

El algoritmo Metropolis-Hastings: problemas inversos, calibración de modelos y el modelo TOMGRO

Trabajo realizado por **Carlos Nosa** ♦

Trabajo dirigido por **Juan Galvis** ♦ y **Rodrigo Gil** *

♦ Departamento de Matemáticas, Universidad Nacional de Colombia, Sede Bogotá

* Departamento de Agronomía, Universidad Nacional de Colombia, Sede Bogotá

18 de febrero del 2025

Contenidos

1 Planteamiento del problema

2 Fundamentos

- Inferencia bayesiana
- Métodos de Montecarlo basados en cadenas de Markov (MCMC)

3 Resultados

- Calibración de modelos
 - Ejemplo: Crecimiento logístico
- Calibración del modelo TOMGRO

4 Conclusiones y Trabajo Futuro

5 Referencias

Pregunta de investigación

¿Cómo estimar los parámetros de modelos de ecuaciones diferenciales, como TOMGRO, utilizando inferencia bayesiana basada en métodos de Montecarlo usando cadenas de Markov?

Pregunta de investigación

¿Cómo estimar los parámetros de modelos de ecuaciones diferenciales, como TOMGRO, utilizando inferencia bayesiana basada en métodos de Montecarlo usando cadenas de Markov?

- Inferencia bayesiana.
- Métodos de Montecarlo basados en cadenas de Markov.
- Modelo de ecuaciones diferenciales TOMGRO (*Tomato Growth*).

Inferencia bayesiana

Consideraremos el modelo

$$Z = \mathcal{F}(\Theta, \mathcal{E})$$

donde

- $Z \in \mathbb{R}^n$ es el **vector aleatorio** de dónde provienen los datos z_{obs} ,
 - $\Theta \in \mathbb{R}^m$ es el **vector aleatorio** de parámetros del modelo,
 - $\mathcal{E} \in \mathbb{R}^k$ es el **vector aleatorio** de ruido de dónde proviene el ruido de los datos ϵ , y,
 - $F : \mathbb{R}^m \times \mathbb{R}^k \rightarrow \mathbb{R}^m$ una función.

Inferencia bayesiana

Sea $\pi_{\Theta, Z}(\theta, z)$ la densidad de probabilidad conjunta de los parámetros y los datos.

- ### ■ Distribución *a priori* del vector de parámetros Θ .

$$\pi_{prior}(\theta) = \pi_\Theta(\theta) = \int_{\mathbb{R}^n} \pi_{\Theta,Z}(\theta, z) dy.$$

- La distribución de probabilidad condicional o función de verosimilitud.

$$\pi_{Z|\Theta}(z|\theta) = \frac{\pi_{\Theta,Z}(\theta, z)}{\pi_\Theta(\theta)}.$$

- Si se tiene que $Z = z_{obs}$, la distribución de probabilidad condicional

$$\pi_{post}(\theta) = \pi_{\Theta|Z}(\theta|Z_{obs}),$$

se llama la distribución **a posteriori** de Θ .

Inferencia bayesiana

- Por el Teorema de Bayes,

$$\pi_{\Theta|Z}(\theta|z_{obs}) = \frac{\pi_{Z|\Theta}(z_{obs}|\theta)\pi_{\Theta}(\theta)}{\pi_Z(z_{obs})}.$$

- Algunas de las siguientes estadísticas son de relevancia:

$$\underbrace{\Theta_{MAP} = \max_{\theta \in \mathbb{R}^m} \{\pi_{\Theta|Z}(\theta|z_{obs})\}}_{\text{Máximo a posteriori}}$$

y

$$\underbrace{\Theta_{CM} = E[\pi_{\Theta|Z}(\theta|z_{obs})]}_{\text{Media condicional}}.$$

Inferencia bayesiana

- El **problema inverso** de estimar los parámetros Θ dados unos datos observados $Z = z_{obs}$ en inferencia bayesiana tiene como solución la distribución posterior $\pi_{post}(\theta)$.
- Se **explora** la distribución a posteriori $\pi_{\Theta|Z}(\theta|z_{obs})$ por medio de

$$\pi_{post}(\theta) = \pi_{\Theta|Z}(\theta|z_{obs}) \propto \pi_{Z|\Theta}(z_{obs}|\theta)\pi_{\Theta}(\theta).$$

- Usando el problema directo e información sobre z_{obs} , se modela la verosimilitud $\pi_{Z|\Theta}(z_{obs}|\theta)$.
- Se construye la distribución a priori $\pi_{prior}(\theta)$ usando conocimiento previo sobre los parámetros Θ .

Introducción a los métodos de Montecarlo basados en cadenas de Markov (MCMC)

- Son un tipo de métodos que se utilizan para obtener muestras de una distribución de probabilidad .
 - Dada una distribución de probabilidad, se puede construir una cadena de Markov cuya **distribución de equilibrio** coincide con la distribución objetivo.
 - En el contexto de la inferencia bayesiana, tiene como objetivo **muestrear la distribución a posteriori** usando el Teorema de Bayes:

$$\underbrace{\pi_{\Theta|Z}(\theta|Z_{obs})}_{A \text{ posteriori}} \propto \underbrace{\pi_{Z|\Theta}(Z_{obs}|\theta)}_{Vero\text{similitud}} \underbrace{\pi_{\Theta}(\theta)}_{A \text{ priori}}.$$

- En este trabajo de investigación se usaron el algoritmo **Metropolis Hastings**, el muestreador de **Gibbs** y el algoritmo **two-walk**, centrándose en el primero.

Teorema. Sea μ una medida de probabilidad en \mathbb{R}^n y $\{X_j\}$ una cadena de Markov homogénea en el tiempo con núcleo de transición P . Si μ es una medida invariante de P , y P es irreducible y aperiódica, entonces:

$$\lim_{N \rightarrow \infty} P^{(N)}(x, B) = \mu(B) \quad \text{para todo } B \in \mathcal{B}.$$

Y para $f \in L^1(\mu(dx))$,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N f(X_j) = \int_{\mathbb{R}^n} f(x) \mu(dx) \quad \text{casi seguramente.}$$

- Para una medida de probabilidad μ absolutamente continua ($\mu(d\theta) = \pi(\theta)d\theta$) se busca un núcleo de transición de probabilidad P de tal manera que:

$$\mu P(B) = \mu(B)$$

para todo conjunto de Borel B . Esta condición se conoce como que la medida μ es **estacionaria** respecto al núcleo de transición de probabilidad P .

- Si se tiene la representación del núcleo P como:

$$\begin{aligned} P(\theta, B) &= \int_B Q(\theta, \theta^*) d\theta^* + r(\theta) \chi_B(\theta) \\ &= \int_B Q(\theta, \theta^*) d\theta^* + \left(1 - \int_{\mathbb{R}^m} Q(\theta, \theta^*) d\theta\right) \chi_B(\theta), \end{aligned}$$

es suficiente que el núcleo de transición Q cumpla las **ecuaciones de balance detallado** para todo θ y θ^* :

$$\pi(\theta^*) Q(\theta^*, \theta) = \pi(\theta) Q(\theta, \theta^*),$$

para que se cumpla que la medida μ es **estacionaria** respecto al núcleo de transición de probabilidad P .

- Dado el kernel de transición Q , si no se satisface que:

$$\pi(\theta^*)Q(\theta^*, \theta) = \pi(\theta)Q(\theta, \theta^*),$$

entonces se construye una función α de tal manera que:

$$\pi(\theta^*)Q(\theta^*, \theta)\alpha(\theta^*, \theta) = \pi(\theta)Q(\theta, \theta^*)\alpha(\theta, \theta^*),$$

una opción es

$$\alpha(\theta, \theta^*) = \min \left(1, \frac{\pi(\theta^*)Q(\theta^*, \theta)}{\pi(\theta)Q(\theta, \theta^*)} \right).$$

Algoritmo Metropolis-Hastings (MH)

Datos: Punto inicial θ_0 , número de muestras $N \in \mathbb{N}$, un kernel de transición Q y densidad π .

Resultado: Muestras $\{\theta_1, \dots, \theta_N\}$ de la distribución π .

Proceso: Para $i = 1, \dots, N$:

- Generar una propuesta θ_{i+1} a partir de la anterior muestra θ_i usando Q . Construcción de la **cadena de Markov**.
 - Comparar las muestras θ_i y θ_{i+1} usando:

$$\alpha(\theta_i, \theta_{i+1}) = \min \left(1, \frac{\pi(\theta_{i+1})Q(\theta_{i+1}, \theta_i)}{\pi(\theta_i)Q(\theta_i, \theta_{i+1})} \right)$$

- Aceptar la propuesta con una probabilidad de α .

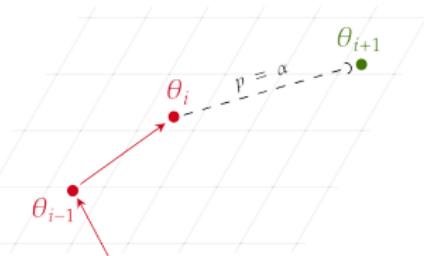


Figura: Camino aleatorio.

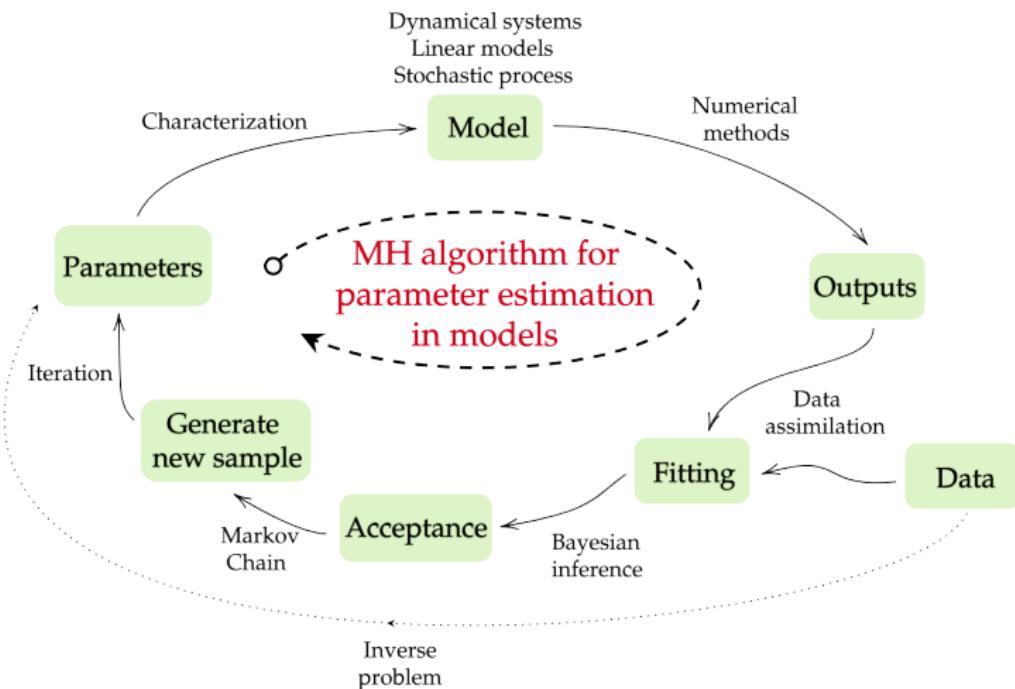


Figura: Uso del enfoque bayesiano para la estimación de parámetros de modelos.
Realización propia.

Descripción del modelo TOMGRO



Figura: Invernadero en el campus de la UN en Bogotá. Mayo 17, 2024.

- Matemáticamente, el modelo es un sistema dinámico de la forma:
$$\frac{dy}{dt} = F(C, \theta, y), \quad y(t_0) = y_0.$$
- El modelo utiliza un enfoque oferta-demanda para repartir los carbohidratos en el crecimiento de diferentes órganos.
- Es usado para crear estrategias óptimas de cultivo.

Crecimiento Logístico

Ejemplo

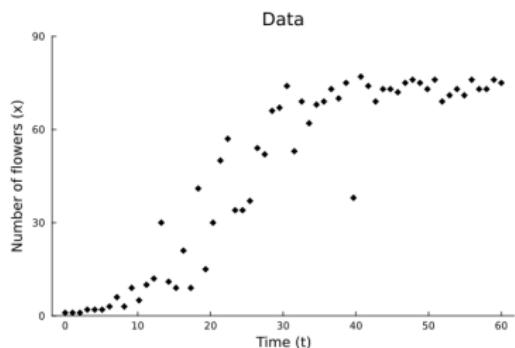


Figura: Datos simulados del modelo logístico.

- Considere el modelo logístico

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{100K}\right), \quad x(t_0) = x_0,$$

donde x es el número de flores, t el tiempo (semana), r representa la razón de crecimiento y $100K$ la capacidad máxima.

- Se simulan datos de este modelo con parámetros de distribución:

$$r \sim Beta(\alpha = 25, \beta = 100), \\ K \sim Beta(\alpha = 300, \beta = 100).$$

Ejecución del algoritmo MH

- Se ejecuta el algoritmo MH con kernel de transición Q proveniente de una distribución gaussiana de media θ_i y varianza fija $\sigma^2 = 0.02$. El punto inicial es $[0.9, 0.1]$ y se hace un total de 5000 iteraciones.
- Para construir la distribución a posteriori se usa lo siguiente:
 - Para construir la verosimilitud se supone que en cada tiempo vienen de una distribución Poisson:

$$Z_t | \Theta \sim \text{Poisson}(x(t; \Theta)).$$

- Para la distribución a priori se supone que los parámetros r y K son independientes y además

$$r \sim \text{Gamma}(1, 1),$$

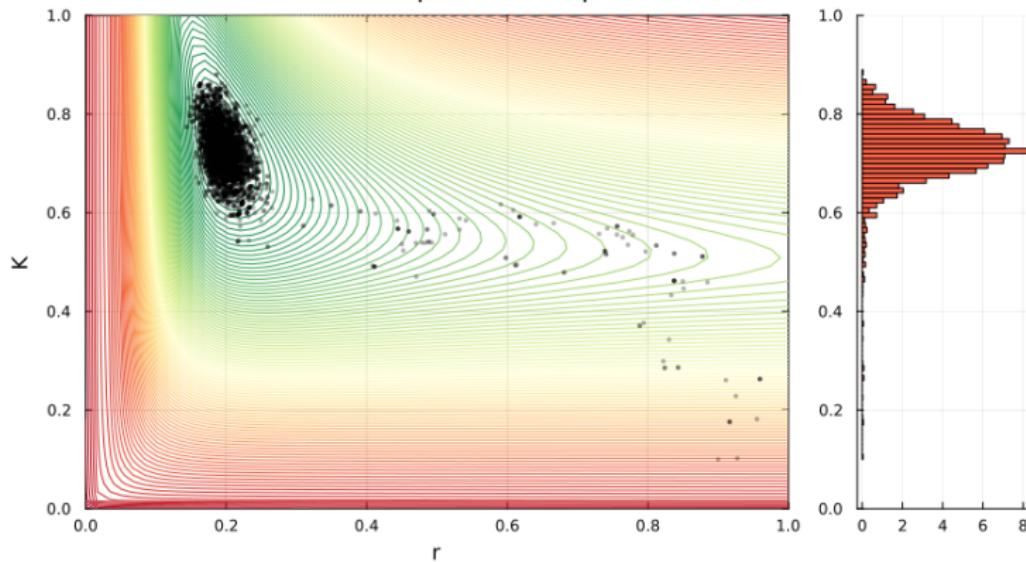
$$K \sim \text{Gamma}(1, 1).$$

Resultado del algoritmo MH

Espacio de parámetros

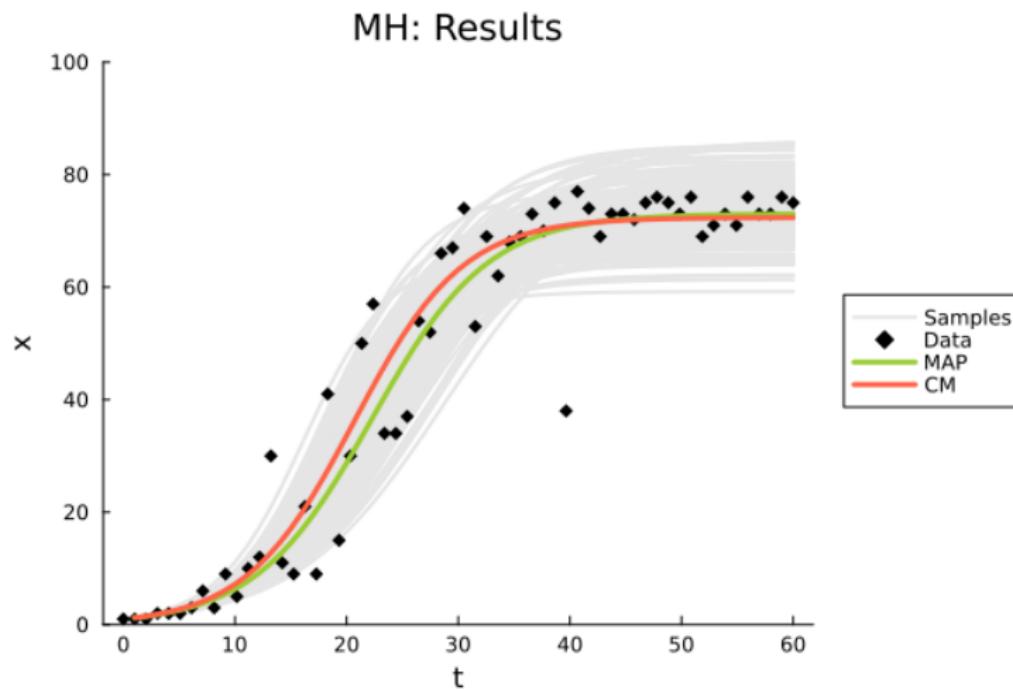


Point cloud in parameter space



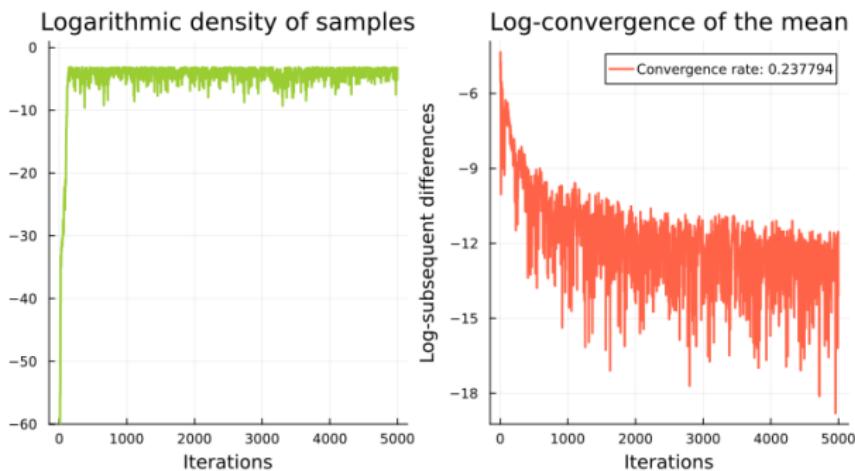
Resultado del algoritmo MH

Modelo y datos



Resultado del algoritmo MH

Convergencia



- La tasa de aceptación de muestras fue del 10.44 %.
- La tasa de convergencia de los parámetros fue de 0.2377.

Ecuaciones del Modelo

Las dinámicas de cada órgano P (frutos, flores, hojas, tallos, racimos) en m clases de edad están descritas por:

$$(1) \begin{cases} \frac{dN_{P,1}}{dt} = p_{1,1} N_{P,1} - p_{1,2} N_{P,1}^2 \\ \frac{dN_{P,k}}{dt} = p_{k-1,2} N_{P,k-1}^2 + p_{k,1} N_{P,k} - p_{k,2} N_{P,k}^2 \\ \frac{dC_{P,1}}{dt} = c_{p,1,1} N_{P,1} C_{P,1} - c_{p,1,2} N_{P,1} C_{P,1}^2 \\ \frac{dC_{P,k}}{dt} = c_{p,k,0} N_{P,k-1} C_{P,k-1}^2 + c_{p,k,1} N_{P,k} C_{P,k} - c_{p,k,2} N_{P,k} C_{P,k}^2 \end{cases}$$

para $k = 2, \dots, m$, donde los coeficientes dependen del tiempo y de otras variables de estado.

Estructura del modelo general

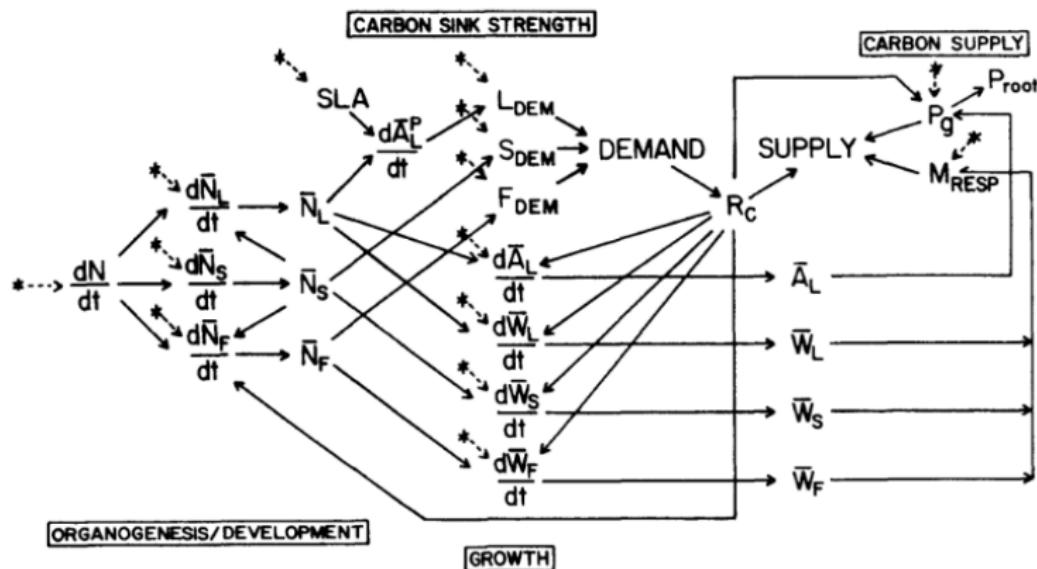


Figura: Relación entre variables del modelo. Tomado de [3].

Estructura del modelo ejemplificado

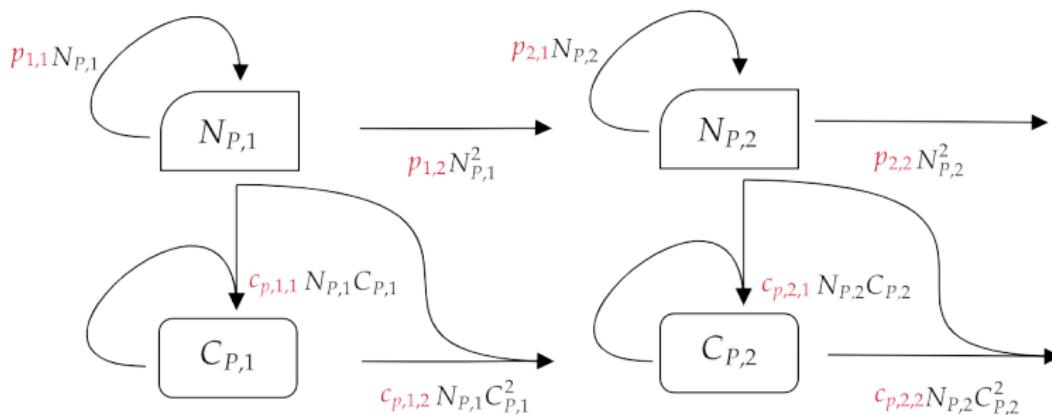


Figura: Relación entre variables del modelo con un órgano y dos clases de edad.
Realización propia.

Ejecución del algoritmo MH en el modelo TOMGRO

- Se generaron 11×4 datos simulados siguiendo la dinámica descrita por la ecuación 1, utilizando el método de Euler para resolver las ecuaciones diferenciales acopladas y la relación recursiva entre dos clases de edad. El resultado del método numérico se perturbó con un ruido multiplicativo proveniente de una distribución Gamma $\Gamma(\alpha = 100, \beta = 1/100)$.
- La función de verosimilitud se modeló como una distribución normal multivariada (asumiendo ruido aditivo):

$$\Sigma = \gamma^2 I,$$

$$\gamma = 1/100,$$

y los parámetros se asumieron uniformemente distribuidos en $(-2.5, 2.5)$.

- Se ejecutó el algoritmo MH con 50000 iteraciones, utilizando un núcleo Gaussiano ($\sigma^2 = 0.01$) y punto inicial en el origen.

Resultado del algoritmo MH

Modelo y convergencia

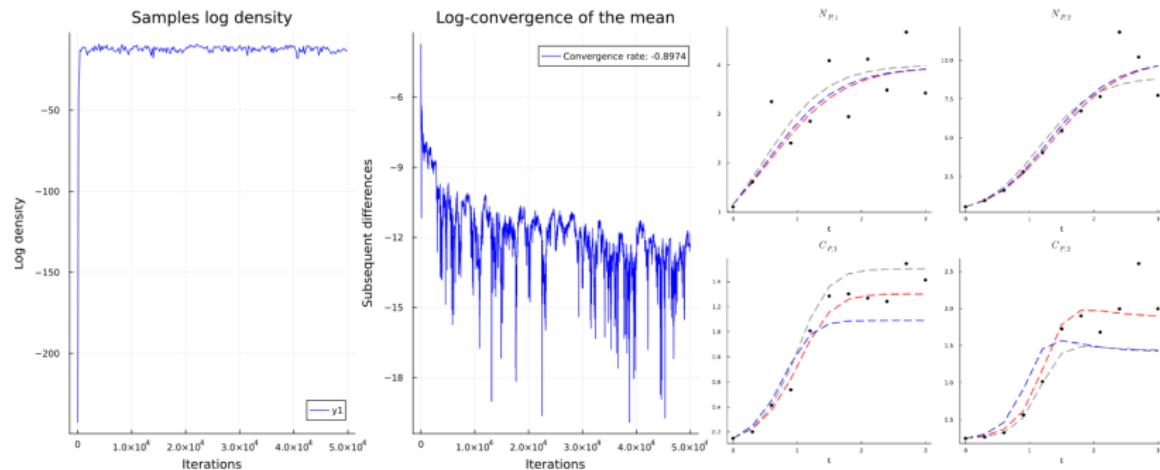


Figura: A la derecha, medidas de convergencia; a la izquierda, resultado de calibración en el modelo (negro: real, azul: CM, rojo: MAP).

Resultado del algoritmo MH

Modelo y convergencia

- La tasa de aceptación de muestras fue del 70.29 %.
- La tasa de convergencia de los parámetros fue de 0.4096.
- El modelo ajustó adecuadamente los datos, con valores de verosimilitud obtenidos de:

$$\frac{\pi_{\text{post}}(\theta_{\text{MAP}})}{\pi_{\text{post}}(\theta_{\text{Real}})} \approx 8.1575 \quad \text{y} \quad \frac{\pi_{\text{post}}(\theta_{\text{CM}})}{\pi_{\text{post}}(\theta_{\text{Real}})} \approx 3.1915$$

Conclusiones y Trabajo Futuro

- Este estudio se centró en el análisis de métodos de estimación de parámetros dentro del marco de los algoritmos de Cadenas de Márkov Monte Carlo (MCMC), en particular el algoritmo de Metropolis-Hastings.
 - Los resultados destacan la efectividad de estas técnicas en distintos escenarios y enfatizan la importancia de seleccionar cuidadosamente las distribuciones de propuesta y ajustar los parámetros del algoritmo para lograr una convergencia eficiente.
 - Una posible línea de estudio es la aplicación de diferentes variaciones del algoritmo de Metropolis-Hastings que incorporen información adicional sobre la distribución objetivo, lo que podría mejorar las tasas de convergencia y la precisión en la estimación.

Conclusiones y Trabajo Futuro

- Otra vía de investigación relevante implica el estudio de las **propiedades de convergencia** del algoritmo de Metropolis-Hastings en la estimación de parámetros para modelos lineales y modelos de ecuaciones diferenciales.
- Además, el problema de predicción puede formularse como un **problema inverso no estacionario** que involucra técnicas de asimilación de datos. En este sentido, metodologías basadas en modelos de Markov, como los filtros de Kalman, ofrecen enfoques prometedores para la integración dinámica de nuevos datos y la mejora del desempeño predictivo.
- Una extensión particularmente interesante es la estimación de parámetros en otros tipos de sistemas dinámicos, incluyendo **ecuaciones diferenciales parciales** (EDPs) y **ecuaciones diferenciales estocásticas** (EDEs).

Referencias

- [1] Cooman, A. (2002). Feasibility of protected tomato cropping in the high altitude tropics using statistical and system dynamic models for plant growth and development [Doctoral thesis]. Katholieke Universiteit Leuven.
- [2] Galvis, J., & Capistrán, M. (2021). Beyond Research: Asimilación de datos en sistemas complejos [Curso].
- [3] J. W. Jones, E. Dayan, L. H. Allen, H. Van Keulen, & H. Challa. (1991). A dynamic tomato growth and yield model(Tomgro). Transactions of the ASAE, 34(2), 0663-0672.
<https://doi.org/10.13031/2013.31715>
- [4] Kaipio, J., & Somersalo, E. (2006). Statistical and computational inverse problems. Springer Science & Business Media.