



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
MAESTRÍA EN ESTADÍSTICA MATEMÁTICA

**Estimación robusta de los parámetros
de la distribución Beta**

Tesis presentada para optar al Título de Magister de la Universidad de
Buenos Aires en Estadística Matemática.

Tesis presentada por: Roberto Carlos Duarte Rivera

Directora de tesis: Dra. Marina Silvia Valdora.

Fecha de presentación del ejemplar:

Fecha de defensa: 21 de Junio del 2023.

Buenos Aires, Argentina 2023



Roberto C. Duarte

Índice general

Índice de figuras	III
Índice de tablas	IV
Agradecimientos	V
Resumen	VI
Abstract	VII
1. Introducción	1
2. Distribución Beta	4
2.1. Introducción	4
2.2. Las funciones Gamma y Beta	4
2.3. La Distribución Beta	6
2.3.1. Momentos y transformaciones de la distribución Beta	10
2.4. Estimadores clásicos	12
2.4.1. Método de los momentos	13
2.4.2. Estimador de máxima verosimilitud	14
3. Estimación Robusta	20
3.1. Introducción	20
3.2. M estimador	21
3.2.1. Propiedades asintóticas del M estimador	22
3.3. Robustez	25
3.3.1. Función de influencia	26
3.3.2. Función de influencia del estimador de máxima verosimilitud para una distribución Beta	27
4. Algunos estimadores robustos aplicados a la estimación de parámetros de una distribución Beta	31
4.1. Introducción	31
4.2. MI Estimador	31
4.2.1. Propiedades asintóticas del MI estimador	33
4.2.2. MI estimador para la distribución Beta	34
4.2.3. Función de influencia del MI estimador para la distribución Beta	35
4.3. Estimador de divergencia de potencia mínima	37
4.3.1. Distribución asintótica del MDP estimador	39
4.3.2. MDP estimador para la distribución Beta	39
4.3.3. Función de influencia del MDP estimador para la distribución Beta	42
5. Resultados de los experimentos y aplicaciones	45
5.1. Eficiencia relativa	45
5.2. Error cuadrático medio	47
5.3. Cubrimiento empírico	49
5.4. Aplicaciones	53
5.4.1. Datos estrés	53
5.4.2. Índices de diversidad	55
A. Demostraciones del capítulo 4	58
Bibliografía	63

Índice de figuras

1.1.	Comportamiento de los estimadores de máxima verosimilitud y momentos ante la presencia de valores atípicos.	2
2.1.	Función poligamma para $m = 0, 1, 2$	5
2.2.	Distribución Beta con $a \neq b$, $a > 1$ y $b > 1$	7
2.3.	Distribución Beta con $a = b \geq 1$	8
2.4.	Distribución Beta con $a \neq b$, $a > 1$ y $b \leq 1$	8
2.5.	Distribución Beta con $a \neq b$, $a < 1$ y $b < 1$	9
2.6.	Distribución Beta con $a = b < 1$	10
3.1.	Funciones de densidad(izquierda) e influencia(derecha) para $a = b = 7$	29
3.2.	Funciones de densidad(izquierda) e influencia(derecha) para $a = 2$, $b = 6$	29
3.3.	Funciones de densidad(izquierda) e influencia(derecha) para $a = 7$, $b = 3$	30
4.1.	Función de influencia del MI estimador cuándo $a = b$	36
4.2.	Función de influencia del MI estimador cuándo $a > b$	36
4.3.	Función de influencia del MI estimador cuándo $a < b$	37
4.4.	Función de influencia para el estimador de divergencia de potencia mínima para una distribución Beta con $a = b = 7$ y diferentes valores de α	43
4.5.	Función de influencia para el estimador de divergencia de potencia mínima para una distribución Beta con $a = 2$, $b = 6$ y diferentes valores de α	43
4.6.	Función de influencia para el estimador de divergencia de potencia mínima para una distribución Beta con $a = 7$, $b = 3$ y diferentes valores de α	44
5.1.	Funciones de densidad consideradas para los experimentos.	47
5.2.	Error cuadrático medio de los estimadores para $a = b = 7$ para diferentes porcentajes de contaminación.	48
5.3.	Error cuadrático medio de los estimadores para $a = 2$ $b = 6$ para diferentes porcentajes de contaminación.	48
5.4.	Gráficos de dispersión para el grupo 1: mujeres con puntuación de ansiedad > 0.01 y grupo 2: mujeres con puntuación de ansiedad ≤ 0.01	53
5.5.	Funciones de densidad graficadas a partir de los parámetros estimados.	54
5.6.	Función de densidad del índice de Pielou de la muestra.	56
5.7.	Funciones de densidad graficadas a partir de los parámetros estimados.	56

Índice de tablas

5.1.	Eficiencia relativa del MI estimador y el estimador de divergencia de potencia mínima respecto al estimador de máxima verosimilitud para $a = b = 7$.	46
5.2.	Eficiencia relativa del MI estimador y el estimador de divergencia de potencia mínima respecto al estimador de máxima verosimilitud para $a = 2, b = 6$.	46
5.3.	Cubrimiento empírico del 95 % para los estimadores bajo contaminación para $a = b = 7$.	51
5.4.	Cubrimiento empírico del 95 % para los estimadores bajo contaminación para $a = 2, b = 6$.	51
5.5.	Estimaciones puntuales e intervalos de confianza de los parámetros de la distribución para el grupo de mujeres con ansiedad mayor a 0.01.	54
5.6.	Probabilidades estimadas a partir de la distribución generada por los estimadores robustos y de máxima verosimilitud.	57

Agradecimientos

A mi madre por siempre apoyarme en todas las decisiones que he tomado en mi vida, sin su apoyo no podría haber logrado todo esto.

A la doctora Marina Valdora por su paciencia y confianza en la elaboración de este trabajo, su orientación me permitió aprender mucho sobre el tema y a motivarme a seguir investigando.

A mis amigos Jesús y Maylin por haber sido un ejemplo a seguir en mis estudios de licenciatura y brindarme su amistad todos estos años.

Por último agradecer a la Universidad de Buenos Aires por brindar educación pública de calidad. Agradezco a cada docente por sus enseñanzas sin las cuales no podría haber finalizado esta tesis.

Roberto C. Duarte

ESTIMACIÓN ROBUSTA DE LOS PARÁMETROS DE LA DISTRIBUCIÓN BETA

Resumen

En la práctica es común encontrarnos con datos en forma de proporciones, tasas o probabilidades, este tipo de datos puede ser modelados mediante una distribución *Beta*; debido a la gran cantidad de aplicaciones que tiene esta distribución es importante estudiar la estimación de sus parámetros. Algunas veces los datos pueden estar contaminados por valores que se encuentran muy alejados de la mayoría, ya sea por algún error en la recolección de la información o por algún fenómeno particular que provoque este comportamiento en los datos, a este tipo de datos se le conocen como valores atípicos. La teoría de estimación clásica propone estimadores como el de momentos o máxima verosimilitud que ante la presencia de estos valores pueden verse muy afectados y proporcionar estimaciones erróneas de los parámetros. Esto puede llevar a realizar una mala toma de decisiones si las muestras están contaminadas y se utilizan este tipo de estimadores.

Por esto surge la necesidad de estudiar estimadores que puedan reducir el efecto de valores atípicos en las estimaciones. Los estimadores que tienen esta propiedad se conocen como estimadores robustos. En esta tesis estudiamos algunos estimadores robustos aplicados a una distribución *Beta* con el fin de realizar una comparación sistemática con el estimador de máxima verosimilitud. Estas comparaciones se realizan a través de medidas como la eficiencia relativa, el error cuadrático medio y el cubrimiento empírico de los intervalos de confianza basados en los estimadores. Al final se muestra una aplicación con datos reales.

Palabras clave: Robustez, M estimador, Transformación integral de probabilidad, Divergencia

ROBUST ESTIMATION OF THE BETA DISTRIBUTION PARAMETERS

Abstract

In practice, it is common to find data in the form of proportions, rates or probabilities; this type of data can be modeled using a *Beta* distribution; Due to the large number of applications that this distribution has, it is important to study the estimation of its parameters. Sometimes the data can be contaminated by values that are very far from the majority, either due to some error in the collection of the information or due to some particular phenomenon that causes this behavior in the data; this type of data is known as outliers. Classical estimation theory proposes estimators such as moments or maximum likelihood that, in the presence of these values, can be greatly affected and provide erroneous estimates of the parameters. This can lead to poor decision making if the samples are contaminated and these types of estimator are used.

This is why the need arises to study estimators that can reduce the effect of outliers on the estimates. Estimators that have this property are known as robust estimators. In this thesis, we study some robust estimators applied to a *Beta* distribution in order to make a systematic comparison with the maximum likelihood estimator. These comparisons are made through measures such as relative efficiency, the mean square error, and the empirical coverage of the confidence intervals based on the estimators. In the end, an application with real data is shown.

Keywords: Robustness, M estimator, Integral probability transformation, Divergence

Capítulo 1

Introducción

La distribución *Beta* es una familia de distribuciones continuas de probabilidad definidas en el intervalo $(0,1)$ parametrizada por dos términos $a > 0$ y $b > 0$ que definen la forma de la distribución. En la práctica es común encontrar datos en forma de proporciones, porcentajes o probabilidades y, debido al rango de valores que toma, la distribución *Beta* es apropiada para modelar este tipo de datos. En la estadística bayesiana, la distribución *Beta* es utilizada como distribución a priori cuando las observaciones siguen una distribución *Binomial*; ver Gelman et al. (2004). En la teoría de estimación clásica se estudia el estimador de máxima verosimilitud, que, asumiendo que los datos siguen la distribución teórica, nos proporciona un estimador con ciertas propiedades deseables como ser consistencia, normalidad asintótica y eficiencia. Sin embargo, en la práctica los modelos que se asumen se cumplen solo aproximadamente, lo que, en ciertos casos, puede causar que las estimaciones no sean fiables. Si los datos contienen observaciones muy alejadas de la mayoría (valores atípicos) esto puede hacer que los estimadores clásicos se vean muy afectados.

Para ilustrar esto, se generó una muestra de tamaño 100 a partir de una distribución *Beta* con parámetros $a = 2$ y $b = 9$. En este caso tenemos una distribución asimétrica hacia la derecha, lo que significa que la mayor densidad se acumula hacia la izquierda y obtener una observación menor a 0.6 tiene una probabilidad aproximada de 0.998, por lo que valores mayores a 0.6 son poco probables. Se elige una contaminación de $x_0 = 0.8$ para reemplazar el 3% de las observaciones simuladas.

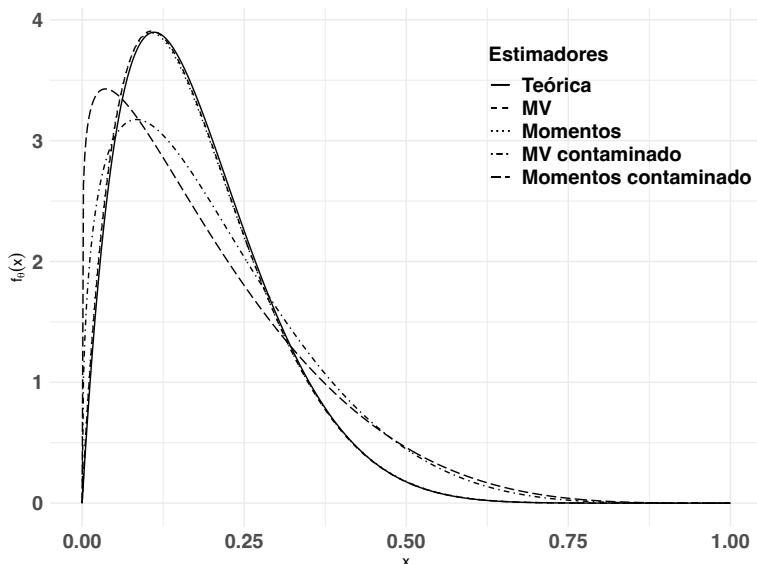


Figura 1.1: Comportamiento de los estimadores de máxima verosimilitud y momentos ante la presencia de valores atípicos.

Este ejemplo fue elegido de forma tal que la estimación por el método de máxima verosimilitud y el método de los momentos fuera buena cuando no hay valores atípicos y poder ver como cambia cuando se reemplaza una pequeña proporción de las observaciones por valores atípicos. Por este motivo surge la necesidad de estudiar estimadores que no se vean tan afectados ante la presencia de estos valores atípicos; los estimadores que tiene esta propiedad se conocen como estimadores robustos.

En la literatura de la estadística robusta diversos autores han propuesto modelos para la estimación de parámetros de una distribución. Thieler et al. (2016) implementaron un método para ajustar de forma robusta una distribución *Beta* utilizando la distancia de Cramér-von-Mises para el cálculo de periodogramas para curvas de luz. Ghosh (2019) propuso un estimador robusto basado en la mínima divergencia de potencia de densidad. Este método fue estudiado de forma general por Basu et al. (1998). En Valdora and Yohai (2020) se hace una propuesta general para estimar robustamente los parámetros de una distribución univariada y se implementa para las distribuciones *Binomial negativa* y *Poisson*.

En esta tesis haremos una comparación sistemática de algunos estimadores robustos existentes en la literatura. Adaptaremos los estimadores basados en la transformación integral de probabilidad propuestos por Valdora and Yohai (2020) junto con los propuestos por Ghosh (2019) para la estimación de parámetros de una distribución *Beta*.

La tesis se organiza de la siguiente forma: en el capítulo 2 definimos los conceptos relacionados con la distribución *Beta* y enunciamos algunas propiedades y teoremas que serán de nuestro interés, también se repasan las técnicas de estimación de parámetros utilizadas en la estadística clásica. En el capítulo 3 estudiamos los conceptos de estimación robusta y definimos una clase de estimadores conocidos como M estimadores y sus propiedades que hacen que estos sean estimadores robustos, al final del capítulo analizamos la falta de robustez del estimador de máxima verosimilitud a través de su función de influencia. En el capítulo 4 estudiamos los estimadores propuestos por Valdora and Yohai (2020) y Ghosh (2019), sus propiedades de robustez y asintóticas en general y luego adaptamos estos para la distribución *Beta*, deduciendo sus funciones de estimación y la forma de su matriz de varianza asintótica. En el capítulo 5 se muestran los resultados de las simulaciones realizadas para hacer la comparación de los estimadores a través de la eficiencia relativa, el error cuadrático medio y el cubrimiento empírico, al final del capítulo se muestran algunas aplicaciones con datos reales.

Capítulo 2

Distribución Beta

2.1. Introducción

En este capítulo iniciaremos estudiando algunas funciones relacionadas con la distribución *Beta*, así como sus propiedades que nos serán de utilidad en las siguientes secciones y capítulos de esta tesis. Después daremos la definición de la distribución *Beta* y algunos teoremas relacionados con sus momentos y transformaciones. Luego terminamos el capítulo estudiando los estimadores de parámetros más utilizados en la teoría de estadística clásica y sus propiedades asintóticas y veremos el problema de sensibilidad a valores atípicos que tienen estos estimadores.

2.2. Las funciones Gamma y Beta

La función de densidad de probabilidad de una variable aleatoria que sigue una distribución Beta involucra las funciones Gamma y Beta, por lo que vamos a dar las definiciones de estas funciones así como las propiedades de las mismas que pueden ser de nuestro interés más adelante.

Definición 2.1 (Función Gamma). *La función Gamma $\Gamma(x) : (0, +\infty) \rightarrow \mathbb{R}$ se define como la integral impropia*

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

La función Gamma tiene la propiedad de poder escribirse de forma recursiva, esto puede demostrarse utilizando integración por partes.

Proposición 2.1. *Para todo $x > 0$, $\Gamma(x + 1) = x\Gamma(x)$.*

En secciones posteriores nos vamos a encontrar con la necesidad de conocer la derivada logarítmica de la función Gamma, esta función se puede generalizar para la derivada parcial de orden m.

Definición 2.2 (Función poligamma). La función poligamma de orden m se define como la $(m+1)$ -ésima derivada logarítmica de la función Gamma.

$$\psi^{(m)}(x) = \frac{d^{m+1}}{dx^{m+1}} \log \Gamma(x).$$

Los casos especiales para $m = 0, 1$ y 2 se conocen como funciones digamma, trigamma y tetragamma, respectivamente. En la Figura 2.1 podemos ver que las funciones poligamma para $m = 0, 1, 2$ son funciones continuas y monótonas para $x > 0$.

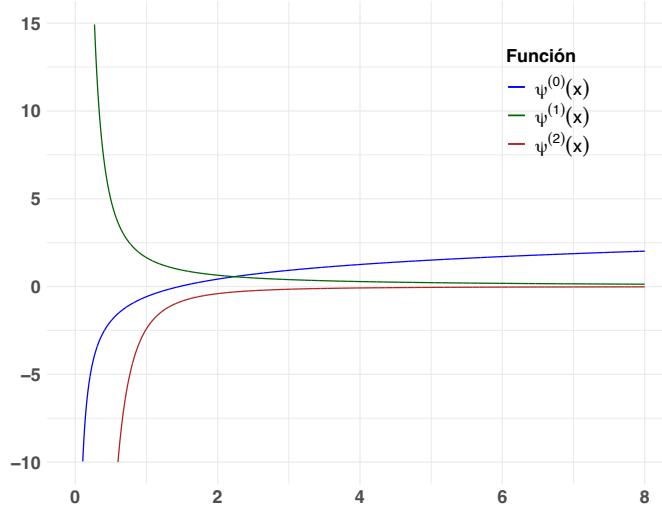


Figura 2.1: Función poligamma para $m = 0, 1, 2$.

Definición 2.3 (Función Beta). La función Beta, $\mathcal{B}(x, y) : (0, +\infty) \times (0, +\infty) \rightarrow \mathbb{R}$ se define como la integral definida

$$\mathcal{B}(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt.$$

Las funciones Gamma y Beta están relacionadas a través de la siguiente identidad:

$$\mathcal{B}(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

La demostración de este resultado puede verse en las páginas 18-19 de Artin (1964). De esta relación podemos encontrar las derivadas logarítmicas de orden m de la función Beta respecto a sus variables x y y , las cuales serán también de nuestro interés en cálculos posteriores.

Primero consideremos el logaritmo de la función Beta

$$\log \mathcal{B}(x, y) = \log \Gamma(x) + \log \Gamma(y) - \log \Gamma(x + y).$$

Luego sus primeras derivadas parciales respecto a x y y respectivamente son:

$$\begin{aligned} \frac{\partial}{\partial x} \log \mathcal{B}(x, y) &= \frac{\partial}{\partial x} \log \Gamma(x) - \frac{\partial}{\partial x} \log \Gamma(x + y) \\ &= \psi^{(0)}(x) - \psi^{(0)}(x + y). \end{aligned} \quad (2.1)$$

$$\begin{aligned} \frac{\partial}{\partial y} \log \mathcal{B}(x, y) &= \frac{\partial}{\partial y} \log \Gamma(y) - \frac{\partial}{\partial y} \log \Gamma(x + y) \\ &= \psi^{(0)}(y) - \psi^{(0)}(x + y). \end{aligned} \quad (2.2)$$

y sus segundas derivadas parciales respecto a x y y son:

$$\frac{\partial^2}{\partial x^2} \log \mathcal{B}(x, y) = \psi^{(1)}(x) - \psi^{(1)}(x + y). \quad (2.3)$$

$$\frac{\partial^2}{\partial y^2} \log \mathcal{B}(x, y) = \psi^{(1)}(y) - \psi^{(1)}(x + y). \quad (2.4)$$

$$\begin{aligned} \frac{\partial^2}{\partial x \partial y} \log \mathcal{B}(x, y) &= \frac{\partial^2}{\partial y \partial x} \log \mathcal{B}(x, y) \\ &= -\psi^{(1)}(x + y). \end{aligned} \quad (2.5)$$

2.3. La Distribución Beta

Definición 2.4. Decimos que una variable aleatoria X tiene una distribución de probabilidad Beta con parámetros $a > 0$ y $b > 0$ si y solo si su función de densidad es

$$f_X(x) = \begin{cases} \frac{1}{\mathcal{B}(a, b)} x^{a-1} (1-x)^{b-1} & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso.} \end{cases} \quad (2.6)$$

Teorema 2.1. Si X es una variable aleatoria con distribución Beta con $a > 0$ y $b > 0$, entonces:

$$E[X] = \frac{a}{a+b} \quad V[X] = \frac{ab}{(a+b)^2(a+b+1)}.$$

Los parámetros a y b controlan la forma de la distribución. Si los valores de a y b son intercambiados la forma de la distribución se refleja alrededor de $1/2$. A continuación se detallan los casos y se muestran algunos ejemplos.

Caso $a \neq b$, $a > 1$ y $b > 1$

La distribución es unimodal y asimétrica. La distribución tiene asimetría hacia la derecha cuando b es mucho más grande que a , pero la distribución se vuelve menos asimétrica y la moda se acerca a $1/2$ a medida que a y b se acercan entre sí. La Figura 2.2 ilustra los casos antes mencionados.

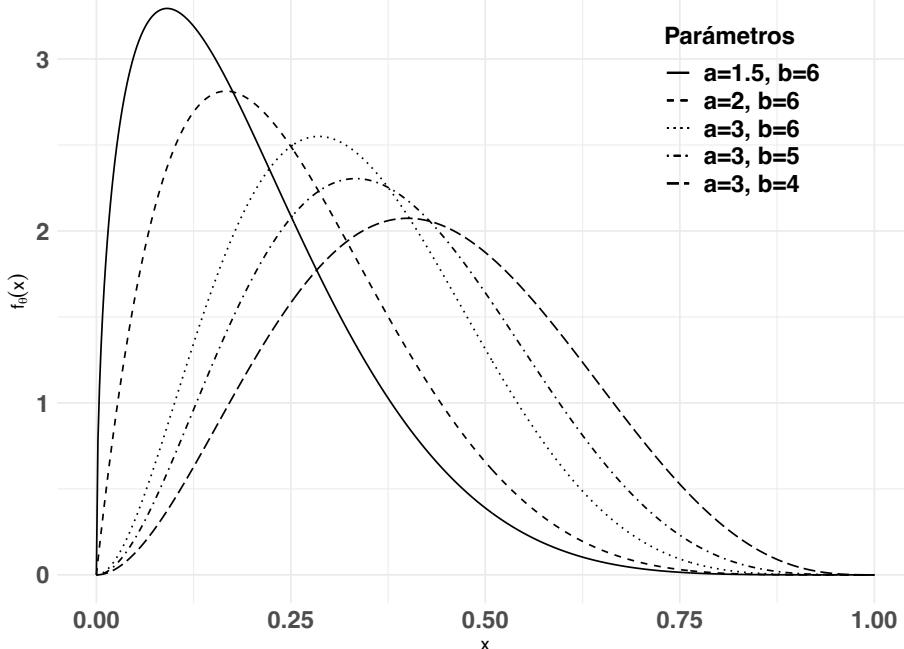
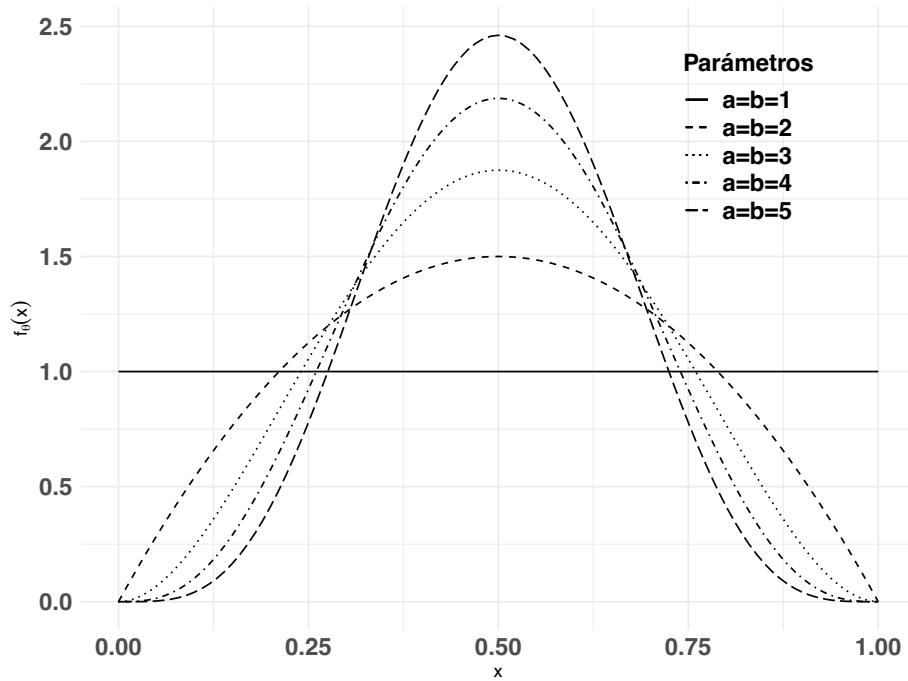


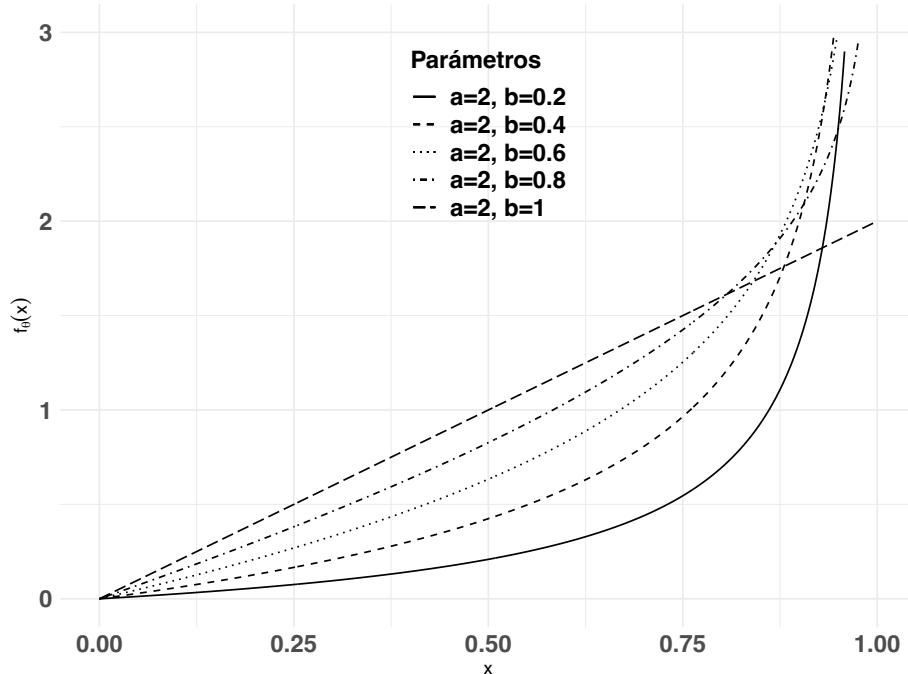
Figura 2.2: Distribución Beta con $a \neq b$, $a > 1$ y $b > 1$

Caso $a = b \geq 1$

La distribución es unimodal y simétrica alrededor de 0.5 , en el caso de $a = b = 1$ equivale a una distribución uniforme($0,1$). La distribución se vuelve más espigada a medida que a y b se hacen más grandes como se muestra en la Figura 2.3.

Figura 2.3: Distribución Beta con $a = b \geq 1$.**Caso $a \neq b$, $a > 1$ y $b \leq 1$**

La distribución es estrictamente creciente con forma de J. La distribución se vuelve más curva a medida que b se vuelve más pequeño. En la Figura 2.4 se pueden observar estos casos.

Figura 2.4: Distribución Beta con $a \neq b$, $a > 1$ y $b \leq 1$

Caso $a \neq b$, $a < 1$ y $b < 1$

La distribución tiene forma de U asimétrica. La distribución se vuelve menos sesgada cuando a y b se acercan entre ellos. En la Figura 2.5 se ilustran estos casos.

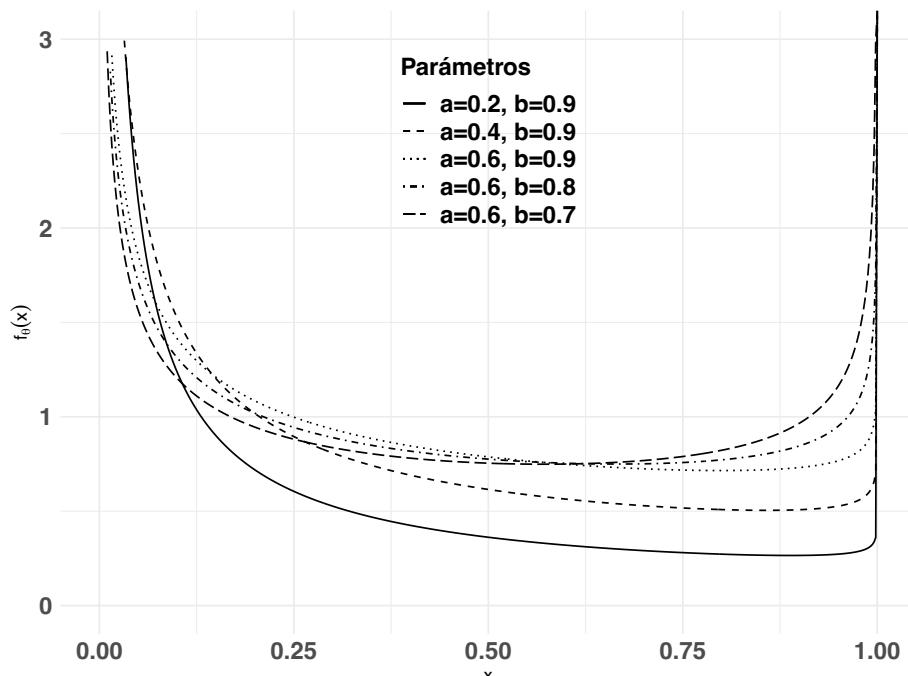


Figura 2.5: Distribución Beta con $a \neq b$, $a < 1$ y $b < 1$

Caso $a = b < 1$

La distribución tiene forma de U y simétrica alrededor de $1/2$. Cuando $a = b = 0.5$ se conoce como distribución arcoseno, utilizada en teoría de la comunicación. La distribución mueve la masa desde el centro hacia las colas a medida que a y b se hacen más pequeñas como se ilustra en la Figura 2.6.

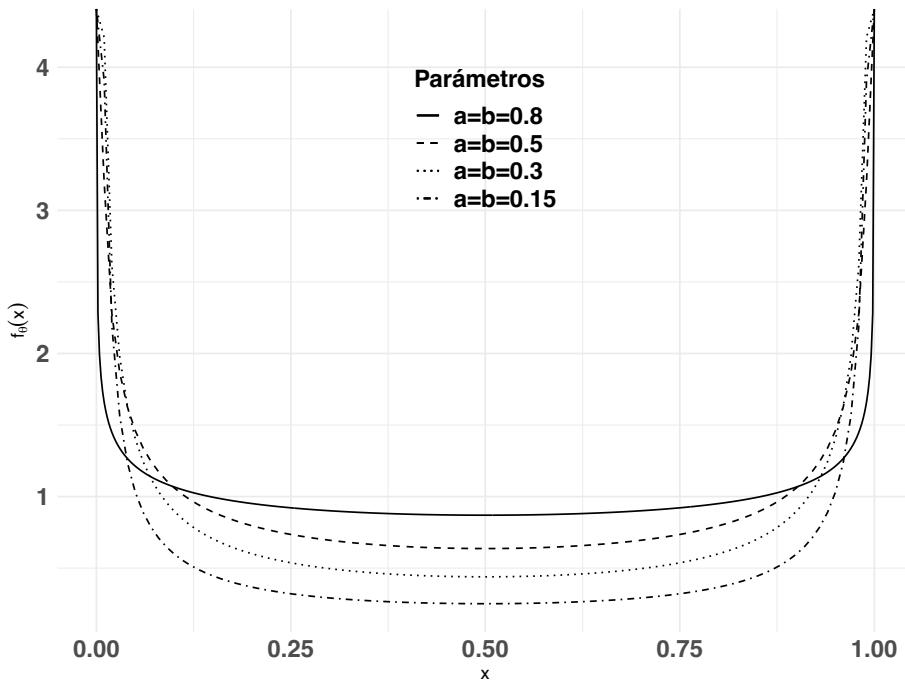


Figura 2.6: Distribución Beta con $a = b < 1$

Notemos que cuando $a < 1$ o $b < 1$ la función de densidad resulta no ser acotada y no garantiza que se cumplan ciertas condiciones de regularidad. En el resto de la tesis vamos a considerar solamente los casos cuando $a > 1$ y $b > 1$. En la siguientes secciones describiremos los métodos clásicos de estimación de parámetros aplicados a la distribución Beta así como las propiedades de los mismos.

2.3.1. Momentos y transformaciones de la distribución Beta

En la estimación de parámetros de la distribución Beta será necesario conocer algunos de sus momentos para deducir los estimadores, además, algunas veces las expresiones de las varianzas asintóticas de estos estimadores dependen de esperanzas de transformaciones de la variable aleatoria. A continuación se muestran resultados que serán de utilidad en secciones posteriores.

Definición 2.5. El k -ésimo momento poblacional de una variable aleatoria X se define como $E(X^k)$ y se denota como μ'_k .

Definición 2.6. El k -ésimo momento muestral de una variable aleatoria X tomada alrededor del origen se define como

$$m'_k = \frac{1}{n} \sum_{i=1}^n X_i^k.$$

Para una variable aleatoria con distribución Beta podemos encontrar una expresión para μ'_k en función de k y sus parámetros a y b .

Teorema 2.2. Sea X una variable aleatoria con distribución $Beta(a,b)$, entonces,

$$E(X^k) = \frac{\Gamma(a+b)\Gamma(a+k)}{\Gamma(a)\Gamma(a+b+k)}.$$

Con el resultado anterior y las propiedades de la función gamma es fácil ver que:

$$E(X) = \frac{a}{a+b} \tag{2.7}$$

$$E(X^2) = \frac{a(a+1)}{(a+b)(a+b+1)}. \tag{2.8}$$

Otras expresiones que nos serán de utilidad son las esperanzas, varianzas y covarianzas de transformaciones logarítmicas de una variable aleatoria con distribución Beta. El siguiente teorema resume las que serán de nuestro interés.

Teorema 2.3. Sea X una variable aleatoria con distribución $Beta(a,b)$ con $a>1$ y $b>1$, entonces:

$$1. \quad E[\log(X)] = \psi^{(0)}(a) - \psi^{(0)}(a+b).$$

$$2. \quad E[\log(1-X)] = \psi^{(0)}(b) - \psi^{(0)}(a+b).$$

$$3. \quad E[\log^2(X)] = [\psi^{(0)}(a) - \psi^{(0)}(a+b)]^2 + \psi^{(1)}(a) - \psi^{(1)}(a+b).$$

$$4. \quad E[\log^2(1-X)] = [\psi^{(0)}(b) - \psi^{(0)}(a+b)]^2 + \psi^{(1)}(b) - \psi^{(1)}(a+b).$$

$$5. \quad E[\log(X)\log(1-X)] = [\psi^{(0)}(a) - \psi^{(0)}(a+b)][\psi^{(0)}(b) - \psi^{(0)}(a+b)] - \psi^{(1)}(a+b).$$

$$6. \quad \text{Var} [\log(X)] = \psi^{(1)}(a) - \psi^{(1)}(a+b).$$

$$7. \quad \text{Var} [\log(1-X)] = \psi^{(1)}(b) - \psi^{(1)}(a+b).$$

$$8. \quad \text{Cov} [\log(X), \log(1-X)] = -\psi^{(1)}(a+b).$$

Demostración. Se hace la demostración del ítem 1; los ítems del 2 al 5 se pueden probar de forma similar.

$$\begin{aligned} E[\log(X)] &= \int_0^1 \log(x) \frac{x^{a-1}(1-x)^{b-1}}{\mathcal{B}(a,b)} dx \\ &= \frac{1}{\mathcal{B}(a,b)} \int_0^1 \frac{\partial}{\partial a} [x^{a-1}(1-x)^{b-1}] dx \\ &= \frac{1}{\mathcal{B}(a,b)} \frac{\partial}{\partial a} \int_0^1 x^{a-1}(1-x)^{b-1} dx \\ &= \frac{1}{\mathcal{B}(a,b)} \frac{\partial}{\partial a} \mathcal{B}(a,b) \\ &= \frac{\partial}{\partial a} \log \mathcal{B}(a,b) \\ &= \frac{\partial}{\partial a} \log \Gamma(a) - \frac{\partial}{\partial a} \log \Gamma(a+b) \\ &= \psi^{(0)}(a) - \psi^{(0)}(a+b). \end{aligned}$$

El intercambio de la derivada por la integral se puede justificar por el teorema de integración bajo el signo integral, ver Teorema 10.39 Apostol (1981). Los ítems del 6 al 8 se pueden probar aplicando la definición $\text{Cov}(Y, Z) = E(YZ) - E(Y)E(Z)$.

□

Los resultados mostrados en esta sección serán utilizados en las siguientes secciones para demostrar algunos teoremas y deducir algunas fórmulas. En la siguiente sección se muestran los métodos clásicos de estimación de parámetros aplicados a una distribución Beta.

2.4. Estimadores clásicos

En la teoría clásica de la estadística los métodos más utilizados para la estimación de parámetros son el método de los momentos y el método de máxima verosimilitud. El método de los momentos es un método cuya forma de calcular sus estimadores es más

sencilla y se basa en igualar los momentos muestrales y poblacionales de una variable aleatoria. Sin embargo, sus propiedades no son tan buenas como las que ofrece el estimador deducido por el método de máxima verosimilitud. A continuación se definen ambos métodos para deducir estimadores y se mencionan sus propiedades.

2.4.1. Método de los momentos

Es uno de los métodos más antiguos para obtener estimadores puntuales y en el caso de la distribución Beta podemos encontrar expresiones cerradas de la forma de los estimadores que proporciona. La idea del método de los momentos es que los momentos muestrales deben dar buenas estimaciones de los momentos poblacionales correspondientes, es decir, m'_k debe ser un buen estimador de μ'_k para $k = 1, \dots$. Como los momentos poblacionales son funciones de los parámetros poblacionales podemos igualar los correspondientes momentos poblacionales y muestrales, y luego despejar los estimadores deseados.

Como la distribución Beta tiene dos parámetros, entonces necesitamos dos ecuaciones que resultan de considerar los primeros dos momentos, luego el estimador de momentos es la raíz del siguiente sistema de ecuaciones

$$\begin{cases} \frac{1}{n} \sum X_i &= \frac{a}{a+b} \\ \frac{1}{n} \sum X_i^2 &= \frac{a(a+1)}{(a+b)(a+b+1)} \end{cases} . \quad (2.9)$$

Resolviendo el sistema anterior para los parámetros a y b podemos encontrar expresiones para los estimadores de momentos:

$$\hat{a}_{mom} = \bar{X} \left[\frac{\bar{X}(1-\bar{X})}{S^2} - 1 \right] \quad \hat{b}_{mom} = (1-\bar{X}) \left[\frac{\bar{X}(1-\bar{X})}{S^2} - 1 \right]. \quad (2.10)$$

Siempre que $S^2 < \bar{X}(1-\bar{X})$, donde \bar{X} y S^2 son la media y varianza muestral respectivamente.

Los estimadores deducidos por el método de los momentos son consistentes y asintóticamente normales, aunque, en general, no son eficientes ni óptimos asintóticamente normales, es decir, que entre los estimadores asintóticamente normales no es el de varianza mínima; ver Alexander and Mood (1969). A pesar que no es el mejor estimador en el sentido de

eficiencia, este puede usarse como una aproximación inicial para encontrar computacionalmente otros estimadores. Otra característica de este estimador es que depende de la media y varianza muestral que son estadísticas poco robustas, ver la sección 1.2 de Maronna et al. (2019) por lo que generan un estimador que puede ser muy afectado ante la presencia de valores atípicos.

2.4.2. Estimador de máxima verosimilitud

El método de máxima verosimilitud es un método de estimación paramétrica que consiste en encontrar el conjunto de parámetros que es más verosímil dados los datos observados. Si X_1, \dots, X_n es una muestra aleatoria independiente e idénticamente distribuida (iid) donde cada X_i tiene función de densidad de probabilidad $f_{\theta}(X)$, entonces el estimador de máxima verosimilitud se define como:

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{MV} &= \arg \max_{\boldsymbol{\theta}} \prod_{i=1}^n f_{\boldsymbol{\theta}}(X_i) \\ &= \arg \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta} | \mathbf{X}).\end{aligned}\tag{2.11}$$

Siempre que se cumplan ciertas condiciones de regularidad, como que el soporte de la variable aleatoria X , no dependa de $\boldsymbol{\theta}$ y que $\boldsymbol{\theta}$ identifica la distribución de X , ver capítulo 6 de Hogg et al. (2019).

Es común definir el estimador de máxima verosimilitud tomando el logaritmo de la función objetivo anterior

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{MV} &= \arg \max_{\boldsymbol{\theta}} \sum_{i=1}^n \log f_{\boldsymbol{\theta}}(X_i) \\ &= \arg \max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta} | \mathbf{X}).\end{aligned}\tag{2.12}$$

La función objetivo en (2.11) se conoce como función de verosimilitud y a (2.12) como función de log verosimilitud. En problemas regulares, el estimador de máxima verosimilitud puede hallarse igualando a 0 las primeras derivadas de la función de log verosimilitud respecto de $\boldsymbol{\theta}$

$$\sum_{i=1}^n \mathbf{u}_{\boldsymbol{\theta}}(X_i) = \mathbf{0}.$$

dónde a la expresión $\mathbf{u}_\theta(X) = \frac{\partial}{\partial \theta} \log f_\theta(X)$ se la conoce como *score*.

Si $X \sim Beta(a, b)$ y llamemos $\theta = (a, b)^T$ podemos escribir la log verosimilitud de una observación como:

$$\log f_\theta(X) = \log \Gamma(a + b) - \log \Gamma(a) - \log \Gamma(b) + (a - 1) \log(X) + (b - 1) \log(1 - X).$$

Luego

$$\frac{\partial}{\partial a} \log f_\theta(X) = \psi^{(0)}(a + b) - \psi^{(0)}(a) + \log(X).$$

$$\frac{\partial}{\partial b} \log f_\theta(X) = \psi^{(0)}(a + b) - \psi^{(0)}(b) + \log(1 - X).$$

Podemos escribir el score como

$$\mathbf{u}_\theta(X) = \left(\frac{\partial}{\partial a} \log f_\theta(X), \frac{\partial}{\partial b} \log f_\theta(X) \right)^T. \quad (2.13)$$

Luego si X_1, \dots, X_n es una muestra aleatoria simple con $X_i \sim Beta(a, b)$ el estimador de máxima verosimilitud para $\theta = (a, b)^T$ es la raíz del sistema de ecuaciones:

$$\begin{cases} \psi^{(0)}(a) - \psi^{(0)}(a + b) &= \frac{1}{n} \sum \log(X_i) \\ \psi^{(0)}(b) - \psi^{(0)}(a + b) &= \frac{1}{n} \sum \log(1 - X_i). \end{cases} \quad (2.14)$$

Es claro que no podemos encontrar expresiones cerradas para el estimador de máxima verosimilitud por lo que la raíz de (2.14) debe ser aproximada mediante métodos numéricos.

Un elemento importante en la teoría de estimación de máxima verosimilitud es la información de Fisher. Esta es una forma de medir la cantidad de información que tiene una variable aleatoria sobre un parámetro desconocido θ de una distribución que modela. A continuación daremos su definición formal.

Definición 2.7 (Información de Fisher). *Sea X una variable aleatoria con función de densidad de probabilidad $f_\theta(X)$, donde $\theta = (\theta_1, \dots, \theta_p)$. La información de Fisher*

esperada de $\boldsymbol{\theta}$ para X se define como:

$$I(\boldsymbol{\theta}) = E \left[-\frac{\partial^2 \log f_{\boldsymbol{\theta}}(X)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right].$$

Existen muchas propiedades que pueden deducirse de la información de Fisher, una de las mas utilizadas es la aditividad de la información con respecto a observaciones independientes.

Lema 2.1. *Sea X_1, \dots, X_n una colección de variables aleatorias independientes con distribución $f_{\boldsymbol{\theta}_i}(X)$, $i = 1, \dots, n$. Sean I e I_i la información esperada proporcionada por (X_1, \dots, X_n) y X_i respectivamente, $i = 1, \dots, n$, entonces*

$$I(\boldsymbol{\theta}) = \sum_{i=1}^n I_i(\boldsymbol{\theta}).$$

Notemos que si particularmente X_1, \dots, X_n son independientes e identicamente distribuidas, entonces

$$I(\boldsymbol{\theta}) = nI_1(\boldsymbol{\theta}).$$

Es decir, que la información para una muestra iid es n veces la información de una variable, a partir de ahora vamos a llamar $I_{\boldsymbol{\theta}}$ a la información de Fisher para una observación.

Bajo ciertas condiciones de regularidad el estimador de máxima verosimilitud es un estimador consistente y asintóticamente normal. La distribución Beta pertenece a una familia de funciones de distribución conocida como familia exponencial y este resultado es de utilidad para probar estas condiciones de regularidad. Las siguientes definiciones son tomadas de van der Vaart (2000).

Definición 2.8 (Familia exponencial). *La familia de distribuciones con función de densidad $f_{\boldsymbol{\theta}}(x)$ pertenece a la familia exponencial p dimensional si $f_{\boldsymbol{\theta}}(X)$ se puede escribir como:*

$$f_{\boldsymbol{\theta}}(X) = d(\boldsymbol{\theta})h(X) \exp [Q(\boldsymbol{\theta})^T t(X)].$$

dónde h y $t = (t_1, \dots, t_p)^T$ son funciones conocidas del espacio muestral, d y $Q = (Q_1, \dots, Q_p)^T$ son funciones de los parámetros.

Definición 2.9 (Familia exponencial de rango completo). Decimos que la familia exponencial es de rango completo si ninguna combinación lineal $\sum_{j=1}^p \lambda_j t_j(X)$ es constante con probabilidad 1, esto es equivalente a que la matriz de covarianza de $t(X)$ es no singular.

Teorema 2.4. Sea X una variable aleatoria continua con distribución $Beta(a,b)$, entonces su función de densidad pertenece a la familia exponencial de rango completo.

Demostración. Si $X \sim Beta(a,b)$ entonces podemos escribir su función de densidad de probabilidad como:

$$\begin{aligned} f_{\theta}(X) &= \frac{1}{B(a,b)} x^{a-1} (1-X)^{b-1} \\ &= \frac{1}{B(a,b)} \exp [(a-1) \log(X) + (b-1) \log(1-X)]. \end{aligned}$$

Luego tenemos que

$$d(\boldsymbol{\theta}) = \frac{1}{B(a,b)}.$$

$$h(X) = 1.$$

$$Q(\boldsymbol{\theta})^T = (a-1, b-1).$$

$$t(X) = (\log(X), \log(1-X))^T.$$

Ahora para probar que es de rango completo consideremos la matriz de covarianza de $t(X)$.

$$\boldsymbol{\Sigma} = \begin{pmatrix} Var[\log(X)] & Cov[\log(X), \log(1-X)] \\ Cov[\log(X), \log(1-X)] & Var[\log(1-X)] \end{pmatrix}.$$

En el Teorema 2.3 se muestran las expresiones de los elementos de la matriz $\boldsymbol{\Sigma}$ en términos de la función poligamma.

$$\boldsymbol{\Sigma} = \begin{pmatrix} \psi^{(1)}(a) - \psi^{(1)}(a+b) & -\psi^{(1)}(a+b) \\ -\psi^{(1)}(a+b) & \psi^{(1)}(b) - \psi^{(1)}(a+b) \end{pmatrix}.$$

Luego

$$\det(\boldsymbol{\Sigma}) = \psi^{(1)}(a) \psi^{(1)}(b) - \psi^{(1)}(a+b) [\psi^{(1)}(a) + \psi^{(1)}(b)]. \quad (2.15)$$

La función $\frac{1}{\psi^{(1)}(x)}$ tiene la propiedad de ser super aditiva, ver Qi (2020), esto implica que

$$\begin{aligned} \frac{1}{\psi^{(1)}(a+b)} &> \frac{1}{\psi^{(1)}(a)} + \frac{1}{\psi^{(1)}(b)} \\ \psi^{(1)}(a)\psi^{(1)}(b) &> \psi^{(1)}(a+b)[\psi^{(1)}(a) + \psi^{(1)}(b)]. \end{aligned}$$

Con este último resultado tenemos que $\det(\Sigma) > 0$, lo que implica que la matriz Σ es definida positiva y por tanto Σ es no singular. \square

El siguiente teorema establece la distribución asintótica del estimador de máxima verosimilitud. Ver van der Vaart (2000).

Teorema 2.5. *Sea la familia de densidades $f_{\theta}(X)$ miembro de la familia exponencial y de rango completo. Sea el espacio de parámetros $\Theta \in \mathbb{R}^p$ abierto y sea $Q : \Theta \mapsto \mathbb{R}^p$ una función uno a uno y continuamente diferenciable en todo Θ con derivadas no singulares. Entonces las ecuaciones de verosimilitud tienen una única solución $\hat{\theta}_n$ con probabilidad 1 y $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N(0, I_{\theta}^{-1})$, para todo θ .*

Con el teorema anterior podemos deducir que el estimador de máxima verosimilitud dada una muestra aleatoria de una distribución *Beta* es asintóticamente normal, donde la matriz de información de Fisher tiene la forma:

$$\begin{aligned} I_{\theta} &= E \begin{pmatrix} -\frac{\partial^2}{\partial a^2} \log f_{\theta}(X) & -\frac{\partial^2}{\partial a \partial b} \log f_{\theta}(X) \\ -\frac{\partial^2}{\partial b \partial a} \log f_{\theta}(X) & -\frac{\partial^2}{\partial b^2} \log f_{\theta}(X) \end{pmatrix} \\ &= E \begin{pmatrix} \psi^{(1)}(a) - \psi^{(1)}(a+b) & -\psi^{(1)}(a+b) \\ -\psi^{(1)}(a+b) & \psi^{(1)}(b) - \psi^{(1)}(a+b) \end{pmatrix} \\ &= \begin{pmatrix} \psi^{(1)}(a) - \psi^{(1)}(a+b) & -\psi^{(1)}(a+b) \\ -\psi^{(1)}(a+b) & \psi^{(1)}(b) - \psi^{(1)}(a+b) \end{pmatrix}. \end{aligned}$$

El método de máxima verosimilitud nos proporciona estimadores, que bajo las condiciones previamente descritas, tiene buenas propiedades; sin embargo, puede hacerse inservible con la presencia de una pequeña proporción de valores atípicos en la muestra, ya que los resultados que proporcionan no dan buena información de la distribución de la que se

obtuvo la muestra. En el siguiente capítulo estudiaremos algunos conceptos de estadística robusta que nos permitirán analizar estimadores para los parámetros de la distribución Beta que no se vean tan afectados ante la presencia de valores atípicos como sucede con el estimador de máxima verosimilitud y de momentos.

Capítulo 3

Estimación Robusta

3.1. Introducción

Bajo el supuesto que los datos pueden modelarse mediante una familia paramétrica de distribuciones con parámetro θ desconocido existen diferentes procedimientos, donde éste puede ser estimado a partir de la información de la muestra; entre estos procedimientos, el más utilizado es el método de máxima verosimilitud, ya que, como vimos en el capítulo anterior, este proporciona propiedades que resultan atractivas al investigador, como ser consistencia, normalidad asintótica y varianza mínima.

En la práctica algunas veces las muestras se encuentran contaminadas con una pequeña fracción de datos que llamaremos valores atípicos o *outliers*. Es decir, que en realidad tendremos una mezcla donde la mayor parte de los datos pueden ajustarse a algún modelo, mientras una pequeña proporción de ellos pueden seguir una distribución arbitraria desconocida. La presencia de estos valores atípicos afecta la estimación de máxima verosimilitud de los parámetros del modelo propuesto y por este motivo surge la necesidad de estudiar estimadores que no se vean afectados ante una pequeña proporción de estos datos. Los estimadores que tienen esta propiedad se conocen como estimadores robustos. En esta sección definimos un tipo de estimador conocido como M estimador que es una generalización del estimador de máxima verosimilitud, que, bajo ciertas condiciones, estos estimadores tienen una distribución normal asintótica y propiedades de robustez que lo hace una buena alternativa para la estimación de parámetros ante la presencia de valores atípicos. Luego estudiaremos una medida de robustez llamada función de influencia. Este capítulo está basado en el capítulo 3 de Maronna et al. (2019).

3.2. M estimador

Suponga que X_1, \dots, X_n es una muestra aleatoria de una distribución paramétrica $F_\theta(X)$. Huber (1964) define un M estimador como la generalización de la estimación de máxima verosimilitud como

$$\hat{\theta}_n = \arg \min_{\theta} \sum_{i=1}^n \rho(x_i; \theta). \quad (3.1)$$

donde ρ es una función arbitraria y $\theta \in \mathbb{R}$. Notemos que si tomamos $\rho(x; \theta) = -\log f_\theta(x)$ tenemos el estimador de máxima verosimilitud. En muchos casos (3.1) puede escribirse como la raíz de la ecuación de estimación:

$$\sum_{i=1}^n \psi(x_i; \theta) = 0. \quad (3.2)$$

donde $\psi(x; \theta) = \frac{\partial}{\partial \theta} \rho(x; \theta)$. A la raíz de (3.2) se le conoce como Z estimador, aunque no siempre un M estimador puede escribirse como un Z estimador, vamos a utilizar el término M estimador para referirnos a ambos ya que los estimadores estudiados en esta tesis pueden ser escritos de ambas formas.

La función ρ o su derivada ψ se pueden elegir de manera que proporcione propiedades deseables del estimador, tanto si los datos siguen exactamente la distribución supuesta o una aproximación de esta. En el caso de parámetro multidimensional tenemos que $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ y el M estimador se puede escribir como solución de

$$\sum_{i=1}^n \boldsymbol{\Psi}(x_i, \boldsymbol{\theta}) = \mathbf{0}. \quad (3.3)$$

donde $\boldsymbol{\Psi} = (\psi_1, \dots, \psi_p)^T$.

En general una función ψ no tiene que ser necesariamente la derivada de una función ρ , sin embargo para garantizar la existencia y unicidad de las soluciones se debe hacer algunos supuestos en la función ψ . En la siguiente sección vamos a estudiar estas condiciones.

3.2.1. Propiedades asintóticas del M estimador

Para poder evaluar el desempeño de un M estimador y construir intervalos de confianza en las estimaciones es necesario conocer su distribución. No siempre es posible encontrar expresiones explícitas que nos permitan conocer la distribución del M estimador por lo que estudiaremos su comportamiento asintótico. A continuación veremos algunos resultados que nos permiten conocer propiedades de consistencia y distribución asintótica del M estimador.

Vamos a considerar ahora la existencia y unicidad de la solución de (3.2). Supongamos que $\theta \in \mathbb{R}$ pertenece a un intervalo (θ_1, θ_2) finito o infinito, el siguiente teorema establece las condiciones en la función ψ para la existencia y unicidad.

Teorema 3.1 (Teorema 10.1 Maronna et al. (2019)). *Asuma que para cada x , $\psi(x, \theta)$ es no creciente como función de θ y*

$$\lim_{\theta \rightarrow \theta_1} \psi(x, \theta) > 0 > \lim_{\theta \rightarrow \theta_2} \psi(x, \theta).$$

Sea

$$g(\theta) = \sum_{i=1}^n \psi(x_i, \theta).$$

entonces

1. *Existe al menos un punto $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ en el cual g cambia de signo, es decir*

$$g(\theta) \geq 0 \text{ para } \theta < \hat{\theta} \text{ y } g(\theta) \leq 0 \text{ para } \theta > \hat{\theta}$$

2. *El conjunto de estos puntos es un intervalo.*

3. *Si ψ es continua en θ , entonces $g(\hat{\theta}) = 0$.*

4. *Si ψ es decreciente, entonces $\hat{\theta}$ es única.*

La demostración de este teorema se puede encontrar en Maronna et al. (2019). La unicidad en el caso multidimensional se puede probar generalizando las condiciones de monotonicidad del Teorema 3.1.

Teorema 3.2. Asuma que para todo x y $\boldsymbol{\theta}$, $\Psi(x, \boldsymbol{\theta})$ es diferenciable y la matriz $\mathbf{D} = \mathbf{D}(x, \boldsymbol{\theta})$ con elementos $\partial \Psi_i / \partial \theta_j$ es definida negativa. Sea x_1, \dots, x_n definamos la función g como

$$g(\boldsymbol{\theta}) = \sum_{i=1}^n \Psi(x_i, \boldsymbol{\theta}).$$

Si existe una solución de $g(\boldsymbol{\theta}) = \mathbf{0}$, entonces esta solución es única.

La demostración de este Teorema se puede encontrar en Maronna et al. (2019). Ahora vamos a estudiar la normalidad asintótica del M estimador. Sea X_1, \dots, X_n iid con distribución $F_{\boldsymbol{\theta}}$, definamos

$$\widehat{\lambda}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \Psi(X_i, \boldsymbol{\theta}) \quad (3.4)$$

$$\lambda(\boldsymbol{\theta}) = E_{F_{\boldsymbol{\theta}}} \Psi(X_1, \boldsymbol{\theta}). \quad (3.5)$$

Sea $\widehat{\boldsymbol{\theta}}_n$ una solución de (3.3), si la solución $\boldsymbol{\theta}_F$ de $\lambda(\boldsymbol{\theta}) = \mathbf{0}$ es única, entonces si $n \rightarrow \infty$, $\widehat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_F$. La sección 6.2 de Ronchetti and Huber (2009) proporciona criterios generales para esta condición. Boos et al. (2013) propone el siguiente teorema sobre la normalidad asintótica del M estimador.

Teorema 3.3. Sea X_1, \dots, X_n una muestra iid con función de distribución $F_{\boldsymbol{\theta}}$, sea $\boldsymbol{\theta}_F$ la única solución de $\lambda(\boldsymbol{\theta}) = \mathbf{0}$ y Ψ' el jacobiano de Ψ . Asuma que:

1. $\Psi(x, \boldsymbol{\theta})$ y sus primeras dos derivadas parciales respecto a $\boldsymbol{\theta}$ existen para todo x en el soporte de $F_{\boldsymbol{\theta}}$ y para todo $\boldsymbol{\theta}$ en un entorno de $\boldsymbol{\theta}_F$.

2. Para cada $\boldsymbol{\theta}$ en un entorno de $\boldsymbol{\theta}_F$, existe una función $g(x)$ tal que para todo j, k y $l \in \{1, \dots, p\}$

$$\left| \frac{\partial^2}{\partial \theta_j \partial \theta_k} \psi_l(x, \boldsymbol{\theta}) \right| \leq g(x)$$

para todo x y $E(g(x)) < \infty$.

3. La matriz

$$\mathbf{A} = E [\Psi(X_1, \boldsymbol{\theta}_F) \Psi(X_1, \boldsymbol{\theta}_F)^T] \quad (3.6)$$

existe y es finita.

4. La matriz

$$\mathbf{B} = E[-\Psi'(X_1, \boldsymbol{\theta}_F)] \quad (3.7)$$

existe y es no singular.

Si $\widehat{\lambda}_n(\boldsymbol{\theta})$ es una función suave y $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_F$ entonces

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_F) \xrightarrow{\mathcal{D}} N_p [\mathbf{0}, \mathbf{B}^{-1} \mathbf{A} \mathbf{B}^{T-1}] .$$

Dada una muestra X_1, \dots, X_n las matrices \mathbf{A} y \mathbf{B} pueden estimarse a partir de la muestra con:

$$\mathbf{A}_n = \frac{1}{n} \sum_{i=1}^n \Psi(X_i, \widehat{\boldsymbol{\theta}}) \Psi(X_i, \widehat{\boldsymbol{\theta}})^T .$$

$$\mathbf{B}_n = \frac{1}{n} \sum_{i=1}^n [-\Psi'(X_i, \widehat{\boldsymbol{\theta}})] .$$

Estos resultados nos permiten evaluar la eficiencia relativa de un M estimador ante otro que usualmente es el estimador de máxima verosimilitud. Si $\widetilde{\boldsymbol{\theta}}_n$ es el estimador de máxima verosimilitud con matriz de covarianza asintótica \mathbf{V}_0 y sea $\widehat{\boldsymbol{\theta}}_n$ otro estimador asintóticamente normal con matriz de covarianza asintótica \mathbf{V} , Maronna et al. (2019) define la eficiencia asintótica de $\widehat{\boldsymbol{\theta}}_n$ como

$$eff(\widehat{\boldsymbol{\theta}}_n) = \min_{\mathbf{c} \neq \mathbf{0}} \frac{\mathbf{c}^T \mathbf{V}_0 \mathbf{c}}{\mathbf{c}^T \mathbf{V} \mathbf{c}} . \quad (3.8)$$

donde $\mathbf{c} \in \mathbb{R}^p$. Se puede probar que (3.8) se puede escribir como

$$eff(\widehat{\boldsymbol{\theta}}_n) = \lambda_1(\mathbf{V}^{-1} \mathbf{V}_0) . \quad (3.9)$$

con $\lambda_1(\mathbf{V}^{-1} \mathbf{V}_0)$ el autovalor más pequeño de la matriz $\mathbf{V}^{-1} \mathbf{V}_0$.

Con los resultados mostrados en esta sección podemos proporcionar estimadores consistentes y asintóticamente normales, cuya varianza asintótica puede estimarse a partir de los datos de una muestra y con esto poder comparar la eficiencia respecto a otros

estimadores. En la siguiente sección estudiamos algunos conceptos de estadística robusta y como evaluar la robustez de un estimador.

3.3. Robustez

El objetivo de la estadística robusta es proporcionar procedimientos confiables ante la presencia de valores atípicos en la muestra. Vamos a considerar un valor atípico como una observación que se aleja de la mayoría de los datos.

Para entender las ideas fundamentales de la estadística robusta es preciso conocer el comportamiento asintótico de un estimador.

Consideremos $\hat{\theta}_n$ un estimador que depende de una muestra X_1, \dots, X_n de una distribución F_θ . En todos los casos de interés práctico existe un valor que depende de F_θ , $\hat{\theta}_\infty = \hat{\theta}_\infty(F_\theta)$ tal que

$$\hat{\theta}_n \xrightarrow{p} \hat{\theta}_\infty(F_\theta). \quad (3.10)$$

Llamamos a $\hat{\theta}_\infty(F_\theta)$ el valor asintótico del estimador en F_θ . Cuando $\hat{\theta}_\infty(F_\theta) = \theta$ decimos que $\hat{\theta}_n$ es un estimador consistente.

En estadística robusta asumimos que los datos no vienen exactamente de una distribución F_θ sino que, más bien, pertenecen a un entorno de esta, es decir, una proporción de los datos siguen F_θ mientras que otra proporción más pequeña puede seguir una distribución arbitraria.

Definición 3.1 (Entorno de contaminación). *Sea F_θ una familia de distribuciones de probabilidad conocida parametrizada por θ y sea $\varepsilon \in [0, 1]$, definimos el entorno de contaminación de F_θ como:*

$$\mathcal{F}(F_\theta, \varepsilon) = \{(1 - \varepsilon)F_\theta + \varepsilon G : G \in \mathcal{G}\}. \quad (3.11)$$

donde \mathcal{G} es una familia de distribuciones arbitraria.

\mathcal{G} puede representar el conjunto de todas las distribuciones, en algunos casos nos interesa particularmente el conjunto de distribuciones de masa puntual δ_{x_0} , esto es la

distribución que concentra toda su masa de probabilidad en un punto $P(X = x_0) = 1$.

La robustez de un estimador puede medirse mediante distintos indicadores, entre los cuales tenemos la función de influencia que determina la robustez frente a la presencia de una proporción infinitesimal de valores atípicos.

3.3.1. Función de influencia

Una forma de estudiar la robustez de un estimador es mediante su curva de sensibilidad, sea x_1, \dots, x_n un conjunto de datos y x_0 una observación, definimos la curva sensibilidad del estimador $\hat{\theta}_n$ como:

$$SC_n(x_0) = \frac{\hat{\theta}_{n+1}(x_1, \dots, x_n, x_0) - \hat{\theta}_n(x_1, \dots, x_n)}{1/(n+1)}.$$

La función de influencia introducida por Hampel (1974) puede verse como la versión asintótica de la curva de sensibilidad. La función de influencia es una aproximación del comportamiento de $\hat{\theta}_\infty$ cuando la muestra contiene una pequeña fracción de valores atípicos y se define como

$$\begin{aligned} IF_{\hat{\theta}}(x_0, F_\theta) &= \lim_{\varepsilon \rightarrow 0^+} \frac{\hat{\theta}_\infty((1-\varepsilon)F_\theta + \varepsilon\delta_{x_0}) - \hat{\theta}_\infty(F_\theta)}{\varepsilon} \\ &= \frac{\partial}{\partial \varepsilon} \hat{\theta}_\infty((1-\varepsilon)F_\theta + \varepsilon\delta_{x_0}) \Big|_{\varepsilon \rightarrow 0^+}. \end{aligned} \quad (3.12)$$

El término $\hat{\theta}_\infty((1-\varepsilon)F_\theta + \varepsilon\delta_{x_0})$ del numerador de la ecuación anterior es el valor asintótico del estimador cuando la distribución es F_θ y una fracción ε de valores atípicos iguales a x_0 , cuando ε es pequeño este valor se puede aproximar como

$$\hat{\theta}_\infty((1-\varepsilon)F_\theta + \varepsilon\delta_{x_0}) \approx \hat{\theta}_\infty(F_\theta) + \varepsilon IF_{\hat{\theta}}(x_0, F_\theta).$$

La función de influencia describe el efecto de una contaminación en el punto x_0 sobre la estimación. Lo ideal es que esta función sea acotada para tener una influencia limitada ante la presencia de una pequeña proporción de valores atípicos en la muestra. En el caso parámetro multidimensional, sean $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ y $\hat{\boldsymbol{\theta}}_\infty$ un estimador con valor asintótico $\hat{\boldsymbol{\theta}}_\infty$, podemos generalizar (3.12) para el caso multidimensional como

$$IF_{\hat{\boldsymbol{\theta}}}(x_0, F_{\boldsymbol{\theta}}) = \mathbf{B}^{-1} \boldsymbol{\Psi}(x_0, \hat{\boldsymbol{\theta}}_\infty). \quad (3.13)$$

donde \mathbf{B} es la matriz dada en (3.7); ver Maronna et al. (2019). Con la función de influencia podemos analizar la robustez de un estimador. A continuación veremos el comportamiento de la función de influencia del estimador de máxima verosimilitud para la distribución Beta, con esto nos daremos cuenta de la falta de robustez de este estimador y la necesidad de estudiar estimadores que presenten mejores propiedades de robustez.

3.3.2. Función de influencia del estimador de máxima verosimilitud para una distribución Beta

Para analizar la función de influencia vamos a escribir el estimador de máxima verosimilitud aplicado a la distribución Beta en forma de M estimador. Sea X_1, \dots, X_n una muestra iid de una distribución $Beta(a, b)$, llamemos $f_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} = (a, b)$ a su función de densidad de probabilidad. En el capítulo 2 describimos cómo calcular el estimador de máxima verosimilitud. Vamos a reescribirlo en forma de M estimador como la solución $\hat{\boldsymbol{\theta}}$ de:

$$\sum_{i=1}^n \Psi^{MV}(x_i, \hat{\boldsymbol{\theta}}) = \mathbf{0} \quad (3.14)$$

donde

$$\begin{aligned} \Psi^{MV}(x, \boldsymbol{\theta}) &= \begin{pmatrix} \psi_1^{MV}(x, \boldsymbol{\theta}) \\ \psi_2^{MV}(x, \boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial a} \log f_{\boldsymbol{\theta}}(x) \\ \frac{\partial}{\partial b} \log f_{\boldsymbol{\theta}}(x) \end{pmatrix} \\ &= \begin{pmatrix} \psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x) \\ \psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x) \end{pmatrix}. \end{aligned} \quad (3.15)$$

Luego $\Psi^{MV'}(x, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \Psi^{MV}(x, \boldsymbol{\theta})$ tiene la forma

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} \Psi^{MV}(x, \boldsymbol{\theta}) &= \begin{pmatrix} \frac{\partial}{\partial a} \psi_1^{MV}(x, \boldsymbol{\theta}) & \frac{\partial}{\partial b} \psi_1^{MV}(x, \boldsymbol{\theta}) \\ \frac{\partial}{\partial a} \psi_2^{MV}(x, \boldsymbol{\theta}) & \frac{\partial}{\partial b} \psi_2^{MV}(x, \boldsymbol{\theta}) \end{pmatrix} \\ &= \begin{pmatrix} \psi^{(1)}(a+b) - \psi^{(1)}(a) & \psi^{(1)}(a+b) \\ \psi^{(1)}(a+b) & \psi^{(1)}(a+b) - \psi^{(1)}(b) \end{pmatrix}. \end{aligned}$$

Notemos que $\frac{\partial}{\partial \boldsymbol{\theta}} \Psi^{\mathbf{MV}}(x, \boldsymbol{\theta})$ no depende de valores aleatorios, por tanto, la matriz \mathbf{B} es

$$\mathbf{B} = -\frac{\partial}{\partial \boldsymbol{\theta}} \Psi^{\mathbf{MV}}(x, \boldsymbol{\theta}).$$

Luego podemos escribir la función de influencia como:

$$IF_{\boldsymbol{\theta}}(x_0, F_{\boldsymbol{\theta}}) = \mathbf{B}^{-1} \Psi^{\mathbf{MV}}(x_0, \boldsymbol{\theta}).$$

$$\begin{aligned} IF_{\boldsymbol{\theta}}(x_0, F_{\boldsymbol{\theta}}) &= \frac{1}{\Delta} \begin{pmatrix} \psi^{(1)}(b) - \psi^{(1)}(a+b) & \psi^{(1)}(a+b) \\ \psi^{(1)}(a+b) & \psi^{(1)}(a) - \psi^{(1)}(a+b) \end{pmatrix} \begin{pmatrix} \psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x_0) \\ \psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x_0) \end{pmatrix} \\ &= \frac{1}{\Delta} \begin{pmatrix} [\psi^{(1)}(b) - \psi^{(1)}(a+b)][\psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x_0)] + \psi^{(1)}(a+b)[\psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x_0)] \\ \psi^{(1)}(a+b)[\psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x_0)] + [\psi^{(1)}(a) - \psi^{(1)}(a+b)][\psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x_0)] \end{pmatrix}. \end{aligned}$$

Dónde $\Delta = \det[-\Psi^{\mathbf{MV}'}(x, \boldsymbol{\theta})] = \psi^{(1)}(a)\psi^{(1)}(b) - \psi^{(1)}(a+b)[\psi^{(1)}(a) + \psi^{(1)}(b)]$, notemos que este resultado es el mismo de (2.15) que ya probamos que es positivo.

Para ver la falta de robustez del estimador de máxima verosimilitud veamos como se comporta la función de influencia cuando x_0 se acerca a los extremos 0 y 1.

$$\lim_{x_0 \rightarrow 0^+} [\psi^{(1)}(b) - \psi^{(1)}(a+b)][\psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x_0)] + \psi^{(1)}(a+b)[\psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x_0)] = -\infty.$$

$$\lim_{x_0 \rightarrow 1^-} [\psi^{(1)}(b) - \psi^{(1)}(a+b)][\psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x_0)] + \psi^{(1)}(a+b)[\psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x_0)] = -\infty.$$

$$\lim_{x_0 \rightarrow 0^+} \psi^{(1)}(a+b)[\psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x_0)] + [\psi^{(1)}(a) - \psi^{(1)}(a+b)][\psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x_0)] = -\infty.$$

$$\lim_{x_0 \rightarrow 1^-} \psi^{(1)}(a+b)[\psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x_0)] + [\psi^{(1)}(a) - \psi^{(1)}(a+b)][\psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x_0)] = -\infty.$$

Notemos que cuando x_0 se acerca a 0 o 1, la función de influencia se va a $-\infty$, esto indica que no es acotada y por tanto el estimador de máxima verosimilitud no es robusto en esos casos.

Consideramos los diferentes casos de la forma de la distribución Beta que fueron mencionados en el capítulo 2 y se grafican sus respectivas funciones de densidad e influencia

para una malla de valores en el intervalo $(0, 1)$ que representan las posibles observaciones atípicas en la muestra.

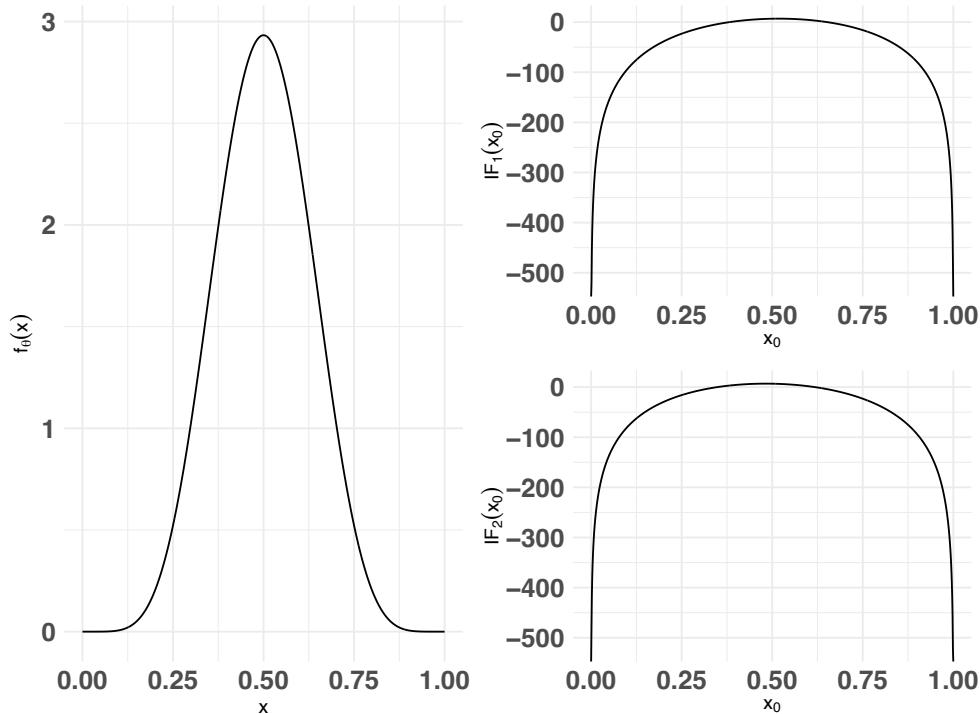


Figura 3.1: Funciones de densidad(izquierda) e influencia(derecha) para $a = b = 7$.

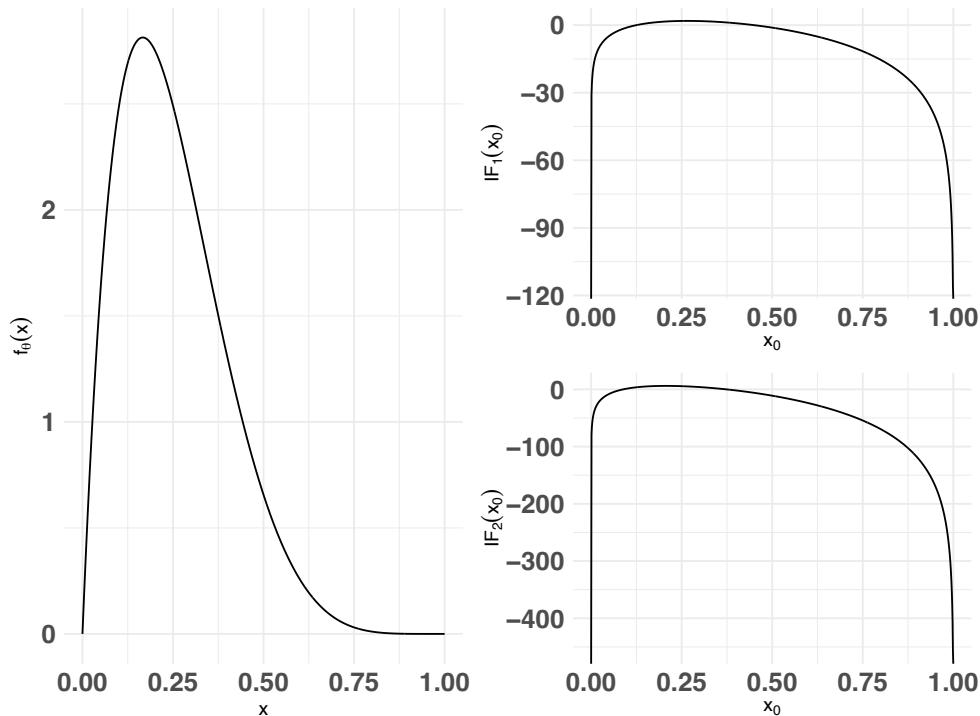


Figura 3.2: Funciones de densidad(izquierda) e influencia(derecha) para $a = 2, b = 6$.

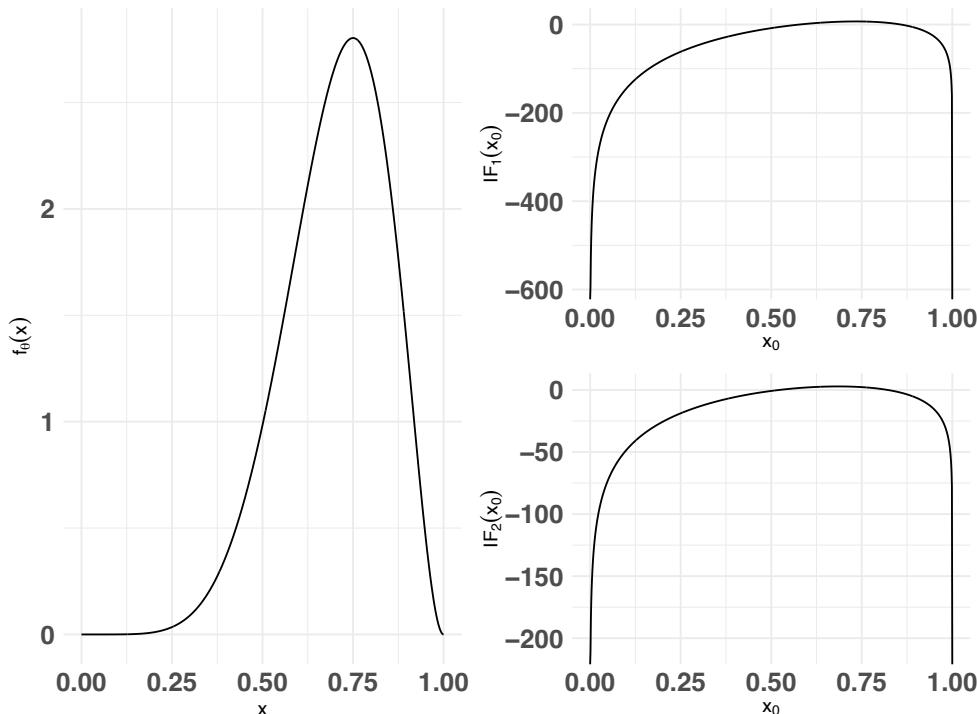


Figura 3.3: Funciones de densidad(izquierda) e influencia(derecha) para $a = 7$, $b = 3$.

Podemos notar en las Figuras 3.1–3.3 que la función de influencia resulta ser no acotada en ningún caso y esto indica la poca robustez del estimador de máxima verosimilitud. En general podemos apreciar que la influencia de una observación es mayor cuando mas lejos está del área donde se acumula mayor densidad. Por este motivo surge la necesidad de estudiar estimadores que ofrezcan mejores propiedades de robustez que las del estimador de máxima verosimilitud. En el siguiente capítulo presentamos dos estimadores robustos que pueden aplicarse a la estimación de parámetros de una distribución Beta.

Capítulo 4

Algunos estimadores robustos aplicados a la estimación de parámetros de una distribución Beta

4.1. Introducción

En la literatura podemos encontrar muchos estimadores robustos aplicados al problema de la estimación de parámetros, en este capítulo vamos a estudiar dos de ellos, un estimador basado en la transformación integral de probabilidad (MI estimador) y el estimador de divergencia de potencia mínima (MDP estimador). Primero vamos a describir de forma general estos estimadores y sus propiedades asintóticas, luego aplicaremos estos resultados al caso particular de una distribución Beta.

En general vamos a estudiar M estimadores definidos por (3.3). Decimos que un estimador es Fisher consistente para la familia $F_{\theta}(X)$ si

$$E_{\theta} \Psi(X, \theta) = \mathbf{0}. \quad (4.1)$$

Veremos que los estimadores que vamos a estudiar cumplen esta propiedad.

4.2. MI Estimador

El MI estimador es un M estimador robusto basado en la transformación integral de probabilidad que fue introducido por Fegyverneki (2003) para distribuciones continuas, Valdora and Yohai (2020) extendieron la propuesta de Fegyverneki para distribuciones discretas y muestran una aplicación para datos de conteo, específicamente para muestras aleatorias que siguen una distribución de Poisson o una Binomial negativa.

Definición 4.1. Sea $F_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \Theta$ con Θ un conjunto abierto en \mathbb{R}^p , una familia continua o discreta de funciones de distribución y $p_{\boldsymbol{\theta}}(k) = P_{\boldsymbol{\theta}}(X = k)$. La transformación integral de probabilidad $T(X)$ se define como

$$T(X) = F_{\boldsymbol{\theta}}(X) - V \times p_{\boldsymbol{\theta}}(X).$$

donde $V \sim U(0, 1)$ independiente de X .

Notemos que de la definición anterior si X es continua la transformación se reduce a $T(X) = F_{\boldsymbol{\theta}}(X)$. El siguiente teorema establece la distribución de la transformación integral de probabilidad.

Teorema 4.1. Si X es una variable aleatoria discreta o continua y T es la transformación integral de probabilidad, entonces

$$T(X) \sim U(0, 1).$$

La demostración del Teorema 4.1 se puede encontrar en la sección 5.3 de Ferguson (1967). Sabiendo que la distribución de la variable aleatoria $T(X)$ es uniforme, su k -ésimo momento es:

$$E_{\boldsymbol{\theta}}[(F_{\boldsymbol{\theta}}(X) - V p_{\boldsymbol{\theta}}(X))^k] = \frac{1}{k+1}.$$

y con esto podemos definir la siguiente función:

$$\psi_k^I(x, \boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[(F_{\boldsymbol{\theta}}(X) - V p_{\boldsymbol{\theta}}(X))^k | X = x] - 1/(k+1). \quad (4.2)$$

Definición 4.2. Sea X_1, \dots, X_n una muestra aleatoria donde cada X_i tiene función de densidad de probabilidad $f_{\boldsymbol{\theta}}(X)$, definimos el MI estimador $\hat{\boldsymbol{\theta}}_n$ de $\boldsymbol{\theta}$ como la solución de

$$\sum_{i=1}^n \psi^I(X_i, \boldsymbol{\theta}) = \mathbf{0}. \quad (4.3)$$

donde $\boldsymbol{\psi}^I(X, \boldsymbol{\theta}) = (\psi_1^I(X, \boldsymbol{\theta}), \dots, \psi_p^I(X, \boldsymbol{\theta}))^T$.

Si la variable aleatoria es continua la ecuación (4.2) se simplifica como:

$$\begin{aligned}\psi_1^I(X, \boldsymbol{\theta}) &= F_{\boldsymbol{\theta}}(X) - 1/2 \\ \psi_2^I(X, \boldsymbol{\theta}) &= F_{\boldsymbol{\theta}}^2(X) - 1/3 \\ &\vdots \\ \psi_k^I(X, \boldsymbol{\theta}) &= F_{\boldsymbol{\theta}}^k(X) - 1/(k+1).\end{aligned}\tag{4.4}$$

4.2.1. Propiedades asintóticas del MI estimador

Notemos que $E_{\boldsymbol{\theta}}(\psi_k^I(X, \boldsymbol{\theta})) = \mathbf{0}$, lo que asegura la consistencia de Fisher del MI estimador. Consideremos X_1, \dots, X_n una secuencia iid de variables aleatorias continuas o discretas con distribución $F_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \Theta$ un conjunto abierto en \mathbb{R}^p . El siguiente teorema establece la consistencia del MI estimador bajo ciertas condiciones en la función de densidad. Su demostración es similar al Teorema 2 de Huber (1967).

Teorema 4.2. *Sea $\hat{\boldsymbol{\theta}}_n$ el MI estimador de $\boldsymbol{\theta}$. Sea $\lambda(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(\Psi^I(\mathbf{X}, \boldsymbol{\theta}))$. Si se cumplen las siguientes condiciones:*

C1 La función $\lambda(\boldsymbol{\theta})$ tiene una única raíz en $\boldsymbol{\theta} = \boldsymbol{\theta}$.

C2 $\Psi^I(\mathbf{x}, \boldsymbol{\theta})$ es continua en $\boldsymbol{\theta}$.

C3 Existe un conjunto compacto $C \in \Theta$ tal que $E_{\boldsymbol{\theta}}(\inf_{\boldsymbol{\theta} \in \Theta - C} \|\Psi^I(\mathbf{X}, \boldsymbol{\theta})\|) > 0$

entonces

$$\hat{\boldsymbol{\theta}} \xrightarrow{c.s.} \boldsymbol{\theta}.$$

Sobre la distribución asintótica del MI estimador, esta sigue inmediatamente del Teorema 3.3.

4.2.2. MI estimador para la distribución Beta

En esta sección calculamos la función $\Psi^I(X, \boldsymbol{\theta})$ y la matriz de covarianza asintótica del MI estimador para la distribución Beta. El espacio de los parámetros de la distribución Beta tiene dimensión 2, la función de estimación tiene la forma

$$\Psi^I(X, \boldsymbol{\theta}) = \begin{pmatrix} \psi_1^I(X, \boldsymbol{\theta}) \\ \psi_2^I(X, \boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} F_{\boldsymbol{\theta}}(X) - 1/2 \\ F_{\boldsymbol{\theta}}^2(X) - 1/3 \end{pmatrix}.$$

Luego el MI estimador es la raíz del sistema de ecuaciones:

$$\begin{cases} \sum_{i=1}^n F_{\boldsymbol{\theta}}(x_i) - 1/2 = 0 \\ \sum_{i=1}^n F_{\boldsymbol{\theta}}^2(x_i) - 1/3 = 0. \end{cases} \quad (4.5)$$

La raíz de (4.5) no puede encontrarse de forma explícita por lo que es necesario utilizar métodos numéricos para aproximarla. Las matrices \mathbf{A} y \mathbf{B} de la varianza asintótica son:

$$\begin{aligned} \mathbf{A} &= E_{\boldsymbol{\theta}}(\Psi^I(X_1, \boldsymbol{\theta}) \Psi^I(X_1, \boldsymbol{\theta})^T) \\ &= E_{\boldsymbol{\theta}} \left[\begin{pmatrix} F_{\boldsymbol{\theta}}(X) - 1/2 \\ F_{\boldsymbol{\theta}}^2(X) - 1/3 \end{pmatrix} (F_{\boldsymbol{\theta}}(X) - 1/2, F_{\boldsymbol{\theta}}^2(X) - 1/3) \right] \\ &= E_{\boldsymbol{\theta}} \begin{pmatrix} (F_{\boldsymbol{\theta}}(X) - 1/2)^2 & (F_{\boldsymbol{\theta}}(X) - 1/2)(F_{\boldsymbol{\theta}}^2(X) - 1/3) \\ (F_{\boldsymbol{\theta}}(X) - 1/2)(F_{\boldsymbol{\theta}}^2(X) - 1/3) & (F_{\boldsymbol{\theta}}^2(X) - 1/3)^2 \end{pmatrix} \\ &= E_{\boldsymbol{\theta}} \begin{pmatrix} F_{\boldsymbol{\theta}}^2(X) - F_{\boldsymbol{\theta}}(X) + 1/4 & F_{\boldsymbol{\theta}}^3(X) - 1/3F_{\boldsymbol{\theta}}(X) - 1/2F_{\boldsymbol{\theta}}^2(X) + 1/6 \\ F_{\boldsymbol{\theta}}^3(X) - 1/3F_{\boldsymbol{\theta}}(X) - 1/2F_{\boldsymbol{\theta}}^2(X) + 1/6 & F_{\boldsymbol{\theta}}^4(X) - 2/3F_{\boldsymbol{\theta}}^2(X) + 1/9 \end{pmatrix}. \end{aligned}$$

Como $F_{\boldsymbol{\theta}}(X) \sim U(0, 1)$ y usando la linealidad de la esperanza obtenemos la expresión final de la matriz \mathbf{A}

$$\mathbf{A} = \begin{pmatrix} 1/12 & 1/12 \\ 1/12 & 4/45 \end{pmatrix}. \quad (4.6)$$

Ahora la matriz \mathbf{B}

$$\mathbf{B} = -E_{\boldsymbol{\theta}} \begin{pmatrix} \frac{\partial}{\partial a} \psi_1^I(X, \boldsymbol{\theta}) & \frac{\partial}{\partial b} \psi_1^I(X, \boldsymbol{\theta}) \\ \frac{\partial}{\partial a} \psi_2^I(X, \boldsymbol{\theta}) & \frac{\partial}{\partial b} \psi_2^I(X, \boldsymbol{\theta}) \end{pmatrix}.$$

La deducción de los elementos de la matriz \mathbf{B} se pueden consultar en el Apéndice A.

4.2.3. Función de influencia del MI estimador para la distribución Beta

Escribamos la matriz \mathbf{B} como:

$$\mathbf{B} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

dónde la forma de los elementos B_{ij} puede consultarse en el Apéndice A, la función de influencia asociada al MI estimador tiene la forma

$$\begin{aligned} IF_{\theta}(x_0, F_{\theta}) &= \frac{1}{\Delta} \begin{pmatrix} -B_{22} & B_{12} \\ B_{21} & -B_{11} \end{pmatrix} \begin{pmatrix} F_{\theta}(x_0) - 1/2 \\ F_{\theta}^2(x_0) - 1/3 \end{pmatrix} \\ &= \frac{1}{\Delta} \begin{pmatrix} -B_{22}(F_{\theta}(x_0) - 1/2) + B_{12}(F_{\theta}^2(x_0) - 1/3) \\ B_{21}(F_{\theta}(x_0) - 1/2) - B_{11}(F_{\theta}^2(x_0) - 1/3) \end{pmatrix}. \end{aligned}$$

dónde $\Delta = B_{11}B_{22} - B_{12}B_{21}$.

Para verificar la robustez del MI estimador su función de influencia debe ser acotada, para probar esto consideremos el límite de la función de influencia cuando x_0 se acerca a los extremos 0 o 1.

$$\lim_{x_0 \rightarrow 0^+} \frac{1}{\Delta} (-B_{22}(F_{\theta}(x_0) - 1/2) + B_{12}(F_{\theta}^2(x_0) - 1/3)) = \frac{1}{\Delta} \left(\frac{1}{2}B_{22} - \frac{1}{3}B_{12} \right).$$

$$\lim_{x_0 \rightarrow 1^-} \frac{1}{\Delta} (-B_{22}(F_{\theta}(x_0) - 1/2) + B_{12}(F_{\theta}^2(x_0) - 1/3)) = \frac{1}{\Delta} \left(-\frac{1}{2}B_{22} + \frac{2}{3}B_{12} \right).$$

$$\lim_{x_0 \rightarrow 0^+} \frac{1}{\Delta} (B_{21}(F_{\theta}(x_0) - 1/2) - B_{11}(F_{\theta}^2(x_0) - 1/3)) = \frac{1}{\Delta} \left(-\frac{1}{2}B_{21} + \frac{1}{3}B_{11} \right).$$

$$\lim_{x_0 \rightarrow 1^-} \frac{1}{\Delta} (B_{21}(F_{\theta}(x_0) - 1/2) - B_{11}(F_{\theta}^2(x_0) - 1/3)) = \frac{1}{\Delta} \left(\frac{1}{2}B_{21} - \frac{2}{3}B_{11} \right).$$

Con los resultados anteriores podemos ver que la función de influencia resulta ser acotada en todos los casos. Las siguientes figuras muestran la función de influencia del MI estimador considerando varios casos.

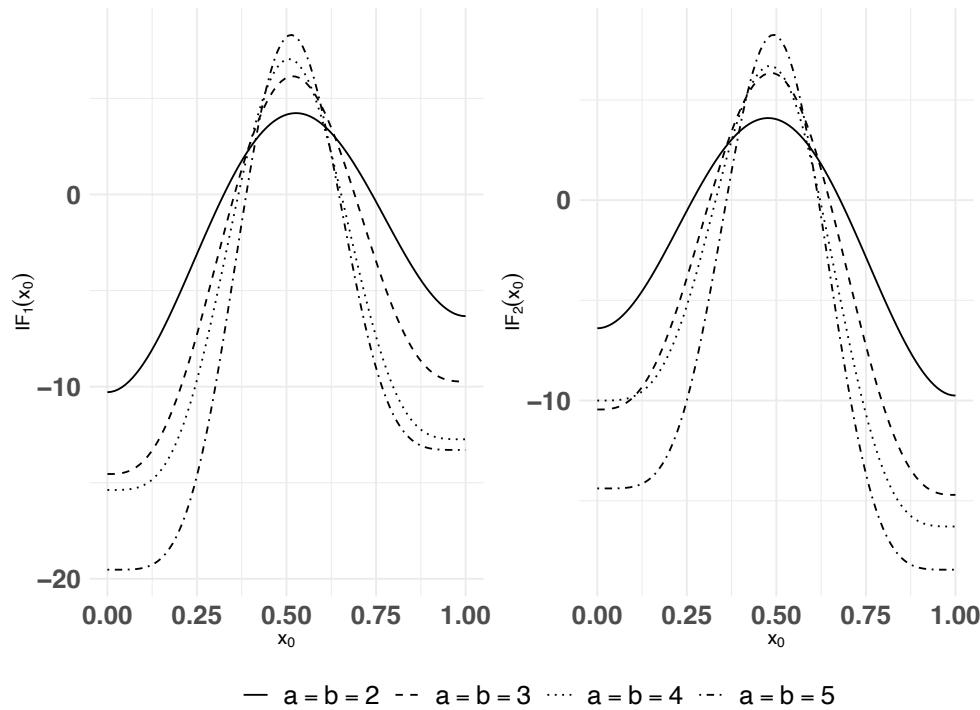


Figura 4.1: Función de influencia del MI estimador cuándo $a = b$.

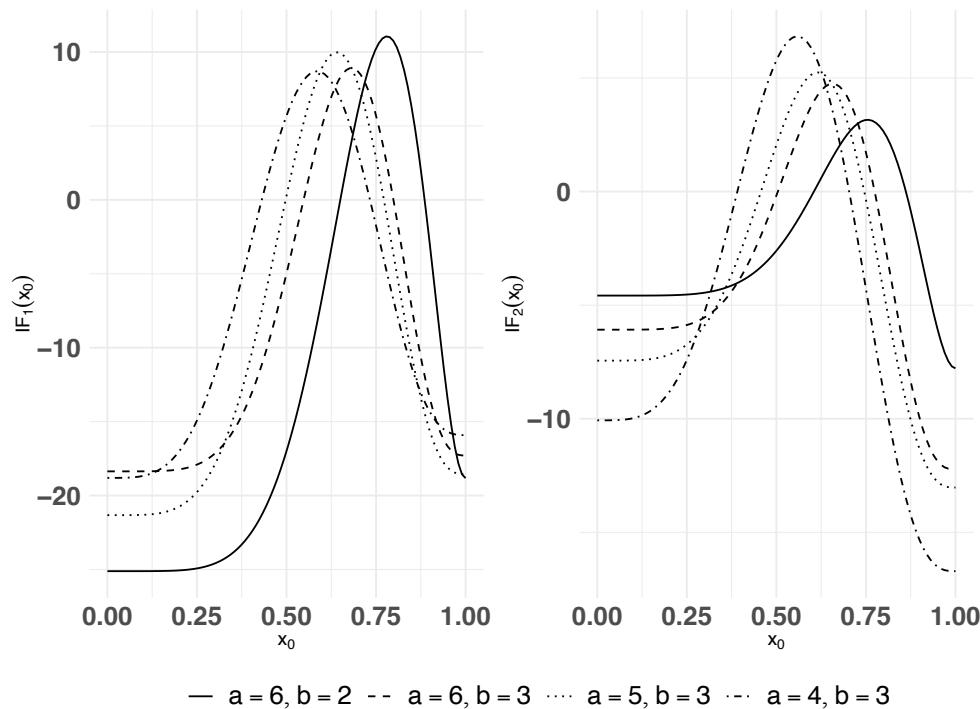


Figura 4.2: Función de influencia del MI estimador cuándo $a > b$.

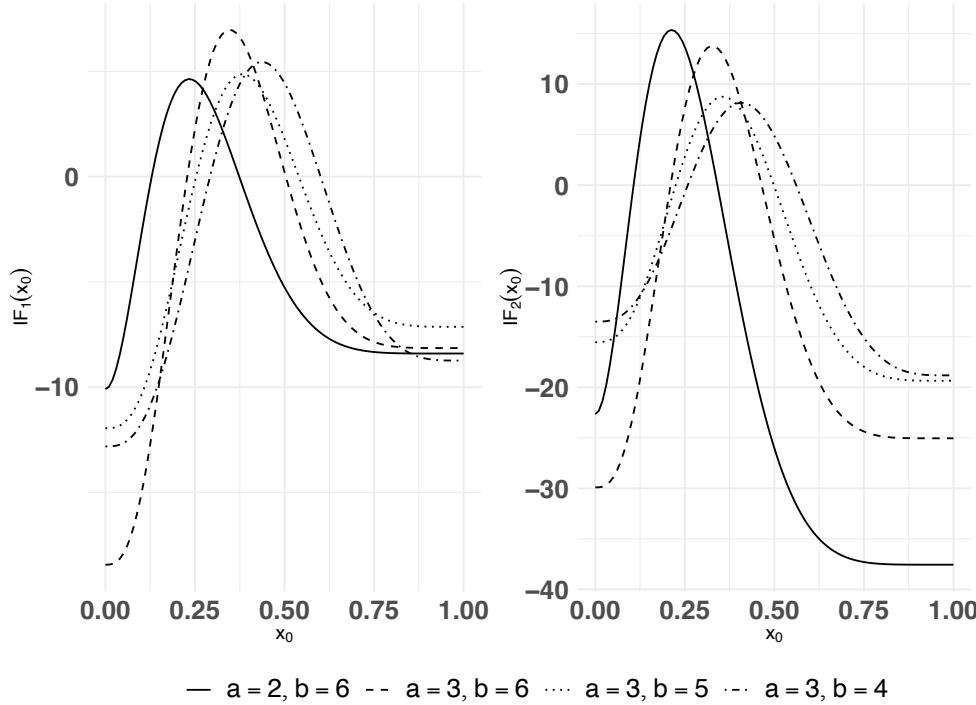


Figura 4.3: Función de influencia del MI estimador cuándo $a < b$.

Podemos notar que las funciones de influencia del MI estimador resultan ser acotadas y con valores mas pequeños que las del estimador de máxima verosimilitud, esto indica que los valores muy alejados de donde se acumula la mayor densidad tiene una influencia mas limitada en el estimador y por lo tanto el MI estimador resulta ser mas robusto.

4.3. Estimador de divergencia de potencia mínima

Basu et al. (1998) proponen un procedimiento robusto de estimación de parámetros que se obtiene minimizando una medida de divergencia sobre el espacio de parámetros, basada en la densidad llamada divergencia de potencia de densidad. Este estimador está indexado por un parámetro $\alpha \geq 0$ que controla el balance entre robustez y eficiencia.

Definición 4.3. Sea $\alpha \geq 0$ y sean g y f dos funciones de densidad, definimos la divergencia de potencia de densidad entre g y f como:

$$d_\alpha(g, f) = \begin{cases} \int \left\{ f^{1+\alpha}(x) - \left(1 + \frac{1}{\alpha}\right) g(x) f^\alpha(x) + \frac{1}{\alpha} g^{\alpha+1}(x) \right\} dx & \alpha > 0 \\ \lim_{\alpha \rightarrow 0} d_\alpha(g, f) = \int g(x) \log \left[\frac{g(x)}{f(x)} \right] dx & \alpha = 0. \end{cases} \quad (4.7)$$

Notemos que si $\alpha = 0$ tenemos la divergencia de Kullback-Lieber. En el proceso de estimación nos interesan valores pequeños de α , específicamente entre 0 y 1, en general se pueden considerar valores de α mayores a 1 pero el procedimiento se vuelve menos eficiente a medida que α aumenta. Consideraremos una familia paramétrica de distribuciones $F_{\boldsymbol{\theta}}$ con densidades $f_{\boldsymbol{\theta}}$ y una muestra aleatoria X_1, \dots, X_n de una distribución G que no pertenece a la familia $F_{\boldsymbol{\theta}}$, con densidad g . Si $f_{\boldsymbol{\theta}}$ juega el rol de f en la ecuación (4.7) y g es la función de densidad estimada de la muestra, entonces podemos reescribir (4.7) como:

$$d_{\alpha}(g, f_{\boldsymbol{\theta}}) = \begin{cases} \int f_{\boldsymbol{\theta}}^{1+\alpha}(x)dx - (1 + \frac{1}{\alpha})E[f_{\boldsymbol{\theta}}^{\alpha}(X)] + \frac{1}{\alpha}E[g^{\alpha}(X)] & \alpha > 0 \\ E[\log g(X)] - E[\log f_{\boldsymbol{\theta}}(X)] & \alpha = 0. \end{cases} \quad (4.8)$$

donde $E[\cdot]$ denota la esperanza respecto a g . Aquí los términos $E[g^{\alpha}(X)]$ y $E[\log g(X)]$ no dependen de $\boldsymbol{\theta}$ y por tanto pueden ignorarse para minimizar la función $d_{\alpha}(g, f_{\boldsymbol{\theta}})$. Por otro lado, dada una muestra X_1, \dots, X_n las esperanzas $E[f_{\boldsymbol{\theta}}^{\alpha}(X, \boldsymbol{\theta})]$ y $E[\log f_{\boldsymbol{\theta}}(X, \boldsymbol{\theta})]$ pueden estimarse mediante $\frac{1}{n} \sum_{i=1}^n f_{\boldsymbol{\theta}}^{\alpha}(X_i, \boldsymbol{\theta})$ y $\frac{1}{n} \sum_{i=1}^n \log f_{\boldsymbol{\theta}}(X_i, \boldsymbol{\theta})$ respectivamente. Con estos resultados podemos definir el estimador de divergencia de potencia mínima (MDP) $\hat{\boldsymbol{\theta}}_n$ con parámetro de ajuste $\alpha > 0$ como:

$$\hat{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta} \in \Theta} H_{\alpha}(\boldsymbol{\theta}) \quad (4.9)$$

donde

$$H_{\alpha}(\boldsymbol{\theta}) = \int f_{\boldsymbol{\theta}}^{1+\alpha}(x)dx - \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^n f_{\boldsymbol{\theta}}^{\alpha}(x_i).$$

Notemos que el caso de $\alpha = 0$ obtenemos el estimador de máxima verosimilitud. Las ecuaciones de estimación tienen la forma

$$\frac{1}{n} \sum \mathbf{u}_{\boldsymbol{\theta}}(x_i) f_{\boldsymbol{\theta}}^{\alpha}(x_i) - \int \mathbf{u}_{\boldsymbol{\theta}}(x) f_{\boldsymbol{\theta}}^{1+\alpha}(x)dx = 0 \quad (4.10)$$

donde $\mathbf{u}_{\boldsymbol{\theta}}(x) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(x)$. Podemos notar que este estimador es de hecho un M estimador con función Ψ correspondiente a

$$\Psi(x, \boldsymbol{\theta}) = \mathbf{u}_{\boldsymbol{\theta}}(x) f_{\boldsymbol{\theta}}^{\alpha}(x) - \int \mathbf{u}_{\boldsymbol{\theta}}(x) f_{\boldsymbol{\theta}}^{1+\alpha}(x)dx. \quad (4.11)$$

4.3.1. Distribución asintótica del MDP estimador

Encontrar una forma cerrada para el estimador (4.9) no es posible, de la misma forma no podemos encontrar formas cerradas para su varianza y esperanza, por este motivo es difícil encontrar su distribución exacta. Basu et al. (1998) propone ciertas condiciones de regularidad bajo las que podemos deducir la distribución asintótica del estimador.

Teorema 4.3. *Bajo ciertas condiciones de regularidad, con probabilidad tendiendo a 1 cuando $n \rightarrow \infty$, existe $\hat{\boldsymbol{\theta}}_n$ tal que:*

1. $\hat{\boldsymbol{\theta}}_n$ es consistente para $\boldsymbol{\theta}$,
2. $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, \mathbf{J}^{-1} \mathbf{K} \mathbf{J}^{-1})$ donde

$$\begin{aligned}\mathbf{J} &= \int \mathbf{u}_{\boldsymbol{\theta}}(x) \mathbf{u}_{\boldsymbol{\theta}}^T(x) f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx, \\ \mathbf{K} &= \int \mathbf{u}_{\boldsymbol{\theta}}(x) \mathbf{u}_{\boldsymbol{\theta}}^T(x) f_{\boldsymbol{\theta}}^{1+2\alpha}(x) dx - \boldsymbol{\xi} \boldsymbol{\xi}^T, \\ \boldsymbol{\xi} &= \int \mathbf{u}_{\boldsymbol{\theta}}(x) f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx.\end{aligned}$$

En la siguiente sección se deducen la forma de las ecuaciones de estimación y matrices de covarianza del MI estimador y del MDP estimador aplicado a una distribución Beta.

4.3.2. MDP estimador para la distribución Beta

En esta sección calculamos la función $\Psi(X, \boldsymbol{\theta})$ y la matriz de covarianza asintótica del MDP estimador para la distribución Beta. Para deducir los resultados del estimador de divergencia de potencia mínima vamos a necesitar la siguiente lista de integrales, su desarrollo puede consultarse en el Apéndice A.

$$\int f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx = \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)}. \quad (4.12)$$

$$\begin{aligned}\int \log(x) f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\ &\times (\psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)).\end{aligned} \quad (4.13)$$

$$\begin{aligned} \int \log(1-x) f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\ &\quad \times (\psi^{(0)}(\alpha b - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)). \end{aligned} \quad (4.14)$$

$$\begin{aligned} \int \log^2(x) f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\ &\quad \times [\psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)]^2 \\ &\quad + \psi^{(1)}(\alpha a - \alpha + a) - \psi^{(1)}(\alpha a + \alpha b - 2\alpha + a + b). \end{aligned} \quad (4.15)$$

$$\begin{aligned} \int \log^2(1-x) f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\ &\quad \times [\psi^{(0)}(\alpha b - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)]^2 \\ &\quad + \psi^{(1)}(\alpha b - \alpha + b) - \psi^{(1)}(\alpha a + \alpha b - 2\alpha + a + b). \end{aligned} \quad (4.16)$$

$$\begin{aligned} \int \log(x) \log(1-x) f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\ &\quad \times [\psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)] \\ &\quad \times [\psi^{(0)}(\alpha b - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)] \\ &\quad - \psi^{(1)}(\alpha a + \alpha b - 2\alpha + a + b). \end{aligned} \quad (4.17)$$

El termino $\mathbf{u}_{\boldsymbol{\theta}}(x)$ tiene la forma

$$\mathbf{u}_{\boldsymbol{\theta}}(x) = \begin{pmatrix} \frac{\partial}{\partial a} \log f_{\boldsymbol{\theta}}(x) \\ \frac{\partial}{\partial b} \log f_{\boldsymbol{\theta}}(x) \end{pmatrix}.$$

Tenemos que

$$\log f_{\boldsymbol{\theta}}(x) = \log \Gamma(a+b) - \log \Gamma(a) - \log \Gamma(b) + (a-1) \log(x) + (b-1) \log(1-x).$$

Luego las derivadas respecto a a y b son, respectivamente:

$$\frac{\partial}{\partial a} \log f_{\boldsymbol{\theta}} = \psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x). \quad (4.18)$$

$$\frac{\partial}{\partial b} \log f_{\boldsymbol{\theta}} = \psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x). \quad (4.19)$$

Con los resultados anteriores podemos encontrar expresiones de la ecuación (4.11) en términos de funciones conocidas, primero veamos respecto al parámetro a y tomando los resultados (4.12), (4.13) y (4.18) tenemos:

$$\begin{aligned}
 \int \frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial a} f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx &= \int [\psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x)] f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx \\
 &= [\psi^{(0)}(a+b) - \psi^{(0)}(a)] \int f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx + \int \log(x) f_{\boldsymbol{\theta}}^{(1+\alpha)}(x) dx \\
 &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\
 &\quad \times [\psi^{(0)}(a+b) - \psi^{(0)}(a) + \psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)].
 \end{aligned} \tag{4.20}$$

ahora respecto a b y considerando (4.12), (4.14) y (4.19)

$$\begin{aligned}
 \int \frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial b} f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx &= \int [\psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x)] f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx \\
 &= [\psi^{(0)}(a+b) - \psi^{(0)}(b)] \int f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx + \int \log(1-x) f_{\boldsymbol{\theta}}^{(1+\alpha)}(x) dx \\
 &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\
 &\quad \times [\psi^{(0)}(a+b) - \psi^{(0)}(b) + \psi^{(0)}(\alpha b - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)].
 \end{aligned} \tag{4.21}$$

La función Ψ del MDP estimador es

$$\Psi(x, \boldsymbol{\theta}) = \begin{pmatrix} \psi_1(x, \boldsymbol{\theta}) \\ \psi_2(x, \boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \frac{\partial \log f_{\boldsymbol{\theta}_0}(x)}{\partial a} f_{\boldsymbol{\theta}}^{\alpha}(x) - \int \frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial a} f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx \\ \frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial b} f_{\boldsymbol{\theta}}^{\alpha}(x) - \int \frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial b} f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx \end{pmatrix}. \tag{4.22}$$

Con las integrales encontradas al inicio podemos encontrar expresiones para (4.22) de la siguiente forma:

$$\begin{aligned}
 \psi_1(x, \boldsymbol{\theta}) &= \frac{\partial \log f_{\boldsymbol{\theta}}}{\partial a} f_{\boldsymbol{\theta}}^{\alpha}(x) - \int \frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial a} f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx \\
 &= [\psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x)] \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{\alpha} x^{\alpha(a-1)} (1-x)^{\alpha(b-1)} \\
 &\quad - \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\
 &\quad \times [\psi^{(0)}(a+b) - \psi^{(0)}(a) + \psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)].
 \end{aligned}$$

$$\begin{aligned}
 \psi_2(x, \boldsymbol{\theta}) &= \frac{\partial \log f_{\boldsymbol{\theta}}}{\partial b} f_{\boldsymbol{\theta}_0}^{\alpha}(x) - \int \frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial b} f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx \\
 &= [\psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x)] \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{\alpha} x^{\alpha(a-1)} (1-x)^{\alpha(b-1)} \\
 &\quad - \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\
 &\quad \times [\psi^{(0)}(a+b) - \psi^{(0)}(b) + \psi^{(0)}(\alpha b - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)].
 \end{aligned}$$

Luego si X_1, \dots, X_n es una muestra aleatoria, el MDP estimador es la raíz del sistema

$$\begin{cases} \sum_{i=1}^n \psi_1(X_i, \boldsymbol{\theta}) = 0 \\ \sum_{i=1}^n \psi_2(X_i, \boldsymbol{\theta}) = 0. \end{cases} \tag{4.23}$$

Como en el caso del MI estimador una vez mas debemos recurrir a métodos numéricos para encontrar la solución de (4.23). Veamos ahora la deducción de la distribución asintótica del estimador. Las matrices \mathbf{J} , \mathbf{K} y $\boldsymbol{\xi}$ dadas en el Teorema 4.3 se pueden encontrar de forma explícita usando los resultados de las integrales calculadas previamente. Las entradas de $\boldsymbol{\xi}$ son directamente las integrales (4.20) y (4.21), mientras que la deducción de las matrices \mathbf{J} y \mathbf{K} las podemos encontrar en el Apéndice A, notemos que el primer término de la matriz \mathbf{K} tiene la misma forma de la matriz \mathbf{J} con 2α en lugar de α .

4.3.3. Función de influencia del MDP estimador para la distribución Beta

Si \mathbf{J} y $\boldsymbol{\xi}$ son finitos la función de influencia es acotada siempre que $\mathbf{u}_{\boldsymbol{\theta}}(x)f_{\boldsymbol{\theta}}^{\alpha}(x)$ sea acotada, Basu et al. (1998) muestra que la función de influencia del estimador de divergencia de potencia mínima tiene la forma:

$$IF_{\hat{\boldsymbol{\theta}}}(x_0, F_{\boldsymbol{\theta}}) = \mathbf{J}^{-1} (\mathbf{u}_{\boldsymbol{\theta}}(x_0)f_{\boldsymbol{\theta}}^{\alpha}(x_0) - \boldsymbol{\xi}).$$

El MDP estimador nos proporciona una familia de estimadores indexados por el parámetro α , donde el estimador de máxima verosimilitud es un caso particular cuando $\alpha = 0$, aumentar el valor de α nos proporciona un estimador mas robusto. En las Figuras 4.4–4.6 podemos notar que la función de influencia se va a acotando con el aumento del valor del parámetro α .

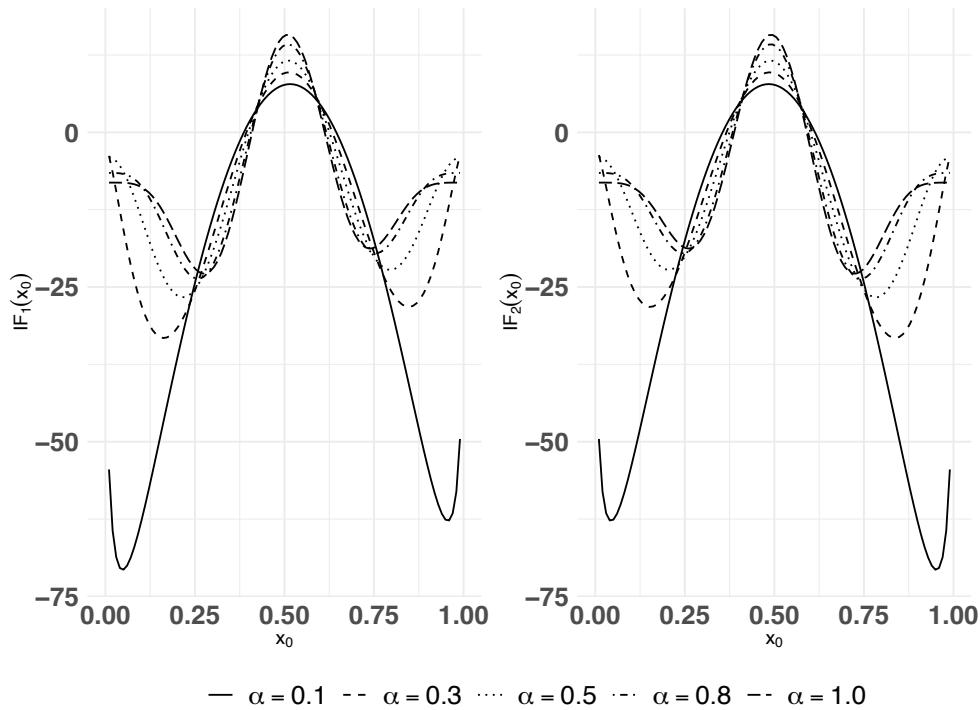


Figura 4.4: Función de influencia para el estimador de divergencia de potencia mínima para una distribución Beta con $a = b = 7$ y diferentes valores de α .

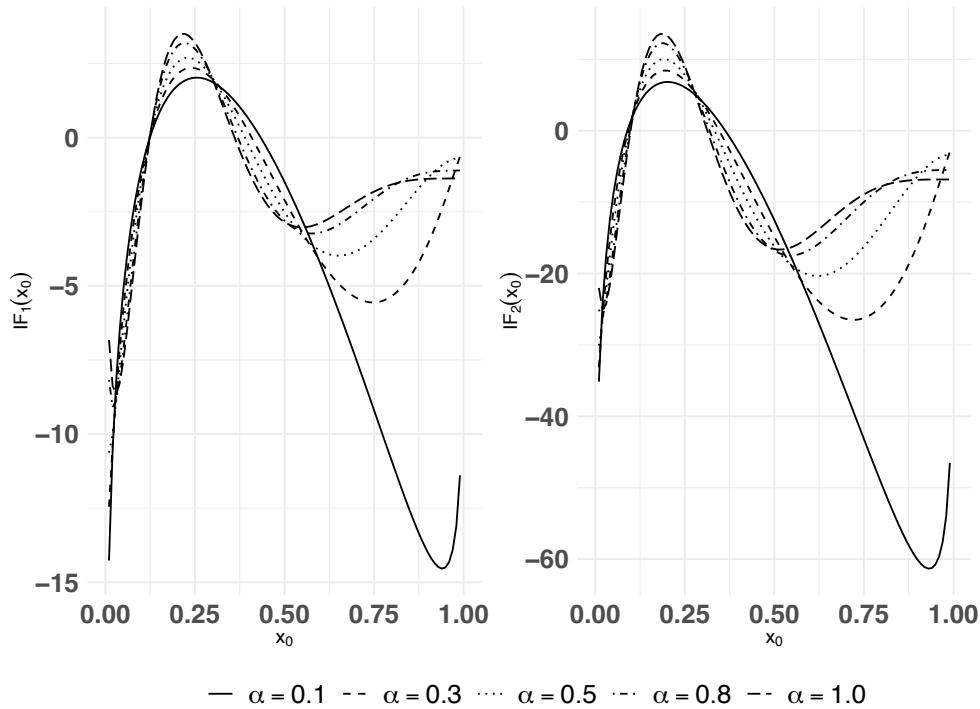


Figura 4.5: Función de influencia para el estimador de divergencia de potencia mínima para una distribución Beta con $a = 2$, $b = 6$ y diferentes valores de α .

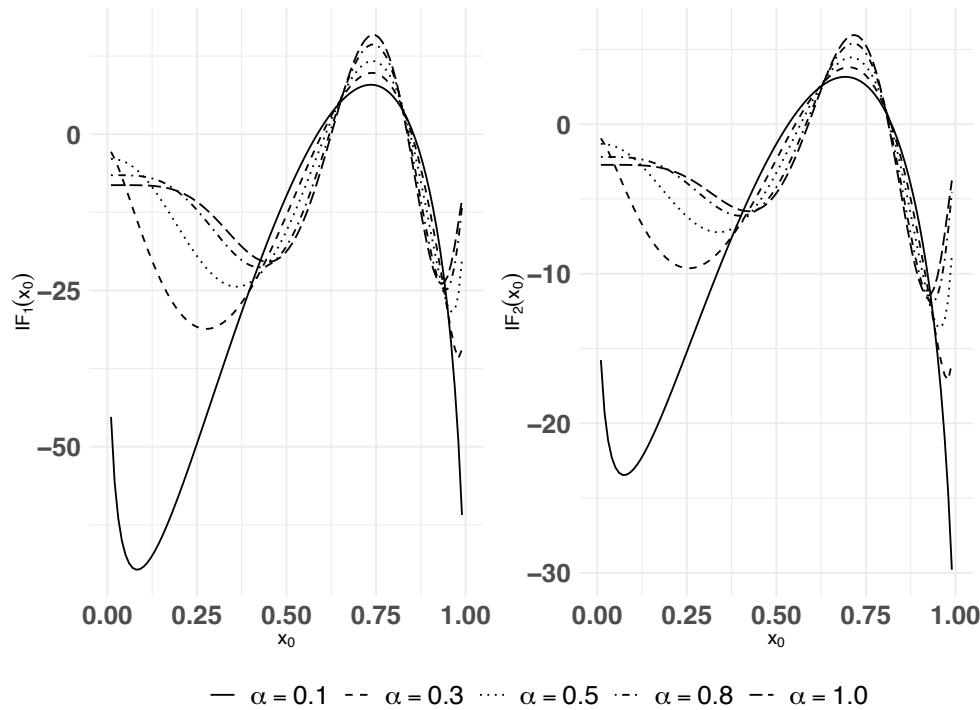


Figura 4.6: Función de influencia para el estimador de divergencia de potencia mínima para una distribución Beta con $a = 7$, $b = 3$ y diferentes valores de α .

Capítulo 5

Resultados de los experimentos y aplicaciones

En este capítulo se realizan las comparaciones del estimador de máxima verosimilitud, MI estimador y MDP estimador, para estimar los parámetros de los diferentes casos de la distribución *Beta* mediante simulaciones en R. Se consideran experimentos sin contaminación en los datos y luego eligiendo contaminaciones adecuadas para cada caso de la distribución *Beta*. Al final del capítulo se muestran algunas aplicaciones de los estimadores usando datos reales.

5.1. Eficiencia relativa

El primer experimento que se realizó fue calcular la eficiencia relativa del MI estimador y MDP estimador respecto al estimador de máxima verosimilitud con dos objetivos: primero, poder comparar la eficiencia de los estimadores robustos ante el estimador de máxima verosimilitud y poder comprobar que estos estimadores brindan buenos resultados ante un escenario de no contaminación en los datos y como segundo objetivo de estos experimentos, elegir un valor de α para el MDP estimador con el cual se hacen los experimentos posteriores.

Se hicieron 1000 repeticiones en las cuales se simularon muestras de tamaño 100 y se calcularon los 3 estimadores para cada muestra generada, con esto obtenemos vectores de estimadores de los parámetros a y b y para cada estimador calculamos el error cuadrático medio empírico definido por

$$ECM(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \theta)^2$$

luego la eficiencia relativa está dada por:

$$ER(\hat{\theta}_{MV}, \hat{\theta}_{ROB}) = \frac{ECM(\hat{\theta}_{MV})}{ECM(\hat{\theta}_{ROB})}.$$

Donde $\hat{\theta}_{MV}$ es el estimador de máxima verosimilitud para a o b y $\hat{\theta}_{ROB}$ es el MI o el MDP estimador para a o b . En la Figura 5.1 se muestran las funciones de densidad consideradas en los experimentos y los resultados se resumen en las tablas siguientes.

Tabla 5.1: Eficiencia relativa del MI estimador y el estimador de divergencia de potencia mínima respecto al estimador de máxima verosimilitud para $a = b = 7$.

Estimador	Eficiencia Relativa	
	a	b
MI estimador	0.7995	0.7907
MDP estimador	0.9526	0.9491
$\alpha = 0.2$	0.8961	0.8924
$\alpha = 0.3$	0.8324	0.8291
$\alpha = 0.4$		

Tabla 5.2: Eficiencia relativa del MI estimador y el estimador de divergencia de potencia mínima respecto al estimador de máxima verosimilitud para $a = 2, b = 6$.

Estimador	Eficiencia Relativa	
	a	b
MI estimador	0.7905	0.8024
MDP estimador	0.9715	0.9538
$\alpha = 0.2$	0.9302	0.8963
$\alpha = 0.3$	0.8803	0.8310
$\alpha = 0.4$		

En general en las Tablas 5.1–5.2 podemos notar que la eficiencia relativa del MI estimador está entre 0.79 y 0.80 dependiendo del modelo y del parámetro a estimar, esto indica que en relación al estimador de máxima verosimilitud, el MI estimador ofrece entre el 79 % y el 80 % de la eficiencia del estimador de máxima verosimilitud. Por otro lado, respecto al MDP estimador tenemos eficiencias relativas mas altas que las que proporciona el MI estimador. Recordemos que si $\alpha = 0$ el MDP estimador es el estimador de máxima verosimilitud y cuanto mas grande se considera el valor de α el estimador se vuelve menos eficiente. Para tener un valor que compense la eficiencia y la robustez nos vamos a quedar con $\alpha = 0.3$ para los experimentos que se harán a continuación.

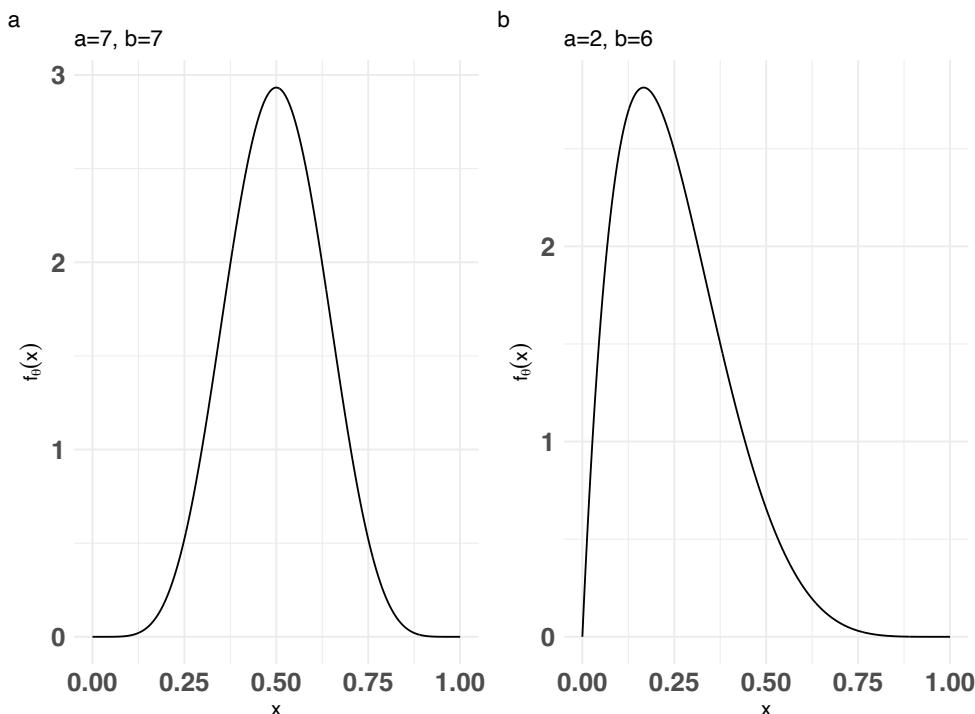


Figura 5.1: Funciones de densidad consideradas para los experimentos.

5.2. Error cuadrático medio

Se realizaron experimentos contaminando las muestras simuladas tomando los valores de la contaminación sobre una malla de valores en el intervalo $(0,1)$ y se consideraron porcentajes de contaminación de 3 %, 5 %, 7 % y 10 %. Se toma un valor de contaminación y se genera una muestra de tamaño 100, luego se reemplaza un porcentaje de las observaciones simuladas por el valor de contaminación y se ajustan los estimadores. Este proceso se repite 1000 veces y con esto calculamos el error cuadrático medio empírico de cada estimación mediante

$$ECM(\hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{i=1}^n \|\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\|^2.$$

donde $\boldsymbol{\theta} = (a, b)^T$ es el vector de parámetros con el cual se simulan los datos y $\hat{\boldsymbol{\theta}}$ es el estimador del vector de parámetros. Las siguientes figuras muestran los gráficos de los errores cuadráticos medios calculados para los casos cuando $a = b = 7$ y $a = 2, b = 6$.

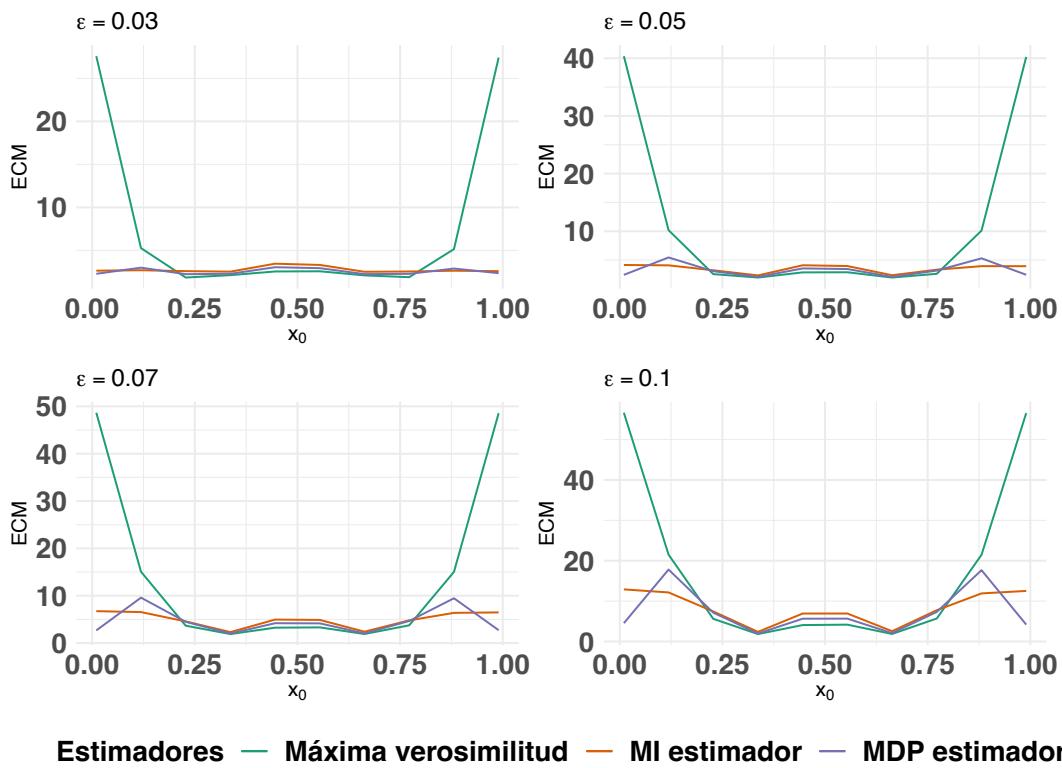


Figura 5.2: Error cuadrático medio de los estimadores para $a = b = 7$ para diferentes porcentajes de contaminación.

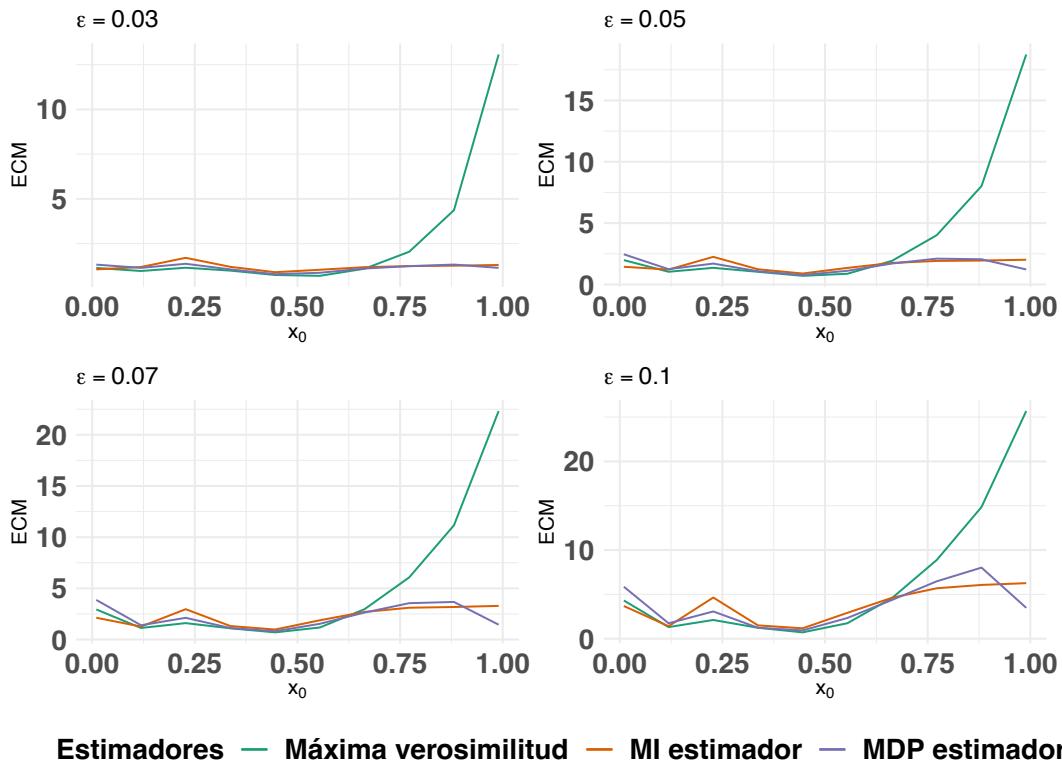


Figura 5.3: Error cuadrático medio de los estimadores para $a = 2$ $b = 6$ para diferentes porcentajes de contaminación.

La Figura 5.2 corresponde a datos generados a partir de la densidad de la Figura 5.1(a) donde podemos apreciar una distribución simétrica centrada en 0.5, si los valores de contaminación se van acercando más a las colas de la distribución vemos que los estimadores robustos proporcionan menor error cuadrático medio, esto indica que hay menos discrepancia en las estimaciones realizadas con los estimadores robustos que con el estimador de máxima verosimilitud y si consideramos los valores de contaminación en el intervalo donde se acumula mayor densidad la diferencia entre el error cuadrático medio proporcionado por los estimadores robustos y el estimador de máxima verosimilitud es similar.

Resultados parecidos a los anteriores notamos en la Figura 5.3 que corresponden a las densidad de la Figura 5.1(b), en este caso tenemos una distribución asimétrica y también vemos que si consideramos contaminaciones más cerca de las colas los estimadores robustos proporcionan mejores resultados en términos del error cuadrático medio.

En general, ante la presencia de valores atípicos, es mejor elegir el MI estimador o el MDP estimador ya que en términos de error cuadrático medio vimos que para valores atípicos muy cercanos a colas con poca probabilidad este es mas pequeño que el error cuadrático medio del estimador de máxima verosimilitud, y si consideramos valores atípicos en las zonas donde hay mayor densidad la diferencia entre el error cuadrático medio del MI estimador y el MDP estimador frente al estimador de máxima verosimilitud no es muy grande.

5.3. Cubrimiento empírico

La teoría de la estimación de parámetros mediante el estimador de máxima verosimilitud, el MI estimador y el MDP estimador nos proporcionan las herramientas suficientes para construir intervalos de confianza en nuestras estimaciones. Si fijamos un nivel de confianza $1 - \alpha$, uno esperaría que si repetimos el experimento de estimar los parámetros muchas veces cerca del $(1 - \alpha)100\%$ de esas veces el verdadero valor del parámetro que se está estimando se encuentre dentro de un intervalo cuya forma es

$$\hat{\theta} \pm z_{\alpha/2} \widehat{SE}(\hat{\theta}).$$

donde $\hat{\theta}$ es el estimador puntual del parámetro θ , $z_{\alpha/2}$ es el cuantil de la distribución normal estándar que deja un área de $\alpha/2$ a la derecha y $\widehat{SE}(\hat{\theta})$ es el error estándar estimador del estimador. Una forma de evaluar qué tanto se está alcanzando este nivel de confianza es mediante el cubrimiento empírico de los intervalos construidos, esto es, repetir muchas veces el experimento de la estimación mediante intervalos y contar que proporción de veces el intervalo contiene el parámetro que se está pretendiendo estimar.

Los siguientes experimentos se realizaron considerando los modelos de la distribución *Beta* que se muestran en la Figura 5.1 considerando porcentajes de contaminación del 1 %, 3 %, 5 %, 7 % y 10 %. Los valores de la contaminación son elegidos de acuerdo a la forma de la densidad. Las simulaciones se realizaron generando en R muestras de tamaño 100 y luego sustituyendo un porcentaje de los datos simulados por la contaminación se construyen los intervalos de confianza. Esto se repite 1000 veces y luego verificamos las veces que el intervalo contiene el verdadero parámetro.

En el primer caso consideramos una distribución beta con $a = b = 7$, aquí tenemos una distribución simétrica como se puede apreciar en la Figura 5.1(a) donde el 0.99 % de la densidad acumulada entre 0.2 y 0.8, por esto elegimos contaminaciones en $x_0 = 0.1$, $x_0 = 0.5$ y $x_0 = 0.9$, el primero y el último casos en las colas de la distribución que tiene poca probabilidad y en $x_0 = 0.5$ el valor en el centro que esperaríamos que no se vean muy afectados los estimadores en ninguno de los casos.

En la Tabla 5.3 observamos los resultados, como era de esperarse para la contaminación en $x_0 = 0.5$ no encontramos efectos notables en la estimación de máxima verosimilitud en comparación con los estimadores robustos. Para las contaminaciones en $x_0 = 0.1$ y $x_0 = 0.9$ vemos una diferencia notable a partir del 3 % de contaminación donde los estimadores robustos dan mejores resultados que los de máxima verosimilitud, a partir del 5 % el nivel de cubrimiento empírico del estimador de máxima verosimilitud es prácticamente nulo. En comparación de los estimadores robustos, notamos que los niveles de cubrimiento empírico del MI estimador están más cerca del nivel nominal cuando $x_0 = 0.1$ y $x_0 = 0.9$ que los niveles de cubrimiento empírico del MDP estimador.

Tabla 5.3: Cubrimiento empírico del 95 % para los estimadores bajo contaminación para $a = b = 7$.

Estimador	x_0	Porcentaje de contaminación											
		1 %		3 %		5 %		7 %		10 %			
		a	b	a	b	a	b	a	b	a	b	a	b
Máxima verosimilitud		0.855	0.883	0.183	0.400	0.000	0.036	0.000	0.000	0.000	0.000	0.000	0.000
MI estimador	0.1	0.942	0.945	0.866	0.895	0.724	0.789	0.524	0.662	0.193	0.348		
MDP estimador $\alpha = 0.3$		0.930	0.937	0.784	0.822	0.501	0.595	0.192	0.286	0.008	0.017		
Máxima verosimilitud		a	b	a	b	a	b	a	b	a	b	a	b
MI estimador	0.5	0.963	0.959	0.970	0.974	0.967	0.971	0.952	0.960	0.921	0.927		
MDP estimador $\alpha = 0.3$		0.964	0.969	0.966	0.965	0.954	0.959	0.938	0.940	0.893	0.897		
Máxima verosimilitud		a	b	a	b	a	b	a	b	a	b	a	b
MI estimador	0.9	0.942	0.939	0.839	0.859	0.800	0.720	0.648	0.505	0.360	0.188		
MDP estimador $\alpha = 0.3$		0.938	0.934	0.833	0.775	0.588	0.485	0.286	0.180	0.020	0.010		

Tabla 5.4: Cubrimiento empírico del 95 % para los estimadores bajo contaminación para $a = 2, b = 6$.

Estimador	x_0	Porcentaje de contaminación											
		1 %		3 %		5 %		7 %		10 %			
		a	b	a	b	a	b	a	b	a	b	a	b
Máxima verosimilitud		0.965	0.950	0.969	0.937	0.969	0.919	0.967	0.893	0.964	0.819		
MI estimador	0.5	0.959	0.947	0.941	0.912	0.920	0.842	0.905	0.751	0.869	0.569		
MDP estimador $\alpha = 0.3$		0.965	0.953	0.960	0.935	0.951	0.895	0.948	0.833	0.931	0.705		
Máxima verosimilitud		a	b	a	b	a	b	a	b	a	b	a	b
MI estimador	0.7	0.952	0.935	0.913	0.862	0.839	0.706	0.726	0.511	0.504	0.185		
MDP estimador $\alpha = 0.3$		0.957	0.938	0.913	0.832	0.827	0.624	0.705	0.346	0.470	0.048		
Máxima verosimilitud		a	b	a	b	a	b	a	b	a	b	a	b
MI estimador	0.9	0.951	0.935	0.910	0.860	0.825	0.703	0.699	0.507	0.457	0.183		
MDP estimador $\alpha = 0.3$		0.954	0.937	0.902	0.850	0.794	0.654	0.545	0.375	0.184	0.087		

Analicemos ahora el caso de $a = 2$ y $b = 6$, en este caso tenemos una distribución asimétrica hacia la derecha lo que indica que la mayoría de valores se van a concentrar a la izquierda como apreciamos en la Figura 5.1(b) hasta $x = 0.6$ se acumula una probabilidad aproximada de 0.98, por tanto obtener valores mayores que 0.6 es cada vez menos probable cuanto mas nos acercamos a 1. Por lo anterior en este modelo se consideraron 3 valores con los que se contamina la muestra $x_0 = 0.5$, $x_0 = 0.7$ y $x_0 = 0.9$. En la Tabla 5.4 se muestran los resultados de los cubrimientos empíricos calculados a partir de las simulaciones. Para el valor de contaminación $x_0 = 0.5$ notamos que no existen demasiadas diferencias en los cubrimientos empíricos entre los estimadores robustos y el de máxima verosimilitud, esto era de esperarse ya que este valor se encuentra dónde esta la mayor concentración de datos, para $x_0 = 0.7$ y $x_0 = 0.9$ empezamos a notar mejores resultados con los estimadores robustos a partir de una contaminación del 3 %, cuando el porcentaje de contaminación aumenta el estimador de máxima verosimilitud se ve acercando mas a cero.

En general podemos decir que el cubrimiento empírico de los estimadores robustos es mejor que el de máxima verosimilitud al considerar contaminaciones en las zonas donde hay menor densidad de los datos, mientras que en las zonas donde se acumula la mayor densidad estas contaminaciones no afectan de manera muy significativa el estimador de máxima verosimilitud y los estimadores robustos. También notamos que los niveles de cubrimiento empírico del MI estimador son mas cercanos al nivel nominal que los niveles de cubrimiento empírico del MDP estimador.

5.4. Aplicaciones

5.4.1. Datos estrés

Se realizó un estudio psicológico basado en 166 mujeres no clínicas en Australia para investigar los efectos del estrés, la ansiedad y la depresión. Se pueden encontrar los detalles en la sección 4.3 de Ghosh (2019). Los autores señalaron que hay una serie de observaciones que podrían influir en las estimaciones. Los datos son puntuaciones de estrés y ansiedad, medidos como una tasa, es decir toman valores entre 0 y 1 por lo tanto asumir una distribución *Beta* para modelar los datos podría ser válido.

Consideramos dos grupos, las mujeres que tienen puntuación de ansiedad mayor que 0.01 y las que tienen puntuación de ansiedad menor o igual a 0.01, a continuación se muestra un gráfico de dispersión y la densidad estimada para cada grupo.

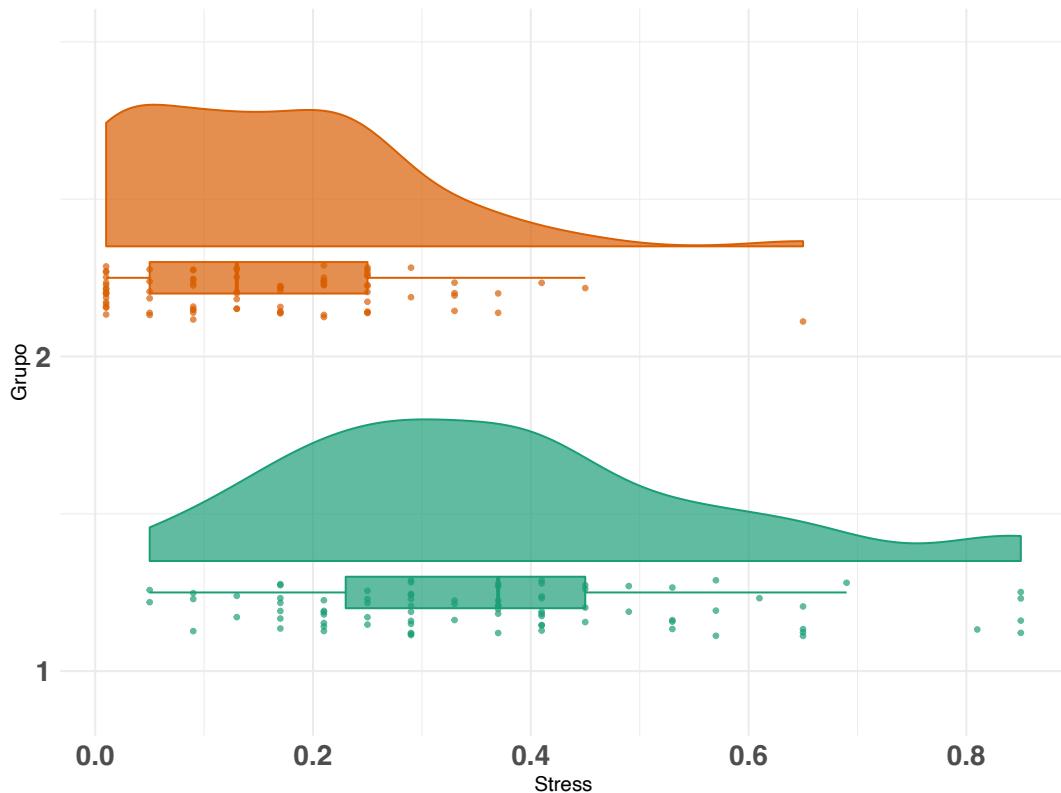


Figura 5.4: Gráficos de dispersión para el grupo 1: mujeres con puntuación de ansiedad > 0.01 y grupo 2: mujeres con puntuación de ansiedad ≤ 0.01 .

En el grupo de mujeres con ansiedad mayor a 0.01 parece tener una distribución *Beta*, además al final hay algunas observaciones que podrían considerarse valores atípicos,

ajustamos los parámetros usando el estimador de máxima verosimilitud, el MI estimador y el MDP estimador los resultados se muestran en la Tabla 5.5.

Tabla 5.5: Estimaciones puntuales e intervalos de confianza de los parámetros de la distribución para el grupo de mujeres con ansiedad mayor a 0.01.

Estimador	Parámetro	
	a	b
MV	2.0635 (1.4735, 2.6535)	3.4128 (2.3902, 4.4354)
MI	2.3698 (1.5239, 3.2158)	4.1543 (2.4685, 5.8401)
MDP	2.2225 (1.5501, 2.8949)	3.8717 (2.6288, 5.1145)

Podemos notar que las estimaciones para los parámetros a y b son mayores en los estimadores robustos que en el estimador de máxima verosimilitud. La verdadera distribución que podría modelar los datos es desconocida, sin embargo si graficamos las densidades generadas a partir de los parámetros estimados notamos en la Figura 5.5 que las curvas que generan los estimadores robustos parecen ajustarse mejor a la densidad empírica que la que es generada por el estimador de máxima verosimilitud.

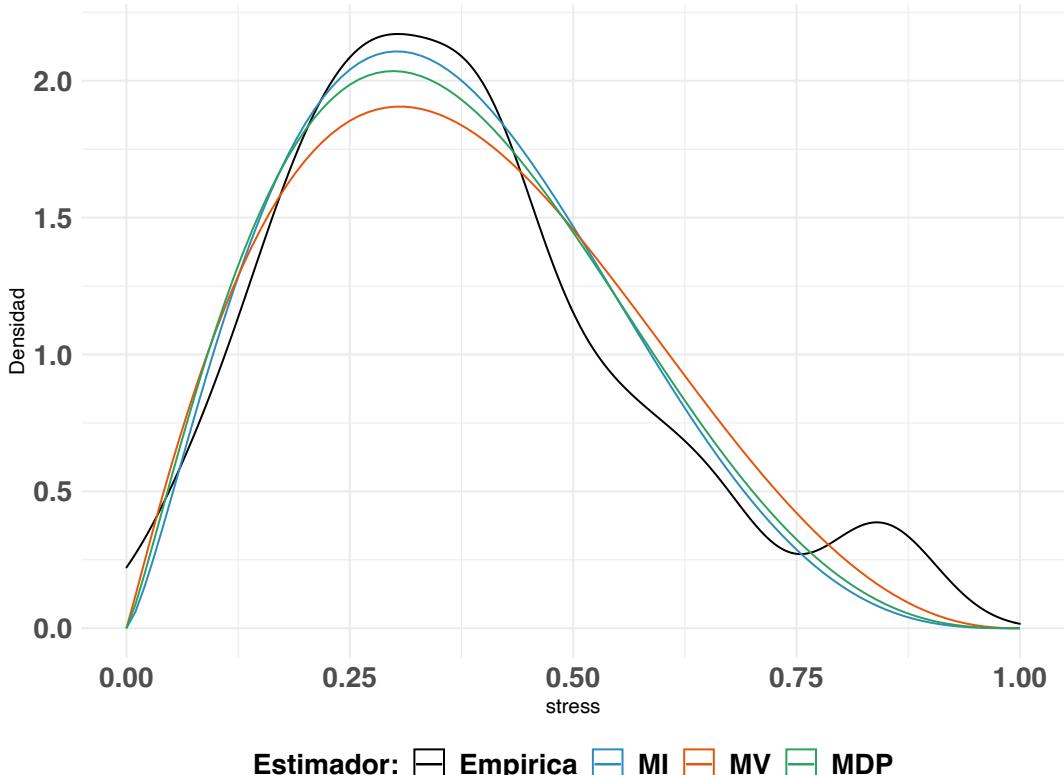


Figura 5.5: Funciones de densidad graficadas a partir de los parámetros estimados.

5.4.2. Índices de diversidad

En ecología un índice de diversidad es una medida cuantitativa que refleja cuántos tipos diferentes (como especies) hay en un conjunto de datos (una comunidad), y que simultáneamente puede tener en cuenta las relaciones filogenéticas entre los individuos distribuidos entre esos tipos, como riqueza, divergencia o uniformidad; ver Tucker et al. (2017). Estos índices son representaciones estadísticas de la biodiversidad en diferentes aspectos (riqueza, uniformidad y dominancia).

El índice de Shannon es uno de los más utilizados para cuantificar la biodiversidad específica. Este índice se basa en el número de especies presentes y su abundancia relativa para cuantificar la heterogeneidad de una comunidad. Su fórmula de cálculo es

$$H = - \sum_{i=1}^S p_i \log p_i.$$

dónde p_i es la proporción de individuos que pertenecen a la i -esima especie en el conjunto de datos de interés y S es el número total de especies. En la mayoría de los ecosistemas naturales el índice de Shannon varía entre 0.5 y 5, aunque su valor normal está entre 2 y 3; valores inferiores a 2 se consideran bajos en diversidad y superiores a 3 son altos en diversidad de especies.

El índice de Pielou estandariza el índice de Shannon en una escala de 0 a 1, esto describe que tan homogéneas son las abundancias relativas de todas las especies en una comunidad su cálculo es de la forma:

$$J = \frac{H}{\log S}.$$

Se tienen datos de una parcela permanente de 50 hectáreas, establecida por el Instituto Smithsonian de Investigaciones Tropicales y la Universidad de Princeton para estudiar la dinámica de la vegetación de los bosques tropicales. El conjunto de datos disponible contiene 50 parcelas, que se alinean entre sí. En cada parcela se registró el número de individuos de todas las especies del bosque y varias variables ambientales. Este conjunto de datos se encuentra en el dataset BCI del paquete vegan de R; ver Oksanen et al. (2013).

Se calcula el índice de Pielou para este conjunto de datos en la Figura 5.6 podemos notar la función de densidad estimada a partir de la muestra.

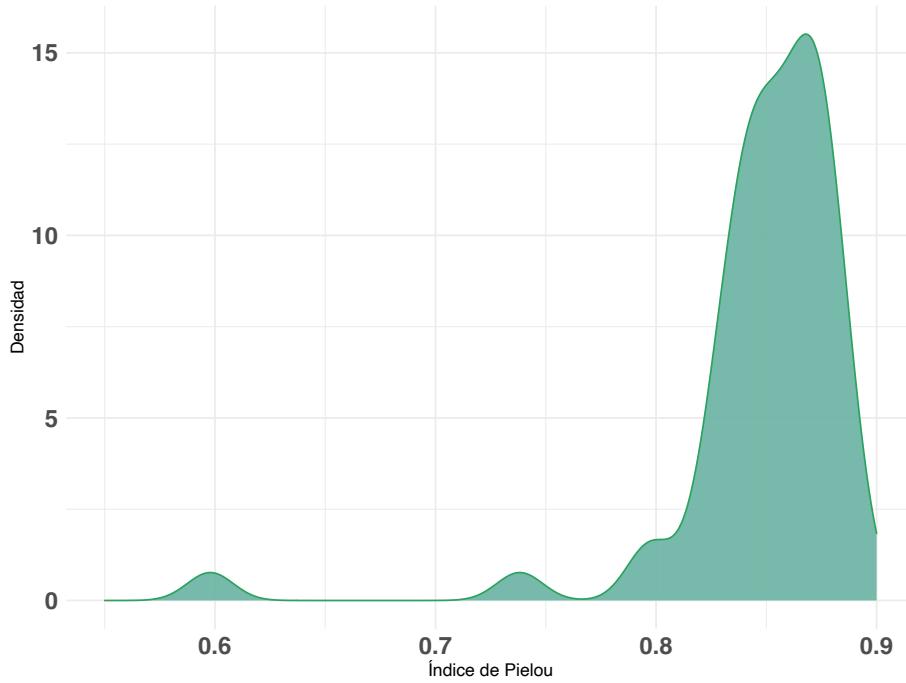


Figura 5.6: Función de densidad del índice de Pielou de la muestra.

Podemos notar que los datos parecen seguir una distribución *Beta* con algunos valores atípicos, hacemos el ajuste de la distribución usando los 3 estimadores y superponemos las distribuciones a partir de los parámetros estimados.

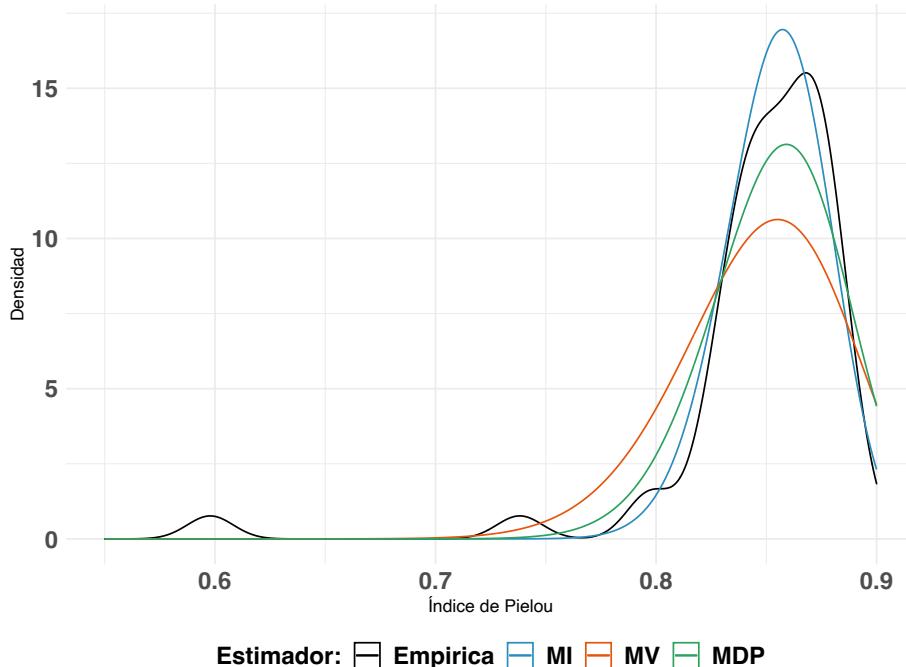


Figura 5.7: Funciones de densidad graficadas a partir de los parámetros estimados.

En la Figura 5.7 podemos notar cómo las distribuciones dadas por los estimadores robustos se ajustan mejor a la forma de la densidad empírica. Podemos usar el ajuste realizado con los estimadores robustos para detectar cuáles valores son atípicos en la muestra para esto estimamos la probabilidad de obtener una observación menor o igual a cada observación de la muestra, luego podemos descartar aquellas observaciones cuya probabilidad acumulada sea muy chica.

Tabla 5.6: Probabilidades estimadas a partir de la distribución generada por los estimadores robustos y de máxima verosimilitud.

Obs	x	MI			MDP			MV			Obs	x	MI			MDP		
		$P(X \leq x)$			$P(X \leq x)$													
1	0.8866	0.9222	0.8615	0.8525	26	0.8752	0.8108	0.7489	0.7593									
2	0.8686	0.7196	0.6700	0.6962	27	0.8662	0.6830	0.6401	0.6724									
3	0.8476	0.3758	0.4001	0.4769	28	0.8315	0.1678	0.2261	0.3195									
4	0.8753	0.8120	0.7499	0.7601	29	0.8281	0.1372	0.1972	0.2906									
5	0.8602	0.5848	0.5622	0.6102	30	0.8419	0.2919	0.3331	0.4188									
6	0.8501	0.4151	0.4308	0.5028	31	0.8575	0.5397	0.5272	0.5820									
7	0.8707	0.7505	0.6960	0.7169	32	0.8453	0.3411	0.3728	0.4535									
8	0.8729	0.7818	0.7230	0.7385	33	0.8397	0.2621	0.3085	0.3968									
9	0.8359	0.2150	0.2684	0.3600	34	0.8452	0.3385	0.3707	0.4517									
10	0.8562	0.5164	0.5092	0.5674	35	0.5979	0.0000*	0.0000*	0.0000*									
11	0.8643	0.6524	0.6155	0.6528	36	0.8506	0.4232	0.4371	0.5081									
12	0.8347	0.2016	0.2567	0.3489	37	0.8469	0.3643	0.3911	0.4692									
13	0.8786	0.8510	0.7864	0.7898	38	0.7979	0.0135*	0.0431	0.1033									
14	0.8762	0.8238	0.7608	0.7689	39	0.7968	0.0122*	0.0405	0.0989									
15	0.8729	0.7818	0.7230	0.7385	40	0.7382	0.0000*	0.0006*	0.0056*									
16	0.8641	0.6501	0.6137	0.6513	41	0.8762	0.8237	0.7607	0.7688									
17	0.8244	0.1087	0.1684	0.2607	42	0.8882	0.9333	0.8747	0.8640									
18	0.8789	0.8540	0.7893	0.7922	43	0.8388	0.2502	0.2985	0.3878									
19	0.8554	0.5039	0.4996	0.5595	44	0.8431	0.3085	0.3466	0.4307									
20	0.8854	0.9136	0.8515	0.8440	45	0.8214	0.0885	0.1468	0.2372									
21	0.8639	0.6469	0.6111	0.6492	46	0.8555	0.5044	0.4999	0.5598									
22	0.8325	0.1784	0.2359	0.3291	47	0.8478	0.3784	0.4022	0.4786									
23	0.8841	0.9035	0.8402	0.8344	48	0.8676	0.7052	0.6582	0.6868									
24	0.8739	0.7941	0.7338	0.7472	49	0.8377	0.2369	0.2873	0.3775									
25	0.8755	0.8154	0.7530	0.7627	50	0.8619	0.6131	0.5843	0.6279									

Si consideramos outliers a las observaciones con cuantil menor o igual a $1/n = 1/50 = 0.02$ entonces el MI detecta 4 outliers (obs 35, 38, 39 y 40) mientras que el MDP y el MV solo detectan 2 (obs 35 y 40).

Apéndice A

Demostraciones del capítulo 4

Elementos de la matriz \mathbf{B} para el MI estimador

Consideremos el elemento \mathbf{B}_{11} , primero calculemos $\frac{\partial}{\partial a} \psi_1^I(X, \boldsymbol{\theta})$, usando las derivadas logarítmicas de la función *Beta* estudiadas en la sección 2.2 tenemos lo siguiente:

$$\begin{aligned}\frac{\partial}{\partial a} \psi_1^I(X, \boldsymbol{\theta}) &= \frac{\partial}{\partial a} F_{\boldsymbol{\theta}}(X) \\ &= \frac{\partial}{\partial a} \left[\mathcal{B}^{-1}(a, b) \int_0^X t^{a-1} (1-t)^{b-1} dt \right] \\ &= -\mathcal{B}^{-1}(a, b) (\psi^{(0)}(a) - \psi^{(0)}(a+b)) \int_0^X t^{a-1} (1-t)^{b-1} dt \\ &\quad + \mathcal{B}^{-1}(a, b) \int_0^X \log(t) t^{a-1} (1-t)^{b-1} dt \\ &= (\psi^{(0)}(a+b) - \psi^{(0)}(a)) F_{\boldsymbol{\theta}}(X) \\ &\quad + \mathcal{B}^{-1}(a, b) \int_0^X \log(t) t^{a-1} (1-t)^{b-1} dt.\end{aligned}$$

Es fácil ver que $|\log(t)t^{a-1}(1-t)^{b-1}|$ puede acotarse por una función integrable en $[0, 1]$, luego la derivación bajo el signo integral puede justificarse usando el teorema 10.39 de Apostol (1981).

Ahora considerando la esperanza

$$\begin{aligned}E_{\boldsymbol{\theta}} \left(\frac{\partial}{\partial a} \psi_1^I(X, \boldsymbol{\theta}_0) \right) &= \frac{1}{2} (\psi^{(0)}(a+b) - \psi^{(0)}(a)) \\ &\quad + \mathcal{B}^{-1}(a, b) E_{\boldsymbol{\theta}} \left(\int_0^X \log(t) t^{a-1} (1-t)^{b-1} dt \right).\end{aligned}\tag{A.1}$$

Notemos que la esperanza en (A.1) es finita ya que es menor o igual que $\mathcal{B}(a, b) E_{\boldsymbol{\theta}}(\log(Z))$, con $Z \sim Beta(a, b)$, calculada en el Teorema 2.3.

Luego con cálculos similares a los anteriores podemos encontrar el elemento \mathbf{B}_{12} .

$$\begin{aligned}E_{\boldsymbol{\theta}} \left(\frac{\partial}{\partial b} \psi_1^I(X, \boldsymbol{\theta}_0) \right) &= \frac{1}{2} (\psi^{(0)}(a+b) - \psi^{(0)}(b)) \\ &\quad + \mathcal{B}^{-1}(a, b) E_{\boldsymbol{\theta}} \left(\int_0^X \log(1-t) t^{a-1} (1-t)^{b-1} dt \right).\end{aligned}\tag{A.2}$$

Ahora veamos los elementos de la segunda fila, primero veamos el \mathbf{B}_{21}

$$\begin{aligned}\frac{\partial}{\partial a} \psi_2^I(X, \boldsymbol{\theta}) &= \frac{\partial}{\partial a} F_{\boldsymbol{\theta}}^2(X) \\ &= 2F_{\boldsymbol{\theta}}(X) \frac{\partial}{\partial a} F_{\boldsymbol{\theta}}(X) \\ &= 2(\psi^{(0)}(a+b) - \psi^{(0)}(a))F_{\boldsymbol{\theta}}^2(X) \\ &\quad + 2F_{\boldsymbol{\theta}}(X)\mathcal{B}^{-1}(a,b) \int_0^X \log(t)t^{a-1}(1-t)^{b-1}dt.\end{aligned}$$

Ahora tomando la esperanza tenemos que

$$\begin{aligned}E_{\boldsymbol{\theta}}\left(\frac{\partial}{\partial a} \psi_2^I(X, \boldsymbol{\theta})\right) &= \frac{2}{3}(\psi^{(0)}(a+b) - \psi^{(0)}(a)) \\ &\quad + 2\mathcal{B}^{-1}(a,b)E_{\boldsymbol{\theta}}\left(F_{\boldsymbol{\theta}}(X) \int_0^X \log(t)t^{a-1}(1-t)^{b-1}dt\right).\end{aligned}\quad (\text{A.3})$$

Por último el elemento \mathbf{B}_{22} es

$$\begin{aligned}E_{\boldsymbol{\theta}}\left(\frac{\partial}{\partial b} \psi_2^I(X, \boldsymbol{\theta})\right) &= \frac{2}{3}(\psi^{(0)}(a+b) - \psi^{(0)}(b)) \\ &\quad + 2\mathcal{B}^{-1}(a,b)E_{\boldsymbol{\theta}}\left(F_{\boldsymbol{\theta}}(X) \int_0^X \log(1-t)t^{a-1}(1-t)^{b-1}dt\right).\end{aligned}\quad (\text{A.4})$$

Las esperanzas que resultan en las Ecuaciones (A.1)–(A.4) no se pueden encontrar analíticamente y pueden ser aproximadas usando la media muestral.

Integrales para el MDP estimador

$$\int f_{\boldsymbol{\theta}}^{1+\alpha}(x)dx = \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]^{1+\alpha} \int_0^1 x^{(1+\alpha)(a-1)}(1-x)^{(1+\alpha)(b-1)}dx.$$

El término que queda adentro de la integral es el núcleo de una distribución $Beta(\alpha a - \alpha + a, \alpha b - \alpha + b)$ completamos las constantes para que la integral resultante sea 1 y luego

$$\int f_{\boldsymbol{\theta}}^{1+\alpha}(x)dx = \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)}.\quad (\text{A.5})$$

Consideremos ahora

$$\int \log(x)f_{\boldsymbol{\theta}}^{1+\alpha}(x)dx = \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]^{1+\alpha} \int_0^1 \log(x)x^{(1+\alpha)(a-1)}(1-x)^{(1+\alpha)(b-1)}dx.$$

de nuevo tenemos el núcleo de una $Beta(\alpha a - \alpha + a, ab - \alpha + b)$ junto al término $\log(x)$.

Completando las constantes tenemos

$$\int \log(x) f_{\theta}^{1+\alpha}(x) dx = \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(ab - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} E[\log(Z)].$$

Con $Z \sim Beta(\alpha a - \alpha + a, ab - \alpha + b)$, luego por el Teorema 2.3 tenemos que

$$E[\log Z] = \psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b). \quad (\text{A.6})$$

$$E[\log(1 - Z)] = \psi^{(0)}(ab - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b). \quad (\text{A.7})$$

$$\begin{aligned} E[\log^2 Z] &= [\psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)]^2 \\ &\quad + \psi^{(1)}(\alpha a - \alpha + a) - \psi^{(1)}(\alpha a + \alpha b - 2\alpha + a + b). \end{aligned} \quad (\text{A.8})$$

$$\begin{aligned} E[\log^2(1 - Z)] &= [\psi^{(0)}(ab - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)]^2 \\ &\quad + \psi^{(1)}(ab - \alpha + b) - \psi^{(1)}(\alpha a + \alpha b - 2\alpha + a + b). \end{aligned} \quad (\text{A.9})$$

$$\begin{aligned} E[\log Z \log(1 - Z)] &= [\psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)] \\ &\quad \times [\psi^{(0)}(ab - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)] \\ &\quad - \psi^{(1)}(\alpha a + \alpha b - 2\alpha + a + b). \end{aligned} \quad (\text{A.10})$$

Luego , considerando (A.6) tenemos

$$\begin{aligned} \int \log(x) f_{\theta}^{1+\alpha}(x) dx &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(ab - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\ &\quad \times (\psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)). \end{aligned} \quad (\text{A.11})$$

un procedimiento similar y considerando (A.7), (A.8), (A.9) y (A.10) respectivamente podemos probar las integrales.

$$\begin{aligned} \int \log(1 - x) f_{\theta}^{1+\alpha}(x) dx &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(ab - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\ &\quad \times (\psi^{(0)}(ab - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)). \end{aligned} \quad (\text{A.12})$$

$$\begin{aligned}
 \int \log^2(x) f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\
 &\quad \times [\psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)]^2 \\
 &\quad + \psi^{(1)}(\alpha a - \alpha + a) - \psi^{(1)}(\alpha a + \alpha b - 2\alpha + a + b). \tag{A.13}
 \end{aligned}$$

$$\begin{aligned}
 \int \log^2(1-x) f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\
 &\quad \times [\psi^{(0)}(\alpha b - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)]^2 \\
 &\quad + \psi^{(1)}(\alpha b - \alpha + b) - \psi^{(1)}(\alpha a + \alpha b - 2\alpha + a + b). \tag{A.14}
 \end{aligned}$$

$$\begin{aligned}
 \int \log(x) \log(1-x) f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \\
 &\quad \times [\psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)] \\
 &\quad \times [\psi^{(0)}(\alpha b - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)] \\
 &\quad - \psi^{(1)}(\alpha a + \alpha b - 2\alpha + a + b). \tag{A.15}
 \end{aligned}$$

Matrices J , K y ξ para el MDP estimador

$$\begin{aligned}
 J_{11} &= \int \left[\frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial a} \right]^2 f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx \\
 &= \int [\psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x)]^2 f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx \\
 &= [\psi^{(0)}(a+b) - \psi^{(0)}(a)]^2 \int f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx + 2[\psi^{(0)}(a+b) - \psi^{(0)}(a)] \int \log(x) f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx \\
 &\quad + \int \log^2(x) f_{\boldsymbol{\theta}}^{1+\alpha}(x) dx.
 \end{aligned}$$

La última expresión está en términos de las integrales (A.5), (A.11) y (A.13) por lo que podemos encontrar de forma cerrada una expresión para J_{11}

$$\begin{aligned}
 J_{11} &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \left[\frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \right] [(\psi^{(0)}(a+b) - \psi^{(0)}(a))^2 \\
 &\quad + 2(\psi^{(0)}(a+b) - \psi^{(0)}(a))(\psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)) \\
 &\quad + (\psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b))^2 \\
 &\quad + \psi^{(1)}(\alpha a - \alpha + a) - \psi^{(1)}(\alpha a + \alpha b - 2\alpha + a + b)].
 \end{aligned}$$

Con cálculos similares y los resultados (A.12), (A.14) y (A.15) podemos encontrar J_{22}

y $J_{12} = J_{21}$.

$$\begin{aligned}
 J_{22} &= \int \left[\frac{\partial \log f_{\theta}(x)}{\partial b} \right]^2 f_{\theta}^{1+\alpha}(x) dx \\
 &= \int [\psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x)]^2 f_{\theta}^{1+\alpha}(x) dx \\
 &= [\psi^{(0)}(a+b) - \psi^{(0)}(b)]^2 \int f_{\theta}^{1+\alpha}(x) dx + 2 [\psi^{(0)}(a+b) - \psi^{(0)}(b)] \int \log(1-x) f_{\theta}^{1+\alpha}(x) dx \\
 &\quad + \int \log^2(1-x) f_{\theta}^{1+\alpha}(x) dx \\
 &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \left[\frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \right] [(\psi^{(0)}(a+b) - \psi^{(0)}(b))^2 \\
 &\quad + 2(\psi^{(0)}(a+b) - \psi^{(0)}(b))(\psi^{(0)}(\alpha b - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)) \\
 &\quad + (\psi^{(0)}(\alpha b - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b))^2 \\
 &\quad + \psi^{(1)}(\alpha b - \alpha + b) - \psi^{(1)}(\alpha a + \alpha b - 2\alpha + a + b)].
 \end{aligned}$$

$$\begin{aligned}
 J_{12} = J_{21} &= \int \left[\frac{\partial \log f_{\theta}(x)}{\partial a} \frac{\partial \log f_{\theta}(x)}{\partial b} \right] f_{\theta}^{1+\alpha}(x) dx \\
 &= \int [\psi^{(0)}(a+b) - \psi^{(0)}(a) + \log(x)] [\psi^{(0)}(a+b) - \psi^{(0)}(b) + \log(1-x)] f_{\theta}^{1+\alpha}(x) dx \\
 &= [\psi^{(0)}(a+b) - \psi^{(0)}(a)] [\psi^{(0)}(a+b) - \psi^{(0)}(b)] \int f_{\theta}^{1+\alpha}(x) dx \\
 &\quad + [\psi^{(0)}(a+b) - \psi^{(0)}(a)] \int \log(1-x) f_{\theta}^{1+\alpha}(x) dx \\
 &\quad + [\psi^{(0)}(a+b) - \psi^{(0)}(b)] \int \log(x) f_{\theta}^{1+\alpha}(x) dx \\
 &\quad + \int \log(x) \log(1-x) f_{\theta}^{1+\alpha}(x) dx \\
 &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^{1+\alpha} \left[\frac{\Gamma(\alpha a - \alpha + a)\Gamma(\alpha b - \alpha + b)}{\Gamma(\alpha a + \alpha b - 2\alpha + a + b)} \right] \\
 &\quad \times [(\psi^{(0)}(a+b) - \psi^{(0)}(a))(\psi^{(0)}(a+b) - \psi^{(0)}(b)) \\
 &\quad + (\psi^{(0)}(a+b) - \psi^{(0)}(a))(\psi^{(0)}(\alpha b - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)) \\
 &\quad + (\psi^{(0)}(a+b) - \psi^{(0)}(b))(\psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)) \\
 &\quad + (\psi^{(0)}(\alpha a - \alpha + a) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)) \\
 &\quad \times (\psi^{(0)}(\alpha b - \alpha + b) - \psi^{(0)}(\alpha a + \alpha b - 2\alpha + a + b)) \\
 &\quad - \psi^{(1)}(\alpha a + \alpha b - 2\alpha + a + b)].
 \end{aligned}$$

Bibliografía

- Alexander, F. G. and Mood, D. B. (1969). *Introduccion a la teoria de la Estadistica.* Introduction to the theory of statistics. Aguilar, not stated edition.
- Ambrosio, B. (2004). *Algoritmos rápidos para computar estimadores robustos.* PhD thesis, Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales.
- Apostol, T. M. (1981). *Mathematical Analysis.* Word Student Series. Addison-Wesley, 2nd edition.
- Artin, E. (1964). *The gamma function.* Athena series; selected topics in mathematics. Holt, Rinehart and Winston, 1st edition.
- Basu, A., Harris, I. R., Hjort, N. L., and Jones, M. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559.
- Boik, R. J. and Robinson-Cox, J. F. (1999). Derivatives of the incomplete beta function. *Journal of Statistical Software*, 3:1–20.
- Boos, D. D., Stefanski, L. A., et al. (2013). *Essential statistical inference.* Springer.
- Fegyverneki, S. (2003). Robust estimators and probability integral transformations. *Mathematical and computer modelling*, 38(7-9):803–814.
- Ferguson, T. S. (1967). *Mathematical statistics a decision theoretic approach.* Probability and mathematical statistics; a series of monographs and textbooks 1. Academic Press.
- Fielitz, B. D. and Myers, B. L. (1975). Concepts, theory, and techniques: Estimation of parameters in the beta distribution. *Decision Sciences*, 6(1):1–13.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian data analysis.* Texts in statistical science. Chapman & Hall/CRC, 2nd ed edition.
- Gerow, K. et al. (1989). Fisher consistency-the evolution of a concept: It's hard to get it right the first time.

- Ghosh, A. (2019). Robust inference under the beta regression model with application to health care studies. *Statistical Methods in Medical Research*, 28(3):871–888.
- Gupta, A. K. and Nadarajah, S. (2004). *Handbook of Beta Distribution and Its Applications*. CRC Press, 1 edition.
- Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the american statistical association*, 69(346):383–393.
- Hogg, R. V., McKean, J. W., and Craig, A. T. (2019). *Introduction to mathematical statistics*. Pearson.
- Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Weather Modification; University of California Press: Berkeley, CA, USA*, page 221.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous univariate distributions, volume 2*, volume 289. John wiley & sons.
- Maronna, R. A., Martin, R. D., Yohai, V. J., and Salibián-Barrera, M. (2019). *Robust statistics: theory and methods (with R)*. John Wiley & Sons.
- Oksanen, J. et al. (2013). Vegan: ecological diversity. *R project*, 368:1–11.
- Owen, C. B. (2008). *Parameter estimation for the beta distribution*. Brigham Young University.
- Pak, R. J. (2014). The minimum density power divergence estimation for the lognormal density. *Communications in Statistics-Theory and Methods*, 43(21):4582–4588.
- Qi, F. (2020). Some properties of several functions involving polygamma functions and originating from the sectional curvature of the beta manifold. *Sao Paulo Journal of Mathematical Sciences*, 14(2):614–630.

BIBLIOGRAFÍA

- Ronchetti, E. M. and Huber, P. J. (2009). *Robust statistics*. John Wiley & Sons.
- Stigler, S. M. (1973). Simon newcomb, percy daniell, and the history of robust estimation 1885–1920. *Journal of the American Statistical Association*, 68(344):872–879.
- Thieler, A. M., Fried, R., and Rathjens, J. (2016). Robper: An r package to calculate periodograms for light curves based on robust regression. *Journal of Statistical Software*, 69:1–37.
- Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., Grenyer, R., Helmus, M. R., Jin, L. S., Mooers, A. O., et al. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, 92(2):698–715.
- Valdora, M. and Yohai, V. (2020). M estimators based on the probability integral transformation with applications to count data. *Statistics & Probability Letters*, 162:108751.
- van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- Wackerly, D. D., Mendenhall III, W., and Scheaffer, R. L. (2008). *Estadística matemática con aplicaciones*. México: Cengage Learning.