

Protocolo de Tesis

César Isaí García Cornejo
Asesor: Dr. José Andrés Christen Gracia

19 de febrero de 2024

Palabras clave— problema inverso, inferencia bayesiana, ecuaciones diferenciales, MCMC

1. Introducción

Los fenómenos de la naturaleza son intrínsecamente causales. Las causas; que son las condiciones, cambios o variables; afligen una reacción a la cual llamamos efecto. El efecto es producto enteramente de la causa. Dicho de otra forma, sin una causa no existe el efecto, por lo que estas se encuentran en un orden tanto lógico como cronológico. Asimismo, los modelos que pretendan describir los fenómenos naturales deben ser causales, respetando orden entre causa y efecto, haciendo viable la posibilidad de generar predicciones elocuentes.

Típicamente, los modelos matemáticos precisan de condiciones iniciales, parámetros u otras formas de variables para determinar unívocamente el problema cuya resolución propicia a la predicción de una o más cualidades del fenómeno modelado. Decimos que aquellas variables previas son las causas, mientras que el efecto son las predicciones dadas por la resolución del problema matemático. El proceso anteriormente descrito se llama problema directo o ‘forward problem’ ya que sigue la dirección de causalidad. Sin embargo, es interesante también el problema inverso, dada ciertas observaciones de las cualidades de un fenómeno (efectos), ¿es posible calcular las causas del modelo que rige el fenómeno?

El problema inverso es uno de los problemas matemáticos más importantes para la ciencia, debido a que ayuda a dilucidar valores de parámetros que no se pueden medir directamente. Además de que tiene un amplio campo de aplicación, entre ellos: óptica, acústica, procesamiento de señales, imágenes medicas, geofísica, oceanografía, astronomía, aprendizaje máquina y un largo etc. A pesar de su gran variabilidad de aplicaciones del problema inverso, las vicisitudes de este lo hacen un problema complejo. Mientras que el forward problem tiene solución única (en caso determinista) el problema inverso no necesariamente. Aunado al hecho de que las observaciones conllevan incertidumbre por errores de medición. Por lo anterior, es que es necesario hacer explícito toda la información a priori de los parámetros del modelo al igual que la aceptable representación de las incertidumbres en los datos.

La teoría más general se obtiene tras representar la información a priori del modelo por una distribución de probabilidad sobre el espacio de modelos, llamada distribución a priori. Dicha formulación se engloba en la estadística bayesiana, que muestra como se transforma la distribución a priori tras la consideración de las observaciones relacionadas al modelo de interés, dando la denominada distribución posterior. La metodología usual para obtener estimaciones de los parámetros se da tras la simulación de realizaciones de la distribución a posterior por métodos monte carlo, siendo el más famoso MCMC Metropolis-Hastings [1].

2. Antecedentes

Sea \mathcal{S} un sistema físico bajo estudio. Ya sea \mathcal{S} una galaxia para un astrofísico, el planeta Tierra para un geofísico, o una partícula cuántica para un físico cuántico. El procedimiento para el estudio de un sistema físico puede ser dividido en los siguientes tres pasos.

1. Parametrización del sistema: descubrimiento del conjunto mínimo de parámetros del modelo cuyos valores caracterizan completamente el sistema.
2. Modelación directa (forward modeling): Descubrimiento de las leyes físicas que nos permiten hacer predicciones, dado valores de parámetros del modelo, de mediciones en observables físicas.
3. Modelación inversa: Usar las mediciones de las observables físicas para inferir los valores de los parámetros del modelo.

Existe una fuerte retroalimentación entre estos pasos. Mientras que los primeros dos pasos son inductivos, el tercero es deductivo. Esto significa que las reglas de razonamiento que se siguen en los primeros dos pasos son difíciles de hacer explícitas. En contraste con la teoría matemática donde la lógica aplica bien en el tercer paso [2].

2.1. Espacio del modelo y espacio de datos

La elección de los parámetros de los modelos suele no ser única. Dos parametrizaciones diferentes de dicen que son equivalentes si existe una biyección que las relaciona.

Independientemente de la parametrización, es posible introducir un espacio de puntos, una variedad, donde cada punto representa un modelo del sistema. La variedad es llamada el espacio del modelo y se denota por \mathfrak{M} . Modelos individuales son puntos del espacio del modelo y se denotan por $\mathcal{M}_1, \mathcal{M}_2, \dots$. Sin embargo, antes de preguntarse si el espacio del modelo es lineal, más importante es la distribución de probabilidad sobre este espacio. Una probabilidad sobre \mathfrak{M} es el mapeo que para cualquier subconjunto \mathcal{A} de \mathfrak{M} , llamada probabilidad de \mathcal{A} , con $\mathbb{P}(\mathfrak{M}) = 1$. Tal distribución de probabilidad puede ser definida sobre cualquier variedad finito dimensional \mathfrak{M} e independientemente de cualquier parametrización de \mathfrak{M} .

Para obtener información en los parámetros del modelo, es necesario obtener observaciones de las observables físicas en el experimento. Podemos definir el espacio de datos como el espacio de todas las posibles respuestas del instrumento de medición. Esto corresponde a otra variedad que se representa por el símbolo \mathfrak{D} . Cualquier resultado de las mediciones corresponde entonces, a un punto \mathcal{D} del espacio \mathfrak{D} [2].

2.2. Problema directo (forward problem)

Los experimentos sugieren teorías físicas, y las teorías físicas predicen los resultados de los experimentos. La comparación entre la predicción y las observaciones permiten mejorar la teoría.

Tomando una perspectiva simplista, resolver el problema directo significa predecir los valores de las observables físicas sin error \mathbf{d} que corresponde a un modelo \mathbf{m} . Esta predicción se denota por

$$\mathbf{m} \mapsto \mathbf{d} = \mathbf{F}(\mathbf{m})$$

donde $\mathbf{d} = \mathbf{F}(\mathbf{m})$ es la notación reducida para $d^i = F^i(m^1, m^2, \dots)$ ($i = 1, 2, \dots$) con m_i y d_i coordenadas en su respectivo espacio. El operador $F(\cdot)$ se llama operador forward y expresa el modelo matemático del sistema físico estudiado.

Los valores predichos no serán, en general, idénticos a los valores observados por las ya mencionas complicaciones: incertidumbre de las mediciones e imperfecciones del modelo. Estas dos diferentes fuentes de error producen incertidumbre del mismo orden de magnitud, debido a que tan pronto como la investigación científica progresa, los métodos experimentales decrecen la incertidumbre en sus mediciones [2].

2.3. El método bayesiano

La metodología tradicional para abordar el problema inverso es por medio de la inferencia bayesiana. Dicho método se sigue de la siguiente manera:

1. Elegimos una densidad para los parámetros $f(\theta)$ llamada distribución a priori, expresa la certidumbre que se tiene acerca del parámetro θ antes de observar los datos.
2. Se selecciona el modelo estadístico $f(x|\theta)$ que refleja la certidumbre de x dado θ .
3. Después de observar los datos X_1, \dots, X_n , actualizamos la incertidumbre calculando la distribución posterior $f(\theta|X_1, \dots, X_n)$.

El tercer paso de esta metodología se obtiene del Teorema de Bayes, sin embargo suele ser que la distribución posterior es conocida salvo una constante de normalización. Dicha problema dificulta la posibilidad de dar estimadores analíticos. De esta forma es que entran Markov Chain Monte Carlo (MCMC) para hacer simulaciones de la distribución posterior, obteniendo estimaciones de la propia distribución así como estimaciones puntuales de los parámetros del modelo [3]. En la próxima sección se ahondará más en los detalles.

3. Planteamiento del problema

Se propone implementar por medio de código una completa resolución al problema inverso para varios modelos físicos establecidos por ecuaciones diferenciales ordinarias. Dicha implementación propone la investigación por medio de experimentación numérica para disminuir el costo de computación al resolver ecuaciones diferenciales por cada parámetro en el espacio de parámetros. Para dicho fin, se propone usar métodos análogos a k-vecinos más cercanos para la estimación de la ecuación diferencial. Los detalles de la propuesta se presentan en las siguientes secciones.

3.1. Consideraciones generales

En un amplio catalogo de modelos físicos, nos encontramos que gran proporción de modelos se basan en ecuaciones diferencial ordinarias dependientes de al menos un parámetro. Consideremos un modelo de este tipo. El forward problem consiste en dar un valor del parámetro o parámetros, para posteriormente resolver la ecuación diferencial. Entonces la trayectoria o solución de la ecuación diferencial será la predicción del modelo, y la causa será el valor de los parámetros.

Consideremos ahora que tenemos pares de observaciones puntuales de una trayectoria que se rige por el modelo establecido por la ODE $d_i, (i = 1, \dots, n)$. Pensemos que son medidas de la trayectoria de una partícula cuya posición evoluciona en el tiempo. Entonces, cada observación $d_i = (x_i, t_i)$ donde interpretamos a x_i como la posición de la partícula al tiempo t_i . Tomemos la discretización del forward map $F_\theta(t_i)$ para el valor del forward map evaluado en el parámetro θ para el tiempo t_i .

Una vez que se ha establecido el forward map es necesario esclarecer como será la incertidumbre por errores de medición así como de deficiencias del modelo, en las observaciones de los datos.

La manera usual para establecer el error es mediante la siguiente ecuación

$$x_i = F_\theta(t_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (1)$$

La ecuación (1) para los errores representa que las diferencias entre lo observado y lo predicho tienen dependencias lineales. Consideramos que la distribución de los errores sigue una distribución normal. Es decir,

$$\varepsilon \sim N(0, \sigma^2) \quad (2)$$

Una estructura alterna para las diferencias entre lo predicho y lo observado sigue la siguiente formulación:

$$x_i = F_\theta(t_i + \varepsilon_i)$$

sin embargo, tal formulación conlleva otras complicaciones por lo que se encuentra en segundo plano de la implementación.

Sea $\theta = (\theta_1, \dots, \theta_m)$ los m parámetros de los cuales depende el modelo a considerar. Por la metodología bayesiana es necesario dar una medida de la certidumbre para cada parámetro. Es decir se considera que los parámetros provienen de una variable aleatoria $\Theta = (\Theta_1, \dots, \Theta_m)$ y según la información a priori del modelo se propone explícitamente la distribución a prior $\pi_\Theta(\theta)$. Por supuesto, las variables entre parámetros son independientes, es decir podemos escribir

$$\pi_\Theta(\theta) = \pi_{\Theta_1}(\theta_1) \cdot \dots \cdot \pi_{\Theta_m}(\theta_m)$$

En consecuencia, para el cálculo de la distribución posterior solo hace falta calcular la verosimilitud del modelo en (1). De la ecuación (1) y (2) se sigue que

$$x_i \sim N(F_\theta(t_i), \sigma^2)$$

por tanto, la verosimilitud es

$$\begin{aligned} \mathcal{L}(\theta|x^n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - F_\theta(t_i))^2} \\ &= \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - F_\theta(t_i))^2} \end{aligned}$$

donde x^n denota (x_1, \dots, x_n) .

Obteniendo así, la distribución posterior

$$\begin{aligned} \pi(\theta|x^n) &\propto \mathcal{L}(\theta|x^n) \pi_\Theta(\theta) \\ &\propto \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - F_\theta(t_i))^2} \pi_{\Theta_1}(\theta_1) \cdot \dots \cdot \pi_{\Theta_m}(\theta_m) \end{aligned}$$

de donde una vez dada las distribuciones a priori, es posible simular variables aleatorias con dicha distribución y así obtener estimaciones de los parámetros del modelo. Dicha simulación se plantea obtener por medio de MCMC Metropolis-Hastings. Finalmente, de la simulación de la distribución posterior, se dan estimaciones puntuales de los parámetros.

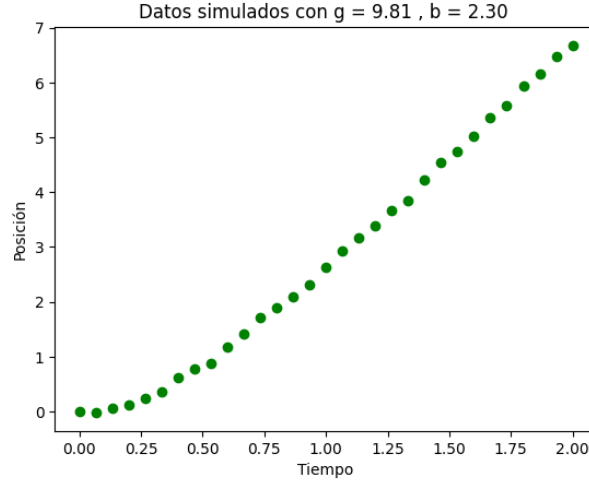


Figura 1: Datos simulados para caída sujeto a fricción.

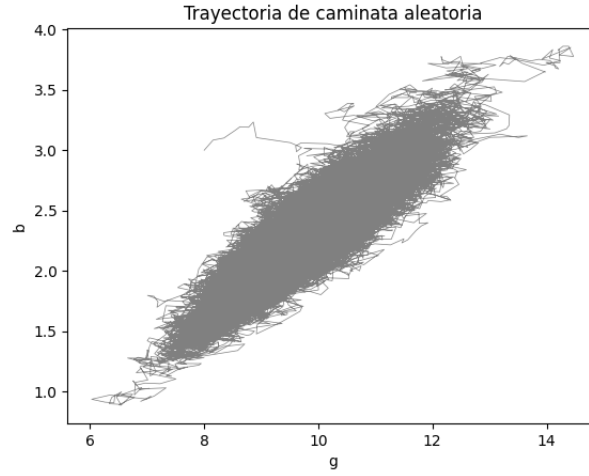


Figura 2: Trayectoria de la caminata aleatoria del MCMC en el espacio de parámetros.

3.2. Consideraciones particulares

Se plantea dar una par de modelos dados por ODE's con parámetros del modelo. Uno de los modelos propuestos para resolver el problema inverso es el dado por la caída de una partícula en un campo gravitatorio constante sujeto a una fuerza de fricción dada por la dinámica clásica. La trayectoria de la partícula $x(t)$ satisface la segunda ley de Newton

$$m\ddot{x} = g - b\dot{x} \quad (3)$$

donde g es la constante de gravitación y b es la constante de fricción. Estas son los parámetros del modelo. El problema forward consiste en dar valores al par de parámetros de modelos y obtener la trayectoria de la posición resolviendo la ecuación diferencial. Por simplicidad se considera de (3) a $m = 1$ y con las condiciones inicial $x(0) = \dot{x}(0) = 0$.

Los datos los obtenemos de simulación, tomaremos dos valores de parámetros g, b y aplicaremos el forward map para obtener puntos en la trayectoria a la cual se procede a agregar ruido gaussiano (Fig. 1).

Las distribuciones a priori tienen que tener soporte en los positivos, por ello es que se proponen a ambos parámetros del modelo con distribuciones gama. Posteriormente, usando Metropolis-Hastings se genera una caminata aleatoria en el espacio de parámetros del modelo, donde a cada paso resuelve el problema forward para los parámetros donde se encuentra la camina y así evaluar lo distribución posterior (Fig. 2)

Con la simulación de la distribución posterior, mostrada en las Fig. 3 y 4 podemos tomar el estimador de Bayes como la media de la distribución, siendo está la estimación para los parámetros.

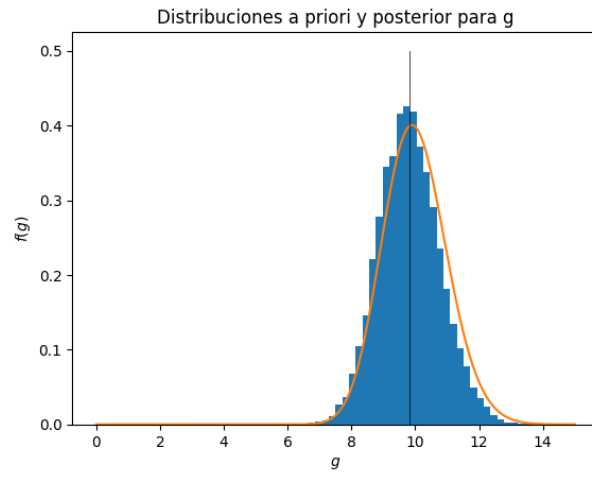


Figura 3: Distribución a priori y distribución posterior para el parámetro g

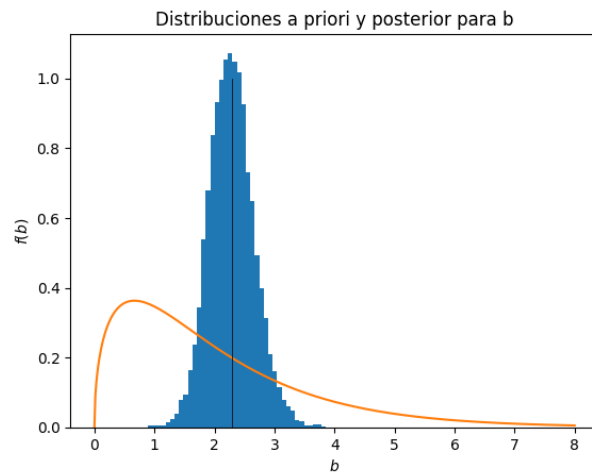


Figura 4: Distribución a priori y distribución posterior para el parámetro b

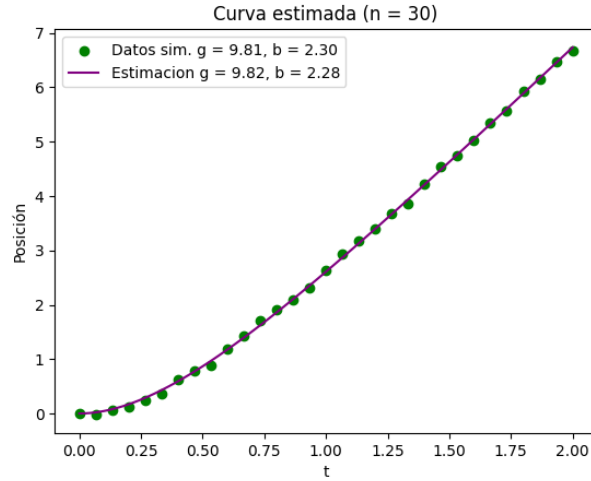


Figura 5: Estimación al problema inverso para el modelo de caída sujeto a fricción.

Finalmente, se coteja la calidad de la estimación con los parámetros con los que desde un principio se simulaban los datos. En la Fig. 5 se contrastan los datos y la curva estimada dada por el problema inverso.

4. Objetivo

A pesar de que el análisis previo nos permite obtener estimaciones bastante precisas para el problema inverso, son computacionalmente pesadas, esto debido a que a cada paso en la cadena requiere que se solucione el problema forward, es decir, se resuelven ecuaciones diferenciales tantas veces como se deje correr la cadena. Es así, que se desea encontrar una especie de *interpolación* para solo solucionar una cantidad fija de veces el problema forward y para cada punto en el espacio de parámetros se pueda estimar la solución en función de las soluciones con parámetros cercanos.

Referencias

- [1] Christian P Robert, George Casella, and George Casella. *Monte Carlo statistical methods*, volume 2. Springer, 1999.
- [2] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.
- [3] Larry Wasserman. *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2004.