TEXTOS ACADÉMICOS

Introducción a la Estadística Bayesiana

Juan Carlos Correa Morales
Carlos Javier Barrera Causil

FONDO EDITORIAL ITM

INTRODUCCIÓN A LA ESTADÍSTICA BAYESIANA

Juan Carlos Correa Morales

Carlos Javier Barrera Causil

Correa Morales, Juan Carlos
Introducción a la Estadística Bayesiana / Juan Carlos Correa Morales, Carlos Javier Barrera Causil. – 1a ed. – Medellín:
Instituto Tecnológico Metropolitano, 2018
222 p. – (Textos Académicos)

Incluye referencias bibliográficas
ISBN 978-958-5414-24-2

1. Estadística bayesiana   I. Barrera Causil, Carlos Javier   II. Título   III. Serie

519.542   SCDD Ed.21

Catalogación en la publicación - Biblioteca ITM
Introducción a la Estadística Bayesiana

Edición: enero 2018
ISBN: 978-958-5414-24-2
Publicación electrónica para consulta gratuita

Autores
JUAN CARLOS CORREA MORALES
CARLOS JAVIER BARRERA CAUSIL

Rectora
MARÍA VICTORIA MEJÍA OROZCO

Directora Editorial
SILVIA INÉS JIMÉNEZ GÓMEZ

Comité Editorial
EDUARD EMIRO RODRÍGUEZ RAMÍREZ, MSC.
JAIME ANDRÉS CANO SALAZAR, PHD.
SILVIA INÉS JIMÉNEZ GÓMEZ, MSC.
YUDY ELENA GIRALDO PÉREZ, MSC.
VIVIANA DÍAZ, ESP.

Corrección de textos
LILA MARÍA CORTÉS FONNEGRA

Secretaria Técnica
VIVIANA DÍAZ

Diagramación
CARLOS JAVIER BARRERA CAUSIL

Diseño de carátula
LEONARDO SÁNCHEZ

Editado en Medellín, Colombia
Instituto Tecnológico Metropolitano
Sello editorial Fondo Editorial ITM
Calle 73 No. 76A 354
Tel.: (574) 440 5197 • Fax: 440 5382
www.itm.edu.co

La incertidumbre está en todas partes
y tú no puedes escapar de ella.
Dennis Lindley

El azar no es, sin embargo, una loca fantasía;
responde a su vez a leyes.
Los dados obedecen a la gravedad
y sólo tienen seis caras.
Juan José Sebreli
Comediantes y mártires: ensayo contra los mitos

Prefacio

La estadística bayesiana es un campo que ha tenido un desarrollo impresionante en los últimos años, en especial desde la introducción de la parte computacional. Muchas ideas han estado circulando desde hace tiempo, pero su imposibilidad práctica hacía que se miraran con cierto pesar, ya que eran muy atractivas pero inaplicables. Esto afortunadamente ha cambiado. Es lamentable que muchos de los libros básicos en estadística no hagan una presentación de los elementos básicos de esta aproximación para la solución de problemas estadísticos. Libros en estadística bayesiana han aparecido en las últimas dos décadas por cantidades apreciables. Antes de los años 90 se tenían libros más enfocados en la teoría de la decisión ([1]; [2]; [3]; [4]; [5]; [6]; [7]; [8]), en aspectos teóricos de la probabilidad subjetiva ([9]; [10]) y algunos pocos a la estadística bayesiana aplicada ([11]; [12]; [13]; [14]). En las últimas dos décadas esto se ha revertido y encontramos libros aplicados de estadística bayesiana en muchas áreas: generales ([15]; [16]; [17]; [18]; [19]; [20]; [21]; [22]; [23]; [24]; [25]), pronósticos [26], econometría [27], bioestadística ([28]; [29]; [30]), ciencias sociales ([31]; [32]), confiabilidad ([33]; [34]), mercadeo [35], aplicaciones en ingeniería civil [36] y otros dedicados a la parte computacional ([37]; [38]; [39]; [40]; [41]; [42]). Samaniego [43] presenta una extensa comparación entre los métodos frecuentistas y los métodos bayesianos. La estadística bayesiana no ha tenido un camino fácil en el mundo del trabajo aplicado. Qin [44] presenta un recuento histórico del uso, discusiones y reservas, de la estadistica bayesiana en econometría, historia que puede ser similar en diferentes áreas de investigación. La Inteligencia Artificial es un área de un fuerte desarrollo tanto teórico y aplicado de gran importancia que hace uso extenso de métodos bayesianos [45].

Aquí vamos a presentar una aproximación eminentemente práctica, esto es, el lector puede aplicar de forma casi inmediata los métodos a problemas reales. El software que se utilizará es de dominio público como el *R* [46] y el *OpenBUGS*. Se requiere familiaridad con el primer programa, al menos a un nivel operativo básico. Haremos énfasis en la parte de construcción de la distribución a priori que resuma el conocimiento previo del experto. Esta parte generalmente no es considerada en los textos de estadística bayesiana moderna, pero consideramos que es la esencia misma del análisis bayesiano y constituye el aporte de este trabajo. Consideramos que si la estadística bayesiana se diferencia en algo de la estadística tradicional (clásica) es en permitirle al usuario incorporar información disponible de una manera transparente y directa.

El programa y lenguaje estadístico R [46] se ha vuelto uno de los estándares para realizar trabajo estadístico, tanto aplicado como el desarrollo de nuevas metodologías. La estadística bayesiana se ha beneficiado enormemente de la flexibilidad y el potencial de este programa, el cual permite crear fácilmente librerías y ponerlas en la red de tal forma que usuarios a nivel mundial puedan usarlas y validarlas, retroalimentando a sus creadores de tal forma que, en muy poco tiempo se tengan subprogramas de gran eficiencia y calidad. Hay ahora muchas librerías que han sido creadas para resolver problemas de tipo general, como son modelos lineales y lineales generalizados, o más generales aún que permiten a un usuario resolver problemas propios mediante el uso de muestreadores proporcionados en estas librerías, como ejemplo tenemos MCMCpack [47] [48], la cual permite ajustar muchos modelos útiles en el trabajo aplicado de una manera simple y directa como se hace en R. Creemos que el éxito de R ha venido en el detrimento de programas como el $WinBUGS$, ya que un investigador prefiere crear programas que por un lado sean más transparentes y, por otro lado, que lleguen a un público más amplio, aunque los estadísticos bayesianos dicen que son complementarios.

Este texto está dirigido a investigadores, estudiantes de pregrado y posgrado en estadística, ingeniería y ciencias, que tengan familiaridad con los métodos estadísticos a un nivel operativo, al menos. Conocimiento de inferencia a un nivel de un texto básico de estadística matemática del estilo de Hogg y Craig [49] o Mood, Graybill y Boes [50] ayuda bastante.

Índice general

Parte I

Elementos básicos

Capítulo 1

Introducción

El problema fundamental del progreso científico, y uno fundamental en la vida diaria, es el de aprender de la experiencia. El conocimiento obtenido de esta manera es parcialmente una descripción de lo que ya hayamos observado, pero una parte consiste en la realización de inferencias de la experiencia pasada para predecir la experiencia futura[9].

La escuela bayesiana en estadística ha tomado fuerza en los últimos años, debido a su potencial para resolver problemas que no se pueden atacar con otros métodos y porque permite incorporar naturalmente información que es útil en la solución del problema enfrentado. Nadie niega que ante un problema en particular debemos utilizar toda la información disponible acerca del mismo o de sucesos similares. Para nuestro caso estadístico, la incertidumbre sobre parámetros poblacionales se resume por medio de distribuciones de probabilidad, que antes de recoger información muestral relevante para ellos, se conoce como 'distribución a priori.' El problema está en la forma de cuantificar esta información sin generar alguna contradicción.

La aproximación bayesiana es una herramienta fundamental en situaciones donde la recolección de información muestral sea muy difícil, por ejemplo en tópicos de alta sensibilidad social, tales como el consumo de drogas ilícitas, o extremadamente costosos o imposibles, como sería el caso de la determinación del riesgo de falla de una nueva nave espacial o cuál es la probabilidad de que haya vida inteligente en nuestra galaxia.

Un problema que se ha planteado cuando se habla de la escuela bayesiana es que dos personas enfrentadas ante un problema y una decisión a tomar, y asumiendo que tengan la misma información muestral, pueden llegar a dos decisiones opuestas si su información adicional es diferente. Greenland [51] afirma que «los epidemiólogos perciben la especificación de la distribución a priori como no práctica y además pocos epidemiólogos emplearían métodos que no están disponibles en paquetes estadísticos líderes». Dienes [52] discute en detalle las posiciones de ambas escuelas.

En estadística realizamos y tratamos de responder preguntas con respecto a las características de una o varias poblaciones. En la aproximación bayesiana tenemos:

- La información sobre un parámetro (puede ser un vector) que se tiene se debe resumir en una distribución de probabilidad, esta será llamada la *distribución a priori*.
- Los parámetros son considerados variables aleatorias (esto no es aceptable en la estadística clásica).
- La información a priori puede provenir de:
  - Estudios previos.
  - Información subjetiva de expertos (la cuantificación de esta información es lo que llamamos *elicitación*).

Albert [53] presenta las siguientes razones por las cuales se debería enseñar estadística desde el punto de vista bayesiano:

- El paradigma bayesiano es un medio natural de implementar el método científico donde la distribución a priori representa sus creencias iniciales acerca del modelo, usted recoge los datos adecuados, y la distribución posterior representa sus creencias actualizadas después de ver los datos.
- Si la incertidumbre acerca de los modelos es expresada utilizando probabilidad subjetiva, entonces la regla de Bayes es la única receta que uno necesita para realizar inferencias de los datos.
- Las afirmaciones inferenciales bayesianas son más fáciles de entender que las basadas en la inferencia tradicional basadas en muestreo repetido. La probabilidad de que un parámetro caiga dentro de un intervalo calculado es igual a 0.95. También, en contraste con los procedimientos tradicionales de pruebas de hipótesis, tiene sentido hablar acerca de la probabilidad de que una hipótesis estadística sea cierta.
- Por el principio de condicionalidad, los únicos datos relevantes para ejecutar inferencias son los datos realmente observados. Uno puede ignorar otros resultados de un espacio muestral que no son observados.
- Los problemas de predicción no son más difíciles que los problemas de estimación de parámetros. Parámetros y observaciones futuras son cantidades desconocidas que son modeladas subjetivamente.

Western y Jackman [54] hacen un recuento de las críticas que dos famosos estadísticos hacen de la aproximación bayesiana (Fisher y Efron). Una de las críticas es la introducción de información subjetiva a priori que haría que los prejuicios de los analistas fueran introducidas en los análisis, dañando los resultados. Tanto Fisher como Efron argumentan que con la inclusión de información subjetiva no es posible realizar un análisis justo de los datos. A lo cual Western y Jackman replican diciendo:

> En la práctica, sin embargo, la información a priori entra en la mayoría de los análisis a través de decisiones de codificación, transformaciones y búsquedas

no reportadas en conjuntos de variables exploratorias para obtener resultados que parezcan significativos en el sentido de caer dentro de un rango de valores esperados. Mientras todos los analistas de datos usan creencias previas, los bayesianos hacen la forma de volver estas aprioris explícitas e integrarlas sistemáticamente en el análisis. Y, reparafrasean a de Finetti quien dijo que el reconocimiento de la subjetividad es el camino a la objetividad.

Kyburg, Jr. [55] nos presenta esta reflexión sobre la incertidumbre:

> Hay dos clases de ignorancias que considero: la clase más simple de ignorancia es la que hace que las loterías sean excitantes; la otra es la que hace que las carreras de caballos sean excitantes.
>
> Una lotería es excitante debido que aunque sepamos exactamente que uno de los números de los posibles se obtendrá, y aunque sepamos que todo lo posible haya sido hecho para garantizar que ninguno estos números tenga ventaja sobre los otros, no sabemos cuál saldrá. Esto es generalmente expresado diciendo que la probabilidad de un estado es la misma que la de cualquier otro estado: o, en el caso particular de la lotería, que la probabilidad de que un tiquete gane es igual a la de cualquier otro tiquete. Esto no es el caso típico en las carreras de caballos; uno no puede organizar las carreras de caballos de tal forma que cada caballo en la carrera tenga (mediante algún consenso general) la misma probabilidad de ganar. Existe una gran cantidad de información acerca de cada caballo que determina la probabilidad que ese caballo gane (si es que tal probabilidad existe del todo), y no hay una forma aceptable de cambiar esas circunstancias- lastimándolo, digamos- tal que las probabilidades sean iguales. Uno encontraría difícil, quizá, elaborar una distinción clara y precisa entre estas dos clases de situaciones que no pudiera ser atacada como artificial; y aún así ellas parecen diferir de una forma importante. Yo consideraría la primera situación como una incertidumbre estadística; y la última, donde las probabilidades dependen fuertemente del conocimiento, como una incertidumbre epistemológica.

1.1. Ejemplos típicos

Ejemplo 1.1. Cálculo de la edad de una persona. En nuestra sociedad es considerado como una forma de mala educación preguntar la edad de una persona. El día que conoce a alguien, usted más o menos puede calcular la edad de esta persona. Este proceso se hace de una manera inconsciente y usualmente llega a un número que aproxima sus creencias sobre la posible edad. Para esto usa la información recolectada previamente sobre ella, por ejemplo, si esta persona tiene una apariencia determinada, si se viste de cierta forma, si se graduó del colegio en cierta época, etc. Si dos personas tienen que calcular la edad de este sujeto, puede que ellas no coincidan en sus resultados, pero no se puede decir cuál de los dos está equivocado (o si los dos lo están), solo hasta el momento en que se conozca la verdadera edad de la persona en cuestión. La incertidumbre que usted tiene acerca de la edad de una persona la podemos expresar en términos probabilísticos con ayuda de la siguiente plantilla (ver Figura 1.1).

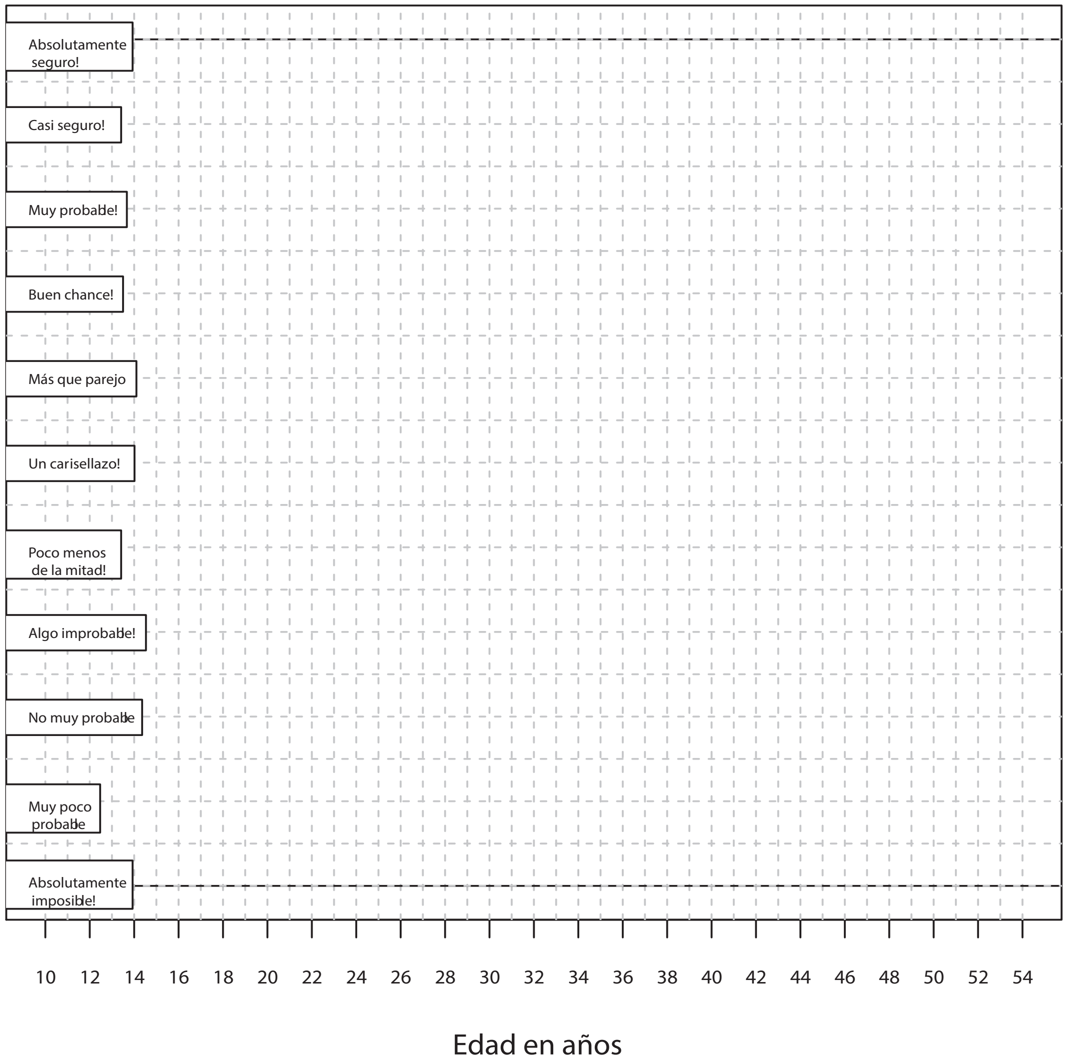

Figura 1.1: *mediante la ayuda de la plantilla podemos 'elicitar' la distribución de probabilidad que nos refleja la incertidumbre que tenemos sobre la edad de una persona. Nota: todas las figuras y tablas del texto son de elaboración propia del autor*

Ejemplo 1.2. La lotería que jugó anoche. Suponga que a usted un amigo le ofrece un billete de lotería, pero con el problema que la lotería jugó anoche. Su amigo, que ha demostrado ser una persona honesta le informa que él no sabe el resultado de la lotería, y usted tampoco. En una situación como esta podemos pensar en una probabilidad de que el billete sea el ganador es la misma que el billete tenía antes de que se jugara la lotería, ¿no lo piensa así?

Ejemplo 1.3. Estatura de los colombianos. Si pensamos en la estatura promedio de los hombres colombianos podemos pensar seriamente que este valor no es mayor que 180 cm., ni menor que 160 cm. Es claro que si conocemos muchos hombres colombianos nuestra información puede utilizarse en un proceso inferencial, pero confiaríamos más si la información sobre la estatura proviene de algún estudio previo realizado sobre el mismo tema.

Ejemplo 1.4. La nota esperada. A un estudiante que acaba de presentar un examen se le puede preguntar cuál es su nota esperada. Con base en su propio conocimiento de su capacidad y de su preparación, de cómo respondió el examen, él puede tener una idea sobre la nota que espera obtener al ser calificado su examen. Obviamente la nota exacta no la conoce ya que existen múltiples factores que entran en una evaluación, pero puede proporcionar un rango dentro del cual se sienta muy seguro.

Ejemplo 1.5. Sobre una proporción. Un estudiante universitario que visite con frecuencia los distintos campus puede intentar estimar el porcentaje de mujeres que estudian en ésta. Él puede establecer valores entre los cuales, cree, cae el porcentaje de mujeres que estudian en la universidad.

Ejemplo 1.6. Porcentaje de estudiantes que consumen una sustancia psicoactiva. Si queremos determinar el porcentaje de estudiantes que consumen un tipo de sustancia psicoactiva podemos utilizar la información que se haya recogido en estudios pasados.

Ejemplo 1.7. Tasa de estudiantes que ejercen la prostitución. Si queremos determinar el porcentaje de estudiantes que ejercen la prostitución en nuestra universidad, no parece fácil resolver esto mediante una simple encuesta, aunque es posible utilizar procedimientos como el de la respuesta aleatorizada, el hecho de enfrentar un encuestador puede llevar a dar respuestas socialmente aceptables.

1.2. Probabilidad personal o subjetiva

Las ideas iniciales de la probabilidad surgieron relacionadas con los juegos de azar y su conceptualización e interpretación son básicamente frecuentistas. Esta formulación frecuentista trabaja bien en muchas situaciones, pero no en todas.

Entre otras, destacamos las tres diferentes interpretaciones que Kyburg, Jr. [55] señala que pueden considerarse respecto a la probabilidad:

1. Interpretación empírico-frecuentista. Esta es la interpretación más común de la probabilidad y hace relación al comportamiento (real o hipotético) de ciertos objetos.

2. Interpretación lógica. Esta interpretación no es común entre los estadísticos y está más bien reservada al mundo de los lógicos. De acuerdo con esta interpretación, hay una relación lógica entre una afirmación (considerada como una hipótesis) y otra afirmación (considerada como evidencia), en virtud de la cual la primera tiene cierta probabilidad relativa a la segunda. Probabilidad lógica es el grado de creencia en proposiciones, que asocian un conjunto de premisas con un conjunto de conclusiones. En la probabilidad lógica , esta relación es única. Fue De Morgan quien primero definió la probabilidad en términos de «grados de creencia»[56].

   Bajo la influencia de Bertrand Russell, Keynes adoptó una *proposición* (en lugar de un evento) «como eso que puede llevar el atributo de la probabilidad». Keynes dice que la probabilidad es *relación lógica* indefinible entre (1)

Capítulo 2

Probabilidad subjetiva «a priori»

El trabajo estadístico descansa en el concepto de probabilidad. La definición matemática es clara: es una función aditiva no negativa, cuyo máximo valor es la unidad [74]. El problema fundamental está en la forma como se determine esa función. Ashby [75] comenta «tres interpretaciones se le pueden dar a las distribuciones a priori: como distribuciones de frecuencia basadas quizá en datos previos, como representaciones normativas y objetivas de lo que es racional creer acerca de un parámetro o como una medida subjetiva de los que un individuo particular realmente cree».

2.1. Clasificación de las distribuciones a priori

$$\text{Distribuciones a priori} = \begin{cases} \text{Propias} \\ \text{Impropias} \end{cases}$$

Definición 2.1 (Distribución a priori propia). *Es una distribución que asigna pesos no negativos y que suman o integran hasta uno, a todos los valores posibles del parámetro.*

Así, una distribución propia satisface las condiciones de función de densidad de probabilidad. Una distribución impropia es la que suma o integra a un valor diferente de uno, digamos K. Si K es finito, entonces la distribución impropia induce una distribución propia normalizando la función. Si K es infinito, entonces la distribución tiene un papel de ponderación o de herramienta técnica para llegar a una distribución posterior.

$$\text{Distribuciones a priori} = \begin{cases} \text{Informativas} \\ \text{No informativas} \end{cases}$$

Definición 2.2 (Distribución a priori no informativa). *Decimos que una distribución a priori es no informativa cuando refleja una ignorancia total o un conocimiento muy limitado sobre el parámetro de interés.*

$$\text{Distribuciones a priori} = \begin{cases} \text{Conjugadas} \\ \text{No conjugadas} \end{cases}$$

Definición 2.3 (Distribución a priori conjugada). *Decimos que una distribución a priori es conjugada, si al proceder a su actualización mediante la información muestral, la distribución a posteriori es igual a la a priori, excepto en los hiperparámetros, es decir, en parámetros distintos al los del modelo muestral.*

2.2. Distribuciones a priori no informativas

En muchas ocasiones sabemos nada o muy poco acerca del parámetro de interés o no queremos involucrar en nuestro estudio información previa, sino más bien dejar que sean los datos los que «hablen por ellos mismos». En este caso la distribución debe reflejar nuestro total desconocimiento de los valores posibles del parámetro. Esta es un área de trabajo que ha crecido enormemente.

Bernardo y Ramón [76] señalan:

> Desde un punto de vista *fundamental*, la derivación de una posterior no subjetiva deberá verse como una parte de un *análisis de sensibilidad* necesario, diseñado para analizar los cambios en la posterior de interés inducidos por los cambios en la a priori: una posterior no subjetiva trata de dar una respuesta a una pregunta sobre qué puede decirse acerca de una cantidad de interés, si el conocimiento a priori de uno estuviera dominado por los datos. Cuando la información subjetiva a priori es especificada, la correspondiente posterior pudiera ser comparada entonces con la a posteiori no informativa para determinar la importancia relativa de las opiniones iniciales en la inferencia final. José Bernardo ha sido un gran investigador junto con sus colaboradores en el área de aprioris no informativas.

Un argumento adicional presentado por Bernardo y Ramón [76] sobre la bondad de usar aprioris no informativas es que la automatización de procesos bayesianos en software estadístico que utiliza ciertos procesos numéricos, estilo MCMC, se vuelven mucho más complejos e imprácticos si no se usan aprioris no informativas.

Evans [77] afirma:

> No es claro para mí por qué los bayesianos son tan insistentes con las aprioris ignorantes. Al menos en ingeniería, la idea debe ser cuantificar lo que el ingeniero conoce, no lo que él no sabe. Muchos artículos en bayesiana comienzan con una o más aprioris ignorantes. Si usted no sabe nada de antemano, qué tiene la bayesiana para ofrecer? Igualmente, si los datos van a tumbar la a priori, qué tiene la bayesiana para ofrecer?...En una situación en ingeniería, rara vez hay una excusa para escoger una a priori ignorante. Si los ingenieros no tienen una buena idea razonablemente de cómo las cosas pueden resultar, todos ellos deben ser despedidos y cada línea de producción debió pararse hace tiempo.

Capítulo 3

Análisis preposterior

Dadas las características del proceso bayesiano, es sano tener una posición crítica con relación a cualquier a priori que obtengamos por cualquier método, deberíamos realizar pruebas que permitan determinar de alguna manera la calidad de la distribución a priori. Martz y Waller [33] recomiendan lo siguiente para garantizar la realización de un buen análisis bayesiano:

1. Una justificación y análisis detallado de la distribución a priori seleccionada, con un claro entendimiento de las implicaciones matemáticas de la a priori. Por ejemplo, la selección de una familia normal para representar nuestro conocimiento a priori sobre un parámetro nos restringe a una clase de distribuciones unimodales.

2. Una documentación completa de las fuentes de datos utilizados en la identificación y selección de la a priori.

3. Un análisis preposterior de la distribución a priori con resultados de prueba hipotéticas.

4. Una distribución a posteriori claramente definida para los parámetros de interés.

5. Un análisis de sensibilidad de las inferencias bayesianas para el modelo a priori seleccionado.

Cada uno de estos puntos debe desarrollarse cuidadosamente tanto en el proceso de elicitación como en el de validación de la información recogida. El segundo paso de la lista de Martz y Waller es ejecutado casi simultáneamente en la misma elicitación. Al experto se le alienta a dar respuestas, pero el elicitado debe proporcionar una retroalimentación oportuna, lo que en términos prácticos significa que es inmediata. Para llevar a cabo esta tarea, se puede generar muestras de la población bajo diferentes condiciones planteadas en la a priori, por ejemplo, podemos tomar la mediana, la media, la moda y cuartiles para generar estas muestras. Si el sujeto se siente cómodo con los valores que las muestras están presentado y él cree que no aparecen valores muy raros, o no observa valores que considera deberían estar presentes en la población, esto no obliga a revisar la elicitación.

Es necesario documentar detalladamente el proceso en sí mismo, bien sea con grabaciones o registro fílmico, para su posterior análisis. Cuando la distribución a priori es obtenida a partir de estudios o datos previos, es necesario justificar plenamente la pertinencia de esos estudios o datos en el problema que estamos enfrentando. Un estudio realizado sobre consumo de drogas en Medellín-Colombia puede no ser útil para uno que se realice en alguna ciudad del exterior, pero ser pertinente para un estudio similar en Cali-Colombia. En otros casos puede darse la pertinencia debido a que se estudian fenómenos biológicos que puedan de alguna manera ser generalizados a poblaciones de muchos lugares, por ejemplo, si queremos determinar la duración promedio de un embarazo.

La selección de la distribución a priori puede realizarse de muchas maneras, unas más fáciles que otras. De cualquier manera es recomendable usar varios procedimientos para su elicitación, los cuales deben coincidir de alguna manera. La distribución elicitada a mano alzada debería ser muy parecida a algún modelo paramétrico ajustado, por ejemplo.

El análisis preposterior debe realizarse con la distribución a priori para verificar la confianza de los resultados obtenidos como resultantes del proceso elicitado. Por ejemplo, si se elicita el número promedio de goles realizados en un partido de fútbol, y además se asume que el número de goles marcados en un partido se puede modelar mediante la distribución de Poisson, valores de la probabilidad de observar ciertos marcadores, condicionados en valores tomados de la a priori sobre la media, deben ser consistentes con lo que el sujeto piensa. Este análisis es equivalente a un análisis exploratorio de datos (EDA), en el cual se buscan características y problemas con la distribución a priori especificada. Luego de este análisis uno podría responder afirmativamente a la siguiente pregunta: ¿estaría confiado, bajo el supuesto que no pueda obtener información muestral, para usar esta distribución para realizar todo el trabajo inferencial?

Otro punto que se analiza en esta etapa es la calidad de la a priori. ¿Es creíble? ¿Se puede utilizar fácilmente a nivel computacional? ¿Está completamente definida sobre el espacio parametral? ¿Cambia radicalmente si se obtienen unos pocos datos?

Una de los asuntos más delicados es la obtención de dos o más aprioris que sean muy diferentes por parte de diferentes expertos o por diferentes métodos de elicitación. En caso de diferencias irreconciliables, es sano trabajar paralelamente con varias aprioris. Hoy en día esto no es una gran problema, debido a la disponibilidad de recursos computacionales.

3.1. Distribución predictiva a priori

La construcción de la distribución predictiva a priori es una de las herramientas que tiene el analista para determinar la calidad de la distribución a priori elicitada, además de entregar información que puede ser útil para determinar los procedimientos de análisis posteriores. Si $\xi(\boldsymbol{\theta})$ representa la distribución a priori sobre $\boldsymbol{\theta}$ elicitada y si se conoce (obviamente hasta cierto nivel) la distribución que genera los datos, siendo $f\left(y\left|\boldsymbol{\theta}\right.\right)$ la distribución de las observaciones futuras, entonces la distribución predictiva a priori se define como,

$$p(y) = \int f\left(y\left|\boldsymbol{\theta}\right.\right)\xi\left(\boldsymbol{\theta}\right)\, d\boldsymbol{\theta}$$

Si los datos generados por esta distribución se «acomodan»a lo que el analista cree deben ser, entonces puede, de cierta manera pensar que la a priori concuerda con lo que piensa y proceder al proceso posterior.

Este análisis se puede realizar con diferentes distribuciones a priori de tal forma que se puede mirar la robustez de los resultados ante cambios de la a priori. Observemos que este análisis no es posible de realizarse con distribuciones a priori no informativas.

Ejemplo 3.1. Número promedio de goles en el fútbol colombiano. Supongamos que elicitamos el número promedio de goles que se marcan en un partido del fútbol profesional en Colombia a un experto y obtenemos que $\theta \sim N\left(2.5, (0.20)^2\right)$. Además si pensamos que el número de goles que se marcan en un partido del rentado colombiano es $Poisson(\theta)$, la distribución predictiva a priori será

$$p(y) = \int_0^\infty \frac{\theta^y \exp(-\theta)}{y!}\frac{1}{\sqrt{2\pi}0.20}\exp\left(-\frac{1}{2\times 0.20^2}(\theta-2.5)^2\right)\, d\theta$$

```
# Simulación de la Distribución Predictiva a priori
# Nro. promedio de goles del campeonato colombiano
# A priori: Normal(2.5, 0.20^2)
# Dist. muestral: Poisson(theta)
# función que muestrea de la predictiva
dist.predictiva<-function(Nsim=10000,media=2.5,dt=0.20){
# genera valores de la Poisson con parámetros de la normal
res<-rpois(Nsim,rnorm(Nsim,mean=media,sd=dt))
# tabla de frecuencias para los valores generados
tabla<-table(res)
print(tabla/sum(tabla))
barplot(tabla)  # diagrama de barras para valores generados
print(summary(res)) # resumen estadístico
print(var(res)) }   # fin de la función

dist.predictiva()
title(main='Distribución predictiva\n campeonato colombiano',
      xlab='Número de Goles', ylab='Frecuencia')
legend(7,2000,c('Media a priori=2.5','Desv. Est. a priori=0.20')
       ,bty='n')
```

Los resultados de la anterior simulación son los siguientes:

Tabla 3.1: *distribución predictiva a priori para el número de goles marcados en el fútbol colombiano*

y	0	1	2	3	4	5	6	7	8	9	10
$p(y)$	0.0849	0.2002	0.2592	0.2101	0.1320	0.0701	0.0288	0.0103	0.0032	0.0010	0.0002

Tabla 3.2: *resumen estadístico de los datos simulados de la predictiva.*

mín	Cuartil 1	Mediana	Media	Cuartil 3	máx	Varianza
0.000	1.000	2.000	2.509	3.000	10.000	2.545

A priori, notamos que la mediana del número de goles para un nuevo partido del rentado colombiano sería de 2 goles.

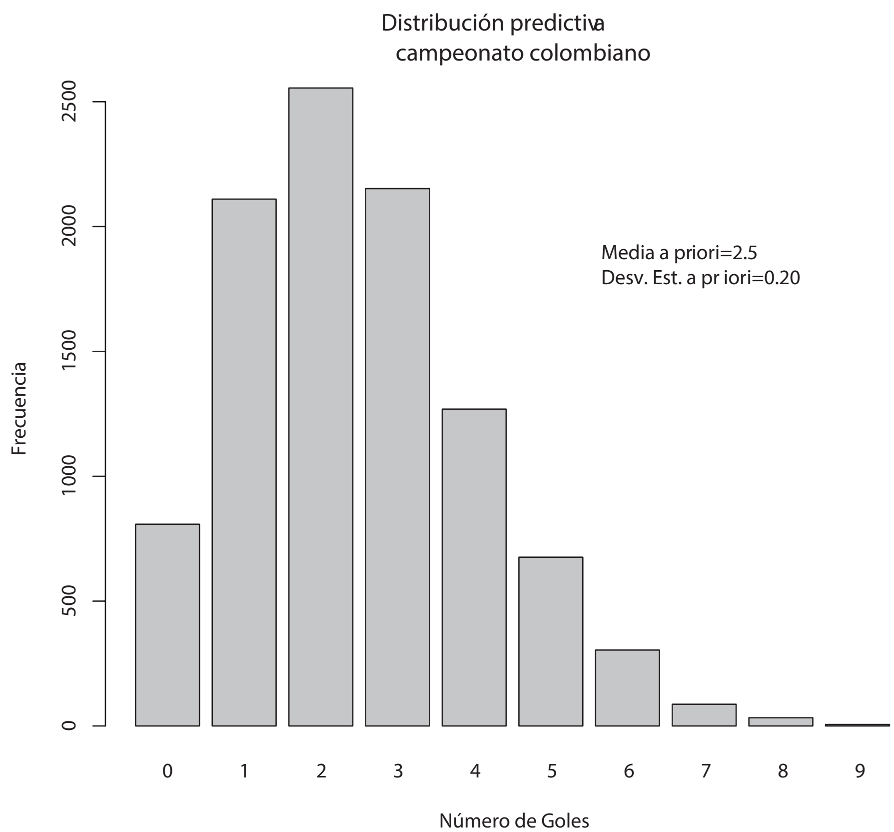

Figura 3.1: *distribución predictiva a priori. Esta distribución no es una distribución Poisson debido a nuestra propia incertidumbre sobre la media del proceso, la cual la expresamos mediante una distribución Normal*

Capítulo 13

Inferencia predictiva

Muchas situaciones aplicadas implican realizar inferencias sobre una observación futura de una variable aleatoria, cuya distribución depende de un número finito de parámetros (desconocidos), esta distribución se conoce como distribución predictiva. Smith [181] argumenta que afirmaciones predictivas acerca de variables aleatorias no observadas tiene más sentido a menudo que la estimación tradicional de parámetros.

13.1. Procedimiento exacto

Asumiendo que $\xi\left(\theta\right)$ es la distribución a priori, y que $\xi\left(\theta|\boldsymbol{x}\right)$ es la posterior, la distribución predictiva bayesiana se calcula como:

$$\begin{aligned}
p\left(z|\boldsymbol{x}\right) &= \frac{p(z,\boldsymbol{x})}{p(\boldsymbol{x})} \\
&= \frac{\int_\Theta p(z,\boldsymbol{x},\theta)\,d\theta}{\int_\Theta p(\boldsymbol{x},\theta)\,d\theta} \\
&= \frac{\int_\Theta p(z,\boldsymbol{x}|\theta)\xi(\theta)\,d\theta}{\int_\Theta p(\boldsymbol{x}|\theta)\xi(\theta)\,d\theta} \\
&= \frac{\int_\Theta p(z|\theta)p(\boldsymbol{x}|\theta)\xi(\theta)\,d\theta}{\int_\Theta p(\boldsymbol{x}|\theta)\xi(\theta)\,d\theta} \\
&= \int_\Theta p(z|\theta)\left\{\frac{p(\boldsymbol{x}|\theta)\xi(\theta)}{\int_\Theta p(\boldsymbol{x}|\theta)\xi(\theta)\,d\theta}\right\}d\theta \\
&= \int p(z|\theta)\xi\left(\theta|\boldsymbol{x}\right)\,d\theta
\end{aligned}$$

Así

$$\begin{aligned}
p\left(z|\boldsymbol{x}\right) &= \int p(z|\theta)\xi\left(\theta|\boldsymbol{x}\right)\,d\theta \\
&= E_{\theta|\boldsymbol{x}}\left[p(z|\theta)\right]
\end{aligned}$$

La función $p(z|\theta)$ es la de verosimilitud de θ evaluada en z.

Ejemplo 13.1. Caso Bernoulli. Suponga que $x_1, \ldots, x_n$ es una muestra aleatoria de una $Bernoulli(\pi)$ y suponga que la distribución a priori de π es una $Beta(\alpha, \beta)$. Encontremos la distribución predictiva de una observación futura z.

Tenemos

$$p\left(z|\boldsymbol{x}\right) = \int p(z|\pi)\xi\left(\pi|\boldsymbol{x}\right)\, d\pi$$

Ahora,

$$p(z|\pi) = \pi^z(1-\pi)^{1-z}, \quad z = 0, 1,$$

y

$$\xi(\pi|\boldsymbol{x}) \propto \pi^{\sum x_i + \alpha - 1}(1-\pi)^{n - \sum x_i + \beta - 1}$$

Ahora, si denotamos por $\alpha^* = \sum x_i + \alpha$ y $\beta^* = n - \sum x_i + \beta$ tenemos que:

$$\begin{aligned}
p(z|\boldsymbol{x}) &= \int_0^1 \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha^*)\Gamma(\beta^*)}\pi^{z+\alpha^*-1}(1-\pi)^{\beta^*+1-z-1}\, d\pi \\
&= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha^*)\Gamma(\beta^*)}\frac{\Gamma(z+\alpha^*)\Gamma(1-z+\beta^*)}{\Gamma(n+\alpha+\beta+1)}
\end{aligned}$$

Así,

$$\begin{aligned}
P(z=0|\boldsymbol{x}) &= \frac{\Gamma(n+\alpha+\beta)\Gamma(1+\beta^*)}{\Gamma(\beta^*)\Gamma(n+\alpha+\beta+1)} \\
&= \frac{\beta^*}{n+\alpha+\beta} \\
&= \frac{\beta^*}{\alpha^*+\beta^*},
\end{aligned}$$

y

$$P(z=1|\boldsymbol{x}) = \frac{\alpha^*}{\alpha^*+\beta^*}$$

Vale la pena notar que:

$$P(z=1|\boldsymbol{x}) = E(\pi|\boldsymbol{x}),$$

es la media posterior.

Ejemplo 13.2. Caso Poisson. Suponga que $x_1, \ldots, x_n$ es una muestra aleatoria de una $Poisson(\theta)$. Además supongamos que la distribución a priori de θ es una $Gamma(\alpha, \beta)$. Encontremos la distribución predictiva $p(z|\boldsymbol{x})$.

Sabemos que la distribución a posteriori es una $Gamma(\alpha^* = \alpha + \sum x_i, \beta^* = \beta + n)$. Ahora,

$$p(z|\boldsymbol{x}) = \frac{\theta^z e^{-\theta}}{z!}$$

Así,

$$\begin{aligned} p(z|\boldsymbol{x}) &= \int_0^\infty \frac{\theta^z e^{-\theta}}{z!} \frac{(\beta^*)^{\alpha^*}}{\Gamma(\alpha^*)} e^{-\beta^* \theta}\, d\theta \\ &= \frac{(\beta^*)^{\alpha^*}}{z!\Gamma(\alpha^*)} \int_0^\infty \theta^{z+\alpha^*-1} e^{-(\beta^*+1)\theta}\, d\theta \\ &= \frac{(\beta^*)^{\alpha^*}}{z!\Gamma(\alpha^*)} \frac{\Gamma(z+\alpha^*)}{(\beta^*+1)^{(z+\alpha^*)}} \\ &= \binom{z+\alpha^*-1}{z} \left(\frac{\beta^*}{\beta^*+1}\right)^{\alpha^*} \left(\frac{1}{\beta^*+1}\right)^z \end{aligned}$$

para $z = 0, 1, 2, \ldots$ Por tanto,

$$z|\boldsymbol{x} \sim Binomial-Negativa\left(\alpha^*, \frac{1}{\beta^*+1}\right)$$

Ejemplo 13.3. Caso Exponencial. Sea $x_1, \ldots, x_n$ una muestra aleatoria de una exponencial con densidad $\theta e^{-\theta x}$, con $x > 0$, $\theta > 0$. Sea Z que denota una observación futura de la misma densidad. Estamos interesados en la probabilidad predictiva que $Z > z$ para algún nivel dado z. Cuando θ es conocido, esto está dado por $\phi = \phi(z|\theta) = e^{-\theta z}$.

Asumiendo que la distribución a priori de θ es $\xi(\theta) \propto \theta^{a-1} e^{-b\theta}$, una a priori Gamma con parámetros (a, b). La distribución a posteriori de θ es también una Gamma con parámetros $(a+n, b+S_n)$, donde $S_n = x_1 + \ldots + x_n$, y la esperanza posterior de ϕ se calcula como:

$$\hat{\phi} = \left(\frac{b+S_n}{b+S_n+z}\right)^{a+n}$$

Cuando $a = b = 0$ se tiene una distribución a priori Jeffreys y la esperanza se reduce a

$$\hat{\phi} = \left(\frac{S_n}{S_n+z}\right)^n$$

Ejemplo 13.4. Distribución Multinomial. En el caso de la distribución multinomial tenemos, bajo una a priori Dirichlet, la a posteriori es también Dirichlet con parámetros $n_i + \alpha_i$, para $i = 1, \ldots, k$. Bajo la distribución a priori de Jeffreys, que corresponde a una Dirichlet con $\alpha_i = 1/2$ para todo $i = 1, \ldots, k$, la distribución predictiva es:

$$p\left(X_i = i\,|\boldsymbol{N}\,\right) = \frac{n_i + \frac{1}{2}}{\sum_{j=1}^{k} n_j + \frac{k}{2}},$$

y, bajo a a priori uniforme,

$$p\left(X_i = i\,|\boldsymbol{N}\,\right) = \frac{n_i + 1}{\sum_{j=1}^{k} n_j + k}$$

13.2. Distribución predictiva vía MCMC

- A veces es difícil resolver la integral para calcular la distribución predictiva

$$\begin{aligned} p\left(z|\boldsymbol{x}\right) &= \int p(z|\theta)\xi\left(\theta|\boldsymbol{x}\right)\, d\theta \\ &= E_{\theta|\boldsymbol{x}}\left[p(z|\theta)\right] \end{aligned}$$

- Una solución es usar MCMC.

13.2.1. Algoritmo

- *(Paso 1)* Genere una muestra de tamaño M, luego de haber quemado n_B muestras de $\xi\left(\theta\,|Datos\,\right)$, puede usar un thin (botar valores intermedios si es necesario para controlar la autocorrelación). Esta muestra la denotamos por comodidad como:

$$\theta_1, \theta_2, \ldots, \theta_M$$

- La distribución predictiva $p\left(z\,|Datos\,\right)$ podemos aproximarla así:

$$\begin{aligned} p\left(z|\boldsymbol{x}\right) &= \int p(z|\theta)\xi\left(\theta|\boldsymbol{x}\right)\, d\theta \\ &= E_{\theta|\boldsymbol{x}}\left[p(z|\theta)\right] \approx \frac{1}{M}\sum_{i=1}^{M} p\left(z\,|\theta_i, Datos\,\right) \end{aligned}$$

- *(Paso 2)* Sacamos al azar un número en $\{1, 2, \ldots, M\}$ con probabilidad $1/M$, digamos m.

- *(Paso 3)* De $p(z|\theta_m, Datos)$ sacamos un número al azar, digamos z.

- *(Paso 4)* Repetimos los pasos 2 y 3 una gran cantidad de veces, digamos K. Al final obtenemos un conjunto de valores

$$z_1, z_2, \ldots z_K$$

- *(Paso 5)* Construimos un estimador de la densidad $p(z|Datos)$. Si z es discreta simplemente calculamos la densidad aproximada como:

$$p(z = j|Datos) \approx \frac{\#\left\{x_k = j\right\}}{K}$$

Ejemplo 13.5. Distribución Discreta. Asumamos:

- $X \sim Poisson(\lambda)$
- $\xi(\lambda)$ es $U(0,3)$
- $x_1, x_2, \ldots, x_n$ es una m.a. de la distribución $Poisson(\lambda)$
- La distribución a posteriori es:

$$\xi\left(\lambda \,|Datos\,\right) = \frac{\lambda^{\sum_{i=1}^{n} x_i} \exp(-n\lambda)}{\prod_{i=1}^{n} x_i!},$$

  para $0 < \lambda < 3$.

- La distribución predictiva de z dado los Datos es:

$$p(z\,|Datos\,) = \int_0^3 \frac{\lambda^z \exp(-\lambda)}{z!} \frac{\lambda^{\sum_{i=1}^{n} x_i} \exp(-n\lambda)}{\prod_{i=1}^{n} x_i!}\, d\lambda$$

$$p(z\,|Datos\,) = \frac{1}{z! \prod_{i=1}^{n} x_i!} \int_0^3 \exp(-\lambda(n+1))\lambda^{z+\sum_{i=1}^{n} x_i}\, d\lambda$$

$$\begin{aligned} p(z\,|Datos\,) &= \frac{1}{z! \prod_{i=1}^{n} x_i!} \frac{\Gamma\left(z + \sum_{i=1}^{n} x_i + 1\right)}{(n+1)^{z+\sum_{i=1}^{n} x_i + 1}} \\ &\times \int_0^3 \frac{(n+1)^{z+\sum_{i=1}^{n} x_i + 1}}{\Gamma\left(z + \sum_{i=1}^{n} x_i + 1\right)} \exp(-\lambda(n+1))\lambda^{z+\sum_{i=1}^{n} x_i}\, d\lambda \end{aligned}$$

Esta última integral corresponde a la función de distribución acumulada de una gamma con parámetros $z + \sum_{i=1}^{n} x_i + 1$ y $n + 1$ evaluada en 3.

Si observamos del proceso 0,0,2,1,2,0,0,2,2,1,1,1,3,4,4,3. Tenemos

```
# Cálculo de la distribución predictiva
# Distr. muestral: Poisson

# A priori: U(0,3)
Datos<-c(0,0,2,1,2,0,0,2,2,1,1,1,3,4,4,3)
p.pred<-function(z,x){
n<-length(x); S.x<-sum(x)
P.x<-prod(factorial(x))
a<-z+S.x+1; b<-n+1
res<-gamma(a)/(factorial(z)*b^a*P.x)*pgamma(3,a,rate=b)
return(res)
}

temp<-p.pred(0:20,Datos)
prob.poste<-temp/sum(temp)
plot(0:20,prob.poste,type='h')
prob.poste
```

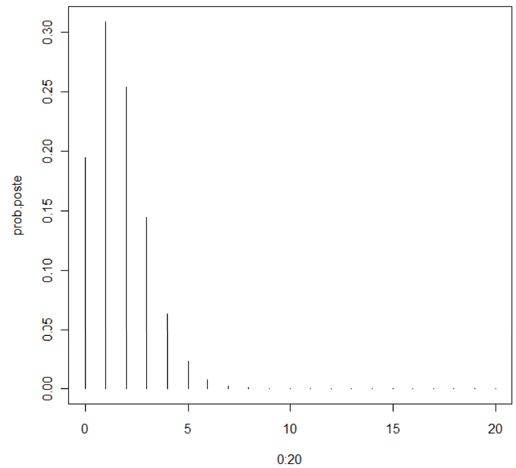

Figura 13.1: *distribución de probabilidad a posteriori de los datos*

Ejemplo 13.6. Distribución continua.

- Suponga $X \sim Gamma(\alpha, \beta)$
- Distribución a priori
$$\xi(\alpha, \beta) \propto 1$$
- Distribución posterior
$$\xi\left(\alpha, \beta \,|Datos\right) \propto \frac{\beta^{n\alpha}}{\left(\Gamma(\alpha)\right)^n} \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} \exp\left(-\beta \sum_{i=1}^{n} x_i\right)$$

```
 # Distribución predictiva para una va continua Gamma
 tiempos<-c(1.2,0.5,1.6,2.0,2.1,2.0)
 prod.tiempos<-prod(tiempos)
 sum.tiempos<-sum(tiempos)
 n<-length(tiempos)
 u<-mean(tiempos)
 v<-var(tiempos)
 a<-u^2/v
 b<-u/v

a
[1] 6.347701
b
[1] 4.051724

veros<-function(a,b,datos){
 res<-1
 for(i in 1:length(datos))
 res<-res*dgamma(datos[i],a,rate=b)
 return(res)
```

```
}

a1<-seq(0.01,16.0,length=50); b1<-seq(0.01,10.0,length=50)
z<-outer(a1, b1, FUN="veros", tiempos)
contour(a1,b1,z,ylab=expression(beta),xlab=expression(alpha))

dist.a.con<-function(a,b,produ,n) exp(n*a*log(b)
                       -n*lgamma(a)+a*log(produ))

# dist.b.con es una gamma(n*a+1,sum.tiempos) # Proceso de muestreo

a.viejo<-a ;b.viejo<-b
result<-c(a,b)
resulta<-matrix(NA,ncol=2,nrow=10000)

for(i in 1:nrow(resulta)){
pesos<-dist.a.con(a1,b.viejo,prod.tiempos,n)
a.nuevo<-sample(a1,1,prob=pesos)
b.nuevo<-rgamma(1,n*a.nuevo+1,sum.tiempos)
resulta[i,]<-c(a.nuevo,b.nuevo)
b.viejo<-b.nuevo }

points(resulta,col='grey')
par(mfrow=c(2,1))
plot(resulta[,1],type='l',ylab=expression(alpha))
plot(resulta[,2],type='l',ylab=expression(beta))
par(mfrow=c(1,1))

# Función que genera muestra de la predictiva

genera.muestra.predictiva<-function(a)rgamma(1,a[1],rate=a[2])
z<-apply(resulta,1,genera.muestra.predictiva)
plot(density(z,from=0),main='Distribución Predictiva')
```

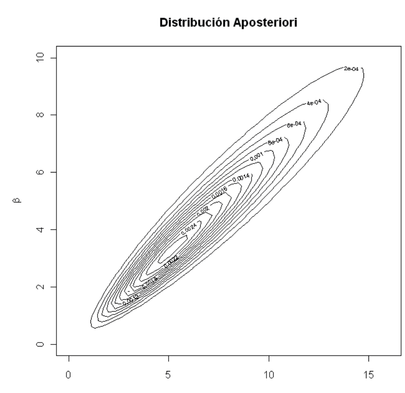

Figura 13.2: *distribución a posteriori para* α *y* β

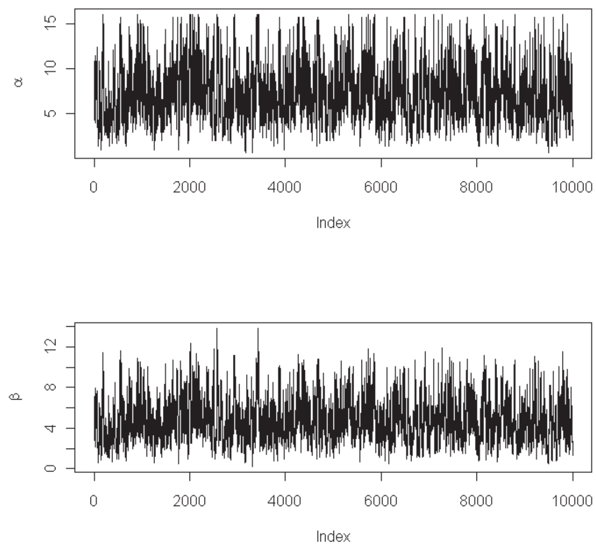

Figura 13.3: *cadenas generadas para* α *y* β

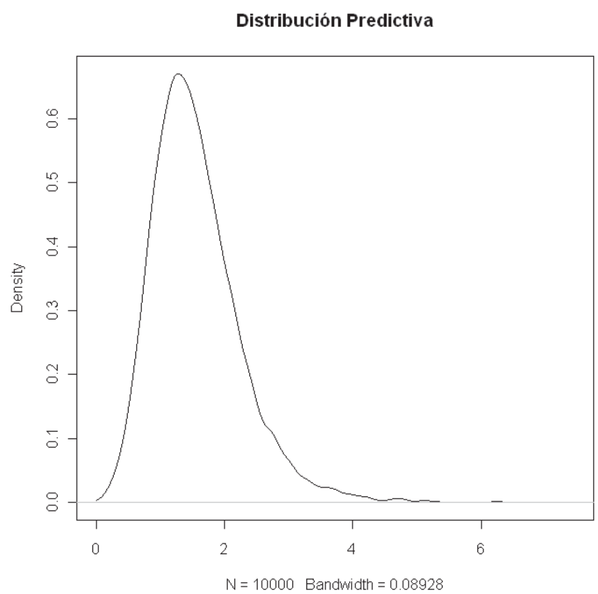

Figura 13.4: *distribución predictiva para una variable aleatoria gamma*

Capítulo 14

Software para estadística bayesiana

14.1. Estadística bayesiana en R

R ha llegado a ser un estándar en el trabajo estadístico aplicado. Su facilidad de extensión por parte de los usuarios lo hace ideal para crear funciones y librerías que ejecuten tareas específicas. Entre las librerías para trabajo bayesiano tenemos:

Librería	Descripción
MCMCpack	Paquete para realizar modelación bayesiana tradicional.
mcmc	Simulación vía MCMC para funciones del usuario.

14.1.1. Librería MCMCpack

Una librería que es muy fácil de utilizar y que posee una amplia base de funciones para ajustar muchos modelos tradicionales.

MCMCregress

Esta función permite ajustar un modelo lineal bayesiano conjugado. Por defecto trabaja con distribuciones a priori no informativas. Esta función asume que la distribución generadora de datos es normal y la distribución de los parámetros es Normal-Gamma.

MCMClogistic

Esta función ajusta modelos de regresión logísticos. Asume que la variable respuesta es 0 o 1. A diferencia de la función $glm()$, no permite considerar respuestas en forma de tabla con columnas fracasos-éxitos.

14.2. Tutorial sobre OpenBUGS

Uno de los inconvenientes que han tenido los métodos bayesianos para ser utilizados en la práctica ha sido la carencia de software especializado. Ninguno de los grandes paquetes en estadística, SAS, SPSS, etc., tienen módulos para hacer estadística bayesiana.

Existe un programa de acceso gratuito al público que permite utilizar simulación estadística basada en cadenas de Markov en una forma simple y efectiva para gran variedad de modelos llamado *BUGS*, que es un acrónimo de *Bayesian analysis Using the Gibbs Sampler* (Muestreador Gibbs, que lo veremos en un capítulo posterior). Este programa está disponible en:

`http://www.openbugs.net/w/FrontPage`

Hay dos versiones de BUGS, WinBUGS y OpenBUGS. La segunda, trabaja bajo Windows con una interfaz muy similar a la de WinBUGS y con una interfaz de texto plano bajo Linux. Además, puede ser llamado desde R para ejecutar el modelo y analizar las cadenas generadas.

Existen otros programas que permiten resolver problemas bayesianos como el BACC, First Bayes, JAGS (Just Another Gibbs Sampler), etc. El *R* trae algunas librerías con soluciones a ciertos problemas específicos, por ejemplo la MCMCPack y CODA.

14.3. ¿Qué se espera de un software para estadística bayesiana?

Koop [182] señala algunos requisitos claves que todo software bayesiano debería cumplir:

1. Debe ser computacionalmente eficiente
2. Debe estar bien documentado
3. El grupo de soporte debe ser amplio y reconocido
4. Debe proporcionar simuladores posteriores para la clase de modelos que los investigadores quieran usar
5. Para los modelos no incluidos, debe ser fácil la inclusión de los simuladores posteriores que se necesitan por parte del usuario
6. Debe tener una base amplia de funciones $g(\theta)$
7. Debe proporcionar medidas del error en la aproximación para las estimadas de $E\left(g(\theta)|Y\right)$ y las verosimilitudes marginales
8. Debe permitir al usuario graficar la a posteriori y la a priori
9. Debe permitirle al usuario realizar un análisis de sensibilidad a priori de una manera fácil
10. Todo lo anterior debe poderse llevar a cabo de una manera simple, transparente y conveniente para el usuario

14.3.1. Utilización de WinBUGS y OpenBUGS

La utilización por primera vez del programa puede ser una experiencia extraña, ya que el programa no funciona en una forma lineal, sino que requiere múltiples pasos que pueden parecer repetitivos, pero que en realidad no lo son.

En $WinBUGS$ y *OpenBUGS* el símbolo $\sim$ significa «distribuido como»y se utiliza para:

- Especificar la distribución de los datos
- Especificar la distribución a priori

Los valores a la izquierda de $\sim$ son llamados «estocásticos».

La flecha (conformada por dos símbolos) a la izquierda `<-` se utiliza como el igual. Por ejemplo `var <- 1/precision` . Los valores a la izquierda de `<-` son llamados «lógicos».

Los pasos en el programa para correr un modelo son:

1. Los comandos anteriores los escribimos en una ventana que abrimos seleccionando *File* y luego *New*. Si usted ya tiene algún archivo con un programa creado y salvado con anterioridad en formato *.odc* puede abrirlo para trabajar con él.

   Si seleccionamos *New* el programa muestra una ventana en blanco en la cual podemos escribir los comandos apropiados, como los que se encuentran enseguida. Con el cursor seleccionamos toda la parte correspondiente al modelo y seleccionamos *Edit* y luego *Copy*.

   ```
   model {
    # likelihood
      n.ij ~ dbin(n, P.ij)
    # prior
      P.ij ~ dbeta(nu, omega)
   } list(k = 201, n = 372, nu = 1, omega = 1)
   ```

   Ejemplo con la longitud máxima del pie de estudiantes universitarios:

   ```
   model
           {
               for(i in 1:N){
                   Y[i] ~ dnorm(mu[i], tau)
                   mu[i] <- beta
               }
               sigma <- 1/sqrt(tau)
               beta ~ dnorm(0, 1.0E-6)
               tau ~ dgamma(1.0E-3, 1.0E-3)
           }
   ```

El programa $WinBUGS$ permite utilizar un lenguaje conciso para expresar un modelo: β y τ son expresados con distribuciones a priori propias pero lo más mínimo informativas que se pueda, mientras que la expresión lógica sigma permite que la desviación estándar sea estimada.

2. Primero seleccionamos el menú *Model*.

3. Abrimos la herramienta *Specification*. Aquí nos aparece una ventana con varias opciones.

4. Señalamos la palabra *check model* en el comienzo de la descripción del modelo. Necesitamos chequear que la descripción del modelo define completamente un modelo de probabilidad. Si el modelo fue especificado correctamente aparece el mensaje *model is syntactically correct* en la parte inferior izquierda de la ventana principal. Sino, nos aparece el tipo de error que tenemos en el modelo.

5. Luego señalamos los datos (los cuales deben estar en un formato especial, estilo $S - Plus$) y los copiamos con *Edit* y luego *Copy*.

6. Nuevamente nos vamos a la ventana *Specification Tool* y seleccionamos *load data*. Si los datos están conformes al modelo, aparece un mensaje en la parte inferior izquierda de la ventana principal donde se informa que los datos fueron cargados. (Estos datos pueden estar copiados en la misma ventana en la cual escribimos nuestro modelo. Lo que hacemos es señalarlos y copiarlos y luego oprimimos el cuadro *load data*).

```
list(Y = c(24.2,25.4,25.0,25.9,25.5,24.4), N = 6)
```

7. El siguiente paso se ejecuta en la ventana *Specification Tool* y seleccionamos *compile*.

8. A continuación en la ventana *Specification Tool* seleccionamos *load inits*. Los valores iniciales para el proceso iterativo (Estos valores iniciales pueden estar copiados también en la misma ventana en la cual escribimos nuestro modelo y los datos. Lo que hacemos es señalarlos y copiarlos y luego oprimimos el cuadro *load inits*).

```
list( beta = 0, tau = 1)
```

   Otra opción nos permite que el programa genere automáticamente valores iniciales, esto lo hace generando números aleatorios de la distribución a priori. El programa permite correr más de una cadena simultáneamente, para lo cual se necesita especificar más de un conjunto de valores iniciales.

9. Del menú *model* seleccione *Update...* y del menú *Inference* seleccione *Samples*. Ahora usted tiene dos nuevas ventanas, una con el nombre *Update Tool* y la otra con el nombre *Sample Monitor Tool*.

La ventana *Update Tool* nos permite generar muestras. En MCMC usualmente hay que dejar correr el muestreador durante algún tiempo (quizá 1000 iteraciones) para asegurarnos de que el proceso está estable antes de guardar valores.

Después de una corrida inicial nos ubicamos en la ventana *Sample Monitor Tool*. Para empezar escribimos los nombres de los nodos (parámetros) que queremos estudiar. Escribimos en la parte de *node* `beta` y seleccionamos luego *set*. Procedemos igual con `tau`.

10. De la ventana *Update Tool* seleccionamos la opción *update*. Esto lo podemos realizar tantas veces como sea necesario para que el proceso converja.

11. De la ventana *Sample Monitor Tool* seleccionamos ya lo que sea de nuestro interés. Por ejemplo, seleccionamos un nodo, digamos `beta` y luego *stats*, nos aparece una nueva ventana con algunos resultados de interés acerca de este parámetro. Lo mismo hacemos para `tau`.

node	mean	sd	MC error	2.5 %	median	97.5 %	start	sample
beta	25.06	0.3443	0.006615	24.34	25.06	25.75	1	3000
sigma	0.778	0.3284	0.0068	0.4095	0.698	1.589	1	3000

Ejemplo 14.1. El caso normal con varianza conocida. Supongamos que tenemos una muestra aleatoria de una normal y asumimos que su varianza es conocida e igual a 1.

```
model normalIID \{ for(in 1:10) \{y[i] ~ dnorm(mu,1) \}

# Distribucion a priori
   mu~dnorm(0,1)
\}

list(y=c(1.84,-0.23,1.12,0.35,-0.24,
            -0.89,1.65,-1.01,2.01,1.12))
```

14.3.2. Algunos de los comandos de WinBUGS y OpenBUGS

Model

Specification. Este comando activa una ventana llamada **Specification Tool**, que nos permite definir completamente el modelo.

check model: chequea el modelo.

load data: carga los datos.

compile: compila el modelo.

load inits: carga los valores iniciales.

gen inits: genera los valores iniciales en caso de que el usuario no los especifica con anterioridad.

num of chains: especifica el número de cadenas a generar.

for chain: establece el salto por cadenas usadas.

Update. Este comando se activa una vez el modelo ha sido compilado e inicializado. El produce la ventana **Update Tool** con los siguientes comandos:

updates: número de actualizaciones MCMC a ser llevadas a cabo.

refresh: el número de actualizaciones entre re-actualizaciones de la pantalla.

thin: las muestras de cada k-ésima iteración será guardada, donde k es el valor de **thin**. Hacer $k > 1$ puede ayudar a reducir la autocorrelación en la muestra.

update: cliquear para comenzar a actualizar el modelo.

over relax: esta selección permite trabajar con una versión más relajada del MCMC.

adapting: esta selección permite un proceso de adaptación inicial para un mejor ajuste de los parámetros. Toda la información generada en este proceso es descartada.

Inference

La opción **Inference** tiene varias opciones, pero la más importante es **Samples**.

Samples...: bajo este comando aparece una ventana con título **Sample Monitor Tool**. Contiene los siguientes campos:

- **node:** se especifica el parámetro o variable de interés para el análisis.
- **chains:** se pueden seleccionar las cadenas con las que se construirán los estadísticos.
- **to:** opera junto con el comando anterior.
- **beg:** cuando se utiliza una submuestra para el análisis este comando nos indica desde dónde empezamos a utilizar los valores originales. Marca el comienzo de la submuestra.
- **end:** marca el final de la submuestra que se inició con el comando anterior.
- **thin:** las muestras de cada k-ésima iteración será utilizada para los estadísticos a producir, donde k es el valor de **thin**.
- **percentiles:**
- **clear:** remueve cualquier valor guardado de las variables.
- **set:** debe utilizarse para empezar a guardar los valores para una variable.
- **trace:** presenta una gráfica del valor de la variable contra el número de la iteración. La traza es dinámica y se está reactualizando.
- **history:** grafica la traza completa para la variable.

density: presenta un gráfico de densidad para la variable si es continua, o un histograma si es discreta,

stats: produce un resumen estadístico para la variable.

coda: produce una representación ASCII del proceso para ser reanalizada con *CODA*.

quantiles:

GR diag: calcula el estadístico para convergencia de Gelman-Rubin.

autoC: grafica la función de autocorrelación de variable hasta un rezago de 50.

Fit...: Fit Tool

Referencias

[1] R. Schlaifer. *Probability and Statistics for Business Decisions: An Introduction to Managerial Economics Under Uncertainty*. McGraw-Hill, New York, 1959.

[2] H. Chernoff and L. E. Moses. *Teoría y Cálculo Elemental de las Decisiones*. Compañía Editorial Continental, S.A., México, D.F., 1959.

[3] H. Raiffa and R. Schlaifer. *Applied Statistical Decision Theory*. Harvard University Press, Boston, 1964.

[4] J. W. Pratt, H. Raiffa, and R. Schlaifer. *Introduction to Statistical Decision Theory*. McGraw-Hill, New York, 1964.

[5] B. W. Morgan. *An Introduction to Bayesian Statistical Decision Processess*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1968.

[6] H. Raiffa. *Decision Analysis: Introductory Lectures on Choice Under Uncertainty*. Addison-Wesley: Reading, Masschusetts, 1970.

[7] M. H. DeGroot. *Optimal Statistical Decisions*. McGraw Hill, New York, 1970.

[8] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. World Scientific Publishing Co., New York, 2 edition, 1985.

[9] H. Jeffreys. *Theory of Probability*. Clarendon Press, Londres, 3 edition, 1961.

[10] L. J. Savage. *Subjective Probability and Statistical Practice. En The Foundations of Statistical Inference, PUBLISHER = G.A. Barnard y D. R. Cox. John Wiley & Sons, YEAR = 1962, edition = , ADDRESS = Londres*.

[11] A. Zellner. *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons, New York, 1971.

[12] R. L. Winkler. *An Introduction to Bayesian Inference and Decision*. Holt, Rinehart and Winston,Inc., New York, 1972.

[13] G. E. P. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, New York, 1973.

[14] S. J. Press. *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*. Dover: Mineola, New York, 2 edition, 1982.

[15] S. J. Press. *Bayesian Statistics: Principles, Models, and Applications*. John Wiley & Sons, New York, 1989.

[16] P. M. Lee. *Bayesian Statistics: An Introduction*. Arnold, Londres, 2 edition, 1997.

[17] T. Leonard and J. S. Hsu. *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researches*. Cambridge University Press, Cambridge, UK, 1999.

[18] C. P. Robert. *The Bayesian Choice: A Decision-Theoretic Motivation*. Springer-Verlag, New York, 1994.

[19] C. P. Robert. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer, New York, 2 edition, 2001.

[20] B. P. Carlin and T. A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC, Boca Raton, 2 edition, 2000.

[21] P. Congdon. *Bayesian Statistical Modelling*. John Wiley & Sons: Chichester, UK, 2001.

[22] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, 2 edition, 2004.

[23] W. M. Bolstad. *Introduction to Bayesian Statistics*. John Wiley & Sons: Hoboken, New Jersey, 2004.

[24] J. K. Ghosh, M. Delampady, and T. Samanta. *An Introduction to Bayesian Analysis: Theory and Mtehods*. Springer, New York, 2006.

[25] J. K. Kruschke. *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Academic Press, 2000.

[26] M. West and J. harrison. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York, 1989.

[27] G. Koop. *Bayesian Econometrics*. Wiley: West Sussex, England, 2003.

[28] G. G. Woodworth. *Biostatistics: A Bayesian Introduction*. John Wiley & Sons, Hoboken, New Jersey, 2004.

[29] D. J. Spiegelhalter, K. R. Abrams, and J. P. Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons, Hoboken, 2004.

[30] A. B. Lawson. *Disease Mapping with WinBUGS and MLwiN*. John Wiley & Sons, Chichester, UK, 2003.

[31] J. Gill. *Bayesian Methods: A Social and Behavioral Sciences Approach*. Chapman & Hall/CRC, Boca Raton, 2002.

[32] S. M. Lynch. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer, New York, 2007.

[33] H. F. Martz and R. A. Waller. *Bayesian Reliability Analysis*. Wiley, New York, 1982.

[34] M. S. Hamada, A. G. Wilson, C. S. Reese, and H. F. Martz. *Bayesian Reliability*. Springer, New York, 2008.

[35] P. E. Rossi, G. M. Allenby, and McCulloch. *Bayesian Statistics and Marketing*. John Wiley & Sons, Chichester, UK, 2005.

[36] K. Yuen. *Bayesian Methods for Structural Dynamics and Civil Engineering*. John Wiley & Sons, Singapore, 2010.

[37] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, 1999.

[38] B. A. Berg. *Markov Chain Monte Carlo Simulations and Their Statistical Analysis With Web-Based Fortran Code*. World Scientific Publishing Co., New Jersey, 2004.

[39] J. Albert. *Bayesian Computation With R*. Springer, Nueva York, 2007.

[40] J. Marin and C. P. Robert. *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer, New York, 2007.

[41] I. Ntzoufras. *Bayesian Modeling Using WinBUGS*. John Wiley & Sons, Hoboken, New Jersey, 2009.

[42] C. P. Robert and G. Casella. *Introducing Monte Carlo Methods with R*. Springer, New York, 2010.

[43] F. J. Samaniego. *A Comparison of the Bayesian and Frequentist Approaches to Estimations*. Springer, New York, 2010.

[44] D. Qin. Bayesian econometrics: The first twenty years. *Econometric Theory*, 12(3):500–516, 1996.

[45] D. Poole, A. Mackworth, and R. Goebel. *Computational Intelligence: A Logical Approach*. Oxford University Press, New York, 1998.

[46] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/, Austria, 2016.

[47] A. D. Martin and K. M. Quinn. Applied bayesian inference in r using mcmcpack. *R News*, 6(1):2–7, 2006.

[48] A. D. Martin, K. M. Quinn, and J. H. park. Mcmcpack: Markov chain monte caro in r. *Journal of Statistical Software*, 42(9):`http://www.jstatsoft.org/`, 2011.

[49] R. V. Hogg and A. T. Craig. *Introduction to Mathematical Statistics*. Collier MacMillan International, New York, 4 edition, 1978.

[50] A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill Kogasakua, Ltd, Tokyo, 3 edition, 1986.

[51] S. Greenland. Putting background information about relative risks into conjugate prior distributions. *Biometrics*, 57(3):663–670, 2001.

[52] Z. Dienes. Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, 6(3):274–290, 2011.

[53] J. Albert. *MATLAB as an Enviroment for Bayesian Computation*. Dept. of Math. and Statistics. Bowling Green State University, 1997.

[54] B. Western and S. Jackman. Bayesian inference for comparative research. *The American Political Science Review*, 88(2):412–423, 1994.

[55] H. E. Kyburg Jr. Probability and decision. *Philosophy of Science*, 33(3):250–261, 1966.

[56] G. K. Chacko. *Decision-Making under Uncertainty*. Praeger Pub., New York, 1991.

[57] J. Franklin. *The Science of Conjecture: Evidence and Probability Before Pascal*. The Johns Hopkins University Press, Baltimore, Maryland, 2001.

[58] D. S. Sivia. *Data Analysis: A Bayesian Tuorial*. Oxford University Press, Oxford, 1996.

[59] T. S. Wallsten and D. V. Budescu. Encoding subjective probabilities: A psychological and psychometric review. *Management Science*, 29(2):151–173, 1983.

[60] B. M. Ayyub. *t Opinions for Uncertainty and Risks. Elicitation of Exper*. CRC Press, Boca Raton, 2001.

[61] J. O. Berger. *Bayesian Analysis: A Look at Today and Thoughts of Tomorrow. Technical Report*. Duke University, 1999.

[62] T. A. Loredo. *From Laplace to Supernova SN 1987A: Bayesian Inference in Astrophysics*. Fougeré. In Maximum Entropy Bayesian Methods. Kluwer Acadaemic Publishers, Dordrecht, 1990.

[63] A. G. Sawyer and J. P. Peter. The significance of statistical significance tests in marketing research. *Journal of Marketing Research*, 20(2):122–133, 1983.

[64] S. Labovitz. The nonutility of significance tests: The significance of tests of significance reconsidered. *The Pacific Sociological Review*, 13(3):141–148, 1970.

[65] B. Lecoutre, M. Lecoutre, and J. Poitevineau. Uses, abuses and misuses of significance tests in the scientific community: Won't the bayesian choice be unavoidable? *International Statistical Review / Revue Internationale de Statistique*, 69(3):399–417, 2001.

[66] F. E. Harrel Jr. *An Introduction to Bayesian Methods with Clinical Applications*. Dept. of Health Evaluation Sciences. School of Medicine, University of Virginia, Charlottesville, 1998.

[67] J. O. Berger and R. L. Wolpert. *The Likelihood Principle.* Institute of Mathematical Statistics: Hayward, California, 2 edition, 1998.

[68] Y. Pawitan. A reminder of the fallibility of the wald statistic: Likelihood explanation. *The American Statistician*, 54(1):54–56, 2000.

[69] L. J. Savage. *The Foundations of the Statistical Inference*. Methuen & Co., Londres, 1962.

[70] R. M. Royall. The effect of sample size on the meaning of significance tests. *The American Statistician*, 40(4):313–315, 1986.

[71] F. M. Rosekrans. Statistical significance and reporting test results. *Journal of Marketing Research*, 6(4):451–455, 1969.

[72] R. L. Winkler. Why bayesian analysis has´nt caught on in healthcare decision making. *International Journal of Technology in Health Care*, 17(1):56–76, 2001.

[73] M. J. Bayarri, M. H. DeGroot, and J. B. Kadane. *Statistical Decision Theory and Realted Topics IV, Gupta, S. S. and Berger, J. O., Eds.* Springer-Verlag, New York, 1988.

[74] S. Jackman. *Bayesian Modelin in the Social Sciences: an Introduction to Markov-Chain Monte Carlo. Technical Report*. Dept. of Political Science, Stanford University, 1999.

[75] D. Ashby. Bayesian statistics in medicine: a 25 year review. *Statistics in Medicine*, 25:3589–3631, 2006.

[76] J. O. Berger and D. Sun. Bayesian analysis for the poly-weibull distribution. *Journal of the American Statistical Association*, 88(424):1412–1418, 1993.

[77] R. A. Evans. *Bayes*. Theory & Practice: The Theory and Applications of Reliability With Emphasis on Bayesian and Nonparametric Methods, New York, 2 edition, 1987.

[78] I. Horowitz. *Introducción al Análisis Cuantitativo de los Negocios*. Ediciones del Castillo, Madrid, 1968.

[79] D. J. Poirier. Frequentist and subjectivist perspectives on the problemof model building in economics. *The Journal of Economic Perspectives*, 2(1):121–144, 1988.

[80] G. Shafer. Probability judgment in artificial intelligence and expert systems. *Statistical Science*, 2(1):3–16, 1987.

[81] D. R. Cox. Some remarks on model criticism. *Methods and Models in Statistics: In Honour of Professor John Nelder, FRS. edited by Adams, N. et al., Imperial College Press: London*, pages 13–21, 2004.

[82] G. D'Agostini. *Role and Meaning of Subjective Probability: Some Comments on Common Misconceptions*. XX International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. Gif sur Yvette, Francia, 2000.

[83] F. J. Anscombe and J. Aumann. A definition of subjective probability. *The Annals of Mathematical Statistics*, 34(1):199–205, 1963.

[84] R. M. Cooke. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press, Oxford, 1991.

[85] R. Jeffrey. *Probability and the Art of Judgement*. Cambridge University Press, New York, 1992.

[86] J. B. Kadane and R. L. Winkler. Separating probability elicitation from utilities. *Journal of the American Statistical Association*, 83(402):357–363, 1988.

[87] A. P. Dawid. Probability, causality and the empirical world: A bayes-de finetti-popper-borel sythesis. *Statistical Science*, 19(1):44–57, 2004.

[88] J. O. Berger, B. Liseo, and R. L. Wolpert. *Integrated Likelihood Methods for Eliminating Nuisance Parameters*. Purdue Univ. Dept. of Statistics Technical Report No. 96-7C, 1998.

[89] A. W. F. Edwards. *Likelihood: Expanded Edition*. The Johns Hopkins University Press, Baltimore, 1992.

[90] B. de Finetti. Sulla proprietá conglomerativa delle probabilitá subordinate. *Atti R. Ist. Lomb. Scienze Lettere*, 63:418–418, 1930.

[91] D. Draper, J. S. Hodges, C. L. Mallows, and D. Pregibon. Exchangeability and data analysis. *Journal of the Royal Statistical Society. Series A*, 156(1):9–37, 1993.

[92] J. M. Dickey. Approximate posterior distributions. *Journal of the American Statistical Association*, 71(355):680–689, 1976.

[93] D. Fink. *A Compdium of Conjugate Priors*. Technical Report. Dept. of Biology. Montana State University, Bozeman, MT, 59717 edition, 1997.

[94] M. Ramoni and P. Sebastiani. *Bayesian Methods for Intelligent Data Analysis. KMi Technical Report*. KMi-TR-67, 1998.

[95] D. Draper. *Bayesian Hierarchical Modeling*. Tutorial 1: ISBA, Crete, 2000.

[96] G. C. G. Wei and M. A. Tanner. Posterior computations for censored regression data. *Journal of the American Statistical Association*, 85(411):829–839, 1990.

[97] A. Gelman. A bayesian formulation of exploratory data analysis and goodness-of-fit testing. *International Statistical Review*, 71(2):369–382, 2003.

[98] E. Damsleth. Conjugate classes for gamma distributions. *Scandinavian Journal of Statistics*, 2(2):80–84, 1975.

[99] G. Meeden. *An Elicitation Procedure Using Piecewise Conjugate Priors*. Bayesian Analysis in Statistics and Econometrics. Goel, P. K. y Iyengar, N. S., Eds. Springer-Verlag, New York, 1992.

[100] J. B. Kadane, M. J. Schervish, and T. Seidenfeld. *Rethinking the Foundations of Statistics*. Cambridge University Press, New York, 1999.

[101] R. L. Winkler. The assessment of prior distributions in bayesian analysis. *Journal of the American Statistical Association*, 62(319):776–800, 1967a.

[102] R. Yang and J. O. Berger. *A Catalog of Noninformative Priors. Technical Report*. Duke University, 1998.

[103] S. K. Bhattacharya. Bayesian approach to life testing and reliability estimation. *Journal of the American Statistical Association*, 62(317):48–62, 1967.

[104] J. Aldrich. The statistical education of harold jeffreys. *International Statistical Review*, 73(2):289–307, 2005.

[105] R. E. Kass and L. Wasserman. *Formal Rules for Selecting Prior Distributions: A Review and Annotated Bibliography. Techinical Report*. Carnegie Mellon University, 1994.

[106] J. G. Ibrahim and P. W. Laud. On bayesian analysis of generalized linear models using jeffreys' prior. *J. Amer. Statist. Assoc.*, 86:981–986, 199i.

[107] F. Tuyl, R. Gerlach, and K. Mengersen. A comparison of bayes-laplace, jeffreys, and other priors: The case of zero events. *The American Statistician*, 62(1):40–44, 2008.

[108] M. Papathomas and R. J. Hocking. Bayesian updating for binary variables: An application in the uk water industry. *The Statistician*, 52(4):483–499, 2003.

[109] V. E. McGee. *Principles of Statistics: Traditional and Bayesian*. Meredith Co., New York, 1971.

[110] M. Zhu and A. Y. Lu. The counter-intuitive non-informative prior for the bernoulli family. *Journal of Statistics Education*, 12(2), 2004.

[111] J. M. Bernardo. Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society, Series B*, 41:113–147, 1979.

[112] J. M. Bernardo. Reference analysis. *Handbook of Statistics. Dey, D. K. y Rao, C. R.*, 25:17–90, 2005.

[113] B. Clarke and D. Sun. Reference priors under the chi-squared distance. *Sankhya: The Indian Journal of Statistics, Series A*, 59(2):215–231, 1997.

[114] B. Liseo. The elimination of nuisance parameters. *Handbook of Statistics. Dey, D. K. y Rao, C. R.*, pages 193–219, 2005.

[115] J. Albert. Nuisance parameters and the use of exploratory graphical methods in a bayesian analysis. *The American Statistician*, 43(4):191–196, 1989.

[116] R. J. Hyndman. Computing and graphing highest density regions. *The American Statistician*, 50(2):120–126, 1996.

[117] A. Elfessi and D. M. Reineke. A bayesian look at classical estimation: The exponential distribution. *Journal of Statistics Education*, 9(1), 2001.

[118] A. J. Rossman, T.H. Short, and M. T. Parks. Bayes estimators for continuous uniform distribution. *Journal of Statistics Education*, 6(3), 1998.

[119] R. J. Serfling. *Aproximation Theorems of Mathematical Statistics*. John Wiley Sons, New York, 1980.

[120] G. Canavos. *Probabilidad y Estadística: Aplicaciones y Métodos*. McGraw Hill, Madrid, 5 edition, 1998.

[121] T. H. Wonnacott and R. J. Wonnacott. *Fundamentos de Estadística para Administración y Economía*. Limusa, México, 1979.

[122] G. G. Roussas. *A First Course in Mathematical Statistics*. Addison-Wesley Publishing Company: Reading, Massachusetts, 1973.

[123] J. G. Kalbfleish. *Probability and Statistical Inference*. Springer-Verlag, New York, 2 edition, 1985.

[124] B. Efron. Computers and theory of statistics: Thinking the unthinkable. *SIAM Review*, 21:460–480, 1979.

[125] T. Wright. When zero defectives appear in a sample: Upper bounds on confidence coefficients of upper bounds. *The American Statitician*, 44(1):40–41, 1990.

[126] R. E. Kass. Bayes factors in practice. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 42(5):551–560, 1993.

[127] S. Sinharay and H.S. Stern. On the sensitivity of bayes factors to the prior distributions. *The American Statistician*, 56(3):196–201, 2002.

[128] F. De Santis and F. Spezzaferri. Methods for default and robust bayesian model comparison: the fractional bayes factor approach. *International Statistical Review*, 67(4):267–286, 1999.

[129] S. K. Sahu. *Bayesian Statistics. Lecture Notes, Faculty of Mathematical Studies*. University of Southhampton, 2000.

Juan Carlos Correa Morales
Docente Asociado, Escuela de Estadística
Universidad Nacional de Colombia Sede Medellín
Medellín-Colombia
Email: `jccorrea@unal.edu.co`

Carlos Javier Barrera Causil
Docente Titular, Facultad de Ciencias Exactas y Aplicadas
Instituto Tecnológico Metropolitano -ITM-
Medellín-Colombia
Email: `carlosbarrera@itm.edu.co`

TEXTOS ACADÉMICOS
Introducción a la Estadística Bayesiana
Fuentes tipográficas: Computer Modern para texto corrido, en 12 puntos, para títulos en Computer Modern, en 24 puntos y subtítulos

La relevancia que ha tomado la estadística bayesiana en distintas áreas lleva a escribir este texto, cuyo objetivo es contribuir en el crecimiento de los métodos bayesianos en América Latina e incentivar a los estudiantes a aplicar dichas herramientas en sus investigaciones. Aquí, se presentan los elementos básicos de la estadística bayesiana, estadística bayesiana computacional y aplicaciones. Esta estructura contiene en total 14 capítulos que ilustran al lector en un gran número de procedimientos. El lector puede solicitar al correo electrónico de los autores la información correspondiente de las bases de datos necesarias para implementar paso a paso los códigos de R y OpenBUGS presentados en esta obra.

The new relevance of Bayesian statistics in different fields led to the creation of this text. Its two goals are contributing to the growth of Bayesian methods in Latin America and encouraging students to use such tools in their research projects. It presents basic elements of Bayesian statistics, computational Bayesian statistics and their applications. It is divided into 14 chapters that instruct in a great deal of procedures. The reader may email the authors requesting the corresponding information of the necessary databases to implement the R and OpenBUGS codes herein step by step.

ITM
Institución Universitaria
Acreditada en Alta Calidad

ISBN 978-958-5414-24-2