

# Los Procesos de Ramificación y su Aplicación al Estudio de la Transmisión de la COVID-19 en Cuba

## Branching Processes and their Application to the Study of Transmission of COVID-19 in Cuba

Marelys Crespo Navas<sup>1\*</sup>, Pedro Pablo de la Cruz Cobas<sup>2</sup>, Ernesto Quevedo Caballero<sup>2</sup>, José E. Valdés Castro<sup>1</sup>

**Resumen** El presente trabajo se dedica al estudio de la transmisión de la Covid-19 en Cuba, utilizando los procesos de ramificación para su modelación. Estos procesos tienen especial interés en la etapa inicial de una epidemia o cuando ocurre un rebrote. En el estudio se utiliza el proceso de Galton-Watson y un proceso de ramificación, el cual denominamos Epi, semejante al proceso de Bellman-Harris, que modela el comportamiento de la epidemia.

Para realizar simulaciones del proceso Epi, se estimó el número medio de individuos contagiados por una persona infectada, cuyo valor empírico es el llamado índice de reproducción básico  $R_0$  en epidemiología, y el tiempo de generación se ajustó mediante una distribución Gamma.

Además, se ilustran con ejemplos resultados asociados al número medio de individuos que se infectan en un instante de tiempo y el número medio de infectados hasta ese instante, la probabilidad de que la transmisión se extinga o se haya extinguido al cabo de un tiempo, el tiempo hasta que la transmisión se extingue con una probabilidad dada, y el tiempo medio hasta su extinción y la probabilidad de extinción de la misma.

**Abstract** The present work is dedicated to the study of transmission of Covid-19 in Cuba, using the branching processes for his modeling. These processes are of especial interest in the initial state of epidemic or when a re-outbreak occurs. The Galton Watson process and a branching process, which we call Epi, similar to the Bellman-Harris process, are used in this study to model the behavior of the epidemic.

To carry out the simulations of the Epi process, the mean number of infected individuals by an infector was estimated, the empiric value of which is the so-called basic reproduction number  $R_0$  in epidemiology, and the generation time was adjusted by a Gamma distribution.

In addition, we illustrate with examples results referred to the mean number of individuals who became infected in an instant of time and the mean number of infected until that moment, the probability that transmission will be extinguished or extinguished after a time, the time until the transmission is extinguished with a given probability, and the mean time until its extinction and the probability of its extinction.

### Palabras Clave

Covid-19, Procesos de Ramificación, Simulación

### Keywords

Covid-19, Branching Processes, Simulation

<sup>1</sup> Departamento de Matemática Aplicada, Universidad de La Habana, La Habana, Cuba, marelys@matcom.uh.cu, vcastro@matcom.uh.cu

<sup>2</sup> Universidad de La Habana, La Habana, Cuba, ppablocrucobas@gmail.com, ernestquevcab@gmail.com

\*Autor para Correspondencia, Corresponding Author

## Introducción

La idea general de los procesos de ramificación es la siguiente. En algún instante de tiempo se identifica un conjunto de individuos que forman la generación 0. Esta generación produce descendientes, cuya colección forma la generación 1. Los descendientes de la generación 1 constituyen la generación 2, y así sucesivamente.

Estos procesos modelan la evolución en el tiempo de procesos con estructuras tipo árbol, las cuales pueden encontrarse en diversas áreas de la ciencia: dinámica de población, física, biología, computación.

El primer estudio sobre procesos de ramificación se realizó en el siglo XIX, y estuvo vinculado al problema de la extinción de apellidos en las familias. Posteriormente este modelo fue nombrado *proceso de Galton-Watson*, en alusión a las dos

personas que iniciaron su estudio. [6] fue el primer libro dedicado a los procesos de ramificación.

Nuestro interés en este tipo de procesos consiste en su aplicación a los estudios sobre epidemias. Parece que uno de los primeros trabajos dedicados a la aplicación de los procesos de ramificación en problemas de epidemias fue [?].

Los procesos de ramificación tienen especial interés en los estudios sobre epidemias cuando la transmisión de la enfermedad se encuentra en una etapa inicial (en la cual las medidas de aislamiento y control aún son débiles), u ocurre un rebrote, ver por ejemplo, la Sección 9.3 de [4] y [8].

En [1] y [10], se realizan estudios sobre la transmisión inicial de la enfermedad Covid-19 en China, en los cuales, entre otros modelos, se utilizan procesos de ramificación. [7] es otro trabajo sobre la transmisión inicial de la enfermedad en China, donde se modela mediante simulación el comportamiento de la misma.

En los procesos de ramificación es esencial la *ley de reproducción*,  $p_k = P(Y = k)$ ,  $k \geq 0$ , de la variable aleatoria  $Y$  que representa el número de individuos a los cuales directamente infecta un individuo ya infectado.

Como en todos los modelos para el estudio de epidemias, una característica esencial en el modelo de Galton-Watson es el valor medio  $m = EY$  de la variable  $Y$ , cuyo valor aproximado empírico es el llamado *índice de reproducción básico*  $R_0$  en epidemiología.

El proceso de Galton-Watson describe la reproducción de individuos infectados a través de generaciones sucesivas. Cuando se requiere una descripción no por generaciones, sino en el tiempo, es necesario tener en cuenta la distribución de una variable aleatoria  $V$  que representa el *tiempo de generación*, es decir, el tiempo que transcurre desde que un individuo es infectado hasta que este infecta a otro. Esto da lugar a procesos de ramificación en tiempo continuo que contienen un proceso de Galton-Watson inmerso. Un proceso de este tipo es el denominado *proceso de Bellman-Harris*.

Los procesos a los cuales hemos hecho referencia permiten hacer estimaciones sobre los siguientes índices:

1. Número medio de individuos que se infecta en un instante dado  $t$ , y número medio de infectados hasta ese instante.
2. Probabilidad de que la transmisión se extinga o se haya extinguido al cabo de un tiempo dado.
3. Tiempo al cabo del cual la transmisión puede extinguirse, con una probabilidad dada.
4. Tiempo medio hasta la extinción y probabilidad de extinción de la transmisión.

En nuestro estudio, los puntos 2, 3 y 4 anteriores se examinan a través de fórmulas conocidas correspondientes a los modelos utilizados, y el punto 1 por medio de la simulación. Además, usando la simulación se exploran diferentes escenarios de propagación de la enfermedad, los cuales podrían ser

utilizados en el diseño de las estrategias de prevención de la misma.

En la Sección 1 se define el proceso de Galton-Watson y se presentan algunas fórmulas y ejemplos que ilustran su uso. La Sección 2 se dedica a las cotas y aproximaciones que permiten realizar estimaciones cuando no se conoce por completo la ley de reproducción de la variable  $Y$ . En la Sección 3 se describen los dos procesos en tiempo continuo de interés en este estudio.

En la Sección 4 se presentan los estimadores del parámetro  $m$ , y también de  $R_0$  y el índice efectivo de reproducción  $R_t$ . En la Sección 5 se realizan simulaciones para la comparación mediante gráficos de los dos modelos considerados, y para ilustrar el comportamiento del número medio de individuos infectados según la variación del parámetro  $m$ . La Sección 6 está dedicada a la validación del modelo y de los programas de simulación. En esta sección se estiman el parámetro  $m$ , y la distribución de los tiempos de generación según los datos de Cuba y de la provincia La Habana, y se compara el resultado de la simulación con la curva real correspondiente al acumulado de casos en la provincia, en el periodo de 30 días a partir del 16 de marzo, fecha en la que se confirmó el primer paciente en esta provincia.

Las definiciones y fórmulas básicas que se utilizarán en las próximas secciones se pueden consultar en [2], [5] y [6]. Cuando es necesario se citan también otras referencias.

## 1. Proceso de Galton-Watson

Asumiremos  $p_0 > 0$ . Denotemos por  $X_n$  la variable aleatoria que representa la población de individuos infectados correspondientes a la generación  $n$ ,  $n = 0, 1, \dots$ . Sean  $Y_1, Y_2, \dots$  variables aleatorias independientes con igual distribución que  $Y$ . Observe que

$$X_n = Y_1 + \dots + Y_{X_{n-1}}, \quad n \geq 1,$$

si  $X_{n-1} > 0$ . El proceso  $\{X_n\}$ ,  $n \geq 0$ , se denomina *proceso de Galton-Watson*.

Supongamos que  $X_0$  es no aleatorio. El número medio de infectados de la generación  $n$  se calcula por la fórmula

$$EX_n = X_0 \cdot m^n, \quad n \geq 0.$$

Sea  $Z_n$  el total de infectados hasta la generación  $n$ , entonces, como  $Z_n = \sum_{i=0}^n X_i$ , se tiene que

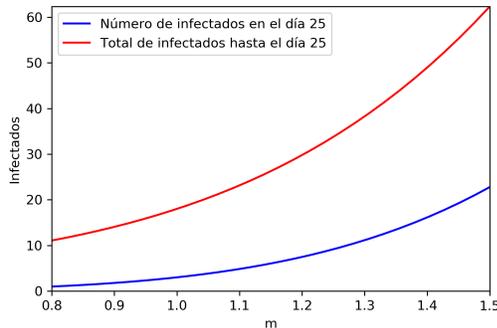
$$EZ_n = X_0 \sum_{i=0}^n m^i, \quad n \geq 1.$$

Las dos fórmulas anteriores son de gran interés práctico, pues aunque se refieren a números medios de infectados en una generación y hasta una generación, respectivamente, y no en el tiempo real, muestran la magnitud de la epidemia para un valor dado de  $m$ .

La relación aproximada entre el tiempo real y las generaciones es la siguiente. Sea  $(0, t_f)$  el intervalo de tiempo en el cual se examina la transmisión de la enfermedad. Denotemos

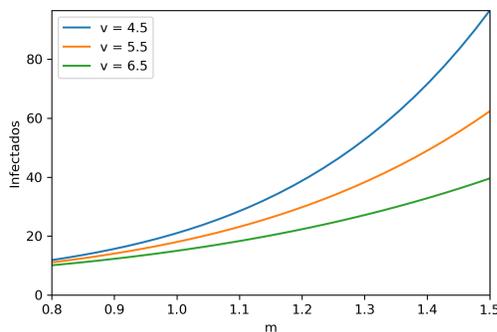
por  $\bar{v}$  el tiempo medio de generación. Entonces el número aproximado  $n$  de generaciones en el intervalo  $(0, t_f)$  se puede tomar como el número entero que redondea el valor de  $t_f/\bar{v}$ .

**Ejemplo 1** Sean  $\bar{v} = 5$ ,  $X_0 = 3$  y  $n = 5$ . Como  $t_f = \bar{v} \times n = 5 \times 5 = 25$ , entonces los valores de  $EX_5$  y  $EZ_5$  constituyen valores aproximados del número de infectados en el instante  $t_f = 25$  y al cabo de  $t_f = 25$  días, respectivamente. Los valores correspondientes se muestran en la Figura 1 para  $m \in [0,8; 1,5]$ .



**Figura 1.** Cantidad media aproximada de individuos infectados en el día 25 y hasta el día 25.

**Ejemplo 2** Sean  $X_0 = 3$  y  $t_f = 25$ . Para los valores  $\bar{v} = 4,5; 5,5$  y  $6,5$ , usando la fórmula  $n = t_f/\bar{v}$ , se obtienen los valores aproximados  $n = 6, 5$  y  $4$ , respectivamente, los cuales corresponden a  $t_f = 25$  días. Para este tiempo se obtiene el total aproximado de individuos infectados para  $m \in [0,8; 1,5]$ . Los resultados se muestran en la Figura 2.



**Figura 2.** Cantidad media aproximada de individuos infectados hasta el día 25.

Sea  $V(Y) = \sigma^2$ . La varianza de  $X_n$  se calcula por la fórmula (ver [5], pag. 20),

$$V(X_n) = \sigma^2 X_0 \frac{m^n(m^n - 1)}{m(m - 1)}.$$

Nótese que cuando  $m > 1$ , el valor medio  $EX_n$  y la varianza  $V(X_n)$  tienen un crecimiento exponencial como función del número de generaciones  $n$ .

Una característica importante del proceso de Galton-Watson es el número de la generación  $T$  hasta la extinción de la epidemia, el cual se define

$$T = \inf\{n \geq 1 : X_n = 0\}.$$

Otra característica de gran importancia es la probabilidad de extinción de la población de infectados. Sea  $Q_n$  la probabilidad de que la epidemia se extinga en la generación  $n$  o antes, es decir,  $Q_n = P(X_n = 0)$ . Puede demostrarse que

$$Q = \lim_n Q_n.$$

Como es conocido, si  $m < 1$ , entonces  $ET < \infty$  y  $Q = 1$ . Es decir, en este caso el número medio de generaciones hasta la extinción de la epidemia es finito, y esta se extingue con probabilidad 1.

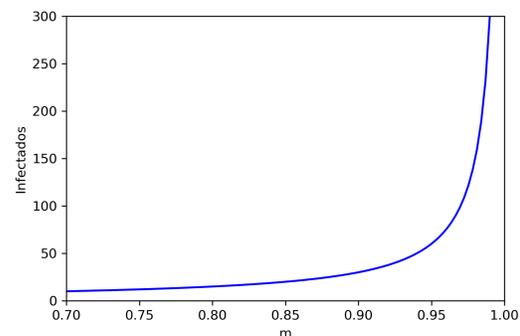
Supongamos que  $m < 1$ , y sea  $N$  el total de individuos infectados hasta la extinción de la epidemia. Si  $X_0 = 1$ , entonces,

$$N = 1 + N_1 + \dots + N_Y,$$

donde  $N_1, N_2, \dots$  son v.a. independientes con igual distribución que  $N$ . Es decir, el número de individuos infectados en total es igual a 1 (el individuo original) más la cantidad de individuos infectados, hasta la extinción, correspondiente a cada uno de los descendientes inmediatos del individuo infectado inicial (estos son  $Y$ ). En general, se halla que

$$EN = \frac{X_0}{1 - m}. \tag{1}$$

**Ejemplo 3** De la fórmula (1) se obtienen los valores de  $EN$  para  $X_0 = 3$  y  $m \in [0,7; 1)$ . Los resultados se muestran en la Figura 3.



**Figura 3.** Cantidad media de infectados hasta la extinción de la epidemia.

Muchas veces se asume como distribución de la variable  $Y$  la distribución de Poisson con parámetro  $m$ . Esta suposición podría justificarse intuitivamente de la siguiente manera. Si el número de individuos de una población es relativamente grande, entonces cada individuo infectado puede infectar a cualquier otro miembro de la población con una probabilidad

pequeña, y de manera aproximadamente independiente. Luego, usando la aproximación de la distribución binomial a la distribución de Poisson, en una etapa inicial, el número de infecciones que causa cada individuo infectado tiene aproximadamente distribución de Poisson. La ventaja en el uso de esta distribución es que depende solo del parámetro  $m$ .

Sea  $X_0 = 1$ . Si  $m > 1$ , la probabilidad de extinción  $Q$  es la solución diferente de 1 de la ecuación

$$z = g(z), \tag{2}$$

donde  $g(z)$ , con  $0 \leq z \leq 1$ , es la función generadora de probabilidades correspondiente a la variable  $Y$ . Para la mayoría de las distribuciones de  $Y$  es necesario hallar la solución numérica de la ecuación (2). Por ejemplo, si  $Y$  tiene distribución de Poisson con parámetro  $m$ , debe hallarse la solución de la ecuación

$$z = e^{-m(1-z)}. \tag{3}$$

Note que cuando  $X_0 = L$ , la probabilidad de extinción de la epidemia es igual a  $Q^L$ . Entonces, para  $m > 1$ , el riesgo  $r$  de un brote de gran magnitud en una localidad, cuando se importan  $L$  casos a la misma, se mide por la probabilidad de no extinción sin intervención ([10], pag.4).

$$r = 1 - Q^L. \tag{4}$$

Por otro lado, la probabilidad  $r_n$  de que la epidemia se extinga después de la generación  $n$  o no se extinga, se calcula por la fórmula

$$r_n = 1 - Q_n^L. \tag{5}$$

Los valores de  $Q_n$  se calculan usando la fórmula (2) de manera recurrente

$$Q_{n+1} = g(Q_n), \quad n \geq 0, \tag{6}$$

donde  $Q_0 = 0$ .

**Ejemplo 4** En la Figura 4 se muestran los valores del riesgo  $r$  como función de  $m$ , para  $L = 1, 3$  y  $5$ , cuando  $Y$  tiene distribución de Poisson.

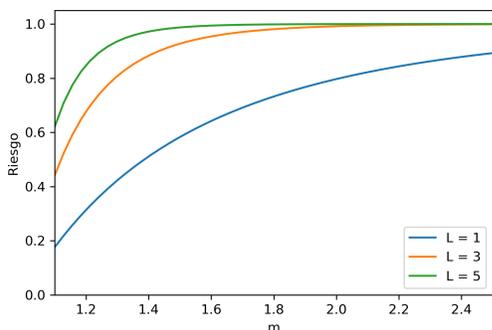


Figura 4. Riesgo  $r$  como función de  $m$ .

**Ejemplo 5** En la Figura 5 se muestran los valores de  $r_n$  en función  $n$ , para  $L = 1, 3$  y  $5$ , cuando  $Y$  tiene distribución de Poisson con parámetro  $m = 1,5$ .

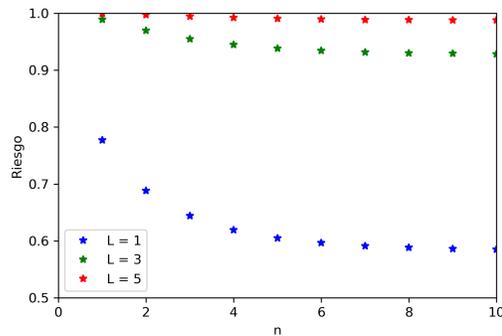


Figura 5. Probabilidad de que la epidemia se extinga después de la generación  $n$  o no se extinga.

## 2. Cotas y aproximaciones

La importancia de las cotas consiste en que permiten estimar características de interés, estimando solo algunos parámetros de la distribución de  $Y$ , sin el conocimiento completo de esta distribución.

Sea  $m < 1$ . Para el caso  $X_0 = 1$  se tienen las siguientes acotaciones para el número medio de generaciones hasta la extinción  $ET$  ([12] y [11]).

$$\frac{1}{p_0} \leq ET \leq \frac{1}{1-m-p_0} \ln\left(\frac{1-m}{p_0}\right), \quad m \neq 1-p_0.$$

Note que en las desigualdades anteriores basta con estimar  $m$  y  $p_0$ .

**Ejemplo 6** Sean  $X_0 = 1$ ,  $p_0 = 0,2$  y  $m = 0,9$ . Usando la desigualdad anterior, se tiene que  $5 \leq ET \leq 7$ . Esta desigualdad sugiere que, si el tiempo medio de generación es igual a 5, entonces el número medio aproximado de días hasta la extinción podría encontrarse en el intervalo  $[25, 35]$ .

Si  $m < 1$ , entonces ([5], pag. 116)

$$1 - X_0 \cdot m^n \leq Q_n \lesssim X_0(1-m)m^{n+1}/\sigma^2.$$

Con la parte izquierda de las desigualdades anteriores se puede estimar la generación  $n$  para la cual la epidemia, con una probabilidad  $\gamma$  dada, se extingue, resolviendo la inecuación

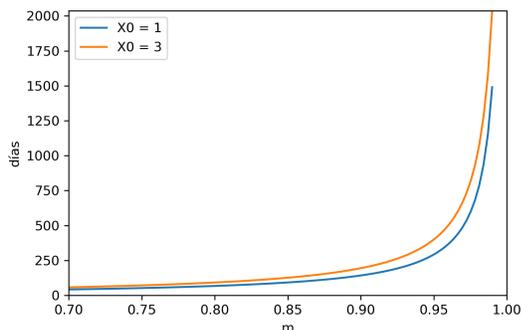
$$1 - X_0 \cdot m^n \geq \gamma,$$

es decir,

$$n \geq \frac{\ln(1-\gamma) - \ln(X_0)}{\ln(m)}. \tag{7}$$

Para  $m < 1$  la fórmula anterior expresa el riesgo de no realizar intervenciones, a través del número de generaciones  $n$  con el cual se extinguiría la epidemia con una probabilidad dada  $\gamma$ .

**Ejemplo 7** Sean  $X_0 = 1$ ,  $m = 0,7$  y  $\gamma = 0,95$ . De la fórmula (7) se halla que  $n \geq 8$ . Luego si el tiempo medio de generación es 5 días, entonces este resultado sugiere, que con una probabilidad 0,95, el tiempo medio mínimo para extinguirse la transmisión sin intervenciones podría ser no menor de 40 días. En la Figura 6 se observan gráficas para distintos valores de  $X_0$  y  $m \in [0,7; 1)$ .



**Figura 6.** Tiempo medio aproximado a partir del cual la epidemia se extingue con probabilidad  $\gamma = 0,95$ .

### 3. Proceso de Bellman-Harris

Consideremos variables aleatorias con igual distribución que una variable aleatoria  $V$ , que representa el tiempo de generación, el cual se define como el tiempo transcurrido desde que un individuo es infectado hasta que este infecta a un grupo de individuos. El tamaño de este grupo tiene distribución igual que una variable aleatoria  $Y$ . Supondremos que los tiempos de generación y los tamaños de los grupos generados son todas variables independientes.

Denotemos por  $X_t$  el número de infectados generados en el instante  $t$  (no hasta  $t$ ). Se llama proceso de Bellman-Harris al proceso  $\{X_t\}$ ,  $t \geq 0$ .

Sea  $M(t) = EX_t$ . Si  $V$  tiene distribución exponencial, entonces el proceso  $X_t$  es un proceso de Markov, llamado proceso de Galton-Watson en tiempo continuo. Sean  $EV = 1/\lambda$  y  $EY = m$ . Puede demostrarse (ver [5] pag. 61) que si  $M(0) = EX_0 = L$ , entonces

$$M(t) = Le^{\lambda(m-1)t}, \quad t \geq 0. \tag{8}$$

Nótese el crecimiento exponencial de  $M(t)$  cuando  $m > 1$ .

Cuando la variable  $V$  no tiene distribución exponencial, para el proceso  $X_t$  solo se obtienen resultados analíticos en casos particulares. En aplicaciones prácticas, en general, es necesario el uso de la simulación.

#### Proceso Epi

El proceso que describiremos a continuación se ajusta mejor al proceso de transmisión de infecciones en epidemias, y su comportamiento medio es muy similar al proceso de Bellman-Harris.

Consideremos ahora que los tiempos de generación son aquellos que transcurren desde que un individuo es infectado hasta que este infecta a otro. Supondremos que estos tiempos son independientes y tienen igual distribución que una variable aleatoria  $V$ , y también son independientes del número de contagiados por cada individuo infectado, cuya distribución es igual a la de una variable aleatoria  $Y$ .

Como antes, sea  $X_t$  el número de infectados generados en el instante  $t$ . Llamaremos proceso Epi (Epi-de epidemia), al proceso  $\{X_t\}$ ,  $t \geq 0$ .

Tanto en el proceso de Bellman-Harris como en el proceso Epi, es de gran interés el proceso  $\{Z_t\}$ ,  $t \geq 0$ , donde  $Z_t$  representa el número de infectados hasta el instante  $t$ .

### 4. $R_t$ . Estimación de $m$

En esta sección asumiremos  $m > 1$ . Recordemos que el valor estimado de  $m$  es  $R_0$ . El número efectivo de reproducción, denotado por  $R_t$ , es el número medio de los individuos que infecta cada individuo en el instante  $t$  de la transmisión.

Si el número de la generación  $n$  es grande, intuitivamente se tiene

$$m \approx \frac{1}{X_n} \sum_{i=1}^{X_n} Y_i = \frac{X_{n+1}}{X_n},$$

y un estimador de  $m$ , como se señala en [3] (Se propuso en [9]) es

$$\bar{m}_n = \frac{X_n}{X_{n-1}}, \quad n \geq 1,$$

si  $X_{n-1} > 0$ . Cuando  $X_n = 0$  se toma  $\bar{m}_n = 1$ . Esta idea se traslada al tiempo real. Si por ejemplo, la unidad de tiempo es día, se expresa

$$R_t \approx \frac{X_t}{X_{t-1}}, \quad t \geq 1.$$

con  $X_{t-1} > 0$ , donde  $X_t$  es el número de infectados correspondientes al instante  $t$ .

Intuitivamente se observa que en la práctica puede ser más útil utilizar una ventana de tiempo  $[t - k, t + k]$  para realizar la estimación,

$$R_t \approx \frac{1}{2k + 1} \sum_{i=t-k}^{t+k} \frac{X_i}{X_{i-1}}, \quad 1 \leq k \leq t. \tag{9}$$

El valor de  $k$  se determina intuitivamente de acuerdo a la situación práctica específica. Podrían tomarse, por ejemplo,  $k = 1$  (3 valores) o  $k = 2$  (5 valores).

Con una idea similar se pueden usar también las ventanas  $(t - k, t]$  o  $[t, t + k)$ . Si se trata de estimar el valor de  $R_0$  al inicio de la transmisión, entonces debe tomarse una ventana del tipo  $(t - k, t]$ ,  $k < t$ .

Según [3], el siguiente estimador de  $m$  es mínimo cuadrático, máximo verosímil y consistente, si la distribución de  $Y$  pertenece a una subclase amplia de la clase de distribuciones exponenciales

$$\hat{m}_n = \frac{\sum_{i=0}^n X_i}{\sum_{i=1}^n X_{i-1}} = \frac{Z_n}{Z_{n-1}}, \quad n \geq 1.$$

Trasladando la idea al caso del tiempo real, se tiene, para la ventana  $[t - k, t + k]$ ,

$$R_t \approx \frac{1}{2k+1} \sum_{i=t-k}^{t+k} \frac{\sum_{j=0}^i X_j}{\sum_{j=1}^i X_{j-1}}, \quad 1 \leq k \leq t. \quad (10)$$

En la literatura se presenta gran variedad de estimadores para  $R_0$ . En el presente estudio, utilizamos las expresiones (9) y (10).

### 5. Simulaciones

Las simulaciones para los procesos de Bellman-Harris y Epi muestran que estos dos procesos se comportan muy similares en términos de sus medias y desviaciones estándar. En la Figura 7 se observa que las medias obtenidas para ambos procesos por medio de la simulación son parecidas. La simulación se realizó usando para  $Y$  una distribución de Poisson con parámetro  $m = 1,45$  y para  $V$  una distribución Gamma con parámetros de forma y escala  $\alpha = 1,52$  y  $\mu = 3,47$ , respectivamente.

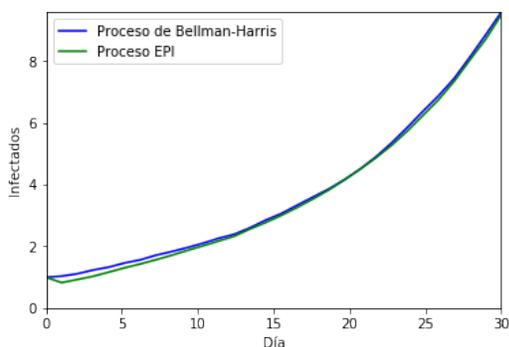


Figura 7. Cantidad media de individuos infectados en el instante  $t$ .

Adoptamos el proceso Epi como modelo de la transmisión de la enfermedad, y por lo tanto para la realización de las simulaciones, puesto que este proceso se ajusta mejor al comportamiento real de la epidemia. A continuación se analiza gráficamente cómo influyen los parámetros  $m$  y  $X_0$  en el desarrollo de la transmisión. En todos los casos se generaron 1000 trayectorias del proceso, donde el tiempo de generación  $V$  tiene distribución Gamma con parámetros de forma y escala  $\alpha = 1,52$  y  $\mu = 3,47$ , respectivamente. Con estos valores de los parámetros, el tiempo medio de generación es  $\bar{v} = 5,28$  días.

El valor de  $m$  tiene una gran influencia en el comportamiento de la epidemia. Asumiendo que en el instante inicial hay 3 individuos infectados, es decir,  $X_0 = 3$ , en la Figura 8 se observa la cantidad media de individuos infectados hasta el instante  $t$ , para diferentes valores de  $m$ .

Otro valor que influye en el comportamiento de la epidemia es la cantidad inicial de individuos infectados. En la Figura 9 se ilustra el comportamiento de la epidemia, a través

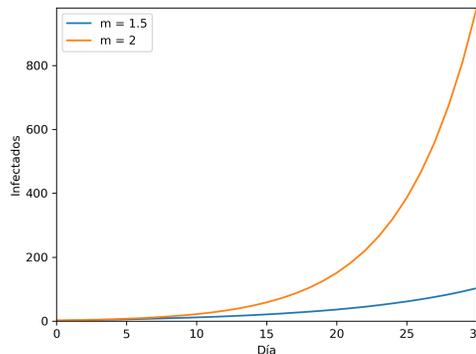


Figura 8. Cantidad media de individuos infectados hasta el instante  $t$  para  $\bar{v} = 5,28$  y  $X_0 = 3$ .

del número medio de infectados, para diferentes valores de  $X_0$  y de  $m$ .

En la Figura 10 se muestran las gráficas de las medias del número de infectados hasta el instante  $t = 30$  y una región de confianza al 90% alrededor de estas medias. Los valores de  $m$  que se tomaron son pequeños, puesto que para valores mayores no es posible determinar una región de confianza aceptable, debido a que en estos casos la varianza del proceso tiene valores muy grandes.

### 6. Validación del modelo y de las simulaciones

Como una forma de verificar la correcta implementación del programa de simulación, se comprobó la coincidencia de los valores medios del número de individuos que se infectan en un instante  $t$ , hallados por medio de la simulación, con la fórmula (8). Esta verificación se realizó asumiendo que los tiempos de generación tienen distribución exponencial, tal como es requerido para deducir (8).

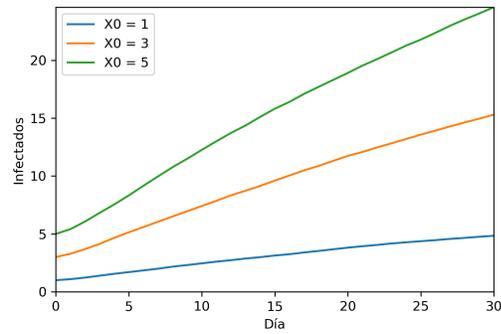
#### 6.1 Tiempos de generación

Debido a la dificultad para medir el tiempo de generación, internacionalmente es aceptado que, como valor aproximado de este, se tome la diferencia entre los instantes de inicio de síntomas del individuo que infecta y el infectado, es decir, el intervalo de serie.

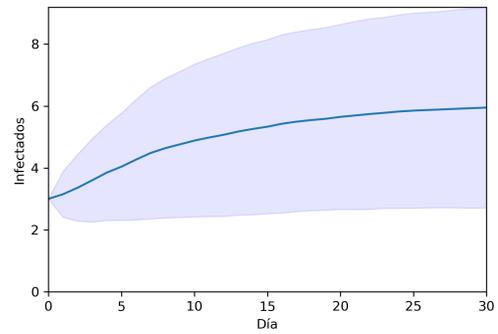
En la Figura 11 se muestra un histograma correspondiente a una muestra de 125 valores de tiempos de generación en Cuba. La media y desviación estándar fueron 5,28 y 4,28 días, respectivamente.

Al realizar la prueba de la bondad de ajuste  $\chi^2$  con la hipótesis de una distribución Gamma, se obtuvo un  $p$ -valor igual a 0,40 aproximadamente.

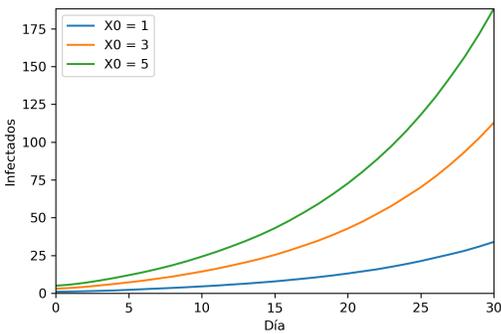
Las estimaciones por el método de los momentos de los parámetros de forma y de escala de la distribución Gamma son  $\mu = 3,47$  y  $\alpha = 1,52$ , respectivamente.



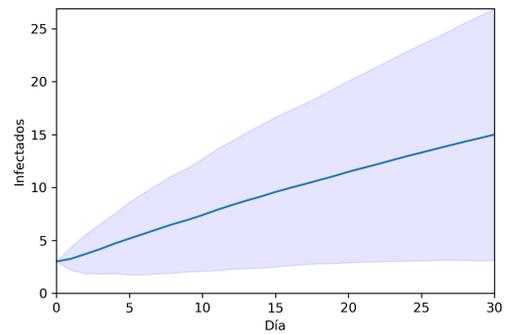
(a)  $m = 0,9$



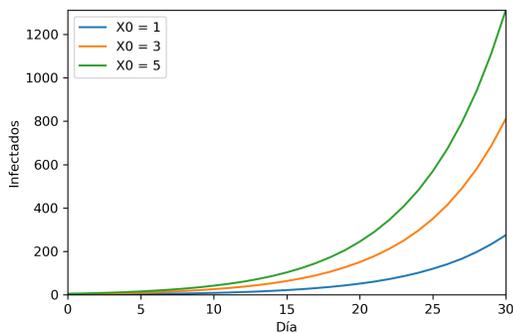
(a)  $m = 0,5$



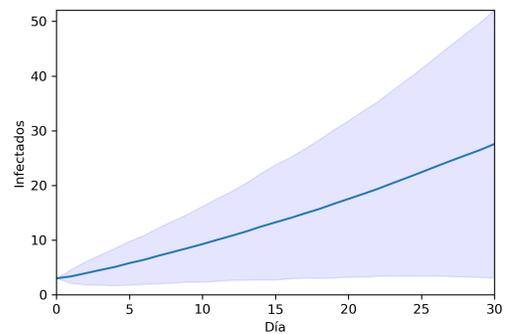
(b)  $m = 1,5$



(b)  $m = 0,9$



(c)  $m = 2$



(c)  $m = 1,1$

**Figura 9.** Cantidad media de individuos infectados hasta el instante  $t$  para  $\bar{\nu} = 5,28$ .

**Figura 10.** Cantidad media de individuos infectados hasta el instante  $t$  para  $\bar{\nu} = 5,28$ , y región de confianza para el 90% ( $\pm 1.64 \cdot \text{stdev}$ ).

### 6.2 La Habana

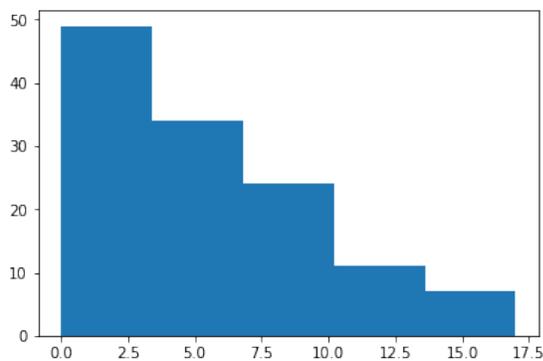
La Habana es la provincia con el mayor número de casos positivos a la enfermedad Covid-19, por lo que es de interés validar el modelo en la capital.

Para la estimación del valor de  $m$  durante las dos primeras semanas desde el inicio de la transmisión, utilizamos los estimadores (9) y (10) con ventana de tiempo  $k = 3$ , y se obtuvo el valor aproximado  $m = 1,42$ .

De una muestra de los tiempos de generación con tamaño 48, se obtuvo que la media y la desviación estándar son 5,67 y

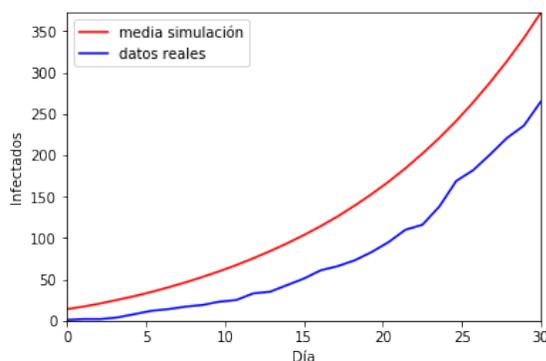
5,28 días, respectivamente. De estos valores, por el método de los momentos, se hallaron las estimaciones de los parámetros,  $\alpha = 1,15$  y  $\mu = 4,92$ . Al realizar la prueba de bondad de ajuste  $\chi^2$ , no se rechaza el ajuste a la distribución Gamma con p-valor igual a 0,20.

En la Figura 12 se observan las curvas correspondientes a la cantidad acumulada de individuos infectados en el instante  $t$  en La Habana, y a la cantidad media de infectados en ese instante obtenida por medio de la simulación del proceso Epi.



**Figura 11.** Histograma de los tiempos de generación para 125 datos en Cuba.

El instante inicial fue el 16 de marzo, fecha de confirmación del primer paciente en la provincia. Se generaron 1000 trayectorias del proceso utilizando los valores mencionados de las estimaciones de los parámetros.



**Figura 12.** Cantidad de individuos infectados hasta el instante  $t$  en La Habana.

Para la simulación se asumió que en el instante inicial habían 14 individuos infectados, que es la cantidad de individuos confirmados con la enfermedad que arribaron a la provincia en los primeros días de inicio de la transmisión.

La curva media obtenida mediante la simulación corresponde a una situación en la cual no se toman medidas de contención de la transmisión, y puesto que se halló con una cantidad limitada de infectados iniciales, infraestima la situación real en que no se toman estas medidas. No obstante, como se observa en la Figura 12, esta curva tiene la tendencia a ser superior a la curva correspondiente al número total de infectados reales, situación en la cual sí existen medidas de contención, aunque por otra parte, debe señalarse que en la curva real posiblemente no se incluyen algunos casos asintomáticos.

## Conclusiones

Los Ejemplos 1-7 de las Secciones 1 y 2, ilustran cómo algunas fórmulas de la teoría de los procesos de ramificación podrían tener utilidad en las estimaciones de diferentes índices sobre el comportamiento de la epidemia.

En la Sección 3 se introdujo el proceso que denominamos Epi, el cual consideramos que es el que mejor modela la epidemia, entre los dos procesos presentados en esa sección.

Las distribuciones, Gamma, Weibull y Lognormal, son utilizadas en los estudios internacionales como modelos de distribución para los tiempos de generación. Tanto para los datos de Cuba, como de La Habana, fue la distribución Gamma la que mejor se ajustó. Por otra parte, asumimos la distribución de Poisson como ley de reproducción, sobre la base de consideraciones teóricas y la experiencia internacional. No obstante, es conveniente analizar el uso de otras distribuciones.

Para la realización de las simulaciones se implementaron programas en el lenguaje Python. La validación del modelo Epi y de su implementación en el lenguaje de programación, realizada con los datos de la provincia La Habana, indica que el programa puede ser utilizado en apoyo a la toma de decisiones sobre las medidas a tomar, ante un posible reinicio de la transmisión de la enfermedad en las diferentes provincias.

## Agradecimientos

Agradecemos a Josué M. Corujo por la colaboración brindada en la primera etapa de la realización de este trabajo.

## Referencias

- [1] Abbott, S.: *The transmissibility of novel Coronavirus in the early stages of the 2019-20 outbreak in Wuhan: Exploring initial point-source exposure sizes and durations using scenario analysis [version 1]*. Wellcome Open Research 2020, 5:17 Last updated: 11 FEB 2020.
- [2] Athreya, K. B. y P. Ney: *Branching Processes*. Springer-Verlag, 1972.
- [3] Becker, N.: *Estimation for Discrete Time Branching Processes with Application to Epidemics*. *Biometrics*, 33:515–522, September 1977.
- [4] Brauer, F. & Castillo-Chávez, C.: *Mathematical Models in Population Biology and Epidemiology*. Springer, 2012.
- [5] Haccou, P., P. Jagers y V. A. Vatutin: *Branching Processes*. Cambridge, 2007.
- [6] Harris, T. H.: *The theory of Branching Processes*. Springer-Verlag, Berlin, 1963.
- [7] Imai, N., Cori, I. A., Dorigatti, M. Baguelin, C. A. Donnelly, S. Riley y N. M. Ferguson: *Report 3: Transmissibility of 2019-nCoV*. Imperial College London COVID-19 Response Team, 25 January 2020.

- [8] Jacob, C.: *Branching Processes: Their Role in Epidemiology*. Int. J. Environ. Res. Public Health, 7:1186–1204, 2010. doi:10.3390/ijerph7031204.
- [9] Nagaev, A. V.: *On estimating the expected number of direct descendants of a particle in a branching process*. Theory of Probability and its Applications, 12:314–320, 1967.
- [10] P., Boldog, Tekeli T., Vizi Z., Dénes A., Bartha F. A. y Rost G.: *Risk assessment of novel coronavirus 2019-nCoV outbreaks outside China*. medRxiv preprint, 2020. <https://doi.org/10.1101/2020.02.04.20020503>.
- [11] Valdés, J. E., Yera Y. G. y L. Zuaznabar: *Bounds for the Expected Time to Extinction and the Probability of Extinction in the Galton-Watson Process*. Communications in Statistics-Theory and Methods, 43:8:1698–1707, 2014.
- [12] Yu, J. y J. Pei: *Extinction of Branching Processes in Varying Environments*. Statistics and Probability Letters, 79:1872–1877, 2009.