Análisis de un indicador para selección de funciones cópula

Camilo Andrés Velasco Mena

Dr. José Rafael Tovar Cuevas, Director
Dr. Miguel Ángel Marmolejo Lasprilla, Codirector

Trabajo presentado como requisito parcial
para optar al título de Matemático

Universidad del Valle
Facultad de Ciencias Naturales y Exactas
Departamento de Matemáticas
Santiago de Cali, Colombia
2025

Resumen

En este trabajo se realiza una descripción detallada de las funciones cópula bivariadas y sus propiedades principales, destacando su relevancia en el estudio de las relaciones de dependencia entre variables aleatorias. Se detallan algunas familias de cópulas paramétricas ampliamente utilizadas y se describen técnicas para la estimación de sus parámetros. Además, se presenta un indicador utlizado para seleccionar una función cópula adecuada dado un conjunto de datos empíricos. Finalmente, se propone una prueba de bondad de ajuste basada en dicho indicador, brindando una herramienta para evaluar la idoneidad de las funciones cópula seleccionadas.

Palabras clave: Funciones cópula, Dependencia, Estimación, Bondad de ajuste.

Agradecimientos

En primer lugar, doy gracias a Dios por ser quien ha guíado mi camino y me ha permitido llegar a este punto. Agradezco inmensamente a mis padres, Andrés y Aura, por su amor, por su apoyo incondicional y por ser quienes han hecho posible la culminación de este proceso. También a mi abuela Graciela, porque estuvo atenta y me acompañó en los inicios de este camino; a mi hermano Samuel y a mi abuela Emira, por escucharme siempre y darme los mejores consejos.

Agradezco también a mis asesores, los profesores Rafael Tovar y Miguel Marmolejo, por haber sugerido el tema y haberme recibido como estudiante en esta etapa; su dedicación y acompañamiento han sido fundamentales para el desarrollo y culminación de este trabajo de grado, y es mucho lo que he aprendido de ellos.

Muchas gracias a las profesoras Paula Bran y Angélica Caicedo, así como al programa de Olimpiadas Regionales de Matemáticas, por haber confiado en mí y haberme dado la oportunidad de crecer y aprender tanto como monitor de ORM. También a las profesoras Heliana Arias y Liliana Posada por aconsejarme y guiarme en varios momentos.

Gracias a mi amigo Steven Sánchez por su disposición permanente para escucharme y animarme en los momentos en que más lo he necesitado, así como por enseñarme tantas cosas en cada conversación que tenemos. Gracias a mis compañeros Catalina Quincosis y Samir Herrera, por su amistad y por todas las largas y productivas sesiones de estudio que tuvimos durante estos años.

Finalmente, agradezco a Julián Peñaranda por enseñarme muchas cosas de estadística, y a Jeniffer Burbano por su apoyo motivacional y su ayuda para superar obstáculos que parecían insalvables.

A todos, mis más sinceros y profundos agradecimientos.

Índice general

Índice de figuras

Índice de tablas

Introducción

La relación de dependencia entre variables aleatorias es uno de los temas más estudiados en probabilidad y estadística (capítulo 5 de Nelsen [21]). En diversos contextos se presentan situaciones en las cuales es importante establecer el grado de dependencia entre múltiples datos que se pueden representar con variables aleatorias. Esto es especialmente relevante en áreas como hidrología, medicina o economía. Por ejemplo, ciertos eventos hidrológicos como inundaciones o tormentas se caracterizan a través de variables aleatorias correlacionadas tales como profundidad, volumen, duración de los flujos de agua, entre otras (Favre et al. [6]). Por su parte, en medicina, se puede determinar la efectividad de un tratamiento para pacientes con cáncer analizando la relación entre variables como tiempo de progresión del tumor, supervivencia libre de enfermedad y supervivencia libre de propagación (Emura et al. [5]). En economía, la dependencia entre variables financieras tales como mercados internaciones y diferentes clases de activos y monedas puede influir en las decisiones de inversión y la evaluación de sus riesgos (Genest et al. [11]).

Existen diferentes herramientas que permiten estudiar relaciones de dependencia; entre estas se encuentran las medidas de asociación, tales como los coeficientes de correlación o de concordancia. Adicionalmente se tienen las funciones cópula. Desde un primer punto de vista, las cópulas son funciones que relacionan la función de distribución multivariada de un vector aleatorio con sus respectivas funciones de distribución marginales. Esto se hace explícito por medio del Teorema de Sklar, que será presentado en la sección 2.2. En este sentido, se puede decir que la función cópula permite analizar la forma en que dos o más variables aleatorias se distribuyen conjuntamente.

Para conocer el grado de dependencia entre variables aleatorias se cuenta con diversas medidas de asociación, como lo son el τ de Kendall y el ρ de Spearman, que se detallan en la sección 2.5. Tal como se muestra en dicha sección, estas medidas se pueden obtener en términos de las funciones cópula. Otro ejemplo de este tipo de medidas es la propuesta en 2005 por Anjos y Kolev [1] y retomada en 2015 por Ledwina [18], la cual se construye a partir de una cópula bivariada y se expone en la sección 3.3.

El desarrollo de diversos procedimientos estadísticos, tales como los que se presentan en la sección 1.2, está basado en el supuesto de que las variables aleatorias en cuestión provienen de una familia de distribuciones indexada por un parámetro.

Sin embargo, en muchos casos no es posible identificar tal familia de distribuciones. Esto supone la necesidad de contar con métodos no paramétricos que permitan establecer un modelo adecuado para una muestra concreta. Dichos métodos se conocen como *pruebas de bondad de ajuste* (capítulos 13 y 14 de Bain y Engelhardt [2]). Para el caso de muestras aleatorias bivariadas, se han desarrollado varias pruebas de bondad de ajuste para cópulas bivariadas, tales como las que aparecen en Deheuvels [4], Fermanian [7], Genest y Favre [10], Genest y Rémillard [12] y Genest et al. [13]. Adicionalmente, en Tovar et al. [24] se contruyó un indicador, basado en la medida presentada en Ledwina [18], con el cual se puede identificar, entre un conjunto de cópulas previamente definidas (tales como las de la sección 2.4), cuál es "la que mejor se ajusta" a las observaciones que se tienen.

De acuerdo a lo anterior, en este trabajo se desarrolla una prueba de bondad de ajuste para funciones cópula, basada en el indicador establecido por Tovar et al. [24]. El capítulo 1 está dedicado a presentar algunos conceptos básicos de probabilidad y estadística necesarios en los capítulos subsiguientes. El capítulo 2 se enfoca en exhibir y explicar lo que son las funciones cópula, presentar algunas familias comunes de ellas e introducir el concepto de medidas de asociación. En el capítulo 3 se muestra detalladamente que la medida presentada en Ledwina [18] es un coeficiente de correlación entre ciertas variables aleatorias; de igual manera se define el concepto de cópula empírica, necesario para comprender el indicador dado por Tovar et al. [24], así como dos métodos ya documentados para la estimación del parámetro de una cópula. Finalmente, en el capítulo 4 se establece una variable aleatoria en términos de la cópula empírica, que facilitará la deducción de la distribución muestral del que será el estadístico de prueba; con esto se presentará una prueba de hipótesis que permite decidir si una función cópula específica resulta adecuada para un conjunto de observaciones bivariadas dadas.

Capítulo 1

Conceptos y definiciones básicas

En este capítulo se presentan algunos conceptos básicos de probabilidad y estadística, los cuales serán necesarios para el desarrollo de los capítulos posteriores.

1.1. Conceptos básicos de Probabilidad

El contenido de esta sección se basa principalmente en Bain y Engelhardt [2], Casella y Berger [3] y Mood et al. [20].

1.1.1. Espacio de probabilidad

Definición 1.1.1. *Un **espacio de probabilidad** es una terna $(\Omega, \mathcal{F}, P)$, donde:*

1. *Ω es un conjunto no vacío (llamado espacio muestral).*
2. *$\mathcal{F} \subseteq \mathcal{P}(\Omega)$ es una σ-álgebra (llamada familia de eventos) que satisface:*
   - *(σ1) $\Omega \in \mathcal{F}$.*
   - *(σ2) Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$.*
   - *(σ3) Si $A_n \in \mathcal{F}$, $n = 1, 2, 3, \ldots$, entonces $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.*
3. *$P : \mathcal{F} \to [0, 1]$ es una función (llamada función de probabilidad) que satisface:*
   - *(P1) Para todo $A \in \mathcal{F}$, $0 \leq P(A) \leq 1$.*
   - *(P2) $P(\Omega) = 1$.*
   - *(P3) Si $A_n \in \mathcal{F}$, $n = 1, 2, 3, \ldots$, es una sucesión disjunta de eventos, esto es, $A_i \cap A_j = \varnothing$ para $i \neq j$, con $i, j = 1, 2, \ldots, n, \ldots$, entonces*

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{n} P(A_n).$$

1.1.2. Probabilidad condicional y eventos independientes

Definición 1.1.2. *Sean $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $B \in \mathcal{F}$ con $P(B) > 0$. Para cada $A \in \mathcal{F}$ se define la* ***probabilidad condicional*** *de A dado B por:*

$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$

Definición 1.1.3. *Dos eventos $A, B \in \mathcal{F}$ son* ***independientes*** *si $P(A \cap B) = P(A)P(B)$. De manera general, se dice que los n eventos $A_1, A_2, \ldots, A_n$ son independientes si para cada conjunto de índices $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, n\}$, $2 \leq k \leq n$, se cumple*

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{i=1}^{k} P(A_{i_j}).$$

1.1.3. Variables aleatorias y función de distribución

Definición 1.1.4. *Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad.*

a) Una función $X : \Omega \to \mathbb{R}$ es una ***variable aleatoria****, si para cada $x \in \mathbb{R}$ se verifica $X^{-1}((-\infty, x]) := \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$.*

b) La ***función de distribución*** *de una variable aleatoria X, denotada por $F_X(\cdot)$ es la función de puntos $F_X : \mathbb{R} \to [0, 1]$ dada por $F_X(x) = P(X \leq x) = P[\{\omega : X(\omega) \leq x\}]$ para todo $x \in \mathbb{R}$.*

Las siguientes son propiedades de la función de distribución de X:

(FD1) Para todo $x \in \mathbb{R}$, $0 \leq F_X(x) \leq 1$.

(FD2) $F_X(\cdot)$ es creciente, i.e., $\forall x, y \in \mathbb{R}$, $x \leq y \Rightarrow F_X(x) \leq F_X(y)$.

(FD3) $F_X(\cdot)$ es continua por la derecha, i.e., $\forall a \in \mathbb{R}$, $F(a^+) \equiv \lim_{x \to a^+} F_X(x) = F_X(a)$.

(FD4) $F_X(\infty) \equiv \lim_{x \to \infty} F_X(x) = 1$ y $F_X(-\infty) \equiv \lim_{x \to -\infty} F_X(x) = 0$.

Tipos de funciones de distribución

Los dos tipos de funciones de distribución más comunes son los siguientes:

1. **Discretas:** Corresponden a funciones de distribución $F_X(\cdot)$ que son escalonadas con saltos en un conjunto finito o infinito contable $S = \{x_1, x_2, x_3, \ldots\} \subseteq \mathbb{R}$, tales que

   a) $P(\{x_k\}) = F_X(x_k) - F_X(x_k^-) > 0$, con $x_k \in S$.

b) $\sum_{k\geq 1} P(\{x_k\}) = 1.$

La función $p : \mathbb{R} \to [0,1]$, definida por

$$p(x) := \begin{cases} F_X(x_k) - F_X(x_k^-), & \text{si } x = x_k \in S, \\ 0, & \text{si } x \notin S. \end{cases}$$

se llama **función de masa** asociada a $F_X(\cdot)$. Se tiene que $F_X(x) = P((-\infty, x]) = \sum_{\{i:x_i \leq x\}} p(x_i)$.

2. **Continuas (absolutamente):** Corresponden a funciones de distribución $F_X(\cdot)$ tales que

$$F_X(x) = \int_{-\infty}^{x} f(t)dt,$$

donde $f : \mathbb{R} \to [0, \infty)$ es una función, llamada **función de densidad**, tal que

a) $f(\cdot) \geq 0.$

b) $\int_{-\infty}^{\infty} f(t)dt = 1.$

Definición 1.1.5. *Sea $p_X(\cdot)$ o $f_X(\cdot)$ la función de masa o densidad de la variable aleatoria X.*

*a) Sea $g(\cdot)$ una función. La **esperanza o media** de $g(X)$, denotada como $E[g(X)] \equiv \mu_{g(X)}$, se define por:*

$$E[g(X)] := \begin{cases} \sum_{i\geq 1} g(x_i)p_X(x_i), & \textit{si } X \textit{ es discreta,} \\ \int_{-\infty}^{\infty} g(x)f_X(x)dx, & \textit{si } X \textit{ es continua,} \end{cases}$$

cuando la serie o la integral converge absolutamente.

*b) La **varianza** de X, denotada como $Var(X) \equiv \sigma_X^2$, se define por:*

$$Var(X) := \begin{cases} \sum_{i\geq 1} (x_i - \mu_X)^2 p_x(x_i), & \textit{si } X \textit{ es discreta,} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)dx, & \textit{si } X \textit{ es continua.} \end{cases}$$

Las siguientes son propiedades de la esperanza y la varianza:

1. $E[c] = c$, $c \in \mathbb{R}$.

2. $E[c_1 g_1(X) + c_2 g_2(X)] = c_1 E[g_1(X)] + c_2 E[g_2(X)]$, $c_1, c_2 \in \mathbb{R}$.

3. Si $g(x) \geq 0$ para todo $x \in \mathbb{R}$, entonces $E[g(X)] \geq 0$.

4. $Var[cX] = c^2 Var[x]$, $c \in \mathbb{R}$.

5. $Var[X] = E[X^2] - E^2[X]$.

6. $Var[aX + b] = a^2 Var[x]$, $a, b \in \mathbb{R}$.

1.1.4. La transformada inversa

Sea X una variable aleatoria continua con función de distribución $F_X(x)$. Si $F_X(x)$ es estrictamente creciente, entonces F_X^{-1} está bien definida por $x = F_X^{-1}(y)$ si y solo si $F_X^{-1}(x) = y$. Si $F_X(x)$ no es estrictamente creciente, es decir, es constante en algún intervalo $[x_1, x_2]$, $F_X^{-1}(x)$ no está bien definida, pues para cualquier $x_1 \leq x \leq x_2$ se tendría que $F_X(x) = y$. Este problema se soluciona definiendo

$$F_X^{-1}(y) = \text{ínf}\,\{x \in \mathbb{R} : y \leq F_X(x)\}. \tag{1.1}$$

Note que ambas definiciones para $F_X^{-1}(y)$ son equivalentes cuando $F_X(x)$ es estrictamente creciente (ver Casella y Berger [3], p. 54).

De acuerdo a lo anterior, se puede enunciar el siguiente teorema:

Teorema 1.1.1. *Si X es una variable aleatoria continua con función de distribución $F_X(\cdot)$, entonces la variable aleatoria $U := F_X(X)$ tiene distribución uniforme en $(0, 1)$. Recíprocamente, si U tiene distribución uniforme en $(0, 1)$, entonces $X := F_X^{-1}(U)$ es una variable aleatoria con función de distribución $F_X(\cdot)$.*

Demostración. Note que $F_U(u) = P[U \leq u] = P[F_X(X) \leq u] = P[X \leq F_X^{-1}(u)] = F_X(F_X^{-1}(u)) = u$, para $0 < u < 1$, la cual es la función de distribución de una variable aleatoria uniforme en $(0, 1)$.

Recíprocamente, note que $P[X \leq x] = P[F_X^{-1}(U) \leq x] = P[U \leq F_X(x)] = F_U(F_X(x)) = F_X(x)$, por lo tanto X tiene función de distribución $F_X(x)$. □

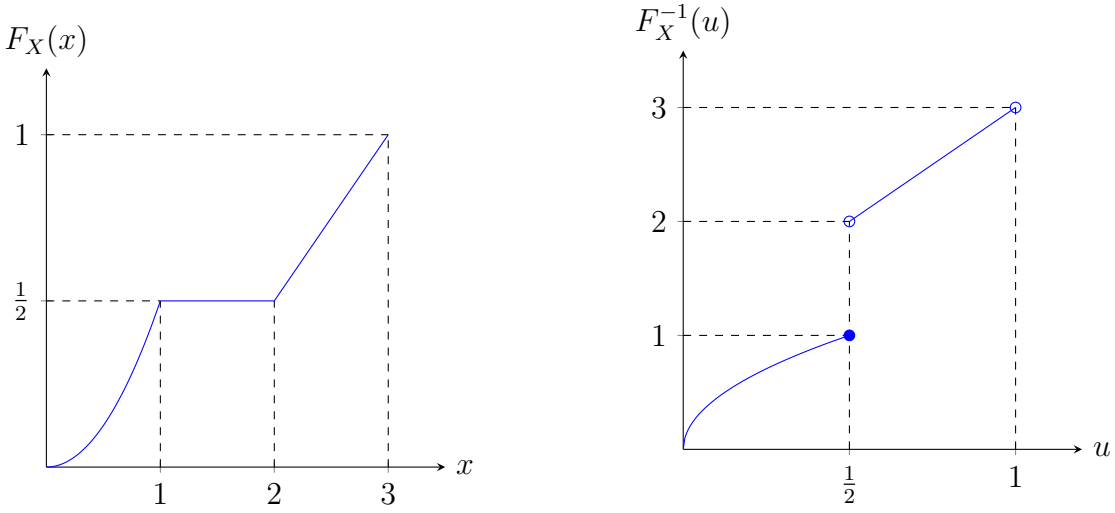
El siguiente ejemplo muestra cómo se deduce la función F_X^{-1} a partir de la definición dada en la ecuación (1.1), cuando F_X no es estrictamente creciente.

Ejemplo 1.1.1. *Sea X una variable aleatoria con función de distribución $F_X(x)$, donde:*

$$F_X(x) = \begin{cases} 0, & si\ x < 0, \\ \frac{x^2}{2}, & si\ 0 \leq x < 1, \\ \frac{1}{2}, & si\ 1 \leq x \leq 2, \\ \frac{1}{2}(x-1), & si\ 2 \leq x \leq 3, \\ 1, & si\ x \geq 3. \end{cases}$$

De acuerdo con la ecuación (1.1) *se tiene que*

$$F_X^{-1}(u) = \begin{cases} 0, & si\ u < 0, \\ \sqrt{2u}, & si\ u < \frac{1}{2}, \\ 1, & si\ u = \frac{1}{2}, \\ 2u+1 & si\ \frac{1}{2} \leq u < 1, \\ 1, & si\ u \geq 1. \end{cases}$$



(a) Función $F_X(x)$. **(b)** Función $F_X^{-1}(u)$.

Figura 1.1: Gráficas de las funciones del Ejemplo 1.1.1.

De acuerdo con el Teorema 1.1.1, $F_X^{-1}(U)$ resulta ser una variable aleatoria con función de distribución $F_X(u)$, si U es una variable aleatoria con distribución uniforme en $(0,1)$.

1.1.5. Vectores aleatorios y función de distribución multivariada

Definición 1.1.6. *Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Un **vector aleatorio** $n-$dimensional es una función $\mathbf{X} \equiv (X_1, X_2, \ldots, X_n) : \Omega \to \mathbb{R}^n$, tal que cada una de sus componentes $X_i : \Omega \to \mathbb{R}$ es una variable aleatoria; $i = 1, 2, \ldots, n$.*

De manera similar al caso de una variable aleatoria, se define la función de distribución multivariada de un vector aleatorio.

Definición 1.1.7. *La **función de distribución** del vector aleatorio* $\mathbf{X}$ *es la función* $F_{\mathbf{X}} : \mathbb{R}^n \to [0,1]$ *definida por*

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &\equiv F_{(X_1,X_2,\ldots,X_n)}(x_1,x_2,\ldots,x_n) := P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n), \\ &\equiv P(X_1 \in (-\infty, x_1], X_2 \in (-\infty, x_2], \ldots, X_n \in (-\infty, x_n]), \\ &\equiv P(\cap_{i=1}^n X_i \in (-\infty, x_i]). \end{aligned}$$

Esta función tiene propiedades similares a las de la función de distribución de la variable aleatoria X. A continuación se presentan estas propiedades para el caso $n = 2$; el caso general es análogo.

(FD1) $0 \leq F_{(X_1,X_2)}(x_1,x_2) \leq 1$ para todo $(x_1,x_2) \in \mathbb{R}^2$.

(FD2) $F_{(X_1,X_2)}(\cdot,\cdot)$ es creciente en cada variable, esto es,

a) $x_1 \leq x_1' \Rightarrow F_{(X_1,X_2)}(x_1,x_2) \leq F_{(X_1,X_2)}(x_1',x_2)$ para $x_1, x_1', x_2 \in \mathbb{R}$.

b) $x_2 \leq x_2' \Rightarrow F_{(X_1,X_2)}(x_1,x_2) \leq F_{(X_1,X_2)}(x_1,x_2')$ para $x_1, x_2, x_2' \in \mathbb{R}$.

(FD3) $F_{(X_1,X_2)}(\cdot,\cdot)$ es continua por la derecha en cada variable, esto es,

$$F_{(X_1,X_2)}(a^+,b) = F_{(X_1,X_2)}(a,b) = F_{(X_1,X_2)}(a,b^+),$$

donde $a, b \in \mathbb{R}$.

(FD4) *a)* $F_{(X_1,X_2)}(-\infty,y) = 0 = F_{(X_1,X_2)}(x,-\infty)$, con $(x,y) \in \mathbb{R}^2$.

b) $F_{(X_1,X_2)}(x_1,\infty) = F_{X_1}(x_1)$ y $F_{(X_1,X_2)}(\infty,x_2) = F_{X_2}(x_2)$.

c) $F_{(X_1,X_2)}(\infty,\infty) = 1$.

(FD5) Para $x_1 < x_1'$ y $x_2 < x_2'$ se verifica

$$\begin{aligned} &P[x_1 < X_1 \leq x_1', x_2 < X_2 \leq x_2'] = \\ &\quad F_{(X_1,X_2)}(x_1',x_2') - F_{(X_1,X_2)}(x_1',x_2) - F_{(X_1,X_2)}(x_1,x_2') + F_{(X_1,X_2)}(x_1,x_2). \end{aligned}$$

Observación 1.1.1. *La propiedad (FD4), literal (b), establece que, si se conoce la función de distribución conjunta* $F_{(X_1,X_2)}(\cdot,\cdot) : \mathbb{R}^2 \to [0,1]$*, entonces se conoce la función de distribución de cada una de las variables aleatorias* X_1 *y* X_2*; dichas funciones* $F_{X_1}(\cdot)$ *y* $F_{X_2}(\cdot)$ *se conocen como las **distribuciones marginales**.*

Tipos de vectores aleatorios

1. **Discretos:** Un vector aleatorio $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ es discreto si su rango $R_{\mathbf{X}} = \{\mathbf{x}^1, \mathbf{x}^2, \ldots\}$ es un conjunto finito o numerable de puntos de $\mathbb{R}^n$. En este caso, la función $p_X : \mathbb{R}^n \to [0,1]$ definida por,

$$p_{\mathbf{X}}(\mathbf{x}) := \begin{cases} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n), & \text{si } \mathbf{x} \in R_{\mathbf{X}}, \\ 0 \qquad\qquad\qquad\qquad\qquad\quad , & \text{si } \mathbf{x} \notin R_{\mathbf{X}}, \end{cases}$$

se llama **función de masa** del vector aleatorio $\mathbf{X}$. Note que $p_{\mathbf{X}}(\mathbf{x}) \in [0,1]$ y $\sum_{i \geq 1} p_{\mathbf{X}}(\mathbf{x}^i) = 1$. Como en el caso de una dimensión, el conocimiento de la función de masa $p_{\mathbf{X}}(\cdot)$ equivale al conocimiento de la función de distribución. En particular, para un vector aleatorio bidimensional con rango $\{(x_1, y_1), (x_2, y_2), \ldots\}$, se tiene que

$$F_{(X,Y)}(x,y) = \sum_{\{i : x_i \leq x, y_i \leq y\}} p_{(X,Y)}(x_i, y_i).$$

2. **Continuos:** Un vector aleatorio $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ es continuo (absolutamente), si existe una función $f_{(X_1,X_2,\ldots,X_n)} : \mathbb{R}^n \to [0, \infty)$ tal que

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_n} \int_{-\infty}^{x_{n-1}} \cdots \int_{-\infty}^{x_1} f_{(X_1,X_2,\ldots,X_n)}(t_1, t_2, \ldots, t_n) dt_1 dt_2 \cdots dt_n,$$

para todo $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. La función $f_{(X_1,\ldots,X_n)}(\cdot, \ldots, \cdot)$ se llama **función de densidad** del vector aleatorio $\mathbf{X}$. Como en el caso de una variable aleatoria, una función de densidad satisface:

   a) $f_{(X_1,X_2,\ldots,X_n)}(x_1, x_2, \ldots, x_n) \geq 0$ para todo $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

   b) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{(X_1,X_2,\ldots,X_n)}(t_1, t_2, \ldots, t_n) dt_1 dt_2 \cdots dt_n = 1.$

   El conocimiento de la función de densidad $f_{\mathbf{X}}(\cdot)$ equivale al conocimiento de la función de distribución $F_{\mathbf{X}}(\cdot)$. En particular, para un vector aleatorio (X, Y) bidimensional,

   - $F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{(X,Y)}(t_1, t_2) dt_1 dt_2,$
   - $f_{(X,Y)}(x,y) = \dfrac{\partial^2 F_{(X,Y)}(x,y)}{\partial x \partial y} = \dfrac{\partial^2 F_{(X,Y)}(x,y)}{\partial y \partial x},$

   en aquellos puntos (x,y) donde $f_{(X,Y)}(\cdot,\cdot)$ es continua. Esta afirmación también es válida para vectores aleatorios $n-$dimensionales.

Observación 1.1.2. *En este contexto, las funciones de densidad $f_X(\cdot)$ y $f_Y(\cdot)$, llamadas* ***funciones de densidad marginales****, se pueden calcular a partir de la función de densidad conjunta $f_{X,Y}(\cdot,\cdot)$:*

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dy \quad y \quad f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dx.$$

A continuación se introducirán las funciones de densidad y distribución condicionadas de una variable aleatoria continua.

Definición 1.1.8. *Sea (X,Y) un vector aleatorio continuo con función de densidad conjunta $f_{(X,Y)}(\cdot,\cdot)$. Para $y \in \mathbb{R}$ tal que $f_Y(y) > 0$, la* ***función de densidad condicionada*** *de la variable aleatoria X dado $Y = y$, se denota por $f_{X|Y}(\cdot|y) : \mathbb{R} \to [0,\infty)$ y se define por $f_{X|Y}(x|y) := \dfrac{f_{(X,Y)}(x,y)}{f_Y(y)}$.*

En este caso, la **función de distribución condicionada** de la variable aleatoria $(X|Y = y)$ es $F_{X|Y}(\cdot|y) : \mathbb{R} \to [0,1]$ dada por

$$F_{X|Y}(x|y) \equiv P[X \leq x|Y = y] = \int_{-\infty}^{x} f_{X|Y}(t|y)dt.$$

De manera análoga, las funciones de densidad y distribución condicionadas de la variable aleatoria $(Y|X = x)$ son, respectivamente,

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)},$$
$$F_{Y|X}(y|x) = \int_{-\infty}^{y} f_{Y|X}(t|x)dt.$$

Para vectores aleatorios n-dimensionales, con $n \geq 3$, se definen de manera análoga las funciones de densidad condicionales.

En general, el conocimiento de las funciones de distribución $F_{X_1}(\cdot)$ y $F_{X_2}(\cdot)$ no permite dar directamente la función de distribución conjunta $F_{(X_1,X_2)}$, excepto en el caso especial en que las variables aleatorias son independientes.

Definición 1.1.9. *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Dichas variables se dice que son* ***independientes****, si para cada $x_1, x_2, \ldots, x_n \in \mathbb{R}$,*

$$F_{(X_1,X_2,\ldots,X_n)}(x_1,x_2,\ldots,x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n).$$

Una definición alternativa pero equivalente para variables aleatorias independientes es la siguiente:

Definición 1.1.10. *Sea $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ un vector aleatorio continuo. Se dice que las variables aleatorias $X_1, X_2, \ldots, X_n$ son* ***independientes****, si para cada $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$:*

$$f_{(X_1,X_2,\ldots,X_n)}(x_1,x_2,\ldots,x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n).$$

Ejemplo 1.1.2. *Supongamos que el vector aleatorio* $(X_1, X_2, \ldots, X_n)$ *tiene función de densidad dada por*

$$\begin{aligned} f_{(X_1,X_2,\ldots,X_n)}(x_1, x_2, \ldots, x_n) &= \lambda^n e^{-\lambda \sum_{i=1}^n x_i}, \\ &= (\lambda e^{-\lambda x_1})(\lambda e^{-\lambda x_2}) \cdots (\lambda e^{-\lambda x_n}). \end{aligned}$$

Note que cada uno de los factores en el producto anterior corresponde a la función de densidad de una variable aleatoria con distribución exponencial de parámetro λ*. Por lo tanto, las variables aleatorias* $X_1, X_2, \ldots, X_n$ *son independientes.*

Esperanza condicionada

Definición 1.1.11. *Sea* (X, Y) *un vector aleatorio y sea* $g(\cdot, \cdot)$ *una función de dos variables. La* ***esperanza condicionada*** *de la variable aleatoria* $g(X, Y)$ *dado* $X = x$*, denotada por* $E[g(X, Y)|X = x]$*, se define como*

$$E[g(X, Y|X = x] = \begin{cases} \sum_j g(x, y_j) p_{Y|X}(y_j|x), & \text{si } (X, Y) \text{ es discreto,} \\ \int_{-\infty}^{\infty} g(x, y) f_{Y|X}(y|x) dy, & \text{si } (X, Y) \text{ es continuo,} \end{cases}$$

cuando la serie o integral converge absolutamente.

Teorema 1.1.2 (Propiedad de la doble esperanza)**.** *Sea* (X, Y) *un vector aleatorio y* $g(\cdot)$ *una función de una variable. Entonces*

$$E[g(X)] = E[E[g(Y)|X]].$$

En particular,

$$E[Y] = E[E[Y|X]].$$

Definición 1.1.12. *Sea* $h(x) = E[Y|X = x]$*. Entonces la función* $h(x)$ *se denomina* ***curva de regresión*** *de* Y *sobre* x*.*

Para más detalles sobre esperanza condicionada y curva de regresión, se puede consultar la sección 4.3 de Mood et al. [20], o la sección 5.4 de Bain y Engelhardt [2].

1.1.6. Covarianza y coeficiente de correlación

Definición 1.1.13. *Sean* X *y* Y *variables aleatorias tales que* $E[X] = \mu_X$*,* $E(Y) = \mu_Y$*,* $Var(X) = \sigma_X^2$ *y* $Var(Y) = \sigma_Y^2$ *existen.*

1. *La* ***covarianza*** *entre* X *y* Y*, denotada por* $Cov(X, Y)$ *o* $\sigma_{X,Y}$*, se define por*

$$\sigma_{X,Y} \equiv Cov(X, Y) := E[(X - \mu_X)(Y - \mu_Y)].$$

2. *El* ***coeficiente de correlación de Pearson*** *entre X y Y, denotado por $\rho(X,Y)$, se define por*

$$\rho(X,Y) := \frac{Cov(X,Y)}{\sigma_X \sigma_Y},$$

donde $\sigma_X > 0$ y $\sigma_Y > 0$.

Las cantidades definidas anteriormente se utilizan para cuantificar el grado de dependencia entre las variables aleatorias X y Y, como se verá en el corolario 1.

Proposición 1.1.1. *Sean X y Y variables aleatorias. Entonces $Cov(X,Y) = E[XY] - \mu_X \mu_Y$.*

Demostración. Por la definición 1.1.13 se tiene que

$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)], \\ &= E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y], \\ &= E[XY] - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y, \\ &= E[XY] - \mu_X \mu_Y. \end{aligned}$$

□

Observación 1.1.3. *En general, la esperanza del producto de dos variables aleatorias se define como*

$$E[g(X) \cdot h(Y)] := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) h(y) f_{(X,Y)}(x,y) dx dy,$$

donde g y h son funciones y $f_{(X,Y)}(\cdot,\cdot)$ denota la función de densidad del vector aleatorio (X,Y).

Teorema 1.1.3 (Desigualdad de Cauchy-Schwarz)**.** *Sean X y Y variables aleatorias. Entonces $E^2[XY] \leq E[X^2]E[Y^2]$ o, equivalentemente, $|E[XY]| \leq (E[X^2]E[Y^2])^{1/2}$. Además, la igualdad se da si y solo si $P(Y = cX) = 1$, donde $c = \dfrac{E[XY]}{E[X^2]}$.*

Demostración. Para $t \in \mathbb{R}$, se define la siguiente función:

$$\begin{aligned} 0 \leq f(t) &:= E[(tX - Y)^2], \\ &= E[t^2 X^2 - 2tXY + Y^2], \\ &= E[X^2]t^2 - 2E[XY]t + E[Y^2]. \end{aligned}$$

La función cuadrática $w = f(t)$ toma valores no negativos. Si $f(t) > 0$ para todo t, entonces $4E^2[XY] - 4E[X^2]E[Y^2] < 0$, es decir, $E^2[XY] < E[X^2]E[Y^2]$. Si $f(t)$ se anula en un punto $t = c$, entonces $0 = f(c) = E[(cX - Y)^2]$; por lo tanto $P[Y = cX] = 1$ y $c = \dfrac{E[XY]}{E[X^2]}$. □

Corolario 1. *Sean X y Y variables aleatorias. Entonces $|\rho(X,Y)| \leq 1$. Además, la igualdad se da si y solo si $\left(\frac{Y-\mu_Y}{\sigma_Y}\right) = \pm\left(\frac{X-\mu_X}{\sigma_X}\right)$ con probabilidad 1.*

Demostración. Aplicando el Teorema 1.1.3 a las variables aleatorias $\tilde{X} := \frac{X-\mu_X}{\sigma_X}$ y $\tilde{Y} := \frac{Y-\mu_Y}{\sigma_Y}$ se obtiene:

1. $|\rho(X,Y)| = \left|\frac{Cov(X,Y)}{\sigma_X\sigma_Y}\right| = |E[\tilde{X}\tilde{Y}]| < (E[\tilde{X}^2]E[\tilde{Y}^2])^{1/2} = 1.$

2. La igualdad en el caso anterior se da si y solo si $P(\tilde{Y} = c\tilde{X}) = 1$, con $c = \frac{E[\tilde{X}\tilde{Y}]}{E[\tilde{X}^2]}$.
   Esto es, $\rho(X,Y) = \pm 1$ si y solo si $\tilde{Y} = c\tilde{X}$ o bien $\left(\frac{Y-\mu_Y}{\sigma_Y}\right) = \pm\left(\frac{X-\mu_X}{\sigma_X}\right)$.

□

1.1.7. Muestras aleatorias

Definición 1.1.14. *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes con función de densidad común $f(\cdot)$. Entonces se dice que $X_1, X_2, \ldots, X_n$ es una **muestra aleatoria** de tamaño n para una población con densidad $f(\cdot)$. En otras palabras, $X_1, X_2, \ldots, X_n$ es una muestra aleatoria si estas son variables aleatorias independientes e idénticamente distribuidas (i.i.d.) con función de densidad $f(\cdot)$.*

Cuando se toma una muestra aleatoria, se pueden obtener otras cantidades asociadas a ella, realizando algunas operaciones con sus elementos. Dichas cantidades pueden expresar información contenida en la muestra de una manera más conveniente o más clara. Matemáticamente, este procedimiento se representa mediante una función $T(X_1, X_2, \ldots, X_n)$, cuyo dominio contiene al conjunto Ω sobre el que están definidas $X_1, X_2, \ldots, X_n$. La función T, que puede ser real o vectorial, define una nueva variable aleatoria $Y = T(X_1, X_2, \ldots, X_n)$. A menudo resulta conveniente conocer la distribución de Y. En la siguiente definición se establece un nombre específico para la función T y su respectiva distribución.

Definición 1.1.15. *Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n y sea $T(X_1, X_2, \ldots, X_n)$ una función real o vectorial cuyo dominio contiene a Ω, el conjunto sobre el que están definidas $X_1, X_2, \ldots, X_n$. Entonces la variable o vector aleatorio $Y = T(X_1, X_2, \ldots, X_n)$ se denomina un **estadístico**. La distribución de probabilidad del estadístico Y se denomina la **distribución muestral** de Y.*

Cuando se toma una muestra aleatoria, se supone que no se conoce la función de distribución de la población. Sin embargo, es posible estimar dicha función utilizando la muestra aleatoria; para ello se tiene la siguiente herramienta.

Definición 1.1.16. *Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con función de distribución $F_X(x)$. La* ***función de distribución empírica****, denotada por $F_n(x)$, se define como:*

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

donde $\mathbb{1}_A(x)$ representa la función indicadora definida como

$$\mathbb{1}_A(x) = \begin{cases} 1, & si\ x \in A, \\ 0, & si\ x \notin A. \end{cases}$$

Convergencia en probabilidad y en distribución

Con frecuencia existe interés en estudiar el comportamiento de algunos estadísticos cuando el tamaño de la muestra es grande ($n \to \infty$). Para ello se establecen dos tipos particulares de convergencia.

Definición 1.1.17 (Convergencia en probabilidad)**.** *Una sucesión de variables aleatorias $X_1, X_2, \ldots$* ***converge en probabilidad*** *a una variable aleatoria X si, para todo $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0 \quad o,\ equivalentemente, \quad \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1.$$

.

Definición 1.1.18 (Convergencia en distribución)**.** *Una sucesión de variables aleatorias $X_1, X_2, \ldots$* ***converge en distribución*** *a una variable aleatoria X si $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ para todos los valores de x en los que $F_X(x)$ sea continua.*

El siguiente teorema establece que la convergencia en probabilidad implica convergencia en distribución.

Teorema 1.1.4. *Si la sucesión de variables aleatorias $X_1, X_2, \ldots$ converge en probabilidad a una variable aleatoria X, entonces la sucesión también converge en distribución a X.*

Demostración. Sea $(\Omega, \mathcal{F}, P)$ el espacio de probabilidad donde están definidas las variables aleatorias X y X_n. Sean $t, \epsilon \in \mathbb{R}$, con $\epsilon > 0$. Supongamos que $X_n > t$ y $|X_n - X| < \epsilon$, entonces $X > t - \epsilon$. Así, si $X \leq t - \epsilon$ se tiene que $X_n \leq t$ o $|X_n - X| \geq \epsilon$, esto es,

$$\{\omega \mid X \leq t - \epsilon\} \subseteq \{\omega \mid X_n \leq t\} \cup \{\omega \mid |X_n - X| \geq \epsilon\}, \quad \omega \in \Omega.$$

Por lo tanto, $P(X \leq t - \epsilon) \leq P(X_n \leq t) + P(|X_n - X| \geq \epsilon)$, de donde,

$$P(X_n \leq t) \geq P(X \leq t-\epsilon) - P(|X_n - X| \geq \epsilon). \tag{1.2}$$

Análogamente, se deduce que

$$P(X_n \leq t) \leq P(X \leq t+\epsilon) + P(|X_n - X| \geq \epsilon). \tag{1.3}$$

De las desigualdades (1.2) y (1.3) se sigue que

$$P(X \leq t-\epsilon) - P(|X_n - X| \geq \epsilon) \leq P(X_n \leq t) \leq P(X \leq t+\epsilon) + P(|X_n - X| \geq \epsilon). \tag{1.4}$$

Ahora, supongamos que $F_X(\cdot)$ es continua en t. Por hipótesis, $X_n \to X$ en probabilidad, esto es, $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$. Aplicando límite a todos los miembros de la desigualdad (1.4) obtenemos

$$P(X \leq t-\epsilon) - 0 \leq \lim_{n\to\infty} P(X_n \leq t) \leq P(X \leq t+\epsilon) + 0.$$

Tomando $\epsilon \to 0$ se deduce que $\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$. □

El siguiente ejemplo muestra que el recíproco del anterior teorema no siempre es válido.

Ejemplo 1.1.3. *Sean X y Y variables aleatorias independientes con distribución Bernoulli, $p = 1/2$; esto es,*

$$p_{(X,Y)}(x,y) = P[X = x]P[Y = y] = \frac{1}{4}\mathbb{1}_R(x,y),$$

donde $R = \{(0,0),(0,1),(1,0),(1,1)\}$. Si $X_n = X$ para todo $n \in \mathbb{N}$, entonces $X_n \xrightarrow{d} Y$ y para $0 < \epsilon < 1$ se tiene $P[|X_n - Y| \geq \epsilon] = P[|X - Y| \geq \epsilon] = P[|X - Y| = 1] = 1/2$. Es decir, $P[|X_n - Y| \geq \epsilon] \nrightarrow 0$.

Sin embargo, en la siguiente observación se señala un caso en el cual sí es válido el recíproco del Teorema 1.1.4.

Observación 1.1.4. *Si $X_n \xrightarrow{d} c$, entonces $X_n \xrightarrow{P} c$, para cualquier $c \in \mathbb{R}$. En efecto,*

$$\begin{aligned} P[|X_n - c| \geq \epsilon] &\leq P[X_n \leq c-\epsilon] + 1 - P[X_n \leq c+\epsilon], \\ &= F_{X_n}(c-\epsilon) + 1 - F_{X_n}(c+\epsilon) \to 0, \quad \textit{cuando } n \to \infty. \end{aligned}$$

Finalmente, se presenta uno de los teoremas más relevantes en probabilidad y estadística.

Teorema 1.1.5 (Teorema del Límite Central)**.** *Si para cada entero positivo n, las variables aleatorias $X_1, X_2, \ldots, X_n$ son independientes e idénticamente distribuidas con media común μ_X y varianza común σ_X^2, entonces para cada $z \in \mathbb{R}$:*

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z),$$

donde $Z_n := \dfrac{\overline{X_n} - E[\overline{X_n}]}{\sqrt{Var(\overline{X_n})}} \equiv \dfrac{\overline{X_n} - \mu_X}{\sigma_X/\sqrt{n}}$ *y $\Phi(\cdot)$ es la función de distribución normal estándar.*

Este teorema indica que para muestras grandes, el estadístico Z_n converge en distribución a una normal estándar, $N(0,1)$. Es de gran utilidad para obtener una distribución muestral aproximada en los casos en que la distribución exacta es desconocida o muy difícil de obtener. Una demostración de este teorema puede consultarse en la página 239 de Bain y Engelhardt [2].

1.2. Conceptos básicos de Estadística

El contenido de esta sección tiene como principales referencias los capítulos 9 y 12 de Bain y Engelhardt [2] y los capítulos 7 y 8 de Casella y Berger [3].

1.2.1. Estimación puntual

Definición 1.2.1. *Un **estimador puntual** es cualquier función $T(X_1, X_2, \ldots, X_n)$ de una muestra aleatoria; es decir, cualquier estadístico es un estimador puntual.*

La definición anterior es bastante general y permite la existencia de múltiples estimadores para un mismo parámetro. Sin embargo, no todos los estimadores son igualmente útiles ni proporcionan buenas aproximaciones al valor verdadero del parámetro de interés. Por ello, es fundamental contar con criterios para evaluar su calidad y así obtener estimadores adecuados. En particular, presentaremos dos enfoques ampliamente utilizados: el **método de momentos** y el **método de máxima verosimilitud**.

- **Método de momentos:**
  Supongamos que $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una población con función de masa o función de densidad $f(x; \theta_1, \ldots, \theta_k)$. El método consiste en igualar los primeros k momentos muestrales con los correspondientes k momentos poblacionales, y resolver el sistema de ecuaciones resultante. En primer lugar, definimos los momentos muestrales, denotados por m_j, y los momentos poblacionales, denotados por μ'_j, $j = 1, \ldots, k$, de la siguiente manera:

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i^1, \qquad \mu'_1 = E[X^1],$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \qquad \mu_2' = E[X^2],$$

$$\vdots$$

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \qquad \mu_k' = E[X^k].$$

Los momentos poblaciones usualmente son funciones de los parámetros $\theta_1, \ldots, \theta_k$, es decir $\mu_j'(\theta_1, \ldots, \theta_k)$. Así, el estimador $(\tilde{\theta_1}, \ldots, \tilde{\theta_k})$ se obtiene resolviendo el siguiente sistema de ecuaciones para $(\theta_1, \ldots, \theta_k)$ en términos de $(m_1, \ldots, m_k)$:

$$m_1 = \mu_1'(\theta_1, \ldots, \theta_k),$$

$$m_2 = \mu_2'(\theta_1, \ldots, \theta_k),$$

$$\vdots$$

$$m_k = \mu_k'(\theta_1, \ldots, \theta_k).$$

- **Método de máxima verosimilitud:**
  Supongamos que $X_1, \ldots, X_n$ es una muestra aleatoria i.i.d. de una población con función de densidad o función de masa $f_X(x;\theta)$, $\theta \in \Omega$. Entonces la función de densidad (o masa) conjunta del vector aleatorio $(X_1, \ldots, X_n)$ es

$$f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n; \theta) = f_X(x_1;\theta) \cdots f_X(x_n;\theta) = \prod_{i=1}^{n} f_X(x_i;\theta).$$

Ahora, supongamos que $(x_1, \ldots, x_n)$ son valores observados de la muestra aleatoria $(X_1, \ldots, X_n)$. La función de verosimilitud, denotada por $L(\theta)$, se define como

$$L(\theta) \equiv L(\theta; x_1, \ldots, x_n) := f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_X(x_i;\theta).$$

Esta función describe cómo varía la probabilidad de observar los datos $(x_1, \ldots, x_n)$ en función del parámetro θ. Para estimarlo, se busca el valor $\hat{\theta}$ que maximiza la verosimilitud, es decir, aquel bajo el cual es más probable observar los datos $(x_1, \ldots, x_n)$. Más concretamente,

Definición 1.2.2. *Sea $x_1, \ldots, x_n$ un conjunto de observaciones y $\hat{\theta}$ el valor donde $L(\theta; x_1, \ldots, x_n)$ alcanza su máximo como función de θ, es decir,*

$$f(x_1, \ldots, x_n; \hat{\theta}) = \max_{\theta \in \Omega} f(x_1, \ldots, x_n; \theta).$$

Entonces $\hat{\theta}$ se denomina un ***estimador de máxima verosimilitud*** *(MLE, por sus siglas en inglés) del parámetro θ.*

Si la función de verosimilitud es diferenciable en θ, los posibles candidatos para el MLE son los valores de θ que resuelven la ecuación

$$\frac{d}{d\theta}L(\theta; x_1, \ldots, x_n) = 0.$$

Para efectos de simplicidad en los cálculos, suele preferirse trabajar con el logaritmo natural de la función de verosimilitud, $\ln L(\theta; x_1, \ldots, x_n)$, en lugar de la verosimilitud misma. Esta transformación da lugar a la función de log-verosimilitud, cuya maximización es equivalente a la de la verosimilitud, ya que el logaritmo es una función estrictamente creciente en $(0, \infty)$. En consecuencia, los puntos donde $L(\theta; x_1, \ldots, x_n)$ alcanza sus extremos coinciden con los de $\ln L(\theta; x_1, \ldots, x_n)$.

Hemos visto que se puede disponer de diversos estimadores para un parámetro. En tal sentido es natural preguntarse cuál de ellos resulta más adecuado. Por esta razón existen diversos criterios que permiten evaluar la calidad de un estimador, algunos de los cuales se presentarán en las definiciones posteriores.

Definición 1.2.3. *El **error cuadrático medio** (MSE, por sus siglas en inglés) de un estimador T de un parámetro θ, denotado por $MSE(T)$, es la función de θ definida por $MSE(T) := E[(T-\theta)^2]$.*

Definición 1.2.4. *El **sesgo** de un estimador puntual T de un parámetro θ, denotado por $b(T)$, es la diferencia entre la esperanza de T y θ, esto es, $b(T) := E[T] - \theta$. Un estimador cuyo sesgo sea exactamente cero se llama un estimador **insesgado** y satisface que $E[T] = \theta$ para todo θ.*

El siguiente teorema presenta una forma importante de interpretar el error cuadrático medio.

Teorema 1.2.1. *Si T es un estimador de θ, entonces*

$$MSE(T) = Var(T) + (b(T))^2.$$

Demostración.

$$\begin{aligned}
MSE(T) &= E[(T-\theta)^2], \\
&= E[(T - E[T] + E[T] - \theta)^2], \\
&= E[(T - E[T])^2 + 2(T - E[T])(E[T] - \theta) + (E[T] - \theta)^2], \\
&= E[(T - E[T])^2] + 2E[(T - E[T])(E[T] - \theta)] + E[(E[T] - \theta)^2], \\
&= Var(T) + 2(E[T] - \theta)(E[T] - E[T]) + (E[T] - \theta)^2, \\
&= Var(T) + (b(T))^2.
\end{aligned}$$

□

Este teorema muestra que el error cuadrático medio relaciona tanto la varianza como el sesgo del estimador. Así, para un estimador insesgado, su error cuadrático medio coincide con su varianza; esto implica que un estimador será preferido desde el punto de vista del error cuadrático medio si es insesgado pero además su varianza es lo más pequeña posible. Esto permite dar la siguiente definición.

Definición 1.2.5. *Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n. Un estimador T^* del parámetro θ se denomina* ***estimador insesgado de mínima varianza*** *(UMVUE, por sus siglas en inglés) de θ si,*

1. *T^* es insesgado para θ.*

2. *Dado cualquier otro estimador insesgado T de θ, $Var(T^*) \leq Var(T)$ para todo θ.*

Para saber si un estimador es el de mínima varianza entre todos los posibles estimadores insesgados del parámetro θ, se dispone del siguiente teorema que establece una cota inferior para dicha varianza.

Teorema 1.2.2 (Cota inferior de Cramér-Rao)**.** *Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria con función de densidad $f(x;\theta)$, y sea $T(\mathbf{X}) = T(X_1, X_2, \ldots, X_n)$ cualquier estimador que satisfaga las siguientes propiedades:*

1. $\dfrac{d}{d\theta}E[T(\mathbf{X})] = \displaystyle\int_\Omega \frac{\partial}{\partial\theta}(T(\mathbf{x})f(x;\theta))d\mathbf{x}$.

2. $Var(T(\mathbf{X})) < \infty$.

Entonces $Var(T(\mathbf{X})) \geq \dfrac{\left(\frac{d}{d\theta}E[T(\mathbf{X})]\right)^2}{nE\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right]}$.

Demostración. Sea $f(x_1, \ldots, x_n;\theta)$ la función de densidad del vector aleatorio $\mathbf{X} = (X_1, \ldots, X_n)$. Definamos la función

$$\begin{aligned} u(x_1, \ldots, x_n;\theta) &:= \frac{\partial}{\partial\theta}\ln f(x_1, \ldots, x_n;\theta), \\ &= \frac{1}{f(x_1, \ldots, x_n;\theta)}\frac{\partial}{\partial\theta}f(x_1, \ldots, x_n;\theta). \end{aligned}$$

Ahora, definamos la variable aleatoria $U = u(X_1, \ldots, X_n;\theta)$; entonces,

$$\begin{aligned} E[U] &= \int \cdots \int u(x_1,\ldots,x_n;\theta) f(x_1,\ldots,x_n;\theta) dx_1 \cdots dx_n, \\ &= \int \cdots \int \frac{\partial}{\partial\theta} f(x_1,\ldots,x_n;\theta) dx_1 \cdots dx_n, \\ &= \frac{d}{d\theta} \int \cdots \int f(x_1,\ldots,x_n;\theta) dx_1 \cdots dx_n, \\ &= \frac{d}{d\theta} 1, \\ &= 0. \end{aligned}$$

De otro lado, note que

$$\begin{aligned} E[T] &= \int \cdots \int T(x_1,\ldots,x_n;\theta) f(x_1,\ldots,x_n;\theta) dx_1 \cdots dx_n, \\ \frac{d}{d\theta}E[T] &= \int \cdots \int T(x_1,\ldots,x_n;\theta) \frac{\partial}{\partial\theta} f(x_1,\ldots,x_n;\theta) dx_1 \cdots dx_n, \\ &= \int \cdots \int T(x_1,\ldots,x_n) u(x_1,\ldots,x_n;\theta) f(x_1,\ldots,x_n;\theta) dx_1 \cdots dx_n, \\ &= E[TU]. \end{aligned}$$

En resumen, se ha mostrado que

$$Var(U) = E[U^2], \qquad E[TU] = \frac{d}{d\theta}E[T]. \tag{1.5}$$

Del corolario 1, se sabe que $|\rho(X,Y)| \leq 1$; empleando la proposición 1.1.1 se obtiene

$$Cov(X,Y)^2 \leq [Var(X)][Var(Y)]. \tag{1.6}$$

Como $E[U] = 0$ entonces $Cov(T,U) = E[TU]$. Usando este hecho, junto con las ecuaciones (1.5) y (1.6) se sigue que

$$Var(T) \geq \frac{E^2[TU]}{E[U^2]} = \frac{\left(\frac{d}{d\theta}E[T]\right)^2}{E\left[\left(\frac{\partial}{\partial\theta}\ln f(x_1,\ldots,x_n;\theta)\right)^2\right]}.$$

Como en este caso $X_1,\ldots,X_n$ son i.i.d., entonces

$$f(x_1,\ldots,x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

Por tanto,

$$u(x_1,\ldots,x_n;\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} f(x_i;\theta).$$

Así las cosas,

$$E[U^2] = Var(U) = nVar\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right) = nE\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right],$$

de donde se sigue el teorema.

□

Note que, de acuerdo a la demostración anterior, la cota inferior de Cramér-Rao se alcanza cuando $\rho(T,U) = \pm 1$, es decir, cuando existe una relación lineal entre las variables T y U. En otras palabras, cuando $T = aU + b$, para algunas constantes $a \neq 0$ y b. Por lo tanto, el Teorema 1.2.2 garantiza la existencia de un UMVUE T cuando la variable aleatoria $U = \sum_{i=1}^{n} \frac{\partial}{\partial\theta}\ln f(X_i;\theta)$ existe y T es función lineal de ella. Ilustramos las ideas anteriores mediante el siguiente ejemplo.

Ejemplo 1.2.1. *Consideremos una muestra aleatoria de una distribución exponencial, $X_i \sim Exp(\theta)$, $i = 1,\ldots,n$. Recordemos que $f_{X_i}(x_i;\theta) = \frac{1}{\theta}e^{-x_i/\theta}$, $x_i > 0$, $\theta > 0$, entonces*

$$\begin{aligned}\ln f_{X_i}(x_i;\theta) &= -\frac{x_i}{\theta} - \ln\theta,\\ \frac{\partial}{\partial\theta}\ln f_{X_i}(x_i;\theta) &= \frac{x_i}{\theta^2} - \frac{1}{\theta},\\ &= \frac{x_i-\theta}{\theta^2}.\end{aligned}$$

Consideremos como estimador $T = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Note que

$$E[\overline{X}] = \frac{1}{n}E\left[\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n}\theta = \theta.$$

Así, $\overline{X}$ es un estimador insesgado para θ. Ahora, verifiquemos si en este caso $\overline{X}$ alcanza la cota inferior de Cramér-Rao. Para ello, observe que la variable aleatoria U está dada por

$$U = \sum_{i=1}^{n}\left(\frac{x_i}{\theta^2} - \frac{1}{\theta}\right) = \frac{1}{\theta^2}\sum_{i=1}^{n} x_i - \frac{n}{\theta}.$$

Luego, deben existir constantes $a \neq 0$ y b tales que

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = a\left[\frac{1}{\theta^2}\sum_{i=1}^{n} x_i - \frac{n}{\theta}\right] + b.$$

En efecto, si $a = \dfrac{\theta^2}{n}$ *y* $b = \theta$ *la igualdad anterior se verifica.*
Por lo tanto, el estimador $\overline{X}$ *alcanza la cota inferior de Cramér-Rao; esto significa que es el de mínima varianza entre los estimadores insesgados, lo que implica que es un UMVUE para* θ*.*

Otro criterio que se tiene para evaluar un estimador es su eficiencia, que se establece a continuación (sección 9.3 de Bain y Engelhardt [2]).

Definición 1.2.6. *La* ***eficiencia relativa*** *de un estimador insesgado* T *de* θ *con respecto a otro estimador insesgado* T^* *de* θ *está dada por*

$$re(T, T^*) := \frac{Var(T^*)}{Var(T)}.$$

Un estimador insesgado T^* *de* θ *se denomina* ***eficiente*** *si* $re(T, T^*) \leq 1$*, para todos los estimadores insesgados* T *de* θ *y todo* $\theta \in \Omega$*. La* ***eficiencia*** *de un estimador insesgado* T *de* θ *está dada por* $e(T) := re(T, T^*)$*, si* T^* *es un estimador eficiente de* θ*.*

Hasta el momento se han presentado criterios para evaluar un estimador basado en una muestra aleatoria de tamaño fijo n. Sin embargo, resulta relevante considerar las propiedades asintóticas de ciertos estimadores, ya que estos podrían no mostrar buenos resultados para muestras pequeñas, pero sí hacerlo cuando el tamaño de muestra se incrementa. En este sentido se introduce el concepto de estimador consistente.

Definición 1.2.7. *Sea* $\{T_n\}$ *una sucesión de estimadores de* θ*. Estos estimadores se denominan* ***consistentes*** *si para todo* $\epsilon > 0$*,*

$$\lim_{n\to\infty} P(|T_n - \theta| < \epsilon) = 1,$$

para todo $\theta \in \Omega$*.*

1.2.2. Pruebas de hipótesis

En la sección anterior se presentó uno de los métodos de la inferencia estadística conocido como estimación puntual. En esta sección se presentará otro método: las pruebas de hipótesis. El objetivo de una prueba de hipótesis es tomar una decisión, con base en datos experimentales, sobre si existe suficiente evidencia para rechazar una hipótesis nula en favor de una hipótesis alternativa.

Definición 1.2.8. *Una* ***hipótesis estadística*** *es una afirmación acerca de un parámetro poblacional. Si la hipótesis especifica completamente el parámetro, se dice que es una* ***hipótesis simple****, de lo contrario se dice que es una* ***hipótesis compuesta****. Las dos hipótesis complementarias de una prueba de hipótesis se denominan* ***hipótesis nula*** *e* ***hipótesis alternativa****, denotadas como* H_0 *y* H_1 *respectivamente.*

Si θ denota un parámetro poblacional, la forma usual en la que se presentan las hipótesis es la siguiente:

$$\begin{cases} H_0 : \ \theta = k, \\ H_1 : \ \theta \neq k, \end{cases} \qquad \text{o} \qquad \begin{cases} H_0 : \ \theta \in \Omega_0, \\ H_1 : \ \theta \in \Omega_0^c, \end{cases}$$

donde Ω_0 es algún subconjunto del espacio parámetrico Ω, mientras que k es algún valor fijo en Ω. En el primer caso, H_0 es simple y H_1 es compuesta; en el segundo caso, ambas hipótesis son compuestas.

Definición 1.2.9. *Una **prueba de hipótesis** es una regla para decidir si se acepta o se rechaza H_0. El subconjunto del espacio muestral para el cual H_0 se rechaza se denomina **región crítica**.*

Generalmente, una prueba de hipótesis se especifica en términos de un estadístico $T(X_1, \ldots, X_n)$, denominado **estadístico de prueba**, que se utiliza para construir la región crítica. Por ejemplo, si se toma $T(X_1, \ldots, X_n) = \overline{X}$, se puede definir la región crítica como $C = \{(x_1, \ldots, x_n) \mid \overline{x} \geq c\}$ para algún c por determinar. Es decir, si $\overline{x} \geq c$, se rechaza H_0.

Cuando se realiza una prueba de hipótesis, existe la posibilidad de cometer dos tipos de errores, denominados error tipo I y error tipo II, a saber:

- **Error tipo I:** Rechazar H_0 cuando esta es verdadera,
- **Error tipo II:** No rechazar H_0 cuando esta es falsa.

Se denota la probabilidad de cometer estos errores de la siguiente forma:

- $P(TI) = P(\text{rechazar } H_0 \mid H_0) = \alpha$,
- $P(TII) = P(\text{aceptar } H_0 \mid H_1) = \beta$.

Definición 1.2.10. *Si la hipótesis nula H_0 es simple, se dice que $\alpha = P(TI)$ es el **nivel de significancia** de la prueba, y si la hipótesis nula H_0 es compuesta, se dice que $\alpha^* := \underset{\theta \in \Omega_0}{\text{máx}} P(\text{rechazar } H_0 \mid H_0)$ es el **tamaño de la prueba**.*

En otras palabras, el tamaño de la prueba α^* es la máxima probabilidad de rechazar H_0 cuando esta es verdadera, maximizando sobre los valores del parámetro bajo H_0. Note que si H_0 es simple, el tamaño de la prueba coincide con el nivel de significancia.

Definición 1.2.11. *La **función potencia** de una prueba, denotada por $\pi(\theta)$, es la probabilidad de rechazar H_0 cuando el verdadero valor del parámetro es θ.*

Para una hipótesis nula simple, digamos $H_0 : \theta = \theta_0$, se tiene que

$$\pi(\theta_0) = P(\text{rechazar } H_0 \mid \theta = \theta_0) = P(\text{rechazar } H_0 \mid H_0) = P(TI) = \alpha.$$

De igual manera, si $H_1 : \theta = \theta_1$, entonces

$$\pi(\theta_1) = P(\text{rechazar } H_0 \mid \theta = \theta_1) = 1 - P(TII) = 1 - \beta.$$

Para el caso de hipótesis compuestas, digamos $H_0 : \theta \in \Omega_0$ contra $H_1 : \theta \in \Omega_0^c$, se tiene que $\alpha = \max_{\theta \in \Omega_0} \pi(\theta)$ y $\pi(\theta) = 1 - \beta$, que depende de θ y nos da la potencia de la prueba. Idealmente, se busca que $\pi(\theta)$ sea siempre cercano a cero.

Definición 1.2.12. *Un valor p, $p(\mathbf{X})$, es un estadístico de prueba que satisface que $0 \leq p(\mathbf{x}) \leq 1$ para toda muestra aleatoria observada $\mathbf{x}$. Valores pequeños de $p(\mathbf{X})$ aportan evidencia de que H_1 es verdadera. Un valor p es válido si, para todo $\theta \in \Omega$ y todo $0 \leq \alpha \leq 1$,*

$$P(p(\mathbf{X}) \leq \alpha \mid \theta) \leq \alpha.$$

En otras palabras, el valor p de una prueba es el valor más pequeño que puede tomar α para rechazar H_0, basado en el valor observado del estadístico de prueba.

1.2.3. Pruebas de bondad de ajuste

En estadística, las pruebas de bondad de ajuste se utilizan para evaluar si un conjunto de datos es coherente con un modelo o una familia de distribuciones de probabilidad específica. Esto es útil cuando no se conoce con certeza el modelo subyacente de la muestra aleatoria y se desea validar si un modelo propuesto puede describir adecuadamente los datos. El procedimiento implica formular una prueba de hipótesis estadística que permita decidir si los datos son consistentes con el modelo considerado. Por ejemplo, se podría tener interés en verificar si un conjunto de observaciones $(x_1, \ldots, x_n)$ podría provenir razonablemente de una distribución normal.

Supongamos que $F_n(\cdot)$ es la función de distribución empírica (Definición 1.1.16), y se desea probar la hipótesis nula simple $F_X(\cdot) = F_X^0(\cdot)$, donde $F_X^0(\cdot)$ es alguna función de distribución específica. Un punto de partida razonable es considerar como estadístico de prueba alguna medida de distancia entre $F_n(\cdot)$ y $F_X^0(\cdot)$. En particular, si d es cualquier métrica en el espacio de las funciones de distribución, entonces $d(F_n(\cdot), F_X^0(\cdot))$ podría servir como estadístico de prueba. Una elección usual para d es

$$d_K(F_X(\cdot), F_Y(\cdot)) := \sup_t |F_X(t) - F_Y(t)|,$$

donde $F_X(\cdot)$ y $F_Y(\cdot)$ son funciones de distribución. La función $d_K(\cdot, \cdot)$ se conoce como **distancia de Kolmogorov-Smirnov**.

De acuerdo a lo anterior, se plantea una de varias pruebas de bondad de ajuste reportadas en la literatura, que se conoce como **Prueba de Kolmogorov-Smirnov** (para más detalles, ver el capítulo 14 de Lehmann y Romano [19]).

En primer lugar, supongamos que $X_1, \ldots, X_n$ es una muestra alatoria de una función de distribución $F_X(\cdot)$ continua. Supongamos que se desea probar la hipótesis nula $F_X(\cdot) = F_X^0(\cdot)$ contra la alternativa $F_X(\cdot) \neq F_X^0(\cdot)$. Para ello, se toma como estadístico de prueba el siguiente:

$$T_n := \sup_{t \in \mathbb{R}} \sqrt{n} |F_n(t) - F_X^0(t)| = \sqrt{n} d_K(F_n(\cdot), F_X^0(\cdot)).$$

La distribución de T_n bajo $F_X^0(\cdot)$ es la misma para todas las funciones de distribución continuas $F_X^0(\cdot)$.

Sea $s_{n,1-\alpha}$ el percentil $1-\alpha$ de la distribución de T_n bajo cualquier función de distribución continua $F_X^0(\cdot)$. Entonces la prueba de Kolmogorov-Smirnov rechaza la hipótesis nula si $T_n > s_{n,1-\alpha}$.

Capítulo 2

Funciones cópula bivariadas

En este capítulo se dará la definición de función cópula bivariada y se enunciarán algunas de sus propiedades y teoremas esenciales, tal como el célebre Teorema de Sklar. Asimismo se presentarán varios ejemplos de familias de cópulas comúnmente utilizadas. Además, se introducirá el concepto de *medidas de asociación* y se presentarán algunas de ellas, en particular el τ de Kendall y el ρ de Spearman. El contenido de este capítulo se fundamenta principalmente en Nelsen [21].

2.1. Definición y propiedades básicas

En adelante $I = [0,1] \subseteq \mathbb{R}$, $I^2 = [0,1] \times [0,1]$, y para una función $G : \mathcal{D} \subseteq \mathbb{R}^n \to \mathbb{R}$, $a, b \in \mathbb{R}$, se define el operador de diferencias en la i-ésima componente como

$$\Delta_{ab}^i G(x_1, \ldots, x_n) := G(x_1, \ldots, x_{i-1}, b, x_{i+1}, \ldots, x_n) - G(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n).$$

Definición 2.1.1. *Una función $C : I^2 \to I$ es una **cópula** si*

(C1) $C(u,v) = uv$ para todo $(u,v) \in \partial I^2$ (la frontera del cuadrado unitario); es decir, para $u, v \in I$ se tiene que

$$C(0,v) = 0 = C(u,0) \qquad y \qquad C(1,v) = v, \quad C(u,1) = u.$$

(C2) Para $0 \leq u_1 \leq u_2 \leq 1$ y $0 \leq v_1 \leq v_2 \leq 1$ se cumple que $\Delta_{u_1 u_2}^1 \Delta_{v_1 v_2}^2 C(x,y) \geq 0$; es decir,

$$\Delta_{u_1 u_2}^1 \Delta_{v_1 v_2}^2 C(x,y) = C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

El siguiente Teorema presenta algunas de las propiedades más importantes de las funciones cópula.

Teorema 2.1.1. *Sea $C : I^2 \to I$ una función cópula. Para $u, v \in I$,*

1. $C(u, v) = \Delta^1_{0u}\Delta^2_{0v}C(x, y)$.
2. $C(\cdot, \cdot)$ *es creciente en cada variable.*
3. *Para $0 \leq u_1 \leq u_2 \leq 1$ y $0 \leq v_1 \leq v_2 \leq 1$:*

$$0 \leq C(u, v_2) - C(u, v_1) \leq v_2 - v_1,$$

$$0 \leq C(u_2, v) - C(u_1, v) \leq u_2 - u_2.$$

4. *Para $(x_1, x_2), (y_1, y_2) \in I^2$:*

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

5. *Para $v \in I$, la derivada parcial $\dfrac{\partial C(u, v)}{\partial u}$ existe para casi todo $u \in I$ y*

$$0 \leq \frac{\partial C(u, v)}{\partial u} \leq 1.$$

Similarmente, para $u \in I$, la derivada parcial $\dfrac{\partial C(u, v)}{\partial v}$ existe para casi todo $v \in I$ y

$$0 \leq \frac{\partial C(u, v)}{\partial v} \leq 1.$$

Demostración. 1.

$$\begin{aligned} \Delta^1_{0u}\Delta^2_{0v}C(x, y) &= \Delta^1_{0u}[C(x, v) - C(x, 0)], \\ &= \Delta^1_{0u}C(x, v), \\ &= C(u, v) - C(0, v), \\ &= C(u, v). \end{aligned}$$

2. Note que

$$\begin{aligned} \Delta^1_{0u}\Delta^2_{v_1v_2}C(x, y) &= \Delta^1_{0u}[C(x, v_2) - C(x, v_1)], \\ &= C(u, v_2) - C(0, v_2) - C(u, v_1) + C(0, v_1), \\ &= C(u, v_2) - C(u, v_1) \geq 0. \end{aligned}$$

Por lo tanto $C(u, v_1) \leq C(u, v_2)$. Similarmente,

$$\Delta^1_{u_1u_2}\Delta^2_{0,v}C(x, y) = C(u_2, v) - C(u_1, v) \geq 0,$$

lo cual implica que $C(u_1, v) \leq C(u_2, v)$.

3. Note que

$$\begin{aligned}\Delta^1_{u1}\Delta^2_{v_1v_2}C(x,y) &= \Delta^1_{u1}[C(x,v_2)-C(x,v_1)],\\ &= C(1,v_2)-C(u,v_2)-C(1,v_1)+C(u,v_1),\\ &= v_2-C(u,v_2)-v_1+C(u,v_1)\geq 0.\end{aligned}$$

Por lo tanto $0\leq C(u,v_2)-C(u,v_1)\leq v_2-v_1$. Similarmente se concluye que $0\leq C(u_2,v)-C(u_1,v)\leq u_2-u_1$.

4. Por el literal *3*, si $0\leq x_2\leq y_2\leq 1$, entonces $0\leq C(x_1,y_2)-C(x_1,x_2)\leq y_2-x_2$, y si $0\leq y_2\leq x_2\leq 1$, entonces $0\leq C(x_1,x_2)-C(x_1,y_2)\leq x_2-y_2$. Por lo tanto $|C(x_1,y_2)-C(x_1,x_2)|\leq |y_2-x_2|$. Análogamente se tiene $|C(x_1,y_2)-C(y_1,y_2)|\leq |y_1-x_1|$. De aquí que

$$\begin{aligned}|C(x_1,x_2)-C(y_1,y_2)| &= |C(x_1,x_2)-C(x_1,y_2)+C(x_1,y_2)-C(y_1,y_2)|,\\ &\leq |C(x_1,x_2)-C(x_1,y_2)|+|C(x_1,y_2)-C(y_1,y_2)|,\\ &\leq |x_2-y_2|+|y_1-x_1|.\end{aligned}$$

5. Por el literal *2*, para $v\in I$ fijo, la función $u\mapsto C(u,v)$ es creciente, por lo que es derivable en casi todo $u\in I$. Además, del literal *3*,

$$0\leq \lim_{h\to 0}\frac{C(u+h,v)-C(u,v)}{h}\leq 1,\quad \text{es decir,}\quad 0\leq \frac{\partial C(u,v)}{\partial u}\leq 1.$$

Similarmente se muestra que $0\leq \dfrac{\partial C(u,v)}{\partial v}\leq 1$.

□

Teorema 2.1.2 (Cotas de *Fréchet-Hoeffding*)**.** *Si $C:I^2\to I$ es una función cópula, entonces*

$$\max\{u+v-1,0\}\leq C(u,v)\leq \min\{u,v\}.$$

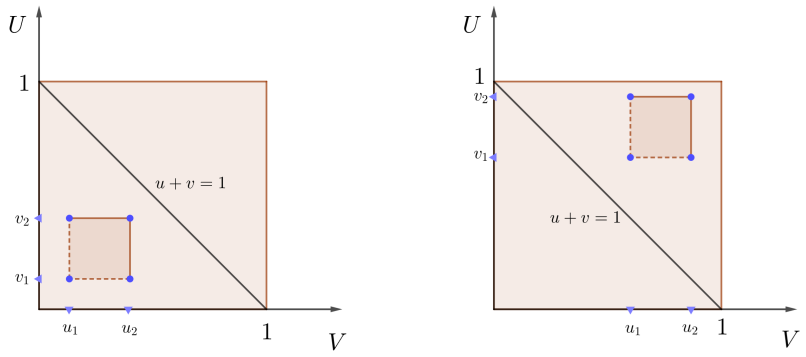
Más aún, $C_{\text{ínf}}(u,v):=\max\{u+v-1,0\}$ y $C_{\text{sup}}(u,v):=\min\{u,v\}$, $(u,v)\in I^2$, son funciones cópula.

Demostración. Por el literal *2* del Teorema 2.1.1, se tiene que $C(u,v)\leq C(u,1)=u$ y que $C(u,v)\leq C(1,v)=v$, por lo cual $C(u,v)\leq \min\{u,v\}$. De otra parte,

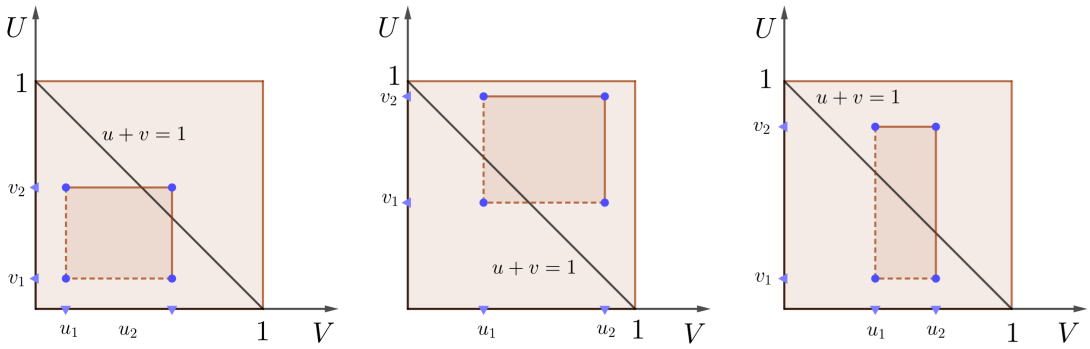
$$\begin{aligned}\Delta^1_{u1}\Delta^2_{v1}C(x,y) &= C(1,1)-C(u,1)-C(1,v)+C(u,v),\\ &= 1-u-v+C(u,v)\geq 0,\end{aligned}$$

lo cual implica que $u+v-1\leq C(u,v)$. Además, como $C(u,v)\geq 0$, se concluye que $\max\{u+v-1,0\}\leq C(u,v)$.

De otro lado, veamos que $C_{\text{ínf}}(u,v)=\max\{u+v-1,0\}$ es una función cópula:



(a) Caso 1.



(b) Caso 2.

Figura 2.1: Verificación de (C2) para $C_{\text{ínf}}(u, v)$.

(C1) Note que $C_{\text{ínf}}(u, 1) = \text{máx}\{u, 0\} = u$, $C_{\text{ínf}}(1, v) = \text{máx}\{v, 0\} = v$, $C_{\text{ínf}}(u, 0) = \text{máx}\{u - 1, 0\} = 0$ y $C_{\text{ínf}}(0, v) = \text{máx}\{v - 1, 0\} = 0$. Por tanto, $C_{\text{ínf}}(u, v) = uv$ para todo $(u, v) \in \partial(I^2)$.

(C2) Sean $0 \leq u_1 \leq u_1 \leq 1$ y $0 \leq v_1 \leq v_2 \leq 1$.

Caso 1: Si $(u_1, u_2] \times (v_1, v_2]$ no corta a la recta $u + v = 1$, como en los casos que se muestran en la Figura 2.1a, entonces $\Delta^1_{u_1u_2}\Delta^2_{v_1v_2}C(x, y) = 0 \geq 0$.

Caso 2: Si $(u_1, u_2] \times (v_1, v_2]$ corta a la recta $u + v = 1$, como en los casos que se muestran en la Figura 2.1b, se tiene que $\Delta^1_{u_1u_2}\Delta^2_{v_1v_2}C(x, y) \geq 0$.

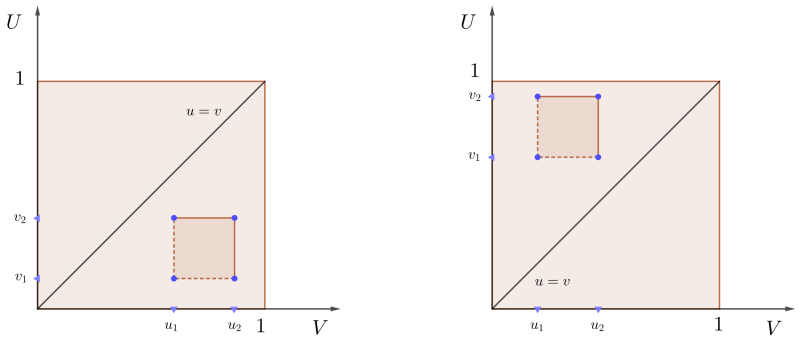
Finalmente, veamos que $C_{\text{sup}}(u, v) = \text{mín}\{u, v\}$ es una función cópula:

(C1) Note que $C_{\text{sup}}(u, 1) = \text{mín}\{u, 1\} = u$, $C_{\text{sup}}(1, v) = \text{mín}\{1, v\} = v$, $C_{\text{sup}}(u, 0) = \text{mín}\{u, 0\} = 0$ y $C_{\text{sup}}(0, v) = \text{mín}\{0, v\} = 0$. Por tanto, $C_{\text{sup}}(u, v) = uv$ para todo $(u, v) \in \partial(I^2)$.

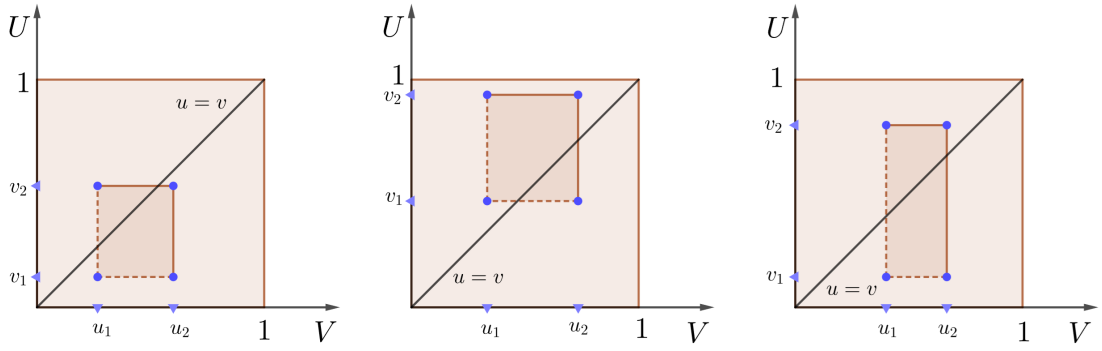
(C2) Sean $0 \leq u_1 \leq u_1 \leq 1$ y $0 \leq v_1 \leq v_2 \leq 1$.

Caso 1: Si $(u_1, u_2] \times (v_1, v_2]$ no corta a la recta $u = v$, como en los casos que se muestran en la Figura 2.2a, entonces $\Delta^1_{u_1u_2}\Delta^2_{v_1v_2}C(x, y) = 0 \geq 0$.

Caso 2: Si $(u_1, u_2] \times (v_1, v_2]$ corta a la recta $u = v$, como en los casos que se muestran en la Figura 2.2b, se tiene que $\Delta^1_{u_1 u_2} \Delta^2_{v_1 v_2} C(x, y) \geq 0$.



(a) Caso 1.



(b) Caso 2.

Figura 2.2: Verificación de (C2) para $C_{\text{sup}}(u, v)$.

□

Observación 2.1.1. *Sea $C : I^2 \to I$ una cópula. La función $F_{(X,Y)}(\cdot, \cdot) : \mathbb{R}^2 \to I$ definida por*

$$F_{(X,Y)}(x, y) := \begin{cases} 0, & si\ x < 0,\ y < 0, \\ C(x, y), & si\ (x, y) \in I^2, \\ x, & si\ y > 1,\ x \in I, \\ y, & si\ x > 1,\ y \in I, \\ 1, & si\ x > 1,\ y > 1. \end{cases}$$

es una función de distribución bivariada tal que

$$F_X(x) = x \mathbb{1}_{[0,1]}(x) + \mathbb{1}_{[1,\infty)}(x),$$
$$F_Y(y) = y \mathbb{1}_{[0,1]}(y) + \mathbb{1}_{[1,\infty)}(y),$$

es decir, las distribuciones marginales $F_X(\cdot)$ y $F_Y(\cdot)$ son uniformes en $[0, 1]$.

De acuerdo a la observación anterior, se puede dar otra definición para funciones cópula, que se presenta a continuación.

Definición 2.1.2. *Una función cópula es la restricción a I^2 de una función de distribución bivariada que tiene distribuciones marginales uniformes en $(0,1)$.*

Ejemplo 2.1.1. *Sea (X,Y) un vector aleatorio tal que $X \sim Unif([0,1])$ y $Y = X$. Entonces $F_{(X,Y)}(\cdot,\cdot)$ es una función de distribución bivariada que tiene distribuciones marginales uniformes en $[0,1]$. Por tanto,*

$$C(u,v) := F_{(X,Y)}(u,v), \quad (u,v) \in I^2,$$

es una función cópula. En efecto, para $(u,v) \in I^2$,

$$\begin{aligned} C(u,v) &= F_{(X,Y)}(u,v), \\ &= P(X \leq u, Y \leq v), \\ &= P(X \leq u, X \leq v), \\ &= P(X \leq \text{mín}\,\{u,v\}), \\ &= \text{mín}\,\{u,v\} \equiv C_{\text{sup}}(u,v). \end{aligned}$$

Ejemplo 2.1.2. *Sea (X,Y) un vector aleatorio tal que $X \sim Unif([0,1])$ y $Y = 1-X$. Como $Y \sim Unif([0,1])$, entonces $F_{(X,Y)}(\cdot,\cdot)$ es una función de distribución bivariada que tiene distribuciones marginales uniformes en $[0,1]$. Por tanto,*

$$C(u,v) := F_{(X,Y)}(u,v), \quad (u,v) \in I^2,$$

es una función cópula. En efecto, para $(u,v) \in I^2$,

$$\begin{aligned} C(u,v) &= F_{(X,Y)}(u,v), \\ &= P(X \leq u, Y \leq v), \\ &= P(X \leq u, 1-X \leq v), \\ &= P(1-v \leq X \leq u), \\ &= \begin{cases} u+v-1, & si\ 1-v \leq u, \\ 0, & si\ u < 1-v, \end{cases} \\ &= \text{máx}\,\{u+v-1, 0\} \equiv C_{\text{ínf}}(u,v). \end{aligned}$$

Otros ejemplos de funciones cópula serán presentados en la Sección 2.4.

2.2. El Teorema de Sklar

Definición 2.2.1. *Sea (X,Y) un vector aleatorio. Se dice que una función cópula $C : I^2 \to I$ está **asociada** a (X,Y) si para todo $(x,y) \in \mathbb{R}^2$, $F_{(X,Y)}(x,y) = C(F_X(x), F_Y(y))$.*

Ejemplo 2.2.1. *Si X y Y son variables aleatorias independientes, entonces $F_{(X,Y)}(x,y) = F_X(x)F_Y(y)$, por lo que una función cópula asociada a (X,Y) es la* ***cópula de la independencia*** *(o* ***cópula producto****), $C_{ind}(u,v) = uv$.*

Proposición 2.2.1. *Sea (X,Y) un vector aleatorio tal que $F_X(\cdot)$ y $F_Y(\cdot)$ son continuas. Entonces existe una única función cópula $C_{X,Y}(\cdot,\cdot)$ asociada a (X,Y) que está dada por*

$$C_{(X,Y)}(u,v) := F_{(X,Y)}(F_X^{-1}(u), F_Y^{-1}(v)). \tag{2.1}$$

Demostración. Por la Transformada inversa de probabilidad (Teorema 1.1.1), se tiene que $U = F_X(X) \sim \text{Unif}([0,1])$ y $V = F_Y(Y) \sim \text{Unif}([0,1])$. De esta manera (U,V) es un vector aleatorio que tiene distribuciones marginales uniformes en $[0,1]$ y su función de distribución, restringida a I^2, es una cópula.

$$\begin{aligned} F_{(U,V)}(u,v) &= P(U \le u, V \le v), \\ &= P(F_X(X) \le u, F_Y(Y) \le v), \\ &= P(X \le F_X^{-1}(u), Y \le F_Y^{-1}(v)), \\ &= F_{(X,Y)}(F_X^{-1}(u), F_Y^{-1}(v)), \\ &=: C_{(X,Y)}(u,v). \end{aligned}$$

Note que, para $(x,y) \in \mathbb{R}^2$,

$$\begin{aligned} C_{(X,Y)}(F_X(x), F_Y(y)) &= F_{(X,Y)}(F_X^{-1}(F_X(x)), F_Y^{-1}(F_Y(y))), \\ &= F_{(X,Y)}(x,y), \end{aligned}$$

por lo tanto $C_{(X,Y)}(\cdot,\cdot)$ es una función cópula asociada a (X,Y), de acuerdo con la Definición 2.2.1.
Ahora, si $\tilde{C}(\cdot,\cdot)$ es otra función cópula asociada a (X,Y), entonces, por definición,

$$\tilde{C}(F_X(x), F_Y(y)) = F_{(X,Y)}(x,y) = C_{(X,Y)}(F_X(x), F_Y(y)), \quad \forall (x,y) \in \mathbb{R}^2.$$

De esta manera, $\tilde{C}(\cdot,\cdot)$ y $C_{(X,Y)}(\cdot,\cdot)$ coinciden en $(0,1)^2$, y por la continuidad uniforme de las funciones cópula (Teorema 2.1.1, literal *4*), coinciden en I^2. □

Observación 2.2.1. *La anterior proposición sugiere un método para construir funciones cópula, por medio de la ecuación (2.1), para $(u,v) \in I^2$.*

Ejemplo 2.2.2. *Sea (X,Y) un vector aleatorio continuo cuya función de densidad está dada por $f_{(X,Y)}(x,y) = (x+y)\mathbb{1}_{[0,1]\times[0,1]}(x,y)$. Su función de distribución, para $(x,y) \in I^2$, es:*

$$\begin{aligned} F_{(X,Y)}(x,y) &= \int_{-\infty}^{y}\int_{-\infty}^{x} f_{(X,Y)}(t_1,t_2)dt_1dt_2, \\ &= \int_{0}^{y}\int_{0}^{x} (t_1+t_2)dt_1dt_2, \\ &= \int_{0}^{y}\left[\frac{x^2}{2}+t_2x\right]dt_2, \\ &= \frac{x^2y}{2}+\frac{xy^2}{2}, \\ &= \frac{xy(x+y)}{2}, \end{aligned}$$

y sus funciones de distribución marginales, para $x, y \in I$*, son:*

$$F_X(x) = \lim_{y\to\infty} F_{(X,Y)}(x,y) = \frac{x(x+1)}{2},$$

$$F_Y(y) = \lim_{x\to\infty} F_{(X,Y)}(x,y) = \frac{y(y+1)}{2}.$$

De aquí se pueden obtener sus inversas:

$$F_X^{-1}(u) = \frac{-1+\sqrt{1+8u}}{2},$$

$$F_Y^{-1}(v) = \frac{-1+\sqrt{1+8v}}{2}.$$

Así, remplazando lo anterior en la ecuación (2.1)*, se obtiene que*

$$\begin{aligned} C_{(X,Y)}(u,v) &= \frac{\left(\frac{\sqrt{1+8u}-1}{2}\right)^2\left(\frac{\sqrt{1+8v}-1}{2}\right)+\left(\frac{\sqrt{1+8u}-1}{2}\right)\left(\frac{\sqrt{1+8v}-1}{2}\right)^2}{2}, \\ &= \frac{1}{16}(\sqrt{1+8u}-1)(\sqrt{1+8v}-1)(\sqrt{1+8u}+\sqrt{1+8v}-2), \end{aligned}$$

la cual es la función cópula asociada a (X,Y)*.*

La proposición anterior se puede extender a vectores aleatorios (X,Y) sin la condición de que $F_X(\cdot)$ y $F_Y(\cdot)$ sean continuas. Esta extensión se establece en el siguiente teorema.

Teorema 2.2.1 (Teorema de Sklar)**.** *Sea* (X,Y) *un vector aleatorio bidimensional. Entonces existe una función cópula* $C_{(X,Y)}(\cdot,\cdot)$ *tal que para todo* $(x,y) \in \mathbb{R}^2$*,* $F_{(X,Y)}(x,y) = C_{(X,Y)}(F_X(x), F_Y(y))$*.*

Una demostración de este teorema se puede consultar en la sección 2.3 de Nelsen [21].

Observación 2.2.2. *Cuando $F_X(\cdot)$ y $F_Y(\cdot)$ no son continuas, pueden existir funciones cópula distintas que verifiquen la igualdad anterior.*

El recíproco del Teorema de Sklar también es válido, y se enuncia a continuación.

Teorema 2.2.2 (Recíproco del Teorema de Sklar)**.** *Sean $C : I^2 \to I$ una función cópula y $F_X(\cdot)$, $F_Y(\cdot)$ funciones de distribución arbitrarias. Entonces $F(x,y) := C(F_X(x), F_Y(y))$, $(x,y) \in \mathbb{R}^2$ es una función de distribución bivariada que tiene distribuciones marginales $F_X(\cdot)$ y $F_Y(\cdot)$.*

Ejemplo 2.2.3. *Suponga que X y Y son variables aleatorias independientes tales que $X \sim Exp(\lambda)$ y $Y \sim Bern(p)$. Entonces (ver Figura 2.3),*

$$F_X(x) = \begin{cases} 0, & si\ x < 0, \\ 1 - e^{-\lambda x}, & si\ x \geq 0, \end{cases} \qquad F_Y(y) = \begin{cases} 0, & si\ y < 0, \\ 1-p, & si\ 0 \leq y < 1, \\ 1, & si\ y \geq 1. \end{cases}$$

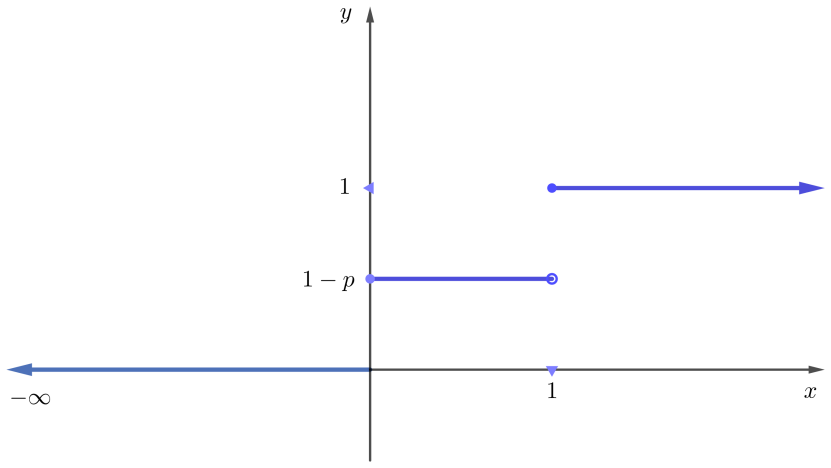


Figura 2.3: Funciones de distribución del ejemplo 2.2.3.

En este caso, la función cópula asociada al vector aleatorio (X,Y) es $C_{ind}(u,v) = uv$. Por lo tanto,

$$\begin{aligned} F_{(X,Y)}(x,y) &= C(F_X(x), F_Y(y)), \\ &= F_X(x)F_Y(y), \\ &= \begin{cases} 0, & si\ x \leq 0,\ y \leq 0, \\ (1-p)(1-e^{-\lambda x}), & si\ x \geq 0,\ 0 \leq y < 1, \\ 1 - e^{-\lambda x}, & si\ x \geq 0,\ y \geq 1. \end{cases} \end{aligned}$$

Se puede verificar que $F_{(X,Y)}(\cdot,\cdot)$ es una función de distribución cuyas marginales son precisamente $F_X(\cdot)$ y $F_Y(\cdot)$.

En la siguiente proposición se presentan otras propiedades básicas de las funciones cópula; entre ellas, se establece que las cópulas son invariantes bajo transformaciones estrictamente crecientes.

Proposición 2.2.2. *Sean X y Y variables aleatorias continuas con función cópula asociada $C_{(X,Y)}(\cdot,\cdot)$. Sean $\alpha(\cdot)$ y $\beta(\cdot)$ funciones estrictamente monótonas sobre el rango de X y el rango de Y, respectivamente. Entonces*

1. *Si $\alpha(\cdot)$ y $\beta(\cdot)$ son estrictamente crecientes, entonces*

$$C_{(\alpha(X),\beta(Y))}(u,v) = C_{(X,Y)}(u,v), \quad (u,v) \in I^2.$$

2. *Si $\alpha(\cdot)$ es estrictamente creciente y $\beta(\cdot)$ es estrictamente decreciente, entonces*

$$C_{(\alpha(X),\beta(Y))}(u,v) = u - C_{(X,Y)}(u,1-v), \quad (u,v) \in I^2.$$

3. *Si $\alpha(\cdot)$ es estrictamente decreciente y $\beta(\cdot)$ es estrictamente creciente, entonces*

$$C_{(\alpha(X),\beta(Y))}(u,v) = v - C_{(X,Y)}(1-u,v), \quad (u,v) \in I^2.$$

4. *Si $\alpha(\cdot)$ y $\beta(\cdot)$ son estrictamente decrecientes, entonces*

$$C_{(\alpha(X),\beta(Y))}(u,v) = u + v - 1 + C_{(X,Y)}(1-u,1-v), \quad (u,v) \in I^2.$$

Demostración. 1. Como $F_{\alpha(X)}(x) = P(\alpha(X) \leq x) = P(X \leq \alpha^{-1}(x)) = F_X(\alpha^{-1}(x))$ y $F_{\beta(Y)}(y) = F_Y(\beta^{-1}(y))$, entonces para $(x,y) \in \mathbb{R}^2$,

$$\begin{aligned} C_{(\alpha(X),\beta(Y))}(F_{\alpha(X)}(x), F_{\beta(Y)}(y)) &= F_{\alpha(X),\beta(Y)}(x,y), \\ &= P(\alpha(X) \leq x, \beta(Y) \leq y), \\ &= P(X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)), \\ &= F_{(X,Y)}(\alpha^{-1}(x), \beta^{-1}(y)), \\ &= C_{(X,Y)}(F_X(\alpha^{-1}(x)), F_Y(\beta^{-1}(y))), \\ &= C_{(X,Y)}(F_{\alpha(X)}(x), F_{\beta(Y)}(y)). \end{aligned}$$

Como X y Y son variables aleatorias continuas, $\mathrm{Ran}(F_{\alpha(X)}(\cdot)) = I = \mathrm{Ran}(F_{\beta(Y)}(\cdot))$ y por tanto $C_{(\alpha(X),\beta(Y))}(u,v) = C_{(X,Y)}(u,v)$ para todo $(u,v) \in I^2$.

4. Como $F_{\alpha(X)}(x) = P(\alpha(X \leq x)) = P(X \geq \alpha^{-1}(x)) = 1 - F_X(\alpha^{-1}(x))$ y $F_{\beta(Y)}(y) = 1 - F_Y(\beta^{-1}(y))$, entonces para $(x,y) \in \mathbb{R}^2$,

$$\begin{aligned}
C_{(\alpha(X),\beta(Y))}(F_{\alpha(X)}(x), F_{\beta(Y)}(y)) &= F_{(\alpha(X),\beta(Y))}(x,y) \\
&= P(\alpha(X) \leq x, \beta(Y) \leq y) \\
&= P(X \geq \alpha^{-1}(x), Y \geq \beta^{-1}(y)) \\
&= 1 - P(X \leq \alpha^{-1}(x)) - P(Y \leq \beta^{-1}(y)) \\
&\quad + P(X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)) \\
&= 1 - F_X(\alpha^{-1}(x)) - F_Y(\beta^{-1}(y)) \\
&\quad + F_{(X,Y)}(\alpha^{-1}(x), \beta^{-1}(y)) \\
&= F_{\alpha(X)}(x) + F_{\beta(Y)}(y) - 1 \\
&\quad + C_{(X,Y)}(F_X(\alpha^{-1}(x)), F_Y(\beta^{-1}(y))) \\
&= F_{\alpha(X)}(x) + F_{\beta(Y)}(y) - 1 \\
&\quad + C_{(X,Y)}(1 - F_{\alpha(X)}(x), 1 - F_{\beta(Y)}(y)).
\end{aligned}$$

Tomando $u = F_{\alpha(X)}(x)$ y $v = F_{\beta(Y)}(y)$ se obtiene que

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

Las demostraciones de los literales *2* y *3* siguen las mismas ideas. □

Recordemos que para un vector aleatorio (X,Y) continuo absolutamente se tiene el concepto de *función de densidad*, cuyo conocimiento es equivalente al de la función de distribución. Dado que las funciones cópula son en particular funciones de distribución bivariadas, es posible asociarles de manera análoga una función de densidad; esta idea se precisa en la siguiente definición.

Definición 2.2.2. *Sea $C(u,v)$ una función cópula. Su* ***función de densidad asociada****, denotada como $c(u,v)$, está dada por* $c(u,v) := \dfrac{\partial^2 C(u,v)}{\partial u \partial v}$.

De la Definición 2.2.1, se sabe que la función de distribución del vector aleatorio (X,Y) está relacionada con su función cópula asociada. Más concretamente,

$$F_{(X,Y)}(x,y) = C(F_X(x), F_Y(y)). \tag{2.2}$$

De acuerdo a esto, es natural suponer que exista también una relación entre la función de densidad del vector (X,Y) y la densidad de su cópula asociada. En efecto, derivando parcialmente a ambos lados de la ecuación (2.2) con respecto a cada variable, tenemos que

$$\frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x,y) = \frac{\partial^2}{\partial x \partial y} C(F_X(x), F_Y(y)),$$

$$f_{(X,Y)}(x,y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y), \tag{2.3}$$

o, equivalentemente,

$$c(F_X(x), F_Y(y)) = \frac{f_{(X,Y)}(x,y)}{f_X(x) f_Y(y)}. \tag{2.4}$$

La ecuación (2.4) se conoce como la **representación canónica** de la cópula $C(u,v)$.

2.3. Una visión heurística de las cópulas

Las funciones cópula son una herramienta fundamental en el estudio de la dependencia entre variables aleatorias. Diferentes autores, como Fermanian et al. [8], Ledwina [18] y Nelsen [21], han señalado que las cópulas permiten describir la estructura de dependencia entre las componentes X y Y de un vector aleatorio (X,Y), independientemente de sus distribuciones marginales $F_X(\cdot)$ y $F_Y(\cdot)$.

Pero, ¿qué significa realmente que una función cópula "capture" la dependencia? En esta sección, exploraremos esta idea desde un enfoque heurístico, complementando lo visto en la sección 2.2, donde el **Teorema de Sklar** muestra que cualquier función de distribución conjunta se puede descomponer en sus distribuciones marginales y una función cópula que las vincula. Además, discutiremos cómo las cópulas se relacionan con las **curvas de regresión**, que fueron introducidas en la sección 1.1.5, proporcionando así una interpretación más intuitiva de su papel en el modelamiento de la dependencia.

En primer lugar, consideremos un vector aleatorio (X,Y), donde X y Y son independientes y ambas tienen distribución uniforme en $I = [0,1]$. Entonces la función de distribución de (X,Y) es

$$F_{(X,Y)}(x,y) = P(X \leq x)P(Y \leq y) = xy,$$

que resulta ser precisamente la cópula de la independencia o cópula producto.

Ahora bien, consideremos una función biyectiva y con derivadas parciales continuas,

$$\begin{array}{rccl} g & : & I \times I & \to & I \times I \\ & & (x,y) & \mapsto & g(x,y) = (g_1(x,y), g_2(x,y)) = (u,v). \end{array}$$

También definamos el vector aleatorio (U,V) dado por $U = g_1(X,Y)$ y $V = g_2(X,Y)$ (ver Figura 2.4).
Por el teorema de las transformaciones (Teorema 13, p. 205, Mood et al. [20]), la función de densidad de este vector aleatorio está dada por

$$f_{(U,V)}(u,v) = |J| \cdot f_{(X,Y)}(g_1^{-1}(u,v), g_2^{-1}(u,v)), \quad (u,v) \in [0,1],$$

donde J representa el jacobiano de la función g y está dado por

$$J = \det \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{pmatrix}.$$

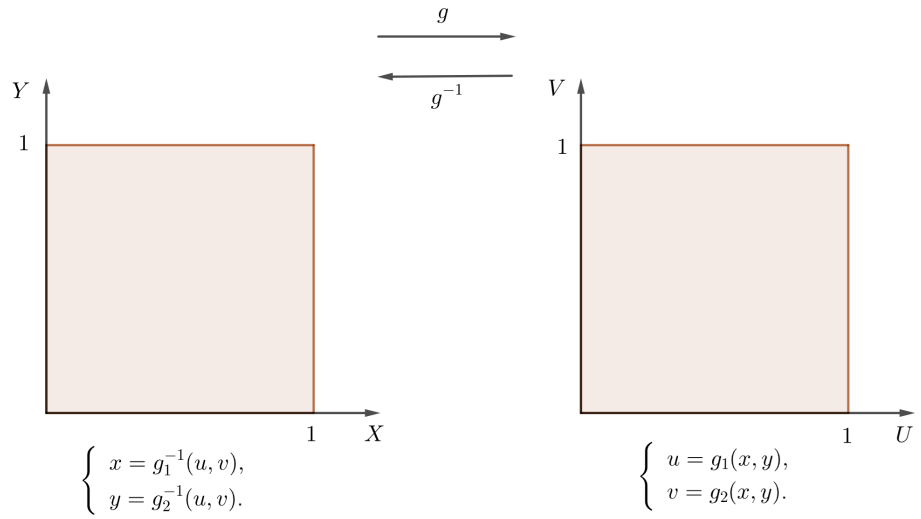


Figura 2.4: Ilustración de la transformación g.

Dado que (X, Y) tiene distribución uniforme en I^2, se sigue que

$$f_{(X,Y)}(g_1^{-1}(u,v), g_2^{-1}(u,v)) = 1;$$

así, $f_{(U,V)}(u,v) = |J|$.

En particular, tomemos $g_1(x,y) = x$, con lo cual $U = X$. De esta manera, la función de densidad de (U,V) es

$$f_{(U,V)}(u,v) = \left|\det \begin{pmatrix} 1 & 0 \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{pmatrix}\right| = \left|\frac{\partial y}{\partial v}\right| = \left|\frac{\partial}{\partial v} g_2^{-1}(u,v)\right|, \quad (u,v) \in I^2,$$

y su función de distribución $F_{(U,V)}(u,v)$ satisface que

- $F_{(U,V)}(0,v) = 0 = F_{(U,V)}(u,0)$;
- $F_{(U,V)}(u,1) = F_U(u) = u$ (ya que U tiene distribución uniforme en I).

Además, de la Definición 1.1.8 se sigue que

$$f_{V|U}(v|u) = \frac{f_{(U,V)}(u,v)}{f_U(u)} = f_{(U,V)}(u,v),$$

puesto que $f_U(u) = 1$.

Ahora bien, tome $u \in I$ fijo y $v^* \in I$ arbitrario. Entonces

$$\begin{aligned} F_{V|U}(v^*|u) &= P(V \leq v^*|U = u), \\ &= \int_0^{v^*} f_{V|U}(v|u)dv, \\ &= \int_0^{v^*} f_{(U,V)}(u,v)dv, \\ &= \int_0^{v^*} \frac{\partial^2 F_{(U,V)}(u,v)}{\partial v \partial u} dv, \\ &= \frac{\partial F_{(U,V)}(u,v)}{\partial u}\Big|_{v=0}^{v=v^*}. \end{aligned}$$

Similarmente, note que

$$\begin{aligned} F_{V|U}(v^*|u) &= \int_0^{v^*} f_{(U,V)}(u,v)dv, \\ &= \int_0^{v^*} \left| \frac{\partial}{\partial v} g_2^{-1}(u,v) \right| dv, \\ &= \int_0^{v^*} \frac{\partial}{\partial v} g_2^{-1}(u,v)dv, \quad \text{(tomando } g_2^{-1}(u,v) > 0\text{)} \\ &= g_2^{-1}(u,v)\Big|_{v=0}^{v=v^*}. \end{aligned}$$

En consecuencia, para casi todo $(u, v^*) \in I^2$, se tiene que

$$g_2^{-1}(u,v^*) = \frac{\partial F_{(U,V)}(u,v^*)}{\partial u} = P(V \leq v^*|U = u) = E[\mathbb{1}_{[0,v^*]}(V)|U = u],$$

que resulta ser la curva de regresión de $\mathbb{1}_{[0,v^*]}(V)$ sobre U.

Si adicionalmente se exige que $V = g_2(X,Y) \sim \text{Unif}([0,1])$, entonces la función $F_{(U,V)} : I^2 \to I$ es una cópula absolutamente continua, ya que en ese caso también satisface que $F_{(U,V)}(1,v) = F_V(v) = v$, verificándose así *(C1)* de la Definición 2.1.1. Observe que *(C2)* ya se tiene por el hecho de que $F_{(U,V)}(u,v)$ es una función de distribución.

De acuerdo a lo anterior, la transformación g^{-1} queda definida como

$$\begin{aligned} g^{-1} \;:\; I \times I &\to I \times I \\ (u,v) &\mapsto g^{-1}(u,v) = (g_1^{-1}(u,v), g_2^{-1}(u,v)) = \left(u, \frac{\partial F_{(U,V)}(u,v)}{\partial u}\right), \end{aligned}$$

y por tanto la transformación g resulta ser

$$\begin{aligned} g \;:\; I \times I &\to I \times I \\ (x,y) &\mapsto g(x,y) = (g_1(x,y), g_2(x,y)) = \left(x, \left(\frac{\partial F_{(U,V)}(u,v)}{\partial u}\right)^{-1}(x,y)\right). \end{aligned}$$

Podemos interpretar esto diciendo que si las variables aleatorias X y Y son i.i.d. con distribución uniforme en I, entonces la transformación

$$U = g_1(X,Y) = X, \qquad V = g_2(X,Y) = \left(\frac{\partial C(u,v)}{\partial u}\right)^{-1}(X,Y),$$

es tal que las variables aleatorias U y V quedan vinculadas o "copuladas" con función de distribución $C(u,v)$. Recíprocamente, si las variables aleatorias U y V tienen como función de distribución $C(u,v)$, $u,v \in I$, entonces la transformación

$$X = g_1^{-1}(U,V) = U, \qquad Y = g_2^{-1}(U,V) = \left(\frac{\partial C(u,v)}{\partial u}\right)(U,V),$$

es tal que X y Y son variables aleatorias i.i.d. con distribución uniforme en I.

En resumen, se puede concluir que cada función cópula absolutamente continua $C(u,v)$ con derivadas parciales induce una transformación biyectiva g que relaciona variables aleatorias independientes con variables aleatorias vinculadas mediante la cópula $C(u,v)$.

Observación 2.3.1. *Una de las aplicaciones que tiene la transformación g es en la simulación de observaciones aleatorias de una función cópula específica. El siguiente es un algoritmo para tal propósito, adaptado de la sección 2.9 de Nelsen [21]:*

1. *Genere dos muestras independientes, $(x_1, x_2, \ldots, x_n)$ y $(y_1, y_2, \ldots, y_n)$, de variables aleatorias uniformes en $I = [0,1]$.*
2. *Elija una función cópula $C(u,v)$ (absolutamente continua y con derivadas parciales) y especifique la transformación g dada por $g(x,y) = \left(x, \left(\frac{\partial C(u,v)}{\partial u}\right)^{-1}(x,y)\right) = (u,v)$.*
3. *Defina $(u_i, v_i) = g(x_i, y_i)$ para cada $i = 1, \ldots, n$.*
4. *Las parejas (u_i, v_i) son observaciones de la cópula $C(u,v)$.*

Así como la transformación g permite dar un algoritmo para simular observaciones aleatorias de una función cópula, su inversa, g^{-1}, permite analizar cómo una variable aleatoria depende de la otra, descomponiendo esta relación en términos de curvas de regresión.
En particular, la relación se hace evidente a través de la expresión

$$g_2^{-1}(u,v) = E[\mathbb{1}_{[0,v]}(V)|U = u],$$

(o alternativamente,

$$g_1^{-1}(u,v) = E[\mathbb{1}_{[0,u]}(U)|V = v]$$

si en la transformación g se toma $y = v$ en lugar de $x = u$).

Cuando v (o u) es fijo, esta función describe la curva de regresión de $\mathbb{1}_{[0,v]}(V)$ sobre U (o $\mathbb{1}_{[0,u]}(U)$ sobre V). Es decir, representa la "mejor" aproximación de una variable en función de la otra en términos de la esperanza condicionada. Estas curvas de regresión son las que explican la estructura de dependencia entre las variables U y V. Ilustraremos la discusión anterior mediante los siguientes ejemplos concretos.

Ejemplo 2.3.1. *Consideremos la función cópula* $C(u,v) = \dfrac{uv}{u+v-uv}$, $(u,v) \neq (0,0)$. *Sus derivadas parciales están dadas por*

$$\partial_1 C(u,v) = \frac{\partial C(u,v)}{\partial u} = \left(\frac{v}{u+v-uv}\right)^2,$$
$$\partial_2 C(u,v) = \frac{\partial C(u,v)}{\partial v} = \left(\frac{u}{u+v-uv}\right)^2.$$

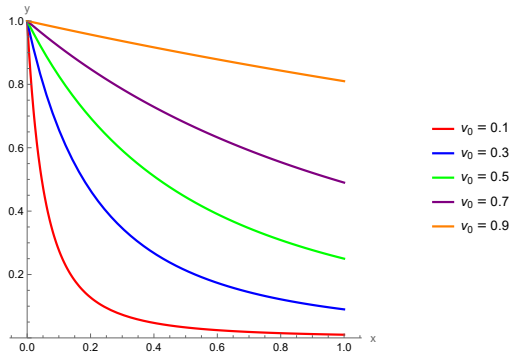
Si fijamos $x = u$, *la transformación* g^{-1} *es*

$$x = u, \qquad y = g_2^{-1}(u,v) = \partial_1 C(u,v_0) = E[\mathbb{1}_{[0,v_0]}(V)|U = u],$$

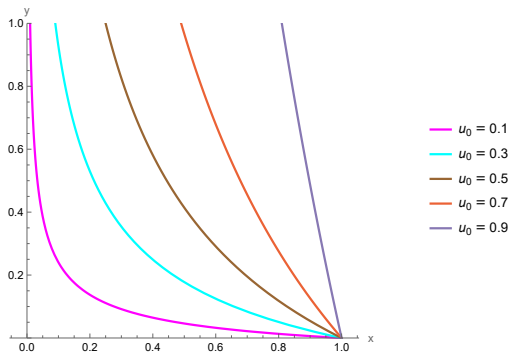
donde $v_0 \in [0,1]$ *es un valor fijo. Similarmente, si establecemos* $y = v$, *tenemos que* g^{-1} *está dada por*

$$x = g_1^{-1}(u,v) = \partial_2 C(u_0,v) = E[\mathbb{1}_{[0,u_0]}(U)|V = v], \qquad y = v,$$

donde $u_0 \in [0,1]$ *es un valor fijo. De esta manera obtenemos las curvas de regresión de* $\mathbb{1}_{[0,v_0]}(V)$ *sobre* U, *y de* $\mathbb{1}_{[0,u_0]}(U)$ *sobre* V. *En la Figura 2.5 se ilustran algunas de ellas, para algunos valores fijos de* v_0 *y* u_0 *respectivamente.*



(a) Curvas de regresión de $\mathbb{1}_{[0,v_0]}(V)$ sobre U.

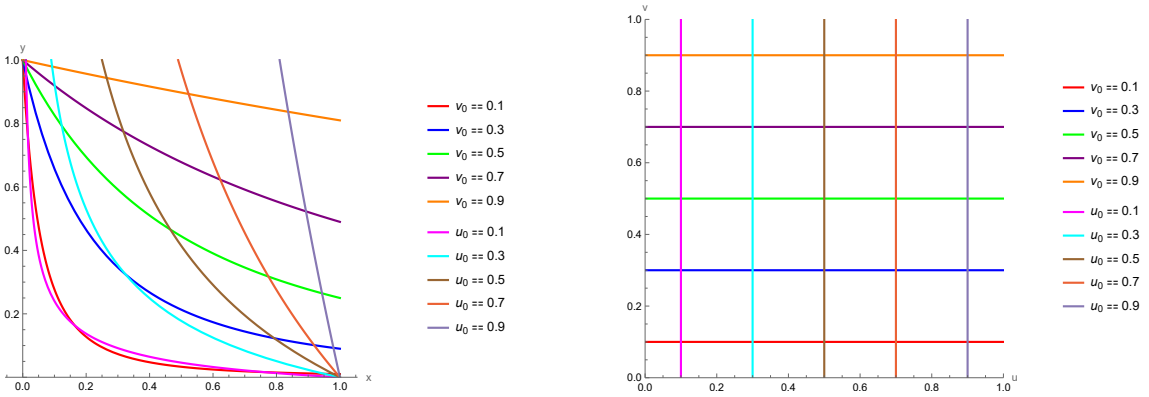


(b) Curvas de regresión de $\mathbb{1}_{[0,u_0]}(U)$ sobre V.

Figura 2.5: Curvas de regresión para la cópula $C(u,v) = \dfrac{uv}{u+v-uv}$, $(u,v) \neq (0,0)$.

Asimismo, en la Figura 2.6 se evidencia una correspondencia entre las curvas de regresión descritas y las rectas $u = u_0$ *y* $v = v_0$, *que surgen cuando se toman los*

valores fijos de u_0 y v_0. Estas correspondencias se representan mediante un esquema de colores: cada curva y su recta asociada comparten el mismo color en ambos gráficos.



(a) Curvas de regresión. **(b)** Rectas $u = u_0$ y $v = v_0$.

Figura 2.6: Correspondencia entre las curvas de regresión y las rectas $u = u_0$ y $v = v_0$ para algunos valores fijos de u_0 y v_0.

Las rectas que se muestran en la Figura 2.6b corresponden a las curvas de regresión cuando las variables aleatorias son independientes. De esta manera, los gráficos 2.6a y 2.6b permiten evidenciar el efecto de la transformación biyectiva g: en un sentido asocia o "copula" las variables aleatorias, mientras que en el otro las separa o "independiza".

Ejemplo 2.3.2. *Consideremos ahora la función cópula $C(u, v) = uv+\theta uv(1-u)(1-v)$, $\theta \in [-1, 1]$. Sus derivadas parciales están dadas por*

$$\partial_1 C(u, v) = \frac{\partial C(u, v)}{\partial u} = v + \theta v(1 - u)(1 - v) - \theta uv(1 - v),$$

$$\partial_2 C(u, v) = \frac{\partial C(u, v)}{\partial v} = u + \theta u(1 - u)(1 - v) - \theta uv(1 - u).$$

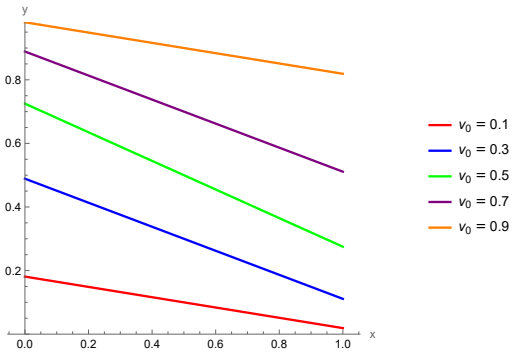
Las posibles transformaciones g^{-1} que se pueden tener son:

$$x = u, \qquad y = g_2^{-1}(u, v) = \partial_1 C(u, v_0) = E[\mathbb{1}_{[0,v_0]}(V)|U = u],$$

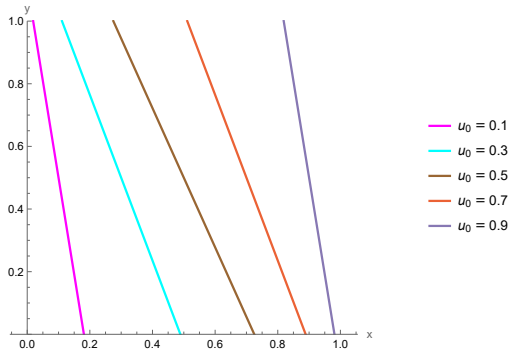
$$x = g_1^{-1}(u, v) = \partial_2 C(u_0, v) = E[\mathbb{1}_{[0,u_0]}(U)|V = v], \qquad y = v,$$

para $u_0, v_0 \in [0, 1]$ fijos. Las curvas de regresión de $\mathbb{1}_{[0,v_0]}(V)$ sobre U, y de $\mathbb{1}_{[0,u_0]}(U)$ sobre V, se presentan en la Figura 2.7, para algunos valores fijos de u_0 y v_0 y $\theta = 0.9$. Se puede observar que existe una relación de correspondencia entre las curvas de regresión descritas antes y las rectas $u = u_0$ y $v = v_0$ para algunos valores fijos de u_0 y v_0. La Figura 2.8 exhibe dicha correspondencia mediante el esquema de colores, donde cada curva y su recta asociada se presentan con el mismo color en ambos planos.

Igual que en el ejemplo anterior, las figuras 2.8a y 2.8b dan cuenta del efecto de la transformación g, que en un sentido asocia las variables aleatorias, mientras que en el otro las separa.

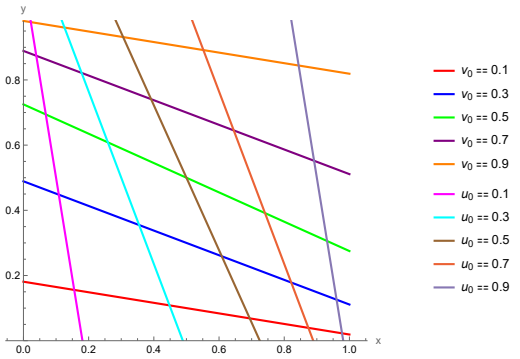


(a) Curvas de regresión de $\mathbb{1}_{[0,v_0]}(V)$ sobre U.

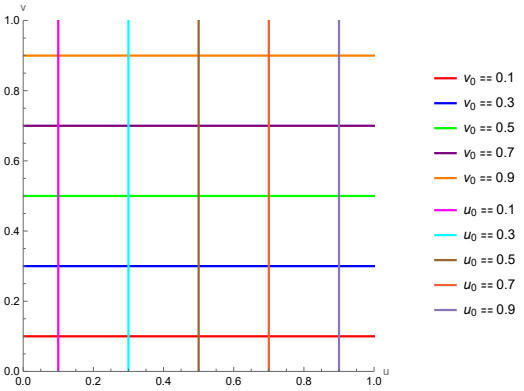


(b) Curvas de regresión de $\mathbb{1}_{[0,u_0]}(U)$ sobre V.

Figura 2.7: Curvas de regresión para la cópula $C(u,v) = uv + \theta uv(1-u)(1-v)$, con $\theta = 0.9$.



(a) Curvas de regresión.



(b) Rectas $u = u_0$ y $v = v_0$.

Figura 2.8: Correspondencia entre las curvas de regresión y las rectas $u = u_0$ y $v = v_0$ para algunos valores fijos de u_0 y v_0.

2.4. Algunas familias especiales de funciones cópula

Existen numerosas funciones cópula diferentes. Muchas de ellas pertenecen a familias paramétricas, donde cada cópula está indexada por uno o más parámetros, comúnmente denominados **parámetros de dependencia**. Este término se debe a que, en general, estos parámetros caracterizan el grado y, en algunos casos, la estructura de dependencia entre las variables del vector aleatorio (X, Y). La interpretación específica de estos parámetros y los valores que pueden tomar dependen de la familia de cópulas considerada.

A continuación se presentan algunas familias paramétricas de funciones cópula ampliamente utilizadas y bien estudiadas. La información expuesta se basa principalmente en Flores de la Fuente [9], Nelsen [21] y Tovar et al. [24].

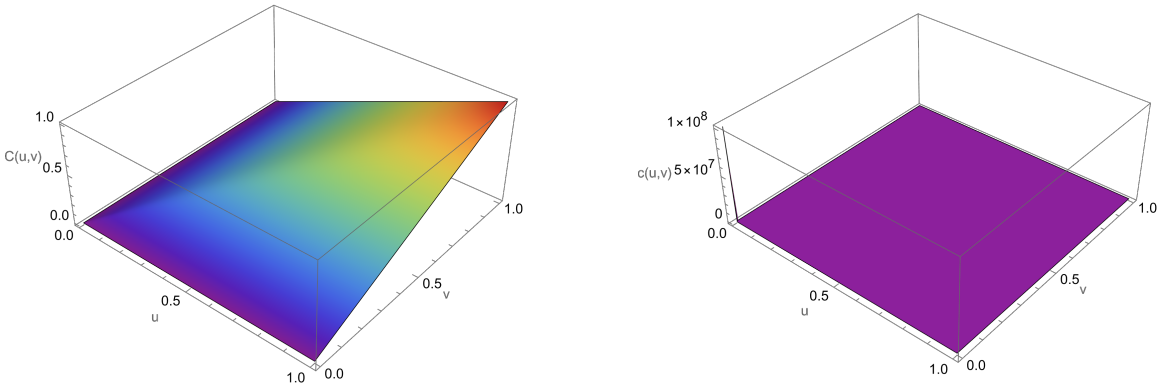
1. **Cópula de Clayton**
   Esta cópula está dada por la siguiente función:

$$C_\theta(u,v) = [\max\{u^{-\theta} + v^{-\theta} - 1, 0\}]^{-1/\theta}, \quad \theta \in [-1,\infty) \setminus \{0\}.$$

   Su función de densidad asociada es:

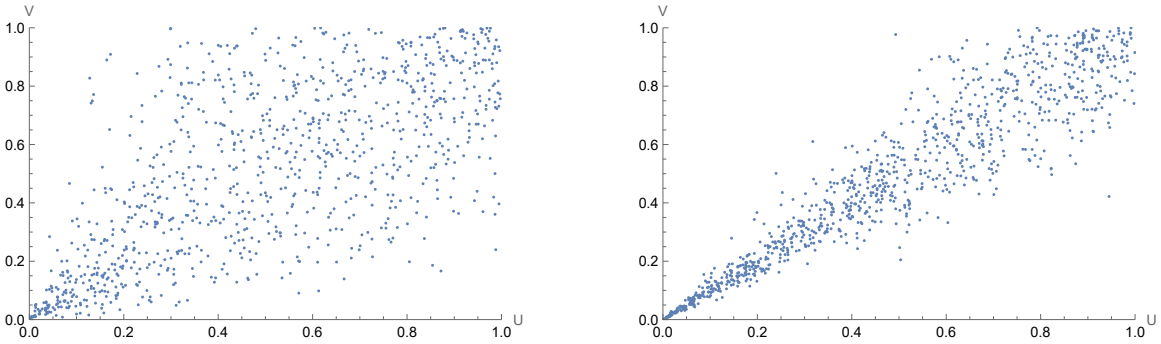
$$c_\theta(u,v) = (1+\theta)(uv)^{-(\theta+1)}(u^{-\theta} + v^{-\theta} - 1)^{-2-\frac{1}{\theta}}.$$

   En la Figura 2.9 se muestran los gráficos de estas funciones.



(a) Función de distribución. **(b)** Función de densidad.

Figura 2.9: Distribución y densidad de la cópula Clayton con $\theta = 8$.



(a) $\theta = 2$. **(b)** $\theta = 8$.

Figura 2.10: Gráficos de dispersión para 1000 datos simulados de la cópula Clayton.

La cópula de Clayton se caracteriza por modelar pares de observaciones asimétricas con una fuerte dependencia en la cola izquierda, es decir, que presentan mayor dependencia para valores cercanos a cero. Cuando $\theta \to \infty$ esta cópula es igual a $C_{\text{sup}}(u,v) = \min\{u,v\}$, y cuando $\theta \to -1$ es igual a $C_{\text{inf}}(u,v) = \max\{u+v-1, 0\}$; en estos casos existe dependencia perfecta entre las variables (dada explícitamente por una función). De otro lado, cuando

$\theta \to 0$, esta cópula tiende a la independencia. La Figura 2.10 muestra dos gráficos de dispersión para 1000 datos simulados de esta cópula, usando $\theta = 2$ y $\theta = 8$.

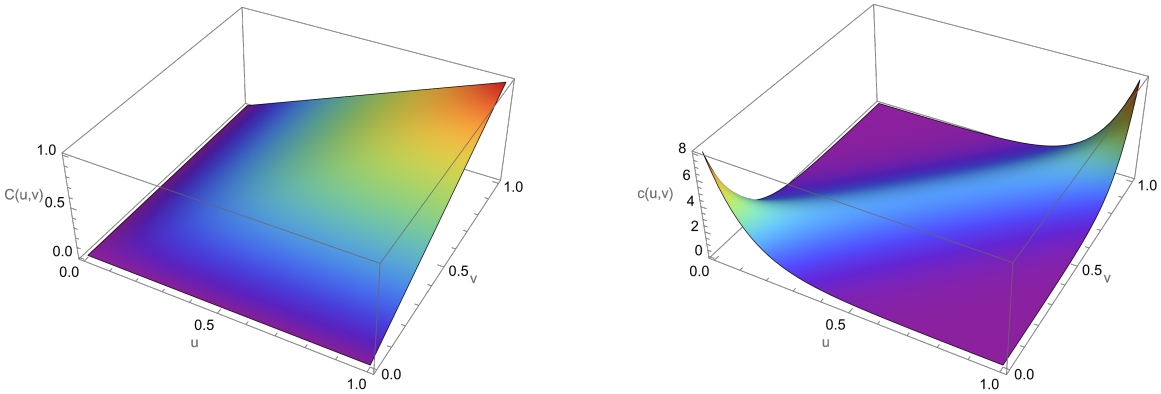
2. **Cópula de Frank**
   Su función está dada por:

$$C_\theta(u,v) = -\frac{1}{\theta}\ln\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right), \quad \theta \in (-\infty,\infty)\setminus\{0\},$$

   y su función de densidad es:

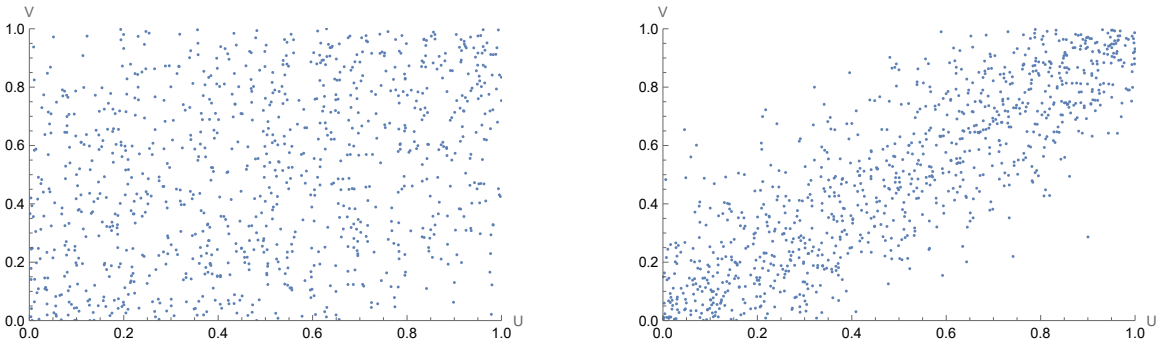
$$c_\theta(u,v) = \frac{-\theta(e^{-\theta}-1)e^{-\theta(u+v)}}{((e^{-\theta u}-1)(e^{-\theta v}-1)+e^{-\theta}-1)^2}.$$

   Los gráficos de estas funciones se presentan en la Figura 2.11.



(a) Función de distribución. **(b)** Función de densidad.

Figura 2.11: Distribución y densidad de la cópula Frank con $\theta = 8$.



(a) $\theta = 2$. **(b)** $\theta = 9$.

Figura 2.12: Gráficos de dispersión para 1000 datos simulados de la cópula Frank.

La cópula de Frank es apropiada para modelar dependencias débiles con tendencia lineal creciente. Cuando $\theta \to -\infty$ esta cópula es igual a $C_{\text{inf}}(u,v)$, y cuando

$\theta \to \infty$ es igual a $C_{\sup}(u, v)$; en estos casos se tiene dependencia funcional. Por su parte, cuando $\theta \to 0$, esta cópula tiende a la independencia. La Figura 2.12 presenta los gráficos de 1000 datos simulados de esta cópula, con $\theta = 2$ y $\theta = 9$.

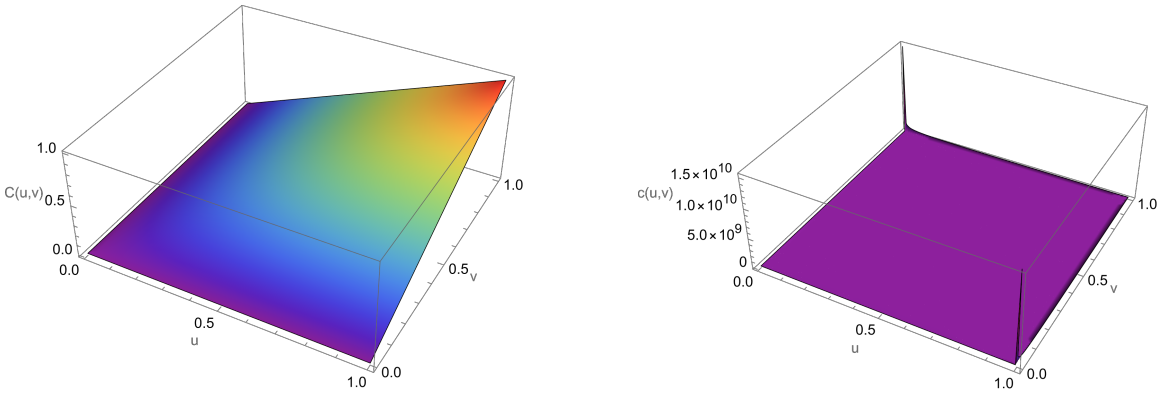
3. **Cópula de Gumbel-Hougaard**
   La forma funcional de esta cópula es:

$$C_\theta(u, v) = e^{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}}, \quad \theta \geq 1.$$

   Su función de densidad está dada por:

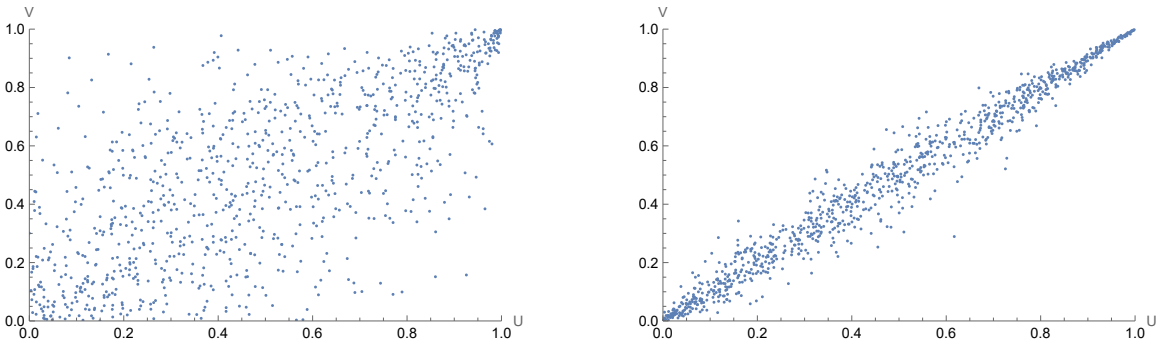
$$c_\theta(u, v) = \frac{C_\theta(u, v)}{uv} \frac{[(-\ln u)(-\ln v)]^{\theta-1}}{[(-\ln u)^\theta + (-\ln v)^\theta]^{2-\frac{1}{\theta}}} \left[((-\ln u)^\theta + (-\ln v)^\theta)^{\frac{1}{\theta}} + \theta - 1\right].$$

   En la Figura 2.13 se ilustran gráficamente las dos funciones anteriores.



(a) Función de distribución. **(b)** Función de densidad.

Figura 2.13: Distribución y densidad de la cópula Gumbel-Hougaard con $\theta = 8$.



(a) $\theta = 2$. **(b)** $\theta = 10$.

Figura 2.14: Gráficos de dispersión para 1000 datos simulados de la cópula Gumbel-Hougaard.

Esta cópula, introducida por Gumbel y Hougaard, es conveniente para modelar pares de observaciones con una dependencia fuerte en la cola superior y débil

en la cola inferior, es decir, cuando la correlación es mayor para valores altos (cercanos a uno) que para valores bajos (cercanos a cero). Como casos especiales, se tiene que esta cópula es igual a la independencia cuando $\theta = 1$, y es igual a $C_{\text{sup}}(u, v)$ cuando $\theta \to \infty$. La Figura 2.14 presenta dos gráficos de dispersión para 1000 datos simulados de esta cópula, con parámetros $\theta = 2$ y $\theta = 10$.

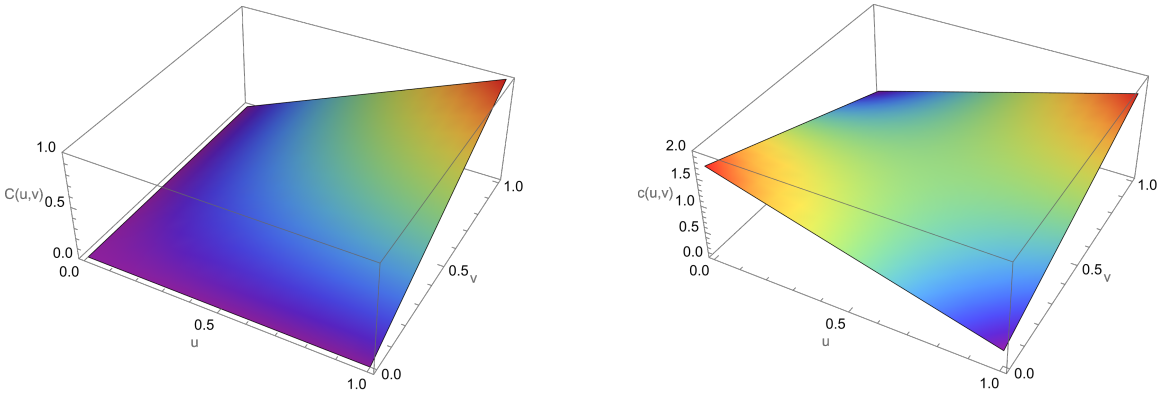
4. **Cópula de Farlie-Gumbel-Morgenstern (FGM)**
   La función de distribución de esta cópula es:

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

   mientras que su función de densidad asociada es:

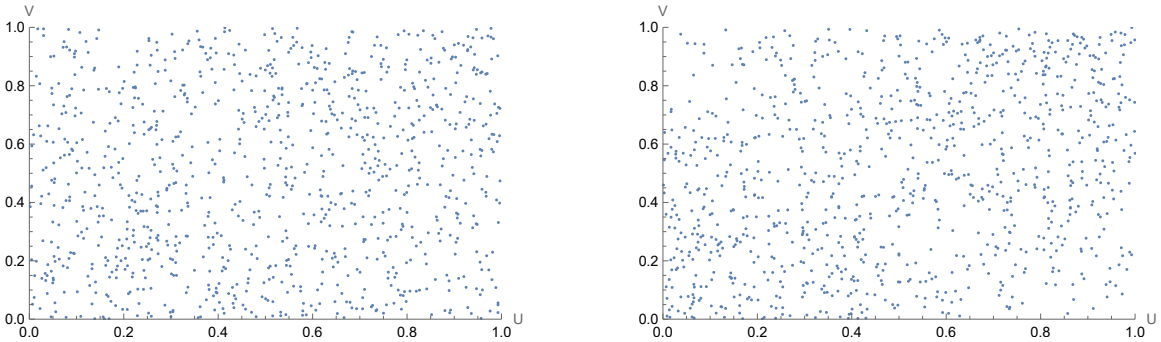
$$c_\theta(u, v) = 1 + \theta(1 - 2u)(1 - 2v).$$

   En la Figura 2.15 se presentan las gráficas de las dos funciones anteriores.



(a) Función de distribución. **(b)** Función de densidad.

Figura 2.15: Distribución y densidad de la cópula FGM con $\theta = 0.7$.



(a) $\theta = 0.4$. **(b)** $\theta = 0.9$.

Figura 2.16: Gráficos de dispersión para 1000 datos simulados de la cópula FGM.

Esta cópula es especialmente utilizada para modelar pares de observaciones con dependencias muy débiles y no lineales. Como caso especial se tiene que esta

cópula es igual a la de independencia cuando $\theta = 0$. En la Figura 2.16 se ilustran 1000 datos simulados de esta cópula para valores $\theta = 0.4$ y $\theta = 0.9$.

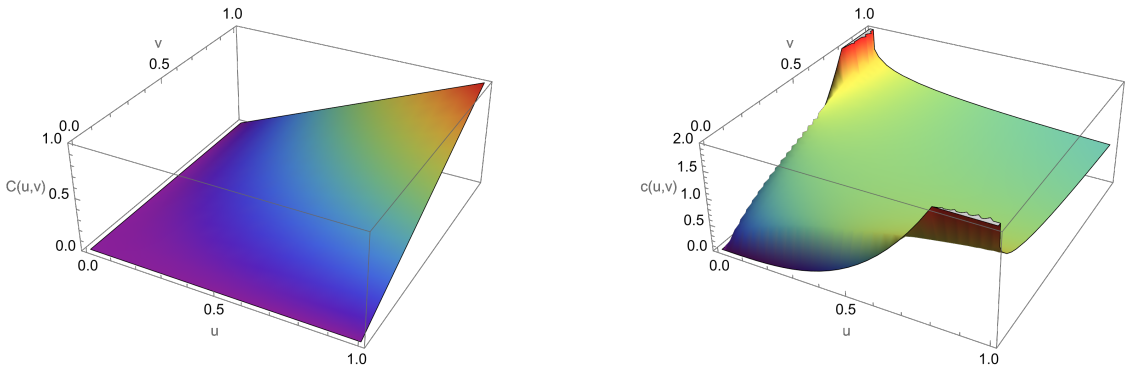
5. **Cópula de Gumbel-Barnett**
   Su forma analítica está dada por:

$$C_\theta(u, v) = uve^{[-\theta \ln u \ln v]}, \quad 0 < \theta \leq 1.$$

   Asimismo, su función de densidad asociada es la siguiente:

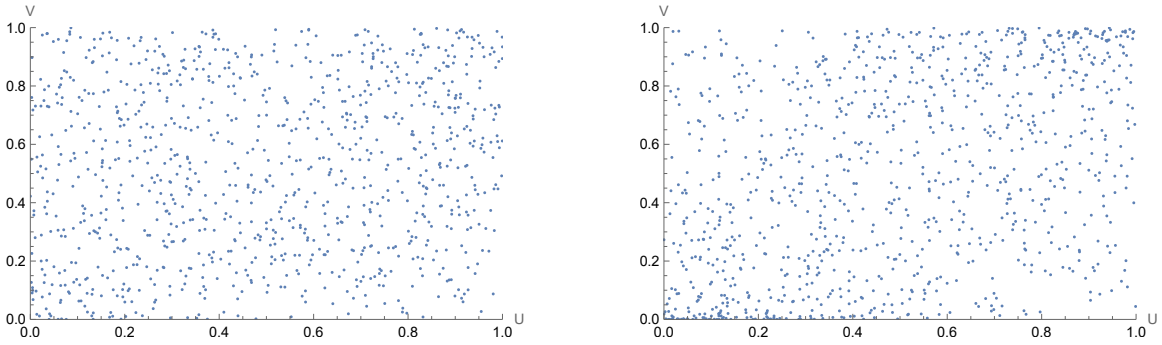
$$c_\theta(u, v) = [1 - \theta - \theta \ln u - \theta \ln v + \theta^2 \ln u \ln v]e^{(-\theta \ln u \ln v)}.$$

   En la Figura 2.17 se muestran los gráficos de las dos funciones anteriores.



(a) Función de distribución. **(b)** Función de densidad.

Figura 2.17: Gráficos de la distribución y densidad de la cópula Gumbel-Barnett con $\theta = 0.2$.



(a) $\theta = 0.2$. **(b)** $\theta = 0.9$.

Figura 2.18: Gráficos de dispersión para 1000 datos simulados de la cópula Gumbel-Barnett.

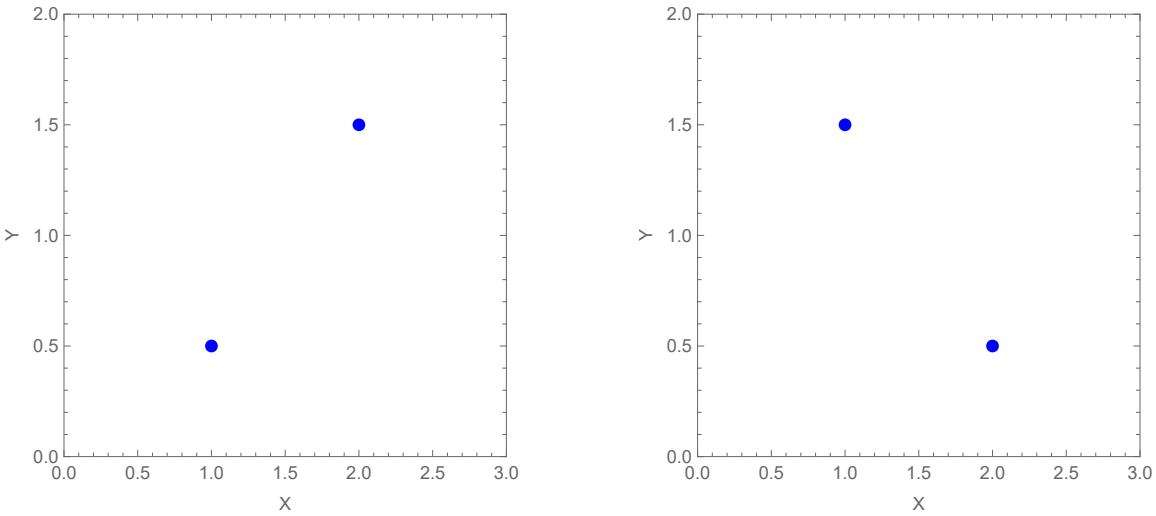
La cópula de Gumbel y Barnett también es conveniente para modelar dependencias débiles no lineales. En particular, se tiene que cuando $\theta = 0$, esta cópula se convierte en la cópula producto, lo que implica que las variables son independientes. En la Figura 2.18 se muestran dos gráficos de dispersión que ilustran 1000 observaciones simuladas de esta cópula con parámetros $\theta = 0.2$ y $\theta = 0.9$.

2.5. Medidas de asociación

Una de las principales utilidades de las funciones cópula tiene que ver con su capacidad para determinar la estructura de dependencia entre las variables que componen un vector aleatorio. Como ya se sabe de la ecuación (2.2), la función de distribución de un vector aleatorio (X, Y) se puede escribir en términos de sus distribuciones marginales y una función cópula. Esto significa que la función de distribución bivariada contiene dos tipos de información acerca del vector aleatorio: sobre el comportamiento de cada variable considerada de manera aislada, y sobre el comportamiento conjunto de las variables (Latorre Lloréns [17]). En este sentido se puede decir que la cópula "captura" la estructura de dependencia que hay entre las variables X y Y.

A partir del conocimiento de la función cópula, se puede intentar medir el grado de dependencia que existe entre las variables; una primera herramienta para abarcar este problema es el coeficiente de correlación de Pearson (Definición 1.1.13). Sin embargo, este coeficiente tiene la limitación de que únicamente es útil cuando las variables aleatorias en cuestión tienen distribución normal, y solo para medir dependencia lineal. Estas restricciones suponen la necesidad de contar con herramientas más robustas que permitan medir la dependencia de manera mucho más general. Por ejemplo, otro tipo de dependencia importante es la **concordancia**, concepto que será introducido en esta sección, al igual que los coeficientes τ de Kendall y ρ de Spearman, los cuales serán esenciales para medir este tipo de relaciones.

Definición 2.5.1. *Sean (x_i, y_i) y (x_j, y_j) dos observaciones de un vector aleatorio (X, Y), donde X y Y son variables aleatorias continuas. Se dice que (x_i, y_i) y (x_j, y_j) son **concordantes** si $x_i < x_j$ y $y_i < y_j$, o si $x_i > x_j$ y $y_i > y_j$. De otro lado, si $x_i < x_j$ y $y_i > y_j$, o $x_i > x_j$ y $y_i < y_j$, se dice que (x_i, y_i) y (x_j, y_j) son **discordantes**.*



(a) Observaciones concordantes. **(b)** Observaciones discordantes.

Figura 2.19: Ejemplo de observaciones concordantes y discordantes.

En otras palabras, se dice que dos variables aleatorias son concordantes si valores grandes (pequeños) de una de ellas tienden a asociarse con valores grandes (pequeños) de la otra. En la Figura 2.19 se ilustran dos observaciones concordantes y dos discordantes.

Observación 2.5.1. *Una forma alternativa de definir la concordancia es la siguiente:* (x_i, y_i) *y* (x_j, y_j) *son concordantes si* $(x_i - x_j)(y_i - y_j) > 0$ *y discordantes si* $(x_i - x_j)(y_i - y_j) < 0$.

Definición 2.5.2. *Una función* $K : \mathcal{D} \to [-1, 1]$, *donde* $\mathcal{D}$ *es el conjunto de vectores aleatorios* (X, Y) *con* X *y* Y *variables aleatorias continuas, es una* ***medida de concordancia****, si:*

1. $K_{(X,Y)} = K_{(Y,X)}$.
2. $K_{(X,X)} = 1$ *y* $K_{(X,-X)} = -1$.
3. *Si* X *y* Y *son variables aleatorias independientes entonces* $K_{(X,Y)} = K_{\Pi} = 0$.
4. $K_{(X,Y)} = -K_{(-X,Y)} = -K_{(X,-Y)}$.
5. *Si* $C_1(u, v) \leq C_2(u, v)$ *para todo* $(u, v) \in I$ *entonces* $K_{C_1} \leq K_{C_2}$.
6. *Si* $\{C_n\}_{n\in\mathbb{N}}$ *es una sucesión de cópulas y* $\lim_{n\to\infty} C_n(u, v) = C(u, v)$ *entonces* $\lim_{n\to\infty} K_{C_n} = K_C$.

Teorema 2.5.1. *Si* K *es una medida de concordancia entonces se tiene lo siguiente:*

1. *Si* $Y = \alpha(X)$, *donde* $\alpha(\cdot)$ *es una función creciente (casi seguramente), entonces* $K_{(X,Y)} = K_{(X,\alpha(X))} = 1$.
2. *Si* $Y = \alpha(X)$, *donde* $\alpha(\cdot)$ *es una función decreciente (casi seguramente), entonces* $K_{(X,Y)} = K_{(X,\alpha(X))} = -1$.
3. *Si* $\alpha(\cdot)$ *y* $\beta(\cdot)$ *son ambas estrictamente crecientes o ambas estrictamente decrecientes (casi seguramente), entonces* $K_{(X,Y)} = K_{(\alpha(X),\beta(Y))}$.

Demostración.

1. Note que $C_{(X,Y)}(u, v) = C_{(X,\alpha(X))}(u, v) = C_{(X,X)}(u, v)$, por el literal *1* de la proposición 2.2.2. Entonces $K_{C_{(X,Y)}} = K_{C_{(X,X)}} = 1$.
2. Por el literal anterior, $K_{(X,-\alpha(X))} = K_{C_{(X,X)}}$. Por el punto *4* de la definición 2.5.2, $K_{(X,-\alpha(X))} = -K_{(X,\alpha(X))}$. De aquí que $K_{(X,\alpha(X))} = -K_{(X,X)} = -1$.
3. Se sigue de los dos literales anteriores.

□

Dos medidas de concordancia bien conocidas son el τ de Kendall y el ρ de Spearman, las cuales serán introducidas a continuación.

Definición 2.5.3. *Sean (X_1, Y_1) y (X_2, Y_2) vectores aleatorios independientes e idénticamente distribuidos con función de distribución común $F_{(X,Y)}(\cdot,\cdot)$. El* ***coeficiente τ de Kendall*** *(en su versión poblacional) es:*

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

Note que la definición anterior corresponde a la probabilidad de concordancia menos la probabilidad de discordancia. Teniendo esto en cuenta, se puede definir una versión muestral para el τ de Kendall, es decir, calculada directamente a partir de una muestra aleatoria. Supongamos que $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ es un conjunto de n observaciones de un vector aleatorio (X, Y), donde X y Y son variables aleatorias continuas. De este conjunto podemos formar $\binom{n}{2}$ pares de observaciones distintas; cada par puede ser concordante o discordante. Denotemos por c el número de parejas concordantes y por d el número de parejas discordantes. Entonces, la probabilidad de que una pareja sea concordante es $\dfrac{c}{c+d} = \dfrac{c}{\binom{n}{2}}$. Equivalentemente, $\dfrac{d}{c+d} = \dfrac{d}{\binom{n}{2}}$ es la probabilidad de que un par sea discordante. Por lo tanto, la probabilidad de concordancia menos la probabilidad de discordancia, expresada en estos términos, es:

$$\hat{\tau} \equiv t = \frac{c-d}{c+d} = \frac{c+d}{\binom{n}{2}}, \tag{2.5}$$

el cual es la versión muestral (o empírica) del τ de Kendall.

Teorema 2.5.2. *Sean (X_1, Y_1) y (X_2, Y_2) vectores aleatorios independientes tales que X_1 y X_2 son variables aleatorias i.i.d. continuas y Y_1 y Y_2 también lo son. Sean $C_1(\cdot,\cdot)$ y $C_2(\cdot,\cdot)$ las funciones cópula asociadas a (X_1, Y_1) y (X_2, Y_2), respectivamente. Sea*

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

Entonces,

$$\begin{aligned} Q \equiv Q(C_1, C_2) &= 4 \iint_{I^2} C_2(u,v)\, dC_1(u,v) - 1, \\ &= 4 \iint_{I^2} C_1(u,v)\, dC_2(u,v) - 1. \end{aligned}$$

La demostración de este teorema se puede consultar en Nelsen [21] (Teorema 5.1.1, p. 159).

De acuerdo a lo anterior, se tiene una forma de calcular el coeficiente τ de Kendall poblacional en términos de la función cópula. Más concretamente,

Teorema 2.5.3. *Sean X y Y variables aleatorias continuas con función cópula asociada $C_{(X,Y)}(u,v)$. Entonces el coeficiente τ de Kendall, denotado por $\tau_{(X,Y)}$, está dado por*

$$\tau_{(X,Y)} = Q(C, C) = 4 \iint_{I^2} C(u,v)\, dC(u,v) - 1. \tag{2.6}$$

Observación 2.5.2. *Como $C(u,v)$, $(u,v) \in I^2$, es la función de distribución de un vector aleatorio (U,V) tal que $U \sim Unif([0,1])$ y $V \sim Unif([0,1])$, entonces*

$$\iint_{I^2} C(u,v)dC(u,v) = E[C(U,V)],$$

y por tanto $\tau_C = 4E[C(U,V)] - 1$.

Para distintas familias de funciones cópula paramétricas (como las descritas en la sección 2.4) es posible obtener una expresión del τ de Kendall poblacional en términos de su parámetro de dependencia, haciendo uso del Teorema 2.5.3. Veamos esto a través del siguiente ejemplo.

Ejemplo 2.5.1. *Consideremos la familia FGM con parámetro θ. Sabemos que $dC_\theta(u,v) = c_\theta(u,v)dudv$ (ver sección 2.4). Entonces, calculando la integral de la ecuación (2.6), se sigue que*

$$\begin{aligned}\iint_{I^2} C_\theta(u,v)dC_\theta(u,v) &= \int_0^1\int_0^1 [uv + \theta uv(1-u)(1-v)][1+\theta(1-2u)(1-2v)]dudv, \\ &= \frac{1}{4} + \frac{\theta}{18}.\end{aligned}$$

Luego, $\tau_\theta = 4\left(\frac{1}{4} + \frac{\theta}{18}\right) - 1 = \frac{2\theta}{9}$.

Otra medida de concordancia conocida es el coeficiente ρ de Spearman.

Definición 2.5.4. *Sean (X_1,Y_1), (X_2,Y_2) y (X_3,Y_3) vectores aleatorios i.i.d. El* ***coeficiente ρ de Spearman*** *(en su versión poblacional) es:*

$$\rho_{(X,Y)} := 3\left(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]\right).$$

Observación 2.5.3. *El vector aleatorio (X_2,Y_3) tiene función de distribución $F_{(X_2,Y_3)}(x_2,y_3) = F_X(x_2)F_Y(y_3)$.*

De manera similar al coeficiente τ de Kendall, se tiene una forma de calcular el ρ de Spearman en términos de la función cópula; esto se hace a través de la función Q introducida en el Teorema 2.5.2.

Teorema 2.5.4. *Sean X y Y variables aleatorias con función cópula asociada $C_{(X,Y)} \equiv C$. Entonces,*

$$\begin{aligned}\rho_{(X,Y)} &= 3Q(C,\Pi), \\ &= 12\iint_{I^2} uvdC(u,v) - 3, \\ &= 12\iint_{I^2} C(u,v)dudv - 3. \qquad (2.7)\end{aligned}$$

También es posible definir una versión muestral para el ρ de Spearman. Supongamos que $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ es un conjunto de n observaciones de un vector aleatorio continuo (X, Y); entonces el ρ de Spearman muestral, denotado por r o $\hat{\rho}$, está dado por

$$\hat{\rho} \equiv r = \frac{12}{n(n+1)(n-1)} \sum_{i=1}^{n} R_i S_i - 3\frac{n+1}{n-1}, \tag{2.8}$$

donde $R_i = \sum_{j=1}^{n} \mathbb{1}_{(-\infty, x_i]}(x_j)$ y $S_i = \sum_{j=1}^{n} \mathbb{1}_{(-\infty, y_i]}(y_j)$.

Observación 2.5.4. *Sean X y Y variables aleatorias continuas con función de distribución $F_X(\cdot)$ y $F_Y(\cdot)$, respectivamente. Por la transformada inversa (Teorema 1.1.1), las variables $U := F_X(X)$ y $V := F_Y(Y)$ tienen distribución uniforme en $(0, 1)$ y su función de distribución conjunta es la cópula $C_{(X,Y)} \equiv C$. Por tanto $E(U) = E(V) = \frac{1}{2}$ y $Var(U) = Var(V) = \frac{1}{12}$. Así, el coeficiente ρ de Spearman es:*

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

el cual es el mismo coeficiente de correlación de Pearson introducido en la Definición 1.1.13.

Haciendo uso del Teorema 2.5.4, es posible obtener expresiones para el ρ de Spearman poblacional en términos del parámetro de dependencia, cuando se trabaja con familias de cópulas paramétricas. Veamos el siguiente ejemplo:

Ejemplo 2.5.2. *Consideremos la familia de funciones cópula FGM con parámetro de dependencia θ (ver sección 2.4). Al calcular la integral de la ecuación (2.7) se obtiene que:*

$$\begin{aligned}
\iint_{I^2} C_\theta(u, v) du dv &= \int_0^1 \int_0^1 [uv + \theta uv(1-u)(1-v)] du dv, \\
&= \frac{1}{4} + \frac{\theta}{36}.
\end{aligned}$$

Luego, $\tau_\theta = 12\left(\frac{1}{4} + \frac{\theta}{36}\right) - 3 = \frac{\theta}{3}$.

En la siguiente tabla, se presentan las expresiones del τ de Kendall y el ρ de Spearman en términos del parámetro de dependencia para las familias de cópulas paramétricas presentadas en la sección 2.4.

Función cópula	τ **de Kendall**	ρ **de Spearman**
Clayton	$\frac{\theta}{\theta+2}$	*
Frank	$1 - \frac{4}{\theta} + \frac{4D_1(\theta)}{\theta}$	$1 - \frac{12}{\theta}[D_1(\theta) - D_2(\theta)]$
Gumbel-Hougaard	$\frac{\theta-1}{\theta}$	*
Gumbel-Barnett	$-e^{-2/\theta}\Gamma(0, \frac{2}{\theta})$	$\frac{12}{\theta}e^{4/\theta}\Gamma(0, \frac{4}{\theta}) - 3$
FGM	$\frac{2\theta}{9}$	$\frac{\theta}{3}$

Tabla 2.1: Expresiones para los coeficientes τ de Kendall y ρ de Spearman en términos del parámetro de dependencia de algunas funciones cópula.

Observación 2.5.5. *En la Tabla 2.1, la función* $D_k(x)$ *es la* ***función de Debye****, que se define, para cualquier entero positivo* k*, por*

$$D_k(x) := \frac{k}{x^k}\int_0^x \frac{t^k}{e^t - 1}dt.$$

Por su parte, la función $\Gamma(a, z)$ *es la* ***función Gamma incompleta****, dada por*

$$\Gamma(a, x) := \int_0^x t^{a-1}e^{-t}dt. \tag{2.9}$$

*El símbolo * en la tabla anterior significa que no se cuenta con una expresión explícita del coeficiente para la cópula respectiva.*

Capítulo 3

Inferencia estadística para funciones cópula

En la práctica, cuando se trabaja con datos reales extraidos de una cierta población, es usual que no se conozca la función de distribución de la variable aleatoria que es objeto de estudio. Lo mismo aplica para datos bivariados y su respectiva función cópula. En ese sentido, el problema principal que se desea tratar en este capítulo es el siguiente: "Dada una muestra aleatoria $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ de un vector aleatorio (X, Y) continuo absolutamente, se desea conocer la función cópula asociada a (X, Y)". Para abordarlo, se pueden considerar dos enfoques: uno *paramétrico* y uno *no paramétrico*. El enfoque paramétrico se emplea cuando se conoce la función de distribución del vector aleatorio (X, Y), digamos $F_{(X,Y)}(\cdot, \cdot)$. En ese caso, se procede de la siguiente manera: en primer lugar, se deducen las funciones las funciones de distribución marginales, $F_X(\cdot)$ y $F_Y(\cdot)$; luego se obtienen sus inversas de acuerdo a la ecuación (1.1) y se aplica el Teorema de Sklar (Teorema 2.2.1), llegando a que $C(u, v) = F_{(X,Y)}(F_X^{-1}(u), F_Y^{-1}(v))$, que resulta ser la función cópula buscada. En el siguiente diagrama se resume el proceso anterior.

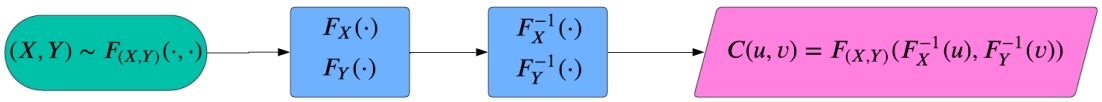


Figura 3.1: Enfoque paramétrico para obtener la función cópula asociada a un vector aleatorio.

Debido a que no siempre se conoce la función de distribución del vector aleatorio (X, Y), se hace necesario disponer de un enfoque no paramétrico, es decir, que esté basado únicamente en la muestra aleatoria. En este capítulo se presentarán las herramientas necesarias para abordar dicho enfoque, que será precisado en el siguiente capítulo a través de una prueba de hipótesis estadística.

En primer lugar, en la sección 3.1 se introducirá la definición de cópula empírica, la cual se calcula a partir de una muestra aleatoria bivariada y resulta ser un estimador de la función cópula exacta. Posteriormente, en la sección 3.2 se presentarán dos métodos para estimar el parámetro de una función cópula, en caso que lo tenga. Además, se expondrá un coeficiente basado en funciones cópula, presentado en Anjos y Kolev [1] y en Ledwina [18], el cual permite determinar el grado de dependencia entre dos variables aleatorias. En adelante, por facilidad, este coeficiente se denominará *Medida de dependencia de Ledwina*, y será detallado en la sección 3.3, donde se evidencia la forma en que se construye y cómo este exhibe un cambio de métrica para el coeficiente de correlación de Pearson, en el sentido de expresarlo como la diferencia entre la cópula y la cópula de independencia. Finalmente, en la sección 3.4 se presentará un indicador desarrollado por Tovar et al. [24], el cual está basado en la medida de Ledwina y que se utiliza para seleccionar una función cópula apropiada para modelar un conjunto de datos empíricos. La elección de dicha cópula se hace entre un conjunto de candidatas, que se establecen a partir de un estudio visual de la dispersión de los datos que se tienen.

3.1. La cópula empírica

Recordemos que en la sección 1.1.7 se dio la definición de Función de Distribución Empírica (Definición 1.1.16), la cual permite estimar la función de distribución de una población a partir de una muestra aleatoria. En el caso bivariado se pueden definir de manera análoga las funciones de distribución marginales empíricas, tal como se establece en Swanepoel y Allison [23], p. 1732.

Definición 3.1.1. *Sea $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ una muestra aleatoria bivariada. Las **funciones de distribución marginales empíricas**, denotadas por $F_X^n(\cdot)$ y $F_Y^n(\cdot)$, se definen como*

$$F_X^n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(X_i) \quad y \quad F_Y^n(y) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,y]}(Y_i),$$

para $x, y \in \mathbb{R}$.

En la práctica, los vectores aleatorios (X_i, Y_i), $i = 1, \ldots, n$, pueden tomar valores en todo $\mathbb{R}^2$. Por esta razón, se hace necesario transformarlos para obtener unas pseudo-observaciones (U_i, V_i) que solo tomen valores en $(0, 1)^2$; esto se requiere para definir más adelante la cópula empírica. Dichas pseudo-observaciones se pueden obtener usando la Transformada Inversa (Teorema 1.1.1), en caso de que se conozcan las distribuciones marginales; cuando este no es el caso, se aplica la siguiente transformación basada en las distribuciones empíricas:

$$U_i := \frac{n}{n+1} F_X^n(X_i) = \frac{1}{n+1} \sum_{j=1}^{n} \mathbb{1}_{(-\infty, X_i]}(X_j), \tag{3.1}$$

$$V_i := \frac{n}{n+1} F_Y^n(Y_i) = \frac{1}{n+1} \sum_{j=1}^{n} \mathbb{1}_{(-\infty, Y_i]}(Y_j). \tag{3.2}$$

La modificación $\frac{n}{n+1}$ que se aplica a las funciones de distribución marginales empíricas permite garantizar que U_i y V_i no tomen el valor de 1, es decir que (U_i, V_i) no se ubique en la frontera del cuadrado unitario. De esta manera se pueden evitar posibles indeterminaciones al momento de realizar operaciones con dichos valores, como se verá más adelante (ver p. 558 de Gijbels et al. [14]).

De acuerdo a lo anterior, las pseudo-observaciones (U_i, V_i) se pueden considerar como una muestra del vector aleatorio $(F_X(X), F_Y(Y))$, cuya función de distribución bivariada es una cópula, ya que $F_X(X)$ y $F_Y(Y)$ son variables aleatorias con distribución uniforme en $[0, 1]$, según la Transformada Inversa.

Con las consideraciones anteriores, se tienen los elementos necesarios para definir la cópula empírica.

Definición 3.1.2. *Sea $(U_1, V_1), (U_2, V_2), \ldots, (U_n, V_n)$ una muestra aleatoria del vector aleatorio (U, V), donde U y V son variables aleatorias con función de distribución uniforme en $[0, 1]$. La **cópula empírica**, denotada por $C_n(\cdot, \cdot)$, se define como*

$$C_n(u, v) := \frac{1}{n} \sum_{i=1}^{n} (\mathbb{1}_{(-\infty, u]}(U_i))(\mathbb{1}_{(-\infty, v]}(V_i)), \qquad (u, v) \in (0, 1)^2.$$

Observación 3.1.1. *Note que para construir la cópula empírica se requiere de una muestra aleatoria (U_i, V_i), $i = 1, \ldots, n$, donde U_i y V_i tienen distribución uniforme en $[0, 1]$, es decir, que solo toman valores entre 0 y 1. Esto muestra la necesidad de obtener las pseudo-observaciones cuando se parte de una muestra aleatoria bivariada cualquiera.*

La definición de cópula empírica dada en Ledwina [18] es la siguiente:

Definición 3.1.3. *Sea $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ una muestra aleatoria de una distribución continua H. Además, sea R_i el rango de X_i en la muestra $X_1, \ldots, X_n$ y S_i el rango de Y_i en la muestra $Y_1, \ldots, Y_n$. Entonces la cópula empírica se define por*

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} \left[\mathbb{1}_{(-\infty, u]} \left(\frac{R_i}{n+1} \right) \cdot \mathbb{1}_{(-\infty, v]} \left(\frac{S_i}{n+1} \right) \right], \quad (u, v) \in [0, 1]^2. \tag{3.3}$$

Esta definición es equivalente a la que se dio líneas arriba. Para comprobarlo, veamos qué significa el rango de una muestra aleatoria.

Definición 3.1.4. *Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución $F_X(\cdot)$ y sean $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ los correspondientes estadísticos de orden. Se dice que X_i tiene rango R_i entre $X_1, X_2, \ldots, X_n$ si $X_i = X_{(R_i)}$.*

Ejemplo 3.1.1. *Consideremos la siguiente muestra aleatoria:*

$$X_1 = 7,\ X_2 = 12,\ X_3 = 1,\ X_4 = 2,\ X_5 = 9.$$

Al ordenar estos valores se obtiene:

$$X_{(1)} = X_3 = 1;\ X_{(2)} = X_4 = 2;\ X_{(3)} = X_1 = 7;\ X_{(4)} = X_5 = 9;\ X_{(5)} = X_2 = 12.$$

Por lo tanto, los rangos son:

$$R_1 = 3,\ R_2 = 5,\ R_3 = 1,\ R_4 = 2,\ R_5 = 4.$$

Note que R_i representa la cantidad de elementos de la muestra aleatoria que son menores o iguales al valor X_i. En otros términos,

$$R_i = \sum_{j=1}^{n} \mathbb{1}_{(-\infty, x_i]}(X_j).$$

Esto significa que

$$\frac{R_i}{n+1} = \frac{1}{n+1}\sum_{j=1}^{n} \mathbb{1}_{(-\infty, x_i]}(X_j) = \frac{n}{n+1}\left(\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{(-\infty, x_i]}(X_j)\right) = \frac{n}{n+1}F_X^n(X_i) = U_i,$$

y, análogamente,

$$\frac{S_i}{n+1} = \frac{n}{n+1}F_Y^n(Y_i) = V_i.$$

Por lo tanto, la ecuación (3.3) se convierte en

$$C_n(u, v) = \frac{1}{n}\sum_{i=1}^{n} (\mathbb{1}_{(-\infty, u]}(U_i))(\mathbb{1}_{(-\infty, v]}(V_i)),$$

que es la misma de la Definición 3.1.2.

Observación 3.1.2. *De acuerdo a la discusión anterior, en adelante se tomará como cópula empírica la presentada en la definición 3.1.2, aún cuando se haga referencia a los resultados de Ledwina [18].*

Ejemplo 3.1.2. *En la Tabla 3.1 se presentan diez observaciones de un vector aleatorio (X, Y) cuya función de distribución es $F_{(X,Y)}(x, y) = (\Phi(x))(\Phi(y))[1 + 0.6(1 - \Phi(x))(1 - \Phi(y))]$, donde $\Phi(\cdot)$ representa la función de distribución normal estándar. Estos valores fueron simulados usando Wolfram Mathematica.*

i	1	2	3	4	5	6	7	8	9	10
x_i	-0.11	0.71	-0.06	-0.82	0.62	-0.96	0.06	0.62	-0.24	-1.27
y_i	-0.41	1.95	-0.80	-0.36	2.89	-0.76	0.78	-1.13	0.59	0.39

Tabla 3.1: Conjunto de datos de ejemplo.

Con estos datos, vamos a construir la cópula empírica. Para ello, lo primero que se necesita es obtener las pseudo-observaciones, empleando las ecuaciones (3.1) y (3.2). Por ejemplo, calculemos u_1:

$$\begin{aligned} u_1 &= \frac{1}{11}[\mathbb{1}_{(-\infty,-0.11]}(-0.11) + \mathbb{1}_{(-\infty,-0.11]}(0.71) + \mathbb{1}_{(-\infty,-0.11]}(-0.06) \\ &\quad +\mathbb{1}_{(-\infty,-0.11]}(-0.82) + \mathbb{1}_{(-\infty,-0.11]}(0.62)] + \mathbb{1}_{(-\infty,-0.11]}(-0.96) \\ &\quad +\mathbb{1}_{(-\infty,-0.11]}(0.06) + \mathbb{1}_{(-\infty,-0.11]}(0.62) + \mathbb{1}_{(-\infty,-0.11]}(-0.24) \\ &\quad +\mathbb{1}_{(-\infty,-0.11]}(-1.27)], \\ &= \frac{1}{11}[1+0+0+1+0+1+0+0+1+1], \\ &= \frac{5}{11}. \end{aligned}$$

De manera similar se procede para los demás valores de u_i y v_i. Los resultados se registran en la Tabla 3.2.

i	1	2	3	4	5	6	7	8	9	10
u_i	5/11	10/11	6/11	3/11	8/11	2/11	7/11	9/11	4/11	1/11
v_i	4/11	9/11	2/11	5/11	10/11	3/11	8/11	1/11	7/11	6/11

Tabla 3.2: Pseudo-observaciones de los datos de la Tabla 3.1.

De acuerdo a lo anterior, se sigue que la cópula empírica es:

$$\begin{aligned} C_{10}(u,v) &= \frac{1}{10}[\mathbb{1}_{(-\infty,u]}(5/11)\,\mathbb{1}_{(-\infty,v]}(4/11) + \mathbb{1}_{(-\infty,u]}(10/11)\,\mathbb{1}_{(-\infty,v]}(9/11) \\ &\quad +\mathbb{1}_{(-\infty,u]}(6/11)\,\mathbb{1}_{(-\infty,v]}(2/11) + \mathbb{1}_{(-\infty,u]}(3/11)\,\mathbb{1}_{(-\infty,v]}(5/11) \\ &\quad +\mathbb{1}_{(-\infty,u]}(8/11)\,\mathbb{1}_{(-\infty,v]}(10/11) + \mathbb{1}_{(-\infty,u]}(2/11)\,\mathbb{1}_{(-\infty,v]}(3/11) \\ &\quad +\mathbb{1}_{(-\infty,u]}(7/11)\,\mathbb{1}_{(-\infty,v]}(8/11) + \mathbb{1}_{(-\infty,u]}(9/11)\,\mathbb{1}_{(-\infty,v]}(1/11) \\ &\quad +\mathbb{1}_{(-\infty,u]}(4/11)\,\mathbb{1}_{(-\infty,v]}(7/11) + \mathbb{1}_{(-\infty,u]}(1/11)\,\mathbb{1}_{(-\infty,v]}(6/11)]. \end{aligned}$$

En la Figura 3.2 se muestra un gráfico de la función anterior.

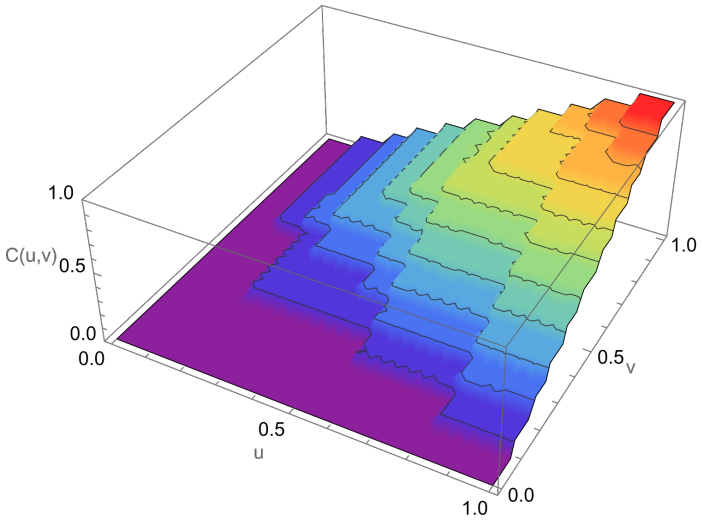


Figura 3.2: Cópula empírica del Ejemplo 3.1.2.

3.2. Estimación del parámetro de una función cópula

En la sección 1.2 se presentaron dos métodos para estimar los parámetros de una distribución de probabilidad a partir de una muestra aleatoria. De manera similar, se puede estimar el parámetro de dependencia de una función cópula a partir de un conjunto de observaciones bivariadas. A continuación se expondrán dos métodos conocidos para realizar este proceso, que han sido documentados en Flores de la Fuente [9] y Genest y Favre [10].

3.2.1. Método del τ de Kendall y el ρ de Spearman

Como se mencionó en el capítulo anterior, es posible tener expresiones para el τ de Kendall y el ρ de Spearman en términos del parámetro de dependencia de una función cópula. En la Tabla 2.1 se presentaron dichas expresiones para algunas cópulas. Asimismo, sabemos que existen versiones muestrales de dichos coeficientes, es decir, que estos se pueden estimar a partir de una muestra aleatoria. La idea de este método consiste en remplazar las versiones poblacionales de τ y ρ por sus versiones muestrales, de manera similar a como se hace en el método de momentos presentado en el capítulo 1. A continuación se describen los pasos para su implementación:

1. Identifique la expresión que relaciona el τ de Kendall o el ρ de Spearman con el parámetro θ (como ejemplo se tienen las expresiones dadas en la Tabla 2.1).

2. Despeje θ de tal manera que este quede expresado como una función de τ o ρ, es decir, $\theta = g(\tau)$ o $\theta = h(\rho)$, para algunas funciones g o h.

3. Calcule las versiones muestrales del τ de Kendall o el ρ de Spearman, $t = \hat{\tau}$ o $r = \hat{\rho}$ respectivamente.

4. Evalúe en t o r las funciones obtenidas en el paso 2 según corresponda, es decir $\tilde{\theta} = g(t)$ o $\tilde{\theta} = h(r)$.

5. $\tilde{\theta}$ es el valor estimado del parámetro θ.

Ilustraremos este método de estimación a través de los siguientes ejemplos.

Ejemplo 3.2.1. *Observe que el vector aleatorio del que fueron generados los datos de la Tabla 3.1 tiene como cópula asociada una FGM con parámetro $\theta = 0.6$. Vamos a calcular una estimación de dicho parámetro por el método del τ de Kendall. Ya sabemos, de la Tabla 2.1, que $\tau = \frac{2}{9}\theta$. Entonces $\theta = \frac{9}{2}\tau$, por lo tanto el estimador de θ en este caso es $\hat{\theta} = \frac{9}{2}t$, donde t está dado por la ecuación (2.5). Para los valores de la Tabla 3.2, se tiene que $c = 26$ y $d = 19$, por lo tanto $t = \frac{26-19}{45} = \frac{7}{45}$. Así, se concluye que*

$$\hat{\theta} = \frac{9}{2} \cdot \frac{7}{45} = \frac{7}{10} = 0.7.$$

Ejemplo 3.2.2. *En este caso vamos a estimar el valor de θ por el método del ρ de Spearman. Sabemos que $\rho = \frac{\theta}{3}$ (ver Tabla 2.1), por lo tanto $\theta = 3\rho$, lo que significa que $\hat{\theta} = 3r$ es el estimador de θ en este caso, donde r está dado por la ecuación (2.8). En la Tabla 3.3 se presentan los valores R_i y S_i, necesarios para el cálculo de r.*

i	1	2	3	4	5	6	7	8	9	10
R_i	5	10	6	3	8	2	7	9	4	1
S_i	4	9	2	5	10	3	8	1	7	6

Tabla 3.3: Valores de R_i y S_i necesarios para el cálculo de r.

Con esta información, se tiene que

$$r = \frac{12}{10 \cdot 11 \cdot 9}(322) - 3\frac{11}{9} = \frac{13}{55}.$$

Esto significa que el valor estimado para θ por este método es $\hat{\theta} = 3\left(\frac{13}{55}\right) \approx 0.709$.

3.2.2. Método de Máxima Verosimilitud

Supongamos que $C_\theta(u, v)$ es una función cópula absolutamente continua con función de densidad asociada $c_\theta(u, v)$. Además, sea $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ una muestra aleatoria del vector aleatorio continuo (X, Y), cuya cópula asociada es $C_\theta(u, v)$

y cuya función de densidad conjunta es $f_{(X,Y)}(\cdot,\cdot)$. Entonces, la función de verosimilitud para esta muestra es:

$$L(\theta; x_i, y_i) = \prod_{i=1}^{n} f_{(X,Y)}(x_i, y_i).$$

Remplazando $f_{(X,Y)}(x_i, y_i)$ por la expresión dada en la ecuación (2.3) se tiene que

$$L(\theta; x_i, y_i) = \prod_{i=1}^{n} c_\theta(F_X(x_i), F_Y(y_i)) f_X(x_i) f_Y(y_i).$$

De aquí se puede obtener la función de log-verosimilitud, la cual está dada por

$$\begin{aligned} l(\theta; x_i, y_i) &= \ln L(\theta; x_i, y_i) \\ &= \sum_{i=1}^{n} \ln\left[c_\theta(F_X(x_i), F_Y(y_i))\right] + \sum_{i=1}^{n} \ln\left[f_X(x_i)\right] + \sum_{i=1}^{n} \ln\left[f_Y(y_i)\right]. \end{aligned}$$

Observe que las expresiones $\sum_{i=1}^{n} \ln\left[f_X(x_i)\right]$ y $\sum_{i=1}^{n} \ln\left[f_Y(y_i)\right]$ no dependen de θ; esto significa que los valores de θ que maximizan la función anterior, también maximizan la siguiente:

$$l(\theta; x_i, y_i) = \sum_{i=1}^{n} \ln\left[c_\theta(F_X(x_i), F_Y(y_i))\right]. \tag{3.4}$$

Por lo tanto, en adelante se considerará esta como la función de log-verosimilitud para la muestra aleatoria. Así, el paso siguiente será determinar el valor $\hat{\theta}$ que maximiza la función $l(\theta; x_i, y_i)$, el cual será precisamente el estimador de máxima verosimilitud para el parámetro θ.

Una observación importante es que en la práctica no siempre se conocen las funciones de distribución marginales $F_X(\cdot)$ y $F_Y(\cdot)$. Para solventar esta limitación, se tienen dos posibilidades:

1. Elegir, por algún método, algunas funciones de distribución $F_X(\cdot)$ y $F_Y(\cdot)$ que parezcan adecuadas para las variables X y Y. Este procedimiento se conoce como **Método IFM** (siglas de *Inference For the Margins*). Una de las dificultades de este proceso es el riesgo que existe de que las marginales elegidas no sean las más adecuadas.

2. Sustituir las funciones $F_X(\cdot)$ y $F_Y(\cdot)$ por sus respectivas funciones de distribución marginales empíricas ajustadas por el factor $\dfrac{n}{n+1}$, tal como se establecieron en las ecuaciones (3.1) y (3.2). Este procedimiento se conoce como **Método CML** (siglas de *Canonical Maximum Likelihood*) o **Método de Máxima Pseudo-verosimilitud**.

A continuación se presenta un ejemplo que ilustra este último procedimiento.

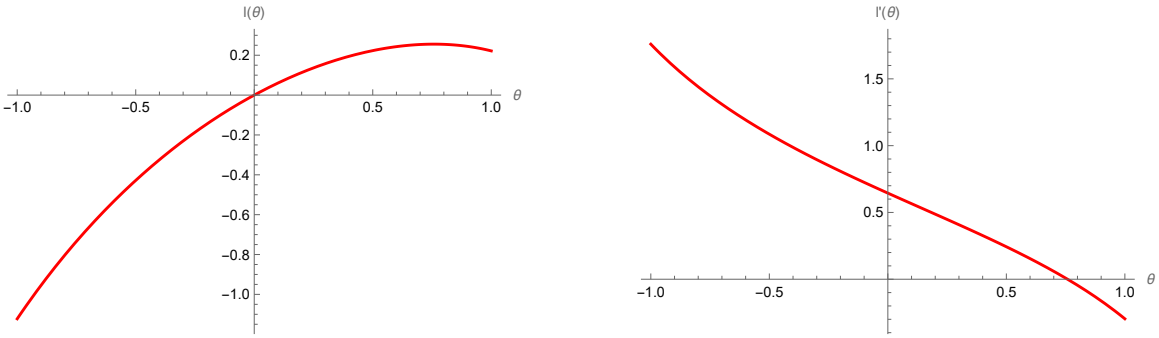
Ejemplo 3.2.3. *Recordemos que los datos de la Tabla 3.1 tienen como cópula asociada una FGM con parámetro θ, el cual estimaremos esta vez por el método de Máxima Pseudo-verosimilitud o CML. En este caso, sabemos que la densidad de la cópula es $c_\theta(u,v) = 1+\theta(1-2u)(1-2v)$. Por lo tanto, la función de log-verosimilitud, de acuerdo con la ecuación (3.4), es*

$$l(\theta) = \sum_{i=1}^{10} \ln\left[1+\theta(1-2u_i)(1-2v_i)\right].$$

Derivando, obtenemos que

$$l'(\theta) = \sum_{i=1}^{10} \frac{(1-2u_i)(1-2v_i)}{1+\theta(1-2u_i)(1-2v_i)}.$$

Remplazando en las ecuaciones anteriores los valores de la Tabla 3.2, se obtiene que la función $l(\theta)$ alcanza un máximo en $\hat{\theta} = 0.7581$. En la Figura 3.3 se presentan los gráficos de las funciones $l(\theta)$ y $l'(\theta)$.



(a) Función $l(\theta)$. **(b)** Función $l'(\theta)$.

Figura 3.3: Gráficas de la función de log-verosimilitud **(a)** y su derivada **(b)**, correspondientes al Ejemplo 3.2.3.

Observación 3.2.1. *El tamaño de muestra de $n = 10$ para los datos de la Tabla 3.1 se eligió con el objetivo de presentar ejemplos sencillos que permitan una visualización más clara de los procedimientos descritos. Sin embargo, esto resulta en que las estimaciones no sean del todo precisas, teniendo en cuenta que los datos fueron generados con un parámetro $\theta = 0.6$. Tomando muestras más grandes los estimadores darán mejores resultados.*

3.3. Medida de dependencia de Ledwina

Sea $C : I^2 \to I$ una cópula. Para $(u, v) \in I^2$ fijo, considere las siguiente funciones:

$$\begin{aligned}
\phi_u(s) &:= -\sqrt{\frac{1-u}{u}}\mathbb{1}_{[0,u]}(s) + \sqrt{\frac{u}{1-u}}\mathbb{1}_{(u,1]}(s), \quad s \in [0,1], \\
&= \begin{cases} -\sqrt{\dfrac{1-u}{u}}, & \text{si } 0 \le s \le u, \\ \sqrt{\dfrac{u}{1-u}}, & \text{si } u < s \le 1. \end{cases}
\end{aligned}$$

$$\begin{aligned}
\phi_v(t) &:= -\sqrt{\frac{1-v}{v}}\mathbb{1}_{[0,v]}(t) + \sqrt{\frac{v}{1-v}}\mathbb{1}_{(v,1]}(t), \quad t \in [0,1], \\
&= \begin{cases} -\sqrt{\dfrac{1-v}{v}}, & \text{si } 0 \le t \le v, \\ \sqrt{\dfrac{v}{1-v}}, & \text{si } v < t \le 1. \end{cases}
\end{aligned}$$

Ahora, considere las siguientes variables aleatorias:

$$Z := \phi_u(U), \qquad W := \phi_v(V),$$

donde $U \sim \text{Unif}([0,1])$, $V \sim \text{Unif}([0,1])$ y $C(u,v)$ es la distribución conjunta de U y V. Entonces:

- En primer lugar, note que

$$\begin{aligned}
E[Z] &= E[\phi_u(U)], \\
&= \left(-\sqrt{\frac{1-u}{u}}\right) P(0 \le U \le u) + \left(\sqrt{\frac{u}{1-u}}\right) P(u \le U \le 1), \\
&= -\sqrt{\frac{1-u}{u}} \cdot u + \sqrt{\frac{u}{1-u}} \cdot (1-u), \\
&= -\sqrt{(1-u)u} + \sqrt{u(1-u)}, \\
&= 0.
\end{aligned}$$

  Similarmente, $E[W] = 0$.

- De otro lado,

$$\begin{aligned}
E[Z^2] &= E[\phi_u^2(U)], \\
&= \left(\frac{1-u}{u}\right) u + \left(\frac{u}{1-u}\right)(1-u), \\
&= (1-u) + u, \\
&= 1.
\end{aligned}$$

Análogamente, $E[W^2] = 1$.

- Finalmente,

$$\begin{aligned}
E[ZW] &= E[\phi_u(U) \cdot \phi_v(V)], \\
&= \sqrt{\frac{1-u}{u}}\sqrt{\frac{1-v}{v}}C(u,v) - \sqrt{\frac{1-u}{u}}\sqrt{\frac{v}{1-v}}[u - C(u,v)] \\
&\quad - \sqrt{\frac{u}{1-u}}\sqrt{\frac{1-v}{v}}[v - C(u,v)] \\
&\quad + \sqrt{\frac{u}{1-u}}\sqrt{\frac{v}{1-v}}[1 + C(u,v) - v - u], \\
&= C(u,v) \cdot \frac{1}{\sqrt{u(1-u)v(1-v)}} - \frac{uv}{\sqrt{u(1-u)v(1-v)}} \\
&\quad - \frac{uv}{\sqrt{u(1-u)v(1-v)}} + \frac{uv}{\sqrt{u(1-u)v(1-v)}}, \\
&= \frac{C(u,v) - uv}{\sqrt{u(1-u)v(1-v)}}.
\end{aligned}$$

Con los resultados anteriores se puede calcular el Coeficiente de correlación de Pearson (Definición 1.1.13, literal *2*) de Z y W.

$$\begin{aligned}
\rho(Z,W) &= \frac{Cov(Z,W)}{\sqrt{Var(Z)Var(W)}}, \\
&= \frac{E[ZW] - E[Z]E[W]}{\sqrt{E[Z^2]E[W^2]}}, \\
&= E[ZW], \\
&= \frac{C(u,v) - uv}{\sqrt{u(1-u)v(1-v)}}, \\
&=: q(u,v).
\end{aligned}$$

En conclusión, la función

$$\begin{array}{rccc} q: & (0,1)\times(0,1) & \longrightarrow & [-1,1] \\ & (u,v) & \longmapsto & q(u,v), \end{array}$$

toma cada punto $(u,v) \in (0,1)^2$ y le asocia el Coeficiente de correlación de Pearson de las variables aleatorias $Z = \phi_u(U)$ y $W = \phi_v(V)$. Dicha función $q(u,v)$ es la que aquí se denomina *Medida de dependencia de Ledwina*, tomada de Ledwina [18].

Esta función resulta ser también una medida de concordancia (ver Proposición 3.1 de Ledwina [18]). Al igual que para el τ de Kendall y el ρ de Spearman, también

se tiene una versión muestral para esta función, que se define a partir de la cópula empírica (Definición 3.1.2); dicha versión muestral, o estimador, es:

$$q^*(u,v) := \frac{C_n(u,v) - uv}{\sqrt{uv(1-u)(1-v)}}. \tag{3.5}$$

El estimador anterior fue utilizado por Tovar et al. [24] para construir un indicador que permite la selección de una función cópula adecuada para unos datos empíricos. Los detalles sobre este indicador se darán en la siguiente sección.

3.4. Un método para seleccionar funciones cópula

Tal como se mencionó al inicio de este capítulo, uno de los problema habituales cuando se trabaja con datos empíricos es que no siempre se tiene certeza de cuál es el modelo del que provienen. En particular, cuando se trabaja con datos bivariados, es deseable saber cuál es la cópula asociada a ellos, de tal forma que se pueda determinar su estructura de dependencia. En ese sentido, se han establecido diversos métodos o pruebas que permiten determinar cuál es la cópula teórica que mejor se ajusta a los datos que se tienen. Uno de estos métodos es el desarrollado en Tovar et al. [24], el cual se describe a continuación.

Sea $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ un conjunto de observaciones de un vector aleatorio continuo (X, Y) cuya cópula asociada no se conoce. En primer lugar, se realiza un estudio visual de la dispersión de los datos anteriores para elegir algunas familias de cópulas que puedan resultar adecuadas. Este estudio se hace mediante un gráfico de dispersión. Por ejemplo, para los datos de la Tabla 3.1, el gráfico de dispersión se presenta en la Figura 3.4.

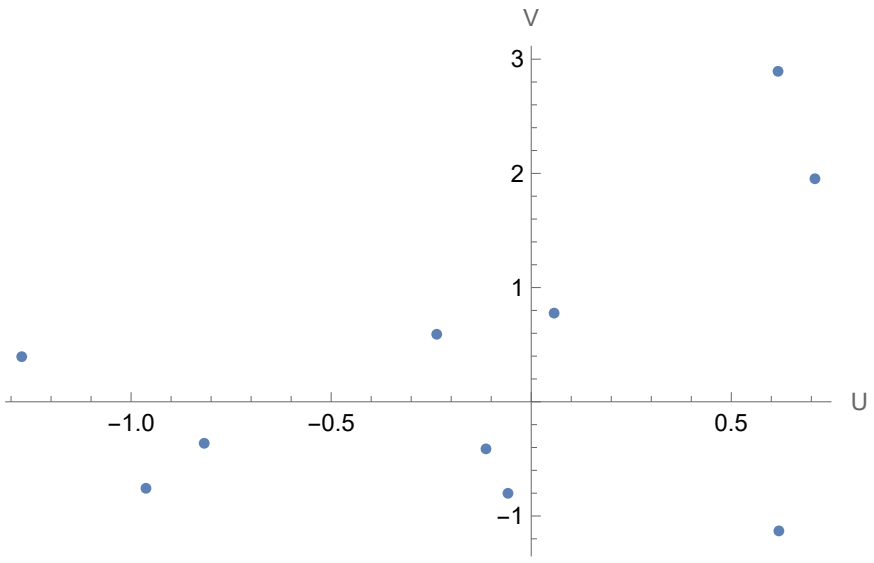


Figura 3.4: Gráfico de dispersión de los datos de la Tabla 3.1.

De este análisis gráfico surge un conjunto de posibles familias de cópulas candidatas, digamos $\mathcal{F} = \{C_\theta^1, C_\theta^2, \ldots, C_\theta^k\}$. La idea es fijar una de ellas como la "función cópula más adecuada" de acuerdo al siguiente procedimiento:

1. A partir de los valores iniciales $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ se deducen las pseudo-observaciones $(u_1, v_1), (u_2, v_2), \ldots, (u_n, v_n)$ empleando las ecuaciones (3.1) y (3.2).

2. Se determina la cópula empírica, $C_n(u, v)$, a partir de las pseudo-observaciones obtenidas en el paso anterior.

3. Se elige una familia de funciones cópula del conjunto $\mathcal{F}$, digamos $C_\theta^1(u, v)$.

4. Se estima el parámetro de $C_\theta^1(u, v)$, usando cualquiera de los métodos presentados en la sección 3.2; denotamos por $\hat{\theta}$ a dicha estimación.

5. Se genera un nuevo conjunto de observaciones $(u_1', v_1'), (u_2', v_2'), \ldots, (u_n', v_n')$, de la familia C_θ^1 con parámetro $\hat{\theta}$.

6. Se evalúan las funciones $q(\cdot, \cdot)$ y $q^*(\cdot, \cdot)$ en cada uno de los (u_i', v_i'), $i = 1 \ldots, n$, obtenidos en el paso anterior, tomando la cópula $C_{\hat{\theta}}^1(u, v)$.

7. Se calcula la siguiente cantidad:

$$I_n := \sqrt{\sum_{i=1}^{n} [q(u_i', v_i') - q^*(u_i', v_i')]^2}. \tag{3.6}$$

Se repiten los pasos 3 al 7 del procedimiento anterior para cada una de las k familias de cópulas del conjunto $\mathcal{F}$, y se elige como "cópula más adecuada" aquella con la cual se obtenga el valor más pequeño de I_n entre todas.

En la Figura 3.5 se presenta un diagrama donde se resume el procedimiento descrito antes.

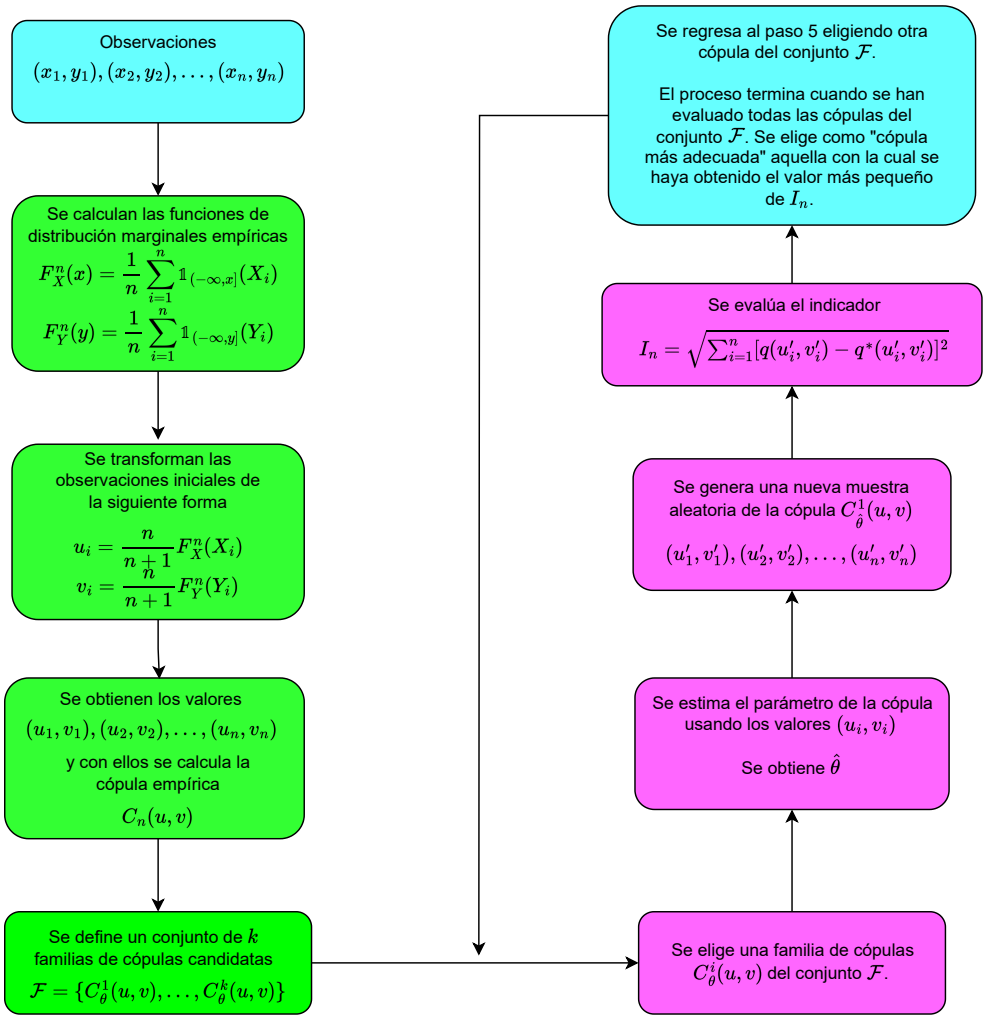


Figura 3.5: Resumen del método para selección de funciones cópula.

Capítulo 4

Una prueba de bondad de ajuste para funciones cópula

En este capítulo se desarrollará el enfoque no paramétrico para abordar el problema que se presentó en el capítulo anterior, que consiste en establecer la cópula asociada a un vector aleatorio continuo (X, Y) a partir de una muestra aleatoria del mismo. Para ello, se construirá una prueba de hipótesis estadística que se pueda utilizar como prueba de bondad de ajuste para funciones cópula, tomando como referencia el indicador analizado en la sección 3.4. En primer lugar, en la sección 4.1 se presentará un resultado importante relacionado con la cópula empírica, algo que será fundamental para determinar la distribución muestral del estadístico de prueba que se presentará en la sección 4.2. Posteriormente, en la sección 4.3 se desarrollará la prueba de bondad de ajuste a partir de dicho estadístico y se expondrán algunas de sus propiedades principales.

4.1. Resultados preliminares

El objetivo de esta sección es exhibir algunos aspectos acerca de la cópula empírica que serán esenciales para la formulación de la prueba de hipótesis de que trata este capítulo. El desarrollo que se presenta a continuación está relacionado con los resultados obtenidos en Fermanian et al. [8] y en la sección 3.10.5.4 de Van der Vaart y Wellner [25].

Sea $C(u, v)$ una función cópula con derivadas parciales continuas, y U y V variables aleatorias continuas con distribución uniforme en $[0, 1]$. Sea $(u, v) \in (0, 1) \times (0, 1)$ fijo y definamos las siguientes variables aleatorias.

- $\xi_1(U, V) = \xi_1 := \mathbb{1}_{[0,u]\times[0,v]}(U, V) = \mathbb{1}_{[0,u]}(U) \cdot \mathbb{1}_{[0,v]}(V)$.
- $\xi_2(U, V) = \xi_2 := \mathbb{1}_{[0,u]\times[0,1]}(U, V) = \mathbb{1}_{[0,u]}(U)$.
- $\xi_3(U, V) = \xi_3 := \mathbb{1}_{[0,1]\times[0,v]}(U, V) = \mathbb{1}_{[0,v]}(V)$.

Observe que:

- $E[\xi_1] = C(u,v)$.
- $E[\xi_2] = C(u,1) = u$.
- $E[\xi_3] = C(1,v) = v$.
- $Var(\xi_1) = C(u,v)[1 - C(u,v)]$.
- $Var(\xi_2) = u(1-u)$.
- $Var(\xi_3) = v(1-v)$.

Además, se tiene que

$$\begin{aligned} Cov(\xi_1,\xi_2) &= E[\xi_1\xi_2] - E[\xi_1]E[\xi_2], \\ &= E[\xi_1] - E[\xi_1]E[\xi_2], \\ &= E[\xi_1](1 - E[\xi_2]), \\ &= C(u,v)(1-u); \end{aligned}$$

$$\begin{aligned} Cov(\xi_1,\xi_3) &= E[\xi_1\xi_3] - E[\xi_1]E[\xi_3], \\ &= E[\xi_1] - E[\xi_1]E[\xi_3], \\ &= C(u,v)(1-v); \end{aligned}$$

$$\begin{aligned} Cov(\xi_2,\xi_3) &= E[\xi_2\xi_3] - E[\xi_2]E[\xi_3], \\ &= E[\xi_1] - E[\xi_2]E[\xi_3], \\ &= C(u,v) - uv. \end{aligned}$$

Por lo tanto, la matriz de covarianzas del vector aleatorio $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}$ es

$$\Sigma_\xi = \begin{pmatrix} C(u,v)[1-C(u,v)] & C(u,v)[1-u] & C(u,v)[1-v] \\ C(u,v)[1-u] & u(1-u) & C(u,v)-uv \\ C(u,v)[1-v] & C(u,v)-uv & v(1-v) \end{pmatrix}. \tag{4.1}$$

Observación 4.1.1. *Observe que el coeficiente de correlación de Pearson de ξ_2 y ξ_3 (Definición 1.1.13), está dado por*

$$\rho(\xi_2,\xi_3) = \frac{Cov(\xi_2,\xi_3)}{\sqrt{Var(\xi_2)Var(\xi_3)}} = \frac{C(u,v)-uv}{\sqrt{uv(1-u)(1-v)}},$$

que resulta ser la función de dependencia de Ledwina, presentada en la sección 3.3.

Ahora, observemos la siguiente variable aleatoria considerada en el ejemplo 3.10.31, p. 538, de Van der Vaart y Wellner [25]:

$$\begin{aligned} \beta(U,V) \equiv \beta &:= \xi_1 - \frac{\partial}{\partial u}C(u,v)\cdot\xi_2 - \frac{\partial}{\partial v}C(u,v)\cdot\xi_3, \\ &= \mathbf{a}^T\cdot\xi, \end{aligned} \tag{4.2}$$

donde $\mathbf{a} = \begin{pmatrix} 1 \\ -\dfrac{\partial C(u,v)}{\partial u} \\ -\dfrac{\partial C(u,v)}{\partial v} \end{pmatrix}$.

La esperanza y varianza de β están dadas por:

$$\begin{aligned} E[\beta] &= E[\xi_1] - \frac{\partial}{\partial u}C(u,v)\cdot E[\xi_2] - \frac{\partial}{\partial v}C(u,v)\cdot E[\xi_2], \\ &= C(u,v) - \frac{\partial}{\partial u}C(u,v)\cdot u - \frac{\partial}{\partial v}C(u,v)\cdot v. \end{aligned} \tag{4.3}$$

$$\begin{aligned} Var(\beta) &= \mathbf{a}^T\cdot\Sigma_\xi\cdot\mathbf{a}, \\ &= C(u,v)[1-C(u,v)] - 2(1-u)C(u,v)\frac{\partial C(u,v)}{\partial u} \\ &\quad -2(1-v)C(u,v)\frac{\partial C(u,v)}{\partial v} + 2\frac{\partial C(u,v)}{\partial u}\frac{\partial C(u,v)}{\partial v}[C(u,v)-uv] \\ &\quad +u(1-u)\left[\frac{\partial C(u,v)}{\partial u}\right]^2 + v(1-v)\left[\frac{\partial C(u,v)}{\partial v}\right]^2. \end{aligned} \tag{4.4}$$

Observación 4.1.2. *De acuerdo con el Teorema 3, página 851 de Fermanian et al. [8], para $(u,v)\in(0,1)^2$, la variable aleatoria $\sqrt{n}(C_n - C)(u,v)$ converge en distribución a una variable aleatoria con distribución normal de media cero y varianza igual a la varianza de la variable aleatoria β, establecida en (4.2). Además, en la sección 4 (p. 1738) de Swanepoel y Allison [23] se desarrolla explícitamente el cálculo de la varianza de la cópula empírica, obteniendo que*

$$Var(C_n(u,v)) = \frac{1}{n}[Var(\beta)] + O\left(\frac{1}{n^2}\right).$$

Cuando $n\to\infty$, se puede tomar $Var(C_n(u,v)) = \dfrac{1}{n}[Var(\beta)]$. Por lo tanto,

$$Var(\beta) = nVar(C_n(u,v)).$$

Ejemplo 4.1.1. *Supongamos que $C(u,v) = uv$. En este caso, $\beta = \xi_1 - v\xi_2 - u\xi_3$ y*

$$\Sigma_\xi = \begin{pmatrix} uv(1-uv) & uv(1-u) & uv(1-v) \\ uv(1-u) & u(1-u) & 0 \\ uv(1-v) & 0 & v(1-v) \end{pmatrix}.$$

De acuerdo a las ecuaciones (4.3) y (4.4) se sigue que

$$E[\beta] = uv - vu - uv = -uv = -C(u,v),$$

$$\begin{aligned} Var(\beta) &= uv(1-uv) - 2uv(1-u)v - 2uv(1-v)u + u(1-u)v^2 + v(1-v)u^2, \\ &= uv(1-u)(1-v). \end{aligned}$$

Por lo tanto, $\rho(\xi_2, \xi_3) = \dfrac{C(u,v) - uv}{\sqrt{uv(1-u)(1-v)}} = 0.$
Además, de acuerdo con el resultado señalado en la Observación 4.1.2, se concluye que $\sqrt{n}q^*(u,v) = \dfrac{\sqrt{n}(C_n(u,v) - uv)}{\sqrt{uv(1-u)(1-v)}} \xrightarrow{d} Z \sim N(0,1).$

4.2. Estadístico de prueba

A partir de los resultados presentados en la sección anterior, se busca proponer un estadístico de prueba similar al de Kolmogorov-Smirnov, con el cual se pueda construir la prueba de bondad de ajuste para funciones cópula. En la sección 4.2.1 se hace un breve resumen acerca de las distribuciones chi-cuadrado y chi y sus principales propiedades; esta información será posteriormente utilizada en la sección 4.2.2 para establecer la distribución muestral de la variable aleatoria que será empleada como estadístico de prueba.

4.2.1. Aspectos básicos de las distribuciones χ^2 y χ

Sea X una variable aleatoria que tiene distribución chi-cuadrado con ν grados de libertad; en símbolos, $X \sim \chi^2_\nu$. Entonces sus funciones de densidad y distribución están dadas por:

$$f_X(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} e^{-x/2} x^{(\nu/2)-1}, \quad x \geq 0,$$

$$F_X(x) = \frac{\Gamma(x/2, \nu/2)}{\Gamma(\nu/2)}, \quad x > 0,\ \nu > 0,$$

donde $\Gamma(\alpha, x)$ corresponde a la **Función Gamma Incompleta** dada en la ecuación (2.9).

La esperanza y varianza para esta distribución están dadas por

$$E[X] = \nu, \qquad Var(X) = 2\nu.$$

Una de las propiedades importantes de la distribución χ^2 es su propiedad reproductiva sobre los grados de libertad, la cual establece que si X y T son variables aleatorias independientes tales que $X \sim \chi^2_k$ y $T \sim \chi^2_l$, entonces $X + T \sim \chi^2_{k+l}$.

Ahora, consideremos la variable aleatoria $Y = \sqrt{X}$. Por el teorema de las transformaciones (Teorema 13, p. 205, de Mood et al. [20]), la función de densidad de Y está dada por:

$$\begin{aligned} f_Y(y) &= f_X(y^2)\left|\frac{d}{dy}y^2\right|, \\ &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)}e^{-y^2/2}y^{\nu/2}(2y), \\ &= \frac{1}{2^{(\nu/2)-1}\Gamma(\nu/2)}e^{-y^2/2}y^{\nu-1}, \quad y>0,\ v>0. \end{aligned}$$

Similarmente, la función de distribución de $Y=\sqrt{X}$ es

$$F_Y(y)=\frac{\Gamma(y^2/2,\nu/2)}{\Gamma(\nu/2)}, \quad y>0,\ \nu>0.$$

La distribución de la variable aleatoria Y se denomina **chi** con ν grados de libertad; en símbolos, $Y\sim\chi_\nu$.

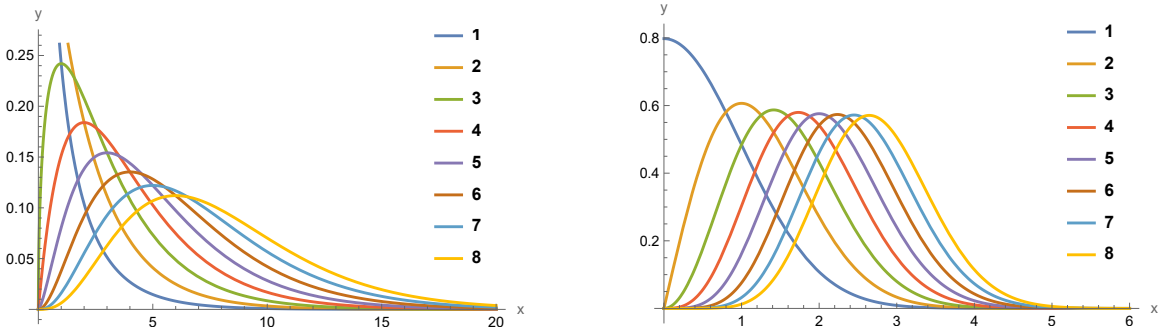
La esperanza y varianza de la distribución chi están dadas por

$$E[Y]=\sqrt{2}\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}, \qquad Var(Y)=\nu-2\left[\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\right]^2.$$

Se sabe que si $X\sim N(0,1)$ entonces $X^2\sim\chi^2_1$. Esto implica que si $X_1,X_2,\ldots,X_n$ son variables aleatorias independientes, todas con distribución normal estándar, entonces $X_1^2+X_2^2+\cdots+X_n^2\sim\chi^2_n$. Luego, si $\mathbf{X}=(X_1,X_2,\ldots,X_n)$, se sigue que

$$||\mathbf{X}||=\sqrt{X_1^2+X_2^2+\cdots+X_n^2}\sim\chi_n.$$

En la Figura 4.1 se presentan los gráficos de la densidad de las distribuciones χ^2 y χ para algunos valores fijos de ν.



(a) Densidad de la distribución χ^2. **(b)** Densidad de la distribución χ.

Figura 4.1: Gráficas de la densidad de las distribuciones χ^2 y χ para algunos valores de ν.

Observe que a medida que el valor de ν se incrementa, las curvas de densidad de las distribuciones χ^2 y χ se asemejan a la densidad de una distribución normal. En efecto, si $X\sim\chi^2_\nu$, entonces

$$X \xrightarrow{d} N(\nu, 2\nu), \quad \text{cuando } \nu \to \infty.$$

Por su parte, cuando $\nu \to \infty$, la distribución χ también converge a una distribución normal. Para verificarlo, definamos $Z = \dfrac{X-\nu}{\sqrt{2\nu}} \xrightarrow{d} N(0,1)$, donde $X \sim \chi^2_\nu$; entonces $X = \sqrt{2\nu}Z + \nu \sim \chi^2_\nu$. Ahora, si tomamos $Y = \sqrt{X}$, se sigue que $Y = \sqrt{\sqrt{2\nu}Z + \nu} \sim \chi_\nu$.

Recordemos que el desarrollo en series de Taylor de una función f en torno a un valor a es

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots .$$

Si tomamos $f(x) = \sqrt{x}$ y $a = \nu$ obtenemos

$$\sqrt{x} = \sqrt{\nu} + \frac{1}{2\sqrt{\nu}}(x-\nu) - \frac{1}{8\nu^{3/2}}(x-\nu)^2 + \cdots .$$

Evaluando en $x = \sqrt{2\nu}Z + \nu$ se obtiene

$$\begin{aligned} Y = \sqrt{\sqrt{2\nu}Z + \nu} &= \sqrt{\nu} + \frac{1}{2\sqrt{\nu}}(\sqrt{2\nu}Z + \nu - \nu) - \frac{1}{8\nu^{3/2}}(\sqrt{2\nu}Z + \nu - \nu)^2 + \cdots , \\ &= \sqrt{\nu} + \frac{1}{2\sqrt{\nu}}(\sqrt{2\nu}Z) - \frac{2\nu Z^2}{8\nu^{3/2}} + \cdots , \\ &= \sqrt{\nu} + \frac{Z}{\sqrt{2}} - \frac{Z^2}{4\sqrt{\nu}} + \cdots . \end{aligned}$$

Note que, cuando $\nu \to \infty$, el tercer sumando y los siguientes en la expresión anterior tienden a cero; así,

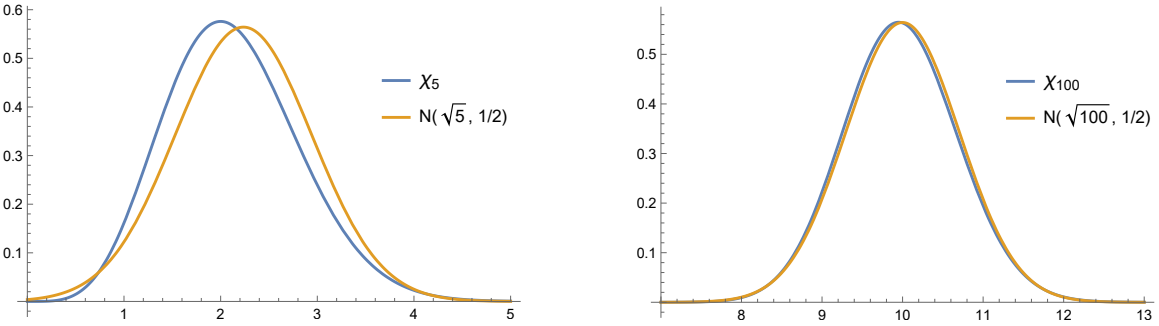
$$Y \approx \sqrt{\nu} + \frac{Z}{\sqrt{2}}.$$

Por lo tanto, $Z \approx \dfrac{Y - \sqrt{\nu}}{\sqrt{1/2}}$. Como $Z \xrightarrow{d} N(0,1)$, entonces

$$Y \xrightarrow{d} N(\sqrt{\nu}, 1/2), \qquad \text{cuando } \nu \to \infty.$$

En la Figura 4.2 se presentan las gráficas de las densidades de las distribuciones χ_ν y $N(\sqrt{\nu}, 1/2)$ para algunos valores de ν, donde se puede apreciar la aproximación señalada.

Observación 4.2.1. *La información de esta sección se fundamenta en el capítulo 18 (p. 415) de Johnson et al. [16], donde se presentan más detalles acerca de estas distribuciones.*



(a) $\nu = 5$. **(b)** $\nu = 100$.

Figura 4.2: Gráficas de la densidad de las distribuciones chi y normal para $\nu = 5$ y $\nu = 100$.

4.2.2. Distribución muestral del estadístico de prueba

Recordemos que el indicador definido en la ecuación (3.6) es:

$$\begin{aligned} I_n &= \sqrt{\sum_{i=1}^{n}[q(u_i,v_i) - q^*(u_i,v_i)]^2}, \\ &= \sqrt{\sum_{i=1}^{n}\left[\frac{C(u_i,v_i) - C_n(u_i,v_i)}{\sqrt{u_iv_i(1-u_i)(1-v_i)}}\right]^2}, \\ &= \sqrt{\sum_{i=1}^{n}\left[\frac{C_n(u_i,v_i) - C(u_i,v_i)}{\sqrt{u_iv_i(1-u_i)(1-v_i)}}\right]^2}. \end{aligned}$$

Ahora, consideremos la siguiente modificación de este indicador:

$$\begin{aligned} I'_n &:= \sqrt{\sum_{i=1}^{n}\left[\frac{\sqrt{n}(C_n(u_i,v_i) - C(u_i,v_i))}{\sqrt{u_iv_i(1-u_i)(1-v_i)}}\right]^2}, \\ &= \sqrt{\sum_{i=1}^{n}(\sqrt{n}q^*(u_i,v_i))^2}, \\ &= ||\sqrt{n}Q^*||, \end{aligned} \tag{4.5}$$

donde $Q^* = \begin{pmatrix} q^*(u_1,v_1) \\ q^*(u_2,v_2) \\ \vdots \\ q^*(u_n,v_n) \end{pmatrix}$ y $||\cdot||$ es la norma euclidiana.

Se sabe que las observaciones $(u_1, v_1), (u_2, v_2), \ldots, (u_n, v_n)$ son independientes por tratarse de una muestra aleatoria; esto implica que las componentes $q^*(u_i, v_i)$, $i = 1, \ldots, n$, del vector Q^* son independientes entre sí. Además, como se vio en el ejemplo 4.1.1, $\sqrt{n}q^*(u_i, v_i) \xrightarrow{d} Z \sim N(0, 1)$ para cada $i = 1, \ldots, n$. De acuerdo a esto se puede concluir que $\sqrt{n}Q^* \xrightarrow{d} \mathbf{Z} \sim N(\mathbf{0}, I_n)$, donde $\mathbf{0}$ representa el vector nulo de n componentes e I_n la matriz identidad de tamaño n.

A continuación se presenta un teorema que es necesario para deducir la convergencia en distribución de I'_n.

Teorema 4.2.1 (Teorema 1.10, p. 59, Shao [22]). *Sean $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}$ vectores aleatorios k-dimensionales definidos en un espacio de probabilidad y sea $g : \mathbb{R}^k \to \mathbb{R}^m$ una función continua en casi todas partes. Entonces $\mathbf{X_n} \xrightarrow{d} \mathbf{X}$ implica que $g(\mathbf{X_n}) \xrightarrow{d} g(\mathbf{X})$.*

Observe que la función norma euclidiana, dada por

$$\begin{array}{rccl} ||\cdot|| & : & \mathbb{R}^n & \to & \mathbb{R} \\ & & \mathbf{x} & \mapsto & ||\mathbf{x}|| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}, \end{array}$$

donde $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, es una función continua. Por lo tanto, aplicando el teorema anterior, los resultados de la sección 4.2.1, y el hecho de que $\sqrt{n}Q^* \xrightarrow{d} \mathbf{Z} \sim N(\mathbf{0}, I_n)$, se concluye que

$$I'_n = ||\sqrt{n}Q^*|| \xrightarrow{d} X \sim \chi_n.$$

De manera general, cuando $C(u, v)$ no es necesariamente la cópula de la independencia, se puede considerar la siguiente variable aleatoria:

$$\begin{aligned} J_n &:= \sqrt{\sum_{i=1}^{n} \frac{n(C_n(u_i, v_i) - C(u_i, v_i))^2}{Var(\beta)}}, \qquad (4.6) \\ &= \sqrt{\sum_{i=1}^{n} \left[\frac{\sqrt{n}(C_n(u_i, v_i) - C(u_i, v_i))}{\sqrt{Var(\beta)}} \right]^2}, \\ &= \sqrt{\sum_{i=1}^{n} \left[\frac{\sqrt{n}(C_n(u_i, v_i) - C(u_i, v_i))}{\sqrt{nVar(C_n(u_i, v_i))}} \right]^2}, \\ &= \sqrt{\sum_{i=1}^{n} \left[\frac{C_n(u_i, v_i) - C(u_i, v_i)}{\sqrt{Var(C_n(u_i, v_i))}} \right]^2}, \\ &= ||L||, \end{aligned}$$

donde $L = \begin{pmatrix} \frac{C_n - C}{\sqrt{Var(C_n)}}(u_1, v_1) \\ \vdots \\ \frac{C_n - C}{\sqrt{Var(C_n)}}(u_n, v_n) \end{pmatrix}$.

Siguiendo el mismo proceso que con I'_n, se llega a la conclusión de que J_n también converge en distribución a $X \sim \chi_n$, para cualquier función cópula continua $C(u, v)$ tal que sus derivadas parciales existan.

Con los resultados anteriores, es posible emplear las variables I'_n y J_n como estadísticos de prueba, el primero para probar independencia, y el segundo para probar cualquier función cópula continua con derivadas parciales.

Adicionalmente, observe que cuando aumenta el tamaño de muestra ($n \to \infty$), la distribución de J_n (o I'_n) se aproxima a una normal con $\mu = \sqrt{\nu}$ y $\sigma^2 = 1/2$, según los resultados de la sección 4.2.1. Por lo tanto, se tiene que $\sqrt{2}(J_n - \sqrt{n}) \xrightarrow{d} N(0, 1)$.

4.3. Prueba de hipótesis

El objetivo de esta sección es presentar formalmente la prueba de bondad de ajuste para funciones cópula, partiendo del estadístico J_n dado en la sección 4.2.2.

Sea $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ una muestra aleatoria cuya cópula asociada $C(u, v)$ es desconocida. Sea $C^1(u, v)$ una cópula fija tal que sus derivadas parciales existen. El objetivo es determinar, a partir de la muestra, si $C^1(u, v)$ es una función cópula adecuada para describir la dependencia entre X y Y. En otras palabras, se desea validar la siguiente hipótesis:

$$\begin{cases} H_0: & C(u, v) = C^1(u, v), \\ H_1: & C(u, v) \neq C^1(u, v). \end{cases}$$

Para ello se utilizará como estadístico de prueba J_n, dado en la ecuación (4.6).

- **Región crítica:** Se toma como región crítica la siguiente:

$$C = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \mid J_n \geq j_{1-\alpha}\}, \tag{4.7}$$

donde α será el nivel de significancia de la prueba y $j_{1-\alpha}$ representa el percentil $1 - \alpha$ de la distribución χ_n. La elección de esta región crítica está justificada por la parte *(i)* del Lema 11.2.1, p. 430, de Lehmann y Romano [19], el cual se enuncia a continuación:

Lema 1. *Sea $\{F_n(\cdot)\}$ una sucesión de funciones de distribución en la recta real que converge a una función de distribución $F(\cdot)$. Suponga que $F(\cdot)$ es continua y estrictamente creciente en $y = F^{-1}(1-\alpha)$. Entonces $F_n^{-1}(1-\alpha) \to F^{-1}(1-\alpha)$.*

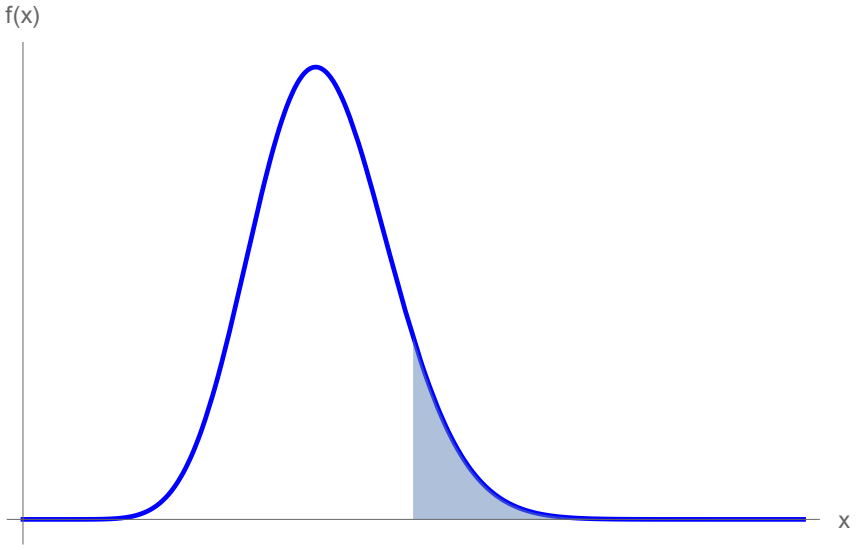


Figura 4.3: Región crítica definida en la ecuación (4.7).

La primera hipótesis del teorema anterior es equivalente a decir que una sucesión de variables aleatorias $\{X_n\}$, cada una con función de distribución $F_n(\cdot)$, converge en distribución a una variable X con función de distribución $F(\cdot)$. La segunda hipótesis es satisfecha por la función de distribución χ_n, por lo tanto se puede aplicar el Lema para la determinación de la región crítica.

Para determinar el valor de $j_{1-\alpha}$, se debe resolver la siguiente ecuación:

$$F_J(j_{1-\alpha}) = \frac{\Gamma(j_{1-\alpha}^2/2, n/2)}{\Gamma(n/2)} = 1 - \alpha;$$

esto se puede realizar por medio de tablas como las que aparecen en Harter [15], o bien usando algún programa de cómputo como *Wolfram Mathematica*.

En la Figura 4.3 se muestra un ejemplo de la región crítica (área sombreada), donde la curva representa la función de densidad de la distribución χ_n.

- **Función potencia:** De acuerdo con la Definición 1.2.11, se obtiene que la función potencia en este caso está dada por

$$\begin{aligned} \pi(C(u,v)) &= P[J_n \in C], \\ &= P[J_n \leq j_{1-\alpha}], \\ &= F_{J_n}(j_{1-\alpha}), \\ &= \frac{\Gamma(j_{1-\alpha}^2/2, n/2)}{\Gamma(n/2)}. \end{aligned}$$

- **Valor p:** Sea k el valor calculado de J_n usando el conjunto de observaciones $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. Teniendo en cuenta la Definición 1.2.12, se deduce que

$$\begin{aligned} p &= P[J_n \geq k], \\ &= 1 - P[J_n \leq k], \\ &= 1 - F_{J_n}(k), \\ &= 1 - \frac{\Gamma(k^2/2, n/2)}{\Gamma(n/2)}. \end{aligned}$$

Observación 4.3.1. *Cuando el tamaño de la muestra aumenta, se puede desarrollar la prueba empleando la aproximación de la distribución χ_n a la normal. En este caso, se tiene que $J_n \xrightarrow{d} N(\sqrt{n}, 1/2)$, luego $\sqrt{2}(J_n - \sqrt{n}) \xrightarrow{d} N(0, 1)$. Por lo tanto, se toma como región crítica la siguiente:*

$$C = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \mid \sqrt{2}(J_n - \sqrt{n}) \geq \Phi^{-1}(1 - \alpha)\}.$$

El valor p también se puede obtener por la aproximación a la distribución normal; en este caso se toma

$$p \approx 1 - \Phi(k'),$$

donde k' es el valor calculado de $\sqrt{2}(J_n - \sqrt{n})$ con la muestra que se tiene, y $\Phi(\cdot)$ es la función de distribución normal estándar.

Ejemplo 4.3.1. *Implementaremos la prueba de hipótesis para los datos de la Tabla 3.1. La hipótesis que se desea validar es la siguiente:*

$$\begin{cases} H_0: & C(u, v) = C_{\hat{\theta}}(u, v), \\ H_1: & C(u, v) \neq C_{\hat{\theta}}(u, v), \end{cases}$$

donde $C_{\hat{\theta}}(u, v)$ es una cópula FGM con parámetro $\hat{\theta}$. Para ello, en primer lugar se genera una nueva muestra de la cópula supuesta con el parámetro $\hat{\theta}$ estimado a partir de la muestra inicial. Del ejemplo 3.2.3 se tiene que $\hat{\theta} = 0.7581$. En la Tabla 4.1 se presenta la nueva muestra simulada, usando el programa Wolfram Mathematica.

i	1	2	3	4	5	6	7	8	9	10
u_i'	0.27	0.30	0.64	0.71	0.62	0.53	0.31	0.66	0.22	0.48
v_i'	0.12	0.33	0.78	0.43	0.22	0.46	0.74	0.71	0.79	0.11

Tabla 4.1: Nueva muestra generada de una cópula FGM con parámetro $\hat{\theta} = 0.7581$.

Para la cópula FGM, la expresión de $Var(\beta)$ (ecuación (4.4)) está dada por:

$$\begin{aligned} Var(\beta)(u,v) &= uv(1-u)(1-v)[1+(1-2u)(1-2v)\theta \\ &\quad +(-v(1-v)-u(1-7v+7v^2)+u^2(1-7v+7v^2))\theta^2 \\ &\quad +2uv(1-3u+2u^2)(1-3v+2v^2)\theta^3]. \end{aligned}$$

Con esta expresión, y con la cópula empírica definida en el ejemplo 3.1.2, se calcula el valor del estadístico de prueba, J_n, evaluándolo en cada uno de los (u_i', v_i') de la Tabla 4.1. De esta manera se obtiene

$$J_n = \sqrt{\sum_{i=1}^{10} \frac{10(C_{10}(u_i', v_i') - C_{\hat{\theta}}(u_i', v_i'))^2}{Var(\beta)(u_i', v_i')}} = 3.78304.$$

Si se elige un nivel de significancia de $\alpha = 0.1$, se tiene que el percentil $1-\alpha$ es $j_{0.9} = 3.9984$. De esta manera se sigue que $J_n < j_{1-\alpha}$, y por tanto no se rechaza la hipótesis nula. Esto permite concluir que los datos de la Tabla 3.1 sí proceden de una cópula FGM, como en efecto sucede.
El valor p para esta prueba es

$$p = 1 - \frac{\Gamma(\frac{3.78304^2}{2}, 5)}{\Gamma[5]} = 0.159255.$$

Usando la aproximación a la distribución normal, se obtiene que

$$\sqrt{2}(J_n - \sqrt{10}) = 0.8779.$$

Por su parte, $\Phi^{-1}(0.9) = 1.28155$. Entonces, como $\sqrt{2}(J_n - \sqrt{10}) < \Phi^{-1}(1-\alpha)$, se llega a la misma conclusión de no rechazar la hipótesis nula. Asimismo, el valor p por medio de esta aproximación es

$$p \approx 1 - \Phi(0.8779) = 0.19.$$

Al tomar un tamaño de muestra más grande, esta aproximación ofrecerá una mayor exactitud.

Ejemplo 4.3.2. *Ahora implementaremos la prueba de hipótesis para 10 pares de observaciones (x_i, y_i), cuyas componentes provienen de variables aleatorias X y Y independientes con distribución normal estándar. La hipótesis que se desea validar es la siguiente:*

$$\begin{cases} H_0: & C(u,v) = uv, \\ H_1: & C(u,v) \neq uv. \end{cases}$$

Los datos que se usarán en este ejemplo fueron generados usando Wolfram Mathematica y se presentan en la Tabla 4.2.

i	1	2	3	4	5	6	7	8	9	10
x_i	−0.99	1.80	−0.63	0.30	−1.68	0.40	−0.21	−0.75	0.54	0.42
y_i	−0.81	1.13	−0.13	−1.09	−0.79	−0.24	−1.44	−0.92	−1.32	−1.45

Tabla 4.2: Muestra de datos independientes con distribución normal estándar.

i	1	2	3	4	5	6	7	8	9	10
u_i	2/11	10/11	4/11	6/11	1/11	7/11	5/11	3/11	9/11	8/11
v_i	6/11	10/11	9/11	4/11	7/11	8/11	2/11	5/11	3/11	1/11

Tabla 4.3: Pseudo-observaciones de los datos de la Tabla 4.2.

Siguiendo el mismo procedimiento descrito en el ejemplo 3.1.2, se obtienen las pseudo-observaciones.

Con los valores de la Tabla 4.3 se obtiene la cópula empírica:

$$\begin{aligned} C_{10}(u,v) \;=\; & \frac{1}{10}[\mathbb{1}_{(-\infty,u]}(2/11)\,\mathbb{1}_{(-\infty,v]}(6/11) + \mathbb{1}_{(-\infty,u]}(10/11)\,\mathbb{1}_{(-\infty,v]}(10/11) \\ & +\mathbb{1}_{(-\infty,u]}(4/11)\,\mathbb{1}_{(-\infty,v]}(9/11) + \mathbb{1}_{(-\infty,u]}(6/11)\,\mathbb{1}_{(-\infty,v]}(4/11) \\ & +\mathbb{1}_{(-\infty,u]}(1/11)\,\mathbb{1}_{(-\infty,v]}(7/11) + \mathbb{1}_{(-\infty,u]}(7/11)\,\mathbb{1}_{(-\infty,v]}(8/11) \\ & +\mathbb{1}_{(-\infty,u]}(5/11)\,\mathbb{1}_{(-\infty,v]}(2/11) + \mathbb{1}_{(-\infty,u]}(3/11)\,\mathbb{1}_{(-\infty,v]}(5/11) \\ & +\mathbb{1}_{(-\infty,u]}(9/11)\,\mathbb{1}_{(-\infty,v]}(3/11) + \mathbb{1}_{(-\infty,u]}(8/11)\,\mathbb{1}_{(-\infty,v]}(1/11)]. \end{aligned}$$

Ahora, se genera una nueva muestra aleatoria de la cópula que se desea probar, en este caso la cópula producto, $C(u,v) = uv$. Los valores, generados con Wolfram Mathematica, se presentan en la Tabla 4.4.

i	1	2	3	4	5	6	7	8	9	10
u'_i	0.12	0.32	0.78	0.43	0.22	0.46	0.74	0.71	0.79	0.11
v'_i	0.38	0.35	0.55	0.73	0.71	0.55	0.23	0.59	0.15	0.63

Tabla 4.4: Nueva muestra generada de una cópula de independencia.

Finalmente, con los valores de la Tabla 4.4, se calcula el estadístico de prueba; en este caso se emplea I'_n (ecuación (4.6)), ya que la hipótesis nula que se desea validar es que los datos son independientes. Se obtiene entonces el siguiente resultado:

$$I'_n = \sqrt{\sum_{i=1}^{10}\left[\frac{\sqrt{10}(C_{10}(u'_i,v'_i) - u'_i v'_i)}{\sqrt{u'_i v'_i (1-u'_i)(1-v'_i)}}\right]^2} = 2.83715.$$

Eligiendo un nivel de significancia de $\alpha = 0.1$, se tiene que el percentil $1-\alpha$ de la distribución χ_{10} es $j_{0.9} = 3.9984$. Así, se concluye que $I'_n < j_{1-\alpha}$, y por lo tanto no se rechaza la hipótesis nula.

Del procedimiento anterior se puede decir que los datos de la Tabla 4.2 sí son independientes, lo cual ya sabemos que es correcto.
El valor p para esta prueba es:

$$p = 1 - \frac{\Gamma(\frac{2.83715^2}{2}, 5)}{\Gamma[5]} = 0.62401.$$

En la Figura 4.4 se presenta un diagrama de flujo que resume el proceso completo para hacer uso de la prueba de bondad de ajuste presentada en esta sección.

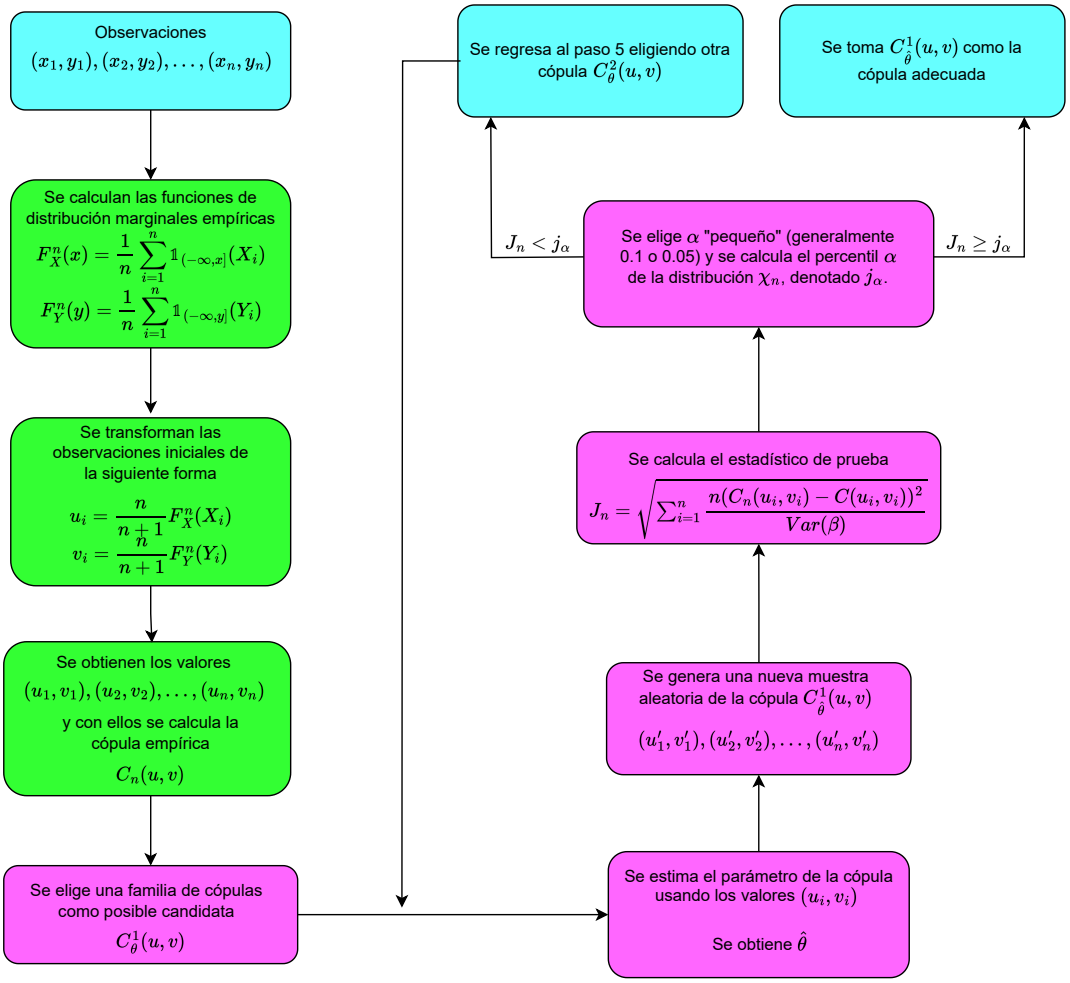


Figura 4.4: Resumen del procedimiento para emplear la prueba de bondad de ajuste.

Bibliografía

[1] U. U. d. Anjos y N. Kolev, "Representation of bivariate copulas via local measure of dependence," *Department of Statistics, University of Sao Paulo*, 2005.

[2] L. J. Bain y M. Engelhardt, *Introduction to probability and mathematical statistics*, 2.ª ed. Duxbury Press Belmont, CA, 1992.

[3] G. Casella y R. L. Berger, *Statistical inference*, 2.ª ed. Cengage Learning, 2002.

[4] P. Deheuvels, "An asymptotic decomposition for multivariate distribution-free tests of independence," *Journal of Multivariate Analysis*, vol. 11, n.º 1, págs. 102-113, 1981.

[5] T. Emura, C. L. Sofeu y V. Rondeau, "Conditional copula models for correlated survival endpoints: Individual patient data meta-analysis of randomized controlled trials," *Statistical Methods in Medical Research*, vol. 30, n.º 12, págs. 2634-2650, 2021.

[6] A.-C. Favre, S. El Adlouni, L. Perreault, N. Thiémonge y B. Bobée, "Multivariate hydrological frequency analysis using copulas," *Water resources research*, vol. 40, n.º 1, 2004.

[7] J.-D. Fermanian, "Goodness-of-fit tests for copulas," *Journal of multivariate analysis*, vol. 95, n.º 1, págs. 119-152, 2005.

[8] J.-D. Fermanian, D. Radulovic y M. Wegkamp, "Weak convergence of empirical copula processes," *Bernoulli*, vol. 10, n.º 5, págs. 847-860, 2004.

[9] E. Flores de la Fuente, "EDAs con Funciones de Cópula," Tesis de Maestría, Centro de Investigación en Matemática (CIMAT), Guanajuato, México, sep. de 2009.

[10] C. Genest y A.-C. Favre, "Everything you always wanted to know about copula modeling but were afraid to ask," *Journal of hydrologic engineering*, vol. 12, n.º 4, págs. 347-368, 2007.

[11] C. Genest, M. Gendron y M. Bourdeau-Brien, "The advent of copulas in finance," *Copulae and multivariate probability distributions in finance*, págs. 1-10, 2013.

[12] C. Genest y B. Rémillard, "Test of independence and randomness based on the empirical copula process," *Test*, vol. 13, págs. 335-369, 2004.

[13] C. Genest, B. Rémillard y D. Beaudoin, "Goodness-of-fit tests for copulas: A review and a power study," *Insurance: Mathematics and economics*, vol. 44, n.º 2, págs. 199-213, 2009.

[14] I. Gijbels, M. Omelka y D. Sznajder, "Positive quadrant dependence tests for copulas," *Canadian Journal of Statistics*, vol. 38, n.º 4, págs. 555-581, 2010.

[15] H. L. Harter, "A new table of percentage points of the Pearson type III distribution," *Technometrics*, vol. 11, n.º 1, págs. 177-187, 1969.

[16] N. Johnson, S. Kotz y N. Balakrishan, *Continuous Univariate Distributions*. Wiley, New York, 1994, vol. 1.

[17] L. Latorre Lloréns, *Teoría de cópulas: Introducción y aplicaciones a Solvencia II*. Madrid: Fundación MAPFRE, 2017.

[18] T. Ledwina, "Visualizing association structure in bivariate copulas using new dependence function," *Stochastic Models, Statistics and Their Applications: Wroclaw, Poland, February 2015*, págs. 19-27, 2015.

[19] E. L. Lehmann y J. P. Romano, *Testing statistical hypotheses*, 3.ª ed. Springer, 2005.

[20] A. M. Mood, F. A. Graybill y D. C. Boes, *Introduction to the Theory of Statistics*. McGraw-Hill, 1974.

[21] R. B. Nelsen, *An introduction to copulas*. Springer, 2006.

[22] J. Shao, *Mathematical statistics*, 2.ª ed. Springer Science & Business Media, 2003.

[23] J. Swanepoel y J. Allison, "Some new results on the empirical copula estimator with applications," *Statistics & Probability Letters*, vol. 83, n.º 7, págs. 1731-1739, 2013.

[24] J. R. Tovar Cuevas, J. Portilla Yela y J. A. Achcar, "A method to select bivariate copula functions," *Revista Colombiana de Estadística*, vol. 42, n.º 1, págs. 61-80, 2019.

[25] A. W. van der Vaart y J. A. Wellner, *Weak Convergence and Empirical Processes*, 2.ª ed. Cham, Suiza: Springer, 2023.