

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

Sebastián Andrés Caamaño Contreras

Universidad de Valparaíso

Profesora Guía

Kerlyns Martínez. Ph.D.

Profesora Co-Guía

Alejandra Christen. Ph.D.

16 de Diciembre 2022



**Universidad
de Valparaíso**
CHILE

Facultad de Ciencias
Instituto de Estadística

- ➊ Motivación
 - Hipótesis
 - Objetivo
- ➋ Estado del arte
- ➌ Datos
- ➍ Validación de calibraciones
 - Caso 1: La enfermedad desaparecerá
 - Caso 2: La enfermedad persiste
- ➎ Análisis de períodos
- ➏ Comparación de estimaciones
- ➐ Conclusión
- ➑ Referencias

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, esta 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, esto con el fin de predecir el comportamiento de un virus o enfermedad en la población.

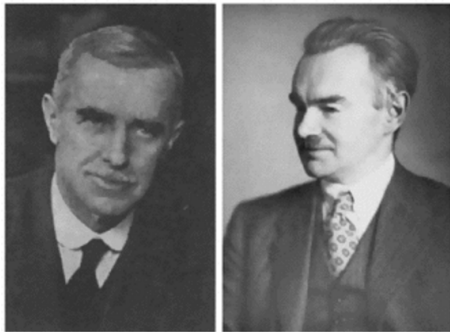


Figura 1: Kermack y McKendrick.

Fuente: (Bacaër, 2011)

El modelo epidemiológico estocástico, incluso en el caso más simple, describirá con mayor precisión la evolución de contagios del SARS-CoV-2 en Chile.

Objetivo general

Comparar modelos epidemiológicos determinista y estocástico, con los parámetros debidamente estimados utilizando los datos del SARS-CoV-2 en Chile.

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 y mutaciones del virus.
- Contrastar el ajuste de los modelos epidemiológicos seleccionados.

Existen dos grandes tipos de modelos epidemiológicos: deterministas y estocásticos (García 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:

$$dX_t = b(X_t)dt, \quad (1)$$

donde la variable de estado es X , que depende del tiempo t y b es una función conocida.

- 2 En un modelo estocástico **no es posible controlar todos los factores**, sino que se pueden calcular la probabilidad de que ocurra cierto evento. Entre los eventos de interés se encuentra **la posibilidad de que la epidemia se extinga**.
Por ejemplo, un modelo estocástico puede ser escrito de la forma:

$$dX_t = b(X_t)dt + B(X_t)dW_t, \quad (2)$$

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.

Modelo determinista SIS

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; γ , tasa de recuperación.

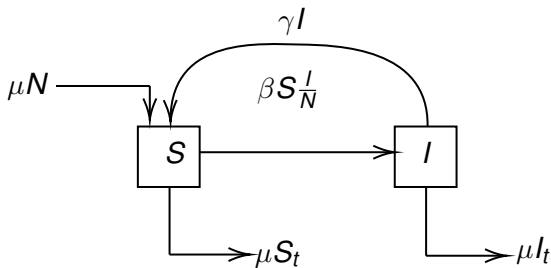


Figura 2: Modelo SIS con nacimientos y muerte

Fuente: Elaboración propia.

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

$$\begin{cases} \frac{dS_t}{dt} = -S_t\beta\frac{I_t}{N} + \gamma I_t - \mu S_t + \mu N, \\ \frac{dI_t}{dt} = S_t\beta\frac{I_t}{N} - \gamma I_t - \mu I_t, \end{cases} \quad \text{con } t \geq 0 \quad (3a)$$

Note que 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:

$$S_t + I_t = N, \quad \text{para todo } t \geq 0 \quad (4)$$

Solución del modelo SIS determinista

A continuación se observa la solución del modelo determinista SIS:

$$\frac{I_t}{N} = \frac{\frac{I_0}{N}(\beta - \gamma - \mu)e^{(\beta - \gamma - \mu)t}}{\beta \frac{I_0}{N}e^{(\beta - \gamma - \mu)t} + \beta \frac{S_0}{N} - (\gamma + \mu)} \quad (5)$$

Con la solución obtenida se evalúa la función al límite en el tiempo para obtener el comportamiento asintótico del modelo epidemiológico SIS determinista.

$$\lim_{t \rightarrow \infty} \frac{I_t}{N} = \begin{cases} 1 - \frac{(\gamma + \mu)}{\beta} & , \text{ si } \beta > \gamma + \mu \\ 0 & , \text{ si } \beta \leq \gamma + \mu \end{cases} \quad (6)$$

En conclusión, 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.

Comportamiento asintótico

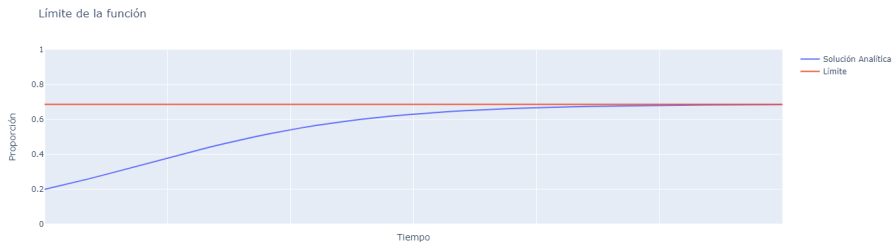


Figura 3: Límite de la función.

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

Fuente: Elaboración propia.

Por otro lado, partiendo de la idea de que la tasa de contagios es aleatoria, el modelo SIS estocástico (Gray, Greenhalgh, Hu, Mao, y Pan, 2011) queda como:

σ , constante de difusión, W_t , proceso de Wiener y los coeficientes μ , β , γ se definen como en el modelo determinista

$$\left\{ \begin{array}{l} \frac{dS_t}{dt} = \left[\mu N - \frac{\beta S_t I_t}{N} + \gamma I_t - \mu S_t \right] dt - \frac{\sigma S_t I_t}{N} dW_t, \end{array} \right. \quad (7a)$$

$$\left\{ \begin{array}{l} \frac{dI_t}{dt} = \left[\frac{\beta S_t I_t}{N} - (\mu + \gamma) I_t \right] dt + \frac{\sigma S_t I_t}{N} dW_t \end{array} \right. \quad (7b)$$

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:

$$S_t + I_t = N, \quad \text{para todo } t \geq 0. \quad (8)$$

La ecuación con las que se hará la estimación será:

$$\frac{dI_t}{Ndt} = \left[\frac{I_t}{N} \left((\beta - \mu - \gamma) - \frac{\beta I_t}{N} \right) \right] dt + \sigma \left(\frac{I_t}{N} - \frac{I_t^2}{N^2} \right) dW_t \quad (9)$$

La base de datos se obtuvo del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación disponibles en GitHub (<https://github.com/MinCiencia/Datos-COVID19>) y corresponde al “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.



Figura 4: Línea de tiempo.

Fuente: Elaboración propia.

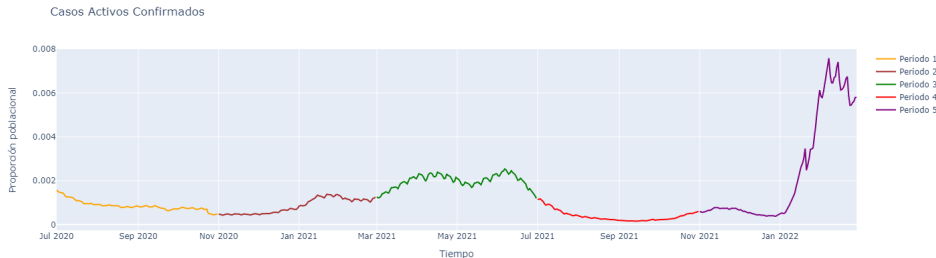


Figura 5: Población de infectados en los diferentes períodos de tiempo.

Fuente: Elaboración propia.

Validación

En el proceso de validación se generarán dos conjunto de datos con parámetros específicos considerando el caso para que la enfermedad persista en el tiempo y el caso en la que se extingue.

Métodos de estimación

- Regresión por mínimos cuadrados
- Método Hamilton Monte Carlo usando muestreo

Método de comparación

- Error cuadrático medio
- Error cuadrático medio estimado

Comportamiento de la enfermedad

- R_0 es un valor que se enfrenta a diferentes procesos para ver si la enfermedad se extingue o perdura en el tiempo

Modelo de calibración determinista

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

$$\mu, \beta, \gamma, \varepsilon \sim \text{Gamma}(\alpha, \lambda), \quad (10)$$

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

$$\pi(\theta | I_t^{obs}) \propto p(\theta) \mathcal{L}(I_t^{obs} | \theta), \quad (11)$$

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 $\pi(\theta | I_t^{obs})$.

Caso 1: Extinción

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 1: Parámetros reales, SIS caso 1.

Fuente: Elaboración propia.

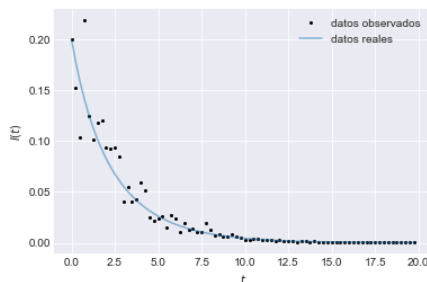


Figura 6: Datos, SIS caso 1.

Fuente: Elaboración propia.

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 2: Parámetros de la cadena, SIS caso 1.

Fuente: Elaboración propia.

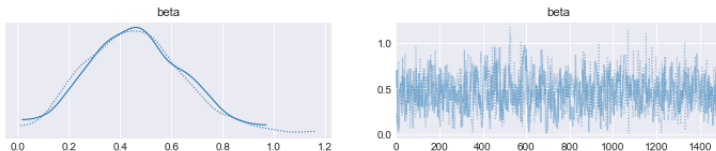


Figura 7: Posteriori y cadenas MCMC para el parámetro β , SIS caso 1.

Fuente: Elaboración propia.

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 3: Estimaciones obtenidas, SIS caso 1.

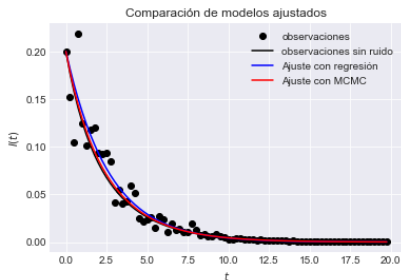


Figura 8: Comparación, SIS caso 1.

Fuente: Elaboración propia.

Modelo de calibración estocástico

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

$$\mu, \beta, \gamma, \sigma \sim \text{Beta}(\alpha, \lambda), \quad (12)$$

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

$$\pi(\theta | I_t^{\text{obs}}) \propto p(\theta) \mathcal{L}(I_t^{\text{obs}} | \theta), \quad (13)$$

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 $\pi(\theta | I_t^{\text{obs}})$.

Caso 2: Persistencia

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: Parámetros reales, SISS caso 2.

Fuente: Elaboración propia.

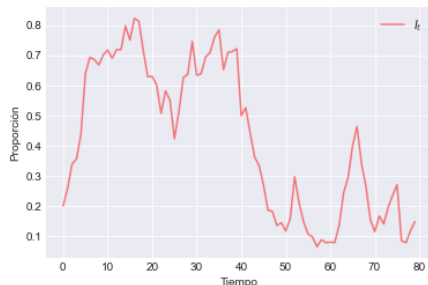


Figura 9: Datos simulados, SISS caso 2.

Fuente: Elaboración propia.

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 5: Parámetros de la cadena, SISS caso 2.

Fuente: Elaboración propia.

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 6: Estimaciones obtenidas, SISS caso 2.

Fuente: Elaboración propia.

Tercer período

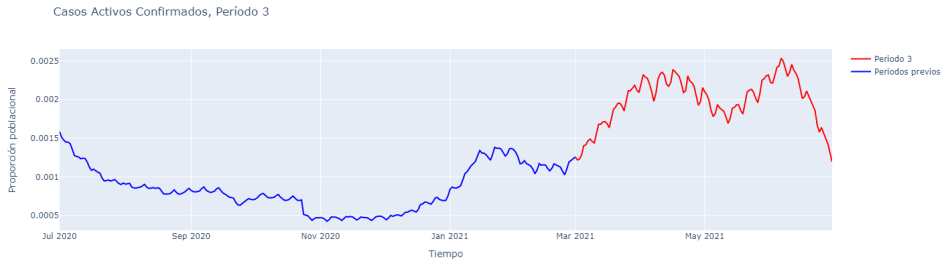


Figura 10: Población de infectados en el tercer período.

Fuente: Elaboración propia.

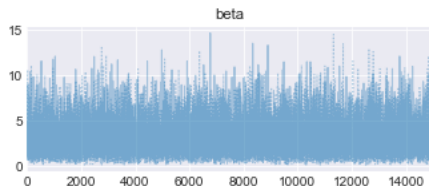
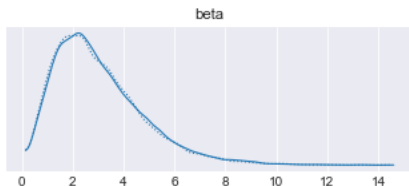


Figura 11: Posteriori y cadena parámetro β , SISS tercer período.

Fuente: Elaboración propia.

Vector de parámetros	Estimación MCMC SIS	Estimación MCMC SISS
θ	[0.06452 , 30.56278]	[0.0033 , 3.0277 , 0.0675]
Error cuadrático medio	6.40288e-08	-
Error cuadrático medio estimado	-	1.0414e-06

Cuadro 7: Estimaciones obtenidas, tercer período.

Fuente: Elaboración propia.

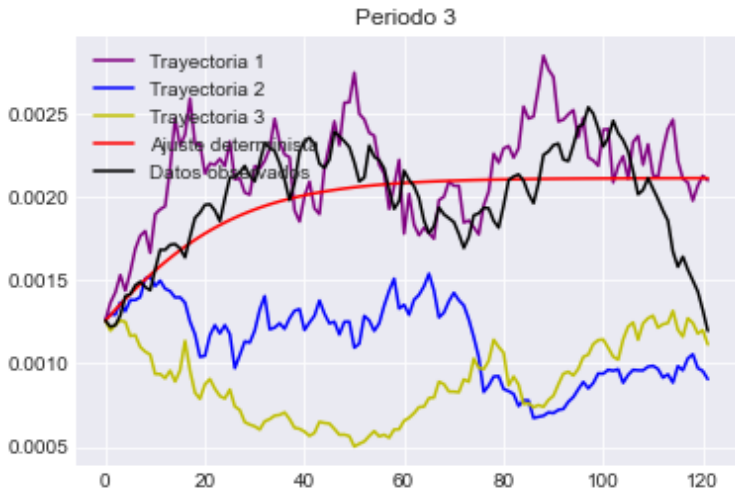


Figura 12: Estimación de infectados en el tercer período.

Fuente: Elaboración propia.

Quinto período



Figura 13: Población de infectados en el quinto período.

Fuente: Elaboración propia.

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 8: Parámetros de la cadena, SIS quinto período.

Fuente: Elaboración propia.

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 9: Parámetros de la cadena, SISS quinto período.

Fuente: Elaboración propia.

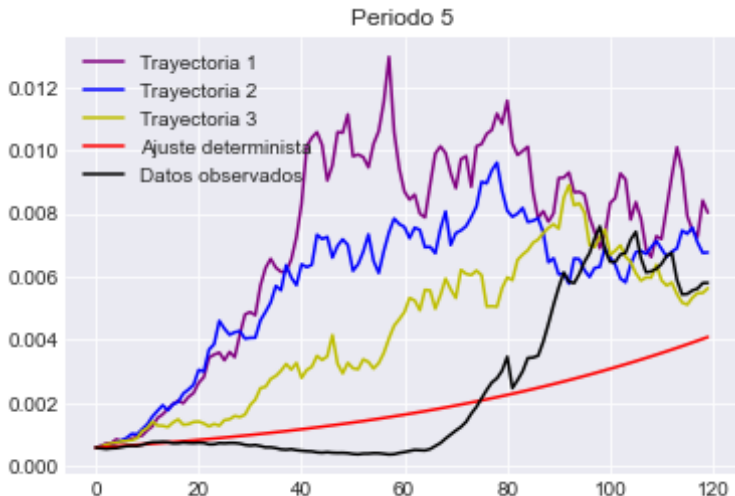
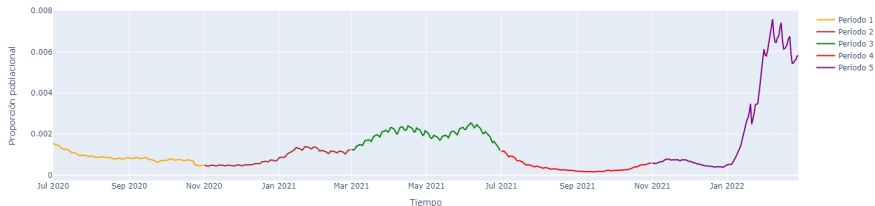


Figura 14: Estimación de infectados en el quinto período.

Fuente: Elaboración propia.

Comparación de estimaciones



	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 10: Resultado de las estimaciones.

Fuente: Elaboración propia.

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

Cuadro 11: Estado de la enfermedad.

Fuente: Elaboración propia.

A lo largo de este trabajo se ha revisado la base teórica de la epidemiología matemática para la modelación de una enfermedad.

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 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 diferentes estimaciones en el modelo determinista y estocástico. El modelo que presentó 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.

- Bacaër, N. (2011). Mckendrick and kermack on epidemic modelling (1926–1927). En *A short history of mathematical population dynamics* (pp. 89–96). Springer.
- Christen, A., Maulén-Yañez, M. A., González-Olivares, E., y Curé, M. (2018). A stochastic disease transmission in an epidemic model considering a hyperbolic incidence rate. *ACTA DE RESUMENES LXXXIII Encuentro Anual Sociedad de Matemática de Chile*, 36.
- de Salud, M. (2022). *COVID-19 en Chile Pandemia 2020-2022*. Gobierno de Chile.
- García, L., y cols. (2017). Modelos matemáticos compartimentales en epidemiología.
- 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.

- 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.