

CÓPULAS Y DEPENDENCIA

Manual didáctico de autoaprendizaje

Dr. Arturo Erdely Ruiz

UNAM FES Acatlán (2025)

PREFACIO

El estudio de la dependencia entre variables aleatorias constituye uno de los temas más relevantes y, al mismo tiempo, más desafiantes para la ciencia estadística. Con frecuencia, los métodos tradicionales se limitan a utilizar modelos lineales o a cuantificar la relación mediante un único coeficiente, lo que puede ocultar aspectos importantes de la interacción entre variables. En este contexto, las funciones cópula ofrecen un marco teórico robusto y flexible para describir, medir y visualizar la dependencia en toda su complejidad.

Este manual surge con el propósito de brindar a estudiantes y profesores de la licenciatura en Actuaría, y de otras disciplinas afines, una introducción accesible y práctica al uso de las cópulas como herramientas para el análisis de dependencia. A lo largo de sus capítulos, el lector encontrará explicaciones breves, ejemplos ilustrativos y representaciones gráficas que facilitan la comprensión de conceptos abstractos, así como ejercicios diseñados para fortalecer la intuición y la aplicación de las ideas presentadas.

En particular, se hace énfasis en la visualización de la dependencia, un recurso pedagógico que permite reconocer patrones y estructuras de dependencia que podrían pasar inadvertidos con enfoques meramente numéricos, o que lucen distorsionados con los tradicionales diagramas de dispersión. De esta forma, se pretende que el manual se convierta en una guía práctica tanto para el autoaprendizaje como para su uso en cursos formales de probabilidad y estadística.

Algunos de los ejemplos abordados en este texto requieren de la realización de cálculos numéricos y generar algunas gráficas. Para tal fin se utilizó el moderno lenguaje de programación **Julia**, y el código de programación de esos ejemplos se encuentra disponible en la plataforma **Github** mediante el siguiente enlace: <https://github.com/aerdely/copdepmanual>

Finalmente, el contenido de este manual no pretende agotar el vasto campo de estudio de las cópulas, sino servir como un punto de partida sólido para quienes se inician en el tema. La combinación de fundamentos teóricos, ejemplos prácticos y recursos visuales busca despertar la curiosidad del lector e invitarlo a profundizar, tanto en aspectos teóricos como en sus aplicaciones. Con ello, se aspira a que este texto sea no solo un material de consulta, sino también una herramienta inspiradora para continuar explorando la riqueza del análisis de dependencia en la estadística contemporánea.

ÍNDICE GENERAL

Prefacio	3
Índice general	5
Índice de figuras	7
1. Dependencia probabilística	11
1. Eventos dependientes	12
2. Variables aleatorias dependientes	19
3. Simulación	36
4. Actividades	44
2. Funciones cópula	45
1. Teorema de Sklar	46
2. Construcción de familias de cópulas	54
2.1. Método de inversión	60
2.2. Cópulas arquimedianas	62
2.3. Mezclando, pegando y rotando	70
3. Simulación	77
4. Estimación	86
4.1. Estimación no paramétrica	87
4.2. Estimación paramétrica	92
5. Actividades	95
3. Medir y visualizar dependencia	97
1. Dependencia de Schweizer-Wolff	98
2. Correlación de Spearman	103
3. Diagramas de dependencia	107
4. Actividades	112
Bibliografía	119

ÍNDICE DE FIGURAS

1.1.	Tipos de funciones de distribución de probabilidades: a) continua, b) discreta, c) mixta.	30
1.2.	Ejemplo 6. <i>Izquierda:</i> Función de distribución F_X monótona creciente. <i>Derecha:</i> Funciones quasi-inversas de F_X	39
2.1.	Curvas de nivel de las cópulas W , Π y M	57
2.2.	Δ_t es la región que contiene todos los conjuntos de nivel t . Curvas de nivel t de las cópulas W (rojo), Π (azul) y M (verde).	58
2.3.	La zona gris indica el conjunto donde siempre están contenidas las secciones horizontales, diagonal y diagonal secundaria de cualquier cópula bivariada. En colores verde, azul y rojo se indican las secciones correspondientes a las cópulas M , Π y W , respectivamente. El caso de las secciones verticales es totalmente análogo al de las horizontales.	59
2.4.	Ejemplo 9. Curvas de nivel de las densidades conjuntas de (X, Y) , (U, V) y (S, T) que tienen la misma cópula subyacente, y por tanto la misma estructura de dependencia, pero con marginales distintas.	62
2.5.	Ejemplo 10. Construcción de la cópula obtenida en (2.27). Para el generador no estricto y su pseudo-inversa se comparan los valores del parámetro $\theta = 1$ (que corresponde a W) versus $\theta = 3$. Las curvas de nivel solo corresponden al caso $\theta = 3$ y tendrían que compararse versus las curvas de nivel de W en la Figura 2.1a. La sección diagonal con $\theta = 3$ (negro) se compara versus las secciones diagonales de las cotas de Fréchet-Hoeffding (rojo y verde) y la cópula independencia (azul).	66
2.6.	Ejemplo 11. Construcción de la cópula obtenida en (2.28). Para el generador estricto y su pseudo-inversa se comparan los valores del parámetro $\theta = 1$ (que corresponde a Π) versus $\theta = 2$. Las curvas de nivel solo corresponden al caso $\theta = 2$ y tendrían que compararse versus las curvas de nivel de Π en la Figura 2.1b. La sección diagonal con $\theta = 2$ (negro) se compara versus las secciones diagonales de las cotas de Fréchet-Hoeffding (rojo y verde) y la cópula independencia (azul).	68
2.7.	Ejemplo 12. Comparando las curvas de nivel de las cotas de Fréchet-Hoeffding W y M con las de la combinación lineal convexa $C = \frac{1}{2}W + \frac{1}{2}M$, así como sus secciones diagonales.	71

2.8. Ejemplo 13. Curvas de nivel y secciones diagonales de los pegados de cópulas $W \boxplus_{0.5} M$ y $M \boxplus_{0.5} W$	73
2.9. Ejemplo 14. Efecto de las rotaciones b)vertical, c)diagonal y d)horizontal, de la densidad de la cópula a) (2.32) con parámetro $\theta = 2$	76
2.10. Ejemplo 15. Simulaciones de vectores aleatorios con la misma cópula subyacente (y por tanto la misma dependencia entre variables) pero con distribuciones marginales distintas.	79
2.11. Ejemplo 16. a) Cúpula C_θ con rotación $\odot 180^\circ$. b) Cúpula C_θ con rotación $\odot 90^\circ$. c) Combinación lineal convexa de a) y b) con ponderación $\alpha = 0.3$ para a). d) Pegado de las cúpulas a) y b) con $u = 0.3$ como punto de pegado.	83
2.12. Ejemplo 17. Simulaciones de vectores aleatorios con la misma cúpula subyacente (gaussiana en este caso), pero con marginales a) Normales $(0, 1)$ y b) Uniformes $(0, 1)$	86
2.13. Ejemplo 18. a) $n = 4,000$ simulaciones de la cúpula paramétrica $V(C_\theta)$, con $\theta = 1$. b) Curvas de nivel de la cúpula paramétrica $V(C_\theta)$. c) Estimación empírica de $V(C_\theta)$. d) Copula Bernstein con $m = 100$, como estimación no paramétrica de $V(C_\theta)$	90
2.14. Ejemplo 19. a) Estimación puntual por máxima verosimilitud del parámetro θ de la familia paramétrica de cúpulas (2.68). b) Similar al inciso a) anterior pero por el método de mínimos cuadrados. c) Histograma de simulaciones a posteriori de θ , junto con estimación puntual (mediana) y por intervalo de probabilidad 95 %.	95
3.1. Ejemplo 20. a) Valores de la cúpula $C_\theta = W \boxplus_\theta M$ en la partición inducida en su dominio. b) Curvas de nivel de la diferencia $C_\theta - \Pi$, resultando no positiva para $u \leq \theta$, y no negativa para $u \geq \theta$. c) En las regiones D_1 y D_2 tenemos que $C_\theta \leq \Pi$, y en las regiones D_3 y D_4 tenemos que $C_\theta \geq \Pi$. d) Dependencia de Schweizer-Wolff bajo la cúpula subyacente C_θ para distintos valores $0 < \theta < 1$	104
3.2. Ejemplo 20. a) Comparando los valores de la medida de dependencia de Schweizer-Wolff $0 \leq \mu_{C_\theta} \leq 1$ versus la correlación de Spearman $-1 \leq \rho_{C_\theta} \leq +1$. b) Comparando μ_{C_θ} versus el valor absoluto $ \rho_{C_\theta} $	107
3.3. Ejemplo 15 del Capítulo 2. Simulaciones de vectores aleatorios con la misma cúpula subyacente (y por tanto la misma dependencia entre variables) pero con distribuciones marginales distintas.	108
3.4. Diagramas de dispersión de simulaciones de pares de variables aleatorias continuas e independientes, pero con funciones de densidad de probabilidades marginales distintas, como se indica en los histogramas en color verde y azul. Se consideraron 5 tipos de densidades marginales, en el siguiente orden: uniforme, decreciente, unimodal, y dos tipos de bimodales. Fuente de la gráfica: Erdely y Rubio-Sánchez, «D-plots: Visualizations for Analysis of Bivariate Dependence Between Continuous Random Variables».	114
3.5. $n = 3000$ simulaciones de las cúpulas W , Π y M aplicando el Teorema 3.2.	115
3.6. Ejemplo 22. a) C_θ muy cercana a Π . b) C_θ en un punto intermedio entre Π y M en términos de su dependencia. c) C_θ muy cercana a M . d) Dependencia de Schweizer-Wolff bajo la cúpula subyacente C_θ para distintos valores $0 < \theta < 1$	116

3.7. Ejemplos de patrones usuales al visualizar diagramas de dependencia (ver el texto para las interpretaciones). Fuente de la gráfica: Erdely y Rubio-Sánchez, «D-plots: Visualizations for Analysis of Bivariate Dependence Between Continuous Random Variables»	117
---	-----

CAPÍTULO 1

DEPENDENCIA PROBABILÍSTICA

1.	Eventos dependientes	12
2.	Variables aleatorias dependientes	19
3.	Simulación	36
4.	Actividades	44

Este capítulo ofrece una breve revisión sobre cómo se formaliza y analiza la dependencia probabilística, comenzando con el concepto de eventos dependientes, seguido por la construcción de variables aleatorias y vectores aleatorios. Se discute cómo la información disponible condiciona toda medida de probabilidad, y cómo dicha información puede actualizarse ante nuevos datos u ocurrencias de eventos. Asimismo, se resumen herramientas como las funciones de distribución y densidades conjuntas, que permiten modelizar y simular dependencia entre variables aleatorias. A través de ejemplos concretos se ilustran los efectos prácticos de la dependencia y se introduce el concepto de independencia como caso límite. Este capítulo sienta las bases conceptuales para el análisis de la dependencia bajo la teoría de funciones cópula, que se abordará en capítulos posteriores.

1. Eventos dependientes

El objeto de estudio de la *Teoría de la Probabilidad* son los experimentos y fenómenos aleatorios, entendiendo por ello procesos cuyo resultado nos resulta incierto a priori. Usualmente se reserva la palabra *experimento* para referirse a procesos controlados o generados por el ser humano, en contraste con un *fenómeno* en donde el ser humano no tiene control alguno y simplemente se limita a observarlo. Recordemos que una *proposición lógica* es cualquier enunciado o afirmación susceptible de ser calificada, en un momento dado, como verdadera o falsa:

Definición 1.1: Evento

Proposición lógica cuya calificación (verdadera o falsa) depende del resultado de un fenómeno o experimento aleatorio.

Decimos que un evento \mathcal{E} *ocurre* si la proposición lógica que lo define resulta verdadera, una vez que se conoce el resultado del fenómeno o experimento aleatorio al cual se refiere, y que *no ocurre* si dicha proposición resulta ser falsa. Como la ocurrencia o no de eventos depende del resultado del fenómeno o experimento aleatorio, que es incierto a priori, se denomina *espacio muestral* al conjunto Ω de todos los resultados posibles u observables de dicho experimento o fenómeno aleatorio. Eso permite asociar a cada evento \mathcal{E} un subconjunto de $E \subseteq \Omega$ que contiene únicamente aquellos elementos de Ω que de observarse nos permitirían afirmar que el evento \mathcal{E} ha ocurrido, es decir:

$$\mathcal{E} \longrightarrow E = \{\omega \in \Omega : \mathcal{E} \text{ es verdadera}\}.$$

Definición 1.2: Espacio de Probabilidad

Representación matemática de un fenómeno o experimento aleatorio mediante una terna $(\Omega, \mathfrak{E}, \mathbb{P})$ cuyos elementos significan lo siguiente:

- Ω = conjunto de resultados posibles u observables del experimento o fenómeno aleatorio en cuestión,
- \mathfrak{E} = colección de subconjuntos de Ω que representan eventos, y
- $\mathbb{P} : \mathfrak{E} \rightarrow [0, 1]$ una función^a que cuantifica el grado de incertidumbre sobre la ocurrencia de cualquier evento representado por un elemento de \mathfrak{E} .

^aDe hecho, una *medida* de probabilidad, desde el punto de vista de la *Teoría de la Medida*. Consultese, por ejemplo: Billingsley, *Probability and Measure*.

La identificación de los distintos elementos de Ω proviene de analizar directamente el experimento o fenómeno aleatorio en cuestión. Respecto a los eventos a considerar en la colección \mathfrak{E} , denominada también *espacio de eventos*, pueden ser todos aquellos (muchos o pocos) que resulten de interés en un momento dado, pero consecuencia de la definición de evento es posible demostrar que cualquier cantidad numerable de operaciones de conjuntos

con elementos de \mathfrak{E} da como resultado un subconjunto de Ω que también representa un evento, y por tanto es también elemento de \mathfrak{E} . Esto se relaciona con el siguiente concepto de la teoría de conjuntos:

Definición 1.3: Sigma álgebra

Sea \mathcal{U} un conjunto no vacío. Una colección \mathfrak{B} de subconjuntos de \mathcal{U} es un σ -álgebra si cumple las siguientes propiedades:

- $\emptyset \in \mathfrak{B}$.
- Si $B \in \mathfrak{B}$ entonces $\mathcal{U} \setminus B \in \mathfrak{B}$.
- Si $B_1, B_2, \dots \in \mathfrak{B}$ entonces

$$\bigcup_{n=1}^{\infty} B_n \in \mathfrak{B}.$$

A partir de la definición anterior, mediante conocidas propiedades de la teoría de conjuntos es posible demostrar que el resultado de cualquier cantidad numerable de operaciones de conjuntos con elementos de un σ -álgebra es siempre un elemento de dicho σ -álgebra, incluyendo el conjunto a partir del cual se define. Con el concepto anterior es inmediato demostrar el siguiente resultado:

Proposición 1.4

Para cualquier espacio de espacio de probabilidad $(\Omega, \mathfrak{E}, \mathbb{P})$ el espacio de eventos \mathfrak{E} es un σ -álgebra de subconjuntos de su espacio muestral Ω .

Definición 1.5: Medida de probabilidad

Para cualquier espacio de probabilidad $(\Omega, \mathfrak{E}, \mathbb{P})$ la función o *medida* de probabilidad $\mathbb{P} : \mathfrak{E} \rightarrow [0, 1]$ cumple las siguientes propiedades:

- $\mathbb{P}(\emptyset) = 0$.
- Si $E_1, E_2, \dots \in \mathfrak{E}$ representan *eventos mutuamente excluyentes*, esto es que $E_j \cap E_k = \emptyset$ para todo $j \neq k$, entonces

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

A partir de la definición anterior se demuestran diversas propiedades que debe cumplir \mathbb{P} en cualquier espacio de probabilidad, entre ellas $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) -$

$\mathbb{P}(E_1 \cap E_2)$, si $E_1 \subseteq E_2$ entonces $\mathbb{P}(E_2 \setminus E_1) = \mathbb{P}(E_2) - \mathbb{P}(E_1)$, entre muchas otras.¹ La siguiente propiedad será de particular relevancia más adelante:

Proposición 1.6

Sea $(\Omega, \mathcal{E}, \mathbb{P})$ un espacio de probabilidad, y sean $E_1, E_2 \in \mathcal{E}$. Entonces:

$$\max\{\mathbb{P}(E_1) + \mathbb{P}(E_2) - 1, 0\} \leq \mathbb{P}(E_1 \cap E_2) \leq \min\{\mathbb{P}(E_1), \mathbb{P}(E_2)\}. \quad (1.1)$$

Para cada experimento o fenómeno aleatorio particular, la especificación de la medida de probabilidad $\mathbb{P} : \mathcal{E} \rightarrow [0, 1]$ depende de la información y comprensión con que se cuente sobre dicho experimento o fenómeno y, por lo tanto, siempre existirá algún grado de subjetividad involucrado. Otro aspecto a considerar es que la especificación de \mathbb{P} no es necesariamente inmutable, ya que corresponde justamente a información con la que se cuenta en un momento dado en el tiempo, pero a la que, con el paso del tiempo, podríaadirse nueva información que no había sido considerada, y que puede llevar a modificar la medida de probabilidad inicial, por una medida de probabilidad más actualizada.

Afirmación: Toda medida de probabilidad es condicional

La medida de probabilidad \mathbb{P} siempre es *condicional* en la información con la cual se especifica, y si dicha información cambia o se actualiza, la medida de probabilidad también debe actualizarse.^a

^aPara una discusión a fondo, consúltese Lindley, «The philosophy of statistics»

Ejemplo 1. Lanzamientos de una moneda aparentemente equilibrada.

Consideremos el experimento aleatorio de lanzar un volado una sola vez con una moneda, a cuyas caras nos referiremos como “águila” y “sol”, y que por comodidad denotaremos con las letras a y s . Decimos que el resultado de un lanzamiento es “águila” si ya en el piso la cara que se aprecia es justamente “águila”.

En este caso el espacio muestral podría denotarse como $\Omega = \{a, s\}$. Nótese que se está aplicando a priori el supuesto de que no es factible que al lanzar la moneda y caer al piso pudiera quedar parada, lo cual sería un tercer resultado posible a considerar, pero que quizás se considera que tiene una probabilidad tan baja que se toma la decisión a priori de descartarlo. Esto constituye una primera decisión subjetiva apelando a una apreciación aparentemente razonable. En este caso, dos eventos que resultan de interés son:

$$\begin{aligned}\mathcal{E}_1 &\equiv \text{el resultado del lanzamiento es “águila”,} \\ \mathcal{E}_2 &\equiv \text{el resultado del lanzamiento es “sol”,}\end{aligned}$$

y por tanto los subconjuntos de Ω que los representan serían $E_1 = \{a\}$ y $E_2 = \{s\}$, respectivamente. Nótese que la colección $\{E_1, E_2\}$ no cumple con la Definición 1.3 de σ -álgebra, pero es inmediato identificar qué elementos le faltan a dicha colección para serlo:

¹Para un análisis exhaustivo consúltense Domínguez-Martínez, *Diseño y análisis de modelos de probabilidad*.

\emptyset y Ω , que de hecho son siempre elementos de cualquier σ -álgebra de subconjuntos de Ω , y que en este caso se justifican también por el hecho de que $E_1 \cup E_2 = \Omega$ y $E_1 \cap E_2 = \emptyset$. Recordemos que Ω representa un evento seguro con probabilidad $\mathbb{P}(\Omega) = 1$, ya que el evento $\mathcal{E}_1 \vee \mathcal{E}_2$ equivalente a “el resultado es águila o sol” es evidentemente algo seguro (bajo el supuesto inicial de que la moneda no podría quedar parada). El evento $\mathcal{E}_1 \wedge \mathcal{E}_2$ es equivalente a “el resultado es águila y sol” lo cual es imposible de forma simultánea. Así, un σ -álgebra de subconjuntos de Ω que incluya la representación de los eventos \mathcal{E}_1 y \mathcal{E}_2 sería:

$$\mathfrak{E} = \{\emptyset, \Omega, E_1, E_2\} = \{\emptyset, \Omega, E_1, \Omega \setminus E_1\},$$

en donde notemos que se aplicó el hecho de que $E_2 = \Omega \setminus E_1$. Nótese que los eventos \mathcal{E}_1 y \mathcal{E}_2 son mutuamente excluyentes y, por tanto, $E_1 \cap E_2 = \emptyset$.

Teniendo ya identificados el espacio muestral Ω y un espacio de eventos \mathfrak{E} , solo falta especificar una medida de probabilidad $\mathbb{P} : \mathfrak{E} \rightarrow [0, 1]$ que nos permita calcular la probabilidad de ocurrencia de cualquiera de los eventos involucrados. De entrada, por la Definición 1.5, cualquier medida de probabilidad debe cumplir $\mathbb{P}(\emptyset) = 0$, y como consecuencia de ella también $\mathbb{P}(\Omega) = 1$. Por lo que solo se requiere, en este caso particular, deducir fórmulas para $\mathbb{P}(E_1)$ y $\mathbb{P}(E_2)$. Como $E_2 = \Omega \setminus E_1$, por propiedades que se deducen de las medidas de probabilidad $\mathbb{P}(E_2) = 1 - \mathbb{P}(E_1)$, así que en este ejemplo basta poder deducir $\mathbb{P}(E_1)$ para que la medida de probabilidad quede completamente especificada sobre cualquier elemento de su dominio \mathfrak{E} .

Con la información inicial que se ha mencionado, denotaremos precisamente mediante \mathbb{P}_0 a la medida de probabilidad que la presente. Pero la medida de probabilidad \mathbb{P}_0 no puede ser cualquiera que satisfaga la Definición 1.5, sino una que además refleje la información y características que inicialmente se conocen del experimento aleatorio en cuestión. Si lo único que sabemos de la moneda es que luce como una moneda usual que al lanzar un volado tiene la misma probabilidad de resultar en “águila” o “sol”, pues consecuentes con ello asignaríamos $\mathbb{P}_0(E_1) = 0.5$ (supuesto de moneda equilibrada) y con ello el espacio de probabilidad inicial $(\Omega, \mathfrak{E}, \mathbb{P}_0)$ quedaría completamente especificado. Nótese que esta forma de especificar \mathbb{P}_0 es condicional en la información que se tiene, y en apreciaciones subjetivas razonadas.

Ahora supongamos que obtenemos información adicional: se nos permite lanzar la moneda, digamos unas 10 veces, y obtenemos los siguientes resultados:

$$(a, a, s, a, s, s, s, a, s, a).$$

Lo anterior es información con la que no se contaba al inicio, y que por tanto la medida de probabilidad \mathbb{P}_0 no contempló, así que procede actualizarla a una nueva medida de probabilidad \mathbb{P}_1 que sí tome en cuenta esta nueva información. En este caso, como se obtuvo igual número de águilas que de soles, la apreciación sería que esto, de hecho, coincide con la apreciación inicial de que se trata de una moneda equilibrada y, por tanto, la actualización consistiría en dejar la misma que se tenía inicialmente, esto es $\mathbb{P}_1 = \mathbb{P}_0$.

Pero el que en diez lanzamientos se obtenga el mismo número de águilas que de soles no es información contundente y definitiva de que se trate de una moneda equilibrada. Bien podría ocurrir lo mismo con una moneda ligeramente cargada a favor de águila, pero que

en 10 lanzamientos no alcanza a notarse claramente. Algo análogo podría ocurrir con una moneda equilibrada, al lanzarla 10 veces podrían no obtenerse el mismo número de águilas que de soles. Es más, bastaría lanzar una moneda equilibrada un número impar de veces, y con toda seguridad nunca se podría obtener el mismo número de águilas que de soles.

Supongamos ahora que nos permiten lanzar la moneda un total de mil veces, y que en tal caso obtenemos 553 águilas y 447 soles. Es un sencillo ejercicio probabilístico verificar que la probabilidad de obtener 553 o más águilas con una moneda equilibrada es aproximadamente igual a 0.00044, y podría con ello decidirse reemplazar el supuesto de moneda equilibrada con el de una moneda con mayor probabilidad de águila que de sol, y actualizar a una medida de probabilidad \mathbb{P}_2 utilizando una estimación puntual de la probabilidad de águila, por ejemplo $\mathbb{P}_2(E_1) = 553/1000 = 0.553$. De los supuestos iniciales solo sobrevive el de dos resultados posibles (descartando que la moneda quede parada en el piso). Hasta el momento hemos tenido una evolución de la medida de probabilidad asociada a un mismo experimento aleatorio, conforme se va teniendo más y más información:

$$\mathbb{P}_0 \longrightarrow \mathbb{P}_1 \longrightarrow \mathbb{P}_2 \longrightarrow \dots$$

ejemplificando cómo **una medida de probabilidad siempre es condicional** en cierto grado de información, y que conforme se obtiene más información debe revisarse y, de ser el caso, actualizarse a una distinta, conscientes de que en un futuro podría llegar más información que implique continuar actualizándola.

Las fuentes de información que nos permiten actualizar una medida de probabilidad son diversas: nuevas apreciaciones subjetivas derivadas de la experiencia o conocimiento del experimento o fenómeno aleatorio en cuestión, recopilación de cierto tipo de mediciones, entre otras. Una fuente común de ello es la información de que determinado evento ha ocurrido, o bien suponer que algún evento ha ocurrido (análisis de un posible escenario), y en tal caso surge la pregunta de cómo proceder en un caso así a la actualización de la medida de probabilidad.

Si en determinado momento tenemos un espacio de probabilidad con una medida de probabilidad inicial, digamos $(\Omega, \mathfrak{E}, \mathbb{P}_0)$, y posteriormente a la especificación de \mathbb{P}_0 ocurre un evento representado por $A \in \mathfrak{E}$ ¿cómo procedemos a actualizar la medida de probabilidad? Denotemos $\mathbb{P}_A : \mathfrak{E} \rightarrow [0, 1]$ a la medida de probabilidad que actualiza a \mathbb{P}_0 ante la ocurrencia de A . Esquemáticamente tenemos:

$$(\Omega, \mathfrak{E}, \mathbb{P}_0) \xrightarrow{A \in \mathfrak{E}} (\Omega, \mathfrak{E}, \mathbb{P}_A = ?)$$

Para cualquier $E \in \mathfrak{E}$ necesariamente la actualización de $\mathbb{P}_0(E)$ a $\mathbb{P}_A(E)$ depende de \mathbb{P}_0 , A y E , esto es, se requiere una transformación Ψ tal que:

$$\mathbb{P}_A(E) = \Psi(\mathbb{P}_0, A, E), \quad \text{para todo } E \in \mathfrak{E},$$

y para motivar la forma de definir Ψ analicemos las siguientes consideraciones:

- Para ser medida de probabilidad, por la Definición 1.5 se debe cumplir que $\mathbb{P}_A(\emptyset) = 0$.

- Aún cuando bajo la medida de probabilidad inicial se tuviera que $\mathbb{P}_0(A) < 1$, ante la ocurrencia del evento que representa A , preguntarse nuevamente por la ocurrencia del evento que representa A implica aludir a un evento seguro, esto es, necesariamente se debe cumplir que $\mathbb{P}_A(A) = 1$.
- Para todo $E \in \mathfrak{E}$ que satisfaga $A \subseteq E$ necesariamente $\mathbb{P}_A(E) = 1$ ya que por la propiedad monótona de las medidas de probabilidad se tiene que $\mathbb{P}_A(A) \leq \mathbb{P}_A(E)$.
- Para todo $E \in \mathfrak{E}$ que satisfaga $E \cap A = \emptyset$ necesariamente $\mathbb{P}_A(E) = 0$ ya que E y A representan eventos mutuamente excluyentes: la ocurrencia del evento que representa A implica la no ocurrencia del evento que representa E , y por tanto se convierte en un evento imposible.
- Solo falta el caso de aquellos $E \in \mathfrak{E}$ que satisfagan $\emptyset \neq E \cap A \subset A$ (inclusión propia, esto es $E \cap A \neq A$). Nuevamente por la propiedad monótona de las medidas de probabilidad, bajo la medida de probabilidad inicial tenemos que $\mathbb{P}_0(E \cap A) \leq \mathbb{P}_0(A)$. Si el evento que representa A ocurre, esto implica que $\mathbb{P}_0(A) > 0$, y ante ello la ocurrencia del evento que representa E será por causa de lo que tiene en común con A , esto es $E \cap A$, y dicha parte tiene una probabilidad inicial $\mathbb{P}_0(E \cap A)$, que en comparación con el nuevo marco de referencia $\mathbb{P}_0(A)$ permite calcular la proporción respecto a este último, esto es:

$$0 < \frac{\mathbb{P}_0(E \cap A)}{\mathbb{P}_0(A)} \leq 1.$$

Definición 1.7: Actualización de una medida de probabilidad

Sea $(\Omega, \mathfrak{E}, \mathbb{P}_0)$ un espacio de probabilidad. Posterior a la ocurrencia de un evento representado por $A \in \mathfrak{E}$ la *actualización* de la probabilidad inicial \mathbb{P}_0 en una medida de probabilidad actualizada ante la ocurrencia del evento representado por A se denota y define:

$$\mathbb{P}_A(E) := \frac{\mathbb{P}_0(E \cap A)}{\mathbb{P}_0(A)}, \quad \text{para todo } E \in \mathfrak{E}.$$

Es inmediato verificar que esta definición satisface las consideraciones que le preceden, así como la Definición 1.5. La referencia común a $\mathbb{P}_A(E)$ es la de *probabilidad condicional* del evento representado por E dada la ocurrencia del evento representado por A , y la notación usual es $\mathbb{P}(E | A)$, con el peligro de que se pierda de vista que lo que realmente está ocurriendo es un cambio (actualización) de medida de probabilidad, de \mathbb{P}_0 a \mathbb{P}_A .

Podemos actualizar una medida de probabilidad tantas veces como nueva información deba considerarse. Si ante la ocurrencia de un evento representado por $A \in \mathfrak{E}$ una medida de probabilidad inicial \mathbb{P}_0 ya fue actualizada a \mathbb{P}_A , y posteriormente a ello ocurre un evento representado por $B \in \mathfrak{E}$, entonces tendremos ahora que actualizar \mathbb{P}_A a $(\mathbb{P}_A)_B$ aplicando nuevamente la Definición 1.7:

$$\forall E \in \mathfrak{E}: (\mathbb{P}_A)_B(E) = \frac{\mathbb{P}_A(E \cap B)}{\mathbb{P}_A(B)} = \frac{\frac{\mathbb{P}_0((E \cap B) \cap A)}{\mathbb{P}_0(A)}}{\frac{\mathbb{P}_0(B \cap A)}{\mathbb{P}_0(A)}} = \frac{\mathbb{P}_0(E \cap (A \cap B))}{\mathbb{P}_0(A \cap B)} = \mathbb{P}_{A \cap B}(E),$$

y por la conmutatividad $A \cap B = B \cap A$ lo anterior implica también:

$$(\mathbb{P}_A)_B = (\mathbb{P}_B)_A = \mathbb{P}_{A \cap B}$$

Lo anterior implica que la actualización de una medida de probabilidad inicial \mathbb{P}_0 ante la ocurrencia sucesiva de los eventos representados por A y B puede hacerse, o bien de forma secuencial:

$$(\Omega, \mathfrak{E}, \mathbb{P}_0) \xrightarrow{A \in \mathfrak{E}} (\Omega, \mathfrak{E}, \mathbb{P}_A) \xrightarrow{B \in \mathfrak{E}} (\Omega, \mathfrak{E}, (\mathbb{P}_A)_B),$$

o bien de forma acumulada ante la ocurrencia de la conjunción de los eventos representados por A y B , representado a su vez por $A \cap B$:

$$(\Omega, \mathfrak{E}, \mathbb{P}_0) \xrightarrow{A, B \in \mathfrak{E}} (\Omega, \mathfrak{E}, \mathbb{P}_{A \cap B}),$$

notando que, de forma acumulada, el orden de ocurrencia de los eventos no afecta el resultado final.

El proceso de actualización de una medida de probabilidad inicial \mathbb{P} en otra \mathbb{P}_A por la ocurrencia de un evento representado por A , implicará cambios en la cuantificación de la incertidumbre sobre otros eventos, esto es que $\mathbb{P}(E_1) \neq \mathbb{P}_A(E_1)$ en algunos casos, pero también pueden existir eventos específicos para los cuales dicha actualización no cambie la incertidumbre sobre ellos, esto es que $\mathbb{P}(E_2) = \mathbb{P}_A(E_2)$. A la colección de eventos representados por:

$$\{E \in \mathfrak{E} : \mathbb{P}_A(E) = \mathbb{P}(E)\}$$

se les denomina *eventos independientes* de la ocurrencia del evento representado por A .

Definición 1.8: Eventos independientes y dependientes

Sea $(\Omega, \mathfrak{E}, \mathbb{P})$ un espacio de probabilidad, y sea $A \in \mathfrak{E}$. Si $E \in \mathfrak{E}$ satisface $\mathbb{P}_A(E) = \mathbb{P}(E)$ se dice que E representa un *evento independiente* del evento que representa A . En caso contrario, se dice que E representa un *evento dependiente* del que representa A .

Consideremos dos eventos representados por $A, B \in \mathfrak{E}$ ambos con probabilidades distintas de cero. Entonces:

$$\begin{aligned} \mathbb{P}(B) = \mathbb{P}_A(B) &\Leftrightarrow \mathbb{P}(B) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} \\ &\Leftrightarrow \mathbb{P}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\ &\Leftrightarrow \mathbb{P}(A) = \mathbb{P}_B(A), \end{aligned}$$

esto es, un evento representado por B es independiente del representado por A si y solo si el representado de A es independiente del representado por B , lo que hace innecesario considerar un orden al respecto, y derivado de lo anterior, se dice simplemente que dos eventos representados por A y B son *eventos independientes* si y solo si:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \quad (1.2)$$

En el caso de que se trate de eventos dependientes, necesariamente se tiene que $\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B)$, ante lo cual surge la pregunta sobre qué tan distante puede llegar a estar $\mathbb{P}(A \cap B)$ del valor $\mathbb{P}(A)\mathbb{P}(B)$ que representa la independencia. Aplicando el resultado (1.1) obtenemos:

$$\max\{\mathbb{P}(A)+\mathbb{P}(B)-1, 0\} - \mathbb{P}(A)\mathbb{P}(B) \leq \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \leq \min\{\mathbb{P}(A), \mathbb{P}(B)\} - \mathbb{P}(A)\mathbb{P}(B),$$

en donde es inmediato verificar que es equivalente a:

$$-\min\{\mathbb{P}(A^c)\mathbb{P}(B^c), \mathbb{P}(A)\mathbb{P}(B)\} \leq \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \leq \min\{\mathbb{P}(A)\mathbb{P}(B^c), \mathbb{P}(A^c)\mathbb{P}(B)\},$$

donde la notación A^c se refiere al complemento de A respecto a Ω , es decir $\Omega \setminus A$. Con lo anterior, una forma de medir el grado de dependencia entre los eventos representados por A y B podría ser justamente la diferencia:

$$\delta(A, B) := \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B), \quad (1.3)$$

de donde tendríamos que los eventos que representan A y B son independientes si y solo si $\delta(A, B) = 0$. Pero si $\delta(A, B) \neq 0$ tenemos el caso de eventos dependientes, y el valor $\delta(A, B)$ es una forma de cuantificar el grado de dependencia, que puede reportar valores positivos o negativos en términos de las probabilidades individuales de A y B , ya que $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ y $\mathbb{P}(B^c) = 1 - \mathbb{P}(B)$:

$$-\min\{\mathbb{P}(A^c)\mathbb{P}(B^c), \mathbb{P}(A)\mathbb{P}(B)\} \leq \delta(A, B) \leq \min\{\mathbb{P}(A)\mathbb{P}(B^c), \mathbb{P}(A^c)\mathbb{P}(B)\}. \quad (1.4)$$

Las cotas para $\delta(A, B)$ en (1.4) pueden ser más o menos amplias, dependiendo de los valores $\mathbb{P}(A)$ y $\mathbb{P}(B)$. Nótese, por ejemplo, que basta con que una de estas probabilidades sea muy cercana a 1 o a 0 para que la cota superior se aproxime mucho a cero por la derecha, y que la cota inferior se aproxime mucho a cero por la izquierda, es decir, queda poco margen para que $\delta(A, B)$ se aleje de cero. Más aún, con que una sola de las probabilidades sea 0 o 1 obtenemos necesariamente $\delta(A, B) = 0$. Recordemos que las probabilidades extremas 0 y 1 representan certidumbre total, sobre la no ocurrencia del evento en caso del 0, y sobre la ocurrencia del evento en el caso de 1. En otras palabras, los eventos seguros o imposibles son siempre independientes de cualquier otro evento.

2. Variables aleatorias dependientes

Una vez especificado un espacio de probabilidad $(\Omega, \mathfrak{E}, \mathbb{P})$ con base en la información que se tiene, en un momento dado, sobre un fenómeno o experimento aleatorio, es posible definir distintas funciones que cuantifiquen aspectos de interés. Se trata de funciones que tienen por dominio al espacio muestral Ω y por codominio al conjunto de los números reales \mathbb{R} , y cuya regla de correspondencia se define de acuerdo con un interés específico. A estas funciones se les conoce como *variables aleatorias*.

Una vez definida una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ sobre un espacio de probabilidad $(\Omega, \mathfrak{E}, \mathbb{P})$, dado que existe incertidumbre sobre el resultado $\omega \in \Omega$ que se puede observar, esto se traduce en incertidumbre también sobre el número real $X(\omega)$ que puede reportar la variable aleatoria X , y por tanto interesa hacer cálculo de probabilidades sobre los valores

que pueda reportar dicha variable aleatoria, ya sea en relación con un valor específico o bien sobre un conjunto de valores. Esto es, calcular la probabilidad de que $X(\omega) = x$ para un valor específico x , o bien la probabilidad de que $X(\omega) \in B$ para algún subconjunto $B \subseteq \mathbb{R}$.

Definición 2.1: Variable Aleatoria

Dado un espacio de probabilidad $(\Omega, \mathfrak{E}, \mathbb{P})$ una **variable aleatoria** es una función $X : \Omega \rightarrow \mathbb{R}$ tal que para *casi cualquier* $B \subseteq \mathbb{R}$ la imagen inversa $X^{(-1)}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ representa a algún evento, es decir tal que $X^{(-1)}(B) \in \mathfrak{E}$.

El uso de la expresión *casi cualquier* $B \subseteq \mathbb{R}$ se refiere a que existen subconjuntos de \mathbb{R} cuya imagen inversa no representa evento alguno, en *teoría de la medida* se les conoce como subconjuntos *no medibles*, probabilísticamente hablando en este caso particular. No abundaremos más en ello porque eso está más allá del objetivo del presente texto, así que de momento nos restringiremos a aquellos subconjuntos de \mathbb{R} que sí son probabilísticamente medibles, colección que se conoce como σ -álgebra de Borel² $\mathfrak{B}(\mathbb{R})$, y a cuyos elementos se les conoce como *conjuntos boreelianos*, que para efectos prácticos resulta más que suficiente para el cálculo de probabilidades.

Al definir una variable aleatoria X sobre un espacio de probabilidad $(\Omega, \mathfrak{E}, \mathbb{P})$ se crea un nuevo espacio de probabilidad $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \mathbb{P}_X)$ donde $\mathbb{P}_X : \mathfrak{B}(\mathbb{R}) \rightarrow [0, 1]$ es la *medida de probabilidad inducida* por la variable aleatoria X , de modo que para cualquier $B \in \mathfrak{B}(\mathbb{R})$ el número real $\mathbb{P}_X(B)$ represente justamente la probabilidad de que la variable aleatoria X reporte un número real $X(\omega)$ que pertenezca a B . Esto último ocurrirá si el resultado del fenómeno o experimento aleatorio es alguno de los elementos del subconjunto $\{\omega \in \Omega : X(\omega) \in B\} = X^{(-1)}(B) \in \mathfrak{E}$, que al representar un evento es posible calcular su probabilidad mediante \mathbb{P} , es decir:

$$\mathbb{P}_X(B) = \mathbb{P}(X^{(-1)}(B)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}). \quad (1.5)$$

En lugar de escribir $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$ se prefiere la notación simplificada $\mathbb{P}(X \in B)$. Se deja como ejercicio demostrar que \mathbb{P}_X cumple la Definición 1.5 con $\Omega = \mathbb{R}$.

Ejemplo 2. Variable Aleatoria Binomial.

Se denominan *Experimentos Aleatorios Binomiales* cualesquiera experimentos o fenómenos aleatorios que cumplen con los siguientes supuestos:

- Una observación o realización del mismo consiste en $n \in \{1, 2, \dots\}$ repeticiones de un procedimiento que en cada repetición solo arroja uno de dos resultados posibles, que denominaremos “éxito” (e) y “fracaso” (f).
- La probabilidad de éxito en cada repetición es un valor constante $0 < \theta < 1$.
- El resultado de cualquier repetición no influye en las otras (independencia entre repeticiones).

²Émile Borel (1871–1956) matemático y político francés.

En este caso, cualquier resultado de este experimento aleatorio consiste en una n -ada de éxitos y fracasos, por lo que el espacio muestral puede denotarse mediante:

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_j \in \{e, f\}\}.$$

Se trata de una espacio muestral finito donde $|\Omega| = 2^n$, pero no es en general equiprobable, salvo el caso particular $\theta = \frac{1}{2}$. Por tratarse de un espacio muestral numerable, para definir la medida de probabilidad correspondiente $\mathbb{P} : \mathfrak{E} \rightarrow [0, 1]$ para cualquier $E \in \mathfrak{E}$, basta poder deducir la probabilidad correspondiente a los eventos simples,^a esto es $\mathbb{P}(\{(\omega_1, \dots, \omega_n)\})$, con base en la descripción de este experimento aleatorio. Consideraremos los eventos $\mathcal{E}_1, \dots, \mathcal{E}_n$ definidos como:

$$\mathcal{E}_k \equiv \text{"Se obtiene éxito en la } k\text{-ésima repetición"}, \quad k \in \{1, \dots, n\},$$

los cuales quedan representados por los subconjuntos de Ω siguientes:

$$\begin{aligned} E_1 &= \{(e, \omega_2, \omega_3, \dots, \omega_n) : \omega_j \in \{e, f\}\}, \\ E_2 &= \{(\omega_1, e, \omega_3, \dots, \omega_n) : \omega_j \in \{e, f\}\}, \\ &\vdots && \vdots && \vdots \\ E_n &= \{(\omega_1, \dots, \omega_{n-1}, e) : \omega_j \in \{e, f\}\}. \end{aligned}$$

Cada uno de estos subconjuntos tiene $|E_k| = 2^{n-1}$ elementos, por lo que con $n \geq 2$ no representan eventos simples, pero la descripción del experimento nos proporciona información sobre la probabilidad que les corresponde: $\mathbb{P}(E_k) = \theta$ para todo $k \in \{1, \dots, n\}$. Más aún, notemos que cualquier evento simple puede expresarse en términos de una conjunción de los eventos $\mathcal{E}_1, \dots, \mathcal{E}_n$ o algunas de sus negaciones. Por ejemplo, para un caso particular $n = 5$, el evento simple representado por $\{(e, f, e, e, f)\}$ es resultado de la siguiente conjunción de eventos:

$$\mathcal{E}_1 \wedge (\neg \mathcal{E}_2) \wedge \mathcal{E}_3 \wedge \mathcal{E}_4 \wedge (\neg \mathcal{E}_5) \longrightarrow E_1 \cap (\Omega \setminus E_2) \cap E_3 \cap E_4 \cap (\Omega \setminus E_5) = \{(e, f, e, e, f)\},$$

y también por la descripción del experimento, en el sentido de que el resultado de cualquier repetición no influye en las otras repeticiones, esto implica la independencia de los eventos $\mathcal{E}_1, \dots, \mathcal{E}_n$ y por tanto:

$$\begin{aligned} \mathbb{P}(\{(e, f, e, e, f)\}) &= \mathbb{P}(E_1 \cap (\Omega \setminus E_2) \cap E_3 \cap E_4 \cap (\Omega \setminus E_5)) \\ &= \mathbb{P}(E_1)\mathbb{P}(\Omega \setminus E_2)\mathbb{P}(E_3)\mathbb{P}(E_4)\mathbb{P}(\Omega \setminus E_5) \\ &= \theta \cdot (1 - \theta) \cdot \theta \cdot \theta \cdot (1 - \theta) = \theta^3(1 - \theta)^2, \end{aligned}$$

y en general, para cualquier valor de n la fórmula general para cualquier n -ada que representa un evento simple:

$$\mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = \theta^r(1 - \theta)^{n-r}, \quad r = \sum_{j=1}^n \mathbf{1}_{\{\omega_j = e\}}, \quad (1.6)$$

donde r representa el número de éxitos (e) en la n -ada $(\omega_1, \dots, \omega_n)$, y por tanto necesariamente $n - r$ es el número de fracasos (f). Al tener una fórmula general para los eventos simples en un espacio muestral numerable, la probabilidad de cualquier evento no simple, representado por algún $E \in \mathfrak{E}$, se calcula mediante:

$$\mathbb{P}(E) = \mathbb{P}\left(\bigcup_{(\omega_1, \dots, \omega_n) \in E} \{(\omega_1, \dots, \omega_n)\}\right) = \sum_{(\omega_1, \dots, \omega_n) \in E} \mathbb{P}(\{(\omega_1, \dots, \omega_n)\}),$$

en donde la suma se justifica por el hecho de que E fue expresado como una unión disjunta numerable de conjuntos que representan eventos simples, aplicando la Definición 1.5. Con todo lo anterior, el espacio de probabilidad $(\Omega, \mathfrak{E}, \mathbb{P})$ para un experimento (o fenómeno) aleatorio binomial ha quedado completamente especificado.

Sobre dicho espacio de probabilidad, podemos definir cuantas variables aleatorias deseemos, mientras cumplan con la Definición 2.1. Una que suele ser de interés es aquella $X : \Omega \rightarrow \mathbb{R}$ que reporta el número de éxitos en las n repeticiones (sin importar el orden), y que es a la que comúnnemente se conoce como *variable aleatoria binomial*:

$$X((\omega_1, \dots, \omega_n)) := \sum_{j=1}^n \mathbf{1}_{\{\omega_j = e\}}, \quad (\omega_1, \dots, \omega_n) \in \Omega, \quad (1.7)$$

y que, por tanto, reporta necesariamente un valor del conjunto $\text{Ran } X = \{0, 1, \dots, n\}$. Entonces para cualquier $x \in \{0, 1, \dots, n\}$ la probabilidad de que la variable aleatoria X reporte el valor $X((\omega_1, \dots, \omega_n)) = x$ se puede calcular mediante (1.5) con $B = \{x\}$. Obtengamos primero su imagen inversa, que denotaremos:

$$E_x := X^{(-1)}(\{x\}) = \left\{ (\omega_1, \dots, \omega_n) \in \Omega : \sum_{j=1}^n \mathbf{1}_{\{\omega_j = e\}} = x \right\}.$$

Notemos que todos los elementos $(\omega_1, \dots, \omega_n) \in E_x$ son n -adas con exactamente x éxitos, y por tanto $n - x$ fracasos, así que aplicando (1.6) obtenemos que $\mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = \theta^x(1 - \theta)^{n-x}$ para todo $(\omega_1, \dots, \omega_n) \in E_x$, de donde obtenemos:

$$\mathbb{P}_X(\{x\}) = \mathbb{P}(E_x) = \sum_{(\omega_1, \dots, \omega_n) \in E_x} \mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = |E_x| \theta^x(1 - \theta)^{n-x}.$$

$|E_x|$ es el número de formas distintas en las que podemos acomodar x éxitos en los n lugares de cualquier n -ada $(\omega_1, \dots, \omega_n) \in E_x$ (los lugares restantes son automáticamente para los $n - x$ fracasos), y esto es justamente el número de combinaciones de n en x , denotado por el coeficiente binomial $\binom{n}{x} = \frac{n!}{x!(n-x)!}$, con lo que concluimos que la fórmula general para calcular la probabilidad de que la variable aleatoria X reporte exactamente x éxitos en n repeticiones, para cualquier $x \in \{0, 1, \dots, n\}$, es la siguiente:

$$\mathbb{P}_X(\{x\}) \equiv \mathbb{P}(X = x) = \binom{n}{x} \theta^x(1 - \theta)^{n-x} \mathbf{1}_{\{0, 1, \dots, n\}}(x). \quad (1.8)$$

Implícitamente $\mathbb{P}(X = x) = 0$ para toda $x \notin \{0, 1, \dots, n\}$ por lo que para cualquier boreliano $B \in \mathfrak{B}(\mathbb{R})$ tenemos que:

$$\mathbb{P}(X \in B) \equiv \mathbb{P}_X(B) = \mathbb{P}_X(B \cap \{0, 1, \dots, n\}) = \sum_{x=0}^n \mathbb{P}(X = x) \mathbf{1}_{\{x \in B\}}. \quad (1.9)$$

^aUn *evento simple* es aquel que puede representarse mediante un conjunto con un solo elemento del espacio muestral, es decir, que es de la forma $\{\omega\}$ para algún $\omega \in \Omega$.

Ejemplo 3. Otra variable aleatoria sobre un experimento binomial.

La llamada *variable aleatoria binomial* X discutida en el Ejemplo 2 anterior no es la única variable aleatoria que puede definirse sobre un mismo experimento o fenómeno aleatorio binomial, podríamos definir también, por ejemplo, $Y : \Omega \rightarrow \mathbb{R}$ que reporte el valor 1 si el resultado de las n repeticiones fue siempre el mismo, y 0 en caso contrario:

$$Y((\omega_1, \dots, \omega_n)) := \mathbf{1}_{\{\omega_1 = \omega_2 = \dots = \omega_n\}}. \quad (1.10)$$

Como en este caso $\text{Ran } Y = \{0, 1\}$ tenemos solo dos valores posibles, basta con obtener la fórmula para calcular $\mathbb{P}(Y = 1)$, y el otro caso se calcula por el complemento, es decir $\mathbb{P}(Y = 0) = 1 - \mathbb{P}(Y = 1)$. Procediendo de forma análoga al Ejemplo 2 obtenemos:

$$\begin{aligned} \mathbb{P}(Y = 1) &\equiv \mathbb{P}_Y(\{1\}) = \mathbb{P}(Y^{(-1)}(\{1\})) \\ &= \mathbb{P}(\{(\omega_1, \dots, \omega_n) \in \Omega : Y((\omega_1, \dots, \omega_n)) = 1\}) \\ &= \mathbb{P}(\{(e, e, \dots, e), (f, f, \dots, f)\}) \\ &= \mathbb{P}(\{(e, e, \dots, e)\}) + \mathbb{P}(\{(f, f, \dots, f)\}) \\ &= \theta^n + (1 - \theta)^n, \\ \mathbb{P}(Y = 0) &= 1 - \theta^n - (1 - \theta)^n. \end{aligned}$$

De hecho, si $n \geq 2$ resulta que Y es lo que se conoce como una *variable aleatoria Bernoulli*, en este caso con parámetro $\theta^n + (1 - \theta)^n$. Nótese que bajo el caso $n = 1$ se tiene que $\mathbb{P}(Y = 1) = 1$, por obvias razones.

Discusión de los Ejemplos 2 y 3

En los Ejemplos 2 y 3 anteriores se definieron dos variables aleatorias distintas sobre un mismo experimento aleatorio binomial, donde cada una cuantifica de forma distinta los posibles resultados del experimento aleatorio. Supongamos que $n \geq 2$. Una vez que se conoce un resultado de dicho experimento aleatorio, digamos $(\omega_1, \dots, \omega_n) \in \Omega$, la variable aleatoria X reporta el valor $X((\omega_1, \dots, \omega_n))$, y la variable aleatoria Y reporta el valor $Y((\omega_1, \dots, \omega_n))$.

Si aún no conocemos un resultado del experimento aleatorio, ya tenemos forma de hacer cálculo de probabilidades con cada una de ellas, de forma individual, en relación con los

posibles valores que cada una puede reportar:

$$\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \mathbf{1}_{\{0,1,\dots,n\}}(x), \quad (1.11)$$

$$\mathbb{P}_Y(\{y\}) = \mathbb{P}(Y = y) = \begin{cases} \theta^n + (1 - \theta)^n, & \text{si } y = 1, \\ 1 - \theta^n - (1 - \theta)^n, & \text{si } y = 0, \\ 0, & \text{en cualquier otro caso.} \end{cases} \quad (1.12)$$

Si solamente tenemos la información explicada en el Ejemplo 2, entonces sin conocer el resultado del experimento aleatorio, digamos *a priori*, la fórmula (1.11) nos permite calcular la probabilidad de que X reporte cada uno de los valores posibles que puede reportar, esto es $x \in \text{Ran } X = \{0, 1, \dots, n\}$.

Ahora supongamos que el experimento aleatorio ya se realizó, y aunque no conocemos el resultado, nos llega la información de que la variable aleatoria Y reportó el valor 1. Esta nueva información implica que debemos revisar, y de ser el caso actualizar, la medida de probabilidad original (o inicial) \mathbb{P}_X ante ello. La información de que $Y = 1$ implica que los valores $\{1, \dots, n-1\}$ ya quedan descartados para X pues $Y = 1$ significa que todos los elementos del resultado $(\omega_1, \dots, \omega_n) \in \Omega$ son iguales, y esto solo es posible si $X = 0$ o bien si $X = n$. Esto significa que para valores $x \in \{1, \dots, n-1\}$ que antes tenían una probabilidad positiva de acuerdo con (1.11) ahora tienen probabilidad cero.

Denotemos $\mathbb{P}_{X|Y=1}$ a la medida de probabilidad para X pero actualizada ante la nueva información de que $Y = 1$. ¿Cómo deducir la fórmula general para $\mathbb{P}_{X|Y=1}(x)$? De entrada ya quedó claro que para $x \in \{1, \dots, n-1\}$ se tiene ahora que $\mathbb{P}_{X|Y=1}(x) = 0$, y solo faltarían los casos $x = 0$ y $x = n$. Aquí es importante recordar que la notación $X = x$ en realidad hace referencia al evento representado por:

$$E_x = X^{(-1)}(\{x\}) = \left\{ (\omega_1, \dots, \omega_n) \in \Omega : \sum_{j=1}^n \mathbf{1}_{\{\omega_j = e\}} = x \right\} \in \mathfrak{E}, \quad x \in \{0, 1, \dots, n\},$$

y de forma análoga $Y = y$ hace referencia al evento representado por $G_y \in \mathfrak{E}$ donde:

$$G_y = Y^{(-1)}(y) = \begin{cases} \{(e, e, \dots, e), (f, f, \dots, f)\}, & \text{si } y = 1, \\ \Omega \setminus \{(e, e, \dots, e), (f, f, \dots, f)\}, & \text{si } y = 0. \end{cases}$$

Entonces, la ocurrencia del evento representado por $G_y \equiv \{Y = y\} \in \mathfrak{E}$ implica la actualización del espacio de probabilidad inicialmente obtenido para el experimento aleatorio binomial en el Ejemplo 2, en el sentido de la Definición 1.7, esto es:

$$(\Omega, \mathfrak{E}, \mathbb{P}) \xrightarrow{G_y \in \mathfrak{E}} (\Omega, \mathfrak{E}, \mathbb{P}_{G_y}),$$

y con ello, la probabilidad de los eventos representados por $E_x \equiv \{X = x\} \in \mathfrak{E}$, que originalmente era $\mathbb{P}(E_x) = \mathbb{P}(X = x)$ como en (1.11), ahora se actualiza a $\mathbb{P}_{G_y}(E_x)$, que de acuerdo con la Definición 1.7 resulta ser:

$$\mathbb{P}(E_x) \longrightarrow \mathbb{P}_{G_y}(E_x) = \frac{\mathbb{P}(E_x \cap G_y)}{\mathbb{P}(G_y)}. \quad (1.13)$$

Para el caso $G_1 \equiv \{Y = 1\} = \{(e, e, \dots, e), (f, f, \dots, f)\}$ tenemos que:

$$E_x \cap G_1 = \begin{cases} \{(f, f, \dots, f)\}, & \text{si } x = 0, \\ \{(e, e, \dots, e)\}, & \text{si } x = n, \\ \emptyset, & \text{si } x \in \{1, \dots, n-1\}, \end{cases}$$

a lo que aplicando (1.6) y (1.12) nos arroja:

$$\mathbb{P}_{G_1}(E_x) = \frac{\mathbb{P}(E_x \cap G_1)}{\mathbb{P}(G_1)} = \begin{cases} \frac{(1-\theta)^n}{\theta^n + (1-\theta)^n}, & \text{si } x = 0, \\ \frac{\theta^n}{\theta^n + (1-\theta)^n}, & \text{si } x = n, \\ 0, & \text{si } x \in \{1, \dots, n-1\}. \end{cases}$$

En la discusión anterior es importante destacar varios aspectos:

- Si se definen distintas variables aleatorias sobre un mismo experimento aleatorio, aunque de forma individual sea posible deducir las medidas de probabilidad que cada una induce, y con ello hacer cálculos de probabilidades con cada una por separado, el conocimiento del valor reportado por una de ellas implica *actualizar* las medidas de probabilidad de las otras.
- Este proceso de actualización involucra cálculo de probabilidades de intersecciones de eventos de las variables aleatorias involucradas, como en el numerador de (1.13):

$$E_x \cap G_y = \{\omega \in \Omega : X(\omega) = x\} \cap \{\omega \in \Omega : Y(\omega) = y\}$$

donde $\omega = (\omega_1, \dots, \omega_n)$, y cuya probabilidad, de forma simplificada se denota:

$$\mathbb{P}(E_x \cap G_y) \equiv \mathbb{P}(X = x, Y = y)$$

y se le conoce como *probabilidad conjunta* de las variables aleatorias X e Y . Esto es, el analizar cómo el valor reportado por una variable aleatoria influye en la otra, requiere de un concepto adicional, como se aborda a continuación:

Definición 2.2: Vector Aleatorio

Dado un espacio de probabilidad $(\Omega, \mathfrak{E}, \mathbb{P})$ un **vector aleatorio** es una función $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ tal que para *casi cualquier* $B \subseteq \mathbb{R}^n$ la imagen inversa $\mathbf{X}^{(-1)}(B) = \{\omega \in \Omega : \mathbf{X}(\omega) \in B\}$ representa a algún evento, es decir tal que $\mathbf{X}^{(-1)}(B) \in \mathfrak{E}$.

El uso de la expresión *casi cualquier* $B \subseteq \mathbb{R}^n$ es análogo a lo explicado respecto a la Definición 2.1.

Al definir un vector aleatorio \mathbf{X} sobre un espacio de probabilidad $(\Omega, \mathfrak{E}, \mathbb{P})$ se crea un nuevo espacio de probabilidad $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n), \mathbb{P}_{\mathbf{X}})$ donde $\mathbb{P}_{\mathbf{X}} : \mathfrak{B}(\mathbb{R}^n) \rightarrow [0, 1]$ es la *medida de probabilidad inducida* por el vector aleatorio \mathbf{X} , de modo que para cualquier $B \in \mathfrak{B}(\mathbb{R}^n)$

el número real $\mathbb{P}_X(B)$ represente justamente la probabilidad de que el vector aleatorio \mathbf{X} reporte un punto en \mathbb{R}^n que pertenezca a B . Esto último ocurrirá si el resultado del fenómeno o experimento aleatorio es alguno de los elementos del subconjunto $\{\omega \in \Omega : \mathbf{X}(\omega) \in B\} = \mathbf{X}^{(-1)}(B) \in \mathfrak{E}$, que al representar un evento es posible calcular su probabilidad mediante \mathbb{P} , es decir:

$$\mathbb{P}_X(B) = \mathbb{P}(\mathbf{X}^{(-1)}(B)) = \mathbb{P}(\{\omega \in \Omega : \mathbf{X}(\omega) \in B\}). \quad (1.14)$$

En lugar de escribir $\mathbb{P}(\{\omega \in \Omega : \mathbf{X}(\omega) \in B\})$ es costumbre escribir simplemente $\mathbb{P}(\mathbf{X} \in B)$. Se deja como ejercicio demostrar que \mathbb{P}_X cumple con la Definición 1.5 con $\Omega = \mathbb{R}^n$.

Para cada $\omega \in \Omega$ tenemos que $\mathbf{X}(\omega)$ es un punto en \mathbb{R}^n y por tanto existen n números reales tales que $\mathbf{X}(\omega) = (x_1, \dots, x_n)$. Más aún, al tratarse de una función $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ existen n funciones componentes X_1, \dots, X_n tales que $X_j : \Omega \rightarrow \mathbb{R}$ para $j \in \{1, \dots, n\}$ de modo que:

$$\mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)). \quad (1.15)$$

La demostración de lo siguiente se deja como ejercicio:

Proposición 2.3

Sea $(\Omega, \mathfrak{E}, \mathbb{P})$ un espacio de probabilidad. Una función $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ es vector aleatorio si y solo si cada una de sus funciones componentes (1.15) son variables aleatorias.

Por simplicidad, con frecuencia nos enfocaremos en vectores aleatorios de dimensión 2, en el entendido de que ideas análogas son válidas en dimensiones mayores. En tal caso utilizaremos la notación (X, Y) para referirnos a un vector aleatorio que tiene como variables aleatorias componentes a X e Y . Así, un vector aleatorio (X, Y) definido sobre un espacio de probabilidad $(\Omega, \mathfrak{E}, \mathbb{P})$ es un función $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ que induce un espacio de probabilidad $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2), \mathbb{P}_{X,Y})$ con *medida de probabilidad conjunta* $\mathbb{P}_{X,Y} : \mathfrak{B}(\mathbb{R}^2) \rightarrow [0, 1]$. Si $A, B \in \mathfrak{B}(\mathbb{R})$ entonces $A \times B \in \mathfrak{B}(\mathbb{R}^2)$, y en tal caso:

$$\begin{aligned} \mathbb{P}_{X,Y}(A \times B) &= \mathbb{P}((X, Y)^{(-1)}(A \times B)) = \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A \times B\}) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\} \cap \{\omega \in \Omega : Y(\omega) \in B\}) \\ &= \mathbb{P}(X^{(-1)}(A) \cap Y^{(-1)}(B)) \\ &\equiv \mathbb{P}(X \in A, Y \in B) \quad (\text{notación simplificada}). \end{aligned} \quad (1.16)$$

A partir de lo anterior, como en particular $\mathbb{R} \in \mathfrak{B}(\mathbb{R})$ entonces aplicando $B = \mathbb{R}$ en (1.16) obtenemos:

$$\mathbb{P}_{X,Y}(A \times \mathbb{R}) = \mathbb{P}(X^{(-1)}(A) \cap \Omega) = \mathbb{P}_X(A),$$

y análogamente $\mathbb{P}_{X,Y}(\mathbb{R} \times B) = \mathbb{P}_Y(B)$, esto es, a partir de la medida de probabilidad conjunta de un vector aleatorio es posible deducir las medidas de probabilidad individuales de las variables aleatorias componentes. A dicho procedimiento se le conoce como *marginalización* de la medida de probabilidad conjunta $\mathbb{P}_{X,Y}$ y por tanto a \mathbb{P}_X y \mathbb{P}_Y se les denomina *medidas de probabilidad marginales* de $\mathbb{P}_{X,Y}$. Esquemáticamente:

$$\mathbb{P}_{X,Y} \longrightarrow \mathbb{P}_X, \mathbb{P}_Y.$$

Pero el solo conocimiento de las marginales \mathbb{P}_X y \mathbb{P}_Y no es suficiente, en general, para reconstruir $\mathbb{P}_{X,Y}$. Esto puede verse aplicando (1.1) a (1.16) obteniendo con ello cotas para la medida de probabilidad conjunta, dadas las medidas de probabilidad marginales:

$$\max\{\mathbb{P}_X(A) + \mathbb{P}_Y(B) - 1, 0\} \leq \mathbb{P}_{X,Y}(A \times B) \leq \min\{\mathbb{P}_X(A), \mathbb{P}_Y(B)\}. \quad (1.17)$$

Esto significa que para un mismo par de medidas de probabilidad marginales \mathbb{P}_X y \mathbb{P}_Y puede existir una infinidad no numerable de medidas de probabilidad conjunta distintas que tengan exactamente las mismas marginales.

En la discusión de los Ejemplos 2 y 3 se motivó la necesidad de calcular medidas de probabilidad conjuntas para poder actualizar la medida de probabilidad de un variable aleatoria, dados los valores de otra. Esto motiva el siguiente concepto:

Definición 2.4: Medida de probabilidad condicional

Sea $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2), \mathbb{P}_{X,Y})$ el espacio de probabilidad inducido por un vector aleatorio (X, Y) definido sobre un espacio de probabilidad $(\Omega, \mathfrak{E}, \mathbb{P})$. Las medidas de probabilidad marginales de la medida de probabilidad conjunta $\mathbb{P}_{X,Y}$ son $\mathbb{P}_X(A) = \mathbb{P}_{X,Y}(A \times \Omega)$ y $\mathbb{P}_Y(B) = \mathbb{P}_{X,Y}(\Omega \times B)$, para cualesquiera A y B en $\mathfrak{B}(\mathbb{R})$, y para los casos en que $\mathbb{P}_X(A) > 0$ y $\mathbb{P}_Y(B) > 0$ las *medidas de probabilidad condicional* son:

$$\mathbb{P}_{X|Y \in B}(A) := \frac{\mathbb{P}_{X,Y}(A \times B)}{\mathbb{P}_Y(B)}, \quad \mathbb{P}_{Y|X \in A}(B) := \frac{\mathbb{P}_{X,Y}(A \times B)}{\mathbb{P}_X(A)}.$$

En otras palabras, ante la ocurrencia de un evento representado por $\{Y \in B\}$ la probabilidad $\mathbb{P}_X(A)$ debe *actualizarse* a $\mathbb{P}_{X|Y \in B}(A)$, y ante la ocurrencia del un evento representado por $\{X \in A\}$ la probabilidad $\mathbb{P}_Y(B)$ debe *actualizarse* a $\mathbb{P}_{Y|X \in A}(B)$. En el caso particular en que, por ejemplo, $\mathbb{P}_X(A) = \mathbb{P}_{X|Y \in B}(A)$ para cualesquiera A y B en $\mathfrak{B}(\mathbb{R})$, tendríamos el caso en que los valores que reporte la variable aleatoria Y no influyen en el cálculo de probabilidades de X . Esto nos lleva a definir un concepto de *independencia*, pero ahora para variables aleatorias, no solo para eventos. Aplicando la Definición 1.8 y (1.2) a eventos $X^{(-1)}(A) \in \mathfrak{E}$ e $Y^{(-1)}(B) \in \mathfrak{E}$ para cualesquiera A y B en $\mathfrak{B}(\mathbb{R})$ podemos plantear lo siguiente:

Definición 2.5: Variables aleatorias independientes

Sea $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2), \mathbb{P}_{X,Y})$ el espacio de probabilidad inducido por un vector aleatorio (X, Y) . Se dice que X e Y son *variables aleatorias independientes*, y lo denotaremos $X \perp Y$, si para cualesquiera A y B en $\mathfrak{B}(\mathbb{R})$ se cumple que:

$$\mathbb{P}_{X,Y}(A \times B) = \mathbb{P}_X(A)\mathbb{P}_Y(B).$$

Un par de observaciones sobre la definición anterior:

- De (1.16) recordemos que $\mathbb{P}_{X,Y}(A \times B) \equiv \mathbb{P}(X \in A, Y \in B)$ por lo que:

$$X \perp Y \Leftrightarrow \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B), \quad \forall A, B \in \mathfrak{B}(\mathbb{R}). \quad (1.18)$$

- $\mathbb{P}_{X,Y}(A \times B) = \mathbb{P}_X(A)\mathbb{P}_Y(B)$ es una de muchas otras medidas de probabilidad conjunta posibles para el vector aleatorio (X, Y) que tiene medidas de probabilidad marginales \mathbb{P}_X y \mathbb{P}_Y , véase (1.17), pero es la medida de probabilidad conjunta que corresponde al caso particular $X \perp Y$.

Trabajar directamente con medidas de probabilidad inducidas como $\mathbb{P}_X : \mathfrak{B}(\mathbb{R}) \rightarrow [0, 1]$ o $\mathbb{P}_{X,Y} : \mathfrak{B}(\mathbb{R}^2) \rightarrow [0, 1]$ usualmente no es práctico, comenzando por el hecho de que sus dominios son colecciones de subconjuntos de \mathbb{R} o \mathbb{R}^2 , así que como funciones no es posible analizarlas con las poderosas herramientas del cálculo diferencial e integral para funciones $\mathbb{R} \rightarrow \mathbb{R}$ o bien $\mathbb{R}^2 \rightarrow \mathbb{R}$, ni siquiera graficarlas. Gracias al siguiente teorema, es posible obtener funciones equivalentes a \mathbb{P}_X , \mathbb{P}_Y y $\mathbb{P}_{X,Y}$ que sí tengan dominios \mathbb{R} y \mathbb{R}^2 sin perder información para el cálculo de probabilidades:

Teorema 2.6: Generador de $\mathfrak{B}(\mathbb{R})$

Sea $\mathcal{J} := \{] -\infty, x] : x \in \mathbb{R}\}$ y denotemos $\sigma(\mathcal{J})$ a la colección de todos los conjuntos resultantes de realizar una cantidad numerable de operaciones de conjuntos con elementos de \mathcal{J} . Entonces:

$$\sigma(\mathcal{J}) = \mathfrak{B}(\mathbb{R}).$$

Demostración: Consúltese, por ejemplo: Clarke, *Random Variables*. □

El teorema anterior significa que casi cualquier subconjunto $B \subseteq \mathbb{R}$ se puede reexpresar en términos de una cantidad numerable de operaciones de conjunto con puros intervalos de la forma $] -\infty, x]$ con valores $x \in \mathbb{R}$. Tal es el caso de cualquier tipo de intervalo en \mathbb{R} . El cálculo de probabilidades con variables aleatorias $\mathbb{P}(X \in B) \equiv \mathbb{P}_X(B)$ más frecuente es justamente en el que B representa un intervalo en \mathbb{R} , que de forma general podríamos denotar como $I =]a, b]$ donde $-\infty < a \leq b < +\infty$, y donde el símbolo $]$ puede denotar abierto o cerrado. Por ejemplo, los intervalos acotados $[a, b]$, $]a, b[$, $]a, b]$, $[a, b[$, $[a, a] = \{a\}$, $]a, a[= \emptyset$, o bien intervalos no acotados como $] -\infty, b]$, $] -\infty, b[$, $[a, +\infty[$, $]a, +\infty[$, $] -\infty, +\infty[= \mathbb{R}$.

Teniendo resuelto cómo calcular probabilidades de que una variable aleatoria X reporte un valor que pertenezca a este tipo de intervalos, el cálculo de probabilidades para uniones disjuntas numerables de intervalos es inmediato a partir de la Definición 1.5, esto es, si I_1, I_2, \dots son intervalos en \mathbb{R} tales que $I_j \cap I_k = \emptyset$ para todo $j \neq k$ entonces:

$$\mathbb{P}\left(X \in \bigcup_{n \geq 1} I_n\right) \equiv \mathbb{P}_X\left(\bigcup_{n \geq 1} I_n\right) = \sum_{n \geq 1} \mathbb{P}_X(I_n) \equiv \sum_{n \geq 1} \mathbb{P}(X \in I_n). \quad (1.19)$$

Por ejemplo, $]a, b] =] -\infty, b] \setminus] -\infty, a]$ y por lo tanto aplicando propiedades de las medidas de probabilidad obtenemos:

$$\mathbb{P}(X \in]a, b]) \equiv \mathbb{P}_X(]a, b]) = \mathbb{P}_X(] -\infty, b] \setminus] -\infty, a]) = \mathbb{P}_X(] -\infty, b]) - \mathbb{P}_X(] -\infty, a]). \quad (1.20)$$

La demostración de los otros casos queda como ejercicio. La relevancia de lo anterior es que entonces basta saber calcular $\mathbb{P}_X(]-\infty, x])$ para cualquier $x \in \mathbb{R}$ y con ello será posible calcular la probabilidad de que la variable aleatoria X reporte un valor $X(\omega)$ que pertenezca a cualquier otro tipo de intervalo. Esto motiva la siguiente:

Definición 2.7: Función de distribución de probabilidades

La **función de distribución de probabilidades** de una variable aleatoria X es una función $F_X : \mathbb{R} \rightarrow [0, 1]$ definida como:

$$F_X(x) := \mathbb{P}_X(]-\infty, x]) \equiv \mathbb{P}(X \leq x).$$

Consecuencia de la definición anterior, y de las propiedades de la medida de probabilidad \mathbb{P}_X inducida por una variable aleatoria X , se obtiene el siguiente teorema, cuya demostración se deja como ejercicio:

Teorema 2.8: Propiedades de una función de distribución

Sea F_X la función de distribución de probabilidades de una variable aleatoria X . Entonces:

- a) F_X es una función monótona creciente, esto es que si $x_1 < x_2$ entonces $F_X(x_1) \leq F_X(x_2)$.
- b) F_X es continua por la derecha, es decir que $\lim_{\delta \rightarrow 0^+} F_X(x + \delta) = F_X(x)$.
- c) $F_X(+\infty) \equiv \lim_{x \rightarrow +\infty} F_X(x) = 1$.
- d) $F_X(-\infty) \equiv \lim_{x \rightarrow -\infty} F_X(x) = 0$.
- e) $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$.
- f) $\mathbb{P}(X < a) = F_X(a-) \equiv \lim_{x \rightarrow a^-} F_X(x)$.
- g) $\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a-)$.
- h) $\mathbb{P}(X > b) = 1 - F_X(b)$.
- i) $\mathbb{P}(X \geq b) = 1 - F_X(b-)$.
- j) $\mathbb{P}(a < X < b) = F_X(b-) - F_X(a)$.
- k) $\mathbb{P}(a \leq X < b) = F_X(b-) - F_X(a-)$.
- l) $\mathbb{P}(X = x) = F_X(x) - F_X(x-)$.

De hecho, si $F : \mathbb{R} \rightarrow [0, 1]$ es cualquier función que satisface las propiedades a), b), c) y d) del Teorema 2.8, se puede demostrar que siempre existe una variable aleatoria que tiene

a F como función de distribución de probabilidades, y los incisos e) a l) son consecuencia de los cuatro primeros, demostración que puede consultarse en Clarke, *Random Variables* o Domínguez-Martínez, *Diseño y análisis de modelos de probabilidad*.

A las variables aleatorias se les puede clasificar en tres categorías, dependiendo de algunas propiedades particulares adicionales de sus funciones de distribución de probabilidades. La categoría que en particular se ocupará en el resto de este texto corresponde al caso en que $\mathbb{P}(X = x) = 0$ para todo valor $x \in \mathbb{R}$, lo cual implicaría que F_X es continua también por la izquierda, y por tanto continua en todo \mathbb{R} , veáse la Figura 1.1a.

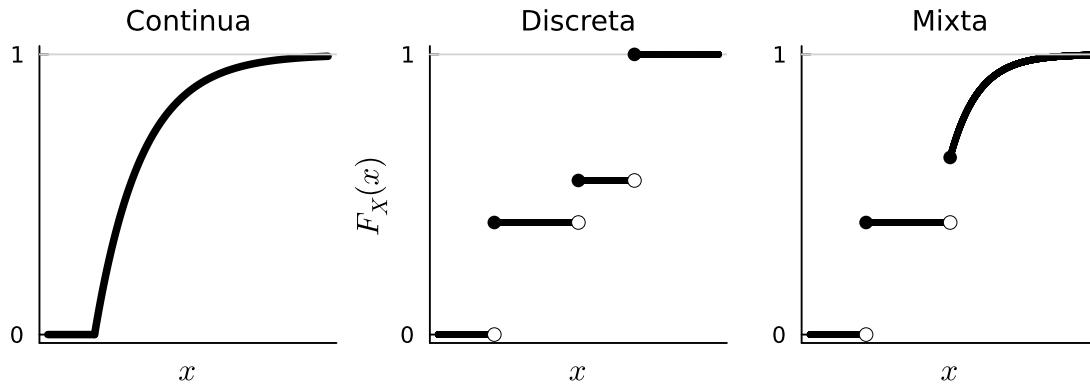


Figura 1.1: Tipos de funciones de distribución de probabilidades: a) continua, b) discreta, c) mixta.

Si F_X resulta ser una función continua se dice que X es una **variable aleatoria continua**, y en tal caso, por consecuencia del Teorema 2.8:

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b), \quad (1.21)$$

$$\mathbb{P}(X > b) = \mathbb{P}(X \geq b), \quad \mathbb{P}(X < a) = \mathbb{P}(X \leq a). \quad (1.22)$$

Definición 2.9: Variable aleatoria absolutamente continua

Sea $|a, b|$ un intervalo cualquiera, abierto, cerrado o semi abierto, acotado o no, es decir $-\infty \leq a \leq b \leq +\infty$. Se dice que X es una **variable aleatoria absolutamente continua** si existe una función integrable $f_X : \mathbb{R} \rightarrow [0, +\infty[$ tal que:

$$\mathbb{P}(X \in |a, b|) = \int_a^b f_X(x) dx,$$

y a la función f_X se le denomina **función de densidad de probabilidades**. \square

Consecuencia inmediata de las Definiciones 2.7 y 2.9, en el caso de que X sea una variable aleatoria absolutamente continua se deduce la siguiente relación entre su función de

distribución de probabilidades F_X y su función de densidad de probabilidades f_X :

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in]-\infty, x]) = \int_{-\infty}^x f_X(t)dt, \quad \text{para todo valor } x \in \mathbb{R}, \quad (1.23)$$

y como $F_X(+\infty) = 1$ esto implica que necesariamente $\int_{-\infty}^{+\infty} f_X(x)dx = 1$. Más aún, en este caso F_X no solo es continua sino que además, como consecuencia del *Teorema Fundamental del Cálculo*, F_X es derivable en cada punto $x \in \mathbb{R}$ donde f_X sea continua, y en tales casos:

$$\frac{d}{dx} F_X(x) = f_X(x). \quad (1.24)$$

Así como cualquier subconjunto $B \in \mathfrak{B}(\mathbb{R})$ puede obtenerse en términos de una cantidad numerable de operaciones de conjuntos con intervalos de la forma $] -\infty, x]$, de forma análoga cualquier subconjunto $B \in \mathfrak{B}(\mathbb{R}^n)$ con $n \geq 2$ puede obtenerse en términos de una cantidad numerable de operaciones de conjuntos con productos cartesianos de intervalos de la forma $] -\infty, x]$ también, es decir con subconjuntos de la clase:

$$\mathcal{J}^n := \{] -\infty, x_1] \times \cdots \times] -\infty, x_n] : (x_1, \dots, x_n) \in \mathbb{R}^n\}, \quad (1.25)$$

lo que implica que para definir la medida de probabilidad $\mathbb{P}_{\mathbf{X}}$ inducida por un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ basta que dicha medida esté definida sobre la clase \mathcal{J}^n , es decir poder calcular para cualquier $(x_1, \dots, x_n) \in \mathbb{R}^n$:

$$\begin{aligned} \mathbb{P}_{\mathbf{X}}(] -\infty, x_1] \times \cdots \times] -\infty, x_n]) &= \mathbb{P}(\mathbf{X}^{(-1)}(] -\infty, x_1] \times \cdots \times] -\infty, x_n])) \\ &= \mathbb{P}\left(\{\omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in] -\infty, x_1] \times \cdots \times] -\infty, x_n]\}\right) \\ &= \mathbb{P}\left(\{\omega \in \Omega : X_1(\omega) \leq x_1 \wedge \cdots \wedge X_n(\omega) \leq x_n\}\right) \\ &= \mathbb{P}\left(\{\omega \in \Omega : X_1(\omega) \leq x_1\} \cap \cdots \cap \{\omega \in \Omega : X_n(\omega) \leq x_n\}\right) \end{aligned} \quad (1.26)$$

en donde es costumbre escribir (1.26) de forma simplificada como $\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$, lo cual motiva la siguiente definición, que resulta análoga a la Definición 2.7:

Definición 2.10: Función de distribución conjunta

La **función de distribución conjunta de probabilidades** de un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ es una función $F_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, 1]$ definida como:

$$F_{\mathbf{X}}(x_1, \dots, x_n) := \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n).$$

A partir de la función de distribución conjunta de probabilidades $F_{\mathbf{X}}$ de un vector aleatorio de probabilidades $\mathbf{X} = (X_1, \dots, X_n)$ es posible deducir las funciones de distribución de probabilidades individuales F_{X_j} de cada una de las variables aleatorias componentes $j \in \{1, \dots, n\}$.

A cada F_{X_j} se le denomina *función de distribución marginal* y al procedimiento para deducirla a partir de la función de distribución conjunta $F_{\mathbf{X}}$ se le denomina *marginalización*:

$$\begin{aligned} F_{X_j}(x_j) &= \mathbb{P}(X_j \leq x_j) = \mathbb{P}(\{\omega \in \Omega : X_j(\omega) \leq x_j\}) \\ &= \mathbb{P}(D_1 \cap \dots \cap D_{j-1} \cap \{\omega \in \Omega : X_j(\omega) \leq x_j\} \cap D_{j+1} \cap \dots \cap D_n) \end{aligned}$$

donde $D_k = \{\omega \in \Omega : X_k(\omega) \leq +\infty\} = \Omega$ para $k \neq j$ y por lo tanto:

$$F_{X_j}(x_j) = F_{\mathbf{X}}(+\infty, \dots, +\infty, x_j, +\infty, \dots, +\infty). \quad (1.27)$$

Sin embargo, a partir de las funciones de distribución marginales F_{X_1}, \dots, F_{X_n} no es posible deducir la función de distribución conjunta del vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ ya que pueden existir distintas funciones de distribución conjunta que tengan esas mismas funciones de distribución marginales, como se ilustra en el siguiente:

Ejemplo 4. Distribuciones conjuntas con mismas marginales

Considérese el caso $n = 2$, primero un vector aleatorio $\mathbf{X} = (X_1, X_2)$ donde X_1 es una variable aleatoria continua uniforme sobre el intervalo abierto $]0, 1[$, y sea la variable aleatoria $X_2 = X_1$, lo que implica que las funciones de distribución marginales son iguales, y dadas por $F_{X_1}(x) = F_{X_2}(x) = F(x) = x\mathbf{1}_{\{0 < x < 1\}} + \mathbf{1}_{\{x \geq 1\}}$. En cuanto a su función de distribución conjunta:

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2) &= \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = \mathbb{P}(X_1 \leq x_1, X_1 \leq x_2) = \mathbb{P}(X_1 \leq \min\{x_1, x_2\}), \\ &= F(\min\{x_1, x_2\}) = F(x_1)\mathbf{1}_{\{x_1 \leq x_2\}} + F(x_2)\mathbf{1}_{\{x_1 > x_2\}} = \min\{F(x_1), F(x_2)\}. \end{aligned}$$

Por otro lado, consideremos ahora el vector aleatorio $\mathbf{Y} = (Y_1, Y_2)$ donde Y_1 es una variable aleatoria continua uniforme sobre el intervalo abierto $]0, 1[$, y donde $Y_2 = 1 - Y_1$. Es inmediato verificar que Y_2 resulta ser también una variable aleatoria continua uniforme sobre el intervalo abierto $]0, 1[$, lo que implica que las funciones de distribución marginales son iguales, y dadas por $F_{Y_1}(y) = F_{Y_2}(y) = F(y) = y\mathbf{1}_{\{0 < y < 1\}} + \mathbf{1}_{\{y \geq 1\}}$. En cuanto a su función de distribución conjunta:

$$\begin{aligned} F_{\mathbf{Y}}(y_1, y_2) &= \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2) = \mathbb{P}(Y_1 \leq y_1, 1 - Y_1 \leq y_2) = \mathbb{P}(1 - y_2 \leq Y_1 \leq y_1), \\ &= [F(y_1) - F(1 - y_2)]\mathbf{1}_{\{1-y_2 < y_1\}} = \max\{F(y_1) + F(y_2) - 1, 0\}. \end{aligned}$$

Tenemos entonces dos vectores aleatorios $\mathbf{X} = (X_1, X_2)$ e $\mathbf{Y} = (Y_1, Y_2)$ cuyas funciones de distribución marginales son iguales, pero con funciones de distribución conjunta diferentes ya que $F_{\mathbf{X}} \neq F_{\mathbf{Y}}$.

En el caso de una variable aleatoria X , por medio de su función de distribución de probabilidades $F_X(x) = \mathbb{P}(X \leq x)$ es relativamente accesible hacer cálculo de probabilidades sobre cualquier tipo de intervalo en \mathbb{R} , como consecuencia del Teorema 2.8. Pero en el caso de un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ con $n \geq 2$, reexpresar un subconjunto de interés $B \in \mathfrak{B}(\mathbb{R}^n)$ por medio de operaciones de conjuntos de la clase \mathcal{J}^n es, por lo general, demasiado complejo y hasta inaccesible en términos prácticos, lo que dificulta el cálculo de

probabilidades por medio de su función de distribución conjunta, salvo en el caso muy particular de que B sea un producto cartesiano de intervalos semiabiertos de la forma $]a, b]$. Por ejemplo, en el caso $n = 2$ se tendría:

$$\mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = F_{X_1, X_2}(b_1, b_2) - F_{X_1, X_2}(b_1, a_2) - F_{X_1, X_2}(a_1, b_2) + F_{X_1, X_2}(a_1, a_2). \quad (1.28)$$

En el resto del texto nos restringiremos al caso de vectores aleatorios bivariados de variables aleatorias continuas, utilizaremos la notación (X, Y) para el vector aleatorio, $F_{X,Y}$ para su función de distribución conjunta (bivariada), y se resume en el siguiente teorema las propiedades que se deducen con base en todo lo anterior:

Teorema 2.11: Propiedades de una distribución conjunta

Sea (X, Y) un vector aleatorio de variables aleatorias continuas, con función de distribución conjunta $F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$. Entonces:

- a) $F_{X,Y}$ es monótona creciente en cada argumento: Si $x_1 < x_2$ entonces $F_{X,Y}(x_1, y) \leq F_{X,Y}(x_2, y)$. Si $y_1 < y_2$ entonces $F_{X,Y}(x, y_1) \leq F_{X,Y}(x, y_2)$.

- b) Las funciones de distribución marginales resultan ser:^a

$$F_X(x) = F_{X,Y}(x, +\infty), \quad F_Y(y) = F_{X,Y}(+\infty, y).$$

- c) $F_{X,Y}$ es continua en todo su dominio \mathbb{R}^2 .

- d) Límites extremos:

$$F_{X,Y}(-\infty, y) = 0 = F_{X,Y}(x, -\infty) = F_{X,Y}(-\infty, -\infty), \quad F_{X,Y}(+\infty, +\infty) = 1.$$

- e) Propiedad 2–creciente: Si $x_1 < x_2$, $y_1 < y_2$, entonces:

$$F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1) \geq 0.$$

- f) Cotas de Fréchet-Hoeffding:^b

$$\max\{F_X(x) + F_Y(y) - 1, 0\} \leq F_{X,Y}(x, y) \leq \min\{F_X(x), F_Y(y)\}.^c$$

- g) $X \perp Y$ si y solo si $F_{X,Y}(x, y) = F_X(x)F_Y(y)$.

^aLa notación $F(+\infty)$ equivale a $\lim_{x \rightarrow +\infty} F(x)$.

^bComo consecuencia de los trabajos por separado de Maurice Fréchet (1878–1973) y Wassily Hoeffding (1914–1991).

^cConsecuencia inmediata de aplicar (1.1) con $E_1 = \{X \leq x\}$ y $E_2 = \{Y \leq y\}$.

Las propiedades f) y g) del teorema anterior ameritan una observación particular que será particularmente relevante en el capítulo siguiente, y por tanto se presenta como el siguiente:

Corolario 2.12

Las funciones $H_j : \mathbb{R}^2 \rightarrow [0, 1]$ definidas como:

$$H_1(x, y) := \max\{F_X(x) + F_Y(y) - 1, 0\}, \quad (1.29)$$

$$H_2(x, y) := F_X(x)F_Y(y), \quad (1.30)$$

$$H_3(x, y) := \min\{F_X(x), F_Y(y)\}, \quad (1.31)$$

cumplen con:

- a) las propiedades del Teorema 2.11,
- b) ser tres funciones de distribución conjunta bivariadas distintas que tienen las mismas funciones de distribución marginales, esto es, para $j \in \{1, 2, 3\}$

$$H_j(x, +\infty) = F_X(x), \quad H_j(+\infty, y) = F_Y(y).$$

Este corolario nos garantiza que existen al menos tres funciones de distribución conjunta bivariadas con exactamente las mismas funciones de distribución marginales, y por tanto el solo conocimiento de las marginales no nos permite, en general, identificar la distribución conjunta que les corresponde. Esquemáticamente:

$$F_{X,Y} \Rightarrow F_X, F_Y$$

$$F_{X,Y} \not\Rightarrow F_X, F_Y$$

Nótese que en el Ejemplo 4 justamente aparecen como funciones de distribución conjunta H_1 y H_3 que son, de hecho, las cotas de Fréchet-Hoeffding del Teorema 2.11.

En el Teorema 2.11 solo se requirió que X e Y fuesen variables aleatorias continuas, no necesariamente absolutamente continuas. Supongamos ahora que lo son y que, por tanto, existen sus correspondientes *funciones de densidad marginales* f_X y f_Y como en la Definición 2.9. En tal caso ¿existe un concepto análogo de función de densidad conjunta del vector aleatorio (X, Y) ? La respuesta es: en ocasiones sí, en otras no.

Definición 2.13: Función de densidad conjunta

Se dice que un vector aleatorio (X, Y) de variables aleatorias continuas es un *vector aleatorio absolutamente continuo* si existe una función integrable $f_{X,Y} : \mathbb{R}^2 \rightarrow [0, +\infty[$ tal que para cualquier $D \in \mathfrak{B}(\mathbb{R}^2)$:

$$\mathbb{P}[(X, Y) \in D] = \iint_D f_{X,Y}(x, y) dx dy,$$

y a la función $f_{X,Y}$ se le denomina *función de densidad conjunta de probabilidades*.

Como necesariamente $\mathbb{P}((X, Y) \in \mathbb{R}^2) = 1$ entonces la definición anterior implica en particular que:

$$\iint_{\mathbb{R}^2} f_{X,Y}(x, y) dx dy = 1. \quad (1.32)$$

A partir de $f_{X,Y}$ es posible deducir las funciones de densidad marginales de cada una de las variables aleatorias:

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\} \cap \Omega) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\} \cap \{\omega \in \Omega : -\infty < Y(\omega) < +\infty\}) \\ &= \mathbb{P}((X, Y) \in]-\infty, x] \times]-\infty, +\infty[) \\ &= \int_{-\infty}^x \int_{-\infty}^{+\infty} f_{X,Y}(s, y) dy ds = \int_{-\infty}^x \psi(s) ds, \end{aligned}$$

donde $\psi(s) := \int_{-\infty}^{+\infty} f_{X,Y}(s, y) dy$, y luego aplicando el *Teorema Fundamental del Cálculo* obtenemos:

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) = \frac{d}{dx} \int_{-\infty}^x \psi(s) ds = \psi(x) \\ &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy, \end{aligned} \quad (1.33)$$

y de manera análoga se obtiene $f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx$. También, a partir de $f_{X,Y}$ podemos obtener la función de distribución conjunta mediante:

$$F_{X,Y}(x, y) = \mathbb{P}((X, Y) \in]-\infty, x] \times]-\infty, y]) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s, t) dt ds, \quad (1.34)$$

y por tanto, como consecuencia del Teorema Fundamental del Cálculo, tenemos que para todo $(x, y) \in \mathbb{R}^2$ donde $f_{X,Y}$ sea continua entonces $F_{X,Y}$ es derivable:

$$\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = f_{X,Y}(x, y). \quad (1.35)$$

Pero el que (X, Y) sea un vector de variables aleatorias absolutamente continuas no implica necesariamente que el vector sea absolutamente continuo, es decir, la existencia de una función de densidad conjunta no está garantizada. Por ejemplo, sea X una variable aleatoria absolutamente continua, $g : \mathbb{R} \rightarrow \mathbb{R}$ una función continua y estrictamente creciente, y defínase una variable aleatoria $Y := g(X)$. Es inmediato verificar que $F_{X,Y}(x, y) = \min\{F_X(x), F_Y(y)\}$ y por tanto $\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = 0$, donde claramente la constante cero no es una función de densidad conjunta de probabilidades.

Ejemplo 5. Vector aleatorio absolutamente continuo.

Consideremos el vector aleatorio (X, Y) con función de densidad conjunta:

$$f_{X,Y}(x, y) = e^{-y} \mathbf{1}_{\{0 < x < y\}}.$$

Entonces por (1.34) su función de distribución conjunta debe ser:

$$\begin{aligned} F_{X,Y}(x,y) &= \int_{-\infty}^x \int_{-\infty}^y e^{-t} \mathbf{1}_{\{0 < s < t\}} dt ds = \dots \\ &= \begin{cases} 1 - e^{-x} - xe^{-y}, & \text{si } 0 < x \leq y, \\ 1 - e^{-y} - ye^{-x}, & \text{si } 0 < y < x, \\ 0, & \text{en cualquier otro caso.} \end{cases} \end{aligned}$$

Claramente $F_{X,Y}(+\infty, +\infty) = 1$, lo que de paso comprueba que en efecto $f_{X,Y}$ es una densidad conjunta ya que es una función no negativa y su integral sobre todo \mathbb{R}^2 es igual a 1, como se mencionó en (1.32). Aplicando el Teorema 2.11, inciso b), deducimos las funciones de distribución marginales:

$$F_X(x) = F_{X,Y}(x, +\infty) = (1 - e^{-x}) \mathbf{1}_{\{x > 0\}}, \quad F_Y(y) = F_{X,Y}(+\infty, y) = (1 - e^{-y} - ye^{-y}) \mathbf{1}_{\{y > 0\}},$$

que al ser diferenciables en todo \mathbb{R} excepto en una cantidad numerable de puntos aislados (en este caso, excepto en $x = 0, y = 0$, respectivamente) corresponden entonces a variables aleatorias absolutamente continuas, con funciones de densidad marginales:

$$f_X(x) = \frac{d}{dx} F_X(x) = e^{-x} \mathbf{1}_{\{x > 0\}}, \quad f_Y(y) = \frac{d}{dy} F_Y(y) = ye^{-y} \mathbf{1}_{\{y > 0\}},$$

que de hecho corresponden a modelos de probabilidad conocidos: marginalmente X resulta ser una variable aleatoria *Exponencial* con parámetro 1, y por otro lado Y resulta ser una variable aleatoria *Gamma* con parámetros (2, 1). También pudimos haber obtenido las densidades marginales directamente a partir de la función de densidad conjunta $f_{X,Y}$, aplicando (1.33), se recomienda hacerlo como ejercicio. Como en este caso claramente:

$$F_X(x)F_Y(y) \neq F_{X,Y}(x, y)$$

entonces por el Teorema 2.11, inciso g), X e Y no son variables aleatorias independientes.

A diferencia de la sección anterior, donde al final se abordó un poco sobre el problema de medir dependencia entre eventos, la discusión sobre cómo medir dependencia entre variables aleatorias la pospondremos hasta el capítulo 3, una vez que se hayan abordado las herramientas adecuadas para ello en el capítulo 2.

3. Simulación

Abordaremos un procedimiento general para simular observaciones de un vector aleatorio (X, Y) de variables aleatorias continuas. Para ello, primero recordaremos un procedimiento general para simular observaciones de una variable aleatoria continua.

Como caso particular del Teorema 2.8, si $F : \mathbb{R} \rightarrow [0, 1]$ es una función de distribución

continua, además de ser monótona creciente, tenemos que:

$$]0, 1[\subseteq \text{Ran } F = \{F(x) : x \in \mathbb{R}\} \subseteq [0, 1]. \quad (1.36)$$

Si en particular F es estrictamente creciente, entonces existe su función inversa $F^{-1} : \text{Ran } F \rightarrow \mathbb{R}$, pero si no es estrictamente creciente, entonces quiere decir que existe al menos un intervalo dentro de su dominio donde la función es constante, y en tal caso estrictamente no existe la función inversa de F . Por ello requerimos un concepto más general:³

Definición 3.1: Función cuasi-inversa

Sea $F : \mathbb{R} \rightarrow [0, 1]$ una función de distribución de probabilidades continua. Una *cuasi-inversa* de F es cualquier función $F^{(-1)} : \text{Ran } F \rightarrow \mathbb{R}$ donde $F^{(-1)}(t)$ es cualquier $x \in \mathbb{R}$ tal que $F(x) = t$, esto es, tal que

$$F(F^{(-1)}(t)) = t.$$

Observación:

- Si la función de distribución F es estrictamente creciente entonces la definición de cuasi-inversa coincide con la de función inversa usual, esto es $F^{(-1)} = F^{-1}$, y por tanto es única.
- Si F no es estrictamente creciente, entonces existe al menos un intervalo $I \subset \mathbb{R}$ tal que para cualesquiera $x_1, x_2 \in I$ se tiene que $F(x_1) = F(x_2) = t_*$ para algún $t_* \in \text{Ran } F$, y en tal caso $F^{(-1)}(t_*) = x_*$ donde x_* es cualquier elemento del intervalo I . De ahí que en la definición se habla de *una cuasi-inversa* y no de *la cuasi-inversa*, es decir que en este caso la cuasi-inversa $F^{(-1)}$ no es única, pues para cada intervalo donde F permanezca constante existe una infinidad de valores para escoger en dicho intervalo y asignarlo como el valor de una cuasi-inversa.
- Si X es una variable aleatoria continua con función de distribución F , entonces para todo intervalo $I =]a, b[$ donde F sea constante se tiene que:

$$\mathbb{P}(X \in I) = \mathbb{P}(X \in]a, b[) = F(b) - F(a) = 0. \quad (1.37)$$

- La cuasi-inversa de una función de distribución continua siempre resulta ser estrictamente creciente, posiblemente con algunas discontinuidades si F no es estrictamente creciente.

Ejemplo 6. Función cuasi-inversa.

Sea X una variable aleatoria absolutamente continua uniforme sobre el conjunto $]0, 1[\cup]2, 3[$, esto es con función de densidad:

$$f_X(x) = \frac{\mathbf{1}_{]0,1[\cup]2,3[}(x)}{2}.$$

³Adaptado de Nelsen, *An Introduction to Copulas*

Entonces su función de distribución es:

$$F_X(x) = \int_{-\infty}^x f_X(s) ds = \begin{cases} 0, & x \leq 0, \\ \frac{x}{2}, & 0 < x < 1, \\ \frac{1}{2}, & 1 \leq x \leq 2, \\ \frac{x-1}{2}, & 2 < x < 3, \\ 1, & x \geq 3, \end{cases}$$

de donde $\text{Ran } F_X = [0, 1]$. Notemos que F_X es constante en tres intervalos:

$$\begin{aligned} \text{Si } x \in I_1 =]-\infty, 0] &\Rightarrow F_X(x) = 0, \\ \text{Si } x \in I_2 = [1, 2] &\Rightarrow F_X(x) = \frac{1}{2}, \\ \text{Si } x \in I_3 = [3, +\infty[&\Rightarrow F_X(x) = 1, \end{aligned}$$

por lo que en este ejemplo la distintas funciones quasi-inversas $F_X^{(-1)} : [0, 1] \rightarrow \mathbb{R}$ que pueden obtenerse quedan parametrizadas mediante la elección de valores $x_j^* \in I_j$, $j \in \{1, 2, 3\}$, de la siguiente forma:

$$F_X^{(-1)}(t) = \begin{cases} x_1^*, & t = 0, \\ 2t, & 0 < t < \frac{1}{2}, \\ x_2^*, & t = \frac{1}{2}, \\ 2t + 1, & \frac{1}{2} < t < 1, \\ x_3^*, & t = 1. \end{cases}$$

En la Figura 1.2 pueden apreciarse las gráficas de F_X y sus quasi-inversas.

Teorema 3.2: Transformada inversa

Sea F una función de distribución de probabilidades continua cualquiera, y sea U una variable aleatoria continua uniforme sobre el intervalo abierto $]0, 1[$. Si se define la variable aleatoria $X := F^{(-1)}(U)$ entonces su función de distribución resulta ser $F_X = F$.

Demostración:

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(F^{(-1)}(U) \leq x) = \mathbb{P}(F(F^{(-1)}(U)) \leq F(x)) \\ &= \mathbb{P}(U \leq F(x)) = F_U(F(x)) = F(x), \end{aligned}$$

donde la última igualdad se justifica por el hecho de que $F_U(u) = u\mathbf{1}_{\{0 < u < 1\}} + \mathbf{1}_{\{u \geq 1\}}$. \square

El Teorema 3.2 anterior, conocido también como el *Teorema de la Transformada Inversa*, es especialmente útil para simular observaciones de una variable aleatoria continua con una

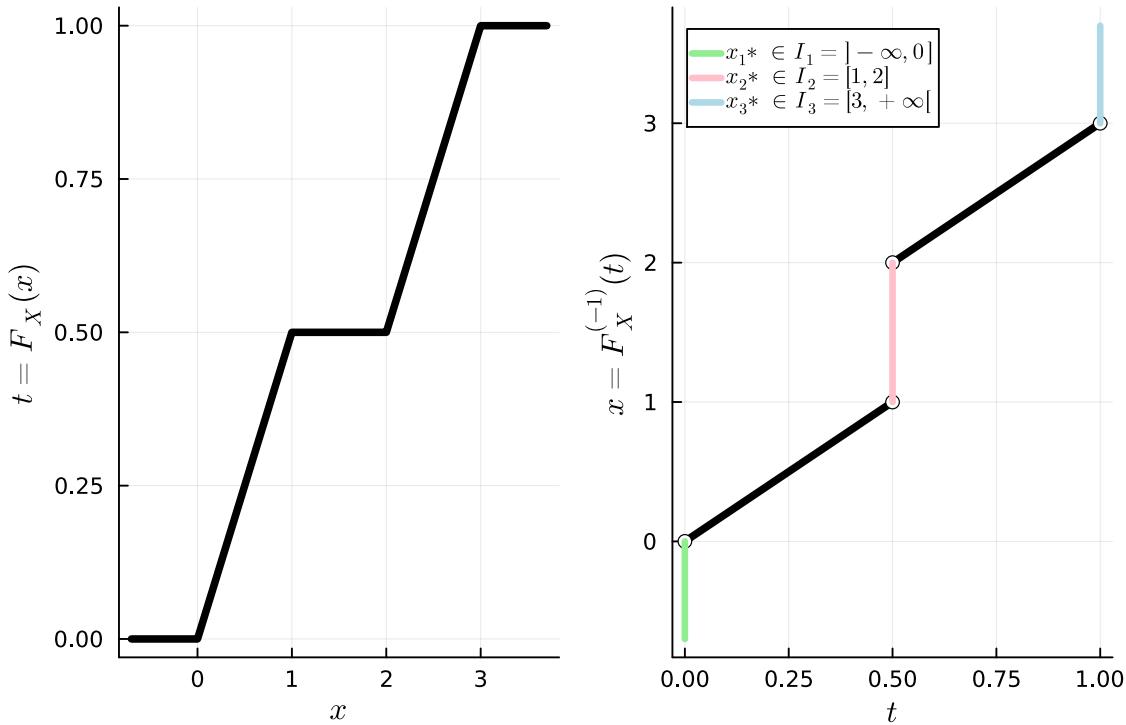


Figura 1.2: Ejemplo 6. *Izquierda:* Función de distribución F_X monótona creciente. *Derecha:* Funciones cuasi-inversas de F_X .

función de distribución dada F , a partir de un simulador de observaciones de una variable aleatoria continua uniforme sobre el intervalo abierto $]0, 1[$. Para simular n observaciones independientes e idénticamente distribuidas (iid) de dicha variable, procedemos como sigue:

Algoritmo 3.3

Para simular n observaciones independientes e idénticamente distribuidas (iid) de una función de distribución continua F_X a partir de un simulador de una variable aleatoria continua $U \sim \text{Uniforme}(0, 1)$:

1. Simule u_1, \dots, u_n iid $\sim \text{Uniforme}(0, 1)$.
2. Para cada $j \in \{1, \dots, n\}$ calcule $x_j = F_X^{(-1)}(u_j)$.
3. x_1, \dots, x_n son iid $\sim F_X$.

Para aquellos valores t en donde el valor de $F_X^{(-1)}(t)$ no sea único, es decir para aquellos casos en los que existe un intervalo $I \subset \mathbb{R}$ tal que $F_X(x) = t$ para todo $x \in I$, se define $F_X^{(-1)}(t) = x_*$ eligiendo un $x_* \in I \cap \overline{\text{Ran } X}$ para garantizar que se simulen valores admisibles para la variable aleatoria X , donde $\overline{\text{Ran } X}$ significa la cerradura del conjunto $\text{Ran } X$.

Ejemplo 7. Simular una variable aleatoria continua.

Continuando con el Ejemplo 6, tenemos que $\text{Ran } X =]0, 1[\cup]2, 3[$ por lo que $\overline{\text{Ran } X} = [0, 1] \cup [2, 3]$. Para definir los valores de la cuasi-inversa en $t \in \{0, \frac{1}{2}, 1\}$, para cada $j \in \{1, 2, 3\}$ elegimos un $x_j^* \in I_j \cap \overline{\text{Ran } X}$, es decir:

$$\begin{aligned} x_1^* &\in]-\infty, 0] \cap ([0, 1] \cup [2, 3]) = \{0\}, \\ x_2^* &\in [1, 2] \cap ([0, 1] \cup [2, 3]) = \{1, 2\}, \\ x_3^* &\in [3, +\infty[\cap ([0, 1] \cup [2, 3]) = \{3\}. \end{aligned}$$

Para $x_1^* = 0$ y $x_3^* = 0$ las opciones resultaron únicas, no así para $x_2^* \in \{1, 2\}$ por lo que arbitrariamente escogemos uno de esos valores, digamos $x_2^* = 1$. Entonces la cuasi-inversa a utilizar para simular observaciones de X por medio del Algoritmo 3.3 se simplifica a:

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

Nótese que la elección de x_2^* cuando $t = \frac{1}{2}$, ya sea como $x_2^* = 1$ o como $x_2^* = 2$, no es relevante desde un punto de vista de la simulación, ya que la probabilidad de tener que evaluar $F_X^{(-1)}(t)$ en $t = \frac{1}{2}$ es $\mathbb{P}(U = \frac{1}{2}) = 0$, por ser U variable aleatoria continua. ■

Ahora consideremos un vector aleatorio (X, Y) de variables aleatorias continuas con función de distribución conjunta $F_{X,Y}$. ¿Cómo simular observaciones iid de (X, Y) ? Solo en el caso de independencia $X \perp Y$ sería válido simplemente deducir las funciones de distribución marginal F_X y F_Y , aplicar el Algoritmo 3.3 para simular por separado $\{x_1, \dots, x_n\} \sim F_X$ y $\{y_1, \dots, y_n\} \sim F_Y$, y simplemente crear los pares $\{(x_1, y_1), \dots, (x_n, y_n)\}$ como observaciones de la función de distribución conjunta $F_{X,Y} = F_X F_Y$, ya que por el supuesto de independencia, los valores de cada variable aleatoria no influyen en los de la otra.

En caso de que las variables aleatorias de un vector aleatorio (X, Y) no sean independientes, no sería válido simular por separado observaciones a partir de F_X y F_Y ya que, recordemos, pueden existir diversas formas de dependencia entre ellas que tengan las mismas marginales. La información de la dependencia se pierde en el proceso de marginalización de $F_{X,Y}$, así que es necesario proceder de forma que se conserve la relación de dependencia existente entre las variables.

A continuación se presenta un método general para simular observaciones de un vector aleatorio de dos variables aleatorias continuas, pero es importante aclarar que para algunos modelos específicos de vectores aleatorios existen también algoritmos específicos que, si bien son equivalentes, pueden resultar más sencillos de implementar y/o más eficientes en algún sentido (velocidad, precisión, etc.).

Algoritmo 3.4

Para simular n observaciones iid de un vector aleatorio (X, Y) de variables aleatorias continuas con función de distribución conjunta $F_{X,Y}$:

1. Deduce una de las funciones de distribución marginales, por ejemplo

$$F_X(x) = F_{X,Y}(x, +\infty).$$

2. Simula x_1, \dots, x_n iid $\sim F_X$ aplicando el Algoritmo 3.3.
3. Deduce la función de distribución condicional para Y dado $X = x$, esto es

$$F_{Y|X}(y | x) = \mathbb{P}(Y \leq y | X = x).$$

4. Para cada x_j simula $y_j \sim F_{Y|X}(y | x_j)$, aplicando el Algoritmo 3.3.
5. $(x_1, y_1), \dots, (x_n, y_n)$ son iid $\sim F_{X,Y}$.

En el algoritmo anterior $F_{Y|X}(y | x) = \mathbb{P}(Y \leq y | X = x)$ es la actualización de la función de distribución $F_Y(y) = \mathbb{P}(Y \leq y)$ ante el hecho de que $\{X = x\}$. Como nos hemos restringido a variables aleatorias continuas, al ser F_X continua entonces por el Teorema 2.8, inciso 1), tenemos que:

$$\mathbb{P}(X = x) = F_X(x) - F_X(x-) = F_X(x) - F_X(x) = 0,$$

y por tanto no podemos aplicar directamente la Definición 2.4 ya que para calcular algo de la forma:

$$\mathbb{P}(Y \in B | X \in A) = \frac{\mathbb{P}(X \in A, Y \in B)}{\mathbb{P}(X \in A)}$$

se requiere que $\mathbb{P}(X \in A) > 0$, y en este caso necesitamos condicionar en $A = \{X = x\}$, donde $\mathbb{P}(X = x) = 0$ por ser X variable aleatoria continua. Pero bastaría con que X sea absolutamente continua, con función de densidad f_X , para poder obtener $F_{Y|X}(y | x)$ por medio de un límite:

$$\begin{aligned} F_{Y|X}(y | x) &= \mathbb{P}(Y \leq y | X = x) = \lim_{\delta \rightarrow 0+} \mathbb{P}(Y \leq y | x \leq X \leq x + \delta) \\ &= \lim_{\delta \rightarrow 0+} \frac{\mathbb{P}(x \leq X \leq x + \delta, Y \leq y)}{\mathbb{P}(x \leq X \leq x + \delta)} = \lim_{\delta \rightarrow 0+} \frac{F_{X,Y}(x + \delta, y) - F_{X,Y}(x, y)}{F_X(x + \delta) - F_X(x)} \\ &= \lim_{\delta \rightarrow 0+} \frac{\frac{F_{X,Y}(x + \delta, y) - F_{X,Y}(x, y)}{\delta}}{\frac{F_X(x + \delta) - F_X(x)}{\delta}} = \frac{\frac{\partial}{\partial x} F_{X,Y}(x, y)}{\frac{d}{dx} F_X(x)} \\ &= \frac{1}{f_X(x)} \frac{\partial}{\partial x} F_{X,Y}(x, y), \quad \forall x : f_X(x) > 0. \end{aligned} \tag{1.38}$$

Si además el vector aleatorio (X, Y) es absolutamente continuo, con función de densidad conjunta $f_{X,Y}$, entonces consecuencia de (1.35) podemos calcular la derivada $\frac{d}{dy}$ en ambos

miembros de (1.38) y obtener así:

$$\begin{aligned}\frac{d}{dy}F_{Y|X}(y|x) &= \frac{d}{dy}\frac{1}{f_X(x)}\frac{\partial}{\partial x}F_{X,Y}(x,y) = \frac{1}{f_X(x)}\frac{\partial^2}{\partial y \partial x}F_{X,Y}(x,y) \\ f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad \forall x : f_X(x) > 0,\end{aligned}\tag{1.39}$$

donde $f_{Y|X}(y|x)$ es la actualización de la función de densidad $f_Y(y)$ condicional en que $X = x$.

Ejemplo 8. Simular el Ejemplo 5.

Desarrollaremos los detalles específicos para aplicar el Algoritmo 3.4 en la simulación de observaciones del vector aleatorio (X, Y) del Ejemplo 5. En dicho ejemplo se dedujo que la distribución de probabilidad marginal de X corresponde a una variable aleatoria *Exponencial* con parámetro 1, es decir con función de distribución de probabilidades:

$$F_X(x) = (1 - e^{-x})\mathbf{1}_{\{x>0\}} = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-x}, & x > 0. \end{cases}$$

Si $x > 0$ entonces $F_X(x) > 0$ es estrictamente creciente, y por tanto su cuasi-inversa es la inversa usual, es decir $F_X^{(-1)}(t) = F_X^{-1}(t) = -\log(1-t)$, con $t \in]0, 1[$. Pero para todo valor $x \in I =]-\infty, 0]$ tenemos que $F_X(x) = 0$ (constante) y por tanto para $t = 0$ una cuasi-inversa es $F_X^{(-1)}(0) = x_*$ para algún $x_* \in I$ que se escoga. Como en este caso $\text{Ran } X =]0, +\infty[$ entonces $\overline{\text{Ran } X} = [0, +\infty[$ y por tanto escogemos $x_* \in I \cap \overline{\text{Ran } X} = \{0\}$, que en este ejemplo solo nos brinda una opción: $x_* = 0$. Así, la cuasi-inversa a utilizar en el paso 2 del Algoritmo 3.4 se simplifica a:

$$F_X^{(-1)}(t) = -\log(1-t), \quad t \in [0, 1].$$

Para obtener la función de distribución de Y condicional en $X = x$ aplicaremos (1.38). A partir de la fórmula explícita de $F_{X,Y}$ obtenida en el Ejemplo 5 se calcula:

$$\frac{\partial}{\partial x}F_{X,Y}(x,y) = \begin{cases} e^{-x} - e^{-y}, & 0 < x < y, \\ 0, & \text{en cualquier otro caso.} \end{cases}$$

La función de densidad de probabilidades de X es:

$$f_X(x) = \frac{d}{dx}F_X(x) = e^{-x}\mathbf{1}_{\{x>0\}} > 0 \Leftrightarrow x > 0,$$

por lo que aplicando (1.38) obtenemos para todo valor $x > 0$:

$$F_{Y|X}(y|x) = \frac{1}{f_X(x)}\frac{\partial}{\partial x}F_{X,Y}(x,y) = \begin{cases} 1 - e^{-(y-x)}, & y > x, \\ 0, & y \leq x. \end{cases}$$

$F_{Y|X}(y|x)$ es estrictamente creciente cuando $y > x$, e igual a la constante cero cuando $y \leq x$. Entonces para obtener una cuasi-inversa $F_{Y|X}^{(-1)}(t|x)$, para el caso $0 < t < 1$

(estRICTAMENTE CRECIENTE) simplemente resolvemos para y la ecuación $1 - e^{-(y-x)} = t$, obteniendo $y = x - \log(1-t)$, y para el caso $t = 0$ tenemos que escoger un valor $y_* \in]-\infty, x] \cap \overline{\text{Ran } Y|X=x} = \{x\}$, de donde finalmente:

$$F_{Y|X}^{(-1)}(t|x) = x - \log(1-t), \quad t \in [0, 1[.$$

Por lo tanto, la implementación del Algoritmo 3.4 para simular n observaciones iid del vector aleatorio (X, Y) de este ejemplo se resume como sigue:

- Simula u_1, \dots, u_n iid $\sim \text{Uniforme}(0, 1)$.
- Calcula $x_j = -\log(1-u_j)$, para cada $j \in \{1, \dots, n\}$.
- Simula v_1, \dots, v_n iid $\sim \text{Uniforme}(0, 1)$.
- Calcula $y_j = x_j - \log(1-v_j)$, para cada $j \in \{1, \dots, n\}$.
- $(x_1, y_1), \dots, (x_n, y_n)$ son iid $\sim F_{X,Y}$.

Observación:

- Los algoritmos 3.3 y 3.4 dependen de que sea posible obtener fórmulas explícitas para la función cuasi-inversa marginal de al menos una de las variables aleatorias, así como de la función cuasi-inversa condicional de la otra variable.
- En caso de que lo anterior no sea posible, pueden utilizarse aproximaciones numéricas de ellas, manteniendo control de la precisión de las mismas.
- Al ser métodos generales, no necesariamente son la forma más sencilla y eficiente de simular en todos los casos. Pueden existir mejores algoritmos que sean equivalentes en algunos casos específicos.

4. Actividades

4.1 Demuestra la Proposición 1.6.

4.2 En la doble desigualdad (1.4), si denotamos $u := \mathbb{P}(A)$, $v := \mathbb{P}(B)$, y $\Delta(u, v) := \delta(A, B)$, podemos reexpresarla como:

$$g(u, v) \leq \Delta(u, v) \leq h(u, v)$$

donde $g(u, v) := -\min\{(1-u)(1-v), uv\}$ y $h(u, v) := \min\{u(1-v), (1-u)v\}$. Grafica las superficies $g, h : [0, 1]^2 \rightarrow [0, 1]$ y calcula los valores máximo y mínimo de la función $\Delta : [0, 1]^2 \rightarrow [0, 1]$.

4.3 Sea X una variable aleatoria Bernoulli con parámetro $0 < \theta_1 < 1$, y sea Y otra variable aleatoria Bernoulli con parámetro $0 < \theta_2 < 1$. Construye la familia de todas las funciones de masa de probabilidades bivariadas distintas que existen para el vector aleatorio discreto (X, Y) . *Sugerencia: Aplica la Proposición 1.6.*

4.4 Utilizando lo desarrollado en los ejemplos 5 y 8:

1. Grafica la función $f_{X,Y}$ de dos formas: como superficie en \mathbb{R}^3 y por medio de conjuntos de nivel en \mathbb{R}^2 .

2. Simula $n = 3,000$ observaciones iid de (X, Y) y grafica lo siguiente:

a) Diagrama de dispersión (scatter plot) de las observaciones, comparándola versus la gráfica de la superficie $\{(x, y, f_{X,Y}(x, y)) : (x, y) \in \mathbb{R}^2\} \subset \mathbb{R}^3$.

b) Histograma estandarizado⁴ de las observaciones de X , comparándolo versus la función de densidad f_X .

c) Histograma estandarizado de las observaciones de Y , comparándolo versus la función de densidad f_Y .

4.5 Considera un vector aleatorio absolutamente continuo (X, Y) con función de densidad constante (positiva) sobre un triángulo $\mathcal{T} \subset \mathbb{R}^2$ con vértices $\{(-1, 0), (0, 1), (2, 0)\}$, e igual a cero fuera de dicho triángulo.

1. Deduce explícitamente la fórmula de la función de densidad conjunta $f_{X,Y}$.

2. Deduce explícitamente las funciones de densidad marginales f_X y f_Y , así como las condicionales $f_{Y|X}$ y $f_{X|Y}$.

3. Simula $n = 3,000$ observaciones iid de (X, Y) y grafica lo siguiente:

a) Diagrama de dispersión (scatter plot) de las observaciones, comparándola versus la gráfica de la superficie $\{(x, y, f_{X,Y}(x, y)) : (x, y) \in \mathbb{R}^2\} \subset \mathbb{R}^3$.

b) Histograma estandarizado de las observaciones de X , comparándolo versus la función de densidad f_X .

c) Histograma estandarizado de las observaciones de Y , comparándolo versus la función de densidad f_Y .

⁴De modo que el área total del mismo sea igual a 1.

CAPÍTULO 2

FUNCIONES CÓPULA

1.	Teorema de Sklar	46
2.	Construcción de familias de cópulas	54
3.	Simulación	77
4.	Estimación	86
5.	Actividades	95

La modelación de la dependencia entre variables aleatorias es un tema fundamental en estadística multivariada. En este contexto, el *Teorema de Sklar* ofrece una herramienta poderosa al haber demostrado que toda función de distribución conjunta de probabilidades de un vector aleatorio puede descomponerse en sus distribuciones marginales y una *función cópula* que captura exclusivamente la estructura de dependencia.

Este capítulo introduce de manera sistemática el concepto de función cópula, explorando sus principales consecuencias teóricas y aplicaciones prácticas. Se presentan las propiedades fundamentales que toda cópula debe cumplir, así como diversos métodos para su construcción, simulación y estimación. A lo largo del capítulo se abordan tanto las cópulas subyacentes –aquellas que emergen naturalmente de vectores aleatorios– como las cópulas bivariadas en sentido abstracto, que pueden definirse sin referencia a distribuciones específicas.

1. Teorema de Sklar

En el capítulo anterior recordamos que a partir de un vector aleatorio (X, Y) con función de distribución conjunta $F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ siempre es posible deducir las funciones de distribución marginal de cada variable:

$$F_X(x) = F_{X,Y}(x, +\infty), \quad F_Y(y) = F_{X,Y}(+\infty, y),$$

pero que el solo conocimiento de las marginales F_X y F_Y no permite, en general, reconstruir $F_{X,Y}$, y que de hecho existe una infinidad no numerable de soluciones posibles con exactamente las mismas marginales, recuérdese el Corolario 2.12 del capítulo 1 (1.29), mismas que se encuentran acotadas por lo que se conoce como las *cotas de Fréchet-Hoeffding*:

$$H_1(x, y) := \max\{F_X(x) + F_Y(y) - 1, 0\} \leq F_{X,Y}(x, y) \leq \min\{F_X(x), F_Y(y)\} =: H_3(x, y). \quad (2.1)$$

H_1 y H_3 , junto con $H_2(x, y) := F_X(x)F_Y(y)$ (que corresponde al caso particular de independencia, es decir $X \perp Y$), tienen las mismas marginales que $F_{X,Y}$, entre muchas otras que podrían definirse. Esto significa que en el proceso de *marginalización* de $F_{X,Y}$ se pierde la información relativa al tipo de particular de dependencia entre las variables aleatorias de (X, Y) . Dicha información se encuentra en una relación funcional entre la función de distribución conjunta y sus marginales, según lo demostró en 1959 Abe Sklar, «*Fonctions de répartition à n dimensions et leurs marges*». El teorema que ahora lleva su apellido fue demostrado en un contexto mucho más general que el que se abordará en el presente texto, ya que es un teorema válido para más de dos variables aleatorias, y de cualquier tipo. En este texto nos enfocamos al caso de dos variables aleatorias continuas, y por tanto la redacción de dicho teorema se adapta a ello:

Teorema 1.1: Teorema de Sklar, 1^a parte

Sea (X, Y) un vector aleatorio de variables aleatorias continuas, con función de distribución conjunta $F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$. Entonces existe una única función $C_{X,Y}$ tal que:

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

^a $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, conjunto extendido de los números reales.

Demostración: Consultese el Teorema 2.3.3 en Nelsen, *An Introduction to Copulas* o bien directamente el artículo de Sklar, «*Fonctions de répartition à n dimensions et leurs marges*».

□

A la función $C_{X,Y}$ del teorema anterior se le denomina **función cópula subyacente** del vector aleatorio (X, Y) . Consecuencia inmediata:

Corolario 1.2

Independencia. Sea (X, Y) un vector aleatorio de variables aleatorias continuas. Entonces:

$$X \perp Y \iff C_{X,Y}(u, v) = uv. \quad (2.3)$$

Demostración:

Recordemos que

$$X \perp Y \Leftrightarrow F_{X,Y}(x, y) = F_X(x)F_Y(y),$$

pero por el Teorema 1.1 existe una única cópula subyacente $C_{X,Y}$ tal que

$$F_{X,Y}(x, y) = C_{X,Y}(F_X(x), F_Y(y))$$

por lo que necesariamente $C_{X,Y}(F_X(x), F_Y(y)) = F_X(x)F_Y(y)$ para cualesquiera $x, y \in \overline{\mathbb{R}}$. Haciendo el cambio de variables $u := F_X(x)$ y $v := F_Y(y)$ se concluye que $C_{X,Y}(u, v) = uv$.

En lo sucesivo, y por su importancia, reservaremos la notación especial Π para referirnos a la cópula subyacente de un par de variables aleatorias continuas independientes, que por lo anterior es única y se define $\Pi(u, v) := uv$, llamándola **cópula producto** o bien **cópula independencia**.

A partir de las propiedades de función de distribución marginal y conjunta (véanse Teoremas 2.8 y 2.11 del capítulo 1) se deducen algunas propiedades que cumple la cópula subyacente de cualquier vector aleatorio bivariado de variables aleatorias continuas:

Teorema 1.3: Algunas propiedades de $C_{X,Y}$

Sea $C_{X,Y}$ la (función) cópula subyacente del vector aleatorio (X, Y) de variables aleatorias continuas con función de distribución conjunta $F_{X,Y}$. Entonces:

- a) $C_{X,Y}$ es una función con dominio $[0, 1]^2$ y codominio $[0, 1]$.
- b) $C_{X,Y}(u, 0) = 0 = C_{X,Y}(0, v)$ para cualesquiera $u, v \in [0, 1]$.
- c) $C_{X,Y}(u, 1) = u$ y $C_{X,Y}(1, v) = v$ para cualesquiera $u, v \in [0, 1]$.
- d) Para cualesquiera $0 \leq u_1 \leq u_2 \leq 1$ y $0 \leq v_1 \leq v_2 \leq 1$

$$C_{X,Y}(u_2, v_2) - C_{X,Y}(u_2, v_1) - C_{X,Y}(u_1, v_2) + C_{X,Y}(u_1, v_1) \geq 0.$$

Demostración: Por el Teorema 1.1 $F_{X,Y}(x, y) = C_{X,Y}(F_X(x), F_Y(y))$. Entonces:

- a) Como X e Y son variables aleatorias continuas, sus funciones de distribución marginales F_X y F_Y son continuas en todo su dominio extendido $\overline{\mathbb{R}}$. Como $F_X(-\infty) = 0$ y $F_X(+\infty) = 1$, además de ser monótona creciente, entonces necesariamente $\text{Ran } F_X =$

$\{F_X(x) : x \in \overline{\mathbb{R}}\} = [0, 1]$, y lo análogo para F_Y , por lo que necesariamente $C_{X,Y}$ está definida en todo el dominio $\text{Ran } F_X \times \text{Ran } F_Y = [0, 1]^2$. Y como $F_{X,Y}(x, y)$ es una probabilidad que se encuentra en el intervalo $[0, 1]$, entonces necesariamente $C_{X,Y}$ reporta valores que pertenecen a dicho conjunto, es decir, su codominio es $[0, 1]$.

b) Como $F_{X,Y}(x, -\infty) = 0$, aplicando el Teorema 1.1 tenemos que

$$0 = F_{X,Y}(x, -\infty) \stackrel{\text{Sklar}}{=} C_{X,Y}(F_X(x), F_Y(-\infty)) = C_{X,Y}(F_X(x), 0),$$

en donde haciendo el cambio de variable $u := F_X(x)$ obtenemos $C_{X,Y}(u, 0) = 0$, para cualquier $u \in [0, 1]$ porque para cada $u \in [0, 1]$ necesariamente existe un $x \in \overline{\mathbb{R}}$ tal que $F_X(x) = u$, por la continuidad de F_X . Y de forma análoga $C_{X,Y}(0, v) = v$ para cualquier $v \in [0, 1]$.

c) Como $F_{X,Y}(x, +\infty) = F_X(x)$, aplicando el Teorema 1.1 tenemos que

$$F_X(x) = F_{X,Y}(x, +\infty) \stackrel{\text{Sklar}}{=} C_{X,Y}(F_X(x), F_Y(+\infty)) = C_{X,Y}(F_X(x), 1),$$

en donde haciendo el cambio de variable $u := F_X(x)$ obtenemos $C_{X,Y}(u, 1) = u$, para cualquier $u \in [0, 1]$, y de forma análoga $C_{X,Y}(1, v) = v$ para cualquier $v \in [0, 1]$.

d) Por la propiedad 2-creciente de las funciones de distribución conjunta bivariadas tenemos que para cualesquiera $x_1 < x_2$, $y_1 < y_2$,

$$F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1) \geq 0,$$

y aplicando el Teorema 1.1 con el cambio de variables $u_j = F_X(x_j)$, $v_j = F_Y(y_j)$ se obtiene que para cualesquiera $0 \leq u_1 \leq u_2 \leq 1$, $0 \leq v_1 \leq v_2 \leq 1$,

$$C_{X,Y}(u_2, v_2) - C_{X,Y}(u_2, v_1) - C_{X,Y}(u_1, v_2) + C_{X,Y}(u_1, v_1) \geq 0.$$

□

Aplicando el Teorema 1.1 a las cotas de Fréchet-Hoeffding (2.1), y haciendo los cambios de variable $u := F_X(x)$ y $v := F_Y(y)$, obtenemos cotas para la cópula subyacente de cualquier vector aleatorio (X, Y) de variables aleatorias continuas:

$$\max\{u + v - 1, 0\} \leq C_{X,Y}(u, v) \leq \min\{u, v\}, \quad \forall (u, v) \in [0, 1]^2. \quad (2.4)$$

El siguiente resultado nos proporciona interpretaciones probabilísticas de las cotas de Fréchet-Hoeffding (2.4):

Teorema 1.4: Cotas de Fréchet-Hoeffding

Sea (X, Y) un vector aleatorio de variables aleatorias absolutamente continuas.

- a) Si $Y = \varphi(X)$, donde φ es una función estrictamente creciente, entonces $C_{X,Y}(u, v) = \min\{u, v\}$.
- b) Si $C_{X,Y}(u, v) = \min\{u, v\}$ entonces existe una función estrictamente creciente φ tal que $\mathbb{P}(Y = \varphi(X)) = 1$ (o equivalentemente, para X en función de Y .)
- c) Si $Y = \psi(X)$, donde ψ es una función estrictamente decreciente, entonces $C_{X,Y}(u, v) = \max\{u + v - 1, 0\}$.
- d) Si $C_{X,Y}(u, v) = \max\{u + v - 1, 0\}$ entonces existe una función estrictamente decreciente ψ tal que $\mathbb{P}(Y = \psi(X)) = 1$ (o equivalentemente, para X en función de Y .)

Demostración: a) Si $Y = \varphi(X)$, donde φ es una función estrictamente creciente, entonces

$$\begin{aligned} F_{X,Y}(x, y) &= \mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x, \varphi(X) \leq y) = \mathbb{P}(X \leq x, X \leq \varphi^{-1}(y)), \\ &= \mathbb{P}(X \leq \min\{x, \varphi^{-1}(y)\}) = F_X(\min\{x, \varphi^{-1}(y)\}), \\ &= \min\{F_X(x), F_X(\varphi^{-1}(y))\} = \min\{F_X(x), F_Y(y)\}, \end{aligned}$$

ya que $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\varphi(X) \leq y) = \mathbb{P}(X \leq \varphi^{-1}(y)) = F_X(\varphi^{-1}(y))$. Pero por el Teorema 1.1 se tiene que $F_{X,Y}(x, y) = C_{X,Y}(F_X(x), F_Y(y))$ por lo que necesariamente $C_{X,Y}(F_X(x), F_Y(y)) = \min\{F_X(x), F_Y(y)\}$, en donde haciendo el cambio de variable $u = F_X(x)$ y $v = F_Y(y)$ se concluye que $C_{X,Y}(u, v) = \min\{u, v\}$.

b) Si la cópula subyacente es $C_{X,Y}(u, v) = \min\{u, v\}$. Entonces por el Teorema 1.1

$$F_{X,Y}(x, y) = C_{X,Y}(F_X(x), F_Y(y)) = \min\{F_X(x), F_Y(y)\},$$

a partir de lo cual, aplicando (1.38), obtenemos la distribución condicional

$$\begin{aligned} F_{Y|X}(y | x) &= \mathbb{P}(Y \leq y | X = x) = \frac{1}{f_X(x)} \frac{\partial}{\partial x} F_{X,Y}(x, y), \quad \forall x : f_X(x) > 0, \\ &= \frac{1}{f_X(x)} \frac{\partial}{\partial x} \min\{F_X(x), F_Y(y)\} = \mathbf{1}_{\{F_X(x) \leq F_Y(y)\}}, \\ &= \mathbf{1}_{\{y \geq F_Y^{(-1)}(F_X(x))\}}, \end{aligned}$$

y por tanto $\mathbb{P}(Y = F_Y^{(-1)}(F_X(x)) | X = x) = 1$. Definamos $\varphi(x) := F_Y^{(-1)}(F_X(x))$. Por definición, $F_Y^{(-1)}$ es estrictamente creciente, pero F_X es monótona creciente, no necesariamente estrictamente creciente, así que, en principio, la composición $F_Y^{(-1)} \circ F_X$ es montotónica creciente. Sin embargo, dicha composición solo se comporta constante sobre conjuntos donde F_X sea constante, que implica que son conjuntos de probabilidad cero, por lo tanto con probabilidad 1 tenemos que φ es estrictamente creciente, de donde se concluye que $\mathbb{P}(Y = \varphi(X)) = 1$.

Las demostraciones c) y d) son análogas, y se dejan como ejercicio. \square

Por su importancia, se acostumbra utilizar notación específica para las cópulas que representan las cotas de Fréchet-Hoeffding: $W(u, v) := \max\{u + v - 1, 0\}$ y $M(u, v) := \min\{u, v\}$. Así, la cópula subyacente de cualquier vector aleatorio está acotada por las cópulas W y M :

$$W(u, v) \leq C_{X,Y}(u, v) \leq M(u, v), \quad \forall (u, v) \in [0, 1]^2. \quad (2.5)$$

Hasta el momento han sido identificadas tres cópulas específicas que se asocian a interpretaciones probabilísticas relevantes: Por un lado, el caso de $\Pi(u, v) = uv$ que es la cópula subyacente a cualquier par de variables aleatorias continuas independientes (ausencia total de dependencia); y en otro extremo, las cotas W y M que son cópulas subyacentes en el caso de total dependencia de las variables, esto es, cuando dado el valor de una, el de la otra queda determinado con probabilidad 1.

El proceso de identificación de las cópulas Π , W y M fue, en los tres casos, a partir de expresiones para $F_{X,Y}$ en términos de sus marginales F_X y F_Y , lo cual al aplicar el Teorema de Sklar 1.1 facilita la identificación de la cópula subyacente $C_{X,Y}$. Pero para los casos en los que esto no sea posible o evidente, tenemos el siguiente corolario, que esencialmente expresa a la cópula subyacente en términos de la función de distribución conjunta y sus marginales:

Corolario 1.5

Sea (X, Y) un vector aleatorio de variables aleatorias continuas con función de distribución conjunta $F_{X,Y}$. Entonces la cópula subyacente $C_{X,Y}$ se obtiene mediante:

$$C_{X,Y}(u, v) = F_{X,Y}\left(F_X^{(-1)}(u), F_Y^{(-1)}(v)\right), \quad (u, v) \in [0, 1]^2. \quad (2.6)$$

Demostración:

Por Teorema 1.1 tenemos que

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

Como F_X y F_Y son funciones continuas entonces $\text{Ran } F_X = \{F_X(x) : x \in \overline{\mathbb{R}}\} = [0, 1]$, y análogamente $\text{Ran } F_Y = [0, 1]$, por lo que

$$\forall (u, v) \in [0, 1]^2 \exists (x, y) \in \overline{\mathbb{R}}^2 : F_X(x) = u, F_Y(y) = v \implies x = F_X^{(-1)}(u), y = F_Y^{(-1)}(v),$$

y por lo tanto

$$F_{X,Y}\left(F_X^{(-1)}(u), F_Y^{(-1)}(v)\right) = C_{X,Y}(u, v), \quad \forall (u, v) \in [0, 1]^2.$$

Corolario 1.6

Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad conjunta $f_{X,Y}$ y cópula subyacente $C_{X,Y}$. Entonces:

$$f_{X,Y}(x, y) = c_{X,Y}(F_X(x), F_Y(y)) f_X(x) f_Y(y), \quad (2.7)$$

en donde

$$c_{X,Y}(u, v) = \frac{\partial^2}{\partial u \partial v} C_{X,Y}(u, v). \quad (2.8)$$

Demostración: Por Teorema 1.1 tenemos que $F_{X,Y}(x, y) = C_{X,Y}(F_X(x), F_Y(y))$, pero por hipótesis y aplicando la regla de la cadena del cálculo diferencial

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} C_{X,Y}(F_X(x), F_Y(y)), \\ &= c_{X,Y}(F_X(x), F_Y(y)) \frac{d}{dx} F_X(x) \frac{d}{dy} F_Y(y) = c_{X,Y}(F_X(x), F_Y(y)) f_X(x) f_Y(y). \end{aligned}$$

□

En el corolario anterior es importante enfatizar que se partió del supuesto que existe una función de densidad conjunta, pero recordemos que es posible tener vectores aleatorios de variables aleatorias continuas que no son absolutamente continuos, es decir, cuya función de densidad conjunta no existe. Tal es el caso, por ejemplo, cuando la cópula subyacente es alguna de las cotas de Fréchet-Hoeffding, véase Teorema 1.4, ya que en tal caso:

$$\frac{\partial^2}{\partial u \partial v} W(u, v) = 0 = \frac{\partial^2}{\partial u \partial v} M(u, v),$$

en contraste cuando la cópula subyacente es $C_{X,Y}(u, v) = \Pi(u, v) = uv$, ya que en tal caso:

$$c_{X,Y}(u, v) = \frac{\partial^2}{\partial u \partial v} uv = 1,$$

y si individualmente X e Y son absolutamente continuas con funciones de densidad marginales f_X y f_Y , respectivamente, entonces por el corolario anterior:

$$f_{X,Y}(x, y) = c_{X,Y}(F_X(x), F_Y(y)) f_X(x) f_Y(y) = f_X(x) f_Y(y),$$

como era de esperarse.

El corolario siguiente nos permite interpretar también a una cópula bivariada como una función de distribución conjunta de dos variables aleatorias marginalmente continuas uniformes sobre $[0, 1]$ cada una, restringida al cuadrado unitario $[0, 1]^2$:

Corolario 1.7

Sea (U, V) un vector aleatorio de variables aleatorias absolutamente continuas, cada una con distribución marginal uniforme sobre el intervalo $[0, 1]$. Entonces:

$$C_{U,V}(u, v) = F_{U,V}(u, v), \quad \forall (u, v) \in [0, 1]^2. \quad (2.9)$$

Demostración:

Como $f_U(u) = \mathbf{1}_{\{0 \leq u \leq 1\}}$ entonces:

$$F_U(u) = \int_{-\infty}^u f_U(t) dt = u \mathbf{1}_{\{0 \leq u \leq 1\}} + \mathbf{1}_{\{u > 1\}},$$

y similar fórmula para F_V . Por el Teorema 1.1 existe una única cópula $C_{U,V}$ tal que:

$$\begin{aligned} F_{U,V}(u, v) &= C_{U,V}(F_U(u), F_V(v)), \quad \forall (u, v) \in \overline{\mathbb{R}}^2, \\ &= \begin{cases} C_{U,V}(u, v), & (u, v) \in [0, 1]^2 \\ C_{U,V}(u, 1) = u, & 0 \leq u \leq 1, v > 1, \\ C_{U,V}(1, v) = v, & u > 1, 0 \leq v \leq 1, \\ 0, & u < 0 \text{ y/o } v < 0. \end{cases} \end{aligned}$$

Si un vector aleatorio (X, Y) con función de distribución conjunta $F_{X,Y}$ es transformado en otro mediante algo de la forma $(S, T) = (\varphi(X), \psi(Y))$, salvo el caso de la transformación identidad, al deducir $F_{S,T}$ con frecuencia resulta una función de distribución conjunta distinta a $F_{X,Y}$. Analizaremos el caso de algunas transformaciones particulares, y su consecuencia en la cópula subyacente:

Teorema 1.8: Invariancia ante transformaciones crecientes

Sea (X, Y) un vector de variables aleatorias continuas. Si $\varphi : \text{Ran } X \rightarrow \mathbb{R}$ y $\psi : \text{Ran } Y \rightarrow \mathbb{R}$ son funciones estrictamente crecientes, entonces la cópula subyacente de (X, Y) y de $(\varphi(X), \psi(Y))$ es la misma, es decir:

$$C_{X,Y} = C_{\varphi(X), \psi(Y)}.$$

Demostración: Definamos las variables aleatorias $S := \varphi(X)$ y $T := \psi(Y)$. Entonces

$$F_S(s) = \mathbb{P}(S \leq s) = \mathbb{P}(\varphi(X) \leq s) = \mathbb{P}(X \leq \varphi^{-1}(s)) = F_X(\varphi^{(-1)}(s)),$$

y de forma análoga $F_T(t) = F_Y(\psi^{-1}(t))$. Aplicando el Teorema 1.1

$$\begin{aligned} C_{S,T}(F_S(s), F_T(t)) &\stackrel{\text{Sklar}}{=} F_{S,T}(s, t) = \mathbb{P}(S \leq s, T \leq t) = \mathbb{P}(\varphi(X) \leq s, \psi(Y) \leq t), \\ &= \mathbb{P}(X \leq \varphi^{-1}(s), Y \leq \psi^{-1}(t)), \\ &= F_{X,Y}(\varphi^{-1}(s), \psi^{-1}(t)), \\ &\stackrel{\text{Sklar}}{=} C_{X,Y}(F_X(\varphi^{-1}(s)), F_Y(\psi^{-1}(t))), \\ &= C_{X,Y}(F_S(s), F_T(t)), \end{aligned}$$

y mediante el cambio de variables $u = F_S(s)$ y $v = F_T(t)$ concluimos que $C_{S,T}(u, v) = C_{X,Y}(u, v)$. \square

También es posible obtener relaciones sencillas entre la cópula subyacente de un vector aleatorio (X, Y) y la cópula subyacente de $(\varphi(X), \psi(Y))$ cuando φ y ψ son estrictamente monótonas, no necesariamente ambas crecientes, que se les conoce como *rotaciones*, por razones que se abordarán más adelante:

Teorema 1.9: Rotaciones

Sea (X, Y) un vector de variables aleatorias continuas. Sean $\varphi : \text{Ran } X \rightarrow \mathbb{R}$ y $\psi : \text{Ran } Y \rightarrow \mathbb{R}$ funciones estrictamente monótonas:^a

a) Si $\varphi \uparrow$ y $\psi \downarrow$ entonces:

$$C_{\varphi(X), \psi(Y)}(u, v) = u - C_{X,Y}(u, 1-v).$$

b) Si $\varphi \downarrow$ y $\psi \uparrow$ entonces:

$$C_{\varphi(X), \psi(Y)}(u, v) = v - C_{X,Y}(1-u, v).$$

c) Si $\varphi \downarrow$ y $\psi \downarrow$ entonces:

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

^aEstrictamente monótona equivale a estrictamente creciente o bien estrictamente decreciente.

Demostración: Definamos las variables $S := \varphi(X)$ y $T := \psi(Y)$.

a) Como $\varphi \uparrow$, de la demostración del Teorema 1.8 tenemos que $F_S(s) = F_X(\varphi^{-1}(s))$, pero como $\psi \downarrow$ entonces para F_T :

$$F_T(t) = \mathbb{P}(T \leq t) = \mathbb{P}(\psi(Y) \leq t) = \mathbb{P}(Y \geq \psi^{-1}(t)) = 1 - F_Y(\psi^{-1}(t)).$$

Aplicando el Teorema 1.1

$$\begin{aligned}
 C_{S,T}(F_S(s), F_T(t)) &\stackrel{\text{Sklar}}{=} F_{S,T}(s, t) = \mathbb{P}(S \leq s, T \leq t) = \mathbb{P}(\varphi(X) \leq s, \psi(Y) \leq t), \\
 &= \mathbb{P}(X \leq \varphi^{-1}(s), Y \geq \psi^{-1}(t)), \\
 &= \mathbb{P}(X \leq \varphi^{-1}(s)) - \mathbb{P}(X \leq \varphi^{-1}(s), Y \leq \psi^{-1}(t)), \\
 &= F_X(\varphi^{-1}(s)) - F_{X,Y}(\varphi^{-1}(s), \psi^{-1}(t)), \\
 &\stackrel{\text{Sklar}}{=} F_X(\varphi^{-1}(s)) - C_{X,Y}(F_X(\varphi^{-1}(s)), F_Y(\psi^{-1}(t))), \\
 &= F_S(s) - C_{X,Y}(F_S(s), 1 - F_T(t)),
 \end{aligned}$$

y mediante el cambio de variables $u = F_S(s)$ y $v = F_T(t)$ concluimos que $C_{S,T}(u, v) = u - C_{X,Y}(u, 1 - v)$.

- b) Totalmente análoga a la del inciso a) anterior.
- c) Del inciso a) utilizamos el caso particular $C_{X,\psi(Y)}(u, v) = u - C_{X,Y}(u, 1 - v)$, y luego a este resultado aplicamos el inciso b) con transformación identidad para $\psi(Y)$,

$$\begin{aligned}
 C_{\varphi(X),\psi(Y)}(u, v) &= v - C_{X,\psi(Y)}(1 - u, v), \\
 &= v - [(1 - u) - C_{X,Y}(1 - u, 1 - v)], \\
 &= u + v - 1 + C_{X,Y}(1 - u, 1 - v).
 \end{aligned}$$

□

2. Construcción de familias de cópulas

La primera parte del Teorema de Sklar (Teorema 1.1) nos dice que, dado un vector aleatorio (X, Y) de variables aleatorias continuas, existe una relación funcional única $C_{X,Y}$ entre la función de distribución conjunta $F_{X,Y}$ y sus marginales F_X y F_Y , y a dicha relación funcional $C_{X,Y}$ la hemos denominado *cópula subyacente del vector aleatorio* (X, Y) . A partir de ello dedujimos algunas propiedades que cumple toda cópula subyacente de un vector aleatorio.

Ahora, sin partir de un vector aleatorio, abordaremos qué puede derivarse por medio de cualquier función que simplemente cumpla las propiedades del Teorema 1.3, esto es, proponiendo la siguiente:

Definición 2.1: Cópula bivariada (o 2-cópula)

Denominaremos **cópula bivariada** a cualquier función $C : [0, 1]^2 \rightarrow [0, 1]$ que cumpla lo siguiente:

- a) $C(u, 0) = 0 = C(0, v)$, para cualesquiera $u, v \in [0, 1]$.
- b) $C(u, 1) = u$ y $C(1, v) = v$, para cualesquiera $u, v \in [0, 1]$.
- c) Propiedad 2-creciente: Para cualesquiera $0 \leq u_1 \leq u_2 \leq 1$ y $0 \leq v_1 \leq v_2 \leq 1$,

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

Respecto a la definición anterior es importante enfatizar que, estrictamente, no involucra concepto probabilístico alguno, no se deriva de un vector aleatorio dado, en contraste con el concepto de la sección anterior denominado *cópula subyacente de un vector aleatorio*, y es por ello que simplemente se utiliza la notación C y no $C_{X,Y}$. Nótese que las propiedades a) y b) definen que todas las cópulas se comportan igual en la *frontera* de su dominio, y por tanto las diferencias entre unas y otras ocurre en el *interior*, con la única restricción de que cumplan la propiedad 2-creciente.

Teorema 2.2: Teorema de Sklar, 2^a parte

Sean $F : \overline{\mathbb{R}} \rightarrow [0, 1]$ y $G : \overline{\mathbb{R}} \rightarrow [0, 1]$ dos funciones de distribución univariadas cualesquiera, y sea $C : [0, 1]^2 \rightarrow [0, 1]$ cópula bivariada cualquiera. Si se define la función $H : \overline{\mathbb{R}}^2 \rightarrow [0, 1]$ mediante:

$$H(x, y) := C(F(x), G(y)) \quad (2.10)$$

entonces H cumple con todas las propiedades de una función de distribución conjunta de dos variables aleatorias continuas (véase Teorema 2.11 del Capítulo 1), que en particular tiene como funciones de distribución marginales a F y G .

Demostración: Aplicando el inciso b) de la Definición 2.1 marginalizamos

$$H(x, +\infty) = C(F(x), G(+\infty)) = C(F(x), 1) \stackrel{\text{b)}}{=} F(x),$$

y lo análogo para la otra marginal. Para el resto de las propiedades consultese el Teorema 2.3.3 en Nelsen, *An Introduction to Copulas* o bien directamente el artículo de Sklar, «*Fonctions de répartition à n dimensions et leurs marges*». \square

Esta segunda parte del Teorema de Sklar resulta especialmente útil para construir distribuciones conjuntas de probabilidad con gran flexibilidad ya que podemos:

- Escoger cualquier cópula bivariada C .

- Escoger cualesquiera marginales univariadas F y G .
- Definir la distribución conjunta $H(x, y) := C(F(x), G(y))$.

A partir de la Definición 2.1 y de conocidos resultados de funciones reales de varias variables reales, pueden deducirse las siguientes propiedades para cualquier cópula bivariada:

Teorema 2.3: Propiedades de cópulas bivariadas

Sea $C : [0, 1]^2 \rightarrow [0, 1]$ una cópula bivariada cualquiera.

- a) Cotas de Fréchet-Hoeffding:

$$W(u, v) \leq C(u, v) \leq M(u, v), \quad \forall (u, v) \in [0, 1]^2,$$

donde las funciones $W(u, v) = \max\{u + v - 1, 0\}$ y $M(u, v) = \min\{u, v\}$ también cumplen con la Definición 2.1 de cópula.

- b) C es uniformemente continua en todo su dominio, es decir para cualesquiera $(u_1, v_1), (u_2, v_2) \in [0, 1]^2$ se cumple:

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

- c) Para cualquier $v \in [0, 1]$ existe la derivada parcial $\frac{\partial}{\partial u} C(u, v)$ para casi cualquier u (es decir, excepto posiblemente para algunos puntos aislados), y en tal caso está acotada:

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

y lo análogo para la otra derivada parcial $\frac{\partial}{\partial v} C(u, v)$.

Demostración: Se deja como ejercicio. Se recomienda consultar las secciones 2.2, 2.3 y 2.4 en Nelsen, *An Introduction to Copulas*. □

Cualquier cópula bivariada $C : [0, 1]^2 \rightarrow [0, 1]$ puede graficarse como una superficie en \mathbb{R}^3 a partir del conjunto de puntos:

$$\{(u, v, C(u, v)) : (u, v) \in [0, 1]^2\} \subset [0, 1]^3,$$

o bien en términos de sus *conjuntos de nivel*:

$$\mathcal{N}_C(t) = \{(u, v) \in [0, 1]^2 : C(u, v) = t\}, \quad t \in [0, 1]. \quad (2.11)$$

En el caso particular de las cópulas W , Π y M los conjuntos de nivel resultan ser *curvas de nivel*, véase la Figura 2.1, pero es factible construir cópulas con conjuntos de nivel que tengan área positiva.

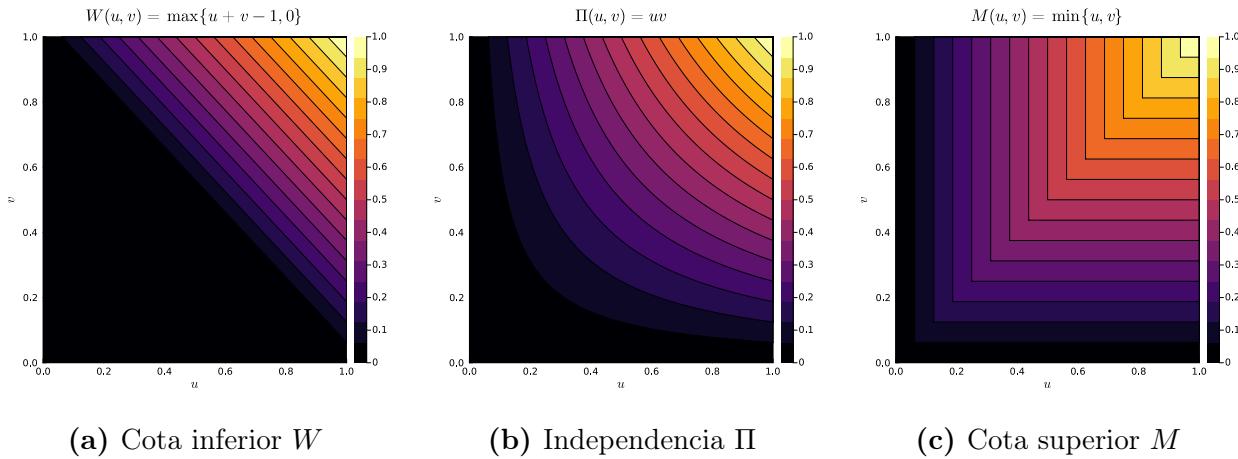


Figura 2.1: Curvas de nivel de las cópulas W , Π y M .

Para un valor específico $0 < t < 1$ el conjunto de nivel t de cualquier cópula C está necesariamente contenido en el triángulo Δ_t con vértices en $\{(t,t), (1,t), (t,1)\}$, esto es $\mathcal{N}_C(t) \subset \Delta_t$, véase la Figura 2.2.

De forma complementaria, también resulta útil analizar y graficar cópulas por secciones, esto es, para cualquier $t \in [0, 1]$, sus *secciones horizontales* $t \rightarrow C(t, v_0)$ para cualquier $v_0 \in [0, 1]$, sus *secciones verticales* $t \rightarrow C(u_0, t)$ para cualquier $u_0 \in [0, 1]$, así como su *sección diagonal* $\delta(t) = C(t, t)$, y su *sección diagonal secundaria* $\lambda(t) = C(t, 1-t)$, véase la Figura 2.3.

Como consecuencia inmediata del Teorema 2.3 se deduce el siguiente:

Corolario 2.4

Sea C una cópula bivariada cualquiera. Entonces sus secciones horizontales, verticales y diagonal son monótonas crecientes y uniformemente continuas, y su sección diagonal secundaria solamente uniformemente continua.

El Corolario 1.7, véase (2.9), nos brinda una interpretación de las cópulas bivariadas como función de distribución conjunta de dos variables aleatorias continuas, cada una marginalmente uniforme sobre el intervalo $[0, 1]$. Recordemos que esto no garantiza que sea absolutamente continua, es decir, que exista una función de densidad conjunta:

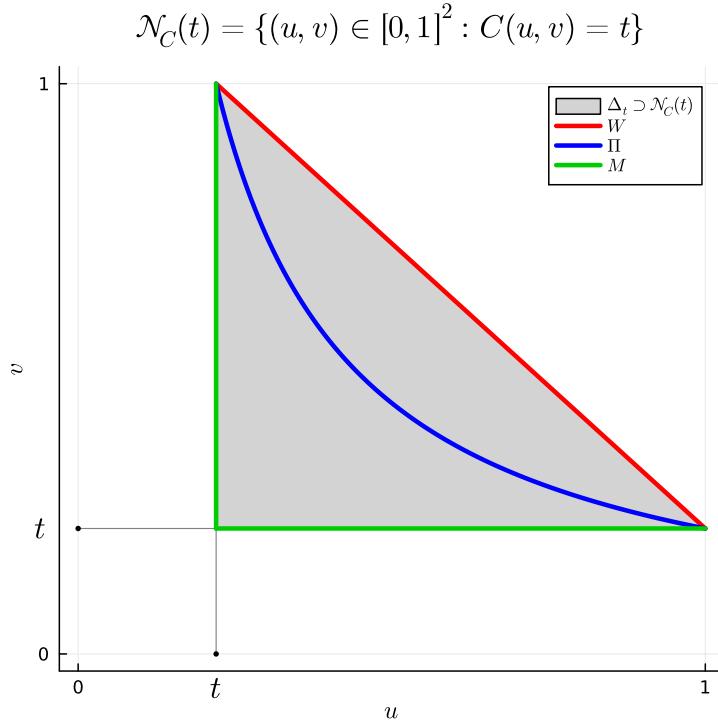


Figura 2.2: Δ_t es la región que contiene todos los conjuntos de nivel t . Curvas de nivel t de las cópulas W (rojo), Π (azul) y M (verde).

Definición 2.5: Cópulas absolutamente continuas y singulares

Sea C una cópula bivariada cualquiera tal que existe su segunda derivada parcial mixta sobre todo el interior de su dominio $[0, 1]^2$, excepto posiblemente sobre algún conjunto de área cero:

- a) Si $\int_0^1 \int_0^1 \frac{\partial^2}{\partial u \partial v} C(u, v) dudv = 1$ se dice que C es una *cópula absolutamente continua con función de densidad*:

$$c(u, v) := \frac{\partial^2}{\partial u \partial v} C(u, v).$$

- b) Si $\frac{\partial^2}{\partial u \partial v} C(u, v) = 0$ se dice que C es una *cópula singular*.

- c) Si $0 < \frac{\partial^2}{\partial u \partial v} C(u, v) < 1$ se dice que C es una cópula con componentes absolutamente continuo y singular.

Con base en el comentario posterior al Corolario 1.6, veáse (2.7), tenemos entonces que la cópula Π es un claro ejemplo de cópula absolutamente continua, mientras que W y M son ejemplos de cópulas singulares. Más adelante se abordará algún ejemplo de cópula con componente tanto singular como absolutamente continuo.

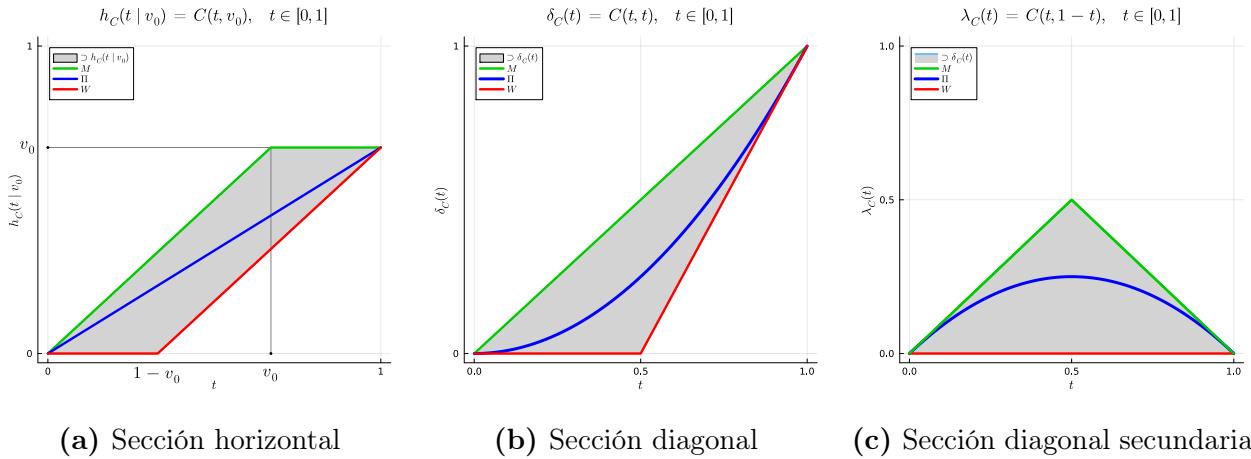


Figura 2.3: La zona gris indica el conjunto donde siempre están contenidas las secciones horizontales, diagonal y diagonal secundaria de cualquier cópula bivariada. En colores verde, azul y rojo se indican las secciones correspondientes a las cópulas M , Π y W , respectivamente. El caso de las secciones verticales es totalmente análogo al de las horizontales.

Corolario 2.6

Sea c la función de densidad de una cópula absolutamente continua. Sean G_1 y G_2 dos funciones de distribución univariadas absolutamente continuas, con funciones de densidad g_1 y g_2 , respectivamente. Entonces la función $h : \mathbb{R}^2 \rightarrow [0, +\infty[$ definida como:

$$h(x, y) := c(G_1(x), G_2(y))g_1(x)g_2(y) \quad (2.12)$$

es una función de densidad conjunta de probabilidades.

Demostración: Sea C la cópula tal que $c(u, v) = \frac{\partial^2}{\partial u \partial v}C(u, v)$. Por Teorema 2.2 tenemos que $H(x, y) := C(G_1(x), G_2(y))$ define una función de distribución conjunta con marginales G_1 y G_2 . Por hipótesis C , G_1 y G_2 son absolutamente continuas, así que aplicando la regla de la cadena del cálculo diferencial obtenemos:

$$\frac{\partial^2}{\partial x \partial y}H(x, y) = \frac{\partial^2}{\partial u \partial v}C(G_1(x), G_2(y))\frac{d}{dx}G_1(x)\frac{d}{dy}G_2(y) = h(x, y).$$

□

Consecuencia inmediata del corolario anterior, y del Corolario 1.7 (véase (2.9)), tenemos que si (U, V) es un vector aleatorio absolutamente continuo de variables aleatorias marginalmente continuas uniformes sobre el intervalo $[0, 1]$ cada una, entonces su función de densidad conjunta resulta igual a la densidad de su cópula subyacente, esto es:

$$f_{U,V}(u, v) = c_{U,V}(u, v) = \frac{\partial^2}{\partial u \partial v}C_{U,V}(u, v), \quad (2.13)$$

es decir, podemos interpretar a la densidad de una cópula (cuando existe) simplemente como la función de densidad conjunta de un par de variables aleatorias que marginalmente son

continuas uniformes sobre $[0, 1]$. Pero si la cópula subyacente de (U, V) es, por ejemplo, la cópula $M(u, v) = \min\{u, v\}$ entonces:

$$\frac{\partial^2}{\partial u \partial v} M(u, v) = 0,$$

es decir, M es una cópula singular, y por tanto no existe la densidad de la cópula ni densidad conjunta para (U, V) , pero lo que siempre existe es su función de distribución conjunta, que en este caso es $F_{U,V}(u, v) = M(u, v)$.

2.1. Método de inversión

Cualquier cópula bivariada puede utilizarse junto con cualquier par de funciones de distribución marginal univariadas para construir una función de distribución conjunta de probabilidades, como consecuencia del Teorema 2.2. Pero cómo proponer una función que cumpla con la Definición 2.1 no es una tarea trivial, especialmente por la propiedad 2-creciente. De aquí la necesidad de contar con diversos métodos para construir familias de cópulas.

Una fuente natural de funciones cópula surge como consecuencia de (2.6) que nos permite extraer la cópula subyacente de cualquier función de distribución conjunta bivariada conocida $F_{X,Y}$ por medio de:

$$C_{X,Y}(u, v) = F_{X,Y}\left(F_X^{(-1)}(u), F_Y^{(-1)}(v)\right), \quad (u, v) \in [0, 1]^2. \quad (2.14)$$

$F_{X,Y}$ tiene por marginales específicas a $F_X(x) = F_{X,Y}(x, +\infty)$ y $F_Y(y) = F_{X,Y}(+\infty, y)$, pero por medio de $C_{X,Y}$ y aplicando el Teorema 2.2 podemos construir una nueva función de distribución conjunta H que tenga el mismo tipo de dependencia que $F_{X,Y}$ pero con funciones de distribución marginal distintas, digamos G_1 y G_2 , esto es:

$$H(x, y) := C_{X,Y}\left(G_1(x), G_2(y)\right),$$

de modo que $C_H = C_{X,Y}$ pero $H \neq F_{X,Y}$.

Ejemplo 9. Misma cópula pero distintas marginales.

En el Ejemplo 5 del Capítulo 1 se abordó el caso de un vector aleatorio (X, Y) absolutamente continuo con función de densidad conjunta:

$$f_{X,Y}(x, y) = e^{-y} \mathbf{1}_{\{0 < x < y\}},$$

y a partir de ella se dedujo su función de distribución conjunta:

$$\begin{aligned} F_{X,Y}(x, y) &= \int_{-\infty}^x \int_{-\infty}^y e^{-t} \mathbf{1}_{\{0 < s < t\}} dt ds = \dots \\ &= \begin{cases} 1 - e^{-x} - xe^{-y}, & \text{si } 0 < x \leq y, \\ 1 - e^{-y} - ye^{-x}, & \text{si } 0 < y < x, \\ 0, & \text{en cualquier otro caso.} \end{cases} \end{aligned}$$

Marginalizando $F_{X,Y}$ se obtuvo:

$$F_X(x) = F_{X,Y}(x, +\infty) = (1 - e^{-x}) \mathbf{1}_{\{x > 0\}}, \quad F_Y(y) = F_{X,Y}(+\infty, y) = (1 - e^{-y} - ye^{-y}) \mathbf{1}_{\{y > 0\}},$$

que corresponden a variables aleatorias absolutamente continuas, con funciones de densidad marginales:

$$f_X(x) = \frac{d}{dx} F_X(x) = e^{-x} \mathbf{1}_{\{x > 0\}}, \quad f_Y(y) = \frac{d}{dy} F_Y(y) = ye^{-y} \mathbf{1}_{\{y > 0\}},$$

y que, de hecho, corresponden a modelos de probabilidad conocidos: marginalmente X resulta ser una variable aleatoria *Exponencial* con parámetro 1, y por otro lado Y resulta ser una variable aleatoria *Gamma* con parámetros (2, 1).

Para $x > 0$ tenemos que $F_X(x) = 1 - e^{-x}$ es estrictamente creciente, y en tal caso existe la función inversa $F_X^{-1}(u) = -\log(1 - u)$ para $0 < u \leq 1$. Si $x \leq 0$ entonces para efectos de una quasi-inversa compatible con $\text{Ran } X$ obtenemos:

$$F_X^{(-1)}(u) = -\log(1 - u), \quad u \in [0, 1].$$

Para $y > 0$ tenemos que $F_Y(y)$ también es estrictamente creciente y por tanto existe la función inversa $F_Y^{-1}(v)$ para $0 < v \leq 1$, aunque en este caso no es posible obtener una fórmula explícita, salvo en el caso $F_Y^{-1}(1) = +\infty$, por lo que tendrá que utilizarse algún método de aproximación numérica. Si $y \leq 0$ entonces para efectos de una quasi-inversa compatible con $\text{Ran } Y$ se utiliza $F_Y^{(-1)}(0) = 0$. Aplicando (2.14) obtenemos:

$$\begin{aligned} C_{X,Y}(u, v) &= F_{X,Y}\left(F_X^{(-1)}(u), F_Y^{(-1)}(v)\right), \\ &= \begin{cases} 1 - (1 - u) + \log(1 - u) \exp(-F_Y^{-1}(v)), & 0 < -\log(1 - u) \leq F_Y^{-1}(v), \\ 1 - \exp(-F_Y^{-1}(v)) - F_Y^{-1}(v) \exp(-F_Y^{-1}(v)), & F_Y^{-1}(v) < -\log(1 - u), \\ 0, & \text{en cualquier otro caso,} \end{cases} \end{aligned}$$

donde $-\log(1 - u) \leq F_Y^{-1}(v) \Leftrightarrow \xi(u) := u + (1 - u) \log(1 - u) \leq v$, por lo que:

$$C_{X,Y}(u, v) = \begin{cases} u + \exp(-F_Y^{-1}(v)) \log(1 - u), & v \geq \xi(u), \\ 1 - (1 + F_Y^{-1}(v)) \exp(-F_Y^{-1}(v)), & v < \xi(u). \end{cases}$$

Calculamos alguna primera derivada parcial, por ejemplo:

$$\frac{\partial}{\partial u} C_{X,Y}(u, v) = \begin{cases} 1 - \frac{\exp(-F_Y^{-1}(v))}{1-u}, & v > \xi(u), \\ 0, & v < \xi(u), \end{cases}$$

y luego

$$c_{X,Y}(u, v) = \frac{\partial^2}{\partial v \partial u} C_{X,Y}(u, v) = \frac{\exp(-F_Y^{-1}(v)) \frac{d}{dv} F_Y^{-1}(v)}{1 - u} \mathbf{1}_{\{v > \xi(u)\}},$$

donde, por propiedades de la derivada de la inversa de una función, y luego aplicando el hecho de que $f_Y(y) = ye^{-y} \mathbf{1}_{\{y > 0\}}$, se deduce que:

$$\frac{d}{dv} F_Y^{-1}(v) = \frac{1}{\frac{d}{dy} F_Y(y)} = \frac{1}{f_Y(F_Y^{-1}(v))} = \frac{1}{F_Y^{-1}(v) \exp(-F_Y^{-1}(v))},$$

por lo que, sustituyendo en la fórmula anterior de $c_{X,Y}$, obtenemos:

$$c_{X,Y}(u, v) = \frac{1}{(1-v)F_Y^{-1}(v)} \mathbf{1}_{\{v > \xi(u)\}}.$$

Además del hecho que $c_{X,Y}$ es la densidad de la cópula subyacente del vector aleatorio (X, Y) con función de densidad conjunta $f_{X,Y}$, recordemos que también podemos interpretar a $c_{X,Y}$ como una función de densidad conjunta con marginales uniformes en $[0, 1]$, que tiene la misma estructura de dependencia que $f_{X,Y}$, pero con marginales distintas.

Más aún, podríamos definir un nuevo vector aleatorio (S, T) que tenga la misma estructura de dependencia que (X, Y) pero con marginales $S \sim \text{Normal}(0, 1)$ y $T \sim \text{Beta}(\frac{1}{2}, \frac{1}{2})$, respectivamente, aplicando el Corolario 2.6, véase (2.12):

$$f_{S,T}(s, t) := c_{X,Y}(F_S(s), F_T(t))f_S(s)f_T(t).$$

Tenemos entonces tres vectores aleatorios (X, Y) , (U, V) y (S, T) , que tienen la misma estructura de dependencia, pero marginales distintas. En la Figura 2.4 pueden apreciarse las curvas de nivel de las tres funciones de densidad. Aunque se aprecian distintas, la dependencia es la misma, y se ven diferentes solamente por tener marginales distintas. ■

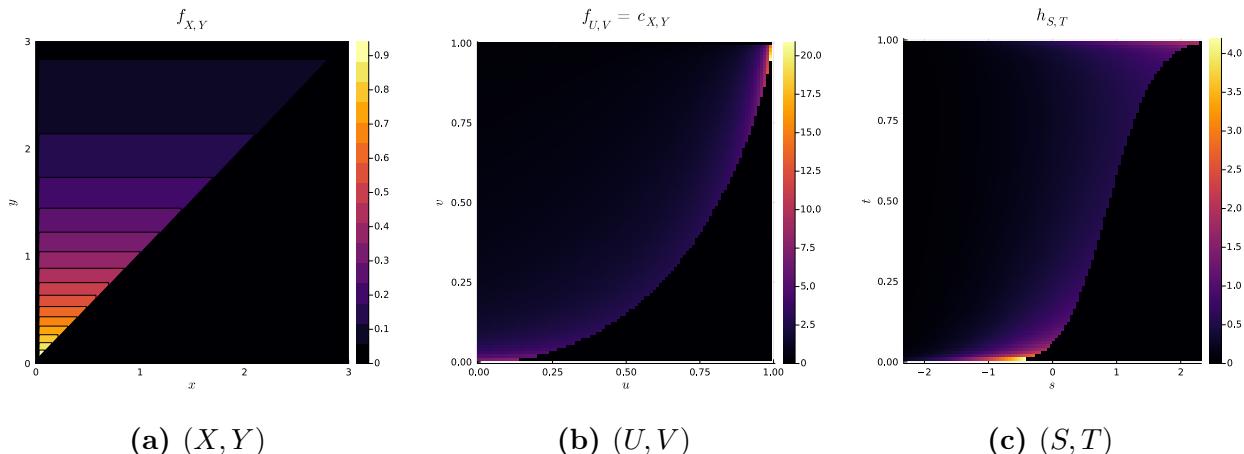


Figura 2.4: Ejemplo 9. Curvas de nivel de las densidades conjuntas de (X, Y) , (U, V) y (S, T) que tienen la misma cópula subyacente, y por tanto la misma estructura de dependencia, pero con marginales distintas.

2.2. Cópulas arquimedianas

Como motivación para introducir esta subsección, recordemos que en una *ecuación funcional* se tienen una o varias funciones que son desconocidas, pero que se asume cumplen ciertas propiedades. Por ejemplo, consideremos el conjunto de los números reales positivos $(\mathbb{R}^+, +, \cdot)$ con las operaciones usuales de suma (+) y multiplicación (\cdot) de números reales, así como la siguiente ecuación funcional:

$$\psi(x \cdot y) = \psi(x) + \psi(y), \quad \forall x, y \in \mathbb{R}^+, \quad (2.15)$$

en donde la función $\psi : \mathbb{R}^+ \rightarrow \mathbb{R}$ es desconocida, pero se supone que es diferenciable. Aplicando el *Teorema Fundamental del Cálculo*¹ se deduce que la solución de la ecuación funcional (2.15) es:

$$\psi(x) = \int_1^x \frac{dz}{z}, \quad \forall x > 0, \quad (2.16)$$

en donde, a pesar de que $\frac{1}{z}$ es una función integrable, no existe su antiderivada explícita. A la solución (2.16) se le conoce como *función logaritmo* y la notación usual es $\log x$.

Ahora consideremos una estructura algebraica sobre el intervalo cerrado $[0, 1]$ nuevamente con la operación binaria usual de suma de números reales ($+$) pero ahora sustituyendo la operación binaria usual de multiplicación de números reales por el operador binario \odot asociado a alguna cópula bivariada C y definido de la siguiente forma:

$$u \odot v := C(u, v), \quad \forall u, v \in [0, 1], \quad (2.17)$$

y ahora sobre la estructura algebraica $([0, 1], +, \odot)$ preguntémonos por la solución de la ecuación funcional:

$$\varphi(u \odot v) = \varphi(u) + \varphi(v), \quad \forall u, v \in [0, 1], \quad (2.18)$$

para alguna función $\varphi : [0, 1] \rightarrow [0, +\infty]$. En el caso particular $C(u, v) = \Pi(u, v) = uv$ tenemos que el operador \odot se convierte en la multiplicación usual de números reales, y en este caso la solución de:

$$\varphi(\Pi(u, v)) = \varphi(u) + \varphi(v) \quad (2.19)$$

sería $\varphi(t) = -\log t$, con $t \in [0, 1]$. La función inversa de φ existe y resulta ser la función $\varphi^{-1} : [0, +\infty] \rightarrow [0, 1]$ dada por $\varphi^{-1}(z) = e^{-z}$, y por tanto la cópula producto en (2.19) podría expresarse en términos de φ , es decir:

$$\Pi(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)). \quad (2.20)$$

El anterior es solo el caso particular de la cópula Π , pero esto motiva la pregunta sobre la posibilidad de generalizarlo para la construcción de una cópula bivariada C en términos de alguna función $\varphi_C : [0, 1] \rightarrow [0, +\infty]$, de modo que:

$$C(u, v) = \varphi_C^{-1}(\varphi_C(u) + \varphi_C(v)). \quad (2.21)$$

¹Consúltese, por ejemplo, el capítulo 6 de Tom M. Apostol (1967) *Calculus. Volume I*. Ed. John Wiley & Sons (Nueva York).

Teorema 2.7: Cópula arquimediana

Sea $\varphi : [0, 1] \rightarrow [0, +\infty]$ una función continua, estrictamente decreciente, convexa,^a y tal que $\varphi(1) = 0$. Entonces la función $C : [0, 1]^2 \rightarrow [0, 1]$ definida como:

$$C(u, v) := \varphi^{[-1]}(\varphi(u) + \varphi(v)) \quad (2.22)$$

es una cópula bivariada, donde la función $\varphi^{[-1]} : [0, +\infty] \rightarrow [0, 1]$ es la *pseudo-inversa* de φ definida como:

$$\varphi^{[-1]}(z) := \begin{cases} \varphi^{-1}(z), & 0 \leq z \leq \varphi(0), \\ 0, & \varphi(0) \leq z \leq +\infty. \end{cases} \quad (2.23)$$

^aUna función g es *convexa* si para cualesquiera x_1 y x_2 en su dominio se cumple:

$$g((1 - \alpha)x_1 + \alpha x_2) \leq (1 - \alpha)g(x_1) + \alpha g(x_2), \quad \forall \alpha \in [0, 1].$$

Demostración: Consultese el Teorema 4.1.4 en Nelsen, *An Introduction to Copulas*. □

En la definición de *pseudo-inversa* (2.23) notemos que:

$$\begin{aligned} \varphi^{[-1]}(\varphi(t)) &= \begin{cases} \varphi^{-1}(\varphi(t)), & 0 \leq \varphi(t) \leq \varphi(0), \\ 0, & \varphi(0) \leq \varphi(t) \leq +\infty, \end{cases} \\ &= t, \end{aligned} \quad (2.24)$$

y también que:

$$\begin{aligned} \varphi(\varphi^{[-1]}(z)) &= \begin{cases} \varphi(\varphi^{-1}(z)), & 0 \leq z \leq \varphi(0), \\ \varphi(0), & \varphi(0) \leq z \leq +\infty, \end{cases} \\ &= \begin{cases} z, & 0 \leq z \leq \varphi(0), \\ \varphi(0), & \varphi(0) \leq z \leq +\infty, \end{cases} \\ &= \min\{z, \varphi(0)\}. \end{aligned} \quad (2.25)$$

A una función φ que cumple la hipótesis del Teorema 2.7 se le denomina *función generadora de cópula bivariada*. Si $\varphi(0) = +\infty$ entonces su pseudo-inversa es la inversa usual, es decir $\varphi^{[-1]} = \varphi^{-1}$, y en tal caso se dice que φ es un *generador estricto*. Pero si $\varphi(0)$ es un valor finito entonces $\varphi^{[-1]}(z)$ es estrictamente decreciente en el intervalo $[0, \varphi(0)]$ y luego es igual a la constante cero para todo $z \geq \varphi(0)$, y se le denomina *generador no estricto*.

El Teorema 2.7 facilita mucho la construcción de familias de cópulas, ya que resulta más sencillo proponer funciones generadoras de una sola variable que cumplan sus supuestos, que imaginar funciones bivariadas que cumplan la propiedad 2-creciente.

Ejemplo 10. Generador no estricto.

Una función tan sencilla como $\varphi(t) = (1-t)\mathbf{1}_{[0,1]}(t)$ cumple con ser un generador, no estricto en este caso porque $\varphi(0) = 1$ es finito, y su pseudo-inversa resulta ser $\varphi^{[-1]}(z) = (1-z)\mathbf{1}_{[0,1]}(z)$, con lo que se genera la siguiente cópula:

$$\begin{aligned} C(u, v) &= \varphi^{[-1]}((1-u) + (1-v)) = (u+v-1)\mathbf{1}_{\{u+v-1 \geq 0\}}, \\ &= \max\{u+v-1, 0\} = W(u, v), \end{aligned} \quad (2.26)$$

es decir la cópula W , que es además la cota inferior de Fréchet-Hoeffding, tiene generador no estricto que ahora denotaremos $\varphi_W(t) = (1-t)\mathbf{1}_{[0,1]}(t)$. Más aún, notemos que $\varphi_W(t)^\theta$ sigue cumpliendo la hipótesis del Teorema 2.7 para cualquier $\theta \geq 1$, permitiéndonos así definir una *familia paramétrica de cópulas* mediante la familia de generadores (no estrictos en este caso) $\{\varphi_\theta(t) = (1-t)^\theta\mathbf{1}_{[0,1]}(t) : \theta \geq 1\}$ que tienen pseudo-inversas $\varphi_\theta^{[-1]}(z) = (1-z^{1/\theta})\mathbf{1}_{[0,1]}(z)$, generando así la familia paramétrica de cópulas:

$$\begin{aligned} C_\theta(u, v) &= \varphi_\theta^{[-1]}((1-u)^\theta + (1-v)^\theta), \\ &= (1 - [(1-u)^\theta + (1-v)^\theta]^{1/\theta})\mathbf{1}_{\{0 \leq (1-u)^\theta + (1-v)^\theta \leq 1\}}, \\ &= \max\{1 - [(1-u)^\theta + (1-v)^\theta]^{1/\theta}, 0\}, \quad \theta \geq 1, \end{aligned} \quad (2.27)$$

familia que incluye a $W = C_1$, véase la Figura 2.5. ■

Ejemplo 11. Generador estricto.

De (2.20) tenemos que $\varphi(t) = (-\log t)\mathbf{1}_{[0,1]}(t)$ también cumple la hipótesis del Teorema 2.7 y es, de hecho, un generador estricto de la cópula Π ya que $\varphi(0) = +\infty$, al que podríamos parametrizar mediante $\varphi_\theta(t) = (-\log t)^\theta$ para valores de θ que resulten admisibles para cumplir con la hipótesis del teorema. De entrada $\theta \neq 0$ porque de lo contrario sería una función constante y violaría el ser estrictamente decreciente. Claramente es una función continua y derivable, por lo que la primera derivada nos determina para qué valores $\theta \neq 0$ la función es estrictamente decreciente, con $0 < t < 1$:

$$\frac{d}{dt}\varphi_\theta(t) < 0 \Leftrightarrow \theta(-\log t)^{\theta-1} \frac{-1}{t} < 0 \Leftrightarrow \theta > 0,$$

es decir, $\varphi_\theta(t)$ es estrictamente decreciente si y solo si $\theta > 0$. Para verificar convexidad requerimos la segunda derivada:

$$\frac{d^2}{dt^2}\varphi_\theta(t) \geq 0 \Leftrightarrow \frac{\theta(-\log t)^{\theta-2}}{t^2}(\theta - 1 - \log t) \geq 0 \Leftrightarrow \theta \geq 1,$$

esto es, $\varphi_\theta(t)$ es convexa si y solo si $\theta \geq 1$. Tenemos entonces ahora una familia de generadores $\{\varphi_\theta(t) = (-\log t)^\theta : \theta \geq 1\}$ que tienen pseudo-inversas (que en realidad son simplemente inversas por provenir de un generador estricto) $\varphi_\theta^{[-1]}(z) = \varphi_\theta^{-1}(z) = \exp(-z^{1/\theta})$, generando así la siguiente familia paramétrica de cópulas:

$$\begin{aligned} C_\theta(u, v) &= \varphi_\theta^{-1}((- \log u)^\theta + (- \log v)^\theta), \\ &= \exp\left(-[(- \log u)^\theta + (- \log v)^\theta]^{1/\theta}\right), \quad \theta \geq 1, \end{aligned} \quad (2.28)$$

familia que incluye a $C_1 = \Pi$, véase la Figura 2.6. ■

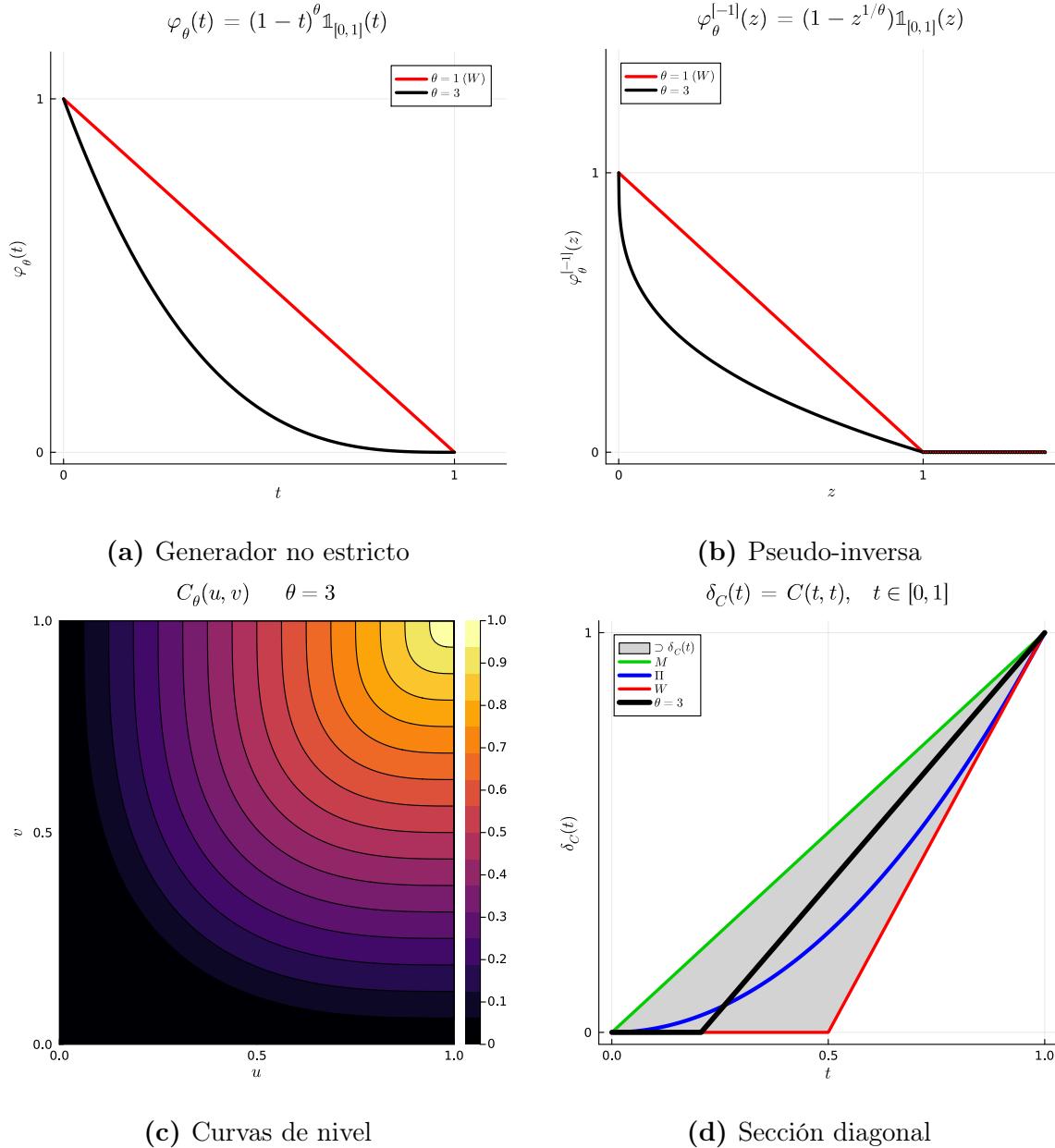


Figura 2.5: Ejemplo 10. Construcción de la cópula obtenida en (2.27). Para el generador no estricto y su pseudo-inversa se comparan los valores del parámetro $\theta = 1$ (que corresponde a W) versus $\theta = 3$. Las curvas de nivel solo corresponden al caso $\theta = 3$ y tendrían que compararse versus las curvas de nivel de W en la Figura 2.1a. La sección diagonal con $\theta = 3$ (negro) se compara versus las secciones diagonales de las cotas de Fréchet-Hoeffding (rojo y verde) y la cópula independencia (azul).

Una propiedad evidente de las cópulas bivariadas generadas mediante el Teorema 2.7 es que son *cópulas simétricas*, esto es que:

$$C_\varphi(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) = \varphi^{[-1]}(\varphi(v) + \varphi(u)) = C_\varphi(v, u).$$

La anterior junto con otras propiedades importantes de las cópulas bivariadas generadas mediante el Teorema 2.7 se resumen en el siguiente:

Teorema 2.8: Propiedades de cópulas arquimedianas

Sea C_φ una cópula bivariada generada por φ como en el Teorema 2.7. Entonces:

- a) C_φ es *simétrica*:

$$C_\varphi(u, v) = C_\varphi(v, u), \quad \forall u, v \in [0, 1].$$

- b) C_φ es *asociativa*:

$$C_\varphi(C_\varphi(u, v), w) = C_\varphi(u, C_\varphi(v, w)), \quad \forall u, v, w \in [0, 1].$$

- c) Para cualquier constante real $\beta > 0$ la función $\varphi_\beta(t) := \beta\varphi(t)$ cumple con la hipótesis del Teorema 2.7, y genera la misma cópula que φ , es decir:

$$C_{\varphi_\beta} = C_\varphi, \quad \forall \beta > 0.$$

- d) Las conjuntos de nivel $\mathcal{N}_{C_\varphi}(t) = \{(u, v) \in [0, 1]^2 : C_\varphi(u, v) = t\}$ para cualquier $0 < t < 1$ son siempre curvas convexas.

- e) Sección diagonal: $\delta_{C_\varphi}(t) < t, \forall 0 < t < 1$.

Demostración: Consúltese la sección 4.3 del libro de Nelsen, *An Introduction to Copulas*. □

Nótese que la propiedad e) del teorema anterior implica que para la cota superior de Fréchet-Hoeffding, la cópula $M(u, v) = \min\{u, v\}$, no existe un generador que pueda construirla, ya que $\delta_M(t) = t$ para todo $t \in [0, 1]$. Esto en contraste con la cópula W que tiene generador (no estricto) $\varphi_W(t) = (1 - t)\mathbf{1}_{[0,1]}(t)$, y la cópula Π que tiene generador (estricto) $\varphi_\Pi(t) = (-\log t)\mathbf{1}_{[0,1]}(t)$, según se dedujo en (2.27) y (2.28), respectivamente.

En la Tabla 4.1 del libro de Nelsen, *An Introduction to Copulas*, se presentan 22 familias paramétricas de cópulas con sus correspondientes generadores. Algunas incluyen a W y/o a Π como casos particulares, como sucedió en (2.27) y (2.28), y aunque no existe generador para la cópula M , resulta que ésta es *cópula límite* de algunas familias de cópulas, esto es que en algunos casos si $\theta \rightarrow +\infty$ entonces $C_\theta \rightarrow M$. Esto puede apreciarse, por ejemplo, de forma indirecta en la Figura 2.6d, notando que la curva negra de la diagonal continuará acercándose más y más a la cota superior conforme θ crece. Formalmente, es posible demostrar que conforme $\theta \rightarrow +\infty$ entonces el límite de C_θ en 2.28 degenera en $C_{+\infty} = M$.

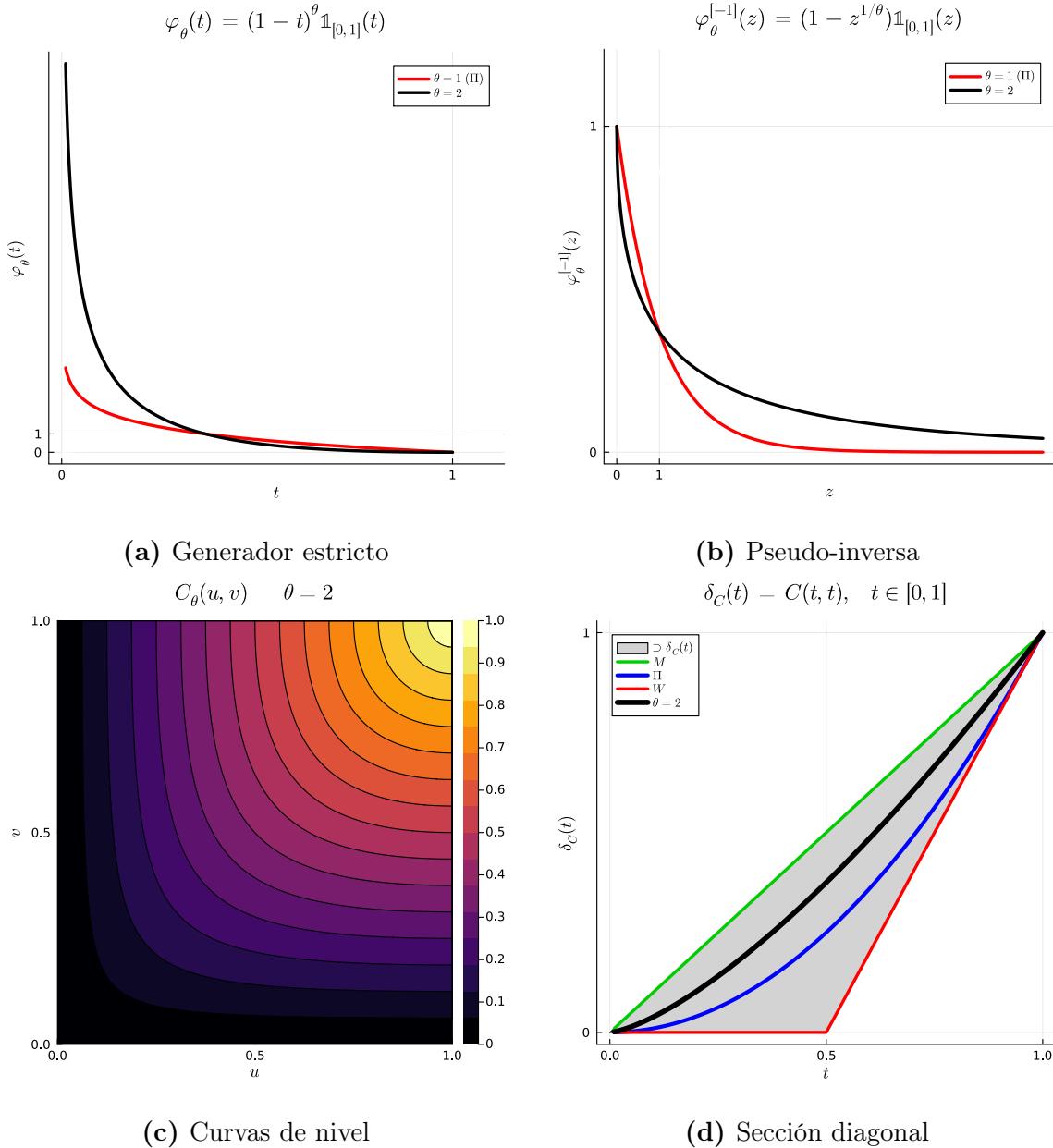


Figura 2.6: Ejemplo 11. Construcción de la cópula obtenida en (2.28). Para el generador estricto y su pseudo-inversa se comparan los valores del parámetro $\theta = 1$ (que corresponde a Π) versus $\theta = 2$. Las curvas de nivel solo corresponden al caso $\theta = 2$ y tendrían que compararse versus las curvas de nivel de Π en la Figura 2.1b. La sección diagonal con $\theta = 2$ (negro) se compara versus las secciones diagonales de las cotas de Fréchet-Hoeffding (rojo y verde) y la cópula independencia (azul).

A las cópulas bivariadas construidas mediante el Teorema 2.7 se les denomina *cópulas arquimedianas*, y para entender la razón de ello haremos las siguientes consideraciones. Al inicio de esta subsección se discutió que, en general, puede utilizarse cualquier cópula C para definir un operador binario cerrado sobre los elementos de $[0, 1]$, esto es definir $u \odot v := C(u, v)$ para cualesquiera $u, v \in [0, 1]$.

Si en particular C_φ es una cópula con generador φ como en el Teorema 2.7 entonces la estructura algebraica $([0, 1], \odot, <)$ tiene (entre otras) las siguientes propiedades,² como consecuencia de la definición de cópula 2.1 y del Teorema 2.8, para cualesquiera $u, v, w \in [0, 1]$:

- Comutatividad: $u \odot v = v \odot u$.
- Asociatividad: $(u \odot v) \odot w = u \odot (v \odot w)$.
- Elemento nulo: $u \odot 0 = 0$.
- Elemento neutro: $u \odot 1 = u$.
- Si $0 < u < 1$ entonces $u^2 := u \odot u < u$.

Nótese que el concepto de *potencia* es respecto al operador binario \odot definiendo como $u^1 := u$, y luego las potencias para $n \in \{2, 3, \dots\}$ se definen recursivamente: $u^n := u \odot u^{n-1}$. Supongamos que φ es un generador estricto. Entonces:

$$\begin{aligned} u^2 &= u \odot u = C_\varphi(u, u) = \varphi^{-1}(2\varphi(u)) \Rightarrow \varphi(u^2) = 2\varphi(u). \\ u^3 &= u \odot u^2 = C_\varphi(u, u^2) = \varphi^{-1}(\varphi(u) + \varphi(u^2)) = \varphi^{-1}(3\varphi(u)) \Rightarrow \varphi(u^3) = 3\varphi(u). \\ &\vdots \\ u^n &= \varphi^{-1}(n\varphi(u)) \Rightarrow \varphi(u^n) = n\varphi(u). \end{aligned}$$

Recordemos la *propiedad arquimediana de los números reales*:³ Para cualesquiera $a, b \in \mathbb{R}^+$ existe un entero positivo n tal que $na > b$, pero:

$$na > b \iff (e^{-a})^n < e^{-b},$$

en donde necesariamente $0 < e^{-a} < 1$ y $0 < e^{-b} < 1$, y la potencia n es respecto a la multiplicación usual de números reales. Entonces para cualesquiera $u, v \in]0, 1[$ tenemos que $\varphi(u), \varphi(v) \in \mathbb{R}^+$ por lo que existe un entero positivo n tal que $n\varphi(u) > v$, pero por ser φ estrictamente decreciente:

$$n\varphi(u) > v \Leftrightarrow \varphi(u^n) > \varphi(v) \Leftrightarrow \varphi^{-1}(\varphi(u^n)) < \varphi^{-1}(\varphi(v)) \Leftrightarrow u^n < v,$$

en donde se aplica ahora la potencia n pero respecto al operador binario \odot inducido por la cópula C_φ . Y a similar conclusión se puede llegar con generadores no estrictos. Entonces, como una cópula generada mediante el Teorema 2.7 induce un operador binario que cumple una propiedad arquimediana bajo dicho operador, es por ello que se les denomina *cópulas arquimedianas*.

²De hecho, $([0, 1], \odot)$ resulta ser lo que se conoce como un *semigrupo conmutativo*.

³Consúltese, por ejemplo, el Teorema I.30 de Tom M. Apostol (1967) *Calculus. Volume I*. Ed. John Wiley & Sons (Nueva York).

2.3. Mezclando, pegando y rotando

En las subsecciones 2.1 y 2.2 anteriores se abordaron algunas formas, entre muchas otras que existen, para deducir o construir distintas familias de cópulas bivariadas. Abordaremos a continuación algunas formas para combinar dos o más cópulas dadas, y crear una nueva.

Teorema 2.9: Combinación lineal convexa

Sean C_1, \dots, C_n cópulas bivariadas cualesquiera. Entonces cualquier combinación lineal convexa de ellas es también una cópula bivariada. Esto es, para cualesquiera números reales $\alpha_1, \dots, \alpha_n$ tales que $\alpha_j \geq 0$, y $\sum_{j=1}^n \alpha_j = 1$, la función $C : [0, 1]^2 \rightarrow [0, 1]$ definida por:

$$C(u, v) := \alpha_1 C_1(u, v) + \cdots + \alpha_n C_n(u, v)$$

es una cópula bivariada.

Demostración: Con $n = 2$ el resultado se sigue verificando que cumple la Definición 2.1, y luego se generaliza mediante inducción matemática. \square

Ejemplo 12. Combinación lineal convexa de W, Π, M .

Sean $0 \leq \alpha \leq 1$ y $0 \leq \beta \leq 1$ tales que $\alpha + \beta \leq 1$. Entonces podemos definir la siguiente combinación lineal convexa:

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

La familia paramétrica de cópulas $\{C_{\alpha,\beta} : \alpha \geq 0, \beta \geq 0, \alpha + \beta \leq 1\}$ incluye además, como casos particulares, las cópulas $C_{1,0} = W$, $C_{0,0} = \Pi$ y $C_{0,1} = M$. De lo anterior se deduce la subfamilia $\beta = 1 - \alpha$, esto es:

$$\begin{aligned} C_\alpha(u, v) &= \alpha W(u, v) + (1 - \alpha) M(u, v), \quad 0 \leq \alpha \leq 1, \\ &= \begin{cases} (1 - \alpha)u, & 0 \leq u \leq \frac{1}{2}, u \leq v \leq 1 - u, \\ (1 - \alpha)v, & 0 \leq u \leq 1, 0 \leq v \leq \min\{u, 1 - u\}, \\ u + \alpha v - \alpha, & 0 \leq u \leq 1, \max\{u, 1 - u\} \leq v \leq 1, \\ \alpha u + v - \alpha, & \frac{1}{2} \leq u \leq 1, 1 - u \leq v \leq u, \end{cases} \end{aligned}$$

que en el caso particular $\alpha = \frac{1}{2}$ representa el punto medio entre las superficies de W y M . En la Figura 2.7 podemos apreciar en las curvas de nivel y sección diagonal de $C_{1/2}$ el efecto de una combinación lineal convexa de W y M . ■

Las figuras 2.7c y 2.7d nos permiten apreciar que, mediante una combinación lineal convexa de cópulas, la cópula resultante de alguna forma mezcla características de ambas, se parece un poco a W pero también a M en ese caso. Veremos a continuación otra forma de combinar cópulas, “pegando” una junto a la otra, en lugar de mezclarlas.

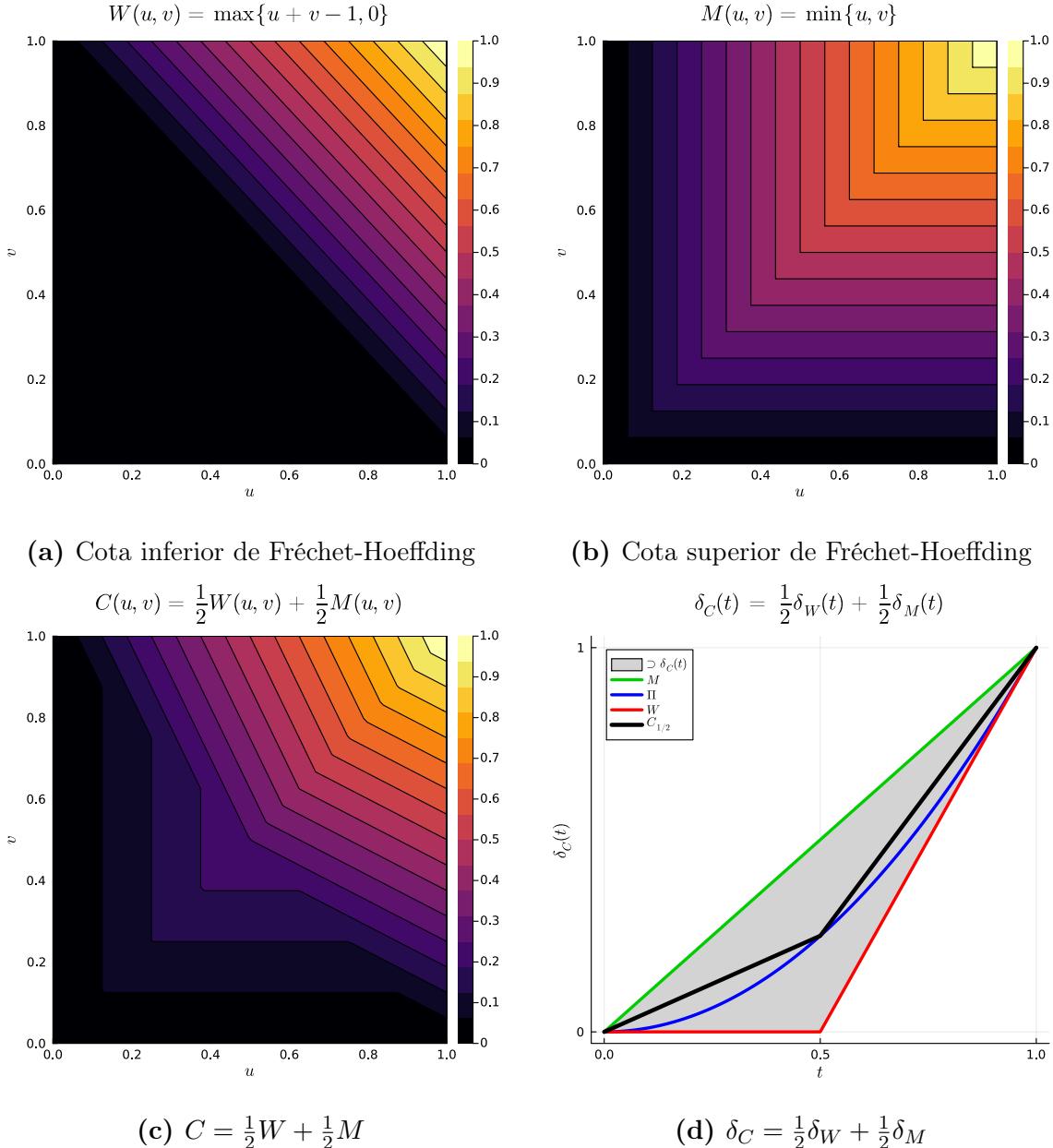


Figura 2.7: Ejemplo 12. Comparando las curvas de nivel de las cotas de Fréchet-Hoeffding W y M con las de la combinación lineal convexa $C = \frac{1}{2}W + \frac{1}{2}M$, así como sus secciones diagonales.

Teorema 2.10: Pegado de cópulas

Sean C_1 y C_2 dos cópulas bivariadas, y un número real $0 < \theta < 1$. Entonces la siguiente función es también una cópula bivariada:

$$C_{1,2,\theta}(u, v) := \begin{cases} \theta C_1\left(\frac{u}{\theta}, v\right), & 0 \leq u \leq \theta, \\ (1 - \theta)C_2\left(\frac{u-\theta}{1-\theta}, v\right) + \theta v, & \theta \leq u \leq 1, \end{cases}$$

que puede denotarse como $C_{1,2,\theta} \equiv C_1 \boxplus_\theta C_2$.

Demostración: Consultese Siburg y Stoimenov, «Gluing copulas».

□

Lo que sucede en la construcción anterior es que la cópula C_1 ha sido deformada desde su dominio original $[0, 1]^2$ al rectángulo $[0, \theta] \times [0, 1]$, C_2 deformada al rectángulo $[\theta, 1] \times [0, 1]$, y luego colocadas una junta a la otra, para completar todo el cuadrado unitario $[0, 1]^2$. El valor $0 < \theta < 1$ juega el papel de “punto de corte” o “punto de pegado” en $u = \theta$.

Esta manera de deformar y pegar cópulas no es, en general, commutativa, es decir, en general $C_1 \boxplus_\theta C_2$ no es necesariamente igual que $C_2 \boxplus_\theta C_1$. Otro aspecto interesante de esta técnica es que no necesariamente da por resultado una cópula simétrica, como sucede con las cópulas arquimedianas, aun pegando cópulas simétricas como se verá en un siguiente ejemplo. Adicionalmente, por inducción matemática, podemos extender el pegado de cópulas a 3 o más cópulas, por ejemplo mediante algo como $(C_1 \boxplus_\theta C_2) \boxplus_\alpha C_3$.

Ejemplo 13. Pegando W y M

Como $W(u, v) = \max\{u + v - 1, 0\}$ y $M(u, v) = \min\{u, v\}$ entonces aplicando el Teorema 2.10 obtenemos, por un lado:

$$(W \boxplus_\theta M)(u, v) = \begin{cases} 0, & 0 \leq u \leq \theta, 0 \leq v \leq 1 - \frac{u}{\theta}, \\ u + \theta v - \theta, & 0 \leq u \leq 1, \max\left\{1 - \frac{u}{\theta}, \frac{u-\theta}{1-\theta}\right\} \leq v \leq 1, \\ v, & \theta \leq u \leq 1, 0 \leq v \leq \frac{u-\theta}{1-\theta}, \end{cases} \quad (2.29)$$

y en orden inverso:

$$(M \boxplus_\theta W)(u, v) = \begin{cases} u, & 0 \leq u \leq \theta, \frac{u}{\theta} \leq v \leq 1, \\ \theta v, & 0 \leq u \leq 1, 0 \leq v \leq \min\left\{\frac{u}{\theta}, \frac{1-u}{1-\theta}\right\}, \\ u + v - 1, & \theta \leq u \leq 1, \frac{1-u}{1-\theta} \leq v \leq 1. \end{cases}$$

Notemos que $W \boxplus_\theta M \neq M \boxplus_\theta W$ y que ninguna de estas cópulas resultantes es simétrica. Veáse la Figura 2.8 para compararlas, y compárense tambien versus la combinación lineal convexa de la Figura 2.7. Nótese que en la sección diagonal de $W \boxplus_{0.5} M$ en la Figura 2.8c el comportamiento para $t \leq 0.5$ es como W , y para $t \geq 0.5$ es como M . Pero en la sección diagonal de $M \boxplus_{0.5} W$ en la Figura 2.8d el comportamiento para $t \leq 0.5$ es como M , y para $t \geq 0.5$ es como W .

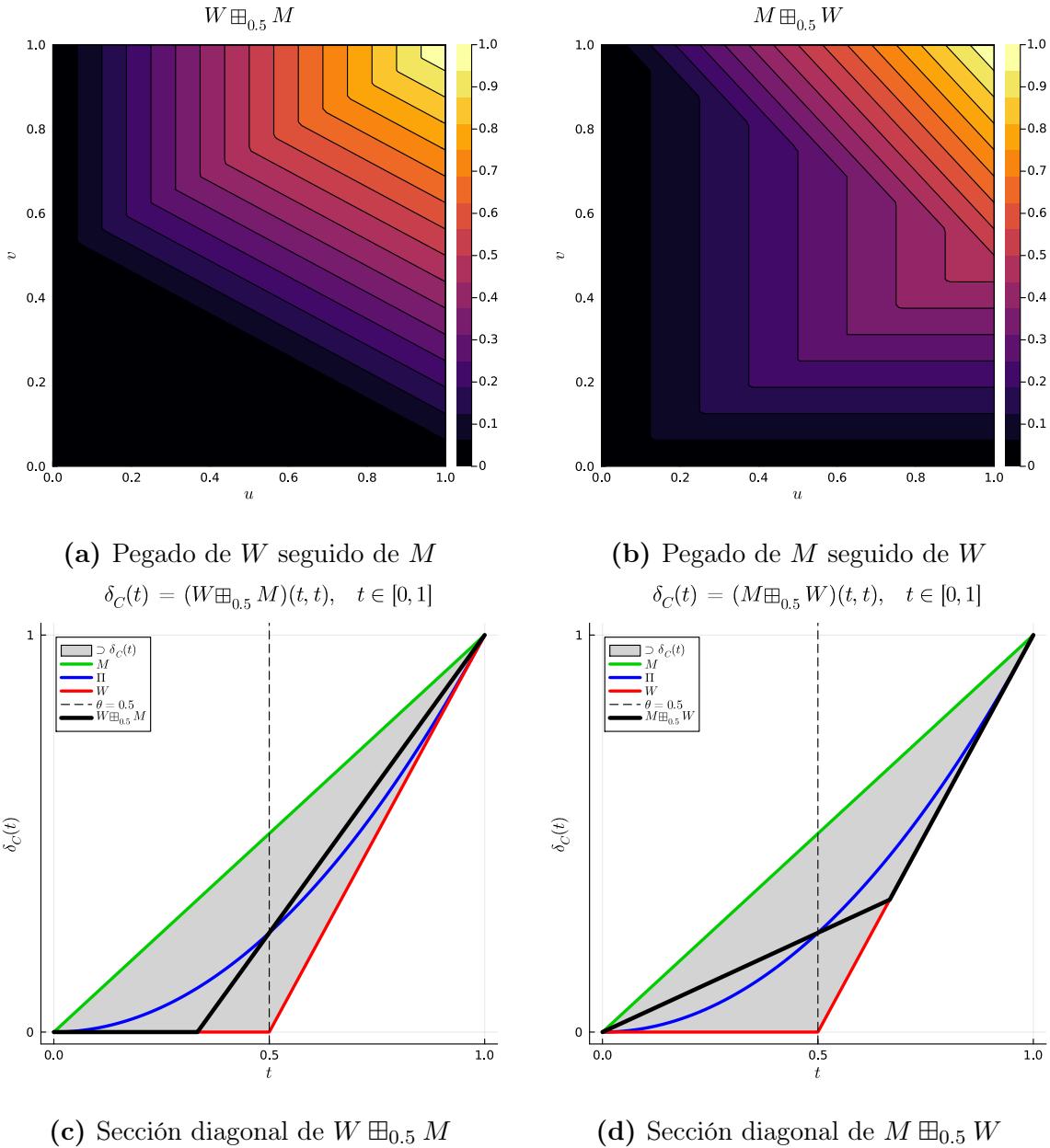


Figura 2.8: Ejemplo 13. Curvas de nivel y secciones diagonales de los pegados de cópulas $W \boxplus_{0.5} M$ y $M \boxplus_{0.5} W$.

Por medio del Teorema 1.9 es posible generar variaciones de una cópula dada. Veremos a continuación por qué se acompañó a dicho teorema con el título de *rotaciones*.

Definición 2.11: Rotaciones

Dada una cópula bivariada C denominamos *rotaciones* de C a las siguientes funciones:

- a) Vertical: $V(C)(u, v) := u - C(u, 1 - v)$.
- b) Horizontal: $H(C)(u, v) := v - C(1 - u, v)$.
- c) Diagonal: $D(C)(u, v) := u + v - 1 + C(1 - u, 1 - v)$.
- d) Simétrica: $S(C)(u, v) := C(v, u)$.

Proposición 2.12

Cualquiera de las rotaciones de una cópula bivariada como en la Definición 2.11 resulta ser también una cópula bivariada.

Demostración: Los casos vertical, horizontal y diagonal son inmediatos a partir del Teorema 1.9, y el caso simétrico es inmediato a partir de la Definición 2.1. \square

La proposición anterior implica que una cópula bivariada C podría transformarse en otra cópula mediante rotaciones, aunque en ocasiones, dependiendo de las características de C , algunas de las rotaciones podrían resultar iguales a C . Por ejemplo, si C es un cópula simétrica entonces obviamente $S(C) = C$. Pero en otros casos puede resultar distinta, por ejemplo $V(M) = W = H(M)$.

Corolario 2.13

Para cualquier cópula bivariada C se cumple lo siguiente:

- a) $D(C) = V(H(C)) = H(V(C))$.
- b) $S(V(C)) = H(S(C))$ y $S(H(C)) = V(S(C))$.
- c) $S(D(C)) = D(S(C))$.

Demostración: Consecuencia inmediata de aplicar la Definición 2.11 en cada caso. \square

Corolario 2.14

Si C es una cópula absolutamente continua (Definición 2.5a) con densidad

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v)$$

entonces su rotaciones también son cópulas absolutamente continuas con las densidades siguientes:

- a) Vertical: $V(c)(u, v) = c(u, 1 - v)$.
- b) Horizontal: $H(c)(u, v) = c(1 - u, v)$.
- c) Diagonal: $D(c)(u, v) = c(1 - u, 1 - v)$.
- d) Simétrica: $S(c)(u, v) = c(v, u)$.

Demostración: Basta calcular $\frac{\partial^2}{\partial u \partial v}$ en cada uno de los casos de la Definición 2.11. □

Ejemplo 14. Rotaciones de una cópula arquimediana.

Consideremos las funciones con parámetro $\theta > 0$ definidas mediante

$$\varphi_\theta(t) := \frac{1}{\theta}(t^{-\theta} - 1)\mathbf{1}_{[0,1]}(t). \quad (2.30)$$

Analizando su primera y segunda derivadas es inmediato verificar que cumple con las características de un generador estricto de cópula arquimediana (Teorema 2.7), que permite construir la familia paramétrica de cópulas $\{C_\theta : \theta > 0\}$ dada por:

$$C_\theta(u, v) = \varphi_\theta^{-1}(\varphi_\theta(u) + \varphi_\theta(v)) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad (2.31)$$

que además resulta ser absolutamente continua (Definición 2.5) con densidad

$$c_\theta(u, v) = \frac{\partial^2}{\partial u \partial v} C_\theta(u, v) = (1 + \theta) u^{-\theta-1} v^{-\theta-1} (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta-2}. \quad (2.32)$$

En la Figura 2.9 podemos apreciar las curvas de nivel de la densidad c_θ con $\theta = 2$, junto con las rotaciones vertical, diagonal y vertical. La rotación simétrica no es necesaria porque toda cópula arquimediana es simétrica (Teorema 2.8a) y por tanto $S(c_\theta) = c_\theta$. Observando la secuencia de las gráficas podemos apreciar cómo la superficie de la cópula original 2.9a va rotando por múltiplos de 90° alrededor del eje vertical $\{(0.5, 0.5, z) : z \in \mathbb{R}\} \subset \mathbb{R}^3$, o equivalentemente sus curvas de nivel van rotando en \mathbb{R}^2 alrededor del punto $(0.5, 0.5)$. ■

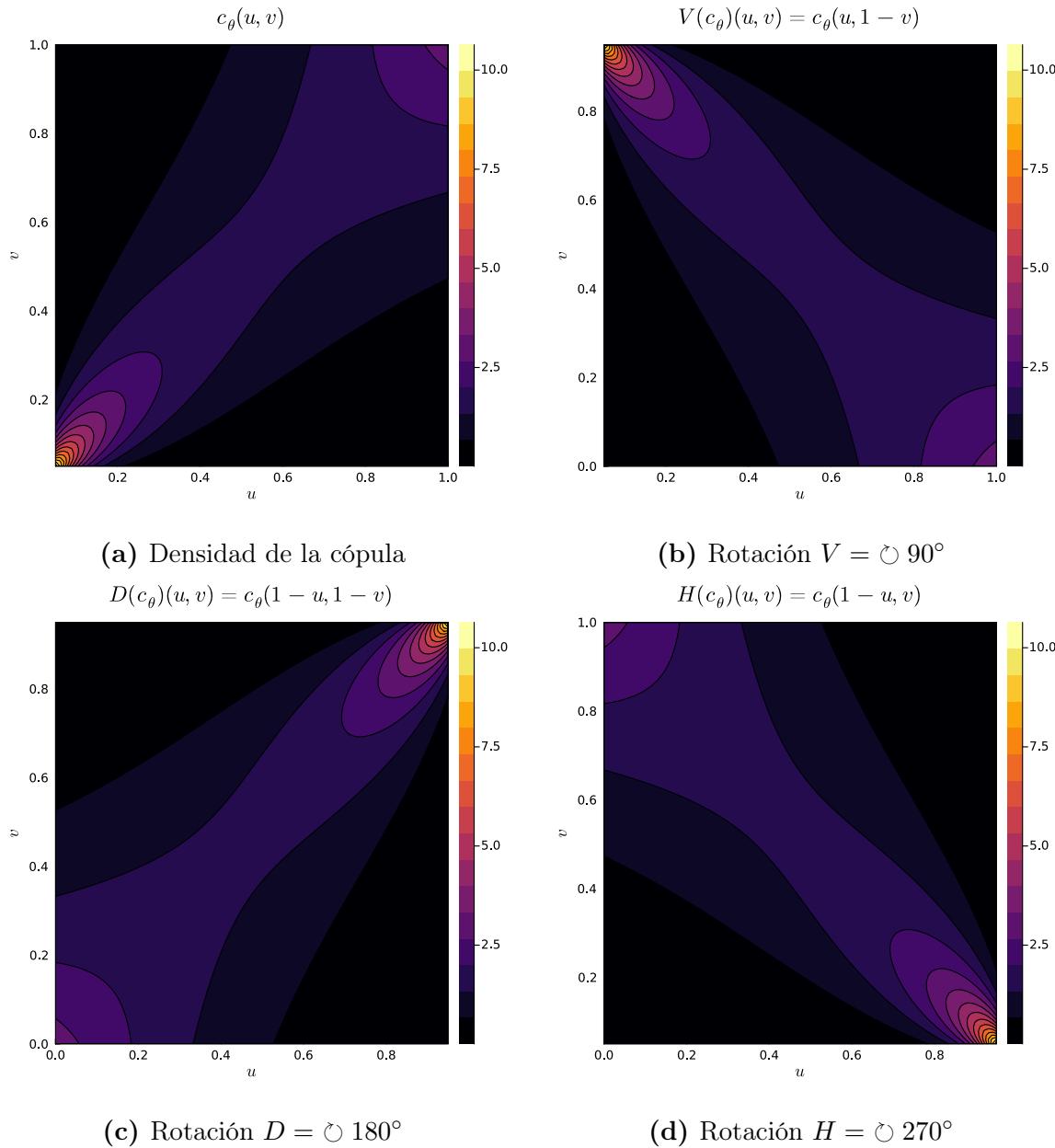


Figura 2.9: Ejemplo 14. Efecto de las rotaciones b)vertical, c)diagonal y d)horizontal, de la densidad de la cópula a) (2.32) con parámetro $\theta = 2$.

3. Simulación

Recordemos que el Corolario 1.7, véase (2.9), nos permite interpretar a una cópula bivariada como la función de distribución conjunta de un vector aleatorio (U, V) de variables aleatorias marginalmente continuas y uniformes sobre el intervalo $[0, 1]$ cada una, y que en tal caso:

$$C_{U,V}(u, v) = F_{U,V}(u, v), \quad \forall (u, v) \in [0, 1]^2.$$

Por lo tanto, simulaciones del vector aleatorio (U, V) equivalen, en este caso, a simulaciones de su cópula subyacente, mismas que pueden generarse aplicando el Algoritmo 3.4 analizado en el Capítulo 1. En particular requerimos una de las funciones de distribución condicional, aplicando (1.38), por ejemplo:

$$F_{V|U}(v | u) = \mathbb{P}(V \leq v | U = u) = \frac{1}{f_U(u)} \frac{\partial}{\partial u} F_{U,V}(u, v) = \frac{\partial}{\partial u} C_{U,V}(u, v), \quad (2.33)$$

y con lo anterior:

- Simula $u_1, \dots, u_n \sim \text{iid Uniforme}(0, 1)$.
- Simula $t_1, \dots, t_n \sim \text{iid Uniforme}(0, 1)$, independientes de los valores u_j .
- Para cada $j \in \{1, \dots, n\}$ calcular $v_j = F_{V|U}^{(-1)}(t_j | u_j)$.
- $(u_1, v_1), \dots, (u_n, v_n)$ son iid $\sim F_{U,V} = C_{U,V}$.

Recordemos también que, por el Teorema de la Transformada Inversa (Teorema 3.2 del Capítulo 1), si $U \sim \text{Uniforme}(0, 1)$ y F_X es la función de distribución de una variable aleatoria continua X , entonces la transformación $F_X^{(-1)}(U)$ tiene la misma distribución de probabilidad que X . Cuasi-inversas como $F_X^{(-1)}$ son siempre estrictamente crecientes (en el caso de variables aleatorias continuas, véase comentario posterior a la Definición 3.1 del Capítulo 1), entonces como consecuencia del Teorema 1.8 tenemos que:

$$C_{U,V} = C_{F_X^{(-1)}(U), F_Y^{(-1)}(V)} = C_{X,Y}. \quad (2.34)$$

Lo anterior fundamenta el siguiente algoritmo general para simular observaciones de un vector aleatorio bivariado de variables aleatorias continuas, con cópula subyacente y marginales dadas:

Algoritmo 3.1

Para simular n observaciones iid de un vector aleatorio (X, Y) de variables aleatorias continuas con cópula subyacente C y funciones de distribución marginal F_X y F_Y :

1. Simula $u_1, \dots, u_n \sim \text{iid Uniforme}(0, 1)$.
2. Define la función $\psi(v | u) := \frac{\partial}{\partial u} C(u, v)$, y deduce su cuasi-inversa $\psi^{(-1)}(t | u)$.
3. Simula $t_1, \dots, t_n \sim \text{iid Uniforme}(0, 1)$, independientes de los valores u_j .
4. Para cada $j \in \{1, \dots, n\}$ calcular $v_j = \psi^{(-1)}(t_j | u_j)$.
5. $(u_1, v_1), \dots, (u_n, v_n)$ son iid $\sim C$.
6. Para cada $j \in \{1, \dots, n\}$ calcula $x_j = F_X^{(-1)}(u_j)$, $y_j = F_Y^{(-1)}(v_j)$.
7. $(x_1, y_1), \dots, (x_n, y_n)$ son iid $\sim F_{X,Y} = C(F_X, F_Y)$.

Los pasos 1 al 5 permiten simular observaciones de una cópula cualquiera C , y ya en el paso 6 se transforman para que tengan las distribuciones marginales que se deseen, pero manteniendo como cópula subyacente a C .

Ejemplo 15. Simulando con cópula arquimediana.

Simularemos observaciones a partir de la cópula arquimediana del Ejemplo 14 (2.31):

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad \theta > 0,$$

a partir de la cual calculamos su derivada parcial:

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

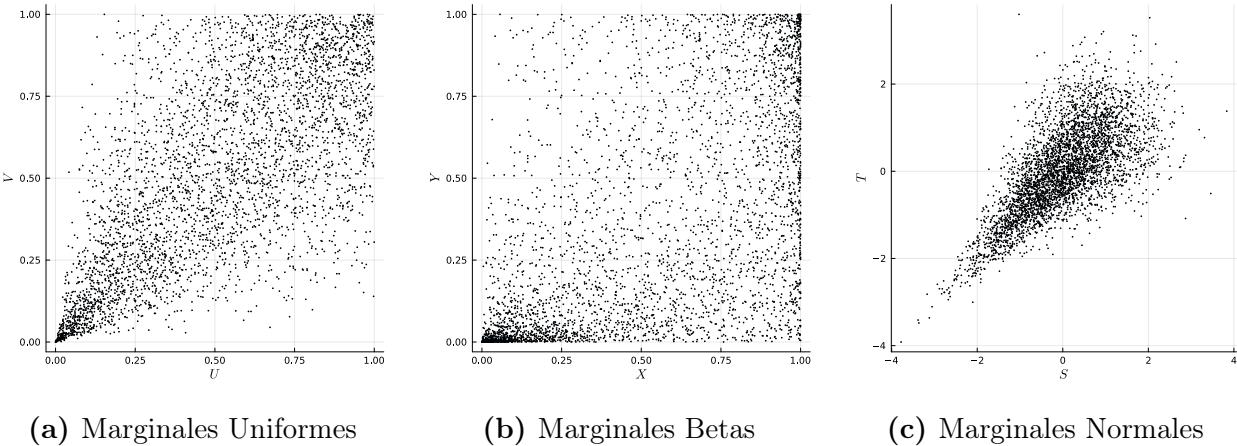
y luego su cuasi-inversa $\psi^{(-1)}$, que en este caso es la inversa usual, por ser ψ estrictamente creciente:

$$\psi_\theta^{-1}(t | u) = (1 + u^{-\theta}(t^{-\theta/(1+\theta)} - 1))^{-1/\theta},$$

con lo que ya podemos aplicar los pasos 1 al 5 del Algoritmo 3.1 para simular observaciones de la cópula C_θ , es decir, de un vector aleatorio (U, V) con marginales $\text{Uniforme}(0, 1)$ y cópula subyacente C_θ . Luego las aprovecharemos para simular observaciones del vector aleatorio (X, Y) con marginales $\text{Beta}(\frac{1}{2}, \frac{1}{2})$ y $\text{Beta}(\frac{1}{3}, \frac{2}{3})$ pero manteniendo la misma cópula subyacente C_θ , y lo análogo para un vector aleatorio (S, T) con ambas marginales $\text{Normal}(0, 1)$ pero igualmente con la misma cópula subyacente C_θ .

En la Figura 2.10 podemos apreciar tres gráficos de dispersión con $n = 4,000$ observaciones simuladas a partir de cada uno de los vectores aleatorios (U, V) , (X, Y) y (S, T) , los tres con cópula subyacente C_θ y $\theta = 2$. Aunque luzcan diferentes, la dependencia es la misma en los tres casos, solo que con distintas distribuciones marginales que no alteran la dependencia entre las variables.

En particular compárese el gráfico de dispersión de (U, V) de la Figura 2.10a versus la Figura 2.9a del Ejemplo 14 que representa la función de densidad c_θ , que al tener mucha densidad cerca del punto $(0, 0)$ claramente explica por qué la mayor concentración de puntos simulados se encuentra cerca de ahí, como era de esperarse.



(a) Marginales Uniformes

(b) Marginales Betas

(c) Marginales Normales

Figura 2.10: Ejemplo 15. Simulaciones de vectores aleatorios con la misma cópula subyacente (y por tanto la misma dependencia entre variables) pero con distribuciones marginales distintas.

Si ya se tiene implementada la forma de simular observaciones de una cópula dada C (pasos 1 al 5 del Algoritmo 3.1), éstas pueden transformarse en observaciones de sus rotaciones (ver Definición 2.11) como consecuencia inmediata del Teorema 1.9:

Teorema 3.2: Simular rotaciones de cópulas.

Sea (U, V) un vector aleatorio con cópula subyacente C y ambas marginales con distribución Uniforme(0, 1). Entonces los siguientes vectores aleatorios también tienen ambas marginales con distribución Uniforme(0, 1) pero con las siguientes cópulas subyacentes:

- a) $(U, 1 - V) \sim V(C)(u, v) = u - C(u, 1 - v).$
- b) $(1 - U, V) \sim H(C)(u, v) = v - C(1 - u, v).$
- c) $(1 - U, 1 - V) \sim D(C)(u, v) = u + v - 1 + C(1 - u, 1 - v).$
- d) $(V, U) \sim S(C)(u, v) = C(v, u).$

Demostración: Es inmediato verificar que si $U \sim \text{Uniforme}(0, 1)$ entonces también la transformación $1 - U \sim \text{Uniforme}(0, 1)$. El resultado se sigue de aplicar este hecho junto con el Teorema 1.9. \square

Lo anterior implica que, si ya tenemos simulaciones $(u_1, v_1), \dots, (u_n, v_n)$ de una cópula C ,

entonces, por ejemplo, $(1 - u_1, 1 - v_1), \dots, (1 - u_n, 1 - v_n)$ son simulaciones de la cópula $D(C)$, no es necesario (aunque sería correcto) aplicar directamente el Algoritmo 3.1 a $D(C)$.

Para simular observaciones de una combinación lineal convexa de cópulas (Teorema 2.9) partiendo de que se tiene implementada la forma de simular observaciones de las cópulas involucradas, se presenta y demuestra el caso de dos cópulas, pero el resultado se generaliza a cualquier cantidad finita de cópulas, mediante inducción matemática:

Teorema 3.3: Simular combinación lineal convexa de cópulas.

Sean (U_1, V_1) y (U_2, V_2) dos vectores aleatorios independientes con cópulas subyacentes C_1 y C_2 , respectivamente, y con marginales Uniforme(0, 1). Entonces, dado un valor $0 < \alpha < 1$, el vector aleatorio definido como:

$$(U, V) := \begin{cases} (U_1, V_1), & \text{con probabilidad } \alpha, \\ (U_2, V_2), & \text{con probabilidad } 1 - \alpha, \end{cases}$$

tiene por cópula subyacente a

$$C_{U,V} = \alpha C_1 + (1 - \alpha) C_2$$

y además $U \sim \text{Uniforme}(0, 1)$ y $V \sim \text{Uniforme}(0, 1)$.

Demostración: Sea B una variable aleatoria Bernoulli con parámetro α , independiente de los vectores aleatorios (U_1, V_1) y (U_2, V_2) . Entonces

$$(U, V) = B(U_1, V_1) + (1 - B)(U_2, V_2) = (BU_1 + (1 - B)U_2, BV_1 + (1 - B)V_2),$$

de donde es inmediato verificar que tiene marginales uniformes y función de distribución conjunta $F_{U,V} = C_{U,V} = \alpha C_1 + (1 - \alpha) C_2$. \square

El resultado anterior es el fundamento del siguiente algoritmo para simular observaciones de una combinación lineal convexa de dos cópulas (fácilmente generalizable a $k > 2$ cópulas):

Algoritmo 3.4

Dadas dos cópulas bivariadas C_1 y C_2 , y un valor $0 < \alpha < 1$, para simular n observaciones iid de un vector aleatorio (U, V) de variables aleatorias marginalmente Uniforme(0, 1) y con cópula subyacente igual a la **combinación lineal convexa**:

$$C_{U,V} = \alpha C_1 + (1 - \alpha) C_2$$

se procede como sigue:

1. Simula $b_1, \dots, b_n \sim Bernoulli(\alpha)$.
2. Simula $(u_{11}, v_{11}), \dots, (u_{1n}, v_{1n}) \sim C_1$.
3. Simula $(u_{21}, v_{21}), \dots, (u_{2n}, v_{2n}) \sim C_2$.
4. Para cada $j \in \{1, \dots, n\}$ calcula

$$(u_j, v_j) = (b_j u_{1j} + (1 - b_j) u_{2j}, b_j v_{1j} + (1 - b_j) v_{2j}).$$

5. $(u_1, v_1), \dots, (u_n, v_n) \sim \alpha C_1 + (1 - \alpha) C_2$.

Para simular observaciones de un pegado de cópulas (Teorema 2.10) partiendo de que se tiene implementada la forma de simular observaciones de las cópulas involucradas, se presenta y demuestra el caso de dos cópulas, pero el resultado se generaliza a cualquier cantidad finita de cópulas, mediante inducción matemática:

Teorema 3.5: Simular pegado de cópulas.

Sean (U_1, V_1) y (U_2, V_2) dos vectores aleatorios independientes con cópulas subyacentes C_1 y C_2 , respectivamente, y con marginales Uniforme(0, 1). Entonces, dado un valor $0 < \theta < 1$, el vector aleatorio definido como:

$$(U, V) := \begin{cases} (\theta U_1, V_1), & \text{con probabilidad } \theta, \\ (\theta + (1 - \theta)U_2, V_2), & \text{con probabilidad } 1 - \theta, \end{cases}$$

tiene por cópula subyacente a:

$$C_{U,V} = (C_1 \boxplus_\theta C_2)(u, v) = \begin{cases} \theta C_1\left(\frac{u}{\theta}, v\right), & 0 \leq u \leq \theta, \\ (1 - \theta)C_2\left(\frac{u-\theta}{1-\theta}, v\right) + \theta v, & \theta \leq u \leq 1, \end{cases}$$

y además $U \sim \text{Uniforme}(0, 1)$ y $V \sim \text{Uniforme}(0, 1)$.

Demostración: El procedimiento es análogo al Teorema 3.3, así que se deja como ejercicio. \square

El resultado anterior es el fundamento del siguiente algoritmo para simular observaciones del pegado de dos cópulas (fácilmente generalizable a $k > 2$ cópulas):

Algoritmo 3.6

Dadas dos cópulas bivariadas C_1 y C_2 , y un valor $0 < \theta < 1$, para simular n observaciones iid de un vector aleatorio (U, V) de variables aleatorias marginalmente Uniforme(0, 1) y con cópula subyacente igual al **pegado de cópulas**:

$$C_{U,V} = C_1 \boxplus_\theta C_2$$

se procede como sigue:

1. Simula $b_1, \dots, b_n \sim Bernoulli(\alpha)$.
2. Simula $(u_{11}, v_{11}), \dots, (u_{1n}, v_{1n}) \sim C_1$.
3. Simula $(u_{21}, v_{21}), \dots, (u_{2n}, v_{2n}) \sim C_2$.
4. Para cada $j \in \{1, \dots, n\}$ calcula

$$\begin{aligned} u_j &= b_j \theta u_{1j} + (1 - b_j)(\theta + (1 - \theta)u_{2j}), \\ v_j &= b_j v_{1j} + (1 - b_j)v_{2j}, \end{aligned}$$

5. $(u_1, v_1), \dots, (u_n, v_n) \sim C_1 \boxplus_\theta C_2$.

Ejemplo 16. Rotando, mezclando y pegando cópulas.

Aprovechando del Ejemplo 15 en donde se implementó la forma de simular observaciones de una cópula C_θ (2.31), simularemos ahora observaciones de las cópulas siguientes, aplicando los Algoritmos 3.4 y 3.6:

- $0.3D(C_\theta) + 0.7V(C_\theta)$ (combinación lineal convexa de cópulas rotadas).
- $D(C_\theta) \boxplus_{0.3} V(C_\theta)$ (pegado de cópulas rotadas).

En la Figura 2.11 se utilizó como valor del parámetro $\theta = 2$, y se simularon $n = 4,000$ observaciones en cada caso. En las Figuras 2.11a y 2.11b se simularon por separado las rotaciones de $\circlearrowleft 180^\circ$ y $\circlearrowleft 90^\circ$ de C_θ (véase Figura 2.10a). En la Figura 2.11c podemos apreciar el resultado de simular una combinación lineal convexa de dichas rotaciones, donde es evidente que la rotación 2.11a tiene menos peso (30 %) que la rotación 2.11b (70 %). En cambio, en la Figura 2.11d es claro el pegado de las cópulas rotadas en el punto de pegado $u = 0.3$.

El Algoritmo 3.1 es totalmente general para simular a partir de cualquier cópula C . Podría incluso aplicarse directamente a cualquier rotación, combinación lineal convexa o pegado de cópulas, aunque en esos casos no es estrictamente necesario, si ya se tiene implementado

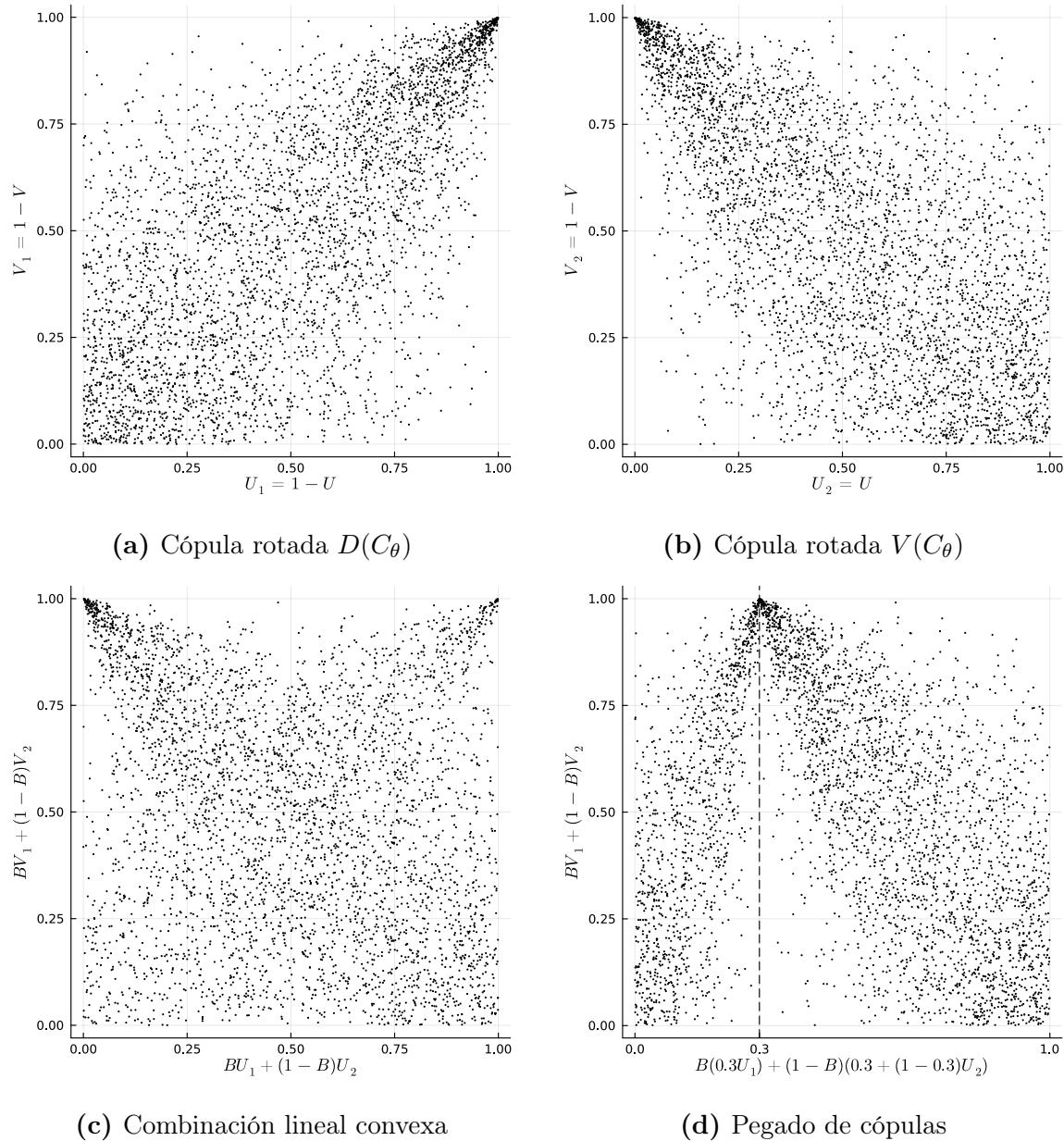


Figura 2.11: Ejemplo 16. **a)** Cúpula C_θ con rotación $\circlearrowleft 180^\circ$. **b)** Cúpula C_θ con rotación $\circlearrowleft 90^\circ$. **c)** Combinación lineal convexa de a) y b) con ponderación $\alpha = 0.3$ para a). **d)** Pegado de las cópulas a) y b) con $u = 0.3$ como punto de pegado.

alguna forma de simular a partir de las cópulas involucradas, como consecuencia de los Teoremas 3.2, 3.3 y 3.5, y los consecuentes Algoritmos 3.4 y 3.6.

En el caso particular de una cópula arquimediana C_φ con generador φ , si bien también es posible aplicar directamente el Algoritmo 3.1 (de hecho así se hizo en el Ejemplo 15), es posible derivar un algoritmo solo en términos de su generador. Por ejemplo, en el caso de un generador estricto, obtenemos:

$$\begin{aligned}\psi(v \mid u) &= \frac{\partial}{\partial u} C_\varphi(u, v) = \frac{\partial}{\partial u} \varphi^{-1}(\varphi(u) + \varphi(v)), \\ &= (\varphi^{-1})'(\varphi(u) + \varphi(v)) \varphi'(u), \\ &= \frac{\varphi'(u)}{\varphi'(\varphi^{-1}(\varphi(u) + \varphi(v)))}.\end{aligned}\tag{2.35}$$

Luego, para obtener la cuasi-inversa resolvemos para v , dado cualquier $0 < t < 1$:

$$\psi(v \mid u) = t \Leftrightarrow \varphi'(\varphi^{-1}(\varphi(u) + \varphi(v))) = \frac{\varphi'(u)}{t},\tag{2.36}$$

$$\Leftrightarrow \varphi^{-1}(\varphi(u) + \varphi(v)) = (\varphi')^{(-1)}\left(\frac{\varphi'(u)}{t}\right),\tag{2.37}$$

$$\begin{aligned}&\Leftrightarrow \varphi(v) = \varphi\left((\varphi')^{(-1)}\left(\frac{\varphi'(u)}{t}\right)\right) - \varphi(u), \\ &\Leftrightarrow v = \varphi^{-1}\left[\varphi\left((\varphi')^{(-1)}\left(\frac{\varphi'(u)}{t}\right)\right) - \varphi(u)\right],\end{aligned}\tag{2.38}$$

notando que del paso (2.36) al (2.37) se tiene que aplicar la cuasi-inversa de φ' , porque la convexidad de φ implica que $\varphi'' \geq 0$, lo que nos dice que φ' es monótona creciente, pero no necesariamente estrictamente creciente. Lo anterior fue para el caso de un generador estricto, pero el caso no estricto es similar, y lo único que cambia en (2.38) es la pseudo-inversa $\varphi^{[-1]}$ en lugar de φ^{-1} . Entonces, el Algoritmo 3.1 para el caso específico de cópulas arquimedianas es el mismo, pero utilizando en el Paso 4 la fórmula:

$$\psi^{(-1)}(t \mid u) = \varphi^{[-1]}\left[\varphi\left((\varphi')^{(-1)}\left(\frac{\varphi'(u)}{t}\right)\right) - \varphi(u)\right].\tag{2.39}$$

Si para una distribución de probabilidad específica de un vector aleatorio (X, Y) de variables aleatorias continuas ya se cuenta con un algoritmo para simular observaciones de dicho vector, mediante una sencilla transformación puede aprovecharse para convertirlas en observaciones de la cópula subyacente, sin necesidad de deducirla para aplicar el Algoritmo 3.1.

Recordemos el concepto de *cuasi-inversa* (ver Capítulo 1, Definición 3.1) de una función de distribución de probabilidad continua $F : \mathbb{R} \rightarrow [0, 1]$, que es cualquier función $F^{(-1)} : \text{Ran } F \rightarrow \mathbb{R}$ donde $F^{(-1)}(t)$ es cualquier $x \in \mathbb{R}$ tal que $F(x) = t$, esto es tal que:

$$F(F^{(-1)}(t)) = t.\tag{2.40}$$

Si F es estrictamente creciente entonces la cuasi-inversa es única y coincide con la inversa usual, es decir $F^{(-1)} = F^{-1}$. Pero si F no es estrictamente creciente (y por tanto solo es monótona creciente) entonces existe una colección numerable intervalos \mathfrak{I} tal que F es constante sobre cada intervalo, esto es:

$$\text{Si } I \in \mathfrak{I} \implies \forall x_1, x_2 \in I : F(x_1) = F(x_2)$$

y en tal caso la cuasi-inversa no es única ya que para cualquier en $x_0 \in I \in \mathfrak{I}$ tendríamos que $F^{(-1)}(F(x_0))$ podría ser cualquier $x \in I$.

Teorema 3.7: Transformación uniforme

Sea X una variable aleatoria continua con función de distribución de probabilidades F_X . Entonces la variable aleatoria $U := F_X(X)$ es continua Uniforme(0, 1).

Demostración: Como $\text{Ran } F_X \subseteq [0, 1]$ entonces $\text{Ran } U \subseteq [0, 1]$ y por tanto $F_U(u) = 0$ para todo valor $u \leq 0$, y $F_U(u) = 1$ para todo $u \geq 1$. Para el caso $0 < u < 1$ primero supongamos que F_X es estrictamente creciente. Entonces existe su inversa F_X^{-1} , misma que es también estrictamente creciente, y por tanto:

$$\begin{aligned} F_U(u) &= \mathbb{P}(U \leq u) = \mathbb{P}(F_X(X) \leq u), \\ &= \mathbb{P}(F_X^{-1}(F_X(X)) \leq F_X^{-1}(u)) = \mathbb{P}(X \leq F_X^{-1}(u)), \\ &= F_X(F_X^{-1}(u)) = u, \quad 0 < u < 1, \end{aligned}$$

pero $F_U(u) = u\mathbf{1}_{\{0 < u < 1\}} + \mathbf{1}_{\{u \geq 1\}}$ es justamente la función de distribución de una variable aleatoria continua Uniforme(0, 1). En el caso de que F_X no sea estrictamente creciente (pero es necesariamente monótona creciente) el argumento es esencialmente el mismo, utilizando alguna cuasi-inversa $F_X^{(-1)}$ en lugar de F_X^{-1} , pero notando que sobre el conjunto $\mathbb{R} \setminus \bigcup \mathfrak{I}$ tenemos que F_X sí es estrictamente creciente y por tanto para cualquier x en dicho conjunto tenemos que $F_X^{(-1)}(F_X(x)) = F_X^{-1}(F_X(x)) = x$. En cualquier otro caso tenemos que el conjunto $\{F_X(x) : x \in \bigcup \mathfrak{I}\}$ es numerable, por ser \mathfrak{I} numerable, y por tanto un conjunto de probabilidad cero para la variable aleatoria continua $F_X(X)$, es decir, con probabilidad 1 la variable aleatoria $F_X(X)$ tiene distribución de probabilidad Uniforme(0, 1). \square

Tenemos entonces que la transformación $F_X(X)$ de una variable aleatoria continua X es estrictamente creciente con probabilidad 1 ya que, de ser el caso, el conjunto de intervalos \mathfrak{I} en donde F_X sea constante, es un conjunto de probabilidad cero bajo la variable aleatoria X , véase (1.37).

Teorema 3.8: Transformación uniforme bivariada

Sea (X, Y) un vector de variables aleatorias continuas, con función de distribución conjunta $F_{X,Y}$ y marginales F_X y F_Y . Entonces la cópula subyacente de (X, Y) es la misma que la del vector aleatorio $(U, V) := (F_X(X), F_Y(Y))$ de marginales Uniforme(0, 1), es decir:

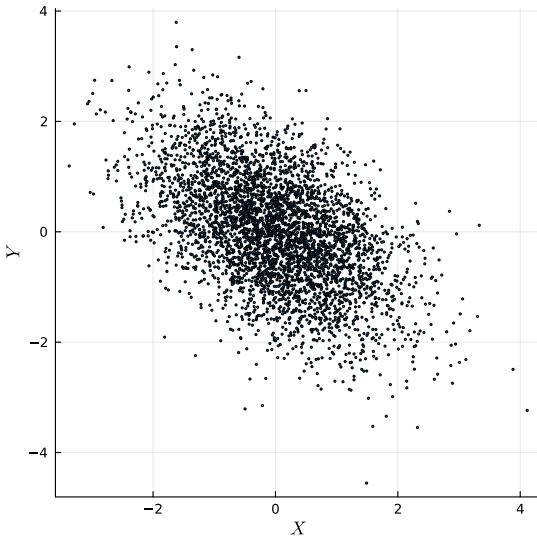
$$C_{X,Y} = C_{F_X(X), F_Y(Y)} = C_{U,V}.$$

Demostración: Consecuencia inmediata del Teorema 1.8 y del Teorema 3.7. □

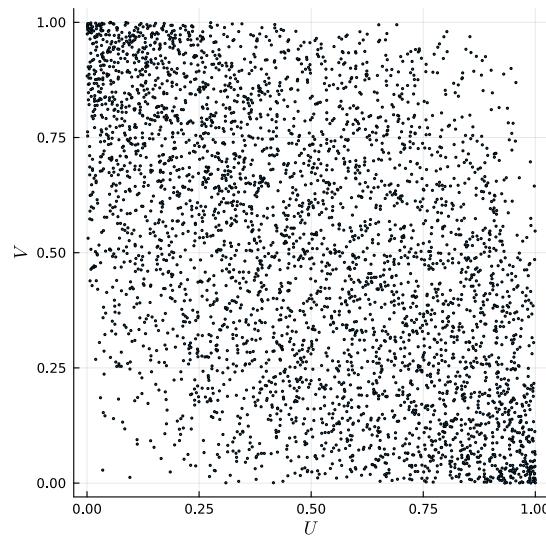
Si ya tenemos un algoritmo para simular n observaciones de un vector (X, Y) de variables aleatorias continuas, digamos $\{(x_1, y_1), \dots, (x_n, y_n)\}$, entonces tendríamos que $\{(u_j, v_j) : u_j = F_X(x_j), v_j = F_Y(y_j), j \in \{1, \dots, n\}\}$ sería un conjunto de observaciones de la cópula subyacente $C_{X,Y}$, como consecuencia del Teorema 3.8 y del Corolario 1.7 (2.9).

Ejemplo 17. Cúpula Normal (o Gaussiana).

Sea (X, Y) un vector aleatorio con distribución Normal bivariada estándar con parámetro de correlación $r_{X,Y} = -0.5$, lo que implica que marginalmente $X \sim \text{Normal}(0, 1)$ e $Y \sim \text{Normal}(0, 1)$. En la Figura 2.12a se muestran $n = 4,000$ pares de observaciones simuladas de (X, Y) , y en la Figura 2.12b se muestran las observaciones de la cúpula subyacente (conocida como cúpula Normal o Gaussiana) aplicando el Teorema 3.8 a las observaciones de (X, Y) .



(a) Marginales Normales



(b) Marginales Uniformes

Figura 2.12: Ejemplo 17. Simulaciones de vectores aleatorios con la misma cúpula subyacente (gaussiana en este caso), pero con marginales **a**) Normales $(0, 1)$ y **b**) Uniformes $(0, 1)$.

4. Estimación

En la sección anterior se abordaron algunas formas de simular observaciones independientes a partir de una cúpula bivariada dada. En esta sección se aborda el problema de dar reversa a dicho proceso, esto es, dado un conjunto de observaciones independientes de una misma cúpula bivariada ¿cómo podemos estimar la cúpula de la cual provienen? Desde un enfoque estadístico, esto puede hacerse de dos formas, dependiendo de la información con que se cuente:

- **Estimación paramétrica:** Se conoce (o se cree conocer) la familia paramétrica de cópulas $\{C_\theta : \theta \in \Theta\}$ de la cual provienen las observaciones, y con base en estas últimas se modeliza probabilísticamente la incertidumbre sobre el parámetro θ , que tiene algún espacio paramétrico Θ (conjunto de valores admisibles de θ).
- **Estimación no paramétrica:** Cuando no se tiene información alguna sobre la cópula de la cual provienen las observaciones, o aun teniéndola, se prefiere que sea solo la información de lo observado lo que se utilice para estimar la cópula.

4.1. Estimación no paramétrica

Cuando nos referimos a una *muestra aleatoria*⁴ observada $D = \{(u_j, v_j) : j \in \{1, \dots, n\}\}$ proveniente de una cópula bivariada C , esto es equivalente a decir que son observaciones iid de un vector aleatorio (U, V) de variables aleatorias marginalmente continuas con distribución Uniforme(0, 1) cada una, cuya función de distribución conjunta $F_{U,V} : \overline{\mathbb{R}}^2 \rightarrow [0, 1]$ restringida en su dominio al cuadrado unitario $[0, 1]^2$ coincide con la fórmula de la cópula subyacente $C_{U,V} = C$, como consecuencia del Corolario 1.7 (2.9), eso es:

$$C(u, v) = C_{U,V}(u, v) = F_{U,V}(u, v), \quad \forall (u, v) \in [0, 1]^2, \quad (2.41)$$

por lo que estimar la cópula desconocida C equivale a estimar a $F_{U,V}$ sobre $[0, 1]^2$, con base en las observaciones $D = \{(u_j, v_j) : j \in \{1, \dots, n\}\}$ de n vectores aleatorios $(U_1, V_1), \dots, (U_n, V_n)$ iid $\sim F_{U,V}$. Para cualquier $(u, v) \in [0, 1]^2$ definamos las variables aleatorias indicadoras:

$$Z_j(u, v) := \mathbf{1}_{\{U_j \leq u, V_j \leq v\}}, \quad j \in \{1, \dots, n\}, \quad (2.42)$$

que solo reportan los valores 0 o 1, lo que implica que son variables aleatorias *Bernoulli* con parámetro igual a su esperanza:

$$\mathbb{E}[Z_j(u, v)] = \mathbb{E}[\mathbf{1}_{\{U_j \leq u, V_j \leq v\}}] = \mathbb{P}(U_j \leq u, V_j \leq v) = F_{U,V}(u, v) = C(u, v). \quad (2.43)$$

Tenemos entonces n variables aleatorias $Z_j(u, v)$ iid Bernoulli con parámetro desconocido $C(u, v)$, por lo que su suma resulta ser una variable aleatoria *Binomial*:

$$S_n := \sum_{j=1}^n Z_j(u, v) \sim \text{Binomial}(n, C(u, v)), \quad (2.44)$$

y por lo tanto su esperanza y varianza son:

$$\mathbb{E}(S_n) = nC(u, v), \quad (2.45)$$

$$\mathbb{V}(S_n) = nC(u, v)[1 - C(u, v)], \quad (2.46)$$

a partir de lo cual es inmediato deducir que:

⁴En la ciencia Estadística, esta expresión se refiere a un conjunto de observaciones independientes e idénticamente distribuidas (iid).

$$\mathbb{E}\left(\frac{1}{n}S_n\right) = C(u, v), \quad (2.47)$$

$$\mathbb{V}\left(\frac{1}{n}S_n\right) = \frac{C(u, v)[1 - C(u, v)]}{n} \xrightarrow{n \rightarrow +\infty} 0. \quad (2.48)$$

Estadísticamente (2.47) significa que la variable aleatoria $\frac{1}{n}S_n$ es un *estimador puntual insesgado* de $C(u, v)$, mientras que (2.48) significa que además es un *estimador consistente* de $C(u, v)$. Dicho de otra forma, conforme n sea más grande la variable aleatoria reportará un valor que, en promedio, cada vez se aleje menos del verdadero valor de $C(u, v)$. Por su importancia denotaremos como:

$$\widehat{C}_n(u, v) := \frac{1}{n}S_n = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{U_j \leq u, V_j \leq v\}}, \quad \forall (u, v) \in [0, 1]^2, \quad (2.49)$$

a la variable aleatoria que resulta ser un estimador puntual insesgado y consistente de la cópula desconocida $C(u, v)$. Una vez que se conocen los valores que reportan los vectores aleatorios $(U_1, V_1), \dots, (U_n, V_n)$, esto es $D = \{(u_1, v_1), \dots, (u_n, v_n)\}$, al valor observado de la variable aleatoria $\widehat{C}_n(u, v)$ lo denotamos y expresamos como:

$$C_n(u, v) := \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{u_j \leq u, v_j \leq v\}}, \quad \forall (u, v) \in [0, 1]^2, \quad (2.50)$$

y que es simplemente la proporción de observaciones (u_j, v_j) en D que cumplen simultáneamente con $u_j \leq u$ y $v_j \leq v$ para un par dado $(u, v) \in [0, 1]^2$. A C_n la llamaremos *estimación empírica* de la cópula C . Algunos autores la llaman “cópula empírica” pero (2.50) no cumple con la Definición 2.1 de cópula. Recordemos que las cópulas son funciones uniformemente continuas, como consecuencia de la propiedad 2-creciente, y C_n como en (2.50) tiene discontinuidades, con saltos de tamaño $\frac{1}{n}$ ya que:

$$\forall (u_j, v_j) \in D : C_n(u_j, v_j) - \lim_{u \rightarrow u_j^-} C_n(u, v_j) = \frac{1}{n} = C_n(u_j, v_j) - \lim_{v \rightarrow v_j^-} C_n(u_j, v).$$

Sin embargo, es posible utilizar la información de C_n para construir una función $\tilde{C}_n^m : [0, 1]^2 \rightarrow [0, 1]$ que sea continua sobre todo su dominio, por medio de *polinomios de Bernstein*, utilizando los valores de C_n sobre el subconjunto $\{0, \frac{1}{m}, \dots, \frac{m-1}{m}, 1\}^2$, con $m \geq 2$:

$$\tilde{C}_n^m(u, v) := \sum_{i=0}^m \sum_{j=0}^m C_n\left(\frac{i}{m}, \frac{j}{m}\right) \binom{m}{i} u^i (1-u)^{m-i} \binom{m}{j} v^j (1-v)^{m-j} \quad (2.51)$$

que de hecho resulta ser una cópula bivariada, conocida como *Cópula Bernstein*, véase Sancetta y Satchell, «[The Bernstein copula and its applications to modeling and approximations of multivariate distributions](#)», y constituye una estimación no paramétrica de la cópula de la cual provienen las observaciones $D = \{(u_1, v_1), \dots, (u_n, v_n)\}$, a la que puede aplicarse, por ejemplo, el Algoritmo 3.1 para simular nuevas observaciones, o bien utilizarse para analizar sus curvas de nivel, o sus secciones verticales, horizontales y diagonales.

Ejemplo 18. Cúpula Bernstein.

Simulamos $n = 4,000$ observaciones iid a partir de la cúpula (2.31) con rotación $\odot 90^\circ$, esto es $V(C_\theta)$, con parámetro $\theta = 1$, ver Figura 2.13a. En la Figura 2.13b tenemos las curvas de nivel de $V(C_\theta)$, mismas que, con base en las simulaciones de la Figura 2.13a, se estiman mediante C_n (2.50), ver Figura 2.13c, y luego se aproximan mediante \tilde{C}_n^m (2.51) con $m = 100$, ver Figura 2.13d.

En el caso que tengamos observaciones iid $\{(x_1, y_1), \dots, (x_n, y_n)\}$ de un vector aleatorio (X, Y) con función de distribución desconocida $F_{X,Y}$, pero marginales conocidas F_X y F_Y , esto es que solamente la cúpula subyacente $C_{X,Y}$ es desconocida, entonces como consecuencia del Teorema 3.8 tenemos que:

$$D := \{(u_j, v_j) : u_j = F_X(x_j), v_j = F_Y(y_j), j \in \{1, \dots, n\}\} \quad (2.52)$$

es el equivalente a observaciones de la cúpula desconocida $C_{X,Y}$ y, por tanto, las estimaciones no paramétricas (2.50) y (2.51) son directamente aplicables.

En el caso que tengamos observaciones iid $\{(x_1, y_1), \dots, (x_n, y_n)\}$ de un vector aleatorio (X, Y) con función de distribución desconocida $F_{X,Y}$, y también marginales desconocidas F_X y F_Y , no es posible aplicar directamente el Teorema 3.8 para obtener observaciones de la cúpula subyacente desconocida $C_{X,Y}$. En este escenario primero necesitamos estimar las marginales, que en el caso no paramétrico puede hacerse mediante el estimador conocido como *función de distribución empírica*. Si X_1, \dots, X_n es una muestra aleatoria (es decir, variables aleatorias iid) de una función de distribución de probabilidades desconocida F , entonces para cualquier $x \in \bar{\mathbb{R}}$ la variable aleatoria:

$$\hat{F}_n(x) := \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{X_j \leq x\}} \quad (2.53)$$

es un estimador puntual insesgado y consistente de $F(x)$, para todo $x \in \bar{\mathbb{R}}$. La manera de verificarlo es totalmente análoga a la justificación para (2.49), notando que $n\hat{F}_n(x)$ es una variable aleatoria discreta Binomial($n, F(x)$). Una vez que se conocen los valores que reportan las variables aleatorias X_1, \dots, X_n , esto es la muestra aleatoria observada $\{x_1, \dots, x_n\}$, obtenemos $F_n : \bar{\mathbb{R}} \rightarrow [0, 1]$, la *función de distribución empírica observada*:

$$F_n(x) := \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{x_j \leq x\}}, \quad (2.54)$$

que cumple con las propiedades de cualquier función de distribución de probabilidades. Más todavía, al ser una función escalonada (constante por intervalos), F_n corresponde a una función de distribución de una variable aleatoria discreta, aún cuando F sea continua. Es decir F_n es una estimación discreta no paramétrica de F , sin importar si F es continua o no. Nótese también que:

$$\sum_{j=1}^n \mathbf{1}_{\{x_j \leq x\}} = \sum_{j=1}^n \mathbf{1}_{\{x_{(j)} \leq x\}} \quad (2.55)$$

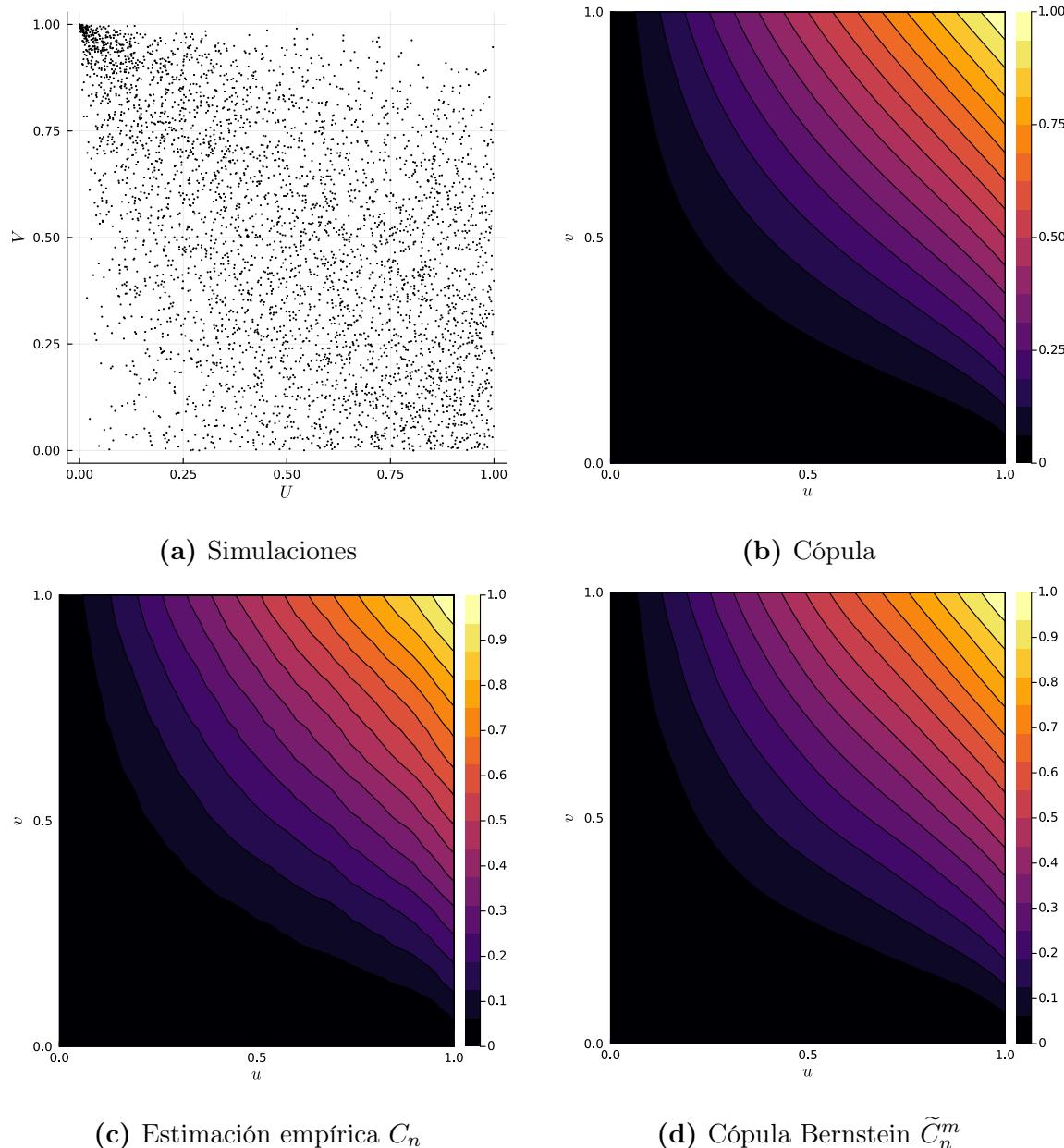


Figura 2.13: Ejemplo 18. **a)** $n = 4,000$ simulaciones de la cópula paramétrica $V(C_\theta)$, con $\theta = 1$. **b)** Curvas de nivel de la cópula paramétrica $V(C_\theta)$. **c)** Estimación empírica de $V(C_\theta)$. **d)** Cúpula Bernstein con $m = 100$, como estimación no paramétrica de $V(C_\theta)$.

donde $x_{(j)}$ es la j -ésima estadística de orden observada, esto es que si se ordena la muestra observada x_1, \dots, x_n de menor a mayor, entonces el valor que queda en la posición j se denota $x_{(j)}$, de modo que:

$$x_{(1)} < x_{(2)} < \dots < x_{(n)},$$

en donde las desigualdades son estrictas porque asumimos que F es continua, ya que no puede haber valores repetidos en tal caso. Entonces (2.54) es equivalente a:

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{x_{(j)} \leq x\}}, \quad \forall x \in \overline{\mathbb{R}} \quad (2.56)$$

y en particular notemos que:

$$F_n(x_{(k)}) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{x_{(j)} \leq x_{(k)}\}} = \frac{k}{n}, \quad \forall k \in \{1, \dots, n\}, \quad (2.57)$$

y que de hecho:

$$F_n(x) = \frac{k}{n}, \quad \forall x \in [x_{(k)}, x_{(k+1)}[, \quad (2.58)$$

lo que significa que F_n va dando saltos siempre de tamaño $\frac{1}{n}$ (cuando las observaciones provienen de una función de distribución continua F), y por lo tanto:

$$\text{Ran } F_n = \{F_n(x) : x \in \overline{\mathbb{R}}\} = \left\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right\}. \quad (2.59)$$

Entonces $\text{Ran } F_n$ es una partición finita uniforme del intervalo $[0, 1]$, lo que abre la puerta a aproximar por polinomios de Bernstein a la cuasi-inversa de F , esto es a $F^{(-1)} : [0, 1] \rightarrow \overline{\mathbb{R}}$ como lo proponen, por ejemplo, Muñoz-Pérez y Fernández-Palacín, «Estimating the quantile function by Bernstein polynomials», utilizando los puntos medios de los escalones de F_n :

$$\tilde{F}_n^{(-1)}(u) := \sum_{j=0}^n \frac{x_{(j)} + x_{(j+1)}}{2} \binom{n}{j} u^j (1-u)^{n-j}, \quad 0 \leq u \leq 1, \quad (2.60)$$

con la convención de elegir como $x_{(0)}$ algún valor adecuado menor o igual que el mínimo muestral observado $x_{(1)}$ (típicamente $x_{(0)} := x_{(1)}$), y como $x_{(n+1)}$ algún valor mayor o igual que el máximo muestral observado $x_{(n)}$ (típicamente $x_{(n+1)} := x_{(n)}$). Sería entonces la inversa de (2.60) una aproximación continua de F , que denotaremos simplemente \tilde{F}_n y que tendrá que aproximarse numéricamente porque usualmente la inversa de un polinomio no tiene una fórmula explícita.

Volviendo al caso en que tengamos una muestra aleatoria observada $\{(x_1, y_1), \dots, (x_n, y_n)\}$ de un vector aleatorio (X, Y) con función de distribución desconocida $F_{X,Y}$, y también marginales desconocidas F_X y F_Y , por medio de (2.60) aproximamos a F_X con \tilde{F}_n y a F_Y con \tilde{G}_n , y transformamos la muestra observada en:

$$\tilde{D} := \{(u_j, v_j) : u_j = \tilde{F}_n(x_j), v_j = \tilde{G}_n(y_j), j \in \{1, \dots, n\}\} \quad (2.61)$$

que constituye una aproximación a observaciones de la cópula subyacente desconocida $C_{X,Y}$, y que se denominan *pseudo-observaciones* de la cópula, y sería con ellas en este caso que se procede a la inferencia estadística, paramétrica o no paramétrica, de la cópula subyacente $C_{X,Y}$.

4.2. Estimación paramétrica

En el caso de que se tenga una muestra aleatoria de un vector aleatorio absolutamente continuo (X, Y) con función de densidad conjunta $f_{X,Y}$, entonces por el Corolario 1.6, véase (2.7), y por la Definición 2.5a, existe una cópula subyacente $C_{X,Y}$ absolutamente continua con densidad $c_{X,Y}(u, v) = \frac{\partial^2}{\partial u \partial v} C_{X,Y}(u, v)$ tal que:

$$f_{X,Y}(x, y) = c_{X,Y}(F_X(x), F_Y(y)) f_X(x) f_Y(y). \quad (2.62)$$

En caso de que las distribuciones marginales y la densidad de la cópula subyacente pertenezcan a alguna familia paramétrica, por ejemplo:

$$\{c_{X,Y}(u, v | \theta) : \theta \in \Theta\}, \quad \{f_X(x | \gamma) : \gamma \in \Gamma\}, \quad \{f_Y(y | \lambda) : \lambda \in \Lambda\}, \quad (2.63)$$

entonces la densidad conjunta $f_{X,Y}$ hereda todos los parámetros, esto es:

$$f_{X,Y}(x, y | \gamma, \lambda, \theta) = c_{X,Y}(F_X(x | \gamma), F_Y(y | \lambda) | \theta) f_X(x | \gamma) f_Y(y | \lambda). \quad (2.64)$$

Dada una muestra aleatoria observada $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ podemos obtener la *función de verosimilitud*:

$$L(\gamma, \lambda, \theta | D) = \prod_{j=1}^n f_{X,Y}(x_j, y_j | \gamma, \lambda, \theta), \quad (2.65)$$

y a partir de ella aplicar los procedimientos usuales de inferencia estadística paramétrica para realizar inferencias sobre el vector de parámetros (γ, δ, θ) , ya sea bajo un enfoque clásico o uno bayesiano. Si solo se desea hacer inferencia paramétrica del parámetro θ de la densidad de la cópula subyacente, se utilizan *observaciones* o bien *pseudo-observaciones* de la cópula subyacente, según sea el caso como se abordó en la Subsección 4.1 anterior, digamos $D = \{(u_1, v_1), \dots, (u_n, v_n)\}$, y entonces la función de verosimilitud correspondiente sería:

$$L(\theta | D) = \prod_{j=1}^n c_{X,Y}(u_j, v_j | \theta), \quad (2.66)$$

y a partir de ella se aplican los procedimientos usuales de inferencia estadística paramétrica, ya sea bajo un enfoque clásico o uno bayesiano.

Lo anterior es válido solo para el caso de cópulas absolutamente continuas, es decir para las que existe la densidad de la cópula (véase Definición 2.5a). Pero si la cópula es singular (Definición 2.5b) o bien tiene tanto componente absolutamente continuo como singular (Definición 2.5c) entonces es necesario proceder de otra manera. Partimos ahora de una muestra aleatoria observada $D = \{(u_1, v_1), \dots, (u_n, v_n)\}$ de observaciones o pseudo-observaciones de una cópula C_θ con parámetro desconocido $\theta \in \Theta$.

En tal caso puede realizarse una *estimación puntual* de θ aplicando el *método de mínimos cuadrados ordinarios*, que en este caso se traduciría en encontrar el valor $\hat{\theta} \in \Theta$ que minimice la función:

$$h(\theta | D) := \sum_{j=1}^n [C_\theta(u_j, v_j) - \tilde{C}_n^m(u_j, v_j)]^2, \quad \theta \in \Theta, \quad (2.67)$$

en donde \tilde{C}_n^m es como en (2.51), aunque en su lugar también podría utilizarse (2.50).

Si se desea realizar inferencias sobre θ más allá de una estimación puntual (estimación por intervalo, contraste de hipótesis o predicción, por ejemplo) entonces se puede aplicar un enfoque bayesiano, por algún *método ABC* (*Approximate Bayesian Computation*)⁵ que en este caso se traduciría en un algoritmo como el siguiente:

Algoritmo 4.1

Método ABC. Sea $D = \{(u_1, v_1), \dots, (u_n, v_n)\}$ un conjunto de observaciones o pseudo-observaciones de una cópula C_θ con parámetro desconocido $\theta \in \Theta$.

1. Simula $m \geq 100,000$ observaciones $\{\theta_1^*, \dots, \theta_m^*\}$ a partir de una *distribución de probabilidad a priori* de θ con soporte en algún subconjunto de Θ que se considere apropiado.
2. Con cada θ_j^* simula una muestra $D_j^* = \{(u_{1j}^*, v_{1j}^*), \dots, (u_{nj}^*, v_{nj}^*)\}$ a partir de $C_{\theta_j^*}$.
3. Opcionalmente (recomendable) a cada muestra simulada D_j^* aplica alguna transformación $\mathcal{T}(D_j^*)$ (en inglés: *summary statistic*) que resuma adecuadamente la información muestral, por ejemplo algún estadístico suficiente.
4. Para cada $j \in \{1, \dots, m\}$ calcula:

$$\varepsilon_j := \delta(\mathcal{T}(D_j^*), \mathcal{T}(D)),$$

donde δ es alguna medida de distancia entre sus argumentos.

5. Elige un número $k \ll m$ (es decir, k significativamente menor que m , por ejemplo $k = 300$) e identifica el conjunto:

$$\{j \in \{1, \dots, m\} : \varepsilon_j \leq \varepsilon_{(k)}\} = \{j_1, \dots, j_k\},$$

esto es identificar aquellas simulaciones D_j^* que generaron los k valores ε_j más pequeños.

6. Los valores simulados $\Theta^* := \{\theta_{j_1}^*, \dots, \theta_{j_k}^*\}$ son (aproximadamente) observaciones de una *distribución de probabilidad a posteriori* de θ dada la muestra osbervada D , y con ellos se pueden realizar inferencias sobre θ .

⁵Véase, por ejemplo: Sisson, Fan y Beaumont, *Handbook of Approximate Bayesian Computation*.

Ejemplo 19. Método ABC.

Utilizaremos las $n = 4,000$ simulaciones $D = \{(u_1, v_1), \dots, (u_n, v_n)\}$ de la cópula en el Ejemplo 18, que sabemos se realizaron con parámetro $\theta = 1$, pero a modo de experimento asumiremos que no conocemos dicho valor, y realizaremos estimación puntual y por intervalo de dicho parámetro. Es el caso de la cópula (2.31) pero con rotación $\circlearrowleft 90^\circ$, esto es:

$$C_\theta(u, v) = u - (u^{-\theta} + (1-v)^{-\theta} - 1)^{-1/\theta}, \quad \theta > 0. \quad (2.68)$$

Aplicando el Corolario 2.14a y a partir de (2.32) la densidad de la cópula correspondiente es:

$$c_\theta(u, v) = (1+\theta)[u(1-v)]^{-(1+\theta)}(u^{-\theta} + (1-v)^{-\theta} - 1)^{-(1/\theta+2)}, \quad \theta > 0.$$

La estimación de máxima verosimilitud de θ se obtiene encontrando el valor $\theta^* \in \Theta$ que maximice la función de verosimilud (2.66), que es lo mismo que maximizar la función de log-verosomilitud, con la ventaja de tener mejor control en la precisión numérica, ya que en lugar de un producto tenemos una suma:

$$\begin{aligned} \log L(\theta | D) &= \sum_{j=1}^n \log c_\theta(u_j, v_j), \quad \theta > 0, \quad \text{donde:} \\ \log c_\theta(u, v) &= \log(1+\theta) + (1+\theta)(\log u + \log(1-v)) - \left(2 + \frac{1}{\theta}\right) \log(u^{-\theta} + (1-v)^{-\theta} - 1). \end{aligned}$$

En la Figura 2.14a tenemos la gráfica de $\log L(\theta | D)$ para algunos valores de θ y que se maximiza en $\theta^* = 1.009$ (aproximadamente). Utilizando (2.67) con C_n en lugar de \tilde{C}_n^m en la Figura 2.14b tenemos la gráfica de $h(\theta | D)$ para algunos valores θ y que se minimiza en $\theta^* = 0.945$ (aproximadamente).

Finalmente, aplicando el Algoritmo 4.1, en la Figura 2.14c se presenta un histograma de simulaciones a posteriori de θ , aplicando el método ABC con $k = 300$, $m = 100000$, utilizando como *summary statistic* a C_n como en (2.50) evaluada en las observaciones de la cópula, y como medida de distancia δ una suma de diferencias cuadráticas, esto es:

$$\begin{aligned} \mathcal{T}(D) &:= \{C_n(u_i, v_i) : i \in \{1, \dots, n\}\}, \\ \delta(\mathcal{T}(D_j^*), \mathcal{T}(D)) &:= \sum_{i=1}^n [C_n(u_{ij}^*, v_{ij}^*) - C_n(u_i, v_i)]^2. \end{aligned}$$

En tal caso podemos utilizar, por ejemplo, la mediana a posteriori como estimación puntual, que en este caso resultó $\theta^* = 0.991$, pero además mediante los cuantiles a posteriori 2,5% y 97,5% obtenemos una estimación por intervalo de probabilidad 95% para θ : [0.867, 1.134].

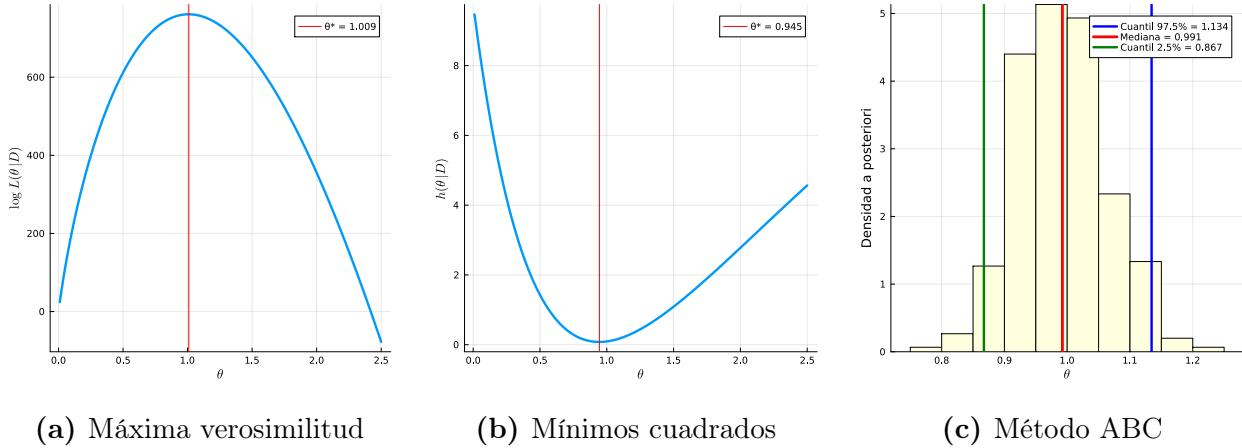


Figura 2.14: Ejemplo 19. a) Estimación puntual por máxima verosimilitud del parámetro θ de la familia paramétrica de cópulas (2.68). b) Similar al inciso a) anterior pero por el método de mínimos cuadrados. c) Histograma de simulaciones a posteriori de θ , junto con estimación puntual (mediana) y por intervalo de probabilidad 95 %.

5. Actividades

- 5.1) Demuestra los incisos c) y d) del Teorema 1.4.
- 5.2) Demuestra el Teorema 2.3. *Se recomienda consultar Nelsen, An Introduction to Copulas.*
- 5.3) Considera la distribución de probabilidad *Normal Bivariada* con vector de parámetros $(\mu_1, \mu_2, \sigma_1, \sigma_2, r)$ donde $\mu_j \in \mathbb{R}$, $\sigma_j \in \mathbb{R}^+$ y $-1 < r < 1$. Este modelo de probabilidad corresponde a un vector aleatorio absolutamente continuo (X, Y) con distribuciones marginales $X \sim \text{Normal}(\mu_1, \sigma_1)$ e $Y \sim \text{Normal}(\mu_2, \sigma_2)$, y donde $-1 < r < 1$ es un parámetro asociado a la dependencia entre dichas variables. Construye una función de distribución conjunta H que tenga la misma cópula subyacente que (X, Y) pero con marginales distintas: Beta($\frac{1}{2}, \frac{1}{3}$) y Pareto(3, 4). Compara gráficamente las curvas de nivel de la función de densidad conjunta de (X, Y) versus la densidad conjunta que se deduce a partir de H , y versus la densidad de la cópula subyacente.
- 5.4) Simula y grafica una muestra aleatoria bivariada de tamaño $n = 5,000$ a partir de la familia de cópulas, que es combinación lineal convexa de las cópulas W , Π y M , ver Ejemplo 12, con $\alpha = 0.6$ y $\beta = 0.3$.
- 5.5) Construye, simula y estima a partir de las siguiente familia paramétrica de cópulas:
 - a) Construye la familia paramétrica:

$$C_\theta(u, v) := \Pi \boxplus_\theta W, \quad 0 < \theta < 1,$$
 y grafica sus curvas de nivel para el caso particular $\theta = 0.6$.

- b) Simula y grafica una muestra tamaño $n = 5,000$ de observaciones de la cópula del inciso a) anterior con parámetro $\theta = 0.6$.
- c) Con las observaciones simuladas del inciso b) anterior calcula la cópula Bernstein (2.51), grafica sus curvas de nivel y compáralas versus las curvas de nivel de la cópula $C_{0.6}$ del inciso a).
- d) Supongamos que no conocemos el verdadero valor de θ , y con las observaciones simuladas del inciso b) estima puntualmente θ por el método de mínimos cuadrados, y luego calcula una estimación puntual y por intervalo a posteriori para θ aplicando un método ABC.

CAPÍTULO 3

MEDIR Y VISUALIZAR DEPENDENCIA

1. Dependencia de Schweizer-Wolff 98
2. Correlación de Spearman . . . 103
3. Diagramas de dependencia . . 107
4. Actividades 112

El análisis de la dependencia entre variables aleatorias no se limita a establecer su existencia, sino que también requiere cuantificarla y representarla de manera clara. En este capítulo se presentan diferentes enfoques para medir el grado de dependencia entre variables aleatorias continuas, destacando medidas como la *dependencia de Schweizer-Wolff* y la *correlación de Spearman*, ambas basadas solamente en la cópula subyacente, lo que garantiza que la cuantificación se centre exclusivamente en la estructura de dependencia y no en las distribuciones marginales. Asimismo, se aborda la importancia de la visualización mediante *diagramas de dependencia*, que permiten interpretar patrones de relación de forma general, sin la distorsión que introducen distribuciones de probabilidad marginales de las variables involucradas.

1. Dependencia de Schweizer-Wolff

Para el caso particular de un vector aleatorio (X, Y) de variables aleatorias continuas, del *Teorema de Sklar* (ver Teorema 1.1 del Capítulo 2) se desprende que toda la información sobre la dependencia entre X e Y está en su cópula subyacente $C_{X,Y}$ y, por lo tanto, cualquier propuesta para medir dicha dependencia debe estar basada explícita o implícitamente en dicha cópula, y solo en ella, porque las funciones de distribución marginal F_X y F_Y solo explican el comportamiento individual de cada variable aleatoria.

Como consecuencia del Corolario 1.2 del Capítulo 2, véase (2.3), tenemos que en el caso de independencia:

$$X \perp Y \iff C_{X,Y} = \Pi, \quad (3.1)$$

donde $\Pi(u, v) = uv$. Entonces, sea cual sea la forma en que se pretenda cuantificar dependencia entre las variables continuas X e Y , está claro que la ausencia total de dependencia (es decir, independencia) queda representada por la cópula Π , y en el caso de que $C_{X,Y} = \Pi$ la dependencia entre las variables debería cuantificarse con el valor mínimo posible, dentro de la escala que se decida utilizar.

Medir o cuantificar implica, en general, utilizar alguna función que asocie a cada objeto que se desea medir un elemento de un conjunto que represente la cuantificación que le corresponde, donde dicho conjunto es habitualmente (aunque no necesariamente) algún subconjunto adecuado de los números reales, que denominaremos *escala*: $\mathbb{D} \subseteq \mathbb{R}$.

Denotemos mediante \mathfrak{X}^2 al conjunto de todos los vectores aleatorios bivariados de variables aleatorias continuas. Pretendemos definir una función $\mu : \mathfrak{X}^2 \rightarrow \mathbb{D}$ que a cada $(X, Y) \in \mathfrak{X}^2$ le asocie un número $\mu((X, Y)) \in \mathbb{D}$ que cuantifique, de alguna manera, el grado de dependencia entre X e Y , que mejor denotaremos como $\mu_{X,Y}$ para una notación más simplificada, o equivalentemente μ_C donde C es la cópula subyacente del vector aleatorio en cuestión.

Si queremos una escala para la cuantificación de la dependencia que tenga un valor mínimo y uno máximo, podríamos elegir a \mathbb{D} como un intervalo cerrado y acotado en \mathbb{R} . El valor mínimo en \mathbb{D} tendría que estar reservado para la mínima dependencia posible, o ausencia total de dependencia (independencia). Si se elige al cero como valor mínimo en la escala \mathbb{D} entonces necesariamente $\mu_\Pi = 0$, esto es:

$$X \perp Y \iff \mu_{X,Y} = 0. \quad (3.2)$$

En caso de que $C_{X,Y} \neq \Pi$, la cuantificación de la dependencia estaría en relación directa con qué tanto se aleja (en algún sentido que habría que especificar) la cópula subyacente $C_{X,Y}$ de la cópula que representa la independencia, y si ya decidimos que el valor mínimo sea $\mu_\Pi = 0$ entonces en el caso de dependencia tendríamos que asignar algún número real $\mu_{X,Y} > 0$. Esto implicaría que la escala sea un intervalo de la forma $\mathbb{D} = [0, b] \subset \mathbb{R}$, en donde el número real $b > 0$ representaría la máxima dependencia posible.

Recordemos las cotas de Fréchet-Hoeffding (2.5), que al ser también cópulas, representan justamente lo más alejado que cualquier cópula puede estar de aquella que representa la independencia:

$$W(u, v) \leq \Pi(u, v) \leq M(u, v), \quad (3.3)$$

y por ello en la definición de μ_C para cualquier cópula subyacente C debería cumplirse que:

$$b = \max \{ \text{dist}\{C, M\}, \text{dist}\{C, W\} \}, \quad (3.4)$$

en donde $\text{dist}\{C_1, C_2\}$ representaría alguna forma de medir distancia entre dos cópulas cualesquiera C_1 y C_2 , que al tratarse de dos superficies continuas y acotadas en el cubo unitario $[0, 1]^3$, es usual hacerlo mediante el volumen total encerrado entre dichas superficies, esto es:

$$\text{dist}\{C_1, C_2\} := \int_0^1 \int_0^1 |C_1(u, v) - C_2(u, v)| dudv, \quad (3.5)$$

lo cual garantiza que $\text{dist}\{C_1, C_2\} = 0$ si y solo si $C_1 = C_2$, y en particular para el caso que nos interesa:

$$\text{dist}\{C, \Pi\} = 0 \iff X \perp Y. \quad (3.6)$$

Resolviendo las integrales correspondientes, es inmediato verificar que:

$$\text{dist}\{\Pi, M\} = \frac{1}{12} = \text{dist}\{\Pi, W\}, \quad (3.7)$$

por lo que, utilizando (3.5) como forma de medir distancias entre superficies, la escala para medir dependencia sería $\mathbb{D} = [0, \frac{1}{12}]$. Por simple comodidad, resulta más práctico utilizar una escala como $[0, 1]$, y esto fácilmente se logra multiplicando por 12 en (3.5), y así definir como **medida de dependencia** para un vector aleatorio (X, Y) de variables aleatorias continuas, con cópula subyacente $C_{X,Y}$, a la distancia (multiplicada por doce) de ésta respecto a Π , esto es:

$$\mu_{X,Y} := 12 \text{dist}\{C_{X,Y}, \Pi\} = 12 \int_0^1 \int_0^1 |C_{X,Y}(u, v) - uv| dudv, \quad (3.8)$$

a la que se le conoce como *medida de dependencia de Schweizer-Wolff*¹ en honor a quienes la propusieron por primera vez.

¹Véase Schweizer y Wolff, «On Nonparametric Measures of Dependence for Random Variables».

Teorema 1.1: Propiedades de la medida de Schweizer-Wolff.

Para cualquier vector aleatorio (X, Y) de variables aleatorias continuas, la medida de dependencia de Schweizer-Wolff $\mu_{X,Y}$ definida en (3.8) cumple con las siguientes propiedades:

- a) Siempre existe y además $0 \leq \mu_{X,Y} \leq 1$.
- b) $\mu_{X,Y} = \mu_{Y,X}$ (no importa el orden de las variables).
- c) $\mu_{X,Y} = 0$ si y solo si $C_{X,Y} = \Pi$.
- d) $\mu_{X,Y} = 1$ si y solo si $C_{X,Y} = W$ o bien $C_{X,Y} = M$.
- e) Si las funciones $\varphi : \text{Ran } X \rightarrow \mathbb{R}$ y $\psi : \text{Ran } Y \rightarrow \mathbb{R}$ son estrictamente monótonas^a entonces:

$$\mu_{\varphi(X),\psi(Y)} = \mu_{X,Y}.$$

^aEsto es, φ estrictamente creciente o estrictamente decreciente, y lo análogo para ψ .

Demostración: Sea $C_{X,Y}$ la única cópula subyacente del vector aleatorio, como consecuencia del Teorema de Sklar.

- a) Consecuencia inmediata de (2.5), (3.6), (3.7) y (3.8).
- b) Consecuencia inmediata de que

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

y del Teorema de Sklar.

- c) Consecuencia inmediata de (3.6).
- d) Consecuencia inmediata de (3.7) y (3.8).
- e) Si ambas φ y ψ son estrictamente crecientes, entonces como consecuencia del Teorema 1.8 del Capítulo 2 la cópula subyacente de (X, Y) y de $(\varphi(X), \psi(Y))$ es la misma, y por tanto el cálculo de (3.8) es el mismo. Si $\varphi \uparrow$ y $\psi \downarrow$ entonces por el Teorema 1.9a del Capítulo 2 tenemos que

$$C_{\varphi(X),\psi(Y)}(u, v) = u - C_{X,Y}(u, 1-v),$$

y al aplicarle (3.8), junto con un sencillo cambio de variable en la integral, se obtiene la conclusión, y análogamente para los otros casos.

□

Observación:

Respecto a las propiedades demostradas en el Teorema 1.1:

- a) El énfasis en decir que siempre existe es porque la popular *correlación de Pearson*^a no siempre existe, ya que ésta depende de la existencia de segundos momentos de ambas variables aleatorias.
- c) Nótese en particular que si $\mu_{X,Y} = 0$ entonces se puede afirmar que las variables aleatorias son independientes, situación que no ocurre con la correlación de Pearson, que bien puede reportar el valor cero sin que necesariamente esto implique independencia.
- d) Como consecuencia del Teorema 1.4 del Capítulo 2, el que $C_{X,Y} = W$ o bien $C_{X,Y} = M$ implica que existe una función estrictamente monótona φ (es decir, estrictamente creciente o estrictamente decreciente) tal que $\mathbb{P}(Y = \varphi(X)) = 1$, lo que significa que una vez que se conoce el valor que reporta una de las variables, digamos $\{X = x\}$, tenemos certeza del valor que corresponde a la otra, en este caso $\{Y = \varphi(x)\}$, y viceversa, porque al ser φ estrictamente monótona existe su función inversa.
- e) $\mu_{X,Y}$ es invariante, tanto bajo transformaciones estrictamente crecientes de sus variables, como ante rotaciones de su cópula subyacente (veáñese los Teoremas 1.8 y 1.9 de Capítulo 2). En el caso de transformaciones estrictamente crecientes, digamos $\varphi \uparrow$ y $\psi \uparrow$, el hecho de que (X, Y) y $(\varphi(X), \psi(Y))$ tengan la misma cópula subyacente implica que la dependencia no cambia, y por tanto la cuantificación de la dependencia debería ser la misma, y de hecho así sucede con $\mu_{X,Y} = \mu_{\varphi(X),\psi(Y)}$, pero no sucede así con la correlación de Pearson, con la que en general $r_{X,Y}$ no es necesariamente igual a $r_{\varphi(X),\psi(Y)}$.

^aLa correlación de Pearson (cuando existe) se calcula mediante:

$$r_{X,Y} = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$$

Ejemplo 20. Cálculo de la dependencia de Schweizer-Wolff.

En el Ejemplo 13 del Capítulo 2, véase (2.29), mediante el método de pegado de cópulas, se obtuvo la siguiente familia paramétrica, pegando a W seguida de M , con parámetro $0 < \theta < 1$, esto es $C_\theta = W \boxplus_\theta M$, donde:

$$(W \boxplus_\theta M)(u, v) = \begin{cases} 0, & 0 \leq u \leq \theta, 0 \leq v \leq 1 - \frac{u}{\theta}, \\ u + \theta v - \theta, & 0 \leq u \leq 1, \max\left\{1 - \frac{u}{\theta}, \frac{u-\theta}{1-\theta}\right\} \leq v \leq 1, \\ v, & \theta \leq u \leq 1, 0 \leq v \leq \frac{u-\theta}{1-\theta}. \end{cases} \quad (3.9)$$

En la Figura 3.1a se aprecia una partición del dominio de C_θ indicando en cada región el comportamiento funcional de dicha cópula. Al calcular la dependencia de Schweizer (3.8), es decir:

$$\mu_{C_\theta} = 12 \int_0^1 \int_0^1 |C_\theta(u, v) - uv| dudv, \quad (3.10)$$

para resolver la integral debemos identificar las regiones del dominio donde $C_\theta \geq \Pi$ y donde $C_\theta \leq \Pi$, para eliminar correctamente el valor absoluto y proceder.

- Si $C_\theta(u, v) = 0$ entonces $C_\theta \leq \Pi$.
- Si $C_\theta(u, v) = v$ entonces $C_\theta \geq \Pi$, por que $v \geq uv$.
- Si $C_\theta(u, v) = u + \theta v - \theta$ tenemos que:

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

véase la Figura 3.1b, así que para resolver la integral doble (3.10) debemos considerar cuatro regiones de integración, véase Figura 3.1c, e identificamos que en D_1 y D_2 se cumple que $C_\theta \leq \Pi$, y que en D_3 y D_4 se cumple $C_\theta \geq \Pi$. Procedemos a calcular la integral doble en cada región, y al final sumamos, esto es:

$$\mu_1 = 12 \iint_{D_1} (uv - 0) dv du = \frac{\theta^2}{2}, \quad (3.11)$$

$$\mu_2 = 12 \iint_{D_2} (uv - (u + \theta v - \theta)) dv du = \frac{\theta^2}{2}, \quad (3.12)$$

$$\mu_3 = 12 \iint_{D_3} (u + \theta v - \theta - uv) dv du = \frac{(1-\theta)^2}{2}, \quad (3.13)$$

$$\mu_4 = 12 \iint_{D_4} (v - uv) dv du = \frac{(1-\theta)^2}{2}, \quad (3.14)$$

$$\mu_{C_\theta} = \mu_1 + \mu_2 + \mu_3 + \mu_4 = \theta^2 + (1-\theta)^2 = 1 - 2\theta(1-\theta). \quad (3.15)$$

En la Figura 3.1d se aprecia la gráfica de valores posibles de (3.15) para distintos valores $0 < \theta < 1$. Nótese que:

$$\lim_{\theta \rightarrow 0+} \mu_{C_\theta} = 1 = \lim_{\theta \rightarrow 1-} \mu_{C_\theta}, \quad (3.16)$$

y que por el Teorema 1.1d $\mu_{C_\theta} = 1$ si y solo si $C_\theta \in \{W, M\}$, y en efecto:

$$\lim_{\theta \rightarrow 0+} C_\theta = M \quad \text{y} \quad \lim_{\theta \rightarrow 1-} C_\theta = W, \quad (3.17)$$

como era de esperarse. Pero notemos que en la escala $[0, 1]$ el mínimo valor de dependencia se alcanza cuando $\theta = \frac{1}{2}$ y resulta ser $\mu_{C_{1/2}} = \frac{1}{2}$, es decir, esta familia paramétrica $\{C_\theta : 0 < \theta < 1\}$ no incluye el caso de independencia. ¿Por qué en este caso $\mu_{C_\theta} \geq \frac{1}{2}$? A partir de (3.9) obtenemos que $\frac{\partial^2}{\partial u \partial v} C_\theta(u, v) = 0$, lo que implica que, según la Definición 2.5b del Capítulo 2, se trata de una cópula singular. Más aún, aplicando (2.33), si (U, V) es un vector aleatorio de variables aleatorias continuas, marginalmente con distribución Uniforme($0, 1$), y con la cópula subyacente C_θ de este ejemplo, entonces:

$$F_{V|U}(v | u) = \mathbb{P}(V \leq v | U = u) = \frac{\partial}{\partial u} C_\theta(u, v) = \mathbf{1}_{\{g(u) \leq v \leq 1\}} \quad (3.18)$$

donde $g(u) := \max \left\{ 1 - \frac{u}{\theta}, \frac{u-\theta}{1-\theta} \right\}$, lo que a su vez implica que:

$$\mathbb{P}(V = g(u) | U = u) = 1, \quad (3.19)$$

y que significa que, con probabilidad 1, el vector aleatorio (U, V) siempre reporta algún par de valores que pertenece al conjunto de puntos $\{(u, g(u)) : 0 \leq u \leq 1\}$, que es la poligonal en línea gruesa que se aprecia en la Figura 3.1a. Esto explica por qué, bajo todos los valores posibles del parámetro θ , la dependencia siempre está significativamente lejos de la independencia: Dado el valor $\{U = u\}$ tenemos certeza de que $\{V = g(u)\}$, y dado el valor $\{V = v\}$, si bien en este caso no tenemos certeza del valor de U , la incertidumbre sobre ello se reduce a solo dos valores posibles:

$$\mathbb{P}(U \in \{(1-v)\theta, \theta + (1-v)\theta\} \mid V = v) = 1. \quad (3.20)$$

Finalmente, como $C_\theta(\theta, 1) = \theta$, de la Figura 3.1a queda claro que con probabilidad θ el vector aleatorio (U, V) reportaría un valor que pertenece al segmento de recta que une los puntos $(0, 1)$ y $(\theta, 0)$, y por complemento, con probabilidad $1 - \theta$ reportaría un valor que pertenece al segmento de recta que une los puntos $(\theta, 0)$ y $(1, 1)$. ■

2. Correlación de Spearman

Lo que podría criticarse a la medida de dependencia de Schweizer-Wolff de la sección anterior, y de hecho considerarse como limitaciones, se resume en lo siguiente:

- $\mu_{X,Y} = 1$ implica que $\mathbb{P}(Y = \varphi(X)) = 1$ para alguna función φ que puede ser estrictamente creciente, o bien estrictamente decreciente, esto es, asigna el mismo valor ante dos extremos opuestos de dependencia.
- $\mu_{X,Y} = \mu_{\varphi(X),\psi(Y)}$ en el caso de las rotaciones $\{\varphi \uparrow, \psi \downarrow\}$ y $\{\varphi \downarrow, \psi \uparrow\}$, a pesar de que claramente cambia el sentido de la dependencia. Como ejemplo particular inmediato tenemos justamente el caso en que $C_{X,Y} = M$ porque entonces ocurre con dichas rotaciones que $\mu_{\varphi(X),\psi(Y)} = W$.

Lo anterior sucede por la diferencia en valor absoluto que (3.8) utiliza para medir la distancia entre la cópula subyacente y la que representa la independencia: al calcular el volumen encerrado entre las superficies $C_{X,Y}$ y Π mediante la integral doble de $|C_{X,Y} - \Pi|$ sobre $[0, 1]^2$ se pierde la información de cuánto de ese volumen corresponde a regiones donde $C_{X,Y} \geq \Pi$ y cuánto corresponde a regiones donde $C_{X,Y} \leq \Pi$. Si en (3.8) eliminamos el valor absoluto y definimos:

$$\rho_{X,Y} := 12 \int_0^1 \int_0^1 (C_{X,Y}(u, v) - uv) dudv, \quad (3.21)$$

obtenemos lo que se conoce como *Correlación de Spearman*.² A partir de lo ya discutido en la sección anterior, es inmediato verificar lo siguiente:

²Véase Spearman, «The Proof and Measurement of Association between Two Things».

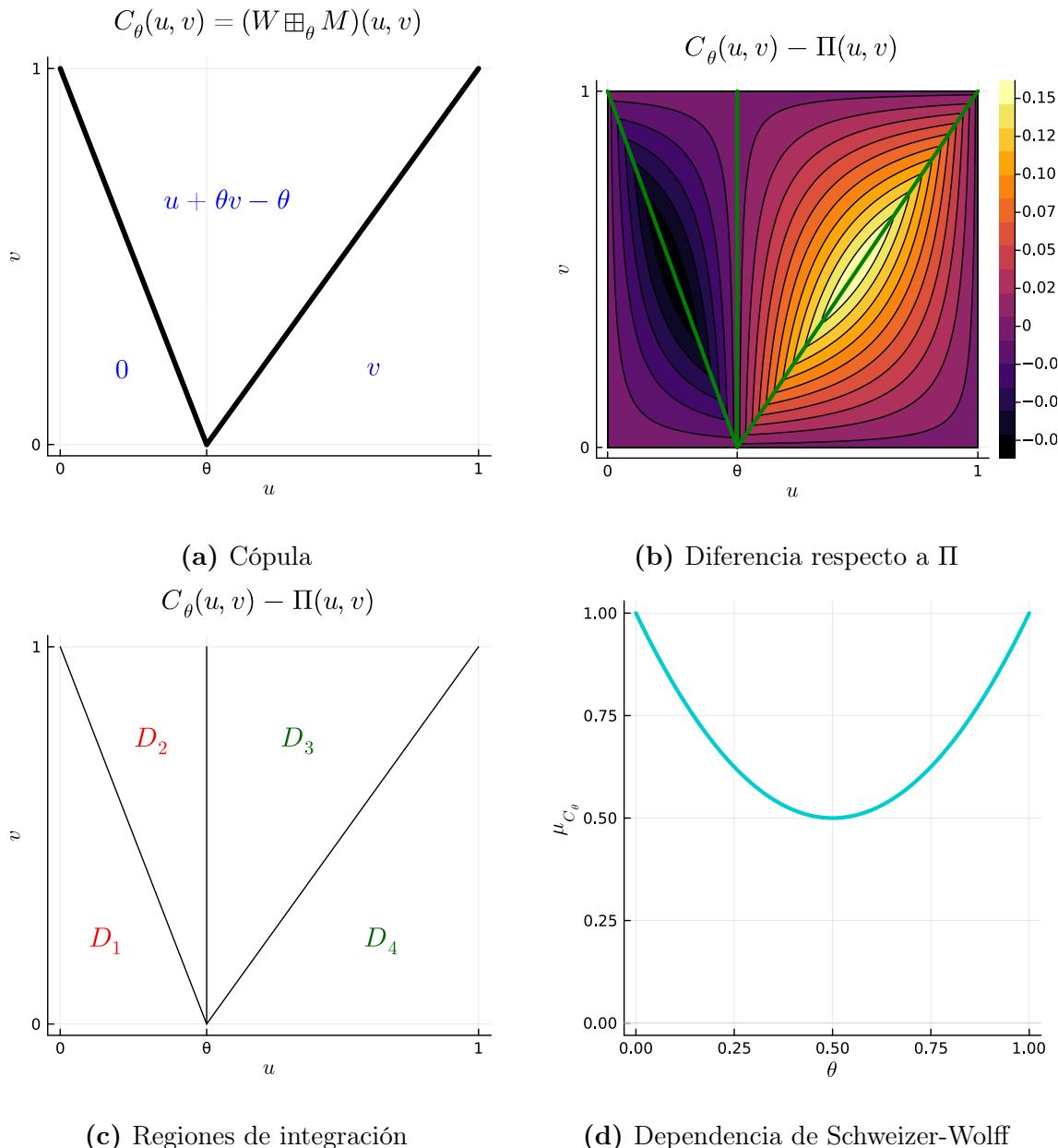


Figura 3.1: Ejemplo 20. **a)** Valores de la cúpula $C_\theta = W \boxplus_\theta M$ en la partición inducida en su dominio. **b)** Curvas de nivel de la diferencia $C_\theta - \Pi$, resultando no positiva para $u \leq \theta$, y no negativa para $u \geq \theta$. **c)** En las regiones D_1 y D_2 tenemos que $C_\theta \leq \Pi$, y en las regiones D_3 y D_4 tenemos que $C_\theta \geq \Pi$. **d)** Dependencia de Schweizer-Wolff bajo la cúpula subyacente C_θ para distintos valores $0 < \theta < 1$.

Teorema 2.1: Propiedades de la correlación de Spearman.

Para cualquier vector aleatorio (X, Y) de variables aleatorias continuas, la correlación de Spearman $\rho_{X,Y}$ definida en (3.21) cumple con las siguientes propiedades:

- Siempre existe y además $-1 \leq \rho_{X,Y} \leq +1$.
- $\rho_{X,Y} = \rho_{Y,X}$ (no importa el orden de las variables).
- Si $C_{X,Y} = \Pi$ entonces $\rho_{X,Y} = 0$, pero no viceversa.
- $\rho_{X,Y} = +1$ si y solo si $C_{X,Y} = M$, y $\rho_{X,Y} = -1$ si y solo si $C_{X,Y} = W$.
- Si las funciones $\varphi : \text{Ran } X \rightarrow \mathbb{R}$ y $\psi : \text{Ran } Y \rightarrow \mathbb{R}$ son ambas estrictamente crecientes, o bien ambas estrictamente decrecientes, entonces:

$$\rho_{\varphi(X),\psi(Y)} = \rho_{X,Y}.$$

- Si $\varphi : \text{Ran } X \rightarrow \mathbb{R}$ es estrictamente creciente, y $\psi : \text{Ran } Y \rightarrow \mathbb{R}$ estrictamente decreciente, o viceversa, entonces:

$$\rho_{\varphi(X),\psi(Y)} = -\rho_{X,Y}.$$

La propiedad d) anterior reserva los valores mínimo y máximo de $\rho_{X,Y}$ para los extremos de dependencia que representan W y M , lo cual constituye una ventaja respecto a $\mu_{X,Y}$, pero la propiedad c) nos dice que $\rho_{X,Y} = 0$ NO implica que $C_{X,Y} = \Pi$ (independencia), una clara desventaja frente a $\mu_{X,Y}$, y esto se explica porque al eliminar el valor absoluto para definir $\rho_{X,Y}$ en (3.21) bien puede ocurrir que el volumen encerrado en la región donde $C_{X,Y} \geq \Pi$ coincida con el volumen encerrado en la región donde $C_{X,Y} \leq \Pi$, y al tener signo contrario, se cancelan. Y aunque no coincidan, a un volumen se le está restando el otro, y por ello valores de $\rho_{X,Y}$ cercanos a cero no dejan claro si se debe a un cercanía con Π o bien a una resta de volúmenes en donde sí hay una dependencia significativa. Y en el caso de la propiedad f) con $\rho_{X,Y}$ se resuelve mediante un cambio de signo el reconocer el cambio en el sentido de la dependencia, cuestión que no es posible con $\mu_{X,Y}$.

Tomando en cuenta las pocas desventajas que por separado tienen la medida de dependencia de Schweizer-Wolff $\mu_{X,Y}$ y la correlación de Spearman $\rho_{X,Y}$, resultan complementarias, es decir, el par de números $(\mu_{X,Y}, \rho_{X,Y})$ constituye una fuente muy valiosa de información sobre la dependencia entre dos variables aleatorias continuas, como consecuencia de las siguientes propiedades que las relacionan, y que son inmediatamente verificables aplicando las definiciones (3.8) y (3.21).

Corolario 2.2

Para cualquier vector aleatorio (X, Y) de variables aleatorias continuas, se cumplen las siguientes propiedades:

- a) $|\rho_{X,Y}| \leq \mu_{X,Y}$;
- b) $\mu_{X,Y} = 0$ implica que $\rho_{X,Y} = 0$, pero no viceversa;
- c) $\rho_{X,Y} = \mu_{X,Y}$ si y solo si $C_{X,Y}(u, v) \geq \Pi(u, v)$ para todo $(u, v) \in [0, 1]^2$;
- d) $\rho_{X,Y} = -\mu_{X,Y}$ si y solo si $C_{X,Y}(u, v) \leq \Pi(u, v)$ para todo $(u, v) \in [0, 1]^2$;
- e) $|\rho_{X,Y}| < \mu_{X,Y}$ si y solo si existe un subconjunto propio no vacío $D \subsetneq [0, 1]^2$ tal que $C_{X,Y}(u, v) \geq \Pi(u, v)$ para todo $(u, v) \in D$, y tal que $C_{X,Y}(u, v) \leq \Pi(u, v)$ para todo $(u, v) \in [0, 1]^2 \setminus D$.

Lo anterior implica que tan solo conociendo el valor del par de números $(\mu_{X,Y}, \rho_{X,Y})$ tenemos información suficiente para saber lo siguiente:

- Si X e Y son o no independientes. Lo son cuando $\mu_{X,Y} = 0$, y por lo tanto $C_{X,Y} = \Pi$, y no lo son cuando $\mu_{X,Y} \neq 0$.
- Si $\rho_{X,Y} = +1$ entonces $C_{X,Y} = M$.
- Si $\rho_{X,Y} = -1$ entonces $C_{X,Y} = W$.
- Si $\rho_{X,Y} = \mu_{X,Y}$ entonces $C_{X,Y} \geq \Pi$ sobre todo su dominio $[0, 1]^2$.
- Si $\rho_{X,Y} = -\mu_{X,Y}$ entonces $C_{X,Y} \leq \Pi$ sobre todo su dominio $[0, 1]^2$.
- Si $|\rho_{X,Y}| < \mu_{X,Y}$ entonces en algunas partes $[0, 1]^2$ tenemos que $C_{X,Y} \geq \Pi$ y en otras $C_{X,Y} \leq \Pi$.

Ejemplo 21. Cálculo de la Correlación de Spearman.

Continuando con el Ejemplo 20, ahora calcularemos $\rho_{X,Y}$ para la familia de cópulas (3.9), que en este caso se facilita mucho gracias al análisis que ya se hizo en dicho ejemplo. En la Figura 3.1c se identificaron las regiones D_1 y D_2 donde $C_\theta \leq \Pi$, así como las regiones D_3 y D_4 donde $C_\theta \geq \Pi$. Esto implica que $|C_\theta - \Pi| = -(C_\theta - \Pi)$ sobre D_1 y D_2 , y que $|C_\theta - \Pi| = (C_\theta - \Pi)$ sobre D_3 y D_4 . Entonces al aplicar la Definición (3.21) y aprovechando los cálculos ya realizados en (3.11), (3.12), (3.13) y (3.14) obtenemos:

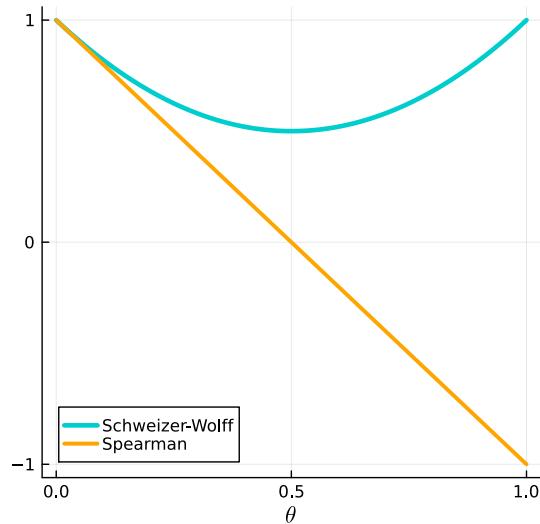
$$\rho_{C_\theta} = -\mu_1 - \mu_2 + \mu_3 + \mu_4 = (1 - \theta)^2 - \theta^2 = 1 - 2\theta, \quad 0 < \theta < 1. \quad (3.22)$$

Si $\theta \rightarrow 0+$ entonces $\rho_{C_\theta} \rightarrow 1$, lo que implicaría que $C_\theta \rightarrow M$, y si $\theta \rightarrow 1$ entonces $\rho_{C_\theta} \rightarrow -1$, lo que implicaría que $C_\theta \rightarrow W$, como era de esperarse, véase (3.17). Por otro lado $\rho_{C_\theta} = 0$ si y solo si $\theta = \frac{1}{2}$, pero dicho valor del parámetro NO implica independencia, porque de hecho la dependencia de Schweizer-Wolff $\mu_{C_{1/2}} = \frac{1}{2} \neq 0$. Como $\rho_{C_\theta} \geq 0$ si y solo

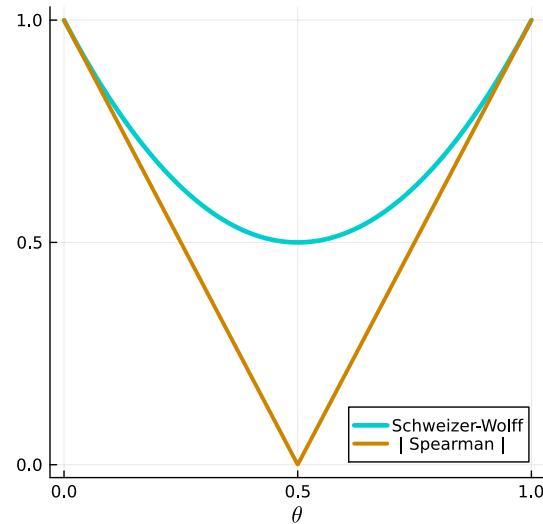
si $\theta \leq \frac{1}{2}$, es inmediato verificar que

$$\begin{aligned} \text{si } \theta \leq \frac{1}{2} &\implies |\rho_{C_\theta}| = \rho_{C_\theta} = 1 - 2\theta < 1 - 2\theta(1 - \theta) = \mu_{C_\theta}, \\ \text{si } \theta \geq \frac{1}{2} &\implies |\rho_{C_\theta}| = -\rho_{C_\theta} = 2\theta - 1 < \mu_{C_\theta}, \end{aligned}$$

con lo que se concluye $|\rho_{C_\theta}| < \mu_{C_\theta}$ para todo $0 < \theta < 1$, y que a su vez significa que en algunas partes de $[0, 1]^2$ se tiene que $C_\theta \geq \Pi$ (de hecho en $D_3 \cup D_4$, veáse Figura 3.1), y en otras $C_\theta \leq \Pi$ (de hecho en $D_1 \cup D_2$). En la Figura 3.2 podemos apreciar gráficas de la correlación de Spearman y la medida de dependencia de Schweizer-Wolff, en función de los distintos valores del parámetro θ .



(a) μ_{C_θ} versus ρ_{C_θ}



(b) μ_{C_θ} versus $|\rho_{C_\theta}|$

Figura 3.2: Ejemplo 20. **a)** Comparando los valores de la medida de dependencia de Schweizer-Wolff $0 \leq \mu_{C_\theta} \leq 1$ versus la correlación de Spearman $-1 \leq \rho_{C_\theta} \leq +1$. **b)** Comparando μ_{C_θ} versus el valor absoluto $|\rho_{C_\theta}|$.

3. Diagramas de dependencia

Cuando se tienen observaciones de un vector aleatorio (X, Y) de variables aleatorias continuas, digamos $\{(x_1, y_1), \dots, (x_n, y_n)\}$, una de las formas más populares de visualizar dicha información es mediante lo que se conoce como *diagrama de dispersión* o también *gráfico de dispersión* (en inglés: *scatter plot*), que consiste en representar cada par (x_j, y_j) como un punto en un plano cartesiano, con la expectativa de obtener una ayuda visual a partir de una nube de puntos que permita conjeturar sobre la intensidad y forma de la dependencia entre las variables X e Y .

Sin embargo, del *Teorema de Sklar* 1.1 del Capítulo 2 se desprende que el uso de diagramas

de dispersión para tal fin no es lo más adecuado, ya que toda la información de la dependencia entre un par de variables aleatorias continuas se encuentra en, y solo en, su cópula subyacente, y los valores observados $\{(x_1, y_1), \dots, (x_n, y_n)\}$ mezclan información de la dependencia que se encuentra en dicha cópula, con información de las distribuciones marginales, que solo explican el comportamiento probabilístico individual de cada variable, pero que no tienen información alguna sobre cómo interactúa una variable con otra, y esto trae como consecuencia una distorsión de lo que se pretende visualizar.

La afirmación del párrafo anterior ya quedó ilustrada en el Ejemplo 15 del Capítulo 2, véase la Figura 3.3, en donde podemos apreciar tres diagramas de dispersión que claramente lucen distintos, pero en donde la cópula subyacente es exactamente la misma, y por tanto la relación de dependencia entre las variables también, y es solo por el hecho de que tienen distribuciones marginales distintas que se ven diferentes, esto es, las distribuciones marginales distorsionan la visualización de la dependencia.

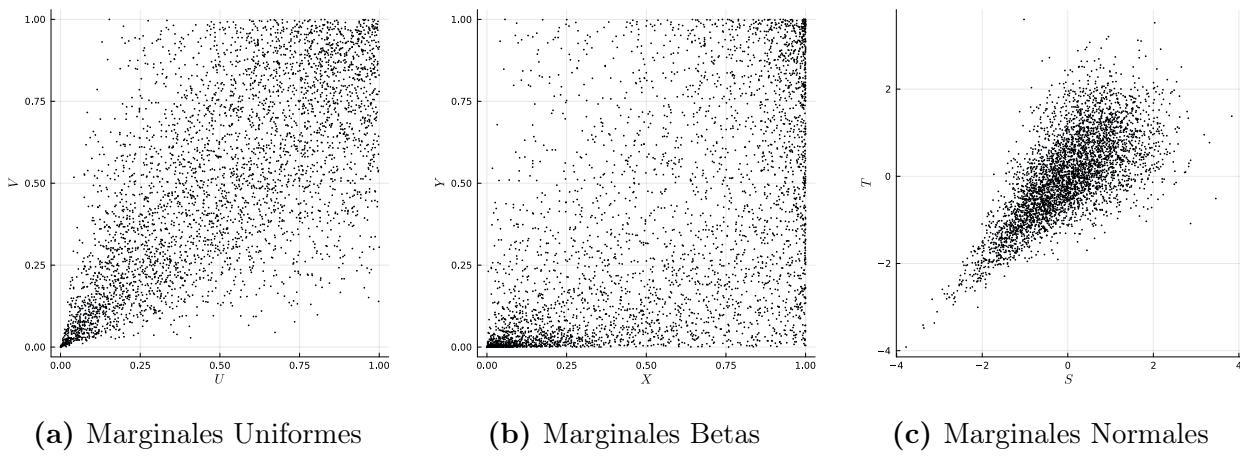


Figura 3.3: Ejemplo 15 del Capítulo 2. Simulaciones de vectores aleatorios con la misma cópula subyacente (y por tanto la misma dependencia entre variables) pero con distribuciones marginales distintas.

Otro ejemplo de lo anterior es el desarrollado por Erdely y Rubio-Sánchez³ en donde se presentan 25 diagramas de dispersión distintos, que en todos los casos se trata de simulaciones de pares de variables aleatorias continuas e independientes, y por tanto la cópula subyacente es la misma (Π en este caso), pero 5 distintos tipos de marginales generan la distorsión suficiente para lograr 25 diagramas de dispersión diferentes, que en realidad representan la misma relación de dependencia (en este caso, independencia), véase la Figura 3.4.

Convencidos de que las distribuciones marginales solo distorsionan la visualización de la dependencia entre dos variables aleatorias continuas, podemos eliminar dicha distorsión aplicando los Teoremas 3.7 y 3.8 del Capítulo 2, ya que si (X, Y) es un vector de variables aleatorias continuas, con función de distribución conjunta $F_{X,Y}$ y marginales F_X y F_Y , entonces la cópula subyacente de (X, Y) es la misma que la del vector aleatorio definido como

³Véase Erdely y Rubio-Sánchez, «D-plots: Visualizations for Analysis of Bivariate Dependence Between Continuous Random Variables».

$(U, V) := (F_X(X), F_Y(Y))$, donde marginalmente tanto U como V tienen distribución de probabilidad Uniforme(0, 1), es decir:

$$C_{X,Y} = C_{F_X(X), F_Y(Y)} = C_{U,V},$$

y como consecuencia del Corolario 1.7 del Capítulo 2, véase (2.9), las observaciones del vector aleatorio (U, V) son, de hecho, observaciones de la cópula subyacente $C_{X,Y}$. Por lo tanto, un diagrama de dispersión de observaciones de $(U, V) = (F_X(X), F_Y(Y))$ permite visualizar la dependencia entre las variables aleatorias X e Y libre de la influencia de las distribuciones marginales F_X y F_Y , ya que siempre ocurre que tanto $F_X(X)$ como $F_Y(Y)$ tienen distribución de probabilidad Uniforme(0, 1).

Por lo anterior, si $\{(x_1, y_1), \dots, (x_n, y_n)\}$ son observaciones de un vector aleatorio (X, Y) con cópula subyacente $C_{X,Y}$ y funciones de distribución marginal F_X y F_Y , entonces los pares de valores:

$$\{(F_X(x_1), F_Y(y_1)), \dots, (F_X(x_n), F_Y(y_n))\} \quad (3.23)$$

son puramente **observaciones de la cópula subyacente** $C_{X,Y}$ ya sin la influencia de las marginales F_X y F_Y , porque marginalmente tanto los valores $\{F_X(x_j) : j \in \{1, \dots, n\}\}$ como $\{F_Y(y_j) : j \in \{1, \dots, n\}\}$ tienen distribución Uniforme(0, 1). Por lo tanto, en la Figura 3.3a lo que tenemos es justamente un diagrama de dispersión de observaciones de la cópula subyacente, y si se aplicara la transformación (3.23) a las Figuras 3.3b y 3.3c sus diagramas de dispersión se verían prácticamente iguales a la Figura 3.3a.

En la Figura 3.4, el gráfico de dispersión en la esquina superior izquierda corresponde, de hecho, a la cópula subyacente (Π en ese caso), y si se aplicara la transformación (3.23) en todos los demás casos se verían todos los diagramas de dispersión como el de la esquina superior izquierda.

Entonces, dada una muestra aleatoria observada $\{(x_1, y_1), \dots, (x_n, y_n)\}$ proveniente de un vector aleatorio (X, Y) de variables aleatorias continuas, con cópula subyacente $C_{X,Y}$ y funciones de distribución marginal F_X y F_Y , una adecuada visualización de la dependencia entre las variables aleatorias X e Y es un gráfico de dispersión de las **observaciones de la cópula** subyacente:

$$\{(u_j, v_j) : u_j = F_X(x_j), v_j = F_Y(y_j), j \in \{1, \dots, n\}\}. \quad (3.24)$$

En caso de que F_X y/o F_Y sean desconocidas, en (3.24) pueden ser reemplazadas por estimaciones (paramétricas y/o no paramétricas) \hat{F}_X y \hat{F}_Y , y en tal caso se les denomina **pseudo-observaciones de la cópula** subyacente, por tratarse de una aproximación vía estimación de las marginales:

$$\{(\hat{u}_j, \hat{v}_j) : \hat{u}_j = \hat{F}_X(x_j), \hat{v}_j = \hat{F}_Y(y_j), j \in \{1, \dots, n\}\}. \quad (3.25)$$

Será siempre importante tener visualmente presente cómo se ven observaciones de tres cópulas importantes: W , Π y M , porque todos los demás casos son algo intermedio entre ellas, incluyendo combinaciones lineales o pegado de cópulas de esos casos intermedios. Recordemos que los pasos 1 al 5 del Algoritmo 3.1 del Capítulo 2 nos indican justamente cómo simular

observaciones de una cópula. Por su importancia y utilidad para los fines de este capítulo lo transcribimos a continuación:

Algoritmo 3.1

Para simular n observaciones iid de una cópula dada C :

1. Simula $u_1, \dots, u_n \sim \text{iid Uniforme}(0, 1)$.
2. Define la función $\psi(v | u) := \frac{\partial}{\partial u} C(u, v)$, y deduce su cuasi-inversa $\psi^{(-1)}(t | u)$.
3. Simula $t_1, \dots, t_n \sim \text{iid Uniforme}(0, 1)$, independientes de los valores u_j .
4. Para cada $j \in \{1, \dots, n\}$ calcular $v_j = \psi^{(-1)}(t_j | u_j)$.
5. $(u_1, v_1), \dots, (u_n, v_n)$ son iid $\sim C$.

Teorema 3.2: Simular las cópulas W , Π y M .

Sean u_1, \dots, u_n valores observados iid de una variable aleatoria continua $U \sim \text{Uniforme}(0, 1)$, y sean también t_1, \dots, t_n valores observados iid de otra variable aleatoria continua $T \sim \text{Uniforme}(0, 1)$, donde U y T son independientes. Entonces:

- a) $\{(u_j, t_j) : j \in \{1, \dots, n\}\}$ son pares de observaciones iid de la cópula Π .
- b) $\{(u_j, u_j) : j \in \{1, \dots, n\}\}$ son pares de observaciones iid de la cópula M .
- c) $\{(u_j, 1 - u_j) : j \in \{1, \dots, n\}\}$ son pares de observaciones iid de la cópula W .

Demostración: Recordemos de (2.33) que si (U, V) es un vector aleatorio de variables aleatorias continuas, con cópula subyacente C y marginales $\text{Uniforme}(0, 1)$, entonces para cualesquiera u y v en $[0, 1]$ la función de distribución condicional:

$$F_{V|U}(v | u) = \mathbb{P}(V \leq v | U = u) = \frac{\partial}{\partial u} C(u, v),$$

además de que marginalmente $F_V(v) = v$ porque $V \sim \text{Uniforme}(0, 1)$.

- a) Si $C(u, v) := \Pi(u, v) = uv$ entonces $F_{V|U}(v | u) = v = F_V(v)$, lo que implica que U y V son independientes, y por tanto simulaciones independientes de U y V que se utilicen para definir pares de valores constituyen observaciones de la cópula Π .
- b) Si $C(u, v) := M(u, v) = \min\{u, v\} = v\mathbf{1}_{\{0 \leq v \leq u\}} + u\mathbf{1}_{\{u < v \leq 1\}}$ entonces $F_{V|U}(v | u) = \mathbf{1}_{\{u < v < 1\}}$, lo que a su vez implica que $\mathbb{P}(V = u | U = u) = 1 = \mathbb{P}(V = U)$, y por lo tanto el vector aleatorio (U, V) tiene por cópula subyacente a M , lo que a su vez implica que $\{(u_j, u_j) : j \in \{1, \dots, n\}\}$ son pares de observaciones iid de la cópula M .

- c) Si $C(u, v) := W(u, v) = \max\{u+v-1, 0\} = (u+v-1)\mathbf{1}_{\{1-u \leq v \leq 1\}}$ entonces $F_{V|U}(v | u) = \mathbf{1}_{\{1-u < v < 1\}}$, lo que a su vez implica que $\mathbb{P}(V = 1 - u | U = u) = 1 = \mathbb{P}(V = 1 - U)$, y por lo tanto el vector aleatorio $(U, 1 - U)$ tiene por cópula subyacente a W , lo que a su vez implica que $\{(u_j, 1 - u_j) : j \in \{1, \dots, n\}\}$ son pares de observaciones iid de la cópula W .

□

En la Figura 3.5 podemos apreciar $n = 3000$ pares de observaciones simuladas a partir de las cópulas W , Π y M , indicando los valores de la medida de Schweizer-Wolff (3.8) y de la Correlación de Spearman (3.21).

Ejemplo 22. $\Pi \leq C_\theta \leq M$.

En el Ejemplo 14 se construyó una familia paramétrica de cópulas arquimedianas, véase (2.31), dada por:

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad \theta > 0.$$

Es un sencillo ejercicio algebraico verificar que para todo $(u, v) \in [0, 1]^2$ y cualquier $\theta > 0$ se cumple que $C_\theta(u, v) \geq uv = \Pi(u, v)$, lo que implica, como consecuencia del Corolario 2.2c, que en este caso $\mu_{C_\theta} = \rho_{C_\theta}$. También es un sencillo ejercicio algebraico demostrar que si $0 < \theta_1 \leq \theta_2$ entonces $C_{\theta_1}(u, v) \leq C_{\theta_2}(u, v)$ para todo $(u, v) \in [0, 1]^2$. Más aún, es un sencillo ejercicio de cálculo de límites demostrar que $\lim_{\theta \rightarrow 0^+} C_\theta = \Pi$ y que $\lim_{\theta \rightarrow +\infty} C_\theta = M$, implicando esto a su vez que $\lim_{\theta \rightarrow 0^+} \mu_{C_\theta} = 0$ y $\lim_{\theta \rightarrow +\infty} \mu_{C_\theta} = 1$. Lo anterior significa que conforme θ decrece entonces la cópula C_θ se va aproximando más a Π , y por tanto μ_{C_θ} se va aproximando más a 0, y que conforme θ crece entonces la cópula se va aproximando más a M , y por tanto μ_{C_θ} se va aproximando más a 1, como se ilustra en las simulaciones de C_θ de la Figura 3.6 (véase el Ejemplo 15 del Capítulo 2 para la fórmula de simulación), con distintos valores del parámetro $0 < \theta < 1$, así como en la gráfica de valores de μ_{C_θ} en función de θ .

Es entonces mediante un diagrama de dispersión de las observaciones (o bien de las pseudo-observaciones, según sea el caso) de la cópula subyacente, obtenidas mediante (3.24) o (3.25), como realmente podemos visualizar la dependencia entre dos variables aleatorias continuas, sin la distorsión de las distribuciones marginales. Los llamaremos **diagramas de dependencia**. En Erdely y Sánchez-Rubio⁴ se presentan 9 patrones usuales, véase la Figura 3.7, con las siguientes interpretaciones:

- R1:** Distribución uniforme (o muy cercano a ello) de puntos sobre $[0, 1]^2$. La cópula subyacente $C \approx \Pi$, y $\mu_C \approx 0$.
- R2:** Los puntos se encuentran más cercanos a la diagonal principal $v = u$. Entonces $C \geq \Pi$, y por lo tanto $\mu_C = \rho_C > 0$.
- R3:** Los puntos se encuentran más cercanos a la diagonal secundaria $v = 1 - u$. Entonces $C \leq \Pi$, y por lo tanto $\mu_C = -\rho_C$, donde $\rho_C < 0$.

⁴Véase Erdely y Rubio-Sánchez, «D-plots: Visualizations for Analysis of Bivariate Dependence Between Continuous Random Variables»

- R4:** Combinación lineal convexa (véase Sección 2.3 del Capítulo 2) de algo del tipo R2 con algo del tipo R3. Habrá regiones en $[0, 1]^2$ en donde $C \geq \Pi$ y otras donde $C \leq \Pi$, con $-\mu_C < \rho_C < +\mu_C$.
- R5:** Combinación lineal convexa de algo del tipo R1 con algo del tipo R2. $C \geq \Pi$, y por lo tanto $\mu_C = \rho_C > 0$.
- R6:** Combinación lineal convexa de algo del tipo R1 con algo del tipo R3. $C \leq \Pi$, y por lo tanto $\mu_C = -\rho_C$, donde $\rho_C < 0$.
- R7:** Pegado de cópulas (véase Sección 2.3 del Capítulo 2) de algo del tipo R2 seguido de algo del tipo R3. Habrá regiones en $[0, 1]^2$ en donde $C \geq \Pi$ y otras donde $C \leq \Pi$, con $-\mu_C < \rho_C < +\mu_C$.
- R8:** Pegado de cópulas de algo del tipo R3 seguido de algo del tipo R2. Habrá regiones en $[0, 1]^2$ en donde $C \geq \Pi$ y otras donde $C \leq \Pi$, con $-\mu_C < \rho_C < +\mu_C$.
- R9:** Pegado de cópulas de algo del tipo R1 seguido de algo del tipo R3. Habrá regiones en $[0, 1]^2$ en donde $C \geq \Pi$ y otras donde $C \leq \Pi$, con $-\mu_C < \rho_C < +\mu_C$.

La lista anterior no es exhaustiva, pero da una idea de patrones comunes con los cuales podrían construirse otros, también mediante las técnicas descritas en la Sección 2.3 del Capítulo 2. Dicha lista es útil para identificar, o bien construir, cópulas paramétricas o no paramétricas que tengan un comportamiento similar, en términos de simulación de observaciones, al del diagrama de dependencia de un conjunto de pares observados que nos interese modelizar.

4. Actividades

- 4.1) Sea C la cópula subyacente de un vector aleatorio (X, Y) de variables aleatorias continuas.

- a) Demuestra que la medida definida como:

$$\Phi_{X,Y} := \left(90 \int_0^1 \int_0^1 (C(u, v) - uv)^2 dudv \right)^{1/2}$$

cumple las mismas propiedades que las ya demostradas para la medida de dependencia de Schweizer-Wolff. A $\Phi_{X,Y}$ se le conoce como medida de dependencia de Hoeffding.⁵

- b) Verifica si la medida definida como:

$$\Lambda_{X,Y} := 4 \sup \{ |C(u, v) - uv| : (u, v) \in [0, 1]^2 \}$$

cumple las mismas propiedades que las ya demostradas para la medida de dependencia de Schweizer-Wolff. (La respuesta es negativa, identifica cuál o cuáles no cumple).

⁵Véase Hoeffding, «[Masstabinvariante Korrelationstheorie](#)».

- 4.2) Desarrolla los detalles para el cálculo de las integrales correspondientes que se requieren en la demostración del Teorema 1.1e y del Teorema 2.1f.
- 4.3) Sea X una variable aleatoria continua Uniforme(0, 1), y se define la variable aleatoria $Y := X^n$ donde n es algún entero positivo. Demuestra y explica por qué en este caso la medida de dependencia de Schweizer-Wolff $\mu_{X,Y} = 1$ y la correlación de Spearman $\rho_{X,Y} = +1$ para todo entero positivo n , y que esto contrasta con la correlación de Pearson, que resulta $r_{X,Y} < +1$ para todo $n \geq 2$, y que además conforme $n \rightarrow +\infty$ entonces $r_{X,Y} \downarrow 0+$.
- 4.4) Para la familia paramétrica de cópulas del Ejemplo 22 demuestra lo siguiente:
- $C_\theta(u, v) \geq uv$, para todo $(u, v) \in [0, 1]^2$ con cualquier $\theta > 0$.
 - Si $0 < \theta_1 \leq \theta_2$ entonces $C_{\theta_1}(u, v) \leq C_{\theta_2}(u, v)$ para todo $(u, v) \in [0, 1]^2$.
 - Cópulas límite de la familia paramétrica:

$$\lim_{\theta \rightarrow 0+} C_\theta = \Pi, \quad \lim_{\theta \rightarrow +\infty} C_\theta = M.$$

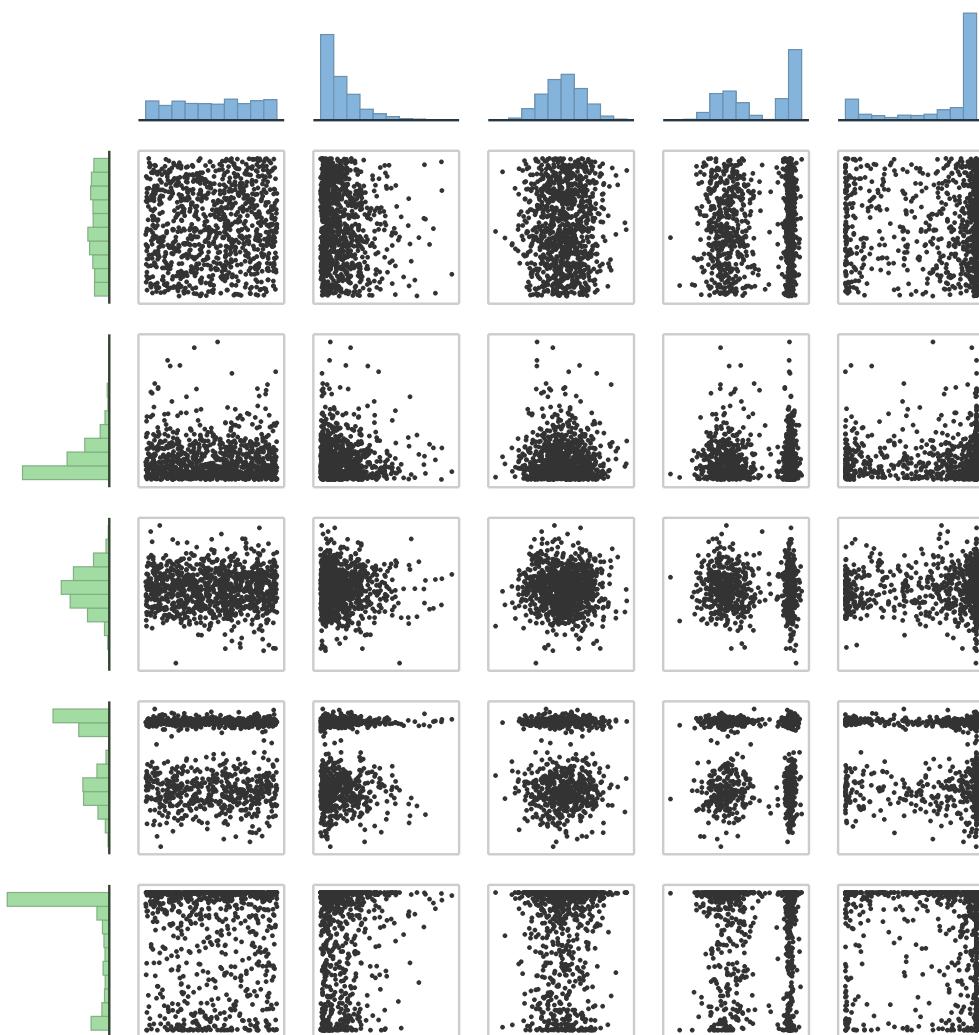


Figura 3.4: Diagramas de dispersión de simulaciones de pares de variables aleatorias continuas e independientes, pero con funciones de densidad de probabilidades marginales distintas, como se indica en los histogramas en color verde y azul. Se consideraron 5 tipos de densidades marginales, en el siguiente orden: uniforme, decreciente, unimodal, y dos tipos de bimodales. Fuente de la gráfica: Erdely y Rubio-Sánchez, «D-plots: Visualizations for Analysis of Bivariate Dependence Between Continuous Random Variables».

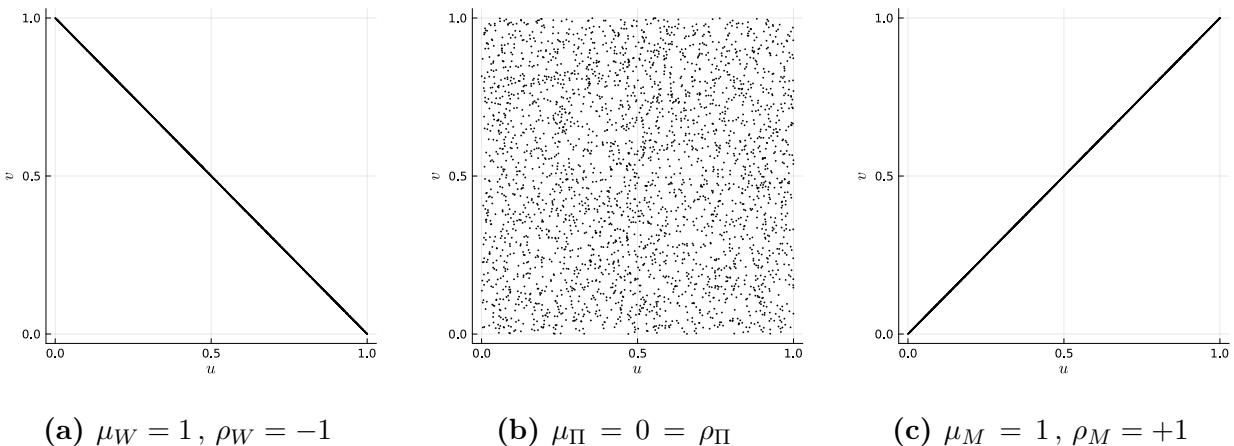


Figura 3.5: $n = 3000$ simulaciones de las cópulas W , Π y M aplicando el Teorema 3.2.

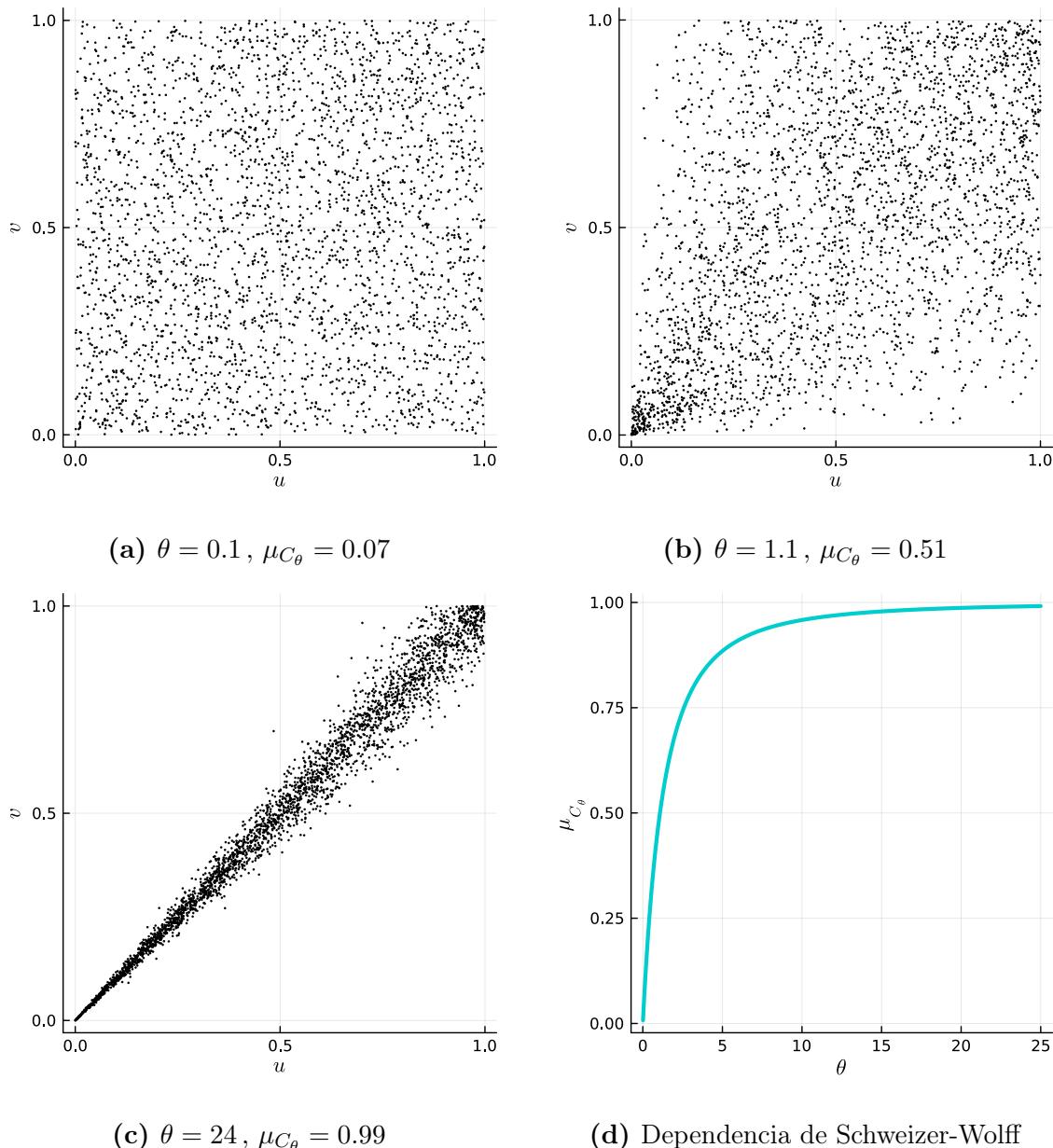


Figura 3.6: Ejemplo 22. **a)** C_θ muy cercana a Π . **b)** C_θ en un punto intermedio entre Π y M en términos de su dependencia. **c)** C_θ muy cercana a M . **d)** Dependencia de Schweizer-Wolff bajo la cópula subyacente C_θ para distintos valores $0 < \theta < 1$.

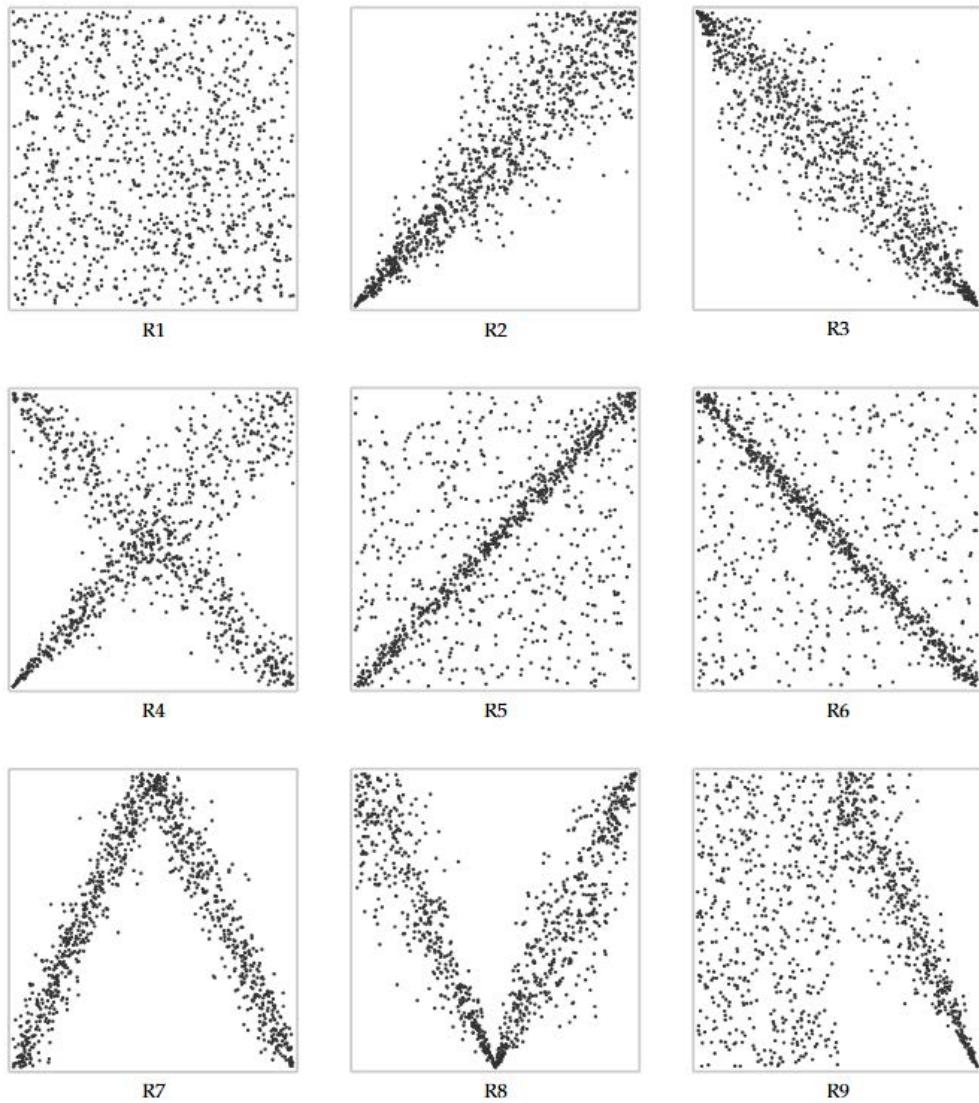


Figura 3.7: Ejemplos de patrones usuales al visualizar diagramas de dependencia (ver el texto para las interpretaciones). Fuente de la gráfica: Erdely y Rubio-Sánchez, «D-plots: Visualizations for Analysis of Bivariate Dependence Between Continuous Random Variables».

BIBLIOGRAFÍA

Spearman: The Proof and Measurement of Association between Two Things Spearman

C. Spearman. «The Proof and Measurement of Association between Two Things». En: *American Journal of Psychology* 15 (1904), págs. 72-101.

Hoeffding: Masstabvariante Korrelationstheorie Hoeffding

Wassily Hoeffding. «Masstabvariante Korrelationstheorie». En: *Schriften des Matematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin* 5.3 (1940), págs. 179-233.

Sklar: Fonctions de répartition à n dimensions et leurs marges Sklar

Abe Sklar. «Fonctions de répartition à n dimensions et leurs marges». En: *Publ Inst Statist Univ Paris* 8 (1959), págs. 229-231.

Clarke: Random Variables Clarke

Lionel Edward Clarke. *Random Variables*. 1.^a ed. Longman, 1975.

Schweizer et al.: On Nonparametric Measures of Dependence for Random Variables Schweizer

Berthold Schweizer y Edward F. Wolff. «On Nonparametric Measures of Dependence for Random Variables». En: *Annals of Statistics* 9 (1981), págs. 879-885.

Muñoz-Pérez et al.: Estimating the quantile function by Bernstein polynomials Muñoz-Pérez

J. Muñoz-Pérez y A. Fernández-Palacín. «Estimating the quantile function by Bernstein polynomials». En: *Comput. Stat. Data Anal.* 5 (1987), págs. 391-397.

Billingsley: Probability and Measure Billingsley

Patrick Billingsley. *Probability and Measure*. 3.^a ed. Wiley, 1995.

Lindley: The philosophy of statistics**Lindley**

Dennis V. Lindley. «The philosophy of statistics». En: *The Statistician* 49.3 (2000), págs. 293-337.

Domínguez-Martínez: Diseño y análisis de modelos de probabilidad Dominguez

Juan-Ignacio Domínguez-Martínez. *Diseño y análisis de modelos de probabilidad*. 1.^a ed. Grupo Editorial Iberoamérica, 2001.

Sancetta et al.: The Bernstein copula and its applications to modeling and approximations of multivariate distributions Sancetta

Alessio Sancetta y Stephen Satchell. «The Bernstein copula and its applications to modeling and approximations of multivariate distributions». En: *Econometric Theory* 20 (2004), págs. 535-562.

Nelsen: An Introduction to Copulas**Nelsen**

Roger B. Nelsen. *An Introduction to Copulas*. 2.^a ed. Springer, 2006.

Siburg et al.: Gluing copulas**Siburg**

K.F. Siburg y P.A. Stoimenov. «Gluing copulas». En: *Commun Stat Theory Methods* 37 (2008), págs. 3124-3134.

Sisson et al.: Handbook of Approximate Bayesian Computation**Sisson**

Scott A. Sisson, Yanan Fan y Mark A. Beaumont. *Handbook of Approximate Bayesian Computation*. 1.^a ed. Chapman y Hall/CRC, 2018.

Erdely et al.: D-plots: Visualizations for Analysis of Bivariate Dependence Between Continuous Random Variables Erdely

Arturo Erdely y Manuel Rubio-Sánchez. «D-plots: Visualizations for Analysis of Bivariate Dependence Between Continuous Random Variables». En: *Stats* 8.43 (2025), págs. 1-24.