LECTURE NOTES ON STOCHASTIC PROCESSES

March 2, 2011

Salha Mamane
University of the Witwatersrand
School of Statistics and Actuarial Science
First Floor, Central Block
East Annex, E58B
Salha.Mamane@wits.ac.za

These course notes are adapted from the notes of Khewal Khesur
(with some additions and corrections from Prof Gordon Kass).
Stochastic Processes: course outline

Course code: STAT3017/* (3004)
Lecturer: Salha Mamane
Office: Central Block East Annexe, E58B
Email: Salha.Mamane@wits.ac.za

Course description

The objective of this course is to introduce stochastic processes, that is, models for systems that evolve unpredictably in time. This type of modelling is essential in diverse fields such as economics, finance, physics, climatology, telecommunication, biology...

The course is composed of three parts. The first part will expose some general facts about stochastic processes and their properties while the two last parts will focus on discrete state space processes with the Markov property.

The goals

At the end of this course, the students should be able to:

- check whether a given process has a certain property: Stationarity, independent increments, stationary increments, Markov property...
- classify the states of a Markov chain or a Markov jump process
- determine the expected sojourn time of a Markov chain or a Markov jump process in given state
- analyse the long term behaviour of Markov chains and Markov jump processes
- compute probabilities of events related to a Poisson process
- fit and simulate Markov chain and Markov jump process models

In general, this course is expected to improve the ability of students to formulate world problems into probability statements and use their knowledge of probability to solve them. Moreover, as the concepts and techniques developed in this course are also found in other courses (Time series analysis, Risk theory,...), the generality of this course should help the students put everything in perspective.

Course schedule

The course is composed of two lectures and one tutorial session per week.

Lectures: 10:15 – 12:00 Mondays (CB15)
Tutorials: 14:15 - 15:00 Wednesdays (CB128)
Consultations: 14:00 - 17:00 Fridays (E58B)

Assessment scheme

Test (18/05/2011): 30%.
Exam: 70%.
# Contents

1 Stochastic Processes ........................................ 2
   1.1 Introduction ............................................. 2
   1.2 Stochastic Processes ...................................... 3
      1.2.1 Characterisation of stochastic processes .......... 5
      1.2.2 Mean and covariance functions .................... 5
   1.3 Stationary processes .................................... 6
      1.3.1 Strict stationarity .................................. 6
      1.3.2 Weak stationarity .................................. 6
      1.3.3 Gaussian processes ................................ 7
   1.4 Independent increments-Stationary increments ............ 8
   1.5 Two examples of independent and stationary increments processes ........ 8
      1.5.1 Random walk ....................................... 8
      1.5.2 Brownian motion ................................... 9
   1.6 Markov processes ....................................... 9
   1.7 Counting processes ................................... 10
      1.7.1 Binomial counting process ......................... 10
      1.7.2 Poisson process ................................... 11
      1.7.3 Renewal processes ................................ 17
   1.8 $\sigma$-algebras and Filtrations ......................... 17
      1.8.1 $\sigma$-algebras ................................... 17
      1.8.2 Random variables .................................. 17
      1.8.3 Filtrations ....................................... 17

2 Markov chains .................................................. 19
   2.1 Definitions and immediate properties ...................... 19
   2.2 The n-step transition probabilities ....................... 21
   2.3 Classification of states ................................ 23
      2.3.1 Decomposition of the state space .................. 23
      2.3.2 Absorbing states .................................. 23
      2.3.3 Recurrent and transient states .................... 24
      2.3.4 Identification of recurrent/transient states ...... 25
      2.3.5 Period of a state .................................. 26
   2.4 Limiting behaviour of Markov chains ....................... 27
      2.4.1 Stationary distributions ........................... 28
      2.4.2 Absorbing probabilities ............................. 28
      2.4.3 Limiting behaviour of ergodic Markov chains ....... 28
      2.4.4 Limiting behaviour of non-ergodic Markov chains .... 29
         Canonical representation of a transition matrix ...... 29
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>Estimation in Markov chains models</td>
<td>31</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Estimation of the transition probabilities</td>
<td>31</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Assessing the goodness of fit</td>
<td>31</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Simulation of a Markov chain</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>Markov jump processes</td>
<td>32</td>
</tr>
<tr>
<td>3.1</td>
<td>Definitions and immediate properties</td>
<td>32</td>
</tr>
<tr>
<td>3.2</td>
<td>Transition intensities and generator matrix</td>
<td>34</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Definitions and basic properties</td>
<td>34</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Interpretation of transition intensities</td>
<td>35</td>
</tr>
<tr>
<td>3.2.3</td>
<td>The Kolmogorov differential equations</td>
<td>36</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Marginal distributions of a Markov jump process</td>
<td>39</td>
</tr>
<tr>
<td>3.3</td>
<td>Sojourn times of a Markov jump process</td>
<td>40</td>
</tr>
<tr>
<td>3.4</td>
<td>Stationary distributions and long-term behaviour of a Markov jump process</td>
<td>43</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Classification of states and stationary distributions</td>
<td>43</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Long-term behaviour of a Markov jump process</td>
<td>44</td>
</tr>
<tr>
<td>3.5</td>
<td>Fitting a Markov jump model</td>
<td>44</td>
</tr>
<tr>
<td>3.6</td>
<td>Examples of Markov jump processes</td>
<td>44</td>
</tr>
<tr>
<td>3.6.1</td>
<td>The Poisson process</td>
<td>44</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Birth and death processes</td>
<td>44</td>
</tr>
</tbody>
</table>
Chapter 1

Stochastic Processes

1.1 Introduction

The objective of this course is to introduce stochastic processes that is mathematical models for systems that evolve randomly in time. This type of modelling is essential in diverse fields such as economics, finance, physics, climatology, telecommunication, biology . . .

If we are only interested in the state of the system at a given time, that can be modelled by a random variable. To model the complete evolution of the system, we need to assign a random variable to each point in time. The collection of all these random variables is called a stochastic process.

When the random variables are independent, the task of studying a stochastic process is relatively easy. But assuming independence is not always possible as it is very often obviously contradicted by observations or intuition. In addition, dependence is sometimes a desirable property as it helps in predicting the future evolution of the system from its past evolution. We therefore need to know how to deal with the dependence of the random variables. Unfortunately, the general setting is too complicated to handle. Therefore, to tackle stochastic processes, we usually need to make some simplifying assumptions about the type of dependence. For example, it is sometimes possible to assume that the future evolution of the system is dependent on the past only through the present state of the system: this is called the Markov property. Some other times it is reasonable to assume that within a certain time duration, the system will behave in the future similarly as it has behaved in the past within same time duration: this is the notion of stationarity. For numerical value processes, it can be reasonable to assume that the growths of the processes within non overlapping time intervals are independent or identically distributed: this is the notion of independent and stationary increments.

The first chapter will expose some general facts about stochastic processes and their properties while the two last chapter will focus on discrete state space processes with the Markov property.
1.2 Stochastic Processes

Consider the exchange rate of South African Rand to US Dollar.

We may be interested in answering questions such as:

1. How do future rates depend on past rates?
2. What is the proportion of time the South African Rand is going up?
3. What is the yearly average rate?
4. What is the probability that the exchange rate will go down to 6.00 given that it is now 7.00?
5. What is the probability that the exchange rate will exceed 9.00?
6. How long would it take before the exchange rate exceeds 9.00?

Let’s play a gambling game. We will toss a coin. If it is head, you give me R1 and if it is tail, I give you R1. Suppose that the current wealth of each of us amounts to R100. Let $X_n$ denote my wealth after $n$ games. We may be interested in determining the following

1. How does my wealth after $n + k$ tosses depend on my wealth after $n$ tosses?
2. What is my average wealth on 100 tosses?
3. What is the probability that I will ever lose all my capital?

4. What is the probability that I will ever reach R150?

To model the time dependence of a random phenomenon, we use the mathematical concept of stochastic process.

**Definition 1.2.1.**

1. A stochastic process \( \{X_t : t \in T\} \) also denoted \((X_t)_{t \in T}\) is a collection of chronologically ordered random variables defined on the same probability space \( \Omega \).

2. The set \( T \subset \mathbb{R} \) is called the time space of the process \((X_t)_{t \in T}\).

3. The set \( S \) of all possible values for \( X_t, t \in T \) is called the state space of the process.

4. The time space is said discrete if the set \( T \) is countable. Otherwise, it is said continuous.

5. The state space is said discrete if for all fixed \( t \), the random variable \( X_t \) is discrete. It is said continuous if for all \( t \), the random variable \( X_t \) is continuous.

**Example 1.2.2.**

1. Let \((A_t)_{t \in T}\) the stochastic process that models the availability of a book at the time of inventory. The librarian classify books as available, misshelved, issued or missing. Then \( S = \{ \text{Available, Misshelved, issued, Missing} \} \) is the state space. If inventories are conducted once every month, then the time space is \( T = \{0, 1, 2, \ldots \} \). Therefore \((A_t)_{t \in T}\) is a discrete time process with discrete state space.

2. Let \((X_t)_{t \in T}\) the stochastic process that models the state of health of policyholders of a life insurance company. The company can classify its policyholders as healthy, sick or dead. Then \( S = \{ \text{Healthy, sick, Dead} \} \) is the state space. If the health states of policyholders are observed continuously, then \( T = [0, \infty) \) is the time space. \((X_t)_{t \in T}\) is a continuous time process with a discrete state space. But the health states of policyholders can also be observed daily; in this case, the time space is \( T = \{0, 1, 2 \ldots \} \) and \((X_t)_{t \in T}\) is a discrete time process with discrete state space.

3. Consider a model for the daily maximal temperatures observed in Johannesburg. The time space is discrete and the state space is continuous since temperature can assume any real value. Therefore the model here is a stochastic process \((X_n)_{n \in \mathbb{N}}\), with \( X_n \) the maximal temperature observed on the \( n \)th day.

4. Let \((X_t)_{t \in T}\) the stochastic process that models the stock price of a company. Then the state space can be considered as continuous \( (S = \mathbb{R}^+) \) if we assume that the stock price can assume any positive real number. The time space \( T \) is discrete.

**Remark 1.2.3.**

1. Many natural random phenomena are more accurately modelled by continuous time stochastic processes. But very often, only discrete data is available. On the other hand, because of the availability of powerful sophisticated mathematical tools, discrete time space processes with continuous state space are sometimes approximated by continuous time space processes with continuous state space.

2. A discrete time space stochastic process with continuous state space is called a time series. Time series are studied in details in another course.
Remark 1.2.4. In our definition of stochastic process we restricted ourselves to temporal processes. A more general definition that for instance allows for spatial processes is the following:

Definition 1.2.5. A stochastic process is a collection of random variables defined on a given probability space and indexed by a parameter \( t \), \( t \in T \).

\[ T \] can be an unidimensional or multidimensional space.

1.2.1 Characterisation of stochastic processes

Definition 1.2.6 (Sample paths and finite-dimensional distributions).

A sample path of the stochastic process \((X_t)_{t \in T}\) is any realisation \( \{x_t, t \in T\} \) of the chronologically ordered random variables \( \{X_t, t \in T\} \).

Sample paths model possible evolutions of the dynamic system. Therefore for a full characterisation of a stochastic process, we need to assign a the probability of observing each sample path. Fortunately, it is sufficient to specify the distributions of all \((X_{t_1}, \ldots, X_{t_n})\), \( t_1, \ldots, t_n \in T \). These distributions are called the finite-dimensional distributions of the stochastic process. In other terms, the full specification of the stochastic process requires that for all \( n \), we know \( P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \ldots, X_{t_n} \leq x_n) \) for all \( t_1, \ldots, t_n \) and \( x_1, x_2, \ldots, x_n \).

1.2.2 Mean and covariance functions

Definition 1.2.7.

1. The mean function of a process \((X_t)_{t \in T}\) is the function \( m \) defined by

\[ m(t) = E(X_t), \quad \forall t \in T. \]

2. The covariance function of a process \((X_t)_{t \in T}\) is the function \( K \) defined by

\[ K(s, t) = \text{cov}(X_s, X_t), \quad \forall (s, t) \in T^2. \]

Remark 1.2.8. \( K(t, t) = \text{var}(X_t) \).

Exercise 1.2.9. Consider a random variable \( X \) uniformly distributed on \([0, \pi]\). Define the stochastic process \((Y_t)_{t \geq 0}\) by

\[ Y_t = X \cos(\omega t). \]

1. Determine the mean function of the process \((Y_t)_{t \geq 0}\).

2. Determine the covariance function of the process \((Y_t)_{t \geq 0}\).

Exercise 1.2.10. Consider two independent random variables \( X \) uniformly distributed on \([0, \pi]\) and \( \varphi \) uniformly distributed on \([-\pi, \pi]\). Define the stochastic process \((Y_t)_{t \geq 0}\) by

\[ Y_t = X \cos(\omega t + \varphi), \]

where \( \omega \in [0, \pi] \).

1. Determine the mean function of the process \((Y_t)_{t \geq 0}\).

2. Determine the covariance function of the process \((Y_t)_{t \geq 0}\).
Classification of stochastic processes

It is clear that assigning directly probabilities to all sample paths is impossible when T is infinite but even when T is finite this is usually an impossible task. Specifying the finite dimensional distributions is less intimidating but still usually impossible. Therefore, we try to achieve this, by assuming that the process has some simplifying probability structure.

1.3 Stationary processes

One interesting property is the following: different observations of the process on intervals of the same length should provide the same information about the process. That is the statistical characteristics of the process do not change in time. This is the notion of stationarity that we formally state below.

1.3.1 Strict stationarity

Definition 1.3.1 (Strict stationarity). A stochastic process is said to be stationary or strictly stationary if the finite-dimensional distributions are invariant under a time shift, that is, for all $s, t_1, t_2, \ldots, t_n$,

$(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ and $(X_{t_1+s}, X_{t_2+s}, \ldots, X_{t_n+s})$ have the same distribution.

Example 1.3.2. Let $(X_t)_{t \in T}$ a stochastic process such that the $X_t, T \in T$ are i.i.d. Then $(X_t)_{t \in T}$ is a stationary process.

Exercise 1.3.3. Let $(X_t)_{t \in T}$ a stationary stochastic process.

1. Show that the mean function is constant.

2. Show that for all $s, t$ the covariance function $K(t, t + s)$ only depends on $s$.

Strict stationarity is usually hard to prove or check. We will thus introduce a related property that is easier to handle.

1.3.2 Weak stationarity

Definition 1.3.4 (Weak stationarity). Consider a stochastic process $(X_t)_{t \in T}$ with finite second order moments. $(X_t)_{t \in T}$ is said to be weakly stationary if its mean function and covariance function are invariant under a time shift.

i.e

$m(t) = \text{constant} = m$ and $K(t, t + s) = \gamma(s)$.

Remark 1.3.5.

1. Weak stationarity is also called second order stationarity.

2. Strict stationarity implies weak stationarity. (Provided the process has finite second moments).

3. Weak stationarity implies constant variance.

4. The function $\gamma$ is called the autocovariance function of the weakly stationary process. It satisfies the following:

$\gamma(-t) = \gamma(t)$. 
Example 1.3.6 (White noise). A white noise is a weakly stationary process with a mean function and an autocovariance function given respectively by

\[ m(t) = 0 \]

\[ \gamma(t) = \begin{cases} 0 & \text{if } t \neq 0 \\ \sigma^2 & \text{if } t = 0 \end{cases} \]

Exercise 1.3.7. Consider two independent identically distributed random variables \( A \) and \( B \) following a standard normal distribution. Set

\[ X_n = A \cos(n \omega) + B \sin(n \omega), \quad \omega \in [0, \pi]. \]

Show that the process \((X_n)_{n \in \mathbb{N}}\) is weakly stationary.

1.3.3 Gaussian processes

Exercise 1.3.8. Consider two independent normally distributed random variables \( X \) and \( Y \).

\[ X \sim N(\mu_1, \sigma_1^2) \quad \text{and} \quad X \sim N(\mu_2, \sigma_2^2). \]

Determine the joint density function of \( X \) and \( Y \).

Definition 1.3.9 (Multivariate normal distribution). A random vector \( X = (X_1, \ldots, X_n) \) is said to be Gaussian if the joint distribution of \((X_1, \ldots, X_n)\) is a multivariate normal distribution. That is the joint density function is given by

\[
    f(x) = f(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \Gamma}} \exp\left[-\frac{1}{2}(x - m)\Gamma^{-1}(x - m)\right]
\]

where \( m = (m_1, \ldots, m_2) \) is a vector and \( \Gamma \) is a symmetric positive definite matrix.

Proposition 1.3.10.

\[ \mathbb{E}(X) = \mathbb{E}(X_1, \ldots, X_n) = m \]

\[ \text{cov}(X_i, X_j) = \Gamma_{ij} \]

Theorem 1.3.11.

A random vector \( X = (X_1, \ldots, X_n) \) is Gaussian if and only if for all \((\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{R}^n, \)

\[
    \sum_{i=1}^{n} \alpha_i X_{t_i} = \alpha_1 X_1 + \alpha_2 X_{t_2} + \ldots + \alpha_n X_{t_n}
\]

has a normal distribution.

Definition 1.3.12. A process \((X_t)_{t \in T}\) is called a Gaussian process if all its finite-dimensional distributions are multivariate normal distributions. That is for all \((t_1, t_2, \ldots, t_n) \in T^n, (X_{t_1}, X_{t_2}, \ldots, X_{t_n})\) has a multivariate normal distribution.

Gaussian processes are fully fully determined by their mean and covariance functions. Therefore the following statement holds.

Proposition 1.3.13. A Gaussian process is strictly stationary if and only if it is weakly stationary.
Proof. We only have to show that for a Gaussian process, second order stationarity implies strict stationarity since the converse holds in general for any stochastic process.

Consider a weakly stationary Gaussian process \((X_t)_{t \in T}\). Then since the process is Gaussian, for all \(t_1, t_2, \ldots, t_n \in T\) and \(s > 0\),
\[
(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) \sim N(m, K)
\]
and
\[
(X_{t_1+s}, X_{t_2+s}, \ldots, X_{t_n+s}) \sim N(\mu, \Gamma).
\]
From the weak stationarity of the process we get:
\[
\mu_i = \mathbb{E}(X_{t_i+s}) = \mathbb{E}(X_{t_i}) = m_i, \quad \forall i
\]
and for all \(i, j\),
\[
\Gamma_{ij} = \text{cov}(X_{t_i+s}, X_{t_j+s}) = \gamma(t_j + s - t_i - s) = \gamma(t_j - t_i) = \text{cov}(X_{t_i}, X_{t_j}) = K_{ij}.
\]

\[
\Box
\]

1.4 Independent increments—Stationary increments

**Definition 1.4.1.**

1. The increment of the process \((X_t)_{t \in T}\) between time \(s\) and \(t\), \(t > s\) is the difference \(X_t - X_s\).

2. A process \((X_t)_{t \in T}\) is said to be an independent increment process if for all \(t_0 < t_1 < t_2 < \ldots < t_n\), the increments
\[
X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}
\]
are independent. In particular,
\[
\text{cov}(X_{t_{i+1}} - X_{t_i}, X_{t_{j+1}} - X_{t_j}) = 0, \quad \forall i \neq j
\]

3. A process \((X_t)_{t \in T}\) is said to have stationary increments if for all \(t, t + h \in T\) the distribution of the increment \(X_{t+h} - X_t\) only depends on \(h\).

\[\text{i.e } \forall s, t \in T, X_t - X_s\text{ has the same distribution as }X_{t+u} - X_{s+u}.\]

**Remark 1.4.2.**

It should be noted that in all the definitions above, the time intervals do not overlap.

1.5 Two examples of independent and stationary increments processes

**1.5.1 Random walk**

A stochastic process \((X_n)_{n \in \mathbb{N}}\) is called a general random walk if
\[
X_n = X_0 + \sum_{i=1}^{n} Y_i,
\]
where \(Y_1, Y_2, \ldots\) are independent identically distributed and independent of \(X_0\).

A random walk has independent and stationary increments.
1.5.2 Brownian motion

Named after the Scottish botanist Robert Brown, a Brownian motion is the continuous version of random walk processes. It plays a prominent role in the general theory of continuous time stochastic processes with continuous state space.

**Definition 1.5.1.** A Brownian motion \((W_t)_{t \in \mathbb{R}}\) also called a Wiener process is a continuous time stochastic process that satisfies the following properties:

1. \(W_0 = 0\);
2. \((W_t)_{t \in \mathbb{R}}\) has independent increments;
3. \(W_{t+s} - W_t \sim N(0, s)\).

**Exercise 1.5.2.** Consider a Brownian motion \((W_t)_{t \in \mathbb{R}}\). Determine the covariance function of \((W_t)_{t \in \mathbb{R}}\).

1.6 Markov processes

We are now going focus on stochastic processes for which the probability of observing a particular future evolution depends only on the current state of the system and not on its past evolution. Such processes are called Markov processes. Markov processes are relatively simple models and can perfectly model many real world phenomena.

**Definition 1.6.1.** A process \((X_t)_{t \in T}\) is said to be a Markov process if it satisfies the following property called the Markov property

\[
P(X_{t_{n+1}} \in B | X_{t_1}, \ldots, X_{t_n}) = P(X_{t_{n+1}} \in B | X_{t_n})
\]

for all \(t_1 < t_2 < \ldots < t_n \in T\) and \(B \subset S\).

A discrete time space Markov process with discrete state space is called a Markov chain. Markov chains will be studied in details in chapter 2.

A continuous time space Markov process with discrete state space is called a Markov jump process. Markov jump processes will be studied in details in chapter 3.

**Proposition 1.6.2.** An independent increment stochastic process has the Markov property.

**Proof.**

Consider an independent increment process \((X_t)_{t \in T}\).

On the one hand, we have

\[
P(X_{t_{n+1}} = x_{n+1} | X_{t_1} = x_1, \ldots, X_{t_n} = x_n) = \frac{P(X_{t_1} = x_1, \ldots, X_{t_{n+1}} = x_{n+1})}{P(X_{t_1} = x_1, \ldots, X_{t_n} = x_n)}
\]

\[
= \frac{P(X_{t_1} = x_1, X_{t_2} - X_{t_1} = x_2 - x_1, \ldots, X_{t_{n+1}} - X_{t_n} = x_{n+1} - x_n)}{P(X_{t_1} = x_1, X_{t_2} - X_{t_1} = x_2 - x_1, \ldots, X_{t_n} - X_{t_{n-1}} = x_n - x_{n-1})}
\]

\[
= \prod_{k=2}^{n+1} \frac{P(X_{t_k} - X_{t_{k-1}})}{P(X_{t_k} - X_{t_{k-1}})} P(X_{t_1} = x_1)
\]

\[
= \prod_{k=2}^{n} \frac{P(X_{t_k} - X_{t_{k-1}})}{P(X_{t_k} - X_{t_{k-1}})} P(X_{t_1} = x_1)
\]

\[
= P(X_{t_{n+1}} - X_{t_n} = x_{n+1} - x_n).
\]
On the other hand, we have
\[ P(X_{n+1} = x_{n+1} | X_n = x_n) = P(X_{n+1} - X_n = x_{n+1} - x_n | X_n = x_n) \]
\[ = P(X_{n+1} - X_n = x_{n+1} - x_n). \]

Therefore
\[ P(X_{n+1} = x_{n+1} | X_1 = x_1, \ldots, X_n = x_n) = P(X_{n+1} - X_n = x_{n+1} - x_n). \]

\[ \square \]

1.7 Counting processes

**Definition 1.7.1** (Counting process). A process \((N_t)_{t \in T}\) is called a counting process if

1. the state space of \((N_t)_{t \in T}\) is \(\mathbb{N}\)
2. \(N_t\) is a non-decreasing function of \(t\).

Counting processes model the number of occurrences of random events.

1.7.1 Binomial counting process

**Definition 1.7.2** (Bernoulli process). A Bernoulli process is a stochastic process \((X_n)_{n \in \mathbb{N}}\) such that the \(X_n\) are iid Bernoulli distributed random variables.

\[ X_n = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases} \]

The Bernoulli process models the occurrence of random events. The process \((N_n)_{n \in \mathbb{N}}\) defined by

\[ N_n = \sum_{i=1}^{n} X_i \]

is thus a counting process that models the number of events that occurred in the interval \([1, n]\). \((N_n)_{n \in \mathbb{N}}\) is called the Binomial counting process and we have

\[ N_n \sim B(n, p). \]

We can also derive the process \((T_n)_{n \geq 1}\) defined by

\[ T_n = \min \{ k, N_k = n \}. \]

\(T_n\) models the time of occurrence of the \(n\)th event.

Another associated process is the process \((Y_n)_{n \geq 1}\) that models the time elapsed between occurrences of two consecutive events. We have

\[ \begin{cases} Y_1 = T_1 \\ Y_n = T_n - T_{n-1}, \text{ for } n \geq 2. \end{cases} \]

**Remark 1.7.3.**

We have
\[ T_n = \sum_{i=1}^{n} Y_i, \quad \forall \ n \geq 1. \]

\[
\begin{cases}
X_1 = N_1 \\
X_n = N_n - N_{n-1}, \quad \text{for } n \geq 2.
\end{cases}
\]

\[ \{N_n \geq m\} \iff \{T_m \leq n\} \]

\[ N_n = \sum_{k=1}^{\infty} I(T_k \leq n) \]

\[ Y_n \sim \text{Geom}(p) \quad \text{i.e. } P(T_n = m) = (1 - p)^m p \]

\[ T_n \] follows a negative binomial distribution \( \mathcal{NB}(n, p) \).

**Exercise 1.7.4.**

1. Determine the distribution of \( Y_n \).
2. Determine the distribution of \( T_n \).
3. Show that \( P(Y_n > y + m | Y_n > m) = P(Y_n > y) \). Comment on the result.

### 1.7.2 Poisson process

**Definition 1.7.5** (Poisson process).

A Poisson process with rate \( \lambda \) is a continuous time counting process \( (N_t)_{t \in T} \) which satisfies the following:

1. \( N_t = 0 \).
2. \((N_t)_{t \in T}\) has independent increments.
3. The number of events that occur in any interval of length \( t \) follows a Poisson distribution with parameter \( \lambda t \).

\[ i.e. \quad P(N_{t+s} - N_t = n) = \frac{e^{-\lambda s}(\lambda s)^n}{n!}. \]

The Poisson process is thus an independent and stationary increments process.

**Exercise 1.7.6.** Let \((X_t)_{t \in T}\) a stochastic process with stationary increments. Is \((X_t)_{t \in T}\) stationary? weakly stationary?

The Poisson process models the number of occurrences of events in a given time interval such that the occurrence of an event in a sub-interval is independent of the occurrence of in any other sub-interval, the probability of occurrence of more than one event at within a short time interval \([t, t+h]\) is very small (negligible with respect to \( h \)), the probability that exactly one event occurs during a short time interval \([t, t+h]\) is proportional to the length \( h \) of the interval. Formally the Poisson process with rate \( \lambda \) is a counting process \((N_t)_{t \geq 0}\) satisfying

1. \((N_t)_{t \geq 0}\) has independent increments
2. \( P(N_{t+h} - N_t > 1) = o(h) \)
Proposition 1.7.7. The previous definition of a Poisson process is equivalent to the definition given in the first chapter (Definition 1.7.5).

Proof.

1. Necessary condition:
   The proof will be complete when we show that
   \[ p_k(t) = \mathbb{P}(N_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}. \]  
   (1.1)

   We have
   \[ p_0(t + h) = \mathbb{P}(N_{t+h} = 0) = \mathbb{P}(N_t = 0, N_{t+h} = N_t) = p_0(t)(1 - \lambda h + o(h)) \]
   \[ p_0(t + h) - p_0(t) \]
   \[ \frac{h}{h} = -\lambda h p_0(t) + o(h) \]
   \[ p_0'(t) = \lim_{h \to 0} \frac{p_0(t + h) - p_0(t)}{h} = -\lambda p_0(t). \]

   And for \( k > 0 \),
   \[ p_k(t + h) = \mathbb{P}(N_{t+h} = k) = \mathbb{P}(N_t = k, N_{t+h} = N_t + 1) = p_k(t)(1 - \lambda h + o(h)) + p_{k-1}(\lambda h + o(h)) \]
   \[ p_k(t + h) - p_k(t) \]
   \[ \frac{h}{h} = -\lambda h p_k(t) + \lambda h p_{k-1} + o(h) \]
   \[ p_k'(t) = \lim_{h \to 0} \frac{p_k(t + h) - p_k(t)}{h} = -\lambda p_k(t) + \lambda p_{k-1}(t). \]

   We thus have to solve the differential equations
   \[ \begin{cases} p_0'(t) = -\lambda p_0(t) \\ p_k'(t) = -\lambda p_k(t) + \lambda p_{k-1}(t), & \forall k > 0 \end{cases} \]

   with the boundary conditions
   \[ \begin{cases} p_0(0) = 1 \\ p_k(0) = 0 \end{cases} \]

   The solution of the first equation is thus
   \[ p_0(t) = e^{-\lambda t}. \]
We verify that \( p_0(t) \) satisfy (1.1). Now suppose that (1.1) is satisfied for a certain \( k \). We have

\[
\begin{align*}
p'_k(t) &= -\lambda p_{k+1}(t) + \lambda p_k(t) \\
&= -\lambda p_{k+1}(t) + \lambda e^{-\lambda t} (\lambda t)^k \frac{k!}{k!} \\
p'_k(t) + \lambda p_k(t) &= \lambda e^{-\lambda t} (\lambda t)^k \\
p'_k(t)e^{\lambda t} + \lambda p_k(t)e^{\lambda t} &= \lambda e^{-\lambda t} (\lambda t)^k \\
\frac{d}{dt} (p_{k+1}(t)e^{\lambda t}) &= \lambda (\lambda t)^k \frac{k!}{k!} = \lambda k+1 \frac{(t)^k}{k!} \\
p_{k+1}(t)e^{\lambda t} &= \lambda \frac{(\lambda t)^{k+1}}{(k+1)!} \\
p_k(t) &= e^{-\lambda t} (\lambda t)^{k+1} \frac{k!}{(k+1)!} \\
p'_{k+1}(t) &= -\lambda p_{k+1}(t) + \lambda e^{-\lambda t} (\lambda t)^k \frac{k!}{k!}
\end{align*}
\]

2. Sufficient condition:

The proof will be complete when we show that

* \( \mathbb{P}(N_{t+h} - N_t > 1) = o(h) \)
* \( \mathbb{P}(N_{t+h} - N_t = 1) = \lambda h + o(h) \)

We have

\[
\begin{align*}
\mathbb{P}(N_{t+h} - N_t = 1) &= \frac{e^{-\lambda h}(\lambda h)^1}{1!} \\
&= \lambda h \left( \sum_{k=0}^{\infty} (-\lambda h)^k \frac{k!}{k!} \right) = \sum_{k=0}^{\infty} (-\lambda h)^{k+1} \frac{k!}{k!} \\
&= \lambda h + o(h).
\end{align*}
\]

\[
\begin{align*}
\mathbb{P}(N_{t+h} - N_t = 0) &= \frac{e^{-\lambda h}(\lambda h)^0}{0!} \\
&= e^{-\lambda h} \\
&= 1 - \lambda h + o(h).
\end{align*}
\]

\[
\begin{align*}
\mathbb{P}(N_{t+h} - N_t > 1) &= 1 - \mathbb{P}(N_{t+h} - N_t = 1) - \mathbb{P}(N_{t+h} - N_t = 1) \\
&= o(h).
\end{align*}
\]
We have the relations
\[ \{ N_t \geq n \} = \{ T_n \leq t \} \]
and
\[ N_t = \max \{ n, T_n \leq t \}. \]
Therefore for all \( n \in \mathbb{N} \),

**Proposition 1.7.8.**

\[
P(N_t = n) = P(N_t \geq n) - P(N_t \geq n + 1) = P(T_n \leq t) - P(T_{n+1} \leq t).
\]

**Proposition 1.7.9.**

Consider a counting process \( (N_t)_{t \in \mathbb{T}} \) with intensity \( \lambda \). Denote the corresponding event times process \( (T_n)_{n \in \mathbb{N}} \). Then \( (N_t)_{t \in \mathbb{T}} \) is a Poisson process if and only if \( (T_n)_{n \in \mathbb{N}} \) has independent exponentially distributed increments.

**Proof.**

1. Necessary condition:
   Let \( Y_n = T_n - T_{n-1}, n \geq 0 \) denote the increments.
   
   \[ P(Y_1 > t) = P(N_t < 1) = P(N_t = 0) = e^{-\lambda t}. \]

   The first interval time is therefore exponentially distributed with parameter \( \lambda \).

   \[ P(Y_2 > t | Y_1 = s) = P(Y_2 + Y_1 > t + s | Y_1 = s) = P(T_2 > t + s | N_s = 1) = P(N_{t+s} - N_s = 0 | N_s = 1) = P(N_{t+s-s} = 0) = e^{-\lambda s}. \]

   Thus \( Y_2 \) is independent of \( Y_1 \) and is exponentially distributed with parameter \( \lambda \).

2. Sufficient condition: we will only prove that \( N_t \sim P(\lambda t) \).

   We have
   \[ T_n = \sum_{i=1}^{n} (T_i - T_{i-1}) = \sum_{i=1}^{n} Y_i \]
   and since the \( Y_i, 1 \leq i \leq n \) are independent and \( Y_i \sim \exp(\lambda) \), then
   \[ T_n \sim \text{Gamma}(n, \lambda). \]

   Therefore
   \[
P(N_t \geq n) = P(T_n \leq t) = \int_{0}^{t} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)} dx
   = \frac{\lambda^n}{\Gamma(n)} \int_{0}^{t} x^{n-1} e^{-\lambda x} dx = \frac{\lambda^n}{\Gamma(n)} I_n
   \]
   where
   \[
   I_n = \int_{0}^{t} x^{n-1} e^{-\lambda x} dx = \left[ \frac{x^n e^{-\lambda x}}{n} \right]_{0}^{t} - \int_{0}^{t} \frac{x^n}{n} e^{-\lambda x} dx
   = \frac{t^n}{n} e^{-\lambda t} + \frac{\lambda}{n} \int_{0}^{t} x^n e^{-\lambda x} dx = \frac{t^n}{n} e^{-\lambda t} + \frac{\lambda}{n} I_{n+1}.
   \]
Therefore

\[ P(N_t \geq n) = \frac{\lambda^n}{\Gamma(n)} \left( \frac{t^n}{n} e^{-\lambda t} + \frac{\lambda}{n} I_{n+1} \right) \]

\[ = \frac{\lambda^n t^n}{n \Gamma(n)} e^{-\lambda t} + \frac{\lambda^{n+1}}{n \Gamma(n)} I_{n+1} \]

\[ = \frac{(\lambda t)^n}{\Gamma(n+1)} e^{-\lambda t} + \frac{\lambda^{n+1}}{\Gamma(n+1)} I_{n+1} \]

\[ = \frac{(\lambda t)^n}{\Gamma(n+1)} e^{-\lambda t} + P(N_t \geq n + 1). \]

Thus

\[ P(N_t = n) = P(N_t \geq n) - P(N_t \geq n + 1) \]

\[ = \frac{(\lambda t)^n}{\Gamma(n+1)} e^{-\lambda t} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \]

Exercise 1.7.10.

The number of shooting stars observed over Johannesburg is a Poisson process with intensity 0.5 \((\text{day}^{-1})\).

1. What is the probability of observing one shooting star on Monday (first day of observation) and two shooting stars over the rest of the week?

2. What is the probability of observing one shooting star on Tuesday if one shooting star was observed on Monday?

3. What is the probability that less than 2 shooting stars had been observed in the first half of April given that 10 shooting stars had been observed in the whole month?

4. (a) On average, how long must one wait to observe 10 shooting stars?
   (b) What is the standard deviation of this waiting time?

5. If 1% of meteorites make it to the surface,
   (a) what is the probability of at least one impact in a year?
   (b) what is the probability of observing one impact in one year and another one in the following year?
   (c) Let \((M_t)_{t \geq 0}\) a Poisson process with intensity 0.5% \((\text{day}^{-1})\). Determine the probability of observing one occurrence of this process in one year and another one the following year. Comment on the result.

Exercise 1.7.11.

1. Simulate a Poisson process with transition rate 0.5.

2. Answer question 4 of the previous exercise using simulations.

Exercise 1.7.12. Determine the covariance function of a Poisson process.

Exercise 1.7.13.
Consider a Poisson process \((N_t)_{t \in T}\).
Prove that given that \(N_t = n\), \(N_s\) has a binomial distribution \(B(n, \frac{s}{t})\) for all \(0 \leq s < t\).

**Exercise 1.7.14.** Consider a Poisson \((N_t)_{t \in T}\).
Show that given that \(N_t = 1\), the first event time \(T_1\) is uniformly distributed on the interval \([0, t]\).

**Exercise 1.7.15 (Link between Poisson process and Uniform distribution).**
Let \(X_1, X_2, \ldots, X_n\) be a random sample from \(U(0, a)\) a uniform distribution on \([0, a]\), \(a > 0\).
Let \(X_{(1)} < X_{(2)} < \ldots < X_{(n)}\) denote the ordered sample.

1. Show that the ordered sample has a joint density function given by
   \[
   f(x_{(1)}, x_{(2)}, \ldots, x_{(n)}) = \begin{cases} \frac{n!}{a^n} x_{(1)} < x_{(2)} < \ldots < x_{(n)} < a \\ 0 & \text{else} \end{cases}
   \]

2. Let \((N_t)_{t \geq 0}\) be a Poisson process with intensity \(\lambda\). Let \(T_i\) be the occurrence time of the \(i\)th event.
   (a) Show that the joint probability density function of \((T_1, T_2, \ldots, T_n)\) is
       \[
       f(t_1, t_2, \ldots, t_n) = \begin{cases} \lambda^n e^{-\lambda t_a} & 0 < t_1 < \ldots < t_n \\ 0 & \text{otherwise} \end{cases}
       \]
   (b) Show that
       \[
       f(t_1, t_2, \ldots, t_n | N_t = n) = \begin{cases} \frac{n!}{t^n} & 0 < t_1 < \ldots < t_n \\ 0 & \text{elsewhere} \end{cases}
       \]

**Exercise 1.7.16 (Compound Poisson process).**
Consider a Poisson process \((N_t)_{t \geq 0}\) with rate \(\lambda\) and \((X_n)_{n \geq 1}\) a stochastic process independent of \((N_t)_{t \geq 0}\).
Let \(S_t = \sum_{i=1}^{N_t} X_i\).

1. Show that
   \[
   \mathbb{E}(S_t) = \mathbb{E}(N_t)\mathbb{E}(X_1).
   \]

2. Show that
   \[
   \mathbb{V}(S_t) = \mathbb{E}(N_t)\mathbb{V}(X_1) + \mathbb{E}^2(X_1)\mathbb{V}(N_t).
   \]

3. The number of claims that arrive at an insurance company follows a Poisson process with rate \(\lambda = 12\) day\(^{-1}\).
   (a) The amounts claimed are assumed to be independent uniformly distributed on \([1000, 10000]\). What is the expected total claim in one week? What is the standard deviation of the total claim in one week?
   (b) Claims are classified in two groups according to the amount claimed: “large claims” and “small claims”. Large claims represent 2% of all claims. What is the probability of arrival of two large claims in one week and two others in the following week?
1.7.3 Renewal processes

Definition 1.7.17 (Renewal processes). A counting process \((N_t)_{t \in T}\) is called a renewal process if the corresponding event times process \((T_n)_{n \in \mathbb{N}}\) has independent stationary increments.

1.8 \(\sigma\)-algebras and Filtrations

1.8.1 \(\sigma\)-algebras

Not always can probabilities be assigned to all subsets of a sample space \(\Omega\). To have a consistent mathematical theory, the definition of probabilities must sometimes be restricted to some subset \(F\) of the power set of \(\Omega\) and this subset \(F\) must satisfy some properties.

Definition 1.8.1. Let \(\Omega\) be a set and let denote \(\mathcal{P}(\Omega)\) the power set of \(\Omega\), that is, the set of all subsets of \(\Omega\). A subset \(F\) of \(\mathcal{P}(\Omega)\) is said to be a \(\sigma\)-algebra on \(\Omega\) if it satisfies the following

1. \(\Omega \in F\)
2. For any \(A \in \mathcal{P}(\Omega)\), if \(A \in F\), the \(A^c \in F\)
3. For any countable collection \((A_i)_{i \in I} \in \mathcal{P}(\Omega)\), if \(A_i \in F\), \(\forall i \in I\), then \(\bigcup_{i \in I} A_i \in F\).

Example 1.8.2. The power set \(\mathcal{P}(\Omega)\) of \(\Omega\) is a \(\sigma\)-algebra on \(\Omega\).

Exercise 1.8.3. Let \(F\) a \(\sigma\)-algebra on \(\Omega\). Show that

1. \(\emptyset \in F\)
2. For any countable collection \((A_i)_{i \in I} \in \mathcal{P}(\Omega)\), if \(A_i \in F\), \(\forall i \in I\), then \(\bigcap_{i \in I} A_i \in F\).

Exercise 1.8.4. Consider \(F_1\) and \(F_2\) two \(\sigma\)-algebras on a set \(\Omega\). Show that \(F_1 \cap F_2\) is a \(\sigma\)-algebra on \(\Omega\).

1.8.2 Random variables

For any random variable there is an underlying sample space \(\Omega\), the set of all possible outcomes and an underlying \(\sigma\)-algebra \(F\) that contains all events to which a probability can be attached. The triple \((\Omega, F, \mathbb{P})\) is called a probability space. Therefore a rigorous definition of a random variable is following

Definition 1.8.5. A random variable on the probability space \((\Omega, F, \mathbb{P})\) is a mapping \(X : \Omega \to \mathbb{R}\) such that for all \(x \in \mathbb{R}\), \(\{X \leq x\} \in F\).

1.8.3 Filtrations

The amount of uncertainty about a stochastic process decreases with time as we observe realisations of the process and realisations of other related processes. The information available at time \(t\) is modelled by a \(\sigma\)-algebra \(F_t\) which contains all events for which it is known at time \(t\) if they have occurred or not. It is clear that for any two instants \(t_1\) and \(t_2\), with \(t_1 < t_2\), we should have \(F_{t_1} \subset F_{t_2}\) as information increases with time.

Definition 1.8.6. A collection of \(\sigma\)-algebras \(\mathcal{F} = (F_t)_{t \in T}\) such that for any \(t_1, t_2 \in T\), with \(t_1 < t_2\), \(F_{t_1} \subset F_{t_2}\) is called a filtration.

For any stochastic process \((X_t)_{t \in T}\), there is
* an underlying sample space $\Omega$ such that each outcome $\omega \in \Omega$ determines a sample path $X_t(\omega)$.

* an underlying $\sigma$-algebra $F$ with respect to which all the random variables are defined.

* an underlying filtration $\mathcal{F} = (F_t)_{t \in T}$ such that $F_t$ contains all the information about the path of the process up to time $t$.

The quadruple $(\Omega, F, (F_t)_{t \in T}, \mathbb{P})$ is called a filtered probability space.

**Remark 1.8.7.** Consider a stochastic process $(X_t)_{t \in T}$ defined on the filtered probability space $(\Omega, F, (F_t)_{t \in T}, \mathbb{P})$. The Markov property can be written

$$
\mathbb{P}(X_t \leq x | F_s) = \mathbb{P}(X_t \leq x | X_s),
$$

for all $s, t \in T, s < t$. 
