



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**SOBRE LAS MEDIDAS DE DEPENDENCIA
DE VARIABLES ALEATORIAS**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

A C T U A R I O

P R E S E N T A :

GERARDO RUBIO HERNANDEZ



DIRECTOR DE TESIS
DRA. MARIA ASUNCION BEGONA FERNANDEZ FERNANDEZ



FACULTAD DE CIENCIAS
SECCION ESCOLAR

2004



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

ESTA TESIS NO SALE
DE LA BIBLIOTECA



UNIVERSIDAD NACIONAL
DE INGENIERÍA
MILITARIA

Envío a la Dirección General de Bibliotecas de la
UNI para su depósito en formato electrónico e impreso el
presente documento de mi trabajo recepcional.

NOMBRE: Gerardo Rubio
Hernández

FECHA: 23 de Enero de 2004

FIRMA: [Firma]

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:
Sobre las Medidas de Dependencia de Variables Aleatorias

realizado por Gerardo Rubio Hernández con número de cuenta 9850869-9

quien cubrió los créditos de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis

Propietario Dra. María Asunción Begoña Fernández Fernández *Begoña Fernández*

Propietario Dr. Federico Jorge O'Reilly Togno *[Firma]*

Propietario Dr. José María González Barrios Murguía *José M. González Barrios*

Suplente Dra. Ana Meda Guardiola *Ana Meda*

Suplente M. en C. Arturo Erdely Ruíz *[Firma]*

Consejo Departamental de Matemáticas

[Firma]
M. en C. José Antonio Flores Díaz

A mi papá y mi mamá

Por fin llegó el día de agradecer a todos ustedes por haber estado de tantas formas conmigo durante todo este tiempo. Antes que a cualquier otro le doy las gracias a mi familia, en especial a ustedes dos, papá y mamá, por haberme dejado ser yo durante todo este tiempo, por comprenderme y apoyarme en lo que yo necesitara. De verdad estoy muy agradecido. Quiero darte las gracias a ti también Begoña, por todo el tiempo que dedicaste a nuestro trabajo y sobre todo por ser mi amiga más que mi asesora. Nelson, hermano, gracias por todo lo que has hecho por mí, porque de verdad creo que gran parte de este trabajo es tuyo también. Les doy las gracias también a mis sinodales: Federico, José María, Ana y Arturo. Gracias por su tiempo y todas las observaciones que hicieron sobre mi trabajo, sé que ayudó a que éste mejorara. Por último y no porque sean menos, sino porque ustedes saben esperar, les quiero dar las gracias a todos mis amigos que de alguna manera tuvieron que ver con mi vida en la Universidad, gracias a ustedes, Jonhy, Liz, David, Ana, Tania, Guillermo, Betsa, Che Che, Gaby, Alicia, Ro, George, Otir, Lu, Luigi, Chabela, Lizzette, Marifi, Lecus, Benja, Yvo, Mampito, Sandra, Marcela, Gaby, Yunis, Gladys, Mike, Martí, Alejandro, a todos los Panchos, Pedro, Franco, Jorginho, Parca, Mike, Marco, Carlitos, Raúl, Alan, Édgar y en especial a Carlos. Muchas gracias a todos por aparecer en mi vida.

Contenido

Introducción	3
1 Preliminares	7
1.1 Introducción	7
1.2 Independencia y Medidas de Dependencia	7
1.3 Definiciones básicas	13
1.4 Cópulas	16
1.5 Teorema de Sklar	20
1.6 Cópulas y variables aleatorias	23
2 Medidas de dependencia	33
2.1 Introducción	33
2.2 Correlación.	34
2.3 Concordancia	54
2.3.1 τ de Kendall.	55
2.3.2 ρ de Spearman	77
2.4 Propiedades de dependencia	86
2.4.1 Dependencia de cuadrante	86
2.4.2 Monotonicidad en la cola	90
2.4.3 Monotonicidad estocástica	94
2.5 Medidas de dependencia	99
2.6 Reordenamientos de M	111
3 Condiciones de “mixing”	115
3.1 Introducción	115
3.2 Medidas de “mixing”	116
3.3 Dependencia en procesos estocásticos	120

4	Aplicaciones	137
4.1	Medida de dependencia muestral	137
4.2	Modelos de series de tiempo	142
4.3	Simulaciones y Aplicaciones	145
	Conclusiones	161
A	Resultados de probabilidad	163
B	Programas	167

Introducción

Uno de los conceptos básicos en la Teoría de Probabilidad, que incluso se estudia en los cursos elementales en la licenciatura, es el de la dependencia de variables aleatorias. Esto no quiere decir de ninguna manera que este concepto sea simple o que esté completamente estudiado, por el contrario, existe una gran cantidad de trabajos recientes en la literatura (véase por ejemplo [3], [9], [10], [17])) que estudian distintos aspectos de la dependencia de variables aleatorias. Recordemos que la dependencia de variables aleatorias se define como la negación de la independencia, más precisamente:

Consideremos X y Y variables aleatorias con función de distribución conjunta H y distribuciones marginales F y G , respectivamente. Decimos que X y Y son independientes si

$$H = FG$$

y decimos que son dependientes si no son independientes. Algunos ejemplos de parejas de variables aleatorias dependientes son:

- La pareja (X, Y) con $Y = \psi(X)$, donde ψ es una función biyectiva.
- Si X y Z son independientes y consideramos la pareja (X, Y) con $Y = X + Z$.
- La pareja (X, Y) con distribución conjuntamente normal y parámetro $\rho \neq 0$.

La finalidad de esta tesis es la de estudiar distintas medidas de dependencia y analizar la información que éstas nos dan acerca de la relación que guardan las variables aleatorias entre sí. Comúnmente las medidas de dependencia se interpretan como un número λ que satisface las siguientes propiedades:

1. λ está definida para cualquier pareja de variables aleatorias X y Y ;

2. $\lambda(X, Y) = \lambda(Y, X)$;
3. existe $0 < M < \infty$ tal que $0 \leq \lambda(X, Y) \leq M$;
4. $\lambda(X, Y) = 0$ si y sólo si X y Y son independientes;
5. $\lambda(X, Y) = M$ si y sólo si X ó Y es una función monótona de Y ó X ;
6. si α y β son funciones estrictamente monótonas en el rango de X y de Y , respectivamente, se cumple $\lambda(X, Y) = \lambda(\alpha(X), \beta(Y))$.

Las propiedades 1 y 2 son naturales. El pedir que sea acotada nos permite tener un parámetro de comparación, que podría no tener sentido si ésta fuera infinita. La propiedad 4 es la que les da sentido a estas medidas; asigna un único valor a la independencia, es decir, nos indica si las variables son independientes o no. La propiedad 5 pretende caracterizar al “extremo opuesto” a la independencia, el problema está en cómo se interprete éste. Tal vez, se podría pedir que el valor $\lambda = M$ corresponda a parejas que se relacionan por una función biyectiva, de tal manera que basta conocer a una para conocer a las dos, sin embargo, se proponen estas propiedades porque es lo más común en la literatura (véase [9] y [15]). La última propiedad pretende reflejar en la medida de dependencia la invarianza de la dependencia bajo transformaciones monótonas. Esta afirmación no es clara, es decir, no es evidente que la dependencia no cambie si aplicamos una transformación monótona, sin embargo, la incluimos por la misma razón que a la propiedad anterior.

Lo primero que observamos es que en general será difícil que un sólo número sea capaz de resumir toda la información que guarda la función de distribución conjunta y en consecuencia, de medir la dependencia. Por lo tanto, en cada caso estudiado, parte de la problemática consistirá en entender qué significa, en relación con la dependencia, cada uno de los valores que toma la medida. En otras palabras, trataremos de contestar si un valor más grande que otro significa una dependencia más “fuerte”.

Otro tipo de medidas que estudiaremos son las medidas de concordancia: decimos que κ es una medida de concordancia si satisface,

1. κ está definida para cualquier pareja de variables aleatorias X y Y ;
2. $\kappa(X, Y) = \kappa(Y, X)$;
3. existe $0 < M < \infty$ tal que $-M \leq \kappa(X, Y) \leq M$;

4. si X y Y son independientes, entonces $\kappa(X, Y) = 0$;
5. $\kappa(X, Y) = M$ si y sólo si X ó Y es una función monótona creciente de Y ó X ;
6. $\kappa(X, Y) = -M$ si y sólo si X ó Y es una función monótona decreciente de Y ó X ;
7. si α y β son funciones estrictamente monótonas en el mismo sentido en el rango de X y de Y respectivamente, se cumple $\kappa(X, Y) = \kappa(\alpha(X), \beta(Y))$, si α y β son funciones estrictamente monótonas en sentido contrario en el rango de X y Y respectivamente, se cumple que $\kappa(X, Y) = -\kappa(\alpha(X), \beta(Y))$.

La finalidad de estas medidas se explica en las propiedades 5 y 6. Éstas caracterizan a aquellas parejas de variables aleatorias que son concordantes o discordantes entre sí. La motivación de las otras propiedades que se piden es similar a la de las medidas de dependencia. La principal diferencia es que mientras las medidas de dependencia caracterizan a la independencia, las de concordancia hacen una diferencia entre la “dependencia positiva” y la “dependencia negativa”. Las medidas que estudiaremos son: la correlación, la τ de Kendall y la ρ de Spearman, que corresponden a medidas de concordancia. También estudiaremos dos medidas de dependencia:

$$\sigma(X, Y) = 12 \iint_{\mathbb{R}^2} |H(x, y) - F(x)G(y)| dx dy;$$

$$\delta(X, Y) = \sup_{x, y \in \mathbb{R}} |H(x, y) - F(x)G(y)|,$$

así como una contraparte estadística de δ . La finalidad de hacer un exhaustivo estudio teórico de las medidas es la de tener los elementos suficientes para, ante la presencia de datos reales, poder deducir de manera más acertada algunas conclusiones sobre su comportamiento.

El trabajo de esta tesis se dividió en 4 capítulos. El capítulo uno contiene las definiciones y el material que necesitaremos a lo largo del trabajo. Daremos y motivaremos la definición de eventos independientes, para de ahí definir en general qué es la independencia en la probabilidad. También daremos las definiciones de medida de dependencia y de medida de concordancia, estudiando las características de cada una y analizando sus diferencias.

Además estudiaremos una familia de funciones llamadas cópulas. Estas funciones las podemos interpretar como distribuciones conjuntas con marginales uniformes en el $(0,1)$. Su importancia recae en el Teorema de Sklar, que dice que la función de distribución de toda pareja de variables aleatorias continuas la podemos escribir como una única cópula compuesta con las funciones de distribución marginales.

El capítulo dos está dedicado a las tres medidas de concordancia más conocidas: la correlación, la τ de Kendall y la ρ de Spearman. Utilizaremos una gran variedad de ejemplos para explicar lo que estas medidas realmente están calculando y en consecuencia, cuándo es adecuado utilizarlas. Asimismo, estudiaremos las dos medidas de dependencia mencionadas antes (σ y δ).

En el capítulo tres, extenderemos el estudio de dependencia a procesos estocásticos. Haremos un breve estudio sobre la independencia que puede haber dentro de éstos mediante el uso de las condiciones llamadas "mixing", además revisaremos algunos ejemplos para tratar de aclararlas.

Para el cuarto y último capítulo, analizaremos un método estadístico para el estudio de la independencia basado en la medida de dependencia δ estudiada en el capítulo dos. También veremos un método para aplicar el estadístico propuesto a modelos de series de tiempo. Por último estudiaremos para parejas de variables aleatorias y modelos de series de tiempo, cómo es el rendimiento de los métodos propuestos mediante el uso de la simulación de datos. Para concluir la tesis, estudiaremos los precios del petróleo y trataremos, con todo lo estudiado, de encontrar un periodo para el que dos observaciones del precio del petróleo sean independientes si distan más de ese periodo. Se incluyen además dos apéndices: en el primero enunciamos algunos resultados de probabilidad. En el segundo se encuentran los programas principales utilizados para la sección de aplicaciones del capítulo cuatro.

Capítulo 1

Preliminares

1.1 Introducción

Este capítulo, como su nombre lo dice, lo dedicaremos a estudiar las herramientas que requeriremos a lo largo de este trabajo. La primera sección corresponde a la definición y un breve estudio de la independencia. Definiremos el concepto de medida de dependencia, y motivaremos dicha definición. Estudiaremos también una familia de funciones conocidas como cópulas que nos serán de mucha utilidad para el estudio de la dependencia. En la Sección 1.3 daremos herramientas que necesitaremos para su estudio; en la Sección 1.4 definiremos lo que es una cópula y revisaremos sus propiedades básicas. En la Sección 1.5 veremos el Teorema de Sklar, que es el teorema que da gran importancia a las cópulas. La Sección 1.6 se encargará de revisar la relación que existe entre las cópulas y las variables aleatorias.

Para la Sección 1.2 se utilizó como referencia [1], [11], [15] y [16]. Para las Secciones 1.3 a la 1.6 se usó [15] como bibliografía.

1.2 Independencia y Medidas de Dependencia

A lo largo de todo el trabajo consideraremos $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad en donde estarán definidas las variables aleatorias. Además, en el caso de que X sea una variable aleatoria, utilizaremos la notación

$$\mathbb{P}\{\{\omega \in \Omega | X(\omega) \in A\}\} = \mathbb{P}\{X \in A\}.$$

Consideremos el lanzamiento de dos monedas justas, es decir, ninguna de sus caras está favorecida al realizarse el volado. Los posibles resultados de los lanzamientos son $\{A, A\}$, $\{A, S\}$, $\{S, A\}$ y $\{S, S\}$, donde a cada uno le asignamos la probabilidad $\frac{1}{4}$. Claramente el resultado del lanzamiento de la primera moneda no influye en el resultado de la segunda. En términos de probabilidad esto lo reflejamos al hacer que la probabilidad de un resultado de la segunda no se altere al conocer el resultado de la primera, por ejemplo, si definimos $E = \{\text{águila en el primer lanzamiento}\}$ y $F = \{\text{sol en el segundo lanzamiento}\}$, entonces $\mathbb{P}[F|E] = \mathbb{P}[F]$, o bien despejando $\mathbb{P}[E \cap F] = \mathbb{P}[E] \mathbb{P}[F]$, lo que sucede pues

$$\mathbb{P}[E \cap F] = \mathbb{P}[\{A, S\}] = \frac{1}{4},$$

mientras que

$$\mathbb{P}[E] = \mathbb{P}[\{A, S\} \cup \{A, A\}] = \frac{1}{2}$$

y de la misma manera

$$\mathbb{P}[F] = \mathbb{P}[\{A, S\} \cup \{S, S\}] = \frac{1}{2}$$

de donde obtenemos que

$$\mathbb{P}[E \cap F] = \mathbb{P}[E] \mathbb{P}[F].$$

En este caso decimos que los eventos son independientes, es decir, no dependen uno del otro. De todas las formas posibles en que dos eventos pueden depender, la más conocida y estudiada es en la que no dependen, es decir, la independencia. Demos entonces una definición formal de lo que estamos hablando.

Definición 1.2.1 (Eventos Independientes). Sean A y B dos eventos, es decir, $A, B \in \mathcal{F}$. Decimos que A y B son independientes si $\mathbb{P}[A \cap B] = \mathbb{P}[A] \mathbb{P}[B]$.

Definición 1.2.2 (σ -álgebras Independientes). Sean \mathcal{A} y \mathcal{B} dos sub σ -álgebras de \mathcal{F} , es decir, $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$. Decimos que \mathcal{A} y \mathcal{B} son independientes si para todo $A \in \mathcal{A}$ y $B \in \mathcal{B}$, A y B son independientes.

A lo largo de este trabajo se estudiarán principalmente variables aleatorias, por lo que necesitamos definir la independencia entre variables aleatorias.

Definición 1.2.3 (Variables Aleatorias Independientes). Sean X y Y dos variables aleatorias. Decimos que X y Y son independientes si $\sigma(X)$ y $\sigma(Y)$ son independientes, donde $\sigma(\cdot)$ denota la σ -álgebra generada por \cdot .

Ésta es la definición general de independencia entre variables aleatorias, el problema está en que comprobar que dos variables aleatorias son independientes es demasiado complicado. La independencia la podemos definir, también, a través de las funciones de distribución.

Sean X y Y dos variables aleatorias. La función de distribución conjunta del vector (X, Y) , $H : \mathbb{R}^2 \rightarrow [0, 1] \times [0, 1]$, la definimos como

$$H(x, y) = \mathbb{P}[X \leq x, Y \leq y],$$

y las funciones de distribución marginales, $F : \mathbb{R} \rightarrow [0, 1]$ y $G : \mathbb{R} \rightarrow [0, 1]$, como

$$F(x) = \mathbb{P}[X \leq x]$$

y

$$G(y) = \mathbb{P}[Y \leq y].$$

Definición 1.2.4 (Independencia). Decimos que las variables X y Y son independientes si para todo $x, y \in \mathbb{R}$, se cumple

$$H(x, y) = F(x)G(y).$$

Aunque ésta no es la definición general de independencia entre variables aleatorias, sí es equivalente (véase Apéndice A) y esta forma es mucho más tratable. Si además las variables aleatorias son continuas, es decir, tienen densidad f y g^1 , respectivamente, son independientes si y sólo si

$$h(x, y) \stackrel{c.s.}{=} f(x)g(y),$$

donde h es la función de densidad conjunta.

La independencia es la forma más conocida de relación entre las variables aleatorias, pero no es la más interesante y mucho menos la única. Uno de los problemas de mayor interés es, ante dos variables aleatorias, ¿cómo se relacionan éstas? La identificación de esta relación no es un problema de ninguna manera trivial, y debe ser tratado con cuidado. El hecho de que las variables aleatorias a simple vista parezcan dependientes, no lo implica de ninguna manera. Veamos algunos ejemplos.

¹Una variable tiene densidad, si existe una función $f \geq 0$ integrable, tal que $\mathbb{P}[X \in A] = \int_A f(x)dx$

Ejemplo 1.2.1. Sean X y Y dos variables aleatorias independientes $N(0,1)$. Consideremos la transformación $W = X + Y$ y $Z = X - Y$. Las variables W y Z , al ser función de X y Y , es de esperarse que no sean independientes entre sí. Calculemos su función de densidad conjunta. El jacobiano está dado por

$$J = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = -2.$$

Como X y Y son independientes, su función de densidad conjunta es

$$f(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$$

y entonces la función de densidad conjunta de W y Z es

$$\begin{aligned} g(w, z) &= \frac{1}{2} f\left(\frac{w+z}{2}, \frac{w-z}{2}\right) \\ &= \frac{1}{4\pi} e^{-\frac{1}{4}\left(\frac{(w+z)^2}{2} + \frac{(w-z)^2}{2}\right)} \\ &= \frac{1}{\sqrt{4\pi}} e^{-\frac{w^2}{4}} \frac{1}{\sqrt{4\pi}} e^{-\frac{z^2}{4}}. \end{aligned}$$

Por lo tanto W y Z son dos variables aleatorias $N(0, 2)$ independientes. \square

Veamos otro ejemplo

Ejemplo 1.2.2. Sean X y Y dos variables aleatorias Gamma de parámetros (α, λ) y (β, λ) , respectivamente. Sean $W = X + Y$ y $Z = \frac{X}{X+Y}$. Otra vez, las variables W y Z no parecen ser independientes. Veamos que sí lo son. El jacobiano es

$$J = \begin{vmatrix} \frac{1}{Y} & \frac{1}{X} \\ \frac{1}{(X+Y)^2} & -\frac{1}{(X+Y)^2} \end{vmatrix} = -\frac{1}{X+Y} = -\frac{1}{W}.$$

Y la transformación inversa es $X = WZ$ y $Y = W(1 - Z)$. La función de densidad conjunta de X y Y es

$$f(x, y) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda(x+y)} x^{\alpha-1} y^{\beta-1} \mathbb{I}_{(0,\infty)}(x) \mathbb{I}_{(0,\infty)}(y),$$

por lo tanto, la densidad conjunta de W y Z queda como

$$\begin{aligned} g(w, z) &= |w|f(wz, w(1-z)) \\ &= \frac{\lambda e^{-\lambda w} (\lambda w)^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)} \mathbb{I}_{(0,\infty)}(w) \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} z^{\alpha-1} (1-z)^{\beta-1} \mathbb{I}_{(0,1)}(z), \end{aligned}$$

y utilizando el hecho que $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, obtenemos que W y Z son dos variables aleatorias independientes; W con distribución Gamma de parámetros $(\alpha + \beta, \lambda)$ y Z con distribución Beta (α, β) . \square

Con los ejemplos anteriores vemos que la independencia es una propiedad de la probabilidad y no significa necesariamente que no exista ninguna relación entre las variables aleatorias, es decir, en general las parejas definidas en los ejemplos anteriores no serán independientes si cambiamos las distribuciones marginales.

Procedamos ahora a definir las medidas de dependencia.

Definición 1.2.5. Una medida numérica λ de asociación entre dos variables aleatorias X y Y (que denotaremos $\lambda(X, Y)$) es una *medida de dependencia* si satisface las siguientes propiedades:

1. λ está definida para cualquier pareja de variables aleatorias X y Y ;
2. $\lambda(X, Y) = \lambda(Y, X)$;
3. existe $0 < M < \infty$ tal que $0 \leq \lambda(X, Y) \leq M$;
4. $\lambda(X, Y) = 0$ si y sólo si X y Y son independientes;
5. $\lambda(X, Y) = M$ si y sólo si X ó Y es una función monótona de Y ó X ;
6. si α y β son funciones estrictamente monótonas en el rango de X y de Y , respectivamente, se cumple $\lambda(X, Y) = \lambda(\alpha(X), \beta(Y))$.

Las propiedades anteriores se explican de la siguiente forma. La conmutatividad es clara pues la dependencia entre (X, Y) es la misma que (Y, X) . Pedimos que exista una cota finita para poder tener una referencia y estandarizar cualquier pareja de variables aleatorias que estudiemos. Si no hay cota, cómo podemos decir que un valor es grande o pequeño. La propiedad 4 la pedimos para poder caracterizar con un sólo valor la independencia, de

tal manera que con ésta podamos decir si las variables son independientes o no entre sí. La propiedad 5, como ya dijimos al principio del trabajo, la pedimos porque es lo más común en la literatura, aunque no queda claro que el “extremo opuesto” a la independencia sea esta forma de relacionarse. Por último, la invarianza bajo transformaciones monótonas la incluimos otra vez por ser lo más común aunque no es claro que la dependencia entre las variables no cambie al aplicar un par de transformaciones monótonas. Observemos que este tipo de medidas no distingue si las variables tienen una relación creciente o decreciente. Para ese caso, alternativamente, podemos pedir a una medida lo siguiente.

Definición 1.2.6. Una medida numérica κ de asociación entre dos variables aleatorias X y Y (que denotaremos $\kappa(X, Y)$) es una *medida de concordancia* si satisface las siguientes propiedades:

1. κ está definida para cualquier pareja de variables aleatorias X y Y ;
2. $\kappa(X, Y) = \kappa(Y, X)$;
3. existe $0 < M < \infty$ tal que $-M \leq \kappa(X, Y) \leq M$;
4. si X y Y son independientes, entonces $\kappa(X, Y) = 0$;
5. $\kappa(X, Y) = M$ si y sólo si X ó Y es una función monótona creciente de Y ó X ;
6. $\kappa(X, Y) = -M$ si y sólo si X ó Y es una función monótona decreciente de Y ó X ;
7. si α y β son funciones estrictamente monótonas en el mismo sentido en el rango de X y de Y respectivamente, se cumple $\kappa(X, Y) = \kappa(\alpha(X), \beta(Y))$, si α y β son funciones estrictamente monótonas en sentido contrario en el rango de X y Y respectivamente, se cumple que $\kappa(X, Y) = -\kappa(\alpha(X), \beta(Y))$.

La motivación para cada una de las propiedades de este tipo de medidas es la misma que en el caso anterior sólo que ahora estamos haciendo la diferencia entre una “dependencia positiva” y una “dependencia negativa”. La principal diferencia es que, mientras que la primera está enfocada hacia la independencia, la segunda lo está hacia cuán concordantes o discordantes son. El tipo de medida adecuado dependerá, entonces, de lo que estemos interesados en saber acerca de la pareja (X, Y) .

1.3 Definiciones básicas

La cópula es una función que relaciona a la distribución conjunta con sus respectivas marginales, también la podemos ver como una distribución conjunta con marginales uniformes, restringida al cuadrado unitario. Claro está que lo anterior no es una definición matemáticamente formal, y eso es lo que haremos en lo que resta de este capítulo.

Pero antes, veamos de qué estamos hablando. Si tenemos una pareja de variables aleatorias (X, Y) , que tienen función de distribución conjunta $H(x, y)$ y marginales $F(x)$ y $G(y)$, tenemos que para cualquier pareja de reales (x, y) , la pareja $(F(x), G(y)) \in [0, 1] \times [0, 1]$ y el número $H(x, y) \in [0, 1]$ están relacionados por una función que llamaremos cópula. Pero para lograr lo anterior, antes necesitamos tratar de generalizar el concepto de función creciente para dos o más dimensiones.

La idea de esta sección es presentar la definición y algunas propiedades que tiene una función de dos dimensiones que llamaremos “bicreciente” o “2-creciente”. La finalidad de una función así, es la de generalizar a dos dimensiones la noción que tenemos en una dimensión de una función creciente. Para esto, denotaremos como $\overline{\mathbb{R}}$ a los reales extendidos $[-\infty, \infty]$. Al intervalo $[0, 1]$ lo denotaremos como \mathbf{I} .

Definición 1.3.1 (H-Volumen). Sea $H : \overline{\mathbb{R}}^2 \rightarrow \mathbb{R}$ una función, cuyo dominio es $S_1 \times S_2$, donde S_1 y S_2 son dos subconjuntos no vacíos de $\overline{\mathbb{R}}$. Sea $B = [x_1, x_2] \times [y_1, y_2]$ un rectángulo cuyos vértices están en el dominio de H . El H -volumen de B , está dado por

$$V_H(B) = H(x_2, y_2) - H(x_1, y_2) - H(x_2, y_1) + H(x_1, y_1).$$

Notemos que la definición anterior no es otra que la medida de un rectángulo, en un espacio de medida, cuya medida está generada por H .

Definamos ahora lo que es una función “2-creciente”.

Definición 1.3.2. Una función $H : \overline{\mathbb{R}}^2 \rightarrow \mathbb{R}$ es 2-creciente si para todo rectángulo B cuyos vértices pertenecen al dominio de H , se cumple que $V_H(B) \geq 0$.

Otra forma de ver esta definición es: una función es bicreciente si asigna una medida mayor o igual a cero a todo rectángulo, es decir, si es una medida positiva.

El comportamiento anterior que tiene una función al “avanzar” en dos dimensiones, ¿tendrá algo que ver con su comportamiento en una sola? La

respuesta es no. Sin algunas restricciones extras, nada tienen que ver, es decir, si una función es bicreciente, no implica que sea creciente en cada una de sus entradas, y viceversa.

Ejemplo 1.3.1. Sea $H(x, y) = (2x - 1)(2y - 1)$ restringida a \mathbf{I}^2 , $H : \mathbf{I}^2 \rightarrow \mathbb{R}$. Veamos primero qué pasa con un cuadrado $B = [x_1, x_2] \times [y_1, y_2] \subseteq \mathbf{I}^2$.

$$\begin{aligned} V_H(B) &= (2x_2 - 1)(2y_2 - 1) - (2x_1 - 1)(2y_2 - 1) - (2x_2 - 1)(2y_1 - 1) \\ &\quad + (2x_1 - 1)(2y_1 - 1) \\ &= 2(- (x_2 + y_2 + x_1 + y_1) + (x_2 + y_2 + x_1 + y_1)) \\ &\quad + 4(x_2y_2 - x_1y_2 - x_2y_1 + x_1y_1) + 2 - 2 \\ &= 4(x_2 - x_1)(y_2 - y_1) \\ &\geq 0. \end{aligned}$$

Por lo tanto H definida así, es una función 2-creciente, pero si consideramos una x fija en el intervalo $[0, \frac{1}{2})$, la función H es decreciente en y . \square

Ejemplo 1.3.2. Sea $H = \max(x, y)$ restringida a \mathbf{I}^2 , $H : \mathbf{I}^2 \rightarrow \mathbb{R}$. Es fácil ver que H es creciente en x y y , sin embargo, si consideramos $B = \mathbf{I}^2$ tenemos

$$V_H(B) = \max(1, 1) - \max(0, 1) - \max(1, 0) + \max(0, 0) = -1 < 0.$$

Por lo tanto, H no es 2-creciente. \square

Lo dos ejemplos anteriores nos muestran que no existe una implicación directa entre una función bicreciente y una función creciente en sus entradas. Nos preguntamos entonces, qué propiedad necesitará una función para ser bicreciente y creciente en sus entradas, y para ello necesitamos un par de lemas y una definición.

Lema 1.3.1. Sean S_1 y S_2 dos subconjuntos no vacíos de $\overline{\mathbb{R}}$ y sea H una función bicreciente con dominio $S_1 \times S_2$. Sean $x_1, x_2 \in S_1$, con $x_1 \leq x_2$, y sean $y_1, y_2 \in S_2$, con $y_1 \leq y_2$. Entonces la función $t \mapsto H(t, y_2) - H(t, y_1)$ es creciente en S_1 y la función $t \mapsto H(x_2, t) - H(x_1, t)$ es creciente en S_2 .

Demostración. Sean $t_1, t_2 \in S_1$, tal que $t_1 \leq t_2$. Si consideramos el cuadrado $B = [t_1, t_2] \times [y_1, y_2]$, por ser H bicreciente, tenemos que $V_H(B) = H(t_2, y_2) - H(t_1, y_2) - H(t_2, y_1) + H(t_1, y_1) \geq 0$, de donde es inmediato que $H(t_1, y_2) - H(t_1, y_1) \leq H(t_2, y_2) - H(t_2, y_1)$, y por lo tanto la función es creciente. La prueba es análoga para la otra función. \square

Para que una función 2-creciente sea creciente en sus entradas nos bastaría que en cada coordenada hubiera un elemento donde la función se anulara, sin importar el valor que toma la otra coordenada, además, éste tendría que ser el mínimo para cumplir las hipótesis del Lema 1.3.1. Lo anterior nos motiva a la siguiente definición.

Definición 1.3.3. Sea $H : \overline{\mathbb{R}}^2 \rightarrow \mathbb{R}$ una función, cuyo dominio es $S_1 \times S_2$, donde S_1 y S_2 son dos subconjuntos no vacíos de $\overline{\mathbb{R}}$. Si S_1 tiene un mínimo elemento a_1 y S_2 tiene un mínimo elemento a_2 , decimos que H es anclada si $H(a_1, y) = 0 = H(x, a_2)$ para todo $(x, y) \in S_1 \times S_2$.

Con la definición anterior y la explicación dada antes, el siguiente Lema es inmediato.

Lema 1.3.2. Sean S_1 y S_2 , dos subconjuntos no vacíos de $\overline{\mathbb{R}}$ y sea H una función 2-creciente y anclada con dominio $S_1 \times S_2$. Entonces H es creciente en cada uno de sus argumentos.

Demostración. Sean a_1 y a_2 los elementos mínimos de S_1 y S_2 respectivamente. Si hacemos $x_1 = a_1$ y $y_1 = a_2$ en el Lema 1.3.1 y consideramos el hecho de que H es anclada, el resultado se sigue inmediatamente. \square

Supongamos ahora que S_1 tiene un máximo elemento b_1 y S_2 tiene un máximo elemento b_2 . Decimos que una función $H : S_1 \times S_2 \rightarrow \mathbb{R}$, tiene marginales F y G si:

$$\text{Dom}F = S_1 \text{ y } F(x) = H(x, b_2) \text{ para toda } x \in S_1$$

$$\text{Dom}G = S_2 \text{ y } G(y) = H(b_1, y) \text{ para toda } y \in S_2.$$

Veamos el siguiente ejemplo.

Ejemplo 1.3.3. Sea H una función con dominio $[-1, 1] \times [0, \infty]$ dada por

$$H(x, y) = \frac{(x+1)(e^y - 1)}{x + 2e^y - 1}.$$

H es anclada ya que $H(x, 0) = 0$ y $H(-1, y) = 0$, además tiene marginales $F(x)$ y $G(y)$ dadas por:

$$F(x) = \frac{x+1}{2} \quad \text{aplicando L'Hopital y}$$

$$G(y) = 1 - e^y.$$

\square

Antes de concluir esta pequeña introducción a las funciones bicrecientes, veamos una propiedad importante que tiene una función bicreciente, anclada y con marginales.

Lema 1.3.3. Sean S_1 y S_2 dos subconjuntos no vacíos de $\overline{\mathbb{R}}$ y sea H una función 2-creciente, anclada y con marginales, con dominio $S_1 \times S_2$. Sean (x_1, y_1) y (x_2, y_2) dos puntos cualesquiera de $S_1 \times S_2$. Entonces

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|.$$

Demostración. Aplicando la desigualdad del triángulo tenemos que

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |H(x_2, y_2) - H(x_1, y_2)| + |H(x_1, y_2) - H(x_1, y_1)|$$

Ahora sabemos por el Lema 1.3.2 que H es creciente en x , por lo que si suponemos que $x_1 \leq x_2$, tenemos que $H(x_2, y_2) - H(x_1, y_2) \geq 0$. Por el Lema 1.3.1 sabemos que la función anterior es creciente para y , y como $y_2 \leq \max S_2$, entonces tenemos que $0 \leq H(x_2, y_2) - H(x_1, y_2) \leq F(x_2) - F(x_1)$. Obtenemos una ecuación similar si $x_1 \geq x_2$. Por lo tanto tenemos en general que $|H(x_2, y_2) - H(x_1, y_2)| \leq |F(x_2) - F(x_1)|$. De manera similar podemos obtener que $|H(x_1, y_2) - H(x_1, y_1)| \leq |G(y_2) - G(y_1)|$, lo que completa la prueba. \square

1.4 Cópulas

Con los elementos vistos en la sección anterior, nos encontramos en posición de definir lo que es una cópula. Para ello primero definiremos lo que es una subcópula, y luego veremos que la cópula es un caso especial de subcópula. No definiremos directamente la cópula porque más adelante veremos que en ciertos casos, como son las variables aleatorias discretas, las subcópulas son más adecuadas.

Definición 1.4.1. Una subcópula de dimensión dos o simplemente una subcópula, es una función C' que tiene las siguientes propiedades:

1. $Dom C' = S_1 \times S_2$, donde S_1 y S_2 son subconjuntos de \mathbf{I} que contienen al 0 y al 1.
2. C' es anclada y 2-creciente.

3. Para todo $u \in S_1$ y $v \in S_2$,

$$C(u, 1) = u \quad \text{y} \quad C(1, v) = v$$

Definición 1.4.2. Una *cópula de dimensión dos* o simplemente una *cópula*, es una subcópula con dominio \mathbf{I}^2

Como habíamos dicho al principio del capítulo, una cópula la podemos interpretar como una función de distribución restringida al cuadrado unitario con marginales $U(0, 1)$, es decir, una cópula es una función $C : \mathbf{I}^2 \rightarrow \mathbf{I}$ con las siguientes propiedades:

1. Para todo $u, v \in \mathbf{I}$,

$$C(u, 0) = 0 = C(0, v)$$

y

$$C(u, 1) = u \quad \text{y} \quad C(1, v) = v.$$

2. Para todo $u_1, u_2, v_1, v_2 \in \mathbf{I}$ tales que $u_1 \leq u_2$ y $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0$$

La idea del uso de las cópulas, es la de encontrar alguna forma de estudiar la dependencia que hay entre variables aleatorias. A continuación veremos un resultado y luego le daremos una interpretación probabilista.

Teorema 1.4.1. *Sea C' una subcópula. Para todo $(u, v) \in \text{Dom}C'$, se cumple*

$$\max(u + v - 1, 0) \leq C'(u, v) \leq \min(u, v). \quad (1.1)$$

Demostración. Sea (u, v) un punto arbitrario de el dominio de C' . Por el Lema 1.3.2 sabemos que $C'(u, v) \leq C'(u, 1) = u$ y $C'(u, v) \leq C'(1, v) = v$, por lo tanto $C'(u, v) \leq \min(u, v)$. Ahora si consideramos el cuadrado $B = [u, 1] \times [v, 1]$, por ser bicreciente, $V_{C'}(B) \geq 0$, de donde podemos ver que $u + v - 1 \leq C'(u, v)$ y por otro lado $C'(u, v) \geq C'(0, v) = 0$ y por lo tanto $\max(u + v - 1, 0) \leq C'(u, v)$ lo que concluye la prueba. \square

Dado que cada cópula es una subcópula, la ecuación anterior se cumple para ellas también. Pero, ¿quiénes son estas cotas de la ecuación (1.1)? Primero veamos que son cópulas. Generalmente se denotan como $M(u, v) = \min(u, v)$ y $W(u, v) = \max(u + v - 1, 0)$.

Ejemplo 1.4.1. La función $M : \mathbf{I}^2 \rightarrow \mathbf{I}$ definida arriba es una cópula ya que:

1. Sea $u \in \mathbf{I}$ entonces

$$M(u, 0) = \min(u, 0) = 0$$

y

$$M(u, 1) = \min(u, 1) = u.$$

Análogo para $v \in \mathbf{I}$.

2. Sea $u_1, u_2, v_1, v_2 \in \mathbf{I}$ tal que $u_1 \leq u_2$ y $v_1 \leq v_2$, entonces

$$M(u_2, v_2) - M(u_1, v_2) - M(u_2, v_1) + M(u_1, v_1) \geq 0$$

ya que

$$\min(u_2, v_2) - \min(u_2, v_1) \geq \min(u_1, v_2) - \min(u_1, v_1) \geq 0$$

Por lo tanto es una cópula. □

Ejemplo 1.4.2. La función $W : \mathbf{I}^2 \rightarrow \mathbf{I}$ definida arriba es una cópula ya que:

1. Sea $u \in \mathbf{I}$ entonces

$$W(u, 0) = \max(u - 1, 0) = 0$$

y

$$W(u, 1) = \max(u, 0) = u$$

Análogo para $v \in \mathbf{I}$.

2. Sea $u_1, u_2, v_1, v_2 \in \mathbf{I}$ tal que $u_1 \leq u_2$ y $v_1 \leq v_2$, entonces

$$W(u_2, v_2) - W(u_1, v_2) - W(u_2, v_1) + W(u_1, v_1) \geq 0$$

ya que

$$\begin{aligned} \max(u_2 + v_2 - 1, 0) - \max(u_2 + v_1 - 1, 0) &\geq \\ \max(u_1 + v_2 - 1, 0) - \max(u_1 + v_1 - 1, 0) &\geq 0 \end{aligned}$$

Por lo tanto es una cópula. \square

Ya tenemos algunas cópulas, veamos a continuación que a partir de dos cópulas podemos formar una infinidad más.

Ejemplo 1.4.3. Si B y D son dos cópulas cualesquiera, entonces la función $C_\lambda = \lambda B + (1 - \lambda)D$ con $\lambda \in \mathbf{I}$ es también una cópula, es decir, cualquier combinación convexa de cópulas es una cópula:

1. $C_\lambda : \mathbf{I}^2 \rightarrow \mathbf{I}$ ya que B y D son cópulas.
2. Sea $u \in \mathbf{I}$ entonces

$$C_\lambda(u, 0) = \lambda 0 + (1 - \lambda)0 = 0$$

y

$$C_\lambda(u, 1) = \lambda u + (1 - \lambda)u = u$$

Análogo para $v \in \mathbf{I}$.

3. Sea $u_1, u_2, v_1, v_2 \in \mathbf{I}$ tal que $u_1 \leq u_2$ y $v_1 \leq v_2$, entonces

$$\begin{aligned} V_{C_\lambda}([u_1, u_2] \times [v_1, v_2]) &= \lambda V_B([u_1, u_2] \times [v_1, v_2]) \\ &\quad + (1 - \lambda)V_D([u_1, u_2] \times [v_1, v_2]) \geq 0. \end{aligned}$$

Por lo tanto C_λ es una cópula. \square

Ya vimos que son cópulas, y entonces, por lo dicho antes, representan un modelo de probabilidad con marginales $U(0,1)$. Pero la pregunta es cuál modelo. Consideremos una variable aleatoria $X \sim U(0,1)$, ahora, la pareja (X, X) tiene distribución conjunta M . El caso de W lo satisface la pareja $(X, 1 - X)$. Lo que tenemos son dos casos extremos de dependencia en el caso de parejas $U(0,1)$, es decir, cualquier pareja de uniformes que consideremos, su distribución está acotada por formas de dependencia muy fuertes, decreciente por debajo, y creciente por encima. Esto nos empieza a hablar del uso de cópulas para medir la dependencia de variables aleatorias. Dado que estamos hablando de dependencia, existe otra cópula importante, la de la independencia, que usualmente se denota por $\Pi(u, v) = uv$.

Otra propiedad importante de las cópulas es la continuidad. El siguiente teorema establece que éstas no son sólo conjuntamente continuas sino Lipschitz continuas.

Teorema 1.4.2. *Sea C' una subcópula. Para todo u_1, u_2, v_1, v_2 en $\text{Dom}C'$, se cumple que*

$$|C'(u_2, v_2) - C'(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

Demostración. La demostración se sigue inmediatamente del Lema 1.3.3 y de las marginales de una subcópula. \square

1.5 Teorema de Sklar

El teorema que da nombre a esta sección es el que da la importancia a las cópulas para su uso probabilístico. El Teorema de Sklar permite ver que la cópula es la función que relaciona a la distribución conjunta con sus marginales. Antes, hablemos un poco de lo que es una función de distribución conjunta.

Definición 1.5.1. Una *función de distribución conjunta*, es una función H con dominio $\overline{\mathbb{R}}^2$ tal que,

1. $H(x, -\infty) = 0 = H(-\infty, y)$ y $H(\infty, \infty) = 1$,
2. H es 2-creciente.

H es anclada, y como $\text{Dom}H = \overline{\mathbb{R}}^2$, entonces H tiene marginales F y G dadas por $F(x) = H(x, \infty)$ y $G(y) = H(\infty, y)$. Además por el Lema 1.3.2 y las propiedades de la distribución conjunta, F y G son funciones de distribución.

Antes de seguir notemos que las distribuciones de las que estamos hablando son más generales que aquellas que aparecen en un modelo de probabilidad. El hecho de no considerar la continuidad por la derecha se debe a que las propiedades anteriores nos bastan para el teorema antes mencionado. Por lo tanto, lo hecho a continuación, servirá también cuando estemos hablando de variables aleatorias y sus respectivas funciones de distribución.

Teorema 1.5.1 (Teorema de Sklar). *Sea H una función de distribución conjunta con marginales F y G . Entonces existe una cópula C , tal que para todo $x, y \in \overline{\mathbb{R}}$, tenemos*

$$H(x, y) = C(F(x), G(y)). \tag{1.2}$$

Si F y G son continuas, entonces C es única; si no, C es única en $\text{Ran}F \times \text{Ran}G$. De la misma manera, si C es una cópula y F y G son funciones de distribución, entonces la función H definida por (1.2) es una función de distribución conjunta con marginales F y G .

Antes de empezar con la demostración, notemos que la contraparte es inmediata de las propiedades de una cópula y de una función de distribución. La importancia del teorema está en que, como su nombre lo dice, es la cópula la que une a las marginales, es decir, es el elemento restante para obtener la distribución conjunta, una vez que se tiene el comportamiento marginal. Es por eso que este resultado hace a la cópula un ente muy importante en el estudio de la dependencia entre variables aleatorias.

La demostración en general se sigue con dos lemas, el primero establece que existe una función que relaciona a la función de distribución conjunta con sus marginales y que esta función resulta ser una subcópula. El segundo lema dice que toda subcópula puede ser extendida a una cópula. La aparición de las subcópulas porque las marginales no son obligatoriamente continuas, en el caso de serlo el resultado no es tan complicado.

Lema 1.5.1. *Sea H una función de distribución conjunta con marginales F y G . Entonces existe una única subcópula C' tal que*

1. $\text{Dom}C' = \text{Ran}F \times \text{Ran}G$,
2. para todo x, y en $\overline{\mathbb{R}}$, $H(x, y) = C'(F(x), G(y))$.

Demostración. La función de distribución H cumple las hipótesis del Lema 1.3.3, por lo tanto, si consideramos $S_1 = S_2 = \overline{\mathbb{R}}$, tenemos que para todos los puntos (x_1, y_1) y (x_2, y_2) en $\overline{\mathbb{R}}^2$ se cumple

$$|H(x_1, y_1) - H(x_2, y_2)| \leq |F(x_1) - F(x_2)| + |G(y_1) - G(y_2)|.$$

De la desigualdad anterior notemos que si $F(x_1) = F(x_2)$ y $G(y_1) = G(y_2)$ entonces $H(x_1, y_1) = H(x_2, y_2)$, es decir, que tenemos una correspondencia donde a cada punto del conjunto $\text{Ran}F \times \text{Ran}G$ le corresponde un único punto del $\text{Ran}H$. Con lo dicho antes, el conjunto de parejas ordenadas

$$\{((F(x), G(y)), H(x, y)) \mid x, y \in \overline{\mathbb{R}}\}$$

definen una función $C' : \overline{\mathbb{R}}^2 \rightarrow \overline{\mathbb{R}}$ con dominio $\text{Ran}F \times \text{Ran}G$. Que esta función C' sea una subcópula se sigue directamente de las propiedades de H ,

F y G . El $\text{Ran}F$ está contenido en \mathbf{I} ya que F es creciente y $H(-\infty, \infty) = F(-\infty) = 0$ y $H(\infty, \infty) = F(\infty) = 1$, por lo que además $0, 1 \in \mathbf{I}$. Análogo para $\text{Ran}G$. C' es anclada ya que $C'(0, G(y)) = H(-\infty, y) = 0 = H(x, -\infty) = C'(F(x), 0)$. El hecho que sea bicreciente es directo de la definición de H . Y por último $C'(F(x), 1) = H(x, \infty) = F(x)$ y similar para la otra marginal. Por lo tanto C' es una subcópula. \square

Notemos que si F y G son continuas tenemos que $\text{Ran}F \times \text{Ran}G = \mathbf{I}^2$, es decir, C' es a una cópula y el Teorema de Sklar queda demostrado. El problema está entonces cuando F ó G no son funciones continuas. La ventaja está en que toda subcópula puede ser extendida a una cópula definiéndola de manera adecuada en los puntos que no pertenecen a su dominio. Para eso es el segundo lema de la demostración.

Lema 1.5.2. *Sea C' una subcópula. Entonces, existe una cópula C tal que $C(u, v) = C'(u, v)$ para toda pareja $(u, v) \in \text{Dom}C'$.*

Demostración. Sea $\text{Dom}C' = S_1 \times S_2$. Por el hecho de que C' es creciente en sus entradas y acotada, podemos extender el dominio a $\overline{S_1} \times \overline{S_2}$, donde $\overline{S_i}$ denota la cerradura de S_i . Gracias al Teorema 1.4.2 esta extensión está bien definida. Sea (a, b) un punto cualquiera en \mathbf{I}^2 , gracias a que el dominio extendido de C' es cerrado, sabemos que existen un máximo y un mínimo elemento $a_1, a_2 \in \overline{S_1}$ tales que $a_1 \leq a \leq a_2$; de la misma manera existen $b_1, b_2 \in \overline{S_2}$ tales que $b_1 \leq b \leq b_2$. Sea

$$\lambda_1 = \begin{cases} \frac{a-a_1}{a_2-a_1}, & \text{si } a \notin \overline{S_1}; \\ 1, & \text{si } a \in \overline{S_1}, \end{cases} \quad \mu_1 = \begin{cases} \frac{b-b_1}{b_2-b_1}, & \text{si } b \notin \overline{S_2}; \\ 1, & \text{si } b \in \overline{S_2}. \end{cases}$$

Definamos con esto

$$C(a, b) = (1 - \lambda_1)(1 - \mu_2)C'(a_1, b_1) + (1 - \lambda_1)\mu_1C'(a_1, b_2) + \lambda_1(1 - \mu_1)C'(a_2, b_1) + \lambda_1\mu_1C'(a_2, b_2). \quad (1.3)$$

Notemos que la interpolación anterior aunque luzca extraña, no es otra cosa que cosiderar el cuadrado $[a_1, a_2] \times [b_1, b_2]$, dividirlo en cuatro secciones con la esquina interior de cada una dada por el punto (a, b) . Luego realizamos un promedio ponderado dependiendo de la cercanía del punto (a, b) a cada una de las esquinas del cuadrado $[a_1, a_2] \times [b_1, b_2]$ asignándole como medida el tamaño del cuadrado opuesto a la esquina en consideración.

De manera directa podemos ver que $\text{Dom}C = \mathbf{I}^2$; que $C(a, b) = C'(a, b)$ para todo $(a, b) \in \text{Dom}C'$; y que C satisface $C(u, 0) = 0 = C(0, v)$ y $C(u, 1) = u$, $C(1, v) = v$ para todo $u, v \in \mathbf{I}$. Por lo tanto, sólo resta probar que esta función sea bicreciente, para ello consideremos otro punto (c, d) en \mathbf{I}^2 que cumpla que $a \leq c$ y $b \leq d$, y sean $c_1, c_2, d_1, d_2, \lambda_2, \mu_2$ los puntos relacionados a (c, d) . Para evaluar $V_C(B)$ para el rectángulo $B = [a, c] \times [b, d]$ tenemos muchos casos ya que entre a y c puede haber cero, uno o más puntos de $\overline{S_1}$, siendo lo mismo para b y d . El caso más sencillo es que entre ninguno de ellos haya un punto de tal forma que tenemos que $a_i = c_i$ y $b_i = d_i$ para $i = 1, 2$. En ese caso tenemos que

$$V_C([a, c] \times [b, d]) = C(c, d) - C(c, b) - C(a, d) + C(a, b)$$

donde para cada miembro del lado derecho aparece una vez $C(a_i, b_j)$ multiplicado por distintos factores, por ejemplo para $C(a_1, b_1)$ tenemos que su factor es

$$\begin{aligned} (1 - \lambda_1)(1 - \mu_1) - (1 - \lambda_1)(1 - \mu_2) \\ - (1 - \lambda_2)(1 - \mu_1) + (1 - \lambda_2)(1 - \mu_2) \\ = (1 - \lambda_1)(\mu_2 - \mu_1) - (1 - \lambda_2)(\mu_2 - \mu_1) \\ = (\lambda_2 - \lambda_1)(\mu_2 - \mu_1). \end{aligned}$$

De la misma manera hacemos para los demás miembros de donde obtenemos que

$$V_C([a, c] \times [b, d]) = (\lambda_2 - \lambda_1)(\mu_2 - \mu_1)V_{C'}([a_1, a_2] \times [b_1, b_2])$$

que es positivo ya que C' es subcópula y $\mu_2 - \mu_1 = (d - b)/(b_2 - b_1) \geq 0$, mismo caso para λ . Éste es el caso más sencillo, de la misma manera se prueba para todos los otros casos. Por lo tanto C es una cópula. \square

1.6 Cópulas y variables aleatorias

Nos encontramos ahora en posición de relacionar a las cópulas con variables aleatorias. Hasta el momento solamente habíamos mencionado que las cópulas las podíamos relacionar con distribuciones conjuntas de variables aleatorias uniformes. Veamos qué pasa con variables aleatorias distintas. Casi todos los resultados de esta sección serán sobre variables aleatorias con

distribución continua y esto se debe a lo visto en el Teorema de Sklar (la no unicidad de la cópula en el caso no continuo). Cuando hablemos de la función de distribución de X , una variable aleatoria, nos referiremos a $F(x) = \mathbb{P}[X \leq x]$, lo mismo con las distribuciones conjuntas.

El teorema de Sklar puede ser visto de nuevo en términos de variables aleatorias.

Teorema 1.6.1. *Sean X y Y dos variables aleatorias con funciones de distribución F y G , respectivamente, y función de distribución conjunta H . Entonces existe una cópula C_{XY} tal que*

$$H(x, y) = C_{XY}(F(x), G(y))$$

para todo $x, y \in \mathbb{R}$. Si F y G son continuas, C_{XY} es única. Si no, C_{XY} es única en $\text{Ran}F \times \text{Ran}G$.

Hemos dicho que las cópulas nos dan la información de cómo se relaciona una pareja de variables aleatorias, por lo que es bastante natural pensar que debe haber una cópula que caracterice la independencia. El siguiente teorema confirma lo dicho y su prueba se sigue directamente del Teorema de Sklar y de saber que X y Y son independientes si y sólo si $H(x, y) = F(x)G(y)$ para todo $x, y \in \mathbb{R}$.

Teorema 1.6.2. *Sean X y Y variables aleatorias continuas. Entonces X y Y son independientes si y sólo si $C_{XY} = \Pi$.*

Si tenemos un vector aleatorio (X, Y) y dos transformaciones α y β estrictamente crecientes, un nuevo vector $(\alpha(X), \beta(Y))$ deberá tener el mismo comportamiento desde el punto de vista conjunto. Claro está que las probabilidades, es decir, las funciones de distribución conjuntas, serán distintas, pero la forma en que las variables se relacionan no tendrá por qué cambiar. Ésta es una propiedad importante dentro de las cópulas, que son invariantes bajo transformaciones estrictamente crecientes, y en general, bajo transformaciones monótonas, la nueva cópula tiene una relación bastante directa con la cópula anterior como veremos adelante.

Teorema 1.6.3. *Sean X y Y dos variables aleatorias continuas con cópula C_{XY} . Si α y β son dos transformaciones estrictamente crecientes en $\text{Ran}X$ y $\text{Ran}Y$ respectivamente, entonces $C_{\alpha(X)\beta(Y)} = C_{XY}$. Por lo tanto, C_{XY} es invariante bajo transformaciones estrictamente crecientes.*

Demostración. Sean F_1, G_1, F_2 y G_2 las respectivas funciones de distribución de $X, Y, \alpha(X)$ y $\beta(Y)$. Como α y β son estrictamente crecientes tenemos que $F_2(x) = \mathbb{P}[\alpha(X) \leq x] = \mathbb{P}[X \leq \alpha^{-1}(x)] = F_1(\alpha^{-1}(x))$ y de la misma forma $G_2(y) = G_1(\beta^{-1}(y))$. Para todo $x, y \in \mathbb{R}$ tenemos

$$\begin{aligned} C_{\alpha(X)\beta(Y)}(F_2(x), G_2(y)) &= \mathbb{P}[\alpha(X) \leq x, \beta(Y) \leq y] \\ &= \mathbb{P}[X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)] \\ &= C_{XY}(F_1(\alpha^{-1}(x)), G_1(\beta^{-1}(y))) \\ &= C_{XY}(F_2(x), G_2(y)) \end{aligned}$$

Como X y Y son continuas, $\text{Ran}F_2 = \text{Ran}G_2 = \mathbf{I}$, por lo tanto $C_{\alpha(X)\beta(Y)} = C_{XY}$ en \mathbf{I}^2 . \square

Si las transformaciones α ó β son estrictamente monótonas, no necesariamente crecientes, tenemos que la cópula $C_{\alpha(X)\beta(Y)}$ es una transformación de la cópula C_{XY} . De manera específica tenemos:

Teorema 1.6.4. *Sean X y Y dos variables aleatorias continuas con cópula C_{XY} . Sean α y β dos transformaciones estrictamente monótonas en $\text{Ran}X$ y $\text{Ran}Y$ respectivamente.*

1. Si α es creciente y β decreciente, entonces

$$C_{\alpha(X)\beta(Y)}(u, v) = u - C_{XY}(u, 1 - v).$$

2. Si α es decreciente y β creciente, entonces

$$C_{\alpha(X)\beta(Y)}(u, v) = v - C_{XY}(1 - u, v).$$

3. Si α y β son decrecientes, entonces

$$C_{\alpha(X)\beta(Y)}(u, v) = u + v - 1 + C_{XY}(1 - u, 1 - v).$$

Demostración. Consideremos las mismas funciones de la prueba del Teorema 1.6.3. Ahora para el primer inciso tenemos que $F_2(x) = F_1(\alpha^{-1}(x))$ y que

$G_2(y) = 1 - G_1(\beta^{-1}(y))$, con lo que obtenemos que

$$\begin{aligned}
 C_{\alpha(X)\beta(Y)}(F_2(x), G_2(y)) &= \mathbb{P}[\alpha(X) \leq x, \beta(Y) \leq y] \\
 &= \mathbb{P}[X \leq \alpha^{-1}(x), Y \geq \beta^{-1}(y)] \\
 &= \mathbb{P}[X \leq \alpha^{-1}(x)] - \mathbb{P}[X \leq \alpha^{-1}(x), Y < \beta^{-1}(y)] \\
 &= F_1(\alpha^{-1}(x)) - C_{XY}(F_1(\alpha^{-1}(x)), G_1(\beta^{-1}(y))) \\
 &= F_2(x) - C_{XY}(F_2(x), 1 - G_2(y))
 \end{aligned}$$

y se concluye igual que para el Teorema 1.6.3. La prueba para los incisos 2 y 3 son análogas a ésta. \square

Lo que en efecto nos dicen los Teoremas 1.6.3 y 1.6.4, es que bajo transformaciones monótonas la estructura de dependencia de las variables aleatorias no cambia, excepto en el sentido de ésta.

Hasta este momento, hemos insistido en la importancia que tienen las cópulas en el sentido de la dependencia de variables aleatorias. Dimos una primera caracterización con cópulas para el caso de independencia, así que es de esperar que suceda lo mismo en los casos extremos, es decir, cuando una variable aleatoria es una función monótona de la otra y por lo tanto el conocimiento de cualquiera de ellas nos permite conocer a la pareja.

En el Teorema 1.4.1 vimos cotas universales para las cópulas, dadas por

$$W(u, v) = \max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v) = M(u, v)$$

para cualquier cópula C y para todo $u, v \in \mathbf{I}$. Por el Teorema de Sklar, si tenemos X y Y variables aleatorias con función de distribución H y marginales F y G respectivamente, tenemos que para todo $x, y \in \mathbb{R}$

$$\max(F(x) + G(y) - 1, 0) \leq H(x, y) \leq \min(F(x), G(y))$$

y como M y W son cópulas, entonces las cotas de arriba son funciones de distribución con marginales F y G , conocidas como las cotas de Fréchet-Hoeffding. Como son funciones de distribución, ¿qué podemos decir de un vector aleatorio (X, Y) cuya función de distribución H sea igual a alguna de estas dos cotas? Para contestar esta pregunta necesitamos definir algunos conceptos.

Definición 1.6.1. Un subconjunto S de $\overline{\mathbb{R}}^2$ es estrictamente creciente si para todo (x, y) y (u, v) en S , $x < u$ implica que $y < v$. De la misma manera, un subconjunto S de $\overline{\mathbb{R}}^2$ es estrictamente decreciente si para todo (x, y) y (u, v) en S , $x < u$ implica que $y > v$.

Los conjuntos definidos arriba son conjuntos que se ven como curvas, no necesariamente continuas, estrictamente monótonas. Si tuviéramos alguna sección del conjunto que no fuera una curva (por ejemplo, un área), siempre podríamos encontrar una pareja de puntos que contradijeran las características de dichos conjuntos.

Lo que vamos a probar es que una función de distribución H es la cota superior de Fréchet-Hoeffding si y sólo si el soporte² de H es un conjunto estrictamente creciente, y es la cota inferior de Fréchet-Hoeffding si y sólo si el soporte de H es un conjunto estrictamente decreciente. Para eso necesitamos probar algunos lemas antes.

Lema 1.6.1. *Sea S un subconjunto de $\overline{\mathbb{R}}^2$. S es estrictamente creciente si y sólo si para cada $(x, y) \in \overline{\mathbb{R}}^2$ o bien,*

1. *para todo $(u, v) \in S$, $u \leq x$ implica que $v \leq y$; o bien,*
2. *para todo $(u, v) \in S$, $v \leq y$ implica que $u \leq x$.*

Demostración. Primero, sea S un conjunto estrictamente creciente, si no ocurre ninguno de los casos del lema tenemos que existen puntos (a, b) y (c, d) en S que cumplen que $a \leq x$ y $b > y$ y que $c \leq y$ y $d > x$ lo que implica que $a < d$ y $b > d$ lo que no puede ocurrir para dos puntos en S .

Segundo, si S no es un conjunto estrictamente creciente, tenemos puntos (a, b) y (c, d) en S que cumplen que $a < c$ y $b \geq d$. Si consideramos el punto $(x, y) = (\frac{a+c}{2}, \frac{b+d}{2})$, no se cumplen ninguna de las dos implicaciones del lema, lo que negándolo nos da la otra implicación de la demostración. \square

Lema 1.6.2. *Sea S un subconjunto de $\overline{\mathbb{R}}^2$. S es estrictamente decreciente si y sólo si para cada $(x, y) \in \overline{\mathbb{R}}^2$ o bien,*

1. *para todo $(u, v) \in S$, $u \leq x$ implica que $v > y$; o bien,*
2. *para todo $(u, v) \in S$, $v > y$ implica que $u \leq x$.*

² S es el soporte de H si para todo conjunto A medible que cumple que $A \cap S = \emptyset$ se tiene que $V_H(A) = 0$.

Demostración. Primero, sea S un conjunto estrictamente decreciente, si no ocurre ninguno de los casos del lema tenemos que existen puntos (a, b) y (c, d) en S que cumplen que $a \leq x$ y $b \leq y$ y que $c > y$ y $d > x$ lo que implica que $a < d$ y $b < c$ lo que no puede ocurrir para dos puntos en S .

Segundo, si S no es un conjunto estrictamente decreciente, tenemos puntos (a, b) y (c, d) en S que cumplen que $a < c$ y $b \leq d$. Si consideramos el punto $(x, y) = (\frac{a+c}{2}, \frac{b+d}{2})$, no se cumplen ninguna de las dos implicaciones del lema, lo que negándolo nos da la vuelta de la demostración. \square

Lo que los lemas anteriores nos dicen es que, por ejemplo, un conjunto es estrictamente creciente si siempre que escojamos un punto en $\overline{\mathbb{R}^2}$, si trazamos una recta vertical y una horizontal que pasen por ese punto, de los cuadrantes que se forman, o el segundo o el cuarto no se intersectan con el conjunto, dependiendo si el punto elegido está por “encima” o por “debajo” de él. Una cosa similar pasará si el conjunto es estrictamente decreciente, sólo que serán el primer o el tercer cuadrante los que no intersectarán con el conjunto.

Lema 1.6.3. Sean X y Y dos variables aleatorias continuas con función de distribución H . H es igual a la cota superior de Fréchet-Hoeffding si y sólo si para todo $(x, y) \in \overline{\mathbb{R}^2}$, o bien $\mathbb{P}[X > x, Y \leq y] = 0$; o bien $\mathbb{P}[X \leq x, Y > y] = 0$.

Demostración. Sean F y G las marginales de H . Entonces

$$\begin{aligned} F(x) &= \mathbb{P}[X \leq x] = \mathbb{P}[X \leq x, Y \leq y] + \mathbb{P}[X \leq x, Y > y] \\ &= H(x, y) + \mathbb{P}[X \leq x, Y > y], \end{aligned}$$

de la misma forma tenemos que

$$G(y) = H(x, y) + \mathbb{P}[X > x, Y \leq y].$$

Por lo tanto, $H(x, y) = \min(F(x), G(y))$ si y sólo si

$$\min(\mathbb{P}[X > x, Y \leq y], \mathbb{P}[X \leq x, Y > y]) = 0,$$

de donde concluimos el lema. \square

Lema 1.6.4. Sean X y Y dos variables aleatorias continuas con función de distribución H . H es igual a la cota inferior de Fréchet-Hoeffding si y sólo si para todo $(x, y) \in \overline{\mathbb{R}^2}$, o bien $\mathbb{P}[X > x, Y > y] = 0$; o bien $\mathbb{P}[X \leq x, Y \leq y] = 0$.

Demostración. Sean F y G las marginales de H . Entonces

$$\begin{aligned} H(x, y) &= \mathbb{P}[X \leq x, Y \leq y] = F(x) - \mathbb{P}[X \leq x, Y > y] \\ &= F(x) - (\mathbb{P}[Y > y] - \mathbb{P}[X > x, Y > y]) \\ &= F(x) + G(y) - 1 + \mathbb{P}[X > x, Y > y], \end{aligned}$$

y

$$H(x, y) = 0 + \mathbb{P}[X \leq x, Y \leq y].$$

Por lo tanto, $H(x, y) = \max(F(x) + G(y) - 1, 0)$ si y sólo si

$$\min(\mathbb{P}[X > x, Y > y], \mathbb{P}[X \leq x, Y \leq y]) = 0,$$

de donde concluimos el lema. \square

Lo que los lemas anteriores nos dicen, es justamente lo que habíamos dicho de un conjunto estrictamente creciente; la distribución conjunta es la cota superior si o el segundo o el cuarto cuadrante, generados por el punto, tienen probabilidad cero, es decir, no intersectan con el soporte. Caso similar para la cota inferior. Ahora estamos en posición de establecer los teoremas que caracterizan a las variables aleatorias cuya cópula es una de las cotas de Fréchet-Hoeffding.

Teorema 1.6.5. *Sean X y Y dos variables aleatorias continuas con función de distribución conjunta H . Entonces, H es la cota superior de Fréchet-Hoeffding si y sólo si el soporte de H es un subconjunto estrictamente creciente de $\overline{\mathbb{R}^2}$.*

Demostración. Sea S el soporte de H y sea (x, y) un punto cualquiera de $\overline{\mathbb{R}^2}$. El inciso 1 del Lema 1.6.1 se cumple si y sólo si $\{(u, v) | u \leq x \text{ y } v > y\} \cap S = \emptyset$; lo que ocurre si y sólo si $\mathbb{P}[X \leq x, Y > y] = 0$. De la misma manera, el inciso 2 del Lema 1.6.1 se cumple si y sólo si $\{(u, v) | u > x \text{ y } v \leq y\} \cap S = \emptyset$; lo que ocurre si y sólo si $\mathbb{P}[X > x, Y \leq y] = 0$. La prueba del teorema se concluye directamente de los Lemas 1.6.1 y 1.6.3 \square

Teorema 1.6.6. *Sean X y Y dos variables aleatorias continuas con función de distribución conjunta H . Entonces, H es la cota inferior de Fréchet-Hoeffding si y sólo si el soporte de H es un subconjunto estrictamente decreciente de $\overline{\mathbb{R}^2}$.*

Demostración. Sea S el soporte de H y sea (x, y) un punto cualquiera de $\overline{\mathbf{R}}^2$. El inciso 1 del Lema 1.6.2 se cumple si y sólo si $\{(u, v) | u \leq x \text{ y } v \leq y\} \cap S = \emptyset$; lo que ocurre si y sólo si $\mathbb{P}[X \leq x, Y \leq y] = 0$. De la misma manera, el inciso 2 del Lema 1.6.2 se cumple si y sólo si $\{(u, v) | u > x \text{ y } v > y\} \cap S = \emptyset$; lo que ocurre si y sólo si $\mathbb{P}[X > x, Y > y] = 0$. La prueba del teorema se concluye directamente de los Lemas 1.6.2 y 1.6.4 \square

Con los teoremas anteriores tenemos que una pareja (X, Y) tiene la cota superior de Fréchet-Hoeffding como cópula si el soporte de su función de distribución conjunta, H , es estrictamente creciente. El soporte, en ese caso, lo podemos expresar como $\{(x, y) | x = \alpha(y)\}$, donde α es una función estrictamente creciente, no necesariamente continua. Pero si el soporte está sobre una curva de la forma $y = \alpha(x)$, qué podemos decir de la relación que tienen las variables aleatorias. Veamos que

$$\mathbb{P}[Y = \alpha(X)] = \iint_{\{(x, y) | x = \alpha(y)\}} dH(x, y) = 1,$$

ya que la integral sobre el soporte con respecto a la distribución es 1. Por lo tanto $Y \stackrel{c.s.}{=} \alpha(X)$. De aquí, concluimos que una variable aleatoria continua es función estrictamente creciente de otra, si y sólo si la cópula que las relaciona es M . La misma discusión se tiene para el caso de la cópula W , solamente que la función es estrictamente decreciente.

Por el Teorema de Sklar, sabemos que si X y Y son dos variables aleatorias continuas con función de distribución conjunta H y distribuciones marginales F y G , entonces existe una cópula C que cumple que $H(x, y) = C(F(x), G(y))$. De esto podemos ver que en el caso de variables aleatorias continuas, la distribución conjunta consta de tres elementos: la cópula, la distribución de X y la distribución de Y . Dado que las distribuciones de X y de Y determinan el comportamiento marginal de cada una de las variables, no debemos entonces pensar que el comportamiento conjunto está en lo que resta de la ecuación, es decir, la cópula. Los tres elementos son indispensables para formar el comportamiento conjunto; la pura cópula no basta para decir cómo se relacionan las variables aleatorias. Por ejemplo, si tenemos X y Z variables aleatorias independientes y definimos $Y = X + Z$, ¿existirá una cópula que resuma la relación que hay entre X y Y , sin importar cómo es el comportamiento marginal? Es de esperarse que no sea así. Veamos el siguiente ejemplo.

Ejemplo 1.6.1. Sean $X \sim N(0, 1)$, $Z \sim N(\mu, \sigma^2)$, independientes y definamos $Y = X + Z$. Intuitivamente conforme la varianza de Z aumente las variables serán cada vez menos dependientes, lo contrario sucederá si la varianza disminuye. La función de densidad conjunta de X y Z está dada por

$$f(x, z) = \frac{1}{2\pi\sigma} \exp \left\{ -\frac{1}{2} \left(x^2 + \frac{(z - \mu)^2}{\sigma^2} \right) \right\}.$$

Si hacemos la transformación $W = X$ y $Y = X + Z$, dado que el jacobiano es 1, la densidad conjunta de W y Y es

$$g(w, y) = \frac{1}{2\pi\sigma} \exp \left\{ -\frac{1}{2} \left(w^2 + \frac{(y - w - \mu)^2}{\sigma^2} \right) \right\}.$$

Sabemos que $W \sim N(0, 1)$ y $Z \sim N(\mu, 1 + \sigma^2)$, así que reacomodando la densidad tenemos que

$$g(w, z) = \frac{1}{2\pi\sqrt{1 + \sigma^2}\sqrt{\frac{\sigma^2}{1 + \sigma^2}}} \exp \left\{ -\frac{1}{2\frac{\sigma^2}{1 + \sigma^2}} \left(w^2 + \frac{(z - \mu)^2}{1 + \sigma^2} - \frac{2}{\sqrt{1 + \sigma^2}} w \frac{z - \mu}{\sqrt{1 + \sigma^2}} \right) \right\}$$

Es decir, W y Z tienen una distribución conjunta normal bivariada con coeficiente de correlación $\rho = \frac{1}{\sqrt{1 + \sigma^2}}$. Entonces $\rho \xrightarrow{\sigma \rightarrow \infty} 0$, lo que implica que la cópula de la pareja (W, Y) converge a Π , mientras que $\rho \xrightarrow{\sigma \rightarrow 0} 1$, lo que implica que la cópula converge a M . Por lo tanto, la dependencia entre X y $X + Z$ no tiene una única cópula sino que esta relación está afectada también por el comportamiento marginal. \square

En resumen, la cópula es necesaria mas no suficiente. Para conocer el comportamiento conjunto basta conocer la función de distribución conjunta, o bien, la cópula y las funciones de distribución marginales.

Veamos otros lemas que serán de utilidad.

Lema 1.6.5. *Sea C una cópula. Si $C(t, t) = t$, para toda $t \in \mathbf{I}$, entonces $C = M$.*

Demostración. Consideremos el cuadrado unitario y dividámoslo a la mitad por la recta identidad. La cópula $M(u, v) = \min(u, v)$ en el sector que queda

por encima de esa recta vale u , mientras que en la parte inferior vale v . Por otro lado, como C es cópula, sabemos que $V_C(B) \geq 0$, para todo rectángulo $B \subseteq \mathbf{I}$. Si elegimos el rectángulo $B = [0, u] \times [u, v]$, con $v > u$, tenemos que

$$C(u, v) - C(u, u) - C(0, v) + C(0, u) = C(u, v) - u \geq 0,$$

de donde tenemos que $C(u, v) \geq u$. Además sabemos que $C(u, v) \leq C(u, 1) = u$, y combinando ambas desigualdades tenemos que $C(u, v) = u$.

De la misma manera podemos ver que $C(u, v) = v$ en el sector que queda por debajo de la recta identidad. Por lo tanto $C(u, v) = \min(u, v) = M(u, v)$ para todo $(u, v) \in \mathbf{I}^2$. \square

Lema 1.6.6. *Sea C una cópula. Si $C(t, 1-t) = 0$, para toda $t \in \mathbf{I}$, entonces $C = W$.*

Demostración. Consideremos el cuadrado unitario y dividámoslo a la mitad por la recta $y = 1 - x$. La cópula $W(u, v) = \max(u + v - 1, 0)$ en el sector que queda por encima de esa recta vale $u + v - 1$, mientras que en la parte inferior vale 0. Por otro lado, como C es cópula, sabemos que $V_C(B) \geq 0$, para todo rectángulo $B \subseteq \mathbf{I}$. Si elegimos el rectángulo $B = [0, u] \times [v, 1 - u]$, con $v < 1 - u$, tenemos que

$$C(u, 1 - u) - C(u, v) - C(0, 1 - u) + C(0, v) = -C(u, v) \geq 0,$$

de donde tenemos que $C(u, v) \leq 0$. Además sabemos que $C(u, v) \geq C(u, 0) = 0$, y combinando ambas desigualdades tenemos que $C(u, v) = 0$.

Ahora, si elegimos el cuadrado $B = [u, 1] \times [1 - u, v]$, con $1 - u < v$, tenemos que

$$C(1, v) - C(1, 1 - u) - C(u, v) + C(u, 1 - u) = v + u - 1 - C(u, v) \geq 0$$

de donde tenemos que $C(u, v) \leq u + v - 1$. Además si consideramos el rectángulo $B = [u, 1] \times [v, 1]$, tenemos que $C(u, v) \geq u + v - 1$ y combinando ambas desigualdades tenemos que $C(u, v) = u + v - 1$. Por lo tanto, $C(u, v) = \max(u + v - 1, 0) = W(u, v)$ para todo $(u, v) \in \mathbf{I}^2$. \square

Lo que los lemas anteriores nos dicen es, que si una cópula es igual a M sobre la diagonal principal, no le queda otra opción que ser M ; mientras que si es igual a W sobre la diagonal secundaria, no le queda otra opción que ser W .

Capítulo 2

Medidas de dependencia

2.1 Introducción

En este capítulo estudiaremos las medidas de dependencia y concordancia más conocidas, poniendo especial atención a las interpretaciones, en ocasiones erróneas, que algunos usuarios de la estadística aplican en el análisis de distintos problemas. Para cada una de las medidas presentaremos ejemplos interesantes en los que se muestran las posibles consecuencias de estas interpretaciones.

En la Sección 2.2 analizaremos la correlación, que es una de las medidas más populares, y las falacias en las que se incurre al considerarla como una medida univocal de independencia. La Sección 2.3 corresponde a la concordancia y las principales medidas de ésta: la τ de Kendall y la ρ de Spearman. En cada caso veremos sus principales propiedades así como las desventajas que pueden tener. Las Secciones 2.4.1, 2.4.2 y 2.4.3 están dedicadas a las propiedades de dependencia que pueden tener las funciones de distribución conjuntas su con algunas medidas de dependencia y concordancia. En la Sección 2.5 estudiaremos algunas medidas de dependencia y analizaremos las diferencias principales con las medidas de concordancia. Por último, en la Sección 2.6, estudiaremos un tipo particular de cópulas, lo reordenamientos de M , y veremos qué nos dicen de las medidas de dependencia.

Para la Sección 2.2 se utilizó [9], [16] y [18]; en la Sección 2.3 se consultó [8], [9], [12], [14] y [15]; las Secciones 2.4.1, 2.4.2 y 2.4.3 se hicieron a partir de [15]; la Sección 2.5 se trabajó a partir de [10], [15] y [17]; la Sección 2.6 se obtuvo de [15].

2.2 Correlación.

El mundo de la estadística estuvo dominado durante mucho tiempo por la distribución Gaussiana: un posible explicación es por la teoría asintótica a la distribución Normal, dada por el Teorema Central de Límite. Por ejemplo, si consideramos un vector aleatorio (X, Y) con distribución normal bivariada, su función de densidad tiene la siguiente forma

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 - 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right] \right\} \quad (2.1)$$

donde a ρ , lo podemos considerar como un parámetro de “independencia”, ya que si $\rho = 0$, la función de densidad conjunta representa un modelo de dos variables aleatorias normales independientes. Podemos pensar además, que ese parámetro ρ “mide” de alguna manera la dependencia, en el sentido de que los parámetros μ_X , σ_X^2 , μ_Y y σ_Y^2 son parámetros que provienen del comportamiento marginal, lo que hace que el único parámetro que corresponde a la función de densidad conjunta, y por lo tanto, a la manera de relacionarse, sea ρ . Más adelante, veremos que la correlación, en el caso normal bivariado está dada por ρ . Lo anterior y lo dicho sobre la importancia de las distribuciones Gaussianas, deben de ser la razones por las cuales la correlación se considera, o se consideraba, tan importante para el conocimiento de la dependencia de dos variables aleatorias.

Si bien es cierto que la dependencia en este caso se sintetiza en ρ , también es cierto que ρ no es otra cosa que un parámetro de la distribución, y que la dependencia realmente está en la estructura de la función de densidad conjunta. El principal problema que tiene la correlación, no es la mucha o poca información que nos da, sino la mala interpretación que se hace de ella. A continuación trataremos de explicar cuándo es que la correlación funciona, qué información nos está dando e ilustrar con ejemplos las grandes equivocaciones en que se puede incurrir.

Empecemos definiendo qué es la correlación.

Definición 2.2.1. Sean X y Y dos variables aleatorias no degeneradas con

segundo momento finito. La correlación entre X y Y se define como:

$$\rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

donde $\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ y σ_X^2, σ_Y^2 denotan las varianzas de X y Y respectivamente.

Para empezar nuestro estudio de la correlación, primero analicemos los valores que ésta toma, para luego tratar de interpretar los significados de éstos. Para ello veamos el siguiente resultado bastante familiar.

Teorema 2.2.1 (Desigualdad de Cauchy-Schwarz). Sean X y Y dos variables aleatorias con segundo momento finito, entonces

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2] \mathbb{E}[Y^2].$$

La igualdad se da si y sólo si existe a tal que $\mathbb{P}[Y = aX] = 1$.

Demostración. Consideremos la variable aleatoria $(tX + Y)^2$. Como dicha variable aleatoria es no negativa, tenemos que

$$0 \leq \mathbb{E}[(tX + Y)^2] = \mathbb{E}[X^2]t^2 + 2\mathbb{E}[XY]t + \mathbb{E}[Y^2].$$

Esto, como polinomio en t , sucede si y sólo si el discriminante es no positivo, es decir

$$(2\mathbb{E}[XY])^2 \leq 4\mathbb{E}[X^2] \mathbb{E}[Y^2],$$

de donde se concluye la primer parte del teorema.

Para la igualdad veamos que si $Y \stackrel{\text{c.s.}}{=} aX$ tenemos que

$$\begin{aligned} (\mathbb{E}[XY])^2 &= (\mathbb{E}[aX^2])^2 = a^2 (\mathbb{E}[X^2])^2 \\ &= \mathbb{E}[X^2] \mathbb{E}[(aX)^2] = \mathbb{E}[X^2] \mathbb{E}[Y^2]. \end{aligned}$$

Para la otra implicación, si $(\mathbb{E}[XY])^2 = \mathbb{E}[X^2] \mathbb{E}[Y^2]$, veamos que

$$\mathbb{E}[(tX + Y)^2] = \mathbb{E}[X^2]t^2 + 2\mathbb{E}[XY]t + \mathbb{E}[Y^2] = 0,$$

si y sólo si

$$\begin{aligned} t &= \frac{-\mathbb{E}[XY] \pm \sqrt{(\mathbb{E}[XY])^2 - \mathbb{E}[X^2] \mathbb{E}[Y^2]}}{\mathbb{E}[X^2]} \\ &= -\frac{\mathbb{E}[XY]}{\mathbb{E}[X^2]}. \end{aligned}$$

Por lo tanto, $\mathbb{E} \left[\left(Y - \frac{\mathbb{E}[XY]}{\mathbb{E}[X^2]} X \right)^2 \right] = 0$, lo que implica que $Y \stackrel{\text{c.s.}}{=} \frac{\mathbb{E}[XY]}{\mathbb{E}[X^2]} X$, lo que completa la prueba. \square

Consideremos dos variables aleatorias X y Y no degeneradas con segundo momento finito, entonces

$$\rho(X, Y) = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2] \mathbb{E}[(Y - \mathbb{E}[Y])^2]}} = \rho(X - \mathbb{E}[X], Y - \mathbb{E}[Y]),$$

y por la desigualdad de Cauchy-Schwarz, aplicándola a las variables $X - \mathbb{E}[X]$ y $Y - \mathbb{E}[Y]$, tenemos que

$$\rho(X, Y)^2 = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]^2}{\mathbb{E}[(X - \mathbb{E}[X])^2] \mathbb{E}[(Y - \mathbb{E}[Y])^2]} \leq 1.$$

Entonces, $\rho \in [-1, 1]$. Una primera duda es, ¿qué pasa en los casos extremos? Para responderla utilicemos la segunda parte del teorema 2.2.1 y lo visto en su demostración.

$$\rho(X, Y)^2 = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]^2}{\mathbb{E}[(X - \mathbb{E}[X])^2] \mathbb{E}[(Y - \mathbb{E}[Y])^2]} = 1,$$

si y sólo si

$$Y - \mathbb{E}[Y] \stackrel{\text{c.s.}}{=} \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\mathbb{E}[(X - \mathbb{E}[X])^2]} (X - \mathbb{E}[X])$$

Entonces, si $\rho(X, Y) = \pm 1$, se tiene que $Y \stackrel{\text{c.s.}}{=} aX + b$ con $a = \frac{\text{Cov}[X, Y]}{\sigma_X^2}$ y $b = \mathbb{E}[Y] - a\mathbb{E}[X]$, es decir, tenemos toda la información necesaria acerca de la dependencia entre X y Y . Los problemas ocurren entonces cuando $\rho \in (-1, 1)$ como veremos más adelante.

Otra pregunta común es, ¿qué podemos decir si la correlación es cero acerca de la dependencia? Veamos el siguiente teorema.

Teorema 2.2.2. Sean X y Y dos variables aleatorias independientes con esperanza finita, entonces $\text{Cov}[X, Y] = 0$.

Demostración. Como X y Y son independientes, entonces por el Teorema A.0.3, $\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]$, de donde se sigue el resultado. \square

El problema es que el recíproco no es cierto en general y ése es uno de los errores más comunes que se comete al considerar la correlación. Primero, veamos dos casos particulares donde obtenemos un si y sólo si, entre la independencia y la correlación cero.

Ejemplo 2.2.1. Sean X y Y dos variables aleatorias conjuntamente normales con parámetros μ_X , σ_X^2 , μ_Y , σ_Y^2 y ρ . La correlación de X y Y es

$$\rho(X, Y) = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} = \mathbb{E} \left[\left(\frac{X - \mu_X}{\sigma_X} \right) \left(\frac{Y - \mu_Y}{\sigma_Y} \right) \right], \quad (2.2)$$

La densidad conjunta de X y Y está dada por la ecuación (2.1), y por la ecuación (2.2), tenemos entonces que

$$\begin{aligned} \rho(X, Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{x - \mu_X}{\sigma_X} \right) \left(\frac{y - \mu_Y}{\sigma_Y} \right) \frac{1}{2\pi\sqrt{1 - \rho^2}} \\ &\quad e^{-\frac{1}{2(1 - \rho^2)} \left[\left(\frac{x - \mu_X}{\sigma_X} \right)^2 + \left(\frac{y - \mu_Y}{\sigma_Y} \right)^2 - 2\rho \frac{(x - \mu_X)(y - \mu_Y)}{\sigma_X \sigma_Y} \right]} dx dy. \end{aligned}$$

Haciendo la transformación $w = (x - \mu_X)/\sigma_X$, $z = (y - \mu_Y)/\sigma_Y$, tenemos que

$$\begin{aligned} \rho(X, Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{wz}{2\pi\sqrt{1 - \rho^2}} e^{-\frac{w^2 + z^2 - 2\rho wz}{2(1 - \rho^2)}} dw dz \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{wz}{2\pi\sqrt{1 - \rho^2}} e^{-\frac{w^2 - 2\rho wz + \rho^2 z^2 + (1 - \rho^2)z^2}{2(1 - \rho^2)}} dw dz \\ &= \int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \int_{-\infty}^{\infty} \frac{w}{\sqrt{2\pi(1 - \rho^2)}} e^{-\frac{(w - \rho z)^2}{2}} dw dz \\ &= \int_{-\infty}^{\infty} \frac{\rho z^2}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \rho. \end{aligned}$$

Por lo tanto, la correlación entre X y Y es el parámetro ρ . Si consideramos, en la ecuación (2.1), que $\rho = 0$, tenemos que dicha función corresponde a la densidad conjunta de una pareja normal bivariada con entradas independientes, entonces X y Y son independientes. En el caso normal bivariado tenemos que X y Y son independientes si y sólo si la correlación es cero. \square

El ejemplo anterior, es el más conocido en donde ocurre esa doble implicación y seguramente la razón por la que existe la confusión acerca de la interpretación que se le debe dar al valor cero de la correlación. No es el único caso en el que esto ocurre. Un ejemplo menos familiar es el siguiente.

Ejemplo 2.2.2. Sean X y Y dos variables aleatorias Bernoulli de parámetros p_1 y p_2 respectivamente. Entonces

$$\text{Cov}[X, Y] = \mathbb{P}[X = 1, Y = 1] - p_1 p_2. \quad (2.3)$$

Si $\text{Cov}[X, Y] = 0$ entonces $\mathbb{P}[X = 1, Y = 1] = p_1 p_2$. Aún más, X y Y son independientes, ya que

$$\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x] \mathbb{P}[Y = y],$$

para toda pareja (x, y) . Por ejemplo, para la pareja $(1, 1)$, lo tenemos por la ecuación (2.3); para la pareja $(0, 1)$ tenemos que

$$\mathbb{P}[X = 0, Y = 1] = \mathbb{P}[Y = 1] - \mathbb{P}[X = 1, Y = 1] = p_2(1 - p_1);$$

y de la misma manera se sigue para todos los otros casos. Por lo tanto, para las variables aleatorias Bernoulli tenemos que la covarianza, y por lo tanto, la correlación cero implica independencia entre las variables aleatorias.

Este resultado, no se da por el hecho de que las variables sean Bernoulli, sino por el hecho de que tomen sólo dos valores, es decir, lo podemos generalizar a variables W y Z , tales que $\mathbb{P}[W = a] = p_1 = 1 - \mathbb{P}[W = b]$, $\mathbb{P}[Z = c] = p_2 = 1 - \mathbb{P}[Z = d]$, donde p_1 y p_2 sean las mismas consideradas antes. Ahora, las variables W y Z las podemos ver en términos de X y Y considerando la transformación $W = (b - a)X + a$ y $Z = (d - c)Y + c$. Entonces

$$\begin{aligned} \text{Cov}(W, Z) &= \mathbb{E}[WZ] - \mathbb{E}[W] \mathbb{E}[Z] \\ &= \mathbb{E}[(b - a)X + a][(d - c)Y + c] - \mathbb{E}[(b - a)X + a] \mathbb{E}[(d - c)Y + c] \\ &= (b - a)(d - c) \mathbb{E}[XY] + a(d - c) \mathbb{E}[Y] + c(b - a) \mathbb{E}[X] + ac \\ &\quad - ((b - a)(d - c) \mathbb{E}[X] \mathbb{E}[Y] + a(d - c) \mathbb{E}[Y] + c(b - a) \mathbb{E}[X] + ac) \\ &= (b - a)(d - c) \text{Cov}[X, Y]. \end{aligned}$$

Por lo tanto, para las parejas de variables aleatorias que toman sólo dos valores tenemos que son independientes si y sólo si la correlación es cero. \square

Sin embargo, en general no es cierto que la correlación cero implique la independencia entre las variables aleatorias, para ver que esto es cierto, veamos algunos contraejemplos.

Ejemplo 2.2.3. Sea $X \sim N(0, 1)$ y $Y = X^2$, como la normal estándar tiene momentos impares cero, entonces $\text{Cov}[X, Y] = \mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2] = 0$ y por lo tanto $\rho(X, Y) = 0$, pero claramente X y Y no son independientes.

Éste es un caso particular de un ejemplo más general, donde basta que X sea simétrica, es decir, que X y $-X$ se distribuyan igual y que tengan cuarto momento finito para que $\rho(X, Y) = 0$. El cuarto momento finito se pide para que la varianza de $Y = X^2$ exista. \square

El ejemplo anterior es bastante claro, tenemos dos variables aleatorias que son función la una de la otra, y sin embargo tienen correlación cero. En el ejemplo 2.2.1 vimos que si dos variables aleatorias X, Y tienen distribución conjunta normal, tenemos que X es independiente de Y si y sólo si $\rho(X, Y) = 0$.

Pero si X y Y se distribuyen normales de manera marginal únicamente, ¿pasará lo mismo? La respuesta es no, para que eso suceda es de gran importancia que lo hagan conjuntamente.

Ejemplo 2.2.4. Sea $X \sim N(0, 1)$ e I una variable aleatoria, independiente de X , que cumple que $\mathbb{P}[I = 1] = \frac{1}{2} = \mathbb{P}[I = -1]$ y sea $Y = IX$. Es claro que las variables aleatorias X y Y no son independientes. Veamos cómo se distribuye Y .

$$\begin{aligned}\mathbb{P}[Y \leq y] &= \mathbb{P}[XI \leq y] \\ &= \frac{1}{2} (\mathbb{P}[X \leq y|I = 1] + \mathbb{P}[X \geq -y|I = -1]) \\ &= \mathbb{P}[X \leq y].\end{aligned}$$

Por lo tanto $Y \sim N(0, 1)$, lo que implica que tenemos una pareja de variables aleatorias normales. Veamos qué pasa con la correlación.

$$\text{Cov}[X, Y] = \mathbb{E}[XY] = \mathbb{E}[X^2I] = \mathbb{E}[X^2]\mathbb{E}[I] = 0$$

de donde concluimos que la correlación es cero. \square

Veamos un ejemplo más

Ejemplo 2.2.5. La función de distribución conjunta F de una pareja de variables aleatorias $N(0,1)$ independientes, la podemos escribir como

$$F(x, y) = \Phi(x)\Phi(y)$$

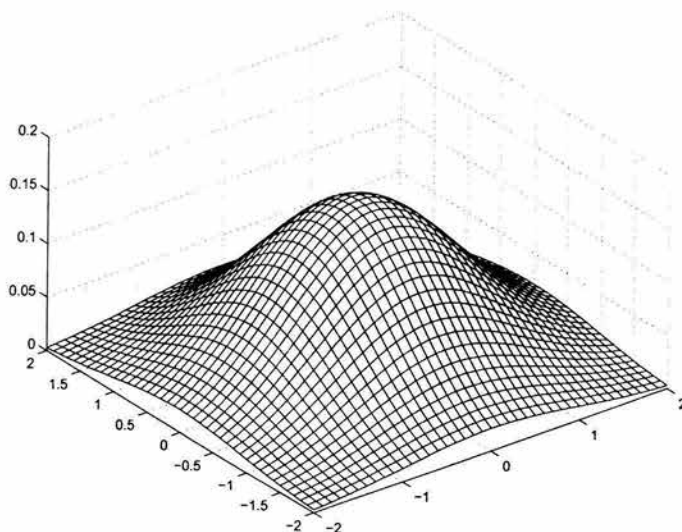


Figura 2.1: Densidad de una pareja de normales estándar independientes.

donde Φ es la función de distribución de una variable aleatoria $N(0, 1)$.

Por el Teorema de Sklar (véase Secciones 1.4 y 1.5), si consideramos cualquier cópula $C \neq \Pi$, y la componemos con la función Φ en cada una de sus entradas, obtenemos una distribución bivariada con marginales normales estándar que no es bivariada normal.

Consideremos el siguiente método de construir una cópula. Sean $f, g : [0, 1] \rightarrow \mathbb{R}$ con $\int_0^1 f(x)dx = \int_0^1 g(y)dy = 0$ y $f(x)g(y) \geq -1$ para todo $x, y \in [0, 1]$. Entonces $h(x, y) = (1 + f(x)g(y))\mathbb{I}_{[0,1]}(x)\mathbb{I}_{[0,1]}(y)$ es una función de densidad bivariada ya que integra uno y es positiva, además, tiene marginales $U(0, 1)$. Por lo tanto

$$C(x, y) = xy + \left(\int_0^x f(u)du \right) \left(\int_0^y g(v)dv \right) \quad (2.4)$$

es una cópula. Ahora, podemos construir una cópula de dicho tipo eligiendo las siguientes funciones

$$f(x) = \mathbb{I}_{(\gamma, 1-\gamma)}(x) + \frac{2\gamma - 1}{2\gamma} \mathbb{I}_{(\gamma, 1-\gamma)^c}(x)$$

$$g(y) = - \left(\mathbb{I}_{(\gamma, 1-\gamma)}(y) + \frac{2\gamma - 1}{2\gamma} \mathbb{I}_{(\gamma, 1-\gamma)^c}(y) \right)$$

donde $\frac{1}{4} \leq \gamma \leq \frac{1}{2}$. Consideremos la función de distribución conjunta $H_\gamma(x, y) = C_\gamma(\Phi(x), \Phi(y))$, donde C_γ , es una cópula como la de la ecuación (2.4). La función de densidad conjunta de dicha distribución, está dada por

$$\begin{aligned} h_\gamma(x, y) &= \frac{\partial^2 H_\gamma(x, y)}{\partial x \partial y} \\ &= \phi(x)\phi(y) + f(\Phi(x))\phi(x)g(\Phi(y))\phi(y) \\ &= \phi(x)\phi(y) [1 - f(\Phi(x))f(\Phi(y))], \end{aligned}$$

donde ϕ es la densidad de una variable normal estándar. La función h_γ se anula en el cuadrado $[\gamma, 1 - \gamma] \times [\gamma, 1 - \gamma]$, lo que hace claro que $H_\gamma(x, y) = C_\gamma(\Phi(x), \Phi(y))$ es totalmente distinta a la función normal bivariada.

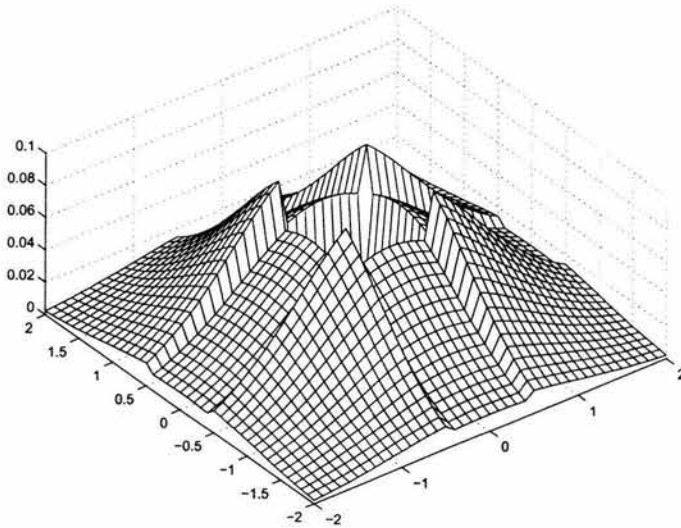


Figura 2.2: Densidad h_γ , con $\gamma = \frac{1}{3}$.

Viendo ambas funciones de densidad conjuntas (figura 2.1 y figura 2.2), donde una tiene el grueso de su masa, la otra tiene un gran hoyo. Veamos qué pasa con la correlación de una pareja (X, Y) que tuviera a H_γ como distribución. Por el Lema de Hoeffding (Lema 2.2.1), que veremos más adelante,

sabemos que

$$\begin{aligned} \text{Cov}[X, Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_{\gamma}(\Phi(x), \Phi(y)) - \Phi(x)\Phi(y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\int_0^{\Phi(x)} f(u) du \right) \left(\int_0^{\Phi(y)} g(v) dv \right) dx dy \\ &= \int_{-\infty}^{\infty} \int_0^{\Phi(x)} f(u) du dx \int_{-\infty}^{\infty} \int_0^{\Phi(y)} g(v) dv dy, \end{aligned}$$

ahora

$$\int_{-\infty}^{\infty} \int_0^{\Phi(x)} f(u) du dx = \int_{-\infty}^{\infty} F(\Phi(x)) dx, \quad (2.5)$$

donde

$$F(s) = \frac{2\gamma - 1}{2\gamma} s \mathbb{I}_{[0, \gamma]}(s) + \left(s - \frac{1}{2} \right) \mathbb{I}_{[\gamma, 1-\gamma]}(s) + \frac{2\gamma - 1}{2\gamma} (s - 1) \mathbb{I}_{(1-\gamma, 1]}(s).$$

La integral de la ecuación (2.5) es cero; para ver que es cierto, veamos que $\int_{-\infty}^0 = -\int_0^{\infty}$ y recordemos que $\Phi(x) = 1 - \Phi(-x)$.

$$\begin{aligned} \int_0^{\infty} F(\Phi(x)) dx &= \int_0^{\infty} \left[\Phi(x) - \frac{1}{2} \right] \mathbb{I}_{[1/2, 1-\gamma]}(\Phi(x)) \\ &\quad + \frac{2\gamma - 1}{2\gamma} (\Phi(x) - 1) \mathbb{I}_{(1-\gamma, 1]}(\Phi(x)) dx \\ &= \int_0^{\infty} \left[\frac{1}{2} - \Phi(-x) \right] \mathbb{I}_{[0, \Phi^{-1}(1-\gamma)]}(x) \\ &\quad + \frac{2\gamma - 1}{2\gamma} (-\Phi(-x)) \mathbb{I}_{(\Phi^{-1}(1-\gamma), \infty)}(x) dx, \end{aligned}$$

y haciendo el cambio de variable $y = -x$, tenemos

$$\begin{aligned} &= \int_{-\infty}^0 \left[\frac{1}{2} - \Phi(y) \right] \mathbb{I}_{[0, \Phi^{-1}(1-\gamma)]}(-y) + \frac{2\gamma - 1}{2\gamma} (-\Phi(y)) \mathbb{I}_{(\Phi^{-1}(1-\gamma), \infty)}(-y) dy \\ &= - \int_{-\infty}^0 \left[\Phi(y) - \frac{1}{2} \right] \mathbb{I}_{[\Phi^{-1}(\gamma), 0]}(y) + \frac{2\gamma - 1}{2\gamma} (\Phi(y)) \mathbb{I}_{(-\infty, \Phi^{-1}(\gamma))}(y) dy. \end{aligned}$$

De donde tenemos que la covarianza es cero y la correlación, entonces, es cero también. Por lo tanto, hemos construido una familia no numerable H_γ de funciones de distribución con marginales normales y correlación cero, distintas de la que representa a la independencia. \square

Con los ejemplos anteriores es claro que la correlación igual a cero no implica que las variables sean independientes. En resumen, vimos casos en que la correlación cero nos decía algo acerca de la independencia y vimos casos en los que no. Por lo tanto, sólo hay que tener conciencia que la correlación cero no nos da ninguna información certera excepto, claro está, en los casos vistos en los ejemplos 2.2.1 y 2.2.2. Lo que sí es cierto, es que si $\rho \neq 0$ entonces la pareja no será independiente.

Vimos antes que si la correlación es 1 ó -1, tenemos una dependencia lineal entre las variables aleatorias, la pregunta es, si tenemos dos distribuciones univariadas, ¿será siempre posible encontrar una distribución conjunta con esas distribuciones marginales que represente un modelo de dependencia lineal? Más aún, para cualesquiera dos distribuciones univariadas, ¿será siempre posible obtener todos los valores entre -1 y 1 para ρ considerando distintas distribuciones conjuntas? La respuesta es no y ésta es la segunda de las creencias erróneas que se tiene de la correlación.

Ejemplo 2.2.6. Sean F y G dos funciones de distribución univariadas tales que, $F(x) = G(x) = 0$ para toda $x < 0$. Sea H una función de distribución conjunta con F y G sus distribuciones marginales. Supongamos que $\rho = -1$. En este caso, por el Teorema 2.2.1, si tenemos que X y Y son las variables asociadas a esta distribución, entonces se cumple que $Y \stackrel{c.s.}{=} aX + b$ con $a < 0$. Pero para todo $y < 0$ tenemos

$$\begin{aligned} G(y) &= \mathbb{P}[Y \leq y] = \mathbb{P}\left[X \geq \frac{y-b}{a}\right] \\ &\geq 1 - F\left(\frac{y-b}{a}\right) > 0, \end{aligned}$$

que contradice que $G(y) = 0$. Es bastante natural lo que pasa; si X y Y se relacionan por una recta de pendiente negativa, no es posible que el rango de ambas sea $[0, \infty)$, de ahí que ocurra una contradicción. \square

Entonces qué pasa con los valores que toma la correlación.

Teorema 2.2.3. *Sean F y G dos funciones de distribución univariadas no degeneradas con segundo momento finito. Entonces:*

1. El conjunto de posibles correlaciones es un intervalo cerrado $[\rho_{\min}, \rho_{\max}]$ que cumple que $\rho_{\min} < 0 < \rho_{\max}$.
2. La correlación ρ alcanza el valor ρ_{\min} si y sólo si F y G se relacionan mediante la cópula W ; ρ alcanza el valor ρ_{\max} si y sólo si F y G se relacionan mediante la cópula M .
3. Sean X y Y variables aleatorias asociadas a F y G . $\rho_{\min} = -1$ si y sólo si existe $a > 0$ y $b \in \mathbb{R}$ de tal manera que Y y $aX + b$ tengan la misma distribución; $\rho_{\max} = 1$ si y sólo si existe $a < 0$ y $b \in \mathbb{R}$ de tal manera que X y $aY + b$ tengan la misma distribución.

Para probar el teorema anterior, necesitamos primero probar un lema.

Lema 2.2.1 (Lema de Hoeffding). Sean X y Y dos variables aleatorias con función de distribución conjunta H , y distribuciones marginales F y G , respectivamente. Si $\mathbb{E}[|XY|]$, $\mathbb{E}[|X|]$ y $\mathbb{E}[|Y|]$ son finitas entonces

$$\text{Cov}[X, Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H(x, y) - F(x)G(y)) dx dy.$$

Demostración. Sean (X_1, Y_1) y (X_2, Y_2) dos vectores aleatorios independientes e idénticamente distribuidos con H , F y G las respectivas funciones de distribución. Entonces

$$\begin{aligned} & 2 \left[\mathbb{E}[X_1 Y_1] - \mathbb{E}[X_1] \mathbb{E}[Y_1] \right] \\ &= \mathbb{E}[X_1 Y_1] - \mathbb{E}[X_1] \mathbb{E}[Y_2] + \mathbb{E}[X_2 Y_2] - \mathbb{E}[X_2] \mathbb{E}[Y_1] \\ &= \mathbb{E}[X_1 Y_1 - X_1 Y_2 + X_2 Y_2 - X_2 Y_1] \\ &= \mathbb{E}[(X_1 - X_2)(Y_1 - Y_2)] = \mathbb{E} \left[\int_{X_2}^{X_1} du \int_{Y_2}^{Y_1} dv \right] \\ &= \mathbb{E} \left[\int_{-\infty}^{\infty} (\mathbb{I}_{(-\infty, u]}(X_1) - \mathbb{I}_{(-\infty, u]}(X_2)) du \int_{-\infty}^{\infty} (\mathbb{I}_{(-\infty, v]}(Y_1) - \mathbb{I}_{(-\infty, v]}(Y_2)) dv \right] \\ &= \mathbb{E} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\mathbb{I}_{(-\infty, u]}(X_1) \mathbb{I}_{(-\infty, v]}(Y_1) - \mathbb{I}_{(-\infty, u]}(X_1) \mathbb{I}_{(-\infty, v]}(Y_2) \right. \\ & \quad \left. - \mathbb{I}_{(-\infty, u]}(X_2) \mathbb{I}_{(-\infty, v]}(Y_1) + \mathbb{I}_{(-\infty, u]}(X_2) \mathbb{I}_{(-\infty, v]}(Y_2)) dudv \right]. \end{aligned}$$

Como $\mathbb{E}[|XY|]$, $\mathbb{E}[|X|]$ y $\mathbb{E}[|Y|]$ son finitas, la expresión anterior es absolutamente integrable y aplicando Fubini tenemos que la expresión es igual a

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 2\mathbb{E} [\mathbb{I}_{(-\infty, u]}(X_1) \mathbb{I}_{(-\infty, v]}(Y_1) - \mathbb{I}_{(-\infty, u]}(X_1) \mathbb{I}_{(-\infty, v]}(Y_2)] dudv \\ & = 2 \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H(u, v) - F(u)G(v)) dudv \right], \end{aligned}$$

lo que da la identidad deseada. \square

Ahora procedamos con la demostración del Teorema 2.2.3.

Demostración del Teorema 2.2.3. Sea H una función de distribución conjunta arbitraria. Por el Teorema 1.4.1 y el Teorema de Sklar (1.5.1), tenemos que

$$\max(F(x) + G(y) - 1, 0) \leq H(x, y) \leq \min(F(x), G(y)).$$

De la desigualdad anterior se sigue que

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\max(F(x) + G(y) - 1, 0) - F(x)G(y)) dx dy \\ & \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H(x, y) - F(x)G(y)) dx dy \quad (2.6) \\ & \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\min(F(x), G(y)) - F(x)G(y)) dx dy. \end{aligned}$$

Mediante el Lema 2.2.1 y la desigualdad (2.6), tenemos que la covarianza entre X y Y se minimiza si éstas se relacionan mediante la cópula W y se maximiza si lo hacen mediante la cópula M . También tenemos que

$$\max(F(x) + G(y) - 1, 0) \leq F(x)G(y) \leq \min(F(x), G(y)). \quad (2.7)$$

Entonces, combinando el lema 2.2.1, la desigualdad (2.7) y lo argumentado arriba, tenemos que $\rho_{\min} \leq 0$ y $\rho_{\max} \geq 0$. Veamos ahora que no pueden ser cero ninguna de las dos. Si $F(x) = \mathbb{I}_{[c, \infty)}(x)$, es decir, una función de distribución asociada a una variables aleatoria degenerada, se puede ver que

$$\max(\mathbb{I}_{[c, \infty)}(x) + G(y) - 1, 0) \equiv \mathbb{I}_{[c, \infty)}(x) G(y) \equiv \min(\mathbb{I}_{[c, \infty)}(x), G(y)),$$

ya que si $x < c$, tenemos que

$$\max(G(y) - 1, 0) = 0 = \min(0, G(y))$$

y si $x \geq c$, entonces

$$\max(G(y), 0) = G(y) = \min(1, G(y)),$$

lo que implica que $\rho_{\min} = \rho_{\max} = 0$. Estamos pidiendo que ninguna de las dos distribuciones sea degenerada. Para que así sea, tienen que existir $x_1, y_1 \in \mathbb{R}$ tales que $0 < F(x_1), G(y_1) < 1$. De no ser así, como las funciones de distribución son crecientes y continuas por la derecha, las funciones F y G tienen que ser de la forma $\mathbb{I}_{[a, \infty)}$ para algún $a \in \mathbb{R}$, lo que implica que son degeneradas. Por lo tanto existen tales puntos. Ahora, como las funciones de distribución son continuas por la derecha y crecientes, tenemos que para $\epsilon = 1 - F(x_1)$, existe un $\delta_1 > 0$, de tal manera que si $0 \leq x - x_1 < \delta_1$, entonces $0 \leq F(x) - F(x_1) < 1 - F(x_1)$. Es decir, existe un intervalo de tamaño $\delta_1 > 0$ para el cual $0 < F(x) < 1$. De la misma manera, para $G(y)$, tenemos que existe un intervalo de tamaño $\delta_2 > 0$ para el cual $0 < G(y) < 1$. Ahora, en esos dos intervalos tenemos que $F(x)G(y) < F(x)$ y que $F(x)G(y) < G(y)$ y por lo tanto $F(x)G(y) < \min(F(x)G(y))$. De aquí podemos ver que

$$\begin{aligned} \rho_{\min} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \min(F(x), G(y)) - F(x)G(y) dx dy \\ &\geq \iint_C \min(F(x), G(y)) - F(x)G(y) dx dy > 0, \end{aligned}$$

donde $C = \{x | 0 \leq x - x_1 < \delta_1\} \times \{y | 0 \leq y - y_1 < \delta_2\}$.

Dentro de la misma región C , tenemos que $(1 - F(x))(1 - G(y)) > 0$, lo que implica que $F(x)G(y) > F(x) + G(y) - 1$; además, $F(x)G(y) > 0$. Por lo tanto, $F(x)G(y) > \max(F(x) + G(y) - 1, 0)$, de donde podemos ver que

$$\begin{aligned} \rho_{\max} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \max(F(x) + G(y) - 1, 0) - F(x)G(y) dx dy \\ &\leq \iint_C \min(F(x), G(y)) - F(x)G(y) dx dy < 0. \end{aligned}$$

Por lo tanto, $\rho_{\min} < 0 < \rho_{\max}$.

Falta probar que todos los valores del intervalo $[\rho_{\min}, \rho_{\max}]$ pueden ser alcanzados. Para ello utilicemos el hecho de que cualquier combinación convexa

de cópulas es una cópula (véase ejemplo 1.4.3). Si $H(x, y) = \lambda W(F(x), G(y)) + (1 - \lambda)M(F(x), G(y))$ tenemos que

$$\begin{aligned} \text{Cov}[X, Y] &= \lambda \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(F(x), G(y)) - F(x)G(y) dx dy \\ &\quad + (1 - \lambda) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} M(F(x), G(y)) - F(x)G(y) dx dy \\ &= \lambda \rho_{\min} + (1 - \lambda) \rho_{\max}, \end{aligned}$$

de donde podemos construir distribuciones conjuntas de tal manera que $\rho \in [\rho_{\min}, \rho_{\max}]$. \square

Ejemplo 2.2.7. Sea F una distribución Lognormal(0, 1) y G una distribución Lognormal(0, σ^2), $\sigma > 0$. Calculemos ρ_{\min} y ρ_{\max} para estas distribuciones. Sabemos que si $Z \sim N(0, 1)$, entonces $e^{aZ} \sim \text{Lognormal}(0, a^2)$ con $a \neq 0$. Por los Teoremas 1.6.5 y 1.6.6, sabemos que si una variable aleatoria es función estrictamente monótona de otra, su función de distribución conjunta es: W compuesta con las marginales, en el caso decreciente; y M compuesta con las marginales, en el caso creciente. Por lo tanto, la pareja $(e^Z, e^{-\sigma Z})$ tiene como cópula a W , y sus distribuciones marginales son las dos Lognormales que queremos. Por el Teorema 2.2.3 alcanza ρ_{\min} . Por otro lado, la pareja $(e^Z, e^{\sigma Z})$ tiene como cópula a M , mismas marginales y alcanza a ρ_{\max} .

Para calcular utilizemos el hecho de que la función generadora de momentos de una variable aleatoria $N(0, 1)$ está dada por $\mathbb{E}[e^{tZ}] = e^{t^2/2}$. Por lo tanto

$$\begin{aligned} \rho_{\min} &= \rho(e^Z, e^{-\sigma Z}) \\ &= \frac{\mathbb{E}[e^{(1-\sigma)Z}] - \mathbb{E}[e^Z] \mathbb{E}[e^{-\sigma Z}]}{\sigma_X \sigma_Y} \\ &= \frac{e^{\frac{(1-\sigma)^2}{2}} - e^{\frac{1+\sigma^2}{2}}}{\sqrt{e^{\sigma+1}(e^{\sigma^2}-1)(e-1)}} \\ &= \frac{e^{-\sigma} - 1}{\sqrt{(e-1)(e^{\sigma^2}-1)}}. \end{aligned}$$

De la misma manera obtenemos que

$$\rho_{\max} = \frac{e^{\sigma} - 1}{\sqrt{(e-1)(e^{\sigma^2}-1)}}.$$

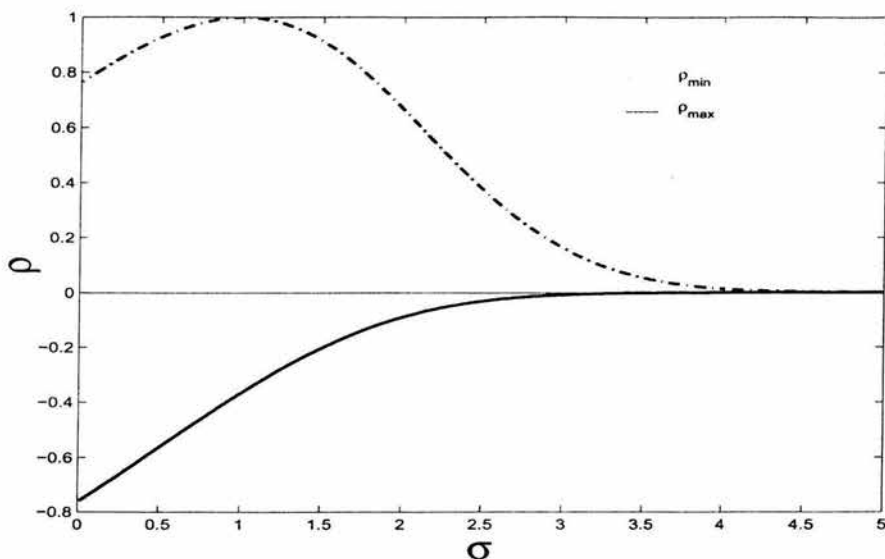


Figura 2.3: ρ_{\min} y ρ_{\max} para un pareja de Lognormales.

Podemos observar (véase figura 2.3) que $\rho_{\min}, \rho_{\max} \xrightarrow{\sigma \rightarrow \infty} 0$. Esto nos enseña que podemos tener vectores (X, Y) cuya correlación sea casi cero, y sin embargo éstos tienen la forma de dependencia más fuerte, creciente o decreciente. Este ejemplo nos muestra entonces que los valores pequeños de la correlación no deben ser interpretados como señal de una dependencia débil entre las variables aleatorias. \square

Hasta este momento con el análisis hecho podemos decir que la información cierta que nos da la correlación se encuentra en los casos extremos; que si conocemos para una pareja de variables aleatorias el rango de su correlación podríamos, en casos ideales de dependencia, con la correlación, afirmar que las variables son función monótona la una de la otra; que si no es el caso, podemos decir que se relacionan dependiendo de lo cercana que ésta se encuentre a los extremos, siempre entendiendo que como en el ejemplo pasado la cercanía puede ser algo muy ambiguo. Sin embargo, tenemos un avance importante al entender la información en el caso extremo que nos da la correlación.

Otro problema común de mala interpretación es sobre si basta conocer los comportamientos marginales y la correlación, para conocer el comportamiento conjunto. Esta duda, seguramente se debe también, a pensar que todo el mundo es Gaussiano. Creo que la respuesta ya ha sido casi contestada, aunque no de manera explícita. En el ejemplo 2.2.4 teníamos dos variables aleatorias $N(0, 1)$ con $\rho = 0$ cuya función conjunta aparentaba ser distinta a aquella de dos variables aleatorias $N(0, 1)$ independientes. Aun así, si tenemos dos variables aleatorias y su correlación, qué podemos decir sobre la función de distribución conjunta.

Ejemplo 2.2.8. Sea F_ρ la función de distribución normal bivariada estándar con coeficiente de correlación ρ . Cualquier combinación convexa $F = \lambda F_{\rho_1} + (1 - \lambda)F_{\rho_2}$ de distribuciones normales bivariadas, es también una función de distribución conjunta con marginales normales $(0, 1)$ y por el lema 2.2.1 con correlación $\lambda\rho_1 + (1 - \lambda)\rho_2$. Ahora, si fijamos $\rho \in (-1, 1)$, de tal manera que $\rho_1 < \rho < \rho_2$, entonces podemos encontrar $\lambda \in (0, 1)$, para que se cumpla que $\rho = \lambda\rho_1 + (1 - \lambda)\rho_2$. Si comparamos ambas funciones de densidad (figura 2.1 y figura 2.4), éstas se parecen mucho, así que, podemos pensar que el comportamiento conjunto de un vector con distribución conjunta F será muy parecido a uno con distribución conjunta F_ρ , lo cual no es cierto.

Para ver que las colas¹ bajo F son más pesadas que bajo F_ρ , supongamos que el vector (X, Y) tiene distribución F_ρ y el vector (U, V) tiene distribución F , y estudiemos cómo se comportan las colas de las variables $X + Y$ y $U + V$. Primero, veamos cómo se distribuye.

Si X y Y son dos variables aleatorias conjuntamente normales estándar con correlación ρ , el vector $X + Y$ se distribuye $N(0, 2(1 + \rho))$. Para ver que es cierto, hagamos la transformación $W = X + Y$ y $Z = X - Y$ ó $X = (W + Z)/2$ y $Y = (W - Z)/2$.

$$|J| = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = \frac{1}{2},$$

¹Si X es variable aleatoria con distribución G , la cola es $\mathbb{P}[X > x] = 1 - G(x)$.

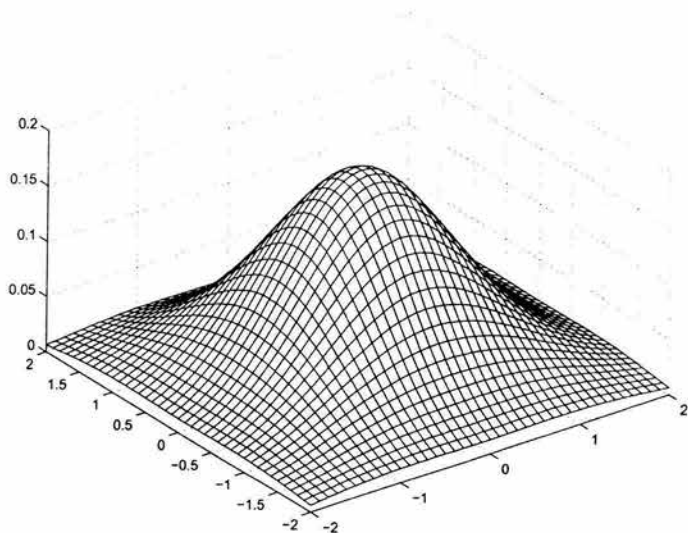


Figura 2.4: Densidad de F , con $\rho_1 = -\frac{1}{2}$ y $\rho_2 = \frac{1}{2}$.

entonces

$$\begin{aligned} f(w, z) &= \frac{1}{2} \frac{1}{2\pi\sqrt{1-\rho^2}} \\ &\quad e^{-\frac{1}{2(1-\rho^2)} \left[\left(\frac{w+z}{2}\right)^2 + \left(\frac{w-z}{2}\right)^2 - 2\rho \frac{(w+z)(w-z)}{4} \right]} \\ &= \frac{1}{4\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\frac{w^2(1-\rho) + z^2(1+\rho)}{2} \right]} \end{aligned}$$

y por lo tanto

$$f(w, z) = \frac{1}{\sqrt{4\pi(1+\rho)}} e^{-\frac{w^2}{4(1+\rho)}} \frac{1}{\sqrt{4\pi(1-\rho)}} e^{-\frac{z^2}{4(1-\rho)}}$$

de donde se obtiene que $W = X + Y \sim N(0, 2(1+\rho))$. Entonces

$$\mathbb{P}[X + Y > a] = 1 - \Phi\left(\frac{a}{2(1+\rho)}\right). \quad (2.8)$$

Para ver la distribución de $U + V$, basta notar que la densidad de F es $\lambda f_{\rho_1} + (1-\lambda)f_{\rho_2}$, donde f_{ρ} es la densidad conjunta de un vector normal con

correlación ρ . Por lo tanto,

$$\mathbb{P}[U + V > a] = \lambda \left[1 - \Phi \left(\frac{a}{2(1 + \rho_1)} \right) \right] + (1 - \lambda) \left[1 - \Phi \left(\frac{a}{2(1 + \rho_2)} \right) \right]. \quad (2.9)$$

Para analizar las colas, apliquemos la regla de l'Hôpital a

$$\frac{1 - \Phi(x)}{x^{-1}\phi(x)},$$

entonces

$$\lim_{x \rightarrow \infty} \frac{1 - \Phi(x)}{x^{-1}\phi(x)} = \lim_{x \rightarrow \infty} \frac{-\phi(x)}{-\phi(x) \left(1 + \frac{1}{x^2}\right)},$$

de donde obtenemos la razón de Mill

$$1 - \Phi(x) = \phi(x) \left(\frac{1}{x} + O \left(\frac{1}{x^2} \right) \right).$$

Usando esta razón y las ecuaciones (2.8) y (2.9), tenemos que

$$\begin{aligned}
\lim_{a \rightarrow \infty} \frac{\mathbb{P}[U + V > a]}{\mathbb{P}[X + Y > a]} &= \lim_{a \rightarrow \infty} \frac{\lambda \left[1 - \Phi \left(\frac{a}{2(1+\rho_1)} \right) \right] + (1 - \lambda) \left[1 - \Phi \left(\frac{a}{2(1+\rho_2)} \right) \right]}{1 - \Phi \left(\frac{a}{2(1+\rho)} \right)} \\
&= \lim_{a \rightarrow \infty} \frac{\lambda \phi \left(\frac{a}{2(1+\rho_1)} \right) \left(\frac{2(1+\rho_1)}{a} + O \left(\frac{1}{a^2} \right) \right)}{\phi \left(\frac{a}{2(1+\rho)} \right) \left(\frac{2(1+\rho)}{a} + O \left(\frac{1}{a^2} \right) \right)} \\
&\quad + \lim_{a \rightarrow \infty} \frac{(1 - \lambda) \phi \left(\frac{a}{2(1+\rho_2)} \right) \left(\frac{2(1+\rho_2)}{a} + O \left(\frac{1}{a^2} \right) \right)}{\phi \left(\frac{a}{2(1+\rho)} \right) \left(\frac{2(1+\rho)}{a} + O \left(\frac{1}{a^2} \right) \right)} \\
&= \lim_{a \rightarrow \infty} \frac{\lambda \phi \left(\frac{a}{2(1+\rho_1)} \right) O \left(\frac{1}{a} \right)}{\phi \left(\frac{a}{2(1+\rho)} \right) O \left(\frac{1}{a} \right)} \\
&\quad + \lim_{a \rightarrow \infty} \frac{(1 - \lambda) \phi \left(\frac{a}{2(1+\rho_2)} \right) O \left(\frac{1}{a} \right)}{\phi \left(\frac{a}{2(1+\rho)} \right) O \left(\frac{1}{a} \right)} \\
&= \lim_{a \rightarrow \infty} \lambda e^{\frac{a^2}{4} \left(\frac{1}{1+\rho} - \frac{1}{1+\rho_1} \right)} + \lim_{a \rightarrow \infty} (1 - \lambda) e^{\frac{a^2}{4} \left(\frac{1}{1+\rho} - \frac{1}{1+\rho_2} \right)} \\
&= \infty + 0 = \infty
\end{aligned}$$

Por lo tanto, aunque tengamos el mismo comportamiento marginal y la misma correlación, el comportamiento del vector bajo los dos distintos modelos cambia considerablemente, ya que las colas son más pesadas en el caso de la combinación convexa de normales. \square

Es clara ahora la respuesta a la última pregunta: para conocer el comportamiento conjunto necesitamos bastante más información que la correlación y las marginales.

Otro problema que tiene la correlación, es que no es invariante bajo transformaciones estrictamente crecientes $T: \mathbb{R} \rightarrow \mathbb{R}$. En general tenemos que

$$\rho(T(X), T(Y)) \neq \rho(X, Y).$$

Veamos el siguiente ejemplo.

Ejemplo 2.2.9. Consideremos un vector aleatorio (X, Y) con distribución normal estándar bivariada y coeficiente de correlación ρ . Si consideramos la transformación $T(x) = \Phi(x)$, con Φ la función de distribución normal estándar, la correlación del vector $(\Phi(X), \Phi(Y))$ es (véase ejemplo 2.3.7)

$$\rho(\Phi(X), \Phi(Y)) = \frac{6}{\pi} \arcsen\left(\frac{\rho}{2}\right)$$

claramente distinta de ρ . □

El Teorema 1.6.3 nos dice que al aplicar dos transformaciones α y β , estrictamente crecientes, a un vector (X, Y) , la cópula no cambia, es decir, $C_{XY} = C_{\alpha(X)\beta(Y)}$, sin embargo el comportamiento marginal sí puede cambiar. Esto muestra evidencias de que la correlación está considerando los tres elementos que forman la distribución conjunta: la cópula y las distribuciones marginales. Esto es más evidente aún si aplicamos una transformación al resultado del Lema 2.2.1

$$\begin{aligned} \text{Cov}[X, Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H(x, y) - F(x)G(y)) dx dy \\ &= \int_0^1 \int_0^1 (C(u, v) - uv) dF^{-1}(u) dG^{-1}(v) \end{aligned}$$

que hace evidente la dependencia con las funciones marginales F y G .

Una última desventaja, evidente desde su definición, es la necesidad de segundos momentos finitos, lo que claramente restringe su uso para algunas variables aleatorias como son la Pareto y la Cauchy, entre otras.

A pesar de todo esto tiene ventajas sobre la sencillez de su cálculo, la directa generalización a más variables y claro está su uso natural como medida de dependencia en el mundo normal multivariado. En conclusión, la información que la correlación nos da solamente será segura cuando estemos en los casos extremos de su rango, siempre y cuando lo conozcamos; los valores $\rho = \pm 1$ nos dejan el problema totalmente resuelto en cuanto a la relación de una variable con la otra; hay que tener cuidado cuando los valores no son estos extremos. Sin embargo, siempre nos puede dar una primera información sobre lo que pasa en cuanto a la dependencia entre dos variables aleatorias.

2.3 Concordancia

El concepto de concordancia y de medidas de concordancia es común en el tema de dependencia. La palabra concordancia, significa *conformidad de una cosa con otra*, y al referir esta palabra a la relación que dos variables aleatorias tienen, pues la conformidad será en el sentido de que si una aumenta, la otra aumente también; o que si una disminuye, la otra lo haga de igual manera. De una manera informal, decimos que una pareja de variables aleatorias es concordante, si los valores grandes de una corresponden a los valores grandes de la otra; o bien, si los valores pequeños de una corresponden a los valores pequeños de la otra. Si una variable es función de la otra, serán concordantes si esa función es creciente y serán discordantes si esa función es decreciente. Como su nombre lo dice, la concordancia dirá qué tanto está de acuerdo una variable aleatoria con respecto a la otra, en el sentido de, qué tanto influye el valor que tome una de ellas sobre el valor que tome la otra.

Las medidas de concordancia más comunes son la τ de Kendall y la ρ de Spearman, que estudiaremos a continuación. Estas medidas se utilizan en pruebas estadísticas para dar criterios de independencia entre las variables aleatorias (véase [8]). El problema está en que si sólo miden la concordancia, ¿podrá ser suficiente para decidir si dos variables aleatorias son independientes entre sí? En otras palabras, la concordancia no tiene porque ser la única forma en que dos variables dependan entre sí. En esta sección trataremos de contestar esta pregunta, así como aclarar lo que es una medida de concordancia.

Formalmente, si tenemos $(x_i, y_i), (x_j, y_j)$, dos parejas de reales, con $x_i \neq x_j$ y $y_i \neq y_j$: decimos que las parejas $(x_i, y_i), (x_j, y_j)$ son concordantes si $x_i < x_j$ y $y_i < y_j$, o si $x_i > x_j$ y $y_i > y_j$; decimos que las parejas $(x_i, y_i), (x_j, y_j)$ son discordantes si $x_i < x_j$ y $y_i > y_j$, o si $x_i > x_j$ y $y_i < y_j$. Notemos que las parejas serán concordantes si $x_i - x_j$ y $y_i - y_j$ tienen el mismo signo; serán discordantes si no. Entonces, las parejas son concordantes si $(x_i - x_j)(y_i - y_j) > 0$ y son discordantes si $(x_i - x_j)(y_i - y_j) < 0$.

Ahora, consideremos n parejas de reales, $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$. Si estamos interesados en saber cómo influye uno de los componentes de la pareja en el otro, una de las cosas que nos interesa saber es si las parejas son concordantes o discordantes entre sí. Si ahora, lo que queremos es resumir esa información en una medida numérica, una primera aproximación, es considerar las $\binom{n}{2}$ posibles formas de tomar dos parejas distintas, y considerar el número $c - d$, donde c es el número de parejas concordantes y d el número de

parejas discordantes. Un problema de este número, es que depende de n , por lo que consideramos entonces $(c - d)/(c + d)$, es decir, normalizamos dicho valor. La medida anterior nos resume, de alguna manera, cómo es que estos valores se relacionan. Entonces, dicha medida t , la podemos escribir como:

$$t = \frac{c - d}{c + d} = (c - d) / \binom{n}{2}. \quad (2.10)$$

Si tenemos una pareja (X, Y) de variables aleatorias, ¿cómo podemos definir una medida, que nos resuma la concordancia que hay entre ellas? Veamos la τ de Kendall.

2.3.1 τ de Kendall.

Notemos que t , definida en la ecuación (2.10), es la probabilidad de concordancia menos la probabilidad de discordancia si consideramos la interpretación clásica de la probabilidad, casos favorables entre casos totales. Ahora, si tenemos un vector aleatorio (X, Y) , con distribución conjunta H , dicha medida, que llamaremos τ de Kendall, se define siguiendo el mismo razonamiento. Sean (X_1, Y_1) y (X_2, Y_2) dos vectores aleatorios independientes e idénticamente distribuidos, con distribución conjunta H . Entonces, la τ de Kendall para el vector (X, Y) se define como:

$$\tau(X, Y) = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Para poder estudiar esta medida, definamos antes una función de concordancia Q , que es la diferencia entre la probabilidad de concordancia y la probabilidad de discordancia de dos vectores aleatorios (X_1, Y_1) y (X_2, Y_2) de variables aleatorias continuas, con distribuciones conjuntas H_1 y H_2 posiblemente distintas, pero con marginales comunes F y G . Lo que a continuación veremos, es que dicha medida Q , depende de las distribuciones de (X_1, Y_1) y (X_2, Y_2) , sólo a través de sus cópulas.

Teorema 2.3.1. Sean (X_1, Y_1) y (X_2, Y_2) dos vectores aleatorios independientes de variables aleatorias continuas, con distribución conjunta H_1 y H_2 , respectivamente, y con marginales comunes F (para X_1 y X_2) y G (para Y_1 y Y_2). Sean C_1 y C_2 las cópulas de (X_1, Y_1) y (X_2, Y_2) , respectivamente, de tal manera que $H_1(x, y) = C_1(F(x), G(y))$ y $H_2(x, y) = C_2(F(x), G(y))$. Sea

Q la diferencia entre la probabilidad de concordancia y de discordancia de las parejas (X_1, Y_1) y (X_2, Y_2) , es decir, sea

$$Q = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Entonces

$$Q = Q(C_1, C_2) = 4 \iint_{\mathbf{R}^2} C_2(u, v) dC_1(u, v) - 1.$$

Demostración. Dado que las variables aleatorias son continuas,

$$\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0] = 1 - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0]$$

y entonces

$$Q = 2\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1. \quad (2.11)$$

Sabemos que

$$\{(X_1 - X_2)(Y_1 - Y_2) > 0\} \stackrel{c.s.}{=} \{X_1 > X_2, Y_1 > Y_2\} \cup \{X_1 < X_2, Y_1 < Y_2\},$$

por lo que

$$\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] = \mathbb{P}[X_1 > X_2, Y_1 > Y_2] + \mathbb{P}[X_1 < X_2, Y_1 < Y_2],$$

donde cada una de estas probabilidades puede ser calculada integrando con respecto a la distribución conjunta de alguno de los vectores aleatorios utilizando la regla de las probabilidades totales y el hecho de que (X_1, Y_1) y (X_2, Y_2) son independientes entre sí. Por lo que tenemos que

$$\begin{aligned} \mathbb{P}[X_1 > X_2, Y_1 > Y_2] &= \mathbb{P}[X_2 < X_1, Y_2 < Y_1] \\ &= \iint_{\mathbf{R}^2} \mathbb{P}[X_2 \leq x, Y_2 \leq y] dC_1(F(x), G(y)) \\ &= \iint_{\mathbf{R}^2} C_2(F(x), G(y)) dC_1(F(x), G(y)) \end{aligned}$$

y haciendo la transformación $u = F(x)$ y $v = G(y)$ obtenemos

$$\mathbb{P}[X_1 > X_2, Y_1 > Y_2] = \iint_{\mathbf{R}^2} C_2(u, v) dC_1(u, v). \quad (2.12)$$

De la misma manera

$$\begin{aligned}\mathbb{P}[X_1 < X_2, Y_1 < Y_2] &= \iint_{\mathbf{R}^2} \mathbb{P}[X_2 > x, Y_2 > y] dC_1(F(x), G(y)) \\ &= \iint_{\mathbf{R}^2} [1 - F(x) - G(y) + C_2(F(x), G(y))] dC_1(F(x), G(y)) \\ &= \iint_{\mathbf{I}^2} [1 - u - v + C_2(u, v)] dC_1(u, v).\end{aligned}$$

Ahora, como C_1 es la distribución conjunta de una pareja (U, V) de variables aleatorias uniformes $(0, 1)$, entonces $E(U) = E(V) = \frac{1}{2}$ y por lo tanto

$$\mathbb{P}[X_1 < X_2, Y_1 < Y_2] = \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v). \quad (2.13)$$

Por lo tanto

$$\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] = 2 \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v).$$

Y sustituyendo en la ecuación (2.11) se obtiene el resultado. \square

La función Q juega un papel importante para el desarrollo de esta sección. Por ello, veamos algunas propiedades importantes.

Corolario 2.3.1.1. Sean C_1, C_2 y Q como en el teorema 2.3.1. Entonces:

1. Q es simétrica en sus argumentos, es decir, $Q(C_1, C_2) = Q(C_2, C_1)$;
2. Q es creciente en cada argumento, es decir, si $C_1 \leq C'_1$ y $C_2 \leq C'_2$ para todo $(u, v) \in \mathbf{I}^2$, entonces $Q(C_1, C_2) \leq Q(C'_1, C'_2)$.

Demostración. El inciso (1) es inmediato de la definición de la función Q . Para el inciso (2), como $C_2 \leq C'_2$ y por monotonía de la integral tenemos que $Q(C_1, C_2) \leq Q(C_1, C'_2)$. Por la misma argumentación y por el inciso (1) tenemos que $Q(C_2, C_1) \leq Q(C'_2, C_1)$ lo que completa la prueba. \square

Para calcular Q para las cópulas más conocidas, necesitamos probar antes el siguiente resultado.

Lema 2.3.1. *Sea C una cópula cuyo soporte en el rectángulo $R = [u_1, u_2] \times [v_1, v_2] \subseteq \mathbf{I}^2$ está sobre una función f que es continua y biyectiva sobre todo el rectángulo R , entonces, para toda función g integrable en \mathbf{I}^2 , tenemos que*

$$\iint_R g(u, v) dC(u, v) = \int_{u_1}^{u_2} g(u, f(u)) du. \quad (2.14)$$

Demostración. Supongamos que f es creciente. Sea $\mathcal{P} = \{a_0, a_1, \dots, a_n\}$ una partición que cumple $u_1 = a_0 < a_1 < \dots < a_n = u_2$. Los n^2 rectángulos de la forma $[a_{k-1}, a_k] \times [f(a_{j-1}), f(a_j)]$, con $k, j = 1, \dots, n$, forman una partición de R . La integral del lado izquierdo de (2.14) tiene como suma de Riemann a

$$\sum_{k=1}^n \sum_{j=1}^n g(a_k, f(a_j)) V_C([a_{k-1}, a_k] \times [f(a_{j-1}), f(a_j)]).$$

Como el soporte de C , está sobre la función $y = f(x)$, todos los rectángulos que no sean de la forma $[a_{k-1}, a_k] \times [f(a_{k-1}), f(a_k)]$, con $k = 1, \dots, n$, tienen C -volumen 0. Por lo tanto, la suma anterior queda de la siguiente forma

$$\begin{aligned} & \sum_{k=1}^n g(a_k, f(a_k)) V_C([a_{k-1}, a_k] \times [f(a_{k-1}), f(a_k)]) \\ &= \sum_{k=1}^n g(a_k, f(a_k)) [C(a_k, f(a_k)) - C(a_{k-1}, f(a_k)) \\ & \quad - C(a_k, f(a_{k-1})) + C(a_{k-1}, f(a_{k-1}))]. \end{aligned}$$

Para calcular esta suma, notemos que el rectángulo $[a_{k-1}, a_k] \times [f(a_k), 1]$ no interseca con el soporte de C , por lo que obtenemos que

$$C(a_k, f(a_k)) - C(a_{k-1}, f(a_k)) = a_k - a_{k-1} = \Delta a_k.$$

Notemos además, que el rectángulo $[a_{k-1}, a_k] \times [0, f(a_{k-1})]$ tampoco interseca con el soporte, y entonces

$$C(a_k, f(a_{k-1})) - C(a_{k-1}, f(a_{k-1})) = 0.$$

Por lo tanto, la suma queda como

$$\sum_{k=1}^n g(a_k, f(a_k)) \Delta a_k,$$

que no es otra cosa que la suma de Riemann de la segunda integral de (2.14), lo que completa la prueba.

Supongamos que f es decreciente. Consideremos la misma partición \mathcal{P} y definamos la partición $\mathcal{P}_1 = \{b_0, b_1, \dots, b_n\}$, con $f(a_k) = b_{n-k}$ para $k = 1, \dots, n$. Entonces, los n^2 rectángulos de la forma $[a_{k-1}, a_k] \times [b_{j-1}, b_j]$ forman una partición de R . Para la primera integral de (2.14) tenemos como suma de Riemann a

$$\sum_{k=1}^n \sum_{j=1}^n g(a_k, b_{j-1}) V_C([a_{k-1}, a_k] \times [b_{j-1}, b_j]).$$

Todos los rectángulos que no sean de la forma $[a_{k-1}, a_k] \times [b_{n-k}, b_{n-(k-1)}]$, tienen C -volumen 0, dejando a la suma como

$$\begin{aligned} & \sum_{k=1}^n g(a_k, b_{n-k}) V_C([a_{k-1}, a_k] \times [b_{n-k}, b_{n-(k-1)}]) \\ &= \sum_{k=1}^n g(a_k, f(a_k)) [C(a_k, f(a_{k-1})) - C(a_{k-1}, f(a_{k-1})) \\ & \quad - C(a_k, f(a_k)) + C(a_{k-1}, f(a_k))], \end{aligned}$$

y mediante argumentos análogos a los usados para f creciente, concluimos la prueba. \square

Teorema 2.3.2. *Sea C una cópula cuyo soporte está sobre una función f que es biyectiva sobre todo \mathbf{I}^2 y continua excepto a lo más en un número finito de puntos, entonces, para toda función g integrable en \mathbf{I}^2 , tenemos que*

$$\iint_{\mathbf{I}^2} g(u, v) dC(u, v) = \int_0^1 g(u, f(u)) du.$$

Demostración. Sean $0 < c_1 < c_2 < \dots < c_n < 1$, los puntos sobre el eje x donde f es discontinua. Sea $R_k = [c_{k-1}, c_k] \times [d_1^k, d_2^k]$, para $k = 1, \dots, n$, con $c_0 = 0$, $d_1^k = \inf_{c_{k-1} < x < c_k} f(x)$ y $d_2^k = \sup_{c_{k-1} < x < c_k} f(x)$. Entonces

$$\iint_{\mathbf{I}^2} g(u, v) dC(u, v) = \sum_{k=1}^n \iint_{R_k} g(u, v) dC(u, v),$$

ya que fuera de esos rectángulos la integral vale 0 puesto que no interseca con el soporte. La función f en cada uno de estos rectángulos es continua

y biyectiva y aplicando el Lema 2.3.1 para cada integral del lado derecho se concluye la prueba. \square

Ejemplo 2.3.1. Consideremos las cópulas W , Π y M y calculemos Q para cualquier pareja de ellas. Para esto, recordemos que el soporte de M está sobre la curva $u = v$ en \mathbf{I}^2 , y por lo tanto, si g es una función integrable cuyo dominio es \mathbf{I}^2 , tenemos por el Teorema 2.3.2

$$\iint_{\mathbf{I}^2} g(u, v) dM(u, v) = \int_0^1 g(u, u) du.$$

Por lo tanto,

$$Q(M, M) = 4 \iint_{\mathbf{I}^2} \min(u, v) dM(u, v) - 1 = 4 \int_0^1 u du - 1 = 1;$$

$$Q(M, \Pi) = 4 \iint_{\mathbf{I}^2} uv dM(u, v) - 1 = 4 \int_0^1 u^2 du - 1 = \frac{1}{3};$$

$$\begin{aligned} Q(M, W) &= 4 \iint_{\mathbf{I}^2} \max(u + v - 1, 0) dM(u, v) - 1 \\ &= 4 \int_{1/2}^1 (2u - 1) du - 1 = 0. \end{aligned}$$

De manera similar, recordemos que el soporte de W está sobre la curva $v = 1 - u$ en \mathbf{I}^2 , y por lo tanto, si g es una función integrable cuyo dominio es \mathbf{I}^2 , tenemos

$$\iint_{\mathbf{I}^2} g(u, v) dW(u, v) = \int_0^1 g(u, 1 - u) du,$$

de donde tenemos

$$Q(W, \Pi) = 4 \iint_{\mathbf{I}^2} uv dW(u, v) - 1 = 4 \int_0^1 u(1 - u) du - 1 = -\frac{1}{3};$$

$$Q(W, W) = 4 \iint_{\mathbf{I}^2} \max(u + v - 1, 0) dW(u, v) - 1 = 4 \int_0^1 0 du - 1 = -1.$$

Por último, como $d\Pi(u, v) = dudv$,

$$Q(\Pi, \Pi) = 4 \iint_{\mathbf{I}^2} uv d\Pi(u, v) - 1 = 4 \int_0^1 \int_0^1 uv dudv - 1 = 0. \quad (2.15)$$

□

τ es un caso particular de Q .

Teorema 2.3.3. Sean X y Y dos variables aleatorias continuas con cópula C . La τ de Kendall para X y Y está dada por

$$\tau(X, Y) = Q(C, C) = 4 \iint_{\mathbf{I}^2} C(u, v) dC(u, v) - 1.$$

Demostración. Basta considerar $C_1 = C_2 = C$ en el Teorema 2.3.1 □

Por el teorema anterior, vemos que a diferencia de lo que vimos con la correlación, el valor de τ sólo depende de la cópula y no del comportamiento marginal. ¿Eso hará que esta medida sea mejor que la correlación? Para tratar de contestar esta pregunta estudiemos qué pasa con τ .

Primero, veamos cuál es el rango de esta medida. Si tenemos una pareja (X, Y) con cópula C , por el Teorema 2.3.3 sabemos que $\tau(X, Y) = Q(C, C)$. Además por el Teorema 1.4.1 sabemos que para cualquier cópula C tenemos que $W \leq C \leq M$, y mediante el Corolario 2.3.1.1 y el ejemplo 2.3.1 podemos ver que $-1 = Q(W, W) \leq Q(C, C) \leq Q(M, M) = 1$. Por lo tanto, si X y Y son dos variables aleatorias continuas

$$-1 \leq \tau(X, Y) \leq 1.$$

Empecemos a analizar la información que nos dan los valores extremos, así como la información que da $\tau = 0$.

Teorema 2.3.4. Sean X y Y dos variables aleatorias continuas con cópula C . Entonces,

1. $\tau(X, Y) = 1$ si y sólo si $C = M$,
2. $\tau(X, Y) = -1$ si y sólo si $C = W$.

Demostración. Supongamos que $\tau(X, Y) = 1$. Por el Teorema 2.3.3, esto pasa si y sólo si $\iint C dC = \frac{1}{2}$. De la demostración del Teorema 2.3.1 podemos ver que, al igual que para la ecuación (2.12), tenemos que

$$\mathbb{P}[X_2 > X_1, Y_2 > Y_1] = \iint_{\mathbb{R}^2} C_1(u, v) dC_2(u, v),$$

y por la ecuación (2.13), vemos que si tenemos dos cópulas C_1 y C_2 , entonces $\iint C_1 dC_2 = \iint C_2 dC_1$. Por lo tanto, $\iint C dC \leq \iint M dC = \iint C dM$. Por el Teorema 2.3.2, para una cópula C , tenemos que

$$\iint_{\mathbb{R}^2} C(u, v) dM(u, v) = \int_0^1 C(u, u) du.$$

Como C , sobre la diagonal principal, es continua y acotada por M , la única forma que $\int_0^1 C(u, u) du = \frac{1}{2}$ es si $C(u, u) = u$, es decir, si C coincide con M sobre la diagonal principal, lo que por el Lema 1.6.5 implica que $C = M$. Basta, por la continuidad de C , encontrar un punto (u_1, u_1) donde $C(u_1, u_1) < u_1$, para encontrar un intervalo abierto donde $C(u, u) < u$, lo que implica que $\iint C dC \leq \int_0^1 C(u, u) du < \frac{1}{2}$. Por lo tanto, si $\tau(X, Y) = 1$, entonces $C = M$.

Supongamos que $\tau(X, Y) = -1$. Otra vez, por el Teorema 2.3.1, esto pasa si y sólo si $\iint C dC = 0$. De la misma manera, tenemos que $\iint C dC \geq \iint W dC = \iint C dW$. Por el Teorema 2.3.2 sabemos que

$$\iint_{\mathbb{R}^2} C(u, v) dW(u, v) = \int_0^1 C(u, 1 - u) du.$$

Como $C \geq 0$ y continua, la única forma que $\int_0^1 C(u, 1 - u) du = 0$ es si $C(u, 1 - u) = 0$, lo que por el Lema 1.6.6, implica que $C = W$. Basta, por la continuidad de C , encontrar un punto $(u_1, 1 - u_1)$ donde $C(u_1, 1 - u_1) > 0$, para encontrar un intervalo abierto donde $C(u, 1 - u) > 0$, lo que implica que $\iint C dC \geq \int_0^1 C(u, 1 - u) du > 0$. Por lo tanto, si $\tau(X, Y) = -1$, entonces $C = W$. \square

Si $|\tau(X, Y)| = 1$, el problema de la dependencia entre las variables está resuelto; tenemos entonces que una variable es función monótona de la otra, y el conocimiento de una bastará para el conocimiento de la pareja.

Además de esta propiedad, como $\tau(X, Y) = Q(C, C)$, es decir, depende sólo de la cópula, hereda las propiedades que ésta tiene, en especial la invarianza bajo transformaciones monótonas (véase Teoremas 1.6.3 y 1.6.4).

Ahora, si $\tau(X, Y) = 0$, ¿qué podemos decir sobre la dependencia de la pareja (X, Y) ? Primero, veamos que si X y Y son dos variables aleatorias continuas e independientes, por el Teorema 1.6.2 sabemos que (X, Y) tendrá como cópula $\Pi(u, v) = uv$. Ahora por la ecuación (2.15) y el Teorema 2.3.3, tenemos que

$$\tau(X, Y) = 4 \iint_{\mathbb{R}^2} uv \, dudv - 1 = 0.$$

Por lo tanto, si X y Y son independientes, $\tau(X, Y) = 0$. El recíproco, en general, no es cierto, aunque existen casos donde sí lo es. Como es de esperarse, el mundo normal bivariado nos permite esta doble implicación.

Ejemplo 2.3.2. Sean X, Y dos variables aleatorias $N(0, 1)$ con coeficiente de correlación ρ . Entonces

$$\tau(X, Y) = \frac{2}{\pi} \arcsen(\rho). \quad (2.16)$$

Antes de realizar el cálculo, notemos que $\tau(X, Y) = 0$ si y sólo si $\rho = 0$ si y sólo si X y Y son independientes (véase ejemplo 2.2.1). Por lo tanto, una vez más, en el caso normal bivariado, $\tau(X, Y) = 0$ si y sólo si X y Y son independientes.

Para probar la ecuación (2.16), veamos lo siguiente.

$$\begin{aligned} \tau(X, Y) &= \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0] \\ &= \mathbb{E} [\mathbb{I}_{(X_1 - X_2)(Y_1 - Y_2) > 0} - \mathbb{I}_{(X_1 - X_2)(Y_1 - Y_2) < 0}] \\ &= \mathbb{E} [\text{sgn}(X_1 - X_2) \text{sgn}(Y_1 - Y_2)], \end{aligned}$$

donde (X_1, Y_1) y (X_2, Y_2) son dos vectores independientes con distribución igual a (X, Y) ; y $\text{sgn}(x) = \mathbb{I}_{(0, \infty)}(x) - \mathbb{I}_{(-\infty, 0)}(x)$ es la función signo.

Para facilitar el trabajo, veamos cómo se distribuye la pareja $(X_1 - X_2, Y_1 - Y_2)$. Como (X_1, Y_1) y (X_2, Y_2) son independientes, la función de densidad conjunta está dada por

$$f(x_1, y_1, x_2, y_2) = \frac{1}{4\pi^2(1 - \rho^2)} e^{-\frac{1}{2(1 - \rho^2)}(x_1^2 + x_2^2 - 2\rho(x_1 y_1 + x_2 y_2) + y_1^2 + y_2^2)}.$$

Ahora consideremos la transformación

$$W_1 = \frac{1}{\sqrt{2}}(X_1 - X_2), \quad W_2 = \frac{1}{\sqrt{2}}(X_1 + X_2),$$

$$Z_1 = \frac{1}{\sqrt{2}}(Y_1 - Y_2), \quad Z_2 = \frac{1}{\sqrt{2}}(Y_1 + Y_2).$$

El determinante del jacobiano de la transformación es

$$|J| = \begin{vmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{vmatrix} = 1.$$

Notemos además que

$$W_1^2 + W_2^2 = \frac{(X_1 - X_2)^2 + (X_1 + X_2)^2}{2} = X_1^2 + X_2^2,$$

y de la misma manera

$$Z_1^2 + Z_2^2 = Y_1^2 + Y_2^2;$$

además

$$\begin{aligned} W_1 Z_1 + W_2 Z_2 &= \frac{(X_1 - X_2)(Y_1 - Y_2) + (X_1 + X_2)(Y_1 + Y_2)}{2} \\ &= \frac{X_1(Y_1 - Y_2 + Y_1 + Y_2) + X_2(Y_2 - Y_1 + Y_1 + Y_2)}{2} \\ &= X_1 Y_1 + X_2 Y_2, \end{aligned}$$

y por lo tanto, la función de densidad de la transformación es

$$\begin{aligned} g(w_1, z_1, w_2, z_2) &= \frac{1}{4\pi^2(1-\rho^2)} e^{-\frac{1}{2(1-\rho^2)}(w_1^2+w_2^2-2\rho(w_1 z_1+w_2 z_2)+z_1^2+z_2^2)} \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(w_1^2-2\rho w_1 z_1+z_1^2)} \times \\ &\quad \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(w_2^2-2\rho w_2 z_2+z_2^2)}. \end{aligned}$$

Por lo tanto, la pareja (W_1, Z_1) se distribuye normal estándar bivariada con coeficiente de correlación ρ . Retomando el cálculo inicial tenemos

$$\begin{aligned} \mathbb{E}[\operatorname{sgn}(X_1 - X_2)\operatorname{sgn}(Y_1 - Y_2)] &= \mathbb{E}\left[\operatorname{sgn}\left(\frac{X_1 - X_2}{\sqrt{2}}\right)\operatorname{sgn}\left(\frac{Y_1 - Y_2}{\sqrt{2}}\right)\right] \\ &= \mathbb{E}[\operatorname{sgn}(W)\operatorname{sgn}(Z)] \\ &= \iint_{\mathbb{R}^2} \operatorname{sgn}(w)\operatorname{sgn}(z) \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(w^2-2\rho w z+z^2)} dw dz. \end{aligned}$$

(2.17)

Para realizar este cálculo utilicemos el siguiente hecho:

$$\operatorname{sgn}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{itx}}{ix} dx. \quad (2.18)$$

Primero, si $t = 0$, entonces $\int_{-\infty}^{\infty} \frac{dx}{x} = 0$, ya que $1/x$ es una función impar. Si $t > 0$ el cálculo se realiza utilizando integrales complejas de la siguiente manera. Para eso, consideremos el valor principal de las integrales, es decir,

$$\int_{-\infty}^{\infty} = \lim_{\substack{c \rightarrow \infty \\ \epsilon \rightarrow 0}} \left(\int_{-c}^{-\epsilon} + \int_{\epsilon}^c \right),$$

y recordemos que si tenemos una curva $\gamma : [a, b] \rightarrow \mathbb{C}$ diferenciable y f una función compleja, entonces la integral de f sobre γ está dada por

$$\int_{\gamma} f dz = \int_a^b f(\gamma(s)) \gamma'(s) ds.$$

Consideremos la integral compleja

$$\int_{\gamma_{R,\epsilon}} \frac{e^{itz}}{iz} dz,$$

donde $\gamma_{R,\epsilon}$ es la curva que va de $-R$ a $-\epsilon$; un semicírculo de radio ϵ sobre el eje real; de ϵ a R ; sube de R a $R + iR$; de $R + iR$ a $-R + iR$; y baja de $-R + iR$ a $-R$. En otras palabras, es una curva que va sobre el eje real excepto en cero, y sube sobre una caja a la parte compleja. Ahora, como en el interior de $\gamma_{R,\epsilon}$, $f(z) = e^{itz}/iz$ es analítica,

$$\int_{\gamma_{R,\epsilon}} f(z) dz = 0.$$

La integral de arriba está dividida en seis partes, según fuimos describiendo la curva. La primera, I_1 , y la tercera, I_3 , convergen, al aplicar los límites de R y ϵ , a la integral que estamos buscando. Calculemos las demás integrales: I_2 , I_4 , I_5 e I_6 . Para I_4 tenemos que

$$I_4 = \int_0^R \frac{e^{it(R+is)}}{R+is} i ds,$$

y calculando el módulo

$$\begin{aligned} |I_4| &\leq \int_0^R \left| \frac{e^{it(R+is)}}{R+is} i \right| ds = \int_0^R \frac{e^{-ts}}{\sqrt{R^2+s^2}} ds \\ &\leq \int_0^R \frac{e^{-ts}}{R} ds = \frac{1-e^{-tR}}{tR} \xrightarrow{R \rightarrow \infty} 0. \end{aligned}$$

De manera similar, para I_5 tenemos que

$$I_5 = \int_R^{-R} \frac{e^{it(s+iR)}}{s+iR} ds,$$

y calculando el módulo

$$\begin{aligned} |I_5| &\leq \int_{-R}^R \left| \frac{e^{it(s+iR)}}{s+iR} \right| ds = \int_{-R}^R \frac{e^{-tR}}{\sqrt{R^2+s^2}} ds \\ &\leq \int_{-R}^R \frac{e^{-tR}}{R} ds = 2e^{-tR} \xrightarrow{R \rightarrow \infty} 0. \end{aligned}$$

Para I_6 podemos ver que

$$I_6 = \int_R^0 \frac{e^{it(-R+is)}}{-R+is} i ds,$$

y calculando el módulo

$$\begin{aligned} |I_6| &\leq \int_0^R \left| \frac{e^{it(-R+is)}}{-R+is} i \right| ds = \int_0^R \frac{e^{-ts}}{\sqrt{R^2+s^2}} ds \\ &\leq \int_0^R \frac{e^{-ts}}{R} ds = \frac{1-e^{-tR}}{tR} \xrightarrow{R \rightarrow \infty} 0. \end{aligned}$$

Por lo tanto, cuando R tiende a infinito, la integral sobre el rectángulo se desvanece. Sólo resta ver qué pasa con I_2 . Para eso, veamos que cerca del cero, $f(z)$ la podemos ver como

$$\begin{aligned} \frac{e^{itz}}{iz} &= \frac{\sum_{n=0}^{\infty} \frac{(itz)^n}{n!}}{iz} = \frac{1}{iz} + \sum_{n=1}^{\infty} \frac{t^n (iz)^{n-1}}{n!} \\ &= \frac{1}{iz} + h(z), \end{aligned}$$

donde $h(z)$ es una función analítica alrededor del cero. Entonces,

$$I_2 = \int_{\Gamma_\epsilon} \frac{dz}{iz} + \int_{\Gamma_\epsilon} h(z) dz.$$

La primera parte de esta integral es

$$\int_{\Gamma_\epsilon} \frac{dz}{iz} = \int_\pi^0 \frac{i\epsilon e^{is}}{i\epsilon e^{is}} ds = -\pi.$$

Para la segunda parte, por ser h analítica, es acotada cerca del 0, digamos que por M , y entonces

$$\left| \int_{\Gamma_\epsilon} h(z) dz \right| \leq M \int_0^\pi \epsilon ds = M\pi\epsilon \xrightarrow{\epsilon \rightarrow 0} 0.$$

Juntando todos los cálculos tenemos que

$$0 = \lim_{\substack{\epsilon \rightarrow 0 \\ R \rightarrow \infty}} \int_{\gamma_{R,\epsilon}} \frac{e^{itz}}{iz} dz = \int_{-\infty}^{\infty} \frac{e^{itx}}{ix} dx - \pi,$$

de donde el resultado (2.18) para $t > 0$ es válido. Si $t < 0$, basta considerar lo calculado para $t > 0$ y hacer un cambio de variable, $u = -x$, en la integral, lo que completa la ecuación (2.18).

Utilizando la ecuación (2.18), en la ecuación (2.17), tenemos que

$$\begin{aligned} \mathbb{E}[\text{sgn}(W)\text{sgn}(Z)] &= \iint_{\mathbf{R}^2} \left(\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{iwx}}{ix} dx \right) \left(\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{izy}}{iy} dy \right) \\ &\quad \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(w^2-2\rho wz+z^2)} dw dz \\ &= \frac{1}{\pi^2} \int_{-\infty}^{\infty} \frac{1}{ix} \int_{-\infty}^{\infty} \frac{1}{iy} \\ &\quad \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{iwx+izy} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(w^2-2\rho wz+z^2)} dw dz \right) dy dx. \end{aligned} \tag{2.19}$$

La expresión que está entre paréntesis, es la función característica conjunta, de una pareja normal estándar bivariada de correlación ρ . Para calcular su

valor, recordemos que, para una variable $N(\mu, \sigma^2)$, su función característica está dada por $\exp(it\mu - \sigma^2 t^2/2)$. Por lo tanto

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{iwx+izy} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(w^2-2\rho wz+z^2)} dw dz \\ &= \int_{-\infty}^{\infty} e^{izy} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iwx} \frac{e^{-\frac{(w-\rho z)^2}{2(1-\rho^2)}}}{\sqrt{2\pi(1-\rho^2)}} dw dz. \end{aligned}$$

La integral de adentro, es la función característica de una variable $N(\rho z, 1-\rho^2)$, evaluada en x , de donde

$$\begin{aligned} &= e^{-\frac{1}{2}(1-\rho^2)x^2} \int_{-\infty}^{\infty} e^{iz(y+\rho x)} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz \\ &= e^{-\frac{1}{2}(1-\rho^2)x^2} e^{-\frac{1}{2}(y+\rho x)^2} \\ &= e^{-\frac{1}{2}(x^2+2\rho xy+y^2)}. \end{aligned}$$

Continuando con la ecuación (2.19), tenemos que

$$\tau(X, Y) = -\frac{1}{\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}(x^2+2\rho xy+y^2)}}{xy} dx dy.$$

Derivando la expresión con respecto a ρ tenemos que

$$\begin{aligned} \frac{\partial \tau(X, Y)}{\partial \rho} &= -\frac{\partial}{\partial \rho} \frac{1}{\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}(x^2+2\rho xy+y^2)}}{xy} dx dy \\ &= -\frac{1}{\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial}{\partial \rho} \frac{e^{-\frac{1}{2}(x^2+2\rho xy+y^2)}}{xy} dx dy \\ &= \frac{1}{\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+2\rho xy+y^2)} dx dy, \end{aligned}$$

y haciendo la transformación $x = u/\sqrt{1-\rho^2}$, $y = v/\sqrt{1-\rho^2}$, tenemos que

$$= \frac{2\pi\sqrt{1-\rho^2}}{\pi^2(1-\rho^2)} = \frac{2}{\pi\sqrt{1-\rho^2}},$$

e integrando se obtiene el resultado deseado en la ecuación (2.16). Por lo tanto, si (X, Y) es un vector gaussiano, entonces X y Y son independientes si y sólo si $\tau(X, Y) = 0$. \square

Nuestra duda sobre la relación entre el valor $\tau = 0$ y la independencia, en general, sigue en pie. Para aclararla veamos algunos ejemplos.

Ejemplo 2.3.3. Sea (X, Y) un vector aleatorio con cópula $C_{\alpha, \beta}$ dada por

$$C_{\alpha, \beta} = \alpha M + (1 - \alpha - \beta)\Pi + \beta W,$$

donde $\alpha, \beta \geq 0$ y $\alpha + \beta \leq 1$. Para calcular τ para esta pareja veamos que

$$dC_{\alpha, \beta} = \alpha dM + (1 - \alpha - \beta)d\Pi + \beta dW,$$

y por lo calculado en el ejemplo 2.3.1 tenemos

$$\begin{aligned} \tau(X, Y) &= \iint C_{\alpha, \beta} dC_{\alpha, \beta} \\ &= \iint (\alpha M + (1 - \alpha - \beta)\Pi + \beta W) \alpha dM \\ &\quad + \iint (\alpha M + (1 - \alpha - \beta)\Pi + \beta W) (1 - \alpha - \beta) d\Pi \\ &\quad + \iint (\alpha M + (1 - \alpha - \beta)\Pi + \beta W) \beta dW \\ &= \left(\alpha + \frac{1 - \alpha - \beta}{3} \right) \alpha + \left(\frac{\alpha}{3} - \frac{\beta}{3} \right) (1 - \alpha - \beta) \\ &\quad + \left(-\frac{1 - \alpha - \beta}{3} - \beta \right) \beta \\ &= \frac{(\alpha - \beta)(1 - \alpha - \beta) + \alpha - \beta + 2(\alpha^2 - \beta^2)}{3} \end{aligned}$$

y por lo tanto

$$\tau(X, Y) = \frac{(\alpha - \beta)(\alpha + \beta + 2)}{3}. \quad (2.20)$$

En la ecuación (2.20) para que el lado derecho sea cero, basta que $\alpha = \beta \in [0, \frac{1}{2}]$. El único caso que representa la independencia, es cuando $\alpha = \beta = 0$. Entonces, para $\alpha = \beta \in (0, 1]$, tenemos cópulas $C_{\alpha, \alpha} = \alpha(M + W) + (1 - 2\alpha)\Pi$, todas ellas distintas de la que representa la independencia, para la cuales $\tau = 0$. \square

Con el ejemplo anterior, vemos que es cierto que en general, $\tau = 0$ no implica independencia, ya que éste nos da una familia de modelos de dependencia, distintos de la independencia, para los cuales τ toma el valor cero.

Sabemos que si el soporte de una función de distribución bivariada de una pareja (X, Y) , está sobre una función, digamos $y = \alpha(x)$, entonces $Y = \alpha(X)$ c.s. . Para dar un ejemplo de dos variables que sean función la una de la otra, pero que cumplan que $\tau = 0$, necesitamos probar el siguiente resultado,

$$\tau(X, Y) = 1 - 4 \iint_{\mathbf{I}^2} \frac{\partial}{\partial u} C(u, v) \frac{\partial}{\partial v} C(u, v) dudv, \quad (2.21)$$

donde C denota la cópula asociada a X y Y . Para esto probemos el siguiente teorema.

Teorema 2.3.5. *Sea C una cópula tal que el producto $(\partial C/\partial u)(\partial C/\partial v)$ es integrable en \mathbf{I}^2 . Entonces*

$$\iint_{\mathbf{I}^2} C(u, v) dC(u, v) = \frac{1}{2} - \iint_{\mathbf{I}^2} \frac{\partial}{\partial u} C(u, v) \frac{\partial}{\partial v} C(u, v) dudv. \quad (2.22)$$

Demostración. Si C es absolutamente continua, tenemos que

$$\iint_{\mathbf{I}^2} C(u, v) dC(u, v) = \int_0^1 \int_0^1 C(u, v) \frac{\partial^2 C(u, v)}{\partial u \partial v} dudv,$$

e integrando por partes se completa el resultado. No siempre C es de esta forma, por lo que para la demostración en general procederemos mediante sumas de Riemann. Empezaremos por el lado izquierdo de la ecuación (2.22). Consideremos la partición del eje x como $0 = u_0 < u_1 < \dots < u_n = 1$, y la del eje y como $0 = v_0 < v_1 < \dots < v_m = 1$. Sea $\Delta u_i = u_i - u_{i-1}$, con $i = 1, \dots, n$, y $\Delta v_j = v_j - v_{j-1}$, con $j = 1, \dots, m$. Entonces, el producto cartesiano de estos intervalos, forman una partición de \mathbf{I}^2 en mn rectángulos de la forma $R_{ij} = [u_{i-1}, u_i] \times [v_{j-1}, v_j]$. Como el C -volumen de estos rectángulos está dado por $V_C(R_{ij}) = C(u_i, v_j) - C(u_{i-1}, v_j) - C(u_i, v_{j-1}) + C(u_{i-1}, v_{j-1})$, entonces la suma de Riemann de la primera integral en (2.22) es

$$\sum_{j=1}^m \sum_{i=1}^n C(u_i, v_j) [C(u_i, v_j) - C(u_{i-1}, v_j) - C(u_i, v_{j-1}) + C(u_{i-1}, v_{j-1})]. \quad (2.23)$$

La suma para el lado derecho de (2.22) tiene que ser de la forma

$$\sum_{j=1}^m \sum_{i=1}^n [C(u_i, v_j) - C(u_{i-1}, v_j)] [C(u_i, v_j) - C(u_i, v_{j-1})],$$

o algo parecido, entonces desarrollando el sumando de (2.23), tenemos que

$$\begin{aligned}
& C(u_i, v_j)[C(u_i, v_j) - C(u_{i-1}, v_j) - C(u_i, v_{j-1}) + C(u_{i-1}, v_{j-1})] \\
&= C(u_i, v_j)^2 - C(u_i, v_j)C(u_{i-1}, v_j) - C(u_i, v_j)C(u_i, v_{j-1}) \\
&\quad + C(u_i, v_j)C(u_{i-1}, v_{j-1}) \\
&= C(u_i, v_j)^2 - C(u_i, v_j)C(u_i, v_{j-1}) \\
&\quad - C(u_i, v_j)[C(u_{i-1}, v_j) - C(u_{i-1}, v_{j-1})] \\
&= C(u_i, v_j)^2 - C(u_{i-1}, v_j)^2 \\
&\quad - C(u_i, v_j)C(u_i, v_{j-1}) + C(u_{i-1}, v_j)C(u_{i-1}, v_{j-1}) \\
&\quad - [C(u_i, v_j) - C(u_{i-1}, v_j)][C(u_{i-1}, v_j) - C(u_{i-1}, v_{j-1})].
\end{aligned}$$

Al calcular la suma, la primera parte será una suma telescópica, es decir

$$\begin{aligned}
& \sum_{j=1}^m \sum_{i=1}^n C(u_i, v_j)^2 - C(u_{i-1}, v_j)^2 \\
&\quad - C(u_i, v_j)C(u_i, v_{j-1}) + C(u_{i-1}, v_j)C(u_{i-1}, v_{j-1}) \\
&= \sum_{j=1}^m C(1, v_j)^2 - C(0, v_j)^2 \\
&\quad - C(1, v_j)C(1, v_{j-1}) + C(0, v_j)C(0, v_{j-1}) \\
&= \sum_{j=1}^m v_j \Delta v_j,
\end{aligned}$$

que no es otra cosa que la suma de Riemann de $\int_0^1 v dv = \frac{1}{2}$.

La segunda parte puede ser rescrita como

$$\sum_{j=1}^m \sum_{i=1}^n \left[\frac{C(u_i, v_j) - C(u_{i-1}, v_j)}{\Delta u_i} \right] \left[\frac{C(u_{i-1}, v_j) - C(u_{i-1}, v_{j-1})}{\Delta v_j} \right] \Delta u_i \Delta v_j,$$

que es la suma de Riemann de la integral del lado derecho de (2.22), lo que completa la prueba. \square

Veamos los ejemplos antes prometidos.

Ejemplo 2.3.4. Sean U y V dos variables aleatorias uniformes $(0,1)$ cuya cópula C_θ , $\theta \in [0, 1]$, tiene el siguiente soporte: una probabilidad θ distribuida

uniformemente sobre la recta que une el punto $(0, 0)$ con el $(\theta, 1)$ y una probabilidad $1 - \theta$ distribuida uniformemente sobre la recta que une el punto $(\theta, 1)$ con el punto $(1, 0)$ (véase figura 2.5), es decir, $V = \frac{U}{\theta} \mathbb{I}_{(0, \theta)}(U) + \frac{1-U}{1-\theta} \mathbb{I}_{[\theta, 1)}(U)$. Si $\theta = 1$ tenemos que $C_1 = M$ y si $\theta = 0$ tenemos que $C_0 = W$. Por lo tanto, dicha familia de cópulas va de la dependencia estrictamente creciente a la estrictamente decreciente, siempre modelando variables totalmente determinadas la una por la otra.

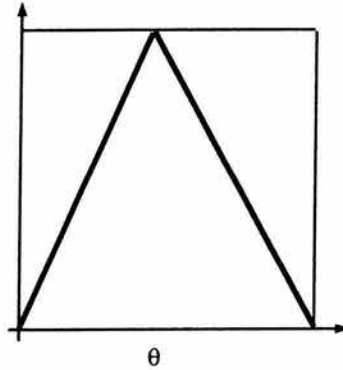


Figura 2.5: Soporte de la cópula C_θ .

Para calcular dicha cópula, consideremos las tres regiones separadas por el soporte empezando de izquierda a derecha. Para la primera, tenemos que $C(u, v) = C(u, 1) = u$, ya que el rectángulo $[0, u] \times [v, 1]$ no interseca el soporte y por lo tanto tiene volumen cero. Para la segunda, $C(u, v) = C(\theta v, v) = C(\theta v, 1) = \theta v$, y en este caso ya que el rectángulo $[\theta v, u] \times [0, v]$ tiene volumen cero. Por último, como el rectángulo $[u, 1] \times [v, 1]$ no interseca el soporte, tenemos que $C(u, v) = u + v - 1$. Resumiendo tenemos que,

$$C_\theta(u, v) = \begin{cases} u, & 0 \leq u \leq \theta v \leq \theta, \\ \theta v, & 0 \leq \theta v < u < 1 - (1 - \theta)v, \\ u + v - 1, & \theta \leq 1 - (1 - \theta)v \leq u \leq 1. \end{cases}$$

Calculemos τ para esta cópula. Si utilizamos la ecuación (2.21) tenemos que

$$\frac{\partial}{\partial u} C_\theta(u, v) \frac{\partial}{\partial v} C_\theta(u, v) = \begin{cases} 1, & \theta \leq 1 - (1 - \theta)v \leq u \leq 1, \\ 0, & \text{en otro caso.} \end{cases}$$

De donde vemos que $\iint_{\mathbf{I}^2} \frac{\partial}{\partial u} C_\theta(u, v) \frac{\partial}{\partial v} C_\theta(u, v) dudv$ es el área del triángulo cuyos vértices son $(\theta, 1)$, $(1, 1)$ y $(1, 0)$, es decir $\frac{1-\theta}{2}$, y por lo tanto

$$\tau_\theta = 2\theta - 1. \quad (2.24)$$

Esta ecuación es cero si $\theta = \frac{1}{2}$, que representa un modelo de variables aleatorias uniformes que cumplen la relación determinista

$$V = \frac{1}{2} \left(U \mathbb{I}_{(0, \frac{1}{2})} (U) + (1 - U) \mathbb{I}_{[\frac{1}{2}, 1)} (U) \right),$$

donde basta conocer el valor de U para conocer el de V . Por lo tanto, el valor $\tau = 0$ puede provenir de parejas cuya relación es muy distinta a la independencia. \square

Veamos más ejemplos.

Ejemplo 2.3.5. Sean U y V dos variables aleatorias uniformes $(0, 1)$ tales que $V = U \oplus \theta^2$, donde $\theta \in [0, 1]$. Para calcular la cópula C_θ de esta pareja, notemos que $V = (U + \theta) \mathbb{I}_{[0, 1-\theta)} (U) + (U + \theta - 1) \mathbb{I}_{[1-\theta, 1]} (U)$, y por lo tanto, el soporte de C_θ , está sobre la función $f(x) = (x + \theta) \mathbb{I}_{[0, 1-\theta)} (x) + (x + \theta - 1) \mathbb{I}_{[1-\theta, 1]} (x)$. Si dividimos el cuadrado unitario en cuatro partes, formadas al trazar las rectas $x = 1 - \theta$ y $y = \theta$ (véase figura 2.6), podemos calcular C_θ en cada uno de esos cuadros. Recordemos que $C_\theta(u, v) = V_{C_\theta}([0, u] \times [0, v])$. Para (u, v) en el cuadrado $[0, 1 - \theta] \times [0, \theta]$, $C_\theta(u, v) = 0$, ya que no interseca con el soporte. Si (u, v) está en $[0, 1 - \theta] \times [\theta, 1]$, tenemos dos casos: si $u < v - \theta$, entonces $C_\theta(u, v) = C_\theta(u, 1) = u$; si $u \geq v - \theta$, entonces $C_\theta(u, v) = C_\theta(v - \theta, v) = v - \theta$. Ahora, si (u, v) está en $[1 - \theta, 1] \times [0, \theta]$, tenemos otra vez dos casos: si $v < u + \theta - 1$, entonces $C_\theta(u, v) = C_\theta(1, v) = v$; si $v \geq u + \theta - 1$, entonces $C_\theta(u, v) = C_\theta(u, u + \theta - 1) = u + \theta - 1$. Por último, si (u, v) está en $[1 - \theta, 1] \times [\theta, 1]$, el cuadrado $[u, 1] \times [v, 1]$ no interseca el soporte, de donde $V_{C_\theta}([u, 1] \times [v, 1]) = 0$, y por lo tanto $C_\theta(u, v) = u + v - 1$. Resumiendo la información calculada arriba tenemos que

$$C_\theta(u, v) = \begin{cases} \min(u, v - \theta) & \text{si } (u, v) \in [0, 1 - \theta] \times [\theta, 1], \\ \min(u + \theta - 1, v) & \text{si } (u, v) \in [1 - \theta, 1] \times [0, \theta], \\ W(u, v) & \text{en otro caso} \end{cases}$$

² $a \oplus b = a + b - [a + b]$, donde $[x]$ denota la parte entera de x

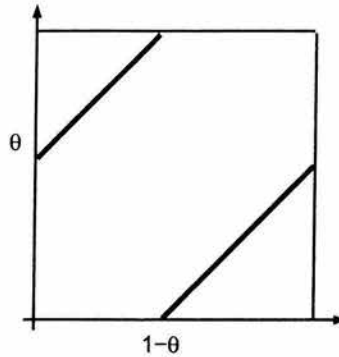


Figura 2.6: Soporte de la cópula C_θ .

Calculemos τ para esta cópula. Dado que su soporte está sobre la curva $v = (u + \theta)\mathbb{I}_{[0, 1-\theta]}(u) + (u + \theta - 1)\mathbb{I}_{[1-\theta, 1]}(u)$, si g es una función integrable cuyo dominio es \mathbf{I}^2 , por el Teorema 2.3.2, tenemos

$$\iint_{\mathbf{I}^2} g(u, v) dC_\theta(u, v) = \int_0^{1-\theta} g(u, u + \theta) du + \int_{1-\theta}^1 g(u, u + \theta - 1) du. \quad (2.25)$$

Entonces

$$\begin{aligned} \tau_\theta &= 4 \iint_{\mathbf{I}^2} C_\theta(u, v) dC_\theta(u, v) - 1 \\ &= 4 \left(\int_0^{1-\theta} \min(u, u) du + \int_{1-\theta}^1 \min(u + \theta - 1, u + \theta - 1) du \right) - 1 \\ &= 4 \left(\int_0^1 u du + \int_{1-\theta}^1 (\theta - 1) du \right) - 1 \\ &= 2 + 4(\theta - 1)\theta - 1, \end{aligned}$$

por lo tanto

$$\tau_\theta = (2\theta - 1)^2. \quad (2.26)$$

La ecuación anterior es cero si $\theta = \frac{1}{2}$. Esto representa a una pareja (U, V) de uniformes cuya relación está dada por $V = U \oplus \frac{1}{2}$, un modelo donde basta

conocer a cualquiera de las dos para conocer a ambas. Por lo tanto, el valor $\tau = 0$ puede corresponder, también, a parejas que estén determinadas la una por la otra. \square

Veamos otro ejemplo.

Ejemplo 2.3.6. Sean U y V dos variables aleatorias uniformes $(0, 1)$, tales que $U \oplus V = \theta$, con $\theta \in [0, 1]$. Notemos que, $U \oplus V = \theta$, lo podemos escribir como $V = (\theta - U)\mathbb{I}_{[0, \theta]}(U) + (\theta + 1 - U)\mathbb{I}_{[\theta, 1]}(U)$, y por lo tanto el soporte de la cópula de la pareja (U, V) , C_θ , está sobre la función $f(x) = (\theta - x)\mathbb{I}_{[0, \theta]}(x) + (\theta + 1 - x)\mathbb{I}_{[\theta, 1]}(x)$. Dividamos el cuadrado unitario en cuatro con las rectas $x = \theta$ y $y = \theta$ (véase figura 2.7), y calculemos en cada cuadrado lo que pasa. Si (u, v) está en el cuadrado $[0, \theta] \times [0, \theta]$, tenemos dos casos: si $u + v < \theta$, entonces $C_\theta(u, v) = 0$, ya que no interseca el soporte; si $u + v \geq \theta$, entonces el rectángulo $[u, 1] \times [v, 1]$ tiene volumen $1 - \theta$, y por lo tanto $1 - u - v + C_\theta(u, v) = 1 - \theta$, de donde obtenemos que $C_\theta(u, v) = u + v - \theta$. Ahora, si (u, v) está en $[0, \theta] \times [\theta, 1]$, entonces $C_\theta(u, v) = C_\theta(u, 1) = u$. Si (u, v) está en el rectángulo $[\theta, 1] \times [0, \theta]$, tenemos que $C_\theta(u, v) = C_\theta(1, v) = v$. Por último, si (u, v) está en el cuadrado $[\theta, 1] \times [\theta, 1]$, tenemos dos casos: si $u + v < 1 + \theta$, entonces $C_\theta(u, v) = \theta$; si $u + v \geq 1 + \theta$, entonces $V_{C_\theta}([u, 1] \times [v, 1]) = 0$, ya que no interseca con el soporte, y por lo tanto $C_\theta(u, v) = u + v - 1$. Resumiendo la información calculada arriba tenemos que

$$C_\theta(u, v) = \begin{cases} \max(0, u + v - \theta) & \text{si } (u, v) \in [0, \theta] \times [0, \theta], \\ \max(\theta, u + v - 1) & \text{si } (u, v) \in [\theta, 1] \times [\theta, 1], \\ M(u, v) & \text{en otro caso} \end{cases}$$

Calculemos la τ para la cópula C_θ . Dado que su soporte está sobre la curva $v = (\theta - u)\mathbb{I}_{[0, \theta]}(u) + (\theta + 1 - u)\mathbb{I}_{[\theta, 1]}(u)$, si g es una función integrable cuyo dominio es \mathbf{I}^2 , entonces tenemos (véase Teorema 2.3.2)

$$\iint_{\mathbf{I}^2} g(u, v) dC_\theta(u, v) = \int_0^\theta g(u, \theta - u) du + \int_\theta^1 g(u, \theta + 1 - u) du. \quad (2.27)$$

Entonces

$$\begin{aligned} \tau_\theta &= 4 \iint_{\mathbf{I}^2} C_\theta(u, v) dC_\theta(u, v) - 1 \\ &= 4 \left(\int_0^\theta \max(0, 0) du + \int_\theta^1 \max(\theta, \theta) du \right) - 1 \\ &= 4\theta(1 - \theta) - 1, \end{aligned}$$

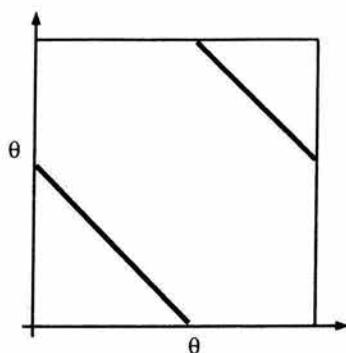


Figura 2.7: Soporte de la cópula C_θ .

por lo tanto

$$\tau_\theta = -(2\theta - 1)^2. \quad (2.28)$$

Otra vez, la ecuación anterior es cero si $\theta = \frac{1}{2}$. En este caso, tenemos una pareja (U, V) , de variables aleatorias uniformes, relacionadas por $U \oplus V = \frac{1}{2}$. Entonces, tenemos otro modelo en el cual los componentes de la pareja son función el uno del otro y $\tau = 0$. \square

Por lo tanto, el valor $\tau = 0$ y la independencia no tienen una implicación recíproca en general. Aprovechemos un poco más estos tres ejemplos. Realmente, ¿qué está midiendo τ ? La respuesta está contestada desde su definición: la probabilidad de concordancia menos la probabilidad de discordancia. Sí, pero ¿qué tiene que ver esto con la dependencia? Olvidemos por un momento nuestra preocupación por el valor $\tau = 0$, ¿qué podemos decir sobre el valor $\tau = c$, $c \in (-1, 1)$, y la dependencia de las variables aleatorias? Según los ejemplos 2.3.4, 2.3.5 y 2.3.6 nada. Notemos que la ecuación (2.24) puede alcanzar cualquier valor entre -1 y 1, mientras que juntando las ecuaciones (2.26) y (2.28), podemos alcanzar todos los valores del rango de τ con parejas de variables que son funciones biyectivas la una de la otra. Entonces, un valor de τ más grande que otro, ¿nos dice que la segunda pareja es más dependiente entre sí que la primera? No. Entonces, nuestra intuición de que conforme nos acerquemos a los valores extremos la dependencia aumentará, es totalmente falsa en general. Pero entendamos un poco más qué significa la probabilidad de concordancia menos la de discordancia. Por ejemplo, si

tenemos dos parejas aleatorias independientes, (U_1, V_1) y (U_2, V_2) , con cópula igual a la calculada en el ejemplo 2.3.5, serán concordantes si U_1 y U_2 están al mismo tiempo en el intervalo $[0, 1 - \theta]$, o bien, en el intervalo $[1 - \theta, 1]$; serán discordantes si se encuentran en intervalos distintos, es decir, serán concordantes si ambas parejas caen en una misma recta, de las dos que componen el soporte de la cópula, y serán discordantes si caen en rectas distintas (véase figura 2.6). Por lo tanto,

$$\begin{aligned}\tau &= \mathbb{P}[\text{concordancia}] - \mathbb{P}[\text{discordancia}] \\ &= (1 - \theta)^2 + \theta^2 - 2\theta(1 - \theta) = (2\theta - 1)^2,\end{aligned}$$

entonces, por ejemplo, en el caso $\theta = \frac{1}{2}$, el valor es cero porque la mitad del “tiempo” son concordantes y la mitad discordantes, aunque todo el tiempo sean variables totalmente determinadas la una por la otra. Una discusión análoga, se puede hacer para la cópula del ejemplo 2.3.6. Por lo tanto, a menos que sepamos algo más sobre la forma en que las variables se relacionan, cualquier valor $\tau \in (-1, 1)$ puede no significar lo que creamos.

En conclusión, los valores extremos de τ nos resuelven por completo el problema de la dependencia; para la distribución normal bivariada τ es una buena medida, aunque no la natural; cualquier valor de $\tau \in (-1, 1)$ debe ser tratado con cuidado, es decir, entender que sólo se trata de medir la concordancia que hay entre dos parejas independientes e idénticamente distribuidas. En otras palabras, los únicos valores en general concluyentes para τ son 1 y -1.

2.3.2 ρ de Spearman

La ρ de Spearman es una medida, al igual que la τ de Kendall, basada en la concordancia y la discordancia. Consideremos otra vez dos parejas de vectores aleatorios (X_1, Y_1) , (X_2, Y_2) , independientes, pero a diferencia de la τ , no idénticamente distribuidos: el vector (X_1, Y_1) con distribución conjunta H y distribuciones marginales F y G , mientras que el vector (X_2, Y_2) , con distribución conjunta FG , es decir, un vector con entradas independientes. La ρ de Spearman, para el vector aleatorio (X, Y) con distribución conjunta H y distribuciones marginales F y G , se define como la probabilidad de concordancia menos la probabilidad de discordancia que existe entre las parejas (X_1, Y_1) y (X_2, Y_2) definidas arriba. Para diferenciarla de la correlación, denotaremos la ρ de Spearman como ρ_s .

$$\rho_s(X, Y) = 3(\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]),$$

donde el “3” es una constante de normalización. Empecemos a ver las propiedades de ρ_s .

Teorema 2.3.6. Sean X y Y dos variables aleatorias continuas con cópula C . La ρ de Spearman para X y Y está dada por

$$\begin{aligned}\rho_s(X, Y) &= 3Q(C, \Pi), \\ &= 12 \iint_{\mathbf{I}^2} uv dC(u, v) - 3, \\ &= 12 \iint_{\mathbf{I}^2} C(u, v) dudv - 3.\end{aligned}\tag{2.29}$$

Demostración. La prueba se sigue directamente de la definición de ρ_s , y del Teorema 2.3.1. \square

Para ver los posibles valores que toma ρ_s , recordemos que para toda cópula C tenemos que $W \leq C \leq M$, y por el Corolario 2.3.1.1, el ejemplo 2.3.1 y la ecuación (2.29), tenemos que $-1 = 3Q(W, \Pi) \leq 3Q(C, \Pi) \leq 3Q(M, \Pi) = 1$, y por lo tanto

$$-1 \leq \rho_s(X, Y) \leq 1.$$

Las mismas preguntas que nos hemos hecho en el caso de la correlación y de τ , las podemos hacer aquí, es decir, ¿qué información dan los casos extremos? y ¿qué información da el valor $\rho_s = 0$? La primera respuesta es sencilla de contestar a partir del teorema 2.3.6.

Teorema 2.3.7. Sean X y Y dos variables aleatorias continuas con cópula C , entonces

1. $\rho_s(X, Y) = 1$ si y sólo si $C = M$,
2. $\rho_s(X, Y) = -1$ si y sólo si $C = W$.

Demostración. $\rho_s(X, Y) = 1$ si y sólo si $\iint_{\mathbf{I}^2} C(u, v) dudv = \frac{1}{3}$ si y sólo si $\iint_{\mathbf{I}^2} C(u, v) dudv$ se maximiza para C , y esto pasa si y sólo si $C = M$. De manera análoga se sigue para $\rho_s = -1$. \square

Entonces, los valores extremos de ρ_s nos dan toda la información necesaria para conocer cómo es el comportamiento conjunto de una pareja de variables aleatorias.

De manera similar que τ , por el Teorema 2.3.6, tenemos que $\rho_s(X, Y)$ sólo depende de la cópula, por lo tanto es invariante bajo transformaciones monótonas (véase Teoremas 1.6.3 y 1.6.4).

Con lo discutido para τ acerca de la relación que guardan el valor cero y la independencia, es de esperarse, al ser ρ_s otra medida de concordancia, que su implicación no sea recíproca en general. Si X y Y son dos variables aleatorias continuas independientes, entonces, por la ecuación (2.29) y el ejemplo 2.3.1 tenemos

$$\rho_s(X, Y) = 3Q(\Pi, \Pi) = 0,$$

es decir, si X y Y son independientes, entonces $\rho_s(X, Y) = 0$. Existen casos donde sí se da una doble implicación, y una vez más, uno de ellos es el de la distribución conjunta Gaussiana.

Ejemplo 2.3.7. Sean X y Y dos variables aleatorias $N(0, 1)$ con coeficiente de correlación ρ . Entonces

$$\rho_s(X, Y) = \frac{6}{\pi} \arcsen\left(\frac{\rho}{2}\right). \quad (2.30)$$

Sabemos que X y Y son independientes si y sólo si $\rho = 0$ (véase ejemplo 2.2.1). Ahora, por la ecuación (2.30), tenemos que $\rho = 0 \Leftrightarrow \rho_s = 0$, y por lo tanto, en el caso normal bivariado, la pareja es independientes si y sólo si $\rho_s = 0$.

Procedamos con el cálculo de la ecuación (2.30). Recordemos que $\text{sgn}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{itx}}{ix} dx$.³

$$\begin{aligned} & \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0] \\ &= \mathbb{E}[\text{sgn}(X_1 - X_2)\text{sgn}(Y_1 - Y_2)] \\ &= \iiint \iiint \text{sgn}(x_1 - x_2)\text{sgn}(y_1 - y_2) dF \\ &= -\frac{1}{\pi^2} \iiint \iiint \int \frac{e^{ia(x_1 - x_2)}}{a} da \int \frac{e^{ib(y_1 - y_2)}}{b} db dF \\ &= -\frac{1}{\pi^2} \iint \frac{1}{ab} \left(\iiint \iiint e^{i(ax_1 + by_1)} e^{-iax_2} e^{-iby_2} dF \right) dadb. \end{aligned}$$

Para calcular la integral que está entre paréntesis, recordemos que para una pareja de normales $(0, 1)$ con correlación ρ , la función característica conjunta

³Para los cálculos de este ejemplo véase el ejemplo 2.3.2

está dada por $\exp(-\frac{1}{2}(a^2 + 2\rho ab + b^2))$, y que los vectores (X_1, Y_1) y (X_2, Y_2) son independientes. Tenemos entonces

$$\begin{aligned} \iiint e^{i(ax_1+by_1)} e^{-iax_2} e^{-iby_2} dF &= e^{-\frac{1}{2}(a^2+2\rho ab+b^2)} e^{-\frac{1}{2}(a^2+b^2)} \\ &= e^{-(a^2+\rho ab+b^2)}. \end{aligned}$$

Retomando el cálculo original tenemos

$$\mathbb{E}[\operatorname{sgn}(X_1 - X_2)\operatorname{sgn}(Y_1 - Y_2)] = -\frac{1}{\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{e^{-(a^2+\rho ab+b^2)}}{ab} dadb,$$

derivando con respecto a ρ la ecuación anterior vemos que

$$\frac{\partial}{\partial \rho} \mathbb{E}[\operatorname{sgn}(X_1 - X_2)\operatorname{sgn}(Y_1 - Y_2)] = \frac{1}{\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(a^2+\rho ab+b^2)} dadb,$$

haciendo la transformación $s = \frac{a}{\sqrt{2(1-(\rho/2)^2)}}$ y $t = \frac{b}{\sqrt{2(1-(\rho/2)^2)}}$ tenemos

$$\begin{aligned} &= \frac{1}{\pi \sqrt{1-(\rho/2)^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi \sqrt{1-(\rho/2)^2}} e^{-\frac{1}{2(1-(\rho/2)^2)(s^2+\rho st+t^2)}} dsdt \\ &= \frac{1}{\pi \sqrt{1-(\rho/2)^2}}, \end{aligned}$$

e integrando y multiplicando por “3” se obtiene el resultado de la ecuación (2.30). \square

Veamos otro ejemplo donde se obtiene esa doble implicación.

Ejemplo 2.3.8. Sean X y Y dos variables aleatorias con cópula C_θ , con $\theta \in [-1, 1]$, dada por

$$C_\theta = uv + \theta uv(1-u)(1-v).$$

Ahora

$$\begin{aligned} \iint_{\mathbb{R}^2} C_\theta dudv &= \left(\int_0^1 u du \right)^2 + \theta \left(\int_0^1 u(1-u) du \right)^2 \\ &= \frac{1}{4} + \frac{\theta}{36}, \end{aligned}$$

y por lo tanto

$$\rho_s(X, Y) = \frac{\theta}{3}.$$

Si $\rho_s = 0$, entonces $\theta = 0$ y $C_0 = uv = \Pi$, es decir, tenemos una pareja de variables aleatorias independientes. Por lo tanto, en este modelo de dependencia paramétrico, tenemos que las variables son independientes si y sólo si $\rho_s = 0$. Notemos además que $\rho_s \in [-\frac{1}{3}, \frac{1}{3}]$, por el teorema 2.3.7 podemos ver, entonces, que este modelo paramétrico no puede contemplar parejas que sean estrictamente monótonas la una de la otra. A estas cópulas se les conoce como la familia Farlie-Gumbel-Morgenstern. \square

Entonces tenemos ejemplos donde ρ_s no sólo da una información certera en los casos extremos, sino también para el valor cero. A pesar de esto, no es cierto que en general $\rho_s = 0$ lleve a un modelo de independencia. Para ver esto, consideremos las mismas cópulas utilizadas en los ejemplos 2.3.3, 2.3.4, 2.3.5 y 2.3.6.

Ejemplo 2.3.9. Sea (X, Y) un vector aleatorio con cópula $C_{\alpha, \beta}$ dada por

$$C_{\alpha, \beta} = \alpha M + (1 - \alpha - \beta)\Pi + \beta W,$$

donde $\alpha, \beta \geq 0$ y $\alpha + \beta \leq 1$. Entonces, por lo visto en el ejemplo 2.3.1 y el Teorema 2.3.1, tenemos que

$$\begin{aligned} Q(C_{\alpha, \beta}, \Pi) &= 4 \iint_{\mathbf{I}^2} (\alpha M + (1 - \alpha - \beta)\Pi + \beta W) d\Pi - 1 \\ &= \alpha Q(M, \Pi) + (1 - \alpha - \beta)Q(\Pi, \Pi) + \beta Q(W, \Pi) \\ &= \alpha \frac{1}{3} + (1 - \alpha - \beta)0 + \beta \left(-\frac{1}{3}\right) = \frac{\alpha - \beta}{3}, \end{aligned}$$

y entonces

$$\rho_s(X, Y) = \alpha - \beta. \quad (2.31)$$

La ecuación anterior es cero si $\alpha = \beta \in [0, \frac{1}{2}]$. En el caso $\alpha = \beta = 0$, tenemos la cópula de la independencia; para cualquier otro caso, $\alpha = \beta \in (0, \frac{1}{2}]$, tenemos cópulas, todas ellas distintas de Π , para las que obtenemos que $\rho_s = 0$ ($C_{\alpha, \alpha} = \alpha(M + W) + (1 - 2\alpha)\Pi$). Por lo tanto, tenemos una familia no numerable de cópulas para las cuales el valor cero y la independencia no tienen relación alguna. Es cierto, entonces, que en general $\rho_s = 0$ no implica independencia. \square

Veamos otros ejemplos.

Ejemplo 2.3.10. Sean U y V dos variables aleatorias uniformes $(0, 1)$, tales que $V = \frac{U}{\theta} \mathbb{I}_{(0,\theta)}(U) + \frac{1-U}{1-\theta} \mathbb{I}_{[\theta,1)}(U)$, con $\theta \in [0, 1]$. Para calcular ρ_s , por la ecuación (2.29) tenemos que

$$\begin{aligned} \iint_{\mathbf{I}^2} C_\theta(u, v) dudv &= \int_0^1 \int_0^{\theta v} u dudv + \int_0^1 \int_{\theta v}^{1-(1-\theta)v} \theta v dudv \\ &\quad + \int_0^1 \int_{1-(1-\theta)v}^1 (u+v-1) dudv \\ &= \int_0^1 \frac{(\theta v)^2}{2} dv + \int_0^1 \theta v(1-v) dv + \int_0^1 (1-\theta^2) \frac{v^2}{2} dv \\ &= \int_0^1 \theta v - \frac{2\theta-1}{2} v^2 dv \\ &= \frac{\theta+1}{6}, \end{aligned}$$

y por lo tanto

$$\rho_{s,\theta} = 2\theta - 1. \quad (2.32)$$

La ecuación (2.32) es igual a cero si $\theta = \frac{1}{2}$. Por lo tanto, el valor $\rho_s = 0$ puede ser alcanzado por variables que sean función una de la otra. \square

Veamos los otros ejemplos.

Ejemplo 2.3.11. Sean U y V dos variables aleatorias uniformes $(0, 1)$ tales que $V = U \oplus \theta$, donde $\theta \in [0, 1]$. Para calcular ρ_s utilicemos las ecuaciones (2.25) y (2.29). Entonces

$$\begin{aligned} \iint_{\mathbf{I}^2} uv dC_\theta(u, v) &= \int_0^{1-\theta} u(u+\theta) du + \int_{1-\theta}^1 u(u+\theta-1) du \\ &= \int_0^1 u(u+\theta) du - \int_{1-\theta}^1 u du \\ &= \frac{1}{3} + \frac{\theta}{2} - \frac{1}{2} + \frac{(1-\theta)^2}{2} \\ &= \frac{1}{3} - \frac{\theta(1-\theta)}{2}, \end{aligned}$$

y por lo tanto

$$\rho_{s,\theta} = 1 - 6\theta(1-\theta). \quad (2.33)$$

La ecuación anterior es cero si $\theta = \frac{1}{2} \pm \frac{1}{\sqrt{12}}$. Tenemos entonces, dos modelos donde U y V son función biyectiva la una de la otra, pero que $\rho_s = 0$. Es claro ahora, que el calcular la probabilidad de concordancia menos la de discordancia que hay entre un vector aleatorio a estudiar y un vector independiente, no nos da más información, acerca de que si su cópula es la de la independencia o no, de la que nos puede dar τ . \square

Ejemplo 2.3.12. Sean U y V dos variables aleatorias uniformes $(0, 1)$, tales que $U \oplus V = \theta$, con $\theta \in [0, 1]$. Para calcular ρ_s utilicemos ahora las ecuaciones (2.27) y (2.29). Entonces

$$\begin{aligned} \iint_{\mathbb{I}^2} uv dC_\theta(u, v) &= \int_0^\theta u(\theta - u) du + \int_\theta^1 u(\theta - u + 1) du \\ &= \int_0^1 u(\theta - u) du + \int_\theta^1 u du \\ &= \frac{\theta}{2} - \frac{1}{3} + \frac{1 - \theta^2}{2} \\ &= \frac{\theta(1 - \theta)}{2} + \frac{1}{6}, \end{aligned}$$

de donde tenemos

$$\rho_{s, \theta} = 6\theta(1 - \theta) - 1. \quad (2.34)$$

Al igual que para la ecuación (2.33), la ecuación anterior es cero si $\theta = \frac{1}{2} \pm \frac{1}{\sqrt{12}}$. Otra vez tenemos dos modelos de dependencia donde las variables son función biyectiva la una de la otra, con $\rho_s = 0$. \square

Juntando las ecuaciones (2.33) y (2.34) podemos, al igual que para τ , alcanzar todos los valores del rango de ρ_s con parejas de variables que son función biyectiva la una de la otra. Por lo tanto, toda la discusión sobre la relación $\rho_s = c$ y la dependencia, para $c \in (-1, 1)$, será la misma que se hizo para la τ . A diferencia de ésta, los ejemplos 2.3.11 y 2.3.12, no ayudan en el caso de ρ_s a entender mejor lo que ésta está calculando, ya que si consideramos una pareja (U_1, V_1) con cópula C_θ , calculada en el ejemplo 2.3.5, y una pareja (U_2, V_2) independiente, la probabilidad de concordancia no es tan sencilla de calcular como en el caso de τ , ya que, si bien la pareja (U_1, V_1) sólo puede estar en alguna de las dos rectas del soporte de su cópula (véase figura 2.6), la pareja (U_2, V_2) puede estar en cualquier punto del cuadrado unitario, por lo que para cualquier valor que tome la primer pareja, la probabilidad

de que la segunda sea concordante o discordante con ella, cambiará. Por lo tanto, veamos un par de interpretaciones, distintas a la de la concordancia, de la ρ de Spearman.

La ecuación (2.29), puede ser escrita de la siguiente forma,

$$\rho_s = 12 \iint_{\mathbf{I}^2} C(u, v) dudv - 3 = 12 \iint_{\mathbf{I}^2} (C(u, v) - uv) dudv. \quad (2.35)$$

Entonces, otra forma de interpretar ρ_s , es como el volumen escalado con signo, que hay entre la cópula de la pareja a considerar y la cópula de la independencia. Entonces, ¿qué tanta información acerca de la dependencia nos puede dar una integral así? Es una forma de resumir la información contenida en la cópula y hay que entender entonces que muchos modelos distintos de dependencia pueden tener el mismo valor para esta integral. También que el valor $\rho_s = 0$ no tiene por qué sólo darse cuando la cópula C a considerar sea Π , sino cuando el volumen que está por debajo de Π es el mismo que el que está por encima.

Cuando hablamos de la correlación, vimos que ésta no es invariante bajo transformaciones estrictamente crecientes (véase ejemplo 2.2.9). También vimos que la cópula sí es invariante bajo transformaciones estrictamente crecientes (véase Teorema 1.6.3). Una forma de corregir lo anterior es la siguiente: si X y Y son variables aleatorias con distribución conjunta H , cópula C y distribuciones marginales F y G , sabemos que $F(X)$ y $G(Y)$ son variables aleatorias uniformes $(0, 1)$ con distribución conjunta C . Por lo tanto, si consideramos, en vez de la correlación entre X y Y , la correlación entre $F(X)$ y $G(Y)$, el problema, de la variación al hacer transformaciones crecientes, se elimina. Ésta es la última interpretación de la ρ_s :

$$\begin{aligned} \rho_s(X, Y) &= 12 \iint_{\mathbf{I}^2} uv dC(u, v) - 3 = 12\mathbb{E}[UV] - 3 \\ &= \frac{\mathbb{E}[UV] - 1/4}{1/12} = \frac{\mathbb{E}[UV] - \mathbb{E}[U]\mathbb{E}[V]}{\sigma_U\sigma_V}, \end{aligned}$$

es decir,

$$\rho_s(X, Y) = \rho(F(X), G(Y)).$$

Entonces, la discusión hecha para ρ_s , se puede extender para la correlación, es decir, sobre qué información nos puede dar una integral como la de la

ecuación (2.35) acerca de la dependencia. La interpretación anterior de la ρ de Spearman, es la razón por la que ésta se conoce como la correlación de rangos.

En conclusión, una vez más los valores extremos nos resuelven el problema de la dependencia que hay entre dos variables aleatorias, no así los cercanos a cero. Si queremos medir la dependencia, lo anterior es lógico, ya que ésta se manifiesta mediante la concordancia cuando las variables son funciones monotonas entre sí. Cuando la dependencia es de otra forma, entonces se ha visto que la concordancia no basta en general.

2.4 Propiedades de dependencia

En las secciones anteriores, revisamos las tres medidas de concordancia más usuales y tratamos de aclarar lo que realmente estaban calculando. En esta sección, estudiaremos algunas propiedades que puede tener la función de distribución de una pareja de variables aleatorias, así como su relación con las medidas antes estudiadas. Empecemos con una de las propiedades más comunes, que se encuentra relacionada de manera directa con la concordancia.

2.4.1 Dependencia de cuadrante

Definición 2.4.1 (Dependencia positiva de cuadrante). Sean X y Y dos variables aleatorias continuas. Se dice que X y Y son *dependientes positivas de cuadrante* (DPC) si para todo $(x, y) \in \mathbb{R}^2$,

$$\mathbb{P}[X \leq x, Y \leq y] \geq \mathbb{P}[X \leq x] \mathbb{P}[Y \leq y], \quad (2.36)$$

o equivalentemente,

$$\mathbb{P}[X > x, Y > y] \geq \mathbb{P}[X > x] \mathbb{P}[Y > y]. \quad (2.37)$$

Veamos la equivalencia entre (2.36) y (2.37):

$$\begin{aligned} \mathbb{P}[X \leq x, Y \leq y] &\geq \mathbb{P}[X \leq x] \mathbb{P}[Y \leq y] \\ &\Leftrightarrow \\ 1 - \mathbb{P}[X > x] - \mathbb{P}[Y > y] + \mathbb{P}[X > x, Y > y] &\geq \\ &\quad (1 - \mathbb{P}[X > x])(1 - \mathbb{P}[Y > y]) \\ &\Leftrightarrow \\ \mathbb{P}[X > x, Y > y] &\geq \mathbb{P}[X > x] \mathbb{P}[Y > y]. \end{aligned}$$

Si tenemos un vector aleatorio (X, Y) que cumple ser DPC, ¿qué interpretación podemos dar a una propiedad así? En otras palabras, ¿qué tiene que ver esto con la forma de relacionarse del vector? Primero notemos que es una propiedad global, es decir, la ecuación (2.36) debe cumplirse para todo \mathbb{R}^2 . Entonces, lo que esta propiedad nos dice es que, es más probable que las variables sean pequeñas de manera conjunta a que lo sean por separado, o reescribiendo la ecuación (2.36) de la siguiente manera

$$\mathbb{P}[X \leq x | Y \leq y] \geq \mathbb{P}[X \leq x],$$

que, dado que Y es menor a un valor cualquiera, eso hace que la probabilidad de que X sea menor a otro cualquiera, sea mayor que sin el conocimiento de Y . Esta propiedad tiene que ver con la concordancia en el sentido de que si tenemos un vector aleatorio, las parejas concordantes son más probables que en el caso de ser independientes, aunque claro está, que ésta, a diferencia de la τ y de la ρ_s , no es una medida de la concordancia sino una propiedad de algunas parejas de variables aleatorias.

Consideremos dos variables aleatorias continuas X y Y con función de distribución conjunta H , con distribuciones marginales F y G , y con cópula C , entonces la ecuación (2.36) puede ser vista como

$$H(x, y) \geq F(x)G(y) \text{ para todo } (x, y) \in \mathbb{R}^2.$$

Si consideramos el Teorema de Sklar (Teorema 1.5.1) y dado que las funciones F^{-1} y G^{-1} son crecientes, obtenemos que X y Y son DPC si y sólo si

$$C(u, v) \geq uv \text{ para todo } (u, v) \in \mathbf{I}^2. \quad (2.38)$$

De la misma manera podemos definir la dependencia negativa de cuadrante (DNC), únicamente cambiando el símbolo \geq por \leq en la ecuaciones (2.36) y (2.37), y la discusión hecha posteriormente es la misma, sólo que ésta se relaciona con la discordancia. Entonces, podemos ver que X y Y son DNC si y sólo si

$$C(u, v) \leq uv \text{ para todo } (u, v) \in \mathbf{I}^2. \quad (2.39)$$

Por lo tanto, la dependencia de cuadrante es una propiedad que queda totalmente comprendida dentro de la cópula.

La dependencia por cuadrante, tanto positiva como negativa, abarca a todas aquellas cópulas, cuya gráfica queda totalmente por encima de Π y a todas aquellas cuya gráfica queda totalmente por debajo de Π . Veamos algunos ejemplos.

Ejemplo 2.4.1. Sean X y Y dos variables aleatorias con cópula C_θ dada por

$$C_\theta(u, v) = uv + \theta uv(1-u)(1-v),$$

con $\theta \in [-1, 1]$.

Notemos que si $-1 \leq \theta_1 \leq \theta_2 \leq 1$, entonces $C_{\theta_1} \leq C_{\theta_2}$; además, si $\theta = 0$, entonces $C_0 = \Pi$. Por lo tanto, si $\theta > 0$ tenemos una familia de vectores aleatorios, todos ellos DPC; si $\theta < 0$ tenemos otra familia de vectores aleatorios que son DNC.

Ejemplo 2.4.2. Sean X y Y dos variables aleatorias con cópula C_θ dada por

$$C_\theta(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)},$$

con $\theta \in [-1, 1]$.

Si $-1 \leq \theta_1 \leq \theta_2 \leq 1$, entonces $C_{\theta_1} \leq C_{\theta_2}$; por otra parte, si $\theta = 0$, $C_0 = \Pi$. Entonces, si $\theta > 0$, tenemos una familia de vectores aleatorios que son DPC; ahora, si $\theta < 0$, tenemos otra familia de vectores aleatorios con DNC.

La dependencia por cuadrante tiene, como lo explicamos, una relación con la concordancia, entonces, las medidas de las que hablamos antes, ¿nos podrán determinar si una pareja de variables aleatorias son dependientes por cuadrante? Es decir, ¿un valor positivo de τ o de ρ_s implicará que la pareja es DPC; o bien, un valor negativo implicará que son DNC? En general, no es cierto como veremos más adelante, pero primero veamos qué pasa con la τ y ρ_s , cuando el vector a considerar es DPC o DNC.

Teorema 2.4.1. *Sean X y Y dos variables aleatorias continuas con cópula C . Si X y Y son DPC, entonces*

$$3\tau(X, Y) \geq \rho_s(X, Y) \geq 0.$$

Demostración. Por la ecuación (2.38), sabemos que $C \geq \Pi$, y por lo tanto $3Q(C, C) \geq 3Q(C, \Pi) \geq 3Q(\Pi, \Pi)$, lo que completa la prueba. \square

De la misma manera, para una pareja DNC, tenemos:

Teorema 2.4.2. *Sean X y Y dos variables aleatorias continuas con cópula C . Si X y Y son DNC, entonces*

$$3\tau(X, Y) \leq \rho_s(X, Y) \leq 0.$$

Veamos que el recíproco no es cierto en general.

Ejemplo 2.4.3. Sean X y Y dos variables aleatorias con cópula

$$C(u, v) = \frac{1}{4}M(u, v) + \frac{3}{4}W(u, v).$$

Consideremos el cuadrado unitario y dividámoslo en cuatro trazando las dos diagonales principales. En el triángulo inferior, $C(u, v) = \frac{1}{4}u$, ahora, si $u \in [\frac{1}{4}, \frac{3}{4}]$, tenemos que para $v \in [0, \frac{1}{4})$, $C(u, v) = \frac{1}{4}u > uv$, mientras

que si, dentro del mismo triángulo, $v \geq \frac{1}{4}$, $C(u, v) \leq uv$. Por lo tanto, esta pareja no es, ni DPC, ni DNC. Por otro lado, por las ecuaciones (2.20) y (2.31), tenemos que $\tau(X, Y) = \rho_s(X, Y) = -\frac{1}{2}$, lo que confirma que un valor negativo de las medidas de concordancia, nada dice sobre si las variables tienen cierta relación por cuadrante.

A pesar de esto, si no pensamos a estas dos propiedades como globales, sino que lo hacemos localmente, es decir, para los puntos $(x, y) \in \mathbb{R}^2$ donde $H(x, y) - F(x)G(y) \geq 0$, X y Y son localmente DPC, mientras que para aquellos donde $H(x, y) - F(x)G(y) \leq 0$, X y Y son localmente DNC, entonces, recordando que

$$\rho_s = 12 \iint_{\mathbb{R}^2} [C(u, v) - uv] dudv,$$

podemos interpretar a ρ_s como una medida del promedio de dependencia por cuadrante (tanto positiva como negativa).

Otra medida que hemos discutido es la correlación. Ésta, al igual que la ρ_s , no implica nada acerca de la dependencia por cuadrante, mas sí sucede de manera recíproca.

Teorema 2.4.3. Sean X y Y dos variables aleatorias continuas con distribución conjunta H , y con marginales F y G , respectivamente.

1. Si X y Y son DPC, entonces $\rho(X, Y) \geq 0$.
2. Si X y Y son DNC, entonces $\rho(X, Y) \leq 0$.

Demostración. Sabemos que X y Y son DPC si $H(x, y) - F(x)G(y) \geq 0$, y por el Lema de Hoeffding (Lema 2.2.1) la prueba se sigue inmediatamente. De manera análoga se hace para el caso DNC. \square

Aunque en general, el hecho de que la correlación sea positiva no implica que la pareja sea DPC, existe un resultado que sí lo implica.

Teorema 2.4.4. Sean X y Y dos variables aleatorias continuas con distribución conjunta H , con marginales F y G , respectivamente, y con cópula C . X y Y son DPC si y sólo si $\text{Cov}[f(X), g(Y)] \geq 0$, para toda función f y g crecientes para las cuales $\mathbb{E}[f(X)]$, $\mathbb{E}[g(Y)]$ y $\mathbb{E}[f(X)g(Y)]$ existan.

Demostración. Si X y Y son DPC, entonces, $f(X)$ y $g(Y)$ también lo son, por lo que $\text{Cov}[f(X), g(Y)] \geq 0$.

Para el regreso, si consideramos $f = \mathbb{I}_{(-\infty, x]}$ y $g = \mathbb{I}_{(-\infty, y]}$, tenemos que

$$\begin{aligned} 0 &\leq \text{Cov}[f(X), g(Y)] \\ &= \mathbb{E} [\mathbb{I}_{(-\infty, x]}(X) \mathbb{I}_{(-\infty, y]}(Y)] - \mathbb{E} [\mathbb{I}_{(-\infty, x]}(X)] \mathbb{E} [\mathbb{I}_{(-\infty, y]}(Y)] \\ &= \mathbb{P}[X \leq x, Y \leq y] - \mathbb{P}[X \leq x] \mathbb{P}[Y \leq y], \end{aligned}$$

lo que completa la prueba. □

2.4.2 Monotonicidad en la cola

La dependencia positiva de cuadrante puede escribirse como

$$\mathbb{P}[Y \leq y | X \leq x] \geq \mathbb{P}[Y \leq y],$$

o como

$$\mathbb{P}[Y \leq y | X \leq x] \geq \mathbb{P}[Y \leq y | X \leq \infty].$$

Podemos hacer más fuerte esta condición pidiendo que, no sólo sea mayor que cuando condicionamos con respecto a $\{X \leq \infty\}$, sino que lo sea cada vez que condicionamos con $\{X \leq x_1\}$, para cualquier $x_1 > x$, es decir, pedir que $\mathbb{P}[Y \leq y | X \leq x]$ sea una función decreciente en x . Lo que estamos haciendo es clasificar a aquellas parejas que cumplan que cada vez que X disminuye, la probabilidad de que Y también lo haga aumenta. De manera formal definimos lo siguiente.

Definición 2.4.2. Sean X y Y dos variables aleatorias.

1. Y es *decreciente en la cola izquierda* de X [que denotamos como $\text{DCI}(Y|X)$] si $\mathbb{P}[Y \leq y | X \leq x]$ es una función decreciente en x para todo y .
2. Y es *creciente en la cola derecha* de X [que denotamos como $\text{CCD}(Y|X)$] si $\mathbb{P}[Y > y | X > x]$ es una función creciente en x para todo y .

Si X y Y tienen función de distribución H y distribuciones marginales F y G respectivamente, tenemos que la pareja es DCI($Y|X$) si $H(x, y)/F(x)$ es decreciente en x (si $F(x) > 0$) y es CCD($Y|X$) si $(1 - F(x) - G(y) + H(x, y))/(1 - F(x))$ es creciente en x . De manera similar podemos definir las propiedades de discordancia intercambiando la palabra decreciente por creciente y viceversa.

Las dos condiciones de monotonicidad implican la dependencia positiva de cuadrante, ya que si, por ejemplo, X y Y cumplen ser DCI($Y|X$), entonces

$$\mathbb{P}[Y \leq y|X \leq x] \geq \mathbb{P}[Y \leq y|X \leq \infty] = \mathbb{P}[Y \leq y],$$

y por lo tanto

$$\mathbb{P}[X \leq x, Y \leq y] = \mathbb{P}[X \leq x] \mathbb{P}[Y \leq y|X \leq x] \geq \mathbb{P}[X \leq x] \mathbb{P}[Y \leq y].$$

De manera análoga, para CCD($Y|X$) tenemos

$$\mathbb{P}[Y > y|X > x] \geq \mathbb{P}[Y > y|X > -\infty] = \mathbb{P}[Y > y],$$

y entonces

$$\mathbb{P}[X > x, Y > y] = \mathbb{P}[X > x] \mathbb{P}[Y > y|X > x] \geq \mathbb{P}[X > x] \mathbb{P}[Y > y].$$

Por lo tanto, si X y Y son dos variables que satisfacen alguna de las condiciones DCI($Y|X$), DCI($X|Y$), CCD($Y|X$) o CCD($X|Y$), entonces X y Y son dependientes positivas por cuadrante.

El recíproco no es cierto como veremos adelante, pero antes necesitamos establecer algunos criterios para estas propiedades.

Teorema 2.4.5. Sean X y Y dos variables aleatorias continuas con cópula C , entonces

1. DCI($Y|X$) si y sólo si para cualquier $v \in \mathbf{I}$, $C(u, v)/v$ es decreciente en u ,
2. CCD($Y|X$) si y sólo si para cualquier $v \in \mathbf{I}$, $[1 - u - v + C(u, v)]/(1 - u)$ es creciente en u , o equivalentemente $[v - C(u, v)]/(1 - u)$ es decreciente en u .

Demostración. Sean H , F y G definidas como siempre. $\text{DCI}(Y|X)$ si y sólo si $H(x, y)/F(x)$ es una función decreciente en x . Como las variables son continuas, tanto F y G como F^{-1} y G^{-1} son crecientes, por lo tanto, al componer $H(x, y)/F(x)$ con $F^{-1}(u)$ y $G^{-1}(v)$, ésta sigue siendo una función decreciente. Lo mismo sucede al componer $C(u, v)/u$ con $F(x)$ y $G(y)$. La prueba para CCD es similar a ésta. \square

Corolario 2.4.5.1. Sean X y Y dos variables aleatorias continuas con cópula C , entonces

1. $\text{DCI}(Y|X)$ si y sólo si para cualquier $v \in \mathbf{I}$, $\partial C(u, v)/\partial u \leq C(u, v)/u$ para u casi seguramente,
2. $\text{CCD}(Y|X)$ si y sólo si para cualquier $v \in \mathbf{I}$, $\partial C(u, v)/\partial u \geq [v - C(u, v)]/(1 - u)$ para u casi seguramente.

Demostración. Por el Teorema 2.4.5 tenemos que DCI si y sólo si $C(u, v)/u$ es decreciente en u si y sólo si $\frac{\partial}{\partial u} \left(\frac{C(u, v)}{u} \right) \leq 0$.

$$\begin{aligned} \frac{\partial}{\partial u} \left(\frac{C(u, v)}{u} \right) &\leq 0, \\ &\Leftrightarrow \\ \frac{u \partial C(u, v)/\partial u - C(u, v)}{u^2} &\leq 0, \\ &\Leftrightarrow \\ \frac{\partial C(u, v)}{\partial u} &\leq \frac{C(u, v)}{u}. \end{aligned}$$

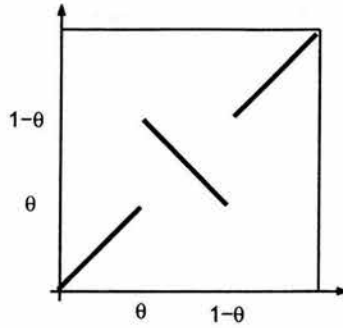
La prueba se sigue igual para el caso CCD. \square

Ahora nos encontramos en posición de dar un ejemplo de una pareja que sea DPC, pero que no sea ninguno de los cuatro posibles casos de monotonicidad.

Ejemplo 2.4.4. Sea C_θ una cópula cuyo soporte está sobre las rectas de la figura 2.8, con $\theta \in (\frac{1}{4}, \frac{1}{2})$.

Entonces

$$C_\theta(u, v) = \begin{cases} \max(\theta, u + v - 1 + \theta) & \text{si } (u, v) \in [\theta, 1 - \theta]^2, \\ M(u, v) & \text{en otro caso.} \end{cases}$$

Figura 2.8: Soporte de la cópula C_θ .

Veamos primero que es DPC. Para esto hay que probar que $C_\theta(u, v) \geq \Pi(u, v) = uv$, para todo $(u, v) \in \mathbf{I}^2$. Si $(u, v) \in [\theta, 1 - \theta]^2$, tenemos dos casos. Primero, si (u, v) está en el triángulo inferior formado por la recta $x + y = 1$, es decir, si cumplen que $u + v \leq 1$, sabemos entonces que $uv \leq \frac{1}{4}$ y como ahí $C_\theta(u, v) = \theta \in (\frac{1}{4}, \frac{1}{2})$, entonces $C_\theta \geq \Pi$. Ahora, si se encuentran en el triángulo de arriba, es decir, que cumplen que $u + v \geq 1$, tenemos lo siguiente: $1 - u + 1 - v \leq 1$, entonces $(1 - u)(1 - v) \leq \frac{1}{4}$, de donde obtenemos que

$$\begin{aligned} C_\theta(u, v) - \Pi(u, v) &= u + v - 1 + \theta - uv \\ &= \theta - (1 - u)(1 - v) \geq 0, \end{aligned}$$

es decir, $C_\theta \geq \Pi$. Ahora, para (u, v) en cualquier otro caso, sabemos que $M \geq \Pi$, y por lo tanto, C_θ cumple ser dependiente positiva de cuadrante.

Veamos ahora que no cumple ninguna de las cuatro posibles condiciones de monotonicidad en la cola. Primero notemos que C_θ es simétrica en u y en v , por lo que bastará probar sólo dos de las condiciones.

Para probar que no es DCI, consideremos u y v que cumplan que $\theta < u, v < 1 - \theta$ y $u + v > 1$, es decir que se encuentren en el triángulo formado sobre la segunda diagonal (véase figura 2.8). Ahí tenemos que

$$\frac{\partial C_\theta(u, v)}{\partial u} = 1,$$

y que

$$\frac{C_\theta(u, v)}{u} = 1 + \frac{v - (1 - \theta)}{u},$$

y como $v < 1 - \theta$, entonces $\partial C_\theta(u, v)/\partial u > C(u, v)/u$, que por el Corolario 2.4.5.1 implica que no es DCI.

Para probar que no es CCD, consideremos u y v que cumplan que $\theta < u, v < 1 - \theta$ y $u + v < 1$, es decir que se encuentren en el triángulo formado por debajo de la segunda diagonal (veáse figura 2.8). Tenemos entonces que

$$\frac{\partial C_\theta(u, v)}{\partial u} = 0,$$

y que

$$\frac{v - C_\theta(u, v)}{1 - u} = \frac{v - \theta}{1 - u},$$

y como $v > \theta$, entonces $\partial C_\theta(u, v)/\partial u < [v - C_\theta(u, v)]/(1 - u)$, y por el Corolario 2.4.5.1, tenemos que nos es CCD.

Por lo tanto, las condiciones de monotonicidad en la cola sí implican la dependencia positiva por cuadrante, mas no sucede de manera inversa. Es decir, la monotonicidad es una condición más fuerte, que la dependencia por cuadrante. \square

Por último, dada la implicación que tiene la monotonicidad sobre la dependencia de cuadrante, sabemos que si tenemos una pareja (X, Y) que cumple ser DCI($Y|X$) o CCD($Y|X$), por el teorema 2.4.1, tenemos que

$$3\tau(X, Y) \geq \rho_s(X, Y) \geq 0.$$

La pregunta es si esta nueva condición nos podrá decir un poco más acerca de τ y ρ_s . La respuesta la contiene el siguiente resultado.

Teorema 2.4.6. *Sean X y Y dos variables aleatorias continuas. Si cumplen ser DCI($Y|X$) y CCD($Y|X$), entonces $\rho_s(X, Y) \geq \tau(X, Y) \geq 0$.*

Para la prueba véase [5].

Entonces, combinando ambos resultados, tenemos que si dos variables son DCI y CCD al mismo tiempo, entonces se cumple que $3\tau \geq \rho_s \geq \tau \geq 0$.

2.4.3 Monotonicidad estocástica

Al hablar de la monotonicidad en la cola, usamos expresiones de la forma $\mathbb{P}[Y > y|X > x]$. Si ahora consideramos, en vez de esa expresión, $\mathbb{P}[Y > y|X = x]$ tenemos una forma de dependencia conocida como “monotonicidad estocástica”.

Definición 2.4.3. Sean X y Y dos variables aleatorias. Y es *estocásticamente creciente* en X [que denotaremos $EC(Y|X)$] si

$\mathbb{P}[Y > y|X = x]$ es una función creciente en x para todo y .

La propiedad de EC puede ser observada, también, a través de

$$\mathbb{P}[Y \leq y|X = x] = 1 - \mathbb{P}[Y > y|X = x],$$

considerando que dicha función sea decreciente en x para todo y .

Al igual que en los casos anteriores, la monotonicidad estocástica, es una propiedad que queda resumida en la cópula en el caso de tener variables aleatorias continuas.

Teorema 2.4.7. *Sean X y Y dos variables aleatorias continuas con cópula C . Entonces $EC(Y|X)$, si y sólo si para cualquier $v \in \mathbf{I}$, $\partial C(u, v)/\partial u$ es una función decreciente en u casi dondequiera.*

Demostración. Sean F y G las funciones de distribución de X y Y respectivamente. Sabemos que $U = F(X)$ y $V = G(Y)$ son variables uniformes $(0,1)$. Entonces $\mathbb{P}[Y \leq y|X = x] = \mathbb{P}[G(Y) \leq G(y)|F(X) = F(x)]$, es decreciente en x para todo y , si y sólo si $\mathbb{P}[V \leq v|U = u]$ es decreciente en u para todo v . Veamos que

$$\begin{aligned} \mathbb{P}[V \leq v|U = u] &= \lim_{\Delta u \rightarrow 0} \mathbb{P}[V \leq v|U \in [u, u + \Delta u]] \\ &= \lim_{\Delta u \rightarrow 0} \frac{\mathbb{P}[U \in [u, u + \Delta u], V \leq v]}{\mathbb{P}[U \in [u, u + \Delta u]]} \\ &= \lim_{\Delta u \rightarrow 0} \frac{C(u + \Delta u, v) - C(u, v)}{\Delta u} \\ &= \frac{\partial C(u, v)}{\partial u}, \end{aligned}$$

lo que completa la prueba. □

La propiedad de monotonicidad estocástica implica las de monotonicidad en la cola y por lo tanto, la de dependencia positiva por cuadrante, aunque el recíproco, en general, no es cierto como veremos más adelante.

Teorema 2.4.8. *Sean X y Y dos variables aleatorias. Si $EC(Y|X)$, entonces $DCI(Y|X)$ y $CCD(Y|X)$.*

Demostración. Sea C la cópula de X y Y . Por el Teorema 2.4.7 tenemos que $\text{EC}(Y|X)$ si y sólo si $C(u, v)$ es una función concava en u para todo v . Como C es concava, si $a, b, \lambda \in \mathbf{I}$, se cumple que $C(\lambda b + (1 - \lambda)a, v) \geq \lambda C(b, v) + (1 - \lambda)C(a, v)$, para toda v . Consideremos $0 \leq u_1 < u_2 \leq 1$, entonces

$$\begin{aligned} C(u_1, v) &= C\left(\frac{u_1}{u_2}u_2 + \left(1 - \frac{u_1}{u_2}\right)0, v\right) \\ &\geq \frac{u_1}{u_2}C(u_2, v) + \left(1 - \frac{u_1}{u_2}\right)C(0, v) \\ &= \frac{u_1}{u_2}C(u_2, v), \end{aligned}$$

lo que implica que $C(u, v)/u$ es decreciente y por el Teorema 2.4.5, tenemos que se cumple $\text{DCI}(Y|X)$.

Además,

$$\begin{aligned} C(u_2, v) &= C\left(\frac{u_2 - u_1}{1 - u_1}1 + \frac{1 - u_2}{1 - u_1}u_1, v\right) \\ &\geq \frac{u_2 - u_1}{1 - u_1}C(1, v) + \frac{1 - u_2}{1 - u_1}C(u_1, v) \\ &= \left(1 - \frac{1 - u_2}{1 - u_1}\right)v + \frac{1 - u_2}{1 - u_1}C(u_1, v), \end{aligned}$$

de donde obtenemos que

$$\frac{v - C(u_2, v)}{1 - u_2} \leq \frac{v - C(u_1, v)}{1 - u_1},$$

lo que implica que $[v - C(u, v)]/(1 - u)$ es decreciente y, de nuevo, por el Teorema 2.4.5, se cumple $\text{CCD}(Y|X)$. \square

Veamos que el recíproco en general no es cierto.

Ejemplo 2.4.5. Sean X y Y dos variables aleatorias continuas con cópula C dada por

$$C(u, v) = \begin{cases} \frac{3uv}{2} - \frac{u+v-1}{2}, & \frac{1}{3} \leq v \leq 1 - u \leq \frac{2}{3}, \\ \frac{3uv}{2}, & \frac{1}{3} \leq 1 - u \leq v \leq \frac{2}{3}, \\ M(u, v), & \text{en otro caso.} \end{cases}$$

Veamos que cumplen ser $\text{DCI}(Y|X)$, $\text{DCI}(X|Y)$, $\text{CCD}(Y|X)$ y $\text{CCD}(X|Y)$. Para eso calculemos lo siguiente

$$\frac{\partial C(u, v)}{\partial u} = \begin{cases} \frac{3v}{2} - \frac{1}{2}, & \frac{1}{3} \leq v \leq 1 - u \leq \frac{2}{3}, \\ \frac{3v}{2}, & \frac{1}{3} \leq 1 - u \leq v \leq \frac{2}{3}, \\ 1, & u < v \text{ y } u, v \in \left[\frac{1}{3}, \frac{2}{3}\right]^2, \\ 0, & \text{en otro caso,} \end{cases}$$

$$\frac{C(u, v)}{u} = \begin{cases} \frac{3v}{2} - \frac{1}{2} + \frac{1-v}{2u}, & \frac{1}{3} \leq v \leq 1 - u \leq \frac{2}{3}, \\ \frac{3v}{2}, & \frac{1}{3} \leq 1 - u \leq v \leq \frac{2}{3}, \\ 1, & u < v \text{ y } u, v \in \left[\frac{1}{3}, \frac{2}{3}\right]^2, \\ \frac{v}{u}, & \text{en otro caso,} \end{cases}$$

$$\frac{v - C(u, v)}{1 - u} = \begin{cases} \frac{3v}{2} - \frac{1}{2}, & \frac{1}{3} \leq v \leq 1 - u \leq \frac{2}{3}, \\ \frac{3v}{2} \left(\frac{2/3 - 3u/2}{1 - u} \right), & \frac{1}{3} \leq 1 - u \leq v \leq \frac{2}{3}, \\ \frac{v - u}{1 - u}, & u < v \text{ y } u, v \in \left[\frac{1}{3}, \frac{2}{3}\right]^2, \\ 0, & \text{en otro caso,} \end{cases}$$

de donde podemos ver que

$$\frac{C(u, v)}{u} \geq \frac{\partial C(u, v)}{\partial u} \geq \frac{v - C(u, v)}{1 - u},$$

y por el corolario 2.4.5.1 tenemos que se cumplen $\text{DCI}(Y|X)$ y $\text{CCD}(Y|X)$. De la misma manera se puede ver que se cumplen $\text{DCI}(X|Y)$ y $\text{CCD}(X|Y)$. Pero, $\partial C(u, v)/\partial u$ es constante en u y por lo tanto no es $\text{EC}(Y|X)$, y de igual manera sucede para $\partial C(u, v)/\partial v$ y entonces, tampoco es $\text{EC}(X|Y)$. Por lo tanto, la monotonicidad estocástica es más fuerte que la monotonicidad en la cola. \square

Hemos visto a lo largo de esta sección, que las medidas citadas antes (ρ , τ y ρ_s), no implican, de manera general ninguna de las propiedades, pero para no variar, el caso Gaussiano vuelve a funcionar de manera adecuada como veremos en el siguiente ejemplo.

Ejemplo 2.4.6. Sean X y Y dos variables aleatorias $N(0, 1)$ con distribución conjunta normal bivariada y coeficiente de correlación ρ . Veamos qué relación

guarda, aquí, la correlación y la propiedad de monotonicidad estocástica. Para eso calculemos la densidad condicional de $Y|X$.

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f(x, y)}{f_X(x)} \\ &= \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 + y^2 - 2\rho xy)\right)}{\frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(\frac{x^2}{2}\right)} \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right), \end{aligned}$$

y por lo tanto $Y|X \sim N(\rho x, 1 - \rho^2)$. Tenemos entonces que

$$\mathbb{P}[Y \leq y|X = x] = \Phi\left(\frac{y - \rho x}{\sqrt{1 - \rho^2}}\right),$$

la cual es una función decreciente en x si y sólo si $\rho > 0$ y es creciente si y sólo si $\rho < 0$. Por lo tanto, en el caso normal bivariado, una vez más, la correlación nos da toda la información que necesitamos en cuanto a la dependencia, en este caso, refiriéndose a las propiedades de EC y, por lo tanto, DCI, CCD y DPC. \square

En conclusión, las tres propiedades están relacionadas con formas de entender la concordancia. A pesar de esto, en ningún caso, τ ó ρ_s nos daban valores que caracterizaran a dichas propiedades, lo que al revés sí sucede. El problema es, que τ y ρ_s se proponen como medidas de concordancia, sin embargo, ninguna logra captar ninguna de estas propiedades, es decir, sin más estudio, un valor positivo de τ ó ρ_s , no nos puede hablar de una propiedad global como éstas.

2.5 Medidas de dependencia

Al analizar a las medidas de concordancia, con todas fuimos muy precisos al indicar que no tienen, en general, un valor que caracterice a la independencia. Vimos que en los casos donde las variables aleatorias en cuestión, estaban relacionadas por una función monótona, las medidas de concordancia lo captaban perfectamente, e incluso indicaban cómo es esta función (creciente o decreciente). Por lo tanto, si estamos interesados en evaluar una condición así, las medidas de concordancia funcionan a la perfección; pero si estamos interesados en la independencia, serán propuestas otras medidas. De esto se encargará esta sección.

Sean X y Y dos variables aleatorias con distribución conjunta H y distribuciones marginales F y G , respectivamente. Sabemos que X y Y son independientes si y sólo si $H(x, y) = F(x)G(y)$ para todo $x, y \in \mathbb{R}$. La forma de generar medidas que nos caractericen a la independencia, será comparando la distribución conjunta (H) con la distribución en el caso de independencia (FG), entonces, cualquier distancia L_p entre H y FG nos servirá, es decir,

$$\left(\iint_{\mathbb{R}^2} |H(x, y) - F(x)G(y)|^p dx dy \right)^{\frac{1}{p}},$$

con $p \in [1, \infty]$. Dado que $|H - FG| \geq 0$, la ecuación anterior es cero si y sólo si $H \stackrel{c.s.}{=} FG$, es decir, es cero si y sólo si X y Y son independientes, con $p \in [1, \infty]$. Son de nuestro particular interés los valores $p = 1$ y $p = \infty$. Por lo tanto, si tenemos X y Y dos variables aleatorias continuas, definimos las siguientes medidas:

$$\sigma(X, Y) = 12 \iint_{\mathbb{R}^2} |H(x, y) - F(x)G(y)| dx dy; \quad (2.40)$$

$$\delta(X, Y) = \sup_{x, y \in \mathbb{R}} |H(x, y) - F(x)G(y)|. \quad (2.41)$$

Si C es la cópula asociada a X y Y , haciendo la transformación $F(x) = u$, $G(y) = v$, obtenemos

$$\sigma(X, Y) = 12 \iint_{\mathbf{I}^2} |C(u, v) - uv| du dv; \quad (2.42)$$

$$\delta(X, Y) = \sup_{u, v \in \mathbf{I}} |C(u, v) - uv|. \quad (2.43)$$

Empecemos analizando las propiedades de σ .

Teorema 2.5.1. *Sean X y Y dos variables aleatorias continuas con cópula C . Entonces se cumple que:*

1. $0 \leq \sigma(X, Y) \leq 1$,
2. $\sigma(X, Y) = 0$ si y sólo si X y Y son independientes;
3. $\sigma(X, Y) = 1$ si y sólo si $X \stackrel{c.s.}{=} \alpha(Y)$ con α estrictamente monótona;
4. si α y β son funciones estrictamente monótonas c.s. en el rango de X y Y , respectivamente, entonces $\sigma(\alpha(X), \beta(Y)) = \sigma(X, Y)$.

Demostración. Para la prueba de la siguiente desigualdad

$$\iint_{\mathbf{I}^2} |C(u, v) - uv| dudv \leq \frac{1}{12}, \quad (2.44)$$

véase [17]. A manera de bosquejo veamos lo siguiente. Sabemos que si $C = M$ ó $C = W$, se cumple la igualdad, como se vio en la prueba del teorema 2.3.7, el problema podría ser que hubiera una cópula C que en la parte que $C > \Pi$, coincidiera con M , y en la parte en que $C < \Pi$ cumpliera que su “distancia” entre ella y Π sea mayor que la que hay entre M y Π o que el problema ocurriera de manera inversa pero para W . Para ver que esto no es posible, veamos qué pasa con $M - \Pi$ y $\Pi - W$. Si dividimos el cuadrado unitario en cuatro trazando sus dos diagonales principales y numeramos los cuadrantes a favor del reloj empezando con el cuadrante de hasta abajo, tenemos que

cuadrante	$M(u, v) - uv$	$uv - W(u, v)$
1	$(1 - u)v$	uv
2	$u(1 - v)$	uv
3	$u(1 - v)$	$(1 - u)(1 - v)$
4	$(1 - u)v$	$(1 - u)(1 - v)$

podemos ver que para el primer cuadrante, tenemos que $M - \Pi > \Pi - W$ si $u < \frac{1}{2}$ y $M - \Pi < \Pi - W$ si $u > \frac{1}{2}$, y de la misma manera podemos analizar qué sucede en cada uno de estos cuadrantes. De esta manera, si

ahora dividimos el cuadrado unitario en cuatro trazando las rectas $x = \frac{1}{2}$ y $y = \frac{1}{2}$ y numeramos a favor del reloj empezando por el cuadrante inferior izquierdo, tenemos que en 1 y 3 $M - \Pi > \Pi - W$, mientras que en 2 y 4 $M - \Pi < \Pi - W$. Sabemos que para cualquier cópula C se cumple que $W \leq C \leq M$, entonces, para que $\iint |C - \Pi| dudv > \frac{1}{12}$, tendría que coincidir con M en los cuadrantes 1 y 3, y en los cuadrantes 2 y 4 estar más abajo de Π de lo que estaría M al reflejarla sobre Π , o lo mismo pero con W . Ninguna cópula C distinta de M puede cumplir eso, ya que si $C = M$ en los cuadrantes 1 y 3, entonces $C = M$ por el Lema 1.6.5, y algo similar pasa en el caso de W . Por lo tanto, (2.44) se cumple y la igualdad se da sólo si $C = M$ ó $C = W$. Entonces se cumple que $\sigma \leq 1$.

Para la segunda parte, basta considerar el Teorema 1.6.2 y la ecuación (2.42). Por los Teoremas 1.6.5 y 1.6.6, tenemos que X y Y son función estrictamente monótona la una de la otra, si y sólo si $C = M$ ó $C = W$, además como ya dijimos, la igualdad en (2.44) se da si y sólo si $C = M$ ó $C = W$, lo que prueba la tercera parte. Para la última parte, recordemos que el Teorema 1.6.4 nos dice que, por ejemplo, si α es creciente y β decreciente, tenemos que $C_{\alpha(X)\beta(Y)}(u, v) = u - C_{XY}(u, 1 - v)$ y por lo tanto $C_{\alpha(X)\beta(Y)}(u, v) - \Pi(u, v) = \Pi(u, 1 - v) - C_{XY}(u, 1 - v)$, y de la misma manera para las distintas combinaciones de α y β . \square

Por la interpretación para ρ_s vista en la ecuación (2.35), es inmediato preguntarse la relación que ésta tendrá con σ . En general, no existe tal relación, pero, por ejemplo, si tenemos X y Y dos variables aleatorias continuas que cumplen ser DPC, por la ecuación (2.38) tenemos que $\sigma(X, Y) = \rho_s(X, Y)$, mientras que si son DNC, por la ecuación (2.39) tenemos que $\sigma(X, Y) = -\rho_s(X, Y)$. Por ejemplo, si X y Y se distribuyen normales estandar bivariadas con coeficiente de correlación ρ , entonces por lo visto en el ejemplo 2.4.6, tenemos que $\sigma(X, Y) = |\rho_s(X, Y)| = \frac{6}{\pi} \arcsin\left(\frac{|\rho|}{2}\right)$, ya que \arcsin es una función impar. Si X y Y no son ni DPC, ni DNC, entonces la cópula asociada a ellas, no será, en general, ni mayor ni menor que Π . Veamos qué pasa con las parejas que vimos en los ejemplos 2.3.4 y 2.3.10.

Ejemplo 2.5.1. Sean U y V dos variables aleatorias uniformes $(0, 1)$, tales que $V = \frac{U}{\theta} \mathbb{I}_{(0, \theta)}(U) + \frac{1-U}{1-\theta} \mathbb{I}_{[\theta, 1)}(U)$, con $\theta \in [0, 1]$. La cópula que calculamos

en el ejemplo 2.3.4, también la podemos ver como

$$C_\theta(u, v) = \begin{cases} M(u, v), & 0 \leq u \leq \theta v \leq \theta, \\ \theta v, & 0 \leq \theta v < u < 1 - (1 - \theta)v, \\ W(u, v), & \theta \leq 1 - (1 - \theta)v \leq u \leq 1. \end{cases}$$

Para calcular σ_θ , sólo resta ver cómo es C_θ con respecto a Π en el triángulo inferior formado por el soporte (véase figura 2.5). Para eso veamos que si $u \leq \theta$, entonces $\theta v \geq uv$, mientras que si $u > \theta$, sucede que $\theta v < uv$. Con esto, procedamos a calcular σ_θ

$$\begin{aligned} \iint_{\mathbf{I}^2} |C_\theta - uv| dudv &= \int_0^1 \int_0^{\theta v} (u - uv) dudv + \int_0^1 \int_{\theta v}^\theta \theta(v - uv) dudv \\ &\quad + \int_0^1 \int_\theta^{1-(1-\theta)v} (uv - \theta v) dudv \\ &\quad + \int_0^1 \int_{1-(1-\theta)v}^1 (uv - u - v + 1) dudv \\ &= \int_0^1 \frac{\theta^2(1-v)v^2}{2} dv + \int_0^1 \frac{\theta^2 v(1-v)^2}{2} dv \\ &\quad + \int_0^1 \frac{(1-\theta)^2 v(1-v)^2}{2} dv + \int_0^1 \frac{(1-\theta)^2 v(1-v)^2}{2} dv \\ &= \frac{\theta^2 + (1-\theta)^2}{2} \left(\int_0^1 (v - v^2) dv \right) \\ &= \frac{\theta^2 + (1-\theta)^2}{12}, \end{aligned}$$

y por la ecuación (2.42) tenemos que

$$\sigma_\theta = 1 - 2\theta + 2\theta^2. \quad (2.45)$$

Si $\theta = 0$ ó $\theta = 1$, la ecuación anterior vale 1, lo que por el Teorema 2.5.1, implica que U y V son funciones monótonas la una de la otra. Por otro lado, si $\theta = 1$, tenemos que $C_1 = M$, mientras que si $\theta = 0$, tenemos que $C_0 = W$, lo que confirma el comentario anterior. Un valor 1 de σ , nos reporta únicamente que las variables son función monótona la una de la otra, pero no nos dice si son crecientes o decrecientes entre sí, a diferencia de lo que τ ó ρ_s reportan. Pero ¿qué pasa con el valor 0? En el Teorema 2.5.1 ya lo

estudiamos, pero recordemos que para estas parejas cuando $\theta = \frac{1}{2}$, tenemos que $\tau_{1/2} = \rho_{s,1/2} = 0$, lo que podría engañarnos acerca de la independencia. Por otro lado, $\sigma_{1/2} = \frac{1}{2}$, que además resulta ser el mínimo de la función σ_θ , es decir, el valor $\sigma_\theta = 0$ no puede ser alcanzado, ya que esta cópula pertenece a una pareja que nunca podrá ser independiente. Entonces, en resumen tenemos que σ nos reporta los casos extremos mas no nos dice de qué tipo son, mientras que caracteriza a la independencia con un único valor. \square

En la ecuación (2.41) definimos otra medida. Veamos sus propiedades.

Teorema 2.5.2. *Sean X y Y dos variables aleatorias continuas con cópula C . Entonces se cumple que:*

1. $0 \leq \delta(X, Y) \leq \frac{1}{4}$,
2. $\delta(X, Y) = 0$ si y sólo si X y Y son independientes;
3. si α y β son funciones estrictamente monótonas c.s. en el rango de X y Y , respectivamente, entonces $\delta(\alpha(X), \beta(Y)) = \delta(X, Y)$.

Demostración. Para probar que $\frac{1}{4}$ es cota superior, consideremos $A \in \sigma(X)$ y $B \in \sigma(Y)$, donde $\sigma(\cdot)$ representa la σ -álgebra generada por la variable aleatoria. Sabemos que $AB \subseteq A, B^4$, de donde obtenemos las siguientes ecuaciones

$$\begin{aligned} \mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B] &\leq \mathbb{P}[A] - \mathbb{P}[A]\mathbb{P}[B] \\ &= \mathbb{P}[A]\mathbb{P}[B^c], \end{aligned}$$

y de la misma manera

$$\mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B] \leq \mathbb{P}[A^c]\mathbb{P}[B],$$

y entonces

$$\begin{aligned} \mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B] &\leq \min(\mathbb{P}[A]\mathbb{P}[B^c], \mathbb{P}[A^c]\mathbb{P}[B]) \\ &\leq \mathbb{P}[A]\mathbb{P}[A^c] \\ &\leq \frac{1}{4}. \end{aligned}$$

⁴ $AB = A \cap B$

Como $\{X \leq x\} \in \sigma(X)$ y $\{Y \leq y\} \in \sigma(Y)$ para todo $x, y \in \mathbb{R}$, entonces $\frac{1}{4}$ es cota superior para $\delta(X, Y)$. Para ver que puede ser alcanzada basta considerar a la cópula M , y al evaluar tenemos que

$$M\left(\frac{1}{2}, \frac{1}{2}\right) - \Pi\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{4},$$

y al ser la cota superior obliga a que el supremo sea $\frac{1}{4}$.

El resto de la prueba se sigue igual que la del Teorema 2.5.1. \square

Notemos que a diferencia de lo que pasa con σ , para δ no tenemos que el valor más grande caracteriza a las parejas de variables que son función monótona la una de la otra, aunque más adelante veremos qué tipo de variables alcanzan dicha cota. Veamos un ejemplo.

Ejemplo 2.5.2. Sean U y V dos variables aleatorias uniformes $(0, 1)$ tales que $V = U \oplus \theta$, donde $\theta \in [0, 1]$, es decir, la pareja introducida en el ejemplo 2.3.5. Notemos que si $\theta = \frac{1}{2}$, la cópula de dicha pareja está dada por

$$C_{\frac{1}{2}}(u, v) = \begin{cases} \min\left(u, v - \frac{1}{2}\right) & \text{si } (u, v) \in \left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right], \\ \min\left(u - \frac{1}{2}, v\right) & \text{si } (u, v) \in \left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right], \\ W(u, v) & \text{en otro caso} \end{cases}$$

Si consideramos el punto $\left(\frac{1}{2}, \frac{1}{2}\right)$, tenemos que

$$C_{\frac{1}{2}}\left(\frac{1}{2}, \frac{1}{2}\right) - \Pi\left(\frac{1}{2}, \frac{1}{2}\right) = -\frac{1}{4},$$

y por lo tanto, $\delta(U, V) = \frac{1}{4}$, donde la pareja (U, V) no son función monótona la una de la otra.

Si ahora consideramos a la pareja introducida en el ejemplo 2.3.6, es decir, U y V variables aleatorias uniformes $(0, 1)$, que cumplen que $U \oplus V = \theta$, con $\theta \in [0, 1]$, tenemos un caso similar si $\theta = \frac{1}{2}$, ya que

$$C_{\frac{1}{2}}(u, v) = \begin{cases} \max\left(0, u + v - \frac{1}{2}\right) & \text{si } (u, v) \in \left[0, \frac{1}{2}\right] \times \left[0, \frac{1}{2}\right], \\ \max\left(\frac{1}{2}, u + v - 1\right) & \text{si } (u, v) \in \left[\frac{1}{2}, 1\right] \times \left[\frac{1}{2}, 1\right], \\ M(u, v) & \text{en otro caso} \end{cases}$$

y al considerar el punto $\left(\frac{1}{2}, \frac{1}{2}\right)$ otra vez, tenemos ahora que

$$C_{\frac{1}{2}}\left(\frac{1}{2}, \frac{1}{2}\right) - \Pi\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{4},$$

y por lo tanto, $\delta(U, V) = \frac{1}{4}$, donde, otra vez, la pareja (U, V) no son función monótona la una de la otra. Entonces, el valor $\delta = \frac{1}{4}$, no es exclusivo para las cópulas M y W como sucede en el caso de σ .

Notemos que tanto las dos cópulas de este ejemplo, como M y W tienen una característica en común: si dividimos el cuadrado unitario en cuatro partiéndolo con las rectas $x = \frac{1}{2}$ y $y = \frac{1}{2}$ y numeramos los cuadrantes en contra del reloj empezando por el que contiene al origen, notemos que, o bien tienen toda su masa en los cuadrantes 1 y 3, o bien la tienen en los cuadrantes 2 y 4. Nos podemos preguntar entonces si ¿ésta será la forma general de todas las cópulas que alcanzan el valor extremo de δ ? Veamos el siguiente teorema \square

Teorema 2.5.3. Sean X y Y variables aleatorias continuas con cópula C . Entonces, $\delta(X, Y) = \frac{1}{4}$ si y sólo si

$$V_C(B_1) = 0, \text{ donde } B_1 = \left\{ \left[0, \frac{1}{2}\right] \times \left[0, \frac{1}{2}\right] \right\} \cup \left\{ \left[\frac{1}{2}, 1\right] \times \left[\frac{1}{2}, 1\right] \right\};$$

o bien

$$V_C(B_2) = 0, \text{ donde } B_2 = \left\{ \left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right] \right\} \cup \left\{ \left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right] \right\}.$$

Demostración. Supongamos que $V_C(B_1) = 0$. Entonces $C\left(\frac{1}{2}, \frac{1}{2}\right) = 0$ lo que implica que $\delta = \frac{1}{4}$, ya que alcanza la cota superior en el punto $\left(\frac{1}{2}, \frac{1}{2}\right)$. Si $V_C(B_2) = 0$, tenemos que $C\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2}$, lo que implica otra vez que $\delta = \frac{1}{4}$.

Supongamos que $\delta = \frac{1}{4}$. Para la demostración veamos primero que el supremo no puede ser alcanzado en ningún punto distinto de $\left(\frac{1}{2}, \frac{1}{2}\right)$ y luego veamos que esa condición es equivalente a que el volumen de las regiones (B_1 ó B_2) sea cero.

Supongamos que el supremo lo alcanza en el punto $(a, b) \in (0, 1) \times (0, 1)$. Para que C alcance la cota tiene que cumplir que $C(a, b) = ab \pm \frac{1}{4}$. Dividamos el cuadrado unitario por las rectas $x = a$ y $y = b$ y numeremos contra el sentido del reloj empezando por el cuadrante que contiene al origen. Supongamos que $C(a, b) = ab - \frac{1}{4}$, entonces el cuadrante 1 tiene volumen $ab - \frac{1}{4}$ y el 3 tiene volumen $(1-a)(1-b) - \frac{1}{4}$, y como C es cópula se tiene que cumplir que

$$ab - \frac{1}{4} \geq 0$$

y

$$(1-a)(1-b) - \frac{1}{4} \geq 0,$$

que es equivalente a

$$\frac{1}{4b} \leq a \leq 1 - \frac{1}{4(1-b)}.$$

La desigualdad de los extremos se cumple si y sólo si

$$-(2b-1)^2 \geq 0,$$

lo que implica que $b = \frac{1}{2}$ y por lo tanto $a = \frac{1}{2}$. En este caso implica que el cuadrante 1 y 3 tienen volumen cero.

Supongamos que $C(a, b) = ab + \frac{1}{4}$. El cuadrante 2 tiene volumen $b(1-a) - \frac{1}{4}$ y el 4 $a(1-b) - \frac{1}{4}$. Como C es cópula, se debe cumplir que

$$b(1-a) - \frac{1}{4} \geq 0$$

y

$$a(1-b) - \frac{1}{4} \geq 0$$

que es equivalente a

$$\frac{1}{4(1-b)} \leq a \leq 1 - \frac{1}{4b}.$$

Considerando únicamente la desigualdad para b , eso sucede si sólo si

$$-(2b-1)^2 \geq 0,$$

lo que implica otra vez que el supremo se alcance en $(\frac{1}{2}, \frac{1}{2})$. En este caso, el volumen del cuadrante 2 y 4 es cero, lo que completa la prueba. \square

Lo que el teorema anterior nos dice es que el valor extremo de δ caracteriza a las variables cuya cópula tiene toda su masa en el primer y tercer cuadrante del cuadrado unitario o en el segundo y cuarto cuadrante.

Sigamos analizando esta medida con más ejemplos. Veamos qué pasa para las variables del ejemplo 2.3.4 y δ .

Ejemplo 2.5.3. Sean U y V dos variables aleatorias uniformes $(0, 1)$, que cumplen que $V = \frac{U}{\theta} \mathbb{I}_{(0, \theta)}(U) + \frac{1-U}{1-\theta} \mathbb{I}_{[\theta, 1)}(U)$, con $\theta \in [0, 1]$. En el ejemplo 2.5.1 vimos que si $u \in [0, \theta]$, $C_\theta(u, v) \geq uv$, donde C_θ es la cópula de la pareja (U, V) , calculada en el ejemplo 2.3.4. Entonces, si $u \leq \theta$, $C_\theta(u, v) - uv$, con v fija, será máximo si $v = \frac{u}{\theta}$. Es fácil ver que el máximo de la función $C_\theta(\theta v, v) - \theta v^2 = \theta(v - v^2)$ estará en el punto $\frac{1}{2}$ y por lo tanto

$$\sup_{0 < u < \theta, 0 < v < 1} |C_\theta(u, v) - uv| = \frac{\theta}{4}.$$

Si $u \in (\theta, 1]$, $C_\theta(u, v) \leq uv$, la función $uv - C_\theta(u, v)$, para v fija, tiene un máximo si $v = \frac{1-u}{1-\theta}$. Mediante cálculo elemental, podemos ver que la función $(1 - (1 - \theta)v)v - C_\theta(1 - (1 - \theta)v, v) = (1 - \theta)(v - v^2)$ tiene su máximo en el punto $\frac{1}{2}$, y por lo tanto

$$\sup_{\theta < u < 1, 0 < v < 1} |uv - C_\theta(u, v)| = \frac{1 - \theta}{4}.$$

Por lo tanto

$$\delta_\theta = \frac{\max(\theta, 1 - \theta)}{4}.$$

Veamos que $\delta_0 = \delta_1 = \frac{1}{4}$, que era de esperarse ya que $C_0 = W$ y $C_1 = M$. Recordemos que para esta cópula, tenemos que si $\theta = \frac{1}{2}$, entonces $\tau_{1/2} = \rho_{s, 1/2} = 0$, mientras que $\delta_{1/2} = \frac{1}{8}$. \square

Veamos otro ejemplo.

Ejemplo 2.5.4. Sean U y V dos variables aleatorias independientes uniformes $(0, 1)$. Para $0 < p < 1$, definamos $X = U$ y $Y = U \mathbb{I}_{(0, p]}(U) + V \mathbb{I}_{(p, 1)}(U)$. Queremos calcular $\delta(X, Y)$: Antes de empezar, analicemos lo que debe pasar conforme p se acerca a los extremos: si p se acerca al cero, tenemos que Y depende cada vez menos de X y por lo tanto δ se tendrá que ir a cero; si p se acerca a uno, Y se hace más parecida a X y por lo tanto δ se irá a un cuarto. Dada la forma en que se relacionan las variables, conforme p sea mayor, X y Y se harán más dependientes por lo que δ debe de ser una función creciente en p . Para calcular, utilicemos la definición vista en (2.41). Lo primero que necesitamos calcular es H , la función de distribución de X y Y .

$$\begin{aligned} H(x, y) &= \mathbb{P}[X \leq x, Y \leq y] \\ &= \mathbb{P}[X \leq x, Y \leq y, U \leq p] + \mathbb{P}[X \leq x, Y \leq y, U > p] \\ &= \mathbb{P}[U \leq x, U \leq y, U \leq p] + \mathbb{P}[U \leq x, V \leq y, U > p], \end{aligned}$$

si $x \leq p$, tenemos que

$$H(x, y) = \min(x, y) + 0 = \min(x, y),$$

mientras que si $x > p$, tenemos que

$$H(x, y) = \min(y, p) + y(x - p),$$

y juntando ambas ecuaciones tenemos

$$H(x, y) = \min(x, y)\mathbb{I}_{(0,p]}(x) + (\min(y, p) + y(x - p))\mathbb{I}_{(p,1)}(x).$$

Para las distribuciones marginales tenemos

$$F(x) = \mathbb{P}[X \leq x] = x\mathbb{I}_{(0,1)}(x) + \mathbb{I}_{[1,\infty)}(x),$$

y

$$\begin{aligned} G(y) &= \mathbb{P}[Y \leq y] \\ &= \mathbb{P}[Y \leq y, U \leq p] + \mathbb{P}[Y \leq y, U > p] \\ &= \mathbb{P}[U \leq y, U \leq p] + \mathbb{P}[V \leq y, U > p], \end{aligned}$$

de donde

$$G(y) = (\min(y, p) + y(1 - p))\mathbb{I}_{(0,1)}(y) + \mathbb{I}_{[1,\infty)}(y).$$

Ahora para obtener δ calculemos

$$\mu := \sup_{0 < x \leq p, 0 < y < 1} |\min(x, y) - x \min(y, p) - xy(1 - p)|$$

y

$$\nu := \sup_{p < x < 1, 0 < y < 1} |\min(y, p) + (x - p)y - x \min(y, p) - xy(1 - p)|,$$

y entonces $\delta(X, Y) = \max(\mu, \nu)$.

Para calcular μ consideremos varios casos. Primero, si $y > p$, tenemos

$$\min(x, y) - x \min(y, p) - xy(1 - p) = x(1 - y)(1 - p).$$

Esta función es creciente en x y decreciente en y , por lo que es máxima si $x = y = p$ y el supremo es $p(1 - p)^2$.

Si $x < y < p$, vemos que

$$\min(x, y) - x \min(y, p) - xy(1 - p) = x(1 - y(2 - p)).$$

Esta función en valor absoluto es creciente para x , por lo que su máximo está en $x = y$. La función $y - y^2(2 - p)$ tiene un punto crítico en $y = \frac{1}{2(2-p)}$, lo que en valor absoluto lleva al máximo relativo $\frac{1}{4(2-p)}$. Además, al evaluar en los extremos del intervalo $(0 y p)$ obtenemos otro posible supremo dado por $p(1 - p)^2$.

El caso $y < x < p$ es similar al anterior y se obtienen los mismos candidatos a supremo.

Procedamos a calcular ν también por casos. Si $y > p$, entonces

$$\min(y, p) + (x - p)y - x \min(y, p) - xy(1 - p) = (1 - x)(1 - y)p,$$

que es una función decreciente, tanto para x como para y , y entonces el supremo se alcanza cuando $x = y = p$ y es $p(1 - p)^2$.

Si $y < p$, tenemos

$$\min(y, p) + (x - p)y - x \min(y, p) - xy(1 - p) = (1 - x)y(1 - p),$$

que es una función decreciente para x y creciente para y ; el supremo se alcanza si $x = y = p$ y es $p(1 - p)^2$.

De primera instancia, podemos pensar que $\delta(X, Y) = \max\left(p(1 - p)^2, \frac{1}{4(2-p)}\right)$.

Ahora, $p(1 - p)^2 > \frac{1}{4(2-p)}$ si y sólo si $-4p^4 + 16p^3 - 20p^2 + 8p - 1 > 0$, lo cual no sucede nunca, ya que dicho polinomio se factoriza como $-(p - r_1)^2(p - r_2)^2$, con $r_1 = 1 - \frac{1}{\sqrt{2}}$ y $r_2 = 1 + \frac{1}{\sqrt{2}}$, y entonces δ sería $\frac{1}{4(2-p)}$ siempre. Según lo que analizamos al principio, si $p = 0$ entonces $\delta = 0$, lo que no pasaría en este caso. El problema está en lo siguiente: cuando calculamos a $\frac{1}{4(2-p)}$ como posible supremo, hicimos el supuesto de que $y < p$, entonces, si el valor $\frac{1}{2(2-p)}$, que es donde se alcanza el supremo, no cumple ser menor que p , no podrá ser supremo, lo que nos lleva a evaluar la condición $\frac{1}{2(2-p)} < p$, que es lo mismo que evaluar la condición $-2p^2 + 4p - 1 > 0$, lo que sucede si $r_1 < p < r_2$, y por lo tanto, tenemos que

$$\delta(X, Y) = \begin{cases} p(1 - p)^2 & \text{si } 0 < p \leq 1 - \frac{1}{\sqrt{2}}, \\ \frac{1}{4(2-p)} & \text{si } 1 - \frac{1}{\sqrt{2}} < p < 1, \end{cases}$$

lo que confirma lo que analizamos al principio de este ejemplo. \square

En conclusión, estas medidas cumplen su principal propósito que es el de identificar a la independencia, aunque pierden la cualidad de caracterizar a las parejas estrictamente monótonas. En el caso de δ su valor extremo lo comparten otras formas de relacionarse, lo que nos indica que ésta mide una dependencia de cuadrante. δ será de nuestro particular interés ya que existe una forma sencilla de proponer un estimador de ella, cosa que haremos más adelante.

2.6 Reordenamientos de M

En esta sección estudiaremos los reordenamientos de M . Un reordenamiento de M , como su nombre lo dice, consiste en particionar el cuadrado unitario en un número finito de franjas verticales y reordenar éstas (véase Figura 2.9). Por ejemplo, la cópula definida en 2.4.4 es un reordenamiento de M . Al sólo estar moviendo el soporte de M , las cópulas generadas siguen representando modelos de probabilidad para parejas de variables uniformes $(0,1)$ que son función biyectiva la una de la otra. La importancia que estas cópulas tienen, es que dada cualquier cópula C , existen reordenamientos de M , C_ϵ , que cumplen que

$$\sup_{u,v \in \mathbf{I}} |C(u,v) - C_\epsilon(u,v)| < \epsilon.$$

Es decir, podemos aproximarnos a cualquier forma de dependencia, tanto como queramos, por parejas de variables que son función la una de la otra. En especial si $C = \Pi$, lo que estamos diciendo es que la independencia puede ser aproximada por formas de dependencia absoluta. El resultado lo probaremos sólo para Π , puesto que es la cópula que nos interesa aproximar.

Teorema 2.6.1. *Para toda $\epsilon > 0$, existe un reordenamiento de M , que denotaremos por Π_ϵ , tal que*

$$\sup_{u,v \in \mathbf{I}} |\Pi(u,v) - \Pi_\epsilon(u,v)| < \epsilon.$$

Demostración. Sea $m \in \mathbb{N}$, tal que $m \geq \frac{4}{\epsilon}$. Por el Teorema 1.4.2, tenemos que para toda cópula C y $u, v, s, t \in \mathbf{I}$,

$$|C(u,v) - C(s,t)| < \frac{\epsilon}{2}, \text{ si } |u-s| < \frac{1}{m} \text{ y } |v-t| < \frac{1}{m}. \quad (2.46)$$

El entero m determina Π_ϵ de la siguiente manera. Sea $n = m^2$ y sea $\{J_n\}$ una partición regular de \mathbf{I} en n pedazos iguales. Sea π un reordenamiento de la partición dado por $\pi(m(j-1) + k) = m(k-1) + j$, para $k, j = 1, \dots, m$. El efecto de esta permutación es el de redistribuir la masa de probabilidad de M de tal manera que cada una de los n cuadros generados por la partición, tengan masa $\frac{1}{n}$ (véase Figura 2.9). Como

$$V_{\Pi_\epsilon} \left(\left[0, \frac{p}{m}\right] \times \left[0, \frac{q}{m}\right] \right) = V_{\Pi} \left(\left[0, \frac{p}{m}\right] \times \left[0, \frac{q}{m}\right] \right),$$

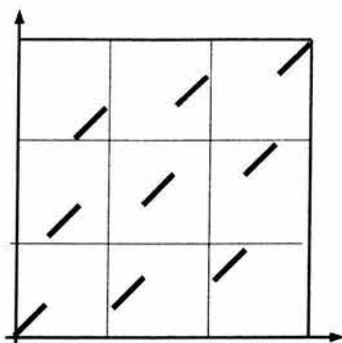


Figura 2.9: Soporte del reordenamiento de M con $m = 3$ y $\pi = (1,4,7,2,5,8,3,6,9)$.

para $p, q = 0, 1, \dots, m$, tenemos que para esos puntos

$$\Pi_{\epsilon} \left(\frac{p}{m}, \frac{q}{m} \right) = \Pi \left(\frac{p}{m}, \frac{q}{m} \right).$$

Sea $(u, v) \in \mathbf{I}^2$, entonces, existen $p, q \in \{0, 1, \dots, m\}$, tales que $|u - \frac{p}{m}| < \frac{1}{m}$ y $|v - \frac{q}{m}| < \frac{1}{m}$. Entonces

$$\begin{aligned} |\Pi(u, v) - \Pi_{\epsilon}(u, v)| &\leq \left| \Pi(u, v) - \Pi \left(\frac{p}{m}, \frac{q}{m} \right) \right| \\ &\quad + \left| \Pi \left(\frac{p}{m}, \frac{q}{m} \right) - \Pi_{\epsilon} \left(\frac{p}{m}, \frac{q}{m} \right) \right| \\ &\quad + \left| \Pi_{\epsilon} \left(\frac{p}{m}, \frac{q}{m} \right) - \Pi_{\epsilon}(u, v) \right| \\ &< \frac{\epsilon}{2} + 0 + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

por (2.46), lo que completa la prueba. \square

Supongamos que X_{ϵ} y Y_{ϵ} son dos variables aleatorias con cópula Π_{ϵ} . Por el teorema anterior y la ecuación (2.42), considerando $\frac{\epsilon}{12}$, tenemos que

$$\sigma(X_{\epsilon/12}, Y_{\epsilon/12}) = 12 \iint_{\mathbf{I}^2} |\Pi_{\epsilon/12}(u, v) - \Pi(u, v)| dudv < \epsilon,$$

y de igual manera, por la ecuación (2.43)

$$\delta(X_{\epsilon}, Y_{\epsilon}) = \sup_{u, v \in \mathbf{I}} |\Pi_{\epsilon}(u, v) - \Pi(u, v)| < \epsilon.$$

Entonces, podemos aproximarnos a cero tanto como nosotros queramos. Esto nos dice que un valor cualquiera en el rango de estas medidas de dependencia puede no decir de ninguna manera lo que nosotros pensemos. En otras palabras, si pensamos que un valor muy pequeño en estas medidas implica que las variables dependen muy poco entre sí, estaremos cometiendo un error, como nos muestra este caso. En conclusión, los valores cercanos a cero sólo implican que la función de distribución conjunta es muy parecida a la de la independencia, pero eso de ninguna manera implica, en general, que las parejas cada vez dependan menos.

Capítulo 3

Condiciones de “mixing”

3.1 Introducción

En el capítulo anterior trabajamos la dependencia que puede haber entre dos variables aleatorias. Pero, ¿qué pasa si queremos extender el problema a más variables, en especial, si queremos estudiar la independencia entre ellas? Si tenemos un proceso $\{X_t\}_{t \in \mathbb{Z}}$ donde las variables no son todas independientes, puede llegar a haber independencia conforme vayamos considerando parejas cada vez más lejanas, según sea el proceso. Por ejemplo, si tenemos $\{Z_t\}_{t \in \mathbb{Z}}$ variables aleatorias independientes e idénticamente distribuidas, y definimos

$$X_t = \sum_{k=0}^m g_k Z_{t-k},$$

donde g_k son constantes, es claro que las variables X_t y X_{t+1} no son independientes, sin embargo, las variables X_t y X_{t+m+1} sí lo son puesto que heredan la independencia de las variables Z_k . El proceso anterior se le conoce como $MA(m)$, por sus siglas en inglés (moving average de orden m). Este es un caso particular de lo que se conoce como la m -dependencia, donde si dos variables “distan” más de m , entonces son independientes. De todas formas, daremos más adelante una definición formal de esto. Entonces, en el caso $MA(m)$, es clara la independencia al alejarnos en el “tiempo”, pero si definimos X_t como

$$X_t = \sum_{k=0}^{\infty} g_k Z_{t-k},$$

no es tan claro cómo será la dependencia de las variables X_t entre sí, a menos que $g_k = 0$ a partir de alguna n . A este proceso se le conoce como $MA(\infty)$. O por ejemplo, si definimos

$$X_t = \theta X_{t-1} + Z_t,$$

tampoco es claro cómo será la dependencia entre dos de ellas. Este proceso se conoce como $AR(1)$, por sus siglas en inglés (autoregression de orden 1). En resumen, en este capítulo trataremos de dar una noción de independencia asintótica, es decir, llevar la m -dependencia al límite, así como tratar de ver ejemplos de cómo es ésta.

En la Sección 3.2 definiremos y estudiaremos las condiciones de mixing que nos servirán para estudiar en la Sección 3.3 la dependencia en los procesos estocásticos.

La bibliografía para este capítulo fue: para la Sección 3.2, [3] y [10], mientras que para la Sección 3.3 se utilizó [3], [4], [6] y [13].

3.2 Medidas de “mixing”

Definición 3.2.1. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Para cualesquiera dos sub σ -álgebras, \mathcal{A} y \mathcal{B} , de \mathcal{F} , definimos las siguientes medidas de dependencia:

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{\substack{A \in \mathcal{A}, \\ B \in \mathcal{B}}} |\mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B]|. \quad (3.1)$$

$$\phi(\mathcal{A}, \mathcal{B}) := \sup_{\substack{A \in \mathcal{A}, \mathbb{P}[A] > 0, \\ B \in \mathcal{B}}} |\mathbb{P}[B|A] - \mathbb{P}[B]|. \quad (3.2)$$

$$\psi(\mathcal{A}, \mathcal{B}) := \sup_{\substack{A \in \mathcal{A}, \mathbb{P}[A] > 0, \\ B \in \mathcal{B}, \mathbb{P}[B] > 0}} \frac{|\mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B]|}{\mathbb{P}[A]\mathbb{P}[B]}. \quad (3.3)$$

Lo primero que nos podemos cuestionar sobre estas medidas es qué valores toman. Es claro que las tres se encuentran acotadas por debajo por 0, pero la cota superior para las tres será distinta. Para eso, veamos el siguiente teorema.

Teorema 3.2.1. *Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, y sean \mathcal{A} y \mathcal{B} dos sub σ -álgebras de \mathcal{F} . Entonces, las medidas definidas en (3.1), (3.2) y (3.3) cumplen:*

$$\alpha(\mathcal{A}, \mathcal{B}) \leq \frac{1}{4}; \quad \phi(\mathcal{A}, \mathcal{B}) \leq 1; \quad \psi(\mathcal{A}, \mathcal{B}) \leq \infty.$$

Demostración. Empecemos por probar que $\frac{1}{4}$ acota a α . Sea $A \in \mathcal{A}$ y $B \in \mathcal{B}$. Sabemos que $AB \subseteq A, B$, de donde obtenemos las siguientes ecuaciones

$$\begin{aligned} \mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B] &\leq \mathbb{P}[A] - \mathbb{P}[A]\mathbb{P}[B] \\ &= \mathbb{P}[A]\mathbb{P}[B^c], \end{aligned}$$

y de la misma manera

$$\mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B] \leq \mathbb{P}[A^c]\mathbb{P}[B],$$

y entonces

$$\begin{aligned} \mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B] &\leq \min(\mathbb{P}[A]\mathbb{P}[B^c], \mathbb{P}[A^c]\mathbb{P}[B]) \\ &\leq \mathbb{P}[A]\mathbb{P}[A^c] \\ &\leq \frac{1}{4}. \end{aligned}$$

Para ver que la cota puede ser alcanzada, consideremos $\mathcal{A} = \mathcal{B} = \{\Omega, \emptyset, A, A^c\}$ y $\mathbb{P}[A] = \frac{1}{2}$. Entonces, si consideramos $A = B$, tenemos que $|\mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B]| = \frac{1}{4}$. Por lo tanto $\alpha \leq \frac{1}{4}$.

Para ϕ , es claro que 1 es una cota superior de dicho conjunto, veamos que puede ser alcanzada. Consideremos el espacio $([0, 1], \mathbb{B}([0, 1]), \mathbb{P})$, donde \mathbb{B} representa a los borelianos y \mathbb{P} es la medida uniforme en el intervalo. Sean $\mathcal{A} = \mathcal{B} = \mathbb{B}([0, 1])$. Si definimos la sucesión de conjuntos $A_n = B_n = [0, \frac{1}{n}]$, tenemos que

$$\mathbb{P}[B_n|A_n] - \mathbb{P}[B_n] = 1 - \frac{1}{n},$$

y entonces

$$\sup_{n \in \mathbb{N}} |\mathbb{P}[B_n|A_n] - \mathbb{P}[B_n]| = 1,$$

por otro lado

$$\sup_{n \in \mathbb{N}} |\mathbb{P}[B_n|A_n] - \mathbb{P}[B_n]| \leq \sup_{\substack{A \in \mathcal{A}, \mathbb{P}[A] > 0, \\ B \in \mathcal{B}}} |\mathbb{P}[B|A] - \mathbb{P}[B]|,$$

de donde concluimos que $\phi \leq 1$.

Por último, para ψ consideremos el mismo caso que para ϕ , de donde tenemos

$$\begin{aligned} \frac{\mathbb{P}[A_n B_n] - \mathbb{P}[A_n] \mathbb{P}[B_n]}{\mathbb{P}[A_n] \mathbb{P}[B_n]} &= \frac{\mathbb{P}[A_n] - \mathbb{P}[A_n]^2}{\mathbb{P}[A_n]^2} \\ &= n - 1, \end{aligned}$$

de donde

$$\sup_{n \in \mathbb{N}} \frac{\mathbb{P}[A_n B_n] - \mathbb{P}[A_n] \mathbb{P}[B_n]}{\mathbb{P}[A_n] \mathbb{P}[B_n]} = \infty,$$

y concluimos igual que para ϕ , y por lo tanto $\psi \leq \infty$. \square

Estas medidas, al igual que σ y δ , están hechas con la idea de identificar a la independencia. Notemos que $\alpha = \phi = \psi = 0$ si y sólo si tenemos independencia entre las dos sub σ -álgebras (véase Sección 1.2).

Dado que $0 \leq \mathbb{P}[\cdot] \leq 1$, es claro que $\alpha \leq \phi \leq \psi$, aunque esta cota puede ser mejorada por

$$2\alpha(\mathcal{A}, \mathcal{B}) \leq \phi(\mathcal{A}, \mathcal{B}) \leq \frac{1}{2}\psi(\mathcal{A}, \mathcal{B}). \quad (3.4)$$

La igualdad se obtiene considerando $\mathcal{A} = \mathcal{B} = \{\Omega, \emptyset, A, A^c\}$ y $\mathbb{P}[A] = \frac{1}{2}$. Para este caso, vemos que $\alpha = \frac{1}{4}$ y por lo tanto $2\alpha = \frac{1}{2}$. Para ϕ , veamos que $\mathbb{P}[A|A] - \mathbb{P}[A] = \frac{1}{2}$. Y por último, $(\mathbb{P}[CD] - \mathbb{P}[C] \mathbb{P}[D]) / (\mathbb{P}[C] \mathbb{P}[D]) = 1$ para todo $C, D \in \mathcal{A}$ con probabilidad positiva y por lo tanto $\frac{1}{2}\psi = \frac{1}{2}$.

Si tenemos dos variables aleatorias X y Y , la medida α la trabajamos sobre $\sigma(X)$ y $\sigma(Y)$, las σ -álgebras generadas por X y Y , es decir,

$$\alpha(\sigma(X), \sigma(Y)) = \sup_{A, B \in \mathbb{B}(\mathbb{R})} |\mathbb{P}[X \in A, Y \in B] - \mathbb{P}[X \in A] \mathbb{P}[Y \in B]|.$$

Por otro lado

$$\delta(X, Y) = \sup_{x, y \in \mathbb{R}} |\mathbb{P}[X \leq x, Y \leq y] - \mathbb{P}[X \leq x] \mathbb{P}[Y \leq y]|.$$

Es claro que α acota a δ , ya que el supremo en el caso de α se toma sobre más conjuntos, pero ¿no será α tan sólo una pretensión de hacer una medida más elegante y más complicada? Es decir, ¿existirán casos donde la desigualdad sea estricta? Para contestar esta pregunta veamos el siguiente ejemplo.

Ejemplo 3.2.1. Sea X una variable aleatoria continua y simétrica con función de distribución F y sea $Y = X^2$. Sea $a > 0$ tal que $\mathbb{P}[|X| \geq a] = \frac{1}{2}$ y definamos $A = \{|X| < a\}$ y $B = A^c = \{|X| \geq a\}$. Notemos que $B = \{|X| \geq a\} = \{Y \geq a^2\}$, es decir, $B \in \sigma(Y)$, mientras que $A \in \sigma(X)$. Veamos además que $\mathbb{P}[AB] = 0$ y $\mathbb{P}[A]\mathbb{P}[B] = \frac{1}{4}$. Por lo tanto, $\alpha(\sigma(X), \sigma(Y)) = \frac{1}{4}$. Ahora, para obtener $\delta(X, Y)$ calculemos para $y \geq 0$

$$\begin{aligned} H(x, y) &= \mathbb{P}[X \leq x, Y \leq y] \\ &= \mathbb{P}[X \leq x, -\sqrt{y} \leq Y \leq \sqrt{y}] \\ &= \mathbb{P}[-\sqrt{y} \leq X \leq \min(x, \sqrt{y})] \mathbb{I}_{(x \geq -\sqrt{y})} \\ &= (F(\min(x, \sqrt{y})) + F(\sqrt{y}) - 1) \mathbb{I}_{(x \geq -\sqrt{y})}. \end{aligned}$$

Veamos además que

$$\begin{aligned} G(y) &= \mathbb{P}[Y \leq y] = \mathbb{P}[-\sqrt{y} \leq X \leq \sqrt{y}] \\ &= 2F(\sqrt{y}) - 1. \end{aligned}$$

Analicemos por casos para calcular $\delta(X, Y)$.

Si $x \geq \sqrt{y}$, entonces

$$\begin{aligned} |2F(\sqrt{y}) - 1 - F(x)(2F(\sqrt{y}) - 1)| &= (2F(\sqrt{y}) - 1)(1 - F(x)) \\ &\leq (2F(\sqrt{y}) - 1)(1 - F(\sqrt{y})). \end{aligned}$$

Si $-\sqrt{y} \leq x < \sqrt{y}$, tenemos

$$\begin{aligned} |F(x) + F(\sqrt{y}) - 1 - F(x)(2F(\sqrt{y}) - 1)| &= |(2F(x) - 1)(1 - F(\sqrt{y}))| \\ &\leq (2F(\sqrt{y}) - 1)(1 - F(\sqrt{y})). \end{aligned}$$

Por último, si $x < -\sqrt{y}$, tenemos que

$$\begin{aligned} |-F(x)(2F(\sqrt{y}) - 1)| &\leq F(-\sqrt{y})(1 - F(\sqrt{y})) \\ &= (2F(\sqrt{y}) - 1)(1 - F(\sqrt{y})). \end{aligned}$$

Por lo tanto

$$\delta(X, Y) = \sup_{y \geq 0} (2F(\sqrt{y}) - 1)(1 - F(\sqrt{y})) = \sup_{\frac{1}{2} \leq u \leq 1} (2u - 1)(1 - u),$$

que se alcanza para $u = \frac{3}{4}$, lo que lleva a

$$\delta(X, Y) = \frac{1}{8} < \frac{1}{4} = \alpha(\sigma(X), \sigma(Y)).$$

El supremo se toma para $u \geq \frac{1}{2}$, ya que $F(y) \geq \frac{1}{2}$ para $y \geq 0$. □

3.3 Dependencia en procesos estocásticos

Para una familia $\{X_s, s \in S\}$ de variables aleatorias, con S un conjunto de índices, sea $\sigma(X_s, s \in S)$ la mínima σ -álgebra generada por ellas. Supongamos que $\{X_t, t \in \mathbb{Z}\}$ es una sucesión de variables aleatorias. Para $-\infty \leq J \leq L \leq \infty$, definimos $\mathcal{F}_J^L := \sigma(X_k, J \leq k \leq L)$.

Decimos que las variables son m -dependientes, si $\mathcal{F}_{-\infty}^J$ y $\mathcal{F}_{J+n}^{\infty}$ son independientes (véase Sección 1.2) para toda $n > m$ y $J \in \mathbb{Z}$. Es decir, si consideramos las variables (X_j, \dots, X_{j+k}) y $(X_{j+k+n}, \dots, X_{j+k+n+l})$, éstas son independientes, siempre y cuando $n > m$.

Ejemplo 3.3.1. Sean $\{Y_t\}_{t \in \mathbb{Z}}$ variables aleatorias independientes e idénticamente distribuidas y sea $X_t = f(Y_t, Y_{t-1}, \dots, Y_{t-m})$ otra variable aleatoria. Entonces, la sucesión de variables $\{X_t\}_{t \in \mathbb{Z}}$ es m -dependiente. Esta clase incluye a las variables que son MA(m). □

En el caso anterior es claro por qué las variables son independientes al alejarse, pero para entenderlo en modelos un poco más complicados, necesitamos definir algunos conceptos. Para toda $n \in \mathbb{N}$ definimos (de acuerdo a las ecuaciones (3.1), (3.2) y (3.3))

$$\alpha(n) := \sup_{J \in \mathbb{Z}} \left\{ \alpha(\mathcal{F}_{-\infty}^J, \mathcal{F}_{J+n}^{\infty}) \right\}, \quad (3.5)$$

$$\phi(n) := \sup_{J \in \mathbb{Z}} \left\{ \phi(\mathcal{F}_{-\infty}^J, \mathcal{F}_{J+n}^{\infty}) \right\}, \quad (3.6)$$

$$\psi(n) := \sup_{J \in \mathbb{Z}} \{ \psi(\mathcal{F}_{-\infty}^J, \mathcal{F}_{J+n}^\infty) \}, \quad (3.7)$$

A partir de lo anterior, las condiciones de “mixing” para una sucesión $\{X_t, t \in \mathbb{Z}\}$, se definen por

$$(X_t)_{t \in \mathbb{Z}} \text{ es strong-mixing o } \alpha\text{-mixing si } \lim_{n \rightarrow \infty} \alpha(n) = 0, \quad (3.8)$$

$$(X_t)_{t \in \mathbb{Z}} \text{ es } \phi\text{-mixing si } \lim_{n \rightarrow \infty} \phi(n) = 0, \quad (3.9)$$

$$(X_t)_{t \in \mathbb{Z}} \text{ es } \psi\text{-mixing si } \lim_{n \rightarrow \infty} \psi(n) = 0, \quad (3.10)$$

Antes de seguir hagamos un par de comentarios:

- Para definir la condición de α -mixing para una sucesión sólo sobre los naturales, la definición de $\alpha(n)$, la modificamos por

$$\alpha(n) := \sup_{J \in \mathbb{N}} \{ \alpha(\mathcal{F}_1^J, \mathcal{F}_{J+n}^\infty) \}.$$

- Para definir la condición de α -mixing para una sucesión estrictamente estacionaria¹, definimos $\alpha(n)$ como

$$\alpha(n) := \alpha(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty).$$

- Los puntos anteriores se siguen igual para el caso de ϕ y ψ -mixing.

Por el Teorema 3.2.1, lo primero que observamos es que si una sucesión es ψ -mixing, entonces es ϕ -mixing; si es ϕ -mixing, entonces es α -mixing, es decir

$$\psi\text{-mixing} \Rightarrow \phi\text{-mixing} \Rightarrow \alpha\text{-mixing}. \quad (3.11)$$

Antes de analizar estas implicaciones, notemos que con las definiciones anteriores, la m -dependencia es extendida al caso límite. Si una sucesión de variables aleatorias cumple con ser m -dependiente, entonces cumple con las condiciones (3.8), (3.9) y (3.10). La pregunta es, ¿existirán sucesiones que cumplan ser α , ϕ ó ψ -mixing, y que no sean m -dependientes? La respuesta es sí. Para demostrarlo, veamos el siguiente ejemplo, que además nos dirá un poco sobre cómo es un proceso de éste tipo.

¹Decimos que $\{X_t\}_{t \in \mathbb{Z}}$ es estrictamente estacionaria si para todo $q \in \mathbb{Z}$ y $n \in \mathbb{N}$, $(X_t, \dots, X_{t+n}) \stackrel{D}{=} (X_{t+q}, \dots, X_{t+q+n})$.

Ejemplo 3.3.2. Sea $\{g_k\}_{k \in \mathbb{N}}$ una sucesión no creciente y no negativa tal que $g_1 \leq \frac{1}{4}$. Lo que haremos es construir un proceso estocástico en el que las parejas no independientes serán escogidas convenientemente de tal manera que al ir midiendo la dependencia, conforme nos vayamos alejando entre las σ -álgebras, ésta vaya comportándose de acuerdo a la sucesión. Consideremos $\{X_n\}_{n \in \mathbb{N}}$ variables aleatorias que cumplen que $\mathbb{P}[X_n = 1] = 1 - \mathbb{P}[X_n = 0] = h$, y que para la pareja (X_{n^2}, X_{n^2+n}) , $\mathbb{P}[X_{n^2} = 1, X_{n^2+n} = 1] = c_n$ donde c_n la elijeremos más adelante. Ahora, la distribución conjunta de las X_n está dada si suponemos que $(X_1, X_2), \dots, (X_{n^2}, X_{n^2+n}), \dots, (X_3), (X_5), \dots$, son independientes entre sí. Notemos que las distribuciones conjuntas están bien definidas puesto que las variables de las parejas no independientes sólo aparecen una vez, es decir, en una sola pareja, ya que para que estén en dos parejas distintas tendría que haber un número natural n que pudiera ser representado por dos números naturales distintos, k y m , por k^2 y por $m^2 + m$, pero esto es imposible ya que la ecuación cuadrática para m , $m^2 + m = k^2$, no tiene raíces enteras. Una condición extra que tenemos que pedir para que la conjunta de la pareja (X_{n^2}, X_{n^2+n}) esté bien definida es que $c_n \leq h$.

Procedamos a calcular $\alpha(k) = \sup_{J \in \mathbb{N}} \{\alpha(\mathcal{F}_1^J, \mathcal{F}_{J+k}^\infty)\}$. Para que se cumpla que $\alpha(\mathcal{F}_1^J, \mathcal{F}_{J+k}^\infty) > 0$, se necesita que una pareja de la forma (X_{n^2}, X_{n^2+n}) quede separada en cada una de las σ -álgebras para que éstas no sean independientes. Entonces, si $J < k^2$, tenemos que $\alpha(\mathcal{F}_1^J, \mathcal{F}_{J+k}^\infty) = 0$.

Para $J \geq k^2$, valdrá lo mismo, siempre y cuando la pareja separada sea la misma. Es decir, la pareja $(X_{(k+m)^2}, X_{(k+m)^2+k+m})$, $m = 0, 1, \dots$, hará que el valor de $\alpha(\mathcal{F}_1^J, \mathcal{F}_{J+k}^\infty)$ sea positivo, si se cumple que

$$(k+m)^2 \leq J \leq (k+m)^2 + k+m$$

y

$$(k+m)^2 \leq J+k \leq (k+m)^2 + k+m,$$

que se traduce en

$$(k+m)^2 \leq J \leq (k+m)^2 + m.$$

Por lo tanto, si $(k+m)^2 \leq J \leq (k+m)^2 + k+m$, $m = 0, 1, \dots$, no se cumple, $\alpha(\mathcal{F}_{-\infty}^J, \mathcal{F}_{J+k}^\infty) = 0$.

Consideremos una J que cumpla la desigualdad anterior, entonces, si $l^2 = (k+m)^2$, se cumple que $\alpha(\mathcal{F}_1^J, \mathcal{F}_{J+k}^\infty) = \alpha(\mathcal{F}_1^{l^2}, \mathcal{F}_{l^2+l}^\infty)$. Procedamos a calcular.

Tomemos dos eventos $B \in \mathcal{F}_1^{l^2}$ y $C \in \mathcal{F}_{l^2+l}^\infty$ y definamos $D = \{X_{l^2} = 1\}$ y $E = \{X_{l^2+l} = 1\}$. Al evento B lo podemos ver como

$$B = B_1 D \cup B_2 D^c$$

donde B_1 y B_2 pertenecen a $\mathcal{F}_1^{l^2-1}$. Entonces

$$\begin{aligned} \mathbb{P}[B] &= \mathbb{P}[B_1 D] + \mathbb{P}[B_2 D^c] \\ &= \mathbb{P}[B_1] h + \mathbb{P}[B_2] (1 - h) \end{aligned}$$

Si elegimos $h \in (0, \frac{1}{2}]$ tenemos que

$$\mathbb{P}[B] \geq (\mathbb{P}[B_1] + \mathbb{P}[B_2])h. \quad (3.12)$$

De igual forma tenemos

$$C = C_1 E \cup C_2 E^c$$

donde C_1 y C_2 pertenecen a $\mathcal{F}_{l^2+l+1}^\infty$. Y de manera similar obtenemos que

$$\mathbb{P}[C] \geq (\mathbb{P}[C_1] + \mathbb{P}[C_2])h. \quad (3.13)$$

Utilizando la independencia entre los eventos y la representación de B y C escribimos

$$\begin{aligned} \mathbb{P}[BC] - \mathbb{P}[B]\mathbb{P}[C] &= \mathbb{P}[B_1 C_1] (\mathbb{P}[DE] - \mathbb{P}[D]\mathbb{P}[E]) \\ &\quad + \mathbb{P}[B_1 C_2] (\mathbb{P}[DE^c] - \mathbb{P}[D]\mathbb{P}[E^c]) \\ &\quad + \mathbb{P}[B_2 C_1] (\mathbb{P}[D^c E] - \mathbb{P}[D^c]\mathbb{P}[E]) \\ &\quad + \mathbb{P}[B_2 C_2] (\mathbb{P}[D^c E^c] - \mathbb{P}[D^c]\mathbb{P}[E^c]) \\ &= (\mathbb{P}[B_1 C_1] - \mathbb{P}[B_1 C_2] - \mathbb{P}[B_2 C_1] + \mathbb{P}[B_2 C_2]) (c_l - h^2) \\ &= (\mathbb{P}[B_1] - \mathbb{P}[B_2]) (\mathbb{P}[C_1] - \mathbb{P}[C_2]) (c_l - h^2) \end{aligned} \quad (3.14)$$

Por lo tanto, $|\mathbb{P}[BC] - \mathbb{P}[B]\mathbb{P}[C]| \leq c_l - h^2$. Si elegimos $B = D$ y $C = E$ obtenemos la igualdad. Definamos $c_l = g_l + h^2$. Entonces, por lo discutido antes, $\alpha(k) = \sup_{m=0,1,\dots} \{\alpha(\mathcal{F}_1^{(k+m)^2}, \mathcal{F}_{(k+m)^2+k+m}^\infty)\}$, y como g_{k+m} es decreciente, tenemos que $\alpha(k) = g_k$. Por lo tanto, formamos un proceso estocástico cuyas $\alpha(k)$ están dadas por una sucesión cualquiera que cumpla las condiciones establecidas.

De la ecuación (3.14), obtenemos que

$$\mathbb{P}[C|B] - \mathbb{P}[C] = \frac{(\mathbb{P}[B_1] - \mathbb{P}[B_2])(\mathbb{P}[C_1] - PC_2)}{\mathbb{P}[B]} g_l.$$

Multiplicando y dividiendo por $\mathbb{P}[B_1] + \mathbb{P}[B_2]$ y utilizando la desigualdad (3.12), concluimos que $|\mathbb{P}[C|B] - \mathbb{P}[C]| \leq \frac{g_l}{h}$. Además, si definimos $B = D$ y $C = E$ obtenemos la igualdad y por lo argumentado antes, $\phi(k) = \frac{g_k}{h}$.

Haciendo un procedimiento análogo, tenemos que

$$\frac{\mathbb{P}[BC] - \mathbb{P}[B]\mathbb{P}[C]}{\mathbb{P}[B]\mathbb{P}[C]} = \frac{(\mathbb{P}[B_1] - \mathbb{P}[B_2])(\mathbb{P}[C_1] - PC_2)}{\mathbb{P}[B]\mathbb{P}[C]} g_l.$$

Ahora, si multiplicamos y dividimos por $(\mathbb{P}[B_1] + \mathbb{P}[B_2])(\mathbb{P}[C_1] + \mathbb{P}[C_2])$, por las ecuaciones (3.12) y (3.13), tenemos que

$$\frac{\mathbb{P}[BC] - \mathbb{P}[B]\mathbb{P}[C]}{\mathbb{P}[B]\mathbb{P}[C]} \leq \frac{g_l}{h^2},$$

y la igualdad se obtiene en el mismo caso. Entonces, $\psi(k) = \frac{g_k}{h^2}$.

Así que si queremos encontrar un proceso que sea α , ϕ ó ψ -mixing, basta elegir una sucesión g_k que converja a cero. \square

En el ejemplo anterior, la sucesión que propusimos cumple las tres condiciones. Entonces, ¿habrá realmente diferencia entre ellas? En otras palabras, ¿podremos encontrar sucesiones que cumplan ser una, pero no la otra? Veamos el siguiente ejemplo.

Ejemplo 3.3.3. Consideremos el mismo proceso definido en el ejemplo 3.3.2. Notemos que los cálculos realizados para $\alpha(\mathcal{F}_1^{l^2}, \mathcal{F}_{l^2+l}^\infty)$, no dependen de que h sea una constante o no, es decir, lo mismo obtendremos si en vez de considerar $\mathbb{P}[X_{l^2} = 1, X_{l^2+l} = 1] = h$, consideramos $\mathbb{P}[X_{l^2} = 1, X_{l^2+l} = 1] = h_l$, siempre y cuando para toda l se cumpla que $h_l \leq \frac{1}{2}$ y $g_l \leq h_l(1 - h_l)$. En ese caso, es claro que $\alpha(k) = g_k$, pero para ϕ y ψ , recordemos que $\phi(\mathcal{F}_1^{l^2}, \mathcal{F}_{l^2+l}^\infty) = \frac{g_l}{h_l}$ y $\psi(\mathcal{F}_1^{l^2}, \mathcal{F}_{l^2+l}^\infty) = \frac{g_l}{h_l^2}$, lo que implica que

$$\phi(k) = \sup_{l \geq k} \{ \phi(\mathcal{F}_1^{l^2}, \mathcal{F}_{l^2+l}^\infty) \} = \sup_{l \geq k} \left\{ \frac{g_l}{h_l} \right\},$$

y

$$\psi(k) = \sup_{l \geq k} \left\{ \psi(\mathcal{F}_1^{l^2}, \mathcal{F}_{l^2+l}^\infty) \right\} = \sup_{l \geq k} \left\{ \frac{g_l}{h_l^2} \right\},$$

no tienen por qué alcanzar el supremo en $l = k$.

Para construir una sucesión de variables que cumpla ser α -mixing, pero que no cumpla ser ϕ -mixing, consideremos la estructura propuesta en este ejemplo para el proceso estocástico del ejemplo 3.3.2, con las sucesiones $g_k = \frac{1}{4k}$ y $h_k = \frac{1}{2k}$. Probemos que cumplen las condiciones para ambas sucesiones: g_k debe cumplir ser decreciente y menor a $\frac{1}{4}$, lo que es claro que cumple; h_k es menor que $\frac{1}{2}$, y además $\frac{1}{4k} \leq \frac{1}{2k} \left(1 - \frac{1}{2k}\right)$ si y sólo si $k \geq 1$. Por lo tanto, las sucesiones cumplen las condiciones necesarias, y en ese caso

$$\alpha(k) = g_k = \frac{1}{4k} \xrightarrow{k \rightarrow \infty} 0$$

lo que implica que $\{X_n\}$ es α -mixing. Por otro lado,

$$\phi(k) = \sup_{l \geq k} \left\{ \frac{g_l}{h_l} \right\} = \sup_{l \geq k} \left\{ \frac{\frac{1}{4l}}{\frac{1}{2l}} \right\} = \frac{1}{2},$$

de donde vemos que $\{X_n\}$ no es ϕ -mixing. Sabemos que, entonces, tampoco será ψ -mixing, lo que se comprueba al ver que

$$\psi(k) = \sup_{l \geq k} \left\{ \frac{g_l}{h_l^2} \right\} = \sup_{l \geq k} \{l\} = \infty.$$

De la misma manera, podemos pensar en un proceso que sea ϕ -mixing, y por lo tanto α -mixing, pero que no sea ψ -mixing. Para eso, proponemos $g_k = \frac{1}{4k^2}$ y $h_k = \frac{1}{2k}$, es claro que cumplen las condiciones necesarias, y en este caso tenemos

$$\phi(k) = \sup_{l \geq k} \left\{ \frac{1}{2l} \right\} = \frac{1}{2k} \xrightarrow{k \rightarrow \infty} 0.$$

Por otro lado

$$\psi(k) = \sup_{l \geq k} \{1\} = 1,$$

lo que implica que no es ψ -mixing. Por lo tanto, la implicación en (3.11) no puede sustituirse por un si y sólo si y tiene sentido definir las distintas condiciones. \square

En el caso de las cadenas de Markov, éstas tiene una forma muy peculiar de dependencia o de independencia condicional. En ese caso tenemos que el cálculo de las $\alpha(n)$ es mucho más sencillo, como podemos ver en el siguiente teorema.

Teorema 3.3.1. *Sea $\{X_t\}_{t \in \mathbb{Z}}$ un cadena de Markov estrictamente estacionaria. Entonces, para toda $n \in \mathbb{N}$*

1. $\alpha(n) = \alpha(\sigma(X_0), \sigma(X_n))$,
2. $\phi(n) = \phi(\sigma(X_0), \sigma(X_n))$,
3. $\psi(n) = \psi(\sigma(X_0), \sigma(X_n))$.

Para la prueba véase [2].

Veamos un ejemplo.

Ejemplo 3.3.4. Sea $\{X_t\}_{t \in \mathbb{Z}}$ un proceso de Markov con matriz de transición $P(i, j)$, tal que $P(i, 1) = P(i, i+1) = \frac{1}{2}$ y distribución estacionaria $\pi(j) = \frac{1}{2^j} \mathbb{1}_{\mathbb{N}}(j)$. Queremos utilizar este proceso para mostrar que si un proceso es ϕ -mixing, no necesariamente lo es si cambiamos el sentido del tiempo. Probemos entonces, que el proceso anterior es ϕ -mixing en el sentido definido, pero no lo es si revertimos el tiempo. Para eso, dado que el proceso es de Markov y estacionario, sabemos que

$$\phi(n) = \phi(\sigma(X_0), \sigma(X_n)).$$

No es complicado ver que la matriz de transición en n pasos está dada por

$$P^{(n)}(i, j) = \begin{cases} \frac{1}{2^j}, & j \leq n, \\ \frac{1}{2^n}, & j = i + n, \\ 0, & \text{en otro caso.} \end{cases}$$

Entonces

$$|P^{(n)}(i, j) - \pi(j)| = \begin{cases} 0, & j \leq n, \\ \frac{1}{2^n} - \frac{1}{2^j}, & j = i + n, \\ \frac{1}{2^j}, & \text{en otro caso.} \end{cases}$$

Sea $A \in \sigma(X_n)$, entonces

$$|P^{(n)}(i, A) - \pi(A)| = \left| \sum_{j \in A} [P^{(n)}(i, j) - \pi(j)] \right|,$$

donde tenemos varios casos para A . Primero, si $A = \{i + n\}$, entonces

$$|P^{(n)}(i, A) - \pi(A)| = \frac{1}{2^n} - \frac{1}{2^j} < \frac{1}{2^n};$$

si $A = \{n+1, n+2, \dots\} \setminus \{i+n\}$, entonces

$$\begin{aligned} |P^{(n)}(i, A) - \pi(A)| &= \left| -\sum_{j \in A} \frac{1}{2^j} \right| \\ &\leq \sum_{k=1}^{\infty} \frac{1}{2^{n+k}} = \frac{1}{2^n}, \end{aligned}$$

que son los conjuntos donde puede tomar el valor más grande, por lo tanto

$$|P^{(n)}(i, A) - \pi(A)| \leq \frac{1}{2^n},$$

para todo $A \in \sigma(X_n)$.

Sea $B \in \sigma(X_0)$, entonces

$$\begin{aligned} |P^{(n)}(B, A) - \pi(A)| &= \frac{|\mathbb{P}^{(n)}(A, B) - \pi(A)\pi(B)|}{\pi(B)} \\ &= \frac{|\sum_{i \in B} [\mathbb{P}^{(n)}(A, i) - \pi(A)\pi(i)]|}{\pi(B)} \\ &= \frac{|\sum_{i \in B} \pi(i)[P^{(n)}(i, A) - \pi(A)]|}{\pi(B)} \\ &\leq \frac{\frac{1}{2^n} \sum_{i \in B} \pi(i)}{\pi(B)} \\ &= \frac{1}{2^n}, \end{aligned}$$

donde $\mathbb{P}^{(n)}(\cdot, \cdot)$ denota la función de probabilidad conjunta de X_0 y X_n .

Por lo tanto $\phi(n) \leq \frac{1}{2^n} \xrightarrow[n \rightarrow \infty]{} 0$, lo que implica que el proceso es ϕ mixing.

Veamos qué pasa con el proceso en reversa.

$$\mathbb{P}[X_n = i | X_{n+1} = j] = \frac{\mathbb{P}[X_{n+1} = j | X_n = i] \mathbb{P}[X_n = i]}{\mathbb{P}[X_{n+1} = j]} = \frac{P(i, j)}{2^{i-j}},$$

entonces, la matriz de transición del proceso en reversa, $Q(i, j)$, está dada por

$$Q(i, j) = \begin{cases} \frac{1}{2^j}, & i = 1, \\ 1, & j = i - 1, \\ 0, & \text{en otro caso,} \end{cases}$$

y otra vez no es complicado ver que en n pasos tenemos

$$Q^{(n)}(i, j) = \begin{cases} \frac{1}{2^j}, & i \leq n, \\ 1, & j = i - n, \\ 0, & \text{en otro caso.} \end{cases}$$

Para calcular $\phi(n)$, notemos que para la pareja $(n+1, 1)$, tenemos que

$$Q^{(n)}(n+1, 1) - \pi(1) = \frac{1}{2},$$

de donde

$$|Q^{(n)}(B, A) - \pi(A)| \geq \frac{1}{2},$$

para toda n , y por lo tanto $\phi(n) \geq \frac{1}{2}$, lo que implica que no es ϕ -mixing. \square

Veamos otro ejemplo.

Ejemplo 3.3.5. Sean $\{Z_t\}_{t \in \mathbb{Z}}$ variables aleatorias continuas, independientes e idénticamente distribuidas, tales que $\gamma = \mathbb{E}[|Z_1|^\delta] < \infty$ para algún $\delta > 0$. Sea $\{X_t\}_{t \in \mathbb{Z}}$ un proceso lineal definido por

$$X_t = \sum_{k=0}^{\infty} g_k Z_{t-k},$$

con $g_0 > 0$. En este ejemplo, veremos que si $\sum_{v=0}^{\infty} v |g_v|^\lambda < \infty$, $\lambda = \frac{\delta}{1+\delta}$, entonces $\{X_t\}$ es strong-mixing.

Notemos que el proceso $\{X_t\}$ es estacionario. Sea C cualquier conjunto boreliano $(p+1)$ -dimensional en el espacio X_{-p}, \dots, X_0 y D una unión de s intervalos m -dimensionales disjuntos en el espacio X_k, \dots, X_{k+m-1} , con p, s, k, m enteros positivos arbitrarios. Sea

$$\mathbf{X} = (X_{-p}, \dots, X_0),$$

$$\mathbf{Y} = (X_k, \dots, X_{k+m-1}),$$

$$\mathbf{y} = (x_k, \dots, x_{k+m-1}),$$

$$D = \bigcup_{j=1}^s D_j,$$

$$D_j = \{\mathbf{y} : a_{jt} < y_t < b_{jt}, k \leq t \leq k + m - 1\},$$

donde D_j son s intervalos ajenos. Definamos

$$W_t = \sum_{j=0}^{t-1} g_j Z_{t-j}, \quad k \leq t \leq k + m - 1,$$

y de aquí

$$V_t = X_t - W_t = \sum_{j=t}^{\infty} g_j Z_{t-j}, \quad k \leq t \leq k + m - 1.$$

Sean

$$\begin{aligned} \pi_{12} &= \mathbb{P}[\mathbf{X} \in C, \mathbf{Y} \in D] \\ \pi_1 &= \mathbb{P}[\mathbf{X} \in C], \quad \pi_2 = \mathbb{P}[\mathbf{Y} \in D]. \end{aligned}$$

y

$$\mathbf{V} = (V_k, \dots, V_{k+m-1}),$$

de tal manera que $\mathbf{Y} = \mathbf{W} + \mathbf{V}$. Definamos también

$$B = \{\mathbf{v} : |v_t| \leq \eta_y, k \leq y \leq k + m - 1\},$$

con $\mathbf{v} = (v_k, \dots, v_{k+m-1})$. Entonces

$$\begin{aligned} \pi_{12} &= \mathbb{P}[\mathbf{X} \in C, \mathbf{W} + \mathbf{V} \in D, \mathbf{V} \in B] \\ &\quad + \mathbb{P}[\mathbf{X} \in C, \mathbf{W} + \mathbf{V} \in D, \mathbf{V} \in B^c]. \end{aligned} \quad (3.15)$$

Como \mathbf{W} es independiente de \mathbf{X} y \mathbf{V} , la primera expresión del lado derecho de (3.15) la podemos ver como

$$\int_{\{\mathbf{x} \in C, \mathbf{v} \in B\}} \chi(\mathbf{v}) dF_{\mathbf{X}, \mathbf{V}}(\mathbf{x}, \mathbf{v}), \quad (3.16)$$

donde $F_{\mathbf{X}, \mathbf{V}}(\mathbf{x}, \mathbf{v})$ es la función de distribución conjunta de \mathbf{X} y \mathbf{V} , $\mathbf{x} = (x_{-p}, \dots, x_0)$ y

$$\chi(\mathbf{v}) = \mathbb{P} \left[\bigcup_{j=1}^s a_{jt} - v_t < W_t < b_{jt} - v_t, k \leq t \leq k+m-1 \right].$$

Notemos que $0 \leq \chi(\mathbf{v}) \leq 1$, además de que $\chi(\mathbf{v})$ es continua para todo $\mathbf{v} \in B$. Para ver que lo anterior es cierto, veamos que \mathbf{W} admite una densidad. Como las variables Z_t son continuas e independientes, el vector (Z_1, \dots, Z_{k+m-1}) tiene una función de densidad conjunta. Por lo que por el teorema de cambio de variable, sólo resta completar el vector \mathbf{W} de manera adecuada y ver que el valor absoluto del determinante del Jacobiano de dicha transformación es siempre positivo. Para eso, definamos las variables W_i con $1 \leq i \leq k-1$, de la misma manera que las variables W_t . Dicha transformación, tiene como Jacobiano una matriz triangular inferior con g_0 sobre la diagonal, lo que implica que el determinante es $g_0^{k+m-1} > 0$ y por lo tanto el vector (W_1, \dots, W_{k+m-1}) admite una densidad conjunta y de la misma manera el vector \mathbf{W} . Por lo tanto, $\chi(\mathbf{v})$ es continua. Como B es acotado y cerrado, entonces existen θ y ξ en B , tales que

$$\chi(\theta) = \min_{\mathbf{v} \in B} \chi(\mathbf{v})$$

$$\chi(\xi) = \max_{\mathbf{v} \in B} \chi(\mathbf{v})$$

Por las ecuaciones (3.15) y (3.16) tenemos

$$\begin{aligned} \pi_{12} &\geq \int_{\{\mathbf{x} \in C, \mathbf{v} \in B\}} \chi(\mathbf{v}) dF_{\mathbf{X}, \mathbf{V}}(\mathbf{x}, \mathbf{v}) \\ &\geq \int_{\{\mathbf{x} \in C, \mathbf{v} \in B\}} \chi(\theta) dF_{\mathbf{X}, \mathbf{V}}(\mathbf{x}, \mathbf{v}) \\ &= \chi(\theta) \mathbb{P}[\mathbf{X} \in C, \mathbf{V} \in B] \\ &\geq \chi(\theta) \mathbb{P}[\mathbf{X} \in C] - \mathbb{P}[\mathbf{V} \in B^c]. \end{aligned}$$

De la misma manera tenemos que

$$\begin{aligned} \pi_{12} &\leq \int_{\{\mathbf{x} \in C, \mathbf{v} \in B\}} \chi(\mathbf{v}) dF_{\mathbf{X}, \mathbf{V}}(\mathbf{x}, \mathbf{v}) + \mathbb{P}[\mathbf{V} \in B^c] \\ &\leq \chi(\xi) \mathbb{P}[\mathbf{X} \in C, \mathbf{V} \in B] + \mathbb{P}[\mathbf{V} \in B^c] \\ &\leq \chi(\xi) \mathbb{P}[\mathbf{X} \in C] + \mathbb{P}[\mathbf{V} \in B^c]. \end{aligned}$$

Juntando ambas desigualdades tenemos

$$\chi(\theta)\pi_1 - \mathbb{P}[\mathbf{V} \in B^c] \leq \pi_{12} \leq \chi(\xi)\pi_1 + \mathbb{P}[\mathbf{V} \in B^c].$$

Como $\mathbb{P}[\mathbf{V} \in B^c] \leq \sum_{t=k}^{k+m-1} \mathbb{P}[|V_t| > \eta_t] = p(\eta)$, digamos, entonces de la desigualdad anterior obtenemos que

$$\chi(\theta)\pi_1 - p(\eta) \leq \pi_{12} \leq \chi(\xi)\pi_1 + p(\eta). \quad (3.17)$$

Si consideramos C como el espacio entero de \mathbf{X} en la ecuación (3.17), tenemos que

$$\chi(\theta) - p(\eta) \leq \pi_2 \leq \chi(\xi) + p(\eta). \quad (3.18)$$

Ahora, si multiplicamos (3.18) por $-\pi_1$ y la sumamos a (3.17), obtenemos

$$\begin{aligned} -\pi_1(\chi(\xi) - \chi(\theta)) - (1 + \pi_1)p(\eta) \\ \leq \pi_{12} - \pi_1\pi_2 \\ \leq \pi_1(\chi(\xi) - \chi(\theta)) + (1 + \pi_1)p(\eta), \end{aligned}$$

de donde,

$$|\pi_{12} - \pi_1\pi_2| \leq 2[\chi(\xi) - \chi(\theta) + p(\eta)]. \quad (3.19)$$

Por la continuidad de la densidad de \mathbf{W} , las parciales de $\chi(\mathbf{v})$ existen en todo B y son continuas. Otra vez, como B es cerrado y acotado, las parciales son acotadas y entonces

$$\left| \frac{\partial \chi(\mathbf{v})}{\partial v_t} \right| \leq M,$$

para toda $\mathbf{v} \in B$. Por el teorema del valor medio tenemos que

$$\chi(\xi) - \chi(\theta) = \nabla \chi(\mathbf{v}_*) \cdot (\xi - \theta).$$

Como $\theta, \xi \in B$, entonces para cada entrada del vector $\xi - \theta$, tenemos que $|\xi_t - \theta_t| \leq 2\eta_t$, $k \leq t \leq k + m - 1$, por lo que

$$\chi(\xi) - \chi(\theta) \leq 2M \sum_{t=k}^{k+m-1} \eta_t. \quad (3.20)$$

Por otro lado

$$\begin{aligned}\mathbb{P}[|V_t| > \eta_t] &= \int \mathbb{I}_{\{|V_t|/\eta_t > 1\}} dF_{\mathbf{V}} \\ &\leq \int_{\{|V_t|/\eta_t > 1\}} \frac{|V_t|^\delta}{\eta_t^\delta} dF_{\mathbf{V}} \\ &\leq \frac{\mathbb{E}[|V_t|^\delta]}{\eta_t^\delta}.\end{aligned}$$

Procedamos a acotar $\mathbb{E}[|V_t|^\delta]$. Primero, si $\delta < 1$, notemos que

$$\left| \sum_{j=t}^{\infty} g_j Z_{t-j} \right|^\delta \leq \sum_{j=t}^{\infty} |g_j Z_{t-j}|^\delta. \quad (3.21)$$

Para probarlo, veamos que si tenemos dos números a y b con $|a| \geq |b|$, entonces probar que

$$|a + b|^\delta \leq |a|^\delta + |b|^\delta,$$

es lo mismo que probar que

$$|1 + x|^\delta \leq 1 + |x|^\delta, \quad |x| \leq 1,$$

factorizando $|a|^\delta$, pero lo anterior es cierto pues

$$|1 + x|^\delta \leq 1 + |x| \leq 1 + |x|^\delta.$$

De manera inductiva podemos ver, entonces, que si tenemos una sucesión $\{a_j\}_{j=1}^{\infty}$,

$$\left| \sum_{j=1}^n a_j \right|^\delta \leq \sum_{j=1}^n |a_j|^\delta,$$

y aplicando límite de ambos lados obtenemos el resultado de (3.21). Entonces

$$\begin{aligned}\mathbb{E}[|V_t|^\delta] &\leq \sum_{j=t}^{\infty} \mathbb{E}[|g_j Z_{t-j}|^\delta] \\ &= \sum_{j=t}^{\infty} |g_j|^\delta \mathbb{E}[|Z_1|^\delta] \\ &= \gamma \sum_{j=t}^{\infty} |g_j|^\delta,\end{aligned}$$

ya que por hipótesis tenemos que $\mathbb{E}[|Z_1|^\delta] = \gamma$. Ahora, si $\delta \geq 1$, tenemos que

$$\mathbb{E}[|V_t|^\delta] \leq \gamma \left(\sum_{j=t}^{\infty} |g_j| \right)^\delta.$$

Para cualquiera de los dos casos de δ , denotemos a la cota como γG_t . Entonces

$$p(\eta) \leq \sum_{t=k}^{k+m-1} \frac{G_t}{\eta_t^\delta} \leq 2M \sum_{t=k}^{k+m-1} \frac{G_t}{\eta_t^\delta}, \quad (3.22)$$

siempre y cuando $M \geq \frac{1}{2}$, si no, elegimos $M = \frac{1}{2}$. Juntando (3.19) con (3.20) y (3.22), tenemos que

$$|\pi_{12} - \pi_1 \pi_2| \leq 2M \sum_{t=k}^{k+m-1} \left(\eta_t + \frac{G_t}{\eta_t^\delta} \right).$$

Como el lado izquierdo no depende de η_t , podemos elegir η_t de tal manera que

$$\eta_t = \frac{G_t}{\eta_t^\delta},$$

o bien

$$\eta_t = G_t^{\frac{1}{1+\delta}},$$

de donde tenemos que

$$|\pi_{12} - \pi_1 \pi_2| \leq 4M \sum_{t=k}^{k+m-1} G_t^{\frac{1}{1+\delta}}.$$

Como $1 + \delta > \delta$, entonces

$$\left(\sum_{j=t}^{\infty} |g_j| \right)^{\frac{\delta}{1+\delta}} \leq \sum_{j=t}^{\infty} |g_j|^{\frac{\delta}{1+\delta}},$$

y de la misma manera, como $1 + \delta > 1$, entonces

$$\left(\sum_{j=t}^{\infty} |g_j|^\delta \right)^{\frac{1}{1+\delta}} \leq \sum_{j=t}^{\infty} |g_j|^{\frac{\delta}{1+\delta}},$$

lo que implica que para todo $\delta > 0$

$$G_t^{\frac{1}{1+\delta}} \leq \sum_{j=t}^{\infty} |g_j|^\lambda,$$

donde $\lambda = \frac{\delta}{1+\delta}$, lo que implica

$$|\pi_{12} - \pi_1 \pi_2| \leq 4M \sum_{t=k}^{k+m-1} \sum_{j=t}^{\infty} |g_j|^\lambda. \quad (3.23)$$

Veamos que

$$\begin{aligned} \sum_{t=k}^{k+m-1} \sum_{j=t}^{\infty} |g_j|^\lambda &\leq \sum_{t=k}^{\infty} \sum_{j=t}^{\infty} |g_j|^\lambda \\ &= \sum_{j=k}^{\infty} \sum_{t=k}^j |g_j|^\lambda \\ &\leq \sum_{j=k}^{\infty} j |g_j|^\lambda. \end{aligned}$$

Si nombramos $h(k) = \sum_{j=k}^{\infty} j |g_j|^\lambda$, por la ecuación (3.23) tenemos que

$$|\pi_{12} - \pi_1 \pi_2| \leq Qh(k), \quad (3.24)$$

donde Q es una constante que no depende de k y el resultado no depende de m . Por lo tanto, ya probamos que para cualquier conjunto boreliano C en el espacio \mathbf{X} y cualquier unión arbitraria de intervalos D en el espacio de $\{X_k, X_{k+1}, \dots\}$ se cumple (3.24). Para probar el resultado en general, usemos el Lema de π , λ -sistema (véase Apéndice A). Sea \mathcal{C} , la clase de todas las imágenes inversas de conjuntos Borelianos en el espacio de $\{X_k, X_{k+1}, \dots\}$ que cumplen con

$$|\mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B]| \leq Qh(k),$$

con $A \in \mathcal{F}_{-\infty}^0$ y $B \in \mathcal{C}$. Queremos probar que dicha clase es un λ -sistema. Sea $B \in \mathcal{C}$, entonces $B^c \in \mathcal{C}$, ya que

$$\mathbb{P}[AB^c] - \mathbb{P}[A]\mathbb{P}[B^c] = \mathbb{P}[A]\mathbb{P}[B] - \mathbb{P}[AB].$$

Sea B_n una sucesión monótona de conjuntos en \mathcal{C} , con $\lim_{n \rightarrow \infty} B_n = B$. Entonces,

$$|\mathbb{P}[AB_n] - \mathbb{P}[A]\mathbb{P}[B_n]| \leq Qh(k),$$

implica

$$|\mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B]| \leq Qh(k),$$

por la continuidad de la probabilidad y el hecho de que la cota no depende de n . Por lo tanto, \mathcal{C} es una λ -sistema. Como la imagen inversa sobre los intervalos en el espacio $\{X_k, X_{k+1}, \dots\}$ es un π -sistema, puesto que es cerrada bajo intersecciones y además por (3.24) está contenida en \mathcal{C} , tenemos que el resultado es válido para \mathcal{F}_k^∞ . Por lo tanto, si $A \in \mathcal{F}_{-\infty}^0$ y $B \in \mathcal{F}_k^\infty$, tenemos que

$$|\mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B]| \leq Qh(k),$$

y como la serie $\sum_{j=0}^{\infty} j|g_j|^\lambda < \infty$, $h(k) \xrightarrow[k \rightarrow \infty]{} 0$, lo que implica que el proceso lineal es α -mixing.

A pesar de que las variables X_t dependen todas entre sí, el hecho de que la serie g_j se vaya "rápido" a cero, implica que las variables Z_k con mayor peso para las X_t sean distintas conforme t crece, lo que provoca que la independencia de las Z_k , se refleje en una independencia asintótica en las X_t . \square

Un último ejemplo.

Ejemplo 3.3.6. Sean $\{Z_t\}_{t \in \mathbb{Z}}$ variables aleatorias independientes e idénticamente distribuidas tales que $\mathbb{P}[Z_t = 0] = \mathbb{P}[Z_t = 1] = \frac{1}{2}$. Definamos las variables $\{X_t\}_{t \in \mathbb{Z}}$ como

$$X_t = \sum_{k=0}^{\infty} \frac{1}{2^{k+1}} Z_{t-k}. \quad (3.25)$$

El proceso es un AR(1) estrictamente estacionario puesto que $X_t = \frac{1}{2}X_{t-1} + \frac{1}{2}Z_t$, aunque también puede ser MA(∞). Para toda t , la variable aleatoria

X_t , se distribuye uniforme $(0,1)$ ya que, en la ecuación (3.25), las variables Z_k son los dígitos en la expansión binaria de X_t . Como además, las variables son independientes y la probabilidad de que tomen el valor 1 ó 0 es $\frac{1}{2}$, entonces, todas las expansiones son igualmente probables. Notemos además que a X_t lo podemos escribir como

$$X_t = 2X_{t+1} - Z_{t+1};$$

si $Z_{t+1} = 0$, entonces $2X_{t+1} \in (0, 1)$, y si $Z_{t+1} = 1$, entonces $2X_{t+1} - 1 \in (0, 1)$. En ambos casos, X_t es la parte fraccional de $2X_{t+1}$, es decir

$$X_t = 2X_{t+1} - [2X_{t+1}],$$

donde $[x]$ representa la parte entera de x . Por lo tanto, X_t es una función boreliana de X_{t+1} , por lo que $\sigma(X_t) \subseteq \sigma(X_{t+1})$. Por lo tanto, para toda $n \geq 1$, X_0 es una función de X_n , y entonces $\sigma(X_0) \subseteq \sigma(X_n)$. Entonces

$$\begin{aligned} \alpha(n) &= \alpha(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty) \\ &\geq \alpha(\sigma(X_0), \sigma(X_n)) \\ &\geq \alpha(\sigma(X_0), \sigma(X_0)) \\ &\geq \mathbb{P} \left[X_0 \leq \frac{1}{2} \right] - \left(\mathbb{P} \left[X_0 \leq \frac{1}{2} \right] \right)^2 = \frac{1}{4}. \end{aligned}$$

Por lo tanto, $\alpha(n) = \frac{1}{4}$ para toda $n \in \mathbb{N}$, lo que implica que el proceso no es α -mixing.

En este ejemplo, a diferencia del anterior, a pesar de que la serie $\frac{1}{2^k}$, converge bastante rápido a 0, el proceso no es mixing. Esto hace importante la condición de que las variables Z_t sean continuas. \square

Las condiciones anteriores se han considerado con el fin de extender los teoremas de límite de sucesiones para que éstas sean cada vez menos restrictivas. En vez de pedir independencia, pedimos ahora una independencia asintótica. En resumen, vimos que las condiciones de mixing no son una extensión ociosa de la m -dependencia, que si bien la sucesión del ejemplo 3.3.2 sólo demostraba la existencia de éstas y explicaba un poco su funcionamiento, no tiene ninguna aplicación práctica, mientras que las sucesiones de los ejemplos 3.3.4 y 3.3.5 sí la tienen. El último ejemplo sirve para detener nuestro paso y recordarnos que las condiciones de cada problema particular son importantes y que éste es un tema complicado, por lo que cada problema debe ser analizado con cuidado.

Capítulo 4

Aplicaciones

En este capítulo analizaremos la parte estadística de la medida de dependencia δ , así como veremos aplicaciones de ésta. En la Sección 4.1 propondremos una contraparte muestral para la medida δ propuesta en la Sección 2.5; en la Sección 4.2 veremos un método, utilizando la medida muestral, para estudiar series de tiempo. Por último, en la Sección 4.3, aplicaremos los métodos propuestos a simulaciones y analizaremos un caso de datos reales con los precios del petróleo.

La Sección 4.1 se basó en [10] y la Sección 4.2 en [7].

4.1 Medida de dependencia muestral

Supongamos que tenemos una muestra de tamaño m , $\mathbf{X}_i = (X_i, Y_i)$, con $i = 1, \dots, m$, provenientes de una distribución común $H(x, y)$, con marginales $F(x)$ y $G(y)$, respectivamente. Denotemos como $H_m(x, y)$ a la función de distribución empírica conjunta y, $F_m(x)$ y $G_m(y)$, a las funciones empíricas marginales de la muestra, es decir,

$$H_m(x, y) = \frac{\sum_{i=1}^m \mathbb{I}_{(-\infty, X_i]}(x) \mathbb{I}_{(-\infty, Y_i]}(y)}{m}, \quad (4.1)$$

$$F_m(x) = \frac{\sum_{i=1}^m \mathbb{I}_{(-\infty, X_i]}(x)}{m}, \quad G_m(y) = \frac{\sum_{i=1}^m \mathbb{I}_{(-\infty, Y_i]}(y)}{m}. \quad (4.2)$$

Proponemos como medida de dependencia de la muestra

$$\hat{\delta}_m(X, Y) := \sup_{(x, y) \in \mathbb{R}^2} |H_m(x, y) - F_m(x)G_m(y)|. \quad (4.3)$$

Esta medida es la versión muestral de aquella propuesta para δ en (2.43). Por el Teorema de Glivenko-Cantelli (véase Teorema A.0.4) sabemos que H_m , F_m y G_m convergen, respectivamente, a H , F y G , lo que da más sentido a la ecuación (4.3). Veamos el siguiente teorema.

Teorema 4.1.1. *Sea (X_i, Y_i) , para $i = 1, \dots, m$, una muestra independiente que proviene de una distribución bivariada común H , con marginales F y G , respectivamente. Sean H_m , G_m y F_m definidas en (4.1) y (4.2). Entonces*

$$\hat{\delta}_m(X, Y) \xrightarrow[m \rightarrow \infty]{c.s.} \delta(X, Y).$$

Demostración. Por simplicidad, en lugar de escribir $H(x, y)$, escribiremos H , y lo mismo para las demás funciones. Para todo $x, y \in \mathbb{R}$, aplicando la desigualdad del triángulo, tenemos que

$$\begin{aligned} |H_m - F_m G_m| &= |H_m - H + H - FG + FG - FG_m + FG_m - F_m G_m| \\ &\leq |H_m - H| + |H - FG| + F|G - G_m| + G_m|F - F_m| \\ &\leq |H_m - H| + |H - FG| + |G - G_m| + |F - F_m|, \end{aligned}$$

de la misma manera tenemos que

$$\begin{aligned} |H - FG| &= |H - H_m + H_m - F_m G_m + F_m G_m - FG_m + FG_m - FG| \\ &\leq |H - H_m| + |H_m - F_m G_m| + G_m|F_m - F| + F|G_m - G| \\ &\leq |H - H_m| + |H_m - F_m G_m| + |F_m - F| + |G_m - G|. \end{aligned}$$

Aplicando supremos y juntando ambas desigualdades tenemos que

$$\begin{aligned} \sup |H - FG| - \sup |H - H_m| - \sup |F_m - F| - \sup |G_m - G| \\ \leq \sup |H_m - F_m G_m| \\ \leq \sup |H_m - H| + \sup |H - FG| + \sup |G - G_m| + \sup |F - F_m|. \end{aligned}$$

Aplicando Glivenko-Cantelli a la ecuación anterior obtenemos

$$\delta(X, Y) \leq \lim_{m \rightarrow \infty} \hat{\delta}_m(X, Y) \leq \delta(X, Y),$$

es decir

$$\lim_{m \rightarrow \infty} \hat{\delta}_m(X, Y) \stackrel{c.s.}{=} \delta(X, Y).$$

□

Para proponer una medida de dependencia muestral con este estadístico, necesitamos encontrar su cota superior, para que, sin importar el tamaño de la muestra, nos dé un valor normalizado. En el caso de la distribución real vimos que estaba acotado por $\frac{1}{4}$ en general, sin embargo en el caso muestral no sucede así. Si $H_m(x, y) = \frac{k}{m}$, entonces $F_m(x), G_m(y) \geq \frac{k}{m}$, y por lo tanto

$$H_m(x, y) - F_m(x)G_m(y) \leq \frac{k}{m} - \left(\frac{k}{m}\right)^2,$$

de donde obtenemos que

$$\hat{\delta}_m(X, Y) \leq \max_{0 \leq k \leq m} \left(\frac{k}{m} - \left(\frac{k}{m}\right)^2 \right) := \kappa_m. \quad (4.4)$$

En el caso de que m sea par, κ_m es igual a $\frac{1}{4}$ ya que el máximo lo alcanza para $k = \frac{m}{2}$. Si m es impar de la forma $m = 2n + 1$, el máximo lo alcanza para $k = n$ ó $k = n + 1$, valiendo en ambos casos $\kappa_m = \frac{n(n+1)}{(2n+1)^2}$. Por lo tanto, para cualquier caso de m , definimos como medida de dependencia muestral

$$\mathcal{C}(X, Y) := \frac{\hat{\delta}_m(X, Y)}{\kappa_m}.$$

Sabemos que $\mathcal{C} \in [0, 1]$. Veamos qué pasa con los valores extremos. Supongamos que tenemos valores $x_1 < x_2 < \dots < x_m$ y $y_1 < y_2 < \dots < y_m$, y que la muestra $(X_i, Y_i)_{i=1}^{m^2}$ está dada por $\{(x_i, y_j) | i, j \in \{1, \dots, m\}\}$. Es decir, la muestra es una cuadrícula de $m \times m$ formada por los dos conjuntos de puntos. Esta forma es la única que puede representar muestras “independientes” en el sentido de que para cualquier valor que obtengamos en la coordenada x , todos los valores de y son igualmente probables, es decir, es

la única forma de generar parejas de valores cuyas distribuciones muestrales resulten ser independientes. Es de esperarse que esta muestra cumpla que $\mathcal{C}(X, Y) = 0$. Como la función de distribución muestral sólo salta en los puntos de la muestra consideremos únicamente lo que vale en estos puntos. Entonces

$$H_m(x_i, y_j) - F_m(x_i)G_m(y_j) = \frac{ij}{m} - \frac{mi}{m^2} \frac{mj}{m^2} = 0,$$

para $i, j = 1, \dots, m$. Por lo tanto, $\mathcal{C}(X, Y) = 0$.

Como esta medida es la versión muestral normalizada de δ , esperamos que el valor extremo sea alcanzado cuando la muestra cumple una relación similar a la que tiene la cópula en el teorema 2.5.3. Siguiendo una prueba muy parecida a la del teorema 2.5.3 podemos ver que si tenemos una muestra de tamaño m , $\mathcal{C}(X, Y) = 1$ si y sólo si para $j = \lfloor \frac{m}{2} \rfloor$, existe $(a, b) \in \mathbb{R}^2$ tal que $H_m(a, b) = \frac{j}{m}$, $F_m(a) = \frac{j}{m}$ y $G_m(b) = \frac{j}{m}$. En otras palabras, si $A = \{(x, y) | x \leq a, y \leq b\}$, $B = \{(x, y) | x > a, y \leq b\}$, $C = \{(x, y) | x > a, y > b\}$ y $D = \{(x, y) | x \leq a, y > b\}$, la muestra está totalmente contenida en $A \cup C$, o lo está en $B \cup D$. En resumen, se comporta de manera similar a δ .

Durante los capítulos anteriores siempre se discutió el problema que hay al querer utilizar medidas como la correlación, la τ de Kendall y la ρ de Spearman para identificar cuándo dos poblaciones son independientes. Se vió, también, que medidas como σ ó δ nos permiten determinar de manera única estos casos. En consecuencia, si queremos identificar la independencia en el caso de una muestra, lo conveniente es usar la versión muestral de alguna de las dos medidas anteriores. La versión muestral de σ es complicada ya que las funciones de distribución muestrales son discontinuas lo que dificulta el cálculo de sus integrales. Por ese motivo nos quedamos con la δ . A continuación, veremos cómo utilizarla para hacer una prueba estadística que nos dé un criterio para decir si los componentes de la muestra son independientes entre sí.

Teorema 4.1.2. *Sea $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ una muestra aleatoria de tamaño m , proveniente de una población (X, Y) , variables aleatorias continuas con distribución $H(x, y)$.*

1. $\hat{\delta}_m(X, Y) \neq 0$ c.s. para toda $m \geq 2$.
2. Sean X y Y variables aleatorias uniformes $(0, 1)$ independientes, entonces siempre es posible calcular la distribución de $\hat{\delta}_m(X, Y)$ para toda $m \geq 2$.

3. Sean X y Y variables aleatorias independientes con distribuciones marginales F y G respectivamente, entonces

$$\hat{\delta}_m(X, Y) \stackrel{D}{=} \hat{\delta}_m(F(X), G(Y)),$$

donde $\stackrel{D}{=}$ representa igualdad en distribución.

Demostración. 1. Como X y Y son continuas, podemos suponer, reordenando, que $x_1 < x_2 < \dots < x_m$, además de que $y_i \neq y_j$ para $i \neq j$, lo que implica que $H_m(x_1, y_1) = \frac{1}{m}$ y $F_m(x_1) = \frac{1}{m}$. Si $H_m(x_1, y_1) - F_m(x_1)G_m(y_1) = 0$, esto implica que $G_m(y_1) = 1$, de donde vemos que $y_1 = \max(y_1, y_2, \dots, y_m)$. Tenemos también que $H_m(x_1, y_2) = 0$ y $\frac{1}{m} \leq G_m(y_2) \leq \frac{m-1}{m}$ lo que implica que $H_m(x_1, y_2) - F_m(x_1)G_m(y_2) \neq 0$. Por lo tanto $\hat{\delta}_m(X, Y) \neq 0$ c.s.

2. Supongamos que $0 < x_1 < x_2 < \dots < x_n < 1$. Definamos $w_i = \frac{i}{m+1}$, para $i = 1, \dots, m$, entonces $0 < z_1 < z_2 < \dots < z_n < 1$. Denotemos por $0 < y_{(1)} < y_{(2)} < \dots < y_{(m)} < 1$ las coordenadas y_i ordenadas. Para todo $i = 1, \dots, m$, existe un único $j_i \in \{1, \dots, m\}$ tal que $y_i = y_{(j_i)}$ y definamos $z_i = \frac{j_i}{m+1}$ para $i = 1, \dots, m$. Entonces para todo $1 \leq k, i \leq m$ tenemos que

$$H_m(x_k, y_i) = H_m(w_k, z_i), \quad F_m(x_k) = F_m(w_k), \quad G_m(y_i) = G_m(z_i),$$

lo que implica que $\hat{\delta}_m$ es igual bajo cualquiera de las dos muestras, (x_i, y_i) ó (w_i, z_i) . Por lo tanto, $\hat{\delta}_m(X, Y)$ sólo depende de la configuración de la muestra pero no de los valores específicos de ésta. Como las variables son independientes, cada una de las $m!$ posibles configuraciones son igualmente probables y entonces

$$\mathbb{P} \left[\hat{\delta}_m(X, Y) = \frac{k}{m^2} \right] = \frac{m(k)}{m!},$$

donde $m(k)$ representa el número de configuraciones para las cuales $\hat{\delta}_m(X, Y) = \frac{k}{m^2}$.

3. Supongamos que $0 < x_1 < x_2 < \dots < x_n < 1$. Como X y Y son continuas, F y G son estrictamente crecientes y entonces $0 < F(x_1) < F(x_2) < \dots < F(x_n) < 1$, además de que $G(y_i)$ preserva el orden de

las y 's. Como $F(X)$ y $G(Y)$ se distribuyen uniformes $(0,1)$ utilizando el mismo argumento de las z 's y w 's, tenemos que

$$\hat{\delta}_m(X, Y) \stackrel{D}{=} \hat{\delta}_m(F(X), G(Y)),$$

lo que completa la prueba. □

Lo que este teorema nos dice es que la distribución de $\hat{\delta}_m$, bajo la hipótesis de independencia, depende únicamente del tamaño de la muestra, mas no de las distribuciones marginales. Si queremos probar independencia entre los componentes de una muestra, basta elegir un nivel de significancia α y rechazar independencia si el valor $\hat{\delta}_m$ estimado es mayor que el cuantil $1 - \alpha$ de la distribución de $\hat{\delta}_m(X, Y)$ en el caso en que X y Y sean independientes. Lo anterior puede tener un problema porque el cálculo de la distribución de $\hat{\delta}_m$ en el caso de independencia, aunque puede ser realizado, es demasiado complicado, por lo que la distribución puede ser aproximada por simulaciones. En resumen, ante una muestra $(X_i, Y_i)_{i=1}^n$ seguimos los siguientes pasos:

- calcular la distribución exacta o una aproximación por simulaciones de $\hat{\delta}_m(U, V)$, para U y V variables aleatorias independientes uniformes $(0,1)$;
- evaluar el estadístico $\hat{\delta}_m(X, Y)$ para la muestra;
- rechazar independencia si el estadístico es mayor que el cuantil $1 - \alpha$ de la distribución exacta o aproximada, en otro caso, no rechazamos independencia.

4.2 Modelos de series de tiempo

En el capítulo anterior discutimos la dependencia que puede haber entre los componentes de una serie de tiempo. Vimos en el ejemplo 3.3.1 que si tenemos $\{Y_t\}_{t \in \mathbb{Z}}$ variables aleatorias independientes e idénticamente distribuidas y $X_t = f(Y_t, Y_{t-1}, \dots, Y_{t-q})$ con $f: \mathbb{R}^{q+1} \rightarrow \mathbb{R}$, entonces, la sucesión de variables $\{X_t\}_{t \in \mathbb{Z}}$ es estacionaria y q -dependiente. Supongamos que tenemos una muestra aleatoria, (X_1, \dots, X_n) , de un proceso de este tipo. Si queremos estudiar cómo es la dependencia entre dos tiempos consecutivos (X_t y X_{t+1}) a partir de la muestra, casi siempre, necesitamos que cada observación sea

independiente. Si conocemos el orden q de este proceso basta considerar las observaciones $(X_1, X_2), (X_{3+q}, X_{4+q}), \dots$, para obtener una muestra independiente de la pareja (X_t, X_{t+1}) . La idea de esta sección es presentar un método estadístico para determinar el orden q de un proceso de estas características utilizando el estadístico $\hat{\delta}_m$ y considerando las observaciones como componentes de vectores bidimensionales independientes

Supongamos que tenemos un modelo de serie de tiempo $\{X_t\}_{t \in \mathbb{Z}}$, donde

$$X_t = f(Y_t, Y_{t-1}, \dots, Y_{t-q}),$$

con $\{Z_t\}$ variables aleatorias continuas, independientes e idénticamente distribuidas. Para determinar el orden q de este modelo, notemos que si q es el orden correcto, entonces para todo $k > q$ y $t \in \mathbb{Z}$, las variables X_t y X_{t+k} son independientes, mientras que para $k \leq q$ las variables X_t y X_{t+k} resultan ser dependientes. Supongamos que nuestra muestra está dada por (X_1, X_2, \dots, X_n) , con n lo suficientemente grande para que permita tomar una muestra de tamaño m , es decir, $n \geq 1 + (2m - 1)(q + 1)$. Si consideramos

$$\begin{aligned} & (X_1, X_{1+(q+1)}) \\ & (X_{1+2(q+1)}, X_{1+3(q+1)}) \\ & \vdots \\ & (X_{1+2(m-1)(q+1)}, X_{1+(2m-1)(q+1)}), \end{aligned} \tag{4.5}$$

tenemos m vectores independientes e idénticamente distribuidos cuyas coordenadas son independientes. Si les aplicamos la prueba de independencia propuesta al final de la sección anterior, la distribución de $\hat{\delta}_m$ sólo depende de m y no de cómo sean las Y_t 's. Si consideramos $k > q$ y suponemos que $n \geq 1 + (2m - 1)k$, los vectores

$$\begin{aligned} & (X_1, X_{1+k}) \\ & (X_{1+2k}, X_{1+3k}) \\ & \vdots \\ & (X_{1+2(m-1)k}, X_{1+(2m-1)k}), \end{aligned} \tag{4.6}$$

se comportarán igual que los de (4.5) y la distribución de $\hat{\delta}_m$ permanecerá

igual. Sin embargo para $0 < k \leq q$, los vectores

$$\begin{aligned} & (X_1, X_{1+k}) \\ & (X_{1+2k}, X_{1+3k}) \\ & \quad \vdots \\ & (X_{1+2(m-1)k}, X_{1+(2m-1)k}), \end{aligned} \tag{4.7}$$

ya no serán, ni independientes entre ellos, ni sus coordenadas entre sí. La distribución de $\hat{\delta}_m$ cambia y es de esperarse que su valor aumente ya que su función de distribución conjunta no será tan parecida a la de la independencia.

En resumen, dada una muestra (X_1, X_2, \dots, X_n) , realizamos los siguientes pasos:

- elegimos $m \in \mathbb{N}$ tal que $n \geq 1 + (2m - 1)K$, donde K se supone ser lo suficientemente grande para que cumpla $K > q$, donde q es el orden correcto del modelo;
- formar muestras para $k = 1, 2, \dots, K$ como las de (4.6) y (4.7);
- para cada muestra utilizar, al nivel α , la prueba de dependencia propuesta en la sección anterior;
- el orden del modelo se toma como el mayor $q \in \{1, 2, \dots, K\}$ para el que fue rechazada la hipótesis de independencia.

4.3 Simulaciones y Aplicaciones

A continuación veremos parejas de variables aleatorias. En cada caso estudiaremos como funciona la prueba de independencia propuesta en la Sección 4.1. El primer caso deberá ser entonces el de la independencia. Se simularon 1000 valores de δ para cada caso, excepto para el caso de la independencia donde se simularon 5000. Los valores de rechazo que se consideraron fueron $\alpha = .05$ y $\alpha = .01$.

Ejemplo 4.3.1. Sean U y V dos variables aleatorias independientes $U(0,1)$. Con la simulación se obtuvo que

$$q_{.95} = .0795, \quad q_{.99} = .092$$

En lo siguiente, utilizaremos $\hat{\delta}_{100}(U, V)$ para denotar el caso de independen-

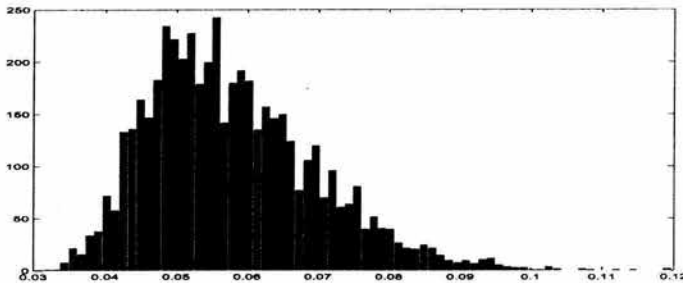


Figura 4.1: Histograma de frecuencias de $\hat{\delta}_{100}(U, V)$.

cia. □

Empecemos con los ejemplos.

Ejemplo 4.3.2. Sean X y Y dos variables aleatorias tales que $X \sim N(0, 1)$ y $Y = X^2$. De la simulación se obtuvo

Sim. \ $q_{1-\alpha}$.05	.01
(X, X^2)	1000	1000

En la tabla se registra el número de veces que se rechazó independencia a los niveles α . En este ejemplo, la prueba detectó en todos los casos la dependencia determinista que las parejas tienen. □

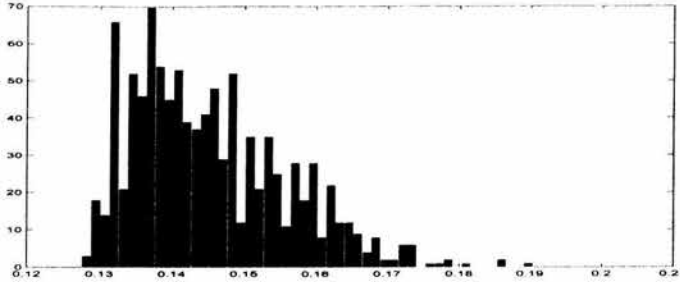


Figura 4.2: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$, (X, X^2) .

El ejemplo anterior es bueno, aunque la dependencia la podíamos haber observado con tan sólo graficar los datos. Trabajemos algunos ejemplos más interesantes. Los dos que veremos a continuación consisten en X y Z independientes, y la pareja (X, Y) dada por $(X, X + Z)$.

Ejemplo 4.3.3. Sea $X \sim U(0, 1)$ y $Z \sim U(0, \theta)$. Definamos $Y = X + Z$. La pareja tiene un elemento de dependencia y uno de independencia. Es de esperarse que cuando $\theta \rightarrow 0$, la cópula de la pareja converja a M , mientras que si $\theta \rightarrow \infty$ converja a la independencia. Se realizaron simulaciones para el caso de $\theta = \frac{1}{2}$, $\theta = 1$ y $\theta = 5$, obteniendo

Sim. \ $q_{1-\alpha}$.05	.01
$\theta = \frac{1}{2}$	1000	1000
$\theta = 1$	1000	1000
$\theta = 5$	104	27

Para los dos primeros casos la prueba funciona correctamente. Para el tercero también, en el sentido de que lo estudiado sugiere que eso tiene que pasar. Dado que la prueba en el caso de las cópulas, se basa en la comparación de la que las relaciona con la de la independencia, es de esperarse que los valores de $\hat{\delta}_{100}(X, Y)$ se parezcan a los de $\hat{\delta}_{100}(U, V)$ conforme θ es más grande.

Incluso si comparamos los histogramas de las Figuras 4.1 y 4.5, vemos que el comportamiento en distribución de $\hat{\delta}_{100}(X, Y)$ y $\hat{\delta}_{100}(U, V)$ es muy parecido. \square

Ejemplo 4.3.4. Sean $X \sim N(0, 1)$ y $Z \sim N(\mu, \sigma^2)$. Definamos $Y = X + Z$. Como podemos ver en el ejemplo 1.6.1, la pareja (X, Y) tiene una distribución normal bivariada con coeficiente de correlación $\rho = \frac{1}{\sqrt{1+\sigma^2}}$. En este caso, si

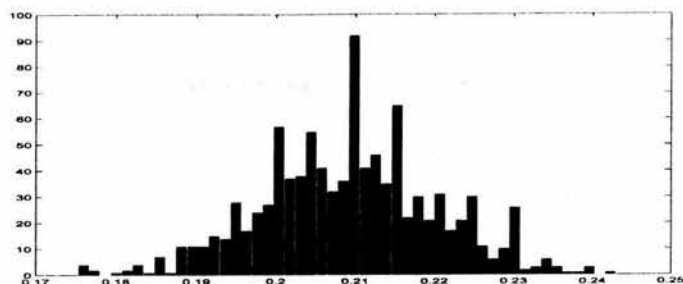


Figura 4.3: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$, $\theta = \frac{1}{2}$.

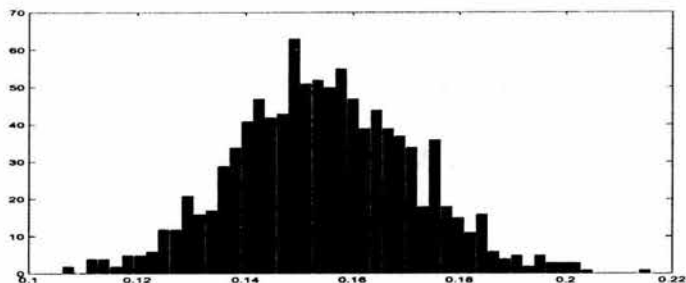


Figura 4.4: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$, $\theta = 1$.

$\sigma \rightarrow 0$, las variables tienden a una pareja $(X, X + \mu)$, mientras que si $\sigma \rightarrow \infty$, $\rho \rightarrow 0$ y por lo tanto, serán independientes. Se simularon los casos $\sigma^2 = 10$, $\sigma^2 = 3$ y $\sigma^2 = 1$ obteniéndose lo siguiente

Sim. \ $q_{1-\alpha}$.05	.01
$\sigma^2 = 10, \rho \approx .3$	126	40
$\sigma^2 = 3, \rho \approx .5$	691	418
$\sigma^2 = 1, \rho \approx .7$	1000	1000

Otra vez como esperabamos, conforme la varianza aumenta, las variables se vuelven menos “dependientes”. Esto se manifiesta en los datos obtenidos. \square

En el Teorema 2.6.1, vimos que con reordenamientos de M , nos podemos aproximar a Π , tanto como nosotros queramos con parejas de variables que son función la una de la otra. Veamos el siguiente ejemplo.

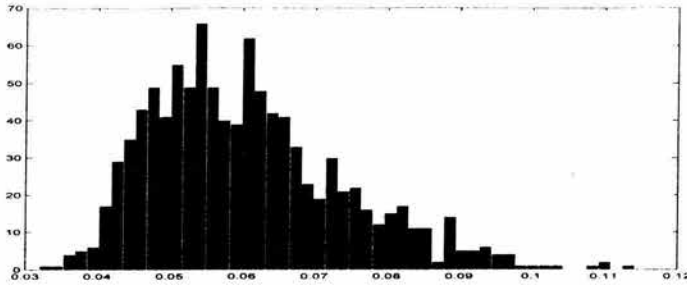


Figura 4.5: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$, $\theta = 5$.

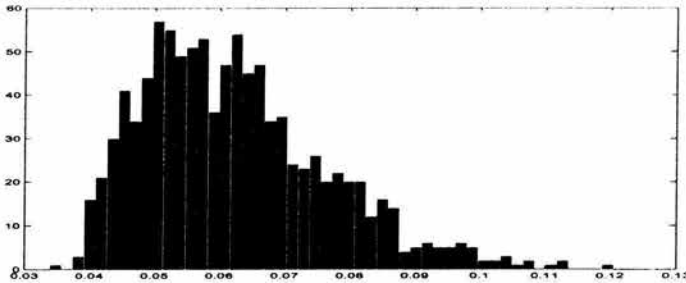


Figura 4.6: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$, $\sigma^2 = 10$.

Ejemplo 4.3.5. Sean X y Y variables aleatorias con cópula Π_ϵ (véase Teorema 2.6.1) para $\epsilon = .05$. Con la simulación se obtuvo que

Sim. \ $q_{1-\alpha}$.05	.01
$\Pi_{.05}$	54	8

Al comparar las Figuras 4.1 y 4.9, podemos ver que el comportamiento de $\hat{\delta}_{100}(U, V)$ y $\hat{\delta}_{100}(X, Y)$ es casi el mismo. Se calculó el supremo de las distancias entre las distribuciones muestrales de δ y está fue de .039, lo que confirma lo observado en los histogramas. Lo que vemos en este ejemplo es bastante lógico porque el sentido de Π_ϵ es parecerse tanto como queramos a Π . \square

También se estudió el comportamiento de la prueba en el caso de procesos q -dependientes. Cada proceso se simuló 1000 veces y en cada caso se eligió la q que el método de la sección anterior propone. Siguiendo la notación

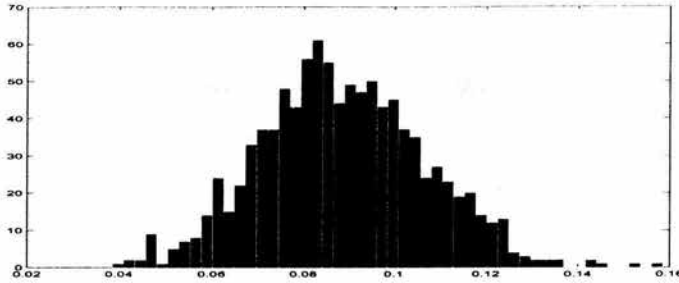


Figura 4.7: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$, $\sigma^2 = 3$.

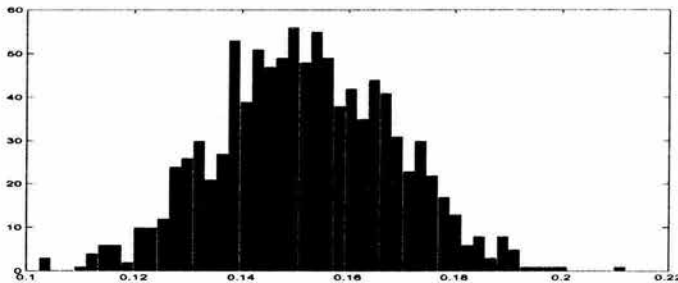


Figura 4.8: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$, $\sigma^2 = 1$.

propuesta, se eligió $m = 100$ y $K = 7$, es decir, muestras de tamaño 100 y hasta el retraso¹ 7. Consideramos otra vez los niveles $\alpha = .05$ y $\alpha = .01$. Veamos los siguientes ejemplos.

Ejemplo 4.3.6. Sea $\{X_t\}_{t \in \mathbb{Z}}$ un proceso definido a partir del proceso independiente $\{Z_t\}_{t \in \mathbb{Z}}$ por

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q},$$

con θ_i , $i = 1 \dots, q$ constantes, es decir, un $\text{MA}(q)$. Denotemos $\bar{\theta} = (\theta_1, \dots, \theta_q)$. A continuación presentamos la tabla de resultados,

¹Con retraso nos referimos a "lag".

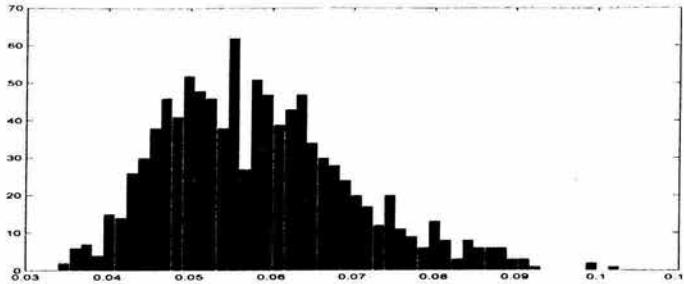


Figura 4.9: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$. $(X, Y) \sim \Pi_{.05}$

$\bar{\theta} \setminus q$	0	1	2	3	4	5	6	7
(1)	18	777	27	25	41	30	30	52
	138	840	7	1	5	1	5	3
(1,1)	0	255	565	36	30	26	35	53
	1	652	330	5	5	2	2	3
(1,1,1)	0	17	501	327	39	35	32	49
	0	136	725	124	5	2	4	4
(1,1,1,1)	0	0	105	539	244	29	35	48
	0	11	411	500	68	3	4	3
$(\frac{1}{2})$	142	653	27	27	35	43	37	36
	505	473	1	4	4	5	4	4
$(\frac{1}{4})$	572	227	32	33	29	35	36	36
	908	69	1	1	4	7	4	6

En la tabla² anterior podemos observar que los casos en que se detecta correctamente el nivel q de dependencia disminuyen por dos motivos: primero, si q aumenta, ya que la pareja X_t y X_{t+q} sólo dependen en Z_t , mientras que sus demás componentes son independientes; segundo, si los valores θ_i disminuyen, ya que el componente que más figura será el que su valor θ_i sea mayor. De todos modos, podemos observar que aunque la prueba falle, su resultado no es tan incorrecto, es decir, si consideramos el vector $\bar{\theta} = (1, 1, 1, 1)$, el valor correcto de q es 4 y aunque la prueba sólo lo detectó 244 veces en el caso $\alpha = .05$, propuso 539 veces el valor $q = 3$. Una última observación, si elegimos un valor K inicial muy alto, corremos el riesgo de equivocarnos en

²Para cada caso de $\bar{\theta}$, el primer renglón representa los resultados al nivel $\alpha = .05$ y el segundo al nivel $\alpha = .01$.

el siguiente sentido: en el caso $\bar{\theta} = (1)$, y $\alpha = .05$, se detectó el valor correcto ($q = 1$) 777 veces, y 205 veces se dijo que $q > 1$. La prueba está diseñada para reportar el q más alto y como se elige un nivel α , es de esperarse que para los retrasos mayores al correcto, es decir, donde hay independencia, aproximadamente $\alpha\%$ de las veces se rechace independencia. Conforme más retrasos tengamos, más probable se vuelve el rechazar independencia cuando sí la haya y por lo tanto equivocarnos. \square

Veamos otros ejemplos donde los procesos no son lineales.

Ejemplo 4.3.7. Sean $\{Z_t\}_{t \in \mathbb{Z}}$ variables independientes e idénticamente distribuidas. Definamos el proceso $\{X_t\}_{t \in \mathbb{Z}}$ como

$$X_t = \frac{Z_t - Z_{t-1}}{Z_{t-1}},$$

es decir, el cambio porcentual del proceso $\{Z_t\}$. En este ejemplo se consideraron distintas distribuciones para Z_t , el valor correcto de q es 1 y los resultados fueron los siguientes

Distr. \ q	0	1	2	3	4	5	6	7
$Z_t \sim N(0, 1)$	666	64	29	38	40	51	57	55
	933	14	6	10	10	8	13	6
$Z_t \sim U(0, 1)$	3	715	35	49	47	51	49	51
	29	915	9	11	14	5	6	11
$Z_t \sim N(1, 1)$	191	553	45	43	37	46	40	45
	575	379	10	7	6	7	7	9
$Z_t \sim U(-\frac{1}{2}, \frac{1}{2})$	647	83	47	28	37	50	60	48
	928	15	11	6	9	11	9	11

Para el caso de $N(0,1)$ y $U(-\frac{1}{2}, \frac{1}{2})$ la prueba falla de manera contundente, mientras que para los otros dos casos la cosa mejora. Esto nos motiva a estudiar qué está pasando: si la prueba está mal diseñada, o tan sólo nos encontramos en casos, que aunque no parezca, se parecen a la independencia. Podemos construir un ejemplo más de parejas de variables aleatorias. \square

Ejemplo 4.3.8. Sean X, Y y Z tres variables normales con media μ y varianza σ^2 , independientes. Definamos

$$A = \frac{X}{Y}, \quad B = \frac{Y}{Z}.$$

Notemos primero que no importa la varianza, ya que equivale a multiplicar y dividir por σ , por lo que las podemos considerar con varianza 1. Se estudio por simulaciones los casos $\mu = 0$, $\mu = 1$ y $\mu = 5$, con los siguientes resultados.

Sim. \ $q_{1-\alpha}$.05	.01
$\mu = 0$	89	21
$\mu = 1$	742	408
$\mu = 5$	983	922

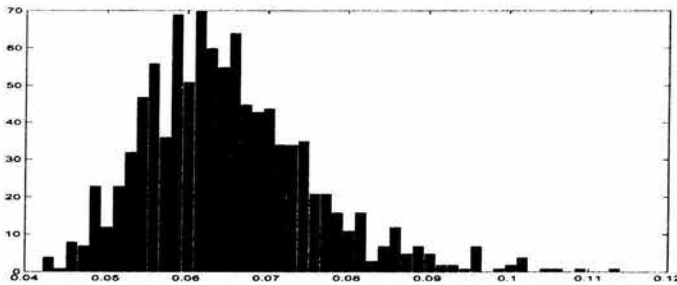


Figura 4.10: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$, $\mu = 0$.

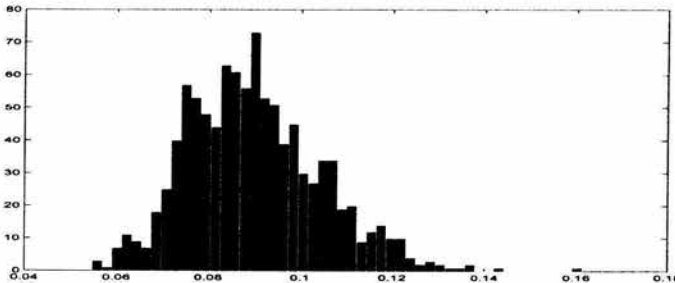


Figura 4.11: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$, $\mu = 1$.

Los histogramas de las Figuras 4.10, 4.11 y 4.12, son muy parecidos excepto que están corridos a la derecha conforme μ aumenta. Al comparar la figura 4.10 con la 4.1, vemos que el comportamiento es parecido, pero no el mismo, mientras una toma valores desde .04, la otra empieza en .03, además de ser

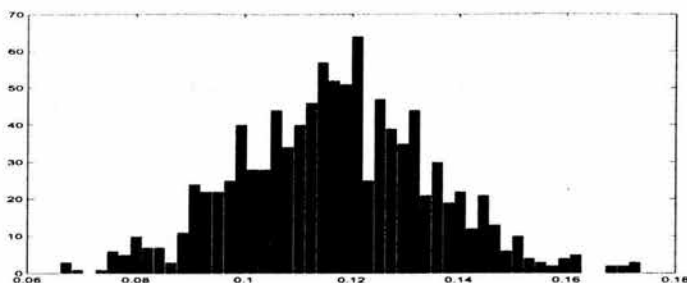


Figura 4.12: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$, $\mu = 5$.

pesada alrededor de .04. Lo que refleja es que el comportamiento de (A, B) es parecido a la independencia en el caso $\mu = 0$. Para tratar de analizar un poco más, dado que sabemos que A y B se distribuyen Cauchy³, calculemos la función de densidad conjunta y tratemos de calcular $\delta(A, B)$, es decir, la real.

Sea $A = \frac{X}{Y}$, $B = \frac{Y}{Z}$ y $C = X$, entonces tenemos que $X = C$, $Y = \frac{C}{A}$ y $Z = \frac{C}{AB}$. El jacobiano de la transformación es

$$J = \begin{vmatrix} 0 & 0 & 1 \\ -\frac{C}{A^2} & 0 & \frac{1}{A} \\ -\frac{C}{A^2B} & -\frac{C}{AB^2} & \frac{1}{AB} \end{vmatrix} = \frac{C^2}{|A^3|B^2},$$

y como X, Y y Z son $N(0,1)$, la densidad conjunta de A, B y C es

$$f(a, b, c) = \frac{c^2}{(2\pi)^{\frac{3}{2}}|a^3|b^2} \exp \left\{ -\frac{c^2}{2} \left(1 + \frac{1}{a^2} + \frac{1}{a^2b^2} \right) \right\}.$$

Para calcular la densidad conjunta de (A, B) , notemos que al integrar con respecto a c , será como calcular la varianza de una variable normal con varianza $\sigma^2 = \frac{a^2b^2}{a^2b^2+b^2+1}$ y por lo tanto, multiplicando y dividiendo por σ , tenemos que

$$f(a, b, c) = \frac{\sigma}{2\pi|a^3|b^2} \frac{c^2}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{c^2}{2\sigma^2} \right\},$$

³La función de densidad está dada por $f(a) = \frac{1}{\pi(a^2+1)}$.

y por lo tanto

$$g(a, b) = \frac{\sigma^3}{2\pi|a^3|b^2}$$

y sustituyendo el valor de σ^2 , obtenemos que la densidad conjunta de (A, B) es

$$g(a, b) = \frac{|b|}{2\pi(a^2b^2 + b^2 + 1)^{\frac{3}{2}}}.$$

Para la función de distribución integremos con respecto a a .

$$\begin{aligned} & \int_{-\infty}^x \frac{|b|}{2\pi(a^2b^2 + b^2 + 1)^{\frac{3}{2}}} da \\ &= \int_{-\infty}^x \frac{|b|}{2\pi|b^3| \left(a^2 + 1 + \frac{1}{b^2}\right)^{\frac{3}{2}}} da \\ &= \frac{1}{2\pi(b^2 + 1)} \left(\frac{x|b|}{\sqrt{x^2b^2 + b^2 + 1}} + 1 \right). \end{aligned}$$

Para calcular la distribución conjunta de (A, B) resta integrar con respecto a b . Lo que queremos calcular es

$$\begin{aligned} \delta(A, B) &= \sup_{(x,y) \in \mathbb{R}^2} \left| \int_{-\infty}^y \frac{1}{2\pi(b^2 + 1)} \left(\frac{x|b|}{\sqrt{x^2b^2 + b^2 + 1}} + 1 \right) db \right. \\ &\quad \left. - \left(\frac{\arctan(x)}{\pi} + \frac{1}{2} \right) \left(\frac{\arctan(y)}{\pi} + \frac{1}{2} \right) \right| \\ &= \sup_{(x,y) \in \mathbb{R}^2} \left| \int_{-\infty}^y \frac{x|b|}{2\pi(b^2 + 1)\sqrt{x^2b^2 + b^2 + 1}} db - \frac{\arctan(x)}{\pi} \left(\frac{\arctan(y)}{\pi} + \frac{\pi}{2} \right) \right| \end{aligned}$$

Dada la complicación de la integral, se procedió a estimarla numéricamente, y se obtuvo $\delta(A, B) \approx .025$. Esto confirma lo dicho antes, la pareja (A, B) , aunque no lo parezca, tiene un comportamiento similar a la independencia si X, Y, Z se distribuyen $N(0,1)$.

Siguiendo el mismo espíritu, se analizó también el caso de las variables A, B cuando X, Y y Z se distribuyen uniforme. Se consideraron los casos $U(0,1)$ y $U(-\frac{1}{2}, \frac{1}{2})$, obteniéndose

Sim. \ $q_{1-\alpha}$.05	.01
$U(0,1)$	999	984
$U(-\frac{1}{2}, \frac{1}{2})$	109	21

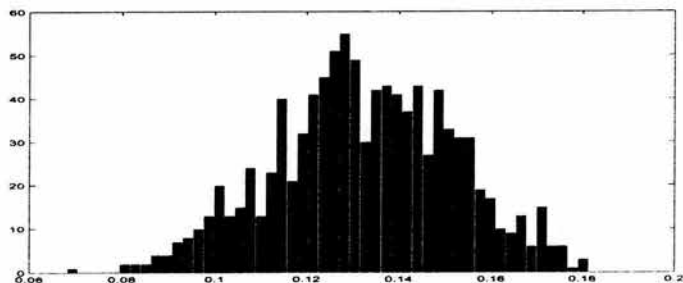


Figura 4.13: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$, $U(0,1)$.

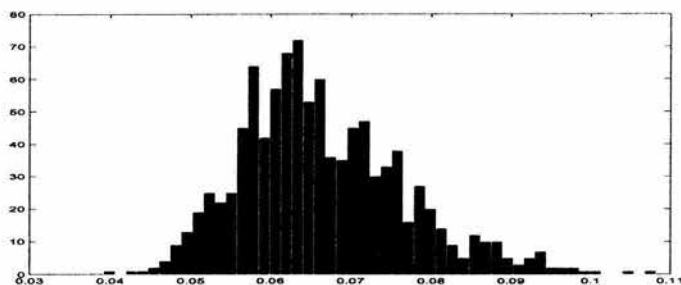


Figura 4.14: Histograma de frecuencias de $\hat{\delta}_{100}(X, Y)$, $U(-\frac{1}{2}, \frac{1}{2})$.

Es de esperarse entonces que las variables, en el caso de $U(-\frac{1}{2}, \frac{1}{2})$, tengan un comportamiento similar a la independencia. \square

Veamos un último ejemplo de un proceso no lineal.

Ejemplo 4.3.9. Sea $\{Z_t\}_{t \in \mathbb{Z}}$ un proceso conformado por variables aleatorias independientes e idénticamente distribuidas. Definamos el proceso $\{X_t\}_{t \in \mathbb{Z}}$ como

$$X_t = Z_t + \theta_1 Z_{t-1}^2 + \dots + \theta_q Z_{t-q}^{q+1},$$

donde θ_i , $i = 1, \dots, q$ son constantes. Sea $\bar{\theta} = (\theta_1, \dots, \theta_q)$. Se realizaron simulaciones para distintas distribuciones de Z_t y para distintos vectores $\bar{\theta}$, y estos son los resultados

Distr.— $\bar{\theta} \setminus q$	0	1	2	3	4	5	6	7
$Z_t \sim U(0,1) - (1)$	2	737	47	39	42	36	51	46
	73	867	16	8	11	10	6	9
$Z_t \sim U(0,1) - (1,1)$	0	273	519	34	35	51	40	48
	0	619	335	6	9	9	9	13
$Z_t \sim U(0,1) - (1,1,1)$	0	23	519	262	39	45	60	52
	0	133	703	123	11	12	10	8
$Z_t \sim N(0,1) - (1)$	293	285	60	59	58	76	82	87
	931	54	1	5	2	2	3	2
$Z_t \sim N(1,1) - (1)$	4	583	48	61	61	86	72	85
	195	785	2	2	3	7	5	1
$Z_t \sim N(-1,1) - (1)$	22	541	55	50	81	78	72	101
	254	723	5	3	4	3	3	5
$Z_t \sim N(5,1) - (1)$	471	98	53	52	78	92	76	80
	961	10	6	7	4	4	5	3
$Z_t \sim N(5,1) - (\frac{1}{5})$	35	535	63	63	78	63	72	91
	415	563	5	3	3	1	5	5
$Z_t \sim U(-\frac{1}{2}, \frac{1}{2}) - (1)$	650	68	45	50	40	44	49	54
	923	20	16	12	9	10	6	4

De la tabla podemos ver que, conforme q aumenta, la dependencia va desapareciendo y la prueba detecta menos veces el nivel correcto de q -dependencia. A diferencia del proceso lineal, aquí la precisión es menor puesto que, al ser las variables Z_t uniformes en el $(0,1)$, al elevarlas a una potencia mayor que 1, su valor decrece bastante y su peso sobre las variables X_t también. Los casos $N(0,1)$ y $U(-\frac{1}{2}, \frac{1}{2})$, a diferencia del $U(0,1)$, fallan desde el primer momento ($q = 1$), y esto, al igual que en el ejemplo 4.3.7, debe tener como explicación que las variables se estén comportando de manera parecida a la independencia. Por último, el caso de la $N(5,1)$ se explica al pensar que para la pareja $(Z_t + Z_{t-1}^2, Z_{t+1} + Z_t^2)$, las partes que dominan son Z_{t-1}^2 y Z_t^2 , independientes, mientras que para la pareja $(Z_t + \frac{1}{5}Z_{t-1}^2, Z_{t+1} + \frac{1}{5}Z_t^2)$, todos los componentes son del mismo orden (alrededor del 5) y por lo tanto es posible identificar su dependencia. \square

Con esto terminamos la parte de simulación. A continuación haremos

una aplicación para el cambio porcentual de los precios del petróleo.

Ejemplo 4.3.10. Se obtuvieron observaciones de la mezcla del Brent del 26 de Mayo de 1983 al 22 de Abril de 2003, siendo un total de 5024 datos. Se generó a partir de estos datos una muestra de 5023 observaciones del cambio porcentual diario, es decir, si $\{P_n\}_{n=1}^{5024}$ son los precios, se calculó

$$X_n = \frac{P_n - P_{n-1}}{P_{n-1}}.$$

Se utilizaron los cambios porcentuales y no los precios diarios o bien las diferencias entre los precios, pues es claro que el precio de mañana depende del precio de hoy, ya que de ahí parte; las diferencias diarias entre los precios no tienen ese problema pero sí tienen el problema de no ser iguales en unidades a lo largo de los 10 años. Por esto se consideró el cambio porcentual.

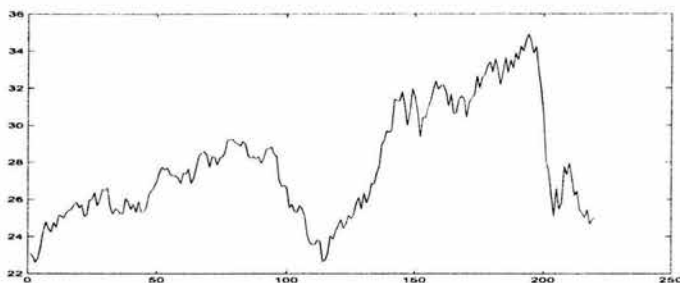


Figura 4.15: Precios diarios del 7 de Junio de 2002 al 14 de Abril de 2003.

Bajo el supuesto de que el proceso $\{X_t\}$ sea de la forma

$$X_t = h(Z_t, Z_{t-1}, \dots, Z_{t-q}),$$

con $\{Z_t\}$ un proceso independiente, se procedió a tratar de identificar un valor de q adecuado. En otras palabras, la pregunta era si se podía identificar un periodo tal que dos observaciones del cambio porcentual en el precio del petróleo fueran independientes si distaban más de ese periodo. Como una aproximación inicial se pretendió estudiar la dependencia del día actual y tres semanas (15 días hábiles) atrás. El tamaño de la muestra general permite generar muestras de parejas (X_n, X_{n+k}) de tamaño 150. Siguiendo el proceso propuesto en la sección anterior, por ejemplo al tomar el salto de tamaño

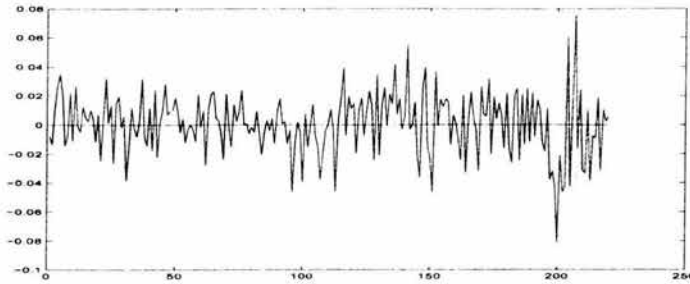


Figura 4.16: Cambio porcentual de los datos de la Figura 4.15.

1, la muestra quedaba de la forma (X_1, X_2) , (X_3, X_4) , etc. Para aprovechar más información se podía generar sólo una muestra más de tamaño 150 y $k = 1$ que no repitiera parejas con la anterior, considerando las parejas de la forma (X_2, X_3) , (X_4, X_5) , etc. Dado que el hacer las muestras de tamaño 150 permite trabajar hasta el retraso 16, se eligieron entonces las parejas de la siguiente manera

$$\begin{aligned}
 &(X_e, X_{e+k}) \\
 &(X_{e+k+s}, X_{e+2k+s}) \\
 &\vdots \\
 &(X_{e+(m-1)(k+s)}, X_{1+mk+(m-1)s}),
 \end{aligned} \tag{4.8}$$

donde e representa el día donde se comienzan las observaciones, k el retraso, es decir, $k = 1, \dots, 16$, $m = 150$ el tamaño de la muestra, y $s = 16$ el salto fijo entre dos observaciones. Con esto se pudo obtener 17 muestras de tamaño 150 desde el retraso 1 al 16 donde ninguna pareja se repite. Para considerar los datos más recientes se tomó e desde 223 a 239. Se aplicó el proceso al nivel $\alpha = .05$, siendo el cuantil $q_{.95} = .0661$. Los resultados fueron los siguientes

Datos \ q	0	1	2	3	4	5	6	7	8
	6	0	1	0	0	0	0	0	0
Datos \ q	9	10	11	12	13	14	15	16	
	0	2	0	2	1	2	2	1	

La tabla anterior⁴ sugiere que el valor $q = 0$ es el más adecuado, es decir, que

⁴La tabla reporta el número de veces que se eligió el valor q de cada columna.

los datos son independientes entre sí. Se calculó también la $\delta_{150}(X_t, X_{t+k})$ promedio ($\bar{\delta}$) obteniéndose

$\bar{\delta} \setminus q$	1	2	3	4	5	6	7	8
	0.0512	0.0512	0.0507	0.0466	0.0471	0.0477	0.0492	0.0499
$\bar{\delta} \setminus q$	9	10	11	12	13	14	15	16
	0.0467	0.0508	0.0451	0.05	0.0471	0.0489	0.0493	0.0484

En ninguno de los casos anteriores se rechaza independencia. Este análisis sugiere que lo más adecuado es decir que $q = 0$ y por lo tanto los datos son independientes. Para tratar de confirmar, se procedió a tomar muestras de tamaño 1000 para dos días consecutivos. Siguiendo la misma notación de antes, los datos permiten para $m = 1000$, tomar e del 23 al 27. En este caso el cuantil del 95% es $q_{.95} = .0266$. Los resultados fueron los siguientes

$\delta \setminus e$	23	24	25	26	27	$\bar{\delta}$
	0.0211	0.0285	0.0296	0.0234	0.0241	0.0253

Esta tabla nos alerta sobre las conclusiones hechas ya que, si bien no se rechazó independencia tres veces, sí se hizo en dos, además de que los valores para los que no se rechazó fueron cercanos al cuantil del 95%. El promedio también queda por debajo del cuantil $q_{.95}$, aunque no demasiado: equivale al cuantil $q_{.92}$. En conclusión, no existe evidencia para rechazar independencia en el cambio porcentual diario de la mezcla del petróleo. A pesar de esto, sería apresurado concluir que existe independencia, nos podemos encontrar en un caso como los vistos en los ejemplos de esta sección.

Conclusiones

La conclusión principal del trabajo es que la dependencia es un tema complicado. Si no conocemos la distribución conjunta y únicamente conocemos alguna de las medidas de concordancia o dependencia, excepto en contados casos, 0 para las medidas de dependencia, 1 y -1 para las de concordancia, absolutamente nada podemos decir sobre cómo es el comportamiento conjunto de las variables aleatorias. Vimos que si para una pareja de variables aleatorias, la correlación es 1 y -1, no sólo nos dice que las variables aleatorias se relacionan por una función monótona, sino también que ésta es una recta y es por esto que se le conoce como una medida de dependencia lineal. Estudiamos muchas de sus ventajas y de sus limitaciones. Su principal ventaja aparece cuando nos encontramos en un modelo normal bivariado. Ahí la correlación funciona a la perfección resumiendo toda la información sobre la dependencia entre las variables aleatorias. De todos modos no debemos perder de vista que el saber que la distribución conjunta es normal, ya nos da mucha información y que la correlación sólo será uno de los parámetros para determinar todo el comportamiento conjunto. Sobre la τ de Kendall podemos concluir que lo que mide es la concordancia que hay entre las variables aleatorias y que no debemos de interpretarla de otra forma, que el decir que un valor más grande que otro, en general se interpreta como un aumento de la dependencia, es un error. La ρ de Spearman la podemos entender como la correlación de rangos, es decir, su esencia es la misma que la de la correlación excepto que de todo el comportamiento conjunto sólo consideramos la cópula. Tanto la τ de Kendall como la ρ de Spearman cumplen todas las propiedades que le pedimos a una medida de concordancia, en especial, los valores 1 y -1 caracterizan a las variables aleatorias que se relacionan entre sí por una función monótona, creciente o decreciente respectivamente. Es importante recordar que ninguna de las medidas anteriores, en general, le otorga un valor exclusivo a la independencia y es por eso que no es lo más adecuado

utilizarlas para estudiar ésta. Para esto es que las medidas de dependencia estudiadas en la tesis son necesarias, siendo ambas, σ y δ , formas bastante naturales de comparar la distribución en cuestión con la de la independencia. Pudimos comprobar que σ cumple todas las propiedades que pedimos a una medida de dependencia, no así δ . Durante el estudio de todas estas medidas fue inevitable observar que las cópulas son una herramienta importantísima en el estudio de la dependencia de las variables aleatorias. Con el Teorema de Sklar, al menos en el caso de variables aleatorias continuas, podemos notar que la cópula resume toda la información conjunta que no proviene de las distribuciones marginales; como dijimos, la función de distribución conjunta, y en consecuencia la dependencia, consta de dos elementos: las distribuciones marginales y la cópula. Todas las medidas estudiadas, excepto la correlación, las podemos ver exclusivamente en términos de la cópula, de igual forma, las propiedades de dependencia quedan resumidas en la misma. Sobre la propiedad de invarianza bajo transformaciones monótonas que le pedimos a las medidas de concordancia y de dependencia, podemos concluir que esto es equivalente a pedir que la medida dependa sólo de la cópula y no de las distribuciones marginales. El decir que en general la dependencia no cambia cuando aplicamos un par de transformaciones monótonas, es lo mismo que decir que la dependencia está concentrada exclusivamente en la cópula y no en la distribución conjunta. Otra propiedad de las medidas de dependencia que cuestionamos al principio es la que dice que el valor extremo debe ser exclusivo para parejas relacionadas por una función monótona. La medida δ no cumple esa propiedad, sin embargo parece ser una medida muy útil para estudiar a la independencia.

En lo que respecta a la parte estadística podemos concluir que el tener una prueba tan universal como la que presentamos aquí, trae consigo el problema de no ser una prueba muy potente. De todos modos, si no tenemos ninguna información extra aparte de los datos, el uso de pruebas de otra índole, como aquellas que utilizan las medidas de concordancia, nos puede llevar a conclusiones erróneas siendo lo más adecuado, en ese caso, utilizar una prueba de este tipo. De todos modos, no debemos perder de vista que existen casos, como vimos en la sección de simulación, sobre los cuales nada podemos decir, o peor aún, tener una probabilidad muy alta de no rechazar independencia y encontrarnos con casos como los reordenamientos de M . Por último, sobre la aplicación del cambio porcentual de los precios del petróleo, parece lo más adecuado concluir únicamente que no se rechaza independencia, sin olvidar que esto no significa que estemos aceptando que hay dependencia.

Apéndice A

Resultados de probabilidad

En este apéndice enunciaremos el Lema de π , λ -sistema y con esto probaremos la equivalencia entre las definiciones de independencia vistas en la Sección 1.2. Además, enunciaremos el Teorema de Glivenko-Cantelli. Se utilizó [1] y [11]. Empecemos con la independencia.

Para ver la equivalencia entre 1.2.4 y 1.2.3, necesitamos definir dos clases de subconjuntos de Ω .

Definición A.0.1 (π -sistema). Sea $\mathcal{C} \subset 2^\Omega$, donde 2^Ω es la potencia de Ω . Decimos que \mathcal{C} es π -sistema, si $A, B \in \mathcal{C}$, entonces $A \cap B \in \mathcal{C}$.

Definición A.0.2 (λ -sistema). Sea $\mathcal{L} \subset 2^\Omega$. Decimos que \mathcal{L} es λ -sistema, si

1. $\Omega \in \mathcal{L}$;
2. $A, B \in \mathcal{L}$ y $A \subset B$ entonces $B \setminus A \in \mathcal{L}$;
3. $(A_n)_{n \in \mathbb{N}} \subset \mathcal{L}$, sucesión creciente de conjuntos, entonces $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{L}$.

Podemos enunciar el Lema de π , λ -sistema, para su demostración véase [1].

Lema A.0.1 (Lema de π , λ -sistema). *Sea \mathcal{C} un π -sistema y \mathcal{L} un λ -sistema, entonces, si $\mathcal{C} \subset \mathcal{L}$ se cumple que $\sigma(\mathcal{C}) \subset \mathcal{L}$.*

Teorema A.0.1. *Si \mathcal{C} y \mathcal{P} son dos π -sistemas independientes¹ entonces $\sigma(\mathcal{C})$ y $\sigma(\mathcal{P})$ son independientes.*

¹La independencia entre π -sistemas la definimos igual que la independencia entre σ -álgebras.

Demostración. Sea $\mathcal{B} = \mathcal{C} \cup \{\Omega\}$ y $\mathcal{A} = \mathcal{P} \cup \{\Omega\}$. Estos siguen siendo claramente π -sistemas. Sea $A \in \mathcal{A}$, definamos $\mathcal{L}_A = \{B \subset \Omega | \mathbb{P}[A \cap B] = \mathbb{P}[A] \mathbb{P}[B]\}$. Probemos que \mathcal{L}_A es un λ -sistema que contiene a \mathcal{B} .

1. $\Omega \in \mathcal{L}_A$ pues claramente $\mathbb{P}[A \cap \Omega] = \mathbb{P}[A] \cdot 1$.

2. Sea $B, C \in \mathcal{L}_A$, con $B \subset C$. Entonces

$$\begin{aligned} \mathbb{P}[A \cap C \setminus B] &= \mathbb{P}[(A \cap C) \setminus (A \cap B)] \\ &= \mathbb{P}[A \cap C] - \mathbb{P}[A \cap B] \\ &= \mathbb{P}[A] (\mathbb{P}[C] - \mathbb{P}[B]) \\ &= \mathbb{P}[A] \mathbb{P}[C \setminus B], \end{aligned}$$

por lo tanto $C \setminus B \in \mathcal{L}_A$.

3. Sea $(B_n) \subset \mathcal{L}_A$, sucesión creciente, entonces

$$\begin{aligned} \mathbb{P}\left[A \cap \bigcup_{n \in \mathbb{N}} B_n\right] &= \mathbb{P}\left[\bigcup_{n \in \mathbb{N}} (A \cap B_n)\right] \\ &= \lim_{n \rightarrow \infty} \mathbb{P}[A \cap B_n] \\ &= \mathbb{P}[A] \lim_{n \rightarrow \infty} \mathbb{P}[B_n] \\ &= \mathbb{P}[A] \mathbb{P}\left[\bigcup_{n \in \mathbb{N}} B_n\right], \end{aligned}$$

por lo tanto $\bigcup_{n \in \mathbb{N}} B_n \in \mathcal{L}_A$.

Por lo tanto, \mathcal{L}_A es λ -sistema. Por hipótesis se cumple que $\mathcal{B} \subset \mathcal{L}_A$, entonces por el Lema A.0.1 tenemos que $\sigma(\mathcal{B}) \subset \mathcal{L}_A$. Como A es cualquier conjunto de \mathcal{A} , entonces tenemos que si $B \in \sigma(\mathcal{B})$, se cumple que $\mathbb{P}[A \cap B] = \mathbb{P}[A] \mathbb{P}[B]$. Sea $B \in \sigma(\mathcal{B})$, definamos $\mathcal{L}_B = \{A \subset \Omega | \mathbb{P}[A \cap B] = \mathbb{P}[A] \mathbb{P}[B]\}$. De la misma manera, podemos probar que \mathcal{L}_B es un λ -sistema que contiene a \mathcal{A} y por lo tanto, por el Lema A.0.1, tenemos que $\sigma(\mathcal{A}) \subset \mathcal{L}_B$ de donde concluimos que si $A \in \sigma(\mathcal{A})$ y $B \in \sigma(\mathcal{B})$, entonces A y B son independientes. Por lo tanto, las σ -álgebras son independientes. \square

Teorema A.0.2. Sean X y Y dos variables aleatorias, H su función de distribución conjunta y F y G las distribuciones marginales. Entonces, X y Y son independientes si y sólo si $H(x, y) = F(x)G(y)$ para todo $x, y \in \mathbb{R}$.

Demostración. Como las variables aleatorias son funciones reales, entonces

$$\sigma(X) = \{A \in \mathcal{F} | A = X^{-1}(B), \text{ con } B \in \mathbb{B}(\mathbb{R})\}$$

y

$$\sigma(Y) = \{A \in \mathcal{F} | A = Y^{-1}(B), \text{ con } B \in \mathbb{B}(\mathbb{R})\}.$$

Los conjuntos

$$\mathcal{C} = \{A \in \mathcal{F} | A = X^{-1}((-\infty, x]), x \in \mathbb{R}\}$$

y

$$\mathcal{P} = \{A \in \mathcal{F} | A = Y^{-1}((-\infty, y]), y \in \mathbb{R}\},$$

son π -sistemas, pues la imagen inversa respeta intersecciones y $(-\infty, a] \cap (-\infty, b] = (-\infty, \min(a, b)]$. Sabemos además que los borelianos pueden ser generados por los intervalos de la forma $(-\infty, a]$, por lo tanto $\sigma(\mathcal{C}) = \sigma(X)$ y $\sigma(\mathcal{P}) = \sigma(Y)$, entonces por el Teorema A.0.1, tenemos que si $H(x, y) = F(x)G(y)$, entonces X y Y son independientes.

La cotraparte es directa del hecho de que los intervalos de la forma $(-\infty, a] \in \mathbb{B}(\mathbb{R})$, lo que completa la prueba. \square

Por lo tanto, por el teorema anterior tenemos que las Definiciones 1.2.4 y 1.2.3 son equivalentes. Veamos otro resultado.

Teorema A.0.3. Sean X y Y dos variables aleatorias, entonces, para toda $f, g : \mathbb{R} \rightarrow \mathbb{R}$ boreliana, tales que $\mathbb{E}[f(X)g(Y)]$, $\mathbb{E}[f(X)]$ y $\mathbb{E}[g(Y)]$ existen, se cumple que

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \mathbb{E}[g(Y)].$$

Demostración. Probemos primero que si las esperanzas existen, entonces $\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]$. Sea H la función de distribución conjunta y F y G las distribuciones marginales. Como X y Y son independientes, entonces $H(x, y) = F(x)G(y)$, aplicando el Teorema de Fubini, tenemos que

$$\begin{aligned} \mathbb{E}[XY] &= \iint xy dF(x, y) = \int \int xy dF(x) dG(y) \\ &= \int x dF(x) \int y dG(y) = \mathbb{E}[X] \mathbb{E}[Y]. \end{aligned}$$

Para probar el resultado, bastará probar que $f(X)$ y $g(Y)$ son independientes y utilizando lo anterior habremos concluido. Pero eso es bastante claro pues $\sigma(f(X)) \subset \sigma(X)$ y $\sigma(g(Y)) \subset \sigma(Y)$, lo que implica que $\sigma(f(X))$ y $\sigma(g(Y))$ son independientes de donde tenemos que $f(X)$ y $g(Y)$ también lo son. Como las esperanzas existen por hipótesis podemos aplicar Fubini y es resultado es válido. \square

Por último, enunciemos el Teorema de Glivenko-Cantelli

Teorema A.0.4 (Teorema de Glivenko-Cantelli). *Sea X una variable aleatoria con función de distribución F y F_n la función de distribución empírica basada en n observaciones independientes de la variable X . Entonces*

$$\sup_{x \in \mathbb{R}} |F(x) - F_n(x)| \xrightarrow[n \rightarrow \infty]{} 0, \text{ c.s.}$$

Apéndice B

Programas

A continuación presentamos los programas utilizados para la parte de simulación. Todos los programas fueron realizados para Matlab.

```
function d=deltamb(x,y)
    % x y y son vectores columna de las parejas de datos
    % Calcula delta para los vectores (x ,y)
    m=sortrows([x y]);
    n=length(x);
    y=sort(y);
    x=sort(x);
    dep=zeros(n,n);
    temp=0.^(abs(y-m(1,2))*ones(n,1));
    t=min(find(temp==1));
    temp=[zeros(t-1,1);ones(n-t+1,1)];
    dep(1,:)=temp';
    for i=2:n
        temp=0.^(abs(y-m(i,2))*ones(n,1));
        t=min(find(temp==1));
        temp=[zeros(t-1,1);ones(n-t+1,1)];
        dep(i,:)=dep(i-1,:)+temp';
    end
    h=find(x(2:n)-x(1:n-1)==0);
    h=flipud(h);
    for i=h'
        dep(i,:)=dep(i+1,:);
    end
```

```

end
ind=distacm(x)*(distacm(y))';
d=max(max(abs(n*dep-ind)))/(n^2);

```

```

function t=distacm(y)
    %devuelve la funcion acumulada para el vector "y" columna
    n=length(y);
    y=sort(y);
    t=sign(y(2:n)-y(1:n-1));
    t=[t;1];
    t=t.*(1:n)';
    h=flipud(find(t==0));
    for j=h'
        t(j)=t(j+1);
    end

```

```

function [x,c]=shufflepi(n,e)
    %genera dos vectores de n parejas de numeros aleatorios
    %de un shuffle que dista de pi menos que e.
    m=ceil(4/e);
    x=rand(n,1);
    c=(x*m^2+(m-1)*(floor(x*m^2-m*floor(x*m))-floor(x*m)))/m^2;

```

Bibliografía

- [1] Billingsley, P. (1995), *Probability and Measure*, (John Wiley & Sons Inc., New York).
- [2] Blum, J.R., Hanson, D.L. y Koopmans, L.H. (1963), "On the strong law of large numbers for a class of stochastic processes", *Z. Wharsch. Verw. Gebiete* **2**, 1-11.
- [3] Bradley, R.C. (1986), "Basic Properties of Strong Mixing Conditions," en *Dependence in Probability and Statistics: A survey in Recent Results*, 165-192, E. Eberlein y M.S. Taqqu, (Eds., Birkhäuser, Boston).
- [4] Brockwell, P.J. y Davis, R.A. (1991), *Time Series: Theory and Methods*, (Springer-Verlag, New York).
- [5] Capéràa, P. y Genest, C. (1993), "Spearman's ρ is larger than Kendall's τ for positively dependent random variables," *J. Nonparametric Statistics* **2**, 183-194.
- [6] Chanda, K.C. (1974), "Strong mixing properties of linear stochastic processes," *J. Appl. Prob.* **11**, 401-408.
- [7] Contreras, A. y González Barrios, J.M. (2002), "A statistical method for the determination of the appropriate order in a general class of time series models", *Preimpreso, IIMAS*.
- [8] Conover, W.J. (1999), *Practical Nonparametric Statistics*, (John Wiley & Sons Inc., New York).
- [9] Embrechts, P., McNeil, A. y Straumann, D. (1999), "Correlation and Dependence in Risk Management: Properties and Pitfalls"

- [10] Fernández, B. y González Barrios, J.M. (2001), "Multidimensional Dependency Measures", *Preimpreso No. 101, IIMAS*.
- [11] Gnedenko, B. (1963), *Theory of Probability*, (Chelsea Publishing Company, USA).
- [12] Kendall, M. (1962), *Rank Correlation Methods*, (Charles Griffin & Company Limited, London).
- [13] Kensten, H. y O'Brien, G.L. (1976), "Examples of mixing sequences," *Duke Math. J.* **43**, 405-415.
- [14] Marsden, J. y Hoffman, M. (1996), *Basic Complex Analysis*, (W.H. Freeman and Company, USA).
- [15] Nelsen, R. (1999) *An Introduction to Copulas*, (Springer-Verlag, New York).
- [16] Ross, S. (1998), *A First Course in Probability*, (Prentice Hall, New Jersey).
- [17] Schweizer, B. y Wolff, E.F. (1981), "On nonparametric measures of dependence for random variables," *Ann. Statistics* **9**, 642-650.
- [18] Shea, G.A. (1983), "Hoeffding's lemma," en *Encyclopedia of Statistical Sciences*, Vol. 3, S. Kotz y N.L. Johnson, editores (John Wiley & Sons, New York), 648-649.