

STOCHASTIC PROCESSES

October 9, 2019

Contents

1	Introduction to Probability Theory	3
1.1	Exercises	3
2	Random variables	5
2.1	Discrete Random variables	5
2.2	Continuous Random variables	7
2.3	Central Limit Theorem	11
2.4	Exercises	11
3	Conditional Expectation	13
3.1	Exercises	14
4	Stochastic processes	17
4.1	Exercises	18
5	Markov Chains	20
5.1	Exercises	22
6	Poisson process	25
6.1	Counting process	25
6.2	Definition of the Poisson process	25
6.3	Interarrival and waiting time distributions	26
6.4	Properties of Poisson process	27
6.5	Nonhomogeneous Poisson process	28
6.6	Compound Poisson Processes	28
6.7	Exercises	28
7	Brownian Motion	31
7.1	Exercises	32
8	Discrete Martingales	34
8.1	Martingales and Brownian motion W_t	35
8.2	Exercises	36
9	Appendix	38
9.1	Eigenvalues and eigenvectors of a matrix	38

1 Introduction to Probability Theory

Denote Ω as a *sample space*. Any subset of the sample space is known as an *event*. We have

$$P(E \cup F) = P(E) + P(F) - P(EF).$$

When E and F are *mutually exclusive* ($EF = \emptyset$) we have

$$P(E \cup F) = P(E) + P(F).$$

Conditional probability is probability that E occurs given that F has occurred and is denoted by $P(E|F)$,

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

Because we know that F has occurred, it follows that F become our new sample space and hence the probability that the event EF occurs will equal the probability of EF relative to the probability of F .

Two events E and F are said to be **independent** if

$$P(EF) = P(E)P(F).$$

This implies that E and F are independent if

$$P(E|F) = P(E).$$

Multiplication Rule

Assuming that all the conditioning events have positive probability, we have

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right).$$

Suppose that F_1, F_2, \dots, F_n are mutually exclusive events such that $\bigcup_{i=1}^n F_i = \Omega$ (exactly one of the events F_1, \dots, F_n will occur). Using the fact that events EF_i are mutually exclusive, we obtain that

$$P(E) = \sum_{i=1}^n P(EF_i) = \sum_{i=1}^n P(E|F_i)P(F_i).$$

Also, we have equation

$$P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)}$$

which is known as *Bayes' formula*.

1.1 Exercises

1. A batch of 25 injection-molded parts contains 5 that have suffered excessive shrinkage.
 - (a) If two parts are selected at random, and without replacement, what is the probability that the second part selected is one with excessive shrinkage?

- (b) If three parts are selected at random, and without replacement, what is the probability that the third part selected is one with excessive shrinkage?

Solution: Let A denote an event that the first part selected has excessive shrinkage and let B denote the event that the second part selected has excessive shrinkage.

(a) $P(A) = \frac{5}{25} = \frac{1}{5}$.

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) = \frac{4}{24} \frac{1}{5} + \frac{5}{24} \frac{4}{5} = 0.2.$$

- (b) Denote by $A_1 = A \cap B$, $A_2 = A \cap \bar{B} + \bar{A} \cap B$ and $A_3 = \bar{A}\bar{B}$. Then

$$P(A_1) = \frac{1}{30}, \quad P(A_2) = \frac{1}{3}, \quad P(A_3) = \frac{19}{30}.$$

Let C denote an event that the third part selected has excessive shrinkage. We have

$$\begin{aligned} P(C) &= P(C|A_1)P(A_1) + P(C|A_2)P(A_2) + P(C|A_3)P(A_3) \\ &= \frac{3}{23} \frac{1}{30} + \frac{4}{23} \frac{1}{3} + \frac{5}{23} \frac{19}{30} = 0.2. \end{aligned}$$

2. One considers 100 people of whom 40 speak Russian, 30 English and 21 French. Then 15 speak Russian and English, 10 Russian and French, 5 French and English and 3 of them speak all three languages. Then one randomly chooses 3 of them.

- (a) What is the probability that neither of 3 people speak foreign language?
 (b) What is the probability that all three people speak Russian language?
 (c) What is the probability that two of them speak some foreign language, but the third person speaks no foreign language?

Solution: $P(\text{person speaks Russian}) = \frac{40}{100}$, $P(\text{person speaks French}) = \frac{21}{100}$,
 $P(\text{person speaks English}) = \frac{30}{100}$, $P(\text{person speaks no foreign language}) = \frac{36}{100}$.

(a) $P(\text{neither of 3 people speak foreign language}) = \frac{36 \cdot 35 \cdot 34}{100 \cdot 99 \cdot 98}$

(b) $P(\text{three people speak Russian language}) = \frac{40 \cdot 39 \cdot 38}{100 \cdot 99 \cdot 98}$

- (c) Let A be an event that two of them speak some foreign language, but the third person speaks no foreign language. $P(A) = \frac{\binom{64}{2} \binom{36}{1}}{\binom{100}{3}}$

2 Random variables

A **random variable** is a real-valued function of the outcome of the experiment.

2.1 Discrete Random variables

A **discrete random variable** is a real-valued function of the outcome of the experiment that can take a finite or countably infinite number of values.

A discrete random variable has an associated **probability mass function (PMF)**, which gives the probability of each numerical value that random variable can take and it is denoted p_X . If x is any real number, the probability mass of x , denoted $p_X(x)$, is the probability of the event $\{X = x\}$ consisting of all outcomes that give rise to a value of X equals to x :

$$p_X(x) = P(\{x = X\}).$$

We have

$$\sum_x p_X(x) = 1$$

and for any set S of possible values of X , we have

$$P(X \in S) = \sum_{x \in S} p_X(x).$$

We define the **expected value** (also called the expectation or the mean) of a discrete random variable X , with PMF p_X , by

$$E(X) = \sum_x xp_X(x).$$

The expected value of the random variable $g(X)$ is given by

$$E(g(X)) = \sum_x g(x)p_X(x).$$

The **variance** $\text{var}(X)$ (or $D(X)$) of a random variable X is defined by

$$\text{var}(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2.$$

Its square root is denoted by σ_X and is called the **standard deviation**.

The Bernoulli Random Variable

The PMF of X which is a Bernoulli random variable with parameter p is given by

$$p_X(k) = P(X = k) = \begin{cases} 1 - p, & x = 0 \\ p, & x = 1 \\ 0, & \text{otherwise} \end{cases}.$$

$$E(X) = p, \quad \text{var}(X) = p(1 - p).$$

The Binomial Random Variable

The PMF of X which is a Binomial random variable with parameters n and p consist of the binomial probabilities that are given by

$$p_X(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$
$$E(X) = np, \quad \text{var}(X) = np(1-p).$$

The Geometric Random Variable

The PMF of X which is a geometric random variable with parameter p is given by

$$p_X(k) = P(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, \dots$$
$$E(X) = \frac{1}{p}, \quad \text{var}(X) = \frac{1-p}{p^2}.$$

The Poisson Random Variable

A Poisson random variable has a PMF of X given by

$$p_X(k) = P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots,$$

where λ is a positive parameter characterizing the PMF.

The Poisson PMF with parameter λ is a good approximation for a binomial PMF with parameters n and p , i.e.

$$e^{-\lambda} \frac{\lambda^k}{k!} \approx \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \quad \text{if } k < n,$$

provided $\lambda = np$, n is very large, and p is very small.

$$E(X) = \lambda, \quad \text{var}(X) = \lambda.$$

Let X and Y be discrete random variables associated with the same experiment.

- The **joint PMF** $p_{X,Y}$ of X and Y is defined by

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

- The **marginal PMFs** of X and Y can be obtained from the joint PMF, using the formulas

$$p_X(x) = P(X = x) = \sum_{R_x} p_{X,Y}(x, y) \quad \text{and} \quad p_Y(y) = P(Y = y) = \sum_{R_y} p_{X,Y}(x, y)$$

where R_x denotes the set of all points in the range of (X, Y) for which $X = x$ and R_y denotes the set of all points in the range of (X, Y) for which $Y = y$.

- If the marginal probability distribution of the discrete random variable X has the probability mass function $p_X(x)$, then

$$E(X) = \sum_x xp_X(x) = \sum_x x \left(\sum_{R_x} p_{X,Y}(x, y) \right) = \sum_R xp_{X,Y}(x, y)$$

and

$$\text{var}(X) = \sum_x (x - E(X))^2 p_X(x) = \sum_R (x - E(X))^2 p_{X,Y}(x, y)$$

where R denotes the set of all points in the range of (X, Y) .

- Given discrete random variables X and Y with joint probability mass function $p_{X,Y}(x, y)$ the **conditional probability mass function** of Y given $X = x$ is

$$p_{Y|\{X=x\}}(y) = \frac{p_{X,Y}(x, y)}{p_X(x)} \text{ for } p_X(x) > 0.$$

- A function $g(X, Y)$ of X and Y defines another random variable, and

$$E(g(X, Y)) = \sum_x \sum_y g(x, y) p_{X,Y}(x, y).$$

If g is linear, of form $aX + bY + c$, we have

$$E(aX + bY + c) = aE(X) + bE(Y) + c.$$

2.2 Continuous Random variables

A random variable X is called **continuous** if there is a nonnegative function φ_X , called **probability density function of X** , or PDF for short, such that

$$P(X \in B) = \int_B \varphi_X(x) dx,$$

for every subset B of the real line. We have

$$P(a \leq X \leq b) = \int_a^b \varphi_X(x) dx, \quad P(X = a) = \int_a^a \varphi_X(x) dx = 0,$$

$$\int_{-\infty}^{\infty} \varphi_X(x) dx = P(-\infty < X < \infty) = 1.$$

Let X be a continuous random variable with PDF φ_X .

- The **expectation of X** is defined by

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

- The expected values rule for a function $g(X)$ has the form

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)\varphi_X(x) dx.$$

- The variance of X is defined by

$$\text{var}(X) = E((X - E(X))^2) = \int_{-\infty}^{\infty} (x - E(X))^2\varphi_X(x) dx.$$

- We have

$$0 \leq \text{var}(X) = E(X^2) - (E(X))^2.$$

- If $Y = aX + b$, where a and b are given scalars, then

$$E(Y) = aE(X) + b, \quad \text{var}(Y) = a^2 \text{var}(X).$$

The **cumulative distribution function** (CDF for short) F_X of a random variable X is defined by

$$F_X(x) = P(X \leq x), \quad \text{for all } x,$$

and has the following properties.

- F_X is monotonically non-decreasing:

$$\text{if } x \leq y \text{ then } F_X(x) \leq F_X(y).$$

- $F_X(x)$ tends to 0 as $x \rightarrow -\infty$ and to 1 as $x \rightarrow \infty$.
- If X is discrete then $F_X(x)$ is a piecewise constant function of x .
- If X is continuous then $F_X(x)$ is a continuous function of x .
- If X is discrete and takes integer values, the PMF and CDF can be obtained from each other by summing or differencing:

$$F_X(k) = \sum_{i=-\infty}^k p_X(i),$$

$$p_X(k) = P(X \leq k) - P(X \leq k - 1) = F_X(k) - F_X(k - 1),$$

for all integer k .

- If X is continuous, the PDF and CDF can be obtained from each other by integration or differentiation:

$$F_X(x) = \int_{-\infty}^x \varphi_X(t) dt, \quad \varphi_X(x) = \frac{dF_X}{dx}(x).$$

(The second equality is valid for those x at which the PDF is continuous.)

Let X and Y be jointly continuous random variables with joint PDF $\varphi_{X,Y}$.

- The **joint PDF** is used to calculate probabilities:

$$P((X, Y) \in B) = \iint_{(x,y) \in B} \varphi_{X,Y}(x, y) dx dy.$$

- The **marginal PDFs** of X and Y can be obtained from the joint PDF, using the formulas

$$\varphi_X(x) = \int_{-\infty}^{\infty} \varphi_{X,Y}(x, y) dy, \quad \varphi_Y(y) = \int_{-\infty}^{\infty} \varphi_{X,Y}(x, y) dx.$$

- The **joint CDF** is defined by $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$, and determines the joint PDF through the formula

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

for every (x, y) at which the joint PDF is continuous.

- Given continuous random variables X and Y with joint probability density function $\varphi_{X,Y}(x, y)$ the **conditional probability density function** of Y given $X = x$ is

$$\varphi_{Y|\{X=x\}}(y) = \frac{\varphi_{X,Y}(x, y)}{\varphi_X(x)}.$$

- A function $g(X, Y)$ of X and Y defines a new random variable, and

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \varphi_{X,Y}(x, y) dx dy.$$

If g is linear, of the form $aX + bY + c$, we have

$$E(aX + bY + c) = aE(X) + bE(Y) + c.$$

The Exponential Random Variable $X : \mathcal{E}(\lambda)$

For $\lambda > 0$,

$$\varphi_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}, \quad F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

$$E(X) = \frac{1}{\lambda}, \quad \text{var}(X) = \frac{1}{\lambda^2}.$$

The Normal Random Variable $X : \mathcal{N}(\mu, \sigma^2)$

For $\sigma > 0$, $-\infty < \mu < \infty$

$$\varphi_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

$$P\{a < X < b\} = F_X(b) - F_X(a) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

$$E(X) = \mu, \quad \text{var}(X) = \sigma^2.$$

Special case: **Standard Normal random variable** $X^* : \mathcal{N}(0, 1)$:

$$\varphi_{X^*}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

$$\Phi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

$$F_{X^*}(x) = 0.5 + \Phi(x)$$

1. $\Phi(0) = 0$
2. $\Phi(-x) = -\Phi(x)$
3. $\Phi(x) = \frac{1}{2}$, for $x \geq 5$
4. $\Phi(x) \approx \frac{1}{2}$ for $3.5 \leq x \leq 5$

Proposition 2.1. *Let Z be a standard normal random variable, let $\mu \in \mathbb{R}$, and let $\sigma^2 > 0$. Then the random variable*

$$X = \sigma Z + \mu$$

is a normal random variable with parameters μ and σ^2 .

The Uniform Random Variable $X : \mathcal{U}(a, b)$

For $a < b$,

$$\varphi_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}, \quad F_X(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1, & x > b \end{cases}.$$

$$E(X) = \frac{a+b}{2}, \quad \text{var}(X) = \frac{(b-a)^2}{12}.$$

2.3 Central Limit Theorem

Theorem 2.1. Let X_1, X_2, \dots be a sequence of independent, identically distributed random variables, each with mean μ and variance σ^2 . Then the distribution of

$$\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \rightarrow \infty$. That is,

$$P\left(\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{x^2}{2}} dx$$

as $n \rightarrow \infty$.

De Moivre-Laplace Formula

Theorem 2.2. Consider a sequence of Bernoulli trials with probability p of success. Let S_n , $n \in \mathbb{N}$, denote the number of successes in the first n trials. For any $a, b \in \mathbb{R}$, with $a < b$,

$$\lim_{n \rightarrow \infty} P\left\{a < \frac{S_n - np}{\sqrt{npq}} < b\right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx = \Phi(b) - \Phi(a).$$

Remark 2.1. De Moivre-Laplace theorem is a special case of the Central Limit Theorem. It is useful in the cases when one has the binomial random variable with n large. More precisely, when n is large and $np \geq 10$ binomial random variable can be approximated by normal random variable.

2.4 Exercises

1. Consider the following joint probability density function

$$\varphi_{X,Y}(x, y) = \begin{cases} c e^{-(2x+3y)}, & x \geq 0, y \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Determine the value of c and then determine the marginal probability density function of X and $P\{X > 1/2 \text{ and } Y > 1/3\}$.

Solution: Since $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi_{X,Y}(x, y) = 1$ we have

$$\begin{aligned} 1 &= \int_0^{\infty} \int_0^{\infty} c e^{-(2x+3y)} dy dx = c \int_0^{\infty} e^{-2x} \int_0^{\infty} e^{-3y} dy dx \\ &= \frac{c}{3} \int_0^{\infty} e^{-2x} (e^0 - e^{-\infty}) dx = \frac{c}{2 \cdot 3} (e^0 - e^{-\infty}) = \frac{c}{6}. \end{aligned}$$

It follows that $c = 6$.

$$\varphi_X(x) = \int_{-\infty}^{\infty} \varphi_{X,Y}(x, y) dy = \int_0^{\infty} 6 e^{-(2x+3y)} dy = 2 e^{-2x}, \quad x \geq 0$$

and $\varphi_X(x) = 0$, for $x < 0$.

$$P\left\{X > \frac{1}{2} \text{ and } Y > \frac{1}{3}\right\} = \int_{1/2}^{\infty} \int_{1/3}^{\infty} 6 e^{-(2x+3y)} dy dx = 6 \frac{1}{6} e^{-2} = e^{-2}.$$

2. At noon on a weekday, we begin recording new call attempts at a telephone switch. Let X denote the arrival time of the first call, as measured by the number of seconds after noon. Let Y denote the arrival time of the second call. In the most common model used in the telephone industry, X and Y are continuous random variables with joint probability density function

$$\varphi_{X,Y}(x,y) = \begin{cases} \lambda^2 e^{-\lambda y}, & 0 \leq x < y, \\ 0, & \text{otherwise,} \end{cases}$$

where $\lambda > 0$ calls/second is the average arrival rate of telephone calls. Find the marginal probability density function of Y and the conditional probability density function $\varphi_{Y|\{X=x\}}(y)$.

Solution: For $x < 0$, $\varphi_X(x) = 0$. For $x \geq 0$ we have

$$\varphi_X(x) = \int_x^\infty \lambda^2 e^{-\lambda y} dy = \lambda e^{-\lambda x}.$$

Also,

$$\varphi_Y(y) = \int_0^y \lambda^2 e^{-\lambda y} dx = \lambda^2 y e^{-\lambda y}, \text{ for } y > 0$$

and $\varphi_Y(y) = 0$, for $y \leq 0$. Then

$$\varphi_{Y|\{X=x\}}(y) = \frac{\varphi_{X,Y}(x,y)}{\varphi_X(x)} = \begin{cases} \lambda e^{-\lambda(y-x)}, & y > x \\ 0, & \text{otherwise.} \end{cases}$$

3. On a multiple-choice exam with three possible answers for each of the five questions, what is the probability that a student would get four or more correct answers just by guessing?
4. Let $X : \mathcal{N}(7, 4)$ be the random variable that represents a lifetime of expensive device. If the device breaks during the first two year the amount of x dinars is paid to insurer, while if it breaks during the third or the fourth year the amount of $\frac{1}{2}x$ dinars is paid to the insurer. The insurance is paid only if the lifetime of the device is less than 4 years. Find the value for x such that the expected payment per device is 5000 dinars.
5. Find the constant $c \in \mathbb{R}$ such that the function

$$\varphi(x) = c|x|e^{-x^2}, x \in \mathbb{R}$$

is probability density function of some random variable X and determine $D(X)$.

3 Conditional Expectation

Let (Ω, \mathcal{F}, P) be a probability space, that is, Ω is a set (finite or infinite), \mathcal{F} is a σ -field and P is a probability measure. σ -field \mathcal{F} is a family of subsets of Ω satisfying the three properties:

- (1) $\emptyset, \Omega \in \mathcal{F}$
- (2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
- (3) $A_n \in \mathcal{F} (n = 1, 2, \dots) \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

A map $X : \Omega \rightarrow \mathbb{R}^n$ is a random variable on (Ω, \mathcal{F}, P) if $X^{-1}(B) \in \mathcal{F}$, for $B \in \mathcal{B}_n$. X is the random variable if and only if X is an \mathcal{F} -measurable function on Ω .

If X is a simple random variable then

$$X(w) = \sum x_k I_{A_k}(w), \quad A_i = \{w | X(w) = x_i\}.$$

$$\mathcal{F}(X) := \{X^{-1}(B), B \in \mathcal{B}_n\}$$

is a σ -algebra.

If $Y = \Phi(X)$ then Y is $\mathcal{F}(X)$ -measurable.

We know that

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

- Let $B \in \mathcal{F}$. Then

$$E(X|B) = \frac{1}{P(B)} \int_B X dP$$

is a number.

- Let X and Y be two random variables.

– If Y is a simple random variable then

$$E(X|Y) := \begin{cases} E(X|A_1), & \text{on } A_1 \\ \vdots & \vdots \\ E(X|A_m), & \text{on } A_m \end{cases} = \begin{cases} \frac{1}{P(A_1)} \int_{A_1} X dP, & \text{on } A_1 \\ \vdots & \vdots \\ \frac{1}{P(A_m)} \int_{A_m} X dP, & \text{on } A_m \end{cases}$$

is a random variable.

$E(X|Y)$ is $\mathcal{F}(Y)$ -measurable because $E(X|Y)$ is a function of Y .

$$\int_A X dP = \int_A E(X|Y) dP, \quad \text{for all } A \in \mathcal{F}(Y)$$

– If Y is arbitrary random variable conditional expectation is each $\mathcal{F}(Y)$ -measurable random variable which satisfy

$$\int_A X dP = \int_A E(X|Y) dP, \quad \text{for all } A \in \mathcal{F}(Y).$$

- Let $\mathcal{V} \subseteq \mathcal{F}$ be a σ -algebra contained in \mathcal{F} . Then the conditional expectation of the random variable X given \mathcal{V} is a \mathcal{V} -measurable random variable such that

$$\int_A X dP = \int_A E(X|\mathcal{V})dP, \quad \text{for all } A \in \mathcal{V}.$$

If X is integrable random variable then for each σ -algebra \mathcal{V} , $\mathcal{V} \subseteq \mathcal{F}$ conditional expectation $E(X|\mathcal{V})$ exists and is unique. We have:

1. $E(X|Y) = E(X|\mathcal{F}(Y))$,
2. $E(X|\mathcal{F}_0) = E(X)$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$,
3. $E(E(X|\mathcal{V})) = E(X)$.

Some properties of a conditional expectation:

1. If X is \mathcal{V} -measurable then $E(X|\mathcal{V}) = X$ a.s. because the conditional expectation given the full σ -algebra, corresponds to complete information about events.
2. $a, b \in \mathbb{R}$ $E(aX + bY|\mathcal{V}) = aE(X|\mathcal{V}) + bE(Y|\mathcal{V})$ a.s.
3. If X is \mathcal{V} -measurable and XY is integrable then $E(XY|\mathcal{V}) = XE(Y|\mathcal{V})$ a.s.
4. If X is independent to \mathcal{V} then $E(X|\mathcal{V}) = E(X)$ a.s.
5. If $X \leq Y$ a.s. $\Rightarrow E(X|\mathcal{V}) \leq E(Y|\mathcal{V})$ a.s.

Let $\bigcup_{i=1}^n A_i = \Omega$, $A_i \cap A_j = \emptyset$, $i \neq j$, $P(A_i) > 0$, for each $i = 1, \dots, n$. Let $\mathcal{F}(A_1, \dots, A_n)$ be a σ -algebra generated by A_1, \dots, A_n . Then

$$E(X|\mathcal{F}(A_1, \dots, A_n)) = \begin{cases} E(X|A_1), & \text{on } A_1 \\ \vdots & \vdots \\ E(X|A_n), & \text{on } A_n \end{cases}$$

$$E(X|\mathcal{F}(A_1, \dots, A_n)) = E(X|\mathcal{F}(Y)) = E(X|Y)$$

Theorem 3.1 (Jensen's Inequality). *Let $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function and let ξ be an integrable random variable on a probability space (Ω, \mathcal{F}, P) such that $\varphi(\xi)$ is also integrable. Then*

$$\varphi(E(\xi|\mathcal{G})) \leq E(\varphi(\xi)|\mathcal{G}) \quad \text{a.s.}$$

for any σ -field \mathcal{G} on Ω contained in \mathcal{F} .

3.1 Exercises

1. A prisoner is trapped in a cell containing three doors. The first door leads to a tunnel that returns him to his cell after two days of travel. The second leads to a tunnel that returns him to his cell after three days of travel. The third door leads immediately to freedom.

- (a) Assuming that the prisoner will always select doors 1, 2, and 3 with probabilities 0.5, 0.3, 0.2, what is the expected number of days until he reaches freedom?
- (b) Assuming that the prisoner is always equally likely to choose among those doors that he has not used, what is the expected number of days until he reaches freedom? (In this version, for instance, if the prisoner initially tries door 1, then when he returns to the cell, he will now select only from doors 2 and 3.)
2. Suppose that 2 batteries are randomly chosen without replacement from the following group of 12 batteries: 3 new, 4 used (working) and 5 defective. Let X denote the number of new batteries chosen and let Y denote the number of used batteries chosen. Find $E(X)$ and $E(X|Y)$.
3. Each day a system is gathering a certain amount of data and sums the money earned that day. Then a program classifies the previous day into a category. Each day can be classified as a Type i , $i \in \{1, \dots, k\}$. It is known that probability the day is classified as Type i is p_i and that

$$\sum_{i=1}^k p_i^2 = 0.5.$$

Also, the expected amount of money earned during the Type i day is $\ln\left(\frac{a}{e^{p_i}}\right)$ millions, $a > 1$. What is the expected amount of money earned on arbitrary chosen day?

4. Suppose that a couple after n year of marriage can have at most n children. Let X denote a random variable which represents the number of children after exactly 3 years of marriage,

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1/6 & 1/3 & 1/3 & 1/6 \end{pmatrix}.$$

If Y is a number of female children after exactly 3 years of marriage, find $E(X|\mathcal{F}(Y))$.

5. Milan and Uroš play a dice game as follows. Each of them throws a dice, independently of the other one. If the sum is 5, 6 or 7, Milan wins. Otherwise, Uroš wins.
- (a) Find a probability that Milan wins.
- (b) What is the probability that Milan's throw resulted in 3, if it is known that Milan won.
6. Suppose that Milan and Uroš play a series of games explained above. The overall winner is the first player to have won two more games than the other. Find the expected number of games played.
7. Students are waiting for the test results. Teacher told them that the results will be visible in exactly 2 hours with probability $\frac{1}{2}$, and in exactly 6 hours with probability $\frac{1}{3}$. Otherwise, they will be able to see the results during the class that is going to start in 20 hours, and it will last for two hours (in that case the waiting time will be uniformly distributed). Consider continuous random variable X that is independent of students waiting time Y . Calculate $E(Y|X + \text{sgn}X)$.

8. Let X be the randomly chosen number from the set $\{1, 2, 3, 4, 5, 6\}$. Y is chosen from the same set among numbers that are integer multiple of X . Find $E(X|\mathcal{F}(Y))$.
9. The game of chance is played as follows. The machine randomly and one by one displays digits 0 and 1 on the screen, with probabilities $1 - p$ and p , respectively. The game ends when the first zero appears. If your stake in one game is $a > 0$, the winning in that game equals the sum of all digits 0 and 1 shown on the screen. What is the expected amount of money you will win in one game? Find the value of p such that the game is fair (winnings are equal to losses).

4 Stochastic processes

A stochastic process $\{X(t), t \in I\}$ is a collection of random variables. That is, for each $t \in I$, $X(t)$ is a random variable. The index t is often interpreted as time and, as a result, we refer to $X(t)$ as the state of the process at time t .

The set I is called the index set of the process. When I is a countable set the stochastic process is said to be a discrete-time process. If I is an interval of the real line, the stochastic process is said to be a continuous-time process. The state space of a stochastic process is defined as the set of all possible values that the random variables $X(t)$ can assume. Thus, a stochastic process is a family of random variables that describes the evolution through time of some (physical) process.

Definition 4.1. Suppose that with each element ω of a sample space S of some random experiment E , we associate a function $X(t, \omega)$, where t belongs to $I \subset \mathbb{R}$. The set $\{X(t, \omega), t \in I\}$ is called a **stochastic (or random process)**.

The function $X(t, \omega)$ is a random variable for any particular value of t .

Definition 4.2. The distribution function of order k of the stochastic process $\{X(t), t \in I\}$ is the joint distribution function of the random vector $X(t_1), \dots, X(t_k)$

$$F_{t_1, \dots, t_k}(x_1, \dots, x_k) = P\{X(t_1) < x_1, \dots, X(t_k) < x_k\}.$$

Definition 4.3. The **mean** $E[X(t)]$ of an s.p. $\{X(t), t \in I\}$ at time t is denoted by $m_X(t)$. Moreover, the **autocorrelation function** and the **autocovariance function** of the process at the point t_1, t_2 are defined, respectively, by

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)],$$

$$K_X(t_1, t_2) = E[X(t_1)X(t_2)] - m_X(t_1)m_X(t_2).$$

The variance of the process at time t is

$$\text{Var}[X(t)] = K_X(t, t).$$

The autocovariance and autocorrelation functions indicate the rate of change of the sample functions of a stochastic process.

Definition 4.4. If the random variables $X(t_4) - X(t_3)$ and $X(t_2) - X(t_1)$ are independent $\forall t_1 < t_2 < t_3 < t_4$, we say that the stochastic process $\{X(t), t \in I\}$ is a **process with independent increments**.

Definition 4.5. If the random variables $X(t_2 + s) - X(t_1 + s)$ and $X(t_2) - X(t_1)$ have the same distribution function for all s , $\{X(t), t \in I\}$ is said to be a **process with stationary increments**.

Definition 4.6. We say that the stochastic process $\{X(t), t \in I\}$ is **stationary, or strict-sense stationary (SSS)**, if its distribution function of order n is invariant under any change of origin:

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_{t_1+s, \dots, t_n+s}(x_1, \dots, x_n)$$

for all s, n and t_1, \dots, t_n .

A stochastic process is stationary if the randomness does not vary with time.

Definition 4.7. We say that the stochastic process $\{X(t), t \in I\}$ is **wide-sense stationary (WSS)** if $m_X(t) = m$ and

$$K_X(t_1, t_2) = K_X(t_2 - t_1), \text{ for all } t_1, t_2 \in I.$$

A stochastic process is wide sense stationary if the expected value is constant with time and the autocovariance function depends only on the time difference between two random variables. Since the expected value is constant, then the statement that autocovariance function depends only on the time difference between two random is equivalent to statement that the autocorrelation function depends only on the time difference between two random

variables

Definition 4.8. We say that the random vector (X_1, \dots, X_n) has a **multi-normal distribution** if each random variable X_k can be expressed as a linear combination of independent random variables Z_1, \dots, Z_m , where $Z_j : \mathcal{N}(0, 1)$, for $j = 1, \dots, m$. That is, if

$$X_k = \mu_k + \sum_{j=1}^m c_{kj} Z_j, \quad \text{for } k = 1, \dots, n$$

where μ_k is a real constant, for all k .

Definition 4.9. A stochastic process $\{X(t), t \in I\}$ is said to be a **Gaussian process** if the random vector $(X(t_1), \dots, X(t_n))$ has a multi-normal distribution, for any n and for all t_1, \dots, t_n .

Proposition 4.1. *If a Gaussian process $\{X(t), t \in I\}$ is such that its mean $m_X(t)$ is a constant m_X and if its autocovariance function $K_X(t, t + s)$ depends only on s , then it is stationary (in the strict sense).*

4.1 Exercises

1. Let W be an exponential random variable with probability density function

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

- (a) Find the cumulative distribution function F_{X_t} of the time delayed ramp process $X_t = t - W$.
- (b) Find the autocovariance function of a process X_t .

2. Consider two independent random variable, X and Y , where $X : \mathcal{E}(1)$ and

$$\varphi_{(X,Y)}(x, y) = \begin{cases} \frac{1}{2}e^{-x}, & x \geq 0, y \in (1, 3), \\ 0, & \text{otherwise.} \end{cases}$$

Find autocovariance function of a stochastic process $N_t = X(tX + Y)$.

3. Consider the following joint probability density function

$$\varphi_{X,Y}(x,y) = \begin{cases} Cye^{-x}, & x,y \in (0,1), \\ 0, & x,y \notin (0,1). \end{cases}$$

- (a) Determine the value of C and then determine the marginal probability density functions for X and Y .
 - (b) Determine the autocovariance function of stochastic process $X_t = X + tY$.
4. Consider continuous and independent random variables X and Y such that $E(X^2) < \infty$ and $E(Y) = 0$. Find the mean of the stochastic process

$$Z_t = t(X^2Y + 1).$$

5 Markov Chains

We consider a stochastic process $\{X_n, n = 0, 1, 2, \dots\}$ that takes on a finite or countable number of possible values. Unless otherwise mentioned, this set of possible values of the process will be denoted by the set of nonnegative integers $\{0, 1, 2, \dots\}$.

If $X_n = x_i$, then the process is said to be in state i at time n . We suppose that whenever the process is in state i , there is a fixed probability p_{ij} that it will next be in state j . That is, we suppose that

$$P\{X_{n+1} = x_j | X_n = x_i, X_{n-1} = x_{i_{n-1}}, \dots, X_1 = x_{i_1}, X_0 = x_{i_0}\} = p_{ij} \quad (1)$$

for all states $i_0, i_1, \dots, i_{n-1}, i, j$ and all $n \geq 0$. Such a stochastic process is known as a Markov chain.

Equation (1) may be interpreted as stating that, for a Markov chain, the conditional distribution of any future state X_{n+1} given the past states X_0, X_1, \dots, X_{n-1} and the present state X_n , is independent of the past states and depends only on the present state.

The value p_{ij} (*one-step transition probability*) represents the probability that the process will, when in state i , next make a transition into state j . Since probabilities are nonnegative and since the process must make a transition into some state, we have that

$$p_{ij} \geq 0, \quad i, j \geq 0, \quad \sum_{j=0}^{\infty} p_{ij} = 1, \quad i = 0, 1, \dots$$

One-step transition probability $p_{ij}^{n, n+1}$ is defined by

$$p_{ij}^{n, n+1} = P\{X_{n+1} = x_j | X_n = x_i\}.$$

If $p_{ij}^{n, n+1}$ does not depend of n Markov chain is **homogeneous**.

The n -step transitional probability $p_{ij}(n)$ is the probability that a process in state i will be in state j after n additional transitions

$$p_{ij}(n) = P\{X_{m+n} = x_j | X_m = x_i\}.$$

The matrix of one-step transition probabilities p_{ij} is denoted by P

$$P = [p_{ij}]_{i,j}$$

The matrix of n -step transitional probabilities is given by

$$P_n = [p_{ij}(n)]_{i,j}.$$

$$p_i(n) = P\{X_n = x_i\}$$

$$\mathbf{p}(0) = [p_1(0), p_2(0), \dots, p_N(0)]$$

For a fixed n we have

$$X_n : \begin{pmatrix} x_1 & x_2 & \dots & x_N \\ p_1(n) & p_2(n) & \dots & p_N(n) \end{pmatrix}$$

The Chapman–Kolmogorov equations provide a method for computing these n -step transition probabilities. These equations are

$$p_{ij}(n+m) = \sum_{k=0}^n p_{ik}(n)p_{kj}(m), \quad P(n+m) = P(n)P(m).$$

$$\mathbf{p}(k) = \mathbf{p}(0) \cdot P^k.$$

For a finite Markov chain with transition matrix P , the n -th step transition matrix is

$$P(n) = P^n.$$

If the state $p(k)$ does not depend of k the chain is *stationary*.

If there exists $n_0 \in \mathbb{N}$ so that $P_{n_0} = P^{n_0}$ has all positive elements the chain is *ergodic*.

State x_j is said to be *accessible* (**dostižno**) from state x_i if $p_{ij}(n_0) > 0$, for some n_0 .

The state x_i is said to be *recurrent* (**apsorbujuće**) if $p_{ii} = 1$ (starting in state x_i , the process will ever reenter state x_i .)

The state x_i is said to be *transient* (**povratno**) if $p_{ii} < 1$.

For a finite Markov chain with initial state probability vector $\mathbf{p}(0)$, the **limiting state probabilities**, when they exist, are defined to be the vector

$$\mathbf{p}^* = \lim_{n \rightarrow \infty} \mathbf{p}(n).$$

We have

$$p_j^* = \lim_{n \rightarrow \infty} p_{ij}(n)$$

The j -th element, p_j^* is the probability the system will be in state j in the distant future. For an ergodic Markov chain $\lim_{n \rightarrow \infty} p_{ij}(n)$ exists and is independent of i . Then we have

$$\mathbf{p}^* = \mathbf{p}^* P$$

or

$$[p_1^* \ p_2^* \ \dots \ p_m^*] = [p_1^* \ p_2^* \ \dots \ p_m^*] P$$

$$\sum_{j=1}^m p_j^* = 1$$

Example 5.1. [A Communications System] Consider a communications system which transmits the digits 0 and 1. Each digit transmitted must pass through several stages, at each of which there is a probability p that the digit entered will be unchanged when it leaves. Letting X_n denote the digit entering the n -th stage, then $\{X_n, n = 0, 1, \dots\}$ is a two-state Markov chain having a transition probability matrix

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

Example 5.2. On any given day Gary is either cheerful (C), so-so (S), or glum (G). If he is cheerful today, then he will be C, S, or G tomorrow with respective probabilities 0.5, 0.4, 0.1. If he is feeling so-so today, then he will be C, S, or G tomorrow with probabilities 0.3, 0.4, 0.3. If he is glum today, then he will be C, S, or G tomorrow with probabilities 0.2, 0.3,

0.5. Letting X_n denote Gary's mood on the n -th day, then $\{X_n, n \geq 0\}$ is a three-state Markov chain (state 0 = C , state 1 = S , state 2 = G) with transition probability matrix

$$P = \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}.$$

5.1 Exercises

1. Ana is doing data analysis and she is receiving data through a communication system. But, she is aware that digit entered might be changed in the process and based of the prediction analysis she knows which digit is unchanged/changed. Previously, she received a transition probability matrix P for a three state Markov chain

$$P = \begin{bmatrix} a & b & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}.$$

Ana knows that, in the long run, 25% of the time system spends in the state 2. Find constants a and b , if possible.

2. The diffusion of electrons and holes across a potential barrier in an electronic devise is modeled as follows. There are m black balls (electrons) in urn A and m white balls (holes) in urn B . We perform independent trials, in each of which a ball is selected at random from each urn and the selected ball from urn A is placed in urn B , while that from urn B is placed in A . Consider the Markov chain representing the number of black balls in urn A immediately after the n -th trial.
 - (a) Describe the one-step transition probabilities of the process.
 - (b) Suppose $m = 2$. Compute the long-run fraction of time when urn A does not contain a black ball.
3. In working with a particular gene for fruit flies, geneticists classify an individual fruit fly as dominant, hybrid or recessive. In running an experiment, an individual fruit fly is crossed with a hybrid, then the offspring is crossed with a hybrid and so forth. The offspring in each generation are recorded as dominant, hybrid or recessive. The probabilities the offspring are dominant, hybrid or recessive depends only on the type of fruit fly the hybrid is crossed with rather than the genetic makeup of previous generations. The offspring of a dominant individual crossed with a hybrid are dominant 50% of the time and hybrid the other 50%. The offspring of a hybrid crossed with a hybrid are dominant 25%, hybrid 50% and recessive 25%, while the offspring of a recessive crossed with a hybrid are hybrid 50% and recessive 50%.
 - (a) Find the transition matrix for this problem.
 - (b) What is the probability the third generation offspring is dominant given the first generation offspring is recessive?
 - (c) If the population of fruit flies initially is 20% dominant, 50% hybrid and 30% recessive, what percentage of the population is dominant after 3 generations?

4. The social status of the n -th generation of some family is given by the Markov chain $X_n, n \in \mathbb{N}_0$. Some family can belong to one of three social classes: 1 - lower, 2 - middle and 3 - upper. The transition probability matrix which describes the change in classes is given by

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}.$$

- (a) Find a probability that a person whose parents are in the middle class ends up in the upper class and that the children of that person are in the lower class.
- (b) Determine a probability that the children of the person whose parents are in the middle class end up in the lower class?
5. Two girls and two boys are playing with the ball. Each boy will toss a ball to the other boy with probability $1/2$, and to each of the girls with probability $1/4$. Each girl will toss the ball to each boy with probability $1/2$ and won't toss the ball to the other girl. If the game last for a long time, how often will each of them receive the ball?
6. Consider a communications system which transmits the digits 0 and 1. Each digit transmitted must pass through several stages, at each of which there is a probability $p \in (0, 1)$ that the digit entered will be unchanged when it leaves. Let X_n denote the digit entering the n -th stage.
- (a) Find the state transition probability matrix.
- (b) Find initial state vector such that the Markov chain $\{X_n, n \geq 0\}$ is stationary.
- (c) Find $P\{X_{n+2} = 1|X_n = 1\}$.
7. Two gamblers play the following game. A fair coin is flipped. If the outcome is heads, player A pays player B 1 dollar, and if the outcome is tails player B pays player A 1 dollar. The game is continued until one of the players goes broke. Suppose that initially player A has 1 dollar and player B has 2 dollars, so a total of 3 dollars is up for grabs. Let X_n denote the number of dollars held by player A after n trials.
- (a) Find the expected number of flips until one of the players goes broke.
- (b) Show that X_n is a Markov chain and find the one-step transition probabilities.
- (c) What is the probability that a game has stopped after only one flip of the coin?
- (d) Find the two-step transition probabilities $P[X_{n+2} = i|X_n = i], i = 1, 2$.
8. Peter takes the course Stochastic Processes this semester on Tuesday, Thursday and Friday. The classes start at 10 am. Peter is used to work until late in the night and consequently, he sometimes misses the class. His attendance behavior is such that he attends class depending only on whether or not he went to the latest class. If he attended class one day, then he will go to class next time it meets with probability $1/2$. If he did not go to one class, then he will go to the next class with probability $3/4$.

- (a) Describe the Markov chain that models Peter's attendance. What is the probability that he will attend class on Thursday if he went to class on Friday?
- (b) Find the probability that Peter attends a class?
- (c) Suppose the course has 30 classes altogether. Give an estimate of the number of classes attended by Peter and explain it.
9. A game is played as follows. The integers $N \geq 2$ and $s \leq N - 1$ are randomly chosen. There are two types of balls in the box: white and black. Initially, the box is filled with N balls. The number of white balls is s . Trials are performed as follows. In each trial one ball is selected. If that ball is white, one returns the selected ball in the box. If the ball is black, one returns the selected ball and adds white ball in the box. Consider the Markov chain representing the number of white balls in the box after the n -th trial.
- (a) Find the one and two step transition probabilities of the Markov chain.
- (b) The game ends when the probability that a number of white balls in the box after d trials increases by d is not greater than $1/30$. What is the minimum number of trials necessary for that to happen if $N = 2$, $s = 1$?
10. In some countries exists a limit on number of children the couple can have in marriage. Suppose that a limit is 4 i.e. a couple cannot have more than 4 children in marriage. Let X_n denote the number of children in marriage after n years. Suppose that the number of children cannot decrease. It is known that the number of children after one year will stay the same with probability 0.5, except in the case when there is 4 children. If in the end of one year the number of children is k , $k = 0, 1, 2$, then at the end of the next year the number of children will be $k + 1$ with probability $\frac{1}{k+2}$. If the couple has a child, the number of children will increase by 2 after 2 year with probability $1/4$.
- (a) Find a one-step transition probability matrix of a Markov chain $\{X_n\}$.
- (b) Is Markov chain $\{X_n\}$ ergodic?
11. Peter decided to lose some weight and sometimes he will go to work by foot. Suppose that day is classified as rainy with probability p , independently of the previous day. Otherwise, the day is classified as a sunny. If the day is rainy, Peter will go to work by car, independently of the previous day. If the day is sunny, Peter decides to base his decision on the previous day. If on the previous day he went to work by car, and the day is sunny he will go to work by foot. If he went to work by foot, and the day is sunny it is equally likely that he will choose to go to work by car and by foot. What proportion of days will Peter go to work by car?
12. A Markov particle moves on points 0,1 and 2 arranged in a circle in the clockwise direction. A step in the clockwise direction occurs with probability p , $0 < p < 1$, and a step in the counter-clockwise direction occurs with probability $1 - p$. If Markov particle is initially positioned on point 0, what proportion of time the particle will spend on point 0? Explain.

6 Poisson process

6.1 Counting process

Definition 6.1. A stochastic process $\{X_t, t \geq 0\}$ is said to be a **counting process** if X_t represents the total number of events that have occurred up to time t . A counting process must satisfy

1. $X_t \geq 0$
2. X_t is integer valued
3. If $s < t$, $X_s \leq X_t$
4. For $s < t$, $X_t - X_s$ equals the number of events that have occurred in the interval $(s, t]$.

λ is called **the rate of the process**.

A counting process is said to possess independent increments if the numbers of events that occur in disjoint time intervals are independent. A counting process is said to possess stationary increments if the distribution of the number of events that occur in any interval of time depends only on the length of the time interval. In other words, the process has stationary increments if the number of events in the interval $(s, s + t)$ has the same distribution for all s .

6.2 Definition of the Poisson process

Definition 6.2. The counting process $\{X_t, t \geq 0\}$ is said to be a Poisson process with rate $\lambda, \lambda > 0$ if

1. $X_0 = 0$.
2. The process has independent increments.
3. The number of events in any interval of length t is Poisson distributed with mean λt . That is, for all $s, t \geq 0$

$$P\{X_{s+t} - X_s = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \dots$$

From condition 3. follows that a Poisson process has stationary increments. Also, for $s = 0$, $X_t - X_0 = X_t$ and

$$P\{X_t = n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

So, $X_t : \mathcal{P}(\lambda t)$ and $E(X_t) = \lambda t$.

But, it is not at all clear how we would determine that condition 3. from the Definition 6.2 is satisfied, and for this reason an equivalent definition of a Poisson process would be useful.

Definition 6.3. The counting process $\{X_t, t \geq 0\}$ is said to be a Poisson process with rate $\lambda, \lambda > 0$ if

1. $X_0 = 0$
2. The process has stationary and independent increments
3. $P\{X_h = 1\} = \lambda h + o(h)$
4. $P\{X_h \geq 2\} = o(h)$

Theorem 6.1. *Definitions 6.2 and 6.3 are equivalent.*

Remark 6.1. *The explicit assumption that the process has stationary increments can be eliminated from Definition 6.3 provided that we change assumptions 3. and 4. to require that for any t the probability of one event in the interval $(t, t + h)$ is $\lambda h + o(h)$ and the probability of two or more events in that interval is $o(h)$.*

Definition 6.4. The counting process $\{X_t, t \geq 0\}$ is said to be a Poisson process with rate $\lambda, \lambda > 0$ if

1. $X_0 = 0$
2. The process has independent increments
3. $P\{X_{t+h} - X_t = 1\} = \lambda h + o(h), h \rightarrow 0$
4. $P\{X_{t+h} - X_t \geq 2\} = o(h), h \rightarrow 0$

6.3 Interarrival and waiting time distributions

Let $\{X_t, t \geq 0\}$ be a Poisson process with a rate λ and

- T_1 – denote the time of the first event
- ...
- $T_n, n \geq 1$ – denote the time between the $(n - 1)$ st and n th event.

The sequence $\{T_n, n \geq 1\}$ is called the **sequence of interarrival times** (*niz vremena zadržavanja u datom stanju*).

Distribution of the T_n ?

The event $\{T_1 > t\}$ takes place if and only if no event of the Poisson process occur in the interval $[0, t]$ and thus

$$P\{T_1 > t\} = P\{X_t = 0\} = e^{-\lambda t}$$

$$F_{T_1}(t) = P\{T_1 < t\} = 1 - P\{T_1 \geq t\} = 1 - P\{T_1 = t\} - P\{T_1 > t\}$$

$$F_{T_1}(t) = 1 - P\{X_t = 0\} = 1 - \frac{(\lambda t)^0}{0!} e^{-\lambda t} = 1 - e^{-\lambda t}, t > 0$$

$$T_1 : \mathcal{E}(\lambda)$$

$$\begin{aligned} P\{T_2 > t | T_1 = s\} &= P\{0 \text{ events in } (s, s + t] | T_1 = s\} \\ &= P\{0 \text{ events in } (s, s + t]\} = e^{-\lambda t} \quad (\text{independent, stationary increments}) \end{aligned}$$

$$T_2 : \mathcal{E}(\lambda)$$

Theorem 6.2. $T_n, n = 1, 2, \dots$ are independent identically distributed exponential random variables having mean $1/\lambda$.

$S_n = \sum_{k=1}^n T_k$ the arrival time of the n th event or **waiting time** until the n th event

$$P\{X_t \geq n\} = P\{S_n \leq t\}$$

$$\varphi_{S_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, t > 0$$

6.4 Properties of Poisson process

$\{X_t, t \geq 0\}$ is Poisson process. Then

$$X_t - X_s : \mathcal{P}(\lambda(t-s)), \lambda > 0$$

If $s = 0$ then

$$X_t : \mathcal{P}(\lambda t) \Rightarrow E(X_t) = D(X_t) = \lambda t.$$

$$K_X(t, s) = \lambda \min\{t, s\}, t, s \geq 0$$

Poisson process is Markov Process.

Consider a Poisson process $\{N_t, t \geq 0\}$ having rate λ , and suppose that each time an event occurs it is classified as either a type I or a type II event. Suppose further that each event is classified as a type I event with probability p or a type II event with probability $1 - p$, independently of all other events.

For example, suppose that customers arrive at a store in accordance with a Poisson process having rate λ and suppose that each arrival is male with probability $1/2$ and female with probability $1/2$. Then a type I event would correspond to a male arrival and a type II event to a female arrival.

Let N_{1t} and N_{2t} denote respectively the number of type I and type II events occurring in $[0, t]$. Note that $N_t = N_{1t} + N_{2t}$.

Proposition 6.1. $\{N_{1t}, t \geq 0\}$ and $\{N_{2t}, t \geq 0\}$ are both Poisson processes having respective rates λp and $\lambda(1 - p)$. Furthermore, the two processes are independent.

We know that if each event of a Poisson process is independently classified as a type I event with probability p and as a type II event with probability $1 - p$ then the counting processes of type I and type II events are independent Poisson processes with respective rates λp and $\lambda(1 - p)$. Suppose now, however, that there are k possible types of events and that the probability that an event is classified as a type i event, $i = 1, \dots, k$, depends on the time the event occurs. Specifically, suppose that if an event occurs at time y then it will be classified as a type i event, independently of anything that has previously occurred, with probability $P_i(y)$, $i = 1, \dots, k$ where $\sum_{i=1}^k P_i(y) = 1$. Now we have

Proposition 6.2. If $X_i(t)$, $i = 1, \dots, k$, represents the number of type i events occurring by time t then $X_i(t)$, $i = 1, \dots, k$, are independent Poisson random variables having means

$$E[X_i(t)] = \lambda \int_0^t P_i(s) ds.$$

6.5 Nonhomogeneous Poisson process

Definition 6.5. The counting process $\{X_t, t \geq 0\}$ is said to be a **nonstationary or nonhomogeneous Poisson process** with intensity function $\lambda(t), t \geq 0$ if

- (i) $X_0 = 0$
- (ii) $\{X_t, t \geq 0\}$ has independent increments
- (iii) $P\{X_{t+h} - X_t = 1\} = \lambda(t)h + o(h), h \rightarrow 0$
- (iv) $P\{X_{t+h} - X_t \geq 2\} = o(h), h \rightarrow 0$

If we let

$$m(t) = \int_0^t \lambda(s) ds,$$

then it can be shown

$$P\{X_{s+t} - X_t = n\} = e^{-(m(s+t)-m(s))} \frac{(m(s+t) - m(s))^n}{n!}.$$

That is, $X_{s+t} - X_t$ is Poisson distributed with mean $m(s+t) - m(t)$.

6.6 Compound Poisson Processes

Definition 6.6. A Stochastic process $\{X_t, t \geq 0\}$ is said to be a **compound Poisson process** if it can be represented, for $t \geq 0$, by

$$X_t = \sum_{i=1}^{N_t} Y_i,$$

where $\{N_t, t \geq 0\}$ is a Poisson process, and $\{Y_i, i = 1, 2, \dots\}$ is a family of independent and identically distributed random variables that is independent of the process $\{N_t, t \geq 0\}$. Thus, if $\{X_t, t \geq 0\}$ is a compound Poisson process then X_t is a compound Poisson random variable.

We have

$$E(X_t) = E(N_t)E(Y_1), \quad D(X_t) = E(N_t)E(Y_1^2).$$

6.7 Exercises

1. A geiger counter is a device to count the radioactive particles emitted by a source. Suppose the particles arrive at the counter according to a Poisson process with rate $\lambda = 1000$ per second. The counter fails to count a particle with probability 0.1, independent of everything else. Suppose the counter registers four particles in 0.01 seconds. What is the probability that at least six particles have actually arrived at the counter during this time period?

2. Events occur according to a nonhomogeneous Poisson process whose mean value function is given by

$$m(t) = t^2 + 2t, \quad t \geq 0.$$

Find the intensity function of this process. What is the probability that n events occur between times $t = 4$ and $t = 5$?

3. In good years, storms occur according to a Poisson process with rate 3 per unit time, while in other years they occur according to a Poisson process with rate 5 per unit time. Suppose next year will be a good year with probability 0.3. Let N_t denote the number of storms during the first t time units of next year.

(a) Find the probability that in the first t time units of next year have occurred n storms.

(b) Is $\{N_t, t \geq 0\}$ a Poisson process?

(c) Does $\{N_t, t \geq 0\}$ have stationary increments? Explain.

(d) If next year starts off with three storms by time $t = 1$, what is the probability it is a good year?

4. The number of trains from Novi Sad to Belgrade which are late more than one hour is described by a Poisson process with a rate one per month, while number of trains from Novi Sad to Subotica which are late more than one hour is an independent Poisson process with a rate 2 per month.

(a) What is the probability that at least 3 trains from Novi Sad to Belgrade are late more than one hour during one month?

(b) What is the probability that 3 trains from Novi Sad to Belgrade and 2 trains from Novi Sad to Subotica are late more than 1 hour during 1 month?

5. All the employees in one company use the same printer. Suppose that the times between two consecutive print requests are independent, identically distributed exponential random variables with parameter 10 per hour. The printer needs exactly 6 seconds to print one paper.

(a) What is the probability that exactly 20 print requests will arrive between 8 : 30 i 10 : 30h?

(b) What is the probability that a print request will arrive while the previous document containing 6 papers is not yet printed?

6. Customers arrive at the automatic teller machine in accordance with a Poisson process with rate 10 per hour. The amount of money withdrawn on each transaction is a random variable with mean \$40 and standard deviation \$50. A negative withdrawal means that money was deposited. The machine is in use for 16 hours daily. Approximate the probability that the total daily withdrawal is less than \$7000.

7. The number of babies born in Serbia during t days is a Poisson process. On an average, a baby is born after every 60 seconds. If it is known that in one of 10 cases baby is not discharged from the hospital after 5 days, find the probability that in two days period, the number of babies which are not discharged from the hospital after 5 days is at most 2.

8. Satellites are launched into space at times distributed according to a Poisson process with rate 2 per year. A time that each satellite independently spends in space before falling to the ground is exponentially distributed with mean 40 years. Find the expected number of satellites on the ground after 20 years.
9. (a) Nikola starts each month with 80 000 dinars on the bank account. On the first day of the month he pays the bills (amount is 10 000 dinars), on the second and the fifteenth day of the month he goes to the supermarket. Amount of money that he spends on those days is a random variable

$$X : \begin{pmatrix} 5000 & 7000 & 8000 & 10000 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}.$$

Amount of money spent on any other day is uniformly distributed in the interval $(0, 2000)$. Denote by X_n the total amount of money available on the n -th day of the month. Suppose that a month has 30 days. Find the expected amount of money available on the n th day of the month if an amount available on $(n-1)$ st day is known and find the expected amount of money spent on arbitrary chosen day.

- (b) Suppose that on average Nikola spends more than 1000 dinars each two days. What is the probability that Nikola spends more than 1000 dinars less than 2 times in 3 days?
10. Consider babies born in the “normal” range of 37–43 weeks gestational age. Extensive data support the assumption that for such babies born in the United States, birth weight is normally distributed with mean 3432 g and standard deviation 482 g . Also, assume that such babies are born at a Poisson rate of 20 babies per minute.
- (a) What is the probability that the birth weight of a randomly selected baby of this type is between 3000 and 4000 g ?
- (b) Calculate the probability that in the next two minutes will be born 40 babies of this type with birth weight between 3000 and 4000 g . Find an average time between two consecutive “normal” births of babies.

11. Consider the emission of γ -photons by a radioactive source. This can be modeled to a close approximation by a nonhomogeneous Poisson process $X(t)$ with intensity function $\lambda(t)$ given by $\lambda(t) = \alpha e^{-\beta t}$, $t \geq 0$, $\alpha, \beta > 0$, where α is a parameter depending on the amount of the radioactive material and the β^{-1} is the mean life of the source. Calculate

$$\frac{\partial}{\partial s} P\{X(s+t) - X(t) = 0\} - \beta P\{X(s+t) - X(t) = 1\}.$$

12. Consider a mechanical device in which “shocks” occur according to a Poisson process with rate 0.1 per hour. The device fails when a total of K shocks occurs. Find the expected lifetime T of the device. Find the probabilities that K and $K-1$ shocks have occurred after $E(T)$ hours and compare those two values.

7 Brownian Motion

Brownian motion is named after English botanist Robert Brown who discovered it. That is the motion exhibited by a small particle which is totally immersed in a liquid or gas.

Consider the symmetric random walk, which in each time unit is equally likely to take a unit step either to the left or to the right. That is, it is a Markov chain with $p_{i,i+1} = p_{i,i-1} = 1/2$, $i = 0, \pm 1, \pm 2, \dots$. Now suppose that we speed up this process by taking smaller and smaller steps in smaller and smaller time intervals. If we now go to the limit in the right manner what we obtain is Brownian motion.

Suppose that each Δt time unit we take a step of size Δx either to the left or the right with equal probabilities. Let $X(t)$ denote the position at time $t = n\Delta t$. Define

$$S_n := \sum_{i=1}^n X_i$$

as the number of steps to the right until $t = n\Delta t$, where

$$X_i : \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}, \quad E(X_i) = \frac{1}{2}, \quad D(X_i) = \frac{1}{4}, \quad i = 1, \dots, n.$$

Now,

$$X(t) = S_n \cdot \Delta x + (n - S_n) \cdot (-\Delta x) = (2S_n - n)\Delta x,$$

$$D(X(t)) = D((2S_n - n)\Delta x) = (\Delta x)^2 D(2S_n - n) = (\Delta x)^2 \cdot n = \frac{(\Delta x)^2}{\Delta t} \cdot t.$$

Denote,

$$\frac{(\Delta x)^2}{\Delta t} = d > 0.$$

Since, $E(S_n) = nE(X_1) = \frac{n}{2}$ and $D(S_n) = nD(X_1) = \frac{n}{4}$, we have

$$X(t) = \left(\frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \right) \sqrt{td}.$$

From De Moivre Laplace theorem follows

$$X(t) : \mathcal{N}(0, td).$$

Without loss of generality take $d = 1$.

Definition 7.1. A stochastic process $\{W_t, t \geq 0\}$ is said to be (*standard*) *Brownian motion* (*Wiener process*) process if

- (i) $W_0 = 0$
- (ii) $\{W_t, t \geq 0\}$ has stationary and independent increments i.e. for each $0 < t_1 < t_2 < \dots < t_{n-1} < t_n < \dots$, $W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$ are independent random variables.
- (iii) $W_t - W_s : \mathcal{N}(0, t - s)$, $t > s$. Specially, if we take $s = 0$ we have

$$W_t - W_0 = W_t : \mathcal{N}(0, t), \quad E[W_t] = 0, \quad D[W_t] = E[W_t^2] = t.$$

Since, $W_t : \mathcal{N}(0, t)$ we have

$$P\{a < W_t < b\} = \int_a^b \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx.$$

Put,

$$g(x, t|y) := \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}.$$

It holds $g(x - y, t|0) = g(x, t|y)$. We claim

$$\begin{aligned} & P\{a_1 < W_{t_1} < b_1, a_2 < W_{t_2} < b_2, \dots, a_n < W_{t_n} < b_n\} \\ &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_n}^{b_n} g(x_1, t_1|0)g(x_2, t_2 - t_1|x_1)g(x_3, t_3 - t_2|x_2) \dots g(x_n, t_n - t_{n-1}|x_{n-1})dx_1 \dots dx_n. \end{aligned}$$

The above formula follows from the theorem:

Theorem 7.1. *Let W_t be Brownian motion process. Then for each $n \in \mathbb{N}$ and for each $0 = t_0 < t_1 < \dots < t_n$ and for all functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$*

$$\begin{aligned} & E\left(f(W_{t_1}, W_{t_2}, \dots, W_{t_n})\right) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n)g(x_1, t_1|0)g(x_2, t_2 - t_1|x_1) \dots g(x_n, t_n - t_{n-1}|x_{n-1})dx_1 \dots dx_n. \end{aligned}$$

Theorem 7.2. *Let W_t be Brownian motion process. Then*

$$K_W(t, s) = \min\{t, s\}.$$

W_t is Markov process.

We say that $\{W_t, t \geq 0\}$ is a *Brownian motion process with drift coefficient μ and variance parameter σ^2* if

- (i) $W_0 = 0$
- (ii) $\{W_t, t \geq 0\}$ has stationary and independent increments
- (iii) W_t is normally distributed with mean μt and variance $t\sigma^2$.

An equivalent definition is to let $\{X_t, t \geq 0\}$ be standard Brownian motion and then define

$$W_t = \sigma X_t + \mu t.$$

7.1 Exercises

1. Show that for any $T > 0$, $V(t) = W(t + T) - W(T)$ is a standard Brownian motion if $W(t)$ is.
2. Let $(W_t)_{t \geq 0}$ be a standard Brownian motion. Find $E(W_4^2|W_2)$.
3. Consider a standard Brownian motion W_t and define

$$X_t := W_t - tW_1, \quad 0 \leq t \leq 1.$$

- (a) Find a autocovariance function $K_X(t, s)$ for a given process X_t .
 (b) Does the process $\{X_t\}_{0 \leq t \leq 1}$ has an independent increments?

4. Find

$$E[\alpha W_t + W_t W_s + \beta W_r^2 W_s | \mathcal{W}_t], \quad 0 < s \leq t < r,$$

if $W_t, t \geq 0$ is standard Brownian motion and \mathcal{W}_t is history of Brownian motion until time t (including time t).

5. Consider the standard Brownian motion $W(t), t \geq 0$.

- (a) Prove that for $s \geq t \geq 0$

$$E(W(s)W^n(t)) = \begin{cases} 0, & \text{if } n = 2k, k = 0, 1, 2, \dots \\ n!! t^{\frac{n+1}{2}}, & \text{if } n = 2k - 1, k = 1, 2, \dots \end{cases}.$$

- (b) Show that $E(|W(t) - W(s)|^2) = |t - s|, t, s \geq 0$.

Hint: Calculate $E(W^{n+1}(t))$.

6. (a) Find the function $f = f(s, t, x)$ such that the following holds

$$P\{W_t < 5\sqrt{t}\}E[W_t^2 W_s^2 | \mathcal{W}_s] + W_s^2 E[W_{2s} | \mathcal{W}_s] = W_s^2 f(s, t, W_s), \quad 0 < s \leq t,$$

if W_t is standard Brownian motion and \mathcal{W}_t is history of Brownian motion until time t .

- (b) Determine $E[f(s, t, W_s)]$.

7. Consider standard Brownian motion W_t . Discuss the value of

$$E(W_t^3 + 2W_s^2 + W_\alpha | \mathcal{W}_\alpha), \quad t > s \geq 0$$

with respect to a parameter $\alpha > 0$. \mathcal{W}_α is history of Brownian motion until time α (including α).

8 Discrete Martingales

As the time increases, so does our knowledge about what happened in the past.

Definition 8.1. A sequence of σ -fields $\mathcal{F}_1, \mathcal{F}_2, \dots$ on Ω such that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \dots \subset \mathcal{F}$ is called a **filtration**.

Here \mathcal{F}_n represents our knowledge at time n . It contains all events A such that at time n it is possible to decide whether A has occurred or not. As n increases, there will be more such events A , i.e. the family \mathcal{F}_n representing our knowledge will become larger.

Suppose that ξ_1, ξ_2, \dots is a sequence of random variables and $\mathcal{F}_1, \mathcal{F}_2, \dots$ is a filtration. They may have nothing in common. However, in practice the filtration will usually contain the knowledge accumulated by observing the outcomes of the random sequence. The condition in the definition below means that \mathcal{F}_n contains everything that can be learned from the values of ξ_1, ξ_2, \dots .

Definition 8.2. We say that a sequence of random variables ξ_1, ξ_2, \dots is **adapted** to a filtration $\mathcal{F}_1, \mathcal{F}_2, \dots$ if ξ_n is \mathcal{F}_n -measurable for each $n = 1, 2, \dots$.

Example 8.1. If $\mathcal{F}_n = \mathcal{F}(\xi_1, \xi_2, \dots, \xi_n)$ is σ -field generated by $\xi_1, \xi_2, \dots, \xi_n$ then ξ_1, ξ_2, \dots is adapted to $\mathcal{F}_1, \mathcal{F}_2, \dots$.

The concept of a martingale has its origin in gambling, namely, it describes a fair game of chance. Some aspects of gambling are inherent in the mathematics of finance, in particular, the theory of financial derivatives such as options. In fact, martingales reach well beyond game theory and appear in various areas of modern probability and stochastic analysis, notably, in diffusion theory.

Definition 8.3. A sequence ξ_1, ξ_2, \dots of random variables is called a **martingale** with respect to a filtration $\mathcal{F}_1, \mathcal{F}_2, \dots$ if

1. ξ_n is integrable for each $n = 1, 2, \dots$,
2. ξ_1, ξ_n, \dots is adapted to $\mathcal{F}_1, \mathcal{F}_2, \dots$,
3. $E(\xi_{n+1}|\mathcal{F}_n) = \xi_n$ a.s. for each $n = 1, 2, \dots$.

Definition 8.4. We say that ξ_1, ξ_2, \dots is a **supermartingale** (resp. **submartingale**) with respect to a filtration $\mathcal{F}_1, \mathcal{F}_2, \dots$ if

1. ξ_n is integrable for each $n = 1, 2, \dots$;
2. ξ_1, ξ_n, \dots is adapted to $\mathcal{F}_1, \mathcal{F}_2, \dots$;
3. $E(\xi_{n+1}|\mathcal{F}_n) \leq \xi_n$ a.s. for each $n = 1, 2, \dots$ (resp. $E(\xi_{n+1}|\mathcal{F}_n) \geq \xi_n$ a.s. for each $n = 1, 2, \dots$)

8.1 Martingales and Brownian motion W_t

Let $\{W_t, t \geq 0\}$ be a Brownian motion process. Denote by

$$\mathcal{W}_s = \mathcal{F}(W_t, 0 \leq t \leq s)$$

the history of Brownian motion until time s .

Theorem 8.1. W_t is martingale with respect to \mathcal{W}_s .

Proof. 1. W_t is integrable for each t , $E[W_t] = 0 < \infty$.

2. follows from construction of \mathcal{W}_s that W_t is adapted to the filtration \mathcal{W}_t .

3. We want to show that $E[W_t|\mathcal{W}_s] = W_s$. We have

$$\begin{aligned} E[W_t|\mathcal{W}_s] &= E[W_t - W_s + W_s|\mathcal{W}_s] \\ &= E[W_t - W_s|\mathcal{W}_s] + E[W_s|\mathcal{W}_s] \\ &= E[W_t - W_s] + W_s = 0 + W_s = W_s. \end{aligned}$$

It was used above that $W_t - W_s : \mathcal{N}(0, t - s)$, W_s is measurable with respect to \mathcal{W}_s and $W_t - W_s$ is independent of \mathcal{W}_s . □

Theorem 8.2. $W_t^2 - t$ is martingale with respect to the history of Brownian motion \mathcal{W}_s .

Proof. 1. $E[W_t^2 - t] = t - t = 0 < \infty$

2. follows from construction of \mathcal{W}_s .

3. We want to show that $E[W_t^2 - t|\mathcal{W}_s] = W_s^2 - s$. We have

$$\begin{aligned} E[W_t^2|\mathcal{W}_s] &= E[(W_t - W_s)^2 + 2W_tW_s - W_s^2|\mathcal{W}_s] \\ &= E[(W_t - W_s)^2|\mathcal{W}_s] + E[2W_tW_s|\mathcal{W}_s] - E[W_s^2|\mathcal{W}_s] \\ &= E[(W_t - W_s)^2] + 2W_sE[W_t|\mathcal{W}_s] - W_s^2 \\ &= t - s + 2W_s^2 - W_s^2 = t - s + W_s^2. \end{aligned}$$

Thus,

$$E[W_t^2 - t|\mathcal{W}_s] = t - s + W_s^2 - t = W_s^2 - s.$$

It was used above that $W_t - W_s : \mathcal{N}(0, t - s)$ and $E[(W_t - W_s)^2] = D[(W_t - W_s)^2] = t - s$. □

Theorem 8.3 (Lévy's characterization of Brownian motion). *Let $\{X_t, t \geq 0\}$ be a stochastic process and denote by $\mathcal{F}_t = \mathcal{F}(X_s, 0 \leq s \leq t)$ a history of the process X_t until time t (t is included). Process $\{X_t, t \geq 0\}$ is Brownian motion if and only if*

1. $X_0 = 0$
2. the sample paths $t \mapsto X_t$ are continuous (X_t has a continuous trajectories)
3. X_t is martingale with respect to filtration $\mathcal{W}_t = \mathcal{F}\{X_s, s \in [0, t]\}$.
4. $X_t^2 - t$ is martingale with respect to filtration \mathcal{W}_t .

8.2 Exercises

1. Show that $e^{W_t}e^{-t/2}$ is a martingale if W_t is standard Brownian motion.
2. Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of independent random variables, such that $P\{X_n = 1\} = p$ and $P\{X_n = -1\} = q$ for all $n \in \mathbb{N}$ and $0 < p < 1, q = 1 - p$. Define $S_n := \sum_{j=1}^n X_j, n \geq 1$.

(a) Compute the probability mass function of the random variable $Y_i := \left(\frac{q}{p}\right)^{X_i}, i \in \mathbb{N}$.

(b) Prove that the sequence of random variables

$$M_n := \left(\frac{q}{p}\right)^{S_n}$$

is a martingale with respect to a filtration $\mathcal{F}_n = \sigma(X_j, 1 \leq j \leq n), n \geq 1$. With $\sigma(X_j, 1 \leq j \leq n), n \geq 1$ we denote σ -algebra generated by random variables $X_j, 1 \leq j \leq n$.

3. X_1, X_2, \dots is a sequence of independent identically distributed random variables with zero expectation and finite variance. Suppose that X_i cannot take absolute values greater than 1. Define

$$S_n := X_1 + \dots + X_n, \quad n \in \mathbb{N}.$$

Find a constant a such that $S_n^2 - aE(S_n^2)$ is a martingale with respect to a filtration \mathcal{F}_n which is a σ -algebra generated by X_1, \dots, X_n .

4. Consider a function $f : [0, \infty) \rightarrow \mathbb{R}$ and define a process

$$X_t := W_t^3 + f(t)W_t, \quad t \geq 0,$$

where W_t is a standard Brownian motion. Determine a function f such that $\{X_t\}$ is a martingale.

5. Find a constant a such that

$$V(t) = W^3(t) + aW^2(t) - 3tW(t) + t, \quad t \geq 0$$

is martingale, if $W(t)$ is standard Brownian motion.

6. Consider a sequence of independent, identically distributed random variables

$$X_n : \begin{pmatrix} -6 & -2 & 2 & 6 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}.$$

Prove that $\gamma_n = \cos(\pi n) \sin\left(\frac{\pi}{2} S_n\right), S_n = X_1 + \dots + X_n, n \in \mathbb{N}$ is a martingale with respect to the filtration $\mathcal{F}_n = \mathcal{F}(X_1, \dots, X_n)$.

7. Consider the sequence X_1, X_2, \dots of independent integrable identically distributed random variables such that

$$\phi(\lambda) = E[e^{\lambda X_1}] < +\infty, \quad \text{for some } \lambda \neq 0.$$

Prove that

$$M_n := \phi^{-n}(\lambda) e^{\lambda S_n}, \quad \text{where } S_n := \sum_{i=1}^n X_i$$

is martingale with respect to filtration $\mathcal{F}_n = \mathcal{F}(X_1, \dots, X_n)$.

8. Let X_t be a stochastic process with stationary and independent increments, zero expectation and variance $\sigma_X^2 t$. Prove that $Z_t = a(X_t^2 - \sigma_X^2 t) + Y_t$, $a \in \mathbb{R}$, is martingale with respect to the history of the stochastic process X_t until time t if Y_t is martingale with respect to the same filtration.
9. Consider a sequence of independent integrable identically distributed random variables Y_1, Y_2, \dots having all positive values. Check if $X_n = (Y_1 Y_2 \cdot \dots \cdot Y_n)^{1/n}$ is martingale with respect to filtration $\{\mathcal{F}_n\}$, where $\mathcal{F}_n = \mathcal{F}(Y_1, \dots, Y_n)$.
10. Consider an urn initially containing r red and b black balls. Repeated drawings are made from this urn as follows: after each drawing one returns a ball and adds a balls of the same color. Here r, b and a are positive integers. Let $\{Y_n\}$ be a sequence of random variables such that $Y_n = 1$ if the n -th ball drawn is red and $Y_n = 0$ if the n -th ball drawn is black. Let r_n and b_n be the number of red and black balls, respectively in the urn after the n -th draw has been completed.
- (a) Define Z_n as the number of red balls at the completion of the n -th draw. Express Z_n as a function of a random variables Y_1, \dots, Y_n . Prove that $\{Z_n\}$ is Markov chain and find the one step transition probabilities. Is $\{Z_n\}$ homogeneous or non-homogeneous?
- (b) Define X_n as the proportion of red balls in the urn at the completion of the n -th draw, that is $X_n = \frac{r_n}{r_n + b_n}$. Prove that $\{X_n\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}$, where $\mathcal{F}_n = \mathcal{F}(Y_1, \dots, Y_n)$.

9 Appendix

9.1 Eigenvalues and eigenvectors of a matrix

For a given square matrix A a **characteristic polynomial** is $p_A(\lambda) = \det(\lambda I - A)$, where I is identity matrix. The spectrum of a matrix A , $\sigma(A) = \{\lambda \in \mathbb{C} : p_A(\lambda) = 0\}$ is the set of its eigenvalues i.e. $\lambda \in \mathbb{C}$ is an eigenvalue of a matrix A is and only if $\lambda \in \sigma(A)$. Vector $x \in \ker(\lambda I - A)$ is an eigenvector corresponding to the eigenvalue λ . The **right eigenvector** is a column vector satisfying $\lambda x = Ax$ and the **left eigenvector** is a row vector satisfying $\lambda x = xA$.

Cayley Hamilton theorem: $p_A(A) = 0$.

The eigendecomposition (or spectral decomposition) of a diagonalizable matrix A is a decomposition of a diagonalizable matrix into a specific canonical form whereby the matrix is represented in terms of its eigenvalues and eigenvectors. Suppose that the eigenvalues of a $n \times n$ matrix A are denoted by $\lambda_1, \lambda_2, \dots, \lambda_n$. The right eigenvector corresponding to λ_i is denoted by r_i and the left eigenvector corresponding to λ_i is denoted by l_i . So,

$$\lambda_i r_i = A r_i, \quad \lambda_i l_i = l_i A.$$

It holds

$$\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n, \quad \text{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n.$$

Then

$$A = S^{-1} D S = T D T^{-1},$$

where $D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, the i -th row of a matrix S is l_i , while the i -th column of a matrix T is r_i . Then

$$A^n = S^{-1} D^n S = T D^n T^{-1}.$$