenous Poisson process with $\lambda' = 1$. Given $N(t) = n$ or $N'[m^{-1}(t')] = n$, we know that the unordered set of arrival times $S_i', i = 1,2,...,n$ in prime-time are i.i.d. uniform over $[0, t']$. That is

$$\Pr(S_i' \leq s') = \begin{cases} \frac{s'}{t'} & s' \leq t' \\ 1 & s' > t'. \end{cases}$$

The corresponding arrival times in the regular time $S_i, i = 1,2,...,n$ are $S_i = m^{-1}(S_i')$, hence we can write

$$\Pr(S_i' \leq s') = \Pr[m^{-1}(S_i') \leq m^{-1}(s')] \qquad \text{<nondecreasing } m(\cdot) \text{ and } m^{-1}(\cdot)\text{>}$$
$$= \Pr(S_i \leq s).$$

Finally, we have

$$\Pr(S_i \leq s) = \begin{cases} \frac{s'}{t'} = \frac{m(s)}{m(t)} & s \leq t \\ 1 & s > t. \text{ QED} \end{cases}$$

CHAPTER 4 INTRODUCTION TO QUEUEING THEORY

Definition 4.1 A *queueing system* consists of arrival streams of customers and a series of servers. If there is an infinite amount of space for waiting customers, then when there are more customers than available servers, the remaining customers wait in queue. If there is only a finite amount of space for waiting customers, then the customers arriving to a full system depart without receiving service, and we speak of a *loss system*.

Single station queues are often described by the notation $A/B/s/K$ (originally due to Kendall), where A stands for the arrival distribution ($D =$ deterministic, $M =$ exponential(memoryless), $\mathrm{E}_k =$ Erlang-k, and $G =$ General) and B stands for the service distribution ($D =$ deterministic, $M =$ exponential(memoryless), $\mathrm{E}_k =$ Erlang-k, and $G =$ General). The interarrival times and service times are assumed to form i.i.d. sequences that are independent of each other. The number of servers in parallel is s and K is the number of customers the system can hold. If K is not given then it is assumed to be infinite. Unless otherwise stated, service order is assumed to be First-In-First-Out (FIFO), also known as First-Come-First-Serve (FCFS). Other common service disciplines are Last-Come-First-Serve (LCFS) and Shortest Expected Processing Time (SEPT).

Theorem 4.1 Consider an $M/G/\infty$ queue with arrival rate λ. Service times have distribution function G and mean $\frac{1}{\mu} < \infty$. Let $X(t)$ be the number of customers in the system at time t. Let $p_t = \frac{1}{t}\int_0^t [1-G(s)]ds$. Then $X(t)$ is distributed as a Poisson r.v. with mean $\lambda t p_t$. The number of customers that have left the system after receiving service by time t is independent of $X(t)$ and is distributed as a Poisson r.v. with mean $\lambda t(1-p_t)$. The limiting distribution of $X(t)$ as $t \to \infty$ is $\text{Poisson}(\frac{\lambda}{\mu})$.

Proof: Suppose that a customer arrives at time s, where $s < t$. Let $A = 1$ stands for the event that this customer is still in system at time t. Let T be the service time with distribution G. Then we have

$$p(s) = \Pr[A = 1 \mid \text{arrives at s}] = \Pr(T > t-s) = 1 - G(t-s).$$

By **Lemma 3.3**, we know that $X(t)$ is Poisson r.v. with mean $\lambda t p_t$, where

$p_t = \frac{1}{t}\int_0^t [1-G(t-s)]ds = \frac{1}{t}\int_0^t [1-G(u)]du,$ with $u = t-s.$ Hence $\mu(t) \equiv \lambda t p_t = \lambda \int_0^t [1-G(u)]du$.

If we take advantage of the result in **Comment 1.6**, we also have

$$\lim_{t\to\infty} \mu(t) = \lambda \int_0^\infty [1-G(u)]du = \lambda \int_0^\infty \Pr(T > u)du = \lambda\,\mathrm{E}(T).$$

Hence, $\lim_{t\to\infty} \Pr[X(t) = n] = \lim_{t\to\infty} \frac{1}{n!} e^{-\mu(t)}[\mu(t)]^n = \frac{1}{n!} e^{-\lambda \mathrm{E}(T)}[\lambda\,\mathrm{E}(T)]^n$, i.e., $X(t) \Rightarrow \text{Poisson}[\lambda\,\mathrm{E}(T)]$.

Example 4.1 Calls arrive to an ambulance station according to a Poisson process at rate λ. The time required for an ambulance to serve a call and return to base has mean $\frac{1}{\mu}$. How many ambulances should be stationed at the base to ensure at least $r\%$ of calls can be answered immediately?

We can consider ambulances as servers in queueing theory. Assume that if a call arrives when all ambulances are busy, then an outside agent attends to the call. Let $X(t)$ be the number of calls in progress at time t. We have $X(t) \approx \text{Poisson}(\frac{\lambda}{\mu})$. Suppose that there are n ambulances stationed, then it suffices to find out n such that $\Pr[\text{Poisson}(\frac{\lambda}{\mu}) \leq n] \geq r\%$.

CHAPTER 5 RENEWAL THEORY

Comment 5.1 We have seen before that the inter-event times for the Poisson process are i.i.d. exponential r.v.s. A natural generalization is to consider a counting process for which the inter-event times are i.i.d. with an arbitrary distribution. Such a counting process is called a *renewal process*. (The formal definition of a renewal process will follow shortly.) We will use the terms "events" and "renewals" interchangeably, and so we say that the n^{th} renewal occurs at time S_n.

Since the inter-event times are i.i.d., it follows that at each event the process probabilistically starts over, or "regenerates" itself. Note that a Poisson process "regenerates" at all times because of the memoryless property. A renewal process doesn't necessarily regenerate at all times, but it does regenerate at event times. Some examples of renewal processes. (1) In reliability theory, we often consider the life time, X_i, of some component that is immediately replaced when it defects. (2) In inventory theory, we often consider a (s,S) system. When the inventory level falls to s or below, we order up to S immediately. If the time to refill the inventory, or *lead time*, is zero, and the demand process has stationary independent increments, then we get a renewal process. (3) Let's consider a $G/G/K$ system in queueing theory. Let $X(t)$ be the number of customer in the system at time t. When a new customer gets into an empty system, the system regenerates the process.

Much of renewal theory aims to give tools for computing limiting expectations and probabilities. As an example, suppose that the lifetime of a part, T, is random and has a density f. When it fails it is immediately replaced. Let $X(t)$ be the age of the part in operation at time t and assume that a new part is installed at time 0. What is $m(t)=\mathrm{E}[X(t)]$?
If we condition on T, the time of first replacement, then

$$m(t)=\mathrm{E}[X(t)]=\int_0^\infty \mathrm{E}[X(t)\mid T=s]\mathrm{f}(s)ds=\int_0^t \mathrm{E}[X(t)\mid T=s]\mathrm{f}(s)ds+\int_t^\infty t\,\mathrm{f}(s)ds$$
$$=\int_0^t \mathrm{E}[X'(t-s)]\mathrm{f}(s)ds+t\cdot\Pr(T>t)=\int_0^t m(t-s)\,\mathrm{f}(s)ds+t\cdot\Pr(T>t).$$

Note that in the derivation above we have used the knowledge that if the part defects at time s and $s<t$, then the age of the part in operation is $X'(t-s)$, where $X'(\cdot)$ is identically distributed as $X(\cdot)$, and that if $s>t$, then the part doesn't fail at time t and thus the age of the part in operation is t.

This equation in $m(t)$ is called *renewal equation*, and it can be solved analytically only in a few special cases. But we can compute $\lim_{t\to\infty} m(t)$ under very weak conditions. In particular, we will eventually prove (among other things) that if $\mathrm{E}(T^2)<\infty$, then $\lim_{t\to\infty} m(t)=\mathrm{E}(T^2)/[2\,\mathrm{E}(T)]$, not the intuitive $\frac{1}{2}\mathrm{E}(T)$. As a matter of fact, $\mathrm{E}(T^2)/[2\,\mathrm{E}(T)]=\frac{1}{2}\mathrm{E}(T)$ holds only if $\mathrm{Var}(T)=0$, that is the lifetime of the parts is deterministic. In addition, the difference between $\mathrm{E}(T^2)/[2\,\mathrm{E}(T)]$ and $\frac{1}{2}\mathrm{E}(T)$ increases with the variance of T.

Definition 5.1 Let $\{X_n : n = 1,2,...\}$ be a sequence of non-negative i.i.d. r.v.s. with c.d.f. F and $F(0) < 1$. Let $S_n = \sum_{i=1}^n X_i$ and $S_0 = 0$. Let $N(t) = \sup\{n : S_n \le t\}$, then $\{N(t) : t \ge 0\}$ is a renewal process.

Comment 5.2 Suppose that $S_4 < t < S_5$, then we have $N(t) = \sup\{n : S_n \le t\} = \sup\{1,2,3,4\} = 4$. Note that X_i can be discrete, continuous or both, but we require that $F(0) = \Pr(X = 0) < 1$, meaning that we don't want to see all events occurring simultaneously. We write "sup" instead of "max" here to avoid the situation where $\max(n)$ doesn't exist. Note further that $\mu = \mathrm{E}(X_1) \in (0,\infty]$, allowing $\mathrm{E}(X_1) = \infty$. For an example of a r.v. X with $\mathrm{E}(X) = \infty$, take X with c.d.f. $F(x) = 1 - \frac{1}{x}, x \ge 1$. We have $\mathrm{E}(X) = \int_1^\infty \Pr(X > x)dx = \int_1^\infty x^{-1}dx = \infty$. (Does $\mathrm{E}(X) = \int_1^\infty \Pr(X > x)dx$ hold for positive continuous r.v. as well? Yes, see **Comment 1.6**.)

Comment 5.3 We know that $\Pr(X_i > 0) > 0$, i.e., $F(0) < 1$, and $\mu = \mathrm{E}(X_1) \in (0,\infty]$. Suppose that we have $S_n = 1 - \frac{1}{n}, \forall n \ge 1$, then we will have $N(1) = \infty$. We don't want to have this situation happen and want to have finite $N(1)$. In general, we should have finite number of events by time t, i.e., $N(t) < \infty$. The formal result is presented in **Proposition 5.1**. (As a matter of fact, X_i's are not i.i.d. in this example and they are not random either.)

Proposition 5.1 Let $\{N(t) : t \ge 0\}$ be a renewal process, then $N(t) < \infty$ for all $t \ge 0$ almost surely and we can write $N(t) = \max\{n : S_n \le t\}$.
Proof: Let's define two sets $A = \{\omega : \frac{S_n(\omega)}{n} \to \mu \text{ as } n \to \infty\}$ and $B = \{\omega : N_\omega(t) = \infty\}$. By the *Strong Law of Large Numbers*, we know $\Pr(A) = 1$ and we want to prove that $\Pr(B) = 0$. Suppose that $\omega \in A \cap B$. $\omega \in B$ implies $S_n(\omega) \le t$ since ∞ events occurred before time t. $\omega \in A$ implies that $\frac{S_n(\omega)}{n} \to \mu$ as $n \to \infty$, which implies in turn that $\frac{S_n(\omega)}{n} > \frac{\mu}{2}$ for n large enough. That is, $S_n(\omega) > \frac{n\mu}{2}$ for n large enough, i.e., $S_n(\omega) \to +\infty$ as $n \to +\infty$ (since $\frac{\mu}{2} > 0$), which is against $S_n(\omega) \le t$. Therefore, $A \cap B = \varnothing$ and thus $1 \ge \Pr(A \cup B) = \Pr(A) + \Pr(B)$. Hence $0 \le \Pr(B) \le 1 - \Pr(A) = 0$ implies $\Pr(B) = 0$, which implies in turn that $N(t) < \infty, \forall t \ge 0, a.s.$ Furthermore, we notice that $S_n \le t$ is equivalent to $N(t) \ge n$, hence we can write $N(t) = \max\{n : S_n \le t\}$.

Propersition 5.2 Let $\{N(t) : t \ge 0\}$ be a renewal process with renewal distribution F. Let F_n be the n-fold convolution of F with itself. Then $\Pr[N(t) \le n] = 1 - F_{n+1}(t)$.
Proof:

$$\Pr[N(t) \le n] = 1 - \Pr[N(t) \ge n+1] = 1 - \Pr[S_{n+1} \le t] \equiv 1 - F_{n+1}(t).$$

We also have

$$\begin{aligned} F_n(t) &= \Pr[S_n \le t] = \Pr(S_{n-1} + X_n \le t) = F \otimes F_{n-1}(t) \qquad \text{<Since } X_n \text{ and } S_{n-1} \text{ are independent>} \\ &= F \otimes F \otimes \cdots \otimes F(t). \text{ QED} \end{aligned}$$

Definition 5.2 Let $\{N(t) : t \ge 0\}$ be a renewal process with renewal distribution F. Define $m(t) \equiv \mathrm{E}[N(t)]$, then $m(t)$ is called the *renewal function*.

Theorem 5.1 $m(t)=\sum_{n=1}^{\infty}F_n(t)$ and $m(t)<\infty, \forall 0\le t<\infty$.

Proof: $m(t)=\mathrm{E}[N(t)]=\sum_{n=1}^{\infty}\Pr[N(t)\ge n]=\sum_{n=1}^{\infty}\Pr[S_n\le t]=\sum_{n=1}^{\infty}F_n(t)$.

To show that $m(t)<\infty, \forall t\in[0,\infty)$, let's consider the following new process. Let $\alpha>0$ and $\Pr(X_n\ge\alpha)>0$. Define $X_n'=0$ for $X_n<\alpha$ and $X_n'=\alpha$ for $X_n\ge\alpha$. Let $N'(t)=\sup\{n:S_n'\le t\}$.

Clearly, the number of events at time 0 is distributed as geometric with mean $\frac{1}{\Pr(X_1\ge\alpha)}-1$. (Here we use the version $\mu=\frac{1-p}{p}$ in **Comment 1.5.(4)** because it is possible that there is zero event at time 0.) Note that the number of events at time $k\alpha, \forall k\ge 1$, is again distributed as geometric with mean $\frac{1}{\Pr(X_1\ge\alpha)}$. (Here we use ther version $\mu=\frac{1}{p}$ in **Comment 1.5.(4)** because at each time $k\alpha, \forall k\ge 1$, the number of events must be no less than 1; otherwise all events will be concertrated at time $(k-1)\alpha$, violating $\Pr(X_n\ge\alpha)>0$.) Therefore, we have $\mathrm{E}[N'(t)]\le\frac{1}{\Pr(X_1\ge\alpha)}\cdot(\lfloor\frac{t}{\alpha}\rfloor+1)$. Because of the design of our new process, we know $N(t)\le N'(t), \forall t\in[0,\infty)$, and thus $\mathrm{E}[N(t)]\le\mathrm{E}[N'(t)]$, i.e., $m(t)\le\mathrm{E}[N'(t)]\le\frac{1}{\Pr(X_1\ge\alpha)}(\lfloor\frac{t}{\alpha}\rfloor+1)$. Hence $m(t)<\infty$. QED

Example 5.1 Prove the renewal equation $m(t)=F(t)+\int_0^t m(t-x)dF(x)$.

Proof: Using the conditional expectation argument, we have $m(t)=\mathrm{E}[N(t)]=\mathrm{E}\{\mathrm{E}[N(t)\mid X_1]\}$. We also have $\{N(t)\mid X_1=x\}=1+N'(t-x)$, where $N'(\cdot)$ is identically distributed as $N(\cdot)$. Hence we write

$$\mathrm{E}[N(t)\mid X_1=x]=1+\mathrm{E}[N'(t-x)]=1+m(t-x),$$

and

$$m(t)=\int_0^t[1+m(t-x)]dF(x)+\int_t^{\infty}0\cdot dF(x)=F(t)+\int_0^t m(t-x)dF(x).$$

Theorem 5.2 The renewal function $m(t)$ uniquely determines the distribution of the renewal process, i.e., F.

Proof: Let's take the Laplace tranform of $m(t)$ as follows:

$$\begin{aligned}\tilde{m}(s)&=\int_0^{\infty}e^{-st}dm(t)=\int_0^{\infty}e^{-st}d[\textstyle\sum_{n=1}^{\infty}F_n(t)]\\&=\textstyle\sum_{n=1}^{\infty}\int_0^{\infty}e^{-st}dF_n(t) \qquad \text{<by Fubini's theorem; non-negativity of } F_n(t)\text{>}\\&=\textstyle\sum_{n=1}^{\infty}\mathrm{E}[e^{-s\cdot S_n}]=\sum_{n=1}^{\infty}\{\mathrm{E}[e^{-s\cdot X_1}]\}^n \qquad \text{<}X_i\text{'s are i.i.d.>}\\&=\textstyle\sum_{n=1}^{\infty}[\tilde{F}(s)]^n \qquad \text{<}\tilde{F}(s)\equiv\mathrm{E}(e^{-s\cdot X_1})\text{>}\\&=\frac{\tilde{F}(s)}{1-\tilde{F}(s)} \qquad \text{<}0<\tilde{F}(s)=\mathrm{E}(e^{-s\cdot X_1})<1\text{>}\end{aligned}$$

Hence we have $\tilde{F}(s)=\frac{\tilde{m}(s)}{\tilde{m}(s)+1}$, i.e., $m(t)$ uniquely determines F. QED

Example 5.2 Suppose that $m(t)=\lambda t$, where $\lambda>0$. What is F corresponding to $m(t)$?

We have $\tilde{m}(s)=\int_0^{\infty}e^{-st}dm(t)=\int_0^{\infty}e^{-st}\lambda dt=\frac{\lambda}{s}$ and thus $\tilde{F}(s)=\frac{\lambda/s}{1+\lambda/s}=\frac{\lambda}{\lambda+s}$, which is the Laplace tranform of $\exp(\lambda)$. Hence a Poisson process is the only renewal process that has a linear renewal function.

Example 5.3 In defining a renewal process, we assume that $\Pr(X_i < \infty) = F(\infty) = 1$. If $F(\infty) < 1$ then after each renewal there is a positive probability $1 - F(\infty)$ that there will be no further renewal. Argue that when $F(\infty) < 1$ the total number of renewals, $N(\infty)$, is such that $1 + N(\infty)$ has a geometric distribution with mean $\frac{1}{1-F(\infty)}$.

Proof: Clearly, we have

$$\Pr[N(\infty) = n] = \Pr[X_1 < \infty, \ldots, X_n < \infty, X_{n+1} = \infty]$$
$$= \Pr(X_1 < \infty) \cdots \Pr(X_n < \infty) \cdot \Pr(X_{n+1} = \infty) = [F(\infty)]^n \cdot [1 - F(\infty)].$$

This is the p.d.f. of a geometric distribution with mean $\frac{1}{1-F(\infty)} - 1$, and thus $1 + N(\infty)$ has a geometric distribution with mean $\frac{1}{1-F(\infty)}$.

Proposition 5.3 $\lim_{t\to\infty} N(t) = \infty$ almost surely.

Proof: Let's denote $N(\infty) = \lim_{t\to\infty} N(t)$.

$$\Pr[N(\infty) < \infty] = \Pr(X_1 < \infty, \ldots, X_{i-1} < \infty, X_i = \infty \text{ for some } i)$$
$$= \Pr\left(\bigcup_{i=1}^{\infty} X_i = \infty\right)$$
$$\leq \sum_{i=1}^{\infty} \Pr(X_i = \infty) \qquad \text{<by Boole's inequality>}$$
$$= \sum_{i=1}^{\infty} 0 = 0$$

Hence $N(\infty) = \infty$. QED

Comment 5.4 We may want to use the following argument to prove **Proposition 5.3**:

$$\Pr[N(t) < \infty] = 1 \Rightarrow \lim_{t\to\infty} \Pr[N(t) < \infty] = 1 \Rightarrow \Pr\{\lim_{t\to\infty}[N(t) < \infty]\} = 1.$$

The problem of the argument above is that we cannot switch $\lim_{t\to\infty}$ with $\Pr$ like that.

Theorem 5.3 $\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}$ almost surely.

Proof: We first translate $\frac{N(t)}{t}$ into S_n and then use the relationship $S_{N(t)} \leq t \leq S_{N(t)+1}$ to reach the goal. Note that $S_{N(t)} \leq t \leq S_{N(t)+1}$ implies $\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{N(t)+1}}{N(t)}$. Although it is possible that $N(t) = 0$ which would break the inequality above, $N(\infty) = \infty$ implies that we are okay here. If we take limits as $t \to \infty$ on the inequality, we have $\lim_{t\to\infty} \frac{S_{N(t)}}{N(t)} = \mu$ almost surely and $\lim_{t\to\infty} \frac{S_{N(t)+1}}{N(t)} = \lim_{t\to\infty}\left[\frac{S_{N(t)+1}}{N(t)+1} \frac{N(t)+1}{N(t)}\right] = \mu$ almost surely. Finally we have $\lim_{t\to\infty} \frac{t}{N(t)} = \mu$ almost surely and thus $\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}$ almost surely. QED

Comment 5.5 The result in **Theorem 5.3** implies that the average slope of the renewal process is $\frac{1}{\mu}$. Note that $\lim_{t\to\infty} \frac{N(t)}{t}$ is often called the "long-run rate" of the renewal process.

Example 5.4 Consider an $M/G/1$ queue with balking, so that if the server is busy when a customer arrives, the customer departs without receiving service, i.e., the customer is lost. What is the long-run fraction of customers that are lost?

Let λ be the arrival rate and $\frac{1}{\mu}$ be the mean service time. Assume that a customer arrives at time 0 to an empty system. Let Y_i be the i^{th} service time, then Y_i's are i.i.d. with mean $\mathrm{E}(Y_i) = \frac{1}{\mu}$. Let X_i be the i^{th} idle time immediately after the service time, then X_i's are i.i.d. $\exp(\lambda)$ with mean $\frac{1}{\lambda}$ (because of the memoryless property). Let Z_i be the length of the i^{th} cycle covering both the service time and the idle time. Then Z_i's are i.i.d. with mean $\mathrm{E}(Z_i) = \mathrm{E}(Y_i) + \mathrm{E}(X_i) = \frac{1}{\mu} + \frac{1}{\lambda}$. Let the long-run rate at which customers arrive be $\lim_{t\to\infty} \frac{N(t)}{t}$, where $\{N(t) : t \geq 0\}$ is a Poisson process. Also denote the long-run rate at which customers are served as $\lim_{t\to\infty} \frac{N_Z(t)}{t}$, where $N_Z(t)$ is the number of customers served at time t, i.e., the number of cycles Z_i. Since X_i's are i.i.d. $\exp(\lambda)$ and Y_i's are i.i.d. G, Z_i's are i.i.d. and $N_Z(t)$ is a renewal process.

(a) The long-run rate that customers enter the bank is

$$\lim_{t\to\infty} \frac{N_Z(t)}{t} = \frac{1}{\mathrm{E}(Z_i)} = \frac{1}{\mathrm{E}(X_i)+\mathrm{E}(Y_i)} = \frac{1}{1/\mu+1/\lambda} = \frac{\lambda\mu}{\lambda+\mu}.$$

(b) Fraction of customers enter the bank is

$$\lim_{t\to\infty} \frac{N_Z(t)}{N(t)} = \lim_{t\to\infty} \frac{N_Z(t)/t}{N(t)/t} = \frac{1/\mathrm{E}(Z_i)}{1/\lambda} = \frac{\lambda^2\mu}{\lambda+\mu}.$$

(c) Fraction of time the server is busy is

$$\frac{\mathrm{E}(Y_i)}{\mathrm{E}(Z_i)} = \frac{1/\mu}{\lambda\mu/(\lambda+\mu)} = \frac{\lambda+\mu}{\lambda\mu^2}.$$

Example 5.5 A part in use is replaced by a new part either when it fails or when it reaches the age of T years. If the lifetimes of successive parts are independent with a common distribution F having density f, show that: (a) the long-run rate at which parts are replaced equals $\left\{\int_0^T x\,\mathrm{f}(x)dx + T[1-F(T)]\right\}^{-1}$; (b) the long-run rate at which parts in use fail equals $F(T)\cdot\left\{\int_0^T x\,\mathrm{f}(x)dx + T[1-F(T)]\right\}^{-1}$.

Proof: Let T_i be the life time of the i^{th} part and X_i be the actual time being used of the i^{th} part, then $X_i = T_i$ if $T_i \leq T$ and $X_i = T$ otherwise.

(a) Let $N(t)$ be the number of replacement through time t and the long run rate at which parts are replaced be K_N. Then we have $K_N = \lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mathrm{E}(X_i)}$. Furthermore, we have

$$\begin{aligned}\mathrm{E}(X_i) = \mathrm{E}[\mathrm{E}(X_i \mid T_i)] &= \int_0^{+\infty} \mathrm{E}(X_i \mid T_i = u)dF(u) = \int_0^T u dF(u) + \int_T^\infty T dF(u) \\ &= \int_0^T u dF(u) + \int_0^\infty T dF(u) - \int_0^T T dF(u) = \int_0^T u dF(u) + T[1-F(T)],\end{aligned}$$

and thus

$$K_N = \left\{\int_0^T x\,\mathrm{f}(x)dx + T[1-F(T)]\right\}^{-1}.$$

(b) Let Y_j be the time between the $(j-1)^{th}$ failure and the j^{th} failure. Then $Y_j = \sum_{j=1}^N X_{ji}$, where N is the number of replacement between the $(j-1)^{th}$ failure and the j^{th} failure. Here $N \sim$ Geometric with probability $\Pr(T_i \leq T) = F(T)$ and N doesn't depend upon future events. So N is a stopping time and we can use the Wald's Equation to write $\mathrm{E}(Y_j) = \mathrm{E}(N)\cdot\mathrm{E}(X_{ji}) = \frac{1}{F(T)}\left\{\int_0^T x\,\mathrm{f}(x)dx + T[1-F(T)]\right\}$.

Let $N_F(t)$ be the number of failed parts through time t. Since Y_j's are i.i.d., we know $N_F(t)$ is a renewal process. Hence we have

$$\lim_{t\to\infty}\tfrac{N_F(t)}{t}=\tfrac{1}{\mathrm{E}(Y_j)}=F(T)\left\{\int_0^T x\,\mathrm{f}(x)dx+T[1-F(T)]\right\}^{-1},$$

which is the long-run rate at which parts in use fail.

Comment 5.6 Some limiting results are listed as follows:

	$N(t)$	$N(t)/t$	$m(t)$	$m(t)/t$
$t<\infty$	$<\infty$, a.s.	$<\infty$, a.s.	$<\infty$	$<\infty$
$\lim_{t\to\infty}$	$=\infty$, a.s.	$=\frac{1}{\mu}$, a.s.	$=\infty$, a.s.	$=\frac{1}{\mu}$, a.s.

Proof for $\lim_{t\to\infty} m(t)=\infty$:

$$\begin{aligned}\lim_{t\to\infty} m(t)&=\lim_{t\to\infty}\sum\nolimits_{k=1}^{\infty}F_k(t)\ge\lim_{t\to\infty}\sum\nolimits_{k=1}^{n}F_k(t),\forall n\\ &=\sum\nolimits_{k=1}^{n}\lim_{t\to\infty}F_k(t)=n\cdot 1=n.\end{aligned}$$

Since n is arbitrary, we have $\lim_{t\to\infty} m(t)=\infty$. QED

What about $\lim_{t\to\infty}\frac{m(t)}{t}$? It seems tempting to do the following argument:

$$t\approx S_{N(t)}=\sum\nolimits_{i=1}^{N(t)}X_i\Rightarrow t=\mathrm{E}(t)\approx\mathrm{E}\left[\sum\nolimits_{i=1}^{N(t)}X_i\right]=\mathrm{E}[N(t)]\cdot\mathrm{E}(X_1)=m(t)\cdot\mu,$$

and thus we have $\frac{m(t)}{t}\approx\frac{1}{\mu}$. However, the argument above was false in that $N(t)$ and X_i's are not independent as required. When X_i's are very small, we know $N(t)$ will be large and vice versa. Hence $N(t)$ and X_i's are correlated with each other. Specifically, $\mathrm{E}\left[\sum\nolimits_{i=1}^{N}X_i\right]=\mathrm{E}(N)\cdot\mathrm{E}(X_1)$ holds only if X_i's are i.i.d., independent of N and $\mathrm{E}\left|X_1\right|<\infty$ and $\mathrm{E}(N)<\infty$. As a matter of fact, we are going to prove in **Theorem 5.6** that $\lim_{t\to\infty}\frac{m(t)}{t}=\frac{1}{\mu}$, using additional tools that we haven't obtained yet.

Example 5.6 Let $X_i=1$ if I win the i^{th} toss of a fair coin and $X_i=-1$ otherwise. Clearly we have $\Pr(X_i=1)=\frac{1}{2}=\Pr(X_i=-1)$. Moreover, my wins after n plays are $\sum\nolimits_{i=1}^{n}X_i$. Let N be the first time that I am ahead by \$1. (If I win for the first game, then $N=1$; otherwise, I need win the next two games to get $N=3$, etc.) By definition, we have $\sum\nolimits_{i=1}^{N}X_i=1$, but $\mathrm{E}(N)\cdot\mathrm{E}(X_1)=\mathrm{E}(N)\cdot 0=0\ne\mathrm{E}\left[\sum\nolimits_{i=1}^{N}X_i\right]$.

As another example, suppose $\Pr(X_i=-1)=q$, $\Pr(X_i=1)=p$ and $p<q$. We are having an unfair game here. Let N be the last time that I break even, then $\sum\nolimits_{i=1}^{N}X_i=0$ and $N\ge 0$, but $\mathrm{E}(N)\cdot\mathrm{E}(X_1)=\mathrm{E}(N)\cdot(p-q)<0=\mathrm{E}\left[\sum\nolimits_{i=1}^{N}X_i\right]$.

Definition 5.3 Let N be a non-negative, integer-value r.v. Then N is a *stopping time* (*Markov time, optional time*) with respect to the sequence of r.v.s. $\{X_n:n\ge 0\}$ if $I(N=k)=\mathrm{f}_k(X_0,\ldots,X_k),\forall k\ge 0$, where f_k is a deterministic function and I is the indicator function. We say that $I(N=k)$ is "determined by" $X_0,\dots,X_k$. Furthermore, if $X_0,\dots,X_k$ are independent, then the event $\{N=n\}$ is independent of $X_{n+1},X_{n+2},\dots,\forall n=0,1,\dots$.

Comment 5.7 Let A be a given set and $\{X_n : n\ge 0\}$ be a stochastic process. Let $N=\inf\{n\ge 0 : X_n\in A\}=\{\text{the first time hit set } A\}$. N is a stopping time if

$$\begin{aligned} I(N=k) &= I[X_0\notin A, X_1\notin A,\dots,X_{k-1}\notin A, X_k\in A] \\ &= \mathrm{f}_k(X_0,X_1,\dots,X_k),\forall k\ge 0 \end{aligned}$$

i.e., the event $\{N=k\}$ doesn't depend future events $X_{k+1},X_{k+2},\dots$.
Is $N'=N-1$ a stopping time? No. Because $I(N'=k)=I(N=k+1)=\mathrm{f}_{k+1}(X_0,X_1,\dots,X_{k+1})$ and thus $I(N'=k)$ does depend upon one of the future events, X_{k+1}.

Definition 5.4 Let N be a non-negative integer-valued r.v. Then N is a *generalized stopping time* (*randomized stopping time*) with respect to the sequence of r.v.s. $\{X_n : n\ge 0\}$ if the event $\{N=n\}$ is independent of $X_{n+1},X_{n+2},\dots$, for all $n=0,1,\dots$.

Theorem 5.4 *Wald's Equation* Let $X_1,X_2,\dots$ be i.i.d. r.v.s. and N be a (generalized) stopping time with restpect to $X_1,X_2,\dots$. Let $S_0=0$ and $S_n=\sum_{i=1}^n X_i$. If either (1) $X_i\ge 0,\forall i$ almost surely ; or (2) $\mathrm{E}[|X_i|]<\infty,\forall i$ and $\mathrm{E}(N)<\infty$, then we have $\mathrm{E}(S_N)=\mathrm{E}(N)\cdot\mathrm{E}(X_i)$.
Proof: Using the old trick, we have $\mathrm{E}\left[\sum_{i=1}^N X_i\right]=\mathrm{E}\left[\sum_{i=1}^\infty X_i\cdot I(N\ge i)\right]$. Note that the event $\{N\ge i\}$ stands for "stop at time i or later," which is equivalent to "don't stop at any time before $i-1$."
That is $I(N\ge i)=1-I(N\le i-1)=1-\sum_{j=1}^{i-1} I(N=j)=1-\sum_{j=1}^{i-1}\mathrm{f}_j(X_1,\dots,X_j)$, which is a function of $X_1,\dots,X_{i-1}$. Hence we have $I(N\ge i)$ independent of X_i.

Now we can write

$$\begin{aligned} \mathrm{E}\left[\sum_{i=1}^\infty X_i\cdot I(N\ge i)\right] &= \sum_{i=1}^\infty \mathrm{E}[X_i\cdot I(N\ge i)] && \langle\text{by Fubini's theorem}\rangle \\ &= \sum_{i=1}^\infty \mathrm{E}[X_i]\cdot\mathrm{E}[I(N\ge i)] && \langle\text{by independence of } X_i \text{ and } I(N\ge i)\rangle \\ &= \mathrm{E}[X_i]\cdot\sum_{i=1}^\infty \Pr(N\ge i) \\ &= \mathrm{E}[X_i]\cdot\mathrm{E}(N). \end{aligned}$$

Why can we use Fubini's theorem above? We can certainly do so if the first condition satisfies, that is $X_i\ge 0,\forall i$ almost surely . Note that we can also do so if it holds that $\mathrm{E}\left[\sum_{i=1}^\infty |X_i\cdot I(N\ge i)|\right]<\infty$, which is implied by condition (2) in the theorem. How? We have $\mathrm{E}\left[\sum_{i=1}^\infty |X_i\cdot I(N\ge i)|\right]=\mathrm{E}\left[\sum_{i=1}^N |X_i|\right]$ and since $|X_i|\ge 0,\forall i$, we can use the condition (1) in the theorem and the result in the theorem to get $\mathrm{E}\left[\sum_{i=1}^N |X_i|\right]=\mathrm{E}(N)\cdot\mathrm{E}|X_i|$. Hence if we have condition (2), $\mathrm{E}(N)<\infty$ and $\mathrm{E}|X_i|<\infty$, then we have $\mathrm{E}\left[\sum_{i=1}^\infty |X_i\cdot I(N\ge i)|\right]=\mathrm{E}(N)\cdot\mathrm{E}|X_i|<\infty$. QED

Comment 5.8 Now let's consider the reason why we cannot use *Wald's Equation* in the two cases of **Example 5.6** above. For the first case, everything is fine except that $\mathrm{E}(N) = \infty$. How? If we have $\mathrm{E}(N) < \infty$, then we would be able to use *Wald's Equation* and the argument used before to reach a contradition $1 = 0$. Therefore, $\mathrm{E}(N) = \infty$. In the second case, we can actually show that $\mathrm{E}(N) < \infty$. Why cannot we use the *Wald's Equation* then? Because N is not a stopping time in this case in that it depends on the future X_i's.

Example 5.7 Suppose a couple use a decision rule to decide when to stop having babies. Let $X_i = 1$ if the i^{th} child is a boy with probability $\frac{1}{2}$ and $X_i = -1$ if the i^{th} child is a girl with probability $\frac{1}{2}$. (Note that we choose –1 and 1 here to make sure the expected sum of children would reflect some interesting features, more girls or boys or equal, which couldn't be reflected if we were to use other numbers.) Let's assume that $X_1, X_2, \ldots$ are i.i.d. and let N be the stopping time according to the decision rule. Apparently it is realistic to assume that N is a stopping time since future events won't matter to the current stopping time.

Some physical limitations would also suggest the natural assumption of $\mathrm{E}(N) < \infty$. Define $S_n = \sum_{i=1}^{n} X_i$. Clearly $\mathrm{E}(X_1) = 0$ and if $\mathrm{E}(N) < \infty$ holds indeed, then we can use *Wald's Equation* to reach the conclusion: $\mathrm{E}(S_n) = \mathrm{E}(N) \cdot \mathrm{E}(X_1) = 0$.

Let's try one decision rule: "Have children until having the first boy." Clearly the stopping time N has a geometric distribution with mean 1 (It is possible that $N = 0$ so we use the version $\mu = \frac{1-p}{p}$ in **Comment 1.5.(4)**.), and thus we have $\mathrm{E}(S_n) = \mathrm{E}(N) \cdot \mathrm{E}(X_1) = 1 \cdot 0 = 0$.

Let's try another rule: "Have children until the k^{th} boy." We can also show that $\mathrm{E}(N) < \infty$ and thus $\mathrm{E}(S_n) = 0$. Clearly, these two rules point to the same conclusion that we cannot change the sex mix in the long run.

What about this one? "Have children until more boys than girls." We will certainly have $\mathrm{E}(S_n) > 0$, but we actually have $\mathrm{E}(N) = \infty$ so that we couldn't use *Wald's Equation*. Why? Suppose not, we have $\mathrm{E}(N) < \infty$. Then we have

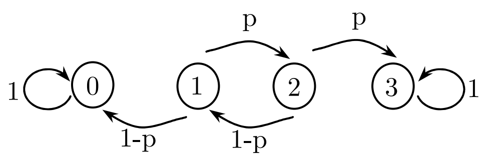
$$\mathrm{E}(N) = \mathrm{E}[\mathrm{E}(N \mid X_1)] = \mathrm{E}(N \mid X_1 = 1) \cdot \tfrac{1}{2} + \mathrm{E}(N \mid X_1 = -1) \cdot \tfrac{1}{2} = 1 \cdot \tfrac{1}{2} + [\mathrm{E}(N') + \mathrm{E}(N'')] \cdot \tfrac{1}{2},$$

where N' stands for the stopping time from "more girls than boys" to "equal number of girls and boys," and N'' stands for the stopping time from "equal number of girls and boys" to "more boys than girls." Moreover, we know that N' and N'' are identically distriuted with N, and thus $\mathrm{E}(N') = \mathrm{E}(N'') = \mathrm{E}(N)$, which implies $\mathrm{E}(N) = \frac{1}{2} + \mathrm{E}(N)$, a contradiction. Therefore, we must have $\mathrm{E}(N) = \infty$.

Example 5.8 *Gambler's Ruin* Suppose that person I and II are gambling. Person I starts with \$1 and person II starts with \$2. During each play, person I wins \$1 from person II with probability p, and loses

\$1 to person II with probability $(1-p)$. Assume that plays are independent and the game will continue until one of the players runs out of money, when is often called a *ruin time*.
Let $X_i = 1$ if person I wins with probability p and $X_i = -1$ if person II wins with probability $(1-p)$. Let N be the number of plays until the ruin time. We can easily show that N is a stopping time. Let F_n be the amount of fortune of person I at time n, so we have $F_n = 1 + \sum_{i=1}^{n} X_i$. Can we use the *Wald's Equation* here? We need to find out $\mathrm{E}(N)$ and see if it is finite.

Let N_i be the time to ruin when person I starts with \$ $i, i = 1,2$. Clearly, in this setup, we have $\mathrm{E}(N) = \mathrm{E}(N_1)$ since person I starts with \$1. Let's consider the following dynamic scheme:



We can write

$$\mathrm{E}(N_1) = \mathrm{E}[\mathrm{E}(N_1 \mid X_1)] = (1-p)\cdot 1 + p\cdot[1+\mathrm{E}(N_2)] = 1 + p\cdot \mathrm{E}(N_2),$$

and

$$\mathrm{E}(N_2) = \mathrm{E}[\mathrm{E}(N_2 \mid X_1)] = p\cdot 1 + (1-p)\cdot[1+\mathrm{E}(N_1)] = 1 + (1-p)\cdot \mathrm{E}(N_1).$$

From the two equation systems, we can solve $\mathrm{E}(N_1) = \frac{1+p}{1-p(1-p)}$, which can be shown to be finite. Therefore, we have $\mathrm{E}(F_n) = 1 + \mathrm{E}(N)\cdot \mathrm{E}(X_1) = 1 + \frac{1+p}{1-p(1-p)}\cdot[p\cdot 1 + (1-p)\cdot(-1)] = 1 + \frac{(2p-1)(1+p)}{1-p(1-p)}$. To give a specific example, let $p = \frac{2}{3}$, then we have $\mathrm{E}(F_n) = \$1.71$.

Definition 5.5 Suppose that $\{X_n : n \geq 1\}$ is a stochastic process with independent X_n that has $\mathrm{E}\left|X_i\right| < \infty$ and $\mathrm{E}(X_i) = \mu, \forall i \geq 1$. Then $\{M_n : n \geq 1\}$ is a margingale with respect to $\{X_n : n \geq 1\}$ if for each $n \geq 1$: (1) $M_n = \mathrm{f}_n(X_1,\dots,X_n)$ is some deterministic function f_n; (2) $\mathrm{E}\left|M_n\right| < \infty$; (3) $\mathrm{E}[M_{n+1} \mid X_1,\dots,X_n] = M_n$.

Define $M_n = \sum_{i=1}^{n}(X_i - \mu)$ and we can verify that it is a martingale.

(1) M_n is clearly a deterministic function of $X_1,\dots,X_n$.

(2) $\mathrm{E}\left|M_n\right| = \mathrm{E}\left|\sum_{i=1}^{n}(X_i-\mu)\right| \leq \mathrm{E}\sum_{i=1}^{n}\left|X_i-\mu\right| = \sum_{i=1}^{n}\mathrm{E}\left|X_i-\mu\right| \leq \sum_{i=1}^{n}\left[\mathrm{E}\left|X_i\right| + |\mu|\right] < \infty$.

(3)

$$\begin{aligned}\mathrm{E}[M_{n+1} \mid X_1,\dots,X_n] &= \mathrm{E}\left[\sum_{i=1}^{n+1}(X_i-\mu) \mid X_1,\dots,X_n\right] = \mathrm{E}\left[\sum_{i=1}^{n}(X_i-\mu) + (X_{n+1}-\mu) \mid X_1,\dots,X_n\right] \\ &= M_n + \mathrm{E}[(X_{n+1}-\mu) \mid X_1,\dots,X_n] = M_n + \mathrm{E}(X_{n+1}-\mu) \qquad \text{<by independence of } X_i\text{>} \\ &= M_n.\end{aligned}$$

Example 5.9 If $\{Z_n : n \geq 1\}$ is a martingale, show that for $1 \leq k < n$, $\mathrm{E}[Z_n \mid Z_1,\dots,Z_k] = Z_k$.
Proof: Clearly, we have $\mathrm{E}[Z_{k+1} \mid Z_1,\dots,Z_k] = Z_k$ from the definition of martingale. Assume that $\mathrm{E}[Z_{k+i} \mid Z_1,\dots,Z_k] = Z_k, \forall i$, then we can write

$$\mathrm{E}[Z_{k+i+1} \mid Z_1,\dots,Z_k] = \mathrm{E}\{\mathrm{E}[Z_{k+i+1} \mid Z_1,\dots,Z_{k+1}] \mid Z_1,\dots,Z_k\} = \mathrm{E}[Z_{k+1} \mid Z_1,\dots,Z_k] = Z_k. \text{ QED}$$

Theorem 5.5 *The Martingale (Optional) Stopping Theorem*

Let $\{M_n : n \geq 1, ...\}$ be a martingale with respect to $\{X_n : n \geq 1\}$, and let N be a stopping time with restpect to $\{X_n : n \geq 1\}$. If $\mathrm{E}(N) < \infty$ and there exists a $K < \infty$ such that $\mathrm{E}\left[\left|M_{n+1} - M_n\right| \mid X_1, X_2, ..., X_n\right] \leq K$ for $n \leq N$, then $\mathrm{E}(M_N) = \mathrm{E}(M_1)$.

Comment 5.9 We are not going to prove this theorem here since we don't have enough tools yet. But we do want to demonstrate the condition for the case $M_n = \sum_{i=1}^{n}(X_i - \mu)$.

$$\mathrm{E}\left[\left|M_{n+1} - M_n\right| \mid X_1, X_2, ..., X_n\right] = \mathrm{E}\left[\left|X_{n+1} - \mu\right| \mid X_1, X_2, ..., X_n\right] = \mathrm{E}\left|X_{n+1} - \mu\right| \leq \mathrm{E}\left|X_{n+1}\right| + \mathrm{E}\left|\mu\right| \equiv K.$$

The point here is that despite the fact that we can get $\mathrm{E}(M_n) = \mathrm{E}(M_1) = 0, \forall n$, but $\mathrm{E}(M_N) = \mathrm{E}(M_1) = 0$ may not hold all the time. This theorem essentially tells us that there is no free lunch.

Example 5.10 Consider a gambler who starts out with no money, and on each gamble is equally likely to win \$1 or lose \$1. Suppose that the gambler will quit playing at time T when his winnings are either A or $-B$, where $A > 0, B > 0$. Let X_i be the gambler's winnings on play number i. (a) Show that $\{S_n : n \geq 1\}$ is a martingale, where $S_n = \sum_{i=1}^{n} X_i$; (b) Use the *optional stopping theorem* to show that $\mathrm{E}(S_T) = 0$ and hence that $\Pr(S_T = A) = \frac{B}{A+B}$; (c) Show that $\{M_n : n \geq 1\}$ is a margtingale, where $M_n = S_n^2 - n$; (d) Use the *optional stopping theorem* to show that $\mathrm{E}(M_T) = 0$ and hence that $\mathrm{E}(T) = AB$.

We have $\mathrm{E}(X_i) = 0$ and $\mathrm{E}(X_i^2) = \mathrm{E}(|X_i|) = 1$.

(a) Clearly S_n is a deterministic function of only $X_1, ..., X_n$.

$$\mathrm{E}(|S_n|) = \mathrm{E}\left[\left|\sum_{i=1}^{n} X_i\right|\right] \leq \mathrm{E}\left[\sum_{i=1}^{n}\left|X_i\right|\right] = \sum_{i=1}^{n} \mathrm{E}\left|X_i\right| = n < \infty.$$

$$\mathrm{E}(S_{n+1} \mid X_1, ..., X_n) = \mathrm{E}(S_n + X_{n+1} \mid X_1, ..., X_n) = S_n + \mathrm{E}(X_{n+1}) = S_n.$$

Therefore, S_n is a martingale.

(b) T is a stopping time because T doesn't depend upon the results of future games.

$$\mathrm{E}\left[\left|S_{n+1} - S_n\right| \mid X_1, ..., X_n\right] = \mathrm{E}\left[\left|X_{n+1}\right| \mid X_1, ..., X_n\right] = \mathrm{E}\left(\left|X_{n+1}\right|\right) = 1 < \infty$$

Hence we can use the *optional stopping theorem* to write

$$\mathrm{E}(S_T) = \mathrm{E}(S_1) = 0 = \Pr(S_T = A) \cdot A + [1 - \Pr(S_T = A)] \cdot (-B) \Rightarrow \Pr(S_T = A) = \tfrac{B}{A+B}.$$

(c) Clearly M_n is a deterministic function of only $X_1, ..., X_n$.

$$\mathrm{E}(|M_n|) = \mathrm{E}\left[\left|S_n^2 - n\right|\right] \leq \mathrm{E}\left[\left|S_n^2\right|\right] + n = \mathrm{E}\left[\sum_{i=1}^{n}\left|X_i^2\right| + 2 \cdot \sum_{i=1}^{n}\sum_{j=1}^{i}\left|X_i X_j\right| + n\right] = n^2 + n < \infty$$

$$\mathrm{E}(M_{n+1} \mid X_1, ..., X_n) = \mathrm{E}[S_{n+1}^2 - (n+1) \mid X_1, ..., X_n] = \mathrm{E}[(S_n^2 - n) + (2S_n X_{n+1} + X_{n+1}^2) - 1 \mid X_1, ..., X_n] = M_n$$

Therefore, M_n is a martingale.

(d) Since

$$\begin{aligned}\mathrm{E}\left[\left|M_{n+1} - M_n\right| \mid X_1, ..., X_n\right] &= \mathrm{E}\left\{\left|[S_{n+1}^2 - (n+1)] - [S_n^2 - n]\right| \mid X_1, ..., X_n\right\} \\ &= \mathrm{E}\left\{\left|(2S_n + X_{n+1})X_{n+1} - 1\right| \mid X_1, ..., X_n\right\} \leq 2 \cdot 0 \cdot 0 + 1 - 1 = 0 < \infty,\end{aligned}$$

we can use the *optional stopping time theorem* to get

$$\mathrm{E}(M_T) = \mathrm{E}(M_1) = \mathrm{E}(X_1^2 - 1) = 0 \text{ or } 0 = \mathrm{E}(T)\cdot \mathrm{E}(S_T^2 - T) \Rightarrow \mathrm{E}(T) = \mathrm{E}(S_T^2) = A^2 \cdot \tfrac{B}{A+B} + B^2 \cdot \tfrac{A}{A+B} = AB.$$

Example 5.11 Let $\{M_n : n \geq 1\}$ be a martingale with respect to $\{X_n : n \geq 1\}$. Suppose that $\{W_n : n \geq 1\}$ is a real-valued sequence such that: (a) $W_i = g_i(X_1,\ldots,X_i)$, $i \geq 1$, for some deterministic function g_i; and (b) $|W_i| \leq c$ where $c > 0$. Set $Z_1 = 0$ and $Z_n = \sum_{i=1}^{n} W_{i-1}(M_i - M_{i-1})$. Show that $\{Z_n : n \geq 1\}$ is a martingale with respect to $\{X_n : n \geq 1\}$.

Proof: It is easy to verify that Z_n is indeed a determinstric function of $X_1,\ldots,X_n$.

$$\mathrm{E}[|Z_n|] = \mathrm{E}\left[\left|\sum_{i=1}^{n} W_{i-1}(M_i - M_{i-1})\right|\right] \leq \mathrm{E}\left[\left|\sum_{i=1}^{n} W_{i-1}M_i\right|\right] + \mathrm{E}\left[\left|\sum_{i=1}^{n} W_{i-1}M_{i-1}\right|\right]$$

$$\leq c\cdot \mathrm{E}\left[\sum_{i=1}^{n} |M_i|\right] + c\cdot \mathrm{E}\left[\sum_{i=1}^{n} |M_{i-1}|\right] < \infty$$

$$\mathrm{E}[Z_{n+1} \mid X_1,\ldots,X_n] = \mathrm{E}[\sum_{i=1}^{n} W_{i-1}(M_i - M_{i-1}) \mid X_1,\ldots,X_n] = \mathrm{E}[Z_n + W_n(M_{n+1} - M_n) \mid X_1,\ldots,X_n]$$

$$= Z_n + \mathrm{E}[W_n(M_{n+1} - M_n) \mid X_1,\ldots,X_n] = Z_n + g_n(X_1,\ldots,X_n)\cdot\{\mathrm{E}[M_{n+1} \mid X_1,\ldots,X_n] - M_n\} = Z_n$$

Lemma 5.1 Let $\{N(t) : t \geq 0\}$ be a renewal process. Then $\mathrm{E}(S_{N(t)+1}) = \mathrm{E}\left[\sum_{i=1}^{N(t)+1} X_i\right] = \mu\cdot[m(t)+1]$.

Proof: We have pondered before whether or not we can have $\mathrm{E}\left[\sum_{i=1}^{N} X_i\right] = \mathrm{E}(N)\cdot \mathrm{E}(X_i)$ under very general conditions, and the answer is given by the *Wald's Equation*. If we can prove that N is a stopping time, then we can certainly apply the *Wald's Equation*. Hence let's consider the following question: Is $I\{N(t) = 2\}$ a function of only X_1, X_2? Given both X_1 and X_2, we cannot safely conclude whether or not $N(t) = 2$ because it is possible that the third event will occur also ahead of t; that is, only if the third event occurs after t, can we safely say $N(t) = 2$. Equivalently, $I\{N(t) = 2\}$ is not a function of only X_1, X_2, and thus $N(t)$ is not a stopping time.

But we can say that $N(t)+1$ is a stopping time, why? Write down $I[N(t)+1 = 2] \overset{?}{=} \mathrm{f}_2(X_1, X_2)$ or $I[N(t) = 1] \overset{?}{=} \mathrm{f}_2(X_1, X_2)$. If t happens to be before the first two events, then clearly we have $I[N(t) = 1] = 0$; if t happens to be after the first two events, then clearly we have $I[N(t) = 1] = 0$; if time t sandwiched between the first two events, then we have $I[N(t) = 1] = 1$. Therefore, $I[N(t) = 1] = \mathrm{f}_2(X_1, X_2)$ and $N(t)+1$ is a stopping time. (Remember that we have proved in **Comment 5.7** that $N-1$ is not a stopping time even though N is?) Formally, we have $\{N(t) = k\} \Leftrightarrow \{X_1 + \ldots + X_k \leq t \text{ and } X_1 + \ldots + X_{k+1} > t\}$, and $\{N(t)+1 = k\} \Leftrightarrow \{N(t) = k-1\}$ $\Leftrightarrow \{X_1 + \ldots + X_{k-1} \leq t \text{ and } X_1 + \ldots + X_k > t\}$, i.e., $I\{N(t)+1 = k\}$ depends on only $X_1,\ldots,X_k$ and thus $N(t)+1$ is a stopping time. Applying the *Wald's Equation*, we have

$$\mathrm{E}(S_{N(t)+1}) = \mathrm{E}\left[\sum_{i=1}^{N(t)+1} X_i\right] = \mathrm{E}(X_1)\cdot \mathrm{E}[N(t)+1] = \mu\cdot[m(t)+1]. \text{ QED}$$

Theorem 5.6 *The Elementary Renewal Theorem* $\lim_{t\to\infty} \frac{m(t)}{t} = \frac{1}{\mu}$.

Proof: Clearly we have $S_{N(t)+1} \geq t$, which implies $\mathrm{E}[S_{N(t)+1}] \geq t$ or $[m(t)+1]\cdot\mu \geq t$, by **Lemma 5.1**. If $\mu < \infty$, we can then write $m(t)+1 \geq \frac{t}{\mu}$ and thus $\frac{m(t)}{t} \geq \frac{1}{\mu} - \frac{1}{t}$. Hence $\liminf_{t\to\infty} \frac{m(t)}{t} \geq \frac{1}{\mu}$. If $\mu = \infty$, then the result holds trivially.

Let's now define a truncated process as following: $\bar{X}_n = X_n$ if $X_n \le M$ and $\bar{X}_n = M$ if $X_n > M$, for some $M > 0$. Now it becomes clear that $S_{\bar{N}(t)+1} \le t + M$ by our design. Thefore, we have $\mathrm{E}[S_{\bar{N}(t)+1}] \le t + M$ or $[\bar{m}(t)+1]\cdot \mu_M \le t + M$, where $\mu_M = \mathrm{E}(\bar{X}_1)$. Clearly $\mu_M \le M$, we can divided both sides of the inequality by μ_M and get $\bar{m}(t)+1 \le \frac{t+M}{\mu_M}$ and thus $\frac{\bar{m}(t)}{t} \le \frac{1}{\mu_M} + \frac{M}{\mu_M \cdot t} - \frac{1}{t}$, which implies $\limsup_{t\to\infty} \frac{\bar{m}(t)}{t} \le \frac{1}{\mu_M}$.

We know that $X_i \ge \bar{X}_i, \forall i$, and thus $N(t) \le \bar{N}(t)$ so that $m(t) \le \bar{m}(t)$. Hence $\limsup_{t\to\infty} \frac{m(t)}{t} \le \limsup_{t\to\infty} \frac{\bar{m}(t)}{t} \le \frac{1}{\mu_M}$. Moreover, since M is arbitrary, so if we let $M \to \infty$, we get $\frac{1}{\mu_M} \to \frac{1}{\mu}$, by the Monotonic Convergence Theorem. Now we have $\frac{1}{\mu} \le \limsup_{t\to\infty} \frac{m(t)}{t} \le \limsup_{t\to\infty} \frac{\bar{m}(t)}{t} \le \frac{1}{\mu}$, which implies that $\lim_{t\to\infty} \frac{m(t)}{t} = \limsup_{t\to\infty} \frac{m(t)}{t} = \limsup_{t\to\infty} \frac{\bar{m}(t)}{t} = \frac{1}{\mu}$. QED

Comment 5.10 What are $\limsup_{t\to\infty} X_t$ and $\liminf_{t\to\infty} X_t$? We know that $\lim_{t\to\infty} X_t$ may not exist, but both $\limsup_{t\to\infty} X_t$ and $\liminf_{t\to\infty} X_t$ will always exist. In particular, $\limsup_{t\to\infty} X_t = \lim_{t\to\infty}\left[\sup_{k\ge t} X_k\right]$ and $\liminf_{t\to\infty} X_t = \lim_{t\to\infty}\left[\inf_{k\ge t} X_k\right]$. We always have $\liminf_{t\to\infty} X_t \le \limsup_{t\to\infty} X_t$ and if $\liminf_{t\to\infty} X_t = \limsup_{t\to\infty} X_t$ then $\lim_{t\to\infty} X_t = \liminf_{t\to\infty} X_t = \limsup_{t\to\infty} X_t$. For example, consider the case $X_t = (-1)^t$. We have $\liminf_{t\to\infty} X_t = -1$ and $\limsup_{t\to\infty} X_t = 1$. But $\lim_{t\to\infty} X_t$ doesn't exist.

Theorem 5.7 *Central Limit Theorem for Renewal Processes* Let $\{N(t): t \ge 0\}$ be a renewal process. Let μ and σ^2, assumed finite, represent the mean and variance of the inter-renewal times. Then $\Pr\left[\frac{N(t)-t/\mu}{\sigma\sqrt{t/\mu^3}} < y\right] \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-x^2/2}dx$ as $t \to \infty$. Note that this implies that, for large enough t, $N(\mathrm{t})$ is approximately normal with mean t/μ and variance $t\sigma^2/\mu^3$.

Example 5.12 Suppose that X_i's are uniformly distributed over $(0,1)$. Suppose we want to know $\Pr[N(6) \ge 10]$. Clearly, $\mu = \mathrm{E}(X_i) = \frac{1}{2}$ and $\mathrm{Var}(X_i) = \frac{1}{12}$. Then by the **Theorem 5.7**, we have $\Pr[N(6) \ge 10] = \Pr\left[\frac{N(6)-\frac{6}{1/2}}{\sqrt{\frac{1}{12}\cdot 6/(\frac{1}{2})^3}} \ge \frac{10-\frac{6}{1/2}}{\sqrt{\frac{1}{12}\cdot 6/(\frac{1}{2})^3}}\right] = \Pr[Z(\cdot) \ge -1] = 1 - \Phi(-1) \approx 0.89.$

Definition 5.6 A non-negative r.v. X is lattice if there is a constant $d > 0$ such that $\sum_{n=0}^{\infty} \Pr(X = nd) = 1$. The largest such d is called the *period* of the lattice. If X is lattice and F is the distribution function of X, then we say that F is lattice.

Comment 5.11 Note that a lattice is a special type of discrete r.v. but a discrete r.v. may not necessarily be a lattice. For example, let $X = 1$ with probability $\frac{1}{2}$ and $X = c$ with probability $\frac{1}{2}$. Then we have X as a discrete r.v. but not necessarily lattice. Specifically, if we use any rational number for c, then the discrete r.v. is a lattice. Otherwise, say, $c = \pi$, it is not a lattice. Proof of the latter case? Suppose that

X is lattice for $c = \pi$, then $1 = nd$ for some rational number n and $\pi = md$ for some rational number m by the definition. But $\pi = \frac{m}{n}$ is rational as the fraction of a rational number to another rational number? The presence of lattice cautiouses us when taking limits. Recall that the *Elementary Renewal Theorem* says that $\frac{m(t)}{t} \to \frac{1}{\mu}$ a.s. as $t \to \infty$. So we know if $m(t) = \frac{t}{\mu}$ then $m(t+a) - m(t) = \frac{a}{\mu}$. The generalized result in in Blackwell's theorem below.

Theorem 5.9 *Blackwell's Theorem*

(1) If F is not lattice, then for all $a \geq 0$, $\lim_{t\to\infty} m(t+a) - m(t) = \frac{a}{\mu}$.

(2) If F is lattice with period d, then for $k = 1,2,\ldots$, $\lim_{n\to\infty} m(nd+kd) - m(nd) = \frac{kd}{\mu}$. For $k = 1$, this implies that $\lim_{n\to\infty} \mathrm{E}(\text{number of renewals at time } nd) = \frac{d}{\mu}$.

Proof: We sketch a proof for part (1) only. Assuming that limits exist. Let $\mathrm{f}(x) = \lim_{t\to\infty} m(t+x) - m(t)$, then we have

$$\mathrm{f}(x+a) = \lim_{t\to\infty}\{[m(t+x+a) - m(t+x)] + [m(t+x) - m(t)]\} = \mathrm{f}(a) + \mathrm{f}(x).$$

Then we have $\mathrm{f}(x) = cx$ for some c. We want to show that $c = \frac{1}{\mu}$. Define $z_t = m(t) - m(t-1)$, then $\lim_{t\to\infty} z_t = \mathrm{f}(1) = c$, which implies that $\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} z_t = c$, that is, $\lim_{n\to\infty} \frac{1}{n}[m(1) - m(0) + m(2) - m(1) + \ldots + m(n) - m(n-1)] = c$, i.e., $\lim_{n\to\infty}\frac{m(n)}{n} - \lim_{n\to\infty}\frac{m(0)}{n} = c$. By the *Elementary Theorem*, we have $\lim_{n\to\infty}\frac{m(n)}{n} = \frac{1}{\mu}$ and we also have $\lim_{n\to\infty}\frac{m(0)}{n} = 0$. Hence we have $c = \frac{1}{\mu}$. QED

Theorem 5.9 *The Key Renewal Theorem*

If F is not lattice and $h(t)$ satisfies: (1) $h(t) \geq 0, \forall t \geq 0$; (2) $h(t)$ is nonincreasing; (3) $\int_0^\infty h(t)dt < \infty$, (any such $h(t)$ is also called "directly Riemann integrable.") then $\lim_{t\to\infty}\int_0^t h(t-x)dm(x) = \frac{1}{\mu}\int_0^\infty h(t)dt$.

Comment 5.12 We know that $\frac{m(t)}{t} \to \frac{1}{\mu}$ a.s. as $t \to \infty$. If $m(t) = \frac{t}{\mu}$, then $dm(t) = \frac{dt}{\mu}$ and thus

$$\lim_{t\to\infty}\int_0^t h(t-x)dm(x) = \lim_{t\to\infty}\frac{1}{\mu}\int_0^t h(t-x)dx = \lim_{t\to\infty}\frac{1}{\mu}\int_0^t h(u)du = \frac{1}{\mu}\int_0^\infty h(u)du.$$

The *Blackwell's Theorem* says that $\lim_{t\to\infty}\frac{m(t+a)-m(t)}{a} = \frac{1}{\mu}$, which implies that $\lim_{a\to 0}\left[\lim_{t\to\infty}\frac{m(t+a)-m(t)}{a}\right] = \lim_{a\to 0}\frac{1}{\mu} = \frac{1}{\mu}$, i.e., $\lim_{t\to\infty}\frac{dm(t)}{dt} = \frac{1}{\mu}$, which is what we need to get the result similar to that in the case where $m(t) = \frac{t}{\mu}$.

Definition 5.7 An equation of the form $g(t) = h(t) + \int_0^t g(t-x)dF(x)$ is called a *renewal type equation*. In convolution the above states that $g = h + g \otimes F$.

Proposition 5.4 A renewal type equation has solution $g(t) = h(t) + \int_0^t h(t-x)dm(x)$.

Proof: Before providing the formal proof, we provide the following intuitive reasoning:

$$g = h + g\otimes F = h + (h + g\otimes F)\otimes F = h + h\otimes F + g\otimes F\otimes F = h + h\otimes\sum_{n=1}^{\infty} F_n = h + h\otimes M.$$

We use the Laplace tranform to prove the result formally. If we take Laplace tranform on both sides of the renewal type equation, we have $\tilde{g}(s) = \tilde{h}(s) + \tilde{g}(s)\tilde{F}(s)$. (Note that Laplace transform is very effectively when calculating convolution.) Hence, we have $\tilde{g}(s)=\frac{\tilde{h}(s)}{1-\tilde{F}(s)}$, since all the Laplace transforms are numbers. Finally, we get $g(t)=h(t)+\int_0^t h(t-x)dm(x)$ from the following decomposition,:

$$\tilde{g}(s)=\frac{\tilde{h}(s)}{1-\tilde{F}(s)}=\frac{\tilde{h}(s)[1-\tilde{F}(s)+\tilde{F}(s)]}{1-\tilde{F}(s)}=\tilde{h}(s)+\tilde{h}(s)\frac{\tilde{F}(s)}{1-\tilde{F}(s)}=\tilde{h}(s)+\tilde{h}(s)\tilde{m}(s).$$

Theorem 5.10 *The Basic Renewal Theorem*

If F is not lattice, $h(t)$ satisfies (1) $h(t)\geq 0,\forall t\geq 0$; (2) $h(t)$ is nonincreasing; (3) $\int_0^\infty h(t)dt<\infty$, and $g(t)$ satisfies the renewal type equation, i.e., $g(t)=h(t)+\int_0^t g(t-x)dF(x)$, then $\lim_{t\to\infty} g(t)=\frac{1}{\mu}\int_0^\infty h(t)dt$.

Proof: Applying the **Proposition 5.4**, we have $g(t)=h(t)+\int_0^t h(t-x)dm(x)$. Hence

$$\lim_{t\to\infty} g(t)=\lim_{t\to\infty} h(t)+\lim_{t\to\infty}\int_0^t h(t-x)dm(x)=\lim_{t\to\infty} h(t)+\frac{1}{\mu}\int_0^\infty h(t)dt \quad \text{<by Theorem 5.9>}$$
$$=\frac{1}{\mu}\int_0^\infty h(t)dt. \quad \text{<by properties of } h(t)\text{> QED}$$

Theorem 5.11 *An Alternating Renewal Process*

Consider a system that can be in one of two states called "on" or "off." Initially it is on for a random time X_1, followed by being off for a random time Y_1. It is then on for X_2 and off for Y_2, etc. Suppose that X_i's are i.i.d. r.v.s. with c.d.f. F and the Y_i's are i.i.d. r.v.s. with c.d.f. G. Although X_i is allowed to be dependent on Y_i, the pairs $\{X_i,Y_i\}$ are i.i.d. Let $p(t)$ be the probability that the system is on at time t. Then if $\mathrm{E}(X_1+Y_1)<\infty$ and X_1+Y_1 is not lattice, then $\lim_{t\to\infty} p(t)=\frac{\mathrm{E}(X_1)}{\mathrm{E}(X_1)+\mathrm{E}(Y_1)}$.

Proof: Let $Z(t)=I\{\text{system is on at time } t\}$, then

$$p(t)=\Pr[Z(t)=1]=\mathrm{E}[Z(t)]=\mathrm{E}\{\mathrm{E}[Z(t)\mid X_1+Y_1]\}=\mathrm{E}\{\Pr[Z(t)=1\mid X_1+Y_1]\}.$$

We also have

$$\Pr[Z(t)=1\mid X_1+Y_1=x]=\begin{cases} p(t-x) & \text{if } x\leq t \\ \Pr(X_1>t\mid X_1+Y_1=x) & \text{if } x>t\end{cases}$$

If we denote the c.d.f. of X_1+Y_1 as H, then

$$p(t)=\mathrm{E}\{\Pr[Z(t)=1\mid X_1+Y_1]\}=\int_0^t p(t-x)dH(x)+\int_t^\infty \Pr(X_1>t\mid X_1+Y_1=x)dH(x).$$
$$\equiv\int_0^t p(t-x)dH(x)+h(t).$$

If X_1+Y_1 is non-lattice and if $h(t)$ satisfy the three conditions, then we can write

$$\lim_{t\to\infty} p(t)=\frac{1}{\mu}\int_0^\infty h(t)dt, \quad \text{<By Theorem 5.10>}$$

where $\mu=\mathrm{E}(X_1)+\mathrm{E}(Y_1)$.

Since $\Pr(X_1>t\mid X_1+Y_1=x)=0$ when $0<x<t$ implies $\int_0^t \Pr(X_1>t\mid X_1+Y_1=x)dH(x)=0$. We can manipulate the $h(t)$ in the following way:

$$h(t)=\int_t^\infty \Pr(X_1>t\mid X_1+Y_1=x)dH(x)+\int_0^t \Pr(X_1>t\mid X_1+Y_1=x)dH(x)$$
$$=\int_0^\infty \Pr(X_1>t\mid X_1+Y_1=x)dH(x)=\mathrm{E}[\Pr(X_1>t\mid X_1+Y_1)]=\Pr(X_1>t).$$

Therefore, the limiting probability of $p(t)$ is

$$\lim_{t\to\infty} p(t) = \tfrac{1}{\mu}\int_0^\infty h(t)dt = \tfrac{1}{\mathrm{E}(X_1)+\mathrm{E}(Y_1)}\int_0^\infty \Pr(X_1 > t)dt = \tfrac{\mathrm{E}(X_1)}{\mathrm{E}(X_1)+\mathrm{E}(Y_1)}.$$

Now let's verify if $h(t)$ satisfies the three conditions required. Clearly, $h(t) = \Pr(X_1 > t) \ge 0, \forall t$ and $h(t) = \Pr(X_1 > t)$ is nonincreasing. Furthermore, we have $\int_0^\infty h(t)dt = \mathrm{E}(X_1) < \infty$. QED

Example 5.13 Let's go back to the example we mentioned when introducing the renewal theory, a part is replaced immediately when it fails. Let $A(t)$ be the age of the part at time t, which is also the time since the last renewal at time t. Let $g(t) = \mathrm{E}[A(t)]$. Then we have

$$g(t) = \mathrm{E}\{\mathrm{E}[A(t) \mid X_1]\} = \int_0^t g(t-x)dF(x) + \int_t^\infty t dF(x)$$
$$= \int_0^t g(t-x)dF(x) + t\cdot[1-F(t)] \equiv \int_0^t g(t-x)dF(x) + h(t)$$

Assume that X_1 is non-lattice and $h(t)$ satisfies the three properties, then we can apply the **Theorem 5.10** to write

$$\lim_{t\to\infty} g(t) = \tfrac{1}{\mathrm{E}(X_1)}\int_0^\infty h(t)dt = \tfrac{1}{\mathrm{E}(X_1)}\int_0^\infty t\cdot \Pr(X_1 > t)dt = \tfrac{1}{\mathrm{E}(X_1)}\cdot\tfrac{\mathrm{E}(X_1^2)}{2} = \tfrac{\mathrm{E}(X_1^2)}{2\,\mathrm{E}(X_1)}.$$

Note that the third equal sign uses the knowledge of **Comment 1.6**.

It is easy to verify that $h(t)$ satisfies the three properties to be "directly Riemann integrable."

Proposition 5.5 Consider an $M/G/1$ queue with arrival rate λ and service rate μ. If $p(t)$ is the probability the server is idle at time t, then $\lim_{t\to\infty} p(t) = 1 - \frac{\lambda}{\mu}$ if $\lambda < \mu$.

Definition 5.8 consider a renewal process. Let $A(t) = t - S_{N(t)}$, then $A(t)$ is referred to as the *age* of the renewal process at time t. Let $Y(t) = S_{N(t)+1} - t$, then $Y(t)$ is referred to as the *excess* (or *residual*) life of the renewal process at time t. Define $X(t) = X_{N(t)+1}$ then $X(t) = A(t) + Y(t)$. Note that $X(t)$ does not generally have the same distribution as X_1.

Proposition 5.6 If the interevent distribution of a renewal process is not lattice and with mean $\mu < \infty$ then $\lim_{t\to\infty}\Pr[Y(t) \le x] = \lim_{t\to\infty}\Pr[A(t) \le x] = \frac{1}{\mu}\int_0^x [1-F(y)]dy$.

Proof: Let's fix x and we can prove this result using the notion of *Alternating Renewal Process* easily. Define $Z(t) \equiv I\{\text{system is on at time } t\} \equiv I\{A(t) \le x\}$. Clearly, the i^{th} span for the system to be on is $\min(x, X_i)$ and the span for the system being off could be zero instantaneously. Therefore, we have

$$\lim_{t\to\infty}\Pr[A(t) \le x] = \lim_{t\to\infty}\Pr[Z(t) = 1] = \tfrac{\mathrm{E}[\min(x,X_1)]}{\mathrm{E}(X_1)}.$$ <By the **Alternating Renewal Process**>

Morover, we have

$$\mathrm{E}[\min(x,X_1)] = \int_0^\infty \Pr[\min(x,X_1) > y]dy = \int_0^x \Pr[\min(x,X_1) > y]dy + \int_x^\infty \Pr[\min(x,X_1) > y]dy$$
$$= \int_0^x \Pr[X_1 > y]dy + 0$$ <note that $0 < y < x$ and $x > X_1$ for the first integral>
$$= \int_0^x [1-F(y)]dy.$$

Finally, we have $\lim_{t\to\infty}\Pr[A(t)\le x]=\frac{1}{\mu}\int_0^x[1-F(y)]dy$.

Since it is easy to recognize that $\{Y(t)\le x\}\Leftrightarrow\{A(t+x)\le x\}$, we have

$$\lim_{t\to\infty}\Pr[Y(t)\le x]=\lim_{t\to\infty}\Pr[A(t+x)\le x]=\lim_{t\to\infty}\Pr[A(t)\le x]=\tfrac{1}{\mu}\int_0^x[1-F(y)]dy.\text{ QED}$$

Proposition 5.7 If the interevent distribution of a renewal process is not lattice and $\mathrm{E}(X_1^2)<\infty$, then $\lim_{t\to\infty}\mathrm{E}[Y(t)]=\lim_{t\to\infty}\mathrm{E}[A(t)]=\frac{\mathrm{E}(X_1^2)}{2\mu}$.

Proof: Since $\mathrm{E}[A(t)]=\int_0^\infty\Pr[A(t)>x]dx$, we have

$$\lim_{t\to\infty}\mathrm{E}[A(t)]=\lim_{t\to\infty}\left\{\int_0^\infty\Pr[A(t)>x]dx\right\}=\int_0^\infty\lim_{t\to\infty}\{\Pr[A(t)>x]dx\}$$

<assuming that we can do this swap, which is not true under general cases.>

$$=\int_0^\infty\{1-\lim_{t\to\infty}\Pr[A(t)\le x]\}dx$$

$$=\int_0^\infty\{\tfrac{\mu}{\mu}-\tfrac{1}{\mu}\int_0^x[1-F(y)]dy\}dx \quad \text{<by the **Proposition 5.6**>}$$

$$=\int_0^\infty\{\tfrac{1}{\mu}\int_0^\infty[1-F(y)]dy-\tfrac{1}{\mu}\int_0^x[1-F(y)]dy\}dx$$

$$=\int_0^\infty\{\tfrac{1}{\mu}\int_x^\infty[1-F(y)]dy\}dx$$

$$=\tfrac{1}{\mu}\int_0^\infty\int_0^y[1-F(y)]dxdy \qquad \text{<**Fubini Theorem**; } 0\le x\le y<\infty\text{>}$$

$$=\tfrac{1}{\mu}\int_0^\infty y[1-F(y)]dy$$

$$=\tfrac{1}{\mu}\cdot\tfrac{1}{2}\cdot\int_0^\infty[1-F(y)]dy^2=\tfrac{1}{2\mu}\cdot\left\{y^2[1-F(y)]\,\big|_0^\infty-\int_0^\infty y^2d[1-F(y)]\right\}$$

$$=\tfrac{1}{2\mu}\cdot\left[0-0+\int_0^\infty y^2dF(y)\right] \qquad \text{<if } m \text{ has finite variance, then } \lim_{m\to\infty}m^2[1-F(m)]=0\text{>}$$

$$=\tfrac{\mathrm{E}(X^2)}{2\,\mathrm{E}(X)}.\text{ QED}$$

Proposition 5.8 If the interevent distribution of a renewal process is not lattice and $\mu<\infty$ then $\lim_{t\to\infty}\Pr[X(t)\le x]=\frac{1}{\mu}\int_0^x ydF(y)$ and $\lim_{t\to\infty}\mathrm{E}[X(t)]=\frac{\mathrm{E}(X_1^2)}{\mu}$.

Proof: Define $Z(t)=I\{\text{system is on at time } t\}=\{X(t)\le x\}$ and note that we have zero length of on-off cycles. Then by the notion of *Alternating Renewal Process*, we have

$$\lim_{t\to\infty}\Pr[X(t)\le x]=\mathrm{E}(\text{on time in a clycle})\,/\,\mathrm{E}(\text{length of a cycle})=\mathrm{E}\{X\cdot I(X\le x)\}\,/\,\mathrm{E}(X)=\tfrac{1}{\mu}\int_0^x ydF(y),$$

where the indicator function rips off the other half of the integral. By using the result from **Proposition 5.7**, we have $\lim_{t\to\infty}\mathrm{E}[X(t)]=\lim_{t\to\infty}\mathrm{E}[A(t)]+\lim_{t\to\infty}\mathrm{E}[Y(t)]=\frac{\mathrm{E}(X_1^2)}{\mu}$. Also note that $\mathrm{E}[X(t)]\to\frac{\mathrm{E}(X^2)}{\mathrm{E}(X)}\ge\mathrm{E}(X)$, reflecting the essence of the "inspection paradox."

Example 5.14 Let $A(t)$ be the age at time t for a renewal process where $X_i\sim F$ with p.d.f. f . $X(t)$ is the length of the cycle in progress at time t . Compute $\lim_{t\to\infty}\Pr[\frac{A(t)}{X(t)}\le x]$ for fixed $x\in(0,1)$.

Define $g(t)=\Pr[\frac{A(t)}{X(t)}\le x]$. Then we can write

$$\begin{aligned}
g(t) &= \mathrm{E}\left\{\Pr\left[\tfrac{A(t)}{X(t)} \le x \mid X_1\right]\right\} \\
&= \int_0^t \Pr\left[\tfrac{A(t)}{X(t)} \le x \mid X_1 = u\right]\cdot \mathrm{f}(u)du + \Pr\left[\tfrac{A(t)}{X(t)} \le x \mid X_1 > t\right]\cdot \Pr(X_1 > t) \\
&= \int_0^t \Pr\left[\tfrac{A(t-u)}{X(t-u)} \le x\right]\cdot \mathrm{f}(u)du + \Pr\left[\tfrac{A(t)}{X(t)} \le x, X_1 > t\right] \\
&= \int_0^t g(t-u)\cdot \mathrm{f}(u)du + \Pr\left[X_1 \ge \tfrac{t}{x}, X_1 > t\right] \qquad \langle A(t) = t \text{ and } X(t) = X_1 \text{ if } X_1 > t\rangle \\
&\equiv \int_0^t g(t-u)\cdot \mathrm{f}(u)du + h(t).
\end{aligned}$$

Since X_i 's are non-lattice, by the renewal theorem, we have $\lim_{t\to\infty} g(t) = \frac{1}{\mathrm{E}(X_1)}\int_0^\infty h(t)dt$. Moreover, we have

$$h(t) = \Pr\left[X_1 \ge \tfrac{t}{x}, X_1 > t\right] = \Pr\left[X_1 \ge \tfrac{t}{x}\right].$$

Therefore, we have

$$\lim_{t\to\infty} g(t) = \tfrac{1}{\mathrm{E}(X_1)}\int_0^\infty h(t)dt = \tfrac{1}{\mathrm{E}(X_1)}\int_0^\infty \Pr\left[X_1 \ge \tfrac{t}{x}\right]dt = \tfrac{1}{\mathrm{E}(X_1)}\int_0^\infty \Pr\left[x\cdot X_1 \ge t\right]dt = \tfrac{\mathrm{E}(x\cdot X_1)}{\mathrm{E}(X_1)} = x.$$

That is, $\lim_{t\to\infty} \Pr\left[\frac{A(t)}{X(t)} \le x\right] = x$ or $\frac{A(t)}{X(t)} \Rightarrow U(0,1)$.

Definition 5.9 Let $\{X_n : n \ge 1\}$ be a sequence of non-negative independent r.v.s. Suppose X_1 has c.d.f G and $X_i, i \ge 2$, has c.d.f. F . Then the counting process with these r.v.s. as interevent times, $\{N_D(t) : t \ge 0\}$, is a *delayed renewal process*.

Proposition 5.9 Let $\{N_D(t) : t \ge 0\}$ be a delayed renewal process, then we have the following results:
(1) $\frac{N_D(t)}{t} \to \frac{1}{\mu}$ a.s. as $t \to \infty$; (2) $\frac{m_D(t)}{t} \to \frac{1}{\mu}$ a.s. as $t \to \infty$;
(3) if F is not lattice, then $m_D(t+a) - m_D(t) \to \frac{a}{\mu}$ a.s. as $t \to \infty$;
(4) if F and G are lattice with the same period d then $\mathrm{E}(\text{number of renewals at } nd) \to \frac{d}{\mu}$;
a.s. as $n \to \infty$;
(5) if F is not lattice, $\mu < \infty$, and h satisfies the three properties of "directly Riemann integrable," then $\lim_{t\to\infty} \int_0^t h(t-x)dm_D(x) = \frac{1}{\mu}\int_0^\infty h(t)dt$.

Definition 5.10 Let $\{X_n : n \ge 1\}$ be a sequence of non-negative i.i.d. r.v.s. forming the interevent times of a renewal process. Suppose that a reward Y_n is earned at the time of the n^{th} renewal where $\{Y_n : n \ge 1\}$ is a sequence of i.i.d. r.v.s. We assume that (X_i, Y_i) are jointly i.i.d. but Y_i can depend on X_i . Then $\{(X_i, Y_i) : i \ge 1\}$ is a *renewal-reward process*.

Theorem 5.12 Let $\{(X_i, Y_i) : i \ge 1\}$ be a renewal-reward process with $\mathrm{E}(X_1) < \infty$ and $\mathrm{E}(Y_1) < \infty$. Let $Y(t) = \sum_{i=1}^{N(t)} Y_i$ be the total reward earned up to time t . Then $\lim_{t\to\infty} \frac{Y(t)}{t} = \frac{\mathrm{E}(Y_1)}{\mathrm{E}(X_1)}$, and $\lim_{t\to\infty} \frac{\mathrm{E}[Y(t)]}{t} = \frac{\mathrm{E}(Y_1)}{\mathrm{E}(X_1)}$ a.s. .

Proof: $\lim_{t\to\infty} \frac{Y(t)}{t} = \lim_{t\to\infty} \frac{\sum_{i=1}^{N(t)} Y_i}{N(t)} \cdot \frac{N(t)}{t} = \mathrm{E}(Y)\cdot \frac{1}{\mathrm{E}(X)} = \frac{\mathrm{E}(Y)}{\mathrm{E}(X)}$ a.s. Note that we used the *Strong Law of Large Numbers* to get the second equal sign above. Next, we have

$$\lim_{t\to\infty} \tfrac{\mathrm{E}[Y(t)]}{t} = \lim_{t\to\infty} \tfrac{\mathrm{E}\left[\sum_{i=1}^{N(t)} Y_i\right]}{t} = \lim_{t\to\infty}\left[\tfrac{\mathrm{E}\left[\sum_{i=1}^{N(t)+1} Y_i\right]}{t} - \tfrac{\mathrm{E}\, Y_{N(t)+1}}{t}\right].$$

We can show that $\lim_{t\to\infty} \frac{E[Y_{N(t)+1}]}{t} = 0$, a proof of which was given in the textbook. We know $N(t)+1$ is a stopping time for $\{X_i : i \geq 1\}$ and $N(t)+1$ is a generalized stopping time for $\{Y_i : i \geq 1\}$ and thus we can use the *Wald's Equation* to write:

$$\lim_{t\to\infty} \frac{E\left[\sum_{i=1}^{N(t)+1} Y_i\right]}{t} = \lim_{t\to\infty} \frac{m(t)+1}{t} \cdot E(Y_i) = \frac{1}{\mu} \cdot E(Y_i).$$ <by the **Elementary Renewal Theorem**>

Note that the limits in this theorem don't change if the reward is earned continuously, instead of at the end of each cycle.

Example 5.15 Suppose that the cumulative demand for a product through time t, $D(t)$, gives a renewal process $\{D(t) : t \geq 0\}$ with i.i.d. interdemand times $\tau_1, \tau_2, ...$ with $E(\tau_1) = \frac{1}{d}$. Suppose that the lead times are 0, so that we reorder whenever inventory hits 0. Suppose further that we order the deterministic quantity Q units. Each order carries a fixed cost $\$K$ and holding costs are $\$h$ per unit per unit time. The objective is to minimize the long-run average cost.

Let Y_i be the cost occurred over the i^{th} cycle and X_i be the length of the i^{th} cycle. Denote the long-run average cost as AC. Then we have $AC = \lim_{t\to\infty} \frac{\sum_{i=1}^{N(t)} Y_i}{t} = \frac{E(Y_1)}{E(X_1)}$. Clearly, $E(X_1) = E\left[\sum_{i=1}^{Q} \tau_i\right] = Q \cdot E(\tau_i) = \frac{Q}{d}$ and $Y_1 = K + h \cdot Q \cdot \tau_1 + h \cdot (Q-1) \cdot \tau_2 + ... + h \cdot \tau_Q$, which implies that

$$E(Y_1) = K + \frac{h \cdot Q}{d} + \frac{h(Q-1)}{d} + ... + \frac{h}{d} = K + \frac{h}{d} \cdot \frac{(1+Q) \cdot Q}{2}.$$

Therefore, we have $AC = E(Y_1) / E(X_1) = [K + \frac{h}{2d} \cdot Q \cdot (Q+1)] / (Q / d) = \frac{Kd}{Q} + \frac{h}{2} \cdot (Q+1)$. And finally $Q^* = \arg\min_Q \{\frac{Kd}{Q} + \frac{h}{2} \cdot (Q+1)\} = \sqrt{\frac{2Kd}{h}}$.

Definition 5.11 Consider a stochastic process $\{X(t) : t \geq 0\}$ with state space $\{0,1,2,...\}$ having the property that there exist time points at which the process (probabilistically) restarts itself. That is, suppose that with probability 1 there exists a time S_1 such that the process beyong S_1 is probabilitistic replica of the whole process starting at 0. Such a stochastic process is known as a *regenerative process*.

Theorem 5.13 Let $\{X(t) : t \geq 0\}$ be a regenerative process. Then $\{S_1, S_2, ...\}$ constitute the events of a renewal process. We say that a cycle is completed every time a renewal occurs. Let $N(t) = \max(n : S_n \leq t)$ denote the number of renewals by time t. If F, the distribution of a cycle, has a density over some interval, and if $E(S_1) < \infty$, then $\lim_{t\to\infty} \Pr[X(t) = j] = \frac{\text{E(amount of time in state } j \text{ during a cycle)}}{\text{E(time of a cycle)}}$.

Proposition 5.10 For a regenerative process with $E(S_1) < \infty$, we have

$$\lim_{t\to\infty} \frac{[\text{amount of time in state } j \text{ during } (0,t)]}{t} = \lim_{t\to\infty} \frac{\text{E[amount of time in state } j \text{ during } (0,t)]}{t} = \frac{\text{E[amount of time in state } j \text{ during a cycle]}}{\text{E(time of a cycle)}}.$$

Proof: Think of a renewal-reward process. Let Y_i be the reward in the i^{th} cycle, i.e., the amount of time spent in state j. Let X_i be the length of the i^{th} cycle. Then the renewal reward theorem tells us that

$$\frac{Y(t)}{t} \to \frac{E(Y_1)}{E(X_1)} \text{ a.s. and } \frac{E[Y(t)]}{t} \to \frac{E(Y_1)}{E(X_1)} \text{ a.s. QED}$$

CHAPTER 6 MARKOV CHAINS

Definition 6.1 A Markov process is a stochastic process $\{X(t): t \in T\}$ such that for all $t_1 < t_2 < ... < t_n < t_{n+1} \in T$, $\Pr[X(t_{n+1}) \le x \mid X(t_1) = x_1,..., X(t_n) = x_n] = \Pr[X(t_{n+1}) \le x \mid X(t_n) = x_n]$, for all $x_1, x_2,..., x_n, x$. Note the property above is called the *Markov property* and in effect says that given the present, the future is independent of the past.

Theorem 6.1 Every stochastic process with independent increments and $X(0) = 0$ is a Markov process.

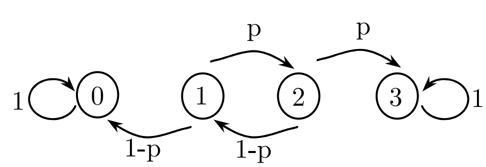
Comment 6.1 Note that in the universe of stochastic process, we consider two subsets, one is Renewal Process and the other is Markov Chain. Any stochastic process with independent increments belongs to Markov Chain, and the set of Market Chain intersects with the set of Renewal Process at a special point, Poisson Process. Note that here we relax the assumption that the inter-renewal times X_i being i.i.d. and restrict them in a way that they satisfy the Markov property.

Definition 6.2 A discrete time Markov process $\{X_n : n \ge 1\}$ is a (discrete time) *Markov chain* if the state space of all possible values of the process, $\mathcal{S}$, is discrete (i.e., finite or countably infinite).

Definition 6.3 If $\Pr(X_{n+1} = j \mid X_n = i)$ does not depend on n, we say that the Markov chain is *time homogeneous*. We will assume time homogeneity for all future chains. What this property says is that the transition probabilities won't be updated (or learned) through time.

Definition 6.4 Let $p_{ij} = \Pr(X_{n+1} = j \mid X_n = i), i, j = 0,1,2,...$. Then p_{ij} is said to be the 1-step transition (or jump) probability from state i to state j. The p_{ij}'s are often written as a single matrix P where the $(i, j)^{th}$ element of P is p_{ij}. P is called the 1-step probability transition matrix.

Definition 6.5 The 1-step transition matrix is often drawn pictorially as a *transition diagram*. For example, we can draw the following transition diagram in the Gambler's Ruin.



In this diagram the nodes represent the different states and the arcs are labeled with the transition probabilities between those states.

Definition 6.6 Let $a_i = \Pr[X(0) = i], i = 0,1,....$ The initial probability vector is $a = (a_0\ a_1\ a_2\ ...)$.

Theorem 6.2 Together a and P completely determine the distribution of the Markov chain.

Proof:

$$\begin{aligned}
&\Pr[X_0 = i_0, X_1 = i_1, ..., X_n = i_n] \text{ <where } \forall i_k, k = 1,...,n \text{ are deterministic>} \\
&= \Pr[X_n = i_n \mid X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i_{n-1}] \cdot \Pr[X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i_{n-1}] \\
&= \Pr[X_n = i_n \mid X_{n-1} = i_{n-1}] \cdot \Pr[X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i_{n-1}] \\
&= p_{i_{n-1}i_n} \cdots p_{i_0,i_1} \cdot a \text{ QED}
\end{aligned}$$

Definition 6.7 $p_{ij}^{(n)} = \Pr[X_n = j \mid X_0 = i] = \Pr[X_{m+n} = j \mid X_m = i]$ is the probability that if a Markov chain starts in state i then after n transitions it will be in state j. This is called the *n-step* or n^{th} *-order transition probability*. Just like the 1-step transitions, these n-step transitions are often written in matrix form as $P^{(n)}$.

Proposition 6.1 The n-step transition matrix may be found by multiplying P (the 1-step transition matrix) by itself n times, i.e., $P^{(n)} = P^n$. Notice that this implies that for any m and n, $P^{m+n} = P^m P^n$, which leads to the *Chapman-Kolmogorov equation* for Markov chains with state space $\mathcal{S}$,
$p_{ij}^{(n+m)} = \sum_{k\in\mathcal{S}} p_{ik}^{(n)} p_{kj}^{(m)}$, for any non-negative integer m and n and all state $k \in \mathcal{S}$.
Proof:

$$\begin{aligned}
p_{ij}^{(n+m)} &= \Pr[X_{n+m} = j \mid X_0 = i] \\
&= \sum_{k\in\mathbf{S}} \Pr[X_{n+m} = j \mid X_0 = i, X_n = k] \cdot \Pr[X_n = k \mid X_0 = i] \\
&= \sum_{k\in\mathbf{S}} \Pr[X_{n+m} = j \mid X_n = k] \cdot \Pr[X_n = k \mid X_0 = i] \\
&= \sum_{k\in\mathbf{S}} p_{kj}^{(m)} \cdot p_{ik}^{(n)} = [P^{(n)} P^{(m)}]_{ij}, \text{ i.e. } , P^{n+m} = P^n P^m. \text{ QED}
\end{aligned}$$

Proposition 6.2 The probability that the Markov chain is in state j after n transitions is $a_j^{(n)} = \sum_{k\in\mathcal{S}} p_{kj}^{(n)} a_k$, and the vector of the state probabilities $a^{(n)}$ may be found by $a^{(n)} = aP^n$.
Proof: Let $a_j = \Pr[X_0 = j]$. We have

$$a_j^{(n)} = \Pr[X_n = j] = \sum_{k\in\mathcal{S}} \Pr[X_n = j \mid X_0 = k] \cdot \Pr(X_0 = k) = \sum_{k\in\mathcal{S}} p_{kj}^{(n)} \cdot a_k = [aP^{(n)}]_j,$$

or $a^{(n)} = aP^{(n)} = aP^n$, where $P^{(n)} = P^n$ follows immediately from **Proposition 6.1**.

Definition 6.8 State j is said to be *accessible* from state i, written $i \to j$, if $p_{ij}^{(n)} > 0$ for some $n \ge 0$ (where $p_{ij}^{(0)} = 1$ for $i = j$ and $= 0$ otherwise), i.e., if the chain is in state i then it is possible that at some later point it will end up in state j. (For example, in the transition diagram for **definition 6.5**, we have $1 \to 0, 1 \to 2, 1 \to 3, 1 \to 1$, but we don't have $3 \to 1$.)

Comment 6.2 Why do we have $p_{ij}^{(0)} = 1$ for $i = j$ and $= 0$ otherwise? Note that $\sum_{k\in\mathcal{S}} p_{ik}^{(n)} = 1$ for $n \ge 0$ and thus $\sum_{k\in\mathcal{S}} p_{ik}^{(0)} = 1$, which prompts us to write $p_{ii}^{(0)} = 1$ and $p_{ik}^{(0)} = 0, \forall k \ne i$.

Definition 6.9 Two states i and j are said to *communicate*, written $i \leftrightarrow j$, if they are accessible from each other. (For example, in the transition diagram for **definition 6.5**, we have $1 \leftrightarrow 2, 0 \leftrightarrow 0, 3 \leftrightarrow 3$.)

Theorem 6.3 Communication is an equivalence relation, i.e.,
(1) $i \leftrightarrow i$ (reflective); (2) $i \leftrightarrow j$ implies $j \leftrightarrow i$ (symmetric); (3) $i \leftrightarrow k$ and $k \leftrightarrow j$ imply $i \leftrightarrow j$ (transitive).

Proof: (1) Take $n = 0$, we have $p_{ii}^{(0)} = 1$ by definition, and thus $i \leftrightarrow i$.

(2) This property is obvious and we are not going to prove it here.

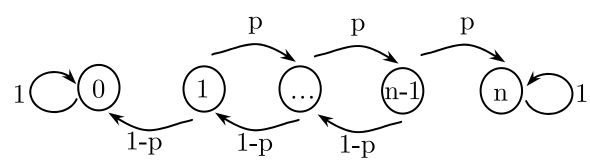
(3) $i \leftrightarrow k \Leftrightarrow \exists m_1, n_1 \ni p_{ik}^{(m_1)} > 0, p_{ki}^{(n_1)} > 0$ and $k \leftrightarrow j \Leftrightarrow \exists m_2, n_2 \ni p_{kj}^{(m_2)} > 0, p_{jk}^{(n_2)} > 0$. Clearly, we have $p_{ij}^{(m_1+m_2)} = \sum_{l \in \mathcal{S}} p_{il}^{(m_1)} p_{lj}^{(m_2)} \geq p_{ik}^{(m_1)} p_{kj}^{(m_2)} > 0$. Similarly, we have $p_{ji}^{(n_1+n_2)} > 0$ and thus we have $i \leftrightarrow j$.

Definition 6.10 We may partition the state space into mutually exclusive and exhaustive *classes* such that two states communicate if and only if they are in the same class. We do this by starting with any state i and forming the class C_i of all states that communicate with i. Then we repeat for any state not in C_i, and so on.

Definition 6.11 A Markov chain is said to be *irreducible* if all states communicate (i.e., there is only one equivalence class); otherwise, it is *reducible*. So a Markov chain is irreducible if, for all states i and j, there is an $n \geq 0$ such that $\Pr[X_n = j \mid X_0 = i] > 0$.

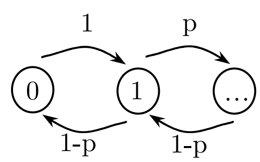
Example 6.1 In all the transition diagrams below, we assume $0 < p < 1$.

(a)



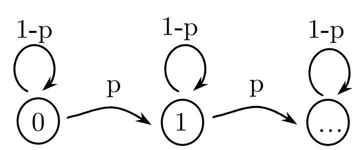
Since $C_1 = \{0\}, C_2 = \{1, 2, \ldots, n-1\}, C_3 = \{n\}$, the Markov chain above is reducible.

(b)



Since $C_1 = \{0, 1, \ldots.\}$, the Markov chain above is irreducible.

(c)

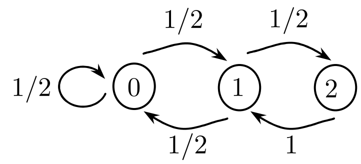


Since $C_1 = \{0\}, C_2 = \{1\}, \ldots$, the Markov chain above is reducible.

Definition 6.12 The *period* of a state j, written $d(j)$, is the greatest common divisor of all $n > 0$ such that $p_{jj}^{(n)} > 0$. If $d(j) = 1$ then state j is said to be *aperiodic*. If $p_{jj}^{(n)} = 0, \forall n > 0$, then we define $d(j) \equiv 0$.

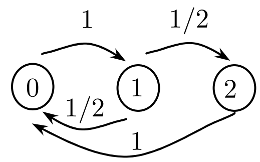
Example 6.2

(a)



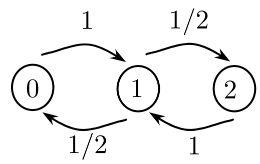
Since $p_{00}^{(n)} > 0,$ if $n \in \{1,2,3,...\}$, we have $d(0) = 1$ and thus the Markov chain above is aperiodic.

(b)



Since $p_{00}^{(n)} > 0,$ if $n \in \{2,3,...\}$, we have $d(0) = 1$ and thus the Markov chain above is aperiodic. Note that although $p_{00}^{(1)} = 0$, we still have $d(0) = 1$.

(c)



Since $p_{00}^{(n)} > 0,$ if $n \in \{2,4,6,...\}$, we have $d(0) = 2$ and thus the Markov chain above is periodic.

Definition 6.13 A property of a state is called a *class property* if either all members of any class share the property or none of the members share the property.

Proposition 6.3 If $i \leftrightarrow j$, then $d(i) = d(j)$ (i.e., periodicity is a class property).

Definition 6.14 Define the *first passage probabilities* as $\mathrm{f}_{ij}^{n} = \Pr[X_n = j, X_k \neq j, k = 1,2,...,n-1 \mid X_0 = i]$, $n = 1,2,....$ So f_{ij}^{n} is the probability that the Markov chain is in state j for the first time after n transitions, having started in state i . Define $\mathrm{f}_{ij}^{0} \equiv 0$ for $i \neq j$.

Proposition 6.4 We have $\mathrm{f}_{ij}^{n} \le p_{ij}^{(n)}, \forall i,j,$ and $n \ge 0$ as well as $p_{ij}^{(n)} = \sum_{k=1}^{n} \mathrm{f}_{ij}^{k}\, p_{jj}^{(n-k)}, \forall i,j,$ and $n \ge 1$.

Definition 6.15 Let $\mathrm{f}_{ij} = \sum_{k=1}^{\infty} \mathrm{f}_{ij}^{k}$, then f_{ij} is the probability of ever going from i to j .

Definition 6.16 State j is *recurrent* if $\mathrm{f}_{jj} = 1$ and is *transient* otherwise.

Proposition 6.5 State j is recurrent if and only if $\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$.

Proof: Let $N = \sum_{k=1}^{\infty} I(X_k = j) =$ the number of visits to state j throughout the time. Then we have

$$
\begin{aligned}
\mathrm{E}(N \mid X_0 = j) &= \mathrm{E}\left[\sum_{k=1}^{\infty} I(X_k = j) \mid X_0 = j\right] \\
&= \sum_{k=1}^{\infty} \mathrm{E}[I(X_k = j) \mid X_0 = j] \qquad \langle \text{by Fubini's Theorem;non-negativity} \rangle \\
&= \sum_{k=1}^{\infty} \Pr[X_k = j \mid X_0 = j] = \sum_{k=1}^{\infty} p_{jj}^{(k)}
\end{aligned}
$$

To prove $\Rightarrow$, we have the following reasoning:

$$\text{recurrent} \Leftrightarrow \mathrm{f}_{jj} = 1 \Leftrightarrow \text{starting from state } j\text{, always come back to state } j$$

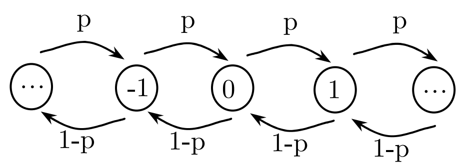
$$\Leftrightarrow (N \mid X_0 = j) = \infty, \text{ a.s. } \Rightarrow \mathrm{E}(N \mid X_0 = j) = \infty \Leftrightarrow \sum_{k=1}^{\infty} p_{jj}^{(k)} = \infty.$$

To prove $\Leftarrow$, we prove "transient $\Rightarrow \sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$". Clearly, we can write

$$\text{transient} \Leftrightarrow \mathrm{f}_{jj} < 1 \Leftrightarrow \Pr[\text{return to } j \mid \text{start in } j] = \mathrm{f}_{jj} < 1.$$

That is, the number of visits N to state j, given starting in state j, is $\text{geometric}(1-\mathrm{f}_{jj})$. Hence $\mathrm{E}(N \mid X_0 = j) = \frac{1}{1-\mathrm{f}_{jj}} < \infty$. QED

Example 6.3 One-dimension random walk ($0 < p < 1$)



It is quite intuitive to think that when $p < \frac{1}{2}$, then the chain will move towards the state $-\infty$ and thus each of the state in the chain is transient; when $p > \frac{1}{2}$, then the chain will move towards the state $+\infty$ and thus each of the state in the chain is again transient; when $p = \frac{1}{2}$, then every state is recurrent. We can prove this intuition as follows.

Let's at first consider the state 0, without loss of generality. Clearly, we have $p_{00}^{(2n+1)} = 0$ since we cannot get back to the original state 0 in an odd number of steps. What is $p_{00}^{(2n)}$ then? It is obvious that we should move n steps to the right of the state 0 and n steps to the left of the state 0, with the particular order and combination of the directions being arbitrary. Equivalently, we can try to fill $2n$ slots with n R's and n L's. There are $\binom{2n}{n}$ combinations of directions R and L, and each combination of directions has probability of $p^n(1-p)^n$. Therefore, $p_{00}^{(2n)} = \binom{2n}{n} p^n (1-p)^n$.

To determine whether state 0 is transient, we need to verify whether $\sum_{k=1}^{\infty} p_{00}^{(k)}$ is finite. It is easy to get $\sum_{k=1}^{\infty} p_{00}^{(k)} = \sum_{n=1}^{\infty} p_{00}^{(2n)} = \sum_{n=1}^{\infty} \frac{(2n)!}{(n!)^2}[p(1-p)]^n$. In order to calculate this sum, we refer to one variation of the *Stirling's Formula*: $\frac{n!}{n^{n+\frac{1}{2}} \cdot e^{-n} \cdot \sqrt{2\pi}} \to 1$ as $n \to \infty$. Hence, we have

$$\sum_{n=1}^{\infty} \frac{(2n)!}{(n!)^2}[p(1-p)]^n \approx \sum_{n=1}^{\infty} \frac{(2n)^{2n+\frac{1}{2}} \cdot e^{-2n} \cdot \sqrt{2\pi}}{n^{2n+1} \cdot e^{-2n} \cdot 2\pi}[p(1-p)]^n = \sum_{n=1}^{\infty} \frac{(2n)^{2n+\frac{1}{2}}}{\sqrt{2\pi n}}[p(1-p)]^n = \sum_{n=1}^{\infty} \frac{[4p(1-p)]^n}{\sqrt{\pi n}}.$$

Because $4p(1-p) < 1$ when $p \neq \frac{1}{2}$ and $4p(1-p) = 1$ when $p = \frac{1}{2}$, we know that

$$\sum_{n=1}^{\infty} \frac{[4p(1-p)]^n}{\sqrt{\pi n}} \begin{cases} = \infty & \text{if } p = \frac{1}{2} \qquad \text{i.e., state 0 is recurrent;} \\ < \infty & \text{if } p \neq \frac{1}{2} \qquad \text{i.e., state 0 is transient.} \end{cases}$$

To see why $\sum_{n=1}^{\infty} \frac{[4p(1-p)]^n}{\sqrt{\pi n}} < \infty$ if $p \neq \frac{1}{2}$, we can write

$$\sum_{n=1}^{\infty} \frac{[4p(1-p)]^n}{\sqrt{\pi n}} < \sum_{n=1}^{\infty} \frac{[4p(1-p)]^n}{\sqrt{\pi}} = \frac{4p(1-p)}{\sqrt{\pi}[1-4p(1-p)]} < \infty.$$

The argument can be easily applied to any other states and get the same conclusion. QED

Comment 6.3 We can extend the result above to random walks of higher dimensions. For example, for the case of two-dimensional random walk, we find that if the probabilities of going up or down or left or right are equal to $\frac{1}{4}$, then each of the state is recurrent; otherwise, each of the state is transient. However, for the case of three-dimensional or higher-dimensional random walk, no matter how we tweak the probabilities, each state is transient.

Proposition 6.6 If i is recurrent and $i \leftrightarrow j$, then j is recurrent (i.e., recurrence is a class property).
Proof: We can use the argument of geometric trial to prove it. The proof is skipped here.

Proposition 6.7 If $i \leftrightarrow j$ and i is recurrent then $\mathrm{f}_{ij} = 1$.
Intuition: By **Proposition 6.6**, we know that $i \leftrightarrow j$ and i, j are both recurrent, so that the number of visits to state j, started in state i, must be infinite, i.e., $\mathrm{f}_{ij} = 1$.

Proposition 6.8 If state j is transient then $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all initial states i.
Intuition: This result is a bit stronger than part (2) of the proof for **Proposition 6.5**.

Corollary 6.1 If state j is transient then $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ for all initial states i.
Intution: This corollary is a direct implication of **Proposition 6.8** because of the following theorem: If $\sum_{n=1}^{\infty} a_n \to s < \infty$ then $\lim_{n\to\infty} a_n = 0$. That is, $\lim_{n\to\infty} a_n = 0$ is a necessary (but not sufficient) condition for the convergence of an infinite series.

Definition 6.17 Let μ_{jj} be the expected number of transitions needed to return to state j from state j, so that $\mu_{jj} = \infty$ if state j is transient and $\mu_{jj} = \sum_{n=1}^{\infty} n\,\mathrm{f}_{jj}^{n}$ if state j is recurrent. Note that μ_{jj} is also referred to as the mean recurrence time for state j.

Theorem 6.4 Let $N_j(t)$ be the number of transitions into state j by time t. If $i \leftrightarrow j$ then $\Pr\left[\lim_{t\to\infty} \frac{N_j(t)}{t} = \frac{1}{\mu_{jj}} \mid X_0 = i\right] = 1$.
Intuition: Clearly $N_j(t) = \sum_{k=1}^{t} I(X_k = j)$ and each time when state j is reached, we consider a renewal occurs. So we can use the results from the renewal theory to reach the conclusion that $\frac{N_j(t)}{t} \to \frac{1}{\mu_{jj}}$ a.s.

Theorem 6.5 If $i \leftrightarrow j$ then $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{n} p_{ij}^{(k)} = \frac{1}{\mu_{jj}}$.
Intuition: This result is very similar to $\lim_{t\to\infty} \frac{m(t)}{t} \to \frac{1}{\mu}$ in the renewal theory.

Theorem 6.6 We have $\lim_{n\to\infty} p_{jj}^{[nd(j)]} = \frac{d(j)}{\mu_{jj}}$ for all cases. In particular, if $i \leftrightarrow j$ and state j is aperiodic, then $\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\mu_{jj}}$.

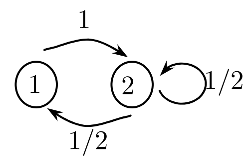
Intuition: (for the aperiodic case only) Let's define that a renewal occurs whenever the state j is visited. Note that in Markov chain, we don't have the case where two renewals occurring in the same time. Instead, the state j is either visisted or not. Hence, we can write

$$p_{jj}^{(n)} = \Pr(\text{renewal at time } n | \text{renew at time } 0) = \mathrm{E}(\# \text{ of renewals through time } n | \text{renew at time } 0).$$

Furthermore, when state j is aperiodic, the Markov chain (in discrete time) is clearly a lattice with $d = 1$, so we can use the result (4) in **Proposition 5.9** to get the conclusion $\lim_{n\to\infty} p_{jj}^{n} \to \frac{1}{\mu_{jj}}$.

For the second part of the theorem, we consider the part of the chain jumping from state i to state j for the first time a delay, after which we have a standard renewal as defined above. So we can use the results for delayed renewal to get the conclusion in the second part.

Example 6.4 Let the transition matrix be $P = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. We can draw the transition diagram as follows.



It is easy to see that there is only one communiting class, $\{1,2\}$, and thus the Markov chain above is irreducible. Since $p_{22}^{(n)} > 0$ for $n = \{1,2,...\}$, we have $d(2) = 1$ and thus $d(2) = d(1) = 1$. The Markov chain is aperiodic. Let's next find out the mean recurrence time μ_{11}. By definition, we have

$$\mu_{11} = 2\cdot 1\cdot \tfrac{1}{2} + 3\cdot 1\cdot \tfrac{1}{2}\cdot \tfrac{1}{2} + 4\cdot 1\cdot \tfrac{1}{2}\cdot \tfrac{1}{2}\cdot \tfrac{1}{2} + ..., \text{ or } 2\mu_{11} = 2 + 3\cdot \tfrac{1}{2} + 4\cdot (\tfrac{1}{2})^2 + ..., \text{ or}$$

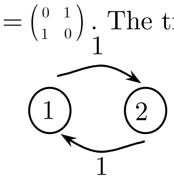
$$2\mu_{11} - \mu_{11} = 2 + \tfrac{1}{2} + (\tfrac{1}{2})^2 + ... = 2 + \tfrac{1}{2}/(1 - \tfrac{1}{2}) = 3.$$

By **Theorem 6.6**, we have $p_{11}^{(n)} \to \frac{1}{\mu_{11}} = \frac{1}{3}$ and $p_{21}^{(n)} \to \frac{1}{\mu_{11}} = \frac{1}{3}$. Hence $p_{12}^{(n)} \to \frac{2}{3}$ and $p_{22}^{(n)} \to \frac{2}{3}$. That is, $P^n \to \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$. If we were to use *first-step analysis* mentioned in **example 6.9**, we can write

$$\mu_{11} = (1 + \mu_{21})\cdot 1 \text{ and } \mu_{21} = (1 + \mu_{21})\cdot \tfrac{1}{2} + 1\cdot \tfrac{1}{2} \Leftrightarrow \mu_{21} = 2 \text{ and } \mu_{11} = 3;$$

$$\mu_{22} = 1\cdot \tfrac{1}{2} + (1 + \mu_{12})\cdot \tfrac{1}{2} \text{ and } \mu_{12} = 1\cdot 1 = 1 \Leftrightarrow \mu_{12} = 1 \text{ and } \mu_{22} = \tfrac{3}{2}.$$

Example 6.5 Let the transition matrix be $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The transition diagram is drawn as follows.



There is only one communicating class, $\{1,2\}$. Since $p_{11}^{(n)} > 0$ for $n = \{2,4,...\}$, we have $d(1) = d(2) = 2$ and thus the Markov chain above is periodic. Since after any even number of steps, the chain returns back to the original state, we have $P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Similarly, after any odd number of steps, the chain visits the other state, not the original state, we have $P^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Hence, we can write $p_{ii}^{(2n)} = 1, i = 1,2$ and the **Theorem 6.6** implies that $\lim_{n\to\infty} p_{ii}^{(2n)} = 1 = \frac{d(i)}{\mu_{ii}} = \frac{2}{\mu_{ii}}$, i.e., $\mu_{ii} = 2, i = 1,2$. Note that because of the alternating nature of the chain, $\lim_{n\to\infty} p_{ii}^{(n)}$ doesn't exists, and thus we need the period $d(i)$ sits on the top as steps in the first part of the **Theorem 6.6**.

Definition 6.18 Define the limiting probability of state j as $\pi_j \equiv \lim_{n\to\infty} p_{jj}^{[nd(j)]} = \frac{d(j)}{\mu_{jj}}$. [Note that in **Example 6.4**, we have $\pi_1 = \frac{1}{3}$ and $\pi_2 = \frac{2}{3}$ and in **Example 6.5**, we have $\pi_1 = 1$ and $\pi_2 = 1$. Why in the second case we don't have $\pi_1 + \pi_2 = 1$ as we do in the first example? It is because the periodicity of the second Markov chain messes up the limiting probability.]

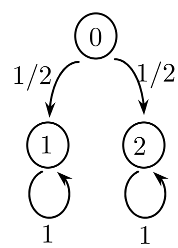
Definition 6.19 A recurrent state j is said to be *positive recurrent* if $\pi_j > 0$ and *null recurrent* if $\pi_j = 0$.

Example 6.6 In **Example 6.3**, we have $p_{jj}^{(2n)} = \frac{(2n)!}{(n!)^2}[p(1-p)]^n \approx \frac{[4p(1-p)]^n}{\sqrt{\pi n}}$. Since $4p(1-p) \in [0,1]$, we have $\lim_{n\to\infty} \frac{[4p(1-p)]^n}{\sqrt{\pi n}} = 0$, i.e., $\pi_j = 0$ and $\mu_{jj} = \infty$. So when $p = \frac{1}{2}$, state j is called *null recurrent*.

Comment 6.4 For an aperiodic Markov chain, we have the following table of results.

State i	State j	$\lim_{n\to\infty} p_{ij}^{(n)}$
Transient	Transient	0
Recurrent	Transient	0
Transient	Recurrent	f_{ij} / μ_{jj}
Recurrent	Recurrent	$\frac{1}{\mu_{jj}}$ if $i \leftrightarrow j$; and 0 if i $\not\leftrightarrow$ j.

When determining the results above, keep in mind the following special case, where state 0 is transient and state 1 and 2 are recurrent. Note that a null recurrent state can never happen in a finite state space.



Proposition 6.9 Positive (or null) recurrence is a class property.

Definition 6.20 A positive recurrent, aperiodic state is called *ergodic*.

Definition 6.21 An irreducible aperiodic positive recurrent Markov chain is an *ergodic Markov chain*.

Definition 6.22 A probability distribution $\hat{\pi} = (\hat{\pi}_0 \ \hat{\pi}_1 \ ...)$ is said to be a *stationary distribution* if:
(1) for all state $j \in \mathcal{S}$, $\hat{\pi}_j = \sum_{i\in\mathcal{S}} \hat{\pi}_i p_{ij}$; (2) $\sum_{i\in\mathcal{S}} \hat{\pi}_i = 1$. That is, $\hat{\pi}$ satisfies $\hat{\pi} = \hat{\pi}P$ and (2).

Example 6.7 In our previous example with $P = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, we have $\pi_1 = \frac{1}{3}, \pi_2 = \frac{2}{3}, d(1) = d(2) = 1$, and thus the chain is ergodic. We could have solve a system of two equations satisfying the properties in **Definition 6.22** for $\hat{\pi}_1$ and $\hat{\pi}_2$. That is, from $\hat{\pi} = \hat{\pi}P$, we have $\hat{\pi}_1 = \hat{\pi}_2 \cdot \frac{1}{2}$ and $\hat{\pi}_2 = \hat{\pi}_1 + \hat{\pi}_2 \cdot \frac{1}{2}$; from $\sum_{i=1,2} \hat{\pi}_i = 1$, we have $\hat{\pi}_1 + \hat{\pi}_2 = 1$. It is no surprise to find some redundant equations from these two requirements, and we can solve for $\hat{\pi}_1 = \frac{1}{3}, \hat{\pi}_2 = \frac{2}{3}$. In our previous example with $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\hat{\pi} = \hat{\pi}P$ implies that $\hat{\pi}_1 = \hat{\pi}_2$ and $\hat{\pi}_1 + \hat{\pi}_2 = 1$. We can thus solve for $\hat{\pi}_1 = \hat{\pi}_2 = \frac{1}{2}$.
In the Gambler's ruin, we have the following transition matix

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

From $\hat{\pi} = \hat{\pi}P$, we have $\hat{\pi}_1 = \hat{\pi}_1 + (1-p)\hat{\pi}_2 \Rightarrow \hat{\pi}_2 = 0$, $\hat{\pi}_2 = (1-p)\hat{\pi}_3 \Rightarrow \hat{\pi}_3 = 0$, $\hat{\pi}_3 = p\hat{\pi}_2$, and $\hat{\pi}_4 = p\hat{\pi}_3 + \hat{\pi}_4 \Rightarrow \hat{\pi}_3 = 0$. From $\sum_{i \in \mathcal{S}} \hat{\pi}_i = 1$, we have $\hat{\pi}_1 + \hat{\pi}_2 + \hat{\pi}_3 + \hat{\pi}_4 = 1 \Rightarrow \hat{\pi}_1 + \hat{\pi}_4 = 1$. That is, we have infinite amount of stationary distributions in this Gambler's ruin.

Theorem 6.7 If a Markov chain is ergodic then the limiting probabilities $(\pi_0 \ \pi_1 \ ...)$ form a stationary distribution and there are no other stationary distributions.

Example 6.8 Let the transition matrix be

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

Clearly, there is one communicating class containing all three states and the Markov chain is irreducible. Since the chain can reach state 1 starting from state 1 with one step, we have $p_{11}^{(n)} > 0$ for $n = \{1,2,3,...\}$ and thus $d(1) = d(2) = d(3) = 1$. The Markov chain is aperiodic and all states are recurrent. Since we cannot have $\mu_{jj} = \infty$, for recurrent states $j = 1,2,3$, when having only finite state spaces, so all states are positive recurrent. Therefore, we have an ergodic Markov chain and by **Theorem 6.7**, we can get the limiting probabilities from sationary ones. That is, we can solve $\hat{\pi} = \hat{\pi}P$ and $\hat{\pi}_1 + \hat{\pi}_2 + \hat{\pi}_3 = 1$ for the limiting probabilities.

Note that all rows and colums of matrix P add up to 1 and this type of stochastic matrix is called "*Doubly Stochastic Matrix*." One of special properties of the doubly stochastic matrix is that the stationary probabilities are equal across all states for an ergodic Markov chain. That is, we have $\hat{\pi}_1 = \hat{\pi}_2 = \hat{\pi}_3 = \frac{1}{3}$. By **Theorem 6.7**, we have $\pi_j = \hat{\pi}_j, j = 1,2,3$ and thus $P_{ij}^n = \frac{1}{3}, \forall i, j = 1,2,3$.

Theorem 6.8 If an irreducible aperiodic Markov chain is not positive recurrent, then no stationary distribution exists. (In this case, we have $p_{jj}^n \to 0, \forall j \in \mathcal{S}$.)

Example 6.9 Let $X = \{X_n : n \geq 0\}$ be an irreducible Markov chain on a finite state space $\mathcal{S}$. Suppose that we want to compute $u(x) = \mathrm{E}[T_A \mid X_0 = x]$, where $T_A = \inf\{n \geq 0 : X_n \in A\}$ is the first hitting time to the set A, and $x \in A^c$. We can apply *first-step analysis* to solve this problem. *First-step analysis* is another name for conditioning on the first state one enters upon leaving x. So we can write

$$u(x) = \mathrm{E}(T_A \mid X_0 = x) = \sum_{y \in \mathcal{S}} \mathrm{E}(T_A \mid X_1 = y, X_0 = x) \cdot \Pr(X_1 = y \mid X_0 = x)$$
$$= \sum_{y \in A} 1 \cdot p_{xy} + \sum_{y \in A^c} [1 + u(y)] \cdot p_{xy} = \sum_{y \in \mathcal{S}} 1 \cdot p_{xy} + \sum_{y \in A^c} u(y) \cdot p_{xy},$$

which is the following, in matrix notation,

$$\mathbf{u} = \mathbf{i} + \mathbf{B}\mathbf{u}.$$

In the equation above, $\mathbf{B}$ is the portion of the transition matrix pertaining to states $j \in A^c$, $\mathbf{u}$ is the vector of mean first-hitting times for all states $j \in A^c$, and $\mathbf{i}$ is a vector of ones. Thus, we can computer $\mathbf{u}$ by solving this linear system of equations. And this technique works for a wide variety of performance measures. Note that the **Example 5.8** is one application of this tool. In particular, we have $A = \{1, 2\}$, $A^c = \{0, 3\}$, $\mathbf{B} = \begin{pmatrix} 0 & p \\ 1-p & 0 \end{pmatrix}$ and $\mathbf{u} = \begin{bmatrix} \mathrm{E}(N_1) \\ \mathrm{E}(N_2) \end{bmatrix}$.

Example 6.10 Suppose that $\mathrm{f} : \mathcal{S} \to \mathcal{R}$ is a function returning the "cost" associated with occupying states in the Markov chain. For $x \in A^c$, define that $v(x) = \mathrm{E}\left[\sum_{k=0}^{T_A - 1} \mathrm{f}(X_k) \mid X_0 = x\right]$. How can we find out $v(x)$? Clearly, we can write

$$v(x) = \mathrm{E}\left[\sum_{k=0}^{T_A - 1} \mathrm{f}(X_k) \mid X_0 = x\right] = \mathrm{f}(x) + \mathrm{E}\left[\sum_{k=1}^{T_A - 1} \mathrm{f}(X_k) \mid X_0 = x\right].$$

Note that if the chain reaches the set A in the first step, then the second term above contains an empty summation and thus zero expected cost. Using the first-step analysis again, we have

$$\mathrm{E}\left[\sum_{k=1}^{T_A - 1} \mathrm{f}(X_k) \mid X_0 = x\right] = \sum_{y \in A} 0 \cdot p_{xy} + \sum_{y \in A^c} v(y) \cdot p_{xy}.$$

Therefore, we have $v(x) = \mathrm{f}(x) + \sum_{y \in A^c} v(y) \cdot p_{xy}$, or in matrix notation, $\mathbf{v} = \mathbf{f} + \mathbf{B}\mathbf{v}$, from which we can solve for $\mathbf{v}$. Note that **Example 6.9** is the same as this example with $\mathrm{f}(x) = 1$, $\forall x \in A^c$.

Example 6.11 Suppose that $0 < \alpha < 1$ and consider the infinite-horizon discounted cost $w(x) = \mathrm{E}\left[\sum_{k=0}^{\infty} \alpha^k \, \mathrm{f}(X_k) \mid X_0 = x\right]$, defined for all $x \in \mathcal{S}$. Show that $\mathbf{w} = \mathbf{f} + \alpha \mathbf{P} \mathbf{w}$.

Using the first-step analysis, we can write

$$w(x) = \mathrm{E}\left[\sum_{k=0}^{\infty} \alpha^k \, \mathrm{f}(X_k) \mid X_0 = x\right]$$
$$= \sum_{y=0}^{\infty} \left\{ \mathrm{E}\left[\sum_{k=0}^{\infty} \alpha^k \, \mathrm{f}(X_k) \mid X_1 = y, X_0 = x\right] \cdot \Pr[X_1 = y \mid X_0 = x] \right\}$$
$$= \sum_{y=0}^{\infty} \left\{ \mathrm{f}(x) \cdot p_{xy} + \mathrm{E}\left[\sum_{k=1}^{\infty} \alpha^k \, \mathrm{f}(X_k) \mid X_1 = y\right] \cdot p_{xy} \right\}$$
$$= \mathrm{f}(x) + \sum_{y=0}^{\infty} \left\{ \alpha \, \mathrm{E}\left[\sum_{k=1}^{\infty} \alpha^{k-1} \, \mathrm{f}(X_k) \mid X_1 = y\right] \cdot p_{xy} \right\}$$
$$= \mathrm{f}(x) + \alpha \cdot \sum_{y=0}^{\infty} \left\{ \mathrm{E}\left[\sum_{j=0}^{\infty} \alpha^j \, \mathrm{f}(X_{j+1}) \mid X_1 = y\right] \cdot p_{xy} \right\}$$
$$= \mathrm{f}(x) + \alpha \cdot \sum_{y=0}^{\infty} \left\{ w(y) \cdot p_{xy} \right\}.$$

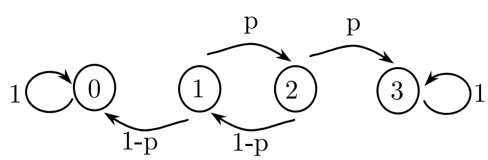
In matrix notation, the relationship above can be written as $\mathbf{w} = \mathbf{f} + \alpha \mathbf{P} \mathbf{w}$.

Example 6.12 Compute $\lim_{n\to\infty} P^n$ when the transition matrix is

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

Clearly, there are two communicating classes in this Markov chain, $C_1 = \{1,2\}$ and $C_2 = \{3,4\}$. Moreover, states 1 and 2 are transient and states 3 and 4 are positive recurrent (it is impossible to have null recurrent states in finite states space). Therefore, in the first two columns of $\lim_{n\to\infty} P^n$, we have all zeros. If we pay attention to the right lower 2×2 corner of the transition matrix, we find that it is a doubly stochastic matrix, so that the limiting probabilities across states 3 and 4 are both equal to $\frac{1}{2}$. Therefore, the last two columns of $\lim_{n\to\infty} P^n$ are full of $\frac{1}{2}$'s.

Example 6.13 Given the following transition diagram with $0 < p < 1$, find out $\lim_{n\to\infty} P^n$.



Clearly, there are three communicating classes in this chain, $C_1 = \{0\}, C_2 = \{1,2\}, C_3 = \{3\}$. It is easy to determine that $d(0) = d(3) = 1$ and $d(1) = d(2) = 2$. Apparently states 1 and 2 are transient and states 0 and 3 are positive recurrent (once again, it is impossible to get null recurrent states in a finite state space). Therefore, we have $\lim_{n\to\infty} p_{i1}^{(n)} = \lim_{n\to\infty} p_{i2}^{(n)} = 0, \forall i = 0,1,2,3$, $\lim_{n\to\infty} p_{03}^{(n)} = \lim_{n\to\infty} p_{30}^{(n)} = 0$, and $\lim_{n\to\infty} p_{00}^{(n)} = \lim_{n\to\infty} p_{33}^{(n)} = 1$. Note that the last limiting probabilities also imply that $\mu_{00} = \mu_{33} = 1$. Now we need determine $\lim_{n\to\infty} p_{10}^{(n)}, \lim_{n\to\infty} p_{13}^{(n)}, \lim_{n\to\infty} p_{20}^{(n)},$ and $\lim_{n\to\infty} p_{23}^{(n)}$, the limiting probabilities from one transient state to a positive recurrent state.

From **Comment 6.4**, it is easy to know that $\lim_{n\to\infty} p_{ij}^{(n)} = \mathrm{f}_{ij}, i = 1,2$ and $j = 0,3$. Now we use the first-step analysis to get $\mathrm{f}_{ij}, i = 1,2$ and $j = 0,3$. We can write

$$\begin{aligned} \mathrm{f}_{10} &= \sum_{k=1}^{\infty} \mathrm{f}_{10}^{k} = \mathrm{E}_y\left[\sum_{k=1}^{\infty} \left(\mathrm{f}_{10}^{k} \mid X_1 = y\right) \cdot \Pr\left(X_1 = y \mid X_0 = 1\right)\right] \\ &= 1 \cdot p_{10} + \sum_{k=2}^{\infty} \mathrm{f}_{20}^{k-1} \cdot p_{12} = (1-p) + \mathrm{f}_{20} \cdot p. \end{aligned}$$

Similarly, we have

$$\mathrm{f}_{13} = 0 \cdot (1-p) + \mathrm{f}_{23} \cdot p,\ \mathrm{f}_{20} = \mathrm{f}_{10} \cdot (1-p) + 0 \cdot p,\ \text{and}\ \mathrm{f}_{23} = \mathrm{f}_{13} \cdot (1-p) + 1 \cdot p.$$

Taking advantage of the facts that $\mathrm{f}_{10} + \mathrm{f}_{13} = 1$ and $\mathrm{f}_{20} + \mathrm{f}_{23} = 1$, the equations above reduce to the following system:

$$\mathrm{f}_{10} = (1-p) + p \cdot \mathrm{f}_{20} \ \text{and}\ \mathrm{f}_{20} = (1-p) \cdot \mathrm{f}_{10},$$

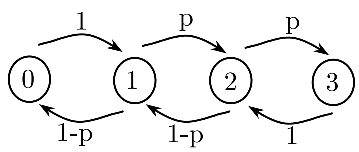
which produces the result

$$\mathrm{f}_{10} = \tfrac{1-p}{1-p(1-p)} \ \text{and}\ \mathrm{f}_{20} = \tfrac{(1-p)^2}{1-p(1-p)}.$$

Therefore, the limiting probabilities matrix is

$$P^n \to \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1-p}{1-p(1-p)} & 0 & 0 & \frac{(1-p)^2}{1-p(1-p)} \\ \frac{(1-p)^2}{1-p(1-p)} & 0 & 0 & \frac{1-p}{1-p(1-p)} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Example 6.14 Given the following transition diagram with $0 < p < 1$, find out $\lim_{n\to\infty} P^n$.



There is only one communicating class containing all four states and $d(0) = d(1) = d(2) = d(3) = 2$. All four states are positive recurrent. One of the most distinguish feature of this Markov chain is that in odd number of steps, state 0 cannot communicate with states 0 and 2, state 2 cannot communicate with states 2 and 0, state 1 cannot communicate with states 1 and 3, and state 3 cannot communicate with states 3 and 1. In even number of steps, the feature is reversed. Therefore, $\lim_{n\to\infty} P^n$ doesn't exist but both $\lim_{n\to\infty} P^{2n}$ and $\lim_{n\to\infty} P^{2n+1}$ exist.

For sufficiently large n, let the odd-number of steps transition matrix be the following,

$$P^{2n-1} = \begin{bmatrix} 0 & a & 0 & 1-a \\ b & 0 & 1-b & 0 \\ 0 & c & 0 & 1-c \\ d & 0 & 1-d & 0 \end{bmatrix}.$$

From P^{2n-1}, we can easily get the next even-number of steps transition matrix as,

$$P^{2n} = P^{2n-1}P$$

$$= \begin{bmatrix} 0 & a & 0 & 1-a \\ b & 0 & 1-b & 0 \\ 0 & c & 0 & 1-c \\ d & 0 & 1-d & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} a(1-p) & 0 & ap+(1-a) & 0 \\ 0 & b+(1-b)(1-p) & 0 & (1-b)p \\ c(1-p) & 0 & cp+(1-c) & 0 \\ 0 & d+(1-d)(1-p) & 0 & (1-d)p \end{bmatrix}.$$

Furthermore, we can do the same procedure one more time to get $P^{2n+1} = P^{2n}P$. Finally by equating $P^{2n+1} = P^{2n-1}$, which should be the case as $n \to \infty$, we can solve for a, b, c, d and thus $\lim_{n\to\infty} P^{2n}, \lim_{n\to\infty} P^{2n+1}$.

We get $a = c = \frac{1-p}{1-p(1-p)}$ and $b = d = \frac{(1-p)^2}{1-p(1-p)}$ and the $\lim_{n\to\infty} P^{2n}$ and $\lim_{n\to\infty} P^{2n+1}$ will be given accordingly.

References

Billingsley, P., *Probability and Measure*, 2nd ed. Wiley 1986. (higher level)

Cinlar, E., *Introduction to stochastic processes*, Prentice-Hall, Englewood Cliffs, New Jersey, 1975 (same level).

Feller, W., *Introduction to Probability Theory and Its Applications*, Vols. 1 and 2, Wiley and Sons, 1970 (higher level).

Heyman D., and Sobel, M., *Stochastic Models in Operations Research*, (Vol. 1), McGraw-Hill, 1982 (same level).

Karlin and Taylor, *A First Course in Stochastic Processes*, 2nd Edition, Academic Press, 1975 (higher level).

Parzen, E., *Stochastic Processes*, Holden Day, 1962 (same to slightly higher level).

Ross, S., *Introduction to Probability Models*, Academic Press, latest (lower level).

Ross, S., *Stochastic Processes*, 2nd Edition (**textbook and comparison basis**)

Wolff, R., *Stochastic Modeling and the Theory of Queues*, Englewood Cliffs, NJ, 1989 (same level).