

# CUANTIFICACIÓN DE INCERTIDUMBRE BAYESIANA APROXIMADA EN PROBLEMAS INVERSOS DE ODE

T E S I S

Que para obtener el grado de  
**Maestro en Probabilidad y Estadística**

**Presenta**

César Isaí García Cornejo

**Director de Tesis:**

Dr. José Andrés Christen Gracia

---

Autorización de la versión final



*A mi madre **Rosa Cornejo** quien me apoya emocionalmente y me incentivo a conseguir mis metas.*

*A mi padre **Julio García** quien con su esfuerzo logro darme la oportunidad de perseguir la profesionalización.*

*A mi novia **Zaira Martínez** quien me dio amor y apoyo durante la elaboración de la tesis así como en la maestría.*



# Agradecimientos

A mis padres que me apoyaron incondicionalmente confiando en mí dándome el impulso que necesitaba para concluir la maestría antes y durante el periodo de la misma. Agradecer a mi asesor Andrés Christen por su valioso apoyo como asesor, quien además se mostró disponible y generoso. Agradecer a mi novia Zaira Martínez por su apoyo, comprensión y disciplina que nos sacó adelante día con día. A mis compañeros y ahora amigos cuyo apoyo mutuo fue central en nuestro crecimiento denotando sus fuertes valores como el compañerismo, empatía y amistad.



# Índice general

<b>Agradecimientos</b>	<b>III</b>
<b>Índice de figuras</b>	<b>VII</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Proceso de modelación . . . . .	1
<b>2. Antecedentes</b>	<b>3</b>
2.1. Problema inverso . . . . .	3
2.1.1. Modelos descritos por ODEs . . . . .	4
2.2. Solución bayesiana a problemas inversos . . . . .	6
2.3. Método MCMC . . . . .	9
2.3.1. Algoritmo Metropolis-Hastings . . . . .	9
2.4. Forward map aproximado . . . . .	11
2.4.1. Construcción del forward map aproximado . . . . .	12
2.4.2. Consistencia y utilidad del forward map aproximado . . . . .	13
<b>3. Implementación</b>	<b>15</b>
3.1. Modelos de trabajo . . . . .	15
3.1.1. Modelo Gravitatorio . . . . .	16
3.1.2. Modelo Logístico . . . . .	19
3.1.3. Modelo SIR . . . . .	22
3.2. Enfoque bayesiano al problema inverso . . . . .	23
3.2.1. Distribución posterior . . . . .	24
3.2.2. Simulación con forward map ordinario . . . . .	26
3.3. Simulación con forward map aproximado . . . . .	38
3.3.1. Regularización de las unidades . . . . .	38
3.3.2. Construcción de la regularización de unidades . . . . .	40
3.3.3. Experimentación con distribuciones posteriores aproximadas .	40
<b>4. Conclusiones y trabajo futuro</b>	<b>51</b>
<b>Referencias</b>	<b>55</b>



# Índice de figuras

3.1. Varias trayectorias con la dinámica gravitación sujeta a fricción para varias valoress de los parámetros g,b. . . . .	19
3.2. Varias trayectorias con la dinámica de crecimiento logístico para varias valoress de los parámetros K,r. . . . .	21
3.3. Trayectoria de cada grupo del modelo SIR con $\beta = 0.009$ y $\gamma = 0.5$ . . . . .	23
3.4. Distribuciones marginal a priori para el parámetro $g$ (izquierda) y $b$ (derecha). . . . .	28
3.5. Muestra $y$ del modelo gravitatorio. . . . .	29
3.6. Estimación de la distribución posterior conjunta por método MCMC Metropolis-Hastings. . . . .	29
3.7. Distribución posterior (azul) y distribución a priori (verde) para el parámetros $g$ y $b$ del modelo gravitatorio. Los valores verdaderos se marcan en negro. . . . .	30
3.8. Distribución predictiva para el modelo gravitatorio . . . . .	31
3.9. Distribución predictiva del modelo gravitatorio para los valores de tiempo de 0.2s y 0.7s. . . . .	31
3.10. Muestra $y$ para el modelo logístico. . . . .	32
3.11. Distribuciones a priori para $\theta_1$ y $\theta_2$ . . . . .	33
3.12. Posterior conjunta para el modelo logístico. . . . .	33
3.13. Distribución posterior (azul) y distribución a priori (verde) para el parámetros $\theta_1$ y $\theta_2$ del modelo logístico. Los valores verdaderos se marcan en negro. . . . .	34
3.14. Distribución predictiva para el modelo logístico. . . . .	34
3.15. Muestra $y$ del modelo SIR. . . . .	35
3.16. Distribuciones marginal a priori para el parámetro $\beta$ (izquierda) y $\gamma$ (derecha). . . . .	36
3.17. Distribución posterior conjunta para el modelo SIR. . . . .	36
3.18. Distribución posterior (azul) y distribución a priori (verde) para el parámetros $\beta$ y $\gamma$ del modelo SIR. Los valores verdaderos se marcan en negro. . . . .	37
3.19. Distribución predictiva para el modelo SIR. . . . .	37
3.20. Malla para el modelo gravitatorio (azul) con los 8 vecinos más cercanos al parámetro $\theta$ (rojo) con unidades estándar (izquierda) y unidades anglosajonas (derecha). . . . .	39

3.21. Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a tres vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo gravitacional. . . . .	42
3.22. Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a cinco vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo gravitacional. . . . .	42
3.23. Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a ocho vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo gravitacional. . . . .	43
3.24. Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a tres vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo logístico. . . . .	45
3.25. Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a cinco vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo logístico. . . . .	45
3.26. Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a ocho vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario. . . . .	46
3.27. Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a tres vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo SIR. . . . .	47
3.28. Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a cinco vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo SIR. . . . .	48
3.29. Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a ocho vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo SIR. . . . .	49
4.1. Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a cinco vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha, sin regularización de las unidades. En azul la distribución posterior con forward map ordinario para modelo SIR. . . . .	52





# Capítulo 1

## Introducción

En una amplia gama de disciplinas se utilizan modelos que pretenden describir diversos fenómenos. La inmensa mayoría de estos modelos son matemáticos y establecen relaciones entre variables expresadas mediante ecuaciones dependientes de ciertos parámetros. El tipo de modelos que vamos a considerar en esta tesis son aquellos cuyas variables se pueden expresar ya sea en una forma funcional o en forma de ecuación diferencial ordinaria (ODE).

### 1.1. Proceso de modelación

El proceso de modelación de un fenómeno involucra tres sectores principales. Al investigar un fenómeno, se identifican, proponen y prueban las variables que influyen en él. Este proceso, conocido como **parametrización**, implica seleccionar las variables relevantes para la descripción del modelo.

Una vez determinadas las variables del modelo, se establecen explícita o implícitamente las relaciones entre ellas. Se pueden proponer modelos lineales generalizados, series temporales, ecuaciones diferenciales o simplemente ecuaciones algebraicas con las variables y parámetros de ajuste, que en adelante llamaremos simplemente parámetros. Este sector del proceso se conoce como **modelación directa**.

El último sector de la modelación es el **problema inverso**. Como su nombre indica, se refiere a inferir los parámetros del modelo a partir de una muestra del fenómeno

modelado. Esto es opuesto al problema directo, que busca describir explícitamente la dinámica del fenómeno a partir de parámetros fijos (Calvetti y Somersalo, 2018).

El proceso de modelación no sigue una línea recta. Es por ello que se optó por llamar a sus partes como sectores de modelación en lugar de etapas de modelación, ya que todos estos sectores siguen un camino cíclico según las virtudes de cada modelo en particular.

Una vez aclarado que los modelos considerados aquí son aquellos que pueden expresarse como una ODE dependientes de parámetros, la tesis se centra en el problema inverso desde el enfoque bayesiano. En el capítulo 2 se describe el paradigma bayesiano aplicado al problema inverso, además de los antecedentes necesarios para hacer de este texto autocontenido.

En el capítulo 3 se describen las implementaciones realizadas para la inferencia bayesiana de los parámetros de tres modelos particulares según el paradigma convencional. Además, se propone un método aproximado para realizar el mismo procedimiento, cuestionando la factibilidad de esta propuesta como sustituto del paradigma original.

El objetivo de esta tesis es experimentar con el uso de modelos sustitutos para el problema directo, es por ello que el método aproximado propuesto aquí es central para el trabajo de tesis. Las inferencias de los modelos sustitutos en sus diferentes variantes son plasmadas en el capítulo 3 donde las conclusiones obtenidas se describen en el capítulo 4.

# Capítulo 2

## Antecedentes

En el capítulo precedente relata la estructura subyacente de la modelación. Formalizaremos y detallaremos los lineamientos del sector referente al problema inverso, cuya gestión no es única para cada caso, pues esta sujeto a las vicisitudes del modelo en cuestión. A diferencia de la mayoría de los enfoques al problema inverso, el enfoque bayesiano goza de ser un método apropiado pues para los modelos contemplados aquí la metodología es análoga ([Stuart, 2010](#)).

### 2.1. Problema inverso

El estudio del problema inverso se remonta a mediados del siglo XVII. Los físicos contemporáneos interesados en la relación causa-efecto de diferentes fenómenos físicos lograron predecir el comportamiento en la dinámica intrínseca del fenómeno en cuestión. Así, para la ley de gravitacional de Newton, una vez obtenidas las masas de los cuerpos celestes (causas) es realizable predecir las trayectorias que siguen los cuerpos en el tiempo (efecto). Estudiar los efectos de un fenómeno dada ciertas causas es el problema directo perteneciente al sector de modelación directa. El problema inverso toma la dirección opuesta, dados los efectos, es de interés investigar las causas que lo ocasionaron. Retomando un ejemplo newtoniano, dado un campo gravitatorio circundando cierto cuerpo celeste, es de interés conocer la masa de dicho cuerpo. Enfaticemos que el problema inverso adolece de no ser único, suele ser el caso

que a ciertos efectos se correspondan varias causas. El hecho de que no exista una mapeo uno a uno entre causa y efecto complica el problema inverso. Sin embargo, la información que se tenga a priori del fenómeno es crucial para determinar con una menor ambigüedad las causas ([Tarantola, 2005](#)).

### 2.1.1. Modelos descritos por ODEs

Para describir fenómenos los modelos matemáticos pueden ser de una suntuosa cantidad de clases o estilos. Con fines pedagógicos, considere que se desea describir los costos de la renta  $Y$  de inmuebles en cierta ciudad (efecto). Supóngase que de análisis previos se concluyó que dicho fenómeno está vinculado al tamaño del inmueble  $X_1$ , número de habitaciones  $X_2$ , ingreso medio por habitante  $X_3$  (causas). Un modelo plausible es por modelos lineales generalizados

$$Y = G(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3) + \varepsilon \quad (2.1)$$

donde  $G$  es la función liga,  $\varepsilon \sim N(0, \sigma^2)$  un error aleatorio y  $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$  son los parámetros del modelo ([Dobson y Barnett, 2018](#)). El problema directo considera predecir el valor de  $Y$  conociendo  $X = (X_1, X_2, X_3)$  y los parámetros del modelo  $\beta$ . Para esta clase de modelos el problema directo no representa complejidad. En contraste, el problema inverso precisa estimaciones de  $\beta$  y  $\sigma$  dada muestras de  $Y$  y  $X$ , siendo un problema no trivial.

El análisis del problema inverso para modelos en general es bastante amplio. Por ello es necesario restringirlo a una clase menor de modelos, siendo estos los modelos dados por ODEs. Considérese un modelo para describir la dinámica de  $y(t)$  a lo largo del tiempo  $t$ . Los modelos considerados son aquellos que se pueden expresar de la forma

$$G(t, y(t), y'(t), y''(t), \dots) = 0$$

que es una ODE con  $G$  una función continua conocida y  $y(0) = y_0, y'(0) = y'_0, \dots$  sus

condiciones iniciales.

Más aún, se puede obtener el análisis del problema inverso para sistemas de ecuaciones diferenciales ordinarias. Esto es

$$\left\{ \begin{array}{l} G_1(t, y_1(t), y'_1(t), \dots, y_2(t), y'_2(t), \dots, y_d(t), y'_d(t), \dots) = 0 \\ G_2(t, y_1(t), y'_1(t), \dots, y_2(t), y'_2(t), \dots, y_d(t), y'_d(t), \dots) = 0 \\ \vdots \\ G_d(t, y_1(t), y'_1(t), \dots, y_2(t), y'_2(t), \dots, y_d(t), y'_d(t), \dots) = 0 \end{array} \right.$$

con  $G_1, G_2, \dots, G_d$  funciones continuas conocidas y condiciones iniciales  $y_i(0) = y_{0i}, y'_i(0) = y'_{0i}, \dots$  para  $i \in \{1, \dots, d\}$  (Apostol, 2019).

Observemos que la restricción a modelos descritos por ODEs implica restringir las variables en un soporte continuo, esto por la continuidad en la solución de las ODEs, por lo que modelos con variables discretas deben de replantearse.

En el capítulo 3 se detallan tres modelos de los cuales se consideraron los análisis del problema inverso. Dentro de estos se consideran modelos como la dinámica de caída libre sujeto a fricción, donde la distancia recorrida en caída se describe por

$$mx''(t) = mg - bx'(t) \quad (2.2)$$

con  $x(0) = x_0$  y  $x'(0) = v_0$  donde  $m, g, b$  parámetros del modelo. En el supuesto de desconocer las condiciones iniciales, estas se consideran dentro del espacio de parámetros, sin embargo en este modelo no es el caso ya que se posse información de estas.

Por el teorema de existencia y unicidad en ecuaciones diferenciales; dadas las condiciones iniciales, los parámetros del modelo definen únicamente el modelo (Kelley, 2010). Consideremos el espacio de parámetros  $\Theta$  que a su vez es la familia de *submodelos* posibles descrito por la ODE. En el caso del modelo (2.2) los parámetros son  $m, g$  y  $b$ , por lo que el espacio paramétrico  $\Theta$  es el conjunto que se conforma por  $\theta = (m, g, b)$  todos los *submodelos* de la familia (2.2).

El problema directo es entonces aquel que dado un  $\theta \in \Theta$  obtiene la trayectoria o

solución de la ecuación diferencial. Dicho problema está bien definido, pues resolver la ODE es posible al menos numéricamente. Denotemos por  $\mathcal{Y}$  al espacio de las soluciones posibles de la ecuación diferencial. De esta forma existe un mapeo del espacio  $\Theta$  a  $\mathcal{Y}$  el cual llamaremos **forward map**. Así, el forward map es

$$\theta \mapsto F[\theta]$$

donde  $\theta \in \Theta$  y  $F[\theta] \in \mathcal{Y}$ .

Para el modelo dado en (2.2) una vez dado  $\theta = (m, g, b)$  la solución  $x(t) \in \mathcal{Y}$  se obtiene del forward map, que en otras palabras es simplemente la solución analítica o numérica de la ecuación diferencial.

## 2.2. Solución bayesiana a problemas inversos

Principalmente existen dos razones para que las observaciones de un modelo no concuerden exactamente con las predicciones dadas por el mismo. La primera se debe a errores de medición por la incertidumbre de los aparatos de medición. El segundo motivo se debe a los defectos propios del modelo, pues a pesar de que el modelo pretende describir el fenómeno de interés estos nunca son lo mismo. Así se considera la vaga interpretación que jacta a los modelos como aproximaciones de los fenómenos. La relevancia de los errores abre una puerta a la cuantificación de la incertidumbre.

Por otro lado, el paradigma bayesiano para problemas inversos se centra en cuantificar la incertidumbre en los parámetros del modelo. Es decir, su objetivo es establecer una medida de probabilidad posterior a las observaciones de la trayectoria del modelo. Equivalentemente, se busca la distribución de probabilidad  $\pi(\theta|\mathbf{y})$  donde  $\mathbf{y} = (y_1, \dots, y_n)$  son las observaciones de la trayectoria a lo largo del tiempo  $t_1, t_2, \dots, t_n$ . Partiendo de la información acerca de los parámetros previo a cualquier observación, se propone una distribución de probabilidad  $\pi(\theta)$  llamada distribución a priori, y tras aplicar el Teorema de Bayes obtener la distribución posterior (Wasserman, 2013).

Como ya se ha mencionado, no se espera que las observaciones de la trayectoria coincidan con las predicciones del modelo. En virtud de lo mencionado, ajustamos los errores conforme a una distribución normal. Para modelos dinámicos, aquellos que evolucionan con el tiempo, la predicción se obtiene del forward map con un vector de parámetros  $\theta \in \Theta$  dado. De esta forma, obtenemos una función en el tiempo  $y(t)$ , Es decir, se tiene la igualdad  $F[\theta](t) = y(t)$ . Luego, se requieren las predicciones dadas por el modelo a tiempos fijos  $t_i$ , la cual simplemente es la evaluación  $y(t_i)$ , para todo  $i \in \{1, \dots, n\}$ . Usando la notación alterna con el forward map, se tiene  $F_\theta(t_i)$  como la predicción a tiempo fijo.

De esta forma, se establece que las discordancias entre observaciones y predicciones siguen una distribución normal de la forma

$$y_i = F_\theta(t_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2),$$

donde las  $y'_i$ s son las observaciones del fenómeno.

Dicho ajuste es elegido por antonomasia, debido a la predilección dentro de los de su clase. Cualquier otra distribución o forma funcional para ajustar los errores entre medición y modelo es un tópico interesante, sin embargo sale del propósito de la tesis por lo que se restringe al estudio clásico de ajuste ([Berger, 2013](#)).

En consecuencia, se ha establecido una distribución para las observaciones  $y_i$  las cuales tienen asociada una verosimilitud sobre  $\theta$  y  $\sigma$ . Como se consideran errores condicionalmente independientes, pues se tratan de errores de medición, la verosimilitud se sigue de

$$\mathcal{L}(\theta, \sigma) = f(\mathbf{y}|\theta, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_i - F_\theta(t_i))^2 \right\},$$

simplificando

$$f(\mathbf{y}|\theta, \sigma) = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - F_\theta(t_i))^2 \right\}, \quad (2.3)$$

Del teorema de Bayes, se obtiene la distribución posterior para los parámetros

$$\pi(\theta, \sigma | \mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{\int f(\mathbf{y}|\theta)\pi(\theta)d\theta}. \quad (2.4)$$

donde la constante de integración  $h(\mathbf{y}) = \int f(\mathbf{y}|\theta)\pi(\theta)d\theta$  es la constante de normalización para la distribución posterior.

Salvo en análisis conjugado, donde la distribución a priori y la distribución posterior pertenecen a la misma familia, la obtención analítica de la constante de normalización  $h(\mathbf{y})$  es un problema complejo. De aquí surge la necesidad de métodos numéricos para determinar la distribución posterior con precisión ([Robert, Casella, y Casella, 1999](#)).

Utilizar métodos numéricos de integración para  $h(\mathbf{y})$  solamente es plausible para modelos de un solo parámetro ( $d = 1$ ), es decir el espacio parametal  $\Theta \subset \mathbb{R}$ . En el caso de más de un parámetro ( $d > 1$ ) aún es posible usar métodos numéricos, sin embargo tienen un desempeño deficiente.

En los años 1990's, la estadística bayesiana inició un crecimiento sin precedentes tras descubrir e implementar métodos Montecarlo. Los métodos Montecarlo son un conjunto de algoritmos iterativos no deterministas con fin de estimar el valor de cierto calculo. El método Markov Chain Monte Carlo (MCMC) resulta ser conveniente para la estadística bayesiana, no solo porque evita el calculo de la constante de integración  $h(\mathbf{y})$  sino por la generalidad que propicia al poder utilizarse para estimar una basta familia de distribuciones. Además, la estimación por este método tiene un error absoluto que decrece como  $\frac{1}{\sqrt{N}}$  en virtud del teorema del límite central ([Casella y Berger, 2024](#)), que es de orden menor que aquellos dados por métodos numéricos de integración. La asiduidad del método MCMC, tras prevalecer en el tiempo, nos da una idea de lo apropiado que es para la estadística bayesiana.

## 2.3. Método MCMC

Obtener la distribución posterior (2.4) con el método de Markov Chain Monte Carlo (MCMC) evita el calculo numérico de  $h(\mathbf{y})$  como se ha mencionado previamen-  
te. En su lugar, MCMC estima la distribución posterior con una generosa muestra  
simulada de la misma distribución posterior salvo la constante de normalización. Por  
ende, con la misma muestra permite obtener estimaciones a cualquier momento de la  
distribución posterior , así como la estimación de las distribuciones marginales. Los  
métodos MCMC buscan encontrar una cadena de Markov  $X^{(1)}, X^{(2)}, \dots$  cuya distri-  
bución estacionaria sea la *distribución objetivo*<sup>1</sup>  $f(x)$  pese a que la misma no esté  
normalizada.

### 2.3.1. Algoritmo Metropolis-Hastings

El algoritmo de Metropolis-Hastings es un método de MCMC para generar mues-  
tras de una distribución objetivo  $f(x)$  partiendo de muestras de una distribución pro-  
puesta  $q(y|x)$  que no son necesariamente simétricas. Al igual que todo método MCMC,  
el algoritmo de Metropolis-Hastings genera una cadena  $X_0, X_1, \dots, X_N$  construyendo  
recursivamente la cadena dependiendo solamente del estado previo, es decir preser-  
vando la propiedad de Markov. El estado inicial  $X_0$  se toma aleatoriamente. Luego,  
teniendo hasta el estado  $X_i$ , el estado  $X_{i+1}$  se obtiene siguiendo

1. Generar una propuesta  $Y \sim q(y|X_i)$ .

2. Calcular  $\rho$

$$\rho = \min \left\{ \frac{f(y)q(x|y)}{f(x)q(y|x)}, 1 \right\}$$

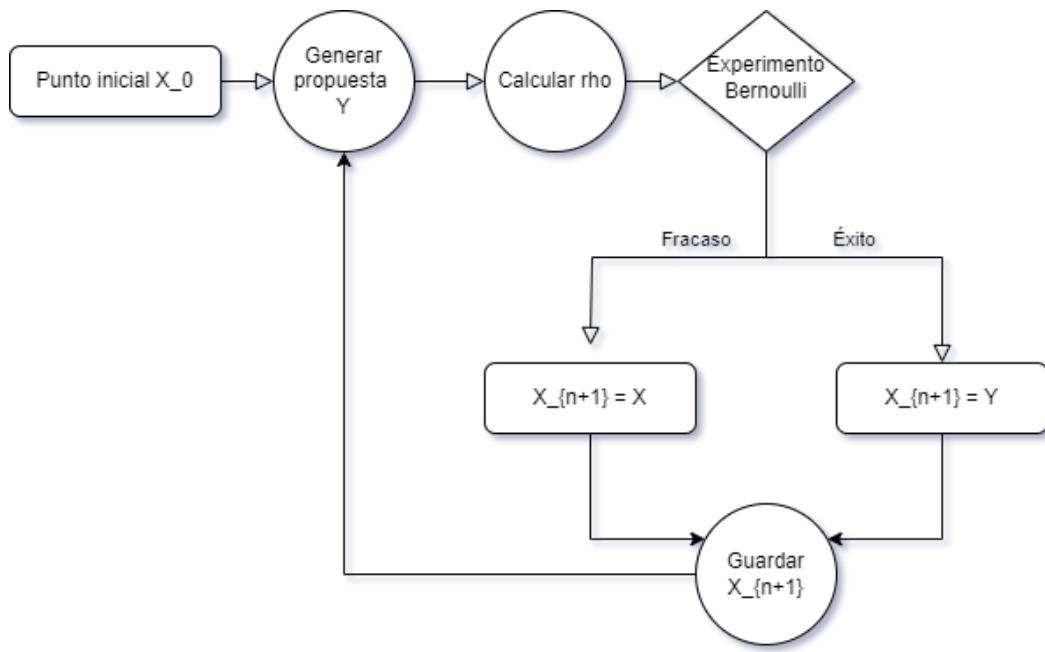
---

<sup>1</sup>En aplicaciones a la estadística bayesiana usando MCMC, la distribución posterior es la distri-  
bución objetivo,  $\pi(\theta, \sigma|\mathbf{y})$  en nuestro caso.

3. Realizar un experimento Bernoulli con probabilidad de éxito  $\rho$ .

$$X_{i+1} = \begin{cases} Y & , \text{ con probabilidad } \rho \\ X_i & , \text{ con probabilidad } 1 - \rho \end{cases}$$

En el siguiente gráfico se muestra un diagrama de flujo del algoritmo Metropolis-Hastings.



El algoritmo previamente esbozado a manera de opúsculo, subrepticiamente converge en distribución estacionaria a la distribución objetivo bajo condiciones generales de regularidad. Indudablemente dicha convergencia está asegurada por el Teorema Ergódico en cadenas de Markov ([Norris, 1998](#)). En beneficio del mencionado teorema, basta con construir una cadena de Markov de estados continuos a tiempos discretos recurrente e irreducible y aperiódica cuyo kernel de transición cumpla con balance detallado con respecto a  $f(x)$ . Para ahondar en detalles se requiere de un extenso estudio en cadenas de Markov a tiempo discreto con estados arbitrarios, donde en lugar de la usual matriz de transición se usa un operador de salto llamado kernel de transición. Un estudio completo del algoritmo de Metropolis-Hastings se encuentra en ([Mengersen y Tweedie, 1996](#)).

Pese a la enrevesada formalización del método, la estructura del algoritmo Metropolis-Hastings basta para dilucidar en ciertos detalles. Primeramente, del paso uno es claro que se debe elegir una distribución asociada a la variable propuesta  $Y$  de forma que no sea un problema simular de esta. Además, dado que la propuesta se acepta o rechaza para ser realización de la variable  $X$  asociada a la distribución objetivo, entonces se debe elegir una distribución de  $Y$  tal que  $\text{supp}\{X\} \subset \text{supp}\{Y\}$ , que denota el soporte de la variable  $Y$  debe contener el soporte de la variable objetivo  $X$ .

Asimismo, el segundo paso del algoritmo Metropolis-Hastings se observa la prescindencia del factor de normalización de la distribución objetivo puesto que solo es relevante el cociente  $\frac{f(y)}{f(x)}$ , un atributo peculiar de la distribución posterior. Paralelamente, el tercer paso del algoritmo se observa la propiedad de Markov, ya que la aceptación de la propuesta para pertenecer a la cadena depende estocásticamente del estado que lo precede únicamente.

## 2.4. Forward map aproximado

Hay que destacar que, una vez abordado el enfoque bayesiano para el problema inverso y el método MCMC para obtener la distribución posterior de los parámetros del modelo bajo estudio, el reto a superar recae en la implementación de la metodología descrita. Es preciso tener presente que pese a ser una metodología ampliamente funcional, no se cierra la puerta a indagar mejoras en su procedimiento. En el metodología ordinaria previamente descrita se reconoce que las multiples evaluaciones del forward map en cada iteración del algoritmo Metropolis-Hastings consume una sustancial cantidad de tiempo computacional para su procesamiento. Por lo anterior es que la propuesta principal del trabajo de tesis es la introducción del forward map aproximado.

El forward map aproximado  $F_\theta^*$  es una construcción multifacética basada en aproximaciones de estadística espacial. Tal aproximación pretende crear una distribución posterior aproximada  $\tilde{\pi}(\theta|\mathbf{y})$  con la sustitución del forward map a su versión aproximada, de forma que la simulación de esta nueva distribución posterior sea compu-

tacionalmente menos pesada y el MCMC pueda llevarse a cabo.

### 2.4.1. Construcción del forward map aproximado

Existen diferentes maneras de hacer las aproximaciones del forward map. La forma propuesta es considerar una discretización del espacio de parámetros  $\Theta$ . Usando coordenadas canónicas, cada  $\theta \in \Theta \subset \mathbb{R}^d$  se puede escribir como  $\theta = (\theta_1, \theta_2, \dots, \theta_d)$ . Consideremos para cada coordenada un intervalo  $[\theta_i^{min}, \theta_i^{max}]$  para toda  $i \in \{1, \dots, d\}$ , donde  $\theta_i^{min}$  y  $\theta_i^{max}$  son cotas para el espacio de parámetros que contenga la masa de probabilidad. Puede pensarse en tomar cuantiles de 0.01 y 0.99 respectivamente de la distribución marginal a priori para  $\theta_i$ . Posteriormente, tomemos una partición equidistante en  $M$  puntos para cada coordenada. Más llanamente, la partición de la coordenada  $i$ -ésima es el conjunto  $\mathcal{M}_i = \{\theta_i^{(1)}, \theta_i^{(2)}, \dots, \theta_i^{(M-1)}, \theta_i^{(M)}\}$  con  $\theta_i^{(1)} = \theta_i^{min}$  y  $\theta_i^{(M)} = \theta_i^{max}$ . De esta forma, la partición crea una malla  $\mathcal{M} = \mathcal{M}_1 \times \dots \times \mathcal{M}_d$  del espacio paramétrico  $\Theta$ . Obsérvese que la cardinalidad de  $\mathcal{M}$  es de  $M^d$ . Denotemos por  $\vartheta$  a los elementos de  $\mathcal{M}$ , así el enmallado se constituye de  $M^d$  parámetros. Es decir, se establece que  $\mathcal{M} = \{\vartheta_1, \vartheta_2, \dots, \vartheta_{M^d}\}$ , nótese que  $\vartheta_j \in \Theta$  para toda  $j \in \{1, \dots, M^d\}$ .

Seguidamente, se requiere evaluar el forward map para cada elemento de  $\mathcal{M}$ , en la notación acordada se requiere  $y_j = F[\vartheta_j]$ , con  $y_j = y_j(t)$  funciones continuas, que al estar restringidos a modelos dados por ODE es necesario un cálculo intermedio. Dicho en otras palabras, solo se resuelven las ecuaciones diferenciales con los parámetros donde el enmallado intersecta.

Para terminar, el forward map aproximado para parámetros  $\theta \notin \mathcal{M}$  se propone buscar a los  $k$  vectores de parámetros  $\vartheta \in \mathcal{M}$  más cercanos en distancia euclídea a  $\theta$ , denotando estos  $k$  vecinos por  $\vartheta^{(1)}, \dots, \vartheta^{(k)}$  cada uno a una distancia  $d_1, \dots, d_k$  de  $\theta$ , respectivamente; con  $d_1 \leq d_2 \leq \dots \leq d_k$ . La aproximación al forward map con  $k$  vecinos en una malla con  $M$  particiones es

$$\tilde{F}_M^k(\theta) = \sum_{j=1}^k \omega_j(\theta) F(\vartheta^{(j)}) \quad (2.5)$$

donde  $\omega_j(\theta) = d_j^{-1} / \sum_{i=1}^k d_i^{-1}$ , que es una suma ponderada inversamente por las distancias, ya que se pretende que parámetros ‘lejanos’ de  $\theta$  tengan menor relevancia que los ‘cercanos’.

### 2.4.2. Consistencia y utilidad del forward map aproximado

La preeminencia potencial del uso del forward map aproximado para el estudio del problema inverso reside parcialmente en un menor tiempo de ejecución de la implementación en contraste con su análogo ordinario. Existen modelos de ecuaciones diferenciales cuya implementación del problema inverso bajo enfoque bayesiano toma días en ejecutarse, véase [Galaviz \(2023\)](#) y [Cotter, Dashti, y Stuart \(2010\)](#). Debido a la reticente necesidad humana en aplicaciones imposergables es imperativo agilizar el tiempo en operación. Concisamente en modelos epidemiológicos la rapidez de acción se torna fundamental como lo fue en la predicción de ocupación hospitalaria por COVID-19 en México ([Capistran, Capella, y Christen, 2021](#)).

Por otro lado, fuera de las aplicaciones prácticas del método esbozado con el forward map aproximado, es de sumo interés matemático el estudio de la consistencia del método propuesto. Recordemos que tenemos espacio de maniobra para la construcción del forward map aproximado al mover la densidad de puntos en la malla  $M$  así como en la cantidad de vecinos  $k$  usados para la suma ponderada;  $\tilde{F}_M^k$ . De esta misma forma, tras sustituir el forward map aproximado (2.5) en la expresión (2.4) obtenemos la distribución posterior aproximada de los parámetros del modelo. Formalizando, salvo un factor de normalización definimos

$$\tilde{\pi}_M^k(\theta|\mathbf{y}) \propto \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - F_M^k(\theta))^2 \right\} \pi(\theta), \quad (2.6)$$

la distribución posterior aproximada dado las observaciones  $\mathbf{y} = (y_1, \dots, y_n)$  al tiempo  $t_1, \dots, t_n$ ; respectivamente. Consideremos que la aproximación a la distribución posterior usando la metodología ordinaria del problema inverso es en efecto una aproximación plausible usando la *distancia de Kullback-Leibler*.

Si  $f(x)$  y  $g(x)$  son funciones de densidad de probabilidad. La distancia de Kullback-

Leibler entre  $f$  y  $g$  se define como

$$D(f, g) = \int f(x) \log \left( \frac{f(x)}{g(x)} \right) dx.$$

Se puede mostrar que  $D(f, g) \geq 0$  y  $D(f, f) = 0$  ([Wasserman, 2006](#)). Definamos la distancia entre las distribuciones posteriores como

$$c_M^k = D(\pi(\theta|\mathbf{y}), \tilde{\pi}_M^k(\theta|\mathbf{y}))$$

Diremos que el forward map aproximado es **consistente** si para una cantidad de vecinos  $k$  fijo y conocido, las distancias

$$c_M^k \rightarrow 0 \quad \text{a medida que} \quad M \rightarrow \infty$$

En el caso de un vecino ( $k = 1$ ) observemos que, desde el punto de vista computacional, las distribuciones posteriores  $\tilde{\pi}_M^k(\theta|\mathbf{y})$  y  $\pi(\theta|\mathbf{y})$  se construyen de forma semejante para  $M$  grande. Sin embargo, demostrar que  $C_M^1 \rightarrow 0$  requiere de una restricción en la hipótesis, como se menciona en lo siguiente.

Mas aún, el estudio para  $k > 1$  requiere de una estructura teórica sobre las cualidades particulares de la ecuación diferencial del modelo en cuestión, además de hipótesis sobre las mismas distribuciones posteriores. Sin embargo, existen teoremas para convergencia en variación total y convergencia débil de las distribuciones posteriores con forward map aproximado para casos particulares, véase [Christen y Pérez-Garmendia \(2021\)](#).

Dada la naturaleza de los modelos propuestos para esta tesis, para indagar sobre la consistencia es necesario usar técnicas heurísticas. En el capítulo siguiente se consideran modelos paramétricos particulares aplicando las metodologías del problema inverso con y sin aproximación para así cotejar empíricamente cierto grado de convergencia.

# Capítulo 3

## Implementación

La teoría desarrollada referente a la convergencia en problemas inversos con forward map aproximado no es suficientemente general para considerarse un estudio integral. En su defecto se ha optado por un desarrollo heurístico centrado en experimentación con una variedad de parámetros o indicadores estratégicamente seleccionados. En adelante encararemos dos frentes, el primero de ellos es la implementación en código por medio del lenguaje de programación Python, donde se tomarán modelos rudimentarios de ecuaciones diferenciales ordinarias. El segundo frente es el estudio de las distribuciones posteriores aproximadas en función de los atributos del forward map aproximado. Cabe enfatizar que parte considerable del trabajo de tesis involucra la ejecución de conocimientos técnicos de programación. La implementación fundamental se encuentra en el [enlace](#) adjunto <sup>1</sup>.

### 3.1. Modelos de trabajo

Existen una variedad de modelos a los que se puede aplicar la metodología anteriormente descrita para problemas inversos. Sin embargo, cada modelo tiene sus bemoles, lo que dificulta calibrar (en general), para cada aspecto en la distribución de los parámetros, como puede ser el caso de tomar las distribuciones a priori adecuadas, la variación de la muestra, los dominios adecuados para graficar las distribuciones a

---

<sup>1</sup>Implementación disponible en <https://github.com/cesarhttpc/Tesis>

posterior, entre otras. Por ello, nos permitiremos enfocarnos en solo tres modelos que son el modelo gravitatorio sujeto a fricción para un partícula en caída, el modelo de crecimiento logístico para una población en crecimiento y finalmente el modelo epidemiológico SIR para brotes de enfermedades de transmisión directa. Más adelante analizaremos el uso de modelos sustitutos.

### 3.1.1. Modelo Gravitatorio

Consideremos una partícula puntual en un campo gravitatorio cercano a la superficie terrestre. Por medio de la dinámica clásica podemos modelar dicha caída con las leyes de Newton. Sea  $x(t)$  la distancia recorrida (unidimensional) por la partícula puntual en el tiempo  $t$ . De la segunda ley de Newton sabemos que

$$\sum F_i = m\ddot{x}(t), \quad (3.1)$$

donde  $\sum F_i$  es la suma de fuerzas ejercida sobre la partícula con masa  $m$ .

Para el modelo de caída libre nos dice que la fuerza ejercida en la partícula es constante y se le conoce como constante de aceleración gravitacional  $g$ . De forma que la trayectoria se rige de (3.1) con únicamente la fuerza gravitacional  $F = mg$ . La ecuación de la dinámica es

$$\ddot{x}(t) = g,$$

bajo las condiciones iniciales  $x(0) = x_0, \dot{x}(0) = v_0$ . La solución a la ecuación dinámica es

$$x(t) = \frac{1}{2}gt^2 + v_0t + x_0.$$

Véase los detalles del modelo en [Alonso y Finn \(1970\)](#)

Sin embargo, dentro del marco establecido para el modelo gravitacional, no es un modelo realista ya que este no está considerando la desaceleración producto de la fricción con el medio que interactúa (aire). Por ello es necesario modelar la fuerza de

fricción de forma adecuada. Un modelo físico plausible para la fricción es considerar una fuerza opuesta a la fuerza de gravedad y que está depende de la velocidad de la partícula. Consideremos dicha fuerza de fricción como  $F_f = -bv(x)$ . Podemos pensar que el modelo para la fuerza de fricción es una aproximación en polinomios de Taylor a primer grado, donde se descarta el termino constante ya que no tiene sentido físico que la fricción obedezca a marcos de referencia.

El modelo que estamos interesados es entonces el **modelo gravitatorio sujeto a fricción**, cuya ecuación dinámica es

$$m\ddot{x} = mg - b\dot{x}, \quad (3.2)$$

que representa la trayectoria de la partícula sujeta a dos fuerzas opuestas, la gravitatoria y la fricción con el medio. Además con las condiciones iniciales  $x(0) = x_0$ ,  $\dot{x}(0) = v_0$ . Notemos que dados los parámetros  $\theta = (g, b)$  podemos determinar unívocamente la trayectoria de interés.

Para poder hacer inferencia bayesiana en el problema inverso, es necesario construir el forward map  $F(\theta)$ . Esto puede ser de dos maneras, podemos resolver la ecuación diferencial y obtener  $x(t)$  en términos de  $\theta = (g, b)$  en caso de que exista dicha solución explícita. Otra forma de construir el forward map es utilizando métodos numéricos para resolver la EDO.

En este modelo sí tenemos solución explícita a la ecuación de la dinámica. Para poder resolverla expresamos la EDO en términos de la velocidad  $v(t) = \dot{x}(t)$ . Teniendo una EDO lineal de primer orden.

La solución a la ecuación dinámica

$$m \frac{dv}{dt} = mg - bv, \quad (3.3)$$

se obtiene del siguiente análisis. Primeramente notemos que la aceleración de la partícula se anula. Es decir, la velocidad tiene una asíntota. Esto corresponde al caso cuando la fuerza gravitacional es igual y opuesta a la fuerza de fricción. Por tanto

podemos definir una velocidad terminal  $v_T$  de la relación

$$mg - bv_T = 0, \quad \Rightarrow \quad v_T = \frac{mg}{b}, \quad (3.4)$$

de reordenar e integrar (3.3)

$$\frac{dv}{dt} = -\frac{b}{m}(v - v_T), \quad (3.5)$$

e integrando con la condición inicial  $v_0 = 0$

$$\int_0^v \frac{dv}{v - v_T} = -\frac{b}{m} \int_0^t dt,$$

obtenemos

$$\log \left( \frac{v_t - v}{v_t} \right) = -\frac{b}{m} t,$$

despejando para  $v$

$$v = v_T \left( 1 - \exp \left\{ -\frac{b}{m} t \right\} \right), \quad (3.6)$$

donde vemos que en efecto la velocidad terminal  $v_T$  es una asíntota debido a que  $v(t)$  ya que se aproxima a  $v_T$  a medida que pasa el tiempo.

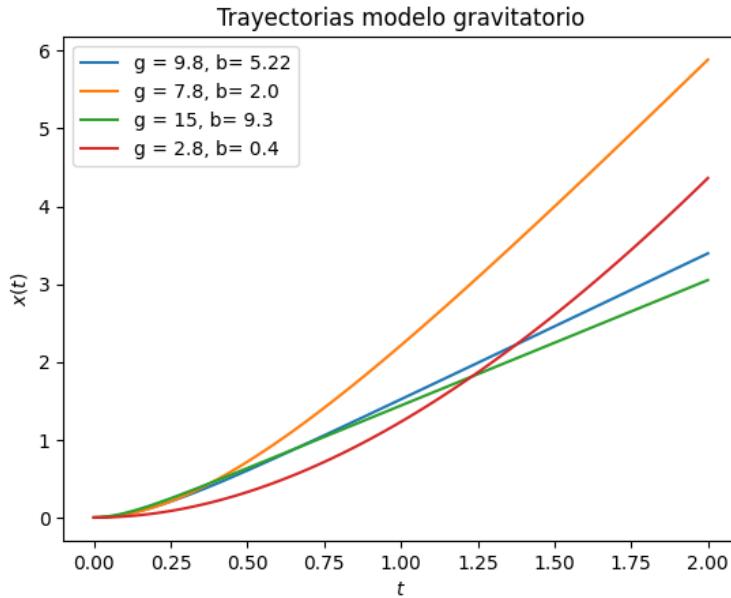
Para obtener la función de la trayectoria, simplemente integramos una vez más obteniendo

$$x(t) = \int_0^t v_T \left( 1 - \exp \left\{ -\frac{b}{m} t \right\} \right) dt + x_0,$$

tomando  $x_0 = 0$ , finalmente nos queda que la trayectoria de la partícula es

$$x(t) = v_T \left[ t - \frac{m}{b} \left( 1 - \exp \left\{ -\frac{b}{m} t \right\} \right) \right], \quad (3.7)$$

apegándose al método de integración dado por Sears y cols. (1986).



**Figura 3.1:** Varias trayectorias con la dinámica gravedad sujeta a fricción para varias valores de los parámetros  $g, b$ .

Por tanto, podemos tomar el forward map  $F(g, b) = x(t)$ , prescindiendo de  $m$  al considerar  $m = 1$ .

Por otro lado, recordemos que la idea principal es resolver el problema inverso. Es decir, dada una muestra de la trayectoria  $x(t_i)$ , se busca hacer inferencia de los parámetros  $\theta = (g, b)$ . La ciencia actual ya da por conocida con buen grado de exactitud la aceleración de la gravedad  $g$ . Sin embargo, vamos a considerar que desconocemos su valor, lo que puede aplicar para constantes en alturas distintas a la superficie terrestre o constantes gravitacionales en otro planeta. Con fines ilustrativos esbozamos en la Fig 3.1 varias trayectorias según sus parámetros  $(g, b)$  y notamos en todas un carácter creciente en el tiempo.

### 3.1.2. Modelo Logístico

Un modelo de crecimiento poblacional simple caracterizado por una ecuación diferencial es aquel que la tasa de cambio de la población es proporcional al tamaño actual de la población. Sea  $P(t)$  el tamaño de la población al tiempo  $t$ . El modelo

descrito es

$$\frac{dP}{dt} = \lambda P(t), \quad (3.8)$$

donde  $\lambda$  es la tasa de crecimiento donde además se tiene la condición  $P(0) = P_0$ .

La solución explicita al modelo en (3.8) es

$$P(t) = P_0 e^{\lambda t},$$

que es una función creciente en el tiempo. Sin embargo, este modelo solo es valioso para intervalos de tiempo pequeños, ya que la tasa de crecimiento de una poblacional real tiende a decaer a medida que se consumen los recursos para su sustento.

Para saldar con la dificultad planteada, podemos modelar la tasa de crecimiento de una población considerando tanto al tamaño de la población como a los recursos aún presentes. Consideremos a  $K$  como la capacidad de sustento, que se puede interpretar como la cantidad máxima de población que los actuales recursos permiten mantener, entonces esperamos que a medida que crece la población, la tasa de crecimiento de la misma decrezca.

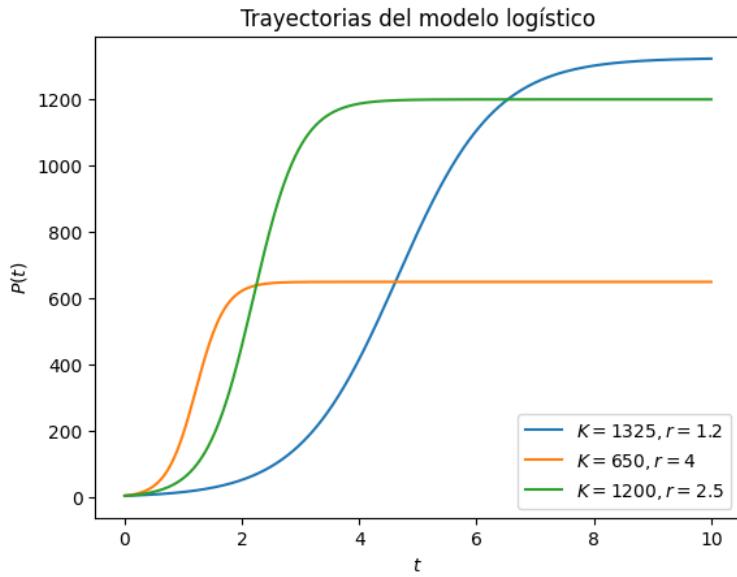
El modelo que se propone es conocido como el modelo de crecimiento logístico. Este nos dice que el cambio de la población es proporcional al tamaño de la población misma así como a la proporción de recursos disponibles. Dicho modelo se puede escribir con la ecuación de Verhulst ([Zill, Cullen, Hernández, y López, 2002](#))

$$\frac{dP}{dt} = rP \left(1 - \frac{P}{K}\right), \quad (3.9)$$

donde  $r$  es la tasa de crecimiento y  $K$  la capacidad de sustento.

A pesar de que la ecuación diferencial del modelo de crecimiento logístico es no lineal tiene solución explicita. Notemos que la EDO pertenece a la familia de las EDO de Bernoulli ([Apostol, 1991](#)), esto es tiene la forma

$$\frac{dy}{dt} + P(x)y = Q(x)y^n,$$



**Figura 3.2:** Varias trayectorias con la dinámica de crecimiento logístico para varias valores de los parámetros  $K, r$ .

y además puede transformar a una EDO lineal de primer orden con el cambio de variable  $u(x) = y^{1-n}$ .

Por tanto, del cambio  $u(t) = P^{-1}$  obtenemos la solución explícita a (3.9)

$$P(t) = \frac{KP_0e^{rt}}{K + P_0(e^{rt} - 1)}, \quad (3.10)$$

donde verificamos que  $\lim_{t \rightarrow \infty} = K$ .

Existe otra reparametrización del modelo logístico que simplifica los cálculos en la metodología del problema inverso. Escribiremos que  $X(t)$  sigue el crecimiento logístico si satisface

$$\frac{dX(t)}{dt} = \theta_1 X(t) (\theta_2 - X(t)). \quad (3.11)$$

Con la parametrización dada por

$$\theta_1 = \frac{r}{K}, \quad \theta_2 = K,$$

recuperamos la expresión dada en (3.9).

En la Fig 3.2 se muestran varias trayectorias del crecimiento poblacional con diferentes parámetros usando la parametrización interpretable en términos de tasa de crecimiento  $r$  y poblacional máxima  $K$ . Además, vemos que cada trayectoria es no decreciente y tiende asintóticamente al parámetro  $K$ .

### 3.1.3. Modelo SIR

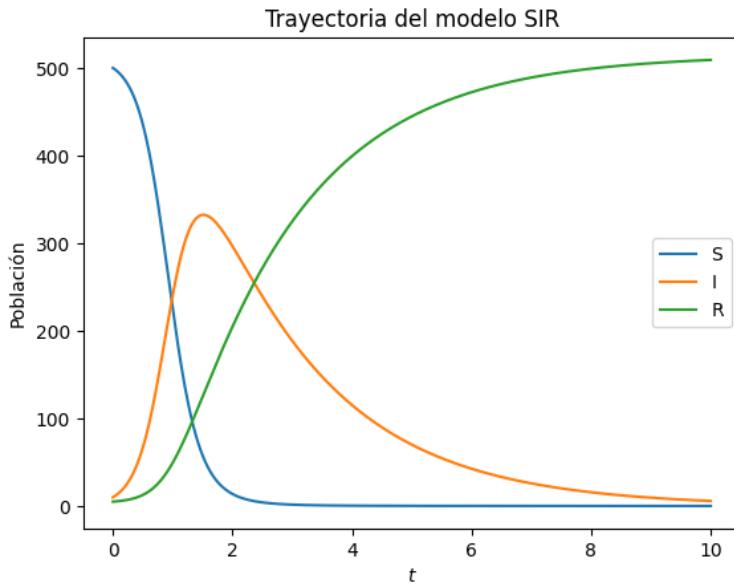
Los modelos epidemiológicos clasifican a la población en clases que distinguen si han tenido o no cierta enfermedad. Un modelo que captura con buena aproximación a varias enfermedades es el modelo SIR. Este propone dividir a las  $N$  personas de una población en tres grupos:

1. Susceptible ( $S$ ): Cantidad de personas que no han tenido la enfermedad ni están enfermas del brote de interés.
2. Infectoso ( $I$ ): Son las personas que actualmente portan el patógeno y son además un vector de transmisión.
3. Recuperado ( $R$ ): Para aquellas personas que se han recuperado de la enfermedad y ya tienen inmunidad o también para aquellas que perecieron.

Así, el modelo propuesto supone que la tasa de cambios de las personas susceptibles decrece proporcionalmente a la cantidad de infectados y la cantidad de susceptibles restantes. De igual forma, la cantidad de recuperados crece con una tasa proporcional a la cantidad de infectados. La dinámica del modelo SIR se da con el siguiente sistema de EDO's

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI, \\ \frac{dI}{dt} &= \beta SI - \gamma I, \\ \frac{dR}{dt} &= \gamma I, \end{aligned} \tag{3.12}$$

donde  $\beta$  es la tasa de infección,  $\gamma$  es la tasa de recuperación y  $S + I + R = N$  (Weiss, 2013).



**Figura 3.3:** Trayectoria de cada grupo del modelo SIR con  $\beta = 0.009$  y  $\gamma = 0.5$ .

La solución explícita para  $S(t)$ ,  $I(t)$ ,  $R(t)$  no existe en general, por ello se requieren métodos numéricos (Mathews y Fink, 2000) para resolver numéricamente para cada grupo y poder construir el forward map . La implementación de python usando la paquetería odeint se puede ver en Jiménez Bedolla (2022) y en Griffiths y Higham (2010). En la Fig. 3.3 tenemos un ejemplo para un solo caso del modelo SIR con tasa de infección  $\beta = 0.009$  y tasa de recuperación  $\gamma = 0.5$

## 3.2. Enfoque bayesiano al problema inverso

Una vez establecidos los modelos sobre los que trabajaremos, nos enfocamos en abordar el problema inverso para los mismos. Para ello, es necesario tener a disposición una muestra de la trayectoria del modelo al cual deseamos inferir sus parámetros.

Según sea el caso, podemos realizar mediciones de caída de un objeto para ciertos tiempos, contar la cantidad de población para cada intervalo de tiempo, o observar los grupos de personas en un brote de cierto patógeno. Denotemos estas observaciones por  $y_1, y_2, \dots, y_n$ .

Sin embargo, no es fácil tener acceso a dicha muestra, por lo que optaremos por simular los datos directamente del modelo seleccionando los parámetros convenientes.

temente y además agregar un ruido gaussiano. Es decir, tomamos  $\theta \in \Theta$  fijo, luego consideramos los tiempos a los cuales corresponderá la muestra, denotados por  $t_1, t_2, \dots, t_n$ . Posteriormente, con el forward map ordinario  $F_\theta$  (sin aproximación por vecinos cercanos) generado según el modelo (ya sea analítico o numérico) evaluamos para cada tiempo (denotado por  $F_\theta(t_i)$ ) y agregamos un ruido  $\varepsilon_i \sim N(0, \sigma^2)$ . Así la muestra considerada se constituye por

$$y_i = F_\theta(t_i) + \varepsilon_i, \quad (3.13)$$

para  $i = 1, 2, \dots, n$ . Dicho de otra forma la muestra contemplada es

$$\{(y_1, t_1), (y_2, t_2), \dots, (y_n, t_n)\}, \quad (3.14)$$

un conjunto de  $n$  tuplas.

Una vez establecida la muestra  $y_i$ , ahora el propósito es abordar el problema inverso, esto es, buscamos  $\theta$  tal que  $y_i \approx F_\theta(t_i)$ . Ya hemos estudiado el enfoque bayesiano al problema inverso en el capítulo 2, desde luego la implementación obtenida aquí sera inicialmente el caso de forward map ordinario. Más adelante estudiaremos el forward map aproximado.

### 3.2.1. Distribución posterior

En la teoría de la estadística bayesiana, el paradigma central se basa en obtener la distribución posterior de los parámetros de interés dada ciertas observaciones a las que hemos denotamos por  $\mathbf{y} = (y_1, y_2, \dots, y_n)$ . El vector de parámetros es  $\theta = (\theta_1, \theta_2, \dots, \theta_d) \in \mathbb{R}^d$ . Recordemos de (2.4) que la distribución posterior satisface la ecuación

$$\pi(\theta|\mathbf{y}) \propto f(\mathbf{y}|\theta)\pi(\theta), \quad (3.15)$$

definida salvo una constante en términos de la verosimilitud y la distribución a priori.

El análisis del calculo de la verosimilitud se trató en la sección 2 del capítulo 2.

De la ecuación (2.3) junto con (3.15)

$$\pi(\theta|\mathbf{y}) \propto \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - F_\theta(t_i))^2 \right\} \pi(\theta), \quad (3.16)$$

rescatamos la distribución posterior en función del forward map, salvo una constante. De esta forma, resta discutir las cualidades deseadas para los modelos anteriormente descritos.

Para nuestro caso particular, todos los modelos propuestos en la sección precedente dependen de parámetros no negativos, es decir,  $\theta_i \geq 0$  para toda  $i = 1, \dots, d$  (inclusive condiciones iniciales). Incluso algunos de ellos se tiene bastante certeza de su valor, que podría pensarse en ajustar una distribución a priori normal. Sin embargo, ajustar una normal conlleva a tener problemas en el soporte, por lo que puede pensarse en considerar una normal truncada a los positivos, sin embargo está complica innecesariamente las expresiones para la distribución posterior.

Finalmente, se opta por proponer distribuciones a priori gamma, debido a que existe una parametrización en la cual se asemeja a una distribución normal pero con el soporte adecuado. La distribución gamma que tiene por máximo a  $\theta_i^*$  y asemeja una normal, aunque asimétrica, es

$$\theta_i \sim \text{Gamma} \left( \alpha_i, \frac{\theta_i^*}{\alpha_i} \right), \quad (3.17)$$

con  $\theta_i^*$  y  $\alpha_i$  parámetros conocidos. Vea por ejemplo la Fig. 3.4.

La parametrización de la  $\text{Gamma}(\alpha, \beta)$  es tal que su función de densidad sea

$$f(\theta|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp \{-\beta\theta\},$$

con  $\alpha, \beta$  conocidos.

Como  $\theta_i$  se proponen independientes, entonces la distribución a priori para  $\theta$  es

$$\begin{aligned}\pi(\theta|\alpha) &= \pi(\theta_1|\alpha_1) \cdots \pi(\theta_d|\alpha_d) \\ &= \prod_{i=1}^n \left[ \frac{1}{\Gamma(\alpha_i)} \left( \frac{\theta_i^*}{\alpha_i} \right)^{\alpha_i} \theta_i^{\alpha_i-1} \exp \left\{ -\frac{\theta_i^*}{\alpha_i} \theta_i \right\} \right],\end{aligned}\quad (3.18)$$

desarrollando

$$\begin{aligned}\pi(\theta|\alpha) &= \\ &\frac{1}{\Gamma(\alpha_1)} \left( \frac{\theta_1^*}{\alpha_1} \right)^{\alpha_1} \theta_1^{\alpha_1-1} \exp \left\{ -\frac{\theta_1^*}{\alpha_1} \theta_1 \right\} \cdots \frac{1}{\Gamma(\alpha_d)} \left( \frac{\theta_d^*}{\alpha_d} \right)^{\alpha_d} \theta_d^{\alpha_d-1} \exp \left\{ -\frac{\theta_d^*}{\alpha_d} \theta_d \right\},\end{aligned}\quad (3.19)$$

con  $\alpha = (\alpha_1, \dots, \alpha_p)$  y  $\theta^* = (\theta_1^*, \dots, \theta_d^*)$  conocidos.

Sustituyendo (3.18) en (3.16) se tiene la forma funcional de la distribución posterior

$$\begin{aligned}\pi(\theta, \sigma|\mathbf{y}) \propto \\ \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - F_\theta(t_i))^2 \right\} \prod_{i=1}^n \left[ \frac{1}{\Gamma(\alpha_i)} \left( \frac{\theta_i^*}{\alpha_i} \right)^{\alpha_i} \theta_i^{\alpha_i-1} \exp \left\{ -\frac{\theta_i^*}{\alpha_i} \theta_i \right\} \right]\end{aligned}\quad (3.20)$$

Notemos que la distribución en (3.20) no pertenece a una familia conocida. Para lidiar con este problema usamos métodos Monte Carlo. Más precisamente, podemos simular de la distribución posterior (3.20).

### 3.2.2. Simulación con forward map ordinario

Al incursionar a la estadística bayesiana se aprende que se puede obtener la distribución posterior por medio del famoso análisis conjugado. Es decir se propone una distribución a priori convenientemente para que la distribución posterior pertenezca a la misma familia. Sin embargo, el análisis conjugado puede ser no trivial o incluso no existir. Por ello es que simular de la distribución posterior con MCMC (Markov Chain Monte Carlo) es la vía necesaria en muchos casos para tener una muestra

$(X_i, i \in \{1, \dots, T\})$  de dicha distribución, con la que podemos estimar cualquier momento de la distribución. Si estamos interesados en obtener la media de  $X$ , en beneficio de la ley de grandes números (Van der Vaart, 2000), podemos estimar como

$$\mathbb{E}[X] \approx \frac{1}{T} \sum_{i=1}^T X_i \quad (3.21)$$

De igual forma si estamos interesados en la media del estadístico  $h(X)$  para un función  $h$  arbitraria, se estima por

$$\mathbb{E}[h(X)] \approx \frac{1}{T} \sum_{i=1}^T h(X_i) \quad (3.22)$$

donde  $X_i$  es la  $i$ -ésima simulación de la distribución posterior por Metropolis-Hastings.

La distribución objetivo que se implementa en el MCMC es la distribución posterior (3.20). Recordemos que la distribución posterior depende del forward map  $F_\theta(t_i)$  obtenido de evaluar los tiempos de la muestra. Recordemos que  $F_\theta(t_i)$  es la versión discretizada de  $F_\theta$ . Sea  $\theta \in \mathbb{R}^d$  la función

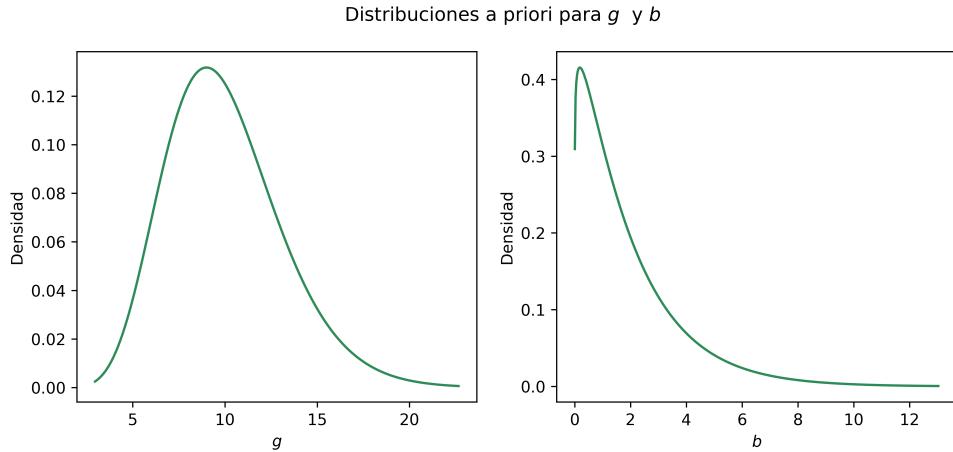
$$\theta \mapsto F_\theta = x(t) \quad (3.23)$$

tal que  $x(t)$  es la solución de la ecuación diferencial ordinaria  $x(t) = G(t, x(t), x'(t), x''(t), \dots)$  del modelo bajo consideración.

### Simulación del modelo gravitatorio

Para poder hacer simulaciones, es necesario determinar completamente la distribución posterior. El Forward map para el modelo gravitatorio es  $F_\theta = F(g, b) = x(t)$  con la ecuación diferencial de la dinámica gravitacional dada en (3.2) o equivalentemente en (3.7), con  $x_0 = 0$  y  $v_0 = 0$  las condiciones iniciales conocidas.

Además, es necesario saber como abordar el caso de la desviación estándar  $\sigma$ . En esencia existen dos formas, pensar en  $\sigma$  como parámetro conocido, por lo que no es necesario hacer inferencia, manteniendo la distribución posterior tal como se



**Figura 3.4:** Distribuciones marginal a priori para el parámetro  $g$  (izquierda) y  $b$  (derecha).

ha detallado. En el otro caso es para  $\sigma$  desconocida, al no tener certeza de su valor esta es agregada a los parámetros a inferir. Siguiendo el paradigma bayesiano, se establece una distribución a priori para  $\sigma$  independiente de la distribución a priori para  $\theta$ . En otras palabras, ahora la distribución posterior depende de  $d+1$  parámetros. Sin embargo para fines de esta tesis, nos abstendremos de los casos en los que  $\sigma$  es conocida.

De esta manera, la distribución posterior (3.20) tendrá como variables únicamente al vector de parámetros  $\theta$ , al ser una EDO dependiente solamente de  $g$  y  $b$  se sigue que  $\theta = (g, b)$  siendo un vector en  $\mathbb{R}^d$  con  $d = 2$ . Además es necesario definir las constantes para las distribuciones a priori gamma.

Consideremos las distribuciones a priori (3.17) como

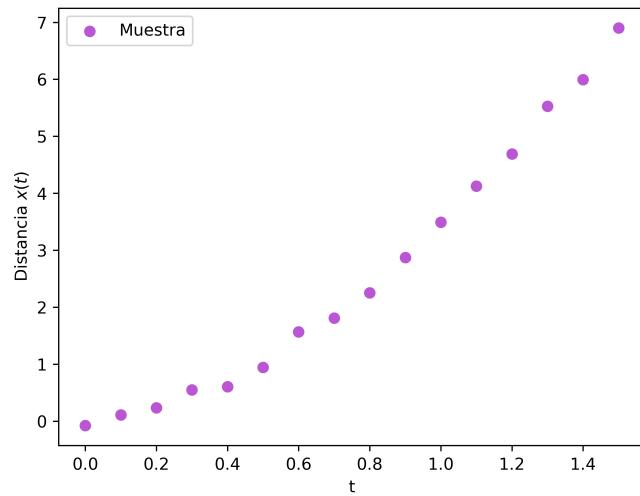
$$g \sim \text{Gamma} \left( \alpha, \frac{\theta_1^*}{\alpha} \right) \quad (3.24)$$

$$b \sim \text{Gamma} \left( \beta, \frac{\theta_2^*}{\beta} \right) \quad (3.25)$$

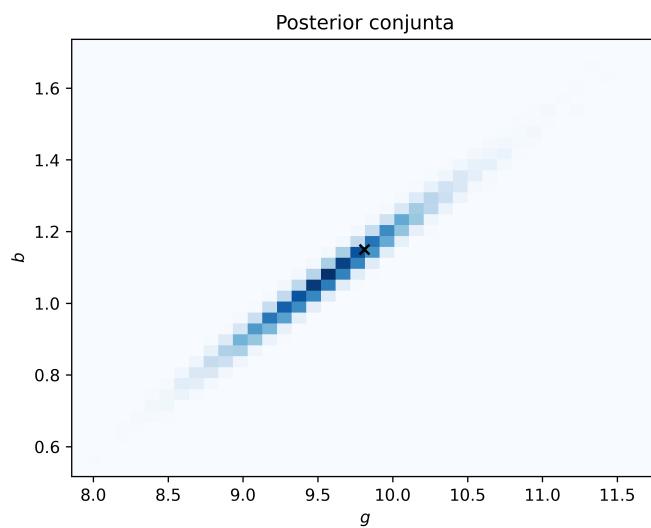
con  $\alpha = 10$ ,  $\beta = 1.1$  con máximo en  $\theta_1^* = 10$  y  $\theta_2^* = 2$ . Podemos observar las distribuciones a priori en la Fig. 3.4.

Ahora, generamos una muestra  $\mathbf{y}$  del forward map gravitatorio y agregamos un ruido gaussiano con  $\sigma = 0.1$  obteniendo la muestra como en la Fig. 3.5.

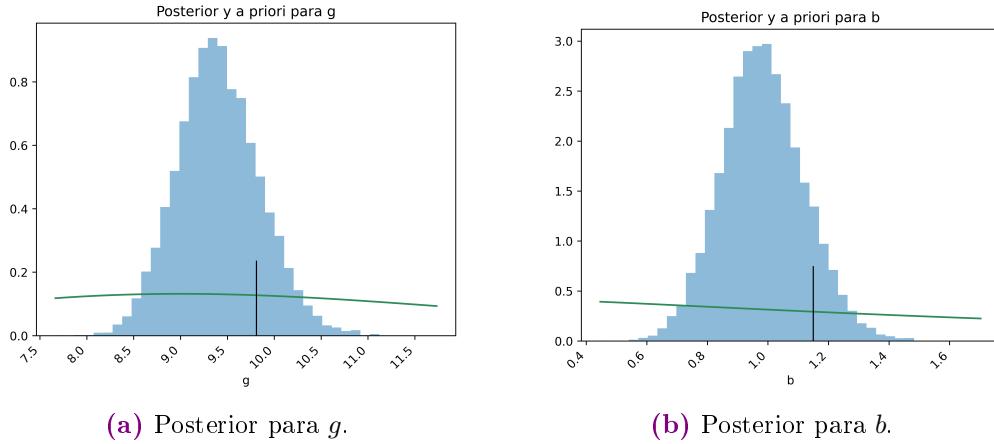
Luego, con la implementación de la metodología propuesta (Kaipio y Somersalo,



**Figura 3.5:** Muestra  $\mathbf{y}$  del modelo gravitatorio.



**Figura 3.6:** Estimación de la distribución posterior conjunta por método MCMC Metropolis-Hastings.

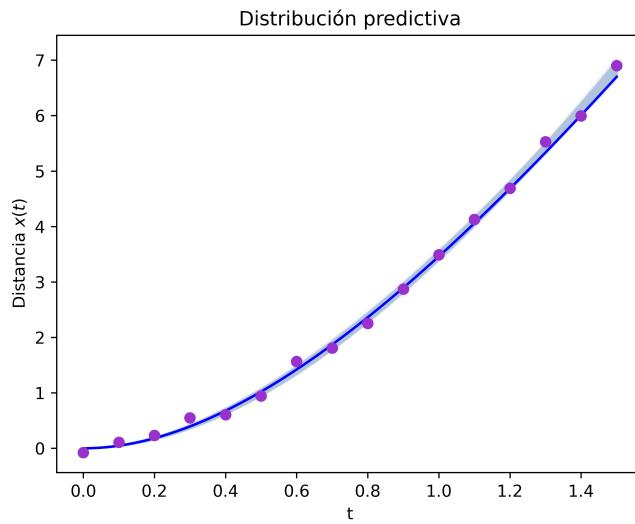


**Figura 3.7:** Distribución posterior (azul) y distribución a priori (verde) para el parámetros  $g$  y  $b$  del modelo gravitatorio. Los valores verdaderos se marcan en negro.

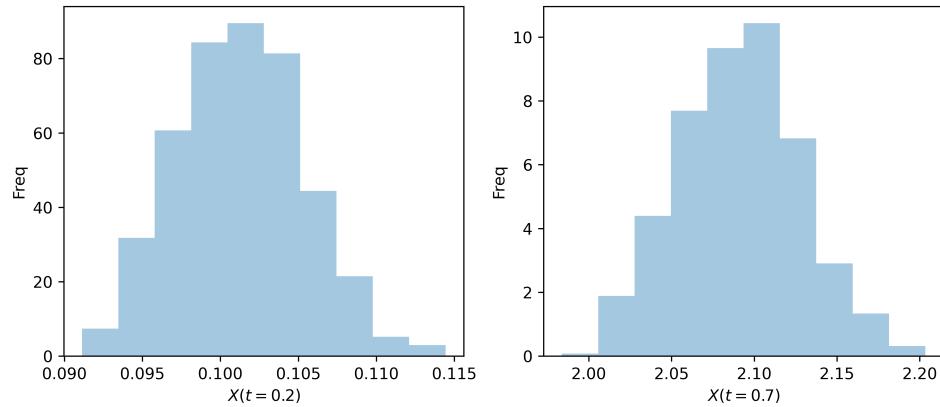
2006), usando el algoritmo MCMC Metropolis-Hastings t-walk (Christen y Fox, 2010) para una cadena de tamaño  $T = 600,000$  y un burn in de 20,000 tenemos que la trayectoria seguida en el espacio de parámetros  $\theta = (g, b)$  para la distribución posterior (conjunta) es el mostrado en la Fig. 3.6. Notemos que dicha posterior tiene alta correlación entre parámetros. Esto tiene la interpretación física que para valores altos de  $g$  es necesario tener más fricción  $b$  con el medio para la misma trayectoria y.

Es de interés las marginales de la distribución posterior. Dado que la simulación de la cadena es bidimensional, de (3.22) tomamos  $h_1(X_1, X_2) = 1_{\{X_1\}}$  y  $h_2(X_1, X_2) = 1_{\{X_2\}}$ , dicho de otra forma, simplemente tomamos el histograma sobre una componente de la cadena dada por MCMC. Podemos ver las estimaciones a las distribuciones marginales posteriores para los parámetros  $g$  y  $b$  en Fig. 3.7a y Fig. 3.7b

Finalmente, es de interés observar la distribución que puede generar las curvas predictivas. Es decir, de la cadena  $X_t$  fruto de MCMC, tomamos submuestreos en intervalos  $X_{[iN, (i+1)N]}$  para tomar estimaciones puntuales (media) de  $g$  y  $b$  y así aplicar el forward map para cada estimación. Dicho de otra forma, la información obtenida de la distribución posterior se desea transmitir directamente a la trayectoria, generando así distribuciones sobre las mismas.



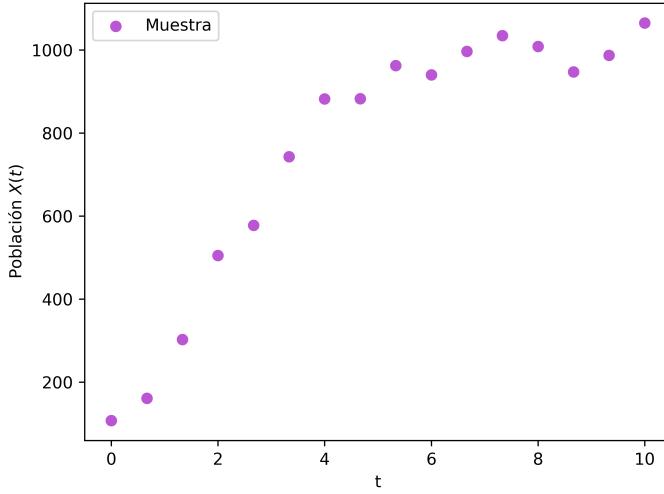
**Figura 3.8:** Distribución predictiva para el modelo gravitatorio



**Figura 3.9:** Distribución predictiva del modelo gravitatorio para los valores de tiempo de 0.2s y 0.7s.

En la Fig. 3.8 se muestra en azul claro los distintos forward map  $F_\theta$  con varias realizaciones de  $\theta|y$  dada por la estimación de la distribución posterior por Metropolis-Hastings, mientras que en azul oscuro se muestra el forward map de partida, es decir, de donde se realizaron las simulaciones para generar  $y$  (morado). Por tanto, notamos que de cierta forma, la información de la distribución posterior contiene a la simulación inicial, es decir al valor verdadero.

Más aún, para ejemplificar la distribución predictiva, considere dos valores fijos de  $t$  y tómese cortes verticales de la distribución predictiva. En la Fig. 3.9 nos muestra



**Figura 3.10:** Muestra y para el modelo logístico.

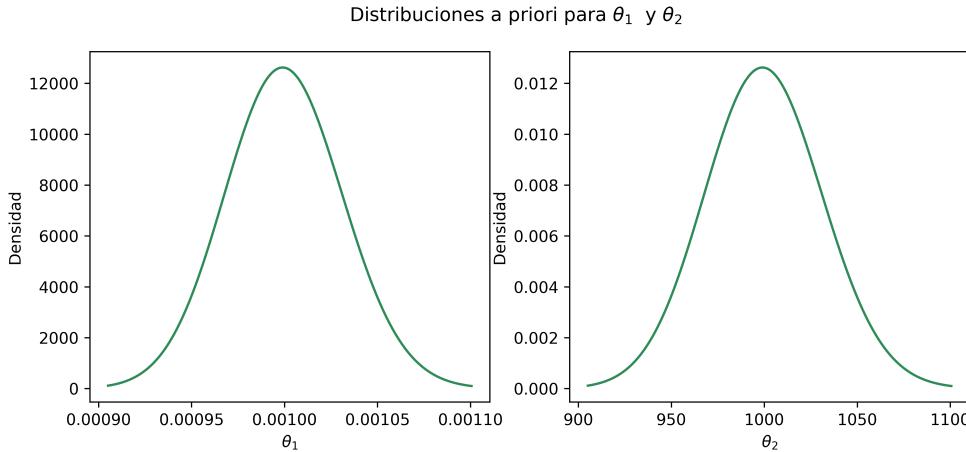
para el tiempo  $t = 0.2s$  y  $t = 0.7s$  la distribución de probabilidad de la posición de la partícula afectada por la gravedad en ese momento de tiempo.

### Simulación del modelo logístico

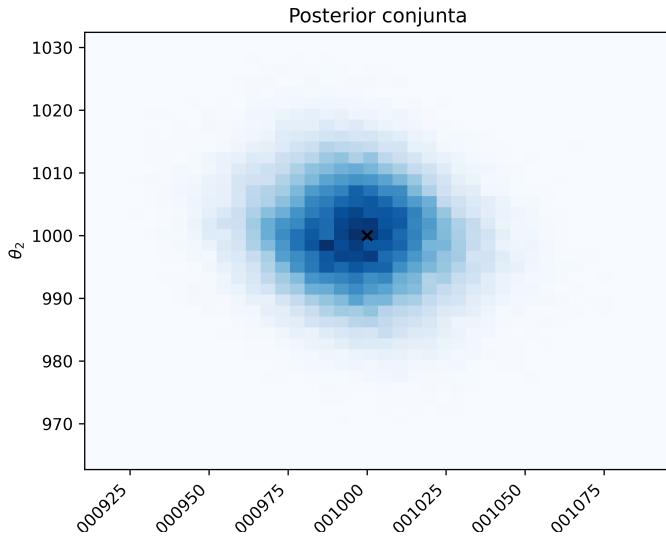
En el modelo gravitacional tenemos un ejemplo de simulación del problema inverso donde las unidades de los parámetros no difieren significativamente en ordenes de magnitud. En secciones posteriores se explica el concepto de regularización de las unidades, donde la metodología aproximada para modelos cuyos ordenes de magnitud difieren significativamente tienen sesgo en la estimación de la distribución posterior. Dicho lo anterior, siguiendo la metodología ordinaria para la resolución del problema inverso en el modelo logístico, con la base de datos simuladas del mismo considerando  $\theta_1 = 0.001$  y  $\theta_2 = 1000$  con un ruido gaussiano aditivo de desviación estándar  $\sigma = 30$  graficado en la Fig. 3.10.

Posteriormente, definamos las constantes de las distribuciones gamma a priori en (3.17). Seleccionemos las constantes semejantes a los valores verdaderos, cuyas gráficas se muestran en la Fig. 3.11

De esta forma, análogo al modelo gravitacional, generamos una cadena con la implementación del algoritmo Metropolis-Hastings dado por Christen y Fox (2010) con



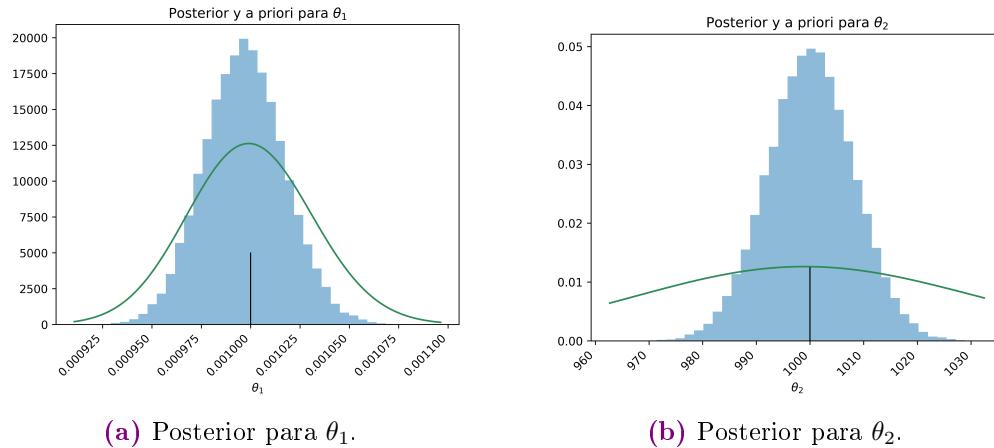
**Figura 3.11:** Distribuciones a priori para  $\theta_1$  y  $\theta_2$



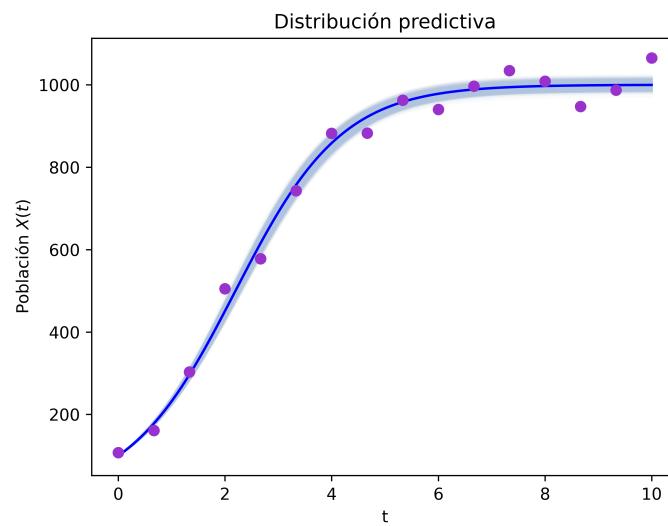
**Figura 3.12:** Posterior conjunta para el modelo logístico.

$T = 600,000$  iteraciones y un burn in de 20,000 estados. La distribución posterior, Fig. 3.12, muestra una forma acampanada ligeramente asimétrica con una sutil correlación negativa entre parámetros. Las distribución marginal para  $\theta_1$  y  $\theta_2$  se muestran en la Fig. 3.13a y 3.13b.

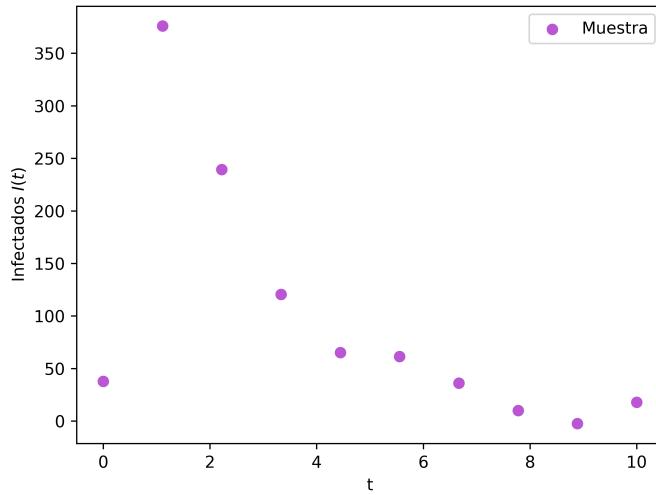
Finalmente, el análisis concluye con la distribución predictiva dada por el forward map tomando estimaciones puntuales de la distribución posterior, en la Fig. 3.14 notamos que la distribución tiene una varianza mayor que en el modelo gravitacional, además de observar que la trayectoria de partida (azul oscuro) se encuentra incluida en la predicción dada por el problema inverso.



**Figura 3.13:** Distribución posterior (azul) y distribución a priori (verde) para el parámetros  $\theta_1$  y  $\theta_2$  del modelo logístico. Los valores verdaderos se marcan en negro.



**Figura 3.14:** Distribución predictiva para el modelo logístico.

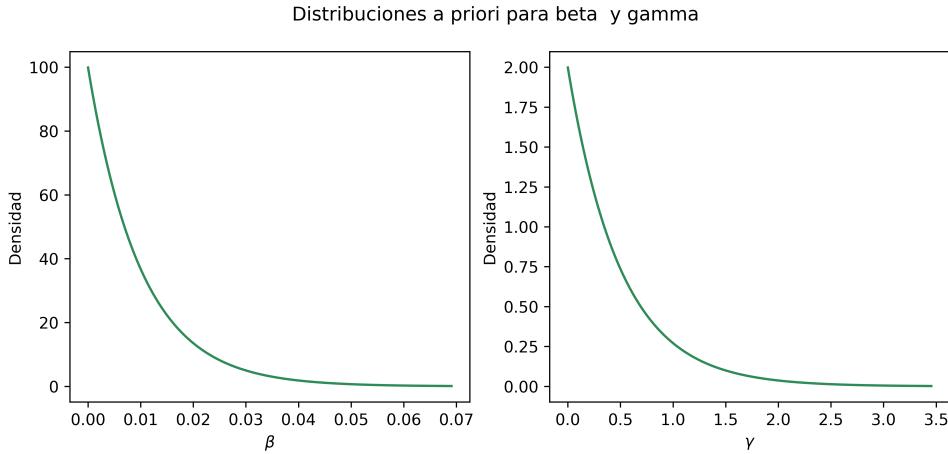


**Figura 3.15:** Muestra y del modelo SIR.

### Simulación del modelo SIR

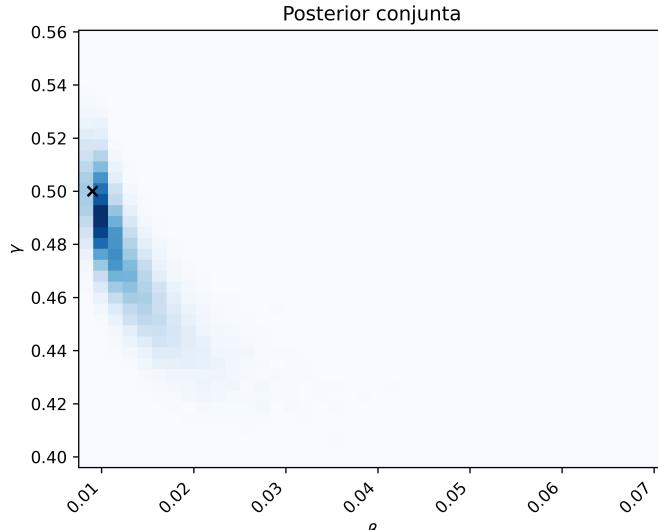
Para el modelo epidemiológico tomamos una muestra de  $n = 10$  observaciones de infectados en una población cerrada a lo largo del tiempo. Al igual que los modelos previos, dichas observaciones son simulaciones del forward map con ruido gaussiano. Tomando una muestra de  $n = 10$  con parámetros  $\beta = 0.009$ ,  $\gamma = 0.5$  y ruido con desviación estándar de  $\sigma = 10$  y dando las condiciones iniciales por conocidas con  $I(0) = 50$ ,  $S(0) = 500$ ,  $R(0) = 5$ . En la Fig. 3.15 vemos la gráfica de la muestra de infectados, notamos que tienen un rápido crecimiento para después decaer suavemente.

Para abordar el problema inverso establezcamos los parámetros que se desea hacer inferencia. El espacio paramétrico se conforma por  $\theta = (\beta, \gamma)$  donde  $\beta > 0$  es la tasa de infección y  $\gamma > 0$  es la tasa de recuperación dados en (3.12). Nuevamente como se trata de parámetros positivos usamos la distribución a priori (3.17). Podemos rescatar un conocimiento a priori sobre su distribución previo a los datos (Weiss, 2013). La elección considera para las distribuciones gamma a priori son mostradas en la Fig. 3.16.



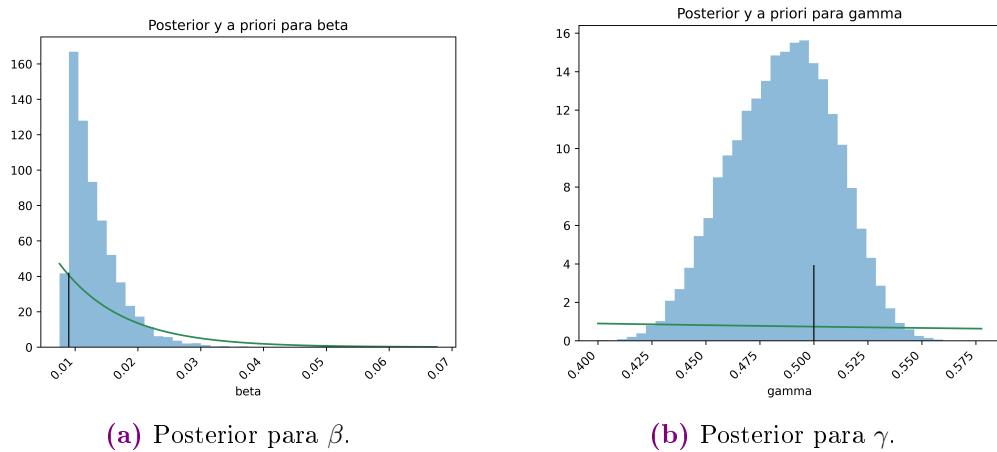
**Figura 3.16:** Distribuciones marginal a priori para el parámetro  $\beta$  (izquierda) y  $\gamma$  (derecha).

Análogamente a los modelos previos, una vez obtenida la forma funcional de la distribución posterior, usando el algoritmo Metropolis-Hastings para una cadena de tamaño  $T = 600,000$ . La distribución posterior dada se muestra en la Fig. 3.17 que muestra una forma de media luna concentrada en el borde izquierdo.



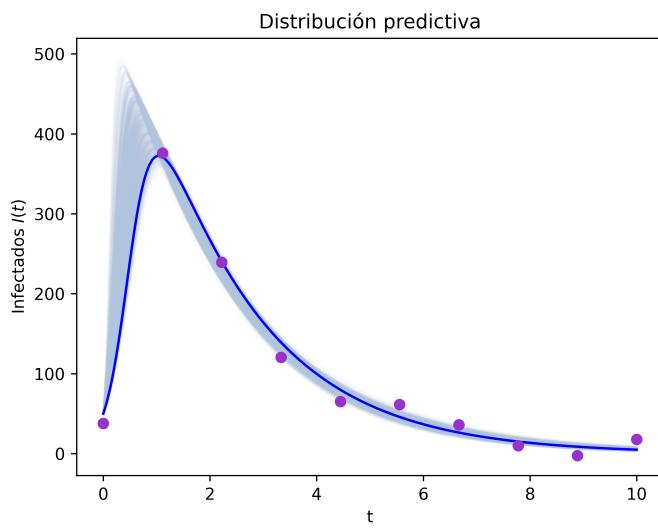
**Figura 3.17:** Distribución posterior conjunta para el modelo SIR.

Para observar la distribución para cada parámetro se toma la distribución marginal, respectivamente. Observemos de la Fig. 3.18a y Fig. 3.18b la distribución de  $\beta$  se aglomera en valores cercanos a cero, mientras que para  $\gamma$  tenemos bastante diferencia respecto a la distribución a priori.



**Figura 3.18:** Distribución posterior (azul) y distribución a priori (verde) para el parámetros  $\beta$  y  $\gamma$  del modelo SIR. Los valores verdaderos se marcan en negro.

Tomando realizaciones  $\theta_j = (\beta_j, \gamma_j)$  de la distribución posterior para luego calcular  $F(\theta_j)$  nos da la distribución vista para la cantidad de infectados en tiempo arbitrario dado por el modelo SIR. Tal gráfico dado en Fig. 3.19 vemos que el modelo (azul fuerte) se encuentra incluido en las distribuciones de infectados.



**Figura 3.19:** Distribución predictiva para el modelo SIR.

### 3.3. Simulación con forward map aproximado

Es fundamental en este proyecto contrastar las distribuciones posteriores fruto de la resolución del problema inverso con forward map ordinario contra aquellas distribuciones posteriores dadas por el forward map aproximado. En esta sección se presentan los resultados del análisis aproximado así como su correspondiente comparación con su semejante ordinario para cada modelo previamente discutido.

Recordemos de la sección 4 del capítulo 2 la construcción del forward map aproximado. Consideremos una partición tipo malla  $\mathcal{M}$ . Sea  $\vartheta \in \mathcal{M}$  un elemento del subconjunto del espacio de parámetros  $\Theta$ . Sea  $\vartheta^{(1)}, \dots, \vartheta^{(k)}$  los  $k$  vecinos mas cercanos a  $\theta$ . Para  $\theta \notin \mathcal{M}$  el forward map aproximado a  $k$  vecinos cercanos de definición (o resolución)  $M$  es:

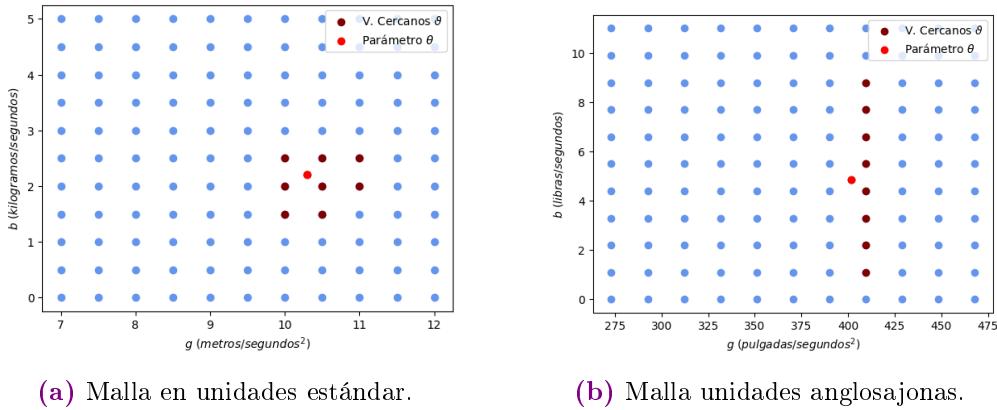
$$\tilde{F}_M^k(\theta) = \sum_{j=1}^k \omega_j(\theta) F(\vartheta^{(j)}) \quad (3.26)$$

donde  $\omega_j(\theta) = d_j(\theta)^{-1} / \sum_{i=1}^k d_i(\theta)^{-1}$  con  $d_j = d(\theta, \vartheta^{(j)})$ . Notemos que el forward map aproximado usa el forward map ordinario  $F(\vartheta)$  para los elementos de la malla  $\mathcal{M}$ . Así la construcción de la distribución posterior aproximada  $\pi_M^k$  se da reemplazando el forward map ordinario en la distribución posterior de los parámetros, tal como en la ecuación (2.6).

Previo a realizar las simulaciones para distintos casos es necesario revisar un último concepto. Proceder a estimar las distribuciones posteriores aproximadas, de la forma detallada a lo largo de este texto, tendría un inconveniente al analizar la convergencia a mediada que se aumenta la resolución de la malla. Este problema se salda con la regularización de las unidades.

#### 3.3.1. Regularización de las unidades

Un paso previo a la obtención del forward map aproximado es la regularización de las unidades. La regularización de las unidades es una aporte fundamental en la metodología del enfoque bayesiano del problema inverso con el forward map apro-



**Figura 3.20:** Malla para el modelo gravitatorio (azul) con los 8 vecinos más cercanos al parámetro  $\theta$  (rojo) con unidades estándar (izquierda) y unidades anglosajonas (derecha).

ximado. Veremos más adelante que según el modelo, las distribuciones posteriores aproximadas son sensibles a las unidades de medida de las observables. Esto significa que el método aproximado propuesto confronta una laguna estructural que amerita rectificación. Para ver esto, observe la ecuación del forward map aproximado (3.26) depende de los  $k$  vecinos más cercanos  $\vartheta^{(j)}$  del parámetro  $\theta = (\theta_1, \dots, \theta_d) \in \Theta$ . Dado que siempre existe reparametrización del modelo, entonces las unidades de  $\theta_1$  y  $\theta_2$ , por ejemplo, se pueden alterar a gusto, lo que termina afectando las distancias y modificando a otros vecinos cercanos.

Concisamente, las unidades del modelo gravitatorio para  $g$  usualmente se dan en *metros/segundo* y para  $b$  en *kilogramos/segundo*. Sin embargo, piense en la reparametrización donde se contemplan kilómetros en lugar de metros y gramos en lugar de kilogramos, pues la distancia euclíadiana resultará en un incremento en tres ordenes de magnitud para  $b$  y un decrecimiento en tres ordenes para  $g$ . De forma que los nuevos vecinos más cercanos terminan por ignorar cualquier propuesta que involucre un cambio para  $b$ . Lo mismo pasa en el caso de que la parametrización de cada parámetro difiera en más de un orden de magnitud.

De igual forma, la aproximación del forward map en el caso del modelo gravitacional esta condicionada a las unidades de medida, veamos el caso concreto de usar unidades estándar (Sistema Internacional) y el caso de unidades anglosajonas. En la Fig. 3.20 se muestra un enmallado para el modelo gravitatorio en unidades estándar

3.20a, mientras que en 3.20b se toma la reparametrización en libras y pulgadas. Nótese que los vecinos cercanos (rojo oscuro) al parámetro  $\theta$  (rojo) se ven ampliamente alterados. En consecuencia, es necesario regular las unidades de forma que el uso de cualquier sistema de unidades sea irrelevante para la aproximación del forward map.

### 3.3.2. Construcción de la regularización de unidades

Como hemos visto, la amplia diferencia en ordenes de magnitud entre parámetros conduce a una construcción patógena del forward map. Por esto mismo, la regularización de las unidades se establece como una técnica para normalizar cada parámetro al intervalo  $[0, 1]$ .

Recordemos la notación usada en la sección 2.4. Sea  $\theta = (\theta_1, \dots, \theta_d) \in \Theta$ . Sea  $\mathcal{M}$  una malla de  $\Theta$ , de forma que  $\mathcal{M} = \mathcal{M}_1 \times \dots \times \mathcal{M}_d$ , donde  $\mathcal{M}_i = \{\theta_i^{Min}, \theta_i^{(2)}, \dots, \theta_i^{(M-1)}, \theta_i^{Max}\}$ . Para cada componente de  $\theta$ , aplicamos la transformación de normalización  $\varphi$  a la malla  $\mathcal{M}$ , donde a cada elemento  $\theta_i^{(j)} \in \mathcal{M}_i$  se aplica la transformación

$$\varphi_i^{(j)} = \frac{\theta_i^{(j)} - \theta_i^{Min}}{\theta_i^{Max} - \theta_i^{Min}}. \quad (3.27)$$

Considerando que  $\mathcal{N}_i = \{\varphi_i^{(1)}, \dots, \varphi_i^{(M)}\}$  para cada  $i \in \{1, \dots, d\}$ . De forma que se crea un enmallado uniforme  $\mathcal{N} = \mathcal{N}_1 \times \dots \times \mathcal{N}_d$ . Con la transformación (3.27) tenemos que el espacio reparametrizado  $\mathcal{N}$  es un subconjunto del espacio  $[0, 1]^d$ , por lo que al usar distancia euclídea los vecinos más cercanos no serán sensibles a reparametrizaciones. En adelante entenderemos a los  $k$  vecinos más cercanos  $v^j$  con la reparametrización  $\phi$  y la malla  $\mathcal{N}$  para la construcción del forward map aproximado en (3.26).

### 3.3.3. Experimentación con distribuciones posteriores aproximadas

Las aproximaciones de la distribución posterior considera aproximaciones del forward map. Estas se hacen con la *interpolación* de  $k$  vecinos más cercanos de la malla

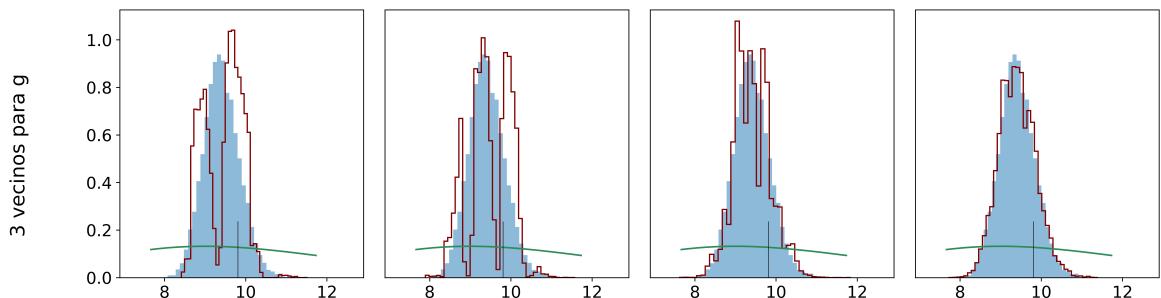
$\mathcal{N}$ . Particularmente se experimentó con vecinos  $k = 3, 5, 8$ . Además, para indagar en la convergencia de la distribución aproximada se toman la resolución de la aproximación de forma creciente. Es decir, se aproxima el forward map con mallas de tamaño  $M \times M$  con  $M = 10, 15, 30, 50$ .

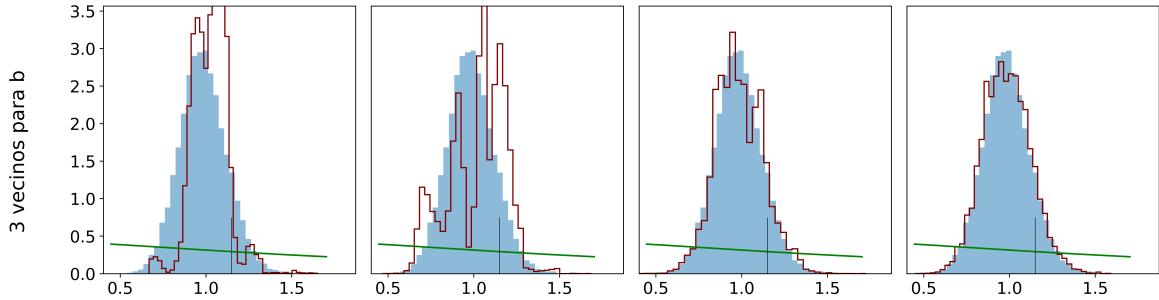
### Forward map aproximado para el modelo gravitatorio

En el análisis bayesiano del problema inverso del modelo gravitatorio se obtuvo la distribución posterior con el forward map ordinario. Esta vez, el análisis bayesiano se hará con el forward map aproximados para cada par  $k$  y  $M$  de las aproximaciones propuestas. Posteriormente, las distribuciones aproximadas  $\tilde{\pi}_M^k(g, b)$  las cotejaremos con su análogo ordinario,  $\pi(g, b)$  de forma que sea visible o no la convergencia al análogo ordinario.

Para empezar, tomemos la aproximación del forward map con 3 vecinos cercanos ( $k = 3$ ). Luego, estimamos la distribución posterior conjunta para cada rejilla ( $M = 10, 15, 30, 50$ ) haciendo MCMC con el t-walk (Christen y Fox, 2010). Posteriormente se toman las distribuciones marginales para cada parámetro  $g, b$ . Al ser un estudio heurístico, el grado de aproximación se hace tras simple comparación de la distribución marginal aproximada para cada parámetro,  $\tilde{\pi}_M^k(g)$  con  $\pi(g)$ , y comparación de  $\tilde{\pi}_M^k(b)$  con  $\pi(b)$ .

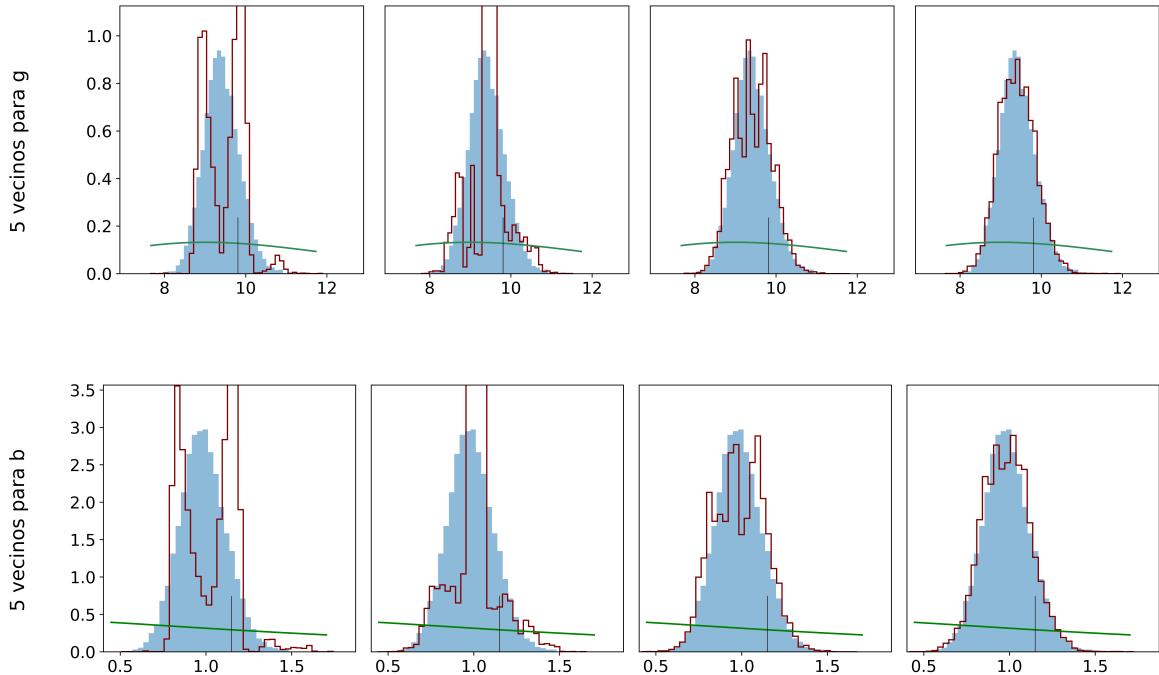
En la Fig. 3.21 se muestran las distribuciones marginales para la aproximación del forward map con  $k = 3$  y de izquierda a derecha una malla de 10x10, 15x15, 30x30, 50x50.





**Figura 3.21:** Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a tres vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo gravitacional.

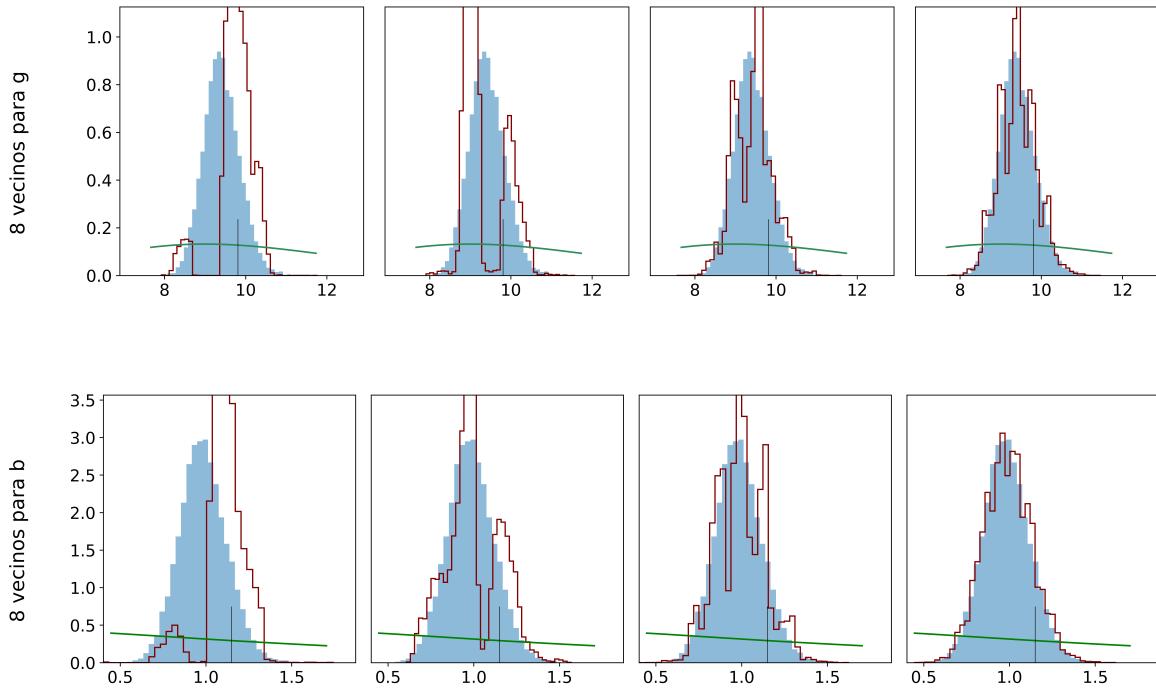
Luego, aumentando la cantidad de vecinos cercanos en la aproximación a  $k = 5$ , podemos ver en la Fig. 3.22 las distribuciones marginales para la aproximación del forward map con  $k = 5$  y de izquierda a derecha una malla de 10x10, 15x15, 30x30, 50x50.



**Figura 3.22:** Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a cinco vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo gravitacional.

En la Fig. 3.23 se muestran las distribuciones marginales para la aproximación del

forward map con  $k = 8$  y de izquierda a derecha una malla de  $10 \times 10$ ,  $15 \times 15$ ,  $30 \times 30$ ,  $50 \times 50$ .



**Figura 3.23:** Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a ocho vecinos cercanos y una malla de resolución  $10, 15, 30, 50$  de izquierdo a derecho. En azul la distribución posterior con forward map ordinario para modelo gravitacional.

Notemos del análisis previo que de manera general las distribuciones posteriores aproximadas  $\tilde{\pi}_M^k(g, b)$  mejoran a medida que la malla se hace más fina. Luego, la mejor aproximación se obtiene con la malla más fina que discretiza el espacio paramétral en  $50 \times 50$  puntos. Queda por definir que cantidad de vecinos hace la mejor aproximación. Nótese que para 8 vecinos ( $k = 8$ ) no ajusta para el parámetro  $g$  lo que descarta esta elección. Finalmente, para 3 y 5 vecinos tenemos distribuciones bastante similares por lo que la elección es indiferente. Así, seleccionamos el forward map aproximado con 3 vecinos en una malla  $50 \times 50$ .

Para saber la factibilidad del método se considera al tiempo de ejecución, que es una variable tangible y de fácil acceso. En la tabla 3.1 se muestran los tiempos de ejecución mencionados para cada aproximación del forward map considerada. Por otro lado, el tiempo de ejecución del método ordinario para el problema inverso en

el modelo gravitacional es de **5 min 57 s**. Esto es considerando un MCMC de  $T = 600,000$  iteraciones y un burn in de 20,000, al igual que cualquier MCMC considerado en esta tesis.

Malla	10 x 10	15 x 15	30 x 30	50 x 50
3 vecinos	4m 52s	4m 55s	5m 52s	<b>5m 00s</b>
5 vecinos	5m 03s	5m 02s	4m 53s	5m 02s
8 vecinos	5m 00s	5m 09s	5m 02s	5m 02s

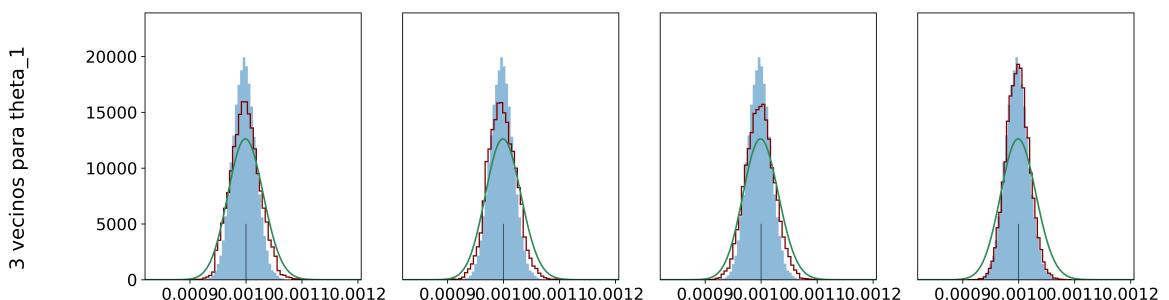
**Tabla 3.1:** Tabla de tiempos de ejecución para el MCMC del modelo gravitatorio con forward map aproximado con  $k$  vecinos (filas) en una malla de  $M \times M$  (columnas).

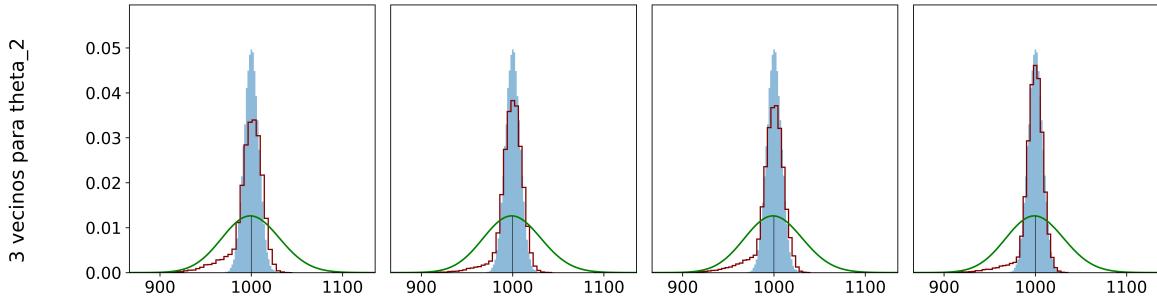
Vemos que la aproximación elegida tiene una ganancia de 57 segundos de ejecución, lo que representa una ganancia de una sexta parte del tiempo.

### Forward map aproximado para el modelo logístico

Es de interés analizar la metodología aproximada propuesta para otros modelos. La experimentación con el modelo logístico nos permite dilucidar el comportamiento de la metodología en cuestión. Al igual que el modelo gravitacional, aquí realizamos los mismos experimentos con  $k = 3, 5, 8$  y una resolución de malla de  $M = 10, 15, 30, 50$ . En todos los casos usamos el t-walk con  $T = 600,000$  iteraciones y un burn in de 20,000.

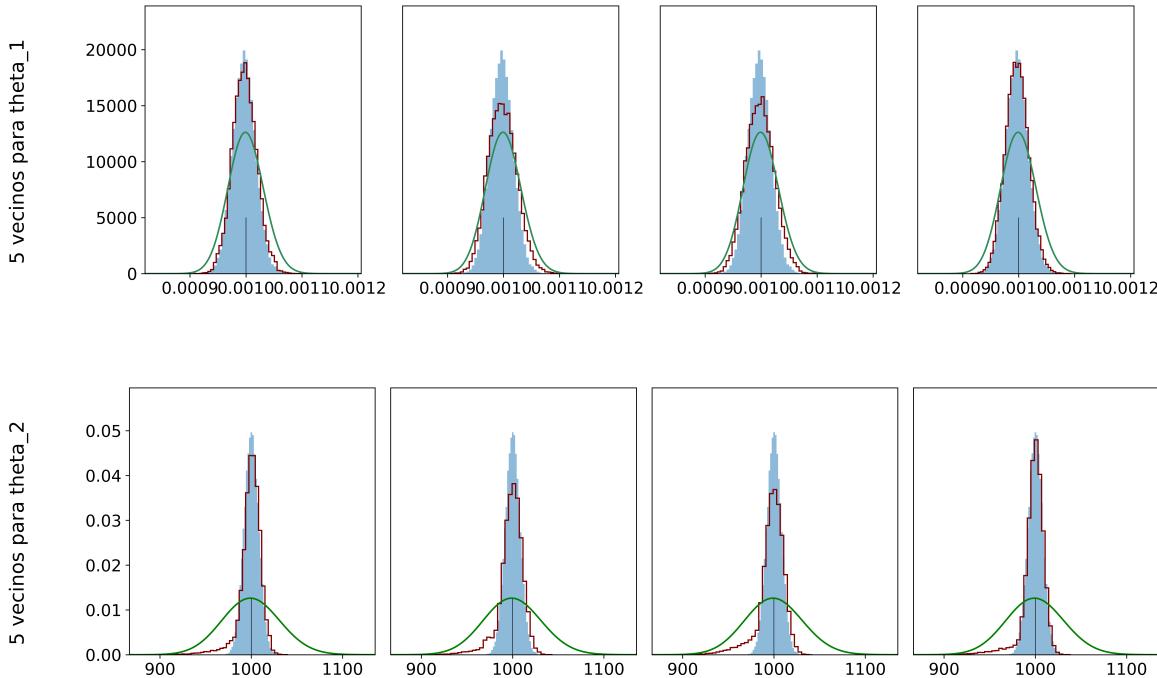
Empezando por tres vecinos cercanos para aproximación del forward map, vemos en la Fig. 3.24 las distribuciones marginales para la aproximación del forward map con  $k = 3$  y de izquierda a derecha una malla de 10x10, 15x15, 30x30, 50x50.





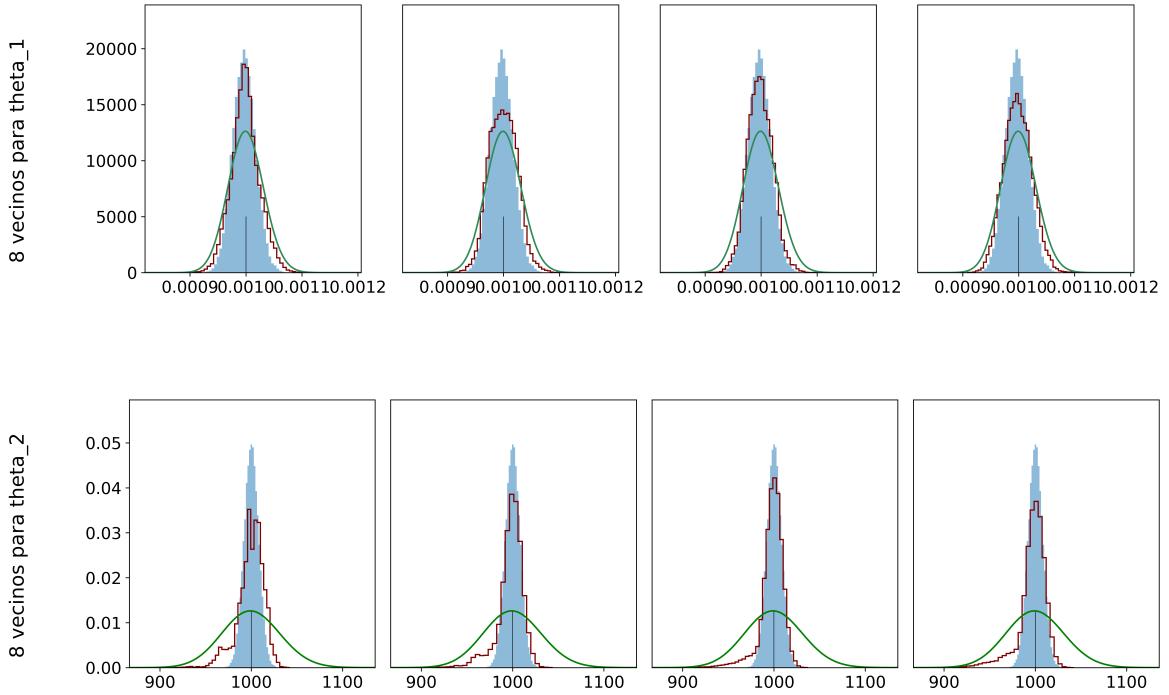
**Figura 3.24:** Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a tres vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo logístico.

Posteriormente, para cinco vecinos cercanos, en la Fig. 3.25 se muestran las distribuciones marginales para la aproximación del forward map con  $k = 5$  y de izquierda a derecha una malla de 10x10, 15x15, 30x30, 50x50.



**Figura 3.25:** Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a cinco vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo logístico.

Finalmente, para ocho vecinos cercanos, en la Fig. 3.26 se muestran las distribuciones marginales para la aproximación del forward map con  $k = 8$  y de izquierda a derecha una malla de 10x10, 15x15, 30x30, 50x50.



**Figura 3.26:** Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a ocho vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario.

El tiempo de ejecución del método ordinario para el problema inverso en el modelo logístico es de **20 min 58 s**. Mientras que en la tabla 3.2 tenemos los tiempos de ejecución del modelo logístico con forward map aproximado para cada caso considerado.

Malla	10 x 10	15 x 15	30 x 30	50 x 50
3 vecinos	10m 58s	9m 36s	9m 25s	<b>6m 02s</b>
5 vecinos	6m 02s	6m 03s	7m 20s	10m 04s
8 vecinos	10m 08s	10m 24s	10m 21s	10m 21s

**Tabla 3.2:** Tabla de tiempos de ejecución para el MCMC del modelo logístico con forward map aproximado con  $k$  vecinos (filas) en una malla de  $M \times M$  (columnas).

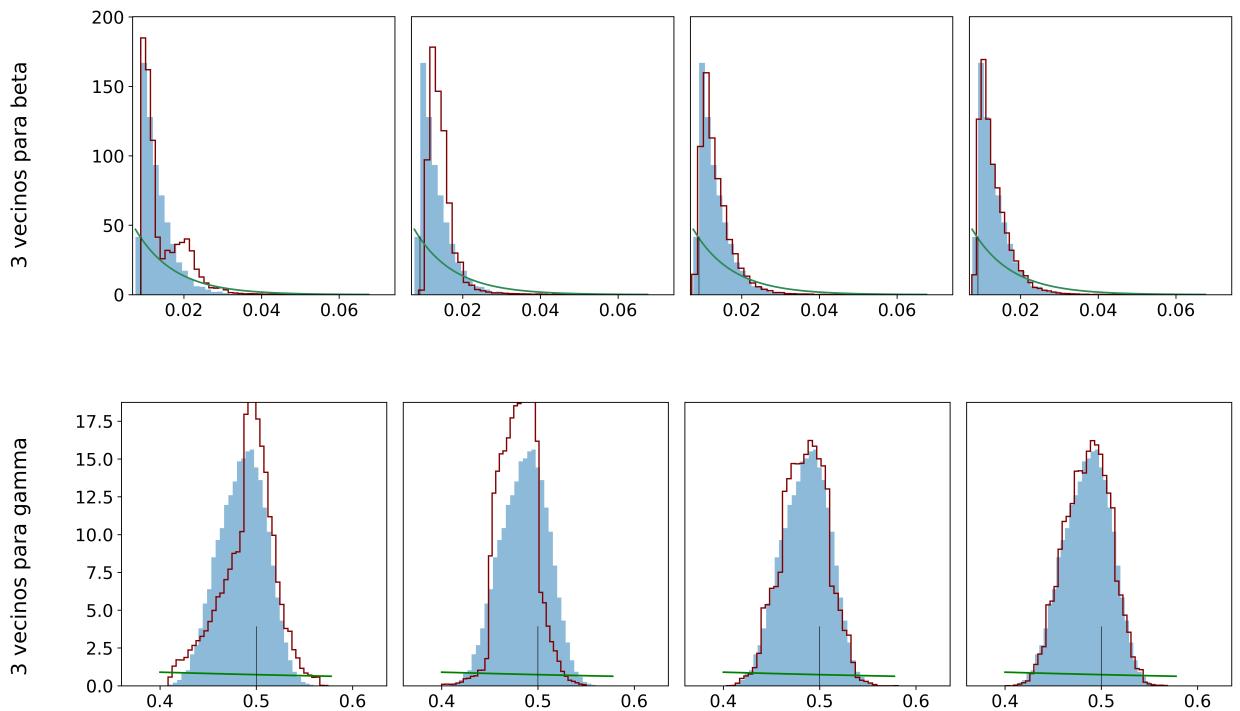
Observemos que el grado de aproximación es plausible para cada experimento. Notemos que el parámetro  $\theta_1$  aproxima mejor que el parámetro  $\theta_2$ . Dichos experimentos muestran comportamientos no esperados, pese a esto, el tiempo de ejecución es significativamente menor que su análogo ordinario puesto que reduce el tiempo de cálculo en un 70 % aproximadamente.

### Convergencia en el modelo epidemiológico

En el modelo epidemiológico SIR, el cual sabemos que su problema directo no tiene solución analítica, tiene relevancia en las diferencias de los ordenes de magnitud para sus parámetros  $\alpha, \beta$ , que es un experimento de mérito para la metodología aproximada propuesta, de regularizar las unidades.

De igual forma que las aproximaciones en modelos precedentes, en este caso se hacen repetidas distribuciones posteriores aproximadas con distintos forward map aproximados. Esto es tomar forward maps que consideren 3, 5, 8 vecinos cercanos con resolución de malla  $M = 10, 15, 30, 50$ . De igual forma que los casos anteriores, se usa el t-walk con  $T = 600,000$  iteraciones y un burn in de 20,000 iteraciones.

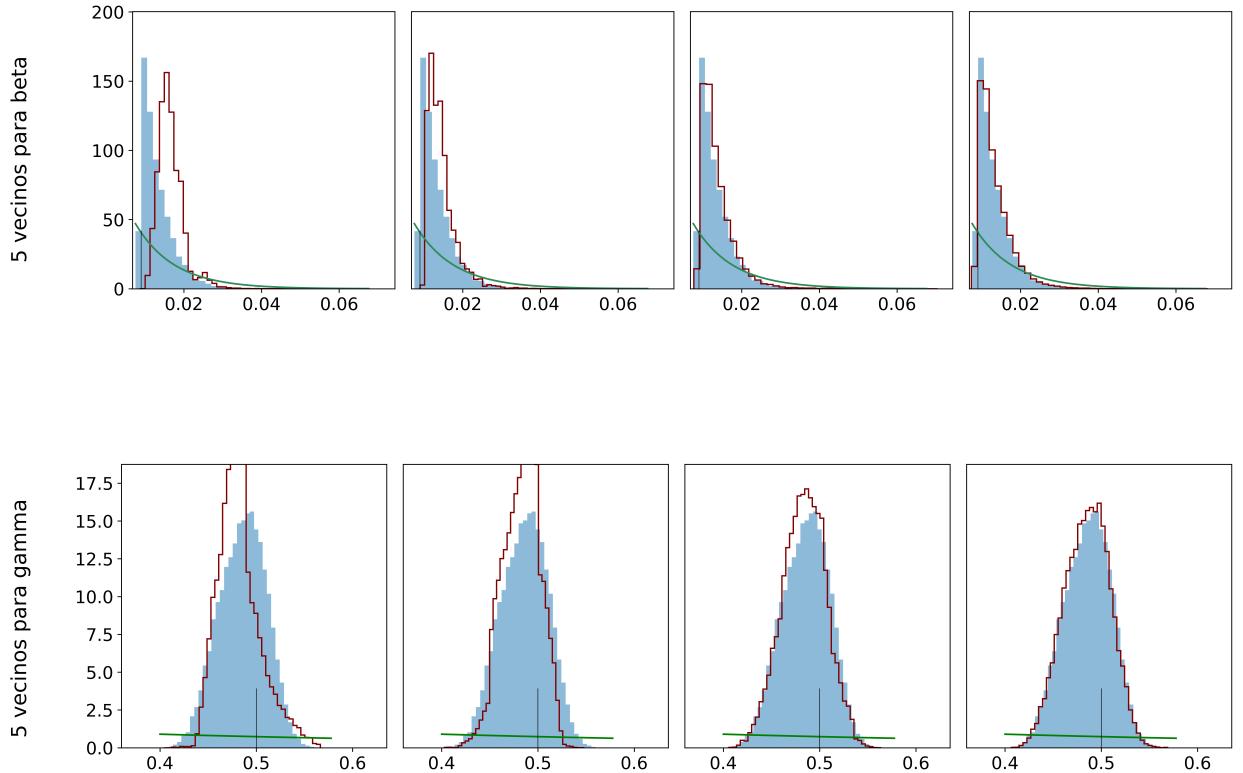
Empezando por tres vecinos cercanos, se muestran en la Fig. 3.27 las distribuciones marginales para la aproximación del forward map con  $k = 5$  y de izquierda a derecha una malla de  $10 \times 10, 15 \times 15, 30 \times 30, 50 \times 50$ .



**Figura 3.27:** Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a tres vecinos cercanos y una malla de resolución  $10, 15, 30, 50$  de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo SIR.

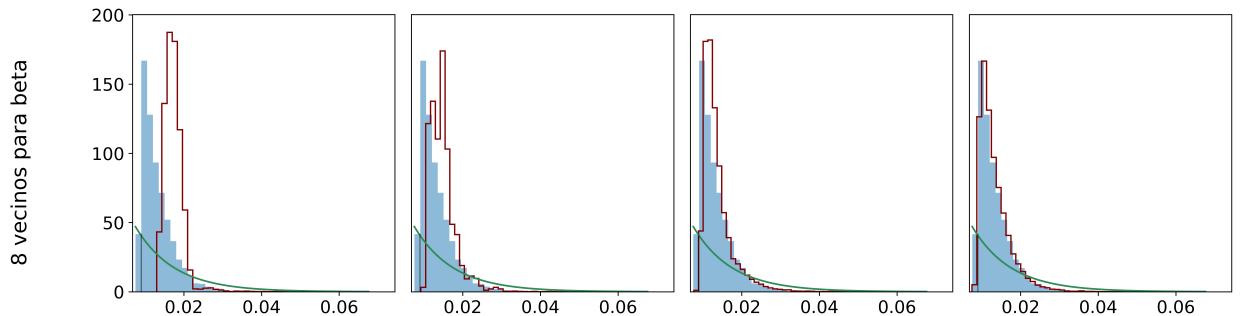
Posteriormente, con cinco vecinos cercanos, en la Fig. 3.28 se muestran las distri-

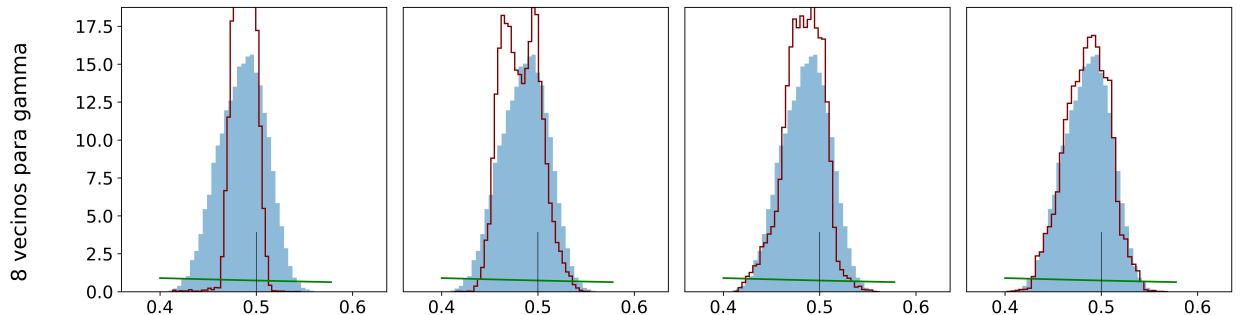
buciones marginales para la aproximación del forward map con  $k = 5$  y de izquierda a derecha una malla de 10x10, 15x15, 30x30, 50x50.



**Figura 3.28:** Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a cinco vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo SIR.

Finalmente, en la Fig. 3.29 se muestran las distribuciones marginales para la aproximación del forward map con  $k = 8$  y de izquierda a derecha una malla de 10x10, 15x15, 30x30, 50x50.





**Figura 3.29:** Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a ocho vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha. En azul la distribución posterior con forward map ordinario para modelo SIR.

El tiempo de ejecución del método ordinario para el problema inverso en el modelo gravitacional es de **14 min 05 seg**. Mientras que los distintos tiempos de ejecución para las posteriores aproximadas se tabulan a continuación.

Malla	10 x 10	15 x 15	30 x 30	50 x 50
3 vecinos	5m 38s	5m 36s	8m 14s	<b>5m 50s</b>
5 vecinos	5m 59s	5m 50s	5m 56s	5m 57s
8 vecinos	6m 38s	7m 26s	7m 02s	6m 09s

**Tabla 3.3:** Tabla de tiempos de ejecución para el MCMC del modelo SIR con forward map aproximado con  $k$  vecinos (filas) en una malla de  $M \times M$  (columnas).

Dado que la aproximación es bastante razonable incluso para mallas no tan finas, una elección predilecta para  $k$  y  $M$  no surge por simple inspección. Sin embargo, debido al tiempo de ejecución es que se toma a la aproximación optima con  $k = 3$  y  $M = 50$ .



# Capítulo 4

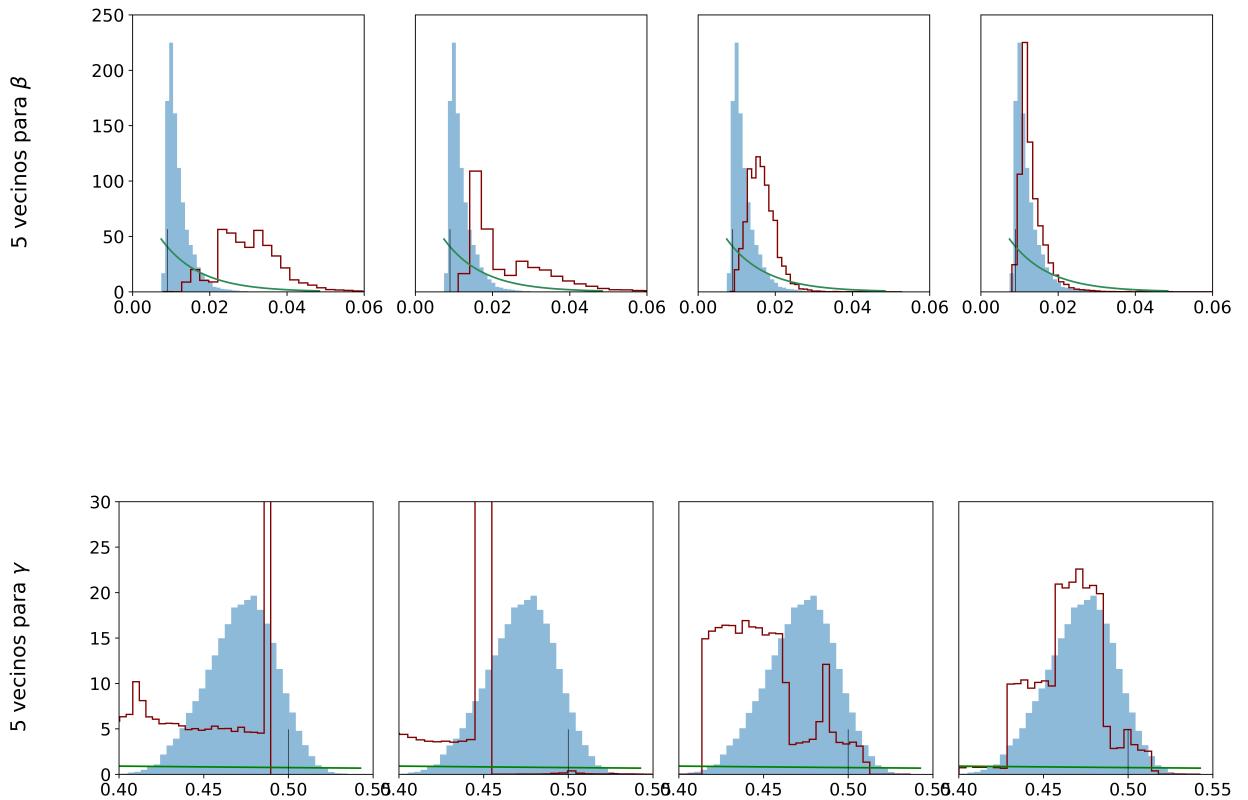
## Conclusiones y trabajo futuro

Las aproximaciones de la distribución posterior dadas por las aproximaciones del forward map son en general útiles y simples de manejar. En cada uno de los tres modelos estudiados se llegó a que existe una aproximación decente con un menor tiempo de ejecución. Sin embargo, aunque es cierto que comparaciones uno a uno el método aproximado es mejor en tiempo de ejecución, no lo es en conjunto. Esto quiere decir, que la investigación de la mejor aproximación lleva una mayor inversión temporal. Más aún, considérese que la experimentación se ha realizado con aproximaciones del forward map de la misma cantidad de vecinos cercanos y misma resolución de malla teniendo un grado de aproximación óptimo con tres vecinos y una definición de malla de 50 en cada uno de los modelos presentados. Por razonamiento inductivo, podemos conjeturar que tendremos una aproximación decente en cualquier otro modelo de EDO's con la misma regla de aproximación, esperando una mejora en el tiempo de ejecución previo a realizar la metodología con forward map ordinario.

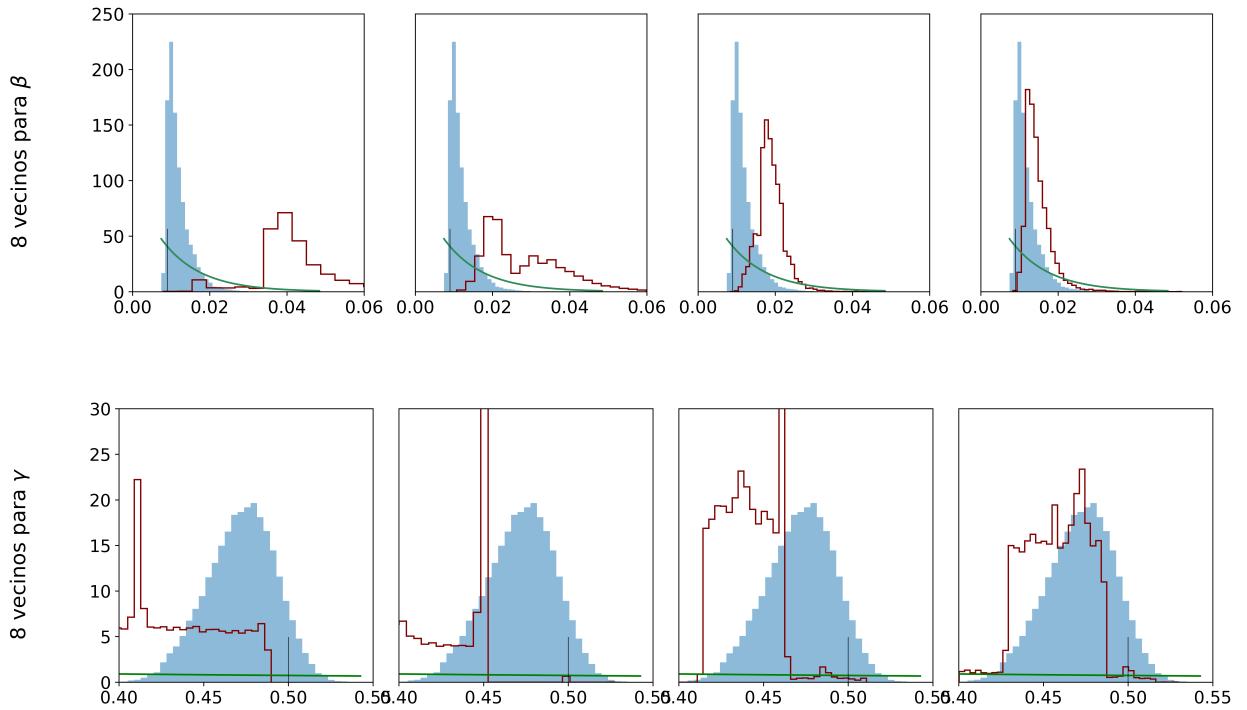
Cabe resaltar que las aproximaciones hechas con más vecinos fueron rechazadas por deficiencia de la convergencia experimental en la distribución posterior. Este caso se atribuye a que agregar muchos vecinos, como lo fue para aproximaciones de 8 vecinos cercanos, añade a la aproximación de la distribución posterior un *ruido* que perturba considerablemente la distribución posterior resultante.

Por otro lado, la incorporación de la regulación de las unidades tomó un papel fundamental para la metodología aproximada. En teoría hemos visto que la ausencia

de esta resulta en aproximaciones sesgadas a aquellos parámetros cuyas unidades sean la de menor orden de magnitud. En la práctica, el uso de la metodología sin la regularización de las unidades produce aproximaciones burdas para los parámetros con mayor orden de magnitud. En el caso del modelo SIR, considerar la resolución del problema inverso sin la regularización de las unidades, ocasiona que la distribución posterior aproximada, solo aproxima marginalmente para  $\beta$ . Esto se debe a que sin la regularización de las unidades y debido a la construcción ponderada inversamente a la distancia en el forward map aproximado, es que el orden de magnitud de las unidades de  $\gamma$  al ser mayores que las mismas de  $\beta$ , son insignificantes para la aproximación ocasionando el sesgo mencionado. Veamos el caso de cinco vecinos cercanos en la Fig. 4.1 y el caso de ocho vecinos cercanos en la Fig. 4.2 sin la regularización de las unidades.



**Figura 4.1:** Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a cinco vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha, sin regularización de las unidades. En azul la distribución posterior con forward map ordinario para modelo SIR.



**Figura 4.2:** Distribuciones marginales posteriores aproximadas (rojo) con forward map aproximado a ocho vecinos cercanos y una malla de resolución 10,15,30,50 de izquierda a derecha, sin regularización de las unidades. En azul la distribución posterior con forward map ordinario para modelo SIR.

Existen multiples formas de mejorar el algoritmo para la solución del problema inverso en modelos dados por EDO's. Una de ellas considerada a trabajo futuro es la creación de mallas no rectangulares, que además se vaya construyendo de forma adaptativa, esto quiere decir que se deje correr la cadena del algoritmo Metropolis-Hastings de forma ordinaria una cantidad fija de iteraciones y en cada paso se construya la malla. Esto con la idea de que los puntos de la discretización que son poco probables no se vean sobre-representados.

También considerado a futuro, es de interés trabajar con distintos modelos de EDO's a los presentados en este proyecto, pues de esta forma se podrían encontrar casos en los que la metodología aproximada propuesta aquí resulte con irregularidades, donde el tratamiento de estas conlleve a la mejora del algoritmo. Además de irregularidades como la que sucede en el modelo logístico, vea la Fig. 3.24, donde se muestra un sesgo a la izquierda de la distribución posterior aproximada pese a que ya se regularizaron las unidades.

Por otro lado, es de interés implementar cuantitativamente el grado de aproximación que tiene las posteriores con su análogo ordinario. Proponiendo la implementación de las distancias de Kullback-Leibler como sustitución de la simple inspección de la aproximación.

# Referencias

- Alonso, M., y Finn, E. J. (1970). Física: Volumen i: Mecánica.
- Apostol, T. M. (1991). *Calculus, volume 1*. John Wiley & Sons.
- Apostol, T. M. (2019). *Calculus ii: cálculo con funciones de varias variables y álgebra lineal, con aplicaciones para ecuaciones diferenciales y probabilidad*. Reverté.
- Berger, J. O. (2013). *Statistical decision theory and bayesian analysis*. Springer Science & Business Media.
- Calvetti, D., y Somersalo, E. (2018). Inverse problems: From regularization to bayesian inference. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(3), e1427.
- Capistran, M. A., Capella, A., y Christen, J. A. (2021). Forecasting hospital demand in metropolitan areas during the current covid-19 pandemic and estimates of lockdown-induced 2nd waves. *PloS one*, 16(1), e0245669.
- Casella, G., y Berger, R. (2024). *Statistical inference*. CRC Press.
- Christen, J. A., y Fox, C. (2010). A general purpose sampling algorithm for continuous distributions (the t-walk).
- Christen, J. A., y Pérez-Garmendia, J. L. (2021). Weak and tv consistency in bayesian uncertainty quantification using disintegration. *Boletín de la Sociedad Matemática Mexicana*, 27(1), 2.
- Cotter, S. L., Dashti, M., y Stuart, A. M. (2010). Approximation of bayesian inverse problems for pdes. *SIAM journal on numerical analysis*, 48(1), 322–345.
- Dobson, A. J., y Barnett, A. G. (2018). *An introduction to generalized linear models*. Chapman and Hall/CRC.
- Galaviz, A. (2023). *A physics-based surrogate model in inverse problem* (PhD thesis).

- Centro de Investigación en Matemáticas, Guanajuato, Guanajuato. (Available at <https://www.cimat.mx/Aplicaciones/biblioteca/TD/Pub/>)
- Griffiths, D. F., y Higham, D. J. (2010). *Numerical methods for ordinary differential equations: initial value problems* (Vol. 5). Springer.
- Jiménez Bedolla, J. C. (2022). *Métodos numéricos usando python*. Universidad Nacional Autónoma de México.
- Kaipio, J., y Somersalo, E. (2006). *Statistical and computational inverse problems* (Vol. 160). Springer Science & Business Media.
- Kelley, W. G. (2010). *The theory of differential equations*. Springer.
- Mathews, J., y Fink, K. (2000). *Métodos numéricos*. Madrid: Prentice Hall.
- Mengersen, K. L., y Tweedie, R. L. (1996). Rates of convergence of the hastings and metropolis algorithms. *The annals of Statistics*, 24(1), 101–121.
- Norris, J. R. (1998). *Markov chains* (n.º 2). Cambridge university press.
- Robert, C. P., Casella, G., y Casella, G. (1999). *Monte carlo statistical methods* (Vol. 2). Springer.
- Sears, F. W., Zemansky, M. W., Young, H. D., Vara, R. H., García, M. G., Güemes, E. R., ... Benites, F. G. (1986). *Física universitaria*. Fondo Educativo Interamericano Naucalpan de Juárez, México.
- Stuart, A. M. (2010). Inverse problems: a bayesian perspective. *Acta numerica*, 19, 451–559.
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.
- Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.
- Wasserman, L. (2006). *All of nonparametric statistics*. Springer Science & Business Media.
- Wasserman, L. (2013). *All of statistics: a concise course in statistical inference*. Springer Science & Business Media.
- Weiss, H. H. (2013). The sir model and the foundations of public health. *Materials matemáticas*, 0001–17.

Zill, D. G., Cullen, M. R., Hernández, A. E. G., y López, E. F. (2002). *Ecuaciones diferenciales con problemas de valores en la frontera* (Vol. 1). Thomson.

