CIMAT

CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS, A.C.

ESTIMACIÓN DE LA FUNCIÓN DE DEPENDENCIA DE LA DISTRIBUCION BIVARIADA DE VALORES EXTREMOS

TESIS

que para obtener el grado de

DOCTOR EN CIENCIAS

presenta

Enrique Raúl Villa Diharce

Guanajuato, Gto., Mayo de 1999

A

Malú, Enrique, Alejandra y Emilio,

que son

una motivación para seguir adelante.

CIMAT
BIBLIOTECA

CIMAT
BIBLIOTECA

015963

AGRADECIMIENTOS

Deseo agradecer al Dr. Victor Pérez Abreu Carrión por su gran apoyo desde el inicio de este trabajo al proponerme el tema de investigación y brindarme condiciones para la realización del mismo, así como por sus valiosos comentarios para el desarrollo del capítulo dos. Al Dr. Javier Rojo por haber llegado justo en el momento preciso para otorgarme su apoyo atinado y afectuoso como asesor de tesis. Su colaboración, opiniones y consejos, le dieron al presente trabajo su forma definitiva. A el Dr. Miguel Nakamura Savoy. por la atención que me brindó desde el inicio del trabajo, y que con sus comentarios concisos y profundos me dió la oportunidad de repensar algunas ideas. A el Dr. Luis A. Escobar, por haberse tomado la tarea de hacer una minuciosa revisión de la tesis, haciendome señalamientos de gran valor para mejorar el trabajo. A los Dres. Fernando Avila Murillo. y Graciela Gonzalez Farías, por sus atinados comentarios y por haber fungido como sinodales en mi examen profesional.

Al Conacyt, debo el apoyo brindado a través de la Catedra Patrimonial del Programa de Apoyo a los Profesores e Investigadores para la obtención del Doctorado, y tambien a través del proyecto Métodos Estadísticos para el Monitoreo y Evaluación del Medio Ambiente, 0328P-E9506.

El ambiente intelectual del CIMAT fué una gran motivación durante el desarrollo del trabajo, debido al permanente flujo de investigadores visitantes, el apoyo para asistir a eventos internacionales de relevancia, pero sobre todo por los amigos y colegas que me han acompañado en este camino. Entre estos, debo mencionar a Armando, Eloisa, Roman, Jorge, Batún y Lupita.

Capítulo 1

Introducción

1.1 Algunas áreas de aplicación de extremos multivariados

La variabilidad en los procesos que se presentan en la naturaleza y en la industria es un hecho reconocido que hace necesaria la búsqueda de modelos para su estudio. Frecuentemente las observaciones que resultan de un proceso determinado, se expresan en forma resumida por medio de algunos estadísticos, que dan información acerca de la tendencia central, la dispersión, la asimetría y la curtosis o exceso de la distribución de los datos. De estos estadísticos, el que más atención ha recibido es la media, como medida de tendencia central, que en muchos casos se interpreta como una representación del conjunto de las observaciones.

En algunas áreas del conocimiento como hidrología,monitoreo ambiental, meteorología, ciencia de materiales, e ingeniería, hay situaciones donde el objetivo es modelar los valores extremos de la sucesión de observaciones, debido a que el comportamiento del máximo o mínimo, determina el estado del sistema considerado, tal como sucede en los siguientes ejemplos.

En hidrología son de primordial interés los valores extremos de variables como, pre-

cipitación pluvial, flujo de ríos y volumen de agua en presas, debido al impacto que tienen en las condiciones de vida. La humanidad siempre ha enfrentado las consecuencias de extremos hidrológicos principalmente a través de inundaciones o sequías. El exceso o la falta de agua en ríos, lagunas, presas, etcétera, ha ejercido una presión en las sociedades para investigar y conocer el comportamiento de los extremos en hidrología.

La probabilidad de inundación en una localidad cercana a un río, depende del comportamiento del valor máximo del nivel de éste. Cuando se construye una protección para evitar que el río inunde una ciudad, es importante para su diseño el tener información histórica de la altura máxima anual, con el objeto de determinar la altura de diseño con un margen de seguridad dado. Si se tienen varios puntos en el recorrido de un río, donde se puede producir una inundación, entonces es de interés conocer el comportamiento simultáneo de los niveles máximos de él en los diferentes puntos. En este caso tenemos un problema multivariado de valores extremos.

Las normas de calidad del aire están definidas por medio de valores máximos permisibles de concentración de contaminantes. En las grandes ciudades el ozono se ha convertido en el contaminante ambiental mas frecuente y su control ha representado un gran reto para los responsables de la gestión ambiental. En NRC (1991), se presenta una amplia discusión sobre el problema del ozono en la contaminación del aire en el medio urbano, así como también algunas propuestas para su control.

Usualmente los índices de contaminación anual, se determinan a partir del valor máximo anual o de estadísticas de orden alto. También se han utilizado estadísticas relacionadas con la cola derecha de la distribución de la concentración del contaminante.

Cuando estudiamos el comportamiento del valor extremo de una serie de concentraciones de un contaminante, medidas por una estación de monitoreo ambiental, tenemos un problema univariado de valores extremos, pero si consideramos las mediciones para un contaminante en varias estaciones de una red de monitoreo, entonces tenemos un problema multivariado de valores extremos, en donde es natural suponer dependencia entre ellos.

Es importante conocer la estructura de dependencia de las concentraciones extremas multivariadas, debido a que tiene una relación estrecha con la dinámica de dispersión de los contaminantes. Smith et al. (1989), analizan series de tiempo ambientales, y Joe (1994) presenta modelos para valores extremos multivariados con aplicaciones a datos ambientales.

En el proceso de diseño en ingeniería estructural, las distribuciones de valores extremos juegan un papel importante debido a que los extremos de cargas y resistencias determinan el riesgo de falla de una estructura. El análisis de las estructuras expuestas a cargas de vientos, se lleva a cabo considerando la velocidad máxima anual de viento como representativa de los vientos que ocurren durante el año. En el caso de las cargas inducidas por temblores, en términos de velocidad y aceleración del suelo, la modelación también se lleva a cabo considerando valores extremos.

En una presa, el esfuerzo que soporta la cortina depende del volumen de lluvia que se recibe en la región de captura que alimenta la presa. En algunas estructuras marinas la falla depende del comportamiento extremo de vientos y olas. Una amplia discusión de métodos estadísticos para problemas de valores extremos multivariados en diseño estructural se encuentran en Coles y Tawn (1994).

La temperatura es uno de los factores limitantes fundamentales en la distribución de la vida en la Tierra. Existen grandes diferenciales de temperatura sobre la Tierra, que se deben a la energía solar y a la distribución de tierra y agua. Por ser la temperatura un mecanismo limitante en la distribución de plantas y animales, sus valores extremos son fundamentales en la determinación de los patrones de una gran cantidad de actividades. Temperaturas altas y bajas extremas, tienen un impacto fuerte en la salud, agricultura, consumo de energía, recursos acuáticos, transportación, etc. Smith et al. (1997), proponen una metodología específica para el análisis de valores extremos de temperatura.

En confiabilidad se tienen sistemas formados por una gran cantidad de componentes conectados en serie, donde la falla del sistema ocurre al fallar el primer componente. De

esta forma el tiempo de falla del sistema es el mínimo de los tiempos de falla de los componentes. Cuando se tienen varios sistemas, que contienen un gran número de componentes conectados en serie funcionando en un ambiente común, entonces la distribución de los tiempos de vida de los sistemas es una distribución multivariada de valores extremos.

1.2 Teoría de extremos multivariados

El propósito de esta sección es presentar un resumen de la herramienta teórica existente en la literatura para abordar la modelación de los problemas a los que nos referimos en la sección anterior.

Suponemos que se miden k características de n elementos de una población. Sea X_{ij} el valor de la característica j medida en el individuo i, donde $1 \leq i \leq n$ y $1 \leq j \leq k$. Representamos las n observaciones multivariadas por $\{\mathbf{X}_i = (X_{i1}, X_{i2}, ..., X_{ik}), 1 \leq i \leq n\}$. Estamos interesados en la distribución asintótica de $\mathbf{M}_n = (M_{n1}, M_{n2}, ..., M_{nk})$, donde $M_{nj} = \max(X_{1j}, X_{2j}, ..., X_{nj})$. Sean $\mathbf{X}_1$, $\mathbf{X}_2$, ... observaciones independientes idénticamente distribuidas, con distribución común $F(\mathbf{x}) = F(x_1, x_2, ..., x_k)$. Supongamos que existen sucesiones $\mathbf{a}_n$, $\mathbf{b}_n$ de vectores tales que cada componente de $\mathbf{b}_n$ es positiva y

$$P\left[\frac{M_{nj} - a_{nj}}{b_{nj}} \leq x_j, 1 \leq j \leq k\right] = F^n(b_{n1}x_1 + a_{n1}, ..., b_{nk}x_k + a_{nk}) \qquad (1.1)$$

converge en distribución a una distribución límite L tal que cada marginal $L_j, 1 \leq j \leq k$, es no degenerada. Cualquier distribución F que satisface lo anterior, se dice que está en el dominio de atracción de la distribución asintótica de valores extremos L.

La distribución límite L satisface la relación de max-estabilidad, Resnick (1987): Para $1 \leq j \leq k$ y para todo $t > 0$ existen funciones $\alpha_j(t), \beta_j(t) > 0$ tales que

$$L^t(\mathbf{x}) = L(\beta_1(t)x_1 + \alpha_1(t), ..., \beta_k(t)x_k + \alpha_k(t)), \text{ para todo } t > 0. \qquad (1.2)$$

La convergencia de la $j-ésima$ marginal en (1.1) nos dá la distribución límite marginal L_j, que es de alguno de los siguientes tres tipos:

$$\Phi(x;\mu,\sigma,\alpha) = \begin{cases} 0, & x \leq \mu \\ \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}\right\}, & x > \mu, \quad \alpha,\sigma > 0, \ -\infty < \mu < \infty \end{cases}$$

$$\Psi(x;\mu,\sigma,\alpha) = \begin{cases} \exp\left\{-\left(-\frac{x-\mu}{\sigma}\right)^{\alpha}\right\}, & x < \mu, \quad \alpha,\sigma > 0, \ -\infty < \mu < \infty \\ 1, & x \geq \mu \end{cases}$$

$$\Lambda(x;\mu,\sigma) = \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, \ -\infty < x < \infty, \ \sigma > 0, \ -\infty < \mu < \infty.$$

Estas distribuciones asintóticas univariadas para máximos se conocen como tipos Fréchet, Weibull y Gumbel respectivamente. Los parámetros μ y σ son de localización y escala. En los primeros dos tipos tenemos además el parámetro de forma α que refleja el comportamiento de F_j en la cola derecha. Una propiedad importante de estas distribuciones de valor extremo univariadas, es que podemos ir de una distribución a la otra con cambios de variables adecuados.

Se tienen resultados similares a los anteriores para la distribución asintótica de mínimos. Las tres distribuciones límites en el caso de mínimos son:

$$\Phi^*(x;\mu,\sigma,\alpha) = \begin{cases} 1 - \exp\left\{-\left(-\frac{x-\mu}{\sigma}\right)^{-\alpha}\right\}, & x < \mu, \quad \alpha,\sigma > 0, \ -\infty < \mu < \infty \\ 1, & x \geq \mu \end{cases}$$

$$\Psi^*(x;\mu,\sigma,\alpha) = \begin{cases} 0, & x \leq \mu, \quad \alpha,\sigma > 0, \ -\infty < \mu < \infty \\ 1 - \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^{\alpha}\right\}, & x > \mu \end{cases}$$

$$\Lambda^*(x;\mu,\sigma) = 1 - \exp\left\{-\exp\left(\frac{x-\mu}{\sigma}\right)\right\}, \ -\infty < x < \infty, \ \sigma > 0, \ -\infty < \mu < \infty.$$

Como las distribuciones límites univariadas para máximos y mínimos pueden transformarse unas en otras, no hay pérdida de generalidad al suponer que las marginales de la distribución asintótica de valores extremos multivariada están distribuidas identicamente, de acuerdo a alguna de las distribuciones limites univariadas.

Debido a que la elección de las marginales es arbitraria, en la literatura se han utilizado tres tipos de distribuciones asintóticas para estudiar propiedades de extremos multivariados. Tiago de Oliveira (1980) y Galambos (1978) utilizan la distribución Gumbel o de valor extremo para máximos, Hann y Resnick (1977) trabajan con la distribución Fréchet, mientras que Pickands (1981) y Deheuvels (1983) prefieren la distribución exponencial.

De las representaciones anteriores para la distribución de extremos multivariados, es la de Pickands (1981) la que utilizamos en este trabajo, debido a que en el caso bivariado nos permite estudiar la estructura de dependencia de los extremos de una manera sencilla, como se expone en el Capítulo tres. A continuación describimos la representación de Pickands para la distribución de extremos multivariados.

Sea $\mathbf{Y}_n = (Y_{n1}, Y_{n2}, ..., Y_{nk})$, donde $Y_{nj} = \min(X_{1j}, X_{2j}, ..., X_{nj})$, y $\mathbf{X}_1$, $\mathbf{X}_2$, ... son las observaciones multivariadas de dimensión k. Supongamos que existen sucesiones $\mathbf{c}_n$, $\mathbf{d}_n$ de vectores tales que cada componente de $\mathbf{d}_n$ es positiva y

$$P\left[\frac{Y_{nj} - c_{nj}}{d_{nj}} > x_j, 1 \leq j \leq k\right] = \bar{F}^n(b_{n1}x_1 + a_{n1}, ..., b_{nk}x_k + a_{nk}) \to G(x_1, ..., x_k) \quad (1.3)$$

cuando $n \to \infty$, donde cada marginal $G_j, 1 \leq j \leq k$ se distribuye exponencialmente, $G(x_j) = \exp(-x_j), x_j > 0, 1 \leq j \leq k$. En este caso, la distribución asintótica G satisface la condición de min-estabilidad: para toda $s \geq 0$ se tiene que

$$G^s(x_1, x_2, ..., x_k) = G(sx_1, sx_2, ..., sx_k), \; x_1 > 0, ..., x_k > 0. \quad (1.4)$$

En el siguiente teorema tenemos una representación para la distribución asintótica de valores extremos multivariados con marginales exponenciales unitarias.

Teorema (Pickands, 1981). La función de supervivencia $G(x_1, x_2, ..., x_k)$ satisface la condición de min-estabilidad si y solo si existe una medida positiva H definida en el simplex unitario,

$$S_k = \left\{(w_1, ..., w_{k-1}) \in [0,1]^{k-1} : w_1 + w_2 + ... + w_{k-1} \leq 1\right\},$$

tal que,

$$G(\mathbf{x}) = \exp\left\{-(x_1 + ... + x_k) \int_{S_k} \max\left[w_1 q_1, ..., w_k q_k\right] dH(q_1, ..., q_{k-1})\right\}, \quad (1.5)$$

donde $q_k = 1 - q_1 - ... - q_{k-1}$, $w_j = x_j/(x_1 + ... + x_k)$. La restricción adicional,

$$\int_{S_k} q_1 dH(q_1, ..., q_{k-1}) = ... = \int_{S_k} q_k dH(q_1, ..., q_{k-1}) = 1, \quad (1.6)$$

es necesaria para que las marginales tengan distribución exponencial unitaria.

La representación anterior de G puede expresarse también como,

$$G(\mathbf{x}) = \exp\left\{-\left(\sum_{j=1}^{k} x_i\right) A\left(x_1/\sum_{j=1}^{k} x_j, x_2/\sum_{j=1}^{k} x_j, .., x_{k-1}/\sum_{j=1}^{k} x_j\right)\right\}, \quad (1.7)$$

donde

$$A(w_1, ..., w_{k-1}) = \int_{S_k} \max\left[w_1 q_1, ..., w_k q_k\right] dH(q_1, ..., q_{k-1}) \quad (1.8)$$

es una función que también caracteriza la estructura de dependencia de $\mathbf{X}$, al igual que la medida H. La función A se conoce como **función de dependencia** y es una función convexa que satisface,

$$\max(w_1, w_2, ..., w_k) \leq A(w_1, w_2, ..., w_{k-1}) \leq 1,$$

y además como consecuencia de (1.6), la función de dependencia toma el valor 1 en los vértices del simplex S_k,

$$A(0,0,...,0) = A(1,0,...,0) = ... = A(0,0,...,1) = 1. \tag{1.9}$$

La representación anterior para $k = 2$, donde (X,Y) tiene distribución de valores extremos bivariados, con marginales exponenciales unitarios, resulta ser,

$$G(x,y) = \exp\left\{-(x+y)\,A\left(\frac{y}{x+y}\right)\right\},\ x>0,\ y>0 \tag{1.10}$$

con función de dependencia

$$A(w) = \int_0^1 \max\left[w\,(1-q)\,,(1-w)\,q\right] dH(q). \tag{1.11}$$

En este caso H es una medida positiva finita en [0,1], y para tener marginales exponenciales unitarias, se requiere que

$$\int_0^1 q dH(q) = \int_0^1 (1-q) dH(q) = 1. \tag{1.12}$$

Esto es, $(1/2)H$ es la función de distribución de una variable aleatoria con media $1/2$ y la función de dependencia A es una función convexa tal que $A(0) = A(1) = 1$, y $\max(w, 1-w) \leq A(w) \leq 1$. Con esta estructura de dependencia las variables X y Y están asociadas positivamente.

Las funciones de dependencia $A(w)$, de distribución característica $H(w)$ y de densidad característica $h(w)$, cuando esta existe, caracterizan la distribución bivariada de valores extremos. En principio cualquiera de las tres funciones se puede elegir para modelar la

distribución. La relación entre estas funciones es

$$A(w) = (1-w) + wH(w) - \int_0^w q dH(q), \tag{1.13}$$

$$H(w) = 1 + \frac{dA(w)}{dw}. \tag{1.14}$$

En la siguiente tabla se presentan las formas de las funciones de dependencia A y distribución característica H para los casos límites de dependencia total e independencia:

	Independencia	Dependencia total
$A(w)$	$1, 0 \leq w \leq 1$	$\max(w, 1-w), 0 \leq w \leq 1$
$H(w)$	$I_{(0,1)}(w) + 2I_{\{1\}}(w)$	$I_{(0,\frac{1}{2})}(w) + 2I_{[\frac{1}{2},1]}(w)$.

La modelación de la estructura de dependencia a través de la medida H ha sido utilizada por varios autores para proponer modelos de la función de dependencia.

Un aspecto importante de la teoría de valores extremos multivariados es que mientras en el caso univariado las distribuciones asintónticas son modelos con un máximo de tres parámetros, en el caso multivariado las distribuciones están parametrizadas por funciones, ya sea la función de dependencia A o la medida característica H. Estas dos funciones pueden considerarse en el caso más general, como parámetros de dimensión infinita.

Para el análisis de datos provenientes de distribuciones de valores extremos multivariados es conveniente tener modelos paramétricos de dimensión finita, que permitan hacer la inferencia por técnicas de verosimilitud. Para esto, es útil tener familias de funciones de dependencia o de distribuciones características, que aproximen adecuadamente a la función de dependencia A o a la distribución característica H.

Para el caso bivariado de valores extremos, se tiene una caracterización adicional de la estructura de dependencia que proviene de la siguiente proposición que establece la

relación entre la función de dependencia $A(w)$ y la función de distribución $F(w)$ del cociente $W = Y/(X+Y)$.

Proposición (Tiago de Oliveira, 1989): La función de dependencia $A(w)$ y la función de distribución $F(w)$ satisfacen,

$$F(w) = w + w(1-w)A^{'}(w)/A(w). \tag{1.15}$$

Demostración : Sea (X,Y) el vector de valores extremos, con función de supervivencia $G(x,y)$ definida en términos de la función de dependencia $A(w)$, de acuerdo a la relación (1.10). Desarrollando la función de supervivencia $S(w)$ para el cociente $W = Y/(X+Y)$, tenemos:

$$\begin{aligned} S(w) &= 1 - F(w) = P[W > w] = P\left[\frac{Y}{X+Y} > w\right] \\ &= P\left[Y > \frac{w}{1-w}X\right] = \int_0^\infty P\left[Y > \frac{w}{1-w}x | X = x\right] \exp(-x)dx. \end{aligned}$$

La supervivencia condicional $G_{Y|X}$ de Y dado $X = x$, se expresa por la supervivencia bivariada G y la supervivencia marginal G_X de X por,

$$G_{Y|X}(y|x) = \frac{1}{G'_X(x)}\frac{\partial G(x,y)}{\partial x},$$

entonces

$$\begin{aligned} P\left[Y > \frac{w}{1-w}x | X = x\right] &= G_{Y|X}\left(\frac{wx}{1-w}|x\right) \\ &= \left[\frac{1}{-\exp(-x)}\right] \exp\left[-\frac{x}{1-w}A(w)\right] \{-A(w) + wA'(w)\}. \end{aligned}$$

Sustituyendo esto en el desarrollo inicial de la supervivencia del cociente, tenemos,

$$\begin{aligned} S(w) &= [A(w) - wA'(w)]\left[\frac{1-w}{A(w)}\right] \int_0^\infty \exp[-z]\,dz \\ &= 1 - w - w(1-w)A'(w)/A(w). \end{aligned}$$

De aquí obtenemos la expresión (1.15)propuesta para la función de distribución del cociente W. ■

De la relación dada en la proposición anterior obtenemos la siguiente representación de la función de dependencia en términos de la función de distribución $F(w)$ del cociente W,

$$A(w) = \exp\left\{\int_0^w \frac{F(s) - s}{s(1-s)}ds\right\}. \tag{1.16}$$

1.3 Modelaje paramétrico de la función de dependencia

La estructura de dependencia de valores extremos multivariados está dada por la función de dependencia $A(w)$, que se define como una función integral determinada por una medida positiva $H(w)$, conocida como medida característica. La estructura de dependencia puede entonces definirse por la función de dependencia o por la medida característica.

Los primeros modelos conocidos para la función de dependencia fueron el logístico y el mixto simétricos dados por Gumbel (1960), antes de conocerse la representación dada por Pickands (1981) para la distribución de valores extremos multivariados, en donde aparece la función de dependencia, así como su caracterización a través de la medida característica. Posteriormente Tawn (1988) propuso las versiones asimétricas de los modelos mixto y logístico. Entre otros modelos de dependencia asimétricos que han

surgido se tienen el modelo bilogístico de Joe, Smith y Weissman (1989), así como el logístico invertido de Joe (1990).

En el Cápitulo dos se propone un modelo adicional para la función de dependencia. Este modelo resulta bastante flexible debido a que obtiene expresando la medida característica $H(w)$ como una mezcla de distribuciones betas.

A continuación se presentan las funciones de dependencia $A(w)$, de supervivencia conjunta $G(x,y)$ y de distribución del cociente $F(w)$ correspondientes a los diferentes modelos de dependencia mencionados. Además se describen los casos límites de independencia y dependencia total.

Modelo simétrico mixto, Gumbel (1960):

$A(w) = \theta w(w-1) + 1,\ 0 \leq \theta \leq 1.$

$G(x,y) = \exp\left\{-(x+y) + \frac{\theta xy}{x+y}\right\}.$

$F(w) = \frac{-\theta w^3 + 2\theta w^2 + (1-\theta)w}{\theta w^2 - \theta w + 1},\ 0 \leq \theta \leq 1.$

Esta familia caracteriza en forma incompleta la estructura de dependencia ya que aunque el caso de independencia se tiene cuando $\theta = 0$, el otro caso límite, $\theta = 1$ no corresponde a dependencia total, sino a un nivel de dependencia media entre X y Y.

Modelo simétrico logístico, Gumbel (1960):

$A(w) = \left\{(1-w)^\alpha + w^\alpha\right\}^{1/\alpha}, 1 \leq \alpha < \infty,$

$G(x,y) = \exp\left\{-\left(x^\alpha + y^\alpha\right)^{1/\alpha}\right\}.$

$F(w) = \frac{1}{1+\left(\frac{1-w}{w}\right)^\alpha},\ 1 \leq \alpha < \infty,$

Los casos de independencia y dependencia completa se tienen cuando $\alpha = 1$ y $\alpha \to \infty$ respectivamente.

Modelo mixto asimétrico, Tawn (1988):

$A(w) = \phi w^3 + \theta w^2 - (\theta + \phi)w + 1,\ 0 \leq \theta, \theta + \phi \leq 1,\ \theta + 2\phi \leq 1,\ \theta + 3\phi \geq 0,$

$G(x,y) = \exp\left\{-(x+y) + xy\left[x(\theta+\phi) + y(\theta+2\phi)\right](x+y)^{-2}\right\}.$

El caso de independencia se tiene cuando $\theta = \phi = 0$, sin embargo esta familia, al igual que el caso simétrico, caracteriza en forma incompleta la estructura de dependencia.

Modelo logístico asimétrico, Tawn (1988):

$A(w) = \left\{\theta^\alpha (1-w)^\alpha + \phi^\alpha w^\alpha\right\}^{1/\alpha} + (\theta - \phi)w + 1 - \theta,\ 0 \leq \theta, \phi \leq 1,\ 1 \leq \alpha < \infty.$

$G(x,y) = \exp\left\{-(1-\theta)x - (1-\phi)y - \left(\theta^\alpha x^\alpha + \phi^\alpha y^\alpha\right)^{1/\alpha}\right\}.$

$F(w) = \frac{\phi^\alpha w^\alpha \left[\theta^\alpha (1-w)^\alpha + \phi^\alpha w^\alpha\right]^{(1/\alpha)-1} + (1-\phi)w}{\left[\theta^\alpha (1-w)^\alpha + \phi^\alpha w^\alpha\right]^{1/\alpha} + (\theta - \phi)w + 1 - \theta}.$

El caso de independencia se tiene cuando $\theta = 0$ ó $\phi = 0$ ó $\alpha = 1$ mientras que el caso de dependencia completa se tiene cuando $\theta = \phi = 1$ y $\alpha \to \infty$.

Modelo logístico invertido, Joe (1990):

$A(w) = 1 - \left[(\phi_1 w)^{-\tau} + \left\{\phi_2 (1-w)\right\}^{-\tau}\right]^{-1/\tau},\ \tau \geq 0,\ 0 \leq \phi_i \leq 1,\ i = 1,2.$

$G(x,y) = \exp\left\{-(x+y) + \left(\phi_1^{-\tau} y^{-\tau} + \phi_2^{-\tau} x^{-\tau}\right)^{-1/\tau}\right\}.$

$F(w) = \frac{w - (\phi_1 w)^{-\tau}\left[(\phi_1 w)^{-\tau} + \left\{\phi_2(1-w)\right\}^{-\tau}\right]^{-1/\tau - 1}}{1 - \left[(\phi_1 w)^{-\tau} + \left\{\phi_2(1-w)\right\}^{-\tau}\right]^{-1/\tau}}.$

Los casos límites $\tau \to 0$ y $\tau \to 1$ corresponden a independencia y dependencia completa.

Modelo bilogístico, Joe, Smith y Weissman (1990):

$A(w) = \int_0^1 \max\left\{(1-\alpha)(1-w)u^{-\alpha}, (1-\beta)w(1-u)^{-\beta}\right\} du, 0 \leq \alpha, \beta \leq 1.$

Los casos limites de independencia y dependencia total cuando $\alpha = \beta \to 0$ y cuando $\alpha = \beta \to 1$.

Una expresión alternativa para esta función de dependencia es,

$A(w) = (1-w)u^{1-\alpha} + w(1-u)^{1-\beta}, 0 \leq \alpha, \beta \leq 1,$

donde $u = u(w, \alpha, \beta)$, es la raíz de la ecuación,

$$(1-\alpha)(1-w)(1-u)^\beta - (1-\beta)wu^\alpha = 0.$$

$G(x,y) = \exp\left\{-xu^{1-\alpha} - y(1-u)^{1-\beta}\right\}.$

$F(w) = \frac{1}{1+\left(\frac{1-\beta}{1-\alpha}\right)\left(\frac{u}{1-u}\right)}, 0 \leq \alpha, \beta \leq 1$, siendo u la función de w y los parámetros α y β definida anteriormente.

Modelo de Mezcla de betas

El modelo de mezcla de betas, presentado en el Capítulo dos, tiene asociada una distribución de W dada por,

$$F(w) = \frac{\sum_{i=1}^{2} \theta_i \left(\frac{\nu-\nu_i}{\nu}\right) \left[wB(w, \nu_i, \nu - \nu_i + 1) + 2\frac{\nu_i(\nu_i+1)}{(\nu+1)(\nu+2)} b(w, \nu_i + 2, \nu - \nu_i + 1)\right]}{(1-w) + \sum_{i=1}^{2} \theta_i \left[wB(w, \nu_i, \nu - \nu_i) - \left(\frac{\nu_i}{\nu}\right) B(w, \nu_i + 2, \nu - \nu_i + 1)\right]},$$

donde b y B son las funciones de densidad y de distribución acumulada beta. La restricción para los parámetros es, $0 \leq \nu_1 \leq \nu/2 \leq \nu_2 \leq \nu$.

1.4 Descripción del trabajo

En este trabajo se proponen y discuten modelos de funciones de dependencia de valores extremos bivariados (X, Y). Se enfatiza de manera especial el análisis estadístico de extremos bivariados en el esquema de cocientes de extremos, $W = Y/(X+Y)$

En el Capítulo 2 se presenta un nuevo modelo paramétrico para la función de dependencia, definiendo la distribución característica como una mezcla de betas. Como todas las distribuciones definidas por mezclas, ésta resulta ser sumamente flexible, además de que puede ser extendida a dimensiones mayores a dos. La función de dependencia $A(w)$ correspondiente, tiene como caso límite a una función convexa poligonal. Con este modelo podemos generar, con el algoritmo dado en el Apéndice, observaciones bivariadas, con una gran variedad de patrones de dependencia. Se ilustra la flexibilidad del modelo propuesto, comparando su ajuste a dos conjuntos de datos, con diferente estructura de dependencia, a los que también se ajusta el modelo logístico asimétrico.

En el Capítulo 3 se propone un nuevo modelo paramétrico para la estructura de dependencia simétrica, definido por una función de distribución simétrica para el cociente

$W = Y/(X+Y)$, con sólo un parámetro. Al definir el modelo de dependencia a través de la distribución del cociente, reducimos el problema originalmente bivariado, a uno univariado, pudiendo con esto, utilizar pruebas sencillas de bondad de ajuste para juzgar los modelos de interés. Los modelos mencionados en la sección anterior tienen formas complicadas para la función de distribución del cociente. Se dá la expresión para el coeficiente de asociación tao de Kendal en términos de la función de distribución del cociente W. El modelo aquí propuesto se encuentra en la familia exponencial, razón por la que podemos hacer inferencia de manera sencilla. Se presentan y discuten estimadores del parámetro del modelo propuesto, basados en verosimilitud, momentos y esquema Bayesiano.

El Capítulo 4 considera la estimación no-paramétrica de la estructura de dependencia de extremos bivariados. Se propone un estimador no-paramétrico para la función de dependencia. Este estimador se obtiene sustituyendo la función de distribución del cociente W por su versión muestral, en la expresión de la función de dependencia definida por la función de distribución del cociente. Este nuevo estimador es convexo, a diferencia del estimador propuesto por Capéraà, Fougerés y Genest (1997), partiendo del mismo principio de sustitución de la distribución teórica del cociente por la versión muestral. Se estudian algunas propiedades asintóticas del estimador y además, para el caso de muestras pequeñas se compara éste con otros estimadores no-paramétricos conocidos, a través de un estudio de simulación. En general el nuevo estimador resulta mejor que los estimadores de Pickands (1981) y de Tiago de Oliveira. (1992).

En el Capítulo 5 se discuten 3 ejemplos con datos reales, donde se tienen 3 niveles de dependencia bivariada: alta, media y baja. En el ejemplo de dependencia alta, los datos son niveles de el flujo de un río en dos puntos diferentes de su recorrido. En este caso se discuten los ajustes de diferentes modelos de dependencia, y se encuentra que los modelos simétricos de dependencia ajustan bien. Se analizan los datos con el modelo simétrico propuesto en el Capítulo tres, siguiendo el esquema de los cocientes. En este caso, se estiman algunas probabilidades conjuntas y condicionales de excedencias. En el ejemplo

de baja dependencia entre los valores extremos, se prueba la hipótesis de independencia de los extremos, siguiendo los enfoques de verosimilitud y Bayesiano.

Finalmente en el Capítulo 6 se presentan las aportaciones del trabajo, así como algunas áreas de oportunidad para la investigación futura en el campo de la estadística de valores extremos bivariados. Se advierte que dentro del esquema de los cocientes, el terreno es bastante fértil, debido a que no ha sido suficientemente investigado. Además, la difusión de técnicas de análisis relacionadas con el esquema de los cocientes, puede realizarse con facilidad entre los usuarios, debido a su sencilléz.

En el Apéndice se presenta un algoritmo para la generación de observaciones multivariadas con distribución de valores extremos. Este algoritmo se basa en la representación de Pickands (1981) para la función de dependencia.

Capítulo 2

Modelo de mezcla de betas para la distribución característica

2.1 Modelo de mezcla de betas

La distribución multivariada de valores extremos puede caracterizarse por una función de dependencia $A(w)$ o por una medida característica $H(w)$, definidas por (1.8) y (1.5) respectivamente. En el caso bivariado, $(1/2)H$ puede interpretarse como la función de distribución de una variable aleatoria con media 1 y soporte $[0,1]$. Una familia flexible de distribuciones con soporte $[0,1]$, que proponemos en este capítulo como modelo de H, es una mezcla de distribuciones beta,

$$H(q) = \sum_{i=1}^{m} \theta_i B(q, \nu_i, \nu - \nu_i), \quad 0 \leq q \leq 1 \tag{2.1}$$

donde $0 \leq \nu_1 \leq \nu_2 \leq ... \leq \nu_m \leq \nu$, $\theta_i > 0, i = 1,2,...,m$, $\theta_1 + \theta_2 + ... + \theta_m = 2$, $\nu_1\theta_1 + \nu_2\theta_2 + ... + \nu_m\theta_m = \nu$ y $B(q, \nu_i, \nu - \nu_i)$ es la función de distribución beta con parámetros ν_i y $\nu - \nu_i$.

De las restricciones (1.12) tenemos para $m = 2$,

$$\begin{aligned} \theta_1 + \theta_2 &= 2, \\ \nu_1\theta_1 + \nu_2\theta_2 &= \nu. \end{aligned}$$

De aquí tenemos que θ_1 y θ_2 pueden expresarse como,

$$\begin{aligned} \theta_1 &= \tfrac{\nu - 2\nu_2}{\nu_1 - \nu_2}, \\ \theta_2 &= \tfrac{\nu - 2\nu_1}{\nu_2 - \nu_1}. \end{aligned} \tag{2.2}$$

Sólo tres parámetros son libres, ν_1, ν_2, y ν, mientras que θ_1 y θ_2 son determinados por los primeros. Para tener θ_1 y θ_2 positivos, (2.2) requiere que los parámetros ν_1, ν_2, y ν satisfagan la relación,

$$0 < \nu_1 < \nu/2 < \nu_2 < \nu. \tag{2.3}$$

Para $m = 3$, tenemos las restricciones,

$$\begin{aligned} \theta_1 + \theta_2 + \theta_3 &= 2, \\ \nu_1\theta_1 + \nu_2\theta_2 + \nu_3\theta_3 &= \nu. \end{aligned}$$

En este caso hay cinco parámetros libres, que pueden ser ν_1, ν_2, ν_3, ν y θ_2, quedando θ_1 y θ_3 en términos de los primeros,

$$\begin{aligned} \theta_1 &= \tfrac{(\nu - 2\nu_1) - (\nu_2 - \nu_1)\theta_2}{\nu_3 - \nu_1} \\ \theta_3 &= \tfrac{(\nu - 2\nu_3) - (\nu_2 - \nu_3)\theta_2}{\nu_1 - \nu_3}. \end{aligned}$$

De acuerdo con (1.13), podemos expresar la función de dependencia A en términos

de la distribución característica H. Introduciendo en (1.13) la medida H definida por (2.1) como una mezcla de betas, obtenemos la correspondiente expresión de la función de dependencia en términos de distribuciones betas,

$$\begin{aligned} A(w) &= (1-w) + wH(w) - \int_0^w q\,dH(q) \\ &= (1-w) + w\sum_{i=1}^{m} \theta_i B(w, \nu_i, \nu - \nu_i) \\ &\quad - \sum_{i=1}^{m} \left(\frac{\nu_i}{\nu}\right) \theta_i B(w, \nu_i + 1, \nu - \nu_i). \end{aligned} \tag{2.4}$$

Este modelo paramétrico de la función de dependencia dá un modelo diferenciable, que aunque no tiene como casos particulares las situaciones de independencia y dependencia total sí los tiene como casos límites.

Proposición En el modelo de mezcla de 2 betas, tenemos los casos de independencia y dependencia total, como casos límites.

i) El caso de independencia se tiene cuando $(\nu_1/\nu) \to 0$, $(\nu_2/\nu) \to 1$ y $\nu \to \infty$.

ii) El caso de dependencia total se tiene cuando, $(\nu_1/\nu) \to 1/2, (\nu_2/\nu) \to 1/2$ y $\nu \to \infty$.

Demostración: De (2.4), para $m = 2$ tenemos el caso límite, cuando $\nu \to \infty$, $(\nu_1/\nu) \to w_1$ y $(\nu_2/\nu) \to w_2$,

$$A_\infty(w) = [1 - w_1\theta_1 J_1(w) - w_2\theta_2 J_2(w)] - w\,[1 - \theta_1 J_1(w) - \theta_2 J_2(w)]\,, \tag{2.5}$$

donde las funciones $J_1(w)$ y $J_2(w)$ son los casos límites de las funciones beta,

$$J_i(w) = \lim_{\substack{\nu\to\infty \\ \nu_i/\nu \to w_i}} B(w, \nu_i, \nu - \nu_i) = \lim_{\substack{\nu\to\infty \\ \nu_i/\nu \to w_i}} B(w, \nu_i - 1, \nu - \nu_i) = \begin{cases} 0 & 0 \le w < w_i \\ 1 & w_i \le w \le 1 \end{cases},$$

mientras que los coeficientes θ_1 y θ_2 tienen los valores límites, $(1 - 2w_2)/(w_1 - w_2)$ y $(1 - 2w_1)/(w_2 - w_1)$, respectivamente.

Para probar (i), observamos que cuando $\nu \to \infty, (\nu_1/\nu) \to 0$ y $(\nu_2/\nu) \to 1$, entonces $w_1 = 0, w_2 = 1, \theta_1 = 1$ y $\theta_2 = 1$. Sustituyendo estos valores en (2.5) tenemos, $A_\infty(w) = 1$, para $0 \leq w \leq 1$.

Para probar (ii), observamos que cuando $\nu \to \infty, (\nu_1/\nu) \to 1/2$ y $(\nu_2/\nu) \to 1/2$, entonces $w_1 = w_2 = 1/2$ y además,

$$J_1(w) = J_2(w) = \begin{cases} 0 & 0 \leq w < 1/2 \\ 1 & 1/2 \leq w \leq 1 \end{cases}.$$

Sustituyendo esto en (2.5), tenemos,

$$\begin{aligned} A_\infty(w) &= [1 - (1/2)(\theta_1 + \theta_2)J_1(w)] - w\,[1 - (\theta_1 + \theta_2)J_1(w)] \\ &= [1 - J_1(w)] - w\,[1 - 2J_1(w)] \\ &= \max(1 - w, w),\ 0 \leq w \leq 1. \end{aligned} \tag{2.6}$$

■

Tenemos el modelo de mezcla de dos betas simétrico cuando imponemos la restricción $\nu_2 = \nu - \nu_1$; en este caso los pesos de la mezcla son $\theta_1 = \theta_2 = 1$, y la expresión para la distribución característica es

$$H(q) = B(q, \nu_1, \nu - \nu_1) + B(q, \nu - \nu_1, \nu_1),$$

la cual es simétrica. La expresión para la función de dependencia correspondiente es,

$$\begin{aligned} A(w) &= (1 - w) + w\,[B(w, \nu_1, \nu - \nu_1) + B(w, \nu - \nu_1, \nu_1)] \\ &\quad + \left(\frac{\nu_1}{\nu}\right) B(w, \nu_1 + 1, \nu - \nu_1) + \left(1 - \frac{\nu_1}{\nu}\right) B(w, \nu - \nu_1 + 1, \nu_1). \end{aligned}$$

Un factor de dependencia es

$$\delta = \frac{2}{\nu} \max\{\nu_1, (\nu - \nu_2)\},$$

tomando valores cercanos a cero en el caso de dependencia baja y cercanos a uno en caso de dependencia alta. Un factor adicional de la fuerza de dependencia o independencia lo es el valor del parámetro ν.

Proposición La función de dependencia para el modelo de mezcla de m betas, tiene como caso límite una función poligonal convexa cuando $\nu \to \infty, (\nu_i/\nu) \to w_i$, para $i = 1, 2, ..., m$.

Demostración: La función característica $H(q)$ definida por (2.1), tiene como forma límite a la función,

$$H_\infty(q) = \sum_{i=1}^{m} \theta_i J_i(w), \tag{2.7}$$

donde

$$J_i(w) = \begin{cases} 0 & 0 \leq w < w_i \\ 1 & w_i \leq w \leq 1 \end{cases},\ i = 1, 2, ..., m. \tag{2.8}$$

La función H_∞ es una función escalonada, con saltos en los puntos $w_1, w_2, ..., w_m$, de magnitud $\theta_1, \theta_2, ..., \theta_m$ respectivamente. Como $0 \leq \nu_1 \leq \nu_2 \leq ... \leq \nu_m \leq \nu$ entonces $w_1 \leq w_2 \leq ... \leq w_m$.

La función de dependencia $A(w)$ está determinada por la función característica $H(w)$ según (1.13),

$$A(w) = 1 - w + wH(w) - \int_0^w q dH(q).$$

Al integrar por partes la integral del lado derecho, obtenemos,

$$A(w) = 1 - w + \int_0^w H(q) dq.$$

Sustituyendo H por H_∞, obtenemos la función de dependencia $A_\infty(w)$, que es una función poligonal, lineal entre cada par de puntos, w_{i-1}, w_i, debido a que H_∞ es constante en cada intervalo $[w_{i-1}, w_i)$.

■

Obtenemos la forma límite A_∞ de la función de dependencia, tomando el límite de (2.4), cuando ν y ν_i crecen, de tal forma que $(\nu_i/\nu) \to w_i$ para $i = 1, 2, ..., m$,

$$A_\infty(w) = \left[1 - \sum_{i=1}^{m} w_i \theta_i J_i(w)\right] - w\left[1 - \sum_{i=1}^{m} \theta_i J_i(w)\right], \tag{2.9}$$

donde $J_i(w), i = 1, 2, ..., m$ están dadas por (2.8).

Esta función poligonal convexa tiene sus puntos de cambio de pendiente en las abscisas $w = w_i$. Los coeficientes $\theta_1, \theta_2, ..., \theta_m$ de la función característica límite H_∞ escalonada, satisfacen las restricciones,

$$\sum_{i=1}^{m} \theta_i = 2, \tag{2.10}$$

$$\sum_{i=1}^{m} \theta_i w_i = 1. \tag{2.11}$$

Para demostrar esto, partimos de las restricciones para una función característica H, citadas en el capítulo 1,

$$\int_0^1 q dH(q) = \int_0^1 (1-q) dH(q) = 1,$$

las cuales en el caso bajo estudio son

$$\begin{aligned} H(1) &= 2, \\ \int_0^1 H(q)dq &= 1. \end{aligned}$$

De la primera expresión, obtenemos la restricción (2.10), mientras que de la segunda tenemos,

$$\int_0^1 H(q)dq = \int_0^1 \sum_{i=1}^{m} \theta_i J_i(q) dq = \int_0^1 \sum_{i=1}^{m} \left(\sum_{k=1}^{i} \theta_k\right) I_{[w_k, w_{k+1}]}(q) dq,$$

donde $w_{m+1} = 1$ y $I_{[a,b]}(q)$ es la función que vale uno en el intervalo $[a, b]$ y cero fuera de él. Esta integral resulta ser, .

$$\begin{aligned} \int_0^1 H(q)dq &= \sum_{i=1}^{m} \left(\sum_{k=1}^{i} \theta_k\right) \int_{w_k}^{w_{k+1}} dq = \sum_{i=1}^{m} \left(\sum_{k=1}^{i} \theta_k\right) (w_{k+1} - w_k) \\ &= \sum_{i=1}^{m} \theta_i (1 - w_i) = 1, \end{aligned}$$

de donde obtenemos (2.11).

Para valores grandes de ν la función de dependencia resulta ser una poligonal suavizada. El modelo con $m = 2$ tiene suficiente flexibilidad para modelar diferentes grados de asimetría y dependencia.

2.2 Verosimilitud

Anteriormente se han utilizado varios métodos ad-hoc para estimar los parámetros de modelos de valores extremos bivariados (Gumbel y Mustafi, 1967; Posner et al., 1969). En este trabajo nos interesa comparar los diferentes modelos citados, por lo tanto uti-

lizamos un método de estimación común, como lo es el método de máxima verosimilitud. Para todos los modelos, estimamos los parámetros de cada modelo de valores extremos, maximizando la verosimilitud basada en la función de densidad conjunta de (X, Y):

$$L(\theta) = \prod_{i=1}^{n} L_i(\theta; x_i, y_i) = \prod_{i=1}^{n} f(x_i, y_i; \theta).$$

Para construir la función de verosimilitud con el modelo propuesto, inicialmente expresamos la función de densidad conjunta bivariada $f(x, y; \theta)$, en términos de la función de distribución característica H, utilizada en la definición de la función de dependencia,

$$\begin{aligned} f(x,y) &= \frac{\partial^2 S(x,y)}{\partial x \partial y} \qquad (2.12) \\ &= \exp\left[-x - yH\left(\frac{y}{x+y}\right) + (x+y)\int_0^{\frac{y}{x+y}} qdH(q)\right] \\ &\times \left\{\frac{xy}{(x+y)^3} H'\left(\frac{y}{x+y}\right) + \left[1 - \int_0^{\frac{y}{x+y}} qdH(q)\right]\left[\int_0^{\frac{y}{x+y}} (1-q)\,dH(q)\right]\right\} \end{aligned}$$

sustituimos H y $h = H'$ por las correspondientes mezclas de betas (2.1) para obtener la densidad conjunta en términos de los parámetros del modelo propuesto, con la distribución característica dada por (2.3) con $m = 2$, y además,

$$H'(w) = h(w) = \sum_{i=1}^{2} \theta_i b(w, \nu_i, \nu - \nu_i),$$

$$\int_0^w qdH(q) = \sum_{i=1}^{2} \theta_i \left(\frac{\nu_i}{\nu}\right) b(w, \nu_i + 1, \nu - \nu_i),$$

$$\int_0^w (1-q)dH(q) = \sum_{i=1}^{2} \theta_i \left(\frac{\nu - \nu_i}{\nu}\right) b(w, \nu_i, \nu - \nu_i + 1).$$

El estimador $\hat{\theta}$ de máxima verosimilitud de $\theta = (\nu_1, \nu_2, \nu)$ es el que maximiza la función de verosimilitud con la restricción,

$$\theta \in \Theta_0 = \{(\nu_1, \nu_2, \nu)\,|0 \leq \nu_1 \leq \nu/2 \leq \nu_2 \leq \nu\}.$$

Para obtener este estimador, maximizamos la función de verosimilitud $\tilde{L}(\theta)$ modificada de la siguiente forma: $\tilde{L}(\theta) = \begin{cases} L(\theta) & si\ \theta \in \Theta_0 \subset R^3 \\ 0 & si\ \theta \in R^3 - \Theta_0. \end{cases}$

Como la función de verosimilitud que resulta al considerar el modelo propuesto o los modelos mencionados, es complicada, resulta que el estimador de máxima verosimilitud no tiene forma cerrada, sino que se obtiene numéricamente.

2.3 Comparación

Con el fin de ilustrar la flexibilidad del modelo de mezcla de dos betas aquí propuesto, se simularon dos conjuntos de datos con diferente estructura de dependencia, ajustando los modelos logístico asimétrico y de mezcla de dos betas. En estos dos ejemplos, observamos que mientras que el modelo propuesto se adapta al logístico asimétrico, éste no tiene la flexibilidad para adaptarse al modelo de mezcla.

El modelo logístico asimétrico de la función de dependencia sólo tiene un punto de cambio, mientras que el modelo de mezcla de betas tiene dos. Esto le da mayor flexibilidad al segundo modelo, como se observa en las Figuras 2.1 y 2.2, donde se muestran dos ejemplos de ajuste de estos modelos a diferentes conjuntos de datos simulados.

En la Figura 2.1, se muestran las gráficas del modelo logístico asimétrico utilizado para simular 100 parejas (X, Y) a las que se ajustaron los modelos logístico asimétrico y mezcla de dos betas cuyas gráficas se presentan aquí. Los modelos ajustados prácticamente

coinciden en todo el intervalo $(0, 1)$.

Tabla 2.1. Parámetros de modelo simulado y modelos estimados.

modelo	param.	param.	param.	log-ver.
Logístico asimétrico simulado	$\theta = 1.0$	$\phi = .8$	$\alpha = 5$	
Mezcla de 2 betas estimado	$\hat{\nu}_1 = .124$	$\hat{\nu}_2 = 20.054$	$\hat{\nu} = 34.579$	-110.802
Logístico asimétrico estimado	$\hat{\theta} = .969$	$\hat{\phi} = .735$	$\hat{\alpha} = 6.240$	-109.921

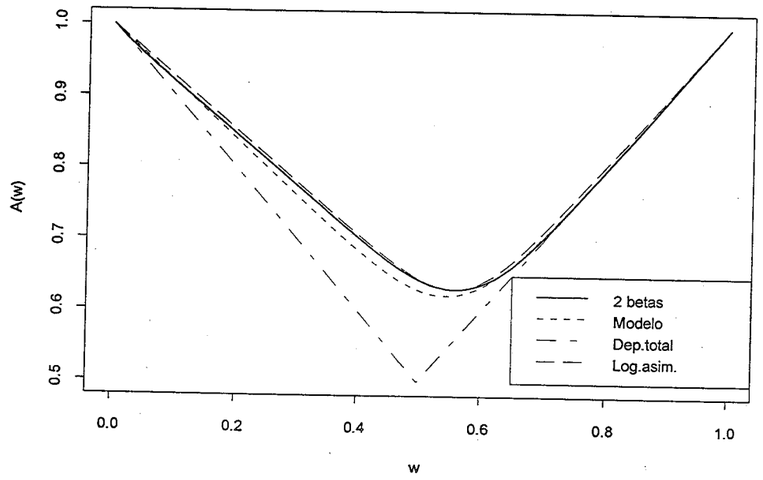


Figura 2.1. Funciones de dependencia simulada y estimadas. Modelo simulado: logístico asimétrico. Modelos estimados: mezcla de 2 betas y logístico asimétrico.

Así mismo se simularon 100 parejas (X, Y) de valores extremos con una función de dependencia mezcla de betas que se presenta en la Figura 2.2, donde también aparecen las funciones de dependencia logística asimétrica y mezcla de betas ajustadas. En esta Figura, a diferencia de la anterior, encontramos que la mezcla de betas estimada

prácticamente coincide con la función de dependencia simulada, mientras que la logística estimada sólo coincide en la rama derecha de la dependencia simulada, esto es en la parte donde la dependencia fuerte es más prolongada. Este comportamiento anómalo se debe a que la logística sólo tiene un punto de cambio en su curvatura.

Tabla 2.2. Parámetros de modelo simulado y modelos estimados.

modelo	param.	param.	param.	log-ver.
Mezcla de 2 betas simulado	$\nu_1 = 20$	$\nu_2 = 70$	$\nu = 100$	
Mezcla de 2 betas estimado	$\hat{\nu}_1 = 27.414$	$\hat{\nu}_2 = 93.689$	$\hat{\nu} = 132.857$	-140.335
Logístico asimétrico estimado	$\hat{\theta} = 1.0$	$\hat{\phi} = .424$	$\hat{\alpha} = 9.275$	-162.835

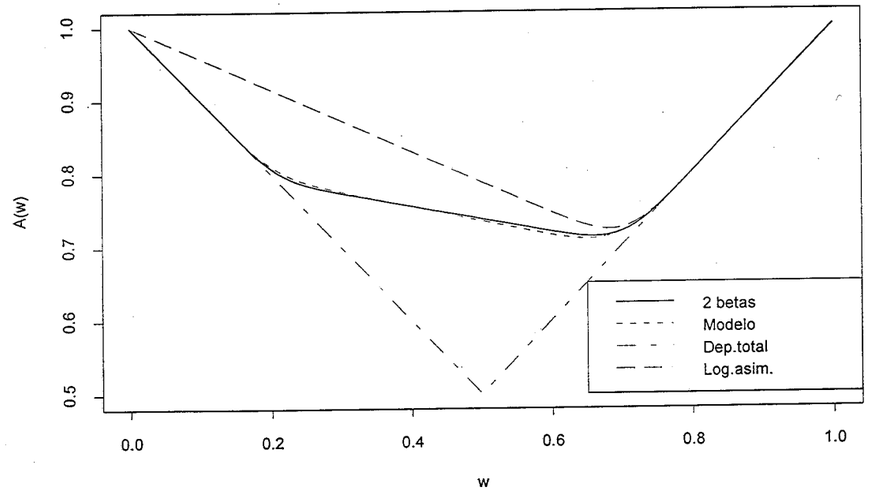


Figura 2.2. Funciones de dependencia simulada y estimadas. Modelo simulado: mezcla de 2 betas. Modelos estimados: mezcla de 2 betas y logístico asimétrico.

Se simularon dos conjuntos de datos con el modelo logístico asimétrico y el de mezcla de dos betas. Al ajustar ambos modelos a los datos generados con el modelo logístico se

observó que ambos modelos ajustan igualmente bien, mientras que al ajustar los modelos a los datos generados con el modelo mezcla de betas, el modelo mezcla de betas ajustó bien y en cambio el modelo logístico asimétrico tuvo un ajuste deficiente. Aquí concluimos que el modelo de mezcla de betas es bastante flexible al ajustar extremos bivariados con dependencia presente.

Cabe destacar que el modelo propuesto en el caso bivariado puede extenderse a tres o más dimensiones, utilizando mezcla de distribuciones Dirichlet.

Capítulo 3

Un modelo de dependencia de extremos bivariados (X,Y), en términos de la distribución del cociente W=Y/(X+Y)

3.1 Caracterización

La estructura de dependencia de los valores extremos bivariados (X, Y) está definida de acuerdo a la caracterización de Pickands (1981) por la función de dependencia $A(w)$, que es función de X y Y a través del cociente $W = Y/(X+Y)$. El patrón de variabilidad de W depende de la estructura de dependencia de las variables exponenciales unitarias X y Y. Así, cuando las dos variables son independientes, W se distribuye uniformemente en el intervalo $(0, 1)$, mientras que si hay un patrón de dependencia total, donde $P(X = Y) = 1$, entonces W tiene en el intervalo $(0, 1)$ una distribución de salto, con $P(W = 1/2) = 1$. En general la relación entre la estructura de dependencia de X y Y, y el patrón de variabilidad de W, está dado por la siguiente relación entre la función de dependencia $A(w)$ y la función de distribución del cociente, cuya demostración se dió en el Capítulo

1:

$$F(w) = w + w(1-w)A'(w)/A(w). \tag{3.1}$$

Esta relación ha sido planteada anteriormente por Tiago de Oliveira (1984), en el contexto de la función de dependencia para el caso de valores extremos bivariados con marginales Gumbel. Fácilmente obtenemos de (3.1) una representación adicional para la función de dependencia $A(w)$ en términos de la función de distribución $F(w)$ del cociente W,

$$A(w) = \exp\left\{\int_0^w \frac{F(s)-s}{s(1-s)}ds\right\}. \tag{3.2}$$

Las restricciones que tiene la función de distribución $F(w)$, son las heredadas por la función de dependencia, mencionadas en el Capítulo 1, a través de la relación (3.1).

De la identidad (3.2) y de las restricciones para $A(w)$, de ser convexa y además $A(0) = A(1) = 1$, tenemos que existe un intervalo $[w_0, w_1]$, con $0 \leq w_0 \leq w_1 \leq 1$, tal que la función

$$g(w) = \int_0^w \frac{F(s)-s}{s(1-s)}ds$$

es decreciente en $[0, w_0]$, se anula en $[w_0, w_1]$ y es creciente en $[w_0, 1]$.

Esto significa que la función de distribución $F(w)$ inicia por debajo de la diagonal, cruza en un punto w_0 a la diagonal, coincidiendo con ella en $[w_0, w_1]$ y posteriormente continua por arriba de la diagonal, de tal forma que la integral de 0 a 1, que aparece en la exponencial de (3.2) es cero.

Tanto la forma de la función de dependencia $A(w)$ como de la función de distribución $F(w)$ nos dan una idea gráfica de la estructura de dependencia de las variables X y Y. La ventaja de analizar la estructura de dependencia de los extremos X, Y a través de los

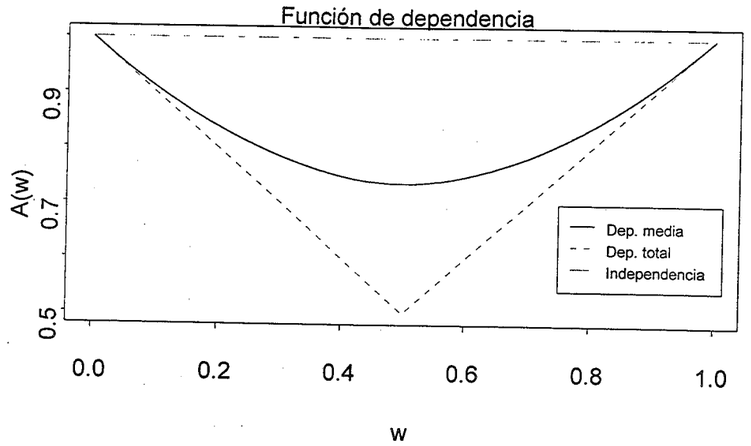
cocientes $W = Y/(X+Y)$, se debe a la sencillez con que podemos visualizar la forma de la función de distribución del cociente, evaluando la función de distribución empírica para la muestra de cocientes $w_1, w_2, ..., w_n$ que tengamos. Además podemos utilizar pruebas de bondad de ajuste basadas en la función de distribución empírica.

Cuando tenemos un modelo paramétrico para la estructura de dependencia de valores extremos bivariados, podemos estimar el modelo considerando el esquema de verosimilitud, bajo cualquiera de las dos caracterizaciones; la función de dependencia o la función de distribución del cociente. En el caso de caracterizar el patrón de dependencia con la función $A(w)$, la estimación se dá en el esquema bivariado (X, Y), mientras que al considerar la dependencia a través de la función $F(w)$, como lo hacemos en este capítulo, la estimación se dá en el esquema univariado W.

En el esquema de la distribución del cociente W, un patrón de dependencia positiva se traduce en acumulación de puntos w_i en la parte central del intervalo $[0, 1]$, mientras que la independencia entre X y Y corresponde a la distribución Uniforme$(0, 1)$ de W.

En el caso de dependencia positiva, la función de dependencia $A(w)$, es una función convexa que se encuentra arriba de $\max(w, 1-w)$ y $A(0) = A(1) = 1$.

En la Figura 3.2 tenemos funciones de supervivencia bivariada para diferentes casos de dependencia. En las gráficas correspondientes a dependencia positiva tenemos mayor probabilidad de encontrar que X y Y ambas sean grandes o pequeñas. Esta caracterización de la estructura de dependencia entre X y Y permite investigar diferentes aspectos del patrón de dependencia estudiando la función de distribución $F(w)$. Por ejemplo, la hipótesis de independencia entre las variables X y Y se reduce a probar la hipótesis de distribución Uniforme$(0, 1)$ para los cocientes $w_1, w_2, ..., w_n$. Igualmente, la hipótesis de simetría alrededor de $1/2$ de la función de dependencia $A(w)$ se traduce en la hipótesis de simetría alrededor de $1/2$ de la función de distribución $F(w)$, que puede probarse con estadísticas basadas en la función de distribución empírica, discutidas en D'Agostino-Stephens (1988).



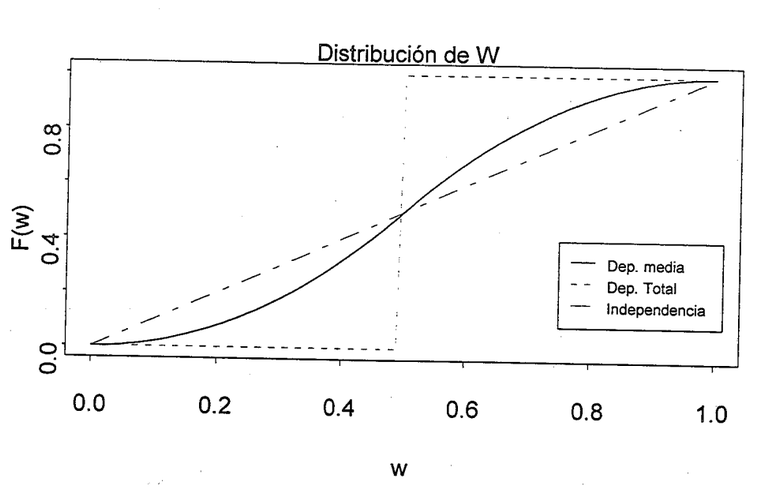
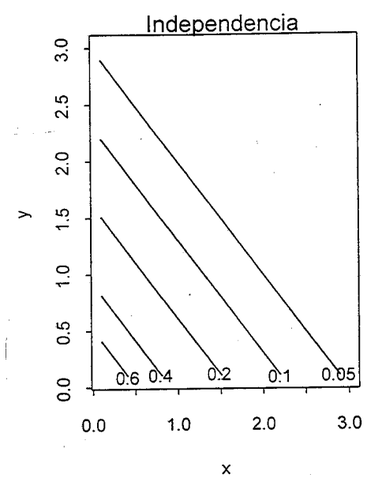
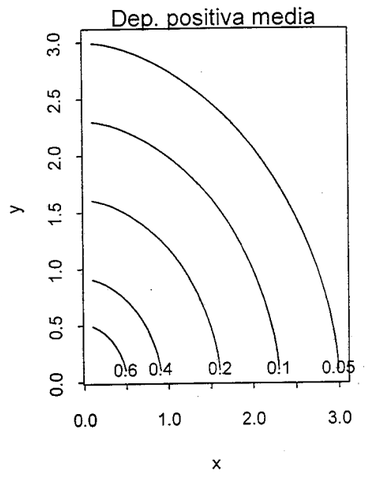


Figura 3.1. Funciones de dependencia $A(w)$, y de distribución $F(w)$ del cociente $W = Y/(X+Y)$, para los casos de independencia, dependencia media y dependencia total.





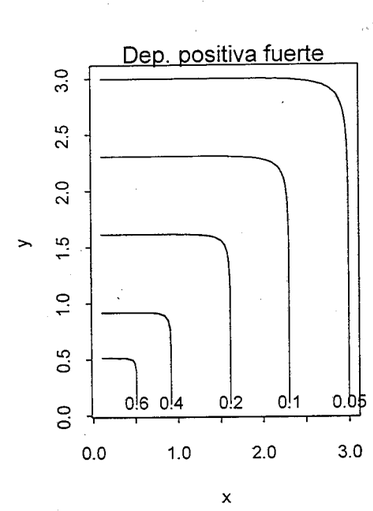


Figura 3.2. Gráficas de contorno de la función de supervivencia bivariada para diferentes casos de dependencia de los valores extremos X y Y.

Las funciones de distribución del cociente W, correspondientes a los modelos paramétricos de la distribución bivariada de valores extremos citados anteriormente, que se se obtienen a partir de la relación (3.1), se presentan en el Capítulo 1.

3.2 Modelo propuesto

Una característica notable de la teoría de valores extremos multivariados es que, a diferencia del caso univariado, no existe una familia paramétrica de dimensión finita para la función de dependencia A. Por esta razón, la modelación paramétrica ha consistido en dar subfamilias paramétricas de dimensión finita, que tienen la flexibilidad suficiente para aproximar los diferentes casos posibles de dependencia.

Considerando que hay una relación biunívoca entre las funciones de dependencia A y las funciones de distribución F del cociente, la estrategia anterior de modelación, se puede seguir para modelar la estructura de dependencia, en términos de la función de distribución del cociente.

En la Figura 3.1 presentamos gráficamente la relación entre la función de dependencia $A(w)$ y la función de distribución $F(w)$, para diferentes casos de dependencia.

El caso de independencia corresponde a la función de dependencia constante $A(w) = 1$, y a la distribución Uniforme$(0,1)$ para el cociente W. El caso de dependencia total, equivale a tener como función de dependencia $A(w)$, las diagonales superiores del cuadrado unitario, mientras que la función de distribución $F(w)$ es la función de salto en $1/2$.

El caso general de dependencia simétrica, corresponde en el caso de la función $A(w)$, a una función convexa simétrica dentro del área acotada por los casos límite de dependencia, mientras que para la función de distribución $F(w)$, tenemos una función sigmoidal simétrica ubicada dentro del área limitada por los casos de independencia y dependencia total.

En este esquema de modelación, proponemos un modelo para la estructura de dependencia simétrica, dado por una subfamilia de distribuciones simétricas, con soporte $[0,1]$, que tiene como un caso límite la distribución singular que asigna probabilidad uno a $w = 1/2$, y también tiene como uno de sus miembros a la distribución Uniforme$(0,1)$. Gráficamente esta familia cubre el área delimitada por los casos extremos de dependencia mostrados en la Figura 3.1.

El modelo propuesto para la estructura de dependencia, está dado por la familia de distribuciones,

$$F(w) = \begin{cases} 2^{a-1}w^a, & 0 \leq w \leq 1/2 \\ 1 - 2^{a-1}(1-w)^a, & 1/2 < w \leq 1, \quad 0 < a < \infty. \end{cases} \tag{3.3}$$

Cuando $a = 1$, tenemos el caso de independencia entre la variables X y Y, ya que $F(w) = w$ y por lo tanto, de (3.2) tenemos $A(w) = 1, 0 \leq w \leq 1$. Cuando $a > 1$, tenemos el caso de dependencia positiva entre las variables X y Y. Este modelo también considera el caso de dependencia negativa entre X y Y, que se tiene cuando $0 < a < 1$. La ventaja de extender el espacio parametral de $[1,\infty)$ a $(0,\infty)$ radica en que el parámetro $a = 1$, es un punto de interés, ya que corresponde a tener independencia; en el espacio parámetral extendido este valor queda como un punto interior, mientras que en el subespacio $[1,\infty)$ es un punto frontera, lo cual crea problema al hacer inferencia sobre el parámetro a, tanto con el esquema de verosimilitud como con el de momentos.

Este modelo es similar al logístico simétrico, sólo que, como veremos en la sección de inferencia, a diferencia del modelo logístico, en el nuevo modelo se tiene que el estimador máximo verosímil del parámetro a tiene una forma cerrada, al igual que su varianza. Además, por la sencillez de la distribución del cociente, fácilmente se obtiene el momento de orden dos, con lo cual se llega a una expresión sencilla para el estimador del parámetro a por el método de momentos.

Aunque el modelo propuesto se define por la función de distribución del cociente dada en (3.3), utilizando la relación (3.2), podemos obtener la expresión correspondiente para la función de dependencia A. Esta función no tiene una forma cerrada, y como se

demuestra enseguida, se expresa en términos de la serie

$$A(w) = \max(w, 1-w) \exp\left[2^{a-1} \sum_{n=0}^{\infty} \frac{(\min(w, 1-w))^{a+n}}{a+n}\right], 0 \leq w \leq 1. \tag{3.4}$$

Para demostrar la identidad (3.4) consideramos primeramente, $0 < w < 1/2$. Desarrollando la integral que se tiene como argumento de la exponencial en (3.2), tenemos,

$$\begin{aligned} \int_0^w \left\{\frac{F(s)-s}{s(1-s)}\right\} ds &= \int_0^w \left\{\frac{2^{a-1}s^a - s}{s(1-s)}\right\} ds \qquad (3.5) \\ &= \log(1-w) + 2^{a-1} \int_0^w \left\{\frac{s^{a-1}}{(1-s)}\right\} ds \\ &= \log(1-w) + \int_0^w \sum_{i=0}^{\infty} s^{a-1+i} ds \\ &= \log(1-w) + \sum_{i=0}^{\infty} \frac{w^{a+i}}{a+i}. \end{aligned}$$

De manera similar, para $1/2 < w \leq 1$, desarrollando la integral que se tiene como argumento de la exponencial en (3.2), tenemos,

$$\int_0^w \left\{\frac{F(s)-s}{s(1-s)}\right\} ds = -\int_w^1 \left\{\frac{F(s)-s}{s(1-s)}\right\} ds = -\int_w^1 \left\{\frac{1-2^{a-1}(1-s)^a - s}{s(1-s)}\right\} ds,$$

y haciendo el cambio de variable, $t = s - w$, obtenemos,

$$\begin{aligned} \int_0^w \left\{\frac{F(s)-s}{s(1-s)}\right\} ds &= \log(w) + 2^{a-1} \int_0^{1-w} \left\{\frac{t^{a-1}}{(1-t)}\right\} dt \qquad (3.6) \\ &= \log(w) + 2^{a-1} \sum_{i=0}^{\infty} \frac{(1-w)^{a+i}}{a+i}. \end{aligned}$$

De (3.5) y (3.6) obtenemos (3.4). ■

Existe otra representación para la función de dependencia $A(w)$, en términos de la función hipergeométrica, que resulta al expresar la integral $\int_0^w \left\{\frac{s^{a-1}}{(1-s)}\right\} ds$ en términos de la función hipergeométrica, como se muestra enseguida.

$$\begin{aligned} \int_0^w \left\{\frac{s^{a-1}}{(1-s)}\right\} ds &= \int_0^1 \left\{\frac{(wt)^{a-1}}{(1-wt)}\right\} d(wt) = w^a \int_0^1 \left\{\frac{t^{a-1}}{(1-wt)}\right\} dt \\ &= \frac{w^a}{a} F(1, a, a+1, w), \end{aligned}$$

donde $F(1, a, a+1, w)$ es la función hipergeométrica (ver Abramowitz & Stegun, 1965). Esta representación adicional, es

$$A(w) = \max(w, 1-w) \exp\left[\frac{2^{a-1}}{a} \min(w, 1-w) F(1, a, a+1, \min(w, 1-w))\right]. \tag{3.7}$$

Para valores grandes de a se tiene el siguiente desarrollo asintótico (ver Abramowitz & Stegun, 1965) de la función hipergeométrica,

$$F(1, a, a+1, w) = \sum_{i=0}^{m} \left(\frac{a}{a+i}\right) w^i + O(|a+1|^{-m-1}).$$

Así se tiene la siguiente representación para la función de dependencia,

$$A(w) = \max(w, 1-w) \exp\left[2^{a-1} \sum_{i=0}^{m} \frac{(\min(w, 1-w))^{a+i}}{a+i} + O\left(|a+1|^{-m-1}\right)\right]. \tag{3.8}$$

El número de términos en la serie puede elegirse de acuerdo al orden de magnitud $O(|a+1|^{-m-1})$. Por ejemplo, especificando una exactitud 10^{-10}, en número de términos depende del valor del parámetro a:

a	1.001	1.01	1.1	1.5	2	3	4	5	6
m	32	31	30	25	20	17	14	12	11.

3.2.1 Medida de asociación

Una medida usual de asociación entre dos variables aleatorias X y Y, es la τ de Kendal τ, que se define como la diferencia entre las probabilidades de concordancia y discordancia para dos parejas de variables aleatorias, (X_1, Y_1) y (X_2, Y_2). La asociación entre los valores extremos (X, Y), medida por la τ de Kendall, se obtiene a partir de la función de dependencia $A(w)$, por Ghoudi, Khoudraji y Rivest (1998),

$$\tau = \int_0^1 \frac{t(1-t)}{A(t)} dA'(t). \tag{3.9}$$

Es posible definir este coeficiente de asociación, en términos de la función de distribución $F(w)$ del cociente W, como lo establecemos a continuación.

Proposición. La asociación entre los valores extremos (X, Y), medida por la τ de Kendall, se obtiene a partir de la función de distribución $F(w)$ del cociente $W = Y/(X+Y)$, por

$$\tau = \int_0^1 \left[\frac{F(t)}{t} - 1\right]\left[\frac{F(t)}{1-t} - 1\right] dt. \tag{3.10}$$

Demostración. De la relación (3.1) entre la función de dependencia $A(w)$ y la función de distribución $F(w)$, tenemos,

$$A'(t) = \frac{F(t)-t}{t(1-t)} \times A(t).$$

Derivando ambos lados de esta igualdad tenemos,

$$\frac{d}{dt}\left[A'(t)\right] = \left[\frac{F(t)-t}{t(1-t)}\right] A'(t) + A(t)\frac{d}{dt}\left[\frac{F(t)-t}{t(1-t)}\right].$$

Desarrollando y sustituyendo en la expresión integral (3.9) para τ tenemos,

$$\begin{aligned}\tau &= \int_0^1 \left\{\frac{[F(t)-t]^2 - [F(t)-t](1-2t)}{[t(1-t)]} + [f(t)-1]\right\} dt \\ &= \int_0^1 \left\{\frac{[F(t)-t][F(t)-1+t]}{[t(1-t)]}\right\} dt + [F(1)-1].\end{aligned}$$

Como el soporte del cociente es el intervalo $[0, 1]$, entonces $F(1) = 1$. Por lo tanto la expresión integral para τ, en términos de la función de distribución del cociente es

$$\tau = \int_0^1 \left\{\left[\frac{F(t)}{t} - 1\right]\left[\frac{F(t)}{1-t} - 1\right]\right\} dt.$$

■

Para el modelo propuesto, definido por la función de distribución del cociente, este coeficiente de asociación está dado en términos de la función hipergeométrica $F(\alpha, \beta, \gamma, w)$ por,

$$\tau = (1 - \frac{1}{a}) + \frac{1}{4a}F(1, 2a, 2a+1, 1/2) - \frac{1}{a+1}F(1, a+1, a+2, 1/2), \tag{3.11}$$

o de la representación en series para la función hipergeométrica, tenemos

$$\tau = (1 - \frac{1}{a}) + \frac{1}{2}\sum_{i=0}^{\infty} \frac{\left(\frac{1}{2}\right)^i}{2a+i} - \sum_{i=0}^{\infty} \frac{\left(\frac{1}{2}\right)^i}{a+1+i}. \tag{3.12}$$

Demostramos la expresión anterior de la τ de Kendall, partiendo de la relación (3.10), incluyendo la función de distribución del cociente $F(w)$ para el modelo propuesto,

$$
\begin{aligned}
\tau &= \int_0^{1/2} \left[\frac{2^{a-1}t^a}{t} - 1\right]\left[\frac{2^{a-1}t^a}{1-t} - 1\right] dt \\
&\quad + \int_{1/2}^{1} \left[\frac{1-2^{a-1}(1-t)^a}{t} - 1\right]\left[\frac{1-2^{a-1}(1-t)^a}{1-t} - 1\right] dt \\
&= 2\int_0^{1/2} \left[\frac{2^{a-1}t^a}{t} - 1\right]\left[\frac{2^{a-1}t^a}{1-t} - 1\right] dt \\
&= 2\int_0^{1/2} \left[\frac{4^{a-1}t^{2a-1}}{1-t} - \frac{2^{a-1}t^a}{1-t} - 2^{a-1}t^{a-1} + 1\right] dt.
\end{aligned}
$$

Los dos primeros términos del integrando no dan una forma cerrada para la integral, por lo tanto, la τ queda en términos de la función hipergeométrica,

$$
\begin{aligned}
\tau &= (1-\frac{1}{a}) + 2\left[4^{a-1}\int_0^{1/2} \frac{t^{2a-1}}{1-t}dt - 2^{a-1}\int_0^{1/2} \frac{t^a}{1-t}dt\right] \qquad (3.13) \\
&= (1-\frac{1}{a}) + \frac{1}{4a}F(1,2a,2a+1,1/2) - \frac{1}{a+1}F(1,a+1,a+2,1/2).
\end{aligned}
$$

■

Tabla 3.1. Relación entre el coeficiente de asociación τ de Kendall y el parámetro a del modelo de dependencia propuesto.

a	1.0	1.1	1.5	2	3	4	5	10	20	50	100
τ	.0	.078	.291	.445	.611	.699	.754	.871	.933	.973	.986

En la Figura 3.3 mostramos la relación entre la τ de Kendall y el parámetro a del modelo propuesto. Aquí observamos que incrementos iniciales del parámetro a correspon-

den a incrementos considerables del coeficiente de asociación τ. Esta relación se muestra numéricamente en la Tabla 3.1 para algunos valores de a y τ.

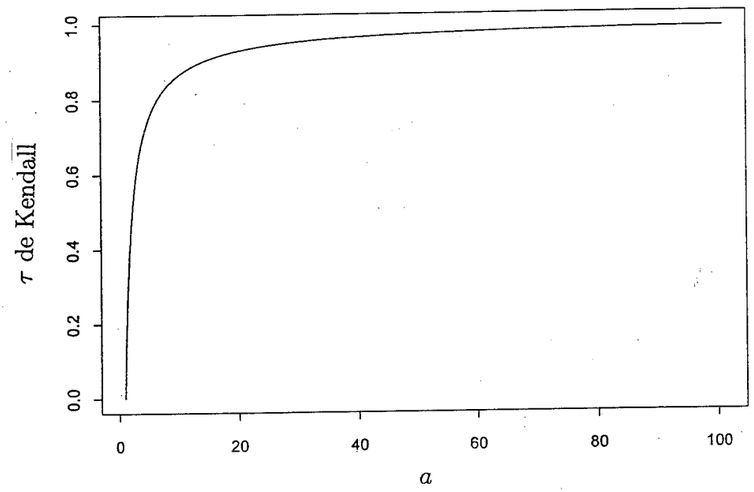


Figura 3.3. Relación entre el coeficiente de asociación τ de Kendall y el parámetro a del modelo de dependencia propuesto.

3.3 Inferencia paramétrica

La estructura de dependencia de los valores extremos (X, Y), se encuentra en la función de dependencia $A(w)$, que parametriza la distribución de los valores extremos (X, Y). Inicialmente la información sobre la estructura de dependencia, se encuentra en la muestra $(X_i, Y_i), i = 1, 2, ..., n$ de valores extremos. Al considerar la caracterización de la.

dependencia en términos de la función de distribución $F(w)$ del cociente, la información muestral se encuentra en los cocientes $W_i, i = 1, 2, ..., n$. El estudio de la estructura de dependencia cuando tenemos un modelo paramétrico, se puede llevar a cabo utilizando métodos de verosimilitud, el método de momentos o el esquema Bayesiano. Estos procedimientos en nuestro caso no presentan gran dificultad, debido a que la función de densidad de probabilidades tiene una forma sencilla,

$$f(w) = \begin{cases} a2^{a-1}w^{a-1}, & 0 \leq w \leq 1/2 \\ a2^{a-1}(1-w)^{a-1}, & 1/2 < w \leq 1. \end{cases} \tag{3.14}$$

$$= a\left[2\min(w, 1-w)\right]^{a-1}, \quad 0 \leq w \leq 1,$$

donde a es un parámetro que se encuentra en el rango $[1,\infty)$.

La familia de distribuciones aquí propuesta pertenece a la familia exponencial, ya que la densidad (3.14) se puede expresar en forma exponencial como,

$$f(w;a) = g(a)b(w)\exp\left\{\phi(a)h(w)\right\}, \tag{3.15}$$

donde, $g(a) = a, b(w) = 1, \phi(a) = -(a-1), h(w) = -\log\left[2\min(w, 1-w)\right]$.

Es importante encontrar que esta distribución está dentro de la familia exponencial, porque con esto, sabemos que comparte las propiedades conocidas para esta extensa familia de distribuciones.

Así, existe un estadístico completo suficiente para el parámetro a. Cuando la información disponible para estimar el parámetro a es la muestra aleatoria $\{W_1, W_2, ..., W_n\}$, entonces el estadístico completo suficiente es,

$$T(W_1, ..., W_n) = -\sum_{i=1}^{n} \log\left[2\min(W_i, 1-W_i)\right]. \tag{3.16}$$

En los modelos paramétricos de valores extremos bivariados que se han propuesto por

varios autores y que se mencionan en el primer Capítulo, la correspondiente distribución del cociente W no tiene una forma sencilla, y además no se encuentran en la familia exponencial. En el Apéndice del presente Capítulo se muestran las funciones de densidad de probabilidad del cociente.

Al comparar diferentes estimadores para el parámetro a, una medida de desempeño adecuada es el error cuadrático medio (ECM) de un estimador U, que está relacionado con su varianza y sesgo,

$$E_a(U-a)^2 = V_a(U) + [E_a(U) - a]^2.$$

Al restringir la comparación de estimadores a la familia de los estimadores insesgados del parámetro a, la varianza y el ECM coinciden, ya que $E_a(U) - a = 0$. Así, la búsqueda de un mejor estimador dentro de los estimadores insesgados, es equivalente a buscar un estimador insesgado con mínima varianza uniformemente (IMVU), esto es, un estimador S del parámetro a, tal que, $E_a(S) = a$ y $V_a(S) \leq V_a(U)$, para cualquier estimador insesgado U de a. Un resultado útil en la búsqueda de un estimador IMVU para a, es el siguiente:

Teorema (Lehmann-Scheffé). Si $T(W_1, W_2, ..., W_n)$ es un estadístico suficiente completo y $S(W_1, W_2, ..., W_n) = s(T)$, una función de T, es un estimador insesgado de a, entonces $S(W_1, W_2, ..., W_n)$ es un estimador IMVU de a

El estadístico suficiente completo T de a, dado en (3.16) no es insesgado para el parámetro a, pero podemos podemos transformarlo para obtener un estadístico insesgado. El siguiente teorema nos da la distribución del estadístico suficiente T.

Teorema. El estadístico suficiente $T(W_1, W_2, ..., W_n)$ dado en (3.16) tiene una distribución Gama, con parámetros, n y a.

Demostración. **Sea** W_i, una variable aleatoria con función de distribución acumulada

dada por (3.3), y sea $X_i = 2\min(W_i, 1-W_i)$. Para $0 \leq x \leq 1$,

$$\begin{aligned} F_X(x) &= P(X_i \leq x) = P(2\min(W_i, 1-W_i) \leq x) = P(\min(W_i, 1-W_i) \leq x/2) \\ &= P(W_i \leq x/2) + P(1 - W_i \leq x/2) = P(W_i \leq x/2) + P(W_i \geq 1 - x/2) \\ &= 2^{a-1}(x/2)^a + 1 - [1 - 2^{a-1}(1 - 1 + x/2)^a] \\ &= x^a. \end{aligned}$$

Sea la variable aleatoria $Y_i = -\ln(X_i)$. Sea $y > 0$.

$$\begin{aligned} F_{Y_i}(y) &= P(Y_i \leq y) = P[-\ln(X_i) \leq y] = P[X_i \geq \exp(-y)] \\ &= 1 - \exp(-ay). \end{aligned}$$

Como $T(W_1, W_2, ..., W_n)$ es una suma de n variables aleatorias $Y_1, Y_2, ..., Y_n$, independientes idénticamente distribuidas exponencialmente con parámetro a, entonces, T tiene distribución Gama(n, a).

■

De acuerdo a la distribución del estadístico completo y suficiente $T(W_1, W_2, ..., W_n)$, su esperanza y su varianza están dadas por,

$$\begin{aligned} E(T) &= n/a \\ V(T) &= n/a^2. \end{aligned}$$

Fácilmente podemos ver que si el estadístico T tiene distribución Gama con parámetros a y n, entonces función $S = (n-1)/T$ del estadístico T, es un estimador insesgado de a.

De acuerdo al teorema de Lehmann-Scheffé, el estimador

$$\begin{aligned} S(W_1, W_2, ..., W_n) &= (n-1)/T(W_1, W_2, ..., W_n) \\ &= -\frac{n-1}{\sum_{i=1}^{n} \log\left[2\min\left(W_i, 1-W_i\right)\right]}, \end{aligned}$$

para el parámetro a, es un estimador IMVU.

Como $T = T(W_1, W_2, ..., W_n)$ tiene distribución Gama(n, a), entonces podemos obtener estimaciones por intervalo a partir de la cantidad pivotal, $2aT$, que tiene distribución ji-cuadrada, con $2n$ grados de libertad.

De acuerdo a lo anterior,

$$\begin{aligned} 1 - \alpha &= P\left[\chi^2_{(2n,\alpha/2)} \leq 2aT \leq \chi^2_{(2n,1-\alpha/2)}\right] \\ &= P\left[\frac{\chi^2_{(2n,\alpha/2)}}{2T} \leq a \leq \frac{\chi^2_{(2n,1-\alpha/2)}}{2T}\right], \end{aligned}$$

por lo cual, un intervalo de $100(1-\alpha)\%$ de confianza para el parámetro a está dado por,

$$\left[\frac{\chi^2_{(2n,\alpha/2)}}{2T}, \frac{\chi^2_{(2n,1-\alpha/2)}}{2T}\right].$$

El límite inferior de este intervalo puede ser menor que 1 con una probabilidad muy alta cuando los datos provienen de una población de extremos bivariados, con una estructura de dependencia negativa o positiva pero débil y que por azar se induce una dependencia negativa en la muestra.

3.3.1 Verosimilitud

En el esquema de verosimilitud, la información del modelo propuesto, que se encuentra en la muestra de cocientes $w_1, w_2, ...w_n$, está dada por la función de verosimilitud,

$$L(a; w_1, ..., w_n) = a^n \prod_{i=1}^{n} [2\min(w_i, 1-w_i)]^{a-1}, \ a > 0. \tag{3.17}$$

En forma equivalente, esta información se encuentra en la función log-verosimilitud,

$$l(a; w_1, ..., w_n) = n\log(a) + (a-1)\sum_{i=1}^{n} \log[2\min(w_i, 1-w_i)]. \tag{3.18}$$

Un estimador puntual del parámetro a, es el estimador de máxima verosimilitud, que se define como el valor de a que maximiza la función log-verosimilitud (3.18). En nuestro modelo, este estimador es,

$$\hat{a} = -n\left\{\sum_{i=1}^{n} \ln[2\min(w_i, 1-w_i)]\right\}^{-1}. \tag{3.19}$$

La información observada está dada por,

$$I(\hat{a}) = -\left.\frac{d^2 l(a)}{da^2}\right|_{a=\hat{a}} = \frac{n}{\hat{a}^2},$$

mientras que la varianza asintótica es,

$$V(\hat{a}) = a^2/n. \tag{3.20}$$

Como la función de densidad de probabilidades correspondiente al modelo propuesto tiene un comportamiento suave, entonces el estimador máximo-verosímil tiene una distribución asintótica normal, con media a y varianza $I^{-1}(a)$.

Las funciones de verosimilitud relativa $R(a) = L(a)/L(\hat{a})$ y log-verosimilitud relativa $r(a) = l(a) - l(\hat{a})$, dan una medida del apoyo de los datos a diferentes valores

del parámetro a. La función $-2\log R(a) = -2r(a)$ se distribuye asintóticamente como ji-cuadrada con un grado de libertad; de aquí obtenemos intervalos de verosimilitud-confianza. Un intervalo de verosimilitud-confianza $(1-\alpha)100\%$, se define como

$$\left\{a > 0 : -2r(a) \leq \chi^2_{(1-\alpha;1)}\right\},$$

donde $\chi^2_{(1-\alpha;1)}$ es el cuantil $(1-\alpha/2)$ de la distribución ji-cuadrada con un grado de libertad.

El límite izquierdo del intervalo de verosimilitud-confianza también puede ser menor que 1 cuando los datos tienen un patrón de dependencia positiva pero débil o negativa.

El estimador de máxima verosimilitud dado por (3.19), es un estimador sesgado, función del estadístico completo suficiente dado por (3.16), con esperanza y varianza,

$$\begin{aligned} E(\hat{a}) &= \left(\frac{n}{n-1}\right)a \\ V(\hat{a}) &= \left(\frac{n^2}{(n-1)^2(n-2)}\right)a^2. \end{aligned}$$

3.3.2 Estimación de a por el Método de Momentos

La estimación por el método de momentos, del parámetro a que posee el modelo propuesto (3.3), requiere de una ecuación donde se tienen las versiones muestral y poblacional, del momento del menor orden que permita su resolución. Como el modelo propuesto es una distribución simétrica con soporte $[0, 1]$, el primer momento $E(W)$ es la constante $1/2$, independiente del parámetro a. Requerimos entonces formular una ecuación con los momentos muestral y poblacional de orden 2. El segundo momento de la distribución (3.3), es

$$E(W^2) = \frac{1}{2} - \frac{a(a+3)}{4(a+1)(a+2)}.$$

Resolviendo la ecuación de segundos momentos $E(W^2) = (w_1^2 + w_2^2 + ... + w_n^2)/n$, tenemos el estimador de momentos para a,

$$\hat{a} = \frac{-3 + (1 + 2/s^2)^{1/2}}{2}, \tag{3.21}$$

donde s^2 es el segundo momento muestral centrado en la media poblacional $1/2,$.

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(w_i^2 - \frac{1}{4}\right).$$

Este estimador está determinado por la variabilidad muestral de los cocientes w_1, w_2,- ..., w_n, y podemos ver fácilmente que cuando el valor de s^2 coincide con la varianza $(1/12)$ de la distribución Uniforme$(0, 1)$, entonces $\hat{a}$ toma el valor 1, que corresponde a el caso de independencia. En cambio, cuando se tiene variabilidad muestral nula alrededor de la media poblacional $1/2$, entonces $\hat{a} = \infty$, lo cual corresponde a tener dependencia total positiva. En cambio si la distribución de los cocientes alrededor de la media poblacional $1/2$, es tal que el segundo momento muestral s^2 toma un valor mayor a la varianza $(1/12)$ de la distribución Uniforme$(0, 1)$, esto es una indicación de que los datos tienen una estructura de dependencia negativa, y el parámetro estimado $\hat{a}$ es menor que 1.

3.3.3 Enfoque Bayesiano

El espacio parametral del modelo propuesto en (3.3) se define en forma extendida, incluyendo valores menores que 1, con el fin de evitar problemas al hacer inferencias sobre el parámetro a. Cuando tenemos la seguridad de que la estructura de dependencia de los extremos bivariados es no negativa, y por tal razón queremos obtener una estimación de a, no menor que 1 y además, una estimación por intervalo que excluya valores menores que 1, podemos utilizar el esquema Bayesiano, para incluir la restricción de ser mayor o igual a 1 para el parámetro a.

Exploramos aquí el enfoque Bayesiano de inferencia acerca del parámetro del modelo propuesto. La información que permite estimar el parámetro o probar hipótesis estadísticas relacionadas con el modelo, se encuentra en la función de densidad de probabilidades a posteriori, dada por

$$f(a|w_1, w_2, ..., w_n) = \frac{f(w_1, w_2, ..., w_n|a)f_0(a)}{\int_A f(w_1, w_2, ..., w_n|a)f_0(a)da},$$

donde $f(w_1, w_2, ..., w_n|a)$ es la densidad conjunta de los datos $w_1, w_2, ..., w_n$ dado el valor a del parámetro, $f_0(a)$ es la función de densidad de probabilidades a priori.

El Teorema de Bayes nos permite combinar información a priori sobre el parámetro a con datos generados de acuerdo al modelo de probabilidad propuesto en (3.3), para obtener información del parámetro a, actualizada por los datos. Este conocimiento actualizado de a, lo tenemos en la función de densidad de a condicionada por los datos, $f(a|w_1, w_2, ..., w_n)$.

En nuestro caso la información de los datos ingresa vía la verosimilitud dada en (3.17), en donde combinamos los datos observados y la función de densidad (3.14) del modelo propuesto.

La información a priori sobre el parámetro a, se tiene en la función de densidad a priori $f_0(a)$. En la determinación de funciones de densidad a priori, dos consideraciones importantes son, la información previa que tenemos sobre el parámetro y la facilidad para implementar el paradigma Bayesiano.

La información que tenemos sobre a, en nuestro caso, es que la dependencia que se tiene entre los extremos X, Y es positiva. Antes de traducir esto en una función de densidad a priori, discutimos una familia de densidades a priori, motivada por conveniencia matemática, que permite obtener con facilidad la función de densidad a posteriori para a. Esta es la familia de densidades a priori conjugada.

Como la función de densidad de probabilidades (3.14) del modelo propuesto se encuentra en la familia exponencial de distribuciones, entonces se tiene una distribución a

priori conjugada que puede encontrarse con relativa facilidad. La forma exponencial de nuestra función de densidad de probabilidades es,

$$f(w;a) = g(a)b(w)\exp\{\phi(a)h(w)\},$$

donde $g(a) = a$, $b(w) = 1$, $\phi(a) = -(a-1)$, $h(w) = -2\log[\min(w, 1-w)]$. Esta familia exponencial es una familia regular ya que el soporte $[1,\infty)$ de la distribución no depende del parámetro a. Cuando la función de densidad de probabilidades es de la forma anterior, la distribución a priori está dada por Bernardo & Smith (1994),

$$f_0(a) = [K(\tau_0,\tau_1)]^{-1}[g(a)]^{\tau_0}\exp\{\phi(a)\tau_1\},$$

donde el vector (τ_0,τ_1) es tal que

$$K(\tau_0,\tau_1) = \int_1^\infty [g(a)]^{\tau_0}\exp\{\phi(a)\tau_1\}\,da < \infty.$$

En nuestro caso, la distribución a priori resulta ser una Gama(τ_0+1,τ_1) truncada por la izquierda en 1, con función de densidad,

$$f_0(a|\tau_0,\tau_1) = \left[\frac{(\tau_1)^{\tau_0+1}}{\Gamma(\tau_0+1)S(1,\tau_0+1,\tau_1)}\right]a^{\tau_0}\exp\{-\tau_1 a\}, \tag{3.22}$$

donde $S(1,\tau_0+1,\tau_1)$ es la integral de 1 a ∞ de la densidad Gama(τ_0+1,τ_1) no truncada.

La densidad a posteriori es,

$$f(a|\mathbf{w},\boldsymbol{\tau}) = \frac{a^n\left[\prod_{i=1}^n 2\min(w_i,1-w_i)\right]^{a-1}\left[\frac{(\tau_1)^{\tau_0+1}a^{\tau_0}\exp\{-\tau_1 a\}}{\Gamma(\tau_0+1)S(1,\tau_0+1,\tau_1)}\right]}{\int_1^\infty a^n\left[\prod_{i=1}^n 2\min(w_i,1-w_i)\right]^{a-1}\left[\frac{(\tau_1)^{\tau_0+1}a^{\tau_0}\exp\{-\tau_1 a\}}{\Gamma(\tau_0+1)S(1,\tau_0+1,\tau_1)}\right]da} \tag{3.23}$$

$$= \frac{a^{\tau_0+n}\exp\{-(t(w)+\tau_1)a\}}{\int_1^\infty a^{\tau_0+n}\exp\{-(t(w)+\tau_1)a\}\,da}$$

$$= \frac{[(t(w)+\tau_1)]^{\tau_0+n+1}a^{\tau_0+n}\exp\{-(t(w)+\tau_1)a\}}{\Gamma(\tau_0+n+1)S(1,\tau_0+n+1,\tau_1+t(w))},$$

donde $t(w)$ es el estadístico suficiente (3.16). Esta es la densidad de la distribución Gama$(\tau_0+n+1,\tau_1+t(w))$, truncada por la izquierda en 1. La esperanza y la varianza para la variable aleatoria $\mathbf{a}$ son,

$$E(\mathbf{a}|\mathbf{w},\boldsymbol{\tau}) = \left(\frac{\tau_0+n+1}{(t(w)+\tau_1)}\right)C_1, \text{ y} \tag{3.24}$$

$$V(\mathbf{a}|\mathbf{w},\boldsymbol{\tau}) = \left(\frac{\tau_0+n+1}{(t(w)+\tau_1)^2}\right)C_2,$$

donde,

$$C_1 = \left(\frac{S(1;\tau_0+n+1,t(w)+\tau_1)}{S(1;\tau_0+n,t(w)+\tau_1)}\right),$$

$$C_2 = \left\{n\left(\frac{S(1;\tau_0+n+2,t(w)+\tau_1)}{S(1;\tau_0+n,t(w)+\tau_1)}\right) - (n-1)\left(\frac{S(1;\tau_0+n+1,t(w)+\tau_1)}{S(1;\tau_0+n,t(w)+\tau_1)}\right)^2\right\}.$$

Si tomamos a $E(\mathbf{a}|\mathbf{w},\boldsymbol{\tau})$ como un estimador Bayesiano para el parámetro a, tenemos que este es una versión corregida del estimador máximo-verosímil, dado por $n/t(w)$. Este último se obtiene también considerando la versión no-informativa de la densidad Gama, donde los parámetros son, $\tau_0 = -1, \tau_1 = 0$.

Ahora determinamos la función de densidad a priori, de acuerdo a la información previa que tenemos sobre el parámetro de dependencia a. Anteriormente comentamos que entre los extremos X y Y se tiene dependencia positiva. Esta información expresada en términos de del coeficiente de asociación τ de Kendall, es $0 \le \tau \le 1$.

Por la relación (3.11) entre el parámetro a del modelo y el coeficiente de asociación τ, tenemos la relación aproximada, $\tau = 1 - 1/a$. Tomando a a y τ como variables

aleatorias, traducimos la información sobre la dependencia positiva entre los extremos X y Y, proponiendo la distribución a priori Uniforme$(0,1)$ para la τ de Kendall. Haciendo un cambio de variable obtenemos fácilmente la densidad de probabilidades a priori para a,

$$f(a) = a^{-2} 1_{[1,\infty)}(a). \tag{3.25}$$

Incluyendo esta densidad a priori, la densidad a posteriori, se obtiene,

$$f(a|(w_1,...,w_n)) \propto a^{n-2} \exp\{-at(w)\} 1_{[1,\infty)}(a).$$

La forma exacta de esta distribución truncada por la izquierda en 1, se obtiene normalizando la expresión de la densidad anterior, con la probabilidad acumulada de uno a infinito, que es la función de supervivencia evaluada en uno, $S(1; n-1, t(w))$, para la distribución Gama$(n-1, t(w))$. Así obtenemos la densidad a posteriori truncada,

$$f(a|(w_1,...,w_n)) = [1/S(1; n-1, t(w))] a^{n-2} \exp\{-at(w)\} 1_{[1,\infty)}(a). \tag{3.26}$$

Es interesante notar que obtenemos una distribución a posteriori de la misma familia conjugada encontrada anteriormente, con los parámetros $\tau_0 = -2$ y $\tau_1 = 0$.

En este caso la esperanza y la varianza de a están dados por

$$\begin{aligned} E(a|w) &= \frac{n-1}{t(w)} \left(\frac{S(1; n, t(w))}{S(1; n-1, t(w))} \right) \\ V(a|w) &= \frac{n-1}{t^2(w)} \left\{ n \left(\frac{S(1; n+1, t(w))}{S(1; n-1, t(w))} \right) - (n-1) \left(\frac{S(1; n, t(w))}{S(1; n-1, t(w))} \right)^2 \right\}. \end{aligned} \tag{3.27}$$

Un estimador puntual Bayesiano del parámetro a, es la esperanza $E(a|w)$, mientras que el estimador Bayesiano de confianza $100(1-\alpha)\%$ de longitud mínima, es el intervalo

(a_1, a_2) de máxima densidad posterior, discutido en Box y Tiao (1973), que se obtiene a partir de la densidad posterior $f(a|w)$, definiendo los límites a_1, a_2, por las restricciones,

$$\begin{aligned} \int_{a_1}^{a_2} f(a|w) da &= 1 - \alpha, \\ f(a_1|w) &= f(a_2|w). \end{aligned}$$

Teniendo la distribución a posteriori del parámetro a, podemos hacer predicciones sobre diferentes cantidades de interés o sobre observaciones futuras, a partir de las observaciones que se tienen.

La probabilidad conjunta a posteriori de exceso está dada por,

$$P(X > x, Y > y|w_1,...,w_n) = \int_0^\infty P(X > x, Y > y|a) f(a|w_1,...,w_n) da. \tag{3.28}$$

Esta estimación considera tanto la variabilidad del valor de a, como la variabilidad residual de X, Y, cuando a es conocido.

Obtenemos numéricamente la predicción anterior, generando la secuencia de realizaciones $a_1, a_2, ..., a_m$ con la distribución a posteriori de a, y calculamos la media de las probabilidades de exceso evaluadas para las diferentes realizaciones de a :

$$\frac{1}{m} \sum_{i=1}^{m} P(X > x, Y > y|a_i). \tag{3.29}$$

3.3.4 Comparación de estimadores

Antes de comparar los estimadores de a, $\hat{a}_L$ de máxima verosimilitud, $\hat{a}_M$ de momentos y $\hat{a}_B$ Bayesiano, comentamos que, mientras que $\hat{a}_L$ y $\hat{a}_B$ quedan definidos en términos del estadístico suficiente completo $T(W_1, W_2, ..., W_n)$, $\hat{a}_M$ es función de la varianza muestral

s^2. Por esta razón tiene mayor varianza el estimador $\hat{a}_M$. Además, fácilmente se obtiene la expresión de la varianza del estimador máximo verosímil, en cambio para la varianza de $\hat{a}_M$ no encontramos una expresión. Podemos estimar la varianza del estimador de momentos siguiendo un método como Jackknife.

La comparación que hacemos aquí de $\hat{a}_L$, $\hat{a}_M$ y $\hat{a}_B$, no va en el sentido de confrontar los tres estimadores, ya que mientras que los estimadores de máxima verosimilitud y de momentos consideran el espacio parametral extendido $(0, \infty)$, el estimador Bayesiano considera la restricción del espacio parametral, al caso de dependencia no negativa entre X y Y, esto es considera el espacio parametral $[1, \infty)$.

Comparamos los estimadores $\hat{a}_L$, $\hat{a}_M$ y $\hat{a}_B$ a través de un estudio de simulación, tomando muestras de tamaño 20 y 50, y considerando diferentes niveles de dependencia positiva. Como criterio de comparación utilizamos el sesgo y el error cuadrático medio de los estimadores. Obtenemos los parámetros de comparación a partir de 10000 muestras simuladas.

En el estudio de simulación, consideramos los casos de dependencia muy baja, baja, mediana y alta, tomando para a, los valores de $1.1, 1.5, 2.0$ y 10 respectivamente. De acuerdo a la relación (3.10), entre el parámetro a y la τ de Kendall, a estos valores de a le corresponden los coeficientes de asociación, $0.078, 0.291, 0.445$ y 0.871 respectivamente.

En la Tabla 3.2 tenemos los valores de sesgo y en la Tabla 3.3 tenemos los valores de error cuadrático medio obtenidos para los tres estimadores que se comparan. En cada una de las Figuras 3.5-3.8, se tienen las densidades estimadas para las 10000 estimaciones de a obtenidas con los tres estimadores estudiados, con muestras de tamaño 20. Las densidades se estimaron en forma no paramétrica (Silverman, B. W., 1986), utilizando la función *density* de S-plus (StatSci, 1996).

En la Tabla 3.2 observamos que los tres estimadores tienen un sesgo positivo, en los cuatro casos de dependencia considerados. Mientras que para los estimadores de máxima verosimilitud y de momentos, la diferencia es muy pequeña, el estimador Bayesiano muestra un sesgo menor.

Tabla 3.2. Sesgo para los estimadores de máxima verosimilitud, de momentos y Bayesiano, del parámetro a del modelo propuesto.

$Sesgo(\hat{a}_L)$ $Sesgo(\hat{a}_M)$ $Sesgo(\hat{a}_B)$	dep. muy baja $a = 1.1$	dep. baja $a = 1.5$	dep. media $a = 2.0$	dep. alta $a = 10.0$
$n = 20$	.058561 .048968 .173216	.076316 .072701 .060026	.110961 .111438 .020554	.500462 .758639 −.024561
$n = 50$	.023631 .018934 .087814	.030218 .025470 .009088	.048219 .049051 −.0022741	.193516 .283505 −.010355

Tabla 3.3. Error cuadrático medio para los estimadores de máxima verosimilitud, de momentos y Bayesiano, del parámetro a del modelo propuesto.

$ECM(\hat{a}_L)$ $ECM(\hat{a}_M)$ $ECM(\hat{a}_B)$	dep. muy baja $a = 1.1$	dep. baja $a = 1.5$	dep. media $a = 2.0$	dep. alta $a = 10.0$
$n = 20$	.079089 .096321 .059434	.141872 .166391 .095653	.260854 .290876 .210330	6.368880 7.873536 5.522475
$n = 50$	.027084 .031595 .018528	.050199 .054570 .042802	.135848 .144031 .128236	2.216371 2.569900 2.092745

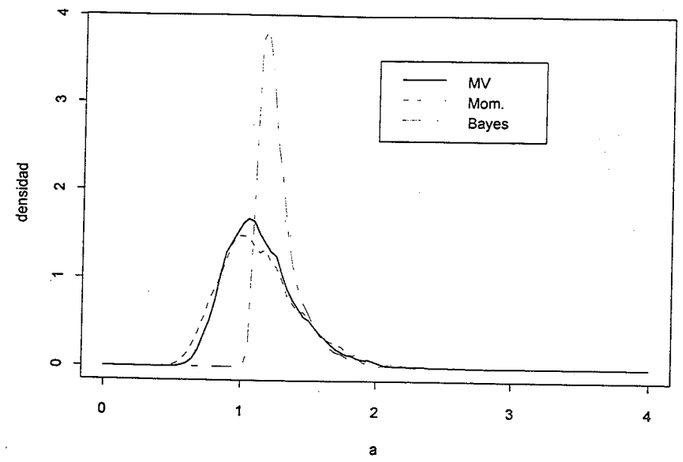


Figura 3.4. Densidades estimadas en forma no paramétrica, para las estimaciones de a obtenidas en el estudio de simulación realizado para comparar los estimadores, con muestras de tamaño 20 y dependencia muy baja, $a = 1.1$ y $\tau = 0.078$.

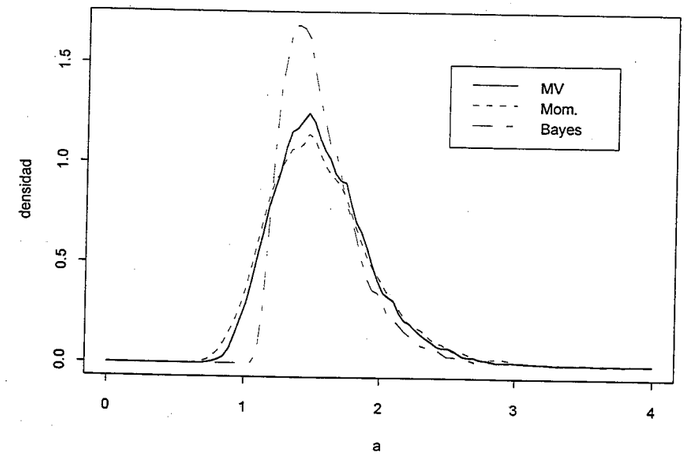


Figura 3.5. Densidades estimadas en forma no paramétrica, para las estimaciones de a obtenidas en el estudio de simulación realizado para comparar los estimadores, con muestras de tamaño 20 y dependencia baja, $a = 1.5$ y $\tau = 0.291$.

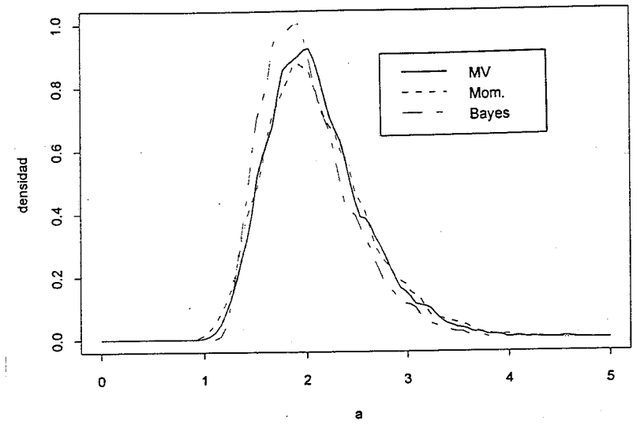


Figura 3.6. Densidades estimadas en forma no paramétrica, para las estimaciones de a obtenidas en el estudio de simulación realizado para comparar los estimadores, con muestras de tamaño 20 y dependencia media, $a = 2.0$ y $\tau = 0.445$.

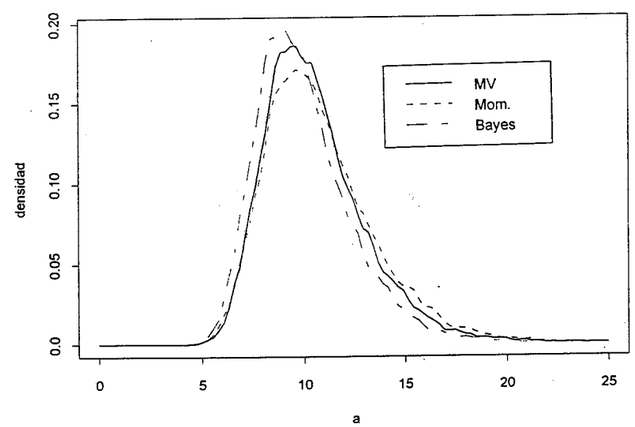


Figura 3.7. Densidades estimadas en forma no paramétrica, para las estimaciones de a obtenidas en el estudio de simulación realizado para comparar los estimadores, con muestras de tamaño 20 y dependencia alta, $a = 10.0$ y $\tau = 0.871$.

Esta disminución del sesgo, se explica por la relación existente entre los estimadores de máxima verosimilitud y Bayesiano, ya que el segundo es menor que el primero por un factor de correción menor que uno,

$$\hat{a}_B = \left(\frac{n-1}{n} \frac{S(1,n,t(w)}{S(1,n-1,t(w)} \right) \hat{a}_L. \tag{3.30}$$

El comportamiento del error cuadrático medio de los estimadores, mostrado en la Tabla 3.3, exhibe una menor variabilidad para el estimador Bayesiano. Esta estabilidad de $\hat{a}_B$, es mayor a medida que se reduce el nivel de dependencia, esto es, a medida que a está mas cerca de uno. Este comportamiento se debe a que la información de dependencia positiva, incorporado en el esquema Bayesiano, cancela la posibilidad de que las estimaciones Bayesianas de a tomen valores menores que uno, que corresponden a un patrón de dependencia negativa. Este acotamiento de $\hat{a}_B$ por la izquierda en uno, no se tiene para los otros dos estimadores $\hat{a}_L$ y $\hat{a}_M$, por lo tanto muestran mayor variabilidad en el caso de patrones de baja dependencia.

El comportamiento de los estimadores discutido en términos del sesgo y el error cuadrático medio, lo observamos en las Figuras 3.4-3.7, donde mostramos la densidad estimada para las estimaciones.

Apéndice: Funciones de densidad para otros modelos

Modelo logístico simétrico: la función de densidad del cociente W es,

$$f(w) = \frac{\theta \left[w\left(1-w\right) \right]^{\theta-1}}{\left[w^{\theta} + \left(1-w\right)^{\theta} \right]^2}.$$

Modelo logístico asimétrico: la función de densidad de probabilidades es,

$$f(w) = \frac{n'(w)d(w) - n(w)d'(w)}{d^2(w)},$$

donde,

$$\begin{aligned}
n(w) &= \phi^{\alpha} w^{\alpha} \left[p(w) \right]^{1/\alpha - 1} + (1-\phi)w \\
d(w) &= \left[p(w) \right]^{1/\alpha} + (\theta - \phi)w + 1 - \theta \\
n'(w) &= \phi^{\alpha} w^{\alpha-1} \left[p(w) \right]^{1/\alpha - 2} \left[\theta^{\alpha}(1-w)^{\alpha-1}(\alpha - w) + \phi^{\alpha} w^{\alpha} \right] + 1 - \phi \\
d'(w) &= \left[p(w) \right]^{1/\alpha - 1} \left[-\theta^{\alpha}(1-w)^{\alpha-1} + \phi^{\alpha} w^{\alpha-1} \right] + \theta - \phi \\
p(w) &= \theta^{\alpha}(1-w)^{\alpha} + \phi^{\alpha} w^{\alpha}
\end{aligned}$$

Esta función de densidad cuando $\theta = \phi$ corresponde a la densidad del modelo logístico simétrico.

Modelo bilogístico: la función de densidad del cociente es,

$$f(w) = \frac{(1-\alpha)(1-\beta)u(1-u)}{w(1-w)\left[(1-\alpha) + (\alpha - \beta)\, u \right]^2 \left[\alpha + (\beta - \alpha)\, u \right]},$$

donde como antes u es solución de la ecuación

$$(1-\alpha)(1-w)(1-u)^{\beta} - (1-\beta)wu^{\alpha} = 0.$$

Modelo logístico invertido: la función de densidad de probabilidades es:

$$f(w) = \frac{n'(w)d(w) - n(w)d'(w)}{[d(w)]^2},$$

donde

$$\begin{aligned}
n(w) &= w - (\phi_1 w)^{-\tau}\left\{(\phi_1 w)^{-\tau} + [\phi_2(1-w)]^{-\tau}\right\}^{-1/\tau-1}, \\
d(w) &= 1 - \left\{(\phi_1 w)^{-\tau} + [\phi_2(1-w)]^{-\tau}\right\}^{-1/\tau}, \\
n'(w) &= 1 - \phi_1^{-\tau}\left[\phi_1^{-\tau} w^{-2\tau-1} - \phi_2^{-\tau} w^{-\tau-1}(1-w)^{-\tau-1}(\tau+w)\right] p^{-1/\tau-2}(w) \\
d'(w) &= p^{-1/\tau-1}(w)\left[-\phi_1^{-\tau} w^{-\tau-1} + \phi_2^{-\tau}(1-w)^{-\tau-1}\right] \\
p(w) &= \left\{(\phi_1 w)^{-\tau} + [\phi_2(1-w)]^{-\tau}\right\}
\end{aligned}$$

Modelo de mezcla de betas: la función de densidad de probabilidades para el cociente W en el caso del modelo de mezcla de betas, es

$$f(w) = \left[1 - w(\alpha^{'}/\alpha)\right]\left[1 + (1-w)(\alpha^{'}/\alpha)\right] + w(1-w)(\alpha^{''}/\alpha)$$

donde,

$$\alpha(w) = (1-w) + w\sum_{i=1}^{m}\theta_i B(w, \nu_i, \nu-\nu_i) - \sum_{i=1}^{m}\left(\frac{\nu_i}{\nu}\right)\theta_i B(w, \nu_i+1, \nu-\nu_i),$$

$$\begin{aligned}
\alpha'(w) &= -1 + \sum_{i=1}^{m}\theta_i B(w, \nu_i, \nu-\nu_i) + w\sum_{i=1}^{m}\theta_i b(w, \nu_i, \nu-\nu_i) \\
&\quad - \sum_{i=1}^{m}\left(\frac{\nu_i}{\nu}\right)\theta_i b(w, \nu_i+1, \nu-\nu_i),
\end{aligned}$$

$$\begin{aligned}
\alpha''(w) &= 2\sum_{i=1}^{m}\theta_i b(w, \nu_i, \nu-\nu_i) + w\sum_{i=1}^{m}\theta_i b'(w, \nu_i, \nu-\nu_i) \\
&\quad - \sum_{i=1}^{m}\left(\frac{\nu_i}{\nu}\right)\theta_i b'(w, \nu_i+1, \nu-\nu_i).
\end{aligned}$$

Siendo B y b las funciones de distribución y de densidad beta respectivamente. Además, la derivada de la función de densidad está dada por,

$$b'(w, r, s) = \frac{[(r-1)+(2-r-s)]}{[\Gamma(r)\Gamma(s)/\Gamma(r+s)]}x^{r-2}(1-x)^{s-2}.$$

Para la optimización de la función log-verosimilitud, utilizamos un procedimiento de optimización con restricción, ya que el el vector de parámetros (ν_1, ν_2, ν), debe satisfacer, $0 < \nu_1 < \nu/2 < \nu_2 < \nu$.

Capítulo 4

Estimador no-paramétrico de la función de dependencia $A(w)$

4.1 Introducción

La distribución conjunta de valores extremos bivariados (X, Y) con marginales exponenciales unitarias, según Pickands (1981), tiene una representación (1.10) en donde el parámetro es la función de dependencia $A(w)$. De acuerdo a esta caracterización, el espacio paramétrico de la distribución es un espacio de dimensión infinita, constituido por todas las funciones convexas definidas en el intervalo $[0, 1]$, con la restricción adicional $\max(w, 1-w) \leq A(w) < 1, 0 \leq w \leq 1$.

Por la flexibilidad de la función de dependencia, es útil tener un estimador no-paramétrico para esta función, que que tenga un mínimo de restricciones y que permita a los datos determinar la forma de la función de dependencia.

Un estimador no paramétrico de la función de dependencia, nos permite analizar una muestra de valores extremos bivariados, extrayendo la información sobre la estructura de dependencia sin la contaminación debida a una restricción impuesta de un patrón de dependencia erróneo.

Además, podemos utilizar la estimación no paramétrica de la función de dependencia

como referencia para juzgar un modelo paramétrico de la estructura de dependencia.

Pickands (1981) fué el primero en proponer un estimador no-paramétrico de la función de dependencia, basándose en la proposición siguiente.

Proposición 4.1. Sean X y Y variables aleatorias con distribución bivariada de valores extremos, con marginales exponenciales unitarias. Para cualquier $0 < w < 1$, la variable aleatoria

$$Z = \min\left(\frac{X}{1-w}, \frac{Y}{w}\right), \tag{4.1}$$

se distribuye exponencialmente, con parámetro media $[A(w)]^{-1}$.

Demostración. Para la variable aleatoria Z definida por (4.1), tenemos

$$\begin{aligned} P(Z > z) &= P\{\min[X/(1-w), Y/w] > z\} = P\{X/(1-w) > z, Y/w > z\} \\ &= P\{X > z(1-w), Y > zw\} = \exp\{-[z(1-w)+zw]A(zw/[z(1-w)+zw])\} \\ &= \exp\{-zA(w)\}. \end{aligned}$$

Por lo tanto, la variable aleatoria Z definida por (4.1), tiene distribución exponencial con media $[A(w)]^{-1}$.

■

De la proposición anterior, tenemos que dado un valor $0 < w < 1$, de las parejas $(x_i, y_i), i = 1, 2, ..., n$ obtenemos la muestra $z_i, i = 1, 2, ..., n$ con $z_i = \min\{x_i/(1-w), y_i/w)$, que proviene de una distribución exponencial con media $[A(w)]^{-1}$. De aquí que la media muestral $(z_1 + z_2 + ... + z_n)/n$ estima la media $[A(w)]^{-1}$. Este razonamiento justifica el estimador de la función de dependencia propuesto por Pickands (1981),

$$\tilde{A}_n(w) = n\left\{\sum_{i=1}^{n} \min\left[x_i/(1-w), y_i/w\right]\right\}^{-1}, \; 0 \leq w \leq 1. \tag{4.2}$$

Este estimador converge con probabilidad 1 a la función de dependencia, pero con probabilidad 1, no es convexo. El autor del estimador propone salvar esta dificultad tomando el máximo minorante convexo de (4.2), de la siguiente manera.

Sea la muestra de valores extremos bivariada $(x_i, y_i), i = 1, 2, ..., n$. La información muestral, en el esquema de los cocientes, esta dada por $w_1, w_2, ..., w_n$. Considerando que la medida característica H que interviene en la definición (1.11) de A, solo tiene información en los puntos w_i, resulta de (1.11) que la función de dependencia es lineal en los intervalos (w_{i-1}, w_i).

Pickands (1981) propone tomar como estimador de A, el máximo minorante convexo $\bar{A}$, definido como $\bar{A}(w) = \max\{g(w) : g(w_i) \leq \tilde{A}_n(w_i), i = 1, 2, .., n;\ g$ es una función convexa y segmentada con puntos de cambio de pendiente dados por , $w_i = y_i/(x_i + y_i),\ i = 1, 2, ..., n\}$.

Smith et al (1990) proponen métodos de Kernel para estimación de densidades, para suavizar el estimador $\tilde{A}_n(w)$ propuesto por Pickands. En su forma más simple el estimador de kernel, está dado por,

$$\hat{A}(w, \lambda) = \frac{1}{\lambda}\int_0^1 \hat{A}(w) K\left(\frac{w-u}{\lambda}\right) du,$$

donde λ es un parámetro de suavizamiento y K es una función de kernel, que usualmente es una función de densidad de probabilidades simétrica alrededor de cero. El comportamiento de este estimador no se conoce bien aún para muestras grandes.

Posteriormente Tiago de Oliveira (1992) propuso un estimador no-paramétrico para la función de dependencia dado por,

$$\hat{A}(w) = 1 - \min\left\{\min(w, 1-w), \frac{1}{1+\log(n)}\sum_{i=1}^{n}\min\left(\frac{1-w}{1+nx_i}, \frac{w}{1+ny_i}\right)\right\}. \tag{4.3}$$

Este estimador converge muy lentamente en media cuadrática a la función de dependencia.

Los estimadores no-paramétricos que han sido propuestos hasta ahora se obtienen fundamentalmente a partir de (4.2). En la siguiente sección se propone otro estimador no-paramétrico de la función de dependencia definido en términos de los cocientes $w_i = y_i/(x_i + y_i)$, siguiendo la idea de sustituir la función de distribución del cociente $F(w)$, por la distribución empírica $F_n(w)$, en la caracterización de la función de dependencia dada en (3.2),

$$A(w) = \exp\left\{\int_0^w \frac{F(s) - s}{s(1-s)} ds\right\}. \tag{4.4}$$

En los primeros estimadores la información de la estructura de dependencia está dada en las parejas (x_i, y_i) mientras que en el nuevo estimador, la información proviene de los cocientes w_i.

El estimador no-paramétrico propuesto se comparará con los estimadores de Pickands (1981) y Tiago de Oliveira (1992), a través de un estudio de Monte Carlo, para muestras pequeñas.

4.2 Estimadores no paramétricos

La representación de la función de dependencia en términos de la función de distribución de los cocientes dada por la expresión (4.4), sugiere el estimador $\hat{A}_n(w)$, el cual se construye simplemente reemplazando $F(s)$ por la distribución empírica $F_n(s)$. Sin embargo esto no es suficiente, ya que la función de distribución empírica generalmente no satisface las restricciones necesarias para pertenecer a la familia de distribuciones $F(w)$ que hacen

de la función $\hat{A}_n(w)$ así obtenida una función de dependencia.

Sea (X_i, Y_i) una muestra aleatoria de la distribución bivariada de valores extremos con marginales exponenciales unitarias, y definimos $W_i = Y_i/(X_i + Y_i), i = 1, 2, ..., n$, como los cocientes. La distribución $F_n(w)$ representa entonces la función de distribución empírica de la muestra $w_1, w_2, ..., w_n$.

El estimador que se obtiene al sustituir $F(w)$ por $F_n(w)$, tiene buenas propiedades asintóticas. Este estimador que fué considerado recientemente por Capéraà, Fougerés y Genest (1997), no es convexo.

Definimos un estimador no-paramétrico $\hat{A}_1(w)$ de $A(w)$ sustituyendo en (4.4) la función de distribución F por la función de distribución empírica F_n, modificando además el integrando al tomar el supremo de el integrando en (4.4) una vez que se haya reemplazado F por F_n, para hacer de (4.4) una función convexa,

$$\hat{A}_1(w) = \exp\left\{\int_0^w \sup_{u \le s}\left[\frac{F_n(u) - u}{u(1-u)}\right] ds\right\}. \tag{4.5}$$

Obtenemos otro estimador $\hat{A}_2(w)$ de $A(w)$, utilizando ínfimo en lugar de el supremo e integrando de derecha a izquierda,

$$\hat{A}_2(w) = \exp\left\{-\int_w^1 \inf_{u \ge s}\left[\frac{F_n(u) - u}{u(1-u)}\right] ds\right\}. \tag{4.6}$$

Al tomar el supremo en el integrando de (4.5), evitamos oscilaciones en el integrando, que se traducen en obtener un estimador de A no convexo. En el caso de (4.6), al tomar el ínfimo en el integrando, tenemos también un integrando monótono, que nos dá un estimador convexo de la función de dependencia A.

Estos estimadores son convexos como se demuestra enseguida. Sin embargo, su comportamiento es diferente en las vecindades de cero y uno. El primer estimador tiene un buen comportamiento en la vecindad de cero, mientras que el segundo estimador se comporta bien en la vecindad de uno. Estas observaciones se hacen más precisas en la

siguiente sección, en donde inicialmente discutimos las propiedades de los estimadores $\hat{A}_1$ y $\hat{A}_2$.

4.3 Propiedades de los estimadores

La función de dependencia A es convexa en $(0, 1)$, y por lo tanto es deseable que cualquier estimador de A refleje esta propiedad. No es claro inmediatamente que los estimadores $\hat{A}_1$ y $\hat{A}_2$ definidos en (4.5) y (4.6) sean convexos. Esta sección examina algunas propiedades de $\hat{A}_1(w)$ y $\hat{A}_2(w)$. Para probar la convexidad de $\hat{A}_1(w)$ y $\hat{A}_2(w)$, utilizamos los siguientes resultados sobre funciones convexas.

Teorema 4.2. (Rockafeller, R.T, 1970). sea $f : (-\infty, +\infty)^n \to (-\infty, +\infty]$, una función convexa y sea $\varphi : (-\infty, +\infty) \to (-\infty, +\infty]$, una función convexa no decreciente, entonces $h(x) = \varphi(f(x))$, es una función convexa, tomando $\varphi(+\infty) = +\infty$.

Teorema 4.3. (Rockafeller, R.T, 1970) Sea $a \in (-\infty, +\infty)$, y sea $\varphi : (-\infty, +\infty) \to (-\infty, +\infty]$, una función no decreciente, tal que $\varphi(a)$ es finita. La función f dada por

$$f(x) = \int_a^x \varphi(t)dt,$$

es una función convexa bien definida.

Estos dos teoremas nos permiten demostrar de una manera breve, la convexidad de los estimadores de A, dados en (4.5) y (4.6) respectivamente. En el siguiente teorema, tenemos que las funciones que componen los estimadores definidos, tienen directamente las propiedades de las funciones mencionadas en los teoremas anteriores.

Teorema 4.4. Los estimadores $\hat{A}_1$ y $\hat{A}_2$ definidos por (4.5) y (4.6) respectivamente, son convexos en $(0, 1)$

Demostración. Consideramos primeramente $\hat{A}_1$. El estimador $\hat{A}_1$ es la composición de la función exponencial y la función integral

$$f(x) = \int_a^x \sup_{u \le s} \left[\frac{F_n(u) - u}{u(1-u)} \right] ds.$$

Como la función exponencial es una función convexa no decreciente, por el Teorema (4.2) es suficiente probar que la función integral $f(x)$, es convexa. Por el Teorema (4.3), esto último es cierto debido a que la función

$$\varphi(x) = \sup_{u \le s} \left[\frac{F_n(u) - u}{u(1-u)} \right], .0 \le s \le x,$$

es una función no decreciente.

Consideramos ahora $\hat{A}_2$. Como $\hat{A}_2$ es una composición de la función exponencial y la función

$$g(x) = \int_x^1 \inf_{u \ge s} \left[\frac{F_n(u) - u}{u(1-u)} \right] ds, \ 0 \le x \le 1,$$

es suficiente probar que $g(x)$ es una función convexa, para probar la convexidad de $\hat{A}_2$.

Haciendo el cambio de variable $x \to y = (1-x)$ tenemos,

$$g(x) = \int_{1-x}^0 - \inf_{u \ge 1-t} \left[\frac{F_n(u) - u}{u(1-u)} \right] dt = \int_0^{1-x} \inf_{u \ge 1-t} \left[\frac{F_n(u) - u}{u(1-u)} \right] dt.$$

El integrando

$$\inf_{u \ge 1-t} \left[\frac{F_n(u) - u}{u(1-u)} \right], \ 0 \le t \le 1,$$

es una función no decreciente debido a que el conjunto donde tomamos el ínfimo $[0, 1-t]$, decrece a medida que t crece de cero a uno. Entonces por el Teorema (4.3), $g(x)$ en una

función convexa así que $\hat{A}_2$ es una función convexa.

■

Notamos, como consecuencia de (1.11) y (1.12), que la función de dependencia A, además de ser convexa, también satisface las siguientes restricciones:

$$A(0) = A(1) = 1, \text{ y } \max(x, 1-x) \le A(x) \le 1, \text{ para } 0 \le x \le 1.$$

Luego es de interés, determinar cuando se cumplen propiedades similares para los estimadores $\hat{A}_1$ y $\hat{A}_2$. El siguiente teorema estudia estas propiedades para los estimadores $\hat{A}_1$ y $\hat{A}_2$.

Teorema 4.5. Sean $\hat{A}_1$ y $\hat{A}_2$ definidos por (4.5) y (4.6) respectivamente. Entonces,
(i) $\hat{A}_1(0) = \hat{A}_2(1) = 1$.
(ii) $\hat{A}_1(x) < 1$, para $x < X_{(1)}$, y $\hat{A}_2(x) < 1$ para $x \ge X_{(n)}$.
(iii) $\hat{A}_1(x) \ge 1 - x$, para $x < 1/2$, y $\hat{A}_2(x) \ge x$ para $1/2 \le x \le 1$.

Demostración. Para probar (i) observamos que para $x < X_{(1)}$,

$$\int_0^x \sup_{y \le t} \left(\frac{F_n(y) - y}{y(1-y)} \right) dt = \int_0^x \sup_{y \le t} \left(\frac{-1}{1-y} \right) dt = -x. \tag{4.7}$$

Entonces, $\lim_{x \to 0} \hat{A}_1(x) = 1$. Similarmente, para $x \ge X_{(n)}$,

$$\int_x^1 \inf_{y \ge t} \left(\frac{F_n(y) - y}{y(1-y)} \right) dt = 1 - x. \tag{4.8}$$

Luego, $\lim_{x \to 1} \hat{A}_2(x) = 1$. Notamos que (ii) es consecuencia inmediata de (4.7) y (4.8), y (iii) sigue de (i) y la convexidad de $\hat{A}_1$ y $\hat{A}_2$.

■

Sin embargo, los requisitos para los estimadores $\hat{A}_1$ y $\hat{A}_2$, de satisfacer la condición

$x \leq \hat{A}_1(x) \leq 1$, para $1/2 \leq x \leq 1$ y $1 - x \leq \hat{A}_2(x) \leq 1$, para $0 \leq x \leq 1/2$, no son satisfechos con probabilidad positiva, como se ilustra por los siguientes argumentos.

Solo analizamos a continuación el estimador $\hat{A}_1$. Para el estimador $\hat{A}_2$ se siguen argumentos similares. Escribimos, para $x > X_{(1)}$,

$$\begin{aligned}\int_0^x \sup_{y \leq t} \left(\frac{F_n(y) - y}{y(1-y)} \right) dt &= \int_0^{X_{(1)}} \sup_{y \leq t} \left(\frac{-1}{1-y} \right) dt + \int_{X_{(1)}}^x \sup_{y \leq t} \left(\frac{F_n(y) - y}{y(1-y)} \right) dt \\ &= -X_{(1)} + \int_{X_{(1)}}^x \sup_{y \leq t} \left(\frac{F_n(y) - y}{y(1-y)} \right) dt.\end{aligned}$$

Pero

$$\begin{aligned}\int_{X_{(1)}}^x \sup_{y \leq t} \left(\frac{F_n(y) - y}{y(1-y)} \right) dt &= \int_{X_{(1)}}^x \max \left\{ \sup_{y < X_{(1)}} \left(\frac{F_n(y) - y}{y(1-y)} \right), \sup_{X_{(1)} \leq y \leq t} \left(\frac{F_n(y) - y}{y(1-y)} \right) \right\} dt \\ &= \int_{X_{(1)}}^x \max \left\{ -1, \sup_{X_{(1)} \leq y \leq t} \left(\frac{F_n(y) - y}{y(1-y)} \right) \right\} dt \\ &\leq \int_{X_{(1)}}^x \max \left\{ -1, \sup_{X_{(1)} \leq y \leq t} \left(\frac{1}{y} \right) \right\} dt \\ &= \int_{X_{(1)}}^x \max \left\{ -1, \frac{1}{X_{(1)}} \right\} dt = \frac{x}{X_{(1)}} - 1.\end{aligned}$$

Luego,

$$\int_0^x \sup_{y \leq t} \left(\frac{F_n(y) - y}{y(1-y)} \right) dt \leq -X_{(1)} - 1 + \frac{x}{X_{(1)}}. \tag{4.9}$$

Haciendo $x = 1$ en la expresión anterior, es claro que el lado derecho de (4.9) es menor que cero con probabilidad positiva, y entonces $\hat{A}_1(1) < 1$ con probabilidad positiva.

También notamos que

$$\begin{aligned}\int_0^1 \sup_{y \leq t} \left(\frac{F_n(y) - y}{y(1-y)} \right) dt &= \int_0^{X_{(1)}} \sup_{y \leq t} \left(\frac{F_n(y) - y}{y(1-y)} \right) dt + \int_{X_{(1)}}^{X_{(n)}} \sup_{y \leq t} \left(\frac{F_n(y) - y}{y(1-y)} \right) dt \\ &\quad + \int_{X_{(n)}}^1 \sup_{y \leq t} \left(\frac{F_n(y) - y}{y(1-y)} \right) dt \\ &\geq \int_0^{X_{(1)}} \frac{-dt}{1-t} + \int_{X_{(1)}}^{X_{(n)}} \frac{\frac{1}{n} - X_{(1)}}{X_{(1)}(1 - X_{(1)})} dt + \int_{X_{(n)}}^1 \sup_{X_{(n)} \leq y \leq t} \left(\frac{F_n(y) - y}{y(1-y)} \right) dt \\ &\geq \ln(1 - X_{(1)}) + \frac{\left(\frac{1}{n} - X_{(1)}\right)\left(X_{(n)} - X_{(1)}\right)}{X_{(1)}(1 - X_{(1)})} - \ln(X_{(n)}).\end{aligned}$$

Luego, si $X_{(1)} + X_{(n)} < 1$ y $X_{(1)} < 1/n$, entonces $\hat{A}_1(1) > 1$.

A pesar de los posibles inconvenientes de que $\hat{A}_1(1) > 1$ y $\hat{A}_1(1) < 1$ con probabilidad positiva, puede probarse que bajo suposiciones adecuadas acerca de las colas derecha e izquierda de la distribución F del cociente $W = Y/(X + Y)$, los estimadores $\hat{A}_1$ y $\hat{A}_2$ tienen la propiedad de consistencia en subintervalos adecuados de $[0, 1]$ cuando $\log A(x)$ es convexa. Por ejemplo usando el Teorema 1.3, parte (iii) en la pagina 266 de Csörgő y Horvath (1993), puede probarse la consistencia uniforme débil de los estimadores en intervalos adecuados $[a_n, 1 - b_n]$, donde $a_n \to 0$, $b_n \to 0$, $na_n \to \infty$, y $nb_n \to \infty$. No obstante, una restricción adicional para las colas (en cero y uno) de la función de distribución F del cociente, garantiza la convergencia uniforme fuerte en $[0, 1]$. Esto se tiene en el siguiente teorema.

Teorema 4.6. Sea la función de dependencia $A(x)$ log convexa, y suponga que las colas izquierda y derecha de F se comportan como las colas de una distribución Beta con parámetros $1 + \delta$ y $1 + \delta^*$ respectivamente. Esto es, $F(y) \backsim y^{1+\delta}$ cuando $y \to 0$,

$1-F(y) \backsim (1-y)^{1+\delta^*}$ cuando $y \to 1$, para algunos $\delta, \delta^* > 0$. Entonces,

$$\sup_{0<x<1} \left|\hat{A}_1(x) - A(x)\right| \to 0, \quad \text{y} \quad \sup_{0<x<1} \left|\hat{A}_2(x) - A(x)\right| \to 0,$$

cuando $n \to \infty$, con probabilidad uno.

Demostración. Consideramos solo el caso de $\hat{A}_1$. Para $\hat{A}_2$, la consistencia uniforme fuerte se prueba de manera similar.

Como $\sup_{0<x<1} \left|\hat{A}_1(x) - A(x)\right| \leq \sup_{0<x<1} \left|\frac{\hat{A}_1(x)}{A(x)} - 1\right|$, y además $\frac{F(y)-y}{y(1-y)}$ es no decreciente por ser A log-convexa, es suficiente probar que

$$\sup_{0<x<1} \left| \exp\left\{ \int_0^x \sup_{0\leq y\leq t} \left(\frac{F_n(y)-y}{y(1-y)}\right) dt - \int_0^x \sup_{0\leq y\leq t} \left(\frac{F(y)-y}{y(1-y)}\right) dt \right\} - 1 \right|$$

converge a cero con probabilidad uno, y por consiguiente es suficiente probar que

$$\sup_{0<x<1} \left| \int_0^x \sup_{0\leq y\leq t} \left(\frac{F_n(y)-y}{y(1-y)}\right) dt - \int_0^x \sup_{0\leq y\leq t} \left(\frac{F(y)-y}{y(1-y)}\right) dt \right| \to 0$$

cuando $n \to \infty$, con probabilidad uno. Pero,

$$\left| \int_0^x \sup_{0\leq y\leq t} \left(\frac{F_n(y)-y}{y(1-y)}\right) dt - \int_0^x \sup_{0\leq y\leq t} \left(\frac{F(y)-y}{y(1-y)}\right) dt \right| \leq \int_0^x \sup_{0\leq y\leq t} \left(\frac{F_n(y)-F(y)}{y(1-y)}\right) dt,$$

siguiendo esta última desigualdad de Lema 1 en Rojo (1998). La última integral es menor que $\sup_{0\leq y\leq 1} \left(\frac{|F_n(y)-F(y)|}{y(1-y)}\right)$. Como en Capéraà *et al (1997)*, utilizamos un resultado de Lai (1974) como sigue: Para $\max\left(\frac{1}{1+\delta}, \frac{1}{1+\delta^*}\right) < \alpha < 1$, tenemos

$$\begin{aligned} \sup_{0\leq y\leq 1} \left(\frac{|F_n(y)-F(y)|}{y(1-y)}\right) &= \sup_{0\leq y\leq 1} \left(\frac{|F_n(y)-F(y)|}{(F(y)(1-F(y)))^\alpha} \frac{(F(y)(1-F(y)))^\alpha}{y(1-y)}\right) \\ &\leq \sup_{0\leq y\leq 1} \frac{|F_n(y)-F(y)|}{(F(y)(1-F(y)))^\alpha} \sup_{0\leq y\leq 1} \frac{(F(y))^\alpha((1-F(y)))^\alpha}{y(1-y)}. \end{aligned}$$

Como $F(y) \backsim y^{1+\delta}$ cuando $y \to 0$, $1-F(y) \backsim (1-y)^{1+\delta^*}$ cuando $y \to 1$, entonces $\frac{(F(y))^\alpha((1-F(y)))^\alpha}{y(1-y)}$ es acotado en vecindades de cero y uno, y como $\sup_{0\leq y\leq 1} \frac{|F_n(y)-F(y)|}{(F(y)(1-F(y)))^\alpha}$ converge a cero con probabilidad uno (Lai (1974)), se sigue el resultado. ■

Sin embargo, aunque como demostramos en el Teorema (4.6), $\hat{A}_1$ y $\hat{A}_2$ son fuertemente consistentes uniformemente en $[0,1]$, optamos por considerar una leve modificación de $\hat{A}_1$ y $\hat{A}_2$. Corrigiendo el comportamiento de $1/[y \times (1-y)]$ en las vecindades de cero y uno, es posible demostrar convergencia uniforme fuerte en subintervalos adecuados de $(0,1)$, sin la suposición de colas tipo beta. Esta es la principal motivación para los siguientes estimadores que son leves modificaciones de $\hat{A}_1$ y $\hat{A}_2$. Sea $a_n \to 0$, a una tasa que se especifica mas adelante, y definimos

$$\hat{A}_3(x) = \begin{cases} 1-x, & x < a_n \\ (1-a_n)\exp\left\{ \int_{a_n}^x \sup_{a_n\leq y\leq t} \left(\frac{F_n(y)-y}{y(1-y)}\right) dt \right\}, & x \geq a_n \end{cases} \tag{4.10}$$

y

$$\hat{A}_4(x) = \begin{cases} (1-a_n)\exp\left\{ -\int_x^{1-a_n} \inf_{t\leq y\leq 1-a_n} \left(\frac{F_n(y)-y}{y(1-y)}\right) dt \right\}, & x < 1-a_n \\ x, & x \geq 1-a_n \end{cases} \tag{4.11}$$

Notamos que $\hat{A}_3$ y $\hat{A}_4$ son continuos. También para $x \geq a_n$,

$$\hat{A}_3(x) = \sup_{a_n\leq y\leq x} \left(\frac{F_n(y)-y}{y(1-y)}\right) \exp\left\{ \int_{a_n}^x \sup_{a_n\leq y\leq t} \left(\frac{F_n(y)-y}{y(1-y)}\right) dt \right\} (1-a_n).$$

Luego,

$$\lim_{x \downarrow a_n} \hat{A}_3'(x) = \left(\frac{F_n(a_n)-a_n}{a_n(1-a_n)}\right)(1-a_n) \geq -1.$$

El resultado $\lim_{x\uparrow 1-a_n} \hat{A}_4'(x) \leq 1$ se obtiene por argumento similar. Luego, para probar la convexidad de $\hat{A}_3$ y $\hat{A}_4$, es suficiente mostrar que $\hat{A}_3'$ y $\hat{A}_4'$ son crecientes en $(a_n, 1)$ y $(0, 1-a_n)$ respectivamente. Esto se hace usando un argumento similar al ya utilizado para probar la convexidad de $\hat{A}_1$ y $\hat{A}_2$, que por lo tanto no repetimos aquí. La convergencia uniforme fuerte de $\hat{A}_3$ y $\hat{A}_4$ también se establece en el siguiente teorema.

Teorema 4.7. Sean $\hat{A}_3$ y $\hat{A}_4$ definidos por (4.10) y (4.11) respectivamente y suponga que $\ln A(x)$ es convexa. Suponga que $a_n \to 0$ con $\sqrt{n}a_n/(\log\log n) \to \infty$ cuando $n \to \infty$ y $\sum n a_n^{1/\alpha} < \infty$ para algún $0 < \alpha < 1$. Entonces, con probabilidad uno,

$$\sup_{0\leq x\leq 1-a_n} \left|\hat{A}_3(x) - A(x)\right| \to 0 \text{ y } \sup_{a_n\leq x\leq 1} \left|\hat{A}_4(x) - A(x)\right| \to 0$$

cuando $n \to \infty$.

Demostración. Nótese que

$$\begin{aligned}\sup_{0\leq x\leq a_n} \left|\hat{A}_3(x) - A(x)\right| &= \sup_{0\leq x\leq a_n} |1 - x - A(x)| \\ &\leq \max\left\{\sup_{0\leq x\leq a_n} |1 - A(x)|, a_n\right\} \to 0 \text{ como } n \to \infty.\end{aligned}$$

Entonces es suficiente considerar el caso $x \geq a_n$. Ahora, para $x \geq a_n$,

$$\begin{aligned}\hat{A}_3(x) - A(x) \leq\ & (1-a_n)\left|\int_{a_n}^{x} \sup_{a_n\leq y\leq t}\left(\frac{F_n(y)-y}{y(1-y)}\right)dt\right. \\ & \left. - \int_0^x \left(\frac{F(t)-t}{t(1-t)}\right)dt\right| \exp(\xi) + a_n \exp\left\{\int_0^x \left(\frac{F(t)-t}{t(1-t)}\right)dt\right\},\end{aligned} \tag{4.12}$$

donde ξ se encuentra entre $\int_{a_n}^{x} \sup_{a_n\leq y\leq t}\left(\frac{F_n(y)-y}{y(1-y)}\right)dt$ y $\int_0^x \left(\frac{F(t)-t}{t(1-t)}\right)dt$.

Puesto que $a_n \to 0$ y $\exp\left\{\int_0^x \left(\frac{F(t)-t}{t(1-t)}\right)dt\right\} = A(x)$, con $0 \leq A(x) \leq 1$ para todo x, se sigue que el último término en (4.12) converge uniformemente a 0. También puesto que ξ

se encuentra entre $\int_{a_n}^{x} \sup_{a_n\leq y\leq t}\left(\frac{F_n(y)-y}{y(1-y)}\right)dt$ y $\int_0^x \left(\frac{F(t)-t}{t(1-t)}\right)dt$, para probar la convergencia uniforme fuerte de $\hat{A}_3$ en $(0, 1-a_n)$, es suficiente probar que

$$\sup_{a_n\leq x\leq 1-a_n}\left|\int_{a_n}^{x} \sup_{a_n\leq y\leq t}\left(\frac{F_n(y)-y}{y(1-y)}\right)dt - \int_0^x \left(\frac{F(t)-t}{t(1-t)}\right)dt\right| \to 0$$

con probabilidad uno como $n \to \infty$, y por lo tanto es suficiente probar que

$$\sup_{a_n\leq x\leq 1-a_n}\left|\int_{a_n}^{x} \sup_{a_n\leq y\leq t}\left(\frac{F_n(y)-y}{y(1-y)}\right)dt - \int_{a_n}^x \left(\frac{F(t)-t}{t(1-t)}\right)dt\right| \to 0$$

con probabilidad uno como $n \to \infty$, ya que $\int_0^{a_n} \left(\frac{F(t)-t}{t(1-t)}\right)dt = \ln A(a_n) \to 0$.

Como $\ln A(x)$ es convexo, entonces $\frac{F(y)-y}{y(1-y)}$ es creciente y luego

$$\frac{F(y)-y}{y(1-y)} = \sup_{a_n\leq x\leq y} \frac{F(x)-x}{x(1-x)}.$$

Entonces,

$$\begin{aligned}&\sup_{a_n\leq x\leq 1-a_n}\left|\int_{a_n}^{x} \sup_{a_n\leq y\leq t}\left(\frac{F_n(y)-y}{y(1-y)}\right)dt - \int_{a_n}^x \left(\frac{F(t)-t}{t(1-t)}\right)dt\right| \\ =\ &\sup_{a_n\leq x\leq 1-a_n}\left|\int_{a_n}^{x} \sup_{a_n\leq y\leq t}\left(\frac{F_n(y)-y}{y(1-y)}\right)dt - \int_{a_n}^x \sup_{a_n\leq y\leq t}\left(\frac{F(y)-y}{y(1-y)}\right)dt\right| \\ \leq\ &\sup_{a_n\leq x\leq 1-a_n}\int_{a_n}^{x}\left|\sup_{a_n\leq y\leq t}\left(\frac{F_n(y)-y}{y(1-y)}\right) - \sup_{a_n\leq y\leq t}\left(\frac{F(y)-y}{y(1-y)}\right)\right|dt \\ \leq\ &\sup_{a_n\leq x\leq 1-a_n}\int_{a_n}^{x}\left|\sup_{a_n\leq y\leq t}\left(\frac{F_n(y)-F(y)}{y(1-y)}\right)\right|dt \\ \leq\ &O\left(n^{-1/2}\log\log n\right)\sup_{a_n\leq x\leq 1-a_n}\int_{a_n}^{x}\sup_{a_n\leq y\leq t}\left(\frac{1}{y(1-y)}\right)dt,\end{aligned}$$

donde la última desigualdad se cumple con probabilidad uno. Pero,

$$\begin{aligned}\sup_{a_n\le x\le 1-a_n}\int_{a_n}^{x}\sup_{a_n\le y\le t}\left(\frac{1}{y(1-y)}\right)dt &\le \int_{a_n}^{1-a_n}\sup_{a_n\le y\le 1-a_n}\left(\frac{1}{y(1-y)}\right)dt \\ &\le (1-2a_n)\max\left(\frac{1}{a_n(1-a_n)},\frac{1}{a_n(1-a_n)}\right).\end{aligned}$$

Así, tenemos que

$$\sup_{a_n\le x\le 1-a_n}\left|\int_{a_n}^{x}\sup_{a_n\le y\le t}\left(\frac{F_n(y)-y}{y(1-y)}\right)dt-\int_{a_n}^{x}\left(\frac{F(t)-t}{t(1-t)}\right)dt\right|\le\frac{O\left(n^{-1/2}\log\log n\right)}{a_n}\frac{(1-2a_n)}{1-a_n},$$

y el resultado es una consecuencia de el hecho que $O\left(n^{-1/2}\log\log n\right)/a_n\to 0$.

Con un argumento similar puede probarse que $\sup_{a_n\le x\le 1}\left|\hat{A}_4(x)-A(x)\right|\to 0$ con probabilidad uno, cuando $n\to\infty$. ■

4.4 Combinación de estimadores

El estimador $\hat{A}_1(w)$ dado en (4.5) es convexo y tiene un buen comportamiento en la parte izquierda del intervalo $[0,1]$, mientras que $\hat{A}_2(w)$ dado en (4.6) además de ser convexo se comporta bien en la parte derecha del intervalo unitario.

Para obtener un estimador no paramétrico de la función de dependencia que se comporte bien tanto del lado izquierdo como del derecho, combinamos los estimadores $\hat{A}_1$ y $\hat{A}_2$, tomando del primero, la información de dependencia en el lado izquierdo del intervalo $[0,1]$, y del segundo la información correspondiente al lado derecho, esto es, tomamos de cada uno lo mejor.

Así, para obtener un estimador que además de ser convexo, se comporte bien en ambos lados, combinamos los dos estimadores $\hat{A}_1(w)$ y $\hat{A}_2(w)$ de la siguiente manera:

Primeramente definimos la función $\tilde{A}(w)$ como,

$$\tilde{A}(w)=\begin{cases}\hat{A}_1(w), & 0\le w<1/2,\\ \\ \hat{A}_2(w), & 1/2<w\le 1.\end{cases}$$

Finalmente definimos $\hat{A}(w)$ como el máximo minorante convexo de $\tilde{A}(w)$, como de describe a continuación.

Sean w_i y w_d los puntos donde $\hat{A}_1(w)$ y $\hat{A}_2(w)$ alcanzan su valor mínimo. Si $\hat{A}_1(w_i)<\hat{A}_2(w_d)$, entonces $\hat{A}(w)=\hat{A}_1(w)$ para $0\le w\le w_i$. Enseguida tomamos el punto w_0 tal que $\hat{A}_2'(w_0)$ es igual a la pendiente del segmento que une los puntos $(w_i,\hat{A}_1(w_i))$ y $(w_0,\hat{A}_2(w_0))$. Definimos entonces $\hat{A}(w)$ como el segmento anterior para $w_i\le w\le w_0$ y como $\hat{A}_2(w)$ para $w_0\le w\le 1$.

Y si $\hat{A}_1(w_i)>\hat{A}_2(w_d)$, entonces $\hat{A}(w)=\hat{A}_2(w)$ para $w_d\le w\le 1$. Tomando ahora el punto w_0 tal que $\hat{A}_1'(w_0)$ es igual a la pendiente del segmento que une los puntos $(w_0,\hat{A}_1(w_0))$ y $(w_d,\hat{A}_2(w_d))$. Definimos entonces $\hat{A}(w)$ como el segmento anterior para $w_0\le w\le w_d$ y como $\hat{A}_1(w)$ para $0\le w\le w_0$.

En la sección final de este capítulo se da el algoritmo para obtener el estimador $\hat{A}(w)$, que se ha programado en el lenguaje Splus, y que se utilizó para las comparaciones que se presentan en la siguiente sección.

4.5 Comparación con otros estimadores en muestras pequeñas

Se realizó un estudio de simulación para comparar el estimador no-paramétrico $\hat{A}(w)$ de la función de dependencia propuesto en la sección anterior, con los estimadores no-paramétricos de Pickands (1981) y Tiago de Oliveira (1992).

Se compararon los estimadores tomando como criterio las funciones de sesgo y error

cuadrático medio En el estudio se simularon tres patrones de dependencia, dados por tres formas de la función de dependencia: simétrica, asimétrica y plana en la parte central. Además en cada caso, se consideraron tres niveles de dependencia: bajo, mediano y alto.

Para simular los patrones simétrico y asimétrico, se utilizó el modelo logístico asimétrico de Tawn (1988),

$$A_l(w) = \{\theta^\alpha(1-w)^\alpha + \phi^\alpha w^\alpha\}^{1/\alpha} + (\theta-\phi)w + 1 - \theta, \;\; 0 \leq \theta, \phi \leq 1, \alpha \geq 1,$$

que por su flexibilidad cubre una amplia gama de situaciones de dependencia.

En la Tabla 4.1 se muestran los parámetros utilizados para la generar las primeras tres funciones de dependencia simétrica, mismas que se muestran el la Figura 4.1. Aquí tenemos el caso simétrico del modelo logístico, ya que $\theta = \phi = 1$.

Los parámetros utilizados en el caso de las las funciones de dependencia asimétricas, generadas con el modelo logístico asimétrico, se encuentran en la Tabla 4.2. En la Figura 4.2 se tienen estas funciones asimétricas.

Las funciones de dependencia planas en la parte central se generaron con el modelo de mezcla de dos betas con parámetros $(\nu_1, \nu-\nu_1)$ y $(\nu_2, \nu-\nu_2)$, discutido en el Capítulo 2.

Podemos interpretar estas funciones de dependencia planas como casos de independencia local, en la parte central del intervalo $(0,1)$.

Tabla 4.1. Valores de los parámetros del modelo logístico asimétrico correspondientes a los casos de dependencia mencionados.

Situación\Parámetros	θ	ϕ	α	τ de Kendall
Dependencia Baja Simétrica	1	1	1.2	.167
Dependencia Media Simétrica	1	1	2	.500
Dependencia Alta Simétrica	1	1	5	.800

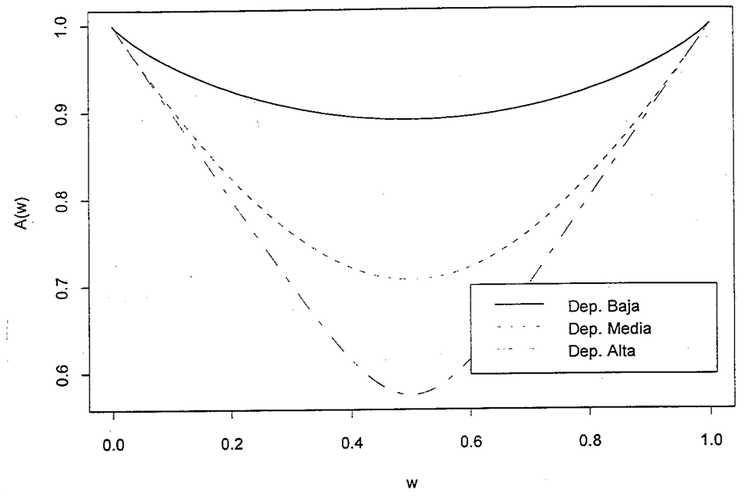


Figura 4.1. Funciones de dependencia simétrica utilizadas en la comparación de estimadores no paramétricos.

En la Tabla 4.3 se tienen los valores de los parámetros utilizados para simular también en este caso tres niveles de dependencia, bajo, mediano y fuerte. En la Figura 4.3 tenemos las gráficas de las funciones de dependencia correspondientes a estos parámetros.

Tabla 4.2. Valores de los parámetros del modelo logístico asimétrico correspondientes a los casos de dependencia mencionados.

Situación\Parámetros	θ	ϕ	α	τ de Kendall
Dependencia Baja Asimétrica	.1	.5	7	.089
Dependencia Media Asimétrica	.3	.6	7	.237
Dependencia Alta Asimétrica	.5	1	5	.439

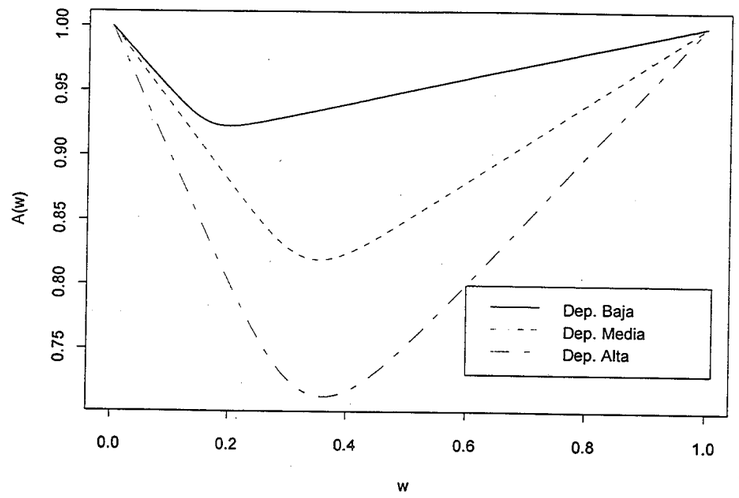


Figura 4.2. Funciones de dependencia asimétrica utilizadas en la comparación de los estimadores no parametricos.

Con el coeficiente de asociación τ de Kendall (3.9), se midió el nivel de asociación entre los valores extremos X y Y para los patrones de dependencia considerados en este estudio. En las Tablas 4.1 - 4.3 se muestran los valores del coeficiente de asociación correspondiente a cada uno de los patrones de dependencia.

Tabla 4.3. Valores de los parámetros del modelo mezcla de betas correspondientes a los casos de dependencia mencionados.

Situación/Parámetros	v_1	v_2	v	τ de Kendall
Dependencia Baja Simétrica	20	180	200	.205
Dependencia Media Simétrica	50	150	200	.492
Dependencia Alta Simétrica	80	120	200	.774

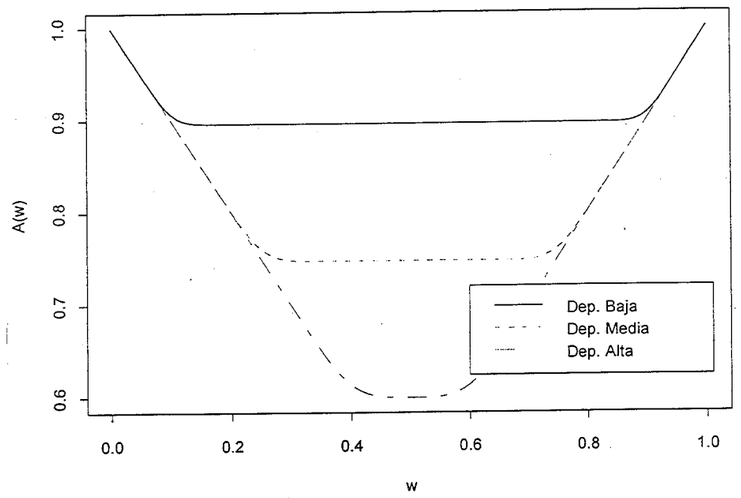


Figura 4.3. Funciones de dependencia planas utilizadas en la comparación de estimadores no parametricos de la función de dependencia.

Para cada estructura de dependencia se simularon 10,000 muestras de tamaño 20, de observaciones bivariadas, $(x_i, y_i), i = 1, 2, ..., 20$, con el algoritmo dado en el apéndice final del trabajo. Para cada muestra simulada se estimó la función de dependencia, con los estimadores no-paramétricos de Pickands (1981), de Tiago de Oliveira (1992) y el propuesto en este capítulo.

En la Figura 4.4 presentamos las funciones de sesgo y error cuadrático medio, correspondientes a los estimadores comparados, para la estructura de dependencia simétrica. En esta figura observamos lo siguiente:

El estimador de Pickands tiene un mal comportamiento cuando hay un nivel bajo de dependencia, mejora cuando el nivel de dependencia aumenta. El estimador que tiene el peor comportamiento es el de Tiago de O., ya que en un nivel bajo de dependencia

la sobrestima y al aumentar la dependencia, tiende a subestimarla en forma creciente. Nuestro estimador se encuentra en una posición intermedia, como lo notamos en las funciones de error cuadrático medio. En la función de sesgo para el caso de fuerte dependencia, observamos que subestima a la función de dependencia en la parte central.

En la Figura 4.5 se tienen las funciones de sesgo y error cuadrático medio, para los casos de dependencia asimétrica. Aquí observamos lo siguiente:

El estimador de Tiago de Oliveira tiene un mal comportamiento en los casos de menor y mayor nivel de dependencia. En estos dos casos de dependencia asimétrica, nuestro estimador es el que tiene un mejor comportamiento. No encontramos aquí un estimador que sea mejor uniformemente, ya que las funciones de error cuadrático medio se cruzan.

En el caso de la estructura de dependencia generada con las funciones de dependencia planas en la parte central, encontramos resultados mostrados en la Figura 4.6, que son similares a los encontrados en el caso de dependencia simétrica.

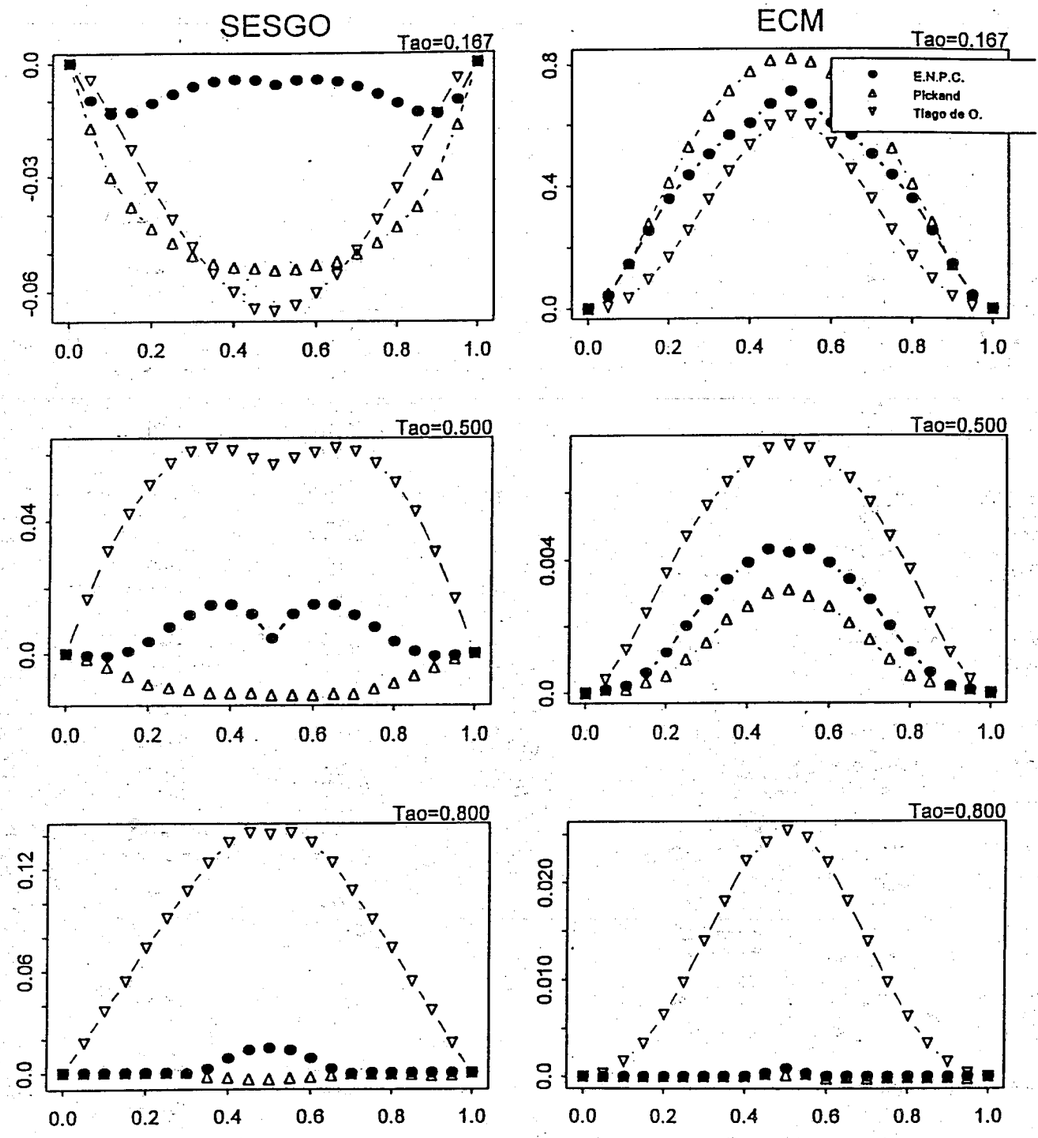


Figura 4.4. Funciones de sesgo y error cuadrático medio para los casos de dependencia baja, media y alta simétrica.

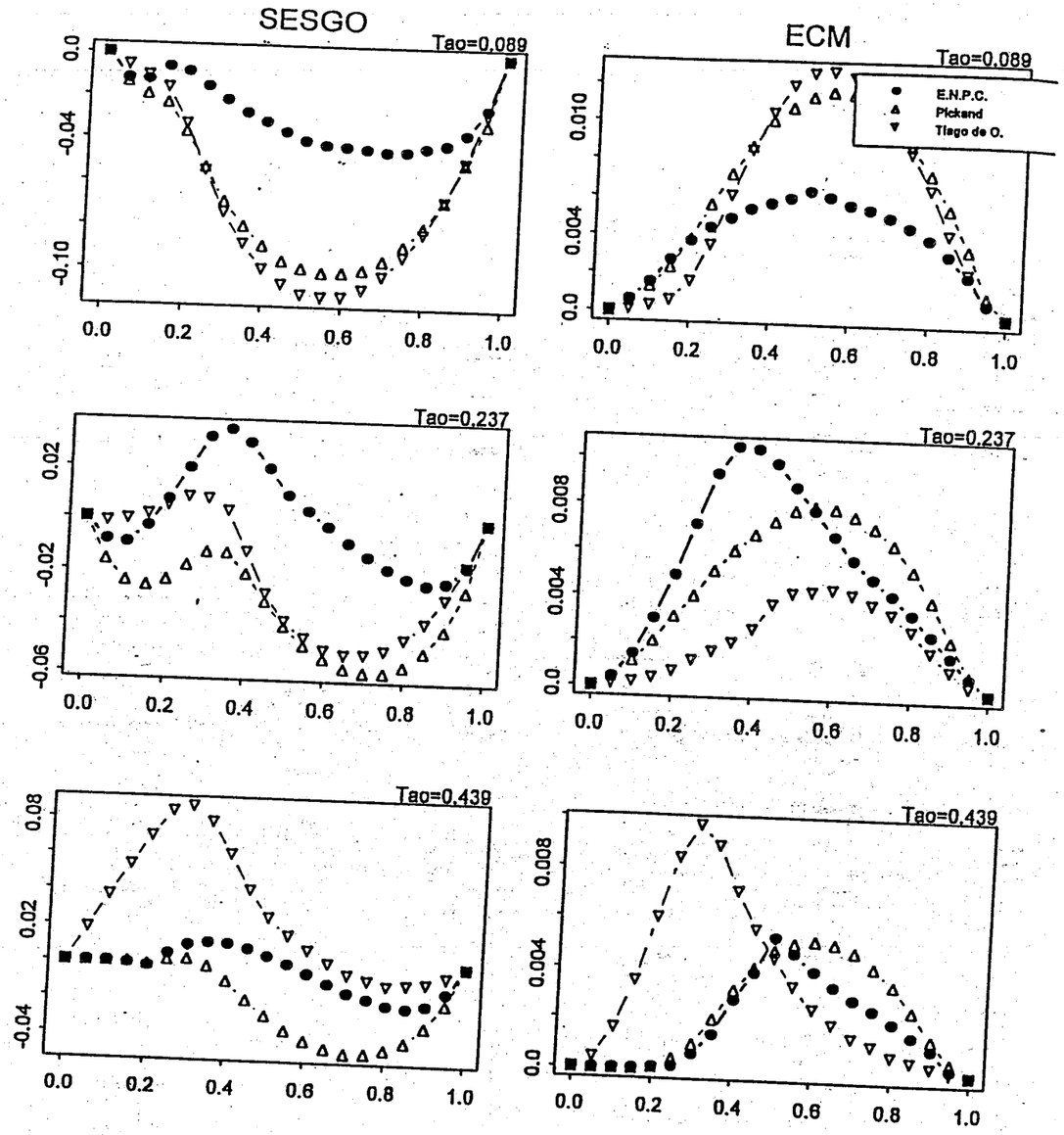


Figura 4.5. Funciones de sesgo y error cuadrático medio para los casos de dependencia baja, media y alta asimétrica.

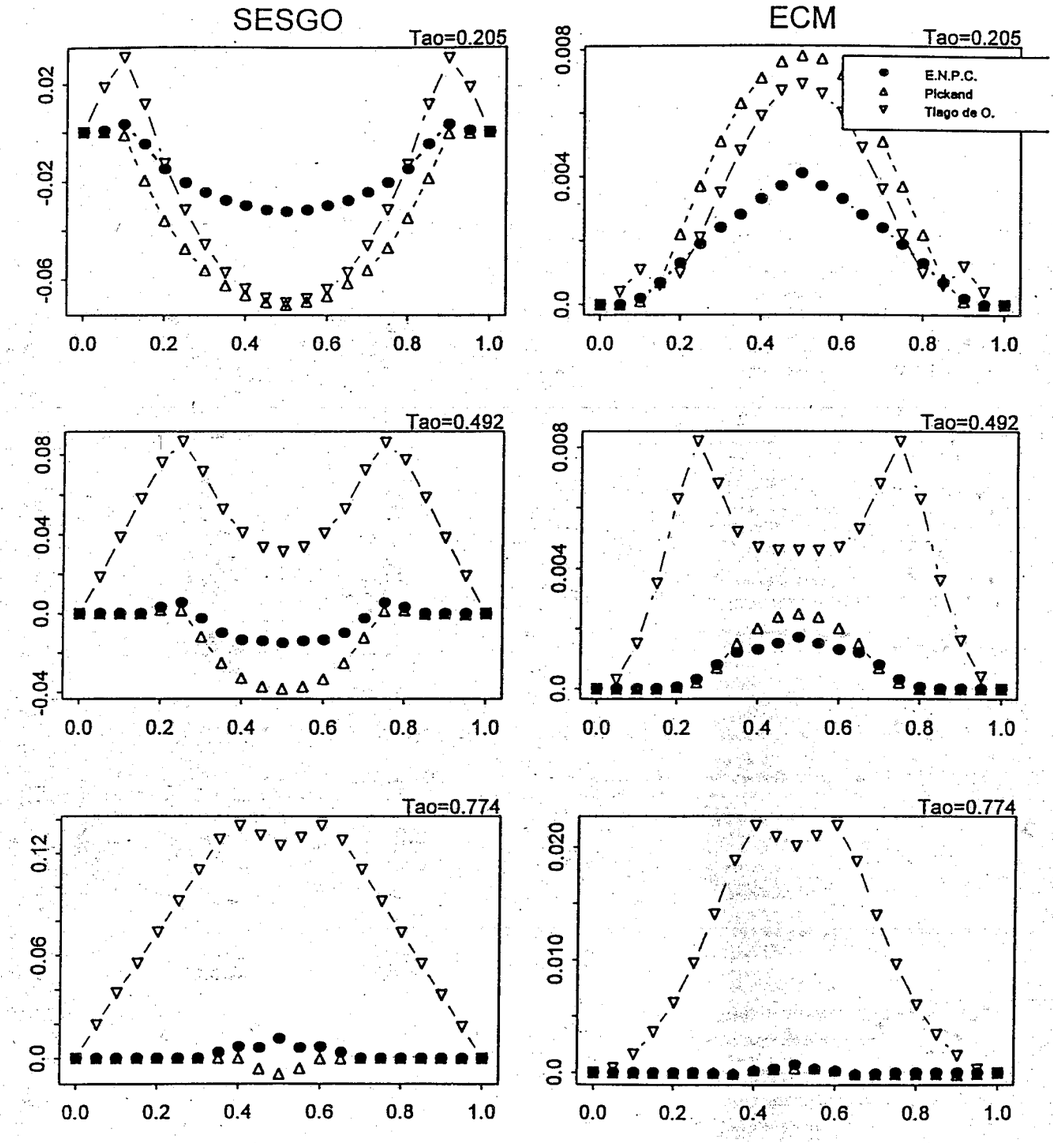


Figura 4.6. Funciones de sesgo y error cuadrático medio para los casos de dependencia baja, media y alta con estructura de dependencia plana en la parte central, independencia local.

4.6 Apéndice: Algoritmo

En esta sección presentamos el algoritmo para obtener un estimador no paramétrico de la función de dependencia:

Sean las estadísticas de orden $X_{(k)}, k = 1, 2, ..., n$, y sean $X_{(0)}$ y $X_{(n+1)}$, tales que,

$$0 = X_{(0)} < X_{(1)} < ... < X_{(n)} < X_{(n+1)} = 1.$$

Paso 1: Definición de $\hat{A}_1(x)$ y $\hat{A}_2(x)$
para $X_{(k)} \leq x < X_{(k+1)}$,

$$\hat{A}_1(x) = \exp\left[\sum_{j=0}^{k-1}(X_{(j+1)} - X_{(j)}) \max\left\{-1, \max_{1\leq i\leq j}\left\{\frac{\frac{i}{n}-X_{(i)}}{X_{(i)}(1-X_{(i)})}\right\}\right\}\right.$$
$$\left.+(x - X_{(k)}) \max\left\{-1, \max_{1\leq i\leq k}\left\{\frac{\frac{i}{n}-X_{(i)}}{X_{(i)}(1-X_{(i)})}\right\}\right\}\right]$$

$$\hat{A}_2(x) = \exp\left[-\sum_{j=k+1}^{n}(X_{(j+1)} - X_{(j)}) \min\left\{1, \min_{j\leq i\leq n-1}\left\{\frac{\frac{i}{n}-X_{(i+1)}}{X_{(i+1)}(1-X_{(i+1)})}\right\}\right\}\right.$$
$$\left.-(X_{(k+1)} - x) \min\left\{1, \min_{(k\leq i\leq n-1)}\left\{\frac{\frac{i}{n}-X_{(i+1)}}{X_{(i+1)}(1-X_{(i+1)})}\right\}\right\}\right]$$

Paso 2: Combinación de $\hat{A}_1$ $\hat{A}_2$.
Definir

$$\tilde{A}(x) = \begin{cases} \max(1 - x, \hat{A}_1(x)) & , 0 \leq x \leq 1/2 \\ \max(x, \hat{A}_2(x)) & , 1/2 < x \leq 1 \end{cases}.$$

Paso 3: Máximo minorante convexo de $\tilde{A}(x)$.
Sean

$$x_i = \arg\min\left\{\tilde{A}(x), 0 \leq x \leq 1/2\right\},$$
$$x_d = \arg\min\left\{\tilde{A}(x), 1/2 < x \leq 1\right\}.$$

Si $\tilde{A}(x_i) < \tilde{A}(x)$ entonces,

3.1: tomar $x_0 \in (1/2, 1)$ tal que $\tilde{A}'(x) = [\tilde{A}(x_0) - \tilde{A}(x_i)]/(x_0 - x_i)$ y definir

$$\hat{A}(x) = \begin{cases} \tilde{A}(x) & , 0 \leq x \leq x_i \\ \tilde{A}(x_i) + \tilde{A}'(x_0)(x - x_i) & , x_i < x \leq x_0 \\ \tilde{A}(x) & , x_0 < x \leq 1 \end{cases}.$$

En otro caso ($\tilde{A}(x_i) > \tilde{A}(x_d)$),

3.2: tomar $x_0 \in (0, 1/2)$ tal que $\tilde{A}'(x) = [\tilde{A}(x_d) - \tilde{A}(x_0)]/(x_d - x_0)$ y definir

$$\hat{A}(x) = \begin{cases} \tilde{A}(x) & , 0 \leq x \leq x_0 \\ \tilde{A}(x_0) + \tilde{A}'(x_0)(x - x_0) & , x_0 < x \leq x_d \\ \tilde{A}(x) & , x_d < x \leq 1 \end{cases}.$$

Capítulo 5

Ejemplos

5.1 Introducción

Los datos utilizados en estos ejemplos provienen de transformaciones realizadas a datos originales, que aquí se describen. Primeramente se identifican las distribuciones de las variables marginales con técnicas gráficas y posteriormente se transforman para tener marginales exponenciales unitarias.

En el primer ejemplo analizamos la estructura de dependencia de los valores máximos anuales de las descargas registradas en dos localidades diferentes de un río. En este ejemplo tenemos un caso de fuerte dependencia, que es importante considerar al calcular probabilidades de tener exceso simultaneo en las descargas máximas anuales.

En el segundo ejemplo tenemos un caso de independencia entre valores extremos. Los datos son las edades máximas de muerte durante cada año, para hombres y mujeres en Suecia. Este conjunto de datos se ha utilizado anteriormente (Gumbel y Goldstein, 1964) para mostrar un caso de independencia entre valores extremos. En este ejemplo observamos que tanto la inferencia basada en verosimilitud, como en el enfoque Bayesiano, concluyen a favor de la hipótesis de independencia entre las edades máximas de muerte de hombres y mujeres.

El tercer ejemplo considera el patrón de dependencia entre los niveles altos de ozono

y velocidades mínimas de viento, registrados en una estación de la red de monitoreo ambiental de la ciudad de México. Este es un caso de dependencia reducida entre valores extremos. Los niveles altos de ozono y las velocidades mínimas de viento se toman por episodios, en el sentido de Smith (1989).

En los tres ejemplos inicialmente se estiman y comparan en el esquema bivariado los modelos logísticos y de mezcla de betas, citados en el Capítulo 2. Posteriormente se hace el análisis en el esquema univariado de los cocientes, presentado en el Capítulo 3.

5.2 Ejemplo 1. Descargas máximas del río Ocmulgee

Los datos en este ejemplo provienen de las descargas registradas del río Ocmulgee en dos puntos diferentes de su recorrido por el estado de Georgia, Estados Unidos. La muestra bivariada está formada por las descargas máximas anuales del río, medidas, río arriba en Hawkinsville y río abajo en Macon, durante los años 1910-1948, de acuerdo a el estudio geológico reportado por Carter (1951).

Estos datos, que son un ejemplo claro de dependencia positiva fuerte, se presentan en la Tabla 5.1. Para estos máximos anuales, se identificó en ambas componentes una distribución Gumbel, con parámetros de localización y escala, $\mu = 18.00103, \sigma = 26.76638$ para Macon y $\mu = 15.05039, \sigma = 23.85584$ para Hawkinsville. Después de identificar la distribución de los datos originales, se transformaron a variables con distribución exponencial unitaria, con la transformación $x = \exp(-(u-\mu)/\sigma)$, aplicada a las observaciones originales u, que son las descargas máximas anuales.

Los datos transformados, con marginales exponenciales unitarias, se presentan en el diagrama de dispersión de la Figura 5.1, en donde observamos, como clara evidencia de una asociación positiva importante entre los valores extremos, una franja delgada donde se encuentran las parejas de observaciones extremas.

Estudiamos la estructura de dependencia en los datos transformados, con los modelos definidos en los capítulos anteriores. Inicialmente ajustamos el modelo de mezcla de

betas, propuesto en el Capítulo 2, además de los modelos propuestos por otra autores que se citan en el Capítulo 1, con el objetivo de comparar los modelos a través de las funciones de dependencia estimadas con los diferentes modelos.

Tabla 5.1. Descargas máximas anuales del río Ocmulgee, (1000 pies cúbicos/seg.).

año	Ma.	Hawk.	año	Ma.	Hawk.	año	Ma.	Hawk.	año	Ma.	Hawk.
1910	28.8	18.8	1920	66.2	45.2	1930	64.4	50.0	1940	14.2	13.3
1911	8.5	5.9	1921	37.0	30.0	1931	10.7	12.2	1941	7.3	6.9
1912	44.8	44.4	1922	48.6	44.0	1932	19.6	16.2	1942	73.4	57.0
1913	51.0	52.0	1923	28.3	30.3	1933	19.0	19.9	1943	44.8	41.6
1914	4.8	5.9	1924	21.0	15.2	1934	16.9	17.4	1944	50.2	46.8
1915	19.1	20.1	1925	72.5	79.0	1935	22.7	13.5	1945	40.4	26.2
1916	47.8	40.4	1926	28.3	19.3	1936	65.3	61.0	1946	57.6	35.4
1917	25.4	27.0	1927	7.9	7.6	1937	33.3	25.8	1947	32.6	34.8
1918	14.3	14.3	1928	47.1	42.4	1938	31.0	33.0	1948	24.0	28.2
1919	31.0	19.5	1929	73.4	70.1	1939	33.9	37.9	1949	84.0	68.0

Ajustamos inicialmente el modelo de mezcla de 2 betas, estimando por máxima verosimilitud los parámetros ν_1, ν_2, ν, con los resultados que aparecen en la Tabla 5.2. El parámetro ν_1 tiene un valor muy cercano a $\nu/2$ y a su vez, ν tiene un valor relativamente grande, lo cual es un indicio de que hay una asociación fuerte y la dependencia es altamente simétrica. Posteriormente ajustamos el modelo de mezcla de 2 betas simétrico, que tenemos al considerar la restricción $\nu_2 = \nu - \nu_1$, con las estimaciones de ν_1 y ν que se dan en la Tabla 5.2, donde resulta nuevamente que ν_1 tiene un valor muy cercano a $\nu/2$, como indicación de la fuerte dependencia entre los extremos.

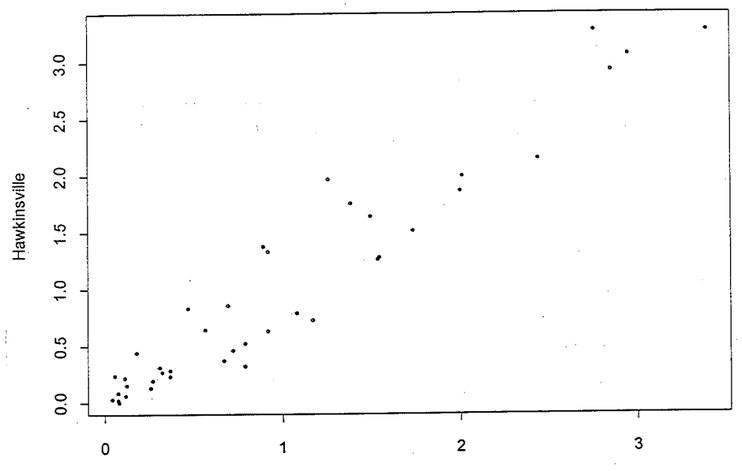


Figura 5.1. Descargas maximas anuales transformadas a exponenciales unitarias, registradas en dos localidades del rio Ocmulgee.

Aquí es importante notar que la desviación estándar de las estimaciones se reduce notablemente al pasar del modelo original al modelo simétrico. El motivo de esta reducción es que el modelo original está sobre-parametrizado, como lo indica la prueba de razón de verosimilitudes para la hipótesis de simetría de la función de dependencia $A(w)$,

$$H_0 \ : \ \nu_2 = \nu - \nu_1,$$

contra la hipótesis alternativa

$$H_0 \ : \ \nu_2 \neq \nu - \nu_1.$$

La estadística de prueba

$$\lambda = -2\left[l(\hat{\nu}_1, \hat{\nu} - \hat{\nu}_1, \hat{\nu}) - l(\hat{\nu}_1, \hat{\nu} - \hat{\nu}_1, \hat{\nu})\right] = -2(-41.65 + 40.561) = 2.008,$$

es notablemente pequeña, por lo tanto hay una diferencia reducida entre el modelo general y el modelo restringido.

Al ajustar el modelo logístico asimétrico encontramos estimaciones de los parámetros que indican una dependencia simétrica fuerte, ya que los parámetros de asimetría θ y ϕ prácticamente son iguales a la unidad. El modelo logístico simétrico se ajusta con resultados muy similares a el caso asimétrico, como era de esperarse debido a que los parámetros de asimetría θ y ϕ estimados, revelan un alto grado de simetría entre los valores extremos.

En la Tabla 5.2, observamos que para el modelo logístico invertido, obtenemos resultados similares, tanto en la versión general como en la simétrica, a los obtenidos para el modelo logístico asimétrico.

Tabla 5.2. Estimaciones de parámetros de varios modelos, por máxima verosimilitud. Los datos son descargas máximas del río Ocmulgee en dos localidades diferentes.

Modelo	$\hat{\theta}_1(d.e.)$	$\hat{\theta}_2(d.e.)$	$\hat{\theta}_3(d.e.)$	Log-ver.
M. 2B	$\nu_1 = 9.869(3.590)$	$\nu_2 = 15.710(5.776)$	$\nu = 20.714(8.005)$	-40.561
M. 2B sim.	$\nu_1 = 6.367(.798)$	$\nu_2 = \nu - \nu_1$	$\nu = 13.425(.196)$	-41.565
L. Inv.	$\phi_1 = 0.999$	$\phi_2 = 1.0$	$\tau = 3.1528$	-43.940
L.Inv.Sim.			$\tau = 3.1531(.503)$	-43.940
L. Asim.	$\theta = 0.998$	$\phi = 1.0$	$\alpha = 3.856$	-43.966
L. Sim.			$\alpha = 3.864(0.512)$	-43.958

En la Figura 5.2 se tienen las funciones de dependencia para los modelos de mezcla de betas simétrico y asimétrico, así como para los modelos logístico general y logístico

invertido general, notando que hay una gran coincidencia gráficas de las diferentes funciones.

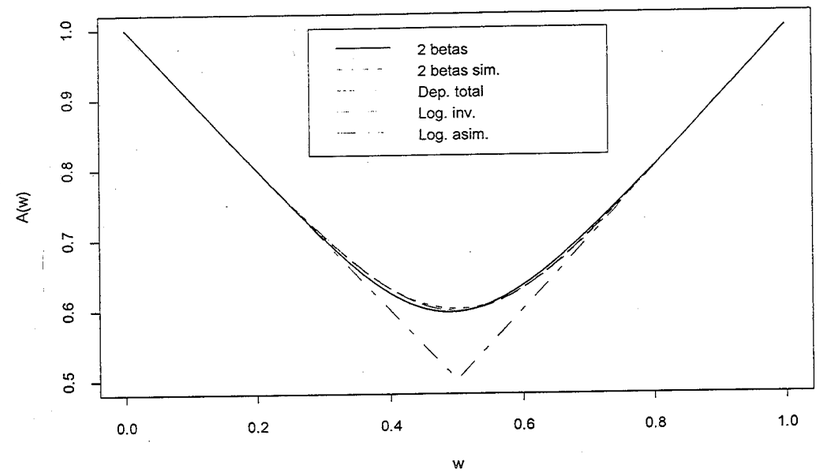


Figura 5.2. Funciones de dependencia para descargas máximas anuales del rio Ocmulgee en dos localidades diferentes. Modelos mezcla de 2 betas asimétrico y simétrico, logístico asimétrico y logístico invertido.

5.2.1 Inferencia en el esquema de cocientes

Enseguida analizamos los datos de descargas máximas anuales del río Ocmulgee, utilizando el esquema univariado de los cocientes, $W = Y/(X + Y)$. En la Figura 5.3 tenemos la gráfica de la función de distribución empírica de los cocientes $w_1, w_2, ..., w_n$, donde observamos una aglomeración de los cocientes alrededor de $1/2$, lo cual es evidencia de un alto grado de asociación entre los valores extremos X y Y.

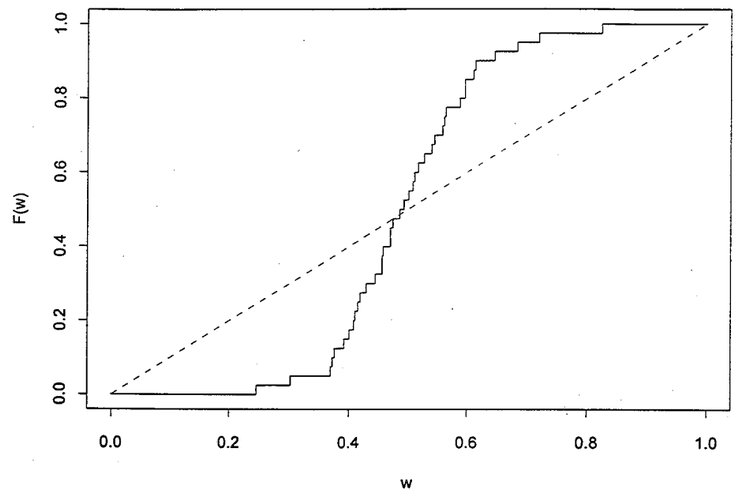


Figura 5.3. Función de distribución empírica para los cocientes de valores extremos provenientes de descargas máximas del rio Ocmulgee en dos localidades diferentes.

Las estimaciones de máxima verosimilitud y de momentos se presentan en la Tabla 5.3, en donde observamos una gran similitud en las estimaciones. En la Figura 5.4 tenemos la función de verosimilitud relativa observada $R(a) = L(a)/L(\hat{a})$, correspondiente al modelo (3.3) propuesto en el Capítulo 3, así como también la verosimilitud relativa normal con media y la varianza correspondientes a las estimaciones máximo-verosímiles, obtenidas aquí.

Tabla 5.3. Estimaciones del parámetro a del modelo propuesto

	$\hat{a}$	$\widehat{Var}(\hat{a})$
Máxima Verosimilitud	4.7934	0.5744
Momentos	4.7850	

La comparación de la verosimilitud observada con la verosimilitud relativa normal, permite juzgar la bondad de la aproximación normal a la distribución del estimador máximo-verosímil, con el fin de realizar inferencias sobre el parámetro a. Esta semejanza entre las verosimilitudes relativas observada y normal indica que las inferencias sobre el parámetro a, basadas en la aproximación normal son adecuadas.

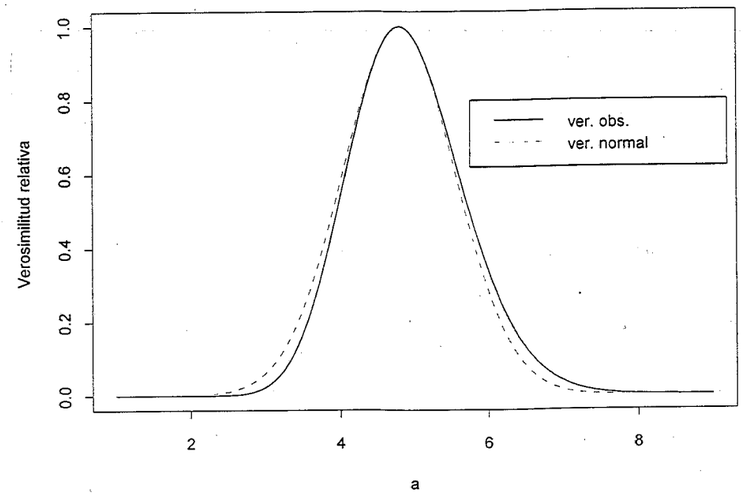


Figura 5.4. Funciones de verosimilitud relativa para el modelo propuesto y normal.

Enfoque Bayesiano

En el enfoque Bayesiano, presentado en el Capítulo 3, una distribución a priori para el parámetro a del modelo, que surge de la consideración de asociación positiva entre los valores extremos, está dada por,

$$f_0(a) = a^{-2} I_{[1,\infty)}(a).$$

La distribución a posteriori de a, (3.26) es la distribución gama truncada por la izquierda en uno, con parámetros de escala 8.45 y de forma 39, y con función de densidad,

$$f(a|w) = [1/S(1, 39, 8.345)]a^{38} \exp\{-8.345a\} I_{[1,\infty)}(a).$$

El estimador puntual Bayesiano de a dado en (3.27) es,

$$\hat{a}_B = 4.6736.$$

Este valor es menor que la estimación máximo-verosímil (4.7934), obtenida anteriormente. Esta relación entre las estimaciones máximo-verosímil y Bayesiana, se encontró anteriormente al comparar los estimadores en términos del sesgo, con los resultados señalados en la Tabla 3.2, para el caso de dependencia fuerte, como es el caso de este ejemplo.

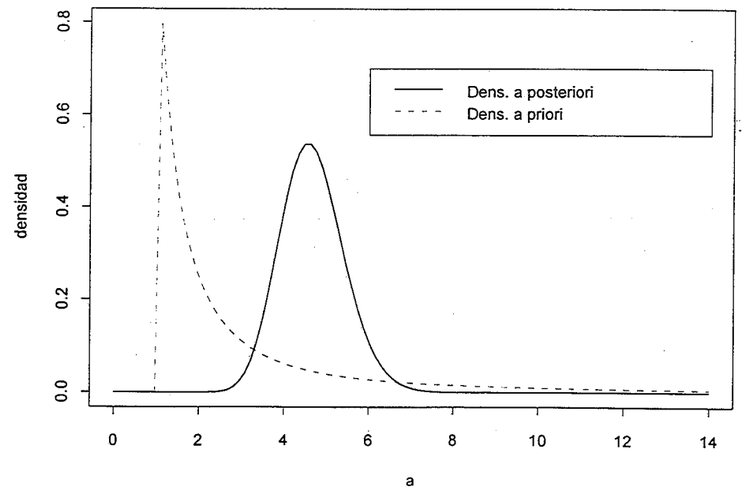


Figura 5.5. Funciones de densidad de las distribuciones a priori y a posteriori de a.

En la Figura 5.5 tenemos las funciones de densidad de las distribuciones a priori y a posteriori para a. La separación que observamos en las funciones de densidad a priori y a posteriori, se debe a que los datos dominan la inferencia.

En la Tabla 5.4 tenemos los intervalos de confianza de longitud mínima para diferentes niveles de probabilidad.

Tabla 5.4. Intervalos de confianza Bayesianos para tres niveles de probabilidad.

Confianza	90	95	99
Int. de confianza	(3.442, 5.882)	(3.253, 6.163)	(2.903, 6.735)

Obtenemos en este problema, predicciones relacionadas con la ocurrencia conjunta de excesos, en la escala original. Por ejemplo, la predicción de que el nivel de las descargas máximas U y V en las dos localidades no sea mayor que u y v, para las dos localidades esta dado por,

$$P(U < u, V < v|w) = P(X > x, Y > y|w) = S(x, y|w),$$

donde X y Y son las variables exponenciales unitarias que resultan al transformar las variables originales U, V tipo Gumbel, como se describe al inicio de esta sección. En la Tabla 5.5 mostramos la predicción anterior para diferentes valores de u y v.

Tabla 5.5. Predicciones de que las descargas máximas en dos localidades, no rebasen diferentes niveles u y v.

u	v	x	y	$S(x, y\|w)$
50	50	.2750	.1760	.7448
80	80	.0519	.0239	.9479
100	100	.0171	.0063	.9827
150	150	.0010	2.4×10^{-4}	.9989
200	200	6.6×10^{-5}	8.2×10^{-6}	.9999

5.2.2 Probabilidades de interés

Si las variables aleatorias U, V son las descargas máximas anuales registradas en Macon y Hawkinsville respectivamente, y u, v son los valores umbrales considerados para ambos lugares, entonces son de interés las probabilidades p_1 y p_2, de que los máximos anuales en Macon y Hawkinsville sobrepasen los valores umbrales u y v respectivamente. Además de estas probabilidades univariadas, tenemos las siguientes probabilidades que consideran la variabilidad conjunta en ambas localidades: la probabilidad p_{1o2}, de tener exceso en al menos una localidad y la probabilidad p_{1y2}, de tener exceso en ambos lugares.

De las probabilidades anteriores podemos deducir las probabilidades condicionales p_{12}, de tener exceso en Hawkinsville dado que hay exceso en Macon y p_{-12}, de tener exceso en Hawkinsville, dado que no hay exceso en Macon.

Para los máximos anuales en Hawkinsville y Macon se identificaron distribuciones de valor extremos para máximos con los parámetros de localización y escala dados al inicio de esta sección, por lo tanto, las probabilidades de tener exceso están dadas por,

$$\begin{aligned} p_1 &= P(U > u) = 1 - \exp\left[-\exp\left(\frac{u - 18.00103}{26.76638}\right)\right], \\ p_2 &= P(V > v) = 1 - \exp\left[-\exp\left(\frac{v - 15.05039}{23.85584}\right)\right]. \end{aligned}$$

Las probabilidades que consideran la variabilidad en ambas localidades dependen de la función de dependencia $A(w)$, de acuerdo a las siguientes expresiones:

$$\begin{aligned} p_{1o2} &= 1 - P(U < u, V < v) \\ &= 1 - [(1-p_1)(1-p_2)]^{A\left(\frac{\log(1-p_2)}{\log(1-p_1)+\log(1-p_2)}\right)}, \\ p_{1y2} &= P(U > u, V > v) = p_1 + p_2 - p_{1o2}, \\ p_{12} &= P(V > v | U > u) = \frac{p_{1y2}}{p_1}, \\ p_{-12} &= \frac{p_2 - p_{1y2}}{1 - p_1}. \end{aligned}$$

En la Tabla 5.6 tenemos estas probabilidades para valores umbrales $u = v = 120$, considerando las estimaciones de la función de dependencia estimada por máxima verosimilitud, momentos y método Bayesiano y además la estimación no-paramétrica aquí propuesta.

En este ejemplo el estimador no paramétrico propuesto detecta una menor dependencia, ya que tiene un valor de $A(w)$ mayor que los otros, esto se refleja en las probabilidades p_{1y2} y p_{21}, ligadas a la ocurrencia simultánea de exceso. En la Gráfica 5.6, presentamos las función de dependencia para el modelo aquí propuesto, con el parámetro de dependencia a estimado por máxima verosimilitud, momentos y en forma Bayesiana. También tenemos la estimación no paramétrica, con el estimador no paramétrico aquí propuesto. En esta gráfica notamos una dependencia menor con el estimador no paramétrico.

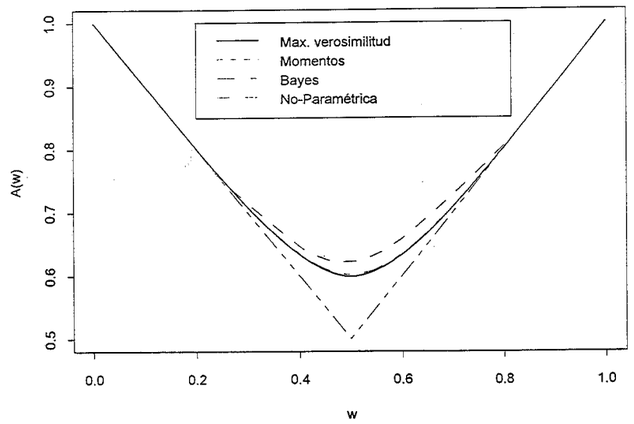


Figura 5.6. Estimaciones de la función de dependencia por los métodos de máxima verosimilitud, momentos y Bayesiano con el modelo paramétrico propuesto, así como también en forma no paramétrica.

Tabla 5.6. Probabilidades marginales, conjuntas y condicionales de tener exceso en las descargas máximas anuales en Hawkinsville y Macon, para los valores umbrales $u = v = 120$.

	Máx.Ver.	Mom.	Bayes	No Param.	Independencia
$A(w)$	0.6622	0.6623	0.6635	0.6749	1.0000
p_1	0.02188	0.02188	0.02188	0.02188	0.02188
p_2	0.01221	0.01221	0.01221	0.01221	0.01221
p_*	0.022533	0.022537	0.02257	0.022947	0.033811
p^*	0.011565	0.011562	0.011522	0.011132	0.00256
p_{12}	0.528381	0.528227	0.526383	0.509040	0.012210
p_{-12}	0.000658	0.000662	0.000703	0.001102	0.012210

En caso de suponer erróneamente independencia entre las descargas máximas anuales tenemos una gran diferencia en las probabilidades ligadas a la ocurrencia de exceso en ambos lugares. Es interesante notar aquí que las probabilidades de la Tabla 5.6, dependen del valor de la función de dependencia evaluada en

$$w = \frac{\log(1-p_2)}{\log(1-p_1)+\log(1-p_2)} = 0.3570,$$

observando con esto que no solo es de interés el conocimiento de la función de dependencia en $w = 1/2$, que corresponden al caso de tener probabilidades iguales en las marginales, esto es, al caso en que $p_1 = p_2$.

5.3 Ejemplo 2. Edades de muerte máximas en Suecia.

En este ejemplo consideramos las edades de muerte máximas anuales registradas en Suecia de 1905 a 1958. Estos datos fueron publicados en Gumbel y Goldstein (1964)

y fueron tomados de la oficina central de estadísticas en Estocolmo, Suecia. Los datos se presentan en la Tabla 5.7. Para estos máximos anuales, su distribución se identificó inicialmente como Fréchet, tanto para hombres como para mujeres. Después de identificar la distribución de los datos originales, se transformaron a exponencial unitaria.

Tabla 5.7. Edades máximas anuales de muerte en Suecia, para hombres y mujeres.

año	hombres	mujeres	año	hombres	mujeres	año	hombres	mujeres
1905	100.88	102.54	1923	101.63	102.32	1941	101.66	106.15
1906	101.17	106.13	1924	103.47	103.56	1942	106.48	104.71
1907	104.65	103.46	1925	105.48	103.86	1943	101.26	103.83
1908	105.12	102.12	1926	104.01	105.87	1944	105.12	105.19
1909	102.57	101.69	1927	105.83	103.31	1945	104.88	105.03
1910	101.70	102.92	1928	105.00	104.37	1946	102.41	105.88
1911	100.49	102.78	1929	102.78	102.72	1947	104.22	107.49
1912	100.90	106.15	1930	102.61	105.01	1948	102.88	105.83
1913	103.06	105.02	1931	105.55	104.40	1949	103.57	103.41
1914	102.63	103.56	1932	102.55	104.87	1950	105.12	105.64
1915	102.69	106.52	1933	103.17	105.98	1951	103.80	103.53
1916	100.82	101.50	1934	103.98	103.38	1952	102.94	107.90
1917	102.52	104.01	1935	106.09	105.32	1953	103.00	104.42
1918	100.08	105.01	1936	103.43	103.77	1954	106.50	104.85
1919	101.67	104.52	1937	105.72	105.86	1955	103.36	103.97
1920	101.41	103.94	1938	103.24	104.27	1956	103.15	107.89
1921	101.76	103.14	1939	103.25	105.45	1957	102.54	104.46
1922	102.57	104.33	1940	103.40	105.71	1958	104.92	104.12

Estos datos son un ejemplo de independencia entre valores extremos, como se observa en el diagrama de dispersión de los datos transformados de la Figura 5.7, en donde observamos las parejas de valores extremos transformados, distribuidas aleatoriamente, sin un patrón de asociación positiva o negativa.

El patrón de independencia entre estos valores extremos, también se aprecia en las estimaciones de los parámetros del modelo de mezcla de 2 betas, presentadas en la Tabla 5.8, donde tenemos que ν tiene un valor grande y además ν_1, ν_2 están ubicados en los extremos del intervalo $(0, \nu)$. Este comportamiento de los parámetros de este modelo corresponde, como mencionamos en el Capítulo 2, a una situación de dependencia muy baja.

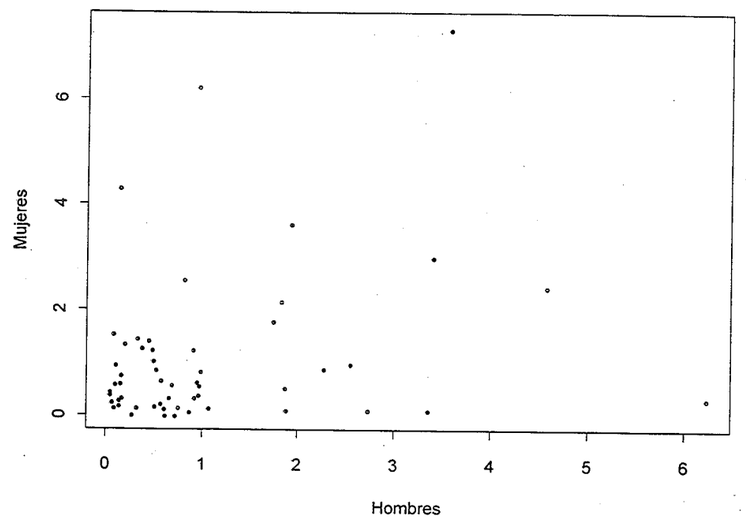


Figura 5.7. Edades de muerte máximas anuales, transformadas a exponenciales, para hombres y mujeres en Suecia.

Tabla 5.8. Estimaciones de parámetros de modelos de dependencia por máxima verosimilitud. Los datos son edades máximas de muerte seleccionadas anualmente en Suecia durante 54 años.

Modelo	$\hat{\theta}_1(d.e.)$	$\hat{\theta}_2(d.e.)$	$\hat{\theta}_3(d.e.)$	*Log.ver.*
M. 2B	$\nu_1 = 7.9 \times 10^{-12}$	$\nu_2 = 22.409$	$\nu = 23.532$	-116.073
M. 2B sim.	$\nu_1 = 1,115 \times 10^{-10}$	$\nu_2 = \nu - \nu_1$	$\nu = 16.7331$	-116.589
L. Inv.	$\phi_1 = .576$	$\phi_2 = .0615$	$\tau = 38.408$	-113.735
L.Inv.Sim.			$\tau = .351(.142)$	-116.122
L. Asim.	$\theta = .577$	$\phi = .0162$	$\alpha = 39.128$	-112.775
L. Sim.			$\alpha = 1.1038(.112)$	-116.283

Los modelos logístico asimétrico y logístico invertido muestran ajustes similares, esto es, reflejan un grado de dependencia muy bajo entre los valores extremos. Al tomar las versiones simétricas de estos modelos notamos que sus parámetros α y τ estimados tienen un cambio abrupto. Esta inestabilidad en el parámetro α del modelo logístico asimétrico y τ del logístico invertido, hacen difícil su interpretación, ya que esta depende fuertemente de los valores estimados de los parámetros de asimetría.

En la Figura 5.8 tenemos las estimaciones de los parámetros de las funciones de dependencia utilizadas para modelar el patrón de dependencia en este ejemplo.

5.3.1 Inferencia en el esquema de cocientes

Enseguida analizamos los datos de edades máximas de muerte, utilizando el esquema univariado de los cocientes, $W = Y/(X + Y)$. En la Figura 5.9 tenemos la gráfica de la función de distribución empírica de los cocientes $w_1, w_2, ..., w_n$, donde observamos una distribución uniforme de los cocientes $w_1, w_2, ..., w_n$, en el intervalo $(0, 1)$, lo cual es evidencia de un patrón de independencia entre los valores extremos X y Y.

Las estimaciones de máxima verosimilitud y de momentos se presentan en la Tabla 5.9, en donde observamos una gran similitud en las estimaciones. En la Figura 5.10 tenemos la función de verosimilitud relativa observada $R(a) = L(a)/L(\hat{a})$, correspondiente al modelo (3.3) propuesto en el Capítulo 3, así como también la verosimilitud relativa normal con media y la varianza correspondientes a las estimaciones máximo-verosímiles, obtenidas aquí.

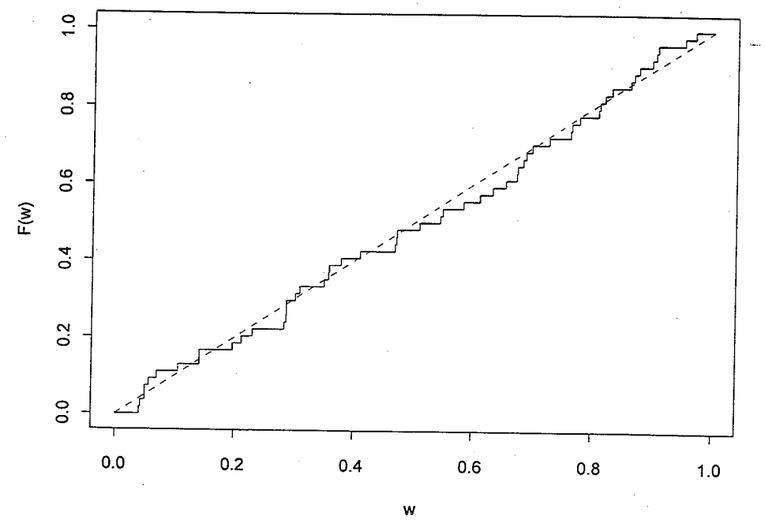


Figura 5.9. Función de distribución empírica para los cocientes de valores extremos provenientes de edades de muerte máximas en Suecia para hombres y mujeres.

La comparación de la verosimilitud relativa observada con la verosimilitud relativa normal, permite juzgar la bondad de la aproximación normal a la distribución del estimador máximo-verosímil, con el fin de realizar inferencias sobre el parámetro a.

Tabla 5.9. Estimaciones del parámetro a del modelo propuesto

	$\hat{a}$	$\widehat{Var}(\hat{a})$
Máxima Verosimilitud	1.0812	.0216
Momentos	.9940	

Tabla 5.10. Intervalos de confianza, para el parámetro a del modelo propuesto, basados en la aproximación normal de la verosimilitud.

Confianza	80	90	95
Int. de confianza	(.8926, 1.2697)	(.8392, 1.3232)	(.7928, 1.3695)

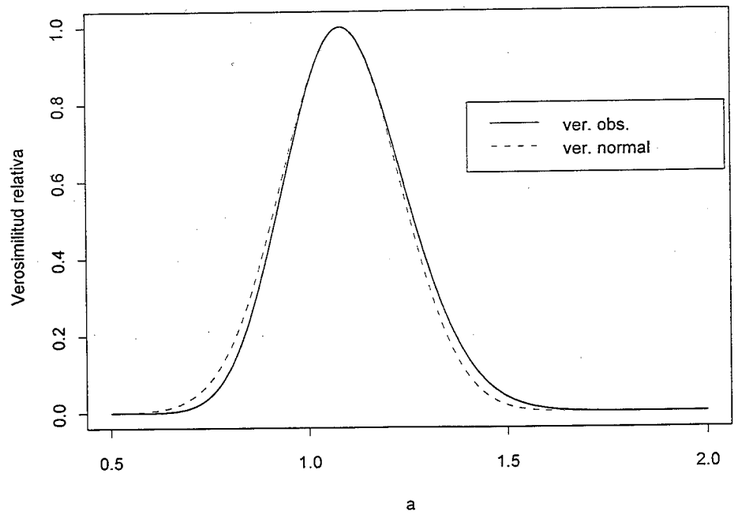


Figura 5.10. Funciones de verosimilitud relativa, observada y normal aproximada, para las edades máximas de muerte anuales.

En este ejemplo encontramos en la verosimilitud observada un sesgo a la derecha similar al señalado en el ejemplo anterior donde analizamos descargas máximas anuales. Este comportamiento de sesgo a la derecha, se observó en el Capítulo 3 en la comparación de estimadores, que se muestra en las Gráficas 3.4-3.7.

La evidencia de los datos en favor de la hipótesis de independencia entre los valores extremos, la apreciamos al obtener intervalos de confianza para el parámetro a del modelo propuesto. Estos intervalos de confianza para diferentes niveles de confianza, que se muestran en la Tabla 5.10, contienen al valor $a = 1$, que corresponde a la situación de independencia entre los valores extremos.

5.3.2 Esquema Bayesiano

Analizamos el conjunto de datos de este ejemplo, con el enfoque Bayesiano, considerando como en el ejemplo anterior, la distribución a priori dada en (3.25). La distribución a posteriori que resulta de esta distribución previa y de los datos es la distribución gama truncada por la izquierda en uno, con parámetros de escala 49.944 y de forma 53. Estas distribuciones a priori y a posteriori se muestran en la Figura 5.11.

El estimador Bayesiano del parámetro a, dado por la esperanza (3.27), es $\hat{a}_B = 1.1428$, que es mayor a la estimación máximo-verosímil calculado anteriormente (1.0812). Cuando se tiene un patrón de dependencia muy baja, como en este ejemplo, la estimación Bayesiana tiende a ser mayor que la estimación máximo-verosímil sin restricción, debido a que en el esquema Bayesiano, la distribución a priori utilizada, asigna probabilidad cero a la región de valores de a menores que uno, mientras que para los valores mayores que uno les asigna un peso decreciente pero importante, como lo podemos notar en la Figura 5.11.

Esta función de densidad a priori genera una distribución a posteriori que asigna pesos importantes a valores de a mayores que uno, por lo cual al tomar la esperanza como un estimador puntual, este tiene un valor mayor que el estimador máximo-verosímil.

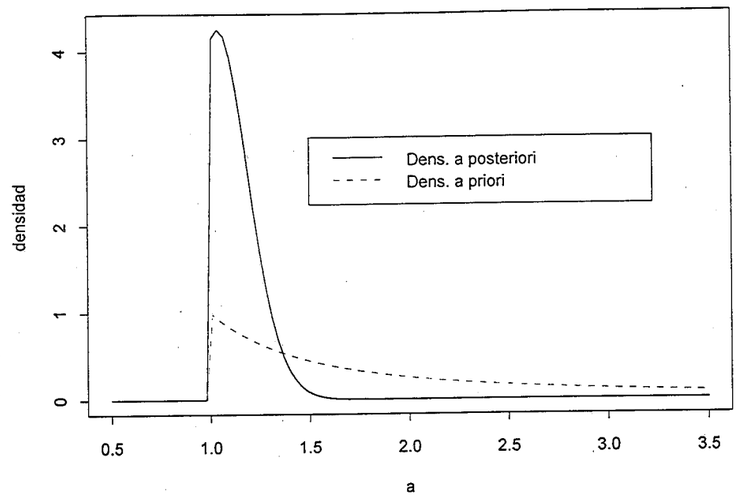


Figura 5.11. Funciones de densidad a priori y a posteriori de a, para el ejemplo de las edades máximas de muerte.

Tabla 5.11. Intervalos de confianza Bayesianos de longitud mínima, para el parámetro a del modelo propuesto.

Confianza	80	90	95
Int. de confianza	(1.000, 1.2269)	(1.000, 1.2902)	(1.000, 1.3459)

Desde el enfoque Bayesiano, los datos también apoyan la hipótesis de independencia de los valores extremos de edades de muerte, ya que cualquier intervalo de confianza Bayesiano de longitud mínima, para probabilidades mayores que $p = 0.4555$, contiene al valor $a = 1$. Esto se debe a que el máximo de la función de densidad a posteriori, está muy cerca del valor de truncamiento a la izquierda. En la Tabla 5.11 se muestran tres intervalos de confianza Bayesianos de longitud mínima.

5.4 Ejemplo 3 Concentración de Ozono y velocidad de viento

En este ejemplo los datos son concentración máxima de ozono y velocidad mínima de viento por episodio, transformados a exponenciales unitarias. El objetivo es estudiar la asociación de los extremos de estas variables. Los datos originales son mediciones diarias, de concentración máxima de ozono y velocidad mínima de viento. Las concentraciones de ozono se registraron en la estación Pedregal de la red de monitoreo ambiental de la ciudad de México, del 1 de enero de 1987 al 31 de mayo de 1992. La velocidades mínimas de viento provienen de lecturas tomadas en la estación meteorológica del aeropuerto de la ciudad de México.

Se formaron episodios en base a las mediciones de ozono, con un umbral de .25 ppm (partes por millon), y una separación de 48 horas. De cada episodio se tomó la máxima concentración de ozono y la mínima velocidad de viento. Para el ozono se identificó una distribución Gumbel, mientras que para el viento fué la distribución Weibull para mínimos. Posteriormente ambas variables se transformaron a exponenciales unitarias.

Para evitar problemas de estacionalidad y tendencia, solo se consideran los datos de las estaciones de primavera y otoño. La idea de no incluir los datos de verano e invierno, se debe a que por un estudio preliminar (Castro, 1994), se determinó que en las estaciones de invierno y verano, por la presencia de los sistema anticiclonicos, la duración de los episodios difería de la duración durante las estaciones de primavera y otoño, esto es, identificaron que había homogeneidad dentro de primavera y otoño, y que en el par de estaciones verano e invierno se tenía otra agrupación de la duración de los episodios.

Para la velocidad mínima de viento, se identificó una distribución Weibul para mínimos, por lo que estos datos originales se transformaron a exponencial unitaria con una función de transformación creciente. En cambio para la máxima concentración de ozono se identificó una distribución Gumbel para máximos, así que las mediciones originales de concentración se transformaron a exponencial unitaria con una función de transformación

decreciente.

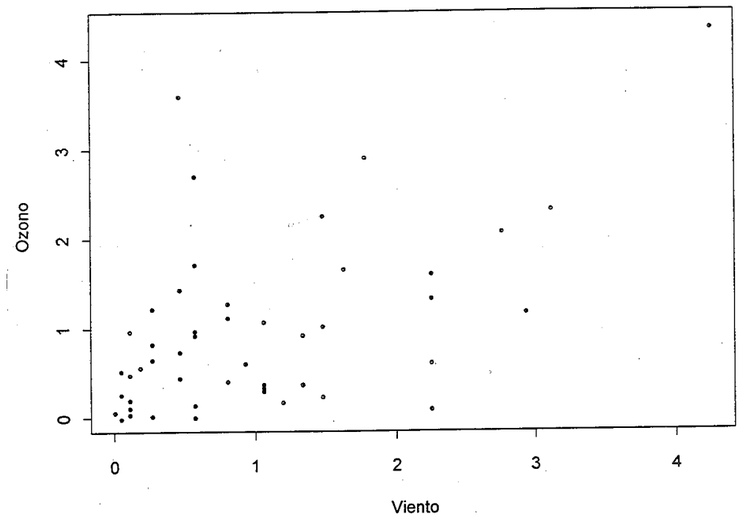


Figura 5.12. Valores extremos de velocidad mínima de viento y concentración máxima de ozono, transformadas a exponenciales unitarias.

Por el tipo de funciones de transformación utilizadas, tenemos que una asociación positiva en nuestros datos transformados, significa que en los datos originales hay una asociación negativa. Así, por ejemplo: aumento en la concentración de ozono y disminución en velocidad mínima de viento. Bajo el supuesto de que la inmovilidad de la masa de aire esta asociada con el aumento de la contaminación, tiene sentido que los datos originales tengan una asociación negativa y por lo tanto los datos transformados tengan una asociación positiva.

En la Figura 5.12, tenemos el diagrama de dispersión de los valores extremos de ozono y viento, transformados a exponenciales unitarias. En esta gráfica notamos un grado de asociación reducido entre las variables.

Tabla 5.12. Estimaciones de parámetros de modelos de dependencia, por máxima verosimilitud. Los datos son concentraciones máximas de ozono y velocidades mínimas de viento.

Modelo	$\hat{\theta}_1(d.e.)$	$\hat{\theta}_2(d.e.)$	$\hat{\theta}_3(d.e.)$	*Log.ver.*
M. 2B	$\nu_1 = .354(.926)$	$\nu_2 = 7.941(7.241)$	$\nu = 8.655(8.190)$	-89.1611
M. 2B sim.	$\nu_1 = .502(2.677)$	$\nu_2 = \nu - \nu_1$	$\nu = 8.439(.307)$	-89.272
L. Inv.	$\phi_1 = .999$	$\phi_2 = .999$	$\tau = .648$	-90.5432
L.Inv.Sim.			$\tau = .653(.147)$	-88.562
L. Asim.	$\theta = .999$	$\phi = .999$	$\alpha = 1.3782$	-89.0834
L. Sim.			$\alpha = 1.3784(.145)$	-89.0834

En la Tabla 5.12, tenemos las estimaciones de los parámetros por máxima verosimilitud, para los diferentes modelos de dependencia considerados en este trabajo. Los parámetros estimados para el modelo de mezcla de betas, reflejan un grado de dependencia reducido, ya que los valores de $\hat{\nu}_1$ y $\hat{\nu}_2$ están ubicados en los extremos del intervalo $(0, \hat{\nu})$.

Comparando el nivel de asociación con el observado en el segundo ejemplo de edades de muerte, encontramos que el nivel de asociación en mayor en este ejemplo de datos de ozono y viento, debido a que el valor de $\hat{\nu}$ es menor para el último ejemplo.

El modelo de mezcla de betas simétrico ajusta de una manera muy similar al general, como lo podemos ver en la Tabla 5.12. Esta similitud entre los modelos general y simétrico, también se observa para los modelos logístico asimétrico y logístico invertido, ya que en la Tabla 5.12, encontramos estimaciones con una diferencia reducida.

Los modelos logísticos detectan un grado de dependencia reducido. En el caso del modelo logístico asimétrico encontramos una estimación del parámetro α (1.3782), que corresponde a un grado de asociación pequeño.

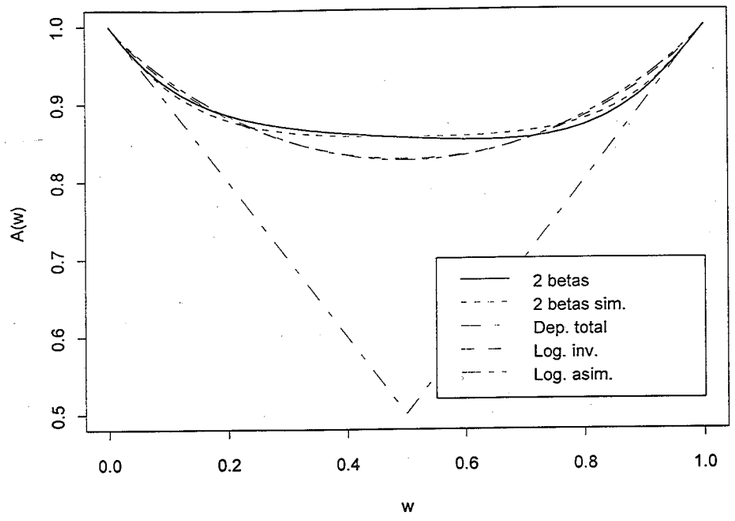


Figura 5.13. Funciones de dependencia para datos de ozono y velocidad de viento. Modelos mezcla de 2 betas asimétrico y simétrico, logístico asimétrico y logístico invertido.

En la Figura 5.13 tenemos las gráficas de las funciones de dependencia ajustadas, y encontramos que los modelos logísticos prácticamente coinciden entre ellos, así como también coinciden los de mezcla de betas general y simétrico. Considerando a $2(1-A(.5))$ como un índice de dependencia tenemos de las gráficas de estas funciones de dependencia, que el modelo de mezcla de betas registra una menor dependencia entre los extremos, que los modelos logísticos.

5.4.1 Inferencia en el esquema de cocientes

Al analizar estos datos en el esquema univariado de los cocientes $W = Y/(X+Y)$, iniciamos observando la Gráfica 5.14 de la función de distribución empírica, en donde

apreciamos un nivel de dependencia positiva reducido, ya que la función de distribución empírica, inicia por debajo de la diagonal y termina por arriba.

Aquí reconocemos que es más fácil detectar gráficamente la asociación positiva, cuando esta es reducida, a través de la función de distribución empírica de los cocientes, que del diagrama de dispersión de las parejas de extremos.

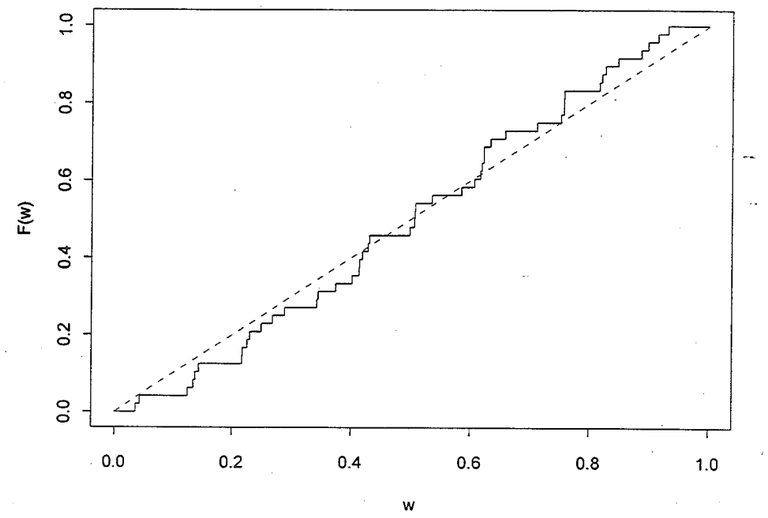


Figura 5.14. Función de distribución empírica para los cocientes w de valores extremos, provenientes de datos de velocidad de viento y concentración de ozono.

En la Figura 5.15, tenemos la función de verosimilitud relativa observada, que tiene un sesgo positivo al compararla con la verosimilitud relativa normal de la misma figura. Este sesgo es consistente con el sesgo encontrado para el estimador de máxima verosimilitud, en el estudio de simulación reportado en el Capítulo 3, con los valores de sesgo reportados en la Tabla 3.2.

En la Tabla 5.13 tenemos los estimadores de máxima verosimilitud y de momentos. Los intervalos de confianza normales aproximados se reportan en la Tabla 5.14 para

tres niveles de confianza. En estos observamos que el valor del parámetro $a = 1$, no se encuentra dentro de los intervalos de confianza dados.

Tabla 5.13. Estimaciones del parámetro a del modelo propuesto

	$\hat{a}$	$\widehat{Var}(\hat{a})$
Máxima Verosimilitud	1.3996	.0408
Momentos	1.3208	

Tabla 5.14. Intervalos de confianza, para el parámetro a del modelo propuesto, basados en la aproximación normal de la verosimilitud.

Confianza	80	90	95
Int. de confianza	(1.1407, 1.6585)	(1.0673, 1.7319	(1.0036, 1.7955)

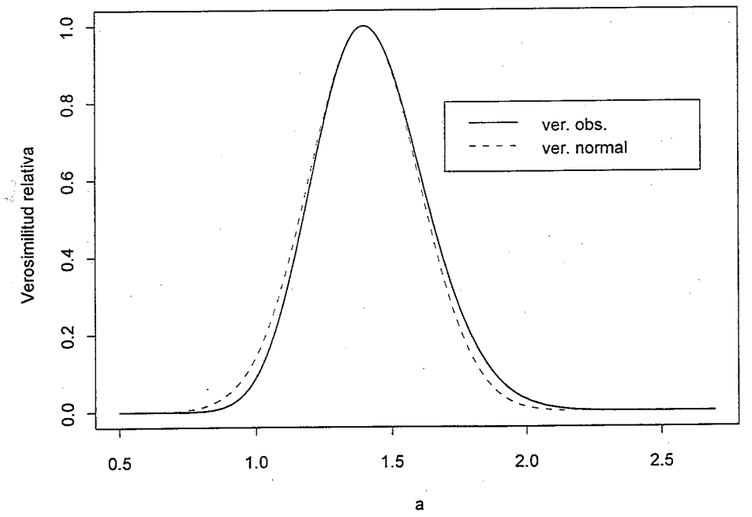


Figura 5.15. Funciones de verosimilitud relativa observada y normal, para el modelo propuesto y los datos de viento y ozono.

5.4.2 Esquema Bayesiano

Para los datos de este ejemplo y la distribución a priori dada en forma general por (3.25), obtenemos la distribución a posteriori gama truncada por la izquierda en uno, con parámetros de escala 34.296 y de forma 47.

En este caso el estimador Bayesiano del parámetro a dado por la esperanza (3.27), tiene un valor $\hat{a}_B = 1.3803$, menor que el de máxima verosimilitud, lo cual es congruente con el estudio de simulación del Capítulo 3. En la Figura 5.16 tenemos las funciones de densidad a priori y a posteriori.

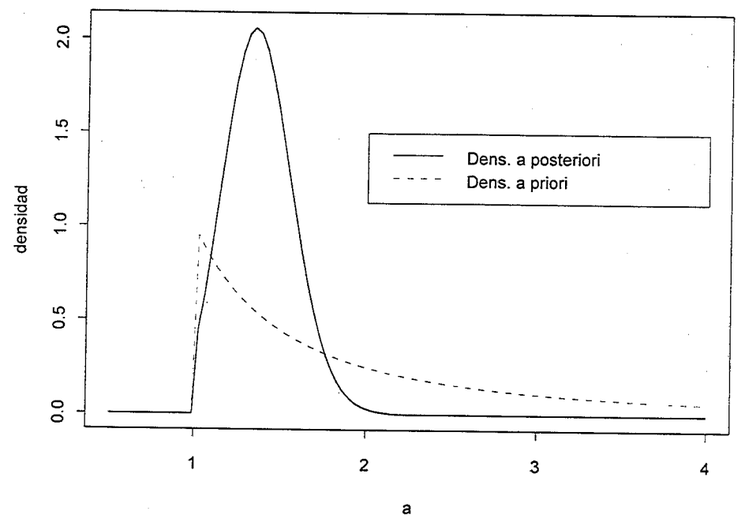


Figura 5.16. Funciones de densidad a priori y a posteriori de a para el modelo propuesto y los datos de viento y ozono.

Los intervalos de confianza Bayesianos de longitud mínima, que no contienen la orilla izquierda ($a = 1$) del espacio paramétrico, se obtienen haciendo cortes en la función de

densidad a posteriori, por arriba del valor $p = .3398$, estos intervalos de longitud mínima tienen probabilidades de contener al parámetro a, menores de .9637.

Capítulo 6

Conclusiones

El presente trabajo enfatiza el análisis de datos de valores extremos bivariados, basado en el esquema univariado de los cocientes $W = Y/(X+Y)$, en donde el patrón de dependencia entre los extremos se define a través de la función de distribución del cociente W. Para esto se presenta el enfoque paramétrico y no paramétrico.

Se propuso un modelo de mezcla de betas para la función de dependencia de valores extremos bivariada. Este modelo es bastante flexible, como todos los modelos de mezclas. La función de dependencia tiene una representación integral, definida por una medida positiva del intervalo [0,1]. Esta medida es equivalente a dos veces una función de distribución. El modelo de mezcla de betas propuesto, se construye dando una distribución de mezcla de betas para la medida positiva, que caracteriza a la función de dependencia. Este modelo permite simular observaciones bivariadas de valores extremos, con una gran variedad de patrones de dependencia.

La función de dependencia también puede definirse en términos de la función de distribución del cociente W. Se dá la relación entre el coeficiente de asociación tao de Kendall y la función de distribución F_W. Además se propone una distribución simétrica alrededor de $1/2$ para el cociente W, de aquí se obtiene un modelo para la función de dependencia simétrica.

Este nuevo modelo está definido por un parámetro a que puede interpretarse como

un parámetro de dependencia. En este modelo el espacio paramétrico es $[1,\infty)$, teniendo independencia cuando $a = 1$ y dependencia positiva cuando $a > 1$.

La distribución que define el nuevo modelo se encuentra en la familia exponencial, por lo cual existe un estadístico suficiente y completo $T(W_1, W_2, ..., W_n)$ que contiene en forma resumida la información de la muestra de los cocientes $W_1, W_2, ..., W_n$. El estimador de máxima verosimilitud es función de este estadístico, razón por la que es de mínima varianza, aunque es sesgado.

Se obtiene el estimador insesgado de mínima varianza uniformemente (IMVU) como función de $T(W_1, W_2, ..., W_n)$. Como este estadístico tiene distribución *Gama*, entonces obtienen estimaciones por intervalo a partir de una cantidad pivotal.

Se propone el esquema Bayesiano de inferencia, incorporando la información muestral a través de la verosimilitud. En la función de densidad de probabilidades a priori introducimos la información previa sobre dependencia positiva de los extremos. La función de densidades a priori dada a partir de la información que se tiene acerca de la estructura de dependencia, se encuentra en la familia conjugada.

La función de densidades a posteriori para a, es una gama truncada por la izquierda en uno, por lo tanto tiene como soporte el conjunto $[1,\infty)$. Entonces el estimador Bayesiano de a que se define a partir de la densidad a posteriori toma siempre valores que corresponden a dependencia positiva o independencia. Esto no siempre ocurre con el estimador de máxima verosimilitud, ya que cuando se tiene un nivel bajo de dependencia, la estimación máximo verosimil puede ser negativa.

Se propone un estimador convexo no paramétrico para la función de dependencia de valores extremos bivariados. Este estimador se construye sustituyendo la función de distribución F del cociente W por la función de distribución empírica correspondiente F_n, en la expresión de la función de dependencia definida en términos de dicha función de distribución.

Segun el estudio de simulación donde se compara el estimador no paramétrico propuesto con otros dos, tiene buen comportamiento en los casos de dependencia baja y

mediana, mientras que en los casos de dependencia alta, subestima el nivel de dependencia en la parte central del intervalo $[0, 1]$.

Se presentan tres ejemplos de aplicación, con niveles de dependencia baja, mediana y alto. En el caso de dependencia grande, se muestran algunas probabilidades de interés.

Como consecuencia de este trabajo, se puede explorar la posibilidad de extender el modelo simétrico de dependencia a uno asimétrico mas general.

CIM
BIBLIO

Apéndice A

Generación de observaciones multivariadas de valores extremos

A.1 Algoritmo de Pickands

En el caso de variables aleatorias univariadas podemos generar observaciones con una cierta distribución F transformando con la inversa F^{-1} de la función de distribución, observaciones generadas previamente con distribución uniforme $(0, 1)$. Cuando tenemos para un vector aleatorio Y una caracterización en términos de una función de variables aleatorias que se pueden generar facilmente entonces podemos generar relizaciones del vector aleatorio Y tomando funciones $Y = g(X_1, X_2, ..., X_l; \mathbf{a})$ de realizaciones de alguna variable aleatoria X , con vector de parametros $\mathbf{a} = (a_1, a_2, ..., a_l)'$ que corresponde de alguna manera a la parametrización de la variable Y que estamos generando. Este método se utiliza para generar observaciones con distribución normal univariada, como combinaciones lineales de variables aleatorias uniformes $(0, 1)$, con la justificación que permite el teorema central de límite. También pueden generarse observaciones con distribución normal multivariada transformando vectores formados por variables aleatorias independientes idénticamente distribuidas, que se pueden generar facilmente. Podemos generar relizaciones de un vector aleatorio con distribución multivariada de valores extremos con

marginales exponenciales, como se describe enseguida.

De acuerdo a la representación de Pickands, un vector aleatorio tiene una distribución multivariada de valores extremos con marginales exponenciales unitarias si y solo si la función de supervivencia multivariada se puede expresar de la siguiente forma

$$-\log S(\mathbf{y}) = -\log P\left(\bigcap_{k=1}^{K}\{Y_k > y_k\}\right) = \int_{S_k} \max_{1\le k\le K}\left(q_k^{(l)} y_k\right) dH(q) \qquad \text{(A.1)}$$

donde H es una medida positiva finita definida sobre el simplex S_K. Podemos deducir un algoritmo para generar observaciones con distribución multivariada de valores extremos a partir de la siguiente construcción de una variable aleatoria Y que tiene la distribución mencionada:

Sea L un entero positivo. Sean $U_1, U_2, ..., U_L$ una muestra aleatoria proveniente de una distribución exponencial unitaria, y sea la matriz $K \times L\ (a_{kl})$, $a_{kl} \geq 0$, donde cada renglón tiene al menos un término no cero, sea además,

$$Y_k = \min_{1\le l\le L} \frac{U_l}{a_{kl}},\ k = 1, 2, ..., K \qquad \text{(A.2)}$$

entonces

$$-\log S(\mathbf{y}) = \sum_{l=1}^{L}\left[\max_{1\le k\le K} y_k a_{kl}\right] = \sum_{l=1}^{L}\left[\max_{1\le k\le K} y_k q_k^{(l)}\right] c_l = \int_{S_k} \max_{1\le k\le K}\left(q_k^{(l)} y_k\right) dH^*(q) \qquad \text{(A.3)}$$

donde $a_{kl} = q_k^{(l)} c_l$, $q^{(l)} = (q_1^{(l)}, q_2^{(l)}, ..., q_K^{(l)})$ son elementos del simplex S_K, y además H^* es una medida positiva finita discreta tal que $H^*(q^{(l)}) = c_l$. Dados L puntos $q^{(1)}, q^{(2)}, ..., q^{(L)}$ en el simplex S_K y conociendo la medida discreta H^* podemos obtener los elementos de la matriz (a_{kl}).

Para generar las observaciones con una distribución multivariada de valores extremos determinada, necesitamos conocer la medida H asociada, que discretizamos de acuerdo a el conjunto $\left\{q^{(l)} \in S_K, l = 1, 2, ..., L\right\}$ para obtener H^* como una aproximación de H. De acuerdo a esto, podemos definir el algoritmo para generar vectores aleatorios de dimensión K con la distribución multivariada de valores extremos:

Algoritmo 2.1

0.- Definir la función de distribución característica H,

1.- Dar el conjunto de L de puntos $\left\{q^{(l)} \in S_K, l = 1, 2, ..., L\right\}$ para la discretización de H.

2.- Dar $c_l = H(q^{(l)})$, $l = 1, 2, ..., L$

3.- Calcular $a_{kl} = c_l q_k^{(l)}, k = 1, 2, ..., K, l = 1, 2, ..., L$

4.- Generar números aleatorios independientes $U_l \sim EXP(1),\ \ l = 1, 2, ..., L$

4.- Obtener $Y_k = \min_{1\le l\le L}\left(\frac{U_l}{a_{kl}}\right), k = 1, 2, ..., K$

Para el caso de observaciones bivariadas, el algoritmo es:

Algoritmo 2.2

Algoritmo para generar vectores aleatorios (X, Y), con distribución de valores extremos bivariada con marginales exponenciales, suponiendo conocida la función de dependencia A y por lo tanto también la función de distribución característica H, de la distribución de valores extremos multivariada.

0.- Definir la función de distribución característica H, y dar el número L de subintervalos para la discretización de H.

1.- Dar $q^{(l)} = l/L$, $l = 1, 2, ..., L$

2.- Calcular $c_l = H(q^{(l)})$, $l = 1, 2, ..., L$

3.- Dar $a_{1l} = c_l q^{(l)},\ \ a_{2l} = c_l[1 - q^{(l)}]$

4.-Generar números aleatorios independientes $U_l \sim EXP(1),\ \ l = 1, 2, ..., L$

4.- Dar $X = \min_{1 \leq l \leq L} \left(\frac{U_l}{a_{1l}} \right), \ Y = \min_{1 \leq l \leq L} \left(\frac{U_l}{a_{2l}} \right)$

La representación 2.2 tiene como casos particulares el caso de independencia cuando $K = L$ y la matriz (a_{kl}) es diagonal, mientras que cuando $L = 1 < K$, las componentes son completamente dependientes.

CIM
BIBLIO

Bibliografía

[1] Abramowitz, M. and Stegun, I. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables,* Dover, New York, 1965.

[2] Bernardo J. M.,and Smith A., F., M. *Bayesian theory,* J. Wiley, 1994.

[3] Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Massachusetts.

[4] Capéraà, P., Fougerés, A.-L., and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika,* **84**, 566-577.

[5] Carter, R. W. (1951). Floods in Georgia, Geological Survey Circular 100, March 1951.

[6] Castro, J.I. (1994). Distribución conjunta de valores extremos de ozono y de algunas variables meteorológicas. *Revista de Estadística,* Vol. VI, No. 8, 83-98.

[7] Coles, S. G. and J. A. Tawn. (1990). Statistical of coastal flood prevention. *Philosophical Transactions of the Royal Society,* A, **332**, 457-476.

[8] Coles, S. G. and J. A. Tawn. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society,* B, **53**, 377-392.

[9] Coles, S. G. and J. A. Tawn. (1994). Statistical Methods for Multivariate Extremes: an Application to Structural Design, *Applied Statistics,* **43**, 1-48.

[10] Csörgő, M and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics,* Wiley & Sons, Chichester.

[11] D'Agostino, R. B. and Stephens, *Goodness-of-Fit Techniques*, Marcel Dekker, 1986.

[12] Deheuvels, P. (1983). Point processes and multivariate extreme values. *Journal of Multivariate Analysis,***13**, 257-271.

[13] Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics.* New York: Wiley.

[14] Ghoudi, K., Khoudraji, A. and Rivest, L-P. (1998), Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles, *The Canadian Journal of Statistics,* **26**, 187-197.

[15] Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, **55**, 698-707.

[16] Gumbel, E. J. (1964). Analysis of empirical bivariate extremal distributions, *Journal of the American Statistical Association*, **59**, 794-816.

[17] Gumbel, E. J. and Goldstein, N. (1964). Analysis of empirical bivariate extremal distributions, *Journal of the American Statistical Association*, **59**, 794-816.

[18] Gumbel, E. J. and Mustafi, C. K. (1967). Some analytical properties of bivariate extremal distributions, *Journal of the American Statistical Association*, **62**, 269-288.

[19] Haan, L. De and S. I. Resnick (1977). Limit theory for multivariate sample extremes. *Zeitschrift fur Wahrscheinl. Theorie* **40**, 317-337.

[20] Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & Probability Letters, 9*, 75-81.

[21] Joe, H. (1993). Parametric family f multivariate distributions with given margins. *Journal of Multivariate Analysis,* **46**, 262-282.

[22] Joe, H. (1994). Multivariate extreme-value distributions with applications to environmental data. *The Canadian Journal of Statistics,* **22**, 47-64.

[23] Joe, H., Smith,R. L. & Weissman, I. (1989), Bivariate Threshold Methods for Extremes. *Journal of the Royal Statistical Society,* B., **54,** 171-183.

[24] Lai, T. L. (1974). Convergence rates in the strong law of large numbers for random variables taking values in Banach spaces, *Bulletin Institute of Mathematics Academy Sinica,* **2**, 67-85.

[25] NRC (National Research Council) (1991). *Rethinking the Ozone Problem in Urban and Regional Air Pollution,* National Academy Press, Washington, D. C.

[26] Pickands, J. (1981). Multivariate extreme value distribution. In *Proceedings of the 43rd Session of the International Statistics Institute,* (Buenos Aires), pp. 859-878.

[27] Posner, E. C., Roderick, E. R., Ashlock, J. C. and Lurie, S. (1969). Application of an estimator of high efficiency in bivariate extreme value theory. *Journal of the American Statistical Association*, **64**, 1403-1414.

[28] Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes.* New York: Springer Verlag.

[29] Rockafellar, R. T. (1970). *Convex Analysis.* New Jersey, Princeton University Press

[30] Rojo, J., Villa-Diharce, E. and Flores, M. (1997). Nonparametric Estimation of the Dependence Function in Bivariate Extreme Value Distributions. (enviado).

[31] Rojo, J. (1998). Estimation in the Quantile Function of an IFRA distribution function. *Scandinavian Journal of Statistics,* **25**, 293-310.

[32] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, London.

[33] Smith, R. L. (1985). Statistics of extreme values. *Bulletin of the International Statistics Institute,* **11**, 1-17.

[34] Smith, R.L. (1989), Extreme value analysis of environmental time series: an example based on ozone data (with discussion). *Statistical Science,* **4,** 367-393.

[35] Smith, R.L., Tawn, J.A. and Yuen, H.K. (1990), Statistics of Multivariate Extremes. *International Statistics Review,* **58,1,** 47-58.

[36] Smith, R.L., Tawn, J.A. and Coles, S.G. (1997), Markov chain models for threshold exceedances. *Biometrika,* **84,** 249-258.

[37] StatSci (1996). S-PLUS *for Windows Users Manual, Version 3.3.* Seattle, Wash.

[38] Tawn, J. A. (1988). Bivariate extreme value theory-models and estimation. *Biometrika,* **75**, 397-415.

[39] Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika,* **77**, 245-253.

[40] Tiago de Oliveira, J. (1980). Bivariate extremes; foundations and statistics. in *Multivariate Analysis V*, ed. P.R. Krishnaiah, pp. 349-66. Amsterdam: North Holland.

[41] Tiago de Oliveira, J. (1984). Bivariate models for extremes. in *Statistical Extremes and Applications* edited by J. Tiago de Oliveira, pp. 131-153. D. Riedel Publishing Company.

[42] Tiago de Oliveira, J. (1989). Intrinsic estimation of the dependence structure for bivariate extremes. *Statistics & Probability Letters, 8*, 213-218.

[43] Tiago de Oliveira, J. (1992). Intrinsic estimation of the dependence function in bivariate extremes: a new and simpler approach. *Communications in Statistics, Theory and Methods, A* **21**, 599-611.

[44] Villa E. (1996). Modelling and Estimating the Dependence Function in Bivariate Extreme Value Distributions. *Comunicación Técnica (I-96-05, PE/CIMAT)*

[45] Yuen, H. K. (1990). *Estimation of multivariate extreme value distributions by a kernel method with an application to non-Gaussian time series.* Ph. D. Thesis, University of Surrey.