STAT 333

Stochastic Processes I

This page intentionally left blank.

Contents

This page intentionally left blank.

1.
Review of Probability

1.1 Probability Spaces

Probability Space

Def'n 1.1 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 $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$ *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 $A \in \mathcal{F}$,

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

Conditional Probability

Def'n 1.2 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,

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

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

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}\left(A_i | \bigcap_{j=1}^{i} A_j\right),$$ *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

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

(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
$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \tag{1.1}$$
for all $x \in \{0, \dots, n\}$. 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
$$p_X(x) = \binom{x-1}{n-1} p^n (1-p)^{x-n} \tag{1.2}$$
for every $x \in \mathbb{N}, x \geq n$.

(a) Note that the apparence 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

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

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. ◁

Continuous Random Variable

Def'n 1.6

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

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

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 $\mathrm{sd}(X)$, by

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

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

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 $X \sim \mathrm{BIN}(n,p)$, 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

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

(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 $i \in \{1, \ldots, n\}$,

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

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

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

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

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

Def'n 1.14 **Joint MGF** of a Random Vector

Let $\mathbf{X} = (X_1, \ldots, X_n)$ 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 $i_1, \ldots, i_n \in \mathbb{N} \cup \{0\}$,

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

1.5 Independence

Independent Random Variables

Def'n 1.15 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 $Y \sim \mathrm{BIN}(\sum_{i=1}^m n_i, p)$. It follows from (1.20) that $\sum_{i=1}^m X_i \sim \mathrm{BIN}(\sum_{i=1}^m n_i, p)$. $\triangleleft$

Convergence of a Sequence of Random Variables

Def'n 1.16 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.

Conditional PMF

Def'n 2.1 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 $p_{X_2}(x_2) > 0$.

Conditional Expectation

Def'n 2.2 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 $X_1\sim\operatorname{BIN}(n_1,p),X_2\sim\operatorname{BIN}(n_2,p)$ 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

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

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

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

Here is an intuitive explanation of why $X_1|(X_1+X_2=m)\sim\operatorname{HG}(n_1+n_2,n_1,m)$. 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 $B_1,\dots,B_{n_1+n_2}$ are successes, and we also know exactly n_1 of $B_1,\dots,B_{n_1}$ 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 $X_1\sim\operatorname{POI}(\lambda_1),\dots,X_m\sim\operatorname{POI}(\lambda_m)$ for some $\lambda_1,\dots,\lambda_m>0$ be independent and let $Y=\sum_{i=1}^m X_i$. Find the conditional distribution of $X_j|(Y=n)$, where $j\in\{1,\dots,m\},n\in\mathbb{N}$.

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

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

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 $\mathrm{var}(X|Y=y)$, as

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

Proposition 2.4

We have

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

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

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

It follows that

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

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

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

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

Answer. By the law of total probability,

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

Moreover,

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

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

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

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 $Z = X \,|\, (X < Y)$. Show that

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

are identically distributed.

Proof. Observe that

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

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

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

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

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

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 $N_1 \sim \mathrm{GEO}(p)$, 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 $x_0,\dots,x_{n+1}$ in the codomain of $X_0,\dots,X_{n+1}$, respectively,

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

(3.2) In other words, the Markov property states that the conditional distribution of a *future* state X_{n+1} given the *past* states $X_0,\dots,X_{n-1}$ 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 $k_1,\dots,k_l \in \{1,\dots,n-1\}$ with $k_1 < \cdots < k_l$,

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

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 $n \in \mathbb{N}, m \in \{0,\dots,n\}$, 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

$$\alpha_n = \alpha_0 P^n. \qquad [3.5]$$

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

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

Similarly,

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

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,

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

(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

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

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

Answer. We have

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

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

Answer. We have

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

3.2 Accessibility and Communication

Def'n 3.7 **Accessble** State

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

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

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.

Def'n 3.8 **Irreducible, Reducible** DTMC

A DTMC is called

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

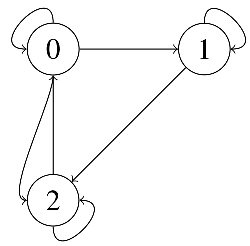
(b) ***reducible*** if not irreducible.

(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.

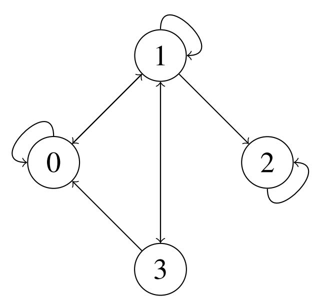
Answer. By drawing a state transition diagram, it is clear that the DTMC is irreducible. ◁

(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

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

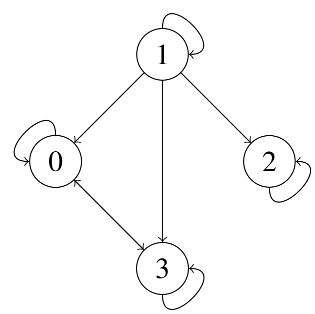
If $d(i) = 1$, then we say i is ***aperiodic***, and if every state of a DTMC is aperiodic, then we call the DTMC ***aperiodic***.

(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

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

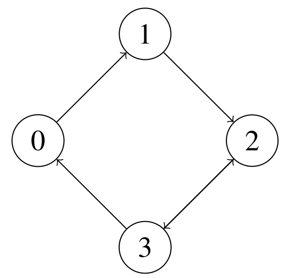
for all $i \in \{0,1,2,3\}$. Thus we conclude that the DTMC is aperiodic. ◁

(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 $i \in \{0,1,2,3\}$. ◁

(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 $d(2) = d(3) = 2$. ◁

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

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

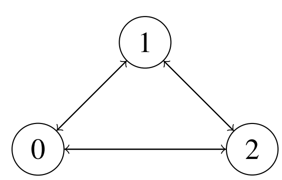
Thus $d(j) = d(i)$, as required. ∎

(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 $f_{i,j}^{(n)}$.

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

Notation 3.10 Consider the setting of (3.20). We write $f_{i,j}^{(n)}$ to denote

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

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:

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

This means

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

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

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

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

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

$f_{i,j}$

Notation 3.11 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

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

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 $\mathcal{S} = \{0, \ldots, N\}$ 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

$$P_{j,i}^{(n)} = 0 \qquad [3.8]$$

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. $\triangleleft$

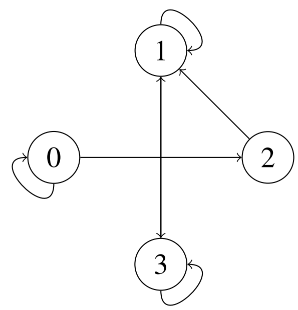
(3.24) As the above example demonstrates, *once a DTMC enters a recurrent class of states, it can never leave that class*. For this reason, a recurrent class is often referred to as a *closed* class.

(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. $\triangleleft$

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

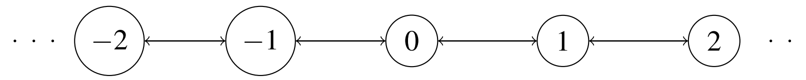
$$X_n = \sum_{k=0}^{n} Y_k,$$

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

$$\mathbb{P}(Y_k = -1) = 1 - p$$
$$\mathbb{P}(Y_k = 1) = p$$

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.



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,

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

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

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