Probability

P1 Probability, random variables and expectation

P1.1 Sample spaces and events

Definition The *sample space* Ω for a random experiment is the set of all possible outcomes of the random experiment.

Definition An *event* relating to an experiment is a subset of Ω.

Example P1.1 The sample space for the toss of two fair coins is

$$\Omega = \{HH, HT, TH, TT\}.$$

Let event A be obtaining at least one head, then

$$A = \{HH, HT, TH\}.$$

Example P1.2 The sample space of an individual's weight change under a diet and exercise regime is

$$\Omega = \{w : -\infty < w < \infty\}.$$

Let event L be the individual loses at least 10 kgs, then

$$L = \{w : w \leq -10\}.$$

Note that Ω may be discrete (finite or countable) or continuous.

P1.2 Probability

Possible interpretations of probability:

1. Classical interpretation. Assuming that all outcomes of an

experiment are equally likely, then the probability of an event $A = \frac{n(A)}{n(\Omega)}$, where $n(A)$ is the number of outcomes satisfying A and $n(\Omega)$ is the number of outcomes in Ω.

2. Frequency interpretation. The probability of an event is the relative frequency of observing a particular outcome when an experiment is repeated a large number of times under similar circumstances.

3. Subjective interpretation. The probability of an event is an individual's perception as to the likelihood of an event's occurrence.

Definition A *probability* (*measure*) is a real-valued set function P defined on the events (subsets) of a sample space Ω satisfying the following three axioms (see Kolmogorov, 1933):

A1. $P(E) \geq 0$ for any event E;

A2. $P(\Omega) = 1$;

A3. If $E_1, E_2, \ldots$ is any infinite sequence of disjoint events (i.e. $E_i \cap E_j = \emptyset$ for all $i \neq j$, then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

Note that all of the other properties of probability (measures) that we use are derived from these three axioms.

Exercise Using only the axioms above, prove:
1. $0 \leq P(E) \leq 1$ for any event E.
2. $P(E^C) = 1 - P(E)$ where E^C is the complement of E.
3. $P(\emptyset) = 0$.
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

P1.3 Conditional probability

Definition The *conditional probability* of an event E *given* an event F is

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}, \quad \text{provided } P(F) > 0.$$

Note if $P(F) > 0$, then $P(E \cap F) = P(E|F)P(F)$.

Example P1.3 Consider the experiment of tossing a fair 6-sided die. What is the probability of observing a 2 if the outcome was even?

Let event T be observing a 2 and let event E be the outcome is even. Find $P(T|E)$.

$$P(T|E) = \frac{P(T \cap E)}{P(E)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Total probability theorem Let $E_1, E_2, \ldots$ be a partition of Ω (i.e. $E_i \cap E_j = \emptyset$ for all $i \neq j$ and $\bigcup_{i=1}^{\infty} E_i = \Omega$) and let $F \subseteq \Omega$ be any event. Then,

$$P(F) = \sum_{i=1}^{\infty} P(F \mid E_i)\, P(E_i).$$

Example P1.4 Suppose a factory uses three different machines to produce tin cans. Machine I produces 50% of all cans, machine II produces 30% of all cans and machine III produces the rest of the cans. It is known that 4% of cans produced on machine I are defective, 2% of the cans produced on machine II are defective and 5% of the cans produced on machine III are defective. If a can is selected at random, what is the probability that it is defective?

Let event M_i be the can is produced by machine i, $i = 1, 2, 3$. Let D be the event that the can is defective.

$$\begin{aligned} P(M_1) &= 0.5 \qquad P(D|M_1) = 0.04 \\ P(M_2) &= 0.3 \qquad P(D|M_2) = 0.02 \\ P(M_3) &= 0.2 \qquad P(D|M_3) = 0.05 \\ P(D) &= \sum_{i=1}^{3} P(D|M_i)P(M_i) \\ &= (0.04 \times 0.5) + (0.02 \times 0.3) + (0.05 \times 0.2) = 0.036. \end{aligned}$$

Bayes' Formula Let $E_1, E_2, \ldots, E_n$ be the partition of Ω (i.e. $E_i \cap E_j = \emptyset$ for all $i \neq j$ and $\bigcup_{i=1}^{n} E_i = \Omega$) such that $P(E_i) > 0$ for all $i = 1, \ldots, n$, and let $F \subseteq \Omega$ be any event such that $P(F) > 0$. Then

$$P(E_k|F) = \frac{P(F|E_k)P(E_k)}{\sum_{i=1}^{n} P(F|E_i)P(E_i)}.$$

Proof: If $P(F) > 0$ and $P(E_k) > 0$, then by definition

$$P(E_k|F) = \frac{P(E_k \cap F)}{P(F)} = \frac{P(F|E_k)P(E_k)}{P(F)}.$$

Since $E_1, E_2, \ldots, E_n$ is a partition of Ω such that $P(E_i) > 0$ for all i, then by the total probability theorem we can rewrite $P(F)$ and obtain

$$P(E_k|F) = \frac{P(F|E_k)P(E_k)}{\sum_{i=1}^{n} P(F|E_i)P(E_i)}.$$

Example P1.5 Consider example P1.4. Suppose now that we randomly select a can and find that it is defective. What is the probability that it was produced by machine I?

$$P(M_1|D) = \frac{P(D|M_1)P(M_1)}{P(D)} = \frac{0.04 \times 0.5}{0.036} = 0.55.$$

P1.4 Independence

Definition Events $E_1, E_2, \ldots, E_N$ are said to be *independent* if, for *any* finite subset $\{i_1, i_2, \ldots, i_n\} \subseteq \{1, \ldots, N\}$,

$$P\left(\bigcap_{j=1}^{n} E_{i_j}\right) = \prod_{j=1}^{n} P(E_{i_j}).$$

Note, in particular, two events E and F are independent if

$$P(E \cap F) = P(E)P(F)$$

Theorem If $P(F) > 0$, two events, E and F, are independent if and only if $P(E|F) = P(E)$.

Proof:

$$\begin{aligned} P(E|F) &= P(E) \\ \frac{P(E \cap F)}{P(F)} &= P(E) \\ P(E \cap F) &= P(E)P(F). \end{aligned}$$

P1.5 Random variables

Definition A random variable (r.v.) X is a function of outcome or a mapping from Ω to $\mathbb{R}$: $X : \Omega \longrightarrow \mathbb{R}$.

For example,
(a) Let X be the number of heads observed when tossing a fair coin three times.

(b) Let T be the length of time you wait to be serviced by a bank teller.

Note: Random variables can be either discrete (i.e. take a finite or countable number of values), continuous, or mixed.

Definition The *(cumulative) distribution function* (c.d.f.) is defined by

$$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

Properties of the c.d.f:

1. $P(X > x) = 1 - F_X(x)$.
2. $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$.

Note the c.d.f. is defined for all random variables regardless of whether they are discrete, continuous or mixed.

Definition If X is a **discrete** random variable, then we can define a function $p_X(x)$, called the *probability mass function* (p.m.f.) such that

$$p_X(x_i) = P(X = x_i) = P(\{\omega : X(\omega) = x_i\}).$$

Example P1.6 Let X be the number of heads observed when tossing a fair coin three times. What is the p.m.f. of X?

$$p_X(x) = \begin{cases} 1/8 & \text{if } x = 0 \\ 3/8 & \text{if } x = 1 \\ 3/8 & \text{if } x = 2 \\ 1/8 & \text{if } x = 3 \\ 0 & \text{otherwise.} \end{cases}$$

Definition Let X be a **continuous** random variable. If there exists some nonnegative function f_X on $\mathbb{R}$ such that for any interval I,

$$P(X \in I) = \int_I f_X(u)du,$$

the function f_X is called the *probability density function* (p.d.f.) of X.

Note that if $F_X(x)$ is the c.d.f. of a continuous random variable X, then the p.d.f. of X is given by

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

Note that

$$F_X(x) = P(X \leq x) = \begin{cases} \sum_{x_i \leq x} p_X(x_i) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{x} f_X(u)du & \text{if } X \text{ is continuous.} \end{cases}$$

P1.6 Expectation

Definition The *expectation* of a random variable X is defined by

$$E[X] = \begin{cases} \sum_{x_i} x_i p_X(x_i) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x f_X(x)dx & \text{if } X \text{ is continuous.} \end{cases}$$

Note that $E[X]$ only exists if $E[|X|] < \infty$ and that $E[X]$ is a measure of the "centre" of the distribution (i.e. "the centre of mass").

Definition If $Y = g(X)$ then the expectation of Y is given by

$$\begin{aligned} E[Y] &= E[g(X)] \\ &= \begin{cases} \sum_{x_i} g(x_i)p_X(x_i) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} g(x)f_X(x)dx & \text{if } X \text{ is continuous.} \end{cases} \end{aligned}$$

Properties of expectation For given constants c, c_i and d,

1. $E[c] = c$;
2. $E[cg(X) + d] = cE[g(X)] + d$;
3. $E\left[\sum_{i=1}^{n} c_i g_i(X)\right] = \sum_{i=1}^{n} c_i E[g_i(X)]$.

Definition The *variance* of X is

$$\mathsf{Var}(X) = E\left[(X - E[X])^2\right].$$

The *standard deviation* of X is $\sqrt{\mathsf{Var}(X)}$.

Properties of variance

1. $\mathsf{Var}(X) = E[X^2] - (E[X])^2$;

2. $\mathsf{Var}(X) \geq 0$;

3. $\mathsf{Var}(cX + d) = c^2 \mathsf{Var}(X)$;

4. If $X_1, \ldots, X_n$ are independent and $c_1, \ldots, c_n$ are any given constants, then

$$\mathsf{Var}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i^2 \mathsf{Var}(X_i).$$

See the supplementary sheet for specific examples of discrete and continuous distributions and some of their specific characteristics.

P2 Joint distribution functions

P2.1 Joint c.d.f. and p.d.f.

Definition The *joint (cumulative) probability distribution function* (joint c.d.f.) of X and Y is defined by

$$\begin{aligned} F_{X,Y}(x,y) &= P(\{\omega : X(\omega) \leq x \text{ and } Y(\omega) \leq y\}) \\ &= P(X \leq x, Y \leq y), \quad x, y \in \mathbb{R}. \end{aligned}$$

Definition Two r.v.'s X and Y are said to be jointly continuous, if there exists a function $f_{X,Y}(x,y) \geq 0$ such that for *every* 'nice' set $C \subseteq \mathbb{R}^2$,

$$P((X,Y) \in C) = \int\int_C f_{X,Y}(x,y)dxdy.$$

The function $f_{X,Y}$ is called the *joint probability density function* (joint p.d.f.) of X and Y.

If X and Y are jointly continuous, then

$$\begin{aligned} F_{X,Y}(x,y) &= P(X \leq x, Y \leq y) \\ &= \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)dudv. \end{aligned}$$

Hence,

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}.$$

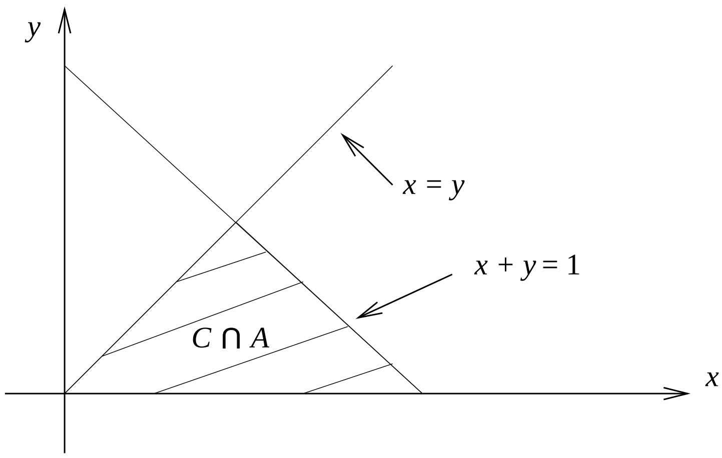
Example P2.1 Suppose that

$$f_{X,Y}(x,y) = \begin{cases} 24x(1-x-y) & x \geq 0, y \geq 0, x+y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find $P(X > Y)$, (ii) Find $P(X > \frac{1}{2})$.

(i) Let $C = \{(x,y) : x > y\}$ and write $A = \{(x,y) : f_{X,Y}(x,y) > 0\}$. Then,

$$C \cap A = \{(x,y); x > 0, y > 0, x+y < 1, x > y\}.$$



$$
\begin{aligned}
P(X > Y) &= P((X, Y) \in C) \\
&= \int\int_C f_{X,Y}(x, y)dxdy \\
&= \int\int_{C\cap A} 24x(1 - x - y)dxdy \\
&= \int_0^{1/2}\int_y^{1-y} 24x(1 - x - y)dxdy \\
&= \int_0^{1/2} \left[12x^2 - 8x^3 - 12yx^2\right]_y^{1-y} dy \\
&\vdots \\
&= \int_0^{1/2} 4 - 12y + 16y^3 dy \\
&= \left[4y - 6y^2 + 4y^4\right]_0^{1/2} \\
&= 2 - \frac{3}{2} + \frac{1}{4} = \frac{3}{4}.
\end{aligned}
$$

(ii) Let $D = \{(x, y) : x > 1/2\}$, then

$$D \cap A = \{(x, y); x > 1/2, y > 0, x + y < 1\}.$$

$$
\begin{aligned}
P(X > 1/2) &= P((X,Y) \in D) \\
&= \int\int_D f_{X,Y}(x,y)dxdy \\
&= \int\int_{D\cap A} 24x(1-x-y)dxdy \\
&= \int_{1/2}^{1}\int_{0}^{1-x} 24x(1-x-y)dydx \\
&= \int_{1/2}^{1} \left[24xy\left(1-x-\frac{1}{2}y\right)\right]_0^{1-x} dx \\
&= \int_{1/2}^{1} 12x(1-x)^2dx \\
&= \left[\frac{12}{2}x^2 - \frac{24}{3}x^3 + \frac{12}{4}x^4\right]_{1/2}^{1} \\
&= 6 - 8 + 3 - \frac{3}{2} + 1 - \frac{3}{16} = \frac{5}{16}.
\end{aligned}
$$

P2.2 Marginal c.d.f. and p.d.f.

Definition Suppose that the c.d.f. of X and Y is given by $F_{X,Y}$, then the c.d.f. of X can be obtained from $F_{X,Y}$ since

$$
\begin{aligned}
F_X(x) &= P(X \le x) = P(X \le x, Y < \infty) \\
&= \lim_{y\to\infty} F_{X,Y}(x,y)
\end{aligned}
$$

F_X is called the *marginal distribution* (marginal c.d.f.) of X.

Definition If $f_{X,Y}$ is the joint p.d.f. of X and Y, then the *marginal probability density function* (marginal p.d.f.) of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$$

Example P2.2 Consider example P2.1. Find the marginal p.d.f. and c.d.f of Y.

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \\
&= \begin{cases} \int_0^{1-y} 24x(1-x-y)dx & 0 \le y \le 1, \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} 4(1-y)^3 & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

Hence,

$$
F_Y(y) = \begin{cases} 0 & y < 0, \\ \int_0^y 4(1-u)^3 du = 1-(1-y)^4 & 0 \le y \le 1, \\ 1 & y > 1. \end{cases}
$$

Example P2.3 Suppose that

$$
f_{X,Y}(x,y) = \begin{cases} e^{-(x+y)} & 0 < x, y < \infty, \\ 0 & \text{otherwise,} \end{cases}
$$

and let $Z = X/Y$. Find the p.d.f. of Z.

Clearly, $Z > 0$. For $z > 0$,

$$
\begin{aligned}
F_Z(z) &= P(Z \le z) = P(X/Y \le z) \\
&= \int\int_{\{(x,y):x/y \le z\}} f_{X,Y}(x,y)dxdy \\
&= \int_0^{\infty}\int_0^{yz} e^{-(x+y)}dxdy \\
&= \int_0^{\infty} -e^{-y(1+z)} + e^{-y}dy \\
&= 1 - \frac{1}{1+z}
\end{aligned}
$$

and so

$$
f_Z(z) = \frac{dF_Z(z)}{dz} = \begin{cases} \frac{1}{(1+z)^2} & z > 0 \\ 0 & z \le 0. \end{cases}
$$

Note that we can extend the notion of joint and marginal distributions to random variables $X_1, X_2, \ldots, X_n$ in a similar fashion.

P2.3 Independent random variables

Definition Random variables X and Y are said to be *independent* if, for all $x, y \in \mathbb{R}$,

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y),$$

i.e., for all $x, y \in \mathbb{R}$, $F_{X,Y}(x, y) = F_X(x)F_Y(y)$.

If X and Y are discrete random variables with joint p.m.f. $p_{X,Y}(x, y)$ and marginal p.m.f.'s $p_X(x)$ and $p_Y(y)$, respectively, then X and Y are independent if and only if for all $x, y \in \mathbb{R}$,

$$p_{X,Y}(x, y) = p_X(x)p_Y(y).$$

If X and Y are continuous random variables with joint p.d.f. $f_{X,Y}(x, y)$ and marginal p.d.f.'s $f_X(x)$ and $f_Y(y)$, respectively, then X and Y are independent if and only if for all $x, y \in \mathbb{R}$,

$$f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

Note that we can extend the notion of independent random variables to random variables $X_1, X_2, \ldots, X_n$

Definition The random variables $X_1, X_2, \ldots, X_n$ are said to *independent and identically distributed* (i.i.d.) if,
1. $X_1, X_2, \ldots, X_n$ are independent.
2. $X_1, X_2, \ldots, X_n$ all come from the same distribution function (i.e. $X_i \sim F$ for all $i = 1, \ldots, n$).

Definition The random variables $X_1, X_2, \ldots, X_n$ are said to be a *random sample* if they are i.i.d.

Example P2.4 Suppose $X_1, X_2, \ldots, X_n$ are a random sample from the Poisson distribution with mean λ. Find the joint p.m.f. of $X_1, X_2, \ldots, X_n$.

If $X_i \sim Poi(\lambda)$, then its p.m.f. is given by

$$p_{X_i}(x_i) = \begin{cases} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} & \text{if } x_i = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

Since $X_1, X_2, \ldots, X_n$ are independent, their joint p.m.f. is given by,

$$\begin{aligned} p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) &= \prod_{i=1}^{n} p_{X_i}(x_i) \\ &= \begin{cases} \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} & \text{if } x_i = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \\ &= \begin{cases} \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} & \text{if } x_i = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

P3 Conditional distribution and conditional expectation

P3.1 Conditional distribution

Recall that for any two events E and F such that $P(F) > 0$, we defined in Section P1.3 that

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Can we extend this idea to random variables?

Definition If X and Y are discrete random variables, the *conditional probability mass function* of X given $Y = y$ is

$$\begin{aligned} p_{X|Y}(x|y) &= P(X = x|Y = y) \\ &= \begin{cases} \frac{p_{X,Y}(x,y)}{p_Y(y)} & \text{if } p_Y(y) > 0 \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

where $p_{X,Y}(x, y)$ is the joint p.m.f. of X and Y and $p_Y(y)$ is the marginal p.m.f. of Y for any x and y such that $p_Y(y) > 0$.

Note that
1. $P(X = x|Y = y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \geq 0$
2.

$$\begin{aligned} \sum_x P(X = x|Y = y) &= \sum_x \frac{p_{X,Y}(x, y)}{p_Y(y)} \\ &= \frac{1}{p_Y(y)} \sum_x p_{X,Y}(x, y) \\ &= \frac{1}{p_Y(y)} p_Y(y) = 1. \end{aligned}$$

This implies that $P(X = x|Y = y)$ is itself a p.m.f.

Definition If X and Y are discrete random variables, the *conditional (cumulative) probability distribution function* of X

given $Y = y$ is

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \sum_{x' \leq x} p_{X|Y}(x'|y).$$

Example P3.1 Suppose the joint p.m.f. of X and Y is given by the following probability table.

		Y			
		0	1	2	3
	0	0	1/42	2/42	3/42
X	1	2/42	3/42	4/42	5/42
	2	4/42	5/42	6/42	7/42

Determine the conditional p.m.f. of Y given $X = 1$.

$$P_{Y|X}(y|x = 1) = \frac{p_{X,Y}(x = 1, y)}{p_X(x = 1)} = \frac{p_{X,Y}(x = 1, y)}{14/42}$$

The conditional p.m.f. of Y given $X = 1$ is therefore

$$P_{Y|X}(y|x = 1) = \begin{cases} \frac{2/42}{14/42} = \frac{2}{14} & \text{if } y = 0, \\ \frac{3/42}{14/42} = \frac{3}{14} & \text{if } y = 1, \\ \frac{4/42}{14/42} = \frac{4}{14} & \text{if } y = 2, \\ \frac{5/42}{14/42} = \frac{5}{14} & \text{if } y = 3. \end{cases}$$

We cannot extend this idea to the continuous case directly since $P_Y(Y = y) = 0$.

Definition If X and Y have a joint p.d.f. $f_{X,Y}$, then the *conditional probability density function* of X, given that $Y = y$, is defined by

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } f_Y(y) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Definition Furthermore, we can define the *conditional (cumulative) probability distribution function* of X, given $Y = y$, as

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \int_{-\infty}^{x} f_{X|Y}(u|y)du.$$

Example P3.2 Suppose the joint p.d.f. of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} 24x(1-x-y) & x \geq 0, y \geq 0, x+y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find (a) the conditional p.d.f. of X given $Y = y$ and (b) the conditional p.d.f. of X given $Y = \frac{1}{2}$.

(a) In example P2.2 we found

$$f_Y(y) = \begin{cases} 4(1-y)^3 & 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \begin{cases} \frac{24x(1-x-y)}{4(1-y)^3} & x \geq 0, y \geq 0, x+y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

(b)

$$\begin{aligned} f_{X|Y}\left(x|\frac{1}{2}\right) &= \frac{f_{X,Y}(x,\frac{1}{2})}{f_Y(\frac{1}{2})} \\ &= \begin{cases} \frac{24x(1/2-x)}{4(1/2)^3} = 48x(\frac{1}{2}-x) & 0 \leq x \leq \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Note that conditional pdf's are themselves pdf's and have all the properties associated with pdf's.

P3.2 Conditional expectation

Definition The *conditional expectation* of X, given $Y = y$, is defined by

$$E[X|Y=y] = \begin{cases} \sum_x x p_{X|Y}(x|y) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx & \text{if } X \text{ is continuous.} \end{cases}$$

Since $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$, then $f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y)$. Consequently, we can reconstruct the joint p.d.f. (p.m.f.) if either
(a) we are given the conditional p.d.f. (p.m.f.) of X given $Y = y$ and the marginal p.d.f. (p.m.f.) of Y, or
(b) we are given the conditional p.d.f. (p.m.f.) of Y given $X = x$ and the marginal p.d.f. (p.m.f.) of X.

Example P3.3 Suppose that the joint p.d.f. of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} e^{-(\frac{x}{y}+y)}y^{-1} & 0 < x, y < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

For $y > 0$, find $P(X > 1|Y = y)$ and $E[X|Y = y]$.

For $y > 0$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_0^{\infty} e^{-(\frac{x}{y}+y)}y^{-1}dx = e^{-y}$$

Hence, for $y > 0$,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} e^{-x/y}y^{-1} & x > 0, \\ 0 & x \leq 0. \end{cases}$$

Thus,

$$\begin{aligned} P(X > 1|Y = y) &= \int_1^\infty f_{X|Y}(x|y)dx \\ &= \int_1^\infty e^{-x/y}y^{-1}dx = e^{-1/y}, \end{aligned}$$

and

$$E[X|Y = y] = \int_{-\infty}^\infty x f_{X|Y}(x|y)dx = \int_0^\infty \frac{x}{y}e^{-x/y}dx = y.$$

P3.3 Independent random variables

See section P2.3. If X and Y are independent continuous random variables, then for any y s.t. $f_Y(y) > 0$

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

$\forall\, x \in \mathbb{R}$.

P4 Expectation, covariance and correlation

P4.1 Expectation of a function of random variables

Definition If $X_1, X_2, \ldots, X_n$ are jointly continuous, then the *expectation of the function* $g(X_1, X_2, \ldots, X_n)$ is

$$E[g(X_1, \ldots, X_n)] = \int \cdots \int_{\mathbb{R}^n} g(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \cdots dx_n$$

Note that if $X_1, X_2, \ldots, X_n$ are discrete, we replace the integral by summations and the joint p.d.f. with the joint p.m.f.

Properties of expectation

1. $E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$.

2. If X and Y are independent, then $E[XY] = E[X]E[Y]$.

3. If X and Y are independent and g and h are any real functions, then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.

P4.2 Covariance

Definition The *covariance* of two random variables, X and Y, is defined by

$$\text{Cov}(X, Y) = E\left[(X - E[X])(Y - E[Y])\right]$$

Properties of covariance

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

2. $\text{Cov}(X, X) = \text{Var}(X)$.

3. $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$.

4.

$$\mathrm{Cov}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \mathrm{Cov}(X_i, Y_j).$$

5. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$.

6.

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i) + 2 \sum_{1\le i<j\le n} a_i a_j \mathrm{Cov}(X_i, X_j).$$

7. If $X_1, X_2, \ldots, X_n$ are independent, then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

8. If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$. The converse is NOT true.

Example P4.1 Suppose X and Y are discrete random variables whose probability mass function is given by the following table:

		X			
		-1	0	1	$p_Y(y)$
Y	0	0	1/3	0	1/3
	1	1/3	0	1/3	2/3
	$p_X(x)$	1/3	1/3	1/3	

What is the covariance of X and Y? Are X and Y independent?

$$\begin{aligned}
E[X] &= -1(1/3) + 0(1/3) + 1(1/3) = 0 \\
E[Y] &= 0(1/3) + 1(2/3) = 2/3 \\
E[XY] &= -1(0)(0) + 0(0)(1/3) + 1(0)(0) \\
&\quad -1(1)(1/3) + 0(1)(0) + 1(1)(1/3) \\
&= 0 \\
\text{Cov}(X,Y) &= E[XY] - E[X]E[Y] = 0 - 0 = 0.
\end{aligned}$$

However, $p_{X,Y}(0,0) = 1/3 \neq p_X(0)p_Y(0) = (1/3)(1/3)$. Therefore, X and Y are not independent, but $\text{Cov}(X,Y) = 0$.

P4.3 Correlation

Definition If $\text{Var}(X) > 0$ and $\text{Var}(Y) > 0$, then the *correlation* of X and Y is defined by

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Properties of correlation

1. $-1 \leq \rho(X,Y) \leq 1$.

2. If X and Y are independent, then $\rho(X,Y) = 0$. Note, again, that the converse is not true.

3. $\rho(aX+b, cY+d) = \begin{cases} \rho(X,Y) & \text{if } ac > 0, \\ -\rho(X,Y) & \text{if } ac < 0. \end{cases}$

P5 Central limit theorem

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables (i.e. a random sample) with finite mean μ and variance σ^2. Let $S_n = X_1 + \cdots + X_n$. Then

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

Alternatively,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1).$$

Example P5.1 Suppose $X_1, X_2, \ldots, X_{100}$ are i.i.d. exponential random variables with parameter $\lambda = 4$.
(a) Find $P(S_{100} > 30)$.
(b) Find limits within which $\bar{X}$ will lie with probability 0.95

(a) Since $X_1, X_2, \ldots, X_{100}$ are i.i.d. exponential random variables with parameter $\lambda = 4$, $E[X_i] = \frac{1}{4}$ and $\text{Var}(X_i) = \frac{1}{16}$. Hence,

$$\begin{aligned} E[S_{100}] &= 100\frac{1}{4} = 25; \\ \text{Var}(S_{100}) &= 100\frac{1}{16} = \frac{25}{4}. \end{aligned}$$

Given $n = 100$, S_{100} is approximately normally distributed by the central limit theorem (CLT). Therefore,

$$\begin{aligned} P(S_{100} > 30) &= P\left(\frac{S_{100} - 25}{\sqrt{\frac{25}{4}}} > \frac{30 - 25}{\sqrt{\frac{25}{4}}}\right) \\ &\approx P(Z > 2) = 1 - P(Z \leq 2) \\ &= 0.0228. \end{aligned}$$

(b) Since $X_1, X_2, \ldots, X_{100}$ are i.i.d. exponential random variables with parameter $\lambda = 4$, $E[X_i] = \frac{1}{4}$ and $\mathrm{Var}(X_i) = \frac{1}{16}$. Therefore, $E[\bar{X}] = \frac{1}{4}$ and $\mathrm{Var}(\bar{X}) = \frac{1/16}{100}$.

Since $n = 100$, $\bar{X}$ will be approximately normally distributed by the CLT, hence

$$\begin{aligned} 0.95 &= P(a < \bar{X} < b) \\ &= P\left(\frac{a - 1/4}{\sqrt{1/1600}} < \frac{\bar{X} - 1/4}{\sqrt{1/1600}} < \frac{b - 1/4}{\sqrt{1/1600}}\right) \\ &\approx P\left(\frac{a - 1/4}{\sqrt{1/1600}} < Z < \frac{b - 1/4}{\sqrt{1/1600}}\right) \\ &= P(-z_{.025} < Z < z_{.025}). \end{aligned}$$

This implies,

$$\begin{aligned} \frac{a - 1/4}{\sqrt{1/1600}} &= -z_{.025} = -1.96, \\ \frac{b - 1/4}{\sqrt{1/1600}} &= z_{.025} = 1.96. \end{aligned}$$

Hence,

$$\begin{aligned} a &= 0.25 - 1.96\frac{1}{40} = 0.201, \\ b &= 0.25 + 1.96\frac{1}{40} = 0.299. \end{aligned}$$

P6 Transformations

P6.1 Univariate case

Suppose X is a continuous random variable with p.d.f. $f(x)$. Let g be a continuous function, then $Y = g(X)$ is a continuous random variable. Our aim is to find the p.d.f. of Y.

The distribution function method has two steps:
1. Compute the c.d.f. of Y, i.e.

$$F_Y(y) = P(Y \leq y).$$

2. Derive the p.d.f. of Y, $f_Y(y)$, using the fact that

$$f_Y(y) = \frac{dF_Y(y)}{dy}.$$

Example P6.1 Let $Z \sim N(0,1)$. Find the p.d.f. of $Y = Z^2$.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(Z^2 \leq y) = P(-\sqrt{y} \leq Z \leq \sqrt{y}) \\ &= P(Z \leq \sqrt{y}) - P(Z \leq -\sqrt{y}) \\ &= F_Z(\sqrt{y}) - F_Z(-\sqrt{y}) \end{aligned}$$

Note that if we want a specific formula for F_Y, then we can evaluate the resulting c.d.f.'s. In this case,

$$F_Z(\sqrt{y}) = \int_{-\infty}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

Therefore,

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} = \frac{d}{dy}F_Z(\sqrt{y}) - \frac{d}{dy}F_Z(-\sqrt{y}) \\ &= \frac{d}{dz}F_Z(z)\frac{d}{dy}(z) - \frac{d}{dz}F_Z(-z)\frac{d}{dy}(-z) \end{aligned}$$

where $z = y^{1/2}$. Now $\frac{d}{dz}F_Z(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$, so

$$\begin{aligned}
f_Y(y) &= \frac{1}{\sqrt{2\pi}}e^{-\frac{(\sqrt{y})^2}{2}}\frac{1}{2\sqrt{y}} - \frac{1}{\sqrt{2\pi}}e^{-\frac{(-\sqrt{y})^2}{2}}\frac{-1}{2\sqrt{y}} \\
&= \frac{1}{2\sqrt{2\pi y}}e^{-\frac{y}{2}} + \frac{1}{2\sqrt{2\pi y}}e^{-\frac{y}{2}} \\
&= \frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}}
\end{aligned}$$

if $y > 0$.

Note that $Y = Z^2$ has a Chi-squared distribution with 1 degree of freedom.

P6.2 Bivariate case

Suppose X_1 and X_2 are continuous random variables with joint p.d.f. given by $f_{X_1,X_2}(x_1, x_2)$ Let $(Y_1, Y_2) = T(X_1, X_2)$. We want to find the joint p.d.f. of Y_1 and Y_2.

Definition Suppose $T : (x_1, x_2) \to (y_1, y_2)$ is a one-to-one transformation in some region of $\mathbb{R}^2$, such that $x_1 = H_1(y_1, y_2)$ and $x_2 = H_2(y_1, y_2)$. The *Jacobian* of $T^{-1} = (H_1, H_2)$ is defined by

$$J(y_1, y_2) = \begin{vmatrix} \frac{\partial H_1}{\partial y_1} & \frac{\partial H_1}{\partial y_2} \\ \frac{\partial H_2}{\partial y_1} & \frac{\partial H_2}{\partial y_2} \end{vmatrix}.$$

Theorem Let $(Y_1, Y_2) = T(X_1, X_2)$ be some transformation of random variables. If T is a one-to-one function and the Jacobian of T^{-1} is non-zero in $T(A)$ where

$$A = \{(x_1, x_2) : f_{X_1,X_2}(X_1, X_2) > 0\},$$

then the joint p.d.f. of Y_1 and Y_2, $f_{Y_1,Y_2}(y_1, y_2)$, is given by

$$f_{X_1,X_2}(H_1(y_1, y_2), H_2(y_1, y_2))|J(y_1, y_2)|$$

if $(y_1, y_2) \in T(A)$ and 0 otherwise.

Example P6.2 Let $X_1 \sim U(0,1)$, $X_2 \sim U(0,1)$ and suppose that X_1 and X_2 are independent. Let

$$Y_1 = X_1 + X_2, \quad Y_2 = X_1 - X_2.$$

Find the joint p.d.f. of Y_1 and Y_2.

The joint p.d.f. of X_1 and X_2 is

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= f_{X_1}(x_1)f_{X_2}(x_2) \\ &= \begin{cases} 1 & 0 \le x_1 \le 1, 0 \le x_2 \le 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Now $T : (x_1,x_2) \mapsto (y_1,y_2)$ is defined by

$$y_1 = x_1 + x_2, \quad y_2 = x_1 - x_2.$$

Hence,

$$x_1 = H_1(y_1,y_2) = \frac{y_1+y_2}{2}, \quad x_2 = H_2(y_1,y_2) = \frac{y_1-y_2}{2}.$$

The Jacobian of T^{-1} is

$$J(y_1,y_2) = \begin{vmatrix} \frac{\partial H_1}{\partial y_1} & \frac{\partial H_1}{\partial y_2} \\ \frac{\partial H_2}{\partial y_1} & \frac{\partial H_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}.$$

Since $A = \{(x_1,x_2) : 0 \le x_1 \le 1, 0 \le x_2 \le 1\}$ and since the lines $x_1 = 0$, $x_1 = 1$, $x_2 = 0$ and $x_2 = 1$ map to the lines $y_1 + y_2 = 0$, $y_1 + y_2 = 2$, $y_1 - y_2 = 0$ and $y_1 - y_2 = 2$, it can be checked that

$$T(A) = \{(y_1,y_2) : 0 \le y_1 + y_2 \le 2, 0 \le y_1 - y_2 \le 2\}.$$

Thus,

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= \begin{cases} \frac{1}{2} f_{X_1,X_2}(H_1(y_1,y_2), H_2(y_1,y_2)) & (y_1,y_2) \in T(A), \\ 0 & \text{otherwise.} \end{cases} \\ &= \begin{cases} \frac{1}{2} & 0 \le y_1 + y_2 \le 2, 0 \le y_1 - y_2 \le 2, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Example P6.3 Suppose that X_1 and X_2 are i.i.d. exponential random variables with parameter λ. Let $Y_1 = \frac{X_1}{X_2}$ and $Y_2 = X_1 + X_2$.
(a) Find the joint p.d.f. of Y_1 and Y_2.
(b) Find the p.d.f. of Y_1.

(a) Since X_1 and X_2 are i.i.d. exponential random variables with parameter λ, the joint p.d.f. of X_1 and X_2 is given by

$$\begin{aligned}
f_{X_1,X_2}(x_1,x_2) &= f_{X_1}(x_1)f_{X_2}(x_2) \\
&= \begin{cases} \lambda e^{-\lambda x_1}\lambda e^{-\lambda x_2} & \text{if } x_1, x_2 > 0 \\ 0 & \text{otherwise.} \end{cases} \\
&= \begin{cases} \lambda^2 e^{-\lambda(x_1+x_2)} & \text{if } x_1, x_2 > 0 \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Solving simultaneously for X_1 and X_2 in terms of Y_1 and Y_2, gives $X_1 = Y_1 X_2$ and

$$Y_2 = X_1 + X_2 = Y_1 X_2 + X_2 = X_2(Y_1 + 1).$$

Hence, $X_2 = \frac{Y_2}{Y_1+1}$ and $X_1 = Y_1 X_2 = \frac{Y_1 Y_2}{Y_1+1}$.

Computing the Jacobian of T^{-1}, we get

$$\begin{aligned}
J(y_1,y_2) &= \begin{vmatrix} \frac{\partial H_1}{\partial y_1} & \frac{\partial H_1}{\partial y_2} \\ \frac{\partial H_2}{\partial y_1} & \frac{\partial H_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{y_2}{(y_1+1)^2} & \frac{y_1}{y_1+1} \\ -\frac{y_2}{(y_1+1)^2} & \frac{1}{y_1+1} \end{vmatrix} \\
&= \frac{y_2}{(y_1+1)^3} + \frac{y_1 y_2}{(y_1+1)^3} \\
&= \frac{y_2}{(y_1+1)^2}.
\end{aligned}$$

Now,

$$\begin{aligned}
A &= \{(x_1,x_2) : f_{X_1,X_2}(x_1,x_2) > 0\} \\
&= \{(x_1,x_2) : x_1 > 0, x_2 > 0\}.
\end{aligned}$$

Therefore, $T(A) \subseteq \{(y_1, y_2) : y_1 > 0, y_2 > 0\}$. Since $x_1 > 0$ and $x_2 > 0$, $y_1 = \frac{x_1}{x_2} > 0$. Furthermore, since $x_1 = \frac{y_1 y_2}{y_1+1} > 0$, then $y_1 y_2 > 0$ implies $y_2 > 0$. Therefore,

$$T(A) = \{(y_1, y_2) : y_1 > 0, y_2 > 0\}.$$

Consequently, the joint p.d.f. of Y_1 and Y_2, $f = f_{Y_1,Y_2}(y_1, y_2)$ is given by

$$\begin{aligned}
f &= f_{X_1,X_2}(H_1(y_1, y_2), H_2(y_1, y_2))|J(y_1, y_2)| \\
&= f_{X_1,X_2}\left(\frac{y_1 y_2}{1+y_1}, \frac{y_2}{1+y_1}\right)\left|\frac{y_2}{(1+y_1)^2}\right| \\
&= \lambda^2 e^{-\lambda\left(\frac{y_1 y_2}{(1+y_1)} + \frac{y_2}{(1+y_1)}\right)} \frac{y_2}{(1+y_1)^2} \\
&= \lambda^2 e^{-\lambda y_2} \frac{y_2}{(1+y_1)^2}
\end{aligned}$$

if $y_1, y_2 > 0$ and 0 otherwise.

(b) The p.d.f. of Y_1 is the marginal p.d.f. of Y_1, therefore,

$$\begin{aligned}
f_{Y_1}(y_1) &= \int_0^\infty \lambda^2 e^{-\lambda y_2} \frac{y_2}{(1+y_1)^2} dy_2 \\
&= \frac{1}{(1+y_1)^2} \int_0^\infty \lambda e^{-\lambda y_2} dy_2 \\
&= \frac{1}{(1+y_1)^2}
\end{aligned}$$

if $y_1 > 0$. So,

$$f_{Y_1}(y_1) = \begin{cases} \frac{1}{(1+y_1)^2} & \text{if } y_1 > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Note that one can extend the method of transformations to the case of n random variables.

P7 Multivariate Normal distribution

Definition A random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$ is said to have an n dimensional normal distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ if the joint p.d.f. of $\boldsymbol{X}$ is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-\frac{n}{2}} \left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\},$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T$ and $\boldsymbol{\Sigma} = (\sigma_{ij})$ is an $n \times n$ real, symmetric, positive definite matrix with all positive eigenvalues. It is denoted by

$$\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Properties of the multivariate Normal distribution

1. If $\boldsymbol{D}$ is a $p \times n$ matrix and $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{X} \sim N_p(\boldsymbol{D}\boldsymbol{\mu}, \boldsymbol{D}\Sigma\boldsymbol{D}^T)$.

2. The marginal distribution of each component X_i is normal with $E[X_i] = \mu_i$ and $\text{Var}(X_i) = \sigma_{ii}$. Note that this is a direct consequence of property (1) with

$$\boldsymbol{D} = (0, \ldots, 0, 1, 0, \ldots, 0),$$

i.e. the ith component equal to 1.

3. The components $X_1, X_2, \ldots, X_n$ of a multivariate normal random vector are independent of each other if and only if $X_1, X_2, \ldots, X_n$ are uncorrelated, i.e. $\sigma_{ij} = \text{Cov}(X_i, X_j) = 0$ for all $i \neq j$.

4. Conditional distributions derived from joint normal distributions are normal.

Bivariate Normal distribution This is a special case with $n = 2$.

Definition The random variables X_1 and X_2 are said to have a *bivariate Normal distribution* with mean $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and variance-covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ if their joint p.d.f. is given by

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right.\right. \\ &\quad \left.\left. -2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}. \end{aligned}$$

Notes

1. $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ and Cov$(X_1, X_2) = \rho\sigma_1\sigma_2$.

2. $(X_1|X_2 = x_2) \sim N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2)\right)$.

3. $(X_2|X_1 = x_1) \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1-\rho^2)\right)$.

Example P7.1 Suppose $\boldsymbol{X} = (X_1, X_2, X_3)^T \sim N_3(\mathbf{0}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

(a) Find the distribution of $Z = X_1 + X_2$. (b) Determine the constant c such that $Z_1 = 2X_1 + cX_2$ and $Z_2 = 2X_1 + cX_3$ are independent.

(a) Let $Z = \boldsymbol{D}\boldsymbol{X} = X_1 + X_2$. Then $\boldsymbol{D} = (1\ 1\ 0)$. By property

(1), $Z \sim N(\boldsymbol{D0}, \boldsymbol{D\Sigma D}^T)$, where $\boldsymbol{D0} = 0$ and

$$\begin{aligned}
\boldsymbol{D\Sigma D}^T &= (1\ 1\ 0) \begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \\
&= (3\ 5\ 0) \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 8.
\end{aligned}$$

Therefore, $Z \sim N(0, 8)$.

(b) Let $\boldsymbol{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \boldsymbol{DX}$. Choose

$$\boldsymbol{D} = \begin{bmatrix} 2 & c & 0 \\ 2 & 0 & c \end{bmatrix}.$$

By property (1), $\boldsymbol{Z} \sim N_2(\boldsymbol{D0}, \boldsymbol{D\Sigma D}^T)$, where $\boldsymbol{D0} = \boldsymbol{0}$ and

$$\begin{aligned}
\boldsymbol{D\Sigma D}^T &= \begin{bmatrix} 2 & c & 0 \\ 2 & 0 & c \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ c & 0 \\ 0 & c \end{bmatrix} \\
&= \begin{bmatrix} 4+c & 2+4c & 0 \\ 4 & 2 & 5c \end{bmatrix} \begin{bmatrix} 2 & 2 \\ c & 0 \\ 0 & c \end{bmatrix} \\
&= \begin{bmatrix} 8+4c+4c^2 & 8+2c \\ 8+2c & 8+5c^2 \end{bmatrix}
\end{aligned}$$

For Z_1 to be independent of Z_2, $\text{Cov}(Z_1, Z_2) = 8 + 2c = 0$. Therefore, $c = -4$.

P8 Moment generating functions

Recall, for any random variable X, $E\left[X^k\right]$ is called the kth moment of X, if $E\left[X^k\right]$ exists (i.e. if $E\left[|X^k|\right] < \infty$). For example,

$E[X] = \mu$ is the first moment of X.

$E\left[X^2\right]$ is the second moment of X.

Definition The kth *central moment* of X is $E\left[(X-\mu)^k\right]$, if $E\left[|(X-\mu)^k|\right] < \infty$.

For example,

$$E\left[(X-\mu)^2\right] = \mathsf{Var}(X)$$

is the second central moment of X;

$$S = \frac{E\left[(X-\mu)^3\right]}{\sigma^3}$$

is a measure of skewness in the distribution;

$$K = \frac{E\left[(X-\mu)^4\right]}{\sigma^4}$$

is a measure of kurtosis of the distribution.

Definition For a random variable X, if $E\left[e^{tX}\right]$ exists for all $t \in (-h, h)$, $h > 0$, then

$$M_X(t) = E\left[e^{tX}\right]$$

is called the *moment generating function* (m.g.f.) of X.

From calculus, recall that the power series expansion of e^X is given by

$$e^X = \sum_{k=0}^{\infty} \frac{X^k}{k!}$$

if $-\infty < X < \infty$. Then for any real t,

$$e^{tX} = \sum_{k=0}^{\infty} \frac{(tX)^k}{k!}$$

if $-\infty < X < \infty$.

Suppose that $E\left[e^{tX}\right]$ exists for all $t \in (-h, h)$, $h > 0$, then $E\left[e^{tX}\right] < \infty$ for all $t \in (-h, h)$, $h > 0$. Therefore,

$$\begin{aligned} E\left[e^{tX}\right] &= E\left[\sum_{k=0}^{\infty} \frac{t^k X^k}{k!}\right] = \sum_{k=0}^{\infty} E\left[\frac{t^k X^k}{k!}\right] \\ &= \sum_{k=0}^{\infty} \frac{E\left[X^k\right] t^k}{k!} \end{aligned}$$

for all $t \in (-h, h)$.

Hence,

$$\begin{aligned} M_X(t) &= E\left[e^{tX}\right] \\ &= 1 + E\left[X\right] t + \frac{E\left[X^2\right] t^2}{2!} + \frac{E\left[X^3\right] t^3}{3!} + \cdots \end{aligned}$$

if $t \in (-h, h)$.

Note that for the m.g.f. of X to exist, all moments of X must exist. Hence not all random variables will have a m.g.f.

Example P8.1 Find the m.g.f. of $X \sim$ Poisson(λ).

If $X \sim$ Poisson(λ), then its p.m.f. is given by

$$p_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & \text{if } x = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$\begin{aligned} M_X(t) &= E\left[e^{tX}\right] = \sum_x e^{tx} p_X(x) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} = e^{-\lambda} e^{e^t\lambda} = e^{\lambda(e^t - 1)} \end{aligned}$$

for all real t.

Example P8.2 Find the m.g.f. of $X \sim U(0,1)$.

The m.g.f. is given by

$$\begin{aligned} M_X(t) &= E\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx = \int_0^1 e^{tx}\, 1\, dx = \frac{1}{t} e^{tx}\Big|_0^1 \\ &= \frac{1}{t}e^t - \frac{1}{t}e^0 = \frac{1}{t}(e^t - 1) \end{aligned}$$

if $t \neq 0$. If $t = 0$,

$$M_X(t) = M_X(0) = E\left[e^{0X}\right] = E[1] = 1.$$

Therefore,

$$M_X(t) = \begin{cases} \frac{1}{t}(e^t - 1) & \text{if } t \neq 0 \\ 1 & \text{if } t = 0. \end{cases}$$

Example P8.3 Find the m.g.f. of $X \sim N(\mu, \sigma^2)$.

$$\begin{aligned} M_X(t) &= E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left\{x^2 - 2x(\mu + t\sigma^2) + \mu^2\right\}} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left\{(x-(\mu+t\sigma^2))^2 - \sigma^2(2t\mu + t^2\sigma^2)\right\}} dx \\ &= e^{t\mu + \frac{1}{2}t^2\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\left\{x - (\mu + t\sigma^2)\right\}^2} dx \\ &= e^{t\mu + \frac{1}{2}t^2\sigma^2}. \end{aligned}$$

Properties of the m.g.f.

1. For any random variable X,

$$M_X(0) = E\left[e^{0X}\right] = E[1] = 1.$$

2. Recall,

$$M_X(t) = 1 + E[X]t + \frac{E\left[X^2\right]t^2}{2!} + \frac{E\left[X^3\right]t^3}{3!} + \cdots$$

if $t \in (-h, h)$. Therefore,

$$M_X'(t) = E[X] + 2\frac{E\left[X^2\right]t}{2!} + 3\frac{E\left[X^3\right]t^2}{3!} + \cdots$$

if $t \in (-h, h)$. Hence,

$$M_X'(0) = E[X].$$

Furthermore,

$$M_X''(t) = E\left[X^2\right] + (3)(2)\frac{E\left[X^3\right]t}{3!} + \cdots$$

if $t \in (-h, h)$. Hence,

$$M_X''(0) = E\left[X^2\right].$$

In general, $M_X^{(n)}(t) = E\left[X^n e^{tX}\right]$ and

$$M_X^{(n)}(0) = E[X^n].$$

Example P8.4 The m.g.f. of a Poisson random variable is given by $M_X(t) = e^{\lambda(e^t-1)}$ for all real t. Find the mean and variance of X.

Since $M_X'(t) = \lambda e^t e^{\lambda(e^t-1)} = \lambda e^{t+\lambda(e^t-1)}$,

$$E[X] = M_X'(0) = \lambda e^0 e^{\lambda(e^0-1)} = \lambda.$$

Also, since $M_X''(t) = \lambda(1+\lambda e^t)e^{t+\lambda(e^t-1)}$,

$$\begin{aligned} E\left[X^2\right] &= M_X''(0) = \lambda(1+\lambda e^0)e^{0+\lambda(e^0-1)} \\ &= \lambda(1+\lambda) = \lambda + \lambda^2. \end{aligned}$$

Hence,

$$\text{Var}(X) = E\left[X^2\right] - (E[X])^2 = \lambda + \lambda^2 - \lambda^2 = \lambda.$$

Theorem There is a one-to-one correspondence between m.g.f.'s and c.d.f.'s. In other words, the moment generating function uniquely determines the distribution function.

Theorem If $X_1, X_2, \dots, X_n$ are independent random variables, then the moment generating function of $X_1 + \cdots + X_n$ is given by

$$M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n M_{X_i}(t).$$

Proof Let $X_1, X_2, \dots, X_n$ be independent random variables with m.g.f.'s $M_{X_i}(t)$, then

$$\begin{aligned} M_{\sum_{i=1}^n X_i}(t) &= E\left[e^{(\sum_{i=1}^n X_i)t}\right] = E\left[\prod_{i=1}^n e^{X_i t}\right] \\ &= \prod_{i=1}^n E\left[e^{X_i t}\right] \end{aligned}$$

since $X_1, X_2, \dots, X_n$ are independent.

Therefore, $M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n M_{X_i}(t)$.

Example P8.5 The m.g.f. of $X \sim N(\mu, \sigma^2)$ is given by $M_X(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}$ for all real t. Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Find the m.g.f. and distribution of $X_1 + \cdots + X_n$.

The m.g.f. of $X_1 + \cdots + X_n$ is

$$\begin{aligned} M_{\sum_{i=1}^n X_i}(t) &= \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n e^{t\mu + \frac{1}{2}t^2\sigma^2} \\ &= e^{tn\mu + \frac{1}{2}nt^2\sigma^2}. \end{aligned}$$

From the theorem above, we know that m.g.f.'s are unique. Hence,

$$M_{\sum_{i=1}^n X_i}(t) = e^{t(n\mu) + \frac{1}{2}t^2(n\sigma^2)}$$

is the m.g.f. of a normal distribution with mean $n\mu$ and variance $n\sigma^2$. Therefore if $X_1, X_2, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$ then $\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$.

P9 Markov Chains

Consider the following example, known as Gambler's Ruin. Suppose a gambler has \$1. Each time he plays he bets \$1. If he wins the game, his bet is returned and he wins an additional \$1. If he loses the game, he loses the \$1 he bet. The probability of him winning any game is 0.6, the probability of him losing any game is 0.4. The gambler continues to play until either he is ruined (i.e. he has no money left) or he has \$4.

Let X_n be the amount of money the gambler has after n games.

Definition The set of all possible outcomes of X_n is called the *state space* S.

In the Gambler's Ruin example, $S = \{0, 1, 2, 3, 4\}$.

Definition A set of random variables $\{X_n : n = 0, 1, 2, \ldots\}$ with a finite or countably infinite state space S is said to be a *Markov Chain*, if for all $i_0, \ldots, i_{n-1}, i, j \in S$ and $n = 0, 1, 2, \ldots,$

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i).$$

Markovian Property Given the state of the Markov chain at present (X_n), its future state (X_{n+1}) is independent of the past states $(X_{n-1}, \ldots, X_1, X_0)$.

Definition A Markov chain with state space S is said to have *stationary transition probabilities*, if for all $i, j \in S$, the probability of transition from state i to state j, in one step, does not depend on the time that the transition will occur. (i.e. $P(X_{n+1} = j | X_n = i)$ does not depend on n.)

Notation If $\{X_n : n = 0, 1, \dots\}$ is a Markov chain with stationary transition probabilities, denote

$$p_{ij} = P(X_{n+1} = j | X_n = i) \text{ for } i, j \in S.$$

Definition If we let

$$P = (p_{ij}) = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

then P is called the *transition probability matrix*.

Note that for each row i, $\sum_{j=0}^{\infty} p_{ij} = 1$. The sum of the columns is not necessarily 1.

Example P9.1 At an intersection, a working traffic light will be out of order the next day with probability 0.07, and an out-of-order traffic light will be working the next day with probability 0.88. Let $X_n = 1$ if on day n the traffic light is working; $X_n = 0$ if on day n the traffic light is out-of-order.
(a) What is the state space of the Markov Chain $\{X_n : n = 0, 1, \dots\}$?
(b) Compute the transition probability matrix P.

(a) The state space is $S = \{0, 1\}$.
(b) The transition probability matrix is

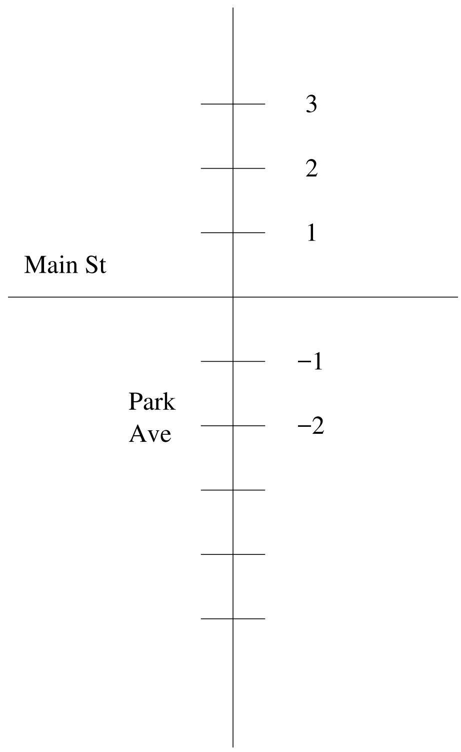
$$P = \begin{pmatrix} 0.12 & 0.88 \\ 0.07 & 0.93 \end{pmatrix}.$$

Example P9.2 Consider again the Gambler's Ruin Example, with state space $S = \{0, 1, 2, 3, 4\}$. Compute the transition probability matrix.

The transition probability matrix is

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.4 & 0 & 0.6 & 0 & 0 \\ 0 & 0.4 & 0 & 0.6 & 0 \\ 0 & 0 & 0.4 & 0 & 0.6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Example P9.3 (One-dimensional random walk) In a large town, Park Avenue is a long north-south avenue with many intersections. A drunken man is wandering along the avenue and does not really know which way he is going. Suppose that at the end of each block, he decides to walk either north with probability p or south with probability $1 - p$. Let X_n denote the intersection where the drunk finds himself after n steps. (Denote the central intersection in the city of Park Avenue with Main Street as the origin O.)



(a) What is the state space of the Markov Chain?
(b) What are the transition probabilities at state i?

(a) The state space is $S = \{\dots, -2, -1, 0, 1, 2, \dots\}$.
(b) The transition probabilities at state i are

$$p_{ij} = \begin{cases} p & \text{if } j = i+1 \\ 1-p & \text{if } j = i-1 \\ 0 & \text{otherwise.} \end{cases}$$

Notation Let $\{i \to j\}$ denote moving from state i to state j, then $\{i \to j \to k\}$ denotes moving first from state i to state j then to state k. Therefore, if $P(\{i \to j\}) = p_{ij}$, then

$$\begin{aligned} P(\{i \to j \to k\}) &= P(\{i \to j\})P(\{j \to k\}) \\ &= p_{ij}p_{jk}. \end{aligned}$$

Let $p_{ij}^{(n)}$ be the probability of moving from state i to j in n steps. Then

$$p_{ij}^{(n)} = P(X_{n+m} = j | X_m = i), n, m \geq 0.$$

Note that

$$\begin{aligned} P_{ij}^{(0)} &= P(X_m = j | X_m = i) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \\ P_{ij}^{(1)} &= P(X_{m+1} = j | X_m = i) = p_{ij}. \end{aligned}$$

Definition The n step *transition probability matrix* is

$$P^{(n)} = \begin{pmatrix} p_{00}^{(n)} & p_{01}^{(n)} & p_{02}^{(n)} & \cdots \\ p_{10}^{(n)} & p_{11}^{(n)} & p_{12}^{(n)} & \cdots \\ p_{20}^{(n)} & p_{21}^{(n)} & p_{22}^{(n)} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Note that in general, $P_{ij}^{(n)} \neq (p_{ij})^n$.

The Chapman-Kolmogorov Equations state that

$$p_{ij}^{(m+n)} = \sum_{k=0}^{\infty} p_{ik}^{(m)} p_{kj}^{(n)}$$

or $P^{(m+n)} = P^{(m)} P^{(n)}$.

Recall Example P9.1, the traffic light example where $S = \{0, 1\}$, state 0 is the traffic lights not working, state 1 is the traffic light working and $P = \begin{pmatrix} 0.12 & 0.88 \\ 0.07 & 0.93 \end{pmatrix}$. Then

$$P^2 = P \times P = \begin{pmatrix} 0.0760 & 0.9240 \\ 0.0735 & 0.9265 \end{pmatrix}.$$

Therefore, the probability of an out-of-order traffic light not working in 2 days is $p_{00}^{(2)} = 0.076$.

Recall Example P9.2, the Gambler's Ruin example where $S = \{0, 1, 2, 3, 4\}$ and

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.4 & 0 & 0.6 & 0 & 0 \\ 0 & 0.4 & 0 & 0.6 & 0 \\ 0 & 0 & 0.4 & 0 & 0.6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

then

$$P^{10} = \begin{pmatrix} 1.000 & 0 & 0 & 0 & 0 \\ 0.575 & 0.013 & 0 & 0.019 & 0.393 \\ 0.300 & 0 & 0.025 & 0 & 0.675 \\ 0.117 & 0.008 & 0 & 0.013 & 0.862 \\ 0 & 0 & 0 & 0 & 1.000 \end{pmatrix},$$

where $p_{ij}^{(10)}$ is the probability of having j dollars after 10 plays if one began with i dollars (at any stage of the game).

Theorem Let $\{X_n : n = 0, 1, 2, \ldots\}$ be a Markov chain with transition probability matrix $P = (p_{ij})$. For $i \geq 0$, let $p(i) = P(X_0 = i)$ be the probability mass function of X_0. Then the probability mass function of X_n is given by

$$P(X_n = j) = \sum_{i=0}^{\infty} p(i) p_{ij}^{(n)} \quad j = 0, 1, 2, \ldots$$

Proof Since states $\{0, 1, 2, \ldots\}$ are disjoint, we can use the law of total probability so that

$$\begin{aligned} P(X_n = j) &= \sum_{i=0}^{\infty} P(X_n = j | X_0 = i) P(X_0 = i) \\ &= \sum_{i=0}^{\infty} p(i) p_{ij}^{(n)}. \end{aligned}$$

Note that a Markov chain is uniquely determined by its transition matrix.

Example P9.4 Recall Example P9.1, the traffic light example. Suppose the probability that a traffic light functions properly on any day is 0.95. What is the probability that the light is out-of-order in 2 days?

Given $P(X_0 = 1) = 0.95$, then $P(X_0 = 0) = 0.05$. We want to find $P(X_2 = 0)$.

$$\begin{aligned} P(X_2 = 0) &= P(X_2 = 0 | X_0 = 0) P(X_0 = 0) \\ &\quad + P(X_2 = 0 | X_0 = 1) P(X_0 = 1) \\ &= p_{00}^{(2)}(0.05) + p_{10}^{(2)}(0.95) \\ &= (0.076)(0.05) + (0.0735)(0.95) \\ &= 0.0038 + 0.069825 = 0.073625. \end{aligned}$$

Example P9.5 Recall Example P9.2, the Gambler's Ruin example. Suppose that it is equally likely that a man begins the game with \$1, \$2 or \$3 what is the probability that he has won the game (has \$4) by the 10th attempt?

Given $P(X_0 = 1) = P(X_0 = 2) = P(X_0 = 3) = \frac{1}{3}$, we want to find $P(X_{10} = 4)$.

$$\begin{aligned} P(X_{10} = 4) &= P(X_{10} = 4|X_0 = 1)P(X_0 = 1) \\ &\quad +P(X_{10} = 4|X_0 = 2)P(X_0 = 2) \\ &\quad +P(X_{10} = 4|X_0 = 3)P(X_0 = 3) \\ &= (0.393 + 0.675 + 0.862)\frac{1}{3} \\ &= 0.643. \end{aligned}$$

Definition A state j of a Markov chain is said to *absorbing* if once the process enters state j it remains there forever. (i.e. A state j of a Markov chain is absorbing if $p_{jj} = 1$).

For example, in the Gambler's Ruin example states 0 and 4 are absorbing since once you've lost you're out and once you've won then you've finished the game. (Note $p_{00} = 1$ and $p_{44} = 1$.)

In the traffic lights example, no states are absorbing. There is always a non-zero probability of moving to the other state.

Some related questions are:
What is the probability of a Markov chain eventually being absorbed into state j?
What is the expected number of steps until one arrives in absorption state j?

Notation Let $\{X_n : n = 0, 1, 2, \dots\}$ be a Markov chain, then we denote the probability that, starting from state i, the process will return to state i for the first time after exactly n steps as $f_{ii}^{(n)}$. Let

$$f_i = \sum_{n=1}^{\infty} f_{ii}^{(n)}$$

be the probability that starting from state i the process will return to state i after a finite number of steps.

Definition A state i is said to be *recurrent* if $f_i = 1$. If $f_i < 1$, then state i is called *transient* (i.e. there is a positive probability that the process will not return to i).

For the traffic light example we observe the following:

$$\begin{aligned}
P &= \begin{pmatrix} 0.12 & 0.88 \\ 0.07 & 0.93 \end{pmatrix} \\
P^2 &= \begin{pmatrix} 0.0760 & 0.9240 \\ 0.0735 & 0.9265 \end{pmatrix} \\
P^3 &= \begin{pmatrix} 0.073800 & 0.926200 \\ 0.073675 & 0.926325 \end{pmatrix} \\
P^4 &= \begin{pmatrix} 0.07369000 & 0.92631000 \\ 0.07368375 & 0.92631625 \end{pmatrix} \\
P^5 &= \begin{pmatrix} 0.0736845000 & 0.9263155000 \\ 0.0736841875 & 0.9263158125 \end{pmatrix}
\end{aligned}$$

It seems that $P^n \to \begin{pmatrix} 0.07368 & 0.92632 \\ 0.07368 & 0.92632 \end{pmatrix}$ as $n \to \infty$. Therefore as n increases, it doesn't matter what our initial state was, the probability of transitioning to the other states becomes the same.

Definition If a Markov chain, for each $j \geq 0$, $\lim_{n\to\infty} p_{ij}^{(n)}$ exists and is independent of i, we say that the Markov chain is in *equilibrium* or *steady state*. The limits

$$\pi_j = \lim_{n \to \infty} p_{ij}^{(n)},$$

$j \geq 0$, are called the *stationary probabilities* of the Markov chain.

If π is the row vector of stationary probabilities, then we can find π by solving

$$\pi = \pi P.$$

For the traffic light example, set $\pi = (\pi_1 \pi_2)$. Then

$$\begin{aligned}
\pi &= \pi P \\
\Leftrightarrow (\pi_1 \ \pi_2) &= (\pi_1 \ \pi_2) \begin{pmatrix} 0.12 & 0.88 \\ 0.07 & 0.93 \end{pmatrix} \\
\Rightarrow \pi_1 &= 0.12\pi_1 + 0.07\pi_2 \\
\text{and } \ \pi_2 &= 0.88\pi_1 + 0.93\pi_2.
\end{aligned}$$

We then use the fact that $\pi_1 = 1 - \pi_2$ to obtain

$$(\pi_1 \ \pi_2) = \left(\frac{0.07}{0.95} \ \ \frac{0.88}{0.95} \right) \approx (0.07368 \ \ 0.92632).$$