Pontifícia Universidade Católica
do Rio de Janeiro

João Pedro Xavier Freitas

Branching processes for epidemics' study

Dissertação de Mestrado

Dissertation presented to the Programa de Pós–graduação em Mecânica Aplicada, do Departamento de Engenharia Mecânica da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Mecânica Aplicada.

Advisor : Profa. Roberta de Queiroz Lima
Co-advisor: Prof. Rubens Sampaio Filho

Rio de Janeiro
September 2023

Pontifícia Universidade Católica
do Rio de Janeiro

João Pedro Xavier Freitas

Branching processes for epidemics' study

Dissertation presented to the Programa de Pós–graduação em Mecânica Aplicada da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Mecânica Aplicada. Approved by the Examination Committee:

Profa. Roberta de Queiroz Lima
Advisor
Departamento de Engenharia Mecânica – PUC-Rio

Prof. Rubens Sampaio Filho
Co-advisor
Departamento de Engenharia Mecânica – PUC-Rio

Prof. José Eduardo Souza de Cursi
INSA - Rouen

Prof. Marcelo Areias Trindade
USP - São Carlos

Rio de Janeiro, September the 29th, 2023

João Pedro Xavier Freitas

Graduated in Civil Engineering by Universidade Federal de Goiás (UFG)

Bibliographic data

Freitas, João Pedro Xavier

Branching processes for epidemics' study / João Pedro Xavier Freitas; advisor: Roberta de Queiroz Lima; co-advisor: Rubens Sampaio Filho. – 2023.
89 f. : il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Mecânica, 2023.
Inclui bibliografia

1. Engenharia Mecânica – Teses. 2. Epidemias. 3. Processos de ramificação. 4. Quantificação de incertezas. 5. Inferência bayesiana. I. Lima, Roberta de Queiroz. II. Sampaio Filho, Rubens. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Mecânica. IV. Título.

CDD: 621

Acknowledgments

I would like to first thank my advisor Roberta, who since the beginning of my master's program was already guiding me through this process. She gave me all the support and motivation necessary to study such an intriguing and fascinating subject that is probability. Then, I would like to thank my co-advisor Rubens for his expertise, suggestions and advice.

Then I wish to thank my family, which is fundamental to me. I appreciate all the unconditional love, support and kindness from my mother Carmen and my father Epitácio. They always told me the importance of studying and gave me the best education. I could not be definitely here without them. They are everything to me.

This paragraph is dedicated to all my friends. First, I would like to thank my friends from Goiânia, where I was born and lived for long years. They are the ones I have shared great part of my joy, laughs and incredible moments. In specifically, I mention two important groups from there: Chimper and my colleagues from university. Last but not least, I would also like to thank Victor, who is my best friend from Rio de Janeiro; Danilo, who is an old friend I had the incredible chance to study again together here in PUC-Rio and Hector Eduardo for all his advice, support and company.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. It is also important for me to thank FAPERJ and PUC-Rio for their financial support during my master program.

Abstract

Freitas, João Pedro Xavier; Lima, Roberta de Queiroz (Advisor); Sampaio Filho, Rubens (Co-Advisor). **Branching processes for epidemics' study**. Rio de Janeiro, 2023. 89p. Dissertação de Mestrado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

This work models an epidemic's spreading over time with a stochastic approach. The number of infections per infector is modeled as a discrete random variable, named here as contagion. Therefore, the evolution of the disease over time is a stochastic process. More specifically, this propagation is modeled as the Bienaymé-Galton-Watson process, one kind of branching process with discrete parameter. In this process, for a given time, the number of infected members, i.e. a generation of infected members, is a random variable. In the first part of this dissertation, given that the mass function of the contagion's random variable is known, four methodologies to find the mass function of the generations of the stochastic process are compared. The methodologies are: probability generating functions with and without polynomial identities, Markov chain and Monte Carlo simulations. The first and the third methodologies provide analytical expressions relating the contagion random variable and the generation's size random variable. These analytical expressions are used in the second part of this dissertation, where a classical inverse problem of bayesian parametric inference is studied. With the help of Bayes' rule, parameters of the contagion random variable are inferred from realizations of the stochastic process. The analytical expressions obtained in the first part of the work are used to build appropriate likelihood functions. In order to solve the inverse problem, two different ways of using data from the Bienaymé-Galton-Watson process are developed and compared: when data are realizations of a single generation of the branching process and when data is just one realization of the branching process observed over a certain number of generations. The criteria used in this work to stop the update process in the bayesian parametric inference uses the L_2-Wasserstein distance, which is a metric based on optimal mass transference. All numerical and symbolical routines developed to this work are written in MATLAB.

Keywords

Epidemics; Branching processes; Uncertainty quantification; Bayesian inference.

Resumo

Freitas, João Pedro Xavier; Lima, Roberta de Queiroz; Sampaio Filho, Rubens. **Processos de ramificação para o estudo de epidemias**. Rio de Janeiro, 2023. 89p. Dissertação de Mestrado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

Este trabalho modela a evolução temporal de uma epidemia com uma abordagem estocástica. O número de novas infecções por infectado é modelado como uma variável aleatória discreta, chamada aqui de contágio. Logo, a evolução temporal da doença é um processo estocástico. Mais especificamente, a propagação é dada pelo modelo de Bienaymé-Galton-Watson, um tipo de processo de ramificação de parâmetro discreto. Neste processo, para um determinado instante, o número de membros infectados, ou seja, a geração de membros infectados é uma variável aleatória. Na primeira parte da dissertação, dado que o modelo probabilístico do contágio é conhecido, quatro metodologias utilizadas para obter as funções de massa das gerações do processo estocástico são comparadas. As metodologias são: funções geradoras de probabilidade com e sem identidades polinomiais, cadeia de Markov e simulações de Monte Carlo. A primeira e terceira metodologias fornecem expressões analíticas relacionando a variável aleatória de contágio com a variável aleatória do tamanho de uma geração. Essas expressões analíticas são utilizadas na segunda parte desta dissertação, na qual o problema clássico de inferência paramétrica bayesiana é estudado. Com a ajuda do teorema de Bayes, parâmetros da variável aleatória de contágio são inferidos a partir de realizações do processo de ramificação. As expressões analíticas obtidas na primeira parte do trabalho são usadas para construir funções de verossimilhança apropriadas. Para resolver o problema inverso, duas maneiras diferentes de se usar dados provindos do processo de Bienaymé-Galton-Watson são desenvolvidas e comparadas: quando dados são realizações de uma única geração do processo de ramificação ou quando os dados são uma única realização do processo de ramificação observada ao longo de uma quantidade de gerações. O critério abordado neste trabalho para encerrar o processo de atualização na inferência paramétrica usa a distância de L_2-Wasserstein, que é uma métrica baseada no transporte ótimo de massa. Todas as rotinas numéricas e simbólicas desenvolvidas neste trabalho são escritas em MATLAB.

Palavras-chave

Epidemias; Processos de ramificação; Quantificação de incertezas; Inferência bayesiana.

Table of contents

List of figures

List of tables

List of Abreviations

cdf – Cumulative distribution function

i.i.d. – Independent and identically distributed

pdf – Probability density function

pgf – Probability generating function

pmf – Probability mass function

BGW – Bienaymé-Galton-Watson

IBM – Individual-based Model

ID – Probability generating functions with polynomial identities

MC – Markov chain

MCS – Monte Carlo simulation

PGF – Probability generating functions without polynomial identities

SAR – Secondary attack rate

SI – Susceptible-infectious

SIR – Susceptible-infectious-recovered

1
Introduction

1.1
Reflection about modeling epidemics

Modeling the spread of a disease over time is a challenging task. The transmissibility of the epidemic [1], i.e., how easy the disease will spread from an infected person (infector) to susceptible ones (infectees) is affected by several parameters, such as the individual behavior and the genomic structures of the pathogen. These parameters are actually hardly traceable and therefore a stochastic interpretation is more appropriate.

However, the feature of transmissibility is usually assessed by deterministic assignments and the most common ones are the parameter of the basic reproduction number R_0 and the secondary attack rate (SAR). The first one represents the average number of individuals infected caused by a single infector at the start of an epidemic. The other is defined as a proportion among the individuals infected and the infectees, also when the epidemic takes off. Both parameters have reported estimates that are heterogeneous even for the same virus. For instance, influenza virus may have R_0 around 1 or above 10. Its counterpart SAR lies between 1% and 38%.

Several different models in the literature are proposed for the dynamics of an infectious disease [2]. The compartmental ones, such as the susceptible-infectious (SI model) and the susceptible-infectious-recovered (SIR model) are categorized as population-level. In these models, the population is divided in groups (named compartments) and an initial value problem characterizes the evolution of the number of individuals in each of the groups [3–5]. For example, in the SIR model the basic reproduction number is fundamental to assign parameters related to the group transitions.

The main problem of these deterministic models is the prediction accuracy and it is a fundamental point for governments to take actions. Accurate predictions of the evolution of the number of infected individuals over time can help, for example, in the organization of hospital supplies. This is one of the reasons that nowadays the individual-level models are getting more attention. They focus on stochastic outbreaks. Transmissibility is then treated in a

stochastic perspective and a more complete prediction is made.

1.2 Introducing branching processes

From the exposed before, it is more reasonable to assign a discrete random variable to the number of infections per infector. In this work, this random variable is named contagion, represented by C. The evolution of the epidemic over time is therefore a discrete state stochastic process and in this work the traditional (discrete state) branching process in discrete time is chosen to model it [6, 7]. Each random variable X_t attached to the branching process represents in this context a generation's size of members infected.

This stochastic process, also named Bienaymé-Galton-Watson (BGW), deals with the so-called demographic stochasticity [8], i.e., the population randomness over time is a consequence of the individual's uncertainty. Thus, it is an individual-based model (IBM) that entails the stochastic features of transmissiblity.

According to [9], the original motivation for this model was the attempt to explain the reason why population of countries were growing exponentially, whilst family names were disappearing. Nowadays, as exposed in [10], a variety of fields works with branching processes: cell biology, population demography, biochemical processes, genetics, epidemiology and actuarial sciences. Propagation phenomena such as those follow some underlying principles. They usually have four main stages [11]: spark, growth, peak and decline.

There are other types of branching process. For instance, the ones continuous in time [12, 13], the ones continuous in state, such as the Feller type, spatial branching process [14], which combines the branching phenomenon with a spatial motion and other superprocesses. Recent studies in the field also focus on the evolution of genealogies and histories and in a more generalized branching property [15].

1.3 Importance of probability mass functions

Once we are using the BGW process, probabilistic descriptions of the size of subsequent generations of infected people are tackled. Therefore, it is meaningful to quantify the uncertainty in the process. Figure 1.1 shows an example of the evolution of mass functions along generations. Since some popular set of statistics, like the mean and standard deviation, the mean and coefficient of variation or the Shannon entropy, are not suitable as a proper measure for uncertainty [16], the cumulative distribution function (cdf) is the

best option. Since the BGW process is discrete in state, finding mass functions (pmf) is another alternative.

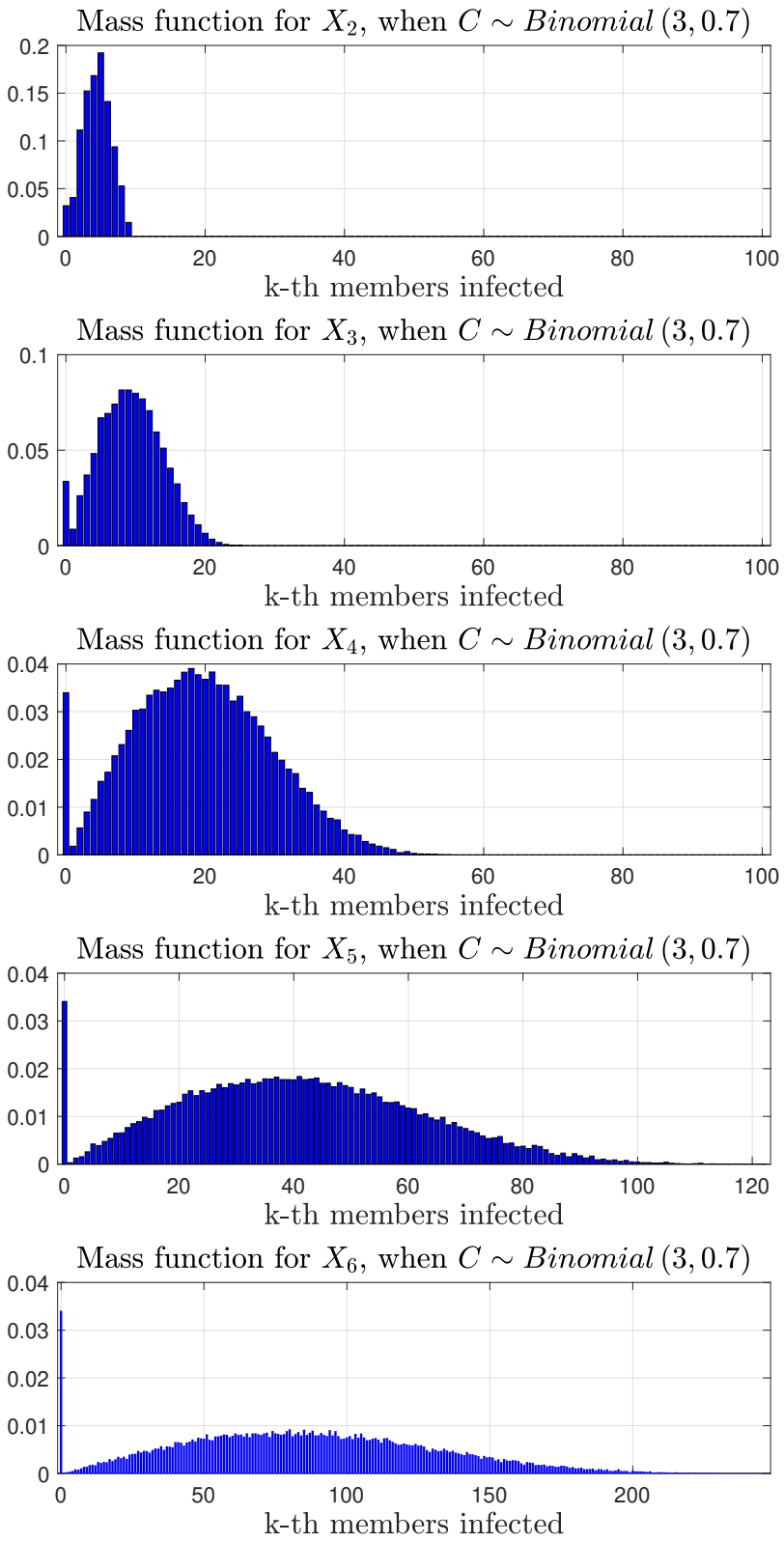


Figure 1.1: Evolution of generation's size mass function in BGW for an example studied and later on presented.

The mass function allows to represent the degree of knowledge of the possible underlying values of a random entity. The latter entails all the probabilistic information and hence finding it is the better option to characterize the problem of the random object regardless the context [17]. The pmfs in this context allow us also to understand the stochastic evolution of the disease over time, evaluate statistics, check the influence of the contagion's random variable and obtain the extinction probability of the disease. Moreover, they provide useful mechanisms to work with bayesian approaches.

This last benefit is a key factor to study this particular problem into two different perspectives, which is indeed our aim: the deductive and the plausible one. In this work, for the former reasoning, the main question is how to find the mass functions of further generations when the contagion's probabilistic model is given. For the latter, we would like to know what we can say about the contagion's probabilistic model when the epidemic is observed.

1.4 Reasoning point of view

The deductive logic is a type of reasoning used to derive effects or outcomes from a knowing cause. This sort of rationale is found in pure mathematics and it is the most desirable one, since it entails for example two strong syllogisms presented in Eq. (1-1) from the implication $A \Rightarrow B$. A and B are in here propositions and the implication $A \Rightarrow B \equiv A = AB$ from the Boolean logic. The function AB is a conjuction or logical product.

$$\begin{cases} \text{if A is } \textbf{TRUE}\text{, then B is } \textbf{TRUE} \\ \text{if B is } \textbf{FALSE}\text{, then A is } \textbf{FALSE} \end{cases} . \tag{1-1}$$

In Eq. (1-1), there are two certainties presented. On the other hand, this situation is not what is usually faced for most scientists. The opposite scenario is indeed the recurrent one: given effects or observations, how can they give a certainty truth to a theory? This question is indeed an open-ended one and it seems that there is no solution. Therefore, the logical nature of scientific reasoning [18] belongs to the plausible reasoning [19], also named inductive logic. Based on the same implication, Eq. (1-2) shows weaker syllogisms this time. The weaker word in this context is used to consider the fact that an extensional logic is in here included.

$$\begin{cases} \text{if B is } \textbf{TRUE}\text{, then A is more plausible} \\ \text{if A is } \textbf{FALSE}\text{, then B is less plausible} \end{cases} . \tag{1-2}$$

The former case in Eq. (1-2) indicates directly that the occurrence of the

logical consequence B reflects a belief that statement A might be true, but there is no way to ensure that. The latter one indicates basically the same thing: A is just one logical cause for statement B. As a consequence, if A is false, the degree of belief in the veracity of B tends to decrease.

The main point of it is that, in plausible or inductive logic, beliefs are somehow measured. Another fundamental aspect is that for instance, based on historical events, for many times that B was true, the proposition A was also true. This way, it is reasonable that the plausibility in A gets even greater in the first syllogism of Eq. (1-2). Hence, the *priori* knowledge has an effect on the degree of beliefs. More than that, in light of new information, the degree of beliefs can still be modified.

These considerations seem to be strictly qualitative to be even quantified and remain purely as common sense, but they are not. They actually follow some quantitative rules: the basic algebra of probability theory in Eq. (1-3)

$$\begin{cases} \mathbb{P}(A) + \mathbb{P}\left(\bar{A}\right) = 1 \\ \mathbb{P}(A, B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B) \end{cases}, \tag{1-3}$$

in which $\bar{A}$ is the negation of the proposition A. The propositions A and B can be interpreted as events in the probability space defined by the triple $(\Omega, \mathcal{F}, \mathbb{P})$, in which Ω is the sample space, $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P}$ is a probability measure on $\mathcal{F}$. The sample space Ω in here is a single extension of all logical truths, and ϕ of all contradictions. The set-operations of $\mathcal{F}$ are equally defined in the Boolean logic: complementation as the negation, closed under union as disjunction, and closed under intersection as conjunction. Therefore, in a probabilistic perspective, the first case in Eq. (1-3) is the normalization condition and the second is the conditional probability.

Both the probabilities $\mathbb{P}(A, B)$ and $\mathbb{P}(B, A)$ represent the same. Then, from the manipulation of second case in Eq. (1-3),

$$\begin{aligned} \mathbb{P}(A, B) &= \mathbb{P}(B, A) \\ \mathbb{P}(A \mid B)\,\mathbb{P}(B) &= \mathbb{P}(B \mid A)\,\mathbb{P}(A) \\ \mathbb{P}(A \mid B) &= \frac{\mathbb{P}(B \mid A)\,\mathbb{P}(A)}{\mathbb{P}(B)}. \end{aligned} \tag{1-4}$$

Eq. (1-4) is known as the Bayes' theorem and it is a mathematical relation with conditional probabilities. The term $\mathbb{P}(B \mid A)$ plays the central role in the transition between the deductive and the plausible reasoning in this work. When the observable entity is related to the event A, we are facing a sampling distribution and it is a statement from the deductive logic. Otherwise, it is a

likelihood function and the inference logic is faced.

The pmfs are indeed sampling distributions. Once they are found, it is possible to build likelihood functions. The opposite could also be done. However, it easier to work in a deductive perspective and that is the reason this work starts at this point.

1.5 Objectives of the dissertation

As said before, this work starts focusing on the deductive reasoning and the central point in here is to answer the question: given the probabilistic model of the contagion's random variable, how can we find the probability mass function for further generations? Four different methodologies are herein presented.

The first is a traditional solution based on probability generating functions (pgfs). There is a novelty in this work at this point that is introducing polynomial identities to the chain rule as a second methodology in order to decrease the runtime spent when working with pgfs. The third is developing the one-step transition matrices for the BGW process when the contagion is modeled as a Binomial family. This is another contribution of this dissertation. Finally, the fourth is a numerical solution based on Monte Carlo simulations. All of them are discussed and compared in the framework of computational costs (runtime and storage), possible applications and individual features.

Now that mass functions for further generations are available, the second part of this work introduces the plausible approach and deals with the opposite scenario: given that a realization of the epidemic modeled as a BGW process happens, what is possible to say about the contagion's random variable? More specific, herein we assume that the contagion's random variable family is known and the aim is to infer its parameters. This is the classical inverse problem of bayesian parametric inference.

There are several new contributions on this perspective. With the help of Bayes' rule, this work introduces ways of building likelihood functions to perform parametric inference when the content of data are realizations of a single generation of the branching process and when data is just one realization of the branching process observed over a certain number of generations. The analysis in here are also focused on the stochastic effect of data sequences.

1.6 Structure of the work

This dissertation is divided according to the following schema: the first chapter, this one, introduces the content of this work. It gives us a glimpse on models for epidemic propagation over time and justifies why a stochastic framework is fundamental to this subject. Moreover, it shows briefly the branching process adopted in here and gives an explanation of the importance of finding mass functions. Finally, it brings the two logical perspectives this work covers and gives its objectives.

The second chapter is responsible to give the mathematical formulation of the BGW process. It also gives to this branching process the proper context in epidemics. This is the chapter related to the model chosen and gives a fundamental information, which is the relation of the contagion random variable with the random variable representing the size of any further generation.

The third chapter focuses on the deductive reasoning. It provides the theory for four methodologies to get mass functions for further generations given the probabilistic model of the contagion: probability generating functions, Markov chain and Monte Carlo simulations. There is also an attempt of improvement in the pgf methodology introducing polynomial identities for the chain rule. Two major comparisons, which are better explained in the chapter are done. One focuses on a local perspective, which is a state by state view and the other on a global framework, which is related to the overall support. All of them rely on computational costs, possible applications and individual features.

The fourth chapter brings the plausible reasoning and studies the bayesian parametric inference problem. First, it shows how to build likelihood functions according to the type of data we are dealing. It also enunciates the L_p-Wasserstein distance, which is an interesting metric on probability spaces used in this work as a convergence criteria. It ends with the analysis of the effects of stochastic data sequences.

The fifth, and last, chapter of this dissertation are the conclusions, which sums up all of the content explored in the previous chapters and brings the main results and understanding of this research.

2
Epidemic propagation modeled as a BGW process

An epidemic takes off with a unique infector, who belongs to the so-called 0th generation of infected members. This person is indicated as individual number 1 in the realization of the BGW process in Figure 2.1. As a consequence of the contact with him/her, another person got sick. This one is individual number 2, who is the only infected member of the 1st generation. This time, individual number 2 spread the disease to two people. They are indicated as individuals number 3 and 4 and form the 2nd generation. Therefore, the size of this generation is two. Following the digraph in Figure 2.1, individual number 3 is responsible for the infections of individuals number 5 and 6, while individual number 4 infected individuals number 7 and 8. Hence, the 3rd generation has size four. The remain network displays the relation of infections up to the 5th generation.

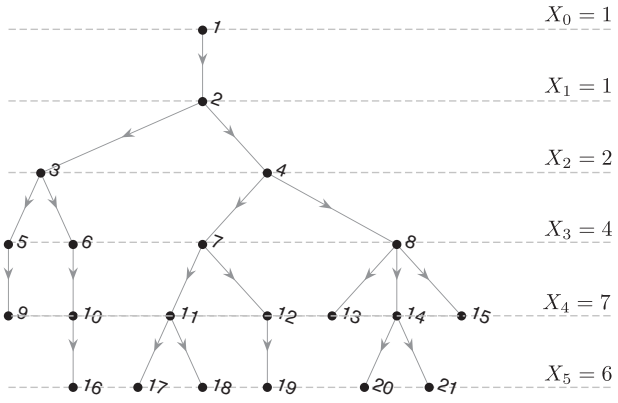


Figure 2.1: Realization of the BGW process up to the fifth generation of infected members.

The directed graph in Figure 2.1 is also known as a ramification tree. Each random variable from the family of random variables of this stochastic process $\mathcal{X} = \{X_t\}$ models the size of the generation of infected members attached to the temporal index t. This is a branching process discrete in time. Hence, $t \in \mathbb{N}^0$, in which $t = 0$ represents when the epidemic takes off with the deterministic statement of a single initial infector, i.e., $X_0 = 1$, and $\mathbb{N}^0$ is the set of all natural numbers including zero. Since it is dealt with population sizes per generation, the state space $\mathbb{S}$ of this branching process is also discrete. At

first, the generation's size can take any non-negative value, then $\mathbb{S} = \mathbb{N}^0$. Later on, when introducing the Markov chain approach, some restrictions on $\mathbb{S}$ are discussed.

The key factor that rules the branching process is the number of successful infections per infector. This aspect is modeled in here also as a discrete random variable that assumes non-negative values. It is named as the contagion random variable C. For instance, in the example of Figure 2.1, for the individual number 6 and number 7 from the 3rd generation, each realization of the contagion resulted in respectively one and two members infected that belongs to the 4th generation. It is important to highlight that in the BGW process, the contagions are independent and identically distributed (i.i.d.) random variables. This means in the current context that the transmissibility does not rely on the population size of the generation nor on the evolution of the disease over time and reflects the demographic stochasticity.

To sum up, the epidemic always begins only with a single infector from the 0th generation. A subsequent size generation X_{t+1} depends on a quantity of realizations of the contagion random variable. This quantity is the population size of the most previous generation X_t. Notice that there is no way to ensure this value at first, unless a realization of X_t is done. Except for the first generation, in which $X_1 = C$, since the previous generation size is unitary from the deterministic statement. Generally speaking, the size X_{t+1} is determined according to Schinazi [20] as

$$X_{t+1} = \sum_{k=0}^{X_t} C,\ t \in \mathbb{N}^0. \tag{2-1}$$

This stochastic process ends when the size of any generation $X_k, k \in \mathbb{N}$ is zero. In this situation, for any further value $t > k$, X_t is also zero, which means the extinction of the disease in the epidemiological context.

3
Mass functions for further generations

3.1
Methodologies to find mass functions for further generations

As explicit in Eq. (2-1), the generation's size X_{t+1} is a sum of a random number related to X_t of i.i.d. contagion random variables C. This chapter focuses on quantifying the uncertainty over the generations of branching processes. Therefore, get the cdf of each random variable from the family $\mathcal{X} = \{X_t\}, t \in \mathbb{N}$ is fundamental. This time, the 0th generation is not included, because there is not any uncertainty inherent in its size. Since this is a discrete state stochastic process, one similar way to achieve this goal is to find instead each pmf for the subsequent generations. In this case, the pmf $\mathbb{P}(X = x)$ is indeed the probability of x, in which $\mathbb{P}$ is the probability measure from the probability space defined by the triple $(\Omega, \mathcal{F}, \mathbb{P})$.

Next, different methodologies to find mass functions for further generations in a BGW process given that the probabilistic model of the contagion C is known are presented. The following approaches do not provide a pmf expressed by a single formula, which is usually found. But they are still able to map the probabilities to each state from the state space $\mathbb{S}$. Another observation is that only further generations, i.e., the ones related to $t \geq 1$ are at this point discussed, because X_1 follows the same law of C.

3.1.1
First methodology: probability generating functions

The probabilities $p_k = \mathbb{P}(X = k)$, $k \in \mathbb{S}$, for a non-negative integers-valued random variable X can be rewritten uniquely in a power series configuration of increasing values of k: $p_0 + p_1 s + p_2 s^2 + \ldots$. One way to generate this sequence is with the help of its probability generating function (pgf). The pgf of a random variable X is the function $G_X(s)$ defined by

$$G_X(s) := \sum_{x=0}^{\infty} s^x \, \mathbb{P}(X = x) = p_0 + p_1 s + p_2 s^2 + \ldots \tag{3-1}$$

The pgf can be related to an expectation value using the law of subconscious statistician, which is enunciated in Theorem 2.29 of Grimmett and

Welsh [21],

$$G_X(s) = \sum_{x=0}^{\infty} s^x \, \mathbb{P}(X = x) = \mathbb{E}\left(s^X\right). \tag{3-2}$$

Two common pgfs are the ones related to the $Binomial(m, p)$ distribution and the $Geometric-0(p)$. The former distribution models in this context a society with strict social distancing rules, in which an infector contacts with m infectees at most and the probability of infection is p for each of them. The latter distribution models a society with weak social distancing rules, in which there is no limit of infectees and the probability of infection is $1 - p$ for each of them. Their pgfs are given as

- $X \sim Binomial(m, p)$:

$$G_X(s) = \sum_{x=0}^{\infty} \binom{m}{x} p^x (1-p)^{m-x} s^x = [(1-p) + p\,s]^m. \tag{3-3}$$

- $X \sim Geometric - 0(p)$:

$$G_X(s) = \sum_{x=0}^{\infty} p(1-p)^x s^x = \frac{p}{1-(1-p)\,s}. \tag{3-4}$$

A useful property coming from the pgfs is the possibility to build a pgf for the sum of a random quantity of i.i.d. random variables, such as the case in Eq. (2-1). As a consequence, it enables to rewrite the probabilities related to X_{t+1} in a sequential way. The key aspect is to describe the pgf $G_{X_{t+1}}(s)$ as a function of $G_C(s)$, which is known, once the contagion's probabilistic model is completely previously defined. This relation is represented in Eq. (3-5).

$$\begin{aligned}
G_{X_{t+1}}(s) &= \sum_{x=0}^{\infty} s^x \, \mathbb{P}(X_{t+1} = x) = \mathbb{E}\left(s^{X_{t+1}}\right) \\
&= \sum_{i=0}^{\infty} \mathbb{E}\left(s^{X_{t+1}} \mid X_t = i\right) \mathbb{P}(X_t = i) \\
&= \sum_{i=0}^{\infty} \mathbb{E}\left(s^{\underbrace{C + C + \ldots + C}_{X_t \text{ times}}}\right) \mathbb{P}(X_t = i) \\
&= \sum_{i=0}^{\infty} \underbrace{G_C(s)}_{\text{argument}}^{\,i} \, \mathbb{P}(X_t = i), \text{ since i.i.d.} \\
&= G_{X_t}(G_C(s)), \text{ as result of the definition in Eq. (3-1)} \\
&= G_C(G_C(\ldots(G_C(s)))), \text{ recurrence happens } t \text{ times.}
\end{aligned} \tag{3-5}$$

Once the pgf of X_{t+1} is described in terms of the pgf of C, it is possible to evaluate the probabilities $\mathbb{P}(X_{t+1} = k), k \in \mathbb{S}$. In order to do that, it is necessary to take the k-th derivative of $G_{X_{t+1}}(s)$, divide it for the factorial of k and then evaluate the analytical expression in $s = 0$. These operations are

summed up in Eq. (3-6)

$$\mathbb{P}\left(X_{t+1}=k\right)=\frac{1}{k!}\left.\frac{d^{(k)}\left[G_C\left(G_C\left(\ldots\left(G_C\left(s\right)\right)\right)\right)\right]}{ds^{(k)}}\right|_{s=0}. \tag{3-6}$$

This approach provides a piecewise function with analytical expressions per generation and state. They map the contagion random variable, through its pgf, with any generation's size probability. There is no need in this methodology to find probabilities of previous generations to get values of the mass function for further ones. This property is so-called time-independency. Moreover, the probabilities of states of the same generation are evaluated individually, which means this methodology is a local one. In terms of computational aspects, an extensive use of symbolic computation is required to find the analytical function multicomposition of the pgfs and their derivatives.

To clarify the use of the pgfs to find probabilities related to further generations, suppose the contagion random variable is modeled as $C \sim Binomial\left(3, 0.5\right)$. The aim is to find the values of the mass function for the 3rd generation's size random variable, for instance the first four ones $k = 0, 1, 2, 3$. Initially, the pgf $G_{X_3}\left(s\right)$ must be rewritten in terms of $G_C\left(s\right)$. According to Eqs. (3-3) and (3-5), the relation between them is

$$G_{X_3}\left(s\right)=\left\{0.5+0.5\left[0.5+0.5\left(0.5+0.5\,s\right)^3\right]^3\right\}^3.$$

The probabilities are then evaluated using Eq. (3-6),

$$\mathbb{P}\left(X_3=0\right)=\frac{1}{0!}\left.\left\{0.5+0.5\left[0.5+0.5\left(0.5+0.5\,s\right)^3\right]^3\right\}^3\right|_{s=0}=0.204$$

$$\mathbb{P}\left(X_3=1\right)=\frac{1}{1!}\left.\frac{d^{(1)}\left\{0.5+0.5\left[0.5+0.5\left(0.5+0.5\,s\right)^3\right]^3\right\}^3}{ds^{(1)}}\right|_{s=0}=0.093$$

$$\mathbb{P}\left(X_3=2\right)=\frac{1}{2!}\left.\frac{d^{(2)}\left\{0.5+0.5\left[0.5+0.5\left(0.5+0.5\,s\right)^3\right]^3\right\}^3}{ds^{(2)}}\right|_{s=0}=0.138$$

$$\mathbb{P}\left(X_3=3\right)=\frac{1}{3!}\left.\frac{d^{(3)}\left\{0.5+0.5\left[0.5+0.5\left(0.5+0.5\,s\right)^3\right]^3\right\}^3}{ds^{(3)}}\right|_{s=0}=0.134$$

Notice that only few values of the support of X_3 were covered. Despite the fact that there is no previous knowledge of the mass function for this generation, the actual support is indeed known. It is $[0, 1, \ldots, 27]$. This is a consequence of the contagion random variable's distribution $C \sim Binomial\left(3, 0.5\right)$.

The values of the mass function for X_3 just evaluated are presented in blue in Fig. 3.1 and the remain ones in gray.

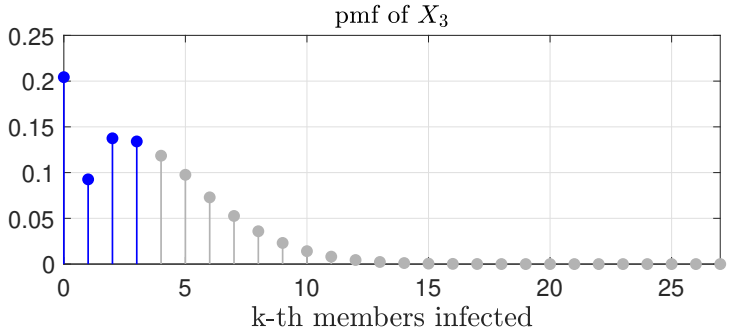


Figure 3.1: Mass function for X_3 when $C \sim Binomial\,(3, 0.5)$. The blue values are the ones evaluated with the pgfs.

3.1.2 Second methodology: polynomial identities for the chain rule

The task of evaluating directly the derivative in Eq. (3-6) may not be feasible. The greater the number of the generation is, the corresponding function of the pgf gets more complex, since the number of recurrences in the function multicomposition increases. Another crucial aspect is that the order of the derivative also increases accordingly to the number of infectors desired. Instead of taking this straight approach, the chain rule could be applied and then the derivatives are done individually for each function which composes the multicomposition pgf. For instance, suppose the 1st and 2nd derivatives of the function composition $\Phi\,(b) = f\,(g\,(b))$ are sought. From the chain rule,

$$\frac{d^{(1)}\left[\Phi\,(b)\right]}{db^{(1)}} = \left.\frac{d^{(1)}\left[f\,(a)\right]}{da^{(1)}}\right|_{a=g(b)} \frac{d^{(1)}\left[g\,(b)\right]}{db^{(1)}}$$

$$\frac{d^{(2)}\left[\Phi\,(b)\right]}{db^{(2)}} = \left.\frac{d^{(1)}\left[f\,(a)\right]}{da^{(1)}}\right|_{a=g(b)} \frac{d^{(2)}\left[g\,(b)\right]}{db^{(2)}} + \left.\frac{d^{(2)}\left[f\,(a)\right]}{da^{(2)}}\right|_{a=g(b)} \left[\frac{d^{(1)}\left[g\,(b)\right]}{db^{(1)}}\right]^2 .$$

Notice that the chain rule gives an identity in which derivatives are done individually for the functions $f\,(a)$ and $g\,(b)$. In order to explain how these identities are generated, some definitions are introduced.

$$\Phi_k := D^{(k)}\Phi\,(b)\,, \quad f_k := \left.D^{(k)} f\,(a)\right|_{a=g(b)}\,, \quad g_k := D^{(k)} g\,(b)\,.$$

The derivatives above are then rewritten in a polynomial structure,

$$\underbrace{\frac{d^{(1)}\left[\Phi\,(b)\right]}{db^{(1)}}}_{\Phi_1} = \underbrace{\left.\frac{d^{(1)}\left[f\,(a)\right]}{da^{(1)}}\right|_{a=g(b)}}_{f_1} \underbrace{\frac{d^{(1)}\left[g\,(b)\right]}{db^{(1)}}}_{g_1}$$

$$\Phi_1 = f_1\, g_1$$

$$\underbrace{\frac{d^{(2)}\left[\Phi\left(b\right)\right]}{db^{(2)}}}_{\Phi_2} = \underbrace{\left.\frac{d^{(1)}\left[f\left(a\right)\right]}{da^{(1)}}\right|_{a=g(b)}}_{f_1} \underbrace{\frac{d^{(2)}\left[g\left(b\right)\right]}{db^{(2)}}}_{g_2} + \underbrace{\left.\frac{d^{(2)}\left[f\left(a\right)\right]}{da^{(2)}}\right|_{a=g(b)}}_{f_2} \left[\underbrace{\frac{d^{(1)}\left[g\left(b\right)\right]}{db^{(1)}}}_{g_1}\right]^2$$

$$\Phi_2 = f_1\, g_2 + f_2\, g_1^2.$$

The k-th derivative of a composition, Φ_k, is then related to a k-th polynomial $\mathcal{B}_k$ composed by the monomials $f_1, f_2, \ldots, f_k, g_1, g_2, \ldots, g_k$:

$$\Phi_k = \mathcal{B}_k\left(f_1, g_1, f_2, g_2, \ldots, f_k, g_k\right) := \mathcal{B}_k\left(\left\{f_k\right\}, \left\{g_k\right\}\right). \tag{3-7}$$

With the help of the Faà di Bruno's formula [22], these polynomials are obtained according to Eq. (3-8):

$$\mathcal{B}_k\left(\left\{f_k\right\}, \left\{g_k\right\}\right) = \sum_{\substack{z=1,\\ r_i:\, i\,\in\,\zeta_z^k}}^{\mathcal{N}(\zeta\vdash k)} \frac{k!}{r_1!r_2!\ldots r_k!} f_r \left[\frac{g_1}{1!}\right]^{r_1} \left[\frac{g_2}{2!}\right]^{r_2} \cdots \left[\frac{g_k}{k!}\right]^{r_k}, \tag{3-8}$$

where $\mathcal{N}\left(\zeta \vdash k\right)$ means the number of partitions $\zeta \vdash k$ of the positive integer k, each $\zeta_z^k, z = 1, 2, \ldots, \mathcal{N}\left(\zeta \vdash k\right)$ is a partition from the integer k, r_i is the number of parts i happening in that partition and finally $r = r_1 + r_2 + \cdots + r_k$ is the sum of all possible numbers of parts of a specific partition. These definitions come from the number theory.

For example, the partitions of the integer 4, $\zeta \vdash 4$, are all the following representations of 4 as a sum of positive integers, in which the order is irrelevant and hence they are written once with non-increasing order of parts:

$$\begin{aligned}
\zeta_1^4 &= 4\\
\zeta_2^4 &= 3 + 1\\
\zeta_3^4 &= 2 + 2\\
\zeta_4^4 &= 2 + 1 + 1\\
\zeta_5^4 &= 1 + 1 + 1 + 1
\end{aligned}$$

The integer 4 has then $\mathcal{N}\left(\zeta \vdash 4\right) = 5$. There is only one part 4 in ζ_1^4 and as a consequence $r_1 = r_2 = r_3 = 0$, $r_4 = 1$ and $r = 1$. The partition ζ_2^4 is composed of a single part 3 and 1 in a way that $r_1 = r_3 = 1$, the others are null and $r = 2$. There are only two parts of the same integer 2 in ζ_3^4 and $r = r_2 = 2$. ζ_4^4 has two parts of the integer 1 and one of 2, then $r_1 = 2$, $r_2 = 1$ and $r = 3$. Finally, ζ_5^4 is composed only of four parts of the same integer, 1, which means that $r = r_1 = 4$.

Eq. (3-8) gives the following polynomial identities:

$$\Phi_1 = f_1\, g_1$$
$$\Phi_2 = f_1\, g_2 + f_2\, g_1^2$$
$$\Phi_3 = f_1\, g_3 + f_2\,(3\, g_2\, g_1) + f_3\, g_1^3$$
$$\Phi_4 = f_1\, g_4 + f_2\left(4\, g_3\, g_1 + 3\, g_2^2\right) + f_3\left(6\, g_2\, g_1^2\right) + f_4\, g_1^4.$$

The Faà di Bruno's formula gives an explicit expression for a specific polynomial. But this is not the only way to get the identities for the chain rule. Another attempt is through a recursive fashion technique that replaces monomials' indexes from previous polynomials generated according to

$$\mathcal{B}_{k+1}\left(f_1, g_1, f_2, g_2, \ldots, f_{k+1}, g_{k+1}\right) = \sum_{j=0}^{k} \binom{k}{j} \mathcal{B}_{k-j}\left(f_2, g_1, f_3, g_2, \ldots, f_{k-j+1}, g_{k-j}\right) g_{j+1}. \tag{3-9}$$

in which $\mathcal{B}_0 = f_1$.

This time Φ_4 is also generated, but in the recursive fashion of Eq. (3-9).

$$\begin{aligned}
\Phi_4 =& \sum_{j=0}^{3} \binom{3}{j} \mathcal{B}_{3-j}\left(f_2, g_1, f_3, g_2, \ldots, f_{3-j+1}, g_{3-j}\right) g_{j+1} \\
=& \binom{3}{0} \mathcal{B}_3\left(f_2, g_1, f_3, g_2, f_4, g_3\right) g_1 + \binom{3}{1} \mathcal{B}_2\left(f_2, g_1, f_3, g_2\right) g_2 + \\
& \binom{3}{2} \mathcal{B}_1\left(f_2, g_1\right) g_3 + \binom{3}{3} \mathcal{B}_0\, g_4 \\
=& \underbrace{f_2\, g_3}_{f_1 \to f_2}\, g_1 + \underbrace{f_3\,(3\, g_2\, g_1)}_{f_2 \to f_3}\, g_1 + \underbrace{f_4\, g_1^3}_{f_3 \to f_4}\, g_1 + 3\, \underbrace{f_2\, g_2}_{f_1 \to f_2}\, g_2 + 3\, \underbrace{f_3\, g_1^2}_{f_2 \to f_3}\, g_2 + \\
& 3\, \underbrace{f_2\, g_1}_{f_1 \to f_2}\, g_3 + f_1\, g_4 \\
=& f_1\, g_4 + f_2\left(4\, g_3\, g_1 + 3\, g_2^2\right) + f_3\left(6\, g_2\, g_1^2\right) + f_4\, g_1^4.
\end{aligned}$$

It is important to highlight that the polynomial identities found are associated to derivatives of a function composition $\Phi(b) = f(g(b))$. This is the case specifically for the 2nd generation, in which $\Phi(s) = G_C(G_C(s))$. For any other function multicomposition, identities for the chain rule can also be generated as in [23], but there is indeed no need of them. Instead of find these other relations, the polynomial identities for a function composition can be applied recurrently. For instance, suppose the 1st derivative of the function multicomposition $\Phi(c) = f(g(h(c)))$ is sought. Firstly, $g(h(c))$ is rewritten as $j(c)$. Then, from the polynomial identities

$$\Phi_1 = \mathcal{B}\left(\{f_1\}, \{j_1\}\right)$$
$$\Phi_1 = f_1\, j_1.$$

But j_1 itself is the first derivative of a function composition $g\left(h\left(c\right)\right)$. Applying again the first polynomial identity,

$$\Phi_1 = f_1\, \mathcal{B}\left(\{g_1\}, \{h_1\}\right)$$
$$\Phi_1 = f_1\, g_1\, h_1,$$

in which

$$f_1 = \left.\frac{d^{(1)}\left[f\left(a\right)\right]}{da^{(1)}}\right|_{a=g(h(c))}, \quad g_1 = \left.\frac{d^{(1)}\left[g\left(b\right)\right]}{db^{(1)}}\right|_{b=h(c)}, \quad h_1 = \frac{d^{(1)}\left[h\left(c\right)\right]}{dc^{(1)}}.$$

Eq. (3-6) can be then rewritten in terms of the polynomial identities for derivatives of composition functions when $k > 0$,

$$\mathbb{P}\left(X_{t+1} = k\right) = \frac{1}{k!}\, \mathcal{B}_k \left(\left\{G_C\left(s\right)_k\right\}, \left\{ \underbrace{G_C\left(G_C\left(\ldots\left(G_C\left(s\right)\right)\right)\right)_k}_{\text{t-1 recurrences}} \right\} \right)\Bigg|_{s=0}. \quad (3\text{-}10)$$

For example, $\mathbb{P}\left(X_3 = 2\right)$ is desired to be known when the contagion random variable is modeled as $C \sim \mathit{Binomial}\left(3, 0.5\right)$. With the help of polynomial identities,

$$\begin{aligned}
\mathbb{P}\left(X_3 = 2\right) &= \frac{1}{2!}\, \mathcal{B}_2 \left(\underbrace{\left\{G_C\left(s\right)_2\right\}}_{\{f_2\}}, \underbrace{\left\{G_C\left(G_C\left(s\right)\right)_2\right\}}_{\{j_2\}} \right)\Bigg|_{s=0} \\
&= \frac{1}{2!} \left(f_1\, j_2 + f_2\, j_1^2 \right)\Big|_{s=0} = \frac{1}{2!} \left(f_1|_{s=0}\ j_2|_{s=0} + f_2|_{s=0}\ {j_1|_{s=0}}^2 \right) \\
&= \frac{1}{2!} \left[f_1|_{s=0}\ \mathcal{B}\left(\underbrace{\left\{G_C\left(s\right)_2\right\}}_{\{g_2\}}, \underbrace{\left\{G_C\left(s\right)_2\right\}}_{\{h_2\}} \right)\Bigg|_{s=0} + f_2|_{s=0} \right. \\
&\quad \left. \mathcal{B}\left(\underbrace{\left\{G_C\left(s\right)_1\right\}}_{\{g_1\}}, \underbrace{\left\{G_C\left(s\right)_1\right\}}_{\{h_1\}} \right)\Bigg|_{s=0} \right] \\
&= \frac{1}{2!} \left[f_1|_{s=0}\ \underbrace{\left(g_1\, h_2 + g_2\, h_1^2\right)\big|_{s=0}}_{2!\,\mathbb{P}(X_2=2)} + f_2|_{s=0}\ \underbrace{{\left(g_1\, h_1\right)|_{s=0}}^2}_{1!\,\mathbb{P}(X_2=1)} \right] \\
&= \frac{1}{2!} \left[D^{(1)}G_C\left(s\right)\Big|_{s=G_C(G_C(0))} \left(D^{(1)}G_C\left(s\right)\Big|_{s=G_C(s)}\ D^{(2)}G_C\left(s\right)\Big|_{s=0} \right.\right. \\
&\quad \left.\left. + D^{(2)}G_C\left(s\right)\Big|_{s=G_C(0)}\ {D^{(1)}G_C\left(s\right)\Big|_{s=0}}^2 \right) + D^{(2)}G_C\left(s\right)\Big|_{s=G_C(G_C(0))} \right. \\
&\quad \left. \left(D^{(1)}G_C\left(s\right)\Big|_{s=G_C(0)}\ D^{(1)}G_C\left(s\right)\Big|_{s=0} \right)^2 \right] \\
&= \frac{1}{2!} \left(0.520 \times 0.475 + 0.884 \times 0.178^2 \right) = 0.138.
\end{aligned}$$

Notice that in order to calculate $\mathbb{P}\left(X_3 = 2\right)$, the probabilities $\mathbb{P}\left(X_2 = 2\right)$

and $\mathbb{P}(X_2 = 1)$ could have been used if their values are beforehand known. Instead of working with this methodology with time-independence, it is chosen to embrace this time-dependence to avoid extra operations. It means that if $\mathbb{P}(X_{t+1} = k), t > 1$ and $k > 1$ is desired to be calculated, the probabilities $\mathbb{P}(X_t = k), k > 1$ will be calculated before. It seems at first hand not a beneficial idea. However, when the aim is to find the whole mass function for subsequent further generations to see how the process develops in a stochastic sense, this option provides advantages in terms of runtime. On the other hand, the analytical expression relating the probability of any state from a further generation and the contagion random variable is partially suppressed. Anyway, this is also a local strategy.

Eqs. (3-8) and (3-9) are responsible to obtain the generalized polynomial identities for the chain rule of a function composition. Once these structures are available, they are saved as a data file. This stage requires an extensive use of symbolic computation. The other stage consists of loading the identities and then performing Eq. (3-10) according to the contagion's pgf.

3.1.3
Third methodology: Markov chain

A stochastic process is a Markov chain if the probability of achieving the following state $i_{t+1} \in \mathbb{S}$ coming from the current state $i_t \in \mathbb{S}$ is independent of its past, i.e.,

$$\begin{aligned}\mathbb{P}(X_{t+1} = i_{t+1} \mid X_0 = i_0, X_1 = i_1, \ldots, X_t = i_t) = \\ \mathbb{P}(X_{t+1} = i_{t+1} \mid X_t = i_t).\end{aligned} \tag{3-11}$$

In other words, given the family of random variables $\mathcal{X} = \{X_t\}_{t \in \mathbb{N}}$ defined on the state space $\mathbb{S}$, it is only necessary to know the current probability distribution, X_t, in order to find the subsequent one, X_{t+1}. For the case of a discrete state Markov chain, such as the BGW process, the mass functions are assessed by two fundamental entities: the mass function for a current distribution λ^{X_t} and the t-th one-step transition matrix $\boldsymbol{T}^{(t)}$. The former is a row vector, whose size is the cardinality of the state space $|\mathbb{S}|$, and the entries are the probabilities of X_t related to each state of $\mathbb{S}$. The latter is a matrix with size $|\mathbb{S}| \times |\mathbb{S}|$, whose elements are known as t-th one-step transition probabilities $p_{i,j}(t) = \mathbb{P}(X_{t+1} = j \mid X_t = i)$. This matrix is called a stochastic matrix since each row follows the normalization condition $\sum_{j \in \mathbb{S}} p_{i,j}(t) = 1$ and it is responsible to link the mass function for the current distribution λ^{X_t} to the next one $\lambda^{X_{t+1}}$, as in Eq. (3-12)

$$\lambda^{X_{t+1}} = \lambda^{X_t} \boldsymbol{T}^{(t)}. \tag{3-12}$$

It is also possible to link the mass function for the subsequent distribution with the first one of the stochastic process. In this context, the pmf $\lambda^{X_{t+1}}$ is the same of λ^{C}, because the size of the first generation is unitary.

$$\lambda^{X_{t+1}} = \lambda^{C}\, \boldsymbol{T}^{(1)}\, \boldsymbol{T}^{(2)} \ldots \boldsymbol{T}^{(t)}. \tag{3-13}$$

The one-step transition matrices from the BGW process are not the same, which means it is a non-homogeneous Markov chain. This is a consequence of the branching process' network evolution. In the epidemic context, the greater the number of the generation is, the number of infectees increases. For instance, the generation $t+1$ has $q_{X_{t+1}}$ infectees and hence there is an one-step transition probability $\mathbb{P}\left(X_{t+1} = q_{X_{t+1}} \mid X_t = j\right) > 0$. On the other hand, the number of susceptible members of the t-th generation is fewer than $q_{X_{t+1}}$, hence $\mathbb{P}\left(X_t = q_{X_{t+1}} \mid X_{t-1} = j\right) = 0, \forall\, j \in \mathbb{S}$.

A meaningful aspect is that the BGW process can have an infinite state space $\mathbb{S}$ if a final index t_f to the family $\mathcal{X} = \{X_t\}_{t\in\mathbb{N}}$ is not set or if the contagion random variable does not have a finite support. In both situations, it is impossible to write completely the one-step transition matrices. Therefore, a final generation to analyze must be imposed previously and the contagion random variable must have finite support. In this case, the support of each random variable from $\mathcal{X} = \{X_t\}_{t\in\mathbb{N}}$ is $\mathbb{1}_{[0,1,\ldots,{q_C}^t]}$, in which q_C is the upper limit of infectees of the contagion random variable. The elements of the t-th one-step transition matrix are defined for the BGW process according to

$$p_{i,j}\,(t) = \begin{cases} 1, \text{ if } \begin{cases} i = j = 0 \\ i > {q_C}^t,\ j = 0 \end{cases} \\ 0, \text{ if } \begin{cases} i = 0,\ j \neq 0 \\ i > {q_C}^t,\ j \neq 0 \\ 0 < i \leq {q_C}^t,\ j > i \times q_C \end{cases} \\ \mathbb{P}\left(\sum_{k=1}^{X_t=i} C = j\right), \text{ otherwise .} \end{cases} \tag{3-14}$$

The last statement of Eq. (3-14) has a sum of a beforehand known deterministic number of times of the contagion random variable. This is not the same as expressed in Eq. (3-5), in which this number is a stochastic object. For now, another relation is established between the pgf of the generation's size random variable attached to index $t+1$ and the contagion one. The connection

is presented next in Eq. (3-15),

$$\begin{aligned} G_{X_{t+1}}(s) &= \sum_{x=0}^{\infty} s^x \, \mathbb{P}(X_{t+1} = x) = \mathbb{E}\left(s^{X_{t+1}}\right) \\ &= \mathbb{E}\left(s^{\underbrace{C+C+\ldots+C}_{i \text{ times}}}\right) = \mathbb{E}\left(s^C s^C \ldots s^C\right) \\ &= \mathbb{E}\left(s^C\right)\mathbb{E}\left(s^C\right)\ldots\mathbb{E}\left(s^C\right), \text{ from independence} \\ &= G_C(s)\, G_C(s) \ldots G_C(s), \text{ by Eq. (3-2)} \\ &= {G_C}^i(s). \end{aligned} \tag{3-15}$$

The last case of the t-th one-step transition probabilities is obtained similarly as in Eq. (3-6). The main difference is that this time the relation between $G_C(s)$ and $G_{X_{t+1}(s)}$ is given by Eq. (3-15) and it is a conditional probability on the state of the current generation's size $X_t = i$.

$$\mathbb{P}\left(\sum_{k=1}^{X_t=i} C = j\right) = \frac{1}{j!} \left.\frac{d^{(j)}\left[{G_C}^i(s)\right]}{ds^{(j)}}\right|_{s=0} \tag{3-16}$$

This methodology is a time-dependent one, because the operation in Eq. (3-13) when dealing with the multiplications coming from the left side results in the sequence of mass functions $\left\{\lambda^{X_t}\right\}_{t>1}$. Each element of this sequence is a row vector covering the whole state space $\mathbb{S}$, so it is not a local approach.

Next, another example is used to illustrate the Markov chain methodology applied into the BGW process. Firstly, the contagion is modeled as $C \sim Binomial(2,\, 0.5)$, so the upper limit number of infectees per generation is $q_{X_t} = 2^t$. The aim is to get the pmf of the third generation's size, hence the state space is $\mathbb{S} = [0, 1, \ldots, 2^3]$. As a consequence, the one-step transition matrices $\boldsymbol{T}^{(1)}$, $\boldsymbol{T}^{(2)}$ have size 9×9 and the initial row vector λ^{X_1} has size 1×9. A few 2nd one-step transition probabilities are calculated in the following. They are related, as an example, to the row $i = 2$, which means here that this row considers the size of the second generation to be beforehand known and to be two. According to Eqs. (3-3) and (3-15), the relation between the pgf of the contagion random variable and the one of the third generation's size given that the second has size two is

$$G_{X_3|X_2=2}(s) = \left[(0.5 + 0.5\, s)^2\right]^2$$

The first five 2nd one-step transition probabilities for this row are

obtained applying Eq. (3-14).

$$p_{2,0}(2) = \frac{1}{0!}\left[(0.5+0.5\,s)^2\right]^2\Big|_{s=0} = 0.063$$

$$p_{2,1}(2) = \frac{1}{1!}\left.\frac{d^{(1)}\left\{\left[(0.5+0.5\,s)^2\right]^2\right\}}{ds^{(1)}}\right|_{s=0} = 0.250$$

$$p_{2,2}(2) = \frac{1}{2!}\left.\frac{d^{(2)}\left\{\left[(0.5+0.5\,s)^2\right]^2\right\}}{ds^{(2)}}\right|_{s=0} = 0.375$$

$$p_{2,3}(2) = \frac{1}{3!}\left.\frac{d^{(3)}\left\{\left[(0.5+0.5\,s)^2\right]^2\right\}}{ds^{(3)}}\right|_{s=0} = 0.250$$

$$p_{2,4}(2) = \frac{1}{4!}\left.\frac{d^{(4)}\left\{\left[(0.5+0.5\,s)^2\right]^2\right\}}{ds^{(4)}}\right|_{s=0} = 0.063.$$

The remain 2nd one-step transition probabilities for $i = 2$ are all zero, since $0 < i \leq {q_C}^t$, $j > i \times q_C$. In other words, it is not possible for $X_3 \geq 5$, once $X_2 = 2$ and each infector in this case is restricted to contact with two infectees. The complete 1st and 2nd one-step transition matrices of this examples are

$$\boldsymbol{T}^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.250 & 0.500 & 0.250 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.063 & 0.250 & 0.375 & 0.250 & 0.063 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{T}^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.250 & 0.500 & 0.250 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.063 & 0.250 & 0.375 & 0.250 & 0.063 & 0 & 0 & 0 & 0 \\ 0.016 & 0.094 & 0.234 & 0.313 & 0.234 & 0.094 & 0.016 & 0 & 0 \\ 0.004 & 0.031 & 0.109 & 0.219 & 0.273 & 0.219 & 0.109 & 0.0313 & 0.004 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The values of the mass function for the 3rd generation's size are given as

$$\lambda^{X_3} = \underbrace{\lambda^C \boldsymbol{T}^{(1)}}_{\lambda^{X_2}} \boldsymbol{T}^{(2)}$$

$$\lambda^{X_3} = \begin{bmatrix} 0.250 & 0.500 & 0.250 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \boldsymbol{T}^{(1)} \boldsymbol{T}^{(2)}$$

$$\lambda^{X_3} = \begin{bmatrix} 0.483 & 0.217 & 0.177 & 0.078 & 0.033 & 0.009 & 0.003 & 4.883\times10^{-4} & 6.104\times10^{-5} \end{bmatrix}.$$

Notice that the values of the mass function for the 2nd generation's size is also found along the operation above. The mass function for this distribution is displayed next in Figure 3.2.

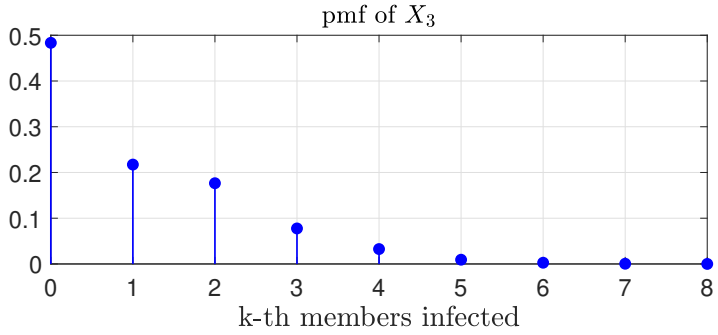


Figure 3.2: Mass function for X_3 when $C \sim Binomial\,(2, 0.5)$.

It is important to highlight that it is not possible to find the values of the mass function for the 4th generation's size for example with the one-step transition matrices and the initial distribution above. In order to do that, they must be resized according to a new state space $\mathbb{S}$ that includes all the possible states of X_4. Moreover, the 3rd one-step transition matrix must also be calculated.

Another remarkable observation coming from the one-step transition matrices of the BGW process is that the furthest matrix has all the elements of the previous ones. This provides a great advantage in terms of computational cost to find all necessary stochastic matrices. It avoids to calculate subsequently all of them. From the last one matrix $\boldsymbol{T}^{(t)}$, its rows ${q_C}^k + 2$ to ${q_C}^t + 1$ should be replaced for row vectors that starts with 1 and is followed by zeros to get $\boldsymbol{T}\,(k)$, $\forall\, 1 \leq k < t$.

3.1.4
Fourth methodology: Monte Carlo simulations

Monte Carlo simulations (MCS) is a numerical methodology based on sampling from random generators to build statistical models. Its way of employment is intrinsically associated to the problem in question. The general steps that are common in the range of all applications are making realizations

from a stochastic object, performing deterministic transformations on each generated experiment, making sample statistics or histograms of them and evaluating its convergence given a tolerance.

The problem in this work consists of finding the values of the mass function for a further generation's size. To tackle it, firstly n_r realizations of the branching process up to a desired generation is done. Each realization is called an experiment and all of them represent a sample. Then, sample statistics are taken from these n_r realizations. They are indeed random objects, since they rely at first on the random sample generator. In order to deal with this uncertainty, a convergence analysis must be made. Then, a tolerance ξ is established. If the sample statistics from the n_r realizations do not satisfy this error, another sample with a greater number of experiments than n_r must be generated and another sample statistics are taken and compared. For this case, the difference between the sample mean $\hat{\mu}_{X_{t+1}}$ and the expectation of the generation's random variable X_{t+1} is compared with the tolerance

$$\hat{\xi}_{t+1} = \left| \mathbb{E}\left(X_{t+1}\right) - \hat{\mu}_{X_{t+1}} \right| < \xi. \tag{3-17}$$

There are some problems in which it is impossible to get the expectation of a specific random variable, but this is not the case $\mathbb{E}\left(X_{t+1}\right)$. In the BGW process, despite the fact that there is no previous knowledge of the probability distribution of any further generation, its expectation can be found as shown in Grimmett and Welsh [21] with the help of the pgfs and Abel's lemma

$$\begin{aligned} \frac{d^{(1)} G_{X_{t+1}}(s)}{ds^{(1)}} &= \frac{d^{(1)}}{ds^{(1)}} \sum_{x=0}^{\infty} s^x \, \mathbb{P}\left(X_{t+1} = x\right) \\ &= \sum_{x=0}^{\infty} \frac{d^{(1)}}{ds^{(1)}} \, s^x \, \mathbb{P}\left(X_{t+1} = x\right) \\ &= \sum_{x=0}^{\infty} x s^{\,x-1} \, \mathbb{P}\left(X_{t+1} = x\right). \end{aligned} \tag{3-18}$$

Taking $s = 1$ in Eq. (3-18), the expectation is found

$$\begin{aligned} \left. \frac{d^{(1)} G_{X_{t+1}}(s)}{ds^{(1)}} \right|_{s=1} &= \sum_{x=0}^{\infty} x \, \mathbb{P}\left(X_{t+1} = x\right) \\ &= \mathbb{E}\left(X_{t+1}\right). \end{aligned} \tag{3-19}$$

The relation between the pgf of a specific generation's size $G_{X_{t+1}}(s)$ and the pgf of the contagion random variable $G_C(s)$ is assessed by Eq. (3-5). This substitution, in which recurrence happens t times, gives

$$\mathbb{E}\left(X_{t+1}\right) = \left.\frac{d^{(1)}\left[G_C\left(G_C\left(\ldots\left(G_C\left(s\right)\right)\right)\right)\right]}{ds^{(1)}}\right|_{s=1}. \tag{3-20}$$

This numerical approach is a time-dependent and not a local methodology. Once the number of experiments n_r is determined, each realization of the branching process gives values of infected members per generation. After that, a normalized histogram of each desired generation is done and the approximation to the mass function along the support is visualized.

An example is explored in the following. The contagion random variable is modeled this time as $C \sim Binomial(3, 0.7)$ distribution and the values of the mass function for the 4th generation's size are sought. Its expectation value is $\mathbb{E}\left(X_4\right) = 19.448$ according to Eq. (3-20). For a tolerance of $\xi = 0.001$, the sample statistic $\hat{\xi}_4$ from 48518 realizations attempts is acceptable. Its convergence analysis is displayed in Fig. 3.3, in which the sample mean is compared with the expectation of the random variable.

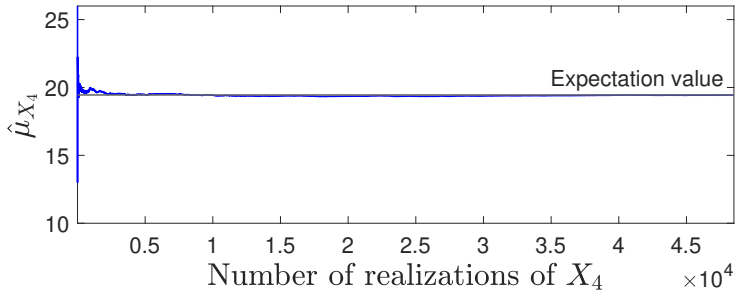


Figure 3.3: Convergence analysis of the number of experiments to approximate the mass function for X_4 when the contagion is modeled as $C \sim Binomial\left(3, 0.7\right)$ and a tolerance $\xi = 0.001$.

Therefore, $n_r = 49000$ are made to build the histogram that approximates to the probability mass function for X_4. Moreover, it is shown in Fig. 3.4 an evolution of histograms from different samples to show the stability of the increasing number of experiments until $n_r = 49000$ reflecting on the histograms' shapes.

Monte Carlo has its foundations in two meaningful theorems: the law of large numbers and the central limit theorem. The former ensures under some circumstances that when increasing the sample size the approximations converge. The latter shows how it converges.

3.1.4.1 The law of large numbers

Suppose a sequence of i.i.d. random variables $Y_1, Y_2, \ldots, Y_n$. Each one of them has the same mean value μ and variance σ^2. Now, another random

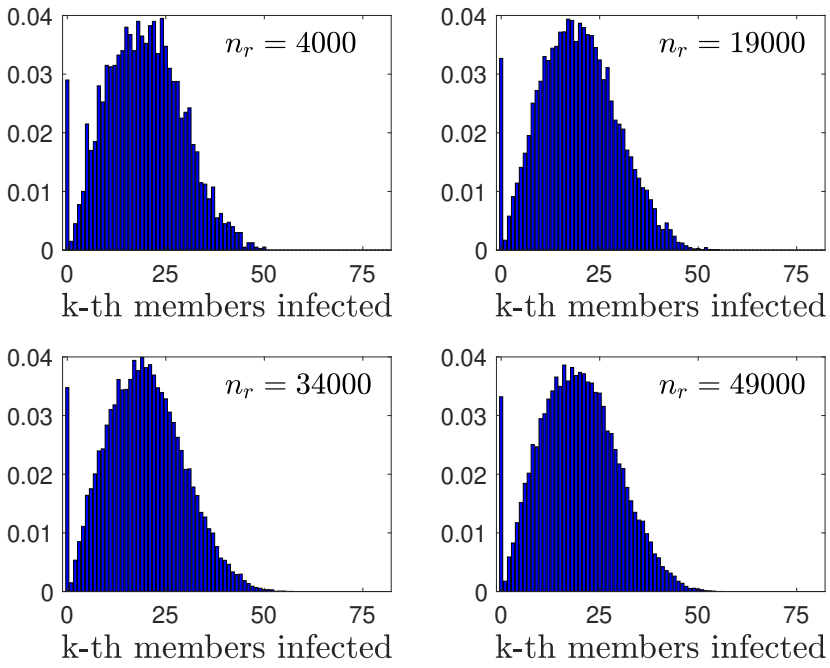


Figure 3.4: Histogram evolution to approximate the mass function for X_4 with different samples when contagion is modeled as $C \sim Binomial\,(3, 0.7)$.

variable is defined as the sum of elements of the previous sequence, $S_n = Y_1 + Y_2 + \ldots + Y_n$. We have that S_n/n converges in mean square to the expectation value μ, $S_n/n \xrightarrow{\text{m.s.}} \mu$, which means

$$\sqrt{\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right]} \to 0, \text{ as } n \to \infty. \tag{3-21}$$

3.1.4.2 The central limit theorem

Again, suppose the same sequence above and its elements' sum S_n. This time, rather than work with S_n, we deal with the standardized version of it,

$$Z_n := \frac{S_n - \mathbb{E}\,(S_n)}{\sqrt{\text{var}\,(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \tag{3-22}$$

We have that Z_n converges in distribution to a random variable, which has normal distribution and its mean is 0 and variance is 1, $Z_n \xrightarrow{\text{d}} \mathcal{N}\,(0, 1)$, which means

$$\mathbb{P}\,(Z_n \leq x) \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du, \forall\, x \in \mathbb{R}, \text{ as } n \to \infty. \tag{3-23}$$

3.2 Comparison among methodologies

In this section, the introduced methodologies are compared in terms of computational cost (runtime and storage), time-dependency and local features, limitations of the contagion's random variable law and other aspects specific from each methodology. It begins by comparing the pgf methodology with and without the help of the polynomial identities of Section 3.1.2. Then, according to the results of this first analysis, the pgf is also compared this time with the Markov chain and Monte Carlo simulations methodologies. All codes were written in MATLAB and they were run on a MacBook Air M2, 16 GB of RAM and 512 GB of storage.

3.2.1 Local comparisons: probability generating functions with and without polynomial identities

Firstly, the two different ways to obtain the general polynomial identities in Section 3.1.2 are analyzed. Figure 3.5 shows the comparison in terms of runtime (CPU time) between them.

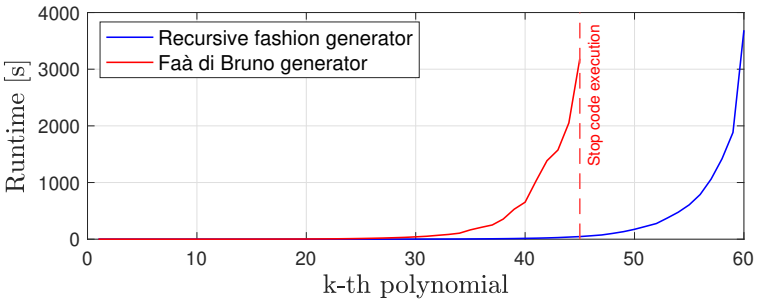


Figure 3.5: Runtime spent to generate each polynomial identity through Faà di Bruno's formula and recursive fashion.

Both approaches lead to almost the same runtime to generate the general polynomial identities up to the case that covers later 30 infections in this work's context. From this number, the recursive fashion generates the next ones identities faster. The execution of the Faà di Bruno's generator was interrupted when the 45th polynomial was obtained, once its runtimes were already longer in comparison. Despite of the runtime benefit of the recursive fashion generator, it was not feasible to find more polynomials than the 60th, because it faced indeed a RAM issue. In order to get the 60th identity, the machine required 11.15 GB of RAM. This is a consequence of the cumulative extensive use of symbolic computation to get each next polynomial. The total runtime spent with the recursive fashion way to get all the 60 polynomials was 11559.943 s.

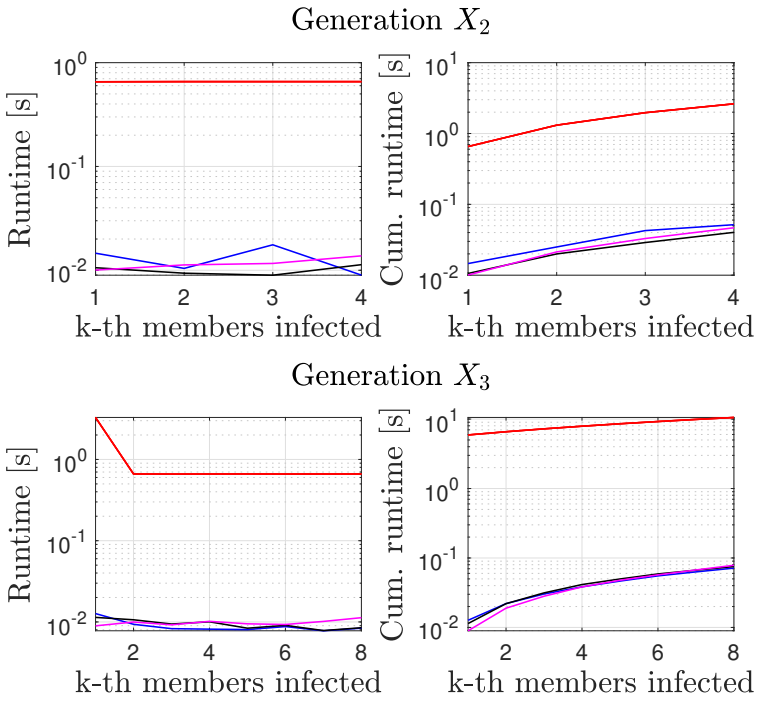
After that, the polynomial identities are saved as MATLAB data files. Finally, they are loaded and the values of the mass functions are obtained according to Eq. (3-10) given the desired probability distribution of the contagion random variable. The step of loading them required 2060.814 s.

The comparison between the pgf with and without the polynomial identities is presented next. Two discrete probability distributions families to model the contagion random variable are analyzed: the Binomial and the Geometric-0. For each case, a parametric study is also presented. It is important to highlight that the runtime spent to generate the polynomial identities and the loading is not being included in the following analysis.

3.2.1.1 Contagion modeled as a Binomial distribution

For the case of $C \sim Binomial\,(m, p)$, the former parameter changes in the range $m = [2, 3]$, while the latter parameter in $p = [0.3, 0.5, 0.7]$. The 2nd generation up to the 6th one are focused in here.

Figure 3.6 shows the runtime and the cumulative runtime to find the values of the mass function for these generations for the $Binomial\,(2, p)$ family of random variables of the contagion. It is noticed that the pgf methodology without the polynomial identities finishes faster the simulation. But, the greater the number of the generation is, this difference decreases. According to the value of the parameter p, the runtime's curves visually vary for the pgf itself, but not with the use of the polynomial identities.



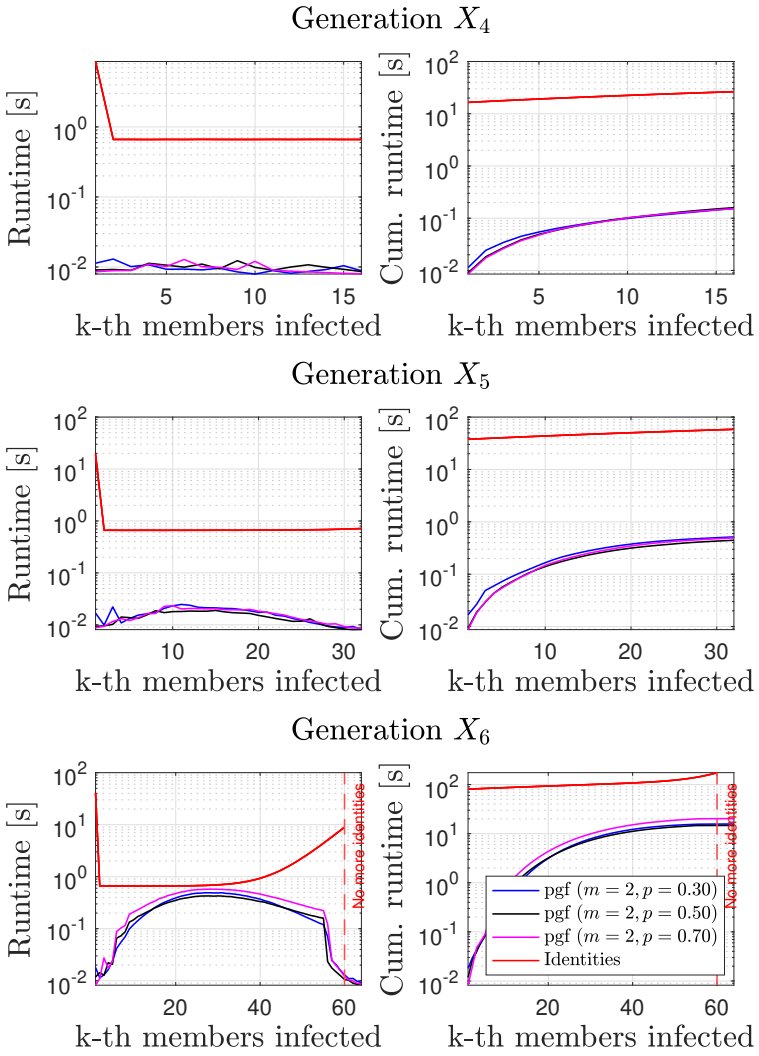


Figure 3.6: Runtime and cumulative runtime for the case of $C \sim Binomial\,(2, p)$ when comparing the pgf itself and with the help of polynomial identities.

The notable peak displayed in the first point of the relation runtime versus number of members infected for the identities approach from the 3rd generation on reflects the time-dependency of this methodology. It is not possible to get the values of the mass function for the 3rd generation without evaluating the ones from the 2nd one for example. Moreover, the number of polynomial identities available cannot cover the whole support of the 6th generation' size. In this situation, the probabilistic information missed is minimum as visualized in the pmf and cdf in Figure 3.7 for $C \sim Binomial\,(2, 0.7)$, which is the worst scenario.

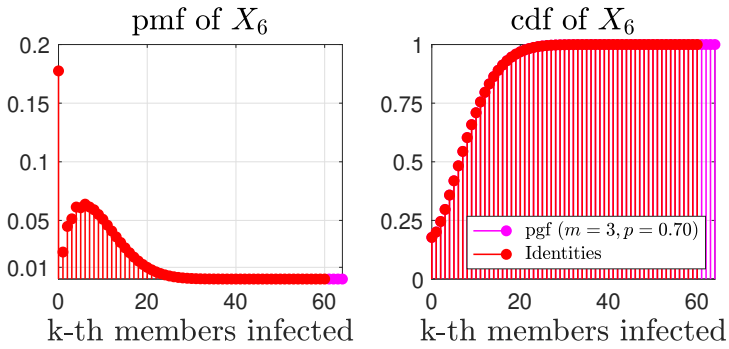
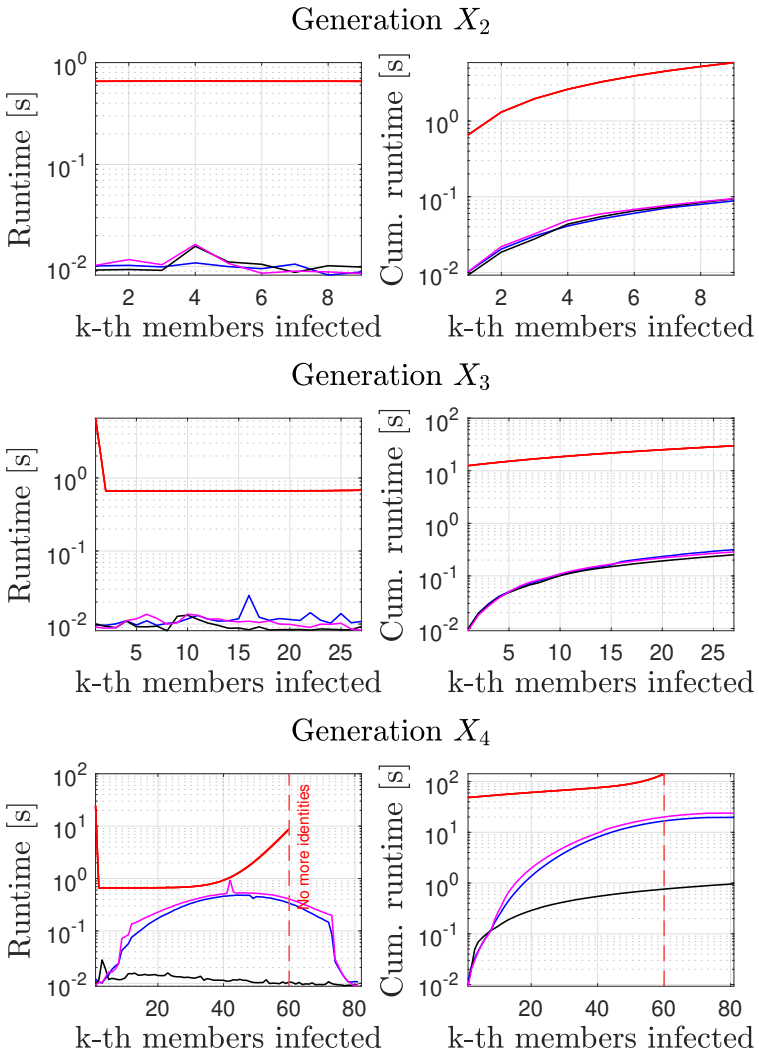


Figure 3.7: Pmf and cdf of X_6 for $C \sim Binomial\,(2, 0.7)$

Figure 3.8 shows this time the runtimes and cumulative runtimes for the BGW process with $C \sim Binomial\,(2, 0.7)$.



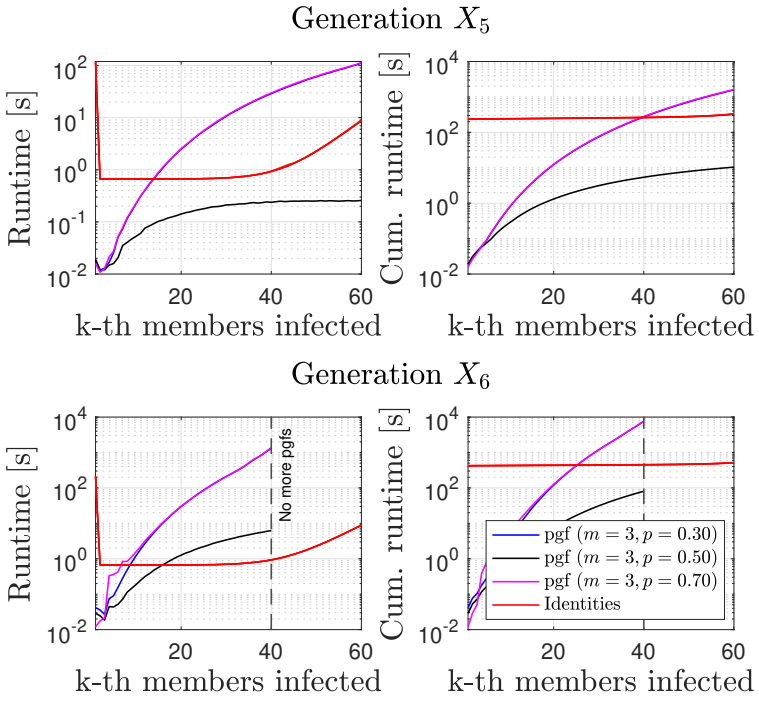
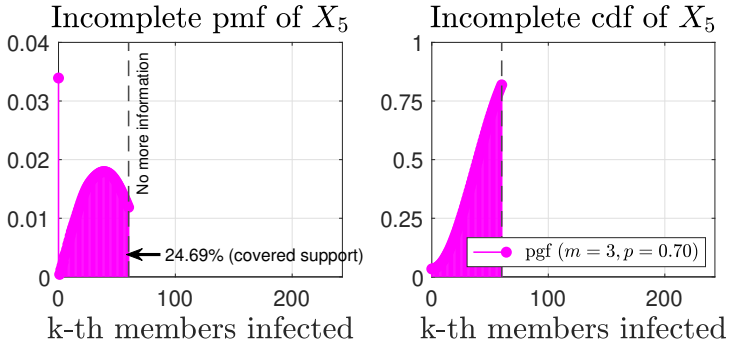


Figure 3.8: Runtime and cumulative runtime for the case of $C \sim Binomial\,(3, p)$ when comparing the pgf itself and with the help of polynomial identities.

The pgf itself is not feasible to cover the entire support of the distribution for generations further than the 5th one, due its already high values of runtime for a few quantity of members infected. The parameter $p = 0.5$ is a exception. Its simulation runs faster than the others, because there is a beneficial factorization in the probability generating function when this value occurs. A forced expansion of the analytical expression provides indeed similar runtimes coming from the other values of p. The polynomial identity methodology for $m = 3$ has in general greater runtimes than the pgf itself in this situation, but it is also not able to find all values of the mass function for these random variables. The missed probabilistic content is now significant. Figure 3.9 illustrates this aspect for the worst scenario of the 5th and 6th generations, which is $C \sim Binomial\,(3, 0.7)$.



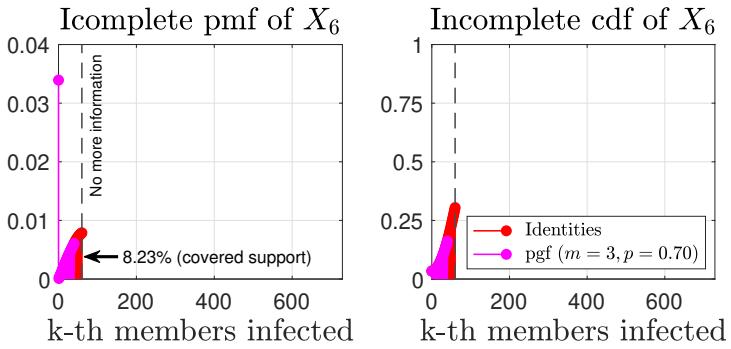


Figure 3.9: Incomplete pmf and cdf of X_5 and X_6 for $C \sim Binomial\,(3, 0.7)$.

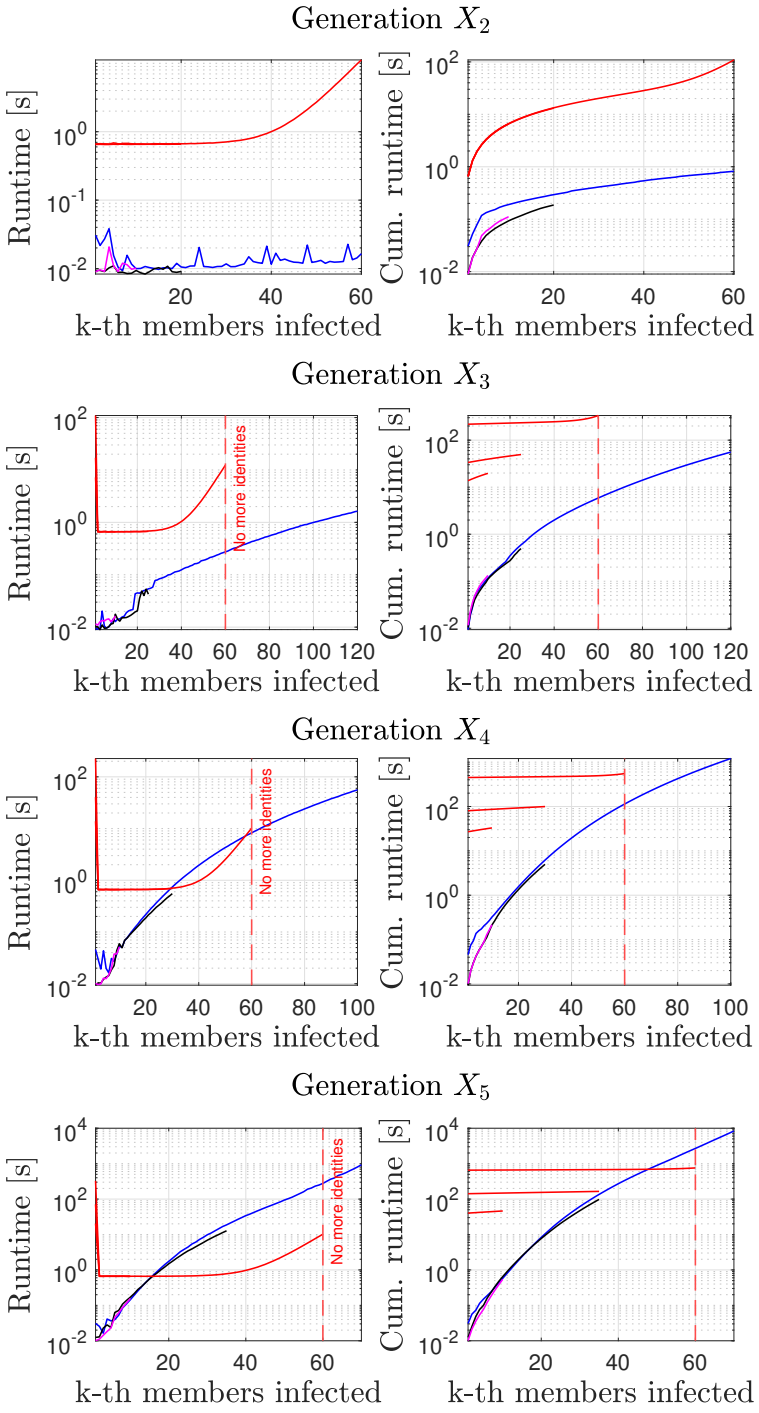
3.2.1.2 Contagion modeled as a Geometric-0 distribution

For the case of $C \sim Geometric - 0\,(p)$, the probability of infection for each member p varies in the range $p = [0.3, 0.5, 0.7]$. Unlike the previous one, this distribution has not an upper limit for its support. Monte Carlo is then first run for all the simulations desired to get a proper upper limit, because this methodology allows with a few number of experiments and computational costs to visualize the main interval inside the support that enclosures significant probabilistic information. These results are displayed in Table 3.1. The values of the mass function for the 2nd generation up to the 6th one are also sought in here.

p	**t-th Generation**				
	2	**3**	**4**	**5**	**6**
0.3	60	120	180	240	300
0.5	20	25	30	35	40
0.7	10	10	10	10	10

Table 3.1: Upper limit of the support according to Monte Carlo simulations.

Figure 3.10 presents the runtime and cumulative runtime of these simulations. This time, the pgf and the polynomial identities cover completely the proposed support of the distributions for $C \sim Geometric - 0\,(0.5)$ and $C \sim Geometric - 0\,(0.7)$. The worst scenario is $C \sim Geometric - 0\,(0.3)$, in which the number of identities contemplates totally only the second generation. Moreover, in this same scenario, the pgf methodology begins to struggle in terms of runtime from the 4th generation on to entail all the support proposed. It turns not feasible anymore to find the values of the mass function.


Generation X_2
Runtime [s]
Cum. runtime [s]
k-th members infected
k-th members infected
Generation X_3
No more identities
Runtime [s]
Cum. runtime [s]
k-th members infected
k-th members infected
Generation X_4
No more identities
Runtime [s]
Cum. runtime [s]
k-th members infected
k-th members infected
Generation X_5
No more identities
Runtime [s]
Cum. runtime [s]
k-th members infected
k-th members infected

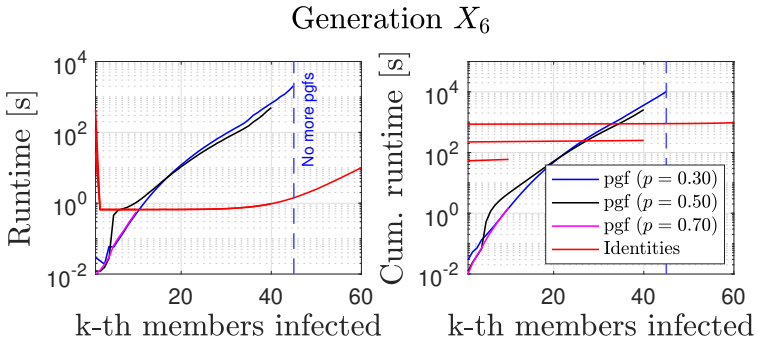


Figure 3.10: Runtime and cumulative runtime for the case of $C \sim Geometric-0\,(p)$ when comparing the pgf itself and with the help of polynomial identities.

The situation gets more complicated in the 6th generation, in which even with the limitation of the number of polynomial identities, the earlier probabilities are way easier assessed by this means. The missing probabilistic content is explored in Figure 3.11.

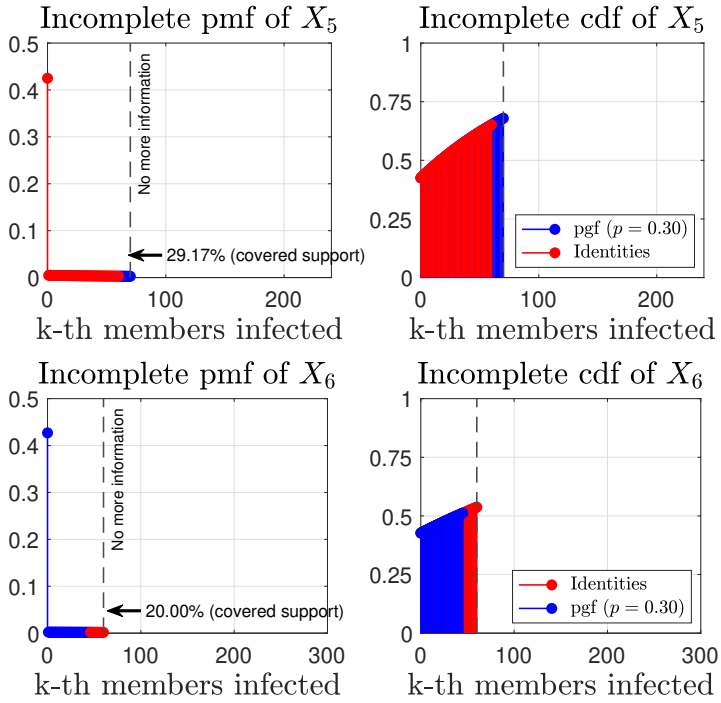


Figure 3.11: Incomplete pmf and cdf of X_5 and X_6 for $C \sim Geometric-0\,(0.3)$.

The different cumulative curves coming from the use of the polynomial identities for the Geometric-0 case is a consequence of the support difference of each contagion random variable evaluated. The runtime spent per member infected is indeed the same per generation regardless of the model of the contagion. The highest-valued curve in the cumulative runtime comparisons over generations are the ones for $p = 0.3$, the mid ones from $p = 0.5$ and the lowest-valued related to $p = 0.7$.

3.2.2 Global comparisons

In this section, the analysis are done in a global perspective of the generation's size. Moreover, the methodologies are also compared in terms of storage, except for the pgf, because the symbolic expressions get too complex and it is not feasible to be stored in a data file. For the contagion modeled as a Binomial random variable, Monte Carlo simulations, Markov chain and the best scenario of the pgf regarding the use of polynomial identities or not based are discussed. This last choice is based primarily on covering a greater support and then on the cumulative runtime spent in case of the first one is the same. For the Geometric-0 case, Markov chain methodology is not available.

3.2.2.1 Contagion modeled as a Binomial distribution

The analysis begins with the parametric study of the contagion random variable modeled as $C \sim Binomial\,(2, p)$. For the Monte Carlo methodology, the tolerance adopted was $\xi = 0.001$ for all simulations. The convergence study was made individually for each generation desired based on $p = 0.7$, and the required number of experiments n_r and the correspondent expectation of its random variables are shown in Table 3.2. The choice of $p = 0.7$ represents the worst scenario, in which the probability of higher numbers of infected people is seen.

t-th Generation	**2**	**3**	**4**	**5**	**6**
n_r	1300	3900	14000	23000	54000
$\mathbb{E}\,(X_t)$	1.960	2.744	3.842	5.378	7.530

Table 3.2: Number of experiments required per generation for $C \sim Binomial\,(2, p)$ and $\xi = 0.001$.

Based mainly on covering as many number of infected members per generation as possible and then, in case of the support covered is the same, on the cumulative runtime spent, Table 3.3 shows when the pgf is studied in a global sense with or without the help of the polynomial identities. In Table 3.3 PGF refers to the case without the polynomials and ID for the opposite.

p	**t-th Generation**				
	2	**3**	**4**	**5**	**6**
0.3	PGF	PGF	PGF	PGF	PGF
0.5	PGF	PGF	PGF	PGF	PGF
0.7	PGF	PGF	PGF	PGF	PGF

Table 3.3: Description of when using probability generating function with or without the help of polynomial identities for $C \sim Binomial\,(2, p)$.

Next, Figure 3.12 displays the runtime spent in a global sense for the three methodologies and for the parametric analysis of p.

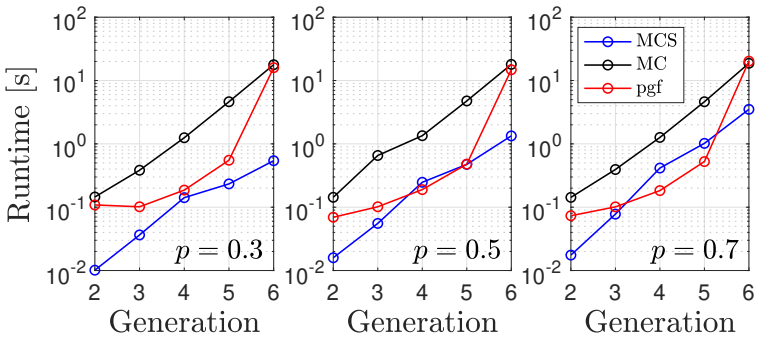


Figure 3.12: Runtime spent in a global sense for $C \sim Binomial\,(2, p)$.

The Monte Carlo simulations's runtime is actually a stochastic object, because it relies on the random generator. Figure 3.12 illustrates only a single realization of the runtime in each generation. For this specific scenario, this methodology shows in general quicker runtimes. The Markov chain approach has intrinsically a deterministic runtime, regardless of computational noises. The same happens to the pgf approach, in which quicker values than the MC one are presented, but it has the highest runtime increase per generation, due the complexity advance of its analytical expressions associated to function multicompositions.

The data file from a Monte Carlo simulations contains the runtime spent for each experiment from the n_r required, the number of members infected per generation, the approximated values of the mass functions from the generation's size random variables. For the Markov chain methodology, the data consists of all necessary one-step transition matrices, the values of the initial distribution, the probabilities of the size from the generations and the runtime spent of the entire simulation. Figure 3.13 shows the comparison in terms of storage.

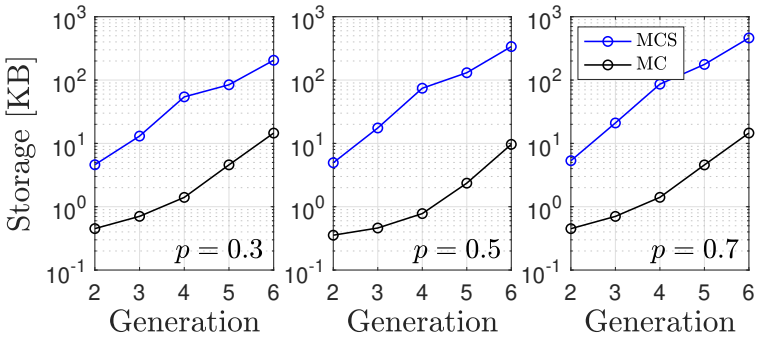


Figure 3.13: Storage spent in a global sense for $C \sim Binomial\,(2, p)$.

The storage is also a random object for the Monte Carlo simulations. Regardless the value of p, the storage of Monte Carlo methodology has the greatest values in these realizations presented in Figure 3.13. The storage size for the Markov chain approach remains almost the same regardless of the value of p, except for $p = 0.5$, and it is a deterministic object. The storage size differences between files with variables of the same amount of RAM is a consequence of MAT-files' features of storing with smaller data types and using a specific data compression technique. More explanations of this subject can be found in [24].

Now, the contagion is modeled as $C \sim Binomial\,(3, p)$. The same criteria and considerations used above are again in here adopted. Table 3.4 shows the number of experiments n_r required and the expectation of the generations' random variable.

t-th Generation	**2**	**3**	**4**	**5**	**6**
n_r	12000	19000	49000	129000	3346000
$\mathbb{E}\,(X_t)$	4.410	9.261	19.448	40.841	85.767

Table 3.4: Number of experiments required per generation for $C \sim Binomial\,(3, p)$ and $\xi = 0.001$.

Table 3.5 presents the scenarios in which the use of polynomial identities have greater results according to the conditions established. This time, it is seen that polynomial identities were at an advantage in general for greater values of p.

p	t-th Generation				
	2	**3**	**4**	**5**	**6**
0.3	PGF	PGF	PGF	ID	ID
0.5	PGF	PGF	PGF	PGF	ID
0.7	PGF	PGF	PGF	ID	ID

Table 3.5: Description of when using probability generating function with or without the help of polynomial identities $C \sim Binomial\,(3, p)$.

Runtime comparisons in a global sense are displayed in Figure 3.14 for $C \sim Binomial\,(3, p)$. The red dashed lines are reminders of the incomplete information of the pmf of the random variables for the pgf case. As a consequence, the runtimes visualized of this methodology are lower than what is indeed expected.

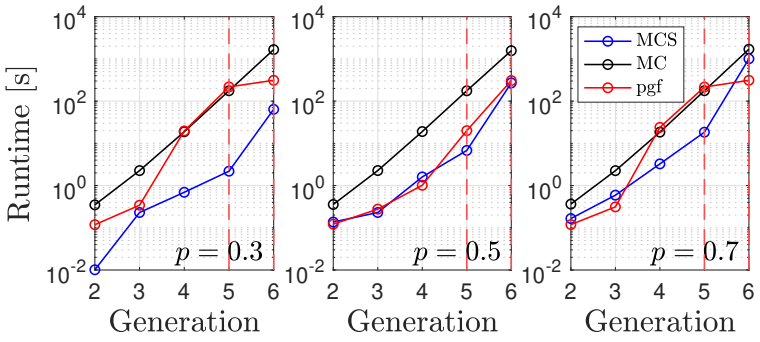


Figure 3.14: Runtime spent in a global sense for $C \sim Binomial\,(3, p)$.

The Monte Carlo simulations methodology and the Markov chain have closer values than the ones in the case $C \sim Binomial\,(2, p)$. But it is important to highlight that the presented MCS values are just from one realization of the runtime for each generation. Next, Figure 3.15 illustrates the storage spent. The same characteristics observed for the case of $C \sim Binomial\,(2, p)$ are somehow found in here given the scale difference aspect.

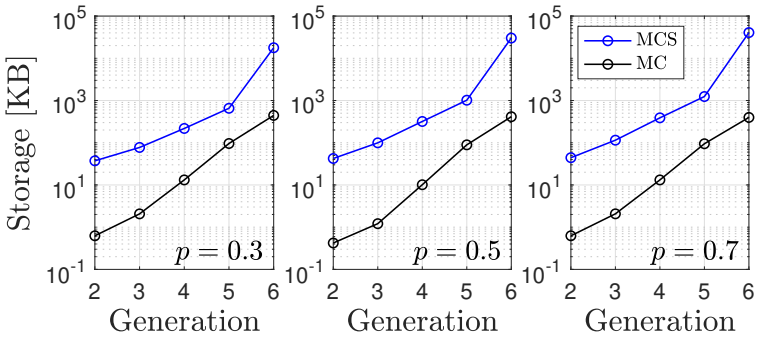


Figure 3.15: Storage spent in a global sense for $C \sim Binomial\,(3, p)$.

Another meaningful feature that has not been discussed so far is the advantage of the analytical expression that links the contagion random variable with the generation's size random variable. Purely numerical methodologies, such as the Monte Carlo simulations, do not get this probabilistic information. As a consequence, it is not possible for example to use this methodology to build a likelihood function in a bayesian inference problem. On the other hand, Markov chain and the pgf approaches are effective to perform either a deductive or inductive logical problem in this context.

3.2.2.2 Contagion modeled as a Geometric-0 distribution

The convergence analysis of Monte Carlo simulations for the parametric study of the contagion modeled as $C \sim Geometric - 0\,(p)$ is presentend in Table 3.6. Now, the convergence study was made individually for each generation desired based on $p = 0.3$ and the tolerance previously imposed is also $\xi = 0.001$.

t-th Generation	**2**	**3**	**4**	**5**	**6**
n_r	12000	38000	98000	190000	989000
$\mathbb{E}\,(X_t)$	5.444	12.704	29.642	69.167	161.384

Table 3.6: Number of experiments required per generation for $C \sim Geometric - 0\,(p)$ and $\xi = 0.001$.

The cases in which the pgf approach with polynomial identities are at an advantage are represented in Table 3.7. The ID status for $p = 0.3$ and X_6 is a better option because the pgf methodology could not cover a greater support due runtime issues, while for $p = 0.3$ and X_5 the cumulative runtime with the use of polynomial identities is lower than without it for the same support.

p	**t-th Generation**				
	2	**3**	**4**	**5**	**6**
0.3	PGF	PGF	PGF	PGF	ID
0.5	PGF	PGF	PGF	PGF	ID
0.7	PGF	PGF	PGF	PGF	PGF

Table 3.7: Description of when using probability generating function with or without the help of polynomial identities $C \sim Geometric - 0\,(p)$.

Figure 3.16 displays the runtime comparison of both the methodologies: MCS and pgfs. The red dashed lines represent this time the generations in which the pgf methodology with or without the polynomial identities could

not cover the proposed support from a preliminary analysis coming from the Monte Carlo simulations exposed in Table 3.1.

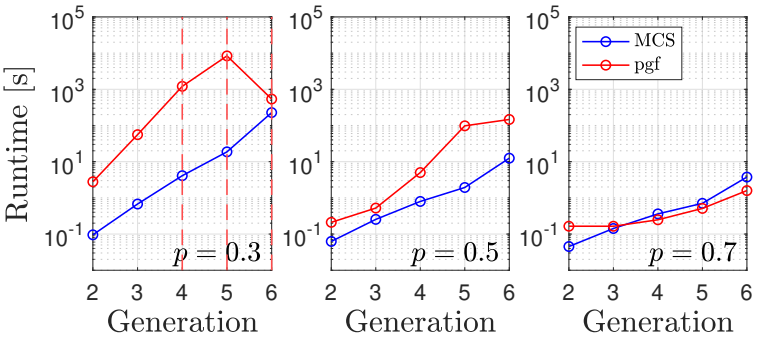


Figure 3.16: Runtime spent in a global sense for $C \sim Geometric - 0\,(p)$.

The pgf approach comes closer to the values of runtime of the MCS methodology for the values of $p = 0.5$ and $p = 0.7$. The main reason is that the upper limit of the support of each generation of them is not that great as in $p = 0.3$. Even not covering the complete proposed support, the runtimes spent for the pgf are greater than the MCS ones for $p = 0.3$

Finally, the storage analysis is presented in Fig. 3.17. In this case, there is not any other methodology to compare with the MCS one. The results presented are again just a realization of this random object. The number of experiments n_r has a meaningful influence on it, because the data collects n_r realizations of each random variable from the BGW process. This is one of the main reasons of the difference between the storage size of X_6 for $C \sim Binomial\,(2, p)$, $C \sim Binomial\,(3, p)$ and $C \sim Geometric - 0\,(p)$.

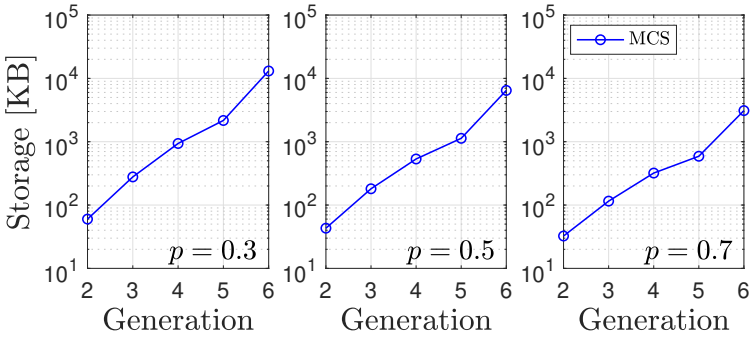


Figure 3.17: Storage spent in a global sense for $C \sim Geometric - 0\,(p)$.

4
Inverse problem of bayesian parametric inference

4.1
Parametric inference problem of the contagion random variable

Let continuous parameters of a general model be assigned with the random vector $\boldsymbol{W} = (W_1, W_2, \ldots, W_k)$ and the data from different sources with another random vector $\boldsymbol{D} = (D_1, D_2, \ldots, D_n)$. Inspired in the idea of Eq. (1-4),

$$\begin{aligned}
\mathbb{P}(\boldsymbol{W} \mid \boldsymbol{D}) &= \int_{\boldsymbol{W}} f_{\boldsymbol{W}|\boldsymbol{D}}(\boldsymbol{w} \mid \boldsymbol{d})\, d\boldsymbol{w} \\
&= \int_{\boldsymbol{W}} \frac{f_{\boldsymbol{W},\boldsymbol{D}}(\boldsymbol{w}, \boldsymbol{d})}{f_{\boldsymbol{D}}(\boldsymbol{d})}\, d\boldsymbol{w} \\
&= \int_{\boldsymbol{W}} \frac{f_{\boldsymbol{D},\boldsymbol{W}}(\boldsymbol{d}, \boldsymbol{w})}{f_{\boldsymbol{D}}(\boldsymbol{d})}\, d\boldsymbol{w} \\
&= \int_{\boldsymbol{W}} \frac{f_{\boldsymbol{D}|\boldsymbol{W}}(\boldsymbol{d} \mid \boldsymbol{w})\ f_{\boldsymbol{W}}(\boldsymbol{w})}{f_{\boldsymbol{D}}(\boldsymbol{d})}\, d\boldsymbol{w} \\
\int_{\boldsymbol{W}} f_{\boldsymbol{W}|\boldsymbol{D}}(\boldsymbol{w} \mid \boldsymbol{d})\, d\boldsymbol{w} &= \int_{\boldsymbol{W}} \frac{f_{\boldsymbol{D}|\boldsymbol{W}}(\boldsymbol{d} \mid \boldsymbol{w})\ f_{\boldsymbol{W}}(\boldsymbol{w})}{f_{\boldsymbol{D}}(\boldsymbol{d})}\, d\boldsymbol{w} \\
f_{\boldsymbol{W}|\boldsymbol{D}}(\boldsymbol{w} \mid \boldsymbol{d}) &= \frac{f_{\boldsymbol{D}|\boldsymbol{W}}(\boldsymbol{d} \mid \boldsymbol{w})\ f_{\boldsymbol{W}}(\boldsymbol{w})}{f_{\boldsymbol{D}}(\boldsymbol{d})}
\end{aligned} \tag{4-1}$$

The development above begins with the probability measure $\mathbb{P}$ assigned to a continuous random vector. For a Lebesgue measure, the idea of probability is related to the cdf of the random object. As a consequence, four functions are related. In a parametric inference problem, the data is an observable feature $\boldsymbol{d} = \hat{\boldsymbol{d}}$. Eq. (4-1) then is rewritten as

$$f_{\boldsymbol{W}|\boldsymbol{D}}\left(\boldsymbol{w} \mid \boldsymbol{d} = \hat{\boldsymbol{d}}\right) = \frac{f_{\boldsymbol{D}|\boldsymbol{W}}\left(\boldsymbol{d} = \hat{\boldsymbol{d}} \mid \boldsymbol{w}\right)\ f_{\boldsymbol{W}}(\boldsymbol{w})}{f_{\boldsymbol{D}}\left(\boldsymbol{d} = \hat{\boldsymbol{d}}\right)} \tag{4-2}$$

Each term in Eq. (4-2) has a special name in literature:

– Posterior distribution is the one identified by $f_{\boldsymbol{W}|\boldsymbol{D}}\left(\boldsymbol{w} \mid \boldsymbol{d} = \hat{\boldsymbol{d}}\right)$. This represents the degree of belief of the underlying possible values of the parameters assigned to the random vector $\boldsymbol{W}$ after the content of the data gets observable, i.e., there is a realization of $\boldsymbol{D}$.

- Likelihood function is the one assigned in $L(\boldsymbol{w}) = f_{\boldsymbol{D}|\boldsymbol{W}}\left(\boldsymbol{d} = \hat{\boldsymbol{d}} \mid \boldsymbol{w}\right)$. It is responsible to change the previous beliefs of the underlying possible values of the parameters given a new current information available in the realization of $\boldsymbol{D}$. Notice that $L(\boldsymbol{w})$ is not necessarily a pdf. Instead of that, it is a dimensionless numerical function, when multiplied by a prior and a normalization constant may turn into a posterior distribution.
- Prior distribution is indicated as $f_{\boldsymbol{W}}(\boldsymbol{w})$. This pdf states the knowledge of the underlying values of the parameters assigned to random vector $\boldsymbol{W}$ before data is available.
- Evidence is related to the quantity in $f_{\boldsymbol{D}}\left(\boldsymbol{d} = \hat{\boldsymbol{d}}\right)$. Its dependency relies only on the data. Therefore, it is a constant in the inference problem and in many situation is not evaluated leading to Eq. (4-3).

$$f_{\boldsymbol{W}|\boldsymbol{D}}\left(\boldsymbol{w} \mid \boldsymbol{d} = \hat{\boldsymbol{d}}\right) \propto f_{\boldsymbol{D}|\boldsymbol{W}}\left(\boldsymbol{d} = \hat{\boldsymbol{d}} \mid \boldsymbol{w}\right) f_{\boldsymbol{W}}(\boldsymbol{w}). \tag{4-3}$$

Eqs. (4-2) and (4-3) provide an easier way to tackle inference problems. The hard task of describing directly the knowledge for the hypothesis regarding the parameters in light of data is rearranged for a combination of more understandable functions. Through the prior function, it is possible to incorporate any source of knowledge of the problem in question before data is available. Moreover, anytime a new information arrives, it is also possible to update the current belief with the help of the likelihood.

The data can be incorporated sequentially or in a batch way. When the data is independent, which means that the measurement of one datum does not affect the outcome of another, the likelihood function can be expressed in the one-step configuration as

$$f_{\boldsymbol{D}|\boldsymbol{W}}\left(\{\boldsymbol{d}_k\} = \left\{\hat{\boldsymbol{d}}_k\right\} \mid \boldsymbol{w}\right) = \prod_{z=1}^{k} f_{\boldsymbol{D}|\boldsymbol{W}}\left(\boldsymbol{d}_{\boldsymbol{z}} = \hat{\boldsymbol{d}}_z \mid \boldsymbol{w}\right). \tag{4-4}$$

In our context, the aim is to make bayesian inferences of parameters from the distribution function which models the contagion random variable, given some set of data. We begin by proposing a discrete random variable family to it, for instance the Binomial distribution. The challenge is now to assign a likelihood function and a prior distribution.

The data available could be for example a collection of different spots of a same region, all of them referring to the first generation's size of members infected. A good technique to assign a proper likelihood function is to identify a sampling distribution $f_{\boldsymbol{D}|\boldsymbol{W}}(\boldsymbol{d} \mid \boldsymbol{w} = \hat{\boldsymbol{w}})$ when available. In this case, the sampling distribution for each data is indeed the mass function from the Binomial contagion random variable

$$f_{X_1|\boldsymbol{W}}\left(x_1 \mid \boldsymbol{w} = \{\hat{m}, \hat{p}\}\right) = \frac{\hat{m}!}{x_1!\,(\hat{m} - x_1)!}\,\hat{p}^{\,x_1}\,(1-\hat{p})^{\hat{m}-x_1}\,.$$

To simplify the problem, we assume that the parameter m is beforehand known, so it is an observable $\hat{m}$. The bayesian inference is applied only to the parameter p. Therefore, given the sampling distribution that is related to the deductive logical analysis, the likelihood function is

$$L\left(p\right) = \frac{\hat{m}!}{\hat{x}_1!\,(\hat{m} - \hat{x}_1)!}\,p^{\,\hat{x}_1}\,(1-p)^{\hat{m}-\hat{x}_1}\,.$$

If the k data collected from different spots are independent, the posterior distribution representing the degree of belief in the underlying values of parameter p is given in a batch sample as

$$f_{P|X_1}\left(p \mid \{x_{1_k}\} = \{\hat{x}_{1_k}\}\right) \propto \prod_{z=1}^{k}\left[\frac{\hat{m}!}{\hat{x}_{1_z}!\,(\hat{m} - \hat{x}_{1_z})!}\,p^{\,\hat{x}_{1_z}}\,(1-p)^{\hat{m}-\hat{x}_{1_z}}\right] f_P\left(p\right)$$

In the absence of any coherent information available to ascribe features to the prior distribution, a flat uniform one reflects the previous state of knowledge. As a consequence, the prior distribution has no longer a dependency on the parameter, regardless of the support $[0, 1]$. It is a constant. The posterior is now proportional only to the likelihood function,

$$f_{P|X_1}\left(p \mid \{x_{1_k}\} = \{\hat{x}_{1_k}\}\right) \propto \prod_{z=1}^{k}\left[\frac{\hat{m}!}{\hat{x}_{1_z}!\,(\hat{m} - \hat{x}_{1_z})!}\,p^{\,\hat{x}_{1_z}}\,(1-p)^{\hat{m}-\hat{x}_{1_z}}\right]$$

When data comes directly from realizations of the first generation's size, the likelihood function for the inverse problem is at first hand available, because a law is assumed for the contagion. The attempts of the bayesian inference in this situation are to find the most suitable probabilistic description of the degree of belief of the underlying values of parameters according to the data. The situation gets more complex when data comes from a further generation or from a realization of the stochastic process observed over a certain number of generations. In each one of these two scenarios, a specific strategy to make the bayesian inference is required. The two developed strategies are explained next.

4.1.1
First strategy: data coming from some further generation

Consider that the source of data is any further generation's size of members infected of different spots of a region. We start again proposing a probabilistic model to the contagion, but how can we build a proper likelihood function? The best attempt is to look at the BGW process in a deductive perspective and find somehow another sampling distribution. As explored in

the previous section, the pgf and Markov chain approaches provide piecewise pmfs for each state from any generation attached to the BGW process. As a novelty, this work shows that they are indeed tools to build the likelihood functions.

The first technique has a sampling distribution according to Eq. (3-6) from the previous section

$$f_{X_t|\boldsymbol{W}}\left(x_t \mid \boldsymbol{w}=\hat{\boldsymbol{w}}\right)=\frac{1}{x_t!}\left.\frac{d^{(x_t)}\left[G_C\left(G_C\left(\ldots\left(G_C\left(s,\boldsymbol{w}=\hat{\boldsymbol{w}}\right)\right)\right)\right)\right]}{ds^{(x_t)}}\right|_{s=0}, \tag{4-5}$$

in which the multicomposition function has $t-1$ recurrence assignments. We emphasize this time that the pgf $G_C\left(s,\boldsymbol{w}=\hat{\boldsymbol{w}}\right)$ has also a dependency on parameters $\boldsymbol{w}$ from the distribution assigned to the contagion. It is usually not represented, because in the deductive logic, the contagion is completely previously defined. Based on Eq. (4-5), the piecewise likelihood function is introduced as

$$L\left(\boldsymbol{w}\right)=\frac{1}{\hat{x}_t!}\left.\frac{d^{(\hat{x}_t)}\left[G_C\left(G_C\left(\ldots\left(G_C\left(s,\boldsymbol{w}\right)\right)\right)\right)\right]}{ds^{(\hat{x}_t)}}\right|_{s=0}. \tag{4-6}$$

The construction of the likelihood function this time based on Markov chain technique has also the limitation that the support of the contagion's probabilistic model must be finite. The sampling distribution according to this method is given in Eq. (3-13).

$$\begin{aligned} f_{X_t|\boldsymbol{W}}\left(x_t \mid \boldsymbol{w}=\hat{\boldsymbol{w}}\right)=&\lambda^C\left(\boldsymbol{w}=\hat{\boldsymbol{w}}\right)\boldsymbol{T}^{(1)}\left(\boldsymbol{w}=\hat{\boldsymbol{w}}\right)\boldsymbol{T}^{(2)}\left(\boldsymbol{w}=\hat{\boldsymbol{w}}\right)\ldots \\ &\boldsymbol{T}^{(t-1)}\left(\boldsymbol{w}=\hat{\boldsymbol{w}}\right) \end{aligned} \tag{4-7}$$

Again, we highlight the dependency on parameters from the contagion random variable. The pmf λ^C is directly dependent on them, whereas the elements of one-step transition matrices rely on pgfs, which are functions of these parameters. Finally, the likelihood function is obtained. Einstein summation convention is adopted here.

$$\begin{aligned} L\left(\boldsymbol{w}\right)=&\lambda_i^C\left(\boldsymbol{w}\right)\boldsymbol{T}_{ij}^{(1)}\left(\boldsymbol{w}\right)\boldsymbol{T}_{jk}^{(2)}\left(\boldsymbol{w}\right)\ldots\boldsymbol{T}_{lm}^{(t-1)}\left(\boldsymbol{w}\right)\delta_{m\,\hat{x}_t} \\ =&\lambda_{\hat{x}_t}^{X_t}\left(\boldsymbol{w}\right) \end{aligned} \tag{4-8}$$

In other words, the likelihood function in this case is the analytical expression that belongs to the position $\hat{x}_t$ from the vector λ_{X_t}, which is a function of the parameters $\boldsymbol{w}$.

In order to clarify how the bayesian parameter inference works and give a better glimpse on how to use data from some further generation for this purpose, an example is presented next. Our source of data comes for instance from different realizations of the 2nd generation from the BGW process. Again, we assume that the contagion's random variable model is $Binomial\left(3,p\right)$.

Therefore, the aim is to estimate the parameter p.

According to the bayesian perspective, the parameter is indeed treated as random variable P. Once more, there is not any coherent information in our hands about the underlying values of this random variable. The priori distribution is then better described as a uniform distribution. The state space from X_2 is in this situation $\mathbb{S} = \{0, 1, \ldots, 9\}$. These are all the possible realizations of the 2nd generation's size random variable. Each one of them associates a dimensionless numerical function relying on p, i.e., a likelihood function.

We could choose in this scenario between Eq. (4-6) and Eq. (4-8) to get the expressions. Selecting the second one, which seems trickier at first hand, we first need the initial distribution vector $\lambda^C(p)$, which has all the probabilities related to the random variable C, but relying on the parameter p. From the pmf of the Binomial distribution,

$$\lambda^C(p) = \Big[\underbrace{-(p-1)^3}_{C=0} \;\; \underbrace{3\,p\,(p-1)^2}_{C=1} \;\; \underbrace{-3\,p^2\,(p-1)}_{C=2} \;\; \underbrace{p^3}_{C=3} \;\; \underbrace{0 \;\ldots\; 0}_{4\leq C\leq 9}\Big].$$

The other fundamental point in here is to write the one-step transition matrices. In order to move from the 1st generation to the 2nd one, only the 1st one-step transition matrix $\boldsymbol{T}^{(1)}$ is required. Each element of it depends for our purpose on a function of $G_C(s, \boldsymbol{w})$, which is obtained in a similar way as presented in Eq. (3-14) combined with Eq. (3-16),

$$p_{i,j}(t, \boldsymbol{w}) = \begin{cases} 1, \text{ if } \begin{cases} i = j = 0 \\ i > {q_C}^t(\boldsymbol{w}),\ j = 0 \end{cases} \\ 0, \text{ if } \begin{cases} i = 0,\ j \neq 0 \\ i > {q_C}^t(\boldsymbol{w}),\ j \neq 0 \\ 0 < i \leq {q_C}^t(\boldsymbol{w}),\ j > i \times q_C(\boldsymbol{w}) \end{cases} \\ \mathbb{P}\left(\sum_{k=1}^{X_t=i} C = j\right) = \dfrac{1}{j!}\,\dfrac{d^{(j)}\left[{G_C}^i(s, \boldsymbol{w})\right]}{ds^{(j)}}\Big|_{s=0}, \text{otherwise.} \end{cases} \tag{4-9}$$

For instance, the first four elements of the third row from the 1st one-step transition matrices $p_{2,0}(1,p), p_{2,1}(1,p), p_{2,2}(1,p), p_{2,3}(1,p)$, which suppose $C = 2$, according to Eq. (4-9) are

$$p_{2,0}(1,p) = \frac{1}{0!}\left[\left[(1-p) + p\,s\right]^3\right]^2\Big|_{s=0} = (p-1)^6$$

$$p_{2,1}(1,p) = \frac{1}{1!}\left.\frac{d^{(1)}\left[\left[(1-p)+p\,s\right]^3\right]^2}{ds^{(1)}}\right|_{s=0} = -6\,p\,(p-1)^5$$

$$p_{2,2}(1,p) = \frac{1}{2!}\left.\frac{d^{(2)}\left[\left[(1-p)+p\,s\right]^3\right]^2}{ds^{(2)}}\right|_{s=0} = 15\,p^2\,(p-1)^4$$

$$p_{2,3}(1,p) = \frac{1}{3!}\left.\frac{d^{(3)}\left[\left[(1-p)+p\,s\right]^3\right]^2}{ds^{(3)}}\right|_{s=0} = -20\,p^3\,(p-1)^3\,.$$

Given the complete first one-step transition matrix $\boldsymbol{T}^{(1)}(p)$, we are now able to find all the possible likelihood expressions per state.

$$\lambda^{X_2}(p) = \lambda^{C}(p)\,\boldsymbol{T}^{(1)}$$

$$\left(\lambda^{X_2}(p)\right)^{\top} = \begin{bmatrix} -(p-1)^3 - 3\,p^2\,(p-1)^7 - p^3\,(p-1)^9 - 3\,p\,(p-1)^5 \\ 9\,p^2\,(p-1)^4 + 18\,p^3\,(p-1)^6 + 9\,p^4\,(p-1)^8 \\ -9\,p^3\,(p-1)^3 - 45\,p^4\,(p-1)^5 - 36\,p^5\,(p-1)^7 \\ 3\,p^4\,(p-1)^2 + 60\,p^5\,(p-1)^4 + 84\,p^6\,(p-1)^6 \\ -45\,p^6\,(p-1)^3 - 126\,p^7\,(p-1)^5 \\ 18\,p^7\,(p-1)^2 + 126\,p^8\,(p-1)^4 \\ -84\,p^9\,(p-1)^3 - 3\,p^8\,(p-1) \\ 36\,p^{10}\,(p-1)^2 \\ -9\,p^{11}\,(p-1) \\ p^{12} \end{bmatrix}$$

Suppose that the k-th first elements of a sequence of data collected from realizations of X_2 are $\{X_2\}_{k,k\in\mathbb{N}^*} = \{1,1,0,0,0,2,\ldots\}$. The first likelihood in this case is given by the second element from the vector above, once $X_2 = 1$. It makes the posterior distribution be proportional to

$$f_{P|X_2}(p \mid \{x_2\}_1 = \{1\}) \propto 9\,p^2\,(p-1)^4 + 18\,p^3\,(p-1)^6 + 9\,p^4\,(p-1)^8.$$

To make the expression above a probability density function (pdf), we should multiply it by the inverse of its integral over the domain of p, which would be exactly the value of the evidence term not mentioned. It is indeed not necessary for merely parametric estimation. Following the sequence of realizations, we have observed again $X_2 = 1$, which makes the likelihood function to be the second element of the vector as expected. The latest posterior probability expressed is now the prior one. Multiplying both the current prior and likelihood function, the updated posterior distribution is now

$$f_{P|X_2}(p \mid \{x_2\}_2 = \{1,1\}) \propto \left[9\,p^2\,(p-1)^4 + 18\,p^3\,(p-1)^6 + 9\,p^4\,(p-1)^8\right]^2.$$

This updating process follows the same way as long as data coming from X_2 is available. The latest posterior turns the brand-new prior in light of new information processed by the corresponding likelihood function in the vector presented. The subsequent normalized posterior distributions are displayed in Figure 4.1.

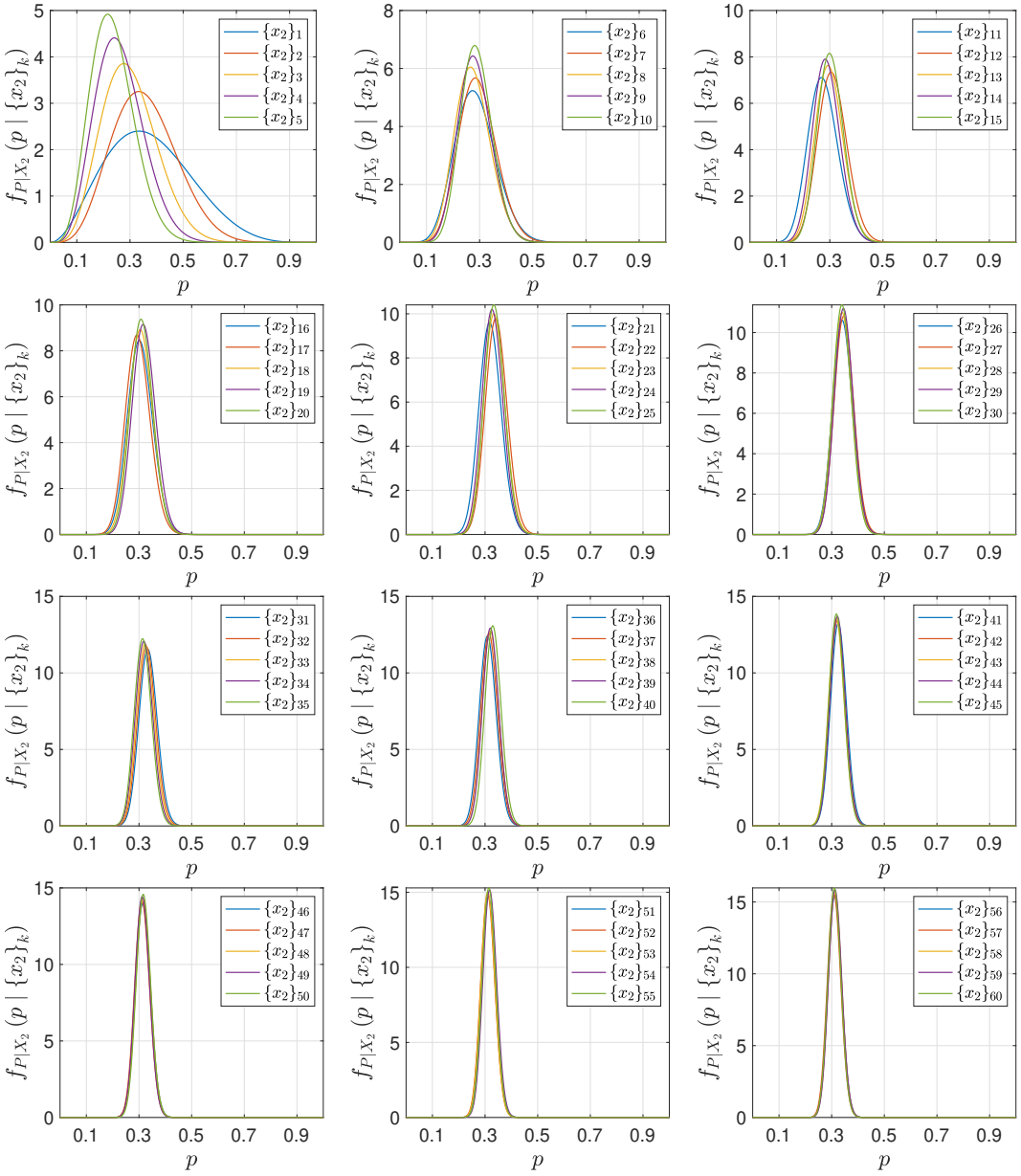


Figure 4.1: Normalized posterior distribution evolution in light of new data coming from realizations of X_2.

When the number of realizations of X_2 increases, the posterior distribution related to the parameter in question tends to become a Dirac delta function as far as the updating process lasts. This property is illustrated in Figure 4.1. The higher degree of beliefs in the possible underlying values of the random variable P concentrates each time in closer confidence intervals and

the best statistic to estimate a value for the parameter p is the mode of its posterior distribution.

4.1.2
Second strategy: data is a realization of the branching process observed over a certain number of generations

Instead of data coming from realizations of a single generation of the branching process, we could consider a different type of data. The data could be just one realization of the branching process observed over a certain number of generations. The main question remains: how to build proper likelihoods? The answer to this question is another novelty of this work. In literature, this sort of data has been used before in bayesian non-parametric inference for multitype BGW processes [25]. This is not the case of this work, since it is a bayesian parametric estimation. Before starting to develop this second strategy, a brief observation about data dependence is quite important.

Eq. (4-6) and Eq. (4-8) provide likelihood functions for data coming from a further generation. We could at first in this new strategy think to use one of these equations applied to each generation from the ramification tree with no concerns. At first sight this seems like a good idea, however, in the previous case, the outcome of some X_t has not any influence on the next outcome of this same random variable, since they are not from the same BGW process. This is not what happens here. The subsequent data collected are structured. There is a dependence, which modifies the updating process and does not allow to use the assumption made in Eq. (4-4).

According to the Eq. (4-3), the first updating process is written as

$$f_{\boldsymbol{W}|\boldsymbol{D}}\left(\boldsymbol{w} \mid \boldsymbol{d}=\hat{\boldsymbol{d}}\right) \propto f_{\boldsymbol{D}|\boldsymbol{W}}\left(\boldsymbol{d}=\hat{\boldsymbol{d}} \mid \boldsymbol{w}\right) f_{\boldsymbol{W}}(\boldsymbol{w}).$$

In order to save notation, we define in here

$$\begin{aligned} f\left(\boldsymbol{w} \mid \hat{\boldsymbol{d}}\right) &:= f_{\boldsymbol{W}|\boldsymbol{D}}\left(\boldsymbol{w} \mid \boldsymbol{d}=\hat{\boldsymbol{d}}\right) \\ f\left(\hat{\boldsymbol{d}} \mid \boldsymbol{w}\right) &:= f_{\boldsymbol{D}|\boldsymbol{W}}\left(\boldsymbol{d}=\hat{\boldsymbol{d}} \mid \boldsymbol{w}\right) \\ f(\boldsymbol{w}) &:= f_{\boldsymbol{W}}(\boldsymbol{w}). \end{aligned} \tag{4-10}$$

For the case of this parametric inference problem, the updating starts with $\boldsymbol{D} = X_1$ and the parameter investigated is $\boldsymbol{W} = P$,

$$f(p \mid \hat{x}_1) \propto f(\hat{x}_1 \mid p)\, f(p) \tag{4-11}$$

The following update comes from data related to X_2, when the most previous posterior distribution is now the prior one,

$$f(p \mid \hat{x}_2, \hat{x}_1) \propto f(\hat{x}_2 \mid \hat{x}_1, p)\, f(p \mid \hat{x}_1) \tag{4-12}$$

Next, new information comes from X_3, then from X_4 and so on. The updates up to some X_t data are presented in the following.

$$f\left(p \mid \hat{x}_3, \hat{x}_2, \hat{x}_1\right) \propto f\left(\hat{x}_3 \mid \hat{x}_2, \hat{x}_1, p\right) f\left(p \mid \hat{x}_2, \hat{x}_1\right) \tag{4-13}$$

$$f\left(p \mid \hat{x}_4, \hat{x}_3, \hat{x}_2, \hat{x}_1\right) \propto f\left(\hat{x}_4 \mid \hat{x}_3, \hat{x}_2, \hat{x}_1, p\right) f\left(p \mid \hat{x}_3, \hat{x}_2, \hat{x}_1\right) \tag{4-14}$$

$$\vdots$$

$$f\left(p \mid \hat{x}_t, \hat{x}_{t-1}, \ldots, \hat{x}_2, \hat{x}_1\right) \propto f\left(\hat{x}_t \mid \hat{x}_{t-1}, \ldots, \hat{x}_2, \hat{x}_1, p\right) f\left(p \mid \hat{x}_{t-1}, \ldots, \hat{x}_2, \hat{x}_1\right). \tag{4-15}$$

The prior distribution from Eq. (4-15) is the posterior when data arrived from X_{t-1}. Defining $f\left(p \mid \{\hat{x}_t\}\right) := f\left(p \mid \hat{x}_t, \hat{x}_{t-1}, \ldots, \hat{x}_2, \hat{x}_1\right)$ and reassigning to each current prior the expression of the latest posterior in a recurrence fashion,

$$\begin{aligned} f\left(p \mid \{\hat{x}_t\}\right) &\propto f\left(\hat{x}_t \mid \hat{x}_{t-1}, \ldots, \hat{x}_2, \hat{x}_1, p\right) f\left(p \mid \{\hat{x}_{t-1}\}\right) \\ &\propto f\left(\hat{x}_t \mid \hat{x}_{t-1}, \ldots, \hat{x}_2, \hat{x}_1, p\right) f\left(\hat{x}_{t-1} \mid \hat{x}_{t-2}, \ldots, \hat{x}_2, \hat{x}_1, p\right) \\ &\quad f\left(p \mid \{\hat{x}_{t-2}\}\right) \\ &\propto f\left(\hat{x}_t \mid \hat{x}_{t-1}, \ldots, \hat{x}_2, \hat{x}_1, p\right) f\left(\hat{x}_{t-1} \mid \hat{x}_{t-2}, \ldots, \hat{x}_2, \hat{x}_1, p\right) \\ &\quad \ldots f\left(\hat{x}_2 \mid \hat{x}_1, p\right) f\left(p \mid \{\hat{x}_1\}\right) \\ &\propto f\left(\hat{x}_t \mid \hat{x}_{t-1}, \ldots, \hat{x}_2, \hat{x}_1, p\right) f\left(\hat{x}_{t-1} \mid \hat{x}_{t-2}, \ldots, \hat{x}_2, \hat{x}_1, p\right) \\ &\quad \ldots f\left(\hat{x}_2 \mid \hat{x}_1, p\right) f\left(\hat{x}_1 \mid p\right) f\left(p\right). \end{aligned} \tag{4-16}$$

Since the BGW process has the Markov property,

$$\begin{aligned} f\left(p \mid \{\hat{x}_t\}\right) \propto & \underbrace{f\left(\hat{x}_t \mid \hat{x}_{t-1}, p\right)}_{\text{(A)}} \underbrace{f\left(\hat{x}_{t-1} \mid \hat{x}_{t-2}, p\right)}_{\text{(A)}} \ldots \underbrace{f\left(\hat{x}_2 \mid \hat{x}_1, p\right)}_{\text{(A)}} \\ & \underbrace{f\left(\hat{x}_1 \mid p\right)}_{\text{(B)}} \underbrace{f\left(p\right)}_{\text{(C)}}. \end{aligned} \tag{4-17}$$

The likelihood functions signed with (A) are obtained according to

$$f\left(\hat{x}_k \mid \hat{x}_{k-1}, p\right) = \frac{1}{\hat{x}_k!} \left. \frac{d^{(\hat{x}_k)} \left[{G_C}^{\hat{x}_{k-1}}\left(s, p\right)\right]}{ds^{(\hat{x}_k)}} \right|_{s=0}. \tag{4-18}$$

Otherwise, the remain likelihood signed with (B) can be found according to Eq. (4-6) or just building it through the sampling distribution of the contagion random variable. Finally, the function marked with (C) is just the prior distribution.

Another example is presented to clarify this way of making bayesian parametric inference. We assume again that the probabilistic model for the contagion random variable is $C \sim Binomial\left(3, p\right)$ and we want to estimate the parameter p. Therefore, it is treated as a continuous random variable P

with support $[0, 1]$, which covers the physical possible values of it. Suppose a realization of the branching process is composed by the following realization of subsequent generations $\{X_k\}_{k\in\mathbb{N}^*} = \{1, 2, 4, 3, 5, 6, \ldots\}$.

Our prior knowledge of P is modeled as a uniform distribution. The first data $X_1 = 1$ provides the following likelihood according to Eq. (4-6)

$$f\left(\hat{x}_1 \mid p\right) = \frac{1}{1!}\left.\frac{d^{(1)}\left[(1-p)+p\,s\right]^3}{ds^{(1)}}\right|_{s=0} = 3\,p\,(p-1)^2\,.$$

As a consequence, the posterior distribution is

$$f\left(p \mid \hat{x}_1\right) \propto 3\,p\,(p-1)^2\,.$$

Based on the conditional statement $X_2 = 2 \mid X_1 = 1$, the next likelihood is built. From Eq. (4-18), the corresponding expression is

$$f\left(\hat{x}_2 \mid \hat{x}_1, p\right) = \frac{1}{2!}\left.\frac{d^{(2)}\left[(1-p)+p\,s\right]^3}{ds^{(2)}}\right|_{s=0} = -3\,p^2\,(p-1)\,,$$

which transforms our posterior distribution into

$$f\left(p \mid \{\hat{x}_2\}\right) \propto -9\,p^3(p-1)^3.$$

The subsequent normalized posterior distributions are displayed in Figure 4.2. It shows the evolution of them in light of new data coming from further generations of the same BGW process.

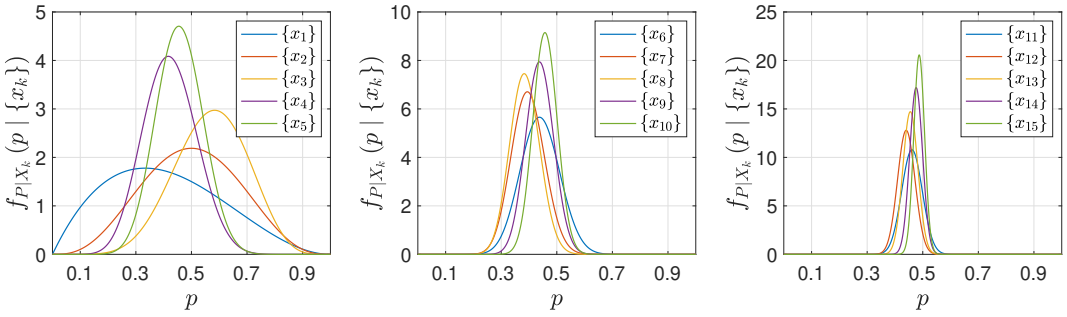


Figure 4.2: Normalized posterior distribution evolution in light of new data coming from realizations of subsequent generations of the same BGW process.

4.2
Measuring convergence in the updating of the posterior distributions

The likelihood function is responsible to change our degree of beliefs of the underlying values of $\boldsymbol{W}$ given a new information $\hat{\boldsymbol{d}}_1$ available. It modifies the shape of prior distributions' pdf and turns in into a posterior distribution. What if the information obtained was for instance $\hat{\boldsymbol{d}}_2$ instead of $\hat{\boldsymbol{d}}_1$? Could the transformation have been more significant? How to measure it?

An interesting way to do this is by calculating a distance named L_p-Wasserstein [26]. This is a probability metric that relates to an optimal mass transference problem. It is also known as the Monge-Kantorovich problem. We start with an example to give a better glimpse. Suppose we have an initial two-dimensional mass function represented by $f_{\boldsymbol{A}}(x_A, y_A)$ and we would like to transform it to a configuration given by the also two-dimensional mass function $f_{\boldsymbol{B}}(x_B, y_B)$. Figure 4.3 shows a two dimensional scatter view of $f_{\boldsymbol{A}}(x_A, y_A)$ in blue color and $f_{\boldsymbol{B}}(x_B, y_B)$ in red color.

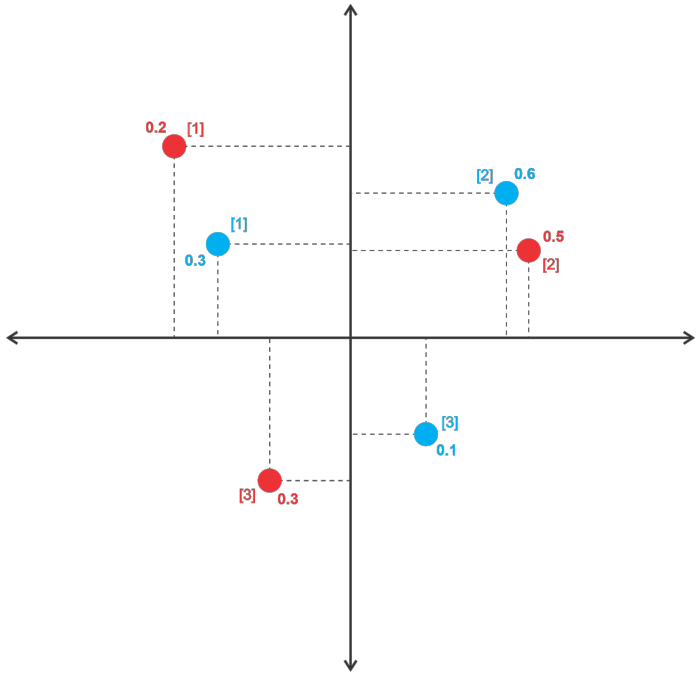


Figure 4.3: Two dimensional scatter view of $f_{\boldsymbol{A}}(x_A, y_A)$ and $f_{\boldsymbol{B}}(x_B, y_B)$.

A possible transport plan to move the content of $f_{\boldsymbol{A}}(x_A, y_A)$ to $f_{\boldsymbol{B}}(x_B, y_B)$ is illustrated in Figure 4.4. When moving some quantity of probability (mass) from a point from the domain of $f_{\boldsymbol{A}}(x_A, y_A)$ into a point from the domain of $f_{\boldsymbol{B}}(x_B, y_B)$, there is also a cost embedded as a distance. Another possible transport plan is illustrated in Figure 4.5.

There are many other possible options for transport plan to move the content from $f_{\boldsymbol{A}}(x_A, y_A)$ to $f_{\boldsymbol{B}}(x_B, y_B)$. They are indeed in this case joined mass functions $f_{\boldsymbol{A},\boldsymbol{B}}(x_A, x_B, y_A, y_B)$, whose marginals are $f_{\boldsymbol{A}}(x_A, y_A)$ and $f_{\boldsymbol{B}}(x_B, y_B)$. However, which one of them is the optimal transport plan $f^*_{\boldsymbol{A},\boldsymbol{B}}(x_A, x_B, y_A, y_B)$ responsible to minimize the cost associated to the transferring?

For now, the L_p-Wasserstein $\mathcal{W}_p(\bullet, \bullet)$ is defined as the total cost given by the optimal transport plan between two pmf (discrete case) or two pdf (continuous case), in which the cost function is a p-norm. Therefore, for the

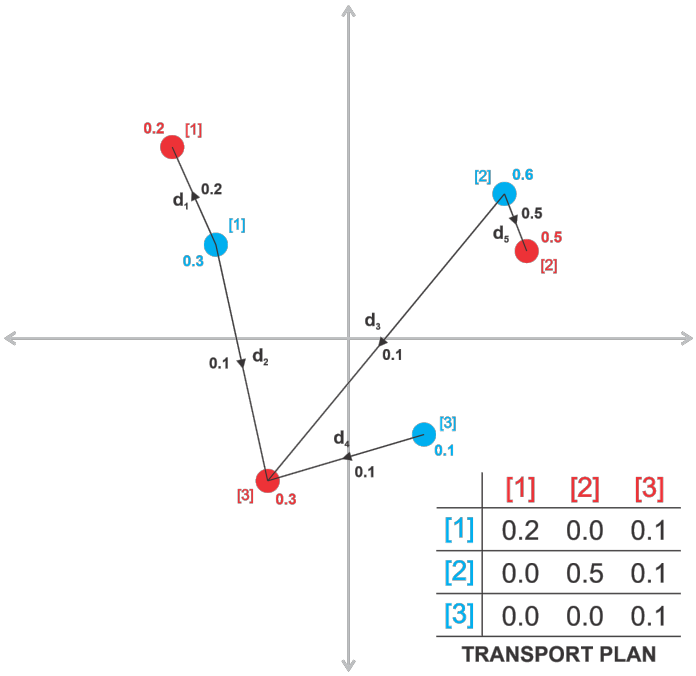


Figure 4.4: First option of transport plan to move $f_{\boldsymbol{A}}(x_A, y_A)$ to $f_{\boldsymbol{B}}(x_B, y_B)$.

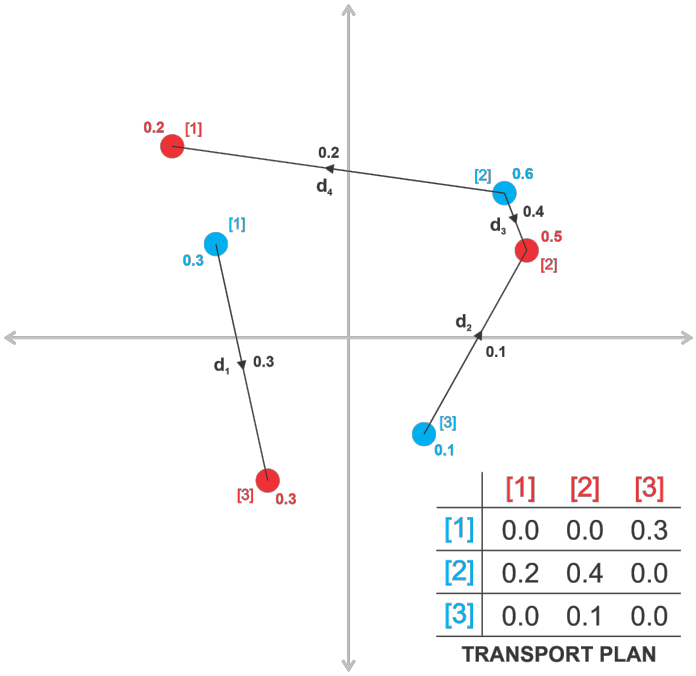


Figure 4.5: Second option of transport plan to move $f_{\boldsymbol{A}}(x_A, y_A)$ to $f_{\boldsymbol{B}}(x_B, y_B)$.

continuous case that is faced in this work, the L_p-Wasserstein between the posterior $f_\alpha\left(p_1 \mid \hat{d}\right)$ and the prior $f_\beta(p_2)$ distributions is

$$\mathcal{W}_p(f_\alpha, f_\beta) = \left[\inf_{f_{\alpha,\beta} \in \Gamma\left(f_{P|D}, f_P\right)} \int_{\Omega\times\Omega} d(p_1, p_2)^p \, df_{\alpha,\beta}(p_1, p_2)\right]^{1/p}. \tag{4-19}$$

It is not a simple task to find the optimal transport plan. However,

there is a closed solution for the L_p-Wasserstein in case both the probability distribution are continuous and one-dimensional [27, 28],

$$\mathcal{W}_p\left(f_\alpha, f_\beta\right) = \left[\int_0^1 \left|F_\alpha^{-1}\left(z\right) - F_\beta^{-1}\left(z\right)\right|^p dz\right]^{1/p}, \tag{4-20}$$

in which, $F_\alpha\left(p_1 \mid \hat{d}\right)$ and $F_\beta\left(p_2\right)$ are the cdf from the posterior and prior distributions.

4.3
Comparison between the two strategies to make the bayesian inference

Given that in this work two strategies to use data from the BGW process to estimate parameters from the contagion random variable are introduced, we would like to understand what are the advantages and disadvantages of each strategy. More than this, we would like to compare them in terms of computational cost and rate of convergence when making the posterior updating. Some of the questions that arise are: is it easier to estimate the parameters using different data from the same generation or from a single realization of the ramification tree? Is it faster (in terms of computational cost) to estimate the parameters using different data from the same generation or from a single realization of the ramification tree observed over a certain number of generations? If we use the first strategy, that is, different realizations from the same generation, what would be the influence of the generation's number in the rate of convergence? In other words, what leads to faster convergence, to use data from an earlier generation or a more advanced?

First, we need to define a stopping criteria for the updating to the bayesian parametric estimation. This is done here using the L_2-Wasserstein distance [29]. If the subsequent information does not change a prior distribution into a posterior one in a value of L_2-Wasserstein greater than $\xi_{\mathcal{W}_2} = 1 \times 10^{-3}$, the updating process is interrupted. An example is showed in the following.

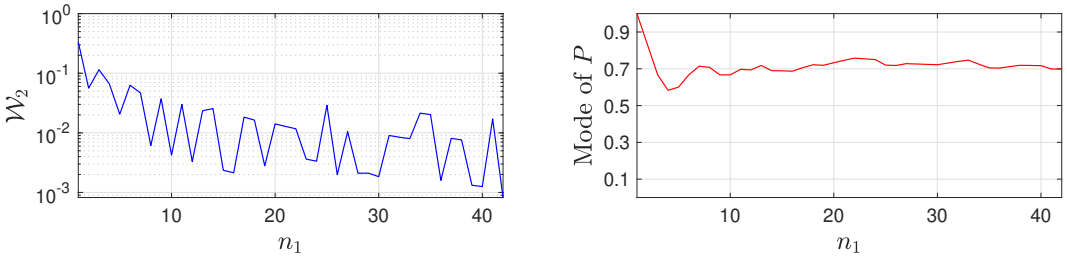


Figure 4.6: Example of convergence for bayesian parametric inference.

Figure 4.6 shows a convergence analysis for data coming from X_1 where the contagion random variable is modeled as $C \sim Binomial\,(3, p)$. Based on the L_2-Wasserstein stopping criteria, the number of updates necessary was $n_1 = 42$, as showed in blue line. For every posterior distribution for the random variable P, its mode was tracked and it was seen in red line that as long as data arrived, the best estimate for P tended to a value of $p \approx 0.7$.

There is a fundamental aspect to be considered of this analysis. A specific data sequence $\{x_1\}_k = \{3, 2, 1, 1, 2, 3, \ldots\}$ is the one responsible for the number of updates n_1, the L_2-Wasserstein path and the mode of P path visualized. If another sequence was observed, different results of them would happen. Therefore, the number of updates is indeed a random variable, here named N, conditional on the data sequence random variable $\{X_1\}_k$. The other features are conditional stochastic processes.

In order to deal with this uncertainty scenario, Monte Carlo simulations are done in order to get the mass function of the random variable N. The stopping criteria in this case is based on the stability of the cumulative mean of the outcomes n_k.

The following analysis are divided in three main general cases: parametric estimation for the contagion modeled as $Binomial\,(3, p)$, $Geometric{-}0\,(p)$ and $Poisson\,(\lambda)$ families. For the first and second of them, a parametric study on artificial data for the cases $p = [0.3, 0.5, 0.7]$ is performed and, for the last one, a parametric study on $\lambda = [0.9, 1.5, 2.1]$ is done. For each of them, we compare the case of data coming from the same 1st up to 4th generations and from a certain number of generations of the same realization of a BGW process.

4.3.1
Contagion modeled as a Binomial distribution

First, as a result of the Monte Carlo simulations, the number of experiments to study the stochastic effect of data sequence on the bayesian parametric estimation was 200 for $C \sim Binomial\,(3, p)$. All the scenarios studied for this random variable family were done with Markov chain technique from Eq. (4-8), because it was previously show in the last chapter that it is faster than the pgf one. The cumulative mean for the outcomes of the number of updates N is showed in Figure 4.7. The initial parameter to generate data to study the case $C \sim Binomial\,(3, p)$ is $p = 0.3$.

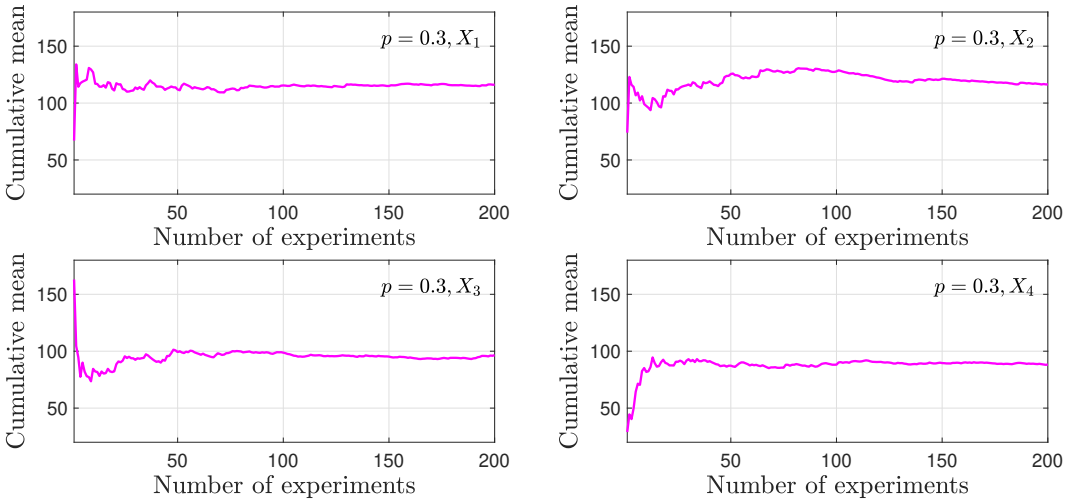


Figure 4.7: Convergence of the cumulative mean of the number of updates for the case of data coming from the same generation when $C \sim Binomial\,(3, 0.3)$.

Despite the fact that the greater the number of the generation is, its support is larger, the cumulative mean comparison shows that the number of updates to achieve the convergence based on L_2-Wasserstein gets slightly fewer in mean. Next, it is presented in Figure 4.8 the mode of P paths from each experiment done.

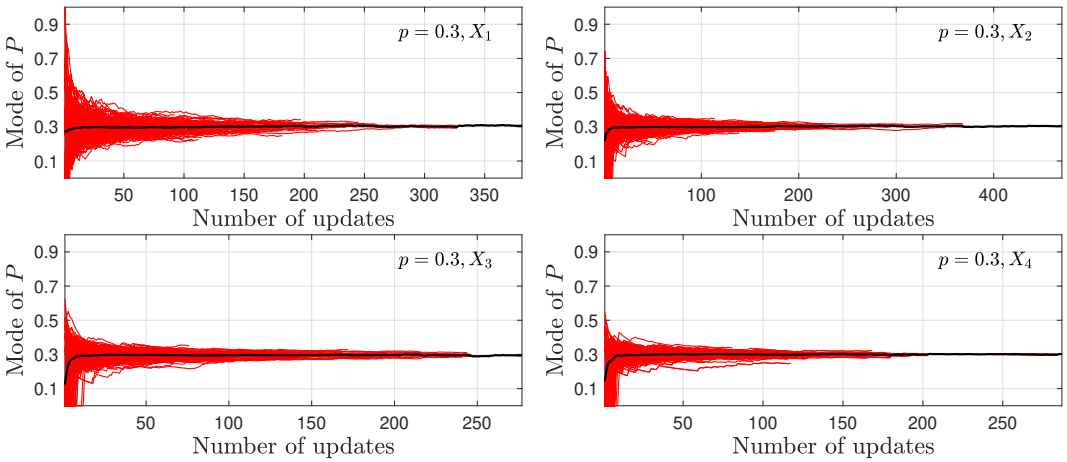


Figure 4.8: Mode of P paths for the case of data coming from the same generation when $C \sim Binomial\,(3, 0.3)$. The black curves are the mean paths.

From the mode of P paths indicated in Figure 4.8, it is observed that, the greater the number of the generation is, the deviations of modes in general decrease for the same number of udpates. Some values of the support are not even once considered as a best estimate for P, which comes from the fact that observed data from further generations tends to concentrate at small intervals in comparison to its total support. This is a consequence of the pmf of them, which affects directly the likelihoods chosen to update priors into posteriors distributions.

Next, the cumulative mean for the outcomes of the number of updates N from data generated by $C \sim Binomial\,(3, 0.5)$ is presented in Figure 4.9.

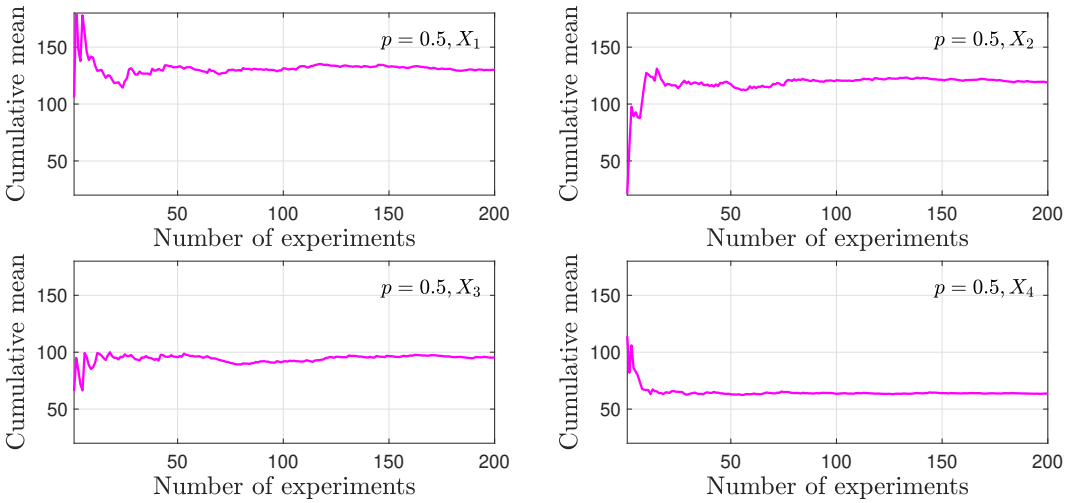


Figure 4.9: Convergence of the cumulative mean of the number of updates for the case of data coming from the same generation when $C \sim Binomial\,(3, 0.5)$.

Likewise the last scenario, the greater the number of the generation is, the cumulative mean indicates that the number of updates to achieve the convergence criteria of L_2-Wasserstein is lower. This time, the 4th generation had a larger fluctuation in its value in comparison with the case $p = 0.3$. The respective mode of P paths per experiment are displayed in Figure 4.10.

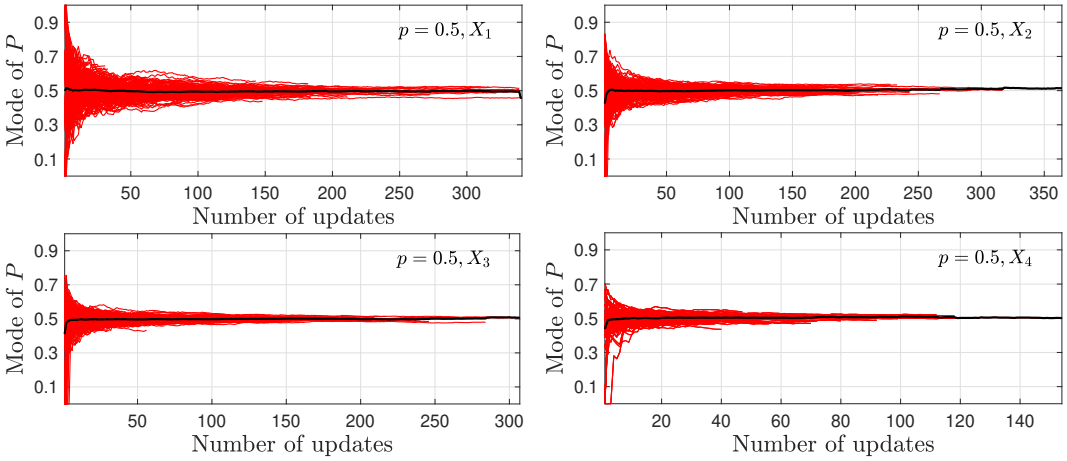


Figure 4.10: Mode of P paths for the case of data coming from the same generation when $C \sim Binomial\,(3, 0.5)$. The black curves are the mean paths.

The mode of P paths have the same behavior as for $p = 0.3$. However, for the 1st generation now, the modes are somehow symetrically distributed around the mean mode of P path. This particularity for X_1 is related to its symmetrical pmf alongside the support.

Finally, the Monte Carlo simulations for the case of $p = 0.7$ is depicted now in Figure 4.11.

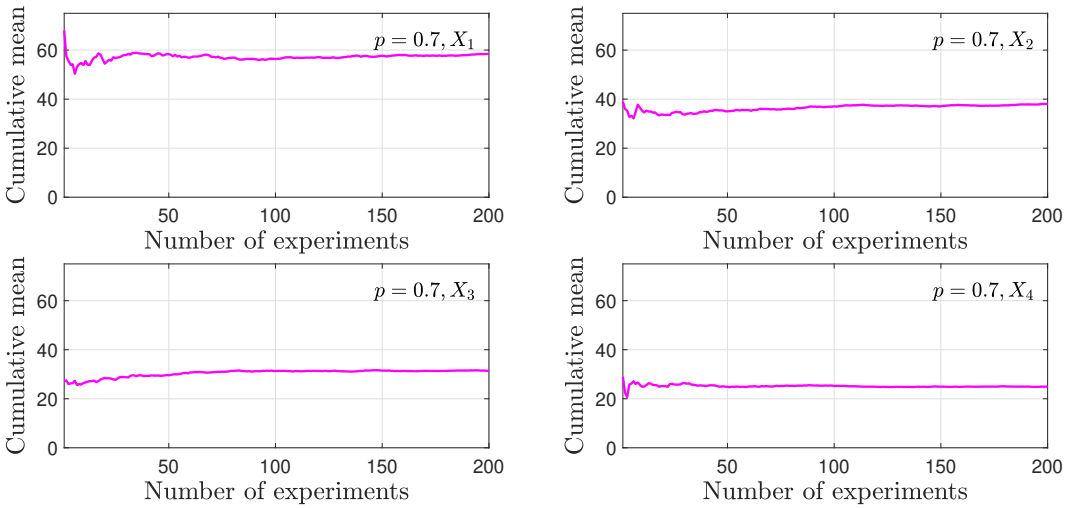


Figure 4.11: Convergence of the cumulative mean of the number of updates for the case of data coming from the same generation when $C \sim Binomial\,(3, 0.7)$.

More infectious diseases, i.e., greater values of p for the contagion random variable modeled as a Binomial family need lower number of updates to achieve the convergence criteria of L_2-Wasserstein. The same is true the greater the number of the generation is in any of the three scenarios in here presented. In the following, the mode of P paths are displayed in Figure 4.12. They follow the same behavior presented in other scenarios.

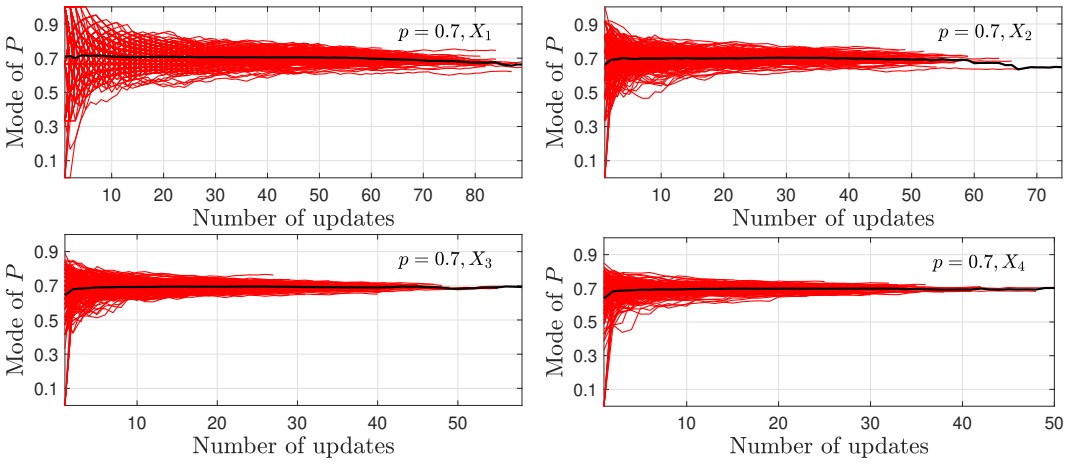


Figure 4.12: Mode of P paths for the case of data coming from the same generation when $C \sim Binomial\,(3, 0.7)$. The black curves are the mean paths.

As exposed above, the greater the number of the generation is, the required number of updates to achieve the convergence criteria of L_2-Wasserstein decreases in mean. However, it is important to highlight that the runtime spent

to work with these likelihoods increases a lot, due their analytical complexity gain.

A better comprehension on why data from further generations are more useful to achieve the L_2-Wasserstein convergence is experienced in the presence of the likelihood functions view. Figure 4.13 displays all possible normalized likelihoods functions coming from the 1st up to the 4th generation.

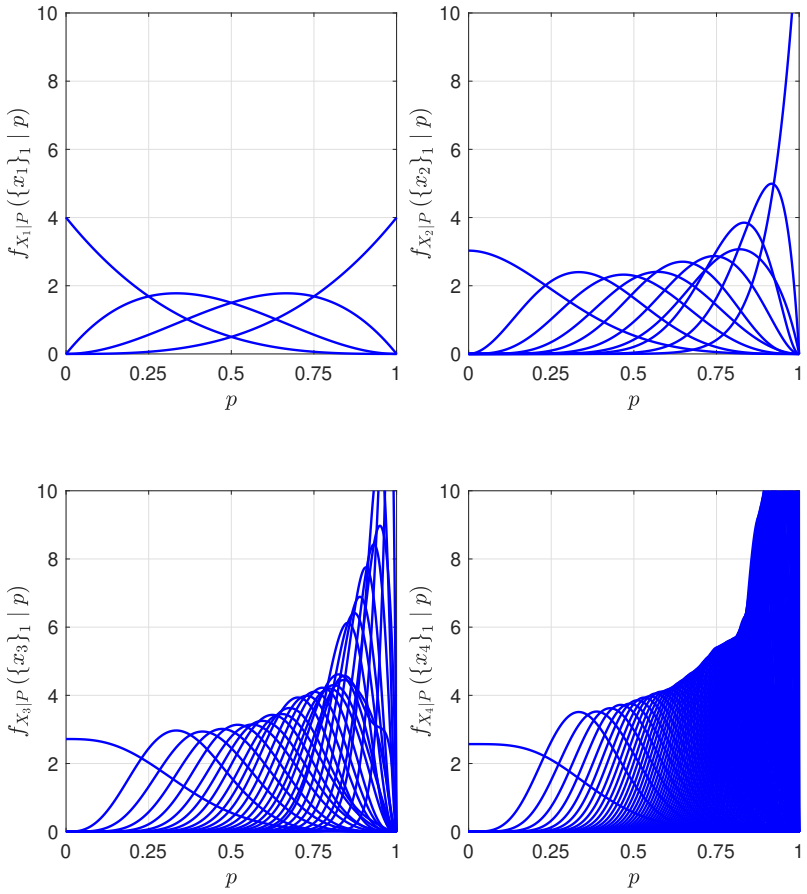


Figure 4.13: Normalized likelihood functions coming from the 1st up to 4th generation for $C \sim Binomial\,(3, p)$.

The greater the number of the generation is, the normalized likelihood functions have a sharper form, so their content gets more concentrated in shorter intervals. Moreover, it is possible to visualize why for higher values of p, the parametric estimation is easier. The function density gets higher and the functions get more specified when closer to end of p.

Now, each sequence of data is composed for elements related to different subsequent generations from the same BGW process. The cumulative means of each realization of N, i.e., number of updates in bayesian convergence, is showed in Figure 4.14.

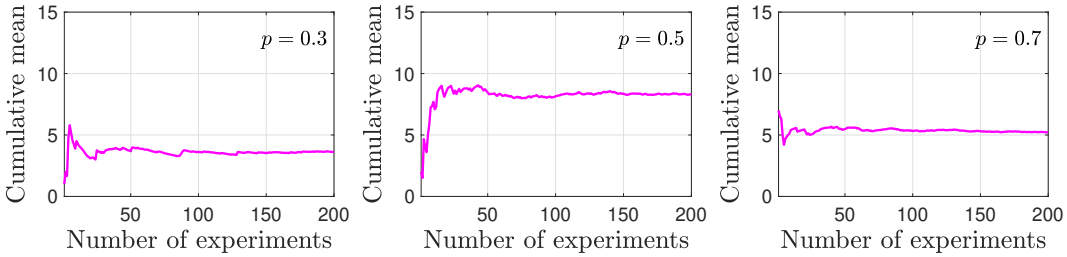


Figure 4.14: Convergence of the cumulative mean of the number of updates for the case of data coming from different subsequent generations when $C \sim Binomial\,(3, 0.3)$, $C \sim Binomial\,(3, 0.5)$ and $C \sim Binomial\,(3, 0.7)$ respectively.

From the Monte Carlo simulations analysis, a sample with 200 hundred experiments is also enough to enlighten the stochastic effect. The cumulative means of each realization of N, i.e., number of updates in bayesian convergence, is showed in Figure 4.14. The likelihood functions were built according to Eq. (4-18). Before discussing these results, the mode of P paths are presented in Figure 4.15.

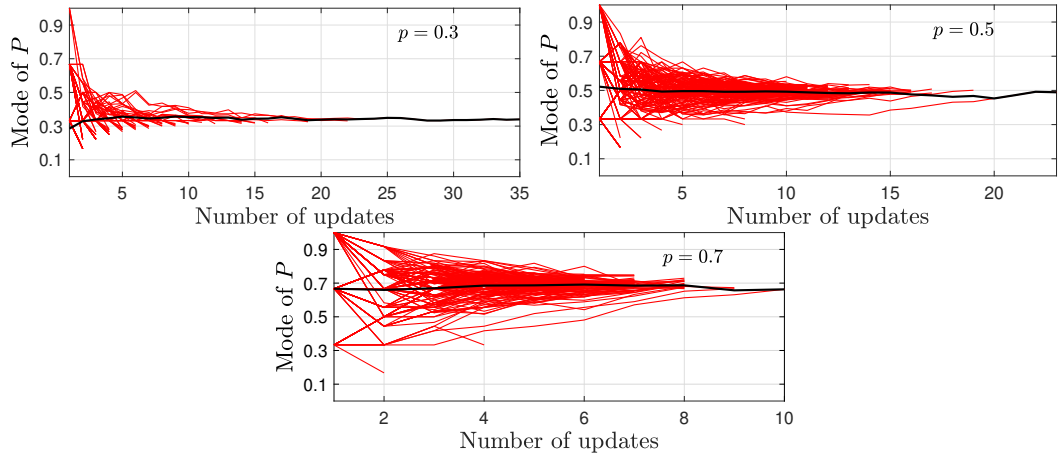


Figure 4.15: Mode of P paths for the case of data coming from different subsequent generations for the same BGW process when $C \sim Binomial\,(3, 0.3)$, $C \sim Binomial\,(3, 0.5)$ and $C \sim Binomial\,(3, 0.7)$ respectively. The black curves are the mean paths.

Different from the scenario where data comes from the same generation, the case $p = 0.3$ now shows fewer updates in mean are required to update the bayesian parametric inference in comparison with $p = 0.5$ and $p = 0.7$. This information itself is tendentious. What is happening indeed is that for $p = 0.3$, the BGW processes has a higher chance to extinct very soon. The probability of extinction is a more critical feature on this way of making bayesian parametric estimation. Therefore, the number of updates are low, because the sequences have in general extinct early.

On the other hand, there is a very important thing to say about the results from $p = 0.7$. It is in fact the scenario where the data incorporated through the likelihood most transform the prior functions into posterior in a positive way of best estimating P as a mode. This is seen through the reduced variances for few updates in the mode of P paths. However, the number of updates were not interrupted because of the L_2-Wasserstein convergence criteria. A floating-point arithmetic precision issue was instead the main responsible for it. The posteriors obtained are still in each update highly modified and quickly get a sharply shape.

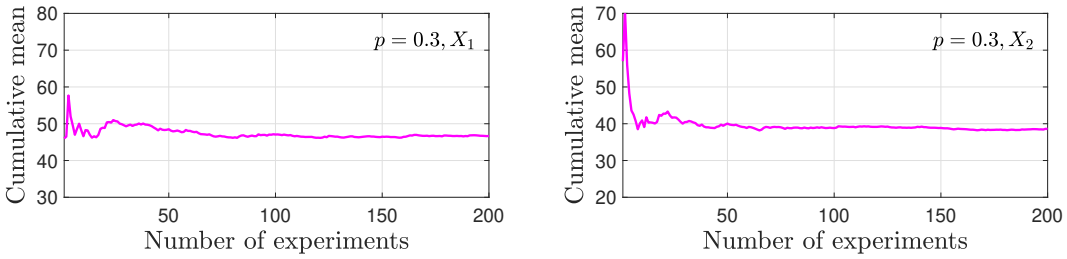
4.3.2
Contagion modeled as a Geometric-0 distribution

It is important to highlight that there is an upper limit imposed in data acquisition to avoid calculating hardly feasible higher-order derivatives for the Geometric-0 analysis. Despite the generation, the maximum possible outcome is 60. As a consequence, this boundary leads to intervals covering the approximate cumulative probabilities presented in Table 4.1.

p	**t-th Generation**			
	1	**2**	**3**	**4**
0.3	100%	99.95%	96.29%	-
0.5	100%	100%	100%	100%
0.7	100%	100%	100%	100%

Table 4.1: Approximate cumulative probabilities for $C \sim Geometric - 0\,(p)$ when data acquisition is truncated at 60 members infected.

Unlike the previous random variable family, the likelihoods in here cannot be found according to the Markov technique from Eq. (4-8). We use the pgf approach from Eq. (4-6) instead. Figure 4.16 shows the study related to the convergence of the number of updates N for $C \sim Geometric - 0\,(0.3)$.



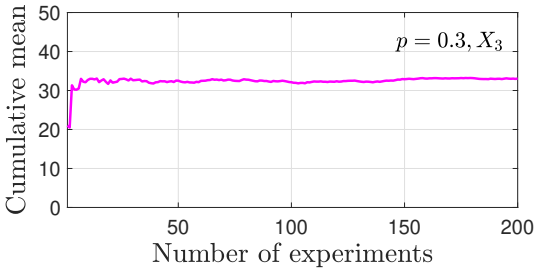


Figure 4.16: Convergence of the cumulative mean of the number of updates for the case of data coming from the same generation when $C \sim Geometric - 0\,(0.3)$.

Similar to the Binomial behavior, the greater the number of the generation is, the number of updates required to comply with the L_2-Wasserstein convergence criteria decreases in mean. The fourth generation is not investigated this time, due its extensive runtime to be performed. Next, Figure 4.17 displays the mode of P paths from each sequence of data obtained.

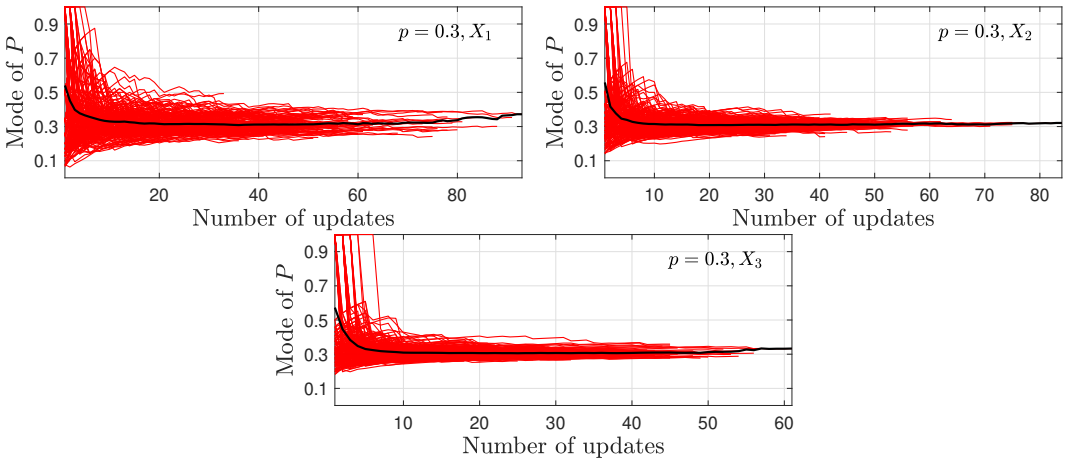


Figure 4.17: Mode of P paths for the case of data coming from the same generation when $C \sim Geometric - 0\,(0.3)$. The black curves are the mean paths.

Some of the sequences lead initially to best estimates of the random variable P as the value one. This is a consequence of the high probabilities associated to the state 0 in the pmfs for the Geometric-0 random variable family. However, with few updates this wrong estimate turns easily to values closer to the real one $p = 0.3$.

This time, the case when data is generated according to $p = 0.5$ is investigated in Figure 4.18. It is a case where the expected behavior according to the previous experiences changes. It starts with a lower value in mean number for N when data comes from the 1st generation. Then it increases the next two generations. Finally, when evaluating the 4th generation, the mean value returns to decrease. Figure 4.19 presents the mode of P paths.

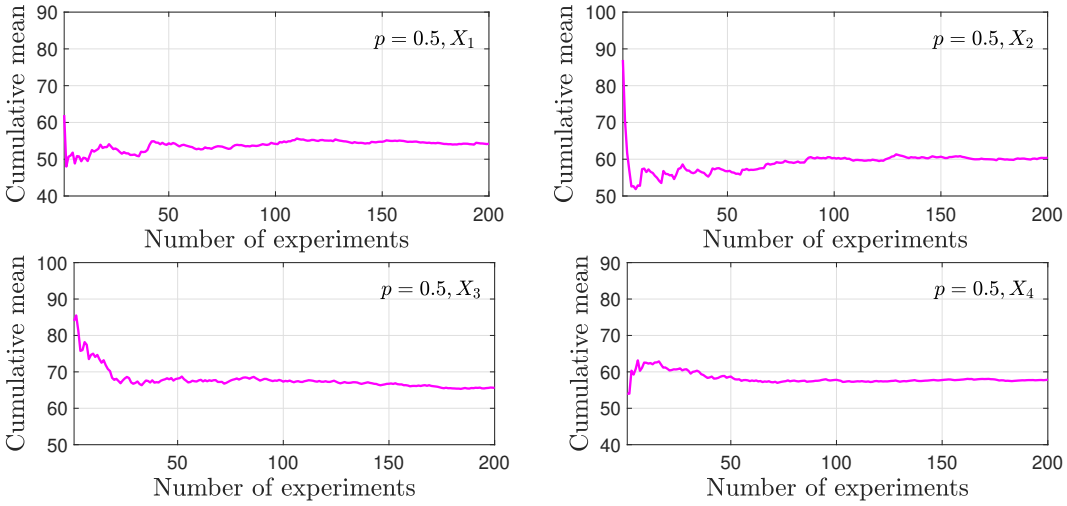


Figure 4.18: Convergence of the cumulative mean of the number of updates for the case of data coming from the same generation when $C \sim Geometric - 0\,(0.5)$.

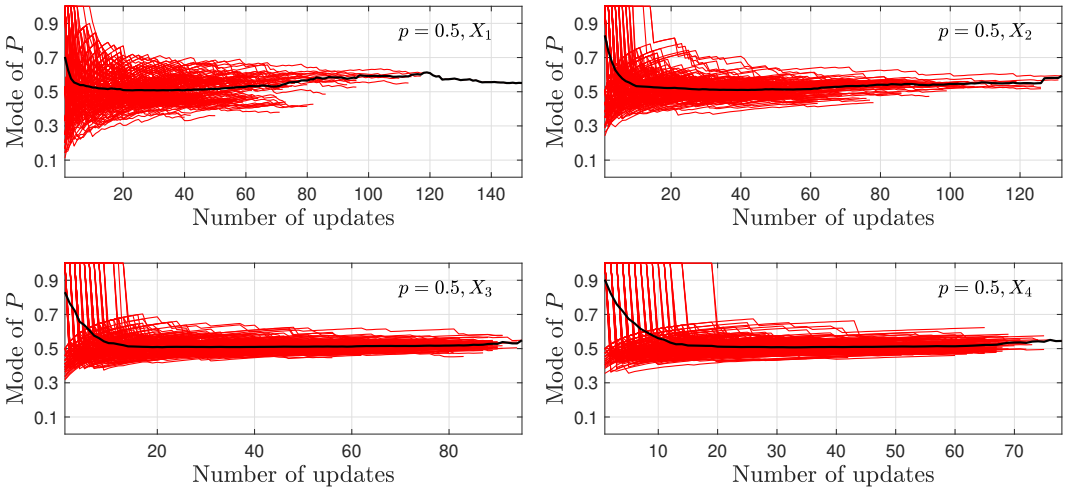
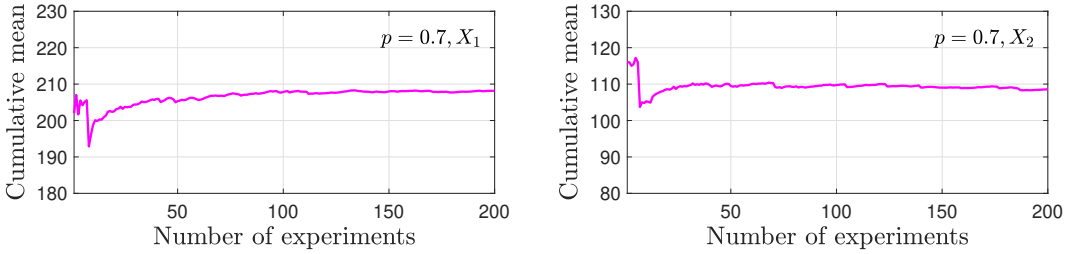


Figure 4.19: Mode of P paths for the case of data coming from the same generation when $C \sim Geometric - 0\,(0.5)$. The black curves are the mean paths.

The last scenario when data comes from the same generation for C $Geometric - 0\,(p)$ is the one with $p = 0.7$, which represents the least infectious disease in comparison with previous cases studied this section. Its number of updates' convergence based on the sequential data is presented in Figure 4.20.



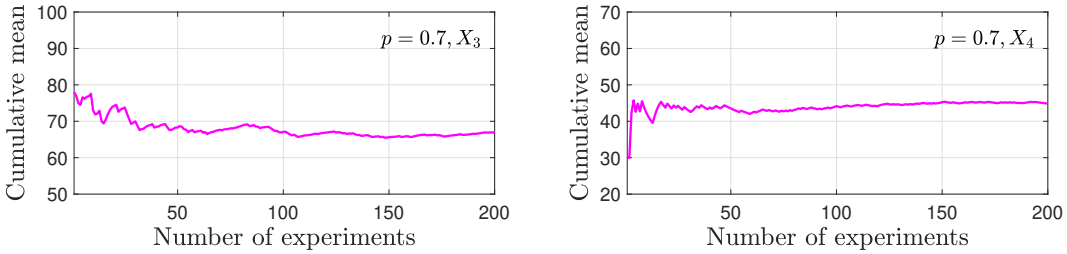


Figure 4.20: Convergence of the cumulative mean of the number of updates for the case of data coming from the same generation when $C \sim Geometric - 0\,(0.7)$.

The number of updates required are the highest in mean per generation when compared to other values of p. The difference between the third and fourth generation now is minimal. Finally, the mode of P paths are displayed in Figure 4.21.

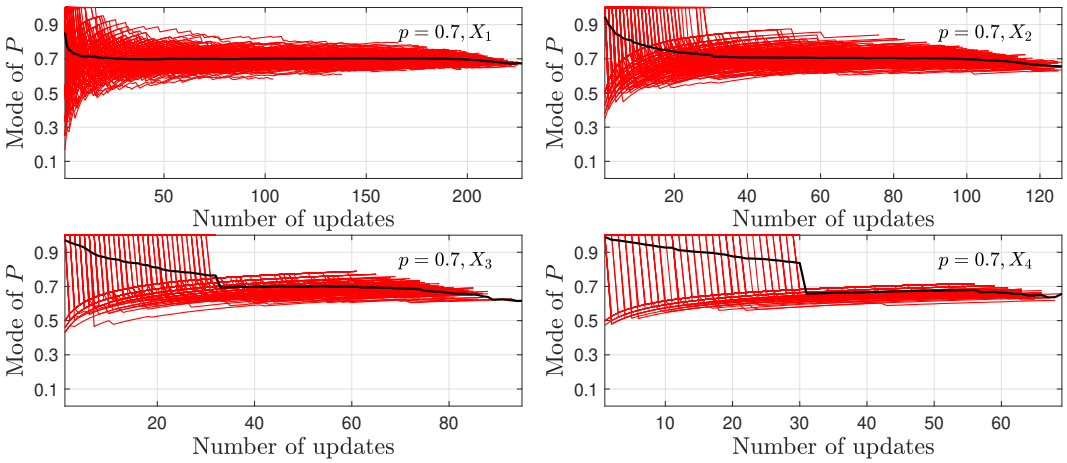


Figure 4.21: Mode of P paths for the case of data coming from the same generation when $C \sim Geometric - 0\,(0.7)$. The black curves are the mean paths.

It is important to emphasize that for the Geometric-0 family, the highest values of p are related to less infectious diseases. This is the reason why the best estimates in Figure 4.21 are many times mapped into $p = 1.0$, which means that the related data sequences begin in these situations with no infections and the following data are also this outcome. Some mode of P paths converge indeed in this situation for $p = 1.0$, which is a wrong conclusion. This is a consequence of the tolerance 0.001 imposed in the updating process for the L_2-Wasserstein criteria.

In order to get a better comprehension on the likelihood building according to the data received, Figure 4.22 displays the normalized likelihood functions according to each possible state up to 60 the data could have taken from the four generations studied. Notice that for the first generation, the

lower outcomes of data are the ones with flatten curves for likelihood function and the higher outcomes map with more insurance that the parameter p is closer to zero. The greater the number of the generation is, the likelihoods related to higher outcomes move away the belief in the most lower values for p. Moreover, the likelihoods associated to lower outcomes gets sharper and more concentrated around some specific value. The only exception is for the case of none infections, in which the likelihood functions seems to change into a stationary flatten configuration the greater the number of the generation is.

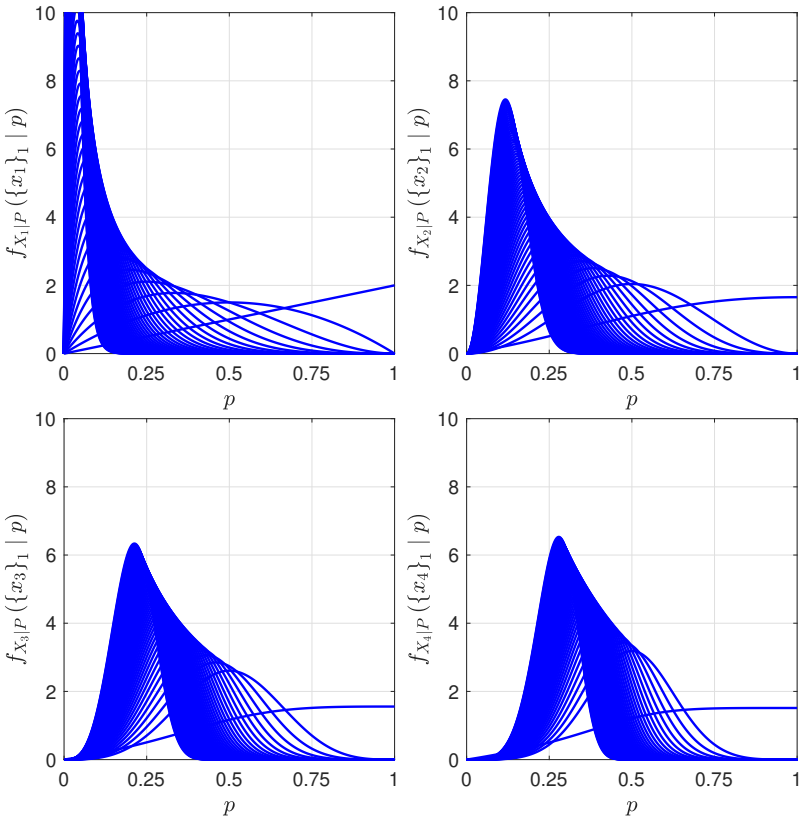


Figure 4.22: Normalized likelihood functions coming from the 1st up to 4th generation for $C \sim Geometric - 0\,(p)$.

Now, each data sequence comes from the realization of subsequent generations that belongs to the same ramification tree. Figure 4.23 shows the convergence analysis of the cumulative sample mean of N when $C \sim Geometric - 0\,(0.3)$, $C \sim Geometric - 0\,(0.5)$ and $C \sim Geometric - 0\,(2.1)$.

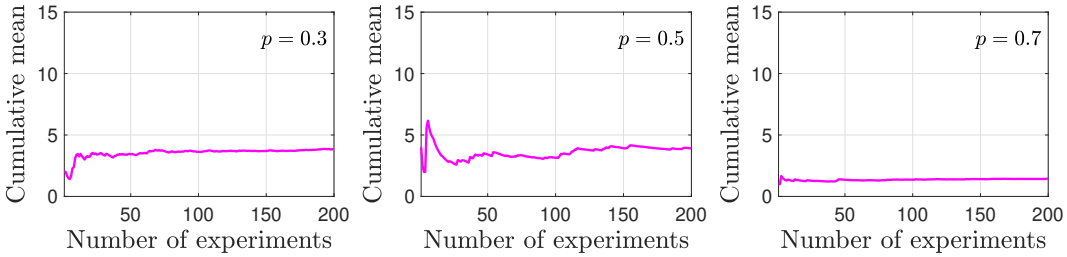


Figure 4.23: Convergence of the cumulative mean of the number of updates for the case of data coming from different subsequent generations when $C \sim Geometric-0\,(0.3)$, $C \sim Geometric-0\,(0.5)$ and $C \sim Geometric-0\,(2.1)$ respectively.

The above study gives us again an incomplete information about what is indeed happening when data is used this way. It actually leads to a misleading interpretation. According to the results seen in Fig 4.23, one can presume that this methodology independently of the value of p requires way fewer updates to achieve the L_2-Wasserstein criteria than the other way of using data. This comparison is not even fair. Before discussing this point, Figure 4.24 presents the mode of P paths, which contribute to analyze this situation.

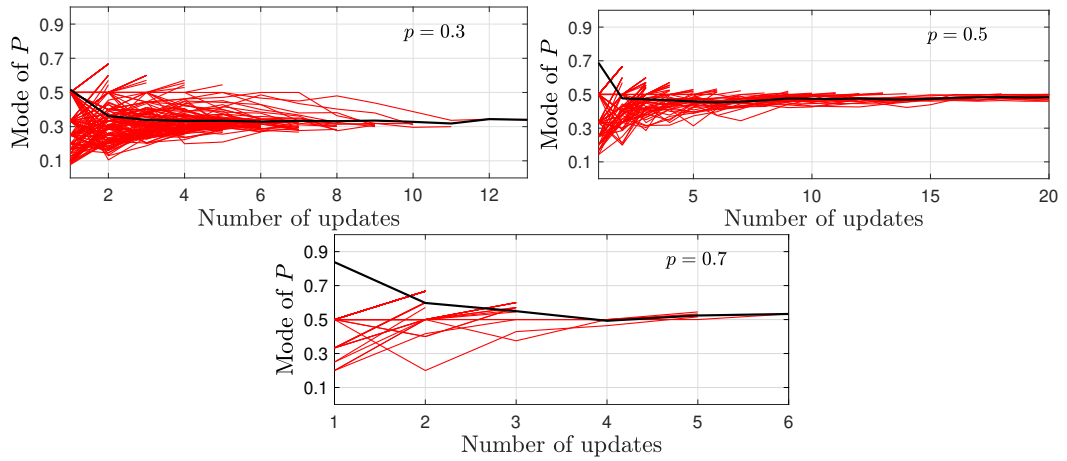


Figure 4.24: Mode of P paths for the case of data coming from different subsequent generations for the same BGW process when $C \sim Geometric - 0\,(0.3)$, $C \sim Geometric - 0\,(0.5)$ and $C \sim Geometric - 0\,(2.1)$ respectively.

Parameter $p = 0.3$ is the one when working with $C \sim Geometric - 0\,(p)$ that in here leads to the most contagious spreading. The probability of extinction is the lowest when compared to the other two cases studied for this probabilistic model. None of the 200 experiments had its bayesian updating ended because L_2-Wasserstein was achieved. Most of them stopped indeed due floating-point arithmetic problems. The other ones were ramification trees that extinct early. The former situation indeed leads to great results for parametric inference. However, the latest case does not. The critical identification happens

to $p = 0.7$. All of the realizations of the BGW process extincts in earlier generations. As a consequence, it does not provide substantial information to work with and so the parametric inference fails.

4.3.3
Contagion modeled as a Poisson distribution

Again there is an upper limit imposed in data acquisition to avoid calculating hardly feasible higher-order derivatives for the Poisson analysis. Despite the generation, the maximum possible outcome is 60. As a consequence, this boundary leads to intervals covering the approximate cumulative probabilities presented in Table 4.2.

p	**t-th Generation**			
	1	**2**	**3**	**4**
0.9	100%	100%	100%	100%
1.5	100%	100%	100%	100%
2.1	100%	100%	100%	-

Table 4.2: Approximate cumulative probabilities for $C \sim Poisson\,(\lambda)$ when data acquisition is truncated at 60 members infected.

The last random variable family studied for the parametric inference problem is the $Poisson\,(\lambda)$ one. We begin the analysis when data comes from the same generation for the contagion modeled as $C \sim Poisson\,(0.9)$. The convergence of its number of updates random variable N is displayed in Figure 4.25

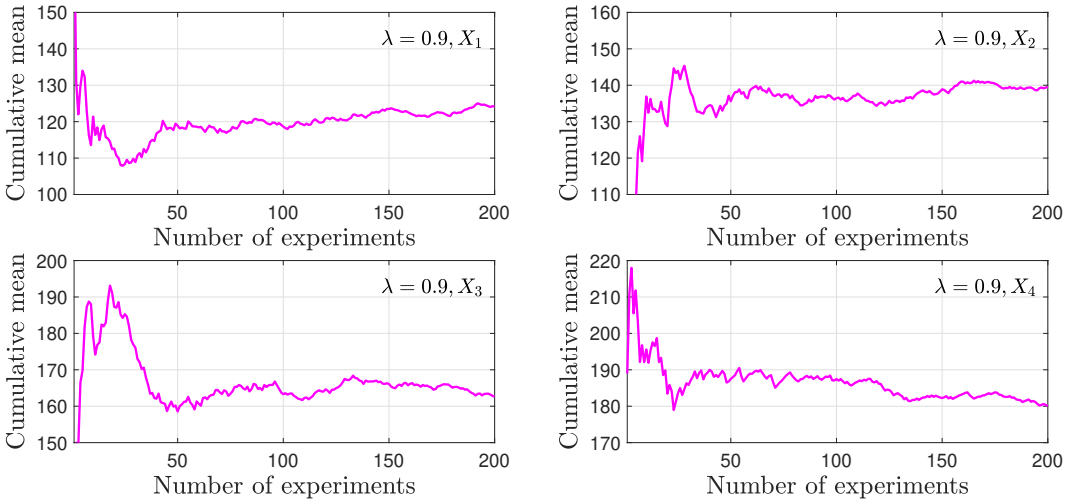


Figure 4.25: Convergence of the cumulative mean of the number of updates for the case of data coming from the same generation when $C \sim Poisson\,(0.9)$.

As we can see from Figure 4.25, the greater the number of the generation is, the sample mean value for the random variable N presented the very opposite behavior that was in the previous cases seen. Now, the sample mean value has increased. This isolated information can lead to a quite tricky conclusion that when data comes from most previous generation, it would require in general few data to infer properly the best estimate for the parameter λ. Let's take a look at Figure 4.26.

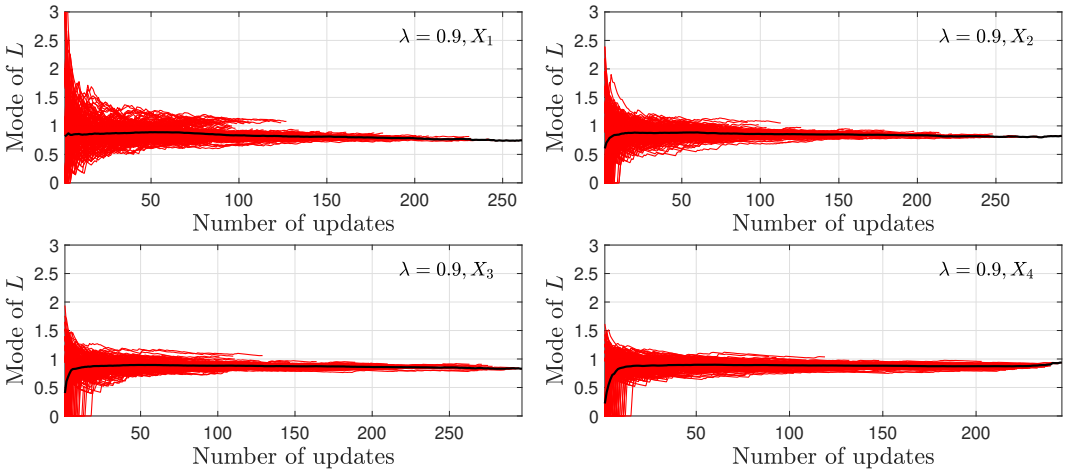
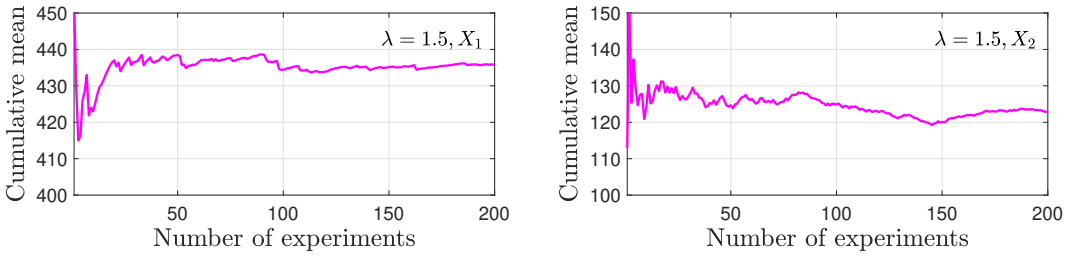


Figure 4.26: Mode of P paths for the case of data coming from the same generation when $C \sim Poisson\,(0.9)$. The black curves are the mean paths.

When looking at the mode of L paths from the 1st generation, it is possible to see that some data sequences lead the convergence in L_2-Wasserstein for a value of λ that is not that close to the real one in comparison with the other results from further generations. The graphic related to X_4 from Figure 4.26 shows that in this scenario most of the data sequences result on a final posterior distribution whose mode is closer to $\lambda = 0.9$. Therefore, the major factor that leads to the results of N in Figure 4.25 was indeed the tolerance.

Data is now generated according to $\lambda = 1.5$, the convergence study of the random variable N related to the number of updates required is displayed in Figure 4.27.



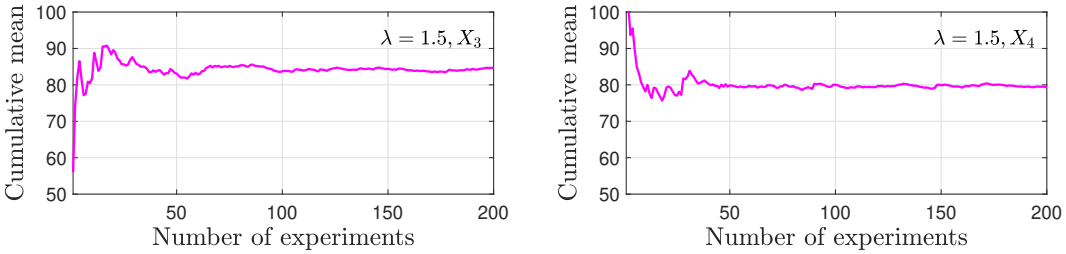


Figure 4.27: Convergence of the cumulative mean of the number of updates for the case of data coming from the same generation when $C \sim Poisson\,(1.5)$.

The general behavior now returns to the usual one. The greater the number of the generation is, the required number of updates to achieve the L_2-Wasserstein convergence decreases in mean.

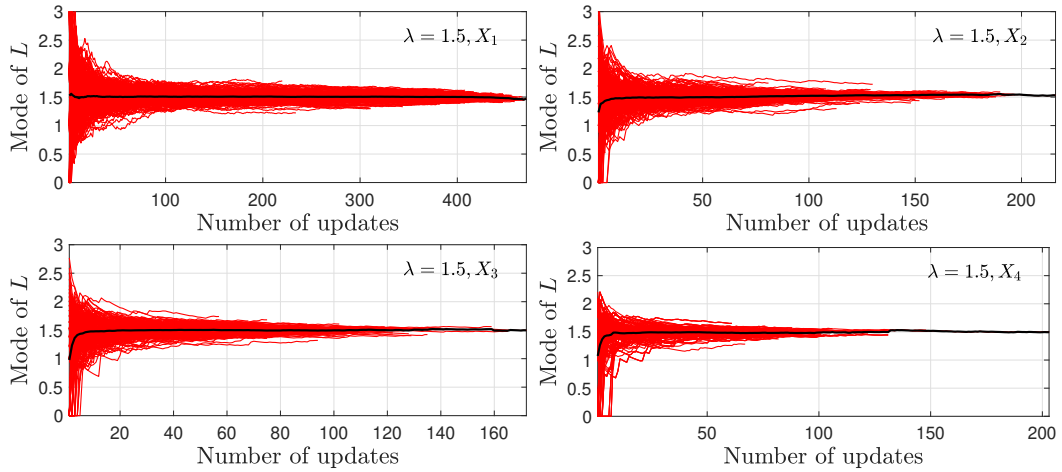
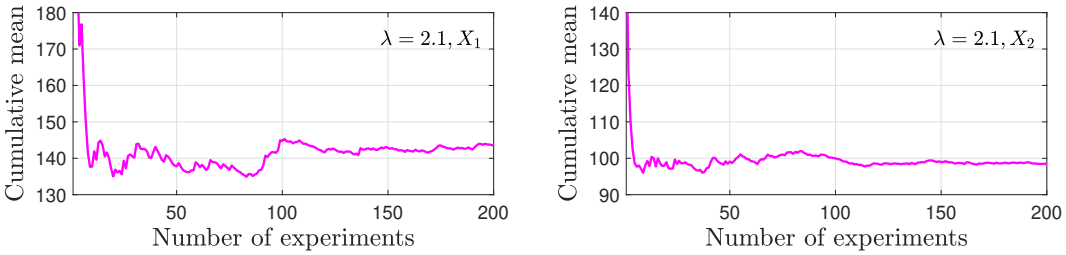


Figure 4.28: Mode of P paths for the case of data coming from the same generation when $C \sim Poisson\,(1.5)$. The black curves are the mean paths.

Figure 4.28 shows the mode of L paths. The results reassure that the greater the number of the generation is, the data provides likelihood functions whose mapping capacity towards the real value of the parameter λ is more efficient. Another visible aspect is that the interval of modes of all the resulting posteriors when the first data arrived gets shorter the greater the number of the generation is.



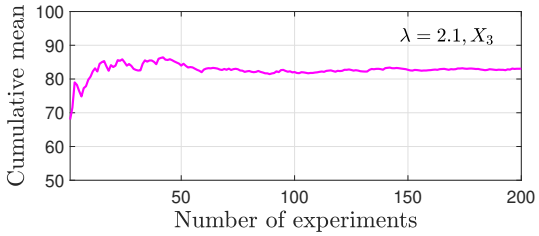


Figure 4.29: Convergence of the cumulative mean of the number of updates for the case of data coming from the same generation when $C \sim Poisson\,(2.1)$.

The last case study is the one for $C \sim Poisson\,(2.1)$. The 4th generation was not possible to analyse this time, due the symbolic gain complexity of its probability generating functions, resulting in non-feasible runtimes. The convergence study for the random variable N is presented in Figure 4.29. Again, the greater the number of the generation is, the number of updates to achieve the L_2-Wasserstein criteria decreases in mean. Figure 4.30 shows the mode of L paths for each one of the two hundred experiments done. The same behavior is seen. The greater the number of the generation is, the modes during the 200 paths are closer to the real value of λ with less updates.

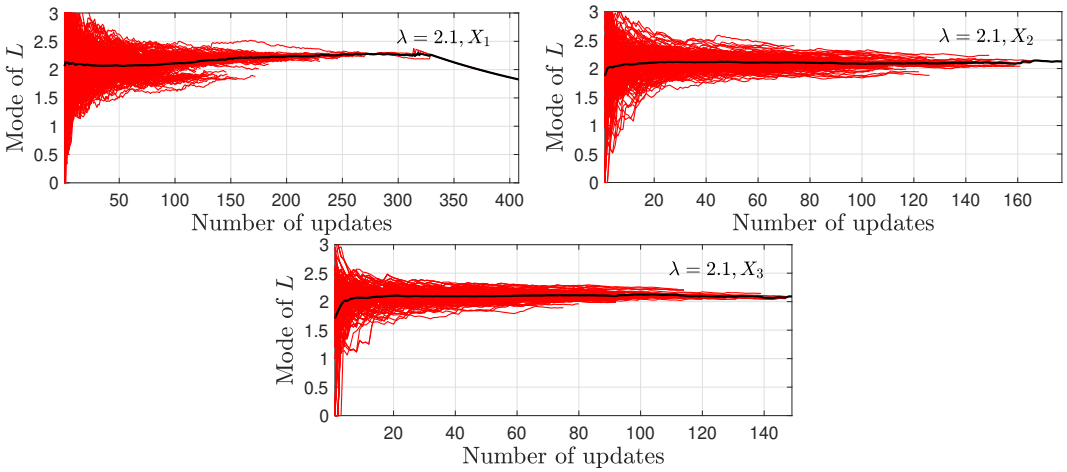


Figure 4.30: Mode of P paths for the case of data coming from the same generation when $C \sim Poisson\,(2.1)$. The black curves are the mean paths.

Now, the overall aspect from the evolution of the likelihoods per generation is discussed. Figure 4.31 shows the normalized likelihood functions related to each possible outcome up to the 40th one when $C \sim Poisson\,(\lambda)$.

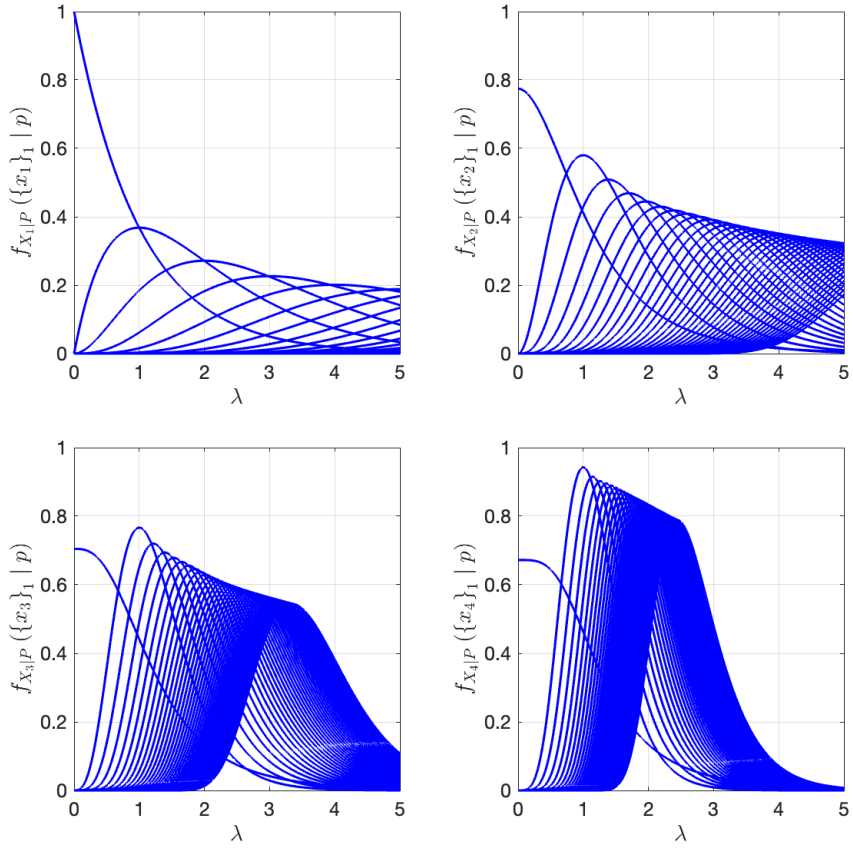


Figure 4.31: Normalized likelihood functions coming from the 1st up to 4th generation for $C \sim Poisson\,(\lambda)$.

The greater the number of the generation is, it is possible to visualize that the normalized likelihoods get also sharpen and they tend to concentrate their significant content in lower values of λ. The mapping capacity towards closer intervals for the underlying possible values for the parameter increases. The only exceptions are the likelihoods related to no infections. Their evolution over generations show that the shape of functions gets more flatten, which is in agreement with the previous results seen from $C \sim Binomial\,(3, p)$ and $C \sim Geometric - 0\,(p)$.

Now, the second way to use data coming from the BGW process is studied for the contagion modeled as a Poisson distribution. It consists on looking at the realizations of subsequent generations from the same realization of branching process. The convergence study for N is displayed in Figure 4.32 for the three cases studied of the Poisson family: $C \sim Poisson\,(0.9)$, $C \sim Poisson\,(1.5)$ and $C \sim Poisson\,(2.1)$.

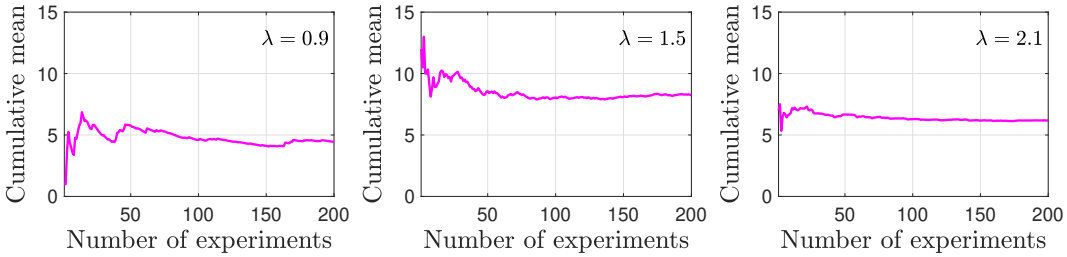


Figure 4.32: Convergence of the cumulative mean of the number of updates for the case of data coming from different subsequent generations when $C \sim Poisson\,(0.9)$, $C \sim Poisson\,(1.5)$ and $C \sim Poisson\,(2.1)$ respectively.

There is a similar behavior in the results seen in Figure 4.32 when compared to the previous ones. The bayesian updating for a single experiment from the 200 hundred studied does not properly end when the L_2-Wasserstein tolerance is achieved. Figure 4.33 shows the mode of L paths.

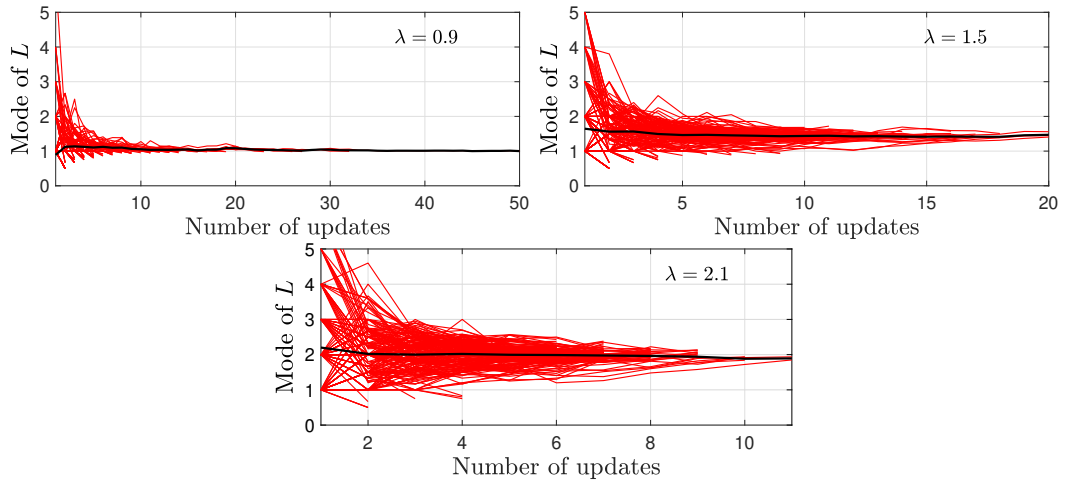


Figure 4.33: Mode of L paths for the case of data coming from different subsequent generations for the same BGW process when $C \sim Poisson\,(0.9)$, $C \sim Poisson\,(1.5)$ and $C \sim Poisson\,(2.1)$ respectively. The black curves are the mean paths.

For $\lambda = 0.9$, what usually happens is that many of the realizations of the BGW process extinct shortly. Therefore, the number of updates in these cases is not high. There was for instance a divergent realization that lasted longer. As we can see, this one took 50 times before the bayesian updating stops. It shows the discrepancy between ramification trees that extinct shortly and the ones that not and how this affects the study. Higher probabilities of extinction affect critically the result of parametric inference when data is used this way. On the other hand, for $\lambda = 2.1$, where the chances of extinctions are lower, the parametric inference is way more efficient, but the L_2-Wasserstein is not a good criteria to define when the bayesian updating should stop. The gain of

good information per update is so strong that the evolution of the posteriors shortly faces floating-point arithmetic precision problems.

4.3.4 Brief discussion on computational costs

There are two different ways explored in here to work with data to make bayesian inference of parameters from the contagion random variable. The first one, which focuses on different realizations of the same generation, is compared in the last section the greater the number of the generation is. The general behavior seen is that the required number of updates to achieve the tolerance imposed by the L_2-Wasserstein convergence criteria decreases in mean for further generations. However, it is important to highlight that the runtime spent to do this increases. The greater the number of the generation is, the analytical expressions to build the likelihood functions turn more complex, regardless the probabilistic model of the contagion random variable.

When the contagion random variable was modeled according to the Geometric-0 or the Poisson family, a limit to the maximum outcome of data was indeed necessary to be imposed to make the analysis feasible. Even though, it was not possible to work with the 4th generation when $C \sim Geometric-0\,(0.3)$ and $C \sim Poisson\,(2.1)$. The most hard-working likelihoods to build are the ones related to $C \sim Poisson\,(\lambda)$. Moreover, a recurrent problem was exceeding RAM capacity. It was usually faced the greater the number of the generation is, regardless the probabilistic model of the contagion. The workaround made was to run the 200 experiments not in a row, but do it instead partially.

If data comes from each subsequent generation of the realization of the same BGW process, the runtimes were not anymore the problem. The fact that probability of extinction was a crucial point is not a computational issue. However, when the realizations of the branching process do not extinct in earlier generations, the boom of information makes really sharpen likelihood functions. As a consequence, floating-point arithmetic precision is compromised and the updating process struggles.

5
Conclusions

Mass functions for further generations play a central role in this work. They are fundamental to get an overall comprehension on the stochastic evolution of a disease over time. The first part of this dissertation focuses on different methodologies to obtain these mass functions.

The first methodology developed to get the mass functions are probability generating functions. It is the methodology that struggles the most with computational costs. The greater the number of the generation is, the analytical expression of the pgfs gets more complex, because the number of recurrence calls in muliticomposition function increases. As a consequence, the use of symbolic computation is higher than the other methodologies. In order to find the probabilities, taking derivatives of these expressions are necessary. The probability associated to a greater number of members infected in any generation requires a higher-order derivative. Two main issues arise in this process. The first disadvantage is the runtime spent to perform high-order derivatives the greatest the number of the generation is. At some level, it is no longer feasible. The second disadvantage is storing in MAT-file data each single expression related to the probability of a specific numbers of members infected in some generation. On the other hand, this technique allows to build likelihood functions regardless the probabilistic model for the contagion random variable.

A novel attempt to improve this methodology was done introducing the polynomial identities. The goal of this second methodology is to avoid taking derivatives of functions multicomposition directly. Instead of that, based on the chain rule, the derivatives would be done for single functions individually. In order to do that, it is fundamental to know how the chain rule splits the derivatives. The polynomial identities are responsible to describe it and two different techniques to get them were discussed. The first one was based on the Faà di Bruno's formula and the second was a recursive fashion way to generate them. In terms of runtime, the former struggles around the 45th polynomial, which is associated to the probability of 45 members infected to any generation. The latter was able to get up to 60 polynomials and the limitation was a result of RAM capacity. The use of these identities provides better results in terms of runtime the greatest the number of the generation is, because the

pgf methodology itself is not feasible to find probabilities associated to higher number of members infections of further generations.

The second methodology used comes from the Markov chain property of the BGW process. The use of one-step transition matrices and a initial distribution to find mass functions for further generations in this branching process is also a novelty in this work. However, there is a limitation to apply it. It is necessary a discrete support of the contagion's probabilistic model. The case, for example, of the Binomial distribution. Different from the pgfs, this is a time-dependent and a global methodology. However, the computational costs (runtime and storage) are in comparison greater. The main reason for it is that the one-step transition probabilities are indeed function of pgfs, but they are not in a function multicomposition structure. This explains the lower runtimes spent. Now, from the storage perspective, the data stored are all numerical information and this require less bytes. Moreover, an important aspect is that this methodology also allows to build likelihood functions.

Monte Carlo simulations was the last methodology used to find mass functions for further generations in a global and time-dependent sense. Different from the other ones analyzed, it is a numerical approach. There is no need to deal with symbolic computation in here. As a consequence, the runtime and storage properties are, the further the generation is, lower than the other methodologies according to the set up for comparison. However, these properties are this time stochastic, since they rely on realizations of the random generator related to the probabilistic model of the contagion random variable. Another important aspect to mention is that, since it is a numerical methodology, it is not able to provide analytical expressions to build likelihood functions.

Now that mass functions for further generations are able to be found and two methodologies provide an analytical relation between the contagion random variable and any further generation's size random variable, the classical inverse problem of parametric inference can be studied with the help of the Bayes' rule. Given that the probabilistic model of the contagion's random variable is known, its parameters are treated as random variables and they are possible to be inferred using data coming from realizations of the BGW process. Two strategies to incorporate data were proposed in this work as novelties and studied. The first one is related to data coming from different realizations of the same generation. The other is related to data coming from the same ramification tree observed over a certain number of generations, i.e., a realization of the BGW process. The criteria to stop the bayesian updating of prior distributions into posterior distributions was the L_2-Wasserstein distance.

The number of updates required is here a discrete random variable indeed, since the data relies on random generators. Therefore, this point was studied in a stochastic sense.

The first strategy to use data shows in general that it requires less updates to achieve the same tolerance of the L_2-Wasserstein criteria when data come from further generations, regardless of the probabilistic model of the contagion's random variable. The main reason for it is that the greater the number of the generation is the normalized likelihoods' shape turn sharper. Therefore, their mapping info is more concentrated in closer intervals. Other conclusion seen is that it is easier to infer the values of the parameters when data is actually generated in situations where the contagion's random variable is more infectious. This is a result of likelihoods tending to be less flat for parameters that lead to more infectious scenarios.

Finally, the second strategy to use data relies significantly on the probability of extinction of the disease over the generations. If the realization of the branching process ends shortly, there is not sufficient information available to do a great inference of the parameters. On the other hand, when the ramification lasts longer, each new update as a likelihood function changes the prior beliefs in a very strong manner. The updating process is way more gainful than the first way to incorporate data. However in this situation, floating-point arithmetic precision is a concern. Moreover, the L_2-Wasserstein is still a good way to measure the impact of data changing a prior into a posterior distribution, but it was not efficient as a stopping criteria to this strategy.

Another contribution of this work is the development in MATLAB of all the symbolic and numerical routines used. Algorithms were developed to find the mass function for further generations according to each methodology: probability generating functions with and without the polynomial identities, Markov chain and Monte Carlo simulations. Moreover, for the case of polynomial identities, the Faà di Bruno's formula and the recursive fashion approach were implemented by the author in MATLAB. The routines related to the second part of the dissertation, which are also novelties, consist on performing parametric bayesian inference with data coming from the BGW process in both strategies using likelihoods constructed according to the pgf methodology and also the Markov methodology. In this part, the implementation of the L_2-Wasserstein was also done.

Besides this dissertation, the following articles were also made:

– "A comparison of different approaches to find the probability distribution of further generations in a branching process" for the 6th International Symposium on Uncertainty Quantification and Stochastic Modeling (Un-

certainties 2023). This article was done with the contribution of Roberta Lima and Rubens Sampaio. It focuses on the comparison among the probability generating function, Markov chain and Monte Carlo simulations methodologies to find mass functions for further generations in a BGW process;

– "Strategies to find the values of the mass functions of further generations of a branching process that models the spread of an epidemiological disease" for the XLII Congresso Nacional de Matemática Aplicada e Computacional (CNMAC 2023). This article was done with the contribution of Roberta Lima and Rubens Sampaio. It focuses on the comparison between the probability generating function with and without the polynomial identities and also on the comparison between the Faà di Bruno's formula and the recursive fashion approach to generate these polynomials;

– "Implementation of polynomial identities from the chain rule in probability generating functions for branching processes" for the XXXIX Congreso Argentino de Mecánica Computacional (MECOM 2023). This article was done with the contribution of Roberta Lima and Rubens Sampaio. It focuses on the mathematical implementation of the polynomial identities to tackle function multicompositions and the advantages and disadvantages of its use when compared to the probability generating functions itself.

Bibliography

[1] LEUNG, N. H. L. **Transmissibility and transmission of respiratory viruses**. Nature Rev Microbiol, 19:528–545, 2021.

[2] VERELST, F.; WILLEM, L.; BEUTELS, P. **Behavioural change models for infectious disease transmission: a systematic review (2010-2015)**. J. R. Soc. Inteface, 13:528–545, 2016.

[3] MARTCHEVA, M. **An Introduction to Mathematical Epidemiology**. Springer, New York, USA, 1st edition, 2015.

[4] LI, M. **An Introduction to Mathematical Modeling of Infectious Diseases**. Springer, Switzerland, 1st edition, 2018.

[5] BRAUER, F.; CASTILLO-CHAVEZ, C.; FENG, Z. **Mathematical Models in Epidemiology**. Springer, New York, USA, 1st edition, 2019.

[6] BORGES, B.; LIMA, R.; SAMPAIO, R. **Análise estocástica de propagação de doenças epidemiológicas**. Revista Mundi Engenharia, Tecnologia e Gestão, 6(3):352–01, 352–11, 2021.

[7] BORGES, B.; LIMA, R.; SAMPAIO, R. **How the spread of an infectious disease is affected by the contagion's probabilistic model**. XIV Encontro Acadêmico de Modelagem Computacional, p. 1–10, 2021.

[8] HACCOU, P.; JAGERS, P.; VATUTIN, V. A. **Branching Processes: Variation, Growth, and Extinction of Populations**. Cambridge University Press, New York, 1st edition, 2005.

[9] BACAËR, N.. **A Short History of Mathematical Poulation Dynamics**. Springer, London, England, 1st edition, 2011.

[10] GONZÁLEZ, M.; MINUESA C.; PUERTO I.; VIDYASHANKAR A.. **Robust estimation in controlled branching processes: Bayesian estimators via disparities**. International Society for Bayesian Analysis, 16:1009–1037, 2021.

[11] KUCHARSKI, A. **The Rules of Contagion**. Profile Books, London, 1st edition, 2020.

[12] BANSAYE, V.; MÉLÉARD, S. **Stochastic Models for Structured Populations: Scaling Limits and Long Time Behavior**. Springer, Switzerland, 1st edition, 2015.

[13] PARDOUX, E. **Probabilistic Models of Population Evolution**. Springer, Switzerland, 1st edition, 2016.

[14] DAWSON, A. D.; GREVEN, A. **State dependent multitype spatial branching processes and their longtime behavior**. Electronic Journal of Probability, 8:1–93, 2003.

[15] GREVEN, A.; RIPPL, T.; GLÖEDE, P. **Branching processes - a general concept**. ALEA, Lat. Am. J. Probab. Math. Stat., 18:635–706, 2021.

[16] LIMA, R.; SAMPAIO, R. **What is uncertainty quantification?** Journal of the Brazilian Society of Mechanical Sciences and Engineering, 40(155), 2018.

[17] CORTÉS, J.; LÓPEZ-NAVARO, E.; ROMERO, J.; ROSELLÓ, M. **Probabilistic analysis of a cantilever beam subjected to random loads via probability density functions**. Computational and Applied Mathematics, 42:42:42, 2023.

[18] HOWSON, C.; URBACH, P. **Scientific Reasoning The Bayesian Approach**. Open Court, Chicago, 3rd edition, 2006.

[19] JAYNES, E. T. **Probability Theory The Logic of Science**. Cambridge University Press, New York, 1st edition, 2003.

[20] SCHINAZI, R. B. **Classical and Spatial Stochastic Processes**. Springer, New York, 2nd edition, 2014.

[21] GRIMMETT, G.; WELSH, D. **Probability: An Introduction**. Oxford University Press, New York, 2nd edition, 2014.

[22] RIORDAN, J. **Introduction to Combinatorial Analysis**. John Wiley & Sons, New York, 1st edition, 1958.

[23] NATALINI, P.; RICCI, P. E. **An extension of the bell polynomials**. Computers and Mathematics with Applications, 47:719–725, 2004.

[24] The MathWorks, Inc. **MAT-File Format**, 2023.

[25] GONZÁLEZ, M.; MARTÍN, J.; MARTÍNEZ R.; MOTA M.. **Non-parametric bayesian estimation for multitype branching processes through simulation-based methods**. Computational Statistics I& Data Analysis, 52:1281–1291, 2008.

[26] RACHEV, S.; KLEBANOV, L.; STOYANOV, S.; FABOZZI, F.. **The Methods of Distances in the Theory of Probability and Statistics**. Springer, New York, 1st edition, 2013.

[27] VALLENDER, S.. **Calculation of the wasserstein distance between probability distributions on the line**. Theory of Probability & Its Applications, 18(4):784–786, 1974.

[28] DEZA, M.; DEZA, E.. **Encyclopedia of Distances**. Springer, Berlin, 4th edition, 2016.

[29] QIN, Q.; HOBERT, J. **Wasserstein-based methods for convergence complexity analysis of MCMC with applications**. Ann. Appl. Probab., 32:124–166, 2022.