



# UNIVERSIDAD DE GRANADA

---

Facultad de Ciencias

GRADO EN FÍSICA

TRABAJO FIN DE GRADO

## **Estudio de Predicción de Series Temporales mediante Estadística de Bayes**

Presentado por:

**D. Javier Jesús González Moreno**

Curso Académico 2023/2024

## **Resumen**

En este trabajo se explican los fundamentos teóricos y prácticos necesarios para entender como se aplica la Estadística Bayesiana para modelar y realizar predicciones en series temporales. Primero se introduce el marco teórico de la Estadística Bayesiana, sus bases y los conceptos necesarios para entender cuál es la utilidad y las ventajas que la diferencian de la Estadística Clásica.

También se detallan las bases de los métodos Monte Carlo basados en cadenas de Markov (MCMC). Se explican los conceptos necesarios para entender la naturaleza matemática de estos procesos y se desarrollan los dos algoritmos más importantes (Metrópolis-Hastings y Muestreador de Gibbs).

Se muestran las principales herramientas para realizar análisis sobre el modelo, además de detallar las estrategias a seguir para optimizar todo lo posible los resultados.

Finalmente, para ilustrar las técnicas y procedimientos que se explican, se realizan tres ejemplos diferentes de series temporales. De esta forma se muestra al lector como es el proceso de modelaje y como se interpretan los resultados obtenidos.

## **Abstract**

This work explains the theoretical and practical foundations necessary to understand how Bayesian Statistics is applied to model and predict time series. First, the theoretical framework of Bayesian Statistics, its principles, and the key concepts that underscore its utility and advantages over classical statistics are covered. This initial discussion provides the foundational knowledge needed to grasp the motivation behind these studies.

Furthermore, the foundations of Monte Carlo Methods based on Markov Chains (MCMC) are detailed. These methods are crucial, as they enable the practical application of Bayesian statistical concepts. The necessary mathematical concepts to understand these processes are explained, and the two most important algorithms (Metropolis-Hastings and Gibbs Sampler) are described in detail. Additionally, the main tools for analyzing the obtained models are presented, along with strategies for optimizing the results as much as possible.

Finally, to illustrate the explained techniques and procedures, three different examples of time series are provided. These examples demonstrate to the reader the modeling process and how to interpret the obtained results.

# Índice

<b>1</b>	<b>Introducción</b>	<b>4</b>
<b>2</b>	<b>Estadística Bayesiana e Inferencia Bayesiana</b>	<b>5</b>
2.1	Teorema de Bayes . . . . .	5
2.2	Distribuciones a priori. . . . .	6
2.2.1	Distribuciones a priori conjugadas. . . . .	7
2.2.2	Distribuciones poco informativas. . . . .	9
2.2.3	Distribuciones a priori propias. . . . .	9
2.2.4	Distribuciones a priori impropias . . . . .	10
2.3	Contraste con la Estadística Clásica y utilidades. . . . .	10
2.4	Ejemplo clásico de la moneda . . . . .	11
<b>3</b>	<b>Método de Monte Carlo mediante Cadenas de Markov (MCMC)</b>	<b>11</b>
3.1	Integración Monte Carlo . . . . .	11
3.2	Cadenas de Markov . . . . .	13
3.2.1	Matriz de Transición . . . . .	13
3.2.2	Propiedades de las Cadenas de Markov . . . . .	13
3.2.3	Distribución Estacionaria . . . . .	14
3.3	Métodos MCMC . . . . .	15
3.3.1	Metropolis-Hastings . . . . .	15
3.3.2	Muestreador de Gibbs . . . . .	16
3.4	Análisis de resultados . . . . .	16
3.4.1	Traza y Burn in . . . . .	17
3.4.2	Gelman-Rubin . . . . .	19
3.4.3	Corner plot . . . . .	20
3.5	Implementación en PyMC . . . . .	20
<b>4</b>	<b>Aplicación para modelar series temporales</b>	<b>20</b>
4.1	Predicción de valores en una serie temporal con modelo multivariable	20
4.2	Estudio temporal de la probabilidad de que ocurra un suceso . . .	24
4.3	Estudio temporal aplicado al campo de la física . . . . .	28
<b>5</b>	<b>Conclusions.</b>	<b>30</b>
	<b>Referencias</b>	<b>32</b>

## 1 Introducción

La predicción de series temporales se refiere al estudio de patrones que se repiten en el tiempo para poder describir el comportamiento que esperamos observar. Se trata de un área muy importante y fundamental en muchos campos como la economía, estudios sanitarios, ciencia...

En el estudio de series temporales es fundamental el concepto de inferencia. La inferencia consiste en el procedimiento que se lleva a cabo para obtener un modelo de predicción a partir de unos datos iniciales.

El estudio de las series temporales puede llegar a resultar altamente complejo, debido a la naturaleza estocástica de algunas series temporales. Incluso con el uso de herramientas computacionales potentes, comprender la naturaleza temporal de un suceso puede ser altamente complicado.

Además, un caso de estudio concreto puede resultar muy diferente a otro, por lo que es interesante considerar diferentes técnicas de inferencia .

Una de estas técnicas, es la inferencia Bayesiana. Estas se fundamentan en los conceptos teóricos de la Estadística Bayesiana, que ofrece una perspectiva alterna a la Estadística Clásica.

Estos estudios se llevan a cabo mediante la programación probabilística, en concreto, los métodos computacionales que se emplean en la inferencia Bayesiana son los métodos de Monte Carlo basados en cadenas de Markov (MCMC).

Los objetivos de este trabajo son:

- Explicar los fundamentos teóricos de la estadística bayesiana y su aplicación en la predicción de series temporales.
- Implementar y facilitar la comprensión del método MCMC para llevar a cabo la inferencia Bayesiana.
- Mostrar las diferentes herramientas de análisis.
- Aplicar los conceptos teóricos a través de estudios prácticos para hacer más fácil la comprensión.

Vamos a definir unos conceptos estadísticos básicos que vamos a emplear durante el trabajo:

1. **Variable estadística:** Característica que puede tomar diferentes valores, los cuales pueden medirse.
2. **Espacio de probabilidad:** Modelo matemático que engloba los posibles resultados al medir una variable.
3. **Distribución estadística:** Modelo teórico que muestra como se distribuyen los posibles valores de una variable según los valores más y menos comunes.
4. **Valor medio:** Promedio que se obtiene al realizar un número alto de medidas de una variable.

## 2 Estadística Bayesiana e Inferencia Bayesiana

Los orígenes de la Estadística Bayesiana se remontan al siglo XVIII, con los estudios de Thomas Bayes y Pierre-Simon Laplace. En los últimos tiempos, su popularidad ha crecido bastante debido a los avances en computación, que hacen posible la implementación de métodos basados en la misma de forma más eficiente.

La Estadística Bayesiana [1] es una rama de la estadística que nos ofrece un enfoque alternativo a la Estadística Clásica. La principal diferencia es que en lugar de tratar los parámetros como valores fijos pero desconocidos, la Estadística Bayesiana los trata como variables aleatorias con una distribución de probabilidad. Esto permite combinar información previa (a priori) con los datos observados para obtener una actualización de la creencia sobre los parámetros del modelo (a posteriori).

Y justamente, esta es la esencia de la Estadística Bayesiana, la combinación de conocimientos previos con los datos observados para actualizar continuamente esos conocimientos. De tal manera que con la incorporación de nuevos datos, las creencias y entendimientos se ajustan y detallan. Esto permite adaptar el modelo ante cualquier factor desconocido que pueda influir en el suceso. Si alguna variable de la que no existe conocimiento cambia, el modelo se actualizará para reflejar los nuevos datos y poder ajustar las predicciones en consecuencia. El poder actualizar de esta manera las predicciones es una de las principales ventajas de la Estadística Bayesiana sobre la Estadística Clásica, que no permite incorporar las actualizaciones de nuevos datos de manera tan dinámica.

### 2.1 Teorema de Bayes

Antes de continuar profundizando, se enuncia a continuación el teorema de Bayes, que es el pilar fundamental de la Estadística Bayesiana. Este teorema permite entender cómo se pueden actualizar las creencias sobre un parámetro de interés a partir de la observación de datos. Se van a desarrollar las ecuaciones necesarias hasta llegar al teorema de Bayes.

Se parte de la definición de probabilidad condicional. La probabilidad condicional de  $\theta$  dado  $X$  se define:

$$P(\theta|X) = \frac{P(\theta \cap X)}{P(X)}. \quad (2.1)$$

De forma similar, la probabilidad condicional de  $X$  dado  $\theta$  :

$$P(X|\theta) = \frac{P(X \cap \theta)}{P(\theta)}. \quad (2.2)$$

Dado que  $P(\theta \cap X)$  es lo mismo que  $P(X \cap \theta)$ , se igualan las dos expresiones:

$$P(\theta|X) \cdot P(X) = P(X|\theta) \cdot P(\theta). \quad (2.3)$$

Si se reorganiza la ecuación 2.3, se obtiene el Teorema de Bayes[2].

**Teorema 1 (Teorema de Bayes)** Sean  $\{\theta_1, \theta_2, \dots, \theta_i, \dots, \theta_n\}$  sucesos excluyente tales que  $P(\theta_i) \neq 0$  para todo  $i = 1, 2, \dots, n$ . Sea  $X$  un suceso cualquiera del que se conocen  $P(X|\theta_i)$ . La probabilidad  $P(\theta_i|X)$  viene dada por la expresión:

$$P(\theta_i|X) = \frac{P(X|\theta_i)P(\theta_i)}{P(X)}, \quad (2.4)$$

donde  $P(X)$  se puede calcular como:

$$P(X) = \sum_{j=1}^n P(X|\theta_j)P(\theta_j). \quad (2.5)$$

- $P(\theta|X)$  es la distribución a posteriori del parámetro  $\theta$  dados los datos  $X$ . Esto representa la probabilidad de que el parámetro  $\theta$  sea verdadero dado  $X$ . Esta es la distribución que se busca obtener, que se trata de la actualización de la distribución inicial que se define para los parámetros.
- $P(X|\theta)$  es el likelihood (o verosimilitud) de los datos  $X$  dada la hipótesis  $\theta$ . Esta se define como la probabilidad de observar los datos  $X$  si  $\theta$  es verdadero. Es una medida de cuán bien el modelo propuesto explica los datos.
- $P(\theta)$  es la distribución a priori del parámetro  $\theta$ , que representa el conocimiento inicial antes de observar los datos. El investigador es el que define esta distribución.
- $P(X)$  es la probabilidad de los datos  $X$ , también conocida como verosimilitud marginal que será una constante de normalización en estos estudios.

Este teorema nos permite observar de forma matemática como funciona la inferencia bayesiana, que consiste en actualizar la distribución del parámetro de interés a partir de la información previa y los datos observados.

## 2.2 Distribuciones a priori.

Una de las bases en las que se fundamenta el enfoque bayesiano es en el uso del conocimiento previo [3], pues la forma final de las distribuciones a posteriori de los parámetros dependerán en gran medida las distribuciones a priori que se escojan. Las distribuciones a posteriori brindan mucha información útil, pues realmente informarán, no solo del mejor modelo que se pueda obtener, si no que con la visualización de la distribución a posteriori se podrá observar como de probables serían otros modelos, viendo que credibilidad tiene cada valor del parámetro en cuestión (mientras más cercano a 1 se encuentre en el eje y, mayor certeza habrá de que ese valor del parámetro sea el adecuado en el modelo, mientras que si un valor posible del parámetro valiera 0 querría decir que se tiene la certeza de que ese valor no sería el adecuado nunca). Elegir una distribución que se pueda adecuar a lo esperado asegurará la convergencia en el proceso de búsqueda y aumentará la eficiencia. Las distribuciones a priori pueden clasificarse de dos maneras diferentes. Por un lado se clasifican como conjugadas o poco informativas y por otro lado pueden ser propias o impropias.

### 2.2.1 Distribuciones a priori conjugadas.

Por definición, una distribución a priori se dice conjugada respecto a la verosimilitud si pertenece a la misma familia que la distribución a posteriori obtenida [4]. En este caso, la distribución a priori y a posteriori son conjugadas entre sí. El poder definir una distribución a priori conjugada dependerá del conocimiento previo que se tenga sobre el estudio. Pues para poder definir de forma cerrada la distribución a priori se debe conocer con mucha certeza cuál será la forma de la distribución final. Las distribuciones conjugadas más usadas son las que observamos en la siguiente tabla.

<i>A priori</i>	<i>Verosimilitud</i>	<i>A posteriori</i>
Normal	Normal	Normal
Beta	Binomial	Beta
Gamma	Poisson	Gamma
Dirichlet	Multinomial	Dirichlet

Se pueden visualizar estas distribuciones en la figura 1.

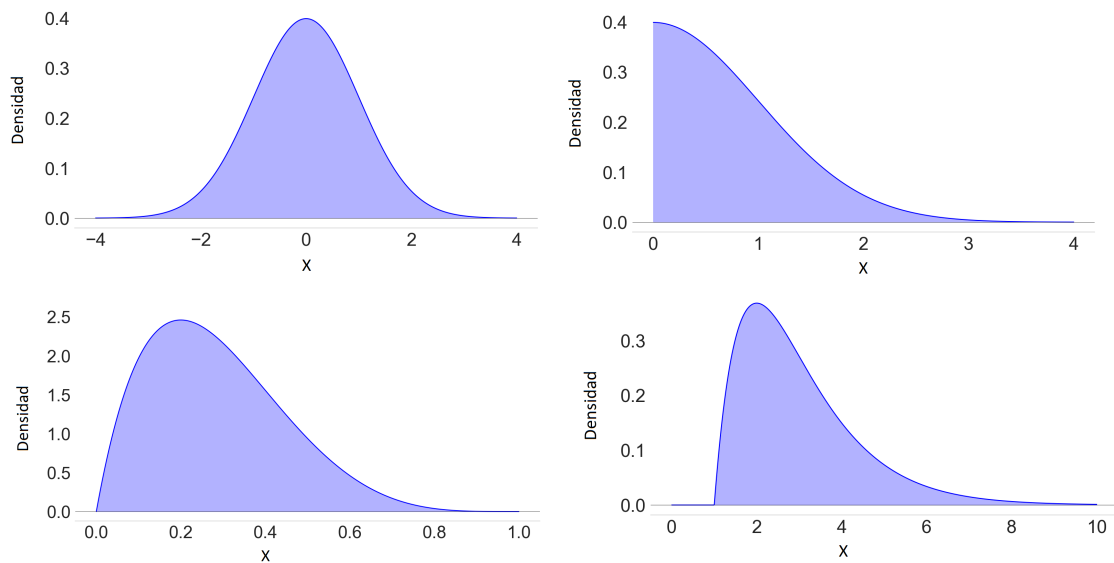


Figura 1: De izquierda a derecha, distribución Normal, distribución Seminormal, distribución Beta y distribución Gamma.

Una de las distribuciones conjugadas que se ha usado en el estudio práctico de este trabajo es la distribución normal. Se puede ver de forma analítica la demostración de que esta distribución es conjugada respecto a la verosimilitud normal.

**Demostracion** El teorema de Bayes establece que la distribución a posteriori  $P(\theta|X)$  de un parámetro  $\theta$  dado un conjunto de datos  $X$  es proporcional al producto de la distribución a priori  $P(\theta)$  y el likelihood  $P(X|\theta)$ :

$$P(\theta|X) \propto P(X|\theta) \cdot P(\theta). \quad (2.6)$$

Suponiendo que la distribución a priori de  $\theta$  es una distribución normal con media  $\mu_0$  y varianza  $\sigma_0^2$ :

$$P(\theta) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(\theta - \mu_0)^2}{2\sigma_0^2}\right). \quad (2.7)$$

De igual manera el likelihood es una distribución normal con media  $\theta$  y varianza  $\sigma^2$ :

$$P(X|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(X - \theta)^2}{2\sigma^2}\right). \quad (2.8)$$

Introduciendo las ecuaciones 2.7 y 2.8 en 2.6, se obtiene:

$$P(\theta|X) \propto \exp\left(-\frac{(\theta - \mu_0)^2}{2\sigma_0^2}\right) \cdot \exp\left(-\frac{(X - \theta)^2}{2\sigma^2}\right). \quad (2.9)$$

Simplificando la expresión anterior:

$$P(\theta|X) \propto \quad (2.10)$$

$$\begin{aligned} & \exp\left(-\frac{1}{2} \left[ \frac{\theta^2 - 2\theta\mu_0 + \mu_0^2}{\sigma_0^2} + \frac{\theta^2 - 2\theta X + X^2}{\sigma^2} \right] \right) \\ P(\theta|X) & \propto \exp\left(-\frac{1}{2} \left[ \frac{\theta^2 - 2\theta\mu_0 + \mu_0^2 + \theta^2 - 2\theta X + X^2}{\sigma_0^2} + \frac{\sigma^2(\theta^2 - 2\theta X + X^2)}{\sigma^4} \right] \right) \\ P(\theta|X) & \propto \exp\left(-\frac{1}{2} \left[ \frac{(\sigma^2 + \sigma_0^2)\theta^2 - 2\theta(\sigma_0^2 X + \sigma^2 \mu_0) + \sigma_0^2 X^2 + \sigma^2 X^2}{\sigma_0^2 \sigma^2} \right] \right) \\ P(\theta|X) & \propto \exp\left(-\frac{1}{2} \left[ \frac{(\theta - \frac{\sigma_0^2 X + \sigma^2 \mu_0}{\sigma^2})^2}{\frac{\sigma_0^2}{\sigma^2}} \right] \right). \end{aligned} \quad (2.11)$$

La expresión que se obtiene es proporcional a la densidad de probabilidad de una distribución normal con media  $\frac{\sigma_0^2 X + \sigma^2 \mu_0}{\sigma^2}$  y varianza  $\frac{\sigma_0^2}{\sigma^2}$ . Se puede concluir en que la distribución a posteriori es normal.

En Estadística, se emplea una notación característica para definir distribuciones de probabilidad. A continuación, se presenta un ejemplo de cómo se define una distribución normal.

$$\theta \sim \mathcal{N}(\mu, \sigma^2), \quad (2.12)$$

donde:

- $\theta$  es el parámetro(variable aleatoria) que sigue una distribución normal.
- $\mu$  es la media de la distribución normal.



- $\sigma^2$  es la varianza de la distribución normal.

La notación  $\theta \sim \mathcal{N}(\mu, \sigma^2)$  se lee como "la variable aleatoria  $\theta$  está distribuida de forma normal con media  $\mu$  y varianza  $\sigma^2$ ".

### 2.2.2 Distribuciones poco informativas.

En los casos en los que la información previa no sea tan sólida se debe definir las distribuciones de forma más genérica. Lo más usual es usar la distribución uniforme, que tiene la forma que se aprecia en la figura 2.

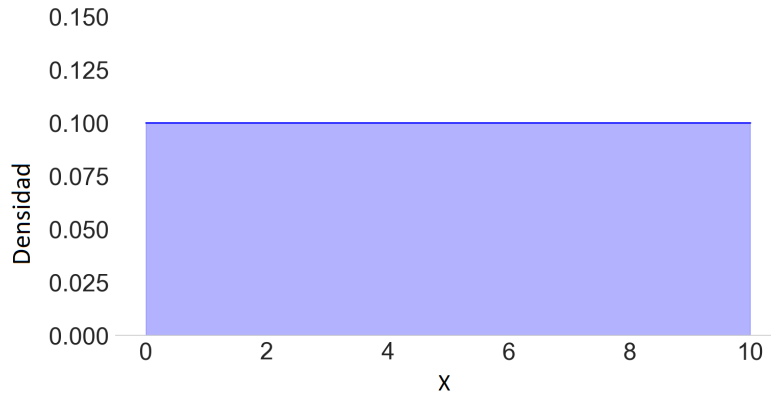


Figura 2: Distribución uniforme, usada en modelos con escasa información previa.

### 2.2.3 Distribuciones a priori propias.

Una distribución a priori propia es una distribución de probabilidad que por definición se considera válida en el campo de la estadística. Esto significa que la integral de la función de densidad de probabilidad (pdf) sobre el espacio de parámetros es igual a 1 (esto es lo mismo que decir que la distribución está normalizada).

**Definición:** Sea  $P(\theta)$  una distribución a priori sobre el parámetro  $\theta$ . La distribución a priori es propia si:

$$\int P(\theta) d\theta = 1.$$

Por ejemplo, se puede ver que una distribución a priori uniforme en el intervalo  $[a, b]$  es una distribución a priori propia:

$$\pi(\theta) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq \theta \leq b, \\ 0 & \text{en otro caso.} \end{cases}$$

Las distribuciones a priori propias son útiles cuando se tiene gran cantidad de información sobre el parámetro.

### 2.2.4 Distribuciones a priori impropias

Una distribución a priori impropia es una función que no está normalizada (su integral en el espacio de parámetros no es finita). Aunque no es una distribución de probabilidad válida, puede utilizarse en Estadística Bayesiana y puede llevar a una distribución a posteriori propia cuando se combina con los datos observados.

**Definición:** Sea  $P(\theta)$  una función sobre el parámetro  $\theta$ . La distribución a priori es impropia si:

$$\int P(\theta) d\theta = \infty.$$

Una distribución a priori uniforme sobre todos los números reales es una distribución a priori impropia:

$$P(\theta) = k, \quad \text{donde } k \text{ es una constante.}$$

Aquí, la integral  $\int_{-\infty}^{\infty} k d\theta$  no converge.

Las distribuciones a priori impropias son útiles cuando se queremos expresar una falta de información previa sobre el parámetro. Son comunes en métodos bayesianos no informativos.

### 2.3 Contraste con la Estadística Clásica y utilidades.

La Estadística Clásica y la Estadística Bayesiana difieren en varios aspectos importantes. Primero, la interpretación de la probabilidad es diferente entre ambos enfoques. En la Estadística Clásica, la probabilidad se interpreta como la frecuencia en el límite de un número infinito de repeticiones del experimento ( $n \rightarrow \infty$ ). Por el contrario, en la Estadística Bayesiana, la probabilidad representa la incertidumbre o creencia sobre un evento. Esto significa que la Estadística Clásica se basa en la observación de cuantas veces se repite un evento para determinar probabilidades, mientras que la Estadística Bayesiana estudia cómo las creencias sobre esos eventos se ajustan con nueva información.

La inferencia también difiere mucho entre ambos puntos de vista. En la inferencia clásica, la incertidumbre se muestra a través de la variabilidad de los estimadores usando intervalos de confianza. En la inferencia bayesiana, la incertidumbre se expresa directamente a través de la distribución a posteriori del parámetro. Esta distribución a posteriori da una imagen más completa de la incertidumbre y permite hacer inferencias más intuitivas sobre los parámetros desconocidos.

Usar la Estadística Bayesiana en lugar de la Estadística Clásica puede ser especialmente beneficioso en los siguientes casos:

1. **Conocimientos previos sólidos:** Si se posee conocimiento sólido basado en observaciones pasadas o modelos ya existentes, es interesante emplear el enfoque bayesiano, porque nos permite integrar esta información en el análisis, lo que aumenta la precisión de los resultados.
2. **Pocos datos:** En casos donde tengamos una cantidad de datos escasa, los enfoques bayesianos son útiles al permitir incorporar la información previa como base para mejorar los resultados.

3. **Alta incertidumbre:** En contextos donde la incertidumbre es alta, los métodos bayesianos nos ofrecen una manera de cuantificar y entender esta incertidumbre.
4. **Comparación de alternativas:** Cuando se necesita comparar con detalle varias opciones, los métodos bayesianos proporcionan unos resultados muy completos para comparar alternativas y poder tomar decisiones con mayor conocimiento.

A la hora de comparar la Estadística Bayesiana con la Estadística Clásica, se suele comentar sobre la subjetividad que albergan los métodos bayesianos. Y es que realmente, en rasgos generales los métodos estadísticos basados en la probabilidad pueden considerarse subjetivos, ya que se apoyan en simplificaciones matemáticas del mundo real. Teniendo esto en cuenta, el enfoque bayesiano suele ser visto como particularmente subjetivo debido a la elección inicial de la distribución a priori, que a menudo refleja las creencias del investigador. Sin embargo, al actualizar esta distribución con datos adicionales y ajustarla en función de la evidencia, se puede mejorar la objetividad del análisis. Esto indica que, aunque inicialmente el enfoque bayesiano pueda parecer subjetivo, su objetividad puede aumentar a medida que se integran más datos.

## 2.4 Ejemplo clásico de la moneda

Para mostrar estos conceptos explicados hasta ahora y poder sintetizarlos se va a explicar el ejemplo clásico del lanzamiento de una moneda. Se supone que se tiene una moneda cuyo sesgo (probabilidad de obtener cara) se desconoce. En la Estadística Clásica se tendría que realizar muchas medidas de datos para poder llegar a determinar la probabilidad de obtener cara. En la Estadística Bayesiana, antes de lanzar la moneda, se usaría la creencia que de que al lanzar la moneda se tiene la misma probabilidad de obtener cara y cruz. Esta sería la distribución a priori. Tras lanzar la moneda varias veces e ir observando los resultados, se va actualizando la creencia inicial de la moneda, lo que acaba resultando en la distribución a posteriori, que muestra la incertidumbre sobre la probabilidad de obtener cara.

# 3 Método de Monte Carlo mediante Cadenas de Markov (MCMC)

## 3.1 Integración Monte Carlo

Los métodos bayesianos han obtenido un gran aumento de popularidad en los últimos tiempos debido a los grandes avances en el campo de la computación y la programación probabilística, pues antes era realmente costoso realizar las inferencias, y hoy en día se poseen las herramientas necesarias para ello. A la hora de entender como se realizan las inferencias, es de vital importancia comprender los métodos Monte Carlo. Estos métodos consisten en la generación masiva de muestras aleatorias para poder aproximar cálculos que son altamente complejos en su

forma analítica[5]. Antes de entrar a describir las funciones de estos métodos dentro de la inferencia bayesiana, se va a detallar los rasgos generales en los que se basan. Considerando la siguiente integral que se quiere resolver:

$$E[h(x)] = \int h(x)f(x)dx. \quad (3.1)$$

El funcionamiento del método Monte Carlo para resolver la integral consiste en generar una serie de muestras aleatorias de la función densidad  $f$  y proponer como aproximación:

$$\hat{h}_n = \frac{1}{n} \sum_{j=1}^n h(x_j). \quad (3.2)$$

Debido a la Ley fuerte de los grandes números tenemos que casi seguramente  $\hat{h}_n$  converge a  $E[h(x)]$ .

Para ilustrar esto, se ha realizado un programa de un clásico ejemplo para obtener el valor del número  $\pi$  con este procedimiento. Para ello se define un cuadrado de lado dos que tendrá un área de 4 unidades. Dentro de este cuadrado se tiene un círculo de radio uno con área  $\pi$ . Comenzamos a generar puntos aleatorios de forma masiva. Si se contabilizan los puntos que quedan dentro del círculo de radio 1, la proporción que se obtenga entre el número de puntos dentro y fuera del círculo será igual al área del círculo ( $\pi$ ) entre el área del cuadrado (4). Se puede observar el proceso en la figura 3.

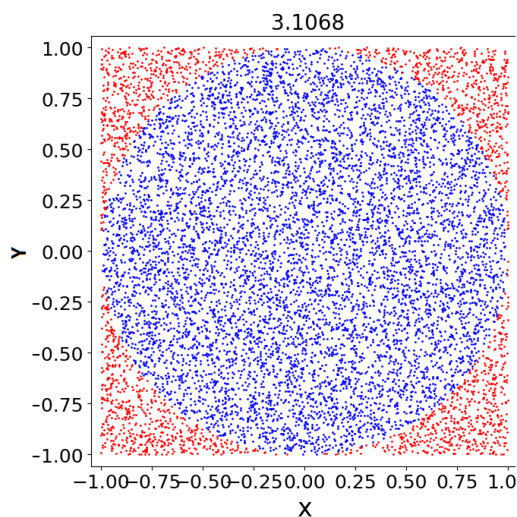


Figura 3: Estimación de  $\pi$  con integración de Monte Carlo. Los puntos en azul representan los valores generados dentro del círculo de radio  $\pi$ , y los rojos los generados fuera.

En este ejemplo se ha obtenido un valor de  $\pi = 3.1068$ . De esta forma se puede comprender de manera gráfica e intuitiva como funciona el Método Monte Carlo para generar estimaciones.

### 3.2 Cadenas de Markov

Los métodos bayesianos giran entorno al concepto de alcanzar una distribución estacionaria, lo que significará que no hay mejor distribución posible para el modelo. En este contexto de convergencia es fundamental comprender las cadenas de Markov [6] [7].

Las cadenas de Markov se definen como modelos matemáticos que describen un sistema que transita de un estado a otro dentro de un espacio. En estas cadenas la probabilidad de que se transite a cualquier estado depende únicamente del estado anterior.

Analíticamente una cadena de Markov se define mediante una secuencia de variables aleatorias  $\{X_t\}_{t \geq 0}$  con la propiedad antes mencionada:

$$P(X_{t+1} = x \mid X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0) = P(X_{t+1} = x \mid X_t = x_t) \quad (3.3)$$

De esta forma se puede empezar a comprender que la idea del uso de estas cadenas consiste en que si se logra construir una cadena de Markov, tal que su distribución estacionaria sea la distribución que estamos interesados en muestrear, tendremos mayor facilidad para hacer estimaciones con la integración Monte Carlo.

#### 3.2.1 Matriz de Transición

Dentro de una cadena de Markov, las probabilidades de transición entre diferentes estados se describen mediante una matriz de transición  $P$ , donde  $P_{ij} = P(X_{t+1} = j \mid X_t = i)$ . Esta matriz cumple las dos siguientes condiciones:

1.  $P_{ij} \geq 0$  para todo par  $i, j$  de estados.
2.  $\sum_j P_{ij} = 1$  para todo estado  $i$ , lo que implica que la cadena se mantenga en algún estado en cada paso.

#### 3.2.2 Propiedades de las Cadenas de Markov

Las cadenas de Markov tienen dos propiedades fundamentales:

- **Irreducibilidad:** Una cadena de Markov es irreducible si es posible llegar de cualquier estado a cualquier otro estado. Matemáticamente, una cadena es irreducible si para cada par de estados  $i$  y  $j$  existe un  $n$  tal que  $(P^n)_{ij} > 0$ , donde  $P^n$  es la matriz de transición elevada a la  $n$ -ésima potencia.
- **Aperiocidad:** Una cadena de Markov se dice que es aperiódica si no hay ciclos determinísticos en el proceso de transición (esto significa que la cadena no sigue un patrón fijo de transiciones).

### 3.2.3 Distribución Estacionaria

Como se ha comentado, una cadena de Markov puede converger a una distribución estacionaria  $\pi$  si cumple ciertas condiciones. La distribución estacionaria cumple la ecuación de equilibrio:

$$\pi Q = \pi, \quad (3.4)$$

donde  $Q$  es la matriz de transición de la cadena. Si una cadena es irreducible y aperiódica, se tiene una distribución estacionaria única y todas las distribuciones iniciales convergerán a ella después de un largo número de iteraciones.

**Demostración** La existencia de una distribución estacionaria para una cadena de Markov irreducible y aperiódica se demuestra utilizando el teorema de Perron-Frobenius.

El teorema de Perron-Frobenius enuncia que cualquier matriz irreducible  $Q$  tiene un único valor propio  $\lambda = 1$  con un vector propio correspondiente  $\pi$  cuyas entradas son todas positivas y que puede ser normalizado para que sumen 1. Este vector  $\pi$  es la distribución estacionaria.

Siendo  $Q$  una matriz de transición irreducible (esto significa que para cualquier par de estados  $i$  y  $j$ , existe un número de pasos  $n$  tal que  $(Q^n)_{ij} > 0$ ) que está normalizada:

$$\sum_{i=1}^n \pi_i = 1. \quad (3.5)$$

Queremos mostrar que  $\pi Q = \pi$ . Suponemos que  $\pi$  es un vector fila de longitud  $n$  (el número de estados en la cadena de Markov). Entonces, para cada componente  $\pi_i$ :

$$\sum_{j=1}^n \pi_j Q_{ji} = \pi_i. \quad (3.6)$$

Esto equivale a:

$$\pi Q = \pi. \quad (3.7)$$

La irreducibilidad y la aperiodicidad son las propiedades que garantizan que la cadena de Markov no tiene patrones de transición definidos y que es posible llegar a cualquier estado desde cualquier otro. Con estas propiedades se ve que la distribución estacionaria es única y que cualquier distribución inicial convergerá a esta distribución estacionaria a largo plazo. Siendo  $\mu$  una distribución inicial, se tiene:

$$\lim_{n \rightarrow \infty} \mu Q^n = \pi. \quad (3.8)$$

### 3.3 Métodos MCMC

Tras haber introducido los métodos de Monte Carlo y las cadenas de Markov, podemos centrarnos en los detalles de estas técnicas dentro de la inferencia bayesiana [8].

La idea para llegar a obtener la distribución a posteriori  $\pi$  es emplear el método Monte Carlo para generar cadenas de Markov cuya distribución de convergencia sea justamente  $\pi$ . Tal y como se ha explicado en la sección anterior, la distribución en cada momento solo dependerá de la anterior. A continuación se desarrollará con más detalle los dos algoritmos más importantes en el contexto bayesiano [9], pero el funcionamiento básico consiste en que en cada iteración se generarán muestras aleatorias de una distribución propuesta y se tendrá que comprobar que tan bien se adecuan los datos reales con el modelo para los parámetros del instante en el que se encuentra la cadena Markov (para ello habrá que evaluar la función likelihood, donde se emplearán de nuevo los métodos Monte Carlo para realizar esta estimación).

#### 3.3.1 Metrópolis-Hastings

El algoritmo de Metrópolis-Hastings es uno de los métodos de Monte Carlo basado en cadenas de Markov más utilizados para generar muestras de una distribución de probabilidad objetivo  $\pi(x)$ . Este será el algoritmo que se empleará en el caso práctico de este trabajo. A continuación, se describe paso a paso el algoritmo:

1. **Inicialización:** Se comienza con un valor inicial  $x^{(0)}$  en el espacio de búsqueda.
2. **Propuesta:** Se genera una propuesta  $x^*$  a partir de una distribución de propuesta  $q(x^*|x^{(t)})$ , donde  $x^{(t)}$  es el valor actual en la cadena de Markov.
3. **Aceptación o Rechazo:** Se calcula la probabilidad de aceptación  $\alpha$  como la razón entre la probabilidad de  $x^*$  bajo la distribución objetivo y la probabilidad de  $x^{(t)}$  bajo la misma distribución:

$$\alpha = \min \left( 1, \frac{\pi(x^*)q(x^{(t)}|x^*)}{\pi(x^{(t)})q(x^*|x^{(t)})} \right). \quad (3.9)$$

Se determina a partir de que valor se acepta la propuesta  $x^*$  según el valor de la probabilidad  $\alpha$ . A veces se define  $\alpha$  de forma que aunque el valor actual no se ajuste mejor que el anterior a los datos, este se acepta para explorar correctamente todos los valores. Si  $x^*$  es aceptado, lo agregamos a nuestra cadena de Markov, es decir,  $x^{(t+1)} = x^*$ ; si no, repetimos el valor actual,  $x^{(t+1)} = x^{(t)}$ .

4. **Iteración:** Repetimos los pasos 2 y 3 para generar una secuencia de muestras  $\{x^{(0)}, x^{(1)}, x^{(2)}, \dots\}$  de la distribución objetivo  $\pi(x)$ .

Para alcanzar la distribución estacionaria será fundamental la correcta elección de la distribución de propuesta  $q(x^*|x^{(t)})$ .

### 3.3.2 Muestreador de Gibbs

El muestreador de Gibbs es el otro método de MCMC que más se emplea para generar muestras de una distribución conjunta de varias variables aleatorias  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ . Este algoritmo es útil cuando queremos generar muestras de una distribución que es difícil de muestrear directamente, pero cuyas distribuciones condicionales son más fáciles de obtener.

Este es el paso a paso del algoritmo:

1. **Inicialización:** Comenzamos con un conjunto inicial de valores para todas las variables  $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$ .
2. **Iteración:** Para cada variable  $x_i$  en  $\mathbf{x}$ , repetimos los siguientes pasos:
  - a) **Muestreo condicional:** Muestreamos  $x_i^{(t+1)}$  de su distribución condicional  $p(x_i | \mathbf{x}_{\setminus i}^{(t)})$ , donde  $\mathbf{x}_{\setminus i}^{(t)}$  es el conjunto de todas las variables excepto  $x_i$  en la iteración  $t$ .
3. **Actualización:** Actualizamos  $\mathbf{x}^{(t)}$  con los nuevos valores que han sido muestreados  $\mathbf{x}^{(t+1)}$ .
4. **Iteración:** Repetimos el paso 2 hasta alcanzar el número de iteraciones seleccionadas.

El muestreador de Gibbs funcionará correctamente siempre que las distribuciones condicionales  $p(x_i | \mathbf{x}_{\setminus i})$  sean muestreables y que el algoritmo tenga un número de iteraciones que sea suficiente para explorar todos los valores.

### 3.4 Análisis de resultados

A la hora de valorar la calidad de un estudio bayesiano con métodos MCMC, tenemos una gran variedad de herramientas de análisis. Con estas se podrá visualizar como ha sido el proceso que se ha llevado a cabo, detectar posibles errores y mejorar ciertos aspectos para optimizar el modelo.

- **Gráfico de Traza:** Este gráfico representa los valores que va tomando un parámetro en cada una de las iteraciones del estudio. Al observar este gráfico se puede determinar si en el estudio se han explorado todos los posibles valores o si en cambio el proceso de búsqueda no está resultando efectivo.
- **Cálculo de Medias y Varianzas:** Una vez descartadas las iteraciones de calentamiento, se pueden realizar estimaciones de las medias y varianzas de las cadenas. Estas estimaciones resumen la información obtenida en la simulación.
- **Cálculo de Intervalos de Confianza:** Además de las medias y varianzas, es posible estimar intervalos de confianza para los parámetros de interés.



### 3.4.1 Traza y Burn in

Como se ha explicado anteriormente, el gráfico de traza es una herramienta fundamental en el análisis bayesiano, utilizada para evaluar la convergencia de las cadenas generadas por el algoritmo. Este gráfico representa los valores de una variable de interés en función del número de iteración, lo que permite visualizar cómo evolucionan las muestras a lo largo del tiempo.

Cuando la cadena tarda largos periodos de tiempo en explorar el espacio de la variable, significará que para obtener la convergencia necesitaremos un número elevado de iteraciones. Este comportamiento también podría indicar que puede que sea necesario realizar cambios en las configuraciones iniciales de los parámetros del método. En cambio, si las iteraciones forman una serie estable en torno a un valor concreto a medida que aumenta el número de iteración, significará que se está alcanzando correctamente la convergencia.

Como se ha explicado, puede que en algunos casos se necesite un número elevado de iteraciones para obtener la convergencia. En estos casos las primeras iteraciones del algoritmo pueden no ser suficientemente representativas, por ello en estos métodos es fundamental el concepto de burn in o período de calentamiento.

El período de calentamiento, se refiere a las iteraciones iniciales de la cadena que se descartan. Estas iteraciones iniciales se eliminan porque no pertenecen a la distribución estacionaria de la que se pretende muestrear, y su inclusión podría empeorar las estimaciones.

El número de iteraciones de calentamiento a seleccionar depende de varios factores. La proximidad entre la distribución propuesta y la distribución estacionaria tiene una gran importancia. Cuanto más similares sean la distribución inicial y la objetivo, menos iteraciones de calentamiento serán necesarias.

**Determinación del Número de Iteraciones de Burn-in** Para determinar el número de iteraciones a descartar se siguen los siguientes pasos:

1. **Análisis Gráfico:** La manera más común de seleccionar  $m$  es observando la traza de la cadena de Markov y determinar cuando se empieza a explorar el espacio al completo.
2. **Métodos Analíticos:** Existen algunos métodos analíticos para determinar el número de iteraciones a descartar, sin embargo estos métodos no son muy usados.
3. **Cadenas en Paralelo:** Una forma efectiva de verificar la convergencia es ejecutar varias cadenas independientes desde diferentes valores iniciales. Si todas las cadenas convergen rápidamente a la misma región del espacio podremos saber que se ha alcanzado la convergencia.

Se pueden visualizar los siguientes gráficos de traza, donde se distinguen las primeras iteraciones descartadas y las siguientes iteraciones seleccionadas para el análisis.

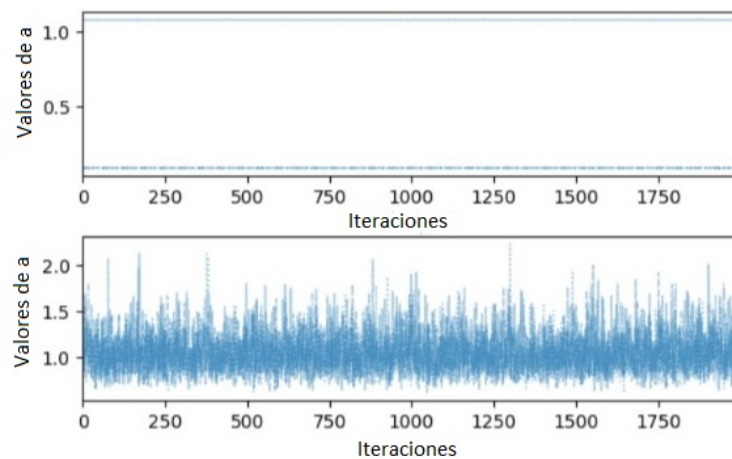


Figura 4: En la imagen superior se muestra el gráfico de la traza(valores del parámetro explorados en función de la iteración) correspondiente a las primeras iteraciones, las cuales serían descartadas. En la imagen inferior se presentan las iteraciones seleccionadas para el análisis.

La convergencia se puede intuir también observando la distribución a posteriori. Como se muestra en la figura 5, cuando se realizan pocas iteraciones, la distribución no tiene un valor máximo bien definido y no tiene una curva de distribución suave, sino que posee picos muy pronunciados. En cambio, para un número de iteraciones elevado, se visualiza que la forma de la distribución sí que tiene un valor máximo localizado y la distribución se encuentra bien definida.

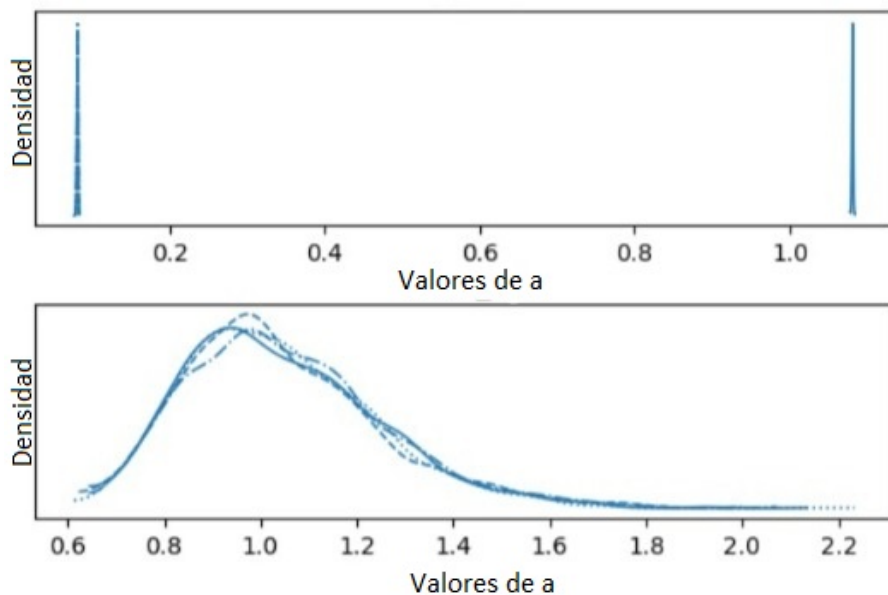


Figura 5: En la imagen superior se muestra la distribución a posteriori del parámetro  $a$  (en el eje  $x$  los valores posibles de un parámetro, en el eje  $y$  la credibilidad para cada valor) correspondiente a las primeras iteraciones, las cuales son descartadas. En la imagen inferior se presenta la distribución a posteriori una vez alcanzada la convergencia.

### 3.4.2 Gelman-Rubin

El diagnóstico de Gelman-Rubin ( $\hat{R}$ ), es una herramienta utilizada para comparar la variabilidad entre varias cadenas de Markov iniciadas en diferentes puntos con la variabilidad que se tiene dentro de cada cadena, para poder ver si todas las cadenas han convergido a la misma distribución estacionaria.

**Pasos del Diagnóstico de Gelman-Rubin** Como se describe en [10], lo primero será generar  $m$  cadenas de Markov, de longitud  $n$ , desde diferentes valores iniciales. Suponiendo  $m$  cadenas  $\{X_t^i\}$  donde  $i = 1, 2, \dots, m$  y  $t = 1, 2, \dots, n$ .

Se calcula la media de los valores generados para cada cadena:

$$\bar{X}^i = \frac{1}{n} \sum_{t=1}^n X_t^i, \quad (3.10)$$

a continuación se calcula la media global:

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m \bar{X}^i, \quad (3.11)$$

así como la variancia entre cadenas  $B$ :

$$B = \frac{n}{m-1} \sum_{i=1}^m (\bar{X}^i - \bar{X})^2, \quad (3.12)$$

y la variancia dentro de cada cadena  $W$ :

$$W = \frac{1}{m} \sum_{i=1}^m \left( \frac{1}{n-1} \sum_{t=1}^n (X_t^i - \bar{X}^i)^2 \right). \quad (3.13)$$

Para finalizar, se calcula  $\hat{R}$ :

$$\hat{R} = \sqrt{\frac{\hat{V}}{W}}, \quad (3.14)$$

donde  $\hat{V}$  es una estimación de la varianza marginal del objetivo:

$$\hat{V} = \left(1 - \frac{1}{n}\right) W + \frac{1}{n} B. \quad (3.15)$$

Si  $\hat{R} \approx 1$ , significará que las cadenas han convergido y están muestreando la misma distribución estacionaria. Si  $\hat{R}$  es notablemente mayor que 1, esto nos indicará que las cadenas no han convergido y que necesitamos más iteraciones en nuestro proceso.

### 3.4.3 Corner plot

Otra herramienta gráfica para evaluar la convergencia es el Corner plot. El Corner plot se emplea en análisis complejos donde se tiene un número elevado de parámetros en el modelo. Es una forma de representar las distribuciones a posteriori obtenidas junto a unos gráficos que indican el grado de correlación que existe entre las distribuciones de los diferentes parámetros entre sí. Si los puntos del diagrama se encuentran dispersos o formando más de una nube de puntos, significará que las distribuciones de los parámetros no están correctamente correlacionadas. En cambio si los puntos se encuentran en nubes estrechas y diagonales se interpreta que las distribuciones a posteriori están correlacionadas, lo que implica que se ha alcanzado la convergencia.

## 3.5 Implementación en PyMC

Para poder aplicar las nociones explicadas a un caso de estudio real se ha usado la librería PyMC en Python. Se trata de una librería que facilita la construcción de modelos bayesianos mediante MCMC. Proporciona herramientas para definir modelos probabilísticos, ejecutar algoritmos MCMC y analizar los resultados[11].

## 4 Aplicación para modelar series temporales

Ahora se van a detallar tres casos de estudio con datos reales, usando diferentes enfoques para poner en práctica los conceptos explicados y visualizar los métodos de análisis. En el primer caso se ha realizado un estudio de una serie temporal en la que se ha podido construir un modelo multivariable. En el segundo caso, se ha empleado un enfoque diferente utilizando la Estadística Bayesiana para estudiar la probabilidad de transitar de un estado a otro diferente. En el último de los casos se mostrará un ejemplo de como se pueden aplicar estas técnicas en el campo de la física. Todos los programas utilizados se encuentran en <https://github.com/javigonzalez20/TFG-Estad-stica-Bayesiana>.

### 4.1 Predicción de valores en una serie temporal con modelo multivariable

En este primer caso, el objetivo es realizar un estudio temporal utilizando un modelo con múltiples parámetros para analizar la esencia de la Estadística Bayesiana. Para esto, se ha decidido estudiar la temperatura media en Málaga a lo largo de los meses. Los datos se han obtenido de la página web Weather Spark [12], abarcando el período desde enero de 2022 hasta junio de 2023. Podemos ver gráficamente los datos en la figura 6 .

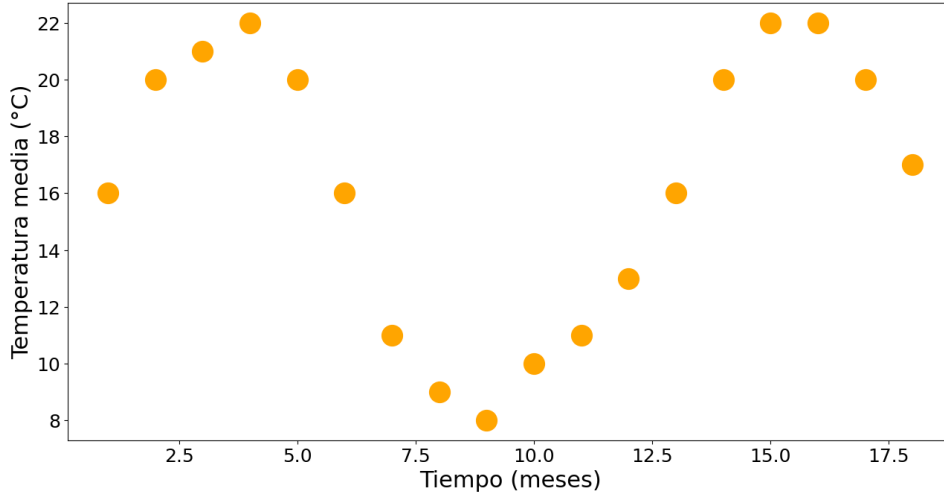


Figura 6: Datos de la temperatura media(°C) de la ciudad de Málaga a lo largo de los meses.

Como se observa, los datos siguen una tendencia sinoidal, por lo que este es un muy buen ejemplo para trabajar con un modelo en el que se definan varias variables. La función a modelar será:

$$T = A \cdot \sin(2\pi f \cdot t) + b. \quad (4.1)$$

Los parámetros presentes en el modelo son:  $A$ , que representa la amplitud de la función sinoidal;  $f$ , que corresponde a la frecuencia de la función; y  $b$ , un término adicional que se suma.

Para elegir la forma del likelihood se asume que los errores en las observaciones siguen una distribución normal alrededor de los resultados predichos por el modelo.

$$\mathcal{L}(\mathbf{y}|\hat{\mathbf{y}}, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \hat{y}_i)^2}{2\sigma^2}\right), \quad (4.2)$$

donde  $\sigma$ , la desviación estándar del ruido es otro parámetro que se define,  $y$  son los datos observados e  $\hat{y}_i$  los valores predichos por el modelo. Se usa el likelihood como medida de que probable es que sea el modelo propuesto con los datos que se tienen.

Las distribuciones a priori que se han seleccionado para cada uno de los parámetros son:

$$\text{Amplitud} \sim \text{Seminormal}(\sigma = 1), \quad (4.3)$$

$$\text{Frecuencia} \sim \text{Normal}(\mu = 0, \sigma = 1), \quad (4.4)$$

$$b \sim \text{Normal}(\mu = 15, \sigma = 1), \quad (4.5)$$

$$\sigma_{ruido} \sim \text{Seminormal}(\sigma = 1). \quad (4.6)$$

Definir una buena distribución para los parámetros es fundamental para el correcto progreso del estudio. En este caso, se observa, por ejemplo, que la función sinoidal está desplazada en el eje vertical a un valor cercano a 15. Por lo tanto, al definir la distribución a priori de  $b$ , se establece como una distribución con media 15, reflejando el desplazamiento observado. La distribución de la frecuencia se define como una distribución normal con media 0, lo que indica que no hay sesgo hacia valores positivos o negativos. Las distribuciones de  $\sigma_{ruido}$  y amplitud se definen como distribuciones seminormales para asegurar que los valores sean positivos. La desviación estándar de las distribuciones seleccionadas dependerá de la confianza que se tenga en que los valores de los parámetros sean similares al valor medio esperado.

Definir una desviación estándar baja indicará que se tiene un alto conocimiento sobre el valor del parámetro, y el seleccionar un valor alto supondría que se tiene más incertidumbre sobre el valor del parámetro. En este caso se ha definido 1000 iteraciones de burn in y 1000 iteraciones a tener en cuenta para el estudio.

Se obtiene la distribución a posteriori y gráfica de traza de los parámetros que se representan en la figura 7. La distribución a posteriori representa en el eje  $x$  los diferentes valores que puede tomar el parámetro, y en el eje  $y$  la credibilidad de que cada valor sea el que modele los datos. En el gráfico de traza se representan los valores que va tomando cada parámetro a lo largo de las iteraciones.

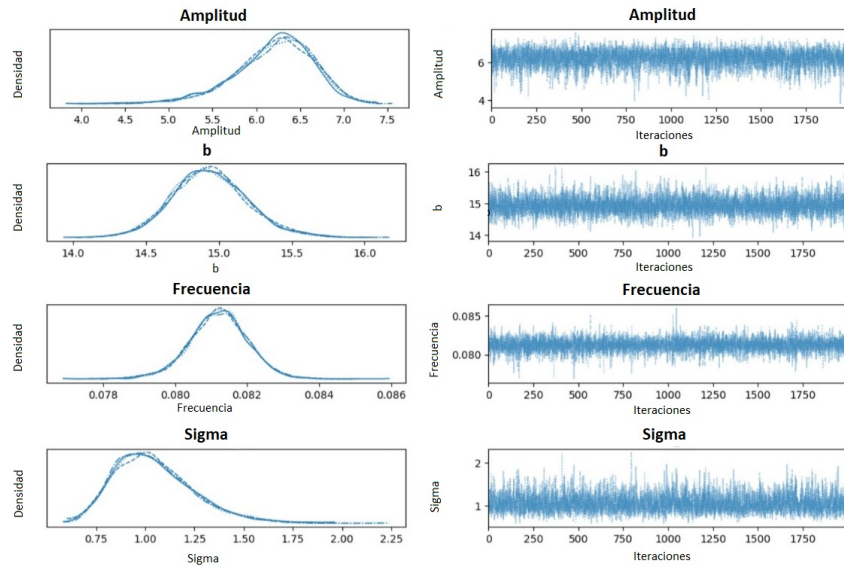


Figura 7: Distribución posterior y gráfico de traza de los diferentes parámetros

Al observar la traza, se observa que el proceso se ha realizado correctamente, explorando todos los valores posibles. Además, la distribución a posteriori ha llegado a converger con valores de credibilidad máxima bien definidos.

Finalmente, tal y como se detalla en la sección 3.4.2, se calculan y se representan los diagnósticos de Gelman-Rubin para cada parámetro. Este resultado es

especialmente útil para sacar conclusiones sobre el análisis, ya que en la traza y la forma de la distribución posterior se puede intuir si existe convergencia, pero con el valor del diagnóstico, se tiene un resultado más exacto de la convergencia de nuestros parámetros. Por ejemplo si se obtiene un valor de  $\hat{R} = 5$ , se observará convergencia gráfica, pero se sabrá que se puede mejorar la elección de la distribución inicial del parámetro para que el diagnóstico sea lo más cercano a 1 posible. Se pueden observar los datos obtenidos en la figura 8.

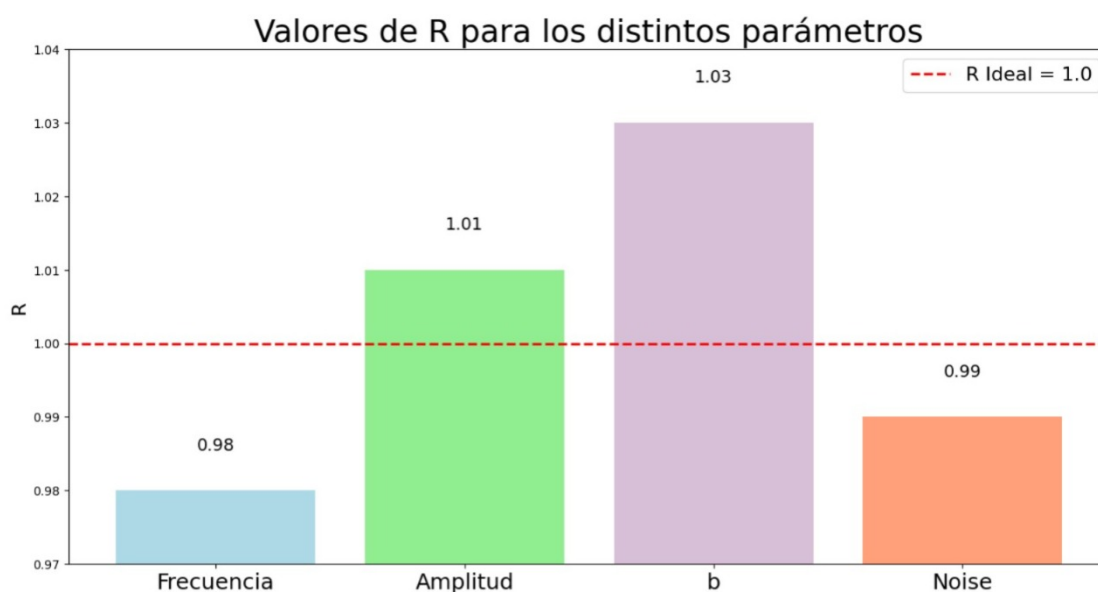


Figura 8: Diagnóstico de Gelman-Rubin para los distintos parámetros. En rojo el valor ideal de convergencia de Gelman-Rubin, 1.

Se observa que para todos los parámetros se obtiene un valor muy cercano a uno, lo que significa que la distribución ha convergido correctamente. Este diagnóstico también permite comparar el grado de convergencia entre los diferentes parámetros. En este caso todos convergen totalmente.

Ahora que se ha comprobado que el proceso cumple con las condiciones necesarias para ser considerado correcto, se va a mostrar el resultado obtenido. Calculando la media y la desviación estándar a lo largo de las cadenas de Markov y usando el valor de máxima credibilidad de los parámetros, se puede realizar la predicción temporal con el intervalo de confianza, como se muestra en la figura 9.

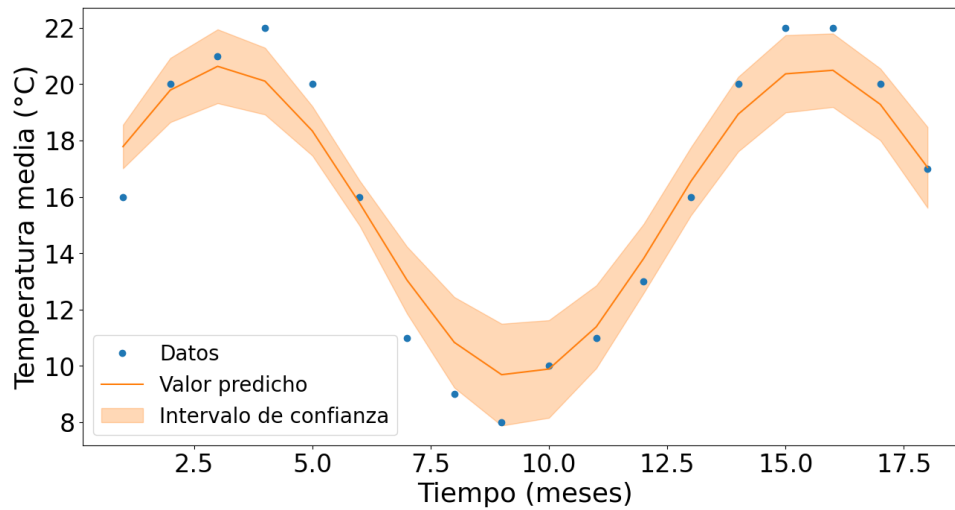


Figura 9: Representación del modelo obtenido de la temperatura a lo largo de los meses junto a los datos y el intervalo de confianza.

El modelo obtenido está bastante próximo a la tendencia que siguen los datos. Para mejorar la precisión se debería definir un modelo más complejo, pero dentro del contexto en el que se ha trabajado se sabe que se ha realizado el mejor ajuste posible debido a la convergencia correcta de las distintas distribuciones.

Gracias al enfoque bayesiano, conforme vayan pasando los meses se podrán ir actualizando las distribuciones de los parámetros con los nuevos datos de temperatura de manera muy dinámica, es decir, si por ejemplo hubiera una gran subida de la temperatura en los siguientes meses, se detectaría que el modelo actual difiere en gran medida de lo observado y se actualizaría de manera notable.

## 4.2 Estudio temporal de la probabilidad de que ocurra un suceso

Ahora, cambiando un poco el enfoque, se busca estudiar la probabilidad de transitar de un estado a otro en un momento determinado. Para ello, se me ocurrió analizar la probabilidad de haber comido en un intervalo de tiempo específico. Hoy en día, controlar los tiempos de las comidas puede ser relevante para diversos objetivos nutricionales, por lo que este puede ser un estudio interesante. Los datos, representados en la figura 10 consisten en un registro de 35 días, donde el eje  $x$  representa la hora desde las 12:00 AM hasta las 5:00 PM en intervalos de 15 minutos, y en el eje  $y$  se indica el estado de ayuno (0) o de haber comido (1).



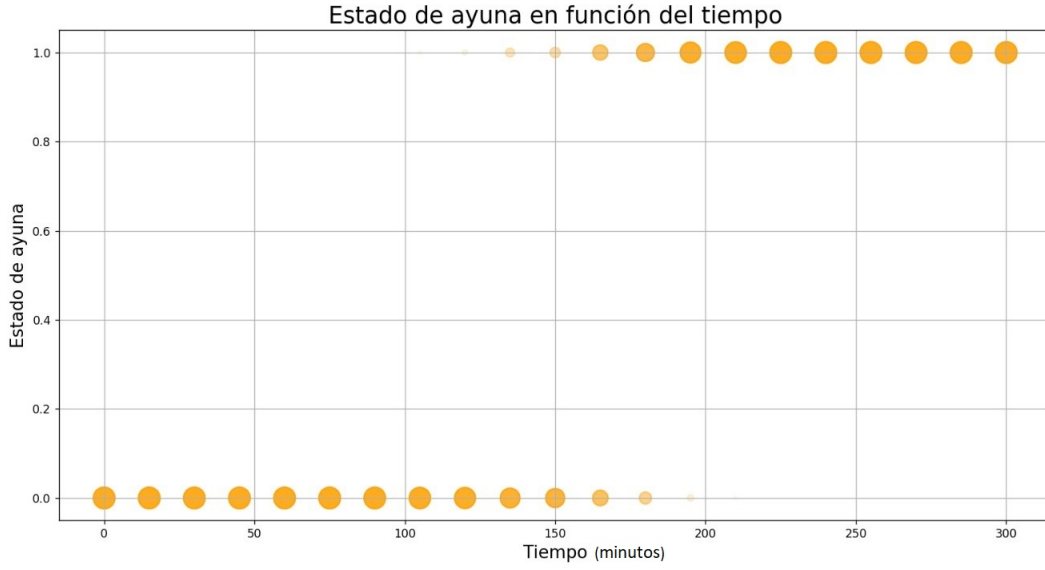


Figura 10: Datos registrados de 35 días sobre el estado de ayuna en función del tiempo.

En la gráfica, los puntos con mayor frecuencia se representan con mayor densidad y tamaño, lo que permite identificar claramente el intervalo con mayor variación. Como he mencionado, en este caso resulta interesante estudiar la probabilidad en función del tiempo,  $P(\text{comer}|\text{tiempo})$ . Se observa que hay determinadas horas en las que es seguro que aún esté en ayunas (como a las 12:00 AM) y otros intervalos en los que también siempre sé que he roto el ayuno. Aunque podría parecer que la probabilidad sigue una función escalón, ya que suelo comer a las 14:30, lo interesante aquí es analizar ese pequeño intervalo que varía según el día por diversas razones. Por ello, la probabilidad, en lugar de ser un porcentaje fijo, se asemeja a la expresión dada por la función logística:

$$\frac{1}{1 + e^{\beta x + \alpha}} \quad (4.7)$$

En este caso los parámetros son  $\beta$  y  $\alpha$ , cuyas distribuciones a priori se definen como normales con media 0 y desviación 0.5:

$$\alpha \sim \mathcal{N}(0.0, 0.05), \quad (4.8)$$

$$\beta \sim \mathcal{N}(0.0, 0.05). \quad (4.9)$$

Para estudiar en este caso el likelihood (definido igual que en el apartado anterior), tenemos que modelar los datos con la distribución de Bernoulli, dado que se está trabajando con resultados de tipo binario:

$$\text{obs} \sim \text{Bernoulli}(p), \quad (4.10)$$

donde,  $p$  es la probabilidad de haber comido (es decir, que la  $y$  sea 1, siendo  $y$  los datos observados).

En términos del likelihood, el programa está modelando la probabilidad de observar los datos y dados los parámetros  $\alpha$  y  $\beta$ .

Definiendo 1000 iteraciones de burn in y 1000 iteraciones a tener en cuenta, se representa en la figura 11 las trazas y distribuciones posteriores de  $\alpha$  y  $\beta$ .

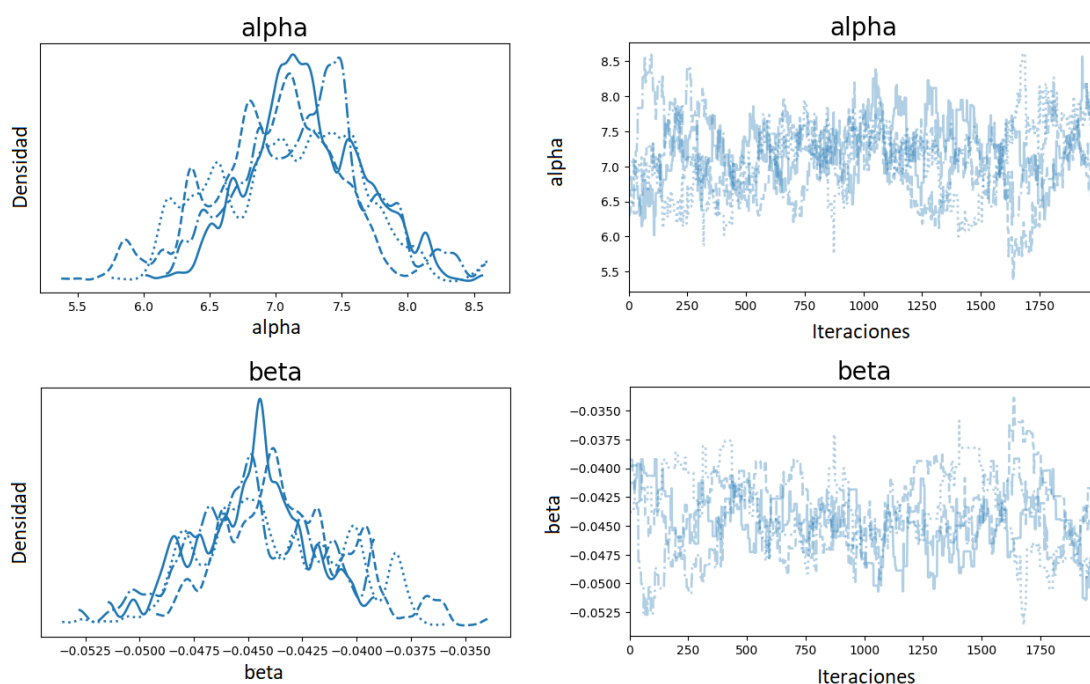


Figura 11: Distribución posterior (en el eje x los diferentes valores que puede tomar el parámetro, en el eje y la credibilidad de que cada valor sea el que modele los datos) y traza (en el eje x las iteraciones, en el eje y los valores que va tomando cada parámetro).

Una vez más, se observa que la distribución posterior converge y que en la traza se explora correctamente el espacio. En este caso las distribuciones a posteriori obtenidas muestran más picos en el caso anterior. Esto se ajusta más a lo que encontramos en los casos de estudios más complejos.

Es relevante destacar que, además de la intuición en la selección de las distribuciones a priori, también se emplea un enfoque de prueba y error. Si se nota que un parámetro presenta una evolución inusual, es un indicio de que se debe reconsiderar la distribución elegida. En este caso he tenido que realizar varias pruebas y analizar la traza para poder encontrar las distribuciones que me permiten realizar el estudio.

Se obtiene también el diagnóstico de Gelman-Rubin, y al igual que antes, representamos gráficamente en la figura 12.

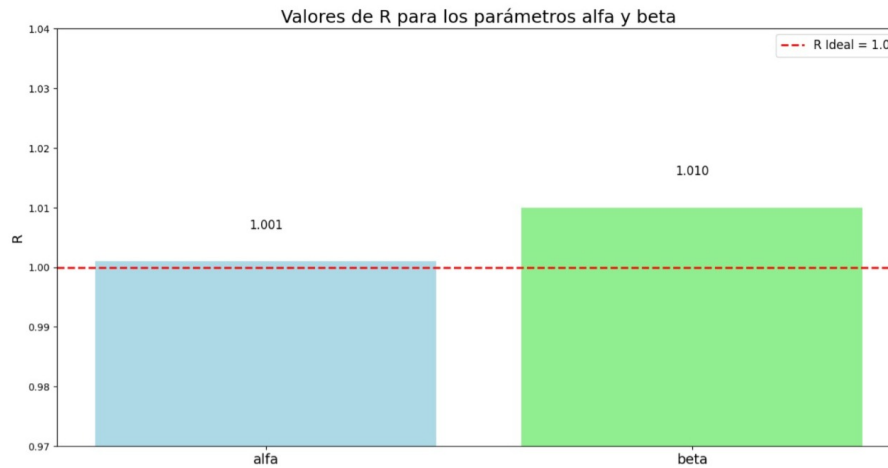


Figura 12: Distribución posterior y traza.

Para ambos parámetros el valor de  $\hat{R}$  es muy cercano a 1, por lo que se puede visualizar de manera numérica que la convergencia para ambos parámetros es total.

Una vez comprobado que el proceso ha sido correcto, usando los valores con mayor credibilidad de las distribuciones obtenidas de los parámetros, se representa en la figura 13 la probabilidad de estar en ayuna en función del tiempo.

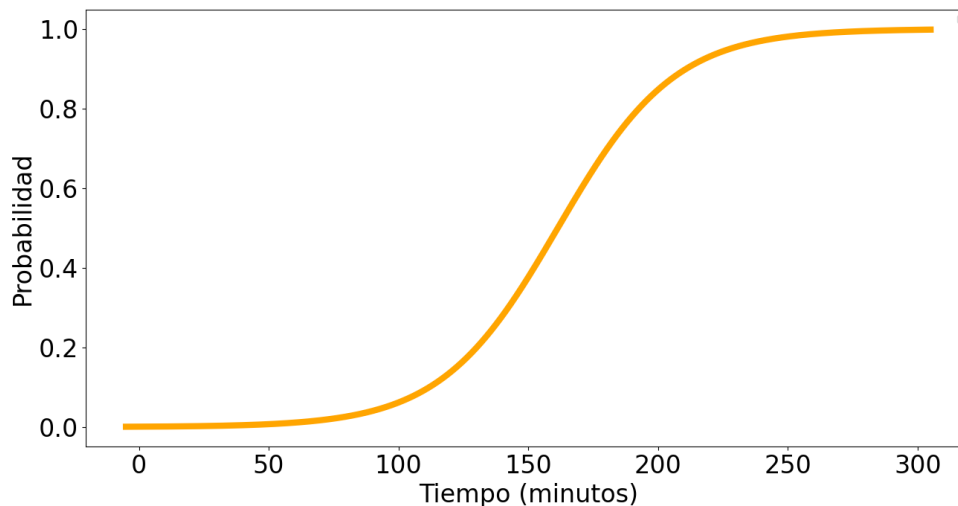


Figura 13: Probabilidad estimada.

Si nos dirigimos a un nutricionista y nos solicita información sobre nuestros hábitos alimenticios para llevar a cabo un estudio, esta metodología nos permitiría proporcionar datos muy detallados y específicos. Además, sería muy útil y eficaz si queremos cambiar nuestra dieta con el tiempo, ya que se podrían actualizar las distribuciones obtenidas al añadir nuevos datos, lo cual es un punto fundamental de la estadística bayesiana. Por ejemplo, si nos apuntáramos a una

nueva actividad, nuestros resultados se adaptarían de manera dinámica a nuestros nuevos hábitos.

### 4.3 Estudio temporal aplicado al campo de la física

Para acabar, tras haber desarrollado paso a paso de manera muy detallada como es el proceso de realizar un estudio bayesiano para modelos más simples, se va a acabar mostrando una aplicación real de estos métodos en el mundo de la física, para visualizar lo útiles que son estas técnicas. Siguiendo la dinámica de [13] y [14], se van a estudiar los ciclos de Milankovitch.

Para ello se ha analizado la desviación de la temperatura (respecto a la media) de la atmósfera terrestre a lo largo de los últimos 80000 años (la temperatura de hace miles de años se obtiene estudiando los núcleos de hielo terrestres). Se observan los datos en la figura 14.

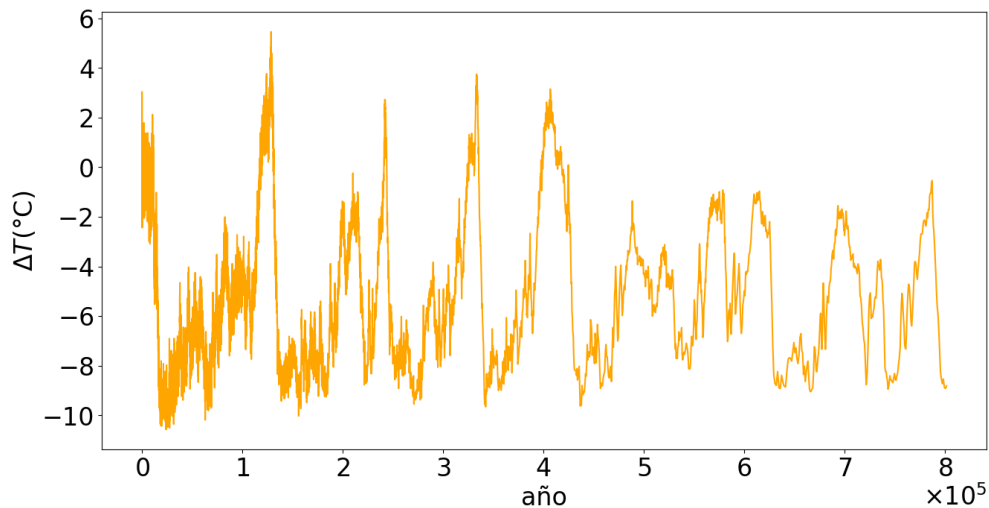


Figura 14: Variación de la desviación de la temperatura(°C ) de la atmósfera terrestre a lo largo de los años.

Se observa que hay cierto patrón cíclico en la variación de la temperatura. Esta periodicidad se debe a los procesos que engloban los ciclos de Milankovitch, entre los que se encuentran la precisión axial, la oblicuidad y la inclinación orbital terrestre.

Siguiendo el razonamiento de la sección 4.1, se define un modelo sinoidal. Sin embargo, como se conocen los periodos de precisión axial (26000 años), de la oblicuidad (41000 años) y de la inclinación orbital (100000 años), se puede emplear esta información para definir el modelo como una suma de tres sinoidales y una constante.

$$\Delta T = a_1 \sin\left(\frac{2\pi t}{p_1}\right) + a_2 \sin\left(\frac{2\pi t}{p_2}\right) + a_3 \sin\left(\frac{2\pi t}{p_3}\right) + T_0, \quad (4.11)$$

donde los parámetros  $a_1, a_2, a_3$  son las amplitudes de las sinoidales,  $p_1, p_2, p_3$  son los períodos de las sinoidales, y  $T_0$  es una constante que ajusta la media de las fluctuaciones.

En este caso las distribuciones a priori de todos los parámetros son definidas como distribuciones uniformes (las 3 amplitudes son definidas uniformemente entre 0 y 5, las frecuencias entre 10000 y 200000, y la  $T_0$  entre -10 y 0). Se han realizado 100 pasos de burn in y 500 iteraciones a tener en cuenta. Como se explica en 3.4.3, en estudios con gran cantidad de parámetros, la forma más eficaz para visualizar los datos es con el Corner plot, en el que se podrán observar las distribuciones a posteriori de cada parámetro y la correlación existente entre estas. Podemos visualizarlo en la figura 15.

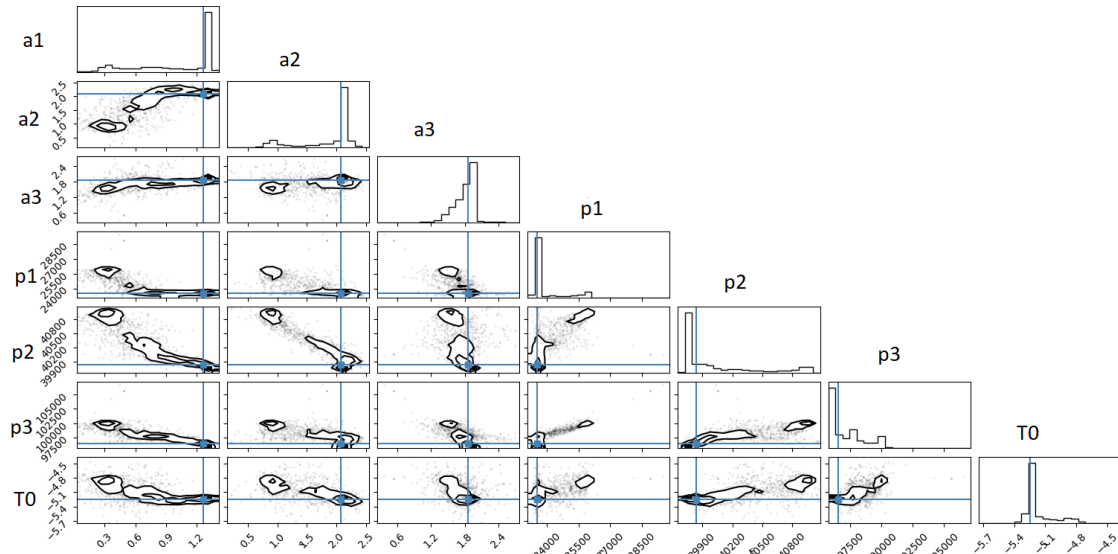


Figura 15: Corner plot, en el que se pueden observar las distribuciones a posteriori (diagonal superior) junto con los diagramas de correlación.

Para interpretar este diagrama hay que fijarse en la correlación de dos parámetros entre sí (por ejemplo el diagrama en la segunda fila y primera columna es el diagrama de correlación entre las variables  $a_2$  y  $a_1$ ). Si los puntos de este diagrama forman una figura estrecha y diagonal, significará que los dos parámetros se encuentran fuertemente correlacionados. En cambio, si los puntos se encuentran más dispersos y menos localizados podría llegar a significar que no hemos alcanzado la convergencia. En este caso se observa que todos los parámetros están correctamente correlacionados entre sí. En la diagonal superior de la figura se puede visualizar la forma de las distintas distribuciones a posteriori de los parámetros. Estas han convergido a una forma muy definida, lo que indica que se ha realizado un buen ajuste.

Finalmente, empleando los valores con mayor credibilidad de las distribuciones de los parámetros, se muestra el modelo obtenido en la figura 16.

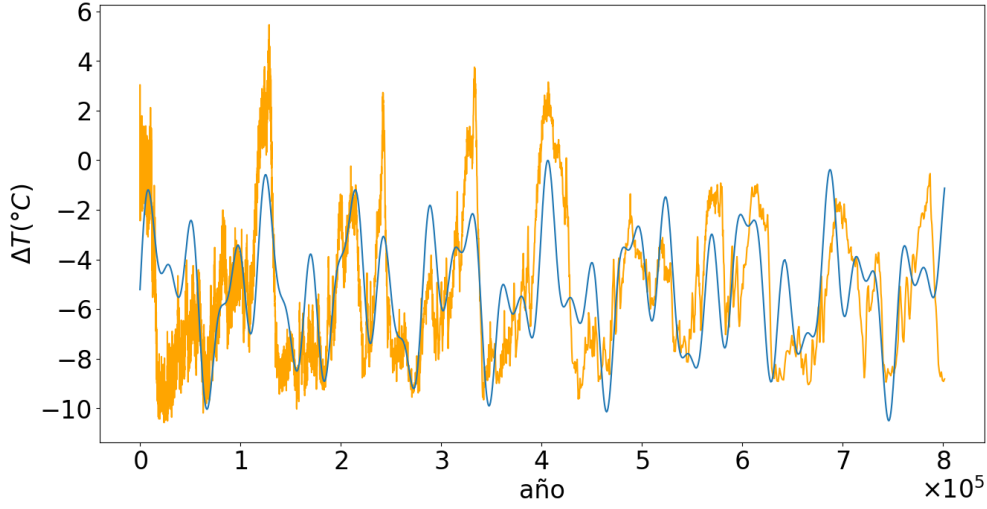


Figura 16: En azul el modelado de la variación de la temperatura a lo largo de los años y en naranja los datos de la variación de la temperatura.

Se ve que el modelo obtenido sigue una tendencia muy similar a la de los datos. Los valores de los periodos con mayor credibilidad son  $p1 = 23517$  años,  $p2 = 39742$  años y  $p3 = 96095$  años, que son bastante similares a los periodos conocidos y comentados anteriormente, por lo que se puede intuir que el razonamiento inicial ha sido correcto. Para obtener mejores resultados se podría seguir añadiendo términos al modelo usando los conocimientos físicos que se tengan sobre el caso de estudio.

## 5 Conclusions.

The aim of this project has been to introduce and elucidate the value of the Bayesian approach for making time series predictions.

The first objective was to provide an initial introduction to Bayesian statistics itself and explain the rationale behind applying these concepts. Since Bayesian reasoning is not commonly used, it is essential to first understand its fundamentals and utilities. To achieve this, the initial definitions and characteristics were introduced, and a comparison with frequentist statistics was made to highlight the fundamental values of this approach. After explaining that the idea in this context is to study the parameters that will form a model to obtain their distributions, we began detailing these distributions and their understanding within this context.

Next, the Monte Carlo Method and Markov chains were explained. These concepts help us understand how inference is practically conducted. Following a discussion on the function and utility of Monte Carlo methods, we defined Markov chains and their mathematical properties, which are crucial as they explain the convergence we aim to achieve in the distributions of our parameters. Various

analysis methods used in these studies to visualize and understand the process (trace, distribution plot, Gelman-Rubin diagnostic and Corner plot) were also explained.

Finally, a practical study of three different cases was conducted to explicitly demonstrate how the previously explained concepts are applied and what such a study would entail. In the first case, we studied the average temperature data in Málaga over the months, observing that the data seemed to follow a sinusoidal trend. We defined four parameters for the study and observed convergence across various types of analysis. By representing with the confidence interval, we were able to compare the obtained model with the data, confirming that the data indeed follow a trend very similar to the one obtained.

In the second case, we studied the probability of a certain event occurring over time. Specifically, we examined the probability of fasting based on the time of day. Observing the trend, we see that the probability resembles the logistic function, for which we defined the parameters  $\alpha$  and  $\beta$ . Once again, we obtained satisfactory results.

In the last case studied, we conducted a study applied to physics. We examined the Milankovitch cycles, modeling the variation in the temperature deviation of the atmosphere over the years. By observing the data fluctuations and knowing the periods of the processes that encompass the Milankovitch cycles, we defined a model composed of the sum of three sinusoidal terms and a constant. To analyze this more complex example, we used the Corner plot for correlation. The resulting model closely follows the trend of the data.

Through this work, we hope to have demonstrated the intriguing potential of Bayesian methods, explaining their principles, and understanding how they could be applied.

## Referencias

- [1] John Kruschke. *Doing Bayesian Data Analysis. A Tutorial with R, JAGS, and Stan*. Dept. of Psychological and Brain Sciences Indiana University, Bloomington, 2022.
- [2] David Pérez Ros. *Métodos Monte Carlo basados en cadenas de Markov*. Universidad de Zaragoza, 2022.
- [3] David J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
- [4] Andrew Gelman. *Bayesian Data Analysis*. Columbia University, 1995.
- [5] Robert, C.P. and Casella, G. *Introducing Monte Carlo Methods with R*. Springer, 2010.
- [6] José Jiménez. *Métodos Monte Carlo basados en cadenas de Markov*. 2015.
- [7] Norris, J. R. *Markov Chains (Cambridge Series in Statistical and Probabilistic Mathematics)*. 1998.
- [8] Conchi Ausín. *Tema 8: Métodos de cadenas de Markov Monte Carlo*. Universidad Carlos III de Madrid, 2012.
- [9] Ricardo Medel Esquivel†, Isidro Gómez Vargas. *An introduction to Markov Chain Monte Carlo*. Instituto de Ciencias Físicas UNAM, Av. Universidad s/n, 2019.
- [10] Ana Paula Palacios y Peter Diko. *Introducción a la estadística bayesiana, aplicaciones y métodos*. Universidad Carlos III de Madrid, 2011.
- [11] Learn PyMC & Bayesian modeling. <https://www.pymc.io/welcome.html>.
- [12] Weather Spark. <https://es.weatherspark.com/>.
- [13] A. Malinverno<sup>1</sup> and S. R. Meyers. *Bayesian Estimation of Past Astronomical Frequencies, Lunar Distance, and Length of Day From Sediment Cycles*. *Geochemistry, Geophysics, Geosystems*, 2024, 25, e2023GC011176.
- [14] Imad Pasha. *Fitting the Earth's Milankovich Cycles*. 2017. [https://github.com/prappleizer/prappleizer.github.io/tree/master/.ipynb\\_checkpoints](https://github.com/prappleizer/prappleizer.github.io/tree/master/.ipynb_checkpoints).