UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

Un enfoque integral para la inferencia estadística paramétrica

TESIS

QUE PARA OBTENER EL TÍTULO DE LICENCIADA EN ACTUARÍA

PRESENTAN:

PRISCILA ANAHÍ GARCÍA FERRO

VERÓNICA MARIEL HERNÁNDEZ MARTÍNEZ

ASESOR: DR. ARTURO ERDELY RUIZ

AGOSTO 2016

SANTA CRUZ, ACATLÁN, NAUCALPAN, EDO. DE MÉX.

Nunca consideres el estudio como una obligación,
sino como una oportunidad para penetrar en
el bello y maravilloso mundo del saber

Albert Einstein

Agradecimientos.

A nuestros padres quienes nos apoyaron en todo momento e hicieron este sueño posible.

A nuestros maestros gracias por los conocimientos transmitidos, su paciencia y apoyo

A nuestro asesor de tesis el Dr. Arturo Erdely Ruiz quien fue nuestra inspiración para realizar este trabajo, gracias por su tiempo, por compartir con nosotras sus conocimientos, por su apoyo, su confianza y por su amistad.

A nuestra querida Universidad.

Índice general

Introducción

¿Es posible desarrollar la teoría básica de la inferencia estadística paramétrica bajo un enfoque integral, sin hacer una separación entre enfoque clásico y bayesiano?

La estadística inferencial es la parte de la estadística que se encarga de estudiar procedimientos para la obtención de conclusiones a partir de la información parcial, típicamente la que se encuentra en una muestra aleatoria.

La inferencia paramétrica resuelve los problemas de las distribuciones de probabilidad cuando ésta es conocida en su totalidad, salvo por uno o varios parámetros.

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

Se parte de la premisa de que se cuenta ya con conocimientos sobre la teoría básica de probabilidades, particularmente sobre variables y vectores aleatorios y sus transformaciones, sin necesidad de antecedente alguno sobre algún enfoque de la estadística, y que lo anterior es suficiente para poder abordar la teoría básica de la inferencia estadística paramétrica de manera integral.

Usualmente, los dos enfoques se abordan de manera separada, por lo que este trabajo pretende presentar de manera integral, original, estructurada y desarrollada, un solo enfoque, despues del análisis de textos esenciales sobre inferencia estadística parámetrica bajos los enfoques clásico y bayesiano, y otros más bajo enfoques integrales, además de analizar diversos ejemplos de aplicación y se desarrollan los que resulten idóneos para aplicar la teoría bajo el enfoque integral propuesto.

Con este trabajo, podemos ver que el estudio de la estadística puede ser en un enfoque integral, que posiblemente facilite el entendimiento de esta área de gran importancia, sin necesidad de hacer enfásis en las diferencias conceptuales y metodológicas de ambos enfoques.

La gran mayoría de los textos sobre inferencia estadística paramétrica (teoría y aplicaciones) para el nivel licenciatura de una carrera como Actuaría abordan solamente (o predominantemente) el enfoque clásico de la estadística, en comparación con aquellos que lo hacen bajo el enfoque bayesiano. Más raros aún son aquellos que abordan ambos enfoques de una forma integral o equilibrada, tal es el caso de los títulos Statistical Inference An Integrated Approach y Principles of Statistical Inference.

Este trabajo pretende ser de mucha ayuda para los interesados en el área, ya que no se cuenta con mucho material en idioma español. Por tal motivo, es de nuestro interés presentar un escrito que muestre de manera clara y de fácil entendimiento los conceptos de la inferencia estadística paramétrica.

Se abordan diferentes problemas que tienen soluciones estadísticas, mostrando los pasos y justificando su desarrollo. Desde ejemplos muy sencillos hasta algunos más complejos, con la intención de ilustrar los temas para un mejor entendimiento.

Capítulo 1

Inferencia estadística paramétrica

1.1. Familias paramétricas

Una familia de funciones de densidad o de masa de probabilidades (fdp o fmp) se denomina **familia exponencial** si puede expresarse de la siguiente manera:

$$f(x \mid \theta) = h(x)\, c(\theta) \exp\left(\sum_{i=1}^{k} w_i(\theta)\, t_i(x) \right),$$

donde: $h(x) \geq 0$, t_i son funciones reales que no dependen de θ, $c(\theta) \geq 0$, w_i funciones reales que no dependen de x, veáse [2].

Ejemplo: Consideremos una variable aleatoria (v.a) binomial(n, θ), n conocida, θ desconocida, $0 < \theta < 1$, con función de densidad:

$$f(x \mid n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} I_{\{0,1,\dots,n\}}(x).$$

Entonces con $k = 1$, $w(\theta) = \log(\frac{\theta}{1-\theta})$, $t(x) = x$ y $e^{x \log(\frac{\theta}{1-\theta})}$, se tiene que el modelo binomial(n, θ) con n conocida, es familia exponencial.

Si n es desconocida en este modelo, no es familia exponencial ya que no podríamos separar la función indicadora ni el coeficiente binomial.

El modelo clásico de regresión lineal simple, bajo el supuesto de normalidad, es decir que $Y_i \sim \text{Normal}(\alpha + \beta x_i, \sigma^2)\,, i = 1, ..., n$ con función de densidad de probabilidad:

$$f_{Y_i}(y_i \,|\, \alpha + \beta x_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{ \frac{-(y_i - (\alpha + \beta x_i))^2}{2\sigma^2} \right\}$$

también es una familia exponencial con $k = 3$, $\theta = (\alpha, \beta, \sigma^2)$ y:

$$h(y_i) = \frac{1}{\sqrt{2\pi}}, \qquad c(\alpha, \beta, \sigma^2) = \frac{1}{\sqrt{\sigma^2}} exp\left(- \frac{(\alpha + \beta x_i)^2}{2\sigma^2} \right)$$

$$w_1(\alpha, \beta, \sigma^2) = \frac{\alpha}{\sigma^2}, \qquad t_1(y_i) = y_i$$

$$w_2(\alpha, \beta, \sigma^2) = \frac{\beta}{\sigma^2}, \qquad t_2(y_i) = x_i y_i$$

$$w_3(\alpha, \beta, \sigma^2) = -\frac{1}{2\sigma^2}, \quad t_3(y_i) = y_i^2$$

Familia Paramétrica	**Parámetros**
Bernoulli	p
Binomial	n conocida, p desconocida
Geométrica	p
Binomial Negativa	r conocida, p
Poisson	λ
Beta	α, β
Exponencial	β
Gamma	α, β
Normal	μ, σ^2
Weibull	γ, β conocida

Cuadro 1.1: Algunas familias exponenciales.

Sea f_Z una función de densidad de una v.a Z. Sea la v.a $X = \alpha + \beta Z$ con $\alpha \in \mathbb{R}$, $\beta > 0$

$$F_X(x) = P(X \leq x) = P(\alpha + \beta Z \leq x)$$

$$= P\left(Z \leq \frac{x-\alpha}{\beta}\right) = F_Z\left(\frac{x-\alpha}{\beta}\right)$$

$$\Rightarrow f_X(x|\alpha, \beta) = \frac{1}{\beta} f_Z\left(\frac{x-\alpha}{\beta}\right)$$

es la **familia de localización y escala** construida a partir de la densidad f_Z, que se denomina densidad estándar de la familia. α es el parámetro de localización y β es el parámetro de escala, veáse [2].

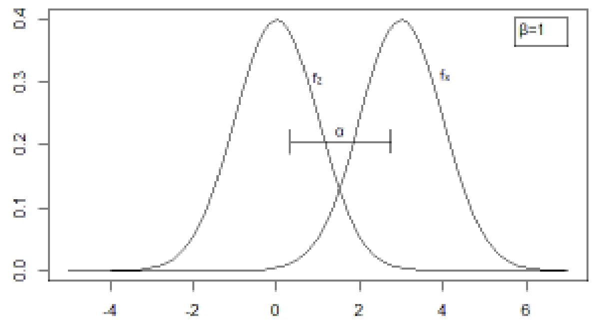

Figura 1.1: Parámetro de localización.

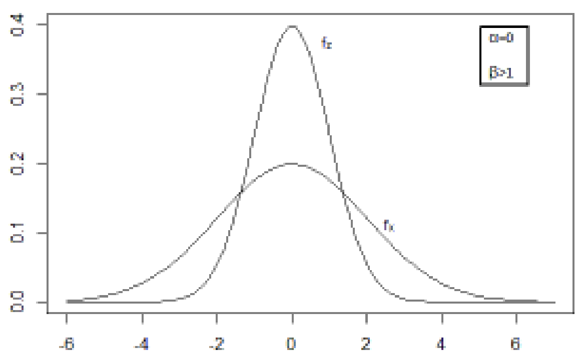

Figura 1.2: Parámetro de escala.

En las figuras 1.1 y 1.2, se observa el efecto de los parámetros de localizacón y escala respectivamente.

Ejemplos: Se tiene

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} exp\left(\frac{-z^2}{2}\right) \sim \mathrm{N}(0,1)$$

Si definimos a $X := \beta\, Z + \alpha$ entonces

$$F_X(x) = P(X \leq x) = P(\alpha+\beta\, Z \leq x) = P\left(Z \leq \left(\frac{X-\alpha}{\beta}\right)\right) = F_Z\left(\frac{X-\alpha}{\beta}\right).$$

Al derivar obtenemos que

$$f_X(x) = \frac{1}{\beta} f_Z\left(\frac{X-\alpha}{\beta}\right).$$

Entonces

$$f_X(x \,|\, \alpha, \beta) = \frac{1}{\sqrt{2\,\pi}\,\beta} exp\left[-\frac{1}{2}\left(\frac{x-\alpha}{\beta}\right)^2\right] \sim N(\alpha, \beta^2).$$

De manera que, el modelo normal con los parámetros usuales α, β^2 es una familia de localización y escala.

Ahora supongamos que se tiene

$$f_Z(z) = \frac{1}{B(\gamma,\omega)} z^{\gamma-1}(1-z)^{\omega-1} I_{\{0\,\leq\, z\,\leq\, 1\}} \sim \text{Beta}(\gamma, \omega)$$

donde

$$B(\gamma,\omega) = \frac{\Gamma(\gamma)\Gamma(\omega)}{\Gamma(\gamma+\omega)}.$$

Si definimos a $X := \beta\, Z + \alpha$ entonces

$$F_X(x) = P(X \leq x) = P(\alpha+\beta\, Z \leq x) = P\left(Z \leq \left(\frac{X-\alpha}{\beta}\right)\right) = F_Z\left(\frac{X-\alpha}{\beta}\right).$$

derivando obtenemos que

$$f_X(x) = \frac{1}{\beta} f_Z\left(\frac{X-\alpha}{\beta}\right)$$

entonces

$$f_X(x \,|\, \alpha, \beta) = \frac{1}{\beta B(\gamma,\omega)} \left(\frac{x-\alpha}{\beta}\right)^{\gamma-1} \left(1 - \frac{x-\alpha}{\beta}\right)^{\omega-1} I_{\{\alpha\,\leq\, x\,\leq\, \beta+\alpha\}}$$

La distribución de probabilidad así obtenida para X se conoce como beta de cuatro parámetros. Ver Figura 1.3.

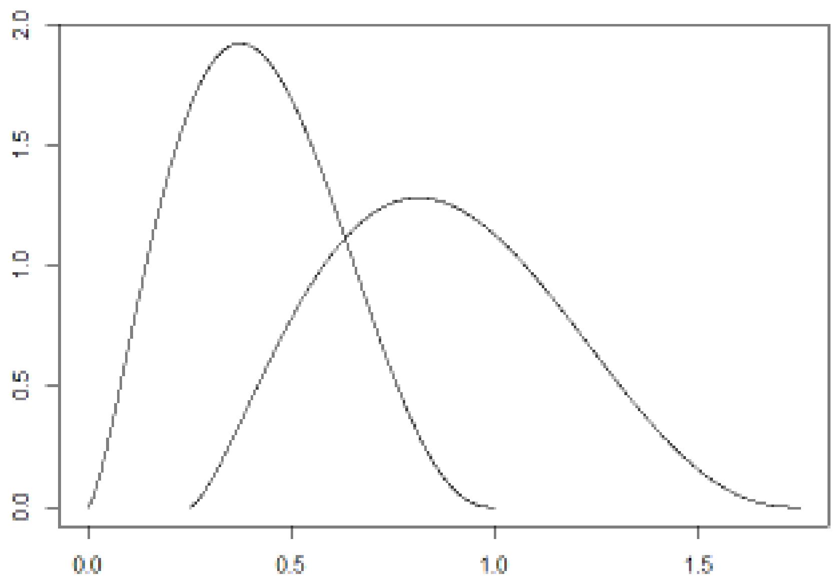

Figura 1.3: Efecto de localización y escala en el modelo beta.

1.2. Muestra aleatoria y sus transformaciones

Una **muestra aleatoria** (m.a) es una sucesión finita $X_1, ..., X_n$ de variables aleatorias independientes e idénticamente distribuidas, y a los valores que reporte se les conoce como muestra aleatoria observada, y la denotamos $\mathbf{x} = (x_1, x_2, ..., x_n)$, veáse [8].

Las muestras aleatorias se obtienen con la intención de inferir propiedades de la totalidad de la población, para lo cual deben ser representativas de la misma. Para cumplir esta característica, la inclusión de sujetos en la muestra debe seguir alguna técnica de muestreo.

Un **estadístico** es una variable o vector aleatorio resultante de transformar una muestra aleatoria, que pretende ser una herramienta útil para hacer inferencia en relación a la población de interés. Por ejemplo, sea una función $t : \mathbb{R}^n \to \mathbb{R}$ y definamos la v.a $T := t(X_1, X_2, ..., X_n)$. Entonces T es un estadístico, y tendrá alguna distribución de probabilidad, que eventualmente puede obtenerse explícitamente o bien aproximarse, veáse [8].

Nombre del estadístico	Fórmula matemática
Media muestral	$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$
Varianza muestral	$S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
Mediana muestral	$X_{(\frac{n+1}{2})}$ si n es impar, $\frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}$ si n es par
Rango muestral	$X_{(n)} - X_{(1)}$
Mínimo	$X_{(1)} := \text{mín}\{X_1, ..., X_n\}$
Máximo	$X_{(n)} := \max\{X_1, ..., X_n\}$

Cuadro 1.2: Algunos estadísticos comunes.

En el Cuadro 1.2 se presentan algunos de los estadísticos más utilizados.

1.3. Paradigma general

Supongamos que se tiene una m.a observada $\mathbf{x} = (x_1, ..., x_n)$, proveniente de una familia paramétrica $f(x \,|\, \theta)$ con parámetro desconocido $\theta \in \Theta$. Típicamente lo que se busca es hacer inferencia acerca de una observación futura, es decir predicción, para así hacer cálculo de probabilidades de interés, mediante:

$$P(X_{n+1} \in A) = \int_A f(x \,|\, \theta)\, dx.$$

Como el parámetro θ es desconocido, procedemos probabilísticamente de la siguiente manera:

$$\wp(x \,|\, \mathbf{x}) = \int_\Theta \wp(x, \theta \,|\, \mathbf{x})\, d\theta = \int_\Theta f(x \,|\, \theta) \wp(\theta \,|\, \mathbf{x})\, d\theta \tag{1.1}$$

de donde $\wp(\theta \,|\, \mathbf{x})$, se obtiene por medio de la regla de Bayes, veáse [6]:

$$\wp(\theta \,|\, \mathbf{x}) = \frac{\wp(\mathbf{x} \,|\, \theta) \wp(\theta)}{\int_\Theta \wp(\mathbf{x} \,|\, \vartheta) \wp(\vartheta)\, d\vartheta} \tag{1.2}$$

La ecuación (1.1) se conoce como distribución predictiva final y la ecuación (1.2) es conocida como distribución final.

Notemos que $L(\theta \,|\, \mathbf{x}) = \wp(\mathbf{x} \,|\, \theta)$ la cual se conoce como función de máxima verosimilitud.

$\wp(\theta)$ es la distribución inicial, basada en información previa, veáse [7]. El denominador es una constante, que en ocasiones resulta difícil calcular de manera explícita, especialmente si la dimensión del espacio paramétrico es mayor a 1. Sin embargo, existen métodos numéricos y de Monte Carlo para su aproximación.

Notemos que

$$L(\theta \,|\, \mathbf{x}) = \prod_{i\,=\,1}^{n} f(x_i \,|\, \theta).$$

Observemos que la distribución final contempla la información muestral así como la información inicial. Una vez hallada ya es posible plantear cualquier problema de inferencia en relación al parámetro desconocido θ y/o a una observación futura.

Ejemplo: Si tenemos una m.a $X_1, ..., X_n \sim$ Bernoulli(θ), $0 < \theta < 1$ con una función de densidad $f(x \,|\, \theta) = \theta^{x}(1-\theta)^{1-x}\, I_{(0\,,\,1)}(x)$, entonces

$$L(\theta \,|\, \mathbf{x}) = \prod_{i\,=\,1}^{n} f(x_i \,|\, \theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}.$$

Supongamos que contamos con información inicial que puede traducirse en las siguientes ecuaciones:

$$\text{Mediana}(\theta) = \theta_0 \tag{1.3}$$

$$P(\theta_1 < \theta < \theta_2) = \gamma \tag{1.4}$$

con θ_0, θ_1, θ_2 y γ conocidos.

Como θ pertenece al intervalo (0,1), cualquier modelo que tenga por lo menos dos parámetros y soporte en (0,1) nos podría resultar útil para incorporar la información inicial de las ecuaciones (1.3) y (1.4).

Por ejemplo, con el modelo Beta(α, β) tendríamos como distribución inicial:

$$\wp_1(\theta \,|\, \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\, I_{(0\,,\,1)}(\theta),$$

donde los parámetros α y β tendrían que calibrarse de acuerdo con la información inicial de las ecuaciones (1.3) y (1.4).

Al utilizar (1.3) tenemos:

$$F_{\wp_1}(\theta_0 \,|\, \alpha, \beta) = \frac{1}{2} \Rightarrow \theta_0 = F_{\wp_1}^{-1}\left(\frac{1}{2} \,\middle|\, \alpha, \beta\right)$$

$$F_{\wp_1}(\theta_0 \,|\, \alpha, \beta) - \frac{1}{2} = 0$$

Por (1.4),

$$F_{\wp_1}(\theta_2 \,|\, \alpha, \beta) - F_{\wp_1}(\theta_1 \,|\, \alpha, \beta) = \gamma$$
$$F_{\wp_1}(\theta_2 \,|\, \alpha, \beta) - F_{\wp_1}(\theta_1 \,|\, \alpha, \beta) - \gamma = 0.$$

Al combinar las ecuaciones anteriores se tiene:

$$h_1(\alpha, \beta) := \left[F_{\wp_1}(\theta_0 \,|\, \alpha, \beta) - \frac{1}{2} \right]^2 + \left[F_{\wp_1}(\theta_2 \,|\, \alpha, \beta) - F_{\wp_1}(\theta_1 \,|\, \alpha, \beta) - \gamma \right]^2.$$

Basta minimizar $h_1(\alpha, \beta)$ para encontrar los valores de α y β que reflejen la información inicial de las ecuaciones (1.3) y (1.4).

Realizaremos la construcción de otro modelo que puede funcionar como distribución inicial, el cual es conocido como distribución Kumaraswamy.

Sea $Y \sim \text{Beta}(1, b)$, $W = Y^{1/a}$, $0 < W < 1$ con $a > 0$ y $b > 0$. Entonces,

$$F_W(w) = P(W \leq w) = P(Y^{1/a}) = P(Y \leq w^{\,a}) = F_Y(w^{\,a}).$$

dado que $Y \sim \text{Beta}(1, b)$ tenemos que

$$f_Y(y \,|\, 1, b) = \frac{\Gamma(1+b)}{\Gamma(1)\Gamma(b)} \; y^0 \,(1-y)^{\,b-1} = \; b\,(1-y)^{\,b-1}\, I_{(0\,,1)}(y).$$

Entonces:

$$f_W(w \,|\, a, b) = a\, w^{\,a-1} f_Y(w^{\,a}) = a\, w^{\,a-1}\, b\,(1 - w^{\,a})^{\,b-1}\, I_{(0,1)}(w^{\,a})$$
$$= a\, b\, w^{\,a-1}(1 - w^{\,a})^{\,b}\, I_{(0\,,1)}(w).$$

Al obtener la función de distribución de W,

$$F_W(w \,|\, a,b) = \int_{-\infty}^{w} f_W(t \,|\, a,b)\, dt = \int_0^w a\, b\, t^{\,a-1} d\,t = 1 - (1 - w^{\,a})^{\,b}.$$

Si se hace $u = 1 - (1 - w^{\,a})^{\,b},$ entonces $w = \left[1 - (1-u)^{\,1/b}\right]^{\,1/a}$.

Por lo tanto:

$$F_W^{-1}(u \,|\, a,b) = \left[1 - (1-u)^{\,1/b}\right]^{\,1/a},\ 0 < u < 1.$$

Con $\wp_2(\theta \,|\, a,b) \sim$ Kumaraswamy(a,b),

Por (1.3):

$$F_{\wp_2}^{-1}\left(\frac{1}{2} \,\middle|\, a,b\right) = \theta_0 \Longleftrightarrow \left(1 - \frac{1}{2^{1/b}}\right)^{1/a} = \theta_0$$

$$\left(1 - \frac{1}{2^{1/b}}\right)^{1/a} - \theta_0 = 0.$$

Al aplicar (1.4)

$$F_{\wp_2}(\theta_2 \,|\, a,b) - F_{\wp_2}(\theta_1 \,|\, a,b) = \gamma \Longleftrightarrow (1 - {\theta_1}^{a})^{\,b} - (1 - {\theta_2}^{a})^{\,b} = \gamma.$$

$$(1 - {\theta_1}^{a})^{\,b} - (1 - {\theta_2}^{a})^{\,b} - \gamma = 0.$$

$$h_2(a,b) := \left[\left(1 - \frac{1}{2^{1/b}}\right)^{1/a} - \theta_0\right]^2 + \left[(1 - {\theta_1}^{a})^{\,b} - (1 - {\theta_2}^{a})^{\,b} - \gamma\right]^2.$$

Ahora se minimizará $h_2(a,b)$.

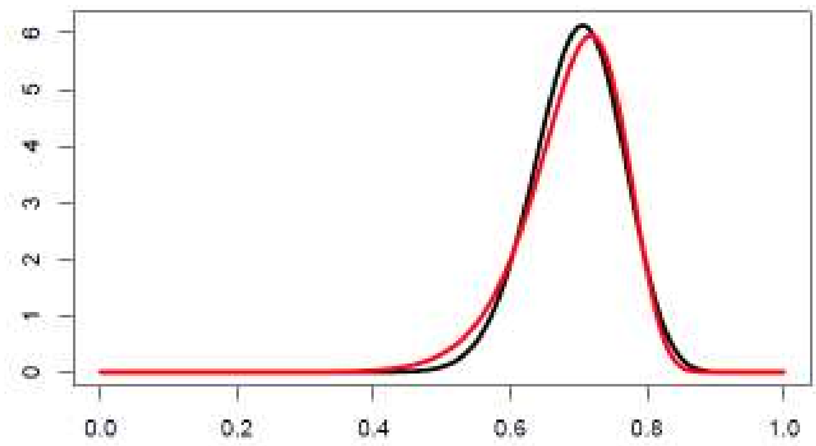

Figura 1.4: Comparación de densidades inicial, Beta en negro y Kumaraswamy en rojo

En la Figura (1.4) se observa que ambos modelos pueden reflejar la información inicial.

Al obtener la distribución final (1.2) con $\wp_1$:

$$L(\theta \,|\, \mathbf{x})\wp_1(\theta) \propto \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\theta^{\alpha-1}(1-\theta)^{\beta-1}\, I_{(0\,,1)}(\theta)$$
$$= \theta^{\alpha+\sum x_i -1}(1-\theta)^{\beta+n-\sum x_i -1} \sim \text{Beta}\bigg(\alpha + \sum x_i, \beta + n - \sum x_i\bigg).$$

Por lo tanto $\wp_1(\theta \,|\, \mathbf{x}) \sim \text{Beta}\bigg(\alpha + \sum x_i, \beta + n - \sum x_i\bigg)$.

Si quisiéramos desvanecer la influencia de la información inicial, obtendríamos la distribución final no informativa, en cuyo caso $\alpha \to 0$ y $\beta \to 0$, es decir:

$$\wp_1^*(\theta \,|\, \mathbf{x}) \sim \text{Beta}\bigg(\sum x_i, n - \sum x_i\bigg).$$

Ahora al obtener la distribución final (1.2) con $\wp_2$,

$$L(\theta\,|\,\mathbf{x})\wp_2(\theta) \propto \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\theta^{a-1}(1-\theta^{\,a})^{b-1}I_{(0\,,\,1)}(\theta)$$

$$= \theta^{a+\sum x_i-1}(1-\theta)^{n-\sum x_i}(1-\theta)^{b-1}\,I_{(0\,,\,1)}(\theta).$$

Nótemos que $\propto$ hace referencia a la proporcionalidad, y resulta más fácil trabajar con ella, ya que al eliminar las constantes se hace los cálculos más sencillos.

Para obtener el denominador es necesario resolver la siguiente integral:

$$K = \int_\Theta \theta^{(a+\sum x_i-1)}(1-\theta)^{n-\sum x_i}(1-\theta)^{b-1}\,I_{(0\,,\,1)}(\theta)d\theta.$$

Es importante observar que no es posible resolver la expresión anterior de forma explícita, por lo cual, se utilizó la herramienta de programación R para su aproximación véase [9], dando como resultado $K = 3.905823e-31$ (véase A.1).

Por lo cual

$$\wp_2(\theta\,|\,\mathbf{x}) = \theta^{a+\sum x_i-1}(1-\theta)^{n-\sum x_i}(1-\theta)^{b-1}\,I_{(0\,,\,1)}(\theta)(K)^{-1}.$$

Ejemplo: Tenemos una m.a $X_1, ..., X_n \sim$ Uniforme$(0, \theta)$, $\theta > 0$, cuya función de densidad es $f(x\,|\,\theta) = \frac{1}{\theta}\,I_{\{0\,<\,x\,<\,\theta\}}$, entonces:

$$L(\theta\,|\,\mathbf{x}) = \prod_{i\,=\,1}^{n} f(x_i\,|\,\theta) = \frac{1}{\theta^n}\,I_{\{\theta\,>\,x_{(n)}\}}.$$

Suponiendo que la información inicial que se tenga puede expresarse mediante una distribución de probabilidad Pareto con parámetros α y β calibrados de acuerdo con ello:

$$\wp(\theta \,|\, \alpha, \beta) = \frac{\beta\alpha^{\beta}}{\theta^{\beta+1}}\, I_{\{\theta > \alpha\}} \text{ con } \alpha, \beta > 0.$$

Entonces:

$$L(\theta \,|\, \mathbf{x})\, \wp(\theta) = \frac{\beta\alpha^{\beta}}{\theta^{\beta+n+1}}\, I_{\{\theta > \alpha \vee x_{(n)}\}}.$$

Luego:

$$\int_{\Theta} L(\theta \,|\, \mathbf{x})\, \wp(\theta)\, d\theta = \frac{\beta\alpha^{\beta}}{(\beta+n)(\alpha \vee x_{(n)})^{\beta+n}}.$$

Al obtener la distribución final (1.2),

$$\wp(\theta \,|\, \mathbf{x}) = \frac{(\beta+n)(\alpha \vee x_{(n)})^{\beta+n}}{\theta^{\beta+n+1}}\, I_{\{\theta > \alpha \vee x_{(n)}\}} \sim \text{Pareto}\big(\alpha \vee x_{(n)},\, \beta+n\big).$$

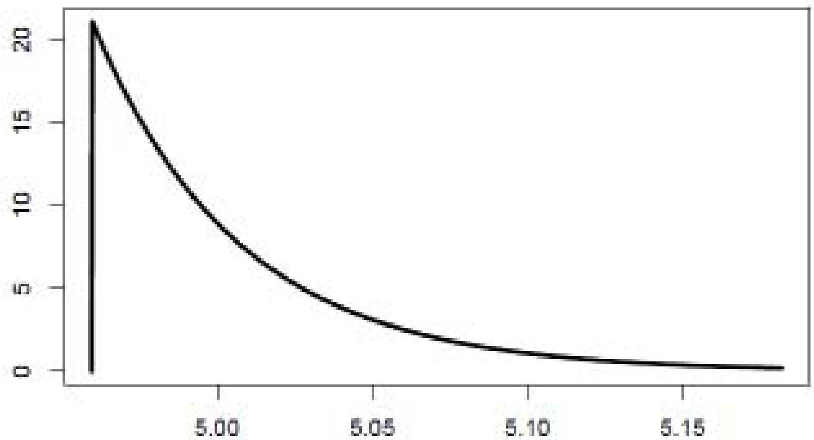

Figura 1.5: Distribución final

Y no informativa $\wp^*(\theta \,|\, \mathbf{x}) \sim \text{Pareto}(x_{(n)},\, n)$

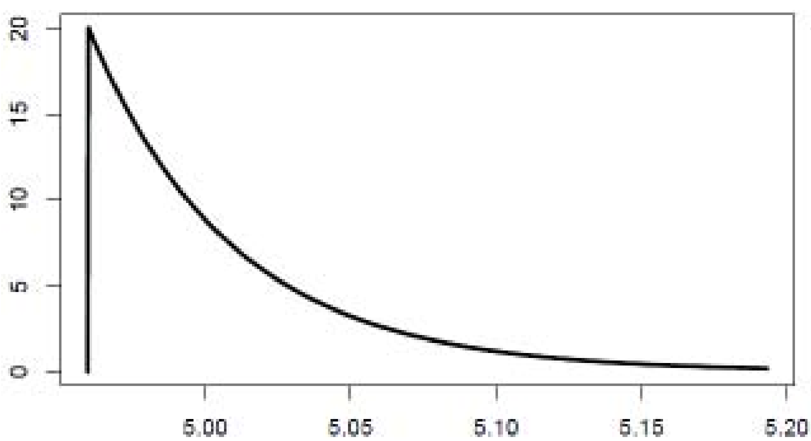

Figura 1.6: Distribución final no informativa

Ejemplo: Tenemos una m.a $X_1, ..., X_n \sim$ Uniforme$(a,\, b)$, $a < b$, con $a,\, b \in \mathbb{R}$ con una función de densidad $f(x \,|\, a,\, b) = \dfrac{1}{b-a}\, I_{\{a < x < b\}}$. Entonces,

$$L(a, b \,|\, \mathbf{x}) = \prod_{i=1}^{n} f(x_i \,|\, a,\, b) = \frac{1}{(b-a)^n}\, I_{\{a < x_{(1)}\}}\, I_{\{b > x_{(n)}\}}.$$

Supongamos que contamos con la siguiente información inicial:

$$\text{Media}(b) = \beta_1 + k_0 \tag{1.5}$$

$$P(a < \beta_0) = 1 = P(b > \beta_1). \tag{1.6}$$

Con k_0 β_0 y β_1 conocidos.

Vamos a construir una distribución inicial que sea **familia conjugada** del modelo Uniforme(a, b). Decimos que la distribución inicial es familia conjugada para el modelo $L(\theta \,|\, \mathbf{x})$, cuando sigue la misma ley de probabilidad que la distribución final, lo que hace los cálculos de probabilidades más sencillos, ya que es un modelo conocido, (veáse [4]):

$$\wp(a, b \,|\, \mathbf{x}) \propto L(a, b \,|\, \mathbf{x}) \wp(a, b).$$

Entonces,

$$\int\int_{\mathbb{R}^2} \frac{1}{(b-a)^n}\, I_{\{a < x_{(1)}\}}\, I_{\{b > x_{(n)}\}}\, db\, da = \int_{-\infty}^{x_{(1)}} \left(\int_{x_{(n)}}^{\infty} (b-a)^{-n}\, db \right) da.$$

Con el cambio de variable $u = b - a, du = db$ tenemos

$$\int_{-\infty}^{x_{(1)}} \int_{x_{(n)}-a}^{\infty} u^{-n}\, du = \int_{-\infty}^{x_{(1)}} \left(\frac{u^{-n+1}}{-n+1} \Big|_{x_{(n)}-a}^{\infty} \right) da = \int_{-\infty}^{x_{(1)}} \frac{(x_{(n)}-a)^{-n+1}}{n-1}\, da.$$

Con el cambio de variable $v = x_{(n)} - a, dv = -da$ tenemos que

$$\begin{aligned} -\int_{\infty}^{x_{(n)}-x_{(1)}} \frac{(x_{(n)}-a)^{-n+1}}{n-1}\, da &= \frac{(x_{(n)}-a)^{-n+2}}{(n-1)(-n+2)} \Big|_{x_{(n)}-x_{(1)}}^{\infty} \\ &= \frac{1}{(n-1)(n-2)(x_{(n)}-x_{(1)})^{n-2}} \end{aligned}$$

Entonces,

$$\int\int_{\mathbb{R}^2} \frac{1}{(b-a)^n}\, I_{\{a<x_{(1)}\}}\, I_{\{b>x_{(n)}\}}\, db\, da = \frac{1}{(n-1)(n-2)(x_{(n)}-x_{(1)})^{n-2}}.$$

Al despejar,

$$\int\int_{\mathbb{R}^2} \frac{(n-1)(n-2)(x_{(n)}-x_{(1)})^{n-2}}{(b-a)^n} I_{\{a<x_{(1)}\}}\, I_{\{b>x_{(n)}\}}\, db\, da = 1 \text{ con } n \geq 3.$$

Sean $\alpha = n-2,\ \beta_1 = x_{(n)},$ y $\beta_0 = x_{(1)}$. Entonces,

$$\wp(a, b \,|\, \alpha, \beta_0, \beta_1) = \frac{\alpha(\alpha+1)(\beta_1-\beta_0)^{\alpha}}{(b-a)^{\alpha+2}} I_{\{a<\beta_0\}}\, I_{\{b>\beta_1\}} \quad \alpha > 0,\ \beta_0 < \beta_1.$$

Hemos construido una distribución inicial que además es familia conjugada para el modelo.

Ahora, al obtener la distribución final (1.2) se tiene,

$$\wp(a, b \,|\, \mathbf{x}) \propto L(a, b \,|\, \mathbf{x})\wp(a, b) = \frac{1}{(b-a)^{n+\alpha+2}}.$$

Al obtener la constante de integración

$$\int\int_{\mathbb{R}^2} \frac{1}{(b-a)^{n+\alpha+2}} I_{\{a < \beta_0 \wedge x_{(1)}\}}\, I_{\{b > \beta_1 \vee x_{(n)}\}} db\, da.$$

Con el cambio de variable $u = b - a, du = db$ tenemos

$$\int_{-\infty}^{\beta_0 \wedge x_{(1)}} \int_{\beta_1 \vee x_{(n)} - a}^{\infty} \frac{1}{u^{n+\alpha+2}} du\, da = \int_{-\infty}^{\beta_0 \wedge x_{(1)}} \left. \frac{u^{-n-\alpha-1}}{-n-\alpha-1} \right|_{\beta_1 \vee x_{(n)} - a}^{\infty} da$$

$$= \int_{-\infty}^{\beta_0 \wedge x_{(1)}} \frac{(\beta_1 \vee x_{(n)} - a)^{-n-\alpha-1}}{-n-\alpha-1} da$$

Con el cambio de variable $v = \beta_1 \vee x_{(n)} - a, dv = -da$ tenemos

$$-\int_{-\infty}^{\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}} \frac{v^{-n-\alpha-1}}{-n-\alpha-1} dv = \left. \frac{v^{-n-\alpha}}{(n+\alpha+1)(n+\alpha)} \right|_{-\infty}^{\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}}$$

$$= \frac{\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}}{(n+\alpha+1)(n+\alpha)}$$

Por lo tanto $\wp(a, b \,|\, \mathbf{x})$ es igual a

$$\frac{(\alpha+n)(\alpha+n+1)(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)})^{\alpha+n}}{(b-a)^{\alpha+n+2}} I_{\{a < \beta_0 \wedge x_{(1)}\}} \, I_{\{b > \beta_1 \vee x_{(n)}\}}$$

$$= \wp(a, b \,|\, \alpha+n,\, \beta_0 \wedge x_{(1)},\, \beta_1 \vee x_{(n)}).$$

Al obtener la predictiva final (1.1)

$$\wp(x \,|\, \mathbf{x}) = \int\int_{\mathbb{R}^2} f(x \,|\, a, b) \wp(a, b \,|\, \mathbf{x}) \, db \, da$$

Para facilitar los cálculos, haremos

$N = (\alpha+n)(\alpha+n+1)(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)})^{\alpha+n}$

$$= \int_{-\infty}^{\text{mín}\{x,\, \beta_0 \wedge x_{(1)}\}} \int_{\text{máx}\{x,\, \beta_1 \vee x_{(n)}\}}^{\infty} \frac{N}{(b-a)^{\alpha+n+3}} db \, da,$$

$$= N \int_{-\infty}^{\text{mín}\{x,\, \beta_0 \wedge x_{(1)}\}} \int_{\text{máx}\{x,\, \beta_1 \vee x_{(n)}\}}^{\infty} (b-a)^{-\alpha-n-3} db \, da.$$

Con el cambio de variable $u = b - a,\ du = db$ tenemos

$$N\int_{-\infty}^{\text{mín}\{x,\beta_0\wedge x_{(1)}\}} \left.\frac{u^{-\alpha-n-2}}{-\alpha-n-2}\right|_{\text{máx}\{x,\beta_1\vee x_{(n)}\}-a}^{\infty} da.$$

$$= N\int_{-\infty}^{\text{mín}\{x,\beta_0\wedge x_{(1)}\}} \frac{(\text{máx}\{x,\ \beta_1\ \vee\ x_{(n)}\}-a)^{-\alpha-n-2}}{\alpha+n+2} da.$$

Con el cambio de variable $t=\text{máx}\{x,\ \beta_1\ \vee\ x_{(n)}\}-a,\ dt=-da$ tenemos

$$\left(\frac{N}{\alpha+n+2}\right)\left.\frac{t^{-\alpha-n-1}}{-\alpha-n-1}\right|_{\text{máx}\{x,\beta_1\vee x_{(n)}\}-\text{mín}\{x,\beta_0\wedge x_{(1)}\}}^{\infty},$$

$$= \frac{N(\text{máx}\{x,\ \beta_1\ \vee\ x_{(n)}\}-\text{mín}\{x,\ \beta_0\ \wedge\ x_{(1)}\})^{-\alpha-n-1}}{\alpha+n+2}.$$

$$= \frac{(\alpha+n)(\beta_1\vee x_{(n)}-\beta_0\wedge x_{(1)})^{\alpha+n}(\text{máx}\{x,\ \beta_1\ \vee\ x_{(n)}\}-\text{mín}\{x,\ \beta_0\ \wedge\ x_{(1)}\})^{-\alpha-n-1}}{\alpha+n+2}.$$

Sean $B_1=\beta_1\vee x_{(n)}$ y $B_0=\beta_0\wedge x_{(1)}$. Entonces

$$\wp(x\,|\,\mathbf{x}) = \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(x\vee B_1-x\wedge B_0)^{\alpha+n+1}}.$$

Con la distribución inicial,

obtenemos la marginal $\wp_A(a)$:

$$\wp_A(a) = \int_{-\infty}^{\infty} \wp(a,b)db \, = \alpha(\alpha+1)(\beta_1-\beta_0)^{\alpha}\, I_{\{a<\beta_0\}} \int_{\beta_1}^{\infty} (b-a)^{-\alpha-2} db.$$

Con el cambio de variable $u = b - a$, $du = db$ tenemos

$$\wp_A(a) = \int_{-\infty}^{\infty} \wp(a,b)db \, = \alpha(\alpha+1)(\beta_1-\beta_0)^{\alpha}\, I_{\{a<\beta_0\}} \left.\frac{u^{-\alpha-1}}{-\alpha-1}\right|_{\beta_1-a}^{\infty}.$$

Entonces

$$\wp_A(a \,|\, \alpha, \beta_0, \beta_1) = \frac{\alpha(\beta_1-\beta_0)^{\alpha}}{(\beta_1-a)^{\alpha+1}}\, I_{\{a\,<\,\beta_0\}} \, \text{con } \alpha > 0, \; \beta_0 < \beta_1.$$

La marginal $\wp_B(b)$ es

$$\wp_B(b) = \int_{-\infty}^{\infty} \wp(a,b)da \, = \alpha(\alpha+1)(\beta_1-\beta_0)^{\alpha}\, I_{\{b>\beta_1\}} \int_{\beta_1}^{\infty} (b-a)^{-\alpha-2} da.$$

Con el cambio de variable $u = b - a$, $du = -da$ tenemos

$$\wp_B(b) = \int_{-\infty}^{\infty} \wp(a,b)da \, = \alpha(\alpha+1)(\beta_1-\beta_0)^{\alpha}\, I_{\{b>\beta_1\}} \left.\frac{u^{-\alpha-1}}{-\alpha-1}\right|_{b-\beta_0}^{\infty}.$$

Entonces

$$\wp_B(b\,|\,\alpha,\beta_0,\beta_1) = \frac{\alpha(\beta_1-\beta_0)^{\alpha}}{(b-\beta_0)^{\alpha+1}}\, I_{\{b>\beta_1\}} \text{ con } \alpha>0,\ \beta_0<\beta_1$$

y la condicional $\wp_{B\,|\,A}(b\,|\,a)$ está dada por

$$\wp_{B\,|\,A}(b\,|\,a) = \frac{\wp(a,b)}{\wp(a)}$$

$$= \left(\frac{\alpha(\alpha+1)(\beta_1-\beta_0)^{\alpha}\, I_{\{a<\beta_0\}}\, I_{\{b>\beta_1\}}}{(b-a)^{\alpha+2}}\right)\left(\frac{(\beta_1-a)^{\alpha+1}}{\alpha(\beta_1-\beta_0)^{\alpha}\, I_{\{a<\beta_0\}}}\right)$$

$$= \frac{(\alpha+1)(\beta_1-a)^{\alpha+1}}{(b-a)^{\alpha+2}}\, I_{\{b>\beta_1\}}.$$

De la distribución final y tomando en cuenta que la distribución inicial es familia conjugada, obtenemos las marginales

$$\wp_A(a\,|\,\mathbf{x}) = \frac{(\alpha+n)(\beta_1\vee x_{(n)}-\beta_0\wedge x_{(1)})^{\alpha+n}}{(\beta_1\vee x_{(n)}-a)^{\alpha+n+1}}\, I_{\{a<\beta_0\wedge x_{(1)}\}} \text{ con } \alpha>0,\ \beta_0<\beta_1$$

$$\wp_B(b\,|\,\mathbf{x}) = \frac{(\alpha+n)(\beta_1\vee x_{(n)}-\beta_0\wedge x_{(1)})^{\alpha+n}}{(b-\beta_0\wedge x_{(1)})^{\alpha+n+1}}\, I_{\{b>\beta_1\vee x_{(n)}\}} \text{ con } \alpha>0,\ \beta_0<\beta_1.$$

Y la condicional

$$\wp_{B\,|\,A}(b\,|\,a) = \frac{(\alpha+n+1)(\beta_1 \vee x_{(n)} - a)^{\alpha+n+1}}{(b-a)^{\alpha+n+2}}\, I_{\{b\,>\,\beta_1 \vee x_{(n)}\}}.$$

Las distribuciones así obtenidas serán de utilidad para hacer inferencia estadística tanto sobre los parámetros como sobre una observación futura (predicción), como se verá en los capítulos siguientes.

Capítulo 2

Estimación puntual y de conjunto

2.1. Estimación puntual

Estimar puntualmente un parámetro θ de un modelo de probabilidad $f(x \mid \theta)$ dada cierta información inicial y/o muestral es escoger un valor $\hat{\theta}$ que pertenezca al espacio paramétrico de θ, mismo que denotaremos con Θ, y que esté cercano al verdadero valor de θ que desconocemos. $\hat{\theta}$ se puede obtener mediante alguna medida de tendencia central de la distribución final $\wp(\theta \mid \mathbf{x})$, por ejemplo la media:

$$\hat{\theta} = E(\theta) = \int_{\Theta} \theta \, \wp(\theta \mid \mathbf{x}) d\theta. \tag{2.1}$$

En el caso de que no se tenga una muestra observada, puede usarse $\wp(\theta)$ en vez de $\wp(\theta \mid \mathbf{x})$, veáse [3].

Ejemplo: Consideremos una m.a $X_1, ..., X_n \sim$ Uniforme$(0, \theta)$, $\theta > 0$, donde la información inicial se representa mediante una distribución de probabilidad Pareto(α, β).

Utilizando la media como medida de tendencia central una estimación puntual de θ inicial es:

$$\hat{\theta} = E(\theta) = \int_{\Theta} \theta \, \wp(\theta \,|\, \alpha, \beta) \, d\theta = \frac{\beta\alpha}{\beta - 1} \quad \text{con } \beta > 1$$

Del capítulo anterior tenemos que $\wp(\theta \,|\, \mathbf{x}) \sim \text{Pareto}(\alpha \vee x_{(n)},\, \beta + n)$, por lo que la estimación puntual final de θ es:

$$\hat{\theta} = E(\theta) = \int_{\Theta} \theta \, \wp(\theta \,|\, \mathbf{x}) \, d\theta = \frac{(\beta + n)(\alpha \vee x_{(n)})}{\beta + n - 1}$$

Si se deseara la influencia de la información inicial, debe hacerse $\alpha \to 0$, $\beta \to 0$. Entonces:

$$\hat{\theta} = \frac{n \, x_{(n)}}{n - 1}.$$

Ahora, utilizando la mediana, tenemos:

$$\int_{-\infty}^{t} \frac{\beta \, \alpha^{\beta}}{\theta^{\beta+1}} \, I_{\{\alpha < \theta\}} \, I_{\{t < \beta\}} \, d\theta + I_{\{t \geq \beta\}} = \beta \, \alpha^{\beta} \, I_{\{t < \beta\}} \int_{\alpha}^{t} \theta^{-\beta-1} \, d\theta + \, I_{\{t \geq \beta\}}$$

$$= -\alpha^{\beta} \, (t^{-\beta} - \alpha^{-\beta}) \, I_{\{t < \beta\}} + \, I_{\{t >= \beta\}} = 1 - \left(\frac{\alpha}{t}\right)^{\beta} I_{\{t < \beta\}} + \, I_{\{t \geq \beta\}}.$$

Entonces

$$1 - \left(\frac{\alpha}{t}\right)^{\beta} = \frac{1}{2}.$$

Al despejar a t, tenemos

$$t = \frac{\alpha}{(1/2)^{1/\beta}}.$$

Ya que la distribución Pareto es familia conjugada para la distribución Uniforme$(0, \theta)$, el estimador puntual buscado es

$$t = 2^{1/(\beta+n)}(\alpha \vee x_{(n)}).$$

Si $\alpha \to 0$, $\beta \to 0$,

$$t = 2^{1/n}\, x_{(n)}.$$

Sabemos que la distribución Pareto es continua, al utilizar la moda, que es donde esta distribución alcanza el supremo, el estimador buscado es $(\alpha \vee x_{(n)})$.

Si $\alpha \to 0$, el estimador que buscamos es $x_{(n)}$.

Nota: Podemos ver que al usar la moda y desvanecer la información inicial, coincide con lo que algunos conocen como *estimador de máxima verosimilitud*, veáse [1]. Decidir qué medida de tendencia central conviene utilizar será tema del capítulo 4.

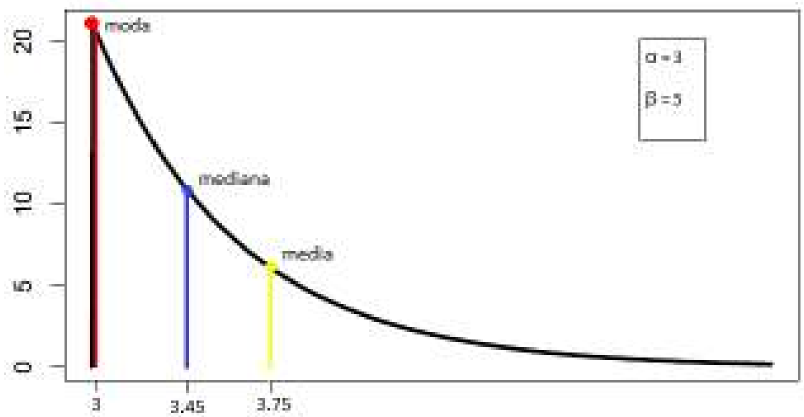

Figura 2.1: Estimadores puntuales Distribución Pareto

Ejemplo: Consideremos una m.a $X_1, ..., X_n \sim$ Bernoulli(θ), $\theta > 0$, donde la información inicial se representa mediante una distribución de probabilidad Beta(α, β).

La estimación puntual de θ inicial es

$$\hat{\theta} = E(\theta) = \int_{\Theta} \theta \, \wp(\theta \,|\, \alpha, \beta) \, d\theta = \frac{\alpha}{\alpha + \beta}.$$

Del capítulo anterior tenemos que $\wp(\theta \,|\, \mathbf{x}) \sim \text{Beta}(\alpha + \sum x_i,\ \beta + n - \sum x_i)$, por lo que la estimación puntual final de θ es

$$\hat{\theta} = E(\theta) = \int_{\Theta} \theta \, \wp(\theta \,|\, \mathbf{x}) \, d\theta = \frac{\alpha + \sum x_i}{\alpha + \beta + n}.$$

Si $\alpha \to 0$, $\beta \to 0$,

$$\hat{\theta} = \frac{\sum x_i}{n}.$$

El estimador puntual es, vía la mediana,

$$\hat{\theta} = F^{-1}(1/2 \,|\, \mathbf{x}).$$

Es importante observar que no es posible obtener una función de distribución explícita para el modelo Beta, por lo cual, se debe aproximar numéricamente.

Sabemos que la distribución Beta es continua. Al emplear la moda, que es donde alcanza el supremo, el estimador buscado es

$$\frac{\alpha + \sum x_i - 1}{\alpha + \beta + n - 2} \text{ con } \alpha,\, \beta > 1 \text{ cuando existe.}$$

Si $\alpha \to 0$, $\beta \to 0$, el estimador que buscamos es

$$\frac{\sum x_i - 1}{n - 2}.$$

Observemos que para valores muy grandes de n el valor del estimador puntual vía la mediana, vía la media o vía la moda no difieren mucho.

Ejemplo: Consideremos una m.a $X_1, ..., X_n \sim$ Uniforme$(a,\, b)$, $a < b$, donde la información inicial se representa mediante una distribución de probabilidad

$$\wp(a,\, b \,|\, \alpha,\, \beta_0,\, \beta_1) = \frac{\alpha(\alpha+1)(\beta_1 - \beta_0)^{\alpha}}{(b-a)^{\alpha+2}}\, I_{\{a < \beta_0\}}\, I_{\{b > \beta_1\}}\ \alpha > 0,\ \beta_0 < \beta_1.$$

Del capítulo anterior tenemos que

$$\wp_A(a \,|\, \alpha, \beta_0, \beta_1) = \frac{\alpha(\beta_1 - \beta_0)^\alpha}{(\beta_1 - a)^{\alpha+1}}\, I_{\{a < \beta_0\}} \text{ con } \alpha > 0,\ \beta_0 < \beta_1$$

$$\wp_B(b \,|\, \alpha, \beta_0, \beta_1) = \frac{\alpha(\beta_1 - \beta_0)^\alpha}{(b - \beta_0)^{\alpha+1}}\, I_{\{b > \beta_1\}} \text{ con } \alpha > 0,\ \beta_0 < \beta_1$$

El estimador puntual inicial de a es

$$\hat{a} = E(A) = \int_A a\, \wp(a \,|\, \alpha, \beta_0, \beta_1)\, da = \alpha(\beta_1 - \beta_0)^\alpha \int_{-\infty}^{\beta_0} a(\beta_1 - a)^{-\alpha-1}\, da.$$

Con el cambio de variable $u = \beta_1 - a,\ a = \beta_1 - u$ y $du = -da$ por lo que

$$\alpha(\beta_1 - \beta_0)^\alpha \int_{\beta_1-\beta_0}^{\infty} (\beta_1 - u)u^{-\alpha-1}\, du = \beta_1 - \frac{\alpha(\beta_1 - \beta_0)}{\alpha - 1}.$$

Por lo tanto la estimación puntual inicial de a es

$$\hat{a} = \beta_1 - \frac{\alpha(\beta_1 - \beta_0)}{\alpha - 1}.$$

El estimador puntual inicial de b es

$$\hat{b} = E(B) = \int_B b\, \wp(b \,|\, \alpha, \beta_0, \beta_1)\, db = \alpha(\beta_1 - \beta_0)^\alpha \int_{\beta_1}^{\infty} b(b - \beta_0)^{-\alpha-1}\, db.$$

Con el cambio de variable $u = b - \beta_0,\ b = u + \beta_0$ y $du = db$ y entonces

$$\alpha(\beta_1 - \beta_0)^{\alpha} \int_{\beta_1-\beta_0}^{\infty} (u + \beta_0) u^{-\alpha-1}\, du = \beta_0 + \frac{\alpha(\beta_1 - \beta_0)}{\alpha - 1}.$$

Por lo tanto, la estimación puntual inicial de b es

$$\hat{b} = \beta_0 + \frac{\alpha(\beta_1 - \beta_0)}{\alpha - 1}.$$

La estimación puntual final de A y B, tomando en cuenta que son familias conjugadas, es pues

$$\hat{a} = (\beta_1 \vee x_{(n)}) - \frac{(\alpha + n)(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)})}{\alpha + n - 1}$$

$$\hat{b} = (\beta_0 \wedge x_{(1)}) + \frac{(\alpha + n)(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)})}{\alpha + n - 1}.$$

Si $\alpha \to 0$, $\beta_0 \to \infty$, $\beta_1 \to 0$, los estimadores que buscamos son

$$\hat{a} = x_{(n)} - \frac{n(x_{(n)} - x_{(1)})}{n - 1}$$

$$\hat{b} = x_{(1)} + \frac{n(x_{(n)} - x_{(1)})}{n - 1}.$$

Con la mediana, la función de distribución para la marginal A es

$$\int_{-\infty}^{t} \wp_A(a \,|\, \alpha, \beta_0, \beta_1)\, da = \alpha(\beta_1 - \beta_0)^{\alpha}\, I_{\{t<\beta_0\}} \int_{-\infty}^{t} (\beta_1 - a)^{-\alpha-1}\, da + \, I_{\{t\geq\beta_0\}}$$

$$= \left(\frac{\beta_1 - \beta_0}{\beta_1 - t}\right)^{\alpha} I_{\{t<\beta_0\}} + \, I_{\{t\geq\beta_0\}}.$$

Entonces

$$\left(\frac{\beta_1 - \beta_0}{\beta_1 - t}\right)^{\alpha} = 1/2.$$

Al despejar a t se sigue que

$$t = \beta_1 - 2^{\,1/\alpha}(\beta_1 - \beta_0).$$

Ya que es familia conjugada, el estimador buscado es

$$t = (\beta_1 \vee x_{(n)}) - 2^{\,1/(\alpha+n)}(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}).$$

Si $\alpha \to 0,\ \beta_0 \to \infty$ y $\beta_1 \to 0$,

$$t = x_{(n)} - 2^{\,1/n}(x_{(n)} - x_{(1)}).$$

La función de distribución para la marginal $B_{(x)}$ es en este caso la integral

$$\int_{-\infty}^{t} \wp_B(b \,|\, \alpha, \beta_0, \beta_1)\, db = \alpha(\beta_1 - \beta_0)^{\alpha}\, I_{\{t>\beta_1\}} \int_{\beta_1}^{t} (b-\beta_0)^{-\alpha-1}\, db + I_{\{t\leq\beta_1\}}$$

$$= 1 - \left(\frac{\beta_1 - \beta_0}{t - \beta_0}\right)^{\alpha} I_{\{t>\beta_1\}} + I_{\{t\leq\beta_1\}}.$$

Entonces

$$= 1 - \left(\frac{\beta_1 - \beta_0}{t - \beta_0}\right)^{\alpha} = 1/2.$$

De donde se sigue que

$$t = \beta_0 + 2^{\,1/\alpha}(\beta_1 - \beta_0).$$

Ya que es familia conjugada, el estimador buscado es:

$$t = (\beta_0 \wedge x_{(1)}) - 2^{\,1/(\alpha+n)}(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}).$$

Si $\alpha \to 0,\ \beta_0 \to \infty$ y $\beta_1 \to 0$

$$t = x_{(1)} - 2^{\,1/n}(x_{(n)} - x_{(1)}).$$

Sabemos que las marginales de a y b son continuas, la de a creciente y la de b decreciente. Al utilizar la moda, que es donde alcanza el supremo, el estimador buscado en cada caso es:

$$\hat{a} = \beta_0 \wedge x_{(1)}$$

y

$$\hat{b} = \beta_1 \vee x_{(n)}.$$

Si $\beta_0 \to \infty$, $\beta_1 \to 0$, los estimadores que buscamos son

$$\hat{a} = x_{(1)} \text{ y } \hat{b} = x_{(n)}.$$

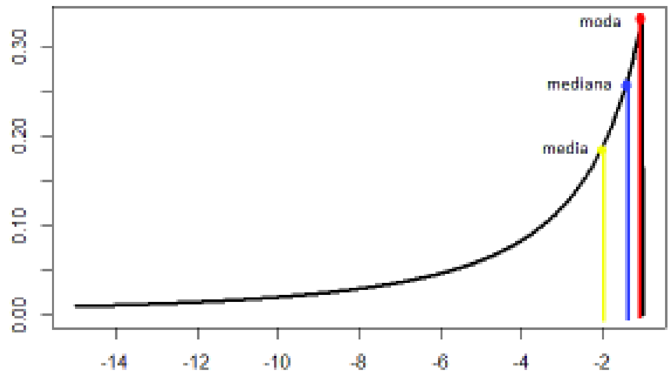

Figura 2.2: Estimadores puntuales de la marginal A

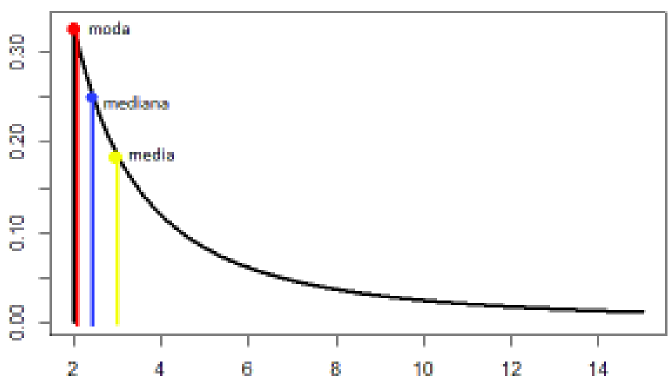

Figura 2.3: Estimadores puntuales de la marginal B

2.2. Estimación de conjunto

Hacer estimación de conjunto es encontrar un subconjunto $\Theta_\gamma \subset \Theta$ tal que $P(\theta \in \Theta_\gamma) = \gamma$, donde la probabilidad γ es típicamente muy cercana a 1.

En el caso particular de un espacio paramétrico de dimensión 1 la estimación de conjunto consiste en encontrar valores $\hat{\theta}_i$, $\hat{\theta}_s$ elementos del espacio paramétrico, $\hat{\theta}_i < \hat{\theta}_s$ dada cierta información inicial y/o muestral tal que $P(\hat{\theta}_i \leq \theta \leq \hat{\theta}_s) = \gamma$, típicamente muy cercana a 1, donde $\hat{\theta}_i$ y $\hat{\theta}_s$ sean solución de la siguiente integral (o suma en el caso discreto):

$$\int_{\hat{\theta}_i}^{\hat{\theta}_s} \wp(\theta \,|\, \mathbf{x}) d\theta = \gamma \tag{2.2}$$

Normalmente la solución para $\hat{\theta}_i, \hat{\theta}_s$ no es única. Por ello se escoge aquella que minimice la longitud del conjunto $[\hat{\theta}_i, \hat{\theta}_s]$, (veáse [2]).

Ejemplo: Consideremos una m.a $X_1, ..., X_n \sim$ Bernoulli(θ), $0 < \theta < 1$, donde la información inicial se representa mediante una distribución de probabilidad Uniforme(0, 1).

La ecuación (1.2) es en este caso

$$\wp(\theta \,|\, \mathbf{x}) = \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i} I_{(0,1)}\theta}{\int_0^1 \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} I_{(0,1)}\theta \, d\theta}.$$

Podemos ver que la constante de integración $K = \int_0^1 \theta^{\sum x_i+1-1}(1-\theta)^{n-\sum x_i+1-1} \, d\theta$ es el kernel de una Beta($1+\sum x_i$, $n - \sum x_i + 1$). Por lo tanto:

$$\wp(\theta \,|\, \mathbf{x}) \sim \text{Beta}\Big(1 + \sum x_i, \, n - \sum x_i + 1\Big).$$

Ahora si observamos la muestra $(0, 0, 0)$, $n = 3$. Entonces $\wp(\theta \,|\, 0, 0, 0) \sim$ Beta$(1, 4)$ y

$$\wp(\theta \,|\, 0, 0, 0) = \frac{\Gamma(5)}{\Gamma(1)\,\Gamma(4)} \theta^{1-1}(1-\theta)^{4-1} I_{(0,1)}\theta = 4(1-\theta)^3 I_{(0,1)}\theta.$$

Dado un valor $\gamma \in (0,1)$ queremos conocer un conjunto de probabilidad γ para θ, es decir, encontrar $0 < \hat{\theta}_i < \hat{\theta}_s < 1$ tal que $P(\hat{\theta}_i \leq \theta \leq \hat{\theta}_s) = \gamma$. Para ello hacemos

$$\int_{\hat{\theta}_i}^{\hat{\theta}_s} \wp(\theta \,|\, 0,0,0)\, d\theta = 4\int_{\hat{\theta}_i}^{\hat{\theta}_s} (1-\theta)^3\, d\theta = (1-\hat{\theta}_i)^4 - (1-\hat{\theta}_s)^4 = \gamma.$$

De donde se sigue que

$$\hat{\theta}_s = -[(1-\hat{\theta}_i)^4 - \gamma]^{1/4} + 1.$$

Hay múltiples soluciones para la expresión anterior por lo que escogemos la de menor longitud. En este caso haremos $\hat{\theta}_i = 0$, por lo cual $\hat{\theta}_s = 1-(1-\gamma)^{1/4}$.

Por lo tanto $\left[0, 1-(1-\gamma)^{1/4}\right]$ es de probabilidad γ para θ.

Si $\gamma = 0.95$, entonces $\hat{\theta}_s = 0.5271$, y por lo tanto $(0, 0.5271)$ es una estimación de conjunto de probabilidad 95 % para θ.

Ejemplo: Consideremos una m.a $X_1, ..., X_n \sim$ Exponencial(θ), $\theta > 0$, donde la información inicial se representa mediante una distribución de probabilidad Gamma$(\alpha,\, \beta)$.

La ecuación (1.2) es

$$\wp(\theta \,|\, \mathbf{x}) = \frac{\beta^{\,\alpha}}{\Gamma(\alpha)}\, \theta^{\,n+\alpha-1}\, e^{-\theta\,(\beta+\sum x_i)} I_{\{\theta>0\}}.$$

Podemos ver que la expresión anterior forma parte de una Gamma$(n+\alpha, \beta+\sum x_i)$. Hagamos $\alpha \to 0$, $\beta \to 0$ para desvanecer la información inicial. Tomando en cuenta una muestra de tamaño $n = 1$,

$$\wp(\theta \,|\, x_1) \sim \text{Gamma}(1,\, x_1) \sim \text{Exponencial}(x_1).$$

Entonces $F(\theta \,|\, x_1) = 1 - e^{-x_1 \theta},\ \theta > 0$.

Buscamos $0 < \hat{\theta}_i < \hat{\theta}_s$ tal que

$$P(\hat{\theta}_i \leq \theta \leq \hat{\theta}_s) = \gamma = F(\hat{\theta}_s \,|\, x_1) - F(\hat{\theta}_i \,|\, x_1)$$

$$\Leftrightarrow \hat{\theta}_s = F^{-1}\Bigg(\gamma + F(\hat{\theta}_i \,|\, x_1)\Bigg) = -\,\frac{1}{x_1}\,log\,(e^{\,-x_1\,\hat{\theta}_i} - \gamma).$$

Entonces $h(\hat{\theta}_i) = \hat{\theta}_s - \hat{\theta}_i = -\,\frac{1}{x_1}\,log\,(e^{\,-x_1\,\hat{\theta}_i} - \gamma) - a$.

Deseamos que $h(\hat{\theta}_i)$ se minimice. Esto se logra haciendo que $\hat{\theta}_i \longrightarrow 0$, por lo que

$$\hat{\theta}_s = -\tfrac{1}{x_1} log\,(1-\gamma).$$

Por lo tanto, $\left[0,\, \dfrac{-log\,(1-\gamma)}{x_1}\right]$ es un conjunto de probabilidad γ para θ.

Ejemplo: Consideremos una m.a $X_1, ..., X_n \sim$ Uniforme$(a,\, b)$, $a < b$, donde la información inicial se representa mediante una distribución de probabilidad

$$\wp(a,\, b \,|\, \alpha,\, \beta_0,\, \beta_1) = \frac{\alpha(\alpha+1)(\beta_1 - \beta_0)^{\,\alpha}}{(b-a)^{\alpha+2}}\, I_{\{a\,<\,\beta_0\}}\, I_{\{b\,>\,\beta_1\}}\ \alpha > 0,\ \beta_0 < \beta_1.$$

Deseamos encontrar un conjunto de longitud mínima con un cierto valor c con probabilidad γ para la marginal A, entonces tenemos

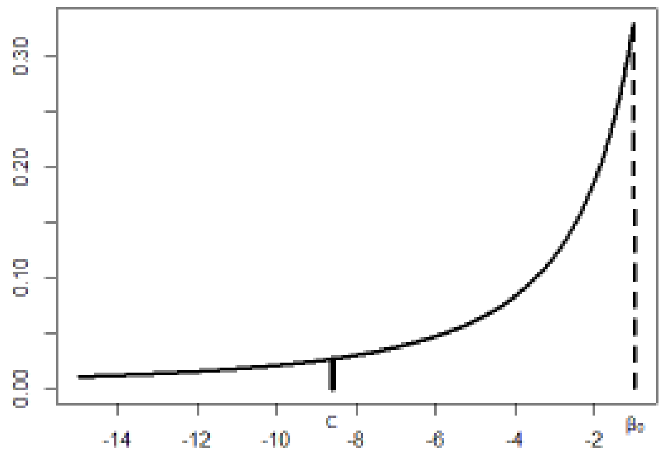

Figura 2.4: Marginal A

$$\int_c^{\beta_0} \wp_A(a \,|\, \alpha, \beta_0, \beta_1)\, da = \int_c^{\beta_0} \frac{\alpha(\beta_1 - \beta_0)^{\alpha}}{(\beta_1 - a)^{\alpha+1}}\, I_{\{a < \beta_0\}}\, da = 1 - \left(\frac{\beta_1 - \beta_0}{\beta_1 - c}\right)^{\alpha}$$

Igualando a la probabilidad Gamma y despejando c

$$1 - \left(\frac{\beta_1 - \beta_0}{\beta_1 - c}\right)^{\alpha} = \gamma$$

$$c = \beta_1 - \frac{\beta_1 - \beta_0}{(1-\gamma)^{1/\alpha}}.$$

Ya que es familia conjugada, el conjunto buscado es

$$\left(\beta_1 \vee x_{(n)} - \frac{\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}}{(1-\gamma)^{1/(\alpha+n)}}\,,\ \ \beta_0 \wedge x_{(1)}\right).$$

Deseamos encontrar un conjunto de área mínima con un cierto valor c con probabilidad γ para la marginal B, entonces tenemos

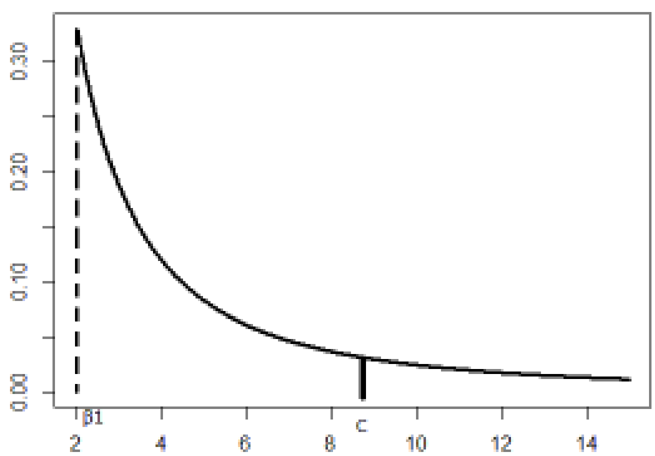

Figura 2.5: Marginal B

$$\int_{\beta_1}^{c} \wp_B(b\,|\,\alpha,\beta_0,\beta_1)\,db = \int_{\beta_1}^{c} \frac{\alpha(\beta_1-\beta_0)^{\alpha}}{(b-\beta_0)^{\alpha+1}}\, I_{\{b>\beta_1\}}\,db = 1 - \left(\frac{\beta_1-\beta_0}{c-\beta_0}\right)^{\alpha}.$$

Igualando a la probabilidad Gamma y despejando c

$$1 - \left(\frac{\beta_1-\beta_0}{c-\beta_0}\right)^{\alpha} = \gamma$$

$$c = \beta_0 + \frac{\beta_1 - \beta_0}{(1-\gamma)^{1/\alpha}}.$$

Ya que es familia conjugada, el conjunto buscado es

$$\left(\beta_1 \vee x_{(n)}\,,\;\; \beta_0 \wedge x_{(1)} + \frac{\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}}{(1-\gamma)^{1/(\alpha+n)}}\right).$$

Deseamos encontrar un volumen de longitud mínima con un cierto radio r con probabilidad γ para la densidad final Uniforme(a, b). Entonces tenemos la integral doble

$$\int_{\beta_1}^{\beta_1+r}\int_{\beta_0-\sqrt{r^2-(b-\beta_1)^2}}^{\beta_0} \frac{\alpha(\alpha+1)(\beta_1-\beta_0)^{\,\alpha}}{(b-a)^{\,\alpha+2}}\, da\, db$$

$$= \alpha(\beta_1-\beta_0)^{\alpha}\left(\int_{\beta_1}^{\beta_1+r}(b-\beta_0)^{-\alpha-1}\, db - \int_{\beta_1}^{\beta_1+r}(b+\sqrt{r^2-(b-\beta_1)^2}-\beta_0)^{-\alpha-1}\, db\right).$$

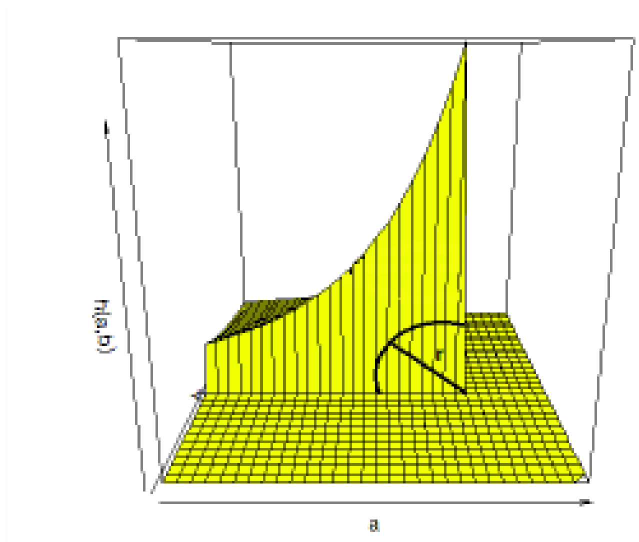

Figura 2.6: Densidad final

Es importante observar que no es posible resolver la expresión anterior de forna explícita y por tanto en necesario recurrir a la aproximación numérica.

Capítulo 3

Contraste de hipótesis

Sea $\{\Theta_i : i = 1, \ldots, m\}$ una partición del espacio paramétrico Θ. Una hipótesis sobre el parámetro θ es de la forma:

$$\mathcal{H}_i : \theta \in \Theta_i, \;\; \Theta_i \subseteq \Theta.$$

Es posible contrastar dos o más hipótesis, ya que a cada una de ellas se le calcula su probabilidad de ocurrencia. Dicha probabilidad se calcula de la siguiente forma

$$P(\mathcal{H}_i) = P(\theta \in \Theta_i) = \int_{\Theta_i} \wp(\theta \,|\, \mathbf{x}) d\theta. \tag{3.1}$$

Una forma de escoger una de las hipótesis es elegir aquella $\mathcal{H}_*$ tal que $P(\mathcal{H}_*) \geq P(\mathcal{H}_i)$ para todo i, (veáse [6]).

En el caso de que no se tenga una muestra observada, puede usarse $\wp(\theta)$ en vez de $\wp(\theta \,|\, \mathbf{x})$.

Ejemplo: Consideremos una m.a $X_1, ..., X_n \sim$ Bernoulli(θ), $0 < \theta < 1$. Supongamos que la información inicial se representa mediante una distribución de probabilidad Uniforme(0, 1). Se observa la muestra $(0, 0, 0)$. Como hipótesis tenemos

$$\mathcal{H}_1 : \theta < 1/3 \qquad \mathcal{H}_2 : 1/3 \leq \theta \leq 2/3 \qquad \mathcal{H}_3 : \theta > 2/3$$

Del capítulo anterior tenemos que

$$\wp(\theta \,|\, 0,0,0) = 4(1-\theta)^3 I_{(0,1)}\theta.$$

Entonces

$$P(\mathcal{H}_1) = P(\theta < 1/3) = \int_0^{1/3} \wp(\theta \,|\, 0,0,0)\, d\theta = \int_0^{1/3} 4(1-\theta)^3 I_{(0,1)}\theta \, d\theta = 0.8024,$$

$$P(\mathcal{H}_2) = P(1/3 \leq \theta \leq 2/3) = \int_{1/3}^{2/3} \wp(\theta \,|\, 0,0,0)\, d\theta = \int_{1/3}^{2/3} 4(1-\theta)^3 I_{(0,1)}\theta \, d\theta$$
$$= 0.1851,$$

$$P(\mathcal{H}_3) = P(\theta > 2/3) = \int_{2/3}^{1} \wp(\theta \,|\, 0,0,0)\, d\theta = \int_{2/3}^{1} 4(1-\theta)^3 I_{(0,1)}\theta \, d\theta = 0.0123.$$

Por lo tanto escogemos la hipótesis $\mathcal{H}_1$, ya que es la de mayor probabilidad.

Ejemplo: Consideremos una m.a $X_1, ..., X_n \sim$ Uniforme(0, θ), $\theta > 0$, supongamos que la información inicial se representa mediante una distribución de probabilidad Pareto(0, 1). Se observa la muestra $(1, 1/2, 1/3)$. Como hipótesis tenemos:

$$\mathcal{H}_1 : \theta < 2 \qquad \mathcal{H}_2 : 2 \leq \theta < 3 \qquad \mathcal{H}_3 : 3 \leq \theta$$

Tenemos que

$$\wp(\theta\,|\,1,1/2,1/3) = \frac{4}{\theta^{\,5}}\,I_{\{\theta>1\}}$$

.

Entonces

$$P(\mathcal{H}_1) = P(\theta < 2) = \int_0^2 \wp(\theta\,|\,1,1/2,1/3)\,d\theta = \int_0^2 \frac{4}{\theta^{\,5}}\,I_{\{\theta>1\}}\,d\theta = 0.9375,$$

$$P(\mathcal{H}_2) = P(2 \leq \theta < 3) = \int_2^3 \wp(\theta\,|\,1,1/2,1/3)\,d\theta = \int_2^3 \frac{4}{\theta^{\,5}}\,I_{\{\theta>1\}}\,d\theta = 0.05015,$$

$$P(\mathcal{H}_3) = P(\theta \geq 3) = \int_3^\infty \wp(\theta\,|\,1,1/2,1/3)\,d\theta = \int_3^\infty \frac{4}{\theta^{\,5}}\,I_{\{\theta>1\}}\,d\theta = 0.01235.$$

Por lo tanto, escogemos la hipótesis $\mathcal{H}_1$ ya que es la de mayor probabilidad.

Ejemplo: Consideremos una m.a $X_1, ..., X_n \sim$ Uniforme$(a,\, b)$, $a < b$, donde la información inicial se representa mediante una distribución de probabilidad

$$\wp(a,\, b\,|\,\alpha,\, \beta_0,\, \beta_1) = \frac{\alpha(\alpha+1)(\beta_1-\beta_0)^{\,\alpha}}{(b-a)^{\alpha+2}}\,I_{\{a<\beta_0\}}\,I_{\{b>\beta_1\}}\;\;\alpha>0,\;\;\beta_0<\beta_1$$

con la marginal A y como hipótesis

$$\mathcal{H}_1 : a < \beta_0/3 \qquad \mathcal{H}_2 : \beta_0/3 \leq a < \beta_0/2 \qquad \mathcal{H}_3 : a \geq \beta_0/2$$

Tenemos que

$$\wp_A(a \,|\, \alpha, \beta_0, \beta_1) = \frac{\alpha(\beta_1 - \beta_0)^\alpha}{(\beta_1 - a)^{\alpha+1}} \, I_{\{a < \beta_0\}} \text{ con } \alpha > 0, \; \beta_0 < \beta_1$$

Entonces

$$P(\mathcal{H}_1) = P(a < \beta_0/3) = \int_{-\infty}^{\beta_0/3} \wp(a \,|\, \alpha, \beta_0, \beta_1) \, da$$
$$= \int_{-\infty}^{\beta_0/3} \frac{\alpha(\beta_1 - \beta_0)^\alpha}{(\beta_1 - a)^{\alpha+1}} \, I_{\{a < \beta_0\}} \, da = \left(\frac{\beta_1 - \beta_0}{\beta_1 - (\beta_0/3)} \right)^\alpha$$

$$P(\mathcal{H}_2) = P(\beta_0/3 \leq a < \beta_0/2) = \int_{\beta_0/3}^{\beta_0/2} \wp(a \,|\, \alpha, \beta_0, \beta_1) \, da$$
$$= \int_{\beta_0/3}^{\beta_0/2} \frac{\alpha(\beta_1 - \beta_0)^\alpha}{(\beta_1 - a)^{\alpha+1}} \, I_{\{a < \beta_0\}} \, da = \left(\frac{\beta_1 - \beta_0}{\beta_1 - (\beta_0/2)} \right)^\alpha - \left(\frac{\beta_1 - \beta_0}{\beta_1 - (\beta_0/3)} \right)^\alpha$$

$$P(\mathcal{H}_3) = P(a \geq \beta_0/2) = \int_{\beta_0/2}^{\beta_0} \wp(a \,|\, \alpha, \beta_0, \beta_1) \, da$$
$$= \int_{\beta_0/2}^{\beta_0} \frac{\alpha(\beta_1 - \beta_0)^\alpha}{(\beta_1 - a)^{\alpha+1}} \, I_{\{a < \beta_0\}} \, da = 1 - \left(\frac{\beta_1 - \beta_0}{\beta_1 - (\beta_0/2)} \right)^\alpha$$

Recordando que es familia conjugada tenemos que

$$P(\mathcal{H}_1) = \left(\frac{\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}}{\beta_1 \vee x_{(n)} - ((\beta_0 \wedge x_{(1)})/3)} \right)^{\alpha+n}$$

$$P(\mathcal{H}_2) = \left(\frac{\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}}{\beta_1 \vee x_{(n)} - ((\beta_0 \wedge x_{(1)})/2)} \right)^{\alpha+n} - \left(\frac{\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}}{\beta_1 \vee x_{(n)} - ((\beta_0 \wedge x_{(1)})/3)} \right)^{\alpha+n}$$

$$P(\mathcal{H}_3) = 1 - \left(\frac{\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}}{\beta_1 \vee x_{(n)} - ((\beta_0 \wedge x_{(1)})/2)} \right)^{\alpha+n}$$

Se escoge aquella hipótesis que tenga la mayor probabilidad.

Ahora con la distribución final y como hipótesis:

$$\mathcal{H}_1 : b < 2a \qquad \mathcal{H}_2 : 2a \leq b < 3a \qquad \mathcal{H}_3 : b \geq 3a$$

tenemos que

$$P(\mathcal{H}_1) = P(b < 2a) = \int_{\beta_1/2}^{\beta_0} \int_{\beta_1}^{2a} \wp(a,\, b \,|\, \alpha,\, \beta_0,\, \beta_1)\, db\, da$$

$$= \int_{\beta_1/2}^{\beta_0} \int_{\beta_1}^{2a} \frac{\alpha(\alpha+1)(\beta_1 - \beta_0)^{\alpha}}{(b-a)^{\alpha+2}}\, I_{\{a < \beta_0\}}\, I_{\{b > \beta_1\}}\ \alpha > 0,\ \beta_0 < \beta_1\, db\, da$$

$$= \left(\frac{\beta_1 - \beta_0}{\beta_0} \right)^{\alpha} - 2\left(\frac{2(\beta_1 - \beta_0)}{\beta_1} \right)^{\alpha} + 1$$

$$P(\mathcal{H}_2) = P(2a \leq b < 3a) = \int_{\beta_1/3}^{\beta_0} \int_{\beta_1}^{3a} \wp(a,\, b \,|\, \alpha,\, \beta_0,\, \beta_1)\, db\, da - P(\mathcal{H}_1)$$

$$= \int_{\beta_1/3}^{\beta_0} \int_{\beta_1}^{3a} \frac{\alpha(\alpha+1)(\beta_1-\beta_0)^{\alpha}}{(b-a)^{\alpha+2}} \, I_{\{a<\beta_0\}} \, I_{\{b>\beta_1\}} \; \alpha > 0, \;\beta_0 < \beta_1 \, db \, da - P(\mathcal{H}_1)$$

$$= \frac{(\beta_1-\beta_0)^{\alpha}}{2^{\alpha+1}(\beta_0)^{\alpha}} - \frac{(\beta_1-\beta_0)^{\alpha}}{2^{\alpha+1}(\beta_1/3)^{\alpha}} - \left(\frac{3(\beta_1-\beta_0)}{2\beta_1}\right)^{\alpha} - \left(\frac{\beta_1-\beta_0}{\beta_0}\right)^{\alpha} + 2\left(\frac{2(\beta_1-\beta_0)}{\beta_1}\right)^{\alpha}$$

$$P(\mathcal{H}_3) = P(b \geq 3a) = 1 - P(\mathcal{H}_1) - P(\mathcal{H}_2)$$

$$= \frac{(\beta_1-\beta_0)^{\alpha}}{2^{\alpha+1}(\beta_1/3)^{\alpha}} + \left(\frac{3(\beta_1-\beta_0)}{2\beta_1}\right)^{\alpha} - \frac{(\beta_1-\beta_0)^{\alpha}}{2^{\alpha+1}(\beta_0)^{\alpha}}$$

Recordando que es familia conjugada tenemos que:

$$P(\mathcal{H}_1) = \left(\frac{\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}}{\beta_0 \wedge x_{(1)}}\right)^{\alpha+n} - 2\left(\frac{2(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)})}{\beta_1 \vee x_{(n)}}\right)^{\alpha+n} + 1$$

$$P(\mathcal{H}_2) = \frac{(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)})^{\alpha+n}}{2^{\alpha+n+1}(\beta_0 \wedge x_{(1)})^{\alpha+n}} - \frac{(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)})^{\alpha+n}}{2^{\alpha+n+1}(\beta_1/3 \vee x_{(n)})^{\alpha+n}}$$

$$-\left(\frac{3(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)})}{2\beta_1 \vee x_{(n)}}\right)^{\alpha+n} - \left(\frac{\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)}}{\beta_0 \wedge x_{(1)}}\right)^{\alpha+n}$$

$$+2\left(\frac{2(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)})}{\beta_1 \vee x_{(n)}}\right)^{\alpha+n}$$

$$P(\mathcal{H}_3) = \frac{(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)})^{\alpha+n}}{2^{\alpha+n+1}(\beta_1/3 \vee x_{(n)})^{\alpha+n}} + \left(\frac{3(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)})}{2\beta_1 \vee x_{(n)}}\right)^{\alpha+n}$$

$$-\frac{(\beta_1 \vee x_{(n)} - \beta_0 \wedge x_{(1)})^{\alpha+n}}{2^{\alpha+n+1}(\beta_0 \wedge x_{(1)})^{\alpha+n}}$$

Se escoge aquella hipótesis que tenga la mayor probabilidad.

Capítulo 4

Teoría de la decisión e inferencia estadística

4.1. Representación formal de un problema de decisión

Resolver un problema de decisión es determinar una relación de preferencia $\precsim$ sobre un conjunto $\mathcal{A}$ de acciones posibles, y con ella elegir la acción óptima. Un problema de decisión puede representarse mediante (ξ, ζ, $\mathcal{A}$, $\precsim$), en donde:

$\xi = \{E_1, E_2, ..., E_k\}$, es un conjunto de eventos mutuamente excluyentes y además relevantes para el problema en cuestión, entendiendo por evento una proposición lógica, cuyo valor de verdad o falsedad se determina con base en el resultado de un fenómeno o experimento aleatorio, cuya incertidumbre se puede modelar mediante una medida de probabilidad P.

$\zeta = \mathcal{A} \times \xi$ se denomina conjunto de consecuencias, mismas que se pueden cuantificar mediante una función de utilidad (o de pérdida) $u : \zeta \to \mathbb{R}$.

Un problema de decisión también puede representarse mediante:

$$((\mathcal{A}, \precsim), (\xi, P), (\zeta, u)).$$

Lo que buscamos es determinar la relación de preferencia $\precsim$ en $\mathcal{A}$, es decir, dados a_1, $a_2 \in \mathcal{A}$, si a_2 es igual o más preferible que a_1, se escribe $a_1 \precsim a_2$ y la decisión óptima sería una $a_* \in \mathcal{A}$ tal que $a \precsim a_*$ para toda $a \in \mathcal{A}$.

La relación de preferencia $\precsim$ induce otras dos relaciones:

$$a_1 \sim a_2 : \text{si } a_1 \precsim a_2 \text{ y } a_2 \precsim a_1.$$

La expresión anterior denota indiferencia al elegir una acción u otra.

Por otro lado la siguiente expresión denota que a_1 es estrictamente preferente a a_2:

$$a_1 \prec a_2 : \text{si } a_1 \precsim a_2 \text{ más no } a_1 \sim a_2.$$

A cada acción $a_i \in \mathcal{A}$ se le puede asociar una v.a X_i con $\operatorname{Ran} X_i = \{u_{i\,1}, ..., u_{i\,k}\}$ con función de masa de probabilidades:

$$P(X_i = u_{i\,j}) = P(E_j).$$

La utilidad esperada de la acción a_i se define como :

$$\overline{u}\,(a_i) = E(X_i) = \sum_{j=1}^{k} u_{i\,j} P(E_j).$$

Existen diversos criterios para la solución de un problema de decisión, entre ellos está *el criterio de la utilidad esperada máxima*, veáse [4], donde la determinación de la relación de preferencia está dada por: $a_1 \precsim a_2$ si y solo si $\overline{u}\,(a_1) \leq \overline{u}\,(a_2)$, y en consecuencia:

$$a_1 \sim a_2 \text{ si y solo si } \overline{u}\,(a_1) = \overline{u}\,(a_2).$$

$$a_1 \prec a_2 \text{ si y solo si } \overline{u}\,(a_1) < \overline{u}\,(a_2).$$

Una acción $a_* \in \mathcal{A}$ se considera óptima si:

$$\overline{u}\,(a_*) \geq \overline{u}\,(a_i) \;\text{ para toda } a_i \in \mathcal{A}.$$

Si la acción no fuera única hablaremos de un subconjunto $\mathcal{A}^* \subset \mathcal{A}$ de acciones óptimas.

Decimos que una acción $a_1 \in \mathcal{A}$ está dominada por otra acción $a_2 \in \mathcal{A}$ si $u\,(a_1, E_j) \leq u\,(a_2, E_j)$ para toda E_j, y además existe un j_0 tal que $u\,(a_1, E_{j_0}) < u\,(a_2, E_{j_0})$ y en tal caso se dice que a_1 es una acción inadmisible, y por tanto debiera ser eliminada de $\mathcal{A}$, veáse [6].

4.2. Contraste de hipótesis como problema de decisión

Una aplicación de teoría de la decisión a la inferencia estadística se da en contraste de hipótesis. Si consideramos una m.a $X_1, ..., X_n \sim f(x\,|\,\theta)$ con θ desconocido, y se tienen las hipótesis: $\mathcal{H}_1 : \theta \in \Theta_1, ..., \mathcal{H}_m : \theta \in \Theta_m$ con $\{\Theta_1, ..., \Theta_m\}$ partición de Θ, se definen las acciones $A = \{a_1, ..., a_m\}$ de modo que a_i represente la acción de actuar como si la hipótesis $\mathcal{H}_i$ fuera la correcta. Se define una función de utilidad u de forma que $u(a_i, \mathcal{H}_j)$ es utilidad o bien consecuencia económica, de escoger la hipótesis i cuando en realidad la hipótesis correcta es $\mathcal{H}_j$.

$P(\mathcal{H}_j)$	$P(\mathcal{H}_1)$	$P(\mathcal{H}_2)$	$\cdots$	$P(\mathcal{H}_m)$	
$u(a_i,\mathcal{H}_j)$	$\mathcal{H}_1$	$\mathcal{H}_2$	$\cdots$	$\mathcal{H}_m$	$\overline{u}(a_i)$
a_1	u_{11}	u_{12}	$\cdots$	u_{1m}	$\overline{u}(a_1)$
a_2	u_{21}	u_{22}	$\cdots$	u_{2m}	$\overline{u}(a_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
a_m	u_{m1}	u_{m2}	$\cdots$	u_{mm}	$\overline{u}(a_m)$

La acción óptima a_* es tal que $\overline{u}(a_*) \geq \overline{u}(a_i)$ para toda i, por lo que escogemos $\mathcal{H}_*$.

Típicamente la función de utilidad debiera cumplir para toda i $u(a_i,\mathcal{H}_i) > u(a_i,\mathcal{H}_j)$ para toda $j \neq i$.

Una función de utilidad simple sería: $u(a_i,\mathcal{H}_j) = I_{\{i=j\}}$

$P(\mathcal{H}_j)$	$P(\mathcal{H}_1)$	$P(\mathcal{H}_2)$	$\cdots$	$P(\mathcal{H}_m)$	
$u(a_i,\mathcal{H}_j)$	$\mathcal{H}_1$	$\mathcal{H}_2$	$\cdots$	$\mathcal{H}_m$	$\overline{u}(a_i)$
a_1	1	0	$\cdots$	0	$P(\mathcal{H}_1)$
a_2	0	1	$\cdots$	0	$P(\mathcal{H}_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
a_m	0	0	$\cdots$	1	$P(\mathcal{H}_m)$

Bajo esta función de la utilidad, la acción óptima es escoger la hipótesis de mayor probabilidad.

Por ejemplo, si consideramos una m.a $X_1,...,X_n \sim$ Bernoulli(θ), $0 < \theta < 1$, supongamos que la información inicial se representa mediante una distribución de probabilidad Uniforme(0, 1). Supongamos que se observa la muestra $(0,0,0)$. Como hipótesis tenemos:

$$\mathcal{H}_1 : \theta < 1/3 \qquad \mathcal{H}_2 : 1/3 \leq \theta \leq 2/3 \qquad \mathcal{H}_3 : \theta > 2/3$$

Del ejemplo del Capítulo 3, tenemos que:

$P(\mathcal{H}_1) = 0.8024$
$P(\mathcal{H}_2) = 0.1851$
$P(\mathcal{H}_3) = 0.0123$

Por lo tanto:

$P(\mathcal{H}_j)$	$P(\mathcal{H}_1)$	$P(\mathcal{H}_2)$	$P(\mathcal{H}_3)$	
$u(a_i, \mathcal{H}_j)$	$\mathcal{H}_1$	$\mathcal{H}_2$	$\mathcal{H}_3$	$\overline{u}(a_i)$
a_1	1	0	0	0.8024
a_2	0	1	0	0.1851
a_3	0	0	1	0.0123

Por lo que la acción óptima es a_1, es decir, escoger $\mathcal{H}_1$.

En el caso de tener sólo dos hipotesis, se puede plantear el problema de la siguiente manera:

Si se tiene una m.a $X_1, ..., X_n \sim f(x \,|\, \theta)$, $\theta \in \Theta$, θ desconocida, $\Theta = \Theta_0 \cup \Theta_1$, $\Theta_0 \cap \Theta_1 = \emptyset$ se contrastan las hipótesis:

$$\mathcal{H}_0 : \theta \in \Theta_0 \text{ vs } \mathcal{H}_1 : \theta \in \Theta_1 = \Theta - \Theta_0$$

de donde $j \in \{0, 1\}$:

$$p_j = P(\mathcal{H}_j) = P(\theta \in \Theta_j) = \begin{cases} \int_{\Theta_j} \wp(\theta)\, d\theta, & \text{inicial.} \\ \\ \int_{\Theta_j} \wp(\theta \,|\, \mathbf{x})\, d\theta, & \text{final.} \end{cases}$$

Con $\wp(\theta \,|\, \mathbf{x})$ informativa y no informativa.

	p_0	p_1		
$u(a_i, \mathcal{H}_j)$	$\mathcal{H}_0$	$\mathcal{H}_1$	$\overline{u}(a_i)$	
a_0	u_{11}	u_{12} Error tipo II	$\overline{u}(a_1)$	$= p_0\, u_{11} + p_0\, u_{12}$
a_1	u_{21} Error tipo I	u_{22}	$\overline{u}(a_2)$	$= p_0\, u_{21} + p_1\, u_{22}$

Lo razonable es $u_{11} > u_{12}, \quad u_{22} > u_{21}$.

Podemos ver que el error tipo I es la consecuencia $(a_1, \mathcal{H}_0)$, que es la probabilidad de rechazar la hipótesis $\mathcal{H}_0$ dado que es cierta, y por otra parte el error tipo II es la consecuencia $(a_0, \mathcal{H}_1)$, que es la probabilidad de aceptar la hipótesis $\mathcal{H}_0$ dado que es falsa.

Es de nuestro interés calcular la probabilidad de cometer los errores mencionados, por lo que:

$$P(\text{Error tipo I}) = \begin{cases} p_1, & \text{si } \overline{u}(a_1) \geq \overline{u}(a_0). \\ \\ 0, & \text{si } \overline{u}(a_1) < \overline{u}(a_0). \end{cases}$$

$$P(\text{Error tipo II}) = \begin{cases} p_0, & \text{si } \overline{u}(a_0) \geq \overline{u}(a_1). \\ \\ 0, & \text{si } \overline{u}(a_0) < \overline{u}(a_1). \end{cases}$$

Ejemplo: Consideramos una m.a $X_1, ..., X_n \sim \text{Uniforme}(0, \theta)$, $\theta > 0$. Supongamos que la información inicial se representa mediante una distribución de probabilidad Pareto(α, β). Como hipótesis tenemos:

$$\mathcal{H}_0 : \theta \leq \theta_0 \quad \text{vs} \quad \mathcal{H}_1 : \theta > \theta_0$$

Del ejemplo del Capítulo 1, tenemos que:

$$\wp(\theta \,|\, \mathbf{x}) = \frac{(\beta+n)(\alpha \vee x_{(n)})^{\beta+n}}{\theta^{\beta+n+1}} \, I_{\{\theta > \alpha \vee x_{(n)}\}} \sim \text{Pareto}(\alpha \vee x_{(n)}, \beta + n)$$

Obteniendo $P(\mathcal{H}_0)$ tenemos:

$$P(\mathcal{H}_0) = P(\theta \leq \theta_0) = \int_{-\infty}^{\text{máx}\{\theta_0, \alpha \vee x_{(n)}\}} \wp(\theta \,|\, \mathbf{x}) \, I_{\{\theta > \alpha \vee x_{(n)}\}} d\theta$$

$$= \int_{\alpha \vee x_{(n)}}^{\text{máx}\{\theta_0, \alpha \vee x_{(n)}\}} \frac{(\beta+n)(\alpha \vee x_{(n)})^{\beta+n}}{\theta^{\beta+n+1}} \, d\theta = 1 - (\alpha \vee x_{(n)})^{\beta+n} \, \text{máx}\{\theta_0, \alpha \vee x_{(n)}\}^{-(\beta+n)}$$

por lo que:

$$P(\mathcal{H}_0) = 1 - \left(\frac{\alpha \vee x_{(n)}}{\text{máx}\{\theta_0, \alpha \vee x_{(n)}\}} \right)^{\beta+n}$$

Ahora bien obteniendo $P(\mathcal{H}_1)$ tenemos:

$$P(\mathcal{H}_1) = P(\theta > \theta_0) = \int_{\theta_0}^{\infty} \wp(\theta \mid \mathbf{x})\, I_{\{\theta > \alpha \vee x_{(n)}\}} d\theta$$

$$= \int_{\text{máx}\{\alpha \vee x_{(n)}, \theta_0\}}^{\infty} \frac{(\beta+n)(\alpha \vee x_{(n)})^{\beta+n}}{\theta^{\beta+n+1}}\, d\theta = \left(\frac{\alpha \vee x_{(n)}}{\text{máx}\{\theta_0, \alpha \vee x_{(n)}\}} \right)^{\beta+n}$$

Por lo tanto:

$$P(\mathcal{H}_0) = 1 - \left(\frac{\alpha \vee x_{(n)}}{\text{máx}\{\theta_0, \alpha \vee x_{(n)}\}} \right)^{\beta+n}$$

$$P(\mathcal{H}_1) = \left(\frac{\alpha \vee x_{(n)}}{\text{máx}\{\theta_0, \alpha \vee x_{(n)}\}} \right)^{\beta+n}$$

Si no deseamos contemplar la información inicial lo hacemos no informativa haciendo $\alpha,\ \ \beta \to 0$ es decir:

$$P(\mathcal{H}_0) = 1 - \left(\frac{x_{(n)}}{x_{(n)} \vee \theta_0} \right)^{n}$$

$$P(\mathcal{H}_1) = \left(\frac{x_{(n)}}{x_{(n)} \vee \theta_0}\right)^n.$$

Entonces, con la función de utilidad trivial tenemos que:

	p_0	p_1	
$u(a_i, \mathcal{H}_j)$	$\mathcal{H}_0$	$\mathcal{H}_1$	$\overline{u}(a_i)$
a_0	1	0 Error tipo II	$1 - \left(\frac{x_{(n)}}{x_{(n)} \vee \theta_0}\right)^n$
a_1	0 Error tipo I	1	$\left(\frac{x_{(n)}}{x_{(n)} \vee \theta_0}\right)^n$

Entonces

$$P(\text{Error tipo I}) = \begin{cases} \left(\frac{x_{(n)}}{x_{(n)} \vee \theta_0}\right)^n, & \text{si } \overline{u}(a_1) \geq \overline{u}(a_0). \\ \\ 0, & \text{si } \overline{u}(a_1) < \overline{u}(a_0). \end{cases}$$

$$P(\text{Error tipo II}) = \begin{cases} 1 - \left(\dfrac{x_{(n)}}{x_{(n)} \vee \theta_0}\right)^n, & \text{si } \overline{u}(a_1) \geq \overline{u}(a_0). \\ \\ 0, & \text{si } \overline{u}(a_1) < \overline{u}(a_0). \end{cases}$$

4.3. Estimación puntual como problema de decisión

En el Capítulo 3, vimos que existen diferentes estimadores puntuales vía la media, mediana y moda, la cuestión es saber cuál elegir. Esto es un problema de inferencia que puede ser planteado como un problema de decisión y para ello es necesario proponer una función de utilidad que cuantifique de cierta manera las consecuencias.

Si se tiene la siguiente función de utilidad:

$$\mu(\hat{\theta},\, \theta) = -(\hat{\theta} - \theta)^2 \tag{4.1}$$

entonces

$$\overline{\mu}(\hat{\theta}) = \int_{\Theta} \mu(\hat{\theta}, \theta)\wp(\theta \mid \mathbf{x})\, d\theta =: E_\theta[\mu(\hat{\theta}, \theta)]$$

$$= E_\theta[-(\hat{\theta} - \theta)^2] = -E_\theta[(\hat{\theta} - E_\theta(\theta) + E_\theta(\theta) - \theta)^2]$$

$$= -E_\theta[(\hat{\theta} - E_\theta(\theta))^2 + 2(\hat{\theta} - E_\theta(\theta))(E_\theta(\theta) - \theta) + (E_\theta(\theta) - \theta)^2]$$

$$= -(\hat{\theta} - E_\theta(\theta))^2 + 2(\hat{\theta} - E_\theta(\theta)) * 0 - V_\theta(\theta).$$

El valor $\hat{\theta}_*$ que maximiza $\overline{\mu}(\hat{\theta})$ es el mismo que minimiza $V_\theta(\theta) + (\hat{\theta} - E_\theta(\theta))^2$, $\hat{\theta} \in \Theta$, es decir cuando $\hat{\theta}_* = E_\theta(\theta)$.

Por tanto el estimador puntual para θ bajo función de utilidad cuadrática es $\hat{\theta}_* = E(\theta)$, es decir, la media. Es importante ver que $\theta \sim \wp(\theta)$ o bien $\theta \sim \wp(\theta \,|\, \mathbf{x})$.

En los ejemplos de estimación puntual del Capítulo 2, se obtuvieron algunos estimadores puntuales vía la media.

Ahora bien supongamos que se tiene la siguiente función de utilidad.

$$\mu(\hat{\theta}, \theta) = -|\hat{\theta} - \theta| \tag{4.2}$$

Notemos que la función anterior se maximiza cuando la estimación es exactamente igual al valor teórico del parámetro.
Entonces

$$\mu(\hat{\theta}, \theta) = [(\hat{\theta} - \theta) I_{\{\hat{\theta} < \theta\}}] - [(\hat{\theta} - \theta) I_{\{\hat{\theta} \geq \theta\}}]$$

$$= \hat{\theta} I_{\{\hat{\theta} < \theta\}} - \theta I_{\{\hat{\theta} < \theta\}} - \hat{\theta} I_{\{\hat{\theta} \geq \theta\}} + \theta I_{\{\hat{\theta} \geq \theta\}}$$

$$= \hat{\theta}[I_{\{\theta>\hat{\theta}\}} - I_{\{\theta\leq\hat{\theta}\}}] + \theta I_{\{\theta\leq\hat{\theta}\}} - \theta I_{\{\theta>\hat{\theta}\}}.$$

Sabemos que:

$$\overline{\mu}(\hat{\theta}) = E_\theta[\overline{\mu}(\hat{\theta},\theta)] = E_\theta(-|\hat{\theta}-\theta|)$$

$$= E_\theta[\hat{\theta}[I_{\{\theta>\hat{\theta}\}} - I_{\{\theta\leq\hat{\theta}\}}] + \theta I_{\{\theta\leq\hat{\theta}\}} - \theta I_{\{\theta>\hat{\theta}\}}]$$

$$= \hat{\theta}[E_\theta(I_{\{\theta>\hat{\theta}\}}) - E_\theta(I_{\{\theta\leq\hat{\theta}\}})] + E_\theta(\theta I_{\{\theta\leq\hat{\theta}\}}) - E_\theta(\theta I_{\{\theta>\hat{\theta}\}}).$$

Notemos que:

$$E_\theta(I_{\{\theta>\hat{\theta}\}}) = P(\theta > \hat{\theta}) = 1 - F_\theta(\hat{\theta})$$

$$E_\theta(I_{\{\theta\leq\hat{\theta}\}}) = P(\theta \leq \hat{\theta}) = F_\theta(\hat{\theta})$$

Así pues:

$$E_\theta(I_{\{\theta>\hat{\theta}\}}) - E_\theta(I_{\{\theta\leq\hat{\theta}\}}) = 1 - 2F_\theta(\hat{\theta}).$$

Por otra parte $\theta \sim \wp(\theta)$ o bien $\theta \sim \wp(\theta \,|\, \mathbf{x})$. Usaremos en este caso $\theta \sim \wp(\theta \,|\, \mathbf{x})$

$$E_\theta(\theta I_{\{\theta\leq\hat{\theta}\}}) = \int_\Theta \theta I_{\{\theta\leq\hat{\theta}\}}\wp(\theta\,|\,\mathbf{x})\,d\theta = \int_{-\infty}^{\hat{\theta}} \theta\wp(\theta\,|\,\mathbf{x})\,d\theta$$

$$E_\theta(\theta I_{\{\theta>\hat{\theta}\}}) = \int_\Theta \theta I_{\{\theta>\hat{\theta}\}}\wp(\theta\,|\,\mathbf{x})\,d\theta = \int_{\hat{\theta}}^{\infty} \theta\wp(\theta\,|\,\mathbf{x})\,d\theta$$

Notemos que:

$$E_\theta(\theta I_{\{\theta\leq\hat{\theta}\}}) + E_\theta(\theta I_{\{\theta>\hat{\theta}\}}) = E_\theta(\theta),$$

por lo que:

$$\overline{\mu}(\hat{\theta}) = \hat{\theta}[1-2F_\theta(\hat{\theta})] + 2\int_{-\infty}^{\hat{\theta}} \theta\wp(\theta\,|\,\mathbf{x})\,d\theta - E_\theta(\theta).$$

Entonces:

$$\frac{d}{d\hat{\theta}}\,\overline{\mu}(\hat{\theta}) = 1-2F_\theta(\hat{\theta}) - 2\hat{\theta}\wp(\hat{\theta}\,|\,\mathbf{x}) + 2\hat{\theta}\wp(\hat{\theta}\,|\,\mathbf{x}) - 0 = 1-2F_\theta(\hat{\theta}).$$

Encontrando el punto crítico:

$$\frac{d}{d\hat{\theta}}\,\overline{\mu}(\hat{\theta}) = 0 \Leftrightarrow 1-2F_\theta(\hat{\theta}) = 0$$

$$\Leftrightarrow F_\theta(\hat{\theta}) = \frac{1}{2} \Leftrightarrow \hat{\theta} = \text{mediana}(\theta).$$

Como:

$$\frac{d^2}{d\hat{\theta}^2}\,\overline{\mu}(\hat{\theta}) = -2\wp(\hat{\theta}) < 0$$

Por lo que en $\hat{\theta} = \text{mediana}(\theta)$ se alcanza un máximo.

En los ejemplos de estimación puntual del capítulo 2, se obtuvieron algunos estimadores puntuales vía la mediana.

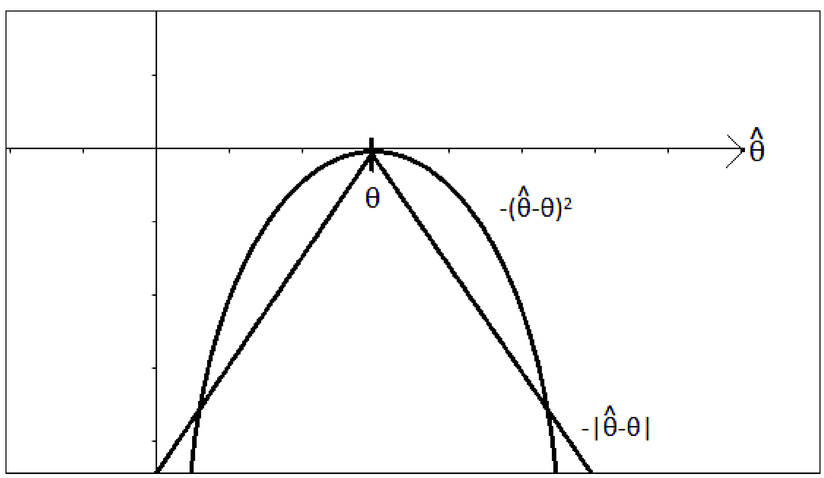

Figura 4.1: Funciones de utilidad media y mediana

Ahora bien supongamos la siguiente función de utilidad.

$$\mu(\hat{\theta}, \theta) = I_{\{\hat{\theta}-\epsilon, \hat{\theta}+\epsilon\}}(\theta) \quad \epsilon > 0 \tag{4.3}$$

Al igual que en los ejemplos anteriores $\hat{\theta} \sim \wp(\theta)$ ó bien $\hat{\theta} \sim \wp(\theta \mid \mathbf{x})$ y usaremos en este caso la distribución final.

Entonces:

$$\overline{\mu}(\hat{\theta}) = E_\theta[\mu(\hat{\theta}, \theta)] = E_\theta[I_{\{\hat{\theta}-\epsilon \leq \theta \leq \hat{\theta}+\epsilon\}}]$$

$$= P(\hat{\theta} - \epsilon \leq \theta \leq \hat{\theta} + \epsilon) = \int_{\hat{\theta}-\epsilon}^{\hat{\theta}+\epsilon} \wp(\theta \mid \mathbf{x})\, d\theta$$

Entonces $\overline{\mu}(\hat{\theta})$ se maximiza cuando $\hat{\theta} = \text{moda}(\theta)$.

Es importante notar que $\wp(\theta \mid \mathbf{x}) \propto L(\theta \mid \mathbf{x})\wp(\theta)$. Sin embargo al tomar el caso no informativo, vemos que éste coincide con el que se conoce como estimador de máxima verosimilitud.

En los ejemplos de estimación puntual del capítulo 2, se obtuvieron algunos estimadores puntuales vía la moda.

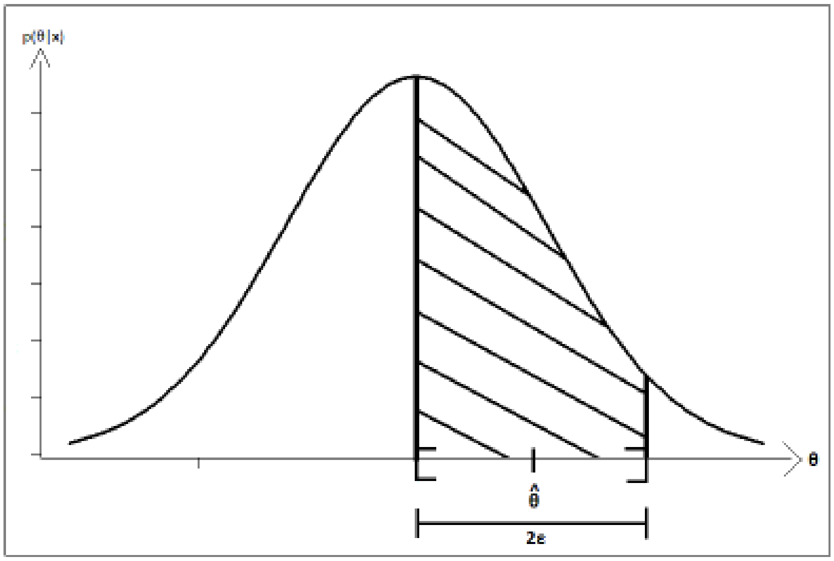

Figura 4.2: Funcion de utilidad moda

4.4. Estimación de conjunto como problema de decisión

La estimación de conjunto también puede plantearse como un problema de decisión, ya que existen múltiples soluciones para C en la ecuación $\int_C \wp(\theta \mid \mathbf{x})\, d\theta = \gamma$.

Al conjunto de soluciones posibles lo denotaremos por:

$$\mathcal{A} = \left\{ C \subset \Theta : \int_C \wp(\theta \mid \mathbf{x})\, d\theta = \gamma \right\},$$

donde $\mathcal{A}$ representa el conjunto de acciones, en el que cada acción es una de los distintos subconjuntos que podemos elegir.

Propongamos una función de utilidad tal que se prefieran las regiones C de menor tamaño posible, el cual se denotará mediante $||C||$, y que contengan el verdadero valor de θ:

$$u(C, \theta) = -k||C|| + 1_C(\theta), \quad k > 0.$$

La función de utilidad esperada para cada $C \in \mathcal{A}$ se obtiene mediante:

$$\overline{u}(C) = \int_{\Theta} u(C, \theta) \wp(\theta \,|\, \mathbf{x}) \, d\theta = -k||C|| + \gamma.$$

Por lo tanto, la región óptima será aquella $C^* \in \mathcal{A}$ tal que su tamaño $||C^*||$ sea el mínimo, donde C^* se conoce como conjunto de probabilidad γ de máxima densidad. En los ejemplos de estimación por regiones del Capítulo 2, se obtuvieron conjuntos de tamaño mínimo.

Capítulo 5

Predicción

Es de nuestro interés hacer el cálculo de probabilidades de una observación futura de la v.a X, el cual hacemos mediante la distribución predictiva final 1.1, veáse [5].

5.1. Estimación puntual predictiva

Ejemplo: Tenemos una m.a $X_1, ..., X_n \sim$ Uniforme$(a,\, b)$ $a < b$ con $a,\, b \in \mathbb{R}$ con una función de densidad $f(x \,|\, a,\, b) = \dfrac{1}{b-a}\, I_{\{a < x < b\}}$ entonces

$$L(a, b \,|\, \mathbf{x}) = \prod_{i=1}^{n} f(x_i \,|\, a,\, b) = \frac{1}{(b-a)^n}\, I_{\{a < x_{(1)}\}}\, I_{\{b > x_{(n)}\}}.$$

Del Capítulo 1, tenemos que la distribución predictiva final es:

$$\wp(x \,|\, \mathbf{x}) = \frac{(\alpha+n)(B_1 - B_0)^{\alpha+n}}{(\alpha+n+2)(x \vee B_1 - x \wedge B_0)^{\alpha+n+1}},$$

donde $B_1 = \beta_1 \vee x_{(n)}$ y $B_0 = \beta_0 \wedge x_{(1)}$

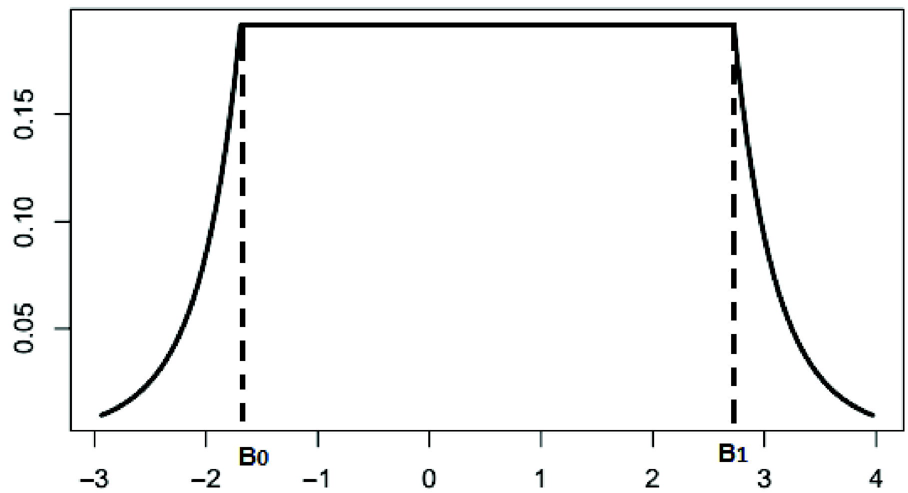

Figura 5.1: Predictiva final

Al utilizar la media como medida de tendencia central una estimación puntual predictiva de x sería:

$$\hat{x} = E(x) = \int_{-\infty}^{\infty} x \, \wp(x \,|\, \mathbf{x}) \, dx$$

$$= \int_{-\infty}^{\infty} x \, \frac{(\alpha+n)(B_1 - B_0)^{\alpha+n}}{(\alpha+n+2)(x \vee B_1 - x \wedge B_0)^{\alpha+n+1}} \, dx$$

.

Caso 1: $x < B_0$.

$$\hat{x} = \frac{(\alpha+n)(B_1 - B_0)^{\alpha+n}}{(\alpha+n+2)} \int_{-\infty}^{B_0} \frac{x}{(B_1 - x)^{\alpha+n+1}} \, dx$$

Con el cambio de variable $u = B_1 - x \quad du = -dx$ tenemos:

$$= -\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \int_{\infty}^{B_1-B_0} (B_1-u)\, u^{-\alpha-n-1}\, du$$

$$= -\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \left(\frac{B_1\ u^{-\alpha-n}}{-(\alpha+n)} \Bigg|_{\infty}^{B_1-B_0} - \frac{u^{-\alpha-n+1}}{-(\alpha+n-1)} \Bigg|_{\infty}^{B_1-B_0} \right)$$

$$= \frac{B_1}{\alpha+n+2} - \frac{(\alpha+n)(B_1-B_0)}{(\alpha+n+2)(\alpha+n-1)}.$$

Por lo tanto el estimador puntual es:

$$\hat{x} = \frac{B_1}{\alpha+n+2} - \frac{(\alpha+n)(B_1-B_0)}{(\alpha+n+2)(\alpha+n-1)}.$$

Caso 2: $B_0 < x < B_1$.

$$\hat{x} = \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \int_{B_0}^{B_1} \frac{x}{(B_1-B_0)^{\alpha+n+1}}\, dx = \frac{(B_1^2-B_0^2)(\alpha+n)}{2(\alpha+n+2)(B_1-B_0)}.$$

Por lo tanto el estimador puntual es:

$$\hat{x} = \frac{(B_1^2-B_0^2)(\alpha+n)}{2(\alpha+n+2)(B_1-B_0)}.$$

Caso 3: $x > B_1$.

$$\hat{x} = \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \int_{B_1}^{\infty} \frac{x}{(x-B_0)^{\alpha+n+1}}\, dx.$$

Con el cambio de variable $u = x - B_0 \quad du = dx$ tenemos:

$$= \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \int_{B_1-B_0}^{\infty} (u+B_0)\, u^{-\alpha-n-1}\, du$$

$$= \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \left(\frac{B_0\ u^{-\alpha-n}}{-(\alpha+n)} \Bigg|_{B_1-B_0}^{\infty} + \frac{u^{-\alpha-n+1}}{-(\alpha+n-1)} \Bigg|_{B_1-B_0}^{\infty} \right)$$

$$= \frac{(\alpha+n)(B_1-B_0)}{(\alpha+n+2)(\alpha+n-1)} + \frac{B_0}{\alpha+n+2}.$$

Por lo tanto el estimador puntual es:

$$\hat{x} = \frac{(\alpha+n)(B_1-B_0)}{(\alpha+n+2)(\alpha+n-1)} + \frac{B_0}{\alpha+n+2}.$$

En el cápitulo anterior vimos que puede plantearse la estimación puntual como un problema de decisión. Con la siguiente función de utilidad se obtiene como estimador puntual la media de una observación futura.

$$\mu(\hat{x},\, x) = -(\hat{x}-x)^2 \tag{5.1}$$

Notemos que esta es la misma expresión obtenida anteriormente.

Ahora se utilizará la mediana:

Encontrando la función de distribución

Caso 1: $x < B_0$.

$$\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)}\, I_{\{t<B_0\}} \int_{-\infty}^{t} \frac{1}{(B_1-x)^{\alpha+n+1}}\, dx \;+\; I_{\{t\geq B_0\}}.$$

Con el cambio de variable: $u = B_1 - x,\ du = -dx$:

$$\begin{aligned} &\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)}\, I_{\{t<B_0\}} \left.\frac{u^{-\alpha-n}}{(\alpha+n)}\right|_{\infty}^{B_1-t} \;+\; I_{\{t\geq B_0\}} \\ &= \frac{(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(B_1-t)^{\alpha+n}}\, I_{\{t<B_0\}} \;+\; I_{\{t\geq B_0\}}. \end{aligned}$$

Entonces:

$$\frac{(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(B_1-t)^{\alpha+n}} = \frac{1}{2}.$$

Al despejar t, tenemos:

$$t = B_1 - \left(\frac{2(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)}\right)^{1/\alpha+n}.$$

Caso 2: $B_0 < x < B_1$.

$$\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)}\, I_{\{B_0<t<B_1\}} \int_{B_0}^{t} \frac{1}{(B_1-B_0)^{\alpha+n+1}}\, dx \;+\; I_{\{t<B_0\}}\, I_{\{t>B_1\}}$$

$$= \frac{(\alpha+n)}{(\alpha+n+2)(B_1-B_0)}(t-B_0)\, I_{\{B_0<t<B_1\}} \;+\; I_{\{t<B_0\}}\, I_{\{t>B_1\}}.$$

Entonces:

$$\frac{(\alpha+n)}{(\alpha+n+2)(B_1-B_0)}(t-B_0) = \frac{1}{2}.$$

Al despejar t, tenemos:

$$t = B_0 + \frac{2(\alpha+n)}{(\alpha+n+2)(B_1-B_0)}.$$

Caso 3: $B_1 < x$.

$$\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)}\, I_{\{t>B_1\}} \int_{B_1}^{t} \frac{1}{(x-B_0)^{\alpha+n+1}}\, dx \;+\; I_{\{t\leq B_1\}}.$$

Con el cambio de variable: $u = x - B_0,\ du = dx$:

$$\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} I_{\{t>B_1\}} \frac{u^{-\alpha-n}}{-(\alpha+n)}\Bigg|_{B_1-B_0}^{t-B_0} + I_{\{t\leq B_1\}}$$

$$= \frac{1}{(\alpha+n+2)} - \frac{(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(t-B_0)^{\alpha+n}} I_{\{t>B_1\}} + I_{\{t\leq B_1\}}.$$

Entonces:

$$\frac{1}{(\alpha+n+2)} - \frac{(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(t-B_0)^{\alpha+n}} = \frac{1}{2}.$$

Al despejar t, tenemos:

$$t = B_0 + \left(\frac{2(B_1-B_0)^{\alpha+n}}{(-\alpha-n)}\right)^{1/\alpha+n}.$$

En el capítulo anterior vimos que puede plantearse la estimación puntual como un problema de decisión. Con la siguiente función de utilidad se obtiene como estimador puntual la mediana de una observación futura.

$$\mu(\hat{x}, x) = -|\hat{x} - x| \tag{5.2}$$

Notemos que esta expresión es la misma obtenida anteriormente.

Ahora utilizando la moda, que es donde se alcanza el supremo, podemos ver que este estimador no es único, ya que existen una infinidad de soluciones en el conjunto $[B_0, B_1]$. Por lo tanto, tomaremos como estimador puntual $\hat{x}$ al punto medio del conjunto.

5.2. Estimación de conjunto predictiva

Del ejemplo anterior tenemos que una estimación de conjunto para la función predictiva final sería:

Caso 1: $x < B_0$.

$$\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \int_{\hat{\theta}_i}^{B_0} \frac{1}{(B_1-x)^{\alpha+n+1}}\, dx.$$

Con el cambio de variable: $u = B_1 - x,\ du = -dx$:

$$\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \left. \frac{u^{-\alpha-n}}{(\alpha+n)} \right|_{B_1-\hat{\theta}_i}^{B_1-B_0} = \frac{1}{\alpha+n+2} - \frac{(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(B_1-\hat{\theta}_i)^{\alpha+n}}$$

Entonces:

$$\frac{1}{\alpha+n+2} - \frac{(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(B_1-\hat{\theta}_i)^{\alpha+n}} = \gamma.$$

De donde se sigue que:

$$\hat{\theta}_i = B_1 - \left(\frac{(B_1-B_0)^{\alpha+n}}{1-(\alpha+n+2)\,\gamma} \right)^{1/\alpha+n},$$

Por lo cual el conjunto buscado es:

$$\left(B_1 - \left(\frac{(B_1-B_0)^{\alpha+n}}{1-(\alpha+n+2)\,\gamma} \right)^{1/\alpha+n},\ B_0 \right).$$

Caso 2: $B_0 < x < B_1$.

$$\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)}\int_{B_0}^{B_1}\frac{1}{(B_1-B_0)^{\alpha+n+1}}\,dx=\frac{\alpha+n}{(\alpha+n+2)}.$$

Notemos que no existe un conjunto para este caso, ya que sólo es una constante.

Caso 3: $x > B_1$.

$$\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)}\int_{B_1}^{\hat{\theta}_s}\frac{1}{(x-B_0)^{\alpha+n+1}}\,dx.$$

Con el cambio de variable $u = x - B_0 \quad du = dx$ tenemos:

$$\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)}\;\frac{u^{-\alpha-n}}{-(\alpha+n)}\bigg|_{B_1-B_0}^{\hat{\theta}_s-B_0}=\frac{1}{\alpha+n+2}-\frac{(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(\hat{\theta}_s-B_0)^{\alpha+n}}.$$

Entonces:

$$\frac{1}{\alpha+n+2}-\frac{(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(\hat{\theta}_s-B_0)^{\alpha+n}}=\gamma.$$

De donde:

$$\hat{\theta}_s = B_0 + \left(\frac{(B_1 - B_0)^{\alpha+n}}{1-(\alpha+n+2)\,\gamma} \right)^{1/\alpha+n}.$$

Por lo cual el conjunto buscado es:

$$\left(B_1 \ , \ B_0 + \left(\frac{(B_1 - B_0)^{\alpha+n}}{1-(\alpha+n+2)\,\gamma} \right)^{1/\alpha+n} \right).$$

En el capítulo anterior vimos que puede plantearse la estimación por regiones como un problema de decisión el cual coincide con lo anterior.

5.3. Contraste de hipótesis predictivo

Del ejemplo anterior tenemos que:

$$\wp(x \,|\, \mathbf{x}) = \begin{cases} \dfrac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(B_1-x)^{\alpha+n+1}} & \text{si } x \leq B_0. \\[2em] \dfrac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(B_1-B_0)^{\alpha+n+1}} & \text{si } B_0 < x < B_1. \\[2em] \dfrac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(x-B_0)^{\alpha+n+1}} & \text{si } x \geq B_1. \end{cases}$$

Tomando las siguientes hipótesis:

$$\mathcal{H}_1 : X_{n+1} \leq -1 \quad \mathcal{H}_2 : -1 < X_{n+1} < 3 \quad \mathcal{H}_3 : X_{n+1} \geq 3$$

Se tiene que:

$$P(\mathcal{H}_1) = P(X_{n+1} \leq -1) = \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \int_{-\infty}^{B_0} \frac{1}{(B_1-x)^{\alpha+n+1}}\,dx$$

$$+\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(B_1-B_0)^{\alpha+n+1}} \int_{B_0}^{-1} dx.$$

Con el cambio de variable: $u = B_1 - x,\ du = -dx$:

$$= -\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \left.\frac{u^{-\alpha-n}}{-(\alpha+n)}\right|_{\infty}^{B_1-B_0} + \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(B_1-B_0)^{\alpha+n+1}}(-1-B_0).$$

Por lo tanto:

$$P(\mathcal{H}_1) = \frac{1}{\alpha+n+2} - \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}(1+B_0)}{(\alpha+n+2)(B_1-B_0)^{\alpha+n+1}}.$$

Por otro lado:

$$P(\mathcal{H}_2) = P(-1 < X_{n+1} < 3) = \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(B_1-B_0)^{\alpha+n+1}} \int_{-1}^{B_1} dx$$

$$+\frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \int_{B_1}^{3} \frac{1}{(x-B_0)^{\alpha+n+1}}\,dx.$$

Con el cambio de variable $u = x - B_0,\ du = dx$:

$$= \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}(B_1+1)}{(\alpha+n+2)(B_1-B_0)^{\alpha+n+1}} + \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \left. \frac{u^{-\alpha-n}}{-(\alpha+n)} \right|_{B_1-B_0}^{3-B_0} dx.$$

Por lo tanto:

$$P(\mathcal{H}_2) = \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}(B_1+1)}{(\alpha+n+2)(B_1-B_0)^{\alpha+n+1}} - \frac{(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(3-B_0)^{\alpha+n}} + \frac{1}{\alpha+n+2}.$$

Y finalmente:

$$P(\mathcal{H}_3) = P(X_{n+1} \geq 3) = \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \int_3^\infty \frac{1}{(x-B_0)^{\alpha+n+1}}\, dx.$$

Con el cambio de variable $u = x - B_0,\ du = dx$:

$$= \frac{(\alpha+n)(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)} \left. \frac{u^{-\alpha-n}}{-(\alpha+n)} \right|_{3-B_0}^{\infty}.$$

Por lo tanto:

$$P(\mathcal{H}_3) = \frac{(B_1-B_0)^{\alpha+n}}{(\alpha+n+2)(3-B_0)^{\alpha+n}}.$$

En el capítulo anterior vimos que puede plantearse el contraste de hipótesis como un problema de decisión.

Por ejemplo, utilizando las hipótesis anteriores y tomando en cuenta que $\alpha = 1, \quad B_0 = -1.5, \quad B_1 = 3.5$ *y* $n = 3$ tenemos:

$P(\mathcal{H}_j)$	$P(\mathcal{H}_1)$	$P(\mathcal{H}_2)$	$P(\mathcal{H}_3)$	
$u(a_i, \mathcal{H}_j)$	$\mathcal{H}_1$	$\mathcal{H}_2$	$\mathcal{H}_3$	$\overline{u}(a_i)$
a_1	1	0	0	0.2333
a_2	0	1	0	0.5126
a_3	0	0	1	0.2540

Por lo que la acción óptima es a_2.

5.4. Cadena de Markov de dos estados

Ejemplo: Tenemos una cadena de Markov homogénea de dos estados $E = \{0, 1\}$ con $\{X_n : n = 0, 1, ...\}$ y una distribución inicial $V_0 = (\gamma \quad 1-\gamma)$ cuya matriz de transición es:

$$\mathcal{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

Con parámetros $\alpha, \beta, \gamma \in (0, 1)$

Obteniendo la función $L(\alpha, \beta, \gamma \,|\, \mathbf{x})$ tenemos:

$$L(\alpha, \beta, \gamma \,|\, \mathbf{x}) = P(X_0 = x_0,\ X_1 = x_1,\ ...,\ X_n = x_n \,|\, \alpha, \beta, \gamma)$$

$$= V_0(x_0)\, P_{x_0 x_1}\, P_{x_1 x_2} ... P_{x_{n-1} x_n} = \gamma^{1-x_0}(1-\gamma)^{x_0}(1-\alpha)^{n_{00}}\alpha^{n_{01}}\beta^{n_{10}}(1-\beta)^{n_{11}}$$

Donde n_{ij} es el número de transiciones de i a j, $P_{i\,j} \in \{\alpha,\, 1-\alpha,\, \beta,\, 1-\beta\}$ y además $n_{00} + n_{01} + n_{10} + n_{11} = n$.

Notemos que $\gamma^{1-x_0}(1-\gamma)^{x_0}$ es el kernel de una Beta$(1 - x_0 + 1, x : 0 + 1)$, $(1-\alpha)^{n_{00}}\alpha^{n_{01}}$ es el kernel de una Beta$(n_{01}+1, n_{00}+1)$ y además $\beta^{n_{10}}(1-\beta)^{n_{11}}$ es el kernel de una Beta$(n_{10} + 1, n_{11} + 1)$.

Propongamos una función de distribución inicial que además sea conjugada de la siguiente manera:

$$\wp(\alpha, \beta, \gamma) = \wp_1(\alpha)\, \wp_2(\beta)\, \wp_3(\gamma),$$

De donde:

$$\begin{aligned}
\wp_1(\alpha) &\sim \text{Beta}(a_1, b_1),\\
\wp_2(\beta) &\sim \text{Beta}(a_2, b_2),\\
\wp_3(\gamma) &\sim \text{Beta}(a_3, b_3).
\end{aligned}$$

Obteniendo la función de distribución final 1.2 tenemos que:

$$\wp(\alpha, \beta, \gamma \,|\, \mathbf{x}) = \text{Beta}(\alpha \,|\, a_1 + n_{01}, b_1 + n_{00}) \cdot \text{Beta}(\beta \,|\, a_2 + n_{10}, b_2 + n_{11})$$
$$\cdot \text{Beta}(\gamma \,|\, a_3 + 1 - x_0, b_3 + x_0),$$

Donde x_0 puede ser 0 ó 1.

ESTIMACIÓN PUNTUAL

Utilizando la media como medida de tendencia central las estimaciones puntuales iniciales son:

$$\hat{\alpha} = E(\wp_1(\alpha)) = \frac{a_1}{a_1 + b_1},$$

$$\hat{\beta} = E(\wp_2(\beta)) = \frac{a_2}{a_2 + b_2},$$

$$\hat{\gamma} = E(\wp_3(\gamma)) = \frac{a_3}{a_3 + b_3}.$$

Y las estimacionnes puntuales finales son:

$$\hat{\alpha} = E(\mathrm{Beta}(a_1 + n_{01}, b_1 + n_{00})) = \frac{a_1 + n_{01}}{a_1 + b_1 + n_{01} + n_{00}},$$

$$\hat{\beta} = E(\mathrm{Beta}(a_2 + n_{10}, b_2 + n_{11})) = \frac{a_2 + n_{10}}{a_2 + b_2 + n_{10} + n_{11}},$$

$$\hat{\gamma} = E(\mathrm{Beta}(a_3 + 1 - x_0, b_3 + x_0)) = \frac{a_3 + 1 - x_0}{a_3 + b_3 + 1}.$$

Si quisieramos que se desvaneciera la influencia de la información inicial, hacemos $a_i \to 0$, $b_i \to 0$

Entonces:

$$\hat{\alpha} = \frac{n_{01}}{n_{01} + n_{00}},$$

$$\hat{\beta} = \frac{n_{10}}{n_{10} + n_{11}},$$

$$\hat{\gamma} = 1 - x_0.$$

Sabemos que la distribución Beta es continua. Entonces utilizando la moda, que es donde alcanza el supremo, los estimadores son:

$$\hat{\alpha} = \frac{a_1 + n_{01} + \sum x_i - 1}{a_1 + n_{01} + b_1 + n_{00} + n - 2},$$

$$\hat{\beta} = \frac{a_2 + n_{10} + \sum x_i - 1}{a_2 + n_{10} + b_2 + n_{11} + n - 2},$$

$$\hat{\gamma} = \frac{a_3 + \sum x_i - x_0}{a_3 + b_3 + n - 1}.$$

Si quisieramos que se desvaneciera la influencia de la información inicial, hacemos a_i y $b_i \to 0$, por lo que los estimadores que buscamos son:

$$\hat{\alpha} = \frac{n_{01} + \sum x_i - 1}{n_{01} + n_{00} + n - 2},$$

$$\hat{\beta} = \frac{n_{10} + \sum x_i - 1}{n_{10} + n_{11} + n - 2},$$

$$\hat{\gamma} = \frac{\sum x_i - x_0}{n-1}.$$

Para encontrar el estimador puntual vía la mediana:

$$\hat{\theta} = F^{-1}(1/2 \,|\, \mathbf{x}).$$

Es importante observar que no es posible obtener una función de distribución explícita para el modelo Beta, como se mecionó anteriormente.

Ahora trabajaremos en hacer el cálculo de probabilidades de una observación futura.

$$P(X_{n+1} = x_{n+1},\, X_{n+2} = x_{n+2},\, ...,\, X_{n+m} = x_{n+m} \,|\, X_n = x_n,\, ...\, X_0 = x_0)$$

$$= P(X_{n+1} = x_{n+1},\, X_{n+2} = x_{n+2},\, ...,\, X_{n+m} = x_{n+m} \,|\, X_n = x_n)$$
$$= P(X_{n+m} = x_{n+m} \,|\, X_{n+m-1} = x_{n+m-1}) ... P(X_{n+1} = x_{n+1} \,|\, X_n = x_n)$$
$$= \alpha^{k_{01}}(1-\alpha)^{k_{00}}\beta^{k_{10}}(1-\beta)^{k_{11}},$$

Donde k_{ij} es el número de transiciones de i a j de datos futuros, $P_{ij} \in \{\alpha,\, 1-\alpha,\, \beta,\, 1-\beta\}$ y además $k_{00} + k_{01} + k_{10} + k_{11} = k$.

Notemos que $\alpha^{k_{01}}(1-\alpha)^{k_{00}}$ es el kernel de una Beta$(k_{01}+1, k_{00}+1)$ y además $\beta^{k_{10}}(1-\beta)^{k_{11}}$ es el kernel de una Beta$(k_{10}+1, k_{11}+1)$.

Obteniendo la función de distribución predictiva final 1.1 tenemos que:

$$\wp(x \,|\, \mathbf{x}) = \int_0^1 \int_0^1 \int_0^1 \alpha^{k_{01}}(1-\alpha)^{k_{00}} \cdot \beta^{k_{10}}(1-\beta)^{k_{11}}$$
$$\cdot \mathrm{Beta}(\alpha \,|\, a_1 + n_{01}, b_1 + n_{00}) \cdot \mathrm{Beta}(\beta \,|\, a_2 + n_{10}, b_2 + n_{11})$$

$$\cdot \text{Beta}(\gamma \,|\, a_3 + 1 - x_0, b_3 + x_0)\, d\alpha\, d\beta\, d\gamma.$$

Notemos que podemos expresar lo anterior como un producto de tres integrales por la independencia de los parámetros, es decir:

$$\wp(x \,|\, \mathbf{x}) = \frac{\Gamma(a_1 + n_{01} + b_1 + n_{00})}{\Gamma(a_1 + n_{01})\Gamma(b_1 + n_{00})} \int_0^1 \alpha^{a_1+n_{01}+k_{01}-1}(1-\alpha)^{b_1+n_{00}+k_{00}-1}\, d\alpha$$
$$\cdot \frac{\Gamma(a_2 + n_{10} + b_2 + n_{11})}{\Gamma(a_2 + n_{10})\Gamma(b_2 + n_{11})} \int_0^1 \beta^{a_2+n_{10}+k_{10}-1}(1-\beta)^{b_2+n_{11}+k_{11}-1}\, d\beta$$
$$\cdot \frac{\Gamma(a_3 + 1 - x_0 + b_3 + x_0)}{\Gamma(a_3 + 1 - x_0)\Gamma(b_3 + x_0)} \int_0^1 \gamma^{a_3+1-x_0}(1-\gamma)^{b_3+x_0-1}\, d\gamma.$$

Puesto que x_0 y x_1 sólo pueden tomar los valores de 0 ó 1 y además sabemos que:

$$\int_0^1 z^{a+1-1}(1-z)^{b+1-1}\, dz = \frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+2)}$$

Se sigue que:

$$\wp(x \,|\, \mathbf{x}) = \frac{\Gamma(a_1 + n_{01} + b_1 + n_{00})}{\Gamma(a_1 + n_{01})\Gamma(b_1 + n_{00})} \cdot \frac{\Gamma(a_1 + n_{01} + k_{01})\Gamma(b_1 + n_{00} + k_{00})}{\Gamma(a_1 + n_{01} + k_{01} + b_1 + n_{00} + k_{00})}$$
$$\cdot \frac{\Gamma(a_2 + n_{10} + b_2 + n_{11})}{\Gamma(a_2 + n_{10})\Gamma(b_2 + n_{11})} \cdot \frac{\Gamma(a_2 + n_{10} + k_{10})\Gamma(b_2 + n_{11} + k_{11})}{\Gamma(a_2 + n_{10} + k_{10} + b_2 + n_{11} + k_{11})}.$$

ESTIMACIÓN PUNTUAL PREDICTIVA

Utilizando la media como medida de tendencia central y tomando en cuenta que x sólo puede tomar el valor de 0 ó 1, las estimaciónes puntuales predictivas son:

$$\hat{x} = E(x) = \sum_{0}^{1} x \wp(x \,|\, \mathbf{x}) = \wp(x \,|\, \mathbf{x})$$

$$= \wp(x \,|\, \mathbf{x}) = \frac{\Gamma(a_1 + n_{01} + b_1 + n_{00})}{\Gamma(a_1 + n_{01})\Gamma(b_1 + n_{00})} \cdot \frac{\Gamma(a_1 + n_{01} + k_{01})\Gamma(b_1 + n_{00} + k_{00})}{\Gamma(a_1 + n_{01} + k_{01} + b_1 + n_{00} + k_{00})}$$
$$\cdot \frac{\Gamma(a_2 + n_{10} + b_2 + n_{11})}{\Gamma(a_2 + n_{10})\Gamma(b_2 + n_{11})} \cdot \frac{\Gamma(a_2 + n_{10} + k_{10})\Gamma(b_2 + n_{11} + k_{11})}{\Gamma(a_2 + n_{10} + k_{10} + b_2 + n_{11} + k_{11})}.$$

Si quisieramos que se desvaneciera la influencia de la información inicial, hacemos $a_i \to 0$, $b_i \to 0$

Entonces:

$$\hat{x} = \wp(x \,|\, \mathbf{x}) = \frac{\Gamma(n_{01} + n_{00})}{\Gamma(n_{01})\Gamma(n_{00})} \cdot \frac{\Gamma(n_{01} + k_{01})\Gamma(n_{00} + k_{00})}{\Gamma(n_{01} + k_{01} + n_{00} + k_{00})}$$
$$\cdot \frac{\Gamma(n_{10} + n_{11})}{\Gamma(n_{10})\Gamma(n_{11})} \cdot \frac{\Gamma(n_{10} + k_{10})\Gamma(n_{11} + k_{11})}{\Gamma(n_{10} + k_{10} + n_{11} + k_{11})}.$$

Ahora bien, utilizando la moda, que es donde alcanza el supremo, el estimador buscado es 1.

ESTIMACIÓN DE CONJUNTO

En el Capítulo 3 vimos en qué consiste la estimación por conjunto:

$$P(\hat{\theta}_i \leq \theta \leq \hat{\theta}_s) = \int_{\hat{\theta}_i}^{\hat{\theta}_s} \wp(\theta \,|\, \mathbf{x}) d\theta = F(\theta_s \,|\, \mathbf{x}) - F(\theta_i \,|\, \mathbf{x}).$$

Pero también sabemos que en el caso del modelo Beta, no se tiene función de distribución explícita, por lo que en este ejemplo, la estimación por conjunto para cada marginal se obtiene numéricamente.

CONTRASTE DE HIPÓTESIS.

Supongamos que se tienen las siguientes hipótesis:

$$\mathcal{H}_0 = \alpha \leq 1 - \beta \qquad \mathcal{H}_1 = \alpha > 1 - \beta$$

Obteniendo $P(\mathcal{H}_0)$ tenemos:

$$P(\alpha \leq 1 - \beta) = \int_0^1 \int_0^1 \wp(\alpha, \beta \,|\, \mathbf{x}) I_{\{\beta \leq 1-\alpha\}} \, d\beta \, d\alpha$$

$$= \int_0^1 \int_0^{1-\alpha} \wp(\alpha, \beta \,|\, \mathbf{x}) \, d\beta \, d\alpha,$$

Donde $\wp(\alpha, \beta \,|\, \mathbf{x}) = \text{Beta}(\alpha \,|\, a_1 + n_{01}, b_1 + n_{00}) \cdot \text{Beta}(\beta \,|\, a_2 + n_{10}, b_2 + n_{11})$.
Por lo que:

$$P(\alpha \leq 1-\beta) = \int_0^1 \int_0^{1-\alpha} \text{Beta}(\alpha \,|\, a_1+n_{01}, b_1+n_{00}) \cdot \text{Beta}(\beta \,|\, a_2+n_{10}, b_2+n_{11}) \, d\beta \, d\alpha.$$

Es importante observar que no es posible resolver la expresión anterior por métodos usuales, por lo cual, se utilizó la herramienta R para su aproximación, dando como resultado de la misma 0.8025721, véase A.3.

La figura 5.2 esquematiza la región de integración.

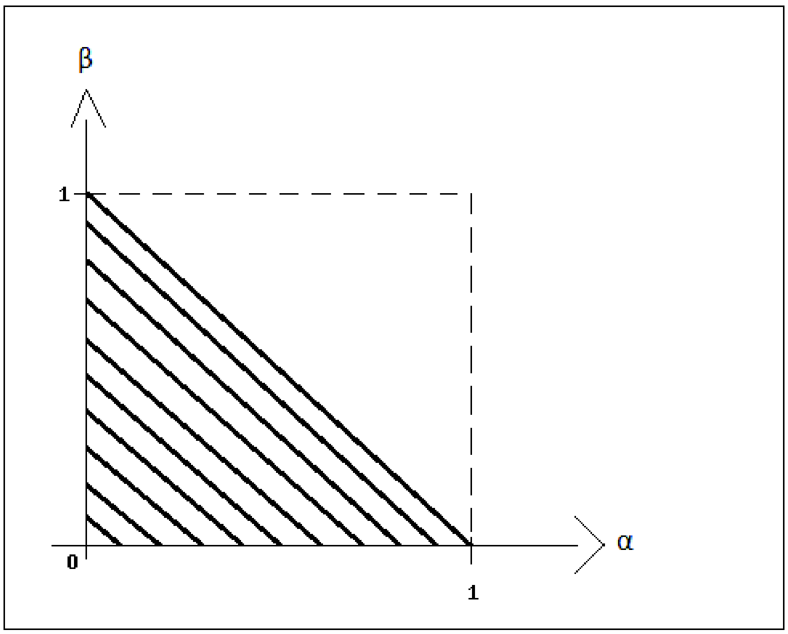

Figura 5.2: Región de integración

Ahora bien:

$$P(\alpha > 1-\beta) = 1 - P(\alpha \leq 1-\beta) = 1 - 0.8025721 = 0.1974279.$$

Por lo tanto:

$$P(\alpha \leq 1-\beta) = 0.8025721.$$

$$P(\alpha > 1-\beta) = 0.1974279.$$

Conclusiones

En conclusión, tras el análisis de diverso material bibliográfico, se logró crear un escrito que aborda los principales temas de la inferencia estadística paramétrica logrando un enfoque integral en contraste con la práctica usual, en idioma español, y que presenta desarrollo conceptual y ejemplos para un mejor entendimiento de los temas abordados.

Este trabajo implicará que la inferencia estadística parámetrica ya no se vea desde dos puntos diferentes, si no de una manera unida, que se entienda que se puede enseñar y aprender en conjunto, sin necesidad de enfatizar tanto entre las diferencias, sino mejor en la relación que pueden llegar a tener, logrando un entendimiento más rápido y una postura equilibrada de ambos enfoques.

Es recomendable el uso de herramientas computacionales, pues estás fueron de mucha ayuda para el desarrollo de algunos temas, ya que no siempre resulta factible el cálculo manual y explícito de las fórmulas.

Existen ya materiales que tratan de presentar un enfoque integral, pero la diferencia con respecto al presente, está principalmente en el idioma y en que en este trabajo siempre se procuró no mencionar ningún enfoque, sino realmente hacer algo integral.

Entre las limitaciones de la investigación se encuentra el poco material, en idioma español, que existe, y que podría utilizarse como elemento bibliográfico, así como el uso de herramientas computacionales que no todos los lectores utilizan y que podrían significar un problema para el entendimiento completo de los temas.

Apéndice A

Código de programación en R

A.1. Inferencia estadística paramétrica

```
#GRAFICA DE LOCALIZACION

curve(dnorm(x,0,1),from=-5, to=7,lwd=3)
curve(dnorm(x,3,1),add=T,from=-5, to=7,lwd=3)

#GRAFICA DE ESCALA

curve(dnorm(x,0,1),from=-5, to=7,lwd=3)
curve(dnorm(x,0,2),add=T,from=-5, to=7,lwd=3)


#GRAFICA DEL MODELO BETA

a <- 2.5; b <- 3.5; alf <- 1/4; bet <- 1.5
z <- seq(0, 1, length = 1000) # Ran Z
x <- seq(0 + alf, bet + alf, length = 1000)  # Ran X
plot(z, dbeta(z, a, b), xlim = c(0, bet + alf), type = "l",lwd=3)
lines(x, (1/bet)*dbeta((x - alf)/bet, a, b), type = "l",lwd=3)

#CONSTANTE K DE INTEGRACION DEL MODELO KUMARASWAMY
```

```
t <- 0.7
r <- 0.5
v <- 0.8
z <- 0.95

# Creando la densidad inicial Kumaraswamy(a, b)

dkuma <- function(w, a, b)
         a*b*(w^(a - 1))*((1 - w^a)^(b - 1))*(0 < w)*(w < 1)
pkuma <- function(w, a, b)
         (1 - (1 - w^a)^b)*(0 < w)*(w < 1) + 1*(w >= 1)
qkuma <- function(u, a, b)
         ((1 - (1 - u)^(1/b))^(1/a))*(0 < u)*(u < 1)

t2 <- function(ab) (qkuma(1/2, ab[1], ab[2]) - t)^2 +
      (pkuma(v, ab[1], ab[2]) - pkuma(r, ab[1], ab[2]) - z)^2
(t2NLM <- nlm(h2, c(1, 1)))
a <- t2NLM$estimate[1]
b <- t2NLM$estimate[2]
pkuma(t, a, b)
pkuma(v, a, b) - pkuma(r, a, b)

# Muestra aleatoria

n <- 100  # probar n = 100, 1000
theta <- 0.6
set.seed(1)  # analizar semillas 1 y 2
x <- rbinom(n, size = 1, prob = theta)

# Verosimilitud

L <- function(theta) (theta^(sum(x)))*(1 - theta)^(length(x) - sum(x))


# Densidad final con densidad inicial Kumaraswamy(a, b)

numerador <- function(theta) L(theta)*dkuma(theta, a, b)
(K <- integrate(numerador, 0, 1))
```

```
#Comprobando que integre 1

dpost2 <- function(theta) numerador(theta)/K\$value
integrate(dpost2, 0, 1)
```

A.2. Estimación puntual y de conjunto

```
# MARGINALES MODELO UNIFORME(a, b)

b0 <- -1
b1 <- 2
k0 <- 3

(aa <- log(2)/log(1 + k0/(b1 - b0)))

#Marginal A

prA <- function(a) (a < b0)*aa*((b1 - b0)^aa)/((b1 - a)^(aa + 1))

a <- seq(from=-15, to=-1, length=1000)

plot(a, prA(a), type = "l", col = "black", lwd = 3)

#Marginal B

prB <- function(b) (b > b1)*aa*((b1 - b0)^aa)/((b - b0)^(aa + 1))

b <- seq(from=2, to=15, length=1000)

plot(b, prB(b), type = "l", col = "black", lwd = 3)


# UNIFORME (a, b)
```

```
b0 <- -1
b1 <- 2
a1 <- 3


# Modelo inicial

(a2 <- log(2)/log(1 + a1/(b1 - b0)))
# Modelo bivariado:
h <- function(a, b, a2, b0, b1)
     (a < b0)*(b > b1)*a2*(a2 + 1)*((b1 - b0)^a2)/((b - a)^(a2 + 2))

# Graficar densidad conjunta inicial

f.biv.eval <- function (f.biv, x.min, x.max, y.min, y.max,
                        num.div = 30, th = 0, ph = 15){
    z <- matrix(0, nrow = (num.div + 1), ncol = (num.div + 1))
    x <- seq(from = x.min, to = x.max, length = (num.div + 1))
    y <- seq(from = y.min, to = y.max, length = (num.div + 1))
    for (i in 1:(num.div + 1)) {
        for (j in 1:(num.div + 1)) {
            z[i, j] <- f.biv(x[i], y[j])
        }
    }
    persp(x, y, z, theta = th, phi = ph, col="yellow", xlab = "a",
      ylab = "b", zlab = "h(a,b)")
    return(list(x, y, z))
}
f.biv <- function(a, b) h(a, b, a2, b0, b1)
hEval <- f.biv.eval(f.biv, -4, 0, 0, 5, 30)
```

A.3. Predicción

```
#GRAFICA PREDICTIVA A POSTERIORI UNIFORME(a, b)
```

```
b0 <- -1
b1 <- 2
k0 <- 3
n <- 10
x <- runif(n, -2, 3)
B0 <- min(b0, min(x))
B1 <- max(b1, max(x))
X <- (B0 + B1)/2
(aa <- log(2)/log(1 + k0/(b1 - b0)))
maxi <- function(c, d) (c + d + abs(c - d))/2
max3 <- function(c, d, e) maxi(maxi(c, d), e)
mini <- function(c, d) (c + d - abs(c - d))/2
min3 <- function(c, d, e) mini(mini(c, d), e)
pred <- function(y)
              (aa + n)*((B1 - B0)^(aa + n))/((aa + n + 2)*((max3(y,b1,max(x))
                - min3(y,b0,min(x)))^(aa + n + 1)))

pp <- 0.99

g <- function(z) integrate(pred, X, z)$value - pp/2
superior <- uniroot(g, c(X, X + 1000))$root
inferior <- X - (superior - X)
y <- seq(inferior, superior, length = 1000)
plot(y, pred(y), type = "l", lwd = 3)

#CONTRASTE DE HIPOTESIS. EJEMPLO CADENA DE MARKOV

library(cubature)

#Tomando la muestra 0,0,0,1,1,0,1,0,1,1

a1 <- 2
a2 <- 3
b1 <- 4
b2 <- 5
n00 <- 3
n01 <- 3
n10 <- 2
```

```
n11 <- 2


#x[1] es alfa y x[2] es beta

f <- function(x) dbeta(x[1], a1+n01, b1+n00)
     *dbeta(x[2], a2+n10, b2+n11)*(x[1]<=(1-x[2]))
adaptIntegrate(f, c(0,0), c(1,1))
```

Bibliografía

[1] P.J. Bickel, K.A. Doksum (1977) *Mathematical Statistics.* Prentice Hall

[2] G. Casella, R.L. Berger (2002) *Statistical Inference.* Duxbury

[3] D.R. Cox (2006) *Principles of Statistical Inference.* Cambridge University Press

[4] D.R. Cox, D.V. Hinkley (2000) *Theoretical Statistics.* Chapman & Hall / CRC

[5] J.K. Ghosh, M. Delampady, T. Samanta (2006) *An Introduction to Bayesian Analysis.* Springer

[6] H.S. Migon, D. Gamerman, F. Louzada (2015) *Statistical Inference An Integrated Approach.* CRC Press

[7] C.P. Robert (2007) *The Bayesian Choice.* Springer

[8] J. Shao (2003) *Mathematical Statistics.* Springer

[9] M.J. Crawley (2013) *The R Book* Wiley