

Article

Estimation of the Instantaneous Reproduction Number and Its Confidence Interval for Modeling the COVID-19 Pandemic

Publio Darío Cortés-Carvajal ¹, Mitzi Cubilla-Montilla ^{2,3,*} and David Ricardo González-Cortés ⁴¹ Independent Researcher, Panama City 0824, Panama; pdcortesc@gmail.com² Departamento de Estadística, Universidad de Panamá, Panama City 0824, Panama³ Investigadora del Sistema Nacional de Investigación de la Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT), Panama City 0824, Panama⁴ Tetrapack Panamá, Panama City 0819, Panama; davidrgc10@gmail.com

* Correspondence: mitzi.cubilla@up.ac.pa

Abstract: In this paper, we derive an optimal model for calculating the instantaneous reproduction number, which is an important metric to help in controlling the evolution of epidemics. Our approach, within a *frequentist framework*, gave us the opportunity to calculate a more realistic *confidence interval*, a fundamental tool for a safe interpretation of the instantaneous reproduction number value, so that health and governmental people pay more attention to it. Our reasoning begins by decoupling the incidence data in mean and Gaussian noise by using practical series analysis techniques; then, we continue with a likely relationship between the present and past incidence data. Monte Carlo simulations and numerical integrations were conducted to complement the analytical proofs, and illustrations are provided for each stage of analysis to validate the analytical results. Finally, a real case study is discussed with the incidence data of the Republic of Panama regarding the COVID-19 pandemic. We have shown that, for the calculation of the confidence interval of the instantaneous reproduction number, it is essential to include all sources of variability, not only the Poissonian processes of the incidences. This proposal is delivered with analysis tools developed with Microsoft Excel.

Keywords: Bayesian framework; COVID-19; generation time; Monte Carlo simulation; Poissonian variation; serial interval; time-since-infection models



Citation: Cortés-Carvajal, P.D.; Cubilla-Montilla, M.; González-Cortés, D.R. Estimation of the Instantaneous Reproduction Number and Its Confidence Interval for Modeling the COVID-19 Pandemic. *Mathematics* **2022**, *10*, 287. <https://doi.org/10.3390/math10020287>

Academic Editor: Victor Leiva

Received: 30 November 2021

Accepted: 14 January 2022

Published: 17 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Throughout history, epidemics have been a threat to humanity due to their rapid ability to spread locally, regionally, and globally. Today, humanity is living one of these epidemics, COVID-19, which was classified as a pandemic by the World Health Organization on 11 March 2020, becoming one of the most devastating pandemic phenomena for human health in the last century. COVID-19 challenged health systems all over the world. To study the reproduction of this virus, health experts designed a series of strategies [1]. Consequently, recent investigations have been developed to model the COVID-19 pandemic [2–4].

The number of infections that occur in an epidemic is associated with the number of cases in the past [5]. This observation has led to mathematical models to explain the evolution of epidemics [6]. Usually, these models propose methods to calculate a frequently used epidemiological indicator, the *instantaneous reproduction number* R_t , which is the average number of secondary cases caused by an infected individual [1,7]. As this indicator comes from incidence (cases per day) data with a lot of variability, we should not take decisions based on the R_t value alone, because it will be affected by this instability. The statisticians dealt with this problem by conceiving the confidence interval, which is a range of values where it is probable to find the true mean (the R_t in our case). Therefore, to take an informed decision we should take care of both the R_t (a sample mean) and its confidence interval. To date, the current methods to calculate the R_t had consistent results between them [8]. Thus, our main concern is the confidence interval because, when it is calculated

by current methods, there is an accepted practice of considering only the Poisson process of the incidences, discarding other kinds of variability (registry errors, delayed tasks, counting errors, etc.). Unfortunately, this practice produces too narrow confidence intervals for the R_t . A narrow confidence interval could be interpreted as positive at first glance; however, in this case, the people responsible for making the decisions are not going to have the complete information available. This is the problem that we are proposing to solve.

In summary, the main objective of this paper is to propose a new model to calculate the R_t confidence interval, addressing not only the Poissonian random process but also considering all sources of variability. We provide a valid link to the repository (Available at <https://github.com/publiodariocortes/seguiepi>, accessed on 30 October 2021) that implements the estimation method described and allows reproduction of the results.

The paper is organized into the following sections. Section 2 comprises the theoretical foundations and the literature review. In Section 3, a model is proposed to simplify the interpretation of the series of incidence, decoupling it into a mean and a Gaussian component of pure noise. The contagion model is described, and a special noisy reproduction number is deduced, obtaining an explicit expression for the estimation of the instantaneous reproduction number and the confidence intervals. Section 4 provides graphical demonstrations that support the assumptions made throughout this paper. At the end of this section, we present a real case with incidence data of the COVID-19 pandemic of the Republic of Panama. Finally, in Section 5, we present our conclusions.

2. Literature Review

Although the modeling of epidemics had its beginning in 1760 with a study about the spread of smallpox [9], it was not until the beginning of the 20th century that modern modeling began with Kermack and MacKendrick [10], who formulated the first *compartmentalized deterministic model* (SIR). These models captured important features of disease transmission, but their discrete stage structuring was in contradiction with the complex medical realities of disease progression [11]. For this reason, as an alternative more adjusted to reality, different models have been proposed that recognize the importance of the *time-since-infection*, the so-called TSI models. Within this category of models, different methods have been devised to calculate the *effective reproduction number*. This is an indicator that has become very relevant to monitor the chain of infections during epidemics, which precisely uses the time-since-infection to calculate it. The distribution of this time is the base to calculate a series of w_s factors, whose values are used to adjust the importance of past incidence to cause current incidence, considering the elapsed time from the instant of infection to the instant when the symptoms appear. According to Fraser [12], the *effective reproduction number* has two modalities. The first, called the *case (or cohort) reproduction number*, was proposed by Wallinga and Teunis [13]. In this modality, this indicator was calculated dividing the incidence that will occur in the future caused by infected individuals (incidence) at the present instant t , by the incidence at instant t . The second modality, called the *instantaneous reproduction number*, was proposed by Fraser himself. In this modality, this indicator was calculated dividing the incidence at time t by an estimate of the incidence before t (past), which could have contributed to the contagion of individuals counted as incidence at time t . In both modalities, it is necessary to relate the infected individuals with those who infected them. Since this relationship is very difficult to establish, it has been necessary to make estimates based on the TSI model.

Many have contributed to calculating the *effective reproduction number* in the modalities identified by Fraser [12], and we briefly review them here. First, as a departure study, we have the work of Fine [14]. He completed a detailed study of the time and place of the infection events, including the time elapsed between infection events of successive individuals, in a chain of transmission, the *transmission interval*. Wallinga and Teunis [13] developed a method to calculate the *case reproduction number*. Svensson [15] proposed a mathematical model that characterized the *serial interval* and *generation times*, clarifying the differences between these two concepts. The *generation times* were measured from the

time individual A was infected, until the moment that an individual B was infected by individual A. On the other hand, the *serial interval* is measured from the time individual A presented symptoms, until the time that an individual B, infected by individual A, exhibited symptoms. Fraser [12] deduced a method to calculate the *instantaneous reproduction number*, mentioned above. Cori et al. [16] proposed another method to calculate the *instantaneous reproduction number* (R_t), by assuming that the distribution of infectivity (or *serial interval*) was independent of the calendar time [16], considering that the *incidence in the observation time* t follows a Poissonian distribution, with a mean given by $R_t \sum_{s=1}^t I_{t-s} w_s$. With this, following a *Bayesian analysis framework*, the R_t and the corresponding *confidence interval* were calculated. Thomson et al. [17] developed a proposal that moved away from the assumption of an invariant *serial interval* with the time, and from the assumption of local transmission of the infections, maintaining the assumption of the Poissonian distribution of the incidences. In this context, the Bayesian model has been adopted by several researchers [18,19].

Finally, below is shown a comparison of the investigations regarding epidemic models (Table 1). It should be noted that only two investigations considered all variability (Fraser [12] and ours). The analytical method for calculating the confidence interval we are using includes all variability, a basic requirement for Statistical Process Control.

Table 1. Comparative table of investigations regarding to epidemics process models. The blank cells mean the criterion was not applied.

Year	Author	Subject	Deterministic (O)/Stochastic (X)	Stochastic Framework	Variability Considered	Method for Confidence Interval Calculation	Potential to Control the Sanitary and Governmental Processes of the Epidemics from the Perspective of Statistical Control Process
1927	Kermack & MacKendrick	SIR	O				
2003	Fine	Transmission interval	X				
2004	Wallinga & Teunis	Effective reproduction (case)	X	Bayesian	Poisson	Not declared	No
2007	Svensson	Generation time and serial interval	X				
2007	Fraser	Instantaneous reproduction number and case (cohort) reproduction number	X	Frequentist	All	Experimental (Bootstrapping)	Could be difficult with bootstrapping techniques
2013	Cori et al.	Instantaneous reproduction number	X	Bayesian	Poisson	Analytically (Gamma distribution)	No
2019	Thompson et al.	Instantaneous reproduction number	X	Bayesian	Poisson	Bayesian method (Credible interval)	No
2020	Peterson & Adhikari	Time since infection models	O				
2020	Zhao et al.	Serial interval	X				
2021	Cortés-Carvajal, P.D.; Cubilla Montilla, M.; González-Cortés, D.R.	Instantaneous reproduction number	X	Frequentist	All	Analytically (Normal distribution, using the central limit theorem)	Yes

3. Materials and Methods

This section provides details about the data simulation, the model used to simplify the interpretation of the series of incidence, the contagion model, the estimation of the instantaneous reproduction number and the confidence intervals.

3.1. Data

As the R_t calculation is based on the I_t incidence serial data and considering we aspire to include all sources of variability, we must have a practical statistical model that describes I_t incidence serial data. This model should facilitate the statistical calculations of statistical concepts such as means, standard deviations, distribution models, and confidence intervals. Taking care of our objective by observing the graphs of some pandemic data (Figure 1) it is common to find an aleatory pattern around the mean, with a bandwidth growing following the mean growing.

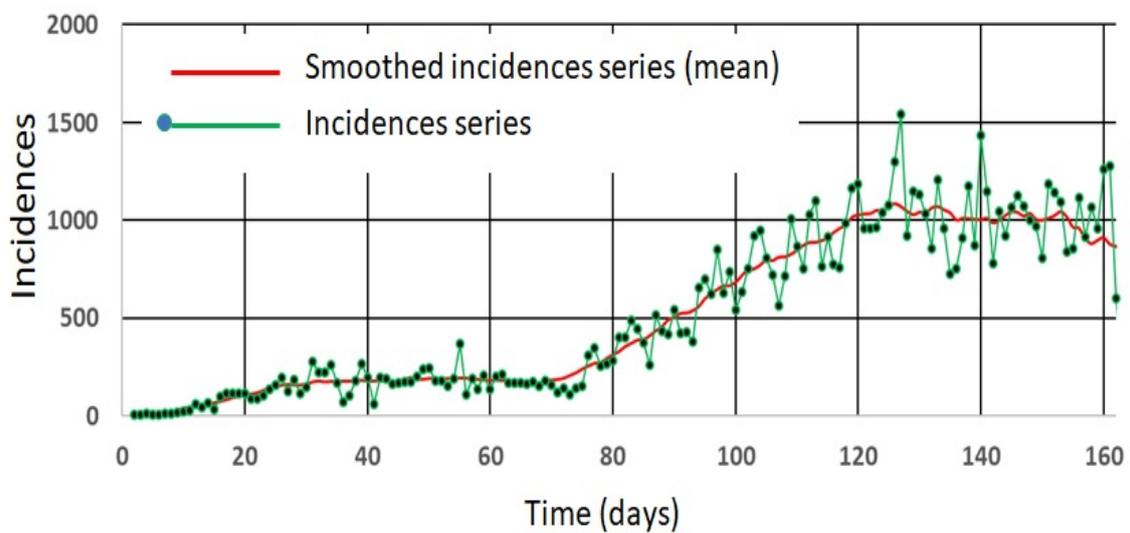


Figure 1. Partial series of incidence and smoothed incidence (mean) of the COVID-19 pandemic, Panama 2020, showing a typical pattern of variation around the mean.

It is unrealistic to think that I_t will have the same type of distribution in all observing instants. However, for simplicity, we need to select only one type, which should be something in the middle (symmetric and mesokurtic). The best candidate for this purpose is normal distribution, which has mathematical properties that simplify the analysis and the results interpretations. For example, the normal distribution is the core of the Central Limit Theorem. The normal distribution is useful modeling the distribution of large integer values, as it is the case of the incidences we can expect in an epidemic. Consequently, it should be tolerable that a few incidences of low integer values have a somewhat distorted statistical representation.

Therefore, we are proposing a model that assumes the existence of a “smooth” curve I_t^μ to which a random sequence of noise e_t has been added. This noise will be a composition of two independent types of noise: *counting error* and *Poissonian variation*. *Counting error* is assumed to follow a *normal distribution*, and *Poissonian variation* is considered to be approximated with a normal distribution also. Therefore, the mix of these noises produce a single noise, e_t , which follows a unified normal distribution, $N(0, \sigma_t^2)$, at each measurement time t . Furthermore, to achieve a more realistic model, the noise, e_t , is assigned a standard deviation, σ_t , that varies according to a function $f\{I_t^\mu\}$, which is all time positive and increasing with I_t^μ .

$$I_t = I_t^\mu + e_t, \quad e_t \sim N(0, \sigma_t^2), \sigma_t = f\{I_t^\mu\} \tag{1}$$

To simplify the presentation and demonstration of the theory that we will be developing in this paper, we will use simulated data constructed as follows:

$$I_t^\mu = 120e^{(\frac{t-20}{16})^2} + 480e^{(\frac{t-60}{20})^2} + 246e^{(\frac{t-80}{10})^2} + 120e^{(\frac{t-100}{12})^2}, \quad 10 < t < 120 \quad (2)$$

The variability of the noise, e_t , is arbitrarily modeled:

$$\sigma_t = f\{I_t^\mu\} = \sqrt{10I_t^\mu} \Rightarrow e_t \sim N(0, 10I_t^\mu) \quad (3)$$

With the previous definitions, it is possible to construct a simulated time series. The simulation strategy will allow us to determine the effectiveness of the applied analysis in extracting the statistical parameters of a time series, which hypothetically behaves like the proposed model (1).

The values of I_t^μ can be easily generated in a spreadsheet. The values of e_t , which arises from a random process, require the application of a simulation process, which can also be implemented in a spreadsheet (Figure 2):

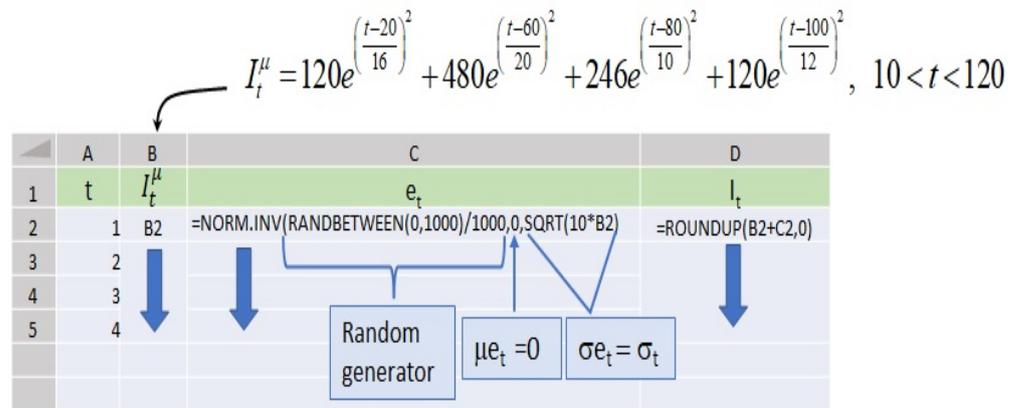


Figure 2. Generation of random incidences I_t applying special functions in an MS Excel spreadsheet.

In this way, it is possible to construct an infinity of graphs of I_t . Figure 3 shows some examples of them:

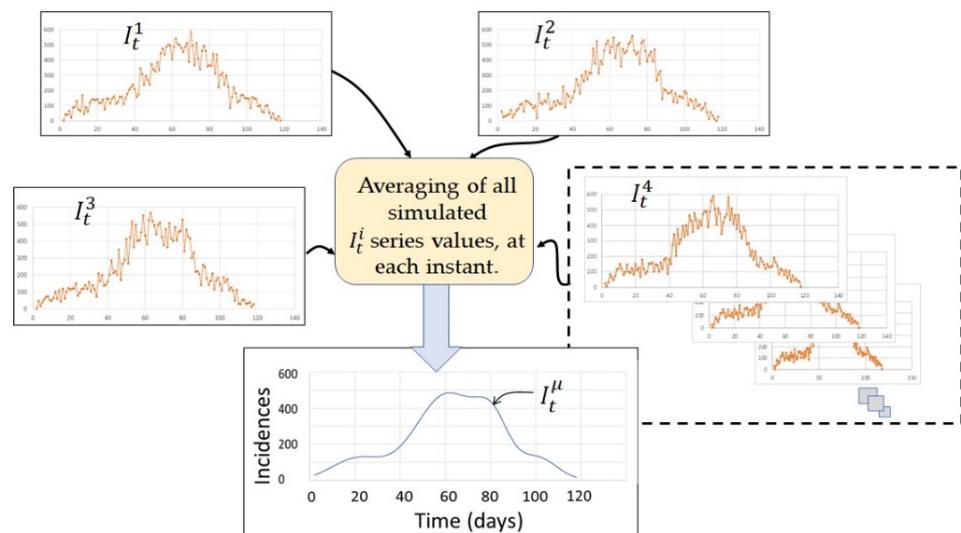


Figure 3. Random incidence series and reconstruction of the mean by averaging of them. The proximity of the calculated mean curve to the real mean curve of the population is better as the number of incidence series added increases.

Appendix A develops the theory we propose for the estimation of the model parameters (1) from real incidence data series. The estimation of the population mean, I_t^H , will be Im_t , and the estimation of the standard deviation, σ_t , will be $\hat{\sigma}_t$. Hence, from now on we will assume that Im_t and $\hat{\sigma}_t$ are known values for all incidences.

3.2. Bayesian Framework

Today the most popular method to calculate the *instantaneous reproduction number* R_t is based on Cori et al. [16], which assumes that I_t follows a Poissonian distribution, with a mean Im_t given by:

$$Im_t = R_t \sum_{s=1}^t w_s I_{t-s} \tag{4}$$

The mean depends on the known values of incidences I_{t-s} , on the infectivity profiles given by the factors w_s calculated based on the *serial interval* distribution (explained below), and on an R_t value to be discovered. Using a *Bayesian framework*, Cori et al. [16] deduced a method to calculate R_t , which follows a Gamma distribution. Now we propose a new method following an alternative approach within the *Frequentist framework*, without the restriction of a Poissonian distribution for the I_t values.

3.3. Serial Interval

As we will be using the *serial interval* in our proposal, we will briefly review this concept. The w_s values based on the *serial interval* are obtained in epidemiological studies for the current epidemic and are independent of the calendar time [12]. The values of w_s are usually modeled with a distribution chosen by the analyst based on the most accepted information. Typically, it is modeled with Gamma (Figure 4), Lognormal, or Weibull distributions [20], facilitating its parametric generation by defining the parameters of the mean μ_s and the standard deviation σ_s .

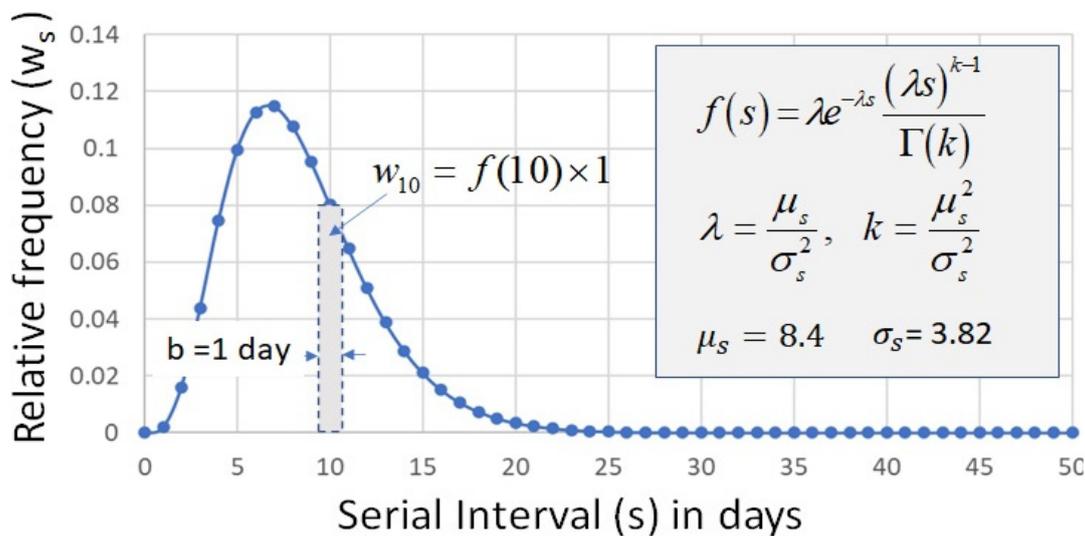


Figure 4. Typical plot of w_s factors vs serial interval constructed based on a gamma distribution. The main parameter values (k and λ) of the distribution can be defined by calculus based on the provided mean (μ_s) and standard deviation (σ_s), which should be obtained by experimentation.

The mean μ_s can be interpreted as the average time elapsed measured from the time a symptomatic individual A infects a susceptible individual B, until the time that individual B manifests symptoms. Therefore, σ_s is the standard deviation of this time.

In the analysis we will be developing, it may be known that a symptomatic individual A infected a susceptible individual B, but usually it will not be known exactly when individual B became infected after individual A became symptomatic. Thus, to simplify

this lack of information, we will assume that individual B was infected just after individual A became symptomatic.

Taking care of the above assumption, the w_s factors are calculated based on the distribution of the serial interval $f(s)$. In this example, the w_s values are calculated by the rectangular rule of integration ($f(s) \times b$). The main parameter values (k and λ) of the distribution can be defined by calculus based on a provided mean (μ_s) and a standard deviation (σ_s), which should be obtained by experimentation.

The w_s factors have a special meaning regarding *one susceptible individual infected* in $s = 0$. An infected individual in $s = 0$ will manifest symptoms on any instant after $s = 0$, and the probability that the symptoms appear in a specific instant $s = e$ in the future is w_e . In Figure 4, for example, we can see that the probability that an infected individual manifests symptoms in $s = 10$, is $w_{10} = 0.08$. Now, suppose that at $s = 0$ we have 100 susceptible individuals just infected, then the expected number of new individuals (infected at $s = 0$) with symptoms at $s = 10$ will be $0.08 \times 100 = 8$.

Not all the susceptible individuals will be infected at $s = 0$. The actual number of susceptible individuals infected at $s = 0$ will depend on the number of infectious individuals at $s = 0$, the number of susceptible individuals just before $s = 0$, and many other factors (population size, health conditions, environment, use of face masks, social distancing, population immunity, and so forth).

3.4. The Contagion Model

Let us consider a likely relation based on the fact that contagion in an epidemic comes from an infected individual:

$$I_t = \sum_{s=1}^t r_{t,s} I_{t-s} \tag{5}$$

The incidence I_t of an instant t results from the individual contributions ($r_{t,s} I_{t-s}$) to the infections of all the I_{t-s} incidences of the past, where the factors $r_{t,s}$ are the specific rates of these contributions.

Equation (5) is too simple, we present it here just as didactical tool to emphasize the exclusive relationship of I_t (present) with the past values of I_{t-s} ($t - s = 1, 2, \dots, t - 1$), but in fact, the situation is more complex. So, to have a more realistic analysis, we will redefine the $r_{t,s}$ factors as follow:

$$r_{t,s} = R_{t,s}^* w_s \tag{6}$$

In this way, we still have a factor that contains “the specific rate” of the contribution of all the I_{t-s} values to the value of I_t , but now we have split $r_{t,s}$ into two factors, $R_{t,s}^*$ and w_s , to which we will be assigning important meanings.

The w_s factors are known from the *serial interval*, representing the effects that have occurred since the elapsed time from the infection in the instant ($t - s$) to the instant t of symptoms manifestation, considering we are constructing a TSI model. It can be said, these factors represent the *natural law that governs the relationship between the infectious agent and its host*. Thus, by definition, these factors must be statistically constants (Figure 4) and independent of t .

The $R_{t,s}^*$ mathematical interpretation will be deduced below. In the meantime, assume it is a convenient factor necessary to make (5) valid.

Replacing $r_{t,s}$ from (6) in (5):

$$I_t \sum_{s=1}^t R_{t,s}^* w_s I_{t-s} = \sum_{s=1}^t w_s (R_{t,s}^* I_{t-s}) \tag{7}$$

We can still say the relation presented in (5) is accomplished, but now, with the specific meaning of w_s , it is possible to deepen the mathematical interpretation of (7). Considering the example presented at the end of Section 3.3 and Equation (7), the product ($R_{t,s}^* I_{t-s}$) could be interpreted as the total number of susceptible individuals just *infected* in the instant

$(t - s)$. Therefore, the product $w_s (R_{t,s}^* I_{t-s})$ is the number of individuals infected in $(t - s)$ who will become symptomatic in t . The sum of all $w_s (R_{t,s}^* I_{t-s})$ products from $s = 1$ to $s = t$ will result precisely in the incidence I_t (all the symptomatic individuals counted in t).

The above is consistent with the example of Section 3.3; our model still works, but we do not have a specific mathematical interpretation of $R_{t,s}^*$ yet. Now, we take one step forward rewriting (7) in two possible forms:

- (a) Keeping $R_{t,s}^*$ with the original dependency on t and s :

$$I_t = \sum_{s=1}^t R_{t,s}^* (w_s I_{t-s}) \tag{8}$$

- (b) Declaring $R_{t,s}^*$ as independent of s , and keeping the dependency on t :

$$I_t = R_t^* \sum_{s=1}^t (w_s I_{t-s}) \tag{9}$$

Both Equations (8) and (9) are valid and maintain the relation proposed in (5). There is no mathematical reason to prefer one or the other. Both Equations (8) and (9) let us know that not all the incidence I_{t-s} will have an *effective responsibility* in the resulting incidence I_t , because of the elapsed time effect. This condition is observed in the product $(w_s I_{t-s})$. Thus, we are free to select one of the Equations (8) or (9) by appealing any other reason instead of a mathematical one. Therefore, we will think of practical and useful reasons that help us to control the evolution of the epidemics. Equation (8) has one $R_{t,s}^*$ per combination of t and s indexes. Equation (8) is useless because there are infinite $R_{t,s}^*$ factors for each t that can be solutions for Equation (8). So, we will discard Equation (8) as impractical. In Equation (9), in each instant t we can find a specific R_t^* . Solving from Equation (9):

$$R_t^* = \frac{I_t}{\sum_{s=1}^t w_s I_{t-s}} \tag{10}$$

There is only one solution for R_t^* , which has an expression (10) that provides an *easy and practical interpretation* of R_t^* . As the summation $\sum_{s=1}^t w_s I_{t-s}$ in (9) can be interpreted as the summation of the *effective incidences* I_{t-s} ($s = 1, 2, \dots, t$), now the practical (and mathematical) interpretation for R_t^* is:

R_t^* is the ratio of I_t with respect to the effective number of infectious individuals in the past. In simple words, it means R_t^* is the number of secondary cases caused by an infected individual. This is a very important concept. As the summation $\sum_{s=1}^t w_s I_{t-s}$ in (9) is something we cannot handle, because all the incidences I_{t-s} are in the past, and the w_s factors represent a natural law, we should act on the R_t^* factor, which represents some elements that can be controlled by us (for example, population education, sanitary and governmental process, quarantines, and so forth), even if there are some other elements that cannot be handled directly (such as weather conditions). Thus, through our actions, we should try to reduce R_t^* as much as possible; to reduce the number of incidences, R_t^* must be less than 1.

3.5. Frequentist Framework

Since the denominator $\sum_{s=1}^t w_s I_{t-s}$ (10) behaves as a weighted average (low variability), the variability of R_t^* is mainly caused by the variability of I_t . Therefore, as R_t^* is a kind of *reproduction number*, we will name it the *noisy reproduction number* (just a name to distinguish it from other reproduction numbers). To stabilize R_t^* for useful purposes, we propose to obtain the mean by the standard *Frequentist method*:

$$E[R_t^*] = \int_{-\infty}^{\infty} R_t^* f(R_t^*) dR_t^* \tag{11}$$

where $f(R_t^*)$ is the probability density function of R_t^* . The deduction of the $f(R_t^*)$ equation is a little complex and extensive, so it is explained in Appendix B. Next, assuming we know $f(R_t^*)$, it is possible to solve Equation (11) numerically, but it is not possible to obtain an explicit expression for R_t^* mean. Therefore, we will try using *expected values operators* and Taylor approximations. Replacing I_t and I_{t-s} with their model components (1) in (10):

$$R_t^* = \frac{I_t^\mu + e_t}{\sum_{s=1}^t w_s (I_{t-s}^\mu + e_{t-s})} = \frac{I_t^\mu + e_t}{\sum_{s=1}^t w_s I_{t-s}^\mu + \sum_{s=1}^t w_s e_{t-s}} \tag{12}$$

Defining auxiliary variables:

$$c_t = \sum_{s=1}^t w_s I_{t-s}^\mu, \quad \varepsilon_t = \sum_{s=1}^t w_s e_{t-s}, \quad g_t(\varepsilon_t) = \frac{1}{c_t + \varepsilon_t} \tag{13}$$

To convert $g_t(\varepsilon_t)$ in a more manageable form, we will replace it with its Taylor approximation, if ε_t has small values around zero (which is true in general):

$$g_t(\varepsilon_t) = \frac{1}{c_t + \varepsilon_t} \approx \frac{1}{c_t} - \left(\frac{1}{c_t^2}\right)\varepsilon_t \tag{14}$$

Applying (14) in (12):

$$R_t^* \approx (I_t^\mu + e_t) \left[\frac{1}{c_t} - \left(\frac{1}{c_t^2}\right)\varepsilon_t \right] = \frac{I_t^\mu}{c_t} - \left(\frac{I_t^\mu}{c_t^2}\right)\varepsilon_t + \frac{e_t}{c_t} - \left(\frac{e_t}{c_t^2}\right)\varepsilon_t = \frac{I_t^\mu}{c_t} + \frac{e_t}{c_t} - \left(\frac{I_t^\mu}{c_t^2}\right)\varepsilon_t - \left(\frac{1}{c_t^2}\right)e_t\varepsilon_t \tag{15}$$

Applying expected value operators:

$$E[R_t^*] \approx \frac{I_t^\mu}{c_t} + \frac{E[e_t]}{c_t} - \left(\frac{I_t^\mu}{c_t^2}\right)E[\varepsilon_t] - \left(\frac{1}{c_t^2}\right)E[e_t\varepsilon_t] \tag{16}$$

In (13) it is implicit that ε_t and e_t are independent normal distributed variables, because e_t is not included in the weighted average of errors; therefore, there is no correlation between them, and consequently $E[e_t\varepsilon_t] = E[e_t]E[\varepsilon_t]$. Moreover, it is noted that $E[e_t] = 0$, and $E[\varepsilon_t] = 0$. Thus:

$$E[R_t^*] \approx \frac{I_t^\mu}{c_t} + \frac{E[e_t]}{c_t} - \left(\frac{I_t^\mu}{c_t^2}\right)E[\varepsilon_t] - \left(\frac{1}{c_t^2}\right)E[e_t]E[\varepsilon_t] = \frac{I_t^\mu}{c_t} \tag{17}$$

Replacing c_t from (13) in (17):

$$E[R_t^*] = R_t^\mu \approx \frac{I_t^\mu}{\sum_{s=1}^t w_s I_{t-s}^\mu} \tag{18}$$

Thus, considering that the estimated value for I_t^μ is Im_t , the estimated value of R_t^μ is:

$$R_t^m = \frac{Im_t}{\sum_{s=1}^t w_s Im_{t-s}} \tag{19}$$

This result we will be tested later in Section 4.2.

This equation is similar to the equation that Fraser [12] proposed, but he used a different approach by starting from a continuous model.

As he called this ratio the instantaneous reproduction number, we named it the same. Thus, this paper mainly proposes some steps forward to find its probability density function and the confidence interval.

3.6. The Probability Density Function of R_t^m

Now we will proceed to calculate the probability density function of the estimator R_t^m (19). To begin with, it is noted that the numerator and denominator (19) are constructed with moving averages of the incidence series I_t (Appendix A), which is a procedure that involves the overlap of “crude” incidence I_t in the calculations of the means Im_t and of Im_{t-s} . Therefore, it cannot be said that the numerator and the denominator are independent. The same occurs between the means Im_{t-s} of the denominator $\sum_{s=1}^t w_s Im_{t-s}$ (from now on denoted as Λm_t). As this lack of independence makes the analytical calculation of a formula for the probability density function of R_t^m impossible, the problem was approached by applying the experimental method of Monte Carlo simulations, for which we worked with the generator of a series of incidence (Figure 2).

With the generator, 320 series of I_t of 128 days were obtained, each of which was subjected to a calculation of moving averages, to finally obtain the curves of R_t^m , using Equation (19). Figure 5 shows a superposition of 10 of these curves.

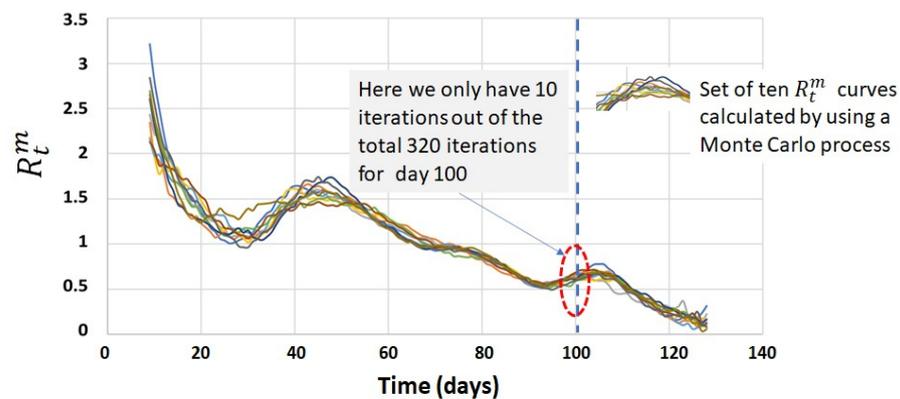


Figure 5. Superposition of 10 randomly generated R_t^m curves. Each R_t^m curve was generated on calculus based on a specific Im_t series and applying (19). Following a Monte Carlo simulation process, the Im_t series were calculated with the I_t series data generated by a random incidence series generator.

For each instant of observation, we obtained a sample of 320 R_t^m values coming from the curves (you can imagine 320 curves in Figure 5) and calculated the mean and the standard deviation of R_t^m from these values. Furthermore, at 10-day intervals, the corresponding histograms were constructed. Figure 6 shows the histogram of 320 R_t^m values, for $t = 100$.

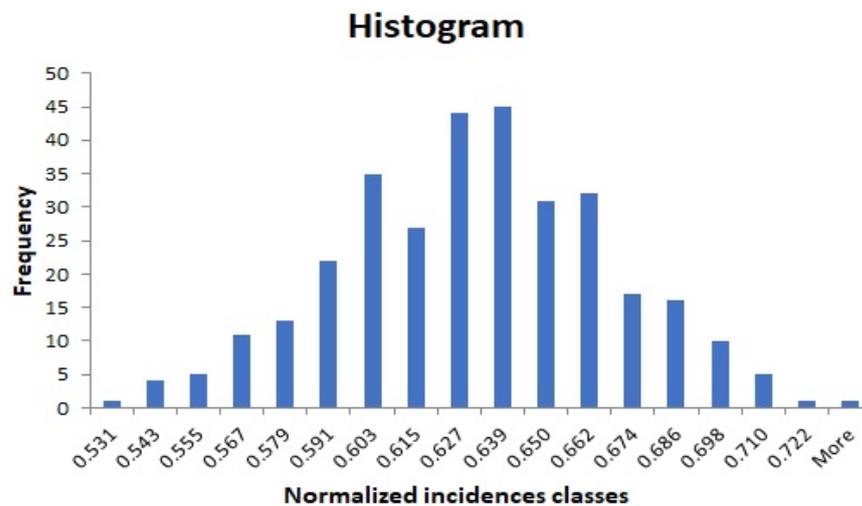


Figure 6. Histogram of the values of 320 random versions of R_t^m , at time $t = 100$, following a Monte Carlo simulation process based on the I_t series data generated with the random incidence series generator.

For practical purposes, this histogram can be said to follow a normal distribution. Thus:

$$R_t^m \sim N\left(R_t^\mu, \sigma_{R_t^m}^2\right) \tag{20}$$

This experimental finding suggests an analytical way to calculate the probability density of R_t^m , even if it is by way of approximation.

We can decompose Im_{t-s} into its model components ($Im_{t-s} = I_{t-s}^\mu + e_{t-s}^m$), as we performed with I_{t-s} (1). Thus, let us replace Im_{t-s} in the denominator of Equation (19):

$$R_t^m = \frac{Im_t}{\sum_{s=1}^t w_{t,s} \left(I_{t-s}^\mu + e_{t-s}^m \right)} = \frac{Im_t}{\sum_{s=1}^t w_s I_{t-s}^\mu + \sum_{s=1}^t w_s e_{t-s}^m} \tag{21}$$

where the e_{t-s}^m values are error components which follow a normal distribution $N\left(0, \frac{\sigma_{I_{t-s}^\mu}^2}{n_t}\right)$, with n_t as the number of incidences defined to calculate the incidence mean in the *moving average* procedure (Appendix A). This means the time series values of e_{t-s}^m delimit a compressed bandwidth compared with the bandwidth of e_{t-s} . Thus, it gives us confidence to assume that $\sum_{s=1}^t w_{t,s} I_{t-s}^\mu \gg \sum_{s=1}^t w_{t,s} e_{t-s}^m$ (which is true in general):

$$R_t^m \approx \frac{Im_t}{\sum_{s=1}^t w_s I_{t-s}^\mu} \tag{22}$$

In summary, considering the denominator $\sum_{s=1}^t w_s I_{t-s}^\mu$ is a constant, an alternative function for the probability density of R_t^m , could be:

$$g(R_t^m) = \frac{1}{\sqrt{2\pi}\sigma_{R_t^m}} e^{-\frac{1}{2}\left(\frac{R_t^m - R_t^\mu}{\sigma_{R_t^m}}\right)^2} \Rightarrow R_t^m \sim N(R_t^\mu, \sigma_{R_t^m}^2) \tag{23}$$

$$R_t^\mu = \frac{I_t^\mu}{\Lambda_t^\mu}, \quad \sigma_{R_t^m} = \frac{\sigma_t}{\Lambda_t^\mu \sqrt{n_t}}, \quad \Lambda_t^\mu = \sum_{s=1}^t w_s I_{t-s}^\mu$$

Consecutively, the estimators for the parameters of Equation (23) are:

$$R_t^m = \frac{Im_t}{\Lambda_t^m}, \quad \hat{\sigma}_{R_t^m} = \frac{\hat{\sigma}_t}{\Lambda_t^m \sqrt{n_t}}, \quad \Lambda_t^m = \sum_{s=1}^t w_s Im_{t-s} \tag{24}$$

3.7. The Confidence Interval of R_t^m

The confidence interval is straightforwardly calculated from Equation (24):

$$R_t^m - t_{\frac{\alpha}{2}, n_t-2} \hat{\sigma}_{R_t^m} < R_t^m < R_t^m + t_{\frac{\alpha}{2}, n_t-2} \hat{\sigma}_{R_t^m} \tag{25}$$

This result will be tested later in Section 4.3.

We use a t-Student distribution to compensate for any effect that could cause the number n_t used in the moving average calculation of Im_t . The degree of freedom ($n_t - 2$) used is a consequence of Im_t (22), because it can be interpreted as the middle point of a line through n_t points (see Appendix A, we use the centered moving average method), the same reasoning as if it were a case of *simple regression analysis*.

Equation (25) specifically is the instantaneous reproduction number confidence interval without uncertainty in the serial interval, because there is another type of confidence interval which considers serial interval uncertainty [16]. This means, although the serial interval is still considered to have stable values in theory, it could be impossible to obtain serial interval data to calculate the w_s factors. This could happen, for example, at the beginning of an epidemic. This condition can be handled by modeling the uncertainty in the serial interval parameters [16], which are the mean (μ_S) and the standard deviation (σ_S). It is assumed that these parameters are random independent normal distributed variables, which they really are not. We follow the same procedure of the Monte Carlo simulation, but instead of changing I_t , we changed μ_S and σ_S simultaneously. To emphasize that μ_S and σ_S will be considered as random variables, they will be replaced by ms and ss , respectively. Thus:

$$ms \sim N(ms_\mu, ms_\sigma^2), \tag{26}$$

where ms_μ and ms_σ^2 are the mean and the variance of ms .

Moreover:

$$ss \sim N(ss_\mu, ss_\sigma^2) \tag{27}$$

where ss_μ and ss_σ^2 are the mean and the variance of ss .

To identify the effect of a specific random pair j of each Monte Carlo iteration, superscripts will be applied:

$$(ms^{(j)}, ss^{(j)}) \rightarrow w_1^{(j)}, w_2^{(j)}, w_3^{(j)} \dots \tag{28}$$

Consider Equation (19), again:

$$R_t^m = \frac{Im_t}{\sum_{s=1}^t w_s Im_{t-s}} \tag{29}$$

For each observation instant t , we propose a Monte Carlo simulation process as follows: The denominator of Equation (29):

1. Generate N random pairs $(ms^{(j)}, ss^{(j)})$.
2. Calculate the weights $w_s^{(j)}$ for each pair j .
3. Calculate the denominator $\sum_{s=1}^t w_s^{(j)} Im_{t-s}$ for each pair j . Apply each pair $(ms^{(j)}, ss^{(j)})$ in the calculation of R_t^m :

- Calculate the values of the means and standard deviations corresponding to each pair:

$$R_{t,j}^m = \frac{Im_t}{\Lambda_{t,j}^m}, \quad \hat{\sigma}_{R_t^m}^{(j)} = \frac{\hat{\sigma}_t}{\Lambda_{t,j}^m \sqrt{n_t}}, \quad \Lambda_{t,j}^m = \sum_{s=1}^t w_s^{(j)} Im_{t-s} \tag{30}$$

- Calculate the general mean of all means $R_{t,j}^m$:

$$R_t^m = \frac{1}{N} \sum_j R_{t,j}^m \tag{31}$$

- Calculate the general standard deviation by using the following formula:

$$\hat{\sigma}_{R_t^m} = \sqrt{\frac{1}{N} \sum_j \left[\left(\hat{\sigma}_{R_t^m}^j \right)^2 + \left(R_{t,j}^m - R_t^m \right)^2 \right]} \tag{32}$$

- Given the mean (31), the standard deviation (32), and a *significance level* α , assuming a Gamma distribution, calculate the lower and upper limits of the confidence interval for the population mean of R_t^m , for each observation instant t.

The derivation and proof of this procedure are in Appendix C.

It should be noticed that this procedure used a Gamma distribution in step 7, when this distribution is in fact a subrogated function of a summatory of normal distributions (Appendix C). To take care of the effect caused by time of moving average (n_t), we should have used Student's t-test distributions instead to obtain a more conservative confidence interval, but unfortunately it cannot be handled with a Gamma distribution as a subrogated density function. Therefore, steps 5, 6, and 7 produce a confidence interval a little narrower than the conservative confidence interval to which we could aspire. It is a tradeoff to speed up the computation process.

4. Results

This section provides graphical demonstrations which support the assumptions completed through Section 3. At the end, a case with real incidence data of a pandemic will be presented.

4.1. Simulated Incidence Data Series

Figure 7 shows a superposition of different types of incidence series. *Crude incidence data* I_t was generated by simulation in the spreadsheet (Figure 2), adding *aleatory noise* e_t to the *ideal model* of I_t^m , following the model Equations (2) and (3). *Moving average incidence series* (Im_t) was generated using 15 crude incidence data per calculation (n_k), as described in Appendix A.

It is important to note there is no appreciable delay between these graphs (Figure 7), especially the *moving average incidence series*. We used a *centered moving average method*.

4.2. Process Calculation of R_t^m

The Im_t series was generated with the moving average of I_t ; then, Λ_t^m was generated with the denominator of the Equation (19). With this, for each t instant, we can calculate R_t^m (19). In fact, in Figure 8, it is easy to say when the R_t^m curve will be greater than 1, or less than 1. Note the smoothness of the Λ_t^m curve. As there is no appreciable randomness in the Λ_t^m curve, we can assume it is statistically constant in each instant t. This condition simplifies the deduction of the *probability density function* of R_t^m (23) and its *confidence interval* (25).

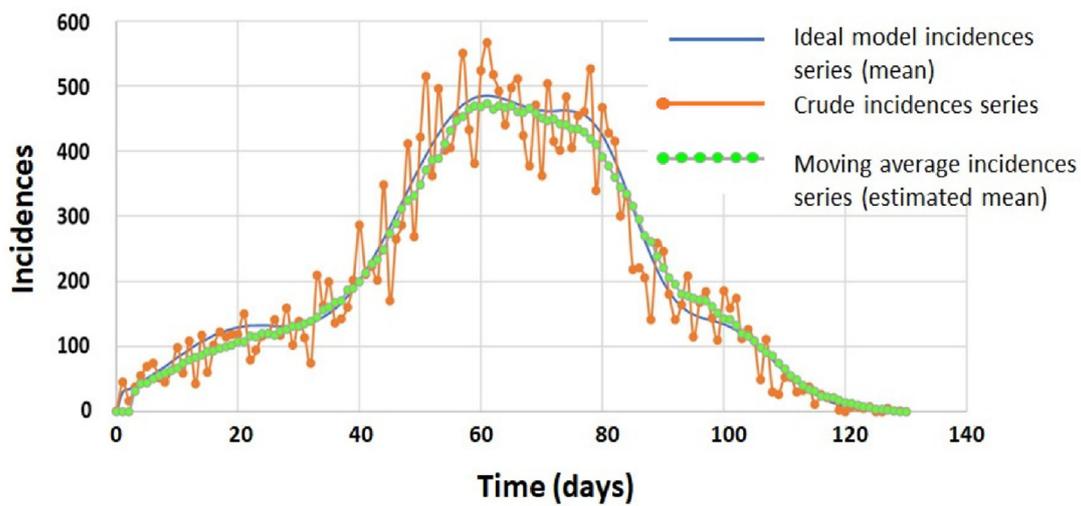


Figure 7. Crude I_t incidence series (orange-zigzag), ideal model incidence (blue-smooth), and moving average incidence (green-smooth dotted). The I_t incidence series was generated with the random incidence series generator. All curves are synchronized.

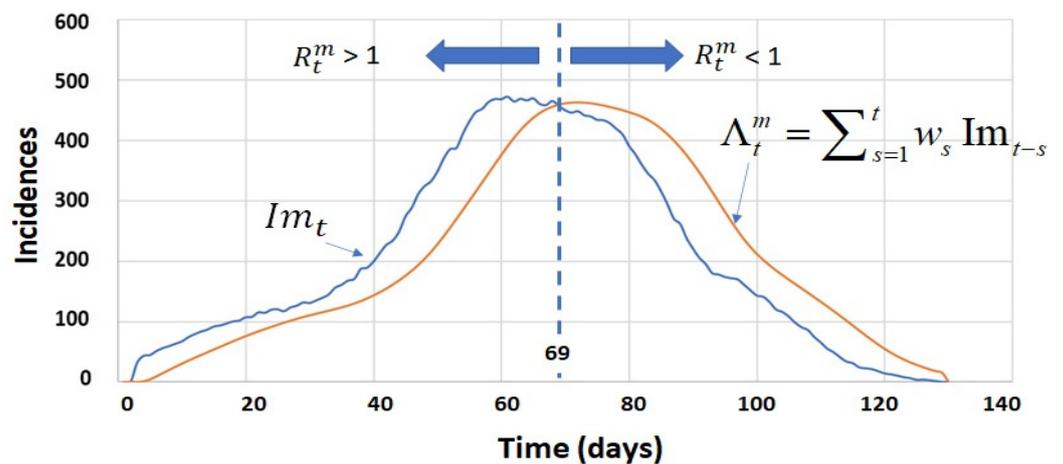


Figure 8. Im_t moving average series (blue) and Λ_t^m weighted average series (orange). The division of Im_t by Λ_t^m produces the R_t^m curve.

In Figure 9, we have the R_t^* calculated with Equation (10), the graph of R_t^m -approximate calculated with Equation (19), and the graph of R_t^m -exact calculated by the numeric integration of Equation (11). There is no visible difference between both types of R_t^m graph; therefore, we can say the Taylor approximation used to obtain Equation (19) was good enough. The numerical integration was performed, for each instant t , with the standard method of the *trapezoidal rule*.

4.3. Process Calculation of the R_t^m Confidence Interval

Table 2, on the right side, has a column of the standard deviation of all the 320 R_t^m iterations for each t instant (days), as described in Section 3.6. This large quantity of values gives us confidence that the standard deviation calculated was almost the *standard deviation of the population*, which we denote as $\sigma_{R_t^m}^{MCS}$ to emphasize it coming from a Monte Carlo simulation. Therefore, considering R_t^m follows a normal distribution (20), we can use the values of $\sigma_{R_t^m}^{MCS}$ to construct an exact confidence interval to test the confidence interval we proposed in Equation (25). Figure 10 shows the superposition of both types of confidence intervals.

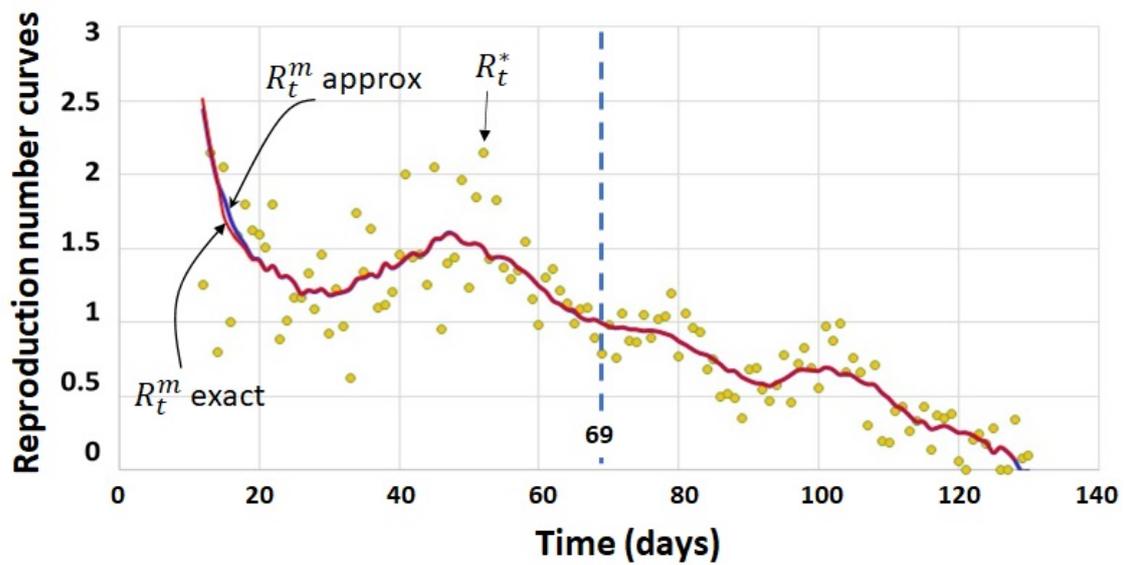


Figure 9. Comparison between reproduction number curves, R_t^* , R_t^m exact, and R_t^m approx.

Table 2. Partial table of 320 Monte Carlo iterations of 128 days R_t^m series.

Day	R_t^m iterations							SD MCS	Standard deviation
	Rtm 314	Rtm 315	Rtm 316	Rtm 317	Rtm 318	Rtm 319	Rtm 320		
11	2.17001	2.26392	2.47387	1.97609	2.00207	2.25481	2.07686	0.542613	
12	2.06127	2.11066	2.41122	1.87018	1.79859	2.13629	1.93210	0.589714	
13	1.91099	2.06799	2.32853	1.74570	1.75767	2.05797	1.89229	0.640457	
14	1.89529	1.86216	2.25943	1.59517	1.69615	1.94120	1.83866	0.69466	
15	1.85282	1.79917	2.14870	1.47665	1.70210	1.86042	1.77167	0.751537	
16	1.86321	1.68052	1.92846	1.42572	1.68435	1.76157	1.60419	0.808612	
17	1.75908	1.62979	1.83304	1.35254	1.57679	1.67131	1.57936	0.866715	
18	1.69122	1.59028	1.75054	1.38187	1.50763	1.55256	1.51021	0.924422	
19	1.59302	1.53044	1.68361	1.35205	1.43605	1.54926	1.40478	0.982537	
20	1.53617	1.46193	1.61977	1.28381	1.42267	1.43287	1.31839	1.040908	
21	1.52046	1.38184	1.61057	1.25251	1.45462	1.36895	1.22546	1.099110	
22	1.45647	1.30902	1.54020	1.23288	1.42035	1.35385	1.13442	1.157020	

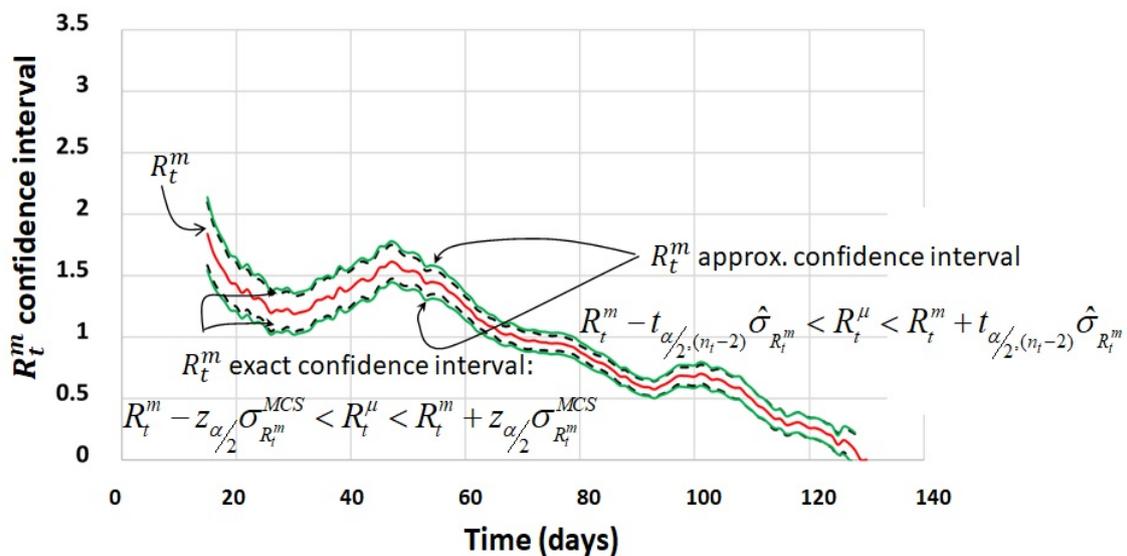


Figure 10. R_t^m confidence intervals: Monte Carlo simulation (exact) and approximated method (25).

Although it is a qualitative graphic test (Figure 10), the approximated method (25) is good enough for practical purposes. The confidence interval width of the approximated method is the widest, because it takes care of the sample size n_t used for calculating the moving average of the incidences.

4.4. Comparison between Bayesian Method and Frequentist Method for Calculation of the Instantaneous Reproduction Number (R_t vs. R_t^m)

All the methods in Figure 11 share the same I_t series and each graph was calculated with the serial interval parameters shown in the figure. The “Length of time steps” (same as time for moving average, n_k) was 15 days. The calculus of the graph data of the Bayesian framework was obtained with the Excel software EpiEstim, developed by Cori et al. [16]. This procedure requires one additional special parameter, the posterior coefficient of variation (CV) = 0.3. Both Bayesian framework graphs, Figure 11a,c, delay the Frequentist framework graphs, Figure 11b,d, by 7 days, which is a number rounded down from the length of time steps divided by 2. It is not a relationship by chance. This happens because we are using a centered moving average (Appendix A) method to calculate the mean of the incidences I_t . We selected this method because, in this way, the mean series is synchronized with the crude incidence series of I_t (Figure 7). Thus, the Frequentist framework method produces R_t^m values synchronized with the steps (days) of observations, which is a great advantage.

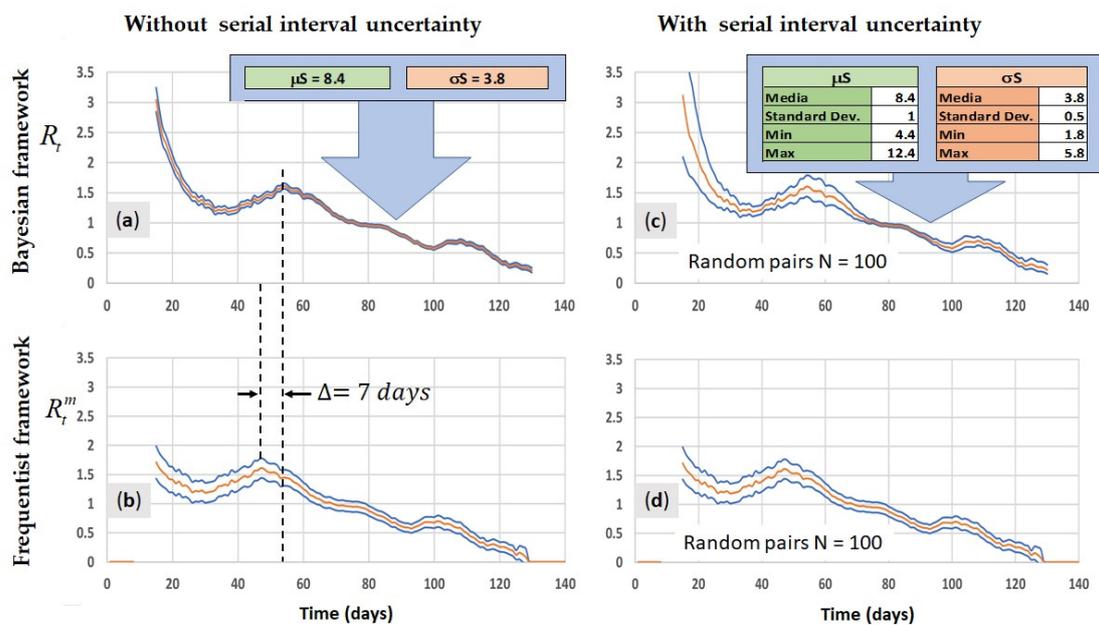


Figure 11. Instantaneous reproduction number graphs calculated by different methodologies, considering confidence intervals of 95%: (a) Bayesian framework without serial interval uncertainty; (b) Frequentist framework without serial interval uncertainty; (c) Bayesian framework with serial interval uncertainty; and (d) Frequentist framework with serial interval uncertainty.

Most of the time, the Frequentist framework graphs, Figure 11b,d, produce wider confidence interval graphs than the Bayesian framework graphs, Figure 11a,c. This has happened because the Frequentist framework method is open to accept any aleatory variation, not only Poissonian variation, as is the case of the Bayesian framework method. Usually, we would like to have a thin confidence interval but, in this case, “thin” means we are not aware of the real variability of the data; therefore, the Frequentist framework had a better performance than the Bayesian framework, because they show us the complete panorama.

Figure 12 shows an interesting relationship between the Bayesian and Frequentist framework methods; when the Frequentist framework method is 7 days delayed, both

graphs are almost the same. This happens because if we make appropriate approximations, the equations of both methods are also nearly the same.

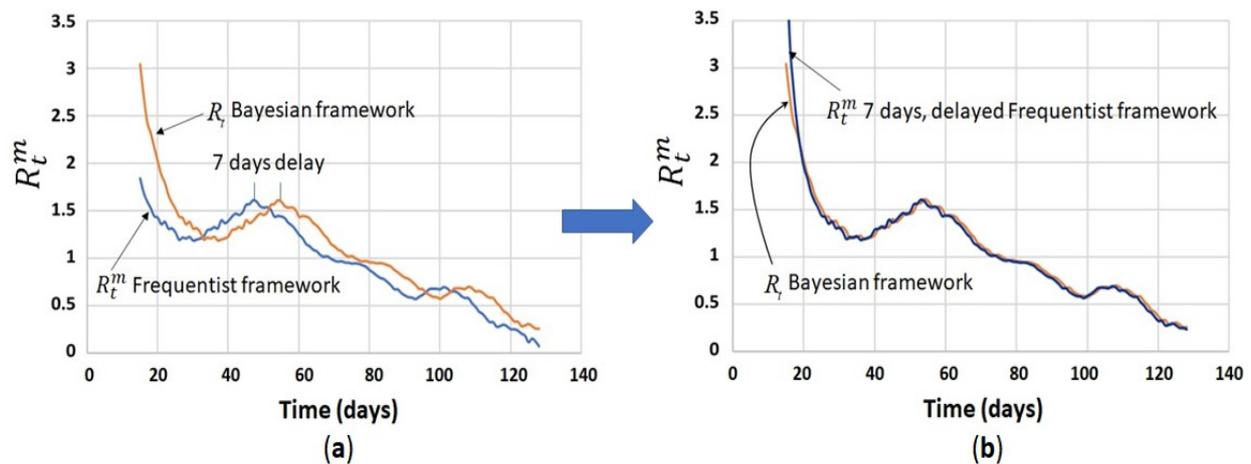


Figure 12. Instantaneous reproduction number calculated through Bayesian and Frequentist framework without *serial interval* uncertainty: (a) Superposition curves of both methods; and (b) Superposition curves delaying the Frequentist method curve by 7 days.

4.5. A Real Case Application: Pandemic COVID-19 in Panama

The data collection of the pandemic of COVID-19 in Panama began in March 2021, with some isolated cases (Figure 13).

Both instantaneous reproduction number graphs were constructed by using n_k time (Frequentist framework) and the “length of time steps” (Bayesian framework) equal to 15 days. As expected, it found a delay of 7 days; moreover, the confidence interval is wider in the Frequentist framework than in the Bayesian framework.

In Panama, the health authorities used the Bayesian framework method to calculate the R_t . The health authorities and scientists knew the pandemic could begin in our country at any moment (by knowing the world epidemic reports from the beginning in China, at the end of 2019), but there was no way to know when, where, and how, and the best procedure to follow. The incidence growth rate was high during the first days, doubling its value every three days, which caused very large R_t^m (from day 1 to 20, Figure 13a). As soon as the first pandemic procedures were implemented, and the population mobility was controlled (around day 50), the pandemic process arrived at a stage the epidemiologists named *community transmission*. This means the contagion was between individuals of the community, locally, following a natural behavior, with the infectious agents moving through the “windows” not closed by the epidemic procedures. On day 85, the new procedures began to show effectiveness, because the R_t^m (Figure 13b) diminished, but the analysts did not notice it until seven days later, because they were using R_t (Figure 13c). In any case, this condition could be difficult to detect by observing incidence data alone (Figure 13a). The ideal condition to control an epidemic, according to epidemiologists, is to obtain a reproductive number below 1. The first time it seemed to be happening was day 60 (Figure 13c), but now, with our R_t^m curve (Figure 13b), we noticed there was too much variability (wide confidence interval) to have enough confidence. For this reason, we must be cautious when the instantaneous reproduction number seems to be diminishing and use the *upper limit* of its *confident interval* to take appropriate decisions. We had four moments when the R_t (Figure 13c) seemed to be less than 1 (days 60, 140, 223, and 312, Figure 13c), considering its thin confidence interval. Retrospectively thinking, it should be noted that day 223 (18 October 2021) was a very risky day to take decisions, considering $R_t < 1$ (Figure 13c), because something happened a few days later when the incidence began to grow, resulting in the largest incidence peak of Panama. Of course, the health system used many other indicators to take decisions, and what was important was the proximity of the

end of the year festivities. If we had used our frequentist method (Figure 13b), the first moment in which the R_t^m would seem to have decreased would have been on day 310 (13 January 2021), using the criterion of the upper confidence limit.

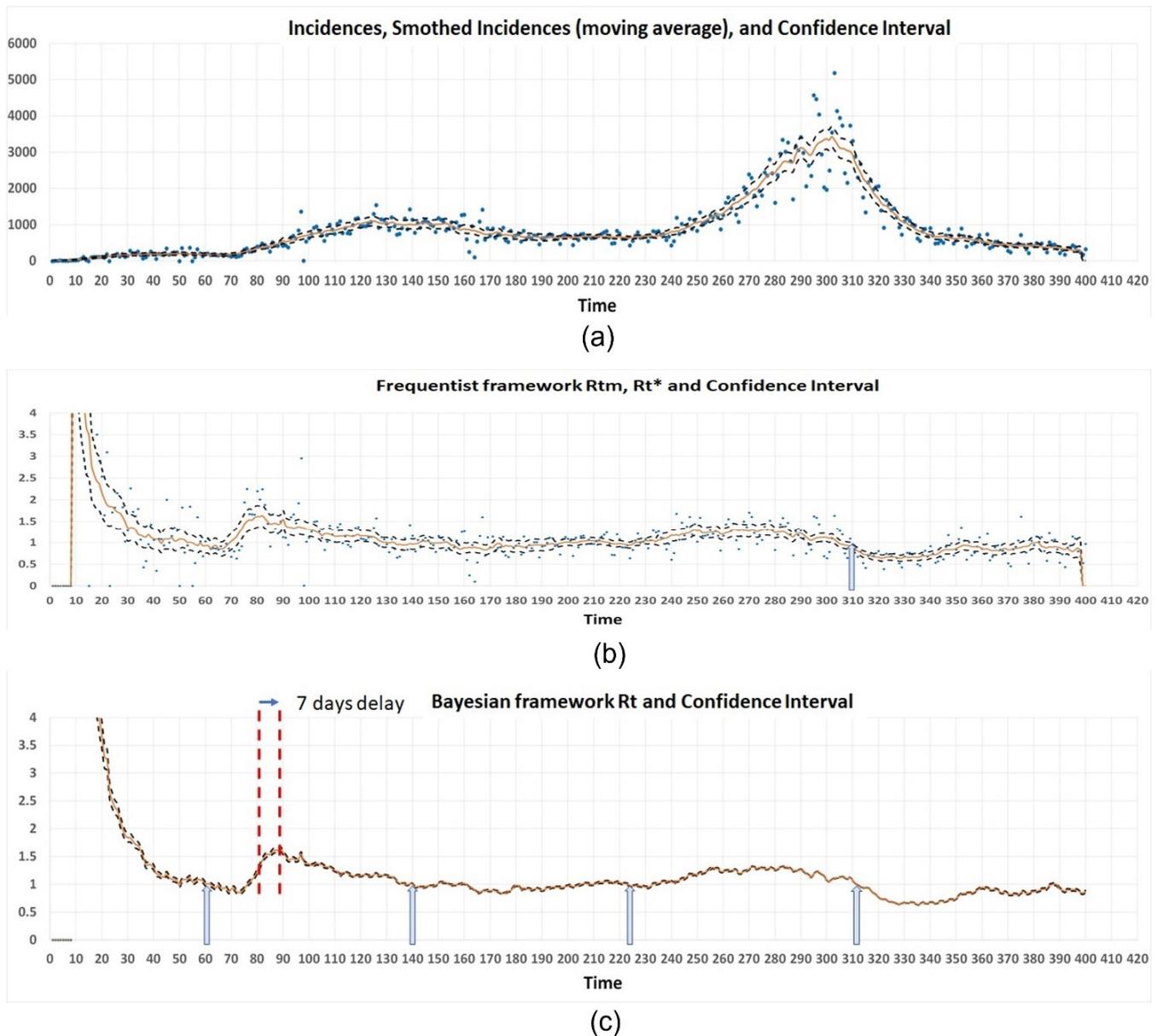


Figure 13. Partial data of the COVID-19 pandemic, Panama (9 March 2020–12 April 2021): (a) Incidence data series; (b) Frequentist framework R_t^m , with its 95% confidence interval; and (c) Bayesian framework R_t , with its 95% confidence interval.

4.6. Computational Aspects

The computational experiments were carried out on a computer with the following hardware characteristics: (i) OS: Windows 10 for 64 bits; (ii) RAM: 4 Gigabytes; and (iii) processor: Intel Core i3 (7th Gen) i3-7020U Dual-core, 2.30 GHz-4 GB DDR4 SDRAM. We used Microsoft Excel, from Office 2019, as our software platform. Here we implemented the theory of this paper with our software SeguiEpi, programmed with standard Excel instructions, including macros. This software is available at: <https://github.com/publiodariocortes/seguiEpi/> Size: 11,349 KB (accessed on 29 November 2021).

After downloading, the user must read the instructions in the worksheet “User manual”. As the software is an application of the theory for R_t calculation, there are two special modes, with their corresponding run times:

- Rt without serial interval uncertainty < 5 s.
- Rt with serial interval uncertainty < 60 s.

5. Conclusions

The results of this study are relevant, since epidemiological models are important to monitor the contagion of diseases and evaluate the effectiveness of health policies, and the measures adopted by governments to guarantee adequate hospital care. Estimating the instantaneous reproduction number is a key factor in detecting changes in disease transmission over time.

The method of calculating the confidence interval of the instantaneous reproduction number developed in this article is a significant change with respect to other proposed methods. The explicit formula for calculating R_t (Bayesian framework) produced a curve delayed with respect to the curve of R_t^m (our Frequentist framework). The delayed time was about a half of the time of moving average time (n_k). Nevertheless, it was not a large issue, because we can adjust our interpretation of the R_t , considering this delay. The situation differed with the *confidence interval* because we included all sources of variability. Therefore, the main contribution of this work is the proposition of an expression for the *confidence interval* of the instantaneous reproduction number, considering all sources of variability of the incidences. We defend our approach affirming it produced a 95% *confidence interval* much more appropriate for using it as a decision tool in controlling the evolution of a pandemic event, such as COVID-19. We hope the software (SeguiEpi) we are providing with the implementation of this theory can help in controlling the evolution of epidemics.

Although our results are encouraging, more tests are still needed with data from other countries and different infectious agents. In any case, it will be necessary to conduct comparative studies between the Bayesian frequentist method and our Frequentist method, such the comparison we made with the case of Panama, including the measures taken by the authorities, the population behavior, and other relevant situations, to find if there are connections with the evolution of the epidemics and for recommending improvements in the decision processes. Moreover, we must continue the investigation of *time series analysis* for decoupling the incidences in their model components and, in relation with this matter, it is necessary to conduct sensitivity analysis to determine the robustness of the R_t^m and its *confidence interval*, regarding the parameters defined for decoupling the incidences. This sensitivity analysis should be performed with simulated and actual incidence data, by using the software we already have. We have shown how to calculate an optimum confidence interval considering serial interval uncertainty, but we still need to find a way to develop a software to perform these tasks, so that we can replace our solution through the subrogated Gamma function. The solution for this matter is already conceptually completed with what we obtained with the summation of normal distribution at Appendix C; we only need to replace the summation of normal distribution with a summation of the Student’s t-test distribution. The only important problem to solve is the implementation of a software that automatically performs all the manual tasks we have in our current method. It will be a great advantage if we could automatically actualize the serial interval parameters by incorporating the new information collected in the local area; moreover, it will be necessary to consider the cases of infected people that come from other regions. This problem has already been solved from a point of view of the Bayesian framework, so we should begin by deeply understanding this approach. In parallel with the proposed investigations, we should perform an important revolution to the method of how to control the epidemics with the implementation of the Statistical Process Control (SPC). It is unquestionable that the controls applied to combat epidemics are a collection of processes. We implemented a special tool in the software we are proposing, a Control Chart designed to monitor the transformed error (Control Panel 3, SeguiEpi Excel software). In this direction, many

changes can be performed, including the critical thinking and organization of the Six Sigma Methodology.

Author Contributions: Conceptualization, P.D.C.-C.; methodology, P.D.C.-C.; formal analysis, P.D.C.-C., D.R.G.-C. and M.C.-M.; software, P.D.C.-C.; writing—original draft preparation, P.D.C.-C., D.R.G.-C. and M.C.-M.; writing—review and editing, P.D.C.-C., D.R.G.-C. and M.C.-M.; visualization, P.D.C.-C. and D.R.G.-C.; supervision, M.C.-M.; funding acquisition, M.C.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was made possible thanks to the support of the Sistema Nacional de Investigación (SNI) of Secretaría Nacional de Ciencia, Tecnología e Innovación (Panama).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available at <https://coronavirus.jhu.edu/about/how-to-use-our-data> (accessed on 23 December 2021), https://github.com/CSSEGISandData/COVID-19/blob/21b4a7275905738d6bb11627e5ffe76f79cb9b8b/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv#L210 (accessed on 23 December 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Modeling Incidence Series with Normal Parameters, Means, and the Standard Deviations

Estimating the model means of the incidences:

The method applied to calculate the Im_t means of the incidence series is the *centered moving average (without lag)* [21]. This method allows the Im_t mean to be calculated with incidence data from the past and future, resulting in a synchronized pattern with the original incidences I_t . The simple way to implement this method does not allow for the calculation of Im_t at the beginning and at the end of the observation period, because the number of I_t incidences considered (n_k) for the mean calculation should be constant. We propose a less constrained method that we have named *extended centered moving average*, which allows a flexible rule to define the number of incidences applied to calculate the mean. In our proposed method (A1), N is the total number of points in the series, n_k is the conventional moving average time defined by the analyst, and n_t is a dynamic moving average time calculated for the observation time t .

$$\begin{aligned}
 & \text{If } 1 < t < \frac{n_k + 1}{2} \Rightarrow n_t = 2(t - 1) + 1, \quad Im_t = \frac{1}{n_t} \sum_{j=1}^{n_t} I_j, \\
 & \text{If } \frac{n_k + 1}{2} \leq t \leq N - \frac{n_k + 1}{2} + 1 \Rightarrow n_t = n_k, \\
 & \quad Im_t \begin{cases} \frac{1}{n_t} \sum_{i=t-\frac{n_t-1}{2}}^{t+\frac{n_t-1}{2}} I_i & n_t \text{ odd} \\ \frac{\sum_{i=t-\frac{n_t}{2}}^{t+\frac{n_t}{2}-1} I_i + \sum_{i=t-\frac{n_t}{2}+1}^{t+\frac{n_t}{2}} I_i}{2n_t} & n_t \text{ even} \end{cases} \\
 & \text{If } N - \frac{n_k + 1}{2} + 1 < t < N \Rightarrow n_t = 2(N - t) + 1, \quad Im_t = \frac{1}{n_t} \sum_{j=1}^{n_t} I_{N-j+1} \quad (A1)
 \end{aligned}$$

The method described by (A1) was developed experimentally by observing the numerical pattern generated after defining an efficient data structure, easy for software implementation, with the property that only two t instants did not have means, $t = 1$ and $t = N$. Figure A1 shows one sample of the patterns, where we constructed it by setting $N = 20$ and $n_k = 9$.

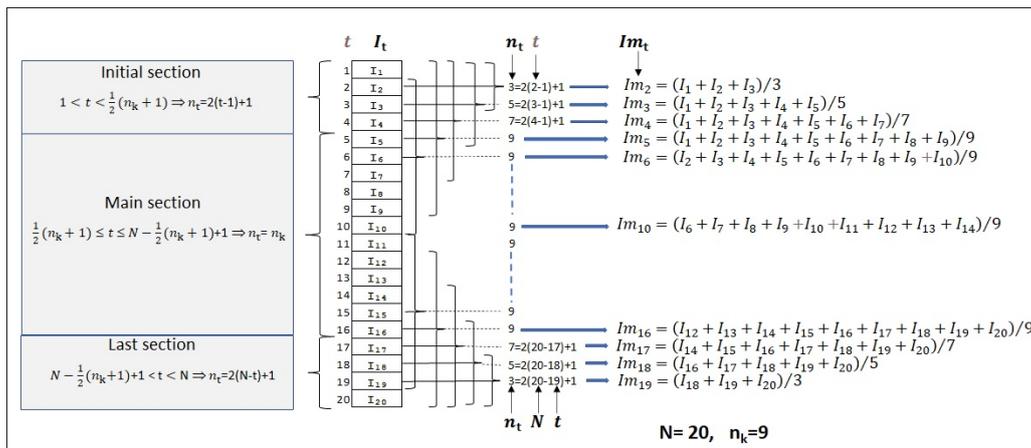


Figure A1. An example of a numerical pattern used to deduce the *extended centered moving average* method to calculate the Im_t mean series. The n_t values of the initial and last sections are sequences of odd numbers, less than n_k . The n_t value of the main section is n_k . There are no defined values for $t = 1$ and $t = N$.

As the maximum n_t values are in the *main section* (Figure A1), it is obvious that this section will have more precision in means calculation. This difference in precision will be considered in confidence interval calculation, where the n_t value is one of the applied parameters.

The analyst must take care of the importance of the moving average time (n_k) defined. Its value must be selected carefully to have Im_t values near to the population means. The software provides a set of tests and a diagram to help on this purpose:

- Randomness of sign test (applied to $e_t = I_t - Im_t$) should be passed.
- Normality test (applied to the *transformed error* e_t^* , below defined) should be passed.
- Histogram plot (applied to the *transformed error* e_t^*) should look like a normal pattern.
- Superimposed graphs of I_t and Im_t , where the Im_t curve should “navigate” through the I_t , revealing no aleatory patterns (for example, week periodicity).

The value of n_k assigned must not be too small to make imprecise the estimate of I_t^H , nor must it be too large to not allow the randomness of the signs of the errors around the mean.

Estimating the model standard deviation of the incidences:

The proposed incidence model is:

$$I_t = I_t^H + e_t, \quad e_t \sim N(0, \sigma_t^2) \tag{A2}$$

The standard deviation we are looking for is precisely the standard deviation of the error e_t . Figure A2 shows the scatter diagram of the error vs. the value of the estimator Im_t of one of the simulated time series of the incidence presented in the paper. Note that the upper contour of the error grows as the value of Im_t grows. Although this is a simulation series, the figure illustrates a typical reality situation.

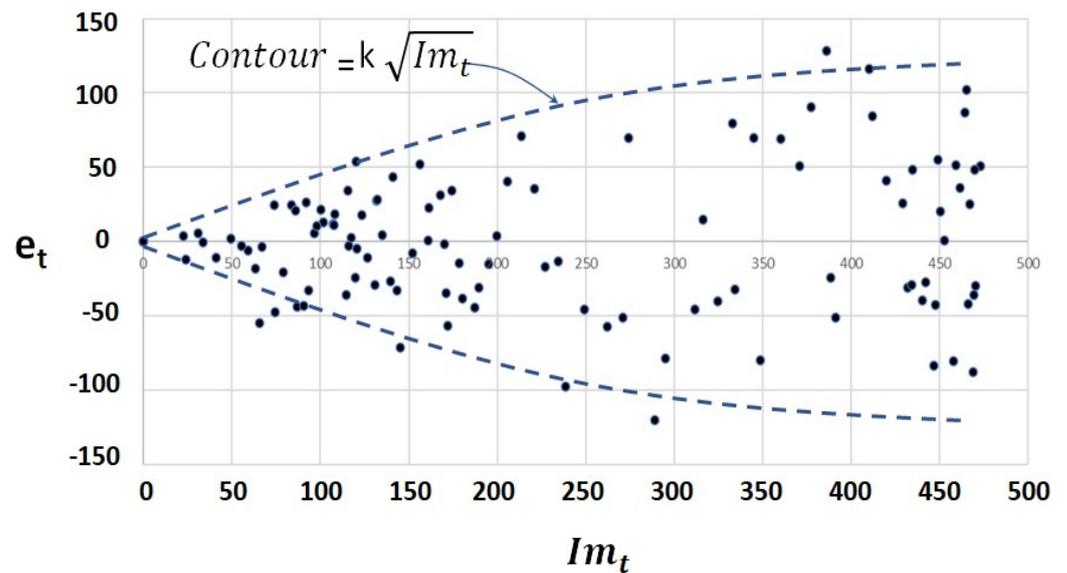


Figure A2. Scatter plot diagram of e_t vs. Im_t .

Following a standard procedure of time series analysis, we simply will sketch the contour of the dots pattern with a general equation:

$$Contour = k \times Im_t^\gamma \tag{A3}$$

The factor of interest in (A3) is Im_t^γ (power function), in which we do not know which is the most appropriate exponent γ . The usual practice to define γ is to divide the errors e_t by the factor Im_t^γ (A4), and test with different values of γ , until the scatter diagram of the transformed error e_t^* vs. Im_t forms a horizontal band (e_t^* becomes homoscedastic in respect to Im_t). All e_t^* values are assumed to be normally distributed at each time t , with a standard deviation denoted by σ_t^* . Thus:

$$e_t^* = \frac{e_t}{Im_t^\gamma} \sim N(0, \sigma_t^*) \tag{A4}$$

After some trials from $\gamma = 0$, it was found that for $\gamma = 0.5$, the scatter diagram e_t^* vs. Im_t assumes the desired horizontal band shape (Figure A3).

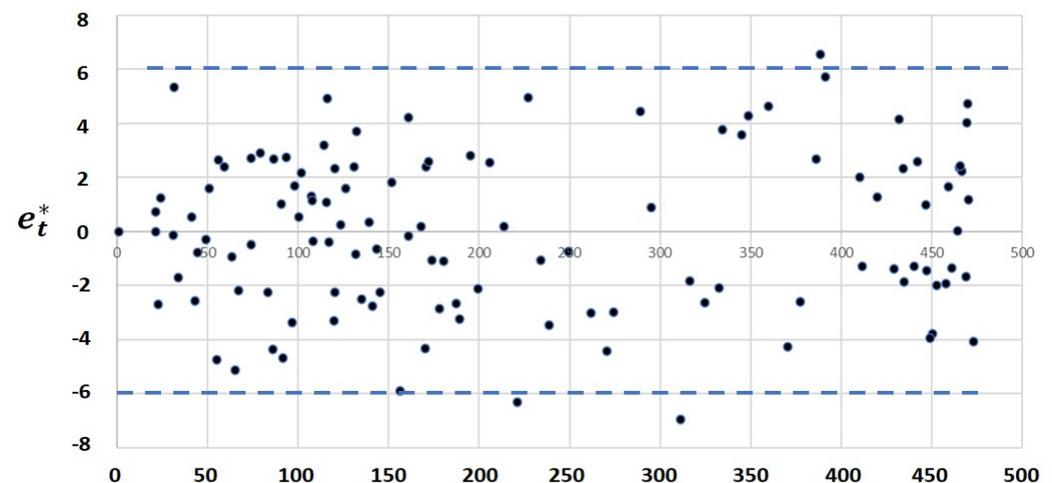


Figure A3. Scatter plot diagram of e_t^* vs. Im_t .

The shape of the error contour in counting processes, such as those of an epidemic, cannot always be modeled with a simple power function. There are other functions necessary to try; for example, logarithm, sine (positive increasing part), hyperbolic tangent, logistic, etc. All these functions share the quality that they are increasing (positive first derivative) in the range of their application. In this proposal, we selected a modified weighted average of the *power function* and the *logistic function*, which we named Logipow (A5).

$$\begin{aligned}
 \text{Logistic}(Im_t) &= \frac{1}{1+e^{4 \times (1-2a \times (\frac{Im_t}{MAX(Im_t)} + b))}} \\
 \text{AdLog}(Im_t) &= \frac{\text{Logistic}(Im_t) - \text{Logistic}(0)}{\text{Logistic}(MAX(Im_t)) - \text{Logistic}(0)} \quad * \text{ Adjusted Logistic.} \quad (A5) \\
 \text{Logipow} &= K \left[\beta \frac{Im_t^\gamma}{MAX(Im_t)^\gamma} + (1 - \beta)(B + (1 - B)\text{AdLog}(Im_t)) \right]
 \end{aligned}$$

The Logipow function is a kind of intuitive function with the special quality that it can be modulated with parameters to follow the contour with more flexibility than the standard power function. Next, we describe the parameter descriptions:

- a (slope factor): Provides the slope of the logistic function in its inflexion point.
- b (horizontal shift): Shifts the logistic function left ($b > 0$) and right ($b < 0$).
- B (vertical shift): Shift the logistic function up and down.
- β (balance): Defines the mix proportion between the logistic and power functions.
- K (scale): This factor does not alter the performance of the function. It provides visual help to the user of the software.

Figure A4 shows the general sections of one sample curve of the Logipow function.

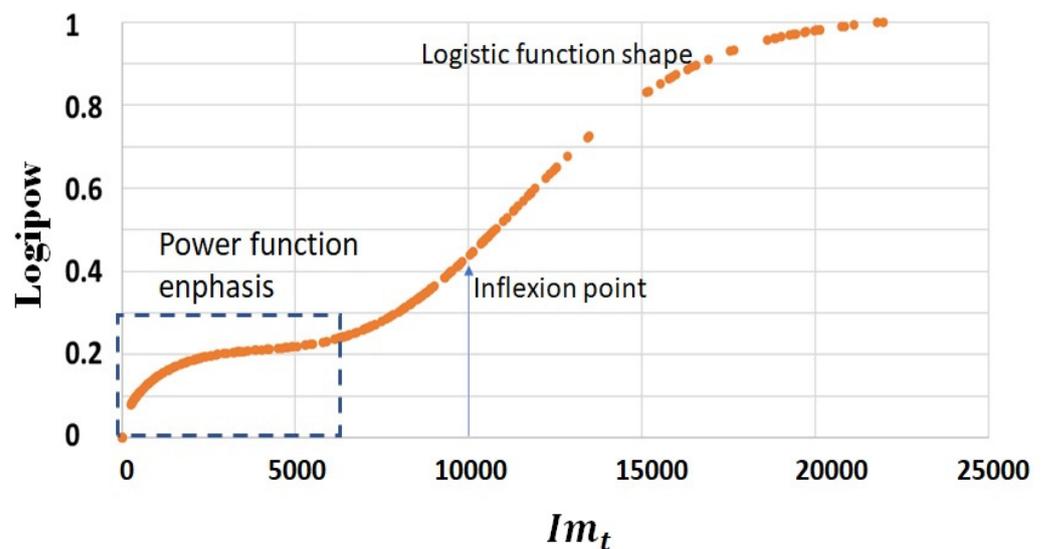


Figure A4. A sample of the Logipow function molded by parameters.

Coming back to the Equation (A4), we now change the *power function* with the more general function we are proposing:

$$e_t^* = \frac{e_t}{\text{Logipow}(Im_t)} \quad (A6)$$

Assuming that Im_t is statistically constant (for a practical simplification of the problem), the estimated standard deviation of e_t^* is straightforward calculated:

$$\hat{\sigma}_t^* = \frac{\hat{\sigma}_t}{\text{Logipow}(Im_t)} \quad (A7)$$

It is proposed two methods for the estimation of σ_t^* :

1. Assuming that σ_t^* is constant within time intervals: For fixed counting process within the time intervals.
2. Assuming that σ_t^* is continually changing over time: For counting processes that we know are changing but do not know when they change.

Assuming that σ_t^* is constant within time intervals:

Let M be the number of time sections of the series of e_t^* . Each section of size N_k (with $k = 1, 2, 3, \dots, M$). Let t_k be the starting instant of each section. The moving range average of the section k is:

$$\overline{RMe}_k^* = \frac{1}{N_k - 1} \sum_{t=t_k+1}^{t_k+N_k} |e_t^* - e_{t-1}^*|, \tag{A8}$$

The estimated standard deviation $\hat{\sigma}_k^*$ of each section is calculated by dividing \overline{RMe}_k^* by a conversion factor $d_2 = 1.128$.

$$\hat{\sigma}_k^* = \left(\frac{\overline{RMe}_k^*}{d_2} \right) = \left(\frac{\overline{RMe}_k^*}{1.128} \right) \tag{A9}$$

Within each section, $\hat{\sigma}_t^*$ will have a constant value of $\hat{\sigma}_k^*$ given by (A9).

This method can be applied to the great section of the entire epidemic time. This is a special option that can be referred to as *assuming that σ_t^* is constant all time*.

The calculation of the standard deviation through the moving range (A9) is the most accurate method when only one value is available at each measurement instant [22].

Assuming that σ_t^* is continually changing over time:

Let e_t^* be an N size series and all its moving ranges given by:

$$RMe_t^* = |e_t^* - e_{t-1}^*|, \quad 2 \leq t \leq N \tag{A10}$$

Taking the *centered moving average* \overline{RMe}_t^* of all these moving ranges, with moving average time n, for each instant t. The estimated standard deviation $\hat{\sigma}_t^*$ is calculated by dividing the moving average of the range \overline{RMe}_t^* by a conversion factor $d_2 = 1.128$.

$$\hat{\sigma}_t^* = \left(\frac{\overline{RMe}_t^*}{d_2} \right) = \left(\frac{\overline{RMe}_t^*}{1.128} \right) \tag{A11}$$

The work to calculate the estimated standard deviation, $\hat{\sigma}_t$, is summarized by calculating $\hat{\sigma}_t^*$ and then solving for $\hat{\sigma}_t$ in (A7):

$$\hat{\sigma}_t = \hat{\sigma}_t^* \text{Logipow}(Im_t) \tag{A12}$$

Finally, it is necessary to point out that parameter estimation is a key step in order to apply the method for calculating the R_t^m , and that we should try to achieve better results, but we will need to obtain experience in such a way that the conditions of normality do not apply too much excessive pressure. We have the intangible help of the Central Limit Theorem applied to the mean variables.

Appendix B. The Probability Density Function of the Noisy Reproduction Number R_t^*

Now we will deduce the probability density function of the *noisy reproduction number* R_t^* needed as a mathematical tool for testing procedures. R_t^* was defined as:

$$R_t^* = \frac{I_t}{\sum_{s=1}^t w_s I_{t-s}} \tag{A13}$$

Before the beginning of the analysis, it is necessary to clarify some important concepts. If the specific values of I_t and all I_{t-s} are known, it is possible to estimate their normal parameters in some way, for example, by the method of Appendix A. The effect of decoupling the I_t and all I_{t-s} in constants I_t^H and I_{t-s}^H , and gaussian noisy components e_t and

all e_{t-s} , is that they become statistically independent variables. The natural correlations are trespassed to the constant's components I_t^μ and I_{t-s}^μ , but it does not matter because they are constants (statistically frozen), and the errors components e_t and all e_{t-s} become independent (it is something that can be verified in the error series).

Continuing with (A13), it is observed that the divisor $\sum_{s=1}^t w_{t,s} I_{t-s}$ is a random variable, henceforth denoted Λ_t , which is normally distributed according to $N(\Lambda_t^\mu, \sigma_{\Lambda_t}^2)$. Λ_t^μ is straightforward calculated:

$$\Lambda_t^\mu = \sum_{s=1}^t w_s I_{t-s}^\mu \tag{A14}$$

Furthermore, considering that the incidences I_{t-s} of (A13) have their means and errors decoupled, the calculation of the variance of Λ_t can be performed knowing that the incidences I_{t-s} are independent variables. Thus:

$$\sigma_{\Lambda_t}^2 = \sum_{s=1}^t w_s^2 \sigma_{I_{t-s}}^2 \tag{A15}$$

With the above, and knowing R_t^* is a ratio (A13) in which the numerator I_t and the denominator Λ_t are independent, the probability density function $f(R_t^*)$ can be calculated following the general theoretical procedure [23]:

Let X and Y be independent continuous random variables, with pdfs $f_X(x)$ and $f_Y(y)$, respectively. Assume that X is zero for at most a set of isolated points. Let $W = Y/X$. Then:

$$f_W(w) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(wx) dx \tag{A16}$$

In summary, the probability density function of R_t^* can be calculated by applying (A16) to the ratio I_t/Λ_t , knowing that I_t and Λ_t are distributed according to $N(I_t^\mu, \sigma_{I_t}^2)$ and $N(\Lambda_t^\mu, \sigma_{\Lambda_t}^2)$, respectively:

$$f(R_t^*) = \frac{1}{2\pi\sigma_{\Lambda_t}\sigma_t} e^{-\frac{\sigma_t^2(\Lambda_t^\mu)^2 + \sigma_{\Lambda_t}^2(I_t^\mu)^2}{2\sigma_{\Lambda_t}^2\sigma_t^2}} \left(\frac{1}{\frac{\sigma_t^2 + \sigma_{\Lambda_t}^2 (R_t^*)^2}{2\sigma_{\Lambda_t}^2\sigma_t^2}} + \frac{\left(\frac{\sigma_t^2\Lambda_t^\mu + \sigma_{\Lambda_t}^2 R_t^* I_t^\mu}{\sigma_{\Lambda_t}^2\sigma_t^2}\right) \sqrt{\pi} e^{-\frac{\left(\frac{\sigma_t^2\Lambda_t^\mu + \sigma_{\Lambda_t}^2 R_t^* I_t^\mu}{\sigma_{\Lambda_t}^2\sigma_t^2}\right)^2}{\frac{\sigma_t^2 + \sigma_{\Lambda_t}^2 (R_t^*)^2}{2\sigma_{\Lambda_t}^2\sigma_t^2}}}}{2\left(\frac{\sigma_t^2 + \sigma_{\Lambda_t}^2 (R_t^*)^2}{2\sigma_{\Lambda_t}^2\sigma_t^2}\right)^{\frac{3}{2}}} \operatorname{erf}\left(\frac{\frac{\sigma_t^2\Lambda_t^\mu + \sigma_{\Lambda_t}^2 R_t^* I_t^\mu}{\sigma_{\Lambda_t}^2\sigma_t^2}}{2\sqrt{\frac{\sigma_t^2 + \sigma_{\Lambda_t}^2 (R_t^*)^2}{2\sigma_{\Lambda_t}^2\sigma_t^2}}}\right) \right) \tag{A17}$$

$\operatorname{erf}(z)$ is the error function: $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$.

Of course, (A17) is a very complex function we should try to simplify (which is performed in the paper).

Appendix C. R_t^m Confidence Interval Calculation with Serial Interval Uncertainty

Let us begin with a hypothetical problem whose solution will help us to solve the specific problem of the R_t^m confidence interval with serial interval uncertainty.

Suppose we have N normal variables x_j ($j = 1, 2, 3, \dots, N$), each of which with means and variances organized in pairs (μ_j, σ_j^2) . Furthermore, each x_j is related to M random values $x_{j,k}$ ($k = 1, 2, 3, \dots, M$). Thus, the values $x_{j,k}$ are the elements of a matrix:

$$\begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,M} \\ x_{2,1} & x_{2,2} & \dots & x_{2,M} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N,1} & x_{N,2} & \dots & x_{N,M} \end{bmatrix} \Rightarrow \begin{matrix} x_1 \sim N(\mu_1, \sigma_1^2) \\ x_2 \sim N(\mu_2, \sigma_2^2) \\ \dots \dots \dots \dots \dots \dots \\ x_N \sim N(\mu_N, \sigma_N^2) \end{matrix} \tag{A18}$$

From this matrix perspective, the specific problem we will try to solve is described as follows:

Conditions:

1. The values of the means μ_j and the variances σ_j^2 are known and come from an unspecified random process.
2. The values $x_{j,k}$ of the row j of this matrix are random versions of the variable $x_j \sim N(\mu_j, \sigma_j^2)$.
3. There is a global variable that we will call x , whose random values are obtained from the matrix $[x_{j,k}]$ and that has a population mean μ and variance σ^2 . The corresponding probability density function $g(x)$ is unknown.

Question:

Find the pdf $g(x)$, and calculate the mean and the variance of the variable x .

Solution:

Since we are assuming that we know the N random versions of the means and variances corresponding to the variables x_j , the calculation of the mean of x is immediate:

$$\mu = \frac{\sum_{j=1}^N \mu_j}{N} \tag{A19}$$

Now this result (A19) can be used to look at the problem from a different perspective. Since the mean parameters μ and μ_j are the expected values of the random variables x and x_j , respectively, equation (A19) can be presented in more detail:

$$E(x) = \frac{1}{N} \sum_{j=1}^N E(x_j) = \frac{1}{N} \sum_{j=1}^N \int_{x_j \rightarrow -\infty}^{x_j \rightarrow +\infty} x_j e^{-\frac{1}{2}(\frac{x_j - \mu_j}{\sigma_j})^2} \frac{dx_j}{\sqrt{2\pi}\sigma_j} \tag{A20}$$

Since all the variables x_j vary within the same range $(-\infty, +\infty)$, the result of the N integrations will not be affected if we make all variables x_j vary, which implies that they are all equal to the same variable, which we will provisionally call y . Thus:

$$E(x) = \frac{1}{N} \sum_{j=1}^N \int_{-\infty}^{+\infty} y e^{-\frac{1}{2}(\frac{y - \mu_j}{\sigma_j})^2} \frac{dy}{\sqrt{2\pi}\sigma_j} \tag{A21}$$

Rearranging factors:

$$E(x) = \int_{-\infty}^{+\infty} y \frac{1}{N} \sum_{j=1}^N \frac{e^{-\frac{1}{2}(\frac{y - \mu_j}{\sigma_j})^2}}{\sqrt{2\pi}\sigma_j} dy \tag{A22}$$

It is noted that, in this way, the expected value of x is not affected:

$$E(x) = \int_{-\infty}^{+\infty} y \frac{1}{N} \sum_{j=1}^N \frac{e^{-\frac{1}{2}(\frac{y-\mu_j}{\sigma_j})^2}}{\sqrt{2\pi}\sigma_j} dy = \frac{\sum_{j=1}^N \mu_j}{N} = \mu \tag{A23}$$

Since the underlined part of the above equation meets all the requirements to be a probability density function (always positive function and the total area under the curve is equal to 1), and the calculated expected value is equal to the value expected of the variable x , we can consider the possibility that $x = y$. Thus, the probability density function of x could be:

$$g(x) = \frac{1}{N} \sum_{j=1}^N \frac{e^{-\frac{1}{2}(\frac{x-\mu_j}{\sigma_j})^2}}{\sqrt{2\pi}\sigma_j} = \frac{1}{N\sqrt{2\pi}} \sum_{j=1}^N \frac{e^{-\frac{1}{2}(\frac{x-\mu_j}{\sigma_j})^2}}{\sigma_j} \tag{A24}$$

If the above equation is right, the variance could be calculated:

$$V(x) = \frac{1}{N\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x - \mu)^2 \sum_{j=1}^N \frac{e^{-\frac{1}{2}(\frac{x-\mu_j}{\sigma_j})^2}}{\sigma_j} dx \tag{A25}$$

Solving the integral:

$$V(x) = \sum_{j=1}^N \left[\frac{\sigma_j^2 + (\mu_j - \mu)^2}{N} \right] = \frac{1}{N} \sum_{j=1}^N [\sigma_j^2 + (\mu_j - \mu)^2] \text{ (Jumping to the solution)} \tag{A26}$$

The above hypothetical problem and its solution is the same as our problem of finding the probability density function of the *instantaneous reproduction number R_t^m with serial interval uncertainty*. We only need to make the following variables redefinitions:

$$\mu_j \rightarrow R_{t,j}^m = \frac{Im_t}{\Lambda_{t,j}^m}, \quad \sigma_j \rightarrow \hat{\sigma}_{R_t^m}^{(j)} = \frac{\hat{\sigma}_t}{\Lambda_{t,j}^m \sqrt{n_t}}, \quad \Lambda_{t,j}^m = \sum_{s=1}^t w_s^{(j)} Im_{t-s} \tag{A27}$$

The general mean and standard deviation are:

$$\mu = \frac{\sum_{j=1}^N \mu_j}{N} \rightarrow R_t^m = \frac{\sum_{j=1}^N R_{t,j}^m}{N} \tag{A28}$$

$$\sqrt{V(x)} = \sqrt{\frac{1}{N} \sum_{j=1}^N [\sigma_j^2 + (\mu_j - \mu)^2]} \rightarrow \hat{\sigma}_{R_t^m} = \sqrt{\frac{1}{N} \sum_{j=1}^N \left[\left(\hat{\sigma}_{R_t^m}^{(j)} \right)^2 + \left(R_{t,j}^m - R_t^m \right)^2 \right]} \tag{A29}$$

Finally, the probability density function of R_t^m is:

$$g(x) = \frac{1}{N\sqrt{2\pi}} \sum_{j=1}^N \frac{e^{-\frac{1}{2}(\frac{x-\mu_j}{\sigma_j})^2}}{\sigma_j} \rightarrow f(R_t^m) = \frac{1}{N\sqrt{2\pi}} \sum_{j=1}^N \frac{e^{-\frac{1}{2}\left(\frac{R_t^m - R_{t,j}^m}{\hat{\sigma}_{R_t^m}^{(j)}}\right)^2}}{\hat{\sigma}_{R_t^m}^{(j)}} \tag{A30}$$

Equation (A30) alone is sufficient to statistically describe R_t^m , but unfortunately, in the calculus of *confidence intervals*, it results in the computational processes that need to be run for each instant t taking a long time to be completed. Therefore, instead of using (A30) we prefer to use a surrogate probability density function such as the Gamma distribution. In Figure A5, we compare the limits of the 95% confidence interval obtained by applying the exact function (A30) and a surrogated Gamma distribution with means and standard deviation given by (A28) and (A29), respectively. The data of the series of incidence are the same as those used in the paper.

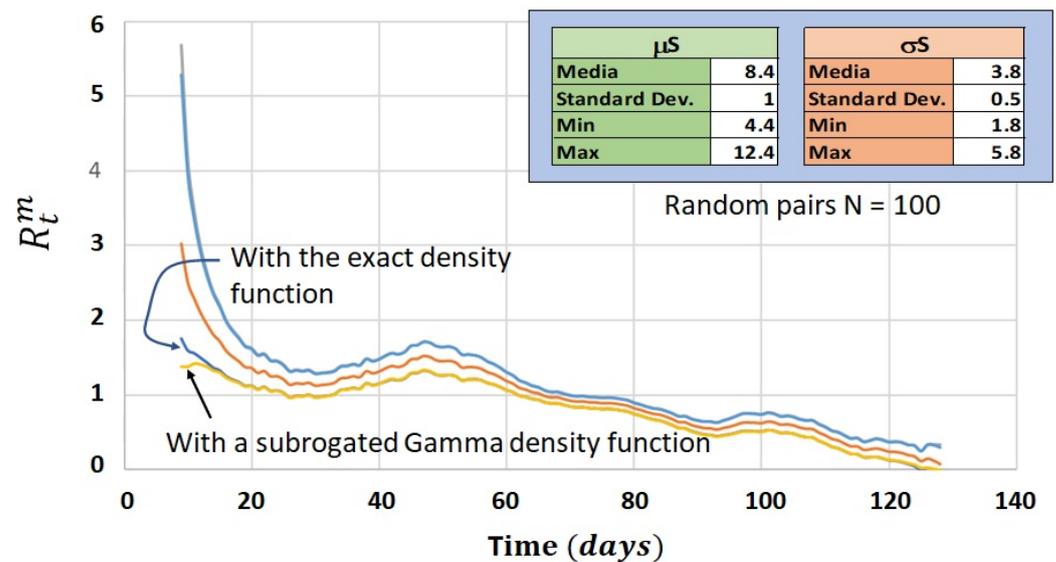


Figure A5. Superposition of 95% confidence intervals calculated with the exact probability density function and a subrogated Gamma density function.

References

- Gostic, K.M.; McGough, L.; Baskerville, E.B.; Abbott, S.; Joshi, K.; Tedijanto, C.; Kahn, R.; Niehus, R.; Hay, J.A.; De Salazar, P.M.; et al. Practical Considerations for Measuring the Effective Reproductive Number, R_t . *PLoS Comput. Biol.* **2020**, *16*, e1008409. [\[CrossRef\]](#)
- Avram, F.; Adenane, R.; Ketcheson, D.I. A Review of Matrix SIR Arino Epidemic Models. *Mathematics* **2021**, *9*, 1513. [\[CrossRef\]](#)
- Hussain, S.; Madi, E.; Khan, H.; Etemad, S.; Rezapour, S.; Sitthiwirattam, T.; Patanarapeelert, N. Investigation of the Stochastic Modeling of COVID-19 with Environmental Noise from the Analytical and Numerical Point of View. *Mathematics* **2021**, *9*, 3122. [\[CrossRef\]](#)
- Alonso-Quesada, S.; De la Sen, M.; Nistal, R. An SIRS Epidemic Model Supervised by a Control System for Vaccination and Treatment Actions Which Involve First-Order Dynamics and Vaccination of Newborns. *Mathematics* **2022**, *10*, 36. [\[CrossRef\]](#)
- Petermann, M.; Wyler, D. A Pitfall in Estimating the Effective Reproductive Number R_t for COVID-19. *Swiss Med. Wkly.* **2020**, *150*, w20307.
- Ganasegeran, K.; Ch'ng, A.S.H.; Looi, I. What Is the Estimated COVID-19 Reproduction Number and the Proportion of the Population That Needs to Be Immunized to Achieve Herd Immunity in Malaysia? A Mathematical Epidemiology Synthesis. *COVID* **2021**, *1*, 3. [\[CrossRef\]](#)
- Na, J.; Tibebe, H.; De Silva, V.; Kondoz, A. Probabilistic Approximation of Effective Reproduction Number of COVID-19 Using Daily Death Statistics. *Chaos Solitons Fractals* **2020**, *140*, 110181. [\[CrossRef\]](#) [\[PubMed\]](#)
- Knight, J.; Mishra, S. Estimating Effective Reproduction Number Using Generation Time versus Serial Interval, with Application to Covid-19 in the Greater Toronto Area, Canada. *Infect. Dis Model.* **2020**, *5*, 889–896. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bacaër, N. *A Short History of Mathematical Population Dynamics*; Springer: Berlin/Heidelberg, Germany, 2011.
- Kermack, W.O.; McKendrick, A.G. A Contribution to the Mathematical Theory of Epidemics. *Proc. R Soc. London Ser. A Contain. Pap. A Math. Phys. Character* **1927**, *115*, 700–721. [\[CrossRef\]](#)
- Peterson, J.D.; Adhikari, R. Efficient and Flexible Methods for Time since Infection Models. *arXiv* **2019**, arXiv:2010.10955. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fraser, C. Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic. *PLoS ONE* **2007**, *2*, e758. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wallinga, J.; Teunis, T. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. *Am. J. Epidemiol.* **2004**, *160*, 509–516. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fine, P.E.M. The Interval between Successive Cases of an Infectious Disease. *Am. J. Epidemiol.* **2003**, *158*, 1039–1047. [\[CrossRef\]](#) [\[PubMed\]](#)
- Svensson, M. A Note on Generation Times in Epidemic Models. *Math. Biosci.* **2007**, *208*, 300–311. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cori, A.; Ferguson, N.M.; Fraser, C.; Cauchemez, S. Practice of Epidemiology/A New Framework and Software to Estimate Time-Varying Reproduction Numbers during Epidemics. *Am. J. Epidemiol.* **2013**, *178*, 1505–1512. [\[CrossRef\]](#) [\[PubMed\]](#)
- Thompson, R.; Stockwind, J.; Van Gaalene, R.; Polonsky, J.; Kamvarg, Z.; Demarsh, P.; Dahlgvist, E.; Lij, S.; Miguelk, E.; Jombartg, T.; et al. Improved Inference of Time-Varying Reproduction Numbers during Infectious Disease Outbreaks. *Epidemics* **2019**, *29*, 100356. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cabras, S. A Bayesian-Deep Learning Model for Estimating Covid-19 Evolution in Spain. *Mathematics* **2021**, *9*, 2921. [\[CrossRef\]](#)

19. Xu, J.; Tang, Y. Mathematics Bayesian Framework for Multi-Wave COVID-19 Epidemic Analysis Using Empirical Vaccination Data Framework for Multi-Wave. *Mathematics* **2022**, *10*, 22. [[CrossRef](#)]
20. Zhao, S.; Cao, P.; Gao, D.; Zhuang, Z.; Cai, Y.; Ran, J.; Chong, M.K.C.; Wang, K.; Lou, Y.; Wang, W.; et al. Serial Interval in Determining the Estimation of Reproduction Number of the Novel Coronavirus Disease (COVID-19) during the Early Outbreak. *J. Travel Med.* **2020**, *27*, 1–3. [[CrossRef](#)]
21. Wheelwright, S.; Makridakis, S.; Hyndman, R. *Forecasting: Methods and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 1998.
22. Montgomery, D. *Introduction to Statistical Quality Control*, 4th ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001.
23. Larsen, R.; Marx, M. *An Introduction to Mathematical Statistics*, 5th ed.; Prentice Hall/Pearson: Hoboken, NJ, USA, 2012.