STAT 333

Stochastic Processes I

This page intentionally left blank.

Contents

This page intentionally left blank.

1. Review of Probability

1.1 Probability Spaces

Def'n 1.1 **Probability Space**

A ***probability space*** is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ such that the following holds.

(a) The ***sample space*** Ω is nonempty.

(b) The ***event space*** $\mathcal{F}$ is a σ-algebra on Ω. That is, $\mathcal{F} \subseteq 2^{\Omega}$ with the following properties:

(i) $\Omega \in \mathcal{F}$;

(ii) for every $A \in \mathcal{F}$, $(\Omega \setminus A) \in \mathcal{F}$; and *closure under complements*

(iii) for every countable $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$. *closure under countable unions*

(c) The ***probability function*** $\mathbb{P} : \mathcal{F} \to [0,1]$ satisfies the following.

(i) For every countable and pairwise disjoint $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i). \qquad \textit{σ-additivity}$$

(ii) $\mathbb{P}(\Omega) = 1$.

(1.1) For simplicity, fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ throughout this section.

(1.2) Probability of the Complement of an Event

A direct consequence of Def'n 1.1 is the following: for every $A \in \mathcal{F}$,

$$\mathbb{P}(\Omega \setminus A) = 1 - \mathbb{P}(A).$$

Def'n 1.2 **Conditional Probability**

Let $A, B \in \mathcal{F}$ be such that $\mathbb{P}(B) \neq 0$. The ***conditional probability*** of A given B occurs, denoted as $\mathbb{P}(A|B)$, is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

(1.3) Let $A, B \in \mathcal{F}$ be such that $\mathbb{P}(B) \neq 0$.

(a) Note that

$$\mathbb{P}(A|\Omega) = \frac{\mathbb{P}(A \cap \Omega)}{\mathbb{P}(\Omega)} = \frac{\mathbb{P}(A)}{1} = \mathbb{P}(A),$$

as expected.

(b) By rearranging,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B). \qquad \textit{multiplication rule}$$

For any finite $\{A_i\}_{i=1}^{n} \subseteq \mathcal{F}$, we can generalize the multiplication rule as follows:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}\left(A_i | \bigcap_{j=1}^{i} A_j\right), \qquad \textit{generalized multiplication rule}$$

provided that $\mathbb{P}\left(\bigcap_{j=1}^{i} A_j\right) \neq 0$ for all $i \in \{1, \ldots, n\}$.

(EX 1.4)
Rolling a Fair Die

Suppose that we roll a fair six-sided die once. Let A denote the event of rolling a number less than 4 and let B denote the event of rolling an odd number. Given that the roll is odd, what is the probability that the number rolled is less than 4?

Answer. Note that we are trying to calculate $\mathbb{P}(A|B)$. By definition, $A = \{1,2,3\}, B = \{1,3,5\}$. So it follows that

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}\{1,3\}}{\mathbb{P}\{1,3,5\}} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}. \qquad \triangleleft$$

Note that we are *implicitly* defining the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as $\left(\{1,\ldots,6\}, 2^{\{1,\ldots,6\}}, |\cdot|\right)$ for (EX 1.4).

Def'n 1.3 **Independent** Events

We say $A, B \in \mathcal{F}$ are ***independent*** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

Theorem 1.1
Law of Total Probability

Let $\mathcal{C} \subseteq \mathcal{F}$ be a countable partition of Ω. Then

$$\mathbb{P}(A) = \sum_{B \in \mathcal{C}} \mathbb{P}(A|B)\,\mathbb{P}(B)$$

for every $A \in \mathcal{F}$.

Corollary 1.1.1
Bayes' Formula

Consider the setting of Theorem 1.1. Then for every $C \in \mathcal{C}$,

$$\mathbb{P}(C|A) = \frac{\mathbb{P}(A|C)\,\mathbb{P}(C)}{\sum_{B \in \mathcal{C}} \mathbb{P}(A|B)\,\mathbb{P}(B)}.$$

1.2 Random Variables

Def'n 1.4 **Random Variable**

A ***random variable*** (or ***rv*** for short) X is a function of the form $X : \Omega \to \mathbb{R}$, where Ω is the sample space of a probability space.

Def'n 1.5 **Discrete** Random Variable

Let X be a random variable. When the image of X is countable, we say X is ***discrete***. There are two important functions that are associated with X.

(a) We define the ***probability mass function*** (or ***pmf*** for short) for X, denoted as p_X, by

$$p_X(x) = \mathbb{P}\{X = x\} \qquad \forall x \in \mathbb{R}.$$

(b) We define the ***cumulative distribution function*** (or ***cdf*** for short) for X, denoted as F_X, by

$$F_X(x) = \mathbb{P}\{X \le x\} = \sum_{y \le x} p_X(x) \qquad \forall x \in \mathbb{R}.$$

(1.5) Let X be a discrete random variable.

(a) Sometimes it is handy to have the ***tail probability function*** (or ***tpf*** for short) for X, denoted as $\overline{F}_X$: it is defined as
$$\overline{F}_X(x) = 1 - F(x) \qquad \forall x \in \mathbb{R}.$$

(b) Let S be the image of X. We can order the elements of S in the increasing order, so that $S = \{x_i\}_{i=1}^{n}$ if S is finite or $S = \{x_i\}_{i=1}^{\infty}$ if S is infinite, where $x_i < x_{i+1}$ for all i. Then note that we can *recover* the pmf p_X of X from F_X by
$$p_X(x_1) = F_X(x_1)$$
and
$$p_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$
for every $i \geq 2$.

(1.6) Bernoulli

A ***Bernoulli trial*** is a random trial with probability $p \in [0,1]$ of being a *success* and probability $1-p$ of being a *failure*. If we let $X = 1$ if the trial is successful and $X = 0$ if it fails, then X is said to be a ***Bernoulli*** random variable with parameter p, denoted as $X \sim \mathrm{B}(p)$. Note that X has a pmf
$$p_X(x) = p^x (1-p)^{1-x}$$
for all $x \in \{0,1\}$.

(1.7) Binomial

A binomial random variable generalizes Bernoulli random variable. Consider the case where we run $n \in \mathbb{N}$ *independent* Bernoulli trials, each with probability $p \in (0,1]$, where we let X denote the number of successes. Then we say X is a ***binomial*** random variable with parameters n, p, denoted as $X \sim \mathrm{BIN}(n,p)$. The pmf of X is given by
$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad [1.1]$$
for all $x \in \{0, \ldots, n\}$. Note that $\binom{n}{x}$ is the *number of distinct x-subsets of a n-set*. Here are some remarks.

(a) A $\mathrm{BIN}(1,p)$ simplifies to become $\mathrm{B}(p)$.

(b) Note that [1.1] is even defined for $n = 0$, in which case $p_X(0) = 1$. Such a distribution is said to be *degenerate* at 0.

(1.8) Negative Binomial

Suppose that we have independent Bernoulli trials, each with success probability $p \in (0,1]$ required to observe $n \in \mathbb{N}$ successes. If we let X denote the number of trials needed, then X is a ***negative binomial*** random variable with parameters n, p, denoted as $X \sim \mathrm{NB}_t(n,p)$. X has a pmf
$$p_X(x) = \binom{x-1}{n-1} p^n (1-p)^{x-n} \qquad [1.2]$$
for every $x \in \mathbb{N}, x \geq n$.

(a) Note that the apperence of $\binom{x-1}{n-1}$ instead of $\binom{x}{n}$ in [1.2]; this is because the final trial (i.e. the nth trial) must always be a success.

(b) Sometimes, a negative binomial distribution is alternatively defined as the number of *failures* observed to achieve n successes. If Y denotes such a random variable and $X \sim \mathrm{NB}_t(n,p)$, then clearly $X = Y + n$, which implies
$$p_Y(y) = \binom{y+n-1}{n-1} p^n (1-p)^y$$
for all $y \in \mathbb{N} \cup \{0\}$. We denote $Y \sim \mathrm{NB}_f(n,p)$.

(1.9) Geometric

A ***geometric*** random variable is a *special case of negative binomial*: that is, if $X \sim \mathrm{NB}_t(1,p)$ for some $p \in (0,1]$, then we say X is a geometric random variable with success probability p, denoted as $X \sim \mathrm{GEO}_t(p)$.[1]

(1.10) Discrete Uniform

If a random variable X is *equally likely* to take on values in a finite set $\{a, a+1, \ldots, b\}$ for some $a, b \in \mathbb{Z}, a \leq b$, then we say X is a ***discrete uniform*** random variable, denoted as $X \sim \mathrm{DU}(a,b)$. X has a pmf

$$p_X(x) = \frac{1}{b-a+1}$$

for every $x \in \{a, a+1, \ldots, b\}$.

(1.11) Hypergeometric

If X denotes the number of success objects in $n \in \mathbb{N}$ draws *without replacement* from a finite set of size $N \in \mathbb{N}$ containing exactly $r \in \mathbb{N}$ success objects, then X is a ***hypergeometric*** random variable with parameters N, r, n, denoted as $X \sim \mathrm{HG}(N,r,n)$. X has a pmf

$$p_X(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

for all $x \in \{\max\{0, n-N+r\}, \ldots, \min\{n,r\}\}$.

(1.12) Poisson

A random variable is called ***Poisson*** with parameter $\lambda > 0$, denoted as $X \sim \mathrm{POI}(\lambda)$, if

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad [1.3]$$

for all $x \in \mathbb{N} \cup \{0\}$. Note that [1.3] is even defined for $\lambda = 0$, in which case $p_X(0) = 1$ (i.e. X is degenerate at 0).

(EX 1.13) Approximating Binomial with Poisson

Show that when $n \in \mathbb{N}$ is large and $p \in (0,1]$ is small, $p_X \sim p_Y$ where $X \sim \mathrm{BIN}(n,p), Y \sim \mathrm{POI}(np)$.

Proof. Let $x \in \{0, \ldots, n\}$. Then

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!}\frac{(n)_x}{x!}\frac{\left(1-\frac{\lambda}{n}\right)^n}{(1-\lambda_n)^x}.$$

where $\lambda = np$. Now note that $(n)_x \sim n^x$, $1 - \frac{\lambda}{n} \sim 1$, and $\left(1-\frac{\lambda}{n}\right)^n \sim e^{-\lambda}$ since n is large and $p = \frac{\lambda}{n}$ is small. Hence

$$p_X(x) = \frac{\lambda^x}{x!}\frac{(n)_x}{x!}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x} \sim e^{-\lambda}\frac{\lambda^x}{x!} = p_Y(x),$$

as required. ◁

Def'n 1.6

Continuous Random Variable

Let X be a random variable. We say X is ***continuous*** if there exists nonnegative $f_X : \mathbb{R} \to \mathbb{R}$ such that

$$\mathbb{P}\{X \in B\} = \int_{x \in B} f_X(x)\,\mathrm{dx}$$

for all measurable $B \subseteq \mathbb{R}$, where f_X is called the ***probability density function*** (***pmf***) of X. We also define

[1] Similar to negative binomial, we write $X \sim \mathrm{GEO}_f(p)$ if $X \sim \mathrm{NB}_f(1,p)$.

the ***cumulative distribution function*** $F_X : \mathbb{R} \to [0,1]$ of X by

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t \qquad \forall x \in \mathbb{R}.$$

(1.14) Let X be a continuous random variable. Then note that

$$f_X = F_X'$$

by the fundamental theorem of calculus.

(1.15) Uniform A random variable X is called a ***uniform*** random variable on an interval $(a,b) \subseteq \mathbb{R}$, denoted as $X \sim \mathrm{U}(a,b)$ if

$$f_X(x) = \frac{1}{b-a} \qquad \forall x \in (a,b).$$

(1.16) Beta A random variable X is called ***Beta*** with parameters $m,n \in \mathbb{N}$, denoted as $X \sim \mathrm{BETA}(m,n)$, if

$$f_X(x) = \frac{(m+n-1)!}{(m-1)!(n-1)!} x^{m-1}(1-x)^{n-1} \qquad \forall x \in (0,1).$$

(1.17) Erlang A random variable X is called ***Erlang*** with parameters $n \in \mathbb{N}, \lambda > 0$, denoted as $X \sim \mathrm{ERLANG}(n,\lambda)$ if

$$f_X(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \qquad \forall x > 0.$$

(1.18) Exponential A random variable X is called ***exponential*** with parameter $\lambda > 0$, denoted as $X \sim \mathrm{EXP}(\lambda)$, if

$$f_X(x) = \lambda e^{-\lambda x} \qquad \forall x > 0.$$

Note that $\mathrm{ERLANG}(1,\lambda)$ simplifies to $\mathrm{EXP}(\lambda)$.

1.3 Expectation

Expectation of a Random Variable

Def'n 1.7 Let X be a random variable. Then we define the ***expectation*** of X, denoted as $\mathbb{E}(X)$, by

$$\mathbb{E}(X) = \begin{cases} \sum_{x \in \mathbb{R}: p_X(x) > 0} x p_X(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x f_X(x)\,\mathrm{d}x & \text{if } X \text{ is continuous} \end{cases}$$

if exists.

***n*th Moment** of a Random Variable

Def'n 1.8 Let X be a random variable. For any $n \in \mathbb{N} \cup \{0\}$, if $\mathbb{E}(X^n)$ exists, then it is called the ***nth moment*** of X.

Variance, Standard Deviation of a Random Variable

Def'n 1.9 Let X be a random variable. We define the ***variance*** of X, denoted as $\mathrm{var}(X)$, by

$$\mathrm{var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).$$

We define the ***standard deviation*** (***stdev***) of X, denoted as $\operatorname{sd}(X)$, by

$$\operatorname{sd}(X) = \sqrt{\operatorname{var}(X)}.$$

Theorem 1.2
Law of the Unconscious Statistician (LOTUS)

Let X be a random variable and let $g : \mathbb{R} \to \mathbb{R}$ be measurable. Then

$$\mathbb{E}(g(X)) = \begin{cases} \sum_{x \in \mathbb{R}: p_X(x) > 0} g(x) p_X(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} g(x) f_X(x) \,\mathrm{d}x & \text{if } X \text{ is continuous} \end{cases}.$$

Corollary 1.2.1

Let X be a random variable and let $a, b \in \mathbb{R}$.

(a) $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.

(b) $\operatorname{var}(aX + b) = a^2 \operatorname{var}(X)$.

Moment Generating Function of a Random Variable

Def'n 1.10 Let X be a random variable. We define the ***moment generating function*** (***mgf***) of X, denoted as φ_X, by

$$\varphi_X(t) = \mathbb{E}\left(e^{tX}\right) \qquad \forall t \in \mathbb{R}.$$

(1.19) Note that

$$\varphi_X(t) = \mathbb{E}\left(e^{tX}\right) = \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right) = \sum_{n=0}^{\infty} \mathbb{E}(X^n) \frac{t^n}{n!},$$

implying that $\mathbb{E}(X^n)$ is the coefficient of $\frac{t^n}{n!}$ in the above series expansion. In particular,

$$\mathbb{E}(X^n) = \varphi_X^{(n)}(0)$$

for all $n \in \mathbb{N}$.

(1.20) It is worth noting that a mgf *uniquely* determines the probability distribution of a random variable.

(EX 1.21) Let $X \sim \operatorname{BIN}(n, p)$, where $n \in \mathbb{N}, p \in (0, 1]$. Find φ_X and use it to calculate $\mathbb{E}(X)$.

Answer. Observe that, for every $t \in \mathbb{R}$,

$$\begin{aligned} \varphi_X(t) = \mathbb{E}\left(e^{tX}\right) &= \sum_{x=0}^{n} e^{tX} p_X(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} \left(e^t p\right)^x (1-p)^{n-x} = \left(e^t p + 1 - p\right)^n, \end{aligned}$$

where the last equality holds by the binomial theorem. It follows that

$$\mathbb{E}(X) = \varphi_X'(0) = \frac{\mathrm{d}}{\mathrm{d}t} \left(e^t p + 1 - p\right)^n \Big|_{t=0} = n\left(e^t p + 1 - p\right)^{n-1} e^t p \Big|_{t=0} = np. \qquad \triangleleft$$

1.4 Joint Distributions

Random Vector

Def'n 1.11 Let $X_1, \ldots, X_n$ be random variables. Then we call the n-tuple $\mathbf{X} = (X_1, \ldots, X_n)$ a ***random vector***.

(a) The ***joint cdf*** of $\mathbf{X}$, denoted as $F_{\mathbf{X}}$, is defined as

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}\{\mathbf{X} \leq \mathbf{x}\} \qquad \forall \mathbf{x} \in \mathbb{R}^n.$$

(b) When $X_1, \ldots, X_n$ is discrete, we say $\mathbf{X}$ is ***jointly discrete***, and define the ***joint pmf*** of $\mathbf{X}$, denoted as $p_{\mathbf{X}}$, by

$$p_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}\{\mathbf{X} = \mathbf{x}\} \qquad \forall \mathbf{x} \in \mathbb{R}^n.$$

(c) When there exists nonnegative $f_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathbb{P}\{\mathbf{X} \in S\} = \int_S f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

for every $S \subseteq \mathbb{R}^n$, then we say $\mathbf{X}$ is ***jointly continuous*** and call $f_{\mathbf{X}}$ a ***joint pdf*** of $\mathbf{X}$.

(1.22) Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector.

(a) Note that, for every $i \in \{1, \ldots, n\}$,

$$F_{X_i}(x_i) = F_{\mathbf{X}}\Bigg(\infty, \ldots, \infty, \underbrace{x_i}_{i\text{th position}}, \infty, \ldots, \infty\Bigg) \qquad \forall x_i \in \mathbb{R}.$$

We call F_{X_i} the ***ith marginal cdf*** of $\mathbf{X}$.

(b) In case $\mathbf{X}$ is jointly discrete, for every $i \in \{1, \ldots, n\}$,

$$p_{X_i}(x_i) = \sum_{\substack{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \in \mathbb{R} \\ : p_{\mathbf{X}}(x_1, \ldots, x_n) > 0}} p_{\mathbf{X}}(x_1, \ldots, x_n) \qquad \forall x_i \in \mathbb{R}.$$

We call p_{X_i} the ***ith marginal pmf*** of $\mathbf{X}$.

(c) In case $\mathbf{X}$ is jointly continuous, each X_i is continuous, and for every $i \in \{1, \ldots, n\}$,

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \underbrace{\int_{-\infty}^{x_i}}_{i\text{th from right}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(t_1, \ldots, t_n)\,\mathrm{d}t_1 \cdots \mathrm{d}t_n \qquad \forall x_i \in \mathbb{R}.$$

We call f_{X_i} the ***ith marginal pdf*** of $\mathbf{X}$. It is worth noting that

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = \frac{\partial^n}{\partial \mathrm{x}_1 \cdots \partial \mathrm{x}_n} F(x_1, \ldots, x_n).$$

Proposition 1.3

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be joinly continuous. Then for any injective C^1 $g : \mathbb{R}^n \to \mathbb{R}^n$ with nowhere vanishing Jacobian determinant,

$$f_{g(\mathbf{X})}(\mathbf{y}) = f_{\mathbf{X}}\left(g^{-1}(\mathbf{y})\right)\left|\mathrm{J}_g\left(g^{-1}(\mathbf{y})\right)\right|^{-1} \qquad \forall \mathbf{y} \in g^{-1}(\mathbb{R}^n).$$

Def'n 1.12

Expectation of a Random Vector

Let $\mathbf{X}$ be a random vector. Then we define the ***expectation*** of $\mathbf{X}$, denoted as $\mathbb{E}(\mathbf{X})$, by

$$\mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n)).$$

Def'n 1.13

Covariance of Two Random Variables

Let X, Y be random variables. Then we define the ***covariance*** of X, Y, denoted as $\operatorname{cov}(X,Y)$, by

$$\operatorname{cov}(X,Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$

(1.23)
Covariance

Let X, Y be random variables. Note that

$$\operatorname{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

In particular, $\operatorname{cov}(X,X) = \operatorname{var}(X)$.

Theorem 1.4
Multivariate LOTUS

Let $\mathbf{X}$ be a random vector and let $g : \mathbb{R}^n \to \mathbb{R}$. Then

$$\mathbb{E}(g(\mathbf{X})) = \begin{cases} \sum_{\mathbf{x} \in \mathbb{R}^n : p_{\mathbf{X}}(\mathbf{x}) > 0} g(\mathbf{x})\, p_{\mathbf{X}}(\mathbf{x}) & \text{if } X \text{ is jointly discrete} \\ \int_{\mathbb{R}^n} g(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x} & \text{if } \mathbf{X} \text{ is jointly continuous} \end{cases}.$$

Corollary 1.4.1
Linearity of Expectation

Let $X_1, \dots, X_n$ be random variables and let $a_1, \dots, a_n \in \mathbb{R}$. Then

$$\mathbb{E}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i \mathbb{E}(X_i).$$

Corollary 1.4.2

Let X_1, X_2 be random variables and let $a_1, a_2 \in \mathbb{R}$. Then

$$\operatorname{var}(a_1 X_1 + a_2 X_2) = a_1^2 \operatorname{var}(X_1) + a_2^2 \operatorname{var}(X_2) + 2 a_1 a_2 \operatorname{cov}(X_1, X_2).$$

Def'n 1.14

Joint MGF of a Random Vector

Let $\mathbf{X} = (X_1, \dots, X_n)$ be a random vector. We define the ***joint mgf*** of $\mathbf{X}$, denoted as $\varphi_{\mathbf{X}}$, by

$$\varphi_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}^T \mathbf{X}} \qquad \forall \mathbf{t} \in \mathbb{R}^n.$$

(1.24)
Joint Moment

Let $\mathbf{X}$ be a random vector. Then for every $i_1, \dots, i_n \in \mathbb{N} \cup \{0\}$,

$$\mathbb{E}\left(X_1^{i_1} \cdots X_n^{i_n}\right) = \left.\frac{\partial^{\sum_{j=1}^n i_j}}{\partial \mathrm{x}_1^{i_1} \cdots \partial \mathrm{x}_n^{i_n}} \varphi_{\mathbf{X}}(\mathbf{x})\right|_{\mathbf{x}=\mathbf{0}}.$$

1.5 Independence

Def'n 1.15 **Independent** Random Variables

Let $X_1, \dots, X_n$ be random variables. We say $X_1, \dots, X_n$ are ***independent*** if

$$F_{(X_1,\dots,X_n)}(x_1,\dots,x_n) = \prod_{i=1}^n F_{X_i}(x_i)$$

for all $(x_1,\dots,x_n) \in \mathbb{R}^n$.

(1.25) Let $X_1, \dots, X_n$ be random variables and let $\mathbf{X} = (X_1,\dots,X_n)$.

(a) If $\mathbf{X}$ is jointly discrete, then Def'n 1.15 is equivalent to saying that $p_{\mathbf{X}}(x_1,\dots,x_n) = \prod_{i=1}^n p_{X_i}(x_i)$ for every $(x_1,\dots,x_n) \in \mathbb{R}^n$.

(b) If $\mathbf{X}$ is jointly continuous, then Def'n 1.15 is equivalent to saying that $f_{\mathbf{X}}(x_1,\dots,x_n) = \prod_{i=1}^n f_{X_i}(x_i)$ for every $(x_1,\dots,x_n) \in \mathbb{R}^n$.

(c) If $n=2$ and X_1, X_2 are independent, note that $\operatorname{cov}(X_1,X_2) = 0$. However the converse does not hold in general.

Theorem 1.5 MGF of the Sum of Independent Random Variables

Let $X_1, \dots, X_n$ be independent random variables. Then

$$\varphi_{\sum_{i=1}^n X_i} = \prod_{i=1}^n \varphi_{X_i}.$$

Corollary 1.5.1

Let $X_1, \dots, X_n$ be iid random variables. Then

$$\varphi_{\sum_{i=1}^n X_i} = \varphi_{X_1}^n.$$

(EX 1.26) Sum of Independent Binomial Random Variables

Let $X_1 \sim \mathrm{BIN}(n_1,p), \dots, X_m \sim \mathrm{BIN}(n_m,p)$, where $n_1,\dots,n_m \in \mathbb{N}, p \in (0,1]$. Find the distribution of $\sum_{i=1}^m X_i$.

Answer. Observe that, for every $t \in \mathbb{R}$,

$$\varphi_{\sum_{i=1}^m X_i}(t) = \prod_{i=1}^m \varphi(t) = \prod_{i=1}^m \left(e^t p + 1 - p\right)^{n_i} = \left(e^t p + 1 - p\right)^{\sum_{i=1}^m n_i} = \varphi_Y(t),$$

where $Y \sim \mathrm{BIN}\left(\sum_{i=1}^m n_i, p\right)$. It follows from (1.20) that $\sum_{i=1}^m X_i \sim \mathrm{BIN}\left(\sum_{i=1}^m n_i, p\right)$. ◁

Def'n 1.16 **Convergence** of a Sequence of Random Variables

Let $(X_n)_{n=1}^\infty$ be a sequence of random variables and let X be a random variable.

(a) We say $(X_n)_{n=1}^\infty$ ***converges*** to X ***in distribution*** if

$$\lim_{n\to\infty} \mathbb{P}\{X_n \le x\} = \mathbb{P}\{X \le x\}$$

for all $x \in \mathbb{R}$ at which F_X is continuous.

(b) We say $(X_n)_{n=1}^\infty$ ***converges*** to X ***in probability*** if

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

for every $\varepsilon > 0$.

(c) We say $(X_n)_{n=1}^{\infty}$ ***converges*** to X ***almost surely*** (***a.s.***) if

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

(1.27) Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables and let X be a random variable. Then

$$\begin{aligned}(X_n)_{n=1}^{\infty} \text{ converges to } X \text{ a.s.} &\implies (X_n)_{n=1}^{\infty} \text{ converges to } X \text{ in probability} \\ &\implies (X_n)_{n=1}^{\infty} \text{ converges to } X \text{ in distribution.}\end{aligned}$$

Theorem 1.6 Strong Law of Large Numbers (SLLN)

Let $(X_n)_{n=1}^{\infty}$ be a sequence of iid random variables with common expectation $\mu \in \mathbb{R}$. Then $\left(\overline{X}_n\right)_{n=1}^{\infty}$ converges to μ almost surely, where for every $n \in \mathbb{N}$,

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}.$$

This page intentionally left blank.

2.
Conditional Distributions

2.1 Jointly Discrete Case

(2.1) For convenience, we shall only consider *bivariate* case. Let X_1, X_2 be discrete random variables and let $x_2 \in \mathbb{R}$ throughout this section.

Def'n 2.1 **Conditional PMF**

If $p_{X_2}(x_2) > 0$, then we define the ***conditional pmf*** of X_1 given $X_2 = x_2$, denoted as $p_{X_1|X_2}(\cdot|x_2)$, is defined by

$$p_{X_1|X_2}(x_1|x_2) = \frac{p_{(X_1,X_2)}(x_1,x_2)}{p_{X_2}(x_2)}$$

for all $x_1 \in \mathbb{R}$. We denote the resulting distribution by $X_1|(X_2 = x_2)$.

(2.2)

(a) We alternatively write $\mathbb{P}(X_1 = \cdot|X_2 = x_2)$ to denote $p_{X_1|X_2}(\cdot|x_2)$. Also note that

$$p_{X_1|X_2}(x_1|x_2) = \mathbb{P}(X_1 = x_1) = \frac{\mathbb{P}(X_1 = x_2, X_2 = x_2)}{\mathbb{P}(X_2 = x_2)} = \frac{p_{(X_1,X_2)}(x_1,x_2)}{p_{X_2}(x_2)}$$

(b) If X_1, X_2 are *independent*, then

$$p_{(X_1,X_2)}(x_1,x_2) = p_{X_1}(x_1)\, p_{X_2}(x_2)$$

for every $x_1, x_2 \in \mathbb{R}$, which means

$$p_{X_1|X_2}(x_1|x_2) = p_{X_1}(x_1)$$

for all $x_1, x_2 \in \mathbb{R}$ such that $p_{X_2}(x_2) > 0$.

Def'n 2.2 **Conditional Expectation**

If $p_{X_2}(x_2) > 0$, then we define the ***conditional mean***, denoted as $\mathbb{E}(X_1|X_2 = x_2)$, of $X_1|(X_2 = x_2)$ by

$$\mathbb{E}(X_1|X_2 = x_2) = \sum\nolimits_{x_1 \in \mathbb{R}: p_{X_1|X_2}(x_1|x_2) > 0} x_1 p_{X_1|X_2}(x_1|x_2).$$

Proposition 2.1

Let $w : \mathbb{R}^2 \to \mathbb{R}$. Then

$$\mathbb{E}(w(X_1,X_2)|X_2 = x_2) = \mathbb{E}(w(X_1,x_2)|X_2 = x_2).$$

Corollary 2.1.1

Given any $g, h : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}(g(X_1)h(X_2)|X_2 = x_2) = \mathbb{E}(g(X_1)h(x_2)|X_2 = x_2).$$

Corollary 2.1.2

Let X_3 be a random variable and let $x_3 \in \mathbb{R}$ be such that $p_{X_3}(x_3) > 0$. Then

$$\mathbb{E}(X_1 + X_2|X_3 = x_3) = \mathbb{E}(X_1|X_3 = x_3) + \mathbb{E}(X_2|X_3 = x_3).$$

Def'n 2.3

Conditional Variance

We define the ***conditional variance*** of $X_1|X_2 = x_2$, denoted as $\operatorname{var}(X_1|X_2 = x_2)$, by

$$\operatorname{var}(X_1|X_2 = x_2) = \mathbb{E}\left((X_1 - \mathbb{E}(X_1|X_2 = x_2))^2 |X_2 = x_2\right).$$

Proposition 2.2

We have

$$\operatorname{var}(X_1|X_2 = x_2) = \mathbb{E}\left(X_1^2|X_2 = x_2\right) - \mathbb{E}(X_1|X_2 = x_2)^2.$$

(EX 2.3) Suppose $X_1 \sim \mathrm{BIN}(n_1, p), X_2 \sim \mathrm{BIN}(n_2, p)$ for some $n_1, n_2 \in \mathbb{N} \cup \{0\}, p \in (0,1]$ are independent and let $m \in \mathbb{N} \cup \{0\}$. Find $p_{X_1|X_1+X_2}(\cdot|X_1 + X_2 = m)$.

Answer. We may assume $m \le n_1 + n_2$, since otherwise $p_{X_1+X_2}(m) = 0$. Then observe that

$$\begin{aligned} p_{X_1|X_1+X_2}(x_1|X_1+X_2=m) &= \frac{\mathbb{P}(X_1 = x_1, X_1+X_2 = m)}{\mathbb{P}(X_1+X_2=m)} \\ &= \frac{\mathbb{P}(X_1 = x_1, X_2 = m - x_1)}{\mathbb{P}(X_1+X_2=m)} \\ &= \frac{\mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = m - x_1)}{\mathbb{P}(X_1+X_2=m)} && \text{since } X_1, X_2 \text{ are independent} \\ &= \frac{\binom{n_1}{x_1} p^{x_1}(1-p)^{n_1-x_1}\binom{n_2}{m-x_1} p^{m-x_1}(1-p)^{1-m+x_1}}{\binom{n_1+n_2}{m} p^m (1-p)^{1-m}} && \text{since } X_1+X_2 \sim \mathrm{BIN}(n_1+n_2, p) \end{aligned}$$

for all $x_1 \in \{0, \ldots, n_1\}$. But note that this is exactly the pmf of $\mathrm{HG}(n_1+n_2, n_1, m)$. That is,

$$X_1|(X_1+X_2=m) \sim \mathrm{HG}(n_1+n_2, n_1, m). \qquad \triangleleft$$

Here is an intuitive explanation of why $X_1|(X_1+X_2=m) \sim \mathrm{HG}(n_1+n_2, n_1, m)$. Consider a sequence of $n_1 + n_2$ Bernoulli trials $(B_i)_{i=1}^{n_1+n_2}$, each with success probability p. We know exactly m of $B_1, \ldots, B_{n_1+n_2}$ are successes, and we also know exactly n_1 of $B_1, \ldots, B_{n_1}$ are successes. But each B_i has success probability p, so we end up with a hypergeometric distribution. See (1.11).

(EX 2.4) Let $X_1 \sim \mathrm{POI}(\lambda_1), \ldots, X_m \sim \mathrm{POI}(\lambda_m)$ for some $\lambda_1, \ldots, \lambda_m > 0$ be independent and let $Y = \sum_{i=1}^m X_i$. Find the conditional distribution of $X_j|(Y = n)$, where $j \in \{1, \ldots, m\}, n \in \mathbb{N}$.

Answer. First note that $X_j, \sum_{i=1, i\neq j}^m X_i$ are independent, since $X_1, \ldots, X_m$ are independent. Fix $x_j \in \{0, \ldots, n\}$. Then

$$\begin{aligned} p_{X_j|Y}(x_j|n) &= \frac{\mathbb{P}(X_j = x_j, Y = n)}{\mathbb{P}(Y=n)} \\ &= \frac{\mathbb{P}(X_j = x_j, \sum_{i=1}^m X_i = n)}{\mathbb{P}(Y=n)} \\ &= \frac{\mathbb{P}\left(X_j = x_j, \sum_{i=1, i\neq j}^m X_i = n - x_j\right)}{\mathbb{P}(Y=n)} \\ &= \frac{\mathbb{P}(X_j = x_j)\mathbb{P}\left(\sum_{i=1, i\neq j}^m X_i = n - x_j\right)}{\mathbb{P}(Y=n)}. && \text{since } X_j, \textstyle\sum_{i=1, i\neq j}^m X_i \text{ are independent} \end{aligned}$$

But

$$Y \sim \text{POI}\left(\sum_{i=1}^{m} \lambda_i\right), \sum_{i=1,i\neq j}^{m} X_i \sim \text{POI}\left(\sum_{i=1,i\neq j}^{m} \lambda_i\right) \qquad [2.1]$$

as sums of random variables, so

$$\begin{aligned}
p_{X_j|Y}(x_j|n) &= \frac{\frac{e^{-\lambda_j}\lambda_j^{x_j}}{x_j!}\frac{e^{-\sum_{i=1,i\neq j}^{m}\lambda_i}\left(\sum_{i=1,i\neq j}^{m}\lambda_i\right)^{n-x_j}}{(n-\lambda_j)!}}{\frac{e^{\sum_{i=1}^{m}\lambda_i}\left(\sum_{i=1}^{m}\lambda_i\right)^{n}}{n!}}. && \text{by [2.1]}\\
&= \binom{n}{x_j}\frac{\lambda_j^{x_j}\left(\sum_{i=1,i\neq j}^{m}\lambda_i\right)^{n-x_j}}{\left(\sum_{i=1}^{m}\lambda_i\right)^{n}}\\
&= \binom{n}{x_j}\frac{\lambda_j^{x_j}(\lambda-\lambda_j)^{n-x_j}}{\lambda^n} && \text{by letting } \lambda=\textstyle\sum_{i=1}^{m}\lambda_i\\
&= \binom{n}{x_j}\left(\frac{\lambda_j}{\lambda}\right)^{x_j}\left(\frac{\lambda-\lambda_j}{\lambda}\right)^{n-x_j}\\
&= \binom{n}{x_j}\left(\frac{\lambda_j}{\lambda}\right)^{x_j}\left(1-\frac{\lambda_j}{\lambda}\right)^{n-x_j}\\
&= \binom{n}{x_j}p^{x_j}(1-p)^{n-x_j}. && \text{by letting } p=\tfrac{\lambda_i}{\lambda}
\end{aligned}$$

Since $0 < \lambda_i \geq \lambda$, $p \in (0,1]$, so it follows that

$$X_j|(Y=n) \sim \text{BIN}\left(n, \frac{\lambda_j}{\sum_{i=1}^{m}}\lambda_i\right). \qquad \triangleleft$$

2.2 Jointly Continuous Case

(2.5) Let X, Y be jointly continuous random variables and let $y \in \mathbb{R}$ throughout this section.

Def'n 2.4 **Conditional PDF**

We define the ***conditional pdf*** of X given $Y=y$, denoted as $f_{X|Y}(\cdot|y)$, by

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f(y)}$$

for all $x \in \mathbb{R}$.

(2.6) Given $a, b \in \mathbb{R}, a \leq b$, observe that

$$\mathbb{P}(a \leq X \leq b|Y=y) = \int_a^b f_{X|Y}(x|y)\,\mathrm{d}x.$$

Def'n 2.5 **Conditional Expectation**

We define the ***conditional expectation*** of X given $Y = y$, denoted as $\mathbb{E}(X|Y=y)$, as

$$\mathbb{E}(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, \mathrm{d}x\,.$$

Proposition 2.3

Let $g : \mathbb{R} \to \mathbb{R}$. Then

$$\mathbb{E}(g(X)|Y=y) = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\, \mathrm{d}x\,.$$

Def'n 2.6 **Conditional Variance**

We define the ***conditional variance*** of X given $Y = y$, denoted as $\operatorname{var}(X|Y=y)$, as

$$\operatorname{var}(X|Y=y) = \mathbb{E}\left((X - \mathbb{E}(X|Y=y))^2 |Y=y\right).$$

Proposition 2.4

We have

$$\operatorname{var}(X|Y=y) = \mathbb{E}\left(X^2|Y=y\right) - \mathbb{E}(X|Y=y)^2\,.$$

2.3 Conditioning

(2.7) Let X, Y be random variables. Then we can define $v : \mathbb{R} \to \mathbb{R}$ by

$$v(y) = \mathbb{E}(X|Y=y)$$

for all $y \in \mathbb{R}$.

Notation 2.7 $\mathbb{E}(X|Y)$

Consider the setting of (2.7). We write $\mathbb{E}(X|Y)$ to denote $v(Y)$.

Since any real-valued function of a random variable is a random variable, so it makes sense to consider the expectation of $\mathbb{E}(X|Y)$:

$$\mathbb{E}(\mathbb{E}(X|Y)) = \begin{cases} \sum_{y \in \mathbb{R}: p_Y(y)>0} \mathbb{E}(X|Y=y)\, p_Y(y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} \mathbb{E}(X|Y=y)\, f_Y(y)\, \mathrm{d}y & \text{if } Y \text{ is continuous} \end{cases}. \qquad [2.2]$$

Theorem 2.5
Law of Total Expectation

Let X, Y be random variables. Then

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))\,.$$

Proof. We shall consider the continuous case only — assume X, Y are jointly continuous. Recall from the definition of $\mathbb{E}(X|Y)$ that

$$\mathbb{E}(\mathbb{E}(X|Y)) = \int_{-\infty}^{\infty} \mathbb{E}(X|Y=y)\, f_Y(y)\, \mathrm{d}y\,.$$

But

$$\begin{aligned}\mathbb{E}(X|Y=y) &= \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} x \frac{f_{(X,Y)}(x,y)}{f_Y(y)}\,\mathrm{d}x.\end{aligned}$$

It follows that

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X|Y)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,\mathrm{d}x\, f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \frac{f_{(X,Y)}(x,y)}{f_Y(y)}\,\mathrm{d}x\, f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{(X,Y)}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x \\ &= \mathbb{E}(X),\end{aligned}$$

as desired. ♦

(2.8) Suppose $X \sim \mathrm{GEO}_t(p)$ where $p \in (0,1]$. Calculate $\mathbb{E}(X)$, $\mathrm{var}(X)$ using the law of total expectation.

Answer. Recall that X is the number of iid Bernoulli trials, each with success probability p, needed to obtain the first success. So let Y be the first trial. Then observe that

$$X|(Y=1) = 1$$

but

$$X|(Y=0) = X+1.$$

By the law of total expeectation,

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X|Y)) = p_Y(0)\,\mathbb{E}(X|Y=0) + p_Y(1)\,\mathbb{E}(X|Y=1) \\ &= (1-p)\,\mathbb{E}(X+1) + p\,\mathbb{E}(1) = (1-p) + (1-p)\,\mathbb{E}(X) + p = 1 + (1-p)\,\mathbb{E}(X),\end{aligned}$$

so rearranging gives

$$\mathbb{E}(X) = \frac{1}{p}.$$

On the other hand,

$$\begin{aligned}\mathbb{E}\left(X^2\right) &= \mathbb{E}\left(\mathbb{E}\left(X^2|Y\right)\right) = p_Y(0)\,\mathbb{E}\left(X^2|Y=0\right) + p_Y(1)\,\mathbb{E}\left(X^2|Y=1\right) \\ &= (1-p)\,\mathbb{E}\left(X^2+2x+1\right) + p\,\mathbb{E}(1) = (1-p)\,\mathbb{E}\left(X^2\right) + 2(1-p)\,\mathbb{E}(X) + 1,\end{aligned}$$

so

$$\mathbb{E}\left(X^2\right) = \frac{2(1-p)\,\mathbb{E}(X)+1}{p} = \frac{\frac{2-p}{p}+1}{p} = \frac{2}{p^2} - \frac{1}{p} + \frac{1}{p} = \frac{2}{p^2}.$$

Thus

$$\mathrm{var}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 = \frac{2}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1}{p^2}.$$ ◁

Note that the obtained expectation and variance agree with the known results.

Notation 2.8

$\text{var}(X|Y)$

Let X, Y be random variables. Let $v : \mathbb{R} \to \mathbb{R}$ be defined by

$$v(y) = \text{var}(X|Y = y)$$

for all $y \in \mathbb{R}$. Then we write $\text{var}(X|Y)$ to denote $v(Y)$.

(2.9) Similar to $\mathbb{E}(X|Y)$, $\text{var}(X|Y)$ is a random variable as a function, v, of a random variable, Y. The following is a *variance analogue* of the law of total probability.

Theorem 2.6
Conditional Variance Formula

Let X, Y be random variables. Then

$$\text{var}(X) = \mathbb{E}(\text{var}(X|Y)) + \text{var}(\mathbb{E}(X|Y)).$$

Proof. First note that, for any $y \in \mathbb{R}$,

$$\text{var}(X|Y = y) = \mathbb{E}\left(X^2|Y = y\right) - \mathbb{E}(X|Y = y)^2,$$

which means

$$\text{var}(X|Y) = \mathbb{E}\left(X^2|Y\right) - \mathbb{E}(X|Y)^2.$$

On the other hand,

$$\text{var}(\mathbb{E}(X|Y)) = \mathbb{E}\left(\mathbb{E}(X|Y)^2\right) - \mathbb{E}(\mathbb{E}(X|Y))^2.$$

It follows from the law of total expectation that

$$\mathbb{E}(\text{var}(X|Y)) + \text{var}(\mathbb{E}(X|Y)) = \mathbb{E}\left(\mathbb{E}\left(X^2|Y\right)\right) - \mathbb{E}(\mathbb{E}(X|Y))^2 = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 = \text{var}(X).$$ ♦

(EX 2.10)
Random Sum

Let $(X_i)_{i=1}^{\infty}$ be an iid sequence of random variables with common mean $\mu \in \mathbb{R}$ and common variance $\sigma^2 \geq 0$ and let N be a nonnegative integer-valued random variable that is independent of $X_1, \ldots$. Let

$$T = \sum\nolimits_{i=1}^{N} X_i.$$

Find $\mathbb{E}(T), \text{var}(T)$.

Answer. By the law of total probability,

$$\begin{aligned}\mathbb{E}(T) = \mathbb{E}(\mathbb{E}(T|N)) = \mathbb{E}(\mathbb{E}(T|N = n)\,|_{n=N}) &= \mathbb{E}\left(\mathbb{E}\left(\sum\nolimits_{i=1}^{N} X_i|N = n\right)|_{n=N}\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(\sum\nolimits_{i=1}^{n} X_i|N = n\right)|_{n=N}\right) = \mathbb{E}\left(\mathbb{E}\left(\sum\nolimits_{i=1}^{n} X_i\right)|_{n=N}\right) = \mathbb{E}\left(\sum\nolimits_{i=1}^{N} X_i\right) \\ &= \mathbb{E}(\mu N) = \mu\,\mathbb{E}(N).\end{aligned}$$

Moreover,

$$\text{var}(T|N = n) = \text{var}\left(\sum\nolimits_{i=1}^{N} X_i|N = n\right) = \text{var}\left(\sum\nolimits_{i=1}^{n} X_i|N = n\right) = \text{var}\left(\sum\nolimits_{i=1}^{n} X_i\right) = n\sigma^2,$$

which means

$$\mathbb{E}(\text{var}(T|N)) = \mathbb{E}\left(N\sigma^2\right) = \sigma^2\,\mathbb{E}(N).$$

On the other hand,

$$\text{var}(\mathbb{E}(T|N)) = \text{var}(\mu N) = \mu^2\,\text{var}(N).$$

Thus
$$\operatorname{var}(T) = \mathbb{E}(\operatorname{var}(T|N)) + \operatorname{var}(\mathbb{E}(T|N)) = \sigma^2\,\mathbb{E}(N) + \mu^2 \operatorname{var}(N)$$
by the conditional variance formula. ◁

(2.11) Recall from [2.2] that, given any random variables X, Y,
$$\mathbb{E}(\mathbb{E}(X|Y)) = \begin{cases} \sum_{y\in\mathbb{R}:p_Y(y)>0} \mathbb{E}(X|Y=y)\,p_Y(y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} \mathbb{E}(X|Y=y)\,f_Y(y)\,\mathrm{d}y & \text{if } Y \text{ is continuous} \end{cases}.$$
Now, suppose that A represents some event of interest and we desire to determine $\mathbb{P}(A)$. Define an *indicator random variable* X by
$$X = \begin{cases} 0 & \text{if } A^C \text{ occurs} \\ 1 & \text{if } A \text{ occurs} \end{cases}.$$
Clearly, $\mathbb{P}(X=1) = \mathbb{P}(A), \mathbb{P}(X=0) = 1-\mathbb{P}(A)$, so that $X \sim \mathrm{B}(\mathbb{P}(A))$. Hence $\mathbb{E}(X) = \mathbb{P}(A)$ and
$$\begin{aligned} \mathbb{E}(X|Y=y) &= \sum\nolimits_{x\in\{0,1\}} x\,\mathbb{P}(X=x|Y=y) \\ &= 0\,\mathbb{P}(X=0|Y=y) + 1\,\mathbb{P}(X=1|Y=y) \\ &= \mathbb{P}(X=1|Y=y) \\ &= \mathbb{P}(A|Y=y)\,. \end{aligned}$$
for any random variable Y. Hence [2.2] becomes
$$\mathbb{P}(A) = \begin{cases} \sum_{y\in\mathbb{R}:p_Y(y)>0} \mathbb{E}(A|Y=y)\,p_Y(y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} \mathbb{E}(A|Y=y)\,f_Y(y)\,\mathrm{d}y & \text{if } Y \text{ is continuous} \end{cases} \qquad [2.3]$$
for all random variable Y.

(EX 2.12) Let X, Y be independent continuous random variables. Show that
$$\mathbb{P}(X<Y) = \int_{-\infty}^{\infty} F_X(y)\,f_Y(y)\,\mathrm{d}y\,. \qquad [2.4]$$

Proof. Let A be the event
$$A = \{X<Y\}\,.$$
Then we have
$$\begin{aligned} \mathbb{P}(X<Y) = \mathbb{P}(A) &= \int_{-\infty}^{\infty} \mathbb{P}(A|Y=y)\,f_Y(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} \mathbb{P}(X<Y|Y=y)\,f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} \mathbb{P}(X<y|Y=y)\,f_Y(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} \mathbb{P}(X<y)\,f_Y(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} \mathbb{P}(X\le y)\,f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} F_X(y)\,f_Y(y)\,\mathrm{d}y\,. \end{aligned}$$
◁

(EX 2.13) Consider the setting of (EX 2.12) and further assume that X, Y are identically distributed. Show that [2.4] simplifies to
$$\mathbb{P}(X<Y) = \frac{1}{2}. \qquad [2.5]$$

Proof. Observe that $f_X = f_Y$ since X, Y are iid, so

$$\mathbb{P}(X < Y) = \int_{-\infty}^{\infty} F_X(y) f_Y(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} F_X(y) f_X(y)\,\mathrm{d}y = \int_0^1 u\,\mathrm{d}u = \frac{1}{2}$$

by the change of variable $u = F_X(y)$. ◁

(EX 2.14) Suppose $X \sim \mathrm{EXP}(\lambda_1), Y \sim \mathrm{EXP}(\lambda_2)$ are independent. Show

$$\mathbb{P}(X < Y) = \frac{\lambda_1}{\lambda_2}. \tag{2.6}$$

Proof. Since $X \sim \mathrm{EXP}(\lambda_1), Y \sim \mathrm{EXP}(\lambda_2)$, we have

$$\begin{cases} f_y(y) & = \lambda_2 e^{-\lambda_2 y} \\ F_X(y) & = 1 - e^{-\lambda_1 y} \end{cases}$$

for all $y > 0$. It follows from [2.4] that

$$\begin{aligned}\mathbb{P}(X < Y) &= \int_{-\infty}^{\infty} F_X(y) f_Y(y)\,\mathrm{d}y = \int_0^{\infty} \left(1 - e^{-\lambda_1 y}\right) \lambda_2 e^{-\lambda_2 y}\,\mathrm{d}y = \lambda_2 \int_0^{\infty} e^{-\lambda_2 y} - e^{-(\lambda_1+\lambda_2)y}\,\mathrm{d}y \\ &= \lambda_2 \left(-\frac{1}{\lambda_2} e^{-\lambda_2 y} + \frac{1}{\lambda_1+\lambda_2} e^{-(\lambda_1+\lambda_2)y}\right)\Bigg|_{y=0}^{\infty} = 1 - \frac{\lambda_2}{\lambda_1+\lambda_2} = \frac{(\lambda_1+\lambda_2)-\lambda_2}{\lambda_1+\lambda_2} = \frac{\lambda_1}{\lambda_1+\lambda_2}.\end{aligned}$$ ◁

(EX 2.15) Suppose W, X, Y are positive independent continuous random variables and let $Z = X \mid (X < Y)$. Show that

$$\begin{aligned} U &= (W, X) \mid (W < X < Y) \\ V &= (W, Z) \mid (W < Z) \end{aligned}$$

are identically distributed.

Proof. Observe that

$$F_U(w,x) = \mathbb{P}(W \le w, X \le x | W < X < Y) = \frac{\mathbb{P}(W \le w, X \le x, W < X, X < Y)}{\mathbb{P}(W < X, X < Y)} \tag{2.7}$$

for every $w, x > 0$. By conditioning on X,

$$\begin{aligned}\mathbb{P}(W < X, X < Y) &= \int_0^{\infty} \mathbb{P}(W < X, X < Y | X = s) f_X(s)\,\mathrm{d}s \\ &= \int_0^{\infty} \mathbb{P}(W < s, s < Y | X = s) f_X(s)\,\mathrm{d}s \\ &= \int_0^{\infty} \mathbb{P}(W < s)\,\mathbb{P}(s < Y) f_X(s)\,\mathrm{d}s, \end{aligned} \tag{2.8}$$

where the last equality follows from the fact that W, X, Y are independent. In a similar manner,

$$\begin{aligned}\mathbb{P}(W \le w, X \le x, W < X, X < Y) &= \int_0^{\infty} \mathbb{P}(W \le w, X \le x, W < X, X < Y | X = s) f_X(s)\,\mathrm{d}s \\ &= \int_0^{\infty} \mathbb{P}(W \le w, s \le x, W < s, s < Y | X = s) f_X(s)\,\mathrm{d}s \\ &= \int_0^{\infty} \mathbb{P}(W \le w)\,\mathbb{P}(s \le x)\,\mathbb{P}(W < s)\,\mathbb{P}(s < Y) f_X(s)\,\mathrm{d}s \\ &= \mathbb{P}(W \le w) \int_0^{x} \mathbb{P}(W < s)\,\mathbb{P}(s < Y) f_X(s)\,\mathrm{d}s. \end{aligned} \tag{2.9}$$

Moreover, for every $z > 0$,

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \le z) = \mathbb{P}(X \le z | X < Y) = \frac{\mathbb{P}(X \le z, X < Y)}{\mathbb{P}(X < Y)} \\ &= \frac{\int_0^\infty \mathbb{P}(X \le z, X < Y | X = s) f_X(s)\,\mathrm{ds}}{\mathbb{P}(X < Y)} = \frac{\int_0^\infty \mathbb{P}(s \le z, s < Y | X = s) f_X(s)\,\mathrm{ds}}{\mathbb{P}(X < Y)} \\ &= \frac{\int_0^z \mathbb{P}(s < Y) f_X(s)\,\mathrm{ds}}{\mathbb{P}(X < Y)}, \end{aligned}$$

so by differentiating with respect to z, we obtain

$$f_Z(z) = \frac{\mathrm{d}}{\mathrm{dz}} \frac{\int_0^z \mathbb{P}(s < Y) f_X(s)\,\mathrm{ds}}{\mathbb{P}(X < Y)} = \frac{\mathbb{P}(z < Y) f_X(z)}{\mathbb{P}(X < Y)}. \tag{2.10}$$

Now note that the cdf of V is given by

$$F_V(w, z) = \mathbb{P}(W \le w, Z \le z | W < Z) = \frac{\mathbb{P}(W \le w, Z \le z, W < Z)}{\mathbb{P}(W < Z)} \tag{2.11}$$

for every $w, z > 0$. Since W independent of X, Y, it is independent of $Z = X | (X < Y)$, so

$$\begin{aligned} \mathbb{P}(W < Z) &= \int_0^\infty \mathbb{P}(W < Z | Z = s) f_Z(s)\,\mathrm{dz} = \int_0^\infty \mathbb{P}(W < s | Z = s) f_Z(s)\,\mathrm{ds} \\ &= \int_0^\infty \mathbb{P}(W < s) f_Z(s)\,\mathrm{ds} = \int_0^\infty \mathbb{P}(W < s) \frac{\mathbb{P}(s < Y) f_X(s)}{\mathbb{P}(X < Y)}\,\mathrm{ds} \\ &\overset{[2.8]}{=} \frac{\mathbb{P}(W < X, X < Y)}{\mathbb{P}(X < Y)}. \end{aligned} \tag{2.12}$$

Furthermore,

$$\begin{aligned} \mathbb{P}(W \le w, Z \le z, W < Z) &= \int_0^\infty \mathbb{P}(W \le w, Z \le z, W < Z | Z = s) f_Z(s)\,\mathrm{ds} \\ &= \int_0^\infty \mathbb{P}(W \le w, s \le z, W < s | Z = s) f_Z(s)\,\mathrm{ds} \\ &= \mathbb{P}(W \le w) \int_0^z \mathbb{P}(W < s) f_Z(s)\,\mathrm{ds} \\ &\overset{[2.10]}{=} \int_0^z \mathbb{P}(W < s) \frac{\mathbb{P}(Y > s) f_X(s)}{\mathbb{P}(X < Y)}\,\mathrm{ds} \\ &\overset{[2.9]}{=} \frac{\mathbb{P}(W \le w, X \le x, W < X, X < Y)}{\mathbb{P}(X < Y)} \end{aligned} \tag{2.13}$$

for every $w, z > 0$. Thus

$$\begin{aligned} F_V(w, z) &\overset{[2.11]}{=} \frac{\mathbb{P}(W \le w, Z \le z, W < Z)}{\mathbb{P}(W < Z)} = \frac{\frac{\mathbb{P}(W \le w, X \le z, W < X, X < Y)}{\mathbb{P}(X < Y)}}{\frac{\mathbb{P}(W < X, X < Y)}{\mathbb{P}(X < Y)}} \\ &\overset{\substack{[2.12] \\ [2.13]}}{=} \frac{\mathbb{P}(W \le w, X \le z, W < X, X < Y)}{\mathbb{P}(W < X, X < Y)} \overset{[2.7]}{=} F_U(w, z) \end{aligned}$$

for every $w, z > 0$, so $V \sim U$. ◁

(EX 2.16) Consider an experiment in which iid trials, each with success probability $p \in (0,1]$, are performed until $k \in \mathbb{N}$ consecutive successes are observed. Dtermine the expectation of the number of trials needed to achieve k consecutive successes.

Answer. For each $l \in \mathbb{N}$, let N_l denote the number of trials required to achieve l consecutive successes, where we desire to find $\mathbb{E}(N_k)$. First note that $N_1 \sim \text{GEO}(p)$, so

$$\mathbb{E}(N_1) = \frac{1}{p}. \tag{2.14}$$

For the general case, the idea is to condition on N_{l-1}: fix $l \in \mathbb{N}, l \geq 2$ and observe that

$$\mathbb{E}(N_l) = \mathbb{E}(\mathbb{E}(N_l | N_{l-1}))$$

from the law of total expectation. Define, for every $n \in \mathbb{N}$,

$$Y | (N_{l-1} = n) = \begin{cases} 0 & \text{if } n+1\text{th trial is a failure} \\ 1 & \text{if } n+1\text{th trial is a success} \end{cases}.$$

Then, for every $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}(N_l | N_{l-1}) &\overset{[2.3]}{=} \sum\nolimits_{y \in \{0,1\}} \mathbb{E}(N_l | N_{l-1} = n, Y = y) \mathbb{P}(Y = y | N_{l-1} = n) \\ &= p \mathbb{E}(N_l | N_{l-1} = n, Y = 1) + (1-p) \mathbb{E}(N_l | N_{l-1} = n, Y = 0) \\ &= p(n+1) + (1-p)(n+1+\mathbb{E}(N_l)) \\ &= n+1+(1-p)\mathbb{E}(N_l), \end{aligned}$$

since

$$\begin{aligned} N_l | (N_{l-1} = n, Y = 0) &\sim n+1+N_l, \\ N_l | (N_{l-1} = n, Y = 1) &\sim n+1. \end{aligned}$$

This implies

$$\mathbb{E}(N_l) = \mathbb{E}(\mathbb{E}(N_l | N_{l-1})) = \mathbb{E}(N_{l-1}) + 1 + (1-p)\mathbb{E}(N_l),$$

so

$$\mathbb{E}(N_l) = \frac{\mathbb{E}(N_{l-1}) + 1}{p}. \tag{2.15}$$

Now the claim is that

$$\mathbb{E}(N_l) = \sum\nolimits_{r=1}^{l} \frac{1}{p^r}. \tag{2.16}$$

To verify this, note that the base case is provided by [2.14]. Moreover, for every $l \in \mathbb{N}, l \geq 2$,

$$\mathbb{E}(N_l) = \frac{\sum_{r=1}^{l-1} \frac{1}{p^r} + 1}{p} = \frac{1}{p} + \sum\nolimits_{r=1}^{l-1} \frac{1}{p^{r+1}} = \sum\nolimits_{r=1}^{l} \frac{1}{p^r}$$

by induction. Thus by [2.16],

$$\mathbb{E}(N_k) = \sum\nolimits_{r=1}^{k} \frac{1}{p^r}.$$

◁

This page intentionally left blank.

3.
Markov Chains

3.1 Markov Chains

Def'n 3.1 **Stochastic Process**

Any collection of random variables (or random vectors) of the form $\{X(t)\}_{t\in\mathcal{T}}$ is called a ***stochastic process***.

(3.1) Given a stochastic process $\{X(t)\}_{t\in\mathcal{T}}$ The index set $\mathcal{T}$ is often interpreted in the context of time. As such, usually $\mathcal{T}\subseteq\mathbb{R}$ and we say $X(t)$ is the ***state*** of the process at time $t\in\mathcal{T}$.

Def'n 3.2 **Continuous-time, Discrete-time** Stochastic Process

Let $\{X(t)\}_{t\in\mathcal{T}}$ be a stochastic process. We say $\{X\}_{t\in\mathcal{T}}$ is

(a) ***continuous-time*** if $\mathcal{T}$ is a (union of) continuum of real numbers; and

(b) ***discrete-time*** if $\mathcal{T}$ is a countable subset of real numbers.[a]

[a] In general, we use $\mathbb{N}\cup\{0\}$ as the index set of discrete-time stochastic processes. In fact, we shall use this convention throughout this note, unless otherwise specified.

Def'n 3.3 **Discrete-time Markov Chain (DTMC)**

We say a discrete-time stochastic process $\{X_n\}_{n\in\mathbb{N}\cup\{0\}}$ is a ***discrete-time Markov chain*** (***DTMC***) if

(a) each X_n is discrete; and

(b) for every $n\in\mathbb{N}\cup\{0\}$ and $x_0,\ldots,x_{n+1}$ in the codomain of $X_0,\ldots,X_{n+1}$, respectively,

$$\mathbb{P}(X_{n+1}=x_{n+1}|X_n=x_n,\ldots,X_0=x_0)=\mathbb{P}(X_{n+1}=x_{n+1}|X_n=x_n). \qquad \textit{Markov property}$$

(3.2) In other words, the Markov property states that the conditional distribution of a *future* state X_{n+1} given the *past* states $X_0,\ldots,X_{n-1}$ and the *present* state X_n is independent of the past states. It is also worth noting that the Markov property ensures that, given any $k_1,\ldots,k_l\in\{1,\ldots,n-1\}$ with $k_1<\cdots<k_l$,

$$\mathbb{P}(X_{n+1}=x_{n+1}|X_{k_l}=x_{k_l},\ldots,X_{k_1}=x_{k_1})=\mathbb{P}(X_{n+1}=x_{n+1}|X_{k_l}=x_{k_l}).$$

Def'n 3.4 **Transition Probability Matrix**

For any pair of states $i,j\in\mathbb{N}\cup\{0\}$, the ***transition probability*** from state i at time n to state j at time $n+1$ is given by

$$\mathbb{P}(X_{n+1}=j|X_n=i)$$

for all $n\in\mathbb{N}\cup\{0\}$. The ***transition probability matrix*** from time n to time $n+1$ is defined as

$$\begin{bmatrix} P_{n,0,0} & P_{n,0,1} & \cdots \\ P_{n,1,0} & P_{n,1,1} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

for all $n\in\mathbb{N}\cup\{0\}$, where $P_{n,i,j}=\mathbb{P}(X_{n+1}=j|X_n=i)$ for all $i,j\in\mathbb{N}\cup\{0\}$.

(3.3) It is clear from the construction that, given any TPM P,

(a) every entry of P is nonnegative; and

(b) for any row of P, the sum of the entries is 1.

Any matrix that satisfies (a), (b) is called ***stochastic***.

Stationary (Homogeneous) DTMC

Def'n 3.5 Let $\{X_n\}_{n\in\mathbb{N}\cup\{0\}}$ be a DTMC. We say $\{X_n\}_{n\in\mathbb{N}\cup\{0\}}$ is ***stationary*** (or ***homogeneous***) if the transition probability is independent of the time.[a] That is, for all times $n, m \in \mathbb{N}\cup\{0\}$ and indices $i, j \in \mathbb{N}\cup\{0\}$,

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_{m+1} = j | X_n = i).$$

[a]We shall only consider stationary DTMCs in this note.

(EX 3.4) On a given day the weather is clear, overcast, or rainy. If the weather is clear today, then it would be clear, overcast, or rainy tomorrow with respective probabilities $0.6, 0.3, 0.1$. If the weather is overcast today, then it would be clear, overcast, or rainy tomorrow with respective probabilities $0.2, 0.5, 0.3$. If the weather is rainy today, then it would be clear, overcast, or rainy tomorrow with respective probabilities $0.4, 0.2, 0, 4$. Construct the underlying DTMC and determine its TPM.

Answer. Note that the weather tomorrow only depends on the weather today, implying that the Markov property holds. Hence, letting

$$X_n = \begin{cases} 0 & \text{if the weather on } n\text{th day is clear} \\ 1 & \text{if the weather on } n\text{th day is overcast} \\ 2 & \text{if the weather on } n\text{th day is rainy} \end{cases},$$

$(X_n)_{n\in\mathbb{N}\cup\{0\}}$ is a 3-state DTMC. Moreover, the TPM is given by

$$\begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.2 & 0.4 \end{bmatrix}.$$ ◁

***n*-step Transition Probability**

Def'n 3.6 Suppose that we have a DTMC $\{X_n\}_{n\in\mathbb{N}\cup\{0\}}$. For every states $i, j \in \mathbb{N}\cup\{0\}$ and $n \in \mathbb{N}\cup\{0\}$, we define the n-***step transition probability***, commonly denoted as $P_{i,j}^{(n)}$, as

$$P_{i,j}^{(n)} = \mathbb{P}(X_{m+n} = j | X_m = i),$$

where $m \in \mathbb{N}\cup\{0\}$.[a] We call

$$P^{(n)} = \left[P_{i,j}^{(n)}\right]_{i,j\in\mathbb{N}\cup\{0\}}$$

the n-***step transition probability matrix*** (n-***step TPM***).

[a]The definition is independent of m since we assumed our DTMC to be stationary. In other words, we may define $P_{i,j}^{(n)} = \mathbb{P}(X_n = j | X_0 = i)$.

(3.5) Consider a DTMC $\{X_n\}_{n\in\mathbb{N}\cup\{0\}}$, its TPM P, and n-step TPMs $P^{(0)},\ldots$.

(a) From the construction, it is evident that

$$P_{i,j}^{(0)} = \delta_{ij}$$

for every states i, j, where δ is the Kronecker delta. It follows that $P^{(0)}$ is the identity matrix.

(b) $P^{(1)} = P$.

(3.6) Chapman-Kolmogorov Equations

For any $n \in \mathbb{N}$, we have

$$P_{i,j}^{(n)} = \sum_{k=0}^{\infty} P_{i,k}^{(n-1)} P_{k,j}. \tag{3.1}$$

Proof. Observe that

$$\begin{aligned}
P_{i,j}^{(n)} &= \mathbb{P}(X_n = j | X_0 = i) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(X_n = j | X_{n-1} = k, x_0 = i)\,\mathbb{P}(X_{n-1} = k | X_0 = i) \\
&= \sum_{k=0}^{\infty} P_{i,k}^{(n-1)}\,\mathbb{P}(X_n = j | X_{n-1} = k, X_0 = i) \\
&= \sum_{k=0}^{\infty} P_{i,k}^{(n-1)}\,\mathbb{P}(X_n = j | X_{n-1} = k) && \text{by Markov property} \\
&= \sum_{k=0}^{\infty} P_{i,k}^{(n-1)} P_{k,j},
\end{aligned}$$

as required. ◁

This in particular implies that,

$$P^{(n)} = P^{(n-1)} P \tag{3.2}$$

for every $n \in \mathbb{N}$, and as a corollary,

Chapman-Komogorov Equations for a DTMC

$$P_{i,j}^{(n)} = \sum_{k=0}^{\infty} P_{i,k}^{(m)} P_{k,j}^{(n-m)} \tag{3.3}$$

for every $i, j \in \mathbb{N}\cup\{0\}$ and $n \in \mathbb{N}, m \in \{0,\ldots,n\}$. In matrix form, this translates to

Chapman-Komogorov Equations in Matrix Form

$$P^{(n)} = P^{(m)} P^{(n-m)}. \tag{3.4}$$

(3.7) Consider the row vector

$$\alpha_n = \begin{bmatrix} \alpha_{n,0} & \alpha_{n,1} & \cdots \end{bmatrix}$$

for every $n \in \mathbb{N}\cup\{0\}$, where

$$\alpha_{n,k} = \mathbb{P}(X_n = k)$$

for every $k \in \mathbb{N}$. In other words, α_n represents the marginal pmf of X_n, and as a consequence,

$$\sum_{k=0}^{\infty} \alpha_{n,k} = 1.$$

In case $n = 0$, α_0 is referred to as the ***initial conditions*** (or ***initial probability row vector***) of the DTMC. Now let us see how we can calculate α_n. For every $n \in \mathbb{N}, m \in \{0, \ldots, n\}$, note that

$$\begin{aligned}
\alpha_{n,k} &= \mathbb{P}(X_n = k) \\
&= \sum_{i=0}^{\infty} \mathbb{P}(X_n = k | X_m = i) \mathbb{P}(X_m = i) \\
&= \sum_{i=0}^{\infty} \alpha_{m,i} \mathbb{P}(X_{n-m} = k | X_0 = i) && \text{since the DTMC is stationary} \\
&= \sum_{i=0}^{\infty} \alpha_{m,i} P_{i,k}^{(n-m)}.
\end{aligned}$$

In matrix form,

$$\alpha_n = \alpha_m P^{(n-m)} = \alpha_m P^{n-m},$$

or

Marginal PDF of X_n

$$\alpha_n = \alpha_0 P^n. \tag{3.5}$$

(3.8) Having knowledge of the initial conditions and the one-step transition probabilities, one can calculate various probabilites of possible interest, such as

$$\begin{aligned}
\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0) &= \mathbb{P}(X_0 = x_0) \mathbb{P}(X_1 = x_1 | X_0 = x_0) \cdots \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) \\
&= \mathbb{P}(X_0 = x_0) \mathbb{P}(X_1 = x_1 | X_0 = x_0) \cdots \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \\
&= \alpha_{0,x_0} P_{x_0,x_1} \cdots P_{x_{n-1},x_n}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbb{P}(X_{n+m} = x_{n+m}, \ldots, X_{n+1} = x_{n+1} | X_n = x_n) \\
&= \frac{\mathbb{P}(X_{n+m} = x_{n+m}, \ldots, X_n = x_n)}{\mathbb{P}(X_n = x_n)} \\
&= \frac{\mathbb{P}(X_n = x_n) \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n) \cdots \mathbb{P}(X_{n+m} = x_{n+m} | X_{n+m-1} = x_{n+m-1}, \ldots, X_n = x_n)}{\mathbb{P}(X_n = x_n)} \\
&= P_{x_n,x_{n+1}} \cdots P_{x_{n+m-1},x_{n+m}}.
\end{aligned}$$

The key observation is that the DTMC is *completely characterized* by its one-step TPM P and the initial conditions α_0.

(EX 3.9) A particle moves along the states $0, 1, 2$ according to a DTMC whose TPM P is given by

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0.5 & 0 & 0.5 \end{bmatrix}.$$

Let X_n denote the position of the particle after the nth move (i.e. the DTMC is $\{X_n\}_{n \in \mathbb{N} \cup \{0\}}$). Suppose that the particle is equally likely to start in any of the three positions.

(a) Calculate $\mathbb{P}(X_3 = 1 | X_0 = 0)$.

Answer. We desire to find $P_{0,1}^{(3)}$. To get this, we proceed to calculate $P^{(3)}$, the 3-step transition TPM, which satisfies

$$P^{(3)} = P^3$$

where P is the TPM. First of all,

$$P^2 = \begin{bmatrix} 0.54 & 0.26 & 0.2 \\ 0.2 & 0.36 & 0.44 \\ 0.6 & 0.1 & 0.3 \end{bmatrix}$$

and

$$P^3 = \begin{bmatrix} 0.478 & 0.264 & 0.258 \\ 0.36 & 0.256 & 0.384 \\ 0.57 & 0.18 & 0.25 \end{bmatrix}$$

by direct calculations. Thus,

$$P_{0,1}^{(3)} = P_{0,1}^3 = 0.264.$$ $\triangleleft$

(b) Calculate $\mathbb{P}(X_4 = 2)$.

Answer. We desire to find

$$\alpha_{4,2} = \mathbb{P}(X_4 = 2).$$

To do so, let us calculate α_4, which satisfies

$$\alpha_4 = \alpha_0 P^4.$$

By a direct calculation,

$$P^4 = \begin{bmatrix} \frac{1159}{2500} & \frac{127}{500} & \frac{353}{1250} \\ \frac{111}{250} & \frac{141}{625} & \frac{413}{1250} \\ \frac{131}{250} & \frac{111}{500} & \frac{127}{500} \end{bmatrix},$$

so

$$\alpha_4 = \alpha_0 P^4 = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{1159}{2500} & \frac{127}{500} & \frac{353}{1250} \\ \frac{111}{250} & \frac{141}{625} & \frac{413}{1250} \\ \frac{131}{250} & \frac{111}{500} & \frac{127}{500} \end{bmatrix} = \begin{bmatrix} \frac{1193}{2500} & \frac{877}{3750} & \frac{2167}{7500} \end{bmatrix}.$$

Thus

$$\alpha_{4,2} = \frac{2167}{7500}.$$ $\triangleleft$

(c) Calculate $\mathbb{P}(X_6 = 0, X_4 = 2)$.

Answer. We have

$$\mathbb{P}(X_6 = 0, X_4 = 2) = \mathbb{P}(X_4 = 2)\,\mathbb{P}(X_6 = 0 | X_4 = 2) = \alpha_{4,2} P_{2,0}^{(2)} = 0.17336.$$ $\triangleleft$

(d) Calculate $\mathbb{P}(X_9 = 0, X_7 = 2 | X_5 = 1, X_2 = 0)$.

Answer. We have

$$\begin{aligned} &\mathbb{P}(X_9 = 0, X_7 = 2 | X_5 = 1, X_2 = 0) \\ &= \mathbb{P}(X_7 = 2 | X_5 = 1, X_4 = 0)\,\mathbb{P}(X_9 = 0 | X_7 = 2, X_5 = 1, X_2 = 0) \\ &= \mathbb{P}(X_7 = 2 | X_5 = 1)\,\mathbb{P}(X_9 = 0 | X_7 = 2) = P_{1,2}^{(2)} P_{2,0}^{(2)} = 0.264. \end{aligned}$$ $\triangleleft$

3.2 Accessibility and Communication

Accessble State

Def'n 3.7 Let i, j be states of a DTMC with n-step TPMs $P^{(n)}$.

(a) We say j is ***accessible*** from state i, denoted as $i \to j$, if there exists $n \in \mathbb{N} \cup \{0\}$ such that $P_{i,j}^{(n)} > 0$.

(b) We say i, j ***communicate***, denoted as $i \leftrightarrow j$ if $i \to j, j \to i$.

(3.10) In terms of accessibility, note that the magnitude of the components of P do not matter. All that matters is which are positive and which are 0. In particular, if state j is not accessible from state i, then $P_{i,j}^{(n)} = 0$ for every $n \in \mathbb{N} \cup \{0\}$, and

$$\mathbb{P}\left(\exists m \in \mathbb{N} \cup \{0\} \, [X_m = j] \, | X_0 = i\right) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N} \cup \{0\}} \{X_n = j\} \, | X_0 = i\right)$$
$$\leq \sum_{n=0}^{\infty} \mathbb{P}\left(X_n = j | X_0 = i\right) = \sum_{n=0}^{\infty} P_{i,j}^{(n)} = 0.$$

In other words, if j is not accessible from i, then the probability that the DTMC ever visits state j given $X_0 = i$ is 0.

(3.11) Communication is an *equivalence relation*. That is, given any states i, j, k,

(a) $i \leftrightarrow i$; *reflexivity*

(b) $i \leftrightarrow j$ implies $j \leftrightarrow i$; and *symmetry*

(c) $i \leftrightarrow j, j \leftrightarrow k$ implies $i \leftrightarrow k$. *transitivity*

Proof. (a), (b) are clear. To show transitivity, we know that there are $n, m \in \mathbb{N} \cup \{0\}$ such that $P_{i,j}^{(n)}, P_{j,k}^{(m)} > 0$. Then by the Chapman-Kolmogorov equations,

$$P_{i,k}^{(n+m)} = \sum_{l=0}^{\infty} P_{i,l}^{(n)} P_{l,k}^{(m)} \geq P_{i,j}^{(n)} P_{j,k}^{(m)} > 0.$$

Hence $i \to k$. Using the same logic, $k \to i$. Thus $i \leftrightarrow k$. ◁

The fact that communication forms an equivalence relation allows us to *partition* all the states of a DTMC into equivalence classes, called ***communication classes***, so that within each class, all states communicate. For any states i, j belong to distinct classes, *at most* one of $i \to j, j \to i$ holds.

Irreducible, Reducible DTMC

Def'n 3.8 A DTMC is called

(a) ***irreducible*** if it has only one communcation class; and

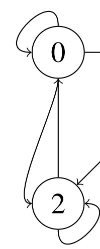
(b) ***reducible*** if not irreducible.

(EX 3.12) Suppose that the TPM P of a DTMC is

$$P = [P_{i,j}]_{i,j=0}^{2} \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0.5 & 0 & 0.5 \end{bmatrix}.$$

Find the communication classes of the DTMC.

Answer. We are going to draw a *state transition diagram*.



Thus $\{0,1,2\}$ is the only communication class of the DTMC; in other words, the DTMC is irreducible. ◁

(EX 3.13) Consider a DTMC with TPM

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

Find the communication classes of this DTMC.

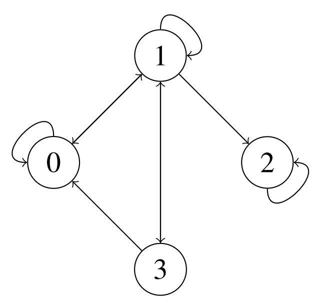
Answer. By drawing a state transition diagram, it is clear that the DTMC is irreducible. ◁

(EX 3.14) Consider a DTMC with TPM

$$P = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & 1 & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 \end{bmatrix}.$$

Find the communication classes of this DTMC.

Answer. Observe that the state transition diagram is



Thus $\{0,1,3\}, \{2\}$ are the communication classes of the DTMC. ◁

3.3 Periodicity

Period of a State of a DTMC

Def'n 3.9 Let P be the TPM of a DTMC. Given a state i of the DTMC, we define the ***period*** of i, denoted as $d(i)$, by

$$d(i) = \begin{cases} \gcd\left\{n \in \mathbb{N} : P_{i,i}^{(n)} > 0\right\} & \text{if there is } n \in \mathbb{N} \text{ such that } P_{i,i}^{(n)} > 0 \\ \infty & \text{otherwise} \end{cases}.$$

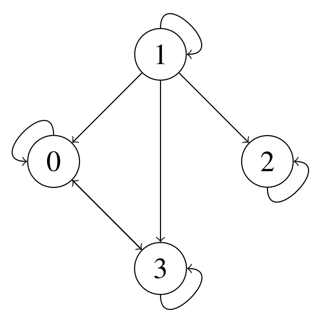
If $d(i) = 1$, then we say i is ***aperiodic***, and if every state of a DTMC is aperiodic, then we call the DTMC ***aperiodic***.

(EX 3.15) Consider a DTMC with TPM

$$P = \begin{bmatrix} \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & 1 & 0 \\ \frac{3}{4} & 0 & 0 & \frac{1}{4} \end{bmatrix}.$$

Determine the communication classes of this DTMC and the period of each state.

Answer. Note that the state transition diagram of the DTMC is the following.



Hence the communication classes are $\{0,3\}, \{1\}, \{2\}$. Moreover, we note that

$$d(0) = d(1) = d(2) = d(3) = 1,$$

since

$$P(1)_{i,i} = P_{i,i} > 0$$

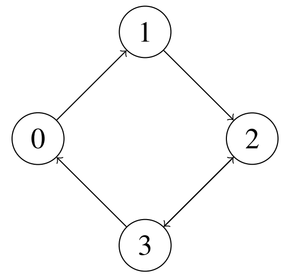
for all $i \in \{0,1,2,3\}$. Thus we conclude that the DTMC is aperiodic. ◁

(EX 3.16) Consider a DTMC with TPM

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

Determine the period of each state.

Answer. The state transition diagram of the DTMC is the following.



Note that it is clear from the diagram that
$$P_{i,i}^{(n)} > 0$$
only if n is even for every $i \in \{0,1,2,3\}$. This means $d(i) \in \{2,4\}$ for all $i \in \{0,1,2,3\}$. For each $i \in \{0,1\}$, note that $P_{i,i}^{(4)}, P_{i,i}^{(6)} > 0$, so
$$d(0) = d(1) = 2.$$
Moreover, for each $i \in \{2,3\}$, note that $P_{i,i}^{(2)}, P_{i,i}^{(4)} > 0$, so
$$d(2) = d(3) = 2.$$

Thus $d(i) = 2$ for all $i \in \{0,1,2,3\}$. $\triangleleft$

(EX 3.17) Consider the DTMC with TPM
$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$
Find the communication classes of this DTMC and determine the period of each state.

Answer. It is clear from the definition of P that
$$0 \leftrightarrow 1, 2 \leftrightarrow 3.$$

Also note that P is a block diagonal matrix of the form
$$P = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$
for some $A, B \in M_{2\times 2}(\mathbb{R})$. This means
$$P^n = \begin{bmatrix} A^n & 0 \\ 0 & B^n \end{bmatrix}$$
for every $n \in \mathbb{N} \cup \{0\}$, so the communication classes are $\{0,1\}, \{2,3\}$. Moreover, $P_{0,0}, P_{1,1} > 0$, so $d(0) = d(1) = 1$. Lastly,
$$B^n = \begin{cases} I_2 & \text{if } n \text{ is even} \\ B & \text{if } n \text{ is odd} \end{cases},$$
so $d(2) = d(3) = 2$. $\triangleleft$

Proposition 3.1
Communication Preserves Periodicity

If i, j are states of a DTMC that communicates, then $d(i) = d(j)$.

Proof. Since the result is clearly true when $i = j$, assume $i \neq j$. Since $i \leftrightarrow j$, we know by definition that

$$P_{j,i}^{(m)}, P_{i,j}^{(n)} > 0$$

for some $n, m \in \mathbb{N}$. Moreover, since $i \leftrightarrow j$ means $i \to j$ and $j \to i$, there exists $s \in \mathbb{N}$ such that

$$P_{j,j}^{(s)} > 0.$$

Note that

(a) $P_{i,i}^{(n+m)} \geq P_{i,j}^{(n)} P_{j,i}^{(m)} > 0$; and

(b) $P_{i,i}^{(n+m+s)} \geq P_{i,j}^{(n)} P_{j,j}^{(s)} P_{j,i}^{(m)} > 0$.

(a), (b) implies that

$$d(i) | s,$$

and in particular that

$$d(i) | d(j).$$

By using the same argument, we also conclude that

$$d(j) | d(i).$$

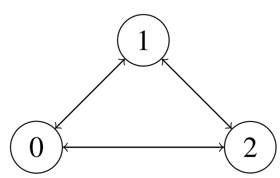
Thus $d(j) = d(i)$, as required. ♦

(EX 3.18) Consider a DTMC with TPM

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

Find the communication classes of this DTMC and determine the period of each state.

Answer. The following is the state transition diagram of the DTMC.



This means $\{0, 1, 2\}$ is the only communication class. Moreover, note that $0 \to 1 \to 0$ and $0 \to 1 \to 2 \to 0$ are cycles of lengths $2, 3$, respectively, so $d(0) = \gcd\{2, 3, \ldots\} = 1$. It follows from Proposition 3.1 that

$$d(1) = d(2) = d(0) = 1.$$ ◁

(3.19) As (EX 3.18) shows, it is possible to observe aperiodic behavior even though the main diagonal components of the TPM are zero. Generally, if $d(i) = k$, then it does not necessarily imply that $P_{i,i}^{(k)} > 0$. Instead, it implies that if the DTMC is in state i at time 0, then it is impossible to observe the DTMC in state i at time $n \in \mathbb{N}$ is n is not a multiplie of k.

3.4 Transience and Recurrence

(3.20) We desire to take a closer look at the likelihood of a DTMC beginning in some state of returning to that particular state. To that end, let us consider the probability that, starting from state i, the first visit of the DTMC to state j occurs at time $n \in \mathbb{N}$, denoted as $f_{i,j}^{(n)}$.

Notation 3.10 $f_{i,j}^{(n)}$

Consider the setting of (3.20). We write $f_{i,j}^{(n)}$ to denote

$$f_{i,j}^{(n)} = \mathbb{P}\left(X_n = j, \forall m \in \{n-1, \dots, 1\} \left[X_m \neq j\right] | X_0 = i\right)$$

for all $i, j \in \mathbb{N} \cup \{0\}$.

It is clear from Notation 3.10 that

$$f_{i,j}^{(1)} = \mathbb{P}\left(X_1 = j | X_0 = i\right) = P_{i,j},$$

where P is the TPM of the DTMC. For every $n \geq 2$, the determination of $f_{i,j}^{(n)}$ becomes more complicated, and so we desire to construct a procedure which will enable us to compute $f_{i,j}^{(n)}$. To do so, we consider the quantity $P_{i,j}^{(n)}$ and condition on the time that the first visit to state j is made:

$$\begin{aligned}
P_{i,j}^{(n)} &= \mathbb{P}\left(X_n = j | X_0 = i\right) \\
&= \sum_{k=1}^{n} \mathbb{P}\left(X_n = j, \text{first visit to } j \text{ occurs at } k | X_0 = i\right) \\
&= \sum_{k=1}^{n} \mathbb{P}\left(X_n = j, X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j | X_0 = i\right) \\
&= \sum_{k=1}^{n} \mathbb{P}\left(X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j | X_0 = j\right) \mathbb{P}\left(X_n = j | X_k = j\right) && \text{by the Markov property} \\
&= \sum_{k=1}^{n} f_{i,j}^{(k)} P_{j,j}^{(n-k)}.
\end{aligned}$$

This means

$$P_{i,j}^{(n)} = f_{i,j}^{(n)} P_{j,j}^{(0)} + \sum_{k=1}^{n-1} f_{i,j}^{(k)} P_{j,j}^{(n-k)} == f_{i,j}^{(n)} + \sum_{k=1}^{n-1} f_{i,j}^{(k)} P_{j,j}^{(n-k)}.$$

Rearranging with respect to $f_{i,j}^{(n)}$ gives

A Recursive Formula for $f_{i,j}^{(n)}$

$$f_{i,j}^{(n)} = P_{i,j}^{(n)} + \sum_{k=1}^{n-1} f_{i,j}^{(k)} P_{j,j}^{(n-k)}. \qquad [3.6]$$

When $n \geq 2$, [3.6] yields a recursive means to compute $f_{i,j}^{(n)}$.

Notation 3.11 $f_{i,j}$

Given a DTMC, let $f_{i,j}$ denote

$$f_{i,j} = \mathbb{P}\left(\exists n \in \mathbb{N}\left[X_n = j\right] | X_0 = i\right).$$

Note that

$$\begin{aligned} f_{i,j} &= \sum_{k=1}^{\infty} \mathbb{P}\left(\exists n \in \mathbb{N}\left[X_n = j\right], X_k = j, X_{k-1} \neq j, \ldots, X_1 \neq j | X_0 = i\right) \\ &= \sum_{k=1}^{\infty} \mathbb{P}\left(X_k = j, X_{k-1} \neq j, \ldots, X_1 \neq j | X_0 = i\right) \\ &= \sum_{k=1}^{\infty} f_{i,j}^{(k)} \\ &\leq 1. \end{aligned}$$

This leads to the following important concept in the study of Markov chains.

Def'n 3.12 **Transient, Recurrent** State

Given a state i of a DTMC, we say i is

(a) ***transient*** if $f_{i,i} < 1$; and

(b) ***recurrent*** if $f_{i,i} = 1$.

(3.21) Characterizing Transience and Recurrence

In what follows, we proceed to look at alternative ways of characterizing the notions of transience and recurrence. As such, let us first define M_i be a random variable which counts the number of (future) times the DTMC visits state i, disregarding the possibility of starting in state i at time 0. If we assume that $f_{i,i} < 1$, then the Markov property and stationary assumption imply that

$$\mathbb{P}\left(M_i = k | X_0 = i\right) = \left(\prod_{n=1}^{k} f_{i,i}\right)\left(1 - f_{i,i}\right) = f_{i,i}^{k}\left(1 - f_{i,i}\right) \qquad [3.7]$$

for every $k \in \mathbb{N} \cup \{0\}$, as the DTMC will return to state i k times with probability $f_{i,i}$ and then never return with probability $1 - f_{i,i}$. But note that [3.7] is the pmf of $\text{GEO}_f\left(1 - f_{i,i}\right)$ (i.e. $M_i | \left(X_0 = i\right) \sim \text{GEO}_f\left(1 - f_{i,i}\right)$). This implies

$$\mathbb{E}\left(M_i | X_0 = i\right) = \frac{1 - \left(1 - f_{i,i}\right)}{1 - f_{i,i}} = \frac{f_{i,i}}{1 - f_{i,i}}$$

since $f_{i,i} < 1$. On the other hand, if $f_{i,i} = 1$, then $\mathbb{P}\left(M_i > k | X_0 = i\right) = 1$ for all $k \in \mathbb{N}$, implying that

$$\mathbb{E}\left(M_i | X_0 = i\right) = \infty.$$

To obtain another characterization, we may define a sequence of indicator random variables $\left(A_n\right)_{n=1}^{\infty}$ by

$$A_n = \begin{cases} 0 & \text{if } X_n \neq i \\ 1 & \text{if } X_n = i \end{cases}$$

for all $n \in \mathbb{N}$. With this definition,

$$M_i = \sum_{n=1}^{\infty} A_n.$$

This means

$$\begin{aligned} \mathbb{E}\left(M_i | X_0 = i\right) = \mathbb{E}\left(\sum_{n=1}^{\infty} A_n | X_0 = i\right) &= \sum_{n=1}^{\infty} \mathbb{E}\left(A_n | X_0 = i\right) \\ &= \sum_{n=1}^{\infty} 0 \cdot \mathbb{P}\left(A_n = 0 | X_0 = i\right) + 1 \cdot \mathbb{P}\left(A_n = 1 | X_0 = i\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(X_n = i | X_0 = i\right) \\ &= \sum_{n=1}^{\infty} P_{i,i}^{(n)}. \end{aligned}$$

We summarize our characterizations into the following proposition.

Proposition 3.2
Characterizations of Transience

Let i be a state of a DTMC. The following are equivalent.[a]

(a) *i is transient.*

(b) *$\mathbb{E}(M_i|X_0=i)$ is finite, where M_i is the number of (future) times the DTMC visits state i.*

(c) *The series $\sum_{n=1}^{\infty} P_{i,i}^{(n)}$ is convergent.*

[a] Of course, negations of (b), (c) are characterizations of recurrence.

In other words, a transient state will only be visited *finitely often*.

Proposition 3.3
Communication Preserves Recurrence

Let i, j be states that communicate. Then i is recurrent if and only if j is recurrent.

Proof. It suffices to show that when i is recurrent, then so is j. So assume that i is recurrent. Since $i \leftrightarrow j$, there exists $m, n \in \mathbb{N} \cup \{0\}$ such that

$$P_{i,j}^{(m)}, P_{j,i}^{(n)} > 0.$$

Also, since i is recurrent, we know that the series $\sum_{k=1}^{\infty} P_{i,i}^{(k)}$ is divergent. Now, note that

$$P_{j,j}^{(n+k+m)} \geq P_{j,i}^{(n)} P_{i,i}^{(k)} P_{i,j}^{(m)}$$

for every $k \in \mathbb{N}$. This means the series

$$\sum_{l=n+m+1}^{\infty} P_{j,j}^{(l)} = \sum_{k=1}^{\infty} P_{j,j}^{(n+k+m)} = P_{j,i}^{(n)} P_{i,j}^{(m)} \sum_{k=1}^{\infty} P_{i,i}^{(k)}$$

is divergent, since $P_{i,j}^{(m)}, P_{j,i}^{(n)} > 0$, so $\sum_{l=1}^{\infty} P_{j,j}^{(l)}$ is also divergent. Thus j is recurrent, as required. ♦

Proposition 3.4

If i, j are states of a DTMC that communicates and i is recurrent, then

$$f_{i,j} = 1.$$

Proof. We may assume $i \neq j$. Since $i \leftrightarrow j$ and i is recurrent, j is recurrent by Proposition 3.3. This means $f_{j,j} = 1$. To prove $f_{i,j} = 1$, suppose $f_{i,j} < 1$ for the sake of contradiction. Since $i \leftrightarrow j$, let

$$n_i = \min\left\{n \in \mathbb{N} : P_{j,i}^{(n)} > 0\right\}.$$

That is, each time the DTMC visits to state j, there is a nonzero probability $P_{j,i}^{(n_i)} > 0$ of being in state i n_i time units later. Since we assumed $f_{i,j} < 1$, then this means that the probability of returning to state j after visiting i in the future is not guaranteed, as $1 - f_{i,j} > 0$. Therefore, we have

$$1 - f_{j,j} = \mathbb{P}(\forall n \in \mathbb{N}\, [X_n \neq j] \,|X_0 = j) \geq \underbrace{P_{j,i}^{(n_i)}}_{>0} \underbrace{(1 - f_{i,j})}_{>0} > 0,$$

so $f_{j,j} < 1$, which is our desired contradiction. Thus $f_{i,j} = 1$, as required. ♦

Proposition 3.5

Every finite-state DTMC has at least one recurrent state.

Proof. We may assume that $\mathcal{S} = \{0, \dots, N\}$ for some $N \in \mathbb{N}$ is the state space of the DTMC. Assume that every state is transient for the sake of contradiction. For each $i \in \mathcal{S}$, since we assumed i is transient, $f_{i,i} < 1$. This means that, after a finite amount of time, T_i, state i will not be visited again. Consequently, after a finite amount of time

$$T = \max_{i \in \mathcal{S}} T_i,$$

has gone by, none of the states will be visited ever again. However, the DTMC must be in some state after T units of time, so we obtain a contradiction. Thus there is a recurrent state of the DTMC. ♦

Corollary 3.5.1

Every irreducible finite-state DTMC is recurrent.[a]

[a] We say a DTMC is ***recurrent*** if every state is recurrent.

(EX 3.22)

Consider the DTMC with TPM

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix}.$$

Determine whether each state is transient or recurrent.

Answer. From (EX 3.13), we know that the DTMC is irreducible. Thus by Corollary 3.5.1, every state of the DTMC is recurrent. ◁

Proposition 3.6

Let i, j be states of a DTMC. If i is recurrent and $i \not\leftrightarrow j$, then $P_{i,j}^{(k)} = 0$ for every $k \in \mathbb{N}$.

Proof. For the sake of contradiction, assume $P_{i,j}^{(k)} > 0$ for some $k \in \mathbb{N}$. Choose

$$k_i = \min \left\{ k \in \mathbb{N} : P_{i,j}^{(k)} > 0 \right\}.$$

Then

$$P_{j,i}^{(n)} = 0 \tag{3.8}$$

for every $n \in \mathbb{N}$, since otherwise $i \leftrightarrow j$. However, this means the DTMC has a nonzero probability of at least $P_{i,j}^{(k_i)}$ of never returning to state i, by the minimality of k_i and [3.8]. This is a contradiction, so we conclude that $P_{i,j}^{(k)} = 0$ for all $k \in \mathbb{N}$. ♦

(EX 3.23)

Consider a DTMC with TPM

$$P = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & 1 & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 \end{bmatrix}.$$

Determine whether each state is transient or recurrent.

Answer. From (EX 3.14), we know that $\{0, 1, 3\}, \{2\}$ are the communication classes of the DTMC. Now observe that $P_{2,2}^{(n)} = 1$ for every $n \in \mathbb{N}$, so the series

$$\sum_{n=1}^{\infty} P_{2,2}^{(n)}$$

diverges. So 2 is recurrent. Since $2 \not\leftrightarrow j$ for every $j \in \{0,1,3\}$, it follows that $0,1,3$ are recurrent by Proposition 3.6. ◁

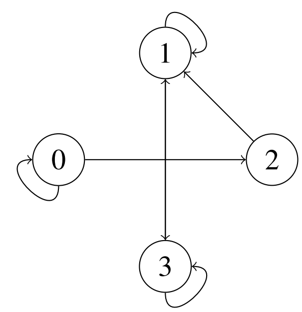
(3.24) As the above example demonstrates, *once a DTMC enters a recurrent class of states, it can never leave that class*. For this reason, a recurrent class is often referred to as a *closed* class.

(EX 3.25) Consider a DTMC with TPM

$$P = \begin{bmatrix} \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 1 & 0 & 0 \\ 0 & \frac{2}{5} & 0 & \frac{3}{5} \end{bmatrix}.$$

Determine whether each state is transient or recurrent.

Answer. The following is the state transition diagram for the DTMC.



This means the communication classes of the DTMC are $\{0\},\{1,3\},\{2\}$. Now observe that

$$\sum_{n=1}^{\infty} P_{0,0}^{(n)} = \sum_{n=1}^{\infty} \left(\frac{1}{4}\right)^n = \frac{1}{3},$$

so 0 is transient. Moreover,

$$\sum_{n=1}^{\infty} P_{2,2}^{(n)} = 0$$

clearly, so 2 is transient. It follows from Proposition 3.3, 3.5 that $1,3$ are recurrent. ◁

3.5 Random Walk

(EX 3.26) Random Walk

Consider a DTMC $\{X_n\}_{n\in\mathbb{N}\cup\{0\}}$ whose state space is $\mathbb{Z}$. Furthermore, suppose that the TPM P for $\{X_n\}_{n\in\mathbb{N}\cup\{0\}}$ satisfies

$$P_{i,i-1} = 1-p$$
$$P_{i,i+1} = p$$

for all $i \in \mathbb{Z}$, where $p \in (0,1)$. As such, X_n can be expressed as

$$X_n = \sum_{k=0}^{n} Y_k,$$

where $\{Y_k\}_{k=0}^{\infty}$ is an independent collection of random variables with $Y_0 = x_0$ and

$$\begin{aligned}\mathbb{P}(Y_k = -1) &= 1-p \\ \mathbb{P}(Y_k = 1) &= p\end{aligned}$$

for all $k \in \mathbb{N}$. Characterize the behavior of this DTMC in terms of its communication classes, periodicity, and recurrence.

Answer. Observe that the state transition diagram of the DTMC is the following.

$$\cdots \quad -2 \longleftrightarrow -1 \longleftrightarrow 0 \longleftrightarrow 1 \longleftrightarrow 2 \quad \cdots$$

Since $p \in (0,1)$, all states communicate with each other, which means $\mathbb{Z}$ is the communication class of the DTMC, and the DTMC is irreducible. Furthermore, starting from state 0, the DTMC cannot visit 0 again in an odd number of transitions. On the other hand, $0 \to 2 \to 0$ is a cycle of length 2. This means the period of 0 is 2. Since periodicity is a class property, it follows that

$$d(i) = 2$$

for all $i \in \mathbb{Z}$. Finally, to determine recurrence of state 0, note that

$$\sum_{m=1}^{\infty} P_{0,0}^{(m)} = \sum_{n=1}^{\infty} P_{0,0}^{(2n)} = \sum_{n=1}^{\infty} \binom{2n}{n} p^n (1-p)^n$$

since $P_{0,0}^{(m)} = 0$ for all odd $m \in \mathbb{N}$. Now note that

$$\begin{aligned}\lim_{n\to\infty} \frac{P^{(2(n+1))}}{P^{(2n)}} &= \lim_{n\to\infty} \frac{\binom{2n+2}{n+1} p^{(n+1)} (1-p)^{n+1}}{\binom{2n}{n} p^n (1-p)^n} = \lim_{n\to\infty} \frac{\frac{(2n+2)!}{(n+1)!(n+1)!} p^{n+1} (1-p)^{n+1}}{\frac{2n!}{n!n!} p^n (1-p)^n} \\ &= \lim_{n\to\infty} \frac{(2n+2)(2n+1)}{(n+1)(n+1)} p(1-p) = \lim_{n\to\infty} 4p(1-p) = 4p(1-p).\end{aligned}$$

This means, when $p \neq \frac{1}{2}$,

$$\lim_{n\to\infty} \frac{P^{(2(n+1))}}{P^{(2n)}} = 4p(1-p) < 1,$$

so the series $\sum_{n=1}^{\infty} P_{0,0}^{(2n)}$ converges by the ratio test. In case $p = \frac{1}{2}$,

$$\lim_{n\to\infty} \frac{P^{(2(n+1))}}{P^{(2n)}} = 1,$$

so the ratio test is inconclusive. To determine what is happening when $p = \frac{1}{2}$, we consider an alternative approach. Recall that

$$f_{i,j} = \mathbb{P}(\exists n \in \mathbb{N}\,[P_n = j]\,|X_0 = i)$$

For convenience, let $q = 1-p$. We condition on the state of the DTMC at time 1:

$$\begin{aligned}f_{0,0} &= \mathbb{P}(\exists n \in \mathbb{N}\,[X_n = 0]\,|X_0 = 0) \\ &= \mathbb{P}(X_1 = -1|X_0 = 0)\,\mathbb{P}(\exists n \geq 2\,[X_n = 0]\,|X_1 = -1, X_0 = 0) \\ &\quad + \mathbb{P}(X_1 = 1|X_0 = 0)\,\mathbb{P}(\exists n \geq 2\,[X_n = 0]\,|X_1 = 1, X_0 = 0) \\ &= \mathbb{P}(X_1 = -1|X_0 = 0)\,\mathbb{P}(\exists n \geq 2\,[X_n = 0]\,|X_1 = -1) \\ &\quad + \mathbb{P}(X_1 = 1|X_0 = 0)\,\mathbb{P}(\exists n \geq 2\,[X_n = 0]\,|X_1 = 1) && \text{by the Markov property} \\ &= qf_{-1,0} + pf_{1,0}.\end{aligned}$$

If we let E represent the event that the DTMC ever makes a future visit to state 0, then

$$E = \bigcup_{i=1}^{\infty} \{X_i = 0\}.$$

So

$$\begin{aligned}
f_{1,0} &= \mathbb{P}(F|X_1 = 0) \\
&= \mathbb{P}(F \cap \{X_1 = 0\} | X_0 = 1) + \mathbb{P}(F \cap \{X_1 = 2\} | X_0 = 1) \\
&= \underbrace{\mathbb{P}(F|X_1 = 0, X_0 = 1)}_{=1} \mathbb{P}(X_1 = 0|X_0 = 1) \\
&\quad + \mathbb{P}(F|X_1 = 2, X_0 = 1)\mathbb{P}(X_1 = 2|X_0 = 1) \\
&= \mathbb{P}(X_1 = 0|X_0 = 1) + \mathbb{P}(F|X_1 = 2)\mathbb{P}(X_1 = 2|X_0 = 1) && \text{by the Markov property} \\
&= q + p\mathbb{P}(E|X_1 = 2) \\
&= q + p\mathbb{P}\left(\bigcup_{i=2}^{\infty} \{X_i = 0\} \bigcup \{X_1 = 0\} | X_1 = 2\right) \\
&= q + p\mathbb{P}\left(\bigcup_{i=2}^{\infty} \{X_i = 0\} | X_1 = 2\right) && \text{since } \mathbb{P}(X_1=0|X_1=2)=0 \\
&= q + p\mathbb{P}(E|X_0 = 2) && \text{by the stationary assumption} \\
&= q + pf_{2,0}.
\end{aligned}$$

Furthermore, it is clear from the defintion of the DTMC that

$$f_{2,0} = f_{2,1}f_{1,0} = f_{1,0}^2,$$

where the last equality holds by the stationary assumption. Hence we obtain that

$$f_{1,0} = (1-p) + pf_{1,0}^2,$$

and by rearranging in terms of $f_{1,0}$ gives

$$pf_{1,0}^2 - f_{1,0} + 1 - p = 0.$$

By applying the quadratic formula,

$$f_{1,0} = \frac{1 \pm \sqrt{1 - 4p(1-p)}}{2p} = 1, \qquad [3.9]$$

since $p = \frac{1}{2}$. By symmetry, $f_{-1,0} = 1$. This means

$$f_{0,0} = (1-p)f_{-1,0} + pf_{1,0} = \frac{1}{2} + \frac{1}{2} = 1,$$

so 0 is recurrent. Thus every state of the DTMC is recurrent, since recurrence is a class property. ◁

Note that, when $p \neq \frac{1}{2}$, the first equality in [3.9] yields

$$f_{1,0} \in \left\{ \frac{1 + |2p-1|}{2p}, \frac{1 - |2p-1|}{2p} \right\}.$$

We may assume $p < \frac{1}{2}$. This means $2p - 1 < 0$, so

$$\frac{1-(1-2p)}{2p} = 1$$
$$\frac{1+(1-2p)}{2p} > 1$$

which means $f_{1,0} = 1$, since a probability cannot be greater than 1. In other words,

$$f_{1,0} = \frac{1-|1-2p|}{2p}.$$

Moreover, it can be shown that

$$f_{-1,0} = \frac{1-|1-2p|}{2(1-p)}.$$

Thus we obtain that

$$f_{0,0} = (1-p)\frac{1-|1-2p|}{2(1-p)} + p\frac{1-(1-2p)}{2p} = 1 - \frac{1}{2}(2-4p) = 1-(1-2p) = 2p < 1$$

when $p < \frac{1}{2}$, which is consistent with our earlier finding that the DTMC is transient when $p \neq \frac{1}{2}$. In general, we have the following formula:

General Formula for $f_{0,0}$ of a Random Walk

$$f_{0,0} = 2\min\{p, 1-p\}. \tag{3.10}$$

3.6 Limiting Behaviors of DTMCs

(3.27) Motivation

The concepts of periodicity and recurrence play an important role in characterizing the limiting behavior of a DTMC. That is, we desire to determine the behavior of the DTMC $(X_n)_{n=1}^{\infty}$ as $n \to \infty$.

Proposition 3.7

For any state i, j of a DTMC, if j is transient, then

$$\lim_{n\to\infty} P_{i,j}^{(n)} = 0.$$

Proof. Recall that

$$f_{i,j}^{(n)} = \mathbb{P}\left(X_n = j, \forall m \in \{n-1,\ldots,1\}\left[X_m \neq j\right] | X_0 = i\right)$$

and that

$$f_{i,j} = \mathbb{P}\left(\exists n \in \mathbb{N}\left[X_n = j\right] | X_0 = i\right).$$

These quantities are related by

$$f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}^{(n)}.$$

Moreover,

$$P_{i,j}^{(n)} = \sum_{k=1}^{n} f_{i,j}^{(k)} P_{j,j}^{(n-k)}.$$

This means

$$\begin{aligned}\sum_{n=1}^{\infty} P_{i,j}^{(n)} &= \sum_{n=1}^{\infty}\sum_{k=1}^{n} f_{i,j}^{(k)} P_{j,j}^{(n-k)} = \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} f_{i,j}^{(k)} P_{j,j}^{(n-k)} \\ &= \sum_{k=1}^{\infty} f_{i,j}^{(k)} \sum_{n=k}^{\infty} P_{j,j}^{(n-k)} = \sum_{k=1}^{\infty} f_{i,j}^{(k)} \sum_{l=0}^{\infty} P_{j,j}^{(l)} \\ &= f_{i,j}\left(1+\sum_{l=1}^{\infty} P_{j,j}^{(l)}\right).\end{aligned}$$

But note that $\sum_{l=1}^{\infty} P_{j,j}^{(l)}$ is convergent, as j is transient. It follows that $\sum_{n=1}^{\infty} P_{i,j}^{(n)}$ is also convergent, which implies

$$\lim_{n\to\infty} P_{i,j}^{(n)} = 0.$$ ♦

(3.28) Mean Recurrent Time

It is worthwhile to determine a set of conditions which ensure the *nice* limiting behavior. To ascertain when such conditions exist, we need to distinguish between two kinds of recurrences. Suppose that we are given a DTMC $(X_n)_{n=1}^{\infty}$ and let

$$N_i = \min\{n \in \mathbb{N} : X_n = i\}$$

for all recurrent state i. Clearly the conditional random variable $N_i | (X_0 = i)$ takes on values in $\mathbb{N}$. Moreover, the conditional pmf is given by

$$p_{N_i|X_0}(n|i) = \mathbb{P}(N_i = n | X_0 = i) = f_{i,i}^{(n)}$$

for all $n \in \mathbb{N}$. We observe that this indeed is a pmf since

$$\sum_{n=1}^{\infty} f_{i,i}^{(n)} = f_{i,i} = 1,$$

as i is recurrent. This leads to the introduction of the following important quantity.

Mean Recurrent Time of a Recurrent State

Def'n 3.13

Consider the setting of (3.28). We define the ***mean recurrent time*** of i, denoted as m_i, by

$$m_i = \mathbb{E}(N_i | X_0 = i).$$

It is immediate from the construction that the following equation holds:

Formula for m_i

$$m_i = \sum_{n=1}^{\infty} n f_{i,i}^{(n)}. \quad [3.11]$$

(a) Note that we may have $m_i = \infty$, in case the right-hand-side of [3.11] diverges to infinity.

(b) In words, m_i represents the *average time* it takes the DTMC to make successive visits to state i.

Two notions of recurrence can now be defined based on the value of m_i.

Positive, Null Recurrence

Def'n 3.14

Let i be a recurrent state of a DTMC. We say i is

(a) ***positive*** recurrent if $m_i < \infty$; and

(b) ***null*** recurrent if $m_i = \infty$.

We shall admit the following facts about positive and null recurrences without any proof, as their proof is lengthy and beyond the scope of this note.

Fact 3.8

Positive recurrence is a class property. That is, given states i, j that communicate, i is positive recurrent if and only if j is.

Fact 3.9

Every recurrent state in a finite-state DTMC is positive recurrent.

Ergodic State

Def'n 3.15

A state of a DTMC is called ***ergodic*** if positive recurrent and aperiodic.

Stationary Distribution of a DTMC

Def'n 3.16

Let $p : \mathbb{N} \cup \{0\} \to [0,1]$ be a pmf. We say p is a ***stationary distribution*** (or ***invariant distribution***, ***steady-state distribution***) of a DTMC if

$$p(j) = \sum\nolimits_{i=0}^{\infty} p(i) P_{i,j}$$

for all $j \in \mathbb{N} \cup \{0\}$.[a]

[a] We usually denote p as a sequence: $p = (p_i)_{i=0}^{\infty}$.

(3.29) Stationary Distribution

Suppose that p is a stationary distribution of a DTMC.

(a) If we use the notation

$$p = (p_i)_{i=0}^{\infty},$$

then Def'n 3.16 can be represented in matrix form as

$$\begin{aligned} p^T e &= 1 \\ p^T &= p^T P, \end{aligned}$$

where $e = (1)_{i=0}^{\infty}$ denotes the sequence whose terms are all 1.[1]

(b) Suppose that the initial conditions of the DTMC $(X_n)_{n=0}^{\infty}$ are given by

$$\alpha_0 = p.$$

As a result, we have that

$$\alpha_{0,j} = \mathbb{P}(X_0 = j) = p_j$$

for all $j \in \mathbb{N} \cup \{0\}$. Now, for any $j \in \mathbb{N} \cup \{0\}$, note that

$$\alpha_{1,j} = \mathbb{P}(X_1 = j) = \sum\nolimits_{i=0}^{\infty} \alpha_{0,i} P_{i,j} = \sum\nolimits_{i=0}^{\infty} p_i P_{i,j} = p_j = \alpha_{0,j}.$$

This indicates $X_1 \sim X_0$, when the initial conditions are set by a stationary distribution. It follows inductively that each X_i, $i \in \mathbb{N}$, is identically distributed to X_0. In words, if a DTMC is started according to a stationary distribution, then the probability of being in a given state remains unchanged (i.e. stationary) over time. This explains the nomenclature *stationary*.

(c) Stationary distribution is not necessarily unique. This happens when a DTMC has more than one positive recurrent communication class. In particular, some examples have an *infinite* number of stationary distributions.

We are also going to accept the following known facts without any formal justification.

[1] We shall use column notation for vectors. That is, whenever we are using sequences as vectors, we are using them as *column* vectors.

Fact 3.10

If a DTMC does not have a positive recurrent state, then there is no stationary distribution.

Fact 3.11

Given any irreducible DTMC, the following are equivalent.

(a) The DTMC is positive recurrent.

(b) There exists a stationary distribution of the DTMC.

(3.30)
Basic Limit Theorem

We are now in position to state the fundamental limiting theorem for DTMCs, generally referred to as the *basic limit theorem*.

Theorem 3.12
Basic Limit Theorem (BLT)

Let $(X_n)_{n=0}^{\infty}$ be a DTMC. If $(X_n)_{n=0}^{\infty}$ is irreducible, recurrent, and aperiodic, then

$$\lim_{n\to\infty} P_{i,j}^{(n)} = \frac{1}{m_j}$$

for all $i, j \in \mathbb{N} \cup \{0\}$. Furthermore, if the recurrence is positive, then $(\pi_j)_{j=0}^{\infty} = \left(\frac{1}{m_j}\right)_{j=0}^{\infty}$ is the unique positive solution to the system of linear equations defined by[a]

$$\begin{aligned} \pi_j &= \sum\nolimits_{i=0}^{\infty} \pi_i P_{i,j} && \forall j \in \mathbb{N} \cup \{0\} \\ \sum\nolimits_{j=0}^{\infty} \pi_j &= 1, \end{aligned}$$

[a]For the definition of m_j, the mean recurrent time, see Def'n 3.13.

A formal proof of the BLT is beyond the scope of this note. But here are some remarks.

(a) If we denote $\pi = (\pi_j)_{j=0}^{\infty}$, then the system of linear inequalities in the BLT can be succintly written as

$$\begin{aligned} \pi^T &= \pi^T P \\ \pi^T e &= 1. \end{aligned}$$

In particular, if a DTMC is irreducible and ergodic, then the BLT states that the limiting probability distribution is the unique stationary distribution.

(b) When a DTMC has a finite number of states, say $N+1 \in \mathbb{N}$ states, the system

$$\begin{aligned} \pi^T &= \pi^T P \\ \pi^T e &= 1 \end{aligned}$$

has $N+2$ equations but $N+1$ indeterminates, of which a unique solution must exist. In fact, the first $N+1$ equations (i.e. $\pi^T = \pi^T P$) are linearly dependent, so we can *drop any one* of the equations and solve the remaining system to obtain the unique solution.

(c) If the conditions of the BLT are satisfied and state j is null recurrent, then $\pi_j = 0$, interestingly similar to the limiting behavior of a transient state.

(EX 3.31) Consider a DTMC with TPM

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

Find the limiting probabilities for this DTMC.

Answer. Note that the DTMC is irreducible, aperiodic, and positive recurrent. Therefore, the limiting probability distribution

$$\pi = (\pi_0, \pi_1, \pi_2)$$

exists by the BLT. To find π, we solve the following system of linear equalities

$$\pi^T = \pi^T P \tag{3.12}$$

subject to $\pi_1 + \pi_2 + \pi_3 = 1$. Note that [3.12] is equivalent to

$$\pi = P^T \pi.$$

In other words, π is the eigenvector of P^T whose sum of entries is 1. Then by employing the techniques from MATH 146, we find that

$$\pi = \begin{bmatrix} \frac{4}{11} \\ \frac{4}{11} \\ \frac{3}{11} \end{bmatrix}.$$

◁

Def'n 3.17 **Doubly Stochastic Matrix**

Let $P \in M_{n\times n}(\mathbb{R})$ be a matrix.[a] If both P, P^T are stochastic, then we say P is ***doubly stochastic***.

[a] $n \in \mathbb{N} \cup \{\infty\}$.

Proposition 3.13 *Suppose that a fintite-state DTMC with state $S = \{0,\ldots,N-1\}$ is irreducible and aperiodic. If the associated TPM P is doubly stochastic, then the limiting probabilities $\pi_0,\ldots,\pi_{N-1}$ exist and are given by*

$$\pi_j = \frac{1}{N}$$

for all $j \in \{0,\ldots,N-1\}$.

Proof. We are given that the DTMC is irreducible and aperiodic. Moreover, every finite irreducible DTMC is positive recurrent, so a unique limiting probability distribution $\pi = (\pi_j)_{j=0}^{N-1}$ exists by the BLT. To determine the limiting distribution, let us propose that

$$\pi = \left(\frac{1}{N}\right)_{j=0}^{N-1}$$

is a solution to the system of linear equalities

$$\pi^T P = \pi^T$$
$$\pi^T e = 1.$$

Clearly the second equation is satisfied:

$$\pi^T e = \begin{bmatrix} \pi_0 & \cdots & \pi_{N-1} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \sum_{j=0}^{N-1} \pi_j = \sum_{j=0}^{N-1} \frac{1}{N} = 1.$$

Moreover, for all $j \in \{0, \dots, N-1\}$, note that

$$\sum_{k=0}^{N-1} \pi_k P_{k,j} = \sum_{k=0}^{N-1} \frac{1}{N} P_{k,j} = \frac{1}{N} \underbrace{\sum_{k=0}^{N-1} P_{k,j}}_{=1} = \frac{1}{N} = \pi_j,$$

where the equality $\sum_{k=0}^{N-1} P_{k,j} = 1$ follows from the fact that P is doubly stochastic. This concludes the proof. ♦

(3.32) Alternative Interpretation of the Limiting Distribution of a DTMC

The primary interpretation of the limiting distribution of a DTMC $(X_n)_{n=0}^{\infty}$ is that after the process has been in operation for a *long* period of time, the probability of finding the process in state j is π_j, assuming the conditions of the BLT are met. In such situations, however, another interpretation exists for π_j: *the long-run mean fraction of time that the process spends in state j*. To see that this interpretation is valid, define the sequence of indicator random variables $(A_k)_{k=1}^{\infty}$ as follows:

$$A_k = \begin{cases} 0 & \text{if } X_k \neq j \\ 1 & \text{if } X_k = j \end{cases}$$

for all $k \in \mathbb{N}$. The fraction of time the DTMC visits state j during the time interval from 1 to n, inclusive, is therefore given by

$$\frac{1}{n} \sum_{k=1}^{n} A_k.$$

Now suppose a state i is given and consider the quantity

$$\mathbb{E}\left(\frac{1}{n} \sum_{k=1}^{n} A_k | X_0 = i\right),$$

which is interpreted as the mean fraction of time spent in state j during the time interval from 1 to n, inclusive, given that the process starts in state i. Note that

$$\begin{aligned} \mathbb{E}\left(\frac{1}{n} \sum_{k=1}^{n} A_k | X_0 = i\right) &= \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left(A_k | X_0 = i\right) \\ &= \frac{1}{n} \sum_{k=1}^{n} 0\,\mathbb{P}\left(A_k = 0 | X_0 = i\right) + 1\,\mathbb{P}\left(A_k = 1 | X_0 = i\right) \\ &= \frac{1}{n} \sum_{k=1}^{n} \mathbb{P}\left(A_k = 1 | X_0 = i\right) \\ &= \frac{1}{n} \sum_{k=1}^{n} P_{i,j}^{(k)}. \end{aligned}$$

But recall that if $(a_n)_{n=1}^{\infty} \in \mathbb{R}^{\mathbb{N}}$ is such that $\lim_{n\to\infty} a_n = a \in \mathbb{R}$, then $\frac{1}{n} \sum_{n=1}^{\infty} a_n = a$. Now we know that

$$\lim_{n\to\infty} P_{i,j}^{(n)} = \pi_j$$

if the conditions of the BLT are satisfied. Thus we obtain

$$\lim_{n\to\infty} \mathbb{E}\left(\frac{1}{n} \sum_{k=1}^{n} A_k | X_0 = i\right) = \pi_j,$$

justifying our interpretation.

(3.33) If $(X_n)_{n=0}^{\infty}$ begins in recurrent state j, then the DTMC spends one unit of time in j every N_j time units. On average, this amounts to one unit of time in state j every $\mathbb{E}(N_j|X_0 = j) = m_j$ time units. If the conditions of the BLT are satisfeid, then it makes sense intuitively that

$$\pi_j = \frac{1}{m_j}$$

as the BLT specifies. We can produce more formal justification in the positive recurrent case. Let $\left(N_j^{(n)}\right)_{n=1}^{\infty}$ be a sequence of random variables where $N_j^{(n)}$ represents the number of transitions between the $(n-1)$th and nth visits into state j. By the Markov property and the stationary assumption of the DTMC, $\left(N_j^{(n)}\right)_{n=1}^{\infty}$ is an iid sequence of random variables with common mean m_j. Therefore, the long-run fraction of time spent in state j can be viewed as

$$\pi_j = \lim_{n\to\infty} \frac{n}{\sum_{i=1}^n N_j^{(i)}} = \lim_{n\to\infty} \frac{1}{\frac{1}{n}\sum_{i=1}^n N_j^{(i)}} = \frac{1}{m_j},$$

where the last equality follows from the SLLN.

3.7 Galton-Watson Branching Process

(3.34) We assume that a population of individuals (e.g. people, organisms, free particles, ...) evolves in discrete time. Specifically, we define

$$X_n = \text{population of the } n\text{th generation}$$

for all $n \in \mathbb{N} \cup \{0\}$, where X_0 is the population of the 0th (i.e. original) generation.

We assume that each individual in a generation produces a random number (possibly 0) of individuals, called *offspring*, which go on and become part of the very next generation. In other words, it is always the offspring of a current generation which go on to form the next generation.

We further assume that individuals produce offspring independently of all others according to the same probability distribution.

Notation 3.18 α_m

For all $m \in \mathbb{N} \cup \{0\}$, let

$$\alpha_m = \mathbb{P}(\text{an individual produces } m \text{ offspring})$$

In addition, we make two additional assumptions: $\alpha_0 \in (0,1)$ and $\alpha_0 + \alpha_1 < 1$.

Notation 3.19 $Z_i^{(j)}$

For each $j \in \mathbb{N} \cup \{0\}$, we write $Z_i^{(j)}$ to denote the number of offspring produced from individual i in the jth generation.

Due to the independence assumptions, $\left(Z_i^{(j)}\right)_{i=1}^{\infty}$ is an iid sequence of random variables with

$$\alpha_m = \mathbb{P}\left(Z_i^{(j)} = m\right)$$

for all $j \in \mathbb{N} \cup \{0\}$. Hence we may define the following.

Notation 3.20 μ, σ

Let $\mu = \mathbb{E}\left(Z_i^{(j)}\right), \sigma^2 = \text{var}\left(Z_i^{(j)}\right)$ represent to the common mean and variance, respectively, of the number of offspring produced by a single individual.

Based on the above assumptions, any Galton-Watson process $(X_n)_{n \in \mathbb{N} \cup \{0\}}$ is a stationary DTMC taking values in the state space $\mathbb{N} \cup \{0\}$, since it follows that

$$X_n = \sum_{i=1}^{X_{n-1}} Z_i^{(n-1)}, \tag{3.13}$$

implying that the Markov property and stationary assumption are both satisfied.

(3.35) Mean and Variance of a Galton-Watson Process

Since [3.13] infers that X_n is expressible as a *random sum*, we can apply the results of (EX 2.10) to obtain

$$\mathbb{E}(X_n) = \mu \, \mathbb{E}(X_n)$$

and

$$\text{var}(X_n) = \sigma^2 \, \mathbb{E}(X_{n-1}) + \mu^2 \, \text{var}(X_{n-1})$$

for all $n \in \mathbb{N}$. Note that the above equations are not explicit, but recursive. Let us henceforth assume that $X_0 = 1$ almost surely. As it is understood that $X_0 = 1$, we will suppress writing the condition $X_0 = 1$ in all expectations and probabilites will follow. This means

$$\mathbb{E}(X_n) = \mu^n$$

and

$$\text{var}(X_n) = \sigma^2 \mu^{n-1} \sum_{i=0}^{n-1} \mu^i = \begin{cases} n\sigma^2 & \text{if } \mu = 1 \\ \sigma^2 \mu^{n-1} \left(\frac{1-\mu^n}{1-\mu}\right) & \text{if } \mu \neq 1 \end{cases}$$

for all $n \in \mathbb{N} \cup \{0\}$.

(3.36) Limiting Probability

In a Galton-Watson branching process, we remakr that $P_{0,0} = 1$, since state 0 is obviously an absorbing state (i.e. no individuals from which offsprings can be born). If we consider state $i \in \mathbb{N}$, then we can easily show that i is transient as follows.

Proof. Note that $0 \not\to i$. But $P_{i,0} = \alpha_0^i > 0$, since $\alpha_0 > 0$. Thus by (the constrapositive) of Proposition 3.6 that i is transient. ◁

Thus, since state 0 is recurrent and states $1, 2, \ldots$ are transient, exactly one of the following happens.

(a) The population will die out eventually.

(b) The population will grow indefinitely.

Notation 3.21 π_0

Let π_0 denote the limiting probability dies out:

$$\pi_0 = \lim_{n \to \infty} \mathbb{P}(X_n = 0).$$

The limiting behaviors can be characterized in terms of μ.

- *Case 1. Suppose* $\mu < 1$. This is referred to as the *subcritical case*. Clearly, as n grows to infinity, both

$$\mathbb{E}(X_n) = \mu^n, \text{var}(X_n) = \sigma^2 \mu^{n-1} \left(\frac{1-\mu^n}{1-\mu}\right)$$

converge to 0. Therefore, we expect that $\pi_0 = 1$. To show this formally, note that

$$\mu^n = \mathbb{E}(X_n) = \sum_{j=1}^{\infty} j\,\mathbb{P}(X_n = j) \geq \sum_{j=1}^{\infty} \mathbb{P}(X_n = j) = \mathbb{P}(X_n \geq 1) = 1 - \mathbb{P}(X_n = 0).$$

This implies that $1 - \mu^n \leq \mathbb{P}(X_n = 0) \leq 1$. Taking the limit leads to

$$1 = \lim_{n\to\infty} (1 - \mu^n) \leq \lim_{n\to\infty} \mathbb{P}(X_n = 0) = \pi_0 \leq 1,$$

so $\pi_0 = 1$.

- *Case 2. Suppose $\mu \geq 1$.* By conditioning on the number of offspring produced by the single individual present in the population at time 0, we obtain

$$\pi_0 = \mathbb{P}(\text{population dies out}) = \sum_{j=0}^{\infty} \mathbb{P}(\text{population dies out} | X_1 = j)\,\alpha_j.$$

However, with $X_1 = j$, the population will eventually die out if and only if each of the j families started by the members of the first generation eventually dies out. As each family is assumed to act independently, and since the probability that any particular family dies out is simply π_0, it follows that

$$\mathbb{P}(\text{population dies out} | X_1 = j) = \pi_0^j,$$

which means

$$\pi_0 = \sum_{j=0}^{\infty} \pi_0^j \alpha_j. \qquad [3.14]$$

Clearly $\pi_0 = 0$ is not a solution and $\pi_0 = 1$ is a solution to [3.14].

Now, it can be shown that, by defining $\alpha : [0,1] \to \mathbb{R}$ by

$$\alpha(z) = \sum_{j=0}^{\infty} z^j \alpha_j,$$

then α admits another root in $(0,1)$ besides 1, say z_0, when $\mu > 1$ (when $\mu = 1$, 1 is the unique root of α). Therefore, we have to determine if $\pi_0 = z_0$ or $\pi_0 = 1$ when $\mu > 1$.

To determine this, let $z^* \in \{z_0, 1\}$. Then by induction, we can show that

$$z^* \geq \mathbb{P}(X_n = 0)$$

for all $n \in \mathbb{N} \cup \{0\}$. As a result, it follows that

$$z^* = \lim_{n\to\infty} z^* \geq \lim_{n\to\infty} \mathbb{P}(X_n = 0) = \pi_0.$$

Thus π_0 is equal to the smallest element of $\{z_0, 1\}$, namely z_0.

It is only when $\mu > 1$ (referred to as *supercritical case*), $\pi_0 \in (0,1)$.

(EX 3.37) Given the following offspring probabilities, what is the probability that the population dies out in the long run assuming that $X_0 = 1$?

(a) $\alpha_0 = \frac{3}{4}, \alpha_1 = \frac{1}{8}, \alpha_2 = \frac{1}{8}$.

Answer. First, we calculate

$$\mu = 0 \cdot \frac{3}{4} + 1 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} = \frac{3}{8}.$$

Since $\mu < 1$, we are in a critical case, so $\pi_0 = 1$. Thus the population will die out almost surely. ◁

(b) Answer. Note that

$$\mu = 0\cdot\frac{1}{5} + 1\cdot\frac{1}{10} + 2\cdot\frac{7}{10} = \frac{3}{2}.$$

Since $\mu > 1$, we are in a supercritical ccase, so $\pi_0 \in (0,1)$. To find π_0, we solve [3.14]:

$$z = \sum_{j=0}^{\infty} z^j \alpha_j = \frac{1}{5} + \frac{1}{10}z + \frac{7}{10}z^2. \tag{3.15}$$

Rearranging [3.15] gives

$$7z^2 - 9z + 2 = 0,$$

whose smallest root is

$$\frac{9-\sqrt{81-56}}{14} = \frac{4}{14} = \frac{2}{7}.$$

Thus $\pi_0 = \frac{2}{7}$. ◁

(3.38) We have some remarks.

(a) In the case when $X_0 = n$ almost surely for some $n \in \mathbb{N}$, the population will die out if and only if the families of each of the n members of the initial generation die out. As a result, it immediately follows out that the *extinction probability* is simply π_0^n.

(b) For certain choices of the offspring distriubtion, the Galton-Watson branching process is not very interesting to analyze. For example, with $X_0 = 1, a_r = 1$ for some $r \in \mathbb{N}\cup\{0\}$, the evolution of the process is almost surely deterministic:

$$\mathbb{P}(X_n = r^n) = 1$$

for all $n \in \mathbb{N}\cup\{0\}$. Another uninteresting case occurs when $\alpha_0, \alpha_1 > 0$ and $\alpha_0 + \alpha_1 = 1$. In this situation, the polulation remains at its initial size $X_0 = 1$ for a random number of generations (according to a geometric distribution), before dying out completely.

3.8 Gambler's Ruin

(3.39) One of the most powerful ideas in the theory of DTMCs is that many fundamental probabilities and expectations can be computed as the solutions of systems of linear equations. We have already seen one such example of this through the application of the BLT. In what follows, we will continue to illustrate this idea by deriving appropirate linear systems for a number of key probabilities and expectations that arise in certain settings.

(EX 3.40) Gambler's Ruin Problem

Consider a gambler who, at each play of a game, has probability $p \in (0,1)$ of winning one unit and probability $q = 1-p$ of losing one unit. Assume that successive plays of the game are independent. If the gambler initially begins with i units, what is the probability that the gambler's fortune will reach $N \in \mathbb{N}$ units before reaching 0 units?

Answer. For each $n \in \mathbb{N}\cup\{0\}$, define X_n as the gambler's fortune after the nth play of the game, with $X_0 = i$ for some $i \in \mathbb{N}$. Clearly, $(X_n)_{n\in\mathbb{N}\cup\{0\}}$ is a stationary DTMC with TPM

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

Note that states $0, N$ are absorbing, so recurrent. States $1, \ldots, N-1$ are in the same communication class, and it is straightforward to show that they are transient. The goal is to determine

$$G_i = \mathbb{P}\left(\text{the gambler's fortune will eventually reach } N | X_0 = i\right)$$

for all $i \in \{0, \ldots, N\}$. This means

$$\lim_{n\to\infty} P^{(n)} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1-G_1 & 0 & \cdots & 0 & G_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-G_{N-1} & 0 & \cdots & 0 & G_{N-1} \\ 0 & 0 & \vdots & 0 & 1 \end{bmatrix}.$$

From above, note that $G_0 = 0, G_N = 1$. Moreover, by conditioning on the outcome of the very first game, we readily obtain for all $i \in \{1, \ldots, N-1\}$ that

$$G_i = pG_{i+1} + qG_{i-1}.$$

But by definition, $p + q = 1$, so

$$pG_i + qG_i = pG_{i+1} + qG_{i-1},$$

rearranging which gives

$$G_{i+1} - G_i = \frac{q}{p}\left(G_i - G_{i-1}\right).$$

But note that

$$\begin{aligned} G_2 - G_1 &= \frac{q}{p}\left(G_1 - G_0\right) = \frac{q}{p} G_1 \\ G_3 - G_2 &= \frac{q}{p}\left(G_2 - G_1\right) = \left(\frac{q}{p}\right)^2 G_1 \\ G_4 - G_3 &= \frac{q}{p}\left(G_3 - G_2\right) = \left(\frac{q}{p}\right)^3 G_1 \\ &\vdots \\ G_{i+1} - G_i &= \frac{q}{p}\left(G_i - G_{i-1}\right) = \left(\frac{q}{p}\right)^i G_1 \end{aligned}$$

by induction. Note that the above i equations are linear, in terms of $G_1, \ldots, G_{i+1}$. Moreover, summing these i equations results in

$$G_{i+1} - G_1 = G_1 \sum\nolimits_{j=1}^{i} \left(\frac{q}{p}\right)^j$$

so

$$G_{i+1} = G_1 \sum\nolimits_{j=0}^{i} \left(\frac{q}{p}\right)^j.$$

Equivalently,

$$G_i = \begin{cases} G_1 \frac{1-\left(\frac{q}{p}\right)^{i-1}}{1-\frac{q}{p}} & \text{if } p \neq \frac{1}{2} \\ iG_1 & \text{if } p = \frac{1}{2} \end{cases}$$

by geometric series. In particular, when $i = N$, we obtain

$$1 = G_N = \frac{1-\left(\frac{q}{p}\right)^N}{1-\left(\frac{q}{p}\right)} G_1$$

for $p \neq \frac{1}{2}$, which means

$$G_1 = \frac{1 - \left(\frac{q}{p}\right)}{1 - \left(\frac{q}{p}\right)^N}.$$

Similarly, for $p = \frac{1}{2}$, we find

$$G_1 = \frac{1}{N}.$$

Combining both cases, we obtain

$$G_i = \begin{cases} \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} & \text{if } p \neq \frac{1}{2} \\ \frac{i}{N} & \text{if } p = \frac{1}{2} \end{cases}$$

for all $i \in \{0, \ldots, N\}$ (the formula we derived for $1, \ldots, N-1$ works for $0, N$ as well). $\triangleleft$

(3.41) We have some remarks.

(a) An interesting question to ask is *what happens to the gambler's probability of winning the jackpot, given an initial fortune of i units, as N grows larger (i.e. $N \to \infty$)?* In other words, we are interested in what happens to the limit of G_i as $N \to \infty$. Lookingg at three cases based on the value of p, we see

- *when* $p = \frac{1}{2}$, $G_i = \frac{1}{N} \to 0$ as $N \to \infty$;
- *when* $p < \frac{1}{2}$, $G_i = \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} \to 0$ as $N \to \infty$; and
- *when* $p > \frac{1}{2}$, $G_i = \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} \to 1 - \left(\frac{q}{p}\right)^i$ as $N \to \infty$.

Simply put, only when $p > \frac{1}{2}$ does a positive probability exist that the gambler's fortune will increase indefinitely. Otherwise, the gambler is sure to go broke.

(b) In our study of the *random walk* in (EX 3.26) featuringg a DTMC on the state space $\mathbb{Z}$ with transition probabilities analogous to those in the gambler's ruin problem, we previously showed that

$$f_{0,0} = (1-p)\, f_{-1,0} + p f_{1,0}.$$

Suppose that $p > \frac{1}{2}$. First of all, note that

$$\begin{aligned} f_{1,0} &= \mathbb{P}\left(\text{random walk ever makes a future visit to state 0 starting from state 1}\right) \\ &= \lim_{N \to \infty} \mathbb{P}\left(\text{gambler's ruin ends up in bankruptcy} | X_0 = 1\right) \\ &= 1 - \left(1 - \left(\frac{1-p}{p}\right)^1\right) = \frac{1-p}{p}. \end{aligned}$$

4.
Poisson Processes

4.1 Exponential Distribution

(4.1) If a random variable X has an exponential distribution with parameter $\lambda > 0$ (i.e. $X \sim \text{EXP}(\lambda)$, where λ is often referred to as the *rate*), then we have the following basic distributional results in place:

$$\begin{aligned} \text{pdf:} \quad & f_X(x) = \lambda e^{-\lambda x} && \forall x > 0 \\ \text{cdf:} \quad & F_X(x) = 1 - e^{-\lambda x} && \forall x \geq 0 \\ \text{mgf:} \quad & \varphi_X(t) = \frac{\lambda}{\lambda - t} && \forall t < \lambda \\ \text{mean:} \quad & \mathbb{E}(X) = \frac{1}{\lambda} \\ \text{variance:} \quad & \text{var}(X) = \frac{1}{\lambda^2} \end{aligned}.$$

(4.2) Minimum of Independent Exponentials

Let $(X_i)_{i=1}^n$ be a finite sequence of independent random variables with

$$X_i \sim \text{EXP}(\lambda_i)$$

for all $i \in \{1, \ldots, n\}$. Let

$$Y = \min(X_1, \ldots, X_n),$$

the *smallest order statistic* of $(X_i)_{i=1}^n$. Clearly, Y takes on possible values in the state space $\mathcal{S} = (0, \infty)$. To determine the distribution of Y, consider the tpf of Y: for all $y \geq 0$,

$$\begin{aligned} \overline{F}_Y(y) &= \mathbb{P}(Y > y) \\ &= \mathbb{P}(\min(X_1, \ldots, X_n) > y) \\ &= \mathbb{P}(X_1 > y, \ldots, X_n > y) \\ &= \prod_{i=1}^n \mathbb{P}(X_i > y) && \text{by independence} \\ &= \prod_{i=1}^n e^{-\lambda_i y} \\ &= e^{-\sum_{i=1}^n \lambda_n y}. \end{aligned}$$

That is, the tpf of Y is the tpf of an $\text{EXP}(\sum_{i=1}^n \lambda_i)$ random variable. Thus,

$$Y \sim \text{EXP}\left(\sum_{i=1}^n \lambda_i\right).$$

When $X_1, \ldots, X_n$ are iid with the common rate $\lambda > 0$, then this simplifies to

$$Y \sim \text{EXP}(n\lambda).$$

(EX 4.3) Let $(X_i)_{i=1}^n$ be a sequence of independent random variables, where

$$X_i \sim \text{EXP}(\lambda_i)$$

for all $i \in \{1, \ldots, n\}$.

(a) For each $j \in \{1, \ldots, n\}$, determine $\mathbb{P}(X_j = \min(X_1, \ldots, X_n))$.

Answer. Note that

$$
\begin{aligned}
\mathbb{P}(X_j = \min(X_1, \ldots, X_n)) &= \mathbb{P}(\forall i \in \{1, \ldots, n\} \setminus \{j\} [X_j < X_i]) \\
&= \int_0^\infty \mathbb{P}(\forall i \in \{1, \ldots, n\} \setminus \{j\} [X_i > x] | X_j = x) \lambda_j e^{-\lambda_j x} \, \mathrm{dx} \\
&= \int_0^\infty \mathbb{P}(\forall i \in \{1, \ldots, n\} \setminus \{j\} [X_i > x]) \lambda_j e^{-\lambda_j x} \, \mathrm{dx} \\
&= \int_0^\infty \prod_{i=1, i \neq j}^n \mathbb{P}(X_i > x) \lambda_j e^{-\lambda_j x} \, \mathrm{dx} \\
&= \int_0^\infty \prod_{i=1, i \neq j}^n e^{-\lambda_i x} \lambda_j e^{-\lambda_j x} \, \mathrm{dx} \\
&= \frac{\lambda_j}{\sum_{i=1}^n \lambda_i} \int_0^\infty \sum_{i=1}^n \lambda_i e^{-\sum_{i=1}^n \lambda_i x} \, \mathrm{dx} \\
&= \frac{\lambda_j}{\sum_{i=1}^n \lambda_i} \int_0^\infty p_Y(x) \, \mathrm{dx} \\
&= \frac{\lambda_j}{\sum_{i=1}^n \lambda_i}
\end{aligned}
$$

where the third equality is obtained by conditioning on X_j, fourth equality is obtained by the fact that X_j is independent of $(X_i)_{i=1, i \neq j}^n$, and the second last equality is by defining $Y \sim \mathrm{EXP}(\sum_{i=1}^n \lambda_i)$. Thus we conclude

$$
\mathbb{P}(X_j = \min(X_1, \ldots, X_n)) = \frac{\lambda_j}{\sum_{i=1}^n \lambda_i}. \qquad \triangleleft
$$

(b) Show that the conditional random variable $X_1 | (X_1 < \cdots < X_n)$ is identically distributed to the random variable $\min(X_1, \ldots, X_n)$.

Proof. Let

$$
Y = X_1 | (X_1 < \cdots < X_n)
$$

for convenience. Then for all $y \geq 0$,

$$
\begin{aligned}
\overline{F}_Y(y) &= \mathbb{P}(X_1 > y | X_1 < \cdots < X_n) \\
&= \frac{\mathbb{P}(y < X_1 < \cdots < X_n)}{\mathbb{P}(X_1 < \cdots < X_n)}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
\mathbb{P}(y < X_1 < \cdots < X_n) &= \int_y^\infty \int_{X_1}^\infty \cdots \int_{X_{n-1}}^\infty \prod_{i=1}^n \lambda_i e^{-\lambda_i x_i} \, \mathrm{dx}_n \cdots \mathrm{dx}_1 \\
&= \prod_{i=1}^{n-1} \lambda_i \int_y^\infty e^{-\lambda_1 x_1} \int_{X_1}^\infty e^{-\lambda_2 x_2} \cdots \int_{X_{n-1}}^\infty \lambda_n e^{-\lambda_n x_n} \, \mathrm{dx}_n \cdots \mathrm{dx}_1 \\
&= \cdots \\
&= \frac{\prod_{i=1}^{n-1} \lambda_i}{\prod_{i=1}^{n-1} \sum_{j=i}^n \lambda_j} e^{-\sum_{i=1}^n \lambda_i y}.
\end{aligned}
$$

But by letting $y = 0$, we obtain $\mathbb{P}(X_1 < \cdots < X_n)$; that is,

$$
\mathbb{P}(X_1 < \cdots < X_n) = \frac{\prod_{i=1}^{n-1} \lambda_i}{\prod_{i=1}^{n-1} \sum_{j=i}^n \lambda_j}.
$$

Thus

$$\overline{F}_Y(y) = e^{-\sum_{i=1}^{n} \lambda_i y}$$

for all $y \geq 0$, which is the tpf of an $\text{EXP}\left(\sum_{i=1}^{n} X_i\right)$ random variable. Since

$$\min(X_1, \ldots, X_n) \sim \text{EXP}\left(\sum_{i=1}^{n} X_i\right)$$

from (4.2), it follows that

$$X_1 | (X_1 < \cdots < X_n) = Y \sim \min(X_1, \ldots, X_n). \qquad \triangleleft$$

In case when $n = 2$, note that the result from part (a) simplifies to become

$$\mathbb{P}(X_1 = \min(X_1, X_2)) = \mathbb{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

Interestingly, looking at the derivation in part (b), we see that

$$\mathbb{P}(X_1 < \cdots < X_n) = \prod_{i=1}^{n-1} \frac{\lambda_1}{\sum_{j=i}^{n} \lambda_j} = \prod_{i=1}^{n-1} \mathbb{P}(X_i = \min(X_i, \ldots, X_n)).$$

Def'n 4.1 **Memoryless** Random Variable

Let X be a random variable. If

$$\mathbb{P}(X > y + z | X > y) = \mathbb{P}(X > z)$$

for all $y, z \geq 0$, then we say X is ***memoryless***.

(4.4) Memoryless Property

Note that if we express $\mathbb{P}(X > y + z | X > y)$ as $\mathbb{P}(X - y > z | X > y)$ and think of X as being the lifetime of some component, then the memoryless property states that the *distribution of the remaining lifetime is independent of the time the component has already lasted.* In other words, such a probability distribution is independent of its history.

Proposition 4.1 Characterization of Memoryless Property

Let X be a random variable. Then X is memoryless if and only if

$$\mathbb{P}(X > y + z) = \mathbb{P}(X > y)\,\mathbb{P}(X > z)$$

for all $y, z \geq 0$.

Proof. Note

$$\begin{aligned} \mathbb{P}(X > y + z | X > y) &= \frac{\mathbb{P}(X > y + z, X > y)}{\mathbb{P}(X > y)} \\ &= \frac{\mathbb{P}(X > y + z)}{\mathbb{P}(X > y)}. \end{aligned}$$

Hence

$$\mathbb{P}(X > y + z | X > y) = \mathbb{P}(X > z)$$

if and only if

$$\mathbb{P}(X > z) = \frac{\mathbb{P}(X > y + z)}{\mathbb{P}(X > y)}$$

if and only if

$$\mathbb{P}(X > y+z) = \mathbb{P}(X > y)\,\mathbb{P}(X > z). \qquad \blacklozenge$$

Proposition 4.2
Exponential Distribution Is Memoryless

Let $X \sim \mathrm{EXP}(\lambda)$. Then X is memoryless.

Proof. For all $y, z \geq 0$, we have

$$\mathbb{P}(X > y+z) = e^{-\lambda(y+z)} = e^{-\lambda y}e^{-\lambda z} = \mathbb{P}(X > y)\,\mathbb{P}(X > z).$$

Thus by Proposition 4.1, X is memoryless. $\blacklozenge$

(EX 4.5) Suppose that a computer has 3 switches which govern the transfer of electronic impulses. These switches operate simulatneously and independently of one another, with lifetimes that are exponentially distributed with mean lifetimes of $10, 5, 4$ years, respectively.

(a) What is the probability that the time until the very first switch breakdown exceed 6 years?

Answer. Let X_i represent the lifetime of switch i, where $i \in \{1,2,3\}$. We know that $X_i \sim \mathrm{EXP}(\lambda_i)$, where $\lambda_1 = \frac{1}{10}, \lambda_2 = \frac{1}{5}, \lambda_3 = \frac{1}{4}$. The time until the first breakdown is defined by the random variable

$$Y = \min(X_1, X_2, X_3).$$

Since the lifetimes are independent of each other,

$$Y \sim \mathrm{EXP}(\lambda),$$

where $\lambda = \frac{1}{10} + \frac{1}{5} + \frac{1}{4} = \frac{11}{20}$ by (4.2). Thus it follows that

$$\mathbb{P}(Y > 6) = e^{-\frac{11}{20}6} = e^{-3.3} \approx 0.0369. \qquad \triangleleft$$

(b) What is the probability that switch 2 outlives switch 1?

Answer. Note that

$$\mathbb{P}(X_1 < X_2) = \frac{\frac{1}{10}}{\frac{1}{10} + \frac{1}{5}} = \frac{1}{3}. \qquad \triangleleft$$

(c) What is the probability that switch 1 has the longest lifetime, followed next by switch 3 and then switch 2?

Answer. Note that

$$\begin{aligned}\mathbb{P}(X_2 < X_3 < X_1) &= \frac{\lambda_2\lambda_3}{(\lambda_2+\lambda_3+\lambda_1)(\lambda_3+\lambda_1)} \\ &= \frac{\left(\frac{1}{5}\right)\left(\frac{1}{4}\right)}{\left(\frac{1}{5}+\frac{1}{4}+\frac{1}{10}\right)\left(\frac{1}{4}+\frac{1}{10}\right)} = \frac{\frac{1}{20}}{\frac{11}{20}\frac{7}{20}} = \frac{20}{77} \approx 0.26. \qquad \triangleleft\end{aligned}$$

(d) If switch 3 is known to have lasted 2 years, what is the probability it will last at most 3 more years?

Answer. Note that

$$\begin{aligned}\mathbb{P}(X_3 \le 5 | X_3 > 2) &= 1 - \mathbb{P}(X_3 > 5 | X_3 > 2) = 1 - \mathbb{P}(X_3 > 2 + 3 | X_3 > 2) \\ &= 1 - \mathbb{P}(X_3 > 3) = 1 - e^{-\left(\frac{1}{4}\right)3} \approx 0.528\end{aligned}$$

due to the memoryless property of exponential distributions. ◁

(e) Considering only switches $1, 2$, what is the expected amount of time until they have both suffered a breakdown?

Answer. We desire to determine

$$\mathbb{E}(\max(X_1, X_2)).$$

We note the following useful identity:

$$\min(X_1, X_2) + \max(X_1, X_2) = X_1 + X_2.$$

Thus

$$\begin{aligned}\mathbb{E}(\max(X_1, X_2)) &= \mathbb{E}(X_1) + \mathbb{E}(X_2) - \mathbb{E}(\min(X_1, X_2)) \\ &= 10 + 5 - \frac{1}{\frac{1}{10} + \frac{1}{5}} = 15 - \frac{10}{3} = \frac{35}{3} \approx 11.667.\end{aligned}$$

◁

(4.6)

(a) The exponential distribution is the *unique* continuous distribution possessing the memoryless property (incidentally, the geometric distribution is the unique discrete distribution which is memoryless).

Proof. Suppose X is a continuous random variable satisfying the memoryless property. Then by Proposition 4.2,

$$\overline{F}_X(y+z) = \overline{F}(y)\overline{F}(z)$$

for all $y, z \ge 0$. It follows immediately that

$$\overline{F}\left(\frac{m}{n}\right) = \left(\overline{F}\left(\frac{1}{n}\right)\right)^m$$

for all $m, n \in \mathbb{N}$. In particular,

$$\overline{F}(1) = \left(\overline{F}\left(\frac{1}{n}\right)\right)^n.$$

This implies

$$\overline{F}(x) = \left(\overline{F}(1)\right)^x$$

for all $x \in \mathbb{Q}, x \ge 0$, and by the continuity of $\overline{F}$, the above equality holds for irrational $x \ge 0$ as well. This implies

$$\overline{F}(x) = e^{-\lambda x}$$

for all $x \ge 0$, where $\lambda = -\ln\left(\overline{F}(1)\right)$. Thus X is exponentially distributed. ◁

(b) The memoryless property of the exponential distribution even holds in a borader setting. Specifically, if $X \sim \text{EXP}(\lambda)$, then

$$\mathbb{P}(X > Y + Z | X > Y) = \mathbb{P}(X > Z) \tag{4.1}$$

for any independently distributed nonnegative random variables Y, Z that are independent of X. We call [4.1] the *generalized memoryless property*.

Proof. Note that

$$\mathbb{P}(X > Y + Z | X > Y) = \frac{\mathbb{P}(X > Y + Z, X > Y)}{\mathbb{P}(X > Y)}.$$

Without loss of generality, assume Y, Z are continuous. This means

$$\begin{aligned}
\mathbb{P}(X > Y + Z, X > Y) &= \int_0^\infty \mathbb{P}(X > Y + Z, X > Y | Y = y) f_Y(y)\,\mathrm{d}y \\
&= \int_0^\infty \mathbb{P}(X > y + Z, X > y) f_Y(y)\,\mathrm{d}y \\
&= \int_0^\infty \mathbb{P}(X > y + Z) f_Y(y)\,\mathrm{d}y \\
&= \int_0^\infty \int_0^\infty \mathbb{P}(X > y + Z | Z = z) f_Z(z)\,\mathrm{d}z\, f_Y(y)\,\mathrm{d}y \\
&= \int_0^\infty \int_0^\infty \mathbb{P}(X > y + z) f_Z(z)\,\mathrm{d}z\, f_Y(y)\,\mathrm{d}y \\
&= \int_0^\infty \int_0^\infty e^{-\lambda(y+z)} f_Z(z)\,\mathrm{d}z\, f_Y(y)\,\mathrm{d}y \\
&= \int_0^\infty \mathbb{P}(X > Z) e^{-\lambda y} f_Y(y)\,\mathrm{d}y \\
&= \mathbb{P}(X > Z) \int_0^\infty e^{-\lambda y} f_Y(y)\,\mathrm{d}y \\
&= \mathbb{P}(X > Z)\,\mathbb{P}(X > Y)
\end{aligned}$$

by conditioning on Y, Z and using the fact that X, Y, Z are independent. Thus

$$\mathbb{P}(X > Y + Z | X > Y) = \mathbb{P}(X > Z),$$

as required. ◁

(c) The generalized memoryless property implies that $(X - Y) \mid (X > Y) \sim \text{EXP}(\lambda)$ regardless of the distribution of Y. To see this, let Z be a random variable with a degenerate distribution at z. In this case, (4.5) becomes

$$\mathbb{P}(X > Y + z | X > Y) = \mathbb{P}(X > z) = e^{-\lambda z},$$

since $X \sim \text{EXP}(\lambda)$. Thus $\mathbb{P}(X - Y > z | X > Y) = e^{-\lambda z}$, and so $(X - Y) \mid (X > Y) \sim \text{EXP}(\lambda)$.

(EX 4.7) Let X_1, X_2 be independent random variables with $X_i \sim \text{EXP}(\lambda_i)$ for all $i \in \{1, 2\}$. Given $X_1 < X_2$, show that $X_1, X_2 - X_1$ are conditionally independent random variables.

Proof. Consider the following conditional joint cdf. For all $x, y \geq 0$,

$$\begin{aligned}
\mathbb{P}(X_1 \leq x, X_2 - X_1 \leq y | X_1 < X_2) &= \frac{\mathbb{P}(X_1 \leq x, X_2 - X_1 \leq y, X_1 < X_2)}{\mathbb{P}(X_1 < X_2)} \\
&= \frac{\mathbb{P}(X_1 \leq x, X_2 - y \leq X_1, X_1 < X_2)}{\frac{\lambda_1}{\lambda_1 + \lambda_2}} \\
&\frac{\lambda_1 + \lambda_2}{\lambda_1} \mathbb{P}(X_2 - y \leq X_1 \leq \min(x, X_2)).
\end{aligned}$$

We break down into two cases.

- *Case 1. Suppose that $x \le y$.* It follows that

$$
\begin{aligned}
&\mathbb{P}(X_2 - y \le X_1 \le \min(x, X_2)) \\
&= \int_0^\infty \mathbb{P}(X_2 - y \le X_1 \le \min(x, X_2) | X_2 = w) f_{X_2}(w)\, \mathrm{dw} \\
&= \int_0^\infty \mathbb{P}(w - y \le X_1 \le \min(x, w)) f_{X_2}(w)\, \mathrm{dw} \\
&= \int_0^x \mathbb{P}(w - y \le X_1 \le \min(x, w)) f_{X_2}(w)\, \mathrm{dw} + \int_x^y \mathbb{P}(w - y \le X_1 \le \min(x, w)) f_{X_2}(w)\, \mathrm{dw} \\
&+ \int_y^{y+x} \mathbb{P}(w - y \le X_1 \le \min(x, w)) f_{X_2}(w)\, \mathrm{dw} + \int_{y+x}^\infty \mathbb{P}(w - y \le X_1 \le \min(x, w)) f_{X_2}(w)\, \mathrm{dw} \\
&= \int_0^x \mathbb{P}(X_1 \le w) \lambda_2 e^{-\lambda_2 w}\, \mathrm{dw} + \mathbb{P}(X_1 \le x) \int_x^y \lambda_2 e^{-\lambda_2 w}\, \mathrm{dw} \\
&+ \int_y^{y+x} (\mathbb{P}(X_1 > w - y) - \mathbb{P}(X_1 > x)) \lambda_2 e^{-\lambda_2 w}\, \mathrm{dw} \\
&= \int_0^x \left(1 - e^{-\lambda_1 w}\right) \lambda_2 e^{-\lambda_2 w}\, \mathrm{dw} + \left(1 - e^{-\lambda_1 x}\right)\left(e^{-\lambda_2 x} - e^{-\lambda_2 y}\right) \\
&+ \int_y^{y+x} \left(e^{-\lambda_1 (w-y)} - e^{-\lambda_1 x}\right) \lambda_2 e^{-\lambda_2} w\, \mathrm{dw} \\
&= \cdots \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2}\left(1 - e^{-\lambda_2 y} - e^{-(\lambda_1 + \lambda_2)x} + e^{-\lambda_2 y} e^{-(\lambda_1 + \lambda_2)x}\right).
\end{aligned}
$$

- *Case 2. Suppose $x > y$.* In a similar fashion to Case 1, it can be shown that

$$
\mathbb{P}(X_2 - y \le X_1 \le \min(x, X_2)) = \frac{\lambda_1}{\lambda_1 + \lambda_2}\left(1 - e^{-\lambda_2 y} - e^{-(\lambda_1 + \lambda_2)x} + e^{-\lambda_2 y} e^{-(\lambda_1 + \lambda_2)x}\right)
$$

  as well.

Therefore, in general, we have

$$
\begin{aligned}
\mathbb{P}(X_1 \le x, X_2 - X_1 \le y | X_1 < X_2) &= \frac{\lambda_1 + \lambda_2}{\lambda_1} \frac{\lambda_1}{\lambda_1 + \lambda_2}\left(1 - e^{-\lambda_2 y} - e^{-(\lambda_1 + \lambda_2)x} + e^{-\lambda_2 y} e^{-(\lambda_1 + \lambda_2)x}\right) \\
&= \left(1 - e^{-\lambda_2 y} - e^{-(\lambda_1 + \lambda_2)x} + e^{-\lambda_2 y} e^{-(\lambda_1 + \lambda_2)x}\right) \\
&= \left(1 - e^{-(\lambda_1 + \lambda_2)x}\right)\left(1 - e^{-\lambda_2 y}\right) \\
&= \mathbb{P}(X_1 \le x | X_1 < X_2)\, \mathbb{P}(X_2 - X_1 \le y | X_1 < X_2)
\end{aligned}
$$

for all $x, y \ge 0$, where the last equality follows from (b) of (EX 4.3) and the generlized memoryless property. Thus, by the definition of independence, $X_1, X_2 - X_1$ are conditionally (i.e. given $X_1 < X_2$) independent. ◁

(4.8) Erlang Distribution

Recall that if $X \sim \text{ERLANG}(n, \lambda)$ where $n \in \mathbb{N}, \lambda > 0$, then its pdf is of the form

$$
f_X(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}
$$

for all $x > 0$. Letting $n = 1$, the above pdf simplifies to become $f_X(x) = \lambda e^{-\lambda x}$ for all $x > 0$, which is the pdf of the $\text{EXP}(\lambda)$. To obtain the corresponding cdf of an $\text{ERLANG}(n, \lambda)$ random variable, we consider the following. For every $x \geq 0$,

$$F_X(x) = \mathbb{P}(X \leq x) = \int_0^x \frac{\lambda^n y^{n-1} e^{-\lambda y}}{(n-1)!} \, \mathrm{dy} = \frac{\lambda^n}{(n-1)!} \int_0^x y^{n-1} e^{-\lambda y} \, \mathrm{dy}.$$

By using the integration by parts $\int u \, \mathrm{dv} = uv - \int v \, \mathrm{du}$ with

$$u = y^{n-1}, \mathrm{dv} = e^{-\lambda y} \, \mathrm{dy},$$

we obtain that

$$\begin{aligned} \int_0^x y^{n-1} e^{-\lambda y} \, \mathrm{dy} &= -\frac{1}{\lambda} y^{n-1} e^{-\lambda y} \Big|_{y=0}^{y=x} + \frac{n-1}{\lambda} \int_0^x y^{n-2} e^{-\lambda y} \, \mathrm{dy} \\ &= -\frac{1}{\lambda} x^{n-1} e^{-\lambda x} + \frac{n-1}{\lambda} \int_0^x y^{n-2} e^{-\lambda y} \, \mathrm{dy}. \end{aligned}$$

So by induction, we obtain that

$$F_X(x) = 1 - e^{-\lambda x} \sum_{j=0}^{n-1} \frac{(\lambda x)^j}{j!}$$

for all $x \geq 0$. In particular, by substituting $n = 1$, we immediately obtain

$$F_X(x) = 1 - e^{-\lambda x}$$

for all $x \geq 0$, the cdf of an $\mathrm{EXP}(\lambda)$ random variable.

(4.9) To determine the mgf of X, we consider the following: for all $t < \lambda$,

$$\begin{aligned} \varphi_X(t) &= \int_0^\infty e^{tx} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \, \mathrm{dx} \\ &= \frac{\lambda^n}{(n-1)!} \int_0^\infty x^{n-1} e^{-\tilde{\lambda} x} \, \mathrm{dx} \\ &= \frac{\lambda^n}{\tilde{\lambda}^n} \int_0^\infty \underbrace{\frac{\tilde{\lambda}^n x^{n-1} e^{-\tilde{\lambda} x}}{(n-1)!}}_{\substack{\text{the pdf of} \\ \mathrm{ERLANG}(n, \tilde{\lambda})}} \, \mathrm{dx} \\ &= \left(\frac{\lambda}{\lambda - t} \right)^n. \end{aligned}$$

However, note that

$$\varphi_X(t) = \left(\frac{\lambda}{\lambda - t} \right)^n = \prod_{i=1}^n \left(\frac{\lambda}{\lambda - t} \right)$$

is the product of n terms, where each term is the mgf of an $\mathrm{EXP}(\lambda)$ random variable.

Now suppose that $(Y_i)_{i=1}^n$ is an iid sequence of $\mathrm{EXP}(\lambda)$ random variables. This means $\varphi_{Y_i}(t) = \frac{\lambda}{\lambda - t}$ for all $t < \lambda, i \in \{1, \ldots, n\}$. Since $\varphi_X = \prod_{i=1}^n \varphi_{Y_i}$, it follows that an Erlang distribution can be viewed as the distribution of a *sum of iid exponential random variables*. As a result, the mean and variance of an $\mathrm{ERLANG}(n, \lambda)$ random variable X are simply obtained as

$$\mathbb{E}(X) = \frac{n}{\lambda}, \mathrm{var}(X) = \frac{n}{\lambda^2}.$$

4.2 Poisson Processes

Def'n 4.2 **Counting Process**

A ***counting process*** $(N(t))_{t \geq 0}$ is a stochastic process in which $N(t)$ represents the number of events that happen by time t, where the index t measures time over a continuous range.

(4.10) Basic Properties of Counting Processes

Let $(N(t))_{t \geq 0}$ be a counting process.

(a) $N(0) = 0$.

(b) $N(t) \in \mathbb{N} \cup \{0\}$ for all $t \geq 0$.

(c) If $0 \leq s < t$, then $N(s) \leq N(t)$.

(d) If $0 \leq s < t$, then $N(t) - N(s)$ counts *the number of events to occur in the time interval* $(s, t]$ *for* $s < t$.

Def'n 4.3 **Independent Increments, Stationary Increments** of a Counting Process

We say a counting process $(N(t))_{t \geq 0}$ has

(a) ***independent increments*** if for every s_1, s_2, t_1, t_2 with

$$(s_1, t_1] \cap (s_2, t_2] = \varnothing,$$

$N(t_1) - N(s_1)$ is independent of $N(t_2) - N(s_2)$; and

(b) ***stationary increments*** if for every $s, t \geq 0$, the distribution of the number of events in $(s, s+t]$ depends only on t, the length of the time interval; that is,

$$N(s+t) - N(s) \sim N(0+t) - N(0) = N(t)$$

for all $s, t \geq 0$.

(4.11) The *assumption of stationary and independent increments* is essentially equivalent to stating that, at any point in time, *the process* $(N(t))_{t \geq 0}$ *probabilistically restarts itself.*

Recall 4.4 $o(h)$

We say a function $f : \mathbb{R} \to \mathbb{R}$ is said to be $o(h)$ if

$$\lim_{h \to 0} \frac{f(h)}{h} = 0.$$

Def'n 4.5 **Poisson Process**

A counting process $(N(t))_{t \geq 0}$ is said to be a ***Poisson process*** at rate λ if the following conditions hold.

(a) $(N(t))_{t \geq 0}$ has both independent and stationary increments.

(b) For all $h > 0$,

$$\mathbb{P}(N(h) = 1) = \lambda h + o(h) \quad [4.2]$$

$$\mathbb{P}(N(h) \geq 2) = o(h). \quad [4.3]$$

(4.12) Let $(N(t))_{t\geq 0}$ be a Poisson process at rate λ.

(a) [4.2] implies that in a *small* interval of time, the probability of a sinlge event occuring is essentially proportional to the length of the interval.

(b) [4.3] implies that two or more events occuring in a *small* interval of time is rare.

(c) [4.2], [4.3] yield

$$\begin{aligned}\mathbb{P}(N(h)=0) &= 1-\mathbb{P}(N(h)>0) = 1-\mathbb{P}(N(h)=1)-\mathbb{P}(N(h)\geq 2)\\ &= 1-(\lambda h+o(h))-o(h) = 1-\lambda h+o(h).\end{aligned}$$

Ultimately, we would like to know the distribution of $N(s+t)-N(s)$, representing the number of events occuring in the interval $(s,s+t]$ for all $s,t\geq 0$. Since a Poisson process has stationary increments this $N(s+t)-N(s)\sim N(t)$. The following proposition specifies the distribution of $N(t)$.

Proposition 4.3 *Let $(N(t))_{t\geq 0}$ be a Poisson process at rate λ. Then for any $t\geq 0$,*

$$N(t)\sim \mathrm{POI}(\lambda t).$$

Proof. For all $t\geq 0$, let φ_t denote $\varphi_{N(t)}$ for convenience. For all $t,h\geq 0$, note that

$$\begin{aligned}\varphi_{t+h}(u) &= \mathbb{E}\left(e^{uN(t+h)}\right) = \mathbb{E}\left(e^{u(N(t+h)-N(t)+N(t))}\right)\\ &= \mathbb{E}\left(e^{u(N(t+h)-N(t))}e^{uN(t)}\right) = \mathbb{E}\left(e^{u(N(t+h)-N(t))}\right)\mathbb{E}\left(e^{uN(t)}\right)\\ &= \mathbb{E}\left(e^{uN(h)}\right)\mathbb{E}\left(e^{uN(t)}\right) = \varphi_h(u)\,\varphi_t(u).\end{aligned}$$

by using stationary and independent increments of $(N(t))_{t\geq 0}$. Note that, for all $h>0, j\geq 2$,

$$0\leq \mathbb{P}(N(h)=j)\leq \mathbb{P}(N(h)\geq 2)$$

so

$$0\leq \frac{\mathbb{P}(N(h)=j)}{h}\leq \frac{\mathbb{P}(N(h)\geq 2)}{h}.$$

But $\lim_{h\to 0}\frac{\mathbb{P}(N(h)\geq 2)}{h}=0$ by Def'n 4.5. Hence by the squeeze theorem,

$$\lim_{h\to\infty}\frac{\mathbb{P}(N(h)=j)}{h}=0,$$

so $\mathbb{P}(N(h)=j)=o(h)$. Using this result, we obtain

$$\begin{aligned}\varphi_h(u) &= \mathbb{E}\left(e^{uN(h)}\right) = \sum_{j=0}^{\infty} e^{uj}\,\mathbb{P}(N(h)=j)\\ &= e^0\,\mathbb{P}(N(h)=0)+e^u\,\mathbb{P}(N(h)=1)+\sum_{j=2}^{\infty} e^{uj}\,\mathbb{P}(N(h)=j)\\ &= (1-\lambda h+o(h))+e^u(\lambda h+o(h))+\sum_{j=2}^{\infty} e^{uj}o(h)\\ &= 1-\lambda h+e^u\lambda h+o(h).\end{aligned}$$

Hence we now have

$$\varphi_{t+h}(u)=\varphi_t(u)(1-\lambda h+e^u\lambda h+o(h))=\varphi_t(u)-\lambda h\varphi_t(u)+e^u\lambda h\varphi_t(u)+o(h)$$

so

$$\varphi_{t+h}(u) - \varphi_t(u) = \lambda h \varphi_t(u)(e^u - 1) + o(h).$$

This means

$$\frac{\mathrm{d}}{\mathrm{dt}}\varphi_t(u) = \lim_{h\to 0}\frac{\varphi_{t+h}(u) - \varphi_t(u)}{h} = \lim_{h\to 0}\frac{\lambda h \varphi_t(u)(e^u - 1) + o(h)}{h} = \lambda\varphi_t(u)(e^u - 1).$$

This is a differential equation, solving which gives

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

for all $u \in \mathbb{R}$, the mgf of a $\mathrm{POI}(\lambda t)$ random variable. ♦

Corollary 4.3.1

Let $(N(t))_{t\geq 0}$ be a Poisson process. Then for all $s, t \geq 0, k \in \mathbb{N} \cup \{0\}$, we have

$$\mathbb{P}(N(s+t) - N(s) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

Def'n 4.6

Interarrival Time Sequence of a Counting Process

Let $(N(t))_{t\geq 0}$ be a counting process. Let

$$T_i = \text{elapsed time between the occurrences of the } i-1\text{th event and the } i\text{th event}$$

for all $i \in \mathbb{N}$. Then we say $(T_i)_{i\in\mathbb{N}}$ is the ***interarrival time sequence*** of $(N(t))_{t\geq 0}$.

Proposition 4.4

Let $(N(t))_{t\geq 0}$ be a Poisson process at rate $\lambda > 0$. Then $(T_i)_{i=1}^{\infty}$ is an iid $\mathrm{EXP}(\lambda)$ sequence.

Proof. We begin by considering T_1. For all $t \geq 0$, note that

$$\begin{aligned}\mathbb{P}(T_1 > t) &= \mathbb{P}(\text{no events occur before time } t)\\ &= \mathbb{P}(N(t) = 0) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t},\end{aligned}$$

which is the tpf of an $\mathrm{EXP}(\lambda)$ random variable. Immediately, $T_1 \sim \mathrm{EXP}(\lambda)$.

Next, for all $s > 0, t \geq 0$,

$$\begin{aligned}\mathbb{P}(T_2 > t | T_1 = s) &= \mathbb{P}(\text{no events occur in } (s, s+t] | \forall w \in [0, s)\, [N(w) = 0], N(s) = 1)\\ &= \mathbb{P}(\text{no events occur in } (s, s+t])\\ &= \mathbb{P}(N(s+t) - N(s) = 0)\\ &= \mathbb{P}(N(t) = 0) = e^{-\lambda t},\end{aligned}$$

where the second equality follows from the fact that $[0, s), (s, s+t]$ are disjoint. But note that

$$\mathbb{P}(T_2 > t | T_1 = s) = e^{-\lambda t}$$

is independent of s, so T_1, T_2 are independent. Moreover, by independence, we have

$$\mathbb{P}(T_2 > t) = \mathbb{P}(T_2 > t | T_1 = s) = e^{-\lambda t},$$

implying that $T_2 \sim \mathrm{EXP}(\lambda)$ as well.

We remark that the general result can be obtained by using induction on i. ♦

Def'n 4.7 **Waiting Time**
Let $(N(t))_{t\geq 0}$ be a counting process. For all $n \in \mathbb{N}$, let S_n be the total elapsed time until the nth event occurs, which is called the ***waiting time*** until the nth event occurs.

(4.13) Waiting Time

By definition, it is clear that

$$S_n = \sum_{i=1}^{n} T_i.$$

Moreover, if $(N(t))_{t\geq 0}$ is a Poisson process at rate λ, then $(T_i)_{i=1}^{\infty}$ is an iid $\text{EXP}(\lambda)$ sequence, so by Proposition 4.4, we have

$$S_n = \sum_{i=1}^{n} T_i \sim \text{ERLANG}(n, \lambda).$$

From our earlier results on the Erlang distribution, we have

$$\begin{aligned}
\mathbb{E}(S_n) &= \frac{n}{\lambda} \\
\text{var}(S_n) &= \frac{n}{\lambda^2} \\
\mathbb{P}(S_n > t) &= e^{-\lambda t} \sum_{j=0}^{n-1} \frac{(\lambda t)^j}{j!}.
\end{aligned}$$

In fact, the tpf of S_N could have been obtained without reference to the Erlang distribution. That is,

$$\begin{aligned}
\mathbb{P}(S_n > t) &= \mathbb{P}(\text{arrival time of the } n\text{th event occurs after time } t) \\
&= \mathbb{P}(\text{at most } n-1 \text{ events occur by time } t) \\
&= \mathbb{P}(N(t) \leq n-1) \\
&= \sum_{j=0}^{n-1} \frac{e^{-\lambda t} (\lambda t)^j}{j!} \\
&= e^{-\lambda t} \sum_{j=0}^{n-1} \frac{(\lambda t)^j}{j!},
\end{aligned}$$

where the second last equality follows from the fact that $N(t) \sim \text{POI}(\lambda t)$.

(4.14) If $(X_i)_{i=1}^{\infty}$ represents an iid $\text{EXP}(\lambda)$ sequence and one constructs a counting process $(N(t))_{t\geq 0}$ defined by

$$N(t) = \max\left\{n \in \mathbb{N} : \sum_{i=1}^{n} X_i \leq t\right\},$$

then $(N(t))_{t\geq 0}$ is actually a Poisson process at rate λ. In particular, we have

$$\mathbb{P}(N(t) \leq k) = \mathbb{P}\left(\sum_{i=1}^{k+1} X_i > t\right) = e^{-\lambda t} \sum_{j=0}^{k} \frac{(\lambda t)^j}{j!}$$

since $\sum_{i=1}^{k+1} X_i \sim \text{ERLANG}(k+1, \lambda)$. This implies

$$\begin{aligned}
\mathbb{P}(N(t) = k) &= \mathbb{P}(N(t) \leq k) - \mathbb{P}(N(t) \leq k-1) \\
&= e^{-\lambda t} \sum_{j=0}^{k} \frac{(\lambda t)^j}{j!} - e^{-\lambda t} \sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!} = \frac{e^{-\lambda t}(\lambda t)^k}{k!}
\end{aligned}$$

for all $k \in \mathbb{N} \cup \{0\}$.

(EX 4.15) At a local insurance company, suppose that fire damage claims come into the company according to a Poisson process at rate 3.8 expected claims per eyar.

(a) What is the probability that exactly 5 claims occur in the time interval $(3.2, 5]$ measured in years?

Answer. Let $N(t)$ be the number of claims arriving to the company in the interval $[0, t]$. Since $(N(t))_{t\geq 0}$ is a Poisson process at rate $\lambda = 3.8$, we have

$$\mathbb{P}(N(5) - N(3.2) = 5) = \mathbb{P}(N(1.8) = 5) = \frac{e^{-3.8(1.8)}(3.8(1.8))^5}{5!} \approx 0.134. \quad \triangleleft$$

(b) What is the probability that the time between the 2nd and 4th claims is between 2 and 5 months?

Answer. Let T be the time between the 2nd and 4th claims. Then

$$T = T_3 + T_4$$

where $(T_i)_{i=1}^{\infty}$ is the interval time sequence of $(N(t))_{t\geq 0}$. This means

$$T_3, T_4 \sim \mathrm{EXP}(3.8)$$

and are independent. This means

$$T \sim \mathrm{ERLANG}(2, 3.8),$$

so

$$\mathbb{P}(T > t) = e^{-3.8t} \sum_{j=0}^{2-1} \frac{(3.8t)^j}{j!} = e^{-3.8t}(1 + 3.8t)$$

for all $t \geq 0$. Thus

$$\begin{aligned}\mathbb{P}\left(\frac{1}{6} \leq T \leq \frac{1}{3}\right) &= \mathbb{P}\left(T > \frac{1}{6}\right) - \mathbb{P}\left(T > \frac{1}{3}\right) \\ &= e^{-3.8\left(\frac{1}{6}\right)}\left(1 + 3.8\left(\frac{1}{6}\right)\right) - e^{-3.8\left(\frac{1}{3}\right)}\left(1 + 3.8\left(\frac{1}{3}\right)\right) \approx 0.337. \quad \triangleleft\end{aligned}$$

Proposition 4.5

Let $(N(t))_{t\geq 0}$ be a Poisson process. Then for all $t > s \geq 0$,

$$N(s) \mid (N(t) = n) \sim \mathrm{BIN}\left(n, \frac{s}{t}\right).$$

Proof. We desire to determine the conditional distribution of $N(s) \mid (N(t) = n)$. Clearly, $N(s) \mid (N(t) = n)$ takes on values in $\{0, \ldots, n\}$. Therefore, for each $m \in \{0, \ldots, n\}$, note that

$$\begin{aligned}\mathbb{P}(N(s) = m \mid N(t) = n) &= \frac{\mathbb{P}(N(s) = m, N(t) = n)}{\mathbb{P}(N(t) = n)} \\ &= \frac{\mathbb{P}(N(s) = m, N(t) - N(s) = n - m)}{\mathbb{P}(N(t) = n)} \\ &= \frac{\mathbb{P}(N(s) = m)\,\mathbb{P}(N(t-s) = n - m)}{\mathbb{P}(N(t) = n)} \qquad \text{independent and stationary increments} \\ &= \frac{\frac{e^{-\lambda s}(\lambda s)^m}{m!} \frac{e^{-\lambda(t-s)}(\lambda(t-s))^{n-m}}{(n-m)!}}{\frac{e^{-\lambda t}(\lambda t)}{n!}} \\ &= \binom{n}{m}\left(\frac{s}{t}\right)^m\left(1 - \frac{s}{t}\right)^{n-m},\end{aligned}$$

which is the pmf of a $\mathrm{BIN}\left(n, \frac{s}{t}\right)$. ♦

(4.16) Suppose now that $(N_1(t))_{t\geq 0}, (N_2(t))_{t\geq 0}$ are independent Poisson processes at rates λ_1, λ_2, respectively. Let $S_n^{(1)}, S_n^{(2)}$ be the arrival times of the nth event for $(N_1(t))_{t\geq 0}, (N_2(t))_{t\geq 0}$, respectively. Based on our knowledge of arrival times, we know that, for each $k \in \{1,2\}$,

$$S_m^{(k)} = \sum\nolimits_{i=1}^{m} T_i^{(k)}$$

where $\left(T_i^{(k)}\right)_{i=1}^{\infty}$ is an iid sequence $\text{EXP}(\lambda_k)$. Moreover, the sequences $\left(T_i^{(1)}\right)_{i=1}^{\infty}, \left(T_i^{(2)}\right)_{i=1}^{\infty}$ are independent.

We are interested in the probability that the mth event from the first process happens before the nth event of the second process, or, equivalently,

$$\mathbb{P}\left(S_m^{(1)} < S_n^{(2)}\right).$$

Before looking at the general case. let us first examine a couple of special cases.

- *Case 1.* $m = n = 1$. Note that

$$\mathbb{P}\left(S_1^{(1)} < S_1^{(2)}\right) = \mathbb{P}\left(T_1^{(1)} < T_1^{(2)}\right) = \frac{\lambda_1}{\lambda_1+\lambda_2}.$$

- *Case 2.* $m = 2, n = 1$. Note that

$$\begin{aligned}
\mathbb{P}\left(S_2^{(1)} < S_1^{(2)}\right) &= \mathbb{P}\left(T_1^{(1)} < T_1^{(2)}\right)\mathbb{P}\left(S_2^{(1)} < S_1^{(2)} | T_1^{(1)} < T_1^{(2)}\right) \\
&\quad + \mathbb{P}\left(T_1^{(1)} > T_1^{(2)}\right)\mathbb{P}\left(S_2^{(1)} < S_1^{(2)} | T_1^{(1)} > T_1^{(2)}\right) \\
&= \mathbb{P}\left(T_1^{(1)} < T_1^{(2)}\right)\mathbb{P}\left(T_1^{(1)} + T_2^{(1)} < T_1^{(2)} | T_1^{(1)} < T_1^{(2)}\right) \\
&= \mathbb{P}\left(T_1^{(1)} < T_1^{(2)}\right)\mathbb{P}\left(T_1^{(2)} - T_1^{(1)} > T_2^{(1)} | T_1^{(2)} < T_1^{(1)}\right) \\
&= \mathbb{P}\left(T_1^{(1)} < T_1^{(2)}\right)\mathbb{P}\left(T_1^{(2)} > T_2^{(1)}\right) \\
&= \frac{\lambda_1}{\lambda_1+\lambda_2}\frac{\lambda_1}{\lambda_1+\lambda_2} \\
&= \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^2,
\end{aligned}$$

  where the second last equality holds by the generalized memoryless property.

In the general case, we realize, through the continued application of the generalized memoryless property, that $\mathbb{P}\left(S_m^{(1)} < S_n^{(2)}\right)$ is equivalent to the probability of observing m *successes* with success probability $\frac{\lambda_1}{\lambda_1+\lambda_2}$ occur before n *failures* with failure probability $\frac{\lambda_2}{\lambda_1+\lambda_2}$ in a sequence of independent Bernoulli trials.

Specifically, in a sequence of $m+j$ Bermoulli trials (where m is the number of successes and j is the number of failures), the $m+j$th trial must always be a success, and the number of failures must be no larger than $n-1$, which ultimately leads to

$$\mathbb{P}\left(S_m^{(1)} < S_n^{(2)}\right) = \sum\nolimits_{j=0}^{n-1}\binom{m+j-1}{m-1}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^m\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^j.$$

(EX 4.17) Consider the setting of (EX 4.15).

(a) If exactly 12 claims have occurred within the first 5 years, how many claims, on average, occurred within the first 3.5 years? How would this change if no claim history of the first 5 years was given?
Answer. We desire to calculate

$$\mathbb{E}(N(3.5)\,|N(5)=12).$$

Using the binomial result of Proposition 4.5 with $s=3.5, t=5, n=12$, we obtain

$$\mathbb{E}(N(3.5)\,|N(5)=12)=12\left(\frac{3.5}{5}\right)=\frac{42}{5}=8.4.$$

On the other hand,

$$\mathbb{E}(N(3.5))=(3.8)(3.5)=13.3>8.4,$$

implying that conditioning on knowledge of $N(5)$ does affect the mean of $N(3.5)$. ◁

(b) At another competing insurance company, suppose that fire damage claims arrive to the company according to a Poisson process with rate 2.9 expected claims per year. What is the probability that 3 claims arrive to this company before 2 claims arrive to the other (i.e. first) company? Assume that the insurance companies operate independently of each other.

Answer. Let $N_1(t), N_2(t)$ denote the number of claims arriving to the first and second company by time $t \geq 0$, respectively. We are assuming here that $(N_1(t))_{t\geq 0}, (N_2(t))_{t\geq 0}$ are independent Poisson processes, with rate $\lambda_1 = 3.8, \lambda_2 = 2.9$, respectively. Then we have

$$\begin{aligned}
&\mathbb{P}(\text{3 claims to arrive to company 2 before 2 claims arrive to company 1})\\
&=\mathbb{P}\left(S_3^{(2)}<S_2^{(1)}\right)\\
&=1-\mathbb{P}\left(S_2^{(1)}<S_3^{(2)}\right)\\
&=\sum\nolimits_{j=0}^{3-1}\binom{2+j-1}{2-1}\left(\frac{3.8}{3.8+2.9}\right)^2\left(\frac{2.9}{3.8+2.9}\right)^j\\
&\approx 0.219.
\end{aligned}$$

◁

(4.18) The next property we examine concerns the classification (i.e. splitting) of events from a Poisson process into (potentially) several different types.

For a Poisson process $(N(t))_{t\geq 0}$ at rate λ, suppose that events can be independently classified as being one of k possible types, with probability $p_i \in [0,1]$ of being of type i, where $i \in \{1,\ldots,k\}$, with $\sum_{i=1}^k p_i = 1$.

Let $(N_i(t))_{t\geq 0}$ be the associated counting process for type-i events, where $i \in \{1,\ldots,k\}$. Clearly, by construction,

$$\sum\nolimits_{i=1}^k N_i(t)=N(t).$$

We now show that each $(N_i(t))_{t\geq 0}$ is a Poisson process.

Proof. Let $s,t \geq 0, k \in \{1,\ldots,k\}$. Then

$$\begin{aligned}
\mathbb{P}(N_i(s+t)-N_i(s)=m_i) &= \sum\nolimits_{n=m_i}^{\infty}\mathbb{P}(N_i(s+t)-N_i(s)=m_i|N(s+t)-N(s)=n)\,\mathbb{P}(N(s+t)-N(s)=n\\
&=\sum\nolimits_{n=m_i}^{\infty}\binom{n}{m_i}p_i^{m_i}(1-p_i)^{n-m_i}\,\mathbb{P}(N(t)=n)\\
&=\sum\nolimits_{n=1}^{\infty}\mathbb{P}(N_i(t)=m_i|N(t)=n)\,\mathbb{P}(N(t)=n)\\
&=\mathbb{P}(N_i(t)=m_i),
\end{aligned}$$

showing that $(N_i(t))_{t\geq 0}$ has stationary increments.

Next, suppose that $(s_1,t_1], (s_2,t_2]$ are disjoint time intervals. For each $i \in \{1,\ldots,k\}$, note that by the independent increments property of $(N(t))_{t\geq 0}$, the number of events in each of these intervals, $N(t_1) - N(s_1), N(t_2) - N(s_2)$ are independent. Therefore, in combination with the fact that the classification of each event is an independent process, it must hold that the number of type-i events to occur in these intervals, $N_i(t_1) - N_i(s_1), N_i(t_2) - N_i(s_2)$, are also independent, implying that $(N_i(t))_{t\geq 0}$ possesses the independent increments property.

Finally,

$$\begin{aligned}
&\mathbb{P}(N_1(t) = m_1, \ldots, N_k(t) = m_k) \\
&= \sum_{n=0}^{\infty} \mathbb{P}(N_1(t), \ldots, N_k(t) = m_k | N(t) = n) \mathbb{P}(N(t) = n) \\
&= \mathbb{P}\left(N_1(t) = m_1, \ldots, N_k(t) = m_k | N(t) = \sum_{j=1}^{k} m_j\right) \mathbb{P}\left(N(t) = \sum_{j=1}^{k} m_j\right) \\
&= \frac{\left(\sum_{i=1}^{k} m_i\right)!}{\prod_{i=1}^{k} m_i!} \prod_{i=1}^{k} p_i^{m_i} \frac{e^{-\lambda t}(\lambda t)^{\sum_{i=1}^{k} m_i}}{\left(\sum_{i=1}^{k} m_k\right)!} \\
&= \prod_{i=1}^{k} \frac{e^{-\lambda p_i t}(\lambda p_i t)^{m_i}}{m_i!} \\
&= \prod_{i=1}^{k} \mathbb{P}(N_i(t) = m_i).
\end{aligned}$$

Thus $N_1(t), \ldots, N_k(t)$ are independent Poisson random variables.

As a result, for each $i \in \{1,\ldots,k\}$, we have

$$\mathbb{P}(N_i(t)) = \cdots = \lambda p_i t + o(t)$$

and

$$\mathbb{P}(N_i(t) \geq 2) = \cdots = o(t).$$

Thus $(N_1(t))_{t\geq 0}, \ldots, (N_k(t))_{t\geq 0}$ are independent Poisson processes with rates $\lambda p_1, \ldots, \lambda p_k$, respectively. ◁

(EX 4.19) Consider the setting of (EX 4.15).

(a) Suppose that fire damage claims can be categorized as being comercial, business, or residual. At the first insurance company, past history suggests that 15% of the claims are commercial, 25% of them are business, and the remaining 60% are residental. Over the course of the next 4 years, what is the probability that the company experiences fewer than 5 claims in each of the 3 categories?

Answer. Let $N_c(t), N_b(t), N_r(t)$ denote the number of commercial, business, and residental claims by time $t \geq 0$, r espectively. It follows that

$$\begin{aligned}
N_c(4) &\sim \mathrm{POI}(3.8 \cdot 0.15 \cdot 4) = \mathrm{POI}(2.28) \\
N_b(4) &\sim \mathrm{POI}(3.8 \cdot 0.25 \cdot 4) = \mathrm{POI}(3.8) \\
N_r(4) &\sim \mathrm{POI}(3.8 \cdot 0.6 \cdot 4) = \mathrm{POI}(9.12).
\end{aligned}$$

Hence

$$\begin{aligned}
\mathbb{P}(N_c(4) < 5, N_b(4) < 5, N_r(4) < 5) &= \mathbb{P}(N_c(4) = 5)\mathbb{P}(N_b(4) = 5)\mathbb{P}(N_r(4) = 5) \\
&= \sum_{i=0}^{4} \frac{e^{-2.28} 2.28^i}{i!} \sum_{i=0}^{4} \frac{e^{-3.8} 3.8^i}{i!} \sum_{i=0}^{4} \frac{e^{-9.12} 9.12^i}{i!} \\
&\approx 0.91857 \cdot 0.66784 \cdot 0.05105 \\
&\approx 0.0313,
\end{aligned}$$

where the first equality is from the fact that $N_c(t), N_b(t), N_r(t)$ are independent for all $t \geq 0$. ◁

(4.20) We also remark that it is also possible to *merge independent Poisson process together.* In particular, if $(N_1(t))_{t\geq 0}, \dots, (N_m(t))_{t\geq 0}$ are m independent Poisson processes at respective rates $\lambda_1, \dots, \lambda_m$, then $(N(t))_{t\geq 0}$ with

$$N(t) = \sum_{i=1}^{m} N_i(t)$$

for all $t \geq 0$ is a Poisson process at rate $\sum_{i=1}^{m} \lambda_i$.

(4.21) Proposition 4.5 indicated that the conditional distribution of $N(s) \mid (N(t) = n)$, where $s < t$, is binomial with n trials and success probability $\frac{s}{t}$. In other words, it is possible to view each event that occurred within $[0,t]$ as being independent of the others, and the probability of any one event landing within the interval $[0,s]$ as being governed by the cdf of a $U(0,t)$ random variable evaluated at s. The idea that we can view $\frac{s}{t}$ as a *uniform probability* is no coincidence. In fact, the following result confirms this notion.

Proposition 4.6

Suppose that $(N(t))_{t\geq 0}$ is a Poisson process at rate λ. Given $N(t) = 1$, the conditional distribution of the first arrival time is uniformly distributed on $(0,t)$. That is,

$$S_1 \mid (N(t) = 1) \sim \mathrm{U}(0,t).$$

Proof. In order to identify the distribution of $S_1 \mid (N(t) = 1)$, consider its cdf: for every $s \in [0,t]$,

$$\begin{aligned}
F_{S_1|(N(t)=1)}(s) &= \mathbb{P}(S_1 \leq s | N(t) = 1) \\
&= \frac{\mathbb{P}(S_1 \leq s, N(t) = 1)}{\mathbb{P}(N(t) = 1)} \\
&= \frac{\mathbb{P}(1 \text{ event in } [0,s] \text{ and } 0 \text{ events in } (s,t])}{\mathbb{P}(N(t) = 1)} \\
&= \frac{\mathbb{P}(N(s) = 1, N(t) - N(s) = 0)}{\mathbb{P}(N(t) = 1)} \\
&= \frac{\mathbb{P}(N(s))\,\mathbb{P}(N(t-s) = 0)}{\mathbb{P}(N(t) = 1)} && \text{independent and stationary increments} \\
&= \frac{\frac{e^{-\lambda s}(\lambda s)^1}{1!} \frac{e^{-\lambda(t-s)}(\lambda(t-s))^0}{0!}}{\frac{e^{-\lambda t}(\lambda t)!}{1!}} \\
&= \frac{s}{t},
\end{aligned}$$

which is the cdf of a $\mathrm{U}(0,t)$ random variable. This concludes the proof. ♦

(4.22) Order Statistics

Proposition 4.6 specifies how S_1 behaves *distributionally* when $N(t) = 1$. A natural question to ask is, *how are the n arrival times $S_1, \dots, S_n$ distributed if it is known that exactly n arrivals have occured by time t (i.e. $N(t) = n$)?* Before we can address this more general question, we must familiarize ourselves with some distributional results about *order statistics*.

In what follows, let $(Y_i)_{i=1}^n$ be an iid sequence of random variables having a common *continuous* distribution on $(0,\infty)$ with cdf $F : [0,\infty) \to [0,1]$ and pdf $f : (0,\infty) \to [0,1]$.

Order Statistic of a Sequence of Random Variables

Recall 4.8 For each $i \in \{1, \dots, n\}$, let $\left(Y_{(i)}\right)_{i=1}^{n}$ be defined as

$$Y_{(i)} = i\text{th smallest among } Y_1, \dots, Y_n.$$

We call $Y_{(i)}$ the ***ith order statistic*** of $(Y_i)_{i=1}^n$ and the sequence $\left(Y_{(i)}\right)_{i=1}^{n}$ the ***order statistics*** of $(Y_i)_{i=1}^n$.

By definition, we observe that

$$Y_{(1)} \leq \cdots \leq Y_{(n)}.$$

Let the joint cdf of $\left(Y_{(i)}\right)_{i=1}^{n}$ be denoted by

$$G(y_1, \dots, y_n) = \mathbb{P}\left(Y_{(1)} \leq y_1, \dots, Y_{(n)} \leq y_n\right)$$

and also define

$$g(y_1, \dots, y_n) = \frac{\partial^n G(y_1, \dots, y_n)}{\partial \mathrm{y}_1 \cdots \partial \mathrm{y}_n}$$

to be the corresponding joint pdf. We desire to determine an expression for g.

Answer. To begin, consider the case when $n = 2$ and assume $0 < y_1 < y_2$. Note that

$$\begin{aligned}\mathbb{P}\left(Y_{(1)} \leq y_1, y_1 < Y_{(2)} \leq y_2\right) &= \mathbb{P}\left(Y_{(1)} \leq y_1, Y_{(2)} \leq y_2\right) - \mathbb{P}\left(Y_{(1)} \leq y_1, Y_{(2)} \leq y_1\right) \\ &= G(y_1, y_2) - G(y_1, y_1),\end{aligned}$$

which means

$$G(y_1, y_2) = \mathbb{P}\left(Y_{(1)} \leq y_1, y_1 < Y_{(2)} \leq y_2\right) + G(y_1, y_1).$$

This means

$$\begin{aligned}g(y_1, y_2) &= \frac{\partial^2 G(y_1, y_2)}{\partial \mathrm{y}_1\, \partial \mathrm{y}_2} \\ &= \frac{\partial^2 \mathbb{P}\left(Y_{(1)} \leq y_1, y_1 < Y_{(2)} \leq y_2\right)}{\partial \mathrm{y}_1\, \partial \mathrm{y}_2} + \frac{\partial^2 G(y_1, y_1)}{\partial \mathrm{y}_1\, \partial \mathrm{y}_2} \\ &= \frac{\partial^2 \mathbb{P}\left(Y_{(1)} \leq y_1, y_1 < Y_{(2)} \leq y_2\right)}{\partial \mathrm{y}_1\, \partial \mathrm{y}_2}.\end{aligned}$$

This fact is true for all $n \in \mathbb{N}$, so that

$$g(y_1, \dots, y_n) = \frac{\partial^n \mathbb{P}\left(Y_{(1)} \leq y_1, y_1 < Y_{(2)} \leq y_2, \dots, y_{n-1} < Y_{(n)} \leq y_n\right)}{\partial y_1 \partial y_2 \cdots \partial y_n}$$

for all $y_1, \dots, y_n \in (0, \infty)$ with $y_1 < \cdots < y_n$.

If we now examine the case when $n = 2$ again with $0 < y_1 < y_2$, we see that

$$\begin{aligned}&\mathbb{P}\left(Y_{(1)} \leq y_1, y_1 < Y_{(2)} \leq y_2\right) \\ &= \mathbb{P}\left(Y_{(1)} \leq y_1, y_1 < Y_{(2)} \leq y_2, \{Y_1 < Y_2\} \cup \{Y_1 > Y_2\}\right) \\ &= \mathbb{P}\left(Y_{(1)} \leq y_1, y_1 < Y_{(2)} \leq y_2, Y_1 < Y_2\right) + \mathbb{P}\left(Y_{(1)} \leq y_1, y_1 < Y_{(2)} \leq y_2, Y_1 > Y_2\right) \\ &= \mathbb{P}(Y_1 \leq y_1, y_1 < Y_2 \leq y_2, Y_1 < Y_2) + \mathbb{P}(Y_2 \leq y_1, y_1 < Y_1 \leq y_2, Y_1 > Y_2) \\ &= \mathbb{P}(Y1 \leq y_1, y_1 < Y_2 \leq y_2) + \mathbb{P}(Y_2 \leq y_1, y_1 < Y_1 \leq y_2) \\ &= 2\mathbb{P}(Y_1 \leq y_1)\mathbb{P}(y_1 < Y_2 \leq y_2) && \text{since } Y_1, Y_2 \text{ are iid} \\ &= 2F(y_1)(F(y_2) - F(y_1)).\end{aligned}$$

As a result,

$$g(y_1, y_2) = \frac{\partial^2}{\partial y_1 \, \partial y_2} 2F(y_1)(F(y_2) - F(y_1)) = 2f(y_1) f(y_2).$$

It can be shown that

$$g(y_1, \dots, y_n) = n! \prod_{i=1}^{n} f(y_i)$$

for all $n \in \mathbb{N}$ with $y_1, \dots, y_n \in (0, \infty)$ where $0 < y_1 < \cdots < y_n$.

We can also find the marginal cdf and pdf of $Y_{(i)}$ as follows:

$$G_i(y_i) = 1 - \sum_{j=0}^{i-1} \binom{n}{j} F(y_i)^j (1 - F(y_i))^{n-j}$$

and

$$g_i(y_i) = \frac{n!}{(n-i)!(i-1)!} F(y_i)^{i-1} f(y_i)(1 - F(y_i))^{n-1}$$

for all $y_i \in (0, \infty)$.

In particular, when $Y_i \sim \mathrm{U}(0, t)$ for each $i \in \{1, \dots, n\}$, then

$$g(y_1, \dots, y_n) = \frac{n!}{t^n}$$

for all $y_1, \dots, y_n \in (0, t)$ with $0 < y_1 < \cdots < y_n < t$ and

$$g_i(y_i) = \frac{n! y_i^{i-1} (t - y_i)^{n-i}}{(n-i)!(i-1)! t^n}$$

for all $y_i \in (0, t)$. ◁

With these results in place, we are now in position to state another important result concerning the Poisson process.

Proposition 4.7

Let $(N(t))_{t \geq 0}$ be a Poisson process at rate λ. Given $N(t) = n$, the conditional joint distribution of the n arrival times is identical to the joint distribution of the n order statistics from the $U(0, t)$ distribution. In other words,

$$(S_1, \dots, S_n) \mid (N(t) = n) \sim \left(Y_{(1)}, \dots, Y_{(n)}\right),$$

where $(Y_i)_{i=1}^{n}$ is an iid $U(0, t)$ sequence.

Proof. Let $s_1, \dots, s_n \in (0, t)$ with $s_1 < \cdots < s_n$. Then

$$\begin{aligned}
&\mathbb{P}(S_1 \leq s_1, s_1 < S_2 \leq s_2, \dots, s_{n-1} < S_n \leq s_n | N(t) = n) \\
&= \frac{\mathbb{P}(S_1 \leq s_1, s_1 < S_2 \leq s_2, \dots, s_{n-1} < S_n \leq s_n, N(t) = n)}{\mathbb{P}(N(t) = n)} \\
&= \frac{\mathbb{P}(N(s_1) = 1, N(s_2) - N(s_1) = 1, \dots, N(s_n) - N(s_{n-1}) = 1, N(t) - N(s_n) = 0)}{\mathbb{P}(N(t) = n)} \\
&= \frac{\mathbb{P}(N(s_1) = 1)\mathbb{P}(N(s_2 - s_1) = 1) \cdots \mathbb{P}(N(s_n - s_{n-1}) = 1)\mathbb{P}(N(t - s_n) = 0)}{\mathbb{P}(N(t) = n)} && \text{independent and stationary increments} \\
&= \frac{\left(e^{-\lambda s_1} \lambda s_1\right)\left(e^{-\lambda(s_2 - s_1)} \lambda (s_2 - s_1)\right) \cdots \left(e^{-\lambda(s_n - s_{n-1})} \lambda (s_n - s_{n-1})\right)\left(e^{-\lambda(t-s)}\right)}{\frac{e^{-\lambda t}(\lambda t)}{n!}} \\
&= \frac{n! s_1 (s_2 - s_1) \cdots (s_n - s_{n-1})}{t^n}.
\end{aligned}$$

Therefore, the joint pdf of $(S_i)_{i=1}^n \mid (N(t)=n)$ can be obtained via partial differentiation:

$$\frac{\partial^n \mathbb{P}\left(S_1 \leq s_1, s_1 < S_2 \leq s_2, \ldots, s_{n-1} < S_n \leq s_n | N(t)=n\right)}{\partial s_1\, \partial s_2 \cdots \partial s_n} = \frac{n!}{t^n},$$

which is the joint pdf of an order statistics of iid $\mathrm{U}(0,t)$ random variables. ♦

(4.23) What this result essentially implies is that under the condition that n events have occured by time t in a Poisson process, the n times at which those events occur are distributed *independently* and *uniformly* over the interval $[0,1]$.

(EX 4.24) Cars arrive to a toll bridge according to a Poisson process at rate λ, where each car pays a toll of \$1 upon arrival. Caclulate the mean and variance of the total amount collected by time t, *discounted* back to time 0 where $\alpha > 0$ is the discount rate per unit time.

Answer. Let $N(t)$ count the number of cars arriving to the toll bridge by time $t \geq 0$, where $(N(t))_{t\geq 0}$ is a Poisson process at rate λ and $N(t) \sim \mathrm{POI}(\lambda t)$ for all $t \geq 0$. Let S_i denote the arrival time of the ith car to the toll bridge, where $i \in \mathbb{N}$. Then the discounted value (i.e. back to time 0) of \$1 paid by the ith arrival is given by

$$1 \cdot e^{-\alpha S_i} = e^{-\alpha S_i}.$$

Let T represent the (discounted) total amount collected by time t, so that

$$T = \sum_{i=1}^{N(t)} e^{-\alpha S_i}.$$

We desire to find $\mathbb{E}(T), \mathrm{var}(T)$.

To find $\mathbb{E}(T)$, note that

$$\begin{aligned}
\mathbb{E}(T) &= \mathbb{E}\left(\sum_{i=1}^{N(t)} e^{-\alpha S_i}\right) \\
&= \sum_{n=1}^{\infty} \mathbb{E}\left(\sum_{i=1}^{N(t)} e^{-\alpha S_i} | N(t)=n\right) \mathbb{P}(N(t)=n) \\
&= \sum_{n=1}^{\infty} \mathbb{E}\left(\sum_{i=1}^{n} e^{-\alpha S_i} | N(t)=n\right) \mathbb{P}(N(t)=n) \\
&= \sum_{n=1}^{\infty} \mathbb{E}\left(\sum_{i=1}^{n} e^{-\alpha Y_{(i)}}\right) \mathbb{P}(N(t)=n) && \text{let } \left(Y_{(i)}\right)_{i=1}^n \text{ be the order statistics from } \mathrm{U}(0,t) \text{ and use Proposition 4.7} \\
&= \sum_{n=1}^{\infty} \mathbb{E}\left(\sum_{i=1}^{n} e^{-\alpha Y_i}\right) \mathbb{P}(N(t)=n) && \text{since we are summing each } e^{-\alpha Y_i} \text{ once anyways} \\
&= \sum_{n=1}^{\infty} n\,\mathbb{E}\left(e^{-\alpha Y_1}\right) \mathbb{P}(N(t)=n) && Y_1, \ldots, Y_n \text{ are iid} \\
&= \sum_{n=1}^{\infty} n \int_0^t e^{-\alpha y} \frac{1}{t}\, \mathrm{d}y\, \mathbb{P}(N(t)=n) \\
&= \sum_{n=1}^{\infty} n \frac{1-e^{-\alpha t}}{\alpha t} \mathbb{P}(N(t)=n) \\
&= \frac{1-e^{-\alpha t}}{\alpha t} \mathbb{E}(N(t)) \\
&= \frac{1-e^{-\alpha t}}{\alpha t} \lambda t \\
&= \frac{\lambda}{\alpha}\left(1-e^{-\alpha t}\right).
\end{aligned}$$

To determine $\operatorname{var}(T)$, we once again apply Proposition 4.7. That is, for all $n \in \mathbb{N}$,

$$\begin{aligned}
\operatorname{var}(T|N(t)=n) &= \operatorname{var}\left(\sum_{i=1}^{n} e^{-\alpha Y_{(i)}}\right) \\
&= \operatorname{var}\left(\sum_{i=1}^{n} e^{-\alpha Y_i}\right) \\
&= n \operatorname{var}\left(e^{-\alpha Y_1}\right) && Y_1, \ldots, Y_n \text{ are iid} \\
&= n\left(\mathbb{E}\left(\left(e^{-\alpha Y_1}\right)^2\right) - \mathbb{E}\left(e^{-\alpha Y_1}\right)^2\right) \\
&= n\left(\mathbb{E}\left(e^{-2\alpha Y_1}\right) - \left(\frac{1-e^{-\alpha t}}{\alpha t}\right)^2\right) \\
&= n\left(\frac{1-e^{-2\alpha t}}{2\alpha t} - \frac{\left(1-e^{-\alpha t}\right)}{\alpha^2 t^2}\right)
\end{aligned}$$

and so

$$\begin{aligned}
\operatorname{var}(T|N(t)) &= \operatorname{var}(T|N(t)=n)\,|_{n=N(t)} \\
&= N(t)\left(\frac{1-e^{-2\alpha t}}{2\alpha t} - \frac{\left(1-e^{-\alpha t}\right)}{\alpha^2 t^2}\right).
\end{aligned}$$

Finally, applying the conditional variance formula, we get

$$\begin{aligned}
\operatorname{var}(T) &= \operatorname{var}(\mathbb{E}(T|N(t))) + \mathbb{E}(\operatorname{var}(T|N(t))) \\
&= \operatorname{var}\left(N(t)\frac{1-e^{-\alpha t}}{\alpha t}\right) + \mathbb{E}\left(N(t)\left(\frac{1-e^{-2\alpha t}}{2\alpha t} - \frac{\left(1-e^{-\alpha t}\right)^2}{\alpha^2 t^2}\right)\right) \\
&= \frac{\left(1-e^{-\alpha t}\right)^2}{\alpha^2 t^2}\operatorname{var}(N(t)) + \left(\frac{1-e^{-2\alpha t}}{2\alpha t} - \frac{\left(1-e^{-\alpha t}\right)^2}{\alpha^2 t^2}\right)\mathbb{E}(N(t)) \\
&= \frac{\lambda\left(1-e^{-2\alpha t}\right)}{2\alpha}.
\end{aligned}$$

◁

(EX 4.25) Sattlites are launched at times according to a Poisson process at rate 3 per year. During the past year, it was observed that only two satellites were launched. What is the joint probability that the first of these two satellites was launched in the first 5 months of the year and the second satellite was launched prior to the last 2 months of the year?

Answer. Let $(N(t))_{t\geq 0}$ be the Poisson process at rate $\lambda = 3$ governing satellite launches. We are interested in calculating

$$\mathbb{P}\left(S_1 \leq \frac{5}{12}, S_2 \leq \frac{5}{6}|N(1)=2\right).$$

To do so, we use (4.22) to obtain the joint conditional pdf g of $(S_1, S_2)\,|N(1)=2$ as follows: for every $s_1, s_2 \in (0,1)$, with $s_1 < s_2$,

$$g(s_1, s_2) = \frac{2!}{1^2} = 2.$$

Thus,

$$
\begin{aligned}
\mathbb{P}\left(S_1 \le \frac{5}{12}, S_2 \le \frac{5}{6} | N(1) = 2\right) &= \int_0^{\frac{5}{12}} \int_{s_1}^{\frac{5}{6}} g(s_1, s_2) \, \mathrm{d}s_1 \, \mathrm{d}s_2 \\
&= \int_0^{\frac{5}{12}} \int_{s_1}^{\frac{5}{6}} 2 \, \mathrm{d}s_1 \, \mathrm{d}s_2 \\
&= 2 \int_0^{\frac{5}{12}} \frac{5}{6} - s_1 \, \mathrm{d}s_1 \\
&= 2\left(\frac{5}{6} \cdot \frac{5}{12} - \frac{\left(\frac{5}{12}\right)^2}{2}\right) \\
&= \frac{25}{48} \approx 0.521.
\end{aligned}
$$

◁

4.3 Generalizations of Poisson Processes

(4.26) Oftentimes, we find the Poisson process difficult to apply in applications of real-life phenomena, largely due to the fact that the Poisson process assumes a *constant* arrival rate of λ for all time. In what follows, we consider a more general type of process in which the arrival rate is allowed to *vary* as a function of time.

Non-homogeneous (Non-stationary) Poisson Process

Def'n 4.9 The counting process $(N(t))_{t \ge 0}$ is a ***non-homogeneous*** (or ***non-stationary***) Poisson process with rate function $\lambda : [0, \infty) \to \mathbb{R}$ if the following three conditions hold: for all $t, h \ge 0$,

(a) $(N(t))_{t \ge 0}$ has independent increments;

(b) $\mathbb{P}(N(t+h) - N(t) = 1) = h\lambda(t) + o(h)$; and

(c) $\mathbb{P}(N(t+h) - N(t) \ge 2) = o(h)$.