Applications in Martingales and Stochastic Processes

[redacted]

Department of Mathematics, Toronto Metropolitan University
Email address: [redacted]

Key words and phrases. *Mathematical Modeling*, *Joint probability*, *Markov Chains*, *Markov Processes*, *Stochastic Calculus*, *Martingales*, *Brownian motion*, *Time Series*

Contents

Preface

These lectures notes are based on the course "Introduction to Stochastic Processes" taught at Toronto Metropolitan University, as part of the undergraduate programs Financial Mathematics and Mathematics and Its Applications. Although the emphasis is towards applications in finance the notes aim to a more general scope of applications. A series of problems in Biology, Medicine , Climate and Actuatrial ae addressed as well.
On the other hand, the notes balance theoretical and practical aspects of stochastic processes.
Basic concepts in Markov chains and Markov processes as reviewed, with details in the Poisson process and some of its many uses.
Special attention is put on the concept of *martingales*, a central piece which leads to the concept of risk neutral measure and provides evaluation methods of financial derivatives. Some elementary developments of stochastic calculus including Ito formula are reviewed, followed to the renowned Black-Scholes equation and beyond.
The final part of the notes are devoted to time series, where techniques and numerous examples are discussed in connection with ARMA and GARCH models.

Aaron Wan

Main Notations and Abbreviations

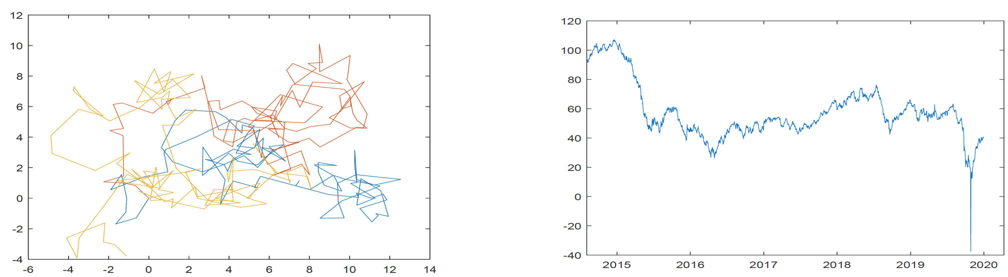
$(\Omega, \mathcal{A}, \mathcal{F}, P)$: stochastic basis,
$\mathcal{F}$: filtration,
$\mathcal{F}^{X_t} := \sigma(X_s, 0 \leq s \leq t) \subset \mathcal{F}_t$: filtration associated with the process $(X_t)_{t\geq 0}$,
$C^k(\mathbb{R})$: functions with k times differentiable with continuous k-th derivative on $\mathbb{R}$ and $[0, t]$,
$C^k(D)$ as the class of real-valued functions defined on the domain $D \subset \mathbb{R}^n$ that are k times differentiable with k-th derivative continuous,
$\hat{f}$: Fourier transform of the function f.
$a_+ = max(a, 0)$
$D_j f$: partial derivative with respect to the j-th variable
$D_t f(t, x)$: partial derivative with respect to t,
$D_x f(t, x)$: partial derivative with respect to x,
$[x]$: integer part of x
1_B: indicator function
$E(X)$: expected value
$Var(X)$: variance
$Sk(X)$: skewness
$K(X)$: Kurtosis
$cov(X, Y)$: covariance
$\rho(X, Y)$: linear correlation coefficient
p.d.f.: probability density function
p.m.f.: probability mass function
c.d.f. : cumulative distribution function
m.l.e. : maximum likelihood estimator
$X \sim F$: X has c.d.f. F
$\overline{X}$: sample mean
S^2: sample variance
S: standard deviation
$\gamma_X(h)$: autocovariance function
$\rho_X(h)$: autocorrelation function

CHAPTER 1

The concept of stochastic process

1.1. Stochastic processes: definitions and examples

Stochastic process can be generally viewed as the dynamic part of probability. The position of complex systems randomly moving upon time can be modeled using a family or collection of random variables, indexed by time or another variable. At any point in time a random variable would describe the position of the system.
One of the first phenomenon of this nature, studied in physical sciences, is the *Brownian motion*, named after the British botanist Robert Brown.
In figure 1 several simulated trajectories of a grain of pollen observed under a microscopic are shown(left). Each colour corresponds to a different trajectory. On the right, daily series of oil future prices, type Western Texas Intermediate (WTI),in US dollars per barrel, during the period 2012-2020, is pictured. What do they have in common? Notice the same erratic and irregular trajectories in both cases.



(a) (b)

FIGURE 1. Left: Trajectories of a grain of pollen under a microscope. Right: Closure Future oil prices of WTI type in NYSE, US dollars per barrel, 2014-2020

Similarly, historic Canadian dollar exchange rates versus US dollar in the period 2012-2019 is shown in figure 2. Any relevant model should capture essential features of the phenomenon under study. As examples of such complex systems we have the stock market, the dynamic of some biological populations, the spread of epidemics, economic series, e.g. unemployment and Gross National Product(GNP), and weather time series, like temperature or precipitations.

We formalize the concept of stochastic process as follows:

DEFINITION 1.1. A stochastic process is a family of random variables $(X_t)_{t\in\mathbb{T}}$ defined on a probability space $(\Omega, \mathcal{A}, P)$ taking values on a set $\mathbb{E}$.

Remark 1.2. i) The probability space is composed by a triplet. Its first element Ω is the space of elementary events, $\mathcal{A}$ is the field of random events endowed with a structure called σ-algebra, to be studied later on, and P is a probability defined on $(\Omega, \mathcal{A})$.

ii) Typically, the subscript t represents time. Hence, $\mathbb{T} = \mathbb{R}_+$, or an interval $[0, T]$ or $\mathbb{T} = \mathbb{N}$ although it may represent another variable of interest such as a geographic coordinate. We will constrain the analysis to the case of a time interpretation of the index.

iii) For a fixed element $w \in \Omega$, $X(w, t), t \in \mathbb{T}$ is a deterministic function of time called *trajectory*. In figure 3 different possible trajectories of a generic stochastic process are shown(left) together with simulated trajectories of a standard Brownian motion (right).

iv) For a fixed $t \in \mathbb{T}$, X_t is a random variable i.e. a function from the sample space Ω into a set $\mathbb{E}$ called *state space*. Typically $\mathbb{E} \subset \mathbb{N}$ or $\mathbb{E} = \mathbb{R}$. In a multidimensional setting $\mathbb{E} \subset \mathbb{N}^d$ or $\mathbb{E} = \mathbb{R}^d$.

In addition to the probability space in the definition of stochastic process we consider an increasing flow of information, described via fields of events or σ-algebra, called *filtration*.

Definition 1.3. *Filtration*
A filtration $(\mathcal{F}_t)_{t \in \mathbb{T}}$ is an increasing collection of fields of events, called also σ-algebra, whose elements

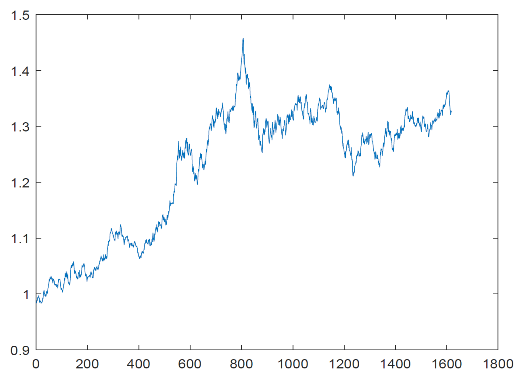
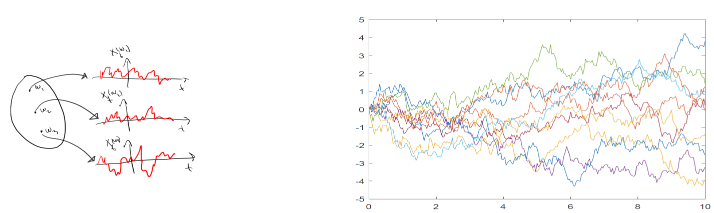


Figure 2. Can/Us dollar exchange Jan. 2012- Jan. 2019 . Source: Canforex



(a) (b)

Figure 3. Left: Possible trajectories of a stochastic process . Right: Simulated trajectories of a Brownian motion.

represent, for a fixed t, all the relevant information about the system in the period $[0, t]$ including the position of the system at time t given by X_t. A filtration must satisfy the condition $\mathcal{F}_s \subset \mathcal{F}_t$, for any $0 \leq s \leq t$.

Thus, for any $t > 0$ the σ-algebra $\mathcal{F}_t$ represents the history of the system on the interval $[0, t]$. For a discussion and examples about σ-algebra see section 3.1.
When a filtration is incorporated to the probability space the quadruplet $(\Omega, \mathcal{A}, P, (\mathcal{F}_t)_{t\in\mathbb{T}})$ is called *stochastic basis*.
In the financial markets asset prices, interest rates and other time-dependent elements are modeled through stochastic processes. The basis of the analysis is the modeling of the dynamic in prices and returns.
In general terms a modeling strategy involves the selection of a simple model, the calibration or estimation of its parameters, testing assumptions made or well-known empirical facts. It is important to note that modeling is a process. The model should capture essential features of the phenomena under study.
Once an adequate mathematical models is established it allows to look into the insight of the phenomenon under study, for example to verify empirically theoretical hypothesis, evaluate associate risks, find the price of derivative instruments, allocate investment in an optimal way according to some criterion and simulate future scenarios.
Some common empirical facts across markets, instruments and asset class, some of them still under discussion, are generally admitted. Among them:

(1) Non-correlated returns : Price returns are significantly (linearly) uncorrelated, which is an advantage in statistical empirical analysis. Their probability distribution is typically non-Gaussian. Both facts combined implies that a non-linear dependence could be in place, see for example GARCH models in chapter 5.
(2) Heavy tail presence: the tail of the probability distributions in prices and returns slowly decays to zero as a power law, with exponent larger the two, describing frequencies of large losses higher than the ones anticipated by a Gaussian distribution. It has a profound implication in risk in risk analysis.
(3) Asymmetry in gains and losses: The probability distribution of losses is skewed to the positive side, meaning that severe losses are less frequent than small ones.
(4) Cluster volatility: Variability in the market is observed for long periods.
(5) Long memory: Persistence of long term dependence.
(6) Self-similarity: Different time scale observations keep the same probability laws.

When a financial position is described by more than one asset, a vector-valued stochastic process is considered.
Stochastic processes appear in connection with biological and medical phenomena. In the study of epidemics stochastic models are intended to describe the evolution of the number of cases along time. In figure 4 number of cases in Canada from January to August 2020 are shown. Examples of stochastic processes are given above.

EXAMPLE 1.4. *White noises.*
A stochastic process $(X_t)_{t\in\mathbb{N}}$ described by a sequence of equally distributed and independent random variables, with finite expected value and variance is called a *white noise*
In particular when $X_t \sim N(0, \sigma^2)$ the process is called a *Gaussian white noise*, when $X_t \sim B(1, p)$ is a *Bernoulli* process.

Random noises are rarely found in nature, engineering, social or economic processes but they are the building blocks for more complex processes.

EXAMPLE 1.5. *Digital transmission*
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 we assume (X_n) is a two-state stochastic process, called *Bernoulli process*.

EXAMPLE 1.6. *Signals*
An acoustic signal is described by the equation:

$$X_t = A\cos(2\pi ft + \theta)$$

where the amplitude A and the phase θ are random variables.
In figure 1.6 possible trajectories of the process describing a signal are shown, when the amplitude changes at random.

EXAMPLE 1.7. *Noisy signals*

$$X_t = A\cos(2\pi ft + \theta) + \xi_t$$

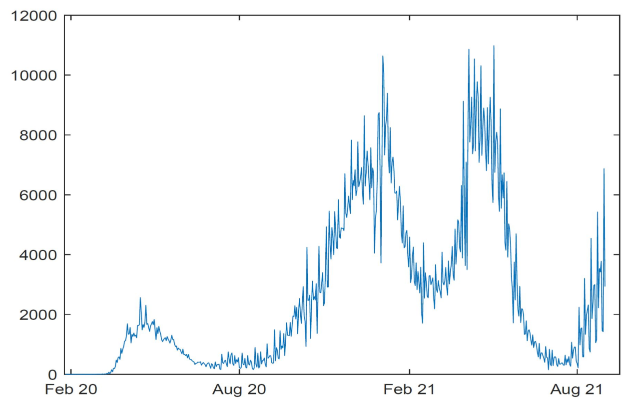
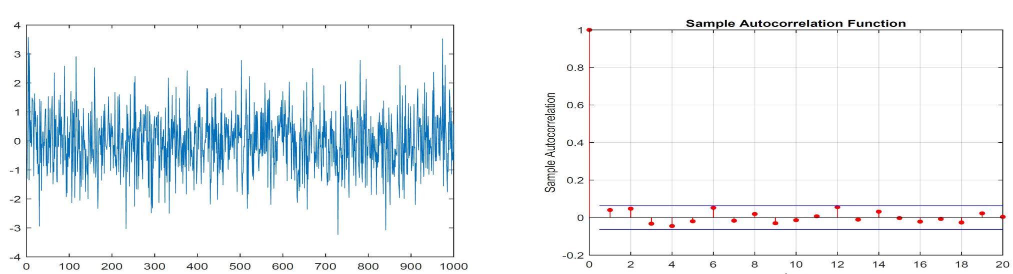


FIGURE 4. Number of new cases in Canada, Jan-Aug. 2020. Source: World Health Organization



(a) (b)

FIGURE 5

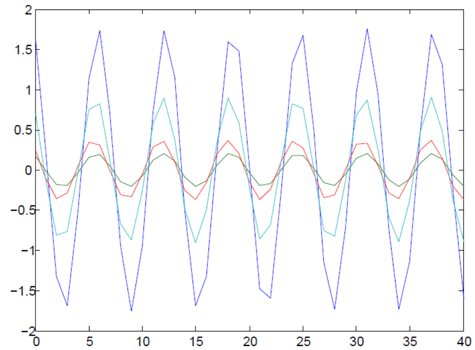


FIGURE 6. Possible trajectories of the process describing a signal are shown, when the amplitude changes at random

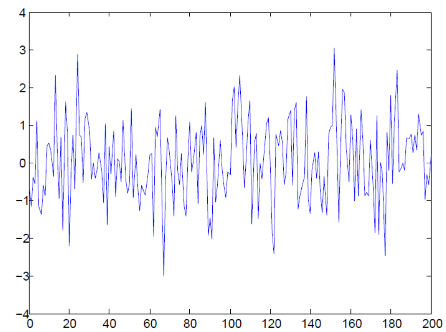


FIGURE 7. The trajectory of a signal with a Gaussian random noise

where the amplitude A and the phase θ are random variables and $(\xi_t)_{t\geq 0}$ is a white noise.
In figure 1.7 a trajectory of a deterministic signal with $A = f = 1$ and $\theta = 0$ with the addition of a Gaussian random noise is shown.

EXAMPLE 1.8. *Random Walks*
Starting at the origin a particle moves one unit to the right with probability p or one unit to the left with probability $1-p$ independently of past moves. Let X_n be the position of the particle after n moves.
$(X_n)_{n\in\mathbb{N}}$ is a stochastic processes on $\mathbb{N}$ given by:

$$X_n = \sum_{k=1}^{n} U_k$$

where $(U_k)_{k\in\mathbb{N}}$ are independent random variables such that:

$$U_k = \begin{cases} 1 & \text{with probability } p \\ -1 & \text{with probability } 1-p \end{cases}$$

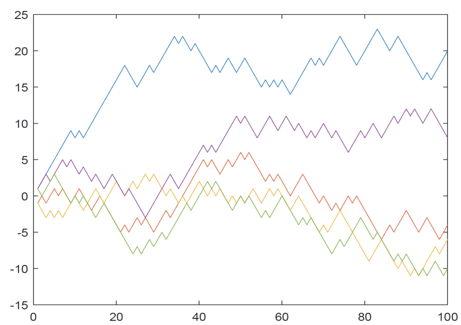


FIGURE 8. Five simulated trajectories of an unrestricted random walk starting $X_0 = 0$ with $p = 0.5$ and 100 steps

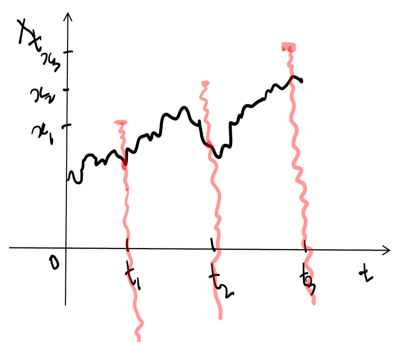


FIGURE 9. Joint events $[X_{t_1} \leq x_1]$, $[X_{t_2} \leq x_2]$ and $[X_{t_3} \leq x_3]$.

As the behavior of a random variable is described by its cumulative distribution function(c.d.f), for stochastic processes a family of c.d.f. is used to describe the joint dependence of the random variables that describe the position of the system upon time.

DEFINITION 1.9. *Family of cumulative distribution function associate to a stochastic process* Let $(X_t)_{t\in\mathbb{T}}$ be a stochastic process with values in the state space $\mathbb{E}$. The family of c.d.f. associated with the process is defined as the set of functions $\{F(t_1,\ldots,t_n;x_1,\ldots,x_n), \text{for } n\in\mathbb{N}, x_k\in\mathbb{E}, t_1 < \ldots < t_n\}$ that verifies:

$$F(t_1,\ldots,t_n;x_1,\ldots,x_n) = P\{X_{t_1}\leq x_1, X_{t_2}\leq x_2,\ldots,X_{t_n}\leq x_n\} \tag{1.1}$$

From this family, marginal c.d.f. and moments can be computed.
For any $t\in\mathbb{T}$ and $x\in\mathbb{E}$, taking $n=1$ in equation (1.1), we get the probability behavior of the position of the system at a fix time t, called marginal distributions. Thus:

$$F_{X_t}(x) = P[X_t \leq x]$$

Taking $n=2$, we can compute the joint probability of the positions of the system at two different time instants by:

$$F_{X_t,X_s}(x,y) = P[X_t\leq x, X_s\leq y]$$

Time	$\mathbb{T} \subset \mathbb{N}$	$\mathbb{T} \subset \mathbb{R}_+$
State space		
$\mathbb{E} \subset \mathbb{N}$	Discrete-valued Discrete-time	Discrete-valued Continuous-time
$\mathbb{E} \subset \mathbb{R}$	Continuous-valued Discrete-time	Continuous-valued Continuous-time

This last function allows to analyze the dependence between the positions at times $s > 0$ and $t > 0$. From the family of joint distribution functions we can obtain the moments of the process, whenever they are finite, as it is shown below in the case of first and second order moments:

i) Expected value:
$$\mu_X(t) = E[X_t]$$
ii) Variance:
$$\sigma_X^2(t) = Var(X_t)$$
iii) Autocovariance function:
$$\gamma_X(s,t) = E[(X_t - E(X_t))(X_s - E(X_s))]$$
iv) Autocorrelation function:
$$\rho_X(s,t) = \frac{\gamma_X(s,t)}{\sigma_{X_t}\sigma_{X_s}}$$

Higher order moments can be defined in a similar way. Notice that now, as opposed to the case of random variables, the moments of a stochastic process are not ocnstants but deterministic function of time. We recall the first moment μ_{X_t} offers information about the average position of the system, while the variance tells us about its variability. The autocovariance and autocorrelation functions are measures of the linear dependence between positions at two different time instants.
In general, only for a few cases of stochastic processes the family of joint probabilities is known.

1.2. Classes of stochastic processes

A general classification of stochastic processes, accordingly to the time index and state space is given in table 1.2.

The class of stationary processes has the property that its probability laws are invariant upon time.

Definition 1.10. *Stationary process (strict sense)*
A stochastic process $(X_t)_{t\in\mathbb{T}}$ defined on a probability space $(\Omega, \mathcal{A}, P)$ is stationary in the strict sense if:

$$F(t_1, \ldots, t_n; x_1, \ldots, x_n) = F(t_1 + T, \ldots, t_n + T; x_1, \ldots, x_n) \tag{1.2}$$

for $n \in \mathbb{N}, x_k \in \mathbb{E}, t_1 < \ldots < t_n$, $T \in \mathbb{T}$

The stationarity condition (1.2) means that the probability laws of the process do not change over time. In particular, it implies that the mean and variance of the process do not depend on t. On the other hand, the autocovariance function depends only on the difference between the final and initial time. To verify both statements take $n = 1$ and $n = 2$ in equation (1.2) and look at the definition of mean and covariance.
Stationary in the sense defined above may be too strict for practical purposes, few processes have this property. An alternative way to view the stationarity condition is to consider only two consequences of this property about the mean and the autocovariance.
Therefore we have:

DEFINITION 1.11. *Stationary process (wide sense)*
A stochastic process $(X_t)_{t\in\mathbb{T}}$ defined on a probability space $(\Omega, \mathcal{A}, P)$ is stationary in the wide sense if:

a) $\mu_{X_t} = m$ is constant.
b) $\gamma_X(t, t+h)$ is a function of h, denoted $\gamma_X(h)$.

EXAMPLE 1.12. *First-order Autoregressive process(AR(1))*
A stationary process that satisfies:

$$X_t = \theta X_{t-1} + Z_t \tag{1.3}$$

where $(Z_t)_{t\in\mathbb{N}}$ is a white noise with variance σ^2, uncorrelated with $(X_t)_{t\in\mathbb{N}}$ and $|\theta| < 1$ is an unknown parameter is called autoregressive process.
Taking expectations on both sides in the equation above we see that the first moment verifies $\mu = E(X_t) = \theta E(X_{t-1}) = \theta\mu$, which implies that $\mu = 0$ for $\theta \neq 0$. On the other hand, if $\theta = 0$ the process is a white noise and $\mu = 0$.
On the other hand, for $h > 0$ (for h negative a similar analysis applies):
$X_t X_{t+h} = \theta X_t X_{t+h-1} + Z_{t+h} X_t$
Taking expectations on both sides of the equation above:

$$\begin{aligned}
\gamma_X(h) = E(X_t X_{t+h}) &= E(\theta X_t X_{t+h-1}) + E(X_t Z_{t+h}) \\
&= \theta cov(X_t, X_{t+h-1}) + cov(Z_{t+h}, X_t) \\
&= \theta\gamma_X(h-1) = \ldots = \theta^h \gamma_X(0)
\end{aligned}$$

Also, for $h = 0$:

$$\gamma_X(0) = Var(X_t) = \theta^2 Var(X_{t-1}) + Var(Z_t) = \theta^2\gamma_X(0) + \sigma^2$$

Solving for $\gamma_X(0)$ we get:

$$\gamma_X(0) = \frac{\sigma^2}{1-\theta^2}$$

In summary:

$$\gamma_X(h) = \begin{cases} \frac{\sigma^2}{(1-\theta^2)} & h = 0 \\ \theta^{|h|}\frac{\sigma^2}{(1-\theta^2)} & h \neq 0 \end{cases} \tag{1.4}$$

Notice that the condition $|\theta| < 1$ implies $\gamma_X(h) \to 0$ when $h \to +\infty$, meaning that the linear dependence between two positions of the system becomes very weak for separate instants in time.

Another important, yet general, class is the one of stochastic processes whose increments are independent and remain stationary in non-overlapping time intervals.

DEFINITION 1.13. *Processes with independent and stationary increments*

i) A process $(X_t)_{t\in\mathbb{T}}$ has independent increments is for any $n \in \mathbb{T}$, $t_0 = 0 < t_1 < t_2 < \ldots t_n$ the random variables $X_{t_k} - X_{t_{k-1}}$, $k = 1, 2, \ldots, n$ are independent.
ii) A process $(X_t)_{t\in\mathbb{T}}$ has stationary increments is for any $t, h \in \mathbb{T}$, the probability distribution of $X_{t+h} - X_t$ does not depends on t.

Counting random event occurrences upon time, such as accidents, earthquakes, radioactive decay are important phenomena. They are studied by counting processes.

DEFINITION 1.14. *Counting process*
A counting process $(N_t)_{t\geq 0}$ is a process with $N_0 = 0$, taking values in $\mathbb{N}$ whose trajectories have only jumps of size one and remain constant between jumps.
The value N_t is interpreted as the number of events occurred on the interval $[0, t]$.

The best known counting process is the Poisson process.

Definition 1.15. *Poisson processes*
A process counting process $(N_t)_{t\in\mathbb{N}}$ with values in $\mathbb{N}$ is a Poisson process if:

p1) $N_0 = 0$.
p2) $(N_t)_{t\in\mathbb{N}}$ has stationary and independent increments.
p3) For all $t > s > 0$ we have $N_t - N_s \sim \mathcal{P}oiss(\lambda(t-s))$, where $\mathcal{P}oiss(\lambda t)$ is the Poisson probability law with parameter λt.

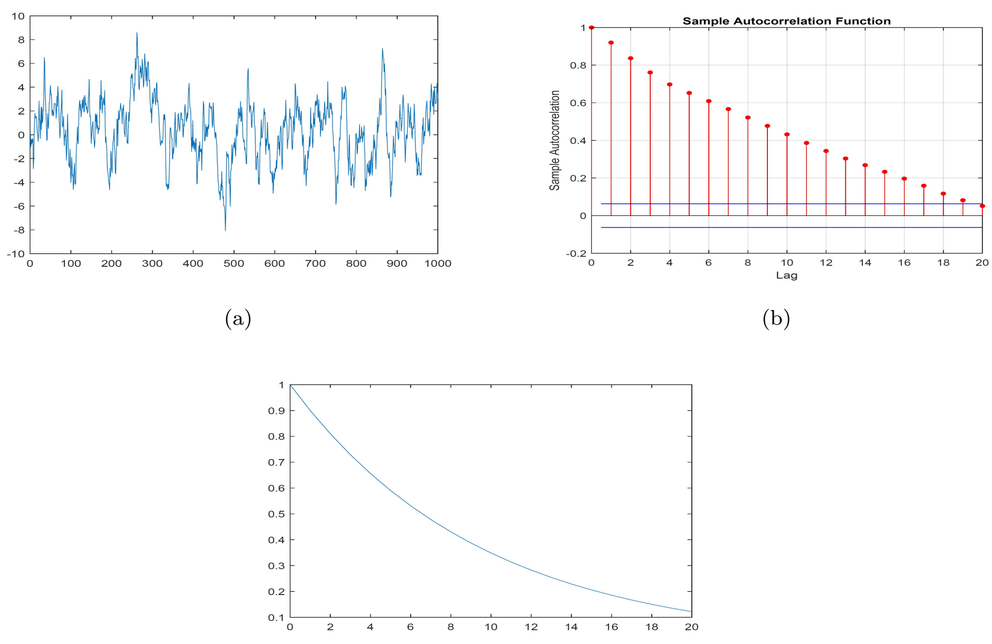
One of the most important stochastic processes is the one describing the Brownian motion introduced above. Formally it is defined as shown in the next statement.

Definition 1.16. *Brownian Motion*
A stochastic process $(B_t)_{t\geq 0}$ such that $B_0 = 0$, is a *Brownian Motion* or *Wiener process* if:

i) $(B_t)_{t\geq 0}$ has stationary and independent increments.
ii) $B_t - B_s \sim N(0, \sigma(t-s))$.

It is called a standard Brownian motion if, additionally, $\sigma = 1$. Unless explicitly said the Brownian motion will be assumed standard.

Here, B_t represents the position of the particle at time t. A more detailed study can be found in chapter 4.



(a) (b)

(c)

Figure 10. Clockwise starting from the upper left quadrant: simulated autoregressive process AR(1) with $\phi = 0.9$, empirical autocorrelation function and theoretical correlation function.

EXAMPLE 1.17. *Autocovariance function of a Brownian motion*
Compute the mean, variance, autocovariance and autocorrelation functions of a Brownian motion.
Solution
For $s < t$:

$$\begin{aligned} \gamma_B(s,t) &= E[B_t B_s] = E[(B_t - B_s + B_s)B_s] \\ &= E[(B_t - B_s)B_s] + E[B_s^2] \\ &= E[B_t - B_s]E[B_s] + s = s \end{aligned}$$

Interchanging the roles of s and t we have:

$$\gamma_B(s,t) = min(s,t)$$

1.3. Probability distributions in finance

Most common multivariate probability distribution laws are presented in this section.

1.3.1. Multivariate normal distribution. *Gaussian distribution*
The Gaussian distribution can be generalized to several dimensions.

DEFINITION 1.18. *Multivariate Normal Distribution*
A random vector $X = (X_1, X_2, \ldots, X_n)$ has a multivariate normal distribution with parameters μ and Σ if its p.d.f. is given by:

$$f(x) = \frac{1}{2^{\frac{n}{2}} det(\Sigma)} exp(-(x-\mu)\Sigma^{-1}(x-\mu)^T, \text{ for } x \in \mathbb{R}^n \tag{1.5}$$

where $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ and Σ is a $n \times n$ symmetric positive semi-definite matrix and $det(\Sigma)$ its determinant. The matrix Σ has full rank.

The characteristic function of a multivariate Gaussian distribution random variable X is:

$$\varphi_X(u) = exp(i\mu u^T - \frac{1}{2} u \Sigma u^T), \quad u \in \mathbb{R}^d \tag{1.6}$$

The notation for a multivariate normal distribution is $X \sim N_d(\mu, \Sigma)$.
The following results holds:

THEOREM 1.19. *The following three proposition are equivalent:*

i) *$X \sim N_d(\mu, \Sigma)$.*
ii) *X has characteristic function given by equation (1.6).*
iii) *For any $a \in \mathbb{R}^d$ the random variable aX^T has a normal multivariate distribution with $E(X) = \mu$ and $Var(X) = a\Sigma a^T$.*

REMARK 1.20. As a consequence of the theorem the components of a multivariate normal vector have normal distribution and the sum of normal random variables is normally distributed. Moreover, conditional distributions on the components is also multivariate normal distribution.

1.3.2. The t-student distribution. A t-student with ν degrees of freedom law has probability density function given by:

$$f(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\Gamma(\nu/2)} \frac{1}{(1 + x^2/\nu)^{(\nu+1)/2}} \tag{1.7}$$

where the special function Γ is defined as:

$$\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x} dx$$

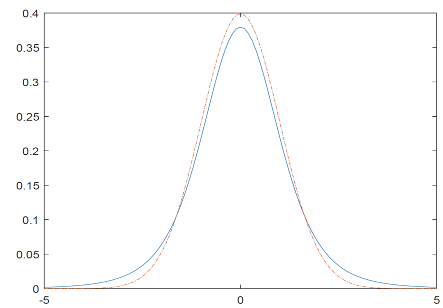


FIGURE 11. P.d. f. of standard normal and t-student with $\nu = 6$.

The notation for a t-student distribution law with ν degrees of freedom is $t \sim t(\nu)$.
Its expected value is $E(t(\nu)) = 0, \nu > 1$ and its variance $Var(t(\nu)) = \frac{\nu}{\nu-2}$ for $\nu > 2$, otherwise the variance is infinity.
The t-student distribution arises in sampling problems. Two related results are:

PROPOSITION 1.21. Let the random variables $Z \sim N(0,1)$ and $\chi^2 \sim \chi^2(\nu)$ be independent. Then:

i)
$$t = \frac{Z}{\sqrt{\frac{\chi^2}{\nu}}} \sim t(\nu).$$

ii) For a random sample $X_1, X_2, \ldots, X_n$ with a Gaussian distribution the random variables $\overline{X}$ and S^2 are independent. Moreover the random variable:
$$t_S = \frac{\overline{X}}{S} \sim t(\nu - 1).$$

The second result is used for small samples.

DEFINITION 1.22. *located and scaled t-student*
A random variable T_S has a scaled and located t-student distribution with ν degrees of freedom if $\frac{T_S - \mu}{\sigma}$ follows a t-student distribution with ν degrees of freedom, where μ is a location parameter and $\sigma > 0$ is a scaling parameter. The notation in this case is $T_S \sim t(\nu, \mu, \sigma)$.

The t-student distribution can be generalized to multiple dimensions as well.

DEFINITION 1.23. A random variable T_S has a multivariate scaled and located t-student distribution with ν degrees of freedom, location parameter μ and scale prameter Σ if it has p.d.f.:

$$f(x) = \frac{\Gamma(\frac{\nu+d}{2})}{(\pi n)^{\frac{d}{2}} \nu^{\frac{d}{2}} \Gamma(\frac{\nu}{2}) det^{\frac{1}{2}}(\Sigma)} \frac{1}{(1 + (x-\mu)\Sigma^{-1}(x-\mu)^T/\nu)^{\frac{\nu+d}{2}}} \tag{1.8}$$

1.3.3. Log-normal distributions. As prices are positive quantities, the Gaussian distribution can not directly be used to model them. On the other hand, exponential Gaussian or *log-normal* random variable offer the possibility to model positive random magnitudes. They arise in the framework of the Black-Scholes model. The definition is the following:

DEFINITION 1.24. *Log-normal distributions*
A random variable Y follows a log-normal distribution with parameters μ and σ^2 if $Y = e^X$ where $X \sim N(\mu, \sigma^2)$ or equivalently if $\log(Y) \sim N(\mu, \sigma^2)$.

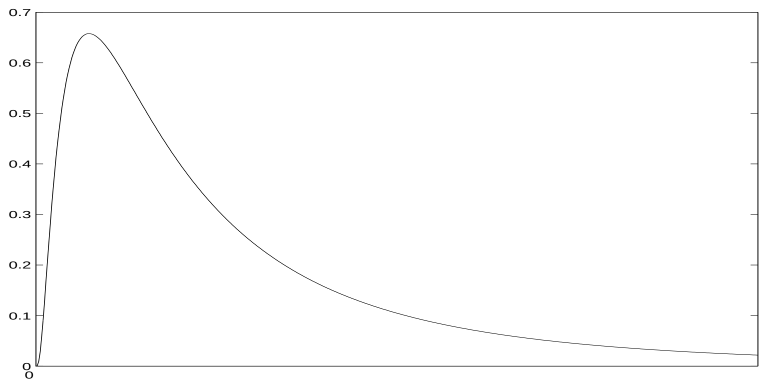


FIGURE 12. Log-normal density with parameters $\mu = 0$ and $\sigma = 1$

Its p.d.f. is given by:

$$f_Y(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2}},\ x > 0$$

The moments and associated function as described in table 1.
Its skewness is:

$$Sk_X = (exp(\sigma^2) + 2)(exp(\sigma^2) - 1)^{\frac{1}{2}} \tag{1.9}$$

and kurtosis:

$$K_X = exp(4\sigma^2) + 2exp(3\sigma^2) + 3exp(2\sigma^2) \tag{1.10}$$

In the figure 12 the graph of a log-normal p.d.f. with parameters $\mu = 0$ and $\sigma = 1$ is shown.

REMARK 1.25. 1) The distribution of a log-normal random variable is right-skewed, the larger σ the more pronounced is the skewness.
2) To estimate and test the parameters μ and σ^2 is a log-normal distribution it is enough to take the logarithm of the original data and apply the corresponding estimation of Gaussian data.
3) The kurtosis of a log-normal distribution verifies $K \geq 7e - 6 \geq 3$. Hence the log-normal distribution has tails heavier than the Gaussian distribution.

1.3.4. Mixing laws. Let $Y = (X, Z)$ be a pair of random variables, such that Z is a discrete random variable with p.m.f.:

$$P(Z = j) = p_j, \text{for } j = 1, 2, \ldots, k$$

Furthermore assume that conditionally to $[Z = j]$, the random variable X has a p.d.f. $f_{X/Z_j}(x)$. Then, the p.d.f. of X is given by:

$$f_X(x) = \sum_{j=1}^{k} p_j f_{X/Z_j}(x)$$

Usually Z is not observable and p_j are unknown parameters. Special likelihood methods such as the *Expectation-Maximization* algorithm, are needed.

In the particular case that the conditional distribution of X given Z are normally distributed with parameters μ_Z and σ_Z is said that X follows a *mixing Gaussian law*. The p.d.f. of X is:

$$f_X(x) = \sum_{j=1}^{k} p_j \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}} \tag{1.11}$$

Its expected value and variance are:

$$\begin{aligned} E(X) &= \sum_{j=1}^{m} p_j \mu_i \\ V(X) &= \sum_{j=1}^{m} p_j \sigma_j^2 \end{aligned}$$

In figure 13 a typical mixing of two distributions is shown. The resulting distribution is bimodal. The general definition of a multivariate mixing Gaussian probability distribution is given below.

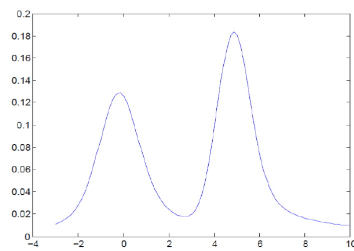


FIGURE 13. Mixing of two pdf

DEFINITION 1.26. *Multivariate Gaussian mixing*
A random vector X has a multivariate Gaussian mixing distribution if:

$$X = \mu + \sqrt{W} A Z \tag{1.12}$$

where Z is a d-dimensional vector with independent components having a standard Gaussian distribution. Hence $Z \sim N_k(0, I_k)$, $W \geq 0$ is a random variable independent of Z and μ and A are respectively deterministic $d \times 1$ and $d \times k$ matrices.

Notice that conditionally on W the random vector X has a $N_d(\mu, W\Sigma)$ distribution with $\Sigma = AA^T$.
The next proposition relates the multivariate Gaussian mixing with a multivariate t-student distribution.

PROPOSITION 1.27. If $W = \frac{\nu}{\chi^2(\nu)}$, where $\chi^2(\nu)$ is a chi-square distribution with ν degrees of freedom then $X = \mu + \sqrt{W} A Z$ has a $t(\nu, \mu, \Sigma)$

The proposition above suggests an algorithm to simulate scaled and located multivariate t-student distributions with parameters μ, Σ and ν.
Algorithm 1.1

1) Generate a chi-square random variable with ν degrees of freedom.
2) Compute $W = \frac{\nu}{\chi^2(\nu)}$.
3) Obtain the Cholesky decomposition $\Sigma = AA^T$.
3) Compute $X = \mu + \sqrt{W}AZ$.

1.3.5. Stable Laws. In the decade of 1960 by Mandelbrot and Fama proposed the α-stable distribution law to model financial data.
The family of stable distributions not only captures heavy tails and asymmetric behavior but also, the dependency on four parameters, instead for two parameters in the Gaussian case, make them more flexible to adapt empirical data for estimation and model testing proposes.
Another nice property is that they have domain of attraction, i.e., limits of sums of i.i.d. random variables have also a stable distribution after a suitable re-normalization, making them robust models, then slight modifications will produce a distribution law that still lies in the family.

Definition 1.28. A random variable X follows a stable law if its characteristic function admits the form:

$$\phi_X u = \begin{cases} \exp(-\sigma^\alpha |u|^\alpha (1 - i\beta sgn(t) \tan \pi\alpha 2) + i\mu t) & \text{for } \alpha \neq 1 \\ \exp(-\sigma|u|(1 + i\beta \frac{\pi}{2} sgn(t) \log|u|) + i\mu u) & \text{for } \alpha = 1 \end{cases} \tag{1.13}$$

where the parameters satisfy the constrains $\alpha \in (0, 2]$,$\sigma \in \mathbb{R}_0^+$, $\beta \in [-1, 1]$,$\mu \in \mathbb{R}$.

A random variable X with a stable distribution is denoted $X \sim S(\alpha, \beta, \sigma, \mu)$.

Theorem 1.29. *The following properties hold for an stable distribution:*

i) *For $\alpha \in (0, 2)$ and $X \sim S(\alpha, \beta, \sigma, 0)$ then:*

$$\lim_{x\to\infty} x^\alpha P(X > x) = C_\alpha (1 + \beta)\sigma^\alpha \tag{1.14}$$

where C_α is a constant depending on α.

ii) *For $\alpha \in (0, 2)$ and $X \sim S(\alpha, \beta, \sigma, 0)$ then $E|X|^p$ exists for $0 < p < \alpha$ and $E|X|^p = \infty$ for $p \geq \alpha$.*

The value α as a tail parameter, β as a coefficient of skewness, σ as an scale parameter and μ as a location parameter.
Properties i) and ii) express the heavy tail behavior .
Notice also that Gaussian law is included in the family with $\alpha = 2$.
In figure 14 it is shown the empirical density obtained by simulation and the one obtained by inverting the characteristic function. The p.d.f. of s stable law is not explicitly know. It can be numerically computed from the characteristic function by inversion Fourier formula. Namely:

$$f_X(x) \quad = \quad \frac{1}{2\pi} \int_0^{+\infty} e^{-ixu} \varphi_X(u) du$$

To create a stable probability in MATLAB use the command:
$pd = makedist('Stable', 'alpha', 0.5, 'beta', 0, 'gam', 1, 'delta', 0)$.

1.3.6. Generalized hyperbolic distributions. Hyperbolic and related classes have been first considered in connection with grain size distribution. Applications to a financial context came in the decade of 1990s. We present a summary with their main characteristics:

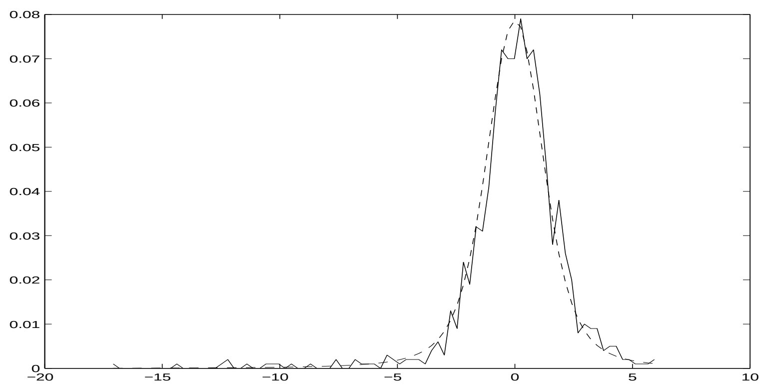


FIGURE 14. density stable law with $\alpha = 1.5$ (discontinuous line)

a) *Generalized hyperbolic law.*
Density function:

$$f(x) = \frac{(\frac{\gamma}{\delta})^{\lambda}}{\sqrt{2\pi}K_{\lambda}(\delta\gamma)} \frac{K_{\lambda-\frac{1}{2}}(-\alpha\sqrt{\delta^2+(x-\mu)^2})}{(\frac{\sqrt{\delta^2+(x-\mu)^2}}{\alpha})^{\lambda-\frac{1}{2}}}. \exp\beta(x-\mu) \tag{1.15}$$

where $\gamma^2 = \alpha^2 - \beta^2$,$\alpha > 0,\ \delta > 0, \alpha^2 > \beta^2, \beta \in \mathbb{R}, \lambda \in \mathbb{R}$. Here K_λ is the Bessel function of third kind given by:

$$K_{\lambda}(x) = \frac{1}{2}\int_0^{+\infty} y^{\lambda-1}e^{-\frac{1}{2}x(y+y^{-1})}dy$$

The mean and the variance are respectively:

$$\begin{aligned} E(X) &= \mu + \frac{\delta\beta K_{\lambda+1}(\delta\gamma)}{\gamma K_{\lambda}(\delta\gamma)} \\ Var(X) &= \frac{\delta\beta K_{\lambda+1}(\delta\gamma)}{\gamma K_{\lambda}(\delta\gamma)} + \frac{\beta^2\delta^2}{\gamma^2}\left(\frac{K_{\lambda+2}(\delta\gamma)}{K_{\lambda}(\delta\gamma)} - \frac{K^2_{\lambda+1}(\delta\gamma)}{K^2_{\lambda}(\delta\gamma)}\right) \end{aligned}$$

b) *Hyperbolic law.*It is a particular case of the density above with $\lambda = 1$.Its density function is:

$$f(x) = \frac{\gamma}{2\alpha\delta K_{\lambda}(\delta\gamma)}\exp(-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu))$$

c) *Normal inverse Gaussian law.* It is also a particular of the Generalized Hyperbolic with $\lambda = -\frac{1}{2}$. Its density is given by:

$$f(x) = \frac{\alpha\delta}{\pi}e^{\delta\gamma}.\frac{K_1(\alpha\sqrt{\delta^2+(x-\mu)^2}}{\sqrt{\delta^2+(x-\mu)^2}}.e^{\beta(x-\mu)},\ x\in\mathbb{R}$$

The mean and variance are:

$$\begin{aligned} E(X) &= \mu + \frac{\delta\beta}{\gamma} \\ Var(X) &= \frac{\delta\alpha^2}{\gamma^3} \end{aligned}$$

d) *Variance-Gammma law(VG)*
Density function is obtained by making $\delta = 0$ in 1.15 as

$$f(x) = \frac{\gamma^{2\lambda}}{\sqrt{\pi}\Gamma(\lambda)(2\alpha)^{\lambda-\frac{1}{2}}} \mid x-\mu \mid^{\lambda-\frac{1}{2}} K_{\lambda-\frac{1}{2}}(\alpha \mid x-\mu \mid) e^{\beta(x-\mu)}$$

with $\alpha >\mid \beta \mid$.
Its mean and variance are:

$$\begin{aligned} E(X) &= \mu + \frac{2\lambda\beta}{\gamma^2} \\ Var(X) &= \frac{2\lambda\beta}{\gamma^2}(1+2(\frac{\beta}{\gamma})^2) \end{aligned}$$

Another class considered is the *Generalized Inverse Gaussian law*(GIG) with density:

$$f(x) = \frac{(\frac{\gamma}{\delta})^\lambda}{2K_\lambda(\delta\gamma)} x^{\lambda-1} \exp\{-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\}, \text{ for} x > 0$$

with $\delta > 0, \gamma > 0, \text{and} \lambda \in \mathbb{R}$.
Note that *Gamma* law is obtained as a limit procedure.
Its moments are:

$$EX^j = (\frac{\delta}{\gamma})^j \frac{K_{\lambda+j}(\delta\gamma)}{K_\lambda(\delta\gamma)}, \text{ for } j = 1, 2, \cdots$$

Distribution	probability/density	Expected Value	Variance
log-normal	$\frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x-\mu)^2}{2}}, \ x > 0$	$exp(\mu + \frac{\sigma^2}{2})$	$(exp(\sigma^2) - 1)exp(2\mu + \sigma^2)$
Gaussian mixing	$\sum_{j=1}^m p_j \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}}$	$\sum_{j=1}^m p_j \mu_i$	$\sum_{j=1}^m p_j \sigma_j^2$
Stable	$\frac{1}{2\pi} \int_0^{+\infty} e^{-ixu} \varphi_X(u) du$	$\mu, \alpha > 1$	d.n.e.
Generalized hyperbolic	See 1.15	$\mu + \frac{\delta\beta K_{\lambda+1}(\delta\gamma)}{\gamma K_\lambda(\delta\gamma)}$	$\frac{\delta\beta K_{\lambda+1}(\delta\gamma)}{\gamma K_\lambda(\delta\gamma)} + \frac{\beta^2\delta^2}{\gamma^2}\left(\frac{K_{\lambda+2}(\delta\gamma)}{K_\lambda(\delta\gamma)} - \frac{K^2_{\lambda+1}(\delta\gamma)}{K^2_\lambda(\delta\gamma)}\right)$

TABLE 1. A list of some probability distribution used in Finance

1.4. Solved problems

(1) For a Bernoulli process $(X_t)_{t\in\mathbb{N}}$ with $0 < p < 1$
  a) Find the first order moment m_{X_t} of the process.
  b) Find the second order moment $\sigma^2_{X_t}$ of the process.
  c) Find the autocovariance and autocorrelation functions $\gamma(s,t)$ and $\rho(s,t)$.
  d) Is the signal $(X_t)_{t\in\mathbb{N}}$ a stationary process in the strick and wide sense? Explain.

  Solution
  a) $m_{X_t} = E(X_t) = p$ (expected value of a binomial random variable).
  b) $\sigma^2_{X_t} = p(1-p)$(variance of a binomial random variable) .
  c) $\gamma(s,t) = \rho(s,t) = 0$ because of the independence.
  d) Yes, it is stationary in both wide and strict sense. The expected value and the autocovariance do not depend on t. The probability laws are time-invariant.

(2) Let an acoustic signal be described by the equation:

$$X_t = A\cos(wt + \theta),\ t \in \mathbb{N}$$

where A is a random variable with zero mean and unit variance, θ is a random variable uniformly distributed on the interval $]-\pi, \pi]$, both independent.

a) Find the first order moment m_{X_t} of the process.
b) Find the second order moment $\sigma^2_{X_t}$ of the process.
c) Find the autocovariance and autocorrelation functions $\gamma(s,t)$ and $\rho(s,t)$.
d) Is the signal $(X_t)_{t\in\mathbb{N}}$ a stationary process in the strict and wide sense? Explain.

Solution

a) $m_{X_t} = E(X_t) = E(A)E(\cos(wt+\theta)) = 0$.
b)

$$\begin{aligned}
\sigma^2_{X_t} &= E(X_t^2) = E(A^2)E(\cos^2(wt+\theta)) \\
&= E(\cos^2(wt+\theta)) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\cos^2(wt+x)\,dx \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\cos(2wt+2x)+1}{2}\,dx \text{ double angle trig. identity} \\
&= \frac{1}{8\pi}\sin y|_{2wt-2\pi}^{2wt+2\pi} + \frac{1}{2} = \frac{1}{2}
\end{aligned}$$

c) Take $h > 0$:

$$\begin{aligned}
\gamma(t,t+h) &= E(X_tX_{t+h}) = E(A^2)E(\cos(wt+\theta)\cos(w(t+h)+\theta)) \\
&= E(\cos(wt+\theta)\cos(w(t+h)+\theta)) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\cos(wt+x)\cos(w(t+h)+x)\,dx \\
&= \frac{1}{4\pi}\int_{-\pi}^{\pi}[\cos(wh)+\cos(w(t+h)+x)]\,dx \\
&= \frac{1}{8\pi}\sin y|_{2wt-2\pi}^{2wt+2\pi} + \frac{1}{2} = \frac{1}{2}\cos(wh)
\end{aligned}$$

We have used the trigonometric identity $cosAcosB = \frac{1}{2}[cos(A+B)+cos(A-B)]$.

$$\gamma(t,t+h) = \frac{\gamma(t,t+h)}{\sigma_{X_t}\sigma_{X_{t+h}}} = \cos(wh)$$

d) Neither the expected value nor the autocovariance function depend on t, therefore it is a process stationary in the wide sense. On the other hand the c.d.f. of X_t depends on t, then it is not stationary in the strict sense.

(3) The pair of random variables (X,Y) denotes the number of defects in two disjoint parts of a composite material. Its joint probability function is given in the following table:

$p_{X,Y}$	$y=1$	$y=2$
$x=1$	$\frac{1}{6}$	$\frac{5}{18}$
$x=2$	$\frac{2}{9}$	$\frac{1}{3}$

a) Find the probability $P[X=1, Y>0.5]$.
b) Find the probability mass function of the random variable $W = X+2Y$.
c) Find the conditional probability mass function at $x=1$ conditional on $[Y=1]$, i.e. $p_{X/Y}(1/1)$.

d) Find the conditional expected value $E[X/Y=1]$.
e) Are the random variables X and Y independent? Explain.
f) Find the expected values of X and Y. Find the covariance $C(X,Y)$.
g) Find the standard deviations of X and Y and the linear correlation coefficient $\rho_{X,Y}$.
h) Calculate $Var(X+2Y)$.

Solution

a) $P[X=1,Y>0.5]=p_{X,Y}(1,1)+p_{X,Y}(1,2)=\frac{1}{6}+\frac{5}{18}=\frac{8}{18}$.
b) Possible values: 3,4,5,6
$p_W(3)=p_{X,Y}(1,1)=\frac{1}{6}$, $p_W(4)=p_{X,Y}(2,1)=\frac{2}{9}$, $p_W(5)=p_{X,Y}(1,2)=\frac{5}{18}$
$p_W(6)=p_{X,Y}(2,2)=\frac{1}{3}$.
c) $p_{X/Y}(1/1)=\frac{p_{X,Y}(1,1)}{p_Y(1)}=\frac{3}{7}$.
d) $E[X/Y=1]=1\times\frac{3}{7}+2\times\frac{4}{7}=\frac{11}{7}$.
e) No, for example $p_{X,Y}(1,1)=\frac{1}{6}\neq p_X(1)p_Y(1)$.
f) $E(X)=1\times\frac{8}{18}+2\times\frac{10}{18}=\frac{14}{9}$
$E(Y)=1\times\frac{7}{18}+2\times\frac{11}{18}=\frac{29}{18}$.
$E(XY)=1\times1\times\frac{1}{6}+1\times2\times\frac{5}{18}+2\times1\times\frac{2}{9}+2\times2\times\frac{1}{3}=\frac{45}{18}$
$Cov(X,Y)=\frac{45}{18}-\frac{14}{9}\frac{29}{18}=-\frac{1}{162}$.
g) $\rho_{X,Y}=\frac{Cov(X,Y)}{\sigma_x\sigma_Y}=\frac{-\frac{1}{162}}{\sqrt{\frac{20}{81}\frac{77}{324}}}=-0.025$.
h) Calculate $Var(X+2Y)=Var(X)+4Var(Y)+4Cov(X,Y)=\frac{20}{81}+4\frac{77}{324}-\frac{4}{162}=1.17$.

(4) *Triangular uniform distribution*
Let (X,Y) a pair of random variables with joint density function:

$$f_{X,Y}(x,y)=\begin{cases}2 & x,y\geq 0, x+y\leq 1\\ 0 & otherwise\end{cases}$$

a) Find the margin probability density functions of the random variables X and Y.
b) Find the conditional probability density function $f_{X/Y}(x/y)$.
c) Find the conditional expected value $E(X/Y)$.
d) Are the random variables X and Y independent?

Solution

a) $f_X(x)=\int_{\mathbb{R}}f_{X,Y}(x,y)dy=\int_0^{1-x}2dy=2(1-x)$, for $0\leq x<1$ and zero otherwise. Similarly $f_Y(y)=2(1-y)$ for $0\leq y\leq 1$ and zero otherwise.

b) $f_{X/Y}(x/y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}=\frac{1}{1-y}$, for $x,y\geq 0, x+y<1$ and zero otherwise.
c) $E(X/Y)=\int_{\mathbb{R}}xf_{X/Y}(x/y)dx=\int_0^{1-y}x\frac{1}{1-y}\,dx=\frac{1}{2(1-y)}$ for $y\geq 0\leq y<1$ and zero otherwise.
d) $f_{X,Y}(x,y)=2\neq f_X(x)f_Y(y)=\frac{1}{(1-x)(1-y)}$.

(5) *Monty Hall* Named after the host of the American television show "Let's make a deal" At the end of a game you have to choose between three closed doors. Two of them have nothing behind, and one contains a prize. You chose one door, without knowing what is behind the door the presenter opens another door that contains nothing. He gives you the choice of changing the door or sticking with the initial choice.
Solution: The key in any such problem is the sample space which has to be complete enough to be able to answer the questions asked. Let D_i be the event that the prize is behind door i. Let S_W be the event that switching wins the prize. It does not matter which door we chose initially. The reasoning is identical with all three doors. So, we assume that

initially we pick door 1. Events $D_i i = 1, 2, 3$ are mutually exclusive, and we can write

$$P(SW) = P(SW|D_1)P(D_1) + P(SW|D_2)P(D_2) + P(SW|D_3)P(D_3)$$

When the prize is behind door 1, since we chose door 1, the presenter has two choices for the door to show us. However, neither would contain the prize, and in either case switching does not result in winning the prize, therefore P(SW—D1) = 0. If the prize is behind door 2, since our choice is door 1, the presenter has no alternative but to show us the other door (3) which contains nothing. Thus, switching in this case results in winning the price. The same reasoning works if the prize is behind door 3. Therefore:

$$P(SW) = 1.1/3 + 1.1/3 + 0.1/3 = 2/3$$

Thus switching has a higher probability of winning than not switching.

(6) Let W be a Gaussian random variable with expected value $\mu = 0$ and $\sigma^2 = 16$. Given the event $C = [W > 0]$,
  a) What is the conditional p.d.f. $f_{W/C}(x)$?
  b) Find the conditional expected value $E[W/C]$.
  c) Find the conditional variance $Var[W/C]$.

<u>Solution</u>

  a) First find $P(C) = P(W > 0) = \Phi(0) = \frac{1}{2}$(alternatively notice that the p.d.f. is symmetric then the areas under this curve for negative and positive values are equals and equal to one half).
  Then we have that the p.d.f. of W is:

$$f_W(x) = \frac{1}{\sqrt{32\pi}} e^{-\frac{x^2}{32}}$$

  Therefore:

$$f_{W/C}(x) = \begin{cases} \frac{2}{\sqrt{32\pi}} e^{-\frac{x^2}{32}} & x \geq 0 \\ 0 & otherwise \end{cases}$$

  b) The conditional expected value of W given C is:

$$E[W/C] = \int_{-\infty}^{+\infty} x f_{W/C}(x)dx = \frac{2}{4\sqrt{2\pi}} \int_0^{+\infty} x e^{-\frac{x^2}{32}} dx$$

  by the substitution $y = \frac{x^2}{2}$ we have:

$$E[W/C] = \frac{8}{\sqrt{2\pi}} \int_0^{+\infty} e^{-y} dy = \frac{8}{\sqrt{2\pi}}$$

  c)

$$\begin{aligned} E[W^2/C] &= \int_{-\infty}^{+\infty} x^2 f_{W/C}(x)dx = 2\int_0^{+\infty} x^2 f_W(x)dx \\ &= \int_{-\infty}^{+\infty} x^2 f_W(x)dx = E[W^2] = \sigma^2 + \mu^2 = 16 \end{aligned}$$

  where for the second equality we have taken into account that $x^2 f_W(x)$ is and even function ,therefore:
  $\int_{-\infty}^{+\infty} x^2 f_W(x)dx = 2\int_0^{+\infty} x^2 f_W(x)dx$.

Finally:

$$Var[X/C] \quad = \quad E[W^2/C] - (E[W/C])^2 = 16 - \frac{32}{\pi}$$

(7) Let X be a uniform random variable with uniform distribution on $(-5, 5)$. Given the event $B = [|X| \leq 3]$:
a) Find the conditional density function $f_{X/B}(x)$.
b) Find the conditional expected value $E[X/B]$.
Solution
a) First we find:

$$P(B) = P[-3 \leq X \leq 3] = \int_{-3}^{3} \frac{1}{10} dx = \frac{3}{5}$$

Notice that the p.d.f. of X is:

$$f_X(x) = \begin{cases} \frac{1}{10} & -5 \leq x \leq 5 \\ 0 & otherwise \end{cases}$$

Then, as $B \subset S_X$:

$$f_{X/B}(x) = \begin{cases} \frac{f_X(x)}{P(B)} = \frac{1}{6} & -3 \leq x \leq 3 \\ 0 & otherwise \end{cases}$$

therefore $Y \sim U(-3, 3)$.
b) By the result about the expected value of a uniform distribution $E[X/B] = \frac{-3+3}{2} = 0$. Alternatively:
$E[X/B] = \int_{-3}^{3} x\frac{1}{6} dx = 0$

(8) A random variable X describes the voltage at the receiver in a modem. When a symbol "0" is transmitted (event B_0), $X \sim N(-5, 2)$. When a symbol "1" is transmitted(event B_1), $X \sim N(5, 2)$. Given he symbols are equally liked to be sent, find the p.d.f. function of X.

Solution

$$f_X(x) \quad = \quad f_{X/B_0}(x)P(B_0) + f_{X/B_1}(x)P(B_1) = \frac{1}{2}\frac{1}{\sqrt{4\pi}}e^{-\frac{(y-5)^2}{4}} + \frac{1}{2}\frac{1}{\sqrt{4\pi}}e^{-\frac{(y+5)^2}{4}}$$

(9) Let Y be an exponential distribution with parameter $\lambda = 0.2$. Given the event $A = [Y < 2]$.
a) What is the conditional p.d.f., $f_{X/A}(x)$?
b) Find the conditional expected value, $E[X/A]$.
Solution
a) We first compute:

$$P[A] = \int_0^2 0.2e^{-0.2x}\, dx = -0.2\frac{1}{0.2}e^{-0.2x}|_0^2 = 1 - e^{-0.4}$$

Then for $0 < x < 2$

$$f_{X/A}(x) \quad = \quad \frac{f_X(x)}{P[A]} = \frac{0.2e^{-0.2x}}{1 - e^{-0.4}}$$

and zero otherwise.
b)

$$\begin{aligned} E[X/A] &= \int_{-\infty}^{+\infty} x f_{X/A}(x)\,dx = \int_0^2 x\frac{0.2e^{-0.2x}}{1-e^{-0.4}}\,dx \\ &= \frac{0.2}{1-e^{-0.4}}\int_0^2 xe^{-0.2x}\,dx \end{aligned}$$

Using the integration by parts formula with $u = x$ and $dv = e^{-0.2x}dx$ we have:

$$\begin{aligned} E[X/A] &= \frac{0.2}{1-e^{-0.4}}\left(-5xe^{-0.2x}|_0^2 + \int_0^2 5e^{-0.2x}\right)\,dx \\ &= \frac{5-7e^{-0.4}}{1-e^{-0.4}} \end{aligned}$$

(10) Flip a coin twice. On each flip, the probability of heads equals p. Let X_i equal the number of heads (either 0 or 1) on flip i. Let $W = X_1 - X_2$ and $Y = X_1 + X_2$. Find $p_{W,Y}(w, y)$ and $p_{W/Y}(w/y)$.
Solution
Notice that the possible values of W are $-1, 0, 1$ and the possible values of Y are $0, 1, 2$. Then, for example:

$$p_{W,Y}(-1, 0) = P(W = -1, Y = 0) = 0$$

because it is impossible to have both values at the same time.
The analysis is completed in the following table:

outcome	prob.	W	Y
hh	p^2	0	2
ht	$p(1-p)$	1	1
th	$p(1-p)$	-1	1
tt	$(1-p)^2$	0	0

It leads to:

$p_{W,Y}(w, y)$	$y = 0$	$y = 1$	$y = 2$
$w = -1$	0	$p(1-p)$	0
$w = 0$	$(1-p)^2$	0	p^2
$w = 1$	0	$p(1-p)$	0

The marginal p.m.f. are computed by adding up rows for $p_W(w)$ and columns for $p_Y(y)$, then:

$$p_Y(y) = \begin{cases} (1-p)^2 & y = 0 \\ 2p(1-p) & y = 1 \\ p^2 & y = 2 \end{cases}$$

The conditional distribution for $y = 1$ is:

$$p_{W/Y}(w/1) = \begin{cases} \frac{p(1-p)}{2p(1-p)} = \frac{1}{2} & w = -1, 1 \\ 0 & otherwise \end{cases}$$

Similarly:

$$p_{W/Y}(w/0) = \begin{cases} \frac{(1-p)^2}{(1-p)^2} = 1 & w = 0 \\ 0 & otherwise \end{cases}$$

$$p_{W/Y}(w/2) = \begin{cases} \frac{p^2}{(p^2} = 1 & w = 0 \\ 0 & otherwise \end{cases}$$

(11) Given a pair of continuous random variables (X, Y) with joint density function:

$$f_{X,Y}(x, y) = \begin{cases} x + y & 0 < x, y < 1 \\ 0 & otherwise \end{cases}$$

a) Find its conditional p.d.f. $f_{X/Y}(x/y)$.
b) Find its conditional p.d.f. $f_{Y/X}(y/x)$.
c) Find the conditional expected value $E[X/Y = y]$.
d) Are the random variables independent?

Solution

a) Find its conditional p.d.f. $f_{X/Y}(x/y)$. We first compute the marginal p.d.f.:

$$\begin{aligned} f_X(x) &= \int_0^1 (x + y)\, dy = x + \frac{1}{2}, 0 \le x \le 1 \\ f_Y(y) &= \int_0^1 (x + y)\, dx = y + \frac{1}{2}, 0 \le y \le 1 \end{aligned}$$

Then for $0 \le y \le 1$:

$$f_{X/Y}(x/y) = \begin{cases} \frac{x+y}{y+\frac{1}{2}} & 0 < x < 1 \\ 0 & otherwise \end{cases}$$

b) Find its conditional p.d.f. $f_{Y/X}(y/x)$. For $0 \le x \le 1$:

$$f_{Y/X}(y/x) = \begin{cases} \frac{x+y}{x+\frac{1}{2}} & 0 < y < 1 \\ 0 & otherwise \end{cases}$$

c) Find the conditional expected value $E[X/Y = y]$.

$$\begin{aligned} E[X/Y = y] &= \int_{-\infty}^{+\infty} x f_{X/Y}(x/y)\, dx \\ &= \frac{1}{y + \frac{1}{2}} \int_0^1 x(x + y)\, dx \\ &= \frac{1}{y + \frac{1}{2}} \left[\int_0^1 x^2\, dx + y \int_0^1 x\, dx \right] \\ &= \frac{1}{y + \frac{1}{2}} \left[\frac{1}{3} + \frac{y}{2} \right] \\ &= \frac{3y + 2}{6y + 3} \end{aligned}$$

for $0 \le y \le 1$ and zero otherwise.

d) We see that:

$$\begin{aligned} f_{X,Y}(x, y) &= x + y \\ &\neq f_X(x) f_Y(y) = (x + \frac{1}{2})(y + \frac{1}{2}) \end{aligned}$$

therefore the random variables are not independent.

(12) Let X have a Beta distribution with parameters a and b. Let Y be distributed as Binomial(n,X).

a) Find $E[Y|X]$.

b) Describe the distribution of Y and give $E[Y]$ and $V[Y]$.
c) What is the distribution of Y in the case when X is uniform on the interval $(0,1)$?

Solution

a) $E[Y|X] = nX$.
b) $p_Y(x) := P(Y = x) = \int_0^1 B(n,x)p_{\beta(a,b)}(x)\,dx$, where $p_{\beta(a,b)}(x)$ is the p.d.f. of a beta distribution with parameters $B(n,x)$ is the p.m.f. of a binomial distribution with parameters n and x.
By the iterative formula $E[Y] = E[E[Y/X] = E[nX] = n\frac{a}{a+b}$.
On the other hand:

$$\begin{aligned} E[Y^2] &= E[E[Y^2/X]] = E[nX(1-X) + n^2X^2] = nE[X] + n(n-1)E[X^2] \\ &= n\frac{a}{a+b} + n(n-1)\frac{a}{(a+b)^2}\left(a + \frac{b}{a+b+1}\right) \\ Var[Y] &= E[Y^2] - (E[Y])^2 = n\frac{a}{a+b} + n(n-1)\frac{a}{(a+b)^2}\left(a + \frac{b}{a+b+1}\right) - n^2\frac{a^2}{(a+b)^2} \end{aligned}$$

c) For $y = 0, 1, \ldots, n$

$$\begin{aligned} p_Y(y) &= P[Y = y] = E[E[1_{[Y=y]}/X]] = \int_0^1 \binom{n}{y} x^y (1-x)^{n-y}\,dx \\ &= \binom{n}{y}\frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)} = \frac{1}{n+1} \end{aligned}$$

(13) For the joint distribution given below calculate the p.d.f. of $Y|X$ and $E[Y|X]$:

$$f(x,y) = \lambda^2 exp(-\lambda y),\ y > x > 0$$

Solution

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{X,Y}(x,y)\,dy = \lambda^2 \int_0^{+\infty} exp(-\lambda(x+y))\,dy \\ &= \lambda^2 exp(-\lambda x)\int_x^{+\infty} exp(-\lambda y)\,dy = \lambda exp(-2\lambda x), x \geq 0 \\ f_{Y/X}(y/x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} = \lambda exp(\lambda x) exp(-\lambda y), y > x \geq 0 \\ E(Y/X) &= \lambda exp(\lambda x)\int_x^{+\infty} y exp(-\lambda y)\,dy \\ &= \lambda exp(\lambda x)\left(\frac{x}{\lambda}exp(-\lambda x) + \frac{1}{\lambda^2}exp(-\lambda x)\right) \\ &= x + \frac{1}{\lambda} \end{aligned}$$

(14) *Memoryless property of the exponential distribution*
Let X be an exponential random variable with parameter λ. Let $a > 0, b > 0$. Find:
a) $P(X > a)$
b) $P(X > a+b, X > a)$
c) $P(X > a+b/X > b)$
d) If X is the time between two consecutive breakdowns of a machine interpret the result.

Solution

a) $P(X > a) = exp(-\lambda a)$.

b) $P(X > a+b, X > a) = P(X > a+b) = exp(-\lambda(a+b))$
c) $P(X > a+b/X > b) = \frac{P(X>a+b, X>b)}{P(X>b)} = exp(-\lambda a)$.
d) Interpretation: given that the machine has not broken for b hours, the probability that the machine does not break in the next a hours equals the probability it does not break in a hours, i.e. the break up of the machine does not depend on its previous behavior.

(15) Let $U \sim U(0,1)$ and F be a continuous distribution function, i.e. a non-decreasing, right-continuous function with $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$. Consider the new random variable $Y = F^{-1}(U)$.
a) Find $F_Y(x)$.
b) Let $F(x) = 1 - e^{\lambda x}$, for $x > 0$ and zero otherwise. Find the p.d.f. of $X = F^{-1}(U)$.
c) Prove that if $U \sim U(0,1)$, then $1 - U \sim U(0,1)$ and hence we have $X = -\frac{1}{\lambda}\log(U) \sim exp(\lambda)$.

This result provides a method to generate numbers coming from a continuous random variable with c.d.f. F, as long as its inverse can be calculated and provided the existence of a mechanism to generate uniformly distributed numbers.

Solution

Let $0 \leq y \leq 1$:
a) $F_Y(x) = P(Y \leq y) = P(F^{-1}(U) \leq y) = P(U \leq F(y)) = F(y)$.
Hence the random variable Y has c.d.f. F. It allows to generate numbers coming from a c.d.f. F by generating uniform numbers and transforming them into $F^{-1}(U)$.
b) By part a) $X = -\frac{1}{\lambda}\log(1-U) \sim exp(\lambda)$.
c) $F_{1-U}(x) = P(1 - U \leq x) = P(U > 1 - x) = 1 - (1 - x) = x$.

(16) Suppose a set $A \subset \Omega$. Calculate $\sigma(A)$ the σ-algebra generated by A.

Solution

By definition Ω is in $\sigma(A)$, so is A. Also its complements $\emptyset$ and A^c are. Unions of these sets reduces to the same set. Thus $\sigma(\Omega) = \{\emptyset, A, A^c, \Omega\}$.

1.5. Problems

(1) Let a noisy acoustic signal be described by the equation:

$$X_t = A\cos(2\pi f t + \theta) + \xi_t,\ t \in \mathbb{N}$$

where A is a random variable with zero mean and unit variance, θ is a random variable uniformly distributed on the interval $]-\pi,\pi]$ and $(\xi_t)_{t\in\mathbb{N}}$ is a white noise with mean zero and variance one independent of the random variables A and θ.
a) Find the first order moment m_{X_t} of the process.
b) Find the second order moment $\sigma^2_{X_t}$ of the process.
c) Find the autocovariance and autocorrelation functions $\gamma(s,t)$ and $\rho(s,t)$.
d) Is the signal $(X_t)_{t\in\mathbb{N}}$ a stationary process in the strict and wide sense? Explain.

(2) If a process $(X_t)_{t\geq 0}$ has stationary and independent increments, finite first and second moments, with $X_0 = 0$, prove that:
a) $m(t+s) = m(t) + m(s)$, $\sigma^2(t+s) = \sigma^2(t) + \sigma^2(s)$, $t, s > 0$
b) $E(X_t) = E(X_1)t$
c) $Var[X_t] = Var[X_1]t$

Hint: Use the following result, a continuous function f that verifies $f(x+y) = f(x) + f(y)$ is linear, i.e. $f(x) = f(1)x$.

(3) Let $(X_t)_{t\in\mathbf{N}}$ be a random noise given by:

$$X_t = Acos(t+\theta) + \xi_t$$

where A is a random variable with mean zero and variance one, θ is a random variable uniformly distributed on $(0,1)$ and $(\xi_t)_{t\in\mathbf{N}}$ is a white noise with mean zero and variance equal to one. The quantities A, θ and ξ_t are independent.

a) For any $t \in \mathbb{N}$ find $E(X_t)$, the expected value of the process.
b) Find the autocovariance function $\gamma_X(s,t)$ at points $s=1$ and $t=2$.
c) Is the process stationary in the wide sense? Explain.

(4) Let $(X_t)_{t\in\mathbf{N}}$ be a random noise given by:

$$X_t = sin(2\pi t+\theta) + 2\xi_t$$

where θ is a random variable uniformly distributed on $(-\pi,\pi)$ and $(\xi_t)_{t\in\mathbf{N}}$ is a white noise with mean zero and variance equal to one. The quantities θ and ξ_t are independent.

a) For any $t \in \mathbb{N}$ find $E(X_t)$, the expected value of the process.
b) Find the variance function $\sigma^2(X_t)$.
c) Is the process stationary in the strict and wide sense? Explain.

(5) *Poisson probability*
Consider a probability space defined by $\Omega = \mathbb{N}$, the power set as the σ-algebra and the probability defined for non-negative integer values as $P(k) = \frac{e^{-\lambda}\lambda^k}{k!}$.

a) Find $P(0)$, $P(1)$, $P(\{0,1\})$.
b) Verify P is a probability, i.e. the Kolmogorov's axioms hold.

(6) *A mixing random variable*
Certain game asks you to toss a coin. If the coin lands Tail, you lose 1$, and if Head then you draw a number from [0, 1] at random and gain that number. Furthermore, suppose that the coin lands a Head with probability p. Let X be the amount of money won or lost after one game. Find the distribution of X.

(7) A hen lays N eggs where N has the Poisson distribution with parameter λ. Each egg hatches with probability p independently of all other eggs. Let K be the number of chicks that hatch. Find E(K—N), E(N), and E(K). Furthermore, find the distribution of K.

(8) A received carrier signal is given by:

$$X_t = A\cos(2\pi ft+\theta)$$

where A is constant and $\theta \sim U(0,2\pi)$. Find the expected value and autocorrelation of process X_t.

(9) A simple model(in degrees Celsius) for the daily temperature process C_t is:

$$C_t = 16(1-\cos\frac{2\pi t}{365}) + 4X_t$$

where $X_1, X_2, \ldots$ are a white noise with $\sigma = 1$.
a) Find $E[C_t]$.
b) Find the autocovariance function $C_C(s,t)$.

CHAPTER 2

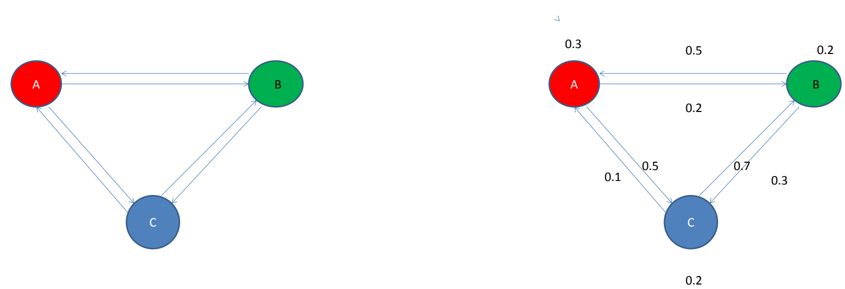
Markov Chains and Markov Processes

In this chapter we study the class of Markov process, named after the Russian mathematician Andrey Markov (1856-1922). The main idea is of a simple dynamic where the future position of the system depends on the past positions only through the present moment. We first introduce the class of discrete-time and finite or countable space models known as Markov chains and move to Poisson and compound Poisson processes as special counting processes. Then, we analyse continuous-time processes with the Markovian property. Finally we cover applications to the modeling of epidemics and population dynamic. We cover also some applications to rating algorithms in Finance.

2.1. Markov Chains

Definition 2.1. *Markov Chain*
A discrete-time stochastic process $(X_n)_{n\in\mathbb{N}}$ defined on a probability space $(\Omega, \mathcal{A}, P)$ with values in a set $\mathbb{E} \subset \mathbb{N}$ is a Markov chain if for all $n \in N$, $t_1 \leq t_2 \leq \ldots \leq t_n \in \mathbb{N}$, $i_1, i_2, \ldots, i_n \in \mathbb{E}$ we have:

$$P\left\{X_{t_n} = i_n / X_{t_{n-1}} = i_{n-1}, \ldots, X_{t_1} = i_1\right\} = P\left\{X_{t_n} = i_n / X_{t_{n-1}} = i_{n-1}\right\} \tag{2.1}$$



(a) (b)

Figure 1. Left: A Markov chain with three states. The system is moving between them following the arrows. Right: Same system and its transition probabilities

In figure 1(left) a Markov chain with three states is shown. Arrows indicate possible moves between states. In figure 1(right) the probabilities associated with these transitions are given. Thus, for example the system once at A will move to B in one step 20% of the time, independent of previous movements.

Remark 2.2. The Markovian property, expressed in equation (2.1), is interpreted saying that the future position of the system, given the present, is independent of the past. A Markov chain is a process with a weak memory. Moving forward it "forgets" past positions, only the current position counts.

The behavior of a Markov chain is determined by the probabilities of transition from one state to another together with the probability to be at a particular initial state.

Definition 2.3. *Transition probabilities*
Let the stochastic process $(X_n)_{n\in\mathbb{N}}$, defined on a probability space $(\Omega,\mathcal{A},P)$, with values in a set $\mathbb{E}\subset\mathbb{N}$ be a Markov Chain. The transition probability from state i at time s to state j at time t is defined as:

$$p_{ij}(s,t)=P\{X_t=j/X_s=i\},\ i,j\in\mathbb{E}$$

The subclass of stationary Markov Chains, where these transition probabilities depends only on the number of steps and not on the particular starting and ending point is introduced below.

Definition 2.4. *Stationary Markov Chain*
A Markov chain $(X_n)_{n\in\mathbb{N}}$ is stationary if for any $s>0,t>0$ and any states $i,j\in\mathbb{E}$ the transition probability $p_{ij}(s,s+t)$ depends only on the length of the interval $(s,s+t$ and not on the initial and final times s and $s+t$. Hence, they can be written as:

$$p_{ij}(t)=P\{X_{t+s}=j/X_s=i\},\ i,j\in\mathbb{E}$$

The quantities $p_{ij}(t)$ are called *t-step transition probabilities* from state i to state j.
For the seek of simplicity we will constrain our analysis to stationary Markov chains with a finite number of states, unless the contrary is indicated. Some phenomena, e.g. climatic, economic systems do not present homogeneous movements upon time, others have a seasonal behavior, hence they can not be modeled by stationary Markov chains. Nonetheless results for stationary chains can be generally adapted to non-stationary processes with some extra notational work.
The t-step transition probability matrix denoted by $P(t)$ is a matrix with entries $P(t)=(p_{ij}(t))_{i,j\in\mathbb{E}}$.
In particular, we denote $P(1)=P$.

Remark 2.5. The sum of the elements of rows in the matrix $P(t)$ is equal one.

$$\sum_{j\in\mathbb{E}}p_{ij}(t)=1,\ t\geq 0,\ i\in\mathbb{E}$$

A matrix with such property is called *stochastic*.
The definition of a stationarity Markov chain is consistent with the concept of strictly stationary process introduced in chapter one.

The probability distribution of the initial state and the transition probability matrices determine the probability of any given trajectory of a stationary Markov chain and therefore the family of joint c.d.f's of the process by simply adding the probabilities of each trajectory . The next result precise the claim.

Proposition 2.6. *Probabilities of a trajectory*
Let $(X_n)_{n\in\mathbb{N}}$ defined on a probability space $(\Omega,\mathcal{A},P)$ be a Markov chain with state space $\mathbb{E}$, initial probabilities $p_j^{(0)}=P[X_0=j]$,$j\in\mathbb{E}$ and transition probability matrix $P(t)$.
Let $n\in\mathbb{N}$, $0<t_1<t_2<\ldots<t_n$, and $i_1,i_2,\ldots,i_n\in\mathbb{E}$.
Then:

$$P[X_0=i_0,X_{t_1}=i_1,\ldots,X_{t_n}=i_n] \quad = \quad p_j^{(0)}\prod_{j=1}^{n}p_{i_{j-1},i_j}(t_j-t_{j-1})$$

PROOF. By the multiplication rule and the Markov property:

$$
\begin{aligned}
& P[X_0 = i_0, X_{t_1} = i_1, \dots, X_{t_n} = i_n] \\
= \;& P[X_{t_n} = i_n / X_{t_{n-1}} = i_{n-1}, \dots, X_0 = i_0] P[X_0 = i_0, X_{t_1} = i_1, \dots, X_{t_{n-1}} = i_{n-1}] \\
= \;& P[X_{t_n} = i_n / X_{t_{n-1}} = i_{n-1}] P[X_0 = i_0, X_{t_1} = i_1, \dots, X_{t_{n-1}} = i_{n-1}] \\
= \;& P[X_{t_n} = i_n / X_{t_{n-1}} = i_{n-1}] P[X_{t_{n-1}} = i_{n-1} / X_{t_{n-2}} = i_{n-2}] P[X_{t_{n-2}} = i_{n-2}, \dots X_{t_1} = i_1, X_0 = i_0] \\
& \dots \\
= \;& P[X_{t_n} = i_n / X_{t_{n-1}} = i_{n-1}] P[X_{t_{n-2}} = i_{n-2} / X_{t_{n-2}} = i_{n-2}] \dots P[X_{t_1} = i_1 / X_{t_0} = i_0] P[X_0 = i_0]
\end{aligned}
$$

$\square$

We show some examples of applications of Markov Chains to different areas.

EXAMPLE 2.7. Examples of Markov chains.

a) *Communication 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, independently of previous transmissions. Letting X_n denote the digit entering the n-th stage. Then, we can assume that $(X_n)_{n\in\mathbb{N}}$ is a two-state Markov chain having a transition probability matrix:

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

The state space is $\mathbb{E} = \{0, 1\}$.

b) *Random Walks*
Starting at the origin a particle moves one unit to the right with probability p or one unit to the left with probability $1-p$ independently of past moves. Let X_n be the position of the particle after n moves, i.e. distance from the origin at time n.
Hence, an unrestricted random walk $(X_n)_{n\in\mathbb{N}}$ is a Markov chain with space of states $\mathbb{E} = \mathbb{Z}$ and one-step transition probabilities:

$$
\begin{aligned}
p_{i,i+1} &= p \\
p_{i,i-1} &= 1-p \\
p_{i,j} &= 0, \text{ for } j \neq i-1, i+1
\end{aligned}
$$

for any $i \in \mathbb{Z}$.
Variants of this process are the random walks with *reflecting* and *absorbing* barriers. The former, with state space $\{0, 1, \dots, a\}$ and reflecting barriers at 0 and a, i.e. $p_{0,1} = p_{a,a-1} = 1$. The process has absorbing barriers at 0 and a if $p_{0,0} = p_{a,a} = 1$. The one-step transition probability for the latter is:

$$
\begin{pmatrix}
1 & 0 & 0 & \dots & 0 & 0 \\
1-p & 0 & p & \dots & 0 & 0 \\
0 & 1-p & 0 & p & \dots & 0 \\
 & & & & \dots & \\
0 & 0 & 0 & \dots & 0 & 1
\end{pmatrix}
$$

c) *Galton-Watson Chain*
A Galton-Watson process is a sequence of random variables $(Z_n)_{n\in\mathbb{N}}$, where Z_n represents

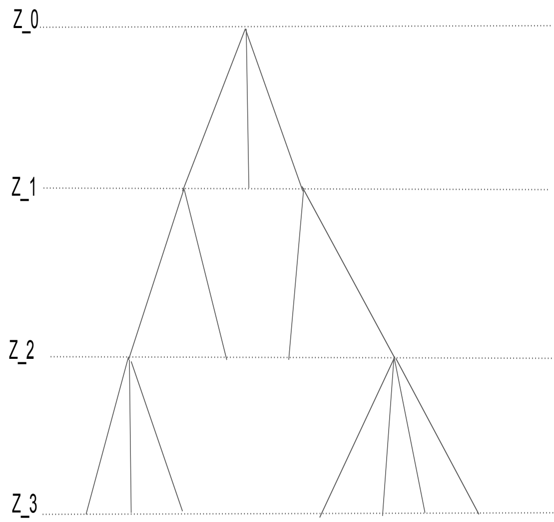


FIGURE 2. A Galton-Watson chain for the dynamic of a population. Z_n is the number of individuals in the n-th generation

the number of individuals in generation n.
The elements in the sequence is related in the following way:

$$\begin{aligned} Z_0 &= 1 \\ Z_{n+1} &= \sum_{k=1}^{Z_n} X_k, \ n = 1, 2, \dots \end{aligned}$$

where for any n, $(X_k)_{k=1,2,\dots,Z_n}$ is an array of independent and identically distributed random variables with finite mean $E(X_k^{(n)}) = \mu$. The random variable X_k is the random number of offsprings coming from the k-th individual in the n-th generation.
In figure 2 a graphic representation of a Galton-Watson chain is observed. The state space is $\mathbb{E} = \mathbb{N}$. The state zero, when the population becomes extinct, is an absorbing state.
Notice that the number individuals The model was originally introduced by the British botanist Francis Galton, motivated by the analysis of the extinction of aristocratic names in England. See [**2**].
Notice that the number of individuals in generation $n + 1$, only depends on the number of individuals in the previous generation and their number of offsprings which are also independent from previous generations. Hence the position of the system at $t = n + 1$, denoted by Z_{n+1} and conditionally to the position at time n, denoted by Z_n, is independent from past positions. It reflects the Markovian condition.

A Galton-Watson is a stationary Markov chain. Its one-step transition probability is:

$$p_{ij} \quad = \quad P(Z_{n+1}/Z_n) = P(\sum_{k=1}^{i} X_k = j) = f_{X_1}^{*i}(j)$$

where $f_{X_1}^{*i}$ is the i-th convolution of probability mass function(p.m.f.) of the number of offsprings, i.e.:

$$\begin{aligned} f_{X_1}^{*1}(j) &:= f_{X_1}(j) = \sum_{k=0}^{\infty} PX_1 = ks^k,\ 0 < s \leq 1 \\ f_{X_1}^{*i}(j) &= f_{X_1}^{*(i-1)} * f_{X_1}(j) := \sum_{k=0}^{j} f_{X_1}^{*(i-1)}(j-k) f_{X_1}(k) \end{aligned}$$

Its probability generating function as:

$$g_n(s) := E(s^{Z_n}) = \sum_{k=0}^{+\infty} P(Z_n = k)s^k, 0 < s \leq 1 \tag{2.2}$$

Table 1: Example one-year transition matrix - Moody's rating system

	Aaa	Aa	A	Baa	Ba	B	Caa	D
Aaa	93.38%	5.94%	0.64%	0.00%	0.02%	0.00%	0.00%	0.02%
Aa	1.61%	90.53%	7.46%	0.26%	0.09%	0.01%	0.00%	0.04%
A	0.07%	2.28%	92.35%	4.63%	0.45%	0.12%	0.01%	0.09%
Baa	0.05%	0.26%	5.51%	88.48%	4.76%	0.71%	0.08%	0.15%
Ba	0.02%	0.05%	0.42%	5.16%	86.91%	5.91%	0.24%	1.29%
B	0.00%	0.04%	0.13%	0.54%	6.35%	84.22%	1.91%	6.81%
Caa	0.00%	0.00%	0.00%	0.62%	2.05%	4.08%	69.19%	24.06%

FIGURE 3. Transition rating matrix one year horizon according to Moody. Source: The J.P. Morgan guide to Credit Derivative

d) *Transition of default probabilities*
Moody is a rating agency that classifies companies according to their chances to default. Rates go from solvent companies with a low probability of default to companies in technical default.
A sequence of random variables $(X_n)_{n \in \mathbb{N}}$, assumed to be a Markov chain, describes the credit worthiness at period n, typically a year, of the company via its rating category. The state space is $\{Aaa, Aa, A, Baa, Ba, C, D\}$ In figure 3 transition probabilities between

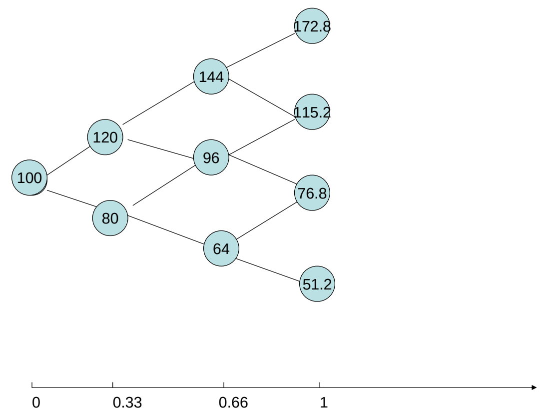


FIGURE 4. A binomial tree of three steps with spot price $S_0 = \$100$, $a = -0.20$, $b = 0.20$

rating categories in a year are presented.
Thus, for example, the entry in the first row and first column, equal to 93.38 % means that a companies rated Aaa has 93.38 % of probability to remain rated Aaa next year. Entries in the probability transition matrix are estimated from the corresponding observed frequencies of events in past years. A last row, corresponding to the state D with entries equal to zero everywhere, except one in the last entry is omitted.

e) *Modeling stocks: Binomial(Cox-Ross-Rubinstein) model*
Consider the process $(S_t)_{t\in\mathbb{N}}$, where S_t is the price of certain asset at time t. The dynamic of the prices is given by the recurrent expressions:

$$S_{t+1} = \begin{cases} S_t\,(1+a) & \text{with prob. } 1-p \\ S_t\,(1+b) & \text{with prob } p \end{cases} \tag{2.3}$$

with $-1 < a < b$.
Notice that because of the Markovian character of the process, changes in the asset price from one period to the next depend only on the price at the initial time.
See figure 4 where a binomial tree of prices is shown. The chain starts with an initial price $S_0 = \$100$. At time $t = \frac{1}{3}$, representing a quarter of a year, the price can go up to $S_1 = \$120$ or down to $S_1 = \$80$. Hence the rates are $a = -0.20$ and $b = 0.20$.

The next result allows to obtain the transition probabilities of a Markov chain in n steps from transition probabilities in one step.

THEOREM 2.8. *Chapman-Kolmogorov equations*
Let $(X_n)_{n\in\mathbb{N}}$ a Markov Chain with state space $\mathbb{E}$. Then, for any $n, m \in \mathbb{N}$ and $i, j \in \mathbb{E}$ the following system of equations hold:

$$p_{ij}(n+m) = \sum_{l\in\mathbb{E}} p_{il}(m)p_{lj}(n)$$

or in matrix form:

$$P(n+m) = P(n)P(m)$$

In particular:

$$P(n) \quad = \quad P^n$$

where $P(n)$ is the n-step transition probability matrix and P^n is the n-th power of one step transition probability matrix P.

Proof. Notice that $\Omega = \bigcup_{i\in\mathbb{E}}[X_{m+k} = l]$. Hence:

$$\begin{aligned} p_{ij}(n+m) &= P(X_{n+m+k} = j/X_k = i) = P(X_{n+m+k} = j, \bigcup_{i\in\mathbb{E}}[X_{m+k} = l]/X_k = i) \\ &= \sum_{l\in\mathbb{E}} P(X_{n+m+k} = j, X_{m+k} = l/X_k = i) \\ &= \sum_{l\in\mathbb{E}} P(X_{n+m+k} = j/X_{m+k} = l, X_k = i)P(X_{m+k} = l/X_k = i) \\ &= \sum_{l\in\mathbb{E}} P(X_{n+m+k} = j/X_{m+k} = l)P(X_{m+k} = l/X_k = i) \\ &= \sum_{l\in\mathbb{E}} p_{lj}(n)p_{il}(m) \end{aligned}$$

The second equality follows from the third Kolmogorov's axiom, the third equality from the property $P(A\bigcap B/C) = P(A/B\bigcap C)$ for random events A, B and C and the forth by the Markov property.

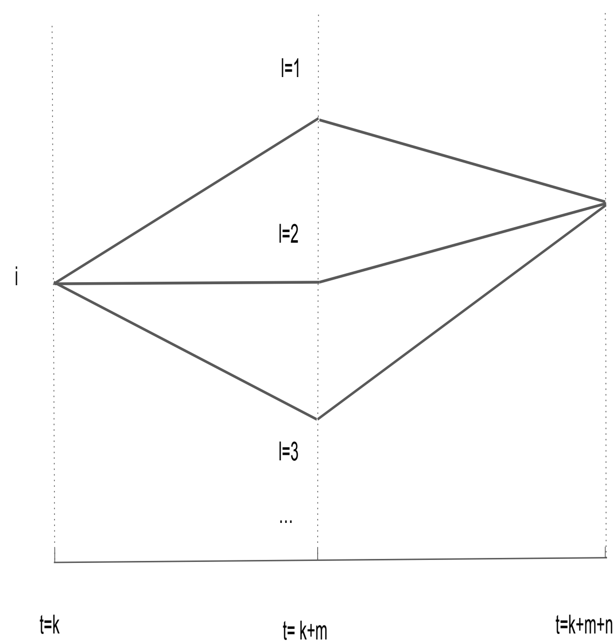


Figure 5. Paths from state i at time $t = k$ to state j at time $t = k + m + n$.

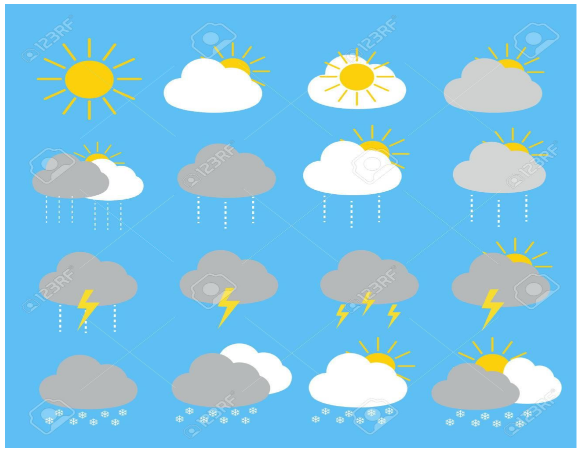


FIGURE 6. Sequence of sunny, cloudy and rainy days

The second equation in the theorem follows from the definition of matrix multiplication, while the third expression is obtained from applying repetitively the matrix form of Chapman-Kolmogorov equations. Applying several times Chapman-Kolmogorov equations with steps $n-1$ and 1 we have:

$$P(n) = PP(n-1) = \ldots = PPP(n-2) = P^n$$

We refer to figure 5 for an illustration. □

Weather sequence

EXAMPLE 2.9. Suppose weather in certain region is described by a Markov chain $(X_n)_{n\in\mathbb{N}}$, where its positions represent the daily states given by three events: sunny (state 1), cloudy (state 2) and rainy (state 3). See figure 6 for a schematic representation. Form previous studies it is known that the one-step transition probability matrix of the process is given by:

$$P = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.4 & 0.2 & 0.4 \\ 0 & 0.6 & 0.4 \end{pmatrix}$$

a) Find the probability transition matrix after 5 days.
b) If today is a rainy day, what is the probability that in 5 days it will be sunny?

Solution
a) After multiplying the one-step transition matrix five times we get:

$$P(5) = \begin{pmatrix} 0.5920 & 0.2550 & 0.1530 \\ 0.5101 & 0.2858 & 0.2042 \\ 0.4589 & 0.3062 & 0.2349 \end{pmatrix}$$

b) 45.89%

Next, we study the behavior of a Markov Chain after a large number of steps have elapsed.

DEFINITION 2.10. *Stationary probabilities*
Given a Markov chain $(X_n)_{n\in\mathbb{N}}$ with state space $\mathbb{E}$, we define the stationary probability set as the sequence $(\pi_j)_{j\in\mathbb{N}}$ verifying:

$$\pi_j = \lim_{n\to\infty} p_{ij}(n), j \in \mathbb{E}$$

$$\sum_{j\in\mathbb{E}} \pi_j = 1$$

REMARK 2.11. Notice that the stationary probabilities do not depend on the initial starting position of the Markov Chain. Intuitively it is consistent with the weak memory condition.

The existence of the stationary probabilities is not longer guaranteed. Below we state sufficient conditions, far from being necessary conditions, for their existence. They are related with the concept of communication states.

DEFINITION 2.12. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov Chain with state space $\mathbb{E}$.
A state j can be reached from another state i if there exists $n \in \mathbb{N}$ such that $p_{ij}(n) > 0$ Two states $i, j \in \mathbb{E}$ are said to communicate each other if there exist $n, m \in \mathbb{N}$ such that:

$$p_{ij}(n)p_{ji}(m) > 0$$

In other terms i can reach from j and viceversa. When i communicates with j we write $i \leftrightarrow j$.
A communicating class is a subset of states that communicate each other.

PROPOSITION 2.13. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov Chain with state space $\mathbb{E}$ where all states communicate each other. Then, the stationary probabilities exist and they verify the system of equations:

$$\begin{aligned} \pi_j &= \sum_{l\in\mathbb{E}} \pi_l p_{lj},\ j \in \mathbb{E} \\ \sum_{l\in\mathbb{E}} \pi_l &= 1 \end{aligned}$$

Or, in matrix form:

$$\Pi = \Pi P$$

where $\Pi = (\pi_1, \pi_2, \ldots,)$ is the vector of stationary probabilities.

PROOF. The proof of the existence is beyond the scope of these note. We proof that the stationary probabilities verifies the system of equations in the case the number of states is finite. Assuming stationary probabilities exist, we have by Chapman-Kolmogorov equations:

$$p_{ij}(n+1) = \sum_{l\in\mathbb{E}} p_{il}(n)p_{lj}$$

Taking limits on both sides, when $n \to \infty$ result is obtained. □

EXAMPLE 2.14. Refer to example 2.9.

a) Find the communicating states.
b) Does the stationary probabilities exist? If affirmative compute them.

Solution

a) All states communicate (see matrix $P(5)$ where all transition probabilities are positive).
b) Yes, because all states communicate. We have to solve the system:

$$\begin{aligned} \pi_1 &= 0.8\pi_1 + 0.4\pi_2 \\ \pi_2 &= 0.2\pi_1 + 0.2\pi_2 + 0.6\pi_3 \\ \pi_3 &= 0.4\pi_2 + 0.4\pi_3 \\ \pi_1 & \quad +\pi_2 + \pi_3 = 1 \end{aligned}$$

Its solution is $\pi_1 = \frac{6}{11}, \pi_2 = \frac{3}{11}, \pi_3 = \frac{2}{11}$

In order to simulate a Markov chain, from $X_n = i$ we have to generate a random variable, with p.m.f. equal to $p_{ij}, j \in \mathbb{E}$. The procedure is summarized in the following algorithm.

Algorithm 2.1

i) Inputs: X_0, P.
ii) At time $t = k$, if $X_k = i$, generate $U \sim U(0,1)$.
iii) If $\sum_{l=1}^{j-1} p_{il} \leq U < \sum_{l=1}^{j} p_{il}$ then $X_{k+1} = j$.

EXAMPLE 2.15. *Random walk*
To simulate m repetitions of an unrestricted random walk with initial position at the origin X_0, a probability p of go right the following code has been created in MATLAB:

```
function X=rwsim(X0,p,n,m)
%random walk simulation
%output: seq of positions
%X0: initial position
%p: probability of going right
%n: no. of steps
%m: no. of repetitions
X0M=X0.*ones(1,m);
U=rand(n,m);
NR=(U < p);
NR1=2*NR-1+X0;
X=cumsum(NR1);
end
```

In figure 8, Chapter 1 five trajectories of an unrestricted random walk starting $X_0 = 0$ with $p = 0.5$ and 100 steps are simulated using a Monte Carlo approach.

2.2. Poisson and compound Poisson processes

Counting random events occurring on an interval are present in a variety of important phenomena, ranging from earthquakes to car accidents. We recall from chapter 1 that a counting process $(N_t)_{t\geq 0}$ is a process starting at $N_0 = 0$, taking values in the set of non-negative integers, with trajectories having discontinuities of size one and remaining constant between jumps. The value N_t is interpreted as the number of events taking place on the interval $[0, t]$.
As the most notable case of counting processes we have the Poisson process. We recall its definition from chapter 1:

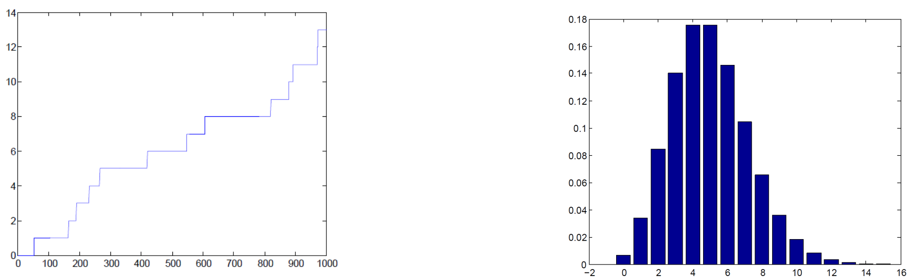
DEFINITION 2.16. A process $(N_t)_{t\geq 0}$ defined on the stochastic basis $(\Omega, \mathcal{A}, P)$, with values in $\mathbb{N}$ is a Poisson process if it satisfies:

p1) $N_0 = 0$.
p2) It has independent increments, i.e. for any $n \in \mathbb{N}$, $t_1 \leq t_2 \leq \ldots \leq t_n$ the random variables $N_{t_j} - N_{t_{j-1}}$ are independent.
p3) It has stationary increments, i.e. for any $s, t > 0$ the probability distribution of $N_{s+t} - N_s$ does not depend on s.
p4) For any $s, t \geq 0$ $N_t - N_s \sim \mathcal{P}(\lambda(t-s))$, where $\mathcal{P}(\lambda t)$ is the Poisson probability law with parameter λt.

Condition p4) implies that:

$$P(N_t - N_s = k) = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!},\ k = 0, 1, \ldots$$

REMARK 2.17. For a Poisson probability distribution with parameter λt, see table 1 in chapter 1, we have that $E(N_t) = Var(N_t) = \lambda t$.
Solving for λ we have $\lambda = \frac{1}{t}E(N_t)$. It allows to interpret the value λ, called intensity of the Poisson process, as the average number of events per time unit.
Notice that the parameter of a Poisson process is the unknown value λ, as the parameter of the random variable N_t is λt.



(a) (b)

FIGURE 7. Figure (a):A simulated trajectory of a Poisson process. Figure (b): The distribution of frequencies of a Poisson distribution

In figure 7a) a simulated trajectory of a Poisson Process with $\lambda = 5$ is shown. Trajectories are constant on random intervals.
A sufficient condition for a counting process with independent and stationary increments to be a Poisson process is given in the result below. Previously, we introduce the convenient functions small "o" and big 'O".

DEFINITION 2.18. *Small and big "o" functions*
Real-valued functions f and g verify the relation $f = o(g)$ if:

$$\lim_{t \to 0} \frac{f(t)}{g(t)} = 0$$

They verify the relation $f = O(g)$ if:

$$|f(t)| \leq M|g(t)|$$

for some constant $M > 0$.

In other words f is a small "o" function of g is it is an infinitesimal of higher order than g. For example if $f(t) = t^2$ and $g(t) = t$, then $f = o(g)$. If t is small the function f is negligible with respect to g.

THEOREM 2.19. *Sufficient condition for a Poisson process*
Let $(N_t)_{t \geq 0}$ be a counting process starting at the origin. Denote by $p_k(t) = P[N_t = k]$. In addition assume that:

i) *The process has stationary and independent increments.*
ii) *The probability that 2 of more events occur in a small interval of time is negligible, i.e.*

$$\sum_{k=2}^{+\infty} p_k(h) = o(h)$$

Then, $(N_t)_{t\geq 0}$ is a Poisson process.

PROOF. For any nonnegative integers m and n by the condition of stationarity and independence of the increments we have:

$$p := P[N_1 = 0] = (P[N_{1/n} = 0])^n$$

Hence:

$$P[N_{1/n} = 0] = p^{\frac{1}{n}}. \text{ Similarly } P[N_{m/n} = 0] = p^{\frac{m}{n}}$$

Then, we take an increasing sequence of rational numbers $\frac{m(n)}{n}$ such that $\frac{m(n)}{n} \to t$ as $n \to \infty$. Notice the numerator depends on n.
Therefore:

$$\begin{aligned} p_0(t) &= P[N_t = 0] = \lim_{\frac{m(n)}{n}\to t} P[N_{m(n)/n} = 0] \\ &= \lim_{\frac{m(n)}{n}\to t} p^{\frac{m(n)}{n}} = p^t = e^{-\lambda t} = 1 - \lambda t + o(t) \end{aligned}$$

with $\lambda = -\log p > 0$.
The last equality follows from a first order Taylor expansion of the exponential function around zero.
Moreover, from the Law of Total Probability and $k \geq 1$:

$$\begin{aligned} p_k(t+h) &= p_k(t)p_0(h) + p_{k-1}(t)p_1(h) + \sum_{j=2}^{+\infty} p_{k-j}(t)p_j(h) \\ &= p_k(t)(1-\lambda h + o(h)) + (\lambda h + o(h))p_{k-1}(t) + o(h) \end{aligned}$$

After subtracting $p_k(t)$ on both sides we have:

$$\frac{p_k(t+h) - p_k(t)}{h} = -\lambda p_k(t) + \lambda p_{k-1}(t) + \frac{o(h)}{h}$$

Taking the limit when $h \to 0^+$ leads to:

$$\frac{dp_k(t)}{dt} = -\lambda p_k(t) + \lambda p_{k-1}(t)$$

Solving the system of ordinary differential equation system in a recursive way, starting with $p_0(t) = e^{-\lambda t}$ we have:

$$p_k(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

□

EXAMPLE 2.20. Data packages transmitted from a modem over a phone line form a Poisson process with $\lambda = 10$ packages/sec. Find the probability that:

a) Seven packages arrive within a second.
b) Twenty packages arrive within two seconds.
c) Seven packages arrive between 8 a.m. a second after 8 a.m. and another seven packages arrive within 8 p.m. a second after 8 p.m.

Solution
Let N_t be the number of packages transmitted on $[0, t]$ by the modem. Assume it is a Poisson process with intensity $\lambda = 10$.

a)

$$P[N_1 - N_0 = 7] = P[N_1 = 7] = e^{-10}\frac{10^7}{7!} = 0.0901$$

b)

$$P[N_2 = 20] \quad = \quad e^{-2\times 10}\frac{20^{20}}{20!} = 0.0888$$

c) Let t' the time in seconds at 8 p.m. when the process stars at 8 a.m. By stationarity and independence of the increments:

$$P[N_1 = 7, N_{t'+1} - N_{t'} = 7] \quad = \quad (P[N_1 = 7])^2 = \left(e^{-10}\frac{10^7}{7!}\right)^2 = 0.0081$$

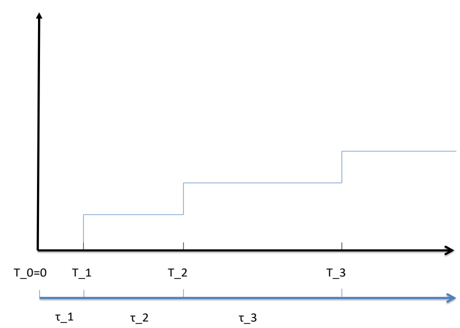


FIGURE 8. Times between consecutive events in a Poisson process

The next result provides the probability distribution of the time between consecutive events in a Poisson process. The situation is illustrated in figure 8.

PROPOSITION 2.21. *Probability density function of the time between jumps*
Let $(N_t)_{t\in\mathbb{N}}$ be a Poisson process with parameter $\lambda > 0$ and $T_0 = 0, T_1, T_2, \dots, T_n, \dots$ its jump times. Then, the time between jumps denoted by $\tau_j = T_j - T_{j-1}, j = 1, 2, \dots$ are independent and exponentially distributed.

PROOF. Let $n \in \mathbb{N}$ and deterministic times $0 \le t_1, t_2, \dots, t_n, t$ then by the stationarity of the increments in a Poisson process and for $t \ge 0$:

$$P(\tau_j > t) \quad = \quad P(N_{T_{j-1}+t} - N_{T_{j-1}} = 0) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$$

Then, $F_{\tau_j}(t) = P(\tau_j \ge t) = 1 - e^{-\lambda t}$ which is the c.d.f. of an exponential distribution of parameter λ.
On the other hand:

$$\begin{aligned}
& P(\tau_1 \ge t_1, \tau_2 \ge t_2, \dots, \tau_n \ge t_n) \\
= \quad & P(N_{t_1} = 0, N_{T_1+t_2} - N_{T_1} = 0, \dots, N_{T_{n-1}+t_n} - N_{T_{n-1}} = 0) \\
= \quad & P(N_{t_1} = 0)P(N_{T_1+t_2} - N_{T_1} = 0)\dots P(N_{T_{n-1}+t_n} - N_{T_{n-1}} = 0) \\
= \quad & e^{-\lambda t_1}e^{-\lambda t_2}\dots e^{-\lambda t_n}
\end{aligned}$$

It proves that the events $[\tau_k \ge t_k], k = 1, 2, \dots, n$ are independent, which is equivalent to the independence of the random variables τ_k. $\square$

REMARK 2.22. We have used a stronger condition for stationarity of the Poisson process. It is verified also for random times, i.e. for any non-negative random variable T and a deterministic value t the probability distribution of $N_{T+t} - N_T$ does not depend on T.

PROPOSITION 2.23. *Conditional jump times*
Let $(N_t)_{n\in\mathbb{N}}$ be a Poisson process with parameter $\lambda > 0$ and $T_1 < T_2, <, \ldots, T_n, \ldots$ the times at which the process changes position. Then, conditionally to $N_t = n$ the jump times $T_1, T_2, \ldots, T_n$ are distributed as $U(1), U(2), \ldots, U(n)$ where the latter is an ordered sample(from smallest to greatest) of $U_1, U_2, \ldots, U_n$, where each U_k is uniformly distributed on $(0, t)$.

There are several ways to simulate trajectories of a Poisson process on an interval $[0, T]$ depending on the information we use. Based on the condition p4) in the definition of a Poisson process. It uses the MATLAB built-in command *Poiss* that allows to generate number coming from a Poisson distribution.

Algorithm 2.2(positions at regular intervals)

1.- Generate $N_k \sim Poiss(\lambda\Delta t)$, for $k = 1, 2, \ldots, n = [\frac{T}{\Delta t}]$
2.- Generate $N_{n+1} \sim Poiss(\lambda(T - \Delta t))$ if necessary.
Here $[x]$ is the integer part of x.

Based on proposition 2.21.

Algorithm 2.3(positions at jumping times)

(1) At step k, generate $\tau_k \sim \exp(\lambda)$
(2) Compute $T_k = \sum_{j=1}^{k} \tau_j$.
(3) If $T_k < T$ set $N_{T_k} = k$.

Finally algorithm 2.4 is based on proposition 2.23.

Algorithm 2.4(Conditional jump times)

(1) Generate $N_T \sim Poiss(\lambda T)$.
(2) Generate $U_1, U_2, \ldots, U_{N_T}$ uniformly distributed.
(3) Order the sample to $U(1), U(2), \ldots, U(N_T)$
(4) $T_k = U(k)$.
(5) Set $N_{T_k} = k$.

A natural extension of a Poisson process is to allow jumps of random magnitude, while keeping the properties of stationarity and independence of its increments.

DEFINITION 2.24. *Compound Poisson process*
A Compound Poisson Process $(Y_t)_{t\geq 0}$ is defined as:

$$Y_t = \sum_{k=1}^{N_t} X_k$$

where $(N_t)_{t\geq 0}$ is a Poisson Process with parameter $\lambda > 0$ and $(X_k)_{k\in\mathbb{N}}$ is a sequence of independent and equally distributed random variables with mean finite μ and variance σ^2, and independent also of the Poisson process.

The expected value of Y_t is computed using the iterative property of the conditional expected value as:

$$\begin{aligned} E(Y_t) &= E(\sum_{i=1}^{N_t} X_i) = E(E(\sum_{i=1}^{N_t} X_i/N_t)) \\ &= E(N_t E(X_i)) = \mu E(N_t) = \mu\lambda t \end{aligned}$$

Similarly:

$$Var(Y_t) = \lambda(\sigma^2 + \mu^2)t$$

The value X_k is the random size of the k-th jump. Notice that the Poisson process is a particular case of a compound Poisson process with deterministic jumps $X_k = 1$.

EXAMPLE 2.25. Let $(Y_t)_{t\geq 0}$ a compound Poisson process with normal jumps $X_k \sim N(\mu_J, \sigma_J^2)$.

a) Find the probability distribution of Y_t conditionally to N_t.
b) Find the unconditional probability distribution of Y_t.
c) Find the expected value, standard deviation, the third and forth moments of the compound Poisson process $(Y_t)_{t\geq 0}$.

Solution

a) Because sum of normal random variables is normally distributed $Y_t/N_t \sim N(N_t\mu_J, N_t\sigma_J^2)$. For $N_t = 0$ it is a degenerated normal distribution where all the probability mass is concentrate at zero.
b) From the Law of Total Probability:

$$\begin{aligned} P(Y_t \leq x) &= \sum_{k=0}^{+\infty} P(Y_t \leq x/N_t = k)P(N_t = k) \\ &= \sum_{k=0}^{+\infty} F_{Y_k/N_t=k}(x)e^{-\lambda t}\frac{(\lambda t))^k}{k!} \\ &= 1_{\{\mathbb{R}_+\}}(x)e^{-\lambda t} + \sum_{k=1}^{+\infty} \Phi\left(\frac{x - k\mu_J}{\sqrt{k}\sigma_J}x\right)e^{-\lambda t}\frac{(\lambda t))^k}{k!} \end{aligned}$$

where Φ is the c.d.f. of a standard normal distribution.
c) $E(Y_t) = \lambda\mu_J$, $Var(Y_t) = \lambda\sigma_J^2$, $Sk_{Y_t} = 0$, $K_{Y_t} = 3\lambda$.

EXAMPLE 2.26. *Claims of an insurance company*
Let $(N_t)_{t\geq 0}$ be the number of claims of certain type in an insurance company in an interval $[0, t]$, described by a Poisson process of parameter $\lambda > 0$. Let $(X_k)_{k\in\mathbb{N}}$ be a sequence of i.i.d. random variables representing the size of the claims(in thousand of dollars). The total claim the company faces on $[0, t]$ it defined by the compound Poisson process :

$$T_t = \sum_{k=1}^{N_t} X_k$$

(a) Assume that the claims have an exponential probability distribution, i.e. $X_k \sim exp(-\mu)$. Find the probability distribution of the total claim T_t when the number of claims in $[0, t]$ is equal to n($N_t = n$).
(b) Find the conditional p.d.f. $f_{T_t}(x/N_t = n)$.
(c) Find the unconditional c.d.f. $F_{T_t}(x)$.

Solution

(a) We prove the statement using the one-to-one correspondence between c.d.f. and characteristic functions. The later can be computed:

$$\begin{aligned}\varphi_{T_t/N_t=n}(u) &= \prod_{k=1}^{n}\varphi_{X_k}(u)=\prod_{k=1}^{n}\frac{\mu}{(\mu-iu)}\\ &= \frac{\mu^n}{(\mu-iu)^n}=\frac{1}{(1-i\frac{1}{\mu}u)^n}\end{aligned}$$

In the later we recognize a Gamma distribution of parameters n and μ. Also called an Erlang distribution when, as it is the case, n is a non-negative integer number.

(b) By part (a) the p.d.f. is:

$$f_{T_t/N_t}(x/n)=\frac{\mu^n}{(n-1)!}x^{n-1}e^{-\mu x},\ x>0,n\geq 1$$

and zero otherwise.
For $n=0$ the random variable T_t is discrete with mass concentrated at 0.
Hence, its conditional c.d.f. is:

$$F_{T_t/N_t}(x/n) = \int_0^x f_{T_t/N_t}(y/n)\,dy=\Gamma(\mu,n,x)$$

The special function $\Gamma(\mu,n,x)$ is known as the incomplete gamma function.

(c) As in the previous example:

$$\begin{aligned}F_{T_t}(x) &= \sum_{k=0}^{+\infty}P(Y_t\leq x/N_t=k)P(N_t=k)\\ &= \sum_{k=0}^{+\infty}F_{Y_k/N_t}(x/k)e^{-\lambda t}\frac{(\lambda t))^k}{k!}\\ &= 1_{\{\mathbb{R}_+\}}(x)e^{-\lambda t}+\sum_{k=1}^{+\infty}\Gamma(\mu,k,x)e^{-\mu x}e^{-\lambda t}\frac{(\lambda t)^k}{k!}\end{aligned}$$

An algorithm to simulate a Compound Poisson process combines any of the algorithms to simulated a Poisson process with the simulation of the random jumps. It is summarized as follows:

Algorithm 2.5(Simulating a compound Poisson process on $[0,T]$**)**

(1) Initialization: Set values λ,μ_J,σ_J,T.
(2) Simulate a Poisson process N_t on $[0,T]$ following any of the algorithms 2.2-2.4.
(3) Simulate N_t i.i.d. variables X_k, $k=1,2,\ldots,N_t$ with a given distribution.
(4) Compute $Y_t=\sum_{k=1}^{N_t}X_k$ at selected points t's.

EXAMPLE 2.27. To simulate a compound Poisson process with normal jumps we adapt algorithm 2.5 to:

Algorithm 2.6(Simulating a compound Poisson process with normal jumps)

(1) Simulate a Poisson process N_t on $[0,T]$.
(2) Simulate N_t i.i.d. variables X_k, $k=1,2,\ldots,N_T$ with a given distribution.
(3) Compute $Y_t=\sum_{k=1}^{N_t}X_k$.

An example of code in MATLAB is the following:

```
function [ p, y]=comppoisson(tsim, lambda, mu,sigma, nrep, delta)
%simulate poisson process trajectories,normal jumps starting at zero
%output y: trajectories
```

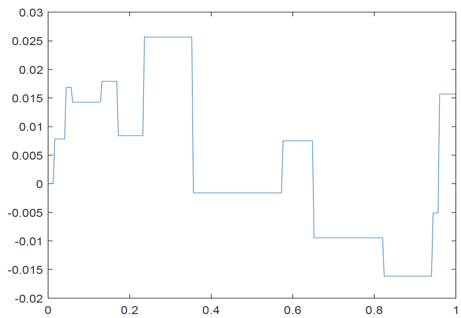


FIGURE 9. Simulation of the trajectory of a compound Poisson process with Gaussian jumps

```
% inputs: intensity parametrer, nosimulations, simulation time,
% delta:simulation interval
% mu , sigma: parameters jumps
%tsim=1; lambda=5; mu=0.01; sigma=0.20; nrep=3; delta=1/365;
nint=fix(tsim/delta); %nro of intervals
NEV=zeros(nrep,nint);
CNEV=zeros(nrep,nint);
N=poissrnd(lambda*tsim,1,nrep);
y=zeros(nrep, nint);
p=zeros(nrep,nint);
ax=0:delta:1;
for k=1:nrep
JT=sort(tsim*rand(1,N(k)));
for j=1:nint
NEV(k,j)=length(JT((JT¡j*delta)& (JT¿=(j-1)*delta)));
if NEV(k,j)==0
CNEV(k,j)=0;
else
CNEV(k,j)=sum(normrnd(mu,sigma,1,NEV(k,j)));
end
end
p(k,:)=cumsum(NEV(k,:));
y(k,:)=cumsum(CNEV(k,:));
end
% plot(ax, [0 y])
end
```

In figure 9 a simulation on the interval $[0, 1]$ of three trajectories of a compound Poisson process with normal jump sizes and parameters $\lambda = 10; \mu = 0.01; \sigma = 0.20$ is shown.

2.3. Continuous-Time Markov Processes

The dynamic of a Poisson process can be generalized by allowing jumps of any integer size. For example, studies in biological populations consider changes of size 1 and -1 on small interval of time while keeping the Markovian property. The latter, for continuous-time processes reads as follows:

Definition 2.28. *Continuous-time Markov processes*
A continuous-time stochastic process $(X_t)_{t\geq 0}$ with values in a set $\mathbb{E} \subset \mathbb{N}$ defined on the probability space $(\Omega, \mathcal{A}, P)$ is a Markov process if for all $n \in N$, $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$, $i_1, i_2, \ldots, i_n \in \mathbb{E}$ we have:

$$P\left\{X_{t_n} = i_n / X_{t_{n-1}} = i_{n-1}, \ldots, X_{t_1} = i_1\right\} = P\left\{X_{t_n} = i_n / X_{t_{n-1}} = i_{n-1}\right\}$$

As in the case of Markov chains transition probabilities are given by:

$$p_{ij}(s,t) = P(X_t = j / X_s = i), \text{ for } 0 \leq s \leq t, i, j \in \mathbb{E}$$

Similarly, it is said a Markov process is stationary if $p_{ij}(s,t) = p_{ij}(t-s)$. In this case we write:

$$p_{ij}(t) = P(X_{t+s} = j / X_s = i), \text{ for } s, t \geq 0$$

Given a Markov process $(X_t)_{t\geq 0}$ there is an embedded Markov chain associated with the former. Namely:

Definition 2.29. Let $(X_t)_{t\geq 0}$ be a Markov process and $(T_n)_{n\in\mathbb{N}}$ the sequence of random times at which the system changes its position. The process $(X_{T_n})_{n\in\mathbb{N}}$ is a Markov chain called *embedded Markov chain* associated with the process.

Due to the topology of the index space $\mathbb{R}_+$ the one-step transition probabilities do not play the central role as in the Markov chains. Instead, we look at the change of the transition probabilities in a neighbourhood of the origin.

Definition 2.30. *Instantaneous transition intensity*
For a stationary Markov process we define the instantaneous transition intensity from state i to state j as:

$$q_{ij} = \lim_{h\to 0^+} \frac{p_{ij}(h) - p_{ij}(0)}{h} = \lim_{h\to 0^+} \frac{p_{ij}(h) - \delta_{ij}}{h} \tag{2.4}$$

where δ_{ij} is the usual Kronecker's number, i.e. equal to one if $i = j$ and zero otherwise.

When $i \neq j$ equation (2.4) can be written:

$$p_{ij}(h) = P(X_{t+h} = j / X_t = i) = q_{ij} h + o(h)$$

On the other hand when $i = j$ the probability of exiting i on an interval $[t, t+h)$ is:

$$1 - p_{ii}(h) = P(X_{t+h} = j / X_t = i) = -q_{ii} h + o(h)$$

If the process has n possible states we build the matrix of transition intensities:

$$A = \begin{pmatrix} q_{11} & q_{12} & \ldots & q_{1n} \\ & & \ldots & \\ q_{n1} & q_{n2} & \ldots & q_{nn} \end{pmatrix}$$

When the process has infinitely many states we get a matrix with infinitely many rows and columns. Form the relation:

$$\sum_{i\in\mathbb{E}} p_{ij}(h) = \sum_{j\neq i} p_{ij}(h) + p_{ii}(h) = 1$$

and after dividing by h and taking limits the transition intensities are related by:

$$q_{ii} = -\sum_{j \neq i} q_{ij}, \; i \in \mathbb{E} \tag{2.5}$$

REMARK 2.31. (1) In view of the expressions of q_{ij} and q_{ii} above the meaning of the instantaneous transition intensity q_{ij} is the following: As in a small interval $[t, t+h)$, the function $o(h)$ becomes negligible compared with h the probability of going from i to j on the interval $[t, t+h)$ is proportional to t, with a transition rate equal to q_{ij}.

(2) Similarly, the interpretation of q_{ii} is that for small h the probability of exiting i, given by $1 - p_{ii}(h)$ is proportional to h, with rate of exit equals to $-q_{ii}$.

EXAMPLE 2.32. *Examples of Markov processes*

(1) *Poisson process*
The Poisson process is a simple particular case of a Markov process, reflecting the fact that in short periods of time only transitions from i to $i+1$ are allowed. From the first order Taylor expansion $exp(-\lambda h) = 1 - \lambda h + o(h)$ we have:

$$\begin{aligned}
P(N_{t+h} = k+1/N_t = k) &= \lambda h exp(-\lambda h) = \lambda h - \lambda^2 h^2 + \lambda h o(h) = \lambda h + o(h) \\
P(N_{t+h} = k/N_t = k) &= exp(-\lambda h) = 1 - \lambda h + o(h) \\
P(N_{t+h} = k+l/N_t = k) &= exp(-\lambda h)\frac{(\lambda h)^k}{k!} = o(h), \text{ for } |l| > 1
\end{aligned}$$

Its instantaneous transition intensities are:

$$q_{k,k+1} = \lambda, q_{kk} = -\lambda, \text{and } q_{kj} = 0, j \neq k, k+1$$

By an abuse of notations we use the symbol $o(h)$ for possible different functions, provided the property holds.
Since a Poisson process has independent increments it can be written, for $s < t$, as $X_t = (X_t - X_s) + X_s$. As the increment $X_t - X_s$ is independent from the positions of the system prior to s, intuitively it is clear that the Markovian property (2.1) holds.
Therefore, the matrix of instantaneous transition intensities has infinitely many rows and columns. it has the shape:

$$A = \begin{pmatrix} -\lambda & \lambda & 0 & \dots \\ 0 & -\lambda & \lambda & \dots \\ & & & \dots \\ 0 & 0 & \dots & \end{pmatrix}$$

(2) *Birth-and-death processes*
Let $(X_t)_{t \geq 0}$ be a continuous-time Markov process, where X_t is the number of individuals in a population at time t. It is a *birth-and-death* process if it has the following transition probabilities:

$$\begin{aligned}
P(X_{t+h} = k+1/X_t = k) &= \lambda_k h + o(h) \\
P(X_{t+h} = k-1/X_t = k) &= \mu_k h + o(h) \\
P(X_{t+h} = k/X_t = k) &= -(\mu_k + \lambda_k)h + o(h) \\
P(X_{t+h} = k+l/X_t = k) &= o(h), \text{ for } |l| > 1
\end{aligned}$$

Its instantaneous transition intensities are:

$$q_{k,k+1} = \lambda_k, q_{k,k-1} = \mu_k, q_{kk} = -(\lambda_k + \mu_k), \text{and } q_{kj} = 0, j \neq k, k-1, k+1$$

In particular, a linear birth-and-death process has $\lambda_k = \lambda k$ and $\mu_k = \mu k$. Its matrix of instantaneous transition intensities is:

$$A = \begin{pmatrix} -\lambda & \lambda & 0 & \dots & \\ \mu & -(\lambda+\mu) & \lambda & \dots & \\ 0 & 2\mu & -2(\lambda+\mu) & 2\lambda & \dots \\ & & \dots & & \end{pmatrix}$$

A rationale for the instantaneous transition intensity rates in a linear birth-and-death

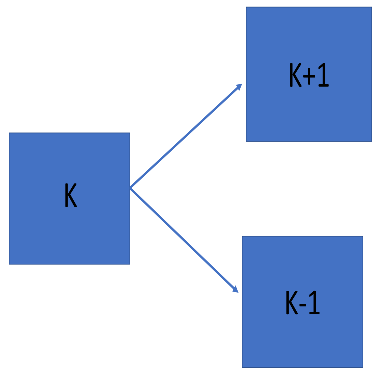


FIGURE 10. Possible transitions in a birth-and-death model

process are given in solved problem 2.10.

(3) *M/M/n Queuing Systems*
Consider a situation where customers arrive to the system following a Poisson process of intensity λ and are immediately served if one of n servers in the system is idle, otherwise

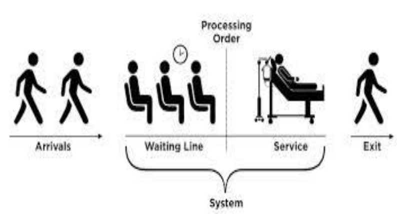


FIGURE 11. A queue system M/M/1

they have to wait in line until being served. See figure 11 for patients arriving to an emergency stomatological clinic. The time of service is an exponentially distributed random variable with parameter μ, independent of the number of individuals in the system. Let $(X_t)_{t\geq 0}$ be a Markov process, where X_t is the number of customers in the system at time t. It is a birth-and-death process taking values on the set $\{0, 1, 2, \dots, \}$ with:

$$\begin{aligned} P(X_{t+h} = k+1/X_t = k) &= e^{-\lambda t}\lambda h = \lambda h + o(h) \\ P(X_{t+h} = k-1/X_t = k) &= \mu k h + o(h), \;\; 1 \leq k \leq n \\ P(X_{t+h} = k-1/X_t = k) &= \mu n h + o(h), \;\; k > n \end{aligned}$$

The second and third equalities follow after noticing that, if $T_1, T_2, \dots, T_k$ define the times until the completion of the service of the k-th individuals being served and

$$T_{m,k} = min(T_1, T_2, \dots, T_k)$$

is the time to the first completion of the service (i.e. first exit from the system) then:

$$\begin{aligned} P(T_{m,k} \geq h) &= P(T_1 > h, T_2 > h, \dots, T_k > h) = P(T_1 > h)P(T_2 > h)\dots P(T_k > h) \\ &= e^{\mu k x} \end{aligned}$$

A series of ordinary differential equations for the transition probabilities is obtained from Chapman-Kolmogorov equations. Two variants of the result are presented above.

Theorem 2.33. *Kolmogorov Equations*
A Markov process $(X_t)_{t\geq 0}$ with values in a finite set $\mathbb{E} \subset \mathbb{N}$, defined on the probability space $(\Omega, \mathcal{A}, P)$ and with transition intensities q_{ij}. Then, the transition probabilities verify:
i)Kolmogorov's backward equation.

$$\frac{dp_{ij}(t)}{dt} = \sum_{l\in\mathbb{E}} q_{il}p_{lj}(t), \;\; i, j \in \mathbb{E}$$

In matrix form:

$$\frac{dP(t)}{dt} = AP(t) \tag{2.6}$$

where $P(t)$ is a matrix with components $p_{ij}(t)$.
Initial conditions are $p_{ij}(0) = \delta_{ij}$.
ii) Kolmogorov's forward equation.

$$\frac{dp_{ij}(t)}{dt} = \sum_{l\in\mathbb{E}} p_{il}(t)q_{lj}, \;\; i, j \in \mathbb{E}$$

Or in matrix form:

$$\frac{dP(t)}{dt} = P(t)A'$$

where A' is transpose of matrix A. Initial conditions are $p_{ij}(0) = \delta_{ij}$.
The solution of equation (2.6) is:

$$P(t) = e^{At} \tag{2.7}$$

where the exponential of a matrix is defined as:

$$e^{At} = \sum_{k=0}^{+\infty} \frac{(At)^k}{k!}$$

PROOF. We refer to figure 12 picturing the decomposition of the event $[X_{t+h} = k]$(bottom line for the Backward Kolmogorov equation).
Thus, for any $t, h \geq 0$, $i, j \in \mathbb{E}$:

$$\begin{aligned}
p_{ij}(t+h) &= \sum_{l\in\mathbb{E}} P[X_{t+h} = j/X_h = l, X_0 = i]P[X_h = l/X_0 = i] = \sum_{l\in\mathbb{E}} p_{il}(h)p_{lj}(t) \\
&= \sum_{l\neq i} p_{lj}(t)p_{il}(h) + p_{ij}(t)p_{ii}(h) \\
&\quad \text{Subtracting } p_{ij}(t) \text{ and dividing by h} \\
\frac{p_{ij}(t+h) - p_{ij}(t)}{h} &= \sum_{l\neq i} p_{lj}(t)\frac{p_{il}(h)}{h} + p_{ij}(t)\frac{p_{ii}(h)-1}{h} \\
\frac{dp_{ij}(t)}{dt} &= \sum_{l\in\mathbb{E}} q_{il}p_{lj}(h)
\end{aligned}$$

where the last equality is obtained after taking the limit when $h \to 0$ in the third equation above.
For part ii), we have:

$$\begin{aligned}
p_{ij}(t+h) &= \sum_{l\in\mathbb{E}} P[X_{t+h} = j/X_t = l, X_0 = i]P[X_t = l/X_0 = i] = \sum_{l\in\mathbb{E}} p_{lj}(h)p_{il}(t) \\
&= \sum_{l\neq i} p_{lj}(h)p_{il}(t) + p_{jj}(h)p_{ij}(t) \\
\frac{p_{ij}(t+h) - p_{ij}(t)}{h} &= \sum_{l\neq i} p_{il}(t)\frac{p_{lj}(h)}{h} + p_{ij}(t)\frac{p_{jj}(h)-1}{h} \\
\frac{dp_{ij}(t)}{dt} &= \sum_{l\in\mathbb{E}} q_{lj}p_{il}(t)
\end{aligned}$$

The proof of equation 2.7 is differed to problem 33. □

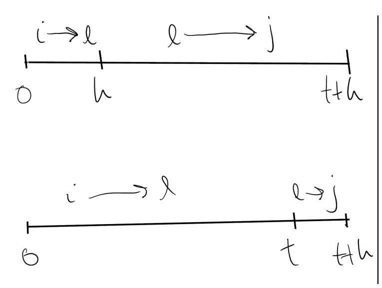


FIGURE 12. Event decomposition on the interval $[0, t)$.

The *master equation* for the transition probabilities in a Markov process is a variant of the Kolmogorov equations when initial conditions are known and instead of transition probabilities the focus is on the state probabilities $p_k(t) = P(X_t = k)$.

Theorem 2.34. *Master equations*
Let $(X_t)_{t\geq 0}$ *be a Markov process with values in a set* $\mathbb{E} \subset \mathbb{N}$*, defined on the probability space* $(\Omega, \mathcal{A}, P)$ *and starting a* $X_0 = a$*. Its instantaneous transition intensities* q_{ij} *are such that* $q_{k,k+|l|} = 0$ *for* $|l| > M \geq 0, k \in \mathbb{E}$*. Then,*

$$\frac{dp_k(t)}{dt} = \sum_{l=-M}^{M} p_{k+l}(t) q_{k+l,k} - q_{kk} p_k(t) \tag{2.8}$$

with initial conditions $p_k(0) = \delta_{k,a}$*.*

Now we discuss how to obtain the stationary probabilities of a Markov process. Recall, from section 2.1 that they are defined as:

$$\pi_j = \lim_{t\to\infty} p_{ij}(t)$$

As in its discrete-time counterpart the quantity π_j is interpreted as the probability that the system stays at state j after a long time. Intuitively, it does not depend on the initial value i because of the Markov property .
If the stationary probabilities exist they must satisfy the system:

$$\begin{aligned} \sum_{l\in\mathbb{E}} q_{lj}\pi_l &= 0 \\ \sum_{l\in\mathbb{E}} \pi_l &= 1 \end{aligned}$$

It is enough to take the limit when $t \to \infty$ in the Kolmogorov's Forward equation, considering also $\frac{p_{ij}(t)}{dt} = 0$.

example 2.35. Consider a birth-and-death process with $\lambda_i > 0, i \geq 0$, $\mu_i > 0, i \geq 1$ and zero otherwise.
The intensity transitions are $q_{i,i+1} = \lambda_i, i \geq 0, q_{i,i-1} = \mu_i, \geq 1, q_{i,i} = -(\lambda_i + \mu_i), i \geq 0$, and zero otherwise.
Combining Kolmogorov master equations with the specific transition verify leads to:

$$\begin{aligned} (\lambda_i + \mu_i)\pi_i &= \lambda_{i-1}\pi_{i-1} + \mu_{i+1}\pi_{i+1}, \ \ i \geq 1 \\ \lambda_0 \pi_0 &= \mu_1 \pi_1 \end{aligned}$$

In order to solve the system we write:

$$\begin{aligned} \pi_1 &= \frac{\lambda_0}{\mu_1}\pi_0 \\ \pi_2 &= \frac{\lambda_1\lambda_0}{\mu_2\mu_1}\pi_0 \\ &\cdots \\ \pi_n &= \frac{\lambda_{n-1}\ldots\lambda_0}{\mu_n\ldots\mu_1}\pi_0 := a_n\pi_0 \end{aligned}$$

Now, from the condition $\sum_{l\in\mathbb{E}} \pi_l = 1$ we can compute π_0 as:

$$\pi_0 = \frac{1}{1 + \sum_{n=1}^{\infty} a_n}$$

and

$$\pi_n = \frac{a_n}{1 + \sum_{n=1}^{\infty} a_n}$$

This section is withheld pending source approval.

This section is withheld pending source approval.

This section is withheld pending source approval.

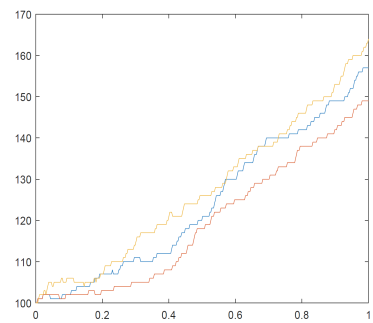


FIGURE 13. Three simulated trajectories of a birth- and-death process are shown. The parameters of the simulation are $T = 1; \lambda = 0.5; \mu = 0.1; m = 1; delta = 1/365; X_0 = 100$

2.5. Applications: Stochastic Epidemic

The spread of an epidemic in a population has been studied form a mathematic perspective. Deterministic and stochastic models have been proposed since the pioneer work of Kemarck and Mckendrick(1927).
In figure **??**) in chapter 1 a time series of weekly new cases, Jan-Aug. 2020 in Canada.

2.5.1. Deterministic S-I-S. Transitions in a S-I-S epidemic model are shown in figure 14 .
The possible transitions and its rates are:

$$\begin{aligned}(S, I) &\longrightarrow (S-1, I+1),\ \lambda(t)SI \\ (S, I) &\longrightarrow (S+1, I-1),\ \beta(t)I\end{aligned}$$

where the first equation represents the event that a susceptible individual is infected and its rate and the second is the event of an infected becoming susceptible. Notice that it is implicitly assumed that the population is homogeneously mixed although with infection and recovery rates, given respectively by $\lambda(t)$ and $\beta(t)$, eventually depend on t. Another implicit assumption is the nonexistence of an immunity period. An infected individual becomes immediately susceptible after recovery.
Moreover, it is also assumed that the number of individuals in the population remains constant upon time. Hence:

$$S_t + I_t = N,\ t \geq 0$$

where N is the population size, S_t and I_t a are the number of susceptible and infected people in the population respectively.
Notice that because the relation $S_t + I_t = N$ an equation for the number of susceptible follows for the corresponding equation for the number of infected.
We focus on the later. The instantaneous rates above to the ordinary differential equation system:

$$\frac{dI_t}{dt} = \lambda(t)S_tI_t \tag{2.13}$$

$$\frac{dS_t}{dt} = -\lambda(t)S_tI_t + \beta(t)I_t \tag{2.14}$$

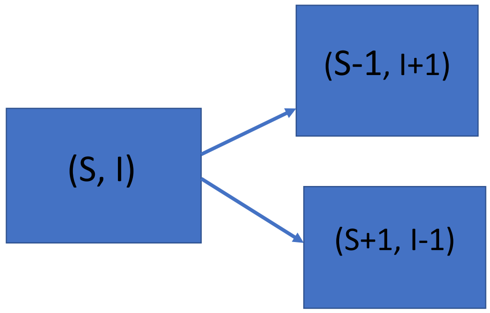


Figure 14. Transitions in a S-I-S models

with initial conditions $S_0 = N - a, I_0 = a$.
Equation (2.13) is rewritten as:

$$\frac{dI_t}{dt} = \lambda(t) I_t (N - I_t)$$

Therefore:

$$\frac{dI_t}{I_t(N - I_t)} = \lambda(t) dt$$

Separating in simple fractions and integrating we have:

$$\begin{aligned} \int_0^t \frac{dI_t}{I_t(N - I_t)} &= \frac{1}{N} \int_0^t \left(\frac{1}{I_t} + \frac{1}{N - I_t} \right) dI_t \\ = \quad & \log \left(\frac{I_t}{N - I_t} \right) - \log \left(\frac{I_0}{N - I_0} \right) \\ = \quad & N \int_0^t \lambda(t) dt \end{aligned}$$

Hence, the solution is:

$$I_t = \frac{I_0 N}{I_0 + (N - I_0) \exp(-\int_0^t \lambda(t) dt)} \tag{2.15}$$

In figure the curve of the number of infected for $\lambda(t) = \lambda$.
From equation (2.15) some features are derived in the proposition above.

Proposition 2.38. For a deterministic S-I-S model:

i) $I_t \longrightarrow N$ when $t \longrightarrow +\infty$.

ii) Hitting time $T_j = \inf\{t/I_t \geq n\}$ solves $I_t = j$ for $t \geq 0$ where I_t is given by equation (2.15).
iii) Epidemic curve (case λ constant):

$$\frac{dI_t}{dt} = \frac{\lambda I_0(N - I_0))}{cosh(\lambda N t) + (1 - 2\frac{I_0}{N})sinh(\lambda N t))^2}$$

2.5.2. Deterministic S-I-R model. The model susceptible-infected-recovered (S-I-R) consider a recovery period for individuals after being infected. Transitions are shown in figure 15 of a S-I-R model. Transition rates in a S-I-R model:

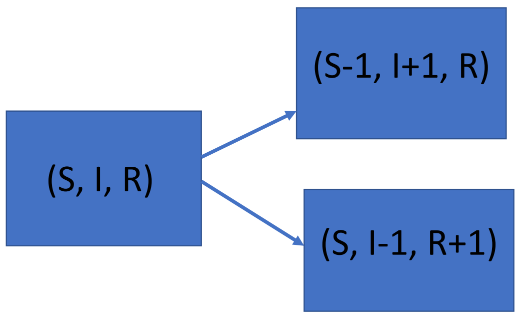


FIGURE 15. Transitions in a S-I-R model

$$\begin{array}{rcl} (S, I, R) & \longrightarrow & (S-1, I+1, R),\ \lambda(t)SI \\ (S, I, R) & \longrightarrow & (S, I-1, R+1),\ \gamma(t)I \end{array}$$

It leads to the following system of ordinary differential equations:

$$\frac{dS_t}{dt} = -\lambda(t)S_t I_t \tag{2.16}$$

$$\frac{dR_t}{dt} = \gamma(t)I_t \tag{2.17}$$

$$\frac{dI_t}{dt} = \lambda(t)S_t I_t - \gamma(t)I_t \tag{2.18}$$

with initial conditions $S_0 = N - a_1, I_0 = a_1, R_0 = 0$, where:
N is the (constant)population size, and S_t, I_t and R_t are the number of susceptible, infected and recovery respectively.
Formally, we divide equation (2.16) by (2.17):

$$\frac{dS_t}{dR_t} = -\frac{1}{\rho(t)}S_t \tag{2.19}$$

whose solution is:

$$S_t = S_0 exp(-\int_0^t \rho^{-1}(s)ds\, R_t) \tag{2.20}$$

with $\rho(t) = \frac{\gamma(t)}{\lambda(t)}$ the relative rate of the infection at time t.
Next, divide equation (2.18) by (2.16) to get:

$$\frac{dI_t}{dS_t} = -1 + \frac{\rho(t)}{S_t}$$

Hence, combining equation (2.20) with (2.17):

$$\frac{dR_t}{dt} = \gamma(t)(N - S_t - R_t) = \gamma(t)[N - R_t - S_0 exp(-\int_0^t \rho^{-1}(s)ds\, R_t)]$$

To simplify we assume the rates $\gamma(t) = \gamma$ and $\lambda(t) = \lambda$ are constant. From a second order Taylor expansion $e^{-cx} \approx 1 - cx + \frac{1}{2}c^2x^2$:

$$\frac{dR_t}{dt} = \gamma(t)(N - S_0 + (\frac{S_0}{\rho}R_t - \frac{S_0}{2\rho^2}R_t^2))$$

whose approximated solution is:

$$R_t \approx \frac{\rho^2}{S_0}(\frac{\rho}{S_0} - 1) + \frac{\alpha\rho^2}{S_0} tanh(\frac{1}{2}\gamma\alpha t - \mu)$$

where:

$$\begin{aligned} \mu &= tanh^{-1}\left(\frac{\frac{S_0}{\rho} - 1}{\alpha}\right) \\ \alpha &= \left(\frac{2S_0}{\rho^2}(N - S_0) + (\frac{\rho}{S_0} - 1)^2\right)^{\frac{1}{2}} \end{aligned}$$

Moreover,

$$R_\infty := \lim_{t\to+\infty} R_t = \frac{\rho^2}{S_0}\left(\frac{\rho}{S_0} - 1 + \alpha\right) \tag{2.21}$$

Comparing with the deterministic S-I-S model we can see that the epidemic will stop without infecting all individuals in the population. This result is known as the *threshold theorem.* (Kemarck and Mckendrick(1927))
The first event reflects the infection of one individual. The equality states the probability of a single infections on an interval $[t, t+h)$ is proportional to the number of infected persons in the population ($I_t = k$) times the number of susceptible individuals, where λ is the infection rate. Transitions $k \to k-1$ reflects that an infected individual becomes healthy, and susceptible again, on the interval $[t, t+h)$. Transitions involving 2 or more individuals on a small interval are $o(h)$.
Figure 16 the possible transitions from state k are shown.

2.5.3. Discrete-time stochastic models. Some discrete-time models have been proposed to describe the spread of infectious diseases in a population. See for example Becker(1989). Reed-Frost and Greenwood models are examples of these models. The Galton-Watson branching model, introduced in chapter one, is another stochastic model that can be used to model early stages of an epidemic.
Let $(S_t, I_t)_{t\in\mathbb{N}}$ par of susceptible and new infective individuals at time t (or generation t). As in

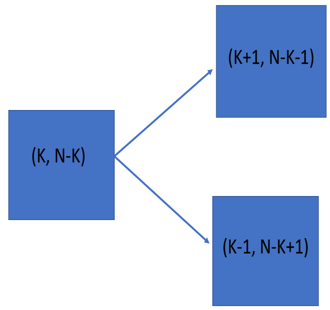


FIGURE 16. Possible transitions from state k.

the subsection above it is assumed the total population is constant upon time and the pair is a stationary Markov chain. Then:

$$P(I_{t+1} = i_{t+1}/S_t = s_t,\ I_t = i_t) = \binom{s_t}{i_{t+1}}(1-q(i_t))^{i_{t+1}}q(i_t)^{s_t - i_{t+1}}$$

The interpretation of the equation above is the following: At time t there are s_t susceptible and i_t infected individuals. A susceptible individual at time t has a probability of an *infectious contact* before time $t+1$ equal to $1-q(i_t)$. Under the assumption of homogeneous and independent contacts it leads to a number of infected people at $t+1$ following a binomial distribution, where the number of experiments is s_t and the probability of success at a single trial is $1-q(i_t)$.
The choice $q(i_t) = q$ leads to the Greenwood model while the choice of $q(i_t) = q^{i_t}$ leads to the Reed-Frost model.

2.5.4. S-I and S-I-S stochastic epidemic model. Consider $(I_t, S_t)_{t\geq 0}$ a Markov processes describing respectively the number of infected and susceptible individuals in a population at time t, as in the deterministic case with a constant total number of individuals in the population $S_t + I_t = N$ and initial number of infective at time zero is $I_0 = a$.
Similarly to the birth-and-death model in an interval $[t, t+h)$ consider the probabilities:

$$\begin{aligned}
P(I_{t+h} = k+1, S_{t+h} = l-1/I_t = k, S_t = l) &= \lambda(t)klh + o(h)\\
P(I_{t+h} = k, S_{t+h} = l/I_t = k, S_t = l) &= 1 - \lambda(t)klh + o(h)\\
P(I_{t+h} = k+m, S_{t+h} = l-m, /I_t = k, S_t = l) &= o(h),\ \text{for } |m| > 1
\end{aligned}$$

The first event reflects the infection of one individual. The equality states that the probability of a single infections in the interval $[t, t+h)$ is proportional to the number of infected persons in the population $(I_t = k)$ times the number of susceptible individuals, where $\lambda(t)$ is the infection rate. Transitions involving two or more individuals on a small interval are of order $o(h)$, i.e. negligible if h are small.
Let's denote by:

$$P(S_t = l, I_t = k) = p_{kl}(t)$$

the state probabilities of the process.
From the equations above and the Law of Total Probability it easily follows that:

$$p_{kl}(t+h) = p_{k+1,l-1}(t)(\lambda(t)(k+1)(l-1)h + o(h)) + p_{k,l}(t)(1-\lambda(t)klh + o(h)) + o(h),\ 1 \leq k, l \leq N-1$$

with some straightforward conditions at the borders $k, l = 0, N$.
The master equation system is obtained by taking the passage to the limit of the incremental quotient. Hence:

$$\frac{dp_{kl}(t)}{dt} = p_{k+1,l-1}(t)\lambda(t)(k+1)(l-1) - p_{k,l}(t)\lambda(t)kl, \ 1 \leq k, l \leq N-1,$$

Since the relation $S_t + I_t = N$ the focus is on one of the two processes, say the infected number of individuals $(I_t)_{t\geq 0}$, with $p_l(t) = P(I_t = l)$. Therefore:

$$\begin{aligned} p_l(t+h) &= p_{l-1}(t)(\lambda(t)(N-l+1)(l-1)h + o(h)) + p_{l+1}(t)(\gamma(t)(l+1)h + o(h)) \\ &+ p_l(t)(1 - \lambda(t)(N-l)lh - \gamma(t)lh + o(h)) + o(h), \ 2 \leq l \leq N-1 \end{aligned}$$

Notice that $l = 0$ is an absorbing state:

$$\begin{aligned} p_0(t+h) &= p_1(t)(\gamma(t)h + o(h)) + p_0(t) + o(h) \\ p_1(t+h) &= p_2(t)(\gamma(t)2h + o(h)) + p_1(t)(1 - \gamma(t)h - \lambda(t)h + o(h)) + o(h) \end{aligned}$$

while:

$$p_N(t+h) = p_{N-1}(t)(\lambda(t)(N-1)h + o(h)) + p_N(t)(1 - \gamma(t)N + o(h)) + o(h)$$

Hence the master equations become:

$$\begin{aligned} \frac{dp_l(t)}{dt} &= \lambda(t)(N-l+1)(l-1)p_{l-1}(t) + \gamma(t)(l+1)p_{l+1}(t) \\ &- (\lambda(t)(N-l)l + \gamma(t)l)p_l(t), \ 2 \leq l \leq N-1 \\ \frac{dp_0(t)}{dt} &= \gamma(t)p_1(t) \\ \frac{dp_1(t)}{dt} &= 2\gamma(t)p_2(t) - (\lambda(t) + \gamma(t))p_1(t) \\ \frac{dp_N(t)}{dt} &= \lambda(t)(N-1)p_{N-1}(t) - \gamma(t)Np_N(t) \end{aligned}$$

Let $G_I(t,s)$ be the p.g.f. of the number of infected individuals at time t. Then:

$$\begin{aligned} \frac{\partial G_I(t,s)}{\partial t} &= \frac{dp_0(t)}{dt} + \frac{dp_1(t)}{dt}s + \frac{dp_N(t)}{dt}s^N + \sum_{l=2}^{N-1} \frac{dp_l(t)}{dt}s^l \\ &= \gamma(t)p_1(t) + 2\gamma(t)p_2(t)s - \gamma(t)p_1(t)s - \lambda(t)p_1(t)s + \lambda(t)(N-1)p_{N-1}(t)s^N - \gamma(t)Np_N(t)s^N \\ &+ \lambda(t)\sum_{l=2}^{N-1}(N-l+1)(l-1)p_{l-1}(t)s^l + \gamma(t)\sum_{l=2}^{N-1}(l+1)p_{l+1}(t)s^l \\ &- \lambda(t)\sum_{l=2}^{N-1}(N-l)lp_l(t)s^l - \gamma(t)\sum_{l=2}^{N-1} lp_l(t)s^l \end{aligned}$$

Regrouping terms:

$$\begin{aligned} \frac{\partial G_I(t,s)}{\partial t} &= \lambda(t)s\sum_{l=1}^{N}(N-l)lp_l(t)s^l + \gamma(t)\frac{1}{s}\sum_{l=0}^{N} lp_l(t)s^l \\ &- \lambda(t)\sum_{l=2}^{N}(N-l)lp_l(t)s^l - \gamma(t)\sum_{l=0}^{N-1} lp_l(t)s^l \end{aligned}$$

Leading to the p.d.e. :

$$\frac{\partial M_I(t,u)}{\partial t} = \lambda(t)(e^u - 1)(N\frac{\partial M_I(t,u)}{\partial u} - \frac{\partial^2 M_I(t,u)}{\partial u^2})$$

Consider now a S-I-S with the following assumptions:

$$\begin{aligned}
P(I_{t+h} = k+1, S_{t+h} = l-1/I_t = k, S_t = l) &= \lambda(t)klh + o(h)\\
P(I_{t+h} = k-1, S_{t+h} = l+1/I_t = k, S_t = l) &= \gamma(t)kh + o(h)\\
P(I_{t+h} = k, S_{t+h} = l/I_t = k, S_t = l) &= 1 - (\lambda(t)kl + \gamma(t)k)h + o(h)\\
P(I_{t+h} = k+m, I_{t+h} = l+m, /I_t = k, S_t = l) &= o(h),\ \text{for } |m| > 1
\end{aligned}$$

The S-I-S model, in addition, considers transitions $k \to k-1$ reflecting that an infected individual becomes healthy, and susceptible again, on the interval $[t, t+h)$. There is not immunity in this case.

$$\begin{aligned}
p_{kl}(t+h) &= p_{k+1,l-1}(t)(\lambda(t)(k+1)(l-1)h + o(h)) + p_{k-1,l+1}(t)(\gamma(t)(l+1)h + o(h))\\
&+ p_{k,l}(t)(1 - \lambda(t)klh - \gamma(t)lh + o(h)) + o(h),\ 1 \le k, l \le N-1
\end{aligned}$$

with some straightforward conditions at the borders $k, l = 0, N$.
The master equation system is obtained by taking the passage to the limit of the incremental quotient. Hence:

$$\begin{aligned}
\frac{dp_{kl}(t)}{dt} &= p_{k+1,l-1}(t)\lambda(t)(k+1)(l-1) + p_{k-1,l+1}(t)\gamma(t)(l+1)\\
&- p_{k,l}(t)(\lambda(t)kl + \gamma(t)l),\ 1 \le k, l \le N-1,
\end{aligned}$$

Since the relation $S_t + I_t = N$ the focus is on one of the two processes, say the infected number of individuals $(I_t)_{t\ge 0}$, with $p_l(t) = P(I_t = l)$. Therefore:

$$\begin{aligned}
p_l(t+h) &= p_{l-1}(t)(\lambda(t)(N-l+1)(l-1)h + o(h)) + p_{l+1}(t)(\gamma(t)(l+1)h + o(h))\\
&+ p_l(t)(1 - \lambda(t)(N-l)lh - \gamma(t)lh + o(h)) + o(h),\ 2 \le l \le N-1
\end{aligned}$$

Notice that $l = 0$ is an absorbing state:

$$\begin{aligned}
p_0(t+h) &= p_1(t)(\gamma(t)h + o(h)) + p_0(t) + o(h)\\
p_1(t+h) &= p_2(t)(\gamma(t)2h + o(h)) + p_1(t)(1 - \gamma(t)h - \lambda(t)h + o(h)) + o(h)
\end{aligned}$$

while:

$$p_N(t+h) = p_{N-1}(t)(\lambda(t)(N-1)h + o(h)) + p_N(t)(1 - \gamma(t)N + o(h)) + o(h)$$

Hence the master equations become:

$$\begin{aligned}
\frac{dp_l(t)}{dt} &= \lambda(t)(N-l+1)(l-1)p_{l-1}(t) + \gamma(t)(l+1)p_{l+1}(t)\\
&- (\lambda(t)(N-l)l + \gamma(t)l)p_l(t),\ 2 \le l \le N-1\\
\frac{dp_0(t)}{dt} &= \gamma(t)p_1(t)\\
\frac{dp_1(t)}{dt} &= 2\gamma(t)p_2(t) - (\lambda(t) + \gamma(t))p_1(t)\\
\frac{dp_N(t)}{dt} &= \lambda(t)(N-1)p_{N-1}(t) - \gamma(t)Np_N(t)
\end{aligned}$$

Let $G_I(t,s)$ be the p.g.f. of the number of infected individuals at time t. Then:

$$\begin{aligned}
\frac{\partial G_I(t,s)}{\partial t} &= \frac{dp_0(t)}{dt} + \frac{dp_1(t)}{dt}s + \frac{dp_N(t)}{dt}s^N + \sum_{l=2}^{N-1}\frac{dp_l(t)}{dt}s^l \\
&= \gamma(t)p_1(t) + 2\gamma(t)p_2(t)s - \gamma(t)p_1(t)s - \lambda(t)p_1(t)s + \lambda(t)(N-1)p_{N-1}(t)s^N - \gamma(t)Np_N(t)s^N \\
&+ \lambda(t)\sum_{l=2}^{N-1}(N-l+1)(l-1)p_{l-1}(t)s^l + \gamma(t)\sum_{l=2}^{N-1}(l+1)p_{l+1}(t)s^l \\
&- \lambda(t)\sum_{l=2}^{N-1}(N-l)lp_l(t)s^l - \gamma(t)\sum_{l=2}^{N-1}lp_l(t)s^l
\end{aligned}$$

Regrouping terms:

$$\begin{aligned}
\frac{\partial G_I(t,s)}{\partial t} &= \lambda(t)s\sum_{l=1}^{N}(N-l)lp_l(t)s^l + \gamma(t)\frac{1}{s}\sum_{l=0}^{N}lp_l(t)s^l \\
&- \lambda(t)\sum_{l=2}^{N}(N-l)lp_l(t)s^l - \gamma(t)\sum_{l=0}^{N-1}lp_l(t)s^l
\end{aligned}$$

Leading to the p.d.e.'s :

$$\begin{aligned}
\frac{\partial G_I(t,s)}{\partial t} &= \lambda(t)(s-1)(N\frac{\partial G_I(t,s)}{\partial x} - \frac{\partial^2 G_I(t,s)}{\partial x^2}) \\
&+ \gamma(t)(s^{-1}-1)\frac{\partial G_I(t,s)}{\partial s} \\
\frac{\partial M_I(t,u)}{\partial t} &= \lambda(t)(e^u-1)(N\frac{\partial M_I(t,u)}{\partial u} - \frac{\partial^2 M_I(t,u)}{\partial u^2}) \\
&+ \gamma(t)(e^{-u}-1)\frac{\partial M_I(t,u)}{\partial u}
\end{aligned}$$

2.5.5. A stochastic S-I-R epidemic model. Consider $(I_t, S_t, R_t)_{t\geq 0}$ a Markov processes describing respectively the number of infected and susceptible individuals in a population at time t , with a constant total number of individuals in the population $S_t + I_t + R_t = N$. We focus on the pair $(S_t, I_t)_{t\geq 0}$. The number of removed individuals is then fixed.
On an interval $[t, t+h)$ consider the probabilities:

$$\begin{aligned}
P(S_{t+h} = k-1, I_{t+h} = l+1, /S_t = l, I_t = k) &= \lambda(t)klh + o(h) \\
P(S_{t+h} = k, I_{t+h} = l-1/I_t = l) &= \gamma(t)lh + o(h) \\
P(S_{t+h} = k, I_{t+h} = l/S_t = k, I_t = l) &= 1 - (\lambda kl + \gamma l)h + o(h) \\
P(S_{t+h} = k+n, I_{t+h} = l+j/I_t = k, S_t = l) &= o(h), \text{ for } |n,j| > 1
\end{aligned}$$

Write $p_{kl}(t) = P(S_t = k, I_t = l)$ to get:

$$\begin{aligned}
p_{kl}(t+h) &= p_{k+1,l-1}(t)(\lambda(t)(k+1)(l-1)h + o(h)) + p_{k,l+1}(t)(\gamma(t)(l+1)h + o(h)) \\
&+ p_{kl}(t)(1 - \lambda(t)klh - \gamma(t)lh + o(h)) + o(h),\ 1 \leq k, l \leq N-1 \\
p_{k0}(t+h) &= (\gamma(t)h + o(h))p_{k1}(t+h) \\
p_{0N}(t+h) &= (\lambda(t)(N-1)h + o(h))p_{1,N-1} + (1 - \gamma(t)Nh + o(h))p_{0N}(t) + o(h) \\
p_{N0}(t+h) &= 0
\end{aligned}$$

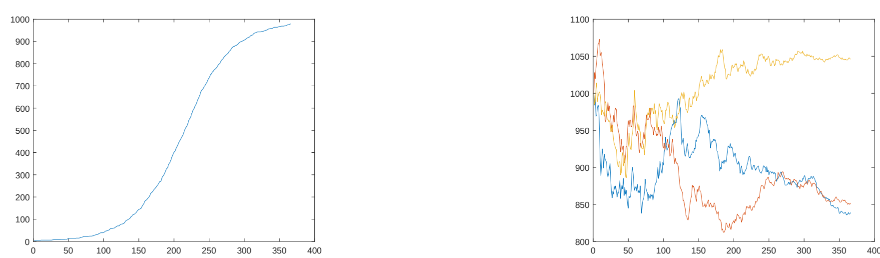
Leading to the master equations:

$$
\begin{aligned}
\frac{dp_{kl}(t)}{dt} &= p_{k+1,l-1}(t)\lambda(t)(k+1)(l-1) + p_{k-1,l+1}(t)\gamma(t)(l+1) \\
&- p_{k,l}(t)(\lambda(t)kl + \gamma(t)l),\ 1 \le k,l \le N-1,
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial G_I(t,s)}{\partial t} &= \lambda(t)(x-1)(N\frac{\partial G_I(t,x)}{\partial x} - \frac{\partial^2 G_I(t,x)}{\partial x^2}) \\
&+ \gamma(t)(x^{-1}-1)\frac{\partial G_I(t,u)}{\partial u} \\
\frac{\partial M_I(t,u)}{\partial t} &= \lambda(t)(e^u-1)(N\frac{\partial M_I(t,u)}{\partial u} - \frac{\partial^2 M_I(t,u)}{\partial u^2}) \\
&+ \gamma(t)(e^{-u}-1)\frac{\partial M_I(t,u)}{\partial u}
\end{aligned}
$$

In figure 17(left) the trajectory of a S-I model is shown. The number of infected people growths until everybody is infected. In figure 17(right) three different trajectories of a S-I-R are shown. The initial number of infected individuals is $I_0 = 1000$.



(a) (b)

FIGURE 17. Left: Trajectory of a S-I model Right:Three different trajectories of a S-I-R

Bibliography

[1] Athreya, K.B., Soumendra N. Lahiri, S.N. Measure Theory and Probability Theory, Springer Science+Business Media, LLC, 2006.

[2] Athreya, Krishna B., Ney, Peter E.(2004). Branching processes. Dover Publications.

[3] Bachelier, L. (1900a), "Theorie de la speculation", Annales Scientifiques de l'Ecole Normale Superieure 3 (17): 21-86.

[4] Becker, N.G. Analysis of Infectious Desease Data. Monnographas on Statistics and Applied Probaiblity, 1989. Chapman and Hill.

[5] Black, F. and M. Scholes (1973) The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, 81(3), 637-654. Statement.

[6] Braumann, Carlos A. Introduction to stochastic differential equations with applications to modelling in biology and finance-Wiley (2019)

[7] Brauer, F. Castillo-Chavez, C. and Feng, Z. Mathematical Models in Epidemiology. Springer. 2019

[8] Billingsley, P. , Probability and Measure, 3rd edn, John Wiley, New York(1995).

[9] Benth, F.E., Vos,L., 2009 A multivariate non-Gaussian stochastic volatility model with leverage for energy markets. Technical Report.Dept. Math.University of Oslo,no. 20, October, 2009. ISSN 0806-2439 .

[10] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics Vol. 31, Issue 3, April 1986, Pages 307-327.

[11] Brzenzniak, Z., Zastawniak, T. Basic Stochastic Processes. A Course through Exercises, Springer-Verlag London, Ltd., London(1999).

[12] Brockwell P.J., Davis R.A., Time Series: Theory and Methods, Springer, New York, 2006.

[13] Cohen, S.N., Elliot, R.J. Stochastic Calculus and Applications, 2nd edition, Birkhauser/Springer, New York, 2015.

[14] Durrett, R. (2004), Probability: Theory and Examples, 3rd edn, Duxbury Press, San Jose, CA.

[15] Glasserman, P 2003 *Monte Carlo Methods in Financial Engineering. Springer-Verlag*

[16] Hull, J 1997 Options, *Futures and Other Derivative Securities. Prentice-Hall*

[17] Dargatz. A Diffusion Approximation for an Epidemic Model.Sonderforschungsbereich 386, Paper 517 (2006). Online unter: http://epub.ub.uni-muenchen.de/

[18] Daley, D. J., Gani, J. Epidemic Modelling. An Introduction-Cambridge University Press (2004).

[19] Ito, K. (1944) Stochastic integral. Proc. Imp. Acad. Tokyo 20 , 519-524.

[20] Jacod,J., Shiriaev , A. (1998) Limit Theorems for Stochastic Processes 2nd Edition.

[21] Jacob, C., Branching Processes: Their Role in Epidemiology. Int. J. Environ. Res. Public Health 2010, 7, 1186-1204; doi:10.3390/ijerph7031204

[22] Doob. Stochastic Processes (1954).

[23] Kolokoltsov V.N.-Markov processes, semigroups and generators. De Gruyter Studies in Mathematics(2011).

[24] Karatzas, I., Shreve, S.E. Brownian Motion and Stochastic Calculus, 2nd edition, Springer-Verlag, New York, 1991.

[25] Kallenberg, O. Foundations of Modern Probability, 2nd edition, Springer- Verlag, New York, 2002.

[26] Attilio Meicci. (2005) Risk and Asset Allocation. Springer-Verlag Berlin Heidelberg 2005.

[27] Øksendal, B., Stochastic Differential Equations. An Introduction with Applications, 6th edition, Springer, Berlin, 2003.

[28] Pascucci, A. PDE and Martingale Methods in Option Pricing. Springer-Verlag Italia 2011.

[29] Rachev, S.T., J Hsu, B. Bagasheva and F. Fabozzi, Bayesian Methods in Finance, John Wiley & Sons, 2008.

[30] Ross, S.M. Stochastic Processes, John Wiley & Sons, Inc., New York, 1996.

[31] Samuelson, P.(1965) Theory of warrant pricing. Industrial Management Review, 13-32, 1965.

[32] Sparatu, A. Analysis and Probability. Elsevier, 2013.

[33] Schoutens, W. Levy Processes in Finance: Pricing Financial Derivatives, Wiley, New York, 2003.