

**UNIVERSIDAD TECNOLÓGICA DE LA MIXTECA**

**“ANÁLISIS BAYESIANO DEL MODELO DE REGRESIÓN  
LINEAL CON UNA APLICACIÓN A DATOS  
ASTRONÓMICOS”**

**T E S I S**

**PARA OBTENER EL TÍTULO DE:  
LICENCIADO EN MATEMÁTICAS APLICADAS**

**PRESENTA:  
ERANDY DONAJÍ GONZÁLEZ LÓPEZ**

**DIRECTOR  
M.C. NORMA EDITH ALAMILLA LÓPEZ**

**CO-DIRECTOR  
DR. LAURENT R. LOINARD CORVAISIER**

**HUAJUAPAN DE LEÓN, OAXACA., OCTUBRE DE 2009**



# Dedicatoria

*Dedico este trabajo a mi madre **Reina María López Gómez**,  
a mi hermano **Natán Eduardo González López**,  
a mi tía **Epifania López Gómez** por ser una segunda madre para mí,  
a mi prima **Cinthya Itzel Cruz López** por ser como una hermana  
y finalmente a mi padre **Salvador González Barradas** (†) que desde el  
cielo cuida mis pasos.*



# Agradecimientos

Agradezco a mi madre **Reina María López Gómez**, a mi hermano **Natán Eduardo González López** y a mi tía **Epifania López Gómez** gracias a su apoyo moral y económico para ayudarme a terminar mi carrera, y no abandonarme en los momentos difíciles.

Le doy gracias a la M.C. Norma Edith Alamilla López mi directora de tesis, por su apoyo y guía en este trabajo, también a mi co-director de tesis Dr. Laurent R. Loinard Corvaisier por su apoyo y tiempo, así como a cada uno de los sinodales: M.C. Vulfrano Tochiuitl Bueno, M.C. Juan Ramón Tijerina González y M.C. José del Carmen Jiménez Hernández.

Gracias a mis amigos Alan, René y Aristeo por su amistad y apoyo. En especial a mi amiga Isabel Franco Jiménez por estar conmigo en los altibajos, a mi amigo Gabriel Rojas Villanueva muchas gracias por tenerme paciencia, escucharme y los buenos consejos.

Finalmente gracias a todos mis maestros durante la carrera por la enseñanza que me brindaron, y a la Universidad Tecnológica de la Mixteca por darme la oportunidad de terminar una carrera.



# Índice general

Dedicatoria	III
Agradecimientos	v
Introducción	1
<b>1. Conceptos bayesianos</b>	<b>5</b>
1.1. Probabilidad subjetiva . . . . .	5
1.2. Análisis bayesiano . . . . .	6
1.3. Función de verosimilitud . . . . .	6
1.3.1. Principio de verosimilitud . . . . .	7
1.4. Distribución a priori . . . . .	8
1.4.1. Determinación subjetiva de la distribución a priori . . . . .	8
1.4.2. Distribución a priori conjugada . . . . .	10
1.4.3. Distribución a priori no informativa . . . . .	12
1.5. Distribución a posteriori . . . . .	15
1.6. Inferencia bayesiana . . . . .	18
1.6.1. Estimación . . . . .	18
1.6.2. Error de estimación. . . . .	19
1.6.3. Estimación multivariada . . . . .	20
1.6.4. Conjuntos creíbles . . . . .	21
1.6.5. Contraste de hipótesis . . . . .	23

1.6.6. Inferencia predictiva . . . . .	27
<b>2. Conceptos astronómicos.</b>	<b>31</b>
2.1. Sistemas de coordenadas. . . . .	31
2.2. Determinación de las ecuaciones. . . . .	35
2.3. Paralaje estelar . . . . .	38
2.4. T Tauri . . . . .	41
<b>3. Análisis de regresión lineal</b>	<b>43</b>
3.1. Regresión lineal simple. . . . .	44
3.1.1. Propiedades de los E.M.V. . . . .	47
3.1.2. Contraste de hipótesis. . . . .	50
3.2. Regresión lineal múltiple. . . . .	56
3.2.1. Estimación de los coeficientes de regresión . . . . .	56
3.2.2. Propiedades de los estimadores. . . . .	60
3.2.3. Contraste de hipótesis . . . . .	65
3.3. Análisis bayesiano del modelo de regresión lineal. . . . .	65
3.3.1. Distribución a posteriori con información a priori no informativa . . . . .	66
3.3.2. Distribución a posteriori con información a priori conjugada . . . . .	70
3.3.3. Inferencia bayesiana para $\beta$ y $\sigma^2$ . . . . .	74
<b>4. Aplicación.</b>	<b>77</b>
4.1. Tratamiento de los datos. . . . .	77
4.1.1. Tratamiento de los datos para la ascensión recta. . . . .	77
4.1.2. Tratamiento de los datos para la declinación. . . . .	81
4.1.3. Comparación de las estimaciones puntuales clásicas y bayesianas . . . . .	86
4.1.4. Intervalos HPD para los parámetros. . . . .	90
4.2. Conclusiones . . . . .	92



---

<b>A. Apéndice</b>	<b>95</b>
A.1. Distribuciones especiales. . . . .	95
A.2. Algunos teoremas relacionados con matrices. . . . .	96
A.3. Integrales . . . . .	96
A.4. Valores esperados y teoría de la distribución. . . . .	97
<b>Bibliografía</b>	<b>97</b>



# Introducción

En la actualidad se reconoce la importancia que tiene la estadística en las diversas áreas del conocimiento. Cada vez son más las distintas áreas de la investigación científica que utilizan la estadística, para la recopilación, presentación, análisis e interpretación de los datos, con el fin de ayudar en una mejor toma de decisiones.

La estadística se ocupa fundamentalmente del análisis de datos que presenten variabilidad con el fin de comprender el mecanismo que los genera, o para ayudar en un proceso de toma de decisiones. En cualquiera de los casos existe una componente de incertidumbre involucrada, por lo que el estadístico se ocupa en reducir lo más posible esa componente, así como también en describirla en forma apropiada. Además, la probabilidad es el único lenguaje posible para describir una lógica que trata con todos los niveles de incertidumbre y no sólo con los extremos de verdad o falsedad.

Se supone que toda forma de incertidumbre debe decidirse por medio de modelos de probabilidad. En el caso de la estadística bayesiana se considera al parámetro, sobre el cual se desea inferir, como un evento incierto. Entonces, como nuestro conocimiento no es preciso y está sujeta a incertidumbre, se puede describir mediante una distribución de probabilidad  $\pi(\theta)$ , llamada distribución a priori del parámetro  $\theta$ , lo que hace que el parámetro tenga el carácter de aleatorio.

La estadística bayesiana es una alternativa a la estadística clásica en los problemas típicos estadísticos como son estimación, prueba de hipótesis y predicción. La estadística bayesiana proporciona un completo paradigma de inferencia estadística y teoría de la decisión con incertidumbre y está basada en el teorema de Bayes que es el que nos proporciona una forma adecuada para incluir información inicial del parámetros mediante la distribución  $\pi(\theta)$  además de la información muestral  $f(x|\theta)$  para así describir toda la informa-

ción disponible del parámetro mediante la distribución a posteriori  $\pi(\theta|x)$ .

La teoría bayesiana plantea la solución a un problema estadístico desde el punto de vista subjetivo de la probabilidad, según el cual, la probabilidad que un estadístico asigna a uno de los posibles resultados de un proceso, representa su propio juicio sobre la verosimilitud de que se tenga el resultado y dicha información esta representada en  $\pi(\theta)$ . Algunos investigadores rechazan que la información inicial se incluya en un proceso de inferencia científica, sin embargo los métodos bayesianos pueden ser aplicados a problemas que han sido inaccesibles a la teoría frecuentista tradicional.

A un problema específico se le puede asignar cualquier tipo de distribución a priori, ya que finalmente al actualizar la información a priori que se tenga acerca del parámetro, mediante el teorema de Bayes se estará actualizando la distribución a posteriori del parámetro y es con esta con las que se hacen las inferencias del mismo.

Cuando un investigador tiene conocimiento previo a un problema, este conocimiento previo puede cuantificarse en un modelo de probabilidad. Si los juicios de una persona sobre la verosimilitud relativa a ciertas combinaciones de resultados satisfacen ciertas condiciones de consistencia, se puede decir que sus probabilidades subjetivas se determinan de manera única.

Dentro de los campos de la investigación en los cuales se utiliza la Estadística se encuentra la Astronomía, en el caso particular de este trabajo de tesis utilizando los modelos de regresión lineal para estimar el valor de parámetros como la paralaje que son importantes al calcular distancias a las estrellas. Para calcular las distancias algunos tipos de observación que se hacen son las de infrarrojo u ondas de radio que nos son visibles a simple vista, utilizando dichas observaciones en modelos que describen la posición de las estrellas en la esfera celeste y resolviendolos con regresión lineal.

El objetivo de esta tesis, es revisar el modelo de regresión lineal con la metodología clásica y también con el enfoque bayesiano y hacer una comparación entre ambos resultados, otro objetivo es resolver las ecuaciones de posición para la estrella T Tau Sb, a través de un modelo de regresión lineal y hacer inferencia para los parámetros del modelo y así determinar el parámetro  $p$  de paralaje y la distancia a que se encuentra la estrella mencionada y compararla con el resultado obtenido en el artículo (veáse [12]).

Para lograr los objetivos se desarrolla la teoría necesaria en los capítulos de la siguiente manera:

En el Capítulo uno, se da un panorama general de lo que es el análisis bayesiano, el teorema de Bayes, función de verosimilitud, las distintas formas de obtener la distribución a priori y la distribución a posteriori, así como la forma de hacer inferencia bayesiana.

En el segundo Capítulo se presenta la información para conocer mas acerca de la estrella T Tau Sb, las ecuaciones de posición y los sistemas coordenados para la ubicación de estrellas en la esfera celeste, se dan las ecuaciones necesarias para determinar la distancia a una estrella, se hace mención de la paralaje trigonométrica y su relación con la distancia, así como una breve descripción del sistema de estrellas doble T Tau y la correspondiente estrella T Tau Sb y se presentan las observaciones de posición de dicha estrella, tomadas de (véase [12]) y una mas proporcionada por Laurent R.Loinard Corvaisier del Centro de Radioastronomía y Astrofísica de la Universidad Nacional Autónoma de México (CRyA-UNAM).

El Capítulo 3 contiene el desarrollo del análisis de regresión, iniciando con los modelos lineal simple y múltiple, para dar paso al análisis bayesiano del modelo de regresión con una distribución a priori no informativa y enseguida el modelo de regresión con distribución a priori conjugada, se determinan las distribuciones a posteriori de los parámetros de regresión, sus estimadores e intervalos HPD.

Por último, el capítulo 4 aborda todo lo concerniente a la aplicación de los modelos, se dan los resultados obtenidos haciendo una comparación entre los resultados clásicos y bayesianos, las estimaciones de los parámetros, y se presentan graficas de las distribuciones a posteriori .

Se obtiene la distancia a la estrella T Tau Sb, utilizando la paralaje trigonométrica obtenida con el análisis de regresión lineal bayesiano y se incluyen las conclusiones obtenidas en esta tesis.



# Capítulo 1

## Conceptos bayesianos

El concepto de probabilidad es utilizado en la vida diaria y a pesar de que es una parte tan común y natural de nuestra experiencia, no existe una unificación de la interpretación de tal concepto, las interpretaciones más utilizadas son, la interpretación clásica, la frecuentista y la subjetiva. La interpretación clásica se basa en decir que los resultados son igualmente verosímiles, la frecuentista en la frecuencia relativa es decir que un proceso se repita un gran número de veces.

A continuación se abordará el enfoque subjetivo de la probabilidad y algunos de los aspectos de inferencia estadística más utilizados, ahora estudiados desde la interpretación subjetiva de la probabilidad.

### 1.1. Probabilidad subjetiva

La idea principal de la probabilidad subjetiva es dejar que la probabilidad de un evento refleje la creencia personal en la ocurrencia de ese evento.

De acuerdo con la interpretación subjetiva de la probabilidad, ésta establece que la probabilidad que un estadístico asigna a cada uno de los posibles resultados de un proceso, representa su propio juicio sobre la verosimilitud de que se obtenga ese resultado.

Este juicio estará basado en opiniones e información acerca del proceso. Además, esta interpretación subjetiva de la probabilidad puede ser formalizada.

Si los juicios de un estadístico acerca de las verosimilitudes relativas a

diversas combinaciones de resultados, satisfacen ciertas condiciones de consistencia, entonces puede demostrarse que sus probabilidades subjetivas para los diferentes sucesos posibles, pueden ser determinadas en forma única.

## 1.2. Análisis bayesiano

En el análisis bayesiano, además de especificar el modelo de los datos observados  $x = (x_1, \dots, x_n)$ , dado un vector de parámetros desconocidos  $\theta$ , usualmente en la forma de la función de probabilidad o función de densidad de probabilidad  $f(x|\theta)$  se supone que  $\theta$  es una cantidad aleatoria y que tiene una distribución a priori  $\pi(\theta)$ . La inferencia concerniente a  $\theta$  está basada en su distribución a posteriori, dada por, el cociente de la distribución conjunta  $h(x, \theta)$  entre la distribución marginal de  $X$ .

$$\pi(\theta|x) = \frac{h(x, \theta)}{m(x)} = \frac{h(x, \theta)}{\int h(x, \theta)d\theta} = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta} \quad (1.1)$$

A la expresión anterior se le llama teorema de Bayes. Nótese la contribución, tanto los datos experimentales (en la forma de la probabilidad  $f$ ) y la opinión a priori (en la forma de la distribución a priori  $\pi$ ) en la ecuación (1.1).

Las inferencias se basan en  $\pi(\theta|x)$  en lugar de  $f(x|\theta)$ ; es decir, en la distribución de probabilidad (del valor desconocido) del parámetro dados los datos observados, en vez de la distribución de los datos dado el valor del parámetro.

## 1.3. Función de verosimilitud

La función de verosimilitud  $l(\theta|x)$  expresa el proceso por el cual pueden aparecer los datos  $x$  en términos del parámetro desconocido  $\theta$ .

**Definición 1.3.1** *Sea una muestra aleatoria  $X_1, \dots, X_n$ , de tamaño  $n$ , se define y se obtiene la función de verosimilitud (o conjunta)  $l(\theta|x)$  como  $\prod_{i=1}^n f(x_i|\theta)$ .*



En la inferencia clásica  $l(\theta|x)$  representa el modelo estadístico del proceso de generación de datos. La estadística clásica suele considerar ese proceso como un muestreo aleatorio sobre la población de datos, y así  $l(\theta|x)$  se interpreta como una probabilidad frecuentista por tanto  $x$  es una variable aleatoria que se utiliza para realizar inferencia.

Desde el punto de vista bayesiano  $l(\theta|x)$  mide los grados de creencia del investigador de que los datos tomen ciertos valores dada la información hipotética de que los parámetros tomen ciertos valores, además de toda la información a priori. El modelo estadístico no es entonces tanto una hipótesis sobre los procesos de generación de datos, sino una afirmación desde el punto de vista subjetivo del investigador acerca del proceso; afirmación que se encuentra condicionada a los valores de los parámetros desconocidos. Es más la función de verosimilitud  $l(\theta|x)$  es una función de  $\theta$  y tan sólo es un instrumento para pasar de la distribución a priori a la distribución a posteriori.

### 1.3.1. Principio de verosimilitud

Un principio muy importante para el paradigma bayesiano es el de verosimilitud que establece en forma explícita la idea natural de que sólo las observaciones actuales  $x$ , deberían ser relevantes para poder establecer conclusiones o tener alguna evidencia sobre  $\theta$ . Este principio se basa en la función de verosimilitud  $l(\theta|x)$  y dice lo siguiente:

*Al hacer inferencia o tomar decisiones sobre  $\theta$  después de que  $x$  se ha observado, toda la información experimental relevante se encuentra en la función de verosimilitud para el  $x$  observado. Es más, dos funciones de verosimilitud contienen la misma información sobre  $\theta$  si una es directamente proporcional a la otra como función de  $\theta$ .*

Algunos aspectos importantes del principio de verosimilitud son los siguientes:

- ◇ La correspondencia de información a partir de funciones de verosimilitud proporcionales se aplica sólo cuando las dos funciones de verosimilitud son para el mismo parámetro.
- ◇ El principio de verosimilitud se cumple sí:

$$\forall x \ l_1(\theta|x) = \alpha l_2(\theta|x) \ \forall \theta.$$

Es decir, la razón de las verosimilitudes para los dos experimentos y cada observación de los datos es constante.

- ◇ El principio de verosimilitud no dice que toda la información sobre  $\theta$  se encuentra en  $l(\theta|x)$ , sino que toda la información experimental es la que se encuentra en  $l(\theta|x)$ .
- ◇ Para los métodos bayesianos, la verosimilitud no es lo más importante, ya que en comparación con los métodos clásicos, en estos la función de verosimilitud es el centro alrededor del cual gira todo tipo de inferencia. Para los métodos bayesianos la función de verosimilitud es tan solo un instrumento para pasar de la distribución a priori  $\pi(\theta)$  a la distribución a posteriori  $\pi(\theta|x)$ .

## 1.4. Distribución a priori

Al hacer inferencias sobre un parámetro, generalmente se cuenta con algún tipo de información (juicios, creencias) acerca de su valor, incluso antes de observar los datos.

La distribución a priori  $\pi(\theta)$  expresa lo que es conocido del parámetro desconocido  $\theta$  antes de observar algún dato.

### 1.4.1. Determinación subjetiva de la distribución a priori

Un elemento importante en las decisiones de muchos problemas es la información a priori que se tenga acerca del parámetro de interés  $\theta$ . La información que el estadístico tiene sobre la verosimilitud de dichos sucesos debe ser cuantificada a través de una medida de probabilidad sobre el espacio paramétrico  $\Theta$ .

Si  $\Theta$  es discreto, el problema se reduce a la determinación de la probabilidad subjetiva para cada elemento de  $\Theta$ . Cuando  $\Theta$  es continuo, el problema de construir  $\pi(\theta)$  es más difícil.

### Histograma de aproximación

Cuando  $\Theta$  es un intervalo de la línea real, la aproximación por histogramas de  $\pi(\theta)$  consiste en

1. Dividir  $\Theta$  en subintervalos.
2. Determinar la probabilidad subjetiva para cada subintervalo.
3. Dibujar el histograma de probabilidad.
4. A partir de este histograma, se esboza una densidad  $\pi(\theta)$ .

Atendiendo la cantidad de subintervalos, se obtendrán desde histogramas burdos (con pocos subintervalos) hasta histogramas muy detallados (gran cantidad de subintervalos). El tipo de histograma dependerá del tipo de problema, pero en cualquier caso se tienen los siguientes inconvenientes:

- No existe ninguna regla que determine la cantidad de subintervalos que deben determinarse.
- Ni tampoco alguna que diga que tamaño deben de tener dichos subintervalos.

Existen dos dificultades para esta aproximación, las funciones así obtenidas son difíciles de trabajar y las funciones de densidad no tienen colas.

### Aproximación por creencias relativas

La aproximación por creencias relativas es muy usada cuando  $\Theta$  es un intervalo de la línea real, esto consiste simplemente en comparar las probabilidades intuitivas para varios puntos en  $\Theta$ , y directamente esbozar la distribución a priori para esta determinación.

La implementación desde el enfoque bayesiano depende del conocimiento previo o subjetivo que se le asigne a la distribución de probabilidad, no solo a las variables como  $x$ , sino también a los parámetros como  $\theta$ .

Históricamente un obstáculo importante para el uso generalizado del paradigma bayesiano ha sido la determinación de la forma apropiada de la distribución a priori  $\pi$  que es usualmente una ardua tarea.

Típicamente, ésta distribución se especifica basándose en la información acumulada de estudios pasados o de la opinión subjetiva de los expertos en el área. Con el fin de racionalizar el proceso de elección de la distribución a

priori, así como simplificar la carga computacional subsecuente, los experimentos solo limitan a menudo esta elección a restringir  $\pi$  a algunas familias de distribuciones familiares.

Una alternativa simple, disponible en algunos casos, es dotar a la distribución a priori con poco contenido informativo, de modo que los datos del presente estudio será la fuerza dominante en la determinación de la distribución a posteriori.

### 1.4.2. Distribución a priori conjugada

Supongamos que  $\theta$  es univariado. Quizás la más simple aproximación para especificar  $\pi(\theta)$  es primero limitar las consideraciones a una colección manejable (a lo mas numerable) de posibles valores considerados de  $\theta$  y subsecuentemente asignar probabilidades de masa a estos valores de tal forma que su suma sea 1, su relativa contribución reflejan al experimentador sus creencias a priori tan cercanas como sea posible.

Si  $\theta$  es un valor discreto, tal aproximación puede ser muy natural. Si  $\theta$  es continuo, debemos asignar a las masas un intervalo en la línea real, en lugar de un solo punto, resultando un histograma a priori para  $\theta$ . Tal histograma (necesariamente sobre una región acotada) parece ser inapropiado, especialmente en lo concerniente a una probabilidad continua  $f(x|\theta)$ , pero puede de hecho ser más apropiada si la integral requiere el cálculo numérico de la distribución a posteriori.

Además, el histograma a priori puede tener tantos subintervalos como sea posible así como también la precisión de la opinión a priori lo permita. Es de vital importancia, sin embargo, que el rango del histograma sea suficientemente grande, ya que como puede verse en (1.1), el soporte de la distribución a posteriori será necesariamente un subconjunto del soporte de la distribución a priori.

Alternativamente, podemos simplemente asumir que la distribución a priori para  $\theta$  pertenece a una familia paramétrica de distribuciones  $\pi(\theta)$ .

Al elegir a la distribución a priori en una familia de distribuciones específicas  $\pi(\theta)$ , algunas elecciones pueden ser convenientemente más calculables unas que otras. En particular puede ser posible seleccionar un miembro de una familia la cual es conjugada para la función de verosimilitud  $l(\theta|x)$ ,

que es, una que conduce a una distribución a posteriori  $\pi(\theta|x)$  que pertenece a la misma familia de distribuciones de la distribución a priori.

**Definición 1.4.1** *Sea la clase  $F$  de funciones de densidad  $f(x|\theta)$ , una clase  $\rho$  de distribuciones a priori se dice que es una familia conjugada para  $F$  si  $\pi(\theta|x)$  está en la clase de  $\rho$ , para toda  $f \in F$  y  $\pi \in \rho$ .*

Usualmente para una clase de densidades  $F$ , una familia conjugada puede ser determinada examinando la función de verosimilitud  $l(\theta|x)$ . Entonces la familia conjugada puede ser escogida como la clase de distribuciones con la misma estructura funcional. A esta clase de familia conjugada se le conoce como distribución a priori conjugada natural.

**Ejemplo 1.4.1** *Supongamos que  $X$  es el número de mujeres embarazadas que llegan a un hospital para tener a sus bebés durante un mes dado. Supongamos además que la tasa de llegada de las mujeres embarazadas tiene la forma de una distribución de probabilidad de Poisson.*

Se tiene  $X_i \sim Poi(\theta)$ .

$$f(x_i|\theta) = \frac{\exp\{-\theta\}\theta^{x_i}}{x_i!},$$

con función de verosimilitud

$$l(\theta|x) = \frac{\exp\{-n\theta\}\theta^{n\bar{x}}}{\prod x_i!}.$$

Para realizar el análisis bayesiano requerimos de una distribución a priori para  $\theta$ , teniendo como soporte la línea real, una elección natural para la distribución a priori la da la distribución gamma,  $l(x|\theta) \propto \theta^{(n\bar{x}-1)+1} \exp\{-n\theta\}$ , es decir una posible familia de distribuciones para  $\pi(\theta)$  es la de la distribución gamma.

$$\pi(\theta) = \frac{\beta^\alpha \theta^{\alpha-1} \exp\{-\beta\theta\}}{\Gamma(\alpha)}, \quad \theta, \alpha, \beta > 0,$$

o  $\theta \sim G(\alpha, \beta)$ , la distribución gamma tiene media  $\alpha\beta$  y varianza  $\alpha\beta^2$ .

### 1.4.3. Distribución a priori no informativa

En ciertos problemas, el conocimiento inicial sobre el verdadero valor del parámetro  $\theta$  puede ser muy débil, o vago, ésto ha llevado a generar un tipo de distribuciones a priori llamada distribuciones a priori no informativas, las cuales reflejan un estado de ignorancia inicial.

Una distribución sobre  $\theta$  se dice que es no informativa si no contiene información sobre  $\theta$ , es decir, no establece si unos valores de  $\theta$  son más favorables que otros.

Por ejemplo, si se establecen dos hipótesis simples sobre el valor de  $\theta$  y asignamos una probabilidad  $1/2$  a cada una de ellas, se tiene una situación no informativa.

#### Método de Jeffreys.

Esta técnica consiste en buscar funciones a priori no informativas invariantes, es decir, si existe ignorancia sobre  $\theta$ , esto implica cierta ignorancia acerca de  $\phi = h(\theta)$ , con la que debería verificarse la siguiente condición de invarianza.

*Si la a priori no informativa sobre  $\theta$  es  $\pi(\theta)$ , la a priori no informativa sobre  $\phi$  debe ser  $\pi(h^{-1}(\phi)) \left| \frac{dh^{-1}(\phi)}{d\phi} \right|$ .*

Jeffreys propuso solucionar esto, definiendo la a priori como la raíz cuadrada de la información esperada de Fisher.

**Definición 1.4.2** *Si  $\theta \in \mathbb{R}$ , se define la información esperada de Fisher como*

$$I(\theta) = -E_{\theta} \left[ \frac{\partial^2}{\partial \theta^2} \log f(x|\theta) \right],$$

en donde

$E_{\theta}$ : el valor esperado de la variable aleatoria que es función de  $x$ .

$f(x|\theta)$ : función de densidad de  $x$  (que depende de  $\theta$ ).

Jeffrey propone que se elija como función a priori no informativa la determinada por la siguiente ecuación,

$$\pi(\theta) \propto |I(\theta)|^{1/2}.$$

**Definición 1.4.3** Si  $\theta \in \mathbb{R}^n$ , se define la matriz de información de Fisher como aquella matriz  $n \times n$  cuyas componentes son:

$$I_{ij}(\theta) = -E_{\theta} \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x|\theta) \right].$$

Jeffrey propone que se elija como función a priori no informativa la determinada por la siguiente expresión:

$$\pi(\theta) \propto [\det I(\theta)]^{1/2}.$$

**Ejemplo 1.4.2** Considere un experimento el cual consiste de la observación de un ensayo Bernoulli.

$$f(x|\theta) = \theta^x (1 - \theta)^{1-x}, \quad x = 0, 1, 0 \leq \theta \leq 1,$$

$$I(\theta) = - \sum_{x=0}^1 f(x|\theta) \frac{\partial^2}{\partial \theta^2} \log f(x|\theta),$$

se tiene que

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \log f(x|\theta) &= \frac{\partial^2}{\partial \theta^2} \log \theta^x (1 - \theta)^{1-x} \\ &= \frac{\partial}{\partial \theta} [x\theta^{-1} - (1-x)(1-\theta)^{-1}] \\ &= -\frac{x}{\theta^2} - \frac{(1-x)}{(1-\theta)^2}, \end{aligned}$$

luego entonces,

$$\begin{aligned} E \left[ \frac{\partial^2}{\partial \theta^2} \log \theta^x (1 - \theta)^{1-x} \right] &= E \left[ -\frac{x}{\theta^2} - \frac{(1-x)}{(1-\theta)^2} \right] \\ &= \sum_{x=0}^1 \theta^x (1 - \theta)^{1-x} \left[ -\frac{x}{\theta^2} - \frac{(1-x)}{(1-\theta)^2} \right] \\ &= -\theta^{-1} (1 - \theta)^{-1}, \\ I(\theta) &= -(-\theta^{-1} (1 - \theta)^{-1}) \\ &= \theta^{-1} (1 - \theta)^{-1}. \end{aligned}$$

Así, la distribución a priori obtenida mediante el método de Jeffrey es:

$$\pi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2},$$

que es el kernel de una densidad beta, por tanto

$$\pi(\theta) = \frac{\Gamma(1)}{\Gamma(1/2)\Gamma(1/2)}\theta^{\frac{1}{2}-1}(1-\theta)^{\frac{1}{2}-1} = \text{Beta}(1/2, 1/2).$$

### Ejemplo 1.4.3

Una densidad de *localización-escala* es una densidad de la forma

$$\sigma^{-1}f((x-\theta)/\sigma) \quad \text{con } \theta \in \mathbb{R}, \sigma > 0,$$

los parámetros  $\theta$  y  $\sigma$  son conocidos, llamados *parámetros de localización y escala* respectivamente.

El parámetro de localización se usa para desplazar una distribución hacia un lado u otro.

El parámetro de escala define cuán dispersa se encuentra la distribución (en el caso de la distribución normal, el parámetro de escala es la desviación típica). La distribución normal es una densidad de *localización-escala*, con los parámetros  $\sigma$  y  $\theta$ .

Una muestra aleatoria se dice que tiene una densidad de *localización-escala* si las densidades de cada una de las variables aleatorias es de *localización-escala*.

Ya que la distribución normal pertenece a este tipo de densidades, se desea encontrar la distribución a priori de  $\eta = (\theta, \sigma)$ .

Se sabe que una distribución normal tiene la forma siguiente,

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



el parámetro  $\eta = (\theta, \sigma)$ , tiene la matriz de información de *Fisher*,

$$\begin{aligned}
 I(\eta) &= -E_{\eta} \begin{bmatrix} \frac{\partial^2}{\partial \theta^2} \ln f(x|\theta) & \frac{\partial^2}{\partial \theta \partial \sigma} \ln f(x|\theta) \\ \frac{\partial^2}{\partial \theta \partial \sigma} \ln f(x|\theta) & \frac{\partial^2}{\partial \sigma^2} \ln f(x|\theta) \end{bmatrix} \\
 &= -E_{\eta} \begin{bmatrix} \frac{\partial^2}{\partial \theta^2} \left( -\frac{(X-\theta)^2}{2\sigma^2} \right) & \frac{\partial^2}{\partial \theta \partial \sigma} \left( -\frac{(X-\theta)^2}{2\sigma^2} \right) \\ \frac{\partial^2}{\partial \theta \partial \sigma} \left( -\frac{(X-\theta)^2}{2\sigma^2} \right) & \frac{\partial^2}{\partial \sigma^2} \left( -\frac{(X-\theta)^2}{2\sigma^2} \right) \end{bmatrix} \\
 &= -E_{\eta} \begin{bmatrix} -\frac{1}{\sigma^2} & \frac{2(\theta-X)}{\sigma^3} \\ \frac{2(\theta-X)}{\sigma^3} & -\frac{3(X-\theta)^2}{\sigma^4} \end{bmatrix} \\
 &= \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{3}{\sigma^2} \end{pmatrix}.
 \end{aligned}$$

De (1.4) se tiene que la distribución a priori para  $\eta$  es,

$$\pi(\eta) = \pi(\theta, \sigma) = \left( \frac{1}{\sigma^2} \cdot \frac{3}{\sigma^2} \right)^{1/2} \propto \frac{1}{\sigma^2}. \quad (1.2)$$

## 1.5. Distribución a posteriori

La información a posteriori de  $\theta$  dado  $x$ , con  $\theta \in \Theta$ , está dada por la expresión  $\pi(x|\theta)$  y expresa lo que es conocido de  $\theta$  después de observar los datos de  $x$ .

Nótese que por la ley multiplicativa de la probabilidad,  $\theta$  y  $x$  tienen la siguiente función de densidad(subjetiva) conjunta,

$$h(x, \theta) = \pi(\theta)l(\theta|x),$$

en donde:

$\pi(\theta)$ : densidad a priori de  $\theta$ .

$l(\theta|x)$ : función de verosimilitud,

y  $x$  tiene función de densidad marginal dada por:

$$m(x) = \int_{\Theta} l(\theta|x)\pi(\theta)d\theta.$$

Si la función marginal  $m(x) \neq 0$ ,

$$\pi(\theta|x) = \frac{h(x, \theta)}{m(x)},$$

o sustituyendo

$$\pi(\theta|x) = \frac{\pi(\theta)l(\theta|x)}{\int l(\theta|x)\pi(\theta)d\theta}.$$

Que es la misma que la ecuación (1.1), es decir, el teorema de Bayes.

Ahora, dada una muestra de  $n$  observaciones independientes, se puede obtener  $l(\theta|x)$  y proceder evaluando en la ecuación (1.1). Pero evaluar esta expresión puede ser simplificada, utilizando un estadístico suficiente para  $\theta$  con función de densidad  $g(S(x)|\theta)$ , y esto se enuncia en el lema siguiente.

**Lema 1.5.1** *Sea  $S(x)$  un estadístico suficiente para  $\theta$  (es decir,  $l(\theta|x) = h(x)g(S(x)|\theta)$ ),  $m(s) \neq 0$  la densidad marginal para  $S(x) = s$ , entonces se cumple que:*

$$\begin{aligned} \pi(\theta|x) &= \pi(\theta|s) \\ &= \frac{g(s|\theta)\pi(\theta)}{m(s)}. \end{aligned}$$

**Demostración.**

$$\begin{aligned} \pi(\theta|x) &= \frac{l(\theta|x)\pi(\theta)}{\int l(x|\theta)\pi(\theta)d\theta} \\ &= \frac{h(x)g(S(x)|\theta)\pi(\theta)}{\int h(x)g(S(x)|\theta)\pi(\theta)d\theta} \\ &= \frac{g(s|\theta)\pi(\theta)}{m(s)} = \pi(\theta|s). \end{aligned}$$

□

Se observa que la ecuación (1.1) se puede expresar de la manera más corta conveniente.

$$\pi(\theta|x) \propto l(x|\theta)\pi(\theta). \quad (1.3)$$

En otras palabras *la distribución a posteriori es proporcional a la verosimilitud por la a priori*, es decir, que la probabilidad es multiplicada por una constante (o una función de  $x$ ), sin alterar a la a posteriori, pues en la ecuación (1.1) el denominador no depende de  $\theta$ .

### Ejemplo 1.5.1

Obtenga la a posteriori para el ejemplo (1.4.2).

Para el ejemplo (1.3.2) se conoce  $\pi(\theta)$ .

Si  $r = \sum_{i=1}^n x_i$ , entonces la a posteriori es

$$\begin{aligned} \pi(\theta|x) &\propto l(\theta|x)\pi(\theta) \\ &\propto \theta^{r-1/2}(1-\theta)^{n-r-1/2}. \end{aligned}$$

De lo anterior se nota, que  $\pi(\theta|x)$  tiene una distribución beta, pues haciendo

$$\alpha = r + \frac{1}{2},$$

$$\beta = n - r + \frac{1}{2}.$$

Se tiene que:

$$\pi(\theta|x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}.$$

### Ejemplo 1.5.2

Obtenga la a posteriori para el ejemplo (1.4.1).

Se tiene que  $\theta \sim G(\alpha, \beta)$ , así la a posteriori utilizando el teorema de Bayes (1.1) es,

$$\begin{aligned} \pi(\theta|x) &\propto l(\theta|x)\pi(\theta) \\ &\propto (\exp\{-n\theta\}\theta^{n\bar{x}})(\theta^{\alpha-1}\exp\{-\beta\theta\}) \\ &\propto \theta^{n\bar{x}+\alpha-1}\exp\{-\theta(n+\beta)\}. \end{aligned}$$

Por lo tanto  $\pi(\theta|x) = G(n\bar{x} + \alpha - 1, n + \beta)$ .

## 1.6. Inferencia bayesiana

Los problemas concernientes a la inferencia de  $\theta$  pueden ser resueltos fácilmente utilizando análisis bayesiano. Dado que la distribución a posteriori contiene toda la información disponible acerca del parámetro, muchas inferencias concernientes a  $\theta$  pueden consistir únicamente de las características de esta distribución.

### 1.6.1. Estimación

#### Estimación puntual

Para estimar  $\theta$ , se pueden aplicar numerosas técnicas de la estadística clásica a la distribución a posteriori. La técnica más conocida es estimación por máxima verosimilitud, en la cual se elige como estimador de  $\theta$  a  $\hat{\theta}$  que es el valor que maximiza a la función de verosimilitud  $l(\theta|x)$ .

Análogamente la estimación bayesiana por máxima verosimilitud se define de la manera siguiente.

**Definición 1.6.1** *La estimación de máxima verosimilitud generalizada de  $\theta$  es la moda más grande  $\hat{\theta}$  de  $\pi(\theta|x)$ . En otras palabras el valor de  $\theta$ ,  $\hat{\theta}$  que maximiza a  $\pi(\theta|x)$ , considerada como función de  $\theta$ .*

#### Ejemplo 1.6.1

Para la siguiente función de verosimilitud, calcular el estimador máximo verosimil de  $\theta$ ,

$$l(\theta|x) = \exp\{-(x - \theta)\}I_{(\theta, \infty)}(x),$$

y la distribución a priori es una Cauchy

$$\pi(\theta) = \frac{1}{\pi}(1 + \theta^2)^{-1}.$$

Luego se tiene que la distribución a posteriori es

$$\pi(\theta|x) = \frac{\exp\{-(x - \theta)\}I_{(\theta, \infty)}(x)}{m(x)(1 + \theta^2)\pi}.$$

Para encontrar el  $\hat{\theta}$ , se tienen dos posibilidades, la primera es que, si  $\theta > x$ , entonces  $I_{(\theta, \infty)} = 0$ , luego  $\pi(\theta|x) = 0$ .

La siguiente es que  $\theta \leq x$ , entonces  $I_{(\theta, \infty)} = 1$  y para esto se tiene que:

$$\pi(\theta|x) = \frac{\exp\{-(x - \theta)\}}{m(x)(1 + \theta^2)\pi}.$$

Para este  $\theta$ , calculamos la derivada con respecto a él,

$$\begin{aligned} \frac{d}{d\theta}\pi(\theta|x) &= \frac{d}{d\theta} \frac{\exp\{-(x - \theta)\}}{m(x)(1 + \theta^2)\pi} \\ &= \frac{\exp\{-x\}}{m(x)\pi} \frac{d}{d\theta} \left[ \frac{\exp\{\theta\}}{(1 + \theta^2)} \right] \\ &= \frac{\exp\{-x\}}{m(x)\pi} [(1 + \theta^2)^{-1} \exp\{\theta\} - 2\theta \exp\{\theta\} (1 + \theta^2)^{-2}] \\ &= \frac{\exp\{-x\}}{m(x)\pi} \left[ \frac{\exp\{\theta\}(1 + \theta^2)}{(1 + \theta^2)^2} - \frac{2\theta \exp\{\theta\}}{(1 + \theta^2)^2} \right] \\ &= \frac{\exp\{-x\}}{m(x)\pi} \left[ \frac{\exp\{\theta\}(1 - 2\theta + \theta^2)}{(1 + \theta^2)^2} \right] \\ &= \frac{\exp\{-x\} \exp\{\theta\}}{m(x)\pi} \frac{(\theta - 1)^2}{(1 + \theta^2)^2}. \end{aligned}$$

Ya que la derivada es siempre positiva  $\pi(\theta|x)$  se decrementa para  $\theta \leq x$ , así se tiene que  $\pi(\theta|x)$  se maximiza en  $\hat{\theta} = x$ . Otro estimador bayesiano común de  $\theta$  es la media de la a posteriori  $\pi(\theta|x)$ .

### 1.6.2. Error de estimación.

Cuando se hace una estimación, es usualmente necesario indicar la precisión de la estimación. La medida bayesiana que se utiliza para medir la precisión de una estimación (en una dimensión) es la varianza a posteriori de la estimación .

**Definición 1.6.2** Si  $\theta$  es un parámetro de valor real con distribución a posteriori  $\pi(\theta|x)$ , y  $\delta$  es el estimador de  $\theta$ , entonces la varianza a posteriori de  $\delta$  es

$$V_{\delta}^{\pi}(x) = E^{\pi(\theta|x)}[(\theta - \delta)^2].$$

Nótese en la definición anterior, que al tener el estimador  $\hat{\theta}$  de  $\theta$ , en este caso  $\delta$ , solo se necesita sustituir a  $\delta$  en la definición de la varianza para obtener la varianza a posteriori.

Cuando  $\delta$  es la media a posteriori,

$$\mu^\pi(x) = E^{\pi(\theta|x)}[\theta],$$

entonces  $V^\pi(x) = V_\mu^\pi(x)$  será llamada varianza a posteriori (y es en efecto la varianza de  $\theta$  para la distribución  $\pi(\theta|x)$ ).

La desviación estándar a posteriori es  $\sqrt{V^\pi(x)}$ . Generalmente se utiliza a la desviación estándar a posteriori  $\sqrt{V_\delta^\pi(x)}$  del estimador  $\delta$ , como el error estándar de la estimación  $\delta$ .

Para simplificar cálculos que serán utilizados más adelante, se puede representar a la varianza a posteriori con la fórmula siguiente:

$$\begin{aligned} V_\delta^\pi(x) &= E^{\pi(\theta|x)}[(\theta - \delta)^2] \\ &= E[(\theta - \mu^\pi(x) + \mu^\pi(x) - \delta)^2] \\ &= E[(\theta - \mu^\pi(x))^2] + E[2(\theta - \mu^\pi(x))(\mu^\pi(x) - \delta)] + E[(\mu^\pi(x) - \delta)^2] \\ &= V^\pi(x) + 2(\mu^\pi(x) - \delta)(E[\theta] - \mu^\pi(x)) + (\mu^\pi(x) - \delta)^2 \\ &= V^\pi(x) + (\mu^\pi(x) - \delta)^2. \end{aligned}$$

Así, entonces

$$V_\delta^\pi(x) = V^\pi(x) + (\mu^\pi(x) - \delta)^2. \quad (1.4)$$

### 1.6.3. Estimación multivariada

La estimación bayesiana de un vector  $\theta = (\theta_1, \dots, \theta_n)$  es la estimación de máxima verosimilitud generalizada (moda a posteriori)  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_n)$ . Sin embargo, por cuestiones de cálculo y porque la unicidad y existencia no se garantizan en el caso multivariado, es más conveniente utilizar la media a posteriori

$$\mu^\pi(x) = (\mu_1^\pi(x), \mu_2^\pi(x), \dots, \mu_n^\pi(x))^t = E^{\pi(\theta|x)}[\theta],$$

como el estimador bayesiano, y la precisión puede ser descrita por la matriz de covarianza a posteriori.

$$V^\pi(x) = E^{\pi(\theta|x)}[(\theta - \mu^\pi(x))(\theta - \mu^\pi(x))^t].$$

El error estándar de la estimación  $\mu_i^\pi(x)$  de  $\hat{\theta}_i$  sería  $\sqrt{V_{ii}^\pi(x)}$ , en donde  $V_{ii}^\pi(x)$  es el  $(i, i)$  elemento de  $V^\pi(x)$ . Análogamente a la ecuación (1.7) se tiene que, el estimador  $\delta$  de  $\theta$ , es:

$$\begin{aligned} V_\delta^\pi(x) &= E^{\pi(\theta|x)}[(\theta - \delta)(\theta - \delta)^t] \\ &= V^\pi(x) + (\mu^\pi(x) - \delta)(\mu^\pi(x) - \delta)^t. \end{aligned}$$

Nótese que la media a posteriori minimiza  $V_\delta^\pi(x)$ .

#### 1.6.4. Conjuntos creíbles

El análogo bayesiano de un intervalo de confianza (IC), suele hacer referencia a un conjunto creíble.

**Definición 1.6.3** *Un  $100 \times (1 - \alpha)$  % conjunto creíble para  $\theta$ , es un subconjunto  $C$  de  $\Theta$  tal que:*

$$1 - \alpha \leq P(C|x) = \begin{cases} \sum_{\theta \in C} \pi(\theta|x) & \text{caso discreto} \\ \int_C \pi(\theta|x) d\theta & \text{caso continuo.} \end{cases}$$

Ya que la distribución a posteriori es actualmente una distribución de probabilidad en  $\Theta$ , se puede decir subjetivamente que la probabilidad de  $\theta$  está en  $C$ . Esta es la diferencia con los procedimientos de la estadística clásica los cuales sólo se interpretan como la probabilidad de que la variable aleatoria  $X$  sea tal que el conjunto creíble  $C(X)$  contenga a  $\theta$ .

Al escoger un conjunto creíble para  $\theta$ , de lo que se trata es siempre de minimizar el tamaño del intervalo. Para realizar esto, se deben incluir en el conjunto, solo aquellos puntos con la densidad a posteriori más grande, es decir, los más probables valores para  $\theta$ .

**Definición 1.6.4** *El conjunto creíble de máxima densidad a posteriori (HPD) al  $100 \times (1 - \alpha)$  %, es el subconjunto  $C$  en  $\Theta$  de la forma*

$$C = \{\theta \in \Theta : \pi(\theta|x) \geq k(\alpha)\},$$

en donde  $k(\alpha)$  es la mayor constante tal que:

$$P(C|x) \geq 1 - \alpha.$$

**Ejemplo 1.6.2**

Supóngase que  $X = (X_1, \dots, X_n)$  es una muestra aleatoria de una distribución Bernoulli con parámetro  $\theta$ . Supóngase además que 16 consumidores han sido seleccionados por una cadena de comida rápida para comparar dos tipos de hamburguesas de carne molida en base al sabor, en la cual una hamburguesa es más cara que la otra. Después de someterlos a la prueba, los consumidores dan sus puntos de vista de estos dos tipos de hamburguesa. El resultado es que 13 de los 16 consumidores prefieren la hamburguesa más cara. Supóngase que  $\theta$  es la probabilidad de que un consumidor prefiera la hamburguesa más cara.

$$Y_i = \begin{cases} 1 & \text{si el consumidor } i \text{ prefiere la hamburguesa más cara} \\ 0 & \text{en otro caso} \end{cases}.$$

Como las decisiones de los consumidores son independientes y  $\theta$  es constante sobre los consumidores, esto forma una secuencia de pruebas Bernoulli. Se define  $X = \sum Y_i$ , se tiene entonces que  $X|\theta \sim \text{Binomial}(16, \theta)$ , entonces la función de densidad de  $X$  es,

$$f(x|\theta) = \binom{16}{X} \theta^x (1 - \theta)^{16-x}.$$

Luego realizando los calculos pertinentes se llega a que

$$\pi(\theta|x) = \text{Beta}(\alpha + x, \beta - x + 16).$$

Utilizando el software ANESBA, se calcula el HPD al 95% y se obtiene el intervalo (0.6867, 0.9387), que quiere decir que la probabilidad de que el estimador de  $\theta$  este en ese intervalo es de 0.95.

En la figura (1.1) se puede ver la distribución a posteriori del consumo de hamburguesas, y en efecto se ve que el estimador se encuentra dentro del intervalo mencionado anteriormente.

**Diferencia entre los conjuntos creíbles y los intervalos de confianza**

Bayesianamente un conjunto creíble  $C$  al  $(1 - \alpha)100\%$  significa que, *la probabilidad de que  $\theta$  esté en  $C$  dados los datos observados es al menos  $(1 - \alpha)$* . En otras palabras, un conjunto creíble da la certeza de que dada la muestra



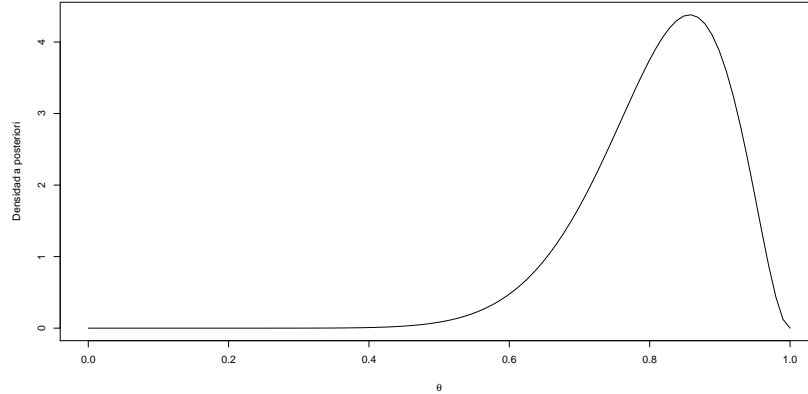


Figura 1.1: Gráfica de la distribución a posteriori derivada del consumo de hamburguesas.

aleatoria tal conjunto  $C$  contenga a  $\theta$  con al menos una probabilidad de  $(1 - \alpha)$ .

Clásicamente un intervalo de confianza al  $(1 - \alpha)100\%$  significa que, *si se pudiera calcular  $C$  para una gran cantidad de conjuntos de muestras aleatorias, cerca del  $(1 - \alpha)100\%$  de ellos cubrirían el valor real de  $\theta$ .*

Es decir, si se generan 100 conjuntos de muestras aleatorias, y se calcula el intervalo de confianza para cada caso (no necesariamente iguales para cada caso), al menos  $(1 - \alpha)100$  cubrirían a  $\theta$ , sin embargo, la desventaja principal que tienen los intervalos de confianza clásicos es que no siempre se pueden generar tantas muestras aleatorias.

### 1.6.5. Contraste de hipótesis

Una hipótesis estadística consiste en una afirmación acerca de un parámetro. Lo que se desea saber al formular una hipótesis es saber si es verdadera o falsa. Usualmente, la hipótesis de la que se pretende averiguar si es verdadera o falsa, se denomina *hipótesis nula*, esta se contrasta con la afirmación contraria a ella llamada *hipótesis alternativa*. En la estadística clásica para el contraste de hipótesis se tienen, la hipótesis *nula*  $H_0 : \theta \in \Theta_0$  y la hipótesis *alternativa*  $H_1 : \theta \in \Theta_1$ . El procedimiento de contraste se realiza en términos de las probabilidades de *error tipo I* y *error tipo II*, estas probabilidades de error representan la posibilidad de rechazar la hipótesis nula dado que es ver-

dadera o de no rechazar la hipótesis nula dado que es falsa, respectivamente.

Bayesianamente, la tarea de decidir entre  $H_0$  y  $H_1$  es más simple. Pues solamente se calculan las probabilidades a posteriori  $\alpha_0 = P(\Theta_0|x)$  y  $\alpha_1 = P(\Theta_1|x)$  y se decide entre  $H_0$  y  $H_1$ . La ventaja es que  $\alpha_0$  y  $\alpha_1$  son las probabilidades reales de las hipótesis de acuerdo a los datos observados y las opiniones a priori.

Se denotará como  $\pi_0$  y  $\pi_1$  a las probabilidades a priori de  $\Theta_0$  y  $\Theta_1$  respectivamente.

**Definición 1.6.5** *Se llamara razón a priori de  $H_0$  contra  $H_1$  al cociente  $\frac{\pi_0}{\pi_1}$ . Análogamente, razón a posteriori de  $H_0$  contra  $H_1$  al cociente  $\frac{\alpha_0}{\alpha_1}$ .*

Si la razón a priori es cercana a la unidad significa que se considera que  $H_0$  y  $H_1$  tienen inicialmente (casi) la misma probabilidad de ocurrir. Si este cociente de probabilidades es grande significa que se considera que  $H_0$  es más probable que  $H_1$  y si el cociente es pequeño ( $< 1$ ) es que inicialmente se considera que  $H_0$  es menos probable que  $H_1$ .

**Definición 1.6.6** *La cantidad*

$$B = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = \frac{\alpha_0\pi_1}{\alpha_1\pi_0},$$

*es llamada **factor de Bayes** en favor de  $\Theta_0$ .*

El factor de Bayes se interpreta como *la razón para  $H_0$  con respecto a  $H_1$  está dada por los datos*. Esto se puede ver fácilmente cuando se tienen hipótesis simples,  $\Theta_0 = \{\theta_0\}$  y  $\Theta_1 = \{\theta_1\}$ , así se tiene

$$\alpha_0 = \frac{\pi_0 f(x|\theta_0)}{\pi_0 f(x|\theta_0) + \pi_1 f(x|\theta_1)}, \alpha_1 = \frac{\pi_1 f(x|\theta_1)}{\pi_0 f(x|\theta_0) + \pi_1 f(x|\theta_1)},$$

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0 f(x|\theta_0)}{\pi_1 f(x|\theta_1)},$$

$$B = \frac{\alpha_0\pi_1}{\alpha_1\pi_0} = \frac{f(x|\theta_0)}{f(x|\theta_1)}.$$

En otras palabras,  $B$  es la razón de probabilidad de  $H_0$  contra  $H_1$ .

Además, nótese que el *factor de Bayes* proporciona una medida de la forma en que los datos aumentan o disminuyen las razones de probabilidades de  $H_0$  respecto de  $H_1$ . Así, si  $B > 1$  quiere decir que  $H_0$  es ahora relativamente más probable que  $H_1$  y si  $B < 1$  ha aumentado la probabilidad relativa de  $H_1$ .

### Contraste de una cola o unilateral

Este sucede cuando  $\Theta \subseteq \mathbb{R}$  y  $\Theta_0$  es totalmente un lado de  $\Theta_1$ , es decir.

$$\begin{aligned} H_0 &: \theta \leq \theta_0 \\ H_1 &: \theta > \theta_0, \end{aligned}$$

o el otro caso

$$\begin{aligned} H_0 &: \theta \geq \theta_0 \\ H_1 &: \theta < \theta_0. \end{aligned}$$

Lo interesante de este tipo de contraste es que el uso del *valor-p* del enfoque clásico tiene una justificación bayesiana.

#### Ejemplo 1.6.3

Supóngase que  $X \sim N(\theta, \sigma^2)$  y  $\theta$  tiene una a priori no informativa  $\pi(\theta) = 1$ , entonces

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

El contraste a realizar es el siguiente:

$$\begin{aligned} H_0 &: \theta \leq \theta_0 \\ H_1 &: \theta > \theta_0, \end{aligned}$$

luego,

$$\alpha_0 = P(\theta \leq \theta_0|x) = \Phi((\theta_0 - x)/\sigma).$$

Clásicamente el *valor-p* es la probabilidad, cuando  $\theta = \theta_0$ . Aquí el *valor-p* es,

$$\text{valor-p} = P(X \geq x) = 1 - \Phi((x - \theta_0)/\sigma),$$

ya que la distribución normal es simétrica, se tiene que el *valor-p* es el mismo que  $\alpha_0$ .

### Contraste de hipótesis nula puntual

El contraste de hipótesis nula puntual se refiere al siguiente contraste

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

que es interesante porque a diferencia del contraste clásico, contiene nuevas características pero principalmente porque las respuestas en el enfoque bayesiano difieren radicalmente de las que da el enfoque clásico.

Casi nunca se da el caso en que  $\theta = \theta_0$  exactamente. Es más razonable tener a la hipótesis nula como  $\theta \in \Theta_0 = (\theta_0 - b, \theta_0 + b)$ , en donde  $b > 0$  es alguna constante elegida de tal forma que para cada  $\theta \in \Theta_0$  puede ser considerada indistinguible de  $\theta_0$ .

Si  $b$  llega a ser muy grande, ya no es recomendable utilizar el contraste por hipótesis nula puntual, porque no da buenos resultados. Dado que se debe hacer el contraste  $H_0 : \theta \in (\theta_0 - b, \theta_0 + b)$ , se necesita saber cuando es apropiado aproximar  $H_0$  por  $H_0 : \theta = \theta_0$ . Desde la perspectiva del análisis bayesiano, la única respuesta a esta pregunta es, *la aproximación es razonable si las probabilidades a posteriori de  $H_0$  son casi iguales en las dos situaciones*. Una condición fuerte cuando éste sea el caso es que la función de distribución conjunta de probabilidad (verosimilitud) observada sea aproximadamente constante en  $(\theta_0 - b, \theta_0 + b)$ .

Para llevar a cabo el contraste bayesiano para una hipótesis nula puntual  $H_0 : \theta = \theta_0$ , no se utiliza una densidad a priori continua, ya que en cualquiera de estos casos  $\theta_0$  dara una probabilidad a priori cero. Se debe utilizar una distribución inicial mixta que asigne probabilidad  $\pi_0$  al punto  $\theta_0$  y distribuya el resto,  $1 - \pi_0$  en los valores  $\theta \neq \theta_0$  es decir la densidad  $\pi_1 g_1(\theta)$ , con una densidad propia  $g_1$ , en donde  $\pi_1 = 1 - \pi_0$ .

Si  $X_1, \dots, X_n$  es una muestra aleatoria simple de la variable aleatoria  $X$ , la densidad marginal de  $X = (X_1, \dots, X_n)$  es:

$$m(x) = \pi_0 f(x|\theta_0) + (1 - \pi_0) m_1(x),$$

donde  $m_1(x) = \int_{\theta \neq \theta_0} f(x|\theta) g_1(\theta) d\theta$  es la densidad bajo  $H_1$ . Entonces la prob-

abilidad a posteriori de  $\theta = \theta_0$  es:

$$\begin{aligned}\alpha_0 &= P(\theta_0|x) \\ &= \frac{f(x|\theta_0)\pi_0}{m(x)} \\ &= \frac{f(x|\theta_0)\pi_0}{f(x|\theta_0)\pi_0 + (1 - \pi_0)m_1(x)} \\ &= \frac{f(x|\theta_0)\pi_0}{f(x|\theta_0)\pi_0 + \pi_1 m_1(x)},\end{aligned}$$

luego

$$\begin{aligned}\frac{\alpha_0}{\alpha_1} &= \frac{P(\theta_0|x)}{1 - \pi(\theta_0|x)} \\ &= \frac{P(\theta_0|x)}{\frac{f(x|\theta_0)\pi_0 + \pi_1 m_1(x) - f(x|\theta_0)\pi_0}{f(x|\theta_0)\pi_0 + \pi_1 m_1(x)}} \\ &= \frac{f(x|\theta_0)\pi_0}{\pi_1 m_1(x)}.\end{aligned}$$

Entonces el factor de Bayes de  $H_0$  contra  $H_1$  es,

$$\begin{aligned}B &= \frac{\alpha_0\pi_1}{\alpha_1\pi_0} \\ &= \frac{\pi_0 f(x|\theta)\pi_1}{\pi_1 m_1(x)\pi_0},\end{aligned}$$

$$B = \frac{f(x|\theta_0)}{m_1(x)}.$$

### 1.6.6. Inferencia predictiva

Los problemas predictivos en estadística son aquellos en los cuales las cantidades desconocidas de interés son las futuras variables aleatorias.

Una situación típica de este tipo es cuando la variable aleatoria  $Z$ , tiene una densidad  $g(z|\theta)$  ( $\theta$  desconocida), y se quiere predecir un valor de  $Z$  cuando existen datos,  $x$ , derivados de una densidad  $h(x|\theta)$ . Por ejemplo,  $x$  pueden ser los datos de un estudio de regresión, y se desea predecir una futura respuesta

de la variable aleatoria  $Z$ , en la regresión.

La idea es predecir a la variable aleatoria  $Z \sim g(z|\theta)$  basado en la observación de  $X \sim h(x|\theta)$ .

Ahora, se supone que  $X$  y  $Z$  son independientes y  $g$  es una densidad (si no fueran independientes, sería necesario cambiar  $g(z|\theta)$  por  $g(z|\theta, x)$ ). La idea de la predicción bayesiana es que, ya que  $\pi(\theta|x)$  es la distribución a posteriori de  $\theta$ , entonces  $g(z|\theta)\pi(\theta|x)$  es la distribución conjunta de  $Z$  y  $\theta$  dado  $X$ , y al integrar sobre  $\theta$  se obtiene la distribución a posteriori de  $Z$  dado  $X$ .

**Definición 1.6.7** La *densidad predictiva* de  $Z$  dado  $X = x$ , cuando la distribución a priori para  $\theta$  es  $\pi$ , está definida por

$$P(z|x) = \int_{\Theta} g(z|\theta)\pi(\theta|x)d\theta.$$

#### Ejemplo 1.6.4

Consideremos el siguiente modelo de regresión lineal.

$$Z = \theta Y + \varepsilon,$$

donde  $\varepsilon \sim N(0, \sigma^2)$  con  $\sigma^2$  conocido, y supóngase que se tiene un conjunto de datos disponibles  $((z_1, y_1), \dots, (z_n, y_n))$ .

Un estadístico suficiente para  $\theta$ , es el estimador por mínimos cuadrados  $X$ , en donde

$$X = \frac{\sum_{i=1}^n Z_i y_i}{\sum_{i=1}^n y_i^2}.$$

Como  $Z_i$  es normal y  $X$  es combinación lineal de normales, entonces  $X$  tendrá una distribución normal.

$$X \sim N(\theta, \sigma^2/Sy),$$

en donde:

$$Sy = \sum_{i=1}^n y_i^2.$$

Si se utiliza una a priori no informativa  $\pi(\theta) = 1$  para  $\theta$ , la distribución a posteriori,  $\pi(\theta|x) \sim N(x, \sigma^2/Sy)$ .

Ahora se obtiene la probabilidad de  $Z$  dado  $X$ :

$$\begin{aligned} P(z|x) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(z - \theta y)^2\right\} \frac{1}{\sqrt{2\pi}\sigma/\sqrt{S_y}} \times \\ &\quad \exp\left\{-\frac{1}{2\sigma^2/S_y}(\theta - x)^2\right\} d\theta \\ &= \frac{\sqrt{S_y}}{2\pi\sigma^2} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}[(z - \theta y)^2 + S_y(\theta - x)^2]\right\} d\theta. \end{aligned}$$

Desarrollando separadamente  $[(z - \theta y)^2 + S_y(\theta - x)^2]$ , se tiene:

$$\begin{aligned} [(z - \theta y)^2 + S_y(\theta - x)^2] &= z^2 + S_y x^2 + \theta^2 y^2 + S_y \theta^2 - 2S_y x \theta + S_y x^2 \\ &= z^2 + S_y x^2 + \theta^2 (y^2 + S_y) - 2\theta (yz + S_y x) \\ &= (y^2 + S_y) \left( \frac{z^2 + S_y x^2}{y^2 + S_y} - \frac{(yz + S_y x)^2}{(y^2 + S_y)^2} + \left( \theta - \frac{yz + S_y x}{y^2 + S_y} \right)^2 \right). \end{aligned}$$

Luego substituyendo nuevamente la expresión anterior en  $p(z|x)$ ,

$$\begin{aligned} P(z|x) &= \frac{\sqrt{S_y}}{2\pi\sigma^2} \exp\left\{-\frac{(y^2 + S_y)}{2\sigma^2} \left[ \frac{z^2 + S_y x^2}{y^2 + S_y} - \frac{(yz + S_y x)^2}{(y^2 + S_y)^2} \right]\right\} \times \\ &\quad \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2/(y^2 + S_y)} \left( \theta - \frac{yz + S_y x}{y^2 + S_y} \right)^2\right\} d\theta \\ &= \frac{\sqrt{S_y}}{2\pi\sigma^2} \exp\left\{-\frac{(y^2 + S_y)}{2\sigma^2} \left[ \frac{z^2 + S_y x^2}{y^2 + S_y} - \frac{(yz + S_y x)^2}{(y^2 + S_y)^2} \right]\right\} \left( \frac{\sigma}{\sqrt{y^2 + S_y}} 2\pi \right) \\ &= \frac{1}{\sqrt{2\pi}\sigma \sqrt{\frac{y^2 + S_y}{S_y}}} \exp\left\{-\frac{1}{2\sigma^2 \left(\frac{y^2 + S_y}{S_y}\right)} (z - xy)^2\right\}. \end{aligned}$$

Finalmente se tiene:

$$P(z|x) = N\left(xy, \frac{\sigma^2(y^2 + S_y)}{S_y}\right).$$





## Capítulo 2

# Conceptos astronómicos.

### 2.1. Sistemas de coordenadas.

La rotación diaria de la Tierra junto con su movimiento alrededor del Sol y la precesión sobre su eje de rotación, aunado al movimiento relativo de las estrellas, planetas y otros objetos, dan lugar al constante cambio de posición aparente de los objetos celestes.

Se puede imaginar que todos los objetos están localizados en la esfera celeste. Para localizar la posición de un objeto en ésta esfera celeste es necesario crear un sistema de coordenadas, el cual consta de tres componentes básicas; un gran círculo, la elección de una dirección positiva del eje de simetría del círculo y un punto de referencia sobre este círculo.

**Definición 2.1.1** *Gran círculo* es la curva resultante de la intersección de una esfera con un plano que pasa por el centro de dicha esfera.

Se necesita un punto por encima del *gran círculo*, que es el que nos va a dar la referencia de Norte y Sur. Y un punto de referencia en el perímetro del *gran círculo*, apartir del cual se tomaran las medidas.

Un ejemplo de esto es el sistema de coordenadas terrestre, con el ecuador como gran círculo, el polo Norte como el punto encima del círculo y finalmente el punto en el que el meridiano de Greenwich que corta al ecuador terrestre como punto de referencia sobre el círculo.

Cabe destacar que los sistemas de coordenadas que se utilizan en la Astronomía son sistemas de coordenadas en dos dimensiones, y no en tres di-

mensiones como se piensa que debería ser, la razón es que no se tiene idea de la medida de profundidad de los objetos celestes.

El sistema mas simple de coordenadas que se puede concebir se basa en el horizonte local del observador. El sistema de coordenadas *altitud-acimut* está basado en la medida del ángulo acimut a lo largo del horizonte y el ángulo de altitud por encima del horizonte.

*Cenit* es el punto justo arriba del observador.

*Vertical* son llamados los gran círculos que pasan por el cenit.

**Definición 2.1.2** La **altura  $h$**  es la medida del ángulo, desde el horizonte hasta la vertical que pasa a través del objeto, se mide en el rango de  $[-90^\circ, +90^\circ]$ , es positiva si el objeto se encuentra por arriba del ecuador celeste y negativa si se encuentra por debajo.

**Definición 2.1.3** La **distancia cenital  $z$**  es el ángulo medido del cenit al objeto, y cumple que  $z + h = 90^\circ$  (complemento de la altura).

**Definición 2.1.4** El **acimut  $A$**  es la medida del ángulo a lo largo del horizonte hacia el este desde el norte en el gran círculo, utilizado para la medida de altitud.

En la Figura 2.1, se puede ver como se ubican estas medidas en la esfera celeste.

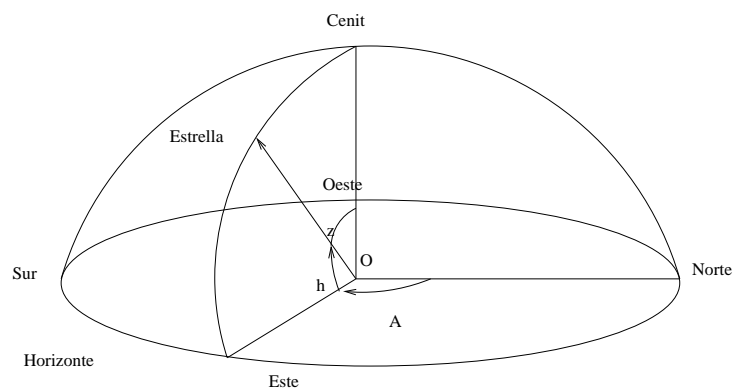


Figura 2.1: Sistema de coordenadas altitud-acimut.

Sin embargo el sistema de coordenadas altitud-acimut es difícil de utilizar en la práctica, porque las coordenadas dependen de la ubicación del observador. Dos personas en dos lugares distintos medirán coordenadas diferentes para el mismo objeto. Para la definición de un sistema más apropiado se necesita definir el *ecuador celeste* y la *eclíptica*.

**Definición 2.1.5** *El **ecuador celeste** es un gran círculo en la esfera celeste en el mismo plano que el ecuador de la Tierra y por tanto perpendicular al eje de rotación de la Tierra. En otras palabras, es la intersección entre el ecuador terrestre en la esfera celeste.*

**Definición 2.1.6** *La **Eclíptica** es la línea curva por donde transcurre el Sol alrededor de la Tierra, en su movimiento aparente visto desde la Tierra. Está formada por la intersección del plano de la órbita terrestre con la esfera celeste.*

Como resultado de la inclinación del eje de rotación de la Tierra, la *eclíptica* está inclinada  $23.5^\circ$  con respecto al *ecuador celeste*.

Los dos puntos de la esfera celeste en los que se corta la eclíptica con el *ecuador celeste* son denominados equinoccios. En este caso el *ecuador celeste* será tomado como un gran círculo, el punto por encima del *ecuador celeste* que se utiliza en este sistema de coordenadas, es llamado *polo norte celeste* y es la intersección entre el eje de rotación de la Tierra y la esfera celeste. Finalmente, el punto de referencia sobre el círculo será el *equinoccio vernal*.

**Definición 2.1.7** *La **declinación**  $\delta$  es equivalente a la altitud y es medida en grados de norte a sur a partir del ecuador celeste.*

**Definición 2.1.8** *La **ascensión recta**  $\alpha$  es análoga al acimut y es medida en dirección este a lo largo del ecuador celeste, desde el equinoccio vernal ( $\gamma$ ) hasta su intersección con el círculo horario del objeto.*

La ascensión recta se mide en horas, minutos y segundos, 24 horas de ascensión recta son equivalentes a  $360^\circ$ , o 1 hora =  $15^\circ$ .

Ya que el sistema de coordenadas ecuatoriales se basa en el ecuador celeste y el equinoccio vernal, cambios en la latitud y longitud del observador no afectan los valores de la ascensión recta y la declinación. Los valores de  $\alpha$

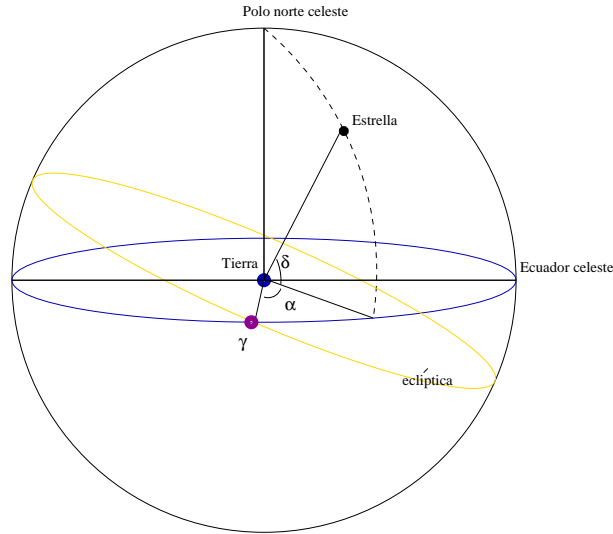


Figura 2.2: Sistema de coordenadas ecuatoriales.

$\alpha$ ,  $\delta$ ,  $\gamma$  son la ascensión recta, declinación y posición del equinoccio vernal respectivamente.

y  $\delta$  no son afectados similarmente por el movimiento de la tierra alrededor del sol. En la Figura 2.2, se muestra el sistema de coordenadas ecuatoriales descrito.

A pesar de las referencias del sistema de coordenadas ecuatorial con el ecuador celeste y su intersección con la *eclíptica* (equinoccio vernal), existe un movimiento llamado *precesión*, el cual hace que el equinoccio vernal no suceda en el mismo lugar cada año, y por lo tanto la ascensión recta y la declinación del objeto cambien un poco.

**Definición 2.1.9** *Precesión* es el lento tambaleo del eje de rotación de la Tierra, debido a la forma no esférica de ésta y a su interacción gravitacional con el Sol y la Luna.

El período de precesión de la Tierra es de 25 770 años y causa que la proyección del polo norte celeste sobre la esfera celeste describa un pequeño círculo (Veáse [6]). Ya que la precesión altera la posición del *equinoccio vernal* a lo largo de la *eclíptica*, es necesario referirse a una época específica (referencia) para definir la *ascensión recta* y la *declinación* de un objeto celeste. El valor actual de  $\alpha$  y  $\delta$ , se pueden calcular basándose en la cantidad de tiempo transcurrido a partir de la época de referencia.

La época utilizada comunmente en la actualidad por los catálogos astronómicos de estrellas, galaxias y otros fenómenos celestes refiere a la posición de un objeto al medio día en Greenwich, Inglaterra el 1 de enero del 2000.

El calendario utilizado comunmente en la mayoría de los países es el *calendario Gregoriano*, sin embargo los astrónomos no se preocupan por los años bisiestos como en el calendario Gregoriano, a ellos más bien lo que les interesa es el número de días (o segundos) entre eventos. Ellos normalmente cuentan los eventos a partir de tiempo 0 que fue el 1 de enero de 4713 A.C. como lo especifica el calendario Gregoriano, y JD 0.0 para el calendario Juliano. La fecha de referencia 1 de enero del 2000 (J2000.0) es JD 2451545.0, y para referirse a las 6 pm del día 1 de enero del 2000 lo que se hace es que se suma la fracción del día entonces esta fecha se designa como JD 2451545.25. El año Juliano consta de 365.25 días (Veáse [6]).

## 2.2. Determinación de las ecuaciones.

Las estrellas tienen un movimiento propio, a través de la bóveda celeste que se puede describir generalmente como un movimiento con una aceleración constante.

Cuando un objeto se encuentra en movimiento, con una aceleración constante, es importante determinar su posición  $x(t)$  al instante  $t$ , la cual está dada por las ecuaciones de la cinemática.

$$x(t) = x_0 + v_0t + \frac{1}{2}at^2, \quad (2.1)$$

donde:

$v_0$ : velocidad del objeto al tiempo  $t = 0$ .

$x_0$ : la posición del objeto al tiempo  $t = 0$ ,

$a$ : aceleración.

El movimiento de las estrellas en la bóveda celeste refelaja un cambio de sus coordenadas ecuatoriales. El Sol, la Luna y los planetas manifiestan un movimiento complejo y relativamente rápido en el cielo. Las estrellas también se mueven con respecto de las demás, aunque su velocidad actual puede ser muy grande, su movimiento relativo aparente es muy difícil de medir por la grandes distancias a las que se encuentran.

Considérese la velocidad de una estrella relativa a un observador (Figura 2.3). Con  $\theta$  el ángulo de cambio en el movimiento,  $D$  distancia a la estrella suponiendo que es constante y  $r$  distancia que se mueve la estrella.

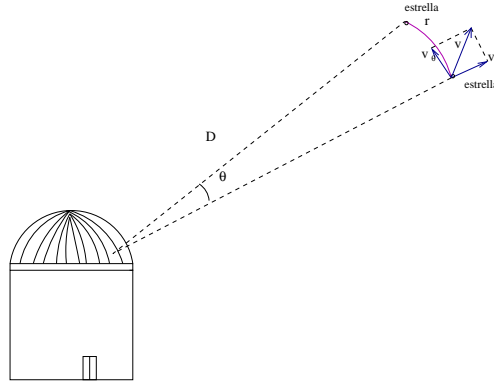


Figura 2.3: Movimiento propio.

El vector velocidad se puede descomponer en componentes mutuamente perpendiculares, una a lo largo de la línea de visión del observador y la otra perpendicular. La componente a lo largo de la línea de visión es la **velocidad radial**,  $v_r$  de la estrella y la segunda componente que es la de nuestro interés es la **velocidad transversal**,  $v_\theta$  de la estrella, a lo largo de la esfera celeste. Esta velocidad transversal aparece como un cambio angular lento en el sistema de coordenadas, conocido como **movimiento propio**.

En un intervalo de tiempo  $dt$  la estrella se mueve en dirección perpendicular a la línea de visión del observador una distancia

$$dr = Dd\theta.$$

Si la distancia del observador a la estrella es  $D$ , entonces el cambio angular en su posición a lo largo de la esfera celeste es  $\frac{d\theta}{dt}$ , que se obtiene al derivar la distancia recorrida por la estrella con respecto al tiempo.

$$\frac{dr}{dt} = \frac{d(D\theta)}{dt} = D\frac{d\theta}{dt},$$

en donde

$$\frac{d\theta}{dt} = \mu : \text{movimiento propio de la estrella.}$$

Al encontrar la segunda derivada de la distancia recorrida por la estrella con respecto al tiempo, se tiene que

$$\frac{dr^2}{dt^2} = \frac{d^2(D\theta)}{dt^2} = D \frac{d^2\theta}{dt^2}.$$

Así como se esta suponiendo una aceleración constante, de acuerdo a la ecuación (2.1),

$$\begin{aligned} r &= r_0 + D \frac{d\theta}{dt} t + D \frac{1}{2} \frac{d^2\theta}{dt^2} t^2, \\ D\theta &= D\theta_0 + D \frac{d\theta}{dt} t + D \frac{1}{2} \frac{d^2\theta}{dt^2} t^2, \\ \theta &= \theta_0 + \frac{d\theta}{dt} t + \frac{1}{2} \frac{d^2\theta}{dt^2} t^2. \end{aligned} \quad (2.2)$$

Ahora aplicando, la ecuación (2.2) se tienen las ecuaciones de posición para la ascensión recta ( $\alpha$ ) y la declinación ( $\delta$ ), respectivamente.

$$\alpha(t) = \alpha_0 + \mu_\alpha t + \frac{1}{2} a_\alpha t^2, \quad (2.3)$$

$$\delta(t) = \delta_0 + \mu_\delta t + \frac{1}{2} a_\delta t^2, \quad (2.4)$$

en donde:

$\alpha_0$ : posición en  $\alpha$  al tiempo 0, en este caso para el año 2000 (J2000),

$\delta_0$ : posición en  $\delta$  al tiempo 0, J2000,

$\mu_\alpha$ : movimiento propio en  $\alpha$ ,

$\mu_\delta$ : movimiento propio en  $\delta$ ,

$a_\alpha$ : aceleración en  $\alpha$ ,

$a_\delta$ : aceleración en  $\delta$ .

Sin embargo, tomando en cuenta que dadas las ascensiones rectas  $\alpha_1$  y  $\alpha_2$  de dos estrellas la diferencia es  $\Delta\alpha = (\alpha_1 - \alpha_2) \cos \delta$ , entonces la ecuación (2.3) queda de la siguiente forma,

$$\alpha(t) = \alpha_0 + (\mu_\alpha \cos \delta) t + \frac{1}{2} (a_\alpha \cos \delta) t^2. \quad (2.5)$$

## 2.3. Paralaje estelar

Para obtener distancias, un método muy utilizado es el de la *paralaje trigonométrica*, que consiste en una técnica de triangulación.

La distancia  $D$  a los planetas en la bóveda celeste puede ser medida de dos puntos de observación separados por una distancia grande en la tierra. Encontrar la distancia siempre a la estrella más cercana requiere una gran línea de referencia más grande que el diámetro de la tierra. Como la tierra orbita alrededor del Sol, dos observaciones de la misma estrella tardan 6 meses en hacer una línea de referencia igual al diámetro de la órbita de la tierra. Ver Figura 2.4.

**Definición 2.3.1** *La medida del ángulo de paralaje  $p$  es la mitad en el máximo cambio en la posición angular.*

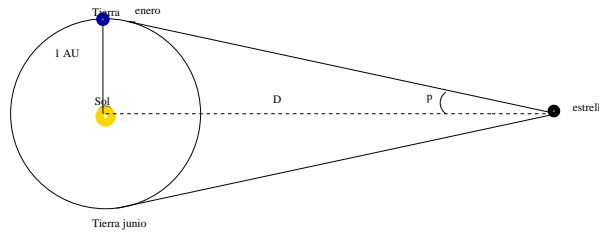


Figura 2.4: Paralaje.

Como se ve en la Figura 2.4 utilizando la técnica de triangulación mencionada y la medida del ángulo de paralaje se puede encontrar la distancia  $D$  a la estrella por medio de la siguiente ecuación,

$$D = \frac{1AU}{\tan p} \simeq \frac{1}{p}AU, \quad (2.6)$$

donde la aproximación  $\tan p \simeq p$  se emplea para el ángulo de paralaje  $p$ , medido en radianes.

Utilizando  $1 \text{ radian} = 57.2957795^\circ = 206264.806''$ , el convertir a  $p$  a  $p''$  en unidades de arcosegundos, nos da una distancia:

$$D = \frac{2206264.806}{p''}AU.$$



**Definición 2.3.2** *Parsec* (*parallax-second en inglés*) es una unidad de medida utilizada para medir distancias en la esfera celeste y está dada por:  
 $1 pc = 2.06264806 \times 10^5 AU = 3.086 \times 10^{16} m$ .

Así de acuerdo a la definición:

$$D = \frac{1}{p''} pc. \quad (2.7)$$

Cuando el ángulo de paralaje es  $p = 1''$ , la distancia a la estrella es  $1 pc$ . Entonces 1 parsec es la distancia a partir de la cual el radio de la órbita de la tierra, 1UA, subtienda un ángulo de  $1''$ , como se ilustra en la Figura 2.5.

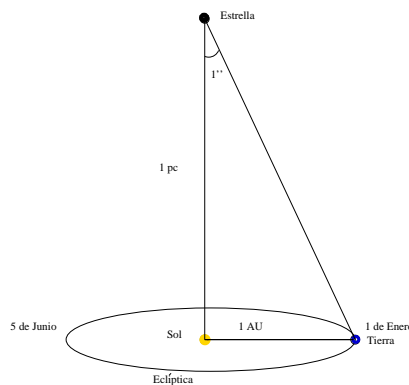


Figura 2.5: El parsec.

Otra unidad de distancia utilizada es el *año-luz* (ly por sus siglas en inglés) y es la distancia que recorre la luz en un año juliano  $1 ly = 9.460730472 \times 10^{15} m$ , un parsec equivale a 3.2615638 ly.

Las estrellas tienen un movimiento propio, sin embargo, vistas desde la Tierra, parecen tener un movimiento aparente, que por el ángulo de paralaje, parece ser un círculo, pero dependiendo de la posición en que se encuentre la estrella, será una línea, un círculo o una elipse.

**Definición 2.3.3** *El cambio en la dirección de un objeto celeste causado por el cambio en la posición del observador se conoce como **movimiento paraláctico**.*

La paralaje de una estrella se conoce comunmente como *paralaje anual*, ya que el movimiento paraláctico varía en un período de un año, Figura 2.6 y

también porque el origen del cambio paraláctico se da en el baricentro del Sistema Solar (el catálogo de posiciones de estrellas y otros objetos fuera del sistema solar por lo general se basan en el baricentro del Sistema Solar).

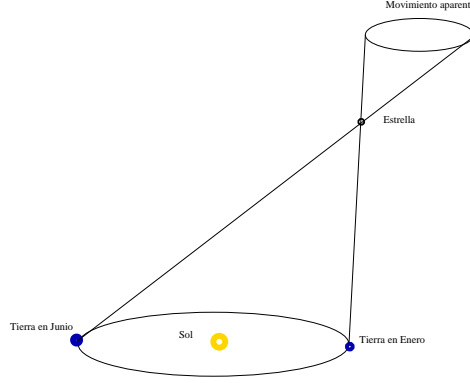


Figura 2.6: Movimiento paraláctico.

El desplazamiento de una estrella en la esfera celeste es la combinación de su paralaje y movimiento propio. Para fuentes aisladas, se considera un movimiento propio uniformemente acelerado, con ascensión recta y declinación que varían respecto al tiempo, cuyas ecuaciones están dadas por:

$$\alpha(t) = \alpha_0 + (\mu_{\alpha 0} \cos \delta)t + \frac{1}{2}(a_{\alpha} \cos \delta)t^2 + pf_{\alpha}(t), \quad (2.8)$$

$$\delta(t) = \delta_0 + \mu_{\delta 0}t + \frac{1}{2}a_{\delta}t^2 + pf_{\delta}(t), \quad (2.9)$$

en donde,

$(\alpha_0, \delta_0)$ : posición baricéntrica en la época de referencia,

$\mu_{\alpha}, \mu_{\delta}$ : componentes del movimiento propio,

$f_{\alpha}, f_{\delta}$ : proyección de la elipse paraláctica sobre  $\alpha$  y  $\delta$  respectivamente,

$a_{\alpha}, a_{\delta}$ : aceleración sobre  $\alpha$  y  $\delta$  respectivamente.

$p$ : paralaje en arcosegundos,

Los valores de  $f_{\alpha}, f_{\delta}$  que son el *paralaje anual* están dados por,

$$f_{\alpha}(t) = (X \sin \alpha_1 - Y \cos \alpha_1)/(15 \cos \delta_1),$$

$$f_{\delta}(t) = (X \cos \alpha_1 \sin \delta_1 + Y \sin \alpha_1 \sin \delta_1 - Z \cos \delta_1)$$

con  $X, Y, Z$ : coordenadas baricéntricas de la Tierra en AU, y

$$\alpha_1 = \alpha - pf_\alpha(t),$$

$$\delta_1 = \delta - pf_\delta(t).$$

## 2.4. T Tauri

*T Tauri* es una estrella joven, que se encuentra dentro de una nube de gas y esto hace imposible conocer todo de ella a simple vista, sin embargo, a través de ondas de radio *T Tauri* se conoce como un sistema de dos estrellas *T Tau* y *T Tau Sb*, observaciones infrarrojas más recientes revelaron que es un sistema triple formado por *T Tau*, *T Tau S* y *T Tau Sb*. En esta tesis los datos que se trabajaran son de *T Tau Sb* y los datos disponibles son los de ondas de radio.

En el artículo *VLBA determination of the distance to nearby star-forming regions. I. The distance to T TAURI with 0.4 % accuracy* de Loinard et. al. (2007) se mencionan los datos, de las *ascensiones rectas* y *declinaciones*, observadas para *T Tauri Sb*, durante doce épocas entre septiembre de 2006 y julio de 2007, se muestran a continuación en la Tabla 2.1.

Tabla 2.1. VLBA Distance to Nearby Star-Forming Regions. I.

Source Position and Flux		
Mean UT Date (yyyy.mm.dd hh:mm)	$\alpha$ (J2000.0)	$\delta$ (J2000.0)
2003.09.24 11:33	04 21 59.4252942	19 32 05.717618
2003.11.18 08:02	04 21 59.4249805	19 32 05.716554
2004.01.15 04:09	04 21 59.4245823	19 32 05.715322
2004.03.26 23:26	04 21 59.4245420	19 32 05.715333
2004.05.13 20:17	04 21 59.4248818	19 32 05.716034
2004.07.08 16:37	04 21 59.4253464	19 32 05.716652
2004.09.16 11:59	04 21 59.4255476	19 32 05.716602
2004.11.09 08:27	04 21 59.4252999	19 32 05.715631
2004.12.28 05:14	04 21 59.4249488	19 32 05.714344
2005.02.24 01:26	04 21 59.4247667	19 32 05.713826
2005.05.09 20:32	04 21 59.4251475	19 32 05.714852
2005.07.08 16:36	04 21 59.4256679	19 32 05.715598
2008.06.12 19:09	04 21 59.4268831	19 32 05.706772*

\*Las primeras doce observaciones fueron tomadas de [12], la última proporcionada por Laurent Loinard C. (CRyA-UNAM).

Nota. Las unidades de  $\alpha$  son horas, minutos y segundos y los de  $\delta$  son grados, arcominutos y arcosegundo.

Tabla 2.2. Julian Dates and Earth Coordinates.

Mean UT Date (yyyy.mm.dd hh:mm)	Julian Date	Earth Barycentric Coordinates (AU)		
		X	Y	Z
2003.09.24 11:33 ...	2,452,906.981522	+1.006064570	+0.012414145	+0.005329883
2003.11.18 08:02 ...	2,452,961.834705	+0.563031813	+0.744728145	+0.322808936
2004.01.15 04:09 ...	2,453,019.672980	-0.401044530	+0.820090543	+0.355473794
2004.03.26 23:26 ...	2,453,091.476395	-0.987542097	-0.107099184	-0.046513144
2004.05.13 20:17 ...	2,453,139.345324	-0.599629779	-0.745571223	-0.323319720
2004.07.08 16:37 ...	2,453,195.192419	+0.297481514	-0.894474582	-0.387883413
2004.09.16 11:59 ...	2,453,264.999583	+1.003677539	-0.099008512	-0.043027999
2004.11.09 08:27 ...	2,453,318.852141	+0.676914299	+0.666368259	+0.288787192
2004.12.28 05:14 ...	2,453,367.718351	-0.111308860	+0.895710071	+0.388210192
2005.02.24 01:26 ...	2,453,425.559664	-0.896015668	+0.377089952	+0.163364585
2005.05.09 20:32 ...	2,453,500.355214	-0.655044089	-0.701050689	-0.304055855
2005.07.08 16:36 ...	2,453,560.191348	+0.293538593	-0.893249243	-0.387385193
2008 06 12 19:09...	2,454,616.298137	-0.369805544	-0.861747353	-0.373653953

## Capítulo 3

# Análisis de regresión lineal

El análisis de regresión es una técnica estadística para investigar y modelar la relación entre variables. Las aplicaciones de la regresión se encuentran en numerosos campos de la investigación como en ingeniería, ciencias físicas y químicas, biología etc. De hecho es una de las técnicas estadísticas más utilizadas.

En este Capítulo, interesa describir la distribución condicional de una variable aleatoria  $Y$  para valores dados de algunas otras variables  $X_1, \dots, X_k$ . Los valores de las variables aleatorias  $X_1, \dots, X_k$  pueden ser observados en un experimento junto con los de  $Y$  o pueden ser variables de control cuyos valores van a ser determinados por el experimentador, además, se puede estudiar la distribución condicional de  $Y$  dadas  $X_1, \dots, X_k$ .

**Definición 3.0.1** Se denomina **función de regresión** de  $Y$  sobre  $X_1, \dots, X_k$ , o simplemente **regresión** de  $Y$  sobre  $X_1, \dots, X_k$  a la esperanza condicional de  $Y$  dado los valores  $X_1 = x_1, \dots, X_k = x_k$ , es decir,

$$E[Y|x_1, \dots, x_k].$$

Supóngase que la función de regresión  $E[Y|x_1, \dots, x_k]$  es una función lineal que tiene la siguiente forma

$$E[Y|x_1, \dots, x_k] = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k.$$

Los coeficientes  $\beta_0, \beta_1, \dots, \beta_k$  se llaman **coeficientes de regresión**, los cuales son desconocidos y se consideran parámetros cuyos valores se estiman. También supóngase que se tienen  $n$  vectores de observaciones. El vector  $i$ -ésimo

$(x_{i1}, \dots, x_{ik}, y_i)$  se compone de un conjunto de valores controlados u observados de  $X_1, \dots, X_k$  y el valor observado de  $Y$ .

### 3.1. Regresión lineal simple.

En el modelo de regresión lineal simple, se tiene un sólo regresor  $X$  que tiene una relación con una variable respuesta  $Y$ , donde la relación es una línea recta.

Este modelo de regresión lineal simple es de la forma,

$$Y = \beta_0 + \beta_1 x + \varepsilon,$$

donde la ordenada al origen  $\beta_0$  y la pendiente  $\beta_1$  son constantes desconocidas, y  $\varepsilon$  es un componente aleatorio del error.

Supóngase que los errores  $\varepsilon_i \sim N(0, \sigma^2)$  y no están correlacionados.

Considere que el regresor  $X$  se tiene como dato, y su error es despreciable, mientras que la respuesta  $Y$  es una variable aleatoria, se tiene entonces una distribución de probabilidad de  $Y$  para cada valor posible de  $X$ , así la media y la varianza de la distribución son

$$\begin{aligned} E[Y|X = x] &= E[\beta_0 + \beta_1 x + \varepsilon] \\ &= \beta_0 + \beta_1 x. \\ \text{Var}[Y|X = x] &= \sigma^2. \end{aligned}$$

A un problema como el descrito anteriormente se le llama de **regresión lineal simple**, y el término simple se emplea por que se está considerando la regresión de  $Y$  sobre una sola variable  $X$ .

Como los parámetros  $\beta_0$  y  $\beta_1$  son desconocidos, se deben estimar con los datos de la muestra. Supóngase que se tiene  $n$  pares de observaciones  $(x_i, y_i)$  con  $i = 1, \dots, n$ , además supóngase que para cualesquiera  $x_1, \dots, x_n$  las variables aleatorias  $Y_1, \dots, Y_n$  son independientes,  $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ ,  $i = 1, \dots, n$ .

La función de distribución de probabilidad conjunta de las  $Y_i$ 's es:

$$f_n(y|x, \beta_0, \beta_1, \sigma^2) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \right]. \quad (3.1)$$

Esta también es la función de verosimilitud de  $\beta_0$ ,  $\beta_1$  y  $\sigma^2$ . Se utiliza ahora (3.1) para hallar los *estimadores máximo verosímil* (E.M.V.) de  $\beta_0$  y  $\beta_1$  (con independencia de  $\sigma^2$ ), que son los que maximizan  $f_n(y|x, \beta_0, \beta_1, \sigma^2)$ , pero minimiza el error

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2.$$

Pero si se maximiza  $f_n(y|x, \beta_0, \beta_1, \sigma^2)$ , también se maximiza  $\ln f_n(y|x, \beta_0, \beta_1, \sigma^2)$ .

$$l(\beta_0, \beta_1, \sigma^2) = \ln f_n(y|x, \beta_0, \beta_1, \sigma^2),$$

$$l(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2.$$

Calculando la parcial de  $l(\beta_0, \beta_1, \sigma^2)$  con respecto a  $\beta_0$  e igualando a cero se tiene que

$$\begin{aligned} \sum_{i=1}^n y_i - n\beta_0 - \beta_1 \sum_{i=1}^n x_i &= 0, \\ \beta_0 &= \frac{1}{n} \sum_{i=1}^n y_i - \frac{\beta_1}{n} \sum_{i=1}^n x_i, \end{aligned}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}. \tag{3.2}$$

Nuevamente calculando la parcial de  $l(\beta_0, \beta_1, \sigma^2)$  con respecto a  $\beta_1$  e igua-

lando a cero se tiene que

$$\begin{aligned}
\sum_{i=1}^n y_i x_i - \beta_0 \sum_{i=1}^n x_i - \beta_1 \sum_{i=1}^n x_i^2 &= 0, \\
\sum_{i=1}^n y_i x_i - \beta_0 n \bar{x} - \beta_1 \sum_{i=1}^n x_i^2 &= 0, \\
\sum_{i=1}^n y_i x_i - (\bar{y} - \beta_1 \bar{x}) n \bar{x} - \beta_1 \sum_{i=1}^n x_i^2 &= 0, \\
\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x} + n \beta_1 \bar{x}^2 - \beta_1 \sum_{i=1}^n x_i^2 &= 0, \\
\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x} + \beta_1 (n \bar{x}^2 - \sum_{i=1}^n x_i^2) &= 0, \\
\sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \bar{x} + \beta_1 (n \bar{x}^2 - \sum_{i=1}^n x_i^2) &= 0,
\end{aligned}$$

$$\begin{aligned}
\beta_1 &= \frac{\sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \bar{x}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}, \\
\beta_1 &= \frac{\sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \bar{x}}{\sum_{i=1}^n x_i^2 - 2n \bar{x} \bar{x} + n \bar{x}^2}, \\
\beta_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n x_i^2 - 2n \bar{x} \bar{x} + n \bar{x}^2}, \\
\beta_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n x_i \bar{x} + \sum_{i=1}^n \bar{x}^2}.
\end{aligned}$$

Finalmente,

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (3.3)$$

Para que sea más fácil el manejo de los términos, se denota a  $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$ , luego (3.5) se escribe como:

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{S_{xx}}.$$



De las ecuaciones (3.2) y (3.3) se tiene que los E.M.V. para  $\beta_0$  y  $\beta_1$  son:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} \quad (3.4)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{S_{xx}}. \quad (3.5)$$

La diferencia entre el valor observado  $y_i$  y el valor ajustado correspondiente  $\hat{y}_i$  se llama **residual**, el  $i$ -ésimo residual es

$$e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i), \quad i = 1, \dots, n.$$

Como la varianza también es desconocida, se puede estimar y su estimación es,

$$\left. \frac{\partial \ln f_n(y|x, \beta_0, \beta_1, \sigma^2)}{\partial \beta_1} \right|_{\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2} = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2(\hat{\sigma}^2)^2} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 0. \quad (3.6)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2. \quad (3.7)$$

Notesé que se han realizado tres supuestos importantes sobre la distribución conjunta de  $Y_1, \dots, Y_n$ , para cualesquiera valores de  $x_1, \dots, x_n$ .

1. *Normalidad*. Se ha supuesto que cada variable  $Y_i$  tiene una distribución normal.
2. *Independencia*. Se ha supuesto que las variables aleatorias  $Y_1, \dots, Y_n$  son independientes.
3. *Homocedasticidad*. Que las variables aleatorias  $Y_1, \dots, Y_n$  tienen la misma varianza  $\sigma^2$ .

### 3.1.1. Propiedades de los E.M.V.

De la ecuación (3.5) se observa que  $\hat{\beta}_1$  es una función lineal de  $Y_1, \dots, Y_n$ , y ya que éstas variables aleatorias son independientes e idénticamente dis-

tribuidas,  $\hat{\beta}_1$  tendrá una distribución normal, con los siguientes valores esperado y varianza correspondientes.

$$\begin{aligned}
 E[\hat{\beta}_1] &= \frac{\sum_{i=1}^n (x_i - \bar{x}) E[Y_i]}{S_{xx}} \\
 &= \frac{\sum_{i=1}^n (x_i - \bar{x}) (\beta_0 + \beta_1 x_i)}{S_{xx}} \\
 &= \beta_0 \frac{\sum_{i=1}^n (x_i - \bar{x})}{S_{xx}} + \beta_1 \frac{\sum_{i=1}^n (x_i - \bar{x}) x_i}{S_{xx}} \\
 &= \beta_0 \frac{n\bar{x} - n\bar{x}}{S_{xx}} + \beta_1 \frac{\sum_{i=1}^n x_i^2 - \bar{x} x_i}{S_{xx}} \\
 &= \beta_1 \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{S_{xx}} \\
 &= \beta_1 \frac{S_{xx}}{S_{xx}},
 \end{aligned}$$

Por lo que se tiene que,

$$E[\hat{\beta}_1] = \beta_1, \quad (3.8)$$

De aquí se puede observar que  $\hat{\beta}_1$  es un estimador insesgado de  $\beta_1$ . Por otro lado, la varianza de  $\hat{\beta}_1$

$$\begin{aligned}
 \text{Var}(\hat{\beta}_1) &= \text{Var}\left(\frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{S_{xx}}\right) \\
 &= \frac{S_{xx}}{S_{xx} S_{xx}} \sigma^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.
 \end{aligned}$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (3.9)$$

Así se tiene que  $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$ .

Ahora puesto que  $\bar{Y}$  y  $\hat{\beta}_1$  son funciones lineales de  $Y_1, \dots, Y_n$  entonces  $\hat{\beta}_0$  también es una función lineal de  $Y_1, \dots, Y_n$ , por lo que  $\beta_0$  tendrá una

distribución normal.

$$\begin{aligned}
 E[\hat{\beta}_0] &= E[\bar{Y} - \hat{\beta}_1 \bar{x}] \\
 &= E[\bar{Y}] - \bar{x}E[\hat{\beta}_1] \\
 &= E\left[\frac{1}{n} \sum_{i=1}^n y_i\right] - \bar{x}\beta_1 \\
 &= \frac{1}{n}E\left[\sum_{i=1}^n y_i\right] - \bar{x}\beta_1 \\
 &= \frac{1}{n} \sum_{i=1}^n (\beta_0 + \beta_1 x_i) - \bar{x}\beta_1 \\
 &= \beta_0 + \beta_1 \bar{x} - \bar{x}\beta_1 \\
 E[\hat{\beta}_0] &= \beta_0.
 \end{aligned} \tag{3.10}$$

También se observa que  $\hat{\beta}_0$  es un estimador insesgado de  $\beta_0$ .

La varianza de  $\bar{Y}$  es  $\text{Var}(\bar{Y}) = \sigma^2/n$ , se puede demostrar que la covarianza entre  $\bar{Y}$  y  $\hat{\beta}_1$  es cero, así se tiene que

$$\begin{aligned}
 \text{Var}(\hat{\beta}_0) &= \text{Var}(\bar{Y} - \hat{\beta}_1 \bar{x}) \\
 &= \text{Var}(\bar{Y}) + \bar{x}^2 \text{Var}(\hat{\beta}_1) - 2\bar{x} \text{Cov}(\bar{Y}, \hat{\beta}_1) \\
 &= \text{Var}(\bar{Y}) + \bar{x}^2 \text{Var}(\hat{\beta}_1) \\
 \text{Var}(\hat{\beta}_0) &= \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right).
 \end{aligned} \tag{3.11}$$

### Teorema de Gauss-Markov para regresión lineal simple.

Un resultado importante acerca de los E.M.V. de  $\beta_0, \beta_1$ , es el teorema de Gauss-Markov, que establece que para el modelo de regresión  $Y = c_0\beta_0 + \beta_1 X + \varepsilon$  con las hipótesis  $E[\varepsilon] = 0$ ,  $\text{Var}(\varepsilon) = \sigma^2$  y errores no correlacionados, los E.M.V. son insesgados y tienen varianza mínima en comparación con todos los demás estimadores insesgados que sean combinaciones lineales de las  $Y_i$ 's

**Teorema 3.1.1** *Sea  $\theta = c_0\beta_0 + c_1\beta_1 + c_2$ , donde  $c_0, c_1, c_2$  son constantes dadas. Entonces, entre todos los estimadores insesgados de  $\theta$  que son combinaciones lineales de las observaciones  $Y_1, \dots, Y_n$ , el estimador  $\hat{\theta} = c_0\hat{\beta}_0 + c_1\hat{\beta}_1 +$*

$c_2$  tiene la menor varianza para todos los valores posibles de los parámetros  $\beta_0, \beta_1$  y  $\sigma^2$ .

Para la demostración de este teorema ver sección (3.2.2).

### 3.1.2. Contraste de hipótesis.

En esta subsección se obtendrá la distribución conjunta de  $\hat{\beta}_0, \hat{\beta}_1$  y  $\hat{\sigma}^2$ . Para valores  $x_1, \dots, x_n, n \geq 3$ , supóngase que  $Y_1, \dots, Y_n$  son independientes y que  $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2), i = 1, \dots, n$ . Haciendo

$$S_x = \left[ \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}.$$

Sean  $a_1 = (a_{11}, \dots, a_{1n})$  y  $a_2 = (a_{21}, \dots, a_{2n})$  vectores  $n$ -dimensionales definidos de la manera siguiente,

$$\begin{aligned} a_{1j} &= \frac{1}{n^{1/2}}, \\ a_{2j} &= \frac{1}{S_x}, \end{aligned}$$

con  $j = 1, \dots, n$ .

Además nótese que

$$\begin{aligned} \sum_{j=1}^n a_{1j}^2 &= \sum_{i=1}^n \left( \frac{1}{n^{1/2}} \right)^2 = 1 \\ \sum_{j=1}^n a_{2j}^2 &= \frac{(x_1 - \bar{x})^2}{S_x} + \dots + \frac{(x_n - \bar{x})^2}{S_x} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{S_x} = \frac{S_x}{S_x} = 1 \\ \sum_{j=1}^n a_{1j} a_{2j} &= \frac{1}{n^{1/2}} \frac{(x_1 - \bar{x})}{S_x} + \dots + \frac{1}{n^{1/2}} \frac{(x_n - \bar{x})}{S_x} = 0. \end{aligned}$$

Ya que los vectores  $a_1$  y  $a_2$ , cumplen las propiedades mencionadas arriba, se puede construir una matriz ortogonal  $A \in M_{n \times n}$ , tal que  $a_1$  y  $a_2$  constituyen

la primera y segunda fila de  $A$ , respectivamente,

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}.$$

Se define ahora un vector aleatorio  $Z$ , por medio de la relación  $Z = AY$ , donde

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}.$$

La distribución conjunta de  $Z_1, \dots, Z_n$  se puede hallar con el siguiente teorema.

**Teorema 3.1.2** *Supóngase que las variables aleatorias  $Y_1, \dots, Y_n$  son independientes y que cada una tiene una distribución normal con la misma varianza  $\sigma^2$ . Si  $A$  es una matriz ortogonal de  $n \times n$  y  $Z = AY$ , entonces las variables aleatorias  $Z_1, \dots, Z_n$  también son independientes y cada una tiene una distribución normal con varianza  $\sigma^2$ .*

**Demostración.** (Veáse [7]).

En un problema de regresión lineal simple, las observaciones  $Y_1, \dots, Y_n$  cumplen las condiciones del teorema anterior, por lo que las componentes del vector aleatorio  $Z = AY$  serán independientes y cada una tendrá una distribución normal con varianza  $\sigma^2$ .

Las dos primeras componentes  $Z_1$  y  $Z_2$  de  $Z$  son,

$$Z_1 = \sum_{j=1}^n a_{1j} Y_j = \frac{1}{n^{1/2}} \sum_{j=1}^n Y_j = \frac{1}{n^{1/2}} n \bar{Y} = n^{1/2} \bar{Y}. \quad (3.12)$$

De la ecuación  $\hat{\beta}_0 = \bar{Y} - \bar{x} \hat{\beta}_1$ .

$$Z_1 = n^{1/2} (\hat{\beta}_0 + \bar{x} \hat{\beta}_1). \quad (3.13)$$

Por otra parte, como  $\hat{\beta}_1 = \frac{\sum_{j=1}^n (x_j - \bar{x})Y_j}{\sum_{j=1}^n (x_j - \bar{x})^2}$ , la segunda componente de la matriz  $A$  es,

$$\begin{aligned} Z_2 = \sum_{j=1}^n a_{2j}Y_j &= \frac{1}{S_x} \sum_{j=1}^n (x_j - \bar{x})Y_j, \\ &= \frac{1}{S_x} \sum_{j=1}^n (x_j - \bar{x})^2 \hat{\beta}_1 = \frac{S_x^2}{S_x} \hat{\beta}_1, \\ Z_2 &= S_x \hat{\beta}_1. \end{aligned} \quad (3.14)$$

Considérese ahora la variable aleatoria  $S^2$ .

$$S^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2. \quad (3.15)$$

Si se demuestra que  $S^2$  es independiente del vector aleatorio  $(\hat{\beta}_0, \hat{\beta}_1)$  entonces  $\hat{\sigma}^2$  es independiente de  $(\hat{\beta}_0, \hat{\beta}_1)$ .

Como se sabe que

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}.$$

$S^2$  se puede reescribir de la siguiente forma.

$$\begin{aligned} S^2 &= \sum_{i=1}^n [(Y_i - \bar{Y}) - \hat{\beta}_1 (x_i - \bar{x})]^2 \\ &= \sum_{i=1}^n (Y_i - \bar{Y})^2 - 2\hat{\beta}_1 \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}) + \hat{\beta}_1^2 S_x^2 \\ &= \sum_{i=1}^n Y_i^2 - 2\bar{Y} \sum_{i=1}^n Y_i + n\bar{Y}^2 - 2\hat{\beta}_1 \sum_{i=1}^n (x_i - \bar{x})Y_i - 2\hat{\beta}_1 \bar{Y} \sum_{i=1}^n (x_i - \bar{x}) + \hat{\beta}_1^2 S_x^2, \end{aligned}$$

pero como  $\hat{\beta}_1 S_x^2 = \sum_{i=1}^n (x_i - \bar{x})Y_i$ ,

$$S^2 = \sum_i^n Y_i^2 - 2n\bar{Y}^2 + n\bar{Y}^2 - 2\hat{\beta}_1^2 S_x^2 - 2\hat{\beta}_1 \bar{Y} (n\bar{x} - n\bar{x}) + \hat{\beta}_1^2 S_x^2,$$

$$S^2 = \sum_i^n Y_i^2 - n\bar{Y}^2 - \hat{\beta}_1^2 S_x^2. \quad (3.16)$$

Puesto que  $Z = AY$ , donde  $A$  es una matriz ortogonal, se cumple que

$$\sum_{i=1}^n Y_i^2 = Y^t Y = Y^t I Y = Y^t A^t A Y = Z^t Z = \sum_{i=1}^n Z_i^2.$$

Utilizando lo anterior y las ecuaciones (3.12) y (3.13), se tiene lo siguiente,

$$S^2 = \sum_{i=1}^n Z_i^2 - Z_1^2 - Z_2^2 = \sum_{i=3}^n Z_i^2. \quad (3.17)$$

Las variables aleatorias  $Z_1, \dots, Z_n$  son independientes y como  $S^2$  es únicamente igual a la suma de los cuadrados de las variables  $Z_3, \dots, Z_n$ , entonces  $S^2$  y el vector aleatorio,  $(Z_1, Z_2)$  son independientes, más aún como  $\hat{\beta}_0$  y  $\hat{\beta}_1$  son funciones de  $Z_1$  y  $Z_2$  únicamente, se tiene que

$$\begin{aligned} \hat{\beta}_0 &= \frac{Z_1}{n^{1/2}} - \frac{\bar{x}Z_2}{S_x}, \\ \hat{\beta}_1 &= \frac{Z_2}{S_x}, \end{aligned}$$

por tanto,  $S^2$  y  $(\hat{\beta}_0, \hat{\beta}_1)$  son independientes.

Para la distribución de  $S^2$ , se tiene que para  $i = 3, \dots, n$ ,

$$Z_i = \sum_{j=1}^n a_{ij} Y_j.$$

$$\begin{aligned} E(Z_i) &= \sum_{j=1}^n a_{ij} E(Y_j) \\ &= \sum_{j=1}^n a_{ij} (\beta_0 + \beta_1 x_j) \\ &= \sum_{j=1}^n a_{ij} [\beta_0 + \beta_1 \bar{x} + \beta_1 (x_j - \bar{x})], \end{aligned}$$

$$E(Z_i) = (\beta_0 + \beta_1 \bar{x}) \sum_{j=1}^n a_{ij} + \beta_1 \sum_{j=1}^n a_{ij}(x_j - \bar{x}). \quad (3.18)$$

Ya que  $A$  es ortogonal, la suma de los productos de los correspondientes términos en dos filas distintas cualesquiera debe ser cero. En particular para  $i = 3, \dots, n$ .

$$\sum_{j=1}^n a_{ij}a_{1j} = 0,$$

$$\sum_{j=1}^n a_{ij}a_{2j} = 0.$$

Por la forma de  $a_{1j}$  y  $a_{2j}$ , se tiene que,

$$\sum_{j=1}^n a_{ij} = 0, \quad \sum_{j=1}^n a_{ij}(x_j - \bar{x}) = 0, \quad (3.19)$$

sustituyendo lo anterior en (3.17),  $E(Z_i) = 0$  para  $i = 1, \dots, n$ .

De acuerdo a todo lo anterior resulta que las  $n - 2$  variables aleatorias  $Z_3, \dots, Z_n$  son independientes y que tienen una distribución normal con media 0 y varianza  $\sigma^2$ , de acuerdo al teorema anterior.

Ya que  $S^2 = \sum_{i=3}^n Z_i$ , la variable aleatoria  $S^2/\sigma^2$  tendrá una distribución  $\chi^2$  con  $n - 2$  grados de libertad.

Además considérese el E.M.V.  $\hat{\sigma}^2$ , como  $\hat{\sigma}^2 = S^2/n$  luego  $n\hat{\sigma}^2 = S^2$  y de los resultados anteriores se deduce que  $\hat{\sigma}^2$  es independiente de los estimadores  $\hat{\beta}_0$  y  $\hat{\beta}_1$  y que,

$$n\hat{\sigma}^2/\sigma^2 \sim \chi_{n-2}^2. \quad (3.20)$$

### Contraste de hipótesis sobre $\beta_0$

Se desea contrastar

$$H_0 : \beta_0 = \beta_0^*,$$

$$H_1 : \beta_0 \neq \beta_0^*.$$

con  $\beta_0^* \in \mathbb{R}$ .

De las ecuaciones (3.10) y (3.11) cuando la hipótesis  $H_0$  es cierta, la siguiente



variable aleatoria tendrá una distribución normal tipificada.

$$W = \frac{\hat{\beta}_0 - \beta_0}{\sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}}$$

Se construirá un contraste  $t$  para que  $H_0$  se rechace cuando  $\beta_0$  se encuentre lejos del valor  $\beta_0^*$  y se acepte  $H_0$  en otro caso. Al ser el valor de  $\sigma$  desconocido no es posible basarse solo en  $W$ . Como  $\hat{\beta}_0$  y  $S^2$  son variables aleatorias independientes, resulta que  $W$  y  $S^2$  también son independientes, entonces cuando  $H_0$  es verdadera se tiene,

$$T = \frac{\frac{\hat{\beta}_0 - \beta_0}{\sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}}}{\sqrt{\frac{n\hat{\sigma}^2}{\sigma^2} / (n-2)}} = \sqrt{\frac{n-2}{n}} \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} \sigma} \sim t_{n-2}$$

De acuerdo al estadístico anterior, se rechaza  $H_0$  si,  $|T| > c_1$  donde  $c_1$  es una constante apropiada cuyo valor se elige para cualquier valor de significancia  $\alpha_0 \in (0, 1)$ .

### Contraste de hipótesis sobre $\beta_1$

Ahora para hacer un contraste con respecto a  $\beta_1$ , con  $\beta_1^* \in \mathbb{R}$

$$\begin{aligned} H_0 : \beta_1 &= \beta_1^*, \\ H_1 : \beta_1 &\neq \beta_1^*. \end{aligned}$$

De la ecuaciones (3.8) y (3.9) cuando  $H_0$  es cierta la variable aleatoria,

$$V = \frac{\hat{\beta}_1 - \beta_1}{\frac{\sigma}{\sqrt{S_{xx}}}} \sim N(0, 1),$$

es tal que el estadístico a utilizar para este contraste es

$$U = \frac{\frac{\hat{\beta}_1 - \beta_1}{\frac{\sigma}{\sqrt{S_{xx}}}}}{\sqrt{\frac{n\hat{\sigma}^2}{\sigma^2} / (n-2)}} = \sqrt{\frac{n-2}{n}} \frac{\hat{\beta}_1 - \beta_1}{\sqrt{S_{xx}} \hat{\sigma}} \sim t_{n-2}.$$

Entonces la hipótesis  $H_0$  se debería rechazar si  $|U| > c_2$  con  $c_2$  una constante apropiada cuyo valor se puede elegir para cualquier valor de significancia  $0 < \alpha_0 < 1$ .

## 3.2. Regresión lineal múltiple.

Supóngase que se tienen  $n$  vectores de observaciones. Para  $i = 1, \dots, n$ , sea  $(x_{i1}, x_{i2}, \dots, x_{ik})$ , el vector de valores observados, ahora lo que se desea es encontrar una función que se ajuste lo mejor posible a estos datos, y esto se realiza mediante un modelo regresión lineal.

Un modelo de regresión lineal en donde intervienen más de una variable regresora se llama *modelo de regresión múltiple*, si el modelo es:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon,$$

entonces se conoce como *modelo de regresión múltiple*, con  $k$  regresores, los parámetros  $\beta_j, j = 0, 1, \dots, k$  se conocen como *coeficientes de regresión*.

Este modelo describe un hiperplano en el espacio de  $k$  dimensiones de las variables regresoras  $X_j$ . Entonces si,  $E[\varepsilon] = 0$ ,  $Var(\varepsilon) = \sigma^2$  y los errores no están correlacionados y si se supone que la función de regresión es una función lineal, entonces

$$E[Y|X_1, \dots, X_k] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k.$$

El parámetro  $\beta_j$  representa el cambio esperado en la respuesta  $Y$  por cambio en la variable regresora  $X_j$ , cuando todas las demás  $X_i (i \neq j)$  permanecen constantes.

### 3.2.1. Estimación de los coeficientes de regresión

#### Estimación de $\beta$ por mínimos cuadrados.

Supóngase que se tienen  $n > k$  observaciones y sea  $y_i$ , la  $i$ -ésima respuesta observada y sea  $x_{ij}$  el  $i$ -ésimo valor de la variable regresora  $X_j$ .

$$\begin{array}{cccccc} i & y & x_1 & \cdots & x_k & \\ 1 & y_1 & x_{11} & \cdots & x_{1k} & \\ 2 & y_2 & x_{21} & \cdots & x_{2k} & \\ \vdots & \vdots & \vdots & \cdots & \vdots & \\ n & y_n & x_{n1} & \cdots & x_{nk} & \end{array}$$

Entonces la  $i$ -ésima observación,

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, \quad i = 1, \dots, n,$$

y la suma de los cuadrados de los errores,

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2.$$

Una forma más sencilla de realizar la estimación por mínimos cuadrados es escribiéndolo en forma matricial, por lo que haciendo

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}, \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}_{n \times r}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{r \times 1}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1},$$

con  $r = k + 1$  el número de parámetros a determinar.

Entonces

$$Y = X\beta + \varepsilon, \quad \varepsilon = Y - X\beta,$$

$$\begin{aligned} S = \varepsilon^t \varepsilon &= (Y - X\beta)^t (Y - X\beta) \\ &= Y^t Y - 2\beta^t X^t Y + \beta^t X^t X \beta. \end{aligned}$$

Si se calcula  $\frac{\partial S}{\partial \beta}$ ,

$$\left. \frac{\partial S}{\partial \beta} \right|_{\hat{\beta}} = -2X^t Y + 2X^t X \hat{\beta}.$$

Si  $X^t X$  es no singular,

$$\hat{\beta} = (X^t X)^{-1} X^t Y.$$

Si  $X^t X$  fuera singular, para resolver el sistema se tendría que calcular la matriz pseudoinversa, lo cual no se desarrollará en esta tesis. En lo sucesivo, se supone que  $X^t X$  es no singular (a menos que se especifique lo contrario).

**Estimación por máxima verosimilitud de  $\sigma^2$** 

El modelo es  $Y = X\beta + \varepsilon$  y los errores se distribuyen en forma normal e independiente con media cero y varianza constante igual a  $\sigma^2$ , es decir,  $\varepsilon \sim N(0, \sigma^2 I)$  ( $I$  matriz identidad de  $n \times n$ ), así

$$f(\varepsilon_i) = \frac{1}{\sqrt{2\sigma}} \exp \left\{ -\frac{\varepsilon_i^2}{2\sigma^2} \right\}.$$

La función de probabilidad conjunta es,

$$f_n(\varepsilon, \beta, \sigma^2) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{\varepsilon^t \varepsilon}{2\sigma^2} \right\}$$

$$f_n(Y, X, \beta, \sigma^2) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{(Y - X\beta)^t (Y - X\beta)}{2\sigma^2} \right\}.$$

Luego, aplicando  $\ln$  a la función anterior y derivando con respecto de  $\sigma^2$ ,

$$\ln f_n(Y, X, \beta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \frac{(Y - X\beta)^t (Y - X\beta)}{2\sigma^2},$$

para un valor fijo de  $\sigma$ , el logaritmo se maximiza, al minimizar,  $(Y - X\beta)^t (Y - X\beta)$ .

$$\frac{\partial \ln f_n(Y, X, \beta, \sigma^2)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{(Y - X\beta)^t (Y - X\beta)}{\sigma^3} = 0.$$

Por tanto el estimador por máxima verosimilitud de  $\sigma^2$  es,

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})^t (Y - X\hat{\beta})}{n}. \quad (3.21)$$

**Estimador insesgado de  $\sigma^2$** 

Se puede estimar  $\sigma^2$ , y encontrar un estimador insesgado de éste, de acuerdo a la suma de cuadrados de los residuales,

$SC_{Res} = e^t e = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ , como  $e = Y - X\hat{\beta}$ , se tiene,

$$\begin{aligned} e^t e &= (Y - X\hat{\beta})^t (Y - X\hat{\beta}) \\ &= Y^t Y - \hat{\beta}^t X^t Y - Y^t X \hat{\beta} + \hat{\beta}^t X^t X \hat{\beta} \\ &= Y^t Y - 2\hat{\beta}^t X^t Y + \hat{\beta}^t X^t X \hat{\beta}. \end{aligned}$$

Como  $X^t X \hat{\beta} = X^t Y$ ,

$$SC_{Res} = Y^t Y - \hat{\beta}^t X^t Y.$$

En seguida se muestra que la suma de cuadrados de los residuales tiene  $n - r$  grados de libertad asociados con ella, porque se estiman  $r$  parámetros en el modelo de regresión.  $SC_{Res}$  se puede también escribir de la siguiente forma

$$\begin{aligned} SC_{Res} &= [Y - X(X^t X)^{-1} X^t Y]^t [Y - X(X^t X)^{-1} X^t Y] \\ &= Y^t [I - X(X^t X)^{-1} X^t] Y, \end{aligned}$$

la matriz  $I - X(X^t X)^{-1} X^t$  es idempotente pues  $I - X(X^t X)^{-1} X^t = [I - X(X^t X)^{-1} X^t]^t [I - X(X^t X)^{-1} X^t]$  y simétrica. Además como  $I \in M_{n \times n}$  y  $X \in M_{n \times r}$  se cumple que

$$\begin{aligned} \text{tr}[I - X(X^t X)^{-1} X^t] &= \text{tr}[I_{n \times n}] - \text{tr}[X^t X (X^t X)^{-1}] \\ &= n - \text{tr}[I_{r \times r}] = n - r. \end{aligned}$$

Por el teorema (A.2.1),

$$\text{ran}[I - X(X^t X)^{-1} X^t] = \text{tr}[I - X(X^t X)^{-1} X^t] = n - r.$$

Aplicando el teorema (A.4.2),

$$\frac{SC_{Res}}{\sigma^2} = \frac{1}{\sigma^2} Y^t [I - X(X^t X)^{-1} X^t] Y \sim \chi_{n-r, \lambda}^2.$$

(Ver punto 1 de A.1.)

El parámetro de no centralidad es,

$$\lambda = \frac{1}{\sigma^2} E[Y]^t [I - X(X^t X)^{-1} X^t] E[Y],$$

como  $E[Y] = X\beta$ , entonces

$$\begin{aligned} \lambda &= \frac{1}{\sigma^2} \beta^t X^t [I - X(X^t X)^{-1} X^t] X \beta \\ &= \frac{1}{\sigma^2} \beta^t [X^t X - X^t X (X^t X)^{-1} X^t X] \beta = 0, \end{aligned}$$

por tanto,

$$\frac{SC_{Res}}{\sigma^2} \sim \chi_{n-r}^2.$$

Aplicando el teorema (A.4.1) el valor esperado de  $SC_{Res}$ ,

$$\begin{aligned} E[SC_{Res}] &= E[Y^t[I - X(X^tX)^{-1}X^t]Y] \\ &= \text{tr}([I - X(X^tX)^{-1}X^t]\sigma^2I) + E[Y^t][I - X(X^tX)^{-1}X^t]E[Y] \\ &= (n - r)\sigma^2. \end{aligned}$$

Haciendo la media de cuadrados residual  $MSC_{Res} = \frac{SC_{Res}}{n-r}$ , se tiene que,

$$E[MSC_{Res}] = E\left[\frac{SC_{Res}}{n-r}\right] = \sigma^2.$$

De la ecuación anterior se tiene que  $MSC_{Res}$  es un estimador insesgado de  $\sigma^2$ .

$$\hat{\sigma}^2 = MSC_{res} = \frac{(Y - X\hat{\beta})^t(Y - X\hat{\beta})}{n - r}.$$

### 3.2.2. Propiedades de los estimadores.

Supóngase que  $Y$  es un vector aleatorio  $n$ -dimensional  $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ .

**Definición 3.2.1** La esperanza  $E[Y]$  del vector aleatorio  $Y$ , se define como el vector  $n$ -dimensional cuyas componentes son las esperanzas de las componentes individuales de  $Y$ . Por tanto

$$E[Y] = \begin{bmatrix} E(Y_1) \\ \vdots \\ E(Y_n) \end{bmatrix}, \quad (3.22)$$

y se denomina vector de medias de  $Y$ .

**Definición 3.2.2** La **matriz de covarianzas** del vector aleatorio  $Y$  se define como la matriz  $n \times n$  tal que para  $i = 1, \dots, n$  y  $j = 1, \dots, n$  el elemento de la  $i$ -ésima fila y  $j$ -ésima columna sea  $\text{Cov}(Y_i, Y_j)$ . Si  $\text{Cov}(Y_i, Y_j) = \sigma_{ij}$ , entonces

$$\text{Cov}(Y) = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1n} \\ \vdots & \dots & \vdots \\ \sigma_{n1} & \dots & \sigma_{nn} \end{bmatrix}. \quad (3.23)$$

De la definición de varianza y covarianza se tiene que para  $i = 1, \dots, n$ ,  $\text{Var}(Y_i) = \text{Cov}(Y_i, Y_i) = \sigma_{ii}$ , luego los elementos de la diagonal de la matriz  $\text{Cov}(Y)$  son las varianzas de  $Y_1, \dots, Y_n$ .

También como  $\text{Cov}(Y_i, Y_j) = \text{Cov}(Y_j, Y_i)$ , entonces  $\sigma_{ij} = \sigma_{ji}$ , y así la matriz  $\text{Cov}(Y)$  es simétrica.

De acuerdo a lo anterior, como  $E(\varepsilon) = 0$ ,  $\text{Var}(\varepsilon) = \sigma^2$  y  $Y = X\beta + \varepsilon$ ,

$$\begin{aligned} E[Y] &= E[X\beta + \varepsilon] = X\beta + E[\varepsilon], \\ E[Y] &= X\beta. \end{aligned} \quad (3.24)$$

Se sabe que las componentes  $Y_1, \dots, Y_n$  de  $Y$  son independientes y la varianza de cada una es  $\sigma^2$ , por lo tanto

$$\text{Cov}(Y) = \sigma^2 I.$$

Ahora para encontrar, las esperanzas y covarianzas de los estimadores, se enuncia el siguiente teorema.

**Teorema 3.2.1** Supóngase que  $Y$  es un vector aleatorio  $n$ -dimensional, para el cual existe el vector de medias  $E[Y]$  y la matriz de covarianzas  $\text{Cov}(Y)$ . Supóngase también que  $A$  es una matriz de tamaño  $p \times n$  cuyos elementos son constantes y que  $W$  es un vector aleatorio  $n$ -dimensional definido por la relación  $W = AY$ . Entonces  $E[W] = AE[Y]$  y  $\text{Cov}(W) = A\text{Cov}(Y)A^t$

**Demostración.**

Sea

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{n1} & \dots & a_{pn} \end{bmatrix},$$

la  $i$ -ésima componente de  $W$  es

$$W_i = \sum_{j=1}^n a_{ij} Y_j,$$

y la  $i$ -ésima componente de  $E(W)$  es

$$E[W_i] = E\left[\sum_{j=1}^n a_{ij} Y_j\right] = \sum_{j=1}^n a_{ij} E(Y_j),$$

pero esta es la  $i$ -ésima componente del vector  $AE(Y)$ , por tanto  $E[W] = AE[Y]$ .

La matriz de covarianzas de  $W$  es una matriz de tamaño  $p \times p$ . Para  $h = 1, \dots, p$  y  $k = 1, \dots, p$

$$\begin{aligned} \text{Cov}(W_h, W_k) &= \text{Cov}\left(\sum_{r=1}^p a_{hr} Y_r, \sum_{s=1}^p a_{ks} Y_s\right) \\ &= \sum_{r=1}^p \sum_{s=1}^p a_{hr} a_{ks} \text{Cov}(Y_r, Y_s). \end{aligned}$$

Pero este elemento es la  $r$ -ésima fila y la  $s$ -ésima columna de matriz  $ACov(Y)A^t$ , entonces

$$\text{Cov}(W) = ACov(Y)A^t.$$

□

Utilizando el teorema anterior, el valor esperado de  $\hat{\beta}$  es,

$$\begin{aligned} E[\hat{\beta}] &= E[(X^t X)^{-1} Y] \\ &= E[(X^t X)^{-1} X^t (X\beta + \varepsilon)] \\ &= E[(X^t X)^{-1} X^t X\beta + (X^t X)^{-1} X^t \varepsilon] \\ &= E[\beta] + (X^t X)^{-1} X^t E[\varepsilon] \\ &= \beta, \end{aligned}$$

el estimador  $\hat{\beta}$  es insesgado, luego

$$E[\hat{\beta}_j] = \beta_j,$$



para  $j = 0, 1, \dots, k$ .

Nuevamente utilizando el teorema anterior, como

$$\hat{\beta} = (X^t X)^{-1} X^t Y,$$

y  $(X^t X)^{-1} X^t$  es una matriz de constantes la covarianza de  $\hat{\beta}$  es la siguiente,

$$\begin{aligned} \text{Cov}(\hat{\beta}) &= [(X^t X)^{-1} X^t] \text{Cov}(Y) [(X^t X)^{-1} X^t]^t \\ &= (X^t X)^{-1} X^t (\sigma^2 I) X (X^t X)^{-1} \\ &= (\sigma^2 I) (X^t X)^{-1} X^t X (X^t X)^{-1} \\ &= \sigma^2 (X^t X)^{-1}, \end{aligned}$$

$$\text{Cov}(\hat{\beta}) = \sigma^2 (X^t X)^{-1}. \quad (3.25)$$

Si  $C = (X^t X)^{-1}$ , entonces para  $j = 1, \dots, n$ ,  $\text{Var}(\hat{\beta}_j)$  será igual a,

$$\text{Var}(\hat{\beta}_j) = \sigma^2 c_{jj}.$$

Para  $i \neq j$ ,  $\text{Cov}(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 c_{ij}$ .

### Teorema de Gauss-Markov para el caso general

Algo que interesa mucho, es garantizar, que nuestra estimación de  $\beta$  tenga una mínima varianza o error de estimación, esta garantía nos la da el teorema de Gauss-Markov.

Supóngase que las observaciones  $Y_1, \dots, Y_n$  no están correlacionadas, que  $E[Y_i] = \beta_0 + \beta_1 x_{i1}, \dots, \beta_k x_{ik}$ , y que  $\text{Var}(Y_i) = \sigma^2$  para  $i = 1, \dots, n$ .

#### Teorema 3.2.2 De Gauss-Markov

Si  $E(Y) = X\beta$  y  $\text{Cov}(Y) = \sigma^2 I$ , entre todos los estimadores insesgados de  $\beta$  que son combinaciones lineales de las observaciones  $Y_1, \dots, Y_n$ , el estimador  $\hat{\beta}$  tiene la menor varianza para todos los valores posibles  $\beta_0, \beta_1, \dots, \beta_k$  y  $\sigma^2$ .

#### Demostración.

Considérese una estimación lineal  $AY$  de  $\beta$ , con  $A$  de tal forma que  $AY$  es un estimador insesgado de mínima varianza de  $\beta$ . Como  $AY$  es un estimador insesgado, se cumple que  $E[AY] = \beta$ , ya que  $E[Y] = X\beta$ , se tiene que,

$$E[AY] = AE[Y] = AX\beta = \beta,$$

lo cual da la condición de insesgades,

$$AX = I, \quad (3.26)$$

ya que la relación  $AX\beta = \beta$ , se tiene para todos los posible valores de  $\beta$  (Veáse [15]).

La matriz de covarianza para el estimador  $AY$  está dada por,

$$\text{Cov}(AY) = A(\sigma^2 I)A^t = \sigma^2 AA^t.$$

Las varianzas de los  $\hat{\beta}_j$  se encuentran en la diagonal de  $\sigma^2 AA^t$ , lo que sigue es determinar quien es la matriz  $A$  que cumple (3.26) y que los elementos de la diagonal de  $AA^t$  sean mínimos. Para relacionar a  $AY$  con  $\hat{\beta} = (X^t X)^{-1} X^t Y$ , se suma y resta  $(X^t X)^{-1} X^t$  en la siguiente expresión,

$$\begin{aligned} AA^t &= [A - (X^t X)^{-1} X^t + (X^t X)^{-1} X^t][A - (X^t X)^{-1} X^t + (X^t X)^{-1} X^t]^t \\ &= [A - (X^t X)^{-1} X^t][A - (X^t X)^{-1} X^t]^t + [A - (X^t X)^{-1} X^t][X(X^t X)^{-1}] + \\ &\quad (X^t X)^{-1} X^t[A - (X^t X)^{-1} X^t]^t + (X^t X)^{-1} X^t X(X^t X)^{-1} \\ &= [A - (X^t X)^{-1} X^t][A - (X^t X)^{-1} X^t]^t AX(X^t X)^{-1} - (X^t X)^{-1} X^t X(X^t X)^{-1} \\ &\quad + (X^t X)^{-1} X^t A^t - (X^t X)^{-1} X^t X(X^t X)^{-1} + (X^t X)^{-1} \\ &= [A - (X^t X)^{-1} X^t][A - (X^t X)^{-1} X^t]^t - 2(X^t X)^{-1} + (X^t X)^{-1} \\ &= [A - (X^t X)^{-1} X^t][A - (X^t X)^{-1} X^t]^t - (X^t X)^{-1}. \end{aligned}$$

La matriz  $[A - (X^t X)^{-1} X^t][A - (X^t X)^{-1} X^t]^t$  es semidefinida positiva (Ver teorema A.2.2), y por el teorema (A.2.3) los elementos de la diagonal son mayores o iguales que cero.

Si se escoge  $A = (X^t X)^{-1} X^t$ , los elementos de la diagonal pueden ser cero, además, este valor de  $A$  satisface la condición de insesgades.

El estimador insesgado de mínima varianza de  $\beta$  es,

$$AY = (X^t X)^{-1} X^t Y,$$

el cual resulta que es el mismo estimador  $\hat{\beta}$ .

□

### 3.2.3. Contráste de hipótesis

Supóngase que se desea contrastar la hipótesis que uno de los coeficientes de la regresión  $\beta_j$  tiene un valor particular  $\beta_j^*$ , es decir.

$$\begin{aligned} H_0 &: \beta_j = \beta_j^*, \\ H_1 &: \beta_j \neq \beta_j^*. \end{aligned}$$

Ya que  $\text{Var}(\hat{\beta}_j) = c_{jj}\sigma^2$ , cuando  $H_0$  es verdadera, la siguiente variable aleatoria  $W_j$  tiene una distribución normal tipificada,

$$W_j = \frac{\hat{\beta}_j - \beta_j}{c_{jj}^{1/2} \sigma}.$$

Como la variable aleatoria  $SC_{Res}/\sigma^2$  tiene una distribución  $\chi^2$  con  $n - r$  grados de libertad, al ser  $SC_{Res}$  independiente de  $\hat{\beta}_j$ , cuando  $H_0$  es cierta, la variable aleatoria  $T_j$  tiene una distribución,

$$T_j = \frac{W_j}{\left[ \frac{1}{n-r} \left( \frac{(SC_{Res})^2}{\sigma^2} \right) \right]^{1/2}} \sim t_{n-r},$$

$$T_j = \left( \frac{n-r}{c_{jj}} \right)^{1/2} \frac{\hat{\beta}_j - \beta_j}{[(Y - X\hat{\beta})^t(Y - X\hat{\beta})]}.$$

$H_0$  se rechaza si  $|T_j| > c$ , donde  $c$  es una constante adecuada cuyo valor se elige para un nivel de significancia  $0 < \alpha_0 < 1$ .

## 3.3. Análisis bayesiano del modelo de regresión lineal.

El modelo bayesiano de regresión lineal múltiple es similar al modelo clásico de regresión lineal múltiple excepto en que en el primero se incluyen especificaciones sobre la distribución a priori de los parámetros.

### 3.3.1. Distribución a posteriori con información a priori no informativa

Se tiene el modelo

$$Y = X\beta + \varepsilon,$$

en donde  $\varepsilon_1, \dots, \varepsilon_n$  son independientes, es decir  $\varepsilon_i \sim N(0, \sigma^2)$ ,  $i = 1, \dots, n$  entonces  $\varepsilon \sim N_n(0, \sigma^2 I)$ .

Así la función de verosimilitud

$$l(\beta, \sigma^2 | Y) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left[ -\frac{1}{2\sigma^2} (Y - X\beta)^t (Y - X\beta) \right]. \quad (3.27)$$

Pero dado que  $\hat{\beta} = (X^t X)^{-1} X^t Y$  y  $\hat{Y} = X\hat{\beta}$ ,

$$\begin{aligned} (Y - X\beta)^t (Y - X\beta) &= (Y - X\beta)^t (Y - X\beta) - 2Y^t \hat{Y} + 2Y^t \hat{Y} \\ &= YY^t - Y^t X\beta - \beta^t X^t Y + \beta^t X^t X\beta - 2Y^t X\hat{\beta} + 2Y^t X\hat{\beta} \\ &= YY^t - 2\beta^t X^t Y + \beta^t X^t X\beta - 2Y^t X\hat{\beta} + 2Y^t X\hat{\beta} \\ &= YY^t + \beta^t X^t X\beta - 2\beta^t X^t X(X^t X)^{-1} X^t Y \\ &\quad - 2Y^t X\hat{\beta} + 2Y^t X(X^t X)^{-1} X^t X\hat{\beta} \\ &= YY^t + \beta^t X^t X\beta - 2\beta^t X^t X\hat{\beta} - 2Y^t X\hat{\beta} + 2\hat{\beta}^t X^t X\hat{\beta} \\ &= YY^t + \beta^t X^t X\beta - \beta^t X^t X\hat{\beta} - \hat{\beta}^t X^t X\beta - Y^t X\hat{\beta} \\ &\quad - \hat{\beta}^t X^t Y + \hat{\beta}^t X^t X\hat{\beta} + \hat{\beta}^t X^t X\hat{\beta} \\ &= YY^t - Y^t X\hat{\beta} - \hat{\beta}^t X^t Y + \hat{\beta}^t X^t X\hat{\beta} + \beta^t X^t X\beta \\ &\quad - \beta^t X^t X\hat{\beta} - \hat{\beta}^t X^t X\beta + \hat{\beta}^t X^t X\hat{\beta} \\ &= (Y - X\hat{\beta})^t (Y - X\hat{\beta}) + (\beta - \hat{\beta})^t X^t X(\beta - \hat{\beta}) \\ &= (n - r) \left( \frac{1}{n - r} \right) (Y - \hat{Y})^t (Y - \hat{Y}) + (\beta - \hat{\beta})^t X^t X(\beta - \hat{\beta}), \end{aligned}$$

$$(Y - X\beta)^t (Y - X\beta) = \nu S^2 + (\beta - \hat{\beta})^t X^t X(\beta - \hat{\beta}) \quad (3.28)$$

con

$$S^2 = \frac{(Y - \hat{Y})^t (Y - \hat{Y})}{\nu}, \quad \nu = n - r.$$

Pero de (3.27), (3.28) puede expresarse como,

$$l(\beta, \sigma^2|Y) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} [\nu S^2 + (\beta - \hat{\beta})^t X^t X (\beta - \hat{\beta})] \right\}. \quad (3.29)$$

Como  $\beta$  es un parámetro de posición y  $\sigma^2$  un parámetro de escala, las distribuciones a priori no informativas son,

$$\pi(\beta) \propto 1 \quad \text{y} \quad \pi(\sigma^2) \propto \frac{1}{\sigma^2},$$

y asumiendo independencia entre  $\beta$  y  $\sigma^2$ ,

$$\pi(\beta, \sigma^2) = \pi(\beta)\pi(\sigma^2) \propto \frac{1}{\sigma^2}.$$

Ahora, por el teorema de Bayes, la distribución a posteriori  $\pi(\beta, \sigma^2|Y)$  es,

$$\pi(\beta, \sigma^2|Y) \propto \pi(\beta, \sigma^2)l(\beta, \sigma^2|Y).$$

De aquí se tiene que,

$$\pi(\beta, \sigma^2|Y) = \frac{1}{\sigma^{n+2}} \exp \left\{ -\frac{1}{2\sigma^2} [\nu S^2 + (\beta - \hat{\beta})^t X^t X (\beta - \hat{\beta})] \right\}. \quad (3.30)$$

Por la ley multiplicativa de la probabilidad, la función anterior puede representarse por el siguiente producto,

$$\pi(\beta, \sigma^2|Y) \propto \pi(\sigma^2|Y)\pi(\beta|Y, \sigma^2), \quad (3.31)$$

utilizando (A.1) se tiene la distribución marginal a posteriori para  $\sigma^2$ ,

$$\begin{aligned}
\pi(\sigma^2|Y) &= \int_{-\infty}^{\infty} \pi(\beta, \sigma^2|Y) d\beta \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma^{n+2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \nu S^2 + (\beta - \hat{\beta})^t X^t X (\beta - \hat{\beta}) \right] \right\} d\beta \\
&= \frac{1}{\sigma^{n+2}} \exp \left\{ -\frac{\nu S^2}{2\sigma^2} \right\} \int_{-\infty}^{\infty} \exp \left[ -\frac{1}{2} (\beta - \hat{\beta})^t \frac{X^t X}{\sigma^2} (\beta - \hat{\beta}) \right] d\beta \\
&= \frac{1}{\sigma^{n+2}} \exp \left\{ -\frac{\nu S^2}{2\sigma^2} \right\} (\sqrt{2\pi})^r \left| \frac{(X^t X)^{-1}}{\sigma^2} \right|^{1/2} \\
&\propto \frac{1}{\sigma^{n+2}} \exp \left\{ -\frac{\nu S^2}{2\sigma^2} \right\} (\sqrt{2\pi})^r \frac{1}{[(\sigma^2)^{2r}]^{-1/2}} \\
&= \frac{1}{\sigma^{n+2-2r}} \exp \left\{ -\frac{\nu S^2}{2\sigma^2} \right\} \\
&= \frac{1}{(\sigma^2)^{\frac{n}{2}+1-\frac{r}{2}}} \exp \left\{ -\frac{\nu S^2}{2\sigma^2} \right\} \\
&= \frac{1}{(\sigma^2)^{\frac{n-r}{2}+1}} \exp \left\{ -\frac{\nu S^2}{2\sigma^2} \right\} \\
&= \frac{1}{(\sigma^2)^{\nu/2+1}} \exp \left\{ -\frac{\nu S^2}{2\sigma^2} \right\},
\end{aligned}$$

lo anterior implica que la distribución marginal a posteriori para  $\sigma^2$  es el kernel de una distribución gamma inversa, por tanto,

$$\pi(\sigma^2|Y) = \frac{\nu S^2}{2\Gamma(\frac{\nu}{2})} \left( \frac{1}{\sigma^2} \right)^{\nu/2+1} \exp \left\{ -\frac{\nu S^2}{2\sigma^2} \right\}, \quad (3.32)$$

es decir,

$$\pi(\sigma^2|Y) = \text{GInv} \left( \frac{\nu}{2}, \frac{\nu S^2}{2} \right). \quad (3.33)$$

(Ver apéndice sección A.1.)

Por otra parte la distribución condicional a posteriori mencionada en (3.31)

está dada por,

$$\begin{aligned}\pi(\beta|\sigma^2, Y) &= \frac{\pi(\beta, \sigma^2|Y)}{\pi(\sigma^2|Y)} \\ &\propto \frac{\frac{1}{\sigma^{n+2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \nu S^2 + (\beta - \hat{\beta})^t X^t X (\beta - \hat{\beta}) \right] \right\}}{\exp \left\{ -\frac{\nu S^2}{2\sigma^2} \right\} (\sigma^2)^{-(\nu/2)-1}} \\ &= \sigma^{-r} \exp \left\{ -\frac{1}{2\sigma^2} (\beta - \hat{\beta})^t X^t X (\beta - \hat{\beta}) \right\}.\end{aligned}$$

Así,

$$\pi(\beta|\sigma^2, Y) = N_r(\hat{\beta}, (X^t X)^{-1} \sigma^2). \quad (3.34)$$

Integrando sobre  $\sigma^2$  la distribución a posteriori conjunta dada en (3.30) y utilizando (A.2) se obtiene la distribución a posteriori marginal de  $\beta$  que es,

$$\pi(\beta|Y) = \int_0^\infty \frac{1}{\sigma^{n+2}} \exp \left\{ -\frac{\nu S^2 + (\beta - \hat{\beta})^t X^t X (\beta - \hat{\beta})}{2\sigma^2} \right\} d\sigma^2, \quad (3.35)$$

Haciendo  $a = \frac{\nu S^2 + (\beta - \hat{\beta})^t X^t X (\beta - \hat{\beta})}{2}$ , sustituyendo en (3.35) y utilizando (A.2)

$$\begin{aligned}\pi(\beta|Y) &= \int_0^\infty (\sigma^2)^{-(\frac{n}{2}+1)} \exp \left\{ -\frac{a}{\sigma^2} \right\} d\sigma^2, \\ &= \left( \frac{a}{2} \right)^{-\frac{n}{2}} \Gamma \left( \frac{n}{2} \right) \\ &= \left[ \frac{\nu S^2 + (\beta - \hat{\beta})^t X^t X (\beta - \hat{\beta})}{2} \right]^{\frac{n}{2}} \Gamma \left( \frac{n}{2} \right) \\ &= \Gamma \left( \frac{\nu + r}{2} \right) 2^{\frac{n}{2}} \left\{ \nu S^2 \left[ 1 + \frac{(\beta - \hat{\beta})^t X^t X (\beta - \hat{\beta})}{\nu S^2} \right] \right\}^{-\frac{1}{2}(\nu+r)} \\ &= \Gamma \left( \frac{\nu + r}{2} \right) (\nu S^2)^{-\frac{1}{2}(\nu+r)} \left[ 1 + \frac{(\beta - \hat{\beta})^t X^t X (\beta - \hat{\beta})}{\nu S^2} \right]^{-\frac{1}{2}(\nu+r)} \\ &\propto \Gamma \left[ \frac{1}{2}(\nu + r) \right] \nu^{-\frac{r}{2}} S^{-r} \left[ 1 + \frac{(\beta - \hat{\beta})^t X^t X (\beta - \hat{\beta})}{\nu S^2} \right]^{-\frac{1}{2}(\nu+r)}\end{aligned}$$

la expresión arriba mencionada es el kernel de una t-Student multivariada, por lo tanto,

$$\pi(\beta|Y) = \frac{\Gamma\left[\frac{1}{2}(\nu+r)\right] |X^t X|^{\frac{1}{2}} S^{-r}}{\left[\Gamma\left(\frac{1}{2}\right)\right]^r \Gamma\left(\frac{1}{2}\nu\right) (\sqrt{\nu})^r} \left[1 + \frac{(\beta - \hat{\beta})X^t X(\beta - \hat{\beta})}{\nu S^2}\right]^{-\frac{1}{2}(\nu+r)} \quad (3.36)$$

y se denota por  $t_r[\hat{\beta}, S^2(X^t X)^{-1}, \nu]$ .

De aquí se tiene que cada parámetro  $\beta_i$ ,  $i = 0, 1, \dots, r$  tiene la siguiente distribución,

$$\pi(\beta_i|Y) = t(\nu, \hat{\beta}_i, h_{ii}),$$

lo cual, es una distribución *t-Student* univariada con  $\nu$  grados de libertad, parámetro de posición  $\hat{\beta}$  y parámetro de escala  $h_{ii}$  que es el elemento  $(i, i)$  de  $S^2(X^t X)^{-1}$ .

### 3.3.2. Distribución a posteriori con información a priori conjugada

Para encontrar la distribución a posteriori, en este caso, los parámetros  $\beta$  y  $\sigma^2$  tienen cada uno una distribución a priori conocida.

Las distribuciones a posteriori utilizadas para  $\beta$  (3.36) y  $\sigma^2$  (3.33) en la sección anterior serán utilizadas ahora como distribuciones a priori para cada parámetro, puesto que la distribución normal-gamma-inversa es conjugada con la verosimilitud (3.29), ya que de (3.34) y (3.42) se tiene

$$\begin{aligned} NGInv(\beta, \sigma^2|Y) &\propto \exp\left\{\frac{1}{2}(\beta - \hat{\beta})^t \frac{(X^t X)}{\sigma^2} (\beta - \hat{\beta})\right\} \left(\frac{1}{\sigma^2}\right)^{\nu/2+1} \exp\left\{-\frac{\nu S^2}{2\sigma^2}\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}[\nu S^2 + (\beta - \hat{\beta})^t (X^t X) (\beta - \hat{\beta})]\right\} \\ &\propto l(\beta, \sigma^2|Y), \end{aligned}$$

correspondiente al vector de datos  $Y$ . Las distribuciones a posteriori  $\pi(\sigma^2|Y)$  y  $\pi(\beta|\sigma^2, Y)$  calculadas de la muestra inicial  $(Y_1, X_1)$  de la sección anterior serán las distribuciones a priori en esta sección, (Veáse [8]).

$$\pi(\sigma^2) = GInv\left(\frac{\nu_1}{2}, \frac{\nu_1 S_1^2}{2}\right) \text{ y } \pi(\beta|\sigma^2) = N_r(\hat{\beta}_1, \sigma^2 M_0^{-1}),$$



con  $\nu_1 = n_1 - r$ ,  $n_1$  es el tamaño de la muestra inicial, utilizando los mismos estimadores  $\hat{\beta}$  y  $\sigma^2$  de la sección anterior.

Supóngase que se tiene una segunda muestra independiente de la primera  $(Y_2, X_2)$  de tamaño  $n_2$ , se tiene la siguiente función de verosimilitud

$$l(\beta, \sigma^2 | Y_2) = \frac{1}{(2\pi)^{n_2/2} \sigma^{n_2}} \exp \left[ -\frac{1}{2\sigma^2} (Y_2 - X_2\beta)^t (Y_2 - X_2\beta) \right].$$

Nótese que al tener una segunda muestra, el estimador  $\hat{\beta}$  tiene la siguiente forma

$$\begin{aligned} \hat{\beta} &= \left( \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}^t \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}^t \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \right) \\ &= \left( X_1^t \quad X_2^t \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right)^{-1} (X_1^t \quad X_2^t) \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \end{aligned}$$

$$\hat{\beta} = (X_1^t X_1 + X_2^t X_2)^{-1} (X_1^t Y_1 + X_2^t Y_2). \quad (3.37)$$

Utilizando el teorema de Bayes, la distribución a posteriori conjunta de  $\beta$  y  $\sigma^2$  es:

$$\begin{aligned} \pi(\beta, \sigma^2 | Y_2) &\propto l(\beta, \sigma^2 | Y_2) \pi(\beta, \sigma^2) \\ &= \frac{1}{(2\pi)^{n_2/2} \sigma^{n_2+r}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \nu_2 S_2^2 + (\beta - \hat{\beta})^t X_2^t X_2 (\beta - \hat{\beta}) \right] \right\} \times \\ &\quad \left( \frac{1}{\sigma^2} \right)^{\frac{\nu_1}{2}+1} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \nu_1 S_1^2 + (\beta - \hat{\beta})^t X_1^t X_1 (\beta - \hat{\beta}) \right] \right\}, \quad (3.38) \end{aligned}$$

con  $\nu_2 = n_2 - r$  y  $S_2^2 = \frac{(Y_2 - X_2 \hat{\beta})^t (Y_2 - X_2 \hat{\beta})}{\nu_2}$ .

Por la ley multiplicativa de la probabilidad, la distribución a posteriori se puede ver también de la siguiente forma,

$$\pi(\beta, \sigma^2 | Y_2) = \pi(\sigma^2 | Y_2) \pi(\beta | \sigma^2, Y_2) \quad (3.39)$$

con  $\pi(\sigma^2|Y_2)$  distribución marginal para  $\sigma^2$  dada la segunda muestra,

$$\begin{aligned}
\pi(\sigma^2|Y_2) &\propto \frac{1}{(\sigma^2)^{\frac{n_2}{2} + \frac{\nu_1}{2} + 1 + r}} \exp\left\{-\frac{\nu_2 S_2 + \nu_1 S_1}{2\sigma^2}\right\} \\
&\quad \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\beta - \hat{\beta})^t \frac{X_2^t X_2 + X_1^t X_1}{\sigma^2} (\beta - \hat{\beta})\right] d\beta \\
&\propto \frac{1}{(\sigma^2)^{\frac{n_2}{2} + \frac{\nu_1}{2} + 1 + r}} \exp\left\{-\frac{\nu_2 S_2 + \nu_1 S_1}{2\sigma^2}\right\} \left\{(\sqrt{2\pi})^r \frac{|X_2^t X_2 + X_1^t X_1|^{-1/2}}{[(\sigma^2)^{2r}]^{-1/2}}\right\} \\
&\propto \frac{1}{(\sigma^2)^{\frac{n_2}{2} + \frac{n_1 - r}{2} + 1 + r - r}} \exp\left\{-\frac{\nu_2 S_2 + \nu_1 S_1}{2\sigma^2}\right\} \\
&\propto \frac{1}{(\sigma^2)^{\frac{\nu}{2} + 1}} \exp\left\{-\frac{\nu S^2}{2\sigma^2}\right\},
\end{aligned}$$

con

$$\nu = n_1 + n_2 - r,$$

$$S^2 = \frac{1}{\nu} \left[ (Y_1 - X_1 \hat{\beta})^t (Y_1 - X_1 \hat{\beta}) + (Y_2 - X_2 \hat{\beta})^t (Y_2 - X_2 \hat{\beta}) \right],$$

luego

$$\pi(\sigma^2|Y_2) = \text{GInv}\left(\frac{\nu}{2}, \frac{\nu S^2}{2}\right). \quad (3.40)$$

Para el segundo término de la ecuación (3.39) que es la distribución condicional a posteriori de  $\beta$  dados  $\sigma^2$  y  $Y_2$  se tiene que,

$$\pi(\beta|\sigma^2, Y_2) = \frac{\pi(\beta, \sigma^2|Y_2)}{\pi(\sigma^2|Y_2)}.$$

$$\begin{aligned}
& \propto \frac{\frac{1}{\sigma^{n_2+n_1+2}} \exp \left\{ -\frac{1}{2\sigma^2} \{ \nu_2 S_2^2 + \nu_1 S_1^2 + (\beta - \hat{\beta})^t [X_2^t X_2 + X_1^t X_1] (\beta - \hat{\beta}) \} \right\}}{\frac{1}{(\sigma^2)^{\frac{\nu}{2}+1}} \exp \left\{ -\frac{\nu S^2}{2\sigma^2} \right\}} \\
& = \frac{(\sigma^2)^{\left(\frac{n_1+n_2-r}{2}+1\right)}}{(\sigma^2)^{\left(\frac{n_1+n_2}{2}+1\right)}} \times \\
& \quad \exp \left\{ -\frac{1}{2\sigma^2} [\nu_2 S_2^2 + \nu_1 S_1^2 + (\beta - \hat{\beta})^t (X_2^t X_2 + X_1^t X_1) (\beta - \hat{\beta}) - \nu_1 S_1^2 - \nu_2 S_2^2] \right\} \\
& \propto \sigma^{-r} \exp \left\{ -\frac{1}{2\sigma^2} (\beta - \hat{\beta})^t (X_2^t X_2 + X_1^t X_1) (\beta - \hat{\beta}) \right\} \\
& = \sigma^{-r} \exp \left\{ -\frac{1}{2} (\beta - \hat{\beta})^t \frac{M}{\sigma^2} (\beta - \hat{\beta}) \right\},
\end{aligned}$$

con

$$M = X_1^t X_1 + X_2^t X_2 = M_0 + X_2^t X_2,$$

así la distribución condicional a posteriori queda definida como,

$$\pi(\beta | \sigma^2, Y_2) = N_r(\hat{\beta}, \sigma^2 M^{-1}). \quad (3.41)$$

Análogamente a lo que se realizó en la sección anterior, para encontrar la distribución marginal a posteriori para  $\beta$ , se tiene que,

$$\pi(\beta | Y_2) = t_r(\nu, \hat{\beta}, S^2 M^{-1}), \quad (3.42)$$

de donde se tiene que para cada  $\beta_i$ ,

$$\pi(\hat{\beta}_i | Y_2) = t(\nu, \hat{\beta}_i, h_{ii}^*), \quad (3.43)$$

donde  $h_{ii}^*$  es el elemento  $(i, i)$  de  $S^2 M^{-1}$ .

Finalmente, para la regresión lineal múltiple bayesiana, con información a priori conjugada, se tienen los siguientes grados de libertad, matriz de datos  $M$  y estadísticos respectivamente.

$$\nu = n_1 + n_2 - r \quad (3.44)$$

$$M = X_1^t X_1 + X_2^t X_2 = M_0 + X_2^t X_2, \quad (3.45)$$

$$\hat{\beta} = (X_1^t X_1 + X_2^t X_2)^{-1} (X_1^t Y_1 + X_2^t Y_1) = M^{-1} (X_1^t Y_1 + X_2^t Y_1), \quad (3.46)$$

$$S^2 = \frac{(Y_1 - X_1 \hat{\beta})^t (Y_1 - X_1 \hat{\beta}) + (Y_2 - X_2 \hat{\beta})^t (Y_2 - X_2 \hat{\beta})}{\nu} \quad (3.47)$$

### 3.3.3. Inferencia bayesiana para $\beta$ y $\sigma^2$ .

Ya que se tiene toda la información para los parámetros en la distribución a posteriori, lo que interesa es realizar inferencias para cada uno de ellos, como estimaciones y cálculo de regiones HPD.

#### Estimación puntual

Para hacer la estimación del parámetro  $\beta$ , se necesita calcular la moda a posteriori, sin embargo, dado que la distribución marginal a posteriori de  $\beta$  es una distribución  $t$ -multivariada la moda a posteriori coincide con la media y la mediana por la simetría de la distribución, entonces el estimador de  $\beta$  es  $\hat{\beta}$  mostrada en la ecuación (3.46).

Por otra parte, para  $\sigma^2$ , la moda de la distribución gamma-inversa es el E.M.V. de  $\pi(\sigma^2|Y)$ , se tiene que

$$\pi(\sigma^2|Y) = \frac{\nu S^2}{2\Gamma(\frac{\nu}{2})} \sigma^{(\frac{\nu}{2}+1)} \exp\left\{-\frac{\nu S^2}{2\sigma^2}\right\},$$

aplicando logaritmo natural  $\ln$  a la distribución anterior,

$$\ln \pi(\sigma^2|Y) = \ln \frac{\nu S^2}{2\Gamma(\frac{\nu}{2})} - \left(\frac{\nu}{2} + 1\right) \ln \sigma^2 - \frac{\nu S^2}{2\sigma^2},$$

derivando con respecto a  $\sigma^2$  para encontrar el valor que maximiza a la función,

$$\frac{d\pi(\sigma^2|Y)}{d\sigma^2} = -\left(\frac{\nu}{2} + 1\right) (\sigma^2)^{-1} + \frac{\nu S^2}{2} (\sigma^2)^{-2} = 0,$$

el estimador es,

$$\sigma_{moda}^2 = \frac{\nu S^2}{\nu + 2}. \quad (3.48)$$

Pero como la distribución gamma-inversa no es simétrica, la media a posteriori, no es la misma que la moda, se elige la que tenga menor varianza, utilizando las propiedades de la función gamma-inversa (Ver apéndice, sección A.1.).

$$\mu^\pi = E^{\pi(\sigma^2|Y)}[\sigma^2] = \frac{\nu S^2/2}{\frac{\nu}{2} - 1} = \frac{\nu S^2}{\nu - 2} \text{ para } \frac{\nu}{2} > 1, \quad (3.49)$$

el error de estimación o varianza a posteriori para  $\mu^\pi$  es,

$$\begin{aligned} V_{\mu^\pi}^\pi &= E^{\pi(\sigma^2|Y)}[(\sigma^2 - \mu^\pi)^2] = \frac{\left(\frac{\nu S^2}{2}\right)^2}{(\nu/2 - 1)^2(\nu/2 - 2)} \\ &= \frac{2\nu^2 S^4}{(\nu - 2)^2(\nu - 4)} \quad \text{para } \frac{\nu}{2} > 2. \end{aligned}$$

el error de estimación para  $\sigma_{moda}^2$  se obtiene utilizando (1.7),

$$\begin{aligned} V_{\sigma_{moda}^2}^\pi &= V^\pi + (\mu^\pi - \sigma_{moda}^2)^2 \\ &= \frac{2\nu^2 S^4}{(\nu - 2)^2(\nu - 4)} + \left(\frac{\nu S^2}{\nu - 2} - \frac{\nu S^2}{\nu + 2}\right)^2 \\ &= \frac{2\nu^2 S^4}{(\nu - 2)^2(\nu - 4)} + \frac{16\nu^2 S^4}{(\nu^2 - 4)^2}. \end{aligned}$$

### Intervalos HPD para $\beta$ .

La distribución a posteriori para  $\beta_i$  es una  $t(\nu, \hat{\beta}_i, h_{ii}^*)$ , una región HPD para  $\beta$  tiene los siguientes límites superiores e inferiores

$$\begin{aligned} Li &= t_{\alpha/2} \sqrt{h_{ii}^*} + \hat{\beta}_i, \\ Ls &= t_{1-\alpha/2} \sqrt{h_{ii}^*} + \hat{\beta}_i = -t_{\alpha/2} \sqrt{h_{ii}^*} + \hat{\beta}_i, \end{aligned}$$

donde  $t_{\alpha/2}$  y  $t_{1-\alpha/2}$  son los cuantiles  $\alpha/2$  y  $1 - \alpha/2$  de la distribución  $t$  de Student con  $\nu$  grados de libertad.



## Capítulo 4

### Aplicación.

#### 4.1. Tratamiento de los datos.

Los modelos de regresión lineal que se analizarán en este capítulo son:

$$\alpha(t) = \alpha_0 + (\mu_{\alpha 0} \cos \delta)t + \frac{1}{2}(a_{\alpha} \cos \delta)t^2 + pf_{\alpha}(t), \quad (4.1)$$

$$\delta(t) = \delta_0 + \mu_{\delta 0}t + \frac{1}{2}a_{\delta}t^2 + pf_{\delta}(t). \quad (4.2)$$

que son las ecuaciones de posición de una estrella, en este caso particular aplicadas a las observaciones de T Tau Sb.

En el caso de los parámetros  $\mu_{\alpha 0} \cos \delta$  y  $a_{\alpha} \cos \delta$ , se tomarán como lineales, pues lo que se quiere determinar es el valor completo de las expresiones.

Se tiene un total de 13 observaciones, 12 tomadas de [12] y una más proporcionada por Laurent Loinard C. del CRyA-UNAM.

Para el modelo con información a priori no informativa, se tomarán las seis primeras observaciones,  $n_1 = 6$ , y para el modelo con información a priori conjugada se anexarán las 7 restantes,  $n_2 = 7$ .

**Nota.** Las unidades de medición de los datos que se manejan a continuación están dadas en grados.

##### 4.1.1. Tratamiento de los datos para la ascensión recta.

Para la primera ecuación se tienen que determinar los parámetros  $\alpha_0$ ,  $\mu_{\alpha 0} \cos \delta$ ,  $a_{\alpha} \cos \delta$ ,  $p$ , los datos disponibles son  $t$ ,  $f_{\alpha}(t)$ .

La matriz  $X_{1,\alpha}$  de datos observados es la siguiente.

$$\begin{aligned}
 X_{1,\alpha} &= \begin{bmatrix} 1 & t_1 & \frac{1}{2}t_1^2 & f_\alpha(t_1) \\ \vdots & \dots & \dots & \vdots \\ 1 & t_6 & \frac{1}{2}t_6^2 & f_\alpha(t_6) \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 3.7289820706 & 6.9526536414 & 0.9659138918 \\ 1 & 3.8791651183 & 7.5239610076 & 0.2158925718 \\ 1 & 4.0375210776 & 8.1507882261 & -0.7481079257 \\ 1 & 4.2341123256 & 8.9638535929 & -0.9063627414 \\ 1 & 4.3651731184 & 9.5273681766 & -0.2508576850 \\ 1 & 4.5180774071 & 10.206511728 & 0.6808492193 \end{bmatrix}.
 \end{aligned}$$

Y el vector  $Y_{1,\alpha}$  de observaciones es.

$$Y_{1,\alpha} = \begin{bmatrix} \alpha(t_1) \\ \vdots \\ \alpha(t_6) \end{bmatrix} = \begin{bmatrix} 65.4976053925000 \\ 65.4976040854167 \\ 65.4976024262500 \\ 65.4976022583333 \\ 65.4976036741667 \\ 65.4976056100000 \end{bmatrix}.$$

Para la distribución a priori conjugada, se toma una segunda muestra, cuya matriz de observaciones es

$$\begin{aligned}
 X_{2,\alpha} &= \begin{bmatrix} 1 & t_1 & \frac{1}{2}t_1^2 & f_\alpha(t_1) \\ \vdots & \dots & \dots & \vdots \\ 1 & t_7 & \frac{1}{2}t_7^2 & f_\alpha(t_7) \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 4.7092031007 & 11.0882969216 & 1.0126420611 \\ 1 & 4.8566465277 & 11.7935077477 & 0.3603317742 \\ 1 & 4.9904377856 & 12.4522346458 & -0.501639227 \\ 1 & 5.1488020626 & 13.2550813401 & -1.031065348 \\ 1 & 5.3535855052 & 14.3304388812 & -0.323953192 \\ 1 & 5.5174114216 & 15.2209143977 & 0.676503019 \\ 1 & 8.4089361723 & 35.3551037747 & -0.022167258 \end{bmatrix}.
 \end{aligned}$$



El vector  $Y_{2,\alpha}$  de observaciones es.

$$Y_{2,\alpha} = \begin{bmatrix} \alpha(t_1) \\ \vdots \\ \alpha(t_6) \end{bmatrix} = \begin{bmatrix} 65.4976064483333 \\ 65.4976054162500 \\ 65.4976039533333 \\ 65.4976031945833 \\ 65.4976047812500 \\ 65.4976069495833 \\ 65.4976120129167 \end{bmatrix}.$$

De la ecuación (3.37), se tiene que las estimaciones bayesianas puntuales de los parámetros del modelo (4.1) son:

$$\begin{bmatrix} \hat{\alpha}_0 \\ \widehat{\mu_\alpha \cos \delta} \\ \widehat{a_\alpha \cos \delta} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} 65.49760355776178 \\ -7.983757763820043e - 07 \\ 4.278631251623810e - 07 \\ 1.893139522913802e - 06 \end{bmatrix},$$

y de la ecuación (3.47) el estadístico para  $\sigma^2$  es:

$$S_\alpha^2 = 3.79247028975891e - 16 \quad (4.3)$$

En las Figuras 4.1, 4.2, 4.3 y 4.4, se tienen las distribución a priori y a posteriori para los parámetros  $\alpha_0$ ,  $\mu_{\alpha_0} \cos \delta$ ,  $a_\alpha \cos \delta$  y  $p$  respectivamente, nótese que la distribución a posteriori es menos dispersa que la a priori, lo que nos indica que se tiene una mejor aproximación al valor del parámetro con la distribución a posteriori. La media, mediana y moda a posteriori de la distribución para  $\alpha_0$  coinciden como se esperaba, por tener  $\alpha_0$  una distribución a posteriori  $t$ , con 9 grados de libertad, como se muestra en la Figura 4.1.

$$\begin{aligned} E[\alpha_0] &= \alpha_{0_{moda}} = \hat{\alpha}_0 = 65.49760355776178 \\ \text{Var}(\alpha_0) &= 1.150195290e - 14, \end{aligned}$$

La media y moda de la distribución a posteriori para  $\mu_\alpha \cos \delta$  como se muestra en la Figura 4.2, coinciden como se esperaba, por tener  $\mu_\alpha \cos \delta$  una distribución a posteriori  $t$  con 9 grados de libertad, ya que esta es una distribución simétrica.

$$\begin{aligned} E[\mu_\alpha \cos \delta] &= \mu_\alpha \cos \delta_{moda} = \widehat{\mu_\alpha \cos \delta} = -7.983757763820043e - 07 \\ \text{Var}(\mu_\alpha \cos \delta) &= 1.403086838e - 15. \end{aligned}$$

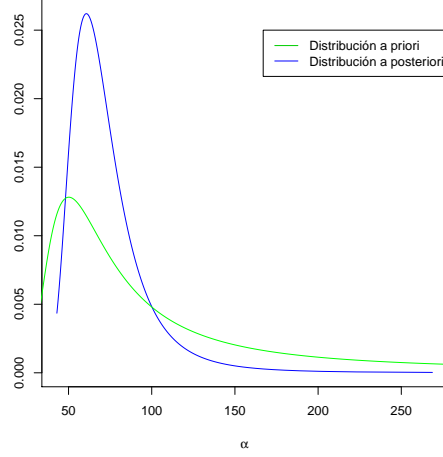


Figura 4.1: Distribución a priori y a posteriori para  $\alpha_0$

En la Figura 4.3 se ve como la media y moda a posteriori  $a_\alpha \cos \delta$  coinciden por tener ésta una distribución a posteriori  $t$ , con 9 grados de libertad.

$$\begin{aligned} E[a_\alpha \cos \delta] &= a_\alpha \cos \delta_{moda} = \widehat{a_\alpha \cos \delta} = 4.278631251623810e - 07 \\ \text{Var}(a_\alpha \cos \delta) &= 3.751050129e - 17 \end{aligned}$$

Para la distribución a posteriori de la paralaje, como se ve en la Figura 4.4 se tiene que,

$$\begin{aligned} E[p] &= p_{moda} = \hat{p} = 1.893139522913802e - 06, \\ \text{Var}(p) &= 6.499137229e - 17 \end{aligned}$$

con una distribución a posteriori  $t_9$ .

Para la varianza  $\sigma^2$  se tiene una distribución  $GInv(4.5, 1.706611630e - 15)$  la cual se muestra en la Figura 4.5, en este caso se graficó, el histograma para diferenciar a  $\sigma^2$  de los parámetros del modelo, como se dijo en el Capítulo 3, se tomará como estimador de  $\sigma^2$  al que tenga menor varianza a posteriori entre la media a posteriori y la moda a posteriori.

$$\begin{aligned} E^\pi[\sigma^2] &= 4.87603323e - 16, \\ V_{E^\pi[\sigma^2]}^\pi &= 9.510280023e - 32 \end{aligned}$$

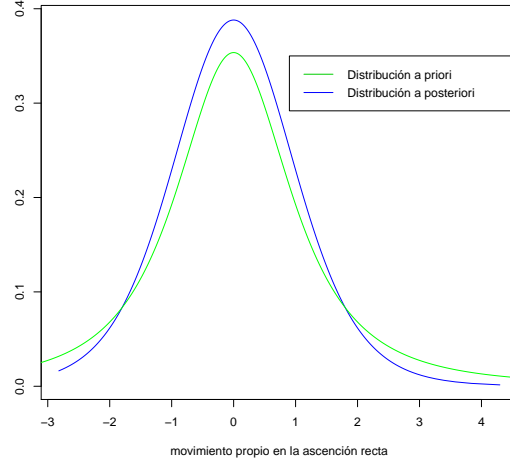


Figura 4.2: Distribución a priori y a posteriori para  $\mu_\alpha \cos \delta$ .

$$\begin{aligned}\sigma_{moda}^2 &= 3.102930237e - 16, \\ V_{\sigma_{moda}^2}^\pi &= 1.265417425e - 31\end{aligned}$$

El estimador para  $\sigma^2$  es la media aposteriori  $E^\pi[\sigma^2]$ , por tener menor error de estimación.

#### 4.1.2. Tratamiento de los datos para la declinación.

Para la ecuación (4.2) se tienen que determinar los parámetros  $\delta_0, \mu_\delta, a_\delta, p$ , los datos disponibles son  $t$  y  $f_\delta(t)$ , entonces la matriz  $X_{1,\delta}$  de datos observados para el modelo es la siguiente.

$$\begin{aligned}X_{1,\delta} &= \begin{bmatrix} 1 & t_1 & \frac{1}{2}t_1^2 & f_\delta(t_1) \\ \vdots & \dots & \dots & \vdots \\ 1 & t_6 & \frac{1}{2}t_6^2 & f_\delta(t_6) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 3.7289820706 & 6.9526536414 & 0.1382735679 \\ 1 & 3.8791651183 & 7.5239610076 & 0.0004501279 \\ 1 & 4.0375210776 & 8.1507882261 & -0.1411006458 \\ 1 & 4.2341123256 & 8.9638535929 & -0.1257018894 \\ 1 & 4.3651731184 & 9.5273681766 & -0.0053006148 \\ 1 & 4.5180774071 & 10.206511728 & 0.1346500422 \end{bmatrix}\end{aligned}$$

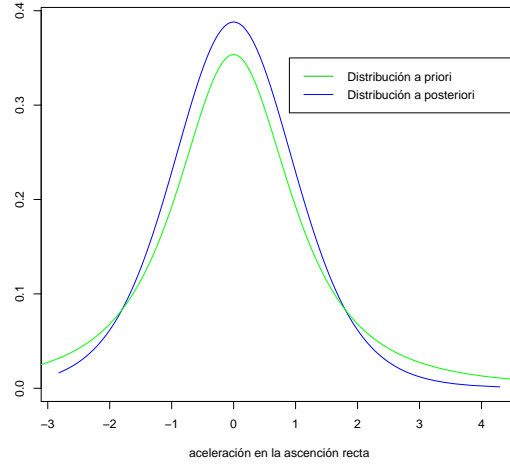


Figura 4.3: Distribución a priori y a posteriori para  $a_\alpha \cos \delta$ .

Y el vector  $Y_{1,\delta}$  de observaciones es.

$$Y_{1,\delta} = \begin{bmatrix} \delta(t_1) \\ \vdots \\ \delta(t_6) \end{bmatrix} = \begin{bmatrix} 19.5349215605556 \\ 19.5349212650000 \\ 19.5349209227778 \\ 19.5349209258333 \\ 19.5349211205556 \\ 19.5349212922222 \end{bmatrix}$$

Con la segunda muestra se obtienen las matrices,

$$X_{2,\delta} = \begin{bmatrix} 1 & t_1 & \frac{1}{2}t_1^2 & f_\delta(t_1) \\ \vdots & \dots & \dots & \vdots \\ 1 & t_7 & \frac{1}{2}t_7^2 & f_\delta(t_7) \end{bmatrix} = \begin{bmatrix} 1 & 4.7092031007 & 11.0882969216 & 0.1496144583 \\ 1 & 4.8566465277 & 11.7935077477 & 0.0244641213 \\ 1 & 4.9904377856 & 12.4522346458 & -0.1087639659 \\ 1 & 5.1488020626 & 13.2550813400 & -0.1634824795 \\ 1 & 5.3535855053 & 14.3304388812 & -0.0175941925 \\ 1 & 5.5174114216 & 15.2209143977 & 0.1340065346 \\ 1 & 8.4089361723 & 35.3551037747 & 0.0386592186 \end{bmatrix}.$$

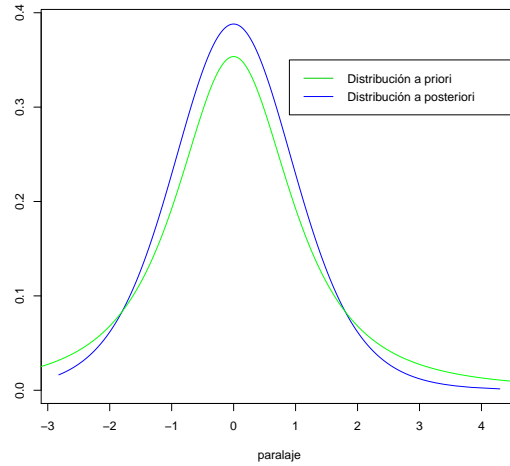


Figura 4.4: Distribución a priori y a posteriori para  $p$ .

y el vector,

$$Y_{2,\delta} = \begin{bmatrix} \delta(t_1) \\ \vdots \\ \delta(t_7) \end{bmatrix} = \begin{bmatrix} 19.5349212783333 \\ 19.5349210086111 \\ 19.5349206511111 \\ 19.5349205072222 \\ 19.5349207922222 \\ 19.5349209994444 \\ 19.5349185477778 \end{bmatrix}.$$

Luego, de la ecuación (3.47), se tiene el estimador bayesiano de los parámetros del modelo de regresión lineal (4.2),

$$\begin{bmatrix} \hat{\delta}_0 \\ \hat{\mu}_\delta \\ \hat{a}_\delta \\ \hat{p} \end{bmatrix} = \begin{bmatrix} 1.9534920557e + 01 \\ 5.3804077298e - 07 \\ -1.8681344597e - 07 \\ 1.9989509606e - 06 \end{bmatrix}, \quad (4.4)$$

utilizando la ecuación (3.37)

$$S_\delta^2 = 1.7645286037e - 15. \quad (4.5)$$

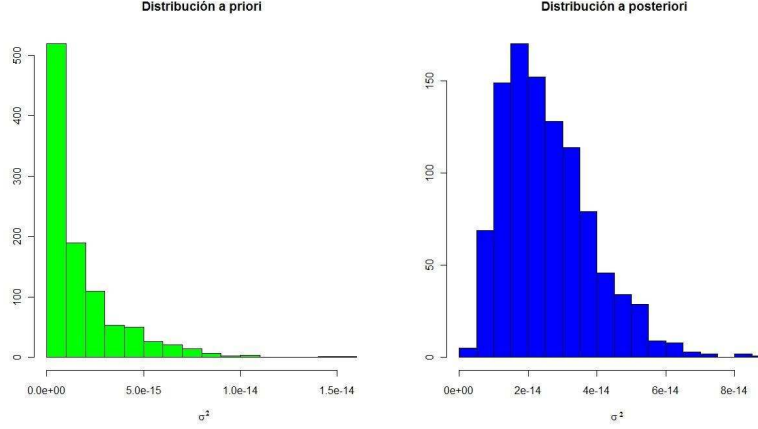


Figura 4.5: Histogramas distribución a priori y a posteriori para  $\sigma^2$ .

Las gráficas de la distribución a priori y a posteriori para cada parámetro, se muestran en las Figuras 4.6, 4.7, 4.8 y 4.9 para  $\delta_0$ ,  $\mu_\delta$ ,  $a_\delta$  y  $p$ , respectivamente. Como en el caso anterior las distribuciones a posteriori, tienen menos dispersion que la distribución a priori.

En la Figura 4.6 se tiene una distribución a posteriori  $t$ , con 9 grados de libertad, con media y varianza a posteriori para  $\delta_0$ ,

$$\begin{aligned} E[\delta_0] &= \delta_{0\text{moda}} = \hat{\delta}_0 = 1.9534920557e + 01 \\ Var(\delta_0) &= 5.29772526e - 14. \end{aligned}$$

También la Figura 4.7 se tiene una distribución a posteriori  $t_9$ , con media y varianza a posteriori para  $\mu_\delta$ ,

$$\begin{aligned} E[\mu_\delta] &= \mu_{\delta\text{moda}} = \hat{\mu}_\delta = 5.3804077298e - 07 \\ Var(\mu_\delta) &= 6.4718237393e - 15. \end{aligned}$$

En la Figura 4.8 se tiene una distribución a posteriori  $t_9$ , con media y varianza a posteriori para  $a_\delta$ ,

$$\begin{aligned} E[a_\delta] &= a_{\delta\text{moda}} = \hat{a}_\delta = -1.8681344597e - 07 \\ Var(a_\delta) &= 1.7344753862e - 16. \end{aligned}$$

La Figura 4.9 muestra una distribución a posteriori  $t$ , con 9 grados de libertad, con media y varianza a posteriori para  $p$ ,

$$\begin{aligned} E[p] &= a_{\delta_{moda}} = \hat{a}_{\delta} = 1.9989509606e - 06 \\ Var(p) &= 1.1597373219e - 14. \end{aligned}$$

Análogo a lo que se realizó en la sección anterior para la distribución de  $\sigma_{\alpha}^2$ , se realizará para  $\sigma_{\delta}^2$ , y esta última tiene una distribución a posteriori  $GInv(4.5, 7.940379e - 15)$ , con media y moda a posteriori,

$$\begin{aligned} E^{\pi}[\sigma_{\delta}^2] &= 2.2686796333e - 15 \\ V_{E^{\pi}[\sigma_{\delta}^2]}^{\pi} &= 2.0587629114e - 30, \end{aligned}$$

$$\begin{aligned} \sigma_{\delta_{moda}}^2 &= 1.4437052212e - 15 \\ V_{\sigma_{\delta_{moda}}^2}^{\pi} &= 2.739345692e - 30. \end{aligned}$$

El estimador de  $\sigma^2$  que mejor lo aproxima es  $E^{\pi}[\sigma_{\delta}^2]$ .

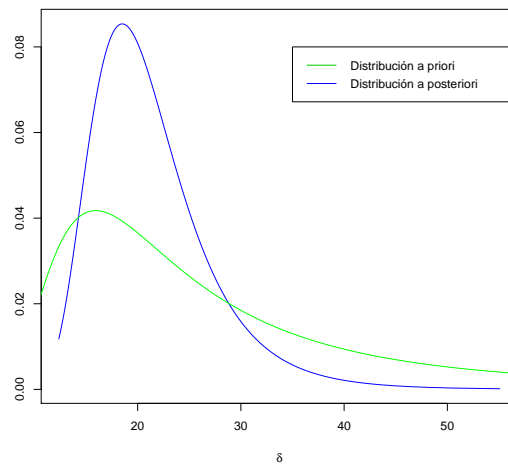


Figura 4.6: Distribución a priori y a posteriori para  $\delta_0$ .

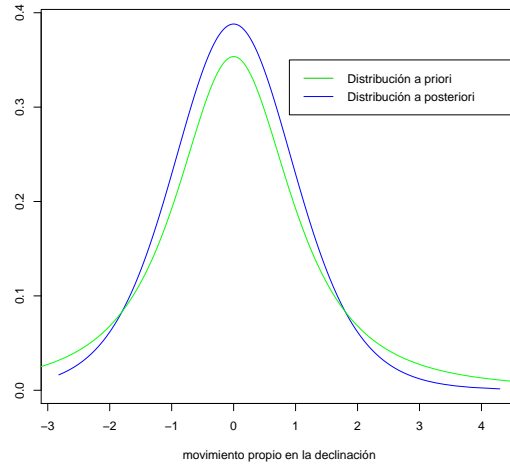


Figura 4.7: Distribución a priori y a posteriori para  $\mu_\delta$ .

### 4.1.3. Comparación de las estimaciones puntuales clásicas y bayesianas

Tabla 4.1. Estimación clásica de los parámetros del modelo de regresión para la ascensión recta.

Parámetro	Valor estimado	Desv. Estándar
$\alpha_0$	6.5497600096e+01	1.805864383e-05
$\mu_\alpha \cos \delta$	9.0058744162e-07	8.7462265125e-06
$a_\alpha \cos \delta$	1.3593543388e-08	2.1087256299e-06
$p$	1.9205624092e-06	9.4635468187e-08
$\sigma^2$	1.0053860731e-15	3.1707823532e-08



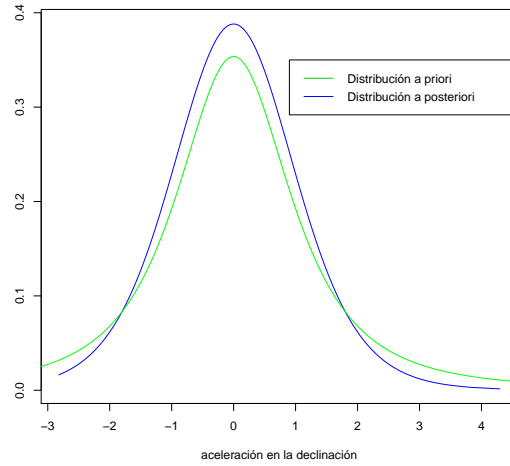
Figura 4.8: Distribución a priori y a posteriori para  $a_\delta$ .

Tabla 4.2. Estimación clásica de los parámetros del modelo de regresión para la declinación.

Parámetro	Valor estimado	Desv. Estándar
$\delta_0$	1.9534903270e+01	1.4046599938e-05
$\mu_\delta$	9.0325224066e-06	6.8407367523e-06
$a_\delta$	-2.2637441511e-06	1.6584489292e-06
$p$	2.4615155509e-06	4.5571251486e-07
$\sigma^2$	4.3051187951e-16	2.0748780193e-08

Tabla 4.3. Estimación bayesiana de los parámetros del modelo de regresión para la ascensión recta.

Parámetro	Valor estimado	Desv. Estándar
$\alpha_0$	6.5497603558e+01	1.0724715802e-07
$\mu_\alpha \cos \delta$	-7.9837577638e-07	3.7457800769e-08
$a_\alpha \cos \delta$	4.2786312516e-07	6.1245817236e-09
$p$	1.8931395229e-06	8.0617226627e-09
$\sigma^2$	Media 4.87603323e-16	3.083874191e-16
	Moda 3.102930237e-16	3.557270618e-16

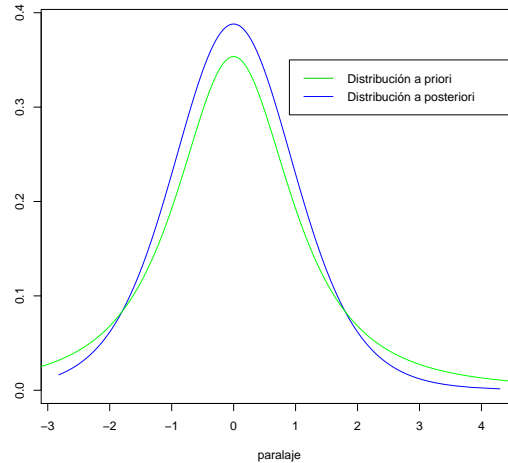


Figura 4.9: Distribución a priori y a posteriori para  $p$ .

Tabla 4.4. Estimación bayesiana de los parámetros del modelo de regresión para la declinación.

Parámetro	Valor estimado	Desv. Estándar
$\delta_0$	1.9534920557e+01	2.3016787917e-07
$\mu_\delta$	5.3804077298e-07	8.0447645952e-08
$a_\delta$	-1.8681344597e-07	1.3169948315e-08
$p$	1.9989509606e-06	1.0769110093e-07
$\sigma^2$	Media	2.2686796333e-15
	Moda	1.4437052212e-15

En las tablas (4.1) y (4.2) se muestran los resultados obtenidos con la metodología clásica y en las tablas (4.3) y (4.4) con la metodología bayesiana, como se observa las estimaciones puntuales son muy parecidas sin embargo, los errores de estimación se aprecian un poco más pequeños en las estimaciones bayesianas.

En las siguientes Tablas 4.5 y 4.6,  $p$  es calculado al promediar el obtenido

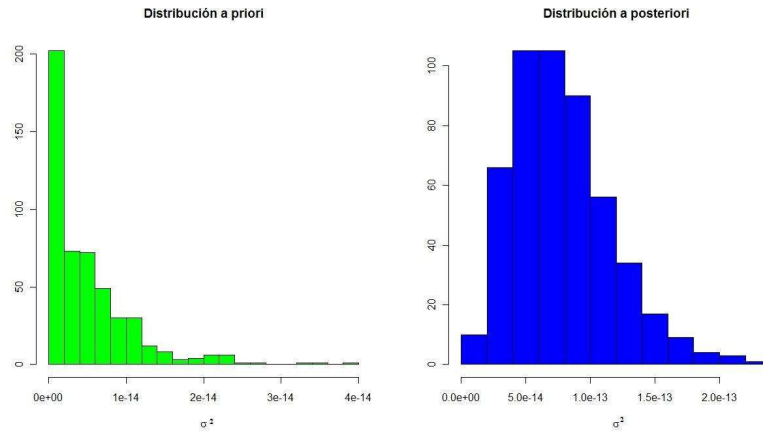


Figura 4.10: Histogramas distribución a priori y a posteriori para  $\sigma^2$ .

para la ascensión recta, con el obtenido para la declinación.

Tabla 4.5. Estimación clásica de los parámetros de regresión en *mas* (mili arco segundo) y *yr* (años luz).

Parámetro	Valor estimado	Desv. Estándar
$\alpha_0$	$4^h 21^m 59.42402161^s$	$0.004334074519^s$
$\mu_\alpha \cos \delta$	$3.2421 \text{ masyr}^{-1}$	$31.48641544 \text{ masyr}^{-1}$
$a_\alpha \cos \delta$	$0.0489 \text{ masyr}^{-2}$	$7.591412 \text{ masyr}^{-2}$
$\delta_0$	$19^\circ 32' 5.651772''$	$0.050567759''$
$\mu_\delta$	$32.51708066 \text{ masyr}^{-1}$	$24.62665231 \text{ masyr}^{-1}$
$a_\delta$	$-8.149478944 \text{ masyr}^{-2}$	$1.640565053 \text{ masyr}^{-2}$
$p$	$7.8877 \text{ mas}$	$0.990626369 \text{ mas}$

Tabla 4.5. Estimación bayesiana de los parámetros de regresión en *mas* (mili arco segundo) y *yr* (años luz).

Parámetro	Valor estimado	Desv. Estándar
$\alpha_0$	$4^h 21^m 59.42485^s$	$0.0002573931^s$
$\mu_\alpha \cos \delta$	$-2.877415 \text{ masyr}^{-1}$	$0.134848 \text{ masyr}^{-1}$
$a_\alpha \cos \delta$	$1.5403072 \text{ masyr}^{-2}$	$0.022048 \text{ masyr}^{-2}$
$\delta_0$	$19^\circ 32' 5.71398''$	$0.00828604''$
$\mu_\delta$	$1.9369467 \text{ masyr}^{-1}$	$0.289611525 \text{ masyr}^{-1}$
$a_\delta$	$-0.672528 \text{ masyr}^{-2}$	$0.047411813 \text{ masyr}^{-2}$
$p$	$7.005762868 \text{ mas}$	$0.208355082 \text{ mas}$

En la segunda columna de la Tabla 4.5 se dan los estimadores para la primera muestra, se nota que en comparación con los de la segunda columna de la Tabla 4.6, los primeros no son buenos estimadores ya que su error de estimación es mayor a los obtenidos bayesianamente.

Con el análisis bayesiano se mejoran las estimaciones, pues se dispone de más información a la que ya se tiene, pero la diferencia no radica tanto en los valores de los estimadores, sino en que bayesianamente, al darle una distribución a priori al parámetro, se afirma, al menos subjetivamente, que estas estimaciones son las mejores.

#### 4.1.4. Intervalos HPD para los parámetros.

Con una probabilidad del 0.95 se tiene los intervalos HPD para los parámetros del modelo de ascensión recta,

$$\begin{aligned}
 \alpha_0 &\in [65.49760331, 65.49760379] \\
 \mu_\alpha \cos \delta &\in [-8.831128132e - 07, 4.136387394e - 07] \\
 a_\alpha \cos \delta &\in [4.140080963e - 07, 4.417181539e - 07] \\
 p &\in [1.874902293e - 06, 1.911376751e - 06],
 \end{aligned}$$

Con una probabilidad del 0.95 se tiene los intervalos HPD para los parámetros del modelo de declinación,

$$\begin{aligned}\delta_0 &\in [19.53492003, 19.53492107] \\ \mu_\delta &\in [3.560521082e - 07, 7.200294376e - 07] \\ a_\delta &\in [-2.16606503, -1.570203888e, 07] \\ p &\in [1.745332152e - 06, 2.232569768e - 06],\end{aligned}$$

#### Determinación de la distancia a T Tau.

Como se puede ver en las Tablas 4.5 y 4.6, los valores de la paralaje  $p$  varían de acuerdo al método utilizado, para el caso de la estimación clásica, es grande y el error de estimación también. El valor de la paralaje que se utilizará es el de la estimación bayesiana, que incorpora información adicional de acuerdo al modelo de regresión lineal con función de distribución a priori conjugada.

Utilizando las ecuaciones descritas en la sección 2.3, se tiene que con una paralaje de  $7.005762868 \text{ mas}$  la distancia a T Tauri es  $142.73963 \text{ pc}$ , y el reportado en el artículo [12] es  $146.7 \text{ pc}$ , los valores son muy parecidos a los obtenidos en [12].

## 4.2. Conclusiones

1. El enfoque bayesiano justifica el uso del conocimiento a priori subjetivo del investigador, en la metodología del modelo de regresión, ya que incorpora de manera coherente las fuentes de información como la información a priori y la información muestral. Es difícil plasmar a través de una distribución a priori el conocimiento que se tiene sobre cierto evento, sin embargo esta dificultad es mas bien práctica, entonces lo que se hizo en esta tesis fue poner atención especial en esta etapa del análisis.
2. Se pudo observar, que cuando no se cuenta con información a priori en el caso de haber asignado distribución a priori no informativa a los parámetros del modelo, la metodología bayesiana y clásica proporcionan resultados similares, sin embargo, no hay que perder de vista que la interpretación desde ambos enfoques es radicalmente diferente, ya que en el análisis bayesiano se trata al parámetro como aleatorio para tener mejor conocimiento sobre su comportamiento además de que se incorpora también la información a priori que se tiene acerca del mismo.
3. Al contar con más información a priori, es decir, cuando se hizo el análisis bayesiano asignando una distribución a priori conjugada, y utilizando un mayor número de muestras, los estimadores obtenidos con la metodología bayesiana redujeron el error de estimación, lo que quiere decir que las estimaciones hechas son buenas.
4. La estimación que se obtuvo de la paralaje para T Tau Sb es 7.005762868 *mas*, por lo tanto, la distancia a esta estrella es de 142.73963 pc, la cual es muy parecida a la obtenida en [12], sin embargo, la variabilidad de las estimaciones obtenidas con el método clásico fueron un poco mayor a las estimaciones obtenidas por el método bayesiano.
5. Los resultados obtenidos son similares, pero el análisis bayesiano es recomendable porque aparte de utilizar la información de la muestra que se tiene, se incluyó la mucha o poca información inicial acerca del parámetro de paralaje y así se determinó la distancia a T Tau Sb, además de que se tiene una mejor idea del comportamiento de dicho parámetro al asignarle una distribución a priori, dicha distribución puede cambiar de acuerdo a la elección del decisor, pero los resultados

---

no deberían ser muy distintos a menos que se trate de asignaciones drásticamente diferentes.





# Apéndice A

## Apéndice

### A.1. Distribuciones especiales.

1. Distribución  $\chi^2$  **no central**.

Si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes y  $X_i \sim N(\delta_i, 1)$  para  $i = 1, \dots, n$ , entonces

$$\sum_{i=1}^n X_i^2 \sim \chi_{n,\lambda}^2,$$

donde  $\chi_{n,\lambda}^2$  es la **distribución  $\chi^2$  no central**, con  $n$  grados de libertad y parámetro  $\lambda = \sum_{i=1}^n \delta_i^2$  de no centralidad.

2. Distribución **Gamma-Inversa**.

La variable aleatoria continua  $X$  tiene una distribución gamma-inversa (*GI**nv*) con parámetros  $a > 0$ ,  $b > 0$  si,

$$GI_{nv}(x|a, b) = cx^{-(a+1)}e^{-b/x}, \quad x > 0,$$

en donde  $c = \frac{b^a}{\Gamma(a)}$  y

$$E[x] = \frac{b}{a-1}, \quad a > 1,$$

$$V[x] = \frac{b^2}{(a-1)^2(a-2)}, \quad a > 2.$$

### 3. Distribución **t-Student multivariada.**

Un vector aleatorio  $X = (X_1, \dots, X_k)$  tiene una distribución t-Student multivariada de dimensión  $k$ , con parámetros  $\mu = (\mu_1, \dots, \mu_k)$ ,  $\lambda$  y  $a$  ( $\mu \in \mathbb{R}^k$ ,  $\lambda$  es una matriz de tamaño  $k \times k$ , simétrica y positiva definida,  $a > 0$ ), si

$$t_k(x|\mu, \lambda, a) = c \left[ 1 + \frac{1}{a}(x - \mu)^t \lambda (x - \mu) \right]^{-(a+k)/2}, \quad x \in \mathbb{R}^k,$$

$$\text{donde } c = \frac{\Gamma[(a+k)/2]}{\Gamma[a/2](a\pi)^{k/2}} |\lambda|^{1/2}.$$

## A.2. Algunos teoremas relacionados con matrices.

**Teorema A.2.1** *Sea  $A$  una matriz idempotente. El rango de  $A$  es su traza.*

**Teorema A.2.2** *Sea  $B$  una matriz de  $q \times n$ ,*

(i) *Si  $\text{ran}(B) = q$ , entonces  $B \times B$  es definida positiva.*

(ii) *Si  $\text{ran}(B) < q$ , entonces  $B \times B$  es semidefinida positiva.*

**Teorema A.2.3** .

(i) *Si  $A$  es definida positiva, entonces todos los elementos de la diagonal principal  $a_{ii}$  son positivos.*

(ii) *Si  $A$  es semidefinida positiva, entonces  $a_{ii} \geq 0$ .*

## A.3. Integrales

Integrales relacionadas con la distribución gamma inversa.

Sea  $\eta$  un vector de constantes de  $n \times 1$  y  $C$  una matriz definida positiva y simétrica,  $x \in \mathbb{R}$ .

$$\int_{-\infty}^{\infty} \exp \left[ \frac{1}{2}(x - \eta)^t C^{-1}(x - \eta) \right] dx = \sqrt{2\pi} |C|^{1/2}. \quad (\text{A.1})$$

Para  $a > 0$ ,  $p > 0$ ,  $x > 0$ .

$$\int_0^{\infty} x^{p-1} e^{-ax^2} dx = \frac{1}{2} a^{-p/2} \Gamma(p/2). \quad (\text{A.2})$$

## A.4. Valores esperados y teoría de la distribución.

**Teorema A.4.1** Sean  $A$  una matriz de  $r \times r$  de constantes y  $Y$  un vector aleatorio de  $r \times 1$ , con media  $\mu$  y matriz no singular  $V$  de covarianza.

1.  $Var[AY] = AVA^t$ .
2.  $E[Y^tAY] = tr(AV) + \mu^tA\mu$ .  
Si  $V = \sigma^2I$ , entonces  $E[Y^tAY] = \sigma^2tr(A) + \mu^tA\mu$ .

**Teorema A.4.2** Sean  $A$  una matriz de  $r \times r$  de constantes y  $Y$  un vector aleatorio normal multivariado de  $r \times 1$ , con media  $\mu$  y matriz no singular  $\sigma^2I$  de covarianza ( $Y \sim N(\mu, \sigma^2I)$ ).

Sea  $U$  la forma cuadrática dada por  $U = Y^tAY$ .

Si  $A$  es idempotente y de rango  $q$ , entonces

$$\frac{U}{\sigma^2} \sim \chi_{q,\lambda}^2, \quad (\text{A.3})$$

donde  $\lambda = \mu^tA\mu/\sigma^2$ .



## Bibliografía

- [1] Alamilla López Norma Edith, *Selección Bayesiana de Modelos: Una Aplicación a los Modelos de regresión Lineal Heteroscedásticos Usando Factores de Bayes Intrínsecos*, Tesis de maestría UNAM, 2004.
- [2] Berger James O., *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, second edition, USA 1985.
- [3] Bernardo José M., Smith Adrian F. M., *Bayesian Theory*, John Wiley & Sons, 1994.
- [4] Box George E.P., Tiao George C., *Bayesian Inferences in Statistical Analysis*, Wiley-Interscience Publication, USA 1992.
- [5] Carlin Bradley P., Louis Thomas A., *Bayesian and Empirical Bayes Methods for Data Analysis*, Chapman & Hall/CRC, segunda edición, 1996.
- [6] Carroll Bradley W., *An Introduction to Modern Astrophysics*, Pearson International Edition, second edition, USA 2007.
- [7] DeGroot Morris H., *Probabilidad y Estadística*, Addison-Wesley, segunda edición.
- [8] Eyzaguirre Pérez Raúl, Chue Gallardo Jorge, *Análisis Bayesiano del Modelo de Regresión Lineal Multiple*, Anales Científicos, Vol. XLIX, 294-318, 2001.
- [9] Graybill Franklin A., *Introduction to Matrices with Applications in Statistics*, Wadsworth Publishing Company, 1969.

- 
- [10] Karttunen H., Kröger P., Oja H., Poutanen M., Donner K.J., *Fundamental Astronomy*, Springer-Verlag, 1996.
- [11] Loinard Laurent, Mioduzseuski Amy J., Rodriguez Luis F., González Rosa A., Rodríguez Mónica I., Torres Rosa M. *Multi-epoch VLBA Observations of T Tauri South*, The Astrophysical Journal, 619: L179 - L182, 2005.
- [12] Loinard Laurent, Torres Rosa M., Mioduzseuski Amy J., Rodriguez Luis F., González-Lópezlira Rosa A., Lachaume Regis, Vázquez Virgilio, González Erandy, *VLBA Determination of the Distance to Nearby Star-Forming Regions. I. the Distance to T Tauri with 0.4% Accuracy*, The Astrophysical Journal, 671:546 - 554, 2007.
- [13] Mendenhall William, Wackerly Dennis D., Scheaffer Richard L., *Estadística Matemática con Aplicaciones*, Grupo editorial Iberoamerica, segunda edición, 1994.
- [14] Montgomery Douglas C., Peck Elizabeth A., Vining G. Geoffrey, *Introducción al Análisis de Regresión Lineal*, Compañía editorial continental, 2002.
- [15] Rencher Alvin C., *Linear Models in Statistics*, A John Wiley & Sons, Inc., Publication, second edition, USA. 2008.
- [16] R Development Core Team (2009). 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>.