



UNIVERSIDAD AUTÓNOMA DE GUERRERO

Facultad de Matemáticas

Maestría en Matemáticas Aplicadas

Estimación Bayesiana en El Modelo Lotka-Volterra

T E S I S

PARA OBTENER EL GRADO DE:

Maestro en Matemáticas Aplicadas

PRESENTA:

Jaime Andrés Cerda Garrido

DIRECTOR DE TESIS:

Dr. Francisco J. Ariza Hernandez

Dr. Jorge Sánchez Ortíz

Abril de 2018.

Dedicatoria

Para ti Olga Perez por siempre creer en que mi vida sería un largo caminar por las matemáticas.

Agradecimientos

A mi Padre Fernando, mi Madre Olga y Hermano Carlos por todo su apoyo incondicional.

A mi novia por su incondicionalidad en este último período.

A mis amigos Andreita y Maca por su preocupación y cariños. A mis compañeros, por todos sus consejos y su apoyo académico ya que sin esté, no hubiera sido posible terminar este pequeño paso.

A mis directores de Tesis:

Martín, Francisco y Jorge, por su apoyo constante y paciencia en la elaboración de este trabajo, pero lo más importante por dejar de ser maestros y pasar a ser grandes amigos.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por haberme brindado el apoyo económico necesario para realizar mis estudios de posgrado.

Índice general

Agradecimientos	II
1. Introducción	1
1.1. Problema de Investigación	1
1.2. Antecedentes	1
1.3. Objetivos	2
1.3.1. Objetivo General	2
1.3.2. Objetivos Específico.	2
1.4. Planteamiento del problema	3
2. Modelo	5
2.1. Modelo de Lotka-Volterra.	5
2.2. Método de Runge-Kutta.	9
3. Problema inverso Bayesiano para el modelo L-V	12
3.1. Teoría Bayesiana	12
3.1.1. Introducción	12
3.1.2. Definiciones y Teoremas Básicos	13
3.1.3. El Teorema de Bayes	14
3.2. El problema de inferencia	16
3.3. Principios fundamentales en Inferencia	17
3.3.1. El Principio de Verosimilitud	17
3.3.2. El Principio de Suficiencia	18
3.3.3. El Principio de Condicionalidad	18
3.4. Modelo de Inferencia para la distribución Normal	18

3.4.1. Modelo con varianza conocida	19
3.5. MCMC	21
3.5.1. Muestreo de Gibbs	22
3.5.2. Algoritmo M-H	23
3.5.3. Algoritmo T-Walk	26
3.6. Diagnósticos de convergencia	28
3.7. Metodología Bayesiana.	31
3.8. Estudio de Simulación	33
3.9. Ejemplos	42
4. Conclusiones	45

Introducción

1.1. Problema de Investigación

Realizaremos estimación de parámetros al sistema de Lotka-Volterra utilizando un enfoque Bayesiano.

1.2. Antecedentes

Resulta muy común pensar que dos especies pueden estar relacionadas de alguna manera, como: por los tipos de alimentos que consumen, el color de su piel, si son carnívoros o herbívoros, etc.

Esto nos lleva a pensar de que manera podrían estar relacionadas dos especies, el depredador y la presa. Los inicios de este problema sucede cuando Lotka (1920) resuelve la ecuación para poder explicar el crecimiento de la bacteria en un matraz. Luego se pregunta: "¿Que sucedería si una especie se quisiera comer a la otra?". Al mismo tiempo el matemático y físico Vito Volterra se había planteado la misma pregunta por una problemática planteada por un familiar cercano, esto los llevo a solucionar un sistema de ecuaciones lineales, siendo así como nace este famoso modelo de Lotka-Volterra.

Desde entonces han utilizado este modelo (presa-depredador) para muchos tipos de competencias de especies logrando buenas aproximaciones, pero la pregunta sería: "¿Es un buen método para calcular su solución?", "¿Existirá otro método para estimarlo?".

La comunidad científica a dado muy buenos métodos para solucionar en forma mas exacta los sistemas de ecuaciones no lineales, uno de estos métodos es el de Runge Kutta de cuarto orden.

El método consiste básicamente en una estructura idéntica a el método de Taylor pero sin necesidad del cálculo de las derivadas de orden superior, este método nos proporcionara el cálculo de las constantes del modelo.

Hay muchos trabajo y publicaciones con este tipo de resolución en revistas del área de las matematicas, salud, biología, etc (12). Pero nosotros proporcionaremos una nueva forma del cálculo de estas soluciones utilizando la estadística Bayesiana.

La estadística Bayesiana tiene muchos ambitos en los cuales hace los procedimientos mas sencillos y simple de resolver. En nuestro caso nuestros parámetros poseen cierta distribución a priori (2) así nosotros obtendremos una nueva distribución posteriori (puede ser una de la misma clase con distintos parámetros u otra distinta) y así poder lograr una mejor estimación de los parámetros.

Los trabajos que se asemejan un poco al que vamos a realizar son (8) y (4). El primero de estos realizan la estimación de parámetros para el modelo de crecimiento poblacional utilizando estadística Bayesiana, en el segundo de ellos logran mejorar el modelo Lotka-Volterra utilizando un sistema ecuaciones diferenciales no lineales. Mezclando ambos conocimientos trataremos de dar solución a este sistema de ecuaciones diferenciales de forma más acertiva que otros métodos utilizados anteriormente.

1.3. Objetivos

1.3.1. Objetivo General

- Estimar los parámetros del modelo clásico de Lotka-Volterra desde una perspectiva Bayesiana.

1.3.2. Objetivos Específico.

- Obtener una solución numérica para el sistema dinámico de Lokta-Volterra.
- Calcular la función de verosimilitud a partir de la solución numérica en cada iteración

Monte Carlo.

- Obtener la distribuciones marginales de cada uno de los parámetros del modelo L-V usando métodos MCMC.
- Mostrar un ejemplo con datos reales que ejemplifique nuestra metodología.

1.4. Planteamiento del problema

En este trabajo, se plantea un modelo general para los problemas inversos que serán estudiados posteriormente.

Modelo General:

1. Sistema Dinámico:

$$\begin{cases} x'_1 &= x_1(a - bx_2), \\ x'_2 &= x_2(-c + dx_1), \\ X(t_0) &= X_0 \end{cases}$$

2. Ecuación de Observación (2):

$$y_i = h(\mathbf{X}(t_i)) + \epsilon_i, \quad i = 1, \dots, n \quad (1.1)$$

Asumiendo que las observaciones $\mathbf{y} = (y_1, y_2, \dots, y_n)$, en tiempos discretos t_1, t_2, \dots, t_n donde y_i corresponden al i -ésimo valor observado bajo incertidumbre, $h : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es la función de observación, $X(t)$ corresponde a la solución numérica del sistema (1.1). Se desea estimar $\boldsymbol{\theta} = (a, b, c, d)$, el cual es un vector de parámetros desconocidos, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$; esto es con la finalidad de caracterizar dicho sistema. La condición de Lipschitz para la función $H : \mathbb{R}^n \times [0, T] \times \Theta \rightarrow \mathbb{R}^n$ asegura la existencia y unicidad de las soluciones del problema con valor inicial (1.1). Existen varios tipos de funciones de observación h que pueden ser consideradas, por ejemplo modelar una sola componente del p -vector $X(t)$ o una combinación lineal de ellas, (4). Aquí, sólo se considera un problema de observaciones uni-dimensionales. $X^*(t)$ representa las observaciones de interés con error de medición para las observaciones en t_i , los cuales son consideradas como variables aleatorias independientes e idénticamente

distribuidas (*i.i.d*) de una distribución normal bivariada con media cero y matriz de varianza y covarianza Σ , denotada por $\epsilon_i \sim N_2(0, \Sigma)$.

La presente tesis está estructurada como sigue: en el Capítulo 2, se presenta el modelo y una descripción de este y su solución numérica. En el Capítulo 3, se aborda el Análisis Bayesiano a problemas inversos, resultados y ejemplos. Y finalmente en el Capítulo 4, analizamos las conclusiones de las estimaciones para nuestro modelo presa-depredador. Después de las referencias bibliográficas y anexos.

Modelo

2.1. Modelo de Lotka-Volterra.

En este capítulo abordaremos el estudio del sistemas dinámico presa-depredador, estos que describen la interacción entre las especies que conviven en un cierto habitad natural. En este caso estudiaremos un modelo en particular que es el de Lotka-Volterra. Este modelo es encontrado especialmente en la biología y la ecología.

Vamos a definir algunos conceptos básicos para tener una mejor comprensión de los sistemas dinámicos.

Sistema autónomo: un sistema de ecuaciones lineales de primer orden se dice que es autónomo cuando se puede escribir de la forma:

$$\frac{dx}{dt} = f(x) \tag{2.1}$$

Donde $x \in \mathbb{R}^n$ y $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, es decir, que la función vectorial f no depende explícitamente del tiempo.

Solución: una solución al sistema (2.1) es una aplicación de $t \rightarrow x(t)$ de algún intervalo $I \subset \mathbb{R} \subset \mathbb{R}^n$, diferenciable en todas sus componentes, y que satisface $\frac{dx(t)}{dt} = f(x(t))$ para

cada una de ellas.

Condición inicial: es una especificación del valor de $\mathbf{x}(\mathbf{t})$ en un instante determinado, habitualmente $\mathbf{t} = \mathbf{0}$. La condición inicial se establece por medio de una ecuación de la forma: $\mathbf{x}(\mathbf{0}) = \mathbf{x}_0$, con $\mathbf{x}_0 \in \mathbb{R}^n$.

Teorema de existencia y unicidad (Picard): Si $\mathbf{G} \subset \mathbb{R}^n$ es un conjunto abierto, $\mathbf{f} : \mathbf{G} \rightarrow \mathbb{R}^n$ es localmente Lipschitziana para todo $\mathbf{x} \in \mathbf{G}$ y $x_0 \in \mathbf{G}$, entonces la ecuación diferencial $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ con $\mathbf{x}(\mathbf{t}_0) = \mathbf{x}_0$ tiene una solución única $\mathbf{x} : \mathbf{I} \rightarrow \mathbf{U}$ para algún intervalo abierto \mathbf{I} que contiene a \mathbf{t}_0 . Habitualmente exigiremos que todas las componentes \mathbf{f} sean diferenciables, lo cual implica la condición de Lipschitz.

Punto crítico: un punto crítico del sistema autónomo es un vector \mathbf{x}_1 tal que $\mathbf{f}(\mathbf{x}_1) = \mathbf{0}$. En este caso es evidente que $\mathbf{x}(\mathbf{t}) = \mathbf{x}_1$ es una solución del sistema autónomo (2.1).

Espacio de fase: es el espacio \mathbb{R}^n donde se representa todas las coordenadas del vector $x(t)$. Si $n = 2$ se denomina plano de fase.

Trayectoria u órbita: es la proyección de la gráfica de la solución $\{t, x(t)\} \in \mathbb{R}^{n+1}$ en el espacio de fase.

Mapa de fase: es el conjunto de trayectorias representadas en el espacio de fase.

Punto crítico estable: un punto crítico z de una ecuación diferencial $\frac{dx}{dt} = f(x)$ se denomina estable (y también Lyapunov-estable) si para cada entorno de U de z , existe otro entorno W de z tal que toda trayectoria que parte de W ésta contenida en U , es decir, que $x_0 \in W$ entonces $x(t) \in U$ para todo $t \geq 0$.

El modelo de Presa-Depredador o de Lotka-Volterra (1):

$$\begin{aligned}x_1' &= x_1(a - bx_2), \\x_2' &= x_2(-c + dx_1),\end{aligned}\tag{2.2}$$

Donde $x_i' := \frac{dx_i(t)}{dt}$, $i = 1, 2$. Como un modelo de competencia de especies, $x_1(t)$ puede ser interpretada como el tamaño de la población presa en el tiempo t y $x_2(t)$ como el tamaño de una población depredadora en el tiempo t . Los parámetros a , b , c y d son las constantes positivas y que son **desconocidas**.

Siendo F el sistema dinámico:

$$F \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 - bx_1x_2 \\ -cx_2 + dx_2x_1 \end{pmatrix} \quad (2.3)$$

Calculamos la matriz Jacobiana:

$$DF \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{df_1}{dx_1} & \frac{df_1}{dx_2} \\ \frac{df_2}{dx_1} & \frac{df_2}{dx_2} \end{pmatrix} = \begin{pmatrix} a - bx_2 & -bx_1 \\ dx_2 & -c + dx_1 \end{pmatrix} \quad (2.4)$$

Igualando cada componente de cada fila a cero, tenemos:

$$\begin{aligned} a - bx_2 &= 0 & bx_1 &= 0 \\ x_2 &= \frac{a}{b} & x_1 &= 0 \end{aligned}$$

La segunda fila:

$$\begin{aligned} dx_2 &= 0 & -c + dx_1 &= 0 \\ x_2 &= 0 & x_1 &= \frac{c}{d} \end{aligned}$$

Cuando $x_1 = x_2 = 0$, tenemos un punto silla con valores propios a y $-c$.

Sabemos que las curvas son estables e inestables, en el eje x_2 (est.) y eje x_1 (inest.).

El punto de equilibrio $(\frac{c}{d}, \frac{a}{b})$.

A partir de este análisis no podemos determinar el comportamiento de las soluciones, posiblemente podrían ir en espiral hacia el punto de equilibrio, o ir en espiral hacia un ciclo límite o salir del espiral hacia el infinito. Para esto buscamos la función de Liapunov L .

Empleando el truco de separación de variables buscamos la función de la forma:

$$L(x_1, x_2) = F(x_1) + G(x_2) \quad (2.5)$$

L : es la derivada con respecto al tiempo, a lo largo de las soluciones:

$$\begin{aligned} dL(x_1, x_2) &= \frac{d}{dt}L(x_1(t), x_2(t)) \\ &= \frac{dF}{dt}x'_1 + \frac{dG}{dt}x'_2 \\ &= x_1 \frac{dF}{dt}(a - bx_2) + x_2 \frac{dG}{dt}(-c + dx_1) \end{aligned}$$

Si $L \equiv 0$, entonces:

$$\frac{x_1 \frac{dF}{dt}}{dx_1 - c} \equiv \frac{x_2 \frac{dG}{dt}}{bx_2 - a}$$

Si x_1 y x_2 son V.I., esto es posible si y solo si:

$$\frac{x_1 \frac{dF}{dt}}{dx_1 - c} \equiv \frac{x_2 \frac{dG}{dt}}{bx_2 - a} = k$$

Siendo k una constante. Si $k = 1$ e integrando, tenemos:

$$\begin{aligned} \frac{dF}{dt} &= d - \frac{c}{x_1} && \longrightarrow F(x_1) = dx_1 - c \ln(x_1) \\ \frac{dG}{dt} &= b - \frac{a}{x_2} && \longrightarrow G(x_2) = bx_2 - a \ln(x_2) \end{aligned}$$

Funcion L : Al considerar los signos de la derivada, podemos notar que el punto de equilibrio $Z \left(\frac{c}{d}, \frac{a}{b} \right)$, es un mínimo absoluto para L . Se deduce que $L[L - L(Z)]$ es una función de Liapunov para el sistema. Por lo tanto, Z es un punto de equilibrio estable. Esto quiere decir que el sistema no posee ciclo límite, porque L no es constante en ningún conjunto abierto.

Teorema: Cada solución del sistema Presa-Depredador es una orbita cerrada(excepto el punto de equilibrio y los ejes).

Demostración: Consideramos una solución $W \neq Z$, donde W no se encuentra en los ejes coordenados. Esta solución gira en espiral alrededor de Z , las soluciones cruzan infinitamente, por lo tanto la secuencia es doblemente infinita $\dots < t_{-1} < t_0 < t_1 < \dots$ tal que $\phi_{t_n}(W)$ esta en la línea $x = \frac{c}{d}$ y $t \rightarrow \pm\infty$ y $n \rightarrow \pm\infty$. Si W no es orbita cerrada, por los puntos $\phi_{t_n}(W)$ serían monotonos a lo largo de la línea.

Como sabemos que no hay ciclos límites ya sea $\phi_{t_n} \rightarrow Z$ y $n \rightarrow \infty$ o $\phi_{t_n} \rightarrow Z$ y $n \rightarrow -\infty$ como L es constante a través de la solución de W esto implica que $L(W) = L(Z)$, pero esto contradice el mínimo de $L(Z)$. Esto completa la prueba.

Concluimos que para cualquier población inicial $(x_1(t_0), x_2(t_0))$ con $x_1(0) \neq 0$ y $x_2(0) \neq 0$ distintos de Z , las poblaciones de presas y depredadores oscilan cíclicamente. No importa cuales sean las poblaciones de P-D, tampoco las especies morirán, y su población no crecera indefinidamente.

2.2. Método de Runge-Kutta.

Los análisis Bayesianos de problemas inversos se concentran en gran medida en el cálculo numérico aproximados de las constantes de la EDO's. Tenemos muchos métodos de aproximaciones numéricas, por las cuales poder abordar y dar solución al problema planteado en el modelo (1) .

El método de Taylor tienen la propiedad de un error local de truncamiento de orden superior, pero la desventaja de requerir el cálculo y la evaluación de las derivadas de $h(x(t), t)$. Esto resulta algo lento y complicado, en la mayoría de los problemas, razón por la cual, en la práctica casi no se utilizan. El método de Euler, lamentablemente requiere de un paso muy pequeño para una precisión razonable.

Los métodos de Runge-Kutta tienen el error local de truncamiento del mismo orden que los métodos de Taylor, pero prescinden del cálculo y evaluación de las derivadas de la función $h((x(t), t)$ con valor inicial $x(0) = x_0$.

$$\begin{cases} \frac{dx(t)}{dt} = G(x(t), t) \\ x(0) = x_0 \end{cases} \quad (2.6)$$

A continuación se entrega la forma general para utilizar el método (4):

$$x_{n+1} = x_n + \frac{\Delta t}{6}(k_1 + 2k_2 + 2k_3 + k_4) \quad (2.7)$$

En donde:

$$\Delta t = t_{n+1} - t_n; n = 1, 2, \dots, N$$

N : es el número de iteraciones.

Si bien con facilidad se pueden deducir otras fórmulas, el algoritmo expresado, se denomina método de Runge-Kutta de orden n . Este algoritmo es de uso extendido, y reconocido como una valiosa herramienta de cálculo, por la buena aproximación que produce. Procederemos a desarrollar hasta el cuarto orden denominado como RK4.

$$k_1 = G(x_n)$$

$$k_2 = G(x_n + \frac{\Delta t}{2}k_1, t_n + \frac{1}{2}\Delta t)$$

$$k_3 = G(x_n + \frac{\Delta t}{2}k_2, t_n + \frac{1}{2}\Delta t)$$

$$k_4 = G(x_n + \Delta tk_3, t_n + \Delta t)$$

$$x_{n+1} = x_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

El error de truncamiento se da en cada paso del método numérico, se expresa como:

$$\varepsilon_n = \| x(t_n) - x(t_{n-1}) - \Delta t k(t_n, x(t_{n-1}), \Delta t, G) \|_2$$

Error global es la diferencia entre la solución calculada y la verdadera Solución a cualquier valor dado de t perteneciente.

$$E_n = \| x_\theta(t_n) - x_{\theta,n} \|$$

Asumimos que la función F toma el papel del lado derecho del problema de valor inicial (2.6) es continuamente diferenciable con respecto a X_θ forma un paralelepípedo $R = \{t_0 \leq t \leq t_0 + a_1, \|X_\theta - X_0\|_2 \leq b\}$, para cada θ . Por lo tanto, existe una solución única del problema del valor inicial en una vecindad de t_0, X_0 para θ . Entonces K es Lipschitz continua en X_θ , es decir, se tiene que $L \in \mathbb{R}^{+*}/(t, x), (t, y) \in \mathbb{R}$, entonces:

$$E_n = \| K(t, x, h, G) - K(t, y, h, G) \|_2 \leq L \| x - y \|_2$$

Proposición 1. Si asumimos (3.3). Supongamos que la anterior suposición se cumple y para el F tenemos un error de truncamiento local e_n es $O(h^{(p+1)})$ Cuando el error de truncamiento global es $O(h^{(p+1)})$.

De la Proposición 1, deducimos que, para cualquier método explícito de un solo paso de orden p como Euler ($p = 1$) y Runge-Kutta ($p = 2$ o $p = 4$) sistema, el error global es de orden $O(h^{p+1})$. Para h lo suficientemente pequeño. Si ponemos $X_\theta^h(t_n) = X_{\theta,n}$, nota que hemos demostrado que:

$$\max_{t \in \{t_1, t_2, \dots, t_n\}} \| X_\theta(t) - X_\theta^h(t) \| \leq C_\theta h^p$$

Con este método ya podemos aproximar los posibles valores de $x^* = x_1, x_2, \dots, x_n$.

Problema inverso Bayesiano para el modelo L-V

3.1. Teoría Bayesiana

3.1.1. Introducción

En general, se usan probabilidades de modo informal para expresar la información o la incertidumbre que se tiene acerca de observaciones de cantidades desconocidas. Sin embargo, el uso de probabilidades para expresar la información se puede hacer de modo formal. Desde el punto de vista matemático se puede demostrar que con el Cálculo de Probabilidades se puede representar de modo numérico el conjunto de racional de creencias, de modo que existe una relación entre probabilidad y e información y la regla de Bayes proporciona un modo natural de actualización de las creencias cuando aparece nueva información. Este proceso de aprendizaje inductivo por medio de la regla de Bayes es la base de la Inferencia Bayesiana. De manera general, los métodos bayesianos son métodos de análisis de datos que se derivan de los principios de la inferencia bayesiana. Estos métodos, proporcionan:

- Estimadores de los parámetros que tienen buenas propiedades estadísticas;
- Una descripción parsimoniosa (simple) de los datos observados;
- Estimación de los datos *missing* y predicciones de futuras observaciones;
- Una metodología computacional potente para la estimación, selección y validación de

modelos.

La metodología bayesiana consta de tres pasos fundamentales:

1. Especificar un modelo de probabilidad que incluya algún tipo de conocimiento previo (*a priori*) sobre los parámetros del modelo dado.
2. Actualizar el conocimiento sobre los parámetros desconocidos condicionando este modelo de probabilidad a los datos observados.
3. Evaluar el ajuste del modelo a los datos y la sensibilidad de las conclusiones a cambios en los supuestos del modelo.

La diferencia fundamental entre la estadística clásica (*frecuentista*) y la bayesiana es el concepto de probabilidad. Para la estadística clásica es un concepto objetivo, que se encuentra en la naturaleza, mientras que para la estadística bayesiana se encuentra en el observador, siendo así un concepto subjetivo. De este modo, en estadística clásica sólo se toma como fuente de información las muestras obtenidas suponiendo, para los desarrollos matemáticos, que se pueden tomar tamaños límite de las mismas. En el caso bayesiano, sin embargo, además de la muestra también juega un papel fundamental la información previa o externa que se posee en relación a los fenómenos que se tratan de modelizar.

3.1.2. Definiciones y Teoremas Básicos

El concepto básico en estadística bayesiana es el de probabilidad condicional: Para dos sucesos A y B ,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \propto P(A \cap B)$$

Se puede aplicar esta definición también a variables discretas o continuas. Desde el punto de vista bayesiano, todas las probabilidades son condicionales porque casi siempre existe algún conocimiento previo o experiencia acerca de los sucesos.

Ley de la Probabilidad Total:

Para un suceso A y una partición B_1, \dots, B_k ,

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$$

Se puede aplicar el teorema a variables discretas:

$$f(x) = \sum_y f(x|Y=y)P(Y=y)$$

o a variables continuas:

$$f(x) = \int f(x|y)f(y)dy$$

3.1.3. El Teorema de Bayes

Se tiene que, para los sucesos A_1, \dots, A_n y B ,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B|A_i)P(A_i)} \propto P(B|A_i)P(A_i)$$

Desde el punto de vista de la estadística bayesiana se tienen dos tipos de cantidades: conocidas y desconocidas, de modo que el objetivo es usar las cantidades conocidas, junto con un modelo paramétrico dado, para hacer inferencias sobre las cantidades desconocidas. Por cantidades desconocidas se puede entender tanto parámetros del modelo como observaciones *missing*. En un modelo básico se tiene un parámetro de interés θ y unos datos observados D y se considera una distribución de probabilidad conjunta para ambos que recoge cómo se relacionan: $p(\theta, D)$. Aplicando la definición de probabilidad condicionada, se tiene que

$$p(\theta, D) = p(\theta)p(\theta|D)$$

$p(\theta)$ se denomina distribución a priori de θ . El segundo término es la *función de verosimilitud*. En realidad, estamos interesados en determinar $p(\theta|D)$ que es la distribución de la cantidad desconocida dada la cantidad conocida. Como $p(\theta|D) = p(D|\theta)$, entonces:

$$p(\theta) \cdot p(D|\theta) = p(D) \cdot p(\theta|D)$$

De modo que

$$p(\theta|D) = \frac{p(\theta) \cdot p(D|\theta)}{p(D)}$$

Esto se puede expresar como

$$\pi(\theta|D) = \frac{p(\theta) \cdot L(\theta|D)}{\int_{\theta} p(\theta) \cdot L(\theta|D) d\theta}$$

Donde $L(\theta|D)$ denota la verosimilitud y

$$P(D) = \int_{\theta} p(\theta) \cdot L(\theta|D) d\theta$$

es la constante normalizadora o la *distribución predictiva a priori*. De manera alternativa, como lo anterior no depende de θ , se puede expresar como:

$$\pi(\theta|D) \propto p(\theta) \cdot L(\theta|D)$$

es decir:

$$Probabilidad a posteriori \propto Probabilidad a priori \times Verosimilitud$$

3.2. El problema de inferencia

Se parte una muestra de datos, $\mathbf{x} = (x_1, \dots, x_n)$, de modo que la variable aleatoria que los genera se asume que depende de unos parámetros $\boldsymbol{\theta}$ y tiene como función de probabilidad f :

$$X|\boldsymbol{\theta} \sim f(\cdot)$$

Se trata de realizar inferencias sobre $\boldsymbol{\theta}$, mediante métodos clásicos o bayesianos.

La **inferencia clásica** presenta las siguientes características:

- El concepto de probabilidad está limitado a aquellos sucesos en los que se pueden definir frecuencias relativas.
- $\boldsymbol{\theta}$ es un valor fijo (pero desconocido).
- Se pueden usar estimadores de máxima verosimilitud (que se justifican asintóticamente) o se pueden usar estimadores insesgados.
- Se usa el concepto de *intervalo de confianza*
- El método de muestreo es muy importante.

La **inferencia bayesiana** se caracteriza por:

- Todos tenemos en nuestras propias probabilidades previas para cualquier suceso: $P(\text{cruz})$, $P(\text{lloverá mañana})$,...
- Nuestras probabilidades pueden ser diferentes porque son nuestras propias medidas de incertidumbre. La única restricción es que han de ser coherentes (que cumplan los axiomas de la probabilidad).
- $\boldsymbol{\theta}$ es un variable y tiene una distribución $f(\boldsymbol{\theta})$. Modificamos nuestras creencias sobre $\boldsymbol{\theta}$ usando el teorema de Bayes.
- La estimación es un problema de decisión. En situaciones distintas se eligen estimadores diferentes: se usa la *teoría de utilidad* para elegirlos.

- Un intervalo de credibilidad del 95 % para θ es un intervalo donde tenemos una probabilidad igual a 0,95 de que contenga θ .
- El método de muestreo no importa. Sólo los datos son importantes.

Comentarios sobre la Teoría Bayesiana

- *Crítica:* θ no está claro que tenga que ser siempre una variable.

Argumento: en la práctica, la distribución $f(\theta)$ muestra los conocimientos acerca de θ . El conocimiento cambia y evoluciona con la recogida de los datos.

- *Crítica:* Falta de objetividad. ¿Cómo se puede elegir la distribución a priori cuando no se tiene información?

Argumento: se pueden establecer métodos *objetivos* para calcular la distribución a priori. A menudo, un análisis clásico equivale a un análisis bayesiano con una distribución apriori no informativa.

- Los aspectos subjetivos son explícitos en un análisis bayesiano, mientras que en los métodos clásicos existen aunque no se muestran claramente.
- Es imprescindible hacer un análisis de sensibilidad. Si los resultados cambian cuando se cambia la distribución inicial, entonces la verosimilitud no da mucha información y la elección de la distribución inicial es fundamental.

3.3. Principios fundamentales en Inferencia

3.3.1. El Principio de Verosimilitud

Para hacer inferencia sobre θ , después de haber observado x , toda la información pertinente está contenida en la función de verosimilitud $l(\theta|x)$. Además, dos funciones de verosimilitud tienen la misma información sobre θ si son proporcionales entre sí.

Los métodos bayesianos cumplen el principio de verosimilitud: Si $l_1(\theta|x) \propto l_2(\theta|x)$ entonces, dada una distribución a priori $\pi(\theta)$,

$$f_1(\boldsymbol{\theta}|\mathbf{x}) \propto \pi(\boldsymbol{\theta})l_1(\boldsymbol{\theta}|\mathbf{x}) \propto \pi(\boldsymbol{\theta})l_2(\boldsymbol{\theta}|\mathbf{x}) \propto f_2(\boldsymbol{\theta}|\mathbf{x})$$

No obstante, se demuestra que los contrastes clásicos con significación a nivel fijo (por ejemplo $\alpha = 0,05$) y los intervalos de confianza no cumplen con este principio.

3.3.2. El Principio de Suficiencia

Un estadístico \mathbf{t} es suficiente para $\boldsymbol{\theta}$ si

$$l(\boldsymbol{\theta}|\mathbf{x}) = h(\mathbf{t}, \boldsymbol{\theta})g(\mathbf{x})$$

que recibe el nombre de teorema de *factorización de Neyman*.

El Principio de Suficiencia afirma que si existe un estadístico suficiente, \mathbf{t} , dadas dos muestras, $\mathbf{x}_1, \mathbf{x}_2$, que cumplen $\mathbf{t}(\mathbf{x}_1) = \mathbf{t}(\mathbf{x}_2)$, las conclusiones basadas en \mathbf{x}_1 y \mathbf{x}_2 deben ser iguales.

Todos los métodos estándar de inferencia estadística cumplen el principio de suficiencia.

3.3.3. El Principio de Condicionalidad

Suponiendo que se pueden hacer dos posibles experimentos E_1 y E_2 en relación a $\boldsymbol{\theta}$, y se elige uno con probabilidad 0,5, entonces la inferencia sobre $\boldsymbol{\theta}$ depende sólo del resultado del experimento seleccionado.

El principio de verosimilitud equivale al principio de suficiencia más el principio de condicionalidad.

3.4. Modelo de Inferencia para la distribución Normal

En este tema se estudian métodos de inferencia para el modelo de la distribución normal. La importancia de este modelo se refleja en la teoría de errores de mediciones, o en la aproximación, a través de la normal, de otras muchas distribuciones como resultado del Teorema Central del Límite.

NOTA: La función de densidad de una v.a. normal con parámetros μ y σ^2 ($X \sim N(\mu, \sigma^2)$)

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

donde $-\infty < x < +\infty$, siendo $E(X) = \mu$ y $Var(X) = \sigma^2$.

3.4.1. Modelo con varianza conocida

En el modelo básico se supone σ conocida, pero se desconoce μ y se desea hacer inferencias sobre μ .

Se tiene una serie de n observaciones y se formula el siguiente modelo:

$$X_1, X_2, \dots, X_n | \theta \sim N(\theta, \sigma_1^2)$$

(con σ_1 conocida) y por razones de flexibilidad y conveniencia matemática, se formula como distribución a priori de μ

$$\theta \sim N(\theta, \sigma_1^2)$$

Dados los datos x_1, x_2, \dots, x_n , se procede del modo habitual y se calcula la distribución a posteriori como:

$$\begin{aligned} \pi(\theta|x_1, x_2, \dots, x_n) &= \frac{f(x_1, x_2, \dots, x_n; \theta)}{f(x_1, x_2, \dots, x_n)} = \frac{f(x_1, x_2, \dots, x_n|\theta)\pi(\theta)}{f(x_1, x_2, \dots, x_n)} \\ f(x_1, x_2, \dots, x_n|\theta)\pi(\theta) &\propto \left(\prod_{i=1}^n \exp\left\{-\frac{(x_i - \theta)^2}{\sigma_1^2}\right\} \right) \cdot \exp\left\{-\frac{1}{2}\left(\frac{\theta - \mu_0}{\sigma_0}\right)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^n \left(\frac{x_i - \theta}{\sigma_1}\right)^2\right\} \cdot \exp\left\{-\frac{1}{2}\left(\frac{\theta - \mu}{\sigma_0}\right)^2\right\} \end{aligned}$$

$$\begin{aligned}
f(x_1, x_2, \dots, x_n | \theta) \pi(\theta) &\propto \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^n \frac{1}{\sigma_1^2} (x_i^2 - 2x_i\theta + \theta^2) + \frac{1}{\sigma_0^2} (\theta^2 - 2\theta\mu_0 + \mu_0^2) \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^n \frac{1}{\sigma_1^2} (-2x_i\theta + \theta^2) + \frac{1}{\sigma_0^2} (\theta^2 - 2\theta\mu_0) \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left(\theta^2 \left(\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2} \right) - 2\theta \left(\frac{\sum_{i=1}^n x_i}{\sigma_1^2} + \frac{\mu_0}{\sigma_0^2} \right) \right) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left(\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2} \right) \cdot \left(\theta^2 - 2\theta \left(\frac{\frac{n}{\sigma_1^2} \bar{x} + \frac{1}{\sigma_0^2} \mu_0}{\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}} \right) \right) \right\}
\end{aligned}$$

NOTA: Si una v.a. Z se distribuye $N(\mu, \sigma)$ entonces su función de densidad se puede escribir como:

$$f(z) \propto \exp \left(-\frac{1}{2} \cdot \frac{1}{s^2} (z - \mu)^2 \right) \propto \left(-\frac{1}{2} \cdot \frac{1}{s^2} (z^2 - 2z\mu) \right)$$

Por tanto, observando la expresión anterior, resulta que la distribución a posteriori es una normal con media y varianza:

$$\mu = \frac{\frac{n}{\sigma_1^2} \bar{x} + \frac{1}{\sigma_0^2} \mu_0}{\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}}$$

$$\sigma^2 = \frac{1}{\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}}$$

esto es $\theta | x \sim N(\mu, \sigma^2)$.

Como consecuencia de todo lo anterior se verifica lo siguiente:

Teorema 1: *La distribución normal es una distribución **conjugada** respecto de la verosimilitud normal, con media desconocida y varianza conocida.*

Distribución no informativa: Son aquellas que proponemos cuando sabemos el soporte de los parámetros a estimar. Estas no nos proporcionan información sobre las observaciones, generan un espectro más amplio para el muestreo de la distribución posterior.

Distribución informativa: Son aquellas que proponemos cuando sabemos el soporte de los parámetros a estimar. Las observaciones nos proporcionan información para proponer los hiper-parámetros para la o las distribuciones asociadas a el o los parámetros a estimar. Esto genera una disminución del espectro para el muestreo de la distribución posterior.

Error Cuadrático Medio: El error cuadrático medio es un estimador T , para estimar θ , se define como:

$$ECM_k = \widehat{V}(\widehat{\theta}_k) + (\widehat{S}_{\theta_k})^2$$

para $k = 1, 2$ donde $\theta_{j,k}$ es el estimador, basado en la j -ésima muestra correspondiente al modelo, del parámetro θ_k^* , el cual es el valor verdadero de los parámetros mencionados arriba y $\widehat{V}(\widehat{\theta}_k)$ es la varianza muestral de $\widehat{\theta}_k$, y el porcentaje de Cobertura es calculada como la proporción de las veces en que el intervalo de credibilidad del 95 % contiene a θ_k^* .

3.5. MCMC

Los métodos Markov Chain Monte Carlo (MCMC) son algoritmos que permiten obtener una muestra de una distribución de probabilidad f sin necesidad de simular directamente dicha distribución. Para ello estos métodos se basan en la construcción de una cadena de Markov ergódica cuya distribución estacionaria es precisamente f (2).

El principio general sobre el cual trabajan estos algoritmos es bastante sencillo. Dado un punto inicial arbitrario $x^{(0)}$ se construye una cadena de Markov ergódica $X_{t \in R}^{(t)}$ cuya distribución estacionaria es la distribución de interés f . Esto garantiza que, para un valor $l \in N$ suficientemente grande, se cumple que $X^{(l)}, X^{(l+1)}, X^{(l+2)}, \dots \sim f$. Obviamente $X^{(l)}$ y $X^{(l+1)}$ no son independientes; sin embargo, para un cierto valor $m \in N$ se puede considerar que $X^{(l)}$ y $X^{(l+m)}$ son aproximadamente independientes. Por lo tanto, si se simula dicha cadena y se toma $Z^{(t)} = X^{(l+mt)}$ entonces se obtiene una muestra aproximadamente independiente de f .

Como la cadena es ergódica, desde el punto de vista teórico el valor inicial $x^{(0)}$ no tiene mayor relevancia. En la práctica cualquier punto inicial dentro del soporte de la distribución

haría que la cadena converja; sin embargo, para ciertos valores iniciales el tiempo de convergencia podría crecer demasiado.

En la terminología de los métodos MCMC, l recibe el nombre de *long run* o *burn in* e indica el tiempo que se necesita para que la cadena converja a la distribución estacionaria. Por otro lado, m se suele llamar *thinning* y determina cada cuántas simulaciones se debe hacer el submuestreo. La calidad de las simulaciones depende fuertemente de estas dos cantidades. Contrario a lo que se esperaría, suele ser más complicado determinar los valores de l y m que construir una cadena de Markov con las propiedades deseadas.

Una de las aplicaciones más comunes de estos algoritmos es el cálculo numérico de integrales multidimensionales. Por esta razón suelen ser de gran utilidad en áreas como la Física y la Estadística. Algunos de los algoritmos MCMC más comunes son: Metropolis-Hastings, Gibbs sampler y Slice sampler.

3.5.1. Muestreo de Gibbs

En la estadística, el muestreo de Gibbs o una muestra de Gibbs es un algoritmo de la cadena de Markov Monte Carlo (MCMC) que se usa para obtener una secuencia de observaciones que se aproximan a partir de una distribución de probabilidad multivariable especificada, cuando el muestreo directo es difícil. Esta secuencia se puede usar para aproximar la distribución conjunta (por ejemplo, para generar un histograma de la distribución); aproximar la distribución marginal de una de las variables, o algún subconjunto de las variables (por ejemplo, los parámetros desconocidos o las variables latentes); o para calcular una integral (como el valor esperado de una de las variables). Por lo general, algunas de las variables corresponden a observaciones cuyos valores son conocidos y, por lo tanto, no necesitan ser muestreados.

El muestreo de Gibbs se usa comúnmente como un medio de inferencia estadística, especialmente la inferencia bayesiana. Es un algoritmo aleatorizado (es decir, un algoritmo que utiliza números aleatorios) y es una alternativa a los algoritmos determinísticos para la inferencia estadística, como el algoritmo de maximización de las expectativas (EM).

Al igual que con otros algoritmos de MCMC, el muestreo de Gibbs genera una cadena de muestras de Markov, cada una de las cuales se correlaciona con muestras cercanas. Como resultado, se debe tener cuidado si se desean muestras independientes. En general, las muestras desde el comienzo de la cadena (el período de quemado) pueden no representar con precisión la distribución deseada y generalmente se descartan. Si es necesario, un posible remedio es reducir la cadena de muestras resultante (es decir, solo tomando cada valor n , por ejemplo, cada 10mo valor). Sin embargo, se ha demostrado que se usa una cadena más larga (por ejemplo, una cadena que es n veces la longitud de la cadena considerada inicialmente con un factor de reducción de n) conduce a mejores estimaciones de la verdadera posterior. Por lo tanto, el aclareo solo debe aplicarse cuando el tiempo o la memoria de la computadora están restringidos (10).

3.5.2. Algoritmo M-H

El método de la cadena de Markov Monte Carlo (MCMC) para obtener una secuencia de muestras aleatorias de una distribución de probabilidad para la cual el muestreo directo es difícil. Esta secuencia se puede usar para aproximar la distribución (p. Ej., Para generar un histograma) o para calcular una integral (como un valor esperado)) Metropolis-Hastings y otros algoritmos de MCMC se utilizan generalmente para tomar muestras de distribuciones multidimensionales, especialmente cuando el número de dimensiones es alto. Para distribuciones unidimensionales, generalmente hay otros métodos disponibles (por ejemplo, muestreo de rechazo adaptativo) que pueden devolver directamente muestras independientes de la distribución, y están libres del problema de muestras autocorrelacionadas que es inherente a los métodos MCMC (11).

Definición: Un método MCMC, para la simulación de una distribución $\pi(\cdot)$, es cualquier método que produce una Cadena de Markov ergódica $(X(t))$ cuya distribución estacionaria es $\pi(\cdot)$, y para la cual existe un algoritmo de simulación.

El principio de los algoritmos MCMC es que dado un valor inicial $x(0)$, la Cadena $(X(t))$ es generada utilizando un kernel de transición con distribución estacionaria, $\pi(\cdot)$, la cual garantiza la convergencia en distribución de $(X(t))$ a $\pi(\cdot)$. Dado que la Cadena es ergódica,

el valor inicial $x^{(0)}$ en principio no es importante.

El algoritmo Metropolis-Hastings comienza muestreando de una densidad de candidatos y una distribución objetivo, pero como se está considerando Cadenas de Markov, la densidad depende del estado actual del proceso, (13). Denotando a la densidad de candidatos por $q(\phi|\theta)$ y suponiendo que $\sum_{\phi} q(\phi|\theta) = 1$. Si resulta que la densidad $q(y|x)$ siempre es la misma, entonces se necesita otra opción. Más sin embargo, si se encuentra que:

$$\pi(\theta)q(\phi|\theta) > \pi(\phi)q(\theta|\phi)$$

entonces parece que el proceso se mueve de θ a ϕ , y de ϕ a θ . Se puede reducir el número de movimientos de θ a ϕ introduciendo una probabilidad $\alpha(\phi|\theta)$, llamada probabilidad de aceptación. Con el fin de alcanzar el tiempo de reversibilidad, se toma $\alpha(\phi|\theta)$ tal que la ecuación

$$\pi(\theta)q(\phi|\theta) = \pi(\phi)q(\theta|\phi)$$

se asegura, y consecuentemente,

$$\alpha(\phi|\theta) = \frac{\pi(\theta)q(\phi|\theta)}{\pi(\phi)q(\theta|\phi)}$$

No se quiere reducir el número de movimientos de ϕ a θ en tal caso, así, se toma $\alpha(\theta|\phi) = 1$, y similarmente $\alpha(\phi|\theta) = 1$ en el caso donde la igualdad se invierte, entonces se tiene

$$\pi(\theta)q(\phi|\theta) < \pi(\phi)q(\theta|\phi)$$

Es claro que una fórmula general es

$$\alpha(\phi|\theta) = \min \left\{ \frac{\pi(\theta)q(\phi|\theta)}{\pi(\phi)q(\theta|\phi)}, 1 \right\}$$

así, la probabilidad de ir de un estado θ a un estado ϕ es $p^*(\phi|\theta) = q(\phi|\theta)\alpha(\phi|\theta)$, mientras que la probabilidad de que la Cadena permanezca en el estado θ es

$$r(\theta) = 1 - \sum_{\phi} q(\phi|\theta)\alpha(\phi|\theta)$$

La matriz de transición de probabilidades es dada por

$$p(\phi|\theta) = p^*(\phi|\theta) + r(\theta)\delta(\phi|\theta) = q(\phi|\theta) + (1 - \sum_{\phi} q(\phi|\theta)\alpha(\phi|\theta))\delta(\phi|\theta)$$

Teniendo en cuenta que es necesario conocer la densidad objetivo $\phi(\theta)$ hasta un múltiplo constante, porque ésta aparece en el numerador y denominador de la expresión para $\alpha(\phi|\theta)$. Además, si la densidad generadora de candidatos $q(\phi|\theta)$ es simétrica, se tiene que $q(\phi|\theta) = q(\theta|\phi)$, y $\alpha(\phi|\theta)$ se reduce a

$$\alpha(\phi|\theta) = \min \left\{ \frac{\pi(\phi)}{\pi(\theta)}, 1 \right\}$$

Se puede resumir el algoritmo Metropolis-Hasting como sigue:

1. Muestrear un candidato θ^* de una distribución propuesta $q(\theta^*|\theta^{(t-1)})$.
2. Calcular:

$$\alpha = \min \left\{ \frac{p(\theta^*)q(\theta^{(t-1)}|\theta^*)}{p(\theta^{(t-1)})q(\theta^*|\theta^{(t-1)})}, 1 \right\}.$$

3. Generar un valor $U \sim \mathcal{U}(0, 1)$.
4. Si $U \leq \alpha$, se define $\theta^{(t)} = \theta^*$; en otro caso se define $\theta^{(t)} = \theta^{(t-1)}$.
5. Se devuelve la secuencia $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n)}\}$.

Más tarde, en Geman y Geman (1984), se presenta un método de simulación, que también genera una Cadena de Markov y que después pasa a ser conocido en la literatura como muestreador de Gibbs. El algoritmo de Metropolis-Hastings y el muestreador de Gibbs forman los dos esquemas básicos de la metodología MCMC a partir de los cuales se han creado otros con fines más específicos y distintas propiedades.

3.5.3. Algoritmo T-Walk

Desarrollamos una nueva muestra MCMC de propósito general para distribuciones arbitrarias continuas que no requieren ajuste. Llamamos a este MCMC el T-walk. El T-walk mantiene dos puntos independientes en el espacio de muestra, y todos los movimientos se basan en propuestas que luego se aceptan con una probabilidad de aceptación estándar de Metropolis-Hastings en el espacio del producto. Por lo tanto, la t-walk es probablemente convergente bajo los requisitos suaves habituales. Restringimos las distribuciones de propuestas, o "movimientos", a aquellas que producen un algoritmo que es invariable a escala, y aproximadamente invariante para afinar las transformaciones del espacio de estado. Por lo tanto, la ampliación de las propuestas, y efectivamente también la transformación de coordenadas, que podrían ser utilizados para aumentar la eficiencia de la toma de muestras, no son necesarios ya que el t-Walk es idéntico en cualquier versión a escala de la distribución de destino. Se dan cuatro movimientos que dan como resultado un algoritmo de muestreo efectivo (9).

Usamos el dispositivo simple de actualizar solo un subconjunto de coordenadas al azar en cada paso para permitir la aplicación del T-Walk a problemas de alta dimensión. En una serie de problemas de prueba a través de dimensiones nos encontramos con que el T-Walk es solamente un pequeño factor menos eficiente que los algoritmos afinados de manera óptima, pero supera significativamente del paseo aleatorio muestreadores generales M-H que no están sintonizados para problemas específicos. Además, la T-Walk sigue siendo efectiva para las distribuciones objetivo para las que no existe una transformación afín óptima, como aquellas en las que la estructura de correlación es muy diferente en diferentes regiones del espacio de estados (9).

Para una función objetivo (distribución posterior) $\pi(x)$, $x \in \mathcal{X}$ (\mathcal{X} es de dimensión n y es un subconjunto de \mathbb{R}^n), donde creamos la nueva función objetivo $f(x, x') = \pi(x)\pi(x')$ en el espacio producto correspondiente $\mathcal{X} \times \mathcal{X}$. Mientras que una propuesta general tiene la forma(9):

$$q\{(y, y')|(x, x')\}.$$

consideramos las dos propuestas restringidas, con igual probabilidad:

$$(y, y') = \begin{cases} (x, h(x', x)) \\ (h(x, x'), x') \end{cases} \quad (3.1)$$

Donde $h(x, x')$ es una variable aleatoria utilizada para formar la propuesta. Es decir, cambiamos solo x por x' en cada paso. Sin embargo, tenemos en cuenta que no estamos considerando dos cadenas paralelas independientes en cada \mathcal{X} ; en cambio, todo el proceso se encuentra en $\mathcal{X} \times \mathcal{X}$. Seleccionaremos al azar entre cuatro propuestas diferentes, que se definirán a continuación, cada una caracterizada por una función particular $h(\cdot, \cdot)$. Primero elegiremos una opción en (3.1) y segundo, crear la propuesta (y, y') simulando de la función h correspondiente(9).

Dentro de un esquema de Metropolis-Hastings, necesitamos calcular la tasa de aceptación correspondiente(9). Denotando la función de densidad de $h(x, x')$ por $g(\cdot | (x, x'))$, esta relación es igual a:

$$\frac{\pi(y') g(x' | y', x)}{\pi(x') g(y' | x', x)}.$$

Para el primer caso en las ecuaciones (3.1) y :

$$\frac{\pi(y) g(x | y, x')}{\pi(x) g(y | x, x')}.$$

Para el segundo caso tenga en cuenta que la restricción a la propuesta (3.1) implica que solo un único se requiere evaluación de la densidad objetivo, en cualquier caso.

Es sencillo mostrar que si la variable aleatoria h es invariante para las transformaciones afines, es decir, $h(\phi x, \phi x')$ para cualquier transformación afín ϕ , entonces también lo son las propuestas (3.1) y la muestra de MCMC resultante. El diseño de un algoritmo de muestreo invariable descansa en la pregunta de si es posible encontrar una o más variables aleatorias h que den un algoritmo de muestreo efectivo. Hemos encontrado que las cuatro opciones para h , dadas a continuación, dan una mezcla adecuada a través de un amplio rango de distribuciones objetivo de dimensión moderada(9).

En muchas aplicaciones, particularmente con correlaciones débiles, encontramos que la mezcla de la cadena se logra principalmente mediante una caminata aleatoria escalada a la que nos referimos como el movimiento de T-Walk (9).

El movimiento de T-Walk, está definida por la función:

$$h_w(x, x')_j = \begin{cases} (x_j + (x_j - x'_j)), & I_j = 1 \\ x_j & I_j = 0. \end{cases}$$

3.6. Diagnósticos de convergencia

Suponiendo que se desea generar una muestra de tamaño N de la distribución $\pi(\boldsymbol{\theta}|\mathbf{y})$. Si para cada uno de N valores iniciales $\boldsymbol{\theta}_1^{(0)}, \dots, \boldsymbol{\theta}_N^{(0)}$ se realiza alguno de los algoritmos discutidos en esta sección, entonces, de un cierto número de iteraciones T suficientemente grande, los valores de $\boldsymbol{\theta}_T^{(0)}, \dots, \boldsymbol{\theta}_N^{(T)}$ pueden considerarse como una muestra de tamaño N de la distribución posterior de $\boldsymbol{\theta}$. Alternativamente se puede generar una sola Cadena y tomar los valores $\boldsymbol{\theta}^{T+K}, \dots, \boldsymbol{\theta}^{T+NK}$ como una muestra de $\pi(\boldsymbol{\theta}|\mathbf{y})$, donde K se elige de manera que la correlación entre las observaciones sea pequeña.

Un método MCMC crea una muestra de la distribución posterior, y por lo general se desea saber si la muestra es suficientemente cercana a la distribución $\pi(\boldsymbol{\theta}|\mathbf{y})$ para ser usada en el análisis de los datos. Existen varias pruebas para verificar la convergencia de la Cadena de Markov, ambas son visuales y estadísticas. Alguna de estas pruebas de inspección visual es la siguiente:

Traza de la Cadena.

La traza es una gráfica del número de iteración contra el valor de la estimación del parámetro en cada iteración. Se puede ver si la Cadena se detiene en determinadas zonas del espacio de parámetros, lo que indica una mala convergencia. Algunas de las pruebas estadísticas para evaluar la convergencia se mencionan a continuación:

- 1 Correr $m \geq 2$ Cadenas de longitud $2N$ de valores iniciales sobredispersados.
- 2 Descartar las primeras N muestras en cada Cadena.

3 Calcular la varianza dentro y entre la cadena.

I Varianza dentro de la Cadena:

$$W = \frac{1}{m} \sum_{j=1}^m s_j^2$$

donde $s_j^2 = \frac{1}{N-1} \sum_{i=1}^N (\theta_{ij} - \bar{\theta}_j)^2$ es simplemente la fórmula para la varianza de la j -ésima Cadena. W es la media de las varianzas de cada cadena.

II varianza entre la Cadena:

$$B = \frac{N}{m-1} \sum_{j=1}^m (\bar{\theta}_j - \bar{\bar{\theta}})^2$$

donde $\bar{\bar{\theta}} = \frac{1}{m} \sum_{j=1}^m \bar{\theta}_j$. Esto es la varianza de las medias de la Cadena multiplicada por N debido a que cada Cadena está basada en N muestras.

4 Calcular la varianza estimada del parámetro como una suma ponderada de la varianza dentro y entre la Cadena.

$$\widehat{Var}(\theta) = \left(1 - \frac{1}{N}\right) W + \frac{B}{N}$$

4 Calcular el factor potencial de reducción de escala. En Gelman et al. (2014), se puede ver que dicho factor es calculado como:

$$\widehat{R} = \sqrt{\frac{\widehat{Var}(\theta)}{W}}$$

donde \widehat{R} , se considera alto si es mayor que 1,1 ó 1,2, entonces se debería correr las Cadenas más tiempo para mejorar la convergencia a la distribución estacionaria. Si se tiene más de un parámetro, entonces se necesita calcular dicho factor para cada parámetro.

Diagnóstico de Geweke.

Este diagnóstico toma dos partes (usualmente el 10 % de la primer mitad, y la segunda mitad) de la Cadena de Markov. Suponiendo que la segunda mitad de la Cadena ha convergido a la distribución estacionaria. Se realiza una prueba de comparación de medias de ambas partes. Si la media del primer 10 % de las muestras no es significativamente diferente del último 50 %, entonces se concluye que estas dos partes provienen de la misma distribución, y por consiguiente la Cadena de Markov ha convergido. El estadístico de prueba es un Z-score estándar con errores estándar ajustados por autocorrelación, (Geweke et al., 1991).

Diagnóstico de Heidelberg y Welch.

Este diagnóstico calcula un estadístico de prueba (basado en el estadístico de prueba de Mises Cramer-von) para aceptar o rechazar la hipótesis nula de que la Cadena de Markov es de una distribución estacionaria. El diagnóstico consiste de dos partes:

1. Primera Parte

- a) Generar una Cadena de N iteraciones y definir un nivel de significación α .
- b) Calcular el estadístico de prueba en toda la Cadena. Aceptar o rechazar la hipótesis nula de que la Cadena es de una distribución estacionaria.
- c) Si la hipótesis nula es rechazada, descartar el primer 10 % de la Cadena. Calcular el estadístico de prueba y aceptar o rechazar la hipótesis nula.
- d) Si la hipótesis nula es rechazada, descartar el próximo 10 % de la Cadena y calcular el estadístico de prueba.
- e) Repetir hasta que la hipótesis nula sea aceptada o el 50 % de la Cadena sea descartada. Si el estadístico todavía rechaza la hipótesis nula, entonces la Cadena falla y se necesita un número mayor de iteraciones.

2. Segunda Parte

Si la Cadena pasa la primera parte del diagnóstico, entonces ésta toma la parte de la Cadena no descartada de la primera parte para probar la segunda parte. La prueba half-width calcula un intervalo de confianza del 95 % para la media, utilizando la porción de la Cadena que pasó la prueba de estacionariedad (primera

parte). La mitad del ancho de este intervalo es comparado con la estimación de la media. Si la razón entre la mitad del ancho y la media es inferior a algún $\epsilon > 0$, entonces la prueba half-width es aceptada. De lo contrario, la longitud de la muestra no es lo suficientemente grande para estimar la media con suficiente precisión.

3.7. Metodología Bayesiana.

La ecuación de observación (2):

$$\mathbf{y}_i = h(\mathbf{x}(t_i)) + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, \quad (3.2)$$

Donde \mathbf{y}_i corresponde a la i -ésimo valor observado en condición de incertidumbre para el sistema (1) en un tiempo discreto $t_i \in [0, T]$, $i = 1, 2, \dots, n$; h es la función de observación y $\boldsymbol{\varepsilon}_i$ son los errores de medición que se suponen independientes idénticamente distribuidos (i.i.d.) de forma normal-bivariada con media cero y varianza Σ , lo denotamos:

$$\boldsymbol{\varepsilon}_i \sim N_2(0, \Sigma); \Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

Sea $\mathbf{y}_i = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$ la función de verosimilitud esta dada por (3):

$$L(\mathbf{y}|\boldsymbol{\theta}) = |\Sigma|^{-1} (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{x}_i^*(t))^T \Sigma^{-1} (\mathbf{y}_i - \mathbf{x}_i^*(t)) \right\} \quad (3.3)$$

Note que la expresión anterior involucra el calculo de $\mathbf{x}^*(t) = (x_1(t), x_2(t))^t$; es decir, la solución del sistema(2.2). Sin embargo, para este modelo, no existe una expresión analítica de la solución. En este trabajo se utilizará un método de solución numérica para resolverlo (1).

Suponiendo independencia en los parámetros, la distribución a priori conjunta puede ser expresada como:

$$p(\boldsymbol{\theta}) = p(a, b, c, d) = p(a)p(b)p(c)p(d).$$

En este trabajo, proponemos distribuciones Gammas como distribuciones a priori para los parámetros. De esta forma, mediante el Teoremas de Bayes, la distribución posterior de los parámetros queda expresada como:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{L(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int_{\Theta} L(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}},$$

$$p(\boldsymbol{\theta}|y) \propto L(\boldsymbol{\theta}; y)p(\boldsymbol{\theta})$$

$$\propto |\Sigma|^{-1} (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{x}_i^*(t))^T \Sigma^{-1} (\mathbf{y}_i - \mathbf{x}_i^*(t)) \right\} \times p(a)p(b)p(c)p(d)$$

donde x^* denota la solución aproximada del modelo (1), obtenida a partir del Método de Runge-Kutta

Usando la función de pérdida cuadrática, el estimador puntual Bayesiano es la media posteriori, y esta dada por:

$$\hat{\theta}_B = E(\theta|\mathbf{y}).$$

Note que la distribución posterior conjunta es analíticamente intratable y las distribuciones marginales posteriores de los los parámetros son complicados. Sin embargo, podemos obtener muestras usando el método de Monte Carlo mediante la técnica de Cadenas de Markov (MCMC). Los métodos MCMC más comunes son el muestreo por Gibbs (5) , (6) y el Metropolis-Hastings (7).

3.8. Estudio de Simulación

Se lleva acabo un estudio de simulación para evaluar el desempeño de los estimadores Bayesianos de los parámetros de interes($\theta = (a, b, c, d)$), los valores de los parámetros fueron $a = 0,4$, $b = 0,02$, $c = 0,8$, $d = 0,025$ y varianza conocida de $\sigma^2 = 10^{-3}$. Para este trabajo, se realizan simulaciones de: 50, 100, 200, 300, 400 y 500 simulaciones, cada simulación tiene 10000 iteraciones, los resultados fueron los siguientes:

Simulaciones n=50					
Parámetros	Sesgo	ECM	Cobertura	I.C.I.	I.C.S.
$a = 0,4$	-0,001327	$9,32 * 10^{-5}$	100 %	0,3768	0,4176
$b = 0,02$	-0,000023	$2,97 * 10^{-7}$	100 %	0,0186	0,0212
$c = 0,8$	0,004354	$3,83 * 10^{-4}$	100 %	0,7641	0,8561
$d = 0,025$	0,000155	$4,03 * 10^{-7}$	100 %	0,0238	0,0267

Cuadro 3.1: *Sesgo estimado, Error Cuadratico Medio(EMC), porcentaje de Cobertura, Intervalos de Credibilidad para los parámetros del modelo Presa-Depredador.*

Simulaciones n=100					
Parámetros	Sesgo	ECM	Cobertura	I.C.I.	I.C.S.
$a = 0,4$	-0,001377	$1,01 * 10^{-4}$	100 %	0,3655	0,4227
$b = 0,02$	-0,000035	$2,98 * 10^{-7}$	100 %	0,0182	0,0212
$c = 0,8$	0,003481	$4,03 * 10^{-4}$	100 %	0,7573	0,8713
$d = 0,025$	0,000111	$4,03 * 10^{-7}$	100 %	0,0236	0,0274

Cuadro 3.2: *Sesgo estimado, Error Cuadratico Medio(EMC), porcentaje de Cobertura, Intervalos de Credibilidad para los parámetros del modelo Presa-Depredador.*

Simulaciones n=200					
Parámetros	Sesgo	ECM	Cobertura	I.C.I.	I.C.S.
$a = 0,4$	-0,000464	$1,02 * 10^{-4}$	100 %	0,3543	0,4292
$b = 0,02$	$-3,89 * 10^{-7}$	$3,05 * 10^{-7}$	100 %	0,0177	0,0214
$c = 0,8$	0,001675	$4,71 * 10^{-4}$	100 %	0,7527	0,9128
$d = 0,025$	$6,23 * 10^{-5}$	$4,54 * 10^{-7}$	100 %	0,0234	0,0283

Cuadro 3.3: *Sesgo estimado, Error Cuadratico Medio(EMC), porcentaje de Cobertura, Intervalos de Credibilidad para los parámetros del modelo Presa-Depredador.*

Simulaciones n=300					
Parámetros	Sesgo	ECM	Cobertura	I.C.I.	I.C.S.
$a = 0,4$	-0,000483	$9,29 * 10^{-5}$	100 %	0,3507	0,4360
$b = 0,02$	$-1,57 * 10^{-5}$	$2,75 * 10^{-7}$	100 %	0,0171	0,0217
$c = 0,8$	0,001935	$3,95 * 10^{-4}$	100 %	0,7337	0,9034
$d = 0,025$	$6,69 * 10^{-5}$	$3,98 * 10^{-7}$	100 %	0,0230	0,0287

Cuadro 3.4: *Sesgo estimado, Error Cuadratico Medio(EMC), porcentaje de Cobertura, Intervalos de Credibilidad para los parámetros del modelo Presa-Depredador.*

Simulaciones n=400					
Parámetros	Sesgo	ECM	Cobertura	I.C.I.	I.C.S.
$a = 0,4$	-0,000358	$9,89 * 10^{-5}$	100 %	0,3616	0,4283
$b = 0,02$	$4,88 * 10^{-6}$	$2,92 * 10^{-7}$	100 %	0,0178	0,0217
$c = 0,8$	0,00195113	$4,14 * 10^{-4}$	100 %	0,7465	0,8925
$d = 0,025$	$5,94 * 10^{-5}$	$4,00 * 10^{-7}$	100 %	0,0233	0,0278

Cuadro 3.5: *Sesgo estimado, Error Cuadratico Medio(EMC), porcentaje de Cobertura, Intervalos de Credibilidad para los parámetros del modelo Presa-Depredador.*

Simulaciones n=500					
Parámetros	Sesgo	ECM	Cobertura	I.C.I.	I.C.S.
$a = 0,4$	$-9,42 * 10^{-4}$	$9,25 * 10^{-5}$	100 %	0,3509	0,4351
$b = 0,02$	$-4,23 * 10^{-5}$	$2,99 * 10^{-7}$	100 %	0,0176	0,0219
$c = 0,8$	$2,91 * 10^{-3}$	$4,07 * 10^{-4}$	100 %	0,7404	0,8935
$d = 0,025$	$9,24 * 10^{-5}$	$3,99 * 10^{-7}$	100 %	0,0231	0,0289

Cuadro 3.6: *Sesgo estimado, Error Cuadratico Medio(EMC), porcentaje de Cobertura, Intervalos de Credibilidad para los parámetros del modelo Presa-Depredador.*

Para analizar la convergencia de las soluciones, realizaremos los diagnósticos de convergencias expuestos anteriormente en la sección 3.6. Los diagnósticos se trabajaran usando 3 cadenas de longitud 100000, con un quemado de 5000 y un adelgazamiento de 10.

1. Diagnóstico de la Traza:

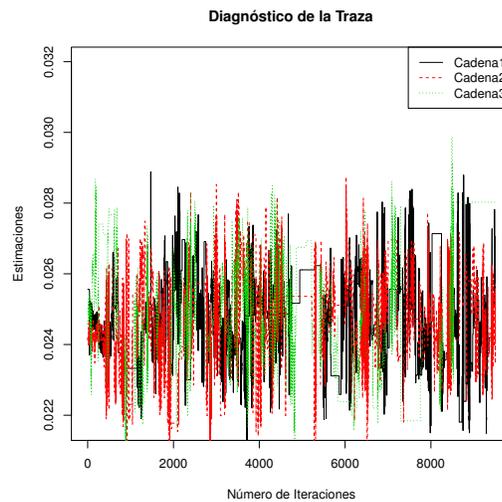


Figura 3.1: *Traza de las cadenas.*

Podemos observar que las 3 cadenas se superponen.

2. Diagnóstico de Gelman y Rubin.

Este Diagnóstico nos muestra lo siguiente:

Test de Gelman y Rubin.		
Parámetros	Punto Estimado	I.C.
a	1,07	1,07
b	1,06	1,06
c	1,08	1,10
d	1,08	1,08
Multivariante=1,07		height

Cuadro 3.7: *Diagnóstico de Gelman y Rubin*

Diagnóstico Gráfico de Gelman y Rubin:

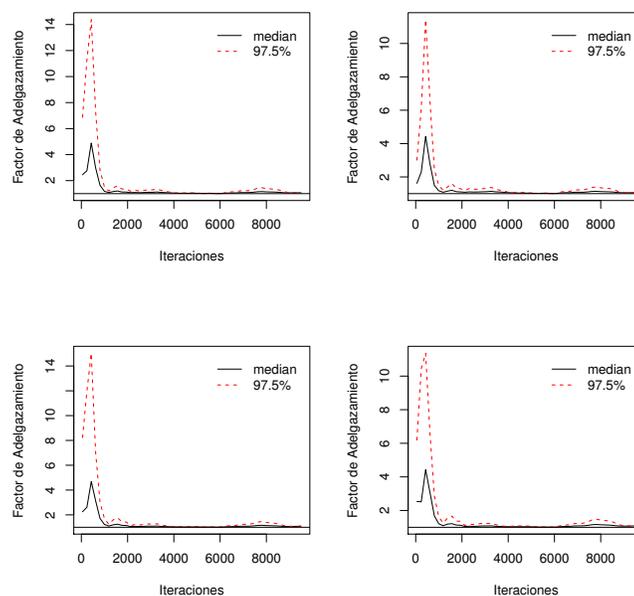


Figura 3.2: *Diagnóstico gráfico de Gelman y Rubin.*

3. Diagnóstico de Geweke.

Este Diagnóstico nos muestra lo siguiente:

Test de Geweke.			
Varianza 1	Varianza 2	Varianza 3	Varianza 4
0,7871	1,0047	-0,9379	-1,0086

Cuadro 3.8: *Diagnóstico de Geweke para la primera cadena*

Test de Geweke.			
Varianza 1	Varianza 2	Varianza 3	Varianza 4
1,491	1,355	-1,918	-1,978

Cuadro 3.9: *Diagnóstico de Geweke para la segunda cadena*

Test de Geweke.			
Varianza 1	Varianza 2	Varianza 3	Varianza 4
-0,8388	-0,8469	0,8135	0,7878

Cuadro 3.10: *Diagnóstico de Geweke para la tercera cadena*

Podemos también observar este diagnóstico gráficamente, entonces:

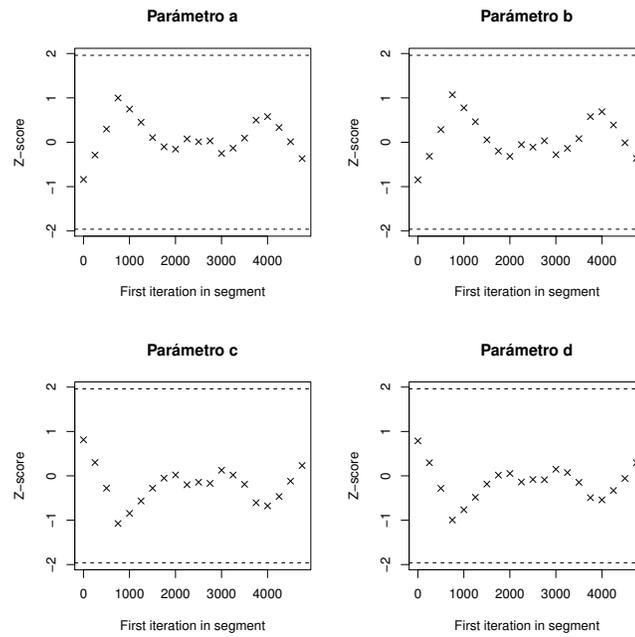


Figura 3.3: *Diagnóstico para la segunda Cadena.*

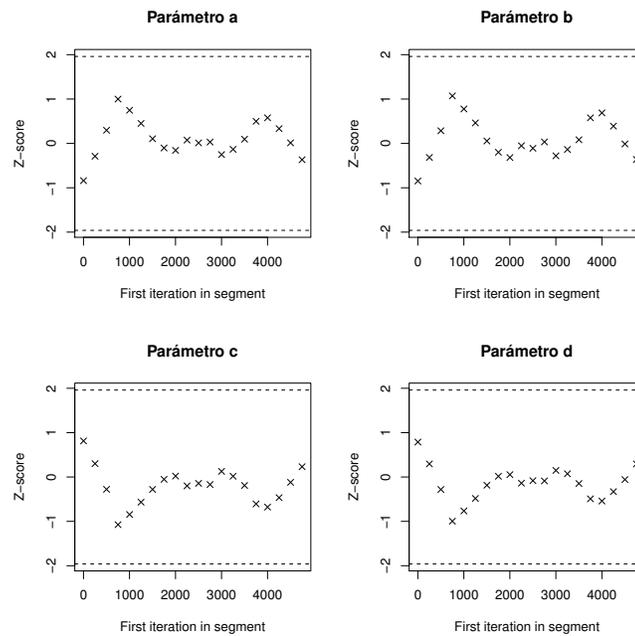


Figura 3.4: *Diagnóstico para la segunda Cadena.*

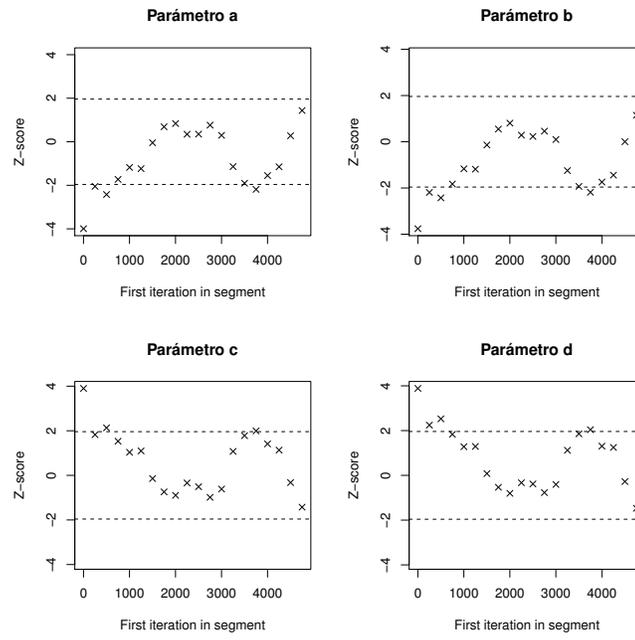


Figura 3.5: *Diagnóstico para la tercera Cadena.*

4. Diagnóstico de Heidelberg y Welch.

Este Diagnóstico nos muestra lo siguiente:

Test de Heidelberg y Welch.						
Parámetros	Prueba de Estacionariedad	Número de Iteraciones	p-valor	Prueba de Halfwidth	Media	Puntuación de Halfwidth
a	pasa	1	0,505	pasa	0,4040	0,004032
b	pasa	1	0,499	pasa	0,0202	0,000216
c	pasa	1	0,584	pasa	0,7921	0,008177
d	pasa	1	0,599	pasa	0,0247	0,000283

Cuadro 3.11: *Diagnóstico de Heidelberg y Welch para la primera cadena.*

Test de Heidelberg y Welch.						
Parámetros	Prueba de Estacionariedad	Número de Iteraciones	p-valor	Prueba de Halfwidth	Media	Puntuación de Halfwidth
a	pasa	1	0,276	pasa	0,4045	0,005358
b	pasa	1	0,216	pasa	0,0202	0,000292
c	pasa	1	0,213	pasa	0,7918	0,009417
d	pasa	1	0,224	pasa	0,0247	0,000325

Cuadro 3.12: *Diagnóstico de Heidelberg y Welch para la segunda cadena.*

Test de Heidelberg y Welch.						
Parámetros	Prueba de Estacionariedad	Número de Iteraciones	p-valor	Prueba de Halfwidth	Media	Puntuación de Halfwidth
a	pasa	1	0,928	pasa	0,4020	0,011793
b	pasa	1	0,921	pasa	0,0200	0,000587
c	pasa	1	0,919	pasa	0,7980	0,023339
d	pasa	1	0,939	pasa	0,0249	0,000752

Cuadro 3.13: *Diagnóstico de Heidelberg y Welch para la tercera cadena.*

3.9. Ejemplos

La problemática es estudiar la interacción entre el Lince Iberico(depredador) y conejo silvestre(presa). Nuestro problema será estimar los parametros del modelo Presa-Depredador. Se presentan los datos a continuación:

Conejo Silvestre: 30, 47.2, 70.2, 77.3 , 36.3, 20.6, 18.1,21.4, 22, 25.4, 27.1, 40.3, 57, 76.5 , 52.3, 19.5,11.2, 7.6, 14.6, 16.2, 24.7

Lince Iberico: 4, 6.1, 9.9, 35.2, 59.4, 41.7, 19, 13, 8.4,9.1, 7.4, 8, 12.3, 19.5, 45.7, 51.1, 29.7, 15.8, 9.6,10.1, 8.6

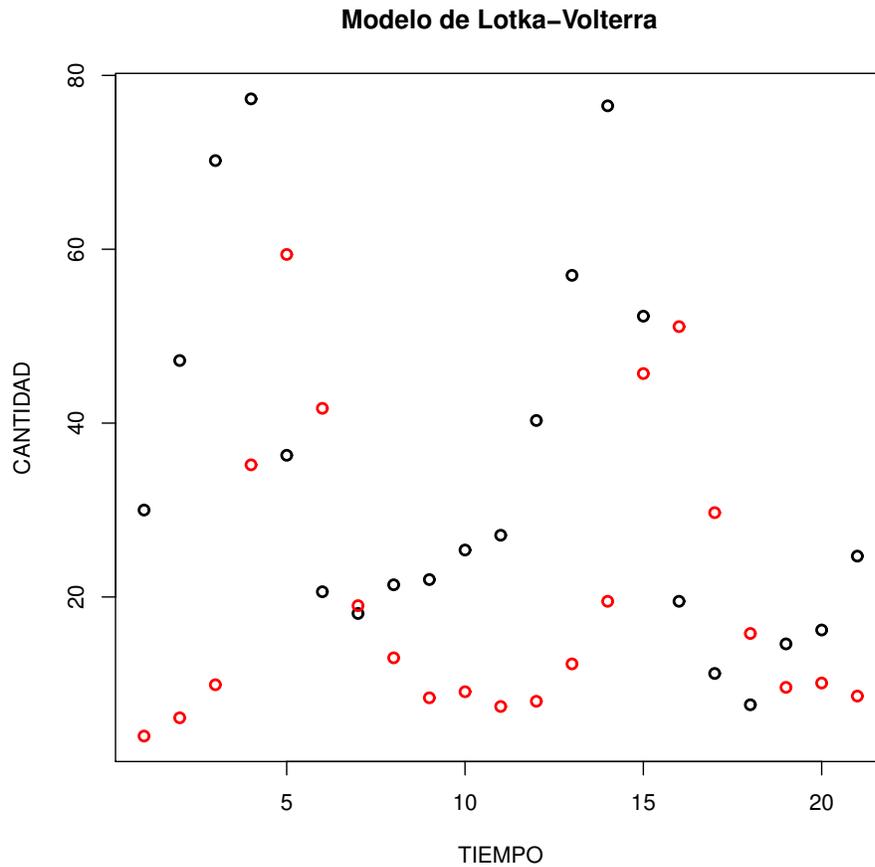


Figura 3.6: Diagrama de dispersión de los lince ibericos y los conejos silvestres.

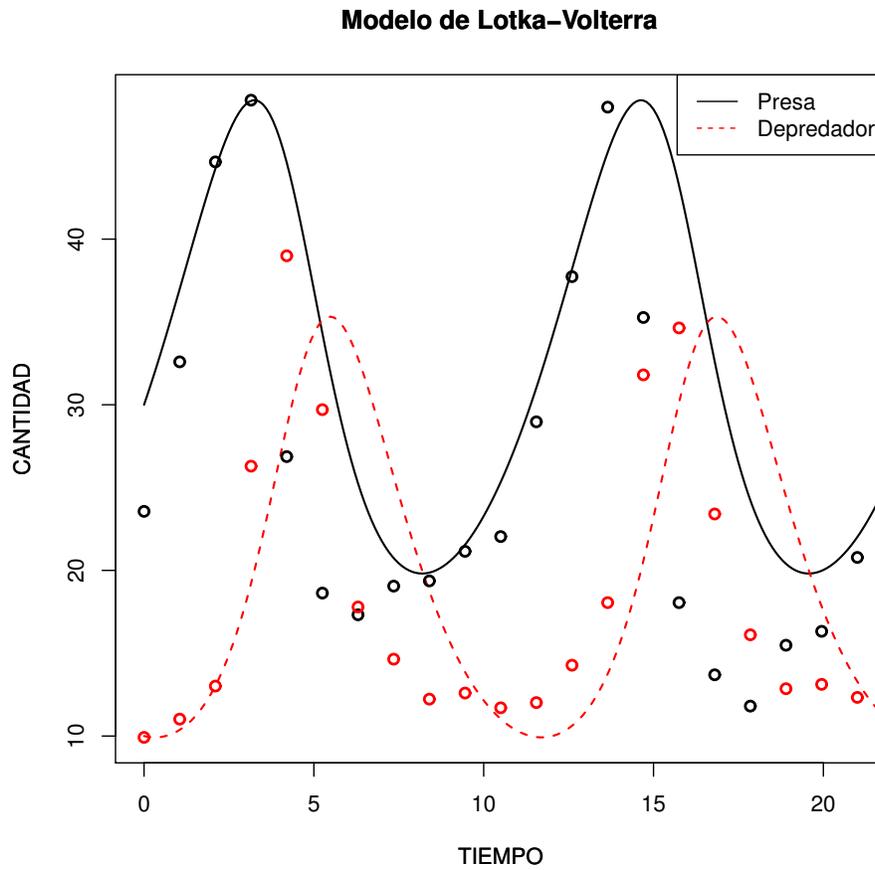


Figura 3.7: *Datos reales y curvas solución obtenidas de la estimación Bayesiana.*

	2.5 %	Media	97.5 %
a	0,07993463	0,21375303	0,45873386
b	0,01126015	0,02249032	0,04729470
c	0,44367405	0,80582582	1,22882474
d	0,02758566	0,03840149	0,05112532

Cuadro 3.14: *Intervalos de credibilidad para los estimadores de los datos reales.*

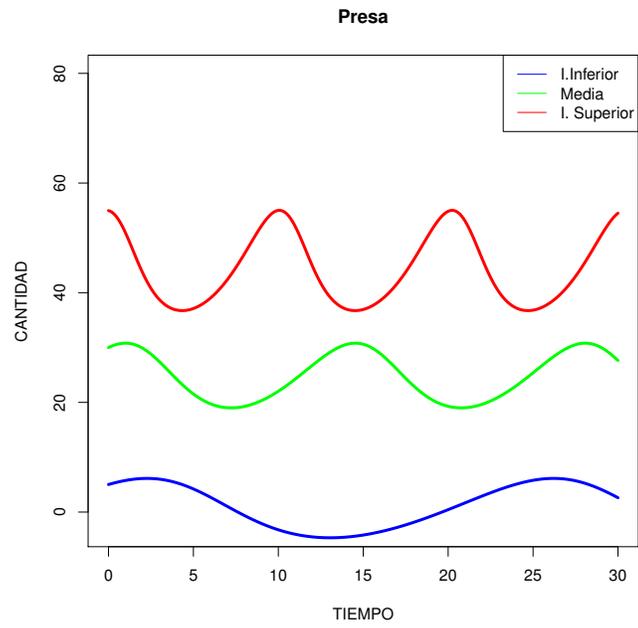


Figura 3.8: Gráfica de intervalos de credibilidad para la presa.

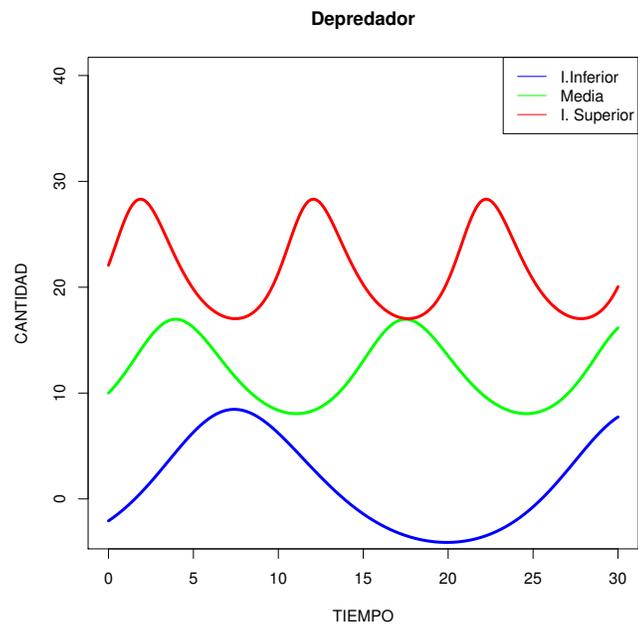


Figura 3.9: Gráfica de intervalos de credibilidad para el depredador.

Conclusiones

Al observar el estudio de las simulaciones, las estimaciones de los parámetros fueron excelentes, sus error cuadrático medio y sesgo son numéricamente pequeños, que es lo que se esperaba obtener a un 100 % de cobertura.

El estudio de convergencia nos muestra que al utilizar 4 diagnósticos distintos, pasa los 4 diagnósticos sin mayor dificultad, esto hace que nuestras estimaciones sean las correctas.

Esta nueva propuesta de estimación de parámetros, es excelente al momento de ser ejecutada, ya que los datos reales nos dicen lo que realmente esta sucediendo entre la Presa y el Depredador, está incertidumbre ya la estamos cuantificando y podemos realizar proyecciones más efectivas para el modelamiento de distintas especies en interacción.

Con respecto al ejemplo desarrollado, los resultados obtenidos en las simulaciones, la convergencia de las soluciones de nuestro ejemplo de la vida cotidiana es acertado y muy conveniente a otros métodos utilizados anteriormente, ya que los datos nos dicen verazmente lo que realmente esta sucediendo en el modelo Lotka-Volterra.

Los métodos MCMC son útiles para solucionar problemas inversos de sistemas ecuaciones diferenciales no lineales, hemos obtenido excelentes resultados, sobre todo cuando se utilizan modelos muy complejos en donde la distribución posterior para las cantidades de interés no tienen forma de las que ya conocemos. El uso de la metodología Bayesiana nos proporciona herramientas mas acertivas para ajustar modelos de datos reales aún cuando los datos se comporten de forma anómala.

El trabajo se enfocó en poder modelar un sistema de ecuaciones diferenciales no lineales

(Presa-Depredador), pero ya podemos concluir que esta herramienta nos permitirá modelar cualquier tipo de sistema de ecuaciones lineales o no lineales, en los cuales tengamos datos reales a utilizar.

El trabajo futuro a realizar, es poder ajustar los mismos datos reales a un sistema de ecuaciones no lineal de orden fraccionario, para esto, debemos enfocarnos en la teoría de ecuaciones diferenciales de orden fraccional y el análisis numérico para resolver estos problemas.

Referencias

- [1] Hirsch, Smale, Devaney, Morris W. Hirsch, Stephen Smale, Robert L. Devaney Differential Equations, Dynamical Systems An introduction to Chaos, Second edition. Elsevier academic press 2004.
- [2] Ming-Hui, Qi-Man, G Ibrahim, 2012, Ming-Hui Chen, Qi-Man Shao, and Joseph G Ibrahim. Monte Carlo methods in Bayesian computation. Springer Science Business Media, 2012
- [3] Casella, Berger, 2002, George Casella, Roger L. Berger, Statistical, Second Edition Duxbury advanced series 2002.
- [4] Capistrán, Donne, 2013, A. Capistrán, J. Andrés Christen, Sophie Donne, Bayesian Analysis of ODE's: solver optimal accuracy and Bayes factors. arXiv:1311.2281v1 [stat.CO] 10 Nov 2013.
- [5] Casella, I. George, 1992, Casella and E. I. George. Explaining the Gibbs Sampler. The American Statistician, 46(3):167-174, 1992.
- [6] Chen, Lee, 1979, H.H. Chen, Y.C. Lee, Internal wave solitons of fluids with finite depth, Phys. Rev. Lett. 43 (1979) 264-266.
- [7] Metropolis, Rosenbluth, Rosenbluth, Teller, Teller, 1953, Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. The Journal of Chemical Physics, 21:1087, 1953.
- [8] F. Ariza, J. Sanchez, M. Arciga, L. Vivas., 2016, Bayesian analysis for a fractional population growth model. Journal of Applied Mathematics. 27 November 2016.

- [9] J. Andrés Christen, Colin Fox, A General Purpose Sampling Algorithm for Continuous Distributions (the t-walk). International Society for Bayesian Analysis. 2010.
- [10] G. Casella, E. I. George, Explaining The Gibbs Sampler. The American Statistician. Aug. 1992.
- [11] S. Chib, E. Greenberg, Understanding the Metropolis Hastings algorithm. The American Statistician. Nov. 1995.
- [12] [M. Capistrán, J. Christen, S. Donnet, 2013]. Bayesian Analysis of ODE's: Solver optimal accuracy and Bayes factors
- [13] [Lee, P. M. (2012)]. Bayesian statistics: an introduction. John Wiley Sons.