2013 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きている. 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.

● **目次**

— 終盤は, やや流動的 (トピックスを選んで講義する予定).

● **成績**

配布プリントに出題する問題 (**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).

● この講義は拙著「確率モデル要論」(牧野書店, 2012 年 6 月発行) にもとづく.
— 講義では, 簡単なレジュメ (英語) だけ配布する.
— ウェッブページには, 過去の講義録 (日本語)・その他の資料も掲載されている.

● 担当者連絡先

尾畑伸明 (情報科学研究科) obata@math.is.tohoku.ac.jp
　　初回 (10 月 3 日) と 11 月以降を担当する
瀬川悦生 (情報科学研究科) segawa@m.tohoku.ac.jp
　　10 月の講義を担当する

● 応用数学連携フォーラム http://www.dais.is.tohoku.ac.jp/~amf/
— 数学と諸分野との連携, ワークショップの開催 (来聴歓迎!)

Contents

Overview

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 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*:

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

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

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

(1) 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 to get at the n-th coin toss.

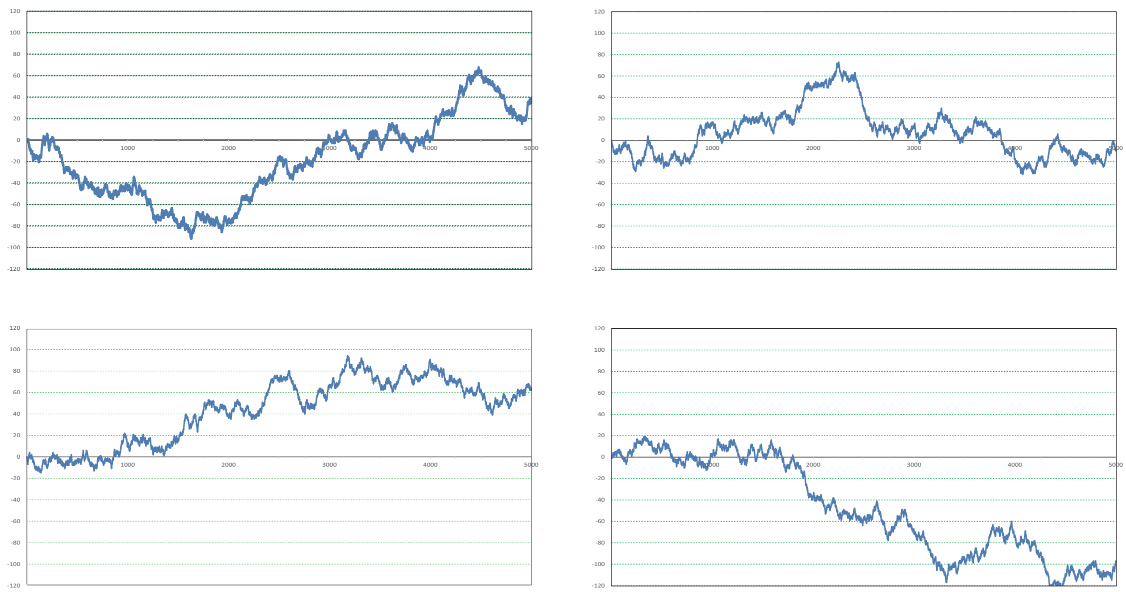
To be precise, we must say that $\{Z_n\}$ is a sequence of independent, identically distributed (iid) random variables with the common distribution

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

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

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

X_n means the net income at the time n. Or it stands for the coordinate of a drunker 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)

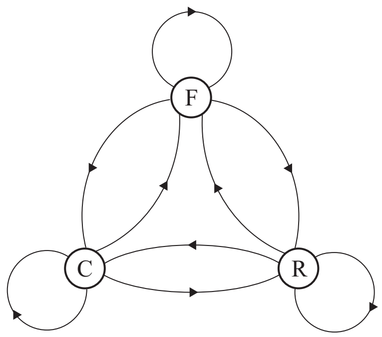
(v) gambler's ruin (random walk with barriers)

(2) Markov chains

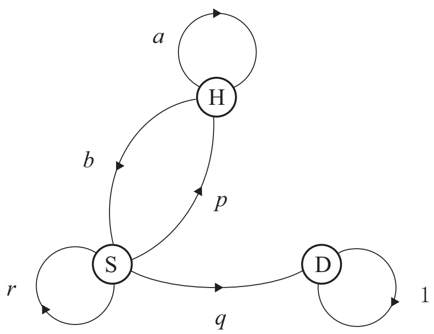
Consider the time evolution of a random phenomenon, where several different *states* are observed at each time step $n = 0, 1, 2, \dots$. For example, for the ever-changing weather, after simplification we observe three states: fine, cloudy, rainy. Collected data look like a sequence of F, C, R:

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

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:

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

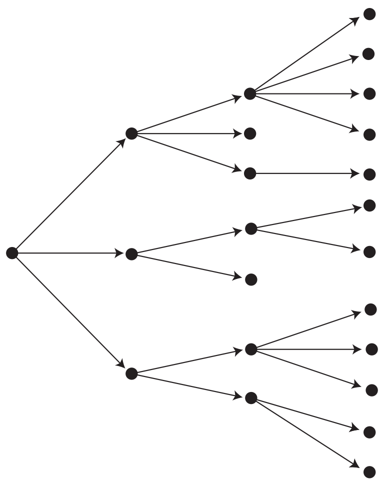
where the right hand side is independent of n (time homogeneous). A random walk is an example of a Markov chain. The theory of Markov chains is one of the best successful theories in probability theory for its simple description and unexpectedly rich structure. We are interested in the following topics:

(i) stationary distribution

(ii) recurrence

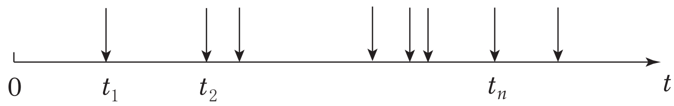
(iii) average life span

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



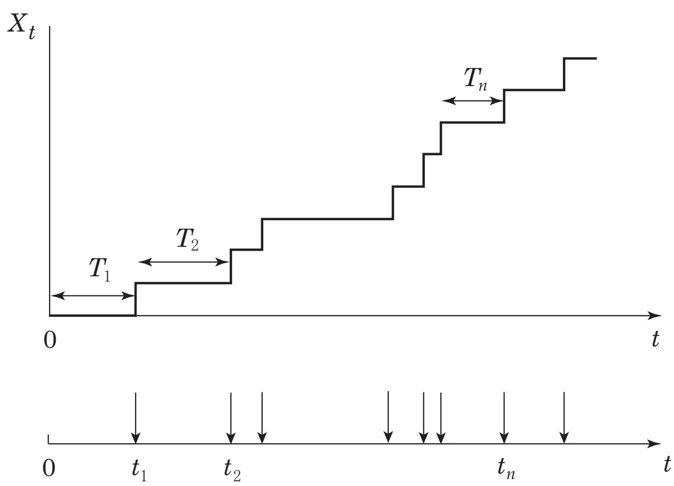
(3) Poisson process

This is one of the most fundamental continuous-time stochastic processes, while the other is the Brownian motion (Wiener process). Both belong to *Markov processes*. 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. The situation is illustrated as



where $t_1, t_2, \ldots$ are the time when E occurs.

Such random phenomena are well modeled by a Poisson process. We count the number of occurrence of the event E during the time interval $[0, t]$ and construct a stochastic process $\{X_t\ ;\ t \geq 0\}$ of which sample path is observed in the real world.



Some interesting subjects are listed below:

(i) statistics between two consecutive occurrence of events (waiting time);

(ii) waiting lines (models of queues);

(iii) birth-and-death processes.

● 参考書

1. 国沢清典 : 確率論とその応用 (岩波全書), 1982.

   本講義はこのレベルをめざす. 少し古いので入手困難かも.

2. 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.

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

4. 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.

5. 佐藤坦 : はじめての確率論　測度から確率へ (共立), 1994.

   コルモゴロフの公理から出発して確率論の基礎を数学的に厳密に展開している. 特に, ルベーグ積分が気になる人は一読してみよ. ただし, 確率モデルという観点は希薄 (この本を読んでから, 適当な本を見てねというスタンス).

6. 舟木直久 : 確率論, 朝倉書店, 2004.

7. 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.

8. 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.

9. 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, 初学者が理論を学ぶには重いか.

10. 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.

11. 鳥脇純一郎 : 工学のための確率論 (オーム社), 2002.

    この本は初学者向けで、確率論を速習するのに使えるが, 講義の内容をフォローするには不足である.

12. 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.

13. 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.

14. 西内啓 : 統計学が最強の学問である, ダイヤモンド社, 2013.

    統計リテラシーの必要性をジャーナリスティックに説く話題の本. 血が沸き立つような書きぶりではあるが, フィッシャー (20 世紀前半の大統計学者) を超えるのは難しい.

1 Random Variables and Probability Distributions

1.1 Random variables

random phenomenon $\Longrightarrow$ possible results $\Longrightarrow$ measurement $\Longrightarrow$ quantitative data

Example 1.1.1 (coin toss) Let's flip a coin. Heads or tails or others?

Example 1.1.2 (random cut) Cut a stick and divide into two segments.

Example 1.1.3 (life span of a person) Estimate the life span of a randomly chosen Japanese people.

A naive account: "variable" x vs "random variable" X
A random variable X is a variable varying over a certain range of real numbers and the probability is assigned to each possible value.

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

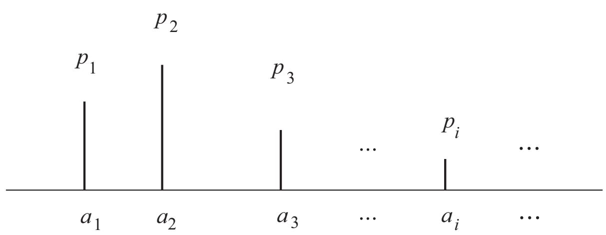
$$P(X = a_i) = p_i, \qquad p_i \ge 0, \qquad \sum p_i = 1.$$

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,

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



Example 1.1.4 (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 1.1.5 (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

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

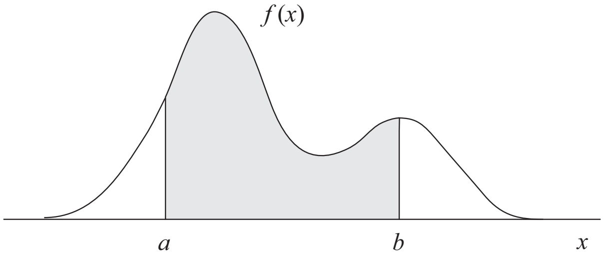
we say that X admits a *probability density function*. Note that

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

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 1.1.6 (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 1.1.7 (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.

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

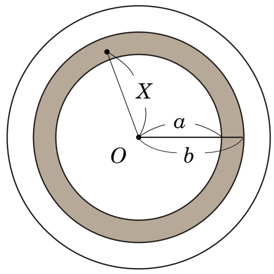


Figure 1.1: Random choice of a point

Example 1.1.8 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

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

1.2 Probability distributions

1.2.1 Mean and variance

Definition 1.2.1 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 1.2.2 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 1.2.3 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,

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

1.2.2 A list of discrete distributions

1. Bernoulli distribution with success probability $0 \le p \le 1$.

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

This is the distribution of coin toss.

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

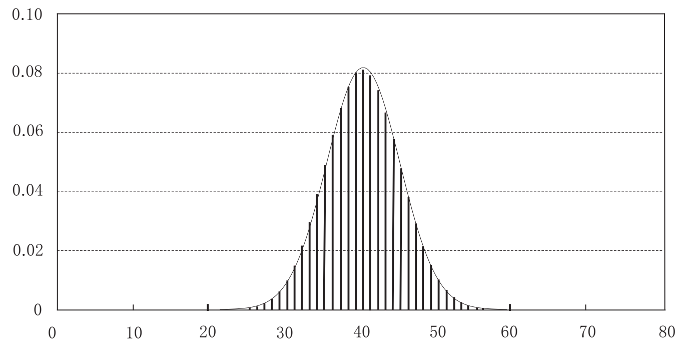
Quiz 1.2.4 Let a, b be distinct real numbers. A general two-point distribution is defined by

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

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

2. Binomial distribution $B(n, p)$ $(0 \le p \le 1, n \ge 1)$.

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



$B(100, 0.4)$

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.

Quiz 1.2.5 Verify that $m = np$ and $\sigma^2 = np(1-p)$ for $B(n, p)$.

3. Geometric distribution with parameter $0 \le p \le 1$.

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

This is the distribution of waiting time for the first heads (Example 1.1.5).

Quiz 1.2.6 Verify that $m = \dfrac{1}{p}$ and $\sigma^2 = \dfrac{1}{p^2}$

Remark 1.2.7 In some literatures, the geometric distribution with parameter p is defined by

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

4. Poisson distribution with parameter $\lambda > 0$.

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

Quiz 1.2.8 Verify that $m = \lambda$ and $\sigma^2 = \lambda$ for the Poisson distribution with parameter λ.

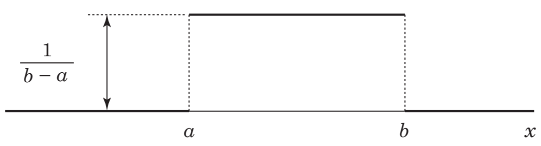
Quiz 1.2.9 (Challenge) Look up the method of generating functions in textbooks:

$$G(z) = \sum_{k=0}^{\infty} p_k z^k, \qquad \text{where} \quad p_k \geq 0, \quad \sum_{k=0}^{\infty} p_k = 1.$$

1.2.3 A list of continuous distributions (density functions)

1. Uniform distribution on $[a, b]$.

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

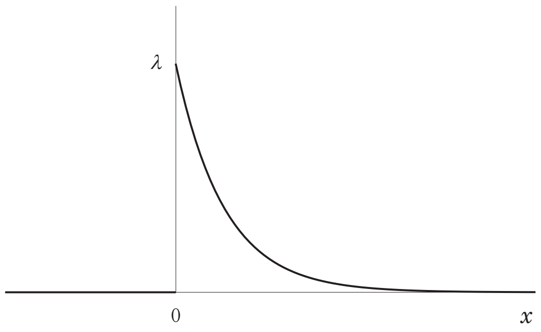


Quiz 1.2.10 Verify that $m = \dfrac{a+b}{2}$ and $\sigma^2 = \dfrac{(b-a)^2}{12}$.

2. Exponential distribution with parameter $\lambda > 0$.

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

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

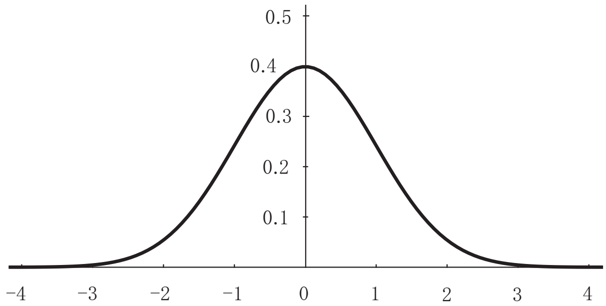


Quiz 1.2.11 Verify that $m = \dfrac{1}{\lambda}$ and $\sigma^2 = \dfrac{1}{\lambda^2}$.

3. Normal (Gaussian) distribution $N(m, \sigma^2)$ $(\sigma > 0, m \in \mathbf{R})$

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

In particular, $N(0, 1)$ is called the *standard normal distribution*.



Quiz 1.2.12 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:

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

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

Quiz 1.2.13 Let S be the length of the shorter segment obtained by randomly cutting the interval $[0, L]$. Calculate the mean and variance of S.

Problem 1 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) Find the probability density function of X.

(2) Calculate the mean and variance of X.

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

Problem 2 (1) Differentiating both sides of the known formula:

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

find the values

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

(2) Determine a constant a in order that

$$f(x) = \begin{cases} ax^2 e^{-x^2}, & x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

is a probability density function on $\mathbf{R}$, and find its mean and variance.

2 Bernoulli Trials

Repeated independent coin tosses are called the *Bernoulli trials*, where the tossed coins are identical in the sense that the probabilities of heads and tails remain the same throughout the trials. The Bernoulli trials form the most fundamental stochastic process.

2.1 Independence

2.1.1 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).

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

2.1.2 Independent random variables

Definition 2.1.9 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 2.1.10 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.

2.2 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 2.2.1 (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 2.2.2 (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 2.2.3 The *covariance* of two random variables X, Y is defined by

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

In particular, $\sigma_{XX} = \sigma_X^2$ becomes the variance of X.

Definition 2.2.4 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 2.2.5 If two random variables X, Y are independent, they are uncorrelated.

Remark 2.2.6 The converse of Theorem 2.2.5 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 $\sigma_{XY} = 0$.

Problem 3 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, this falls into a very particular situation.]

Theorem 2.2.7 (Additivity of variance) Let $X_1, X_2, \dots, X_n$ be random variables, any pair of which is uncorrelated. Then

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

Definition 2.2.8 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 $\sigma_Y > 0$.

Theorem 2.2.9 $-1 \le \rho_{XY} \le 1$ for two random variables X, Y with $\sigma_X > 0, \sigma_Y > 0$.

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

(1) Calculate the covariance σ_{LS} and the correlation coefficient r_{LS}.

(2) Are L, S independent?

2.3 Bernoulli trials

Definition 2.3.1 A sequence of random variables (or a discrete-time stochastic process) $\{Z_1, Z_2, \dots, Z_n, \dots\}$ is called the *Bernoulli trials* with success probability p $(0 \le p \le 1)$ if they are independent and have the same distribution as

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

By definition we have

$$P(Z_1 = \xi_1, Z_2 = \xi_2, \dots, Z_n = \xi_n) = \prod_{k=1}^{n} P(Z_k = \xi_k) \qquad \text{for all } \xi_1, \xi_2, \dots, \xi_n \in \{0, 1\}.$$

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

Let $\{Z_n\}$ be Bernoulli trials with success probability p. Define

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

The stochastic process $\{X_n\}$ is called the *binomial process*. Since X_n counts the number of success during the first n trials,

$$X_n \sim B(n, p).$$

The asymptotic properties of the binomial process will be studied in the following chapters.

Quiz 2.3.2 The waiting time for the first heads is given by $T = \inf\{n \geq 1\,;\, X_n = 1\}$. Find the distribution of T.

Problem 5 Let $\{X_n\}$ be a binomial process with success probability p.

(1) Find the conditional probability $P(X_{n+1} = j | X_n = i)$.

(2) Calculate the covariance $\sigma_{X_m, X_{m+n}}$ and the correlation coefficient $\rho_{X_m, X_{m+n}}$.

2.4 Random walks

Consider slightly modified Bernoulli trials $\{Z_n\}$ as follows:

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

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

Set

$$X_n = \sum_{k=1}^{n} Z_k \,. \tag{2.3}$$

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

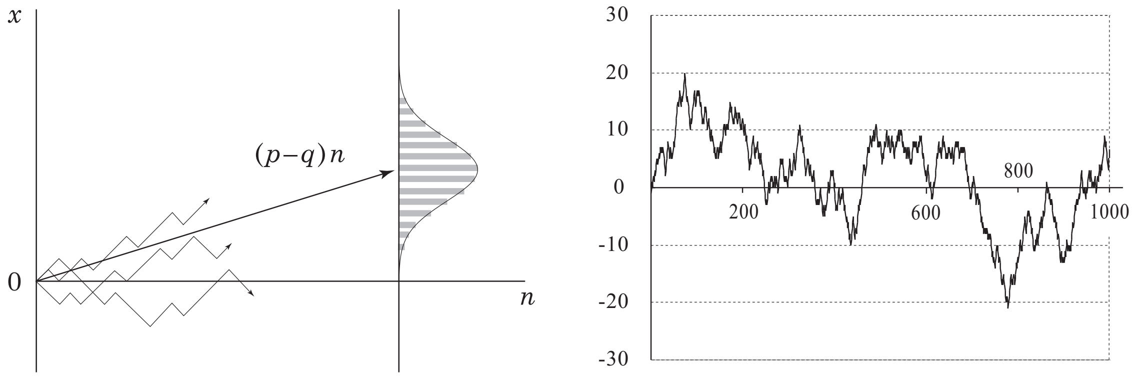


Figure 2.1: Random walk

Problem 6 Let $\{X_n\}$ be the random walk as above.

(1) Show that

$$P(X_n = 2k - n) = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, 2, \dots, n. \tag{2.4}$$

(2) Find $\mathbf{E}[X_n]$ and $\mathbf{V}[X_n]$.

(3) Find the conditional probability $P(X_{n+1} = j | X_n = i)$.

(4) Calculate the covariance $\sigma_{X_m, X_{m+n}}$ and the correlation coefficient $\rho_{X_m, X_{m+n}}$.

[Cf. Problem 5]

3 Law of Large Numbers and Central Limit Theorem

3.1 Observation

Let X_n be the result of the n-th trial of coin-toss:

$$X_n = \begin{cases} 1, & \text{heads}, \\ 0, & \text{tails}. \end{cases} \tag{3.1}$$

Obviously,

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

counts the number of heads during the first n trials. Therefore,

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

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

Computer simulation is easy. The following is just one example showing that the relative frequency of heads S_n/n tends to $1/2$. It is our aim to show this mathematically.

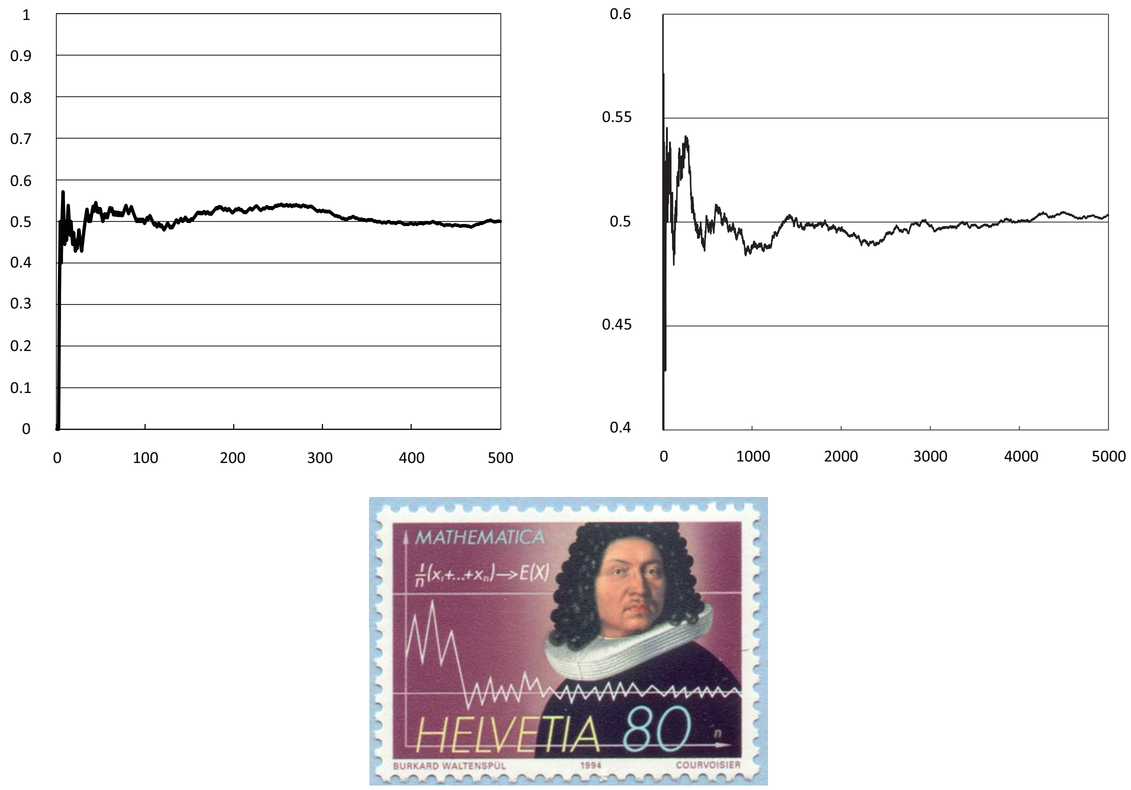


Figure 3.1: Relative frequency of heads S_n/n

However, we cannot accept a naive formula:

$$\lim_{n\to\infty} \frac{S_n}{n} = \frac{1}{2} \tag{3.2}$$

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 (3.2) for *all* samples ω.

Therefore, to show the empirical fact (3.2) we need some *probabilistic formulation*.

3.2 Weak Law of Large Numbers

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 By definition we have

$$m = \mathbf{E}[X] = \int_\Omega X(\omega)\, P(d\omega),$$
$$\sigma^2 = \mathbf{E}[(X - m)^2] = \int_\Omega (X(\omega) - m)^2 P(d\omega).$$

The above integral for the variance is divided into two parts as follows:

$$\sigma^2 = \int_\Omega (X(\omega) - m)^2 P(d\omega)$$
$$= \int_{|X-m|\ge\epsilon} (X(\omega) - m)^2 P(d\omega) + \int_{|X-m|<\epsilon} (X(\omega) - m)^2 P(d\omega)$$

Then we have

$$\sigma^2 \ge \int_{|X-m|\ge\epsilon} (X(\omega) - m)^2 P(d\omega) \ge \int_{|X-m|\ge\epsilon} \epsilon^2 P(d\omega) = \epsilon^2 P(|X - m| \ge \epsilon),$$

as desired. ∎

Proof of 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

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

Let us compute the variance. Since $\mathbf{E}[X_k X_l] = \mathbf{E}[X_k]\mathbf{E}[X_l]$ $(k \neq l)$ by assumption of being uncorrelated, we have

$$\begin{aligned}
\mathbf{E}[Y^2] &= \frac{1}{n^2}\sum_{k,l=1}^{n}\mathbf{E}[X_k X_l] \\
&= \frac{1}{n^2}\left\{\sum_{k=1}^{n}\mathbf{E}[X_k^2] + \sum_{k\neq l}\mathbf{E}[X_k X_l]\right\} \\
&= \frac{1}{n^2}\left\{\sum_{k=1}^{n}\left(\mathbf{V}[X_k] + \mathbf{E}[X_k]^2\right) + \sum_{k\neq l}\mathbf{E}[X_k]\mathbf{E}[X_l]\right\} \\
&= \frac{1}{n^2}\left\{n\sigma^2 + nm^2 + (n^2-n)m^2\right\} \\
&= \frac{\sigma^2}{n} + m^2.
\end{aligned}$$

Therefore,

$$\mathbf{V}[Y] = \mathbf{E}[Y^2] - \mathbf{E}[Y]^2 = \frac{\sigma^2}{n}\,.$$

On the other hand, applying Chebyshev inequality, we have

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

Consequently,

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

as desired. ∎

Example 3.2.4 (Coin-toss)

3.3 Strong Law of Large Numbers

Theorem 3.3.1 (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 mutually 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 3.3.2 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 3.3.1 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.

3.4 De Moivre–Laplace Theorem

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

$$B(n, p) \approx N(np, np(1 - p)) \tag{3.3}$$

We see that the matching becomes better for larger n.

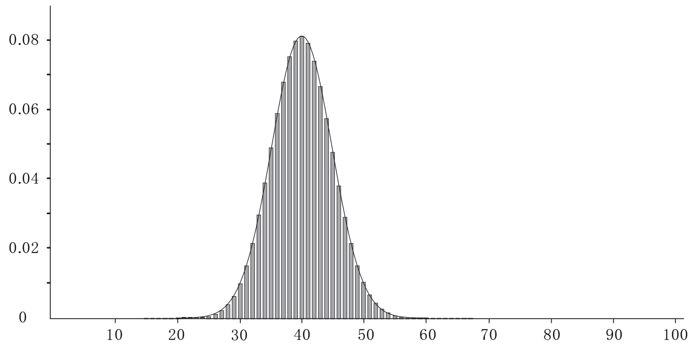


Figure 3.2: The normal distribution whose mean and variance are the same as $B(100, 0.4)$

The approximation (3.3) means that distribution functions are almost the same: For a random variable S obeying the binomial distribution $B(n, p)$ we have

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

Changing the variables, we come to

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

Noting the obvious identity:

$$P(S \le x) = P\left(\frac{S - m}{\sigma} \le \frac{x - m}{\sigma}\right)$$

and replacing $(x - m)/\sigma$ with x, we obtain

$$P\left(\frac{S - np}{\sqrt{np(1 - p)}} \le x\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt. \tag{3.4}$$

The right-hand side is an integral with respect to the normal law $N(0, 1)$ and is independent of n. The identity (3.4) provides the best formulation of the fact that (3.3) becomes better approximation for larger n.

Theorem 3.4.1 (de Moivre–Laplace theorem) Let $0 < p < 1$. Let S_n be a random variable obeying the binomial distribution $B(n, p)$. Then,

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{np(1 - p)}} \le x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt. \tag{3.5}$$

In short, the binomial distribution $B(n, p)$ is close to the normal distribution $N(np, np(1-p))$ as n tends to infinity.

The proof is omitted, see the relevant books.

3.5 Central Limit Theorem

We start with an alternative form of Theorem 3.4.1. 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.

Since we have

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

(3.5) becomes

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

Indeed, the above limit formula holds for general iid random variables.

Theorem 3.5.1 (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 3.5.2 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 $\mu(dx)$.

Theorem 3.5.3 (Glivenko) 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 $z \in \mathbf{C}$, 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.

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.,

$$\varphi_n(z) = \mathbf{E}\left[\exp\left\{ \frac{iz}{\sqrt{n}} \sum_{k=0}^{n} X_k \right\} \right]. \tag{3.6}$$

On the other hand, it is known that the characteristic function of $N(0,1)$ is given by $e^{-z^2/2}$ (Problem 7). By virtue of Glivenko's theorem it is sufficient to show that

$$\lim_{n\to\infty} \varphi_n(z) = e^{-z^2/2}, \qquad z \in \mathbf{R}. \tag{3.7}$$

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

$$\varphi_n(z) = \prod_{k=1}^{n} \mathbf{E}\left[\exp\left\{\frac{iz}{\sqrt{n}} X_k\right\}\right] = \varphi\Big(\frac{z}{\sqrt{n}}\Big)^n. \tag{3.8}$$

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 (3.8) becomes

$$\varphi_n(z) = \left(1 - \frac{z^2}{2n} + \mathbf{E}[R_n(z)]\right)^n. \tag{3.9}$$

4) We note a general limit theorem for the exponential function (Problem 10).

5) We need to prove that

$$\lim_{n\to\infty} n\mathbf{E}[R_n(z)] = 0. \tag{3.10}$$

In fact, by 4) we obtain

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

6) By means of the useful inequality:

$$\left|e^{ix} - \left(1 + ix + \frac{(ix)^2}{2!}\right)\right| \le \min\left\{\frac{|x|^3}{6}, |x|^2\right\}, \qquad x \in \mathbf{R}, \tag{3.11}$$

we obtain

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

Then

$$|n\mathbf{E}[R_n(z)]| \le \mathbf{E}[n|R_n(z)|] \le |z|^2\mathbf{E}\left[\min\left\{\frac{|z|}{6\sqrt{n}}|X_1|^3, |X_1|^2\right\}\right]. \tag{3.12}$$

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 (3.10). ∎

Remark 3.5.4 In the above proof we did not require $\mathbf{E}[|X_1|^3] < \infty$. If $\mathbf{E}[|X_1|^3] < \infty$ is satisfied, (3.10) follows more easily without appealing to the Lebesgue convergence theorem.

Problem 7 Calculate the characteristic function of the standard normal distribution:

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

Problem 8 Let $Z_1, Z_2, \ldots$ be the Bernoulli trials with success probability p and $\bar{Z}_1, \bar{Z}_2, \ldots$ their normalization. Calculate explicitly the characteristic function of

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \bar{Z}_k$$

and find its limit as $n \to \infty$ directly.

Problem 9 Prove (3.11).

Problem 10 Let $a \in \mathbf{C}$ and let $\{\epsilon_n\}$ be a sequence of complex numbers converging to 0. Prove that

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

Problem 11 (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

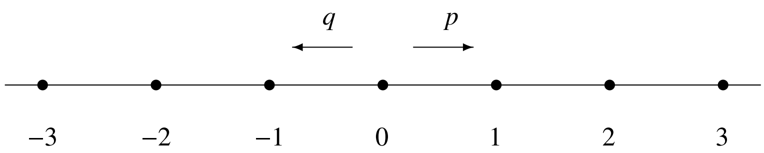
$$\int_0^1 f(x)dx.$$

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

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*.

Let $\{Z_n\}$ be iid random variables such that

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

This is also referred to as the *Bernoulli trials* (the usual Bernoulli trials take values in $\{0, 1\}$). Then we have

$$X_n = Z_1 + Z_2 + \cdots + Z_n\,, \qquad X_0 = 0.$$

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

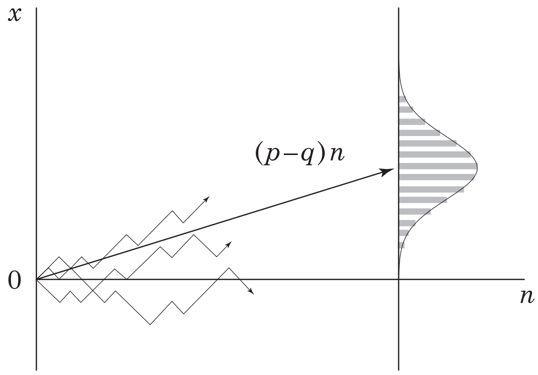
$$\mathbf{E}[X_n] = \sum_{k=1}^{n} \mathbf{E}[Z_k] = (p - q)n.$$

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 men 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).



Problem 12 Let $\{X_n\}$ be the random walk as above. Calculate the covariance $\sigma_{X_m,X_{m+n}}$ and the correlation coefficient $\rho_{X_m,X_{m+n}}$.

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

$$R = P\left(\bigcup_{n=1}^{\infty}\{X_{2n} = 0\}\right). \tag{4.1}$$

It is important to note that

$$\bigcup_{n=1}^{\infty}\{X_{2n} = 0\}$$

is not the sum of disjoint events.

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

$$p_{2n} = P(X_{2n} = 0) = \binom{2n}{n} p^n q^n = \frac{(2n)!}{n!n!}\, p^n q^n, \qquad n = 1, 2, \dots. \tag{4.2}$$

For convenience set $p_0 = 1$. Note that the right hand side of (4.1) is not the sum of p_{2n}. Instead, we need to consider the probability that the random walker returns to the origin after $2n$ steps but not before:

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

Remind that the difference between p_{2n} and q_{2n}. Equivalently, letting

$$T = \inf\{n \geq 1\, ;\, X_n = 0\} \tag{4.3}$$

be the *first hitting time* to the origin, we have

$$q_{2n} = P(T = 2n). \tag{4.4}$$

Remark 4.2.1 The symbol inf in the right-hand side of (4.3) covers the case of $\{n \geq 1\, ;\, X_n = 0\} = \emptyset$. In that case the value is understand to be $+\infty$. So, according to our definition in Chapter 1, T is not a random variable. It is, however, commonly accepted that a random variable takes values in $(-\infty, +\infty) \cup \{\pm\infty\}$.

The return probability is given by

$$R = \sum_{n=1}^{\infty} q_{2n}. \tag{4.5}$$

A direct calculation of q_{2n} will be given in the next section. Here we apply the method of generating functions.

The key relation between $\{p_{2n}\}$ and $\{q_{2n}\}$ is given by

$$p_{2n} = \sum_{k=1}^{n} q_{2k} p_{2n-2k}, \qquad n = 1, 2, \dots, \tag{4.6}$$

which is easily verified by observing when the random walker returning to the origin after $2n$ steps hits first the origin. Define the generating functions of $\{p_{2n}\}$ and $\{q_{2n}\}$ by

$$g(z) = \sum_{n=0}^{\infty} p_{2n} z^{2n}, \qquad h(z) = \sum_{n=1}^{\infty} q_{2n} z^{2n}. \tag{4.7}$$

These are convergent in $|z| \leq 1$. Multiplying z^{2n} to both sides of (4.6) and summing up over n, we obtain

$$\begin{aligned} g(z) - 1 &= \sum_{n=1}^{\infty} \sum_{k=1}^{n} q_{2k} z^{2k} p_{2n-2k} z^{2n-2k} \\ &= \sum_{k=1}^{\infty} \sum_{n=0}^{\infty} q_{2k} z^{2k} p_{2n} z^{2n} \\ &= h(z) g(z). \end{aligned}$$

Hence,

$$h(z) = 1 - \frac{1}{g(z)}. \tag{4.8}$$

On the other hand, by the formula of the binomial coefficients we can compute $g(z)$ explicitly (Problem 13). In fact, we have

$$g(z) = \sum_{n=0}^{\infty} p_{2n} z^{2n} = \sum_{n=0}^{\infty} \binom{2n}{n} p^n q^n z^{2n} = \frac{1}{\sqrt{1 - 4pqz^2}}$$

so that (4.8) becomes

$$h(z) = 1 - \frac{1}{g(z)} = 1 - \sqrt{1 - 4pqz^2}. \tag{4.9}$$

Letting $z \to 1 - 0$, we see that

$$R = h(1) = \sum_{n=1}^{\infty} q_{2n} = 1 - \sqrt{1 - 4pq} = 1 - |p - q|.$$

Here we used a general property stated in Problem 14.

Summing up,

Theorem 4.2.2 Let R be the probability that a random walker starting from the origin returns to the origin in finite time. Then we have

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

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

Theorem 4.2.3 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.

Theorem 4.2.4 The mean recurrent time of the isotropic, one-dimensional random walk is infinity.

Proof Let T be the first hitting time to the origin. The mean recurrent time is given by

$$\mathbf{E}(T) = \sum_{n=1}^{\infty} 2nq_{2n}. \tag{4.10}$$

On the other hand, in view of (4.7) and (4.9) we see that the generating function for $p = q = 1/2$ is given by

$$h(z) = \sum_{n=1}^{\infty} q_{2n} z^{2n} = 1 - \sqrt{1 - z^2}\,.$$

Differentiating with respect to z, we have

$$h'(z) = \sum_{n=1}^{\infty} 2nq_{2n} z^{2n-1} = \frac{z}{\sqrt{1 - z^2}}\,.$$

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

$$\mathbf{E}(T) = \sum_{n=1}^{\infty} 2nq_{2n} = \lim_{z \to 1-0} h'(z) = \lim_{z \to 1-0} \frac{z}{\sqrt{1 - z^2}} = +\infty.$$

This completes the proof. ∎

Remark 4.2.5 We will study the recurrence of a random walk within the framework of a general Markov chain.

Problem 13 Let α be a real constant. Using the binomial expansion:

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

prove that

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

Problem 14 Let $a_n \geq 0$ for $n = 0, 1, 2, \ldots$ and set

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

Show that

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

holds including the case of $\infty = \infty$ whenever the radius of convergence of $f(x)$ is ≥ 1.

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 $(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ 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 $y = x$.

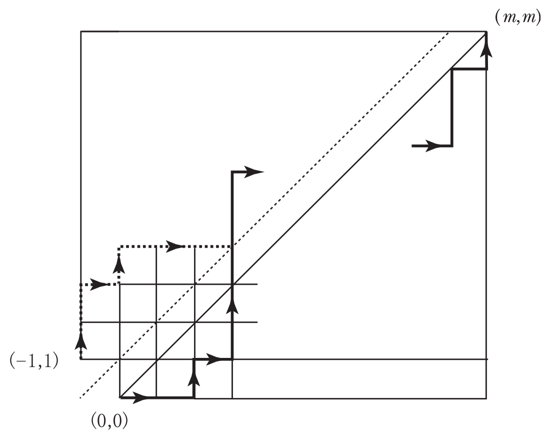
Theorem 4.3.3 (Catalan number)

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

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. ∎



Lemma 4.3.4 The generating function of the Catalan numbers C_n is given by

$$f(z) = \sum_{n=0}^{\infty} C_n z^n = \frac{1 - \sqrt{1 - 4z}}{2z}. \tag{4.11}$$

Proof Problem 15. ∎

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 4.3.5 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.

Theorem 4.3.6 Let $\{X_n\}$ be the random walk starting from 0 with right-move probability p and left-move probability q. Then

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

Proof Obviously, we have

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

In view of Fig. 4.3 we see that

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

Then the result is immediate. ∎

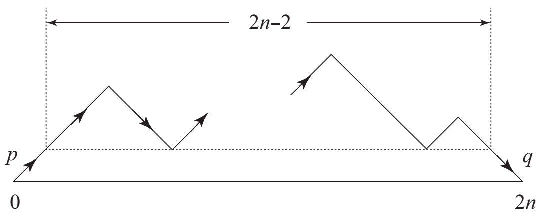


Figure 4.1: Calculating $P(X_1 > 0,\ X_2 > 0, \dots, X_{2n-1} > 0,\ X_{2n} = 0)$

We now come into an alternative proof of Theorem 4.2.2.

Proof [Theorem 4.2.2] We know by Theorem 4.3.6 that

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

Using the generating function of the Catalan numbers, we obtain

$$R = 2pqf(pq) = 2pq\,\frac{1-\sqrt{1-4pq}}{2pq} = 1-\sqrt{1-4pq} = 1-|p-q|,$$

as desired. ∎

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

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

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

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

(3) Find $f(z)$.

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

Problem 16 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

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

Problem 17 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

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

4.4 The Arcsine Law

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

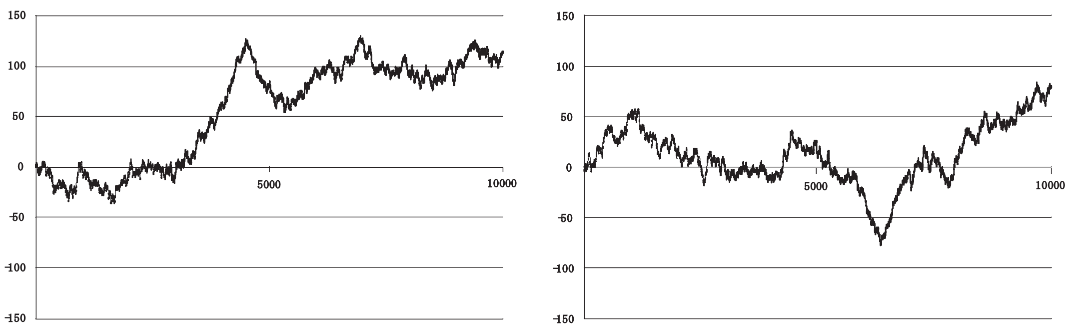


Figure 4.2: Sample paths of a random walk up to time 10000

By the law of large numbers we know that ±1 occur almost 5000 times. 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.

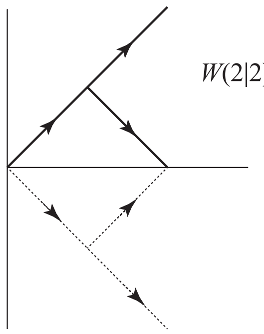
We say that a random walk stays in the positive region in the time interval $[i, i+1]$ if $X_i \geq 0$ and $X_{i+1} \geq 0$. Similarly, we say that a random walk stays in the negative region in the time interval $[i, i+1]$ if $X_i \leq 0$ and $X_{i+1} \leq 0$. Let

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

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

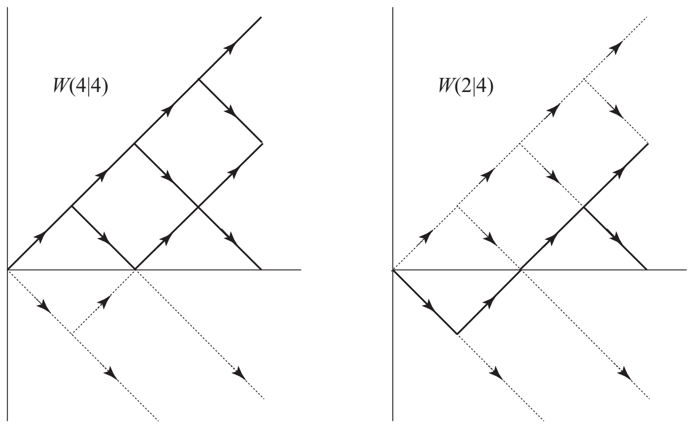
Remind that in this section we only consider an isotropic random walk ($p = q = 1/2$). For $n = 1$ we have

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



Similarly, we have

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



For general $W(2k|2n)$ we start with some simple calculations.

Lemma 4.4.1 For $n = 1, 2, \dots$ we have

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

(2) $f_{2n} \equiv P(X_1 > 0, X_2 > 0, \dots, X_{2n-1} > 0, X_{2n} = 0) = C_{n-1}\left(\frac{1}{2}\right)^{2n} = \frac{1}{4n}\, p_{2n-2}$.

Proof Known fact. ∎

Lemma 4.4.2 For $n = 1, 2, \dots$ we have

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

Proof We see from Lemma 4.4.1 (1) that

$$\begin{aligned}
p_{2n-2} - p_{2n} &= \binom{2n-2}{n-1}\left(\frac{1}{2}\right)^{2n-2} - \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} \\
&= \left(\frac{1}{2}\right)^{2n}\left\{\frac{(2n-2)!2^2}{(n-1)!(n-1)!} - \frac{(2n)!}{n!n!}\right\} \\
&= \left(\frac{1}{2}\right)^{2n}\frac{(2n-2)!}{n!n!}\{4n^2 - (2n)(2n-1)\} \\
&= \left(\frac{1}{2}\right)^{2n}\frac{(2n-2)!}{n!n!} \times 2n \\
&= \frac{1}{2n}\,\frac{(2n-2)!}{(n-1)!(n-1)!}\left(\frac{1}{2}\right)^{2n-2} \\
&= \frac{1}{2n}\,p_{2n-2} \\
&= 2f_{2n}\,.
\end{aligned}$$

Namely, we have

$$p_{2n-2} - p_{2n} = 2f_{2n}\,, \qquad n = 1, 2, \dots . \tag{4.12}$$

Now consider the complement of $\{X_1 \neq 0, X_2 \neq 0, \dots, X_{2n-1} \neq 0, X_{2n} \neq 0\}$.

$$\begin{aligned}
\{X_1 \neq 0, X_2 \neq 0, \dots, X_{2n-1} \neq 0, X_{2n} \neq 0\}^c &= \bigcup_{k=1}^{2n}\{X_k = 0\} \\
&= \bigcup_{k=1}^{n}\{X_{2k} = 0\} \\
&= \bigcup_{k=1}^{n}\{X_1 \neq 0, X_2 \neq 0, \dots, X_{2k-1} \neq 0, X_{2k} = 0\},
\end{aligned}$$

where the last is the disjoint union. Hence, by Lemma 4.4.1 (2) we have

$$\begin{aligned}
&P(X_1 \neq 0, X_2 \neq 0, \dots, X_{2n-1} \neq 0, X_{2n} \neq 0) \\
&= 1 - \sum_{k=1}^{n} P(X_1 \neq 0, X_2 \neq 0, \dots, X_{2k-1} \neq 0, X_{2k} = 0) \\
&= 1 - \sum_{k=1}^{n} 2f_{2k}\,.
\end{aligned}$$

Using (4.12), we obtain

$$\begin{aligned}
&P(X_1 \neq 0, X_2 \neq 0, \dots, X_{2n-1} \neq 0, X_{2n} \neq 0) \\
&= 1 - \sum_{k=1}^{n}(p_{2k-2} - p_{2k}) \\
&= 1 - (p_0 - p_{2n}) = p_{2n},
\end{aligned}$$

which completes the proof. ∎

Lemma 4.4.3 For $n = 1, 2, \dots$ we have

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

Proof Taking the complement into account, we only need to use the relation:

$$p_{2n-2} - p_{2n} = 2f_{2n} = \frac{1}{2n}\, p_{2n-2}\,.$$

∎

Theorem 4.4.4 For $n = 1, 2, \dots$ it holds that

$$W(2k|2n) = p_{2k}p_{2n-2k} = \binom{2k}{k}\binom{2n-2k}{n-k}\Big(\frac{1}{2}\Big)^{2n}, \qquad k = 0, 1, \dots, n.$$

Proof [outline] We first show that

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

Then the assertion is proved by induction on k, n (Problem19). ∎

We find a good approximation when $n \to \infty$. For $0 < a < b < 1$ we have

$$\begin{aligned}
&P(a < \text{ratio of time staying in the positive region} < b)\\
&= \sum_{k=an}^{bn} W(2k|2n)\\
&= \sum_{k=0}^{n} \chi_{[an,bn]}(k) W(2k|2n)\\
&= \sum_{k=0}^{n} \chi_{[a,b]}\Big(\frac{k}{n}\Big)\binom{2k}{k}\binom{2n-2k}{n-k}\Big(\frac{1}{2}\Big)^{2n},
\end{aligned}$$

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:

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

we obtain

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

Therefore,

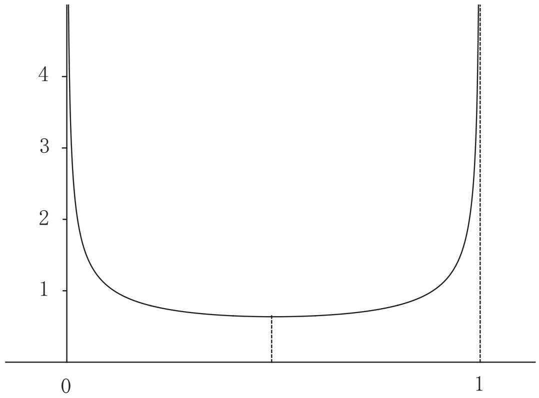
$$\begin{aligned}
&P(a < \text{ratio of time staying in the positive region} < b)\\
&\sim \sum_{k=0}^{n} \chi_{[a,b]}\Big(\frac{k}{n}\Big)\frac{1}{\pi\sqrt{k(n-k)}}\\
&= \sum_{k=0}^{n} \chi_{[a,b]}\Big(\frac{k}{n}\Big)\frac{1}{\pi\sqrt{\frac{k}{n}(1-\frac{k}{n})}}\,\frac{1}{n}\\
&\to \int_0^1 \chi_{[a,b]}(x)\frac{dx}{\pi\sqrt{x(1-x)}}\,.
\end{aligned}$$

Definition 4.4.5 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

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



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 18 Prove Lemmas 4.4.2 and 4.4.3.

Problem 19 Prove Theorem 4.4.4 in detail.

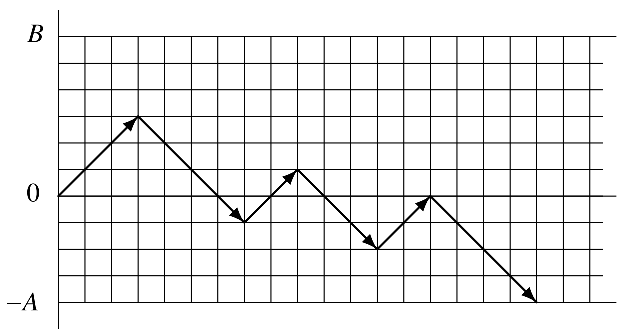
4.5 Gambler's Ruin

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 ($A \geq 1$, $B \geq 1$). 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

$$X_0 = 0, \qquad X_n = \begin{cases} X_{n-1} + Z_n\,, & -A < X_{n-1} < B, \\ -A, & X_{n-1} = -A, \\ B, & X_{n-1} = B. \end{cases} \tag{4.13}$$

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 4.5.1 $\{R_k\,;, -A \le k \le B\}$ fulfills the following difference equation:

$$R_k = pR_{k+1} + qR_{k-1}\,, \qquad R_{-A} = 1, \qquad R_B = 0. \tag{4.14}$$

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

$$S_k = pS_{k+1} + qS_{k-1}\,, \qquad S_{-A} = 0, \qquad S_B = 1. \tag{4.15}$$

Theorem 4.5.2 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 (4.13). 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 4.5.2 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 $n \ge 0$ 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 4.5.2 that the winning probability of A and B are given by

$$P(A) = \frac{A}{A+B}\,, \qquad P(B) = \frac{B}{A+B}\,, \tag{4.16}$$

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 4.2.2. 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. (4.16) 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 4.5.3 Let $\{X_n\}$ be the same as in Theorem 4.5.2. The expected life time of this random walk until absorption is given by

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

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

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

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

$$\mathbf{E}(Y_{-A}) = \mathbf{E}(Y_B) = 0. \tag{4.17}$$

For $-A < k < B$ we have

$$\mathbf{E}(Y_k) = \sum_{j=1}^{\infty} jP(Y_k = j). \tag{4.18}$$

In a similar manner as in the proof of Theorem 4.5.2 we note that

$$P(Y_k = j) = pP(Y_{k+1} = j-1) + qP(Y_{k-1} = j-1). \tag{4.19}$$

Inserting (4.19) into (4.18), we obtain

$$\begin{aligned}\mathbf{E}(Y_k) &= p\sum_{j=1}^{\infty} jP(Y_{k+1} = j-1) + q\sum_{j=1}^{\infty} jP(Y_{k-1} = j-1)\\ &= p\mathbf{E}(Y_{k+1}) + q\mathbf{E}(Y_{k-1}) + 1.\end{aligned} \tag{4.20}$$

Thus, $\mathbf{E}(Y_k)$ is the solution to the difference equation (4.20) with boundary condition (4.17). This difference equation is solved in a standard manner and we find

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

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.

Remark 4.5.4 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, $A \ge 1$, $B \ge 1$. Define $X_0, X_1, X_2, \dots$ by

$$X_0 = 0, \qquad X_n = \begin{cases} X_{n-1} + Z_n\,, & -A < X_{n-1} < B,\\ -A+1, & X_{n-1} = -A,\\ B-1, & X_{n-1} = B. \end{cases} \tag{4.21}$$

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

Problem 20 Solve the difference equation (4.20) with boundary condition (4.17).

Problem 21 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 $P(X_{2n} = 0)$.

確率過程, 特にマルコフ連鎖とランダム・ウォークを学ぶために

1. 国沢清典 : 確率論とその応用 (岩波全書), 1982.
   すでに紹介したとおり.

2. W. Feller: An Introduction to Probability Theory and Its Applications, Vol. 1, Vol. 2, Wiley, 1957.

   これもすでに紹介したとおり. ランダムウォークに関連する統計量の計算では, 事実上, カタラン数が現れているが, それをカタラン数と呼んではいない. 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.

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

   これもすでに紹介済み. 邦訳もある.

   B. V. グネジェンコ (鳥居一雄訳): 確率論教程 I, II, 森北出版, 1971, 1972.

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

   これもすでに紹介済み.

5. 志賀徳造：ルベーグ積分から確率論 (共立), 2000.

   最後の章である種の平面グラフ上のランダム・ウォークを議論している.

6. 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.

   マルコフ連鎖の基本的性質を速習するのに十分であろう.

7. Z. Brzeźniak and T. Zastawniak: Basic Stochastic Processes, Springer, 1999.

   確率過程入門として易しく書かれている.

8. 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.

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

   マルコフ連鎖に関するプロ向きの本.

5 Markov Chains

Let us recall a typical property of a random walk: the next position 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*.

5.1 Conditional Probability

For two events A, B we define

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{5.1}$$

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 2.1.1.

Theorem 5.1.1 For two events A, B the following five properties are equivalent:

(i) A and B are independent;

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

(ii′) $P(A|B) = P(A)$;

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

(iii′) $P(A|B) = P(A|B^c)$.

Here it is assumed that the conditional probabilities are well defined.

Formula (5.1) is often used in the following form:

$$P(A \cap B) = P(B)P(A|B) \tag{5.2}$$

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. 5.1.

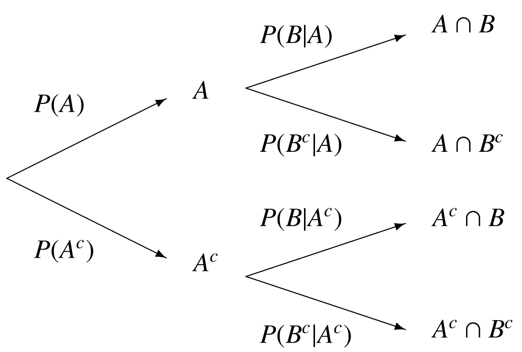


Figure 5.1: Tree diagram

(5.2) is generalized as follows.

Theorem 5.1.2 (Compound probabilities) For events $A_1, A_2, \dots, A_n$ we have

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}). \tag{5.3}$$

Proof Straightforward by induction on n. ∎

Problem 22 Prove Theorem 5.1.1.

5.2 Markov Chains

Let S be an at most 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 5.2.1 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

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

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

Definition 5.2.2 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 called 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 5.2.3 A matrix $P = [p_{ij}]$ with index set S is called a *stochastic matrix* if

(i) $p_{ij} \ge 0$.

(ii) $\sum_{j\in S} p_{ij} = 1$.

Theorem 5.2.4 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 $p(i, j) > 0$.

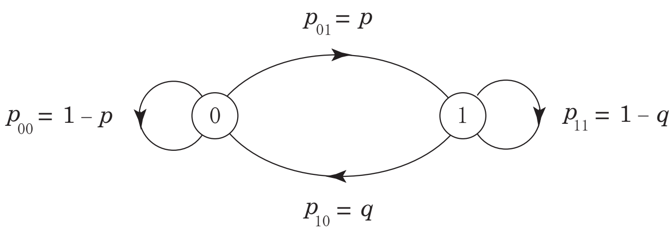
Example 5.2.5 (2-state Markov chain) A Markov chain over the state space $\{0, 1\}$ is determined by the transition probabilities:

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

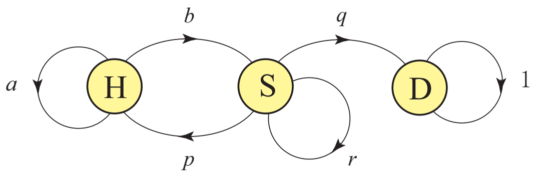
The transition matrix is defined by

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

The transition diagram is as follows:



Example 5.2.6 (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 5.2.7 (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 5.2.8 (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 $-A < i < B$ のときは,

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

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 5.2.9 (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

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

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 of the assumption of being time homogeneous.

Theorem 5.2.10 (Chapman–Kolmogorov equation) For $0 \le r \le n$ we have

$$p_n(i, j) = \sum_{k \in S} p_r(i, k) p_{n-r}(k, j). \tag{5.4}$$

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 (5.4). ∎

Applying (5.4) repeatedly and noting that $p_1(i, j) = p(i, j)$, we obtain

$$p_n(i, j) = \sum_{k_1, \dots, k_{n-1} \in S} p(i, k_1) p(k_1, k_2) \cdots p(k_{n-1}, j). \tag{5.5}$$

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

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

Remark 5.2.12 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).)

5.3 Stationary Distributions

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

$$\pi_i(n) = P(X_n = i), \qquad i \in S.$$

Then the row vector defined by

$$\pi(n) = [\cdots \ \pi_i(n) \ \cdots]$$

is called the *distribution* of X_n. In particular, $\pi(0)$, the distribution of X_0, is called the *initial distribution*.

Theorem 5.3.1 We have

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

In other words,

$$\pi_i(n) = \sum_{j_1, j_2, \dots, j_n} \pi_{j_1}(0) p(j_1, j_2) p(j_2, j_3) \dots p(j_{n-1}, j_n) p(j_n, i).$$

Proof We first note that

$$\pi_i(n) = P(X_n = i) = \sum_{k \in S} P(X_n = i | X_0 = k) P(X_0 = k) = \sum_{k \in S} \pi_k(0) p_n(k, i),$$

where $p_n(k, i) = (P^n)_{ki}$ and the right-hand side is a multiplication of a vector and matrix. Hence,

$$\pi_i(n) = (\pi(0) P^n)_i,$$

that is,

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

as desired. ∎

Definition 5.3.2 In general, a row vector $\pi = [\cdots \pi_i \cdots]$ satisfying

$$\pi_i \geq 0, \qquad \sum_{i \in S} \pi_i = 1$$

is called a *distribution* on S. A distribution π on S is called *stationary* (or *invariant*) if

$$\pi = \pi P. \tag{5.6}$$

Note that a stationary distribution may not exist. If a stationary distribution exists, it is not necessarily unique.

Example 5.3.3 (2-state Markov chain) Let $\{X_n\}$ be the Markov chain introduced in Example 5.2.5. The transition matrix is given by

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

The eigenvalues of P are $1, 1-p-q$. Suppose that $p+q>0$. We have

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

Consequently, if $|r|<1$, or equivalently $0<p+q<2$, we have

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

Hence

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

Consequently, when $0<p+q<2$, the distribution of the Markov chain converges to the stationary distribution independently of the initial distribution.

Example 5.3.4 (3-state Markov chain) We discuss the Markov chain $\{X_n\}$ introduced in Example 5.2.6. The stationary distribution is unique and given by $\pi = [0\ 0\ 1]$.

In order to describe the uniqueness of a stationary distribution we need the following

Definition 5.3.5 We say that a state j *can be reached from* a state i if there exists some $n \geq 0$ such that $p_n(i,j)>0$. By definition every state i can be reached from itself. We say that two states i and j *intercommunicate* if i can be reached form j and j can be reached from i, i.e., there exist $m \geq 0$ and $n \geq 0$ such that $p_n(i,j)>0$ and $p_m(j,i)>0$.

Lemma 5.3.6 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.

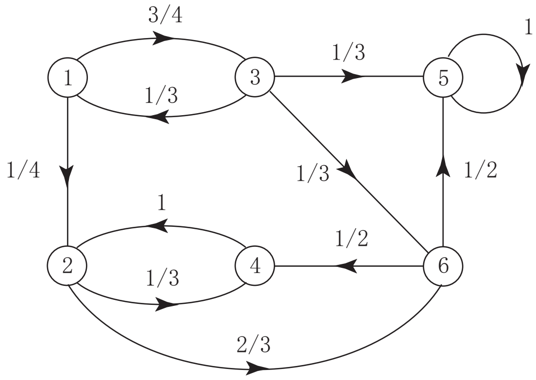
Definition 5.3.7 A state i is called *absorbing* if

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

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

Definition 5.3.8 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 5.3.9 Examine the equivalence relation among the states of a Markov chain described by the following transition diagram:



Example 5.3.10 (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 $p + q > 0$. If $p = q = 0$, then any $\pi = [\pi_0\, , \pi_1]$ is a stationary distribution.

The following result is fundamental, however, we omit the proof, see 国沢「確率論とその応用」, シナジ「マルコフ連鎖から格子確率モデルへ」, 拙著「確率モデル要論」etc.

Theorem 5.3.11 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

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

Recall that every state of an irreducible Markov chain on a finite state space is positive recurrent (Theorem 5.4.9). Therefore,

Theorem 5.3.12 An irreducible Markov chain on a finite state space S admits a unique stationary distribution $\pi = [\pi_i]$. Moreover, $\pi > 0$ for all $i \in S$.

Example 5.3.13 (One-dimensional RW) When S is infinite, Theorem 5.3.12 does not hold in general. In fact, one-dimensional random walk is an example. Let $[\cdots\ \pi(k)\ \cdots]$ be a distribution on $\mathbf{Z}$. If it is stationary, we have

$$\pi(k) = p\pi(k-1) + q\pi(k+1), \qquad k \in \mathbf{Z}. \tag{5.7}$$

The characteristic 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 (5.7) 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 (5.7) 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 5.3.14 (One-dimensional RW with reflection barrier) There is a unique stationary distribution when $p < q$. In fact,

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

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

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

Example 5.3.15 (2-state Markov chain) Let

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

be a transition matrix with $p+q > 0$. The eigenvalues are $1, 1-p-q$, which are distinct by assumption $p+q > 0$. Using the eigenvectors we set

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

Then,

$$PT = T \begin{bmatrix} 1 & 0 \\ 0 & 1-p-q \end{bmatrix},$$

in other words,

$$P^n = T \begin{bmatrix} 1 & 0 \\ 0 & (1-p-q)^n \end{bmatrix} T^{-1}.$$

Letting $n \to \infty$, we obtain

$$\lim_{n\to\infty} P^n = T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} T^{-1} = \frac{1}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix}.$$

On the other hand, for an initial distribution $\pi(0) = [\pi_0(0)\ \pi_1(0)]$ the distribution of the Markov chain at time n is given by Theorem 5.3.1 as follows:

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

Letting $n \to \infty$, we have

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

This coincides with the stationary distribution π, see Example 5.3.10. Namely, for a two-state Markov chain with $p + q > 0$, the distribution of X_n tends to the stationary distribution as $n \to \infty$ independent of an initial distribution.

It is important to know under which condition the distribution of a Markov chain tends to a stationary distribution after long time. Unfortunately, the situation is not simple. It is noted that the unique existence of a stationary distribution (see e.g., Theorem 5.3.12) is not sufficient to claim the convergence to the stationary distribution, as is seen by the following simple example.

Example 5.3.16 Consider a Markov chain determined by the transition matrix:

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

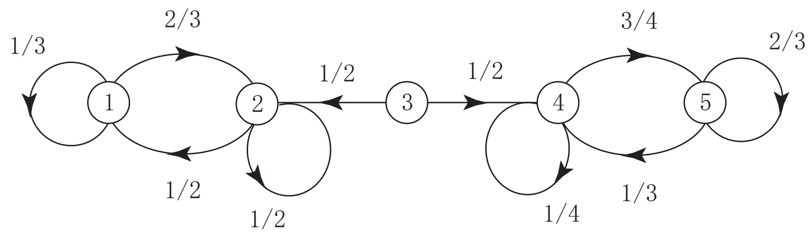
We first note that there exists a unique stationary distribution. But for a given initial distribution $\pi(0)$ it is not necessarily true that $\lim_{n\to\infty} \pi(n)$ converges to the stationary distribution.

Roughly speaking we need to avoid the periodic transition. We only mention the following result, of which the proof is found in 国沢「確率論とその応用」, シナジ「マルコフ連鎖から格子確率モデルへ」, etc.

Theorem 5.3.17 Let π be a stationary distribution of an irreducible Markov chain on a finite state space (It is unique, see Theorem 5.3.12). If $\{X_n\}$ is aperiodic, for any $j \in S$ we have

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

Problem 23 Consider a Markov chain determined by the transition diagram below.



(1) Is the Markov chain irreducible?

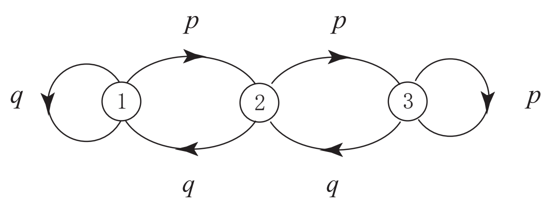
(2) Study the stationary distributions.

Problem 24 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:

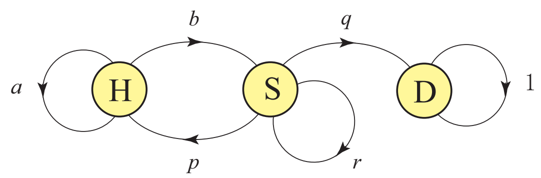
$$\mathbf{E}[X_n], \quad \mathbf{V}[X_n], \quad \mathrm{Cov}\,(X_{m+n}, X_n) = \mathbf{E}[X_{m+n}X_n] - \mathbf{E}[X_{m+n}]\mathbf{E}[X_n], \quad \rho(X_{m+n}, X_n) = \frac{\mathrm{Cov}\,(X_{m+n}, X_n)}{\sqrt{\mathbf{V}[X_{m+n}]\mathbf{V}[X_n]}}$$

Problem 25 There are two parties, say, A and B, and their supporters of a constant ratio exchange at every election. Suppose that after an election, 30% 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.

Problem 26 Find a stationary distribution of the Markov chain defined by the following transition diagram, where $p > 0$ and $q = 1 - p > 0$.



Problem 27 Let $\{X_n\}$ be the Markov chain introduced in Example 5.2.6:



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

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

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

(2) Solving the above linear system, prove the identities for the mean life times:

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

5.4 Recurrence

Definition 5.4.1 Let $i \in S$ be a state. Define the *first hitting time* or *first passage time* to i by

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

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

Theorem 5.4.2 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

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

$p_n(i,j)$ 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

$$p_n(i,j) = \sum_{r=1}^{n} f_r(i,j) p_{n-r}(j,j), \qquad i, j \in S, \quad n = 1, 2, \dots. \tag{5.8}$$

We next introduce the generating functions:

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

In view of (5.8) we see easily that

$$G_{ij}(z) = p_0(i, j) + F_{ij}(z)G_{jj}(z). \tag{5.9}$$

Setting $i = j$ in (5.9), 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 5.4.3 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 5.4.4 (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:

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \tag{5.10}$$

we have

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

Hence, if $p \neq q$ we have

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

If $p = q = \frac{1}{2}$, we have

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

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

Remark 5.4.5 Let $\{a_n\}$ and $\{b_n\}$ be sequences of positive numbers. We write $a_n \sim b_n$ if

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

Problem 28 (1) Let $\{a_n\}$ and $\{b_n\}$ be sequences of positive numbers such that $a_n \sim b_n$. Prove that there exist two constant numbers $c_1 > 0$ and $c_2 > 0$ such that

$$c_1 a_n \le b_n \le c_2 a_n\,.$$

(2) Notations and assumptions being as in (1), prove that $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ converge or diverge at the same time.

(3) Explain the details of Example 5.4.4 filling the gaps of arguments therein.

Example 5.4.6 (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

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

Employing the formula for the binomial coefficients:

$$\sum_{i=0}^{n} \binom{n}{i}^2 = \binom{2n}{n}, \tag{5.11}$$

which is a good exercise for the readers, we obtain

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

Then, by using the Stirling formula, we see that

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

so that

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

Consequently, two-dimensional random walk is recurrent.

Example 5.4.7 (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

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

We note the following two facts. First,

$$\sum_{i+j+k=n} \frac{n!}{i!j!k!} = 3^n. \tag{5.12}$$

Second, the maximum value

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

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

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

by the Stirling formula. Then we have

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

Therefore.

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

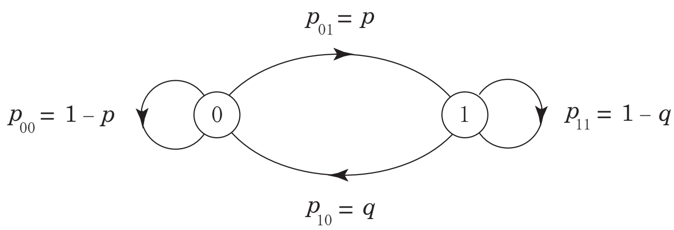
which implies that the random walk is not recurrent (i.e., transient).

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 4.2.4), 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 5.4.8 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 5.4.9 For an irreducible Markov chain on a finite state space S, every state is positive recurrent.

Problem 29 Let $\{X_n\}$ be a Markov chain described by the following transition diagram.



(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

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

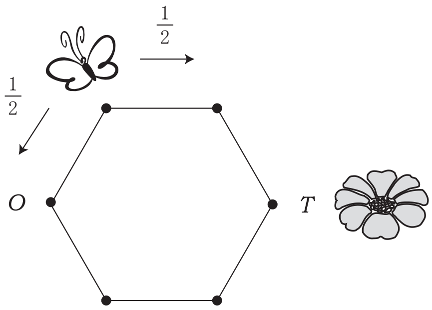
Problem 30 A butterfly flies over the hexagon with a trap T. It starts at the origin O at time $t = 0$ and chooses one direction at probability $1/2$. But when it arrives at T, it is killed.

(1) Find the transition matrix.

(2) Is this Markov chain irreducible?

(3) Find the probability that the butterfly is alive at time $t = 12$.

(4) Observe the situation as $t \to \infty$.



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 16 番のうち 2 題
   17 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：2014 年 2 月 3 日（月）–2 月 7 日（金）

Examination

1. Choose and answer 2 problems among No. 1–16 and 2 problems among No. 17– 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 3 (Mon)– 7 (Fri), 2014.

6 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.

6.1 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. 6.1.

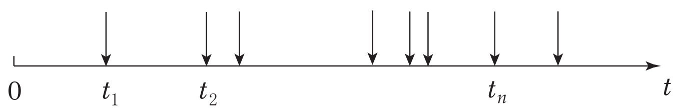


Figure 6.1: Recording when the event E occurs

There are two quantities which should be measured.

(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.

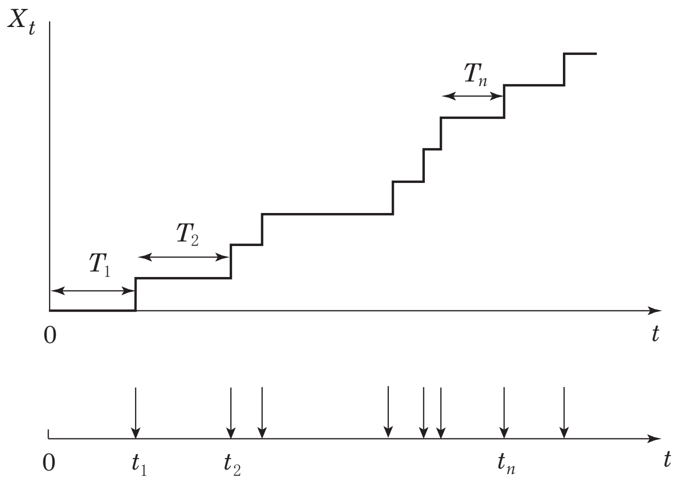
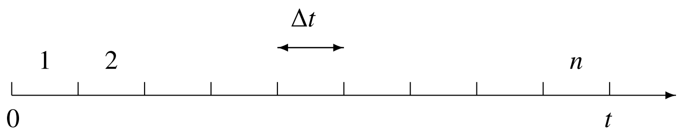


Figure 6.2: The counting process and waiting times



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

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

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

(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

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

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

$$X_t = \lim_{\Delta t \to 0} \sum_{i=1}^{n} Z_i \,. \tag{6.1}$$

Thus $\{X_t\}$ is a continuous time stochastic process which gives the number of occurrence of the event E up to time t. This is called the *Poisson process* with parameter $\lambda > 0$.

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

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

In other words, X_t obeys the Poisson distribution with parameter λt.

Remark 6.1.1 The essence of the above argument is the Poisson's law of small numbers which says that the binomial distribution $B(n, p)$ is approximated by Poisson distribution when n is large and p is small.

Problem 31 Let X be a random variable obeying the binomial distribution $B(n, p)$ and $\lambda = np = \mathbf{E}[X]$ the mean value. Prove that the distribution of X is approximated by the Poisson distribution with parameter λ when n is very big. Moreover, show by numerical calculation that $B(100, 0.05)$ is approximated by the Poisson distribution with parameter $\lambda = 5$.

Theorem 6.1.2 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 $0 \le t_1 < t_2 < \dots < t_k$, 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

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

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 $n_1 < \dots < n_k$. 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

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

∎

Example 6.1.3 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 32 The average number of customers visiting a certain service gate is 20 per hour. Using the Poisson model, calculate the following probabilities.

(1) The probability that one customer visits during the first two minutes after the gate opens.

(2) The probability that one customer visits during the first two minutes and that two customers visit during the next three minutes.

(3) The probability that the server waits for more than ten minutes until the first customer visits.

Problem 33 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.

6.2 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. It is convenient to remind that the Poisson process counts the number of events occurring up to time t. First we set

$$T_1 = \inf\{t \ge 0\,;\, X_t \ge 1\}. \tag{6.2}$$

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

$$T_n = \inf\{t \ge 0\,;\, X_t \ge n\} - T_{n-1}, \qquad n = 2, 3, \dots. \tag{6.3}$$

Theorem 6.2.1 Let $\{X_t\}$ be a Poisson process with parameter λ. Define the waiting time T_n by (6.2) and (6.3). Then, $\{T_n\,;\, n = 1, 2, \dots\}$ becomes a sequence of iid random variables, of which distribution is the exponential distribution with parameter λ.

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

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

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 6.2.2 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 6.2.1 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 34 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 6.2.1. Calculate $P(S_2 \le t)$ and find the probability density function of S_2. [In general, S_n obeys a gamma distribution.]

6.3 The Rigorous Definition of Poisson Processes

The "definition" of a Poisson process in (6.1) is intuitive and instructive for modeling random phenomena. However, strictly speaking, the argument is not sufficient to define a stochastic process $\{X_t\}$. For example, the probability space $(\Omega, \mathcal{F}, P)$ on which $\{X_t\}$ is defined is not at all clear.

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

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

and for $t \geq 0$,

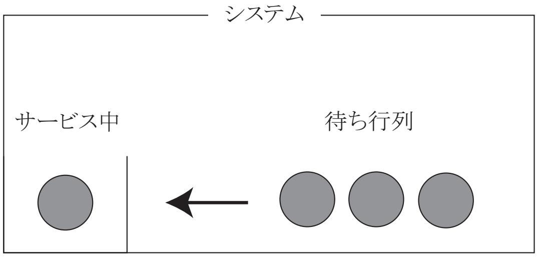
$$X_t = \max\{n \geq 0\,;\, S_n \leq t\}.$$

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

Starting with the above definition one can prove the properties in mentioned Theorem 6.1.2.

6.4 $M/M/1$ Queue

Waiting lines or queues have been extensively studied under the name of *queuing theory*. A queue in the real world is modeled in terms of a system of servers and visiting customers. A customer arrives at a system containing servers. If the server is free, the customer may get the service at once. Otherwise, the customer waits for a vacant server by making a waiting queue. A customer arrives at random and the service time is also random. The main purpose is to construct a stochastic process $\{X(t)\}$ on $\{0, 1, 2, \dots\}$, where the value corresponds to the number of customers in the system at time t (including any currently in service), and then to obtain statistical quantities and characteristic features from $\{X(t)\}$.



A quite a few models have been proposed and studied extensively. Kendall's notation $A/B/c/K/m/Z$, introduced first by David G. Kendall in 1953, is commonly used 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 (the buffer size),
m: calling population,
Z: queue's discipline (priority order)

The most fundamental model is an $M/M/1$ queue (the letter "M" stands for "Markov"). In fact, an $M/M/1$ queue is a continuous time Markov chain on $\{0, 1, 2, 3, ...\}$, which reflects the following situation:

(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;

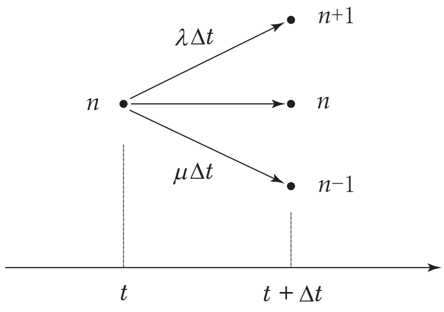
(iii) The system contains a single server;

(iv) The buffer size is infinite;

(v) (First in first out) A single server serves customers one at a time from the front of the queue, according to a first-come, first-served discipline. Customers are served one at a time and that the customer that has been waiting the longest is served first

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 us consider the change of the system during the small time interval $t \to t + \Delta t$. 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, $P(X(t) = n)$ fulfills the following equation:

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

Setting $p_n(t) = P(X(t) = n)$, we have

$$\begin{aligned}
&p_n'(t) = \lambda p_{n-1}(t) - (\lambda + \mu) p_n(t) + \mu p_{n+1}(t), \quad n = 1, 2, \dots, \\
&p_0'(t) = -\lambda p_0(t) + \mu p_1(t).
\end{aligned} \tag{6.4}$$

Important properties of the queue are derived from the above linear system. We are interested in the equilibrium solution (limit transition probability), i.e., $p_n(t)$ in the limit $t \to \infty$. We set

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

under the assumption that the limit exists. In the limit the derivative of the left hand side of (6.4) is 0. Hence,

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

The above linear system is easily solved. If $\lambda \neq \mu$, then the general solution is given by

$$p_n = A \left(\frac{\lambda}{\mu}\right)^n \quad (A \text{ is a constant}).$$

If $\lambda = \mu$, we see that $p_n = A$ (constant). Since p_n should be a probability distribution, we have $\sum_{n=0}^{\infty} p_n = 1$. 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 parameter λ/μ.

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

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

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

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

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

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

Example 6.4.1 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

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

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

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

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

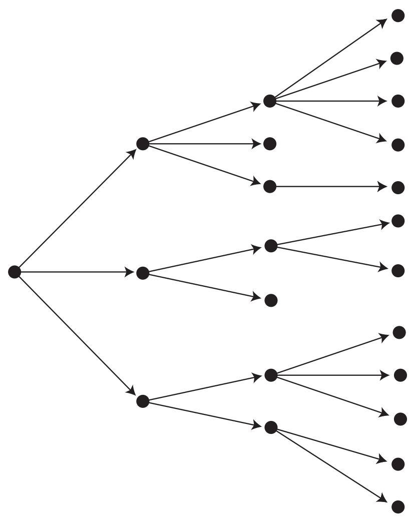
$$1 - \frac{2}{5} - \frac{6}{25} = \frac{9}{25} = 0.36.$$

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

Remark 6.4.2 The Markov process $X(t)$ appearing in the $M/M/1$ queuing model is studied more generally within the frame works of *birth-and-death process*.

7 Galton-Watson Branching Processes

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



References

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

7.1 Definition

We will construct a Markov chain $\{X_n\,;\, n = 0, 1, 2, \dots\}$, where X_n models the number of individuals of the n-th generation. Modeling the situation that each individual in any generation gives birth to a certain number of children and dies, we assume that the number of children obeys a probability distribution common for all individuals and is independent of individuals and of generation.

Now 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. Let $Y_1, Y_2, \dots$ be independent identically distributed random variables, of which the distribution is the same as Y. Then, we define the transition probability by

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

and

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

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

For simplicity we assume that $X_0 = 1$. Moreover, we are interested in the case of

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

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

Problem 35 (One-child policy) Consider the Galton-Watson branching process with offspring distribution satisfying $p_0 + p_1 = 1$. Calculate the probabilities

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

and find the extinction probability

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

7.2 Generating Functions

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

Define the generating function of the offspring distribution by

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

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

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

Lemma 7.2.1

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

Proof By definition,

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

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

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

Hence,

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

which proves the assertion. ∎

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

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

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

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

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

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

Since

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

by assumption of induction, we obtain

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

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

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

which proves the claim for $n + 1$. ∎

Since $X_0 = 1$,

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

In particular,

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

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

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

Then we have

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

Proof Differentiating (7.1), we obtain

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

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

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

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

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

Differentiating both sides, we come to

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

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

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

Therefore,

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

which means that

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

∎

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

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

7.3 Extinction Probability

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

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

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

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

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

Therefore, q satisfies the equation:

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

Proof It follows from Lemma 7.2.2 that

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

Hence,

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

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

Lemma 7.3.2 Assume that the offspring distribution satisfies the conditions:

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

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

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

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

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

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

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

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

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

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

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

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

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

Problem 37 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 $f_n(s)$.