



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS
MATEMÁTICAS Y DE LA ESPECIALIZACIÓN EN
ESTADÍSTICA APLICADA

ESTIMACIÓN DE LA CORRELACIÓN MEDIANTE LA
CÓPULA MUESTRAL

T E S I S

QUE PARA OPTAR POR EL GRADO DE
MAESTRA EN CIENCIAS

PRESENTA
TRINIDAD GONZÁLEZ BONILLA

DIRECTOR DE LA TESIS
Dr. JOSÉ MARÍA GONZÁLEZ-BARRIOS MURGUÍA
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS
APLICADAS Y EN SISTEMAS

MÉXICO D.F., ENERO 2016



Universidad Nacional
Autónoma de México



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.

Agradecimientos

A la Universidad Nacional Autónoma de México por haberme brindado un espacio de desarrollo humano y académico y a las personas que han luchado por mantener la gratuidad de esta casa de estudios.

Al Dr. José María González-Barrios por toda su entrega, su constancia y su incansable labor como profesor e investigador, porque estuvo en todo momento al tanto de los avances de este trabajo y me brindó siempre su apoyo.

A los miembros del jurado: Dra. Ana Meda Guardiola, Dr. Raúl Rueda Díaz del Campo, Dr. Alberto Contreras Cristán y Dr. Pablo Padilla Longoria por su valiosa revisión y comentarios.

Le agradezco a mi madre Alejandra por haber sido madre y padre, por ser mi amiga, mi apoyo y mi confidente, por enseñarme que a pesar de las adversidades la vida es bella y vale la pena vivirla. Porque no hay mejor ejemplo de amor, mamá, que tu vida.

A mis hermanos Lenin y Saúl, porque aún en la distancia me acompaña siempre su cariño.

A Yunuen por ser mi eterna cómplice, mi pilar y ejemplo, con quien puedo compartir por igual alegrías y penas. A Cynthia por la serenidad que me da con su palabra. Gracias por ser las hermanas que la vida me regaló y por todo lo compartido.

A Linda por todo su apoyo, amistad y solidaridad, porque en su enorme corazón siempre hay un huequito para mí, por siempre brindarme su consejo y su cariño.

A mis amigos Aldo, Felipe, Paco, David, Marco, Javier, Dan-Producción, Hugo y Arturo, gracias infinitas por su amistad y por cada risa compartida.

A mis compañeros Roberta, Charlie y Erwin por hacer tan ameno el trabajo.

Y a ti Ricardo, gracias por enseñarme que no hay fecha de caducidad para los sueños, que amor se escribe con cada acción, gracias por caminar conmigo y darme esta hermosa familia que ahora somos. A ti todo mi amor.

Índice

Introducción	I
1. Preliminares	1
1.1. Cópulas	1
1.2. d-cópula muestral de orden m	11
2. Estimación de la correlación de la distribución normal	19
2.1. Distribución normal	19
2.2. Propuesta del estimador	24
3. Resultados	27
3.1. Propiedades del estimador	27
3.2. Robustez	33
3.3. Corrección por Von Misses	40
4. Conclusiones y posibles extensiones	46
Referencias	50

Introducción

Sir Francis Galton definió el concepto teórico de la correlación bivariada en 1885, una década después Karl Pearson publicó la fórmula para el coeficiente de correlación muestral, también conocida como la ρ de Pearson. El coeficiente de correlación muestral es aún la medida más común de correlación ya que no asume información sobre las medias y las varianzas y es el estimador de máxima verosimilitud para el coeficiente de correlación de la distribución normal bivariada.

El coeficiente de correlación muestral de Pearson que hasta ahora se usa tiene varios inconvenientes, uno de ellos es que no es un estimador robusto.

En el presente trabajo se usa la definición de d -cópula muestral propuesta por José M. González y María M. Hernández en el artículo *The sample d -cópula of order m* [7] para definir un estimador de la correlación que resulta ser robusto.

El trabajo está compuesto por tres capítulos, en el primero se revisan los conceptos básicos de cópulas, la definición de d -cópula muestral y los principales teoremas y resultados relacionados con estos temas.

En el segundo capítulo se estudian las propiedades de la distribución normal y, mediante la d -cópula muestral, se propone un estimador de la ρ .

En el tercer capítulo se hace un análisis del estimador y sus propiedades, se demuestra que es asintóticamente insesgado y también se demuestra que es un estimador robusto, finalmente se propone una corrección para que sea insesgado en un intervalo alrededor del cero. En este último capítulo se muestran ejemplos de simulación comparando las estimaciones obtenidas con las usuales de correlación muestral de Pearson.

En el último capítulo se dan algunas conclusiones y posibles extensiones para dimensión tres proponiendo un par de métodos de estimación de los parámetros.

1. Preliminares

1.1. Cópulas

Sean X_1, X_2, \dots, X_d , $d \geq 2$, variables aleatorias continuas con funciones de distribución F_1, \dots, F_d , respectivamente. La función

$$H(x_1, x_2, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d), \quad x_1, \dots, x_d \in \mathbb{R}$$

es la distribución conjunta de las variables aleatorias X_1, X_2, \dots, X_d , la caracterización de la estructura de dependencia de sus marginales está dada por la función C del vector (U_1, \dots, U_d) , llamada cópula, con $U_j = F_j(X_j) \sim U(0, 1)$ para $j = 1, 2, \dots, d$.

Definición 1.1.1 Una d -cópula es una función $C : [0, 1]^d \rightarrow [0, 1]$ para un entero $d \geq 2$ que satisface:

- $C(u_1, \dots, u_d) = 0$ si existe al menos un entero $i \in \{1, \dots, d\}$ tal que $u_i = 0$
- $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ para todo $i \in \{1, \dots, d\}$ y para toda $u_i \in [0, 1]$
- C es una función d -creciente, es decir, para cualquier d -caja, definida como $R = \prod_{i=1}^d [u_i, v_i]$ tal que $R \subset [0, 1]^d$, se cumple que

$$V_C(R) := \sum_{\{\underline{c} \in [0, 1]^d \mid \underline{c} \in \text{Vert}(R)\}} \text{sgn}(\underline{c}) C(\underline{c}) \geq 0, \quad (1)$$

donde

$$\text{sgn}(\underline{c}) = \begin{cases} 1 & \text{si } c_i = u_i \text{ para un número par de } i\text{'s} \\ -1 & \text{si } c_i = u_i \text{ para un número impar de } i\text{'s.} \end{cases}$$

y $\text{Vert}(R)$ es el conjunto de vértices de R .

Por lo tanto, C es una d -cópula si y solo si C es la restricción en $[0, 1]^d$ de una función de distribución de un vector d -dimensional, $\underline{U} = \langle U_1, \dots, U_d \rangle$ con funciones de distribución marginales uniformes $U(0, 1)$. En particular, si $d = 2$ es común decir que C es cópula, en lugar de decir que es una 2-cópula.

Definición 1.1.2 Se definen W y M , dos cópulas, llamadas cotas inferior y superior de Fréchet-Hoeffding, respectivamente

$$\begin{aligned} W(u, v) &= \max\{u + v - 1, 0\} \text{ y} \\ M(u, v) &= \min\{u, v\} \quad \forall \langle u, v \rangle \in [0, 1]^2, \end{aligned}$$

que satisfacen $W(u, v) \leq C(u, v) \leq M(u, v)$ para todo $\langle u, v \rangle \in [0, 1]^2$.

Para dimensiones mayores, $d > 2$, también se pueden definir las cotas de Fréchet-Hoeffding W^d y M^d como

$$\begin{aligned} W^d(u_1, \dots, u_d) &= \max\{u_1 + \dots + u_d - d + 1, 0\} \\ M^d(u_1, \dots, u_d) &= \min\{u_1, \dots, u_d\}. \end{aligned}$$

Que, de manera análoga, satisfacen

$$W^d(u_1, \dots, u_d) \leq C(u_1, \dots, u_d) \leq M^d(u_1, \dots, u_d) \quad \forall \langle u_1, \dots, u_d \rangle \in [0, 1]^d.$$

Es importante notar que cuando $d > 2$, W^d no es una d -cópula, a diferencia de M^d , que es siempre una d -cópula, ver [10].

Un resultado importante que relaciona una función de distribución continua H con sus marginales es el Teorema de Sklar.

Teorema 1.1.3 (Teorema de Sklar) Sea H una d -función de distribución con marginales F_1, F_2, \dots, F_d . Entonces existe una d -cópula C tal que para toda $\underline{x} = \langle x_1, \dots, x_d \rangle \in \bar{\mathbb{R}}^d$

$$H(x_1, x_2, \dots, x_d) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)),$$

si F_1, F_2, \dots, F_d son continuas, entonces C es única; de lo contrario, C es única solo en $\text{Ran}F_1 \times \text{Ran}F_2 \times \dots \times \text{Ran}F_d$. De igual forma, si C es una cópula y F_1, F_2, \dots, F_d son funciones de distribución, entonces la función H , definida como $H(x_1, x_2, \dots, x_d) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d))$ es una función de distribución conjunta con marginales F_1, F_2, \dots, F_d .

Definición 1.1.4 Sea X una variable aleatoria con función de distribución F_X y sea $\langle x_1, \dots, x_n \rangle$ una muestra aleatoria de X . La función de distribución empírica se define como

$$F_n(x) = \sum_{i=1}^n \frac{1}{n} \mathbb{I}_{(-\infty, x]}(x_i),$$

es decir, $F_n(x)$ cuenta el número de observaciones x_i tales que $x_i \leq x$ y lo divide entre el tamaño de muestra.

La función de distribución empírica tiene las siguientes propiedades:

1. Es una función escalonada, la magnitud de los saltos es mayor o igual a $\frac{1}{n}$ en cada x_i .
2. Si $x_{[1, n]} \leq x_{[2, n]} \leq \dots \leq x_{[n, n]}$ son los estadísticos de orden de la muestra aleatoria $\langle x_1, \dots, x_n \rangle$, es decir, la muestra ordenada, entonces F_n es una función de distribución discreta que tiene saltos de magnitud $\frac{1}{n}$ en cada $x_{[i, n]}$, con $i \in \{1, 2, \dots, n\}$, si F_X es continua.

Teorema 1.1.5 (Teorema de Glivenko-Cantelli) Sea X una variable aleatoria con función de distribución F y sea $F_n(x)$ la función de distribución empírica de una muestra aleatoria $\langle x_1, \dots, x_n \rangle$ de tamaño n de la variable aleatoria X . Entonces

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0 \quad \text{c.s. con respecto a } \mathbb{P},$$

donde \mathbb{P} es la medida de probabilidad en $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ inducida por F .

Definición 1.1.6 Sea (X, Y) un vector aleatorio con función de distribución conjunta H y sea $\{(x_i, y_i)\}_{i=1}^n$ una muestra aleatoria de H . La función de distribución empírica bivariada H_n se define como

$$H_n : \mathbb{R}^2 \rightarrow \mathbb{R} \quad \text{tal que} \quad H_n(x, y) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{(-\infty, x] \times (-\infty, y]}(x_i, y_i).$$

Definición 1.1.7 (Función de Rango) Sean X_1, \dots, X_n variables aleatorias y $X_{(1)}, \dots, X_{(n)}$ sus estadísticas de orden. Se define la función de rango $R : \{1, 2, \dots, n\} \times \mathbb{R}^n \rightarrow \{1, 2, \dots, n\}$ como

$$R(j, X_1, \dots, X_n) = k, \quad \text{si} \quad X_j = X_{(k)} \quad (j, k \in \{1, 2, \dots, n\}, X_1, \dots, X_n \in \mathbb{R}).$$

Definición 1.1.8 (Estadísticas de Rango) Sean $\{X_1, \dots, X_n\}$ una muestra aleatoria con $X_i = (X_{i1}, \dots, X_{id}) \in \mathbb{R}^d$, $i = 1, \dots, n$. Sea $i \in \{1, \dots, n\}$, se define la i -ésima estadística de rango $Y_i = (Y_{i1}, \dots, Y_{id})$, como

$$Y_{ij} = \frac{1}{n} R(i, X_{1j}, \dots, X_{nj}) \quad (j = 1, \dots, d).$$

Al conjunto $\{Y_1, \dots, Y_n\}$, donde $Y_i = (Y_{i1}, \dots, Y_{id}) \in [0, 1]^d$, se le conoce como la muestra modificada.

Definición 1.1.9 (Cópula empírica) Se define la cópula empírica asociada a la muestra X_1, \dots, X_n en \mathbb{R}^d como

$$C_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{Y_{i1} \leq x_1, \dots, Y_{id} \leq x_d\}}(x) \quad (x = (x_1, \dots, x_d) \in [0, 1]^d). \quad (2)$$

La cópula empírica es la función de distribución empírica de la muestra modificada.

Definición 1.1.10 (Cópula Empírica Modificada)

$$C_n(U_1, \dots, U_d) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbb{I}_{[F_{j,n}(X_{ij}) \leq u_j]},$$

donde $F_{j,n}$ es $\frac{n}{n+1}$ veces la función empírica de X_j basada en la muestra aleatoria $(x_{11}, \dots, x_{1d}), \dots, (x_{n1}, \dots, x_{nd})$ de la distribución H , C_n es, de hecho, una función de los rangos de estas distribuciones, pues

$$(n+1)F_{j,n}(X_{i,j}) = R_{ij} = \sum_{l=1}^n \mathbb{I}_{[X_{lj} \leq X_{ij}]},$$

con $1 \leq i \leq n$ y $1 \leq j \leq d$ y tal que $C_n(1, \dots, 1) = 1$.

Observación 1.1.11 Se debe notar que al hacer la modificación propuesta por Genest [6] los saltos coinciden con los valores esperados, es decir, si (U_1, \dots, U_n) son variables aleatorias tales que $U_i \sim U(0, 1)$ para toda $i \in \{1, \dots, n\}$ y si $U_{[1]} \leq \dots \leq U_{[n]}$ denota las estadísticas de orden, entonces $\mathbb{E}(U_{[i]}) = \frac{i}{n+1}$.

Ejemplo 1.1.12 Transformación de Rango y Cópula empírica modificada. Sea $\{(-1, 2), (-2, 0), (0, 4), (1, 3), (-3, 1)\}$ una muestra aleatoria.

P_i	(R_{1i}, R_{2i})	(F_{1i}, F_{2i})
$(-1, 2)$	$(3, 3)$	$(\frac{3}{6}, \frac{3}{6})$
$(-2, 0)$	$(2, 1)$	$(\frac{2}{6}, \frac{1}{6})$
$(0, 4)$	$(4, 5)$	$(\frac{4}{6}, \frac{5}{6})$
$(1, 3)$	$(5, 4)$	$(\frac{5}{6}, \frac{4}{6})$
$(-3, 1)$	$(1, 2)$	$(\frac{1}{6}, \frac{2}{6})$

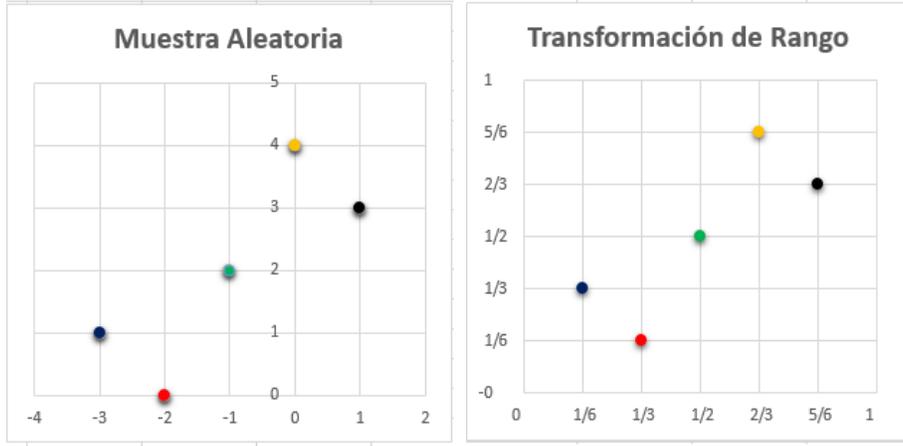


Figura 1: Transformación de Rango

Definición 1.1.13 (Matriz de Transformación) Una matriz de transformación es una matriz real no negativa $T_{n \times m} = (t_{ij})_{\langle i, j \rangle \in I_n \times I_m}$, donde $I_n = \{1, \dots, n\}$, tal que $\max\{n, m\} \geq 2$, $\sum_{i,j} t_{ij} = 1$, $\sum_{i \in I_n} t_{ij} > 0 \quad \forall j \in I_m$ y $\sum_{j \in I_m} t_{ij} > 0 \quad \forall i \in I_n$, es decir, la suma de cualquier renglón o cualquier columna es positiva.

Se definen dos particiones de $[0, 1]$, $\{p_0, p_1, \dots, p_n\}$ y $\{q_0, q_1, \dots, q_m\}$, donde $p_0 = q_0 = 0$, y para $i \in I_n$, sea $p_i = \sum_{i'=1}^i \sum_{j \in I_m} t_{i'j}$, y para $j \in I_m$ sea $q_j = \sum_{j'=1}^j \sum_{i \in I_n} t_{ij'}$.

Sea

$$R_{ij} = (p_{i-1}, p_i] \times (q_{j-1}, q_j] \quad \forall \langle i, j \rangle \in I_n \times I_m,$$

donde si $i = 1$ o $j = 1$ se toman intervalos cerrados en lugar de intervalos abiertos por la derecha. Se observa que $\{R_{ij}\}_{\langle i, j \rangle \in I_n \times I_m}$ es una partición de I^2 .

Sea C una cópula, se define la transformación $T(C)$ usando la partición de I^2 y la matriz de transformación T , donde para cada $\langle i, j \rangle \in I_n \times I_m$, $T(C)$ reparte la masa t_{ij} en R_{ij} reescalando toda la masa de C , esto es, si $\langle u, v \rangle \in R_{ij}$ sea

$$T(C)(u, v) = \sum_{i' < i, j' < j} t_{i'j'} + \frac{u - p_{i-1}}{p_i - p_{i-1}} \sum_{j' < j} t_{ij'} + \frac{v - q_{j-1}}{q_j - q_{j-1}} \sum_{i' < i} t_{i'j} + t_{ij} C \left(\frac{u - p_{i-1}}{p_i - p_{i-1}}, \frac{v - q_{j-1}}{q_j - q_{j-1}} \right),$$

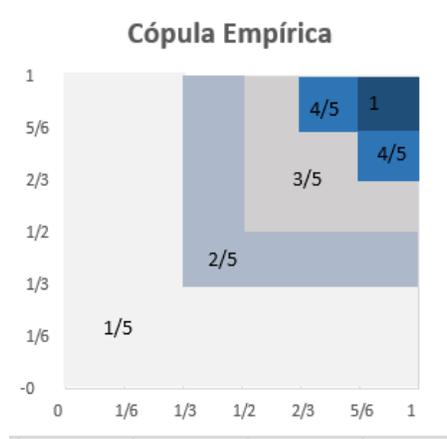


Figura 2: Cópula Empírica

donde la sumas vacías se definen como cero. Entonces $T(C)$ es siempre una cópula. Se define iterativamente

$$T^2(C) = T(T(C)) \quad \text{y} \quad T^{n+1}(C) = T(T^n(C)) \quad \forall n > 2.$$

De hecho $T^n(C) = (\otimes^n T)(C)$, donde \otimes^n es el producto tensorial [2] de T con el mismo n veces. Se puede verificar, mediante inducción, que si T es una matriz de transformación de orden $n \times m$ entonces $\otimes^k T$ también lo es pero de orden $n^k \times m^k$ para toda $k \geq 2$. Entonces se tiene que para cualquier matriz de transformación T existe una única cópula C_T , tal que $T(C_T) = C_T$. Más aún, $C_T = \lim_{n \rightarrow \infty} T^n(C)$ para cualquier cópula C . Como C_T no depende de la cópula C se puede restringir al límite de la sucesión $\{T^n(\Pi)\}_{n \geq 1}$, de hecho se dice que C es invariante si $C = C_T$ para alguna matriz de transformación T .

También se puede observar que si $\pi_1 = \{p_0, p_1, \dots, p_n\}$ y $\pi_2 = \{q_0, q_1, \dots, q_m\}$ son particiones cualesquiera del intervalo $[0, 1]$ y se define $t_{ij} = (p_i - p_{i-1})(q_j - q_{j-1})$ para todo $\langle i, j \rangle \in I_n \times I_m$, entonces $T = (t_{ij})_{\langle i, j \rangle \in I_n \times I_m}$ es una matriz de transformación que genera las particiones π_1 y π_2 y se cumple $C_T = \Pi^2$, la cópula producto. Estos resultados están dados en Fredericks y Nelsen [5].

Se dice que una matriz real cuadrada $\mathbf{P} = (P_{ij})_{i,j=1}^k$, $k \geq 2$, es una matriz doblemente estocástica si y solo si $p_{ij} \geq 0$ y $\sum_{j=1}^k p_{ij} = \sum_{i=1}^k p_{ij} = 1$ para toda $i, j \in \{1, 2, \dots, k\}$.

Los siguientes resultados aparecen en González-Barrios y Hernández-Cedillo [7].

Sea m.t. la notación para matriz de transformación, se define

$\mathcal{T} = \{T_{n \times m} | T_{n \times m} \text{ es una m.t. } n, m \geq 2, t_{ij} \in \mathbb{Q} \forall \langle i, j \rangle \in I_n \times I_m\}$.

Lema 1.1.14 Sea $T_{n \times m} \in \mathcal{T}$ entonces existe $k \geq 2$ y $P_{k \times k} = (p_{ij})_{i,j=1}^k$, una matriz doblemente estocástica tal que si definimos $S_{k \times k} = (\frac{p_{ij}}{k})_{i,j=1}^k$ entonces $T(\Pi^2) = S(\Pi^2)$.

Ejemplo 1.1.15 Como un ejemplo sencillo del lema anterior, se considera la matriz de transformación

$$T = \begin{pmatrix} 0 & \frac{1}{3} \\ \frac{2}{3} & 0 \end{pmatrix}.$$

Entonces, si se define

$$S = \begin{pmatrix} 0 & 0 & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{6} & 0 \\ \frac{1}{6} & \frac{1}{6} & 0 \end{pmatrix},$$

se tiene que $P = 3 \cdot S$ es una matriz doblemente estocástica y $T(\Pi^2) = S(\Pi^2)$.

Una matriz cuadrada d -dimensional \mathbf{P} , para $d \geq 2$ es un arreglo de números reales de la forma $\mathbf{P} = (p_{i_1 i_2 \dots i_d})_{i_1, i_2, \dots, i_d=1}^k$ para alguna $k \geq 2$. Se dice que \mathbf{P} es una **matriz estocástica de dimensión d** si y solo si $0 \leq p_{i_1 i_2 \dots i_d} \leq 1$ para todo $i_1, i_2, \dots, i_d \in \{1, 2, \dots, k\}$ y para todo $1 \leq j_1 < j_2 < \dots < j_{d-1} \leq d$ sucede que

$$\sum_{i_{j_1}=1}^k \sum_{i_{j_2}=1}^k \dots \sum_{i_{j_{d-1}}=1}^k p_{i_1 i_2 \dots i_d} = 1,$$

donde el índice faltante es fijo y tomado en $\{1, \dots, k\}$. Se tiene que

$$\sum_{i_{j_1}=1}^k \sum_{i_{j_2}=1}^k \dots \sum_{i_{j_d}=1}^k p_{i_1 i_2 \dots i_d} = k.$$

Y, para este caso, se cumple que una matriz estocástica de dimensión 2 es una matriz doblemente estocástica. Sea $C_{i_1, i_2, \dots, i_d}^1 = \frac{p_{i_1 i_2 \dots i_d}}{k}$ para todo $i_1, i_2, \dots, i_d \in \{1, \dots, k\}$, y para toda $A \in \mathcal{B}([0, 1]^d)$ y para toda $n \geq 1$ se define

$$\begin{aligned} & \mu_n(A) \tag{3} \\ &= k^{dn} \sum_{i_1, \dots, i_d=1}^{k^n} C_{i_1, \dots, i_d=1}^n \lambda^d \left(\left(\left[\frac{i_1-1}{k^n}, \frac{i_1}{k^n} \right] \times \dots \times \left[\frac{i_d-1}{k^n}, \frac{i_d}{k^n} \right] \right) \cap A \right), \end{aligned}$$

para todo $i_1, i_2, \dots, i_d \in \{1, \dots, k\}$ para todo $i'_1, i'_2, \dots, i'_d \in \{1, \dots, k\}$ y para toda $n \geq 1$,

$$C_{k(i_1-1)+i'_1, \dots, k(i_d-1)+i'_d}^{n+1} = \frac{p_{i'_1 \dots i'_d}}{k^{d-1}} \cdot C_{i_1, \dots, i_d}^n.$$

Proposición 1.1.16 Sea $d \geq 2$, sea \mathbf{P} una matriz cuadrada de dimensión d de orden $k \geq 2$, la cual es una matriz estocástica de dimensión d . Sea $n \geq 1$ y se define μ_n como en la ecuación 3, entonces $([0, 1]^d, \mathcal{B}([0, 1]^d), \mu_n)$ es un espacio de probabilidad. Además, si se define

$$C_n(u_1, \dots, u_d) = \mu_n([0, u_1] \times \dots \times [0, u_d]) \text{ para todo } u_1, \dots, u_d \in [0, 1].$$

Entonces C_n es una d -cópula para toda $n \geq 1$. Si se define $\mu_{\mathbf{P}} = \lim_{n \rightarrow \infty} \mu_n$, entonces $\mu_{\mathbf{P}}$ existe respecto a la convergencia débil y es una medida de probabilidad en $([0, 1]^d, \mathcal{B}([0, 1]^d))$. Aún más, la medida $\mu_{\mathbf{P}}$ induce una d -cópula $C_{\mathbf{P}}$ definida como

$$C_{\mathbf{P}} = \mu_{\mathbf{P}}([0, u_1] \times \dots \times [0, u_d]) \text{ para todo } u_1, \dots, u_d \in [0, 1], \quad (4)$$

si \mathbf{P} incluye ceros entonces $\mu_{\mathbf{P}}$ es una medida singular.

En el caso de $d = 2$ con $0 < a < 1$

$$\mathbf{P} = \begin{pmatrix} \frac{a}{2} & \frac{1-a}{2} \\ \frac{1-a}{2} & \frac{a}{2} \end{pmatrix}.$$

Entonces $2\mathbf{P}$ es una matriz doblemente estocástica y $C_{\mathbf{P}}$ es una cópula singular si $a \neq \frac{1}{2}$, ver Cuculescu and Theodorescu [3].

Teorema 1.1.17 Sea C una d -cópula para alguna $d \geq 2$, entonces para toda $\epsilon > 0$ existe \mathbf{P} , una matriz estocástica de dimensión d tal que si se construye la cópula $C_{\mathbf{P}}$ definida como en la ecuación (4), entonces

$$d_{sup}(C, C_{\mathbf{P}}) = \sup_{u_1, \dots, u_d \in [0, 1]} |C(u_1, \dots, u_d) - C_{\mathbf{P}}(u_1, \dots, u_d)| < \epsilon$$

A continuación se dará una generalización de matrices de transformación en dimensión d encontrado en Trutschnig y Fernández-Sánchez [11].

Sea $I_n = \{1, 2, \dots, n\}$ para $n \geq 1$. Para $d \geq 2$ sean $m_1, \dots, m_d \in \mathbb{N}$, se define $\mathcal{I}^d = \prod_{i=1}^d I_{m_i}$. Sea τ una medida de probabilidad en $(\mathcal{I}^d, 2^{\mathcal{I}^d})$, entonces llamamos a τ **una matriz de transformación generalizada** si para todo $j \in \{1, \dots, d\}$ y para toda $k \in \{1, \dots, m_j\}$

$$\sum_{\mathbf{i} \in \mathcal{I}^d, i_j = k} \tau(\mathbf{i}) > 0, \quad (5)$$

donde $\mathbf{i} = \langle i_1, \dots, i_{j-1}, i_j = k, i_{j+1}, \dots, i_d \rangle \in \mathcal{I}^d$.

Observe que la ecuación (5) es una extensión natural de las condiciones de transformación de matrices en Fredricks [5] y que τ puede ser escrita como una matriz T de dimensión d , usando que

$$\tau(\mathbf{i}) = t_{i_1, i_2, \dots, i_d} \text{ si } \mathbf{i} = \langle i_1, i_2, \dots, i_d \rangle \in \mathcal{I}^d. \quad (6)$$

En lo que resta de esta sección solo se usará $m_1 = m_2 = \dots = m_d = m \geq 2$. En cuyo caso, si todos los valores $\tau(\mathbf{i})$ son racionales, se observa que la ecuación (5) es equivalente a decir que τ induce la existencia de una $m_0 \geq m$ y una matriz cuadrada de dimensión d , T_0 tal que $m_0 \cdot T_0$ es una matriz estocástica de dimensión d . Esto es una consecuencia de la extensión multivariada del Lema 1.1.14 para cualquier medida de probabilidad τ con valores racionales en $(\mathcal{I}^d, 2^{\mathcal{I}^d})$ como en la ecuación (6).

Sea $m \geq 2$ y para toda $\mathbf{i} = \langle i_1, \dots, i_d \rangle \in \mathcal{I}_m := \prod_{j=1}^d I_m$ se define

$$R_{\mathbf{i}} = \left(\frac{i_1 - 1}{m}, \frac{i_1}{m} \right] \times \left(\frac{i_2 - 1}{m}, \frac{i_2}{m} \right] \times \dots \times \left(\frac{i_d - 1}{m}, \frac{i_d}{m} \right], \quad (7)$$

donde si para alguna $j \in \{1, \dots, d\}$, $i_j = 1$, entonces se toman intervalos cerrados en lugar de los intervalos abiertos por la izquierda. Entonces $\{R_{\mathbf{i}}\}_{\mathbf{i} \in \mathcal{I}_m}$ es una partición de $[0, 1]^d$ llamada **partición uniforme de orden m** de $[0, 1]^d$.

Sea C una d -cópula y se define para toda $\mathbf{i} = \langle i_1, \dots, i_d \rangle \in \mathcal{I}_m$ una m -partición

$$t_{i_1, i_2, \dots, i_d} = V_C(R_{\mathbf{i}}) \text{ y } T^C = (t_{i_1, i_2, \dots, i_d})_{i_1, i_2, \dots, i_d=1}^m. \quad (8)$$

Entonces T^C es una matriz cuadrada de dimensión d con entradas no negativas que genera la aproximación de tablero (*checkerboard approximation*) [9], además es una matriz de transformación generalizada, pues en la Ecuación (6), si se toma cualquier $j \in \{1, \dots, d\}$ y cualquier $k \in \{1, \dots, m\}$ por la definición de d -cópula se tiene

$$\begin{aligned}
\sum_{\mathbf{i} \in \mathcal{I}_m, i_j=k} \tau(\mathbf{i}) &= \sum_{i_1=1}^m \cdots \sum_{i_{j-1}=1}^m \sum_{i_{j+1}=1}^m \cdots \sum_{i_d=1}^m V_C(R_{\langle i_1, \dots, i_{j-1}, k, i_{j+1}, \dots, i_d \rangle}) \quad (9) \\
&= V_C \left([0, 1] \times \cdots \times [0, 1] \times \left[\frac{k-1}{m}, \frac{k}{m} \right] \times [0, 1] \times \cdots \times [0, 1] \right) \\
&= C(1, \dots, 1, \frac{k}{m}, 1, \dots, 1) - C(1, \dots, 1, \frac{k-1}{m}, 1, \dots, 1) \\
&= \frac{1}{m} > 0.
\end{aligned}$$

Se observa que en la ecuación anterior para toda d -cópula C , para toda $j \in \{1, \dots, d\}$ y para toda $k \in \{1, \dots, m\}$, $\sum_{\mathbf{i} \in \mathcal{I}_m, i_j=k} \tau(\mathbf{i}) = \frac{1}{m}$ solo depende de m .

También se observa que $m \cdot T^C$ es una matriz estocástica cuadrada de dimensión d .

Ahora, si $T = (t_{i_1, \dots, i_d})_{i_1, \dots, i_d=1}^m$ es una matriz de transformación generalizada cuadrada de dimensión d , definimos $p_{1,0} = p_{2,0} = \cdots = p_{d,0} = 0$ y para cada $j \in \{1, \dots, d\}$ y para toda $k \in \{1, \dots, m\}$ se define

$$p_{j,k} = \sum_{i_j=1}^k \sum_{i_1=1}^m \cdots \sum_{i_{j-1}=1}^m \sum_{i_{j+1}=1}^m \cdots \sum_{i_d=1}^m t_{i_1, \dots, i_d}. \quad (10)$$

Entonces $0 = p_{j,0} < p_{j,1} < \cdots < p_{j,m-1} < p_{j,m} = 1$ es la partición con $m+1$ puntos de $[0, 1]$ inducida por T , la cual corresponde a la j -ésima coordenada. Sea C una d -cópula, se define $T(C)(u_1, \dots, u_d)$ para $\langle u_1, \dots, u_d \rangle \in [0, 1]^d$ usando una fórmula similar a la usada para dos dimensiones en la Ecuación (3), entonces $T(C)$ es siempre una d -cópula. En particular si $C = \Pi^d$ la d -cópula producto, $T(\Pi)(u_1, \dots, u_d)$ tiene una expresión mucho más simple. Por ejemplo, si $d = 3$ y $\langle u_1, u_2, u_3 \rangle \in \mathcal{R}_{\langle i_1, i_2, i_3 \rangle} = \mathcal{R}_{\mathbf{i}} = (p_{1, i_1-1}, p_{1, i_1}] \times (p_{2, i_2-1}, p_{2, i_2}] \times (p_{3, i_3-1}, p_{3, i_3}]$ para alguna $\mathbf{i} \in \mathcal{I}_m$ entonces

$$\begin{aligned}
T(\Pi)(u_1, u_2, u_3) &= \sum_{i < i_1, j < i_2, k < i_3} t_{i,j,k} \\
&+ \left(\frac{u_1 - p_{1,i_1-1}}{p_{1,i_1} - p_{1,i_1-1}} \right) \sum_{j < i_2, k < i_3} t_{i_1,j,k} \\
&+ \left(\frac{u_2 - p_{2,i_2-1}}{p_{2,i_2} - p_{2,i_2-1}} \right) \sum_{i < i_1, k < i_3} t_{i,i_2,k} \\
&+ \left(\frac{u_3 - p_{3,i_3-1}}{p_{3,i_3} - p_{3,i_3-1}} \right) \sum_{i < i_1, j < i_2} t_{i,j,i_3} \\
&+ \left(\frac{u_1 - p_{1,i_1-1}}{p_{1,i_1} - p_{1,i_1-1}} \right) \left(\frac{u_2 - p_{2,i_2-1}}{p_{2,i_2} - p_{2,i_2-1}} \right) \sum_{k < i_3} t_{i_1,i_2,k} \\
&+ \left(\frac{u_1 - p_{1,i_1-1}}{p_{1,i_1} - p_{1,i_1-1}} \right) \left(\frac{u_3 - p_{3,i_3-1}}{p_{3,i_3} - p_{3,i_3-1}} \right) \sum_{j < i_2} t_{i_1,j,i_3} \\
&+ \left(\frac{u_2 - p_{2,i_2-1}}{p_{2,i_2} - p_{2,i_2-1}} \right) \left(\frac{u_3 - p_{3,i_3-1}}{p_{3,i_3} - p_{3,i_3-1}} \right) \sum_{j < i_2} t_{i,i_2,i_3} \\
&+ t_{i_1,i_2,i_3} \left(\frac{u_1 - p_{1,i_1-1}}{p_{1,i_1} - p_{1,i_1-1}} \right) \left(\frac{u_2 - p_{2,i_2-1}}{p_{2,i_2} - p_{2,i_2-1}} \right) \left(\frac{u_3 - p_{3,i_3-1}}{p_{3,i_3} - p_{3,i_3-1}} \right).
\end{aligned}$$

Se observa que $T(\Pi^3)$ es una 3-cópula que asigna masa uniforme t_{i_1,i_2,i_3} a cada caja $R_{\langle i_1, i_2, i_3 \rangle}$ para todo $\langle i_1, i_2, i_3 \rangle \in \{1, \dots, m\}$. Por lo tanto, la matriz de transformación generalizada T se puede ver como la densidad ponderada de la cópula $T(\Pi^3)$, dada por $t_{\mathbf{i}}/\lambda^d(R_{\mathbf{i}})$ para todo $\mathbf{i} \in \mathcal{I}_m$ inducido por las particiones de las 3-cajas que genera.

1.2. d-cópula muestral de orden m

Ahora se usarán las ideas de la sección anterior para definir la d -cópula muestral de orden m .

Definición 1.2.1 (d -Cópula muestral de orden m) Sea $m \geq 2$ y sea $U_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ una muestra independiente de tamaño n , $n \geq m$, de una d -cópula C , tal que $\underline{\mathbf{x}}_k = \langle x_{k,1} \cdots x_{k,d} \rangle \in [0, 1]^d$ para toda $k \in \{1, \dots, n\}$.

Se define, para toda $\mathbf{i} = \langle i_1, \dots, i_d \rangle \in \mathcal{I}_m$, usando la ecuación (7)

$$s_{i_1, \dots, i_d}^n = \frac{|R_{\mathbf{i}} \cap U_n|}{n}, \quad (11)$$

donde $|\cdot|$ denota la cardinalidad del conjunto.

Se define

$$S_m^n = (s_{i_1, i_2, \dots, i_d}^n)_{i_1, i_2, \dots, i_d=1}^m. \quad (12)$$

Entonces S_m^n es una matriz cuadrada de dimensión d tal que

$$\sum_{i_1, i_2, \dots, i_d=1}^m (s_{i_1, i_2, \dots, i_d}^n) = 1. \quad (13)$$

Ahora se define

$$S^+ = \{S_m^n | S_m^n \text{ es una matriz de transformación generalizada}\}. \quad (14)$$

Bajo el supuesto que $S_m^n \in S^+$ se define para todo $j \in \{1, \dots, d\}$, las particiones de $[0, 1]$, $\pi_j^n := \{p_{j,0}^n, \dots, p_{j,m}^n\}$ dadas en la Ecuación (10) y se define la d -**cópula muestral de orden m** , denotada por C_m^n por

$$C_m^n(u_1, \dots, u_d) = \begin{cases} S_m^n(\Pi)(u_1, \dots, u_d) & \text{si } S_m^n \in S^+ \\ \Pi^d(u_1, \dots, u_d) & \text{si } S_m^n \notin S^+ \end{cases}, \quad (15)$$

para todo $\langle u_1, \dots, u_d \rangle \in [0, 1]^d$.

Ejemplo 1.2.2 Se genera una muestra de tamaño cuatro de la cópula Π^2 , $m = 2$, $\langle 0.13587, 0.78362 \rangle$, $\langle 0.29310, 0.21312 \rangle$, $\langle 0.66104, 0.73981 \rangle$ y $\langle 0.88332, 0.43167 \rangle$, entonces $s_{1,1}^4 = s_{2,1}^4 = s_{1,2}^4 = s_{2,2}^4 = \frac{1}{4}$. Así, claramente S_2^4 es una matriz de transformación que genera una partición uniforme dada en la ecuación (7) y C_2^4 es simplemente Π^2 , es decir, se recupera la 2-cópula original.

A continuación se analizan las propiedades de S_m^n .

Proposición 1.2.3 Sea $m \geq 2$, sea C una d -Cópula y sea $U_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ una muestra independiente de tamaño $n \geq m$ de C , se define $q_{\mathbf{i}} = V_C(\mathbf{R}_{\mathbf{i}})$ para toda $\mathbf{i} \in \mathcal{I}_m$. Entonces la matriz cuadrada de dimensión d , $n \cdot S_m^n$, tiene asociada una distribución multinomial con parámetros n y $\{q_{\mathbf{i}}\}_{\mathbf{i} \in \mathcal{I}_m^C}$, donde $\mathcal{I}_m^C = \{\mathbf{i} \in \mathcal{I} | q_{\mathbf{i}} > 0\}$. Además, se tiene que $0 \leq q_{\mathbf{i}} \leq \frac{1}{m}$ y $\sum_{\mathbf{i} \in \mathcal{I}_m, i_j=k} q_{\mathbf{i}} = \frac{1}{m}$ para toda $j \in \{1, \dots, d\}$ y para toda $k \in \{1, \dots, m\}$.

Corolario 1.2.4 Si $C = \Pi^d$ en la proposición anterior, entonces

$$P\left(S_m^n = (s_{i_1, \dots, i_d}^n)_{i_1, \dots, i_d=1}^m\right) = \left(\frac{n!}{\prod_{\mathbf{i} \in \mathcal{I}_m} (n \cdot s_{\mathbf{i}}^n)!}\right) \left(\frac{1}{m^d}\right)^n \quad (16)$$

y si $C = M^d$, donde $M^d(u_1, \dots, u_d) = \min\{u_1, \dots, u_d\}$, entonces

$$P\left(S_m^n = (s_{i_1, \dots, i_d}^n)_{i_1, \dots, i_d=1}^m\right) = \left(\frac{n!}{\prod_{\mathbf{i} \in \mathcal{I}_m^{M^d}} (n \cdot s_{\mathbf{i}}^n)!}\right) \left(\frac{1}{m}\right)^n. \quad (17)$$

Lema 1.2.5 Sea $m \geq 2$ y sea $C = \Pi^d$, la d -cópula producto, y si se supone que el tamaño de la muestra U_n satisface que $n = m$. Entonces

$$P(s_m^n \in \mathcal{S}^+) = \frac{(m!)^d}{(m^d)^m}. \quad (18)$$

Teorema 1.2.6 Sea $m \geq 2$, $n \geq m$ y sea U_n una muestra independiente de tamaño n de una d -cópula C para alguna $d \geq 2$ fija. Se define C_m^n como en la ecuación (15). Sea S_m^n una matriz cuadrada de dimensión d inducida por la muestra U_n . Entonces para toda $\mathbf{i} = \langle i_1, \dots, i_d \rangle \in \mathcal{I}_m$ con m fija,

$$\lim_{n \rightarrow \infty} s_{i_1, \dots, i_d}^n = V_C(R_{\mathbf{i}}), \text{ casi seguramente.} \quad (19)$$

Los elementos en la partición $\{p_{j,0}^n, p_{j,1}^n, \dots, p_{j,m}^n\}$ dada en la ecuación (10), satisfacen que para toda $j \in \{1, \dots, d\}$ y para toda $k \in \{1, \dots, m\}$,

$$\lim_{n \rightarrow \infty} p_{j,k}^n = \frac{k}{m}, \text{ casi seguramente.} \quad (20)$$

Por tanto, si se define el mallado $K_m = \{0, \frac{1}{m}, \dots, \frac{m-1}{m}, 1\}^d$, la d -cópula muestral C_m^n es tal que

$$\lim_{n \rightarrow \infty} C_m^n(u_1, \dots, u_d) = C(u_1, \dots, u_d) \text{ para todo } \langle u_1, \dots, u_d \rangle \in K_m \text{ c.s.} \quad (21)$$

Finalmente, si $m \rightarrow \infty$ con valores de $m \approx n^{\frac{1}{2d}}$ entonces

$$C_m^n \text{ converge uniformemente y casi seguramente a } C. \quad (22)$$

Se observa que de la ecuación (21), si $C_m = \lim_{n \rightarrow \infty} C_m^n$, entonces C_m coincide con μ_1 en la ecuación (3), para $k = m$ y $\mathbf{P} = (p_{i_1, \dots, i_d})_{i_1, \dots, i_d=1}^m$, donde $p_{i_1, \dots, i_d} = m \cdot V_C(R_{\langle i_1, \dots, i_d \rangle})$ para todo $\langle i_1, \dots, i_d \rangle \in \mathcal{I}_m$.

En la definición de C_m^n es muy importante verificar que la matriz cuadrada de dimensión d , S_m^n pertenezca a S^+ .

Otra forma de obtener una cota superior para $P(S_m^n \in S^+)$ es usar la aproximación de Polya. Se consideran k cajas y $n \geq k$ bolas, por cada bola se selecciona una caja de manera uniforme y colocamos esa bola en esa caja, se repite ese procedimiento de forma independiente hasta que las n bolas se terminen. Se quiere encontrar la probabilidad de que no se tengan cajas vacías, este evento se denotará como E_k^n . Esta probabilidad es conocida como el problema de ocupación de Maxwell-Boltzmann cuya fórmula está dada por

$$P(E_k^n) = \sum_{j=0}^k (-1)^j \binom{k}{j} \left(1 - \frac{j}{k}\right)^n.$$

Observe que si se tiene una muestra de tamaño n de la cópula producto Π y se toma $2 \leq m \leq n$, entonces las m^d cajas usadas en la construcción de la cópula empírica de tamaño m tienen la misma probabilidad $\frac{1}{m^d}$. Si se asume que $n \geq m^d$ y $k = m^d$, en el problema de ocupación anterior es claro que la matriz $S_m^n \in S^+$ si cada caja tiene cuando menos una bola. Entonces $P(E_k^n) \leq P(S_m^n \in S^+)$. Así que si se encuentra un valor n , dependiendo de k , tal que $P(E_k^n) \approx 1$, entonces S_m^n es una matriz de transformación generalizada con una probabilidad bastante alta.

Cópula Muestral de orden m para una función de distribución continua

Sea $m, d \geq 2$ enteros fijos y sea H una función de distribución continua en \mathbb{R}^d . Sea $V_n = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ una muestra aleatoria de tamaño $n \geq m$ proveniente de H . Sea $U_n = \{\underline{\mathbf{x}}_1, \dots, \underline{\mathbf{x}}_n\}$ la muestra modificada usual. Esto es, si $j \in \{1, 2, \dots, n\}$ y $\underline{\mathbf{z}}_j = \langle z_{j,1}, \dots, z_{j,d} \rangle \in \mathbb{R}^d$, se define para $k \in \{1, 2, \dots, d\}$

$$R_{j,k} = \sum_{l=1}^n 1_{\{z_{l,k} \leq z_{j,k}\}},$$

donde $R_{j,k}$ es el rango de la observación $z_{j,k}$ para l variando desde 1 hasta n . Ahora definimos para cada $j \in \{1, 2, \dots, n\}$, $\underline{\mathbf{x}}_j = \langle x_{j,1}, \dots, x_{j,d} \rangle$, donde

$$x_{j,k} = \frac{R_{j,k}}{n}.$$

Por lo tanto $U_n = \{\underline{\mathbf{x}}_1, \dots, \underline{\mathbf{x}}_n\} \subset \left\{\frac{1}{n}, \dots, \frac{n-1}{n}, 1\right\}^d \subset [0, 1]^d$. Como se está asumiendo que H es continua, los rangos en la definición de $\underline{\mathbf{x}}_j$ son todos distintos para toda $j \in \{1, 2, \dots, n\}$, casi seguramente.

Recordemos que la **d-cópula empírica** para la muestra modificada U_n está definida para toda $\langle u_1, \dots, u_d \rangle \in [0, 1]^d$ por

$$C^m(u_1, \dots, u_d) = \frac{1}{n} \sum_{j=1}^n 1_{\{x_{j,1} \leq u_1, \dots, x_{j,d} \leq u_d\}}.$$

Se observa que C^m , la cópula empírica, no es una d -cópula. De hecho solo podemos decir que C^m es una d -subcópula con dominio $\left\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\right\}^d$.

En este caso la estructura de la muestra modificada U_n facilita enormemente la estructura de la d -cópula muestral de orden m , como se puede observar a continuación.

Teorema 1.2.7 Sea U_n una muestra aleatoria modificada obtenida de una muestra aleatoria original V_n de una función de distribución continua, H en \mathbb{R}^d . Sea $s_{i_1, \dots, i_d}^n, S_m^n$ y C_m^n como en las ecuaciones anteriores. Entonces, para toda $2 \leq m \leq n$,

$$S_m^n \in S^+.$$

Es decir, S_m^n es siempre una matriz de transformación generalizada. Además, si $2 \leq m \leq n$ y definimos para toda $j \in \{1, 2, \dots, d\}$ las particiones $\pi_j^n = \{0 = p_{j,0}, p_{j,1}, \dots, p_{j,m-1}, p_{j,m} = 1\}$, entonces

$$p_{j,k} = \frac{\lfloor \frac{k \cdot n}{m} \rfloor}{n} \text{ para todo } j \in \{1, 2, \dots, d\} \text{ y para todo } k \in \{1, 2, \dots, m\}.$$

En particular, cuando $n = m$, $\pi_j^n = \{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\}$ para $j \in \{1, 2, \dots, d\}$. De hecho, cuando $n = m \geq 2$ la cópula muestral C_n^n es

$$C_n^n(u_1, \dots, u_d) = C^n(u_1, \dots, u_d) \forall \langle u_1, \dots, u_d \rangle \in \left\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\right\}^d.$$

Es decir, se recupera la cópula empírica definida en el mallado $\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\}^d$. Cabe señalar que en este último teorema, si n es un múltiplo de m , entonces $p_{j,k} = \frac{k}{m}$ para todo $j \in \{1, 2, \dots, d\}$ y para toda $k \in \{1, 2, \dots, m\}$, es decir, recuperamos la partición original de $[0, 1]^d$. En el caso en que n no es un múltiplo de m la partición dada sigue siendo una buena aproximación de la partición original dada por $\{R_{\mathbf{i}}\}_{\mathbf{i} \in \mathcal{I}_m}$ en la ecuación (7).

Observación 1.2.8 Como ya se mencionó, la cópula empírica C^n es una d -subcópula, entonces si se usa la extensión multivariada del lema 2.3.5 (ver Nelsen [10]), usada en la prueba del Teorema de Sklar, se puede extender la cópula empírica C^n a una d -Cópula C_*^n usando interpolación multilinear estándar. En este caso $C_*^n = C_n^n$, la d -cópula muestral de orden n .

A partir de la Observación 1.2.8 se tiene que la densidad de la 2-Cópula $C_*^4 = C_4^4$ del Ejemplo 1.2.2 está dada por

$$C_*^4 = \begin{cases} 4 & \text{si } \langle u, v \rangle \in R_{1,4} \cup R_{2,1} \cup R_{3,3} \cup R_{4,2} \\ 0 & \text{En otro caso} \end{cases}.$$

Entonces el supremo de las distancias entre C_*^4 y Π^2 , la real, es $\frac{3}{16}$. Si se emplea cualquier otra distancia se tiene que la diferencia entre la cópula muestral de orden 2 y la cópula real será de cero.

Observación 1.2.9 Cuando se toma una muestra de una d -Cópula C y S_m^n no es una matriz de transformación generalizada para el valor de m requerido en

un proceso estadístico, se recomienda obtener la muestra modificada y, usando el Teorema 1.2.7, la matriz S_m^n modificada siempre será una matriz de transformación generalizada, incluso en el caso $n = m$.

Ejemplo 1.2.10 (Prueba de independencia) Se tienen observaciones de una normal multivariada de dimensión cuatro con media $\mu = 0$ y varianzas $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$ y matriz de covarianza

$$\begin{pmatrix} 1 & \rho_{1,2} & \rho_{1,3} & \rho_{1,4} \\ \rho_{1,2} & 1 & \rho_{2,3} & \rho_{2,4} \\ \rho_{1,3} & \rho_{2,3} & 1 & \rho_{3,4} \\ \rho_{1,4} & \rho_{2,4} & \rho_{3,4} & 1 \end{pmatrix}.$$

Para realizar la siguiente prueba de hipótesis (prueba de independencia)

$$H_0 : \rho_{i,j} = 0 \text{ Para todo } i, j \text{ vs } H_1 : \text{ Existe } i, j \text{ tal que } \rho_{i,j} \neq 0$$

Se deben obtener las muestras modificadas de la normal tetravariada, si se toma $m = 2$ se deben obtener las 4-cómulas muestrales de orden 2. Se observa que H_0 se cumple si y solo si las coordenadas son independientes. Así que la cópula asociada a las observaciones es la cópula producto Π^4 , en este caso podemos probar, alternativamente, lo siguiente

$$H_0 : V(R_{\mathbf{i}}) = \frac{1}{24} \text{ Para todo } \mathbf{i} \in \mathcal{I}_2 \text{ VS } H_1 : \text{ Existe } \mathbf{i} \in \mathcal{I}_2 \text{ tal que } V(R_{\mathbf{i}}) \neq \frac{1}{24}$$

Para encontrar la potencia de esta prueba se toman $\rho_{1,2} \neq 0$ y $\rho_{i,j} = 0$ para el resto. Se utiliza la χ^2 de Pearson para probar esta hipótesis.

La cópula empírica tiene grandes restricciones en términos de evaluaciones computacionales, pues los arreglos necesarios crecen muy rápido y superan la capacidad de cómputo, la cópula muestral, en cambio, solo requiere arreglos de orden m^d que resulta mucho más manejable para valores pequeños de m .

Aplicaciones de la Copula muestral

Una de las aplicaciones de la cópula muestral es la estimación de parámetros cuando se tienen muestras de una familia paramétrica $\{C_\theta | \theta \in \Theta\}$ con $\Theta \in \mathbb{R}$.

Para algunas familias paramétricas de dimensión d es posible hacer una buena estimación del parámetro con el uso de la d -cópula muestral de orden m , incluso cuando $m = 2$.

Por ejemplo, en el caso de algunas cópulas Arquimedianas multivariadas, se observa que $V_{C_\theta}([0, \frac{1}{2}]^d)$ es una función continua estrictamente monótona f del

parámetro $\theta \in \Theta \subset \mathbb{R}$ para cualquier $d \geq 2$. Este es el caso para la familias Clayton con $\theta \in (0, \infty)$, Frank, Ali-Mikhali-Haq, Gumbler-Hougard, etcétera.

Entonces, la estimar el volumen $V_{C_\theta}([0, \frac{1}{2}]^d)$ usando $s_{1,1,\dots,1}^n$ en la matriz de transformación generalizada s_m^n como la definimos en la ecuación (12) se puede encontrar un valor único para $\hat{\theta}$ tal que $f(\hat{\theta}) = s_{1,1,\dots,1}^n$. En general se debe encontrar f^{-1} para poder dar $\hat{\theta}$ pero f^{-1} no siempre tiene una expresión analítica pero puede ser aproximanda mediante procesos numéricos.

Cuando $m = 2$ se trata de estimar el valor de $\theta \in \Theta$ tal que $f(\theta) = V_{C_\theta}(R_{i_0}) = p_{i_0}$ tal que $p_{i_0} = C_\theta(\frac{1}{2}, \dots, \frac{1}{2})$, donde $i_0 = \langle 1, 1, \dots, 1 \rangle$, con base en una muestra de C_θ de tamaño n . Se sabe, por las propiedades básicas de las distribución multinomial, que el número de las observaciones que caen en la d - caja R_{i_0} , digamos X_{i_0} se distribuye como una binomial con parámetros n, p_{i_0} . Por lo tanto $s_{1,1,\dots,1}^n = \frac{X_{i_0}}{n}$ se distribuye como una binomial rescalada con valores en $\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\}$ y para n lo suficientemente grande $s_{1,1,\dots,1}^n$ es un buen estimador de $f(\theta)$, por lo tanto $\hat{\theta} = f^{-1}(s_{1,1,\dots,1}^n)$ es un buen estimador de θ . El proceso a seguir es el siguiente:

1. Encontrar la imagen directa $f[\theta] = \{f(\theta) = C_\theta(\frac{1}{2}, \dots, \frac{1}{2}) | \theta \in \Theta\} \subset I$ para la familia $\{C_\theta | \theta \in \Theta\}$.
2. Dada una muestra U_n encontrar el valor de $s_{1,1,\dots,1}^n$ en la construcción de la d -cópula muestral de orden $m = 2$. Si $s_{1,1,\dots,1}^n \in \text{int}f[\Theta]$ proceder con el siguiente paso.
3. Si f^{-1} tiene una expresión analítica, definir $\hat{\theta} = f^{-1}(s_{1,1,\dots,1}^n)$ y terminamos. En otro caso, si Θ está acotado, dé un mallado adecuado de Θ para aproximar $f[\Theta]$.

Otra aplicación es una prueba de bondad de ajuste. Si se toma una muestra de una d -cópula de tamaño n y un entero fijo $2 \leq m \leq n$. Sea $R_{\underline{i}}$ para $\underline{i} \in \mathcal{I}_m$ una partición de I^d en la construcción de la muestra de la d -cópula y supongamos que $S_m^n = (s_{i_1, \dots, i_d}^n)_{i_1, \dots, i_d=1}^m$ es una matriz de transformación generalizada, se sabe que S_m^n tiene una distribución multinomial con parámetros n y $\{q_{\underline{i}}\}_{\underline{i} \in \mathcal{I}_m^C}$, donde $\mathcal{I}_m^C = \{\underline{i} \in \mathcal{I}_m | q_{\underline{i}} > 0\}$. Por lo tanto se quiere probar la siguiente hipótesis

$$H_0 : n \cdot S_m^n \rightsquigarrow \text{Mult} \left(n, \{q_{\underline{i}}\}_{\underline{i} \in \mathcal{I}_m^C} \right).$$

Para ello se usa una X^2 de Pearson que asintóticamente se distribuye como una χ -cuadrada con $k - 1$ grados de libertad, donde k denota la cardinalidad de \mathcal{I}_m^C .

Observe que la hipótesis nula propuesta no caracteriza a una única cópula, solo nos da el volumen de las d -cajas necesarias para la construcción de la d -cópula muestral de orden m , sin embargo cuando m es lo suficientemente grande la hipótesis nula se aproxima bastante a la cópula subyacente C .

Otra aplicación importante es una propuesta para probar dos sencillas hipótesis:

$$H_0 : \underline{X} \rightsquigarrow C_0 \text{ vs } H_1 : \underline{X} \rightsquigarrow C_1,$$

donde C_0 y C_1 son dos d -cópulas totalmente determinadas que son diferentes. Sea $m \geq 2$ fijo y denotemos por $V_0 1_0(R_i)$ y $V_0 1_1(R_i)$ el volumen de la partición uniforme dada en la ecuación (7) de $[0, 1]^d$ bajo H_0 y H_1 , respectivamente. Como se trata de dos cópulas distintas, se tiene que $V_0 1_0(R_{i_1}) \neq V_0 1_1(R_{i_1})$ y $V_0 1_0(R_{i_2}) \neq V_0 1_1(R_{i_2})$. Sea $U_n = \{x_1, \dots, x_n\}$ una muestra aleatoria de la d -cópula C_0 y se obtiene la matriz S_m^n , si es una matriz de transformación generalizada, usando la proposición 1.2.3 y el teorema de Neyman-Pearson, se puede encontrar $\frac{L(\theta_0; n, S_m^n)}{L(\theta_1; n, S_m^n)}$, donde $\theta_0 = \{Vol_0(R_i)\}_{i \in \mathcal{I}_m}$ y $\theta_1 = \{Vol_1(R_i)\}_{i \in \mathcal{I}_m}$, el cual está dado por

$$T_{m,n}^d = \prod_{i \in \mathcal{I}_m} \left(\frac{Vol_0(R_i)}{Vol_1(R_i)} \right)^{n \cdot s_i^n}.$$

Entonces la región crítica ideal para probar H_0 vs H_1 está dada por

$$D = \{U_n | U_n \text{ es una muestra de tamaño } n \text{ de } C_0 \text{ tal que } T_{m,n}^d \leq K_\alpha\}.$$

Donde $P(T_{m,n}^d \leq K_\alpha | H_0 \text{ es cierta}) = \alpha$ y α es la probabilidad del error Tipo 1.

Como la distribución de $T_{m,n}^d$ no es conocida, se puede estimar K_α simulando muestras de C_0 un número de veces L que sea grande, se estiman los cuantiles requeridos de la distribución de $T_{m,n}^d$, se recomienda usar $L \geq 50000$ para estimar los valores de K_α .

2. Estimación de la correlación de la distribución normal

2.1. Distribución normal

Distribución Normal Univariada

La densidad de una variable aleatoria que sigue una distribución normal con media μ y desviación estándar σ está dada por

$$N(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (23)$$

La función de distribución de una normal $N(x, 0, 1)$ no tiene forma analítica cerrada, usualmente se denota por Φ

$$\Phi(t) = \int_{-\infty}^t N(x, 0, 1) dx. \quad (24)$$

En el caso general, si se hace el cambio de variable $x' = \frac{x-\mu}{\sigma}$ en la integral $F(t) = \int_{-\infty}^t N(x, \mu, \sigma) dx$, la función de distribución de $N(x, \mu, \sigma)$ se puede ver como

$$F(x) = \Phi((x - \mu) / \sigma).$$

Notemos que $F(x)$ es invertible, con inversa igual a

$$F^{-1}(y) = \sigma\Phi^{-1}(y) + \mu.$$

Definición 2.1.1 Se define la distancia Mahalanobis de un punto x a la Gaussiana que compone la densidad de la Normal como

$$r = \frac{|x - \mu|}{\sigma}. \quad (25)$$

Esta distancia indica cuantas desviaciones estándar hay entre el punto x y la Gaussiana. De modo que si $\sigma = 4$ y $\mu = 2$ entonces para $x = 6$ la distancia Mahalanobis resulta ser $r = 1$ pues está a una desviación estándar de la media, análogamente, para $x = 10$, $r = 2$.

Dado $R > 0$, se calcula la probabilidad de que un punto caiga a una distancia Mahalanobis $r \leq R$ de la Gaussiana. Esta probabilidad es igual a

$$\int_{\mu-\sigma R}^{\mu+\sigma R} N(x, \mu, \sigma) dx,$$

ya que la Gaussiana es simétrica alrededor de su media, la probabilidad de este evento es

$$\begin{aligned}
\int_{\mu-\sigma R}^{\mu+\sigma R} N(x, \mu, \sigma) dx &= 2 \int_{\mu}^{\mu+\sigma R} N(x, \mu, \sigma) dx \\
&= 2[F(\mu + \sigma R) - F(\mu)] \\
&= 2F(\mu + \sigma R) - 1.
\end{aligned}$$

De manera inversa, si la probabilidad de un punto de caer a cierta distancia Mahalanobis R de la Gaussiana es conocida, es posible calcular R de forma explícita

$$\begin{aligned}
p &= 2F(\mu + \sigma R) - 1 \\
\Rightarrow F(\mu + \sigma R) &= \frac{p+1}{2} \\
\Rightarrow R &= \left[F^{-1}\left(\frac{p+1}{2}\right) - \mu \right] / \sigma \\
\Rightarrow R &= \Phi^{-1}\left(\frac{p+1}{2}\right).
\end{aligned}$$

Distribución Normal Bivariada

Sea $z = (x, y)$, la densidad de una normal bivariada con media $\mu = (\mu_1, \mu_2)$ y matriz de varianzas, covarianza Σ es

$$p(x, y) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(z - \mu)\Sigma^{-1}(z - \mu)^t\right], \quad (26)$$

donde Σ es una matriz simétrica y definida positiva

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

con σ_1 y σ_2 distintos de 0 y $|\rho| < 1$. Su matriz inversa Σ^{-1} es

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

La distancia Mahalanobis r de un punto $z = (x, y)$ a la Gaussiana bivariada es

$$r^2 = (z - \mu)\Sigma^{-1}(z - \mu)^t, \quad (27)$$

o de forma explícita

$$r^2 = \frac{1}{1 - \rho^2} \left[\left(\frac{x - \mu_1}{\sigma_1}\right)^2 - 2\rho \left(\frac{x - \mu_1}{\sigma_1}\right) \left(\frac{y - \mu_2}{\sigma_2}\right) + \left(\frac{y - \mu_2}{\sigma_2}\right)^2 \right].$$

Dado que $|\rho| < 1$, la región de puntos para los cuales la distancia r es constante son elipses. La parametrización para una curva de nivel con r^2 fija está dada por

$$x = r\sigma_1 \cos(\theta) + \mu_1 \quad (28)$$

$$y = r\sigma_2 \left(\rho \cos(\theta) + \sqrt{1 - \rho^2} \sin(\theta) \right) + \mu_2. \quad (29)$$

Sea M la descomposición de Cholesky de Σ , es decir, M es una matriz triangular superior tal que $M^t M = \Sigma$, entonces la parametrización anterior puede ser escrita como

$$(x, y) = r(\cos(\theta), \sin(\theta))M + (\mu_1, \mu_2),$$

donde M es

$$M = \begin{pmatrix} \sigma_1 & \rho\sigma_2 \\ 0 & \sigma_2\sqrt{1 - \rho^2} \end{pmatrix}.$$

Esta forma de parametrización de Cholesky también es válida para dimensiones mayores y se pueden obtener de forma eficiente mediante algoritmos computacionales.

Definición 2.1.2 Se dice que \mathbf{M} es una matriz definida positiva si para todo $\mathbf{x} \neq 0$ se tiene que $\mathbf{x}^t \mathbf{M} \mathbf{x} > 0$. Equivalentemente, \mathbf{M} es definida positiva si todos sus valores propios son positivos.

Definición 2.1.3 Se dice que \mathbf{M} es definida semipositiva si todos sus valores propios son mayores o iguales que cero.

Toda matriz de covarianza es definida semipositiva, es decir, que para todo vector $y \in \mathbb{R}^d$ se satisface que $y^t \Sigma y \geq 0$. Sin embargo, para poder realizar la descomposición de Cholesky, es necesario que la matriz sea definida positiva, es decir, que sea una matriz no singular, esto restringe los posibles valores que puede tomar la correlación, como se verá a continuación.

Sea $Z = (X, Y)$ un vector aleatorio en \mathbb{R}^2 que se distribuye Normal con función de densidad definida como en la ecuación (26) con $\mu = \underline{0}$ y $\sigma_1 = \sigma_2 = 1$, para el caso de dimensión dos se observa que ρ puede tomar cualquier valor en el intervalo $(-1, 1)$ pues

$$\text{Det} \left| \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right| = 1 - \rho^2 > 0 \text{ si y solo si } -1 < \rho < 1.$$

Sin embargo para el caso de tres dimensiones se tienen más restricciones y por lo tanto el espacio en donde pueden tomar valores $\rho_{1,2}$, $\rho_{1,3}$ y $\rho_{2,3}$ no es $(-1, 1)^3$,

como se verá a continuación:

$$\text{Det} \left| \begin{pmatrix} 1 & \rho_{1,2} & \rho_{1,3} \\ \rho_{1,2} & 1 & \rho_{2,3} \\ \rho_{1,3} & \rho_{2,3} & 1 \end{pmatrix} \right| = 1 - (\rho_{1,2}^2 + \rho_{1,3}^2 + \rho_{2,3}^2) + 2\rho_{1,2}\rho_{1,3}\rho_{2,3},$$

para dar una descripción de la región en donde $\rho_{1,2}, \rho_{1,3}$ y $\rho_{2,3}$ pueden tomar valores, a fin de que $|\Sigma| > 0$, se renombra a las variables y se analiza la superficie que describen.

$$1 - (x'^2 + y'^2 + z'^2) + 2x'y'z' = 0,$$

donde $\rho_{1,2} = x', \rho_{1,3} = y'$ y $\rho_{2,3} = z'$.

Sea $z' = a$ una constante tal que $-1 < a < 1$.

Entonces se tiene la ecuación $x'^2 + y'^2 - 2ax'y' + (a^2 - 1) = 0$ y mediante la transformación

$$\begin{aligned} x' &= x \cos(\theta) - y \sin(\theta) \\ y' &= x \sin(\theta) + y \cos(\theta) \end{aligned}$$

se tiene que

$$\begin{aligned} 0 &= (x \cos(\theta) - y \sin(\theta))^2 + (x \sin(\theta) + y \cos(\theta))^2 \\ &\quad - 2a(x \cos(\theta) - y \sin(\theta))(x \sin(\theta) + y \cos(\theta)) + (a^2 - 1) \end{aligned}$$

si y solo si

$$x^2(1 - 2a \cos(\theta) \sin(\theta)) + y^2(1 + 2a \cos(\theta) \sin(\theta)) = 1 - a^2$$

y

$$2xya(\sin(\theta)^2 - \cos(\theta)^2) = 0$$

y si $a \neq 0$ entonces $\sin(\theta)^2 = \cos(\theta)^2$ es decir, $\theta = 45'$.

De manera que la ecuación bajo esta transformación sería

$$\begin{aligned} x^2(1 - a) + y^2(1 + a) &= 1 - a^2 \\ \frac{x^2}{1 + a} + \frac{y^2}{1 - a} &= 1 \end{aligned}$$

Que son elipses rotadas 45 grados. Esto indica que el espacio en donde $\rho_{1,2}, \rho_{1,3}$ y $\rho_{2,3}$ pueden tomar valores es el interior de la región descrita en la figura (3). Es importante notar que el espacio posible de las $\rho_{i,j}$'s se reduce conforme aumenta la dimensión pues se tienen más restricciones.

Sea $p(x, y)$ la densidad de la normal bivariada general. En el artículo de Michaël Bensimhou "N-Dimensional Cumulative Function, and other useful facts about

Gaussians and Normal Densities" [1] dan la función de distribución de la normal bivariada en términos de r , la distancia de Mahalanobis (ecuación 27), es decir, calculan la probabilidad de que un punto caiga dentro de la elipse dada por la parametrización de la ecuación (28), para una distancia de Mahalanobis dada $r = R$.

Para ello se requiere el cálculo del Jacobiano de la transformación

$$\begin{aligned}
 J &= \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} \\
 &= \begin{vmatrix} \sigma_1 \cos \theta & -\sigma_1 r \sin \theta \\ \sigma_2(\rho \cos(\theta) + \sqrt{1 - \rho^2}) \sin(\theta) & \sigma_2 r(-\rho \sin(\theta) + \sqrt{1 - \rho^2}) \cos(\theta) \end{vmatrix} \\
 &= \sqrt{1 - \rho^2} \sigma_1 \sigma_2 r \\
 &= \sqrt{|\Sigma|} r.
 \end{aligned}$$

Sin pérdida de generalidad supongamos que $\mu = \underline{0}$ (pues las integrales son invariantes bajo traslaciones). La integral para obtener entonces la distribución es

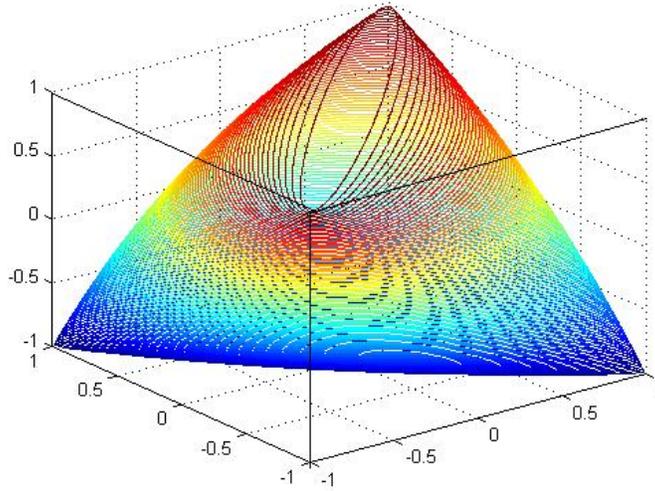


Figura 3: espacio de rhos posibles

$$\begin{aligned}
\frac{1}{2\pi\sqrt{|\Sigma|}} \int_0^{2\pi} \int_0^R \sqrt{|\Sigma|} r e^{-\frac{r^2}{2}} dr d\theta &= \int_0^R r e^{-\frac{r^2}{2}} dr \\
&= \int_0^R e^{-\frac{r^2}{2}} dr^2 / 2 \\
&= \int_0^{R^2/2} e^{-r} dr \\
&= 1 - e^{-R^2/2}.
\end{aligned}$$

Por lo tanto la función de distribución está dada por

$$F(r) = 1 - e^{-r^2/2}. \quad (30)$$

2.2. Propuesta del estimador

Proposición 2.2.1 $p = \int_{-\infty}^0 \int_{-\infty}^0 f_{XY}(x, y) dx dy$ tiene forma cerrada dada por $p = \frac{\arccos(-\rho)}{2\pi}$.

Demostración

Para la demostración se usarán los principales resultados del artículo de Michaël Bensimhoun. Se quiere calcular la integral de la densidad sobre la región $(-\infty, 0] \times (-\infty, 0]$.

Usando la parametrización dada en (28) y suponiendo, sin pérdida de generalidad, que $\mu = \underline{0}$ y $\sigma_1 = \sigma_2 = 1$ se tiene la transformación denotada por ϕ .

$$x = r \cos(\theta) \quad (31)$$

$$y = r \left(\rho \cos(\theta) + \sqrt{1 - \rho^2} \sin(\theta) \right). \quad (32)$$

Por la simetría de la distribución Normal resulta indiferente si se integra sobre $(-\infty, 0] \times (-\infty, 0]$ o sobre $B = [0, \infty) \times [0, \infty)$. Se realizará la integral sobre la región B , para ello se requiere cambiar los límites de integración y después tomar el límite cuando $R \rightarrow \infty$.

Si $y = 0$ entonces $\sqrt{1 - \rho^2} \sin(\theta) = -\rho \cos(\theta)$

lo cual implica que $(1 - \rho^2) \sin^2(\theta) = \rho^2 \cos^2(\theta)$
por lo tanto $\theta = -\arcsin(\rho)$

Análogamente, cuando $x = 0$ se tiene que $\cos(\theta) = 0$ de manera que $\theta = \pi/2$.

Entonces

$$\begin{aligned}
 \frac{1}{2\pi\sqrt{|\Sigma|}} \int_{-\arcsin(\rho)}^{\pi/2} \int_0^R \sqrt{|\Sigma|} r e^{-\frac{r^2}{2}} dr d\theta &= \frac{\pi/2 + \arcsin(\rho)}{2\pi} \int_0^R r e^{-\frac{r^2}{2}} dr \\
 &= \left(\frac{1}{4} + \frac{\arcsin(\rho)}{2\pi} \right) \int_0^R e^{-\frac{r^2}{2}} dr^2 / 2 \\
 &= \left(\frac{1}{4} + \frac{\arcsin(\rho)}{2\pi} \right) \int_0^{R^2/2} e^{-r} dr \\
 &= \left(\frac{1}{4} + \frac{\arcsin(\rho)}{2\pi} \right) (1 - e^{-R^2/2}).
 \end{aligned}$$

La expresión anterior es la distribución de la porción de la elipse que se encuentra en la región B , para obtener la expresión final se toma el límite cuando $R \rightarrow \infty$, entonces

$$\begin{aligned}
 p &= \int_{-\infty}^0 \int_{-\infty}^0 f_{XY}(x, y) dx dy \\
 &= \int_0^{\infty} \int_0^{\infty} f_{XY}(x, y) dx dy \\
 &= \lim_{R \rightarrow \infty} \frac{1}{2\pi\sqrt{|\Sigma|}} \int_{-\arcsin(\rho)}^{\pi/2} \int_0^R \sqrt{|\Sigma|} r e^{-\frac{r^2}{2}} dr d\theta \\
 &= \lim_{R \rightarrow \infty} \left(\frac{1}{4} + \frac{\arcsin(\rho)}{2\pi} \right) (1 - e^{-R^2/2}) \\
 &= \frac{1}{4} + \frac{\arcsin(\rho)}{2\pi}.
 \end{aligned}$$

Utilizando las identidades siguientes

$$\begin{aligned}
 \arccos(x) &= \frac{\pi}{2} - \arcsin(x) \\
 \arccos(\rho) + \arccos(-\rho) &= \pi
 \end{aligned}$$

podemos concluir que

$$\begin{aligned}
 p &= \int_{-\infty}^0 \int_{-\infty}^0 f_{XY}(x, y) dx dy \\
 &= \frac{1}{4} + \frac{\arcsin(\rho)}{2\pi} \\
 &= \frac{1}{4} + \frac{\pi/2 - \arccos(\rho)}{2\pi}
 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{4} + \frac{\pi/2 - (\pi - \arccos(-\rho))}{2\pi} \\
&= \frac{1}{4} + \frac{-\pi/2 + \arccos(-\rho)}{2\pi} \\
&= \frac{\arccos(-\rho)}{2\pi}.
\end{aligned}$$

■

Tenemos una muestra aleatoria de tamaño n , $\{x_1, x_2, \dots, x_n\}$ que proviene de una Normal bivariada con media $\mu = \mathbf{0}$, $\sigma_1 = \sigma_2 = 1$ y correlación conocida ρ . Con base en el resultado de la Proposición 2.2.1 se propone un estimador para ρ dado por

$$\hat{\rho} = -\cos(2\pi Y), \quad (33)$$

donde $Y = \frac{X}{n}$ y X sigue una distribución binomial con parámetros (m, p) con $m = \frac{n}{2}$ y p representa la probabilidad de caer en la primera caja $[0, \frac{1}{2}]^2$ de las observaciones transformadas a la cópula empírica.

3. Resultados

3.1. Propiedades del estimador

Usando el resultado 2.2.1 se sabe que la probabilidad de caer en la caja $[0, \frac{1}{2}]^2$ es $p = \frac{\arccos(-\rho)}{\pi}$.

Proposición 3.1.1 $\mathbb{E}(\hat{\rho}) = -\operatorname{Re}\{(pz + 1 - p)^m, z = \cos(\frac{\pi}{n}) + i \cdot \sin(\frac{\pi}{n}) \in \mathbb{C}$.

Demostración

$$\begin{aligned}
 \mathbb{E}(\hat{\rho}) &= \mathbb{E}\left(-\cos\left(2\pi \cdot \frac{X}{n}\right)\right) \\
 &= \sum_{j=0}^m -\cos\left(2\pi \frac{j}{n}\right) \binom{m}{j} p^j (1-p)^{m-j} \\
 &= \operatorname{Re}\left(\sum_{j=0}^m -\cos\left(2\pi \frac{j}{n}\right) \binom{m}{j} p^j (1-p)^{m-j}\right. \\
 &\quad \left.+ i \cdot \sum_{j=0}^m -\sin\left(2\pi \frac{j}{n}\right) \binom{m}{j} p^j (1-p)^{m-j}\right) \\
 &= \operatorname{Re}\left(-\sum_{j=0}^m z^j \binom{m}{j} p^j p^{m-j}\right) \\
 &= \operatorname{Re}\left(-\sum_{j=0}^m \binom{m}{j} (zp)^j (1-p)^{m-j}\right) \\
 &= -\operatorname{Re}\{(pz + 1 - p)^m\},
 \end{aligned}$$

donde $z = \cos(\pi \cdot \frac{1}{n}) + i \cdot \sin(\pi \cdot \frac{1}{n}) \in \mathbb{C}$ y por las propiedades de los números complejos se tiene que $z^j = \cos(\pi \cdot \frac{j}{n}) + i \cdot \sin(\pi \cdot \frac{j}{n})$ tal que $z^j \in \mathbb{C}$.

■

De aquí se sigue que la esperanza del estimador propuesto tiene forma cerrada y que es la parte real del número complejo $(pz + 1 - p)^m$.

Proposición 3.1.2 $(pz + 1 - p) = w$ con $w = r(\cos(\theta) + i \cdot \sin(\theta)) \in \mathbb{C}$ tal que

$$\begin{aligned}
 r &= \sqrt{2p(p-1)\left(1 - \cos\left(\frac{\pi}{n}\right)\right) + 1} \\
 \theta &= \arctan\left(\frac{2p \cdot \tan\left(\frac{\pi}{2n}\right)}{\sec\left(\frac{\pi}{2n}\right)^2 - 2p \cdot \tan\left(\frac{\pi}{2n}\right)^2}\right)
 \end{aligned}$$

Demostración Primero se obtiene la expresión del módulo r y el argumento θ .

$$\begin{aligned} pz + 1 - p &= p \left(\cos \left(\frac{\pi}{n} \right) + i \cdot \sin \left(\frac{\pi}{n} \right) \right) + 1 - p \\ &= p \left(\cos \left(\frac{\pi}{n} \right) - 1 \right) + 1 + ip \cdot \sin \left(\frac{\pi}{n} \right). \end{aligned}$$

De modo que

$$\begin{aligned} r &= \sqrt{\left(p \left(\cos \left(\frac{\pi}{n} \right) - 1 \right) + 1 \right)^2 + \left(p \cdot \sin \left(\frac{\pi}{n} \right) \right)^2} \\ &= \sqrt{2p^2 - 2p^2 \cos \left(\frac{\pi}{n} \right) + 2p \cos \left(\frac{\pi}{n} \right) + 1 - 2p} \\ &= \sqrt{2p \cos \left(\frac{\pi}{n} \right) [1 - p] - 2p [1 - p] + 1} \\ &= \sqrt{[1 - p] \left(2p \cos \left(\frac{\pi}{n} \right) - 2p \right) + 1} \\ &= \sqrt{2p(p - 1) \left(1 - \cos \left(\frac{\pi}{n} \right) \right) + 1}. \end{aligned}$$

Ahora, para el argumento θ se tiene que

$$\begin{aligned} \theta &= \arctan \left(\frac{p \cdot \sin \left(\frac{\pi}{n} \right)}{p \cdot \cos \left(\frac{\pi}{n} \right) - p + 1} \right) \\ &= \arctan \left(\frac{2p \sin \left(\frac{\pi}{2n} \right) \cos \left(\frac{\pi}{2n} \right)}{-2p \left(\sin \left(\frac{\pi}{2n} \right) \right)^2 + 1} \right) \\ &= \arctan \left(\frac{2p \cdot \cos \left(\frac{\pi}{2n} \right)^2 \tan \left(\frac{\pi}{2n} \right)}{1 - 2p \cdot \tan \left(\frac{\pi}{2n} \right)^2 \cos \left(\frac{\pi}{2n} \right)^2} \right) \\ &= \arctan \left(\frac{2p \cdot \tan \left(\frac{\pi}{2n} \right)}{\frac{1}{\cos \left(\frac{\pi}{2n} \right)^2} - 2p \cdot \tan \left(\frac{\pi}{2n} \right)^2} \right) \\ &= \arctan \left(\frac{2p \cdot \tan \left(\frac{\pi}{2n} \right)}{\sec \left(\frac{\pi}{2n} \right)^2 - 2p \cdot \tan \left(\frac{\pi}{2n} \right)^2} \right). \end{aligned}$$

Entonces w está dado por

$$\begin{aligned}
w &= \left(\sqrt{2p(p-1) \left(1 - \cos\left(\frac{\pi}{n}\right)\right) + 1} \right) \cdot \\
&\quad \left(\cos \left(\arctan \left(\frac{2p \cdot \tan\left(\frac{\pi}{2n}\right)}{\sec\left(\frac{\pi}{2n}\right)^2 - 2p \cdot \tan\left(\frac{\pi}{2n}\right)^2} \right) \right) \right) \\
&\quad + i \cdot \sin \left(\arctan \left(\frac{2p \cdot \tan\left(\frac{\pi}{2n}\right)}{\sec\left(\frac{\pi}{2n}\right)^2 - 2p \cdot \tan\left(\frac{\pi}{2n}\right)^2} \right) \right).
\end{aligned}$$

■

De modo que

$$\begin{aligned}
\mathbb{E}(\hat{\rho}) &= -\mathbf{Re}\langle w^n \rangle \\
&= - \left(\sqrt{2p(p-1) \left(1 - \cos\left(\frac{\pi}{n}\right)\right) + 1} \right)^n \cdot \\
&\quad \left(\cos \left(n \cdot \arctan \left(\frac{2p \cdot \tan\left(\frac{\pi}{2n}\right)}{\sec\left(\frac{\pi}{2n}\right)^2 - 2p \cdot \tan\left(\frac{\pi}{2n}\right)^2} \right) \right) \right).
\end{aligned} \tag{34}$$

Proposición 3.1.3 El estimador $\hat{\rho}$ es insesgado para $\rho \in \{-1, 0, 1\}$.

Demostración

■ $\rho = 1$

Cuando $\rho = 1$, la probabilidad es $p = 1$ ya que las realizaciones caerían solo en la caja C_1 . Bajo estas hipótesis

$$\begin{aligned}
\mathbb{E}(\hat{\rho}) &= - \left(\sqrt{2(1-1) \left(1 - \cos\left(\frac{\pi}{n}\right)\right) + 1} \right)^n \cdot \\
&\quad \left(\cos \left(n \cdot \arctan \left(\frac{2 \tan\left(\frac{\pi}{2n}\right)}{\sec\left(\frac{\pi}{2n}\right)^2 - 2 \tan\left(\frac{\pi}{2n}\right)^2} \right) \right) \right) \\
&= -1 \cdot \left(\cos \left(n \cdot \frac{\pi}{n} \right) \right) \\
&= -(\cos(\pi)) \\
&= 1.
\end{aligned}$$

■ $\rho = 0$

En este caso es claro que $p = \frac{1}{2}$, es decir, las cajas son equiprobables, entonces

$$\begin{aligned}
\mathbb{E}(\hat{\rho}) &= - \left(\sqrt{1 - \left(-\frac{1}{2}\right) \left(1 - \cos\left(\frac{\pi}{n}\right)\right) + 1} \right)^n \cdot \\
&\quad \left(\cos \left(n \cdot \arctan \left(\frac{\tan\left(\frac{\pi}{2n}\right)}{\sec\left(\frac{\pi}{2n}\right)^2 - \tan\left(\frac{\pi}{2n}\right)^2} \right) \right) \right) \\
&= - \left(\sqrt{\frac{1 + \cos\left(\frac{\pi}{n}\right)}{2}} \right)^n \cdot \left(\cos \left(n \cdot \arctan \left(\tan\left(\frac{\pi}{2n}\right) \right) \right) \right) \\
&= - \left(\cos\left(\frac{\pi}{2n}\right) \right)^n \cdot \left(\cos\left(\frac{\pi}{2}\right) \right) \\
&= - \left(\cos\left(\frac{\pi}{2n}\right) \right)^n \cdot (0) \\
&= 0.
\end{aligned}$$

■ $\rho = -1$

Cuando $\rho = -1$, la probabilidad es $p = 0$, entonces

$$\begin{aligned}
\mathbb{E}(\hat{\rho}) &= - (\sqrt{0+1})^n \cdot (\cos(n \cdot \arctan(0))) \\
&= -1 \cdot (\cos(n \cdot 0)) \\
&= -1.
\end{aligned}$$

■

Para los casos de $\rho = \{-1, 0, 1\}$ el estimador es insesgado, sin embargo para otros valores de ρ el estimador es asintóticamente insesgado, ver Figura (4).

Proposición 3.1.4 El estimador $\hat{\rho}$ es asintóticamente insesgado, es decir que $\lim_{n \rightarrow \infty} \mathbb{E}(\hat{\rho}) = \rho$.

Demostración

$$\blacksquare \lim_{n \rightarrow \infty} \left(\sqrt{2p(p-1) \left(1 - \cos\left(\frac{\pi}{n}\right)\right) + 1} \right)^n = 1$$

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \left(\sqrt{2p(p-1) \left(1 - \cos\left(\frac{\pi}{n}\right)\right) + 1} \right)^n = 1 \\
\iff &\lim_{n \rightarrow \infty} \ln \left(\left(\sqrt{2p(p-1) \left(1 - \cos\left(\frac{\pi}{n}\right)\right) + 1} \right)^n \right) = 0.
\end{aligned}$$

Entonces

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \ln \left(\left(\sqrt{2p(p-1) \left(1 - \cos\left(\frac{\pi}{n}\right)\right)} + 1 \right)^n \right) \\
= & \lim_{n \rightarrow \infty} \frac{n}{2} \cdot \ln \left(2p(p-1) \left(1 - \cos\left(\frac{\pi}{n}\right)\right) + 1 \right) \\
= & \lim_{n \rightarrow \infty} \frac{\ln \left(2p(p-1) \left(1 - \cos\left(\frac{\pi}{n}\right)\right) + 1 \right)}{\frac{2}{n}} \\
= & \lim_{n \rightarrow \infty} \frac{\frac{-2p(p-1)\pi \sin(\pi/n)}{n^2}}{\frac{-2}{n^2}} \\
= & \lim_{n \rightarrow \infty} p(p-1)\pi \sin(\pi/n) \\
= & 0 \\
\Rightarrow & \lim_{n \rightarrow \infty} \left(\sqrt{2p(p-1) \left(1 - \cos\left(\frac{\pi}{n}\right)\right)} + 1 \right)^n = 1.
\end{aligned}$$

■ $\lim_{n \rightarrow \infty} n \cdot \arctan \left(\frac{2p \cdot \tan\left(\frac{\pi}{2n}\right)}{\sec\left(\frac{\pi}{2n}\right)^2 - 2p \cdot \tan\left(\frac{\pi}{2n}\right)^2} \right) = \arccos(-\rho)$

$$\begin{aligned}
& \lim_{n \rightarrow \infty} n \cdot \arctan \left(\frac{2p \cdot \tan\left(\frac{\pi}{2n}\right)}{\sec\left(\frac{\pi}{2n}\right)^2 - 2p \cdot \tan\left(\frac{\pi}{2n}\right)^2} \right) \\
= & \lim_{n \rightarrow \infty} \frac{\arctan \left(\frac{2p \cdot \tan\left(\frac{\pi}{2n}\right)}{\sec\left(\frac{\pi}{2n}\right)^2 - 2p \cdot \tan\left(\frac{\pi}{2n}\right)^2} \right)}{1/n} \\
= & \lim_{n \rightarrow \infty} \frac{d \left(\arctan \left(\frac{2p \cdot \tan\left(\frac{\pi}{2n}\right)}{\sec\left(\frac{\pi}{2n}\right)^2 - 2p \cdot \tan\left(\frac{\pi}{2n}\right)^2} \right) \right)}{\frac{dn}{-1/n^2}} \\
= & \lim_{n \rightarrow \infty} \frac{\frac{p \cdot \pi \left(2p \sin\left(\frac{\pi}{2n}\right)^2 - 2 \sin\left(\frac{\pi}{2n}\right)^2 + 1 \right)}{n^2 \left(4p^2 \sin\left(\frac{\pi}{2n}\right)^2 - 4p \sin\left(\frac{\pi}{2n}\right)^2 + 1 \right)}}{\frac{-1}{n^2}} \\
= & \lim_{n \rightarrow \infty} \frac{p \cdot \pi \left(2p \sin\left(\frac{\pi}{2n}\right)^2 - 2 \sin\left(\frac{\pi}{2n}\right)^2 + 1 \right)}{\left(4p^2 \sin\left(\frac{\pi}{2n}\right)^2 - 4p \sin\left(\frac{\pi}{2n}\right)^2 + 1 \right)} \\
= & p \cdot \pi \cdot \lim_{n \rightarrow \infty} \frac{\left(2p \sin\left(\frac{\pi}{2n}\right)^2 - 2 \sin\left(\frac{\pi}{2n}\right)^2 + 1 \right)}{\left(4p^2 \sin\left(\frac{\pi}{2n}\right)^2 - 4p \sin\left(\frac{\pi}{2n}\right)^2 + 1 \right)} \\
= & p \cdot \pi \cdot \frac{2p \cdot 0 - 2 \cdot 0 + 1}{4p^2 \cdot 0 - 4p \cdot 0 + 1} \\
= & p \cdot \pi \\
= & \arccos(-\rho).
\end{aligned}$$

por lo tanto

$$\lim_{n \rightarrow \infty} (w)^n = -e^{-i \cdot \arccos(\rho)}, \quad (35)$$

que es la notación de Euler para los complejos, finalmente

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}(\hat{\rho}) &= -\lim_{n \rightarrow \infty} \mathbf{Re} \langle w^n \rangle \\ &= -\mathbf{Re} \langle -e^{-i \cdot \arccos(\rho)} \rangle \\ &= \mathbf{Re} \langle \cos(-\arccos(\rho)) + i \sin(-\arccos(\rho)) \rangle \\ &= \cos(-\arccos(\rho)) \\ &= \rho. \end{aligned}$$

■

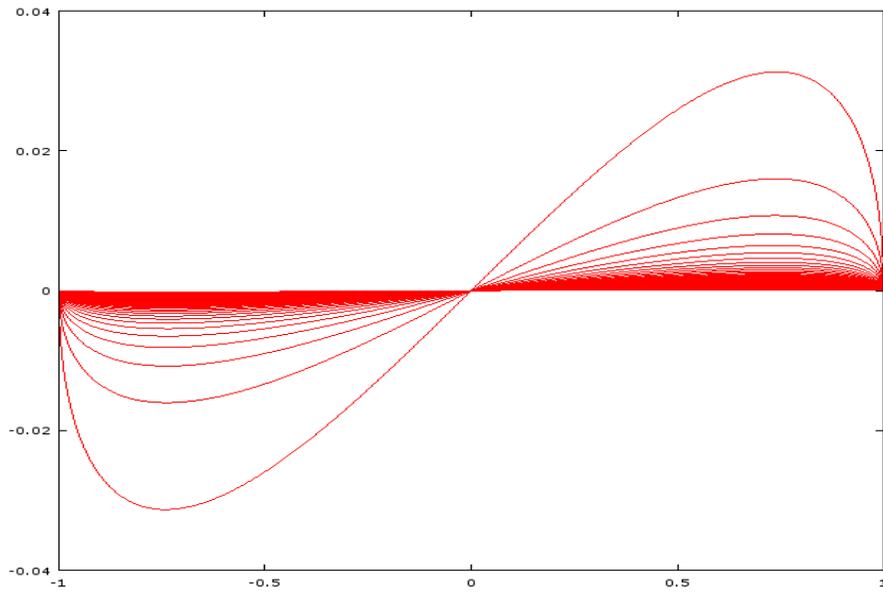


Figura 4: Asintóticamente insesgado

3.2. Robustez

Se dice que las medidas que caracterizan a una distribución, por ejemplo la media o la varianza muestral, son **robustas** si cambios ligeros en la distribución tienen cambios relativamente ligeros en su valor. En este sentido la media muestral y la varianza muestral no son estimadores robustos.

Las observaciones atípicas (*outliers*) han resultado ser un problema práctico serio pues estas observaciones pueden aumentar el error estándar de la media muestral. Los métodos modernos de robustez proporcionan un camino efectivo para lidiar con este problema.

Sea Δ_x una distribución tal que el valor x ocurre con probabilidad 1, es decir, una función de distribución cuya masa se concentra en el punto x . Ahora consideremos una mezcla de dos distribuciones donde una observación es aleatoriamente seleccionada de la distribución F con probabilidad $1 - \epsilon$ y de la distribución Δ_x con probabilidad ϵ . La distribución resultante es

$$F_{x,\epsilon} = (1 - \epsilon)F + \epsilon\Delta_x.$$

Esta nueva distribución representa la “contaminación” de la distribución F en un grado ϵ . Se observa que cuando ϵ es pequeña, $F_{x,\epsilon}$ es similar a F , tomando como medida la distancia de Kolmogorov.

Si ahora se define un funcional T que mapea cualquier distribución a un número real (por ejemplo la media, $T(X) = \mathbb{E}(X)$), entonces se puede definir la curva de influencia.

Definición 3.2.1 Dada la distribución F y Δ_x , se define la **curva de influencia** del operador T en F como

$$IF(x) = \lim_{\epsilon \downarrow 0} \frac{T(F_{x,\epsilon}) - T(F)}{\epsilon}.$$

Decimos que $T(F)$ es infinitesimalmente robusto si $IF(x)$ es acotada.

La curva de influencia, también conocida como la función de influencia, nos da una idea precisa de cómo responde nuestro estimador ante una contaminación pequeña en cualquier punto. La función de influencia en un punto x puede verse como una aproximación del cambio relativo del estimador causada por agregar una pequeña proporción de observaciones adulteradas en x .

La función mixta $F_{x,\epsilon} = (1 - \epsilon)F + \epsilon\Delta_x$ es un modelo apropiado para contaminación pues al muestrear de $F_{x,\epsilon}$ se incurre en el riesgo de obtener una “mala” observación en x con una probabilidad pequeña ϵ .

La función de influencia $IF(x)$ es una derivada direccional de T en F en la dirección de $\Delta_x - F$.

Proposición 3.2.2 La correlación de Pearson no es infinitesimalmente robusta.

Demostración Sea F una distribución normal bivariada $N\left(\bar{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ y sea

$$\Delta_{\langle x, y \rangle}(z, w) = \begin{cases} 1 & \text{si } z > x \text{ y } w > y \\ 0 & \text{en otro caso} \end{cases},$$

y sea $H = (1 - \epsilon)F + \epsilon\Delta_{\langle x, y \rangle}$ y $T(F) = \text{corr}(F)$.

Se denota como $\rho_{Z, W}^H$ a la correlación inherente a la distribución H que está dada por

$$\rho_{Z, W}^H = \frac{\mathbb{E}_H(ZW) - \mathbb{E}_H(Z)\mathbb{E}_H(W)}{\sqrt{\text{Var}_H(Z)}\sqrt{\text{Var}_H(W)}}.$$

Para cualquier función $g(Z)$ se cumple que

$$\mathbb{E}_H(g(Z)) = (1 - \epsilon)\mathbb{E}_F(g(Z)) + \epsilon g(x),$$

pues

$$\begin{aligned} \mathbb{E}_H(g(z)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(z) dH(z, w) \\ &= \int_{-\infty}^x \int_{-\infty}^y g(z) dH(z, w) + \int_{-\infty}^x \int_y^{\infty} g(z) dH(z, w) + \\ &\quad \int_x^{\infty} \int_{-\infty}^y g(z) dH(z, w) + \int_x^{\infty} \int_y^{\infty} g(z) dH(z, w) \\ &= (1 - \epsilon) \int_{-\infty}^x \int_{-\infty}^y g(z) dF(z, w) + (1 - \epsilon) \int_{-\infty}^x \int_y^{\infty} g(z) dF(z, w) + \\ &\quad (1 - \epsilon) \int_x^{\infty} \int_{-\infty}^y g(z) dF(z, w) + (1 - \epsilon) \int_x^{\infty} \int_y^{\infty} g(z) dF(z, w) + \\ &\quad \epsilon \int_x^{\infty} \int_y^{\infty} g(z) d\Delta(z, w) \\ &= (1 - \epsilon)\mathbb{E}_F(g(z)) + \epsilon g(x). \end{aligned}$$

Para el caso particular donde $g(Z) = Z$ tenemos que

$$\begin{aligned} \mathbb{E}_H(z) &= (1 - \epsilon)\mathbb{E}_F(z) + \epsilon x \\ &= \epsilon x. \end{aligned}$$

Análogamente, cuando $g(Z) = Z^2$ tenemos que

$$\begin{aligned} \mathbb{E}_H(Z^2) &= (1 - \epsilon)\mathbb{E}_F(Z^2) + \epsilon x^2 \\ &= (1 - \epsilon) + \epsilon x^2. \end{aligned}$$

Por otro lado

$$\begin{aligned}
Var_H(Z) &= \mathbb{E}_H(Z^2) - \mathbb{E}_H^2(Z) \\
&= (1 - \epsilon) + \epsilon x^2 - (\epsilon x)^2 \\
&= (1 - \epsilon)(1 + \epsilon x^2).
\end{aligned}$$

De forma análoga tenemos que

$$\mathbb{E}_H(g(Z, W)) = (1 - \epsilon)\mathbb{E}_F(g(Z, W)) + \epsilon g(x, y).$$

Ahora bien, por la definición de la distribución F sucede que $Corr_F(Z, W) = \rho = \mathbb{E}(ZW)$. Finalmente, la curva de influencia está dada por

$$\begin{aligned}
IF(x, y) &= \lim_{\epsilon \downarrow 0} \frac{T(F_{(x,y),\epsilon}) - T(F)}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{\rho_{z,w}^H - \rho_{z,w}^F}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{\frac{(1-\epsilon)(\rho+\epsilon xy)}{\sqrt{(1-\epsilon)^2(1+\epsilon x^2)(1+\epsilon y^2)}} - \rho}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{\rho + \epsilon xy - \rho \left(\sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)} \right)}{\epsilon \sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)}} \\
&= \lim_{\epsilon \downarrow 0} \frac{\rho \left(1 - \sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)} \right)}{\epsilon \sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)}} + \lim_{\epsilon \downarrow 0} \frac{xy}{\sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)}} \\
&= \lim_{\epsilon \downarrow 0} \frac{\rho \left(1 - \sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)} \right)}{\epsilon \sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)}} + xy.
\end{aligned}$$

Pero el límite de la expresión $\frac{\rho(1 - \sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)})}{\epsilon \sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)}}$, cuando $\epsilon \downarrow 0$, se indetermina (tanto el límite del denominador como el del numerador son cero), por lo que, usando L'Hopital se tiene que

$$\begin{aligned}
&\lim_{\epsilon \downarrow 0} \frac{\rho \left(1 - \sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)} \right)}{\epsilon \sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)}} \\
&= \lim_{\epsilon \downarrow 0} \frac{d/d\epsilon \left(\rho \left(1 - \sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)} \right) \right)}{d/d\epsilon \left(\epsilon \sqrt{(1 + \epsilon x^2)(1 + \epsilon y^2)} \right)}
\end{aligned}$$

$$\begin{aligned}
&= \lim_{\epsilon \downarrow 0} \frac{-\rho \frac{x^2+y^2+2\epsilon x^2 y^2}{2\sqrt{(1+\epsilon x^2)(1+\epsilon y^2)}}}{\epsilon \left(\frac{x^2+y^2+2\epsilon x^2 y^2}{2\sqrt{(1+\epsilon x^2)(1+\epsilon y^2)}} \right) + \sqrt{(1+\epsilon x^2)(1+\epsilon y^2)}} \\
&= -\rho \left(\frac{x^2+y^2}{2} \right).
\end{aligned}$$

Entonces

$$\begin{aligned}
IF(x, y) &= xy + \lim_{\epsilon \downarrow 0} \frac{\rho \left(1 - \sqrt{(1+\epsilon x^2)(1+\epsilon y^2)} \right)}{\epsilon \sqrt{(1+\epsilon x^2)(1+\epsilon y^2)}} \\
&= xy - \rho \left(\frac{x^2+y^2}{2} \right).
\end{aligned}$$

■

Proposición 3.2.3 El estimador propuesto $\hat{\rho}$ es infinitesimalmente robusto.

Demostración

Sea F la distribución Normal bivariada, $F \sim N\left(\bar{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ y sea $\Delta_{\langle x_0, y_0 \rangle}$, con $\langle x_0, y_0 \rangle \in \mathbb{R}^2$ fijos, definida como

$$\Delta_{\langle x_0, y_0 \rangle}(x, y) = \begin{cases} 1 & \text{si } x \geq x_0 \text{ y } y \geq y_0 \\ 0 & \text{si } x < x_0 \text{ ó } y < y_0 \end{cases}.$$

Se define la distribución de contaminación como $F_{x,\epsilon} = (1-\epsilon)F + \epsilon\Delta_x$ y el operador será el estimador propuesto, de modo que $T(F) = \hat{\rho} = -\cos(\pi P_F)$, donde $P_F = \int_0^\infty \int_0^\infty dF(x, y) + \int_{-\infty}^0 \int_{-\infty}^0 dF(x, y)$.

Si se define a la región A como $A = (-\infty, 0]^2 \cup [0, \infty)^2$, entonces

$$T(\Delta_{\langle x_0, y_0 \rangle}) = \begin{cases} -\cos(\pi) = 1 & \text{si } \langle x_0, y_0 \rangle \in A \\ -\cos(0) = -1 & \text{si } \langle x_0, y_0 \rangle \notin A \end{cases}.$$

Para el caso donde $\langle x_0, y_0 \rangle \in A$, la curva de influencia está dada por

$$\begin{aligned}
IF &= \lim_{\epsilon \downarrow 0} \frac{-\cos\left(\pi \left(\iint_A (1-\epsilon)dF + \iint_A \epsilon d\Delta_{\langle x_0, y_0 \rangle} \right)\right) - (-\cos(\pi P_F))}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{-\cos\left(\pi \left((1-\epsilon)P_F + \epsilon \iint_A d\Delta_{\langle x_0, y_0 \rangle} \right)\right) + \cos(\pi P_F)}{\epsilon}
\end{aligned}$$

$$\begin{aligned}
&= \lim_{\epsilon \downarrow 0} \frac{-\cos(\pi((1-\epsilon)P_F + \epsilon)) + \cos(\pi P_F)}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{-\cos(\pi P_F + \epsilon\pi(1-P_F)) + \cos(\pi P_F)}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{-(\cos(\pi P_F)\cos(\epsilon\pi(1-P_F)) - \sin(\pi P_F)\sin(\epsilon\pi(1-P_F))) + \cos(\pi P_F)}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{\cos(\pi P_F)(1 - \cos(\epsilon\pi(1-P_F))) + \sin(\pi P_F)\sin(\epsilon\pi(1-P_F))}{\epsilon} \\
&= \cos(\pi P_F) \lim_{\epsilon \downarrow 0} \frac{(1 - \cos(\epsilon\pi(1-P_F)))}{\epsilon} + \sin(\pi P_F) \lim_{\epsilon \downarrow 0} \frac{\sin(\epsilon\pi(1-P_F))}{\epsilon}
\end{aligned}$$

usando L'Hopital tenemos que

$$\begin{aligned}
&= \cos(\pi P_F) \lim_{\epsilon \downarrow 0} \frac{(\sin(\epsilon\pi(1-P_F))\pi(1-P_F))}{1} + \\
&\quad \sin(\pi P_F) \lim_{\epsilon \downarrow 0} \frac{\cos(\epsilon\pi(1-P_F))\pi(1-P_F)}{1} \\
&= \sin(\pi P_F)\pi(1-P_F),
\end{aligned}$$

que es una constante no negativa.

Ahora, para el caso en donde $\langle x_0, y_0 \rangle \notin A$ la curva de influencia es

$$\begin{aligned}
IF &= \lim_{\epsilon \downarrow 0} \frac{-\cos\left(\pi\left(\iint_A (1-\epsilon)dF + \iint_A \epsilon d\Delta_{\langle x_0, y_0 \rangle}\right)\right) - (-\cos(\pi P_F))}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{-\cos\left(\pi\left((1-\epsilon)P_F + \epsilon \iint_A d\Delta_{\langle x_0, y_0 \rangle}\right)\right) + \cos(\pi P_F)}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{-\cos(\pi((1-\epsilon)P_F)) + \cos(\pi P_F)}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{-\cos(\pi P_F - \epsilon\pi P_F) + \cos(\pi P_F)}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{-(\cos(\pi P_F)\cos(\epsilon\pi P_F) + \sin(\pi P_F)\sin(\epsilon\pi P_F)) + \cos(\pi P_F)}{\epsilon} \\
&= \lim_{\epsilon \downarrow 0} \frac{\cos(\pi P_F)(1 - \cos(\epsilon\pi P_F)) + \sin(\pi P_F)\sin(\epsilon\pi P_F)}{\epsilon} \\
&= \cos(\pi P_F) \lim_{\epsilon \downarrow 0} \frac{(1 - \cos(\epsilon\pi P_F))}{\epsilon} + \sin(\pi P_F) \lim_{\epsilon \downarrow 0} \frac{\sin(\epsilon\pi P_F)}{\epsilon},
\end{aligned}$$

usando L'Hopital tenemos que

$$\begin{aligned}
&= \cos(\pi P_F) \lim_{\epsilon \downarrow 0} \frac{(\sin(\epsilon \pi P_F)) \pi P_F}{1} + \sin(\pi P_F) \lim_{\epsilon \downarrow 0} \frac{\cos(\epsilon \pi P_F) \pi P_F}{1} \\
&= \sin(\pi P_F) \pi P_F.
\end{aligned}$$

Entonces

$$IF_{T_F}(x_0, y_0) = \begin{cases} \sin(\pi P_F) \pi (1 - P_F) & \text{si } \langle x_0, y_0 \rangle \in A \\ \sin(\pi P_F) \pi P_F & \text{si } \langle x_0, y_0 \rangle \notin A \end{cases},$$

por lo tanto $\hat{\rho}$ es infinitesimalmente robusto. ■

Los siguientes resultados pueden encontrarse en [10].

Teorema 3.2.4 Sea X y Y dos variables aleatorias continuas cuya cópula es C . Entonces la versión poblacional de la rho de Spearman para X y Y , denotada como ρ_C , esta dada por

$$\begin{aligned}
\rho_C &= 3Q(C, \Pi) \\
&= 12 \iint_{I^2} uv dC(u, v) - 3 \\
&= 12 \iint_{I^2} C(u, v) du dv - 3.
\end{aligned}$$

Definición 3.2.5 Una medida de asociación numérica k entre dos variables aleatorias continuas X y Y cuya cópula es C es una **medida de concordancia** si satisface las siguientes propiedades:

1. k está definido para todo par X, Y de variables aleatorias continuas.
2. $-1 \leq k_C \leq 1$, $k_{X, X} = 1$ y $k_{X, -X} = -1$.
3. $k_{X, Y} = k_{Y, X}$.
4. Si X y Y son independientes entonces $k_{X, Y} = k_{\Pi} = 0$.
5. $k_{X, -Y} = k_{-X, Y} = -k_{X, Y}$.
6. Si C_1 y C_2 son cópulas tales que $C_1 \prec C_2$, entonces $k_{C_1} \leq k_{C_2}$.
7. Si $\{(X_n, Y_n)\}$ es una sucesión de variables aleatorias continuas con cópulas C_n , y si $\{C_n\}$ converge puntualmente a C , entonces $\lim_{n \rightarrow \infty} k_{C_n} = k_C$.

Teorema 3.2.6 Si X y Y son variables aleatorias continuas cuya cópula asociada es C , entonces la rho de Spearman, definida como en el Teorema 3.2.4, es una medida de concordancia.

La rho de Spearman es usualmente llamada el coeficiente de correlación de “grado”. Los grados son los análogos poblacionales para rangos, es decir, si x y y son observaciones de dos variables aleatorias X y Y con funciones de distribución F y G , respectivamente, entonces los *grados* de x y y están dados por $u = F(x)$ y $v = G(y)$. Notemos que los grados u y v son observaciones de las variables aleatorias uniformes $(0, 1)$ $U = F(X)$ y $V = G(Y)$ cuya distribución conjunta es C . Como U y V tienen media $\frac{1}{2}$ y varianza $\frac{1}{12}$, la expresión para ρ_C puede ser reescrita como

$$\begin{aligned} \rho_{X,Y} = \rho_C &= 12 \iint_{I^2} uv dC(u, v) - 3 \\ &= 12\mathbb{E}(UV) - 3 \\ &= \frac{\mathbb{E}(UV) - \frac{1}{4}}{1/12} \\ &= \frac{\mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V)}{\sqrt{Var(U)}\sqrt{Var(V)}}. \end{aligned}$$

Como consecuencia, la rho de Spearman para un par de variables aleatorias continuas X y Y es idéntica al coeficiente de correlación de Pearson para los grados de X y Y , es decir, las variables aleatorias $U = F(X)$ y $V = G(Y)$.

Simulación

Ya se demostró que el estimador que se propone en el presente trabajo es robusto, esto se refleja en las simulaciones, se realizaron con una contaminación $\epsilon = 0.05$, los éxitos son casi siempre cercanos al 100 por ciento, veamos a continuación un ejemplo.

Ejemplo 3.2.7 Se toma una muestra de la distribución Normal Bivariada con media $\mu = \underline{0}$ y desviación estándar $\sigma_1 = \sigma_2 = 1$, con una probabilidad de $\epsilon = 0.05$ la muestra se contamina con observaciones provenientes de una normal bivariada con media $\mu_1 = 4$ y $\mu_2 = 6$ y $\sigma_1 = 2$ y $\sigma_2 = 4$ y $\rho_c = -0.9$. Los resultados de las simulaciones se muestran en la tabla siguiente.

Aunque el ejemplo anterior parece extremo se observa que variando solo las medias de la función que contamina obtenemos ya una mejor estimación, esto es evidente si vemos un gráfico de una muestra contaminada.

ρ	Porcentaje de éxitos(*)
0	100
0.1	100
0.2	100
0.3	100
0.4	93
0.5	74
0.6	53
0.7	78
0.8	98
0.9	100
1	100

Cuadro 1: Resultados simulación bajo contaminación

(*)Se considera éxito si la diferencia entre el estimador propuesto y la ρ real es menor que la diferencia entre la ρ de Pearson y la ρ real.

Figura 5: muestra n=500 contaminada

3.3. Corrección por Von Mises

Los funcionales estadísticos fueron introducidos por von Mises en 1936 y son empleados en la teoría de estimación. Von Mises propuso que una expansión de Taylor puede ser utilizada para aproximar funcionales estadísticos por funcionales de una forma más simple y utilizar esta aproximación para obtener resultados sobre las distribuciones asintóticas.

En la sección 2.1. se demostró que el estimador $\hat{\rho}$ es asintóticamente insesgado, es decir, para un tamaño de muestra n fijo se comete un error cuando $\rho \notin \{-1, 0, 1\}$.

Sea $f(\rho, n) = \mathbb{E}(\hat{\rho}) - \rho$ la función que denota el error esperado que se comete. Este error se comporta como una función senoidal (Figura 6).

En esta sección se utiliza el cálculo de von Mises para hacer un ajuste al estimador $\hat{\rho}$ y así minimizar el error $f(\rho, n)$.

La función de influencia de T en F , como en la Definición 3.2.1, está íntimamente ligada a la normalidad asintótica de la secuencia asociada $\{T(F_n)\}$. Se puede decir que existe una forma fuerte de la derivada de T en F , la sucesión $\{T(F_n)\}$ es asintóticamente normal. Más aún, esta forma de la derivada coincide con la función de influencia de F en T . Esto nos lleva a la conclusión de que la “Función de influencia” estima los errores estándar de $T(F_n)$.

Definición 3.3.1 Sea X_1, \dots, X_n una muestra de una población con función de distribución F y sea $T_n = T_n(X_1, \dots, X_n)$ una estadística. Si T_n puede ser escrito como un funcional T de la distribución empírica F_n , $T_n = T(F_n)$, donde

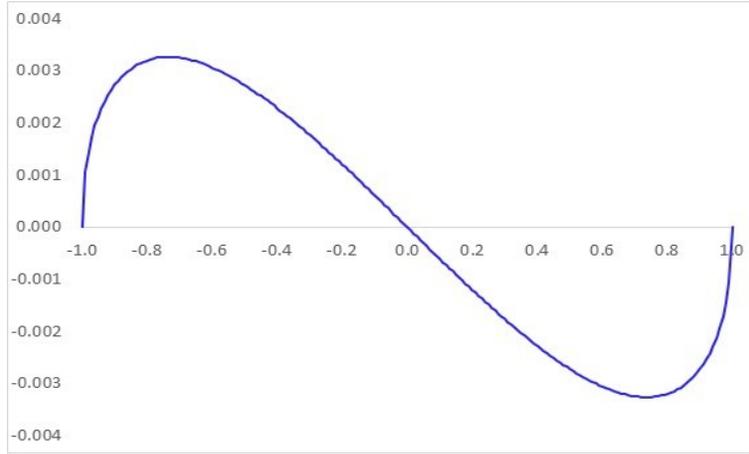


Figura 6: Error $f(\rho, n)$

T no depende de n , entonces T es llamado *funcional estadístico*. Se asume que el dominio de T contiene a la función de distribución empírica F_n para toda $n \geq 1$ y a la función de distribución poblacional F .

Ejemplo 3.3.2 Sea φ una función que toma valores reales y sea

$$T_n(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n \varphi(X_i).$$

Entonces, para una función de distribución G , el funcional definido por

$$T(G) = \int \varphi(x) dG(x)$$

satisface $T_n(X_1, \dots, X_n) = T(F_n)$.

Definición 3.3.3 Sea X_1, \dots, X_n una muestra y sea T un funcional definido en un conjunto convexo de funciones de distribución el cual contiene a la función de distribución empírica F_n y a la función de distribución poblacional F . Sea G un elemento de este conjunto convexo. Entonces la derivada de von Mises T'_F de T en F está definida por

$$T'_F(G - F) = \left. \frac{d}{dt} T(F + t(G - F)) \right|_{t=0}$$

si existe una función que toma valores reales φ_F (independiente de G) tal que

$$T'_F(G - F) = \int \varphi_F(x) d(G - F)(x).$$

La función φ_F se define de manera única exceptuando funciones que difieren por una constante debido a que $d(G - F)$ tiene medida total igual a cero. Se normaliza considerando

$$\int \varphi_F(x) dF(x) = 0.$$

La función φ_F es llamada la curva de influencia o función de influencia de T en F . Se define por la siguiente expresión:

$$\varphi_F(x) = \left. \frac{d}{dt} T(F + t(\Delta_x - F)) \right|_{t=0},$$

donde Δ_x es la función de distribución de masa puntual en x . La notación utilizada para indicar la curva de influencia es la siguiente:

$$\text{IC}(x; F, T) = \varphi_F(x).$$

Para valores grandes de n la curva de influencia mide el efecto que genera en T_n el incluir una nueva observación con valor x . La curva de influencia también proporciona la varianza asintótica cuando T_n se distribuye asintóticamente como una normal.

Definición 3.3.4 (Expansión de von Mises.) Sea T un estadístico funcional, F una función de distribución y G un elemento del dominio de T . Usando el hecho que una función A , k veces diferenciable, definida por

$$A(t) = T(F + t(G - F)), \quad t \in [0, 1]$$

puede ser representada con base en su expansión de Taylor de orden k alrededor de $t = 0$ como

$$A(t) = A(0) + A'(0)t + \dots + \frac{A^{(k)}(0)t^k}{k!} + \text{Rem}_k,$$

donde Rem_k es el residuo de k -ésimo orden, se establece la correspondencia para la siguiente expansión de primer orden del funcional estadístico T

$$T(G) = T(F) + T'_F(G - F) + \text{Rem}(G - F),$$

(se considera $t = 1$). Para $G = F_n$ se tiene la siguiente expansión:

$$\begin{aligned} T(F_n) &= T(F) + T'_F(F_n - F) + \text{Rem}(F_n - F) \\ &= T(F) + \int \varphi_F(x) d(F_n - F)(x) + \text{Rem}(F_n - F) \\ &= T(F) + \int \varphi_F(x) dF_n(x) + \text{Rem}(F_n - F) \end{aligned}$$

debido a que $\int \varphi_F(x) dF(x) = 0$. Esta última igualdad recibe el nombre de *expansión de von Mises* de T en F .

Observación 3.3.5 (Método Delta.) El término lineal de la expansión de von Mises es

$$\int \varphi_F(x) dF_n(x) = \frac{1}{n} \sum_{i=1}^n \varphi_F(X_i),$$

por lo que se cumple

$$\sqrt{n}(T(F_n) - T(F)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi_F(X_i) + \sqrt{n} \text{Rem}(F_n - F).$$

Si

$$0 < \mathbb{E}[\varphi_F^2(X)] = \sigma^2 < \infty$$

y si

$$\sqrt{n} \text{Rem}(F_n - F) \xrightarrow{\mathcal{P}} 0$$

entonces, por el Teorema Central del Límite y el Teorema de Slutsky se cumple

$$\sqrt{n}(T(F_n) - T(F)) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

si $n \rightarrow \infty$.

Proposición 3.3.6 La expresión de primer orden para modelar el error está dada por

$$\text{Ajuste} = \pi \sin(\pi P_F) (p - P_F), \quad (36)$$

donde $p = \int_{-\infty}^0 \int_{-\infty}^0 f_{X,Y}(x,y) dx dy$ y P_F denota el estimado de p .

Demostración

Sea $A(t) = T(F + t \cdot (G - F))$ con $F = F_n$ y $G = F$, utilizando la expansión de von Mises de la Definición 3.3.4, para la aproximación de orden uno tenemos que $A(t) = A(0) + A'(0)t$. Entonces

$$T(F) = T(F_n) + \frac{d}{dt} T(F_n + t(F - F_n))|_{t=0}.$$

Entonces

$$\text{Error} = T(F) - T(F_n) = \frac{d}{dt} T(F_n + t(F - F_n))|_{t=0},$$

con $T(H) = -\cos(\pi \int_{\Omega^*} dH)$, en nuestro caso $\Omega^* = (-\infty, 0] \times (-\infty, 0]$. Entonces

$$\begin{aligned}
T(F_n + t(F - F_n)) &= T((1-t)F_n + tF) \text{ combinación convexa} \\
&= -\cos\left(\pi \iint_{\Omega^*} (1-t) dF_n + \pi \iint_{\Omega^*} t dF\right).
\end{aligned}$$

Entonces

$$\begin{aligned}
\frac{d}{dt}T(F_n + t(F - F_n)) &= \sin\left(\pi \iint_{\Omega^*} (1-t) dF_n + \pi \iint_{\Omega^*} t dF\right) \cdot \\
&\quad \left(-\pi \iint_{\Omega^*} dF_n + \pi \iint_{\Omega^*} dF\right).
\end{aligned}$$

Evaluando en $t = 0$

$$\begin{aligned}
\frac{d}{dt}T(F_n + t(F - F_n))|_{t=0} &= \pi \sin\left(\pi \iint_{\Omega^*} dF_n\right) \cdot \left(\iint_{\Omega^*} dF - \iint_{\Omega^*} dF_n\right) \\
&= \pi \sin(\pi P_F) \cdot (p - P_F).
\end{aligned}$$

■

De manera que podemos hacer una corrección a nuestro estimador

$$\begin{aligned}
\hat{\rho} &= -\cos(2\pi Y) + Ajuste \\
&= -\cos(2\pi Y) + \pi \sin(\pi P_F) (p - P_F).
\end{aligned}$$

Se puede ver en la figura 7 que con el primer término ya se comporta bastante parecido a $f(n, \rho)$. Esta gráfica se hizo con una muestra de la normal bivariada de tamaño 200. Seguimos cometiendo un error pero es bastante más pequeño que antes, si vemos ahora la figura 8 observamos que para el mismo tamaño de muestra $n = 200$ el error máximo cometido es del orden de 0.0000025. Una observación importante es que el nuevo gráfico para el error que se comete muestra que en el cero y en un intervalo alrededor de él nuestro estimador comete un error de cero, lo cual lo hace insesgado ya no solo en el cero si no en una vecindad que lo contenga.

Simulación

Se realizaron simulaciones con el fin de comparar ambas estimaciones, la usual ρ de pearson y el estimador que se propone con la corrección de von Mises. Las simulaciones se hicieron con tamaños de muestra que van desde 30 hasta 1000. En la tabla siguiente podemos observar que cuando tenemos una muestra de tamaño 30 el estimador propuesto es mejor en el intervalo $[0, 0.4)$, dicho intervalo se incrementa a medida que el tamaño de muestra aumenta, podemos

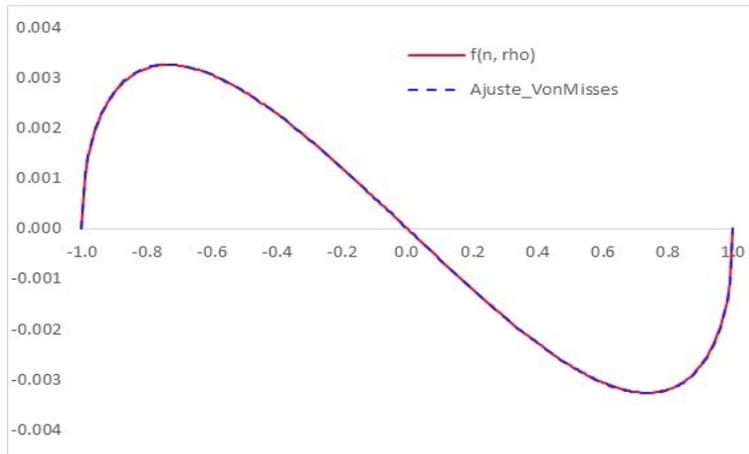


Figura 7: Ajuste

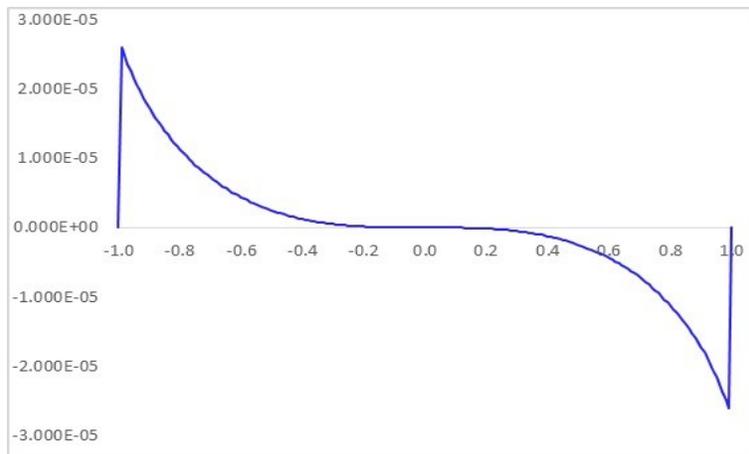


Figura 8: Error bajo el nuevo ajuste

ver que para el tamaño de muestra $n = 1000$ el estimador es más eficaz que la ρ de Pearson en el intervalo $[0, 0.8)$, véase cuadro 2. Este efecto se puede observar en la figura 9, pues para el tamaño de muestra 1,000 (T1000) vemos que la línea queda siempre por encima de 0.5, lo que significa que cuando la muestra crece el estimador que se propone es más cercano al valor real de la ρ que la ρ de Pearson.

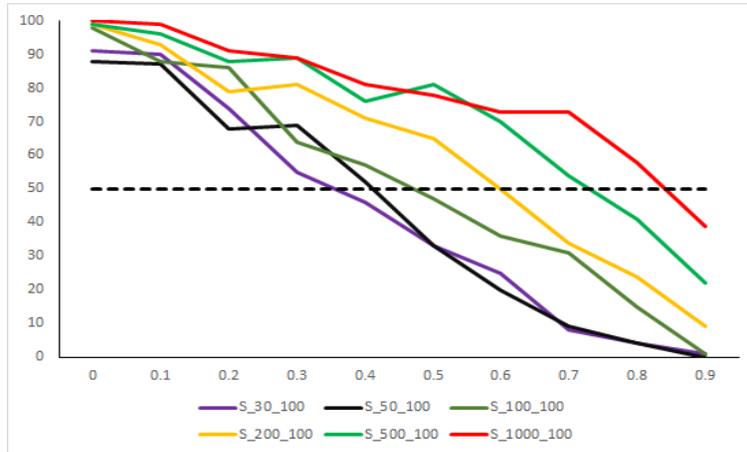


Figura 9: Porcentaje de éxitos

ρ	$n = 30$	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
0	91	88	98	99	99	100
0.1	90	87	88	93	96	99
0.2	74	68	86	79	88	91
0.3	55	69	64	81	89	89
0.4	46	52	57	71	76	81
0.5	33	33	47	65	81	78
0.6	25	20	36	50	70	73
0.7	8	9	31	34	54	73
0.8	4	4	15	24	41	58
0.9	1	0	1	9	22	39

Cuadro 2: Resultados simulación bajo corrección von Mises

4. Conclusiones y posibles extensiones

En este trabajo se propuso un estimador para la correlación de una muestra que proviene de una distribución Normal Bivariada, este estimador tiene algunas ventajas respecto a la Rho de Pearson que suele usarse, como el hecho de ser un estimador Robusto. Si bien nuestro estimador no es insesgado se propone una corrección (la de primer orden de Von Mises) que logra ser insesgado en un intervalo alrededor del cero, este intervalo es mayor a medida que el tamaño de muestra crece y la varianza del estimador es mucho menor en este intervalo.

Todo el desarrollo que se muestra es para una Normal Bivariada, sin embargo el proceso se puede extender a dimensiones mayores tomando pares de las marginales y repitiendo el proceso aquí descrito.

Como ejemplo, sea (X, Y, Z) un vector normal multivariado con media $(0, 0, 0)$

y matriz de varianza covarianza

$$\begin{pmatrix} 1 & \rho_{x,y} & \rho_{x,z} \\ \rho_{x,y} & 1 & \rho_{y,z} \\ \rho_{x,z} & \rho_{y,z} & 1 \end{pmatrix}.$$

Para estimar las tres correlaciones primero se realizan simulaciones de la normal multivariada, asegurándonos que los valores de las correlaciones cumplen con estar dentro de la región descrita en el capítulo 2, posteriormente se toman los vectores de la muestra en parejas, como si de una normal bivariada se tratara, y se estima la correlación correspondiente.

Se realizó una prueba con un tamaño de muestra $n = 1,000$ y 200 iteraciones para $\rho_{x,y} = -0.01$, $\rho_{x,z} = 0.02$ y $\rho_{y,z} = 0.13$, es decir, un punto de la región cercana al cero y se obtuvieron los siguientes resultados.

	ρ	Pearson	Estimador	Éxito
$\rho_{x,y}$	-0.01	-0.008990054	-0.009987653	1
$\rho_{x,z}$	0.02	0.023591647	0.019980360	1
$\rho_{y,z}$	0.13	0.128669083	0.130229299	1

Cuadro 3: Resultados simulación en R^3

También se realizaron mil corridas para analizar la eficacia del estimador, con tamaños de muestra $t = 100, 250$ y 500 y siempre con 200 iteraciones en cada corrida. A continuación se presentan los resultados de algunos valores.

	ρ	$t = 100$	$t = 250$	$t = 500$
$\rho_{x,y}$	-0.1	87.7	91.1	92.7
$\rho_{x,z}$	0.2	72.2	82.4	88.1
$\rho_{y,z}$	0.8	2.3	13	29.1

Cuadro 4: Corrida 1

	ρ	$t = 100$	$t = 250$	$t = 500$
$\rho_{x,y}$	0	97	99	99.5
$\rho_{x,z}$	0.5	35.8	54.1	63.5
$\rho_{y,z}$	0.34	52.5	70.6	77.3

Cuadro 5: Corrida 2

	ρ	$t = 100$	$t = 250$	$t = 500$
$\rho_{x,y}$	0.2	73.3	84	86.4
$\rho_{x,z}$	0.7	11.3	27	41.8
$\rho_{y,z}$	0.5	35.3	54.1	66

Cuadro 6: Corrida 3

	ρ	$t = 100$	$t = 250$	$t = 500$
$\rho_{x,y}$	0.7	14.9	26	43.4
$\rho_{x,z}$	0.7	12.3	26	40.4
$\rho_{y,z}$	0.5	36	52.3	65.3

Cuadro 7: Corrida 4

	ρ	$t = 100$	$t = 250$	$t = 500$
$\rho_{x,y}$	0.7	13.5	27.2	42.3
$\rho_{x,z}$	0.85	1	7.2	17.5
$\rho_{y,z}$	0.5	37.6	51	64.9

Cuadro 8: Corrida 5

Esto indica que, como en el caso bivariado, se logra mejorar la estimación en una vecindad del cero, vecindad que aumenta a medida que el tamaño de muestra crece, quedaría como tarea pendiente de este trabajo la caracterización de dicha región y cómo varía cuando el tamaño de muestra cambia.

Otra manera de realizar la estimación de los parámetros es usar una expresión cuya deducción es análoga a la de la Proposición 2.2.1

$$F_{X,Y,Z}(0,0,0) = \frac{1}{2^3} + \frac{1}{4\pi} (\arcsin(\rho_{x,y}) + \arcsin(\rho_{x,z}) + \arcsin(\rho_{y,z})). \quad (37)$$

En este caso se usarán las dos estimaciones más cercanas a los valores reales de las correlación y con la expresión anterior se despejará la tercera.

Para este segundo método se deben contemplar dos etapas, en una primera etapa se estima correctamente el volumen de la caja $[0, \frac{1}{2}]^3$, lo que equivale a estimar $F_{X,Y,Z}(0,0,0)$, posteriormente se estiman las dos ρ 's mas pequeñas, se escogen

las más pequeñas porque ya hemos visto que es en donde mejor estimamos los coeficientes de correlación. Estas dos ρ 's se estiman usando el método bivariado y para la tercera ρ simplemente hacemos el despeje de la Ecuación 37.

Veamos un ejemplo, se toma un tamaño de muestra $t = 50$ y $\rho_{x,y} = 0.06$, $\rho_{x,z} = 0.2$ y $\rho_{y,z} = 0.9$ con 100 iteraciones, se espera que la peor estimación sea la de $\rho_{y,z}$ pues es la más alejada del cero.

	ρ	Pearson	Estimador	Éxito
$\rho_{x,y}$	0.06	0.06836937	0.06098201	1
$\rho_{x,z}$	0.2	0.21010171	0.20330565	1
$\rho_{y,z}$	0.9	0.89698803	0.91093044	0

Cuadro 9: Resultados simulación en R^3

Como se observa en el cuadro anterior, nuestro estimador de $\rho_{y,z}$ es el único en donde no se tiene una mejora respecto al estimador de Pearson. Se estima que el valor de $F_{X,Y,Z}(0,0,0)$ y ahora despejamos ρ_{yz} .

$$\begin{aligned}
 \widehat{\rho}_{yz} &= \sin \left(\left(F_{X,Y,Z}(0,0,0) - \frac{1}{23} \right) (4\pi) - (\arcsin(\widehat{\rho}_{xy}) + \arcsin(\widehat{\rho}_{xz})) \right) \\
 &= \sin \left((0.10990 \cdot 4\pi) - (\arcsin(0.06098201) + \arcsin(0.20330565)) \right) \\
 &= 0.8980915
 \end{aligned}$$

Obtenemos así un mejor estimador también para el caso de valores alejados del cero, aunque se debe aclarar que en el caso de este sencillo ejemplo resultó ser así pues los valores de $\rho_{x,y}$ y $\rho_{x,z}$ están muy cercanas al cero y por tanto su estimación es mejor.

Finalmente, este trabajo puede también extenderse mediante la comparación del estimador propuesto con algunos otros que tienen amplio uso práctico, por ejemplo el estimador bayesiano que asume como distribución a priori la uniforme o arcoseno, el estimador de máxima verosimilitud (MLE), etcétera.

Referencias

- [1] Bensimhoun, M.(2009), N-dimensional cumulative function, and other useful facts about Gaussians and normal densities.
- [2] Bourbaki, N. (1989), *Elements of Mathematics*, Algebra I, Springer-Verlag.
- [3] Cuculescu, I. and Theodorescu, R. (2001) Copulas: Diagonals, tracks. *Rev. Roumaine Math. Pures Appl.* **6**,731-742.
- [4] Fosdick, B. K. and Raftery A.E. (2012), Estimating the correlation in bivariate normal data with known variances and small sample sizes, *Amer. Statist.* **66** (1), 34-41.
- [5] Fredericks, G.A., Nelsen, R.B., and Rodríguez-Lallena, J.A. (2005), Copulas with fractal supports. *Insurance Math. Econom.*, **44**, 199-213.
- [6] Genest, C. and Rémillard B. (2014), Test of independence and randomness based on the empirical copula process, *Sociedad de Estadística e Investigación Operativa* **13** (2), 335-369.
- [7] González-Barrios, J. M. and Hernández-Cedillo M. M. (2013), Sample d -copula of order m . *Kibernetika*, **49** (5), 663-691.
- [8] González-Barrios, J. M. (2011) *Lecture notes on Probability Theory*. Monografías, IIMAS.
- [9] Mikusiński, P. and Taylor, M.D. (2010). Some approximations of n -copulas. *Metrika*, **72**, 385-414.
- [10] Nelsen, R.B. (2006), *An introduction to copulas (In Springer Series in Statistics)* Springer, New York, second edition.
- [11] Trutschnig, W. and Fernández-Sánchez, J. (2012), Idempotent and multivariate copulas with fractal support, *J. Statist. Plan. Infer.* **142**, 3086-3096.