2015 年度後期

確率モデル論 (情報科学研究科)
応用解析学 (工学研究科)
確率モデル論 (国際高等研究教育院)

● **授業科目の目的・概要及び達成目標等**

理工系科学・生命系科学をはじめ人文社会系科学に至るまで、ランダム現象の数理解析はますます重要になっている。本講義では、そのために必要不可欠となる確率論の基礎概念からはじめ、確率モデルの構成と解析手法を学ぶ。特に、ランダム現象の時間発展を記述する確率過程として、ランダムウォーク・マルコフ連鎖・マルコフ過程の典型例をとりあげて、その性質と幅広い応用を概観する。

Mathematical analysis is important for the understanding of random phenomenon apperaing in various fields of natural, life and social sciences, and the probabilistic approach is essential. We start with fundamental concepts in probability theory and learn basic tools for probabilistic models. In particular, for the time evolution of random phenomenon we study basic properties of random walks, Markov chains, Markov processes, and take a bird's-eye view of their wide applications.

● **Topics**

1. Random variables and probability distributions
2. Bernoulli trials
3. Random walks
4. Markov chains
5. Poisson processes
6. Queues
7. Galton-Watson branching processes
8. Birth-and-death processes .... etc

● **成績**

(1) 配布プリントの問題 (**Problem** として通し番号がつく) から数題を選択してレポートを作成し, 1 月後半に提出してもらう予定. 詳細は後日発表する (コピーレポートは零点).

(2) 例外的な取り扱いは一切しない.

● レジュメ

1. レジュメ (英語) を担当者のウェッブサイト (www.math.is.tohoku.ac.jp/˜obata) からダウンロードして, 各自準備する.
   On participating in the lectures, download the materials from Obata's website.
2. ウェッブサイトには, 過去の講義録・その他の資料も掲載されている.
3. レジュメは毎回の講義内容の項目を示すもので**テキストではない**から, レジュメだけ読んでも勉強にはならない. 書物をひも解いて勉強すること.
4. レジュメは講義の進度に従って準備するので, まとめて事前に配布することはしない.

● Basic References

1. 拙著：確率モデル要論 (牧野書店), 2012.
   過年度の講義をまとめたもので, 本講義の内容もおおむねこの本にしたがう.
2. D. L. Minh: Applied Probability Models, Duxbury, 2001.
   英語で読むのなら, この本が便利である.

● Further Reading

1. S. M. Ross: Introduction to Probability Models, 11th Ed. Academic Press, 2014.

   初版は 1972 年のロングセラー. だが, 800 ページに迫る大部. 取り扱っている内容は [Minh] とオーバーラップする部分が多いが、より初等的なレベルから解説していて読みやすい.

2. J. R. Norris: Markov Chains, Cambridge UP, 1998.

   いろいろな具体例を扱い, 大変読みやすい定評のある教科書.

3. W. Feller: An Introduction to Probability Theory and Its Applications, Vol. 1, Wiley, 1957.

   名著の誉れ高い. この本は講義内容をカバーし, さらに詳しいことがたくさん書かれている (Vol. 2 もある!). 邦訳もある.

   W. フェラー (河田龍夫他訳) : 確率論とその応用 (紀伊国屋). こちらは 4 分冊.

4. B. V. Gnedenko: The Theory of Probability and the Elements of Statistics, AMS Chelsea Publishing Co., 6th ed. 1989.

5. R. Durrett: Probability: Theory and Examples, Duxbury Press, 1996.

   この 2 冊も講義内容をカバーし, さらに詳しいことがたくさん書かれている. この程度の知識があれば, 確率モデルを本格的に研究に生かせるだろう.

6. 志賀徳造：ルベーグ積分から確率論 (共立), 2000.

   前半はルベーグ積分を展開しているが, 後半でランダムウォークを取り扱って確率モデルへの入門をはたす.

7. 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, 1999.

   マルコフ連鎖の基礎理論, ゴルトン・ワトソン過程, 出生死亡過程を含み, 手頃である.

8. 国沢清典 : 確率論とその応用 (岩波全書), 1982.

   少し古いが, 本講義はこのレベルをめざす.

9. 舟木直久：確率論, 朝倉書店, 2004.

10. 西尾真喜子：確率論, 実教出版, 1978.

    この 2 冊はさらに高度なところまで数学理論として展開している.

11. P. ブレモー (釜江哲朗監修, 向井久訳)：モデルで学ぶ確率入門 (新装版), シュプリンガー東京, 2004.

    実用の場面を想定したさまざまな確率モデルが取り上げられている. 例題を通して数学的な枠組を学ぶ形式で書かれている. 個々の事例は興味深いが, 理論を知った上で見ないと難しいかもしれない.

12. F. Spitzer: Principles of Random Walk, Springer, 2nd Ed., 1976.

    ランダムウォークに関するプロ向きの本.

13. K. L. Chung: Markov Chains, Springer, 1960.

    マルコフ連鎖に関するプロ向きの本.

14. イアン・ハッキング (石原英樹・重田園江訳)：偶然を飼いならす, 木鐸社, 1999.

    「この博物誌的な書物を好奇心に満ちたすべての読者に捧げる」とある. 確率統計が 20 世紀の科学に中でいかに成功してきたかを科学史的な視点で論ずる. かなり興味深い.

15. イアン・ハッキング (広田すみれ・森元良太訳)：確率の出現, 慶應義塾大学出版会, 2013.

16. キース・デブリン (原 啓介訳) : 世界を変えた手紙 — パスカル、フェルマーと〈確率〉の誕生, 岩波書店, 2010.

    確率論の始まりをさまざまなエピソードとともに語る. 個人的にはカルダーノに大変興味がある.

Overview

0.1 Stochastic Processes

We will study the probability models for time evolution of random phenomena. Measuring a certain quantity of the random phenomenon at each time step $n = 0, 1, 2, \dots$, we obtain a sequence of real values:

$$x_0, x_1, x_2, \dots, x_n, \dots.$$

Because of randomness, we consider x_n as a realized value of a random variable X_n. Here a random variable is a variable taking several different values with certain probabilities. Thus, the time evolution of a random phenomenon is modeled by a sequence of random variables

$$\{X_n\,;\, n = 0, 1, 2, \dots\} = \{X_0, X_1, X_2, \dots, X_n, \dots\},$$

which is called a *discrete-time stochastic process*. If the measurement is performed along with continuous time, we need a *continuous-time stochastic process*:

$$\{X_t\,;\, t \geq 0\}$$

It is our purpose to construct stochastic processes modeling typical random phenomena and to demonstrate their properties within the framework of modern probability theory.

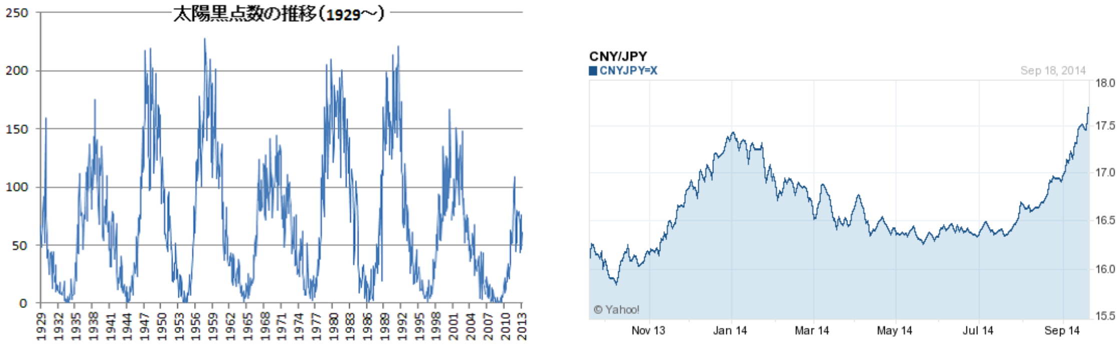


Figure 1: Solar spots and exchanging rates

We hope that you will obtain basic concepts and methods through the following three subjects:

0.2 One-Dimensional Random Walk and Gambler's Ruin Problem

Let us consider coin tossing. We get $+1$ if the heads appears, while we get -1 (i.e., lose $+1$) if the tails appears. Let Z_n be the value of the n-th coin toss.

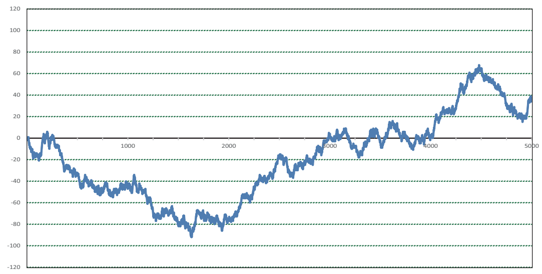
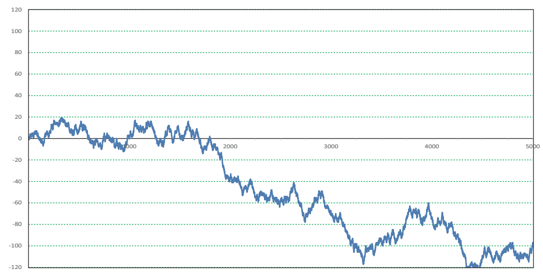
To be precise, we must say that $\{Z_n\}$ is a sequence of independent, identically distributed (iid) random variables with the common distribution

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

In short, $\{Z_n\}$ is called the *Bernoulli trials* with success probability $1/2$. Define

$$X_0 = 0, \quad X_n = \sum_{k=1}^{n} Z_k \quad n = 1, 2, \dots.$$

Then X_n means the net income at the time n, or the coordinate of a drunken walker after n steps. The discrete time stochastic process $\{X_n\}$ is called *one-dimensional random walk*.

(i) law of large numbers

(ii) diffusion speed (central limit theorem)

(iii) recurrence

(iv) long leads (law of happy time)

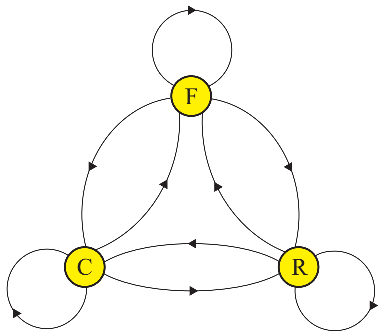
(v) gambler's ruin (random walk with barriers)

0.3 Markov Chains

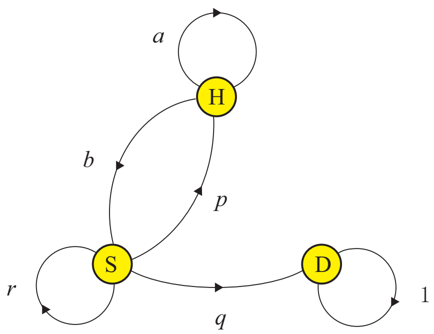
Consider the time evolution of a random phenomenon, where several different *states* are observed at each time step $n = 0, 1, 2, \dots$. For example, for the ever-changing weather, after simplification we observe three states: fine, cloudy, rainy. Collected data look like a sequence of F, C, R:

$$F\,F\,C\,R\,F\,C\,C\,F\,R\,\dots$$

from which we may find the conditional probability $P(X|Y)$ of having a weather X just after Y. Then we come to the transition diagram, where each arrow $Y \to X$ is asigned the conditional probability $P(X|Y)$.



The above diagram describes a general Markov chain over the three states because the transitions occur between every possible pair of states. According to our purpose, we may consider variations. For example, we may consider the following diagram for analysis of life span.



A Markov chain $\{X_n\}$ is a discrete-time stochastic process over a state space $S = \{i, j, \dots\}$ (always assumed to be finite or countably infinite), which is governed by the one-step transition probability:

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

where the right hand side is independent of n (time homogeneous). A random walk is an example of a Markov chain. The theory of Markov chains is one of the best successful theories in probability theory for its simple description and unexpectedly rich structure. We are interested in the following topics:

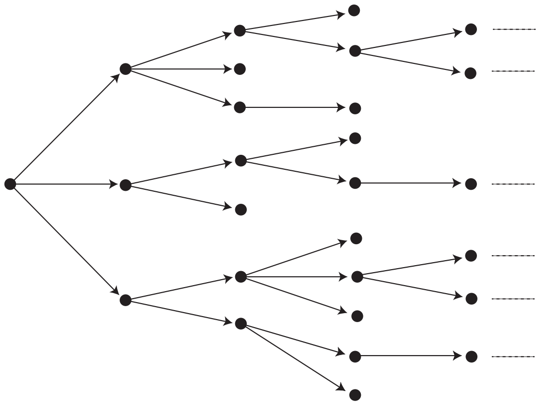
(i) stationary distribution

(ii) recurrence

(iii) average life span

(iv) survival of family names (Galton-Watson tree)

(v) birth-and-death chains



0.4 Poisson Process

Let us imagine that an event E occurs repeatedly at random as time goes on. For example, alert of receiving an e-mail, passengers making a queue at a bus stop, customers visiting a shop, occurrence of defect of a machine, radiation from an atom, etc.

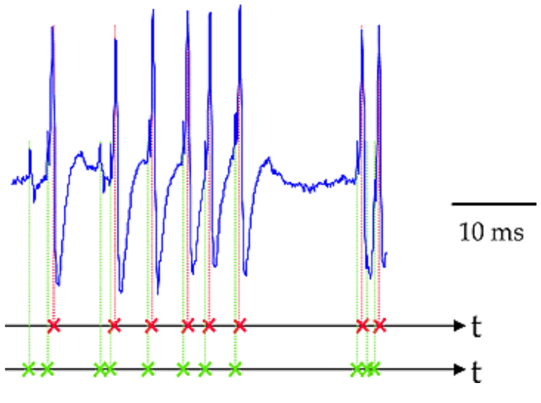
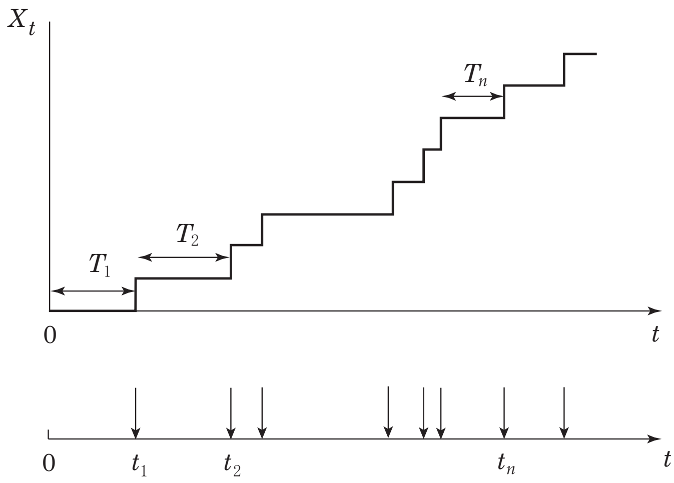


Figure 2: Nerve impulses

To obtain a stochastic process, we count the number of occurrence of the event E during the time interval $[0, t]$, which is denoted by X_t. Then we obtain a stochastic process $\{X_t\,;\, t \ge 0\}$. The situation is illustrated as follows, where $t_1, t_2, \dots$ are the time when E occurs.



A fundamental case is described by a Poisson process, where the event happens independently each other. The first to check is the statistics between two consecutive occurrence of events (waiting time).

(i) Applications to queuing theory (waiting lines are modeled by a Poisson process).

(ii) A birth-and-death process as generalization.

Poisson process is one of the fundamental examples of (continuous-time) *Markov processes*. Another is the Brownian motion (Wiener process).

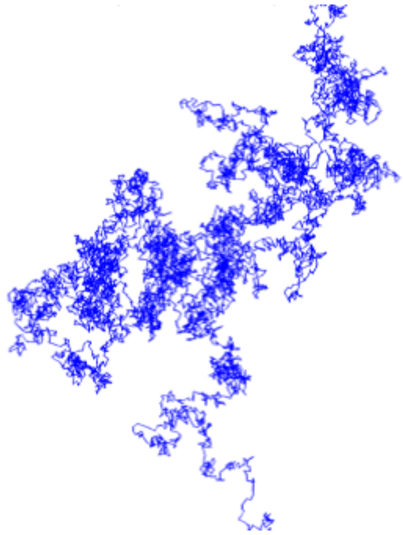


Figure 3: Two-dimensional Brownian Motion (simulation)

Contents

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

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

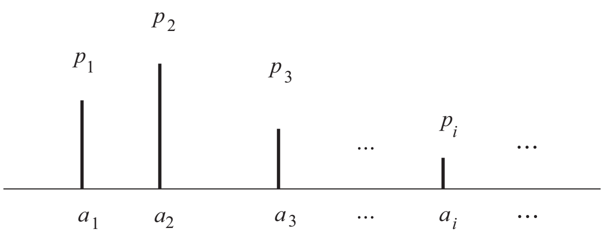
In this case

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

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

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

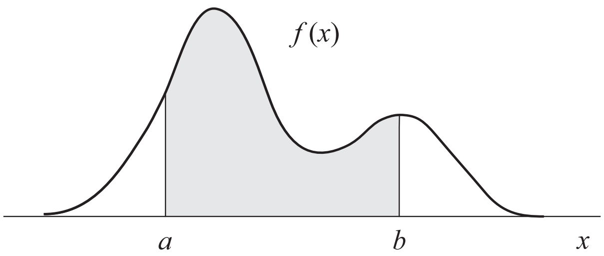
we say that X admits a *probability density function*. Note that

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

In this case,

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

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



It is useful to consider the *distribution function*:

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

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).

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

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

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

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

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

so the probability density function is given by

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

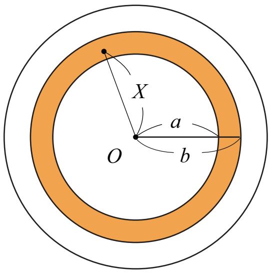


Figure 1.1: Random choice of a point

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

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

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

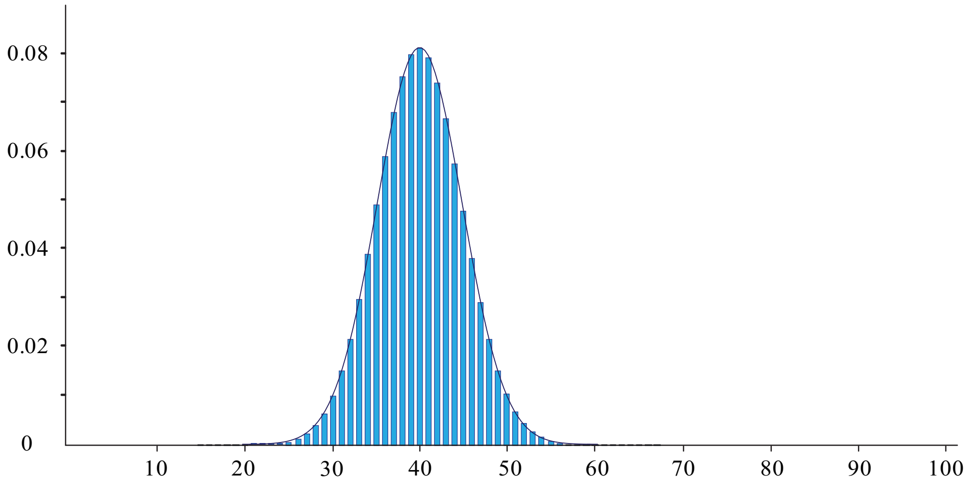
where $0 \le p \le 1$ and $p + q = 1$. Determine the two-point distribution having mean 0, variance 1.

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).

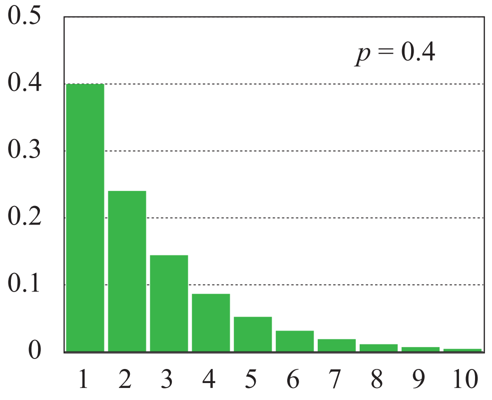


Figure 1.2: Geometric distribution with parameter $p = 0.4$

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

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

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

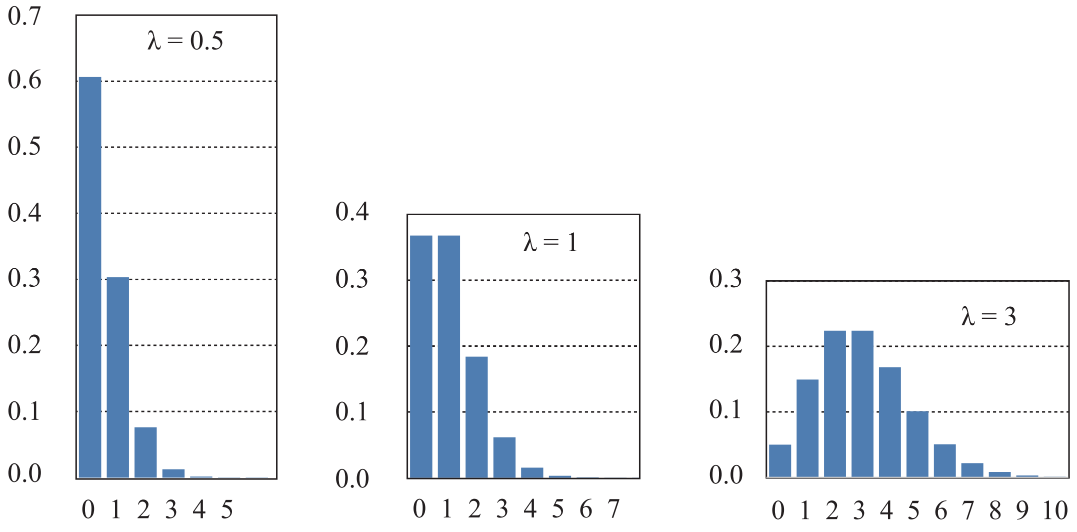


Figure 1.3: Poisson distribution $\lambda = 1/2, 1, 3$

Problem 1 Set

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

(1) By using $G'(1)$ show that the mean value of the Poisson distribution with parameter λ is given by $m = \lambda$.

(2) By using $G''(1)$ show that the variance of the Poisson distribution with parameter λ is given by $\sigma^2 = \lambda$.

(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.

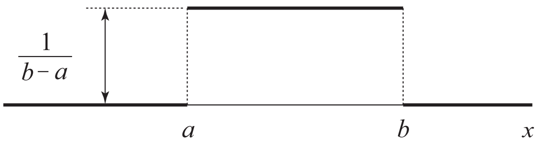
1.3 Continuous Distributions (Density Functions)

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

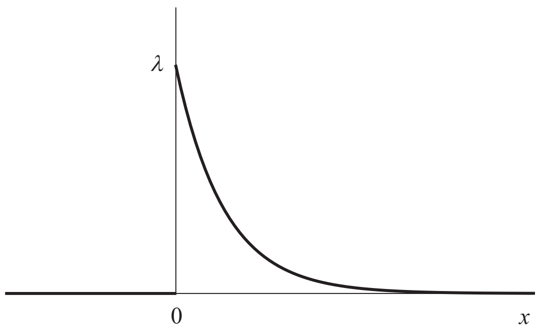
$$m = \int_a^b x \frac{dx}{b-a} = \frac{a+b}{2}, \qquad \sigma^2 = \int_a^b x^2 \frac{dx}{b-a} - m^2 = \frac{(b-a)^2}{12}.$$

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).



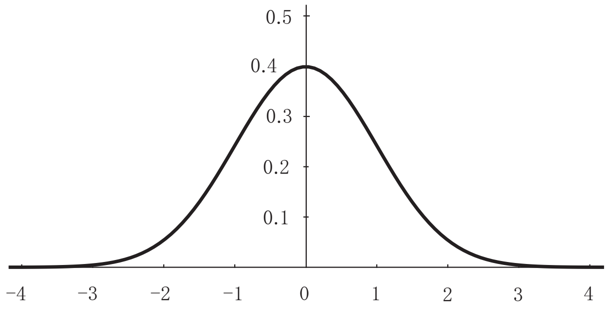
Exercise 1.3.1 Verify that $m = \dfrac{1}{\lambda}$ and $\sigma^2 = \dfrac{1}{\lambda^2}$.

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:

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

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

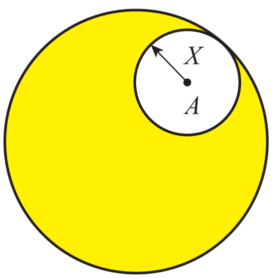
Problem 2 Choose randomly a point A from the disc with radius one and let X be the radius of the inscribed circle with center A.

(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 $S = \pi X^2$.



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.

Example 2.1.4 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 $A_i = \{X_i = 1\}$. Then $\{A_1, A_2, A_3\}$ is pairwise independent but is not independent.

Remark 2.1.5 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 $A = \emptyset$. Similarly, $P(A) = 1$ does not imply $A = \Omega$ (whole event).

Exercise 2.1.6 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 2.1.7 (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 2.1.8 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. ∎

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 2.3.3 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 2.3.4 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 2.3.5 If two random variables X, Y are independent, they are uncorrelated.

Remark 2.3.6 The converse of Theorem 2.3.5 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 2.3.5 is valid (see Problem 5).

Theorem 2.3.7 (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 2.3.8 $-1 \le \rho_{XY} \le 1$ 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 3 Throw two dice and let L be the larger spot and S the smaller. (If double spots, set $L = S$.) Calculate the covariance σ_{LS} and the correlation coefficient ρ_{LS}. Then explain the meaning of the signature of ρ_{LS}.

Problem 4 Let $\{X_n\}$ be Bernoulli trials with success probability p and set

$$S_n = \sum_{k=1}^{n} X_k \qquad S_0 = 0. \tag{2.3}$$

The stochastic process $\{S_n\}$ is called the *binomial process*.

(1) Show that S_n obeys the binomial distribution $B(n, p)$.

(2) Find the covariance $\mathbf{Cov}\,(S_{m+n}, S_m)$ for $m \ge 0$ and $n \ge 1$.

(3) Find the correlation coeficient ρ_{S_{m+n}, S_m} for $m \ge 0$ and $n \ge 1$.

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 $\sigma_{XY} = 0$. [In general, uncorrelated random variables are not necessarily independent. Hence, the above falls into a very particular situation.]

3 Limit Theorems

3.1 Simulation of Coin Toss

Let $\{X_n\}$ be a Bernoulli trial with success probability $1/2$, namely, tossing a fair coin, and consider the binomial process defined by

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

Since S_n counts the number of heads during the first n trials,

$$\frac{S_n}{n} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

gives the relative frequency of heads during the first n trials.

The following is just one example showing that the relative frequency of heads S_n/n tends to $1/2$.

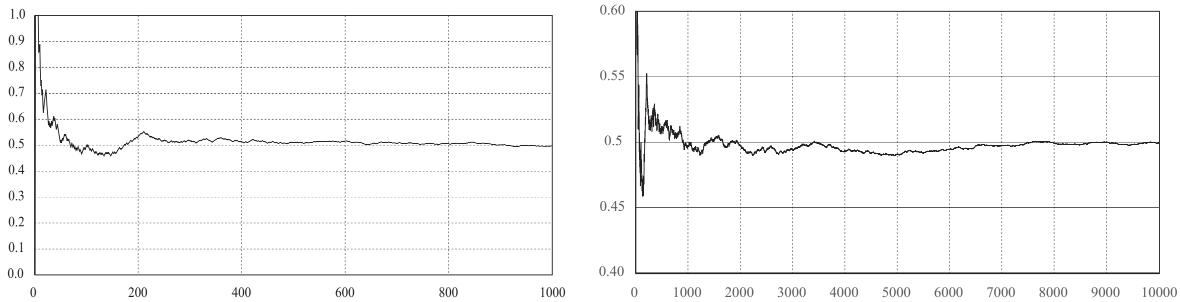


Figure 3.1: Relative frequency of heads S_n/n

It is our aim to show this mathematically. However, we cannot accept a naive formula:

$$\lim_{n\to\infty} \frac{S_n}{n} = \frac{1}{2} \tag{3.1}$$

because

1. Notice that S_n/n is a random variable taking values in $\{0, 1/n, 2/n, \dots, 1\}$.

2. From one series of trials $\omega = (\omega_1, \omega_2, \dots)$ we obtain a sequence of relative frequencies:

$$S_1(\omega), \ \frac{S_2(\omega)}{2}, \ \frac{S_3(\omega)}{3}, \ \dots, \ \frac{S_n(\omega)}{n}, \ \dots.$$

3. For example, for $\omega = (1, 1, 1, \dots)$, S_n/n converges to 1; For $\omega = (0, 0, 0, \dots)$, S_n/n converges to 0. Moreover, for any $0 \le t \le 1$ there exists ω such that S_n/n converges to t; there exists ω such that S_n/n does not converge (oscillating).

4. Namely, it is impossible to show the limit formula (3.1) for *all* samples ω.

Therefore, to show the empirical fact (3.1) we need some *probabilistic formulation*.

3.2 Law of Large Numbers (LLN)

Theorem 3.2.1 (Weak law of large numbers) Let $X_1, X_2, \dots$ be identically distributed random variables with mean m and variance σ^2. (This means that X_i has a finite variance.) If $X_1, X_2, \dots$ are uncorrelated, for any $\epsilon > 0$ we have

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - m\right| \ge \epsilon\right) = 0.$$

We say that $\frac{1}{n}\sum_{k=1}^{n} X_k$ converges to m in probability.

Remark 3.2.2 In many literatures the weak law of large numbers is stated under the assumption that $X_1, X_2, \dots$ are independent. It is noticeable that the same result holds under the weaker assumption of being uncorrelated.

Theorem 3.2.3 (Chebyshev inequality) Let X be a random variable with mean m and variance σ^2. Then, for any $\epsilon > 0$ we have

$$P(|X - m| \ge \epsilon) \le \frac{\sigma^2}{\epsilon^2}.$$

Proof Set $A = \{|X - m| \ge \epsilon\}$ and let 1_A be the indicator random variable. Then

$$\mathbf{E}[1_A] = P(A).$$

By definition we have

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

Now we calculate the right-hand side as follows:

$$\begin{aligned}\mathbf{E}[(X - m)^2] &= \mathbf{E}[(X - m)^2 1_A + (X - m)^2 1_{A^c}] \\ &\ge \mathbf{E}[(X - m)^2 1_A] \\ &\ge \mathbf{E}[\epsilon^2 1_A] = \epsilon^2 P(A).\end{aligned}$$

Then we have

$$\sigma^2 \ge \epsilon^2 P(A),$$

which proves the assertion. ∎

Proof [Theorem 3.2.1 (Weak Law of Large Numbers)] For simplicity we set

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

The mean value is given by

$$\mathbf{E}[Y] = \frac{1}{n}\sum_{k=1}^{n} \mathbf{E}[X_k] = m.$$

Since $X_1, X_2, \dots$ are pairwise uncorrelated, the variance is computed by using the additive property of variance. In fact, we have

$$\mathbf{V}[Y] = \frac{1}{n^2}\mathbf{V}\left[\sum_{k=1}^{n} X_k\right] = \frac{1}{n^2}\sum_{k=1}^{n} \mathbf{V}[X_k] = \frac{1}{n^2} \times n\sigma^2 = \frac{\sigma^2}{n} .$$

On the other hand, applying Chebyshev inequality, we have

$$P(|Y - m| \ge \epsilon) \le \frac{\mathbf{V}[Y]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} .$$

Consequently,

$$\lim_{n\to\infty} P(|Y_n - m| \ge \epsilon) = 0,$$

as desired. ∎

Example 3.2.4 (Coin toss)

In fact, we have a stronger result.

Theorem 3.2.5 (Strong law of large numbers) Let $X_1, X_2, \dots$ be identically distributed random variables with mean m. (This means that X_i has a mean but is not assumed to have a finite variance.) If $X_1, X_2, \dots$ are pairwise independent, we have

$$P\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = m\right) = 1.$$

In other words,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = m \quad \text{a.s.}$$

Remark 3.2.6 Kolmogorov proved the strong law of large numbers under the assumption that $X_1, X_2, \dots$ are independent. In many literatures, the strong law of large numbers is stated as Kolmogorov proved. Its proof being based on the so-called "Kolmogorov's almost sure convergence theorem," we cannot relax the assumption of independence. Theorem 3.2.5 is due to N. Etemadi (1981), where the assumption is relaxed to being mutually independent and the proof is more elementary, see also books by Sato, by Durrett, etc.

3.3 De Moivre–Laplace Theorem

We know that the binomial distribution $B(n, p)$ is close to the normal distribution having the same mean $m = np$ and the variance $\sigma^2 = np(1 - p)$:

$$B(n, p) \approx N(np, np(1 - p)) \tag{3.2}$$

The matching becomes better for larger n.

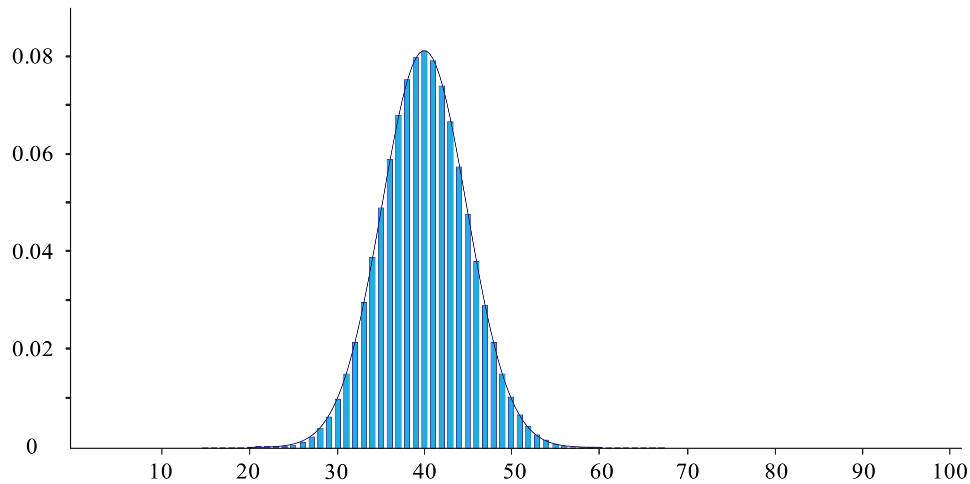


Figure 3.2: The normal distribution whose mean and variance are the same as $B(100, 0.4)$

The approximation (3.2) means that distribution functions are almost the same: For a random variable S obeying the binomial distribution $B(n, p)$ we have

$$P(S \le x) \approx \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-(t-m)^2/2\sigma^2}\, dt, \quad m = np, \quad \sigma^2 = np(1 - p).$$

3.4 Central Limit Theorem (CLT)

Theorem 3.4.1 (Central Limit Theorem) Let $X_1, X_2, \dots$ be iid random variables with mean 0 and variance 1. Then, for any $x \in \mathbf{R}$ it holds that

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k \le x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt.$$

In short, the distribution of $\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k$ converges weakly to the standard normal distribution $N(0,1)$ as $n \to \infty$.

For the proof we need the characteristic function of a distribution.

Definition 3.4.2 The *characteristic function* of a random variable X is defined by

$$\varphi(z) = \mathbf{E}[e^{izX}] = \int_{-\infty}^{+\infty} e^{izx}\mu(dx), \qquad z \in \mathbf{R},$$

where $\mu(dx)$ is the distribution of X. We also say that $\varphi(z)$ is the characteristic function of $\mu(dx)$.

Lemma 3.4.3 (Glivenko's theorem) Let $\mu_1, \mu_2, \dots, \mu$ be a sequence of probability distributions and $\varphi_1, \varphi_2, \dots, \varphi$ their characteristic functions. If $\lim_{n\to\infty} \varphi_n(z) = \varphi(z)$ holds for all $z \in \mathbf{R}$, then μ_n converges weakly to μ. In other words, letting $F_1, F_2, \dots, F$ be distribution functions of $\mu_1, \mu_2, \dots, \mu$, we have

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all continuous point x of F.

Lemma 3.4.4 Let $a \in \mathbf{C}$ and $\{\epsilon_n\}$ a sequence of complex numbers converging to 0. Then we have

$$\lim_{n\to\infty}\left(1 + \frac{a}{n} + \frac{\epsilon_n}{n}\right)^n = e^a.$$

Proof of Central Limit Theorem (outline) 1) Let $\varphi_n(z)$ be the characteristic function of $\frac{1}{\sqrt{n}}\sum_{k=0}^{n} X_k$, i.e.,

$$\varphi_n(z) = \mathbf{E}\left[\exp\left\{\frac{iz}{\sqrt{n}}\sum_{k=0}^{n} X_k\right\}\right]. \tag{3.3}$$

On the other hand, it is known that the characteristic function of $N(0,1)$ is given by $e^{-z^2/2}$ (Exercise 3.4.6). By virtue of Glivenko's theorem it is sufficient to show that

$$\lim_{n\to\infty} \varphi_n(z) = e^{-z^2/2}, \qquad z \in \mathbf{R}. \tag{3.4}$$

2) The characteristic functions of $X_1, X_2, \dots$ are identical, since they have the same distribution. We set

$$\varphi(z) = \mathbf{E}[e^{izX_1}].$$

Since $X_1, X_2, \dots$ are independent, we have

$$\varphi_n(z) = \prod_{k=1}^{n} \mathbf{E}\left[\exp\left\{\frac{iz}{\sqrt{n}}X_k\right\}\right] = \varphi\left(\frac{z}{\sqrt{n}}\right)^n. \tag{3.5}$$

3) By Taylor expansion we write

$$e^{i\frac{z}{\sqrt{n}}X_1} = 1 + i\frac{z}{\sqrt{n}}X_1 - \frac{z^2}{2n}X_1^2 + R_n(z)$$

and take the expectation

$$\varphi\Big(\frac{z}{\sqrt{n}}\Big) = \mathbf{E}[e^{i\frac{z}{\sqrt{n}}X_1}] = 1 - \frac{z^2}{2n} + \mathbf{E}[R_n(z)],$$

where $\mathbf{E}[X_1] = 0$ and $\mathbf{V}[X_1] = 1$ are taken into account. Hence (3.5) becomes

$$\varphi_n(z) = \Big(1 - \frac{z^2}{2n} + \mathbf{E}[R_n(z)]\Big)^n. \tag{3.6}$$

4) If we have

$$\lim_{n\to\infty} n\mathbf{E}[R_n(z)] = 0, \tag{3.7}$$

applying Lemma 3.4.4, we obtain

$$\lim_{n\to\infty} \varphi_n(z) = \lim_{n\to\infty} \Big(1 - \frac{z^2}{2n} + \mathbf{E}[R_n(z)]\Big)^n = e^{-z^2/2}.$$

5) We need to show (3.7). We prepare a useful inequality:

$$\left|e^{ix} - \Big(1 + ix + \frac{(ix)^2}{2!}\Big)\right| \le \min\left\{\frac{|x|^3}{6}, |x|^2\right\}, \qquad x \in \mathbf{R}. \tag{3.8}$$

(Try to prove!) Then we obtain

$$|R_n(z)| \le \min\left\{\frac{1}{6}\left|\frac{z}{\sqrt{n}}X_1\right|^3, \left|\frac{z}{\sqrt{n}}X_1\right|^2\right\}$$

and

$$|n\mathbf{E}[R_n(z)]| \le \mathbf{E}[n|R_n(z)|] \le |z|^2\mathbf{E}\left[\min\left\{\frac{|z|}{6\sqrt{n}}|X_1|^3, |X_1|^2\right\}\right]. \tag{3.9}$$

Note that

$$\min\left\{\frac{|z|}{6\sqrt{n}}|X_1|^3, |X_1|^2\right\} \le |X_1|^2$$

and $\mathbf{E}[|X_1|^2] < \infty$ by assumption. Then, applying the Lebesgue convergence theorem we come to

$$\lim_{n\to\infty} \mathbf{E}\left[\min\left\{\frac{|z|}{6\sqrt{n}}|X_1|^3, |X_1|^2\right\}\right] = \mathbf{E}\left[\lim_{n\to\infty} \min\left\{\frac{|z|}{6\sqrt{n}}|X_1|^3, |X_1|^2\right\}\right] = 0,$$

which shows (3.7). ∎

Remark 3.4.5 In the above proof we did not require $\mathbf{E}[|X_1|^3] < \infty$. If $\mathbf{E}[|X_1|^3] < \infty$ is satisfied, (3.7) follows more easily without appealing to the Lebesgue convergence theorem.

Exercise 3.4.6 Calculate the characteristic function of the standard normal distribution:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{izx} e^{-x^2/2} dx = e^{-z^2/2}, \qquad z \in \mathbf{R}.$$

Remark 3.4.7 The de Moivre–Laplace theorem is just a corollary of the central limit theorem. In fact, let $Z_1, Z_2, \dots$ be the Bernoulli trials with success probability p. Define the normalization by

$$\bar{Z}_k = \frac{Z_k - p}{\sqrt{p(1-p)}}.$$

Then $\bar{Z}_1, \bar{Z}_2, \dots$ become iid random variables with mean 0 and variance 1. Apply the central limit theorem, we see that

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \bar{Z}_k = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{Z_k - p}{\sqrt{p(1-p)}} = \frac{1}{\sqrt{np(1-p)}} \sum_{k=1}^{n} (Z_k - p)$$

obeys $N(0,1)$ in the limit as $n \to \infty$. Then,

$$\sum_{k=1}^{n}(Z_k - p) = \sum_{k=1}^{n} Z_k - pn$$

obeys $N(0, np(1-p))$, and hence

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

obeys $N(np, np(1-p))$ for a large n.

Problem 6 (Monte Carlo simulation) Let $x_1, x_2, \ldots$ is a sequence taken randomly from $[0,1]$. Then for a continuous function $f(x)$ on the interval $[0,1]$, the mean

$$\frac{1}{n}\sum_{k=1}^{n} f(x_k)$$

is considered as a good approximation of the integral

$$\int_0^1 f(x)dx.$$

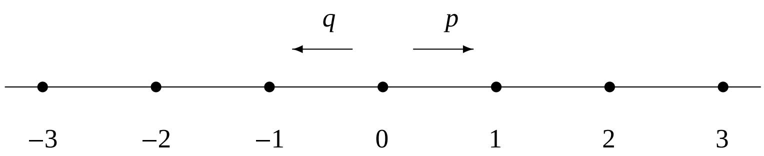
(1) Explain the above statement by means of law of large numbers and central limit theorem.

(2) By using a computer, verify the above fact for $f(x) = \sqrt{1-x^2}$.

4 Random Walks

4.1 One-Dimensional Random Walks

Let us model a drunken man (random walker) walking along a straight road. Suppose that the random walker chooses the direction (left or right) randomly at each step. Let the probability of choosing the right-move be p and the left-move q ($p > 0$, $q > 0$, $p + q = 1$). These are assumed to be independent of the position and time.



Let X_n denote the position of the random walker at time n. We assume that the random walker sits at the origin 0 at time $n = 0$, namely, $X_0 = 0$. Then $\{X_n\}$ becomes a discrete time stochastic process, which is called the *one-dimensional random walk*. To be more precise, introduce a stochastic process $\{Z_n\}$ satisfying

(i) $P(Z_n = 1) = p$ and $P(Z_n = -1) = q = 1 - p$ with $0 < p < 1$;

(ii) $Z_1, Z_2, \dots$ are independent.

We call $\{Z_n\}$ Bernoulli trials too. Define

$$X_0 = 0, \qquad X_n = \sum_{k=1}^{n} Z_k\,, \quad n \geq 1. \tag{4.1}$$

The stochastic process $\{X_n\}$ is called the *one-dimensional random walk* with right-move probability p and the left-move probability $q = 1 - p$.

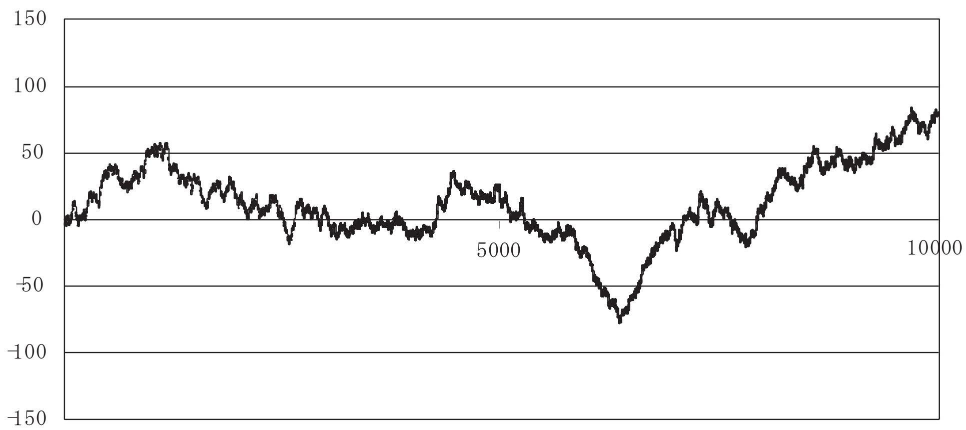


Figure 4.1: Random walk ($p = q = 1/2$)

Theorem 4.1.1 X_n is a random variable taking values in $\{-n, -n+2, \dots, n-2, n\}$. The distribution of X_n is given by

$$P(X_n = n - 2k) = \binom{n}{k} p^{n-k} q^k, \quad k = 0, 1, 2, \dots, n.$$

Proof Let $k = 0, 1, 2, \dots, n$. We observe that

$$X_n = Z_1 + Z_2 + \cdots + Z_n = n - 2k = (n-k) - k$$

if and only if the number of i's such that $Z_i = -1$ is k, and the one such that $Z_i = 1$ is $n-k$. Therefore,

$$P(X_n = n - 2k) = \binom{n}{k} p^{n-k} q^k,$$

as desired. ∎

Theorem 4.1.2 It holds that

$$\mathbf{E}[X_n] = (p-q)n, \qquad \mathbf{V}[X_n] = 4pqn.$$

Proof Note first that

$$\mathbf{E}[Z_k] = p - q, \qquad \mathbf{V}[Z_k] = 4pq.$$

Then, by linearity of the expectation we have

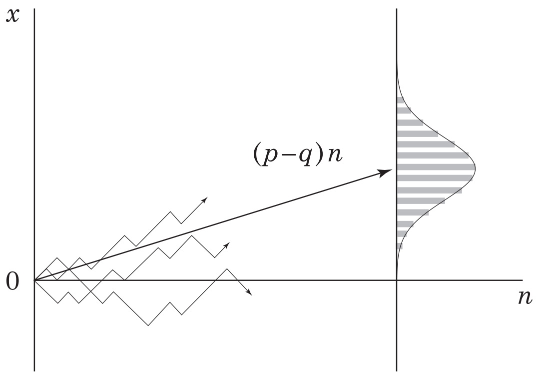
$$\mathbf{E}[X_n] = \sum_{k=1}^{n} \mathbf{E}[Z_k] = (p-q)n.$$

Since $\{Z_n\}$ is independent, by the additivity of variance we have

$$\mathbf{V}[X_n] = \sum_{k=1}^{n} \mathbf{V}[Z_k] = 4pqn.$$

∎

The distribution of X_n tells us where the random walker at time n is found. It has fluctuation around the mean value $(p-q)n$. The range of X_n grows as $n \to \infty$ and so does the variance. It is noticeable that the growth of variance is promotional to n. Finally, we note that the distribution is approximated by the normal distribution $N((p-q)n, 4pqn)$ for a large n (de Moivre–Laplace theorem).



Theorem 4.1.3 Let $\{X_n\}$ be the random walk as above. The covariance is

$$\mathrm{Cov}(X_m, X_{m+n}) = \mathbf{E}[(X_m - \mathbf{E}[X_m])(X_{m+n} - \mathbf{E}[X_{m+n}])] = 4pqm$$

and the correlation coefficient is

$$\rho(X_m, X_{m+n}) = \frac{\mathrm{Cov}(X_m, X_{m+n})}{\sqrt{\mathbf{V}[X_m]}\,\sqrt{\mathbf{V}[X_{m+n}]}} = \sqrt{\frac{m}{m+n}}\,.$$

Problem 7 Let $\{X_n\}$ be the random walk defined by (4.1). A time point $1 \le k \le n-1$ is called *turn* if

$$X_{k-1} < X_k > X_{k+1} \quad \text{or} \quad X_{k-1} > X_k < X_{k+1}$$

Find the distribution of the number of turns up to a fixed time n. In other words, letting N be the number of turns up to a fixed time n, find $P(N = j)$ for $j = 0, 1, 2, \dots$.

4.2 Recurrence

Will a random walker return to the origin in finite time? More precisely, we are interested in the probability that a random walker will return to the origin in finite time.

As in the previous section, let X_n be the position of a random walker starting from the origin (i.e., $X_0 = 0$) with right-move probability p and left-move probability q. Since the random walker returns to the origin only after even steps, we need to calculate

$$R = P\left(\bigcup_{n=1}^{\infty}\{X_{2n} = 0\}\right). \tag{4.2}$$

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,

$$p_{2n} = P(X_{2n} = 0) = \binom{2n}{n} p^n q^n = \frac{(2n)!}{n!n!}\, p^n q^n, \qquad n = 1, 2, \dots. \tag{4.3}$$

For convenience set

$$p_0 = 1.$$

Note that the right hand side of (4.2) 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 difference between p_{2n} and q_{2n}.

Definition 4.2.1 We set

$$T = \inf\{n \geq 1\,;\, X_n = 0\}, \tag{4.4}$$

where $T = +\infty$ for $\{n \geq 1\,;\, X_n = 0\} = \emptyset$. 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

$$P(T = 2n) = q_{2n} \tag{4.5}$$

and therefore, the return probability is given by

$$R = P(T < \infty) = \sum_{n=1}^{\infty} q_{2n}. \tag{4.6}$$

We will calculate $P(T = 2n)$ and R in the next section. A general method for computing (4.6) by means of generating functions will be studied in Chapter 6.

4.3 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 (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 4.3.1 The n-th *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 $n = 0, 1, 2, 3, ...$ are

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

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 4.3.2 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 $y = x$.

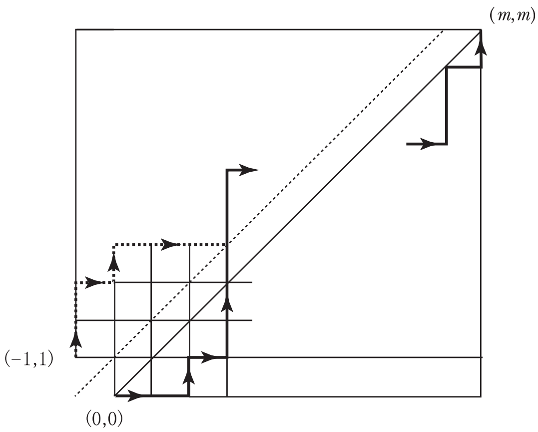
Theorem 4.3.3 (Catalan number)

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

Proof For $n = 0$ it is apparent by the definition $0! = 1$. Suppose $n \geq 1$. We see from Fig. 4.3 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 4.3.4 Let $n \geq 1$. 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.

4.4 Calculating the Return Probability

Theorem 4.4.1 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

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

Proof Obviously, we have

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

In view of Fig. 4.4 we see that

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

Then the result is immediate. ∎

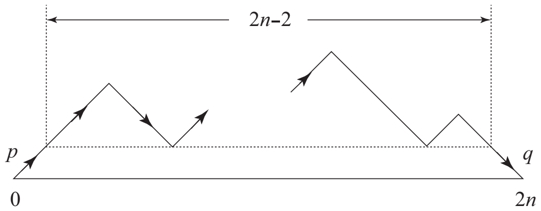


Figure 4.2: Calculating $P(X_1 > 0,\ X_2 > 0, \ldots, X_{2n-1} > 0,\ X_{2n} = 0)$

Remark 4.4.2 There are some noticeable relations between $\{p_{2n}\}$ and $\{q_{2n}\}$.

$$q_{2n} = \frac{2pq}{n}\, p_{2n-2}\,, \qquad n \geq 1,$$
$$q_{2n} = 4pqp_{2n-2} - p_{2n}\,, \qquad n \geq 1.$$

Lemma 4.4.3 The generating function of the Catalan numbers C_n is given by

$$f(z) = \sum_{n=0}^{\infty} C_n z^n = \frac{1 - \sqrt{1-4z}}{2z}\,. \tag{4.7}$$

Proof Problem 10. ∎

Theorem 4.4.4 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 4.4.1 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 4.4.3, 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 4.4.5 A random walk is called *recurrent* if $R = 1$, otherwise it is called *transient*.

Theorem 4.4.6 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 4.4.7 (Null recurrence) The mean recurrent time of the isotropic, one-dimensional random walk is infinity: $\mathbf{E}[T] = +\infty$.

Proof In view of Theorem 4.4.1, setting $p = q = 1/2$, we obtain

$$\mathbf{E}(T) = 4\sum_{n=1}^{\infty} nC_{n-1}\left(\frac{1}{4}\right)^n = \sum_{n=0}^{\infty}(n+1)C_n\left(\frac{1}{4}\right)^n. \tag{4.8}$$

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

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

and hence $\mathbf{E}[T] = +\infty$ as desired, see also Remark 4.4.8. ∎

Remark 4.4.8 Let $a_n \geq 0$ for $n = 0, 1, 2, \ldots$ 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:

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

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

Problem 8 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 9 Let $\{X_n\}$ be a random walk starting from 0 with right-move p and left-move q. Show that

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

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

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

Problem 10 (Lemma 4.4.3) (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 11 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

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

5 Topics in One-Dimensional Random Walks

5.1 The Law of Long Lead

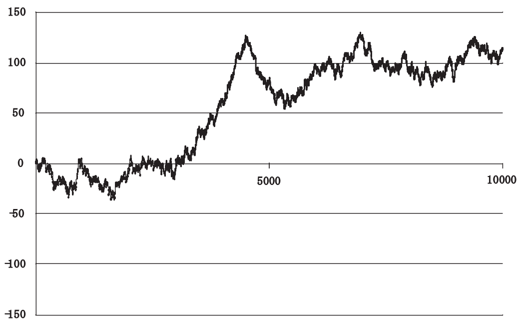
Let us consider an isotropic random walk $\{X_n\}$, namely, letting $\{Z_n\}$ be the Bernoulli trials such that

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

we set

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

Fig. 5.1 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.



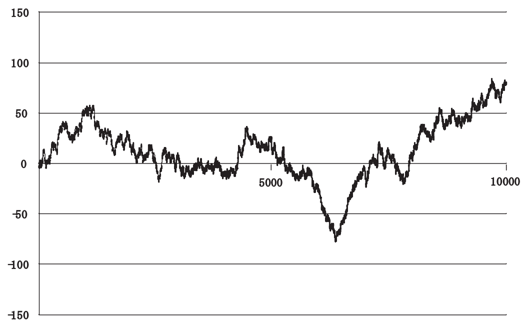


Figure 5.1: Sample paths of a random walk up to time 10000

By the law of large numbers we know that ± 1 occur almost 5000 times. In fact, we see from the value of X_{10000} that ± 1 occur 5000 ± 50 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

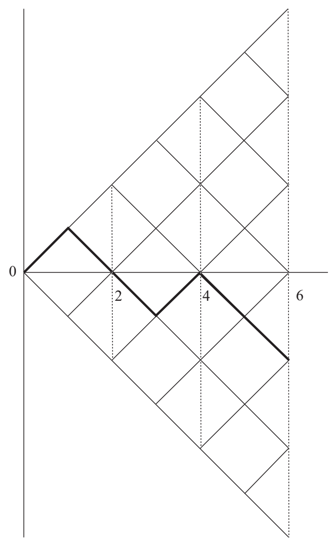
$$W(4|4) = 6 \times \left(\frac{1}{2}\right)^4, \qquad W(2|4) = 4 \times \left(\frac{1}{2}\right)^4, \qquad W(0|4) = 6 \times \left(\frac{1}{2}\right)^4,$$

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

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

Theorem 5.1.1 For $n = 1, 2, \dots$ it holds that

$$W(2k|2n) = \binom{2k}{k}\binom{2n-2k}{n-k}\left(\frac{1}{2}\right)^{2n}, \qquad k = 0, 1, \dots, n. \tag{5.1}$$



Recall that

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

Then (5.1) becomes

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

The proof is, however, not so simple. (In fact, a very tricky simple observation might lead to the result.) The complete proof is found in [Feller]. As before, we set

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

Observing that

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

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 $\frac{H_{2n}}{2n}$ rather than H_{2n} it self. As we have already obtained

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

for $0 < a < 1$ we see that

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

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:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \qquad n \to \infty,$$

we obtain

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

Then,

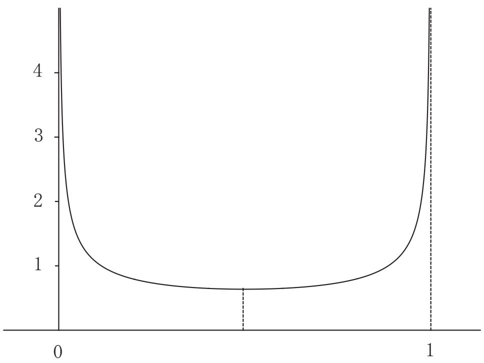
$$\begin{aligned} P\left(a \le \frac{H_{2n}}{2n} \le b\right) &\sim \sum_{k=0}^{n} \chi_{[a,b]}\left(\frac{k}{n}\right)\frac{1}{\pi\sqrt{k(n-k)}} \\ &= \sum_{k=0}^{n} \chi_{[a,b]}\left(\frac{k}{n}\right)\frac{1}{\pi\sqrt{\frac{k}{n}(1-\frac{k}{n})}}\,\frac{1}{n} \\ &\to \int_0^1 \chi_{[a,b]}(x)\frac{dx}{\pi\sqrt{x(1-x)}} = \int_a^b \frac{dx}{\pi\sqrt{x(1-x)}}. \end{aligned}$$

Definition 5.1.2 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 5.1.3 The distribution of $\dfrac{H_{2n}}{2n}$ converges weakly to the arcsine law:

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

For example,

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

Namely, during the long game, the probability that the ratio of winning time exceeds 90% is $1 - F(0.9) \approx 0.205$, which sounds larger than one expects.

Problem 12 For the isotropic random walk ($p = q = 1/2$) derive the following relations.

1. $P(X_1 \neq 0, X_2 \neq 0, \dots, X_{2n-1} \neq 0, X_{2n} \neq 0) = p_{2n}$ for $n = 1, 2, \dots$.
2. $P(X_1 \ge 0, X_2 \ge 0, \dots, X_{2n-1} \ge 0, X_{2n} \ge 0) = p_{2n}$ for $n = 1, 2, \dots$.

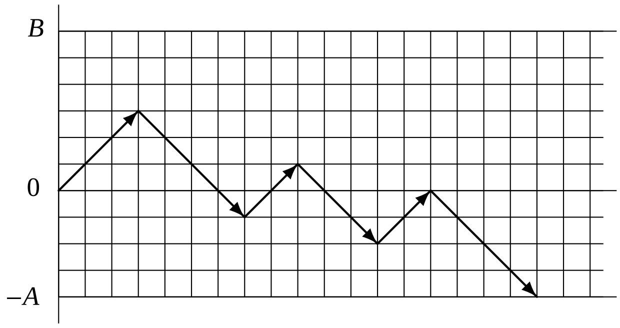
5.2 Gambler's Ruin

We introduce one-dimensional random walks with boundaries. Let us consider a random walker starting from the origin 0 at time $n = 0$. Now there are barriers at the positions $-A$ and B ($A \ge 1$, $B \ge 1$). If the random walker touches the barrier, it remains there afterward. In this sense the positions $-A$ and B are called *absorbing barriers*.

Let $Z_1, Z_2, \dots$ be Bernoulli trials with success probability $0 < p < 1$. Define a discrete time stochastic process $X_0, X_1, X_2, \dots$ by

$$X_0 = 0, \qquad X_n = \begin{cases} X_{n-1} + Z_n\,, & -A < X_{n-1} < B, \\ -A, & X_{n-1} = -A, \\ B, & X_{n-1} = B. \end{cases} \tag{5.2}$$

This $\{X_n\}$ is called a *random walk with absorbing barriers*.



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 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 5.2.1 $\{R_k\,;\,, -A \le k \le B\}$ fulfills the following difference equation:

$$R_k = pR_{k+1} + qR_{k-1}\,, \qquad R_{-A} = 1, \qquad R_B = 0. \tag{5.3}$$

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

$$S_k = pS_{k+1} + qS_{k-1}\,, \qquad S_{-A} = 0, \qquad S_B = 1. \tag{5.4}$$

Theorem 5.2.2 Let $A \ge 1$ and $B \ge 1$. Let $\{X_n\}$ be the random walk with absorbing barriers at $-A$ and B, which is defined by (5.2). Then the probabilities that $\{X_n\}$ is absorbed at the barriers are given by

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

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

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

An interpretation of Theorem 5.2.2 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 5.2.2 that the winning probability of A and B are given by

$$P(A) = \frac{A}{A+B}, \qquad P(B) = \frac{B}{A+B}, \tag{5.5}$$

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 4.4.6. 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. (5.5) 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 5.2.3 Let $\{X_n\}$ be the same as in Theorem 5.2.2. 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

$$\mathbf{E}(Y_{-A}) = \mathbf{E}(Y_B) = 0. \tag{5.6}$$

For $-A < k < B$ we have

$$\mathbf{E}(Y_k) = \sum_{j=1}^{\infty} jP(Y_k = j). \tag{5.7}$$

In a similar manner as in the proof of Theorem 5.2.2 we note that

$$P(Y_k = j) = pP(Y_{k+1} = j-1) + qP(Y_{k-1} = j-1). \tag{5.8}$$

Inserting (5.8) into (5.7), we obtain

$$\begin{aligned} \mathbf{E}(Y_k) &= p\sum_{j=1}^{\infty} jP(Y_{k+1} = j-1) + q\sum_{j=1}^{\infty} jP(Y_{k-1} = j-1) \\ &= p\mathbf{E}(Y_{k+1}) + q\mathbf{E}(Y_{k-1}) + 1. \end{aligned} \tag{5.9}$$

Thus, $\mathbf{E}(Y_k)$ is the solution to the difference equation (5.9) with boundary condition (5.6). 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 13 (1) Explain about the general solutions to the difference equation:

$$x_{n+2} + ax_{n+1} + bx_n = c,$$

where a, b, c are constant numbers.

(2) Solve the difference equation (5.9) with boundary condition (5.6).

5.3 Variants of Boundary Conditions

There is another type of barrier called a *reflecting barrier*. A random walk touches the reflecting barrier, it changes the direction in the next step and continue to move. Let $Z_1, Z_2, \dots$ be Bernoulli trials with success probability $0 < p < 1$. Consider barriers at positions $-A$ and B, $A \geq 1$, $B \geq 1$. Define $X_0, X_1, X_2, \dots$ by

$$X_0 = 0, \qquad X_n = \begin{cases} X_{n-1} + Z_n, & -A < X_{n-1} < B, \\ -A + 1, & X_{n-1} = -A, \\ B - 1, & X_{n-1} = B. \end{cases} \tag{5.10}$$

Then $\{X_n\}$ is called a random walk with reflecting barriers.

Problem 14 Let $\{X_n\,;\, n = 0, 1, 2, \dots\}$ be an isotropic random walk on the half line $\{0, 1, 2, \dots\}$ starting from the origin 0 at time $n = 0$, where the origin is a reflecting barrier. Find $P(X_{2n} = 0)$.

As another boundary condition, we consider a random walk $\{X_n\}$ on the half-line $\{0, 1, 2, \dots\}$ starting from the origin 0 at time $n = 0$. When the random walker sits at one of $\{1, 2, \dots\}$, it moves to the right with probability p and to the left with $q = 1 - p$. When it sits at 0, it moves to the right with probability p and stay there with $q = 1 - p$. Let T be the first hitting time of $\{X_n\}$ to 0, i.e.,

$$T = \inf\{n \geq 1\,;\, X_n = 0\}.$$

Then we have

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

Problem 15 Let $\{X_n\,;\, n = 0, 1, 2, \dots\}$ be the random walk as above.

(1) Show that $P(T < \infty) = 1$ for $q \geq 1/2$ and $P(T < \infty) = 2q$ for $q < 1/2$.

(2) Calculate $\mathbf{E}[T]$.

6 Markov Chains

Let us recall a typical property of a random walk: the next position of the walker is determined probabilistically only by the present position. Namely, the next-step movement is independent of the past trajectories. As the position of the one-dimensional random walk is described in terms of the usual coordinate system $\mathbb{Z} = \{\cdots, -1, 0, 1, 2, \dots\}$, the random walk is formulated as a discrete time stochastic process $\{X_n\}$ taking values in $\mathbb{Z}$. In this sense we call $\mathbb{Z}$ a *state space*. For wider applications a state space is not necessarily a set of numbers, but may be an arbitrary set. Keeping the typical property of the random walk and generalizing the state space, we come to the concept of *Markov chain*.

6.1 Conditional Probability

For two events A, B we define

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{6.1}$$

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 (6.1) is often used in the following form:

$$P(A \cap B) = P(B)P(A|B) \tag{6.2}$$

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. 6.1.

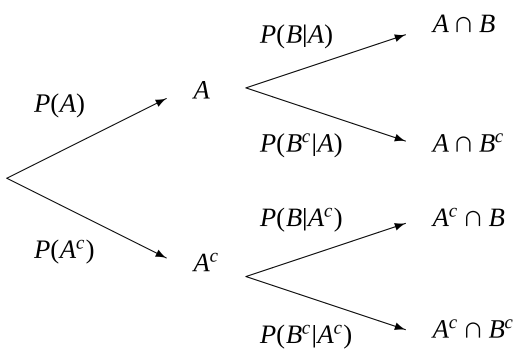


Figure 6.1: Tree diagram

Theorem 6.1.1 (Compound probabilities) For events $A_1, A_2, \dots, A_n$ we have

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}). \tag{6.3}$$

Proof Straightforward by induction on n. ∎

6.2 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 6.2.1 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 6.2.2 Let $r \ge 1$ and $s \ge 1$ 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 n-th trial and $X_n = 0$ if a white ball is picked up at the n-th trial. Then $\{X_1, X_2, \dots, X_N\}$ is a stochastic process. We note that

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

and

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

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

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

which stands for the number of black balls picked up during the first n trials. We the see easily that $\{Y_n\}$ is a Markov chain.

Definition 6.2.3 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 6.2.4 A matrix $P = [p_{ij}]$ with index set S is called a *stochastic matrix* if

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

Theorem 6.2.5 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 $p(i, j) > 0$.

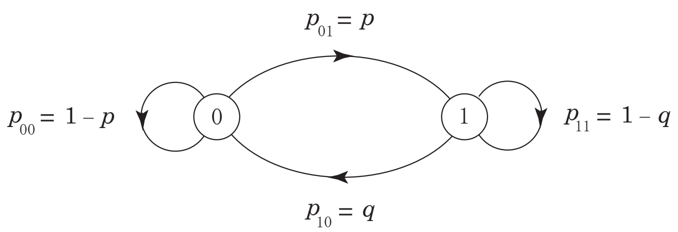
Example 6.2.6 (2-state Markov chain) A Markov chain over the state space $\{0, 1\}$ is determined by the transition probabilities:

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

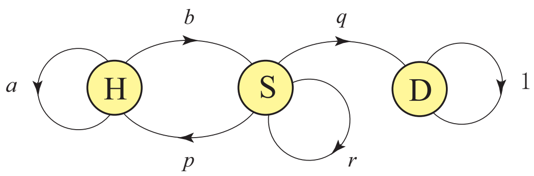
The transition matrix is defined by

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

The transition diagram is as follows:



Example 6.2.7 (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 6.2.8 (Random walk on $\mathbb{Z}^1$) 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 6.2.9 (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 6.2.10 (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

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

6.3 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

$$\pi_i \geq 0 \quad \text{and} \quad \sum_{i \in S} \pi_i = 1. \tag{6.4}$$

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 6.3.1 (Chapman–Kolmogorov equation) For $0 \le r \le n$ we have

$$p_n(i, j) = \sum_{k \in S} p_r(i, k) p_{n-r}(k, j). \tag{6.5}$$

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 (6.5). ∎

Applying (6.5) repeatedly and noting that $p_1(i, j) = p(i, j)$, we obtain

$$p_n(i, j) = \sum_{k_1, \dots, k_{n-1} \in S} p(i, k_1) p(k_1, k_2) \cdots p(k_{n-1}, j). \tag{6.6}$$

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 6.3.2 For $m, n \ge 0$ 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 6.3.1. ∎

Remark 6.3.3 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 6.3.4 We have

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

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 6.3.5 (2-state Markov chain) Let $\{X_n\}$ be the Markov chain introduced in Example 6.2.6. 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

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

Here let us observe the limit as $n \to \infty$. Assume that $0 < p+q < 2$, or equivalently, $|r| < 1$. Then

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

and

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

Note that

$$\begin{bmatrix} \frac{q}{p+q} & \frac{p}{p+q} \end{bmatrix} P = \begin{bmatrix} \frac{q}{p+q} & \frac{p}{p+q} \end{bmatrix}.$$

This means that the distribution $\begin{bmatrix} \frac{q}{p+q} & \frac{p}{p+q} \end{bmatrix}$ is invariant under the Markov chain.

Problem 16 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. When will the party B command a majority? Moreover, find the final ratio of supporters after many elections if the same situation continues.

7 Topics in Markov Chains

7.1 Stationary Distributions

Definition 7.1.1 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

$$\pi = \pi P, \tag{7.1}$$

or equivalently,

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}\,, \qquad j \in S. \tag{7.2}$$

Thus, to find a stationary distribution we need to solve (7.1) (or equivalently (7.2)) together with (6.4). If S is a finite set, finding stationary distributions is reduced to a simple linear system.

Example 7.1.2 (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,

$$\pi = \left[\frac{q}{p+q}\,, \frac{p}{p+q} \right]$$

is a distribution on $S = \{0, 1\}$, so it is a stationary distribution. In this case a stationary distribution is unique. Note that the stationary distribution is obtained as a limit distribution, see Example 6.3.5. In the uninteresting case of $p = q = 0$, any $\pi = [\pi_0\,, \pi_1]$ is a stationary distribution.

Example 7.1.3 (3-state Markov chain) We discuss the Markov chain $\{X_n\}$ introduced in Example 6.2.7. If $q > 0$ and $b > 0$, a stationary distribution is unique and given by $\pi = [0\ \ 0\ \ 1]$.

Example 7.1.4 (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

$$\pi(k) = p\pi(k-1) + q\pi(k+1), \qquad k \in \mathbf{Z}. \tag{7.3}$$

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 (7.3) 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 (7.3) 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 7.1.5 (One-dimensional RW with reflection barrier) There is a unique stationary distribution when $p < q$. In fact,

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

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

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

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

Theorem 7.1.6 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.

We are going into a discussion about unique existence of a stationary distribution.

Definition 7.1.7 We say that a state j *can be reached from* a state i if there exists some $n \ge 0$ 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 $m \ge 0$ and $n \ge 0$ such that $p_n(i, j) > 0$ and $p_m(j, i) > 0$.

Lemma 7.1.8 For two states $i, j \in S$ we define a binomial relation $i \sim j$ when they intercommunicate. Then $\sim$ becomes an equivalence relation on S, namely,

(i) $i \sim i$;

(ii) $i \sim j$ implies $j \sim i$;

(iii) If $i \sim j$ and $j \sim k$, then $i \sim k$.

Proof (i), (ii) are obvious by definition. (iii) is verified by the Chapman-Kolmogorov equation. ∎

Thereby the state space S is classified into a disjoint set of equivalence classes determined by the above $\sim$. Namely, each equivalence class consists of states which intercommunicate each other.

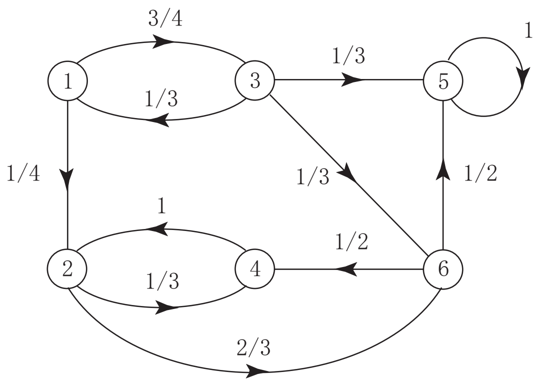
Definition 7.1.9 A state i is called *absorbing* if

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

In particular, an absorbing state is a state which constitutes an equivalence class by itself.

Definition 7.1.10 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.

Example 7.1.11 Examine the equivalence relation among the states of a Markov chain described by the following transition diagram:



Then we have the following fundamental result.

Theorem 7.1.12 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 7.2.9).

Finally, the distribution of a Markov chain does not necessarily converge to a stationary distribution even if it exists uniquely.

Example 7.1.13 Consider a Markov chain determined by the transition matrix:

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

We first note that there exists a unique stationary distribution. But for a given initial distribution $\pi(0)$ it is not necessarily true that $\lim_{n\to\infty} \pi(n)$ 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 7.1.14 For a state $i \in S$,

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

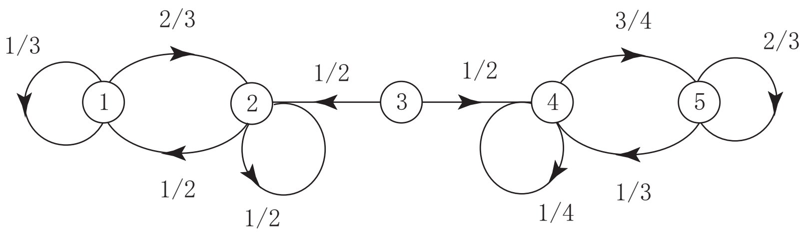
is called the *period* of i. (When the set in the right-hand side is empty, the period is not defined.) A state $i \in S$ is called *aperiodic* if its period is one.

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

Theorem 7.1.15 Let π be a stationary distribution of an irreducible Markov chain on a finite state space (It is unique, see Theorem 7.1.12). If $\{X_n\}$ is aperiodic, for any $j \in S$ we have

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

Problem 17 Consider a Markov chain determined by the transition diagram below.



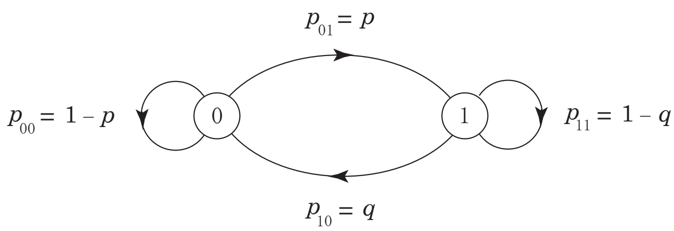
(1) Is the Markov chain irreducible?

(2) Find all stationary distibutions.

Problem 18 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 $\pi_0 = [q/(p+q) \;\; p/(p+q)]$. Calculate the following statistical quantities:

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

Problem 19 Let $\{X_n\}$ be a Markov chain described by the following transition diagram:



where $p > 0$ and $q > 0$. For a state $i \in S$, 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 $T_i = \infty$.)

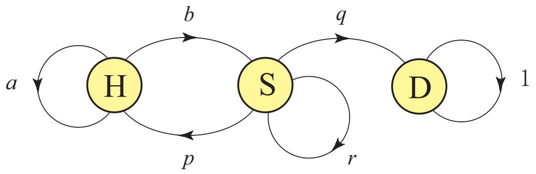
(1) Calculate

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

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

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

Problem 20 Let $\{X_n\}$ be the Markov chain introduced in Example 6.2.7:



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

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

(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:

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

7.2 Recurrence

Definition 7.2.1 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 7.2.2 A state $i \in S$ is recurrent if and only if

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

Proof (basically the same as the proof of recurrence of one-dimensional random walk) We first put

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

$p_n(i, j)$ 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

$$p_n(i, j) = \sum_{r=1}^{n} f_r(i, j) p_{n-r}(j, j), \qquad i, j \in S, \quad n = 1, 2, \ldots. \tag{7.4}$$

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 (7.4) we see easily that

$$G_{ij}(z) = p_0(i, j) + F_{ij}(z) G_{jj}(z). \tag{7.5}$$

Setting $i = j$ in (7.5), we obtain

$$G_{ii}(z) = 1 + F_{ii}(z) G_{ii}(z).$$

Hence,

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

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. ∎

During the above proof we have already established the following

Theorem 7.2.3 If a state i is transient, we have

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

and

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

Example 7.2.4 (random walk on $\mathbb{Z}$) 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 $p_{2n}(0,0)$. On the other hand, we know that

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

see Chapter (4.1.1). Using the Stirling formula:

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \tag{7.6}$$

we have

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

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 7.2.5 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 $c_1 a_n \leq b_n \leq c_2 a_n$. Hence $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ converge or diverge at the same time.

Example 7.2.6 (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:

$$\sum_{i=0}^{n} \binom{n}{i}^2 = \binom{2n}{n}, \tag{7.7}$$

wich 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 7.2.7 (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,

$$\sum_{i+j+k=n} \frac{n!}{i!j!k!} = 3^n. \tag{7.8}$$

Second, the maximum value

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

is attained when $\dfrac{n}{3} - 1 \le i, j, k \le \dfrac{n}{3} + 1$ 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).

A state i is called *recurrent* if $P(T_i < \infty | X_0 = i) = 1$. In this case we are interested in the mean value $\mathbf{E}(T_i | X_0 = i)$ (mean recurrent time). As we have already shown (Theorem **??**), the mean recurrent time of the one-dimensional isotropic random walk is infinity although it is recurrent. In this case the state is called *null recurrent*. On the other hand, if $\mathbf{E}(T_i | X_0 = i) < \infty$ the state i is called *positive recurrent*.

Theorem 7.2.8 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 7.2.9 For an irreducible Markov chain on a finite state space S, every state is positive recurrent.

レポート提出要領

1. 講義中に出題したレポート問題のうち

   1 番から 11 番のうち 2 題, 12 番から最終番のうち 2 題

   の合わせて 4 題を選択して解答せよ (1 題 25 点で採点). コピーレポートは 0 点.
2. 提出先：情報科学研究科 1 F 事務室前のメールボックスに設置するレポート提出専用のボックス
3. 提出期間 (期間外には受け取りません)：2016 年 2 月 1 日（月）–2 月 5 日（金）

Examination

1. Choose and answer 2 problems among No. 1–11 and 2 problems among No. 12– the last. Each is allotted 25 points. Plagiarizing is excluded from evaluation.
2. Submit to: the mailbox prepared for report submission at 1F in front of the administrative office of GSIS.
3. Submission period: February 1 (Mon)– 5 (Fri), 2016.

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.

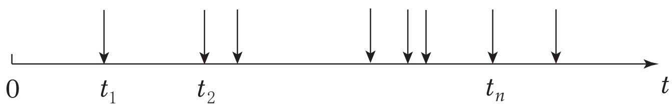


Figure 8.1: Recording when the event E occurs

There are two quantities which we measure.

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

(ii) The waiting time of the n-th 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.

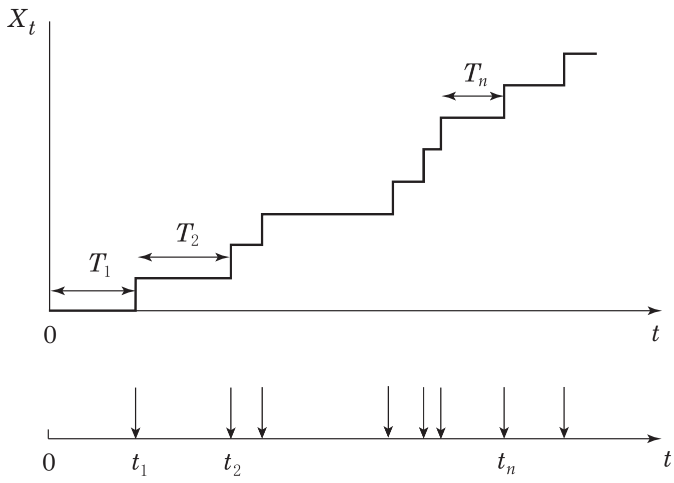
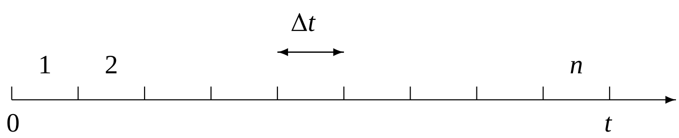


Figure 8.2: The counting process and waiting times



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 i-th 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 21 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 22 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 $\{X_t\,;\,t \ge 0\}$ 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

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

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.,

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

In a similar fashion, we set

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

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,

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

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 23 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 $P(S_2 \le t)$ 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 $t \ge 0$,

$$X_t = \max\{n \ge 0\,;\,S_n \le t\}.$$

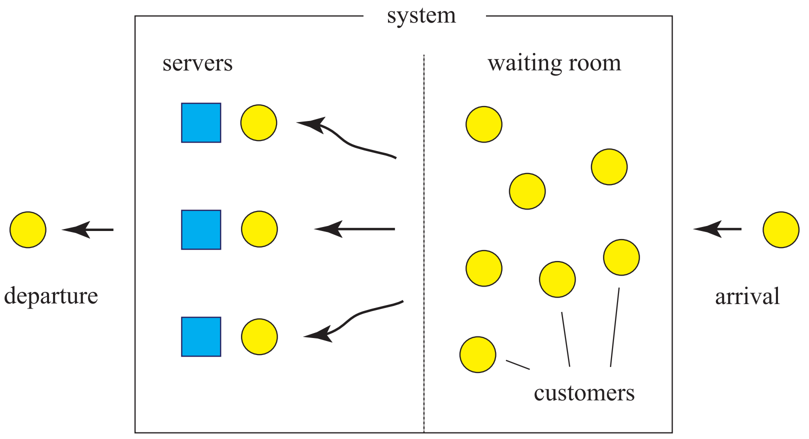
It is obvious that for each $t \ge 0$, X_t is a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. In other words, $\{X_t\,;\,t \ge 0\}$ becomes a continuous time stochastic process. This is called *Poisson process* with parameter λ by definition.

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:

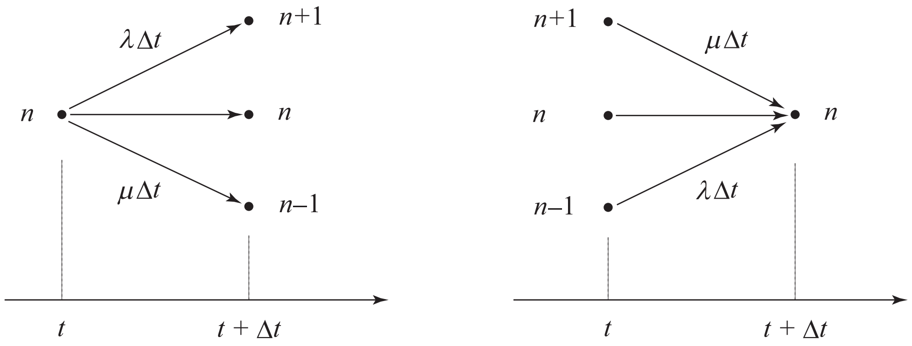
(i) arrivals occur according to a Poisson process with parameter λ;
(ii) service times obey an exponential distribution with parameter μ;
(iii) arrivals of customers and service times are independent;
(iii) the system contains a single server;
(iv) the size of waiting room is infinite;
(v) (FIFO = First 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:

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

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

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

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 24 ($M/M/1/1$ 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:

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

10 Brownian Motion — An Intuitive Introduction

In 1827 Robert Brown (1773–1858, Scottish botanist) observed a continuous jittery motion of small particles spouting from pollen of the plant *Clarkia pulchella* in water under a microscope. For a long period the mechanism of this motion was unknown. In 1905 (known as the miracle year in physics) Albert Einstein published a paper that the Brownian motion was caused by individual water molecules and was given a mathematical description along with physical discussion. The original article is collected in A. Einstein: "Investigations on the Theory of the Brownian Movement," (Dover, 2011). It is probably fair to refer to Marian Smoluchowski (1872–1917, Polish physicist) who also gave a similar mathematical model of Brownian motion. Although his paper was published in 1906, it is said in the exhibition at Krakow University, Poland, that his manuscript was sent to Einstein before his work.

After the physical investigations mathematical study of Brownian motion or more general stochastic processes started. The contributions by Norbert Wiener (1894–1964) and Paul Lévy (1886–1971) were most essential. Thereby Brownian motion is also called the *Wiener process*. In 1940s Kiyoshi Itô (1915–2008, Japanese probabilist) initiated the theory of stochastic differential equations which is nowadays commonly called the *Itô calculus*. During the last 60 years the Itô calculus has developed drastically for vast applications. It is only a small part of the story that financial engineering without Itô formula is impossible and Itô became the most famous Japanese in Wall Street.

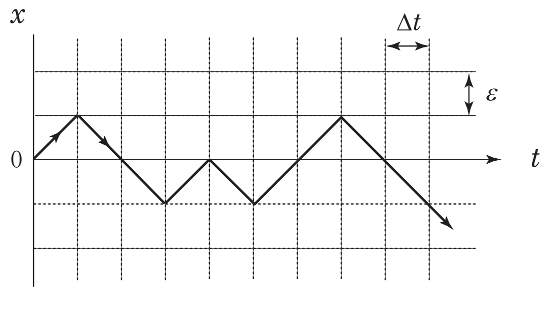
10.1 From Random Walk to Brownian Motion

Consider one-dimensional random walk, where the random walker starts from the origin $x = 0$ at time $t = 0$, and tosses a fair coin every short time interval Δt, and move a very short distance of ϵ to the right or left. Let $X(t) = X(t; \Delta t, \epsilon)$ be the position of the above random walker at time $t \geq 0$. It is convenient to express $X(t)$ by means of Bernoulli trials. Let $\{Z_k\}$ be a Bernoulli trial with success probability $1/2$, i.e.,

$$P(Z_k = +1) = P(Z_k = -1) = \frac{1}{2}.$$

Let n be the number of coin tosses during the unit time interval, i.e., $n\Delta t = 1$. Then we have

$$X(t) = X(t; \Delta t, \epsilon) = \sum_{k=1}^{nt} \epsilon Z_k\,. \tag{10.1}$$



We see easily that

$$\mathbf{E}[X(t)] = \epsilon \sum_{k=1}^{nt} \mathbf{E}[Z_k] = 0, \tag{10.2}$$

$$\mathbf{V}[X(t)] = \epsilon^2 \sum_{k=1}^{nt} \mathbf{V}[Z_k] = \epsilon^2 nt = \frac{\epsilon^2}{\Delta t}\, t. \tag{10.3}$$

We are interested in the limit as $\Delta t \to 0$ and $\epsilon \to 0$, whereas (10.3) suggests that a reasonable limit is obtained under the condition

$$\frac{\epsilon^2}{\Delta t} \to \alpha \text{ (constant)}. \tag{10.4}$$

In fact, a stochastic process obtained by

$$B(t) = \lim X(t; \Delta t, \epsilon) = \lim \sum_{k=1}^{nt} \epsilon Z_k \,, \tag{10.5}$$

where the limit is taken in such a way that

$$\Delta t \to 0, \quad \epsilon \to 0, \quad \frac{\epsilon^2}{\Delta t} \to 1,$$

is called the *Brownian motion* or the *Wiener process*. Since

$$X(t; \Delta t, \epsilon) = \sum_{k=1}^{nt} \epsilon Z_k = \epsilon \sqrt{n} \, \frac{1}{\sqrt{n}} \sum_{k=1}^{nt} Z_k = \sqrt{\frac{\epsilon^2}{\Delta t}} \, \frac{1}{\sqrt{n}} \sum_{k=1}^{nt} Z_k \,,$$

the Brownian motion is given by

$$B(t) = \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} Z_k \,. \tag{10.6}$$

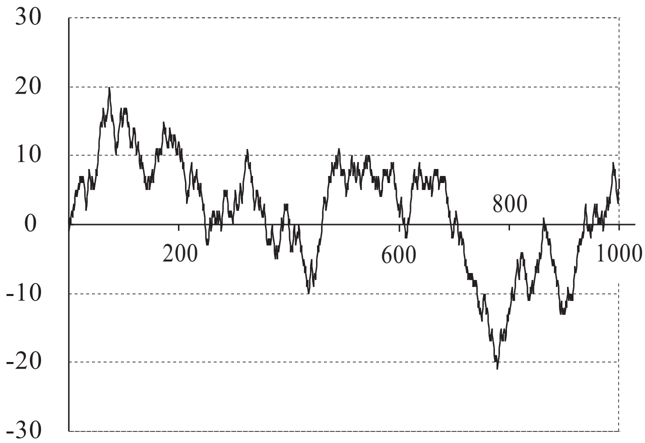


Figure 10.1: Random walk as a simulation of Brownian motion

Remark 10.1.1 The above "construction" of Brownian motion from the random walk is heuristic and instructive. The important step was omitted, namely, we did not argue that the limit of random walks $X(t; \Delta t, \epsilon)$ exists in the sense of stochastic process. In fact, we may give rigorous proof to this point but mathematically totally non-trivial. On the other hand, in most textbooks the Brownian motion is introduced independently of random walk, where the argument becomes much simpler but loses the intuition of the jittery movement of Brownian motion.

(B1) $B(0) = 0$;

(B2) $\mathbf{E}[B(t)] = 0$;

(B3) $\mathrm{Cov}(B(s), B(t)) = \mathbf{E}[B(s)B(t)] = \min\{s, t\}$, in particular, $\mathbf{V}[B(t)] = t$.

In fact, (B1) and (B2) are obvious from (10.6). (B3) is derived as follows:

$$\mathrm{Cov}\,(B(s), B(t)) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{[ns]} \sum_{k=1}^{[nt]} \mathrm{Cov}\,(Z_j, Z_k) = \lim_{n\to\infty} \frac{1}{n} \min\{[ns], [nt]\} = \min\{s, t\}.$$

Applying the central limit theorem to (10.6), we see that $B(t) \sim N(0, t)$. Moreover, we have

(B4) $\{B(t)\,;\, t \geq 0\}$ forms a Gaussian system, i.e., for any finite number of time points $t_1, t_2, \ldots, t_n$ the random vector $(B(t_1), B(t_2), \ldots, B(t_n))$ obeys a n-dimensional normal law (may be degenerate).

(B5) (Continuous sample path) $t \mapsto B(t)$ is continuous almost surely.

In fact, (B5) follows from (B1)–(B4) by the famous continuous version theorem due to Kolmogorov. In most literatures a stochastic process satisfying (B1)–(B5) is called a *Brownian motion* by definition. In this sense a stochastic process satisfying (B1)–(B4) is often called a *weak Brownian motion*.

Finally, we mention the following basic properties.

(B6) (Independent increments) for $0 \le t_1 < t_2 < \cdots < t_n$, the random variables

$$B(t_1),\ B(t_2) - B(t_1),\ B(t_3) - B(t_2),\ \dots,\ B(t_n) - B(t_{n-1})$$

are independent.

(B7) $\{B(t)\}$ is a Markov process. (For the precise definition of Markov process we need some advanced knowledge of conditional probability.)

10.2 Exponential Growth

Let x_n be the quantity to be measured at time $n = 0, 1, 2, \dots$. Suppose that x_n is obtained by constant multiplication of x_{n-1}, that is,

$$x_n = (1 + a)x_{n-1}\,, \quad n = 1, 2, \dots, \tag{10.7}$$

where a is a constant and $x_0 > 0$ is an initial value. The difference equation (10.7) is easily solved explicitly:

$$x_n = (1 + a)^n x_0\,. \tag{10.8}$$

If $a > 0$, then $\{x_n\}$ becomes an increasing sequence, while if $-1 < a < 0$, it becomes a decreasing sequence. In both cases a is called the *growth rate*.

An anecdote of Sorori Shin Zaemon (曽呂利新左衛門) Being asked by Hideyoshi, Shin Zaemon hoped to receive, as a reward, grains of rice everyday for one hundred days in such a way that 1 grain today, 2 grains on the next day, 4 grains on the next-to-next day, and in this manner, twice as many grains as on the day before. Shin Zaemon's desire was accepted by Hideyoshi easily, however, after a few days Hideyoshi recognized that his huge rice granary would become empty before the 100th day. In fact, setting $a = 1$ and $x_0 = 1$, we see that the total amount of the grains to be gifted to Shin Zaemon is

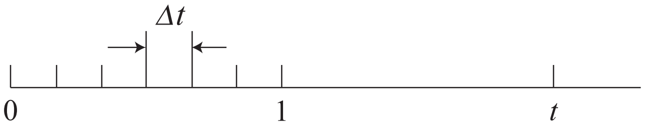
$$\sum_{n=0}^{99} 2^n = 2^{100} - 1 \approx 1.27 \times 10^{30} \approx 2.8 \times 10^{25}\text{kg}$$

(In recent years Japan's production of rice per year is about 8×10^9 kg.) This is an example of a catastrophic result caused by the *exponential growth*.

We would like to derive a continuous-time model from the discrete-time one discussed above. Let $x(t)$ be the quantity under consideration (e.g., population) at time $t \ge 0$. Dividing the unit time interval into n small intervals of length Δt, we approximate $x(t)$ by the discrete-time model $x_{\Delta t}(t)$ defined by

$$x_{\Delta t}(t) = (1 + a)^{nt} x_0\,, \tag{10.9}$$

where $a = a(\Delta t)$ is the growth rate for the small interval of length Δt and $n\Delta t = 1$.



Then the continuous-time model would be obtained by

$$x(t) = \lim x_{\Delta t}(t) = \lim(1 + a)^{nt} x_0\,, \tag{10.10}$$

where the limit is taken in such a way that $n \to \infty$ ($\Longleftrightarrow \Delta t \to 0$) with keeping $n\Delta t = 1$. In order to get a reasonable limit of (10.10), as is suggested by the formula

$$\lim_{n\to\infty}\left(1+\frac{z}{n}\right)^n = e^z,$$

we set

$$a = a(\Delta t) = \alpha\Delta t = \frac{\alpha}{n}\,, \tag{10.11}$$

where α is called the *continuous growth rate*. Then we have

$$x(t) = \lim_{n\to\infty}\left(1+\frac{\alpha}{n}\right)^{nt} x_0 = e^{\alpha t}x_0\,, \tag{10.12}$$

which is the so-called *exponential growth*. Of course, if $\alpha > 0$ it is exponentially increasing, while if $\alpha < 0$ it is exponentially decreasing. Note also that $x(t)$ in (10.12) obeys the differential equation:

$$\frac{dx(t)}{dt} = \alpha x(t), \qquad x(0) = x_0, \tag{10.13}$$

which shares a formal similarity with (10.7).

10.3 Exponential Growth with Fluctuation

We will generalize the discrete-time model (10.7), where the growth rate a is not a constant but is a random variable. As a simple case we assume that the growth rate is given by $a \pm \epsilon$ where $\pm$ is chosen randomly at each time step. To formulate the situation, letting $Z_1, Z_2, \dots$ be independent identically distributed random variables such that

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

we set

$$X_n = (1 + a + \epsilon Z_n)X_{n-1}\,, \qquad X_0 = x_0 \quad \text{(positive constant)}. \tag{10.14}$$

The difference equation (10.14) is easily solved:

$$X_n = x_0 \prod_{k=1}^{n} (1 + a + \epsilon Z_k). \tag{10.15}$$

Note that the mean value and the variance of the growth rate are given by

$$\mathbf{E}[a + \epsilon Z_n] = a, \qquad \mathbf{V}[a + \epsilon Z_n] = \epsilon^2.$$

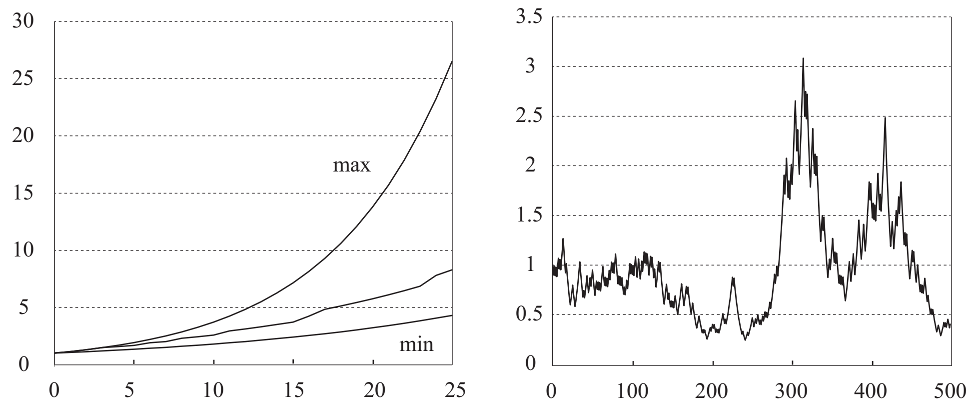


Figure 10.2: Exponential growth with growth rate $a = 0.1 \pm 0.04$ and $a = \pm 0.05$

10.4 Geometric Brownian Motion

We follow the argument in Section 10.2. Let $X_{\Delta t}(t)$ be the discrete-time model corresponding to the small time interval Δt. The growth rate is taken to be

$$\alpha\Delta t + \epsilon Z_k\,.$$

where the first term is already suggested by the argument in Section 10.1, while ϵ is to be specified. It then follows from (10.15) that

$$X_{\Delta t}(t) = x_0 \prod_{k=1}^{nt} (1 + \alpha\Delta t + \epsilon Z_k).$$

We first calculate the mean value of $X_{\Delta t}(t)$:

$$\mathbf{E}[X_{\Delta t}(t)] = \mathbf{E}\Big[x_0 \prod_{k=1}^{nt} (1 + \alpha\Delta t + \epsilon Z_k)\Big] = x_0 \prod_{k=1}^{nt} \mathbf{E}[1 + \alpha\Delta t + \epsilon Z_k] = x_0(1+\alpha\Delta t)^{nt},$$

where we used the independence of $\{Z_k\}$. Hence we have

$$\lim \mathbf{E}[X_{\Delta t}(t)] = x_0 e^{\alpha t}\,.$$

We next consider the variance. We start with

$$\mathbf{E}[X_{\Delta t}(t)^2] = x_0^2 \prod_{k=1}^{nt} \mathbf{E}[(1 + \alpha\Delta t + \epsilon Z_k)^2] = x_0^2\{(1+\alpha\Delta t)^2 + \epsilon^2\}^{nt}\,.$$

In view of $n\Delta t = 1$ we obtain

$$\lim \mathbf{E}(X_{\Delta t}(t)^2) = x_0^2 \lim \left(1 + \frac{2\alpha}{n} + \frac{\alpha^2}{n^2} + \epsilon^2\right)^{nt}\,.$$

Therefore, in order to obtain a reasonable limit we set

$$\epsilon^2 = \frac{\sigma^2}{n}\,.$$

Then we have

$$\lim \mathbf{E}[X_{\Delta t}(t)^2] = x_0^2 e^{(2\alpha+\sigma^2)t}$$

and

$$\lim \mathbf{V}[X_{\Delta t}(t)] = \lim \mathbf{E}[X_{\Delta t}(t)^2] - \lim \mathbf{E}[X_{\Delta t}(t)]^2 = x_0^2 e^{(2\alpha+\sigma^2)t} - x_0^2 e^{2\alpha t} = x_0^2 e^{2\alpha t}(e^{\sigma^2 t} - 1).$$

Thus it is reasonable to assume that the growth rate during the small time interval Δt should be of the form:

$$\alpha\Delta t + \sigma\sqrt{\Delta t}\,Z_k\,, \tag{10.16}$$

where $\alpha \in \mathbf{R}$ and $\sigma > 0$ are constant. We note that the fluctuation of the growth rate is proportional to $\sqrt{\Delta t}$.

Having specified the growth rate as in (10.16), we come back to the continuous model. The discrete-time approximation is given by

$$X_{\Delta t}(t) = x_0 \prod_{k=1}^{nt} (1 + \alpha\Delta t + \sigma\sqrt{\Delta t}\,Z_k). \tag{10.17}$$

We are interested in the limit $X(t) = \lim X_{\Delta t}(t)$. We see from (10.17) that

$$\log \frac{X_{\Delta t}(t)}{x_0} = \sum_{k=1}^{nt} \log(1 + \alpha\Delta t + \sigma\sqrt{\Delta t}\,Z_k). \tag{10.18}$$

Using the Taylor expansion of $log(1 + x)$ and noting that $Z_k^2 = 1$ ($Z_k = \pm 1$), we have

$$\begin{aligned}\log(1 + \alpha\Delta t + \sigma\sqrt{\Delta t}\,Z_k) &\approx (\alpha\Delta t + \sigma\sqrt{\Delta t}\,Z_k) - \frac{1}{2}(\alpha\Delta t + \sigma\sqrt{\Delta t}\,Z_k)^2 \\ &\approx \left(\alpha - \frac{\sigma^2}{2}\right)\Delta t + \sigma\sqrt{\Delta t}\,Z_k\,.\end{aligned}$$

Then (10.18) becomes

$$\log \frac{X_{\Delta t}(t)}{x_0} \approx \left(\alpha - \frac{\sigma^2}{2}\right) t + \sigma \sqrt{\Delta t} \sum_{k=1}^{nt} Z_k \,. \tag{10.19}$$

Recall that

$$B_t = \lim_{n\to\infty} \sqrt{\Delta t} \sum_{k=1}^{nt} Z_k = \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=1}^{nt} Z_k \,, \qquad t > 0,$$

is the *Brownian motion*, see Section 10.1. Then we see from (10.19) that

$$\lim \log \frac{X_{\Delta t}(t)}{x_0} = \left(\alpha - \frac{\sigma^2}{2}\right) t + \sigma B_t \,,$$

and consequently,

$$X(t) = x_0 \exp\left\{\left(\alpha - \frac{\sigma^2}{2}\right) t + \sigma B_t\right\}. \tag{10.20}$$

Definition 10.4.1 A stochastic process of the form $\exp\{at + \sigma B_t\}$ is called a *geometric Brownian motion*.

Definition 10.4.2 The probability distribution of $e^{\sigma Z}$ with $Z \sim N(0,1)$ is called the *log-normal distribution*. The density function of $e^{\sigma Z}$ is given by

$$f(x) = \begin{cases} \dfrac{x^{-1}}{\sqrt{2\pi\sigma^2}} \exp\left\{\dfrac{-(\log x)^2}{2\sigma^2}\right\}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

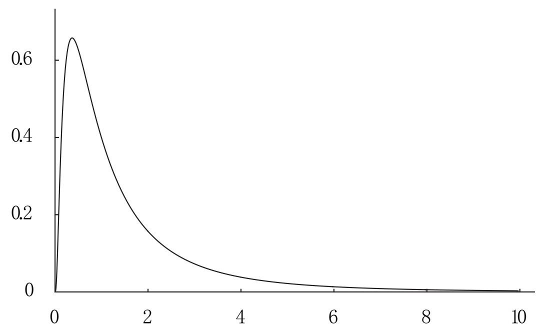
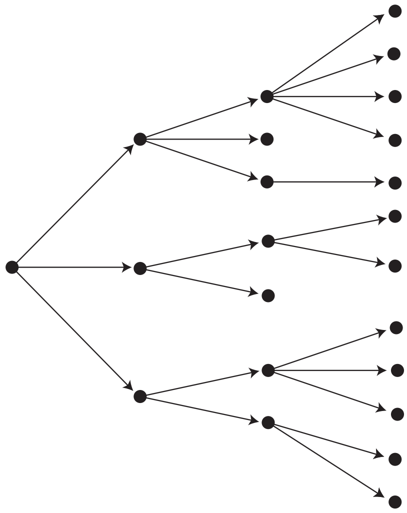


Figure 10.3: Log-normal distribution ($\sigma = 1$)

Remark 10.4.3 The geometric Brownian motion is called the *Black–Scholes model* in Mathematical Finance. It models the time series of stock price and is the basis of the so-called Black–Scholes formula for the price of European-style options, which appeared first in the paper by Fischer Black and Myron Scholes in 1973. Later Robert C. Merton published a paper expanding the mathematical understanding of the options pricing model, and coined the term Black–Scholes options pricing model. Merton and Scholes received the 1997 Nobel Memorial Prize in Economic Sciences for their work. Though ineligible for the prize because of his death in 1995, Black was mentioned as a contributor by the Swedish Academy.

11 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)

11.1 Definition

Let X_n be the number of individuals of the n-th 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

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

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

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

and

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

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 25). 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 25 (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

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

11.2 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

$$f(s) = \sum_{k=0}^{\infty} p_k s^k. \tag{11.1}$$

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 11.2.1

$$\sum_{j=0}^{\infty} p(i, j) s^j = [f(s)]^i, \qquad i = 1, 2, \dots. \tag{11.2}$$

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

$$\sum_{j=0}^{\infty} p_n(i, j)s^j = [f_n(s)]^i, \qquad i = 1, 2, \dots. \tag{11.3}$$

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

$$\sum_{j=0}^{\infty} p(i, j)s^j = [f(s)]^i, \qquad i = 1, 2, \dots, \tag{11.4}$$

Which was shown in Lemma 11.2.1. Suppose that $n \geq 1$ and the claim (11.3) 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 (11.4) 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 11.2.3 Assume that the mean value of the offspring distribution is finite:

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

Then we have

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

Proof Differentiating (11.1), we obtain

$$f'(s) = \sum_{k=0}^{\infty} k p_k s^{k-1}, \qquad |s| < 1. \tag{11.5}$$

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

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

On the other hand, setting $i = 1$ in (11.3), we have

$$\sum_{j=0}^{\infty} p_n(1, j)s^j = f_n(s) = f_{n-1}(f(s)). \tag{11.6}$$

Differentiating both sides, we come to

$$f_n'(s) = \sum_{j=0}^{\infty} j p_n(1,j) s^{j-1} = f_{n-1}'(f(s)) f'(s). \tag{11.7}$$

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

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

∎

In conclusion, the mean value of the number of individuals in the n-th 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 26 Assume that the variance of the offspring distribution is finite: $\mathbf{V}[Y] = \sigma^2 < \infty$. By similar argument as in Theorem 11.2.3, prove that

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

11.3 Extinction Probability

The event $\{X_n = 0\}$ means that the family died out until the n-th 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

$$q = \lim_{n\to\infty} P(X_n = 0). \tag{11.8}$$

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 11.3.1 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:

$$q = f(q). \tag{11.9}$$

Proof It follows from Lemma 11.2.2 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 (11.8). The second assertion follows since $f(s)$ is a continuous function on $[0, 1]$. ∎

Lemma 11.3.2 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 11.3.3 (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 11.3.4 $f_1(0) \le f_2(0) \le \cdots \to q$.

Theorem 11.3.5 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 27 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 $f_n(s)$.

Probabilistic Models・Applied Analysis

Preparing Reports

1. You are requested to show mathematical answers of the proposed problems.
2. It is not supposed to solve the problems without seeking any references. Rather you are suggested to thouroughly study the problems with proper references.
3. It is also encouraged to discuss with your collegues. However, a copy report will be excluded from marking. You must know the essential difference between discussing with collegues and copying results.
4. Seeking similar problems in some websites or books, you might copy the answer. Posting a question in websites, you might expect someone to answer. The proposed problems are original in the sense that they are not cited directly from any references; however, being elementary and fundamental, similar problems are easily created by experts.
5. Understanding the above situation, each problem is marked out of 25 points.

Appraisal standard

Example: Two dice are rolled. Let X be the sum of the slots, and Y the product. Calculate the correlation coefficient of X, Y.

A typical answer would be as follows.

(1) Describe the joint probability $P(X = x, Y = y)$.

(2) Calculate the marginal distributions and obtain the table for $P(X = x)$ and $P(Y = y)$.

(3) With the above tables calculate $\mathbf{E}[X]$, $\mathbf{E}[X^2]$, $\mathbf{V}[X]$, $\mathbf{E}[Y]$, $\mathbf{E}[Y^2]$, $\mathbf{V}[Y]$, $\mathbf{E}[XY]$ and then

$$\mathbf{Cov}\,(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y], \qquad \rho(X, Y) = \frac{\mathbf{Cov}\,(X, Y)}{\sqrt{\mathbf{V}[X]}\,\sqrt{\mathbf{V}[Y]}}.$$

1. Clearing the above level, 20 points will be given.
2. If there is a wrong calculation or if the explanation is not sufficient, some points will be deducted. If the outline is acceptable, 10 or 15 points will be given.
3. On the contrary, if there is an essential error or unlogical explanation without sufficient understanding basic concepts, the point will reduced nearly to zero.

To get the full score 25 points, you must add some selling points.

1. An elegant calculation. In the above example, letting Z_1, Z_2 be the slots of two dice and calculate the statistics by means of $X = Z_1 + Z_2$ and $Y = Z_1 Z_2$.
2. Your own consideration on the problem. In the above example,
   (a) Compare the result with realistic correlation coefficients found in real data. In this case the source of data must be referred to.
   (b) Generalize or extend the problem. For example, discuss what happens if the dice has N faces. Calculation for a few concrete N is also acceptable.
   (c) Discuss similar problems when rolling N dice.
   (d) The slots of rolled dice is discrete random variables. How about a continuous random variable? For example, discuss a similar problem for randomly chosen two numbers in the interval $[0, 1]$.

確率モデル論・応用解析学

レポート心得

1. 基本的には, 出題された「数学の問題を解く」ことが求められる.
2. レポート課題なのだから, ブラインドで解くことは前提としていない. むしろ, 文献等を利用して勉強しながら問題を解くことが推奨される.
3. 友人と議論することも構わない (ルーツを同一とみなされるレポート（コピー）は 0 点！議論をともにすることと, レポートを写すことは似て非なるもの).
4. しかしながら, その気になれば, ネットや文献等で類題を検索して, その解答をコピペできるかもしれない. 質問箱に書き込んで誰かが教えてくれることを期待してもよい. 問題はすべてオリジナル (他所からの引き写しではないという意味) であるが, なにせ基本的な概念を確認するレベルであるから, それなりの専門家なら誰でも似たような問題を思いつく.
5. 以上の状況を一切踏まえた上で, 本レポート課題では 1 題 25 点満点で採点する.

※ 採点基準を例示する.

レポート問題：サイコロを 2 個投げたとき, 出た目の和を X, 積を Y とする. このとき, X, Y の相関係数を計算せよ.

典型的な解答はたぶんこんな感じ：

(1) 結合分布 $P(X = x, Y = y)$ を表に書く. むろん, x, y の可能な値だけを考えればよい.

(2) 上の表で周辺分布を計算して, $P(X = x)$, $P(Y = y)$ の一覧表を得る.

(3) 準備した確率分布をもとに, $\mathbf{E}[X]$, $\mathbf{E}[X^2]$, $\mathbf{V}[X]$, $\mathbf{E}[Y]$, $\mathbf{E}[Y^2]$, $\mathbf{V}[Y]$, $\mathbf{E}[XY]$ を次々に計算して, 最後に,

$$\mathbf{Cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y], \qquad \rho(X, Y) = \frac{\mathbf{Cov}(X, Y)}{\sqrt{\mathbf{V}[X]}\sqrt{\mathbf{V}[Y]}}$$

を求める.

1. ここまで, 正しく計算できれば, 20 点は獲得することになる.
2. 計算ミスや説明文が不足していれば, 相応に減点されるが, 筋がよければ 10〜15 点位は大丈夫.
3. 逆に, 本質的におかしい場合, 概念を理解しているとは思えない展開などがあれば, 限りなく 0 点に近くなる.

※ 正しい答えを導くだけでは, 満点に 5 点足りない. そこがレポートならではのポイント.

1. 途中に, 気の利いた論理や計算の工夫があれば加点する. ただし, これはオプションである. 上の例題なら, サイコロの出目を Z_1, Z_2 として, $X = Z_1 + Z_2$, $Y = Z_1 Z_2$ とおいて, 問題の統計量を計算するなど.
2. 得られた答に対して, 自分ならではの解釈・視点・発展などを述べた部分に対して加点する.

※ 上記の例で言えば, 次のようなことを議論すれば, その出来に従って加点する.

1. 得られた相関係数をより実感のあるものと比較する. たとえば, 英語と数学の得点の相関係数, 身長と体重の相関係数,... (ただし, 出典を明記しない引用は不可).
2. 6 面体のさいころではなくて N 面体のサイコロならどうなるか? 一般の N が難しければ, 具体的な N で計算する. 相関係数は N とともにどう変化するか?
3. サイコロ 2 個ではなく, サイコロ N 個振ったときの和 X と積 Y の相関係数. 一般の N が難しければ, 具体的な N で計算する. 相関係数は N とともにどう変化するか?
4. サイコロの目は離散型確率変数である. 連続型でも考えることはできるだろうか? たとえば, $[0, 1]$ からランダムに 2 個の数を選び, 和を X, 積を Y とする. このとき, X, Y の相関係数を調べる.