Thesis: Applications on Branching Processes.

Name: Mitrofani Ioanna

Commitee:

Vasilis Koutras, Assistant Professor

Professor Agapios Platis

Professor George Dounias

Chios

2021

Abstract

Branching processes are stochastic processes with main difference among other stochastic processes that the systems that they can model have a special construction: a single individual lives for a unit of time and by its death produces n identical copies of itself. According to that principle and based on the relevant branching literature, branching processes are considered as a classical approximation for epidemics, biology, physics etc. The aim of this master thesis is to study in depth a type of stochastic processes, the branching processes, and focus on the existing applications, which would allow us to examine how these processes specifically are implemented and detect the strengths of branching processes comparing to other stochastic models. Additionally, primary scope constitutes the implementation of those processes in two different application areas.

After an extensive search of the relevant branching processes' literature, we obtained that modeling and evaluating system reliability by using these processes seems to be poorly documented. This is the reason why the first part of this research focuses on the application of branching process on system reliability. In particular, we examine a refinery pump system reliability through branching processes. A Markov model is used to formulate real data from a petrochemical industry and be able to use them as inputs to the branching model. Through this application, a refinery pump system availability is discussed in an alternative perspective than in typical reliability theory. The probability of ultimate extinction of that peculiar population consisting of pumps as well as the failure probability of the system during a year are estimated as typical properties of branching processes. Among other findings, that are concentrated to system availability, is that comparing to other stochastic models, a branching process approximation of the system reliability could be profitable for the maintenance departments, as an alternative perspective, because through the expected number of working components and the probability of extinction, the reliability of an entire industry unit is discussed.

Moreove, motivated by the sudden outbreak of the Covid-19 pandemic incidence and based on the fact that branches processes are extensively used to model dynamics of epidemics, another application these processes, refering to a typical branching approximation of the coronavirus (covid-19) spread in Greece is also presented. For this epidemiology model application, by using branching processes and their main properties, important factors are estimated for the virus transmission in Greece, such as the basic and effective reproduction numbers along with the probabilities of the extinction and an outbreak. Based on these factors and on an additional non-mitigation scenario, the effectiveness of control measures is discussed. Overviewing the results revealed that the virus transmission was aggressive, however the control measures were effective. This statement is supported by the values of the aforementioned indicators.

In general, the contribution of this research is based on three pillars. First of all, an analytical theoretical framework of branching processes is presented. Secondly, it provides knowledge about how to formulate and provide reliable results of a real problem in an alternative area of application for these processes, such as the mechanical system reliability. Finally, based on the fact that the presented branching models are consisting of different populations and thus the notion of parameters such as the expected number of population and the probability of extinction differ, we can assume that this research gives a comprehensive view of branching processes dynamics compared to other stochastic processes.

Finally, according to the findings of those implementations, ideas for future work are extensively presented

Contents

Acknowledgments

Chapter 1: Introduction.

In probability theory, there exist numerous models and processes concerned with the analysis of evolution of random phenomena. The basic and most crucial function of models is to formulate all the knowledge we have about a system, so to analyze it and provide useful information or solutions to contemporary problems in a specific field (Haccou, Jagers and Vatutin (2007)). On the contrary to verbal models, that usually involve the risk of not being undrestood by people of different cultures and usually may disregard important factors, mathematical models present a stable, globaly understood, basis for the majority of scientists (e.g. statitians, engineers, computer scientists, biologists, social scientists etc.) regardless their field of research (Haccou, Jagers and Vatutin (2007); Lange (2010)). Especially when the examined system consists of populations, the analysis becomes a complex task and the notion of stochasticity is introduced because of the randomness of populations' evolution. For instance, the evolution of a population consisting of individuals is based on the stochastic fate of individuals and whereas the result of a random event (such as the birth/reproduction or death of a particle) may be only one of the possible results, the actual event is determined by chance (Haccou, Jagers and Vatutin (2007); Siegmund (2020)). Thus, stochastic processes are commonly used to model and analyze systems whose operation is characterized by random factors. Branching processes, among other stochastic processes, form a typical field of applied probability and are considered as a classical approximation for the analysis of population dynamics and growth (Asmussen and Hering (1983); Haccou, Jagers and Vatutin (2007); (Lange (2010)).

Branching processes (BPs) are stochastic processes with main aspect that the systems that a BPs models have a special construction: a single individual lives for a unit of time and by its death produces n identical particles; all these individuals compose a process (Athreya and Ney (1972); Schinazi (1999)). It seems that based on this principle, branching processes can model systems of classical areas (e.g. biology, epidemiology, chemistry, physics, geneology-sociology etc.) as well as alternative phenomena (e.g. cascading failures on power transmission, ventilation or, generally, complex systems) (Kemeny and Snell (1978); Dobson, Carreras and Newman (2005); Pázsit and Pál (2008)). Thourgout this thesis, we focus on the theoretical framework of that field and possible corresponding applications. However, the main aim is to present the basic concepts and formulation of branching theory and to conduct applications on alternative areas, such as reliability of mechanical, industrial or financial related systems, considering the existing relevant literature. To this direction, two different applications of BPs are examined. The first application concerns the mathematical modelling of the reliability of a petrochemical industry plant unit consisting of pumps by using BPs. However, without overlooking the aforementioned ideas, due to the sudden outbreak of the Covid-19 pandemic and based on the fact that BPs are extensively used to model dynamics of epidemics, the second application refers to the coronavirus (covid-19) spread in Greece.

The first presented application attempts to combine Markov processes and branching processes to model a plant unit consisting of pumps. The aim is to evaluate the reliability of the plant as well as some of the system characteristics. Maintenance of equipment, nowadays, is of primary importance and great necessity for industry, comparing to the past where the prevailing idea was that there is no need for further attention (Deepak Prabhakar and Dharmaraj (2018); Raghavaiah and HariPrasad (2019)). In addition, the relevant reliability theory is a key factor to properly maintain a system. To that side, Markov models are classes of stochastic processes that are broadly proposed for reliability system analysis as well

as supporting decision tools for maintenance strategies, on the contrary to BPs, where modeling and evaluating system reliability by using them seems to be poorly documented in the relevant literature (Dawid, McMillan and Revie (2015); Gabe, Freire and de Andrade (2019); Ye, et al. (2020)). Real data from the industry are used to examine the proposed approach for real-life system. Such an analysis is considered as an alternative analysis of systems' reliability and could be profitable for maintenance departments.

The second application of branching processes refers to an epidemiology branching model, so to analyze the initial and latter stages of covid-19 transmission. Mathematical modeling constitutes an important tool to estimate key parameters of a disease transmission and predict the dynamic of a virus (Allen (2015)). More precisely, in the relevant literature, epidemiology is considered as a classical application area of branching processes, which are stochastic individual-based processes. The novel coronavirus (covid-19) was initially identified at the end of 2019 and caused a global health care crisis (Ndaïrou, et al. (2020)). The increased transmissibility of the virus, that led to high mortality, raises the interest of scientists worldwide. Thus, various methods and models have been extensively discussed, so to study and control covid-19 transmission (Ndaïrou, et al. (2020); Ivorra, et al. (2020)). As a second application, a classical Galton-Watson branching process approach is developed for the covid-19 spread in Greece. The primary aim of this application is to model the transmission stages through BPs in order to analyze the first semiannual spread of the pandemic. In order to give a comprehensive view of the pandemic, the effectiveness of control measures is discussed, based on a simple exponential smoothing model, which is used to build a non-mitigation scenario.

To sum up, this thesis is structured in 5 chapters. Chapter 2 analytically presents the typical branching theory and the mathematical formulation of branching models, which constitute important pillars of the following applications that are presented in Chapter 3 and 4 respectively. In these chapters (3 & 4, respectively for each application), each problems' formulation is separately discussed as well as the outputs and conclusions. Finally, the last chapter (Chapter 5) gives a broadly overview of the branching processes viewed through the presented applications and additionally ideas for future work are considered and firmly discussed.

Chapter 2 : Literature Review on Branching Processes.

2.1. Theoretical framework of Bienaymé-Galton-Watson processes.

Branching Processes are individual-based models that consider the growth of populations, whose particles' reproduction follows stochastic laws, and have a long history (Dorman, Sinsheimer and Lange (2004)). The term "branching processes" has been coined in 1847, when the population mathematics adopted probability theory, by A. N. Kolmogorov and N. A. Dmitriev, but the subject is much older, approximately one century ago (Kendall (1966); Gonzalez and del Puerto (2010)). The initial motivation event it is considered to be the extinction of surnames of European aristocratic families, that Francis Galton documented in 1867 and Reverend Henry William Watson formulated in 1873 in Educational Times (Gonzalez and del Puerto (2010); Almudevar, Oakes and Hall (2020)). The initial communication of the problem was published in 1873 to a periodical, that was characterized by Galton as "*a well-known mathematical periodical of a high-class Educational times*" (Kendall (1966)):

"PROBLEM 4001: A large nation, of whom we will only concern ourselves with adult males, N in number, and who each bear separate surnames colonise a district. Their law of population is such that, in each generation, a0 percent of the adult males have no male children who reach adult life; a1 have one such male child; a2 have two; and so on up to a5 who have five.

Find (1) what proportion of their surnames will have become extinct after r generations; and (2) how many instances there will be of the surname being held by m persons."

The formulation that Watson proposed, is based on probability generating functions and stands until nowadays. Removing the restriction of $k \leq 5$, writing p_k instead of Galton's a_k and if the input is

$$f(s) = \sum_{k=0}^{\infty} p_k \cdot s^k \quad (0 \leq s \leq 1) \tag{1}$$

$$and \quad f_1(s) = f(s), \quad f_{n+1}(s) = f(f_n(s)) = f_n(f(s)) \quad (n = 1, 2, \dots) \tag{2}$$

then the power series for f_n have as coefficients the terms of the probability distribution for the total number of males in n^{th} generation, whereas the first male constitutes the zero generation (Kendall (1966)). In addition, Watson detected that the probability of extinction, q_n, satisfies the equations:

$$q_1 = p_0 \quad and \quad q_{n+1} = f(q_n)$$

$$\text{and that if } q_n \to q \text{ when } n \to \infty, \text{ then } q = f(q) \tag{3}$$

and always has as a root $q = 1$ (Kendall (1966)). Based on that Watson believed that the extinction of the male line is an inevitable circumstance. However, a few years later, Agner Krarup Erlang considering the problem of Watson, revised the formulation of Watson's equation (3) and interpreted the probability of extinction q equals to : (i) the chance of no male in first generation, plus to (ii) the chance of one male in first generation, followed by extinction, plus to (iii) the chance of two males in first generation, followed by both ultimate extinction and so on (Kendall (1966)):

$$q = p_0 + p_1 q + p_2 q^2 + \cdots \tag{4}$$

To that side, Erlang noticed that equation (3) can have two roots in the interval [0,1]; one root is $q = 1$, and there is exactly one more root in [0,1] if and only if the expected number of sons per parent is greater than unity:

$$m = \sum_0^\infty k p_k \tag{5}$$

In fact, that observation led to a basic theorem, which until nowadays is considered as an important property of branching processes and stands as categorization parameter of the criticality of a process (Kendall (1966)): "*it is always the smallest root of equation (3) which is the appropriate one; thus extinction is always certain for subcritical populations with $m < 1$ and for critical population with $m = 1$, but there is always a positive chance of survival for supercritical populations with $1 < m \leq \infty$*". However, I. J. Bienaymé had noticed and formulated the family extinction problem before Watson did, thus branching processes are also known as Galton-Watson or Bienaymé-Galton-Watson processes (Kendall (1966); Gonzalez and del Puerto (2010)).

According to the aforementioned information about these processes genealogy and in general social studies area, have been extensively used branching processes, thus it is believed that this area is considered as a classical application area of branching processes (Gonzalez and del Puerto (2010); Lange (2010)). In addition, these processes had been broadly used to biological field applications such as ecology, epidemiology, genetics, evolution, cell biology (cell division), population biology (survival of mutant genes), as well as to physics (neutron chain reaction) (Jagers (1975); Macken and Perelson (1985); Schinazi (1999); Kimmel and Axelrod (2002); Lange (2010)). Apart from the epidemics that will be furtherly discussed in next chapters, all the aforementioned fields are not part of that thesis. However, there exist a considerable number of manuscripts and books that focus on branching processes and their applications to these areas, such as the book of Kimmel and Axelrod (2002), Haccou, Jagers and Vatutin (2007), Gonzalez and del Puerto (2010) and books that set an analytical theoretical framework such as those of Harris (1964), Athreya and Ney (1972).

Nevertheless, in last decades many researchers work on modelling peculiar populations by using branching processes and create applications in non-classical areas, such as the failures of several technological systems (Dobson, Carreras and Newman (2004,2005); Kim and Dobson (2011)). Contemporary studies attempt to approach reliability of mechanical multi-parted systems through branching process models. According to the relevant literature review in classical and contemporary conducted studies, this thesis deals with a typical epidemiology model and a model of system reliability. The next two sub-sections (2.1.1.-2.1.2.) give a broad review in these fields. In addition, the relevant reference in Markov processes is integrated, because a Markov model is used in the branching application so to form real data to use them.

2.2. Elementary branching theory and formulation.

A branching process is a stochastic process which consists of collections of random variables and the main purpose is to serve as a mathematical model of a population in which each individual in generation n produces some random number of individuals in generation $n + 1$. These individuals in next generation produce other individuals. All individuals reproduce independently but under the same offspring distribution. Branching processes are used to model reproduction, for example, the individuals might correspond to bacteria, each of which generates 0, 1, or 2 offspring with some probability in a single time unit. Branching processes can also be used to model other systems with similar dynamics such as the spread of surnames in genealogy or the propagation of neutrons in a nuclear reactor (Haccou, Jagers and Vatutin (2007)).

A branching process has a "boom or bust" distribution: either the population will take off (grows up quickly) or will fail altogether. Additionally, an important information is that by the time the population size starts decreasing, the extinction of the BP will occurr soon by the first stages. One of the crucial questions in the theory of branching processes is the probability of ultimate extinction, where no individuals exist after some finite number of generations, or the probability of survive, as well as the conditions under which the population became extinct (Athreya and Ney (1972)).

The classification of BP could be made in various bases, the majority of authors uses the following: according to their time parameter and their criticality condition (Kimmel and Axelrod (2002)). Based on their criticality there are three cases (subcritical, critical and supercritical), but first we need to get introduced to branching processes/GWP so to carry more information about that classification, thus, we will analyze criticality conditions below (Kimmel and Axelrod (2002); Haccou, Jagers and Vatutin (2007)).

Based on lifetime (time parameter) we have gradient classification. If lifetimes of all individuals are identical and equal to 1, it follows that the process can be described using a discrete-time index, identical to the number of successive generations (Athreya and Ney (1972); (Kemeny and Snell (1978); (Kimmel and Axelrod (2002); Pázsit and Pál (2008)). This is the Galton-Watson classical approach (the simplest form of Discrete Time Branching Processes (DTBP)). On the other hand, if lifetimes of each individual are not constant (one unit time as in classical approach), there are two general categories (Haccou, Jagers and Vatutin (2007)), (Athreya and Ney (1972)):

- Markov Process: if the particles lifetime distribution is exponential (we have a continuous-time Markov process)

- Age-Dependent (or Bellman-Harris) Process: if lifetime of individuals is an arbitrary non-negative random variable.

As in the majority of stochastic processes, there exist discrete-time and continuous-time branching processes. In discrete-time branching models, time is represented by integers that indicate reproduction periods (Haccou, Jagers and Vatutin (2007)). Thus, in the simplest forms, there are no overlaps between generations and reproduction occurs only in separate periods (Haccou, Jagers and Vatutin (2007)). Continuous-time models are more complex than Discrete-time models, however, as a discrete-time branching process has small time-steps, it can be viewed as an approach of a continuous

branching model (Haccou, Jagers and Vatutin (2007)). In fact, discrete-time branching models are usually used so to extract useful information about the corresponding continuous-time model.

The oldest and simplest discrete time branching processes are the simple Galton-Watson single-type branching processes, that are also named Bienaymé-Galton-Watson processes or simple branching processes. As it was already mentioned, we have only one type of individuals and there are non-overlaps between generations, as it is confirmed by Figure 1 that is a simple illustration of the process.
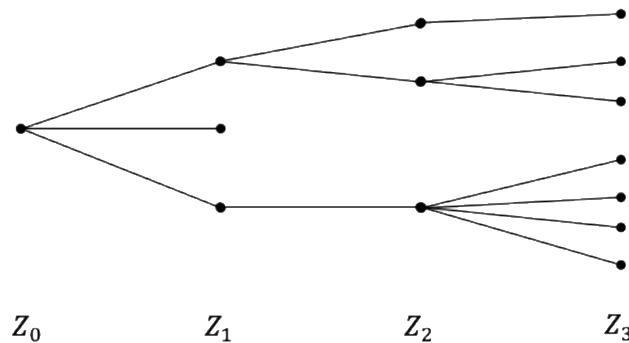

Figure 1. Discrete Time Branching Process tree

The basic branching model or Galton-Watson simple branching process is a discrete time process with non-overlaps between generations and consisting of only one type (or single type) of individuals (Haccou, Jagers και Vatutin (2007)). The size of the population in n^{th} generation is denoted by Z_n. Each individual lives exactly one unit of time, produces a random number of offspring (ξ) and dies. All individuals reproduce independently. The population starts with one individual ($Z_0 = 1$) at $n = 0$ generation and the population size of next generation can be determined by summing the numbers of offspring (Schinazi (1999)). The number of offspring of the ith particle is denoted as ξ_i, and ξ_i with $i = 1, 2, 3, \ldots$ are independent and identically distributed random variables (i.i.d.) and the family size distribution is $P(\xi = k) = p_k$, the probability of producing k offspring (Haccou, Jagers και Vatutin (2007)). Additionally, the population size of (n+1)^th generation (Z_{n+1}) arises as the sum of the offspring delivered by the particles of nth generation :

$$Z_{n+1} = \xi_1 + \xi_2 + \cdots + \xi_{Z_n}$$

So, the number of individuals at time n, Z_n, is equal to the total number of offspring of the individuals $1,2,3,\ldots,Z_{n-1}$. Then the branching process is $\{Z_0, Z_1, Z_2, \ldots,\} = \{Z_n : n \in \mathbb{N}\}$

$$Z_n = \sum_{i=1}^{Z_{n-1}} \xi_i \qquad (6)$$

The state of branching process at time n is Z_n, it takes values 0,1,2, … but $Z_0 = 1$ always. Once Z_n takes the value 0 this means that no individual is alive, then the process is dead and this state is considered as an absorbing. However, the aforementioned formulation and assumptions compose the simplest type of branching models. There exist more complex branching models, such as the multi-type or several type branching processes, where the population consists of more than 1 species and one of the

major differences is that same-type particles are assumed to have identically distributed number of offspring, however different type particles may differ (Haccou, Jagers and Vatutin (2007)). Other more complex cases of branching models are those in which each particle of the population does not necessarily live only one unit of time, not all individuals have the same life spans or particles can give birth at different ages (Haccou, Jagers and Vatutin (2007)). In addition, some cases that start in zero generation with more than one particles are considered as different branching processes. The assumption about the independence of reproduction allows each individual to compose its new branching process, for instance in Figure 1, the first particle of the first generation constitutes a branching process that starts with one individual (itself) that produces two identical copies of itself, these particles compose the next generation and produce other individuals.

Two major issues are important to discuss considering populations and reproductions: the expected number of population size and the probability of extinction (Athreya and Ney (1972); Schinazi (1999); Haccou, Jagers and Vatutin (2007)). In probability and statistics, the population mean, or expected value, is a measure of the central tendency either of a probability distribution or of the random variable characterized by that distribution. So, it seems pretty reasonable to calculate the mean of Z_n ($E(Z_n)$), because we need to find out what we expect the population to be in n generations time (Schinazi (1999)). On the other hand, variance is the expectation of the squared deviation of a random variable from its mean. Informally, it measures how far a set of (random) numbers is spread out from the average value. As we mentioned above for Z_n, the population either will take off or will fail altogether. In fact, if the population size fails, it is likely to do so very quickly, within the first few generations. This explains why we are interested in $Var(Z_n)$. High variance denotes that the population will not evolve rapidly around the mean values, and respectively, low variance alerts us that the population will probably take off around the mean value.

Both $E(Z_n)$ and $Var(Z_n)$ can be expressed in terms of the mean and variance of the number of offspring. Thus, let $E(\xi) = m$ and $Var(\xi) = \sigma^2$. These are the mean and the variance of a single individual (Schinazi (1999)). In order to estimate the expected number of population size, let $\{Z_0, Z_1, Z_2, \dots, \}$ be a branching process with $Z_0 = 1$, let ξ denote the number of offspring and suppose that $E(\xi) = m$. Then,

$$E(Z_n) = E(Z_n|Z_0 = 1) = m^n \tag{7}$$

Proof (Schinazi (1999); Haccou, Jagers and Vatutin (2007)):
We will do the proof by induction. Firstly note that

$$E(Z_1|Z_0 = 1) = E(\xi) = m = m^1$$

The formula works for $n = 1$. Assume that the formula works for n. By conditioning Z_n we get

$$E(Z_{n+1}|Z_0 = 1) = \sum_{k \geq 1} E(Z_{n+1}|Z_n = k) \, P(Z_n = k|Z_0 = 1) \tag{8}$$

By Markov property we are using that

$$E(Z_{n+1}|Z_0 = 1, Z_n = k) = E(Z_{n+1}|Z_n = k) \text{ for every } k \geq 1. \text{ Now we have:}$$

$$E(Z_{n+1}|Z_n = k) = E\left(\sum_{i=1}^{k} \xi_i\right) = km \text{ , because we assumed that } E(\xi) = m.$$

So, equation (8) becomes:

$$E(Z_{n+1}|Z_0 = 1) = \sum_{k \geq 1} km \cdot P(Z_n = k|Z_0 = 1) = mE(Z_n|Z_0 = 1)$$

Because we know that $\sum_{k \geq 1} k \, \mathbb{P}(Z_n = k|Z_0 = 1) = E(Z_n|Z_0 = 1)$.

By the induction hypothesis we have $E(Z_n|Z_0 = 1) = m^n$, so we can conclude that

$$E(Z_{n+1}|Z_0 = 1) = mm^n = m^{n+1}$$

$$E(Z_{n+2}|Z_0 = 1) = mm^{n+1} = m^{n+2} \text{ and so on.}$$

Thus, by the induction analysis it stands that $E(Z_n|Z_0 = 1) = m^n$, which the expected number of population size $E(Z_n)$ and is related to the mean value of offspring.

$$\square$$

In order to estimate the variance, let $\{Z_0, Z_1, Z_2, \dots,\}$ be a branching process with $Z_0 = 1$ and let Y denote the family size distribution and suppose that $E(Y) = m$ and $\text{Var}(Y) = \sigma^2$. Then,

$$Var(Z_n) = \begin{cases} \sigma^2 n & if \ m = 1 \\ \sigma^2 \mu^{n-1}\left(\dfrac{m^n - 1}{m - 1}\right) & if \ m \neq 1 \end{cases} \tag{9}$$

Proof (Schinazi (1999); Haccou, Jagers and Vatutin (2007)):

Starting from $Var(Z_n) = Var\big(E(Z_n|Z_{n-1})\big) + E\big(Var(Z_n|Z_{n-1})\big)$ (Law of total variance). In the previews proof, it has been defined that $E(Z_{n+1}|Z_n = k) = E\left(\sum_{i=1}^{k} Y_i\right) = km$, so $E(Z_n|Z_{n-1} = k) = E\left(\sum_{i=1}^{k} Y_i\right) = km$. Similarly, and because of the independence of Y_i we can get

$$Var(Z_n|Z_{n-1} = k) = E\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} Var(Y_i) = \sigma^2 k$$

Applying the law of total variance we get:

$$Var(Z_n) = Var(mZ_{n-1}) + E(\sigma^2 Z_{n-1}) = m^2 Var(Z_{n-1}) + \sigma^2 m^{n-1} \ for \ n \geq 1$$

By induction and if we assume that $Var(Z_0) = 0$, we get

For Z_1: $Var(Z_1) = m^2 Var(Z_0) + \sigma^2 m^0 = \sigma^2$

For Z_2: $Var(Z_2) = m^2 Var(Z_1) + \sigma^2 m^1 = m^2 \sigma^2 + \sigma^2 m = \sigma^2 \mu(1 + m)$

For Z_3: $Var(Z_3) = m^2 Var(Z_2) + \sigma^2 m^2 = m^3 \sigma^2 (1 + m) + \sigma^2 m^2 = \sigma^2 m^2 (1 + m + m^2)$

For Z_4: $Var(Z_4) = m^2 Var(Z_3) + \sigma^2 m^3 = \sigma^2 m^4 (1 + m + m^2) + \sigma^2 m^3 = \sigma^2 m^3 (1 + m + m^2 + m^3)$

$$\vdots$$

So the general pattern is proved by induction on and it is:

$$Var(Z_n) = \sigma^2 m^{n-1} \sum_{k=0}^{n-1} m^k = \begin{cases} \sigma^2 m & if\ m = 1 \\ \sigma^2 m^{n-1}\left(\dfrac{m^n - 1}{m - 1}\right) & if\ m \neq 1 \end{cases}$$

$\square$

The following examples give briefly a simple implementation of the aforementioned formulas:

1. Family size $Y \sim Geometric(p = 0.3)$

   So, $\quad m = E(\xi) = \dfrac{q}{p} = \dfrac{0.7}{0.3} = 2.33$

   $$\sigma^2 = \dfrac{q}{p^2} = \dfrac{0.7}{(0.3)^2} = 7.78$$

   For the expected population in 10th generation (n=10):
   $$E(Z_{10}) = m^{10} = 2,33^{10} = 4,715.87$$

   Since $m = 2.33 \neq 1$
   $$Var(Z_{10}) = (7.78)^2 (2.33)^2 \left(\dfrac{2.33^{10} - 1}{2.33 - 1}\right) = 43.4 \times 10^7$$

2. Family size $Y \sim Geometric(p = 0.5)$

   So, $\quad \mu = E(\xi) = \dfrac{q}{p} = \dfrac{0.5}{0.5} = 1$

   $$\sigma^2 = \dfrac{q}{p^2} = \dfrac{0.5}{(0.5)^2} = 2$$

   For the expected population in 10th generation (n=10):
   $$E(Z_{10}) = 1^{10} = 1$$

Since $\mu = 1$

$$Var(Z_{10}) = \sigma^2 n = 2 \times 10 = 20$$

Within these examples we can define how mean and variance can be used as prediction tools to see what is going on with the extinction of the population before calculating the probability of extinction. A low variance alerts us to the fact that the process clusters closely around the mean values. Is the variance of a population enough to decide about the extinction of a population? The answer is no, because we need to know the mean value around of which the process clusters. For instance, the population of example 2 is more likely to become extinct than the population of example 1. In example 1, the mean and the variance are high, so the values are spread widely around the "high" mean. In example 2, the mean and the variance are low, which means that the values are spread closely to one. Thus, population of example 2 is more likely to be extinct than the population of example 1. In addition, that assumption explains, firstly, the fact that if the population size fails, it is likely to do so very quickly, and secondly why we are interested in $Var(Z_n)$. An important information, that also proves the assumption we made above about the mean-variance-extinction, is that the mean (m) is an inference about the probability of extinction (Schinazi (1999)):

- If $m \leq 1$ then the expected number of offspring tends fast to 0 and the extinction probability is 1 ($m < 1$ is considered as subcritical case and respectively $m = 1$ is a critical case)
- If $m > 1$ the extinction probability is less than 1 but not necessarily 0 and it is calculated by the PGF of Z_n (supercritical case)

In addition, by knowing or formulating the related probability generating function, the estimation of the expected number of population size could be much simpler. Considering the Watson's initial formulation of equations (1)-(3), the definition of the process (equation (6)), and basic properties of probability generating functions, the expected number of population size is the first derivative of the probability generating function to 1 (Athreya and Ney (1972); Lange (2010)): If the progeny function (PGF) is $G(z)$, then the expected number of population size is $G'(1)$. As it is already mentioned, one of the main properties of branching models is that the probability of ultimate extinction is also estimated by using the progeny functions.

By extinction we mean that the random sequence $\{Z_n\}$ consists of zeros for all, but finite, number of values. Since Z_n is integer-valued, extinction is an event that forces Z_n go to 0 ($Z_n \to 0$). The population is extinct by generation n if $Z_n = 0$, and after that time the population is extinct forever: $Z_t = 0 \; for \; all \; t \geq n$ (Haccou, Jagers and Vatutin (2007)). In addition, if the population once becomes extinct, it is always extinct. Thus, if we assume that extinction occurs at time $n = 2$, we have that $\{Z_2 = 0\}$ and that enforces $\{Z_3 = 0\}$ and so on. The probability of ultimate extinction refers to the probability that the population will be extinct by generation n, for any value of n and if q_n represents the probability of extinction at generation n, then(Schinazi (1999)), (Haccou, Jagers and Vatutin (2007)):

$$P(ultimate\ extinction) = q = \lim_{n \to \infty} q_n$$

Considering the initial formulation of Watson-Erlang (equations (1)-(4)), the progeny function is formulated as follows:

$$G(q) = p_0 + p_1 q + p_2 q^2 + \cdots \tag{10}$$

where p_i is the probability to have $i = 1, 2, 3, \ldots$ individuals and q is the probability of ultimate extinction. Let q_n be the probability of extinction in the n$^{\text{th}}$ time period (generation) and $q_0 = 0$; as $n \to \infty$ then $q_n \to q$ ($0 \leq q \leq 1$). In order for the population to become extinct in period n, all individuals should die at least on the previews period $(n-1)$. Therefore,

$$q_n = p_0 + p_1 q_{n-1} + p_2 q_{n-1}^2 + \cdots \tag{11}$$

Recalling the progeny function (equation (10)), equation (11) can be written as $q_n = G(q_{n-1})$ and according to Watson, as $q_n \to q$, the probability of extinction should satisfy $G(q) = q$. Whereas $\sum p_i = 1$, the solution $q = 1$ is always a root of equation $G(q) = q$. In addition, the first and second derivatives of the PGF are formed respectively as follows, with respect to q:

$$G'(q) = p_1 + 2p_2 q + 3p_3 q^2 + \cdots \tag{12}$$

$$G''(q) = 2p_2 + 3 \cdot 2p_3 q + 4 \cdot 3p_4 q^2 + \cdots \tag{13}$$

It is obvious that both derivatives are non-negative for $q \geq 0$ and it turns out that $G(q)$ is a non-decreasing function that concaves upwards. Thus, $y = z$ and $y = G(z)$ should intersect at most on two points and one of them is $(1,1)$ as $q = 1$ is always a root of equation. To that side there exist three scenarios, illustrated in Figure. (2) (Athreya and Ney (1972)) (Fazlollahtabar and Niaki (2018)):

- 2 roots $(q, 1)$, where $0 \leq q < 1$ (Figure. (2a))
- 1 root $q = 1$ (Figure. (2b))
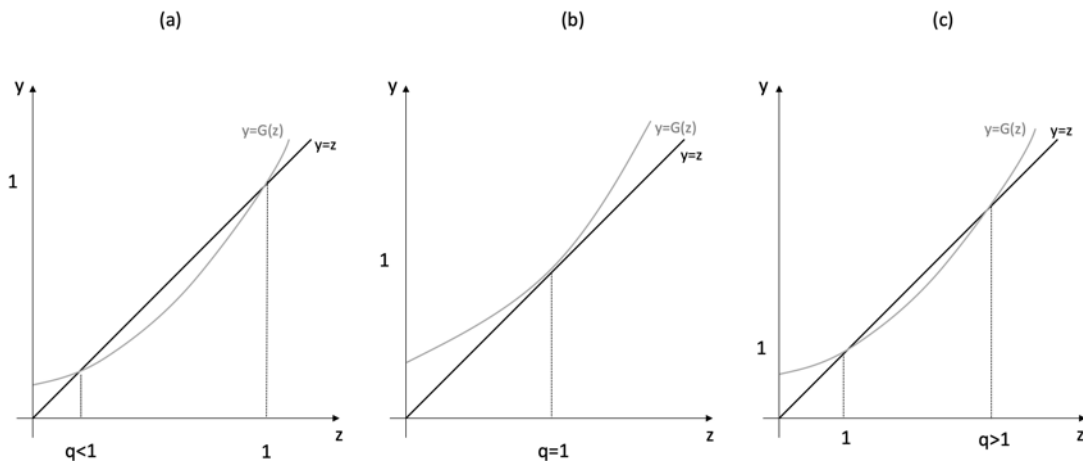- 2 roots $(1, q)$, where $q > 1$ (Figure. (2c))

Figure 2. Graphs of scenarios of intersection points of $y = z$ and $y = G(z)$.

While q refers to the probability of extinction should be $0 \leq q \leq 1$, thus in cases (b), (c) the only solution is $q = 1$. However, recalling that in general $G'(1) = m$, the first derivative of the progeny function at 1

gives the expected number of population size, the case (a), presented in Figure. (3), corresponds to a supercritical case ($G'(1) = m > 1$).
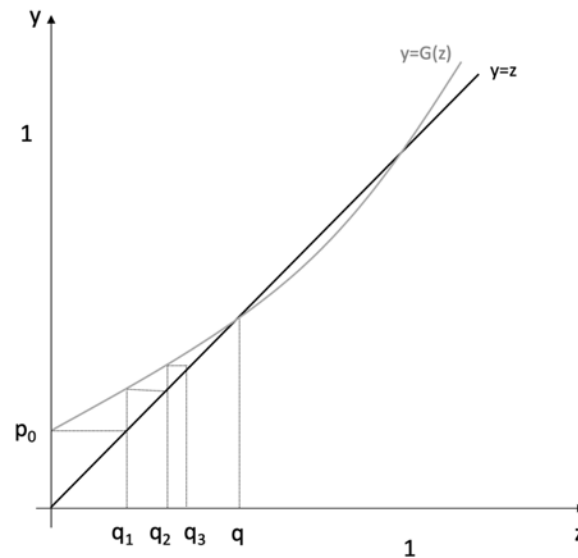

Figure 3. Intersection points of $y = z$ and $y = G(z)$ in a Supercritical case scenario.

To that side, and because $q_0 = 0, q_1 = G(q_0), \dots, q_n = G(q_{n-1})$, the points $(q_i, G(q_i))$ will always lie above the line $y = z$ (Fazlollahtabar and Niaki (2018)). Hence, there exist one more solution that should be less than 1 ($q < 1$) that is estimated as the smallest root of equation $G(z) = z$.

The probability of extinction is related to the mean (m). Some values of m guarantee the death of the branching process with probability 1. If the mean number of offspring per individual m is 1 or less, then the ultimate extinction is certain. Other m guarantee that the probability of extinction will be less than 1. If the mean number of offspring per individual m is higher than 1, then the ultimate extinction is not certain but it might occur (Lange (2010)). According to the aforementioned analysis, let $\{Z_0, Z_1, Z_2, \dots, \}$, with ($Z_0 = 1$), be a branching process and m be the expected number of population size. Let also q be the probability of ultimate extinction. Then (Athreya and Ney (1972); Haccou, Jagers and Vatutin (2007)):

- If $m > 1$, then $q < 1$: extinction is not guaranteed (supercritical case)
- If $m < 1$, then $q = 1$: extinction is guaranteed (subcritical case)
- If $m = 1$, then $q = 1$: extinction is not guaranteed unless there is only one particle (critical case)

To sum up till this point, in a branching process model all the effort is concentrated to build and formulate the relevant probability generating function. Once the progeny function is formulated, we can define the following data about a process:

1. The probability of ultimate extinction as the smallest root of $G(z) = z$.
2. The probability of the population to become extinct in a specific generation (q_n).
3. The conditions of extinction (the expected number of population as $G'(1)$.

2.2.1. Continuous Time Branching Processes (CTBP)

As it was referred before, in a (Galton-Watson) branching process, the future states are independent of the past. This means that these processes satisfy the property of Markov processes, where the future depends only upon the present and not at all upon the past. In addition, in classical Galton-Watson approach, each individual lives for exactly one unit of time. We assume that individuals' lifetime is exponentially distributed and thus we have a continuous-time Markov process. Markov branching processes are the counterparts of GWBP in continuous time (Haccou, Jagers and Vatutin (2007)).

The main difference between continuous and discrete time branching processes is that births and deaths occur at random times for the CTBP. In addition, continuous time branching processes have the Markov property if and only if birth and death times are exponentially distributed (Haccou, Jagers and Vatutin (2007)).

Furthermore, DTBP are much simpler than CTBP. In DTBP, in classical approach, the population, at a time, contains individuals only from one generation. In continuous time this is not true; individuals of different generation may be alive at the same time. This occurs because of the difference in living time spans. Figure 4 shows a family tree of a continuous-time branching process. The individual dies after a single reproduction event, which is the simplest form of CTBP. The numbers at the branches (life-times) denote the generation of each particle. In this general example, at time t_1 the population consists of one particle from the second generation, three particles of third generation.
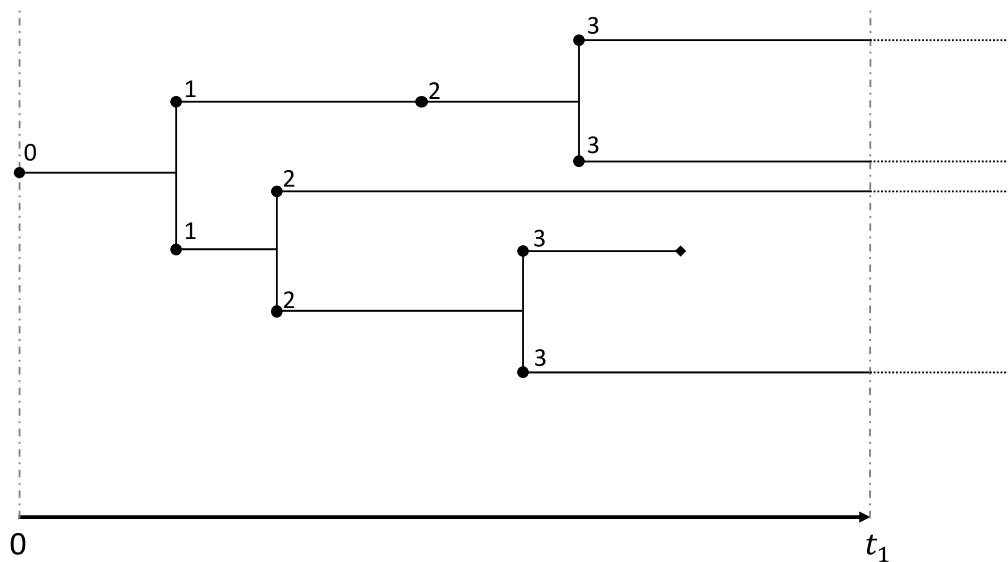
Figure 4. Continuous Time Branching Process tree.

In CTBP an embedded generation process is commonly used so to extract some fundamental results from Discrete Time Branching Processes and attach them to the corresponding Continuous Time Branching Process. The embedded process counts the number of particles of different ages (Haccou, Jagers and Vatutin (2007)). So, this process consists of random integer-valued variables and it is commonly denoted as ζ_n. For instance, in Figure 2 we had $\zeta_0 = 1, \zeta_1 = 2, \zeta_2 = 3$. Many authors do not use a different

symbol for the embedded process, they just call a discrete time branching process Z_n (Haccou, Jagers and Vatutin (2007)).

The embedded generation process is a Discrete Time Branching process. Assuming that $\xi_1, \xi_2, \ldots, \xi_n$ are the total offspring numbers of the various individuals in n^{th} generation during all their lives, we recall the corresponding equation

$$\zeta_{n+1} = \sum_{i=1}^{\zeta_n} \xi_i$$

In addition, the assumption that all individuals reproduce independently of each other and have the same distribution (of reproduction), leads to the outcome that the embedded process is a Galton-Watson Branching Process (Athreya and Ney (1972)).

An embedded generation process does not carry over other useful information about the original process (CTBP), but it is widely used, as mentioned before, to simplify the original process and through this we minimize the computational effort so to carry out fundamental results. For example, the extinction probability at a CTBP is determined by the much simpler embedded process. Besides, the extinction of the original process occurs if and only if the embedded process dies out (Haccou, Jagers and Vatutin (2007)).

In literature there are multi-type and single-type processes (Athreya and Ney (1972)), (Haccou, Jagers and Vatutin (2007)). The single-type processes are characterized by the life-time distribution that gives birth-event ages and numbers of offspring. The simplest cases in CTBP are single-type processes which have only one reproduction event, that happens by the death of the individual and the reproduction is independent of the lifetime. In this form we have the Markov branching and birth-and-death process (and the continuous time binary branching processes). If we let the reproduction distribution to depend on life-spans we have the age-dependent processes or Bellman-Harris Processes. In addition, we have the Sevastyanov Processes where not only the hazard rate but also the distribution of offspring numbers depends on the age of the ancestor. Finally, there are processes in which individuals may reproduce several times during their lifetime and additionally the multi-type processes (Athreya and Ney (1972); Schinazi (1999); Kimmel and Axelrod (2002); Haccou, Jagers and Vatutin (2007)).

Process with one reproduction in a life-span

These processes are usually considered as splitting processes, because the individual after a life-time is replaced by its offspring. In this section we will analyze the processes in which the reproduction is independent of the ancestor's life span and the individuals reproduce independently of each other and processes that reproduction is age-dependent (Haccou, Jagers and Vatutin (2007)).

(a)Markov Branching processes and Birth-and-Death processes

Markov branching processes are counterparts of CTBP. Branching processes in continuous time are splitting Markov processes with some characteristics. Firstly, there is one type of individuals, the reproduction happens individually between particles and the particles reproduce with the same distribution. The concept is that an individual lives a life span which is exponentially distributed and then it splits into ξ particles. The reproduction procedure of each particle is independent of each other and additionally, as mentioned before, the reproduction distribution is independent of the age of the parent.

Markov branching processes are characterized by the parameter λ of the exponentially distributed life-time and the distribution of ξ offspring values. Here, the parameter λ counts the chance per time-unit of splitting and is constant because of the exponential life-time distribution. The parameter λ is named, differently by many authors, as death rate, hazard or intensity ((Athreya and Ney (1972); (Schinazi (1999); (Haccou, Jagers and Vatutin (2007)).

The stochastic process $(Z_t)_{t \geq 0}$, that gives the population size at any time $t \geq 0$, is determined by the structure described above and the starting condition Z_0. The exponential distribution has a probability density function form $\lambda e^{-\lambda t}, t > 0$ for some $\lambda > 0$. In mathematical symbols, a real-valued, nonnegative random variable T follows an exponential distribution if and only if (Haccou, Jagers and Vatutin (2007)):

$$P(T > t + u | T > u) \; = \; P\;(T > t) \qquad for\ all\ t, u \; \geq 0$$

In other words, the probability distribution of the remaining lifetime is independent of the future/next, given that a determined age has been attained. Individuals with exponentially distributed life time do not age. The non-aging property, thus, is extracted from the Markov property. As in Galton-Watson processes we are interested in the expected value of the population size. In addition, if we overlook the risk that more than one death occurs in each generation, we assume that Z_u is ξ or 1. That depends on whether the ancestor of the process has died by u or not (and then obtained ξ children). As it was on GWBP, $m = E[\xi]$, so, we conclude to (Athreya and Ney (1972); Schinazi (1999); Haccou, Jagers and Vatutin (2007)):

- If $m > 1$ the population size has an exponential increase

- If $m < 1$ the population size has an exponential decrease

Markov branching processes are usually related to **birth-and-death processes** which are continuous time models. These processes are commonly defined as integer-valued Markov processes Z_t with the property that the intensity of jumping one step upward from the population size $Z_t = j$ is jb, and the intensity of decrease by one step is jd. The positive numbers d, b are usually called respectively birth and death rates. The concept is that if the population size is $Z_t = j$, we have j particles at present, each one with birth rate b. So, the whole population jumps up with rate jb and down with rate jd. On the individual level the birth-death process is a binary splitting Markov branching process. Life-times are distributed exponentially with the parameter $(b + d)$ (Schinazi (1999); Haccou, Jagers and Vatutin (2007)). Thus, the mean life-time is $1/(b + d)$. The offspring distribution is

$$P(\xi = 0) = \frac{d}{b+d} \quad \text{and} \quad P(\xi = 2) = \frac{b}{b+d}$$

Hence, this model rises the assumption that an individual has a chance per time unit of dying without offspring, equal to $\frac{d}{b+d}$ and a chance per unit time of splitting into two of $\frac{b}{b+d}$. Additionally, the birth-and-death processes are sometimes referred to as linear processes (opposed to Markov) with a growth rate of μ_j, if $X_t = j$. That corresponds to branching processes dependent on population size in continuous time (Athreya and Ney (1972); Schinazi (1999); Haccou, Jagers and Vatutin (2007)).

The Continuous Time Binary Branching Process

We define a continuous time binary process by the following properties:

- The number of individuals at time t is denoted by Z_t and we start with just one single type of reproductive individual, $Z_0 = 1$.

- Independence between individuals

- Each individual gives birth to a new individual with rate λ or dies with rate 1.

Each individual in the population has two independent exponential random variables attached to it. One random variable has rate λ (birth) and the other has rate 1(death). The rule is that if the random variable with λ exponential happens before random variable with rate 1, then the individual is replaced by two individuals (splits into two). Otherwise, the particle dies out with no offspring. As it happens in birth-death processes, we notice that the problem of survival for a CTBP is identical to the survival problem for the corresponding DTBP. Now, we know that a DTBP survives if $E(\xi_1) > 1$. We can compute the expected value of ξ_1, even if we do not know the distribution of it (Haccou, Jagers and Vatutin (2007)).

As next step, we want to define a differential equation for $M(t)$ (the expected size of the population at time t i.e. $E(Z_t)$) so to see what happens to the population in future. At time 0 we have a single individual. We need to see what happens within a small time span h. We have three possible scenarios:

1. The single particle gives birth between times 0 and h, with probability $\lambda h + o(h)$ (because with rate λ an individual gives birth/splits into two individuals)

2. The individual dies out with no offspring, with probability $1 \cdot h + o(h)$ (because with rate 1 an individual dies)

3. Nothing happens at all, with probability $1 - \lambda h - h + o(h)$, in better form $1 - (\lambda + 1)h + o(h)$

In general, we can assume that, if $m = E(\xi) > 1$, the process survives, otherwise (if $m = E(\xi) < 1$) the population dies out. A remarkable note here is that, considering a continuous time branching process with $Z_0 = 1$ and assuming that $Z_1 = 3$, each of these three particles was possible to be appeared in different times between 0 and 1. Nevertheless, by assuming that $(Z_n)_{n \geq 0}$ is a discrete time branching process, we claim that each one of these particles, that have the same progeny distribution, starts its process at time 1. That arises from the Markov property (the memoryless property of the exponential distribution). As we mentioned before, individuals with exponentially distributed life time do not age. It only matters that an individual is present at time 1 and not at all how old is it at time 1 (Athreya and Ney (1972); Haccou, Jagers and Vatutin (2007)).

(b) Bellman-Harris Processes (Age-dependent)

Bellman-Harris processes constitute branching processes in continuous time in which individuals have an arbitrary life-span distribution. The hazard rate is not constant in this case. Thus, we have hazard rate $\lambda(\alpha)$, where a denotes the age of the ancestor. So, these processes are commonly named as age-dependent processes, because the reproduction depends upon the age of the ancestor. In addition, in Bellman-Harris processes there is a single-type of individuals which reproduce independently of each other and the reproduction happens with the same distribution for each particle. A useful note about hazard rate, as Haccou P. et al. notice (Haccou, Jagers and Vatutin (2007)), is that the hazard rate does not affect the expected lifetime reproduction, so for the criticality of the process it does not matter if we have a Markov or a Bellman-Harris branching process, as long as we have the same distribution of offspring numbers. However, the expected population size at a given time is not the same. In addition the population age structure as a function of time differs in the two types of processes (Haccou, Jagers and Vatutin (2007)).

2.3. Branching processes on system reliability analysis.

Stochastic processes and especially Markov chains are powerful mathematical tools and classical approaches in modelling systems transitioning among states (Ye, et al. (2020)). Based on that, Ye et al. (2020) used a continuous time Markov model to represent failures and repairs of processing units (mainly consisting of compressors and pumps) to optimize system's design, reliability and maintenance. A significant number of studies examine reliability of systems by semi-Markov models (Dui, et al. (2015)). To this side, recent as well as classical scientific literature of modelling systems reliability by Markov models are extensively documented and Dawid, McMillan and Revie (2015) had shown a broadly perspective of studies in this field. However, considering the wide variation of stochastic models, latter studies convey the impression that the alternative branching processes fit well to modelling complex systems.

The initial study of Markov models and generating functions that Hermann and Pfaffelhuber (2019) conducted, builds a mathematical framework for the branching processes application to non-biological systems. Despite the fact they use traditional mathematical techniques, they prove crucial branching processes theorems for time-homogenous branching processes with arbitrary distributions with binomial disasters, time-inhomogenous birth-death processes with time-dependent binomial disasters and continuous-state branching processes with binomial disasters.

Cascading failures of electric power transmission systems form classical applications of branching processes in last decades. Dobson, Carreras and Newman (2004) identified the need for models that capture and quantify the risks of load-dependent cascading failure blackouts of power transmission systems due to the impact on society of widespread blackouts occurred in North America in 1996 and 2003. They proved that a branching process is a profitable approach to deal with these risks, while in 2005 as a continued work, they proposed two Galton-Watson branching process approaches, in continuous and discrete time, through a case study with real data from the blackout failures (Dobson, Carreras and Newman (2004, 2005)). Kim and Dobson (2011) extended the aforementioned research by introducing a

saturating Galton-Watson branching process, where cascading failures were modelled. This approach proved very beneficial mainly comparing with other cascading failure models (Kim and Dobson (2011)).

Branching random walks have been also used in reliability theory. Yarovaya (2010) sets a useful theoretical framework for the application of branching random walks in reliability theory, within an analytical birth-death model. Furthermore, these models had been implemented to model ventilation systems in order to study and improve system's reliability and durability. The key concept of this approach, that Gorshkov, Koroleva and Ezhov (2019) proposed, is that the system that was under study, had been segregated in four subsystems and the process was described through degradation groups. These groups seems that were the "memory" of a "memoryless" process, solving the major limitation of using a Markov process in real systems modelling (Gorshkov, Koroleva and Ezhov (2019)).

To carry out reliability analysis of a complex system and prevent breakdowns, Fazlollahtabar and Niaki (2018) used a classical Galton-Watson approach. They modelled a complex system consisting of single and multitype robots and through three numerical applications they suggest how to use extinction probability to predict potential future conditions of the system. A different approach is to model the failures of a repairable mechanical system and not the system itself. Roche-Carrier et al. (2019) introduced an application of branching Poisson process to study the system's reliability and among their findings was that the branching Poisson process offered a physical notion of the fluctuations for the time between failures. Chapter 3 provides a branching process model for the reliability analysis of a refinery pump system.

2.4. Epidemiology models based on branching theory.

In the relevant literature of mathematical modeling of populations considering epidemics, the deterministic models constitute powerful tools when the examined population is large enough. On the contrary, stochastic models can provide useful information when the population size is rather small (Allen (2015)). The theory of branching processes is simple: these processes are often used to model physical systems in which a single-type individual lives for a unit of time and by its death produces *n* identical copies of itself. More complex extensions of that simple process are for instance the multi-type branching processes (as described in sub-section 2.2) (Haccou, Jagers and Vatutin (2007)). The specific construction of the aforementioned process justifies that these processes are usually applied in biology, physics and epidemiology. So, any individual gives rise to a family size, where family sizes are independent and identically distributed (i.i.d) random variables (Athreya and Ney (1972)).

Considering populations, two major issues are important: the expected number of population size and the extinction (more precisely the probability of extinction) (Athreya and Ney (1972); Haccou, Jagers and Vatutin (2007)). If d_k is the probability of extinction at the kth-time period, then the smallest positive root of the probability generating function $h(d) = d$ as d_k tends to d, is detected as the probability of ultimate extinction d. In addition, the first derivate of the probability generating function, that is usually formed as $Q(s) = q_k s^k$, is used to calculate the expected number of population size (m) (Athreya and Ney (1972); Haccou, Jagers and Vatutin (2007)). According to the relevant literature, the expected size of a population (m) could be used as a measure to categorize a branching process into cases and predict the future condition of the system, as follows: If $m > 1$ then $d < 1$: extinction is not guaranteed (supercritical

case), if $m < 1$, then $d = 1$: extinction is guaranteed (subcritical case) and finally if $m = 1$, then $d = 1$: extinction is not guaranteed unless the family size equals 1 (critical case) (Athreya and Ney (1972); Haccou, Jagers and Vatutin (2007)).

S-I-R and S-E-I-R models are two of the basic models for infectious diseases and commonly the branching theory is attached to these models in order to extract useful information about populations such as the probability of transmission extinction or the probability of an outbreak (Allen (2015)). More specifically, a simple single-type branching process could be applied in a S-I-R model in the infectious stage as a birth-death process (Allen (2015)). In a S-I-R model, initially the entire closed population (*N*) is considered as totally susceptible (*S*), a small number of initially infectious individuals introduced into the large susceptible population and cause an outbreak, thus some individuals become infected and infectious (*I*), and *R* means that these individuals are removed from the population either by recovery or death (Jacob (2010); Allen (2015)). Some of the limitations of this approximation is the immunity assumption and the uncertainty based on data (Allen (2015); Bertozzi, et al. (2020)).

A related to S-I-R model is an S-E-I-R model, in which *E* represents the exposed population and implies the assumption of a delay between the exposure and the infectiousness (Bertozzi, et al. (2020)). An S-E-I-R epidemic could be modeled as a multi-type branching process and through the probability generating functions, the probability of an outbreak can be estimated (Allen (2015)). According to Kucharski, et al. (2020), where a mathematical model for the early dynamics of transmission of covid-19 was presented, a major limitation of that kind of approaches is the uncertainty based on data. However, S-E-I-R models although that are more complex compared to S-I-R models, remain simple and provide reliable information about the potential spread of a pandemic disease such as the Severe Acute Respiratory Syndrome (SARS) (Allen (2015)).

The relevant literature of branching processes-based epidemiology models consists of variable complexity models, however data availability in such emerging situations commonly limits the selection power and leads to simple approaches. Chapter 4 provides a branching process epidemiology model.

Chapter 3: Modelling a refinery pump system reliability using Branching Processes.

Maintenance of equipment, nowadays, is of primary importance and great necessity for industry, comparing to the past where the prevailing idea was that there is no need for further attention (Deepak Prabhakar and Dharmaraj (2018); Raghavaiah and HariPrasad (2019)). Maintenance of industrial equipment involves numerous and varied activities (e.g. inspection, testing, replacement etc.), especially when it concerns large-scale plants with complex systems. The increased complexity and the focus on plant viability, efficiency and working safety are factors that force to organize carefully these activities (usually into projects also known as Shutdown/Turnaround projects) and predict the upcoming faults precisely (Langone, et al. (2015); Deepak Prabhakar and Dharmaraj (2018); Raghavaiah and HariPrasad (2019)). In the majority of cases, the proper system maintenance implies the assumption "new" because by the time the equipment is overhauled to manufacturer's standards it is expected to perform as new (Jardine and Tsang (2013); Kumar and Narula (2020)). In addition, reliability, availability and maintainability (RAM analysis) of a system are key factors to evolve a proper reliability-centered maintenance (RCM) plan, which maximizes equipment's life-cycle and minimizes its expenses and maintenance costs (Kumar and Narula (2020)).

Stochastic processes are commonly used to model various technological and industrial systems, calculate reliability indicators (such as the availability, maintainability etc.) and consequently support the development of maintenance plans. Markov models are classes of stochastic processes that are broadly proposed for reliability system analysis as well as supporting decision tools for maintenance strategies (Dawid, McMillan and Revie (2015); Gabe, Freire and de Andrade (2019);Ye, et al. (2020)). However, Markov chains are characterized by limitations; for instance, the most obvious is the compulsory use of exponential distribution, which is usually an unrealistic assumption (Gabe, Freire and de Andrade (2019)). According to Roche-Carrier , et al. (2019) there are four stochastic processes that are widely used to approximate the reliability of a repairable system: the renewal process, the homogenous and non-homogenous Poisson processes and the branching Poisson process. Despite the fact that branching processes are usually applied in biology, physics and epidemiology, in recent years researchers have introduced approaches for modeling technological systems so to confront with the uncertainty of failures and their impact (Kim and Dobson (2010); Fazlollahtabar and Niaki (2018); Gorshkov, Koroleva and Ezhov (2019); Roche-Carrier , et al. (2019)).

This application combines a classical approach of Markov process with a branching process to propose an alternative perspective on reliability system analysis. An industrial unit consisting of three types of pumps is modelled. Raw data from industry are used to calculate the failure probability of each type of pumps through Markov models. Then, within branching process theory, the failure probability of the entire system is evaluated. Numerical applications amplify the outcomes by evaluating: (i) the probability of ultimate extinction of the system and the expected number of working components, (ii) the failure probability for every month during a year, both for each type of pumps and the entire unit.

3.1. System description.

The heart of a petrochemical industry is the refinery, which usually consists of multiple and complex systems with various operations. Fluid moving machines, such as compressors and pumps, add energy to a system and transfer the proper amount of gas and fluids, respectively, through pipelines. Pumps are considered as a very important part of these industries and several types of them are broadly used (Azadeh, Ebrahimipour and Bavar (2009); Chaudhuri (2010)). Centrifugal and positive displacement pumps are two of the principal types of pumps (Chaudhuri (2010)).
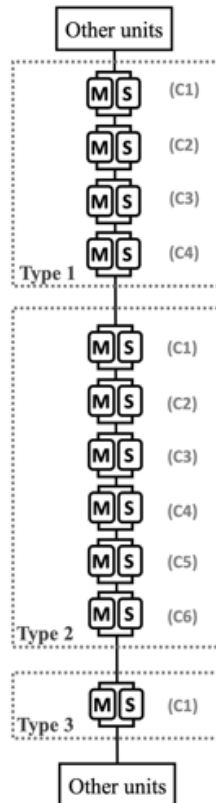
Figure 5. Construction of the pump system (unit).

In petrochemical industries, where the equipment is constantly exposed to weather conditions and dirt, the damage is inevitable. The risk of damage is higher for some equipment, like pumps that impulse liquid, due to their nature. For instance, the positive displacement pumps that are run by sticky fluids, could be completely corrupted due to cylinder or piston damage (Chaudhuri (2010)). According to Azadeh et al. (2009), the majority of failures in a plant are correlated to comphressors, piping and pumps. However, these industries are familiarized with concepts like maintainability or availability and have already adopt maintenance strategies (Azadeh, Ebrahimipour and Bavar (2009)).

This paper deals with a unit consisting of pumps located in a large-scale plant and classified in three types: 8 vertical centrifugal pumps with multiplier (type 1), 12 horizontal centrifugal pumps (type 2) and 2 positive displacement pumps (type 3). Each pump, in any type, comes along with its spare pump, for instance in type 1 there are 4 main and 4 spare pumps etc. In addition, each set of main and spare

pumps is considered as a component. Figure 5, shows in detail the examined system. The unit requires only 4, 6 and 1 pumps of each type respectively to operate. However, in a refinery where the production is continuous, the availability of the units and consequently of the pumps is essential. Thus, spare pumps, which are initiated by the time the main pumps fail, are of great necessity. However, despite redundancies, a failure free system operation is not guaranteed.

3.2. Mathematical modeling and formulation.

Our aim is to model a real system with scope to study main principles of systems reliability theory. The primary pursuit is to propose a model that uses a classical approach of Markov processes to identify the current situation of the system, and the alternative methodology of branching processes to predict future situation and potential failures. So, the proposed model is based on two main pillars, the Markov processes, and the branching processes.

3.2.1. Markov model.

Given that raw data consist of the times that a failure occurs between same-type pumps and the corresponding repair times, we grouped data by the nature of the failures in order to estimate via the use of Maximum Likelihood Estimators (MLE) the mean time between failures and the mean time to repair a failure that is caused by a specific failure of pump. Consequently, we estimated $\hat{\lambda}_i$, $\hat{\mu}_i$ for every type of pumps, where *i* stands for every possible failure.

Each type of pumps is attached with *i* pairs of λ_i and μ_i, where *i=1,2,3,4,5* is related to the possible failed pump's state. If *n* defines the number of possible failures of a same-type pump and each state is related to failures, then every pump could be in *n+1* states. State 0 represents the operational state of a pump, though in states *1, 2, …, n* the pump is non-operational due to a certain type of failure. Based on experts' opinion centrifugal pumps (vertical and horizontal) are characterized by five main failures/repairs and positive displacement pumps by four.

Vertical centrifugal pumps (type1): gear box, seal, check valve, general repair and other. Horizontal centrifugal pumps (type2): gear box, seal, general repair, cobbler and other. Positive displacement pumps (type3): filter, general repair, ball bearing, cobbler. "General repair" and "other" depict respectively sets of major and minor failures that occurred together. Figure 6, incorporates all the aforementioned information.

We define as state probabilities $\pi_i(t), i \in S$ for any $t \geq 0$ as applies in equations (14), (15) :

$$\pi_i(t) = P(X(t) = i) \tag{14}$$

$$\sum_{i \in S} \pi_i(t) = 1 \tag{15}$$

where $\{X(t), t \geq 0\}$ is a continuous time Markov process with state space S.
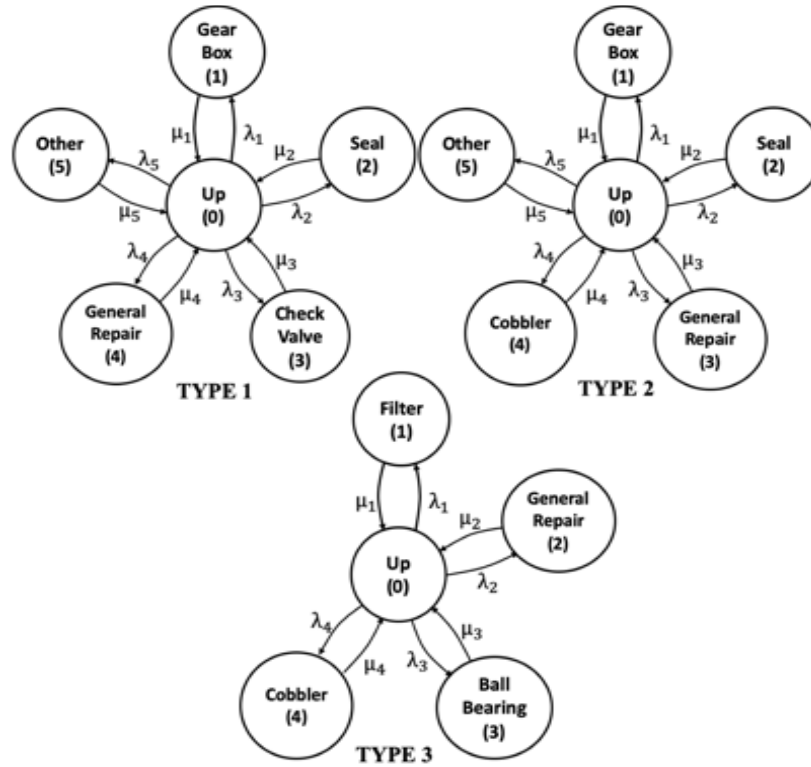
Figure 6. State transition diagrams of type 1, type2, type 3.

The estimation of $\pi_i(t)$ is usually rough. However, since the examined unit is designed to operate continuously in time, we are mainly interested in deriving the asymptotic state probability distribution $\boldsymbol{\pi} = [\pi_0, \pi_1, \dots]$ by solving the following system of linear equations :

$$\boldsymbol{\pi Q} = \boldsymbol{0}$$

$$\sum_{i \in S} \pi_i = 1 \tag{16}$$

where $\boldsymbol{Q}$ is the rate transition kernel among system states. The relation shown in equation (17) provides the failure probability of each pump according to stationary distribution. These probabilities are needed for calculating failure probabilities and reliability of components for every type of pumps by using equation (18) and (19), based on reliability theory (Frankel (1984)).

$$p_{Fail} = 1 - \pi_0 \tag{17}$$

$$p_{fail\ ci} = 1 - p_{fail}^2 \tag{18}$$

$$R_{Ci} = 1 - (1 - p_{failci})^2 \tag{19}$$

where $p_{fail\ Ci}$, $i = 1, 2, 3$ is the failure probability for a set of a main and a spare pump connected in parallel (thereafter called *component*) for pumps of type i. The estimation of the probability to have $0, 1, 2, \dots$ components available ($p_i, i = 0, 1, 2, \dots$) arises from binomial distributions.

23

3.2.2. Branching model.

Branching processes are often used to model physical systems in which an individual lives for a unit of time and by its death produces *n* identical copies of itself. So, any individual gives rise to a family size, where family sizes are independent and identically distributed random variables (i.i.d) (Fazlollahtabar et al. (2018)). Note that, two major issues are important to discuss considering populations: the expected number of population size and the extinction (Athreya and Ney 1972).

Considering that instead of individuals our population consists of sets of pumps, the probability of extinction will determine the probability that the system stops working, while the expected number of population size will define the expected number of available pumps. If we have *n* available pumps in k^{th} time period, then in (*k+1*) time period we will have $X_1 + X_2 + \cdots + X_k$ available pumps, where X_k are i.i.d random variables. We define the probability that the system stops working at k^{th} period as d_k. In order for the system to stop working in period *k*, one component should die at least in the (*k-1*) period. Thus, we determine d_k by equation (20) and reform it in equation (21):

$$d_k = (p_0 + p_1 + p_2 + \cdots) + p_i d_{k-1} \tag{20}$$

$$d_k = h(d_{k-1}) \tag{21}$$

For estimating the probability of ultimate extinction *d*, and as d_k tends to *d*, we need to solve Eq. (22).

$$d = h(d) \tag{22}$$

Since we are interested in the probability of extinction, *d* is detected as the smallest positive root (as it is referred in sub-section 2.2) of equation (22). For instance, for the system that it is addresses, the generating functions for each type are formed respectively as follows:

$$h(d_{t1}) = (p_0 + p_1 + p_2 + p_3) + p_4 d_{t1} \tag{23}$$

$$h(d_{t2}) = \left(\sum_{i=1}^{5} p_i\right) + p_6 d_{t2} \tag{24}$$

$$h(d_{t3}) = p_0 + p_1 d_{t3} \tag{25}$$

where $d_{ti}, i = 1, 2, 3$ is the extinction probability of each subset. The probability of extinction *d*, for the entire system, according to reliability theory for components linked in series, is estimated by equation (26) (Frankel (1984)):

$$d = 1 - (1 - d_{t1})(1 - d_{t2})(1 - d_{t3}) \tag{26}$$

In addition, the derivate of probability generating function is used to calculate the expected number of working pumps (*m*) (Athreya and Ney (2004)). For the pump system, the equations that give the expected number of working pumps for each set are formed as follows:

$$m_1 = p_4 \tag{27}$$

$$m_2 = p_6 \tag{28}$$

$$m_3 = p_1 \tag{29}$$

According to the relevant literature, the expected size of a population (m) could be used as a measure to categorize a branching process into cases and predict the future condition of the system, as follows (Athreya and Ney (2004); Haccou et al. (2007)):

- If $m > 1$, then $d < 1$ extinction is not guaranteed (supercritical case)

- If $m < 1$, then $d = 1$ extinction is guaranteed (sub-critical case)

- If $m = 1$, then $d = 1$: extinction is not guaranteed unless the family size is constant to 1 (critical case)

3.3. Numerical application and results.

Initially, we set the Markov model of each type of pumps according to mathematical formulation. The aim of this modelling approach is to obtain the steady-state probability distribution and eventually the failure probability for each pump of every type. In order to have a definite formatting of the models, we have to estimate the failure rate λ, the repair rate μ, and illustrate the transition diagrams for every pump type (Figure. 2). Table 1 presents the failure and repair rates that are estimated according to mathematical formulation and additionally to Fig. 2, provides a comprehensive view of the system.

Based on the models for the three subsystems, we can calculate through equations (15) and (16) the steady-state probabilities π_i, where *i=0,1,2,3,4,5* corresponding to the systems' states. Table 2 presents the steady state probability distribution for each type of pumps.

According to this analysis, type 1, 2 and 3 pumps operate with probabilities 0.884, 0.9686 and 0.99431 respectively. Thus, we can now evaluate the failure probability of any pump, by the corresponding formulas in equation (17). Consequently, a vertical centrifugal pump (type1) fails with probability 0.1776, a horizontal centrifugal pump (type2) fails with 0.0314 and a positive displacement pump fails with probability 0.0057. In addition, Table 3 presents the failure probability and reliability of a component for all types of pumps according to equation (18) and equation (19) respectively, crucial elements in order to evaluate the probability to have *i* working components.

3.3.1. The probability of ultimate extinction.

The probability of ultimate extinction and the expected size of population are two of the most critical information that can be extracted from a branching process application. Based on these measures we can predict the future condition of the system with satisfying accuracy. The output data provided by the Markov analysis are the failure probabilities for each pump, classified by their type. Recall that every pump of type 1, 2, 3 fails with probabilities calculated by equation (17).

Recall also that we have 8 pumps of type 1, 12 pumps of type 2, and 2 pumps of type 3, where the unit requires 4, 6 and 1 operational pumps respectively. By the time a main pump fails, its spare pump starts working immediately (we assume perfect control switch among the primary and the spare pumps). It is irrelevant if the functioning pumps are all main, spare or a combination of them as long as the aforementioned requirement is met.

We then define as a set, the pumps of the same type, so set 1 refers to pumps of type 1 and so on. The probability needed to calculate the probability of ultimate extinction for each set of pumps is the probability to have 0, 1, 2,… working components, which is referred as p_i, for $i = 0,1,2,3,4,5,6$, where i is the number of working components. These probabilities are calculated based on binomial distribution and presented in Table 4.

Table 1. Failure and repair rates (per month) of the system, where i=1,2,3,4,5.

Type 1		Type 2		Type 3	
$\hat{\lambda}_i$	$\hat{\mu}_i$	$\hat{\lambda}_i$	$\hat{\mu}_i$	$\hat{\lambda}_i$	$\hat{\mu}_i$
0.088	2.31	0.1	12.3	0.0047	30.43
0.087	5.67	0.133	14.2	0.0046	30.43
0.114	22.8	0.031	4.95	0.0163	6.09
0.17	1.88	0.02	30.4	0.0167	6.09
0.513	7.73	0.126	14.9		

Table 2. State probabilities π_i, for each type of pumps.

	Type 1	Type 2	Type 3
π_0	0.8224	0.9686	0.99431
π_1	0.02616	0.0079	0.000153
π_2	0.01265	0.0091	0.000151
π_3	0.0041	0.0055	0.002663
π_4	0.07479	0.0006	0.002724
π_5	0.05463	0.0082	-

Table 3. Reliability and failure probability of a component for each type of pumps (R_{Ci} and $P_{fail,Ci}$, where $i = 1,2,3$).

	Type 1	Type 2	Type 3
p_{failCi}	0.03154176	0.00098596	0.0000324
R_{Ci}	0.062088637	0.001970948	0.0000648

Table 4. Probabilities of *i* working components for each type.

	Type 1	Type 2	Type 3
p_0	0.773832898	0.988232429333	0.9999352
p_1	0.204907338	0.0117096067109	0.0000648
p_2	0.020346939	$5.78115048 \cdot 10^{-5}$	
p_3	$8.9796 \cdot 10^{-4}$	$1.52224644 \cdot 10^{-7}$	
p_4	$1.4861 \cdot 10^{-5}$	$2.2546451 \cdot 10^{-10}$	
p_5		$1.7810255 \cdot 10^{-13}$	
p_6		$5.862068 \cdot 10^{-17}$	

Each set of pumps could possibly have 4, 6 and 1 components maximum, working at a time, thus empty columns appear in Table 4. To evaluate the probability of ultimate extinction for every type, three generating functions, one for each type respectively, are formed as in equations (23), (24) and (25). Solving equation (22) by integrating equations (23), (24) and (25), provides the smallest positive real solution as probabilities of extinction for each set of pumps. So,

$$d_{t1} = d_{t2} = d_{t3} = 1$$

and $d = 1$, based on equation (26). Knowing d for each type of pumps enables us to simply evaluate the expected number of working pumps by using the derivatives of equations (27), (28), and (29). Therefore, the expected number of type 1, 2, 3 working pumps is:

$$m_1 = 0.00001486$$

$$m_2 = 5.86206788 \cdot 10^{-17}$$

$$m_3 = 0.0000648$$

Values of probability of extinction for each type seem reasonable as we observe a non-maintained unit in infinity. As it was predicted, the expected number of working components remains small, as long as the probability of ultimate extinction of the unit is high. In addition, these results conform to branching process theory, which indicated that if $m < 1$, then the extinction is guaranteed. A supplementary approach, analyzed in subsection below, will provide an intuitive idea of the future condition of the unit.

3.3.2. Failure probability during a year.

A reasonable question now that the probability of ultimate extinction is known, is what happens to the system in shorter time intervals? This approach adds information about the unit because it aims to demonstrate the impact of the time to the entire system. Failure probabilities for every type in 12 time units (1 year) are presented in Table 5-7 and illustrated in Figures 7-9. The entire system stops working when one component stops operating. Thus, the probability of the system to stop working arises by equation (26) that evaluates the probabilities of three types for 12 time periods. Table 8 presents the results and Figure 10 illustrates them in a bar-chart.

By Tables 5-8, we observe that after one period the probability of each subsystem remains stable to 1 and subsequently corresponding probability of the unit is 1 for the entire year. The main reason of this impact is that the initial probabilities are too low. Despite that, this approach stands useful because it offers an easy and fast identification of the time that the maintenance of the system, or subsystems, should be applied and slightly supports decisions about which maintenance strategy to adapt for this unit. For instance, pumps of type 1 and 3, which in first three months have probability to fail lower than 1, may be excluded from a shutdown project of preventive maintenance in a scenario where other equipment is more critical, because it seems that small/routine proactive maintenance actions are enough for the smooth operation of the unit.

Table 5. Failure probability of type 1 during a year.

Period	Probability of type 1 components stop working
1	0.999985138983549
2	0.999999999779150
3	0.999999999999997
4	1.00000000000000
5	1.00000000000000
6	1.00000000000000
7	1.00000000000000
8	1.00000000000000
9	1.00000000000000
10	1.00000000000000
11	1.00000000000000
12	1.00000000000000

Table 6. Failure probability of type 2 during a year.

Period	Probability of type 2 components stop working
1	1. 00000000000
2	1. 00000000000
3	1. 00000000000
4	1. 00000000000
5	1. 00000000000
6	1. 00000000000
7	1. 00000000000
8	1. 00000000000
9	1. 00000000000
10	1. 00000000000
11	1. 00000000000
12	1. 00000000000

Table 7. Failure probability of type 3 during a year.

Period	Probability of type 3 components stop working
1	0.99993520000000
2	0.99999999580096
3	0.99999999999973
4	1. 0000000000000
5	1. 0000000000000
6	1. 0000000000000
7	1. 0000000000000
8	1. 0000000000000
9	1. 0000000000000
10	1. 0000000000000
11	1. 0000000000000
12	1. 0000000000000

Table 8. Failure probability of system during a year.

Period	Probability of the unit stop working
1	1. 00000000000
2	1. 00000000000
3	1. 00000000000
4	1. 00000000000
5	1. 00000000000
6	1. 00000000000
7	1. 00000000000
8	1. 00000000000
9	1. 00000000000
10	1. 00000000000
11	1. 00000000000
12	1. 00000000000
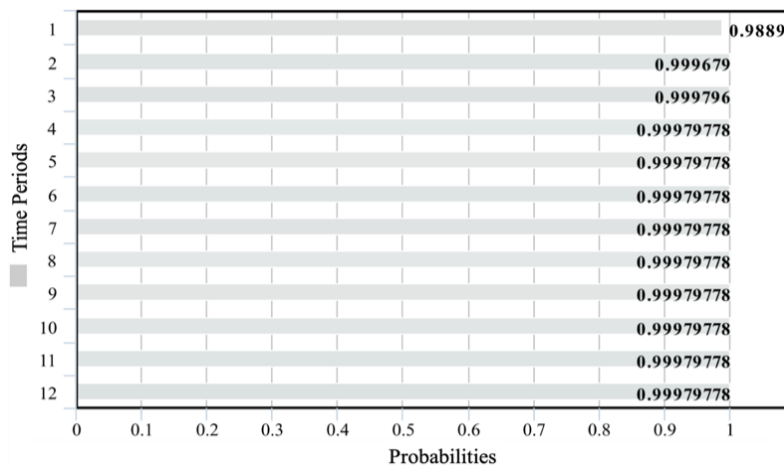
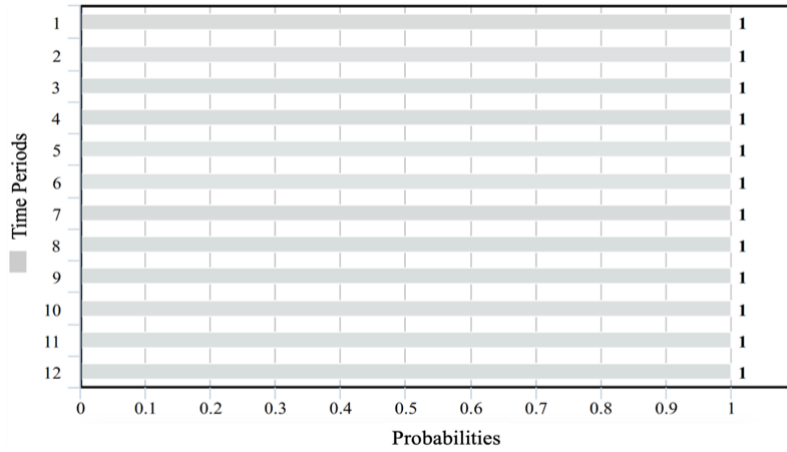Figure 7. Probabilities bar chart of type 1 pumps.

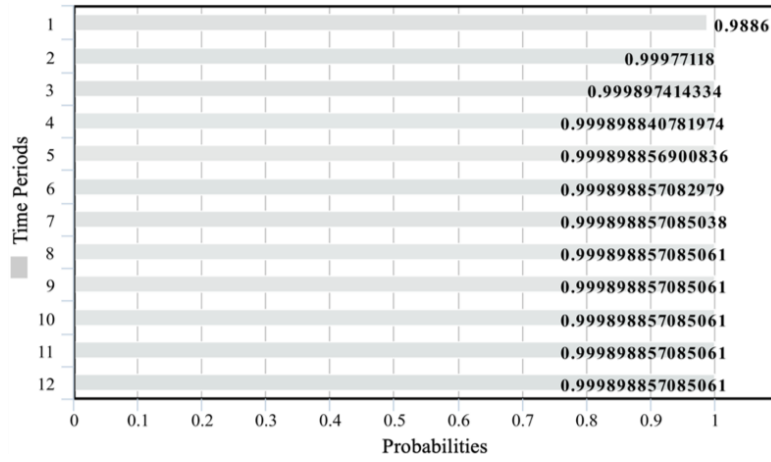Figure 8. Probabilities bar chart of type 2 pumps.
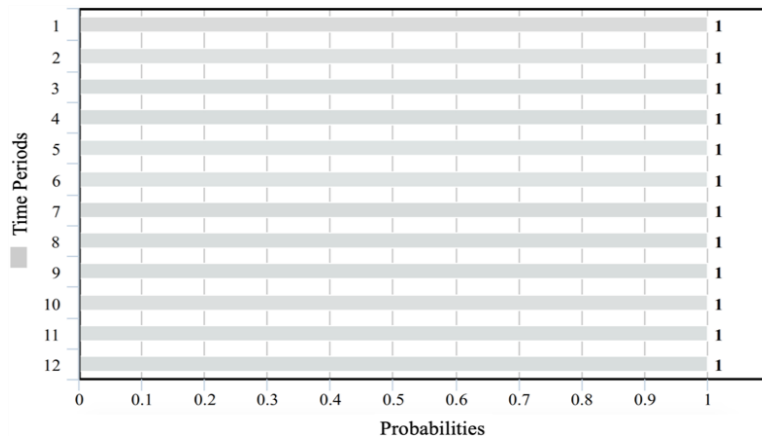

Figure 9. Probabilities bar chart of type 3 pumps.

Figure 10. Probabilities bar chart of the entire system.

3.3.3. Application overview and results discussion.

This application considers reliability analysis of a plant unit consisting of mechanical repairable through branching process modelling. The system is constructed in sets that represent types of main and spare pumps, where each set is considered as a component. A Markov model was built to process raw data from industry and set the basis for the branching model. The probability of ultimate extinction that was evaluated both for each type of pumps and the entire unit, the probability of extinction for 12 time units (1 year) and the expected number of working components could be used as reliability measures to predict future potential conditions of the system.

The proposed approach is suggested for the reliability analysis of large-scale units consisting of single or multi-type equipment, because comparing to other stochastic models it is simpler and provides reliable results while minimizes the computational effort. Thus, we conclude that the contribution of this analysis is beneficial for maintenance departments, because supports maintenance decisions and slightly offers scope optimization of industrial maintenance projects.

Chapter 4: A Branching Process model for the novel coronavirus (covid-19) spread in Greece.

Coronavirus disease 2019 (covid-19) is an infectious disease caused by severe acute respiratory syndrome (SARS). It was initially detected in December 2019 and has been rapidly spread globally inducing the ongoing pandemic (Ndaïrou, et al. (2020)). This novel coronavirus pandemic is deliberated as the biggest worldwide threat and on January 30th, 2020 World Health Organization (WHO) advocated that this new situation is a Public Health Emergency of International Concern, because of the thousands infected cases that were reported and deaths around the globe (Ndaïrou, et al. (2020); Ivorra, et al. (2020)). In particular, by April 1st, 2020, 872,481 confirmed cases and 43,275 deaths were reported (Ivorra, et al. (2020)). The first case in Greece was confirmed on February 26th (2020) and by April 21st, 2,401 cases and 121 deaths were reported, even though health and state authorities had applied small-scale control measures, such as the suspension of educational institutions and the closing of cafes, bars, restaurants sports facilities etc. (NPHO (2020)). On March 22nd, the Greek government announced a general lockdown with movement restrictions and after 42 days of quarantine, when the number of daily reported cases decreased to 10, state authorities gradually repealed the restrictions (NPHO (2020)). These control measures, that were among the strictest in Europe, were initially considered as highly effective and whereas the pandemic was internationally ongoing, the case of Greece was treated as a success story (Tugwell and Nikas (2020)). However, at the time of this revision the numbers have been increased to 82,034 total cases and 1,288 deaths. These fluctuations on numbers attract the interest of researchers to model the transmission of the virus to evaluate and quantify the dynamic of the pandemic (Ndaïrou, et al. (2020)).

Mathematical models of infectious disease transmission effectively describe and simply depict the evolution of diseases by providing quantitative data in epidemiology (Ivorra, et al. (2020); Wang (2020); Ahmed, et al. (2020)). Moreover, using mathematical and statistical tools enables to conduct long or short-term forecasts about the spread in population so to support decisions of intervention strategies, outbreak and healthcare management and policy development (Ivorra, et al. (2020); Wang (2020); Ahmed, et al. (2020); Overton, et al. (2020)). There exist several mathematical models in epidemiology, starting from the simple S-I-R models (Susceptible-Infected-Recovered models), to S-E-I-R models (Susceptible-Exposed-Infected-Recovered models), to more complex proposals (Allen, Jang and Roeger (2017); Ndaïrou, et al. (2020); Wang (2020)). A classical approach of data analytics of virus transmission is described by Galton-Watson branching process and a promising approach that uses advanced mathematical modelling to connect models with machine learning was introduced by Yang *et al.* (Wang (2020); Cruz (2020); Yang, et al. (2020)). However, the majority of analyses focus on human-to-human transmission and this implies two important limitations (Wang (2020); Cruz (2020); Helmy, et al. (2020)):

1. contact tracing data are required and data availability in such emerging cases is limited.

2. the assumption that transmission rates are considered as fixed for mathematical analysis simplicity seems unrealistic.

The main goal of this application is to develop a simple mathematical model using branching processes in order to analyze the initial and latter stages of the outbreak of covid-19 in Greece and

evaluate crucial epidemiological indicators such as the reproduction number and the probability of extinction of the transmission. Finally, through an exponential smoothing model, in a complete absence of control measures a scenario is developed in order to effectively describe the virus transmission in case of Greece and discuss the effectiveness of control measures.

4.1. Branching epidemiology model.

A branching process approach is considered as classical approximation for epidemics (Lange (2010)). In addition, whereas the number of infected people is small, it is believed that people behave independently, hence branching processes can sufficiently model early stages of an epidemic (Lange (2010)). If an infected individual causes $0, 1, 2, …$ new infections with probabilities $q_0, q_1, q_2, …$, then the progeny generating function is formed as $Q(s) = q_k s^k$ and represents the offspring distribution, where q_k is the probability that an infected individual causes k new infections, in a total susceptible population, before s/he dies or recovers from the infection (Lange (2010); Blumberg, Funk and Pulliam (2014)). It is proven that a two-parameter offspring distribution is more beneficial over one-parameter distribution to model large datasets of infectious diseases, because in such a case the model depicts better the actual transmission of the virus assuming the adaption of the parameters (White and Pagano (2008)). Thus, generally it is considered that the offspring distribution is a negative binomial distribution with parameters (R_0, k) (Blumberg and Lloyd-Smith (2013); Zhang, Ye and Lord (2019)). To that side, assuming that variable X refers to a sample of counts of infected cases, the pdf of X is formed as shown in equation (30) and consequently the progeny function is formed as shown in equation (31) (Blumberg and Lloyd-Smith (2013); Zhang, Ye and Lord (2019)):

$$P(X = x) = \frac{\Gamma(x + k)}{x! \, \Gamma(k)} \left(\frac{R_0}{R_0 + k}\right)^x \left(\frac{k}{R_0 + k}\right)^k \tag{30}$$

$$with \, \Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$$

$$Q(z) = (1 + \frac{R_0}{k}(1 - z))^{-k} \tag{31}$$

The basic reproduction number, R_0, is the average number of secondary cases infected by one individual in a total susceptible population, and k is the dispersion parameter which measures the transmission heterogeneity of the virus (Zhang, et al. (2020)). In fact, the dispersion parameter as a transmission heterogeneity measure, quantifies the variability in the number of secondary cases and could be described as a measure for the impact of the superspreading events (lower values of k signifies higher heterogeneity and higher impact of superspreading events) (Zhang, et al. (2020); Riou and Althaus (2020)). However, according to the relevant literature, the offspring distribution is additionally formed as a Poisson or a geometric distribution, which are nested cases of the negative binomial distribution, with $k \to \infty$ and $k \to 1$ respectively (Blumberg and Lloyd-Smith (2013); Waxman and Nouvellet (2019)).

As it is already mentioned, the expected number of population size and the probability of extinction are considered as crucial information about the evolution of a population. Moreover, taking

into account that we focus on the spread of a disease, the expected number of population size stands for R_0 and the probability of extinction refers to the extinction of transmission. Thus, if z is the extinction probability, then $(1-z)$ is the probability of an outbreak. According to branching processes theory, in order to estimate the probability of ultimate extinction z, we need to solve equation (32):

$$Q(z) = z \tag{32}$$

Since we are interested in the probability of extinction, z is detected as the smallest positive root of equation (32) and the derivate of the probability generating function is used to calculate the expected population size. However, in order to compute the probability of extinction, we need to estimate the parameters R_0 and k firstly, because these parameters are critical elements of the probability generating function (equation (30)), and secondly because R_0 is a typical threshold to detect the transmissibility and determines the epidemic potentiality of virus in the absence of control measures (WHO (2003)). More precisely, if the basic reproduction number is $R_0 > 1$, in a total susceptible population, then eventually an epidemic occurs; contrariwise, if $R_0 < 1$, the transmission is self-limited and if $R_0 = 1$, the state is characterized as endemic (endemic equilibrium) (Delamater, et al. (2019); De Serres, Gay and Farrington (2000)).

Similar to the basic reproduction number R_0, R_{eff} is the effective reproduction number, that represents the average number of secondary infected cases under control measures (WHO (2003)). Likewise R_0, R_{eff} determines the epidemic potentiality of virus, but under control measures. Consider an epidemic scenario where an outbreak of a virus starts. In the initial stages there is absence of mitigate interventions and as the transmission continues, the government decides to adapt control measures (such as quarantine or massive vaccination etc.) in order to limit the spread of the disease. In fact, the government attempts to reduce an $R_0 > 1$ to an $R_{eff} < 1$ to bring the outbreak under control. So, we can intuitively assume that as long as the basic reproduction number categorizes the spread of a virus as an epidemic or not, the effective reproduction number evaluates measures' effectiveness (WHO (2003)). Despite the fact that the reproduction number is evaluated differently, the entire process for estimating the probability of ultimate extinction remains identical.

4.1.1. Estimating the parameters: R_0, R_{eff}, k.

Considering a population consisting of particles that are able to reproduce independently (offspring) and a typical predecessor i could produce N_i ancestors, then X_g is the size of gth generation, where $g = 0, 1, 2, \dots, G$ for the first G generations and X_g can be computed as the sum of all offspring that each particle produced a generation before $\left(X_g = \sum_{i=0}^{X_{g-1}} N_i \right)$. Then, the described process is a discrete time branching process and N_i are independently and identically distributed (i.i.d) random variables. In addition, the process X_g, $\{X_g : g = 0, 1, 2, \dots, G\}$, has the Markov property, that is the value of X_g depends only on X_{g-1} (Yan and Chowell (2019)). One of the major properties that arises form branching process theory is that $R_0 = E[N_i]$. This property along with the Markovian property of the process lead to an efficient estimator for R_0, the Harris estimator (White and Pagano (2008); Yan and Chowell (2019)):

$$\widehat{R_0} = \frac{\sum_{g=1}^{G} X_g}{\sum_{g=1}^{G} X_{g-1}} \tag{33}$$

However, in the relevant literature there exist several methods to estimate basic reproduction number R_0, such as the Maximum Likelihood Estimator (MLE) (Obadia, Haneef and Boëlle (2012)). Despite the fact that the notions of R_{eff} and R_0 are close, their evaluation differs. In order to allow the calculation of R_{eff}, the referring population should be divided into chains or clusters of transmission and in some cases contacting data are needed (Blumberg, Funk and Pulliam (2014)). Nevertheless, a common obstacle for researchers that conduct studies for diseases and spread of viruses such as the corona virus that causes severe respiratory syndrome (SARS), is the lack of data (Helmy, et al. (2020)). In this approach, due to limited data availability, we consider one cluster with a limited number of chains. Thus, we assume that the entire cluster has only one R_{eff} (and k, which constitute the cluster's parameters), is evaluated as follows:

$$R_{eff} = k\left(\frac{1}{p} - 1\right) \tag{34}$$

where k is the dispersion parameter and *p* is the scale parameter of the negative binomial distribution of X_g. The smoothest way to evaluate the parameters of a negative binomial distribution is by using the method of moments (Method of Moments Estimator-MME), where by equating the sample mean $\bar{y}$ to the population mean μ, and the sample variance S^2 to the population variance σ^2, the dispersion and scale parameter are estimated through data by solving equations (35) and (36) respectively (Al-Khasawneh (2010)):

$$k = \frac{\bar{y}^2}{S^2 - \bar{y}} \tag{35}$$

$$p = \frac{S^2 - \bar{y}}{S^2} \tag{36}$$

Despite the fact that this approach simply and effectively evaluates dispersion parameter through data, it is proven that several limitations appear, such as (Al-Khasawneh (2010)):

- If the sample variance is higher than the sample mean, then *k* is very large.
- If the sample variance is less than the sample mean, then *k* is negative.
- If the sample variance equals the sample mean, then *k* is not defined.

4.2. Numerical application: the case of Greece.

The first imported case of COVID-19 in Greece was confirmed on February 26th, when an individual came back from a visit to Northern Italy. Most of the subsequent cases in February and early March were related to people that had been to Italy, Israel and Egypt or their contacts. Until the 22nd of March, when Greek government announced a general lockdown, 624 cases and 2 deaths were reported. The basic reproduction number in March 22nd was estimated to $R_0 = 1.18$ according to (3). Table 9 and Figure 11 present all reproduction number from the 27th of February to the 22nd of March.
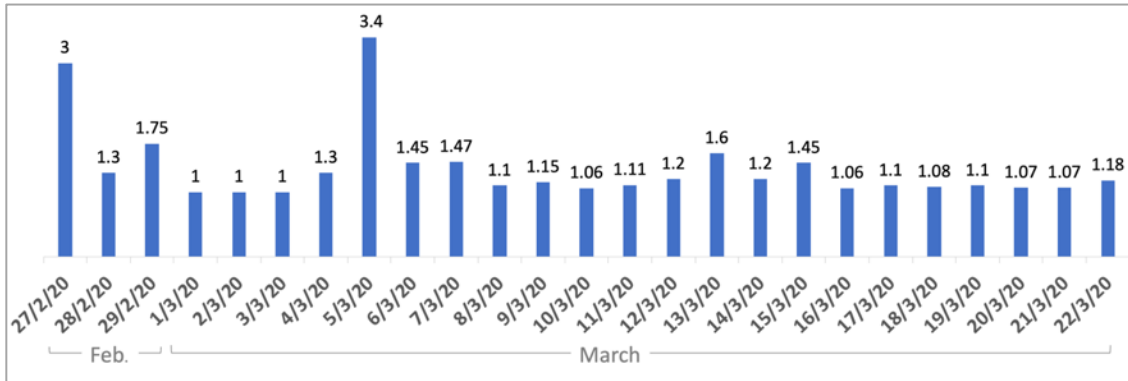
Figure 11. Reproduction numbers from 27th February to March 22.

Table 9. Reproduction numbers from 27th February to March 23.

Date	R0	95% CI	k
27/2/20	3		
28/2/20	1.33	(-1.06 , 3.72)	
29/2/20	1.75	(0.6 , 2.9)	
1/3/20	1	(-0.42 , 2.42)	
2/3/20	1	(0.01 - 2.01)	
3/3/20	1	(0.1 , 1.9)	
4/3/20	1.29	(-0.59 , 1.98)	
5/3/20	3.44	(-0.02 , 6.9)	
6/3/20	1.45	(-0.03 , 2.93)	
7/3/20	1.47	(0.14 , 2.79)	
8/3/20	1.1	(-0.38 , 2.58)	
9/3/20	1.15	(-0.15 , 2.45)	
10/3/20	1.06	(-0.23 , 2.35)	0.74
11/3/20	1.11	(-0.04 , 2.26)	
12/3/20	1.18	(0.15 , 2.21)	
13/3/20	1.6	(0.14 , 3.06)	
14/3/20	1.2	(0.15 , 2.25)	
15/3/20	1.45	(0.33 , 2.57)	
16/3/20	1.06	(-0.21 , 2.33)	
17/3/20	1.1	(-0.02 , 2.22)	
18/3/20	1.08	(-0.03 , 2.19)	
19/3/20	1.11	(0.1 , 2.12)	
20/3/20	1.07	(0.04 , 2.1)	
21/3/20	1.07	(0.08 , 2.06)	
22/3/20	1.18	(0.99 , 1.37)	

As it can be obtained in Table 9, the basic reproduction number is firmly greater than 1 for the entire period before quarantine. More precisely, the day of the announcement of the quarantine, R_0 was estimated to 1.18 (95% CI: (0.99, 1.37)) according to equation (33), the dispersion parameter *k* was 0.74 according to equation (35) and based on these, the probability of spread extinction was estimated to 0 according to equation (32). To this side, the probability of an outbreak was 1, a fact that according to the relevant branching processes literature, advocates the adoption of mitigation measures such as a general lockdown and quarantine. Greek government, through these measures, attempted to reduce the basic reproduction number to $R_{eff} < 1$, so to bring the outbreak under control.

In order to use branching processes for the estimation of R_{eff}, the population was divided into clusters/chains of transmission. We assume that Greece is considered as one cluster and each region as one transmission chain. In fact, due to limited data availability, we presume that the entire Greece composes one cluster consisting of 13 transmission chains, one for each region. Figure 12 depicts the percentage geographical distribution of total cases per region through quarantine period.
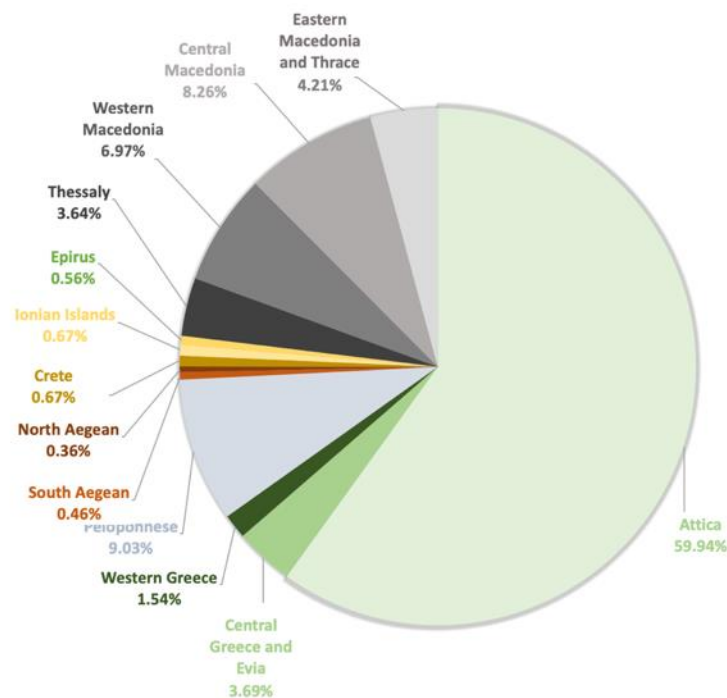

Figure 12. Geographical distribution of total cases per region during quarantine period.

By the end of quarantine, the 4th of May that the new cases per day were under 10, the effective reproduction number R_{eff} was estimated to 0.0003 (95% CI: (-6.7 , 6.7)) according to equation (34) which indicates that the number of secondary infections caused by one infected individual is noticeably reduced and the dispersion parameter (*k*) was 0.23. According to the new estimation of the parameters, the probability of the spread extinction was 1, an approximation that forces the probability of an outbreak to

be 0. These findings prove that the general lockdown was a highly effective measure so to control a general spread of the virus.

However, the available data, from National Public Health Organization (NPHO) of Greece for the summer period, revealed a different evolution of the virus transmission despite the fact that Greek government has done smaller mitigation interventions such as the closing of bars and restaurants at 12p.m. and the extensive usage of masks. More precisely, by the end of the summer period almost 300 new cases per day were reported. Repealing the general lockdown along with the opening of country's borders and allowing the normal operation of stores relieved Greek economy but led to an uncontrolled rise of the transmission, brought new imported cases and transmission chains roughly detected. Within this widespread of corona virus in Greece during summer period, data was noisy and deprived models' accuracy. This scenario could have been seen in the complete absence of initial control measures. Thus, in order to scheme an assumptive state of the virus transmission in which no mitigation actions were adopted, we conduct an additional analysis.

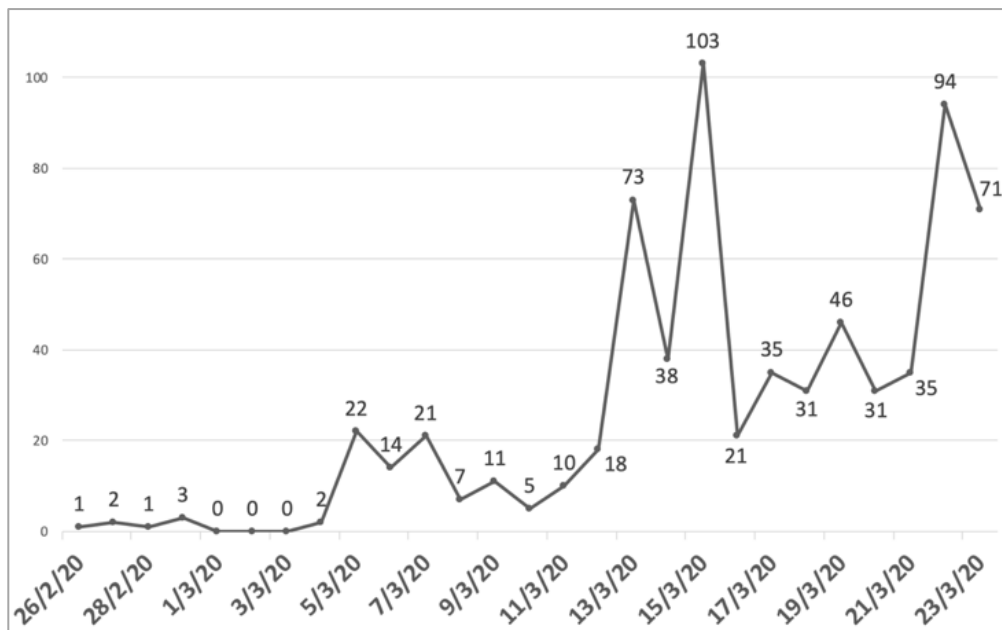### 4.2.1. The non-mitigation scenario.

Figure 13. Real data of reported new cases of corona virus transmission without mitigation actions.

A simple Exponential Smoothing (ETS) model is used to forecast the transmission of corona virus in Greece in total absence of control measures. This model is actually used in order highlight the necessity and the effectiveness of quarantine measures. Real data, from the initial phase of the transmission (Figure 13) in Greece until the 23th of March, was the pillars of the model and the output is presented in Table 10 as new cases per day for the period: 24th of March to 31st of May. According to this approach, by the end of May there would be 247 new cases per day and the total cases are estimated to be close to 10,000. The

outputs underline that in the total absence of control measures, the virus transmission would tend to be uncontrollable since the number of new cases per day increases importantly.

Table 10. Forecasting of new cases/day using Exponential Smoothing (ETS).

Date	New cases/day (fcst)	Date	New cases/day (fcst)
24/3/20	67	28/4/20	159
25/3/20	69	29/4/20	162
26/3/20	72	30/4/20	164
27/3/20	75	1/5/20	167
28/3/20	77	2/5/20	170
29/3/20	80	3/5/20	172
30/3/20	82	4/5/20	175
31/3/20	85	5/5/20	178
1/4/20	88	6/5/20	180
2/4/20	90	7/5/20	183
3/4/20	93	8/5/20	186
4/4/20	96	9/5/20	188
5/4/20	98	10/5/20	191
6/4/20	101	11/5/20	194
7/4/20	104	12/5/20	196
8/4/20	106	13/5/20	199
9/4/20	109	14/5/20	201
10/4/20	112	15/5/20	204
11/4/20	114	16/5/20	207
12/4/20	117	17/5/20	209
13/4/20	119	18/5/20	212
14/4/20	122	19/5/20	215
15/4/20	125	20/5/20	217
16/4/20	127	21/5/20	220
17/4/20	130	22/5/20	223
18/4/20	133	23/5/20	225
19/4/20	135	24/5/20	228
20/4/20	138	25/5/20	231
21/4/20	141	26/5/20	233
22/4/20	143	27/5/20	236
23/4/20	146	28/5/20	238
24/4/20	149	29/5/20	241
25/4/20	151	30/5/20	244
26/4/20	154	31/5/20	246
27/4/20	157	**Total cases:**	10,800

Figure 15 provides a comprehensive view of the scenario in which no mitigation action was adopted and complies with the aforementioned assumption about the roughly controllable spread of the virus. In this graph, the red line depicts the forecasted new cases in a complete absence control measures according to the aforementioned model and the dashed lines refer to the upper/lower confidence bounds of the forecasted data, while the gray line shows the actual data of new cases for the same period. The positive slope of the red line, prove that quarantine was an efficient control measure because in the absence of control measures the number of new cases per day would be increased, contrarily to the mitigation scenario in which this number actually decreases. In addition, in order to achieve a smooth and clear view of the results, the data for 69 days is formed into 10-day time periods. To that side, Table 11. presents the alternatively formatted outputs and Figure 14 visualizes the results.

Table 11. Forecasting of new cases/time period.

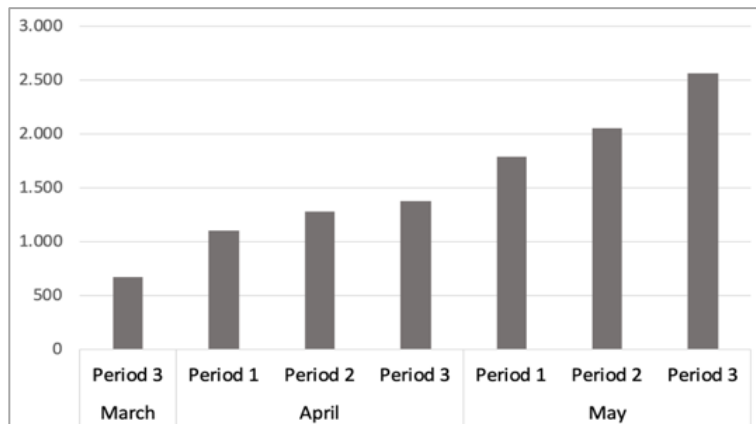| New cases per time period		
March	Period 3	678
	Period 1	1,111
April	Period 2	1,288
	Period 3	1,385
	Period 1	1,790
May	Period 2	2,054
	Period 3	2,565
	Total Cases:	10,800

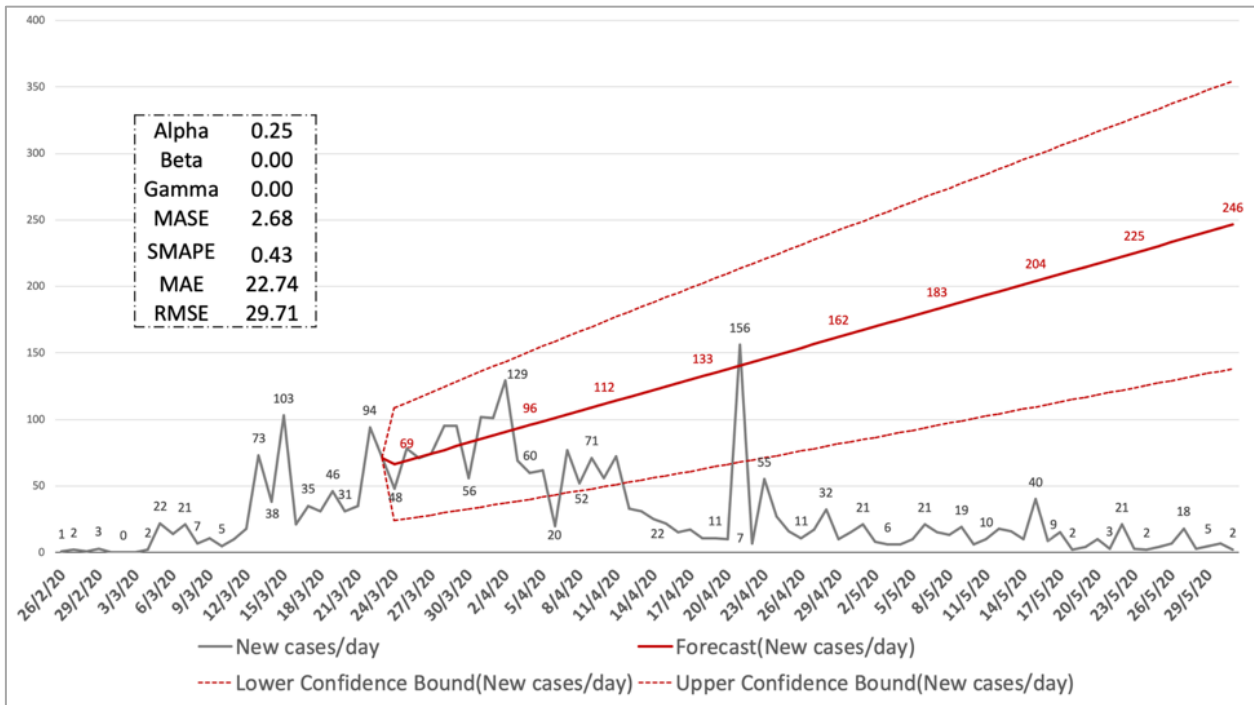Figure 14. Forecasting of new cases/time period.

Figure 15. Graph of real data of new cases/day before and at the quarantine period till the end of May and forecasted new cases/day in a complete absence of control measures.
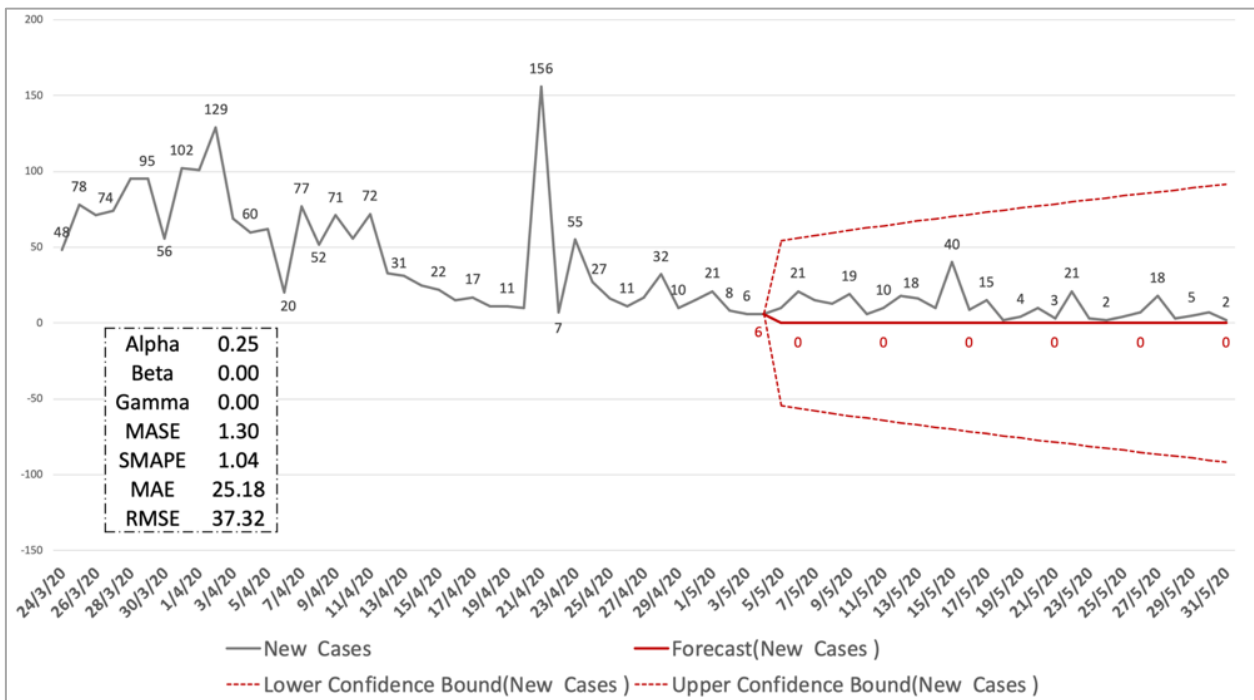

Figure 16. Graph of real data of new cases/day from quarantine period till the end of May and forecasted new cases/day for the prolongation of quarantine period scenario.
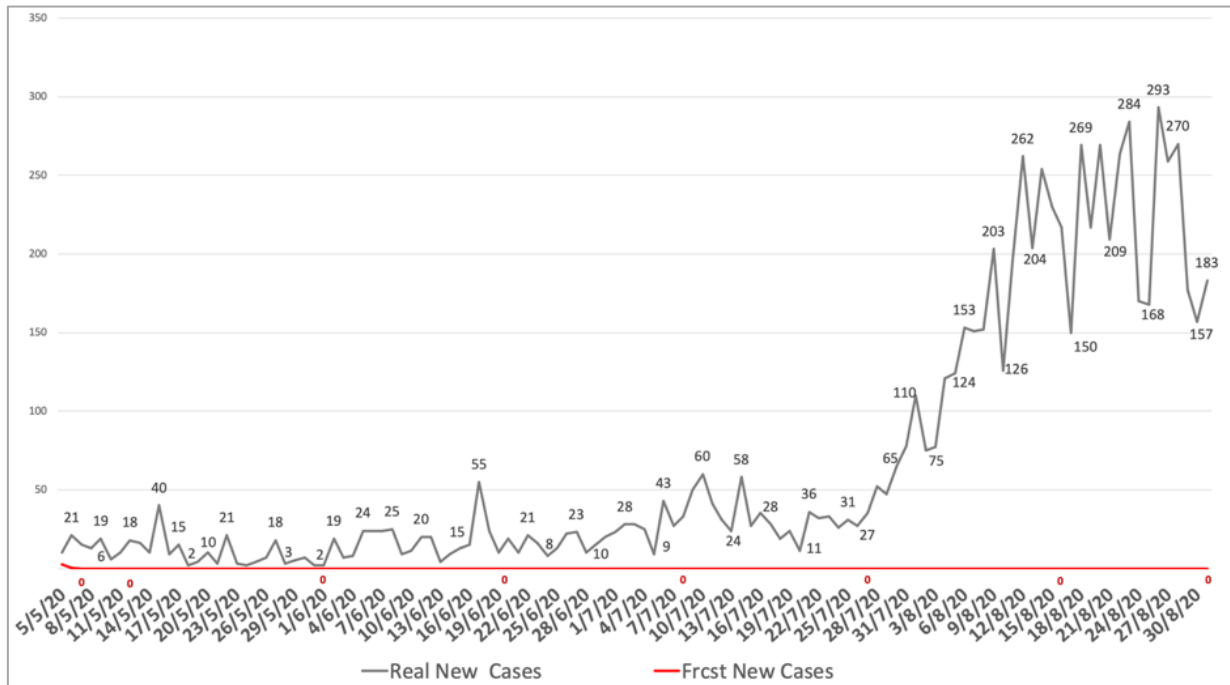
Figure 17. Graph of real data of new cases/day and forecasted new cases/day for the prolongation of quarantine period scenario (May-August).

However, real data of new cases per day for the summer period revealed a different evolution of the virus transmission despite quarantine or smaller mitigation interventions. Figure 16 shows that some days after repealing quarantine (4th of May) the transmission was still under control, but if quarantine was extended until the end of May, based on the forecast model (red line), the transmission would have stopped. Repealing the general lockdown along with the opening of country's borders and allowing the normal operation of stores led to an uncontrolled rise of the transmission, brought new imported cases and transmission chains roughly detected.

More precisely, in Figure 17 the red line refers to new cases per day in extended-lockdown scenario and arises as forecast based on the simple ETS model. The gray line depicts real data of new cases per day for the summer period. As it can be obtained, the positive slope of real data (gray line) along with the negative slope of the red line prove that the prolongation of quarantine would be an efficient choice because the number of new cases/day tends fast to zero, so the transmission of the virus would have been stopped. In addition, we can assume that the aforementioned smaller mitigation actions were not efficient enough so to set under control the transmission of corona virus, especially comparing to a general lockdown which was proven efficient.

4.2.2. Application overview and results discussion.

This application considers the development of a simple mathematical model using branching processes to analyze the initial and latter transmission stages of the covid-19 pandemic outbreak in

Greece. The results depict a reasonably increased R_0 that implies an aggressive spread of the virus. However, the adopted control measures issued by health authorities and adopted by Greek government were proven efficient because the R_0=1.18 was reduced to an R_{eff} = 0.0003. To that side, the additional approach of non-mitigation scenario based on Exponential Smoothing (ETS) model complies with the aforementioned admission. Repealing the movement restrictions and restarting business activity along with opening of Greek borders to tourists, so to support Greek economy, significantly raised the number of reported cases and led to new transmission chains roughly detected.

The lack of available data induced to consider the entire Greece as one cluster which is a slightly unrealistic assumption and the major limitation of the proposed model. However, despite the limitations and the fact this approximation is simple, we conclude that the proposed approach is proven beneficial and meets the initial aims.

Chapter 5: Conclusions and future work.

Branching Processes are individual-based models that consider the growth of populations, whose individuals' reproduction follows mathematical distributions. Thus, the systems that BPs model have a special construction which leads to a specific areas of research, such as biology. The main aim of this thesis was to study in depth this type of stochastic processes with scope to investigate how it can be applied in a wider variety of applications. The preveling idea was to conduct applications on alternative areas, such as reliability of mechanical, industrial or financial related systems, considering the existing literature. To this end, two different applications were examined. The first application confronded with the mathematical modelling through BPs of the reliability of a petrochemical industry plant unit consisting of pumps. However, without overlooking the aforementioned ideas, due to the sudden outbreak of the Covid-19 pandemic the second application refered to the coronavirus (covid-19) spread in Greece. For both applications the relevant literature review was discussed along with the typical branching theory of Galton-Watson processes.

The first application referred to the reliability analysis of a plant unit constructed in sets of main and spare pumps (components). A Markov model was built to process real raw data and set the basis for the branching model. The probability of ultimate extinction that was evaluated both for each type of pumps and the entire unit, the probability of extinction for 12 time units (1 year) and the expected number of working components could be used as reliability measures to predict future potential conditions of the system. The main benefit of this approach is considered that it uses a typical approximation of Markov model but it estimates the reliability and the availability of a mechanical system by using the branching processes. Thus, we can assume that gives a contemporary perspective of the implementation system reliability theory.

Discussing about the second application, that presents the development of a simple mathematical model used to analyze the initial and latter transmission stages of the covid-19 pandemic in Greece, the findings are mainly focused on the outputs of the model, because epidemics consist a typical branching application area. However, the emerging situation that covid-19 caused, raised the interest of scientists as well as of the individuals in general about the epidemic indicators such as the basic or effective reproduction numbers. One of the most crucial findings is the effectiveness of control measures, that Greek government adopted in deferent stages, which was discussed based on the outputs of the branching model along with the additional non-mitigation scenario (simple Exponential Smoothing (ETS) model).

Appling branching processes in biology so to create epidemiology model approximations and examine the spread of diseases is usual and thus it is proven that these processes compose beneficial approaches for that scope. However, based on the system reliability application presented in that thesis, we can assume that branching processes approximations could be effective as a theoretical framework and formulation, because it gives an idea about how to formulate real data and a branching model so to study the reliability of a system, as well as for maintenance departments, because supports maintenance decisions and provides reliable results while minimizes the computational effort comparing to other stochastic methods.

As future work the financial area is highly considered. The implementation of branching processes in financial engineering applications is a complicated task. As in engineering, we firstly have to investigate different systems and study deeply the branching processes area in order to conclude to a specific field of research. Financial engineering area confronts with high risk and uncertainty. Hence, modelling and examination of these types of systems has been of great interest in last decades, mostly because it is a potential way to deal with risk and uncertainty. Nevertheless, the implementation of branching processes in economics and financial fields is, also, poorly documented in the relevant literature. However, there is also a considerable work on apply BP in financial area and economics (Duffie, Filipovic and Schachermayer (2003); Li, Branching Processes and Applications in Macroeconomy (2017)). Stock prices and pricing in markets are important financial areas. Many of the prominent features of stock prices can be represented by a simple branching process (Epps (1996)). In addition, a study, that refers to economic development of product innovations as an evolutionary branching process (Frenken and Boschma (2007)), provides a theoretical framework of how to apply BPs on these areas and explains why is that profitable. The area of product innovations is always up to date and raises the interest of modern industries. Finally, considering the aforementioned brief analysis of the applications on that field, convey the impression that financial and/or financial engineering constitute promising areas of research through branching processes.

References

Ahmed, Ayub, Bashdar Salam, Mahmud Mohammad, Ali Akgul, and Sarbaz H. A. Khoshnaw. (2020). "Analysis coronavirus disease (COVID-19) model using numerical approaches and logistic model." *AIMS Bioengineering* 7 (3): 130-146.

Al-Khasawneh, Mohanad F. (2010). "Estimating the negative binomial dispersion parameter." *Asian Journal of Mathematics and Statistics* 1-15.

Allen, L., S. Jang, and L. Roeger. (2017). "Predicting population extinction or disease outbreaks with stochastic models." *Letters in Biomathematics* 4 (1): 1-22.

Allen, Linda J.S. (2015). *Stochastic population and epidemic models: persistence and extinction.* Switzerland: Springer.

Almudevar, Anthony , David Oakes, and Jack Hall. (2020). *Statistical Modeling for Biological Systems In Memory of Andrei Yakovlev.* Springer.

Asmussen, S., and H. Hering. (1983). *Branching Processes.* New York: Springer Science+Business Media.

Athreya, K.B., and P.E. Ney. (1972). *Branching Processes.* New York: Springer-Verlag Berlin Heidelberg.

Bertozzi, Andrea L., Elisa Franco, George Mohler, Martin B. Short, and Daniel Sledge. (2020). "The challenges of modeling and forecasting the spread of COVID-19." *PNAS* 117 (29): 16732–16738.

Blumberg, Seth , and James O. Lloyd-Smith. (2013). "Inference of R0 and Transmission Heterogeneity from the Size Distribution of Stuttering Chains." *PLOS Computational Biology.*

Blumberg, Seth , Sebastian Funk, and Juliet R. C. Pulliam. (2014). "Detecting Differential Transmissibilities That Affect the Size of Self-Limited Outbreaks." *PLOS Pathogens.*

Cruz, Jo~ao Pires da. (2020). "The data analytics of Covid-19 (or any other virus for that purpose)." *closer*, March 13.

David Aldous, Lara Dolecek. 2003. "Lecture 2: Branching Processes." In *Stat206: Random Graphs and Complex Networks*.

Dawid, R., D. McMillan, and M. Revie. (2015). "Review of Markov Models for Maintenance Optimization in the Context of Offshore Wind." *ANNUAL CONFERENCE OF THE PROGNOSTICS AND HEALTH MANAGEMENT SOCIETY 2015.* PHM Society.

De Serres, Gaston , Nigel J. Gay, and Paddy Farrington. (2000). "Epidemiology of Transmissible Diseases after Elimination." *American Journal of Epidemiology* 1039-1048.

Deepak Prabhakar , P., and A. Dharmaraj. (2018). "MODERN PLANT MAINTENANCE AND RELIABILITYMANAGEMENT METHODS − A REVIEW." *International Journal of Mechanical and ProductionEngineering Research and Development (IJMPERD)* 791-802.

Delamater, Paul L., Erica J. Street, Timothy F. Leslie, Tony Yang, and Kathryn H. Jacobsen. (2019). "Complexity of the Basic Reproduction Number (R0)." *Emerging Infectious Diseases Journal.*

Dobrow, RobertP. 2016. "Branching processes." In *Introduction to stochastic processes with R*. New Jersey: WILEY.

Dobson , Ian, Benjamin A. Carreras, and David E. Newman. (2005). "Branching process models for the exponentially increasing portions of cascading failure blackouts." *Proceedings of the 38th Hawaii International Conference on System Sciences.* Hawaii.

Dorman, Karin S., Janet S. Sinsheimer, and Kenneth Lange. (2004). "In the Garden of Branching Processes." *SIAM Review* 202-229.

DUFFIE, D. , D. FILIPOVIC, and W. SCHACHERMAYER. 2003. "Affine Processes andApplications in Finance." In *The Annals of Applied Probability*, 984–1053. Institute of Mathematical Statistics.

Duffie, D., D. Filipovic, and W. Schachermayer. (2003). "Affine Processes andApplications in Finance." *The Annals of Applied Probability, Institute of Mathematical Statistics* 984-1053.

Epps, T.W. (1996). "Stock prices as branching processes." *Journal Communications in Statistics. Stochastic Models* 4 (12): Pages 529-558.

Fazlollahtabar , Hamed , and Seyed Taghi Akhavan Niaki. (2018). "Modified branching process for the reliability analysis of complex systems: Multiple-robot systems." In *Communications in Statistics - Theory and Methods*. Taylor and Francis.

Frenken, Koen, and Ron Boschma. (2007). "A theoretical framework for Evolutionary Economic Geography: Industrial dynamics and urban growth as a branching process." *Journal of Economic Geography.*

Gabe, C. A., L. O. Freire, and D. A. de Andrade. (2019). "MODELING DYNAMIC SCENARIOS FOR SAFETY, RELIABILITY, AVAILABILITY AND MAINTAINABILITY ANALYSIS." *2019 International Nuclear Atlantic Conference - INAC 2019.* Brazil. 5393-5400.

Gonzalez, Miguel, and Ines Ma del Puerto. (2010). "Branching processes: Genealogy and Evolution." *Bolet́ın de Estad́ıstica e Investigaci ́on Operativa* 107-123.

Haccou, P., P. Jagers, and V. Vatutin. (2007). *Branching Processes: variation, growth and extinction of populations.* Cambridge university press.

Harris, Theodore E. (1964). *The theory of branching processes.* USA: Springer-Verlang.

Helmy, Yosra A., Mohamed Fawzy, Ahmed Elaswad, Ahmed Sobieh, Scott P. Kenney, and Awad A. Shehata. (2020). "The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control." *Journal of Clinical Medicine.*

Ivorra, B., M.R. Ferrández, M. Vela-Pérez, and A.M. Ramos. (2020). "Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) considering its particular characteristics. The case of China." *Communications in Nonlinear Science and Numerical Simulation.*

Jacob, Christine. (2010). "Branching Processes: Their Role in Epidemiology." *Int. J. Environ. Res. Public Health* 7: 1186 - 1204.

Jagers, Peter. (1975). *Branching Processes with biological applications.* New York: Wiley.

Kemeny, J.G., and J.L. Snell. (1978). *Mathematical Models in Social Sciences.* Cambrige, Massachusetts, London, England: MIT Press.

Kendall, David G. (1966). "Branching Processes Since 1873." *The Journal of The London Mathematical Society Volume XLI* 385-406.

Kim, J., and I. Dobson. (2011). "Approximating a loading-dependent cascading failure model with a branching process." *IEEE Transactions on Reliability.*

Kimmel, Marek, and David.E. Axelrod. (2002). *Branching Processes in Biology.* New York: Springer.

Kucharski, Adam J , T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, and R. Eggo. (2020). "Early dynamics of transmission and control of COVID-19: a mathematical modelling study." *Lancet Infect Dis*, March 11: 553-558. Accessed (2020). www.thelancet.com/infection.

La Roche-Carrier, Nicolas , Guyh Dituba Ngoma, Yasar Kocaefe, and Fouad Erchiqui. 2019. "Branching Poisson Process Modelling for Reliability Analysis of Repairable Mechanical System."

Lange, Kenneth. (2010). *Applied Probability.* Springer.

Li, Zenghui. 2017. *Branching Processes and Applications in Macroeconomy.* Ontario: Lakehead University.

Macken, C. A., and A. S. Perelson. (1985). "Branching Processes Applied to Cell Surface Aggregation Phenomena." In *Lecture Notes in Biomath*. New York: Springer-Verlang.

National Public Health Organization (NPHO) of Greece. 2020. *Covid-19 gov (NPHO).* Accessed 11 18, 2020. https://covid19.gov.gr/.

Ndaïrou, Faïçal, Iván Area, Juan J. Nieto, and Delfim F.M. Torres. (2020). "Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan." *Chaos, Solitons and Fractals* ((135)).

Obadia, Thomas , Romana Haneef, and Pierre-Yves Boëlle. (2012). "The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks." *BMC Medical Informatics and Decision Making.*

Overton, Christopher E., Helena B. Stage, Shazaad Ahmad, Jacob Curran-Sebastian, Paul Dark, Rajenki Das, Elizabeth Fearon, et al. (2020). "Using statistics and mathematical modelling to understandinfectious disease outbreaks: COVID-19 as an example." *Infectious Disease Modelling* 5.

Pázsit, Imre , and Lénárd Pál . (2008). *Neutron Fluctuations.* New York: Elsevier.

Raghavaiah , N V, and I HariPrasad. (2019). "MAINTENANCE AND RELIABILITY STRATEGY OF MECHANICALEQUIPMENT IN INDUSTRY." *International Research Journal of Engineering and Technology (IRJET).*

Riou, Julien, and Christian L. Althaus. (2020). "Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020." *Euro Surveill.*

Schinazi, Rinaldo B. (1999). *Classical and Spatial Stochastic Processes with apprications to Biologu.* Birkhäuser.

Siegmund, David O. . (2020). "Encyclopaedia Britannica." *Probability theory.* November 12. Accessed January 16, (2021). https://www.britannica.com/science/probability-theory.

Tugwell, Paul , and Sotiris Nikas. (2020). *Bloomberg.* April 17. Accessed November 18, 2020. https://www.bloomberg.com/news/articles/2020-04-17/humbled-greeks-show-the-world-how-to-handle-the-virus-outbreak.

Wang, Jin . (2020). "Mathematical models for COVID-19: applications, limitations, and potentials." *J Public Health Emerg.* 4.

Waxman, David , and Pierre Nouvellet. (2019). "Supplementary Information Sub- or supercritical transmissibilities: Symmetry in outbreak properties of a disease when conditioned on extinction." *Journal of Theoretical Biology* 80-86.

White, L. Forsberg, and M. Pagano. (2008). "A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic." *Stat Med.*

World Health Organization. (2003). "Consensus document on the epidemiology of severe acute respiratory syndrome (SARS)." *World Health Organization - DEPARTMENT OF COMMUNICABLE DISEASE SURVEILLANCE AND RESPONSE.* May 16-17. Accessed May 25, 2020. https://www.who.int/csr/sars/en/WHOconsensus.pdf.

Yan, Ping, and Gerardo Chowell. (2019). *Quantitative Methods for Investigating Infectious Disease Outbreaks.* Switzerland: Springer.

Yang, Zifeng, Zhiqi Zeng, Ke Wang, Sook-San Wong, Wenhua Liang, Mark Zanin, Peng Liu, et al. (2020). "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions." *Journal of Thoracic Disease* 12 (3): 165-174.

Ye, Y, I.E Grossmann, J. M. Pinto, and S. Ramaswamy. (2020). "MINLP Model for Reliability Optimization of System Design and Maintenance Based on Markov Chain Representation." *E 30th European Symposium on Computer Aided Process Engineering (ESCAPE30).* Milano: Elsevier.

Zhang, Yunjun , Yuying Li, Lu Wang, Mingyuan Li, and Xiaohua Zhou. (2020). "Evaluating Transmission Heterogeneity and Super-Spreading Event of COVID-19 in a Metropolis of China." *International Journal of Enviromental Research and Public Health.*

Zhang, Yunlong, Zhirui Ye, and Dominique Lord. (2019). *Estimating dispersion parameter of negative binomial distribution for analysis of crash data: Boostrapped method maximum likelihood method.* http://www.paper.edu.cn.