



Universidad Autónoma de Guerrero

Unidad Académica de Matemáticas

Maestría en Matemáticas Aplicadas

**Análisis Bayesiano en problemas
inversos para ecuaciones
diferenciales**

T E S I S

PARA OBTENER EL GRADO DE:

Maestría en Matemáticas Aplicadas

PRESENTA:

Luis Xavier Vivas Cruz

DIRECTORES DE TESIS

Dr. Francisco J. Ariza Hernández

Dr. Jorge Sánchez Ortiz

Abril de 2016.



Universidad Autónoma de Guerrero

Unidad Académica de Matemáticas

Maestría en Matemáticas Aplicadas

**Análisis Bayesiano en problemas
inversos para ecuaciones
diferenciales**

T E S I S

PARA OBTENER EL GRADO DE:

Maestría en Matemáticas Aplicadas

PRESENTA:

Luis Xavier Vivas Cruz

DIRECTORES DE TESIS:

Dr. Francisco J. Ariza Hernandez

Dr. Jorge Sánchez Ortíz

Abril de 2016.

Agradecimientos

Quiero agradecer primeramente a Dios por darme la oportunidad de haber concluido este grado de estudios.

A la *Universidad Autónoma de Guerrero* (UAGro), por brindarme el espacio necesario para llevar a cabo este logro.

Al *posgrado de Matemáticas Aplicadas*, por permitirme ser un estudiante en este programa.

Gracias al *Consejo Nacional de Ciencia y Tecnología* (CONACyT) por haberme apoyado económicamente durante los dos años de estudio.

A mis Padres y esposa, por todo su apoyo incondicional y comprensión ofrecidos en todo momento.

A todos los profesores del posgrado de Matemáticas Aplicadas, por las verdaderas clases impartidas, por su motivación, conocimiento y apoyo considerable durante la estancia en esta Maestría.

A mis directores de Tesis:

Dr. Francisco J. Ariza Hernandez, Dr. Jorge Sánchez Ortíz y Dr. Martín P. Árciga Alejandre, por su apoyo constante y paciencia en la elaboración de este trabajo.

¡MUCHAS GRACIAS A TODOS y que *la Fuerza os acompañe!*

Resumen

Uno de los principales objetivos de la ciencia, es tratar de describir y predecir eventos que suceden en el mundo, esto típicamente se hace a través de representaciones matemáticas. Por ejemplo, muchas de las leyes de la naturaleza pueden ser expresadas como sistemas de ecuaciones diferenciales determinísticas, las cuales típicamente contienen parámetros que caracterizan y/o describen dicho sistema. Sin embargo, tan pronto se incorporen datos experimentales, un grado de incertidumbre se presenta debido a múltiples factores, entre estos, errores de medición o de modelamiento. Los problemas inversos se resuelven a través de un conjunto de métodos, que a partir de datos observados o resultados de algunas mediciones, nos sirven para inferir sobre los valores de los parámetros que caracterizan el fenómeno (físico, biológico, etc.) bajo estudio. En este trabajo, se usa la teoría de la inversión estadística para dar una solución sistemática a problemas inversos, la cual los reformula como problemas de inferencia estadística, bajo el enfoque de la estadística Bayesiana. Particularmente, se estudian dos problemas inversos en presencia de incertidumbre para ecuaciones diferenciales. El primero es el modelo de Kotter, el cual es usado en el análisis de deslizamiento de laderas, y el segundo es un modelo de crecimiento exponencial con derivada fraccionaria en el tiempo. Para éste último modelo, se ajustan dos conjuntos de datos reales y se realiza un estudio de simulación para obtener propiedades de los estimadores Bayesianos en términos de su Error Cuadrático Medio, Sesgo y Cobertura estimada. Por último, se realiza una selección Bayesiana de modelos entre un modelo de crecimiento fraccionario contra uno ordinario, usando el Factor de Bayes.

Abstract

One of the main aims of science is to try to describe and predict events that happen in the world, this is typically done through mathematical representations. For example, many of the laws of nature can be expressed as deterministic systems of differential equations, which typically contain parameters that characterize and / or describe the system. However, experimental data as soon incorporate a degree of uncertainty arises due to multiple factors including errors of measurement or modeling. Inverse problems are solved using a set of methods, from observed data or results of some measurements, help us to infer the values of the parameters that characterize the phenomenon (physical, biological, etc.) under study. In this paper, the statistical inversion theory is used to give a systematic solution to inverse problems, which reformulates as problems of statistical inference, under the approach of Bayesian statistics. In particular, two inverse problems in the presence of uncertainty for differential equations are studied. The first is the Kotter model, which is used in the analysis of slope slides, and the second is an exponential growth model with fractional time derivative. For the latter model, two real data sets are adjusted and a simulation study is performed to obtain properties of the Bayesian estimators in terms of Mean Square Error, bias and estimated coverage. Finally, a Bayesian model selection is performed between a growth model to one ordinary fractional using the Bayes Factor.

Índice general

1	Introducción	1
1.1	Planteamiento del problema	4
2	Los problemas inversos Bayesianos	6
2.1	La teoría de la inversión estadística	8
2.2	Fundamentos de la estadística Bayesiana	9
2.2.1	El teorema de Bayes	10
2.2.2	Inferencia Bayesiana	11
2.3	Métodos Monte Carlo vía Cadenas de Markov	13
2.3.1	Diagnósticos de convergencia	17
2.3.2	Implementación de software para MCMC	20
3	Problema inverso Bayesiano para el modelo de Kotter	23
3.1	Introducción	23
3.2	Métodos existentes para determinar una superficie de deslizamiento	24
3.3	Estimación	25
3.4	Conclusión	31
4	Problema inverso Bayesiano para un modelo de crecimiento poblacional fraccionario	32
4.1	Modelo	32
4.2	Solución para el sistema dinámico	33
4.3	Selección de un modelo de crecimiento poblacional fraccionario	34
4.3.1	Estudio de simulación	37

4.3.2	Resultados	39
4.4	Ejemplos de aplicación	41
4.5	Conclusión	57
	Referencias	57
	Anexos	62

Índice de figuras

3.1	<i>Interpretación de las ecuaciones de Kotter.</i>	26
3.2	<i>Conjunto de datos simulados de (3.3) para los 3 tipos de suelo.</i>	29
3.3	<i>Traza y distribución posterior para el parámetro R (circulo de deslizamiento).</i>	30
3.4	<i>Diagnóstico de convergencia de Gelman y Rubin; y de Geweke para el parámetro R (Cadena 1 y 2), respectivamente.</i>	30
4.1	<i>Diagrama de Cajas y Bigotes del DIC en ambos modelos para los casos del parámetro X_0.</i>	41
4.2	<i>Crecimiento de pastos: Observaciones v.s. tiempo.</i>	42
4.3	<i>Curva ajustada para el crecimiento de pastos para el caso de X_0 conocida.</i>	43
4.4	<i>Trazas y distribuciones posteriores para los parámetros a, α y σ^2, respectivamente.</i>	44
4.5	<i>Diagnóstico de convergencia de Gelman y Rubin para los parámetros a, α y σ^2 respectivamente.</i>	45
4.6	<i>Diagnóstico de convergencia de Geweke para los parámetros a, α y σ^2 en las Cadenas 1 y 2, respectivamente.</i>	45
4.7	<i>Curva ajustada para el crecimiento de pastos para el caso de X_0 desconocida.</i>	46
4.8	<i>Trazas y distribuciones posteriores para los parámetros a, α, σ^2 y X_0.</i>	48
4.9	<i>Diagnóstico de convergencia de Gelman y Rubin para los parámetros a, α, σ^2 y X_0 respectivamente.</i>	48

4.10	<i>Diagnóstico de convergencia de Geweke para los parámetros a, α, σ^2 y X_0 respectivamente, Cadena 1.</i>	49
4.11	<i>Diagnóstico de convergencia de Geweke para los parámetros a, α, σ^2 y X_0 respectivamente, Cadena 2.</i>	50
4.12	<i>Porcentaje de la demanda de la Web móvil v.s. tiempo.</i>	53
4.13	<i>Curva ajustada para las observaciones del porcentaje de la demanda de la Web móvil.</i>	54
4.14	<i>Trazas y distribuciones posteriores para los parámetros a, α y σ^2.</i>	55
4.15	<i>Diagnóstico de convergencia de Gelman y Rubin para los parámetros a, α y σ^2 respectivamente.</i>	56
4.16	<i>Diagnóstico de convergencia de Geweke para los parámetros a, α y σ^2 respectivamente, Cadena 1.</i>	56
4.17	<i>Diagnóstico de convergencia de Geweke para los parámetros a, α y σ^2 respectivamente, Cadena 2.</i>	56

Introducción

Una de las principales razones por la que las ecuaciones diferenciales son importantes es debida a la amplia gama de fenómenos físicos o biológicos que éstas pueden describir. Uno de los problemas que se abordan en esta tesis, es el deslizamiento de laderas, que sin duda constituye uno de los desastres geológicos más destructivos a nivel mundial, causando pérdidas económicas y también lamentables pérdidas humanas. La importancia de realizar un análisis de este fenómeno, radica en determinar la superficie de deslizamiento para algunos tipos de suelo de una zona urbana en particular, pues un deslizamiento en estas zonas puede causar enormes pérdidas humanas y materiales.

En este trabajo, se pretende utilizar una herramienta matemática poco usada para modelar tal deslizamiento, que es mediante ecuaciones diferenciales ordinarias. Para lograr esto, analizaremos el problema inverso para las ecuaciones de Kotter, el parámetro de interés en este modelo, es el radio, R , de la superficie de deslizamiento, tal problema será estudiado mediante la teoría de la inversión estadística.

Un segundo modelo que se estudia, es el de crecimiento poblacional exponencial. Este puede ser modelado por una ecuación diferencial ordinaria, esta ecuación involucra ciertos parámetros que caracterizan la evolución en el tiempo de dicho sistema. Una forma común de analizar este fenómeno, es realizar observaciones en el transcurso de algún período de tiempo, registrarlas para obtener una ecuación que mejor las ajuste, (Samson y Donnet, 2009). Dicha ecuación falla al ajustar datos que presentan un com-

portamiento de crecimiento anómalo. Por lo que la población crece más lento o más rápido. Para abordar este problema, las ecuaciones diferenciales fraccionarias pueden ser consideradas, ya que pueden ofrecer un mejor ajuste a los datos experimentales, (Jin y Rundell, 2015).

En la literatura existen pocos modelos de crecimiento poblacional construidos en el contexto de los problemas inversos bajo el punto de vista Bayesiano, en donde una de las metas consiste en que a partir de una colección de datos de crecimiento sujetos a incertidumbre, se desea caracterizar la dinámica del crecimiento.

En los últimos años, el enfoque estadístico de los problemas inversos desde un punto de vista Bayesiano ha sido de gran interés y popularidad para el desarrollo de la cuantificación de incertidumbre en presencia de datos. Libros, tales como Kaipio y Somersalo (2006); Banks et al. (2014) y Tarantola (2005) ofrecen una extensa literatura sobre aspectos estadísticos de problemas inversos, en particular el enfoque Bayesiano. Donnet et al. (2010), desarrolló inferencia Bayesiana de los parámetros de ecuaciones diferenciales estocásticas, deducidas a partir de una ecuación determinística estándar de crecimiento, adicionando efectos aleatorios para la dinámica de crecimiento en una población. En Capistrán et al. (2013), también se desarrolló un análisis Bayesiano de problemas inversos en ecuaciones diferenciales ordinarias. Calvetti et al. (2014), y entre otros, mencionan varios trabajos sobre el desarrollo del marco de referencia para problemas inversos Bayesianos en una amplia gama de aplicaciones, tales como ingeniería, geofísica, ciencias de la tierra y economía. En Dashti y Stuart (2015), se discute el enfoque Bayesiano para problemas inversos en ecuaciones diferenciales desde una perspectiva matemática computacional para la formulación y desarrollo de algoritmos; éste artículo también describe métodos Monte Carlo vía Cadenas de Markov (MCMC) y métodos Monte Carlo secuenciales, además de medidas de preservación reversibles de ecuaciones diferenciales estocásticas (SDE) en espacios de dimensión infinita.

Para verificar si los datos existentes provienen de un modelo FEG o de un modelo OEG, implementamos un método Bayesiano para la selección de modelos basado en el factor de Bayes utilizando la aproximación de la verosimilitud marginal (ML). El factor de Bayes es usado como una herramienta práctica para cuantificar la evidencia a favor de un modelo específico y proporcionar una manera de incorporar información

externa en la evaluación de la evidencia a cerca de una hipótesis, (Kass y Raftery, 1995). Para calcular ML, podemos integrar el producto de la función de verosimilitud y la distribución a priori, pero a veces resulta muy difícil evaluar esta integral, (Gelfand y Dey, 1994). Una forma de estimar ML está basada en las muestras de la distribución posterior que resultan del método MCMC; Chib (1995), hace uso de estas muestras, y mediante el estimador de Gelfand y Dey da una solución alternativa para el cálculo de ML.

La distribución posterior conjunta de los parámetros, como solución a un problema inverso, tiene una forma analítica que no es conocida. Entonces, para explorar dicha distribución, los métodos MCMC son implementados, (Chen et al., 2012).

La selección del modelo Bayesiano en el marco de referencia de problemas inversos ha sido poco investigado. Donnet et al. (2010), sugiere validar el enfoque de SDE a través de criterios basados en la distribución predictiva posterior. Para comparar una distribución posterior numérica con una teórica, en Capistrán et al. (2013), se propone utilizar el factor de Bayes, considerando a ambos como modelos para los datos.

1.1. Planteamiento del problema

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

Modelo general:

1) Sistema dinámico:

$$\begin{cases} \frac{dX(t)}{dt} = H(X(t), t, \xi) \\ X(t_0) = X_0, \end{cases} \quad (1.1)$$

2) Ecuación de observación:

$$y_i = h(X(t_i)) + \varepsilon_i, \quad i = 1, \dots, n. \quad (1.2)$$

Asumiendo un proceso de observaciones $\mathbf{y} = (y_1, y_2, \dots, y_n)^t$ a tiempo discreto t_1, \dots, t_n , donde y_i corresponde al i -ésimo valor observado bajo incertidumbre, $h : \mathbb{R}^p \rightarrow \mathbb{R}^k$ es la función de observación, $X(t)$ corresponde a la solución del sistema (1.1). Se desea estimar las cantidades $\boldsymbol{\theta} = (X_0, \xi, \nu)$, el cual es el vector de parámetros desconocidos, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$; esto con la finalidad de caracterizar dicho sistema. La condición de Lipschitz para la función $H : \mathbb{R}^p \times [0, T] \times \Theta \rightarrow \mathbb{R}^p$ asegura la existencia y unicidad de la solución del problema de 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, (Capistrán et al., 2013). Aquí, sólo se considera un problema de observaciones unidimensionales. El parámetro ν representa las cantidades de interés contenidas en el error de medición para las observaciones al tiempo t_i , los cuales son considerados como variables aleatorias independientes e idénticamente distribuidas (*i.i.d*) de una distribución Normal, con media cero y varianza constante σ^2 , denotada por $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

La presente tesis está estructurada como sigue: en el Capítulo 2, se presenta una breve introducción a los problemas inversos que serán estudiados mediante la teoría de la inversión estadística usando la estadística Bayesiana. En el Capítulo 3, se aborda

el problema inverso para un modelo de deslizamiento de laderas. Y finalmente en el Capítulo 4, se estudia el problema inverso para un modelo de crecimiento poblacional fraccionario, así como una selección de modelos; y una aplicación a ejemplos con datos reales. Después de las referencias bibliográficas, se presentan dos anexos; el anexo 1, comprende la derivación del modelo de Kotter, que es utilizado en el capítulo 3, y en el anexo 2, se da un ejemplo de la implementación en JAGS, el cual es el código usado en el capítulo 4 .

Los problemas inversos Bayesianos

Los problemas inversos son definidos, como el propio término lo indica, como el inverso de problemas directos. Es evidente que esta definición es vacía a menos que se defina el concepto de problemas directos. Los problemas inversos se encuentran típicamente en situaciones en las que se hacen observaciones indirectas de una cantidad de interés.

Las teorías físicas permiten hacer predicciones: dada una descripción completa de un sistema físico, se puede predecir el resultado de algunas mediciones. Este problema de predecir el resultado de mediciones se llama *problema de modelización*, *problema de simulación*, o *problema directo*, (Tarantola, 2005). El problema inverso consiste en utilizar el resultado existente de algunas mediciones para inferir valores de los parámetros que caracterizan el sistema.

Mientras que un problema directo tiene una única solución, un problema inverso podría tener múltiples soluciones, lo que se convierte en un problema inestable y mal planteado. Se dice que un problema inverso,

$$\mathbf{y} = h(\xi) + \varepsilon,$$

está bien planteado (*well-posed*) si:

1. Existe una solución para alguna \mathbf{y} en el espacio de datos observados.
2. La solución es única.

3. La solución depende sensiblemente de las condiciones iniciales.

Debido a esto, en problemas inversos, se necesita tener información inicial disponible acerca de las cantidades que son de interés. Además, se necesita tener cuidado en la representación de la *incertidumbre* en los datos.

La teoría más general (y simple), es obtenida cuando se usa un punto de vista probabilístico, donde la información disponible de una cantidad es representada por una distribución de probabilidad. Esta teoría explica como esta información inicial es transformada en una distribución posterior de probabilidad, mediante un modelo y los resultados actuales de observaciones (con sus incertidumbres).

Con el fin de cuantificar la incertidumbre en las estimaciones de las cantidades de interés, se formula un modelo estadístico de la forma:

$$\mathbf{Y} = h(t; \boldsymbol{\theta}_0) + \boldsymbol{\varepsilon}, \quad (2.1)$$

donde $\boldsymbol{\theta}_0$ son los verdaderos valores hipotéticos de los parámetros desconocidos, y $\boldsymbol{\varepsilon}$ es un vector aleatorio que representa el error de medición para las variables medidas al tiempo t . Las propiedades para los errores son definidas como:

$$\begin{aligned} E(\boldsymbol{\varepsilon}) &= 0, \\ \text{Var}(\boldsymbol{\varepsilon}) &= V_0 = \text{diag}(\sigma_{0,1}^2, \sigma_{0,2}^2, \dots, \sigma_{0,m}^2), \\ \text{Cov}(\varepsilon_i, \varepsilon_i) &= \text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j. \end{aligned}$$

En la ecuación de observación (1.2) se tiene que ε_i tiene distribución Normal, por lo tanto las observaciones y_i también tendrán una distribución similar, la cual es denotada como $y_i \sim \mathcal{N}(h(X(t_i)), \sigma^2)$. Si existen n observaciones (*i.i.d.*), y_1, y_2, \dots, y_n , con ésta distribución conocida, entonces la función de verosimilitud para las observaciones dadas es:

$$L(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} (y_i - h(X(t_i)))^2 \right\}.$$

Cualquier decisión estadística de los datos, como la estimación, predicción o selección de modelo se basa en ésta función, la cual, para un conjunto de observaciones $\mathbf{y} = (y_1, y_2, \dots, y_n)^t$ puede ser expresada de la forma

$$L(\mathbf{y}|\boldsymbol{\theta}) = \sigma^{-n} (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - h(X(t_i)))^2 \right\}. \quad (2.2)$$

Note que (2.2) involucra el cálculo de $X(t)$, una solución de (1.1).

Añadiendo la información inicial acerca de θ , el problema inverso, entonces se puede establecer como un problema de combinar toda esta información.

La *teoría de la inversión estadística* resuelve los problemas inversos sistemáticamente en una manera tal que toda la información disponible es propiamente incorporada en el modelo.

2.1. La teoría de la inversión estadística

Esta teoría reformula los problemas inversos como problemas de inferencia estadística mediante la estadística Bayesiana, en la cual todas las cantidades son modeladas como variables aleatorias (*v.a.*'s).

La aleatoriedad, la cual refleja la incertidumbre ante el observador es codificada en distribuciones de probabilidad de las cantidades de interés.

En esta sección, se explica la aproximación Bayesiana hacia problemas inversos. Se tienen cantidades observables directamente y otras que no pueden ser observadas. En los problemas inversos, algunas de las cantidades no observables son de mayor interés. Estas cantidades dependen una de la otra a través de modelos.

El objetivo de la inversión estadística, es extraer información y evaluar la incertidumbre acerca de las variables basadas en todo el conocimiento disponible de un proceso de observación.

La aproximación de la inversión estadística está basada en los siguientes principios:

1. Todas las variables incluídas en el modelo son modeladas como *v.a.*'s.
2. La aleatoriedad describe nuestro grado de información concerniente a las observaciones.
3. El grado de información es codificado mediante distribuciones de probabilidad.
4. **La solución del problema inverso es la distribución posterior de probabilidad de las cantidades de interés.**

Esta distribución describe el grado de confianza acerca de la cantidad de interés θ después de realizadas las observaciones \mathbf{y} .

En la mayoría de los problemas inversos, uno de los desafíos es la enorme dimensión del problema, y en consecuencia tal distribución posterior también es de dimensión alta; para esto es posible utilizar métodos de simulación mediante Cadenas de Markov (descritos en la sección 2.3) para explorar esta distribución.

2.2. Fundamentos de la estadística Bayesiana

La estadística Bayesiana es un enfoque particular de la estadística en general. El término “Bayesiana” hace referencia a Thomas Bayes (1702-1761), un matemático británico y ministro presbiteriano, cuya obra más conocida es el teorema que lleva su nombre, “*El teorema de Bayes*”, el cual se refiere a la probabilidad de un evento condicionado por la ocurrencia de otro evento.

Más específicamente con su teorema se resuelve el problema conocido como “probabilidad inversa”. Se trata de probabilidad “inversa” en el sentido de que la “directa” sería la probabilidad de observar algo supuesto que rigen ciertas condiciones.

Desde el punto de vista de la aproximación Bayesiana, existe una relación entre probabilidad e información, en donde el teorema de Bayes proporciona un modo natural de actualización de las creencias a cerca de una cantidad de interés cuando nueva información es incorporada al modelo. Tal proceso, es la base de la inferencia Bayesiana.

Dos aspectos importantes que caracterizan el enfoque de la aproximación Bayesiana son:

- a) Considerar que todas las formas de incertidumbre son expresadas en términos de una medida de probabilidad. Se piensa que la probabilidad es una medida de lo que se sabe acerca de un evento.
- b) Las probabilidades de un evento son actualizadas mediante evidencias.

El teorema de Bayes es la herramienta que permite explicar este enfoque. Este teorema será de utilidad para la solución a los problemas inversos que se tratarán en este trabajo.

2.2.1. El teorema de Bayes

Suponiendo que todas las variables aleatorias son absolutamente continuas, esto es, sus distribuciones de probabilidad pueden ser expresadas en términos de densidades de probabilidad. Como en el caso de problemas inversos clásicos, suponiendo que se están observando cantidades

$$\mathbf{y} \in \mathbb{R}^k,$$

con el fin de obtener información de alguna otra cantidad

$$\boldsymbol{\theta} \in \mathbb{R}^d.$$

Con el fin de relacionar estas dos cantidades, se necesita un modelo para su dependencia. Este modelo puede ser inapropiado y puede contener parámetros que no son conocidos. Además, las observaciones \mathbf{y} siempre contienen error.

Desde la perspectiva de la inversión estadística, se llama a Y como la *v.a.* observable (mediciones) y su realización, \mathbf{y} , en el proceso de medición.

La *v.a.* no observable $\boldsymbol{\theta}$ que es de nuestro interés, es llamada *desconocida*. Aquellas variables que no son observables ni de interés, son llamadas *ruído*.

Asumiendo que antes de realizar las mediciones de \mathbf{y} , se tiene alguna información acerca de la *v.a.* $\boldsymbol{\theta}$. Del teorema de Bayes, ésta información puede ser introducida por una densidad de probabilidad inicial $p(\boldsymbol{\theta})$, la cual expresa que conocimiento se tiene acerca de la cantidad de interés.

Asumiendo que después de analizar las observaciones \mathbf{y} e incorporando la información disponible acerca de $\boldsymbol{\theta}$, la densidad de probabilidad conjunta de $\boldsymbol{\theta}$ y \mathbf{y} es denotada por $\pi(\boldsymbol{\theta}, \mathbf{y})$.

La función de densidad marginal de $\boldsymbol{\theta}$ debe ser

$$\int_{\mathbb{R}^k} \pi(\boldsymbol{\theta}, \mathbf{y}) d\mathbf{y} = p(\boldsymbol{\theta}).$$

Finalmente, si los datos \mathbf{y} son dados. La distribución de probabilidad condicional

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\boldsymbol{\theta}, \mathbf{y})}{\pi(\mathbf{y})},$$

es llamada la *distribución posterior* de $\boldsymbol{\theta}$.

Se resume lo mencionado en el siguiente teorema, el cual puede ser referido como:

Teorema 1 (teorema de Bayes). *Suponiendo que la variable $\boldsymbol{\theta} \in \mathbb{R}^d$ tiene una densidad de probabilidad inicial $p(\boldsymbol{\theta})$ y los datos consisten de valores observados \mathbf{y} de una v.a. observada $Y \in \mathbb{R}^k$ tal que $\pi(\mathbf{y}) > 0$. Entonces la distribución de probabilidad de $\boldsymbol{\theta}$, dados los datos \mathbf{y} es:*

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\pi(\mathbf{y})}. \quad (2.3)$$

El objetivo de la aproximación Bayesiana, es suministrar una metodología para estudiar adecuadamente la información mediante análisis de datos y usar la teoría de la decisión para actuar de la mejor manera. Dicha metodología Bayesiana puede ser idealizada por tres aspectos importantes:

1. Establecer un modelo completo de probabilidad. Una distribución conjunta de probabilidad para todas las cantidades observables y no observables en un problema. El modelo debe ser consistente con el conocimiento acerca del problema en estudio y el proceso de colección de datos.
2. Condicionamiento de los datos observados: calcular e interpretar la distribución posterior apropiada. La distribución condicional de probabilidad de las cantidades no observables de interés dados los datos observados.
3. Evaluar el ajuste del modelo y las implicaciones de la distribución posterior resultante.

A continuación se presentarán los elementos para la construcción de un modelo estadístico Bayesiano mediante la inferencia Bayesiana.

2.2.2. Inferencia Bayesiana

Los métodos Bayesianos han sido generalizados particularmente porque son útiles para la solución de problemas en la toma de decisiones.

La estadística Bayesiana proporciona cantidades conocidas y desconocidas, lo cual permite incorporar los datos de los que se tiene conocimiento dentro de la estimación de los parámetros dados inicialmente, logrando así un proceso de estimación más rico en información, realizando inferencias sobre aquellas cantidades desconocidas.

Una vez que el modelo es definido, es objetivo principal de la estadística Bayesiana es realizar inferencias acerca de los parámetros desconocidos $\boldsymbol{\theta}$. Es posible utilizar información previa acerca de los valores de éstos para incorporarlos al análisis de los datos.

En el contexto de la teoría de la inversión estadística, la solución a un problema inverso es la distribución posterior de las cantidades de interés, $\boldsymbol{\theta}$, dado que toda la información disponible se ha incorporado en el modelo, (Kaipio y Somersalo, 2006).

Así, por el teorema de Bayes, la distribución posterior de los parámetros de interés está dada por

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{L(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{m_{\mathbf{Y}}(\mathbf{y})}, \quad (2.4)$$

donde $p(\boldsymbol{\theta})$ es la distribución inicial para $\boldsymbol{\theta}$ y

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

es una constante de normalización, también llamada la *verosimilitud marginal* (ML) de los datos \mathbf{y} , Θ denota el espacio paramétrico de $\boldsymbol{\theta}$.

Suponiendo que $\boldsymbol{\theta}$ es una variable aleatoria que tiene una distribución a priori denotada por $p(\boldsymbol{\theta})$. La información concerniente a $\boldsymbol{\theta}$ está basada en la distribución posterior, la cual es obtenida por el teorema de Bayes. Tal distribución de $\boldsymbol{\theta}$ en presencia de las observaciones \mathbf{y} ésta dada por

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{L(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\Theta} L(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (2.5)$$

Es claro que $\pi(\boldsymbol{\theta}|\mathbf{y})$ es proporcional a la función de verosimilitud multiplicada por la distribución a priori,

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto L(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \quad (2.6)$$

y así, ésta involucra una contribución de los datos observados a través de $L(\mathbf{y}|\boldsymbol{\theta})$, y una contribución de la información a priori cuantificada a través de $p(\boldsymbol{\theta})$.

El teorema de Bayes actualiza el conocimiento de $\boldsymbol{\theta}$ extrayendo información útil de interés, la cual está contenida en el conjunto de observaciones \mathbf{y} .

El uso de la inferencia Bayesiana tiene mucha ventaja en la aplicación a fenómenos físicos, pero presenta un gran esfuerzo al momento de seleccionar la distribución a priori adecuada, además de la complejidad en los cálculos de la distribución posterior.

Una de las limitaciones de la inferencia Bayesiana se concentra en la representación explícita de la distribución posterior de las cantidades de interés, ya que resulta analíticamente intratable y las distribuciones posteriores marginales de los parámetros son complicadas. Sin embargo, en recientes años los estadísticos han desarrollado métodos de muestreo Monte Carlo vía Cadenas de Markov (MCMC) para simular tales distribuciones. Los métodos más comunes son Metropolis-Hastings, (Metropolis et al., 1953; Chib y Greenberg, 1995) y el muestreo Gibbs, (Gelfand y Smith, 1990; Casella y George, 1992).

2.3. Métodos Monte Carlo vía Cadenas de Markov

Los métodos Monte Carlo, basados en Cadenas de Markov, se han caracterizado como una herramienta útil para resolver una infinidad de problemas que surgen en la modelación paramétrica y que son intratables analíticamente.

Para problemas numéricos de altas dimensiones, los métodos MCMC son con frecuencia más eficientes que los métodos numéricos convencionales, (Hastings, 1970). Estos métodos están basados en la simulación de Cadenas de Markov, cuya distribución estacionaria corresponde a la distribución posterior de los parámetros $\pi(\boldsymbol{\theta}|\mathbf{y})$.

La cuestión clave es cómo se producen las transiciones de un estado a otro de la Cadena para que se produzca la convergencia.

Algoritmo de Metropolis-Hastings.

El algoritmo básico propuesto por Metropolis-Hastings, el cual fue desarrollado por Metropolis, Rosenbluth y Teller en 1953 y subsecuentemente generalizado por Hastings en 1970 (Chib y Greenberg, 1995), constituye un método para simular distribuciones multivariadas. Se dará una definición de lo que es un Método Monte Carlo vía Cadenas de Markov.

Definición 1. *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, (Lee, 2012). 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,

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

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

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

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

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

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 $U \sim \mathcal{U}(0, 1)$ el cual es uniformemente distribuido en $(0, 1)$.

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.

Muestreador de Gibbs.

El muestreador de Gibbs puede verse como un caso particular del algoritmo de Metropolis-Hastings, sin embargo, como menciona Robert y Casella (2013), tiene algunas características particulares que le dan entidad propia.

El algoritmo Gibbs permite simular una Cadena de Markov $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(k)}$, con distribución de equilibrio $\pi(\boldsymbol{\theta}|\mathbf{y})$ de dimensión d . En este caso, sin embargo, cada nuevo

valor de la Cadena se obtiene a través de un proceso iterativo que sólo requiere generar muestras de distribuciones cuya dimensión es menor que d y que en la mayoría de los casos tienen una forma más sencilla que la de $\pi(\boldsymbol{\theta}|\mathbf{y})$.

Sea $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k)$ una partición del vector $\boldsymbol{\theta}$, donde $\boldsymbol{\theta}_i \in \mathbb{R}^{d_i}$ y $\sum_{i=1}^k d_i = d$. Las densidades

$$\begin{aligned} & \pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_k, \mathbf{y}) \\ & \quad \vdots \\ & \pi(\boldsymbol{\theta}_i|\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \dots, \boldsymbol{\theta}_k, \mathbf{y}) \quad (i = 2, \dots, k-1) \\ & \quad \vdots \\ & \pi(\boldsymbol{\theta}_k|\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{k-1}, \mathbf{y}) \end{aligned}$$

se conocen como *densidades condicionales completas* y en general pueden identificarse fácilmente al inspeccionar la forma de la distribución posterior $\pi(\boldsymbol{\theta}|\mathbf{y})$. De hecho, para cada $i = 1, \dots, k$,

$$\pi(\boldsymbol{\theta}_i|\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \dots, \boldsymbol{\theta}_k, \mathbf{y}) \propto \pi(\boldsymbol{\theta}|\mathbf{y}),$$

donde $\pi(\boldsymbol{\theta}|\mathbf{y}) = \pi(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k|\mathbf{y})$ es vista sólo como función de $\boldsymbol{\theta}_i$.

Dado un valor inicial $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_1^{(0)}, \dots, \boldsymbol{\theta}_k^{(0)})$, el algoritmo de Gibbs simula una Cadena de Markov en la que $\boldsymbol{\theta}^{(t+1)}$ se obtiene a partir de $\boldsymbol{\theta}^{(t)}$ de la siguiente manera:

$$\begin{aligned} & \text{generar una observación } \boldsymbol{\theta}_1^{(t+1)} \text{ de } \pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2^{(t)}, \boldsymbol{\theta}_3^{(t)}, \dots, \boldsymbol{\theta}_k^{(t)}, \mathbf{y}); \\ & \quad \vdots \\ & \text{generar una observación } \boldsymbol{\theta}_2^{(t+1)} \text{ de } \pi(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_3^{(t)}, \dots, \boldsymbol{\theta}_k^{(t)}, \mathbf{y}); \\ & \quad \vdots \\ & \text{generar una observación } \boldsymbol{\theta}_k^{(t+1)} \text{ de } \pi(\boldsymbol{\theta}_k|\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t+1)}, \dots, \boldsymbol{\theta}_{k-1}^{(t+1)}, \mathbf{y}). \end{aligned}$$

La sucesión $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^k$ así obtenida, es entonces una realización de una Cadena de Markov cuya distribución de transición está dada por

$$\pi(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)}) = \prod_{i=1}^k \pi(\boldsymbol{\theta}_i^{(t+1)}|\boldsymbol{\theta}_1^{(t+1)}, \dots, \boldsymbol{\theta}_{i-1}^{(t+1)}, \boldsymbol{\theta}_{i+1}^{(t)}, \dots, \boldsymbol{\theta}_k^{(t)}, \mathbf{y}).$$

2.3.1. 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, después de un cierto número de iteraciones T suficientemente grande, los valores $\boldsymbol{\theta}_1^{(T)}, \dots, \boldsymbol{\theta}_N^{(T)}$ 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:

Diagnóstico de secuencia múltiple de Gelman y Rubin.

Para cada parámetro se deben realizar los siguientes pasos:

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.

1) 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.

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

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

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

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.

2.3.2. Implementación de software para MCMC

En la actualidad, muchos de los algoritmos MCMC ya se han implementado en programas de computadora, tales como WinBUGS (Spiegelhalter, D. J. et al, 2003), JAGS (Plummer, 2012), Stan (Stan Development Team, 2014) y t-walk, (Christen y Fox, 2010). Todos estos programas proporcionan paquetes para el modelado Bayesiano mediante la simulación de la distribución posterior dada la información existente y un modelo específico.

En esta tesis, se utiliza el programa JAGS para obtener muestras de las distribuciones posteriores marginales de interés.

Introducción a JAGS.

JAGS (Just Another Gibbs Sampler), es un programa para análisis de modelos Bayesianos usando métodos MCMC. Este programa fue desarrollado con tres objetivos: tener una herramienta para el lenguaje BUGS que se ejecuta en Unix; ser extensible, permite a los usuarios escribir sus propias funciones, distribuciones y muestreadores; y ser una plataforma para la experimentación con ideas en modelación Bayesiana.

Ejecutar un modelo en JAGS significa generar muestras de la distribución posterior de los parámetros del modelo. Esto se lleva a cabo en cinco pasos:

1. *Definición del modelo.*

Hay dos partes para definir un modelo en JAGS, una descripción de éste y la definición de los datos.

- 1) El modelo se define en un archivo de texto usando un lenguaje BUGS. Consiste en una serie de relaciones dentro de un bloque delimitado por llaves `{}` y precedido por la palabra: `modelo`. Cada relación define un nodo en términos de otros, estos son referidos como nodos principales.

Las relaciones pueden ser de dos tipos, una *relación estocástica* (\sim) define

un nodo estocástico, representando una variable aleatoria en el modelo; una *relación determinística* (\leftarrow) que define un nodo determinístico, el valor del cual está determinado exactamente por los valores entre paréntesis.

- II) Los datos son definidos en un archivo separado de la definición del modelo. Los valores de los datos deben ser proporcionados por nodos estocásticos y constantes. Es un error proporcionar un dato en un nodo determinístico. Los nodos estocásticos no observados, están referidos como los parámetros del modelo.

2. *Compilación.*

Cuando un modelo es compilado, un gráfico que es representado por el modelo es creado en la memoria de la PC. La compilación puede fallar por las razones siguientes:

- I) El gráfico contiene un ciclo dirigido. Estos son prohibidos en JAGS.
- II) Un parámetro de nivel superior no está definido. Algún nodo que es usado del lado derecho de la relación, pero no es definido del lado izquierdo de tal relación, es supuesto por ser un nodo constante.
- III) El modelo usa una función o distribución que no ha sido definida en algún otro modulo cargado.

Las Cadenas paralelas que se ejecutan en JAGS deben producir una secuencia de muestras independientes de la distribución posterior.

3. *Inicialización.*

Antes de ejecutar un modelo, es necesario inicializarlo. Existen tres pasos para esta tarea:

- I) Se establecen los valores iniciales de los parámetros del modelo.
- II) Un generador de números aleatorios (RNG) es seleccionado para cada Cadena paralela, y su semilla se define.
- III) Las muestras se eligen para cada parámetro en el modelo.

4. *Adaptación y burn-in.*

En teoría, la salida de las muestras MCMC convergen a la distribución posterior cuando el número de iteraciones tiende a infinito. Por convención, la salida del método MCMC se divide en dos partes: un periodo inicial “burn-in”, el cual es descartado, y el resto es considerado por converger a la distribución posterior.

5. *Monitoreo.*

Es un objeto que registra los valores muestreados. El monitoreo más simple es una traza de monitoreo, que almacena el valor muestreado de un nodo en cada iteración.

JAGS no puede monitorear un nodo a menos que éste haya sido definido en el archivo del modelo.

Las siguientes etapas de análisis son hechas fuera de JAGS: diagnósticos de convergencia, crítica del modelo y el resumen de las muestras debe ser hecho usando otros paquetes más adecuados para esta tarea. Hay varios paquetes en R diseñados para analizar la salida MCMC. Los paquetes de R (R Core Team, 2014), tales como R2WinBUGS (Sturtz et al., 2005), R2jags (Su y Yajima, 2015) y rjags (Plummer, 2015) permiten ejecutar WinBUGS y JAGS dentro del ambiente de R.

Problema inverso Bayesiano para el modelo de Kotter

3.1. Introducción

El análisis de estabilidad de laderas es estudiado para evaluar la seguridad en el deslizamiento de una cierta masa de tierra cuando se rompe o pierde el equilibrio de una porción de los materiales que componen una ladera, produciéndose una disminución en la resistencia al esfuerzo cortante debido a que se alcanza la tensión máxima en todos los puntos sobre una superficie de deslizamiento, (Juárez y Rico, 1975).

La ocurrencia de deslizamientos de laderas es un fenómeno sujeto a incertidumbre. Debido a que las grandes masas de tierra han ocasionado pérdidas humanas y materiales en el Estado de Guerrero y en el país, se tiene el interés de entender parte de este fenómeno, mediante la determinación de una superficie de deslizamiento y el uso de un modelo matemático propuesto por ecuaciones diferenciales ordinarias.

De acuerdo a la experiencia, se ha observado que las superficies de deslizamiento siguen aproximadamente el arco de un círculo, la mayoría de las técnicas analíticas comúnmente utilizadas implican el supuesto de un arco circular de falla. Estas técnicas involucran las condiciones físicas en las que se encuentra el suelo, por lo que se requiere la forma que adopta la superficie de deslizamiento.

Además, la necesidad de considerar esfuerzos en suelos y una estratificación en la estructura interna de laderas, implica la utilización de métodos bien fundamentados para realizar el análisis respectivo. En este capítulo, un método de análisis que es de nuestro interés es presentado mediante un modelo determinístico, el cual toma en cuenta las propiedades mecánicas del suelo para determinar los esfuerzos que actúan sobre la superficie de deslizamiento. Trabajos relativos al estudio de superficies de falla, han sido desarrollados por Morgenstern y Price (1965).

3.2. Métodos existentes para determinar una superficie de deslizamiento

En el análisis de la estabilidad de laderas o taludes, la localización de la superficie de falla crítica de una pendiente del suelo es difícil de determinar (para una ilustración de esta superficie, ver la Figura 3.1). Trabajos relativos a la localización de esta superficie de falla fue realizada por Fellenius (1936). En el método que propone Fellenius, el centro de la superficie crítica, O , es supuesta como la intersección de dos líneas establecidas fuera de la base y la parte superior de la pendiente. Un punto P , se fija a una altura $2H$ (siendo H la altura de la ladera) por debajo de la parte superior de la pendiente, y $4.5H$ horizontalmente desde el pie de la ladera. De acuerdo a Fellenius, el centro de la superficie crítica de falla se encuentra a lo largo de una línea que une los puntos P y O , y es obtenida a prueba y error. Este procedimiento es válido únicamente para suelos cohesivos.

Sin embargo, otro método de mucha popularidad es el método de la “malla”, en el que la localización del centro de la superficie crítica es usualmente encontrado por una malla cuadrada de 3×3 (9 puntos). La ubicación del centro de la malla se desplaza durante la búsqueda, hasta que su centro corresponde al centro de la superficie de deslizamiento. El espaciado de los puntos en la malla es reducido durante el proceso de la búsqueda del centro de la superficie. El uso de este método resulta muy tedioso y no garantiza exactamente la superficie crítica.

Con la llegada de PC's de mayor capacidad en memoria, se desarrollaron técnicas

que han sido muy efectivas para la búsqueda del centro de la superficie crítica en el análisis de taludes y laderas. En el artículo Greco (1996), se presentan técnicas basadas en métodos Monte Carlo tipo caminatas aleatorias para localizar la superficie crítica. Las soluciones son aleatoriamente generadas y comparadas con la mejor solución, sin embargo la implementación de este método en una búsqueda automática requiere demasiadas restricciones.

Malkawi et al. (2001), también desarrolló una aproximación para localizar la superficie crítica basada en técnicas Montecarlo (caminatas aleatorias y saltos al azar), éstas son simplemente estructuradas como técnicas de optimización.

En este trabajo, se propone el modelo de Kötter, que está compuesto por dos ecuaciones diferenciales, las cuales son usadas para analizar los esfuerzos tangenciales y normales, en la estabilidad de un talud o ladera a lo largo de una superficie de deslizamiento, (Juárez y Rico, 1975). Se estudia este modelo como un problema inverso bajo un enfoque Bayesiano, con la finalidad de estimar el radio del círculo de deslizamiento para algunos tipos de suelo (como ejemplo se analizan 3 tipos).

3.3. Estimación

El modelo a estudiar, como fue definido en (1.1) y (1.2), está compuesto por:

1) Sistema dinámico:

$$\frac{d\tau}{d\beta} = 2\tau \tan \phi - \gamma R \sin \phi \sin(\beta - \phi), \quad (3.1)$$

$$\frac{d\varphi}{d\beta} = 2c + 2\varphi \tan \phi - \gamma R \cos \phi \sin(\beta - \phi), \quad (3.2)$$

donde:

τ = Esfuerzo tangencial que actúa a lo largo de la superficie de deslizamiento,

φ = Esfuerzo normal que actúa a lo largo de la superficie de deslizamiento,

β = Ángulo que determina la posición del elemento en estudio sobre el arco circular, con respecto a la vertical,

ϕ = Ángulo de fricción interna del suelo,

γ = Peso específico del suelo,

R = Radio de la superficie circular de deslizamiento,

c = Cohesión aparente del suelo.

Una deducción detallada para éstas ecuaciones son presentadas en el Anexo 1. En este capítulo, se hace una modificación en la notación dada en tal anexo, las variables $\tilde{\tau}$ y $\tilde{\sigma}$ son reemplazadas por τ y φ , respectivamente.

En la siguiente figura, se ilustra una interpretación para el modelo de Kotter, en la que se aprecian los esfuerzos actuantes sobre la superficie de deslizamiento.

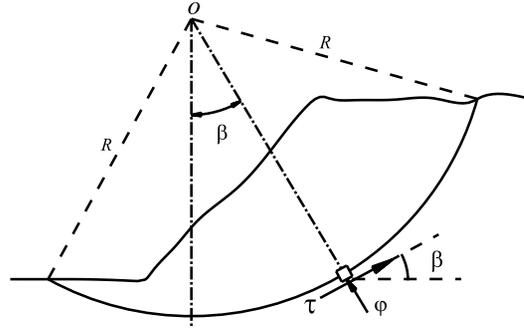


Figura 3.1: Interpretación de las ecuaciones de Kotter.

2) Ecuación de observación:

$$\mathbf{y}_i = \mathbf{h}_i + \mathbf{e}_i, \quad i = 1, \dots, n, \quad (3.3)$$

donde,

$$\mathbf{y}_i = \begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix}, \quad \mathbf{h}_i = \begin{bmatrix} \tau(\beta_i) \\ \varphi(\beta_i) \end{bmatrix}, \quad \mathbf{e}_i = \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix}; \quad \varepsilon_{1i}, \varepsilon_{2i} \sim_{i.i.d.} \mathcal{N}(0, \sigma^2).$$

Las cantidades de interés para el modelo de Kotter son: $\boldsymbol{\theta} = (R, \sigma^2)$.

La función de verosimilitud como se definió en (2.2), para las observaciones \mathbf{y}_1 está dada por:

$$\begin{aligned} L_1(\mathbf{y}_1 | \boldsymbol{\theta}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} (y_{1i} - \tau(\beta_i))^2 \right\}, \\ &= \sigma^{-n} (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{1i} - \tau(\beta_i))^2 \right\}, \end{aligned}$$

donde,

$$\tau(\beta_i) = \tau_0 e^{2\beta_i \tan \phi} - \frac{R\gamma(\cos \beta_i - 3 \cos(\beta_i - 2\phi)) \sin(2\phi)}{2(5 - 3 \cos(2\phi))}.$$

Para incorporar el conocimiento disponible acerca de $\boldsymbol{\theta}$, se asignan las siguientes distribuciones a priori no informativas:

$$\pi(R) \sim \mathcal{G}(r, \lambda), \quad \pi(\tau_*) \sim \mathcal{G}(\eta, \delta),$$

siendo $\tau_* = 1/\sigma^2$ una parametrización de la varianza.

Por lo que, la distribución posterior $\pi_1(\boldsymbol{\theta}|\mathbf{y}_1) \propto L_1(\mathbf{y}_1|\boldsymbol{\theta})p(\boldsymbol{\theta})$ queda definida como:

$$\begin{aligned} \pi_1(\boldsymbol{\theta}|\mathbf{y}_1) &\propto \sigma^{-n} (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{1i} - \tau(\beta_i))^2 \right\} \times \frac{\lambda^r}{\Gamma(r)} R^{r-1} e^{-\lambda R} \times \\ &\quad \frac{\delta^\eta}{\Gamma(\eta)} \tau_*^{\eta-1} e^{-\eta\tau_*} \\ &\propto \exp - \left\{ \frac{\tau_*}{2} \sum_{i=1}^n (y_{1i} - \tau(\beta_i))^2 + \lambda R + \eta\tau_* \right\} \times R^{r-1} \times \tau_*^{\eta-1}. \end{aligned}$$

Análogamente, para las observaciones \mathbf{y}_2 se tiene:

$$L_2(\mathbf{y}_2|\boldsymbol{\theta}) = \sigma^{-n} (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{2i} - \varphi(\beta_i))^2 \right\},$$

con

$$\begin{aligned} \varphi(\beta_i) &= \varphi_0 e^{2\beta_i \tan \phi} + \frac{\cos \phi (R\gamma \cos \beta_i (\cos \phi + 3 \cos 3\phi))}{2(5 - 3 \cos 2\phi)} \\ &\quad - \frac{4 \cos \phi (c \cos \phi \cot \phi + 4c \sin \phi - 3R\gamma \cos^2 \phi \sin \beta_i \sin \phi)}{2(5 - 3 \cos 2\phi)}, \end{aligned}$$

la distribución posterior $\pi_2(\boldsymbol{\theta}|\mathbf{y}_2)$ resulta:

$$\begin{aligned} \pi_2(\boldsymbol{\theta}|\mathbf{y}_2) &\propto \sigma^{-n} (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{2i} - \varphi(\beta_i))^2 \right\} \times \frac{\lambda^r}{\Gamma(r)} R^{r-1} e^{-\lambda R} \times \\ &\quad \frac{\delta^\eta}{\Gamma(\eta)} \tau_*^{\eta-1} e^{-\eta\tau_*} \\ &\propto \exp - \left\{ \frac{\tau_*}{2} \sum_{i=1}^n (y_{2i} - \varphi(\beta_i))^2 + \lambda R + \eta\tau_* \right\} \times R^{r-1} \times \tau_*^{\eta-1}. \end{aligned}$$

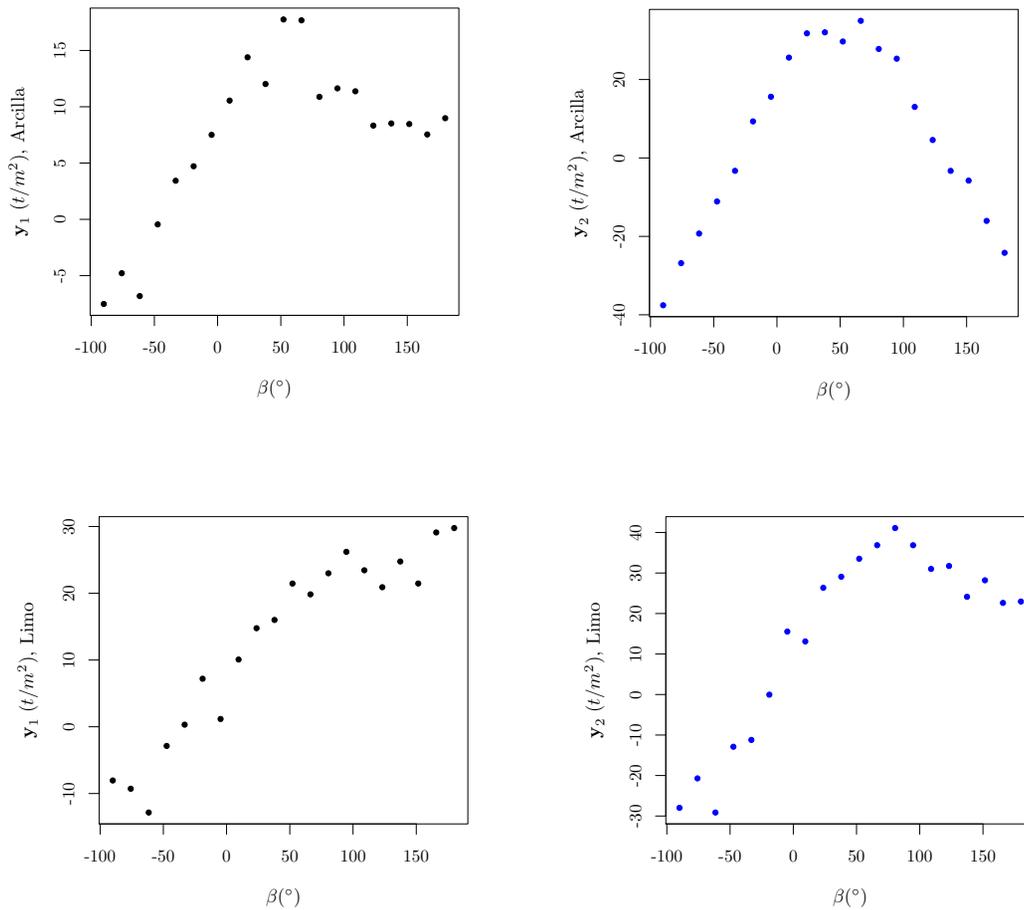
Un método MCMC será empleado usando JAGS (Plummer, 2012) en R (R Core Team, 2014) para determinar $\pi_1(\boldsymbol{\theta}|\mathbf{y}_1)$ y $\pi_2(\boldsymbol{\theta}|\mathbf{y}_2)$.

Cuadro 3.1: *Propiedades mecánicas para 3 tipos de suelo.*

Tipo de suelo	$\gamma(t/m^3)$	$c(t/m^2)$	$\phi(^{\circ})$
Arcilla	1.54	2.5	16
Limo	1.70	1.2	25
Arena	1.92	0.01	32

Para la solución al problema inverso en el modelo de Kotter, se presenta un ejemplo con datos sintéticos, mediante la simulación $n = 20$ datos usando (3.3) para distintos valores de γ , c y ϕ (propiedades mecánicas del suelo) mostrados en la Tabla 3.1.

Los datos simulados se muestran en la Figura 3.2, en donde para distintos ángulos β , se tienen observaciones \mathbf{y}_1 y \mathbf{y}_2 para los esfuerzos tangencial y normal, respectivamente.



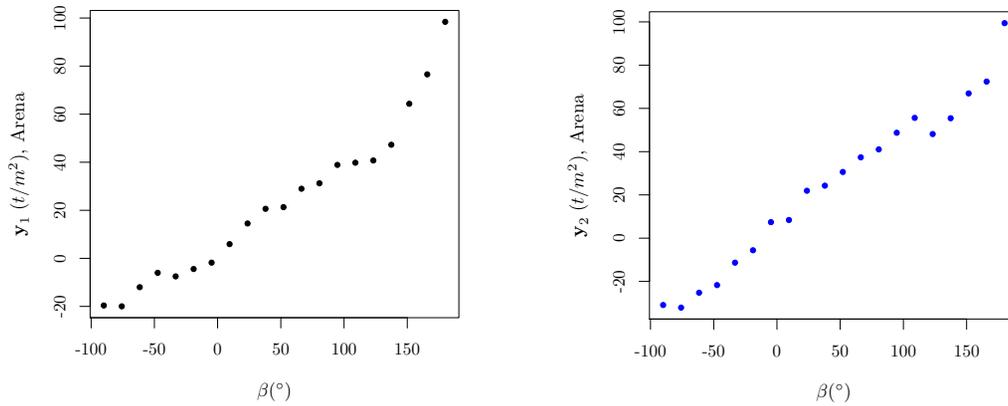


Figura 3.2: *Conjunto de datos simulados de (3.3) para los 3 tipos de suelo.*

En la figura de arriba, se ilustran datos simulados de los valores en los esfuerzos para los tres tipos de suelo utilizados. Se nota una gran diferencia en cada caso, esto debido a que cada tipo de suelo tiene una estructura interna de acuerdo a sus características mecánicas.

Para el análisis Bayesiano en el modelo de Kotter, se utilizan dos Cadenas de Markov con un total de 20,000 iteraciones, considerando 2,000 de éstas como burn-in en cada una. Además, considerando una varianza en los datos de 1×10^3 ; esto por la razón de distinguir la variabilidad en los datos.

Assumiendo, una distribución inicial no informativa para R (radio de la superficie de deslizamiento) denotada por:

$$\pi(R) \sim \mathcal{G}(0.9, 0.03),$$

$\mathcal{G}(\cdot)$ es la distribución Gamma. La razón de los hiperparámetros asignados, es que el radio, R , usado para la generación de datos fue de 30 metros, para esto, se utilizó el software estadístico R, (R Core Team, 2014).

La forma de la distribución posterior para el parámetro R en los tres tipos de suelo, es similar, por lo que sólo se mostrará para la arcilla, así como unos de los diagnósticos visuales de convergencia (traza de la Cadena de Markov):

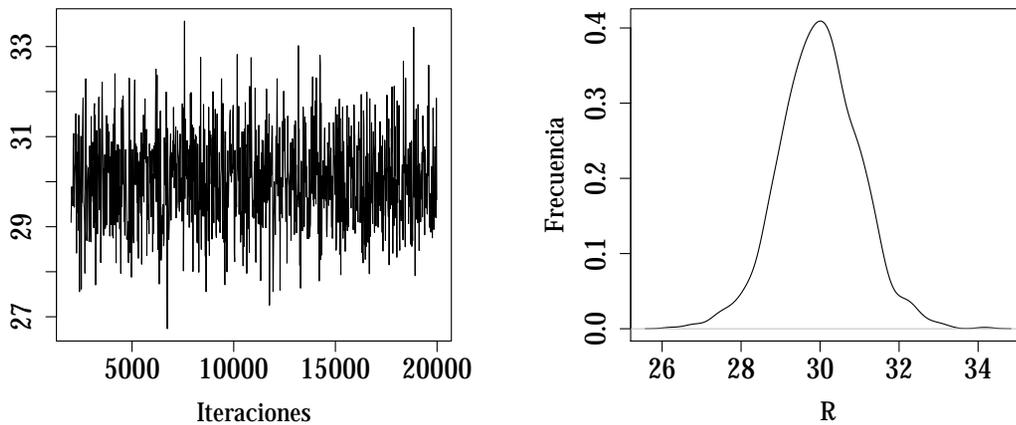


Figura 3.3: *Traza y distribución posterior para el parámetro R (circulo de deslizamiento).*

La traza indica que las Cadenas de Markov empleadas muestran buena convergencia durante el monitoreo y el uso del programa JAGS. Se observa la trayectoria de la Cadena para el parámetro de interés después de 2,000 iteraciones, para las dos Cadenas paralelas, así como la distribución estacionaria de éstas. Algunos de los diagnósticos estadísticos de convergencia descritos en la sección 2.3.1 se ilustran a continuación:

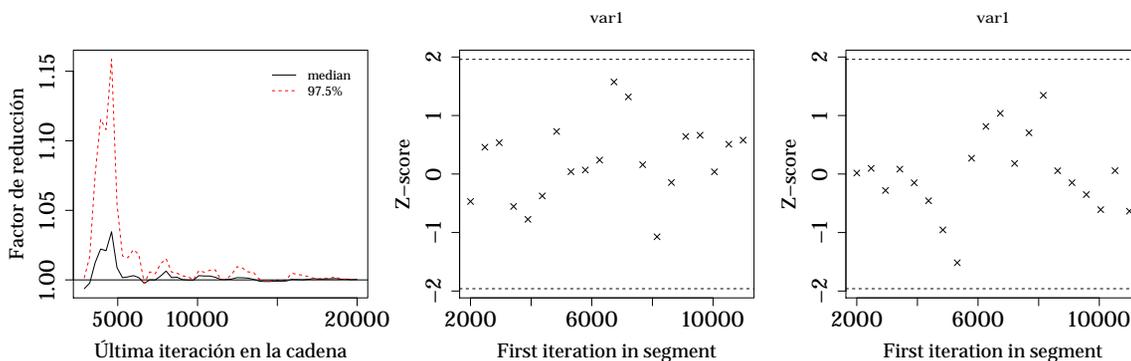


Figura 3.4: *Diagnóstico de convergencia de Gelman y Rubin; y de Geweke para el parámetro R (Cadena 1 y 2), respectivamente.*

En la Tabla 3.2, se muestran las estimaciones del radio de la superficie de deslizamiento para la arcilla, así también para los tipos de suelo restantes antes mencionados.

Cuadro 3.2: *Valores de R estimado para los 3 tipos de suelo.*

Tipo de suelo	\widehat{R}
Arcilla	29.911 m.
Limo	30.014 m.
Arena	29.983 m.

3.4. Conclusión

La finalidad de este ejemplo, es ilustrar la posibilidad de estimar una superficie de deslizamiento en laderas en presencia de datos tales como esfuerzo tangencial y normal; y con la utilización de un modelo determinístico que incluye propiedades mecánicas de suelos, como son las ecuaciones de Kotter.

Cabe señalar que este resultado obtenido es aplicable cuando las laderas están compuestas por un suelo homogéneo y cuando el deslizamiento se debe a una falla rotacional. El uso de métodos MCMC con ayuda del programa JAGS es justificado, debido a que la integración analítica de la distribución final para los parámetros es intratable.

Problema inverso Bayesiano para un modelo de crecimiento poblacional fraccionario

En distintas ramas de la ciencia y en muchas aplicaciones, existen fenómenos que presentan un comportamiento distinto a las leyes físicas que los gobiernan; debido a ésto, la dinámica en los problemas directos cambia dramáticamente. Las ecuaciones diferenciales fraccionarias son adecuadas para explicar este tipo de fenómenos. En este capítulo, se estudia un modelo relacionado al crecimiento exponencial fraccionario, ajustándolo a un conjunto de datos que presentan un crecimiento anómalo. Así también, una selección de modelos con derivada entera y fraccionaria que mejor se ajusta al conjunto de datos.

4.1. Modelo

El modelo matemático a considerar, es como se definió en (1.1) y (1.2) y está compuesto por:

1. El sistema dinámico

$$\begin{cases} \mathcal{D}^\alpha X(t) = aX(t) + f(t), & a > 0, \quad \alpha \in (0, 1]; \\ X(t_0) = X_0, \end{cases} \quad (4.2)$$

donde $X(t)$ es el tamaño de la población al tiempo t , X_0 es el tamaño inicial, a es la tasa de cambio y $f(t)$ es una función que adiciona un comportamiento externo

al sistema en (4.2). Aquí, \mathcal{D}^α es la derivada ordinaria para $\alpha = 1$, y la derivada fraccionaria de Caputo para $\alpha \in (0, 1)$, la cual está definida por

$$\mathcal{D}^\alpha X(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{X'(\tau)}{(t-\tau)^\alpha} d\tau,$$

donde $\Gamma(\cdot)$ es la función Gama.

2. Ecuación de observación:

$$y_i = X(t_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (4.3)$$

donde ε_i son los errores de medición, denotados por $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

El correspondiente vector de parámetros en este modelo es $\boldsymbol{\theta} = (a, \alpha, \sigma^2, X_0)$.

Para poder realizar un proceso de observaciones, se necesita la solución del sistema dinámico dado en (4.2), la cual es desarrollada en la siguiente sección.

4.2. Solución para el sistema dinámico

En esta sección, se da una solución explícita para (4.2). La transformada de Laplace y su inversa están definidas por

$$\widehat{X}(s) = \int_0^\infty e^{-st} X(t) dt, \quad X(t) = \frac{1}{2\pi i} \int_{s_0-i\infty}^{s_0+i\infty} e^{st} \widehat{X}(s) ds,$$

respectivamente. Aplicando la transformada directa a (4.2), se obtiene

$$\widehat{X}(s) = \frac{1}{s^\alpha - a} (s^{\alpha-1} X_0 + \widehat{f}(s)). \quad (4.4)$$

Ahora, usando la siguiente propiedad:

$$[t^{\beta-1} E_{\alpha,\beta}(at^\alpha)]^\wedge(s) = \frac{s^{\alpha-\beta}}{(s^\alpha - a)},$$

donde $E_{\alpha,\beta}$ es la función Mittag-Leffler definida por

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)}, \quad \alpha > 0, \quad \beta > 0, \quad z \in \mathbb{C}, \quad (4.5)$$

y aplicando la transformada de Laplace inversa a (4.4) se llega a

$$X(t) = X_0 E_{\alpha,1}(at^\alpha) + \int_0^t (t-\tau)^{\alpha-1} E_{\alpha,\alpha}(a(t-\tau)^\alpha) f(\tau) d\tau. \quad (4.6)$$

Notando que, si $\alpha = 1$ en esta ecuación, la solución bien conocida para (4.2), está dada por

$$X(t) = X_0 e^{at} + \int_0^t e^{a(t-\tau)} f(\tau) d\tau, \quad (4.7)$$

en la que se hace referencia a un crecimiento exponencial.

Para el caso cuando $\alpha \in (0, 1)$, (4.6) se refiere a un tipo de crecimiento anómalo.

4.3. Selección de un modelo de crecimiento poblacional fraccionario

Método para la selección del modelo.

La selección del modelo Bayesiano usa las reglas de la teoría de probabilidad para seleccionar entre diferentes modelos, una manera de realizar esto, es proponer modelos para datos y observar que modelo se ajusta mejor. En estadística Bayesiana, un método común utilizado para seleccionar entre varios modelos es mediante el factor de Bayes (Kass y Raftery, 1995). Suponiendo que hay \mathcal{K} modelos, $\mathcal{M}_1, \dots, \mathcal{M}_{\mathcal{K}}$. Para el modelo \mathcal{M}_k , la distribución posterior toma la forma

$$\pi(\boldsymbol{\theta}_k | \mathbf{y}, \mathcal{M}_k) \propto L(\mathbf{y} | \boldsymbol{\theta}_k, \mathcal{M}_k) p(\boldsymbol{\theta}_k | \mathcal{M}_k), \quad (4.8)$$

donde $L(\mathbf{y} | \boldsymbol{\theta}_k, \mathcal{M}_k)$ es la función de densidad de los datos $\mathbf{y} = (y_1, \dots, y_n)^t$ bajo el modelo \mathcal{M}_k , y $p(\boldsymbol{\theta}_k | \mathcal{M}_k)$ es la distribución inicial de $\boldsymbol{\theta}_k$. Para comparar diferentes modelos, se calcula la ML $m(\mathbf{y} | \mathcal{M}_k)$ para $k = 1, 2, \dots, \mathcal{K}$, dada por

$$m(\mathbf{y} | \mathcal{M}_k) = \int_{\Theta_k} L(\mathbf{y} | \boldsymbol{\theta}_k, \mathcal{M}_k) p(\boldsymbol{\theta}_k | \mathcal{M}_k) d\boldsymbol{\theta}_k. \quad (4.9)$$

Notando que ML en la ecuación de arriba es la constante de normalización de la distribución posterior, así, ésta puede ser escrita como

$$m(\mathbf{y} | \mathcal{M}_k) = \frac{L(\mathbf{y} | \boldsymbol{\theta}_k, \mathcal{M}_k) p(\boldsymbol{\theta}_k | \mathcal{M}_k)}{\pi(\boldsymbol{\theta}_k | \mathbf{y}, \mathcal{M}_k)}, \quad (4.10)$$

donde el numerador es sólo el producto de la densidad de muestreo y la distribución inicial, incluídas todas las constantes de integración; y el denominador es la densidad posterior de $\boldsymbol{\theta}_k$. Haciendo mención de que la aproximación de la ML es esencialmente la misma como el enfoque del factor de Bayes, la cual puede ser vista a través de la siguiente identidad para la comparación de modelos \mathcal{M}_i y \mathcal{M}_j :

$$FB_{i,j} = \exp \{ \ln[m(\mathbf{y}|\mathcal{M}_i)] - \ln[m(\mathbf{y}|\mathcal{M}_j)] \}, \quad (4.11)$$

donde $FB_{i,j}$ es el factor de Bayes.

En general, ML en (4.9) no puede ser representada en términos de funciones elementales, y para obtener una aproximación numérica usualmente se requiere de un gran esfuerzo computacional. Por otro lado, para calcular ML utilizando (4.10) se necesita la distribución posterior para cada parámetro en el modelo k . Se usará una aproximación de ML mediante el estimador de Gelfand y Dey (Chib, 1995), dado por

$$\hat{m}_{GD} = \left\{ \frac{1}{G} \sum_{g=1}^G \left(\frac{\pi(\boldsymbol{\theta}_k^{(g)})}{L(\mathbf{y}|\boldsymbol{\theta}_k^{(g)}, \mathcal{M}_k) p(\boldsymbol{\theta}_k^{(g)}|\mathcal{M}_k)} \right) \right\}^{-1}. \quad (4.12)$$

Sea $\{\boldsymbol{\theta}_k^{(g)}\} \equiv \{\boldsymbol{\theta}_k^{(1)}, \dots, \boldsymbol{\theta}_k^{(G)}\}$ G muestras de la distribución posterior $\pi(\boldsymbol{\theta}_k|\mathbf{y}, \mathcal{M}_k)$ obtenidas usando un método MCMC. $\pi(\boldsymbol{\theta}_k)$ es una densidad con cola más delgada que el producto de la distribución inicial y la verosimilitud. Se pueden hacer selecciones obvias para escoger $\pi(\cdot)$, una densidad Normal o una densidad t con media y varianza igual a la media y varianza posterior.

En este trabajo, el interés es comparar dos modelos, el modelo FEG (\mathcal{M}_2) vs el modelo OEG (\mathcal{M}_1). El modelo \mathcal{M}_1 está definido como:

$$\begin{aligned} \frac{d}{dt}X(t) &= aX(t), & a > 0; \\ X(t_0) &= X_0, \end{aligned}$$

y la correspondiente ecuación de observación, para n observaciones, es:

$$y_i^{(1)} = X_0 e^{at_i} + \varepsilon_i, \quad i = 1, \dots, n. \quad (4.13)$$

El modelo \mathcal{M}_2 está definido como:

$$\begin{aligned} \mathcal{D}^\alpha X(t) &= aX(t), & a > 0, \quad \alpha \in (0, 1); \\ X(t_0) &= X_0, \end{aligned}$$

y

$$y_i^{(2)} = X_0 E_{\alpha,1}(at_i^\alpha) + \varepsilon_i, \quad i = 1, \dots, n, \quad (4.14)$$

donde $E_{\alpha,1}(\cdot)$ está definida en (4.5). Notando que para ambos modelos la función fuente $f(t)$ es supuesta por ser cero, y la función h es tomada como la función identidad. Se desea estimar desde un punto de vista Bayesiano los parámetros involucrados en los modelos \mathcal{M}_1 y \mathcal{M}_2 .

Para el modelo OEG, el vector de parámetros es $\boldsymbol{\theta}_1 = (a, \sigma^2, X_0)$ y para el modelo FEG es $\boldsymbol{\theta}_2 = (a, \alpha, \sigma^2, X_0)$. También, se hará una estimación del factor de Bayes $\widehat{FB}_{2,1}$, como en (4.11), vía aproximación de la verosimilitud marginal.

Suponiendo independencia inicial de los parámetros, se escribe la distribución inicial conjunta como $p(\boldsymbol{\theta}) = p(a)p(\alpha)p(\sigma^2)p(X_0)$.

En la sección 4.3.1, se propone cada una de estas distribuciones iniciales, las cuales no son conjugadas para la verosimilitud observada. Así, la distribución posterior es analíticamente intratable y las distribuciones posteriores marginales de los parámetros son complicadas; Sin embargo, se pueden obtener muestras de (2.4) utilizando técnicas MCMC.

Uno de los objetivos de esta sección, es identificar el tipo de crecimiento relacionado a un conjunto de datos dado, es decir, seleccionar el modelo que mejor ajusta a los datos. En la siguiente sección, se realiza esta tarea mediante datos simulados, así como un ejemplo aplicado usando datos reales.

Para estimar los parámetros mencionados, se hace uso del teorema de Bayes, para esto, se debe definir la distribución posterior dada en (4.8) para cada modelo; para el modelo (OEG) y las distribuciones a priori,

$$a \sim \mathcal{G}(r, \lambda), \quad \tau_* \sim \mathcal{G}(\eta, \delta), \quad X_0 \sim \mathcal{N}(\mu_*, \sigma_*^2), \quad (4.15)$$

denotando a τ_* (precisión) como una parametrización de la varianza, ($\tau_* = 1/\sigma^2$), la

distribución posterior tiene la siguiente forma:

$$\begin{aligned}
\pi(\boldsymbol{\theta}_1|\mathbf{y}, \mathcal{M}_1) &\propto \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_i - X_0 e^{at_i})^2}{2\sigma^2}\right\} \right) \times \frac{\lambda^r}{\Gamma(r)} a^{r-1} e^{-\lambda a} \times \frac{\delta^\eta}{\Gamma(\eta)} \tau_*^{\eta-1} e^{-\eta\tau_*} \\
&\quad \times \frac{1}{\sqrt{2\pi}\sigma_*} \exp\left\{-\frac{(X_0 - \mu_*)^2}{2\sigma_*^2}\right\} \\
&\propto \tau_*^{n/2} \exp - \left\{ \frac{\tau_*}{2} \sum_{i=1}^n (y_i - X_0 e^{at_i})^2 + \lambda a + \eta\tau_* + \frac{(X_0 - \mu_*)^2}{2\sigma_*^2} \right\} \times a^{r-1} \tau_*^{\eta-1}.
\end{aligned} \tag{4.16}$$

De igual manera, para el modelo (FEG) y adicionando la distribución a priori para el parámetro α , denotada como $\alpha \sim \mathcal{U}(0, 1)$, la distribución posterior es:

$$\begin{aligned}
\pi(\boldsymbol{\theta}_2|\mathbf{y}, \mathcal{M}_2) &\propto \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_i - X_0 E_\alpha(at_i^\alpha))^2}{2\sigma^2}\right\} \right) \times \frac{\lambda^r}{\Gamma(r)} a^{r-1} e^{-\lambda a} \times \frac{\delta^\eta}{\Gamma(\eta)} \tau_*^{\eta-1} e^{-\eta\tau_*} \\
&\quad \times \frac{1}{\sqrt{2\pi}\sigma_*} \exp\left\{-\frac{(X_0 - \mu_*)^2}{2\sigma_*^2}\right\} \times \frac{1}{1-\alpha} \\
&\propto \tau_*^{n/2} \exp - \left\{ \frac{\tau_*}{2} \sum_{i=1}^n (y_i - X_0 E_\alpha(at_i^\alpha))^2 + \lambda a + \eta\tau_* + \frac{(X_0 - \mu_*)^2}{2\sigma_*^2} \right\} \times a^{r-1} \tau_*^{\eta-1}.
\end{aligned} \tag{4.17}$$

Como se puede observar, dichas distribuciones no tienen una forma explícita conocida, para esto, se recurre a métodos MCMC descritos en la sección 2.3 y así lograr el objetivo para la selección del modelo mediante un estudio de simulación.

4.3.1. Estudio de simulación

Se lleva a cabo un estudio de simulación para evaluar el desempeño de los estimadores Bayesianos de los parámetros de interés ($\boldsymbol{\theta}_k$, $k = 1, 2$) y el rendimiento del factor de Bayes utilizando el estimador de Gelfand y Dey (Chib, 1995), para la verosimilitud marginal (4.10). Para esta tarea, se simulan $M=1,000$ conjuntos de datos del modelo FEG con dos tamaños de muestra, $n = 10$ y 30 . Usando JAGS dentro del ambiente de R para este estudio, bajo el siguiente algoritmo:

- 1) Se simula un conjunto de datos sintéticos usando los valores verdaderos de los parámetros $a = 0.2$, $\alpha = 0.5$, $\sigma^2 = 30$ y $X_0 = 20$.

- II) Se ajustan los modelos \mathcal{M}_1 y \mathcal{M}_2 para el conjunto de datos bajo dos casos: X_0 conocido y desconocido. Las siguientes distribuciones iniciales para los parámetros en los modelos son de la siguiente manera:

$$a \sim \mathcal{G}(1, 1), \quad \alpha \sim \mathcal{U}(0, 1), \quad \tau_* \sim \mathcal{G}(9 \times 10^{-4}, 3 \times 10^{-2}), \quad X_0 \sim \mathcal{N}(20, 1),$$

donde $\mathcal{G}(r, \lambda)$ denota la distribución Gamma con parámetro de forma r y parámetro de escala λ , \mathcal{U} es la distribución uniforme continua en el intervalo $(0,1)$. Los hiperparámetros, se dan para conveniencia numérica del experimento. Mediante el uso de métodos MCMC, y usando una función de pérdida cuadrática, una estimación puntual de $\boldsymbol{\theta}$ está dada por $\widehat{\boldsymbol{\theta}} = \frac{1}{M} \sum_{j=1}^M \boldsymbol{\theta}^{(j)}$, donde $\boldsymbol{\theta}^{(j)}$, $j = 1, \dots, M$ es una muestra de la distribución posterior en (4.8) de tamaño M .

Utilizando dos Cadenas, cada una con 5,000 iteraciones con un calentamiento (burn-in) de 1,000 iteraciones y una tasa de adelgazamiento de 4, así, se guarda un total de 2,000 iteraciones para hacer inferencias acerca de los parámetros de interés, $\boldsymbol{\theta}_k$, $k = 1, 2$.

- III) Se calculan los intervalos de credibilidad del 95% para cada parámetro en cada modelo; estos intervalos están basados en el cuantil 0.025 y 0.975 de las correspondientes muestras de la distribución posterior.
- IV) Se obtiene el Criterio de Información de la Desviación (DIC¹) para ambos modelos.
- V) Se calcula ML utilizando el estimador de Gelfand y Dey dado en (4.12) usando las muestras MCMC de la densidad posterior $p(\boldsymbol{\theta}_k | \mathbf{y}, \mathcal{M}_k)$, $k = 1, 2$.
- VI) Se obtiene una estimación del factor de Bayes de \mathcal{M}_2 *v.s* \mathcal{M}_1 , denotado por $\widehat{FB}_{2,1}$. Si $\widehat{FB}_{2,1} \geq 1$, entonces se prefiere el modelo \mathcal{M}_2 respecto al modelo \mathcal{M}_1 .

Se repite $M=1,000$ veces los pasos del (I)-(VI) para determinar las propiedades de los estimadores de los parámetros, tales como el Sesgo,

$$\widehat{S}_{\boldsymbol{\theta}_k} = \frac{1}{M} \sum_{j=1}^M \widehat{\boldsymbol{\theta}}_{j,k} - \boldsymbol{\theta}_k^*,$$

¹DIC es una estimación del error predictivo esperado (entre más bajo sea el DIC, es mejor)

El Error Cuadrático Medio (ECM),

$$\text{ECM}_k = \widehat{V}(\widehat{\boldsymbol{\theta}}_k) + (\widehat{S}_{\boldsymbol{\theta}_k})^2,$$

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

Este estudio de simulación es presentado en el Anexo 2, mediante una implementación en R usando paquetes de JAGS.

4.3.2. Resultados

En esta sección, se presentan los resultados del estudio de simulación y algunas conclusiones acerca de este trabajo. En las Tablas 4.1 y 4.2, se presenta un resumen estadístico del Sesgo estimado, el ECM y el porcentaje de Cobertura para ambos modelos, cuando X_0 es conocida y cuando no se conoce, respectivamente, considerando dos tamaños de muestra ($n = 10, 30$) para cada caso. Este procedimiento fue aplicado a los datos del modelo \mathcal{M}_2 . En estas Tablas, se observó que en todos los casos, la estimación de las propiedades de los estimadores están a favor de los parámetros del modelo \mathcal{M}_2 , es decir, el Sesgo estimado y el ECM son más bajos para \mathcal{M}_2 respecto a \mathcal{M}_1 , y las Coberturas estimadas son mayores para \mathcal{M}_2 que las de \mathcal{M}_1 . En la Figura 4.1, se muestra un resumen de los DIC's para los casos considerados. Se puede observar que los DIC's para \mathcal{M}_2 son más bajos que los de \mathcal{M}_1 , lo cual indica que \mathcal{M}_2 es el mejor.

Para comparar entre el modelo \mathcal{M}_1 y el \mathcal{M}_2 , se calculó el FB para cada conjunto de datos sintéticos. Para el caso con X_0 conocido, el porcentaje de los factores de Bayes estimados a favor de \mathcal{M}_2 fue de 96.5 % con $n = 10$ observaciones; y del 99.4 % para $n = 30$. Para el caso en que X_0 es desconocido, el porcentaje de los factores de Bayes estimados a favor de \mathcal{M}_2 fue de 96.1 % con $n = 10$ observaciones, y de 99.1 % para $n = 30$. Entonces, se ha obtenido evidencia estadística para concluir que el factor de Bayes es un buen criterio para elegir entre un modelo FEG con el orden de la derivada $\alpha \in (0, 1)$ y un modelo OEG, con $\alpha = 1$, como se definió en (4.2).

Cuadro 4.1: *Sesgo estimado, ECM y porcentaje de Cobertura para los parámetros de los modelos \mathcal{M}_1 y \mathcal{M}_2 , con X_0 conocido.*

$n = 10$						
Parámetros	\mathcal{M}_1			\mathcal{M}_2		
	Sesgo	ECM	Cobertura	Sesgo	ECM	Cobertura
a	-0.105	0.011	0.0 %	0.008	0.002	97.5 %
α	—	—	—	0.015	0.020	97.3 %
σ^2	24.184	1173.738	87.3 %	8.536	404.331	96.2 %
$n = 30$						
a	-0.136	0.018	0.0 %	-0.001	0.0004	90.3 %
α	—	—	—	0.007	0.002	90.4 %
σ^2	79.296	6716.946	0.1 %	2.141	81.007	95.0 %

Cuadro 4.2: *Sesgo estimado, ECM y porcentaje de Cobertura para los parámetros de los modelos \mathcal{M}_1 y \mathcal{M}_2 , con X_0 desconocido.*

$n = 10$						
Parámetros	\mathcal{M}_1			\mathcal{M}_2		
	Sesgo	ECM	Cobertura	Sesgo	ECM	Cobertura
a	-0.088	0.008	0.0 %	0.007	0.002	98.6 %
α	—	—	—	0.019	0.019	97.5 %
σ^2	22.128	1067.233	89.0 %	9.610	459.266	94.5 %
X_0	0.484	0.296	100.0 %	0.047	0.019	100.0 %
$n = 30$						
a	-0.136	0.018	0.0 %	-0.002	0.0005	97.2 %
α	—	—	—	0.011	0.002	95.5 %
σ^2	46.624	2529.658	14.5 %	2.119	79.893	94.8 %
X_0	2.330	5.513	6.0 %	0.072	0.106	100.0 %

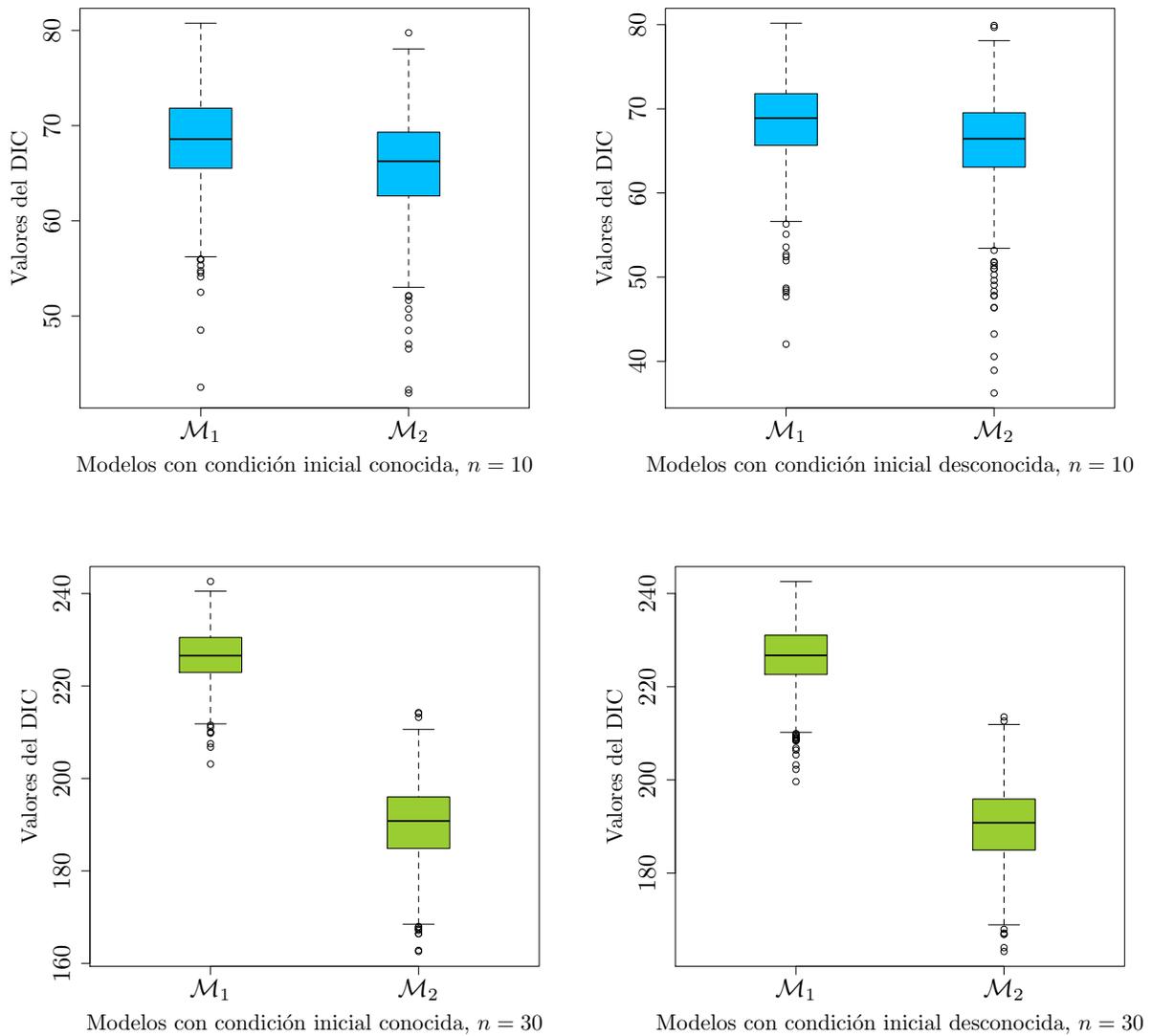


Figura 4.1: Diagrama de Cajas y Bigotes del DIC en ambos modelos para los casos del parámetro X_0 .

4.4. Ejemplos de aplicación

Ejemplo 1.

El siguiente ejemplo muestra datos sobre el crecimiento de pastos, en donde, el objetivo es determinar una curva de crecimiento.

En Biología, las curvas de crecimiento son de gran interés. En Huet (2004), puede

consultarse este ejemplo de aplicación; se observa el rendimiento en el crecimiento de pastos, Q , contra el tiempo, t , desde la última siembra. Los resultados son reportados en la Tabla 4.3 y graficados en la Figura 4.2. El siguiente modelo es supuesto como: el rendimiento al tiempo t , las realizaciones de la variable aleatoria Q , son escritas como en (1.2),

$$y_i = h(Q(t_i)) + \varepsilon_i, \quad i = 1, \dots, n, \quad (4.18)$$

donde $h(Q(t))$ describe la relación entre el rendimiento y el tiempo. Los errores ε , son *v.a.*'s tales que $\varepsilon_i \sim \mathcal{N}(0, 1)$. Debido a que las mediciones están en diferentes unidades experimentales, los errores pueden ser supuestos independientes. Además, suponiendo que la varianza de ε es igual a σ^2 .

Cuadro 4.3: *Datos de rendimiento de pastos v.s. tiempo (las unidades no se mencionan).*

Tiempo (t)	9	14	21	28	42	57	63
Rendimiento (Q)	8.93	10.8	18.59	22.33	39.35	56.11	61.73

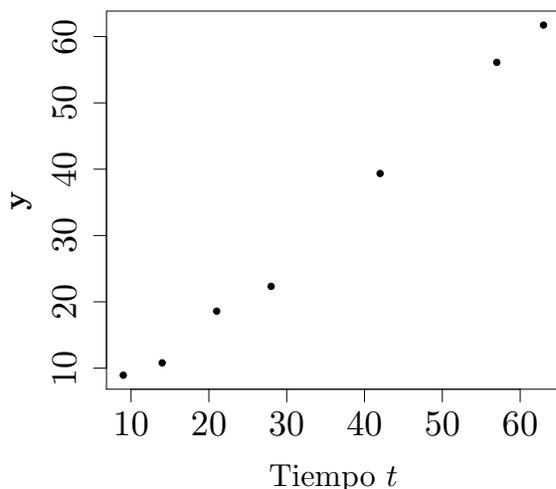


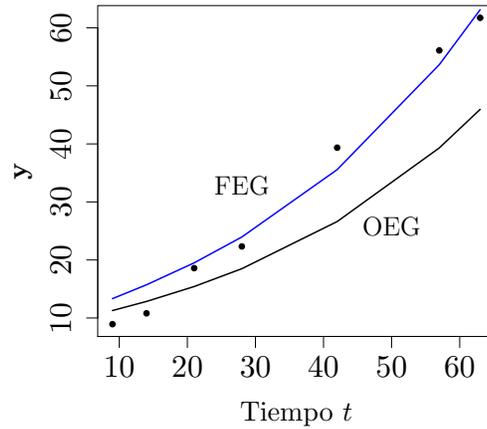
Figura 4.2: *Crecimiento de pastos: Observaciones v.s. tiempo.*

En la Figura 4.2 se observa que los datos presentan un tipo de crecimiento, este puede ser del tipo exponencial, pero también podría tratarse el caso de un crecimiento

anómalo. Para definir que tipo de modelo se ajusta mejor, se realiza una selección de modelos mediante el factor de Bayes (FB). El vector de parámetros para el modelo OEG es $\theta_1 = (a, \sigma^2)$, y para el FEG es $\theta_2 = (a, \alpha, \sigma^2)$. Para este conjunto de datos, el FB resultó de 10.42, el cual está a favor de un modelo FEG.

Considerando a X_0 conocida que corresponde en este caso a $X_0 = 8.93$ al tiempo $t_0 = 9$, las estimaciones de los parámetros para el modelo FEG son:

Parámetros	Valores estimados
a	0.069
α	0.736
σ^2	19.391



La curva ajustada $Q(t; \hat{\theta}_2)$ es mostrada en la Figura 4.3 y es definida como:

$$X(t) = 8.93E_{0.736,1}(0.069t^{0.736})$$

Figura 4.3: Curva ajustada para el crecimiento de pastos para el caso de X_0 conocida.

Las estimaciones para el modelo OEG son:

$$\hat{a} = 0.026, \quad \hat{\sigma}^2 = 26.407,$$

y su correspondiente curva ajustada queda definida como: $X(t) = 8.93e^{0.026t}$.

Se puede notar en la Figura 4.3 que la curva que mejor ajusta al conjunto de datos es la que corresponde al modelo FEG (línea azul), mientras que el modelo OEG (línea negra) tiene deficiencias en el ajuste, por lo que no puede ser considerado como un buen modelo para este tipo de datos que presentan un crecimiento anómalo.

Las distribuciones posteriores para los parámetros a , α y σ^2 se muestran a continuación, así como uno de los diagnósticos visuales de convergencia (traza de la Cadena de Markov):

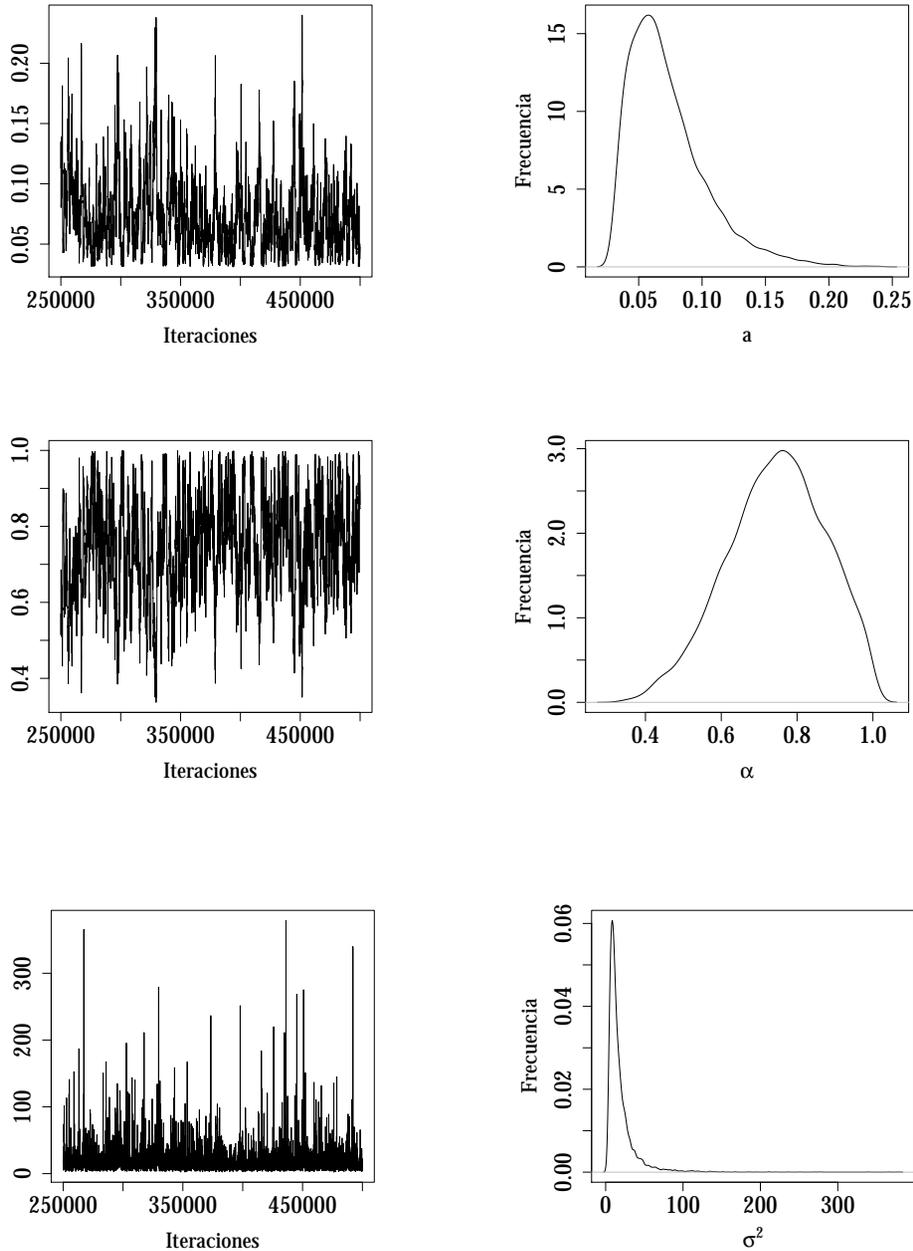


Figura 4.4: *Trazas y distribuciones posteriores para los parámetros a , α y σ^2 , respectivamente.*

Se nota que el centro de cada Cadena parece estar cercano con muy pequeñas fluctuaciones al valor estimado de cada parámetro. Esto indica que cada Cadena alcanzó la convergencia y por lo tanto se obtuvieron las distribuciones posteriores marginales. Algunos de los diagnósticos estadísticos de convergencia descritos en la sección 2.3.1 se ilustran a continuación:

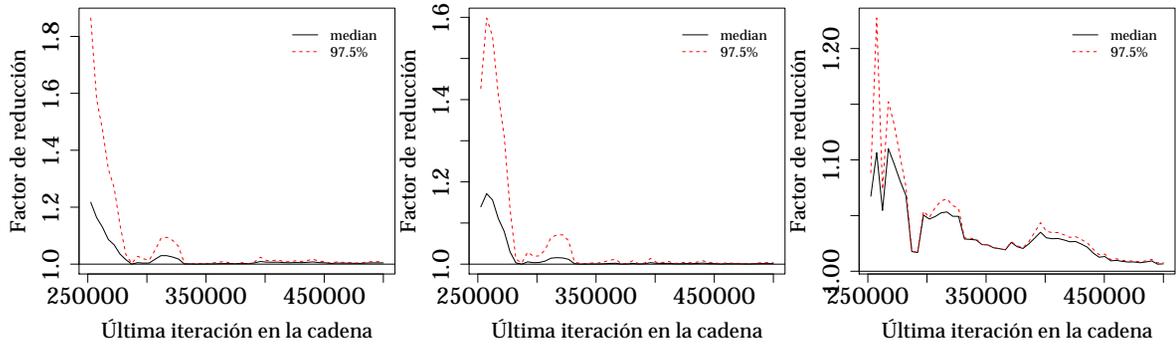


Figura 4.5: *Diagnóstico de convergencia de Gelman y Rubin para los parámetros a , α y σ^2 respectivamente.*

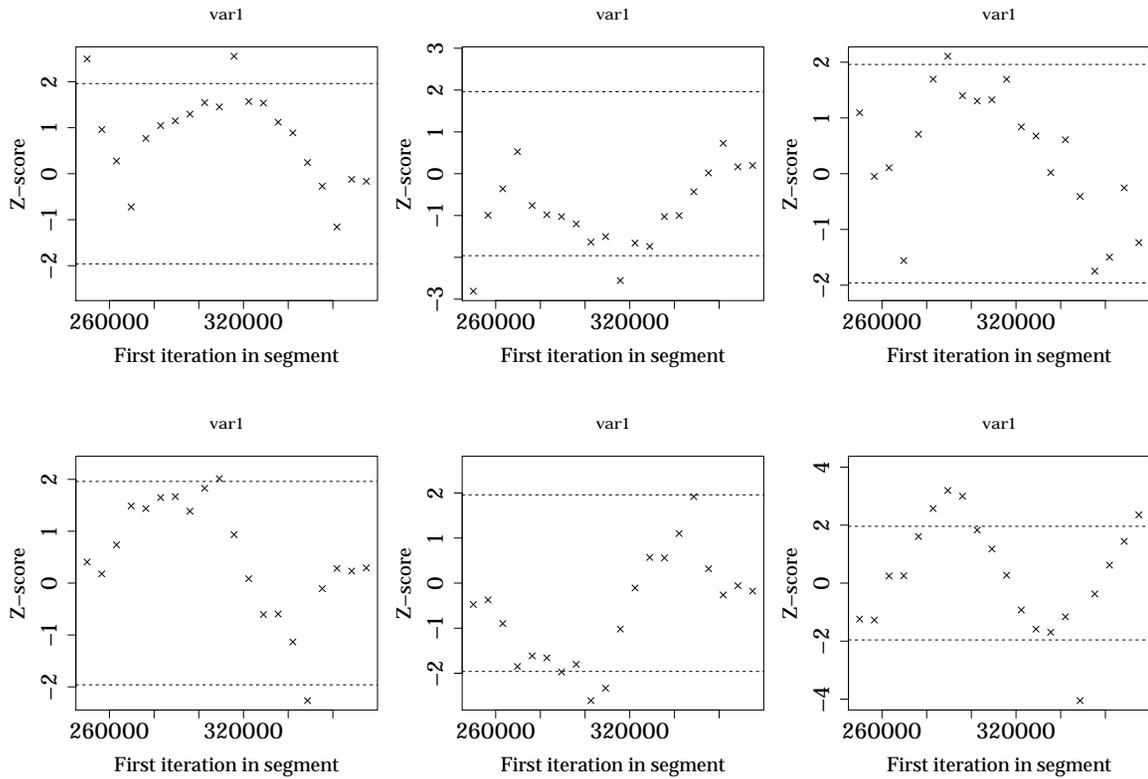


Figura 4.6: *Diagnóstico de convergencia de Geweke para los parámetros a , α y σ^2 en las Cadenas 1 y 2, respectivamente.*

Ahora se considera el caso cuando X_0 es desconocida. Una de las interpretaciones para esta situación, es que dado un conjunto de datos de algún fenómeno que presenta un tipo de crecimiento, se desea estimar el valor inicial con que comenzó la dinámica

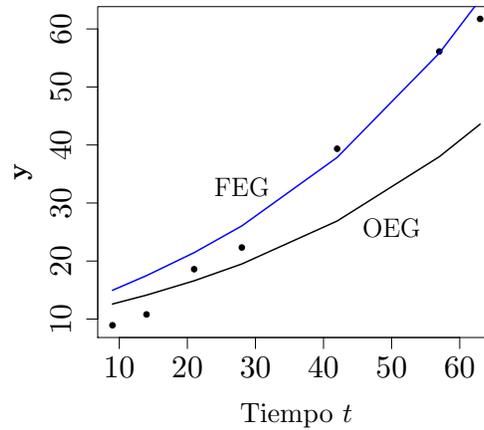
del sistema.

Esto es muy común, sobre todo cuando se estudian ecosistemas en los que existe un cierto número de individuos de los cuales se desconoce la cantidad inicial con que se empezó a poblar el hábitat.

Para este segundo problema inverso, se realizó el mismo mecanismo descrito anteriormente, en el que el FB resultó de 8.42, el cual presenta evidencia sustancial a favor de un modelo FEG.

Las estimaciones de los parámetros para el modelo FEG son:

Parámetros	Valores estimados
a	0.065
α	0.778
σ^2	28.607
X_0	10.236



La curva ajustada $h(t; \hat{\theta})$ es mostrada en la Figura 4.7 y es definida como:

$$X(t) = 10.236E_{0.778,1}(0.065t^{0.778})$$

Figura 4.7: Curva ajustada para el crecimiento de pastos para el caso de X_0 desconocida.

Las estimaciones para el modelo OEG son:

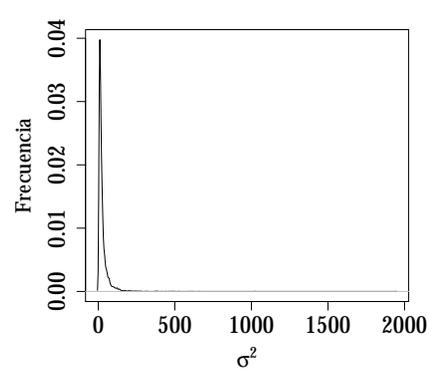
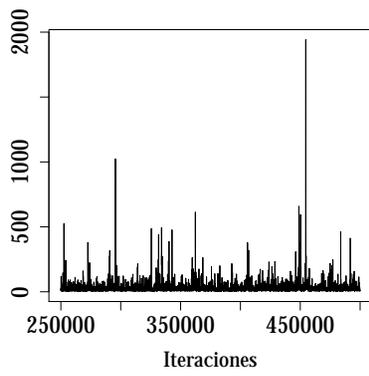
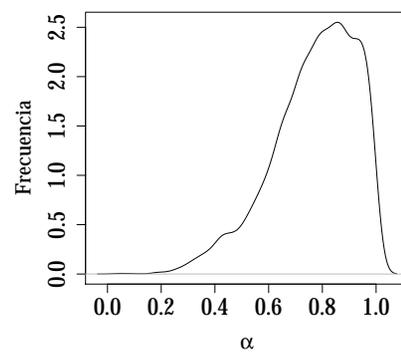
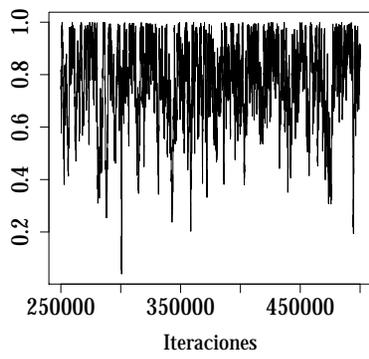
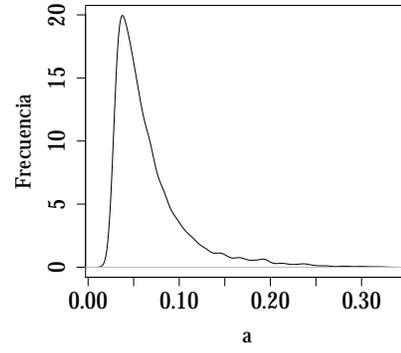
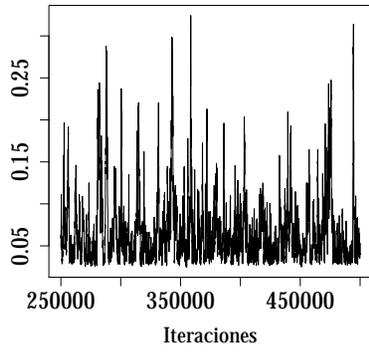
$$\hat{a} = 0.023, \quad \hat{\sigma}^2 = 28.381,$$

y su correspondiente curva ajustada queda definida como: $X(t) = 10.236e^{0.023t}$.

Se puede notar en la Figura 4.7, que la curva que mejor ajusta al conjunto de datos nuevamente es la que corresponde al modelo FEG (línea azul). El valor X_0 resulta muy próximo al verdadero.

Las distribuciones posteriores para los parámetros a , α , σ^2 y X_0 se muestran a continuación, así como uno de los diagnósticos visuales de convergencia (traza de la

Cadena de Markov):



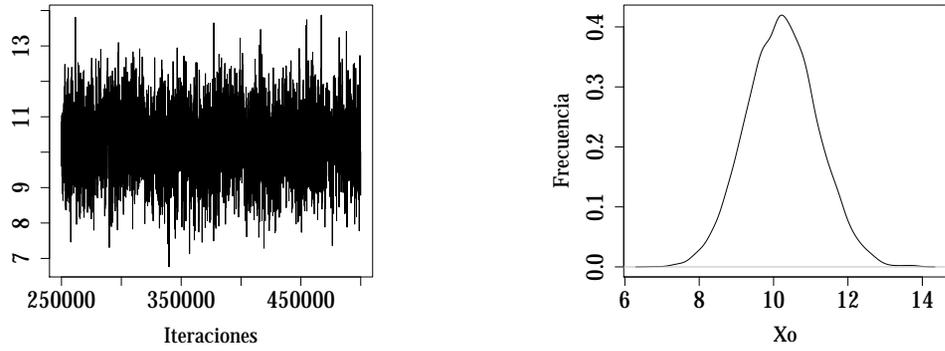


Figura 4.8: *Trazas y distribuciones posteriores para los parámetros a , α , σ^2 y X_0 .*

Se nota que cada Cadena alcanzó la convergencia y por lo tanto se obtuvieron las distribuciones posteriores marginales correspondientes. Algunos de los diagnósticos estadísticos de convergencia descritos en la sección 2.3.1 se ilustran a continuación:

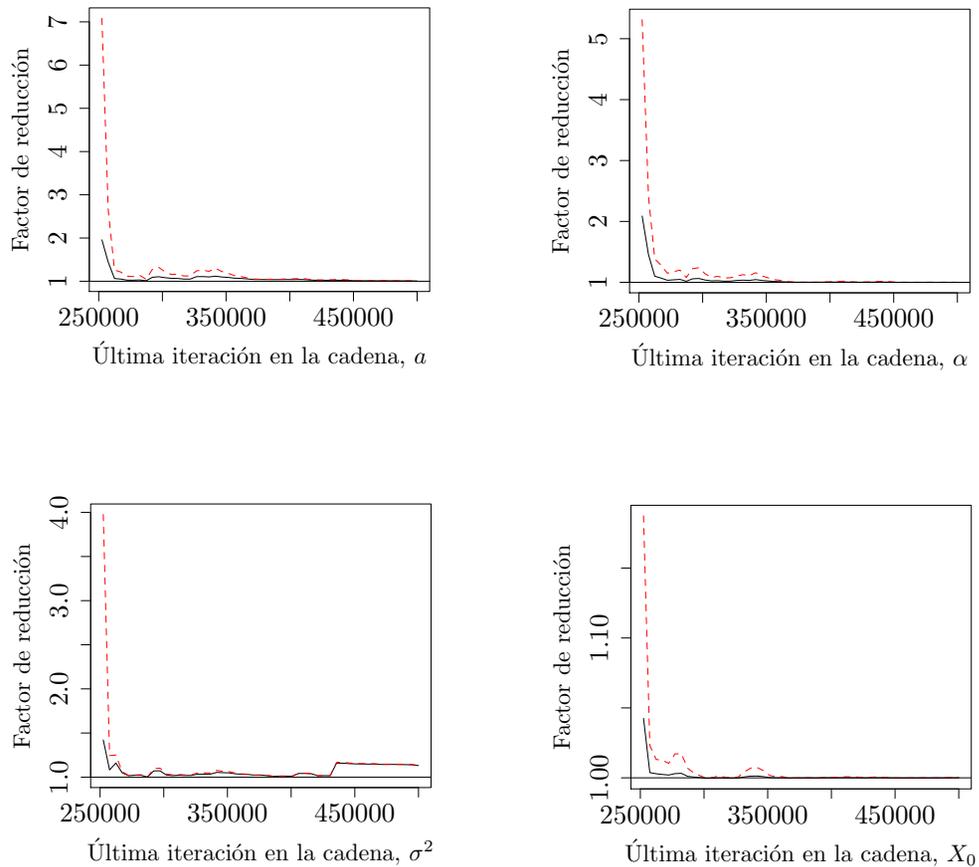


Figura 4.9: *Diagnóstico de convergencia de Gelman y Rubin para los parámetros a , α , σ^2 y X_0 respectivamente.*

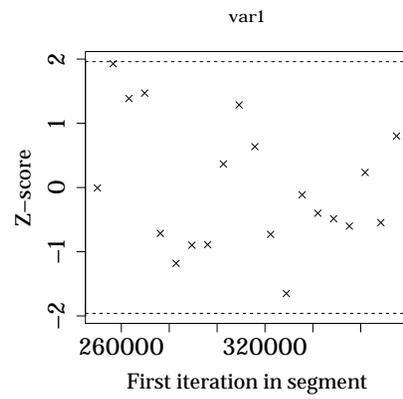
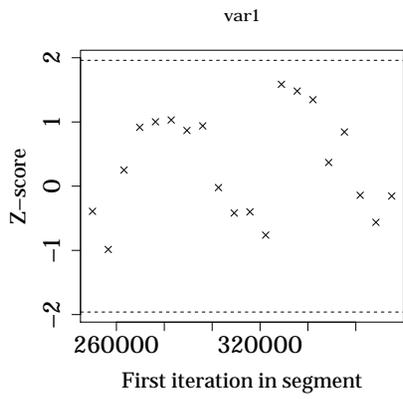
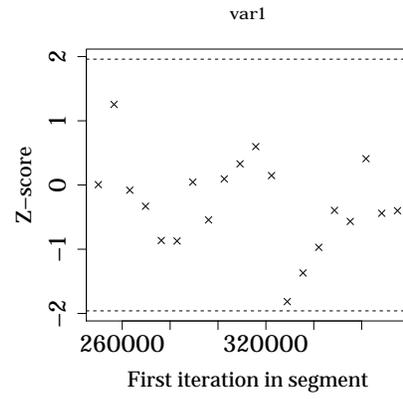
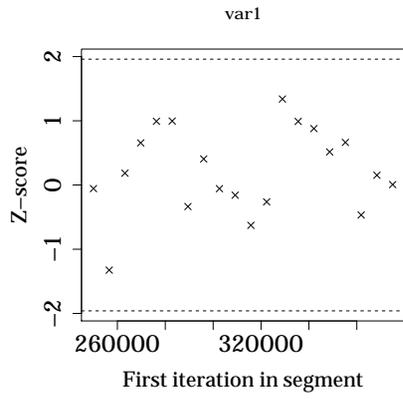
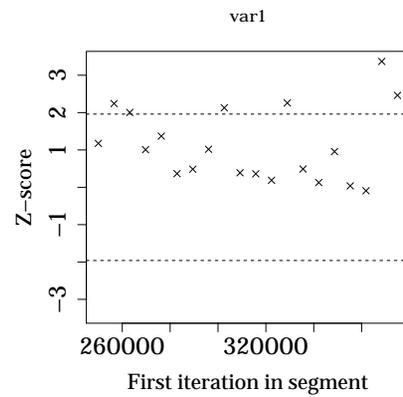
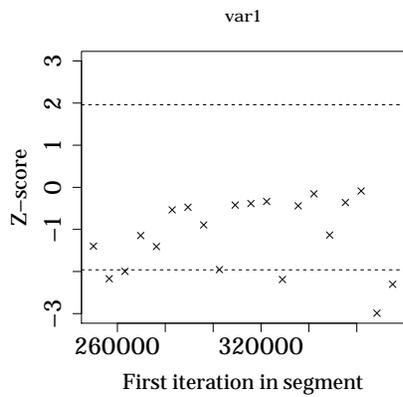


Figura 4.10: *Diagnóstico de convergencia de Geweke para los parámetros a , α , σ^2 y X_0 respectivamente, Cadena 1.*



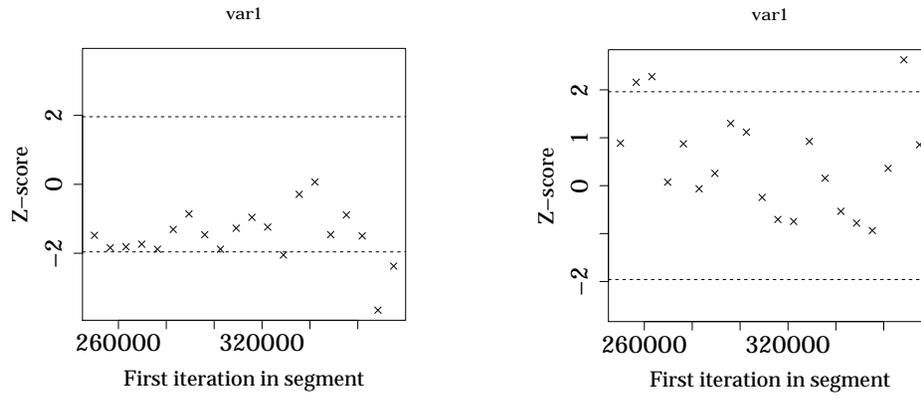


Figura 4.11: *Diagnóstico de convergencia de Geweke para los parámetros a , α , σ^2 y X_0 respectivamente, Cadena 2.*

Ejemplo 2.

El siguiente ejemplo, muestra datos actuales disponibles sobre el porcentaje en la demanda del uso de la Web móvil, comprendidos a partir de Diciembre de 2008 hasta Agosto de 2014 (http://stats.areppim.com/stats/stats_mobiwebsubstxttime.htm). El objetivo es determinar una curva de crecimiento que mejor ajuste este conjunto de datos.

El uso de la Web móvil está creciendo muy rápido desde el año 2008, mientras que la Web por medio de una PC de escritorio está continuamente perdiendo popularidad. Investigaciones por parte de la empresa *Gartner*, indican que los teléfonos móviles superarán a las PC's en el servicio de acceso Web más común en todo el mundo.

En este ejemplo de aplicación, se observa la demanda en el mercado para la Web móvil, X (porcentaje), contra el tiempo, t (meses) en el periodo Dic/2008 - Ago/2014. Los datos son reportados en la Tabla 4.4 y la Figura 4.12. El siguiente modelo es supuesto como: la demanda al tiempo t , las realizaciones de la variable aleatoria X , son escritas como en (1.2),

$$y_i = h(u(t_i)) + \varepsilon_i, \quad i = 1, \dots, n, \quad (4.19)$$

donde $h(u(t))$ describe la relación entre el rendimiento y el tiempo. Los errores ε , son *v.a.*'s tales que $\varepsilon_i \sim \mathcal{N}(0, 1)$. Además, se supone que la varianza de ε es igual a σ^2 .

Cuadro 4.4: *Porcentaje de la demanda en el mercado de la Web móvil v.s. tiempo (valores actuales Diciembre 2008 a Agosto 2014).*

Fecha	% demanda	Fecha	% demanda	Fecha	% demanda
Dic 2008	0.60	Dic 2010	4.10	Dic 2012	14.55
	0.67		4.30		14.13
	0.69		4.45		14.35
	0.80		4.70		14.44
	0.86		5.21		13.90
	0.86		5.75		14.62
Jun 2009	0.94	Jun 2011	6.53	Jun 2013	16.08
	1.05		7.02		17.35
	1.12		7.12		18.00
	1.12		6.74		17.81
	1.15		6.55		19.67
	1.21		6.95		20.04
Dic 2009	1.28	Dic 2011	8.04	Dic 2013	23.41
	1.56		8.49		23.77
	1.72		8.53		24.67
	1.96		8.99		25.42
	2.18		9.58		25.02
	2.32		10.11		27.04
Jun 2010	2.57	Jun 2012	10.40	Jun 2014	28.49
	2.86		11.09		29.48
	3.21		11.78	Ago 2014	30.10
	3.50		12.03		
	3.81		12.30		
	4.02		13.08		

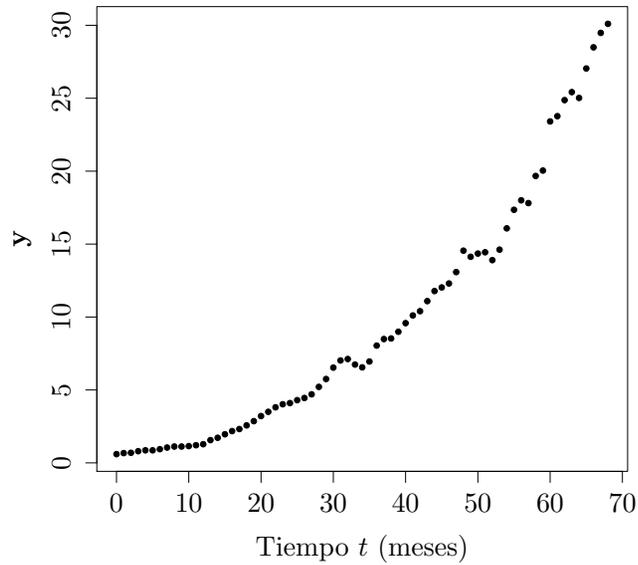


Figura 4.12: *Porcentaje de la demanda de la Web móvil v.s. tiempo.*

En la Figura 4.12 se observa que los datos presentan un tipo de crecimiento, a primera vista el crecimiento parece ser del tipo exponencial, pero también podría tratarse el caso de un crecimiento anómalo. Para definir que tipo de modelo se ajusta mejor, se realizará una selección de modelos mediante el factor de Bayes (FB).

Para este conjunto de datos, el FB resultó de 7.94×10^{27} , el cual está a favor de un modelo FEG. El DIC para el modelo OEG es 293.2 y para el modelo FEG es 156.9. Considerando a X_0 conocida que corresponde en este caso a $X_0 = 0.60$ al tiempo $t_0 = 0$ (Dic 2008), las estimaciones de los parámetros para el modelo FEG se muestran junto a la curva de la siguiente figura:

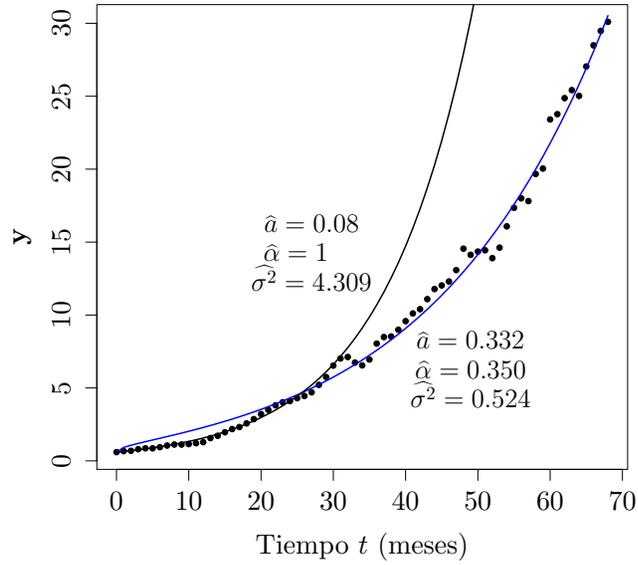


Figura 4.13: *Curva ajustada para las observaciones del porcentaje de la demanda de la Web móvil.*

La curva ajustada $h(u(t))$ es mostrada en la Figura 4.13 y es definida como:

$$X(t) = 0.60E_{0.350,1}(0.332t^{0.350}).$$

Se puede observar que los datos presentan un tipo de crecimiento anómalo, por lo que el modelo FEG tuvo un mejor ajuste respecto al OEG. En la figura anterior se muestran las curvas correspondientes a los modelos ajustados y se aprecia una gran variabilidad en los datos, en los que al parecer en un determinado intervalo de tiempo éstos tienen un crecimiento exponencial y posteriormente el crecimiento es más lento.

Las distribuciones posteriores para los parámetros a , α y σ^2 se muestran a continuación, así como uno de los diagnósticos visuales de convergencia (traza de la Cadena de Markov):

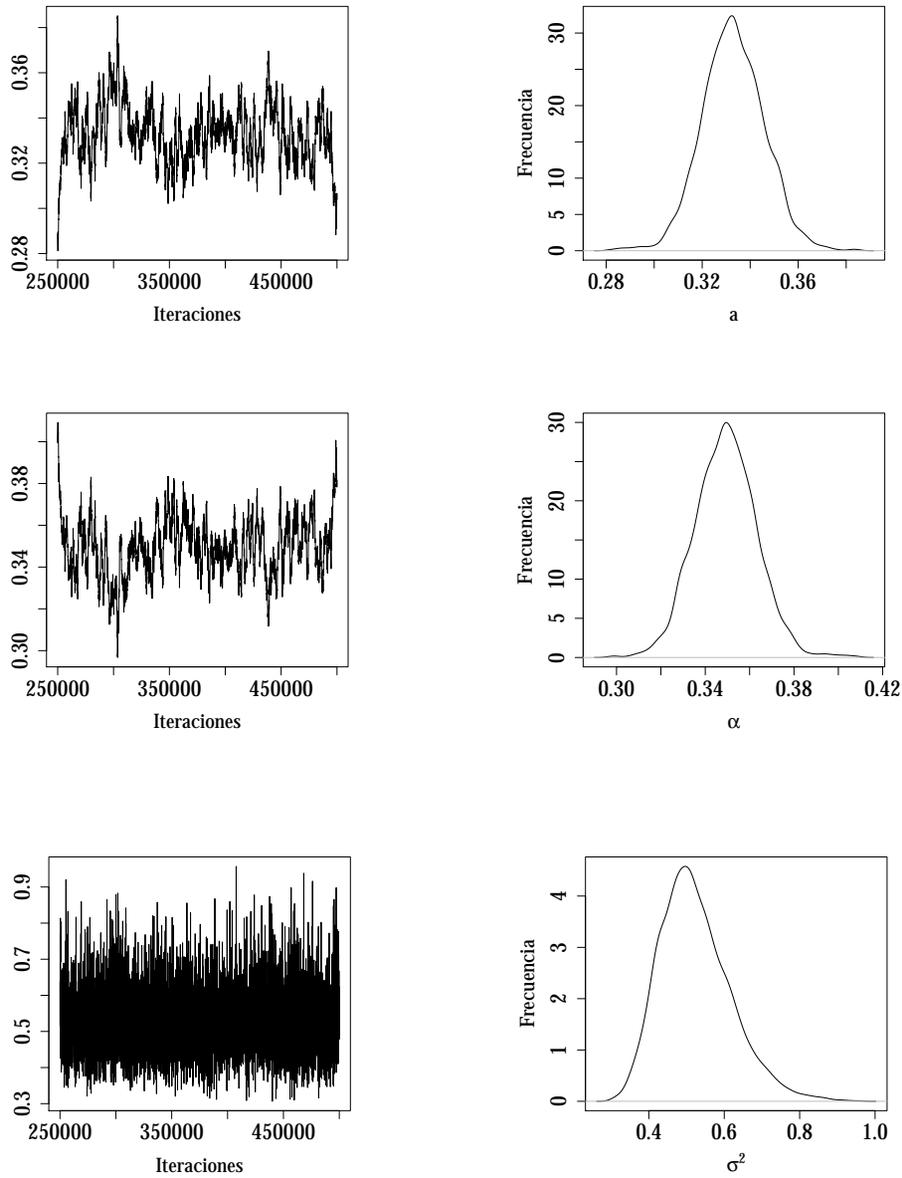


Figura 4.14: *Trazas y distribuciones posteriores para los parámetros a , α y σ^2 .*

Se nota que las Cadenas alcanzaron la convergencia y por lo tanto se obtuvieron las distribuciones posteriores marginales correspondientes.

Algunos de los diagnósticos estadísticos de convergencia descritos en la sección 2.3.1 se ilustran a continuación:

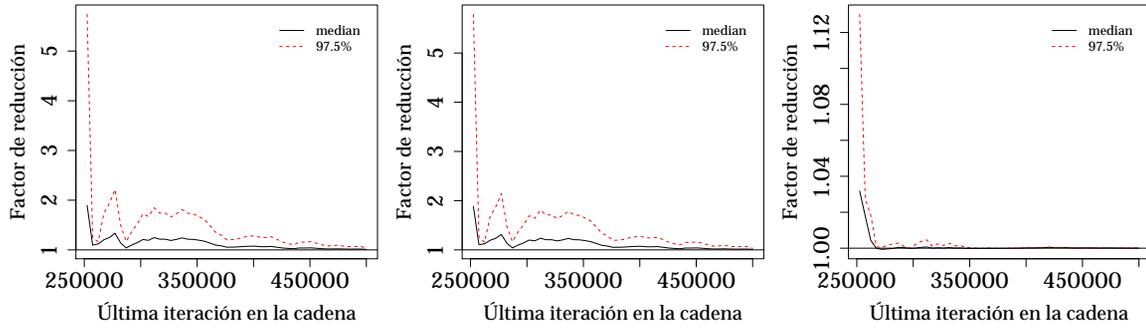


Figura 4.15: *Diagnóstico de convergencia de Gelman y Rubin para los parámetros a , α y σ^2 respectivamente.*

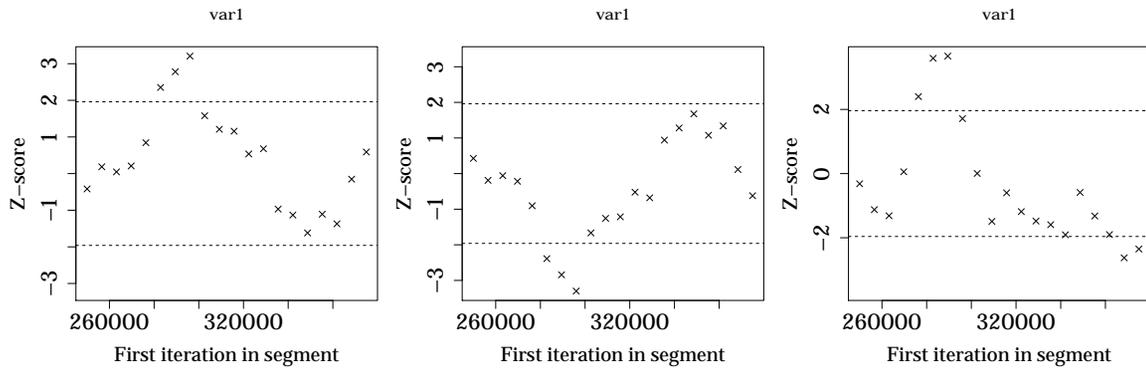


Figura 4.16: *Diagnóstico de convergencia de Geweke para los parámetros a , α y σ^2 respectivamente, Cadena 1.*

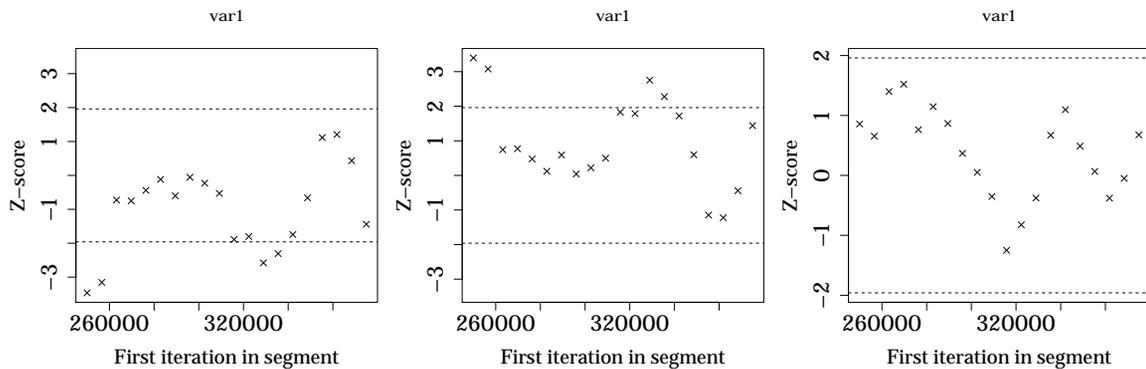


Figura 4.17: *Diagnóstico de convergencia de Geweke para los parámetros a , α y σ^2 respectivamente, Cadena 2.*

4.5. Conclusión

El uso de los métodos MCMC para la solución de problemas inversos en ecuaciones diferenciales, ha tenido buenos resultados, sobre todo cuando se utilizan modelos muy complejos en donde las distribuciones posteriores para las cantidades de interés no son convencionales. El uso de herramientas de la estadística Bayesiana para ajustar un modelo a datos reales, es una buena opción, aún cuando éstos presentan un comportamiento anómalo. En estos ejemplos, se observó que los datos presentan un cierto tipo de crecimiento anómalo, en los que a simple vista pareciera que estos podrían modelarse mediante una curva exponencial, pero mediante el factor de Bayes se seleccionó el modelo adecuado.

Este trabajo se enfocó únicamente en modelos que presentan un crecimiento exponencial. Un trabajo a futuro puede definirse para el modelo logístico de crecimiento, en donde también se pueden presentar datos para los que se requiere seleccionar un modelo con ayuda de los métodos mencionados.

Bibliografía

- Banks, H. T., Hu, S., y Thompson, W. C. (2014). *Modeling and Inverse Problems in the Presence of Uncertainty*. CRC Press. Taylor & Francis Group.
- Calvetti, D., Kaipio, J. P., y Somersalo, E. (2014). Inverse problems in Bayesian framework. *Inverse Problems*, 30(11):110301.
- Capistrán, M., Christen, J. A., y Donnet, S. (2013). Bayesian analysis of ODE's: solver optimal accuracy and Bayes factors. *arXiv preprint arXiv:1311.2281*.
- Casella, G. y George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, 46(3):167–174.
- Chen, M.-H., Shao, Q.-M., y Ibrahim, J. G. (2012). *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media.
- Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90(432):1313–1321.
- Chib, S. y Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The american statistician*, 49(4):327–335.
- Christen, J. A. y Fox, C. (2010). A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Analysis*, 5(2):263–282.
- Dashti, M. y Stuart, A. M. (2015). The Bayesian approach to inverse problems. Lecture notes to appear in Handbook of Uncertainty Quantification, Editors R. Ghanem, D. Higdon and H. Owhadi, Springer, 2016.
- Donnet, S., Foulley, J.-L., y Samson, A. (2010). Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics*, 66(3):733–741.

- Fellenius, W. (1936). Calculation of the stability of earth dams. In *Transactions of the 2nd congress on large dams, Washington, DC*, volume 4, pages 445–463. International Commission on Large Dams (ICOLD) Paris.
- Gelfand, A. E. y Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 501–514.
- Gelfand, A. E. y Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.
- Gelman, A., Carlin, J. B., Stern, H. S., y Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Taylor & Francis.
- Geman, S. y Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741.
- Geweke, J. et al. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA.
- Greco, V. R. (1996). Efficient monte carlo technique for locating critical slip surface. *Journal of Geotechnical Engineering*, 122(7):517–525.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.
- Huet, S. (2004). *Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples*. Springer Series in Statistics. Springer.
- Jin, B. y Rundell, W. (2015). A tutorial on inverse problems for anomalous diffusion processes. *Inverse Problems*, 31(3):035003.
- Juárez, E. y Rico, A. (1975). *Mecánica de Suelos (Teoría y aplicaciones de la mecánica de suelos)*, volume II. Limusa.
- Kaipio, J. y Somersalo, E. (2006). *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media.

- Kass, R. E. y Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430):773–795.
- Lee, P. M. (2012). *Bayesian statistics: an introduction*. John Wiley & Sons.
- Malkawi, A. I. H., Hassan, W. F., y Sarma, S. K. (2001). Global search method for locating general slip surface using monte carlo techniques. *Journal of geotechnical and geoenvironmental engineering*, 127(8):688–698.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., y Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087.
- Morgenstern, N. y Price, V. E. (1965). The analysis of the stability of general slip surfaces. *Geotechnique*, 15(1):79–93.
- Plummer, M. (2012). Jags: A program for analysis of Bayesian graphical models using gibbs sampling.
- Plummer, M. (2015). *rjags: Bayesian Graphical Models using MCMC*. R package version 3-15.
- Plummer, M., Best, N., Cowles, K., y Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Robert, C. y Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.
- Samson, A. y Donnet, S. (2009). Estimation of parameters in incomplete data models defined by dynamical systems. *Journal of Statistical Planning and Inference*, 137(9):2815–2831.
- Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5.
- Spiegelhalter, D. J. et al (2003). *WinBUGS Version 1.4 User Manual*.
- Stan Development Team (2014). *Stan Modeling Language: User’s Guide and Reference Manual. Version 2.2.0*.

Sturtz, S., Ligges, U., y Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, 12(3):1–16.

Su, Y.-S. y Yajima, M. (2015). *R2jags: A Package for Running 'JAGS'from R*. R package version 0.05-03.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. siam.

Anexo 1. Derivación del modelo de Kotter

En el presente anexo se proporciona la derivación de las ecuaciones 3.1 y 3.2 que corresponden al modelo de Kotter.

Se dice que un medio continuo está sometido a un *estado de esfuerzo plano continuo*, cuando los esfuerzos normales y tangenciales paralelos a la normal a ese plano determinado son nulos en todos los puntos del medio ($\sigma_x = \tau_{xy} = \tau_{xz} = 0$).

Un medio continuo está sometido a un estado continuo de *deformación plana* cuando, para todos los puntos del medio puede determinarse un plano en el cuál las deformaciones normales asociadas a él sean nulas, ver Figura 1.a:

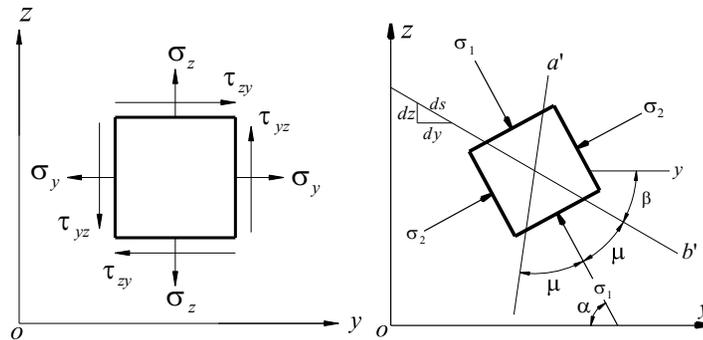


Figura 1.a: Volumen elemental sujeto a un estado de esfuerzo plano; esfuerzos principales y orientación de líneas de deslizamiento.

Conocidos los esfuerzos en un punto P arbitrario, se conoce el estado de esfuerzo plano en ese punto. En efecto, se consideran los planos XY y XZ cuyas trazas con el plano YZ son los ejes Y y Z.

Transformación de estados de esfuerzo plano

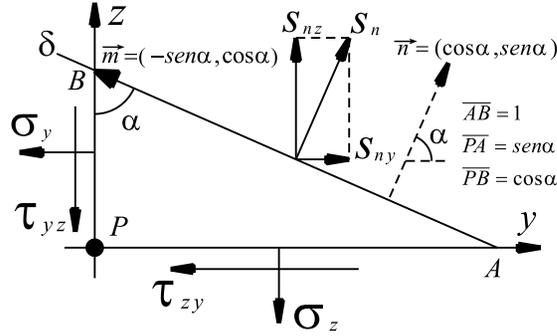


Figura 1.b: Esfuerzos en un punto P en estado de esfuerzo plano.

Las componentes normal (σ_n) y tangencial (τ_n) del esfuerzo total (S_n) asociado al plano \overline{AB} , definido por el vector $\vec{n} = (\cos \alpha, \sin \alpha)$ pueden obtenerse con los productos escalares: $\sigma_n = \vec{S}_n \cdot \vec{n}$ y $\tau_n = \vec{S}_n \cdot \vec{m}$, donde \vec{n} y \vec{m} son los vectores unitarios normal y tangente al plano \overline{AB} , con las direcciones indicadas en la Figura 1.b. Escribiendo las componentes del esfuerzo total S_n en función de las componentes de los factores se llega a:

$$\sigma_n = \sigma_y \cos^2 \alpha + \sigma_z \sin^2 \alpha + 2\tau_{yz} \sin \alpha \cos \alpha, \quad (1)$$

$$\tau_n = (\sigma_z - \sigma_y) \sin \alpha \cos \alpha + \tau_{yz} (\cos^2 \alpha - \sin^2 \alpha). \quad (2)$$

Dado un estado de esfuerzos plano, hay dos planos principales con su correspondiente esfuerzo principal ligado, denotando como σ_1 al esfuerzo principal mayor y σ_2 el menor. Por lo tanto,

$$\sigma_1 = \frac{\sigma_y + \sigma_z}{2} + \sqrt{\left(\frac{\sigma_y - \sigma_z}{2}\right)^2 + (\tau_{yz})^2},$$

$$\sigma_2 = \frac{\sigma_y + \sigma_z}{2} - \sqrt{\left(\frac{\sigma_y - \sigma_z}{2}\right)^2 + (\tau_{yz})^2}.$$

La orientación de los planos principales de esfuerzo es:

$$\tan 2\alpha = -\frac{2\tau_{yz}}{\sigma_z - \sigma_y}. \quad (3)$$

El ángulo α no queda precisamente definido por (3), por lo que ésta ecuación se puede descomponer en las expresiones:

$$\sin 2\alpha = -\frac{2\tau_{yz}}{\sqrt{(\sigma_z - \sigma_y)^2 + 4(\tau_{yz})^2}} \quad \text{y} \quad \cos 2\alpha = \frac{\sigma_z - \sigma_y}{\sqrt{(\sigma_z - \sigma_y)^2 + 4(\tau_{yz})^2}}.$$

Para el lugar geométrico, a cada elemento del conjunto de vectores unitarios \vec{n} le corresponde un elemento del conjunto de parejas ordenadas (σ_n, τ_n) .

Las ecuaciones (1) y (2) ahora se escriben en un sistema de planos principales, correspondientes a esfuerzos principales:

$$\sigma_n = \sigma_1 \cos^2 \alpha + \sigma_2 \sin^2 \alpha,$$

$$\tau_n = (\sigma_2 - \sigma_1) \sin \alpha \cos \alpha,$$

siendo σ_1 el esfuerzo principal mayor, y σ_2 , el menor. Usando identidades trigonométricas básicas, las ecuaciones pueden transformarse como:

$$\sigma_n - \frac{\sigma_1 + \sigma_2}{2} = \frac{\sigma_1 - \sigma_2}{2} \cos 2\alpha,$$

$$\tau_n = -\frac{\sigma_1 - \sigma_2}{2} \sin 2\alpha.$$

Elevando al cuadrado las dos expresiones anteriores y sumándolas se obtiene finalmente:

$$\left(\sigma_n - \frac{\sigma_1 + \sigma_2}{2}\right)^2 + \tau_n^2 = \left(\frac{\sigma_1 - \sigma_2}{2}\right)^2,$$

lo cual es el lugar geométrico de un círculo de centro $(\frac{\sigma_1 + \sigma_2}{2}, 0)$ y radio $(\frac{\sigma_1 - \sigma_2}{2})$ llamado *círculo de Mohr*.

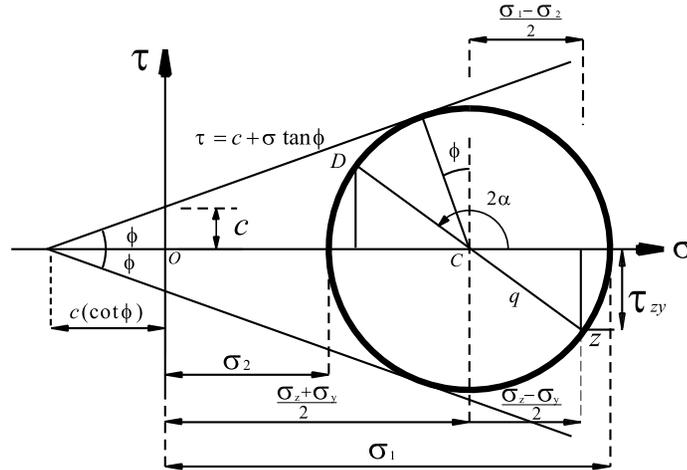


Figura 1.c: *Círculo de Mohr-Coulomb*.

Teoría de la línea de deslizamiento

Para determinar la forma o característica de la línea de deslizamiento, se reescribe el criterio de Mohr-Coulomb en términos de un Estado General de Esfuerzos:

$$\left[\frac{1}{2}(\sigma_z + \sigma_y) + c(\cot \phi)\right] \sin \phi - \left[\frac{1}{4}(\sigma_z - \sigma_y)^2 + \tau_{yz}^2\right]^{\frac{1}{2}} = 0, \quad (4)$$

donde, de la Figura 1.c, cada término de (4) es el radio del círculo de Mohr-Coulomb.

Las fuerzas de cuerpo que actúan son el peso, que se considera unitario; así las ecuaciones de equilibrio al elemento prismático son:

$$\frac{\partial \sigma_y}{\partial y} + \frac{\partial \tau_{zy}}{\partial z} = 0, \quad (5)$$

$$\frac{\partial \tau_{zy}}{\partial y} + \frac{\partial \sigma_z}{\partial z} = \gamma. \quad (6)$$

De la Figura 1.a, la ecuación para la línea de deslizamiento es:

$$\frac{dz}{dy} = \tan \beta = \tan(\alpha \pm \mu).$$

De la Figura 1.c, se ve que $\sigma_m = q(\csc \phi) - c(\cot \phi)$.

$$\sigma_y = (q \csc \phi - c \cot \phi) + q \cos 2\alpha,$$

$$\sigma_z = (q \csc \phi - c \cot \phi) - q \cos 2\alpha,$$

$$\tau_{zy} = q \sen 2\alpha.$$

Sustituyendo estas últimas ecuaciones en (5) y (6), resulta:

$$\begin{aligned} \frac{\partial q}{\partial y} (\csc \phi + \cos 2\alpha) + \frac{\partial q}{\partial z} \sen 2\alpha - 2q \left(\sen 2\alpha \frac{\partial \alpha}{\partial y} - \cos 2\alpha \frac{\partial \alpha}{\partial z} \right) &= 0, \\ \frac{\partial q}{\partial z} (\csc \phi - \cos 2\alpha) + \frac{\partial q}{\partial y} \sen 2\alpha - 2q \left(\sen 2\alpha \frac{\partial \alpha}{\partial z} + \cos 2\alpha \frac{\partial \alpha}{\partial y} \right) &= \gamma. \end{aligned}$$

Usando identidades trigonométricas y la definición de derivada total, las ecuaciones anteriores se reducen a:

$$\cot \phi (dq) - 2q(d\beta) - (\cos \phi dz - \sen \phi dy)\gamma = 0. \quad (7)$$

Los esfuerzos en estado crítico que tocan la envolvente de falla son:

$$\tilde{\tau} = q \cos \phi \text{ y } \tilde{\tau} = \tilde{\sigma} \tan \phi + c,$$

que sustituidos en (7) se obtiene:

$$\begin{aligned} \frac{d\tilde{\tau}}{d\beta} &= 2\tilde{\tau} \tan \phi - \gamma R \sen \phi \sen(\beta - \phi), \\ \frac{d\tilde{\sigma}}{d\beta} &= 2c + 2\tilde{\sigma} \tan \phi - \gamma R \cos \phi \sen(\beta - \phi), \end{aligned}$$

que son llamadas las **ecuaciones de Kötter**. La interpretación para éstas, se ilustran en la Figura 3.1.

Anexo 2. Ejemplo de implementación en JAGS

Los modelos presentados en los capítulos 3 y 4 son implementados en JAGS (Plummer, 2012) usando los paquetes R2jags (Su y Yajima, 2015), coda (Plummer et al., 2006), lattice (Sarkar, 2008), R2WinBUGS (Sturtz et al., 2005) y rjags (Plummer, 2015) dentro del ambiente de R (R Core Team, 2014). El paquete R2jags permite ajustar modelos y realizar diagnósticos de convergencia. Se usa el conjunto de datos dado para los ejemplos en cada capítulo. Para la sección 3.3, en la cual sólo una cantidad es de interés, se omite el código utilizado, ya que se sigue una secuencia similar a los pasos que se enunciarán a continuación:

1. Paquetes.

```
library(R2jags)
library(coda)
library(lattice)
library(R2WinBUGS)
library(rjags)
```

2. Datos. Se leen y se generan los datos en el directorio de trabajo

```
setwd("C:/Users/LuisXavier/TESIS")
# Se generan datos sintéticos provenientes del modelo FEG
# Valores para la simulación
n <- 10
a <- 0.2
t <- 1:n
xo <- 20
sigma2 <- 30
alpha <- 0.5
z <- a*t^alpha

# Leer el archivo para la función Mittag-Leffler
```

```

source("C:/Users/LuisXavier/TESIS/ML.R")

# Vector de media para la función de observación
mu2 <- numeric()
for(i in 1:length(z)){
mu2[i] <- xo*E.ab(alpha,1,z[i])
}

# Generando la muestra
Y <- numeric(0)
for(j in 1:n){
Y[j] <- rnorm(1,mean=mu2[j],sd=sqrt(sigma2))}
Y

# Preparando los datos
N <- length(Y)
datos <- list("Y","t","N")

```

3. **El modelo.** Se escriben los modelos OEG y FEG como un código BUGS y se guardan con terminación `.jags` en el directorio de trabajo.

```

#Codigo BUGS para OEG
cat("
model {
for (i in 1:N){
Y[i] ~ dnorm(mu[i], tau)
mu[i] <- xo*exp(a*t[i])
}
xo <- 20
a ~ dgamma(1,1)
sigma2 <- 1/tau
tau ~ dgamma(0.0009, 0.03) # media =0.03 y var=1

```

```

)",file="M_1.jags",fill=T)

#Codigo BUGS para FEG
cat("
data{
# Leer el archivo para la serie de la función Mittag-Leffler
source("C:/Users/LuisXavier/TESIS/mittag_leffler.R")
}
model {
for (i in 1:N){
Y[i] ~ dnorm(mu[i], tau)
mu[i] <- xo*Mitag(z)
}
xo <- 20
a ~ dgamma(1,1)
sigma2 <- 1/tau
tau ~ dgamma(0.0009, 0.03) # media =0.03 y var=1
alpha ~ dunif(0,1)
)",file="M_2.jags",fill=T)

```

4. **Parámetros.** Se definen los parámetros de interés.

```

#Parámetros a estimar de OEG
par.m1 <- c("a","sigma2")

#Parámetros a estimar de FEG
par.m2 <- c("a","sigma2","alpha")

```

5. **Valores iniciales.** Se definen los valores iniciales para cada Cadena que son introducidos en JAGS.

```

# Val. iniciales para OEG
units1.1 <- list("a"=0.1,"tau"=0.02)

```

```

units2.1 <- list("a"=1,"tau"=.05)
inits2.1 <- list(units1.1, units2.1)

# Val. iniciales para OEG
units1.2 <- list("a"=0.1,"tau"=0.021,"alpha"=0.8)
units2.2 <- list("a"=0.9,"tau"=.05,"alpha"=0.5)
inits2.2 <- list(units1.2, units2.2)

```

6. **Ajuste.** Se ajusta el modelo en JAGS.

```
#Ajustando el modelo en JAGS para OEG
```

```
fit1 <- jags(data=datos, inits=inits2.1, parameters.to.save=par.m1,
n.chains = 2,n.iter=5000, n.burnin=1000,model.file="M_1.jags")
```

```
#Ajustando el modelo en JAGS para FEG
```

```
fit2 <- jags(data=datos, inits=inits2.2, parameters.to.save=par.m2,
n.chains = 2,n.iter=5000, n.burnin=1000,model.file="M_2.jags")
```

7. **Diagnósticos de convergencia.** Se convierte la salida del modelo en un objeto MCMC para obtener varios diagnósticos de convergencia.

```
#Convirtiendo el modelo OEG en un objeto MCMC
```

```
fit.mcmc1 <- as.mcmc(fit1)
```

```
summary(fit.mcmc1)
```

```
densityplot(fit.mcmc1)
```

```
gelman.plot(fit.mcmc1)
```

```
gelman.plot(fit.mcmc1)
```

```
geweke.diag(fit.mcmc1)
```

```
geweke.plot(fit.mcmc1)
```

```
raftery.diag(fit.mcmc1)
```

```
raftery.plot(fit.mcmc1)
```

```
heidel.diag(fit.mcmc1)
```

Similarmente se realiza para el modelo FEG.

8. **Estimación de la verosimilitud marginal.** Se estima ML para ambos modelos como en (4.12).

```
#muestras de la dist. posterior modelo OEG:
a.mc1 <- c(fit.mcmc1[[1]][,1],fit.mcmc1[[2]][,1])
sigma2.mc1 <- c(fit.mcmc1[[1]][,3],fit.mcmc1[[2]][,3])

#Matriz con los parametros "a y sigma2".
Theta.M1 <- matrix(NA,length(a.mc1),2)
Theta.M1[,1] <- a.mc1
Theta.M1[,2] <- sigma2.mc1

#Datos:
myData <- Y

#Funcion de densidad de los datos "log-verosimilitud" (log_f):
L.M1<-function(Y,t,theta,xo)
#Y es vector de datos, theta=(a,sigma2,xo) es vector de parámetros
{ a<-theta[1]; sigma2<-theta[2];
mu=xo*exp(a*t)
return(sum(dnorm(Y,mu,sqrt(sigma2),log=T)))}

# Densidad inicial de los parametros "log-priori" (log_pi):
prior.M1<-function(theta,y1) #y1 es el valor 1 del vector de datos Y
{ dgamma(theta[1],1,1,log=T) + dgamma(1/theta[2],0.0009,0.03,log=T)}

#Densidad con cola mas delgada que el producto de la inicial y la
verosimilitud "log-densidad" (log_P):
ptheta.M1<-function(theta,mu.mc,sigma.mc)
{ dnorm(theta[1],mean=mu.mc[1],sd=sigma.mc[1],log=T) +
dnorm(theta[2],mean=mu.mc[2],sd=sigma.mc[2],log=T) }
```

```

#Tomando exp() del argumento:
mGD.M1<-function(theta,Y,t,mu,sigma,xo)
{ y1<-Y[1]
exp(ptheta.M1(theta,mu,sigma)-L.M1(Y,t,theta,xo)-prior.M1(theta,y1))}
MC1<-length(a.mc1)

#vectores de media y varianza de las muestras posterior.
mu.m1 <- apply(Theta.M1,2,mean)
sigma.m1 <- apply(Theta.M1,2,sd)

#vector que contiene los argumentos ptheta.M1/L.M1*priori.M1
mGD1 <- numeric(0)
xo <- 20
for(i in 1:MC1)
{ mGD1[i]<-mGD.M1(Theta.M1[i,],Y,t,mu=mu.m1,sigma=sigma.m1,xo)}

#Estimador de Gelfand & Dey:
m.GD.M1 <- 1/mean(mGD1)

#muestras de la dist. posterior modelo FEG:
a.mc2 <- c(fit.mcmc2[[1]][,1],fit.mcmc2[[2]][,1])
alpha.mc2 <- c(fit.mcmc2[[1]][,2],fit.mcmc2[[2]][,2])
sigma2.mc2 <- c(fit.mcmc2[[1]][,4],fit.mcmc2[[2]][,4])

#Matriz con los parametros "a,sigma2,xo".
Theta.M2 <- matrix(NA,length(a.mc2),3)
Theta.M2[,1] <- a.mc2
Theta.M2[,2] <- alpha.mc2
Theta.M2[,3] <- sigma2.mc2

```

```

#Datos:
myData <- Y

#Funcion de densidad de los datos "log-verosimilitud" (log_f):
L.M2<-function(Y,t,theta,xo)
#Y vector de datos, theta=(a,alpha,sigma2,xo) vector de parámetros
{ a<-theta[1]; alpha<-theta[2]; sigma2<-theta[3];
mu=xo*E.ab(alpha,1,a*t^alpha)
return(sum(dnorm(Y,mu,sqrt(sigma2),log=T)))}

# Densidad inicial de los parametros "log-priori" (log_pi):
prior.M2<-function(theta,y1) #y1 es el valor 1 del vector de datos Y
{ dgamma(theta[1],1,1,log=T) + dunif(theta[2],0,1,log=T) +
dgamma(1/theta[3],0.0009,0.03,log=T)}

#Densidad con cola mas delgada que el producto de la inicial y la
verosimilitud "log-densidad" (log_P):
ptheta.M2<-function(theta,mu.mc,sigma.mc)
{ dnorm(theta[1],mean=mu.mc[1],sd=sigma.mc[1],log=T) +
  dnorm(theta[2],mean=mu.mc[2],sd=sigma.mc[2],log=T)+
  dnorm(theta[3],mean=mu.mc[3],sd=sigma.mc[3],log=T)}

#Tomando exp() del argumento
mGD.M2<-function(theta,Y,t,mu,sigma,xo)
{ y1<-Y[1]
exp(ptheta.M2(theta,mu,sigma)-L.M2(Y,t,theta,xo)-prior.M2(theta,y1))}

MC2<-length(a.mc2)

#vectores de media y varianza de las muestras posterior.
mu.m2 <- apply(Theta.M2,2,mean)

```

```

sigma.m2 <- apply(Theta.M2,2,sd)

#vector que contiene los argumentos ptheta.M1/L.M1*priori.M1
mGD2<-numeric(0)
x0<-20
for(i in 1:MC2)
{ mGD2[i]<-mGD.M2(Theta.M2[i,],Y,t,mu=mu.m2,sigma=sigma.m2,xo)}

#Estimador de Gelfand & Dey:
m.GD.M2 <- 1/mean(mGD2)

```

9. **Factor de Bayes.** Se realiza la selección del modelo mediante el FB.

```

# Para los modelos OEG y FEG el FB es:
FB <- m.GD.M2/m.GD.M1

```