La importancia de las cópulas.

Juan Vazquez M.[a], Hortensia Reyes C., Bulmaro Juárez H.[c].
[a,b,c] *Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias Físico Matemáticas, Puebla, Puebla, México.*
[a] javqzquez0314@gmail.com, [b] hreyes@fcfm.buap.mx, [c] bjuarez@fcfm.buap.mx .

Resumen

En este trabajo se presenta una descripción general de la teoría básica de las cópulas y sus aplicaciones en la estadística y probabilidad, el cual es un tema novedoso y moderno. El escrito es divulgativo, en el cual se explica las cópulas para dos variables, su relación con las funciones de distribución conjunta (Teorema de Sklar), ventajas y aplicaciones de ellas. Una vez introducida la noción de cópula y el Teorema de Sklar, se explica como a partir de las funciones de distribución $\boldsymbol{F}$ y $\boldsymbol{G}$, de las variables $\boldsymbol{X}$ y $\boldsymbol{Y}$ respectivamente, se puede obtener una función de distribución bidimensional $\boldsymbol{H}$, cuyas funciones marginales son $\boldsymbol{F}$ y $\boldsymbol{G}$, a partir de una cópula. Esta construcción, evita hacer la suposición fuerte de la independencia entre $\boldsymbol{X}$ y $\boldsymbol{Y}$, dando una función $\boldsymbol{H}(\boldsymbol{x},\boldsymbol{y})$ distinta al producto $\boldsymbol{F}(\boldsymbol{x})\boldsymbol{G}(\boldsymbol{y})$, siempre y cuando se tome una cópula distinta a la cópula producto.

Palabra claves: Cópula, Márgenes fijos, Distribución Multivariante.

Introducción

En un modelo estadístico multivariado se captura la estructura de dependencia de las variables aleatorias involucradas, cualesquiera que sean las funciones de distribución de las variables aleatorias individuales. También se permite la construcción de familias de distribuciones bivariadas o multivariadas. Fisher (2016) menciona que, "Las cópulas son de interés para los estadísticos por dos razones principales: en primer lugar, como una forma de estudiar medidas de dependencia sin escala, y en segundo lugar, como punto de partida para la construcción de familias de distribuciones multivariadas, a veces con miras a la simulación".

La palabra cópula es un sustantivo latino que significa "un vínculo, lazo, enlace", y se usa en gramática y lógica para describir "la parte de una proposición que conecta el sujeto y el predicado". La palabra cópula fue empleada por primera vez en un sentido matemático por Sklar (1959) en el teorema (que ahora lleva su nombre) que describe las funciones que "unen" funciones de distribución unidimensionales para formar funciones de distribución multivariadas, ver el trabajo de Nelsen (2011). En el trabajo solo se utilizará la noción de cópula para dos variables, y para casos más generales ver Sempi (2007).

Preliminares

Definiciones generales

Para entender la definición de cópula, se introduce una notación previa. Se denota por $\mathbb{R}$ a los reales, $\overline{\mathbb{R}}$ a reales extendidos y $\overline{\mathbb{R}}^2$ al plano $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$. Un rectángulo B en $\overline{\mathbb{R}}^2$ es el producto cartesiano de dos intervalos cerrados, es decir, $B = [x_1 : x_2] \times [y_1 : y_2]$. Los vértices de un rectángulo denotados por los puntos (x_1, y_1), (x_1, y_2), (x_2, y_1) y (x_2, y_2). La unidad cuadrada $\boldsymbol{I}^2$ es el producto $\boldsymbol{I} \times \boldsymbol{I}$, donde $\boldsymbol{I} = [0: 1]$.

El primer concepto que se define es el de H-volumen.

Definición 1. Sean S_1 y S_2 subconjuntos no vacíos de $\mathbb{R}$, y sea H una función real tal que $Dom\, H = S_1 \times S_2$. Sea $B = [x_1 : x_2] \times [y_1 : y_2]$ un rectángulo cuyos vértices están en $Dom\, H$. Entonces el H-volumen de B viene dado por

$$\mathrm{V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1).}$$

Cópulas

Con la definición anterior, ya se puede hablar del concepto de cópula y el Teorema de Sklar.

Definición 2. Una función $C: \boldsymbol{I}^2 \to \boldsymbol{I}$ es una cópula si

a) $C(u, v) = 0$ si $u = 0$ o $v = 0$,
b) $C(u, 1) = u$ y $C(1, v) = v$,
c) el V_C-volumen de cualquier rectángulo B, con vértices en $\boldsymbol{I}^2$, es no negativo, es decir, $V_C(B) \geq 0$.

La propiedad c) generalmente se conoce como la "propiedad creciente de una cópula". Cada cópula es la restricción al cuadrado unitario $\boldsymbol{I}^2$ de una función de distribución que concentra toda la masa de probabilidad en $\boldsymbol{I}^2$ y que tiene márgenes uniformes (y esto también puede servir como una definición equivalente).

Teorema de Sklar

El teorema de Sklar es fundamental para la teoría de las cópulas y es la base de varias de las aplicaciones de la estadística. El Teorema de Sklar aclara el papel que juegan las cópulas en la relación entre las funciones de distribución multivariadas y sus marginales univariadas.

Teorema 1. Sea H una función de distribución conjunta con márgenes F y G. Entonces existe una cópula C tal que para todo x y y en $\overline{\mathbb{R}}$,

$$H(x, y) = C(F(x), G(y)). \qquad (1)$$

Si F y G son continuos, entonces C es única; de lo contrario, C se determina únicamente en $Ran\, F \times Ran\, G$. Por otro lado, si C es una cópula y F y G son funciones de distribución, entonces la función H definida en (1) es una función de distribución conjunta con marginales F y G.

La demostración de este hecho se encuentra en Nelsen (2007). La segunda parte del Teorema ("inversa") del Teorema de Sklar es especialmente importante en la construcción de modelos estadísticos, ya que permite proceder en dos pasos separados:

- Se eligen las funciones de distribución unidimensionales F y G que describen el comportamiento de las cantidades estadísticas individuales (variables aleatorias) X y Y que aparecen en el modelo.
- Encajar F y G por medio de una cópula C del Teorema de Sklar que captura las relaciones de dependencia entre X y Y.

Estos dos pasos son independientes en el sentido de que, una vez elegida una cópula C, cualquier elección de las funciones de distribución F y G es posible.

Además, la ecuación (1) da una expresión para las funciones de distribución conjunta en términos de una cópula y dos funciones de distribución univariadas. Sin embargo, (1) se puede invertir para expresar cópulas en términos de una función de distribución conjunta y las "inversas" de los dos márgenes.

Por otro lado, si una marginal no es estrictamente creciente, entonces no posee una inversa en el sentido habitual. Por esta razón, se necesita definir las "quasi-inversas" de funciones de distribución.

Definición 3. Sea F una función de distribución. Entonces una quasi-inversa de F es una función $F^{(-1)}$ con dominio $\boldsymbol{I}$, tal que

1. Si t está en el $Ran\, F$, entonces $F^{(-1)}(t)$ es algún número x en $\overline{\mathbb{R}}$ tal que $F(x) = t$, es decir, para todo t en el $Ran\, F$,

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

2. Si t no está en el $Ran\, F$, entonces

$$F^{(-1)}(t) = ínf\{x|F(x) \geq t\} = \sup\{x|F(x) \leq t\}.$$

Por la definición 3, si F es estrictamente creciente, entonces solo tiene una quasi-inversa, que es la inversa ordinaria, para el cual se usa la notación habitual F^{-1}.

Ahora se está en condiciones para formular el Lema siguiente:

Lema 1. Sea $\boldsymbol{H}$, una función de distribución conjunta con marginales continuas $\boldsymbol{F}$ y $\boldsymbol{G}$, y C la cópula del Teorema de Sklar, entonces para cualquier $(\boldsymbol{u}, \boldsymbol{v})$ en el $\boldsymbol{Dom\, C}$,

$$\boldsymbol{C}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{H}\left(\boldsymbol{F}^{(-1)}(\boldsymbol{u}), \boldsymbol{G}^{(-1)}(\boldsymbol{v})\right).$$

Como se mencionó anteriormente, el Lema 1 proporciona un método para construir cópulas, esto a partir de funciones de distribución conjuntas.

Con una extensión apropiada de su dominio a $\overline{\mathbb{R}}^2$, cada cópula es una función de distribución conjunta, con marginales uniformes en $\boldsymbol{I}$. Para más claridad, sea C una cópula y se define la función H_C en $\overline{\mathbb{R}}^2$ mediante

$$H_C(x, y) = \begin{cases} 0, & x < 0 \; o \; y < 0, \\ C(x, y), & (x, y) \in \boldsymbol{I}^2, \\ x, & y > 1, x \in \boldsymbol{I}, \\ y, & x > 1, y \in \boldsymbol{I}, \\ 1, & x > 1, y > 1. \end{cases}$$

Entonces $\boldsymbol{H_C}$ es una función de distribución cuyas marginales son $\boldsymbol{U}(\mathbf{0}, \mathbf{1})$.

Cópulas y variables aleatorias

Teorema 2. Sean X y Y variables aleatorias con funciones de distribución F y G, respectivamente, y función de distribución conjunta H. Entonces existe una cópula C tal que (1) es cierta. Si F y G son continuas, C es única. De lo contrario, C es únicamente determinada en $Ran\, F \times Ran\, G$.

La cópula C del Teorema 2 se denominará cópula de X y Y , y se denotará C_{XY} cuando su identificación con las variables aleatorias X y Y sea ventajosa.

Hay diversos modelos de cópulas, por ejemplo, las cópulas gaussianas, t-student de Gumbel, etc., explicadas en Novales (2017). En este documento, el Teorema siguiente muestra que la cópula del producto $\Pi(u, v) = uv$ caracteriza las variables aleatorias independientes cuando las funciones de distribución son continuas.

Teorema 3. Sean X y Y variables aleatorias continuas. Luego, X y Y son independientes si y sólo si $C_{XY} = \Pi$.

Demostración. La demostración se deriva del Teorema 2 y del supuesto de que X y Y son independientes si y solo si $H(x,y) = F(x)G(y)$ para todo x y y en $\overline{\mathbb{R}}^2$.

Gran parte de la utilidad de las cópulas en el estudio de la estadística no paramétrica se deriva del manejo de las transformaciones estrictamente monótonas de las variables aleatorias, que poseen cópulas que son invariantes o cambian de manera predecible, como lo indica el Teorema 3.

Teorema 4. Sean X e Y variables aleatorias continuas con cópula C_{XY}. Sean α y β estrictamente monótonas en $Ran\, X$ y $Ran\, Y$, respectivamente.

1. Si α y β son ambas estrictamente crecientes, entonces

$$C_{\alpha(X),\beta(Y)} = C_{XY}.$$

2. Si α es estrictamente creciente y β es estrictamente decreciente, entonces

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

3. Si α es estrictamente decreciente y β es estrictamente creciente, entonces

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

4. Si α y β son ambas estrictamente decrecientes, entonces

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

La demostración de este Teorema, se basa en la definición de las funciones de distribución y propiedades de transformaciones estrictamente monótonas. El primer punto se demuestra en Nelsen (2007), y la demostración de los puntos restantes es similar.

Conclusiones

La introducción de cópulas en la literatura estadística ha permitido facilitar el construir modelos, que se obtienen por medio de dos pasos:

(i) la especificación de las leyes marginales de las variables aleatorias involucradas y
(ii) la introducción de una cópula que describa la estructura de dependencia entre las variables.

En muchas aplicaciones esto ha permitido evitar el supuesto de independencia, el cual es un proceso matemáticamente elegante y fácil de manejar, pero generalmente injustificado.

Gracias a las cópulas se ha permitido estudiar de manera estadística diversos fenómenos en los que dos o más variables están correlacionadas. Trabajos actuales que tienen como sustento matemático el concepto de cópulas, son Vásquez y Escarela (2019), para estudiar los datos de riesgo, y Cherubine (2011) que explica las aplicaciones a las finanzas.

Actualmente ya existen algoritmos para la simulación de cópulas, por ejemplo, Mia y Scherer (2017), por lo cual se pueden ya hacer muchas aplicaciones en las ciencias, como ya se había mencionado en este escrito. Trabajos futuros de los autores, incluyen encontrar dos componentes químicos en cierta región, los cuales se les pueda comprobar que existe correlación. Con ayuda de las cópulas, se espera completar datos faltantes, a partir de los datos existentes.

Referencias

Cherubini U. (2011). Copulas in Finance. **International Encyclopedia of Statistical Science**. Springer. Pp. 305-309.

Fisher N.I. (2016). Copulas. **Encyclopedia of Statistical Sciences**, 2da edición. Vol. 2. Wiley, New York, Pp. 1363-1367.

Mia, F. & Scherer, M. (2017). **Simulating copulas: stochastic models, sampling algorithms, and applications**. Springer. 2da. Edition. Vol. 6.

Nelsen, R. (2007). **An introduction to Copulas**. 2da edición. Springer Science & Business Media.

Novales A. (2017). **Cópulas**. Departamento de Economía, Universidad Complutense.

Sempi C. (2011). Copulas. **International Encyclopedia of Statistical Science**. Springer. Pp. 302-305.

Sklar, M. (1959). **Fonctions de repartition an dimensions et leurs marges**. Publ. inst. statist. univ. Paris, Vol. 8, pp. 229-231.

Vásquez, A. & Escarela, G. (2019). Parametric and semiparametric copula-based models for the regression analysis of competing risks. **Communications in Statistics-Theory and Methods**, pp. 1-17.