

Doi.

Artículo de investigación

Un modelo de mezclas Bayesiano para describir rebrotes del COVID-19 en el estado de Guerrero, México

A Bayesian mixture model to describe COVID-19 outbreaks in the state of Guerrero, Mexico

Um modelo de mistura Bayesiana para descrever surtos de COVID-19 no estado de Guerrero, México.

Yaineris Ferrán-Yera ID. 0000-0002-7851-8198

Francisco J. Ariza-Hernández* ID. 0000-0002-1414-8490

Martín P. Árciga-Alejandro ID. 0000-0002-1637-8513

Jorge Sánchez-Ortiz ID. 0000-0002-9228-7285

Facultad de Matemáticas, Universidad Autónoma de Guerrero. Av. Lázaro Cárdenas S/N. Ciudad Universitaria Sur, 39086, Chilpancingo de los Bravo, Guerrero, México.

*Autor de correspondencia arizahfj@uagro.mx

Enviado: 03/02/2024

Revisado: 06/03/2024

Aprobado: 08/05/2024

Publicado: 15/06/2024

Resumen

El COVID-19 ha puesto en alerta al mundo entero debido al gran número de contagios, hospitalizaciones y defunciones provocadas. Desde su inicio, en diciembre de 2019, han estado ocurriendo “olas” de incrementos considerables del número de contagios y muertes, conocidos como rebrotes. En este trabajo, se modela matemáticamente el comportamiento diario de infecciones, hospitalizaciones y defunciones por COVID-19 en el estado de Guerrero del 15 de marzo de 2020 al 27 de febrero de 2022, haciendo uso de un modelo de mezclas Bayesiano con distribuciones gaussianas. Se estimaron los parámetros de interés involucrados en el modelo mediante el método Bayesiano, se obtuvieron las distribuciones posteriores de cada uno de ellos usando Cadenas de Markov vía Monte Carlo y se validaron las estimaciones con los análisis de convergencia correspondientes. Con el modelo de mezclas se identificaron cinco rebrotes para los casos infectados, y cuatro para los hospitalizados y defunciones. Finalmente, la implementación del modelo se realizó con la ayuda del software estadístico R y el paquete JAGS.

Palabras clave: Mezcla de distribuciones, Cadenas de Markov, Monte Carlo, Estimación Bayesiana, Rebotes de COVID-19

Abstract

The COVID-19 pandemic has put the entire world on alert due to the large number of infections, hospitalizations, and deaths caused by this disease. Since its beginning, in December 2019, outbreaks of considerable increases in the number of infections and deaths, have been occurring. In this work, the daily behavior of infections, hospitalizations, and deaths from COVID-19 in the state of Guerrero is mathematically modeled using a Bayesian mixture model with Gaussian distributions from March 15, 2020 to February 27, 2022. The parameters of interest were estimated using the Bayesian method, the posterior distributions of each of them were obtained using Markov Chains via Monte Carlo and the estimates were validated with the corresponding convergence analyses. With the mixture model, five outbreaks were identified for infected cases, and four for both hospitalized and deaths cases. Finally, the implementation of the model was carried out with the help of the R statistical software and the JAGS package.

Keywords: Mixture of distributions, Markov Chain, Monte Carlo, Bayesian estimation, COVID-19 outbreaks.

Resumo

A COVID-19 colocou o mundo inteiro em alerta devido ao grande número de infecções, hospitalizações e mortes que causou. Desde o seu início, em dezembro de 2019, houve ondas sucessivas de aumentos consideráveis no número de infecções e mortes, conhecidas como surtos. Este artigo modela matematicamente o comportamento diário das infecções, hospitalizações e mortes por COVID-19 no estado de Guerrero, de 15 de março de 2020 a 27 de fevereiro de 2022, utilizando um modelo de mistura Bayesiano com distribuições Gaussianas. Os parâmetros de interesse envolvidos no modelo foram estimados utilizando o método Bayesiano, as distribuições posteriores de cada um foram obtidas utilizando cadeias de Markov via Monte Carlo, e as estimativas foram validadas com as respectivas análises de convergência. O modelo de mistura identificou cinco surtos de infecções e quatro de hospitalizações e mortes. Finalmente, o modelo foi implementado utilizando o software estatístico R e o pacote JAGS.

Palavras-chave: Mistura de distribuições, Cadeias de Markov, Monte Carlo, Estimação Bayesiana, surtos de COVID-19

Introducción

La pandemia por COVID-19 ha puesto en alerta al mundo entero desde su aparición en diciembre del 2019 en la ciudad de Wuhan, China, se estima que alrededor de 235,667,597 personas han sido infectadas y 3,924,164 han muerto en todo el continente americano por esta enfermedad; tan solo en México, se tiene el dato de 7,450,992 infectados y 332,986 muertes reportados oficialmente hasta febrero de 2023 (PAHO, 2020) y en el estado de Guerrero se reportan 118,827 infectados y 6,878 defunciones. A pesar de todas las investigaciones médicas y estrategias trazadas para contener el contagio, en determinados momentos han ocurrido “olas” de incremento considerables del número de casos y defunciones, a estas etapas de transición de la pandemia se le conoce como rebrotes. Dichos rebrotes se pueden describir matemáticamente como curvas con características o especificaciones diferentes tanto en intensidad como duración en el tiempo. Las contramedidas de salud pública para interrumpir y controlar la transmisión se basaron en modelos de predicción y en la comprensión del comportamiento de la enfermedad.

Esta enfermedad sugiere desafíos en el uso de modelos matemáticos para describir y predecir la dinámica de la pandemia. Se han propuesto una gran cantidad de modelos matemáticos basados en ecuaciones diferenciales ordinarias y/o parciales diseñados específicamente para

examinar la propagación de COVID-19, los cuales han abordado puntos importantes relacionados con la estimación de los parámetros (Calvetti *et al.*, 2020); por ejemplo, Tejeira-Huacani (2020) realiza un análisis de la propagación de la pandemia usando una modificación del modelo Susceptible-Infectado-Recuperado (SIR), donde se estudia la estabilidad lineal, se muestra que los estados asociados a los puntos fijos llegan a ser inestables y la dinámica del sistema considera medidas de contención. Concluye que estas son útiles para disminuir el número de personas infectadas.

Beira y Sebastião (2021) realizan un análisis de ajuste de un modelo compartimental para el caso portugués, utilizando una plataforma en línea de acceso abierto con la capacidad integrada de resolver sistemas de ecuaciones diferenciales. La cual permitió la validación basada en datos del modelo utilizado y fue la base para proyecciones robustas de diferentes escenarios futuros. Koufi y Koufi (2021), presentan un modelo estocástico de la epidemia de COVID-19 para una población con cinco compartimentos, dan un análisis numérico del modelo estocástico propuesto y realizan una comparación con los resultados del modelo determinista correspondiente. Baleanu, Mohammadi y Rezapour (Baleanu *et al.*, 2020), presentan un modelo de orden fraccionario con derivada Caputo-Fabrizio para modelar la transmisión de COVID-19, quienes implementaron el método homotópico de transformación, que

combina el método de análisis de homotopía y la transformada de Laplace, resolvieron el problema y dan una solución aproximada en series convergentes. Además, exponen resultados numéricos para simular la transmisión del virus y comparar los resultados usando la derivada de Caputo. Manevski *et al.* (2020) proponen un modelo semi paramétrico para describir y pronosticar la pandemia de COVID-19, usando inferencia bayesiana estimaron el número de reproducción efectivo, la tasa de fatalidad y la proporción de la población infectada en el país de Eslovenia; sin embargo, estos autores no consideran el modelamiento conjunto de los rebrotes a través del tiempo.

Por lo tanto, no es una tarea sencilla describir mediante un modelo genérico, el comportamiento de los incrementos de contagio de esta pandemia desde sus inicios. Las mezclas de distribuciones estadísticas representan una herramienta adecuada para analizar el comportamiento de N subpandemias con distintos comportamientos como un todo. Un modelo de mezclas de funciones de distribución permite determinar la presencia de subpoblaciones dentro de una población general, esto sin requerir observaciones individuales de cada una de ellas (Ghojogh *et al.*, 2019).

A la fecha no se han encontrado evidencias literarias de estudios realizados en el estado de Guerrero que utilicen un modelo matemático donde se incluyan varias etapas de transición de esta pandemia. Por esta razón, este trabajo tiene como objetivo, modelar matemáticamente el comportamiento de los rebrotes de infecciones diarias, de hospitalizados y de muerte por COVID-19, a través de un modelo de mezcla de distribuciones gaussianas, desde sus inicios hasta febrero de 2022 para el estado de Guerrero, estos modelos son analizados desde el punto de vista de la Estadística Bayesiana, la cual nos permite actualizar el conocimiento previo sobre las cantidades de interés, utilizando los datos observados y calcular las distribuciones posteriores marginales de dichas variables, que se obtienen numéricamente usando la metodología de Cadena de Markov vía Monte Carlo (Robert y Casella 2005).

Materiales y Métodos

Modelo de Mezclas

En la mayoría de los estudios estadísticos se supone que se tiene una muestra de observaciones que provienen de la misma distribución, por esta razón el interés con los datos muestrales consiste en buscar una sola distribución que describa el comportamiento de los datos. Sin embargo, es posible que se tengan datos que provienen de más de una distribución, por lo que se deben considerar otros modelos que describan mejor ese tipo de datos.

A toda variable aleatoria se le considera que proviene de una distribución de probabilidad, ya sea una distribución conocida o no. Algunas variables aleatorias se extraen de una única distribución, como la Normal. Sin embargo, existen otras variables aleatorias que pueden haberse generado a partir de una mezcla de varias distribuciones y no de una sola. En el contexto estadístico, se utiliza un modelo de mezclas de distribuciones como un modelo probabilístico para representar la presencia de subpoblaciones dentro de una misma población. Los modelos de mezclas finitas proporcionan una metodología que puede identificar la heterogeneidad no observada en los datos, que se debe a la influencia de diferentes variables o grupos que pueden causar sobre dispersión o heteroscedasticidad en un modelo tradicional. En general, un modelo de mezcla finita proporciona una alternativa paramétrica que describe la distribución desconocida en términos de mezclas de distribuciones conocidas, que hace a los modelos ser más flexibles.

La distribución de la mezcla es una suma ponderada de K distribuciones de probabilidad $g_1(x; \theta_1), g_2(x; \theta_2), \dots, g_K(x; \theta_K)$, con pesos $\pi_1, \dots, \pi_K$ que suman uno, y cada distribución de la mezcla tiene su propio parámetro θ_i . La distribución de la mezcla se presenta como:

$$p(\theta) = \sum_{i=1}^K \pi_i g_i(x; \theta_i),$$

con $\sum_{i=1}^K \pi_i = 1$, $\pi_i > 0$, donde:

X – es la muestra poblacional de tamaño K ,
 $g_i(x|\theta_i)$ - densidad de la distribución i ,
 θ_i - vector de parámetros de la distribución i ,
 π_i - peso de la mezcla de la distribución i .

El conjunto de parámetros que definen a estos modelos puede ser estimado mediante diferentes técnicas como métodos gráficos, método de Esperanza-Maximización (EM), método de momentos, método de máxima verosimilitud, aproximaciones Bayesianas, entre otras.

Modelo de ajuste de curvas

Sea X una variable aleatoria con función de densidad de probabilidad (fdp) Normal con media μ y varianza σ^2 , la cual denotamos por $X \sim N(\mu, \sigma^2)$, dicha fdp está dada por:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad -\infty < \mu < \infty, \sigma^2 > 0,$$

donde $x \in \mathbb{R}$. La función de distribución acumulada (fda) de X , está dada por:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(u) du$$

La fdp es también conocida como distribución gaussiana, cuya representación gráfica de la densidad y de la distribución acumulada se muestra en la Figura 1. La distribución normal es muy utilizada en modelos de regresión lineal y de análisis de varianza (ANOVA), los cuales son aplicados en distintas áreas de la ciencia, como la Biología, Medicina, Química, Agronomía, etc. Algunas de sus propiedades más importantes es que el valor esperado de X , es $E(X) = \mu$, el cual representa el valor mas probable de la variable X , y la varianza es $Var(X) = \sigma^2$, el cual representa la anchura de la curva gaussiana.

Los modelos de ajuste de curvas son modelos matemáticos útiles para predecir la trayectoria de pandemias a lo largo del tiempo t . Una curva comúnmente utilizada para tales modelos es una curva en forma de campana definida por la función gaussiana:

$$f(t) = P \exp\left(-\frac{(t - t_m)^2}{2\sigma^2}\right) \quad (1)$$

donde t_m es el tiempo donde la curva alcanza el pico máximo, P es la magnitud del pico, y σ

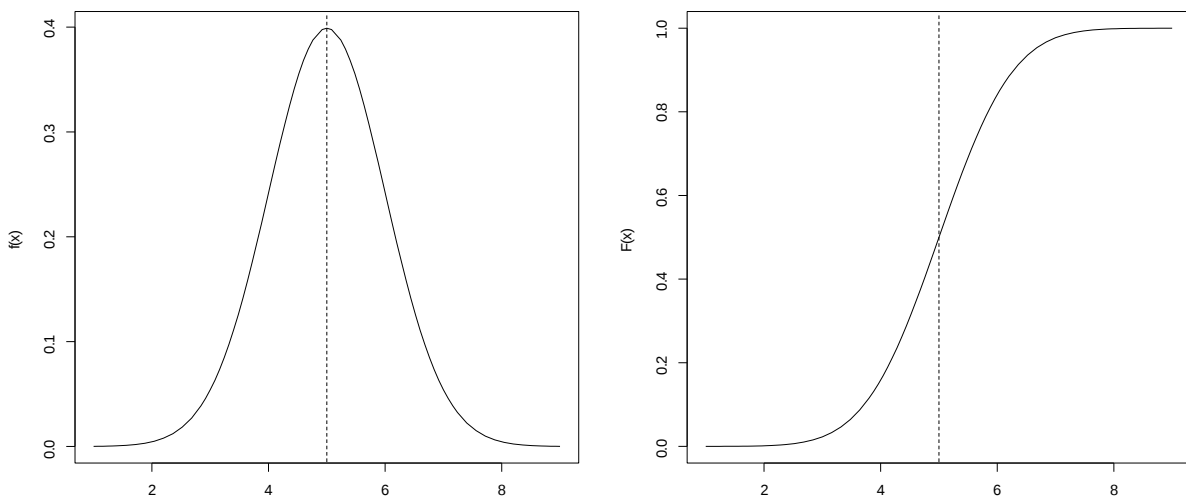


Figura 1. Función de densidad de probabilidad (izquierda) y Función de distribución acumulada (derecha) de $N(5,1)$.

refleja la anchura de la curva. Alternativamente, $f(t)$ en (1) puede expresarse como:

$$f(t) = (P\sqrt{2\pi\sigma}) \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(t - t_m)^2}{2\sigma^2}\right) \\ = M\phi(t, t_m, \sigma)$$

donde $M = P\sqrt{2\pi\sigma}$ representa el número total de infectados y $\phi(t, t_m, \sigma)$ es la función de densidad de probabilidad Normal con media t_m y desviación estándar σ , la cual rige cómo se distribuye M a lo largo del tiempo; es decir, el número de casos esperado en el momento t está dado por $M\phi(t, t_m, \sigma)$ (Kaciroti *et al.*, 2021).

Metodología

Para que un modelo de ajuste de curvas funcione bien, la forma de la curva elegida es importante, debe tener un buen ajuste con los datos observados y una justificación teórica subyacente adecuada. La curva definida por una función Gaussiana tiene una forma simétrica alrededor del pico, y las tendencias futuras se pronostican extrapolando la tendencia observada, forzando el componente posterior al pico de la trayectoria para que sea simétrico con la parte anterior al pico (Kaciroti *et al.*, 2021). Estas previsiones suponen que los factores relacionados con la pandemia no cambian con el tiempo, por lo que son adecuadas cuando la propagación de la pandemia es relativamente homogénea y no hay medidas atenuantes de cambios de comportamiento en cada curva o en cada rebrote.

Aunque la suposición de homogeneidad a lo largo del tiempo puede ser razonable en periodos de tiempo cortos, no es realista para periodos de tiempo largos y áreas grandes, ya que los cambios en la progresión de la pandemia darán lugar a cambios de comportamiento a nivel individual, y a cambios en las políticas y prácticas a nivel local, estatal y federal (por ejemplo, distanciamiento social y órdenes de permanecer en casa). Estos cambios pueden alterar, a veces de forma sustancial, el curso natural de la enfermedad.

Para tener en cuenta estos cambios, proponemos un modelo en el que la curva de la trayectoria de los infectados diarios (hospitalizados o defunciones) $\mu(t)$ en el día t se exprese como

una mezcla de múltiples subcurvas homogéneas. Se puede considerar que cada subcurva capta implícitamente una subepidemia (rebrote) o una tendencia específica en la trayectoria durante un periodo de tiempo determinado, que se relaciona con factores subyacentes (conocidos o desconocidos). Esto es particularmente relevante para la pandemia de COVID-19 en el estado de Guerrero, en donde los cambios en el cumplimiento del distanciamiento social dieron lugar a varios aumentos de casos en el periodo aquí analizado. Específicamente, se tiene que los infectados diarios (hospitalizados o defunciones) en el día t está comprendido por:

$$y(t) = \sum_{i=1}^K y_i(t)$$

donde $y_i(t)$ representa el número de infectados (hospitalizados o defunciones) en el día t atribuidas al rebrote $i = 1, \dots, K$, $K \geq 1$ es el número de rebrotes identificados empíricamente. La trayectoria $\mu_i(t) = E(y_i(t))$, para el rebrote i puede modelarse mediante una función Gaussiana homogénea $\mu_i(t) = C_i\phi(t; t_{m,i}, \sigma_i)$, donde C_i es el número de casos infectados (hospitalizados o defunciones) globales, $t_{m,i}$ es el tiempo donde se alcanza el pico y σ_i representa la amplitud del rebrote i . La curva de infectados (hospitalizados o defunciones) para toda la pandemia es $E(y(t)) = \mu(t)$, la cual se puede descomponer como:

$$\mu(t) = \sum_{i=1}^K \mu_i(t) = \sum_{i=1}^K C_i \phi(t; t_{m,i}, \sigma_i)$$

Sea $C = \sum_{i=1}^K C_i$ los infectados (hospitalizados o defunciones) globales de la pandemia, entonces $\pi_i = C_i/C$ es la proporción de todos los infectados (hospitalizados o defunciones) atribuidas a la curva i , con $C_i = C\pi_i$. Por lo tanto, proponemos el siguiente modelo de mezcla basado en curvas gaussianas para modelar la trayectoria de los casos diarios (hospitalizados diarios y defunciones diarias), de COVID-19 para el estado de Guerrero, definido por:

$$\mu(t) = C \sum_{i=1}^K \pi_i \phi(t; t_{m,i}, \sigma_i)$$

donde C son los casos de infectados (hospitalizados o defunciones) globales de toda la pandemia y $\phi(t; t_{m,i}, \sigma_i)$ representa la parte de la curva relacionada con algún(os) factor(es) subyacente(s), π_i es la proporción de infectados (hospitalizados o defunciones) totales que se atribuyen a dicho(s) factor(es) indexado(s) por i , y K es el número de subcurvas que componen la mezcla. El número de parámetros que identifican la curva es:

$$3K: C, \pi_1, \dots, \pi_{K-1}, (t_{m,1}, \sigma_1), \dots, (t_{m,K}, \sigma_K)$$

La $\phi(t; t_{m,i}, \sigma_i)$ es la función de densidad de probabilidad Gaussiana, definida por el parámetro de localización $t_{m,i}$, que representa el tiempo desde el primer infectado hasta la cúspide, y el parámetro de escala σ_i , que representa la dispersión de la curva.

En este trabajo, se usa el enfoque Bayesiano (Gelman et al., 2004) para estimar los parámetros de interés, entre ellos, los parámetros de la curva y sus intervalos de credibilidad. Suponemos que la distribución de los datos de infectados (hospitalizados o defunciones) observados $y(t)$ en el momento t es una Binomial Negativa (BN); es decir,

$$Y(t) \sim BN(N, p_t)$$

donde N es el tamaño de la población del estado. Se utiliza una función de distribución Binomial Negativa para $Y(t)$, ya que es apropiada para datos de conteo, con $E(Y(t)) = \mu(t) N p_t / (1 - p_t)$. Los métodos de Cadenas de Markov vía Monte Carlo Bayesianos (Green 1995; Koufi y Koufi, 2021) se utilizan para realizar extracciones de la distribución posterior marginal de los parámetros desconocidos.

Distribuciones a priori

Se utilizaron distribuciones previas débilmente informativas para todos los parámetros del

modelo. Las distribuciones a priori para nuestro estudio en general tienen la siguiente forma:

$$\begin{aligned} t_{m_1} &\sim N(\mu_1, \sigma^2) \\ t_{m_2} &\sim N_T(t_{m_1}, \mu_2, \sigma^2) \\ t_{m_3} &\sim N_T(t_{m_2}, \mu_3, \sigma^2) \\ &\vdots \\ t_{m_K} &\sim N_T(t_{m_{K-1}}, \mu_K, \sigma^2) \end{aligned}$$

Note que se tiene la restricción $t_{m_1} < t_{m_2} < t_{m_3} < \dots < t_{m_K}$, donde $N_T(t_l, \mu_K, \sigma^2)$ es una función de densidad de probabilidad Normal Truncada por la izquierda definida por:

$$f_T(t) = \frac{1}{\sigma} \frac{\phi\left(\frac{t-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{t_l - \mu}{\sigma}\right)}, \quad t_l < t < \infty$$

donde $\phi(\cdot)$ y $\Phi(\cdot)$ son las funciones de densidad y de distribución acumulada de probabilidad de la normal estándar. Para los parámetros σ_i y π las distribuciones a priori se muestran de la siguiente forma:

$$\begin{aligned} \sigma_i &\sim \Gamma(\alpha, \lambda), \quad i = 1, \dots, K, \\ \pi &\sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K) \end{aligned}$$

Distribución a posteriori

El cálculo de las densidades a posteriori suele verse facilitado por la siguiente expresión:

$$\begin{aligned} p(X) &\propto p(\theta) p(\theta) \\ &\propto \prod_{i=1}^n [G p_X(\theta_1, \dots, \theta_n)] \times p(\theta_1) p(\theta_2) \dots p(\theta_n) \end{aligned}$$

La información proporcionada por los datos $p(\theta)$ es llamada función de verosimilitud, $p(X)$ es la distribución a posteriori y $p(\theta)$ es la distribución a priori.

Resultados

En la Figura 2, se muestra el comportamiento a través del tiempo de los casos infectados diarios por COVID-19 en el estado de Guerrero a partir del 15 de marzo de 2020 al 27 de febrero de 2022,

en la figura se puede apreciar cinco rebrotes, el primero a principios de julio de 2020, el segundo a finales de septiembre del mismo año, el tercero a mediados del mes de enero de 2021, el cuarto a principios de agosto de 2021 y finalmente el quinto a mediados de enero de 2022.

En la Figura 3, se muestran los casos de hospitalizaciones diarias a partir del 08 de marzo de 2020 al 30 de diciembre de 2021, en la figura se puede apreciar cuatro rebrotes, en los meses de junio y octubre de 2020, y en febrero y agosto de 2021.

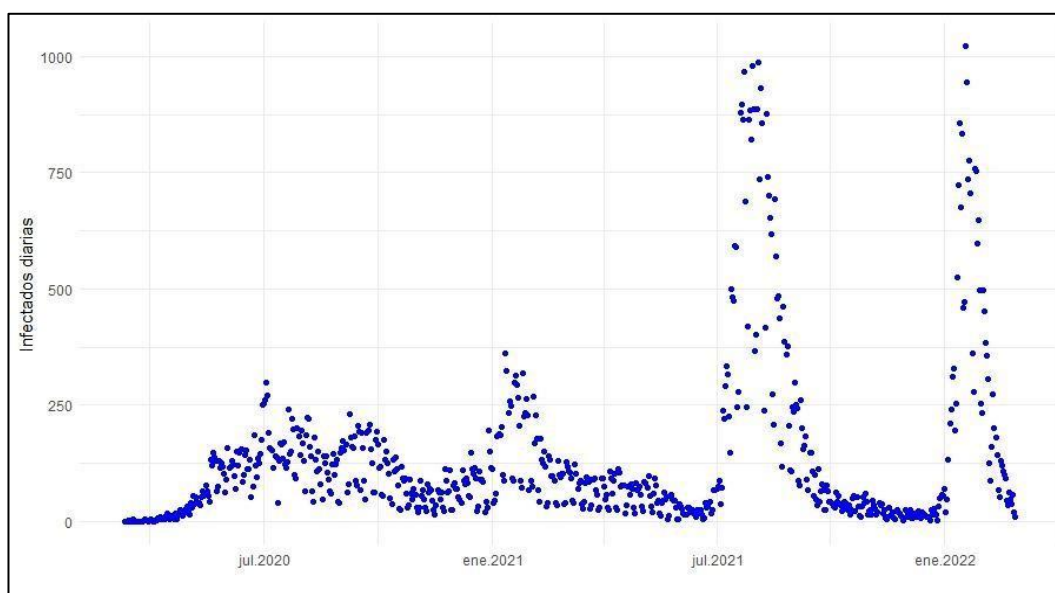


Figura 2. Casos de infectados diarios por COVID-19 en el estado de Guerrero.

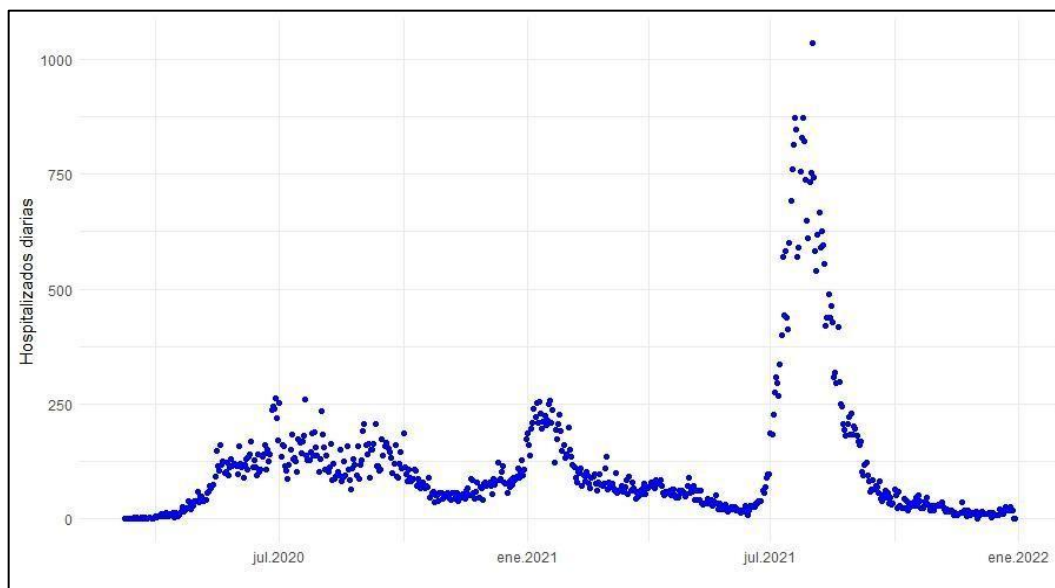


Figura 3. Casos de hospitalizados diarios por COVID-19 en el estado de Guerrero.

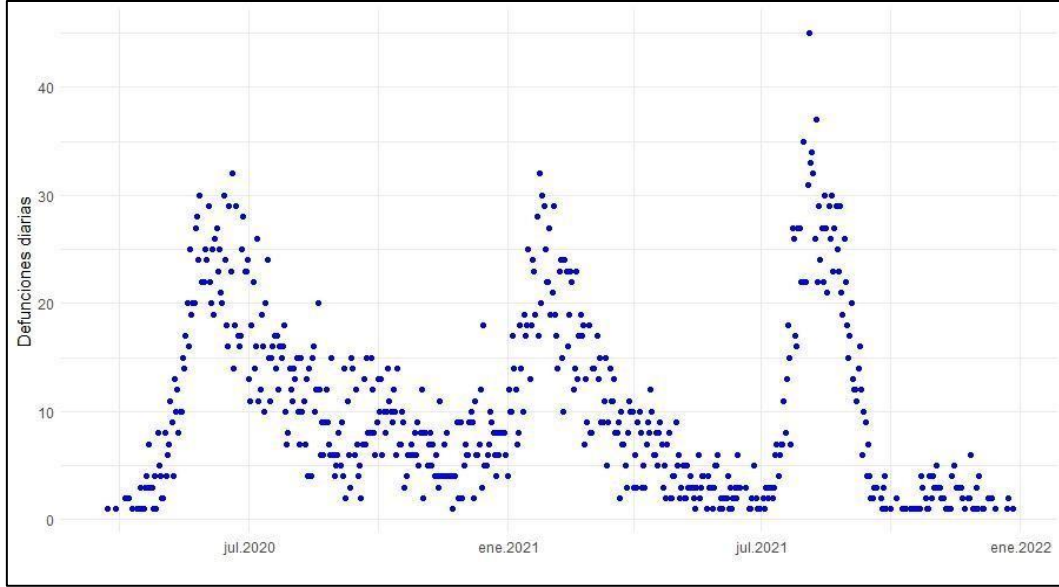


Figura 4. Casos de defunciones diarias por COVID-19 en el estado de Guerrero.

Respecto al caso de defunciones diarias, como se muestra en la Figura 4, se observan defunciones diarias a partir del 22 de marzo de 2020 al 27 de diciembre de 2021, en la figura se puede apreciar cuatro rebrotes en los meses junio y octubre de 2020 y en febrero y agosto de 2021.

En cada uno de los casos anteriores, se identificaron rebrotes de forma empírica y se propusieron los siguientes modelos:

Infectados:

$$Y(t) \sim BN(M, p_t)$$

$$p_t = \frac{M}{\mu(t) + M}$$

$$\mu(t) = M \sum_{i=1}^5 \pi_i \phi(t; t_{m,i}, \sigma_i)$$

$$= M [\pi_1 \phi(t; t_{m,1}, \sigma_1)$$

$$+ \pi_2 \phi(t; t_{m,2}, \sigma_2)$$

$$+ \pi_3 \phi(t; t_{m,3}, \sigma_3)$$

$$+ \pi_4 \phi(t; t_{m,4}, \sigma_4)$$

$$+ \pi_5 \phi(t; t_{m,5}, \sigma_5)]$$

Se propusieron las siguientes distribuciones a priori para los parámetros:

$$t_{m_1} \sim N(130, \frac{1}{100})$$

$$t_{m_2} \sim N_T(t_{m_1}, 210, \frac{1}{100})$$

$$t_{m_3} \sim N_T(t_{m_2}, 325, \frac{1}{100})$$

$$t_{m_4} \sim N_T(t_{m_3}, 530, \frac{1}{100})$$

$$t_{m_5} \sim N_T(t_{m_4}, 690, \frac{1}{100})$$

$$\sigma_i \sim \Gamma(0.1, 0.1), \quad i = 1, 2, 3, 4, 5$$

$$\pi_i \sim \text{Dir}(1), \quad i = 1, 2, 3, 4, 5$$

Hospitalizados:

$$Y(t) \sim BN(M, p_t)$$

$$p_t = \frac{M}{\mu(t) + M}$$

$$\begin{aligned}
\mu(t) &= M \sum_{i=1}^4 \pi_i \phi(t; t_{m,i}, \sigma_i) \\
&= M[\pi_1 \phi(t; t_{m,1}, \sigma_1) \\
&\quad + \pi_2 \phi(t; t_{m,2}, \sigma_2) \\
&\quad + \pi_3 \phi(t; t_{m,3}, \sigma_3) \\
&\quad + \pi_4 \phi(t; t_{m,4}, \sigma_4)]
\end{aligned}$$

Se propusieron las siguientes distribuciones a priori para los parámetros:

$$\begin{aligned}
t_{m_1} &\sim N(100, \frac{1}{100}) \\
t_{m_2} &\sim N_T(t_{m_1}, 170, \frac{1}{100}) \\
t_{m_3} &\sim N_T(t_{m_2}, 295, \frac{1}{100}) \\
t_{m_4} &\sim N_T(t_{m_3}, 510, \frac{1}{100}) \\
\sigma_i &\sim \Gamma(0.1, 0.1), \quad i = 1, 2, 3, 4 \\
\pi_i &\sim \text{Dir}(1), \quad i = 1, 2, 3, 4
\end{aligned}$$

Defunciones:

$$\begin{aligned}
Y(t) &\sim BN(M, p_t) \\
p_t &= \frac{M}{\mu(t) + M} \\
\mu(t) &= M \sum_{i=1}^4 \pi_i \phi(t; t_{m,i}, \sigma_i) \\
&= M[\pi_1 \phi(t; t_{m,1}, \sigma_1) \\
&\quad + \pi_2 \phi(t; t_{m,2}, \sigma_2) \\
&\quad + \pi_3 \phi(t; t_{m,3}, \sigma_3) \\
&\quad + \pi_4 \phi(t; t_{m,4}, \sigma_4)]
\end{aligned}$$

Se propusieron las siguientes distribuciones a priori para los parámetros:

$$\begin{aligned}
t_{m_1} &\sim N(50, \frac{1}{100}) \\
t_{m_2} &\sim N_T(t_{m_1}, 180, \frac{1}{100}) \\
t_{m_3} &\sim N_T(t_{m_2}, 300, \frac{1}{100}) \\
t_{m_4} &\sim N_T(t_{m_3}, 470, \frac{1}{100})
\end{aligned}$$

$$\begin{aligned}
\sigma_i &\sim \Gamma(0.1, 0.1), \quad i = 1, 2, 3, 4 \\
\pi_i &\sim \text{Dir}(1), \quad i = 1, 2, 3, 4
\end{aligned}$$

Mediante la elaboración de un programa en R ([R Core Team, 2021](#)) y uso del paquete JAGS ([Plummer, 2019](#)) se obtuvieron las distribuciones marginales posteriores de cada uno de los parámetros de interés mediante métodos de Cadenas de Markov Monte Carlo. Las estimaciones de los parámetros se obtuvieron como la media aritmética de las cadenas generadas de la distribución objetivo en cada uno de los parámetros. Para cada estimación de los

modelos propuestos, se generaron dos cadenas de tamaño 50,000 iteraciones, de las cuales se tomó un *burn-in* de 10,000 y un adelgazamiento de las cadenas de 40, obteniéndose 1,000 iteraciones por cada cadena. Finalmente, se generaron diagnósticos de convergencia de las cadenas correspondientes.

En las Figuras 5, 6 y 7 se presenta el ajuste del modelo de la mezcla de distribuciones gaussianas (en línea negra) sobre los datos observados (puntos azules) y su respectivo intervalo de credibilidad del 95% (en línea discontinua negra).

Para el caso de infecciones diarias se ajustó un modelo de mezclas con $K = 5$, y para el caso de defunciones y muertes diarias un modelo con $K = 4$. El primer pico representa la dinámica inicial de la pandemia, tras la introducción de medidas de control en marzo, y los otros picos capturan el aumento que ocurrió después de que las medidas fueran eliminadas en diversos grados. En la Tabla 1, se presentan las estimaciones Bayesianas de $t_{m,i}$, $i = 1, \dots, K$ para los tres casos estudiados en el estado de Guerrero; es decir, la fecha estimada del pico de cada rebrote.

La primera fecha de pico para infectados, hospitalizados y defunciones se localiza entre los meses de junio y julio de 2020. La segunda fecha de pico varía entre los meses de agosto de 2020 y octubre de 2020. La tercera fecha de pico se vio reflejada en el modelo en diciembre de 2020 para los casos hospitalizados y enero de 2021 para los casos de infectados y defunciones. Se identificó un cuarto rebrote para las tres bases de datos, con fecha de pico en julio de 2021 para las defunciones y en el mes de agosto del 2021 para

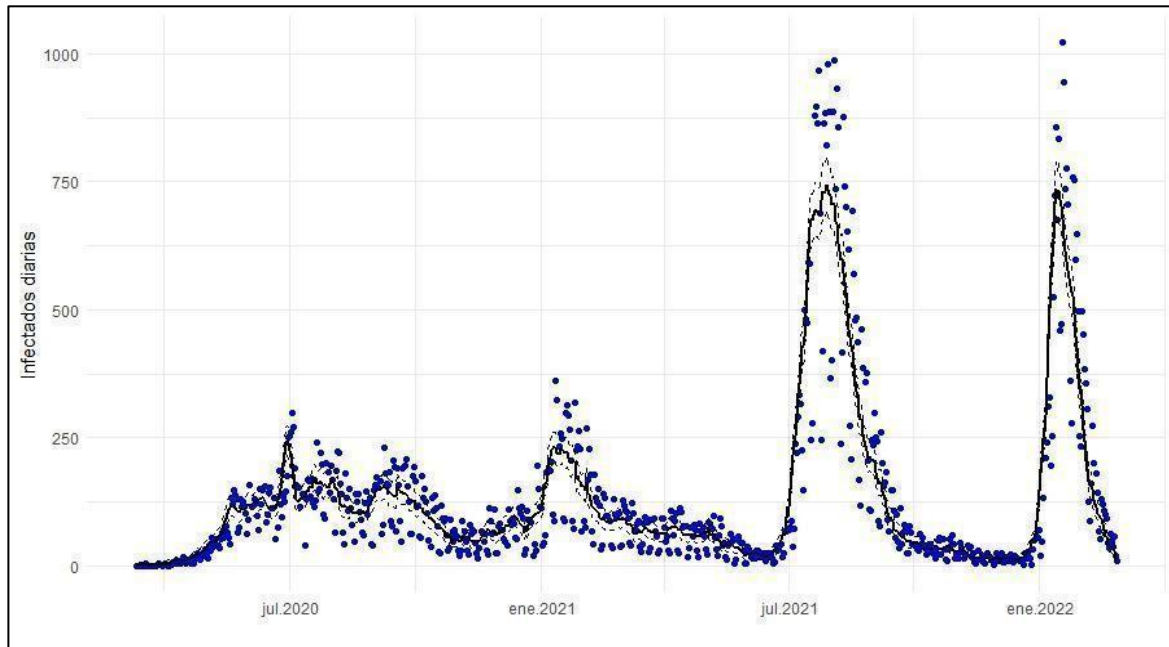


Figura 5. Ajuste del modelo de mezclas a los datos de casos de infectados diarios por COVID-19 en el estado de Guerrero.

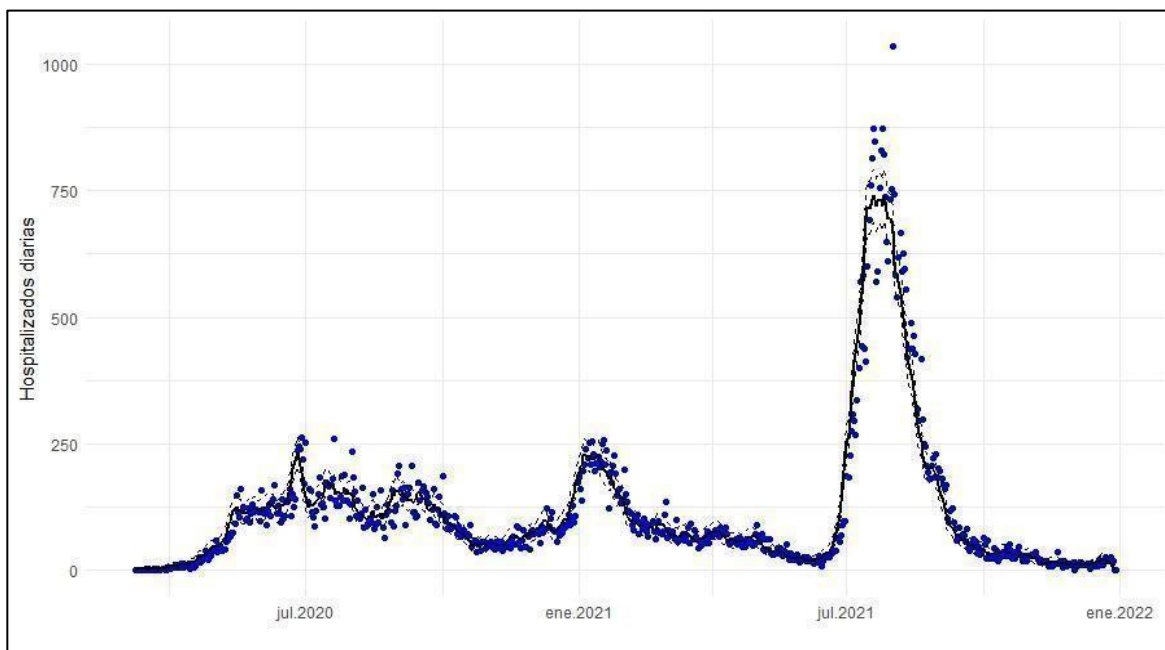


Figura 6. Ajuste del modelo de mezclas a los datos de casos hospitalizados por COVID-19 en el estado de Guerrero.

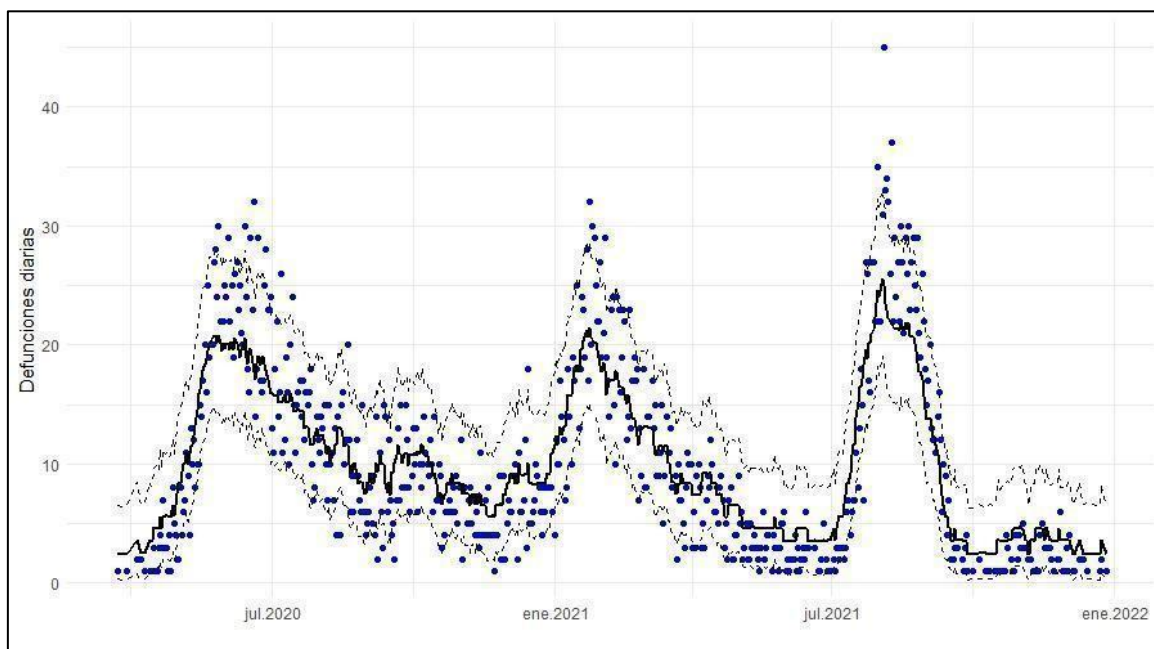


Figura 7. Ajuste del modelo de mezclas a los datos de casos de muertes por COVID-19 en el estado de Guerrero.

Tabla 1. Estimaciones del pico máximo de casos COVID-19 en Guerrero.

Casos	Rebote 1	Rebote 2	Rebote 3	Rebote 4	Rebote 5
Infectados	Julio 9, 2020	Sept 22, 2020	Ene 15, 2021	Agost 8, 2021	Ene 15, 2022
Hospitalizados	Jun 23, 2020	Agost 27, 2020	Dic 30, 2020	Agost 2, 2021	
Defunciones	Jun 6, 2020	Oct 3, 2020	Ene 31, 2021	Jul 30, 2021	

los casos diarios y para los hospitalizados. Por último, el modelo de mezcla logra captar un quinto rebrote para los datos de infectados localizando el máximo número de casos en el mes de enero de 2021.

Conclusiones

En este trabajo, se modeló matemáticamente el comportamiento de los rebrotes de infecciones

diarias, de hospitalizados y de muerte por COVID-19, a través de un modelo de mezcla de distribuciones gaussianas, desde los inicios de la pandemia hasta febrero de 2022 con datos reportados para el estado de Guerrero. Los modelos propuestos fueron analizados desde el punto de vista de la Estadística Bayesiana y medible el uso de software estadístico R y el paquete JAGS. Las estimaciones obtenidas fueron adecuadas, observando que los picos máximos se

obtuvieron durante o después de los periodos vacacionales de verano y diciembre.

Referencias

- Baleanu, D., Mohammadi, H., Rezapour, S. (2020). A fractional differential equation model for the COVID-19 transmission by using the Caputo–Fabrizio derivative. *Advances in Difference Equations*, 2020, 299.
<https://doi.org/10.1186/s13662-020-02762-2>
- Beira, M.J., Sebastião, P.J. (2021). A differential equations model-fitting analysis of COVID-19 epidemiological data to explain multi-wave dynamics. *Scientific Reports*, 11, 16312.
<https://doi.org/10.1038/s41598-021-95494-6>
- Calvetti1, D., Hoover, A., Rose, J., Somersalo, E. (2020). Bayesian dynamical estimation of the parameters of an SE(A)IR COVID-19 spread model.
[\[2005.04365\] Bayesian dynamical estimation of the parameters of an SE\(A\)IR COVID-19 spread model \(arxiv.org\)](https://arxiv.org/abs/2005.04365)
- Gelman, A., Chew, G.L., Shnaidman, M. (2004). Bayesian Analysis of Serial Dilution Assays. *Biometrics*, 60, 407-417.
<https://doi.org/10.1111/j.0006-341x.2004.00185.x>
- Green, P.H. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82, 711-732.
<https://doi.org/10.1093/biomet/82.4.711>
- Ghojogh, B., Ghojogh, A., Crowley, M., Karray, F. (2019). Fitting A Mixture Distribution to Data: Tutorial. *arXiv (Cornell University)*.
<http://export.arxiv.org/pdf/1901.06708>
- Kaciroti, N.A., Lumeng, C., Parekh, V., Boulton, M.L. (2021). A Bayesian Mixture Model for Predicting the COVID-19 Related Mortality in the United States. *American Journal of Tropical Medicine and Hygiene*, 104, 1484-1492.
<https://doi.org/10.4269/ajtmh.20-1147>
- Koufi, A.E., Koufi, N.E. (2021). Stochastic differential equation model of Covid-19: Case study of Pakistan. *Results in physics*, 34, 105218.
<https://doi.org/10.1016/j.rinp.2022.105218>
- Manevski, D., Gorenjec, N.R., Kejžar, N., Blagus, R. (2020). Modeling COVID-19 pandemic using Bayesian analysis with application to Slovene data. *Mathematical biosciences*, 329, 108466.
<https://doi.org/10.1016/j.mbs.2020.108466>
- Robert, C.P., Casella, G. (2005). *Monte Carlo Statistical Methods* (Springer Texts in Statistics). Springer-Verlag New York, Inc. eBooks.
<https://dl.acm.org/citation.cfm?id=1051451>
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
<https://www.R-project.org/>
- PAHO (2020). Epidemic diseases – Cumulative suspected and confirmed COVID-19 cases reported by countries and territories in the Americas. Pan American Health Organization.
[covid19-cumulative cases-03.31.20.pdf \(paho.org\)](https://paho.org/covid19-cumulative-cases-03.31.20.pdf)
- Plummer, M. (2019). *rjags: Bayesian Graphical Models using MCMC*. R package version 4-10.
<https://CRAN.R-project.org/package=rjags>
- Tejeira-Huacani, J.D. (2020). Análisis de un modelo matemático simple para la descripción de propagación de COVID-19. *Revista Boliviana de Física*, 37, 41-43.
http://www.scielo.org.bo/pdf/rbf/v37n37/v37n37_a06.pdf