

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

Agradecimientos	III
Índice de figuras	VII
1. Introducción	1
2. Antecedentes	3
2.1. Problema inverso	3
2.2. Solución bayesiana a problemas inversos	6
2.3. Método MCMC	8
2.4. Forward map aproximado	11
3. Implementación	15
3.1. Modelos de trabajo	15
3.1.1. Modelo Gravitatorio	16
3.1.2. Modelo Logístico	19
3.1.3. Modelo SIR	21
3.2. Enfoque bayesiano al problema inverso	22
3.2.1. Distribución posterior	23
3.2.2. Simulación con MCMC	26
3.2.3. Simulación del modelo SIR	40
3.3. Inferencia con aproximación.	43
3.3.1. Método aproximado al problema inverso	44
4. Conclusiones y trabajo futuro	53
Referencias	55
A. Anexo	57

Índice de figuras

3.1. Varias trayectorias con la dinámica gravitación sujeta a fricción para varias valores de los parámetros g,b.	19
3.2. Varias trayectorias con la dinámica de crecimiento logístico para varias valores de los parámetros K,r.	21
3.3. Trayectoria de cada grupo del modelo SIR con $\beta = 0.009$ y $\gamma = 0.5$. . .	22
3.4. Distribuciones a priori para $\theta_1 = g$ y $\theta_2 = b$	28
3.5. Muestra y del modelo gravitatorio.	28
3.6. Cadena obtenida por MCMC para la distribución posterior conjunta.	29
3.7. Distribución posterior para el parámetros g	29
3.8. Distribución posterior para el parámetro b	30
3.9. Distribución predictiva para el modelo gravitatorio con σ conocido. .	30
3.10. Distribuciones a priori para $\theta_1 = g$ y $\theta_2 = b$	31
3.11. Muestra y del modelo gravitatorio.	32
3.12. Cadena obtenida por MCMC para la distribución posterior conjunta.	32

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 consideraros son aquellos cuyas variables se pueden expresar ya sea en una forma funcional o en forma de ecuación diferencial ordinaria (ODE).

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.

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 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 describe 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.

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 ([Tarantola \(2005\)](#).)

2.1. Problema inverso

El estudio del problema inverso se remonta a mediados del siglo 17. Los físicos contemporáneos interesados en la relación causa-efecto de diferentes fenómenos físicos lograron predecir el comportamiento en la dinámica intrínseca. 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 \(2005\)](#)).

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). Supongase 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 β . Para esta clase de modelos el problema directo no representa complejidad. En contraste, el problema inverso precisa estimaciones de β y σ 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. Considerese 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.

Más aún, se puede obtener el análisis del problema inverso para sistemas de ecua-

ciones diferenciales ordinarias. Esto es

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

con G_1, G_2, \dots, G_d funciones continuas conocidas (Zill, Cullen, Hernández, y López (2002)).

Observemos que la restricción implica restringir las variables en un soporte continuo, por lo que modelos con variables discretas deben de modificarse. 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 m, g, b parámetros del modelo.

Por el teorema de existencia y unicidad en ecuaciones diferenciales; dadas las condiciones iniciales, los parámetros del modelo definen únicamente el modelo (Kelley (2010)). Consideraremos el espacio de parámetros Θ que a su vez es la familia de *submodelos* posibles descrito por la ODE.

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} a las soluciones posibles de la ecuación diferencial. De esta forma existe un mapeo del espacio Θ 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 numéricamente 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 $F(\theta) = y(t)$. 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

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

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 θ y σ . Como se consideran errores independientes, pues se tratan de errores de medición, la verosimilitud se sigue de

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

simplificando

$$\mathcal{L}(\theta, \sigma) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - F_\theta(t_i))^2\right\},$$

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

$$\pi(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{\int f(\mathbf{y}|\theta)\pi(\theta)d\theta}, \quad (2.3)$$

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.

A mediados del siglo XX, 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 $\frac{1}{N}$ 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 (2.3) con el método de Markov Chain Monte Carlo (MCMC) evita el calculo numérico de $h(\mathbf{y})$ como se ha mencionado previamente. 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 asimismo la estimación de las distribuciones marginales. Los métodos MCMC busca encontrar una cadena de Markov $X^{(1)}, X^{(2)}, \dots$ cuya distribución estacionaria sea la *distribución objetivo*¹ $f(x)$ pese a que la misma no esté normalizada.

¹En aplicaciones a la estadística bayesiana usando MCMC, la distribución posterior es la distribución objetivo.

Algoritmo Metropolis-Hastings

El algoritmo de Metropolis-Hastings es un método de MCMC para generar muestras de una distribución objetivo $f(x)$ partiendo de muestras de una distribución propuesta $q(y|x)$ 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 preservando 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 $Y \sim q(y|X_i)$.

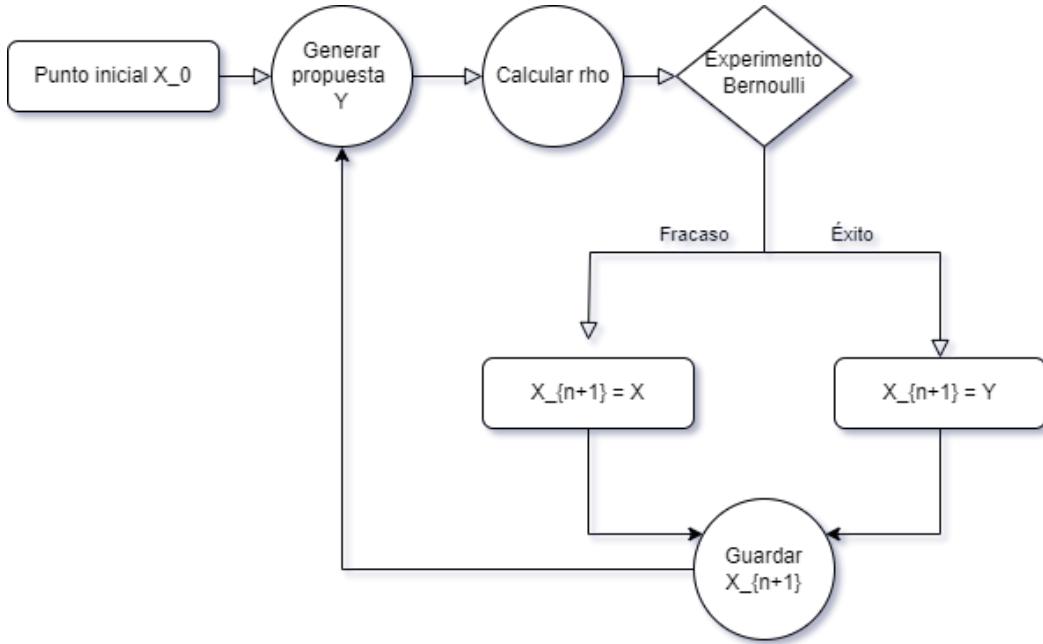
2. Calcular ρ

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

3. Realizar un experimento Bernoulli con probabilidad de éxito ρ .

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

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. 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 una cantidad numerable no finita de estados, 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 inquierir mejoras en su procedimiento. Es por ello que la propuesta principal del trabajo de tesis es la introducción del forward map aproximado.

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

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 Θ . 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 θ_i^{min} y θ_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 θ_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 parametral Θ . Observese que la cardinalidad de \mathcal{M} es de M^d . Denotemos por ϑ a los elementos de \mathcal{M} , así el enmallado se constituye de M^d parámetros. Es decir, se establece que $\mathcal{M} = \{\vartheta_1, \vartheta_2, \dots, \vartheta_{M^d}\}$, note 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 calculo intermedio. Dicho en otras palabras, solo se resuelven las ecuaciones diferenciales con los parámetros donde el enmallado interseca.

Para terminar, el forward map aproximado para parámetros $\theta \notin \mathcal{M}$ se propone buscar a los k vectores de parámetros ϑ más cercanos en distancia euclídea a θ , denotando estos k vecinos por $\vartheta^{(1)}, \dots, \vartheta^{(k)}$ cada uno a una distancia d_1, \dots, d_k de θ , 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

$$\tilde{F}_M^k(\theta) = \sum_{j=1}^k \omega_j F(\vartheta^{(j)}) \quad (2.4)$$

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

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 \(2023\)](#). Debido a la reticente necesidad humana en

aplicaciones imposergables es imperativo agilizar el tiempo en operación. Concisa-mente en modelos epidemiológicos la rapidez de acción se torna fundamental como lo fue en la predicción de ocupación hospitalaria por COVID-19 en la zona metropolitana de la CDMX ([Capistrán, Capella, y Christen \(2020\)](#)).

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. De esta misma forma, tras sustituir el forward map aproximado (2.4) en la expresión (2.3) obtenemos la distribución posterior aproximada de los parámetros del modelo. Formalizando, salvo un factor de normalización definimos

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

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 ($k = 1$) la consistencia es trivial puesto que al hacer la malla más fina, el forward map aproximado será idéntico al forward map ordinario.

El estudio para $k > 1$ requiere de una estructura teórica sobre las cualidades particulares de la ecuación diferencial del modelo en cuestión. De forma que 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

Debido a que nos encontramos en un terreno carente de fundamentos teóricos, nuestra mejor arma para probar que grado de aproximación de los problemas inversos son consistentes es por medio de experimentación metodológica. Esto nos permite revelar obstáculos del nuevo método así como también las grandes ventajas.

Gran parte del trabajo de tesis se centra primordialmente en la experimentación con la propuesta innovadora para la solución bayesiana del problema inverso. Se hacen los experimentos con una implementación en Python cuyo código se encuentra adjunto.

3.1. Modelos de trabajo

Existen una variedad de modelos a los que se puede aplicar la metodología descrita para problemas inversos. De hecho se puede implementar para cualquier modelo descrito por una ecuación diferencial ordinaria. 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 posterior, entre otras. Por ello, nos permitiremos enfocarnos en solo un par de modelos que son el modelo gravitatorio sujeto a fricción para un partícula en caída y el modelo de crecimiento logístico para una población en crecimiento.

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 [Sears y cols. \(1986\)](#)

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

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 de los primeros polinomios de Taylor a primer grado, donde se descarta el término 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 únicamente 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

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

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 $x(t_i)$ como podemos inferir los parámetros $\theta = (g, b)$. La ciencia actual ya es 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

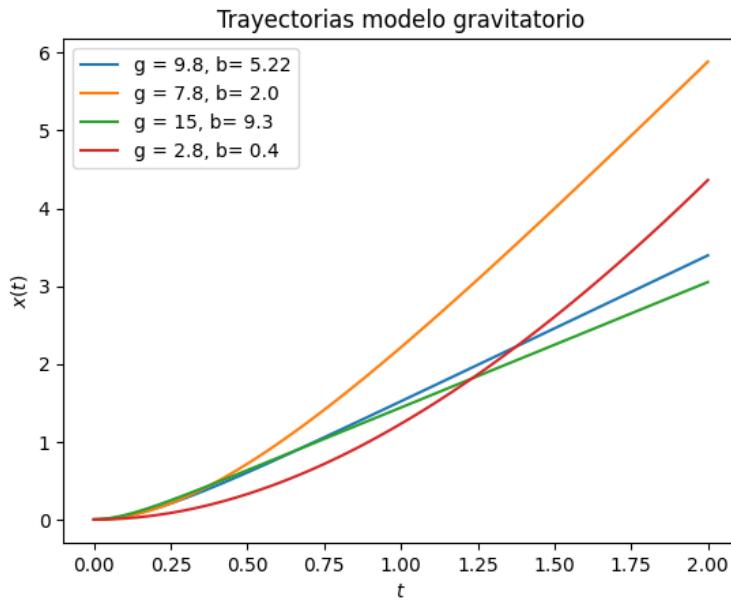


Figura 3.1: Varias trayectorias con la dinámica gravitación sujeta a fricción para varias valores de los parámetros g, b .

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 λ es la tasa de crecimiento y con la condición $P(0) = P_0$.

La solución explícita al modelo en (3.8) es

$$P(t) = P_0 e^{\lambda t} \quad (3.9)$$

que es una función creciente en el tiempo. Sin embargo, este modelo solo es valioso para pequeños intervalos de tiempo, ya que el crecimiento en una la tasa de cambio

de la población tiende a decaer a medida que se consumen los recursos.

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

$$\frac{dP}{dt} = rP \left(1 - \frac{P}{K}\right) \quad (3.10)$$

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 explícita. Notemos que la EDO pertenece a la familia de las EDO de Bernoulli, esto es tiene la forma

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

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

Por tanto, del cambio $u(t) = P^{-1}$ obtenemos la solución explícita a (3.8)

$$P(t) = \frac{KP_0e^{rt}}{K + P_0(e^{rt} - 1)} \quad (3.11)$$

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

Podemos ver en la Fig 3.2 que sus trayectorias son no decrecientes y tiene asintóticamente al parámetro K .

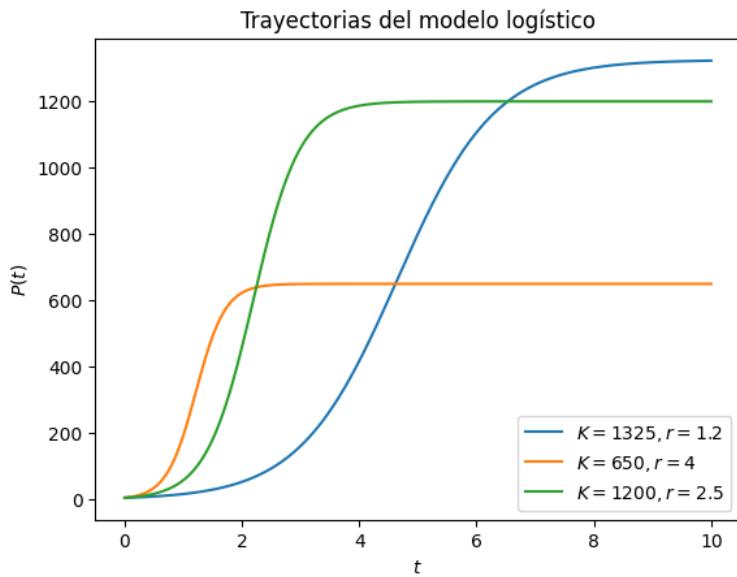


Figura 3.2: Varias trayectorias con la dinámica de crecimiento logístico para varias valores de los parámetros K, r .

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 se es el modelo SIR. Este propone dividir a las N personas de una población en tres grupos:

1. Susceptible (S): Todas las personas que no han tenido la enfermedad ni están enfermas del brote de interés.
2. Infectado (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

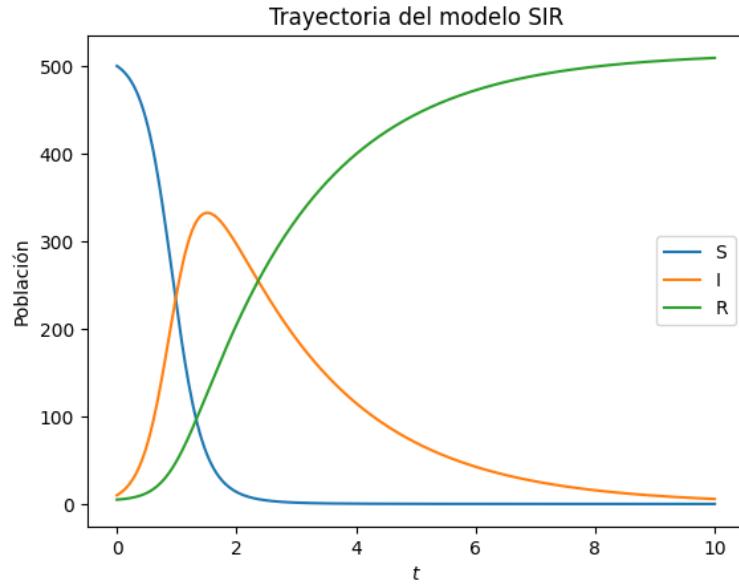


Figura 3.3: Trayectoria de cada grupo del modelo SIR con $\beta = 0.009$ y $\gamma = 0.5$.

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 β es la tasa de infección, γ es la tasa de recuperación y $S + I + R = N$.

La solución explícita para $S(t)$, $I(t)$, $R(t)$ no existe en general, por ello se requieren métodos numéricos para resolver numéricamente para cada grupo y poder construir el forward map. En la Fig. 3.3 tenemos un ejemplo para un 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 tomar mediciones de caída de un objeto para ciertos tiempo, o podemos contar la población para cada intervalo de tiempo, 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 convenientemente y además agregar un ruido gaussiano. Es decir, tomamos $\theta \in \mathbb{R}^d$ 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 F_θ 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

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

un conjunto de n tuplas.

Una vez establecida la muestra y_i , ahora nuestro propósito es abordar el problema inverso, esto es buscamos θ tal que $y_i \approx F_\theta(t_i)$. Del capítulo 2 sabemos que hay varias maneras de abordar el problema, en adelante solamente se usará el enfoque bayesiano del problema inverso.

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 denotamos por $\mathbf{y} = (y_1, y_2, \dots, y_n)$. Luego, sabemos que salvo el caso de análisis conjugado podemos acceder directamente a dicha distribución posterior. En el caso del problema inverso no es posible determinar la distribución por medio de análisis

conjugado en caso general, debido a que el forward map somete a la distribución posterior a una familia diferente a la familia de la distribución a priori.

Para el vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_d) \in \mathbb{R}^d$, recordemos la ecuación ([ecu](#))

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

que nos da la distribución posterior salvo una constante en términos de la verosimilitud y la distribución a priori. Para obtener la verosimilitud, observemos de la ecuación ([3.13](#)) que

$$\mathbf{y}_i \sim N(F_\theta(t_i)), \sigma^2$$

por lo que la verosimilitud para θ es la correspondiente al modelo normal, es decir

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

simplificando

$$\mathcal{L}(\theta, \sigma) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - F_\theta(t_i))^2 \right\}, \quad (3.17)$$

que es la verosimilitud en función de los parámetros θ y la desviación estándar σ .

Por otro lado, la distribución a priori representa el conocimiento previo a observar los datos. Para nuestro caso particular, los modelos propuestos en la sección ([citar sección](#)) dependen de parámetros no negativos, es decir, $\theta_i \geq 0$ para toda $i = 1, \dots, d$. Incluso algunos de ellos se tiene bastante certeza de su valor verdadero, 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 el siguiente paso es considerar una normal truncada.

Finalmente, se optó 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 θ_i^* y asemeja una normal es

$$\theta_i \sim \text{Gamma} \left(\alpha_i, \frac{\theta_i^*}{\alpha_i} \right) \quad (3.18)$$

con θ_i^* y α_i parámetros conocidos.

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 α, β conocidos.

Como θ_i se proponen independientes, entonces la distribución a priori para θ es

$$\begin{aligned} f(\theta|\alpha) &= f(\theta_1|\alpha_1) \cdots f(\theta_d|\alpha_d) \\ &= \prod_{i=1}^n \left[\frac{1}{\Gamma(\alpha_i)} \left(\frac{\theta_i^*}{\alpha_i} \right)^{\alpha_i} \theta_i^{\alpha_i-1} \exp\left\{-\frac{\theta_i^*}{\alpha_i}\theta_i\right\} \right] \end{aligned} \quad (3.19)$$

desarrollando

$$\begin{aligned} f(\theta|\alpha) &= \\ &\frac{1}{\Gamma(\alpha_1)} \left(\frac{\theta_1^*}{\alpha_1} \right)^{\alpha_1} \theta_1^{\alpha_1-1} \exp\left\{-\frac{\theta_1^*}{\alpha_1}\theta_1\right\} \cdots \frac{1}{\Gamma(\alpha_d)} \left(\frac{\theta_d^*}{\alpha_d} \right)^{\alpha_d} \theta_d^{\alpha_d-1} \exp\left\{-\frac{\theta_d^*}{\alpha_d}\theta_d\right\} \end{aligned} \quad (3.20)$$

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

Sustituyendo (3.17) y (3.19) en (3.15) se tiene la forma funcional de la distribución posterior

$$\pi(\theta, \sigma|y) \propto \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - F_\theta(t_i))^2\right\} \prod_{i=1}^n \left[\frac{1}{\Gamma(\alpha_i)} \left(\frac{\theta_i^*}{\alpha_i} \right)^{\alpha_i} \theta_i^{\alpha_i-1} \exp\left\{-\frac{\theta_i^*}{\alpha_i}\theta_i\right\} \right] \quad (3.21)$$

Notemos que la distribución en (3.21) no pertenece a una familia conocida. Para lidiar con este problema usamos métodos Monte Carlo. Más precisamente, podemos

simular de la distribución posterior (3.21).

3.2.2. Simulación con MCMC

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. Otra forma es usando métodos numéricos para calcular la integral del denominador, sin embargo no es óptimo para funciones de varias variables. Por ello es que simular de la distribución posterior con MCMC (Markov Chain Monte Carlo) es la vía óptima para tener una muestra de dicha distribución, con dicha muestra podemos estimar cualquier momento de la distribución. Si estamos interesados en obtener la media de X , en beneficio de la ley de grandes números, podemos estimar como

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

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

$$\mathbb{E}[h(X)] \approx \frac{1}{n} \sum_{i=1}^n h(X_i) \quad (3.23)$$

donde X_i es la simulación de la distribución posterior.

La distribución objetivo que se implementa en el MCMC es la distribución posterior (3.21). 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_θ . Sea $\theta \in \mathbb{R}^d$ la función

$$\theta \mapsto F_\theta = x(t) \quad (3.24)$$

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). Además, es necesario saber como abordar el caso de la desviación estándar σ . Existen dos formas, pensar en σ como parámetro conocido, por lo que no es necesario hacer inferencia, el otro caso es para σ desconocida y veremos los detalles después de explicar el primer caso.

Para el caso σ **conocido**, la distribución posterior (3.21) tendrá como variables únicamente al vector de parámetros θ , 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 (3.18) como

$$g \sim \text{Gamma} \left(\alpha, \frac{\theta_1^*}{\alpha} \right) \quad (3.25)$$

$$b \sim \text{Gamma} \left(\beta, \frac{\theta_2^*}{\beta} \right) \quad (3.26)$$

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.

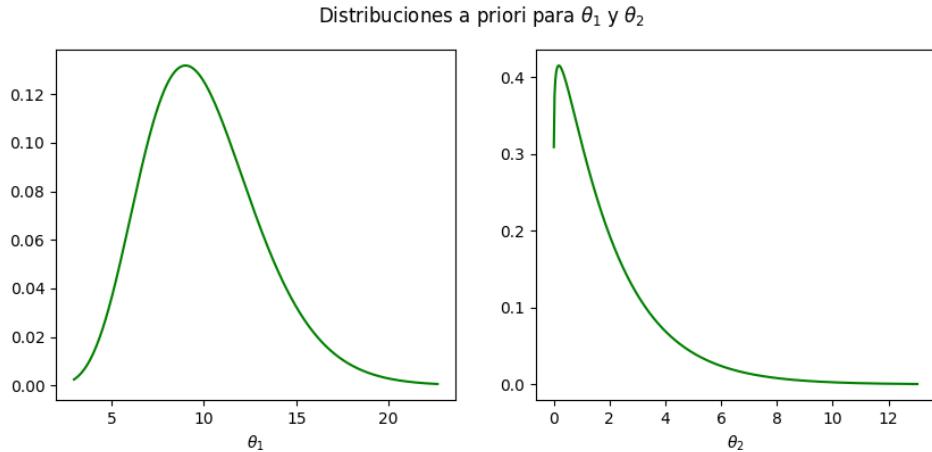


Figura 3.4: Distribuciones a priori para $\theta_1 = g$ y $\theta_2 = b$.

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

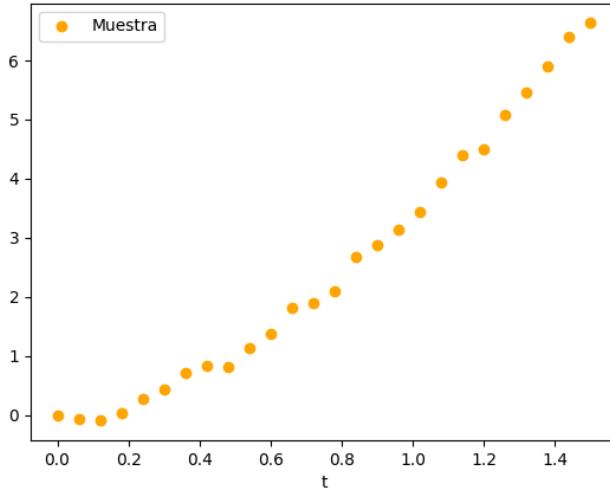


Figura 3.5: Muestra \mathbf{y} del modelo gravitatorio.

Luego, con la metodología propuesta, usando el algoritmo MCMC Metropolis-Hastings para una cadena de tamaño $T = 500,000$ y un burn in de 20,000 tenemos que la trayectoria seguida en el espacio de parámetros $\theta = (g, b)$ para la distribución posterior es el mostrado en la Fig. 3.6. Notemos que dicha posterior tiene alta correlación entre parámetros. Para valores altos de g es necesario tener más fricción b con el medio para la misma trayectoria \mathbf{y} .

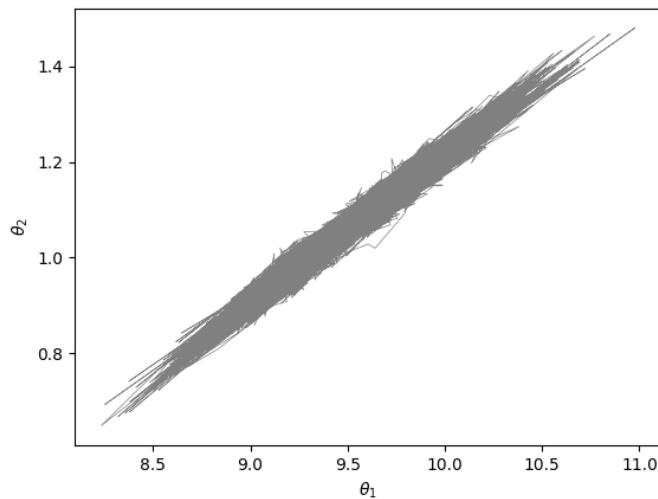


Figura 3.6: Cadena obtenida por MCMC para la distribución posterior conjunta.

Nos es de interés las marginales de la distribución posterior. Dado que la simulación de la cadena es bidimensional, de (3.23) 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 la Fig.

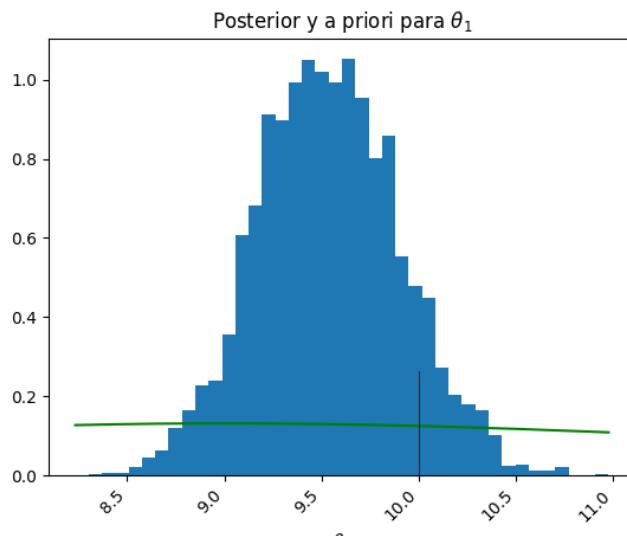


Figura 3.7: Distribución posterior para el parámetro g .

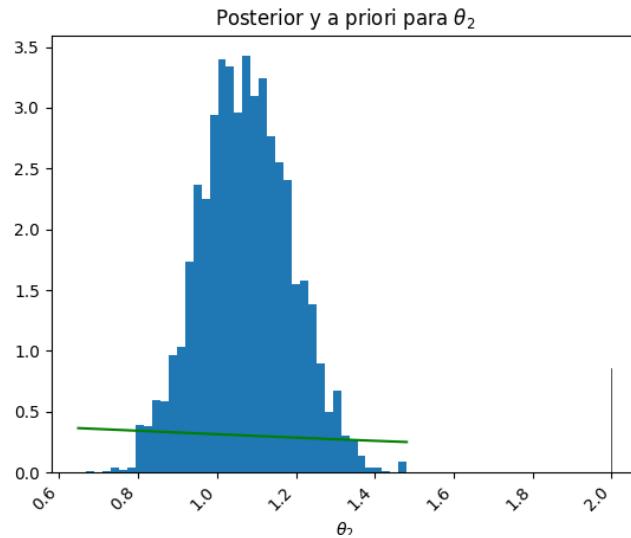


Figura 3.8: Distribución posterior para el parámetro b .

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 de g y b y así aplicar el forward map para cada estimación. En la Fig.

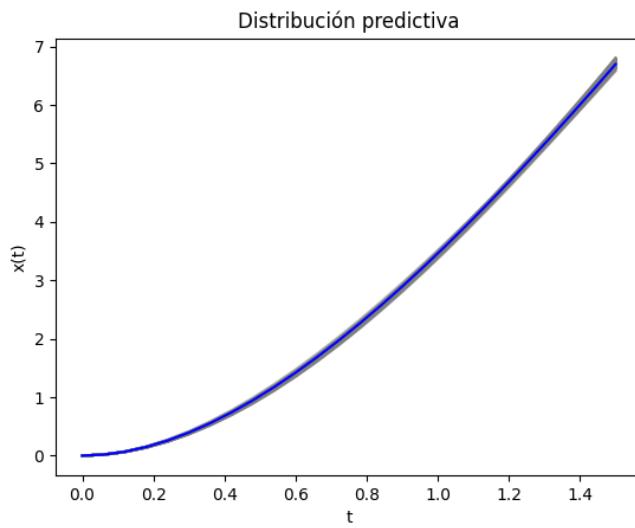


Figura 3.9: Distribución predictiva para el modelo gravitatorio con σ conocido.

Por otro lado, el análisis previo se dio con σ conocido. Sin embargo, este supuesto es irreal. Pues a pesar que el supuesto de normalidad de los errores es plausible, no

lo es considerar la desviación estándar como parámetro conocido.

El caso de σ **desconocido** se considera a dicho parámetro como uno más de la distribución posterior dada en (3.21).

En el estudio de la distribución posterior para este caso, análogo al estudio previo con σ conocido, proponemos las siguientes distribuciones a priori para g y b mostradas en la Fig. 3.10

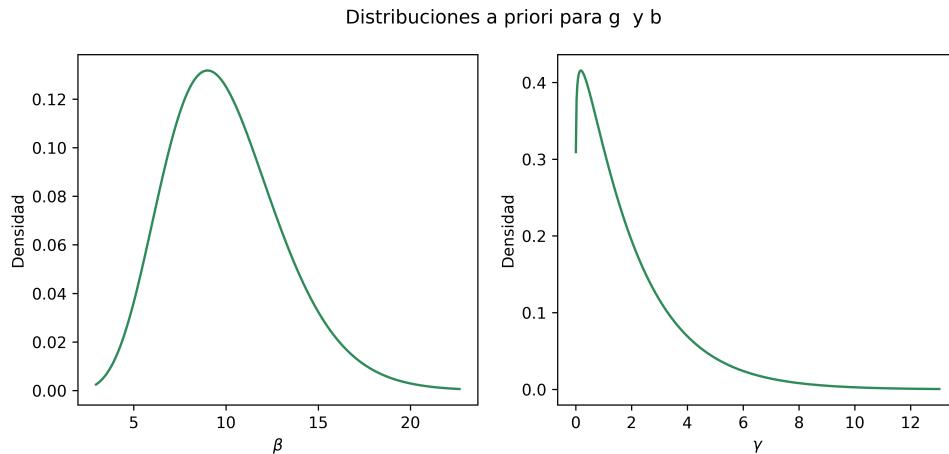


Figura 3.10: Distribuciones a priori para $\theta_1 = g$ y $\theta_2 = b$.

En la Fig. (3.11) tenemos una muestra del modelo gravitatorio. Así, con el algoritmo de MCMC Metropolis-Hastings, la trayectoria de la cadena para la distribución posterior conjunta es 3.12

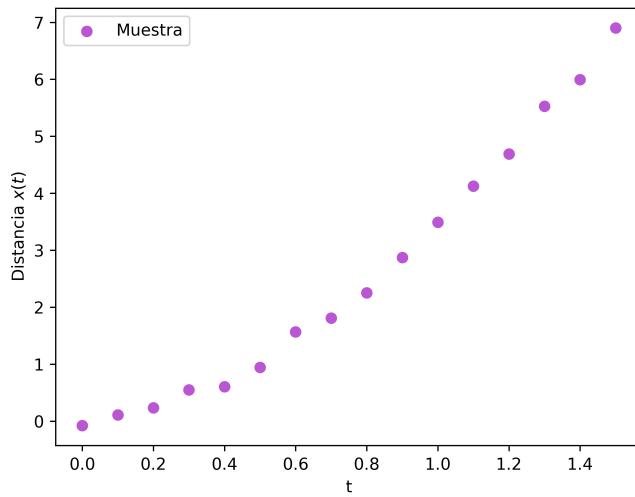


Figura 3.11: Muestra y del modelo gravitatorio.

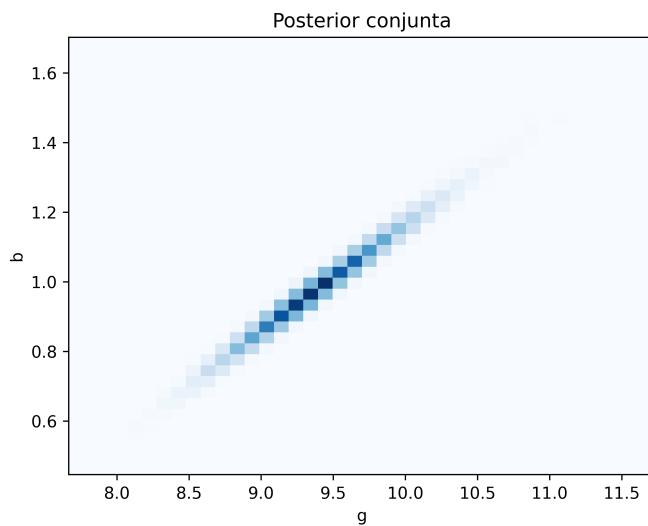
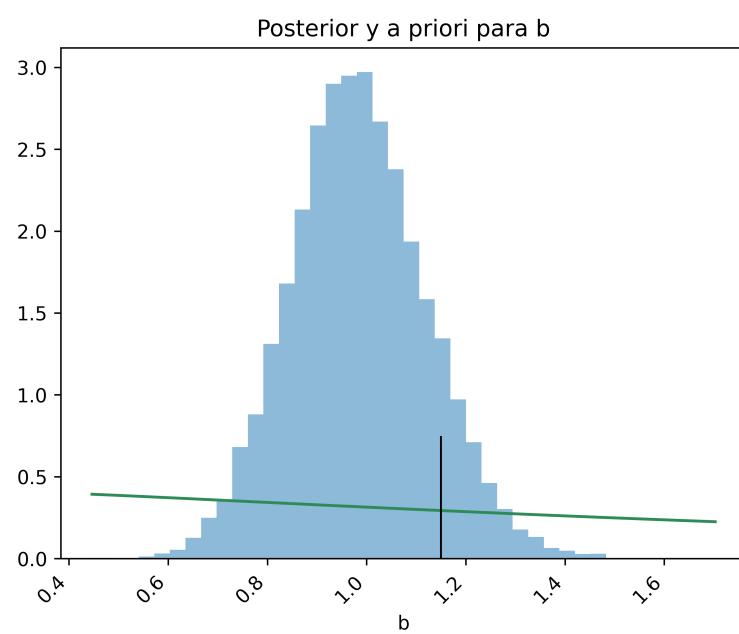
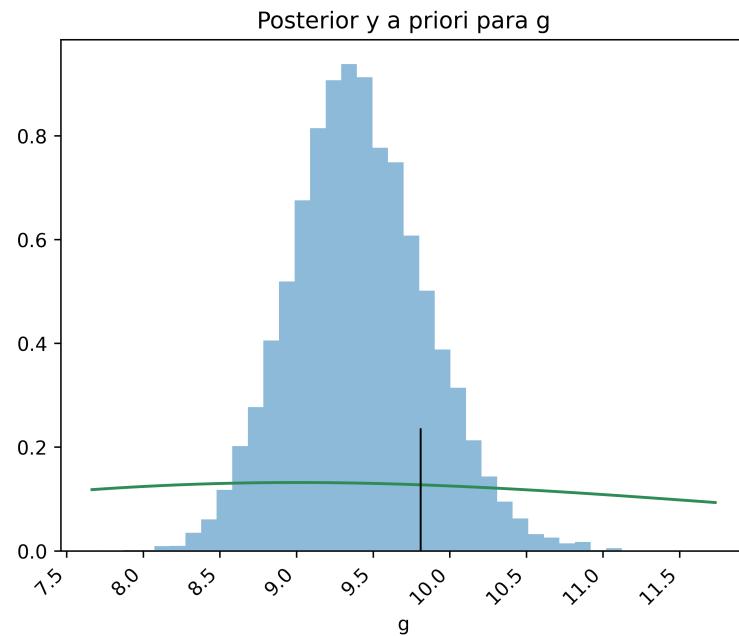
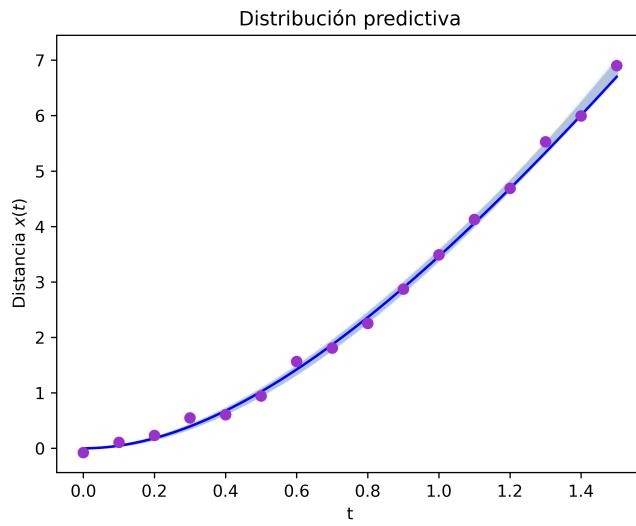
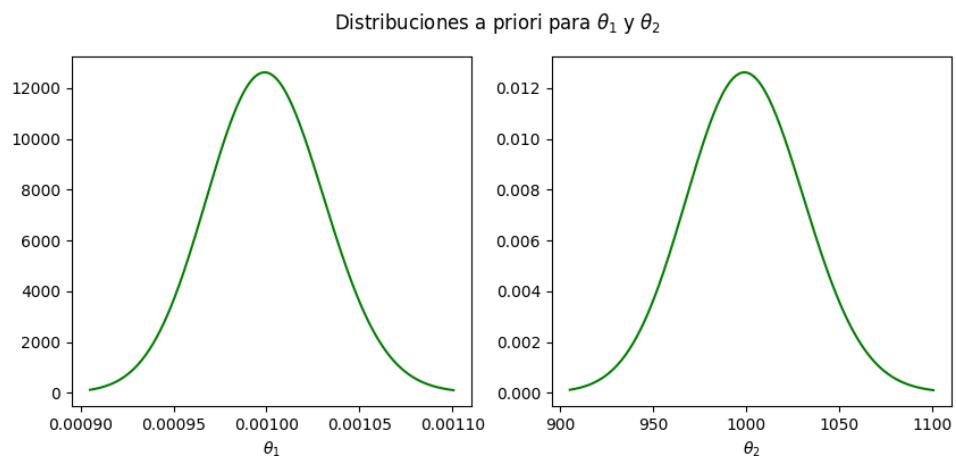


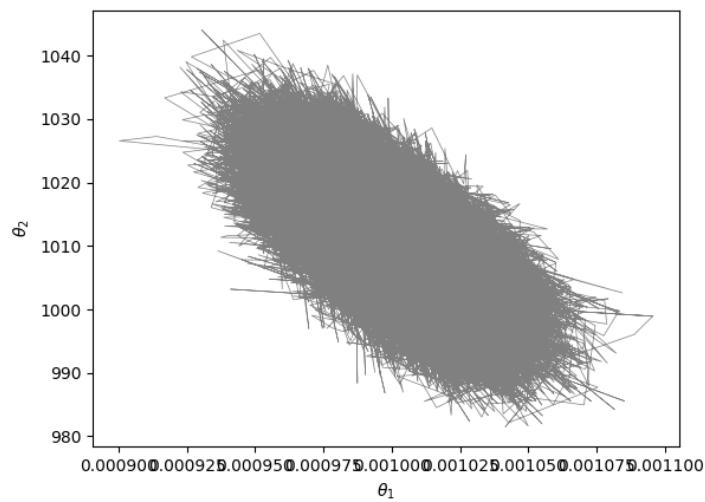
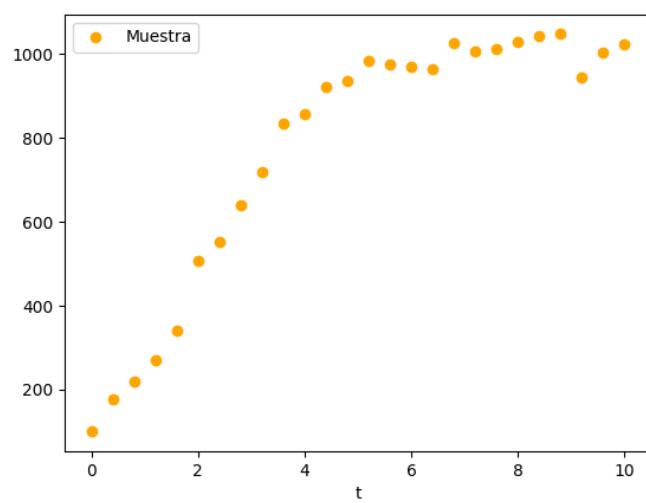
Figura 3.12: Cadena obtenida por MCMC para la distribución posterior conjunta.

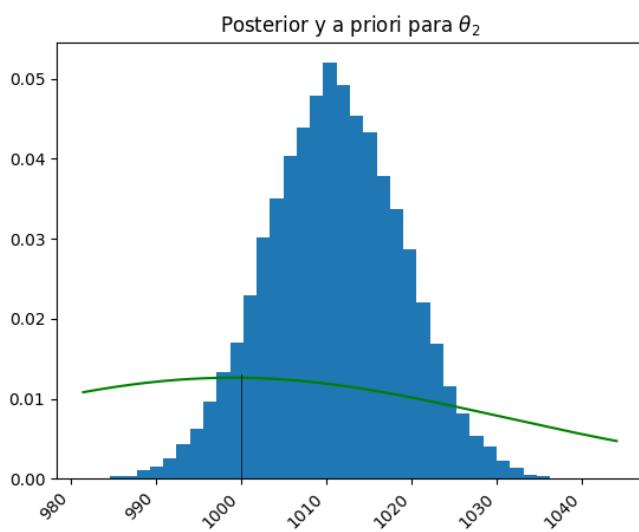
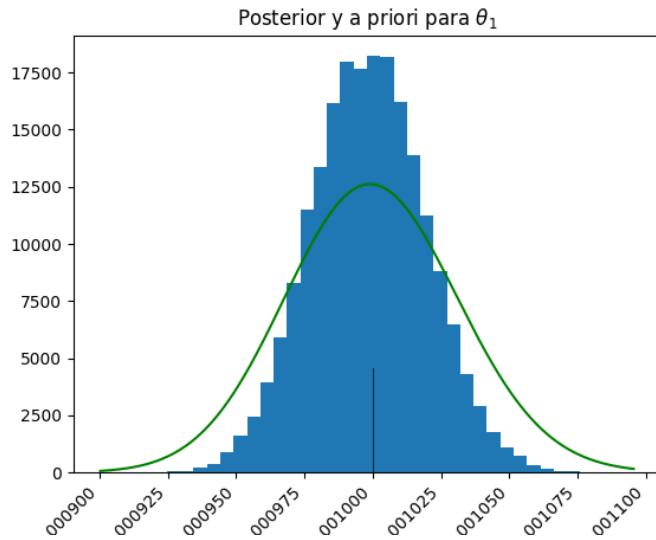


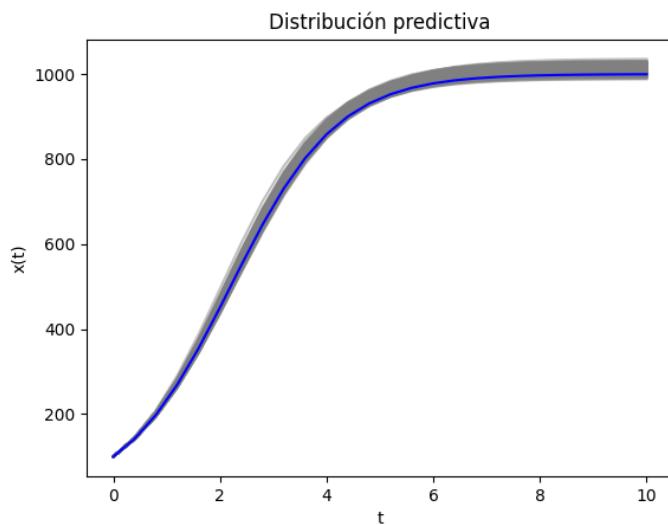


Modelo logístico

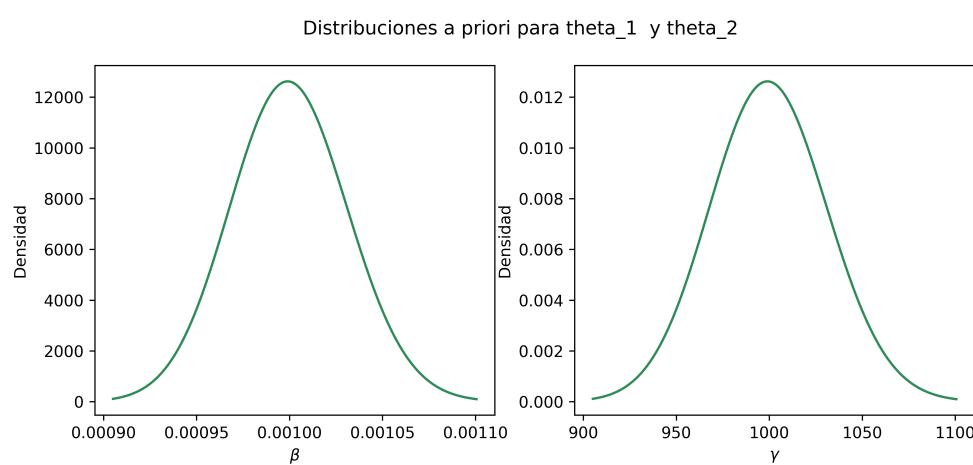


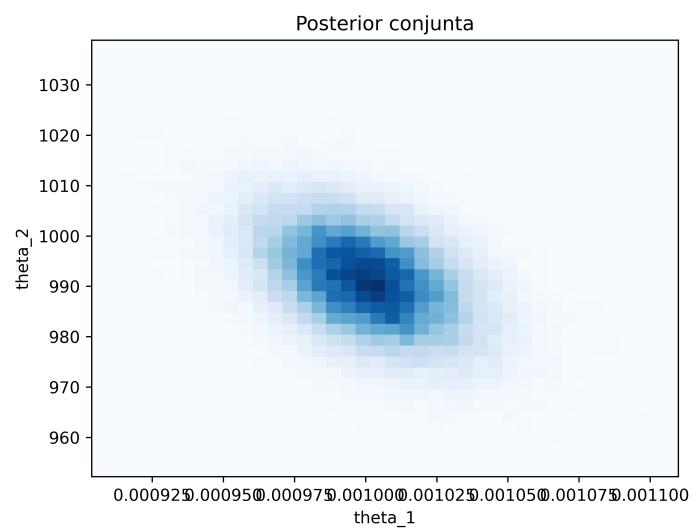
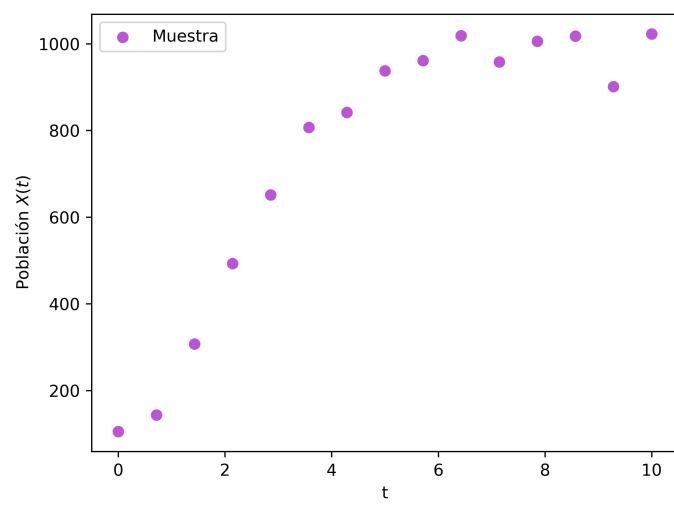


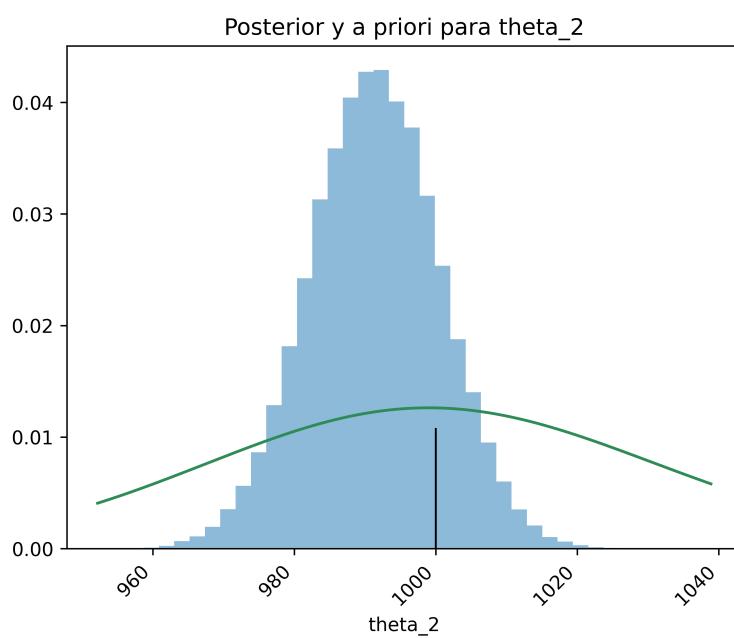
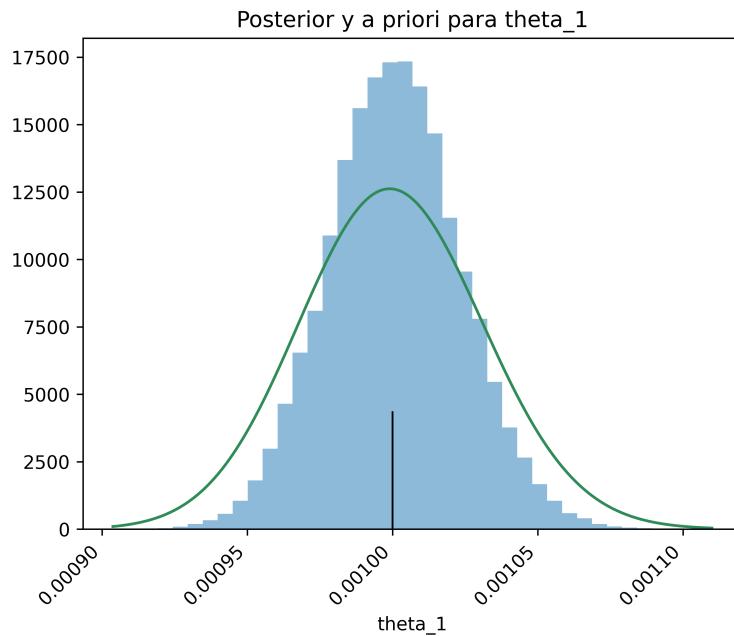


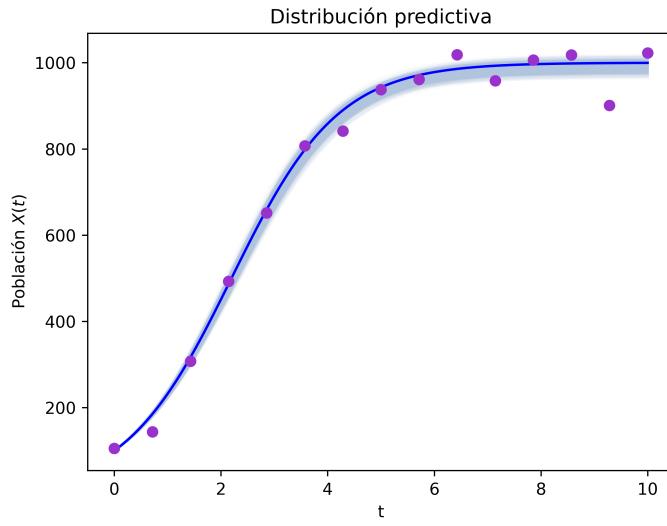


Sigma desconocido



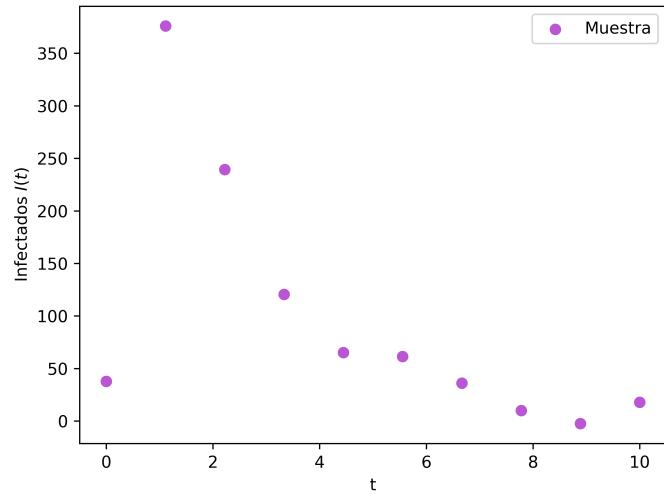






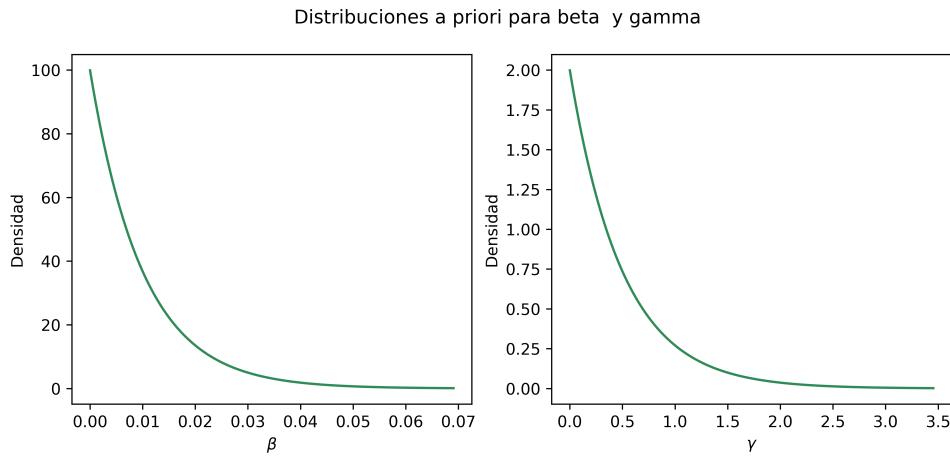
3.2.3. Simulación del modelo SIR

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. En la Fig ?? vemos la grafica de la muestra de infectados, notamos que tienen un rápido crecimiento para después decaer suavemente.

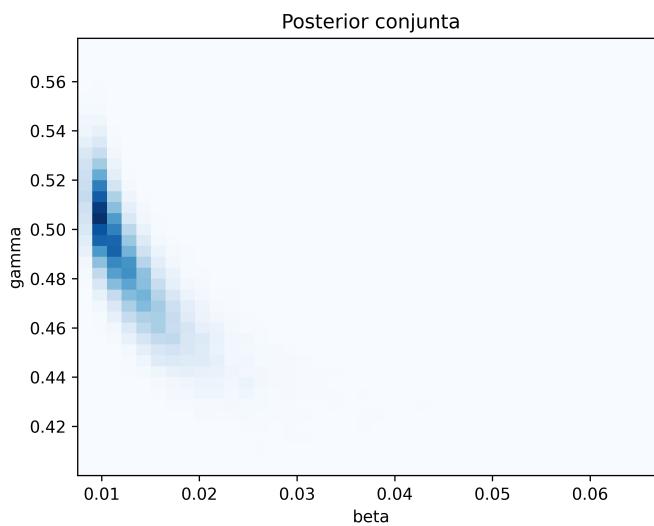


Para abordar el problema inverso establezcamos los parámetros que se desea hacer

inferencia. El espacio paramétral se conforma por $\theta = (\beta, \gamma)$ donde $\beta > 0$ es la tasa de infección y $\gamma > 0$ es la tasa de recuperación. Al tratarse de parámetros positivos usamos las distribuciones a priori (3.18). Al tratarse de parámetros estudiados en la literatura, podemos rescatar un conocimiento a priori sobre su distribución previo a los datos, mostrada en la Fig. ??

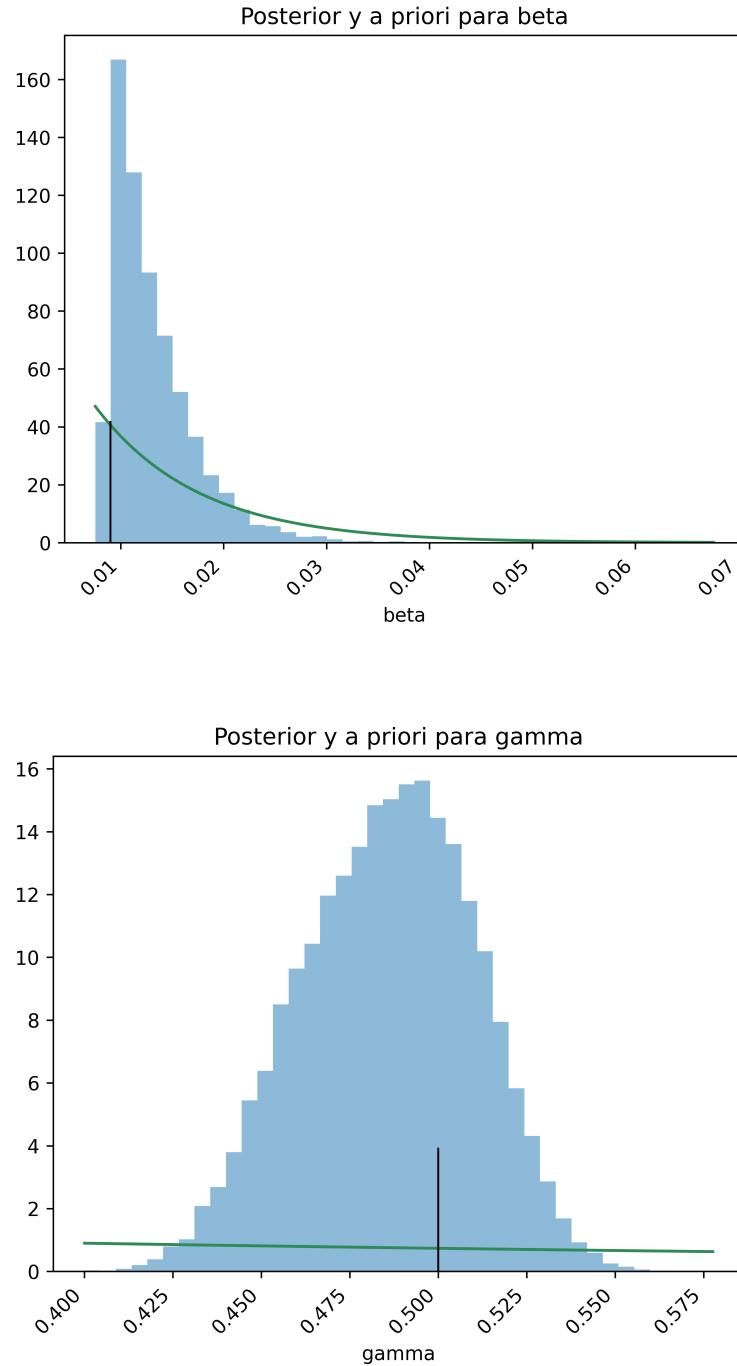


Analogamente a los modelos previos, una vez obtenida la forma funcional de la distribución posterior, usando el algoritmo Metropolis-Hastings para una cadena de tamaño $B = 600000$. La distribución posterior dada se muestra en la Fig. ?? que muestra una forma de media luna concentrada en el borde izquierdo.



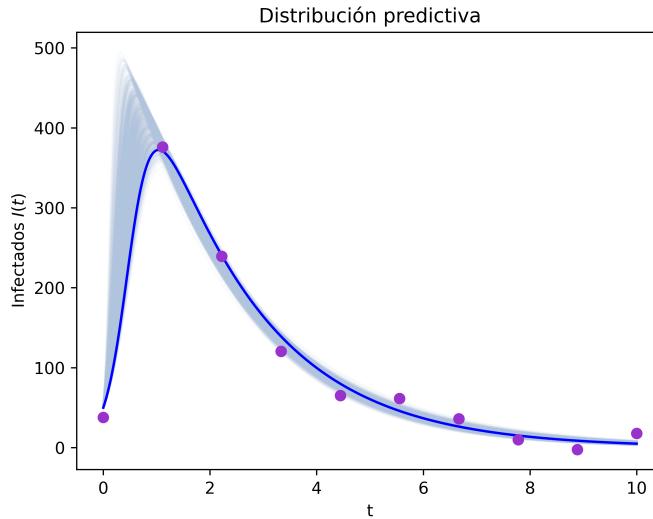
Para observar la distribución para cada parámetro se toma la distribución mar-

ginal, respectivamente. Observemos de la Fig. ?? y Fig. ?? la distribución de β se aglomera en valores cercanos a cero, mientras que para γ tenemos bastante diferencia respecto a la distribución a priori.



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. ?? vemos que el modelo (azul fuerte) se encuentra incluido en las distribuciones de infectados.



3.3. Inferencia con aproximación.

Con la propuesta aproximada para el análisis bayesiano para el problema inverso que resulta ser bastante conveniente dado que ya hemos visto que la metodología desarrollada que la distribución posterior depende del forward map (ordinario), y por ello el algoritmo de Metropolis-Hastings requiere que a cada paso en el espacio de estados Θ en la cadena X_t , se solucione la ecuación diferencial del modelo para los parámetros propuestos por MCMC. Sin embargo, tenemos cadenas de medio millón de pasos, lo que complica el análisis bayesiano. Por tanto, tenemos que el análisis aproximado ahorra tiempo de computo al evitar calcular el forward map iterativamente.

Con la regla de interpolación dada en ([referencia](#)) aproximamos el forward map repetidamente. Se pretende buscar la mejor aproximación a la distribución posterior modificando el forward aproximado tanto en la cantidad de puntos por rejilla así como en la cantidad de vecinos cercanos contemplados para la interpolación.

Las aproximaciones del forward map se hacen con inferencias de los k puntos más cercanos. Particularmente se experimentó con vecinos $k = 3, 5, 8$. Además, para la densidad de puntos de la rejilla, se toman sucesivamente rejillas de tamaño $m \times m$ con

$m = 10, 15, 30, 50$.

Imagen discretización forward map

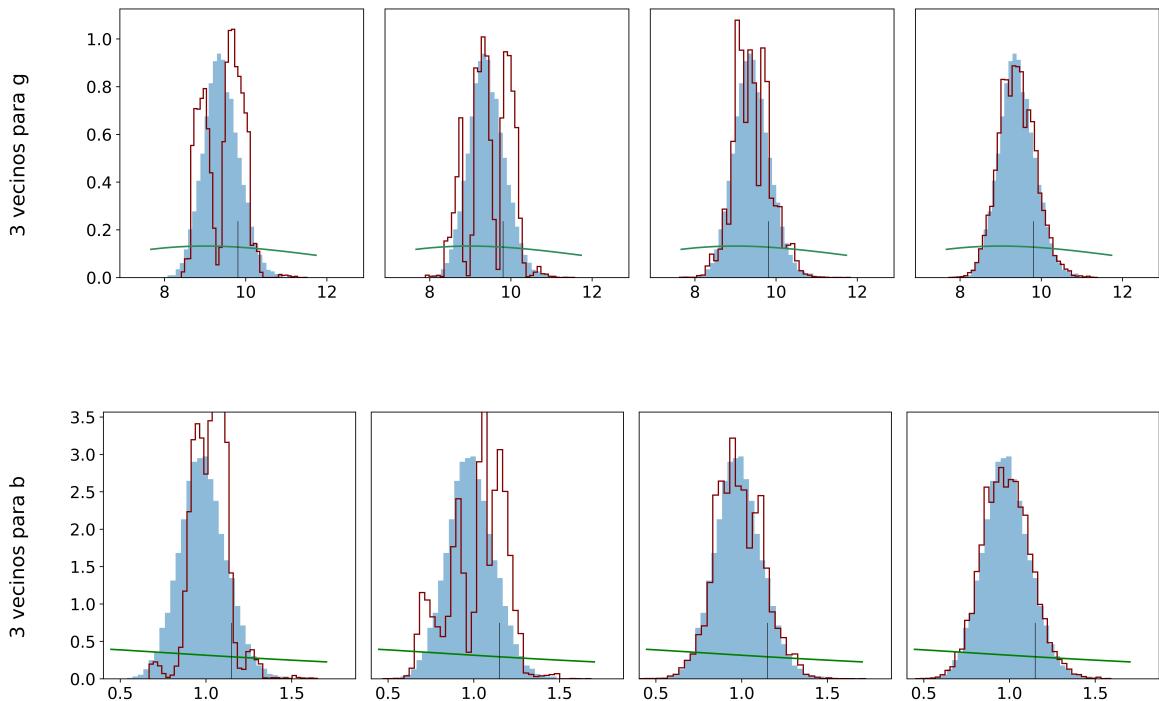
3.3.1. Método aproximado al problema inverso

Forward map aproximado para el modelo gravitatorio

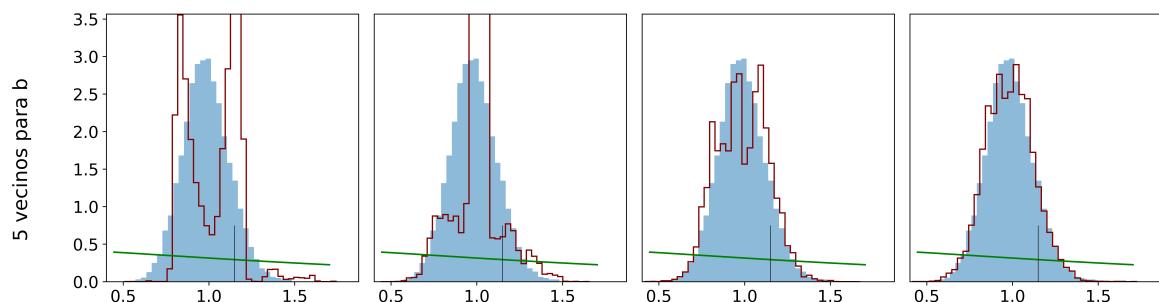
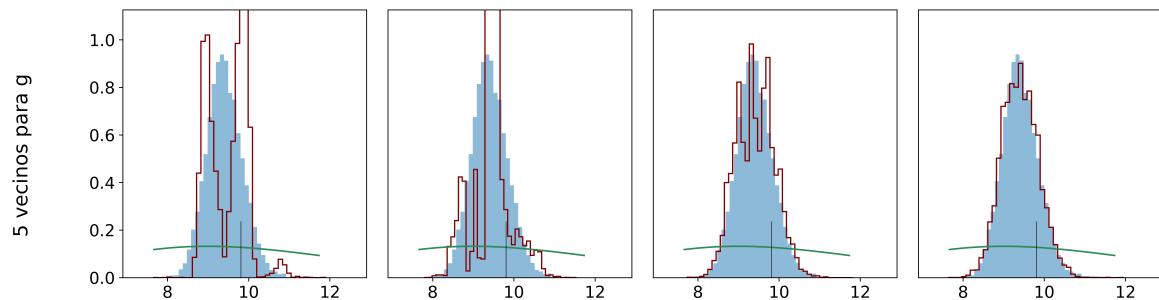
En el análisis bayesiano sobre el modelo gravitatorio se obtuvo la distribución posterior con el forward map ordinario para cada una de las aproximaciones propuestas.

Tomemos la aproximación del forward map con 3 vecinos cercanos ($k = 3$). Luego, calculemos la distribución posterior para cada rejilla ($m = 10, 15, 30, 50$). Tenemos la distribución bidimensional conjunta. Luego se toma 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, $\pi^*(g)$ con $\pi(g)$, y comparación de $\pi^*(b) = \pi(b)$.

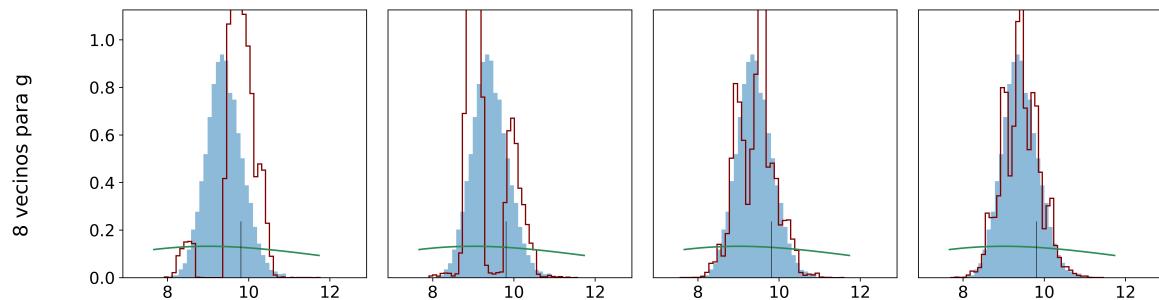
En la Fig. () se muestran las distribuciones marginales para la aproximación del forward map con $k = 3$ y de izquierda a derecha una malla de 10x10, 15x15, 30x30, 50x50.

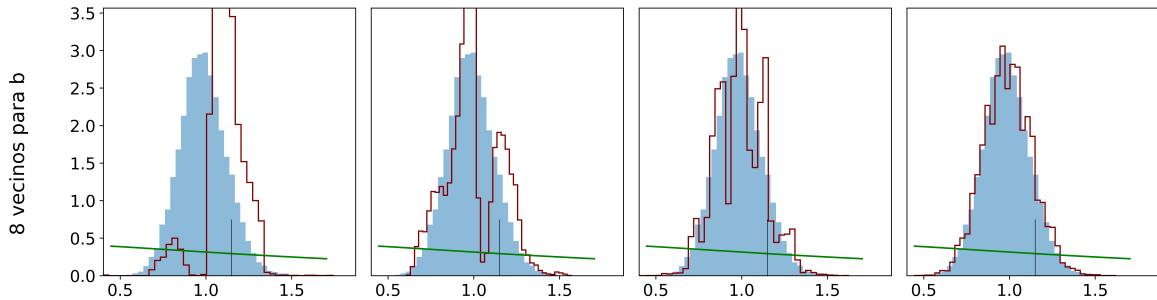


En la Fig. () se muestran las distribuciones marginales para la aproximación del forward map con $k = 5$ y de izquierda a derecha una malla de 10×10 , 15×15 , 30×30 , 50×50 .



En la Fig. () se muestran las distribuciones marginales para la aproximación del forward map con $k = 8$ y de izquierda a derecha una malla de 10×10 , 15×15 , 30×30 , 50×50 .





Notemos del análisis previo que de manera general las distribuciones posteriores aproximadas $\pi^*(g, b)$ mejora 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 50x50 puntos. Queda definir que cantidad de vecinos hace la aproximación mejor. Notese 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 50x50.

Es necesario establecer una estructura para realizar comparaciones entre el método ordinario y aproximado. Para saber la factibilidad del método se considera al tiempo de ejecución, que es una variable tangible.

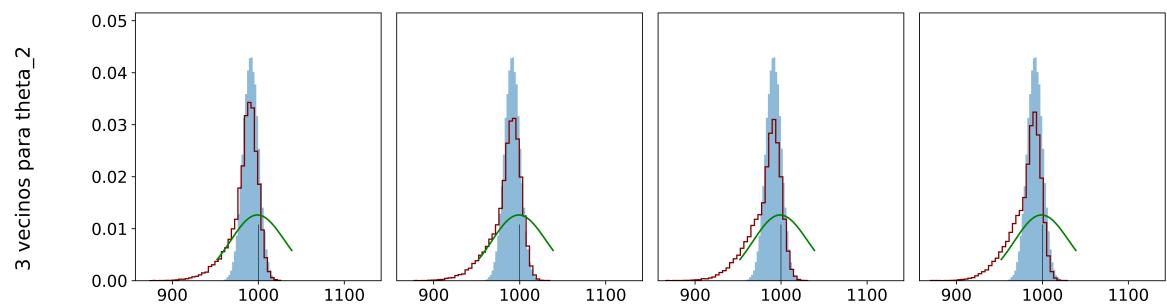
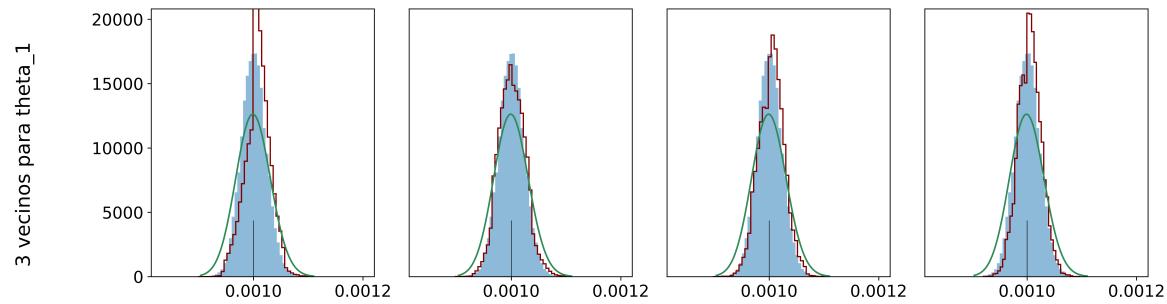
El tiempo de ejecución del método ordinario para el problema inverso en el modelo gravitacional es de **5 min 57 s**

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

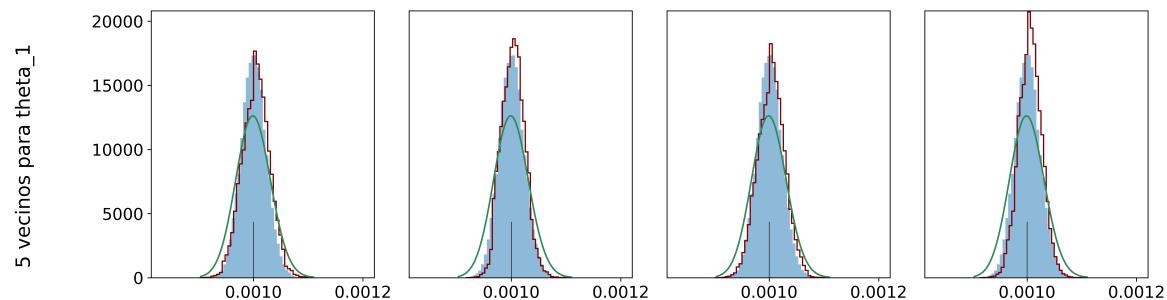
Vemos que la aproximación elegida tiene una ganancia de 57 segundos de ejecución, lo que representa una ganancia de una sexta parte del tiempo.

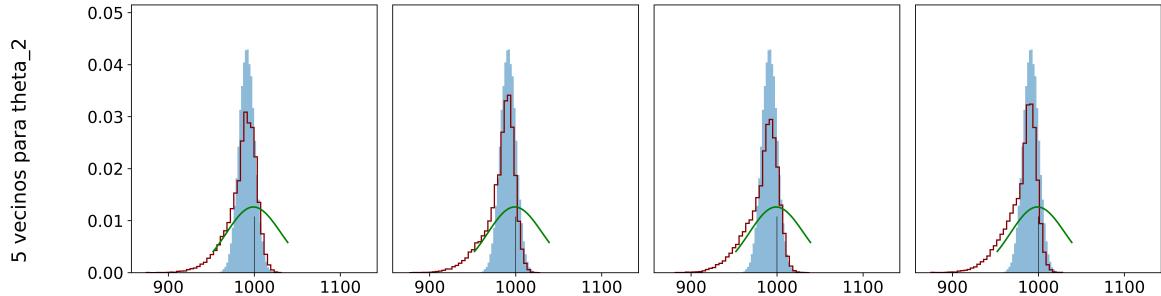
Forward map aproximado para el modelo logístico

En la Fig. () se muestran las distribuciones marginales para la aproximación del forward map con $k = 3$ y de izquierda a derecha una malla de 10×10 , 15×15 , 30×30 , 50×50 .

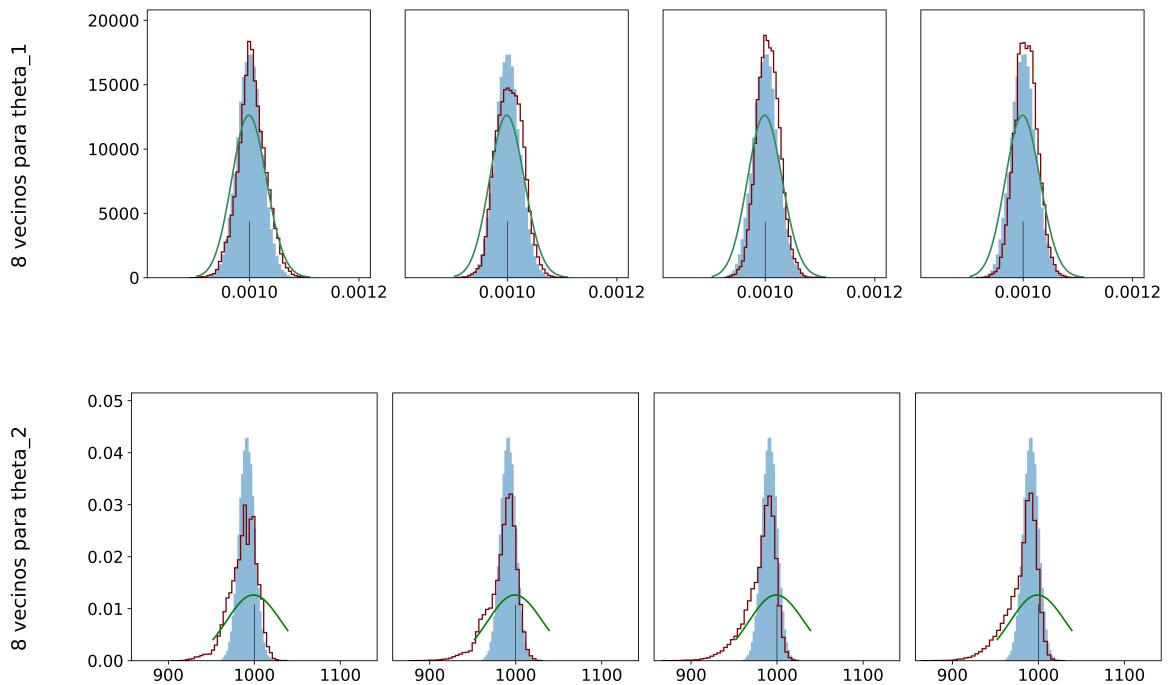


En la Fig. () se muestran las distribuciones marginales para la aproximación del forward map con $k = 5$ y de izquierda a derecha una malla de 10×10 , 15×15 , 30×30 , 50×50 .





En la Fig. () se muestran las distribuciones marginales para la aproximación del forward map con $k = 8$ y de izquierda a derecha una malla de 10×10 , 15×15 , 30×30 , 50×50 .



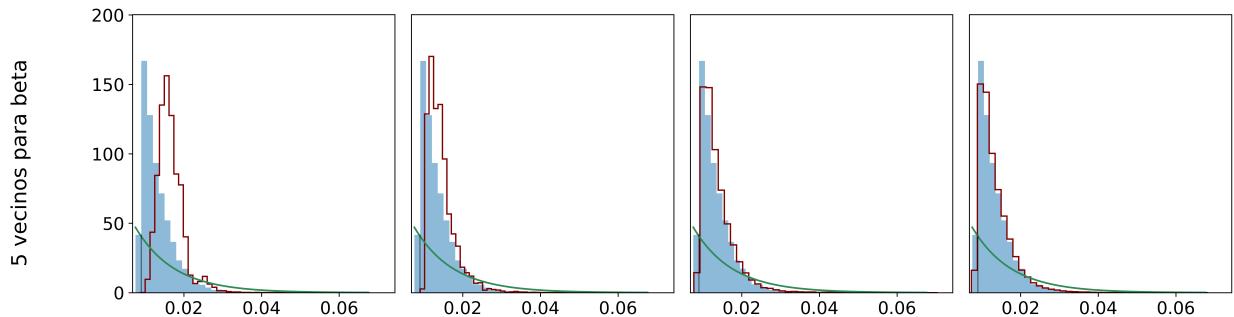
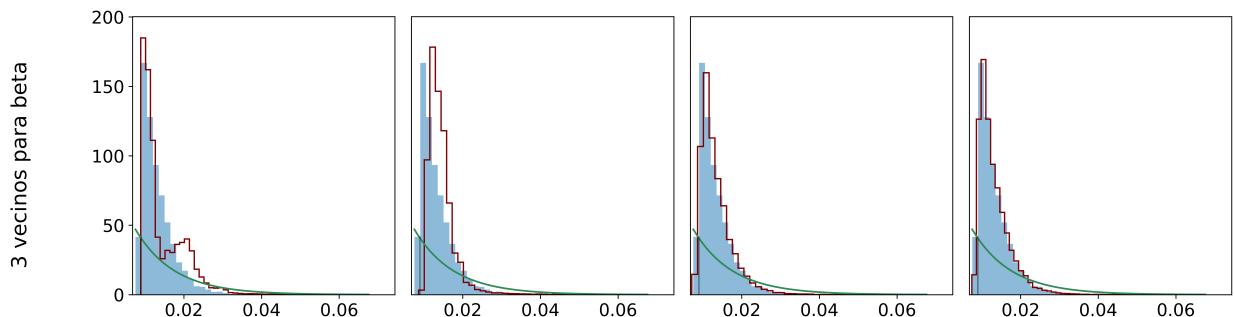
El tiempo de ejecución del método ordinario para el problema inverso en el modelo logístico es de **20 min 58 s**

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

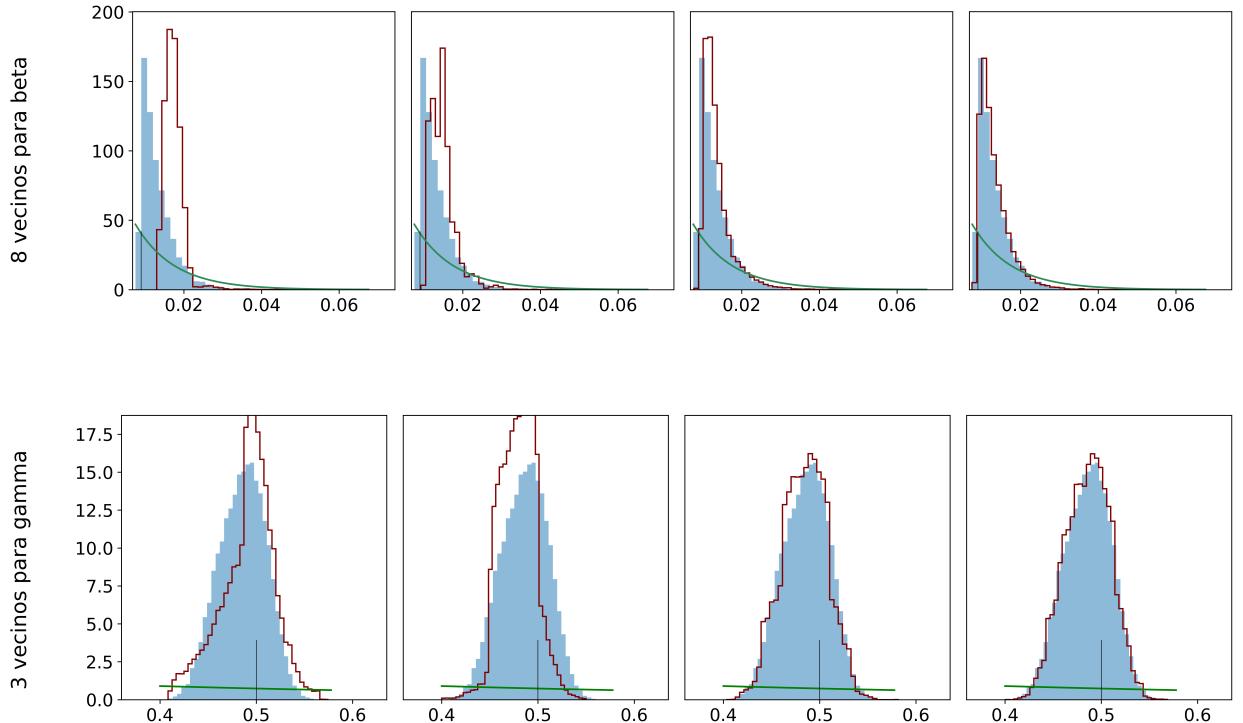
Convergencia en el modelo epidemiológico

El problema inverso en el modelo SIR se observa la relevancia de la aproximación del forward map. Recordemos que el forward map del modelo SIR no es posible obtenerse de manera analítica, lo que conlleva un consumo computacional añadido al momento de analizar los pormenores. Una dificultad para el forward map aproximado radica en el gran cambio de unidades, puesto que las unidades entre los parámetros β y γ difieren por un orden de magnitud, lo que es necesario del ajuste al espacio de parámetros.

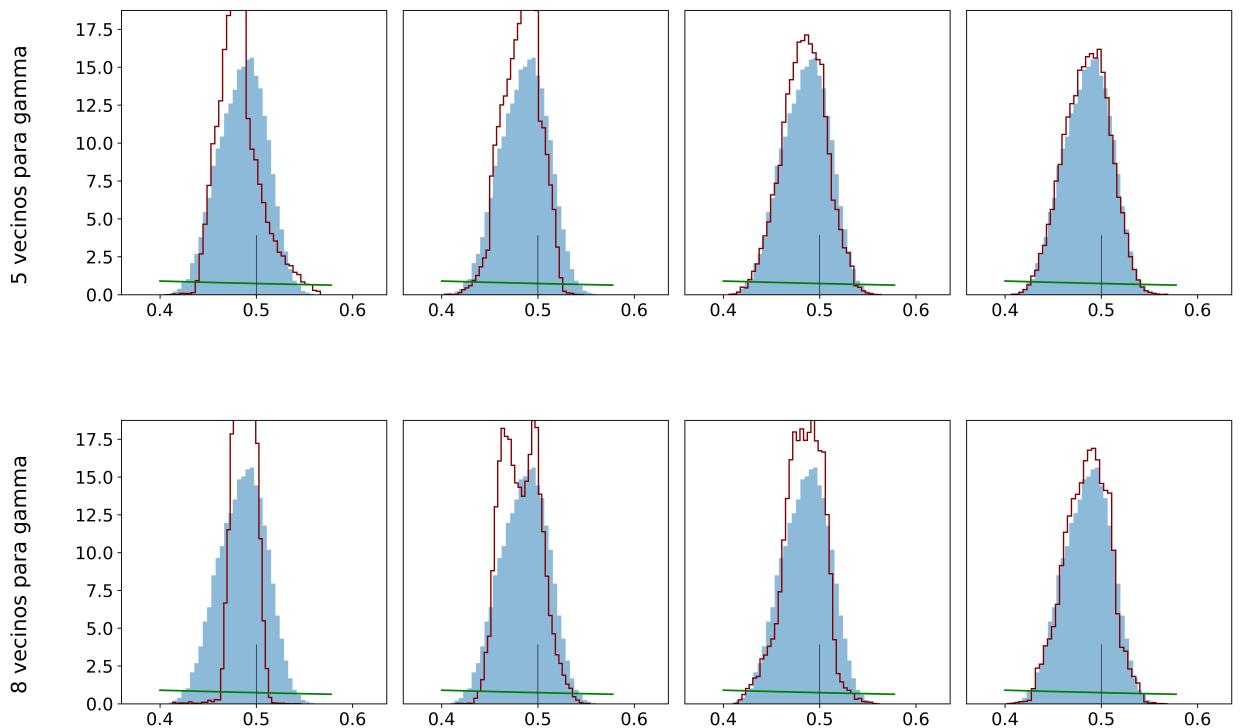
En la Fig. () se muestran las distribuciones marginales para la aproximación del forward map con $k = 5$ y de izquierda a derecha una malla de 10x10, 15x15, 30x30, 50x50.



En la Fig. () 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.



En la Fig. () se muestran las distribuciones marginales para la aproximación del forward map con $k = 5$ y de izquierda a derecha una malla de 10×10 , 15×15 , 30×30 , 50×50 .



El tiempo de ejecución del método ordinario para el problema inverso en el modelo gravitacional es de **14 min 05 seg**

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

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, debe de considerarse también el tiempo en conjunto necesario para conocer la mejor aproximación. Puesto que se realizaron los experimentos del forward map con la misma cantidad de vecinos y número de puntos en enmallado, y además encontramos que en cada uno de ellos se obtuvo una aproximación buena con la malla de tamaño 50x50 con 3 vecinos, suponemos que esto sucede en una amplia gamma de modelos. Así, al analizar un nuevo modelo, tomamos simplemente la aproximación mencionada esperando una mejora sustancial del tiempo de ejecución.

Existen varias maneras de implementar un mejor algoritmo, una de ellas considerada a 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.

Referencias

- Berger, J. O. (2013). *Statistical decision theory and bayesian analysis*. Springer Science & Business Media.
- Capistrán, M. A., Capella, A., y Christen, J. A. (2020). Forecasting hospital demand during covid-19 pandemic outbreaks. *arXiv preprint arXiv:2006.01873*.
- Casella, G., y Berger, R. (2024). *Statistical inference*. CRC Press.
- 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/>)
- Kelley, W. G. (2010). *The theory of differential equations*. Springer.
- 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.
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.
- 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.
- 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.

Apéndice A

Anexo

