UNIVERSIDAD CENTROCCIDENTAL
"LISANDRO ALVARADO"

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas

"Inferencia Estadística Bajo el Enfoque Bayesiano"

Trabajo Especial de Grado presentado por

Br. Freddy Campo

como requisito final
para obtener el título de Licenciado
en Ciencias Matemáticas
Área de Conocimiento: Probabilidad y Estadística.
Tutor: MCs. Luz Rodriguez.

Barquisimeto, Venezuela. Junio de 2014

Universidad Centroccidental
"Lisandro Alvarado"
Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas

ACTA
TRABAJO ESPECIAL DE GRADO

Los suscritos miembros del Jurado designado por el Jefe del Departamento de Matemáticas del Decanato de Ciencias y Tecnología de la Universidad Centroccidental "Lisandro Alvarado", para examinar y dictar el veredicto sobre el Trabajo Especial de Grado titulado:

"Inferencia Estadística Bajo el Enfoque Bayesiano"

Presentado por el ciudadano Br. Freddy Campo titular de la Cédula de Identidad Nº 16.041.209. Con el propósito de cumplir con el requisito académico final para el otorgamiento del título de Licenciado en Ciencias Matemáticas.
Luego de realizada la Defensa y en los términos que imponen los Lineamientos para el Trabajo Especial de Grado de la Licenciatura en Ciencias Matemáticas, se procedió a discutirlo con el interesado habiéndose emitido el veredicto que a continuación se expresa:

[1] ______________________

Con una calificación de ______________________ puntos.
En fe de lo expuesto firmamos la presente Acta en la Ciudad de Barquisimeto a los _____ días del mes de ______________________ de __________.

______________________	______________________
TUTOR	FIRMA
______________________	______________________
PRINCIPAL	FIRMA
______________________	______________________
PRINCIPAL	FIRMA

OBSERVACIONES:

____

____

____

[1] Aprobado ó Reprobado

A Dios Todopoderoso y a mis Padres

Agradecimientos

A Dios primeramente por darme la vida, la sabiduría y la fortaleza para llevar esta carrera adelante y culminar esta importante etapa de mi vida, brindandome su cobertura especial con su sangre preciosa para librarme de todo mal.

A mi Padres Fredis y Amalia por el apoyo incondicional y siempre creer en mi, dándome ánimo para no desmayar en los momentos dificiles, por estar en constante oración para pedir a Dios por mi, por sus sabios consejos y enseñarme que para alcanzar las metas hay que luchar con todas nuestras fuerzas.

A mi hermanas Rosa Linda y Nancy Elena, por darme palabras de aliento y ayudarme con los quehaceres del hogar mientras yo estaba estudiando, por acompañarme en esas largas noches de estudio.

A mis compañeras desde el inicio de esta nueva etapa, Gabriela Galavis, Gabriela Gonzalez (la prima) y Maria Luisa, por compartir sus conocimientos e información en esas mañanas, tardes e incluso noches de estudios en la universidad y en el apartamento, también por su valiosa e incondicional amistad. Al pana Oramys (papucho) por su ayuda con el Latex y en algunas materias, ya culminando con la carrera.

A todas esas personas que hicieron placentero mi viaje de la preparación profesional, mis amigos de Infomática, Analisis de Sistemas y de Civil. A esos colegas que de alguna manera dieron su apoyo.

A mi muy querida Lic. Dasha por darme palabras de apoyo, y ayudarme de forma incondicional en los momentos mas duros de mi estadía en la universidad. A mi querida Yilvric (chocolatico) por esos enormes favores en tu área de trabajo.

Muy especialmente a mi tutora por aceptarme para trabajar con ella, por tener esa confianza en mi, mucha paciencia ante mis dudas respecto al tema y por siempre motivarme a seguir adelante.

A esta prestigiosa casa de estudio por abrirme sus puertas y darme la oportunidad de prepararme como profesional, a todos los profesores que con su sabiduría y conocimiento contribuyeron con mi formación de cada semestre. Al personal de bienestar estudiantil por estar siempre atento y dispuesto a prestar sus servicios, a todo el personal administrativo y obrero que allí labora.

Índice

Introducción

La inferencia estadística tiene como principal objetivo conocer o inferir para una población, o conjunto total, las características obtenidas a partir de una muestra o subconjunto de aquella población. Tradicionalmente los problemas de inferencia estadística se dividen en problemas de estimación y en prueba de hipótesis, ambos problemas están relacionados con la teoría de la decisión y sus cuestiones fueron formalizadas por la estadística matemática. La diferencia fundamental entre estos dos problemas está en su orientación. Mientras que en los problemas de estimación el objetivo es determinar el valor de un parámetro desconocido de una determinada distribución, en las prueba de hipótesis se debe decidir si se acepta o rechaza un valor especifico de un parámetro desconocido. Es importante señalar que la prueba de hipótesis puede estar relacionada no solo con el parámetro sino con el tipo o la naturaleza de la distribución.

El origen de los estudios relacionados con la prueba de hipótesis estadísticas, se puede encontrar en investigaciones realizadas por el físico y matemático Daniel Bernoulli. El autor recurre a la estadística para comprobar unas hipótesis en el campo de la astronomía. Este sería un simple apunte que se desarrollaría más tarde por tres autores: Ronald Fisher, Jerzy Neyman y Egon Pearson. La formulación de hipótesis estadísticas surge ante la necesidad de realizar análisis, con objetividad suficiente en las diferentes decisiones que se tomaban. Los trabajos desarrollados en esta línea, por los tres autores señalados, se agrupan en dos tendencias diferentes: Fisher por un lado y Neyman y Pearson por el otro. Fisher desarrolló su teoría que denominó pruebas de significación y Neyman y Pearson las llamadas prueba de hipótesis. La principal diferencia entre estas dos teorías está en que el modelo de Fisher no se acepta la idea de dos hipótesis estadísticas contrapuestas. Es importante señalar que si analizamos la evolución de la prueba de hipótesis estadísticas, podemos decir que el origen de ambas tendencias se encuentra en la prueba Ji-cuadrado de Karl Pearson, pero el paso del tiempo ha generado un modelo mixto que tiene como base las dos teorías.

La estadística Bayesiana es un término aplicado al cuerpo de las técnicas inferenciales que usan el teorema de Bayes para combinar los datos observados con opiniones subjetivas o personales. El proceso inferencial se puede resumir en la verificación de algunas afirmaciones o conjeturas acerca de un parámetro desconocido o una distribución de probabilidad asociada a un conjunto de datos. Por ejemplo, uno puede estar interesado en verificar si una moneda es justa, si una colección de cantidades son independientes o si distintas poblaciones son probabilísticamente iguales. Cada una de las afirmaciones anteriores constituye una hipótesis y pueden ser asociadas a un modelo. Esto acá significa que se puede parametrizar de alguna forma.

Típicamente, una prueba de hipótesis es un problema de decisión con un número de posibles acciones. Si el investigador toma la decisión equivocada incurre en una pena o sufre una pérdida. El objetivo es minimizar su pérdida de alguna manera. Por ejemplo, bajo el enfoque Bayesiano se tratará de minimizar la pérdida esperada. Una regla para decidir si la hipótesis es aceptada es el llamado procedimiento de ensayo o simplemente una prueba y se denotara por ψ. Uno puede definir por ejemplo que $\psi = i$ si la hipótesis aceptada es H_i. Desde la perspectiva Bayesiana, uno puede tener varias hipótesis alternativas $H_1, \cdots, H_k$ que pueden ser comparadas a través de su respectiva posteriori, $p(H_i|x)$, $i = 1, \cdots, k$.

En este trabajo, desarrollamos los procedimientos inferenciales clásicos y Bayesianos, estableciendo una conexión entre pruebas de hipótesis e intervalo de confianza, describiendo las pruebas basadas en los métodos y técnicas empleadas en el tratamiento de los fundamentos estadísticos matemáticos del problema planteado, tanto desde la perspectiva clásica como la Bayesiana.

Capítulo 1

Preliminares

El objetivo de este capítulo es dar una breve presentación de algunos conceptos y resultados auxiliares que nos permiten establecer las condiciones mínimas necesarias para una buena comprensión del trabajo.

§1.1. Variables aleatorias (v.a.)

La relación entre los sucesos del espacio muestral y el valor numérico que se les asigna se establece a través de variables aleatorias.

Definición 1.1. *Un evento, es un subconjunto de un espacio muestral, es decir, un conjunto de posibles resultados que se pueden dar en un experimento aleatorio. Son aquellos hechos en los que no se sabe con certeza lo que va a suceder, dependen del azar y no se puede determinar sus resultados aun repitiéndolo en varias ocasiones.*

Definición 1.2. *Un espacio muestral, consiste en el conjunto de todos los posibles resultados individuales de un experimento aleatorio. Se simboliza con la letra E y los elementos que lo forman se escriben entre llaves.*

Definición 1.3. *Una v.a. es una función que asigna un valor numérico a cada suceso elemental del espacio muestral.*

Es decir, una v.a. es una variable cuyo valor numérico está determinado por el resultado del experimento aleatorio. La v.a. la notaremos con letras mayúsculas X, Y, ... y *con las letras minúscula x, y, ... sus valores.*

La v.a. puede tomar una cantidad numerable o no numerable de valores, dando lugar a dos tipos de v.a.: discretas y continuas.

Definición 1.4. *Se dice que una variable aleatoria X es discreta si puede tomar un número finito o infinito, pero numerable, de posibles valores.*

Definición 1.5. *Se dice que una variable aleatoria X es continua si puede tomar un número infinito (no numerable) de valores, o bien, si puede tomar un número infinito de valores correspondientes a los puntos de uno o más intervalos de la recta real.*

§1.2. Función de distribución acumulada (Fda)

Sean X, Y dos variables aleatorias definidas conjuntamente, es decir; X e Y tienen una distribución de probabilidad conjunta cuya Función de distribución acumulada (Fda) conjunta está dada por:

$$F(x, y) = P\{X \leq x, Y \leq y\}.$$

De manera general, cuando $X' = (X_1, ..., X_p)$ es un vector de variable aleatorias que son distribuidas conjuntamente la Fda está dada por:

$$F(x) = F(x_1, ..., x_p) = P\{X_1 \leq x_1, ..., X_p \leq x_p\}.$$

§1.3. Densidad

Supongamos que $F(X)$ es continua; entonces de la función de densidad conjunta (fdc) de X es:

$$f(x) = f(x_1, ..., x_p) = \frac{\partial^p F(x)}{\partial x_1 ... \partial x_p} \tag{1.1}$$

Análogo al caso univariado, esto es una relación para la probabilidad de un evento (o conjunto de valores en el espacio de dimensión p) en términos de la densidad conjunta para $X : p \times 1$

$$P\{X \subseteq R\} = \int_R ... \int f(x)dx \tag{1.2}$$

para una región R.

§1.4. Distribución marginal

En el análisis de datos multivariado , es típico comenzar con un vector con muchas componentes; y luego, encontrar posteriormente un subvector de interés. En tal caso, la distribución marginal de los subvectores es importante para la inferencia proporcional.

Sea $X' = (Y', Z')$, donde Y y Z son subvectores de $X : p \times 1$. [por ejemplo, $Y' \equiv (X_1, X_2), Z' \equiv (X_3, ..., X_p)$] entonces, si $g(y), h(z)$ denotan las densidades de Y, Z respectivamente, y si $f(x) = f(y, z)$ denota la densidad de X se tiene que:

$$g(y) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(y, z)dz \tag{1.3}$$

y

$$h(z) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(y, z)dy \tag{1.4}$$

donde todas las integrales son tomadas sobre $(\infty, -\infty), g(y)$ y $h(z)$ son llamadas las densidades marginales de Y y Z.

§1.5. Distribución condicional

La distribución condicional es de interés, y ocurre cuando un grupo de variables aleatorias están siendo estudiadas mientras un segundo grupo se mantiene fijo.

Sean A y B dos eventos que pueden ocurrir en un espacio de 2-dimensiones, entonces por definición, la probabilidad condicional de A dado B ésta dada por:

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

si $P(A) \neq 0$. Si A es un evento donde la variable aleatoria X está en el intervalo $a \leq X \leq b$, y B es un evento donde la variable aleatoria Y esta en el intervalo $c \leq Y \leq d$, entonces

$$P\{c \leq Y \leq d | a \leq X \leq b\} = \frac{P\{a \leq X \leq b, c \leq Y \leq d\}}{P\{a \leq X \leq b\}}$$

y por tanto

$$P\{c \leq Y \leq d | a \leq X \leq b\} = \frac{\int_c^d \int_a^b f(x,y)dxdy}{\int_a^b g(x)dx}$$

donde $f(x,y)$ es la densidad conjunta de X, Y y $g(x)$ es la densidad marginal de X.

La densidad condicional de Y dado $X = x$ está definida como:

$$h(y|x) = \frac{f(x,y)}{g(x)}. \tag{1.5}$$

Así,

$$P\{c \leq Y \leq d | X = x\}\Lambda = \int_c^d h(y|x)dy. \tag{1.6}$$

Generalizado a una dimensión p, sean $X' = (X_1, ..., X_p), Y' = (X_1, ..., X_k)$ y $Z' = (X_{k+1}, ..., X_p)$ los vectores aleatorios y con letra minúscula denotaremos los valores observados. La densidad condicional de Y dado Z está dada por:

$$g(Y|Z) = \frac{f(y,z)}{h(z)} = \frac{f(x)}{h(z)} \tag{1.7}$$

donde $f(x)$ denota la densidad del vector aleatorio X, y $h(z)$ denota la densidad marginal del vector Z.

§1.6. Prueba de hipótesis

Definición 1.6. *Una Hipótesis Estadística (H) es una afirmación o una conjetura acerca de la distribución de una ó mas V.A.*

§1.6.1. Tipos de hipótesis

- Por la forma de la distribución de probabilidad de una determinada población.
- Por el valor del parámetro o parámetros desconocidos de una población.

Definición 1.7. *El espacio de parámetros Θ, es el conjunto de todos los posibles valores del parámetro deconocido θ sobre el cual vamos a plantear las hipótesis. Cuando planteamos una hipótesis estadística sobre el parámetro de una población, se particiona Θ es dos subconjuntos Θ_0 y Θ_1.*

Además también se requiere de los siguientes elementos:

1) La hipótesis nula (H_0). Así, $H_0 : \theta \in \Theta_0$

2) La hipótesis alternativa (H_1). Así, $H_1 : \theta \in \Theta_1$

3) El estadístico de prueba.

4) La región de rechazo.

Definición 1.8. *Una hipótesis estadística es una hipótesis simple, si el subespacio correspondiente contiene un sólo punto. En caso contrario diremos que la hipótesis es compuesta.*

Una vez planteadas las hipótesis H_0 y H_1, se debe decidir de acuerdo con los resultados de una muestra que decisión tomar. Para ello, se escoge una estratégia o función de decisión que lleve a rechazar o no la hipótesis nula. Si denotamos por S el espacio muestral, se tiene que la función de decisión divide al espacio muestral (S) en dos subconjuntos:

$$S_0 = \{x \in S | x \quad \text{lleva a aceptar a} \quad H_0\} \quad \text{y} \quad S_1 = \{x \in S | x \quad \text{lleva a rechazar a} \quad H_0\}$$

Si $x_1, ..., x_n$ son los valores observados de una m.a. de tamaño n entonces $x = (x_1, ..., x_n)$ es un punto del espacio muestral n-dimensional de un experimento. Por tanto, las v.a. de la muestra se resumen en un estadístico $\widehat{\theta} = g(x_1, ..., x_n) = g(x)$.

Así;

- Cuando $\widehat{\theta} \in S_0$ se acepta H_0.
- Cuando $\widehat{\theta} \in S_1$ se rechaza H_0.

Definición 1.9. *Una prueba o test de una hipótesis estadística, es una función de decisión o estratégia a seguir para decidir si rechazamos o no hipótesis nula.*

Cuando una hipótesis estadística es rechazada o aceptada en base a una m.a., siempre existe la posibilidad de tomar una decisión equivocada. Por ello, existen dos tipos de errores que se pueden cometer:

- **Error tipo I:** Se comete cuando se rechaza H_0 siendo verdadera y se denota;

$$\alpha = \mathrm{P}(\text{error tipo I}) = P(\widehat{\theta} \in S_1 | \theta \in \Theta_0).$$

- **Error tipo II:** Se comente cuando se acepta H_0 siendo falsa y se denota;

$$\beta = \mathrm{P}(\text{error tipo II}) = P(\widehat{\theta} \in S_0 | \theta \in \Theta_1).$$

Observación 1.1. *La probabilidad α se conoce también como nivel de significación del contraste.*

Se pueden escoger muchas funciones de decisión o contrastes, el mejor se podría definir como aquel que minimiza simultáneamente α y β. Debido a la imposibilidad de que una decisión minimice α y β a la vez, surge otro principio para elegir la decisión más adecuada.

Fijemos α, es decir, la probabilidad máxima de cometer un error de tipo I que se esté dispuesto a aceptar, y entre todos los contrastes que tengan ese valor α se elige aquel que minimice β o que se maximice $1 - \beta, \forall \theta \in \Theta_1$. Al fijar α, se está fijando el tamaño de la región de rechazo, el cual se presenta en tres casos:

- Toda la región de rechazo (RR) queda en la parte derecha del espacio muestral. Además, esta corresponde al contraste de hipótesis de una cola por la derecha.

  $$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0.$$

  Si $\widehat{\theta}$ es el estadístico de la prueba entonces:

  $$RR = \{\widehat{\theta} > k\}$$

  donde k es tal que $\alpha = P(\widehat{\theta} > k)$.

- Toda la región de rechazo (RR) queda en la parte izquierda del espacio muestral. Además, esta corresponde al contraste de hipótesis de una cola por la izquierda.

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta < \theta_0.$$

Si $\widehat{\theta}$ es el estadístico de la prueba entonces:

$$RR = \{\widehat{\theta} < k\}$$

donde k es tal que $\alpha = P(\widehat{\theta} < k)$.

- La región de rechazo (RR) está parte a la derecha y parte a la izquierda del espacio muestral. Además, ésta corresponde al contraste de hipótesis de dos colas.

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0.$$

Si $\widehat{\theta}$ es el estadístico de la prueba entonces:

$$RR = \{\widehat{\theta} < k_1 \quad \text{o} \quad \widehat{\theta} > k_2\}$$

donde k_1 y k_2 son tales que $\alpha = P(\{\widehat{\theta} < k_1 \quad \text{o} \quad \widehat{\theta} > k_2\})$.

§1.7. Función de potencia

Definición 1.10. *La función de potencia de un contraste o test se define como:*

$$\pi(\theta) = P(\textit{rechazar} \quad H_0|\theta), \forall\theta \in \Theta.$$

Definición 1.11. *Un contraste que maximice la función $\pi(\theta), \forall\theta$, se conoce como el test uniformemente más potente (U.M.P.).*

Capítulo 2

Inferencia Bayesiana

§2.1. Teorema de Bayes, distribuciones a priori y posteriori

Sean X, Θ vectores aleatorios de dimensión p y k respectivamente, distribuidos conjuntamente con densidad condicional de X dado Θ denotada por $f(x|\theta)$, y la densidad marginal de Θ dada por $g(\theta)$; los datos han sido generados al observar X, para algún Θ fijo no-observados. Nos gustaría hacer inferencia sobre Θ considerando tanto nuestro prejuicio (creencia priori), como también las observaciones de $(X|\Theta)$ que indirectamente relacionan a ésta. El teorema de bayes proporciona un mecanismo formal para llevar a cabo esto. En terminología bayesiana $g(\theta)$ es llamada la priori de Θ o densidad a priori Θ ya que esta es la densidad de Θ priori a los datos observados.

Teorema 2.1. *(Bayes) Sean $g_1(\theta)$ la densidad a priori de Θ, y $f(x|\theta)$ la densidad condicional de X dado θ, entonces la densidad de Θ dado X=x, es dada por:*

$$h(\theta|x) = \frac{f(x|\theta)g_1(\theta)}{\int_{-\infty}^{+\infty} f(x|\theta)g_1(\theta)d\theta}. \tag{2.1}$$

Observemos que en (2.1) la integral depende solo de x y éste es fijo y conocido, así dicha integral es constante. Por lo tanto:

$$h(\theta|x) \propto f(x|\theta)g_1(\theta).$$

Una distribución para $\Theta : k \times 1$ posee todas la propiedades habituales de distribuciones de variables aleatorias observadas, excepto que Θ no es observada. Este tipo de distribuciones son llamadas distribuciones de probabilidad subjetiva.

Ejemplo 2.1. *Se desea estimar la probabilidad, θ, de un evento, a parti del resultado de una sucesión de n ensayos Bernoulli, esto es, datos $x_1, x_2, ..., x_n$ que son iguales a uno si ocurre el evento (éxito) y cero si no ocurre. Sea x el número total de éxitos en la muestra de n ensayos. En este caso, el modelo muestral establece que:*

$$f(x|\theta) = Bin(x|n, \theta) = \begin{pmatrix} n \\ x \end{pmatrix} \theta^x (1-\theta)^{n-x}.$$

Suponiendo que $g_1(\theta)$ es uniforme en el intervalo [0,1], se tiene que:

$$h(\theta|x) \propto \theta^x (1-\theta)^{n-x}$$

o lo que es lo mismo, la distribución no normalizada tiene un kernel equivalente a la distribución Beta (β), es decir:

$$\theta|x \sim \beta(x+1, n-x+1).$$

§2.2. Densidad general de la distribución Normal multivariada

Sea $X : p \times 1$ un vector aleatorio con función de densidad $f(x)$. X tiene una distribución Normal multivariada (p variada) no-singular con vector de media $\theta : p \times 1$ y matriz covarianza $\Sigma : p \times p$ si:

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp\left[-\frac{1}{2}(x-\theta)'\Sigma^{-1}(x-\theta)\right] \tag{2.2}$$

para $\Sigma > 0$. Si $|\Sigma| = 0$, la distribución de X es llamada singular o normal degenerada y la densidad no existe.
Denotaremos esta distribución de X por:

$$\mathcal{L}(X) = N(\theta, \Sigma).$$

§2.2.1. Estandarización

Si $\mathcal{L}(X) = N(\theta, \Sigma)$, la distribución de X puede ser estandarizada por la transformación $Y = \Sigma^{-1/2}(X - \theta)$; es decir, $\mathcal{L}(Y) = N(0, I)$.

Observemos que la transformación $Y = \Sigma^{-1/2}(X - \theta)$ es inyectiva y diferenciable, la inversa existe, es diferenciable y es dada por $Y\Sigma^{1/2} + \theta = X$; además
$J(X \to Y) = |\Sigma|^{1/2}$.
Luego,

$$f(y) = \frac{1}{(2\pi)^{p/2}} exp\left[-\frac{1}{2}y'y\right]$$

$\therefore \mathcal{L}(Y) = N(0, I)$.

§2.2.2. Distribución normal bivariada

Sea $X : 2 \times 1$ un vector aleatorio bivariado con $\mathcal{L}(X) = N(\theta, \Sigma)$ y $\Sigma > 0$. Sean $\theta = (\theta_i)$ y $\Sigma = (\sigma_{ij})$ para $i, j = 1, 2$. Para simplificar tomemos $\sigma_{11} = \sigma_1^2, \sigma_{22} = \sigma_2^2$ y $\sigma_{12} = \rho\sigma_1\sigma_2$ donde ρ es el coeficiente de correlación entre X_1 y X_2. Si escribimos (2.2) para $p = 2$ tenemos que la densidad bivariada es dada por:

$$\begin{aligned} f(x) &= f(x_1, x_2) \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\times \; exp\left\{-\frac{1}{2(1-\rho^2)}\left[(\frac{x_1-\theta_1}{\sigma_1})^2 - 2\rho(\frac{x_1-\theta_1}{\sigma_1})(\frac{x_2-\theta_2}{\sigma_2}) + (\frac{x_2-\theta_2}{\sigma_2})^2\right]\right\}. \end{aligned} \tag{2.3}$$

Acá

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \qquad \Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2(1-\rho^2)} & -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} \\ -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} & \frac{1}{\sigma_2^2(1-\rho^2)} \end{pmatrix}.$$

La expresión entre corchetes de (2.3) controla la varianza de $f(x)$, es decir; si la expresión entre corchetes es constante entonces $f(x)$ es constante y viceversa. La expresión (2.3) muestra que si $f(x)$ es constante entonces los puntos (x_1, x_2) están a lo largo de la elipse concéntrica con centro (θ_1, θ_2) con un menor y mayor ángulo respecto a los eje (x_1, x_2), los ángulos depende de los valores $(\rho, \sigma_1, \sigma_2)$.

§2.2.3. Independencia

Sea $\mathcal{L}(X) = N(\theta, \Sigma)$ y $X : 2 \times 1$. Entonces, si $\rho = 0$ en (2.3), X_1 y X_2 no sólo están no-correlacionados, también son independientes. Es fácil ver que al sustituir $\rho = 0$

en (2.3) $f(x_1, x_2)$ se reduce al producto de una función de x_1 y una función de x_2. Por supuesto, lo contrario es también cierto, es decir; sí X_1 y X_2 son independientes, entonces X_1 y X_2 son no-correlacionados; en este sentido, el resultado se cumple para toda distribución bivariada (mientras que en el otro sentido, la falta de correlación generalmente no implica la independencia, aunque si para la distribución normal).

§2.3. Principio de verosimilitud

En el caso en que ”y” se conozca o se considere fijo, $p(y|\theta)$, como función de θ, se le conoce como la *función de verosimilitud.*

La inferencia Bayesiana obedece el principio de verosimilitud, es decir, para una muestra de datos, dos modelos $p(y|\theta)$ con la misma función de verosimilitud producen la misma inferencia sobre θ, o dicho de otra manera, la función de verosimilitud contiene toda la información relevante que aportan los datos. Usar el principio de verosimilitud es aceptar que la inferencia está *condicionada* a los datos observados, pues la verosimilitud está parametrizada por los datos. Esto contrasta con la inferencia basada en las distribuciones de muestreo, donde se considera un estimador $\widehat{\theta} = f(y)$, el cual, de acuerdo con el tipo de experimento de muestreo, tiene una distribución muestral que resume las propiedades de estimador previo a la observación de los datos y por tanto, irrelevante para hacer inferencias luego que se han observado los datos.

§2.4. Distribución a priori no-informativa

Supongamos que $\mathcal{L}(X|\Theta, \Sigma) = N(\theta, \Sigma)$. El problema de la subjetividad es para estimar una distribución a priori para (θ, Σ). Primero consideramos θ_j, una componente de θ *donde* $-\infty < \theta_j < \infty$, $j = 1, ..., p$.

Una manera de expresar la información vaga es usando la noción de que todos los valores de las variables aleatorias, pueden estar sobre la recta real, las cuales son igualmente probables. Así suponemos que la distribución de una variable aleatoria es uniforme en el intervalo (a, b), donde a y b son valores muy grandes negativo y positivo respectivamente. Sabemos que la distribución a posteriori $h(\theta|x) \propto f(x|\theta)g(\theta)$, es dada por el producto de la densidad a priori y la densidad condicional de la variable aleatoria observadas (función de verosimilitud) es importante que para determinar la distribución posteriori, solamente es necesario que el intervalo (a, b) extendido sobre la región en la cual la función de verosimilitud es apreciablemente distinta de cero.

Por ejemplo, cuando la función de verosimilitud está basada en la distribución Normal, una distribución a priori uniforme está definida en un rango más o menos ubicado entre las 3 desviaciones estándar de la media en la distribución Normal. Fuera de este rango, las probabilidades a priori del parámetro no son importantes ya que éstas se obtienen multiplicando las colas de la función de densidad Normal, y por consiguiente no afecta significativamente la inferencia basada en la distribución a posteriori.

§2.5. Conjugada a priori natural

Si la densidad a priori es fácil de calcular, ésta puede ser hallada por medio del problema de una densidad a priori informativa. Una forma, para hallar la densidad a priori es usando la conjugada natural.

La idea básica es escribir la densidad como la función de verosimilitud para las variables aleatorias observadas. Entonces intercambiando los roles de las variables aleatorias observadas con los parámetros, asumimos que éste último es aleatorio y el anterior es fijo, modificando la proporcionalidad constante apropiadamente para que la nueva "densidad" al integrarla sea igual a uno. En resumen, sea $L(x|\theta)$ la verosimilitud obtenida de la familia F. Si existe una constante k de tal manera que la función sea una función de densidad, $p(\theta) = kL(x|\theta)$, entonces la familia F de las densidades p es la conjugada natural de la familia con respecto a la distribución de muestreo de verosimilitud L.
Las ventajas de usar la densidad de la conjugada a priori natural son:

1) La distribución a posteriori pertenece a la misma familia que la de la priori.

2) La distribución a posteriori es fácil de manejar matemáticamente.

Capítulo 3

Prueba de Hipótesis usando métodos Bayesianos

§3.1. Prueba de hipótesis clásica

La teoría general de la prueba de hipótesis viene de la labor pionera de Neyman y Pearson (1928). La característica de la probabilidad de una prueba clásica puede describirse mediante la especificación $\pi(\theta)$, la probabilidad de que la prueba conduzca al rechazo de H_0, para cada valor de $\theta \in \Theta$. La función de π es llamada la potencia de la prueba. Si C es la región crítica entonces π se define por:

$$\pi(\theta) = P(X \in C), \forall \theta \in \Theta.$$

El tamaño o nivel de significancia α de un procedimiento de prueba es definido por:

$$\alpha \geq \sup_{\theta \in \Theta} \pi(\theta).$$

Al igual que en el caso de nivel de confianza, la desigualdad anterior es un requisito técnico. Es más útil en espacios muestrales discretos donde no todos los valores de $[0, 1]$ son las probabilidades posibles. Como se verá en breve, se quiere utilizar un valor de α tan pequeño como sea posible. En la práctica, esto significa que se usa una igualdad.

§3.1.1. Hipótesis simple

Es útil para comenzar el estudio de la teoría con el caso de dos hipótesis simples $H_0 : \theta = \theta_0$ y $H_1 : \theta = \theta_1$. Preferiblemente, se desea hallar un procedimiento de prueba para la cual los dos errores probabilísticos sean tan pequeños como sea posible. En la práctica, es imposible hallar una prueba para la cual estas probabilidades sean minimizadas simultáneamente. Como alternativa, uno puede tratar de construir una prueba que minimice combinaciones lineales de α y β.

Teorema 3.1 (Prueba Óptima)**.** *Supongamos que* $X = (x_1, ..., x_n)$ *es una muestra aleatoria de* $p(x|\theta)$, $H_0 : \theta = \theta_0$ *y* $H_1 : \theta = \theta_1$. *Sea* ψ^* *una prueba de* H_0 *contra* H_1 *tal que* H_0 *es aceptada si* $p_0/p_1 > k$ *y* H_0 *es rechazada si* $p_0/p_1 < k$, *donde* $p_i = p(x|\theta_i), i = 0, 1$ *y* $k > 0$. *(Si* $p_0/p_1 = k$ *no se concluye nada). Entonces, cualquier otra prueba* ψ *será tal que:*

$$a\alpha(\psi^*) + b\beta(\psi^*) \leq a\alpha(\psi) + b\beta(\psi),$$

donde $\alpha(\psi)$ *y* $\beta(\psi)$ *respectivamente denotan las probabilidades de los errores de tipo I y II de la prueba* ψ, *para algún* $a, b \in \mathbb{R}^+$.

Demostración. Sea C la región crítica de una prueba arbitraria ψ y definamos $p_i = p(x|\theta_i), i = 0, 1$. Entonces, para $a, b \in \mathbb{R}^+$ se tiene:

$$\begin{aligned} a\alpha(\psi) + b\beta(\psi) &= a\int_C p(x|\theta_0)\,dx + b\int_{\overline{C}} p(x|\theta_1)\,dx \\ &= a\int_C p(x|\theta_0)\,dx + b[1 - \int_C p(x|\theta_1)\,dx] \\ &= b + \int_C (ap_0 - bp_1)\,dx. \end{aligned}$$

Así, la minimización de $a\alpha(\psi) + b\beta(\psi)$ es equivalente a elegir la región crítica C de manera que el valor de la integral sea mínimo. Esto ocurre si la integración se realiza a través de un conjunto que incluya todos los puntos x tal que $ap_0 - bp_1 < 0$ y no icluya los puntos x tal que $ap_0 - bp_1 > 0$. Por lo tanto, la minimización de $a\alpha(\psi) + b\beta(\psi)$ se lleva a cabo teniendo la región crítica C que incluye sólo los puntos x tal que $ap_0 - bp_1 < 0$ (si la distribución muestral es continua y $ap_0 - bp_1 = 0$, este punto no contribuye). Esto completa la demostración porque $ap_0 - bp_1 < 0$ si y sólo si $p_0/p_1 < k = b/a$, la cual corresponde a la descripción de la prueba ψ^*. □

Observación 3.1. *La relación* p_0/p_1 *se llama la razón de verosimilitud. El teorema establece que una prueba que minimiza* $a\alpha(\psi) + b\beta(\psi)$ *rechaza* H_0 *cuando la razón de* *verosimilitud es pequeña y acepta H_0 cuando la razón de verosimilitud es grande. Por lo general, la hipótesis nula H_0 y el error de tipo I son privilegiados. Por tanto, se considera sólo las pruebas ψ tal que $\alpha(\psi)$ no puede ser más grande que un nivel α_0 preespecificado, y entre ellos buscar algunos que minimicen $\beta(\psi)$.*

Lema 3.2 (Neyman - Pearson)**.** *Supongamos que $X = (x_1, ..., x_n)$ es una muestra aleatoria de $p(x|\theta), H_0 : \theta = \theta_0$ y $H_1 : \theta = \theta_1$. Sea ψ^* una prueba de H_0 contra H_1 tal que H_0 es aceptada si $p_0/p_1 > k$ y H_0 es rechazada si $p_0/p_1 < k$, donde $p_i = p(x|\theta_i), i = 0, 1$ y $k > 0$. (Si $p_0/p_1 = k$ no se concluye nada). Entonces para alguna prueba: $\alpha(\psi) \leq \alpha(\psi^*) \Rightarrow \beta(\psi) \geq \beta(\psi^*)$, también se cumple que $\alpha(\psi) < \alpha(\psi^*) \Rightarrow \beta(\psi) > \beta(\psi^*)$.*

Demostración. Siguiendo la definición de la prueba optima ψ^* en el teorema, se deduce alguna otra prueba ψ dada por:

$$\alpha(\psi^*) + k\beta(\psi^*) \leq \alpha(\psi) + k\beta(\psi) \quad \text{para} \quad k > 0,$$

donde;

$$\begin{aligned} \alpha(\psi) \leq \alpha(\psi^*) \quad &\Rightarrow \quad \alpha(\psi) + k\beta(\psi^*) \leq \alpha(\psi^*) + k\beta(\psi^*) \leq \alpha(\psi) + k\beta(\psi) \\ &\Rightarrow \quad k\beta(\psi^*) \leq k\beta(\psi) \\ &\Rightarrow \quad \beta(\psi^*) \leq \beta(\psi). \end{aligned}$$

También, si $\alpha(\psi) < \alpha(\psi^*)$ entonces $\beta(\psi^*) < \beta(\psi)$, lo que completa la demostración. □

Observación 3.2. *El lema (3.2) sólo toma en cuenta la aceptación ó rechazo de H_0 sin referencia a H_1, α_0 juega el rol de un nivel de significación. Recordando que $\pi(\theta_1) = 1 - \beta(\psi)$, la minimización de β implica la maximización de π. Por lo tanto, el lema de Neyman - Pearson muestra que todas las pruebas con un nivel de significancia dado, que se basa en la razon de verosimilitud son más grandes y más potentes.*

Ejemplo 3.1. *Sea X una v. a. para $N(\theta, \sigma^2)$ con σ^2 conocido y $x_1, x_2, ..., x_n$ una m. a. de $N(\theta, \sigma^2)$, consideremos la prueba $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, con $\theta_0 < \theta_1$, entonces:*

$$\begin{aligned}
\frac{p_0}{p_1} = \frac{p(X|\theta_0)}{p(X|\theta_1)} &= \frac{(2\pi\sigma)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta_0)^2\right\}}{(2\pi\sigma)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta_1)^2\right\}} \\
&= \exp\left\{\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta_0)^2\right)-\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta_1)^2\right)\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta_0)^2+\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta_1)^2\right\} \\
&= \exp\left\{\frac{1}{2\sigma^2}\left[-\sum_{i=1}^{n}(x_i^2-2x_i\theta_0+\theta_0^2)+\sum_{i=1}^{n}(x_i^2-2x_i\theta_1+\theta_1^2)\right]\right\} \\
&= \exp\left\{\frac{1}{2\sigma^2}\left[2\theta_0\sum_{i=1}^{n}x_i-n\theta_0^2-2\theta_1\sum_{i=1}^{n}x_i+n\theta_1^2\right]\right\} \\
&= \exp\left\{\frac{1}{2\sigma^2}\left[2(\theta_0-\theta_1)\sum_{i=1}^{n}x_i+n(\theta_1^2-\theta_0^2)\right]\right\} \\
&= \exp\left\{\frac{1}{2\sigma^2}\left(2(\theta_0-\theta_1)\sum_{i=1}^{n}x_i\right)\right\}\exp\left\{\frac{1}{2\sigma^2}n(\theta_1^2-\theta_0^2)\right\} \\
&\propto \exp\left\{\frac{n(\theta_0-\theta_1)\overline{x}}{\sigma^2}\right\},
\end{aligned}$$

donde la constante de proporcionalidad depende de las constantes θ_0, θ_1 y σ^2.

La prueba de razón de verosimilitud acepta H_0 cuando $p_0/p_1 > k$, entonces:

$$\frac{p_0}{p_1} > k \Leftrightarrow \exp\left\{\frac{n(\theta_0-\theta_1)\overline{x}}{\sigma^2}\right\} > c_3 \Leftrightarrow \frac{n(\theta_0-\theta_1)\overline{x}}{\sigma^2} > c_2 \Leftrightarrow \overline{x} < c_1,$$

ya que $\theta_0 < \theta_1$, para constantes c_1, c_2, c_3. Como el mejor estimador de θ es $\overline{X}$, la media muestral, cuando se prueba H_0 contra H_1, uno espera que la prueba acepte H_0 para valores pequeños de $\overline{X}$. Este es exactamente el resultado de la prueba óptima, para la razón de verosimilitud. El siguiente paso es determinar el valor de c_1. Para ello, observemos que la prueba tiene un nivel α y por tanto:

$$\alpha = P(\text{rechazo de } H_0|H_0 \text{ es verdadera}) = P(\overline{X} > c_1|\theta = \theta_0).$$

Pero,

$$\overline{X}|\theta_0 \sim N\left(\theta_0, \tfrac{\sigma^2}{n}\right) \qquad \text{ó} \qquad z = \tfrac{\overline{X}-\theta_0}{\sigma/\sqrt{n}} \sim N(0,1).$$

Notemos que,

$$\alpha = P(\overline{X} > c_1|\theta = \theta_0) \Rightarrow \alpha = P\left(\frac{\overline{X}-\theta_0}{\sigma/\sqrt{n}} > \frac{c_1-\theta_0}{\sigma/\sqrt{n}}\right) \Rightarrow \alpha = P\left(z > \frac{(c_1-\theta_0)\sqrt{n}}{\sigma}\right),$$

por lo que:

$$\frac{(c_1-\theta_0)\sqrt{n}}{\sigma} = z_\alpha \Rightarrow c_1 = \theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}}.$$

Así, la prueba con nivel de significación α, acepta H_0 si $\overline{X} < \theta_0 + \dfrac{\sigma z_\alpha}{\sqrt{n}}$.

Notemos que la prueba es completamente especificada y no depende del valor de θ_1 por el hecho de que $\theta_1 > \theta_0$. Esto significa que la prueba es la misma para algún valor de θ_1 tal que $\theta_1 > \theta_0$. Por lo tanto, la prueba de razón de máxima verosimilitud es también una prueba más potente para H_0 versus $H_1 : \theta_1 > \theta_0$.

El hecho que la prueba no dependa del valor de θ_1 puede también causar problemas. Consideremos el caso cuando H_0 es rechazado pero $\overline{x}$ está mucho más cerca de θ_0 que de θ_1, esto es:

$$\overline{x} - \theta_0 << \theta_1 - \overline{x}.$$

En este caso, el sentido común sugiere que dadas las elecciones de H_0 y H_1 se puede escoger H_0. Esta es otra razón para evitar compromisos con aceptación o rechazo de H_1 y que la prueba no proporciona información sobre esto. Comentarios similares se aplican si las distancias de $\overline{x}$ a θ_0 y θ_1 son mucho más largas que la distancia entre θ_0 y θ_1. El razonamiento intuitivo basado en el argumento frecuentista es que, si el muestreo de X se repite muchas veces, en solamente $100\alpha\,\%$ de ellos, H_0 puede ser rechazado erroneamente.

La potencia de la prueba $\pi(\theta) =$ P(rechazo de $H_0|\theta$) está dada por:

$$\pi(\theta) = P\left[\overline{X} > \theta_0 + \frac{\theta z_\alpha}{\sqrt{n}}|\theta > \theta_0\right].$$

Pero $\overline{X}|\theta > \theta_0 \sim N\left(\theta, \tfrac{\sigma^2}{n}\right)$ y por lo tanto, $\dfrac{\sqrt{n}(\overline{x}-\theta)}{\sigma} \sim N(0,1)$ entonces:

$$\begin{aligned}
\pi(\theta) &= P\left[\frac{\sqrt{n}(\overline{x}-\theta)}{\sigma} > \frac{\theta_0 + (\sigma/\sqrt{n})z_\alpha - \theta}{\sigma/\sqrt{n}} | \theta > \theta_0\right] \\
&= P\left[\frac{\sqrt{n}(\overline{x}-\theta)}{\sigma} > \frac{\theta_0}{\sigma/\sqrt{n}} + \frac{(\sigma/\sqrt{n})z_\alpha}{\sigma/\sqrt{n}} - \frac{\theta}{\sigma/\sqrt{n}} | \theta > \theta_0\right] \\
&= P\left[\frac{\sqrt{n}(\overline{x}-\theta)}{\sigma} > \frac{(\theta_0-\theta)\sqrt{n}}{\sigma} + z_\alpha | \theta > \theta_0\right] \\
&= 1 - \Phi\left(z_\alpha - \frac{(\theta-\theta_0)\sqrt{n}}{\sigma}\right),
\end{aligned}$$

la cuál es una función creciente de θ. Por lo tanto, el más distante es el valor del parámetro en la hipótesis alternativa, la más pequeña son las probabilidades de un error de tipo II.

Esta prueba no es sólo la prueba más potente de H_0 versus $H_1 : \theta > \theta_0$, también la prueba $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_1$, porque el nivel de la prueba con la nueva hipótesis es:

$$\max_{\theta \leq \theta_0} \pi(\theta) = \max_{\theta \leq \theta_0} P\left[\overline{x} > \theta_0 + \frac{\sigma z_\alpha}{\sqrt{n}}\right].$$

Como vemos, π es una función creciente de θ y el máximo en la región $\{\theta : \theta \leq \theta_0\}$ está dado por el valor de θ_0 y el valor de π en este punto es $1 - \Phi(z_\alpha) = \alpha$. También, en el ejemplo se puede evaluar el tamaño en que se debe rechazar H_0 luego de observar $X = x$, esto es, evaluar γ tal que $\theta_0 - \frac{\sigma z_\gamma}{\sqrt{n}} = \overline{X}$.

Esta da una descripción más precisa de la fuerza de la evidencia de los datos a favor o en contra de la hipótesis. Pequeños valores de γ indica poca probabilidad ó un error de tipo I y por tanto más evidencia a favor de H_0. Igualmente grandes valores de γ indica una alta probabilidad ó un error de tipo I y por tanto más evidencia contra H_0.

En general, supongamos que se tiene una prueba donde H_0 es rechazada cuando el estadístico de la prueba T pertenece a la región de la forma $[T > c]$ y sea t el valor observado de T. Entonces, la evaluación de $Pr(T > t|H_0)$ da una idea de como el extremo del valor observado está bajo H_0. Esta probabilidad es usualmente conocida como el $p-valor$. En el ejemplo anterior, el $p-valor$ está dado por $1-\Phi\left(\frac{\sqrt{n}(\overline{x}-\theta_0)}{\sigma}\right)$.

La noción de un $p - valor$ es útil para determinar la talla en la que se puede rechazar H_0 basada en la información actualmente obtenida por:

$$H_0 \quad \text{es rechazado si y sólo si} \quad p - valor \leq \alpha,$$

donde α es un nivel pre-especificado de la prueba. Hay que destacar que bajo una perspectiva frecuentista, las probabilidades no pueden ser asociadas a las hipótesis, cuando θ no es aleatorio. Por lo tanto, ninguna asociación entre $p - valor$ y la probabilidad de H_0 puede ser hecha porque una probabilidad simplemente no puede ser definida. La noción del $p - valor$ se puede colocar en un conjunto general siempre que tenga sentido especificar el borde de la región crítica en términos del valor observado en la muestra.

Retornando al caso de la población discreta, no siempre es posible obtener pruebas de algún nivel pre-especificado exactamente. Para ser exacto, significa tener $\alpha = n$ donde:

$$n = \sup_{\theta \in \theta_0} P(\text{rechazar} H_0 | \theta \in \Theta_0).$$

Aquí la noción del $p - valor$ pasa a ser importante.

Hay un enfoque alternativo que permite obtener una prueba de nivel exacto para una distribución discreta. Esta alternativa es conocida como una prueba aleatoria donde un nivel pre-especificado se obtiene después de realizar un experimento de Bernoulli independiente adicional con probabilidad de éxito convenientemente elegido para completar la diferencia entre α y n.

§3.1.2. Hipótesis compuesta

Consideremos de nuevo la prueba ψ de $H_0 : \theta = \Theta_0$ contra $H_1 : \theta = \Theta_1$. Sea α el nivel de significancia fijo y $\pi_\psi(\theta) \leq \alpha$ para cada $\theta \in \Theta_0$.

Definición 3.1. *Una prueba ψ^* es uniformemente más potente (UMP en resumen) para $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$ con un nivel de significancia α si:*

1) $\alpha(\psi^*) \leq \alpha$

2) $\forall \psi$ *con* $\alpha(\psi) \leq \alpha, \pi_\psi(\theta) \leq \pi_{\psi^*}(\theta), \forall \theta \in \Theta_1.$

La prueba del ejemplo anterior es UMP para la prueba $\theta = \theta_0$ versus $\theta > \theta_0$ y también para la prueba $\theta \leq \theta_0$ versus $\theta > \theta_0$.

Teorema 3.3. *Sea $X = (x_1, ..., x_n)$ una muestra aleatoria de $p(x|\theta)$ y $p(x|\theta)$ pertenece a la familia paramétrica exponencial con densidad*

$$p(x|\theta) = a(x)\exp\{\phi(\theta)T(x) + b(\theta)\},$$

y sea ϕ una función de θ estríctamente creciente. Entonces la prueba uniformemente más potente de nivel α de la prueba $H_0 = \theta \leq \theta_0$ contra la prueba $H_1 = \theta > \theta_0$ está dada por la región crítica $T(X) > c$, donde c es tal que $\alpha \geq P(T(X) > c|\theta_0)$, (con igualdad para el caso continuo). La potencia de esta prueba es una función de θ creciente. Si las hipótesis son intercambiadas ó ϕ es una función estríctamente decreciente, entonces la prueba uniformemente más potente de nivel α rechaza H_0 si $T(X) < c$, donde c es tal que $\alpha \geq P(T(X) < c|\theta_0)$ y la potencia de esta prueba es de nuevo una función de θ creciente. Si las dos condiciones mencionadas anteriormente se cumplen simultaneamente, la prueba uniformemente más potente permanece inalterada.

Demostración. Consideremos el caso estándar donde ϕ es estrictamente creciente y las hipótesis son $H_0 : \theta = \theta_0$ y $H_1 : \theta = \theta_1 > \theta_0$. En este caso, el lema de Neyman-Pearson asegura que la prueba más potente rechaza H_0 cuando:

$$\begin{aligned}
\frac{p(x|\theta_0)}{p(x|\theta_1)} < c_4 \quad &\Leftrightarrow \quad \frac{a(x)\exp\{\phi(\theta_0)T(x) + b(\theta_0)\}}{a(x)\exp\{\phi(\theta_1)T(x) + b(\theta_1)\}} < c_4 \\
&\Leftrightarrow \quad \exp\{\phi(\theta_0)T(x) + b(\theta_0) - \phi(\theta_1)T(x) - b(\theta_1)\} < c_4 \\
&\Leftrightarrow \quad \exp\{\phi(\theta_0)T(x) - \phi(\theta_1)T(x) + b(\theta_0) - b(\theta_1)\} < c_4 \\
&\Leftrightarrow \quad \exp\{\phi(\theta_0)T(x) - \phi(\theta_1)T(x)\}\exp\{b(\theta_0) - b(\theta_1)\} < c_4 \\
&\Leftrightarrow \quad \exp\{(\phi(\theta_0) - \phi(\theta_1))T(x)\} < c_3 \\
&\Leftrightarrow \quad \log(\exp\{\phi(\theta_0) - \phi(\theta_1)T(x)\}) < \log(c_3) \\
&\Leftrightarrow \quad (\phi(\theta_0) - \phi(\theta_1))T(x) < c_2 \\
&\Leftrightarrow \quad T(x) > c_1,
\end{aligned}$$

donde c_4, c_3, c_2, c_1 son constantes tales que $\alpha = P(T(X) > c|\theta_0)$. Para la prueba más potente, debe ser verdad que $\pi(\theta_0) \leq \pi(\theta_1)$. Así, la función potencia es una función creciente de θ y

$$\pi(\theta_0) = \sup_{\theta:\theta<\theta_0} \pi(\theta).$$

Así, la prueba es uniformemente más potente para $H_0 : \theta \leq \theta_0$. Como en el cálculo anterior sólo fue usada la condición $\theta_1 > \theta_0$, así los resultados deben ser ciertos para cualquier valor de θ_1. Por lo tanto, la prueba es uniformemente más potente para $H_1 : \theta > \theta_0$.

En el caso de una función ϕ estrictamente decreciente, se tiene

$$(\phi(\theta_0) - \phi(\theta_1))T(x) < c_1 \Leftrightarrow T(x) < c$$

y la región crítica sería $\{x : T(x) < c\}$.

En el caso de hipótesis intercambiadas, todas las desigualdades deben reservarse porque se debe trabajar con el radio $\dfrac{p(x|\theta_1)}{p(x|\theta_0)}$ en lugar de $\dfrac{p(x|\theta_0)}{p(x|\theta_1)}$. Esto conduce a la región crítica de la forma $\{x : T(x) < c\}$.

Finalmente, en el caso de una función ϕ estrictamente decreciente e hipótesis intercambiadas, doble inversión de las desigualdades se conserva como era y la región crítica permanece de la forma $\{x : T(x) < c\}$. $\square$

Ejemplo 3.2. *Sea $x_1, ..., x_n$, una m.a. de la distribución de Bernoulli(θ). Notemos que si $X \sim Ber(\theta)$ entonces $p(x) = \theta^x(1-\theta)^{1-x}$, entonces:*

$$\begin{aligned}
p(x|\theta) &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} I_{x_i}(\{0,1\}) \\
&= \prod_{i=1}^{n} \exp\{x_i \log\theta\} \times \exp\{(1-x_i)\log(1-\theta)\} \times I_{x_i}(\{0,1\}) \\
&= \prod_{i=1}^{n} \exp\{x_i \log\theta + (1-x_i)\log(1-\theta)\} \times I_{x_i}(\{0,1\}) \\
&= \prod_{i=1}^{n} \exp\left\{x_i \log\left(\frac{\theta}{1-\theta}\right) + \log(1-\theta)\right\} \times I_{x_i}(\{0,1\}) \\
&= \exp\left\{\sum_{i=1}^{n} x_i \log\left(\frac{\theta}{1-\theta}\right) + n\log(1-\theta)\right\} \times I_x(\{0,1\}^n).
\end{aligned}$$

Haciendo $a(x) = I_x(\{0,1\}^n), b(\theta) = n\log(1-\theta)$ *y* $\phi(\theta) = \log\left[\frac{\theta}{1-\theta}\right]$ *tenemos que* $T(x) = \sum_{i=1}^{n} x_i$ *es un estadístico suficiente para* θ*, por el criterio de factorización. Por lo tanto, la prueba uniformemente más potente para* $H_0 : \theta \leq \theta_0$ *versus* $H_1 : \theta > \theta_0$ *tiene una región crítica de la forma* $\sum_{i=1}^{n} x_i > c$*.*

Observación 3.3. *La propiedad que garantiza la existencia de una prueba uniformemente más potente en la familia exponencial es de hecho más general. Esta encuentra un lugar apropiado bajo familias con razón de verosimilitud monótona.*

Definición 3.2. *La familia de distribuciones* $\{p(x|\theta), \theta \in \Theta\}$ *se dice que tiene razón de verosimilitud monótona si hay un estadístico* $T(X)$ *tal que* $\forall \theta_1, \theta_2 \in \Theta$ *con* $\theta_1 < \theta_2$ *la razón de verosimilitud*

$$\frac{p(x|\theta_2)}{p(x|\theta_1)}$$

es una función monótona de T(X).

La distribución uniforme del intervalo $[0, \theta]$ no pertene a la familia exponencial. Sin embargo, tiene una razón de verosimilitud monótona porque la razón de las densidades muestrales es una función monótonamente decreciente de $T(X) = \text{máx}_i X_i$.

Los resultados que acabamos de demostrar para la familia exponencial se pueden extender para familias con razón de verosimilitud monótona. Por lo tanto, si la razón de verosimilitud es una función creciente de $T(X)$, entonces la prueba uniformemente más potente de nivel α para $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$ está dada por la región crítica de la forma $T(X) < c$ donde c es tal que $\alpha = P(T(X) < c|\theta_0)$ y la potencia de esta prueba es una función creciente de θ. Así mismo, si las hipótesis son intercambiadas o la razón de verosimilitud es una función decreciente de $T(X)$, la prueba uniformemente más potente de nivel α rechaza H_0 si $T(X) > c$ donde c es tal que $\alpha \geq P(T(X) > c|\theta_0)$ y la potencia de esta prueba es de nuevo una función creciente de θ. Si las dos condiciones descritas anteriormente se cumplen simultáneamente, la prueba uniformemente más potente no cambia.

Estos resultados tienen un sentido intuitivo. Cuanto mayor sea la razón de verosimilitud, es más verosímil el valor de θ_0 con respecto a θ_1. Si la razón de verosimilitud es una función creciente de $T(X)$, el mismo razonamiento es cierto para $T(X)$. Por lo tanto, una región de rechazo razonable para H_0 podría estar dada por valores pequeños de $T(X)$.

Ejemplo 3.3. *Sea $X_1, ..., X_n$ una muestra aleatoria con distribución $Ber(\theta)$. Entonces, del ejemplo 2.2 tenemos $p(X|\theta) = \theta^T(1-\theta)^T$ con $T = \sum_{i=1}^{n} X_i$. Si $\theta_1 < \theta_2$, la razón de verosimilitud es:*

$$
\begin{aligned}
\frac{p_0}{p_1} &= \frac{p(x|H_0)}{p(x|H_1)} = \frac{p(x|\theta_2)}{p(x|\theta_1)} = \frac{\theta_2^T(1-\theta_2)^{n-T}}{\theta_1^T(1-\theta_1)^{n-T}} \\
&= \frac{\theta_2^T(1-\theta_2)^{-T}}{\theta_1^T(1-\theta_2)^{-T}} . \frac{(1-\theta_2)^n}{(1-\theta_1)^n} = \frac{\theta_2^T(1-\theta_1)^T}{\theta_1^T(1-\theta_2)^T} . \left(\frac{1-\theta_2}{1-\theta_1}\right)^n \\
&= \left[\frac{\theta_2(1-\theta_1)}{\theta_1(1-\theta_2)}\right]^T . \left(\frac{1-\theta_2}{1-\theta_1}\right)^n = \xi^T.\eta^n,
\end{aligned}
$$

con

$$
\xi = \frac{\theta_2(1-\theta_1)}{\theta_1(1-\theta_2)} \quad y \quad \eta = \frac{1-\theta_2}{1-\theta_1}.
$$

Por lo tanto, como $\theta_1 < \theta_2 \Rightarrow -\theta_1 > -\theta_2 \Rightarrow 1-\theta_1 > 1-\theta_2$ así, $\xi > 0$ y la razón de verosimilitud es creciente en T, confirmando los resultados obtenidos anteriormente en el ejemplo.

Hasta aquí, sólo una cola de las pruebas ha sido considerada. Estas son pruebas donde las regiones paramétricas que definen las hipótesis son dadas por una desigualdad estrícta. Un ejemplo de interés de una hipótesis que no es de una sola cola es $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Esta prueba puede ser útil cuando se comparan tratamientos competentes. Consideremos ahora la observación de una muestra de la distribución $N(\theta, \sigma^2)$ con σ^2 conocido, consideremos las tres pruebas de $H_0 : \theta = \theta_0$ basadas en $\overline{X}$:

1) Rechazo H_0 si $|\overline{X} - \theta_0| > 1{,}645\frac{\sigma}{\sqrt{n}}$;

2) Rechazo H_0 si $\overline{X} - \theta_0 > 1{,}282\frac{\sigma}{\sqrt{n}}$;

3) Rechazo H_0 si $|\overline{X} - \theta_0| < 0{,}126\frac{\sigma}{\sqrt{n}}$.

Calculando la probabilidad de rechazar H_0 vemos que las tres pruebas tienen nivel 0,1. El siguiente paso es proceder con la evaluación de la potencia de cada una de las pruebas. Se puede observar facilmente que ninguna de las pruebas es UMP sobre las otras dos. Sin embargo, la primera prueba es la única con $\min_{\Theta}, \pi(\theta) > \pi(\theta_0)$. Esto significa que las probabilidad de rechazar es más grande bajo la alternativa que bajo la hipótesis nula. Esto garantiza que las probabilidades de rechazar H_0 son grandes cuando H_0 es falso. Esto parece una probabilidad razonable a exigir en las pruebas.

Definición 3.3. *Una prueba ψ para $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$ se dice que es insesgado si para cada par $(\theta, \theta^{'})$ donde $\theta \in \Theta_0$ y $\theta^{'} \in \Theta_1$, entonces $\pi_{\psi}(\theta) \leq \pi_{\psi}(\theta^{'})$. La función de potencia es al menos tan grande en Θ_1 como lo es en Θ_0. Si la prueba no satisface las condiciones anteriores, se dice que es sesgada.*

Uno puede tratar de construir una prueba uniformemente más potente para $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$ dentro de las clases de pruebas insesgadas. En la familia paramétrica exponencial, se puede demostrar que ϕ es una función estríctamente creciente de θ, la prueba insesgada uniformemente más potente de nivel α para $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$ acepta H_0 cuando $c_1 < T(X) < c_2$ con $P(c_1 < T(X) < c_2 | \theta_0) = 1 - \alpha$ y c_1, c_2 es un intervalo de densidad muestral más alta de $T(X)$. Las distribuciones muestrales de $T(X)$ no son necesariamente simétricas sobre (θ_0), como en el caso normal anteriormente presentado. Para tales pruebas, tenemos dos $p-valores$ distintos. En el caso general, se usa el más pequeño.

Tal vez no sea posible, en general, hallar pruebas insesgadas. Un procedimiento general para probar $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$ está basado en la razón de máxima verosimilitud estadística dada por

$$\lambda(X) = \frac{\sup_{\theta \in \Theta_0} p(X|\theta)}{\sup_{\theta \in \Theta_1} p(X|\theta)}.$$

El caso más común para el uso de este procedimiento es cuando Θ_0 y Θ_1 son excluyentes y exhaustivos y Θ_0 es de dimensión más pequeña que Θ_1. En estos casos, el denominador se sustituye por el supremo sobre todo el espacio paramétrico Θ, el cual es más fácil de evaluar. Formalmente, el estadístico $\lambda(X)$ es reemplazado por $\max\{\lambda(X), 1\}$. En algún caso, $\lambda(X)$ es una variable aleatoria en función de la muestra. La prueba de razón de máxima verosimilitud (o generalizada) para H_0 de nivel α acepta H_0 si $\lambda(X) > c$ donde c satisface

$$\alpha \geq \sup_{\theta \in \Theta_0} P(\lambda(X) < c|\theta).$$

Una vez más, α es tomado igual al supremo de las probabilidades anteriores siempre que sea posible. La potencia de la prueba está dada por $\pi(\theta) = P(\lambda(X) < c|\theta)$. Esta prueba rechaza la hipótesis nula H_0 si el valor máximo de la probabilidad bajo H_0 está distante del valor máximo global. Esto es un indicador de que hay una gran mejora en la verosimilitud por considerar puntos fuera de H_0 y esta hipótesis no proporciona una buena descripción de los datos. En este caso, tiene sentido rechazar H_0.

Es importante distinguir entre la prueba anterior y la prueba basada en razones de verosimilitud monótona. Aunque, la prueba de razón de máxima verosimilitud goza de buenas propiedades asintóticas, no siempre es insesgada ó uniformemente más potente. Las principales dificultades asociadas con ellas es el cálculo de la máxima verosimilitud de forma cerrada y la determinación de su distribución muestral.

Otras propiedades deseables en los procedimientos de pruebas son la similaridad e invarianza. Considere el problema de la prueba $H : \theta = \theta_0$ en la presencia de un parámetro inusual ϕ. Una prueba se dice que es similar si el nivel de la prueba es el mismo cualquiera que sea el valor de la distribución del parámetro. Un ejemplo de una prueba similar es la prueba t, que veremos más tarde en esta sección. Una prueba se dice invariante (bajo la familia específica de transformaciones) si la distribución de alguna transformación de las observaciones dentro de la familia permanece en la misma familia. Esto permitirá a las hipótesis permanecer inalteradas con alguna transformación operada sobre la data. Las pruebas que se presentan a continuación en esta sección son todas invariantes bajo transformaciones lineales de las observaciones.

§3.1.3. Prueba de hipótesis con distribución normal

Las pruebas más comunes para muestras de una distribución normal serán presentan acá.

Sean $X_1, ..., X_n$ m.a. independientes e idénticamente distribuidos con $X_i \sim N(\theta, \sigma^2)$ y suponga que se desea probar $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$. Supongamos inicialmente que σ^2 es conocido. En este caso, debemos mostrar que la prueba insesgada uniformemente más potente de nivel α esta dada por la región crítica $\dfrac{\sqrt{n}(\overline{X} - \theta_0)}{\sigma} > z_{\alpha/2}$. Ahora obtendremos la prueba de razón de máxima verosimilitud.

Como $\Theta_0 = \{\theta_0\}$ entonces

$$\sup_{\theta \in \Theta_0} p(x|\theta) = p(x|\theta_0).$$

Para $\theta \in \Theta_1 = \Theta - \{\theta_0\}$, el máximo de $p(x|\theta)$ es obtenido de $\widehat{\theta}$, el estimador de máxima verosimilitud de θ, el cuál en este caso es la media muestral $\overline{X}$. Note que para cualquier valor de $\theta, P(\widehat{\theta} = \theta_0) = 0$ y por lo tanto al considerar el máximo sobre Θ en lugar de sobre Θ_1 en la expresión de razón de máxima verosimilitud no produce ningún cambio. Entonces la razón de máxima verosimilitud estadística esta dada por:

$$\lambda(X) = \frac{p(X|\theta_0)}{p(X|\widehat{\theta})} \tag{3.1}$$

Observemos que:

$$\begin{aligned}
p(X|\theta_0) &= (2\pi\sigma)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta_0)^2 \right\} \\
&= (2\pi\sigma)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \overline{X} + \overline{X} - \theta_0)^2 \right\} \\
&= (2\pi\sigma)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (x_i - \overline{X})^2 + \sum_{i=1}^{n} 2(x_i - \overline{X})(\overline{X} - \theta_0) + \sum_{i=1}^{n} (\overline{X} - \theta_0)^2 \right] \right\} \\
&= (2\pi\sigma)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1^n} (x_i - \overline{X})^2 \right\} \\
&\times \exp\left\{ -\frac{1}{2\sigma^2} \left[2(\overline{X} - \theta_0) \sum_{i=1}^{n} (x_i - \overline{X}) + n(\overline{X} - \theta_0)^2 \right] \right\}
\end{aligned} \tag{3.2}$$

y

$$
\begin{aligned}
p(X|\widehat{\theta}) = p(X|\overline{X}) &= (2\pi\sigma)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\widehat{\theta})^2\right\} \\
&= (2\pi\sigma)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{X})^2\right\}. \qquad (3.3)
\end{aligned}
$$

Luego sustituyendo en (3.1) tenemos:

$$
\begin{aligned}
\lambda(X) &= \frac{p(X|\theta_0)}{p(X|\widehat{\theta})} \\
&= \frac{(2\pi\sigma)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1^n}(x_i-\overline{X})^2\right\}\exp\left\{-\frac{1}{2\sigma^2}\left[2(\overline{X}-\theta_0)\sum_{i=1}^{n}(x_i-\overline{X})+n(\overline{X}-\theta_0)^2\right]\right\}}{(2\pi\sigma)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{X})^2\right\}} \\
&= \exp\left\{-\frac{1}{2\sigma^2}2(\overline{X}-\theta_0)\left(\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\overline{X}\right)\right\}\exp\left\{-\frac{1}{2\sigma^2}n(\overline{X}-\theta_0)^2\right\} \\
&= \exp\left\{-\frac{1}{\sigma^2}(\overline{X}-\theta_0)(n\overline{X}-n\overline{X})\right\}\exp\left\{-\frac{1}{2\sigma^2}n(\overline{X}-\theta_0)^2\right\} \\
&= \exp\left\{-\frac{n}{2\sigma^2}(\overline{X}-\theta_0)^2\right\}.
\end{aligned}
$$

Nótese que bajo $H_0, \overline{X} \sim N(\theta_0, \sigma^2/n)$ y por lo tanto $z = \dfrac{\sqrt{n}(\overline{X}-\theta_0)}{\sigma} \sim N(0,1)$. La razón de máxima verosimilitud está dada por $\lambda(X) = \exp\left\{-\dfrac{z^2}{2}\right\}$. Así, la prueba de razón de máxima verosimilitud rechaza H_0 si:

$$
\begin{aligned}
\lambda(X) < c &\Leftrightarrow \exp\left\{-\frac{z^2}{2}\right\} < c \\
&\Leftrightarrow -\frac{z^2}{2} < c_1 \\
&\Leftrightarrow \frac{z^2}{2} > c_2 \\
&\Leftrightarrow z^2 > c_3 \\
&\Leftrightarrow |z| > c_4.
\end{aligned}
$$

Dado un nivel de significancia $\alpha, \alpha = P(|Z| > c_4|H_0)$ o $c_4 = z_{\alpha/2}$. La potencia de la prueba es $\pi(\theta) = P(|Z| > z_{\alpha/2}|H_1)$.

Bajo $H_1, \overline{X} \sim N(\theta, \sigma^2/n)$. Así

$$\overline{X} - \theta_0 \sim N\left(\theta - \theta_0, \frac{\sigma^2}{n}\right) \Leftrightarrow W = \frac{\sqrt{n}}{\sigma}[\overline{X} - \theta_0 - (\theta - \theta_0)] \sim N(0,1).$$

Entonces

$$W = \frac{\sqrt{n}}{\sigma}(\overline{X} - \theta_0 - (\theta - \theta_0)) = \frac{\sqrt{n}}{\sigma}(\overline{X} - \theta_0) - \frac{\sqrt{n}}{\sigma}(\theta - \theta_0) = Z - \frac{\sqrt{n}}{\sigma}(\theta - \theta_0),$$

y por lo tanto, la potencia de la prueba es:

$$\begin{aligned}
\pi(\theta) &= 1 - P(-z_{\alpha/2} < Z < z_{\alpha/2}|\theta) = 1 - P\left(-z_{\alpha/2} < W + \frac{\sqrt{n}(\theta-\theta_0)}{\sigma} < z_{\alpha/2}|\theta\right) \\
&= 1 - P\left(-z_{\alpha/2} - \frac{\sqrt{n}(\theta-\theta_0)}{\sigma} < W < z_{\alpha/2} - \frac{\sqrt{n}(\theta-\theta_0)}{\sigma}|\theta\right) \\
&= 1 - \left[P\left(W < z_{\alpha/2} - \frac{\sqrt{n}(\theta-\theta_0)}{\sigma}\right) - P\left(W < -z_{\alpha/2} - \frac{\sqrt{n}(\theta-\theta_0)}{\sigma}\right)\right] \\
&= 1 + \Phi\left(-z_{\alpha/2} - \frac{\sqrt{n}(\theta-\theta_0)}{\sigma}\right) - \Phi\left(z_{\alpha/2} - \frac{\sqrt{n}(\theta-\theta_0)}{\sigma}\right) > \alpha.
\end{aligned}$$

La prueba de razón de máxima verosimilitud es insesgada como $\pi(\theta) > \alpha, \forall\theta \neq \theta_0$. La tasa de crecimiento de la potencia depende de σ. La más pequeña es σ (distribución más precisa, población más concentrada), la más rápida es la tasa de crecimiento de la potencia hacia 1.

En el caso donde σ^2 es desconocido, $\Theta_0 = \{(\theta, \sigma^2) : \theta = \theta_0, \sigma^2 > 0\}$ y $\Theta = \{(\theta, \sigma^2) : \theta \in \mathbb{R}, \sigma^2 > 0\}$. Ya que la dimensión de Θ_0 es más pequeña que la dimensión de Θ_1 y de Θ, trabajaremos con el último. Como se habia mencionado antes, este cambio podría afectar solamente a un conjunto con probabilidad cero, es decir:

$$\sup_{(\theta,\sigma^2)\in\Theta_0} p(X|\theta, \sigma^2) = p(X|\theta_0, \widehat{\sigma}_0^2)$$

donde

$$\widehat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \theta_0)^2$$

y

$$\sup_{(\theta,\sigma^2)\in\Theta} p(X|\theta,\sigma^2) = p(X|\widehat{\theta},\widehat{\sigma}^2)$$

donde

$$\widehat{\theta} = \overline{X} \quad \text{y} \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Por lo tanto, el estadístico de razón de máxima verosimilitud esta dado:

$$\begin{aligned}
\lambda(X) &= \frac{p(X|\theta_0,\widehat{\sigma}_0^2)}{p(X|\widehat{\theta},\widehat{\sigma}^2)} = \frac{(2\pi\widehat{\sigma}_0^2)^{-n/2}\exp\left\{-\frac{1}{2\widehat{\sigma}_0^2}\sum_{i=1}^{n}(X_i-\theta_0)^2\right\}}{(2\pi\widehat{\sigma}^2)^{-n/2}\exp\left\{-\frac{1}{2\widehat{\sigma}^2}\sum_{i=1}^{n}(X_i-\overline{X})^2\right\}} \\
&= \frac{(2\pi)^{n/2}(\widehat{\sigma}^2)^{n/2}\exp\left\{-\frac{1}{2\widehat{\sigma}_0^2}n\widehat{\sigma}_0^2\right\}}{(2\pi)^{n/2}(\widehat{\sigma}_0^2)^{n/2}\exp\left\{-\frac{1}{2\widehat{\sigma}^2}n\widehat{\sigma}^2\right\}} \\
&= \frac{(\widehat{\sigma^2})^{n/2}\exp\left\{-\frac{n}{2}\right\}}{(\widehat{\sigma}_0^2)^{n/2}\exp\left\{-\frac{n}{2}\right\}} \\
&= \left(\frac{\widehat{\sigma}^2}{\widehat{\sigma}_0^2}\right)^{n/2}.
\end{aligned}$$

También se puede escribir:

$$\begin{aligned}
\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2} &= \frac{\sum_{i=1}^{n}\frac{(X_i-\theta_0)^2}{n}}{\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{n}} = \frac{\sum_{i=1}^{n}(X_i-\overline{X}+\overline{X}-\theta_0)^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2} \\
&= \frac{\sum_{i=1}^{n}\left[(X_i-\overline{X})^2 + 2(X_i-\overline{X})(\overline{X}-\theta_0) + (\overline{X}-\theta_0)^2\right]}{\sum_{i=1}^{n}(X_i-\overline{X})^2} \\
&= \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2 + 2(X_i-\theta_0)\sum_{i=1}^{n}(X_i-\overline{X}) + \sum_{i=1}^{n}(\overline{X}-\theta_0)^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2 + n(\overline{X}-\theta_0)^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2} = 1 + \frac{n(\overline{X}-\theta_0)^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2} \\
&= 1 + \frac{n(\overline{X}-\theta_0)^2}{(n-1)S^2} = 1 + \frac{T^2}{n-1},
\end{aligned}
$$

donde

$$
T = \frac{\sqrt{n}(\overline{X}-\theta_0)}{S}
$$

y la razón de máxima verosimilitud puede escribirse como $\lambda(X) = \left(1+\frac{T^2}{n-1}\right)^{-n/2}$. Por lo tanto, la prueba de razón de máxima verosimilitud acepta H_0 si $T^2 < c_1$ o $|T| < c$. Como $T|H_0 \sim t_{n-1}(0,1)$ el valor de c de la prueba de nivel α es $t_{\alpha/2}(n-1)$. Esta prueba es usualmente conocida como la prueba t (t-student) y es posiblemente la más usada en estadística. También se puede probar que la potencia de la prueba es una función estrictamente creciente de $|\theta-\theta_0|$. Una consecuencia inmediata es el insesgamiento de la prueba ya que el valor más pequeño de la potencia ocurre cuando $\theta = \theta_0$.

Esta prueba es similar porque ninguna de las propiedades de la prueba es afectada por el valor del parámetro σ^2. Esto ocurrió por el reemplazo de σ^2 por su estimador S^2 y la existencia de una cantidad pivote T. Esta prueba es también invariante bajo transformaciones lineales.

Otra prueba común es la prueba de igualdad de dos medias. Considere dos muestras aleatorias $X_1 = (X_{11},...,X_{1n_1})$ de la distribución $N(\theta_1,\sigma^2)$ y $X_2 = (X_{21},...,X_{2n_2})$ de la distribución $N(\theta_2,\sigma^2)$. Asumimos por simplicidad que las varianzas son iguales. Entonces, la hipótesis de interés esta definida por el espacio de parámetros: $\Theta_0 = \{(\theta_1,\theta_2,\sigma^2) : \theta_1 = \theta_2 = \theta \in \mathbb{R}, \sigma^2 > 0\}$. Observemos que bajo H_0 el problema contiene sólo una muestra simple de tamaño n_1+n_2 de la distribución $N(\theta,\sigma^2)$. El vector parámetro puede ser más útil descrito por tres nuevos parámetros de interés $\theta_2-\theta_1$, alguna otra transformación de parámetros θ_1 y θ_2 no múltiplos de $\theta_2-\theta_1$ y σ^2. Los dos últimos parámetros son ruidosos.

Una vez más, la razón de máxima verosimilitud está basada en la maximización sobre Θ y Θ_0. Este operación esta dada por:

$$\sup_{(\theta_1,\theta_2,\sigma^2)\in\Theta_0} p(X_1,X_2|\theta,\sigma^2) = \left(\frac{1}{2\pi\widehat{\sigma}_0^2}\right)^{\frac{(n_1+n_2)}{2}} \exp\left\{-\frac{1}{2\widehat{\sigma}_0^2}\left[\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\widehat{\theta})^2\right]\right\}$$

donde

$$\begin{aligned}
\widehat{\theta} = \overline{X}_{12} &= \frac{1}{n_1+n_2}\left[(X_{11}+X_{12}+...+X_{1n_1})+(X_{21}+X_{22}+...+X_{2n_2})\right]\\
&= \frac{1}{n_1+n_2}\left[\sum_{i=1}^{n_1}X_{1i}+\sum_{i=1}^{n_2}X_{2i}\right]
\end{aligned}$$

y

$$\begin{aligned}
\widehat{\sigma}_0^2 &= \frac{1}{n_1+n_2}\left[\big((X_{11}-\theta)^2+(X_{12}-\theta)^2+...+(X_{1n_1}-\theta)^2\big)\right.\\
&+ \left.\big((X_{21}-\theta)^2+(X_{22}-\theta)^2+...+(X_{2n_2}-\theta)^2\big)\right]\\
&= \frac{1}{n_1+n_2}\left[\sum_{i=1}^{n_1}(X_{1i}-\widehat{\theta})^2+\sum_{i=1}^{n_2}(X_{2i}-\widehat{\theta})^2\right].
\end{aligned}$$

Por otro lado,

$$\sup_{(\theta_1,\theta_2,\sigma^2)\in\Theta} p(X_1,X_2|\theta_1,\theta_2,\sigma^2) = \left(\frac{1}{2\pi\widehat{\sigma}_0^2}\right)^{\frac{(n_1+n_2)}{2}} \exp\left\{-\frac{1}{2\widehat{\sigma}^2}\left[\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\widehat{\theta}_i)^2\right]\right\}$$

donde

$$\widehat{\theta}_i = \overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}X_{ij} \quad \text{con} \quad i=1,2;$$

y

$$\begin{aligned}
\widehat{\sigma}^2 &= \frac{1}{n_1+n_2}\left[\Big((X_{11}-\widehat{\theta}_1)^2+(X_{12}-\widehat{\theta}_1)^2+...+(X_{1n_1}-\widehat{\theta}_1)^2\Big)\right.\\
&+ \left.\Big((X_{21}-\widehat{\theta}_2)^2+(X_{22}-\widehat{\theta}_2)^2+...+(X_{2n_2}-\widehat{\theta}_2)^2\Big)\right]\\
&= \frac{1}{n_1+n_2}\left[\sum_{i=1}^{n_1}(X_{1i}-\widehat{\theta}_1)^2+\sum_{i=1}^{n_2}(X_{2i}-\widehat{\theta}_2)^2\right].
\end{aligned}$$

Por lo tanto, la razón de máxima verosimilitud esta dada por:

$$
\begin{aligned}
\lambda(X_1, X_2) &= \frac{\left(\frac{1}{2\pi\widehat{\sigma}_0^2}\right)^{\frac{(n_1+n_2)}{2}} \exp\left\{-\frac{1}{2\widehat{\sigma}_0^2}\left[\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\widehat{\theta})^2\right]\right\}}{\left(\frac{1}{2\pi\widehat{\sigma}^2}\right)^{\frac{(n_1+n_2)}{2}} \exp\left\{-\frac{1}{2\widehat{\sigma}^2}\left[\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\widehat{\theta}_i)^2\right]\right\}} \\
&= \frac{\frac{1}{(\widehat{\sigma}_0^2)^{\frac{(n_1+n_2)}{2}}} \exp\left\{-\frac{1}{2\widehat{\sigma}_0^2}(n_1+n_2)\widehat{\sigma}_0^2\right\}}{\frac{1}{(\widehat{\sigma}^2)^{\frac{(n_1+n_2)}{2}}} \exp\left\{-\frac{1}{2\widehat{\sigma}^2}(n_1+n_2)\widehat{\sigma}^2\right\}} \\
&= \frac{(\widehat{\sigma}^2)^{\frac{(n_1+n_2)}{2}} \exp\left\{-\frac{(n_1+n_2)}{2}\right\}}{(\widehat{\sigma}_0^2)^{\frac{(n_1+n_2)}{2}} \exp\left\{-\frac{(n_1+n_2)}{2}\right\}} \\
&= \left(\frac{\widehat{\sigma}^2}{\widehat{\sigma}_0^2}\right)^{\frac{n_1+n_2}{2}}.
\end{aligned}
$$

Notemos que;

$$
\begin{aligned}
\overline{X}_1 - \widehat{\theta} &= \overline{X}_1 - \left(\frac{1}{n_1+n_2}(n_1\overline{X}_1 + n_2\overline{X}_2)\right) \\
&= \frac{\overline{X}_1 n_1 + \overline{X}_1 n_2 - n_1\overline{X}_1}{n_1+n_2} - \frac{n_2\overline{X}_2}{n_1+n_2} \\
&= \frac{n_2}{n_1+n_2}(\overline{X}_1 - \overline{X}_2),
\end{aligned}
$$

y

$$
\begin{aligned}
\overline{X}_2 - \widehat{\theta} &= \overline{X}_2 - \left(\frac{1}{n_1+n_2}(n_1\overline{X}_1 + n_2\overline{X}_2)\right) \\
&= \frac{\overline{X}_2 n_1 + \overline{X}_2 n_2 - n_1\overline{X}_1}{n_1+n_2} - \frac{n_2\overline{X}_2}{n_1+n_2} \\
&= \frac{n_1}{n_1+n_2}(\overline{X}_2 - \overline{X}_1).
\end{aligned}
$$

Esto implíca que:

$$
\begin{aligned}
& \sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\widehat{\theta})^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}((X_{ij}-\overline{X}_i)+(\overline{X}_i-\widehat{\theta}))^2 \\
= & \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left[(X_{ij}-\overline{X}_i)^2+2(X_{ij}-\overline{X}_i)(\overline{X}_i-\widehat{\theta})+(\overline{X}_i-\widehat{\theta})^2\right] \\
= & \sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2+2\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)(\overline{X}_i-\widehat{\theta})+\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\overline{X}_i-\widehat{\theta})^2 \\
= & \sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2+2(\overline{X}_1-\widehat{\theta})\sum_{j=1}^{n_1}(X_{1j}-\overline{X}_1)+2(\overline{X}_2-\widehat{\theta})\sum_{j=1}^{n_2}(X_{2j}-\overline{X}_2) \\
+ & n_1(\overline{X}_1-\widehat{\theta})^2+n_2(\overline{X}_2-\widehat{\theta})^2 \\
= & \sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2+2(\overline{X}_1-\widehat{\theta})\left[\sum_{j=1}^{n_1}X_{1j}-\sum_{j=1}^{n_1}\overline{X}_1\right]+2(\overline{X}_2-\widehat{\theta})\left[\sum_{j=1}^{n_2}X_{2j}-\sum_{j=1}^{n_2}\overline{X}_2\right] \\
+ & n_1\left(\frac{n_2}{n_1+n_2}(\overline{X}_1-\overline{X}_2)\right)^2+n_2\left(\frac{n_1}{n_1+n_2}(\overline{X}_2-\overline{X}_1)\right)^2 \\
= & \sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2+\frac{n_1n_2^2}{(n_1+n_2)^2}(\overline{X}_1-\overline{X}_2)^2+\frac{n_2n_1^2}{(n_1+n_2)^2}(\overline{X}_2-\overline{X}_1)^2 \\
= & \sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2+\frac{n_1n_2^2+n_2n_1^2}{(n_1+n_2)^2}(\overline{X}_1-\overline{X}_2)^2 \\
= & \sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2+\frac{n_1n_2(n_2+n_1)}{(n_1+n_2)^2}(\overline{X}_1-\overline{X}_2)^2 \\
= & \sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2+\frac{n_1n_2}{n_1+n_2}(\overline{X}_1-\overline{X}_2)^2.
\end{aligned}
$$

Por lo tanto,

$$
\lambda(X_1,X_2) = \left[\frac{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2+\frac{n_1n_2}{n_1+n_2}(\overline{X}_1-\overline{X}_2)^2}{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2}\right]^{-\frac{n_1+n_2}{2}}
$$

$$= \left[1 + \frac{\frac{n_1 n_2}{n_1+n_2}(\overline{X}_1 - \overline{X}_2)^2}{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2} \right]^{-\frac{n_1+n_2}{2}}$$

$$= \left[1 + \frac{\left(\frac{n_1+n_2}{n_1 n_2}\right)^{-1}(\overline{X}_1 - \overline{X}_2)^2}{\left(\sum_{j=1}^{n_1}(X_{1j} - \overline{X}_1)^2 + \sum_{j=1}^{n_2}(X_{2j} - \overline{X}_2)^2\right)} \right]^{-\frac{n_1+n_2}{2}}$$

$$= \left[1 + \frac{\left(n_2^{-1} + n_1^{-1}\right)^{-1}(\overline{X}_1 - \overline{X}_2)^2}{(n_1+n_2-2)\frac{(n_1+n_2)\widehat{\sigma}^2}{n_1+n_2-2}} \right]^{-\frac{n_1+n_2}{2}}$$

$$= \left[1 + \frac{\left(n_2^{-1} + n_1^{-1}\right)^{-1}(\overline{X}_1 - \overline{X}_2)^2}{(n_1+n_2-2)S^2} \right]^{-\frac{n_1+n_2}{2}}$$

$$= \left[1 + \frac{(\overline{X}_1 - \overline{X}_2)^2}{(n_2^{-1} + n_1^{-1})(n_1+n_2-2)S^2} \right]^{-\frac{n_1+n_2}{2}}$$

$$= \left[1 + \frac{T^2}{n_1+n_2-2} \right]^{-\frac{n_1+n_2}{2}},$$

donde,

$$T^2 = \frac{(\overline{X}_1 - \overline{X}_2)^2}{(n_1^{-1} + n_2^{-1})S^2} \Rightarrow T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{n_1^{-1} + n_2^{-1}}S} \quad \text{con} \quad S^2 = \frac{(n_1+n_2)\widehat{\sigma}^2}{n_1+n_2-2}.$$

Así, la prueba de razón de máxima verosimilitud acepta H_0 cuando $\lambda(X_1, X_2) > c_1$ o cuando $|T| < c$. Sabemos que, $T|H_0 \sim t_{n_1+n_2-2}(0,1)$. Para una prueba de nivel $\alpha, c = t_{\alpha/2, n_1+n_2-2}$. Análogamente, se puede probar que la potencia de esta prueba es una función estrictamente creciente de $|\theta_1 - \theta_2|$. Por lo tanto, esta prueba es insesgada por las mismas razones que la prueba t anterior. Esto es similar porque ninguna de sus propiedades son afectadas por los dos parámetros de ruido y son invariantes bajo transformaciones lineales en las observaciones.

§3.2. Prueba de hipótesis Bayesiana

En el contexto Bayesiano, el problema de decidir cual hipótesis aceptar es conceptualmente más simple. En general, uno podría comparar las hipótesis $H_1, ..., H_k$ a través de su respectiva posteriori, obtenida por el teorema de Bayes como:

$$p(H_i|x) \propto p(x|H_i)p(H_i).$$

De nuevo, esto se puede catalogar como un problema de decisión. Adicionalmente a las probabilidades (posteriori) de las hipótesis, una estructura de la pérdida asociada con las posibles acciones pueden ser incorporadas.

Al volver a los casos especiales de dos hipótesis, supongamos que se quiere probar $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$. Es suficiente examinar las probabilidades posteriori $p(H_0|x)$ y $p(H_1|x)$. Si $p(H_0|x) > p(H_1|x)$, entonces H_0 podría ser aceptada como la hipótesis más adecuada para θ. En este caso, se puede decir que H_0 es preferible a H_1. En otro caso, H_1 es preferible a H_0. Existe una regla clara para la escogencia de las hipótesis, la cuál no siempre es verdad bajo un enfoque clásico.
Como

$$p(H_0|x) \propto p(x|H_0)p(H_0),$$
$$p(H_1|x) \propto p(x|H_1)p(H_1),$$

y observando que la constante de la proporcionalidad es la misma en ambas expresiones,

$$\frac{p(H_0|x)}{p(H_1|x)} = \frac{p(H_0)p(x|H_0)}{p(H_1)p(x|H_1)}.$$

La razón $\frac{p(H_0)}{p(H_1)}$ es llamada probabilidad a priori entre H_0 y H_1 y la razón $\frac{p(H_0|x)}{p(H_1|x)}$ es llamada la probabilidad posteriori entre H_0 y H_1. La razón

$$\frac{p(x|H_0)}{p(x|H_1)}$$

es llamada factor de Bayes y se denota por $BF(H_0 : H_1)$. Este concepto fue introducido por Jeffreys (1961). Nótese que esto es de alguna manera una razón de verosimilitud. Así, la posteriori está dada por el producto de la priori y el factor de Bayes. Una vez más, la razón de verosimilitud introduce la influencia de las observaciones en el conjunto de las pruebas de hipótesis. En general, las verosimilitudes acá son marginales en el sentido que se obtienen después de la integración de algunos parámetros no asociados con las especificaciones de las hipótesis.

A pesar de su simplicidad notacional, no es fácil en algunos casos especificar $p(H_j)$ cuando H_j es una hipótesis simple, $j = 0, 1$ y θ es continua. Si f es la densidad a priori especificada para $\theta \in \Theta$ tendrémos que $p(H_j) = p(H_j|x) = 0$. En estos casos, una solución es atribuir una probabilidad a priori global π a la hipótesis simple, digamos H_0, para $\pi \in (0, 1)$. Es decir, si H_1 es el complemento de una hipótesis simple H_0, entonces $p(H_1) = 1 - \pi$ y esta probabilidad es distribuida sobre los diferentes valores de θ bajo H_1, generando la distribución a priori de $\theta|H_1$. Esta distribución tendrá densidad f sobre Θ_1.

Como H_0 es una hipótesis simple, se tiene que $p(x|H_0) = p(x|\theta_0)$, la densidad marginal de X dada H_0. Esto también puede ser referido como la probabilidad marginal de H_0 basado en X. La probabilidad marginal de H_1 basada en X es:

$$\begin{aligned} p(x|H_1) &= \int_{\Theta-\{\theta_0\}} p(x, \theta|H_1) d\theta \\ &= \int_{\Theta-\{\theta_0\}} p(x|\theta, H_1) p(\theta|H_1) d\theta \\ &= \int_{\Theta} p(x|\theta) f(\theta) d\theta \end{aligned}$$

observando que la ultima integral se lleva a cabo en todo el espacio de parámetros Θ pues un sólo punto no altera su valor. El factor de Bayes de reduce a:

$$BF(H_0 : H_1) = \frac{p(x|\theta_0)}{\int p(x|\theta) f(\theta) d\theta}.$$

Nótese que esto proporciona las probabilidades relativas entre H_0 y H_1 sin tomar en cuenta las probabilidades a priori. Esto es una medida Bayesiana de la bondad de ajuste de un modelo dado para el conjunto de datos. Un factor de Bayes mayor que 1 indica que H_0 ajusta los datos mejor que H_1 y así tenemos mayor probabilidad. Además, la priori marginal de θ se mezcla y por tanto la distribución marginal de X es:

$$\begin{aligned} p(x) &= \int p(x|\theta)dF(\theta) \\ &= \pi p(x|\theta_0) + (1-\pi)\int p(x|\theta)f(\theta)d\theta \\ &= \pi p(x|\theta_0) + (1-\pi)p(x|H_1). \end{aligned} \tag{3.4}$$

Sea $X = (X_1, ..., X_n)$ una muestra aleatoria de la distribución $N(\theta, \sigma^2)$ y supóngase que uno desea probar $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$. Consideremos solamente el caso cuando la varianza σ^2 es conocida. Entonces, con una probabilidad a priori π de H_0 y considerando que $\theta|H_1 \sim N(\mu, \tau^2)$ es dada, se tiene:

$$p(x|H_1) = \int p(x|\theta)p(\theta)d\theta$$

donde;

$$\begin{aligned} p(x|\theta) &= \prod_{i=1}^{n} p(x_i|\theta) = \prod_{i=1}^{n}(2\pi\sigma^2)^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}(x_i-\theta)^2\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\theta)^2\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(x_i-\overline{x})^2+2(x_i-\overline{x})(\overline{x}-\theta)+(\overline{x}-\theta)^2\right]\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\overline{x})^2+2(\overline{x}-\theta)\sum_{i=1}^{n}(x_i-\overline{x})+\sum_{i=1}^{n}(\overline{x}-\theta)^2\right)\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\overline{x})^2+n(\overline{x}-\theta)^2\right)\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right\}\times\exp\left\{-\frac{n}{2\sigma^2}(\overline{x}-\theta)^2\right\}, \end{aligned} \tag{3.5}$$

y como $\theta|H_1 \sim N(\mu, \tau^2)$

$$\begin{aligned} p(\theta) = f(\theta) = p(\theta|H_1) &= (2\pi\tau^2)^{-1/2}\exp\left\{-\frac{1}{2\tau^2}(\theta-\mu)^2\right\} \\ &= \frac{1}{\sqrt{2\pi}\tau}\exp\left\{-\frac{1}{2\tau^2}(\theta-\mu)^2\right\}. \end{aligned} \tag{3.6}$$

Luego, de (3.5) y (3.6)se tiene que:

$$\begin{aligned}
p(x|H_1) &= \int p(x|\theta)\, p(\theta)\, d\theta \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right\} \\
&\times \int \exp\left\{-\frac{n}{2\sigma^2}(\overline{x}-\theta)^2\right\} \times \frac{1}{\sqrt{2\pi}\tau} \exp\left\{-\frac{1}{2\tau^2}(\theta-\mu)^2\right\} d\theta. \qquad (3.7)
\end{aligned}$$

Desarrollemos ahora el integrando:

$$\begin{aligned}
& \exp\left\{-\frac{n}{2\sigma^2}(\overline{x}^2-2\overline{x}\theta+\theta^2)\right\} \times \exp\left\{-\frac{1}{2\tau^2}(\theta^2-2\theta\mu+\mu^2)\right\} \\
= \; & \exp\left\{-\frac{n\overline{x}^2}{2\sigma^2}\right\} \exp\left\{-\frac{n}{2\sigma^2}(-2\overline{x}\theta+\theta^2)\right\} \exp\left\{-\frac{1}{2\tau^2}(\theta^2-2\theta\mu)\right\} \exp\left\{-\frac{\mu^2}{2\tau^2}\right\} \\
= \; & \exp\left\{-\frac{n\overline{x}^2}{2\sigma^2}-\frac{\mu^2}{2\tau^2}\right\} \exp\left\{-\frac{n(-2\overline{x}\theta)}{2\sigma^2}-\frac{n\theta^2}{2\sigma^2}-\frac{\theta^2}{2\tau^2}+\frac{2\theta\mu}{2\tau^2}\right\} \\
= \; & \exp\left\{-\frac{n\overline{x}^2}{2\sigma^2}-\frac{\mu^2}{2\tau^2}\right\} \exp\left\{-\frac{1}{2}\left[\frac{n\theta^2}{\sigma^2}+\frac{\theta^2}{\tau^2}-\frac{2n\overline{x}\theta}{\sigma^2}-\frac{2\theta\mu}{\tau^2}\right]\right\} \\
= \; & \exp\left\{-\frac{n\overline{x}^2}{2\sigma^2}-\frac{\mu^2}{2\tau^2}\right\} \exp\left\{-\frac{1}{2}\left[\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\theta^2-2\left(\frac{n\overline{x}}{\sigma^2}+\frac{\mu}{\tau^2}\right)\theta\right]\right\} \\
= \; & \exp\left\{-\frac{n\overline{x}^2}{2\sigma^2}-\frac{\mu^2}{2\tau^2}\right\} \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\left[\theta^2-2\frac{\left(\frac{n\overline{x}}{\sigma^2}+\frac{\mu}{\tau^2}\right)}{\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)}\theta\right]\right\} \\
= \; & \exp\left\{-\frac{n\overline{x}^2}{2\sigma^2}-\frac{\mu^2}{2\tau^2}\right\} \exp\left\{-\frac{1}{2}\left(\frac{n\tau^2+\sigma^2}{\sigma^2\tau^2}\right)\left[\theta^2-2\frac{\left(\frac{n\overline{x}\tau^2+\mu\sigma^2}{\sigma^2\tau^2}\right)}{\left(\frac{n\tau^2+\sigma^2}{\sigma^2\tau^2}\right)}\theta\right]\right\} \\
= \; & \exp\left\{-\frac{n\overline{x}^2}{2\sigma^2}-\frac{\mu^2}{2\tau^2}\right\} \exp\left\{-\frac{1}{2}\left(\frac{n\tau^2+\sigma^2}{\sigma^2\tau^2}\right)\left[\theta^2-2\left(\frac{n\overline{x}\tau^2+\mu\sigma^2}{n\tau^2+\sigma^2}\right)\theta\right]\right\}, \qquad (3.8)
\end{aligned}$$

observemos que para resolver la integral en (3.7) es necesario completar cuadrados, con respecto a θ, en la ecuación (3.8):

$$\begin{aligned}
& \theta^2-2\left(\frac{n\overline{x}\tau^2+\mu\sigma^2}{n\tau^2+\sigma^2}\right)\theta \\
= \; & \theta^2-2\left(\frac{n\overline{x}\tau^2+\mu\sigma^2}{n\tau^2+\sigma^2}\right)\theta+\left(\frac{n\overline{x}\tau^2+\mu\sigma^2}{n\tau^2+\sigma^2}\right)^2-\left(\frac{n\overline{x}\tau^2+\mu\sigma^2}{n\tau^2+\sigma^2}\right)^2 \\
= \; & \left[\theta-\left(\frac{n\overline{x}\tau^2+\mu\sigma^2}{n\tau^2+\sigma^2}\right)\right]^2-\left(\frac{n\overline{x}\tau^2+\mu\sigma^2}{n\tau^2+\sigma^2}\right)^2,
\end{aligned}$$

así,

$$\int \frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}}} \exp\left\{-\frac{1}{2\left(\sqrt{\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}}\right)^2}\left[\theta - \left(\frac{n\overline{x}\tau^2+\mu\sigma^2}{n\tau^2+\sigma^2}\right)\right]^2\right\} d\theta \;=\; 1.$$

Haciendo $nS^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2$ y reemplazando los cálculos anteriores en (3.7) tenemos que:

$$\begin{aligned}
& p(x|H_1) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{nS^2}{2\sigma^2}\right\} \\
\times \quad & \frac{1}{\sqrt{2\pi}\tau}\int \exp\left\{-\frac{n}{2\sigma^2}(\overline{x}-\theta)^2\right\}\exp\left\{-\frac{1}{2\tau^2}(\theta-\mu)^2\right\} d\theta \\
= \quad & \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{nS^2}{2\sigma^2}\right\}\exp\left\{-\frac{n\overline{x}^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2} + \frac{1}{2}\left(\frac{n\tau^2+\sigma^2}{\sigma^2\tau^2}\right)\left(\frac{n\overline{x}\tau^2+\mu\sigma^2}{n\tau^2+\sigma^2}\right)^2\right\} \\
\times \quad & \int \frac{\sqrt{\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}}}{\sqrt{2\pi}\tau\sqrt{\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}}} \exp\left\{-\frac{1}{2\left(\sqrt{\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}}\right)^2}\left[\theta - \left(\frac{n\overline{x}\tau^2+\mu\sigma^2}{n\tau^2+\sigma^2}\right)\right]^2\right\} d\theta \\
= \quad & \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{nS^2}{2\sigma^2}\right\}\left(\frac{\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}}{\tau^2}\right)^{1/2} \\
\times \quad & \exp\left\{-\frac{1}{2}\left(\frac{n\overline{x}^2\tau^2+\mu^2\sigma^2}{\sigma^2\tau^2}\right) + \frac{1}{2}\frac{(n\tau^2+\sigma^2)}{\sigma^2\tau^2}\frac{(n\overline{x}\tau^2+\mu\sigma^2)^2}{(n\tau^2+\sigma^2)^2}\right\} \\
= \quad & \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{nS^2}{2\sigma^2}\right\}\left(\frac{\sigma^2}{n\tau^2+\sigma^2}\right)^{1/2} \\
\times \quad & \exp\left\{-\frac{1}{2}\left(\frac{n\overline{x}^2\tau^2+\mu^2\sigma^2}{\sigma^2\tau^2}\right) + \frac{1}{2}\frac{(n\overline{x}\tau^2+\mu\sigma^2)^2}{\sigma^2\tau^2(n\tau^2+\sigma^2)}\right\}.
\end{aligned}$$

Finalmente, resolviendo

$$\begin{aligned}
& \exp\left\{-\frac{1}{2}\left(\frac{n\overline{x}^2\tau^2+\mu^2\sigma^2}{\sigma^2\tau^2}\right)+\frac{1}{2}\frac{(n\overline{x}\tau^2+\mu\sigma^2)^2}{\sigma^2\tau^2(n\tau^2+\sigma^2)}\right\} \\
= & \exp\left\{-\frac{1}{2\sigma^2\tau^2}\left((n\overline{x}^2\tau^2+\mu^2\sigma^2)-\frac{(n\overline{x}\tau^2+\mu\sigma^2)^2}{(n\tau^2+\sigma^2)}\right)\right\} \\
= & \exp\left\{-\frac{1}{2\sigma^2\tau^2}\left(\frac{n^2\overline{x}^2\tau^4+n\overline{x}^2\tau^2\sigma^2+\mu^2\sigma^2n\tau^2+\mu^2\sigma^4-n^2\overline{x}^2\tau^4-2n\overline{x}\tau^2\mu\sigma^2-\mu^2\sigma^4}{(n\tau^2+\sigma^2)}\right)\right\} \\
= & \exp\left\{-\frac{1}{2\sigma^2\tau^2}\left(\frac{n\overline{x}^2\tau^2\sigma^2-2n\overline{x}\tau^2\mu\sigma^2+\mu^2\sigma^2n\tau^2}{(n\tau^2+\sigma^2)}\right)\right\} \\
= & \exp\left\{-\frac{1}{2\sigma^2\tau^2}\left(\frac{\sigma^2\tau^2n(\overline{x}^2-2\overline{x}\mu+\mu^2)}{(n\tau^2+\sigma^2)}\right)\right\} \\
= & \exp\left\{-\frac{1}{2}\left(\frac{n(\overline{x}-\mu)^2}{n\tau^2+\sigma^2}\right)\right\},
\end{aligned}$$

tenemos que:

$$p(x|H_1) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left\{-\frac{nS^2}{2\sigma^2}\right\}\left(\frac{\sigma^2/n}{\tau^2+\sigma^2/n}\right)^{1/2}\exp\left\{-\frac{1}{2}\frac{(\overline{x}-\mu)^2}{(\tau^2+\sigma^2/n)}\right\}. \tag{3.9}$$

Por otro lado, con un procedimiento análogo al realizado en (3.5) se prueba que:

$$\begin{aligned}
p(x|\theta_0) &= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta_0)^2\right\} \\
&= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{nS^2}{2\sigma^2}\right\}.\exp\left\{-\frac{n}{2\sigma^2}(\overline{x}-\theta_0)^2\right\}. \qquad (3.10)
\end{aligned}$$

Luego, de (3.9) y (3.10) el factor de Bayes está dado por:

$$\begin{aligned}
BF(H_0:H_1) &= \frac{p(x|\theta_0)}{\int p(x|\theta)f(\theta)d\theta} \\
&= \frac{(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{nS^2}{2\sigma^2}\right\}.\exp\left\{-\frac{n}{2\sigma^2}(\overline{x}-\theta_0)^2\right\}}{\left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left\{-\frac{nS^2}{2\sigma^2}\right\}\left(\frac{\sigma^2/n}{\tau^2+\sigma^2/n}\right)^{1/2}\exp\left\{-\frac{1}{2}\frac{(\overline{x}-\mu)^2}{(\tau^2+\sigma^2/n)}\right\}} \\
&= \left(\frac{\tau^2+\sigma^2/n}{\sigma^2/n}\right)^{1/2}\exp\left\{\frac{1}{2}\frac{(\overline{x}-\mu)^2}{(\tau^2+\sigma^2/n)}-\frac{n}{2\sigma^2}(\overline{x}-\theta_0)^2\right\} \\
&= \left(\frac{\tau^2+\sigma^2/n}{\sigma^2/n}\right)^{1/2}\exp\left\{\frac{n}{2}\left(\frac{(\overline{x}-\mu)^2}{n\tau^2+\sigma^2}-\frac{(\overline{x}-\theta_0)^2}{\sigma^2}\right)\right\}.
\end{aligned}$$

Como se esperaba, el factor de Bayes sólo depende de la muestra a través de $\overline{x}$. Para obtener el valor de la muestra que maximice el factor de Bayes, es suficiente resolver el problema de maximización con respecto a $\overline{x}$. Tomando logaritmo en la expresión anterior tenemos:

$$\begin{aligned}
\log BF &= \log\left[\left(\frac{\tau^2+\sigma^2/n}{\sigma^2/n}\right)^{1/2}\exp\left\{\frac{n}{2}\left(\frac{(\overline{x}-\mu)^2}{n\tau^2+\sigma^2}-\frac{(\overline{x}-\theta_0)^2}{\sigma^2}\right)\right\}\right]\\
&= \log\left(\frac{\tau^2+\sigma^2/n}{\sigma^2/n}\right)^{1/2}+\log\left[\exp\left\{\frac{n}{2}\left(\frac{(\overline{x}-\mu)^2}{n\tau^2+\sigma^2}-\frac{(\overline{x}-\theta_0)^2}{\sigma^2}\right)\right\}\right]\\
&= \log\left(\frac{\tau^2+\sigma^2/n}{\sigma^2/n}\right)^{1/2}+\frac{n}{2}\left(\frac{(\overline{x}-\mu)^2}{n\tau^2+\sigma^2}-\frac{(\overline{x}-\theta_0)^2}{\sigma^2}\right).
\end{aligned}$$

Derivando con respecto a $\overline{x}$,

$$\frac{\partial \log BF}{\partial \overline{x}} = \frac{n}{2}\left(\frac{2(\overline{x}-\mu)}{n\tau^2+\sigma^2}-\frac{2(\overline{x}-\theta_0)}{\sigma^2}\right)$$

y

$$\begin{aligned}
\frac{\partial^2 \log BF}{\partial^2 \overline{x}} &= \frac{n}{2}\left(\frac{2}{n\tau^2+\sigma^2}-\frac{2}{\sigma^2}\right)\\
&= \frac{n}{2}\left(\frac{2\sigma^2-2\sigma^2-2n\tau^2}{\sigma^2(\sigma^2+n\tau^2)}\right)\\
&= \frac{n}{2}\left(-\frac{2n\tau^2}{\sigma^2(\sigma^2+n\tau^2)}\right)<0,
\end{aligned}$$

resolviendo la ecuación $\dfrac{\partial \log BF}{\partial \overline{x}} = 0$, obtenemos el máximo. Esto es;

$$\begin{aligned}
& \frac{n}{2}\left(\frac{2(\overline{x}-\mu)}{\sigma^2+n\tau^2}-\frac{2(\overline{x}-\theta_0)}{\sigma^2}\right)=0\\
\Rightarrow\ & \frac{2\sigma^2(\overline{x}-\mu)-2(\sigma^2+n\tau^2)(\overline{x}-\theta_0)}{\sigma^2(\sigma^2+n\tau^2)}=0\\
\Rightarrow\ & 2\sigma^2\overline{x}-2\sigma^2\mu-2\sigma^2\overline{x}+2\sigma^2\theta_0-2n\tau^2\overline{x}+2n\tau^2\theta_0=0\\
\Rightarrow\ & -2\sigma^2\mu+2\sigma^2\theta_0-2n\tau^2\overline{x}+2n\tau^2\theta_0=0\\
\Rightarrow\ & \overline{x}=\frac{2(\sigma^2\theta_0+n\tau^2\theta_0-\sigma^2\mu)}{2n\tau^2}\\
\Rightarrow\ & \overline{x}=\frac{\sigma^2\theta_0}{n\tau^2}+\theta_0-\frac{\sigma^2\mu}{n\tau^2}\\
\Rightarrow\ & \overline{x}=\frac{\sigma^2}{n\tau^2}(\theta_0-\mu)+\theta_0.
\end{aligned}$$

Así, la solución es

$$\overline{x}_{\text{máx}} = \theta_0 + \frac{\sigma^2}{n\tau^2}(\theta_0 - \mu)$$

y el valor máximo del factor de Bayes (sustituyendo $\overline{x}_{\text{máx}}$) es:

$$\begin{aligned}
BF(H_0 : H_1) &= \left(\frac{\tau^2 + \sigma^2/n}{\sigma^2/n}\right)^{1/2} \exp\left\{\frac{n}{2}\left(\frac{(\overline{x}-\mu)^2}{n\tau^2+\sigma^2} - \frac{(\overline{x}-\theta_0)^2}{\sigma^2}\right)\right\} \\
&= \left(\frac{n\tau^2+\sigma^2}{\sigma^2}\right)^{1/2} \exp\left\{\frac{n}{2}\left(\frac{\left(\theta_0 + \frac{\sigma^2}{n\tau^2}(\theta_0-\mu)-\mu\right)^2}{n\tau^2+\sigma^2} - \frac{\left(\theta_0 + \frac{\sigma^2}{n\tau^2}(\theta_0-\mu)-\theta_0\right)^2}{\sigma^2}\right)\right\} \\
&= \left(1+\frac{n\tau^2}{\sigma^2}\right)^{1/2} \exp\left\{\frac{n}{2}\left(\frac{\left((\theta_0-\mu)\left(\frac{\sigma^2}{n\tau^2}+1\right)\right)^2}{n\tau^2+\sigma^2} - \frac{\left(\frac{\sigma^2}{n\tau^2}(\theta_0-\mu)\right)^2}{\sigma^2}\right)\right\} \\
&= \left(1+\frac{n\tau^2}{\sigma^2}\right)^{1/2} \exp\left\{\frac{n}{2}\left(\frac{(\theta_0-\mu)^2\left(\frac{\sigma^2+n\tau^2}{n\tau^2}\right)^2}{n\tau^2+\sigma^2} - \frac{\left(\frac{\sigma^2}{n\tau^2}\right)^2(\theta_0-\mu)^2}{\sigma^2}\right)\right\} \\
&= \left(1+\frac{n\tau^2}{\sigma^2}\right)^{1/2} \exp\left\{\frac{n}{2}\left[(\theta_0-\mu)^2\left(\frac{\frac{(\sigma^2+n\tau^2)^2}{n^2\tau^4}}{n\tau^2+\sigma^2} - \frac{\frac{\sigma^4}{n^2\tau^4}}{\sigma^2}\right)\right]\right\} \\
&= \left(1+\frac{n\tau^2}{\sigma^2}\right)^{1/2} \exp\left\{\frac{n}{2}\left[(\theta_0-\mu)^2\left(\frac{\sigma^2+n\tau^2}{n^2\tau^4} - \frac{\sigma^2}{n^2\tau^4}\right)\right]\right\} \\
&= \left(1+\frac{n\tau^2}{\sigma^2}\right)^{1/2} \exp\left\{\frac{n}{2}\left(\frac{(\theta_0-\mu)^2}{n^2\tau^4}(n\tau^2)\right)\right\} \\
&= \left(1+\frac{n\tau^2}{\sigma^2}\right)^{1/2} \exp\left\{\frac{(\theta_0-\mu)^2}{2\tau^2}\right\} > 1.
\end{aligned}$$

Cuanto más grande es el tamaño de la muestra n, más grande las probabilidades de H_0 y más grande es τ^2, más grande será el máximo del factor de Bayes. La incertidumbre a priori juega un papel crucial en la comparación Bayesiana de la hipótesis nula. En el límite, cuando $\tau^2 \to +\infty$, el factor de Bayes $BF(H_0 : H_1)$ también crece indefinidamente. Este límite es el resultado que se obtiene cuándo se usa una priori no informativa de $\theta|H_1$. Esto fue notado por primera vez por *Lindley*(1957) y se le conoce como la paradoja de *Lindley*. Esto tiene por objeto el estudio de la comparación entre lo Bayesiano y el enfoque frecuentista de las hipótesis de prueba.

Para una comparación fácil entre los enfoques, tomemos $\mu = \theta_0$ y $\tau^2 = \sigma^2$. La primera suposición se centra en la distribución a priori alternativa sobre el valor simple de H_0

y la segunda suposición toma la varianza priori en la alternativa igual a la varianza observacional. Entonces,

$$\begin{aligned}
BF(H_0:H_1) &= \left(\frac{\tau^2+\sigma^2/n}{\sigma^2/n}\right)^{1/2}\exp\left\{\frac{n}{2}\left(\frac{(\overline{x}-\mu)^2}{n\tau^2+\sigma^2}-\frac{(\overline{x}-\theta_0)^2}{\sigma^2}\right)\right\}\\
&= \left(\frac{\sigma^2+\sigma^2/n}{\sigma^2/n}\right)^{1/2}\exp\left\{\frac{n}{2}\left(\frac{(\overline{x}-\theta_0)^2}{n\sigma^2+\sigma^2}-\frac{(\overline{x}-\theta_0)^2}{\sigma^2}\right)\right\}\\
&= \left(\frac{\sigma^2}{\sigma^2/n}+1\right)^{1/2}\exp\left\{\frac{n}{2}\left((\overline{x}-\theta_0)^2\left(\frac{1}{n\sigma^2+\sigma^2}-\frac{1}{\sigma^2}\right)\right)\right\}\\
&= (n+1)^{1/2}\exp\left\{\frac{n}{2}\left((\overline{x}-\theta_0)^2\left(-\frac{n\sigma^2}{\sigma^2(n\sigma^2+\sigma^2)}\right)\right)\right\}\\
&= (n+1)^{1/2}\exp\left\{\frac{n}{2}\left(-\frac{n(\overline{x}-\theta_0)^2}{\sigma^2(n+1)}\right)\right\}\\
&= (n+1)^{1/2}\exp\left\{-\frac{n}{2(n+1)}\;z^2\right\},
\end{aligned}$$

donde $z^2=\dfrac{n(\overline{x}-\theta_0)^2}{\sigma^2}$.
Si $p(H_0)=\pi$, entonces

$$\frac{p(H_0|x)}{p(H_1|x)}=\frac{p(H_0)\times p(x|H_0)}{p(H_1)\times p(x|H_1)}=\frac{\pi\times p(x|H_0)}{(1-\pi)\times p(x|H_1)}=\frac{\pi}{1-\pi}BF(H_0:H_1)$$

si y sólo si

$$p(H_0|x)=\left\{1+\left[\frac{\pi}{1-\pi}BF(H_0:H_1)\right]^{-1}\right\}^{-1}.$$

En efecto, de la ecuación (3.4) tenemos que

$$p(x|H_1)=\frac{p(x)-\pi p(x|\theta_0)}{1-\pi}.$$

Así,

$$BF(H_0:H_1)=\frac{p(x|\theta_0)}{p(x|H_1)}=\frac{p(x|\theta_0)}{\int p(x|\theta)f(\theta)d\theta}=\frac{p(x|\theta_0)(1-\pi)}{p(x)-\pi p(x|\theta_0)}.$$

Luego,

$$
\begin{aligned}
[BF(H_0 : H_1)]^{-1} &= \left[\frac{p(x|\theta_0)(1-\pi)}{p(x) - \pi p(x|\theta_0)}\right]^{-1} \\
\Rightarrow \quad \left[\frac{\pi}{1-\pi} BF(H_0 : H_1)\right]^{-1} &= \frac{(1-\pi)}{\pi} \times \frac{p(x) - \pi p(x|\theta_0)}{p(x|\theta_0)(1-\pi)} \\
\Rightarrow \quad \left[\frac{\pi}{1-\pi} BF(H_0 : H_1)\right]^{-1} &= \frac{p(x) - \pi p(x|\theta_0)}{\pi p(x|\theta_0)} \\
\Rightarrow \quad \left[\frac{\pi}{1-\pi} BF(H_0 : H_1)\right]^{-1} &= \frac{p(x)}{\pi p(x|\theta_0)} - 1 \\
\Rightarrow \quad \left[\frac{\pi}{1-\pi} BF(H_0 : H_1)\right]^{-1} + 1 &= \frac{p(x)}{\pi p(x|\theta_0)} \\
\Rightarrow \quad \frac{\pi p(x|\theta_0)}{p(x)} &= \left\{1 + \left[\frac{\pi}{1-\pi} BF(H_0 : H_1)\right]^{-1}\right\}^{-1} \\
\Rightarrow \quad p(H_0|x) = \frac{p(H_0)p(x|\theta_0)}{p(x)} &= \left\{1 + \left[\frac{\pi}{1-\pi} BF(H_0 : H_1)\right]^{-1}\right\}^{-1}.
\end{aligned}
$$

Asumiendo que las prioris son iguales ($p(H_0) = p(H_1) = 1/2$) tenemos;

$$
p(H_0|x) = \left\{1 + \left[(n+1)^{1/2} \exp\left\{-\frac{n}{n+1}\frac{z^2}{2}\right\}\right]^{-1}\right\}^{-1}.
$$

La probabilidad posteriori de H_0 se puede calcular para diferentes valores de n y z ya que ambas pruebas la clásica y la Bayesiana están basadas en z^2. Trabajando con los valores más comunes, asociados con los $p - valores$ 0,05 y 0,01 dados en la tabla de la $N(0,1)$.

Las probabilidades de H_0 son más pequeñas para valores muy grandes de $|z|$ lo cual es razonable. Lo que es menos razonable es el valor que se obtiene para estas probabilidades pero refuerzan la idea que los $p - valores$ podrían no ser tomados como probabilidades asociadas con H_0. Otro resultado interesante es que para algún valor de z dado, la probabilidad posteriori puede variar sustancialmente desde un valor muy bajo que conduciría al rechazo de H_0 a un valor muy alto que conduciría a aceptar a H_0. Una posible forma de conciliar estos resultados con el nivel de significancia es que el nivel podría ser cambiado con el tamaño de la muestra. Grandes tamaños de muestras deberían reducir el nivel y viceversa. Una discusión más detallada del tema se puede hallar en *Berger* y *Sellke*(1987).

§3.3. Prueba de hipótesis e intervalos de confianza

Es notable la fuerte conexión entre la prueba de hipótesis y el intervalo de confianza. Esta conexión es más clara con el enfoque frecuentista pero es también relevante bajo el enfoque Bayesiano como veremos más adelante. Comencemos por el enfoque clásico, considere una prueba de nivel α para hipótesis $\theta = \theta_0$. Supongamos que, de acuerdo al tamaño dado de la muestra, la prueba produce una región de aceptación $A = A(\theta_0)$ donde es importante denotar explícitamente la dependencia de la región sobre θ_0. Así,

$$P(X \in A(\theta_0)|\theta_0) = 1 - \alpha.$$

Variando el valor de θ_0 en Θ generamos una familia de regiones Θ, todas dependiendo de θ y con probabilidad $1 - \alpha$. Esto automáticamente implica que, luego de observar el valor de x para X, la región

$$\{\theta : x \in A(\theta)\}$$

tendrá una confiabilidad de $1 - \alpha$.

Ejemplo 3.4. *Sea $X = (x_1, ..., x_n)$ una muestra aleatoria de la distribución $N(\theta, \sigma^2)$ con σ^2 conocido. La prueba insesgada uniformemente más potente de nivel α del contraste de $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ tiene región de aceptación:*

$$\left\{X : \theta_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \theta_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}.$$

Este intervalo puede ser reescrito como:

$$\left\{X : \overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \theta_0 \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}.$$

Reemplazando ahora θ_0 por θ dado el nivel de confianza $1 - \alpha$ para θ, tenemos:

$$\left\{\theta : \overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \theta \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}.$$

Por supuesto, este es intervalo de confianza para θ que se esperaba obtener.

Ejemplo 3.5. *Tomando de nuevo la muestra aleatoria* $X = (x_1, ..., x_n)$ *para la distribución* $N(\theta, \sigma^2)$ *con* σ^2 *conocido y priori* $\theta \sim N(\mu, \tau^2)$ *obtenemos la posteriori* $p(\theta|x) \propto p(x|\theta)p(\theta)$*, la cual es* $N(\mu_1, \tau_1^2)$ *donde* $\mu_1 = \dfrac{\overline{x}\tau^2 + \sigma^2\mu}{\tau^2 + \sigma^2}$ *y* $\tau_1 = \dfrac{\sigma^2\tau^2}{\tau^2 + \sigma^2}$*. Así, tomando el cuantil* $z_{\alpha/2}$ *de la* $N(0,1)$*, tenemos que:*

$$P\left(-z_{\alpha/2} \leq \frac{\theta - \mu_1}{\tau_1} \leq z_{\alpha/2}\right) = 1 - \alpha$$

esto implica que

$$P(\mu_1 - \tau_1 z_{\alpha/2} \leq \theta \leq \mu_1 + \tau_1 z_{\alpha/2}) = 1 - \alpha.$$

Por lo tanto, el intervalo de confianza del $100(1-\alpha)\,\%$ *para* θ *esta dado por:* $[\mu_1 - \tau_1 z_{\alpha/2}, \mu_1 + \tau_1 z_{\alpha/2}]$*. La hipótesis* $\theta = \theta_0$ *se puede aceptar si* θ_0 *pertenece al intervalo anterior. En el caso no-informativo,* $p(\theta) \equiv$ *constante por lo que* $\theta|x \sim N\left(\overline{x}, \dfrac{\sigma^2}{n}\right)$*, es decir,* $\mu_1 \to \overline{x}$ *y* $\tau_1 \to \frac{\sigma}{\sqrt{n}}$ *y la hipótesis es aceptada si* θ_0 *pertenece al intervalo:*

$$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

coincidiendo con la prueba clásica.

En la sección anterior, la prueba Bayesiana de una hipótesis simple de la forma $\theta = \theta_0$ fue definida para atribuir una probabilidad priori π a esta hipótesis y el $1 - \pi$ restante fue distribuido de acuerdo a algunas distribuciones de probabilidad de $\theta|H_1$. Esta especificación es criticada por dar algunas veces, estatus diferentes a los valores de θ_0. Esta discusión es de hecho más amplia y va más allá de los límites del pensamiento Bayesiano. Esto tiene que ver con la capacidad de una hipótesis simple para representar adecuadamente la situación bajo estudio. Algunos autores sugieren que estas hipótesis son siempre a lo sumo una aproximación útil de la hipótesis más realista donde θ pertenece actualmente a una vecindad de θ_0.

Conclusión

La estadística bayesiana es una alternativa a la estadística clásica o frecuentistas para la solución de problemas típicos como estimación, contraste de hipótesis y predicción.
La teoría bayesiana plantea la solución a un problema estadístico desde el punto de vista subjetivo (densidad o probabilidad a priori) el cual está basado en opiniones e información acerca del proceso; esto es una desventaja pues algunos investigadores rechazan que la información inicial se incluya en un proceso de inferencia científica. Sin embargo, esta situación se puede evitar introduciendo una distribución a priori no informativa o de referencia.
Así, la diferencia entre ambos enfoques radica precisamente en que el enfoque clásico se basa únicamente en los datos observados, mientras que en el enfoque bayesiano combina los datos observados con opiniones subjetivas.
Finalmente, el buen tratamiento de una problema, ya sea a través de métodos frecuentistas o bayesianos, va a depender en gran medida, de la mucha o poca honestidad del investigador a la hora de diseñar el estudio, incorporar los datos, analizarlos y obtener los resultados.
En general, los métodos bayesianos permiten llegar a algunas conclusiones que resultan más intuitivas y cercana al sentido común que las que se obtienen, algunas veces, de los métodos frecuentistas.

Referencias

[1] Anderson, T.W.(1958). An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons.

[2] Anderson, Hair, Tatham y Black. (2005). Análisis Multivariante. 5ta. Edición. Pearson Prentice Hall.

[3] Berger J.O. (1985). Statistical Decision Theory and Bayesian Analysis. 2nd ed. New York: Springer.

[4] Bernardo, J.M. y Smith A.F.M. (1994). Bayesian Theory. New York: Wiley.

[5] Box, G.E.P y Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis. Addison-Wesley Publishing Co.

[6] Bradley, P.C. y Thomas, A.L. (1996). Bayes and Empirical Bayes Methods for Data Analysis. Chapman & Hall.

[7] Casella, G. y Berger, R.L. (1990). Statistical Inference. Thomson.

[8] Congdon Peter (2001). Statistical Modelling. New York: John Wiley and Sons.

[9] Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press.

[10] Johnson, R. y Wichern, D. (1998). Applied Multivariate Statistical Analysis. 4ta. Edition. Prentice Hall.

[11] Migon and Gamerman (1999). Statistical Inference: An Integrated Approach. New York.

[12] Press, James S. (1982). Applied Multivariate Analysis: using Bayesian and Frequentist Methods of Inference. Second Edition. Malabar, Florida.