



Facultad de Ciencias
Instituto de Estadística
Ingeniería en Estadística

Calibración de modelos epidemiológicos aplicados a datos de contagio del virus SARS-CoV-2 en Chile

Sebastián Andrés Caamaño Contreras
Diciembre 2022

Profesora Guía

Kerlyns Martínez. Ph.D.

Instituto de Estadística, Universidad de Valparaíso

Profesora Co-Guía

Alejandra Christen. Ph.D.

Instituto de Estadística, Universidad de Valparaíso

Proyecto de titulación para optar al:

grado académico de: *Licenciado en Estadística*

título profesional de: *Ingeniero en Estadística*

minor en: *Minería de Datos*

Resumen

En este proyecto de título se analizará el comportamiento de un modelo epidemiológico determinista donde se muestra la solución analítica y el comportamiento asintótico de los modelos SIS. Además, se expresará el modelo epidemiológico SIS estocástico, identificando los diferentes parámetros presentes en los sistemas epidemiológicos, y de esta forma poder calibrar y comparar los modelos.

El proyecto cuenta con diferentes períodos de estudio, a los cuales se les realizará una estimación de parámetros utilizando los datos reales del COVID-19 o SARS-CoV-2 en Chile desde Julio 2020 a Febrero 2022.

Agradecimientos

Quiero agradecer a toda mi familia, a mis padres Carlos Caamaño Sepúlveda y Etelvina Contreras Jara, por el apoyo que siempre me han brindado pese a la distancia, siempre me sentí apoyado en esta etapa de mi vida donde se presentaron alegrías y tristezas, ustedes son ese respaldo que me motivo a culminar con este proceso.

A mi hermano Rodrigo Caamaño Contreras, quiero agradecer de manera especial todo el apoyo que me has brindado tus consejos, tus enseñanzas me han permitido llegar lejos, eres la motivación y un ejemplo a seguir, solo puedo decir a echarle ganas porque nada es imposible si nos apoyamos.

A mis amigos, gracias por formar parte de mi vida, han sido el apoyo constante en los momentos de tristeza, alegrías, son parte fundamental de mi vida aunque mis acciones no lo demuestren, mis memes lo hacen.

Agradecer también de manera especial a mi profesora guía por todo el tiempo dedicado, los importantes aportes y sugerencias que contribuyeron al desarrollo de esta tesis. Su apoyo hizo que este proceso sea grato y enriquecedor, además, quiero agradecer el apoyo de los diferentes profesores que compartieron su conocimiento para guiarme y apoyarme cuando más lo necesite, gracias a ustedes todo esto es posible.

Índice general

Resumen	2
Agradecimientos	3
1. Introducción	6
1.1. Fundamentación teórica	6
1.1.1. Sistema dinámico	6
1.1.2. Modelos compartimentales	6
1.1.3. Modelos epidemiológicos	7
1.1.4. Modelo SIS determinista	8
1.1.5. Modelo SIS estocástico	13
1.1.6. Virus SARS-CoV-2	16
1.2. Revisión bibliográfica	17
2. Objetivos	19
2.1. Objetivo general	19
2.2. Objetivos específicos	19
3. Metodología	20
3.1. Ecuaciones diferenciales	20
3.1.1. Clasificación de ecuaciones diferenciales	20
3.1.2. Ecuaciones diferenciales estocásticas	21
3.2. Procedimientos de estimación y aproximación en ecuaciones diferenciales	23
3.2.1. Método de Euler	23
3.2.2. Método Euler–Maruyama	25
3.2.3. Cadenas de Markov	25
3.3. Modelo estadístico en el proceso de calibración	25
3.3.1. Modelo de calibración determinista	25
3.3.2. Modelo de calibración estocástica	25
3.3.3. Metropolis–Hastings	26
3.3.4. Hamiltonian Monte Carlo (HMC)	26
3.3.5. No U-Turn Sampler (NUTS)	27
3.3.6. PyMC 3	27
4. Base de datos y validación	28
4.1. Base de datos	28
4.2. Validación	29
4.2.1. Validación del proceso de calibración para el modelo determinista	29

4.2.2. Validación del proceso de calibración para el modelo estocástico	35
5. Resultados	41
5.1. Estimación de parámetros para el primer período	41
5.1.1. Estimación determinista	42
5.1.2. Estimación estocástica	43
5.2. Estimación de parámetros para el segundo período	44
5.2.1. Estimación determinista	44
5.2.2. Estimación estocástica	45
5.3. Estimación de parámetros para el tercer período	47
5.3.1. Estimación determinista	48
5.3.2. Estimación estocástica	49
5.4. Estimación de parámetros para el cuarto período	50
5.4.1. Estimación determinista	51
5.4.2. Estimación estocástica	52
5.5. Estimación de parámetros para el quinto período	53
5.5.1. Estimación determinista	53
5.5.2. Estimación estocástica	54
5.6. Comparación de estimaciones	56
6. Conclusiones	59
A. Código	60
Referencias	66

Capítulo 1

Introducción

La transmisión de una enfermedad es un fenómeno complejo donde distintos factores interfieren en la propagación de un virus. Entre las disciplinas que estudian estos fenómenos nos encontraremos con la epidemiología matemática, área donde se modela el comportamiento en el tiempo de una enfermedad con la ayuda de ecuaciones diferenciales.

El modelo de [Kermack y McKendrick \(1927\)](#) ha tenido gran influencia en el desarrollo de modelos epidemiológicos con la implementación del modelo SIR, introduciendo generalidades como lo son las tasas variables y recuperaciones, con el fin de predecir el comportamiento de un virus o enfermedad en la población.

El problema fundamental en epidemiología se puede resumir de la siguiente manera: uno o más infectados son introducidos en una población con cierto grado de susceptibilidad a la enfermedad. Ésta comienza a propagarse desde los individuos infecciosos a los individuos susceptibles; Los primeros dejan de serlo debido a que se recuperan o fallecen, y el número de personas que no son afectadas por la enfermedad va disminuyendo. Finalmente, después de un período de tiempo, la epidemia se extingue. Cabe preguntarse si ésta se detuvo debido a que ya no existen individuos susceptibles o por la interacción de varios factores, como facilidad de transmisión del agente infeccioso, recuperación del virus y mortalidad, por este motivo se estudiarán dos tipos de modelos para ver el comportamiento del virus SARS-CoV-2 en Chile.

1.1. Fundamentación teórica

1.1.1. Sistema dinámico

Los sistemas dinámicos nos permiten analizar y describir el cambio que experimenta una variable a lo largo del tiempo. Este cambio estará modelado por una ecuación (o sistemas de ecuaciones) que usualmente se expresarán como ecuaciones diferenciales, y permitirá predecir el comportamiento del sistema en cualquier instante del tiempo.

1.1.2. Modelos compartimentales

Los modelos epidemiológicos compartimentales están constituidos por una variedad de compartimientos, estos se identifican con letras, estas tienen en cuenta diferentes consideraciones, entre estas nos encontraremos con las letras S, I, R, V, M, E, entre otras; por ejemplo:

- S: todos los individuos **S**usceptibles al virus o enfermedad.
- I: todos los individuos **I**nfectados.
- R: todos los individuos **R**ecuperados y que quedan temporal o permanente inmunes.
- M: individuos con inmunidad pasiva. Esta inmunidad es temporal para algunas enfermedades, se presenta principalmente en recién nacidos gracias a los anticuerpos de la **M**adre.
- V: todos los individuos **V**acunados que son inmunes durante un periodo de tiempo o permanentemente.

Como se puede observar, existe una variedad de etiquetas compartimentales que ayudarán a construir un modelo epidemiológico. La complejidad del modelo dependerá de la cantidad de compartimientos que uno utilice para obtener mayor información. A medida que se aumentan los compartimientos, más complejo será el sistema de ecuaciones y encontrar la respuesta no será fácil. En ocasiones nos encontraremos con que el sistema no tendrá solución analítica sino que se requerirán métodos de aproximación numérica para obtener una solución aproximada.

El sistema compartimental debe estar balanceado, esto significa que la suma de las ecuaciones debe dar cero. Si deseamos agregar nuevos parámetros necesitaremos balancear la ecuación para que los modelos compartimentales puedan predecir las características de propagación de una enfermedad de manera correcta.

1.1.3. Modelos epidemiológicos

Existen dos grandes tipos de modelos epidemiológicos: deterministas y estocásticos ([García Roviра y cols., 2017](#)).

1. En un modelo determinista es **posible controlar todos los factores** que intervienen en el estudio del fenómeno, y **predecir sus resultados con exactitud**.
Por ejemplo, un modelo determinista puede ser escrito de la forma:

$$dX_t = b(X_t)dt, \quad (1.1)$$

donde la variable de estado es X y b es una función conocida.

2. En un modelo estocástico **no es posible controlar todos los factores**, sino se pueden calcular las probabilidades que ocurran ciertos eventos, entre estos podemos encontrar **la posibilidad de que la epidemia se extinga**.
Por ejemplo, un modelo estocástico puede ser escrito de la forma:

$$dX_t = b(X_t)dt + B(X_t)dW_t, \quad (1.2)$$

donde la variable de estado es X y las funciones conocidas son b , B y $(W_t; t \geq 0)$ es un proceso de Wiener (ver sección 3).

1.1.4. Modelo SIS determinista

Modelo SI, con $\mu \neq 0$

El siguiente modelo considera dos compartimientos, uno para la población de individuos susceptibles y otro para los individuos infectados, donde entran y salen individuos del compartimiento a diferentes tasas.

En el modelo SI nos encontraremos con: S_t , el número de individuos susceptibles; I_t , el número de individuos infectados; N , el número total de individuos; μ , la tasa de mortalidad y β , la tasa de contagio.

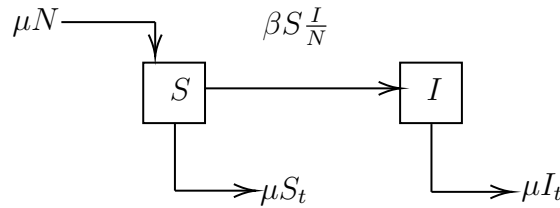


Figura 1.1: Modelo SI $\mu \neq 0$.
Fuente: Elaboración propia.

El modelo presentado en la Figura 1.1 se puede describir matemáticamente como un sistema de ecuaciones diferenciales:

$$\begin{cases} \frac{dS_t}{dt} = N\mu - S_t\beta\frac{I_t}{N} - \mu S_t, & \text{con } t \geq 0 \\ \frac{dI_t}{dt} = \mu S_t + S_t\beta\frac{I_t}{N} - N\mu. \end{cases} \quad (1.3a)$$

$$(1.3b)$$

La población se mantiene constante, con N el número total de individuos de la población, es decir:

$$S_t + I_t = N, \quad \text{para todo } t \geq 0. \quad (1.4)$$

Resolveremos el sistema de ecuaciones (1.3). Para esto despejamos S_t en (1.4) y sustituimos en (1.3b).

$$\begin{aligned} \frac{dI_t}{dt} &= \mu(N - I_t) + (N - I_t)\beta\frac{I_t}{N} - N\mu \\ \frac{dI_t}{dt} &= \beta I_t \left(-\frac{\mu}{\beta} + 1 - \frac{I_t}{N} \right). \end{aligned} \quad (1.5)$$

La ecuación (1.5) corresponde a una ecuación diferencial ordinaria (EDO) de variables separables. Para resolverla separaremos las variables e integraremos a ambos lados para encontrar la solución:

$$\begin{aligned}
\frac{dI_t}{\beta I_t \left(-\frac{\mu}{\beta} + 1 - \frac{I_t}{N} \right)} &= dt \\
\frac{1}{\beta} \int \frac{dI_t}{I_t \left(-\frac{\mu}{\beta} + 1 - \frac{I_t}{N} \right)} &= \int dt \\
\frac{1}{\beta} \int \frac{N\beta dI_t}{I_t (-N\mu + N\beta - \beta I_t)} &= t + c \\
-N \int \frac{dI_t}{I_t^2 \left(\frac{N(\mu - \beta)}{I_t} + \beta \right)} &= t + c
\end{aligned} \tag{1.6}$$

Calculamos la integral (1.6) por cambio de variable

$$\begin{aligned}
u &= \frac{N(\mu - \beta)}{I_t} + \beta, \quad \frac{du}{dI_t} = -\frac{N(\mu - \beta)}{I_t^2}, \quad dI_t = \frac{-I_t^2}{N(\mu - \beta)} du \\
-N \frac{\ln \left(\frac{N\mu - N\beta + \beta I_t}{I_t} \right)}{-N(\mu - \beta)} &= t + c \\
-\ln \left(\frac{N\mu - N\beta + \beta I_t}{I_t} \right) &= (\beta - \mu)(t + c).
\end{aligned} \tag{1.7}$$

De ahora en adelante, c se denotará como una constante arbitraria que podría cambiar entre líneas debido a manipulaciones algebraicas. Despejando I_t de la igualdad anterior obtenemos:

$$\begin{aligned}
\frac{I_t}{N\mu - N\beta + \beta I_t} &= ce^{(\beta - \mu)t} \\
I_t (1 - c\beta e^{(\beta - \mu)t}) &= ce^{(\beta - \mu)t} (N(\mu - \beta)) \\
I_t &= \frac{Nce^{(\beta - \mu)t}(\mu - \beta)}{1 - c\beta e^{(\beta - \mu)t}}
\end{aligned} \tag{1.8}$$

Luego de obtener I_t en la ecuación (1.8) procederemos a calcular la constante c , usando la condición inicial $I_0 = I(0)$:

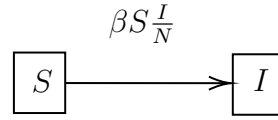
$$\begin{aligned}
I_0 &= \frac{Nc(\mu - \beta)}{1 - c\beta} \\
I_0(1 - c\beta) &= Nc(\mu - \beta) \\
I_0 &= c(N\mu - N\beta + \beta I_0) \\
c &= \frac{I_0}{N\mu - \beta S_0}
\end{aligned} \tag{1.9}$$

Obtenida la constante en (1.9), reemplazaremos en la Ecuación (1.8), donde la solución de (1.5) viene dada por

$$I_t = \frac{e^{(\beta - \mu)t}(\mu - \beta)NI_0}{N(\mu - \beta) + \beta I_0(1 - e^{(\beta - \mu)t})}. \tag{1.10}$$

Modelo SI, con $\mu = 0$

El modelo SI que no considera nacimientos y/o muertes es un caso particular del modelo planteado en el punto anterior. En la Figura 1.2 tenemos un diagrama que ilustra el comportamiento del sistema, donde se observa que los Susceptibles solo se convierten en Infectados y estos no rotan entre compartimientos, es decir, los Susceptibles una vez infectados no pueden recuperarse de la enfermedad y/o virus, lo que genera que el virus se extinga a medida que pase el tiempo.

Figura 1.2: Modelo SI con $\mu = 0$.

Fuente: Elaboración propia.

El diagrama anterior (Figura 1.2) se puede describir matemáticamente como un sistema de ecuaciones diferenciales de la siguiente forma:

$$\begin{cases} \frac{dS_t}{dt} = -S_t \beta \frac{I_t}{N}, & \text{con } t \geq 0 \\ \frac{dI_t}{dt} = S_t \beta \frac{I_t}{N}. \end{cases} \quad (1.11a)$$

$$(1.11b)$$

La población se mantiene constante con N el número total de individuos, de la población, es decir, la suma de los compartimientos será igual al total poblacional:

$$S_t + I_t = N, \quad \text{para todo } t \geq 0. \quad (1.12)$$

Como estamos en un caso particular del modelo SI con nacimientos y muertes visto previamente, el proceso de obtención de la solución es análogo, es por esta razón que solo se mostrará la solución de la ecuación:

$$I_t = \frac{N I_0 e^{\beta t}}{S_0 + I_0 e^{\beta t}}. \quad (1.13)$$

Modelo SIS con nacimientos y muertes

El modelo SIS contiene dos compartimientos, uno para la población de individuos susceptibles y otro para los individuos infectados, sistemas donde entran y salen individuos del compartimiento a diferentes tasas. Este modelo además contempla que toda la población se pueda re-infectar y volver al compartimiento de susceptibles, es decir, los individuos no generarán una inmunidad a la enfermedad o virus que se está estudiando.

En el modelo SIS nos encontraremos con: S_t , el número de individuos susceptibles; I_t , el número de individuos infectados; N , el número total de individuos; μ , la tasa de mortalidad; β , la tasa de contagio y γ , tasa de recuperación.

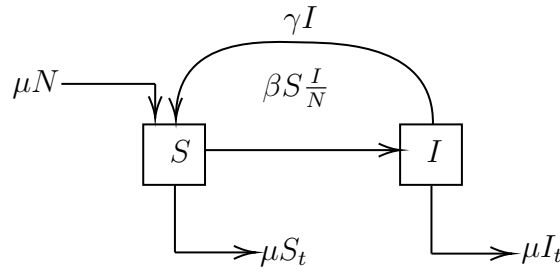


Figura 1.3: Modelo SIS con nacimientos y muertes.
Fuente: Elaboración propia.

El diagrama anterior (Figura 1.3) se puede escribir matemáticamente como un sistema de ecuaciones diferenciales de la siguiente forma:

$$\begin{cases} \frac{dS_t}{dt} = -S_t \beta \frac{I_t}{N} + \gamma I_t - \mu S_t + \mu N, & \text{con } t \geq 0 \\ \frac{dI_t}{dt} = S_t \beta \frac{I_t}{N} - \gamma I_t - \mu I_t. \end{cases} \quad (1.14a)$$

$$(1.14b)$$

La población se mantiene constante, con N el número total de individuos de la población, es decir, la suma de los compartimientos será igual al total poblacional.

$$S_t + I_t = N, \quad \text{para todo } t \geq 0. \quad (1.15)$$

Resolveremos el sistema de ecuaciones (1.14), para esto despejamos S_t en (1.15) y sustituimos en (1.14b).

$$\begin{aligned} \frac{dI_t}{dt} &= (N - I_t) \beta \frac{I_t}{N} - \gamma I_t - \mu I_t \\ \frac{dI_t}{dt} &= \beta I_t - \beta \frac{I_t^2}{N} - \gamma I_t - \mu I_t, \end{aligned} \quad (1.16)$$

que es una ecuación diferencial ordinaria (EDO) de segundo orden de tipo Bernoulli, con

$$P_0(t) \frac{dI}{dt} + P_1(t) I = Q(t) I^n \quad (1.17)$$

donde

$$n = 2, \quad P_0(t) = 1, \quad P_1(t) = -(\beta - \gamma - \mu), \quad Q(t) = \frac{-\beta}{N}.$$

Realizamos el cambio de $I_t = u^{-1}$ y $I'_t = -u^{-2}u'$ en la ecuación (1.16)

$$\begin{aligned} -u^{-2}u' - u^{-1}(\beta - \gamma - \mu) &= -\frac{\beta}{N}(u^{-1})^2 \quad \text{Multiplicamos por } (-u)^2 \\ u' + u(\beta - \gamma - \mu) &= \frac{\beta}{N} \end{aligned} \quad (1.18)$$

La EDO anterior es lineal y de primer orden del tipo $u' + P(t)u = Q(t)$, con

$$y' = u', \quad P(t) = \beta - \gamma - \mu, \quad Q(t) = \frac{\beta}{N}$$

Definimos el factor integrante como $Z(t) = e^{\int P(t)dt}$ y su solución es $Z(t) = e^{(\beta-\gamma-\mu)t}$

Multiplicando la ecuación (1.18) por $Z(t)$, notamos que $uZ(t) = \int Z(t)Q(t)dt$, de donde,

$$\begin{aligned} ue^{(\beta-\gamma-\mu)t} &= \frac{\beta}{N(\beta-\gamma-\mu)}e^{(\beta-\gamma-\mu)t} + c \\ u &= \frac{\beta e^{(\beta-\gamma-\mu)t} + cN(\beta-\gamma-\mu)}{e^{(\beta-\gamma-\mu)t}N(\beta-\gamma-\mu)}. \end{aligned} \quad (1.19)$$

Regresamos el cambio de variable:

$$\begin{aligned} \frac{1}{I_t} &= \frac{\beta e^{(\beta-\gamma-\mu)t} + cN(\beta-\gamma-\mu)}{e^{(\beta-\gamma-\mu)t}N(\beta-\gamma-\mu)} \\ I_t &= \frac{e^{(\beta-\gamma-\mu)t}N(\beta-\gamma-\mu)}{\beta e^{(\beta-\gamma-\mu)t} + cN(\beta-\gamma-\mu)} \end{aligned} \quad (1.20)$$

Luego de obtener I_t , procederemos a calcular la constante c teniendo en cuenta la condición inicial que $I_0 = I(0)$:

$$\begin{aligned} I_0 &= \frac{N(\beta-\gamma-\mu)}{\beta + cN(\beta-\gamma-\mu)} \\ \beta + cN(\beta-\gamma-\mu) &= \frac{N(\beta-\gamma-\mu)}{I_0} \\ cN(\beta-\gamma-\mu) &= \frac{N(\beta-\gamma-\mu) - \beta I_0}{I_0} \\ c &= \frac{N(\beta-\gamma-\mu) - \beta I_0}{NI_0(\beta-\gamma-\mu)} \\ c &= \frac{\beta(N - I_0) - N(\gamma + \mu)}{NI_0(\beta-\gamma-\mu)} \\ c &= \frac{\beta S_0 - N(\gamma + \mu)}{NI_0(\beta-\gamma-\mu)} \end{aligned} \quad (1.21)$$

Obtenida la constante en (1.21), reemplazaremos en la ecuación (1.19), donde se deduce:

$$I_t = \frac{NI_0(\beta-\gamma-\mu)e^{(\beta-\gamma-\mu)t}}{\beta I_0 e^{(\beta-\gamma-\mu)t} + \beta S_0 - N(\gamma + \mu)}. \quad (1.22)$$

Finalmente, el número de susceptibles vendrá dado por $S_t = N - I_t$.

Análisis del comportamiento asintótico

Se evaluará el comportamiento asintótico del modelo SIS debido a que es una generalización de los modelos discutidos anteriormente:

$$\begin{aligned} \frac{I_t}{N} &= \frac{\frac{I_0}{N}(\beta-\gamma-\mu)e^{(\beta-\gamma-\mu)t}}{\frac{\beta I_0 e^{(\beta-\gamma-\mu)t} + \beta S_0 - N(\gamma + \mu)}{N}} \\ \frac{I_t}{N} &= \frac{\frac{I_0}{N}(\beta-\gamma-\mu)e^{(\beta-\gamma-\mu)t}}{\beta \frac{I_0}{N} e^{(\beta-\gamma-\mu)t} + \beta \frac{S_0}{N} - (\gamma + \mu)} \end{aligned} \quad (1.23)$$

Evaluando el límite en el tiempo

$$\lim_{t \rightarrow \infty} \frac{I_t}{N} = \lim_{t \rightarrow \infty} \frac{\frac{I_0}{N}(\beta - \gamma - \mu)e^{(\beta - \gamma - \mu)t}}{\beta \frac{I_0}{N}e^{(\beta - \gamma - \mu)t} + \beta \frac{S_0}{N} - (\gamma + \mu)}, \quad (1.24)$$

obtenemos

$$\lim_{t \rightarrow \infty} \frac{I_t}{N} = \begin{cases} 1 - \frac{(\gamma + \mu)}{\beta} & , \text{ si } \beta > \gamma + \mu \\ 0 & , \text{ si } \beta \leq \gamma + \mu. \end{cases} \quad (1.25)$$

En conclusión en el modelo SIS determinista, si la tasa de contagio β es menor a la tasa de regeneración (la tasa de nacimientos más la tasa de recuperados, $\mu + \gamma$) entonces el virus desaparecerá.

Observando la Figura 1.4, notaremos el comportamiento asintótico de la función. A medida que el tiempo aumenta, el número de individuos infectados alcanzarán su punto máximo.

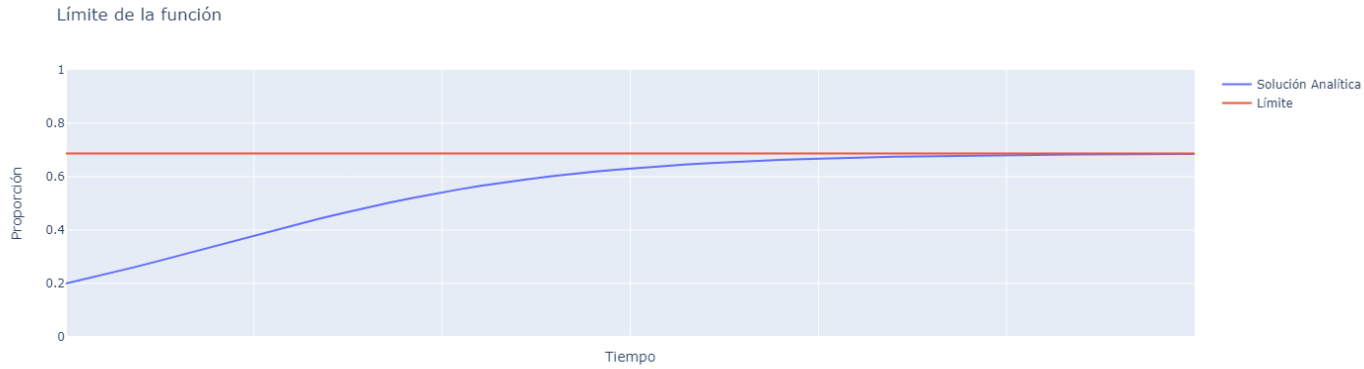


Figura 1.4: Límite de la función

Parámetros: $S_0 = 0.8$ $I_0 = 0.2$ $\mu = 0.15$ $\beta = 0.8$ $\gamma = 0.1$

Fuente: Elaboración propia

1.1.5. Modelo SIS estocástico

El modelo SIS estocástico contiene dos compartimientos, uno para la población de individuos susceptibles y otro para los individuos infectados, donde entran y salen individuos del compartimiento a diferentes tasas, con una perturbación que permitirá que la ecuación varíe en el tiempo. Este modelo además contempla que toda la población se pueda re-infectar y volver al compartimiento de susceptibles; los individuos no generarán una inmunidad a la enfermedad o virus que se está estudiando.

En el modelo SIS estocástico nos encontraremos con: S_t , el número de individuos susceptibles; I_t , el número de individuos infectados; N , el número total de individuos; μ , la tasa de mortalidad; β , la tasa de contagio; γ , tasa de recuperación; σ , coeficiente de variación y W_t ; proceso Browniano estándar.

Para mayor información sobre el proceso Browniano ver sección de metodología.

A continuación se observa el sistema de ecuaciones del modelo estocástico.

$$\begin{cases} \frac{dS_t}{dt} = \left[\mu N - \frac{\beta S_t I_t}{N} + \gamma I_t - \mu S_t \right] dt - \frac{\sigma S_t I_t}{N} dW_t & (1.26a) \\ \frac{dI_t}{dt} = \left[\frac{\beta S_t I_t}{N} - (\mu + \gamma) I_t \right] dt + \frac{\sigma S_t I_t}{N} dW_t & (1.26b) \end{cases}$$

La población se mantiene constante. De esta forma denotamos N a la población total de individuos, donde la suma de los compartimientos será igual al total poblacional:

$$S_t + I_t = N, \quad \text{para todo } t \geq 0. \quad (1.27)$$

La ecuación con las que se hará la estimación será:

$$\frac{dI_t}{N dt} = \left[\frac{I_t}{N} \left((\beta - \mu - \gamma) - \frac{\beta I_t}{N} \right) \right] dt + \sigma \left(\frac{I_t}{N} - \frac{I_t^2}{N^2} \right) dW_t, \quad (1.28)$$

es decir, se trabajará con la proporción de infectados.

Existencia de un modelo estocástico

El modelo SIS estocástico es un modelo especial. Para que el modelo tenga sentido, necesitamos asegurarnos que el modelo estudiado tiene una solución global única y que también la solución está dentro de $(0, N)$ siempre que se parta de ahí. En este contexto, [Gray, Greenhalgh, Hu, Mao, y Pan \(2011\)](#) establecen el siguiente resultado de existencia y unicidad de una solución:

Teorema 1 *Para cualquier valor inicial $I(0) = I_0 \in (0, N)$, la ecuación diferencial estocástica (1.28) tienen una solución global positiva con probabilidad uno, es decir:*

$$\mathbb{P}\{I(t) \in (0, N) \forall t \geq 0\} = 1. \quad (1.29)$$

Al igual que en el caso determinista, existen dos casos los cuales una enfermedad se puede extinguir o persistir en el tiempo. Los siguientes teoremas los extraemos del [Gray y cols. \(2011\)](#):

Extinción

Teorema 2 *Si*

$$R_0^s := R_0^D - \frac{\sigma^2 N^2}{2(\mu + \gamma)} = \frac{\beta N}{\mu + \gamma} - \frac{\sigma N^2}{2(\mu + \gamma)} < 1 \quad \text{y} \quad \sigma^2 \leq \frac{\beta}{N}, \quad (1.30)$$

entonces para cualquier valor inicial dado $I(0) = I_0 \in (0, N)$, entonces la solución de la ecuación estocástica (1.28) obedece

$$\lim_{t \rightarrow \infty} \sup \frac{1}{t} \log(I(t)) \leq \beta N - \mu - \gamma - 0.5\sigma^2 N^2 < 0 \quad \text{a.s.}; \quad (1.31)$$

es decir, $I(t)$ tiende a cero exponencialmente a.s. En otras palabras, la enfermedad se extingue con probabilidad uno.

En el Teorema 2 se requiere que la intensidad del ruido $\sigma \leq \frac{\beta}{N}$. El siguiente teorema cubre el caso cuando $\sigma^2 > \frac{\beta}{N}$.

Teorema 3 Si

$$\sigma^2 > \max\left(\frac{\beta}{N}, \frac{\beta^2}{2(\mu + \gamma)}\right) \quad (1.32)$$

entonces para cualquier valor inicial dado $I(0) = I_0 \in (0, N)$, entonces la solución de la ecuación estocástica (1.28) obedece

$$\lim_{t \rightarrow \infty} \sup \frac{1}{t} \log(I(t)) \leq -\mu - \gamma + \frac{\beta^2}{2\sigma^2} < 0 \quad a.s.; \quad (1.33)$$

es decir, $I(t)$ tiende a cero exponencialmente a.s. En otras palabras, la enfermedad se extingue con probabilidad uno.

Persistencia

Teorema 4 Si

$$R_0^s := \frac{\beta N}{\mu + \gamma} - \frac{\sigma^2 N^2}{2(\mu + \gamma)} > 1, \quad (1.34)$$

entonces para cualquier valor inicial dado $I(0) = I_0 \in (0, N)$, entonces la solución de la ecuación estocástica (1.28) obedece

$$\lim_{t \rightarrow \infty} \sup I(t) \geq \xi \quad a.s.; \quad (1.35)$$

y

$$\lim_{t \rightarrow \infty} \inf I(t) \leq \xi \quad a.s.; \quad (1.36)$$

donde

$$\xi = \frac{1}{\sigma^2} \left(\sqrt{\beta^2 - 2\sigma^2(\mu + \gamma)} - (\beta - \sigma^2 N) \right), \quad (1.37)$$

que es la única raíz en $(0, N)$ de

$$\beta N - \mu - \gamma - \beta \xi - 0.5\sigma^2(N - \xi)^2 = 0, \quad (1.38)$$

es decir, $I(t)$ se elevará hasta o por encima del nivel ξ con una frecuencia infinita y con probabilidad uno.

Estimación de la variación cuadrática

Se utilizará el estimador propuesto por Miao (2004), lo cual ayudará a establecer un punto de partida a la cadena de Markov asociada del proceso de estimación para obtener la variación real del proceso. Para establecer la consistencia de los estimadores de los parámetros de difusión, se necesitan estos dos Teoremas.

Teorema 5 Dado $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} X$, y $P(X = 0) = 0$, entonces $\frac{X_n}{Y_n} \xrightarrow{P} 1$.

Teorema 6 Dejamos X_t sea una solución fuerte de la ecuación diferencial estocástica $dX_t = \mu(X_t)dt + \sigma(X_t)dB_t$, y Γ , una partición de $[0, T]$, entonces

$$\sum_{\Gamma} |X(t_{j+1}) - X(t_j)|^2 \xrightarrow{P} \int_0^T \sigma^2(X_t)dt \quad (1.39)$$

como $|\Gamma| \rightarrow 0$.

dada la ecuación $dX_t = \mu(X_t)dt + \sigma g(X_t)dB_t$ y X_t tiene una solución fuerte, entonces

$$\frac{\sum_{\Gamma} |X(t_{j+1}) - X(t_j)|^2}{\sum_{\Gamma} g(X(t_j))(t_{j+1} - t_j)} \xrightarrow{P} \sigma^2 \quad \text{como } |\Gamma| \rightarrow 0. \quad (1.40)$$

El estimador de la variación cuadrática ayudara a tener un punto de partida al momento de estimar con los valores originales del SARS-CoV-2.

1.1.6. Virus SARS-CoV-2

La enfermedad por coronavirus (COVID-19) es una enfermedad infecciosa causada por el virus SARS-CoV-2 el cual pertenece a la familia de los coronavirus debido a que su superficie tienen puntas de coronas. La infección por el SARS-CoV-2 se detectó por primera vez en 2019.

Los coronavirus infectan a seres humanos y algunos animales. La mayoría de las personas infectadas por el virus experimentarán una enfermedad respiratoria de leve a moderada y se recuperarán sin requerir un tratamiento especial. Sin embargo, algunas enfermarán gravemente y requerirán atención médica. Las personas mayores y las que padecen enfermedades subyacentes, tienen más probabilidades de desarrollar una enfermedad grave. Cualquier persona, de cualquier edad, puede contraer la COVID-19 y enfermar gravemente o morir ([OMS, 2021](#)).

La mejor manera de prevenir y ralentizar la transmisión del virus es estar bien informado sobre la enfermedad y cómo se propaga el virus, mantener la distancia mínima de un metro con las personas, usar adecuadamente la mascarilla, es decir bien ajustada, sin que salga la nariz, lavándose las manos con desinfectante con frecuencia y vacunándose.

El virus puede propagarse desde la boca o nariz de una persona infectada, a través de pequeñas partículas producidas al toser, estornudar, hablar o respirar. Es importante adoptar buenas prácticas respiratorias, por ejemplo, tosiendo en la parte interna del codo flexionado, y quedarse en casa y autoaislarse hasta recuperarse si se siente mal ([OMS, 2021](#)).

Para prevenir la enfermedad del Coronavirus COVID-19 en Chile, la primera medida adoptada por el Gobierno fue crear un comité de crisis, encabezado por el Presidente y autoridades de salud, posteriormente en el mes de marzo del 2020 informan que se cerraran todas las fronteras terrestres, marítimas y áreas del país para el tránsito de personas extranjeras. Otra medida adoptada por el Gobierno fue instaurar el Estado de Excepción Constitucional de Catástrofe, este decreto permitió contar con mayores herramientas para combatir la pandemia, ya que permitía restringir la libertad y movilidad de las personas mediante medidas como cuarentenas, cordones sanitarios y toques de queda ([de Salud, 2022](#)).

1.2. Revisión bibliográfica

El modelo matemático que formulan [Kermack y McKendrick \(1927\)](#) es un modelo complejo para describir la epidemia de peste que sufría la India, al cual denominaron SIR (Susceptibles, Infectados y Recuperados).

En este modelo se establecieron los siguientes postulados básicos ([García Piñera y cols., 2014](#)):

- a) La enfermedad que iban a estudiar debía ser viral o bacteriana, debía transmitirse por contacto directo de persona a persona.
- b) Al inicio de la epidemia, solamente una fracción de la población estará contagiada.
- c) La población sería una población cerrada con excepción de las pocas personas inicialmente enfermas, todas las demás eran susceptibles de enfermarse.
- d) El individuo sufriría el curso completo de la enfermedad para al final recuperarse, adquiriendo inmunidad, o muriendo.
- e) La población total de personas sería constante.

[Lu \(2009\)](#) estudia un modelo SIRS (Susceptible-Infectado-Removido-Susceptible) con o sin retraso temporal distribuido e influenciado por perturbaciones aleatorias, que generaliza el modelo de [Tornatore, Buccellato, y Vetro \(2005\)](#) permitiendo que los individuos removidos vuelvan a ser susceptibles. Este modelo extiende los resultados incluyendo la posibilidad de que la inmunidad sea sólo temporal.

El trabajo de [Gray y cols. \(2011\)](#) generaliza el modelo epidémico SIS determinista a uno estocástico y describe su correspondiente ecuación diferencial estocástica, con el fin de determinar el comportamiento en tiempo largo del compartimiento de infectados. Se demuestra que la solución de la EDE es única y positiva, y se establecen las condiciones de extinción y persistencia de los infectados en el tiempo. Además del estudio analítico de la EDE se ilustran con simulaciones computacionales y aplicación a una enfermedad en la vida real.

Por otro lado, el artículo de [Zhao y Jiang \(2014\)](#) trata el comportamiento a largo plazo de un modelo epidemiológico determinista y estocástico que contempla la vacunación. Los autores buscan determinar la condición exacta para extinguir la pandemia en el caso determinista mientras que para el estocástico se busca determinar un umbral que tendrá la enfermedad. Posteriormente se realiza una simulación para corroborar los resultados obtenidos.

Para encontrar una aproximación a las diferentes ecuaciones estocásticas, [Higham \(2001\)](#) en su artículo ofrece una introducción a los métodos numéricos para las ecuaciones diferenciales estocásticas, donde aborda temas que incluyen la integración estocástica, el método de Euler-Maruyama, el método de Milstein convergencia fuerte y débil, estabilidad lineal y la regla de la cadena estocástica, todo esto realizado en MATLAB 10, esto nos permite tener una perspectiva de como programar los métodos antes mencionados.

El trabajo de [Cockayne y Duncan \(2021\)](#) presenta diferentes formas de calibrar modelos de ecuaciones diferenciales a gran escala con los datos observados. En este trabajo además expone la existencia de un cuello de botella al momento de realizar calibraciones con métodos avanzados. Al momento de calcular las derivadas de la función de pérdida con respecto a los

parámetros estimados necesitan varias soluciones numéricas de un sistema subyacente de ecuaciones diferenciales parciales u ordinarias. En este trabajo se presenta un nuevo enfoque que ayudará a calcular las sensibilidades de forma probabilística, lo cual ayudará a reducir el costo computacional y proporcionará una cuantificación probabilística de la incertidumbre que se estudia.

Capítulo 2

Objetivos

2.1. Objetivo general

Comparar modelos epidemiológicos calibrados deterministas y estocástico utilizando los datos del SARS-CoV-2 en Chile.

2.2. Objetivos específicos

Para poder alcanzar el objetivo general se plantean los siguientes objetivos específicos:

- Delimitar los períodos de tiempo donde se aplicarán los modelos epidemiológicos.
- Calibrar los modelos epidemiológicos.
- Analizar posibles desajustes del modelo debido a la implementación de la vacunación.
- Contrastar el ajuste de los modelos epidemiológicos seleccionados.

Capítulo 3

Metodología

3.1. Ecuaciones diferenciales

Las siguientes definiciones de ecuaciones diferenciales son extraídas del libro [Zill \(2009\)](#).

Una ecuación que contiene derivadas de una o más variables respecto a una o más variables independientes, a esto se le llama ecuación diferencial (ED).

Existen distintas clasificaciones para las ecuaciones diferenciales, entre estas tenemos las ecuaciones diferenciales por tipo, orden y linealidad.

3.1.1. Clasificación de ecuaciones diferenciales

Clasificación por tipo

Si una ecuación contiene sólo derivadas de una o más variables dependientes respecto a una sola variable independiente se dice que es una ecuación diferencial ordinaria (EDO). Una ecuación que involucra derivadas parciales de una o más variables independientes se llamará ecuación diferencial parcial (EDP).

Clasificación por orden

El orden de una ecuación diferencial (ya sea EDO o EDP) es el orden de la mayor derivada en la ecuación, la podemos expresar simbólicamente como una ecuación diferencial de n -ésimo orden con una variable dependiente por la forma general:

$$F(t, x, x', \dots, x^{(n)}) = 0, \quad (3.1)$$

donde F es una función con valores reales de $n + 2$ variables: $t, x, x', \dots, x^{(n)}$.

Clasificación por linealidad

Una ecuación diferencial de n -ésimo orden (3.1) se dice que es lineal si el campo escalar F es lineal en $x, x', \dots, x^{(n)}$. Esto significa que una EDO de n -ésimo orden es lineal cuando se expresa con una ecuación dada por:

$$a_n(t) \frac{d^n x}{dt^n} + a_{n-1}(t) \frac{d^{n-1} x}{dt^{n-1}} + \dots + a_1(t) \frac{dx}{dt} + a_0(t)x = g(t) \quad (3.2)$$

3.1.2. Ecuaciones diferenciales estocásticas

A lo largo de este documento, definiremos el espacio de probabilidad como $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$, con el espacio de probabilidad completo con una filtración $\{\mathcal{F}_t\}_{t \geq 0}$ que satisface las condiciones habituales, es decir, es creciente y continua por la derecha, mientras que \mathcal{F}_0 contiene todos los conjuntos \mathbb{P} -nulos definidos.

La σ -álgebra \mathcal{F}_t contiene a todos los eventos que involucran a las variables del proceso hasta el tiempo t .

Proceso Estocástico

Un proceso estocástico es

- I) una colección $\{X(t) | t \geq 0\}$ de variables aleatorias
- II) para cada punto $w \in \Omega$, el mapeo de $t \mapsto X(t, w)$ es una trayectoria muestral.

Si realizamos un experimento y observamos los valores aleatorios de $X(\cdot)$ a medida que avanza en el tiempo, estaremos viendo la trayectoria muestral de $\{X(t, w) | t \geq 0\}$ para un $w \in \Omega$ fijo. Si volvemos a realizar el experimento, es probable que observemos una trayectoria muestral diferente debido a que es un proceso aleatorio (Evans, 2006).

Una ecuación diferencial estocástica (EDE) es una ecuación diferencial que incluye la posibilidad de efectos aleatorios. Desde la notación utilizada en Carmona (2009), podemos imaginar una EDE como una perturbación de una EDO:

$$\begin{aligned} \frac{dX}{dt} &= b(X_t) + B(X_t)\kappa(t), \quad 0 \leq t \leq T, \\ X(0) &= Z \end{aligned} \tag{3.3}$$

donde $b : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times 1}$, $B : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ y $\kappa(t) : \text{denota el ruido blanco } m \text{ dimensional}$. La solución $X(t)$ será entonces un proceso estocástico $X(t) = X(t, s)$.

Escrito de su forma integral:

$$X(t) = Z + \int_0^t b(X_s)ds + \int_0^t B(X_s)dW_s, \tag{3.4}$$

donde $dW = \kappa(t)dt$ o sea, $W(t) = \int_0^t \kappa(s)ds$. A $W(t)$ se le denomina proceso de Wiener o movimiento Browniano y como su definición lo sugiere, es la variable que acumula el ruido producido de 0 a t .

Lo anterior corresponde a una forma heurística de ver una EDE. Formalmente, un movimiento Browniano estándar unidimensional es un proceso estocástico $\{W_t | t \geq 0\}$ tal que:

- a) $W_0 = 0$ casi seguramente.
- b) Las trayectorias $t \rightarrow W_t$ son continuas.
- c) El proceso tiene incrementos independientes.
- d) La variable $W_t - W_s$ tiene distribución $N(0, t - s)$ para $0 \leq s < t$.

El movimiento Browniano surge como nombre gracias a los aportes de Robert Brown, un biólogo y botánico que en (1826-27), mientras observaba las partículas de polen en el agua, señaló que las partículas se movían a través del líquido. Brown no fue capaz de determinar los mecanismos que provocaron este movimiento; sin embargo, señaló que la trayectoria de una partícula determinada es muy irregular, no tiene tangente en ningún punto, además, los movimientos de dos partículas distintas parecerán ser independientes.

En 1900, Louis Bachelier intentó describir matemáticamente las fluctuaciones de las cotizaciones bursátiles y, esencialmente, descubrió en primer lugar ciertos resultados que más tarde fueron retomados y ampliados por Albert Einstein en 1905 (Evans, 2006). Norbert Wiener durante los años 20 (1920) hacia adelante colocaría la teoría con una sólida fundación matemática y se conocería como Proceso de Wiener.

El movimiento Browniano físico se presenta en tres dimensiones y es completamente errático. En la Figura 3.1 puede apreciarse una trayectoria Browniana cuando ésta se proyecta en el tiempo.

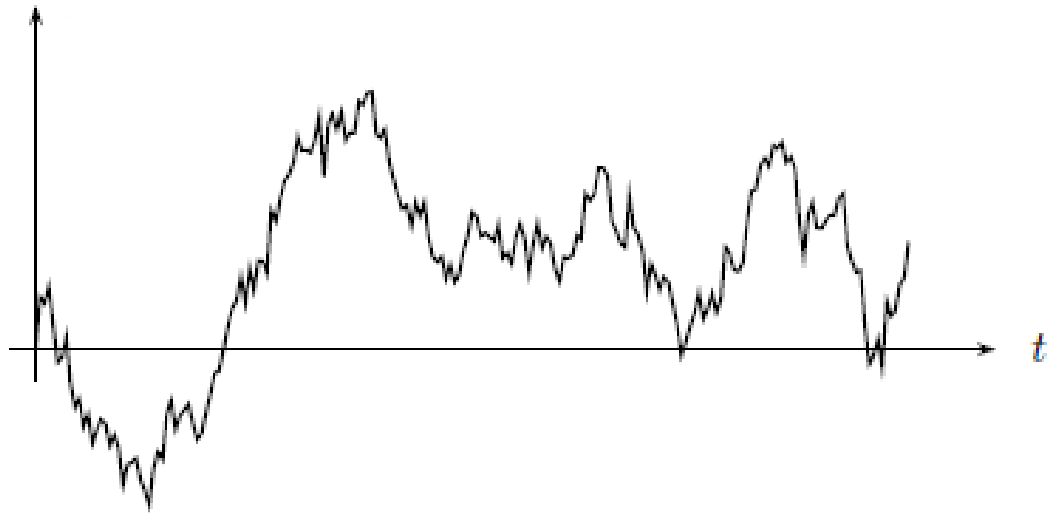


Figura 3.1: Una trayectoria del movimiento Browniano unidimensional.

Fuente: Rincón (2006).

Integral de Itô

La integral de Itô es fundamental para el estudio de cálculos estocásticos. La podemos apreciar en la ecuación (3.4), la cual es:

$$\int_0^t B(X_s) dW_s, \quad (3.5)$$

donde $B(X)$ es un proceso cuadrado-integrable ($\mathbb{E}(\int_0^t B^2(X_s) ds) < \infty$) adaptado a la filtración generada por movimiento Browniano W . El resultado de la integral es otro proceso estocástico, debido a que la integral desde 0 hasta algún valor particular t es una variable aleatoria.

La trayectoria de un movimiento Browniano no satisface los requisitos necesarios del cálculo infinitesimal tradicional, es por esto que las trayectorias son no diferenciables en casi todo punto y tienen variación infinita. Por lo tanto, es necesario construir la integral estocástica con respecto al movimiento Browniano.

La integral de Itô puede definirse de manera similar a la integral de Riemann-Stieltjes ([Oksendal, 2013](#)), que es el límite probabilista de una suma de Riemann, dicho límite no existe en general en términos de trayectorias. Si Π_n es una sucesión de particiones del intervalo $[0, t]$ con el diámetro del elemento más grande de la partición tendiendo a cero, entonces la integral de Itô de B con respecto al proceso Browniano W viene dada por el límite:

$$\int_0^t B(X_s) dW_s = \lim_{n \rightarrow \infty} \sum_{[t_{i-1}, t_i] \in \Pi_n} B_{t_{i-1}} (W_{t_i} - W_{t_{i-1}}). \quad (3.6)$$

3.2. Procedimientos de estimación y aproximación en ecuaciones diferenciales

Existen diferentes formas de resolver ecuaciones diferenciales. En ocasiones nos encontraremos con ecuaciones o sistema de ecuaciones diferenciales que no se pueden resolver de manera analítica o no tienen solución aparente, y es por eso que surgen métodos de aproximación numérica que ayudarán a obtener una aproximación a la solución real a través de procesos iterativos.

3.2.1. Método de Euler

El método de Euler se utiliza para aproximar soluciones con métodos numéricos donde se necesita un valor inicial para seguir avanzando en la dirección indicada por el campo direccional. Esta se detiene después de un corto tiempo estipulado, para examinar la pendiente en la nueva ubicación, y sigue avanzando en esa dirección y luego sigue replicando los pasos anteriores. Este método produce una aproximación a la solución real dado un valor inicial ([Stewart, 2012](#)). Los valores aproximados para la solución del problema con valor inicial $x' = F(t, x)$, $x(t_0) = x_0$, con tamaño de paso Δ , es decir, $t_n = t_{n-1} + \Delta$, son:

$$x_n = x_{n-1} + \Delta F(t_{n-1}, x_{n-1}) \quad n = 1, 2, 3, \dots \quad (3.7)$$

En la Figura 3.2 se observa la aproximación del modelo SIS (1.14), donde se aplica la ecuación (3.7) con un $\Delta = 0.6$. Al escoger un valor pequeño para el tamaño de paso, notamos que la aproximación se sobrepone a la solución analítica, donde la diferencia es mínima a simple vista.

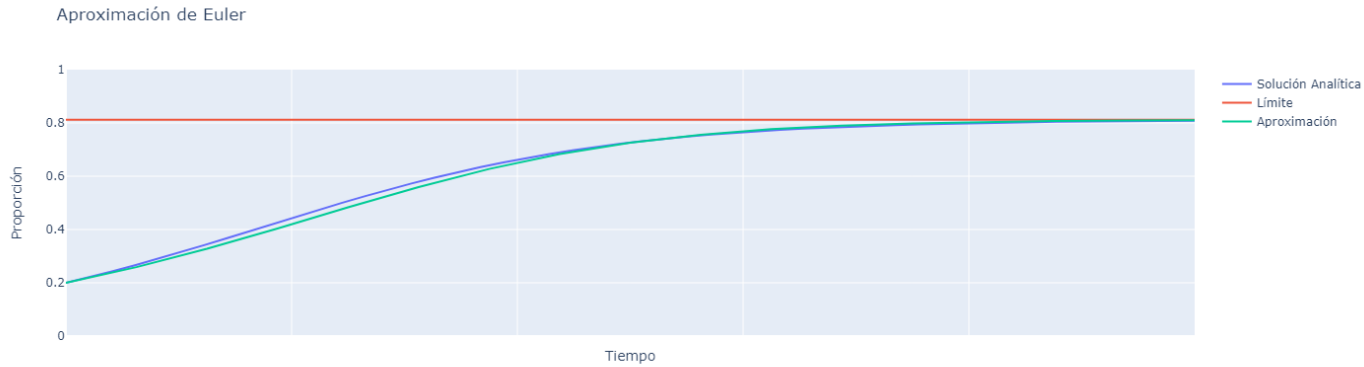


Figura 3.2: Aproximación de Euler

Parámetros: $S_0 = 0.8$ $I_0 = 0.2$ $\mu = 0.15$ $\beta = 0.8$ $\gamma = 0$ $\Delta = 0.6$

Fuente: Elaboración propia.

En la aproximación del modelo SIS (1.14), en la Figura 3.3, se utiliza el $\Delta = 1.9$. De aquí se puede observar que al definirlo con un paso de tiempo alto, la aproximación no se sobrepone a la solución analítica y el margen de error es mayor en comparación con la Figura 3.2.

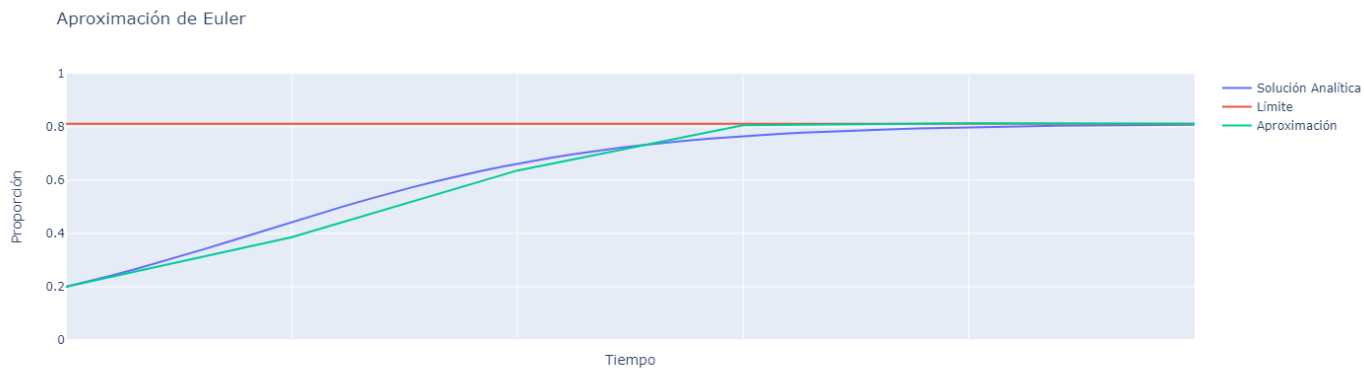


Figura 3.3: Aproximación de Euler

Parámetros: $S_0 = 0.8$ $I_0 = 0.2$ $\mu = 0.15$ $\beta = 0.8$ $\gamma = 0$ $\Delta = 1.9$

Fuente: Elaboración propia.

3.2.2. Método Euler–Maruyama

El método de Euler-Maruyama es una de las aproximaciones numéricas más sencillas para la ecuación diferencial estocástica (EDE) (Bayram, Partal, y Orucova Buyukoz, 2018). Si truncamos la fórmula de Itô de la serie estocástica de Taylor después de los términos de primer orden, obtenemos el método de Euler o método de Euler-Maruyama de la siguiente forma:

$$X(t_{i+1}) = X(t_i) + b(X(t_i))\Delta t + B(X(t_i))\Delta W_i \quad (3.8)$$

para todo $i = 0, 1, 2, \dots, N - 1$ con la condición inicial $X(t_0) = X_0$, $\Delta W_i \sim N(0, \Delta t)$.

3.2.3. Cadenas de Markov

Una cadena de Markov se dice que es un tipo especial de proceso estocástico discreto en la que la probabilidad de que ocurra un evento depende únicamente del pasado inmediato (Gil Sáez y cols., 2020). Esto se debe a que la cadena de Markov es una sucesión de variables aleatorias que toman valores en un conjunto numerable S , el conjunto de estados, tales que

$$P(\Theta_{n+1} = \theta | \Theta_1 = \theta_1, \Theta_2 = \theta_2, \dots, \Theta_n = \theta_n) = P(\Theta_{n+1} = \theta | \Theta_n = \theta_n) \quad (3.9)$$

para cualquier $n \geq 0$ y $\theta, \theta_1, \dots, \theta_{n-1} \in S$. Además supondremos que la probabilidad indicada es independiente en n .

3.3. Modelo estadístico en el proceso de calibración

3.3.1. Modelo de calibración determinista

La idea es considerar $I^{\text{obs}}(t) = \hat{I}(t, \theta) + \varepsilon$, ε un ruido observacional y donde $\hat{I}(t)$ denota la salida del modelo determinista al tiempo t y θ corresponde al vector de parámetros del modelo, Considerando distribuciones a priori $p(\theta)$:

$$\mu, \beta, \gamma, \varepsilon \sim \text{Gamma}(\alpha, \lambda), \quad (3.10)$$

α, λ hiperparámetros arbitrarios, queremos actualizar la distribución tomando en cuenta las observaciones $\{I_t^{\text{obs}}, 0 \leq t \leq T\}$:

$$\pi(\theta | I_t^{\text{obs}}) \propto p(\theta) \mathcal{L}(I_t^{\text{obs}} | \theta), \quad (3.11)$$

con \mathcal{L} la verosimilitud de I_t^{obs} dado θ . Para esto se construye una Cadena de Markov $\{\theta_n; n \geq 0\}$ que converge (en distribución) a la distribución posteriori $\pi(\theta | I_t^{\text{obs}})$

3.3.2. Modelo de calibración estocástica

Para el modelo estocástico se considerara $I^{\text{obs}}(t) = \hat{I}(t, \theta)$, donde $\hat{I}(t)$ denota la salida del modelo estocástico al tiempo t y θ corresponde al vector de parámetros para dicho modelo, Considerando distribuciones a priori $p(\theta)$:

$$\mu, \beta, \gamma, \sigma \sim \text{Beta}(\alpha, \lambda), \quad (3.12)$$

α, λ hiperparámetros arbitrarios, queremos actualizar la distribución tomando en cuenta las observaciones $\{I_t^{obs}, 0 \leq t \leq T\}$:

$$\pi(\theta|I_t^{obs}) \propto p(\theta)\mathcal{L}(I_t^{obs}|\theta), \quad (3.13)$$

con \mathcal{L} la verosimilitud de I_t^{obs} dado θ . Para esto se construye una Cadena de Markov $\{\theta_n; n \geq 0\}$ que converge (en distribución) a la distribución posteriori $\pi(\theta|I_t^{obs})$

3.3.3. Metropolis–Hastings

El algoritmo de Metropolis–Hastings es un método de *Markov chain Monte Carlo* (MCMC). Se utiliza principalmente como una forma de simular observaciones de distribuciones difíciles de manejar (Hitchcock, 2003). El algoritmo produce una cadena de Markov cuya distribución límite de miembros es la densidad objetivo $\pi(\theta)$. En la etapa j , una observación θ_j se genera a partir de una densidad instrumental $q(\cdot|\theta_i)$ que suele ser fácil de simular.

La observación candidata se convierte en el siguiente valor de la cadena de Markov con probabilidad:

$$\rho = \min \left\{ \frac{\pi(\theta_j)q(\theta_i|\theta_j)}{\pi(\theta_i)q(\theta_j|\theta_i)}, 1 \right\}; \quad (3.14)$$

con probabilidad $1 - \rho$, se establece $\theta_j = \theta_i$ como el valor anterior de la cadena de Markov.

El problema que tiene el algoritmo de Metropolis–Hastings es relacionado con el valor inicial. Dado a que elegimos aleatoriamente un valor, es muy posible que se encuentre en una región bastante pequeña (pensar que estamos en la cola de nuestra distribución). Si comienza aquí, podría pasar una cantidad desproporcionada de tiempo atravesando valores de baja densidad. Esto nos dará una sensación incorrecta de los valores, porque crearemos que estos valores ocurren más a menudo.

3.3.4. Hamiltonian Monte Carlo (HMC)

Para obtener los parámetros de una ecuación diferencial ordinaria se utilizará la simulación de Hamiltonian Monte Carlo (HMC) que es un algoritmo de Monte Carlo con cadena de Markov (MCMC) que evita el comportamiento de paseo aleatorio y la sensibilidad a los parámetros correlacionados que afectan a muchos métodos MCMC, tomando una serie de pasos informados por la información del gradiente de primer orden. Estas características le permiten converger a distribuciones objetivo de alta dimensión mucho más rápidamente que otros métodos más sencillos, como el paseo aleatorio de Metrópolis o el muestreo de Gibbs (Hoffman, Gelman, y cols., 2014).

Introducimos una variable de momento auxiliar r_d para cada variable del modelo d . En la implementación habitual, estas variables de momento se extraen independientemente de la distribución normal estándar, lo que da lugar a la densidad conjunta no normalizada (Hoffman y cols., 2014).

El proceso matemático se puede escribir de la siguiente forma:

$$p(\theta, r) \propto \exp\left\{\mathcal{L}(\theta) - \frac{1}{2}r \cdot r\right\}, \quad (3.15)$$

donde $r \cdot r$ es el producto punto usual entre vectores.

3.3.5. No U-Turn Sampler (NUTS)

No-U-Turn Hamiltonian Monte Carlo tiene como objetivo idear un muestreo de Markov chain Monte Carlo (MCMC) que conserve la capacidad de Hamiltonian Monte Carlo (HMC) de suprimir el comportamiento de paseo aleatorio sin necesidad de establecer el número L de pasos de salto que el algoritmo toma para generar una propuesta.

NUTS supera este problema mediante un algoritmo recursivo que recuerda al procedimiento de duplicación ideado por Neal (2003) para el muestreo de trozos, comienza introduciendo una variable de corte u con distribución condicional $p(u|\theta, r) = U(u; [0, \exp\{\mathcal{L}(\theta) - \frac{1}{2}r \cdot r\}])$, lo que hace que la distribución condicional $p(\theta, r|u) = U(\theta, r; \{\theta', r' \exp\{\mathcal{L}(\theta) - \frac{1}{2}r \cdot r\} \geq u\})$.

Tras el remuestreo $u|\theta, r$, en el artículo Neal (2003) describe el proceso de NUTS como un integrador de salto para trazar un camino hacia adelante y hacia atrás en el tiempo ficticio, primero corriendo hacia adelante o hacia atrás 1 paso, luego hacia adelante o hacia atrás 2 pasos, luego hacia adelante o hacia atrás 4 pasos, etc. Este proceso de duplicación construye implícitamente un árbol binario balanceado cuyos nodos hoja corresponden al estado de momento de posición. La duplicación se detiene cuando la subtrayectoria desde el nodo más a la izquierda hasta el más a la derecha de cualquier subárbol equilibrado del árbol binario global comienza a duplicarse sobre sí misma (es decir, la partícula ficticia comienza a dar una vuelta en U). En este punto, NUTS detiene la simulación y toma muestras de entre el conjunto de puntos calculados durante la simulación, teniendo cuidado de preservar el equilibrio detallado.

3.3.6. PyMC 3

PyMC 3, es una librería de Python que se utiliza para ejecutar modelos estadísticos bayesianos y de aprendizaje automático probabilístico. Esta librería se centra en algoritmos avanzados de Markov chain Monte Carlo (MCMC) y ajustes de variación.

En la librería nos encontraremos con diferentes algoritmos de MCMC basados en gradientes y los no basados en gradientes para la inferencia bayesiana, métodos bayesianos estocásticos y variaciones basados en gradientes para la inferencia bayesiana aproximada.

PyMC se ha utilizado para resolver diferentes problemas de inferencia en varios ámbitos científicos. Entre las disciplinas que utilizan esta librería nos entraremos con la astronomía, epidemiología, biología molecular, química, entre otros.

Los algoritmos que se basan en MCMC y están presentes en la librería son: No-U-Turn sampler (NUTS), Metropolis–Hastings, Secuencia de Monte Carlos para aproximaciones bayesianas. Nosotros nos concentraremos en el muestreo a partir del método NUTS.

Capítulo 4

Base de datos y validación

4.1. Base de datos

En el presente trabajo se utilizará la base de datos del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación disponibles en [GitHub](#) o en el url: <https://github.com/MinCiencia/Datos-COVID19> que se basa en el “Informe Epidemiológico” y el “Reporte Diario” emitido por el Ministerio de Salud de Chile, conforme las normas disponibles en www.minsal.cl específicamente las dispuestas para datos de carácter estadístico.

Al momento de trabajar con el conjunto de datos se tomará en cuenta, el estado de excepción constitucional de catástrofe, fecha de vacunación y las fechas en las cuales llegaron nuevas variantes a Chile, esto es de vital importancia debido a que pueden desajustar el modelo epidemiológico. Los períodos de tiempos se encuentran en formato (año-mes), y han sido establecidos según ([de Salud, 2022](#)):

1. Estado de Excepción Constitucional de Catástrofe (2020-03 al 2021-09)
2. Fecha de vacunación masiva (2021-03) hasta la actualidad.
3. Fecha de variantes
 - a) Inglaterra (B.1.1.7) (2020-12)
 - b) Delta (2021-06)
 - c) Omicron (2021-12)

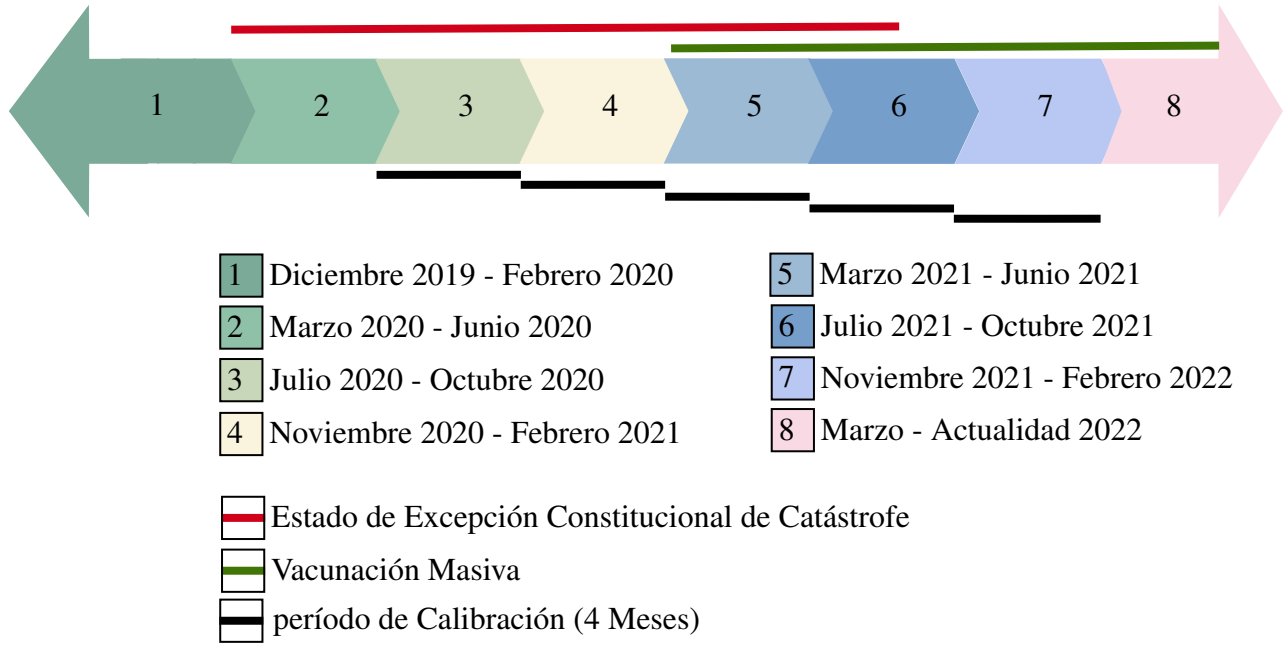


Figura 4.1: Línea Temporal.
Fuente: Elaboración propia.

4.2. Validación

4.2.1. Validación del proceso de calibración para el modelo determinista

Para validar el proceso de calibración, la estimación de los parámetros del modelo determinista obtenida en el Modelo SIS con nacimientos y muertes (1.14), se generará un conjunto de datos aleatorios con parámetros fijos que serán remplazados con la solución de la ecuación del modelo SIS. Con esto se estimarán los parámetros teóricos del conjunto de datos con la librería Pymc3 en el cual se aplicará el método NUTS. Además, se realizará una regresión por mínimos cuadrados para obtener las estimaciones de los parámetros. De esta forma se busca comparar las estimaciones realizadas.

A fin de optimizar el proceso de simulación se generará un vector θ que facilitará el proceso de simulación, debido a que se reduce la dimensión a 2 parámetros. Así, si el primer parámetro $[\beta - (\mu + \gamma)] \leq 0$, la enfermedad desaparecerá y si $[\beta - (\mu + \gamma)] > 0$ la enfermedad persistirá en el tiempo. El segundo parámetro del vector es β :

$$\theta = ([\beta - (\mu + \gamma), \beta]) \quad (4.1)$$

Con el propósito de validar el proceso de calibración se generarán dos conjuntos de datos que cumplen con la persistencia y el fin de la enfermedad en el tiempo.

Caso 1: persistencia

Para iniciar con el proceso de calibración se genera un conjunto de datos que cumple con la condición de persistencia de la enfermedad en el tiempo donde los parámetros del modelo se detallan en el Cuadro 4.1.

Parámetro	Valor
β	0.7
γ	0.1
μ	0.3
θ	[0.3 , 0.7]
I_0	0.2
ε	0.25
N	80

Cuadro 4.1: Parámetros originales, SIS caso 1.

Fuente: Elaboración propia.

En la Figura 4.2 se observa un diagrama de dispersión del conjunto de datos simulados. A estos datos se le agregó un ruido observacional para simular un caso real. Para lograr esto se tomaron muestras aleatorias del conjunto de datos simulados y se le aplica una transformación Log-normal. La figura presenta una linea azul que representa la solución de la ecuación diferencial ordinaria. Esto nos permite ver la dispersión de los valores simulados.

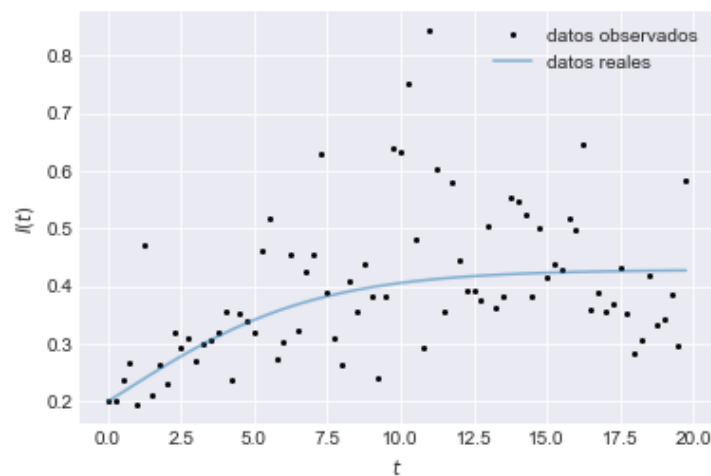


Figura 4.2: Datos simulados, SIS caso 1.

Fuente: Elaboración propia.

Al aplicar el ajuste de mínimos cuadrados al conjunto de datos simulados (ver Cuadro 4.2), los valores estimados son muy cercanos a los parámetros reales, pese a la dispersión que presenta el conjunto de datos.

Parámetros	Valor real	Estimación
$\beta - (\mu + \gamma)$	0.3	0.371
β	0.7	0.832

Cuadro 4.2: Estimación por mínimos cuadrados, SIS caso 1.

Fuente: Elaboración propia.

Para el proceso de simulación con cadenas de Markov Monte Carlo se utilizó como distribución a priori para el primer parámetro una distribución normal $\mathcal{N}(0, 1)$ para el segundo una distribución Cauchy acotada (1,1). El criterio de selección de apriori sera utilizar distribuciones que no presenten un 10 % de divergencia en sus estimaciones, si sobrepasa ese umbral se buscara otra distribución apriori.

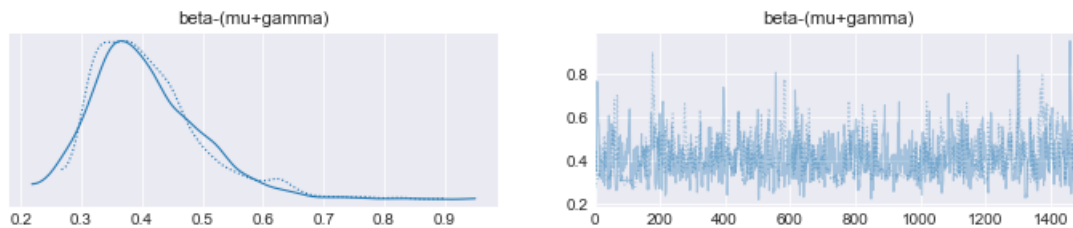
Se crearán dos cadenas con 3000 iteraciones y se aplicará el método NUTS. En el Cuadro 4.3 se encuentra un resumen de la cadena creada.

	Mean	Sd	2.5 %	97.5 %	Mcse mean	Mcse sd	R hat
$\beta - (\mu + \gamma)$	0.410	0.092	0.259	0.583	0.005	0.003	1.0
noise hiperparámetro	0.254	0.022	0.216	0.294	0.001	0.001	1.0
β	0.966	0.241	0.589	1.438	0.013	0.009	1.0

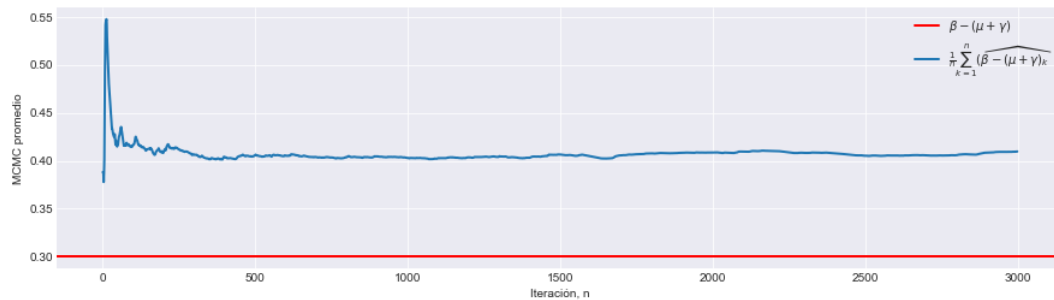
Cuadro 4.3: Resumen de simulación NUTS, SIS caso 1.

Fuente: Elaboración propia.

En la Figura 4.3 (a) se observa la distribución posteriori y la cadena de valores que genera NUTS. Para el primer parámetro del vector, la cadena se observa el comportamiento durante el proceso de iteraciones, donde toma diferentes valores para actualizar su trayectoria. En la Figura 4.3 (b) se aprecia el valor medio que obtuvo la cadena durante las iteraciones.



(a) Cadena



(b) Media de MCMC

Figura 4.3: MCMC del parámetro $[\beta - (\mu + \gamma)]$, SIS caso 1.

Fuente: Elaboración propia.

En la Figura 4.4, se muestran los resultados de las estimaciones en el parámetro β . En la cadena se observa que el proceso fue seleccionando diferentes valores para converger en el valor real del modelo SIS. En la Figura b, se muestra el valor medio de la cadena en cada iteración del proceso, donde los primeros valores de la media son valores extremos que se utilizan para actualizar la cadena y lograr la convergencia al valor real de los parámetros.

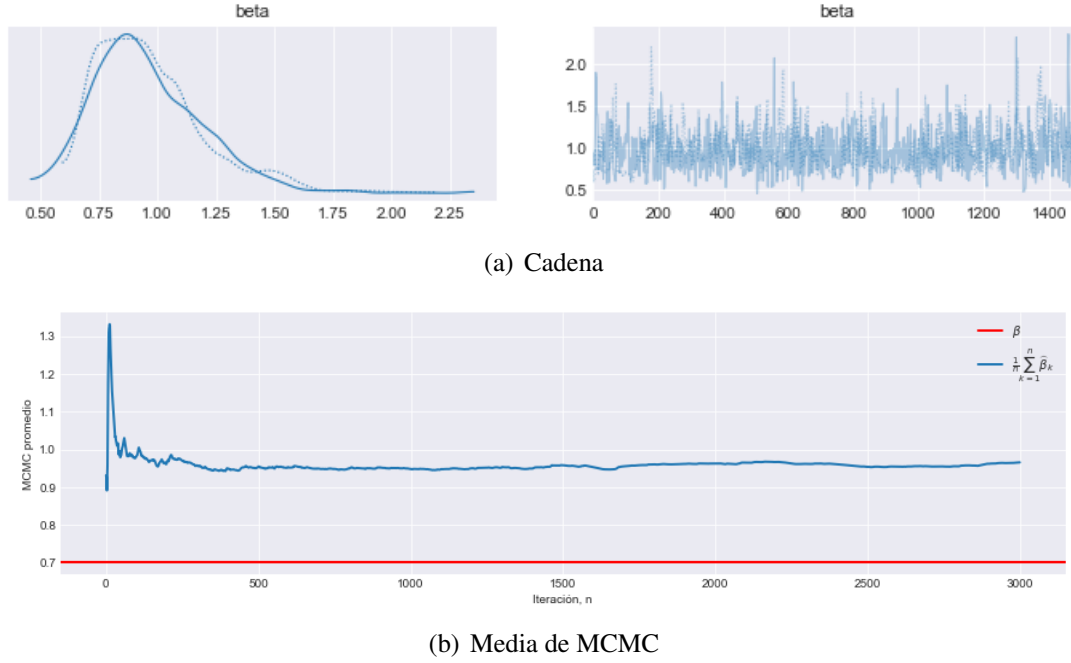


Figura 4.4: MCMC parámetro β , SIS caso 1.

Fuente: Elaboración propia.

En cuanto al comportamiento de los parámetros en la Figura 4.5, se observa que las distribuciones seleccionadas apriori tienen un buen comportamiento con la cadena de Markov, en base a que las posteriores contienen al valor real dentro del intervalo de confianza.

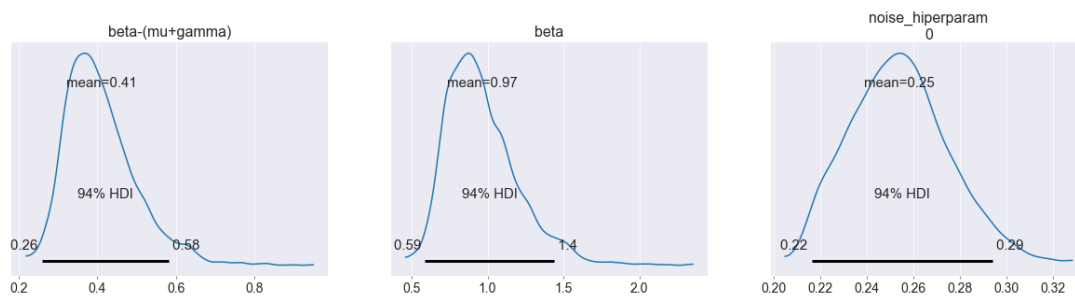


Figura 4.5: posteriori del modelo, SIS caso 1.

Fuente: Elaboración propia.

Los parámetros obtenidos en el Cuadro 4.4 hacen notar que la estimación por regresión de mínimos cuadrados es la más certera al momento de estimar los valores reales. Además, el error cuadrático medio es bajo con ambas metodologías, lo que nos dice que los errores que se están cometiendo para estimar los parámetros son insignificantes.

Vector de parámetro	Valor real	Estimación por Regresión	Estimación por MCMC
θ	[0.3 , 0.7]	[0.371 , 0.832]	[0.409 , 0.965]
Error cuadrático medio, Observado		0.011887	0.012128
Error cuadrático medio, Real		0.000599	0.000252

Cuadro 4.4: Parámetros obtenidos, SIS caso 1.

Fuente: Elaboración propia.

Si bien la regresión obtuvo parámetros más cercanos al valor real, la estimación por cadenas de Markov se ajusta mejor a largo plazo como lo observamos en la Figura 4.6. El ajuste con cadenas de Markov se sobrepone a la curva del valor original a largo plazo, mientras que los ajustes obtenidos con regresión no se sobrepone en ningún momento al valor original. Esto es constante con el error cuadrático medio respecto a los valores reales (sin ruido observacional) obtenidos en el Cuadro 4.4.

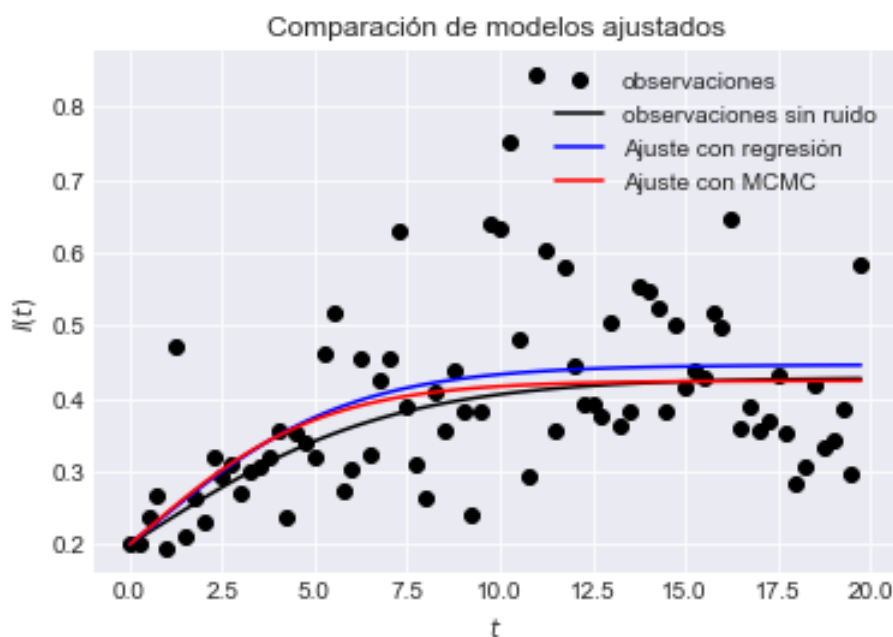


Figura 4.6: comparación de modelos, SIS caso 1.

Fuente: Elaboración propia.

Caso 2: extinción

Para iniciar se genera un conjunto de datos que cumple con la condición de extinción de la enfermedad, se utilizarán los mismos pasos para generar un conjunto de datos con ruido observacional y lograr obtener las estimaciones de los parámetros seleccionados que se muestran en el Cuadro 4.5.

Parámetro	Valor
β	0.7
γ	0.4
μ	0.65
θ	[-0.35 , 0.7]
I_0	0.2
ε	0.25
N	80

Cuadro 4.5: Parámetros del modelo SIS, caso 2.

Fuente: Elaboración propia.

Al momento de realizar las estimaciones por cadenas de Markov se utilizará:

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Normal (0,1)	1500	1500	3000
β	Cauchy (1,1)	1500	1500	3000

Cuadro 4.6: Parámetros de la cadena, SIS caso 2.

Fuente: Elaboración propia.

Las cadenas de Markov generadas con los parámetros del Cuadro 4.6 se comportaron de buena manera; no se presentaron divergencias al momento de simular los estados de la cadena. Los resultados de los diferentes procesos de estimación los encontramos en el Cuadro 4.7, donde podemos observar que en el primer parámetro ambos procesos de calibración se encuentran cerca del valor real. La mayor diferencia la encontraremos en el parámetro β donde la estimación más cercana al valor real se obtiene de la estimación bayesiana. Pero aún así esta muy alejada del valor real.

Vector de parámetros	Valor real	Estimación por regresión	Estimación MCMC
θ	[-0.35 , 0.7]	[-0.3885 , -0.0509]	[-0.3600 , 0.4595]
Error cuadrático medio, Observado		0.00015	0.00016
Error cuadrático medio, Real		0.0000168	0.0000017

Cuadro 4.7: Estimaciones obtenidas, SIS caso 2.

Fuente: Elaboración propia.

En la Figura 4.7, encontramos los dos ajustes del modelo una vez estimados para estimar los parámetros, donde se puede apreciar que el ajuste resultante de estimación por cadenas de Markov Monte Carlo presenta un mayor ajuste a la curva real.

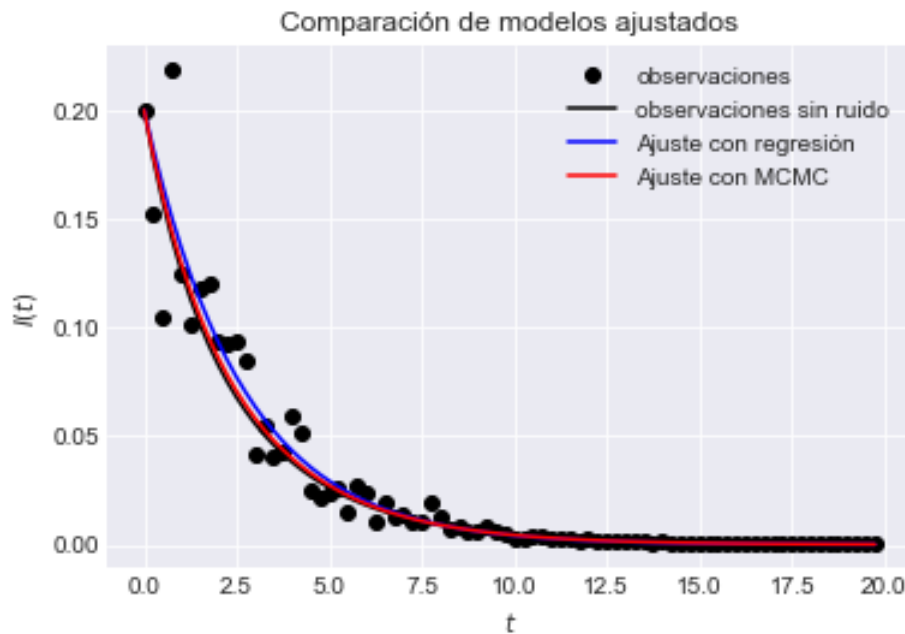


Figura 4.7: Comparación.
Fuente: Elaboración propia.

4.2.2. Validación del proceso de calibración para el modelo estocástico

Para validar el proceso de calibración, la estimación de los parámetros de la ecuación diferencial estocásticas, se generará un conjunto de datos aleatorios con los parámetros del modelo epidemiológico.

Caso 1: persistencia

Para iniciar con el proceso de calibración se genera un conjunto de datos que cumple con la condición de persistencia de la enfermedad en el tiempo (caso 1), donde los parámetros reales se presentan en el Cuadro 4.8, con $R_0 = 1.3103 > 1$.

Parámetro	Valor
β	0.8
γ	0.1
μ	0.24
σ	0.842
θ	[0.46 , 0.8 , 0.842]
R_0	1.3103
I_0	0.2
N	80

Cuadro 4.8: Parámetros reales, SISS caso 1.
Fuente: Elaboración propia.

En la Figura 4.8 se observa una trayectoria generada considerando los parámetros del Cuadro 4.8. Este conjunto de datos se obtuvo por el método de Euler Maruyama.

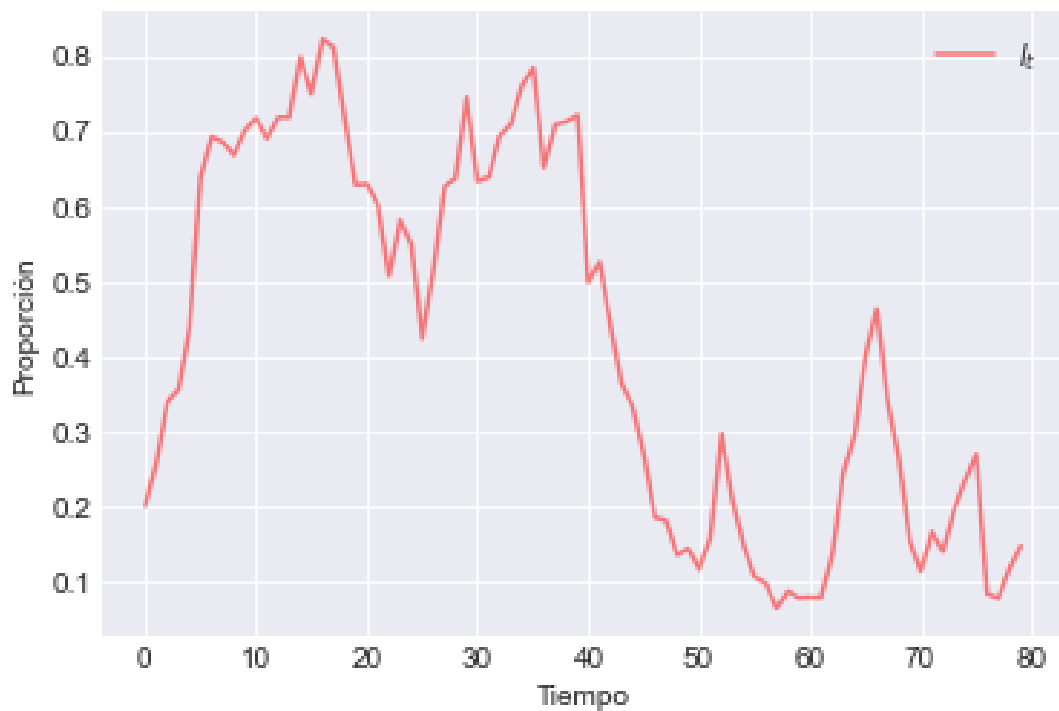


Figura 4.8: Datos simulados, SISS caso 1.

Fuente: Elaboración propia.

Con el fin de estimar los parámetros reales se crearán dos cadenas Markov Monte Carlo, las cuales presentan las siguientes características:

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Normal (0,1)	15000	5000	20000
β	Weibull (5,1)	15000	5000	20000
σ	HalfNormal (1)	15000	5000	20000

Cuadro 4.9: Parámetros de la cadena, SISS caso 1.

Fuente: Elaboración propia.

El comportamiento del primer parámetro en la Figura 4.9. En la Figura (a) encontramos la distribución posteriori al lado izquierdo y la cadena de Markov Monte Carlo en el lado derecho. En la cadena se observa el comportamiento durante las iteraciones, esta toma diferentes valores para actualizar su trayectoria y converger al valor real. En la Figura (b), la línea azul representa el promedio de la cadena en el tiempo y la línea roja es el valor real del parámetro estimado.

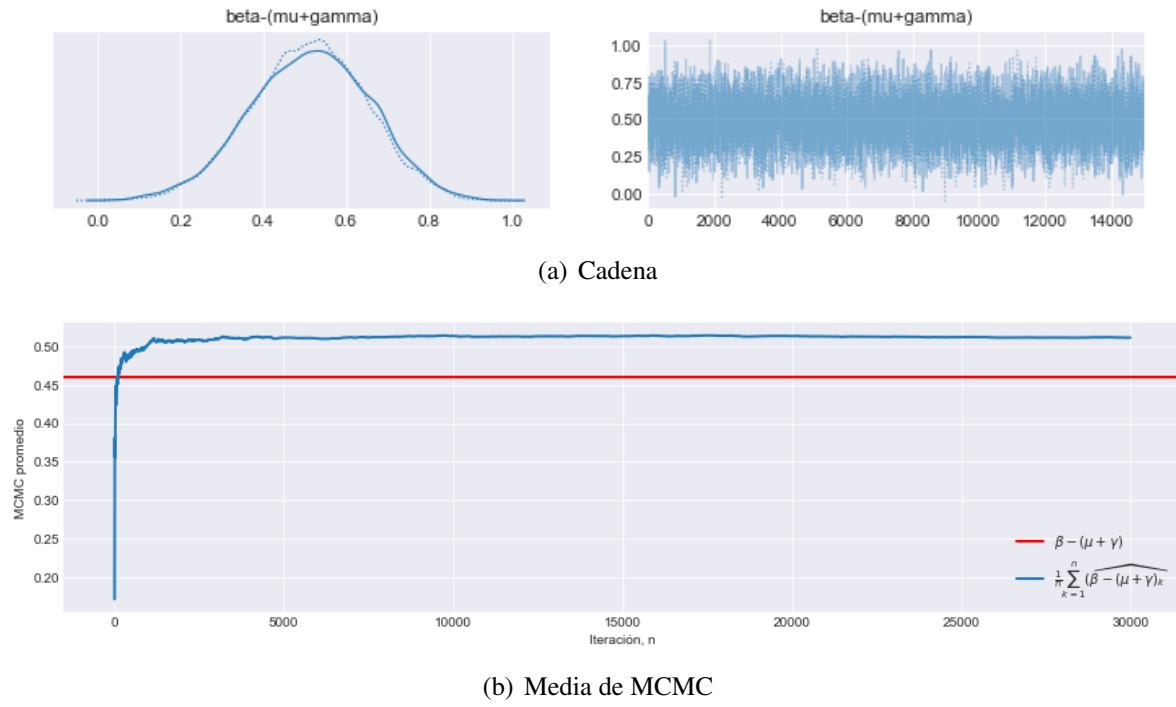


Figura 4.9: MCMC parámetro $[\beta - (\mu + \gamma)]$, SISS caso 1.
Fuente: Elaboración propia.

La estimación del parámetro β (ver Figura 4.10), no presentó complicaciones al momento de explorar valores, debido a la apriori que se seleccionó. Esta le permitió al parámetro explorar diferentes valores para lograr la convergencia. Los valores explorados se concentraron entre 0.5 y 1.5

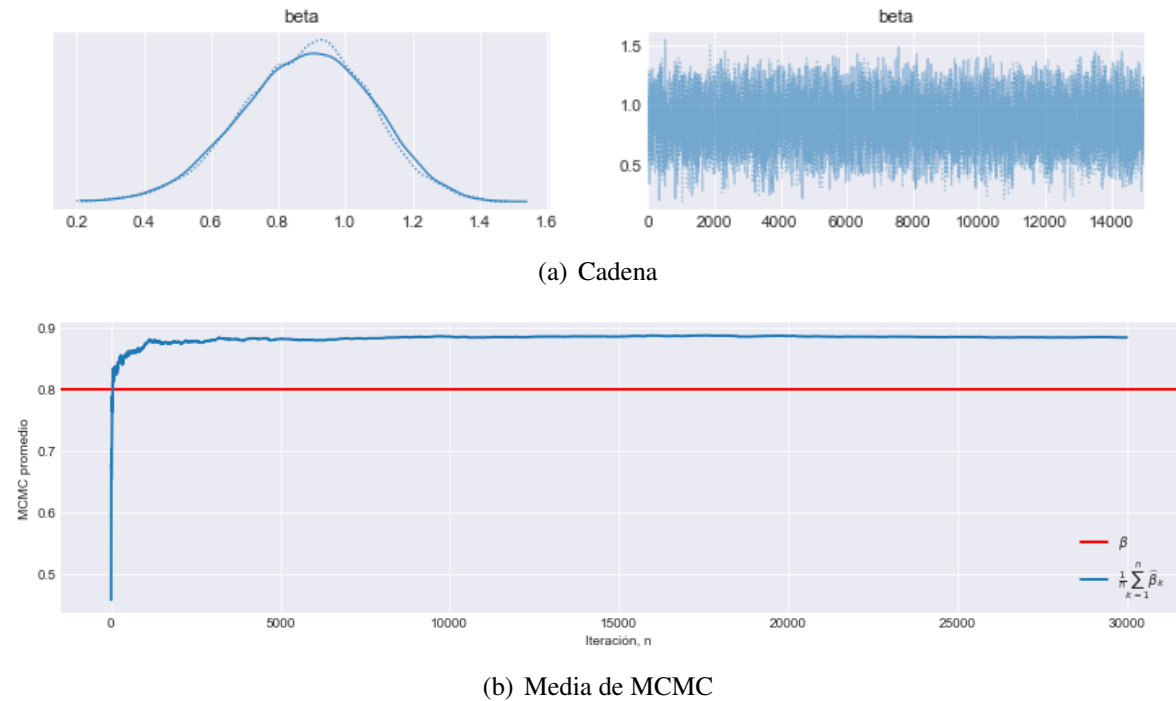


Figura 4.10: MCMC parámetro β , SISS caso 1.
Fuente: Elaboración propia.

Al igual que las otras estimaciones, el parámetro σ no presentó complicaciones para converger al valor real, debido a que se inicializó la cadena con la estimación de la variación cuadrática que ayudó al proceso a converger.

Parámetros	Valor real	Estimación MCMC
$[\beta - (\mu + \gamma)]$	0.46	0.511
β	0.8	0.884
σ	0.842	0.796
R_0	1.310	1.518

Cuadro 4.10: Estimaciones obtenidas, SISS caso 1.

Fuente: Elaboración propia.

El Cuadro 4.10 contiene los parámetros con los valores reales y las estimaciones realizadas por las cadenas de Markov Monte Carlo. Los resultados estimados son cercanos al valor real, lo cual se debe a que la selección de la apriori para cada parámetro fue lo suficientemente informativa para que el proceso converja al valor real.

Caso 2: extinción

Para iniciar con el proceso de calibración se genera un conjunto de datos que cumple con la condición de extinción de la enfermedad en el tiempo (caso 2), donde los parámetros originales son los del Cuadro 4.11.

Parámetro	Valor
β	0.7
γ	0.4
μ	0.65
σ	0.842
θ	$[-0.35, 0.7, 0.842]$
R_0	0.329
I_0	0.2
N	80

Cuadro 4.11: Parámetros originales, SISS caso 2.

Fuente: Elaboración propia.

En la Figura 4.11 se observa una trayectoria generada por los valores del Cuadro 4.11 que rápidamente se extingue. Al igual que en la validación anterior este conjunto de datos se obtuvo por el método de Euler Maruyama.

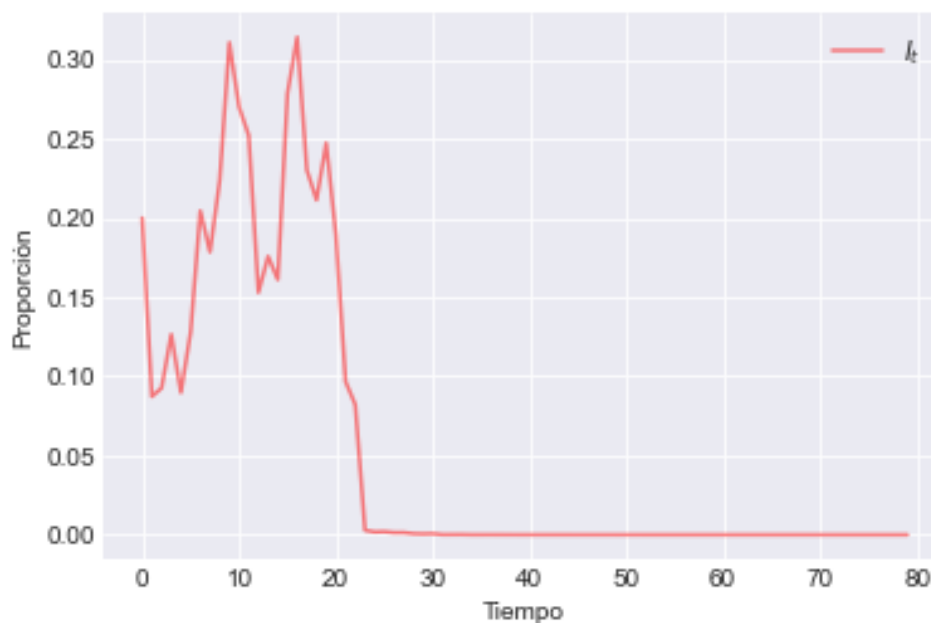


Figura 4.11: Datos simulados, SISS caso 2.

Fuente: Elaboración propia.

Con el fin de estimar los parámetros reales se crearán dos cadenas Markov Monte Carlo, estas presentan las siguientes características:

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Cauchy (-3,1)	15000	5000	20000
β	Gamma (2,2)	15000	5000	20000
σ	HalfNormal (1)	15000	5000	20000

Cuadro 4.12: Parámetros de la cadena, SISS caso 2.

Fuente: Elaboración propia.

El comportamiento de las cadenas de Markov Monte Carlo no presentaron problemas de convergencia debido a las apriori seleccionadas; estas convergieron y no presentaron divergencias.

En la Figura 4.12 podemos observar las distribuciones a posteriori que se generaron.

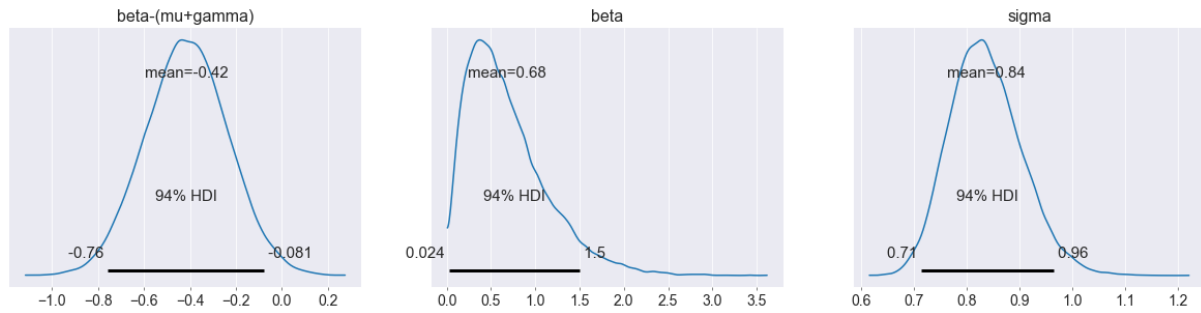


Figura 4.12: Posteriori, SISS caso 2.

Fuente: Elaboración propia.

Los resultado de las estimaciones las encontramos en el Cuadro 4.13. Notamos que las cadenas de Markov Monte Carlo convergieron a los valores reales, con una leve diferencia al valor real.

Parámetros	Valor real	Estimación MCMC
$[\beta - (\mu + \gamma)]$	-0.35	-0.415
β	0.7	0.683
σ	0.842	0.835
R_0	0.329	0.304

Cuadro 4.13: Estimaciones obtenidas, SISS caso 2.

Fuente: Elaboración propia.

Capítulo 5

Resultados

Luego de haber realizado las validaciones en los casos deterministas y estocásticos procedemos a aplicar los procesos de estimación de parámetros a los casos activos del Covid-19 en Chile, con el fin de obtener los diferentes parámetros de los modelos que ayuden en la toma de decisiones respecto de la pandemia.

5.1. Estimación de parámetros para el primer período

El primer período de tiempo seleccionado es de Julio - Octubre del 2020. En este lapso de tiempo Chile se encontraba en los inicios de la pandemia. La medida que contempló el gobierno es el Estado de Excepción Constitucional de Catástrofe que permitió restringir las libertades de los habitantes de todo el país, aplicando medidas como toques de queda, cordones sanitarios, restringir las libertades de locomoción y reunión.



Figura 5.1: Primer período.
Fuente: Elaboración propia.

5.1.1. Estimación determinista

Luego de realizar pruebas con diferentes distribuciones apriori para estimar el primer período de estudio, se seleccionaron las apriori que no generaron divergencias en las cadenas (ver Cuadro 5.1).

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Normal (0,1)	3000	1000	4000
β	Cauchy (1,2)	3000	1000	4000

Cuadro 5.1: Parámetros de la cadena, SIS primer período.

Fuente: Elaboración propia.

Los resultados de las diferentes estimaciones las encontraremos en el Cuadro 5.2. El primer parámetro nos dice que representa la tasa de regeneración ($\beta - (\gamma + \mu)$), la cual dice que la enfermedad persistirá en el tiempo debido a que es un número mayor a cero. El segundo parámetro explica la tasa de transmisión de la enfermedad (β), el valor de la tasa de contagio es demasiado alta y esto tiene sentido por que las medidas preventivas de Chile recién empiezan a tener efecto.

Vector de parámetros	Estimación por MCMC
θ	[0.0164 , 27.7986]
Error cuadrático medio	4.9230e-09

Cuadro 5.2: Parámetros obtenidos, SIS primer período.

Fuente: Elaboración propia.

En la siguiente Figura, nos encontramos con el ajuste de la estimación de cadenas de Markov Monte Carlo, donde notamos que el comportamiento de la curva se ajusta a las observaciones de casos activos en el primer período de tiempo.

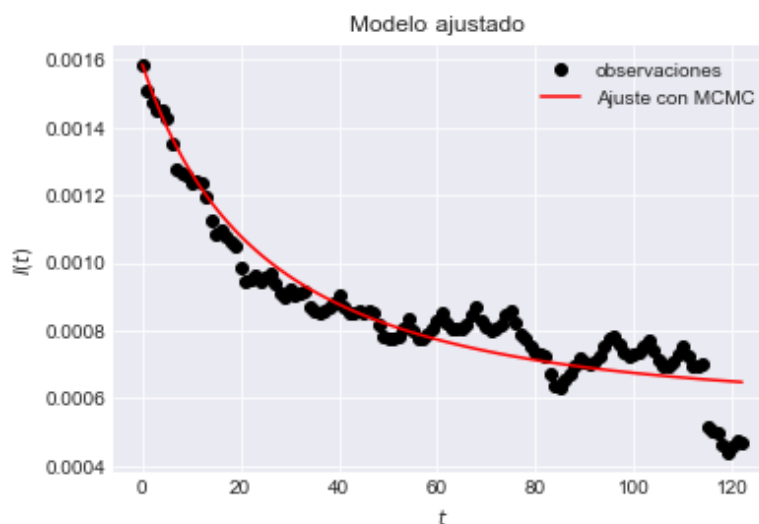


Figura 5.2: Modelo ajustado, SIS primer período.

Fuente: Elaboración propia.

5.1.2. Estimación estocástica

Utilizando la información previamente obtenida en el apartado de validación estocástica, se realizan las estimaciones con la información del Cuadro 5.3, donde la variación del proceso estocástico es el único parámetro con una estimación previa para inicializar la cadena de Markov y encontrar una estimación más cercana al valor real.

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Normal (0,1)	15000	5000	20000
β	Gamma (7.5,1)	15000	5000	20000
σ	HalfNormal (1)	15000	5000	20000

Cuadro 5.3: Parámetros de la cadena, SISS primer período.

Fuente: Elaboración propia.

Las distribuciones a priori seleccionadas no presentaron problemas de convergencias, los parámetros que se estimaron los encontramos en el Cuadro 5.4.

Vector de parámetros	Estimación por MCMC
θ	[0.006, 7.486, 0.018]
Error cuadrático medio estimado	1.820e-07
R_0	1.0007

Cuadro 5.4: Parámetros obtenidos, SISS primer período.

Fuente: Elaboración propia.

En la siguiente Figura nos encontraremos con los datos observados de contagios, la estimación generada del modelo SIS determinista y tres trayectorias generadas por el modelo SISS. Este gráfico recrea el comportamiento de los casos activos de Covid-19 en Chile durante los meses de Julio a Octubre 2020.

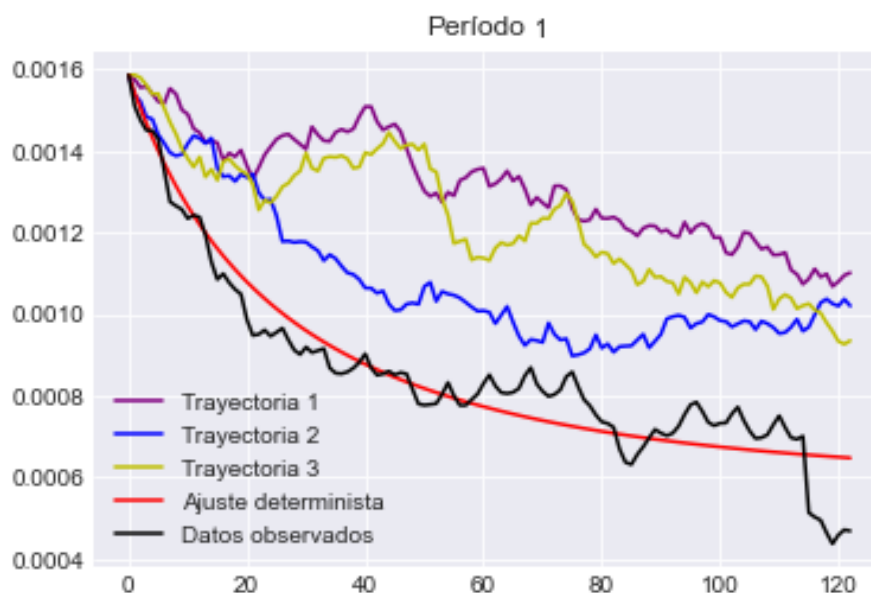


Figura 5.3: Estimaciones del primer período.

Fuente: Elaboración propia.

5.2. Estimación de parámetros para el segundo período

En el segundo período de tiempo de Noviembre 2020 - Febrero del 2021, en Chile se tomaban medidas restrictivas a toda las personas provenientes del extranjero que venían de países que registren transmisión del virus; tendrán que realizar una cuarentena obligatoria al momento de ingresar al país. Se anuncia la apertura de frontera en Aeropuerto Internacional de Santiago y se recibe el primer cargamento de vacunas contra el Covid-19 que se aplica al personal de salud.

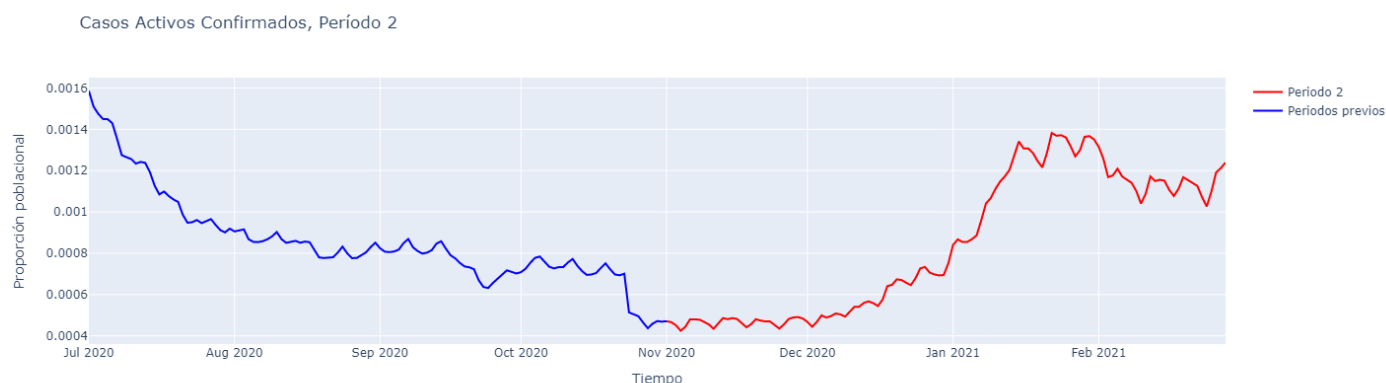


Figura 5.4: Segundo período.

Fuente: Elaboración propia.

En enero se realizaron cambios al Plan Paso a Paso (plan de gobierno para controlar la pandemia). Se disminuyen los aforos y permisos generales que tenían las personas para salir de su residencia. Además, el Presidente Piñera presenta oficialmente el Plan Nacional de Vacunación contra el Covid-19. Para el mes de febrero se inicia el proceso de vacunación masiva con personas mayores de 90 años, trabajadores de recintos educacionales, personas mayores de 65 años, dializadas y trasplantadas.

5.2.1. Estimación determinista

Luego de realizar pruebas con diferentes distribuciones apriori para estimar el primer período de estudio, se seleccionaron las apriori que no generaron divergencias en las cadenas, (ver Cuadro 5.5).

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Normal (0,1)	3000	1000	4000
β	Beta (2,5)	3000	1000	4000

Cuadro 5.5: Parámetros de la cadena, SIS segundo período.

Fuente: Elaboración propia.

La Figura 5.5, contiene las distribución posteriori y la cadena de Markov con las 3000 iteraciones. En las cadenas no se presentan divergencias.

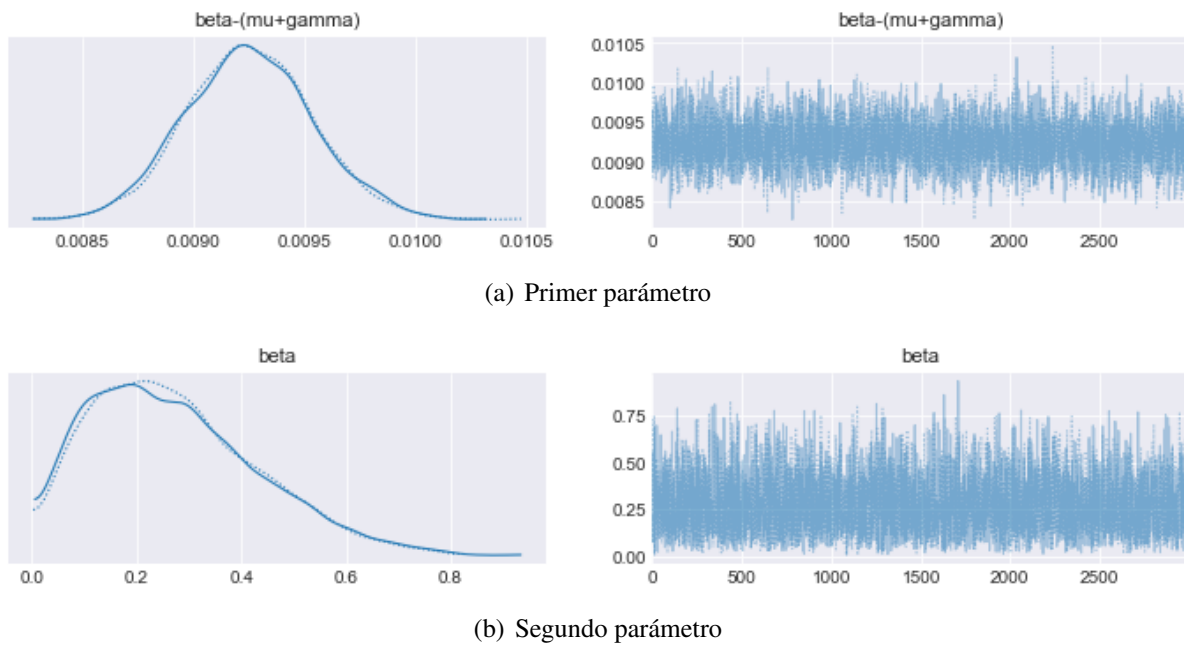


Figura 5.5: Posteriori y cadena de Markov, segundo período.

Fuente: Elaboración propia.

Los resultados de las estimaciones las encontraremos en el Cuadro 5.6. El primer parámetro, nos dice que la enfermedad persistirá en el tiempo debido a que su estimación es positiva; el segundo parámetro explica la tasa de transmisión de la enfermedad. Note que $\hat{\beta}$ disminuye en comparación al primer período, a pesar que la curva de contagios crece.

Vector de parámetros	Estimación por MCMC
θ	[0.0092 , 0.2739]
Error cuadrático medio	2.9335e-08

Cuadro 5.6: Parámetros obtenidos, SIS segundo período.

Fuente: Elaboración propia.

5.2.2. Estimación estocástica

Utilizando la información previamente obtenida en el apartado de validación estocástica, se realizan las estimaciones con la información del Cuadro 5.7, donde la variación del proceso estocástico es el único parámetro con una estimación previa para inicializar la cadena de Markov y encontrar una mejor estimación.

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Normal (0,1)	15000	5000	20000
β	TruncatedNormal (9,10,0)	15000	5000	20000
σ	HalfNormal (1)	15000	5000	20000

Cuadro 5.7: Parámetros de la cadena, SISS segundo período.

Fuente: Elaboración propia.

Note que los valores estimados para β y σ , ambos crecen en comparación al período anterior, lo que podría ajustarse más al hecho de estar en medio de la segunda ola de contagios, que fue más fuerte, aunque menos fatal.

Las distribuciones a priori seleccionadas no presentaron problemas de convergencias. Las estimaciones de los parámetros del modelo los encontramos en el Cuadro 5.8.

Vector de parámetros	Estimación por MCMC
θ	[0.0196 , 12.3633 , 0.0750]
Error cuadrático medio estimado	1.7908e-07
R_0	1.0013

Cuadro 5.8: Parámetros obtenidos, SISS segundo período.

Fuente: Elaboración propia.

En la siguiente Figura nos encontraremos con los datos observados de contagios, la estimación generada del modelo SIS determinista y tres trayectorias generadas por el modelo SISS. Este gráfico recrea el comportamiento de los casos activos de Covid-19 en Chile durante Noviembre 2020 - Febrero 2021.

La trayectoria que recrea el comportamiento de los datos observados es la número tres, esta se comporta de manera similar a la curva de contagios en Chile.

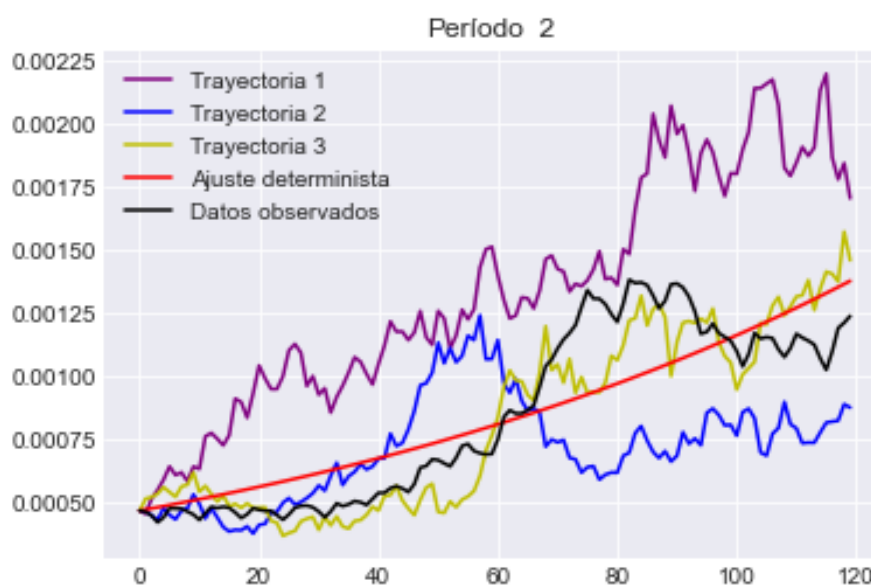


Figura 5.6: Estimaciones del segundo período.

Fuente: Elaboración propia.

5.3. Estimación de parámetros para el tercer período

El tercer período de estudio es de Marzo 2021 - Junio 2021. En marzo Chile superó los 6.5 millones de personas vacunadas con la primera dosis y más de 3.5 millones de personas inoculadas con la segunda dosis.

Entre los sucesos importantes en abril nos encontramos con que el Ministerio de Salud planea trabajo en conjunto con los países del Cono Sur para enfrentar la pandemia; llegan cargamentos con diferentes vacunas que se distribuirán en las diferentes zonas del país y la cantidad de vacunados con la primera dosis asciende a más de 8 millones de personas.

En mayo el Ministerio de Salud informa que la vacuna Sinovac tiene 90,3 % de efectividad para prevenir el ingreso a UCI (Unidad de Cuidados Intensivos). Comienza la inoculación a personas menores de 30 años; además, empieza a regir el Pase de Movilidad a las personas que se ha vacunado, este Pase de Movilidad permite mayor libertad de movimiento a las personas que se han administrado las vacunas contra el Covid-19.

Para finalizar, en el mes de junio el Ministerio de Salud informa resultados del “Análisis de efectividad de vacunas administradas en Chile al completar el esquema de vacunación”. Sinovac y Pfizer-BioNTech son un 90 % y 98 % efectivas, respectivamente, para prevenir ingreso a UCI ([de Salud, 2022](#)). Por otro lado el Ministerio de Salud confirmó el primer caso de la variante Delta.

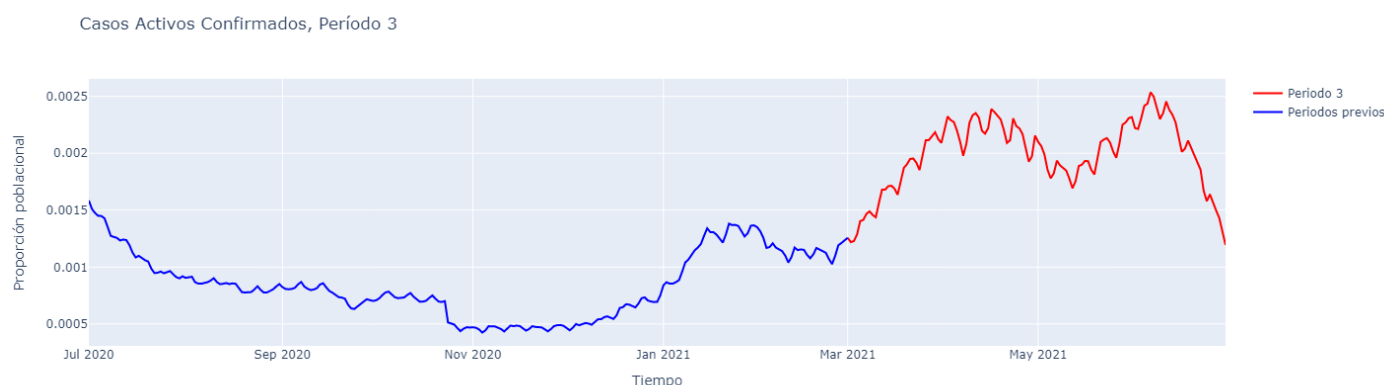


Figura 5.7: Tercer período.
Fuente: Elaboración propia.

5.3.1. Estimación determinista

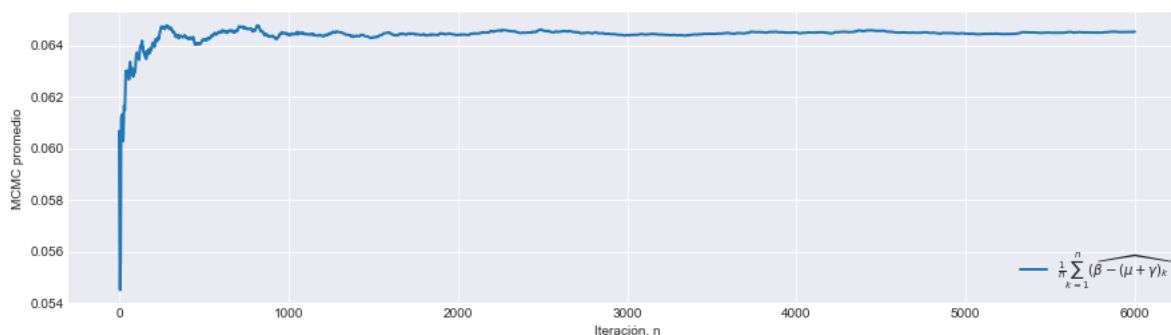
Luego de realizar pruebas con diferentes distribuciones apriori para estimar el tercer período de estudio, se seleccionaron las apriori que generaron menos de un 10 % de divergencias en las cadenas, ver Cuadro 5.9.

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Normal (0,1)	3000	1000	4000
β	Gamma (3,1)	3000	1000	4000

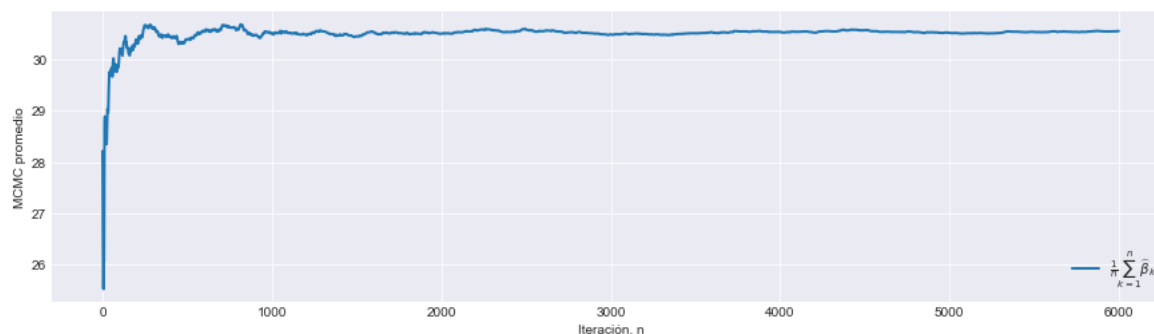
Cuadro 5.9: Parámetros de la cadena, SIS tercer período.

Fuente: Elaboración propia.

La Figura 5.8, contiene las convergencia de la cadena de Markov con las 3000 iteraciones. En esta se aprecia el proceso de convergencia de mil iteraciones. Al ver que la cadena convergen en las primeras mil iteraciones se podría acotar la cantidad asi reducir el costo computacional y el tiempo empleado para encontrar la estimación del parámetro.



(a) Primer parámetro



(b) Segundo parámetro

Figura 5.8: Convergencia de las cadenas, SIS tercer período.

Fuente: Elaboración propia.

Los resultados de las estimaciones las encontraremos en el Cuadro 5.10, el primer parámetro nos dice que la enfermedad persistirá en el tiempo debido a que es un número mayor a cero. El segundo parámetro explica la tasa de transmisión de la enfermedad.

Vector de parámetros	Estimación por MCMC
θ	[0.06452 , 30.56278]
Error cuadrático medio	6.40288e-08

Cuadro 5.10: Parámetros obtenidos, SIS tercer período.

Fuente: Elaboración propia.

En el parámetro β observamos un incremento significativo en el valor de estimación. Debemos notar que el período en los dos primeros meses aumentó la cantidad de casos activos; luego de esta alza el período empezó con una disminución de personas contagiadas. Estos cambios se ven reflejados por el esquema de vacunación en Chile, en donde han vacunado a más de 8 millones de personas.

5.3.2. Estimación estocástica

Utilizando la información previamente obtenida en el apartado de validación estocástica, se realizan las estimaciones con la información del Cuadro 5.11, donde la variación del proceso estocástico es el único parámetro con una estimación previa para inicializar la cadena de Markov y encontrar una estimación más cercana al valor real.

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Normal (0,1)	15000	5000	20000
β	Gamma(3,1)	15000	5000	20000
σ	HalfNormal (1)	15000	5000	20000

Cuadro 5.11: Parámetros de la cadena, SISS tercer período.

Fuente: Elaboración propia.

Las distribuciones a priori seleccionadas no presentaron problemas de convergencias. Los parámetros que se estimaron los encontramos en el Cuadro 5.12

Vector de parámetros	Estimación por MCMC
θ	[0.0033 , 3.0277 , 0.0675]
Error cuadrático medio estimado	1.0414e-06
R_0	1.0003

Cuadro 5.12: Parámetros obtenidos, SISS tercer período.

Fuente: Elaboración propia.

En la siguiente Figura nos encontraremos con los datos observados de contagios, la estimación generada del modelo SIS determinista y tres trayectorias generadas por el modelo SISS. Este gráfico recrea el comportamiento de los casos activos de Covid-19 en Chile durante los meses de Marzo a Junio 2021.

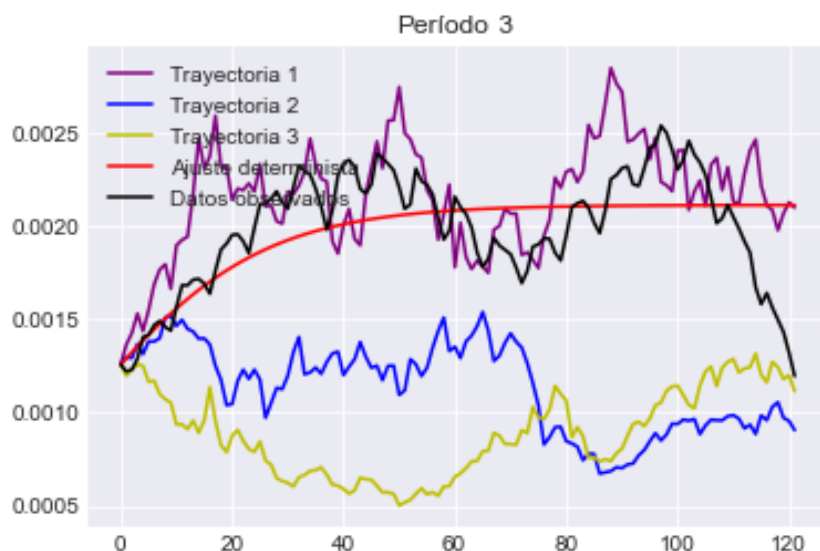


Figura 5.9: Estimaciones del tercer período.

Fuente: Elaboración propia.

La estimación que recrea el comportamiento de contagios en Chile, es la primera trayectoria, esta tiene un compartimiento muy similar a las observaciones emitidas por el Ministerio de Salud.

5.4. Estimación de parámetros para el cuarto período

En los meses de julio a octubre del año 2021, sucedieron diversos acontecimientos el país. Entre los eventos relevantes nos encontramos con el fin del Estado de Excepción Constitucional el 30 de septiembre. Esto permitió a las personas moverse libremente en todo Chile; además, el Ministerio de Salud empieza el proceso de vacunación a la población menor de 17 años y la inoculación a las personas extranjeras. También se anuncian un nuevo proceso de vacunación donde ahora será necesario contar con tres vacunas para tener el pase de movilidad.

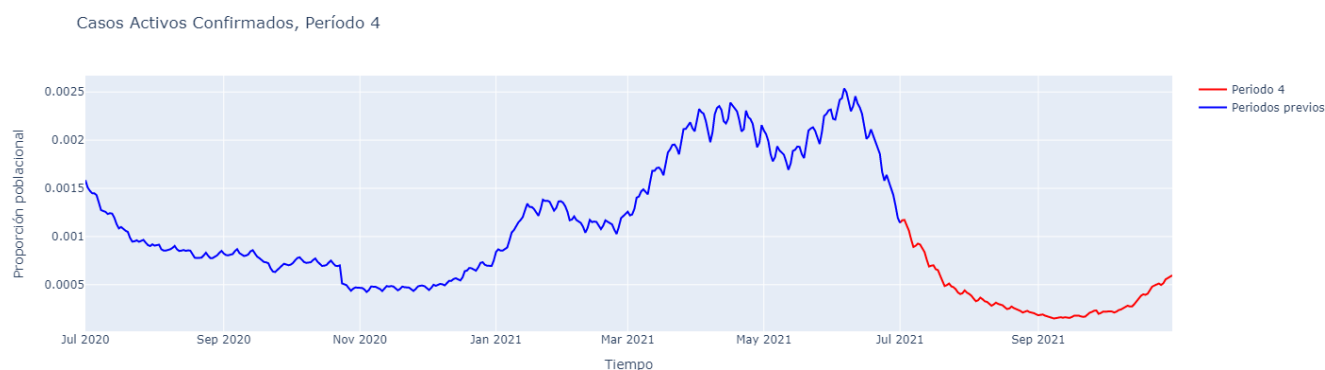


Figura 5.10: Cuarto período.

Fuente: Elaboración propia.

5.4.1. Estimación determinista

Las distribuciones apriori para estimar el cuarto período de tiempo las encontramos en el Cuadro 5.13.

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Normal (0,1)	3000	1000	4000
β	HalfStudentT (1,30)	3000	1000	4000

Cuadro 5.13: Parámetros de la cadena, SIS cuarto período.

Fuente: Elaboración propia.

Los resultados de las estimaciones las encontraremos en el Cuadro 5.14. El primer parámetro nos dice que la enfermedad persistirá en el tiempo debido a que es un número mayor a cero en ambos procesos; el segundo parámetro explica la tasa de transmisión de la enfermedad, la cual es alta en ambas estimaciones. Esto se le puede atribuir a las diferentes acciones que ha realizado el gobierno flexibilizar las medidas tomadas para controlar la pandemia.

Vector de parámetros	Estimación por MCMC
θ	[0.0277 , 116.3588]
Error cuadrático medio	1.7661e-08

Cuadro 5.14: Parámetros obtenidos, SIS Cuarto período.

Fuente: Elaboración propia.

El ajuste de la estimación obtenida por Cadenas de Markov Monte Carlo podemos ver en la siguiente Figura. Los procesos se ajustan de buena manera a los datos observados, como se pudo apreciar en la validación del modelo determinista. El proceso que tuvo un mejor ajuste fue la estimación con cadenas de Markov debido a que se ajusta a la solución real a largo plazo.

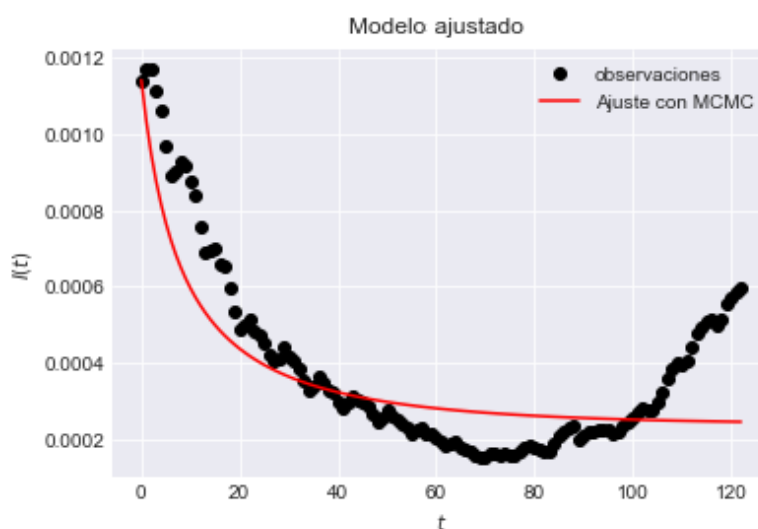


Figura 5.11: Comparación de modelos ajustados.

Fuente: Elaboración propia.

5.4.2. Estimación estocástica

Utilizando la información previamente obtenida en el apartado de validación estocástica, se realizan las estimaciones con la información del Cuadro 5.15, donde la variación del proceso estocástico es el único parámetro con una estimación previa para inicializar la cadena de Markov y encontrar una estimación más cercana al valor real.

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Normal (0,1)	15000	5000	20000
β	ChiSquared (6)	15000	5000	20000
σ	HalfNormal (1)	15000	5000	20000

Cuadro 5.15: Parámetros de la cadena, SISS cuarto período.

Fuente: Elaboración propia.

Las distribuciones a priori seleccionadas no presentaron problemas de convergencias. Los parámetros que se estimaron los encontramos en el Cuadro 5.16.

Vector de parámetros	Estimación por MCMC
θ	[0.0029 , 5.9961 , 0.0215]
Error cuadrático medio estimado	3.6574e-07
R_0	1.0004

Cuadro 5.16: Parámetros obtenidos, SISS cuarto período.

Fuente: Elaboración propia.

En la siguiente Figura nos encontraremos con los datos observados de contagios, la estimación generada del modelo SIS determinista y tres trayectorias generadas por el modelo SISS. Este gráfico recrea el comportamiento de los casos activos de Covid-19 en Chile durante los meses de Julio a Octubre 2021.



Figura 5.12: Estimaciones del cuarto período.

Fuente: Elaboración propia.

5.5. Estimación de parámetros para el quinto período

En los meses de noviembre 2021 a febrero 2022, nos encontraremos con diferentes acontecimientos que destacaron; entre estos, nos encontramos con que Chile se convierte en el primer país de la OCDE (Organización para la Cooperación y el Desarrollo Económicos) en superar las 200 dosis administradas por cada cien habitantes, donde se han administrado más de 40 millones de dosis de vacunas. Esto se ha logrado ya que el ISP (Instituto de Salud Pública) ha autorizado la vacunación de distintos grupos etarios donde la edad mas baja es de tres años.

El 4 de diciembre el Ministerio de Salud confirma primer caso de variante Ómicron y se confirmó su presencia en la Región de Valparaíso, correspondiente a un viajero proveniente de África.

Las autoridades de salud confirman que más de 14 millones de personas han completado esquema de vacunación y anuncian un nuevo proceso de vacunación donde se administrará una cuarta dosis para combatir la pandemia en el mes de febrero.



Figura 5.13: Quinto período.
Fuente: Elaboración propia.

5.5.1. Estimación determinista

Las distribuciones apriori para estimar el quinto período de estudio las encontramos en el Cuadro 5.17.

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Normal (0,1)	3000	1000	4000
β	Weibull (5,1)	3000	1000	4000

Cuadro 5.17: Parámetros de la cadena, SIS quinto período.
Fuente: Elaboración propia.

El resultado de la estimación la encontraremos en el Cuadro 5.18. El primer parámetro dice que la enfermedad persistirá en el tiempo debido a que es un número mayor a cero en ambos procesos. El segundo parámetro explica la tasa de transmisión de la enfermedad en el cual se

observa que el parámetro obtenido en la regresión es un valor negativo, es decir, la enfermedad no se estaría transmitiendo.

Vector de parámetros	Estimación por MCMC
θ	[0.0176, 0.8017]
Error cuadrático medio	2.8825-06

Cuadro 5.18: Parámetros obtenidos, SIS quinto período.

Fuente: Elaboración propia.

5.5.2. Estimación estocástica

Utilizando la información previamente obtenida en el apartado de validación estocástica, se realizan las estimaciones con la información del Cuadro 5.19, donde la variación del proceso estocástico σ no se estimará con cadenas de Markov Monte Carlo, debido a que este período presenta problemas al inicializar el proceso. Es por esto que se utilizara la estimación planteada por Miao (2004).

Parámetros	Distribución a priori	Tamaño de la cadena de Markov	Iteraciones de prueba	Total
$[\beta - (\mu + \gamma)]$	Normal (0,1)	15000	5000	20000
β	TruncatedNormal (5,10,0)	15000	5000	20000
σ	-	-	-	-

Cuadro 5.19: Parámetros de la cadena, SISS quinto período.

Fuente: Elaboración propia.

En la Figura 5.14 se presenta la distribución a posteriori que generó el proceso de estimación. También cuenta con la cadena de estimación donde se puede observar que no presenta problemas de exploración. No existen espacios en blanco que demuestren divergencias en la cadena.

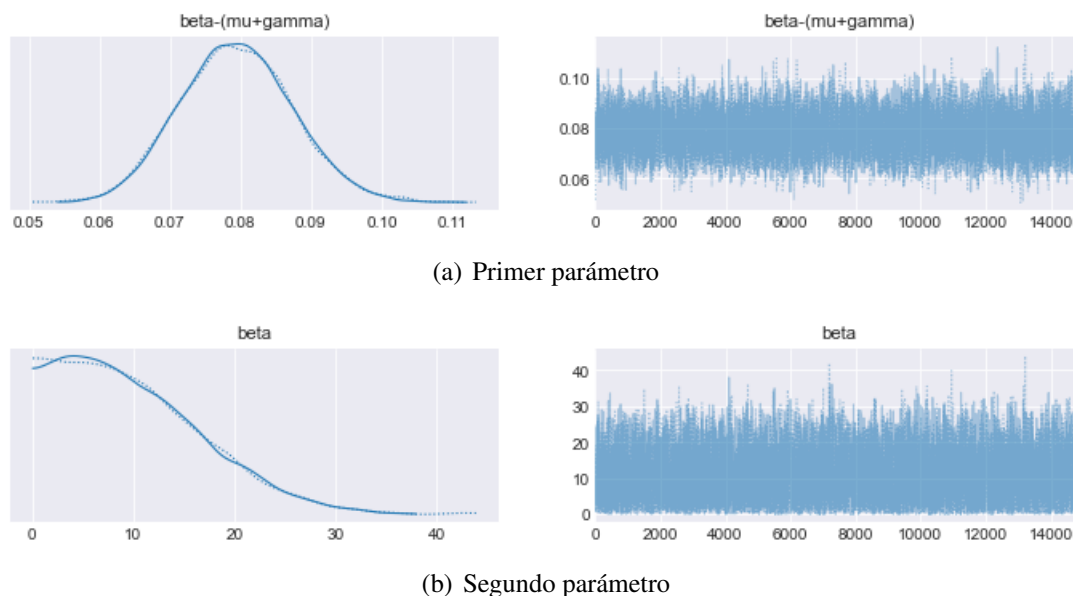


Figura 5.14: Distribución posteriori y cadena de Markov, SISS quinto período.

Fuente: Elaboración propia.

Las estimaciones generadas por las cadenas de Markov Monte Carlo son las siguientes:

Vector de parámetros	Estimación por MCMC
θ	[0.0794 , 10.1577 , 0.0716]
Error cuadrático medio estimado	1.5341e-05
R_0	1.0076

Cuadro 5.20: Parámetros obtenidos, SISS quinto período.

Fuente: Elaboración propia.

En la siguiente Figura nos encontraremos con los datos observados de contagios, la estimación generada del modelo SIS determinista y tres trayectorias generadas por el modelo SISS. Este gráfico recrea el comportamiento de los casos activos de Covid-19 en Chile durante los meses de Noviembre 2021 a Febrero 2022.

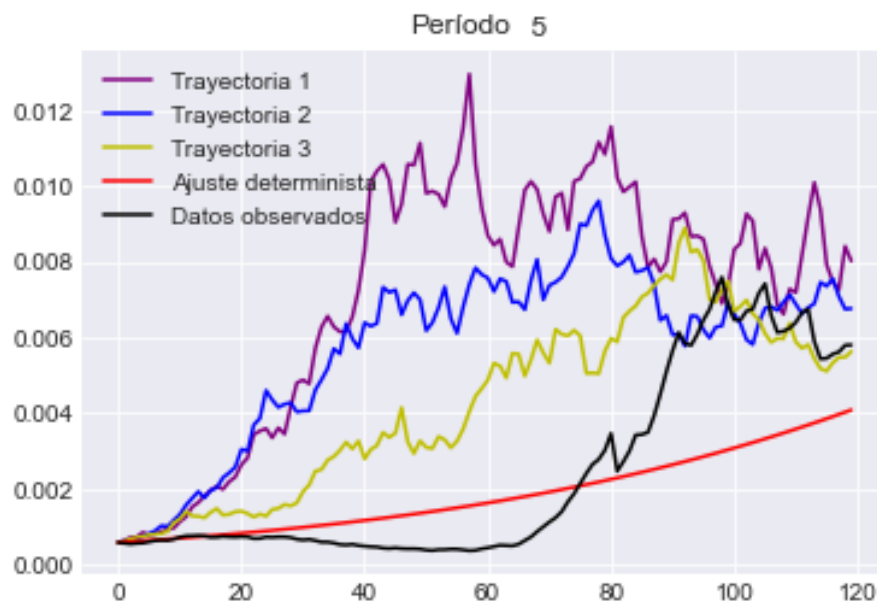


Figura 5.15: Trayectoria del quinto período.

Fuente: Elaboración propia.

La trayectoria que asemeja los valores observados por el Ministerio de Salud en Chile es la tercera. El comportamiento es similar, esta crece en un período determinado para luego bajar, podemos apreciar del día 100 al 120, la trayectoria se sobrepone al valor observado.

5.6. Comparación de estimaciones

Dado que la validación de la estimación de β puede tomar valores negativos en la estimación por regresión, dado a que esta representa una tasa de contagio donde valores deben ser mayores o igual a cero, se dejó fuera de la calibración de los modelos que presentan los datos reales del COVID-19 en Chile, en los períodos de Julio 2020 - Febrero 2022.

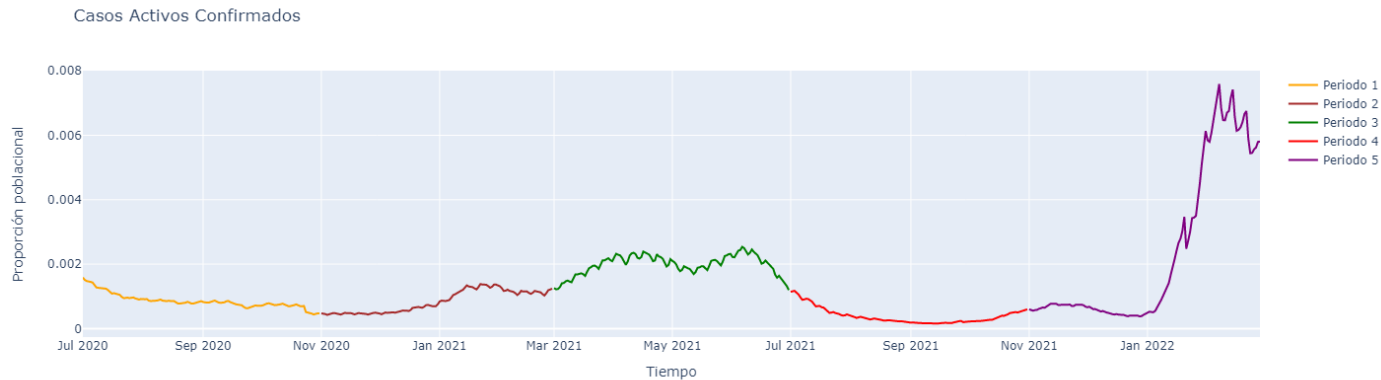


Figura 5.16: Periodos de tiempo.
Fuente: Elaboración propia.

Analizando los resultados del Cuadro 5.21, nos encontraremos que el error cuadrático medio en el período determinista tiene un error muy bajo a comparación del caso estocástico. Esto nos dice que ambos modelos se están ajustando bien. Mirando más de cerca, el comportamiento de las estimaciones de los parámetros del modelo estocástico son más consistentes con las observaciones y los eventos en cada período de tiempo.

	Determinista		Estocástico	
	MCMC	Error	MCMC	Error
	$[\beta - (\mu + \gamma), \beta]$		$[\beta - (\mu + \gamma), \beta, \sigma]$	
Primer período	[0.0164 , 27.7986]	4.9230e-09	[0.0060 , 7.4864 , 0.0184]	1.8201e-07
Segundo período	[0.0092 , 0.2739]	2.9335e-08	[0.0196 , 12.3633 , 0.0750]	1.7908e-07
Tercer período	[0.0645 , 30.5627]	6.4028e-08	[0.0033 , 3.0277 , 0.0675]	1.0414e-06
Cuarto período	[0.0277 , 116.3588]	1.7661e-08	[0.0029 , 5.9961 , 0.0215]	3.6574e-07
Quinto período	[0.0176 , 0.8017]	2.8825e-06	[0.0794 , 10.1577 , 0.0716]	1.5341e-05

Cuadro 5.21: Resultado de las estimaciones.
Fuente: Elaboración propia.

En el primer periodo nos encontramos con una tasa de transmisión es 7.4864, donde el gobierno tomo medidas drásticas para tomar el control de la pandemia con el Estado de Excepción Constitucional de Catástrofe, esta medida no fue la más eficiente para controlar, pero ayudo a mantener bajo la tasa de contagios con una desviación estándar del periodo de tiempo es de 0.0184 y la tasa de renovación del proceso es de 0.0060

Segundo periodo se observa un aumento de en tasa de contagios donde el valor estimado es el más alto con 12.3633 con una variación estimada de 0.0750. Entre las acciones que tomó el gobierno nos encontramos con la apertura de la frontera del Aeropuerto Internacional de Santiago en el cual aplicaron la política de que las personas provenientes de países que presenten el virus deberán realizar cuarentena obligatoria. Además, disminuyeron los permisos para salir de sus hogares y los aforos de estos con el fin de evitar la propagación del virus. Con la aplicación de estas medidas se espera una disminución de contagiados.

Para el tercer periodo de estudio nos encontramos con una tasa de contagiados estimada de 3.0277 y una desviación estimada de 0.0675. Con estas estimaciones y con el inicio del esquema de vacunación se espera que en los próximos meses la tasa de contagios disminuya, pero esto se ve afectado por la llegada de la variante delta.

Tras la llegada de la variante delta y con un proceso de vacunación masiva, Chile en el cuarto periodo presenta una tasa de transmisión estimada de 5.9961 y una desviación estimada de 0.0215. Debido a la baja de casos activos se pone fin al Estado de Excepción Constitucional de catástrofe. Esto permitió a las personas retomar las libertades de movimiento, pero aun mantienen el paso de movilidad y aforos restringidos para evitar eventos masivos que trasmitan la enfermedad.

El Quinto período nos encontramos con más de 14 millones de personas que han completado su esquema de vacunación (contemplando dosis de refuerzo). La tasa de contagios estimada es 10.1577 con una variación estimada de 0.0716. Uno esperaría que los valores fueran bajos por las diferentes acciones tomas para controlar la pandemia, pero el verdadero motivo de tener una tasa de contagio alta es por la llegada de la variante Omicron, una de las mutaciones del virus más contagiosa había arribado a Chile en el mes de diciembre. Esto hizo que la tasa de contagio aumentara abruptamente en un breve período de tiempo.

En el Cuadro 5.22, nos encontraremos que los diferentes modelos de estimación predicen que la enfermedad perdurara en el tiempo.

	Determinista	Estocástico
Primer período	Persistencia	Persistencia
Segundo período	Persistencia	Persistencia
Tercer período	Persistencia	Persistencia
Cuarto período	Persistencia	Persistencia
Quinto período	Persistencia	Persistencia

Cuadro 5.22: Estado de la enfermedad.

Fuente: Elaboración propia.

Las vacunas han sido increíblemente efectivas para reducir las enfermedades graves que genera y la muerte, pero no son perfectas. Las nuevas variantes han puesto a prueba las defensas de la vacunación; después de unos meses, se nota el efecto positivo con la reducción de los contagiados, complicaciones médicas graves y la muerte.

Capítulo 6

Conclusiones

A lo largo de este trabajo se ha revisado la base teórica de la epidemiología matemática, para la modelación de una enfermedad. Entre estas nos encontramos con los modelos deterministas y modelos estocásticos a los que se le han realizado los algoritmos correspondientes para estimar los parámetros de cada caso. Para verificar las estimaciones se realizaron simulaciones para ambos casos existentes, cuando la enfermedad persiste en el tiempo y cuando esta desaparece.

Al momento de aplicar los procesos el costo computacional de estimación se ve considerablemente afectado cuando se esta en presencia del modelo determinista. Los tiempos de ejecución son en promedio por cadena de 40 minutos y este proceso se realiza dos veces con 4000 iteraciones. Este es un proceso que se debe repetir varias veces para calibrar las distribuciones apriori; por el contrario, el modelo estocástico es más rápido, tarda 1 minuto por cadena y se generan 20000 iteraciones lo cual es 5 veces más que el proceso determinista. Al ser un proceso extremadamente rápido a comparación del modelo determinista es mucho más sencillo de implementar.

El aporte de esta tesis la podemos resumir en los siguientes puntos:

1. Se han implementado en el software Python procesos de validación y estimación del modelo SIS y SIS Estocástico.
2. Se validaron los modelos epidemiológicos a través de simulaciones creadas en Python.
3. Se aplicaron los modelos epidemiológicos a los datos del virus SARS-CoV-2 en Chile.

Si bien los diferentes períodos de estudio tuvieron resultados distintos en las estimaciones del modelo determinista y estocástico. El modelo que presento las mejores estimaciones fue el SISS que es un proceso estocástico, por lo que no se espera un ajuste preciso sino un buen ajuste del comportamiento medio. En ambos modelos de estimación, los resultados para cada periodo hacen inferencia de que la enfermedad persiste en los distintos períodos de tiempo.

Cabe señalar que el modelo epidemiológico SIS es un modelo compartimental. Este modelo contempla que la población se puede re-infectar y volver a ser susceptible, y los individuos no generarán una inmunidad a la enfermedad. Si se desea tener mayores antecedentes para explicar el comportamiento del virus será necesario agregar más compartimientos que expliquen la enfermedad. Al agregar compartimientos estaremos aumentando el grado de dificultad y es posible que no tenga una solución analítica.

Apéndice A

Código

Validación Determinista

```
import arviz as az
import matplotlib.pyplot as plt
import numpy as np
import pymc3 as pm
import theano
from sklearn.metrics import mean_squared_error
from pymc3.ode import DifferentialEquation
from scipy.integrate import odeint
from random import seed
import scipy.optimize as opt
plt.style.use("seaborn-darkgrid")

# From now on we'll work with proportions instead of integer numbers
# i -> infectados/N
# Parameters:
# beta = disease transmission rate
# gamma = recovery rate
# mu = mortality rate
# mu + beta = renewal rate
# theta = [beta - (mu+gamma), beta]
# Thus, if theta[0]<0, the disease will vanish and if theta[0]>0 we have persistence of the
# disease

i0 = 0.2 # initial proportion of infected
Tf = 20 # Terminal time
T0 = 0 # Initial time
Dt = 0.25 # time step
beta, gamma, mu = 0.7, 0.4, 0.65
theta = np.asarray([beta - (mu+gamma), beta])
theta_t = theta
times = np.arange(T0, Tf, Dt) # time vector

# ODE model
def SIS(i, t, theta):
    di = theta[0]* i[0] - theta[1]* i[0]*i[0]
    return di

sis_model = DifferentialEquation(
    func=SIS,
    times = np.arange(T0, Tf, Dt),
    n_states=1,
    n_theta=2,
    t0=0,
)

# ODE solution (analytical solution computed before)
def solution_SIS(t,*param):
    aux_exp = np.exp(param[0]*t)
    aux = param[0] + param[1]*(i0)*(aux_exp - 1)
    return i0*param[0]*aux_exp/aux
```

```
# Create true ODE solution
infected_t = odeint(SIS, t=times, y0=i0, args=(theta,), rtol=1e-8)
# Simulate observations (true data+log-normal noise with parameter sigma=0.5 in this case)
infected_obs = np.random.lognormal(mean=np.log(infected_t[1:]), sigma=0.25)
# we add initial condition after because we now the exact initial condition (it doesn't have
↳ noise)
infected_obs = np.concatenate((np.asarray([[i0]]), infected_obs))

# Plot observations and true data
plt.plot(times, infected_obs, marker="o", linestyle="none", color='k', markersize=2,
↳ label=f"datos observados")
plt.plot(times, infected_t, color="C0", alpha=0.5, label=f"datos reales")
plt.xlabel("$t$")
plt.ylabel("$I(t)$")
plt.legend()
plt.show()

# REGRESION MODEL -----
# This suppose to be our benchmark in the case of the deterministic model
theta0 = np.asarray([0.1, 0.1]) #punto de partida para la busqueda del optimo
reg_estim, _ = opt.curve_fit(solution_SIS, xdata = times, p0=theta0, ydata =
↳ infected_obs.flatten(), maxfev=5000)
print("\n Parmetros reales", theta_t)
print("\n Parmetros estimados con regresin", reg_estim)
reg_model = solution_SIS(times, *reg_estim) # curve obtained from regresion estimates

# COMPARE REGRESION MODEL VERSUS TRUE MODEL AND OBSERVATIONS-----
plt.plot(times, infected_obs.flatten(), "ko", label="observaciones")
plt.plot(times, infected_t.flatten(), "k-", label="observaciones sin ruido")
plt.plot(times, reg_model, "b-", label="Ajuste con regresin")
plt.xlabel("$t$")
plt.ylabel("$I(t)$")
plt.title("Modelo ajustado con mnimos cuadrados")
plt.legend(loc="upper right")
plt.show()

# Define pymc model and sampling the MC Markov chain
nb_chains = 2 #choose >1 to run some convergence diagnostics on the chains
len_chains = 1500
len_burning_states = 1500

with pm.Model() as model4:
    # Define prior distributions
    noise_hiperparam = pm.HalfCauchy("noise_hiperparam", 1, shape=1) # noise hiperparameter
    # This prior does not assume a prior knowledge on the persistence or extinction of the
    ↳ disease
    # But if the user knows if the disease persists, the prior can be feed with more info on
    ↳ the support
    param = [0,0]
    param[0] = pm.Normal("beta-(mu+gamma)", 0, 1)
    param[1] = pm.Bound(pm.Cauchy, lower=0) ("beta", 1, 1)
    # define output of the computational model
    sis_curves = sis_model(y0=[i0], theta=param)
    #Define statistical model for observations
    Y = pm.Lognormal("Y", mu=pm.math.log(sis_curves), sigma=noise_hiperparam,
    ↳ observed=infected_obs)
    trace = pm.sample(len_chains, tune=len_burning_states, cores=1, chains=nb_chains)
    data = az.from_pymc3(trace=trace)

# CONVERGENCE DIAGNOSTICS -----
print("\n Parmetros reales", theta_t)
print("\n Resumen de los resultados para la cadena de Markov construida:")

az.summary(trace)

# plot the exploration for each chain and convergence of the ergodic mean
# beta - (mu+gamma)
az.plot_trace({"beta-(mu+gamma)": trace.get_values(varname="beta-(mu+gamma)", combine=False)})
mean_theta0 = [np.mean(trace['beta-(mu+gamma)'][:i]) for i in np.arange(1,
↳ len(trace['beta-(mu+gamma)']))]
```

```
plt.figure(figsize=(15, 4))
plt.axhline([theta_t[0]], lw=2, color="red", label=r"$\beta - (\mu + \gamma)$")
plt.plot(mean_theta0, lw=2, label=r"$\frac{1}{n} \sum_{k=1}^n \widehat{(\beta - (\mu + \gamma))}_k$")
plt.xlabel("Iteracin, n")
plt.legend()
plt.ylabel("MCMC promedio")

# beta
az.plot_trace({"beta": trace.get_values(varname="beta", combine=False)})
mean_theta1 = [np.mean(trace['beta'][i]) for i in np.arange(1, len(trace['beta']))]
plt.figure(figsize=(15, 4))
plt.axhline([theta_t[1]], lw=2, color="red", label=r"$\beta$")
plt.plot(mean_theta1, lw=2, label=r"$\frac{1}{n} \sum_{k=1}^n \widehat{\beta}_k$")
plt.xlabel("Iteracin, n")
plt.legend()
plt.ylabel("MCMC promedio")

# Define bayes model with Bayes estimators
MCMC_model = solution_SIS(times, mean_theta0[-1], mean_theta1[-1])

# Plot posterior densities
az.plot_posterior(data,
                  var_names=["beta-(mu+gamma)", "beta", "noise_hiperparam"]);
plt.show()

# VALIDATION OF PYMC MODEL-----
print("\n Parmetros reales:", theta_t)
print("Parmetros estimados con regresin:", reg_estim)
print("Parmetros estimados con MCMC:", [mean_theta0[-1], mean_theta1[-1]])

# COMPARE MODELS -----
plt.plot(times, infected_obs.flatten(), "ko", label="observaciones")
plt.plot(times, infected_t.flatten(), "k-", label="observaciones sin ruido")
plt.plot(times, reg_model, "b-", label="Ajuste con regresin")
plt.plot(times, MCMC_model, "r-", label="Ajuste con MCMC")
plt.xlabel("$t$")
plt.ylabel("$I(t)$")
plt.title("Comparacin de modelos ajustados")
plt.legend(loc="upper right")
plt.show()

#MSE with respect to observed data
print("\n Error cuadrático medio para regresin:", mean_squared_error(infected_obs, reg_model))
print("Error cuadrático medio para MCMC:", mean_squared_error(infected_obs, MCMC_model))

#MSE with respect to true data
print("\n Error cuadrático medio para regresin:", mean_squared_error(infected_t, reg_model))
print("Error cuadrático medio para MCMC:", mean_squared_error(infected_t, MCMC_model))
```

Validación Estocástica

```
import arviz as az
import matplotlib.pyplot as plt
import numpy as np
import pymc3 as pm
import scipy
import theano
from sklearn.metrics import mean_squared_error
from pymc3.ode import DifferentialEquation
from scipy.integrate import odeint
from random import seed
import scipy.optimize as opt
plt.style.use("seaborn-darkgrid")
from pymc3.distributions.timeseries import EulerMaruyama
import datetime
import pandas as pd
import plotly.express as px
import scipy
px.defaults.height=600
```

```

# From now on we'll work with proportions instead of integer numbers
# i -> infectados/N
# Parameters:
# beta = disease transmission rate
# gamma = recovery rate
# mu = mortality rate
# mu + beta = renewal rate
# theta = [beta - (mu+gamma), beta]
# Thus, if theta[0]<0, the disease will vanish and if theta[0]>0 we have persistence of the
↪ disease

i0 = 0.2 # initial proportion of infected
Tf = 20 # Terminal time
T0 = 0 # Initial time
Dt = 0.25 # time step
beta, gamma, mu, sigma = 0.7, .1, 0.3, 0.842 #add a known sigma to be able to estimate the
↪ true value
theta = np.asarray([beta - (mu+gamma), beta, sigma])
times = np.arange(T0, Tf, Dt) # time vector
theta_t = theta

R_0 = 1 + (theta[0] - 0.5 * theta[2] ** 2) / (theta[1] - theta[0])

if R_0 > 1:
    print('la enfermedad persiste')
else:
    if theta[2] ** 2 <= theta[1]:
        print('la enfermedad se extingue')

if theta[2] ** 2 > theta[1] * max(1, (0.5 * theta[1]) / (theta[1] - theta[0])):
    print('la enfermedad se extingue')

# time series
N = len(times) - 1
x = i0 # initial proportion of infected
x_t = []

# simulate
for i in range(N):
    x = x + (x * (theta[0] - (x * theta[1]))) * Dt + theta[2] * (x - x ** 2) *
    ↪ np.random.normal() * np.sqrt(Dt)
    x_t.append(x)
x_t = np.array(x_t) # convert the values into an array
x_t = np.insert(x_t, 0, i0, axis=0) # add the initial condition at the beginning of the array

# Plotting the Euler-Maruyama process
plt.plot(x_t, "r", label=r"$I_t$", alpha=0.5)
plt.xlabel("Tiempo")
plt.legend()
plt.ylabel("Proporcin")

# ODEs model
def SISS(i, theta):
    di = (i * (theta[0] - theta[1] * i))
    signal = theta[2] * (i - i ** 2)
    return di, signal

sigma_2_ep = sum(np.diff(x_t) ** 2) / (Dt * sum((x_t ** 2) * (1 - x_t) ** 2))
sigma_ep = np.sqrt(sigma_2_ep)

# Define pymc model and sampling the MC Markov chain
nb_chains = 2 #choose >1 to run some convergence diagnostics on the chains
len_chains = 15000
len_burning_states = 5000

with pm.Model() as model:
    theta_m = [0, 0, 0]
    theta_m[0] = pm.Normal("beta - (mu + gamma)", 0, 1)
    theta_m[1] = pm.Weibull("beta", 5, 1)
    theta_m[2] = pm.HalfNormal("sigma", 1)
    eu = EulerMaruyama("EM", Dt, SISS, (theta_m,), shape=x_t.shape, observed=x_t)

```

```

with model:
    trace = pm.sample(len_chains, tune=len_burning_states, cores=1, chains=nb_chains,
        ↪ start={"beta-(mu+gamma)":0.2,"beta":0.3,"sigma":sigma_ep})
    data = az.from_pymc3(trace=trace)

az.summary(trace)

# Plot posterior densities
az.plot_posterior(data, var_names=["beta-(mu+gamma)", "beta","sigma"]);
plt.show()

# plot the exploration for each chain and convergence of the ergodic mean
# beta - (mu+gamma)
az.plot_trace({"beta-(mu+gamma)": trace.get_values(varname="beta-(mu+gamma)", combine=False)})
mean_theta0 = [np.mean(trace['beta-(mu+gamma)'][i]) for i in np.arange(1,
    ↪ len(trace['beta-(mu+gamma)']))]
plt.figure(figsize=(15, 4))
plt.axhline([theta_t[0]], lw=2, color="red", label=r"$\beta-(\mu+\gamma)$")
plt.plot(mean_theta0, lw=2, label=r"$\frac{1}{n}\sum_{k=1}^n\widehat{(\beta-(\mu+\gamma))_k}$")
plt.xlabel("Iteracin, n")
plt.legend()
plt.ylabel("MCMC promedio")

# beta
az.plot_trace({"beta": trace.get_values(varname="beta", combine=False)})
mean_theta1 = [np.mean(trace['beta'][i]) for i in np.arange(1, len(trace['beta']))]
plt.figure(figsize=(15, 4))
plt.axhline([theta_t[1]], lw=2, color="red", label=r"$\beta$")
plt.plot(mean_theta1, lw=2, label=r"$\frac{1}{n}\sum_{k=1}^n\widehat{\beta}_k$")
plt.xlabel("Iteracin, n")
plt.legend()
plt.ylabel("MCMC promedio")

# sigma
az.plot_trace({"sigma": trace.get_values(varname="sigma", combine=False)})
mean_theta2 = [np.mean(trace['sigma'][i]) for i in np.arange(1, len(trace['sigma']))]
plt.figure(figsize=(15, 4))
plt.axhline([theta_t[2]], lw=2, color="red", label=r"$\sigma$")
plt.plot(mean_theta2, lw=2, label=r"$\frac{1}{n}\sum_{k=1}^n\widehat{\sigma}_k$")
plt.xlabel("Iteracin, n")
plt.legend()
plt.ylabel("MCMC promedio")

theta_est = [mean_theta0[-1],mean_theta1[-1],mean_theta2[-1]]

R_0 = 1+(theta_est[0]-0.5*theta_est[2]**2)/(theta_est[1]-theta_est[0])

if R_0 > 1:
    print('la enfermedad persiste')
else:
    if theta_est[2]**2 <= theta_est[1]:
        print('la enfermedad se extingue')

if theta_est[2]**2 > theta_est[1]*max(1, (0.5*theta_est[1])/(theta_est[1]-theta_est[0])):
    print('la enfermedad se extingue')

# generate trace from posterior
ppc_trace = pm.sample_posterior_predictive(trace, model=model)

# plot with data
plt.figure(figsize=(10, 3))
plt.plot(np.percentile(ppc_trace["EM"], [2.5, 97.5], axis=0).T, "k", label=r"$I_{95\%}$")
plt.plot(x_t, "r", label=r"$I_t$")
plt.legend()

def trayectorias(infected_obs,theta_est):
    Dt = 1 # time step
    N = len(infected_obs)-1
    i0 = infected_obs[0]
    x = i0 # initial proportion of infected
    x_t_est = []

```

```

# simulate
for i in range(N):
    x = x + (x*(theta_est[0]-(x*theta_est[1])))*Dt + theta_est[2] * (x-x**2) *
    ↪ np.random.normal()*np.sqrt(Dt)
    x_t_est.append(x)
x_t_est = np.array(x_t_est) # convert the values into an array
x_t_est=np.insert(x_t_est,0,i0,axis=0) # add the initial condition at the beginning of
    ↪ the array
return np.abs(x_t_est)

aa=trayectorias(x_t,theta_est)
msr_aa=mean_squared_error(x_t,aa)

# plot of the trajectories
plt.plot(trayectorias(x_t,theta_est), "r", label=r"Trayectoria $I_t$")
plt.legend()
plt.title("Validacion SISS Mayor, Trayectoria 1 \n error cuadratico medio: "+ str(msr_aa)+"\n
    ↪ muestreo de advis")

```

Referencias

- Bayram, M., Partal, T., y Orucova Buyukoz, G. (2018). Numerical methods for simulation of stochastic differential equations. *Advances in Difference Equations*, 2018(1), 1–10.
- Carmona, A. (2009). *Ecuaciones diferenciales estocásticas* (Tesis Doctoral no publicada). Thesis. Universidad de Buenos Aires, Argentina.
- Cockayne, J., y Duncan, A. (2021). Probabilistic gradients for fast calibration of differential equation models. *SIAM/ASA Journal on Uncertainty Quantification*, 9(4), 1643–1672.
- de Salud, M. (2022). *COVID-19 en Chile Pandemia 2020-2022*. Gobierno de Chile.
- Evans, L. C. (2006). An introduction to stochastic differential equations version 1.2. *Lecture Notes, UC Berkeley*.
- García Piñera, A., y cols. (2014). Modelos de ecuaciones diferenciales para la propagación de enfermedades infecciosas.
- García Rovira, L., y cols. (2017). Modelos matemáticos compartimentales en epidemiología.
- Gil Sáez, B., y cols. (2020). Aplicación de cadenas de markov.
- Gray, A., Greenhalgh, D., Hu, L., Mao, X., y Pan, J. (2011). A stochastic differential equation SIS epidemic model. *SIAM Journal on Applied Mathematics*, 71(3), 876–902.
- Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM review*, 43(3), 525–546.
- Hitchcock, D. B. (2003). A history of the metropolis–hastings algorithm. *The American Statistician*, 57(4), 254–257.
- Hoffman, M. D., Gelman, A., y cols. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1), 1593–1623.
- Kermack, W. O., y McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772), 700–721.
- Lu, Q. (2009). Stability of SIRS system with random perturbations. *Physica A: Statistical Mechanics and Its Applications*, 388(18), 3677–3686.
- Miao, W.-C. (2004). *Quadratic variation estimators for diffusion models in finance*. University of Southern California.
- Neal, R. M. (2003). Slice sampling. *The annals of statistics*, 31(3), 705–767.
- Oksendal, B. (2013). *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media.
- OMS. (2021). Coronavirus. Descargado de https://www.who.int/es/health-topics/coronavirus#tab=tab_1
- Rincón, L. (2006). Introducción a las ecuaciones diferenciales estocásticas. *UNAM. México*.
- Stewart, J. (2012). *Cálculo de varias variables: trascendentes tempranas* (Inf. Téc.). Cengage Learning.
- Tornatore, E., Buccellato, S. M., y Vetro, P. (2005). Stability of a stochastic SIR system. *Physica A: Statistical Mechanics and its Applications*, 354, 111–126.

- Zhao, Y., y Jiang, D. (2014). The threshold of a stochastic SIS epidemic model with vaccination. *Applied Mathematics and Computation*, 243, 718–727.
- Zill, D. (2009). *Ecuaciones diferenciales con aplicaciones de modelado*. Thomson Learning.