Probability Model / Applied Analysis 2019 [1]

by

Reika Fukuizumi [2]

Probability models are essential in mathematical analysis of random phenomena. In these lectures, we focus on Markov chains as basic models of random time evolution. Starting with fundamental concepts in probability theory (random variables, probability distributions, etc.), we study fundamentals on Markov chains (transition probability, recurrence, stationary distributions, etc.). Background knowledge on elementary probability is required.

The lectures will follow the book *Essence of Probability models* by Nobuaki Obata, (published from Makino shoten in 2012, in Japanese). Other references including English/French texts are for example:

1. W. Feller, "An Introduction to Probability Theory and Its Applications," Vol.1 Wiley, 1957.
2. R. Durrett, "Probability: Theory and Examples" fourth edition, Cambridge University Press 2010.
3. T. Bodineau, "PROMENADE ALÉATOIRE : Chaînes de Markov et martingales," MAP432, Ecole Polytechnique 2013.

To obtain this course's credit, you are required to choose by yourself five problems among the problems given during the lectures, and to submit answers to the five problems as a report.

The report should be handed in to the report box aside the educational affairs office of GSIS.

Deadline: January 27, 2020.

No exception will be made.

The 1-4 sections of this course consist of a review of what we have already learned in the undergraduate program (at least at Tohoku University we hold a course "Mathematical Statics.")

[1] updated: Dec.20 2019

[2] Email: fukuizumi@math.is.tohoku.ac.jp

1. Probability Spaces and Random Variables

Ω: sample space consisting of elementary events (or sample points).
$\mathcal{F}$: the set of events.

Definition (Probability). A function $\mathbb{P} : A \in \mathcal{F} \mapsto \mathbb{P}(A) \in [0,1]$ is said to be the probability on $\mathcal{F}$ (or with domain $\mathcal{F}$) if the following (P1)-(P3) are satisfied:

(P1) $0 \leq \mathbb{P}(A) \leq 1$ for any $A \in \mathcal{F}$.
(P2) $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$.
(P3) For $A_1, A_2, ... \in \mathcal{F}$ (infinite sequence) with $A_j \cap A_k = \emptyset$ when $j \neq k$,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Examples: Coin toss, Dise throwing, random cut.

Definition (Probability Space). Let Ω be a non-empty set, and $\mathbb{P}$ be a probability on $\mathcal{F}$. We call $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space.

Theorem 1.1 Let $A_1, A_2, ...$ be a sequence of events.

(1) If $A_1 \subset A_2 \subset A_3 \subset \cdots$, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} A_n.$$

(2) If $A_1 \supset A_2 \supset A_3 \supset \cdots$, then

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} A_n$$

.

Definition (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,... and corresponding nonnegative numbers p_1, p_2, ... such that

$$\mathbb{P}(X = a_i) = p_i, \quad p_i \geq 0, \quad \sum_i p_i = 1.$$

In this case,

$$\mu_X := \sum p_i \delta_{a_i}$$

is called the (probability) distribution of X. Here, for a Borel set $B \subset \mathbb{R}$,

$$\delta_a(B) = \begin{cases} 1, & a \in B \\ 0, & \text{otherwise} \end{cases}$$

is the Dirac measure at $a \in \mathbb{R}$. Obviously,

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

Examples. Coin toss, Waiting time.

Definition (Continuous random variables). A random variable X is called continuous if $\mathbb{P}(X = a) = 0$ for all $a \in \mathbb{R}$. If there exists a function $f(x)$ such that

$$\mathbb{P}(a \le X \le b) = \int_a^b f(x)dx, \quad a < b, \quad a, b \in \mathbb{R},$$

we say that X admits a probability density function $f(x)$, and denote $f(x)$ by $f_X(x)$. Note that

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1, \quad f_X(x) \ge 0$$

In this case,

$$\mu_X(dx) := f_X(x)dx$$

is called the (probability) distribution of X.

It is useful to consider the distribution function

$$F_X(x) := \mathbb{P}(X \in (-\infty, x]) = \int_{-\infty}^{x} f_X(t)dt, \quad x \in \mathbb{R}.$$

Then, if F_X is continuous and piecewise differentiable, we have

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

Remark.

(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.

Examples. random cut.

Definition (mean value). The mean or expectation value of a random variable X is defined by

$$\begin{aligned} m = \mathbb{E}[X] &:= \int_{-\infty}^{+\infty} x\mu_X(dx) \\ &= \begin{cases} \displaystyle\sum_i a_i p_i & \text{if } X \text{ is discrete,} \\ \displaystyle\int_{-\infty}^{+\infty} x f_X(x)dx & \text{if } X \text{ is continuous} \\ & \quad \text{which admits a probability density function } f_X(x) \end{cases} \end{aligned}$$

Proposition 1.1. For a (measurable) function $\varphi(x)$ we have

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

For example,

- (m-th moment) $\mathbb{E}[X^m] = \int_{-\infty}^{+\infty} x^m \mu_X(dx)$.
- (characteristic function) $\mathbb{E}(e^{itX}) = \int_{-\infty}^{+\infty} e^{itx}\mu(dx)$, $t \in \mathbb{R}$.

Definition (variance). The variance of a random variable X is defined by

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

i.e.

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

Examples. Waiting time, random cut.

2. Probability Distributions

We introduce some classical examples of one-dimensional distributions.

- Discrete distributions
  - Bernoulli distribution
    For $0 \leq p \leq 1$, the distribution $(1-p)\delta_0 + p\delta_1$ is called Bernoulli distribution with success probability p.

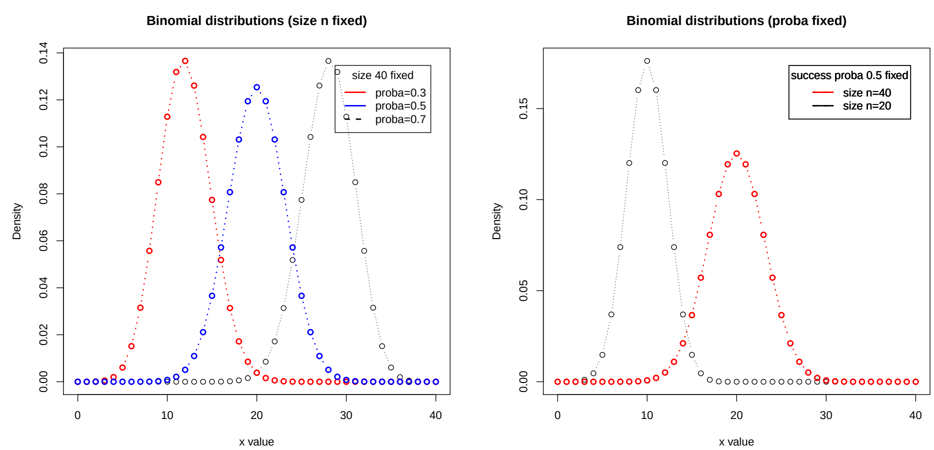
$$m = p, \quad \sigma^2 = p(1-p)$$

  - Binomial distribution $B(n, p)$
    For $0 \leq p \leq 1$ and $n \geq 1$, the distribution

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

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

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



  - Geometric distribution
    For $0 \leq p \leq 1$, the distribution

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

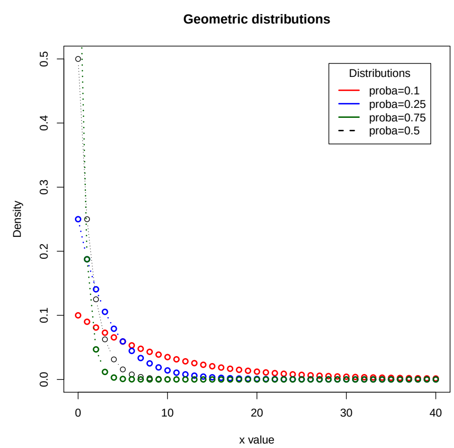
is called the geometric distribution with success probability p.

$$m = \frac{1}{p}, \quad \sigma^2 = \frac{1-p}{p^2}$$

by the computation of the probability generating function,

$$G(z) = \sum_{k=1}^{\infty} p(1-p)^{k-1} z^k, \quad G'(z) = \frac{p}{\{1-(1-p)z\}^2}, \quad G''(z) = \frac{2p(1-p)}{\{1-(1-p)z\}^3}.$$

$$m = G'(1) = 1/p, \quad \sigma^2 = G''(1) + G'(1) - \{G'(1)\}^2 = (1-p)/p^2.$$



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

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

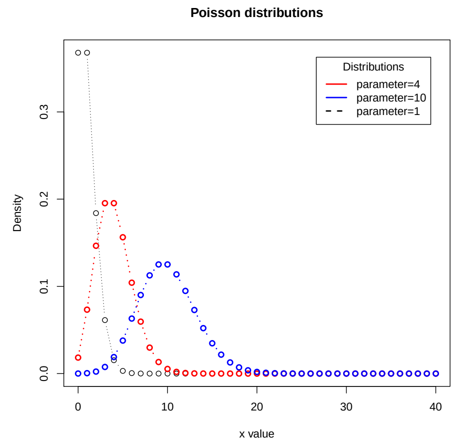
In this case, the mean is $(1-p)/p$ and the variance is $(1-p)/p^2$.

– Poisson distribution
  For $\lambda > 0$ the distribution

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

is called the Poisson distribution with parameter λ.

$$m = \lambda, \quad \sigma^2 = \lambda$$



Problem 1. Find the mean value and variance of the discrete distributions introduced above.

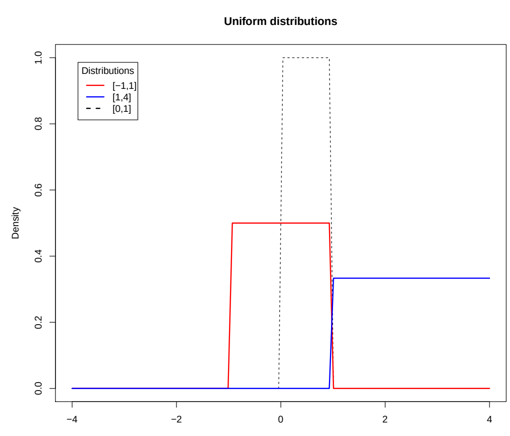
- Continuous distributions
  - Uniform distribution

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

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

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

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

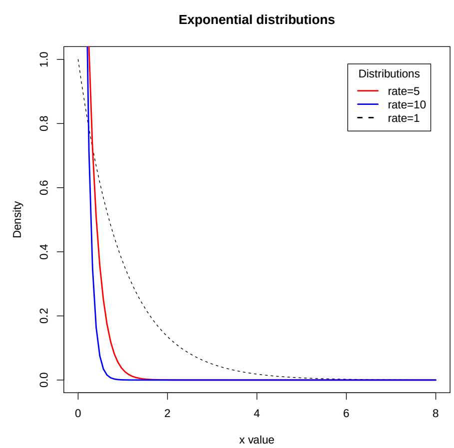


– Exponential distribution
The exponential distribution with parameter $\lambda > 0$ is defined by the density function

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

The mean value and variance are

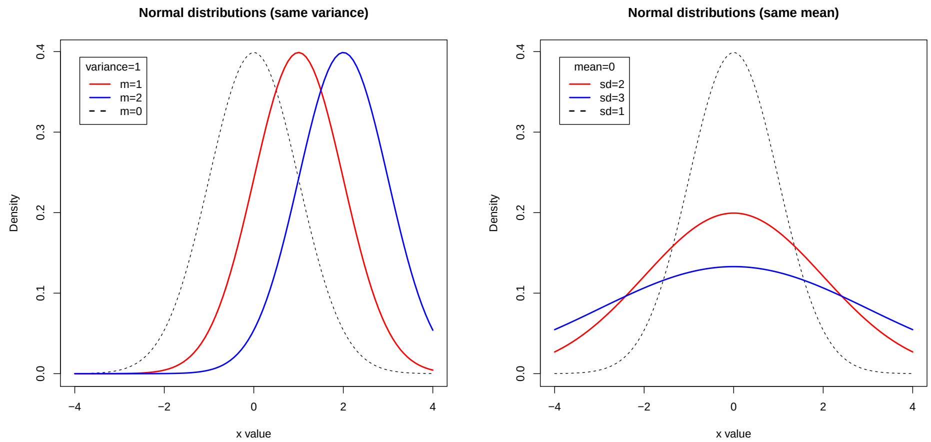
$$m = \frac{1}{\lambda}, \quad \sigma^2 = \frac{1}{\lambda^2}.$$



– Normal distribution $\mathcal{N}(m, \sigma^2)$
For $m \in \mathbb{R}$ and $\sigma > 0$, we see that

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

becomes a density function. The distribution defined by this density function is called the normal distribution or Gaussian distribution. When $m = 0$ and $\sigma^2 = 1$, i.e. $\mathcal{N}(0, 1)$ is called the standard normal distribution or the standard Gaussian distribution.



Recall: $\int_0^{+\infty} e^{-tx^2}dx = \frac{\sqrt{\pi}}{2\sqrt{t}}$.

Thus,

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

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

(1) Find the probability $\mathbb{P}(X \leq x)$, $x \geq 0$.
(2) Find the probability density function $f_X(x)$ of X.
(3) Calculate the mean and variance of X.
(4) Calculate the mean and variance of the area of the inscribed circle: $S = \pi X^2$.

3. Independence and Dependence

- Independent events and conditional probability

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

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

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

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

Example. Drawing randomly a card from a deck of 52 cards.

Remark. It is allowed to consider whether the sequence of events $\{A, A\}$ is independent or not. If they are independent, by definition we have $\mathbb{P}(A) = \mathbb{P}(A)\mathbb{P}(A)$, from which $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$ follows. Notice that $\mathbb{P}(A) = 0$ does not imply $A = \emptyset$ (empty event). Similarly, $\mathbb{P}(A) = 1$ does not imply $A = \Omega$ (whole event).

Definition (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

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{whenever} \quad \mathbb{P}(B) > 0.$$

Theorem. Let A, B be events with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$. A and B are independent iff

$$\mathbb{P}(A|B) = \mathbb{P}(A), \quad \text{and} \quad \mathbb{P}(B|A) = \mathbb{P}(B).$$

- Independent random variables

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

In other words, a (finite or infinite) sequence of random variables $X_1, X_2, ...$ is independent if for any finite $X_{i_1}, ..., X_{i_n}$ $(i_1 < i_2 < .... < i_n)$ and constant numbers $a_1, ..., a_n$, their joint probability has the following property.

$$\mathbb{P}(X_{i_1} \leq a_1, X_{i_2} \leq a_2, ..., X_{i_n} \leq a_n) = \mathbb{P}(X_{i_1} \leq a_1)\mathbb{P}(X_{i_2} \leq a_2) \cdots \mathbb{P}(X_{i_n} \leq a_{i_n}). \quad (0.1)$$

Similar assertion holds for the pairwise independence. If random variables $X_1, X_2, ...$ are discrete, (0.1) may be replaced with

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

Example. Choose at random a point from the rectangle.

Problem 3.

- (1) A box contains four balls with numbers $112, 121, 211, 222$. We draw a ball at random and let X_1 be the first digit, X_2 the second digit, and X_3 the last digit. For $i = 1, 2, 3$ we define an event A_i by $A_i = \{X_i = 1\}$. Show that $\{A_1, A_2, A_3\}$ is pairwise independent but is not independent.
- (2) Two dice are tossed. Let A be the event that the first die gives a 4, B be the event that the sum is 6, and C be the event that the sum is 7. Calculate $\mathbb{P}(B|A)$ and $\mathbb{P}(C|A)$, and study the independence among $\{A, B, C\}$.

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

$$\mathbb{P}(X_n = 1) = p, \quad \mathbb{P}(X_n = 0) = 1 - p.$$

By definition of independence, we have

$$\mathbb{P}(X_1 = a_1, X_2 = a_2, ..., X_n = a_n) = \Pi_{k=1}^{n} \mathbb{P}(X_k = a_k),$$

for all $a_1, ..., a_n \in \{0, 1\}$.

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

- Covariance and correlation coefficients

Recall that the mean of a real-valued (1-dim) random variable X is defined by

$$m = \mathbb{E}(X) = \int_{-\infty}^{\infty} x \mu_X(dx).$$

If $X = (X_1, ..., X_n) \in \mathbb{R}^n$, for a measurable function $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}(\varphi(X)) = \int_{\mathbb{R}^n} \varphi(x) \mu_X(dx), \quad dx = dx_1 dx_2 ....dx_n.$$

Theorem. For two random variables X, Y and two constant numbers a, b it holds that

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

Theorem. If random variables $X_1, X_2, ..., X_n$ are independent, we have

$$\mathbb{E}(X_1 X_2 \cdots X_n) = \mathbb{E}(X_1)\mathbb{E}(X_2)\cdots\mathbb{E}(X_n).$$

Remark. $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ is not a sufficient condition for the random variables X and Y being independent.

Definition (Covariance). The covariance of two random variables X, Y is defined by

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

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

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

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

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

Remark. The converse of this Theorem is not true in general. (see Problem 5.)

Theorem. Let $X_1, X_2, ..., X_n$ be independent random variables. Then

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

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

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

- (1) Show the joint probability of (L, S) by a table.
- (2) Calculate the correlation coefficient ρ_{LS} and explain the meaning of the signature of ρ_{LS}.

<u>Problem 5.</u> Let X and Y be random variables such that

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

where a, b, c, d are constant numbers and $0 < p_1 < 1$, $0 < p_2 < 1$. Show that X, Y are independent if and only if $\sigma_{XY} = 0$. Explain the significance of this case. [Hint: In general, uncorrelated random variables are not necessarily independent.]

4. Limit Theorems

Let $\{X_k\}$ be a Bernoulli trial with success probability 1/2, and consider the binomial process defined by

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

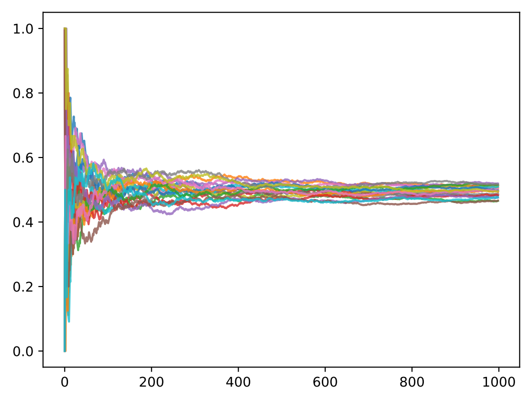
Since S_n counts the number of heads during the first n trials,

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

gives the relative frequency of heads during the first n trials. The following figure is 40 samples randomly chosen showing that the relative frequency of heads $\frac{S_n}{n}$ tends to 1/2. It is our question how to describe this phenomenon mathematically. A naive formula

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

is not acceptable.



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

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

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

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

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

$$\mathbb{P}(|X - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Theorem (Strong law of large numbers). Let $X_1, X_2, ...$ 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 $X_1, X_2, ...$ are pairwise independent, we have

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

In other words,

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

Remark. Kolmogorov proved the strong law of large numbers under the assumption that $X_1, X_2, ...$ are independent. In many literatures, the strong law of large numbers is stated as Kolmogorov proved. Theorem above is due to N. Etemadi (1981), where the assumption is relaxed to being pairwise independent and the proof is more elementary.

Now, consider $X_1, X_2, ...$ independent identically distributed random variables whose mean m. Let $a > m$, and take $\varepsilon = a - m$ in Theorem of weak law of large numbers. Then

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n} \sum_{k=1}^{n} X_k \geq a \right) = 0.$$

In fact, we can see that this convergence is exponential.

Theorem (Cramér). Let $X_1, X_2, ...$ independent identically distributed random variables. Assume that for all $t \in \mathbb{R}$, $\psi(t) := \mathbb{E}(e^{tX_1}) < +\infty$. Then, for any $a > m = \mathbb{E}(X_1)$ and $n = 1, 2, ...,$

$$\mathbb{P}\left(\frac{1}{n} \sum_{k=1}^{n} X_k \geq a \right) \leq e^{-I(a)n},$$

with $I(a) = \sup_{t\in\mathbb{R}}\{at - \log\psi(t)\}$.

Theorem (Central Limit Theorem). Let $Z_1, Z_2, ...$ be independent identically distributed (iid) random variables with mean 0 and variance 1. Then, for any $x \in \mathbb{R}$ it holds that

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

In short,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} Z_k \to \mathcal{N}(0,1), \quad \text{weakly as} \quad n\to\infty.$$

Remark. (The theorem of de Moivre-Laplace —special case of CLT). Let $X_1, X_2, ...$ be a Bernoulli trials with success probability p. Set

$$Z_k = \frac{X_k - m}{\sigma}, \quad m = \mathbb{E}(X_k) = p, \quad \sigma^2 = \mathbb{V}(X_k) = p(1-p).$$

Thus, $Z_1, Z_2, ...$ are iid random variables with 0 and variance 1. Apply the central limit theorem for this $\{Z_k\}$, we have

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

Setting $y = nm + x\sigma\sqrt{n}$, we see that

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt = \frac{1}{\sqrt{2\pi n\sigma^2}}\int_{-\infty}^{y} e^{-\frac{(\theta-nm)^2}{2n\sigma^2}}\,d\theta,$$

which implies for large n

$$\sum_{k=1}^{n} X_k \sim \mathcal{N}(nm, n\sigma^2) = \mathcal{N}(np, np(1-p)).$$

On the other hand, we know that $\sum_{k=1}^{n} X_k$ obeys $B(n,p)$, of which the mean value and variance are given by np and $np(1-p)$. Consequently, for a large n we have

$$B(n,p) \sim \mathcal{N}(np, np(1-p)) :$$

their distribution functions are almost same for large n.

<u>Problem 6</u> (Monte Carlo simulation) Let $f(x)$ be a continuous function on the interval $[0,1]$ and consider the integral

$$\int_0^1 f(x)dx. \tag{0.2}$$

(1) Let X be a random variable obeying the uniform distribution on $[0,1]$. Give expressions of the mean value $\mathbb{E}(f(X))$ and variance $\mathbb{V}(f(X))$ of the random variable $f(X)$.

(2) Let $x_1, x_2, ...$ is a sequence random numbers taken from $[0, 1]$. Explain that the arithmetic mean

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

is a good approximation of the integral (0.2).

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

5. Markov Chains

Recall the conditional probability (see Section 3): 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

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{whenever} \quad \mathbb{P}(B) > 0,$$

i.e.

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

Theorem. For events $A_1, A_2, ..., A_n$, we have

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

Remark. Tree diagram in computation of probability.

Markov chains

Let S be a finite or countable set. Consider a discrete time stochastic process $\{X_n : n = 0, 1, 2...\}$ 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

$$S = \{0, 1\}, \quad S = \{1, 2, ..., N\}, \quad S = \{0, 1, 2, ...\}$$

Definition (Markov Chains). Let $\{X_n : n = 0, 1, 2...\}$ be a discrete time stochastic process over S. It is called a Markov chain over S if

$$\mathbb{P}(X_m = j|X_{n_1} = i_1, X_{n_2} = i_2, \cdots, X_{n_k} = i_k, X_n = i) = \mathbb{P}(X_m = j|X_n = i) \tag{0.3}$$

holds for any $0 \le n_1 < n_2 < \cdots < n_k < n < m$ and $i_1, i_2, ..., i_k, i, j \in S$.

Remark. The property (0.3) is called Markov property. Markov property is weaker than the independence.

Theorem (multiplication rule). Let $\{X_n\}$ be a Markov chain over S . Then, for any $0 \le n_1 < n_2 < \cdots < n_k$ and $i_1, i_2, \cdots i_k \in S$, we have

$$\mathbb{P}(X_{n_1} = i_1, X_{n_2} = i_2, \cdots, X_{n_k} = i_k) =$$
$$\mathbb{P}(X_{n_1} = i_1)\mathbb{P}(X_{n_2} = i_2|X_{n_1} = i_1)\mathbb{P}(X_{n_3} = i_3|X_{n_2} = i_2) \cdots \mathbb{P}(X_{n_k} = i_k|X_{n_{k-1}} = i_{k-1}).$$

Definition (Transition probability). For a Markov chain $\{X_n\}$ over S,

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

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.

Hereafter a Markov chain is *always assumed to be time homogeneous.* In this case the transition probability is denoted by

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

and $P := [p_{i,j}]$ is called the transition matrix.

Definition. A matrix $P = [p_{i,j}]$ with index set $S \times S$ is called a stochastic matrix if

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

Theorem. 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.

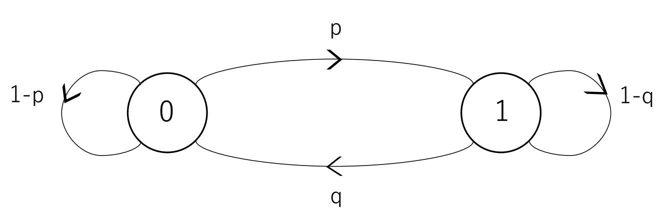
Example 5.1 (2-state Markov chain). A Markov chain over the state space $\{0, 1\}$ is determined by the transition probabilities:

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

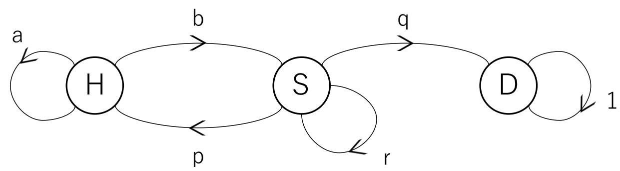
The transition matrix is defined by

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

The transition diagram is as follows:



Example 5.2 (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

$$\begin{bmatrix} a & b & 0 \\ p & r & q \\ 0 & 0 & 1 \end{bmatrix}$$

where $a + b = 1$ and $p + q + r = 1$.

Example 5.3 (Random walk on $\mathbb{Z}^1$). The transition probabilities are given by

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

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

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

Example 5.4 (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 $S = \{-A, -A + 1, ..., B - 1, B\}$. Then the transition probabilities are given as follows.

For $-A < i < B$,

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

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

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

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

In a matrix form we have

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

Example 5.5 (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 $S = \{-A, -A+1, \ldots.., B-1, B\}$. The transition probabilities are given as follows.

For $-A < i < B$,

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

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

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

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

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

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

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

$$\pi(n) = [....., \pi_i(n), ...], \quad \pi_i(n) = \mathbb{P}(X_n = i)$$

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

$$\pi(0) = [......, 0, 1, 0, ...]$$

where 1 occurs at i th position. In this case the Markov chain $\{X_n\}$ starts from the state i.

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

$$p_N(i, j) = \mathbb{P}(X_{n+N} = j | X_n = i), \quad i.j \in S, \quad N = 0, 1, 2, ..$$

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

Theorem (Chapman-Kolmogorov equation). For $0 \leq r \leq n$, we have

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

Recall $P = [p_{ij}]$: the transition matrix (independent of n).

$$\begin{aligned} & \mathbb{P}(X_m = i, X_{m+1} = i_1, \cdots, X_{m+n-1} = i_{n-1}, X_{m+n} = j) \\ = \; & \mathbb{P}(X_m = i) \mathbb{P}(X_{m+1} = i_1 | X_m = i) \cdots \mathbb{P}(X_{m+n} = j | X_{m+n-1} = i_{n-1}) \\ = \; & \mathbb{P}(X_m = i) p(i, i_1) p(i_1, i_2) \cdots p(i_{n-1}, j). \end{aligned}$$

Take the sum with respect to $i_1, \cdots, i_{n-1} \in S$ on the both side, and we obtain the following important result.

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

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

Theorem. We have

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

or equivalently,

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

Remark. Therefore, $\pi(n) = \pi(0)P^n$.

Example 5.6 (2-state Markov chain). Let $\{X_n\}$ be the Markov chain introduced in Example 5.1. The transition matrix has the eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 1-p-q$, and $\lambda_1 \neq \lambda_2$ if $p+q>0$. We consider this case, i.e., the case that P has two distinct eigenvalues. By standard argument, we obtain

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

where we put $r = 1-p-q$.

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

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

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

(1) When will the party B command a majority?
(2) Find the final ratio of supporters after many elections if the same situation continues.

Problem 8. Study the n-step transition probability of the three-state Markov chain introduced in Example 5.2. Explain that every animal dies within finite time if $b>0$ and $q>0$.

Problem 9. 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 $\pi(0) = [\frac{q}{p+q}, \frac{p}{p+q}]$. Calculate the following statistical quantities:

$$\mathbb{E}(X_n), \quad \mathbb{V}(X_n), \quad \mathrm{Cov}(X_{m+n}, X_n).$$

6. Stationary distributions

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

$$\pi = \pi P \tag{0.4}$$

or equivalently if

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}, \quad j \in S. \tag{0.5}$$

Thus, in order to find a stationary distribution of a Markov chain, we need to solve the linear system 0.4 (or equivalently (0.5)) together with the conditions:

$$\sum_{i \in S} \pi_i = 1, \quad \text{and} \quad \pi \geq 0 \quad \text{for all} \quad i \in S$$

Examples. 2-state Markov chain, Random walk on $\mathbb{Z}^1$.

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

(For the proof see the textbooks.)

Remark. Note that the stationary distribution mentioned in the above theorem is not necessarily unique.

Definition. We say that a state j can be reached from a state i if there exists some $n \geq 0$ such that $p_n(i, j) > 0$. By definition every state i can be reached from itself. We say that two states i and j intercommunicate if i can be reached form j and j can be reached from i, i.e., there exist $m \geq 0$ and $n \geq 0$ such that $p_n(i, j) > 0$ and $p_m(j, i) > 0$. For $i, j \in S$ we introduce a binary relation $i \sim j$ when they intercommunicate. Then $\sim$ becomes an equivalence relation on S:

(i) $i \sim i$
(ii) $i \sim j \Longrightarrow j \sim i$
(iii) $i \sim j$, $j \sim k \Longrightarrow i \sim k$.

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

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

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

Now we recall the example of 2-state Markov chain. If $p+q>0$, the distribution of the above Markov chain converges to the unique stationary distribution. Consider the case of $p=q=1$, i.e., the transition matrix becomes

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

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

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

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

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

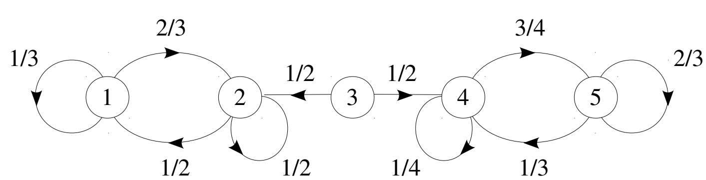
is called the period of i. (When the set in the right-hand side is empty, the period is not defined.) A state $i\in S$ is called aperiodic if its period is one.

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

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

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

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



Problem 11. Let $\{X_n\}$ be the Markov chain introduced in Example 5.2.

(1) Find that if $q>0$ and $b>0$, a stationary distribution is unique and given by $\pi=[0,0,1]$.

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

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

$$\begin{cases} H_n &= aH_{n-1}+bS_{n-1},\\ S_n &= pH_{n-1}+rS_{n-1},\end{cases}$$

where $n \geq 2$, $H_1 = 0$, $S_1 = q$.

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

$$\mathbb{E}[H] = \frac{b+p+q}{bq}, \quad \mathbb{E}[S] = \frac{b+p}{bq}.$$

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

$$p_{i,j} = \begin{cases} \dfrac{1}{\deg(i)}, & \text{if} \quad i \to j \quad \text{and} \quad \deg(i) \neq 0, \\ 0, & \text{if} \quad i \not\to j \quad \text{and} \quad i \neq j, \\ 1 & \text{if} \quad i \to j \quad \text{and} \quad j = i \quad \text{and} \quad \deg(i) = 0. \end{cases}$$

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

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

$$Q = [q_{i,j}],$$
$$q_{i,j} = dp_{i,j} + \varepsilon, \qquad \varepsilon = \frac{1-d}{N}.$$

If $0 \leq d < 1$, the Markov chain determined by Q has necessarily a unique stationary distribution. Choosing a suitable $d < 1$, we may understand the stationary distribution $\pi = [\pi_i]$ as the page rank among the websites. In real application d should not be close to 0 and $d \approx 0.85$ is often taken.

7. Recurrence

Definition. Let S be a state space and $i \in S$. The first hitting time or first passage time to i is defined by

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

If $\{n \geq 1; X_n = i\}$ is an empty set, we define $T_i = +\infty$. A state $i \in S$ is called recurrent if $\mathbb{P}(T_i < +\infty | X_0 = i) = 1$. It is called transient if $\mathbb{P}(T_i = +\infty | X_0 = i) > 0$.

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

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

If a state $i \in S$ is transient, we have

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

and moreover,

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

Examples. Random walk on $\mathbb{Z}^1$, $\mathbb{Z}^2$ and $\mathbb{Z}^3$. The following notation will be used:
Let $\{a_n\}$ and $\{b_n\}$ be sequences of positive numbers. We write $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$. In this case, there exist two constant numbers $c_1 > 0$ and $c_2 > 0$ such that $c_1 a_n \leq b_n \leq c_2 a_n$. Hence $\sum_n a_n$ and $\sum_n b_n$ converge or diverge at the same time.

Definition. If a state $i \in S$ is recurrent, i.e., $\mathbb{P}(T_i < \infty | X_0 = i) = 1$, *the mean recurrent time* is defined by

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

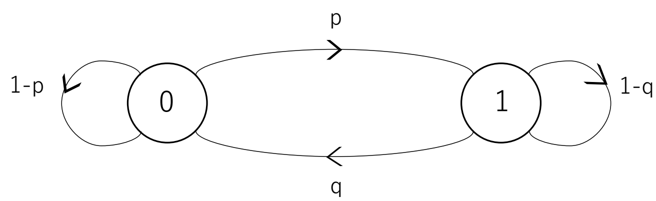
The state i is called positive recurrent if $\mathbb{E}(T_i | X_0 = i) < \infty$, and null recurrent otherwise.

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

Example. The mean recurrent time of the one-dimensional isotropic random walk is infinity, i.e., the one-dimensional isotropic random walk is null recurrent.

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



where $p > 0$ and $q > 0$. For a state $i \in S$ let $T_i = \inf\{n \geq 1; X_n = i\}$ be the first hitting time to i.

(1) Calculate

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

(2) Find $\mathbb{P}(T_0 = n|X_0 = 0)$ and calculate

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

8. Bienaymé-Galton-Watson Branching Process

Consider a simplified family tree where each individual gives birth to offspring (children) and dies. The number of offsprings is random. We are interested in whether the family survives or not.

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

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

$$\mathbb{P}(Y = k) = p_k, \quad k = 0, 1, 2, ...$$

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

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

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

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

and

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

The above Markov chain $\{X_n\}$ over the state space $\{0, 1, 2, ...\}$ is called the Bienaymé-Galton-Watson branching process with offspring distribution $\{p_k : k = 0, 1, 2, ...\}$. For simplicity we assume that $X_0 = 1$. When $p_0 + p_1 = 1$, the famility tree is reduced to just a path without branching so the situation is much simpler (Problem 13).

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

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

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

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

Lemma.

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

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

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

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

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

Then, we have

$$E[X_n] = m^n.$$

In conclusion, the mean value of the number of individuals in the n-th generation, $\mathbb{E}(X_n)$, 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 $m < 1$.

The event $\{X_n = 0$ occurs for some $n \ge 1\}$ means that the family died out before or at the n-th generation. Thus,

$$q = \mathbb{P}\Big(\bigcup_{n=1}^{\infty} \{X_n = 0\}\Big) = \lim_{n\to\infty} \mathbb{P}(X_n = 0)$$

is the probability of extinction of the family. 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. Let $f(s)$ be the generating function of the offspring distribution. Then we have

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

Therefore, q satisfies the equation:

$$q = f(q)$$

Assume that the offspring distribution satisfies the conditions: $p_0 + p_1 < 1$.

Lemma. The generating function $f(s)$ satisfies 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

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

Lemma.

(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 $f(s) = s$.

Theorem. The extinction probability q of the Bienaymé-Galton-Watson branching process as above coincides with the smallest s such that $s = f(s)$, $0 \le s \le 1$. Moreover, if $m \le 1$ we have $q = 1$, and if $m > 1$ we have $q < 1$.

The Bienaymé-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 13 (One-child policy). Consider the Bienaymé-Galton-Watson branching process with offspring distribution satisfying $p_0 + p_1 = 1$. Calculate the probabilities

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

and find the extinction probability

$$\mathbb{P}(X_n = 0 \text{ occurs for some } n \ge 1).$$

Problem 14. Let b, p be constant numbers such that $b > 0$, $0 < p < 1$ and $b + p < 1$. Suppose that the offspring distribution given by

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

(1) Find the generating function $f(s)$ of the offspring distribution.
(2) Set $m = 1$ and find $f_n(s)$.