1 Random Variables and Probability Distributions

1.1 Random Variables

1.1.1 Discrete random variables

A random variable X is called *discrete* if the number of values that X takes is finite or countably infinite. To be more precise, for a discrete random variable X there exist a (finite or infinite) sequence of real numbers $a_1, a_2, \dots$ and corresponding nonnegative numbers $p_1, p_2, \dots$ such that

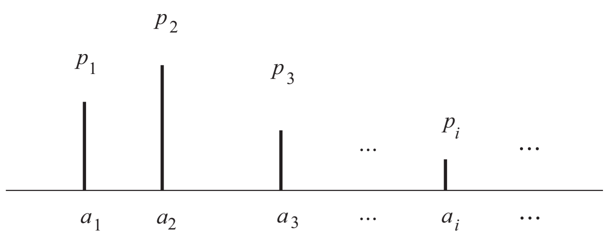
$$P(X = a_i) = p_i, \qquad p_i \geq 0, \qquad \sum p_i = 1.$$

In this case

$$\mu_X(dx) = \sum_i p_i \delta_{a_i}(dx) = \sum_i p_i \delta(x - a_i) dx$$

is called the *(probability) distribution* of X. Obviously,

$$P(a \leq X \leq b) = \sum_{i: a \leq a_i \leq b} p_i$$



Example 1.1.1 (coin toss) We set

$$X = \begin{cases} 1, & \text{heads}, \\ 0, & \text{tails}. \end{cases}$$

Then

$$P(X = 1) = p, \qquad P(X = 0) = q = 1 - p.$$

For a fair coin we set $p = 1/2$.

Example 1.1.2 (waiting time) Flip a fair coin repeatedly until we get the heads. Let T be the number of coin tosses to get the first heads. (If the heads occurs at the first trial, we have $T = 1$; If the tails occurs at the first trial and the heads at the second trial, we have $T = 2$, and so on.)

$$P(T = k) = (1 - p)^{k-1} p, \qquad k = 1, 2, \dots.$$

1.1.2 Continuous random variables

A random variable X is called *continuous* if $P(X = a) = 0$ for all $a \in \mathbf{R}$. We understand intuitively that X varies continuously.

If there exists a function $f(x)$ such that

$$P(a \leq X \leq b) = \int_a^b f(x) dx, \qquad a < b,$$

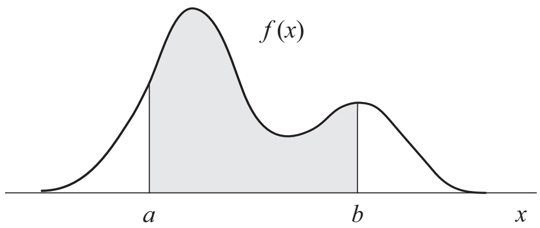
we say that X admits a *probability density function*. Note that

$$\int_{-\infty}^{+\infty} f(x) dx = 1, \qquad f(x) \geq 0.$$

In this case,

$$\mu_X(dx) = f(x)dx$$

is called the *(probability) distribution* of X.



It is useful to consider the *distribution function*:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt, \qquad x \in \mathbf{R}.$$

Then we have

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

Remark 1.1.3 (1) A continuous random variable does not necessarily admit a probability density function. But many continuous random variables in practical applications admit probability density functions.

(2) There is a random variable which is neither discrete nor continuous. But most random variables in practical applications are either discrete or continuous.

Example 1.1.4 (random cut) Divide the interval $[0, L]$ ($L > 0$) into two segments.

(1) Let X be the coordinate of the cutting point (the length of the segment containing 0).

$$F_X(x) = \begin{cases} 0, & x < 0; \\ x/L, & 0 \leq x \leq L; \\ 1, & x > L. \end{cases}$$

(2) Let M be the length of the longer segment.

$$F_M(x) = \begin{cases} 0, & x < L/2; \\ (2x - L)/L, & L/2 \leq x \leq L; \\ 1, & x > L. \end{cases}$$

Example 1.1.5 Let A be a randomly chosen point from the disc with radius $R > 0$. Let X be the distance between the center O and A. We have

$$P(a \leq X \leq b) = \frac{\pi(b^2 - a^2)}{\pi R^2} = \frac{1}{R^2}\int_a^b 2x dx, \qquad 0 < a < b < R,$$

so the probability density function is given by

$$f(x) = \begin{cases} 0, & x \leq 0, \\ \dfrac{2x}{R^2}, & 0 \leq x \leq R, \\ 0, & x > R. \end{cases}$$

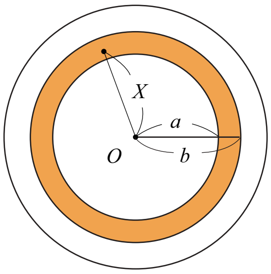


Figure 1.1: Random choice of a point

1.1.3 Mean and variance

Definition 1.1.6 The *mean* or *expectation value* of a random variable X is defined by

$$m = \mathbf{E}[X] = \int_{-\infty}^{+\infty} x\mu_X(dx)$$

- If X is discrete, we have
$$\mathbf{E}[X] = \sum_i a_i p_i\,.$$
- If X admits a probability density function $f(x)$, we have
$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx.$$

Remark 1.1.7 For a function $\varphi(x)$ we have

$$\mathbf{E}[\varphi(X)] = \int_{-\infty}^{+\infty} \varphi(x)\mu(dx).$$

For example,

$$\mathbf{E}[X^m] = \int_{-\infty}^{+\infty} x^m\mu(dx) \qquad (m\text{th moment}),$$
$$\mathbf{E}[e^{itX}] = \int_{-\infty}^{+\infty} e^{itx}\mu(dx) \qquad (\text{characteristic function}).$$

Definition 1.1.8 The *variance* of a random variable X is defined by

$$\sigma^2 = \mathbf{V}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2,$$

or equivalently,

$$\sigma^2 = \mathbf{V}[X] = \int_{-\infty}^{+\infty} (x - \mathbf{E}[X])^2\mu(dx) = \int_{-\infty}^{+\infty} x^2\mu(dx) - \left(\int_{-\infty}^{+\infty} x\mu(dx)\right)^2.$$

Exercise 1.1.9 (see Example 1.1.2) Calculate the mean and variance of the waiting time T.

Exercise 1.1.10 Let S be the length of the shorter segment obtained by randomly cutting the interval $[0, L]$. Calculate the mean and variance of S.

1.2 Discrete Distributions

1.2.1 Bernoulli distribution

For $0 \le p \le 1$ the distribution

$$(1-p)\delta_0 + p\delta_1$$

is called *Bernoulli distribution with success probability* p. This is the distribution of coin toss. The mean value and variance are given by

$$m = p, \qquad \sigma^2 = p(1-p)$$

Exercise 1.2.1 Let a, b be distinct real numbers. A general two-point distribution is defined by

$$p\delta_a + q\delta_b\,,$$

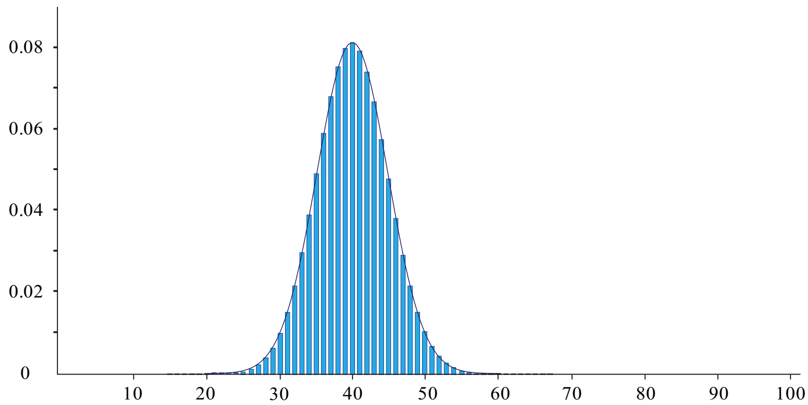
where $0 \le p \le 1$ and $p + q = 1$. Determine the two-point distribution having mean 0, variance 1.

1.2.2 Binomial distribution

For $0 \le p \le 1$ and $n \ge 1$ the distribution

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}\,\delta_k$$

is called the *binomial distribution* $B(n, p)$. The quantity $\binom{n}{k} p^k (1-p)^{n-k}$ is the probability that n coin tosses with probabilities p for heads and $q = 1 - p$ for tails result in k heads and $n - k$ tails.



$B(100, 0.4)$

Exercise 1.2.2 Verify that $m = np$ and $\sigma^2 = np(1-p)$ for $B(n, p)$.

1.2.3 Geometric distribution

For $0 \le p \le 1$ the distribution

$$\sum_{k=1}^{\infty} p(1-p)^{k-1}\delta_k$$

is called the *geometric distribution with success probability* p. This is the distribution of waiting time for the first heads (Example 1.1.2).

Exercise 1.2.3 Verify that $m = \dfrac{1}{p}$ and $\sigma^2 = \dfrac{1}{p^2}$

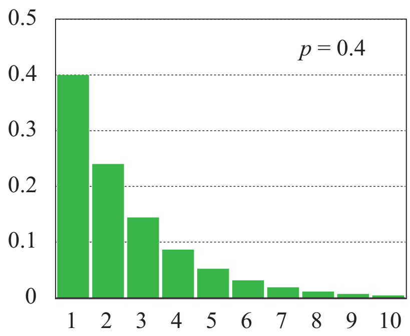


Figure 1.2: Geometric distribution with parameter $p = 0.4$

Remark 1.2.4 In some literatures, the geometric distribution with parameter p is defined by

$$\sum_{k=0}^{\infty} p(1-p)^k \delta_k$$

1.2.4 Poisson distribution

For $\lambda > 0$ the distribution

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k$$

is called the *Poisson distribution with parameter λ*. The mean and variance are given by

$$m = \lambda, \qquad \sigma^2 = \lambda.$$

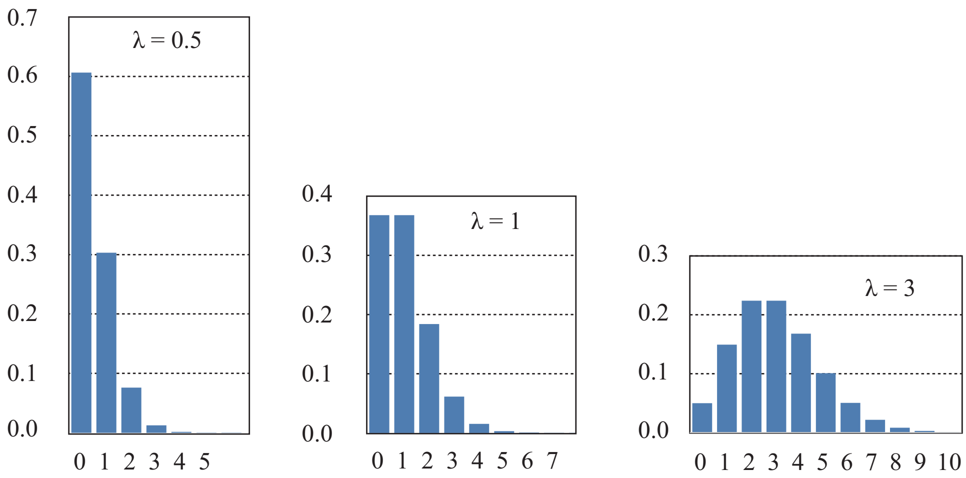


Figure 1.3: Poisson distribution $\lambda = 1/2, 1, 3$

Problem 1 The *probability generating function* of the Poisson distribution is defined by

$$G(z) = \sum_{k=0}^{\infty} p_k z^k, \qquad p_k = e^{-\lambda} \frac{\lambda^k}{k!}.$$

(1) Find a concise expression of $G(z)$.

(2) By using $G'(1)$ and $G''(1)$ find the mean value and the variance of the Poisson distribution with parameter λ.

(3) Show that

$$\sum_{k:\text{odd}} p_k < \sum_{k:\text{even}} p_k \,.$$

In other words, the probability of taking even values is greater than that of odd values.

(4) Discuss relevant topics.

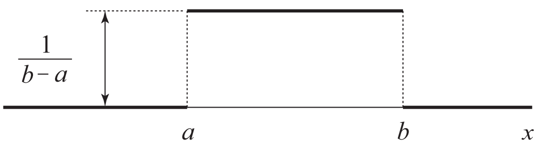
1.3 Continuous Distributions (Density Functions)

1.3.1 Uniform distribution

For a finite interval $[a, b]$,

$$f(x) = \begin{cases} \dfrac{1}{b-a}\,, & a \le x \le b, \\ 0, & \text{otherwise} \end{cases}$$

becomes a density function, which determines the *uniform distribution* on $[a, b]$.



The mean value and the variance are given by

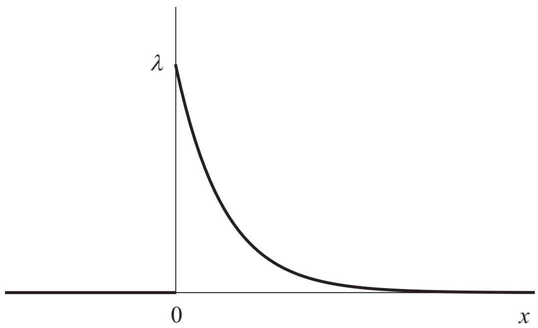
$$m = \int_a^b x \frac{dx}{b-a} = \frac{a+b}{2}\,, \qquad \sigma^2 = \int_a^b x^2 \frac{dx}{b-a} - m^2 = \frac{(b-a)^2}{12}\,.$$

1.3.2 Exponential distribution

The *exponential distribution* with parameter $\lambda > 0$ is defined by the density function

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

This is a model for waiting time (continuous time).



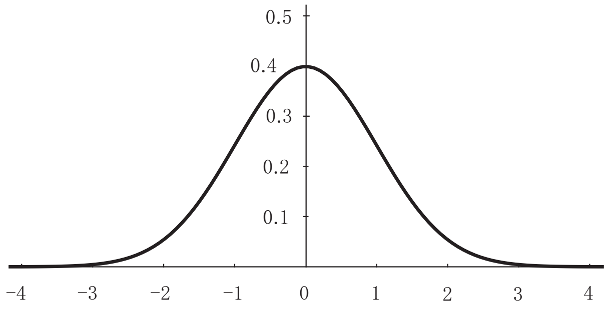
Exercise 1.3.1 Verify that $m = \dfrac{1}{\lambda}$ and $\sigma^2 = \dfrac{1}{\lambda^2}$.

1.3.3 Normal distribution

For $m \in \mathbb{R}$ and *sigma* > 0 we may check that

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}$$

becomes a density function. The distribution defined by the above density function is called the *normal distribution* or *Gaussian distribution* and denoted by $N(m, \sigma^2)$. In particular, $N(0, 1)$ is called the *standard normal distribution* or the *standard Gaussian distribution*.



Exercise 1.3.2 Differentiating both sides of the known formula:

$$\int_0^{+\infty} e^{-tx^2} dx = \frac{\sqrt{\pi}}{2\sqrt{t}}, \qquad t > 0,$$

find the values

$$\int_0^{+\infty} x^{2n} e^{-x^2} dx, \qquad n = 0, 1, 2, \dots.$$

Exercise 1.3.3 Prove that the above $f(x)$ is a probability density function. Then prove by integration that the mean is m and the variance is σ^2:

$$m = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} dx,$$

$$\sigma^2 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-m)^2 \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} dx$$

Problem 2 Choose randomly a point A from the disc with radius one and let X be the radius of the inscribed circle with center A.

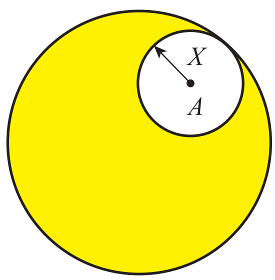
(1) For $x \geq 0$ find the probability $P(X \leq x)$.

(2) Find the probability density function $f_X(x)$ of X. (Note that x varies over all real numbers.)

(3) Calculate the mean and variance of X.

(4) Calculate the mean and variance of the area of inscribed circle $S = \pi X^2$.

(5) Discuss similar questions for a ball.



2 Independence and Dependence

2.1 Independent Events

Definition 2.1.1 (Pairwise independence) A (finite or infinite) sequence of events $A_1, A_2, \dots$ is called *pairwise independent* if any pair of events A_{i_1}, A_{i_2} $(i_1 \neq i_2)$ verifies

$$P(A_{i_1} \cap A_{i_2}) = P(A_{i_1})P(A_{i_2}).$$

Definition 2.1.2 (Independence) A (finite or infinite) sequence of events $A_1, A_2, \dots$ is called *independent* if any choice of finitely many events $A_{i_1}, \dots, A_{i_n}$ $(i_1 < i_2 < \cdots < i_n)$ satisfies

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_n}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_n}).$$

Example 2.1.3 Consider the trial to randomly draw a card from a deck of 52 cards. Let A be the event that the result is an ace and B the event that the result is spades. Then A, B are independent.

Problem 3 An urn contains four balls with numbers 112, 121, 211, 222. We draw a ball at random and let X_1 be the first digit, X_2 the second digit, and X_3 the last digit. For $i = 1, 2, 3$ we define an event A_i by $A_i = \{X_i = 1\}$. Show that $\{A_1, A_2, A_3\}$ is pairwise independent but is not independent.

Remark 2.1.4 It is allowed to consider whether the sequence of events $\{A, A\}$ is independent or not. If they are independent, by definition we have

$$P(A \cap A) = P(A)P(A).$$

Then $P(A) = 0$ or $P(A) = 1$. Notice that $P(A) = 0$ does not imply $A = \emptyset$ (empty event). Similarly, $P(A) = 1$ does not imply $A = \Omega$ (whole event).

Exercise 2.1.5 For A we write $A^\#$ for itself A or its complementary event A^c. Prove the following assertions.

(1) If A and B are independent, so are $A^\#$ and $B^\#$.

(2) If $A_1, A_2, \dots$ are independent, so are $A_1^\#, A_2^\#, \dots$.

Definition 2.1.6 (Conditional probability) For two events A, B the *conditional probability of A relative to B* (or *on the hypothesis B*, or *for given B*) is defined by

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

whenever $P(B) > 0$.

Theorem 2.1.7 Let A, B be events with $P(A) > 0$ and $P(B) > 0$. Then, the following assertions are equivalent:

(i) A, B are independent;

(ii) $P(A|B) = P(A)$;

(iii) $P(B|A) = P(B)$;

2.2 Independent Random Variables

Definition 2.2.1 A (finite or infinite) sequence of random variables $X_1, X_2, \dots$ is *independent* (resp. *pairwise independent*) if so is the sequence of events $\{X_1 \le a_1\}, \{X_1 \le a_2\}, \dots$ for any $a_1, a_2, \dots \in \mathbf{R}$.

In other words, a (finite or infinite) sequence of random variables $X_1, X_2, \dots$ is independent if for any finite $X_{i_1}, \dots, X_{i_n}$ $(i_1 < i_2 < \dots < i_n)$ and constant numbers $a_1, \dots, a_n$

$$P(X_{i_1} \le a_1 \,, X_{i_2} \le a_2 \,, \dots, X_{i_n} \le a_n) = P(X_{i_1} \le a_1)P(X_{i_2} \le a_2) \cdots P(X_{i_n} \le a_n) \tag{2.1}$$

holds. Similar assertion holds for the pairwise independence. If random variables $X_1, X_2, \dots$ are discrete, (2.1) may be replaced with

$$P(X_{i_1} = a_1 \,, X_{i_2} = a_2 \,, \dots, X_{i_n} = a_n) = P(X_{i_1} = a_1)P(X_{i_2} = a_2) \cdots P(X_{i_n} = a_n).$$

Example 2.2.2 Choose at random a point from the rectangle $\Omega = \{(x, y)\,;\, a \le x \le b,\ c \le y \le d\}$. Let X denote the x-coordinates of the chosen point and Y the y-coordinates. Then X, Y are independent.

Example 2.2.3 (Bernoulli trials) This is a model of coin-toss and is the most fundamental stochastic process. A sequence of random variables (or a discrete-time stochastic process) $\{X_1, X_2, \dots, X_n, \dots\}$ is called the *Bernoulli trials* with success probability p $(0 \le p \le 1)$ if they are independent and have the same distribution as

$$P(X_n = 1) = p, \qquad P(X_n = 0) = q = 1 - p.$$

By definition we have

$$P(X_1 = \xi_1, X_2 = \xi_2, \dots, X_n = \xi_n) = \prod_{k=1}^{n} P(X_k = \xi_k) \qquad \text{for all } \xi_1, \xi_2, \dots, \xi_n \in \{0, 1\}.$$

In general, statistical quantity in the left-hand side is called the *finite dimensional distribution* of the stochastic process $\{X_n\}$. The total set of finite dimensional distributions characterizes a stochastic process.

2.3 Covariance and Correlation Coefficient

Recall that the mean of a random variable X is defined by

$$m_X = \mathbf{E}(X) = \int_{-\infty}^{+\infty} x \mu_X(dx).$$

Theorem 2.3.1 (Linearity) For two random variables X, Y and two constant numbers a, b it holds that

$$\mathbf{E}(aX + bY) = a\mathbf{E}(X) + b\mathbf{E}(Y).$$

Theorem 2.3.2 (Multiplicativity) If random variables $X_1, X_2, \dots, X_n$ are independent, we have

$$\mathbf{E}[X_1 X_2 \cdots X_n] = \mathbf{E}[X_1] \cdots \mathbf{E}[X_n]. \tag{2.2}$$

Proof We first prove the assertion for $X_k = 1_{A_k}$ (indicator random variable). By definition $X_1, \dots, X_n$ are independent if and only if so are $A_1, \dots, A_n$. Therefore,

$$\begin{aligned}\mathbf{E}[X_1 \cdots X_n] &= \mathbf{E}[1_{A_1 \cap \dots \cap A_n}] = P(A_1 \cap \dots \cap A_n) \\ &= P(A_1) \cdots P(A_n) = \mathbf{E}[X_1] \cdots \mathbf{E}[X_n].\end{aligned}$$

Thus (2.2) is verified. Then, by linearity the assertion is valid for X_k taking finitely many values (finite linear combination of indicator random variables). Finally, for general X_k, coming back to the definition of Lebesgue integration, we can prove the assertion by approximation argument. ∎

Remark 2.3.3 $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$ is not a sufficient condition for the random variables X and Y being independent. It is merely a necessary condition!

The *variance* of X is defined by

$$\sigma_X^2 = \mathbf{V}(X) = \mathbf{E}[(X - m_X)^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2.$$

By means of the distribution $\mu(dx)$ of X we may write

$$\mathbf{V}(X) = \int_{-\infty}^{+\infty} (x - m_X)^2 \mu(dx) = \int_{-\infty}^{+\infty} x^2 \mu(dx) - \left(\int_{-\infty}^{+\infty} x\mu(dx) \right)^2.$$

Definition 2.3.4 The *covariance* of two random variables X, Y is defined by

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

In particular, $\sigma_{XX} = \sigma_X^2$ becomes the variance of X. The *correlation coefficient* of two random variables X, Y is defined by

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y},$$

whenever $\sigma_X > 0$ and $\sigma_Y > 0$.

Definition 2.3.5 X, Y are called uncorrelated if $\sigma_{XY} = 0$. They are called positively (resp. negatively) correlated if $\sigma_{XY} > 0$ (resp. $\sigma_{XY} < 0$).

Theorem 2.3.6 If two random variables X, Y are independent, they are uncorrelated.

Remark 2.3.7 The converse of Theorem 2.3.6 is not true in general. Let X be a random variable satisfying

$$P(X = -1) = P(X = 1) = \frac{1}{4}, \quad P(X = 0) = \frac{1}{2}$$

and set $Y = X^2$. Then, X, Y are not independent, but $\sigma_{XY} = 0$. On the other hand, for random variables X, Y taking only two values, the converse of Theorem 2.3.6 is valid (see Problem 5).

Theorem 2.3.8 (Additivity of variance) Let $X_1, X_2, \dots, X_n$ be random variables, any pair of which is uncorrelated. Then

$$\mathbf{V}\left[\sum_{k=1}^n X_k\right] = \sum_{k=1}^n \mathbf{V}[X_k].$$

Theorem 2.3.9 $-1 \le \rho_{XY} \le 1$ for two random variables X, Y with $\sigma_X > 0$, $\sigma_Y > 0$.

Proof Note that $\mathbf{E}[\{t(X - m_X) + (Y - m_Y)\}^2] \ge 0$ for all $t \in \mathbf{R}$. ∎

Problem 4 Throw two dice and let L be the larger spot and S the smaller. (If double spots, set $L = S$.)

(1) Show the joint probability of (L, S) by a table.

(2) Calculate the correlation coefficient ρ_{LS} and explain the meaning of the signature of ρ_{LS}.

Problem 5 Let X and Y be random variables such that

$$P(X = a) = p_1, \quad P(X = b) = q_1 = 1 - p_1, \quad P(Y = c) = p_2, \quad P(Y = d) = q_2 = 1 - p_2,$$

where a, b, c, d are constant numbers and $0 < p_1 < 1$, $0 < p_2 < 1$. Show that X, Y are independent if $\sigma_{XY} = 0$. [Notice: In general, uncorrelated random variables are not necessarily independent. Hence, the situation in this problem falls into a very particular one.]

3 Markov Chains

3.1 Conditional Probability

For two events A, B we define

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{3.1}$$

whenever $P(B) > 0$. We call $P(A|B)$ the *conditional probability of A relative to B* It is interpreted as the probability of the event A assuming the event B occurs, see Section 2.1.

Formula (3.1) is often used in the following form:

$$P(A \cap B) = P(B)P(A|B) \tag{3.2}$$

This is the so-called theorem on compound probabilities, giving a ground to the usage of tree diagram in computation of probability. For example, for two events A, B see Fig. 3.1.

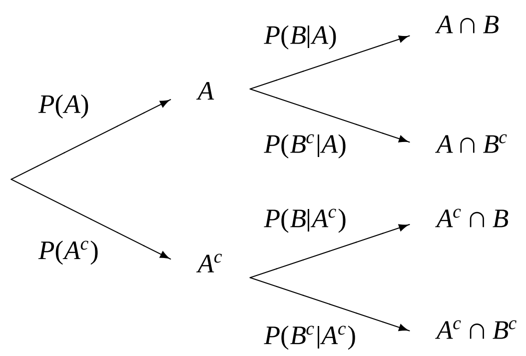


Figure 3.1: Tree diagram

Theorem 3.1.1 (Compound probabilities) For events $A_1, A_2, \dots, A_n$ we have

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}). \tag{3.3}$$

Proof Straightforward by induction on n. ∎

3.2 Markov Chains

Let S be a finite or countable set. Consider a discrete time stochastic process $\{X_n ; n = 0, 1, 2, \dots\}$ taking values in S. This S is called a *state space* and is not necessarily a subset of $\mathbb{R}$ in general. In the following we often meet the cases of $S = \{0, 1\}$, $S = \{1, 2, \dots, N\}$ and $S = \{0, 1, 2, \dots\}$.

Definition 3.2.1 Let $\{X_n ; n = 0, 1, 2, \dots\}$ be a discrete time stochastic process over S. It is called a *Markov process* over S if

$$P(X_n = b|X_{i_1} = a_1, X_{i_2} = a_2, \dots, X_{i_k} = a_k, X_i = a) = P(X_n = b|X_i = a)$$

holds for any $0 \le i_1 < i_2 < \cdots < i_k < i < n$ and $a_1, a_2, \dots, a_k, a, b \in S$.

If $\{X_1, X_2, \dots\}$ are independent random variables with values in S, obviously they form a Markov chain. Hence the Markov property is weaker than independence.

Example 3.2.2 Let $r \geq 1$ and $s \geq 1$ such that $r + s = N$. There are r black balls and s white balls in a box. We pick up balls in the box one by one and set $X_n = 1$ if a black ball is picked up at the nth trial and $X_n = 0$ if a white ball is picked up at the nth trial. Then $\{X_1, X_2, \dots, X_N\}$ is a stochastic process. We note that

$$P(X_n = 1 | X_1 = a_1,\ X_2 = a_2,\ \dots, X_{n-1} = a_{n-1}) = \frac{1}{N-(n-1)} \left[r - \sum_{k=1}^{n-1} a_k \right] \tag{3.4}$$

and

$$P(X_n = 1 | X_{n-1} = a_{n-1}) = \frac{r - a_{n-1}}{N-1}\,, \tag{3.5}$$

for $a_1, \dots, a_{n-1} \in \{0, 1\}$. Hence $\{X_n\}$ is not a Markov chain.

Problem 6 We keep the notations and assumptions in Example 3.2.2.

(1) Prove (3.4) and (3.5).

(2) Let Y_n be the number of black balls picked up during the first n trials, i.e.,

$$Y_n = \sum_{k=1}^{n} X_k\,.$$

Show that $\{Y_n\}$ is a Markov chain.

Definition 3.2.3 For a Markov chain $\{X_n\}$ over S,

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

is called the *transition probability* at time n from a state i to j. If this is independent of n, the Markov chain is called *time homogeneous*. In this case we write

$$p_{ij} = p(i, j) = P(X_{n+1} = j | X_n = i)$$

and simply call it the transition probability. Moreover, the matrix

$$P = [p_{ij}]$$

is called the *transition matrix*.

Obviously, we have for each $i \in S$,

$$\sum_{j \in S} p(i, j) = \sum_{j \in S} P(X_{n+1} = j | X_n = i) = 1.$$

Taking this into account, we give the following

Definition 3.2.4 A matrix $P = [p_{ij}]$ with index set S is called a *stochastic matrix* if

$$p_{ij} \geq 0 \quad \text{and} \quad \sum_{j \in S} p_{ij} = 1.$$

Theorem 3.2.5 The transition matrix of a Markov chain is a stochastic matrix. Conversely, given a stochastic matrix we can construct a Markov chain of which the transition matrix coincides with the given stochastic matrix.

It is convenient to use the *transition diagram* to illustrate a Markov chain. With each state we associate a point and we draw an arrow from i to j when $p(i, j) > 0$.

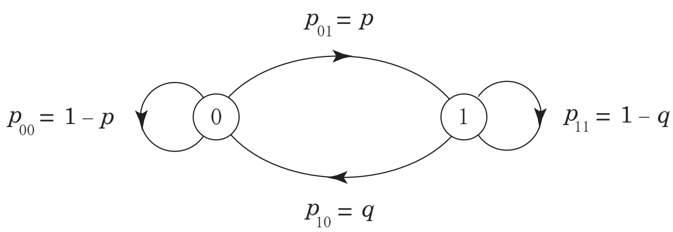
Example 3.2.6 (2-state Markov chain) A Markov chain over the state space $\{0, 1\}$ is determined by the transition probabilities:

$$p(0,1) = p, \quad p(0,0) = 1 - p, \quad p(1,0) = q, \quad p(1,1) = 1 - q.$$

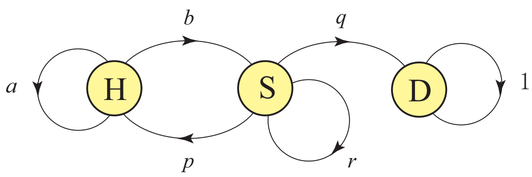
The transition matrix is defined by

$$\begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}.$$

The transition diagram is as follows:



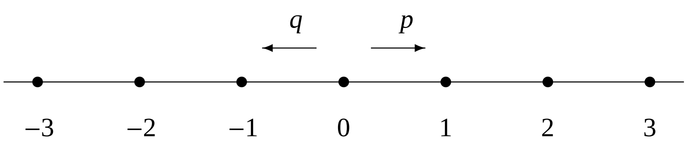
Example 3.2.7 (3-state Markov chain) An animal is healthy, sick or dead, and changes its state every day. Consider a Markov chain on $\{H, S, D\}$ described by the following transition diagram:



The transition matrix is defined by

$$\begin{bmatrix} a & b & 0 \\ p & r & q \\ 0 & 0 & 1 \end{bmatrix}, \qquad a + b = 1, \quad p + q + r = 1.$$

Example 3.2.8 (Random walk on $\mathbb{Z}^1$) The random walk on $\mathbb{Z}^1$ is illustrated as



The transition probabilities are given by

$$p(i,j) = \begin{cases} p, & \text{if } j = i + 1, \\ q = 1 - p, & \text{if } j = i - 1, \\ 0, & \text{otherwise.} \end{cases}$$

The transition matrix is a two-sided infinite matrix given by

$$\begin{bmatrix} \ddots & \ddots & \ddots & \ddots & & & & & \\ \ddots & q & 0 & p & 0 & & & & \\ & 0 & q & 0 & p & 0 & & & \\ & & 0 & q & 0 & p & 0 & & \\ & & & 0 & q & 0 & p & \ddots & \\ & & & & \ddots & \ddots & \ddots & \ddots & \end{bmatrix}$$

Example 3.2.9 (Random walk with absorbing barriers) Let $A > 0$ and $B > 0$. The state space of a random walk with absorbing barriers at $-A$ and B is $S = \{-A, -A+1, \dots, B-1, B\}$. Then the transition probabilities are given as follows. For $-A < i < B$,

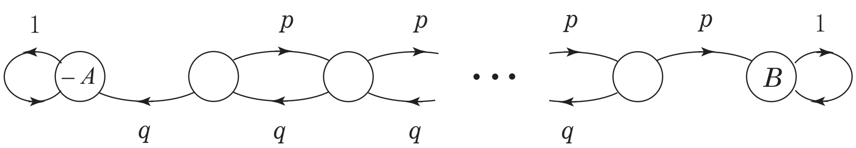
$$p(i,j) = \begin{cases} p, & \text{if } j = i+1, \\ q = 1-p, & \text{if } j = i-1, \\ 0, & \text{otherwise.} \end{cases}$$

For $i = -A$ or $i = B$,

$$p(-A,j) = \begin{cases} 1, & \text{if } j = -A, \\ 0, & \text{otherwise,} \end{cases} \qquad p(B,j) = \begin{cases} 1, & \text{if } j = B, \\ 0, & \text{otherwise.} \end{cases}$$

In a matrix form we have

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & 0 & \cdots & 0 \\ 0 & q & 0 & p & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & q & 0 & p & 0 \\ 0 & 0 & \cdots & 0 & q & 0 & p \\ 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{bmatrix}$$



Example 3.2.10 (Random walk with reflecting barriers) Let $A > 0$ and $B > 0$. The state space of a random walk with absorbing barriers at $-A$ and B is $S = \{-A, -A+1, \dots, B-1, B\}$. The transition probabilities are given as follows. For $-A < i < B$,

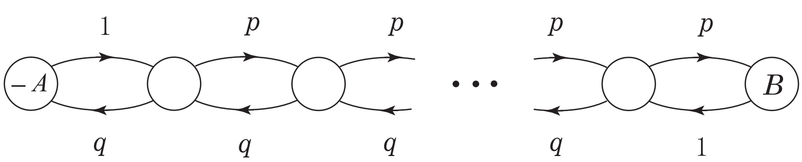
$$p(i,j) = \begin{cases} p, & \text{if } j = i+1, \\ q = 1-p, & \text{if } j = i-1, \\ 0, & \text{otherwise.} \end{cases}$$

For $i = -A$ or $i = B$,

$$p(-A,j) = \begin{cases} 1, & \text{if } j = -A+1, \\ 0, & \text{otherwise,} \end{cases} \qquad p(B,j) = \begin{cases} 1, & \text{if } j = B-1, \\ 0, & \text{otherwise.} \end{cases}$$

In a matrix form we have

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & 0 & \cdots & 0 \\ 0 & q & 0 & p & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & q & 0 & p & 0 \\ 0 & 0 & \cdots & 0 & q & 0 & p \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 \end{bmatrix}$$



3.3 Distribution of a Markov Chain

Let S be a state space as before. In general, a row vector $\pi = [\cdots \pi_i \cdots]$ indexed by S is called a *distribution* on S if

$$\pi_i \geq 0 \quad \text{and} \quad \sum_{i \in S} \pi_i = 1. \tag{3.6}$$

For a Markov chain $\{X_n\}$ on S we set

$$\pi(n) = [\cdots \ \pi_i(n) \ \cdots], \quad \pi_i(n) = P(X_n = i),$$

which becomes a distribution on S. We call $\pi(n)$ the *distribution* of X_n. In particular, $\pi(0)$, the distribution of X_0, is called the *initial distribution*. We often take

$$\pi(0) = [\cdots \ 0, 1, 0, \cdots], \qquad \text{where 1 occurs at } i\text{th posotion.}$$

In this case the Markov chain $\{X_n\}$ starts from the state i.

For a Markov chain $\{X_n\}$ with a transition matrix $P = [p_{ij}]$ the *n-step transition probability* is defined by

$$p_n(i, j) = P(X_{m+n} = j | X_m = i), \qquad i, j \in S.$$

The right-hand side is independent of n because our Markov chain is assumed to be time homogeneous.

Theorem 3.3.1 (Chapman–Kolmogorov equation) For $0 \leq r \leq n$ we have

$$p_n(i, j) = \sum_{k \in S} p_r(i, k) p_{n-r}(k, j). \tag{3.7}$$

Proof First we note the obvious identity:

$$p_n(i, j) = P(X_{m+n} = j | X_m = i) = \sum_{k \in S} P(X_{m+n} = j, X_{m+r} = k | X_m = i).$$

Moreover,

$$\begin{aligned} P(X_{m+n} = j, X_{m+r} = k | X_m = i) &= \frac{P(X_{m+n} = j, X_{m+r} = k, X_m = i)}{P(X_{m+r} = k, X_m = i)} \times \frac{P(X_{m+r} = k, X_m = i)}{P(X_m = i)} \\ &= P(X_{m+n} = j | X_{m+r} = k, X_m = i) P(X_{m+r} = k | X_m = i). \end{aligned}$$

Using the Markov property, we have

$$P(X_{m+n} = j | X_{m+r} = k, X_m = i) = P(X_{m+n} = j | X_{m+r} = k)$$

so that

$$P(X_{m+n} = j, X_{m+r} = k | X_m = i) = P(X_{m+n} = j | X_{m+r} = k) P(X_{m+r} = k | X_m = i).$$

Finally, by the property of being time homogeneous, we come to

$$P(X_{m+n} = j, X_{m+r} = k | X_m = i) = p_{n-r}(k, j) p_r(i, k).$$

Thus we have obtained (3.7). ∎

Applying (3.7) repeatedly and noting that $p_1(i, j) = p(i, j)$, we obtain

$$p_n(i, j) = \sum_{k_1, \dots, k_{n-1} \in S} p(i, k_1) p(k_1, k_2) \cdots p(k_{n-1}, j). \tag{3.8}$$

The right-hand side is nothing else but the multiplication of matrices, i.e., the n-step transition probability $p_n(i, j)$ is the (i, j)-entry of the n-power of the transition matrix P. Summing up, we obtain the following important result.

Theorem 3.3.2 For $m, n \geq 0$ and $i, j \in S$ we have

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

Proof Immediate from Theorem 3.3.1. ∎

Remark 3.3.3 As a result, the Chapman-Kolmogorov equation is nothing else but an entrywise expression of the obvious relation for the transition matrix:

$$P^n = P^r P^{n-r}$$

(As usual, $P^0 = E$ (identity matrix).)

Theorem 3.3.4 We have

$$\pi(n) = \pi(n-1)P, \qquad n \geq 1,$$

or equivalently,

$$\pi_j(n) = \sum_i \pi_i(n-1) p_{ij}\,.$$

Therefore,

$$\pi(n) = \pi(0)P^n.$$

Proof We first note that

$$\pi_j(n) = P(X_n = j) = \sum_{i \in S} P(X_n = j | X_{n-1} = i) P(X_{n-1} = i) = \sum_{i \in S} p_{ij} \pi_i(n-1),$$

which proves $\pi(n) = \pi(n-1)P$. By repeated application we have

$$\pi(n) = \pi(n-1)P = (\pi(n-2)P)P = (\pi(n-2)P^2 = \cdots = \pi(0)P^n,$$

as desired. ∎

Example 3.3.5 (2-state Markov chain) Let $\{X_n\}$ be the Markov chain introduced in Example 3.2.6. The eigenvalues of the transition matrix

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

are $1, 1-p-q$. These are distinct if $p+q > 0$. Omitting the case of $p+q = 0$, i.e., $p = q = 0$, we assume that $p+q > 0$. By standard argument we obtain

$$P^n = \frac{1}{p+q} \begin{bmatrix} q + pr^n & p - pr^n \\ q - qr^n & p + qr^n \end{bmatrix}, \qquad r = 1-p-q.$$

Let $\pi(0) = [\pi_0(0)\ \pi_1(0)]$ be the distriution of X_0. Then the distribution of X_n is given by

$$\pi(n) = [P(X_n = 0)\,,\ P(X_n = 1)] = [\pi_0(0)\ \pi_1(0)]P^n = \pi(0)P^n.$$

Problem 7 There are two parties, say, A and B, and their supporters of a constant ratio exchange at every election. Suppose that just before an election, 25% of the supporters of A change to support B and 20% of the supporters of B change to support A. At the beginning, 85% of the voters support A and 15% support B.

(1) When will the party B command a majority?

(2) Find the final ratio of supporters after many elections if the same situation continues.

(3) Discuss relevant topics.

Problem 8 Let $\{X_n\}$ be a Markov chain on $\{0, 1\}$ given by the transition matrix $P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$ with the initial distribution $\pi_0 = \left[\dfrac{q}{p+q}\,, \dfrac{p}{p+q}\right]$. Calculate the following statistical quantities:

$$\mathbf{E}[X_n], \quad \mathbf{V}[X_n], \quad \mathrm{Cov}\,(X_{m+n}, X_n) = \mathbf{E}[X_{m+n}X_n] - \mathbf{E}[X_{m+n}]\mathbf{E}[X_n], \quad \rho(X_{m+n}, X_n) = \frac{\mathrm{Cov}\,(X_{m+n}, X_n)}{\sqrt{\mathbf{V}[X_{m+n}]\mathbf{V}[X_n]}}$$

4 Stationary Distributions

4.1 Definition and Examples

Definition 4.1.1 Let $\{X_n\}$ be a Markov chain on S with transition probability matrix P. A distribution π on S is called *stationary* (or *invariant*) if

$$\pi = \pi P, \tag{4.1}$$

or equivalently,

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}, \qquad j \in S. \tag{4.2}$$

Thus, to find a stationary distribution we need to solve (4.1) (or equivalently (4.2)) together with (3.6). If S is a finite set, finding stationary distributions is reduced to a simple linear system.

Example 4.1.2 (2-state Markov chain) Consider the transition matrix:

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

Let $\pi = [\pi_0 \, \pi_1]$ and suppose $\pi P = \pi$. Then we have

$$[\pi_0 \, \pi_1] \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} = [(1-p)\pi_0 + q\pi_1 \;\; p\pi_0 + (1-q)\pi_1] = [\pi_0 \, \pi_1],$$

which is equivalent to the following

$$p\pi_0 - q\pi_1 = 0.$$

Together with $\pi_0 + \pi_1 = 1$, we obtain

$$\pi_0 = \frac{q}{p+q}, \quad \pi_1 = \frac{p}{p+q},$$

whenever $p + q > 0$. Indeed, $\pi_0 \geq 0$ and $\pi_1 \geq 0$, so this is a stationary distribution.

The following properties are noteworthy:

(i) If $p + q > 0$, a stationary distribution is unique.

(ii) If $p = q = 0$, the stationary distribution is not uniquely determined. In fact, any distribution $\pi = [\pi_0 , \pi_1]$ is stationary.

Moreover, we see from Example 3.3.5 that if $0 < p + q < 2$, or equivalently, if $|r| < 1$, we have

$$\lim_{n\to\infty} P^n = \frac{1}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix}.$$

Then

$$\lim_{n\to\infty} \pi(n) = \lim_{n\to\infty} \pi(0)P^n = [\pi_0(0) \; \pi_1(0)] \times \frac{1}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix} = \begin{bmatrix} \dfrac{q}{p+q} & \dfrac{p}{p+q} \end{bmatrix}.$$

Thus we get the stationary distribution as a limit distribution.

Example 4.1.3 (3-state Markov chain) We discuss the Markov chain $\{X_n\}$ introduced in Example 3.2.7. If $q > 0$ and $b > 0$, a stationary distribution is unique and given by $\pi = [0 \;\, 0 \;\, 1]$.

Example 4.1.4 (One-dimensional RW) Consider the 1-dimensional random walk with right-move probability $p > 0$ and left-move probability $q = 1 - p > 0$. Let $[\cdots \pi(k) \cdots]$ be a distribution on $\mathbf{Z}$. If it is stationary, we have

$$\pi(k) = p\pi(k-1) + q\pi(k+1), \qquad k \in \mathbf{Z}. \tag{4.3}$$

The characteristic equation of the above difference equation is

$$0 = q\lambda^2 - \lambda + p = (q\lambda - p)(\lambda - 1)$$

so that the eigenvalues are $1, p/q$.

(Case 1) $p \neq q$. Then a general solution to (4.3) is given by

$$\pi(k) = C_1 1^k + C_2 \left(\frac{p}{q}\right)^k = C_1 + C_2 \left(\frac{p}{q}\right)^k, \qquad k \in \mathbf{Z}.$$

This never becomes a probability distribution for any choice of C_1 and C_2. Namely, there is no stationary distribution.

(Case 2) $p = q$. In this case a general solution to (4.3) is given by

$$\pi(k) = (C_1 + C_2 k)1^k = C_1 + C_2 k, \qquad k \in \mathbf{Z}.$$

This never becomes a probability distribution for any choice of C_1 and C_2. Namely, there is no stationary distribution.

Example 4.1.5 (One-dimensional RW with reflection barrier) There is a unique stationary distribution when $p < q$. In fact,

$$\pi(0) = Cp, \quad \pi(k) = C\left(\frac{p}{q}\right)^k, \quad k \geq 1,$$

where C is determined in such a way that $\sum_{k=0}^{\infty} \pi(k) = 1$. Namey,

$$C = \frac{q - p}{2pq}.$$

If $p \geq q$, then there is no stationary distribution.

On stationary distributions of a Markov chain we question:

(1) Is there a stationary distribution?

(2) If yes, is it unique? If not, how to classify?

(3) Does the distributions of a Markov chain converge to a stationary distribution?

4.2 Existence

Theorem 4.2.1 A Markov chain over a finite state space S has a stationary distribution.

A simple proof is based on the Brouwer's fixed-point theorem saying that every continuous function from a convex compact subset of a Euclidean space to itself has a fixed point. In fact, the set of distributions on S is a convex compact subset of a Euclidean space and the map $\pi \mapsto \pi P$ is continuous. Note that the stationary distribution mentioned in the above theorem is not necessarily unique.

Definition 4.2.2 We say that a state j *can be reached from* a state i if there exists some $n \geq 0$ such that $p_n(i, j) > 0$. By definition every state i can be reached from itself. We say that two states i and j *intercommunicate* if i can be reached form j and j can be reached from i, i.e., there exist $m \geq 0$ and $n \geq 0$ such that $p_n(i, j) > 0$ and $p_m(j, i) > 0$.

For $i, j \in S$ we introduce a binary relation $i \sim j$ when they intercommunicate. Then $\sim$ becomes an equivalence relation on S:

$$\text{(i) } i \sim i; \quad \text{(ii) } i \sim j \Longrightarrow j \sim i; \quad \text{(iii) } i \sim j,\ j \sim k \Longrightarrow i \sim k.$$

In fact, (i) and (ii) are obvious by definition, and (iii) is verified by the Chapman-Kolmogorov equation. Thereby the state space S is classified into a disjoint set of equivalence classes. In each equivalence class any two states intercommunicate each other.

Definition 4.2.3 A state i is called *absorbing* if $p_{ii} = 1$. In particular, an absorbing state is a state which constitutes an equivalence class by itself.

Definition 4.2.4 A Markov chain is called *irreducible* if every state can be reached from every other state, i.e., if there is only one equivalence class of intercommunicating states.

Theorem 4.2.5 An irreducible Markov chain on a finite state space S admits a unique stationary distribution $\pi = [\pi_i]$. Moreover, $\pi_i > 0$ for all $i \in S$.

In fact, the proof owes to the following two facts:

(1) For an irreducible Markov chain the following assertions are equivalent:

(i) it admits a stationary distribution;

(ii) every state is positive recurrent.

In this case the stationary distribution π is unique and given by

$$\pi_i = \frac{1}{\mathbf{E}(T_i|X_0 = i)}, \qquad i \in S.$$

(2) Every state of an irreducible Markov chain on a finite state space is positive recurrent (Theorem 5.1.8).

4.3 Convergence

Example 4.3.1 (2-state Markov chain) We recall Examples 3.3.5 and 4.1.2. If $p + q > 0$, the distribution of the above Markov chain converges to the unique stationary distribution. Consider the case of $p = q = 1$, i.e., the transition matrix becomes

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

The stationary distribution is unique. But for a given initial distribution $\pi(0)$ it is not necessarily true that $\pi(n)$ converges to the stationary distribution.

Roughly speaking, we need to avoid the periodic transition in order to have the convergence to a stationary distribution.

Definition 4.3.2 For a state $i \in S$,

$$\mathrm{GCD}\{n \geq 1\,;\, P(X_n = i|X_0 = i) > 0\}$$

is called the *period* of i. (When the set in the right-hand side is empty, the period is not defined.) A state $i \in S$ is called *aperiodic* if its period is one.

Theorem 4.3.3 For an irreducible Markov chain, every state has a common period.

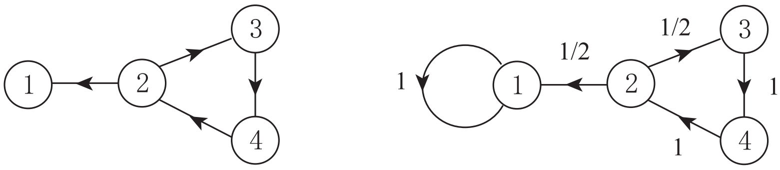
Theorem 4.3.4 Let π be a stationary distribution of an irreducible Markov chain on a finite state space (It is unique, see Theorem 4.2.5). If $\{X_n\}$ is aperiodic, for any $j \in S$ we have

$$\lim_{n\to\infty} P(X_n = j) = \pi_j.$$

Example 4.3.5 (page rank) The hyperlinks among N websites give rise to a digraph (directed graph) G on N vertices. It is natural to consider a Markov chain on G, which is defined by the transition matrix $P = [p_{ij}]$, where

$$p_{ij} = \begin{cases} \dfrac{1}{\deg i} & \text{if } i \to j, \\ 0, & \text{if } i \nrightarrow j \text{ and } i \neq j, \\ 1, & \deg i = 0 \text{ and } j = i, \end{cases}$$

where $\deg i = |\{j\,;\, i \to j\}|$ is the *out-degree* of i.



There exists a stationary state but not necessarily unique. Taking $0 \le d \le 1$ we modify the transition matrix:

$$Q = [q_{ij}], \qquad q_{ij} = dp_{ij} + \epsilon, \qquad \epsilon = \frac{1-d}{N}.$$

If $0 \le d < 1$, the Markov chain determined by Q has necessarily a unique stationary distribution. Choosing a suitable $d < 1$, we may understand the stationary distribution $\pi = [\pi(i)]$ as the page rank among the websites.

Problem 9 Consider the page rank introduced in Example 4.3.5.

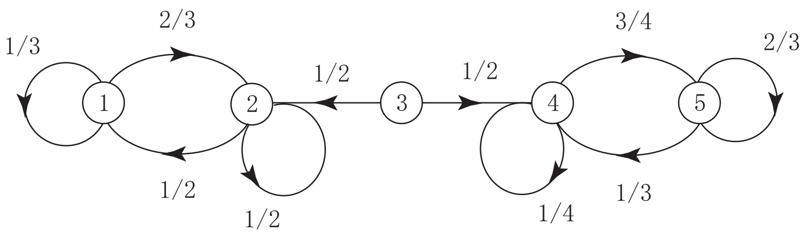
(1) Let $\pi(i)$ be the page rank of a site i. Show that $\pi(i)$ satisfies the following relation

$$\pi(i) = \frac{1-d}{N} + d \sum_{j:j\to i} \frac{\pi(j)}{\deg j}$$

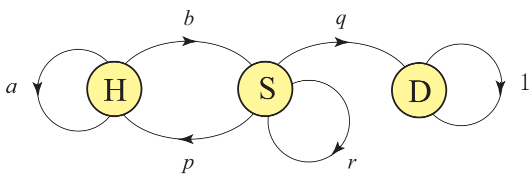
and explain the meaning.

(2) Show more examples of the page rank and discuss the role of sites which have no hyperlinks, that is, $\deg i = 0$ (in terms of $P = [p_{ij}]$ such sites correspond to absorbing states).

Problem 10 Find all stationary distributions of the Markov chain determined by the transition diagram below. Then discuss convergence of distributions.



Problem 11 Let $\{X_n\}$ be the Markov chain introduced in Example 3.2.7:



For $n = 1, 2, \dots$ let H_n denote the probability of starting from H and terminating at D at n-step. Similarly, for $n = 1, 2, \dots$ let S_n denote the probability of starting from S and terminating at D at n-step.

(1) Show that $\{H_n\}$ and $\{S_n\}$ satisfies the following linear system:

$$\begin{cases} H_n = aH_{n-1} + bS_{n-1}, \\ S_n = pH_{n-1} + rS_{n-1}, \end{cases} \qquad n \ge 2; \qquad H_1 = 0, \quad S_1 = q.$$

(2) Let H and S denote the life times starting from the state H and S, respectively. Solving the linear system in (1), prove the following identities for the mean life times:

$$\mathbf{E}[H] = \sum_{n=1}^{\infty} nH_n = \frac{b+p+q}{bq}, \qquad \mathbf{E}[S] = \sum_{n=1}^{\infty} nS_n = \frac{b+p}{bq}.$$

5 Topics in Markov Chains

5.1 Recurrence

Definition 5.1.1 Let $i \in S$ be a state. Define the *first hitting time* or *first passage time* to i by

$$T_i = \inf\{n \geq 1\,;\, X_n = i\}.$$

If there exists no $n \geq 1$ such that $X_n = i$, we define $T_i = \infty$. A state i is called *recurrent* if $P(T_i < \infty | X_0 = i) = 1$. It is called *transient* if $P(T_i = \infty | X_0 = i) > 0$.

Theorem 5.1.2 A state $i \in S$ is recurrent if and only if

$$\sum_{n=0}^{\infty} p_n(i, i) = \infty.$$

If a state i is transient, we have

$$\sum_{n=0}^{\infty} p_n(i, i) < \infty \quad \text{and} \quad \sum_{n=0}^{\infty} p_n(i, i) = \frac{1}{1 - P(T_i < \infty | X_0 = i)}.$$

Proof We first put

$$p_n(i, j) = P(X_n = j | X_0 = i), \quad n = 0, 1, 2, \dots,$$
$$f_n(i, j) = P(T_j = n | X_0 = i) = P(X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j | X_0 = i), \quad n = 1, 2, \dots.$$

$p_n(i, j)$ is nothing else but the n step transition probability. On the other hand, $f_n(i, j)$ is the probability that the Markov chain starts from i and reach j first time after n step. Dividing the set of sample paths from i to j in n steps according to the number of steps after which the path reaches j for the first time, we obtain

$$p_n(i, j) = \sum_{r=1}^{n} f_r(i, j) p_{n-r}(j, j), \qquad i, j \in S, \quad n = 1, 2, \dots. \tag{5.1}$$

We next introduce the generating functions:

$$G_{ij}(z) = \sum_{n=0}^{\infty} p_n(i, j) z^n, \qquad F_{ij}(z) = \sum_{n=1}^{\infty} f_n(i, j) z^n.$$

In view of (5.1) we see easily that

$$G_{ij}(z) = p_0(i, j) + F_{ij}(z) G_{jj}(z). \tag{5.2}$$

Setting $i = j$ in (5.2), we obtain

$$G_{ii}(z) = 1 + F_{ii}(z) G_{ii}(z) \quad \Longleftrightarrow \quad G_{ii}(z) = \frac{1}{1 - F_{ii}(z)}.$$

On the other hand, since

$$G_{ii}(1) = \sum_{n=0}^{\infty} p_n(i, i), \qquad F_{ii}(1) = \sum_{n=1}^{\infty} f_n(i, i) = P(T_i < \infty | X_0 = i)$$

we see that two conditions $F_{ii}(1) = 1$ and $G_{ii}(1) = \infty$ are equivalent. The second statement is readily clear. ∎

Example 5.1.3 (random walk on $\mathbb{Z}$) Since the random walk starting from the origin 0 returns to it only after even steps, for recurrence we only need to compute the sum of $p_{2n}(0,0)$. We start with the obvious result:

$$p_{2n}(0,0) = \frac{(2n)!}{n!n!}\, p^n q^n, \qquad p+q=1.$$

Then, using the Stirling formula:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \tag{5.3}$$

we obtain

$$p_{2n}(0,0) \sim \frac{1}{\sqrt{\pi n}}\,(4pq)^n.$$

Hence,

$$\sum_{n=0}^{\infty} p_{2n}(0,0) \begin{cases} < \infty, & p \neq q, \\ = \infty, & p = q = 1/2. \end{cases}$$

Consequently, one-dimensional random walk is transient if $p \neq q$, and it is recurrent if $p = q = \dfrac{1}{2}$.

Remark 5.1.4 Let $\{a_n\}$ and $\{b_n\}$ be sequences of positive numbers. We write $a_n \sim b_n$ if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 1.$$

In this case, there exist two constant numbers $c_1 > 0$ and $c_2 > 0$ such that $c_1 a_n \le b_n \le c_2 a_n$. Hence $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ converge or diverge at the same time.

Example 5.1.5 (random walk on $\mathbb{Z}^2$) Obviously, the random walk starting from the origin 0 returns to it only after even steps. Therefore, for recurrence we only need to compute the sum of $p_{2n}(0,0)$. For two-dimensional random walk we need to consider two directions along with x-axis and y-axis. We see easily that

$$p_{2n}(0,0) = \sum_{i+j=n} \frac{(2n)!}{i!i!j!j!}\left(\frac{1}{4}\right)^{2n} = \frac{(2n)!}{n!n!}\left(\frac{1}{4}\right)^{2n} \sum_{i+j=n} \frac{n!n!}{i!i!j!j!} = \binom{2n}{n}\left(\frac{1}{4}\right)^{2n} \sum_{i=0}^{n} \binom{n}{i}^2.$$

Employing the formula for the binomial coefficients:

$$\sum_{i=0}^{n} \binom{n}{i}^2 = \binom{2n}{n}, \tag{5.4}$$

which is a good exercise for the readers, we obtain

$$p_{2n}(0,0) = \binom{2n}{n}^2 \left(\frac{1}{4}\right)^{2n}.$$

Then, by using the Stirling formula, we see that

$$p_{2n}(0,0) \sim \frac{1}{\pi n}$$

so that

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

Consequently, two-dimensional random walk is recurrent.

Example 5.1.6 (random walk on $\mathbb{Z}^3$) Let us consider the isotropic random walk in 3-dimension. As there are three directions, say, x, y, z-axis, we have

$$p_{2n}(0,0) = \sum_{i+j+k=n} \frac{(2n)!}{i!i!j!j!k!k!}\left(\frac{1}{6}\right)^{2n} = \frac{(2n)!}{n!n!}\left(\frac{1}{6}\right)^{2n} \sum_{i+j+k=n} \frac{n!n!}{i!i!j!j!k!k!} = \binom{2n}{n}\left(\frac{1}{6}\right)^{2n} \sum_{i+j+k=n} \left(\frac{n!}{i!j!k!}\right)^2.$$

We note the following two facts. First,

$$\sum_{i+j+k=n} \frac{n!}{i!\,j!k!} = 3^n. \tag{5.5}$$

Second, the maximum value

$$M_n = \max_{i+j+k=n} \frac{n!}{i!\,j!k!}$$

is attained when $\frac{n}{3} - 1 \le i, j, k \le \frac{n}{3} + 1$ so

$$M_n \sim \frac{3\sqrt{3}}{2\pi n}\, 3^n$$

by the Stirling formula. Then we have

$$p_{2n}(0,0) \le \binom{2n}{n}\left(\frac{1}{6}\right)^{2n} 3^n M_n \sim \frac{3\sqrt{3}}{2\pi\sqrt{\pi}}\, n^{-3/2}.$$

Therefore.

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

which implies that the random walk is not recurrent (i.e., transient).

If a state i is recurrent, i.e., $P(T_i < \infty | X_0 = i) = 1$, the *mean recurrent time* is defined:

$$\mathbf{E}(T_i | X_0 = i) = \sum_{n=1}^{\infty} nP(T_i = n | X_0 = i).$$

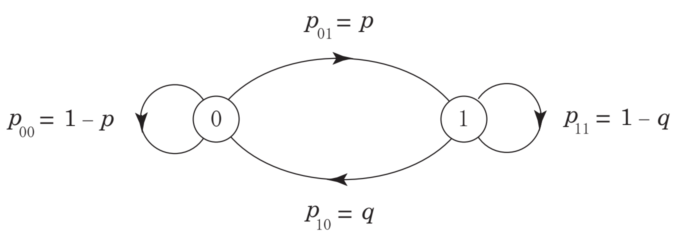
The state i is called *positive recurrent* if $\mathbf{E}(T_i | X_0 = i) < \infty$, and *null recurrent* otherwise.

Theorem 5.1.7 The states in an equivalence class are all positive recurrent, or all null recurrent, or all transient. In particular, for an irreducible Markov chain, the states are all positive recurrent, or all null recurrent, or all transient.

Theorem 5.1.8 For an irreducible Markov chain on a finite state space S, every state is positive recurrent.

Example 5.1.9 The mean recurrent time of the one-dimensional isotropic random walk is infinity, i.e., the one-dimensional isotropic random walk is null recurrent. The proof will be given in Section **??**.

Problem 12 Let $\{X_n\}$ be a Markov chain described by the following transition diagram:



where $p > 0$ and $q > 0$. For a state $i \in S$ let T_i be the first hitting time to i defined by

$$T_i = \inf\{n \ge 1\,;\ X_n = i\}.$$

(1) Calculate

$$P(T_0 = 1 | X_0 = 0), \quad P(T_0 = 2 | X_0 = 0), \quad P(T_0 = 3 | X_0 = 0), \quad P(T_0 = 4 | X_0 = 0).$$

(2) Find $P(T_0 = n | X_0 = 0)$ and calculate

$$\sum_{n=1}^{\infty} P(T_0 = n | X_0 = 0), \qquad \sum_{n=1}^{\infty} nP(T_0 = n | X_0 = 0).$$

5.2 Absorption

A state i is called *absorbing* if $p_{ii} = 1$ and $p_{ij} = 0$ for all $j \neq i$. Once a Markov chain hits an absorbing state, it stays thereat forever.

Let us consider a Markov chain on a finite state space S with some absorbing states. We set

$$S = S_a \cup S_0 ,$$

where S_a denotes the set of absorbing states and S_0 the rest. According to the above partition, the transition matrix is written as

$$P = \begin{bmatrix} 1 & & & & 0 & \cdots & 0 \\ & 1 & & & 0 & \cdots & 0 \\ & & \ddots & & \vdots & \ddots & \vdots \\ & & & 1 & 0 & \cdots & 0 \\ & & & & & & \\ & & * & & & * & \\ & & & & & & \end{bmatrix} = \begin{bmatrix} I & 0 \\ S & T \end{bmatrix}.$$

Then

$$P^n = \begin{bmatrix} I & 0 \\ S & T \end{bmatrix}^n = \begin{bmatrix} I & 0 \\ S_n & T^n \end{bmatrix},$$

where $S_1 = S$ and $S_n = S_{n-1} + T^{n-1}S$. To avoid inessential tediousness we assume the following condition

(C1) For any $i \in S_0$ there exist $j \in S_a$ and $n \geq 1$ such that $(P^n)_{ij} > 0$.

In other words, the Markov chain starting from $i \in S_0$ has a positive probability of absorption. Since S is finite by assumption, the n in (C1) is chosen independently of $i \in S_0$. Hence (C1) is equivalent to the following

(C2) There exists $N \geq 1$ such that for any $i \in S_0$ there exist $j \in S_a$ with $(P^N)_{ij} > 0$.

Lemma 5.2.1 Notations and assumptions being as above, $\lim_{n\to\infty} T^n = 0$.

Proof We see from the obvious relation

$$1 = \sum_{j\in S} (P^N)_{ij} = \sum_{j\in S_0} (P^N)_{ij} + \sum_{j\in S_a} (P^N)_{ij}$$

and condition (C2) that

$$\sum_{j\in S_0} (P^N)_{ij} < 1, \qquad i \in S_0 .$$

Note that for $i, j \in S_0$ we have $(P^N)_{ij} = (T^N)_{ij}$. We choose $\delta < 1$ such that

$$\sum_{j\in S_0} (T^N)_{ij} \leq \delta < 1 \qquad i \in S_0 .$$

Now let $i \in S_0$ and $n \geq N$. We see that

$$\sum_{j\in S_0} (T^n)_{ij} = \sum_{j,k\in S_0} (T^{n-N})_{ik} (T^N)_{kj} = \sum_{k\in S_0} (T^{n-N})_{ik} \sum_{j\in S_0} (T^N)_{kj} \leq \delta \sum_{k\in S_0} (T^{n-N})_{ik} = \delta \sum_{j\in S_0} (T^{n-N})_{ij}.$$

Repeating this procedure, we have

$$\sum_{j\in S_0} (T^n)_{ij} \leq \delta^k \sum_{j\in S_0} (T^{n-kN})_{ij} \leq \delta^k \sum_{j\in S} (P^{n-kN})_{ij} \leq \delta^k,$$

where $0 \leq n - kN < N$. Therefore,

$$\lim_{n\to\infty} \sum_{j\in S_0} (T^n)_{ij} = 0,$$

from which we have $\lim_{n\to\infty} (T^n)_{ij} = 0$ for all $i, j \in S_0$. ∎

Remark 5.2.2 It is shown that every state $i \in S_0$ is transient.

Theorem 5.2.3 Let $\pi_0 = [\alpha\ \beta]$ be the initial distribution (according to $S = S_a \cup S_0$). Then the limit distribution is given by

$$[\alpha + \beta S_\infty\ \ 0], \qquad \text{where} \quad S_\infty = (I - T)^{-1} S.$$

Proof The limit distribution is given by

$$\lim_{n\to\infty} \pi_0 P^n = \lim_{n\to\infty} [\alpha\ \beta] \begin{bmatrix} I & 0 \\ S_n & T^n \end{bmatrix} = \lim_{n\to\infty} [\alpha + \beta S_n\ \ \beta T^n].$$

We see from Lemma 5.2.1 that

$$\lim_{n\to\infty} \beta T^n = 0.$$

On the other hand, since $S_n = S_{n-1} + T^{n-1} S$ we have

$$S_n = (I + T + T^2 + \cdots + T^{n-1}) S$$

and

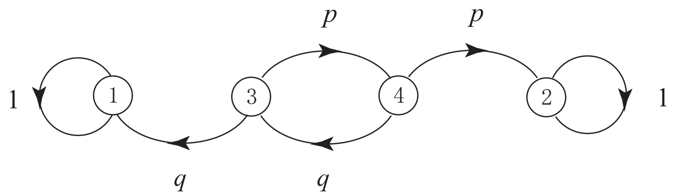
$$(I - T) S_n = (I - T^n) S.$$

Hence

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} (I - T)^{-1} (I - T^n) S = (I - T)^{-1} S,$$

which shows the result. ∎

Example 5.2.4 Consider the Markov chain given by the transition diagram, which is a random walk with absorbing barriers.



The transition matrix is given by

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ q & 0 & 0 & p \\ 0 & p & q & 0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ S & T \end{bmatrix}, \qquad S = \begin{bmatrix} q & 0 \\ 0 & p \end{bmatrix}, \quad T = \begin{bmatrix} 0 & p \\ q & 0 \end{bmatrix}.$$

Then

$$S_\infty = (I - T)^{-1} S = \frac{1}{1 - pq} \begin{bmatrix} q & p^2 \\ q^2 & p \end{bmatrix}$$

Suppose that the initial distribution is given by $\pi_0 = [\alpha\, \beta\, \gamma\, \delta]$. Then the limit distribution is

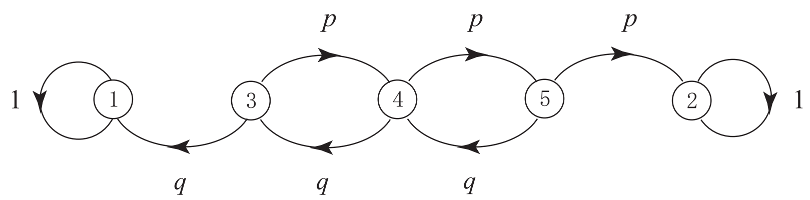
$$\left[\alpha + \frac{q\gamma + q^2\delta}{1 - pq} \quad \beta + \frac{p^2\gamma + p\delta}{1 - pq} \quad 0 \quad 0 \right].$$

In particular, if the Markov chain starts at the state 3, setting $\pi_0 = [0\ 0\ 1\ 0]$, we obtain the limit distribution

$$\left[\frac{q}{1 - pq} \quad \frac{p^2}{1 - pq} \quad 0 \quad 0 \right],$$

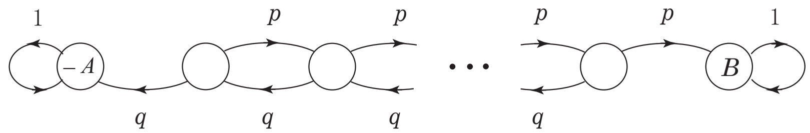
which means that the Markov chain is absorbed in the states 1 or 2 at the ratio $q : p^2$.

Problem 13 Following Example 5.2.4, study the Markov chain given by the following transition diagram, where $p + q = 1$.



5.3 Gambler's Ruin

We consider a random walk with absorbing barriers at $-A$ and B, where $A > 0$ and $B > 0$. This is a Markov chain on the state space $S = \{-A, -A + 1, \dots, B - 1, B\}$ with the transition diagram as follows:



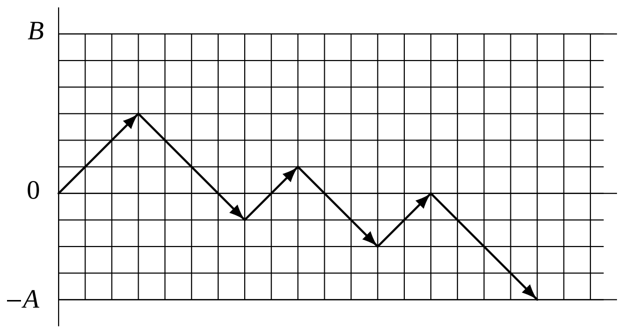
We are interested in the absorbing probability, i.e.,

$$R = P(X_n = -A \text{ for some } n = 1, 2, \dots) = P\left(\bigcup_{n=1}^{\infty} \{X_n = -A\}\right),$$

$$S = P(X_n = B \text{ for some } n = 1, 2, \dots) = P\left(\bigcup_{n=1}^{\infty} \{X_n = B\}\right).$$

Note that the events in the right-hand sides are not the unions of disjoint events.

A sample path is shown in the following picture:



A key idea is to introduce a similar random walk starting at k, $-A \le k \le B$, which is denoted by $X_n^{(k)}$. Then the original one is $X_n = X_n^{(0)}$. Let R_k and S_k be the probabilities that the random walk $X_n^{(k)}$ is absorbed at $-A$ and B, respectively. We wish to find $R = R_0$ and $S = S_0$.

Lemma 5.3.1 $\{R_k\,;\,, -A \le k \le B\}$ fulfills the following difference equation:

$$R_k = pR_{k+1} + qR_{k-1}, \qquad R_{-A} = 1, \qquad R_B = 0. \tag{5.6}$$

Similarly, $\{S_k\,;\,, -A \le k \le B\}$ fulfills the following difference equation:

$$S_k = pS_{k+1} + qS_{k-1}, \qquad S_{-A} = 0, \qquad S_B = 1. \tag{5.7}$$

Theorem 5.3.2 Let $A \geq 1$ and $B \geq 1$. Let $\{X_n\}$ be the random walk with absorbing barriers at $-A$ and B, and with right-move probability p and left-move probability q $(p + q = 1)$. Then the probabilities that $\{X_n\}$ is absorbed at the barriers are given by

$$P(X_n = -A \text{ for some } n) = \begin{cases} \dfrac{(q/p)^A - (q/p)^{A+B}}{1 - (q/p)^{A+B}}, & p \neq q, \\ \dfrac{B}{A+B}, & p = q = \dfrac{1}{2}, \end{cases}$$

$$P(X_n = B \text{ for some } n) = \begin{cases} \dfrac{1 - (q/p)^A}{1 - (q/p)^{A+B}}, & p \neq q, \\ \dfrac{A}{A+B}, & p = q = \dfrac{1}{2}. \end{cases}$$

In particular, the random walk is absorbed at the barriers at probability 1.

An interpretation of Theorem 5.3.2 gives the solution to the *gambler's ruin problem*. Two players A and B toss a fair coin by turns. Let A and B be their allotted points when the game starts. They exchange 1 point after each trial. This game is over when one of the players loses all the allotted points and the other gets $A + B$ points. We are interested in the probability of each player's win. For each $n \geq 0$ define X_n in such a way that the allotted point of A at time n is given by $A + X_n$. Then $\{X_n\}$ becomes a random walk with absorbing barrier at $-A$ and B. It then follows from Theorem 5.3.2 that the winning probability of A and B are given by

$$P(A) = \frac{A}{A+B}, \qquad P(B) = \frac{B}{A+B}, \tag{5.8}$$

respectively. As a result, they are proportional to the initial allotted points. For example, if $A = 1$ and $B = 100$, we have $P(A) = 1/101$ and $P(B) = 100/101$, which sounds that almost no chance of A's win.

In a fair bet the recurrence is guaranteed by Theorem 6.1.11. Even if one has much more losses than wins, continuing the game one will be back to the zero balance. However, in reality there is a barrier of limited money. (5.8) tells the effect of the barrier.

It is also interesting to know the expectation of the number of coin tosses until the game is over.

Theorem 5.3.3 Let $\{X_n\}$ be the same as in Theorem 5.3.2. The expected life time of this random walk until absorption is given by

$$\begin{cases} \dfrac{A}{q-p} - \dfrac{A+B}{q-p}\,\dfrac{1 - (q/p)^A}{1 - (q/p)^{A+B}}, & p \neq q, \\ AB, & p = q = \dfrac{1}{2}. \end{cases}$$

Proof Let Y_k be the life time of a random walk starting from the position k $(-A \leq k \leq B)$ at time $n = 0$ until absorption. In other words,

$$Y_k = \min\{j \geq 0\,;\ X_j^{(k)} = -A \text{ または } X_j^{(k)} = B\}.$$

We wish to compute $\mathbf{E}(Y_0)$. We see by definition that

$$\mathbf{E}(Y_{-A}) = \mathbf{E}(Y_B) = 0. \tag{5.9}$$

For $-A < k < B$ we have

$$\mathbf{E}(Y_k) = \sum_{j=1}^{\infty} jP(Y_k = j). \tag{5.10}$$

In a similar manner as in the proof of Theorem 5.3.2 we note that

$$P(Y_k = j) = pP(Y_{k+1} = j - 1) + qP(Y_{k-1} = j - 1). \tag{5.11}$$

Inserting (5.11) into (5.10), we obtain

$$\begin{aligned} \mathbf{E}(Y_k) &= p\sum_{j=1}^{\infty} jP(Y_{k+1} = j-1) + q\sum_{j=1}^{\infty} jP(Y_{k-1} = j-1) \\ &= p\mathbf{E}(Y_{k+1}) + q\mathbf{E}(Y_{k-1}) + 1. \end{aligned} \tag{5.12}$$

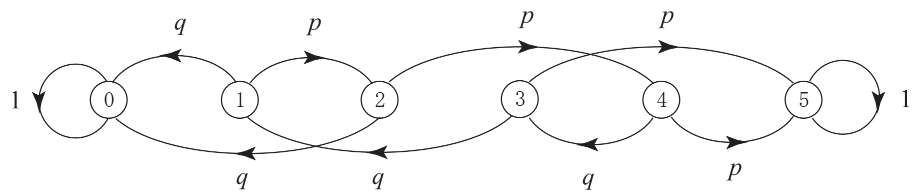
Thus, $\mathbf{E}(Y_k)$ is the solution to the difference equation (5.12) with boundary condition (5.9). This difference equation is solved in a standard manner and we find

$$\mathbf{E}(Y_k) = \begin{cases} \dfrac{A+k}{q-p} - \dfrac{A+B}{q-p}\,\dfrac{1-(q/p)^{A+k}}{1-(q/p)^{A+B}}, & p \neq q, \\ (A+k)(B-k), & p = q = \dfrac{1}{2}. \end{cases}$$

Setting $k = 0$, we obtain the result. ∎

If $p = q = 1/2$ and $A = 1$, $B = 100$, the expected life time is $AB = 100$. The gambler A is much inferior to B in the amount of funds (as we have seen already, the probability of A's win is just $1/101$), however, the expected life time until the game is over is 100, which sounds longer than one expects intuitively. Perhaps this is because the gambler cannot quit gambling.

Problem 14 (A bold gambler) In each game a gambler wins the dollars he bets with probability p, and loses with probability $q = 1 - p$. The goal of the gambler is to get 5 dollars. His strategy is to bet the difference between 5 dollars and what he has. Let X_n be the amount he has just after nth bet.



(1) Analyze the Markov chain $\{X_n\}$ with initial condition $X_0 = 1$.

(2) Compare with the steady gambler discussed in this section, who bets just 1 dollar in each game.

6 Topics in Random Walks

6.1 The Catalan Number

The Catalan number is a famous number known in combinatorics (Eugène Charles Catalan, 1814–1894). Richard P. Stanley (MIT) collected many appearances of the Catalan numbers (R. P. Stanley: "Catalan Numbers," Cambridge University Press, 2015; http://www-math.mit.edu/ rstan/ec/).

We start with the definition. Let $n \geq 1$ and consider a sequence $(\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ of ± 1, that is, an element of $\{-1, 1\}^n$. This sequence is called a *Catalan path* if

$$\begin{aligned} \epsilon_1 &\geq 0 \\ \epsilon_1 + \epsilon_2 &\geq 0 \\ &\cdots \\ \epsilon_1 + \epsilon_2 + \cdots + \epsilon_{n-1} &\geq 0 \\ \epsilon_1 + \epsilon_2 + \cdots + \epsilon_{n-1} + \epsilon_n &= 0. \end{aligned}$$

It is apparent that there is no Catalan path of odd length.

Definition 6.1.1 The nth *Catalan number* is defined to be the number of Catalan paths of length $2n$ and is denoted by C_n. For convenience we set $C_0 = 1$.

The first Catalan numbers for $n = 0, 1, 2, 3, \dots$ are

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, 208012, 742900, 2674440, \dots$$

We will derive a concise expression for the Catalan numbers by using a graphical representation. Consider $n \times n$ grid with the bottom-left corner being given the coordinate $(0, 0)$. With each sequence $(\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ consisting of ± 1 we associate vectors

$$\epsilon_k = +1 \leftrightarrow u_k = (1, 0) \qquad \epsilon_k = -1 \leftrightarrow u_k = (0, 1)$$

and consider a polygonal line connecting

$$(0, 0),\ u_1,\ u_1 + u_2,\ \dots,\ u_1 + u_2 + \cdots + u_{n-1},\ u_1 + u_2 + \cdots + u_{n-1} + u_n$$

in order. If $\epsilon_1 + \epsilon_2 + \cdots + \epsilon_{n-1} + \epsilon_n = 0$, the final vertex becomes

$$u_1 + u_2 + \cdots + u_{n-1} + u_n = (n, n)$$

so that the obtained polygonal line is a shortest path connecting $(0, 0)$ and (n, n) in the grid.

Lemma 6.1.2 There is a one-to-one correspondence between the Catalan paths of length $2n$ and the shortest paths connecting $(0, 0)$ and (n, n) which do not pass the upper region of the diagonal $y = x$.

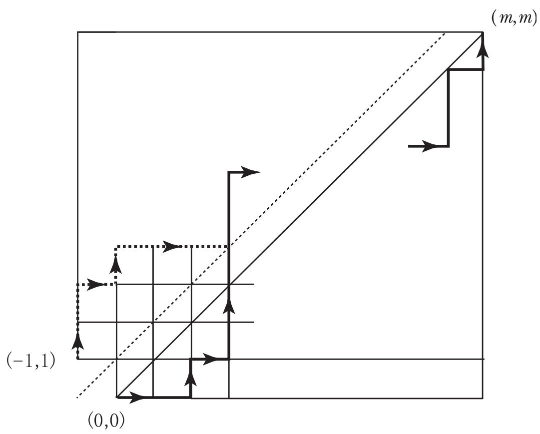
Theorem 6.1.3 (Catalan number)

$$C_n = \frac{(2n)!}{(n+1)!n!}, \qquad n = 0, 1, 2, \dots,$$

Proof For $n = 0$ it is apparent by the definition $0! = 1$. Suppose $n \geq 1$. We see from Fig. 6.1 that

$$C_n = \binom{2n}{n} - \binom{2n}{n+1} = \frac{(2n)!}{n!(n+1)!},$$

as desired. ∎



An alternative representation of the Catalan paths: Consider in the xy-plane a polygonal line connecting the vertices:

$$(0,0),\ (1,\epsilon_1),\ (2,\epsilon_1+\epsilon_2),\ \dots,\ (n-1,\epsilon_1+\epsilon_2+\cdots+\epsilon_{n-1}),\ (n,\epsilon_1+\epsilon_2+\cdots+\epsilon_{n-1}+\epsilon_n)$$

in order. Then, there is a one-to-one correspondence between the Catalan paths of length $2n$ and the sample paths of a random walk starting 0 at time 0 and returning 0 at time $2n$ staying always in the half line $[0,+\infty)$. Therefore,

Lemma 6.1.4 Let $n \ge 1$. The number of sample paths of a random walk starting 0 at time 0 and returning 0 at time $2n$ staying always in the half line $[0,+\infty)$ is the Catalan number C_n.

Let $\{X_n\}$ be a random walk on $\mathbb{Z}$ with right-move probability p and left-move probability $q = 1-p$. We assume that the random walk starts at the origin, i.e., $X_0 = 0$. Since the random walker returns to the origin only after even steps, the return probability is given by

$$R = P\left(\bigcup_{n=1}^{\infty}\{X_{2n}=0\}\right). \tag{6.1}$$

It is important to note that $\bigcup_{n=1}^{\infty}\{X_{2n}=0\}$ is not the sum of disjoint events.

Let p_{2n} be the probability that the random walker is found at the origin at time $2n$, that is,

$$p_{2n} = P(X_{2n}=0) = \binom{2n}{n}p^nq^n = \frac{(2n)!}{n!n!}\,p^nq^n, \qquad n=1,2,\dots. \tag{6.2}$$

For convenience set

$$p_0 = 1.$$

Note that the right hand side of (6.1) is not the sum of p_{2n}. Instead, we need to consider the probability that the random walker returns to the origin after $2n$ steps but not before:

$$q_{2n} = P(X_2 \neq 0,\ X_4 \neq 0,\ \dots, X_{2n-2} \neq 0,\ X_{2n}=0) \qquad n=1,2,\dots.$$

Notice the essential difference between p_{2n} and q_{2n}.

Definition 6.1.5 We set

$$T = \inf\{n \ge 1\,;\, X_n = 0\}, \tag{6.3}$$

where $T = +\infty$ for $\{n \ge 1\,;\, X_n = 0\} = \emptyset$. We call T the *first hitting time* to the origin. (Strictly speaking, T is not a random variable according to our definition in Chapter 1. It is, however, commonly accepted that a random variable takes values in $(-\infty,+\infty)\cup\{\pm\infty\}$.)

By definition we have

$$P(T = 2n) = q_{2n} \tag{6.4}$$

and therefore, the return probability is given by

$$R = P(T < \infty) = \sum_{n=1}^{\infty} q_{2n}. \tag{6.5}$$

We will calculate $P(T = 2n)$ and thereby R.

Theorem 6.1.6 Let $\{X_n\}$ be the random walk starting from 0 with right-move probability p and left-move probability q. Let T be the first hitting-time to 0. Then

$$q_{2n} = P(T = 2n) = 2C_{n-1}(pq)^n, \qquad n = 1, 2, \dots.$$

Proof Obviously, we have

$$\begin{aligned} q_{2n} &= P(X_2 \neq 0,\ X_4 \neq 0,\ \dots, X_{2n-2} \neq 0,\ X_{2n} = 0) \\ &= P(X_1 > 0,\ X_2 > 0,\ X_3 > 0,\ \dots, X_{2n-2} > 0,\ X_{2n-1} > 0,\ X_{2n} = 0) \\ &\quad + P(X_1 < 0,\ X_2 < 0,\ X_3 < 0,\ \dots, X_{2n-2} < 0,\ X_{2n-1} < 0,\ X_{2n} = 0). \end{aligned}$$

In view of Fig. 6.1 we see that

$$P(X_1 > 0,\ X_2 > 0,\ X_3 > 0,\ \dots, X_{2n-2} > 0,\ X_{2n-1} > 0,\ X_{2n} = 0) = p \times C_{n-1}(pq)^{n-1} \times q.$$

Then the result is immediate. ∎

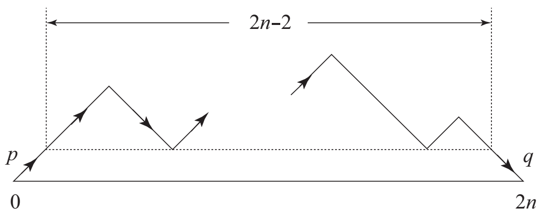


Figure 6.1: Calculating $P(X_1 > 0,\ X_2 > 0,\ \dots, X_{2n-1} > 0,\ X_{2n} = 0)$

Remark 6.1.7 There are some noticeable relations between $\{p_{2n}\}$ and $\{q_{2n}\}$.

$$\begin{aligned} q_{2n} &= \frac{2pq}{n}\, p_{2n-2}\,, \qquad n \geq 1, \\ q_{2n} &= 4pq p_{2n-2} - p_{2n}\,, \qquad n \geq 1. \end{aligned}$$

Lemma 6.1.8 The generating function of the Catalan numbers C_n is given by

$$f(z) = \sum_{n=0}^{\infty} C_n z^n = \frac{1 - \sqrt{1-4z}}{2z}\,. \tag{6.6}$$

Proof Problem 17. ∎

Theorem 6.1.9 Let R be the probability that a random walker starting from the origin returns to the origin in finite time. Then we have

$$R = 1 - |p - q|.$$

Proof We know from Theorem 6.1.6 that the return probability R is given by

$$R = \sum_{n=1}^{\infty} P(T = 2n) = \sum_{n=1}^{\infty} 2C_{n-1}(pq)^n.$$

Using the generating function of the Catalan numbers in Lemma 6.1.8, we obtain

$$R = 2pq \sum_{n=0}^{\infty} C_n(pq)^n = 2pq \times \frac{1 - \sqrt{1-4pq}}{2pq} = 1 - \sqrt{1-4pq}\,.$$

Since $p + q = 1$ we have

$$\sqrt{1-4pq} = \sqrt{(p+q)^2 - 4pq} = \sqrt{(p-q)^2} = |p-q|,$$

which completes the proof. ∎

Definition 6.1.10 A random walk is called *recurrent* if $R = 1$, otherwise it is called *transient*.

Theorem 6.1.11 The one-dimensional random walk is recurrent if and only if $p = q = 1/2$ (isotropic). It is transient if and only if $p \neq q$.

When a random walk is recurrent, it is meaningful to consider the mean recurrent time defined by

$$\mathbf{E}(T) = \sum_{n=1}^{\infty} 2nP(T = 2n) = \sum_{n=1}^{\infty} 2nq_{2n},$$

where T is the first hitting time to the origin.

Theorem 6.1.12 (Null recurrence) The mean recurrent time of the isotropic, one-dimensional random walk is infinity: $\mathbf{E}[T] = +\infty$.

Proof In view of Theorem 6.1.6, setting $p = q = 1/2$, we obtain

$$\mathbf{E}(T) = 4\sum_{n=1}^{\infty} nC_{n-1}\Big(\frac{1}{4}\Big)^n = \sum_{n=0}^{\infty}(n+1)C_n\Big(\frac{1}{4}\Big)^n. \tag{6.7}$$

On the other hand, the generating function of the Catalan numbers is given by

$$f(z) = \sum_{n=0}^{\infty} C_n z^n = \frac{1-\sqrt{1-4z}}{2z}\,.$$

Then

$$2zf(z) = 2\sum_{n=0}^{\infty} C_n z^{n+1} = 1 - \sqrt{1-4z}$$

and differentiating by z, we have

$$2(zf(z))' = 2\sum_{n=0}^{\infty}(n+1)C_n z^n = \frac{2}{\sqrt{1-4z}}\,.$$

Letting $z \to 1/4$ we have

$$2\sum_{n=0}^{\infty}(n+1)C_n\Big(\frac{1}{4}\Big)^n = \lim_{z\uparrow 1/4} = +\infty,$$

and hence $\mathbf{E}[T] = +\infty$ as desired, see also Remark 6.1.13. ∎

Remark 6.1.13 Let $a_n \geq 0$ for $n = 0, 1, 2, \dots$ and consider the power series:

$$f(x) = \sum_{n=0}^{\infty} a_n x^n .$$

If the radius of convergence of the above power series is ≥ 1, we have

$$\lim_{x \to 1-0} f(x) = \sum_{n=0}^{\infty} a_n$$

including the case of $\infty = \infty$. The verification is by elementary calculus based on the following two inequalities:

$$\liminf_{x \to 1-0} f(x) \geq \sum_{n=0}^{N} a_n , \qquad N \geq 1,$$

$$f(x) \leq \sum_{n=0}^{\infty} a_n , \quad x < 1.$$

Problem 15 Find the Catalan numbers C_n in the following steps.

(1) Prove that $C_n = \sum_{k=1}^{n} C_{k-1} C_{n-k}$ by using graphical expressions.

(2) Using (1), prove that the generating function of the Catalan numbers $f(z) = \sum_{n=0}^{\infty} C_n z^n$ verifies

$$f(z) - 1 = z\{f(z)\}^2.$$

(3) Find $f(z)$.

(4) Using Taylor expansion of $f(z)$ obtained in (3), find C_n.

Problem 16 Let $\{X_n\}$ be a random walk starting from 0 with right-move p and left-move q. Show that

$$\begin{aligned} &P(X_1 \geq 0, X_2 \geq 0, \dots, X_{2n-1} \geq 0) \\ &\quad = P(X_1 \geq 0, X_2 \geq 0, \dots, X_{2n} \geq 0) = 1 - q \sum_{k=0}^{n-1} C_k (pq)^k \end{aligned}$$

for $n = 1, 2, \dots$, where C_k is the Catalan number. Using this result, show next that

$$P(X_n \geq 0 \text{ for all } n \geq 1) = \begin{cases} 1 - \dfrac{q}{p}, & p > q, \\ 0, & p \leq q. \end{cases}$$

Problem 17 (Lemma 6.1.8) (1) Using the well-known formula for binomial expansion:

$$(1 + x)^{\alpha} = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n, \qquad |x| < 1,$$

prove that

$$\sum_{n=0}^{\infty} \binom{2n}{n} z^n = \frac{1}{\sqrt{1 - 4z}} , \qquad |z| < \frac{1}{4}.$$

(2) Let C_n be the Catalan number given by

$$C_n = \frac{(2n)!}{n!(n+1)!} , \qquad n = 0, 1, 2, \dots .$$

Prove that

$$\sum_{n=0}^{\infty} C_n z^n = \frac{1 - \sqrt{1 - 4z}}{2z} , \qquad |z| < \frac{1}{4}.$$

Problem 18 In the $m \times (m+n)$ grid consider a shortest path connecting $(0,0)$ and $(m+n, m)$ which do not pass the region upper than the line connecting $(0,0)$ and (m,m). Show that the number of such paths is given by

$$\frac{(2m+n)!(n+1)}{m!(m+n+1)!}.$$

6.2 The Law of Long Lead

Let us consider an isotropic random walk $\{X_n\}$, namely, letting $\{Z_n\}$ be the Bernoulli trials such that

$$P(Z_n = +1) = P(Z_n = -1) = \frac{1}{2},$$

we set

$$X_0 = 0, \qquad X_n = \sum_{k=1}^{n} Z_k .$$

Fig. 6.2 shows sample paths of $X_0, X_1, X_2, \dots, X_{10000}$. We notice that these are just two examples among many different patterns.

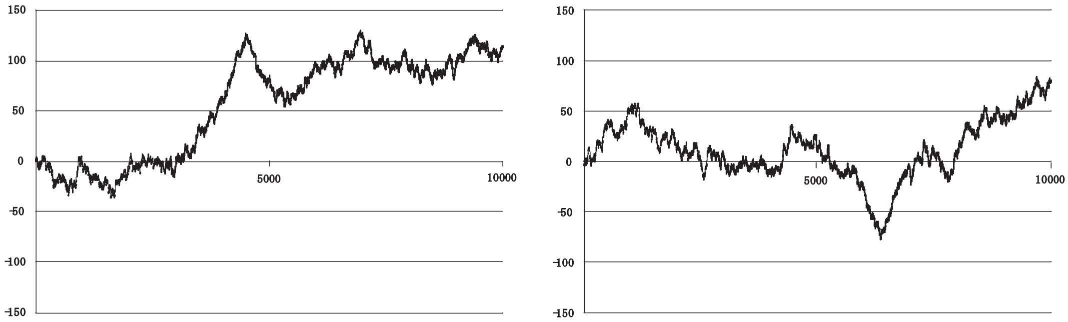


Figure 6.2: Sample paths of a random walk up to time 10000

By the law of large numbers we know that ± 1 occur almost 5000 times during 10000 coin toss. In fact, it follows from the de Moivre–Laplace theorem (more generally, central limit theorem) that

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k$$

obeys $N(0,1)$ in the limit as $n \to \infty$. Hence for a large n we have approximately $X_n \sim N(0,n)$. For example, $P(|X_n| \le 2\sqrt{n}) \approx 95.4\%$ for a large n, so X_{10000} lies in the interval ± 200 at probability 95.4%. In this case ± 1 occurs 5000 ± 100 times. In other words, along the polygonal line the up-move and down-move occur almost the same times, however, the polygonal line stays more often in the upper or lower half region.

We say that a random walk stays in the positive region in the time interval $[i, i+1]$ if $X_i \ge 0$ and $X_{i+1} \ge 0$. Similarly, we say that a random walk stays in the negative region in the time interval $[i, i+1]$ if $X_i \le 0$ and $X_{i+1} \le 0$. Let

$$W(2k|2n), \qquad n = 1, 2, \dots, \quad k = 0, 1, \dots, n,$$

be the probability that the total time of the random walk staying in the positive region during $[0, 2n]$ is $2k$.

Remind that in this section we only consider an isotropic random walk ($p = q = 1/2$). For $n = 1$ we have

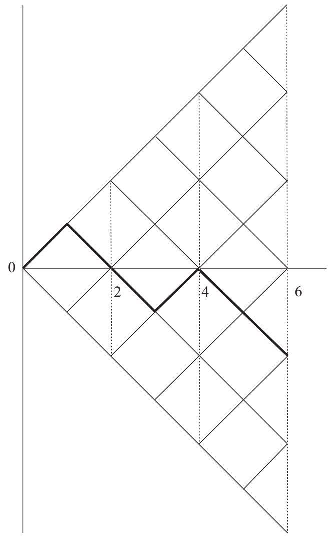
$$W(2|2) = 2 \times \left(\frac{1}{2}\right)^2 = \frac{1}{2}, \qquad W(0|2) = 2 \times \left(\frac{1}{2}\right)^2 = \frac{1}{2}.$$

Similarly, we have

$$W(4|4) = 6 \times \left(\frac{1}{2}\right)^4, \qquad W(2|4) = 4 \times \left(\frac{1}{2}\right)^4, \qquad W(0|4) = 6 \times \left(\frac{1}{2}\right)^4,$$

$$W(6|6) = 20 \times \left(\frac{1}{2}\right)^6, \qquad W(4|6) = 12 \times \left(\frac{1}{2}\right)^6, \qquad W(2|6) = 12 \times \left(\frac{1}{2}\right)^6, \qquad W(0|6) = 20 \times \left(\frac{1}{2}\right)^6,$$

For general $W(2k|2n)$ we have the following somehow surprisingly simple result.



Theorem 6.2.1 For $n = 1, 2, \dots$ it holds that

$$W(2k|2n) = \binom{2k}{k}\binom{2n-2k}{n-k}\left(\frac{1}{2}\right)^{2n}, \qquad k = 0, 1, \dots, n, \tag{6.8}$$

or equivalently,

$$W(2k|2n) = p_{2k}p_{2n-2k}\,. \tag{6.9}$$

That (6.8) and (6.9) are equivalent follows immediately from

$$p_{2n} \equiv P(X_{2n} = 0) = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}, \quad n = 0, 1, 2, \dots.$$

The proof of Theorem 6.2.1 is, however, not so simple as is expected by a nice expression (6.9). One would expect a very tricky simple observation leading the result. The complete proof is found in [Feller]. As before, we set

$$q_{2n} = P(T = 2n), \qquad n = 1, 2, \dots.$$

Observing that

$$q_{2n} = 2P(X_1 > 0, X_2 > 0, \dots, X_{2n-1} > 0, X_{2n} = 0) = 2C_{n-1}\left(\frac{1}{2}\right)^{2n} = \frac{1}{2n}\,p_{2n-2}\,,$$

one can get an obvious relation:

$$W(2k|2n) = \sum_{r=1}^{k}\frac{q_{2r}}{2}\,W(2k-2r|2n-2r) + \sum_{r=1}^{n-k}\frac{q_{2r}}{2}\,W(2k|2n-2r).$$

The assertion is then proved by induction on k, n.

We will find a good approximation for $W(2k|2n)$ when $n \to \infty$. For a fixed n let H_{2n} be the total time that the random walker stays in the positive region up to time $2n$. It is convenient to consider the ratio $\dfrac{H_{2n}}{2n}$ rather than H_{2n} it self. As we have already obtained

$$P(H_{2n} = 2k) = W(2k|2n) = \binom{2k}{k}\binom{2n-2k}{n-k}\left(\frac{1}{2}\right)^{2n},$$

for $0 < a < 1$ we see that

$$P\Big(a \le \frac{H_{2n}}{2n} \le b\Big) = \sum_{k=an}^{bn} W(2k|2n)$$

$$= \sum_{k=0}^{n} \chi_{[an,bn]}(k) W(2k|2n) = \sum_{k=0}^{n} \chi_{[a,b]}\Big(\frac{k}{n}\Big)\binom{2k}{k}\binom{2n-2k}{n-k}\Big(\frac{1}{2}\Big)^{2n},$$

where $\chi_I(x)$ is the indicator function of an interval I, that is, takes 1 for $x \in I$ and 0 otherwise. Using the Stirling formula:

$$n! \sim \sqrt{2\pi n}\Big(\frac{n}{e}\Big)^n, \qquad \text{as } n \to \infty,$$

we obtain

$$\binom{2k}{k}\Big(\frac{1}{2}\Big)^{2k} \sim \frac{1}{\sqrt{\pi k}}\,.$$

Then,

$$P\Big(a \le \frac{H_{2n}}{2n} \le b\Big) \sim \sum_{k=0}^{n} \chi_{[a,b]}\Big(\frac{k}{n}\Big)\frac{1}{\pi\sqrt{k(n-k)}} = \sum_{k=0}^{n} \chi_{[a,b]}\Big(\frac{k}{n}\Big)\frac{1}{\pi\sqrt{\frac{k}{n}(1-\frac{k}{n})}}\,\frac{1}{n}$$

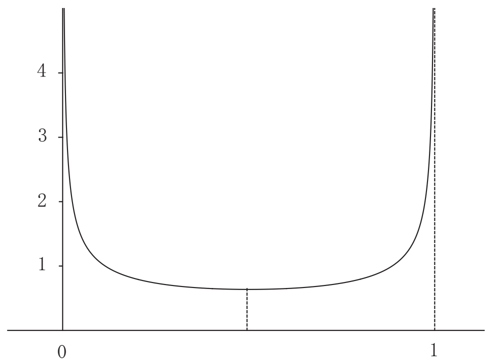
$$\to \int_0^1 \chi_{[a,b]}(x)\frac{dx}{\pi\sqrt{x(1-x)}} = \int_a^b \frac{dx}{\pi\sqrt{x(1-x)}}\,.$$

Definition 6.2.2 The probability distribution defined by the density function:

$$\frac{dx}{\pi\sqrt{x(1-x)}}\,. \qquad 0 < x < 1,$$

is called the *arcsine law*. The distribution function is given by

$$F(x) = \int_0^x \frac{dt}{\pi\sqrt{t(1-t)}} = \frac{2}{\pi}\arcsin\sqrt{x} = \frac{1}{2} + \frac{1}{\pi}\arcsin(2x-1).$$



Theorem 6.2.3 The distribution of $\dfrac{H_{2n}}{2n}$ converges weakly to the arcsine law:

$$\lim_{n\to\infty} P\Big(a \le \frac{H_{2n}}{2n} \le b\Big) = \int_a^b \frac{dx}{\pi\sqrt{x(1-x)}}\,, \qquad 0 \le a < b \le 1.$$

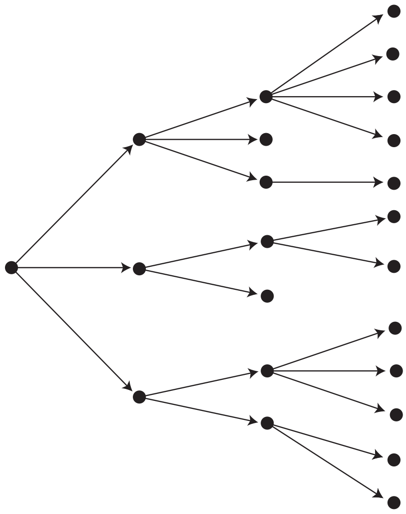
For example,

$$F(0.9) = \frac{2}{\pi}\arcsin\sqrt{0.9} \approx 0.795.$$

Namely, during many ganes the probability that the ratio of winning time (or happy time) exceeds 90% is $1 - F(0.9) \approx 0.205$, which would be much larger than one expects by intuition.

7 Galton-Watson Branching Processes

Consider a simplified family tree where each individual gives birth to offspring (children) and dies. The number of offsprings is random. We are interested in whether the family survives or not. A fundamental model was proposed by F. Galton in 1873 and basic properties were derived by Galton and Watson in their joint paper in the next year. The name “Galton-Watson branching process” is quite common in literatures after their paper, but it would be more fair to refer to it as “BGW process.” In fact, Irénée-Jules Bienaymé studied the same model independently already in 1845.



References

1. R. B. シナジ（今野紀雄・林俊一訳): マルコフ連鎖から格子確率モデルへ, シュプリンガー東京, 2001.
2. K. B. Athreya and P. E. Ney: Branching Processes, Dover 2004 (original version, Springer 1972)

7.1 Definition

Let X_n be the number of individuals of the nth generation. Then $\{X_n\,;\, n = 0, 1, 2, \dots\}$ becomes a discrete-time stochasic process. We assume that the number of children born from each individual obeys a common probability distribution and is independent of individuals and of generation. Under this assupmtion $\{X_n\}$ becomes a Markov chain.

Let us obtain the transition probability. Let Y be the number of children born from an individual and set

$$P(Y = k) = p_k\,, \qquad k = 0, 1, 2, \dots.$$

The sequence $\{p_0, p_1, p_2, \dots\}$ describes the distribution of the number of children born from an individual. In fact, what we need is the condition

$$p_k \ge 0, \qquad \sum_{k=0}^{\infty} p_k = 1.$$

We refer to $\{p_0, p_1, \dots\}$ as the *offspring distribution*. Let $Y_1, Y_2, \dots$ be independent identically distributed random variables, of which the distribution is the same as Y. Then, we define the transition probability by

$$p(i, j) = P(X_{n+1} = j | X_n = i) = P\left(\sum_{k=1}^{i} Y_k = j\right), \qquad i \ge 1, \quad j \ge 0,$$

and

$$p(0, j) = \begin{cases} 0, & j \ge 1, \\ 1, & j = 0. \end{cases}$$

Clearly, the state 0 is an absorbing one. The above Markov chain $\{X_n\}$ over the state space $\{0, 1, 2, \dots\}$ is called the *Galton-Watson branching process* with offspring distribution $\{p_k\,;\, k = 0, 1, 2, \dots\}$.

For simplicity we assume that $X_0 = 1$. When $p_0 + p_1 = 1$, the family tree is reduced to just a path without branching so the situation is much simpler (Problem 19). We will focus on the case where

$$p_0 + p_1 < 1, \quad p_2 < 1, \quad \dots, \quad p_k < 1, \quad \dots.$$

In the next section on we will always assume the above conditions.

Problem 19 (One-child policy) Consider the Galton-Watson branching process with offspring distribution satisfying $p_0 + p_1 = 1$. Calculate the probabilities

$$q_1 = P(X_1 = 0), \quad q_2 = P(X_1 \neq 0, X_2 = 0), \quad \dots, \quad q_n = P(X_1 \neq 0, \dots, X_{n-1} \neq 0, X_n = 0), \quad \dots$$

and find the extinction probability

$$P = \left(\bigcup_{n=1}^{\infty}\{X_n = 0\}\right) = P(X_n = 0 \text{ occurs for some } n \geq 1).$$

7.2 Generating Functions

Let $\{X_n\}$ be the Galton-Watson branching process with offspring distribution $\{p_k\,;\, k = 0, 1, 2, \dots\}$. Let $p(i, j) = P(X_{n+1} = j | X_n = i)$ be the transition probability. We assume that $X_0 = 1$.

Define the generating function of the offspring distribution by

$$f(s) = \sum_{k=0}^{\infty} p_k s^k. \tag{7.1}$$

The series in the right-hand side converges for $|s| \leq 1$. We set

$$f_0(s) = s, \qquad f_1(s) = f(s), \qquad f_n(s) = f(f_{n-1}(s)).$$

Lemma 7.2.1

$$\sum_{j=0}^{\infty} p(i, j) s^j = [f(s)]^i, \qquad i = 1, 2, \dots. \tag{7.2}$$

Proof By definition,

$$p(i, j) = P(Y_1 + \cdots + Y_i = j) = \sum_{\substack{k_1+\cdots+k_i=j \\ k_1\geq 0,\dots,k_i\geq 0}} P(Y_1 = k_1, \dots, Y_i = k_i).$$

Since $Y_1, \dots, Y_i$ are independent, we have

$$p(i, j) = \sum_{\substack{k_1+\cdots+k_i=j \\ k_1\geq 0,\dots,k_i\geq 0}} P(Y_1 = k_1) \cdots P(Y_i = k_i) = \sum_{\substack{k_1+\cdots+k_i=j \\ k_1\geq 0,\dots,k_i\geq 0}} p_{k_1} \cdots p_{k_i}.$$

Hence,

$$\begin{aligned}
\sum_{j=0}^{\infty} p(i, j) s^j &= \sum_{j=0}^{\infty} \sum_{\substack{k_1+\cdots+k_i=j \\ k_1\geq 0,\dots,k_i\geq 0}} p_{k_1} \cdots p_{k_i} s^j \\
&= \sum_{k_1=0}^{\infty} p_{k_1} s^{k_1} \cdots \sum_{k_i=0}^{\infty} p_{k_i} s^{k_i} \\
&= [f(s)]^i,
\end{aligned}$$

which proves the assertion. ∎

Lemma 7.2.2 Let $p_n(i, j)$ be the n-step transition probability of the Galton-Watson branching process. We have

$$\sum_{j=0}^{\infty} p_n(i, j)s^j = [f_n(s)]^i, \qquad i = 1, 2, \dots. \tag{7.3}$$

Proof We prove the assertion by induction on n. First note that $p_1(i, j) = p(i, j)$ and $f_1(s) = f(s)$ by definition. For $n = 1$ we need to show that

$$\sum_{j=0}^{\infty} p(i, j)s^j = [f(s)]^i, \qquad i = 1, 2, \dots, \tag{7.4}$$

Which was shown in Lemma 7.2.1. Suppose that $n \geq 1$ and the claim (7.3) is valid up to n. Using the Chapman-Kolmogorov identity, we see that

$$\sum_{j=0}^{\infty} p_{n+1}(i, j)s^j = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} p(i, k)p_n(k, j)s^j .$$

Since

$$\sum_{j=0}^{\infty} p_n(k, j)s^j = [f_n(s)]^k$$

by assumption of induction, we obtain

$$\sum_{j=0}^{\infty} p_{n+1}(i, j)s^j = \sum_{k=0}^{\infty} p(i, k)[f_n(s)]^k.$$

The right-hand side coincides with (7.4) where s is replaced by $f_n(s)$. Consequently, we come to

$$\sum_{j=0}^{\infty} p_{n+1}(i, j)s^j = [f(f_n(s))]^i = [f_{n+1}(s)]^i,$$

which proves the claim for $n + 1$. ∎

Since $X_0 = 1$,

$$P(X_n = j) = P(X_n = j|X_0 = 1) = p_n(1, j).$$

In particular,

$$P(X_1 = j) = P(X_1 = j|X_0 = 1) = p_1(1, j) = p(1, j) = p_j .$$

Theorem 7.2.3 Assume that the mean value of the offspring distribution is finite:

$$m = \sum_{k=0}^{\infty} kp_k < \infty.$$

Then we have

$$\mathbf{E}[X_n] = m^n.$$

Proof Differentiating (7.1), we obtain

$$f'(s) = \sum_{k=0}^{\infty} kp_k s^{k-1}, \qquad |s| < 1. \tag{7.5}$$

Letting $s \to 1 - 0$, we have

$$\lim_{s\to 1-0} f'(s) = m.$$

On the other hand, setting $i = 1$ in (7.3), we have

$$\sum_{j=0}^{\infty} p_n(1, j)s^j = f_n(s) = f_{n-1}(f(s)). \tag{7.6}$$

Differentiating both sides, we come to

$$f_n'(s) = \sum_{j=0}^{\infty} j p_n(1, j) s^{j-1} = f_{n-1}'(f(s)) f'(s). \tag{7.7}$$

Letting $s \to 1 - 0$, we have

$$\lim_{s\to 1-0} f_n'(s) = \sum_{j=0}^{\infty} j p_n(1, j) = \lim_{s\to 1-0} f_{n-1}'(f(s)) \lim_{s\to 1-0} f'(s) = m \lim_{s\to 1-0} f_{n-1}'(s).$$

Therefore,

$$\lim_{s\to 1-0} f_n'(s) = m^n,$$

which means that

$$\mathbf{E}(X_n) = \sum_{j=0}^{\infty} jP(X_n = j) = \sum_{j=0}^{\infty} j p_n(1, j) = m^n.$$

∎

In conclusion, the mean value of the number of individuals in the nth generation, $\mathbf{E}(X_n)$, decreases and converges to 0 if $m < 1$ and diverges to the infinity if $m > 1$, as $n \to \infty$. It stays at a constant if $m = 1$. We are thus suggested that extinction of the family occurs when $m < 1$.

Problem 20 Assume that the variance of the offspring distribution is finite: $\mathbf{V}[Y] = \sigma^2 < \infty$. By similar argument as in Theorem 7.2.3, prove that

$$\mathbf{V}[X_n] = \begin{cases} \dfrac{\sigma^2 m^{n-1}(m^n - 1)}{m - 1}, & m \neq 1, \\ n\sigma^2, & m = 1. \end{cases}$$

7.3 Extinction Probability

The event $\{X_n = 0\}$ means that the family died out until the nth generation. So

$$q = P\left(\bigcup_{n=1}^{\infty} \{X_n = 0\}\right)$$

is the probability of extinction of the family. Note that the events in the right-hand side is not mutually exclusive but

$$\{X_1 = 0\} \subset \{X_2 = 0\} \subset \cdots \subset \{X_n = 0\} \subset \dots.$$

Therefore, it holds that

$$q = \lim_{n\to\infty} P(X_n = 0). \tag{7.8}$$

If $q = 1$, this family almost surely dies out in some generation. If $q < 1$, the survival probability is positive $1 - q > 0$. We are interested in whether $q = 1$ or not.

Lemma 7.3.1 Let $f(s)$ be the generating function of the offspring distribution, and set $f_n(s) = f(f_{n-1}(s))$ as before. Then we have

$$q = \lim_{n\to\infty} f_n(0).$$

Therefore, q satisfies the equation:

$$q = f(q). \tag{7.9}$$

Proof It follows from Lemma 7.2.2 that

$$f_n(s) = \sum_{j=0}^{\infty} p_n(1, j)s^j.$$

Hence,

$$f_n(0) = p_n(1, 0) = P(X_n = 0|X_0 = 1) = P(X_n = 0),$$

where the last identity is by the assumption of $X_0 = 1$. The assertion is now straightforward by combining (7.8). The second assertion follows since $f(s)$ is a continuous function on $[0, 1]$. ∎

Lemma 7.3.2 Assume that the offspring distribution satisfies the conditions:

$$p_0 + p_1 < 1, \quad p_2 < 1, \quad \dots, \quad p_k < 1, \quad \dots.$$

Then the generating function $f(t)$ verifies the following properties.

(1) $f(s)$ is increasing, i.e., $f(s_1) \le f(s_2)$ for $0 \le s_1 \le s_2 \le 1$.

(2) $f(s)$ is strictly convex, i.e., if $0 \le s_1 < s_2 \le 1$ and $0 < \theta < 1$ we have

$$f(\theta s_1 + (1-\theta)s_2) < \theta f(s_1) + (1-\theta)f(s_2).$$

Proof (1) is apparent since the coefficient of the power series $f(s)$ is non-negative. (2) follows by $f''(s) > 0$. ∎

Lemma 7.3.3 (1) If $m \le 1$, we have $f(s) > s$ for $0 \le s < 1$.

(2) If $m > 1$, there exists a unique s such that $0 \le s < 1$ and $f(s) = s$.

Lemma 7.3.4 $f_1(0) \le f_2(0) \le \dots \to q$.

Theorem 7.3.5 The extinction probability q of the Galton-Watson branching process as above coincides with the smallest s such that

$$s = f(s), \qquad 0 \le s \le 1.$$

Moreover, if $m \le 1$ we have $q = 1$, and if $m > 1$ we have $q < 1$.

The Galton-Watson branching process is called *subcritical*, *critical* and *supercritical* if $m < 1$, $m = 1$ and $m > 1$, respectively. The survival is determined only by the mean value m of the offspring distribution. The situation changes dramatically at $m = 1$ and, following the terminology of statistical physics, we call it *phase transition*.

Problem 21 Let b, p be constant numbers such that $b > 0$, $0 < p < 1$ and $b + p < 1$. For the offspring distribution given by

$$p_k = bp^{k-1}, \qquad k = 1, 2, \dots,$$
$$p_0 = 1 - \sum_{k=1}^{\infty} p_k,$$

find the generating function $f(s)$. Moreover, setting $m = 1$, find $f_n(s)$.

8 Poisson Processes

Let $T \subset \mathbf{R}$ be an interval. A family of random variables $\{X(t)\,;\, t \in T\}$ is called a *continuous time stochastic process*. We often consider $T = [0,1]$ and $T = [0,\infty)$. As $X(t)$ is a random variable for each $t \in T$, it has another variable $\omega \in \Omega$. When we need to explicitly refer to ω, we write $X(t,\omega)$ or $X_t(\omega)$. For fixed $\omega \in \Omega$, the function

$$t \mapsto X(t,\omega)$$

is called a *sample path* of the stochastic process $\{X(t)\}$. It is the central idea of stochastic processes that a random evolution in the real world is expressed by a single sample path selected randomly from all the possible sample paths.

The most fundamental continuous time stochastic processes are the Poisson process and the Brownian motion (Wiener process). In the recent study of mathematical physics and mathematical finance, a kind of composition of these two processes, called the Lévy process (or additive process), has received much attention.

8.1 Heuristic Introduction

Let us imagine that the number of objects changes as time goes on. The number at time t is modelled by a random variable X_t and we wish to construct a stochastic process $\{X_t\}$. In this case X_t takes values in $\{0,1,2,\dots\}$. In general, such a stochastic process is called a *counting process*.

There are many different variations of randomness and so wide variations of counting processes. We below consider the simple situation as follows: We focus an event E which occurs repeatedly at random as time goes on. For example,

(i) alert of receiving an e-mail;

(ii) telephone call received a call center;

(iii) passengers making a queue at a bus stop;

(iv) customers visiting a shop;

(v) occurrence of defect of a machine;

(vi) traffic accident at a corner;

(vii) radiation from an atom.

Let fix a time origin as $t = 0$. We count the number of occurrence of the event E during the time interval $[0,t]$ and denote it by X_t. Let $t_1, t_2, \dots$ be the time when E occurs, see Fig. 8.1.

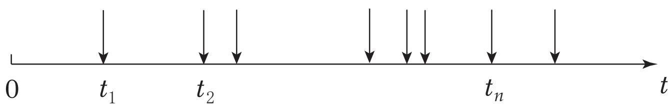


Figure 8.1: Recording when the event E occurs

There are two quantities which we measure.

(i) The number of occurrence of E up to time t, say, X_t. Then $\{X_t\,;\, t \ge 0\}$ becomes a counting process.

(ii) The waiting time of the nth occurrence after the $(n-1)$th occurrence, say, T_n. Here T_1 is defined to be the waiting time of the first occurrence of E after starting the observation. Then $\{T_n\,;\, n = 1,2,\dots\}$ is a sequence of random variables taking values in $[0,\infty)$.

We will introduce heuristically a stochastic process $\{X_t\}$ from the viewpoint of (i). It is convenient to start with discrete time approximation. Fix $t > 0$ and divide the time interval $[0,t]$ into n small intervals. Let

$$\Delta t = \frac{t}{n}$$

be the length of the small intervals and number from the time origin in order.

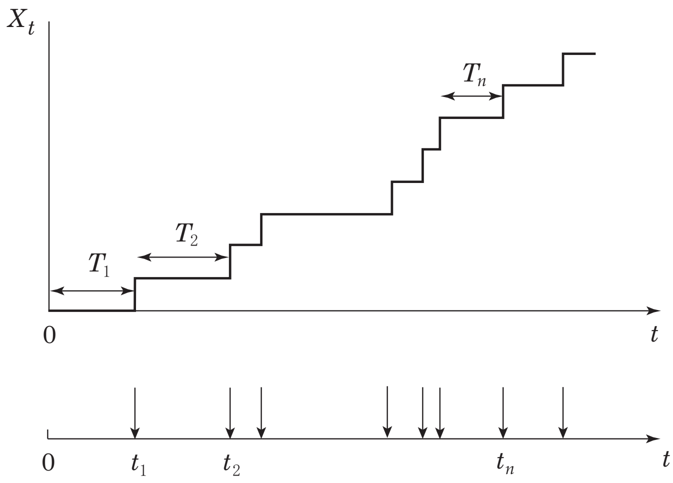
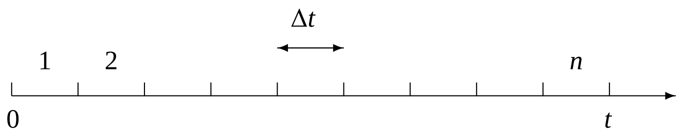


Figure 8.2: The counting process and waiting times



We assume the following conditions on the occurrence of the event E:

(1) There exists a constant $\lambda > 0$ such that

$$
\begin{aligned}
&P(E \text{ occurs just once in a small time interval of length } \Delta t) = \lambda\Delta t + o(\Delta t),\\
&P(E \text{ does not occur in a mall time interval of length } \Delta t) = 1 - \lambda\Delta t + o(\Delta t),\\
&P(E \text{ occurs more than once in a small time interval of length } \Delta t) = o(\Delta t).
\end{aligned}
$$

(2) Occurrence of E in disjoint time intervals is independent.

Some more accounts. Let us imagine the alert of receiving an e-mail. That

$$P(E \text{ occurs more than once in a small time interval of length } \Delta t) = o(\Delta t)$$

means that two occurrences of the event E is always separated. That

$$P(E \text{ occurs just once in a small time interval of length } \Delta t) = \lambda\Delta t + o(\Delta t)$$

means that when Δt is small the probability of occurrence of E in a time interval is proportional to the length of the time interval. We understand from (2) that occurrence of E is independent of the past occurrence.

Let Z_i denote the number of occurrence of the event E in the ith time interval. Then $Z_1, Z_2, \dots, Z_n$ become a sequence of independent random variables with an identical distribution such that

$$P(Z_i = 0) = 1 - \lambda\Delta t + o(\Delta t), \qquad P(Z_i = 1) = \lambda\Delta t + o(\Delta t), \qquad P(Z_i \geq 2) = o(\Delta t).$$

The number of occurrence of E during the time interval $[0, t]$ is given by

$$\sum_{i=1}^{n} Z_i\,.$$

The length Δt is introduced for a technical reason and is not essential in the probability model so letting $\Delta t \to 0$ or equivalently $n \to \infty$, we define X_t by

$$X_t = \lim_{\Delta t \to 0} \sum_{i=1}^{n} Z_i \,. \tag{8.1}$$

Although the limit does require matyhematical justification, we obtain heuristically a continuous time stochastic process $\{X_t\}$, which gives the number of occurrence of the event E up to time t. This is called a *Poisson process* with parameter $\lambda > 0$. A Poisson process belongs to the calss of continuous time Marokov chains.

Theorem 8.1.1 A Poisson process $\{X_t\,;\, t \ge 0\}$ satisfies the following properties:

(1) (counting process) X_t takes vales in $\{0, 1, 2, \dots\}$;

(2) $X_0 = 0$;

(3) (monotone increasing) $X_s \le X_t$ for $0 \le s \le t$;

(4) (independent increment) if $0 \le t_1 < t_2 < \cdots < t_k$, then

$$X_{t_2} - X_{t_1}\,, \quad X_{t_3} - X_{t_2}\,, \quad \dots, \quad X_{t_k} - X_{t_{k-1}}\,,$$

are independent;

(5) (stationarity) for $0 \le s < t$ and $h \ge 0$, the distributions of $X_{t+h} - X_{s+h}$ and $X_t - X_s$ are identical;

(6) there exists a constant $\lambda > 0$ such that

$$P(X_h = 1) = \lambda h + o(h), \qquad P(X_h \ge 2) = o(h).$$

(7) In that case X_t obeys the Poisson distribution with parameter λt.

Proof (1) Since X_t obeys the Poisson distribution with parameter λt, it takes values in non-negative integers almost surely.

(2) Obvious by definition.

(3) Let $s = m\Delta t$, $t = n\Delta t$, $m < n$. Then we have obviously

$$X_s = \lim_{\Delta t \to 0} \sum_{i=1}^{m} Z_i \le \lim_{\Delta t \to 0} \sum_{i=1}^{n} Z_i = X_t \,.$$

(4) Suppose $t_1 = n_1\Delta t, \dots, t_k = n_k\Delta t$ with $n_1 < \cdots < n_k$. Then we have

$$X_{t_2} - X_{t_1} = \lim_{\Delta t \to 0} \sum_{i=1}^{n_2} Z_i - \lim_{\Delta t \to 0} \sum_{i=1}^{n_1} Z_i = \lim_{\Delta t \to 0} \sum_{i=n_1+1}^{n_2} Z_i \,.$$

In other words, $X_{t_2} - X_{t_1}$ is the sum of Z_i's corresponding to the small time intervals contained in $[t_2, t_1)$. Hence, $X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}}$ are the sums of Z_i's and there is no common Z_i appearing in the summands. Since $\{Z_i\}$ are independent, so are $X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}}$.

(5) Since $X_{t+h} - X_{s+h}$ and $X_t - X_s$ are defined from the sums of Z_i's and the numbers of the terms coincide. Therefore the distributions are the same.

(6) Recall that X_h obeys the Poisson distribution with parameter λh. Hence,

$$P(X_h = 0) = e^{-\lambda h} = 1 - \lambda h + \cdots = 1 - \lambda h + o(h),$$
$$P(X_h = 1) = \lambda h e^{-\lambda h} = \lambda h(1 - \lambda h + \dots) = \lambda h + o(h).$$

Therefore we have

$$P(X_h \ge 2) = 1 - P(X_h = 0) - P(X_h = 1) = o(h).$$

(7) We note that

$$P\left(\sum_{i=1}^{n} Z_i = k\right) = \binom{n}{k}(\lambda\Delta t)^k(1 - \lambda\Delta t)^{n-k} + o(\Delta t).$$

In view of $\Delta t = t/n$ we let n tend to the infinity and obtain

$$P(X_t = k) = \lim_{\Delta t \to 0} \frac{(\lambda t)^k}{k!} \frac{n(n-1)\dots(n-k+1)}{n^k} \left(1 - \frac{\lambda t}{n}\right)^{n-k} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

This proves the assertion. ∎

Remark 8.1.2 The essence of the above argument in (7) is the *Poisson's law of small numbers* which says that the binomial distribution $B(n, p)$ is approximated by Poisson distribution with parameter $\lambda = np$ when n is large and p is small. The following table shows the distributions of $B(100, 0.02)$ and the Poisson distribution with parameter $\lambda = 2$.

k	0	1	2	3	4	5	6	$\cdots$
Binomial	0.1326	0.2707	0.2734	0.1823	0.0902	0.0353	0.0114	$\cdots$
Poisson	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.0120	$\cdots$

Example 8.1.3 The average number of customers visiting a certain service gate is two per minute. Using the Poisson model, calculate the following probabilities.

(1) The probability that no customer visits during the first two minutes after the gate opens.

(2) The probability that no customer visits during a time interval of two minutes.

(3) The probability that no customer visits during the first two minutes after the gate opens and that two customers visit during the next one minute.

Let X_t be the number of visitors up to time t. By assumption $\{X_t\}$ is a Poisson process with parameter $\lambda = 2$.

(1) We need to calculate $P(X_2 = 0)$. Since X_2 obeys the Poisson distribution with parameter $2\lambda = 4$, we have

$$P(X_2 = 0) = \frac{4^0}{0!} e^{-4} \approx 0.018.$$

(2) Suppose that the time interval starts at t_0. Then the probability under discussion is given by $P(X_{t_0+2} - X_{t_0} = 0)$. By stationarity we have

$$P(X_{t_0+2} - X_{t_0} = 0) = P(X_2 - X_0 = 0) = P(X_2 = 0),$$

which coincides with (1).

(3) We need calculate the probability $P(X_2 = 0, X_3 - X_2 = 2)$. Since X_2 and $X_3 - X_2$ are independent,

$$P(X_2 = 0, X_3 - X_2 = 2) = P(X_2 = 0)P(X_3 - X_2 = 2).$$

By stationarity we have

$$= P(X_2 = 0)P(X_1 = 2) = \frac{4^0}{0!} e^{-4} \times \frac{2^2}{2!} e^{-2} \approx 0.00496.$$

Problem 22 Let $\{X_t\}$ be a Poisson process. Show that

$$P(X_s = k | X_t = n) = \binom{n}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}, \quad k = 0, 1, \dots, n,$$

for $0 < s < t$. Next give an intuitive explanation of the above formula.

Problem 23 The average number of arrivals of e-mails is 216 per one day. Using the Poisson model, calculate the following probabilities.

(1) The probability that no mail arrives during 10 minutes.

(2) The probability that 4 mails arrive during 30 minutes and 8 mails arrive during the next 30 minutes.

8.2 Waiting Time

Let $\{X_t\,;\, t \geq 0\}$ be a Poisson process with parameter λ. By definition $X_0 = 0$ and X_t increases by one as time passes. Recall that the Poisson process counts the number of events occurring up to time t. First we set

$$T_1 = \inf\{t \geq 0\,;\, X_t \geq 1\}. \tag{8.2}$$

This is the waiting time for the first occurrence of the event E. Let T_2 be the waiting time for the second occurrence of the event E after the first occurrence, i.e.,

$$T_2 = \inf\{t \geq 0\,;\, X_t \geq 2\} - T_1\,.$$

In a similar fashion, we set

$$T_n = \inf\{t \geq 0\,;\, X_t \geq n\} - T_{n-1}, \qquad n = 2, 3, \dots. \tag{8.3}$$

Theorem 8.2.1 Let $\{X_t\}$ be a Poisson process with parameter λ. Define the waiting time T_n by (8.2) and (8.3). Then, $\{T_n\,;\, n = 1, 2, \dots\}$ becomes a sequence of iid random variables, of which distribution is the exponential distribution with parameter λ. In particular, the waiting time for occurence of an event in the Poisson process obeys the exponential distribution with parameter λ.

Proof Set $t = n\Delta t$ and consider the approximation by refinement of the time interval. Recall that to each small time interval of length Δt a random variable Z_i is associated. Then we know that

$$P(T_1 > t) = \lim_{\Delta t \to 0} P(Z_1 = \cdots = Z_n = 0) = \lim_{\Delta t \to 0} (1 - \lambda \Delta t)^n = \lim_{\Delta t \to 0} \left(1 - \frac{\lambda t}{n}\right)^n = e^{-\lambda t}.$$

Therefore,

$$P(T_1 \leq t) = 1 - e^{-\lambda t} = \int_0^t \lambda e^{-\lambda s} ds,$$

which shows that T_1 obeys the exponential distribution with parameter λ. The distributions of $T_2, T_3, \dots$ are similar. ∎

Remark 8.2.2 Let $\{X_t\}$ be a Poisson process with parameter λ. We know that $\mathbf{E}(X_1) = \lambda$, which means the average number of occurrence of the event during the unit time interval. Hence, it is expected that the average waiting time between two occurrences is $1/\lambda$. Theorem 8.2.1 says that the waiting time obeys the exponential distribution with parameter λ so its mean value is $1/\lambda$. Thus, our rough consideration gives the correct answer.

Problem 24 Let $\{X_t\}$ be a Poisson process with parameter λ. The waiting time for n occurrence of the events is defined by $S_n = T_1 + T_2 + \cdots + T_n$, where T_n is given in Theorem 8.2.1. Calculate $P(S_2 \leq t)$ and find the probability density function of S_2. [In general, S_n obeys a gamma distribution.]

8.3 The Rigorous Definition of Poisson Processes

The "definition" of a Poisson process in (8.1) is intuitive and instructive for modeling random phenomena. However, strictly speaking, the argument is not sufficient to define a stochastic process $\{X_t\}$. For example, the probability space $(\Omega, \mathcal{F}, P)$ on which $\{X_t\}$ is defined is not at all clear.

We need to start with the waiting time $\{T_n\}$. First we prepare a sequence of iid random variables $\{T_n\,;\, n = 1, 2, \dots\}$, of which the distribution is the exponential distribution with parameter $\lambda > 0$. Here the probability space $(\Omega, \mathcal{F}, P)$ is clearly defined. Next we set

$$S_0 = 0, \qquad S_n = T_1 + \cdots + T_n\,, \quad n = 1, 2, \dots,$$

and for $t \geq 0$,

$$X_t = \max\{n \geq 0\,;\, S_n \leq t\}.$$

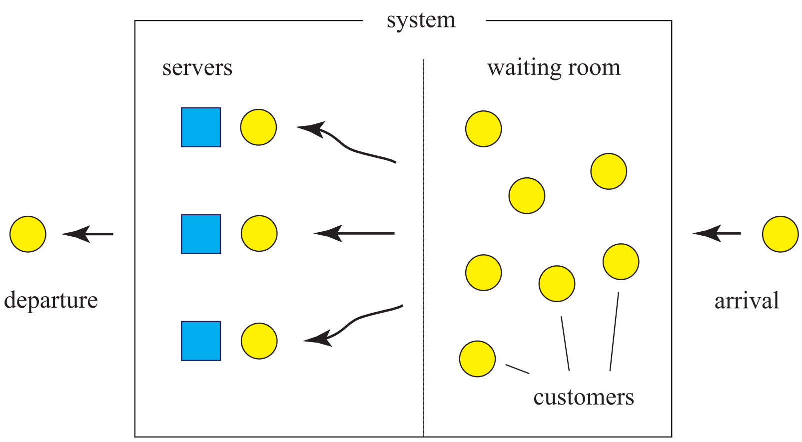
It is obvious that for each $t \geq 0$, X_t is a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. In other words, $\{X_t\,;\, t \geq 0\}$ becomes a continuous time stochastic process. This is called *Poisson process* with parameter λ by definition.

Starting with the above definition one can prove the properties in mentioned Theorem 8.1.1.

9 Queueing Theory

9.1 Modeling Queues

In our daily life, we observe often waiting lines or queues of customers for services. Agner Krarup Erlang (1878–1929, Danish engineer at the Copenhagen Telephone Exchange) published in 1909 the paper entitled: *The Theory of Probabilities and Telephone Conversations*, which opened the door to the research field of *queueing theory*. Such a queue is modeled in terms of a system consisting of servers and a waiting room. Customers arriving at the system are served at once if there is an idle server. Otherwise, the customer waits for a vacant server in a waiting room. After being served, the customer leaves the system.



In most of the qeueing models, a customer arrives at random and the service time is also random. So we are interested in relevant statistics such as

(1) sojourn time (time of a customer staying in the system)

(2) waiting time (= sojourn time - service time)

(3) the number of customers in the system

Apparently, many different conditions may be introduced for the queueing system. In 1953, David G. Kendall introduced the so-called *Kendall's notation*

$$A/B/c/K/m/Z$$

for describing the characteristics of a queuing model, where

A: arrival process,
B: service time distribution,
c: number of servers,
K: number of places in the system (or in the waiting room),
m: calling population,
Z: queue's discipline or priority order, e.g., FIFO (First In First Out)

The first model analyzed by Erlang in 1909 was the $M/D/1$ queue in Kendall's notation, where M means that arrivals occur according to a Poisson process, and D stands for deterministic (i.e., service time is not random but constant).

Most of queueing models are classified into four categories by the behavior of customers as follows:

(I) Delay models: customers wait in line until they can be served.

Example: $M/M/c$ queue, where

(i) customers arrive according to a Poisson process with rate λ;
(ii) there are c servers and there is an infinite waiting space;
(iii) each customer requires an exponential service time with mean $1/\mu$;
(iv) customers who upon arrival find all servers busy wait in line to be served.

(II) Loss models: customers leave the system when they find all servers busy upon arrival.

Example: Erlang's loss model $M/G/c/c$, where

(i) customers arrive according to a Poisson process with rate λ;
(ii) there are c servers and the capacity of the system is limited to c customers, i.e., there is no waiting space;
(iii) each customer requires a generally distributed service time;
(iv) customers who upon arrival find all servers busy are rejected forever.

(III) Retrial models: customers who do not find an idle server upon arrival leave the system only temporarily, and try to reenter some random time later.

Example: the Palm/Erlang-A queue, where

(i) customers arrive according to a Poisson process with rate λ;
(ii) there are c servers and there is an infinite waiting space;
(iii) each customer requires an exponential service time with mean $1/\mu$;
(iv) customers who upon arrival find all servers busy wait in line to be served;
(v) customers wait in line only an exponentially distributed time with mean $1/\theta$ (patience time).

(IV) Abandonment models: customers waiting in line will leave the system before being served after their patience time has expired.

References

1. 塩田茂雄・河西憲一・豊泉洋・会田雅樹：待ち行列理論の基礎と応用, 共立出版, 2014. (新刊書)
2. 宮沢政清：待ち行列の数理とその応用, 牧野書店, 2006.

9.2 $M/M/1$ Queue

This is the most fundamental model, which satisfies the following conditions:

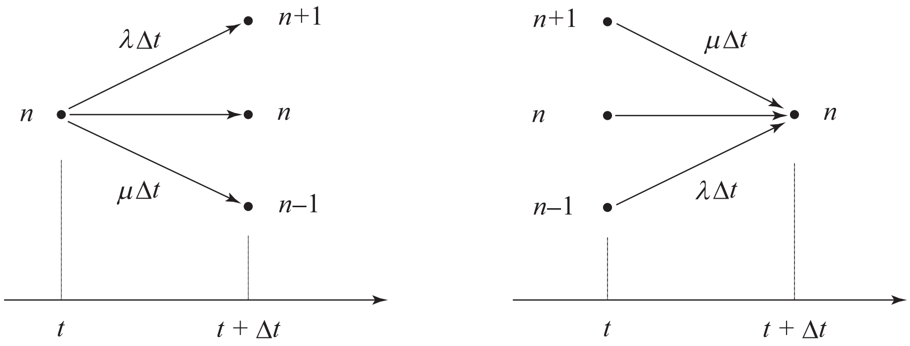
(i) arrivals occur according to a Poisson process with parameter λ;
(ii) service times obey an exponential distribution with parameter μ;
(iii) arrivals of customers and service times are independent;
(iii) the system contains a single server;
(iv) the size of waiting room is infinite;
(v) (FIFO = First In First Out) customers are served from the front of the queue, i.e., according to a first-come, first-served discipline.

Thus there are two parameters characterizing an $M/M/1$ queue, that is, the parameter $\lambda > 0$ for the Poisson arrival and the one $\mu > 0$ for the exponential service. In other words, a customer arrives at the system with average time interval $1/\lambda$ and the average service time is $1/\mu$. In the queuing theory λ is called the *mean arrival rate* and μ the *mean service rate*. Let $X(t)$ be the number of customers in the system at time t. It is the proved that $\{X(t)\,;\, t \geq 0\}$ becomes a continuous time Markov chain on $\{0, 1, 2, 3, ...\}$. In fact, the letter "M" stands for "Markov" or "memoryless".

Our main objective is

$$p_n(t) = P(X(t) = n|X(0) = 0),$$

i.e., the probabbility of finding n customers in the system at time $t > 0$ subject to the initial condition $X(0) = 0$. Let us consider the change of the system during the small time interval $[t, t + \Delta t]$. It is assumed that during the small time interval Δt only one event happens, namely, a new customer arrives, a customer under service leaves the system, or nothing changes. The probabilities of these events are given by $\lambda\Delta t$, $\mu\Delta t$, $1 - \lambda\Delta t - \mu\Delta t$.



Therefore, $P(X(t) = n|X(0) = 0)$ fulfills the following equation:

$$\begin{aligned}
P(X(t + \Delta t) = n|X(0) = 0) &= P(X(t + \Delta t) = n|X(t) = n - 1)P(X(t) = n - 1|X(0) = 0)\\
&\qquad + P(X(t + \Delta t) = n|X(t) = n)P(X(t) = n|X(0) = 0)\\
&\qquad + P(X(t + \Delta t) = n|X(t) = n + 1)P(X(t) = n + 1|X(0) = 0)\\
&= \lambda\Delta t P(X(t) = n - 1|X(0) = 0)\\
&\qquad + (1 - \lambda\Delta t - \mu\Delta t)P(X(t) = n|X(0) = 0)\\
&\qquad + \mu\Delta t P(X(t) = n + 1|X(0) = 0),\\
P(X(t + \Delta t) = 0|X(0) = 0) &= (1 - \lambda\Delta t)P(X(t) = 0|X(0) = 0) + \mu\Delta t P(X(t) = 1|X(0) = 0).
\end{aligned}$$

Hence for $p_n(t) = P(X(t) = n|X(0) = 0)$ we have

$$\begin{aligned}
&p_n'(t) = \lambda p_{n-1}(t) - (\lambda + \mu)p_n(t) + \mu p_{n+1}(t), \quad n = 1, 2, \dots,\\
&p_0'(t) = -\lambda p_0(t) + \mu p_1(t).
\end{aligned} \tag{9.1}$$

The initial condition is as follows:

$$p_0(0) = 1, \qquad p_n(0) = 0 \quad \text{for } n \geq 1. \tag{9.2}$$

Solving the linear system (9.1) with the initial condition (9.2) is not difficult with the help of linear algebra and spectral theory. However, the explicit solution is not so simple and is omitted. We only mention that most important characteristics are obtained from the explicit $p_n(t)$.

Here we focus on the equilibrium solution (limit transition probability), i.e.,

$$p_n = \lim_{t\to\infty} p_n(t)$$

whenever the limit exists. Since in the equilibrium the derivative of the left hand side of (9.1) is 0, we have

$$\begin{aligned}
&\lambda p_{n-1} - (\lambda + \mu)p_n + \mu p_{n+1} = 0 \quad n = 1, 2, \dots,\\
&-\lambda p_0 + \mu p_1 = 0.
\end{aligned} \tag{9.3}$$

A general solution to (9.3) is easily derived:

$$p_n = \begin{cases} C_1 + C_2 \left(\dfrac{\lambda}{\mu}\right)^n, & \lambda \neq \mu,\\ C_1 + C_2 n, & \lambda = \mu. \end{cases}$$

Since p_n gives a probability distribution, we have $p_n \geq 0$ and $\sum_{n=0}^{\infty} p_n = 1$. This occurs only when $\lambda < \mu$ and we have

$$p_n = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n, \quad n = 0, 1, 2, \dots.$$

This is the geometric distribution with parameter λ/μ.

In queuing theory, the ratio of the mean arrival rate λ and the mean service rate μ is called the *utilization*:

$$\rho = \frac{\lambda}{\mu}.$$

Utilization stands for how busy the system is. It was shown above that the number of customers in the system after long time obeys the geometric distribution with parameter ρ. If $\rho < 1$, the system functions well. Otherwise, the queue will continue to grow as time goes on. After long time, i.e., in the equilibrium the number of customers in the system obeys the geometric distribution:

$$(1-\rho)\rho^n, \qquad n = 0, 1, 2, \dots.$$

In particular, the probability that the server is free is $1-\rho$ and the probability that the server is busy and the customer need to wait is ρ. This is the origin of the term *utilization*. Note also that the mean number of the customers in the system is given by

$$\sum_{n=0}^{\infty} n p_n = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}.$$

Example 9.2.1 There is an ATM, where each customer arrives with average time interval 5 minutes and spends 3 minutes in average for the service. Using an $M/M/1$ queue, we know some statistical characteristics. We set

$$\lambda = \frac{1}{5}, \quad \mu = \frac{1}{3}, \quad \rho = \frac{\lambda}{\mu} = \frac{3}{5}.$$

Then the probability that the ATM is free is $p_0 = 1-\rho = \frac{2}{5}$. The probability that the ATM is busy but there is no waiting customer is

$$p_1 = \frac{2}{5} \times \frac{3}{5} = \frac{6}{25}.$$

Hence the probability that the ATM is busy and there is some waiting customers is

$$1 - p_0 - p_1 = 1 - \frac{2}{5} - \frac{6}{25} = \frac{9}{25} = 0.36.$$

So, roughly speaking, a customer needs to make a queue once per three visits.

Remark 9.2.2 The Markov process $X(t)$ appearing in the $M/M/1$ queuing model is studied more generally within the framework of *birth-and-death process*.

Problem 25 ($M/M/1/1$ queue) There is a single server and no waiting space. Customers arrive according to the Poisson process with parameter λ, and their service time obeys the exponential distribution with parameter μ. Let $Q(t)$ be the number of customers in the system at time t. In fact,

$$Q(t) = \begin{cases} 1, & \text{server is busy,} \\ 0, & \text{server is idle,} \end{cases}$$

(1) Find

$$p_0(t) = P(Q(t) = 0 | Q(0) = 0),$$
$$p_1(t) = P(Q(t) = 1 | Q(0) = 0)$$

by solving a linear system satisfied by those $p_0(t)$ and $p_1(t)$.

(2) Using the results in (1), calculate

$$\bar{p}_0 = \lim_{t\to\infty} p_0(t), \qquad \bar{p}_1 = \lim_{t\to\infty} p_1(t),$$

(3) Find the mean number of customers in the system in the long time limit:

$$\lim_{t\to\infty} \mathbf{E}[Q(t)|Q(0) = 0].$$