

**METODOLOGÍA PARA MODELAR LA ASOCIACIÓN ENTRE
DOS VARIABLES ALEATORIAS CONTINUAS A TRAVÉS DE
FUNCIONES CÓPULA: APLICACIÓN EN DATOS
BIOMÉDICOS.**

Maira Alejandra Martínez Hernández



**UNIVERSIDAD DE CÓRDOBA
FACULTAD DE CIENCIAS BÁSICAS
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA
MONTERÍA - CÓRDOBA
2023**

METODOLOGÍA PARA MODELAR LA ASOCIACIÓN ENTRE DOS VARIABLES ALEATORIAS CONTINUAS A TRAVÉS DE FUNCIONES CÓPULA: APLICACIÓN EN DATOS BIOMÉDICOS.

Maira Alejandra Martínez Hernández

Trabajo presentado como requisito para optar al título de
ESTADÍSTICO

Director:

OSNAMIR ELIAS BRU CORDERO



**UNIVERSIDAD DE CÓRDOBA
FACULTAD DE CIENCIAS BÁSICAS
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA
MONTERÍA - CÓRDOBA
2023**

UNIVERSIDAD DE CÓRDOBA
FACULTAD DE CIENCIAS BÁSICAS
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA

Los jurados abajo firmantes certifican que han leído y recomiendan a la **Facultad de Ciencias Básicas** aprobar el trabajo de grado titulado: “**Metodología para modelar la asociación entre dos variables aleatorias continuas a través de funciones cópula: aplicación en datos biomédicos**”, el cual es presentado por **Maira Alejandra Martínez Hernández**, como requisito para optar al título de **Estadístico**.

Fecha: 2 de Octubre de 2023

Director: _____
Osnamir Elias Bru Cordero

Jurado: _____
Guillermo Domingo Martínez Flórez

Jurado: _____
Hugo Alberto Brango García

A **Dios**, autor y protagonista de mi vida.

A mis padres **Miguel Martínez Medina y Maira Luz Hernández Herrera**.

Mis hermanos **Karen Yaneth, Sarai, Elisabeh, Natalia y Jacobo Miguel**.

A mi esposo **Elkin Darío Quintero Morales**.

Índice general

Resumen	VII
Agradecimientos	IX
1. Introducción	1
2. Funciones cópula	4
2.1. Definición	4
2.2. Familias de cópulas	5
2.2.1. Cópulas elípticas	6
2.2.1.1. Cópula gaussiana	6
2.2.1.2. Cópula t-student	6
2.2.2. Cópulas arquimedianas	7
2.2.2.1. Cópula clayton	7
2.2.2.2. Cópula gumbel	8
2.2.2.3. Cópula frank	8
2.2.2.4. Cópula joe	9
2.2.3. Cópula de supervivencia	9
2.2.4. Cópulas de valor extremo	11
2.3. τ de Kendall	11
2.4. Métodos gráficos de bondad de ajuste	13
2.4.1. Método Chi-plot	13
2.4.2. Método K -plot	14

3. Modelos de validación cruzada	15
3.1. Criterio de información de Akaike (AIC)	17
3.2. Criterio de información bayesiano de Schwarz (BIC)	17
4. Esquemas de simulación y aplicación	18
4.1. Modelo con marginales weibull	22
4.2. Métodos gráficos para detectar dependencia	24
4.3. Aplicación de validación cruzada para selección de modelos cópulas	27
4.4. Aplicación en datos biomédicos	29
5. Conclusiones y recomendaciones	34
5.1. Conclusiones	34
5.2. Recomendaciones y futuras investigaciones	35
A. Código en R para el proceso de simulación	36
B. Marginales normal	39
C. Marginales log-normal	56
D. Marginales t-student	73
E. Marginales weibull	90
F. Simulación para datos biomédicos	107

Lista de Tablas

2.1. Propiedades de familias de cópulas elíticas bivariadas	12
2.2. Propiedades de familias de cópulas arquimedianas bivariadas	12
4.1. Nombres de cópulas bivariadas	19
4.2. Resumen de simulación para marginales normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$	20
4.3. Resumen de simulación para marginales log-normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$	21
4.4. Resumen de simulación para marginales t-student con parámetros $v_1 = v_2 = n - 1$	21
4.5. Resumen de simulación para marginales weibull con parámetros $\beta_1 = \beta_2 = 1$ y $\lambda_1 = \lambda_2 = 1$	22
4.6. Resumen de simulación para marginales weibull con parámetros $\beta_1 = \beta_2 = 2$ y $\lambda_1 = \lambda_2 = 1$	23
4.7. Resumen de simulación para marginales weibull con parámetros $\beta_1 = \beta_2 = 0.5$ y $\lambda_1 = \lambda_2 = 1$	23
4.8. AIC y BIC para cópulas ajustadas para marginales normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$	27
4.9. AIC y BIC para cópulas ajustadas para marginales log-normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$	27
4.10. AIC y BIC para cópulas ajustadas para marginales t-student con parámetros $v_1 = v_2 = n - 1$	28
4.11. AIC y BIC para cópulas ajustadas para marginales weibull con parámetros $\beta_1 = \beta_2 = 1$ y $\lambda_1 = \lambda_2 = 1$	28

4.12. AIC y BIC para cópulas ajustadas para marginales weibull con parámetros $\beta_1 =$ $\beta_2 = 2$ y $\lambda_1 = \lambda_2 = 1$	28
4.13. AIC y BIC para cópulas ajustadas para marginales weibull con parámetros $\beta_1 =$ $\beta_2 = 0.5$ y $\lambda_1 = \lambda_2 = 1$	29
4.14. Resumen de simulación para datos biomedicos	32
4.15. AIC y BIC para cópulas ajustadas para datos biomédicos	33
B.1. Pares de variables aleatorias marginales normal $n = 50, r = 20$	39
B.2. Pares de variables aleatorias marginales normal $n = 50, r = 40$	40
B.3. Pares de variables aleatorias marginales normal $n = 50, r = 100$	41
B.4. Pares de variables aleatorias marginales normal $n = 100, r = 20$	44
B.5. Pares de variables aleatorias marginales normal $n = 100, r = 40$	45
B.6. Pares de variables aleatorias marginales normal $n = 100, r = 100$	46
B.7. Pares de variables aleatorias marginales normal $n = 500, r = 20$	50
B.8. Pares de variables aleatorias marginales normal $n = 500, r = 40$	51
B.9. Pares de variables aleatorias marginales normal $n = 500, r = 100$	52
C.1. Pares de variables aleatorias marginales log-normal $n = 50, r = 20$	56
C.2. Pares de variables aleatorias marginales log-normal $n = 50, r = 40$	57
C.3. Pares de variables aleatorias marginales log-normal $n = 50, r = 100$	58
C.4. Pares de variables aleatorias marginales log-normal $n = 100, r = 20$	61
C.5. Pares de variables aleatorias marginales log-normal $n = 100, r = 40$	62
C.6. Pares de variables aleatorias marginales log-normal $n = 100, r = 100$	64
C.7. Pares de variables aleatorias marginales log-normal $n = 500, r = 20$	67
C.8. Pares de variables aleatorias marginales log-normal $n = 500, r = 40$	68
C.9. Pares de variables aleatorias marginales log-normal $n = 500, r = 100$	69
D.1. Pares de variables aleatorias marginales t-student $n = 50, r = 20$	73
D.2. Pares de variables aleatorias marginales t-studen $n = 50, r = 40$	74
D.3. Pares de variables aleatorias marginales t-studen $n = 50, r = 100$	75

D.4. Pares de variables aleatorias marginales t-studen $n = 100, r = 20$	78
D.5. Pares de variables aleatorias marginales t-studen $n = 100, r = 40$	79
D.6. Pares de variables aleatorias marginales t-studen $n = 100, r = 100$	81
D.7. Pares de variables aleatorias marginales t-studen $n = 500, r = 20$	84
D.8. Pares de variables aleatorias marginales t-studen $n = 500, r = 40$	85
D.9. Pares de variables aleatorias marginales t-studen $n = 500, r = 100$	86
E.1. Pares de variables aleatorias marginales weibull $n = 50, r = 20$	90
E.2. Pares de variables aleatorias marginales weibull $n = 50, r = 40$	91
E.3. Pares de variables aleatorias marginales weibull $n = 50, r = 100$	92
E.4. Pares de variables aleatorias marginales weibull $n = 100, r = 20$	95
E.5. Pares de variables aleatorias marginales weibull $n = 100, r = 40$	96
E.6. Pares de variables aleatorias marginales weibull $n = 100, r = 100$	98
E.7. Pares de variables aleatorias marginales weibull $n = 500, r = 20$	101
E.8. Pares de variables aleatorias marginales weibull $n = 500, r = 40$	102
E.9. Pares de variables aleatorias marginales weibull $n = 500, r = 100$	103
F.1. aplicación datos bimédicos	107

Lista de Figuras

4.1. Métodos gráficos para la cópula gaussina con $\rho = 0.88$, para marginales normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$	24
4.2. Métodos gráficos para la cópula frank con $\theta = 4.438$, para marginales log-normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$	24
4.3. Métodos gráficos para la cópula gaussina con $\rho = 0.87$, para marginales t-student con parámetros $v_1 = v_2 = n - 1$	25
4.4. Métodos gráficos para la cópula frank con $\theta = 4.0136$, para marginales weibull con parámetros $\beta_1 = \beta_2 = 1$ y $\lambda_1 = \lambda_2 = 1$	25
4.5. Métodos gráficos para la cópula frank con $\theta = 3.3525$, para marginales weibull con parámetros $\beta_1 = \beta_2 = 2$ y $\lambda_1 = \lambda_2 = 1$	26
4.6. Métodos gráficos para la cópula survival BB7 con $\theta_1 = 1.61$ y $\theta_2 = 4.48$, para marginales weibull con parámetros $\beta_1 = \beta_2 = 0.5$ y $\lambda_1 = \lambda_2 = 1$	26
4.7. Gráfico de dispersión entre tiempo oficio y micronúcleo	30
4.8. Histograma y densidad para tiempo oficio	31
4.9. Histograma y densidad para micronúcleo	31
4.10. Métodos gráficos para la cópula joe rotada 90 grados para datos biomédicos	32

Resumen

La estimación de la fuerza de la dependencia entre dos variables aleatorias es importante en el análisis de los datos. Un método para hallar dicha asociación son las funciones cópulas, representada como una forma paramétrica conveniente para modelar la estructura de dependencia en distribuciones conjuntas de variables aleatorias. Existe una amplia gama de funciones cópulas, por lo que la elección de una cópula adecuada es uno de los grandes retos al que se enfrenta el investigador. Con este trabajo se busca proponer un mecanismo de selección de cópula usando pruebas analíticas y gráficas, entre las pruebas gráficas se encuentran los gráficos *Chi-Plot* y *K-Plot*. Entre sus resultados relevantes se encontró que para marginales normal, la mejor cópula es gaussiana con parámetro $\rho = 0.88$, para marginales log-normal, la cópula que mejor ajustó fue una frank con un parámetro de 4.438, para las marginales t-student, la cópula que mejor ajustó fue una gaussiana con parámetro $\rho = 0.87$ y para para marginales weibull, la copula que mejor ajustó fue frank. Haciendo uso de esta metodología, fue posible encontrar la asociación de dos variables aleatorias, por medio de las funciones cópula.

Palabras Claves: Función Cópula, Dependencia, Datos Bivariados, τ de Kendall, Métodos Gráficos.

Abstract

Estimating the strength of the dependence between two random variables is important in data analysis. One method for finding such an association is copula functions, represented as a convenient parametric form for modeling the dependence structure in joint distributions of random variables. There is a wide range of copula functions, so the choice of a suitable copula is one of the great challenges facing the researcher. With this work we seek to propose a copula selection mechanism using analytical and graphical tests, among the graphical tests are the *Chi-Plot* and *K-Plot* graphs. Among its relevant results it was found that for normal marginals, the best copula is Gaussian with parameter $\rho = 0.88$, for log-normal marginals, the best fitting copula was a frank with a parameter of 4.438, for t-student marginals, the best fitting copula was a Gaussian with parameter $\rho = 0.87$ and for weibull marginals, the best fitting copula was frank. Using this methodology, it was possible to find the association of two random variables by means of the copula functions.

Keywords: Copula Function, Dependence, Bivariate Data, Kendall's τ , Graphical Methods.

Agradecimientos

Agradezco, en primer lugar, a Dios, Él me ha permitido llegar hasta este punto y como dice Romanos 9:16 “Así que no depende del que quiere, ni del que corre, sino de Dios que tiene misericordia”.

A mi familia, padres y hermanos, por brindarme su apoyo incondicional. A mi esposo Elkin Darío Quintero Morales, por apoyarme en lo último de mi proceso.

A mi director, Osnamir Elías Bru Cordero, por compartir su conocimiento conmigo y dirigirme en el último paso para adquirir mi título como Estadístico.

A los profesores de estadística en la Universidad de Córdoba, por impartirme su saber. A mis compañeros Jhon Jeiller Hoyos Vega, Fabiana Ballesteros López, Luis Miguel Buelvas Pestana, Eduardo Javier Hernández Martínez y Elkin David Velásquez Román, por estar conmigo desde el principio de la carrera y tenerme paciencia.

A la profesora Erika María López López, por ser la persona que me instruyó este camino. A Tulia Ester Cordero de Bru, por acogerme en su hogar el tiempo de estudio. A todas esas personas que sin proponérselo fueron parte de este proceso.

Capítulo 1

Introducción

Obtener la asociación entre variables se ha vuelto de gran interés para el análisis de datos, ya que facilita la inferencia. En los últimos tiempos surge con mucha fuerza una metodología llamada cópula. En el contexto de la teoría de probabilidad las funciones llamadas cópulas son en realidad funciones de distribución que representan las relaciones de dependencia entre las variables aleatorias. La cópula representa una forma paramétrica conveniente para modelar la estructura de dependencia en distribuciones conjuntas de variables aleatorias, vale la pena resaltar que no es la única técnica usada para ese objetivo, pero en nuestro trabajo hacemos uso sólo de dicha técnica (Nelsen, 2006).

Las cópulas son funciones de distribución multivariadas cuyas marginales unidimensionales son uniformes en el intervalo $(0, 1)$. Las cópulas representarán una aproximación muy útil a la comprensión y la modelización de las variables aleatorias cuya dependencia se desea estudiar, dado que nos permiten centrarnos explícitamente en la estructura de dependencia entre ellas (Ortego, 2014). Las funciones cópula son de hecho funciones que justifican la correspondencia entre un valor dado de la función de distribución conjunta y pares ordenados de las funciones de distribución marginales. Convirtiéndose esta en una seria alternativa para modelar la dependencia de dos variables aleatorias (Sepúlveda and Garzón, 2017).

Existe una amplia gama de cópulas. En términos generales, las cópulas pueden agruparse como cópulas elípticas, cópulas arquimedianas, cópulas de valor extremo, entre otras. Las cópulas también pueden ser agrupadas por paramétricas y no paramétricas (Gómez, 2017).

Las cópulas son funciones bivariadas que juntan o “copulan” dos funciones de distribución univariadas para construir funciones de distribución bivariadas continuas. Las cópulas pueden extraer la estructura de dependencia de la función de distribución conjunta y, al mismo tiempo, permiten separar la estructura de dependencia del comportamiento marginal (Morales, 2013).

Por otro lado, la validación cruzada hace uso de distintos subconjuntos de los datos disponibles para realizar el entrenamiento del modelo y su posterior validación (Arlot and Celisse, 2010). Esta técnica se utiliza para detectar el sobreajuste, es decir, cuando no se logra encontrar un patrón en el modelo Services (2022). Teniendo en cuenta, si el modelo funciona bien con los datos de prueba y brinda una buena precisión, se supone que el modelo no se sobreajustó a los datos de entrenamiento y puede usarse para la predicción (Datapeaker, 2022).

En la práctica, no existe una forma estándar de seleccionar una cópula entre los diversos modelos posibles, por lo que elegir la cópula adecuada es uno de los mayores desafíos para el investigador, por tanto el objetivo de este estudio es proponer un mecanismo de selección de cópula usando pruebas analíticas y gráficas, y mediante la técnica de validación cruzada. Usando el parámetro de la cópula para obtener el grado de asociación τ de Kendall y comparar las funciones copulas ajustadas a través de las métricas de medición AIC y BIC. Entre las pruebas gráficas se encuentran los gráficos *Chi-Plot* y *K-Plot*, ambas metodologías son una herramienta útil para estudiar la dependencia entre dos variables aleatorias y como método analítico usaremos el método *K-Fold* de validación cruzada tomando como métricas de medición el criterio de información de Akaike (AIC) y criterio de información bayesiano (BIC).

Sabemos que en muchos casos no existe dependencia lineal entre dos variables, sin embargo, no podemos descartar la idea de que estas estén asociadas, teniendo en cuenta que el coeficiente de asociación mide la fuerza de una relación, lo cual significa que las variables tienen elementos comunes o algún grado de concordancia, por tanto, a través de la función cópula mostramos una alternativa para encontrar dicha asociación.

Considerando que los datos biomédicos permiten, entre otras cosas, analizar la efectividad o efectos secundarios de un fármaco o tratamiento, identificar proactivamente la necesidad de atención especializada del paciente y obtener patrones de comportamiento para pacientes y profesionales de la salud, en esta investigación se realizó una aplicación de cópula con datos

biomédicos, donde se busca modelar la asociación de dos variables aleatorias con dicha característica.

El capítulo 2 define los conceptos básicos utilizados en la teoría de cópulas, como lo son; las funciones cópula y sus características más importantes, el parámetro de dependencia τ de Kendall y algunos métodos gráficos para detectar dependencia utilizados en este trabajo. El capítulo 3 explica el concepto de validación cruzada, algunos métodos usados en esta técnica de remuestreo y la definición de algunas métricas de medición.

El capítulo 4 muestra los resultados obtenidos vía simulación de los métodos expuestos en los capítulos 2 y 3 usando el software estadístico R. Finalmente el capítulo 5 presenta las conclusiones del estudio y trabajos futuros.

Capítulo 2

Funciones cópula

En este capítulo ilustraremos todo lo referente a la metodología de las cópulas, para tratar de caracterizar un fenómeno en particular, el cual es modelado por la familia gaussiana o familia arquimediana. Inicialmente presentaremos la definición formal de una cópula.

2.1. Definición

Una cópula, C , es una función de distribución multivariante cuyas distribuciones marginales se distribuyen uniforme entre $[0, 1]$. En el caso bivalente, $C(u; v) = P[U \leq u; V \leq v]$ es una función definida en $[0, 1]^2 \implies [0, 1]$ que verifica las siguientes tres propiedades (Gallardo, 2010):

- $C(u; v)$ es una función creciente para cada una de sus componentes.
- $C(u; 1) = u$ y $C(1; v) = v$.
- $\forall a_1 \leq a_2$ y $\forall b_1 \leq b_2$, entonces $C(a_1; b_1) + C(a_2; b_2) - C(a_1; b_2) - C(a_2; b_1) \geq 0$.

Teorema de Sklar: Sea H una función de distribución n -dimensional con marginales F_1, \dots, F_n . Entonces, existe una cópula n -dimensional C tal que $\forall X \in \bar{R}^n$;

$$H(x_1 \dots x_n) = C(F_1(x_1), \dots, F_n(x_n)),$$

si F_1, \dots, F_n son todas continuas, entonces C es única, por tanto, está unívocamente determinado en $\text{ran}(F_1) \times \dots \times \text{ran}(F_n)$.

Cuando escribimos,

$$F(x; y) = C(F_X(x); F_Y(y)),$$

repartimos la probabilidad conjunta entre las marginales y una cópula, de forma que esta última solamente representa la asociación entre X e Y . Las cópulas separan el comportamiento marginal (representado por las F_i) del conjunto, en contra de lo que ocurre en la representación usual de probabilidades conjuntas vía función de distribución. Por esta razón, las cópulas son denominadas funciones de dependencia. Si las distribuciones marginales son continuas, la cópula es única. Por tanto, a partir de las cópulas, es posible crear distribuciones bivariantes con distribuciones marginales definidas de esta forma, si C es una cópula (Gallardo, 2010).

2.2. Familias de cópulas

Existen muchos tipos de funciones cópula y es difícil encontrar en la literatura una clasificación clara de todas ellas dado que existen diversos criterios para hacerlo: en función de la dependencia y del tipo de relación que reflejan (cópulas elípticas, cópulas arquimedianas, cópulas de valor extremo, entre otras). Por ello, en vez de presentar un esquema general que permita ubicar cada cópula de acuerdo a una jerarquía concreta, enumeraremos algunos de estos criterios y citaremos algunos ejemplos asociados a las clases que resultan de su aplicación (Gallardo, 2010).

Existen cópulas en función del conocimiento explícito de su forma, ya que las cópulas se pueden clasificar en función de que su expresión responda o no a una ecuación paramétrica, pudiendo distinguir entre:

- **Cópulas paramétricas:** todas las cópulas que responden a una misma ecuación paramétrica definen una familia de cópulas. El parámetro (uniparamétricas) o parámetros (multiparamétricas) cuantifican la relación de dependencia entre las variables que asocian.
- **Cópulas no paramétricas:** de igual manera existen familias de cópulas no paramétricas, en cuya definición no participa ningún parámetro sino que, por su estructura empírica, se ajustan de forma local a los datos.

A continuación, mostraremos los tipos de cópulas en función de la relación de dependencia.

2.2.1. Cópulas elípticas

Se definen como las cópulas asociadas a las distribuciones elípticas. Su principal característica es la forma como representan relaciones de dependencia simétricas, sin importar que se tomen en cuenta las colas de las distribuciones.

Las cópulas normales y de t-student son cópulas elípticas, son simétricas y su implementación es sencilla, ya que se conocen bien las distribuciones que están asociadas.

Las cópula elípticas tienen la forma:

$$C_{\rho}(u_1, u_2) = \frac{1}{\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi_{g,1}^{-1}(u_1)} \int_{-\infty}^{\Phi_{g,2}^{-1}(u_2)} g\left(\frac{x^2 - 2\rho xy + y^2}{\sqrt{1-\rho^2}}\right) dx dy = H_{\rho}\left(\Phi_{g,1}^{-1}(u_1), \Phi_{g,2}^{-1}(u_2)\right).$$

La función H es la distribución conjunta de variables X e Y , con $\Phi_{g,1}^{-1}(u_1)$, $\Phi_{g,2}^{-1}(u_2)$ funciones cuantiles respectivamente y ρ su coeficiente de correlación.

Entre las cópulas elípticas tenemos: gaussiana y t-studen.

2.2.1.1. Cópula gaussiana

La cópula gaussiana o normal es la función de dependencia asociada a la distribución normal multivariada.

Sea ρ una matriz diagonal definida positiva con $\text{diag}(\rho) = 1$ y Θ_{ρ} la distribución normal bivariada *standard*. La cópula normal se define de la siguiente forma:

$$C(u_1, u_2; \rho) = \Phi_{\rho}\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right).$$

La densidad de la cópula normal se escribe:

$$c(u_1, u_2; \rho) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{2(1-\rho^2)} + \frac{x_1^2 + x_2^2}{2}\right),$$

con $x_1 = \Phi^{-1}(u_1)$ y $x_2 = \Phi^{-1}(u_2)$ esta relación se muestra utilizando la expresión de la densidad de la distribucion normal Φ bivariada:

$$\varphi_{\rho}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{2(1-\rho^2)}\right).$$

2.2.1.2. Cópula t-student

Es la función de dependencia asociada a la distribución t multivariada. Con $t_{\rho,v}$ la distribución t-student bivariada con v grados de libertad y matriz de correlación ρ . La cópula t se define de

la siguiente forma:

$$C(u_1, u_2; \rho) = \mathbf{t}_{\rho, v} \left(t_v^{-1}(u_1), t_v^{-1}(u_2) \right).$$

La densidad de la copula t está dada por:

$$c(u_1, u_2; \rho, v) = \frac{v}{2\sqrt{1-\rho^2}} \cdot \frac{\Gamma\left(\frac{v}{2}\right)^2}{\Gamma\left(\frac{v+1}{2}\right)^2} \cdot \frac{\left(1 + \frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{v(1-\rho^2)}\right) - \frac{v+2}{2}}{\left[\left(1 + \frac{x_1^2}{v}\right)\left(1 + \frac{x_2^2}{v}\right)\right] - \frac{v+1}{2}},$$

donde $x_1 = t_v^{-1}(u_1)$ y $x_2 = t_v^{-1}(u_2)$, para mostrar ésta relación utilizamos la expresión de la distribución t bivariada:

$$t_{\rho, v}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \left[\frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{y_1^2 + y_2^2 - 2\rho y_1 y_2}{v(1-\rho^2)} \right) - \frac{v+2}{2} \right] dy_1 dy_2.$$

2.2.2. Cópulas arquimedianas

Existe una gran diversidad de familias que pertenecen a la clase arquimediana y gracias a esta variedad permiten recoger muchos tipos de estructuras de dependencia adicionales y son fáciles de construir. Se han creado muchos tipos de cópulas que forman parte de la familia de cópulas arquimedianas entre los que se encuentran las cópulas de frank, clayton y gumbel, entre otras (Nelsen, 2006).

2.2.2.1. Cópula clayton

La función de distribución para la cópula clayton es:

$$C_{\theta}(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 1 \right)^{-\frac{1}{\theta}},$$

donde $\theta > 0$.

La función de densidad es:

$$c(u_1, u_2) = (1 + \theta) (u_1 \cdot u_2)^{-1-\theta} \left[u_1^{-\theta} + u_2^{-\theta} - 1 \right]^{-\frac{1}{\theta}-2}.$$

El coeficiente de correlación de Kendall, en función de su parámetro θ , se define como:

$$\tau(\theta) = \frac{\theta}{\theta + 2},$$

cuando θ tiende a ∞ , se presenta independencia.

2.2.2.2. Cópula gumbel

La función de distribución para la cópula gumbel es:

$$C_{\theta}(u_1, u_2) = \exp \left(- \left[(-\ln u_1)^{\theta} + (-\ln u_2)^{\theta} \right]^{\frac{1}{\theta}} \right)$$

donde $\theta \geq 0$.

La función de densidad es:

$$c_{\theta}(u_1, u_2) = \frac{C_{\theta}(u_1, u_2)}{u_1 u_2} \frac{[\ln u_1 \cdot \ln u_2]^{\theta-1}}{[(-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}]^{2-\frac{1}{\theta}}} \left[[(-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}]^{-\frac{1}{\theta}} + \theta - 1 \right].$$

El coeficiente de correlación de Kendall, en función de su parámetro θ , se define como:

$$\tau(\theta) = 1 - \frac{1}{\theta},$$

cuando θ tiende a 1, se presenta independencia.

2.2.2.3. Cópula frank

La función de distribución para la cópula frank es:

$$C_{\theta}(u_1, u_2) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right),$$

donde $\theta \in \mathbb{R} \setminus \{0\}$.

Si hacemos $g_{u_i} = e^{-\theta u_i} - 1$ y $g_1 = e^{-\theta} - 1$, la derivada de la cópula respecto de la componente u_1 resulta:

$$c = \left(\frac{\partial C_{\theta}(u_1, u_2)}{\partial u_1} \right) = \frac{g_{u_1} \cdot g_{u_2} + g_{u_2}}{g_{u_1} \cdot g_{u_2} + g_1^2}$$

Y la función de densidad:

$$c(u_1, u_2) = -\theta g_1 \frac{e^{-\theta(u_1+u_2)}}{(g_1 + g_{u_1} \cdot g_{u_2})^2}$$

$$= \frac{-\theta (e^{-\theta} - 1) e^{-\theta(u_1+u_2)}}{((e^{-\theta} - 1) + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1))^2}.$$

Para evaluar el grado de asociación entre las marginales en el modelo generado por la cópula de frank, el coeficiente de correlación de Kendall correspondiente está dada por:

$$\tau(\theta) = 1 - \frac{4}{\theta} + \frac{4}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt.$$

2.2.2.4. Cópula joe

La función de distribución para la cópula joe es:

$$C_\theta(u_1, u_2) = 1 - \left[(1 - u_1)^\theta + (1 - u_2)^\theta - (1 - u_1)^\theta (1 - u_2)^\theta \right]^{\frac{1}{\theta}},$$

donde $\theta > 0$.

La densidad de la copula joe está dada por:

$$\begin{aligned} c(u_1, u_2) &= \left[(1 - u_1)^\theta + (1 - u_2)^\theta - (1 - u_1)^\theta (1 - u_2)^\theta \right]^{\frac{1}{\theta} - 2} \\ &\quad \times (1 - u_1)^{\theta - 1} (1 - u_2)^{\theta - 1} \\ &\quad \times \left[\theta - 1 + (1 - u_1)^\theta + (1 - u_2)^\theta - (1 - u_1)^\theta (1 - u_2)^\theta \right]. \end{aligned}$$

El coeficiente de correlación de Kendall, en función de su parámetro θ , se define como:

$$\tau(\theta) = 1 + \frac{4}{\theta^2} \int_0^1 t \log(t) (1 - t)^{\frac{2(1-\theta)}{\theta}} dt.$$

2.2.3. Cópula de supervivencia

En muchas aplicaciones, las variables aleatorias de interés representan la vida de individuos u objetos en alguna población. La probabilidad de que un individuo viva o sobreviva más allá del tiempo x está dada por la función de supervivencia $\bar{F}(x) = P[X > x] = 1 - F(x)$, donde F denota la función de distribución de X . Cuando se trata de tiempos de vida, el rango natural de la variable suele ser $[0, \infty)$; sin embargo, utilizaremos el término “función de supervivencia”, incluso cuando el rango es \bar{R} .

Para un par (X, Y) de variables aleatorias con función de distribución conjunta H , la función de supervivencia conjunta viene dada por $\bar{H}(x, y) = P[X > x, Y > y]$. las marginales de \bar{H}

son las funciones $\bar{H}(x, \infty)$ y $\bar{H}(y, \infty)$, las cuales son las funciones de supervivencia univariadas \bar{F} y \bar{G} , respectivamente. Una pregunta natural es la siguiente: ¿Existe una relación entre las funciones de supervivencia univariadas y conjuntas análoga a la que existe entre las funciones de distribución univariadas y conjuntas, tal como se expresa en el teorema de Sklar? Para responder a esta pregunta, suponga que la cópula de X e Y es C . Entonces tenemos siguiente:

$$\begin{aligned}\bar{H}(x, y) &= 1 - F(x) - G(y) + H(x, y) \\ &= \bar{F}(x) - \bar{G}(y) - 1 + C(F(x), G(y)) \\ &= \bar{F}(x) - \bar{G}(y) - 1 + C(1 - \bar{F}(x), 1 - \bar{G}(y)),\end{aligned}$$

de modo que si definimos una función \hat{C} de I^2 en I por,

$$\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v),$$

tenemos

$$\bar{H}(x, y) = \hat{C}(\bar{F}(x), \bar{G}(y)),$$

observe que \hat{C} copula la función de supervivencia conjunta a sus marginales univariadas de una manera completamente análoga a la forma en que una cópula conecta la función de distribución conjunta hasta sus marginales (Nelsen, 2006).

Haciendo la aplicación para la cópula ya definida, clayton, tenemos: *survival cópula*, de la cópula clayton.

La expresión de la función de distribución para la cópula survival clayton es:

$$C_{\theta}(u_1, u_2) = u_1 + u_2 - 1 \left[(1 - u_1)^{-\frac{1}{\theta}} + (1 - u_2)^{-\frac{1}{\theta}} - 1 \right]^{-\theta}.$$

La función de densidad de la cópula es:

$$c(u_1, u_2) = \left(1 + \frac{1}{\theta}\right) \left[(1 - u_1)^{-\frac{1}{\theta}} + (1 - u_2)^{-\frac{1}{\theta}} - 1 \right]^{-\theta-2} ((1 - u_1)(1 - u_2))^{-1-\frac{1}{\theta}}.$$

El coeficiente de correlación de Kendall, en función de su parámetro θ , es el mismo que para la cópula de clayton.

2.2.4. Cópulas de valor extremo

Las cópulas de valor extremo son de gran utilidad para representar relaciones que tienen mayor énfasis entre los sucesos "cola" (extremos) de las distribuciones marginales. Las cópulas de valor extremo son los posibles límites (en caso de que existan) de cópulas asociadas a los máximos de muestras independientes e idénticamente distribuidas. Sea una muestra de variables aleatorias bidimensionales $(X_1, Y_1), \dots, (X_n, Y_n)$ independientes e idénticamente distribuidas de acuerdo a unas mismas marginales F_X y G_Y y a una misma distribución conjunta H_{XY} que, en virtud del teorema de Sklar lleva asociada una cópula C :

$$H_{XY}(x, y) = C(F_X(x), G_Y(y)).$$

Sean las variables $M_n = \max\{X_1, X_2, \dots, X_n\}$ y $N_n = \max\{Y_1, Y_2, \dots, Y_n\}$ cuyas funciones de distribución vienen dadas por $F^n(x) = P[M_n \leq x]$ y $G^n(y) = P[N_n \leq y]$ y con distribución conjunta $H_n(x) = P[M_n \leq x, N_n \leq y]$.

Si C es la cópula asociada al par (M_n, N_n) y a su posible límite cuando " n " tiende a infinito, se dice entonces que C es una cópula de valor extremo (CVE).

Un coeficiente de asociación mide la fuerza de una relación, lo cual significa que las variables tienen elementos comunes o algún grado de concordancia. Estos coeficientes suelen oscilar entre 0 y 1, donde 0 es sin relación y 1 es una relación perfecta. Sin embargo, algunas medidas de asociación varían de -1 a 1, donde -1 indica una relación inversa perfecta. En este trabajo utilizaremos el coeficiente de asociación τ de Kendall, el cual es utilizado para variables nominales.

2.3. τ de Kendall

La versión muestral de la medida de asociación conocida como τ de Kendall está definida en términos de concordancia, como sigue: Sea $(x_1, y_1), \dots, (x_n, y_n)$ una muestra aleatoria de n observaciones de un vector (X, Y) de variables aleatorias continuas. Hay $\binom{n}{2}$ pares distintos (x_i, y_i) y (x_j, y_j) de observaciones en la muestra, y cada par es concordante o discordante. Sea c el número de pares concordantes y d el número de pares discordantes. Entonces el τ de Kendall para la muestra está definido como:

$$\tau = \frac{c - d}{c + d} = \frac{c - d}{\binom{n}{2}}.$$

De manera equivalente, τ es la probabilidad de concordancia menos la probabilidad de discordancia para pares de observaciones (x_i, y_i) y (x_j, y_j) que es seleccionada aleatoriamente de la muestra. Otra versión del τ de Kendall para un vector aleatorio continuo (X, Y) con función de distribución conjunta H está definida similarmente. Sea (X_1, Y_1) y (X_2, Y_2) vectores aleatorios independientes e idénticamente distribuidos cada uno con función de distribución conjunta H . Entonces, la definición formal del τ de Kendall es la probabilidad de concordancia menos la probabilidad de discordancia.

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

Para la cópula gaussiana, la correlación de Pearson es el parámetro de la cópula. El coeficiente de correlación de Spearman y el τ de Kendall están dados por: $\rho_S = \frac{6}{\pi} \sin^{-1}(\frac{1}{2\rho})$ y $\tau = \frac{2}{\pi} \sin^{-1}(\rho)$ respectivamente.

Tabla 2.1: Propiedades de familias de cópulas elíticas bivariadas

Familia	Nombre	Rango del parametro	τ de Kendall	Dependencia de la cola
1	gaussiana	$\rho \in (-1, 1)$	$\frac{2}{\pi} \sin^{-1}(\rho)$	0
2	t-studen	$\rho \in (-1, 1), v > 2$	$\frac{2}{\pi} \sin^{-1}(\rho)$	$2t_{v+1} \left(-\sqrt{v+1} \sqrt{\frac{1-\rho}{1+\rho}} \right)$

Tabla 2.2: Propiedades de familias de cópulas arquimedianas bivariadas

Familia	Nombre	Generador de funciones	Rango del parametro	τ de Kendall	Dependencia de la cola
3	clayton	$\frac{1}{\theta} (t^{-\theta} - 1)$	$\theta > 0$	$\frac{\theta}{\theta+2}$	$(2^{-\frac{1}{\theta}}, 0)$
4	gumbel	$(-\log t)^\theta$	$\theta \geq 0$	$1 - \frac{1}{\theta}$	$(0, 2 - 2^{\frac{1}{\theta}})$
5	frank	$-\log \left[\frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right]$	$\theta \in \mathbb{R} \setminus \{0\}$	$1 - \frac{4}{\theta} + 4 \frac{D_1(\theta)}{\theta}$	(0,0)
6	joe	$-\log [1 - (1-t)^\theta]$	$\theta > 1$	$1 + \frac{4}{\theta^2} \int_0^1 t \log(t) (1-t)^{\frac{2(1-\theta)}{\theta}} dt$	$(0, 2 - 2^{\frac{1}{\theta}})$
7	BB1	$(t^{-\theta} - 1)^\delta$	$\theta > 0, \delta \geq 1$	$1 - \frac{2}{\delta(\theta+2)}$	$(2^{-\frac{1}{\theta\delta}}, 2 - 2^{\frac{1}{\theta\delta}})$
8	BB6	$(-\log [1 - (1-t)^\theta])^\delta$	$\theta \geq 1, \delta \geq 1$	$1 + \frac{4}{\delta\theta} \int_0^1 (-\log (1 - (1-t)^\theta)) x(1-t) (1 - (1-t)^{-\theta}) dt$	$(0, 2 - 2^{\frac{1}{\theta\delta}})$
9	BB7	$(1 - (1-t)^\theta)^{-\delta} - 1$	$\theta \geq 1, \delta > 0$	$1 + \frac{4}{\delta\theta} \int_0^1 \left(- (1 - (1-t)^\theta)^{\delta+1} x \left[\frac{(1-(1-t)^\theta)^{-\delta} - 1}{(1-t)^{\theta-1}} \right] dt \right)$	$(2^{-\frac{1}{\delta}}, 2 - 2^{\frac{1}{\theta}})$
10	BB8	$-\log \left[\frac{1-(1-\delta t)^\theta}{1-(1-\delta)^\theta} \right]$	$\theta \geq 1, \delta \in (0, 1]$	$1 + \frac{4}{\delta\theta} \int_0^1 (-\log \left(\frac{(1-t\delta)^\theta - 1}{(1-t)^\theta - 1} \right)) x(1-t\delta) (1 - (1-t\delta)^{-\theta}) dt$	(0,0)

donde $D_1(\theta) = \int_0^\theta \frac{t}{e^t - 1} dt$.

En las Tablas 2.1 y 2.2 se muestran las cópulas elípticas y arquimedianas, con sus respectivas características.

Propiedad de invarianza del τ de Kendall: Sean (X_1, Y_1) y (X_2, Y_2) dos variables aleatorias bivariadas independientes, cada una con la distribución bivariada común de (X, Y) , y sean g y h dos funciones reales crecientes, entonces $\tau[g(X), h(Y)] = \tau(X, Y)$.

2.4. Métodos gráficos de bondad de ajuste

2.4.1. Método Chi-plot

El gráfico Chi-Plot fue propuesto originalmente por Fisher y Switzer (Fisher and Switzer, 1985). Su construcción está basada en el estadístico Chi-cuadrado para independencia. Sea $(X_1, Y_1), \dots, (X_n, Y_n)$ una muestra aleatoria de una función de distribución conjunta y continua H de la variable aleatoria (X, Y) , y sea $I(A)$ la función característica del suceso A . Para cada observación (x_i, y_i) se realiza el siguiente procedimiento (Moreno, 2012)

$$H_i = \frac{1}{n-1} \sum_{j \neq i} I(X_j \leq X_i, Y_j \leq Y_i),$$

$$F_i = \frac{1}{n-1} \sum_{j \neq i} I(X_j \leq X_i)$$

y

$$G_i = \frac{1}{n-1} \sum_{j \neq i} I(Y_j \leq Y_i).$$

Ninguna de estas cantidades dependen exclusivamente de los rangos de las observaciones, (Fisher and Switzer, 1985) proponen graficar los pares (λ_i, X_i) , donde:

$$X_i = \frac{H_i - F_i G_i}{\sqrt{F_i(1 - F_i)G_i(1 - G_i)}}$$

y

$$\lambda_i = 4 \text{sign}(\tilde{F}_i, \tilde{G}_i) \max(\tilde{F}_i^2, \tilde{G}_i^2)$$

donde $\tilde{F}_i = F_i - \frac{1}{2}$, $\tilde{G}_i = G_i - \frac{1}{2}$ para $i \in 1, \dots, n$.

El Chi-Plot es un diagrama de dispersión de los pares (λ_i, X_i) , $i = 1, \dots, n$ donde λ_i es una medida de distancia de la observación (X_i, Y_i) al centro de los datos (Moreno, 2012).

Según Cintas (2007), todos los valores de λ_i deben estar en el intervalo $[-1, 1]$. En caso de que los datos constituyan una muestra bivariada con marginales continuas independientes, los valores de λ_i se distribuyen uniformemente. Sin embargo, si X e Y están asociadas, los valores de λ_i se presentarán formando agrupaciones, en particular, valores positivos de λ_i indican que X_i e Y_i son relativamente grandes o relativamente pequeñas respecto a sus medianas, mientras que λ_i negativos corresponden a X_i e Y_i situados en lados opuestos respecto a sus medianas (Cintas, 2007).

2.4.2. Método K -plot

Los K -plot (Kendall-plot) fueron creados por (Genest and Boies, 2003), esta herramienta también se construye sobre los rangos de las observaciones utilizando la transformación integral de probabilidades multivariantes, produciendo un gráfico similar al QQ-plot convencional (Moreno, 2012).

Sea $(X_1, Y_1), \dots, (X_n, Y_n)$ una muestra aleatoria de una función de distribución conjunta y continua H de la variable aleatoria (X, Y) , para construir el gráfico K -Plot se procede de la siguiente manera:

1. Para cada $1 \leq i \leq n$ se calcula H_i (Como en el gráfico Chi-Plot).
2. Se ordenan los valores de H_i de tal forma que $H_{(1)} \leq \dots \leq H_{(n)}$.
3. Se representan los pares $(W_{i:n}, H_i)$, donde $W_{i:n}$ representa la esperanza del estadístico de orden i -ésimo en una muestra de tamaño n , calculando de la siguiente manera:

$$W_{i:n} = n \binom{n-1}{i-1} \int_0^1 w [K_0(w)]^{i-1} [1 - K_0(w)]^{n-i} dK_0(w)$$

con

$$K_0(w) = w - w \log(w) \quad 0 \leq w \leq 1$$

A medida que la representación $H_{(i)}$ contra $W_{i:n}$ se desvía con respecto a la diagonal, se puede asumir dependencia estadística entre las dos variables involucradas.

Capítulo 3

Modelos de validación cruzada

La validación cruzada es una técnica de remuestreo con una iniciativa importante para dividir el conjunto de datos en dos partes: un subconjunto de datos de entrenamiento y un subconjunto de datos de prueba. El primer subconjunto se usan para modelar y el segundo subconjunto de datos de prueba invisibles se usan para pronosticar. Si el modelo funciona bien con el subconjunto de datos de prueba y brinda una buena precisión, se supone que el modelo no se sobreajustó a los datos de entrenamiento y puede usarse para la predicción (Datapeaker, 2022) y (Pérez-Planells et al., 2015)

.

Así mismo, el número de observaciones disponibles n , se divide en dos subgrupos (subconjuntos), de modo que $n = n_c + n_v$, donde n_c es el número de observaciones utilizadas para la construcción del modelo y n_v es el número de observaciones utilizadas para la validación del modelo o el cálculo de los residuos de predicción (error de predicción) (Kozak and Kozak, 2003).

Entre los métodos de validación cruzada, se tienen; *Hold-Out*, *K-Fold* y *Leave One Out*. Inicialmente, *Hold-out* es el método de evaluación más simple y se usa ampliamente en proyectos de aprendizaje automático. Aquí, el conjunto de datos completo (población) se divide en 2 subconjuntos: conjunto de entrenamiento y conjunto de prueba. Los datos se pueden dividir en 70 – 30 o 60 – 40, 75 – 25 o 80 – 20, o incluso 50 – 50 según el caso de uso. Como regla general, la proporción de datos de entrenamiento debe ser mayor que los datos de prueba (Datapeaker, 2022)

.

Por otro lado, el otro método de validación cruzada *K-Fold*, asume que el conjunto de datos

está dividido por un número K . En la terminología de la validación cruzada, cuando se hace referencia a 5 veces o $K = 5$. En este escenario, el procedimiento divide el conjunto de datos en cinco subconjuntos. Se prueba el modelo utilizando el primer subconjunto en la primera iteración y se ejecuta con el subconjunto de datos restante. El segundo subconjunto es útil para probar el conjunto de datos y el otro apoya el proceso de entrenamiento. El mismo proceso se repite hasta que el grupo de prueba usa los cinco subconjuntos (SCIENCE, 2020).

En esta técnica la medida del error de todas las iteraciones se calcula de la siguiente manera.

$$E = \frac{1}{k} \sum_{i=1}^k MSE_i,$$

donde $k < n$, siendo n el tamaño del conjunto de datos.

Finalmente, en el método *Leave One Out*, se divide los datos en subconjuntos de prueba y de entrenamiento, pero con un cambio a las técnicas anteriores. En lugar de dividir los datos en 2 subconjuntos, seleccionamos una única observación como datos de prueba, y el resto se etiqueta como datos de entrenamiento y el modelo está entrenado. Ahora se selecciona la segunda observación como datos de prueba y el modelo se entrena con los datos restantes. Este proceso continúa n veces y el promedio de todas estas iteraciones se calcula y estima como el error del conjunto de prueba (Datapeaker, 2022).

Para el método *Leave One Out* la medida del error es:

$$E = \frac{1}{n} \sum_{i=1}^n MSE_i.$$

La técnica de validación cruzada que se elige depende en gran escala de la situación en donde se desee aplicar y de la estabilización entre el error y la varianza, ya que estos factores no tienen el mismo comportamiento en cada modelo o conjunto de datos. En este trabajo utilizaremos el método *K-Fold*.

El criterio de información de Akaike (AIC) y criterio de información bayesiano (BIC) son medidas de la calidad relativa de un modelo estadístico, para un conjunto dado de datos, proporcionado un medio para la selección del modelo. Dado un conjunto de modelos candidatos para los datos, el modelo preferido es el que tiene el valor mínimo en AIC y BIC.

3.1. Criterio de información de Akaike (AIC)

Este criterio, definido por Akaike (1974) como "An Information Criterion", se basa en la medida de información de Kullback-Leible (1951), la cual permite interpretar la distancia entre dos distribuciones a partir de la log-verosimilitud de un modelo. Esta definida como:

$$AIC = -2l(\theta_{n,k}) + 2k,$$

donde $l(\theta_{n,k})$ corresponde al máximo valor de la función de verosimilitud para el modelo estimado, k al número de parámetros y n al tamaño del conjunto de datos.

3.2. Criterio de información bayesiano de Schwarz (BIC)

Dentro del conjunto de los criterios utilizados en la práctica para la selección de modelos, el AIC es sin duda el más conocido, pero existen diferentes criterios contruidos con el mismo fin, pero con características diferentes. Uno de ellos y que se relaciona directamente con el criterio AIC es el criterio BIC ("Bayesian Information Criterion") derivado del criterio AIC al introducir una modificación bayesiana en la construcción del criterio AIC revisada anteriormente. El criterio BIC fue propuesto por Schwarz (1978) y ha sido uno de los métodos más populares usado para la selección de modelos. Este es un criterio de evaluación de modelos en términos de sus probabilidades posteriores. Este criterio está definido como:

$$BIC = -2l(\theta_{n,k}) + k\log(n),$$

donde $l(\theta_{n,k})$ corresponde al máximo valor de la función de verosimilitud para el modelo estimado, k al número de parámetros y n al tamaño del conjunto de datos.

Si comparamos el AIC y el BIC vemos que la diferencia básica entre ambos criterios radica en que este último penaliza más los modelos con un número mayor de parámetros estimados (debido a la sustitución del 2 por $\log(n)$), obteniéndose así modelos de orden inferior a los obtenidos a partir del AIC y corrigiendo, por tanto, la tendencia a la sobrestimación observada en éste último (Jiménez, 2018).

Capítulo 4

Esquemas de simulación y aplicación

En este capítulo se presenta la realización de varias simulaciones con la función *BiCopSelect* del paquete VineCopula del software estadístico R. Considerando pares de variables aleatorias de tamaño, n cada uno. Se generó, en R, datos de las distribuciones marginales gaussianas normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$, log-normal $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$ y t-student $v_1 = v_2 = n - 1$, además, se analizó el caso para marginales weibull con parámetro de forma $\beta = 1$, $\beta = 2$ y $\beta = 0.5$, respectivamente y parámetro de escala $\lambda = 1$ en cada caso.

La elección de la cópula bivariada está entre las elípticas (gaussiana y t-student), las cópulas arquimedianas (clayton, gumbel, frank, joe, BB1, BB6, BB7 y BB8), las cópulas de supervivencia y las cópulas de valor extremo, para cubrir una amplia gama de patrones de dependencia. Para las familias de cópulas arquimedianas, se incluyeron versiones rotadas para cubrir también la dependencia negativa. Para la simulación se realizaron los siguientes pasos:

- Se generó, en R, datos de las dos marginales de las distribuciones con sus respectivos parámetros.
- Se aplica la función *BiCopSelect* (ver A) a vectores de los datos generados de tamaño 50, 100 y 500, con repeticiones de 20, 40 y 100 para cada tamaño y se calculó el porcentaje de las cópulas candidatas a modelar la asociación entre las marginales.
- Se verifica a través de los métodos gráficos, criterio de información de Akaike (AIC) y criterio de información bayesiano (BIC) si las cópulas con mayor porcentaje mostradas en la tabla de resumen de las simulaciones presentan mejor ajuste sobre las demás candidatas.

Tabla 4.1: Nombres de cópulas bivariadas

Nº	Símbolo	Cópula
0	I	independence
1	N	gaussian
2	t	t
3	C	clayton
4	G	gumbel
5	F	frank
6	J	joe
7	BB1	clayton-gumbel
8	BB6	joe-gumbel
9	BB7	joe-clayton
10	BB8	frank-joe
13	SC	Survival clayton
14	SG	Surviva gumbel
16	SJ	Survival joe
17	SBB1	Survival BB1
18	SBB6	Survival BB6
19	SBB7	Survival BB7
20	SBB8	Survival BB8
23	C90	Rotated clayton 90 degrees
24	G90	Rotated gumbel 90 degrees
26	J90	Rotated joe 90 degrees
27	BB1_90	Rotated BB1 90 degrees
28	BB6_90	Rotated BB6 90 degrees
29	BB7_90	Rotated BB7 90 degrees
30	BB8_90	Rotated frank-joe 90 degrees
33	C270	Rotated clayton 270 degrees
34	G270	Rotated gumbel 270 degrees
36	J270	Rotated joe 270 degrees
37	BB1_270	Rotated BB1 270 degrees
38	BB6_270	Rotated BB6 270 degrees

39	BB7_270	Rotated BB7 270 degrees
40	BB8_270	Rotated frank-joe 270 degrees
104	Tawn	Tawn type 1
114	Tawn180	Rotated Tawn type 1 180 degrees
124	Tawn90	Rotated Tawn type 1 90 degrees
134	Tawn270	Rotated Tawn type 1 270 degrees
204	Tawn2	Tawn type 2
214	Tawn2_180	Rotated Tawn type 2 180 degrees
224	Tawn2_90	Rotated Tawn type 2 90 degrees
234	Tawn2_270	Rotated Tawn type 2 270 degrees

En la Tabla 4.1 se muestra los nombres de las familias de cópulas e incluyendo versiones rotadas para cubrir también la dependencia negativa, con sus respectivas abreviaturas del paquete VineCopula.

Tabla 4.2: Resumen de simulación para marginales normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$

n	r	N	SBB7	SG
50	20	70,00	20,00	10,00
	40	62,50	30,00	7,50
	100	66,00	29,00	5,00
100	20	55,00	45,00	0,00
	40	60,00	40,00	0,00
	100	54,00	46,00	0,00
500	20	50,00	50,00	0,00
	40	50,00	50,00	0,00
	100	52,00	48,00	0,00

La Tabla 4.2 contiene el resumen de simulaciones para marginales normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$, que se encuentran en el Apéndice B, la cual indica que, para una corrida de 20, 40 y 100 pares de variables aleatorias con marginales normales con diferentes tamaños ($n = 50, 100, 500$), las cópulas elípticas (gaussianas) presenta más probabilidades de ajustarse, sin embargo, a medida que crece el n , la probabilidad disminuye.

Tabla 4.3: Resumen de simulación para marginales log-normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$

n	r	BB7	frank	joe	SBB7
50	20	5,00	50,00	40,00	5,00
	40	0,00	47,50	50,00	2,50
	100	1,00	47,00	46,00	6,00
100	20	0,00	55,00	35,00	10,00
	40	0,00	35,00	57,50	7,50
	100	1,00	56,00	35,00	8,00
500	20	0,00	55,00	30,00	15,00
	40	5,00	42,50	47,50	5,00
	100	0,00	54,00	44,00	2,00

La Tabla 4.3 contiene el resumen de simulaciones para marginales log-normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$, que se encuentran en el Apéndice C, la cual indica que, las cópulas arquimedianas presentan mayor probabilidad de ajuste en variables aleatorias con marginales log-normal, en específico la cópula frank y la cópula joe, en los diferentes tamaños de n .

Tabla 4.4: Resumen de simulación para marginales t-student con parámetros $v_1 = v_2 = n - 1$

n	r	N	SBB7	SG	BB7
50	20	60,00	30,00	10,00	0,00
	40	62,50	30,00	2,50	5,00
	100	63,00	28,00	7,00	2,00
100	20	75,00	25,00	0,00	0,00
	40	42,50	57,50	0,00	0,00
	100	39,00	60,00	0,00	1,00
500	20	35,00	65,00	0,00	0,00
	40	60,00	40,00	0,00	0,00
	100	55,00	45,00	0,00	0,00

La Tabla 4.4 contiene el resumen de simulación para marginales t-student con parámetros $v_1 = v_2 = n - 1$, que se encuentran en el Apéndice D, la cual indica que, para pares de variables aleatorias, las cópulas elípticas presentan mayores probabilidades de ajuste, en específico la

cópula gaussiana, sin embargo, para $n = 100$, la cópula arquimediana Survival BB7 tiene una probabilidad superior.

4.1. Modelo con marginales weibull

Se realizaron simulaciones para marginales weibull, variando el valor del parámetro. Las Tablas 4.5, 4.6 y 4.7 muestran los resúmenes de los resultados de las simulaciones para marginales weibull con parámetro de forma $\beta = 1$, $\beta = 2$ y $\beta = 0.5$, respectivamente y parámetro de escala $\lambda = 1$.

Tabla 4.5: Resumen de simulación para marginales weibull con parámetros $\beta_1 = \beta_2 = 1$ y $\lambda_1 = \lambda_2 = 1$

n	r	BB7	BB8	frank	N	I	joe	Tawn90	Tawn270	SBB7	t
50	20	35,00	0,00	5,00	5,00	15,00	5,00	0,00	0,00	0,00	35,00
	40	17,50	5,00	10,00	12,50	2,50	32,50	0,00	2,50	12,50	5,00
	100	18,00	6,00	11,00	6,00	3,00	29,00	7,00	4,00	1,00	15,00
100	20	10,00	10,00	45,00	0,00	0,00	25,00	0,00	0,00	5,00	5,00
	40	10,00	20,00	17,50	2,50	0,00	30,00	0,00	0,00	0,00	20,00
	100	8,00	10,00	36,00	6,00	1,00	24,00	0,00	2,00	0,00	13,00
500	20	0,00	15,00	70,00	0,00	0,00	15,00	0,00	0,00	0,00	0,00
	40	0,00	12,50	67,50	0,00	0,00	20,00	0,00	0,00	0,00	0,00
	100	1,00	19,00	46,00	0,00	0,00	32,00	0,00	0,00	0,00	2,00

Para pares de variables aleatorias con marginales weibull con parámetros $\beta_1 = \beta_2 = 1$ y $\lambda_1 = \lambda_2 = 1$, las cópulas arquimedianas presentan mayor probabilidad de ajuste. A medida que el tamaño de n y el número de repeticiones crecen la cópula frank y la cópula joe aumentan la probabilidad de ajuste, esto se refleja en la Tabla 4.5.

Tabla 4.6: Resumen de simulación para marginales weibull con parámetros $\beta_1 = \beta_2 = 2$ y $\lambda_1 = \lambda_2 = 1$

n	r	BB7	BB8	frank	N	I	joe	SBB7	SC	t
50	20	10,00	0,00	5,00	15,00	10,00	25,00	10,00	0,00	25,00
	40	17,50	2,50	17,50	5,00	0,00	22,50	10,00	0,00	25,00
	100	13,00	1,00	13,00	0,00	0,00	26,00	9,00	0,00	38,00
100	20	25,00	0,00	30,00	10,00	0,00	15,00	0,00	0,00	20,00
	40	25,00	2,50	22,50	0,00	0,00	25,00	7,50	0,00	17,50
	100	11,00	2,00	25,00	1,00	0,00	30,00	2,00	1,00	28,00
500	20	20,00	10,00	30,00	0,00	0,00	35,00	5,00	0,00	0,00
	40	17,50	5,00	50,00	0,00	0,00	22,50	0,00	0,00	5,00
	100	12,00	5,00	45,00	0,00	0,00	30,00	6,00	0,00	2,00

Al igual que para los pares de variables aleatorias con marginales weibull con parámetros $\beta_1 = \beta_2 = 1$ y $\lambda_1 = \lambda_2 = 1$, en la Tabla 4.5, para marginales weibull con parámetros $\beta_1 = \beta_2 = 2$ y $\lambda_1 = \lambda_2 = 1$ las cópulas arquimedianas presentan mayor probabilidad de ajuste. A medida que el tamaño de n y el número de repeticiones crecen la cópula frank y la cópula joe aumentan la probabilidad de ajuste, esto se refleja en la Tabla 4.6.

Tabla 4.7: Resumen de simulación para marginales weibull con parámetros $\beta_1 = \beta_2 = 0.5$ y $\lambda_1 = \lambda_2 = 1$

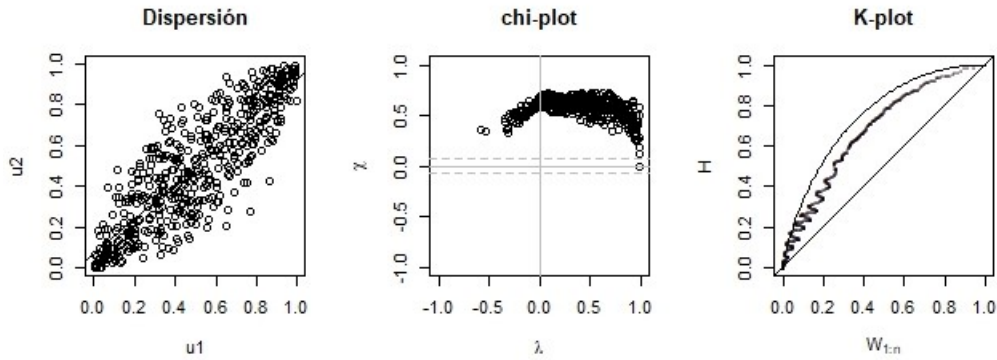
n	r	BB7	frank	N	joe	BB1_90	BB7_270	BB7_90	G270	G90	J270	J90	SBB7	SC	t
50	20	0,00	5,00	0,00	10,00	0,00	10,00	10,00	0,00	5,00	0,00	0,00	40,00	0,00	20,00
	40	0,00	7,50	5,00	10,00	0,00	5,00	17,50	0,00	7,50	0,00	0,00	30,00	2,50	15,00
	100	1,00	6,00	9,00	13,00	1,00	3,00	7,00	6,00	8,00	1,00	1,00	28,00	5,00	11,00
100	20	0,00	0,00	0,00	5,00	0,00	20,00	5,00	0,00	10,00	0,00	0,00	50,00	0,00	10,00
	40	0,00	2,50	2,50	5,00	0,00	0,00	0,00	2,50	0,00	0,00	0,00	52,50	2,50	32,50
	100	0,00	0,00	6,00	6,00	1,00	4,00	9,00	1,00	1,00	0,00	0,00	42,00	0,00	30,00
500	20	0,00	0,00	0,00	0,00	0,00	0,00	5,00	0,00	0,00	0,00	0,00	55,00	0,00	40,00
	40	0,00	0,00	5,00	0,00	0,00	0,00	2,50	0,00	0,00	0,00	0,00	40,00	0,00	52,50
	100	0,00	0,00	1,00	0,00	0,00	2,00	2,00	0,00	0,00	0,00	0,00	46,00	0,00	49,00

La Tabla 4.7 indica que, para pares de variables aleatorias con marginales weibull con parámetros $\beta_1 = \beta_2 = 0.5$ y $\lambda_1 = \lambda_2 = 1$ la cópula survival BB7 presenta mayor probabilidad de ajuste.

4.2. Métodos gráficos para detectar dependencia

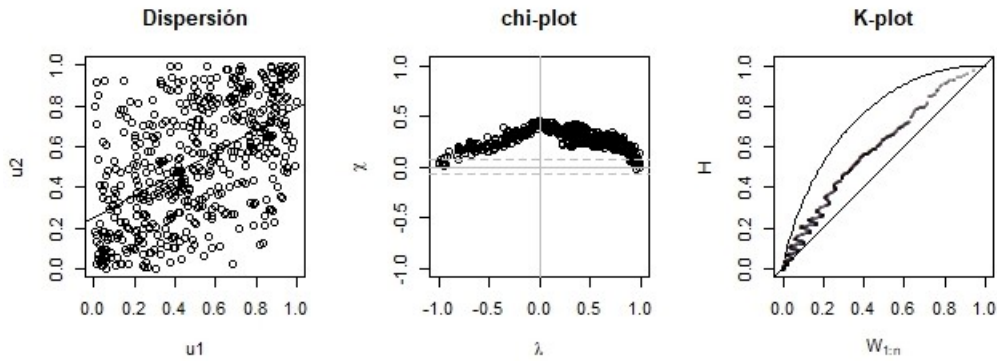
Testeamos los gráficos de dispersión, Chi-plot y K-plot, para la cópula con mayor probabilidad de selección y parámetros obtenidos en cada simulación.

Figura 4.1: Métodos gráficos para la cópula gaussina con $\rho = 0.88$, para marginales normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$



En el testeo gráfico para la cópula gaussiana con parámetro $\rho = 0.88$ para marginales normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$, el gráfico de dispersión indica una relación entre las variables aleatorias, en el gráfico Chi-plot los puntos quedan por fuera de las bandas esto da a indicar una dependencia, al igual que el caso del gráfico K-plot, los puntos se alejan consistentemente de la diagonal, lo que indica dependencia, como se muestra en la Figura 4.1 con una dependencia $\tau = 0.6849$.

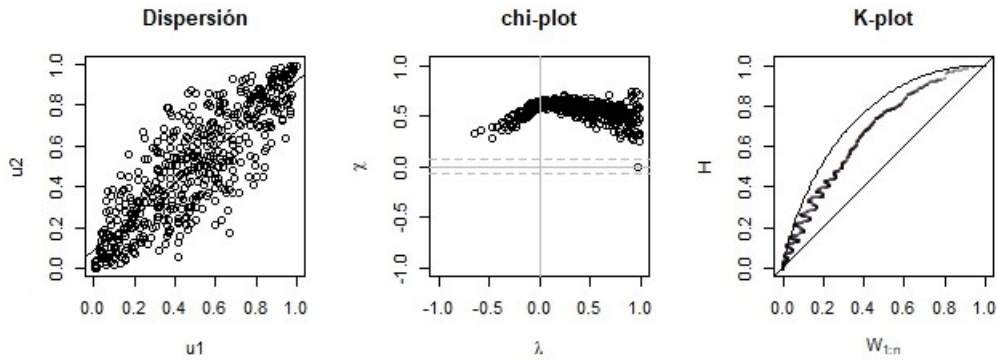
Figura 4.2: Métodos gráficos para la cópula frank con $\theta = 4.438$, para marginales log-normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$



Para marginales log-normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$, la cópula con mayor probabilidad de selección fue frank con un parámetro de 4.438. En la Figura 4.2 se muestran

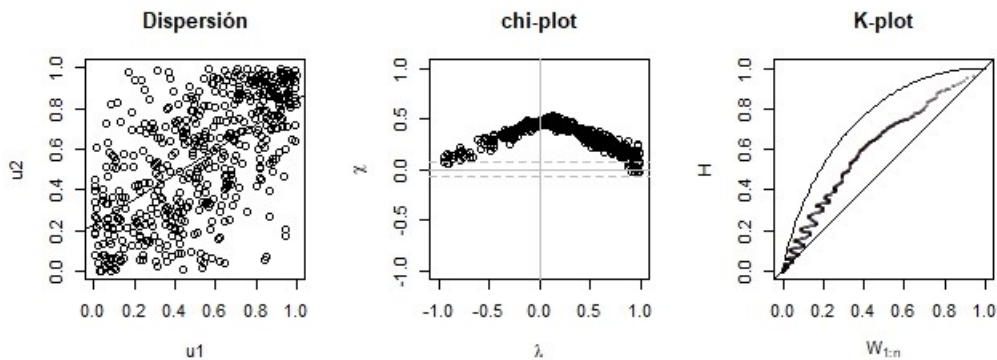
gráficos para detectar dependencia entre las variables. En el gráfico de dispersión se muestra una baja dependencia, sin embargo, en el gráfico Chi-plot los puntos quedan por fuera de las bandas y en el gráfico K-plot, los puntos se alejan consistentemente de la diagonal, en ambos casos refleja dependencia, con $\tau = 0.4189$.

Figura 4.3: Métodos gráficos para la cópula gaussiana con $\rho = 0.87$, para marginales t-student con parámetros $v_1 = v_2 = n - 1$



Para la cópula gaussiana con parámetro $\rho = 0.87$ para marginales t-student con parámetros $v_1 = v_2 = n - 1$, el gráfico de dispersión indica una alta dependencia entre las variables aleatorias, en el gráfico Chi-plot los puntos quedan por fuera de las bandas esto da a indicar dependencia, al igual que el caso del gráfico K-plot, los puntos se alejan consistentemente de la diagonal, lo que indica dependencia, como se muestra en la Figura 4.3 con una dependencia $\tau = 0.6718$.

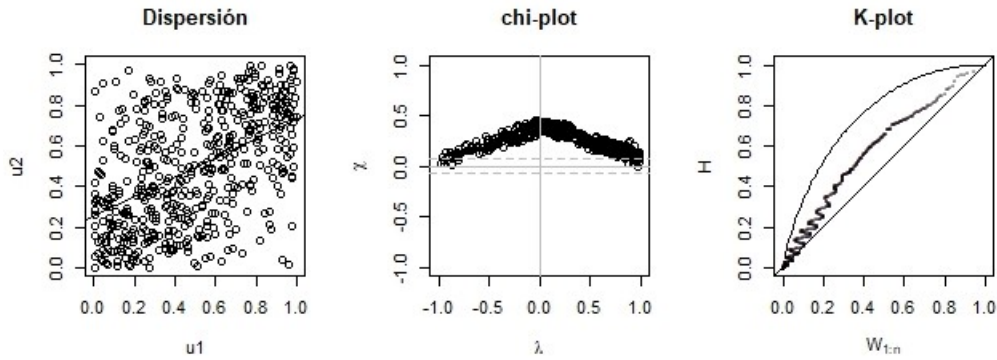
Figura 4.4: Métodos gráficos para la cópula frank con $\theta = 4.0136$, para marginales weibull con parámetros $\beta_1 = \beta_2 = 1$ y $\lambda_1 = \lambda_2 = 1$



Para marginales weibull con parámetros $\beta_1 = \beta_2 = 1$ y $\lambda_1 = \lambda_2 = 1$, la cópula con mayor probabilidad de selección fue frank con un parámetro de 4.0136. En la Figura 4.4 los gráficos

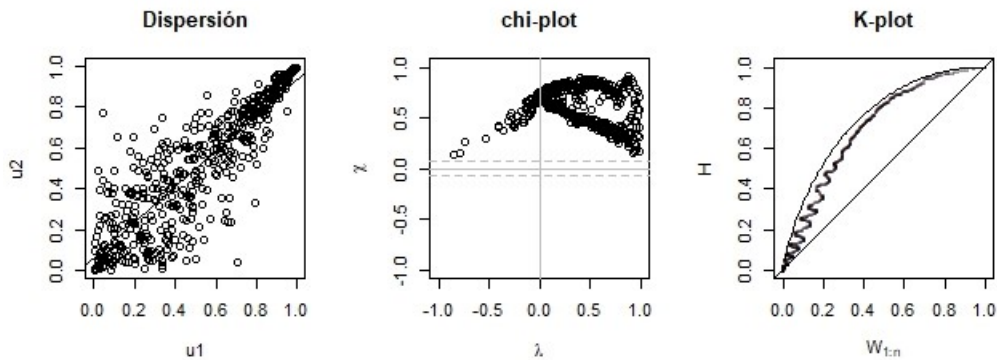
Chip-plot y K-plot detectan dependencia entre las variables aleatorias, sin embargo, en el gráfico de dispersión la dependencia es baja, con $\tau = 0.3891$.

Figura 4.5: Métodos gráficos para la cópula frank con $\theta = 3.3525$, para marginales weibull con parámetros $\beta_1 = \beta_2 = 2$ y $\lambda_1 = \lambda_2 = 1$



Para marginales weibull con parámetros $\beta_1 = \beta_2 = 2$, $\lambda_1 = \lambda_2 = 1$ y la cópula con mayor probabilidad de selección fue frank, al igual que para las marginales weibull con parámetros $\beta_1 = \beta_2 = 1$ y $\lambda_1 = \lambda_2 = 1$, con un parámetro de 3.3525. En la Figura 4.5 los gráficos Chip-plot y K-plot detectan dependencia entre las variables aleatorias, sin embargo, en el gráfico de dispersión la dependencia es baja, con $\tau = 0.3369$.

Figura 4.6: Métodos gráficos para la cópula survival BB7 con $\theta_1 = 1.61$ y $\theta_2 = 4.48$, para marginales weibull con parámetros $\beta_1 = \beta_2 = 0.5$ y $\lambda_1 = \lambda_2 = 1$



Para marginales weibull con parámetros $\beta_1 = \beta_2 = 0.5$, $\lambda_1 = \lambda_2 = 1$, la cópula con mayor probabilidad de selección fue survival BB7 con parámetros de $\theta_1 = 1.61$ y $\theta_2 = 4.48$. En la Figura 4.6 los gráficos Chip-plot y K-plot detectan dependencia entre las variables aleatorias, al igual que en el gráfico de dispersión, con $\tau = 0.6883$.

4.3. Aplicación de validación cruzada para selección de modelos cópulas

Se utilizó el método *K-Fold* de validación cruzada, con $K = 5$, para probar analíticamente la mejor cópula, entre las tres cópulas con mayor porcentaje de ajuste en la simulación con $n = 500$, tomando como métricas de medición el criterio de información de Akaike (AIC) y criterio de información bayesiano (BIC).

Tabla 4.8: AIC y BIC para cópulas ajustadas para marginales normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$

cópula	AIC	BIC
gaussiana ($\rho = 0.88$)	524.5224	527.1276
survival BB7 ($\theta_1 = 5, \theta_2 = 3.30$)	591.4130	596.6233
Survival Gumbel ($\theta = 4.66$)	647.0167	649.6219

La Tabla 4.8 indica que, la cópula elíptica gaussiana presenta mejor ajuste según el criterio de información de Akaike y criterio de información bayesiano en variables aleatorias con marginales normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$.

Tabla 4.9: AIC y BIC para cópulas ajustadas para marginales log-normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$

cópula	AIC	BIC
frank ($\theta = 4.438$)	51.4665	54.0717
joe ($\theta = 3.93$)	262.2161	264.8212
Survival BB7 ($\theta_1 = 1.01, \theta_2 = 3.24$)	288.2866	293.4970

La Tabla 4.9 indica que, la cópula arquimediana frank presenta mejor ajuste según criterio de información de Akaike y criterio de información bayesiano en variables aleatorias con marginales log-normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$.

Tabla 4.10: AIC y BIC para cópulas ajustadas para marginales t-student con parámetros $v_1 = v_2 = n - 1$

cópula	AIC	BIC
gaussiana ($\rho = 0.87$)	407.5406	410.1458
survival BB7 ($\theta_1 = 5, \theta_2 = 3.1$)	543.7185	548.9288
Survival Gumbel ($\theta = 4.89$)	649.9594	652.5646

Según criterio de información de Akaike y criterio de información bayesiano en variables aleatorias con marginales t-student($v_1 = n - 1$) – t-student($v_2 = n - 1$), la cópula elítica gaussiana presenta mejor ajuste, como se muestra en la Tabla 4.10.

Tabla 4.11: AIC y BIC para cópulas ajustadas para marginales weibull con parámetros $\beta_1 = \beta_2 = 1$ y $\lambda_1 = \lambda_2 = 1$

cópula	AIC	BIC
frank ($\theta = 4.0136$)	42.1049	44.7101
joe ($\theta = 3.08$)	157.6756	160.2808
BB8 ($\theta_1 = 3.28, \theta_2 = 1$)	183.0415	188.2519

Según criterio de información de Akaike y criterio de información bayesiano en variables aleatorias con marginales weibull con parámetros $\beta_1 = \beta_2 = 1$ y $\lambda_1 = \lambda_2 = 1$, la cópula arquimediana frank presenta mejor ajuste, como se muestra en la Tabla 4.11.

Tabla 4.12: AIC y BIC para cópulas ajustadas para marginales weibull con parámetros $\beta_1 = \beta_2 = 2$ y $\lambda_1 = \lambda_2 = 1$

cópula	AIC	BIC
frank ($\theta = 3.3525$)	29.4010	32.0062
joe ($\theta = 2.95$)	142.9679	145.5731
BB7 ($\theta_1 = 2.94, \theta_2 = 0.03$)	144.5246	149.7349

La Tabla 4.12 indica que, la cópula arquimediana BB7 presenta mejor ajuste según criterio de información de Akaike y criterio de información bayesiano en variables aleatorias con marginales weibull con parámetros $\beta_1 = \beta_2 = 2$ y $\lambda_1 = \lambda_2 = 1$.

Tabla 4.13: AIC y BIC para cópulas ajustadas para marginales weibull con parámetros $\beta_1 = \beta_2 = 0.5$ y $\lambda_1 = \lambda_2 = 1$

cópula	AIC	BIC
t-student ($\theta_1 = -0.87, \theta_2 = n - 1$)	452.3503	454.9555
joe ($\theta = 6.16$)	580.3195	582.9246
survival BB7 ($\theta_1 = 1.61, \theta_2 = 4.48$)	427.5960	432.8065

La Tabla 4.12 indica que, la cópula arquimediana survival BB7 presenta mejor ajuste según criterio de información de Akaike y criterio de información bayesiano en variables aleatorias con marginales weibull con parámetros $\beta_1 = \beta_2 = 0.5$ y $\lambda_1 = \lambda_2 = 1$.

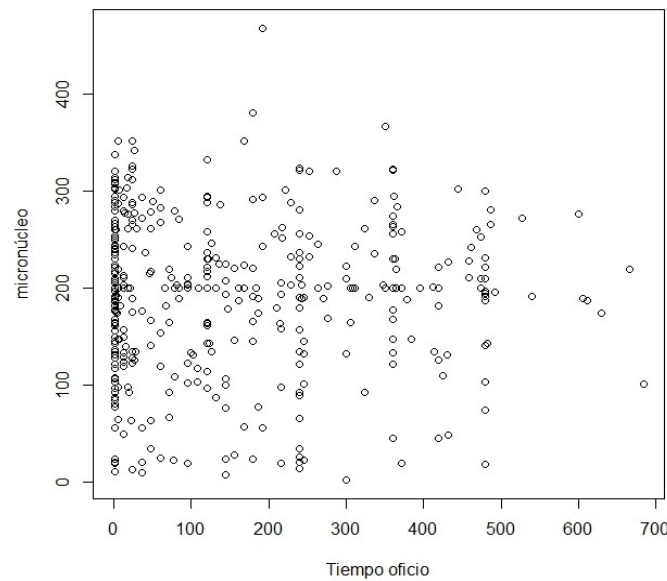
4.4. Aplicación en datos biomédicos

En esta sección estaremos mostrando una aplicación de las funciones cópula para datos biomédicos. Las variables a utilizar son el tiempo de oficio y micronúcleo.

Para la aplicación se tomaron los datos colectados en el marco del proyecto, Evaluación de la exposición a mezclas de contaminantes ambientales y efectos en la salud en población rural en tres departamentos de Colombia 2020 , realizado por la Universidad del Sinú, Universidad Industrial de Santander, y Universidad del Rosario, bajo el patrocinio de Minciencias. Estos datos corresponden a una evaluación de riesgo de cáncer, realizada a 407 individuos pertenecientes a una población expuesta ocupacionalmente a minería y agroquímicos, en los municipios de Montería, Montelíbano (Córdoba), Nechí (Antioquia) y Aránzazu (Caldas), donde 99 individuos son de Montelíbano, 103 de Aranzazu, 104 de Nechí y 101 de Montería, bajo un conteo de daño celular, medido en 2000 células de cada uno de los participantes siguiendo el ensayo Cytokinesis Block Micronucleus Cytome (CBMN Cyt) Assay Biomarkers (Fenech et al., 2003). Se analizaron sistemas mineros de gran, mediana y pequeña escala en Montelíbano, la minería artesanal y de pequeña escala en Nechí y un sistema minero cerrado en Aranzazu. La zona de referencia sin actividad minera se estableció en Montería. ICP-MS (Espectrometría de Masas con Plasma Acoplado Inductivamente) midió la exposición a múltiples elementos en el cabello y la inestabilidad cromosómica, se evaluó mediante la técnica de micronúcleos de bloqueo de citocinesis (Pastor-Sierra et al., 2023).

El objetivo del análisis fue cuantificar la asociación entre el conteo de Micronucleos y el tiempo de oficio, entendido este como el tiempo en meses de exposición a los contaminantes durante el desarrollo de las actividades laborales, y con esto poder determinar la relación verdadera entre este par de variables y nos dará indicios de posibles modelos estadísticos para estudios posteriores. Adicionalmente, se quiere conocer la cópula que nos de cuenta de la relación de dichas variables.

Figura 4.7: Gráfico de dispersión entre tiempo oficio y micronúcleo



El coeficiente de correlación de Pearson para el tiempo oficio y micronúcleo es de $\rho = -0.0135$, indicando baja correlación entre las variables. La Figura 4.7 muestra que no existe correlación lineal entre el tiempo oficio y micronúcleo.

Figura 4.8: Histograma y densidad para tiempo oficio

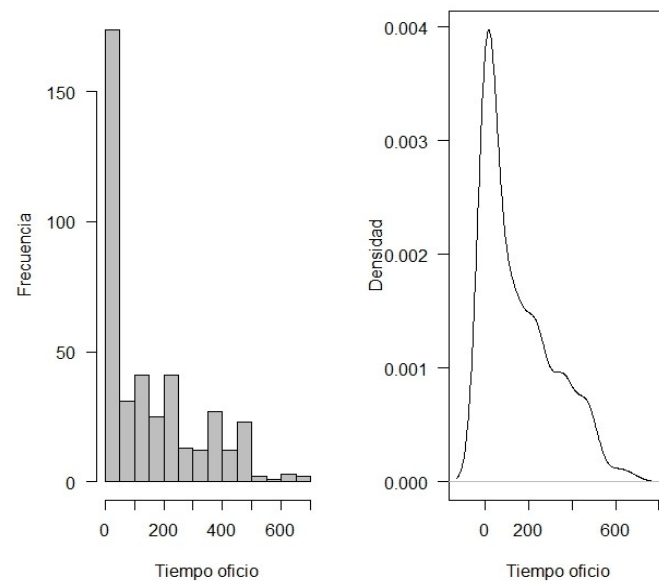
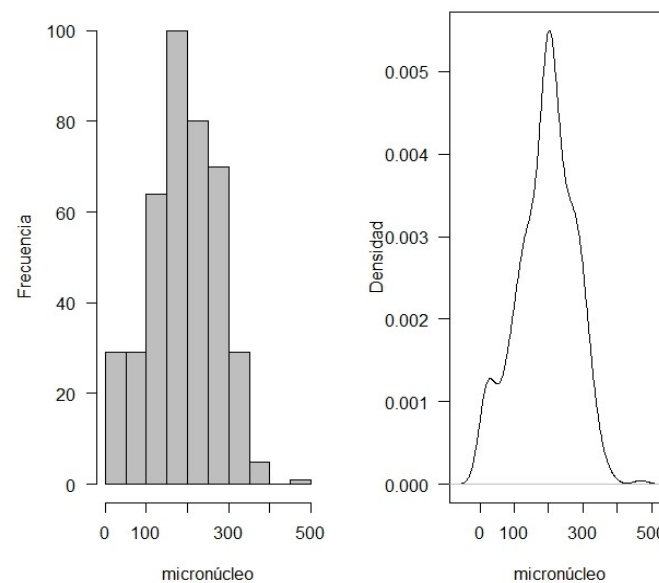


Figura 4.9: Histograma y densidad para micronúcleo



Las Figuras 4.8 y 4.9 muestran el histograma y la densidad para las variables tiempo oficio y micronúcleo. Después de hacer el análisis exploratorio, estimamos los parámetros de la distribución, que en este caso es la logística, a partir de la función *ftdistr* perteneciente al paquete *MASS*.

Así el parámetro de localización para la distribución logística es 195.3834 y de escala es 46.9239, hacemos las pruebas de bondad de ajuste Kolmogorov Smirnov y Anderson Darling, las cuales

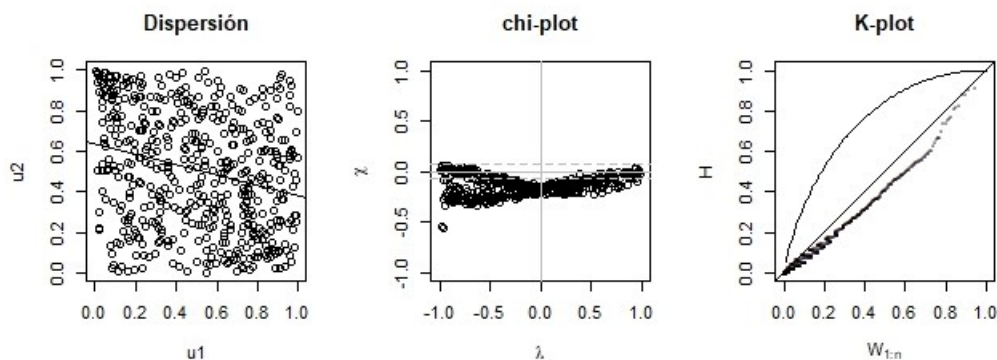
dan como resultado un p-valor de 0.08185 y 0.113 respectivamente, teniendo como hipótesis nula que los datos siguen una distribución logística. Por tanto, con un nivel de significancia $\alpha = 0.05$ no se rechaza la hipótesis nula, es decir que los datos siguen una distribución logística. Teniendo la distribución de los datos con sus respectivos parámetros, procedemos a aplicar la función *BiCopSelect* para ajustar una función cópula que modele el tiempo oficio y el micronúcleo.

Tabla 4.14: Resumen de simulación para datos biomedicos

Tipo de cópula	porcentaje (%)
Frank	6,00
Joe	2,00
J270	6,00
J90	32,00
Tawn180	12,00
Tawn270	10,00
Tawn2_180	12,00
Tawn2_270	4,00
Tawn	4,00
Tawn2	12,00
Total general	100,00

La Tabla 4.14 contiene el resumen de las simulaciones para tiempo oficio y micronúcleo, que se encuentran en el Apéndice F, la cual indica que, las cópulas arquimedianas presentan mayor probabilidad de ajuste, en específico la cópula joe rotada 90 grados.

Figura 4.10: Métodos gráficos para la cópula joe rotada 90 grados para datos biomédicos



En el testeo gráfico para la cópula joe rotada 90 grados con parámetro $\theta = -1.3863$, el gráfico de dispersión indica que no existe relación lineal entre las variables aleatorias, en el gráfico Chi-plot, al igual que el gráfico K-plot, los puntos se alejan de la diagonal, lo que indica dependencia, como se muestra en la Figura 4.10 con una dependencia $\tau = -0.2193$.

Tabla 4.15: AIC y BIC para cópulas ajustadas para datos biomédicos

familia	AIC	BIC
J90 ($\theta = -1.3863$)	4,5393	6,9386
Tawn180 ($\theta_1 = 2.4967, \theta_2 = 0.085$)	7,1505	11,9492
Tawn2_180 ($\theta_1 = 12.105, \theta_2 = 0.1183$)	70,3865	75,1852
Tawn2 ($\theta_1 = 9.3467, \theta_2 = 0.033$)	15,6290	20,4277

La Tabla 4.15 indica que, la cópula arquimediana joe rotada 90 grados con parámetro $\theta = -1.3863$, presenta mejor ajuste según el criterio de información de Akaike y criterio de información bayesiano para tiempo oficio y micronúcleo.

Capítulo 5

Conclusiones y recomendaciones

5.1. Conclusiones

Este trabajo de investigación tuvo como objetivo proponer mecanismos de selección de una cópula para modelar la asociación entre dos variables aleatorias continuas mediante pruebas gráficas (k-plot, chi-plot) y modelos de validación cruzada *K-Fold*.

- Se establecieron las funciones de distribución marginales gaussianas normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$, log-normal $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$ y t-student $v_1 = v_2 = n - 1$, además, se analizó el caso para marginales weibulls con parámetro de forma $\beta = 1$, $\beta = 2$ y $\beta = 0.5$, respectivamente y parámetro de escala $\lambda = 1$ en cada caso.
- Se realizó simulaciones con la función *BiCopSelect* del paquete VineCopula del software estadístico R, identificando las funciones cópulas candidatas para modelar conjuntamente la asociación entre las marginales de las variables aleatorias continuas.
- Se verificó la dependencia a través de los gráficos k-plot y chi-plot vía simulación para la función cópula que mejor ajustó a las marginales establecidas. Se comparó a través de las métricas de medición AIC y BIC las funciones cópulas candidatas para ajustar las marginales establecidas.

Haciendo uso de esta metodología, hemos encontrado que la cópula que mejor ajusta a marginales normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$, fue gaussiana con parámetro $\rho = 0.88$

y una dependencia de $\tau = 0.6849$, para marginales log-normal con parámetros $\mu_1 = \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 0$, la cópula que mejor ajustó fue una frank con un parámetro de 4.438 y dependencia de $\tau = 0.4189$ y para las marginales t-student con parámetros $v_1 = v_2 = n - 1$, la cópula que mejor ajustó fue una gaussiana con parámetro $\rho = 0.87$ y dependencia $\tau = 0.6718$. En el caso de las weibull con parámetros $\beta_1 = \beta_2 = 1$ y $\lambda_1 = \lambda_2 = 1$ la copula que mejor ajustó fue frank con un parámetro de 4.0136 y dependencia de $\tau = 0.3891$, para weibull con parámetros $\beta_1 = \beta_2 = 2$ y $\lambda_1 = \lambda_2 = 1$ la copula que mejor ajustó fue frank con un parámetro de 3.3525 y dependencia de $\tau = 0.3369$ y para marginales weibull con parámetros $\beta_1 = \beta_2 = 0.5$ y $\lambda_1 = \lambda_2 = 1$ la copula que mejor ajustó fue survival BB7 con parámetros $\theta_1 = 1.61$ y $\theta_2 = 4.48$ y dependencia de $\tau = 0.6883$.

Se aplicó la metodología planteada en este trabajo a las variables tiempo oficio y micronúcleo de datos biomédicos y se encontró que la cópula que mejor ajusta la unión de las variables es la cópula arquimediana joe rotada 90 grados con parámetro $\theta = -1.3863$ y una dependencia $\tau = -0.2193$.

A raíz de lo dicho, hemos encontrado que, con los métodos analíticos y gráficos mostrados en este trabajo, fue posible encontrar la asociación de dos variables aleatorias, por medio de las funciones cópula.

5.2. Recomendaciones y futuras investigaciones

Para futuras investigaciones sería de gran ayuda poder usar cópulas discretas y se lograría abarcar problemas donde se involucren conteos.

Adicionalmente, en las funciones cópulas, se sugiere aplicar dicho método cuando las marginales tienen distintas distribuciones.

Apéndice A

Código en R para el proceso de simulación

```
library(VineCopula)
library(CDVineCopulaConditional)

copula=function(n,r){
  ale=1:1000
  for (i in sample(ale,r)) {
    #set.seed(i)
    Nsim=n
    x_1=runif(Nsim)
    x_2=runif(Nsim)

    x1=plnorm(x_1)
    x2=plnorm(x_2)
    selectedCopula <- BiCopSelect(x1,x2,familyset=NA)
    a= selectedCopula$familyname
    b= selectedCopula$family
    c=round(selectedCopula$par,2)
    d=round(selectedCopula$par2,2)
    print(c(b,a,c,d))
  }
}
```

```

#Métodos gráficos de bondad de ajuste
#Grafico de Dispersion
plot(gaussiana,xlab='u1',ylab='u2',main='Dispersión')
abline(lm(ga2~ ga1),lty=1)

#Grafico Chi-plot:
BiCopChiPlot(ga1, ga2, PLOT=TRUE, mode="NULL",main='chi-plot',
             xlim=c(-1,1),ylim=c(-1,1))

#Grafico K-Plot:
BiCopKPlot(ga1, ga2, PLOT=TRUE,main='K-plot')

#función validación cruzada k-folds

library(copula)

# Función para dividir una matriz en k partes y aplicar la función cópula a cada parte
k_folds_copula<- function(mat, k, cop_model,p) {
  # Verificar que la matriz tenga dos columnas
  if (ncol(mat) != 2) {
    stop("La matriz debe tener exactamente dos columnas.")
  }

  # Dividir la matriz en k partes
  mat_parts <- split(mat, f = rep(1:k, length.out = nrow(mat)))

  # Crear una lista para almacenar los resultados
  copula_results <- vector("list", k)
  AIC<-vector("list", k)
  BIC<-vector("list", k)

  # Aplicar la función cópula a cada parte
  for (i in 1:k) {

```

```

# Obtener la parte actual
part <- pobs((matrix(mat_parts[[i]], ncol = 2)))
n<-dim(part)[1]

# Ajustar el modelo de cópula a la parte actual
log_ver=sum(log(dCopula(part, cop_model)))
aic<- -2*log_ver + 2*p
bic <- -2*log_ver + log(n)* p

# Almacenar el resultado en la lista
copula_results[i] <- log_ver
AIC[i]<- aic
BIC[i]<- bic
}

copula_results <- unlist(copula_results)
AIC<- unlist(AIC)
BIC<- unlist(BIC)
return(c(mean(copula_results),mean(AIC),mean(BIC)))
}

#Codigo para hallar las distribución de una variable
\begin{verbatim}
library(MASS)
ajuste<-fitdistr(micronucleo, "logistic");ajuste

> location      scale
195.383403      46.923941
( 4.063375) ( 1.933770)

```

Apéndice B

Marginales normal

Tabla B.1: Pares de variables aleatorias marginales normal $n = 50$, $r = 20$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	1	Gaussian	0.88	0
2	19	Survival BB7	5	3.76
3	1	Gaussian	0.92	0
4	1	Gaussian	0.91	0
5	1	Gaussian	0.91	0
6	19	Survival BB7	5	3.49
7	19	Survival BB7	5	3.72
8	14	Survival Gumbel	4.97	0
9	1	Gaussian	0.9	0
10	1	Gaussian	0.91	0
11	1	Gaussian	0.91	0
12	1	Gaussian	0.92	0
13	1	Gaussian	0.94	0
14	1	Gaussian	0.92	0
15	19	Survival BB7	5	4.1
16	1	Gaussian	0.91	0
17	1	Gaussian	0.87	0
18	1	Gaussian	0.9	0
19	1	Gaussian	0.9	0
20	14	Survival Gumbel	5.61	0

Tabla B.2: Pares de variables aleatorias marginales normal $n = 50$, $r = 40$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	1	Gaussian	0.91	0
2	1	Gaussian	0.91	0
3	1	Gaussian	0.92	0
4	1	Gaussian	0.93	0
5	19	Survival BB7	5	2.94
6	1	Gaussian	0.9	0
7	1	Gaussian	0.93	0
8	19	Survival BB7	5	3.2
9	1	Gaussian	0.91	0
10	19	Survival BB7	5	4.38
11	1	Gaussian	0.9	0
12	19	Survival BB7	5	3.12
13	1	Gaussian	0.96	0
14	1	Gaussian	0.89	0
15	1	Gaussian	0.87	0
16	1	Gaussian	0.91	0
17	1	Gaussian	0.9	0
18	14	Survival Gumbel	4.66	0
19	1	Gaussian	0.92	0
20	19	Survival BB7	5	3.66
21	1	Gaussian	0.9	0
22	1	Gaussian	0.91	0
23	19	Survival BB7	5	4.44
24	1	Gaussian	0.9	0
25	1	Gaussian	0.9	0
26	1	Gaussian	0.91	0
27	14	Survival Gumbel	5.27	0
28	19	Survival BB7	2.49	6
29	19	Survival BB7	5	3.01

30	19	Survival BB7	5	3.86
31	1	Gaussian	0.92	0
32	1	Gaussian	0.92	0
33	1	Gaussian	0.89	0
34	19	Survival BB7	5	4.58
35	1	Gaussian	0.91	0
36	19	Survival BB7	5	4.28
37	19	Survival BB7	5	2.88
38	14	Survival Gumbel	4.83	0
39	1	Gaussian	0.93	0
40	1	Gaussian	0.9	0

Tabla B.3: Pares de variables aleatorias marginales normal $n = 50$, $r = 100$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	19	Survival BB7	5	4.03
2	1	Gaussian	0.9	0
3	19	Survival BB7	5	3.79
4	1	Gaussian	0.89	0
5	1	Gaussian	0.93	0
6	1	Gaussian	0.94	0
7	1	Gaussian	0.91	0
8	1	Gaussian	0.9	0
9	19	Survival BB7	5	4.34
10	19	Survival BB7	2.37	6
11	1	Gaussian	0.93	0
12	19	Survival BB7	5	3.32
13	19	Survival BB7	5	3.62
14	1	Gaussian	0.91	0
15	1	Gaussian	0.93	0
16	1	Gaussian	0.89	0
17	1	Gaussian	0.92	0

18	19	Survival BB7	5	4.86
19	1	Gaussian	0.91	0
20	1	Gaussian	0.91	0
21	1	Gaussian	0.92	0
22	14	Survival Gumbel	5.44	0
23	14	Survival Gumbel	4.51	0
24	1	Gaussian	0.9	0
25	1	Gaussian	0.93	0
26	1	Gaussian	0.91	0
27	19	Survival BB7	5	5.09
28	19	Survival BB7	5	3.03
29	19	Survival BB7	5	3.57
30	1	Gaussian	0.91	0
31	19	Survival BB7	5	3.92
32	1	Gaussian	0.91	0
33	1	Gaussian	0.93	0
34	1	Gaussian	0.92	0
35	1	Gaussian	0.91	0
36	1	Gaussian	0.89	0
37	19	Survival BB7	5	3.48
38	19	Survival BB7	5	3.42
39	1	Gaussian	0.91	0
40	19	Survival BB7	5	3.45
41	1	Gaussian	0.91	0
42	1	Gaussian	0.92	0
43	1	Gaussian	0.91	0
44	1	Gaussian	0.93	0
45	1	Gaussian	0.9	0
46	1	Gaussian	0.92	0
47	1	Gaussian	0.93	0
48	14	Survival Gumbel	4.82	0

49	19	Survival BB7	5	3.46
50	1	Gaussian	0.89	0
51	19	Survival BB7	5	3.15
52	1	Gaussian	0.91	0
53	1	Gaussian	0.91	0
54	1	Gaussian	0.92	0
55	1	Gaussian	0.91	0
56	1	Gaussian	0.88	0
57	1	Gaussian	0.9	0
58	1	Gaussian	0.9	0
59	19	Survival BB7	5	2.82
60	1	Gaussian	0.92	0
61	19	Survival BB7	5	3.88
62	1	Gaussian	0.93	0
63	1	Gaussian	0.91	0
64	19	Survival BB7	5	4.28
65	1	Gaussian	0.93	0
66	19	Survival BB7	5	3.16
67	1	Gaussian	0.91	0
68	1	Gaussian	0.93	0
69	19	Survival BB7	5	3.71
70	1	Gaussian	0.95	0
71	1	Gaussian	0.91	0
72	1	Gaussian	0.92	0
73	19	Survival BB7	5	2.83
74	19	Survival BB7	5	3.47
75	19	Survival BB7	5	3.04
76	1	Gaussian	0.93	0
77	1	Gaussian	0.91	0
78	1	Gaussian	0.9	0
79	14	Survival Gumbel	5.18	0

80	1	Gaussian	0.94	0
81	1	Gaussian	0.89	0
82	1	Gaussian	0.91	0
83	1	Gaussian	0.92	0
84	19	Survival BB7	5	4.89
85	19	Survival BB7	5	4.37
86	1	Gaussian	0.91	0
87	19	Survival BB7	5	3.37
88	1	Gaussian	0.92	0
89	1	Gaussian	0.91	0
90	1	Gaussian	0.92	0
91	1	Gaussian	0.9	0
92	19	Survival BB7	5	4.63
93	1	Gaussian	0.89	0
94	1	Gaussian	0.89	0
95	1	Gaussian	0.94	0
96	1	Gaussian	0.94	0
97	1	Gaussian	0.87	0
98	14	Survival Gumbel	4.66	0
99	1	Gaussian	0.92	0
100	19	Survival BB7	5	3.24

Tabla B.4: Pares de variables aleatorias marginales normal $n = 100$, $r = 20$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	1	Gaussian	0.92	0
2	1	Gaussian	0.92	0
3	1	Gaussian	0.91	0
4	1	Gaussian	0.91	0
5	1	Gaussian	0.9	0
6	1	Gaussian	0.9	0
7	19	Survival BB7	5	2.91

8	19	Survival BB7	5	3.67
9	1	Gaussian	0.91	0
10	1	Gaussian	0.9	0
11	19	Survival BB7	5	3.25
12	19	Survival BB7	5	4.14
13	19	Survival BB7	5	3.29
14	19	Survival BB7	5	3.41
15	19	Survival BB7	5	4.43
16	19	Survival BB7	5	3.16
17	1	Gaussian	0.91	0
18	19	Survival BB7	5	3.9
19	1	Gaussian	0.89	0
20	1	Gaussian	0.9	0

Tabla B.5: Pares de variables aleatorias marginales normal $n = 100$, $r = 40$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	19	Survival BB7	5	3.63
2	1	Gaussian	0.92	0
3	1	Gaussian	0.91	0
4	1	Gaussian	0.9	0
5	19	Survival BB7	5	3.07
6	1	Gaussian	0.91	0
7	19	Survival BB7	5	3.22
8	19	Survival BB7	5	2.9
9	1	Gaussian	0.89	0
10	1	Gaussian	0.9	0
11	1	Gaussian	0.9	0
12	19	Survival BB7	5	2.8
13	1	Gaussian	0.89	0
14	1	Gaussian	0.91	0
15	19	Survival BB7	5	3.42

Pares	Familia	Tipo de cópula	θ_1	θ_1
16	1	Gaussian	0.92	0
17	1	Gaussian	0.91	0
18	19	Survival BB7	5	2.86
19	19	Survival BB7	5	3.68
20	19	Survival BB7	5	4.65
21	19	Survival BB7	5	3.98
22	1	Gaussian	0.92	0
23	1	Gaussian	0.91	0
24	1	Gaussian	0.89	0
25	19	Survival BB7	5	3.56
26	19	Survival BB7	5	3.38
27	1	Gaussian	0.89	0
28	1	Gaussian	0.91	0
29	19	Survival BB7	5	3.61
30	19	Survival BB7	5	2.93
31	1	Gaussian	0.9	0
32	1	Gaussian	0.89	0
33	1	Gaussian	0.91	0
34	1	Gaussian	0.9	0
35	1	Gaussian	0.91	0
36	1	Gaussian	0.91	0
37	19	Survival BB7	5	2.9
38	1	Gaussian	0.89	0
39	19	Survival BB7	5	3.67
40	1	Gaussian	0.91	0

Tabla B.6: Pares de variables aleatorias marginales normal $n = 100$, $r = 100$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	1	Gaussian	0.91	0
2	19	Survival BB7	5	3.92

Pares	Familia	Tipo de cópula	θ_1	θ_1
3	1	Gaussian	0.92	0
4	1	Gaussian	0.91	0
5	19	Survival BB7	5	2.77
6	19	Survival BB7	5	3.51
7	19	Survival BB7	5	3.21
8	19	Survival BB7	5	2.88
9	1	Gaussian	0.92	0
10	19	Survival BB7	5	3.3
11	1	Gaussian	0.92	0
12	19	Survival BB7	5	3.08
13	1	Gaussian	0.92	0
14	1	Gaussian	0.9	0
15	19	Survival BB7	5	3.09
16	1	Gaussian	0.88	0
17	19	Survival BB7	5	3.27
18	19	Survival BB7	5	3.34
19	19	Survival BB7	5	3.77
20	1	Gaussian	0.91	0
21	19	Survival BB7	5	3.75
22	1	Gaussian	0.91	0
23	1	Gaussian	0.9	0
24	1	Gaussian	0.91	0
25	19	Survival BB7	5	2.92
26	1	Gaussian	0.9	0
27	1	Gaussian	0.91	0
28	19	Survival BB7	5	3.85
29	19	Survival BB7	5	4.31
30	1	Gaussian	0.9	0
31	1	Gaussian	0.9	0
32	1	Gaussian	0.94	0
33	19	Survival BB7	5	3.48

Pares	Familia	Tipo de cópula	θ_1	θ_1
34	1	Gaussian	0.89	0
35	1	Gaussian	0.9	0
36	1	Gaussian	0.92	0
37	1	Gaussian	0.9	0
38	19	Survival BB7	5	3.71
39	19	Survival BB7	5	3.4
40	1	Gaussian	0.91	0
41	19	Survival BB7	5	3.54
42	19	Survival BB7	5	3.03
43	19	Survival BB7	5	3.41
44	1	Gaussian	0.9	0
45	19	Survival BB7	5	4.21
46	1	Gaussian	0.9	0
47	1	Gaussian	0.9	0
48	19	Survival BB7	5	2.83
49	1	Gaussian	0.89	0
50	1	Gaussian	0.9	0
51	1	Gaussian	0.91	0
52	19	Survival BB7	5	3.16
53	1	Gaussian	0.94	0
54	19	Survival BB7	5	3.45
55	19	Survival BB7	5	3.6
56	1	Gaussian	0.9	0
57	19	Survival BB7	5	3.88
58	19	Survival BB7	5	3.19
59	19	Survival BB7	5	3.6
60	1	Gaussian	0.9	0
61	1	Gaussian	0.91	0
62	1	Gaussian	0.91	0
63	19	Survival BB7	5	3.59
64	19	Survival BB7	5	3.41

Pares	Familia	Tipo de cópula	θ_1	θ_1
65	1	Gaussian	0.91	0
66	19	Survival BB7	5	3.93
67	1	Gaussian	0.91	0
68	19	Survival BB7	5	3.24
69	1	Gaussian	0.9	0
70	1	Gaussian	0.91	0
71	19	Survival BB7	5	3.37
72	1	Gaussian	0.9	0
73	1	Gaussian	0.91	0
74	1	Gaussian	0.91	0
75	1	Gaussian	0.88	0
76	19	Survival BB7	5	2.92
77	19	Survival BB7	5	3.92
78	1	Gaussian	0.91	0
79	1	Gaussian	0.89	0
80	19	Survival BB7	5	3.5
81	19	Survival BB7	5	3.88
82	1	Gaussian	0.91	0
83	19	Survival BB7	5	3.38
84	19	Survival BB7	5	3.13
85	19	Survival BB7	5	4.21
86	1	Gaussian	0.89	0
87	19	Survival BB7	5	2.71
88	19	Survival BB7	5	3.82
89	1	Gaussian	0.91	0
90	19	Survival BB7	5	3.36
91	1	Gaussian	0.91	0
92	1	Gaussian	0.9	0
93	1	Gaussian	0.9	0
94	1	Gaussian	0.9	0
95	1	Gaussian	0.89	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
96	1	Gaussian	0.91	0
97	19	Survival BB7	5	3.56
98	1	Gaussian	0.89	0
99	1	Gaussian	0.91	0
100	19	Survival BB7	5	3.51

Tabla B.7: Pares de variables aleatorias marginales normal $n = 500$, $r = 20$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	1	gaussiano	0.91	0
2	1	gaussiano	0.9	0
3	1	gaussiano	0.91	0
4	19	Survival BB7	5	2.98
5	1	gaussiano	0.91	0
6	19	Survival BB7	5	2.94
7	19	Survival BB7	5	3.31
8	1	gaussiano	0,92	0
9	1	gaussiano	0.91	0
10	19	Survival BB7	5	3.05
11	1	gaussiano	0.91	0
12	19	Survival BB7	5	3.11
13	19	Survival BB7	5	3.03
14	19	Survival BB7	5	3.78
15	1	gaussiano	0.91	0
16	1	gaussiano	0.91	0
17	19	Survival BB7	5	3.26
18	1	gaussiano	0.91	0
19	19	Survival BB7	5	3.1
20	19	Survival BB7	5	3.17

Tabla B.8: Pares de variables aleatorias marginales normal $n = 500$, $r = 40$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	19	Survival BB7	5	3.26
2	1	Gaussian	0.92	0
3	1	Gaussian	0.92	0
4	1	Gaussian	0.91	0
5	1	Gaussian	0.9	0
6	19	Survival BB7	5	3.2
7	19	Survival BB7	5	3.01
8	19	Survival BB7	5	3.14
9	1	Gaussian	0.91	0
10	1	Gaussian	0.91	0
11	1	Gaussian	0.91	0
12	19	Survival BB7	5	2.96
13	19	Survival BB7	5	3.46
14	19	Survival BB7	5	3
15	19	Survival BB7	5	3.16
16	1	Gaussian	0.9	0
17	1	Gaussian	0.91	0
18	19	Survival BB7	5	3.13
19	1	Gaussian	0.91	0
20	19	Survival BB7	5	3.2
21	19	Survival BB7	5	3.16
22	1	Gaussian	0.91	0
23	19	Survival BB7	5	3.12
24	19	Survival BB7	5	3.28
25	1	Gaussian	0.91	0
26	19	Survival BB7	5	3.03
27	19	Survival BB7	5	3.31
28	1	Gaussian	0.92	0
29	1	Gaussian	0.91	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
30	1	Gaussian	0.91	0
31	1	Gaussian	0.91	0
32	19	Survival BB7	5	3.11
33	19	Survival BB7	5	3.27
34	1	Gaussian	0.91	0
35	1	Gaussian	0.91	0
36	1	Gaussian	0.91	0
37	1	Gaussian	0.91	0
38	19	Survival BB7	5	3.29
39	19	Survival BB7	5	2.96
40	19	Survival BB7	5	3.66

Tabla B.9: Pares de variables aleatorias marginales normal $n = 500$, $r = 100$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	19	Survival BB7	5	3.09
2	1	Gaussian	0.91	0
3	19	Survival BB7	5	3.43
4	19	Survival BB7	5	3.11
5	1	Gaussian	0.91	0
6	19	Survival BB7	5	3.51
7	19	Survival BB7	5	3.28
8	1	Gaussian	0.92	0
9	1	Gaussian	0.91	0
10	19	Survival BB7	5	3.09
11	19	Survival BB7	5	3.07
12	19	Survival BB7	5	2.97
13	1	Gaussian	0.92	0
14	1	Gaussian	0.91	0
15	19	Survival BB7	5	3.49
16	1	Gaussian	0.91	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
17	19	Survival BB7	5	3
18	19	Survival BB7	5	3.38
19	19	Survival BB7	5	3.44
20	19	Survival BB7	5	3.14
21	1	Gaussian	0.91	0
22	1	Gaussian	0.92	0
23	19	Survival BB7	5	3.54
24	1	Gaussian	0.91	0
25	19	Survival BB7	5	3.28
26	1	Gaussian	0.91	0
27	1	Gaussian	0.91	0
28	1	Gaussian	0.91	0
29	19	Survival BB7	5	3.51
30	1	Gaussian	0.91	0
31	19	Survival BB7	5	3.29
32	19	Survival BB7	5	3.3
33	1	Gaussian	0.91	0
34	19	Survival BB7	5	3.33
35	19	Survival BB7	5	3.2
36	1	Gaussian	0.92	0
37	1	Gaussian	0.91	0
38	1	Gaussian	0.91	0
39	19	Survival BB7	5	3.01
40	19	Survival BB7	5	3.02
41	1	Gaussian	0.91	0
42	19	Survival BB7	5	3.25
43	1	Gaussian	0.9	0
44	1	Gaussian	0.91	0
45	19	Survival BB7	5	3.01
46	1	Gaussian	0.91	0
47	1	Gaussian	0.91	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
48	19	Survival BB7	5	2.97
49	1	Gaussian	0.91	0
50	1	Gaussian	0.91	0
51	1	Gaussian	0.91	0
52	19	Survival BB7	5	2.94
53	19	Survival BB7	5	3.45
54	1	Gaussian	0.91	0
55	1	Gaussian	0.91	0
56	19	Survival BB7	5	3.43
57	1	Gaussian	0.91	0
58	19	Survival BB7	5	3.4
59	19	Survival BB7	5	3.39
60	1	Gaussian	0.91	0
61	1	Gaussian	0.92	0
62	19	Survival BB7	5	3.38
63	1	Gaussian	0.91	0
64	1	Gaussian	0.91	0
65	1	Gaussian	0.91	0
66	1	Gaussian	0.91	0
67	19	Survival BB7	5	3.22
68	1	Gaussian	0.9	0
69	19	Survival BB7	5	2.97
70	19	Survival BB7	5	3.2
71	1	Gaussian	0.91	0
72	1	Gaussian	0.91	0
73	19	Survival BB7	5	3.03
74	19	Survival BB7	5	3.03
75	1	Gaussian	0.91	0
76	19	Survival BB7	5	3.26
77	19	Survival BB7	5	3.14
78	19	Survival BB7	5	3.21

Pares	Familia	Tipo de cópula	θ_1	θ_1
79	1	Gaussian	0.91	0
80	19	Survival BB7	5	3.19
81	19	Survival BB7	5	3.07
82	19	Survival BB7	5	3.19
83	1	Gaussian	0.91	0
84	1	Gaussian	0.91	0
85	1	Gaussian	0.91	0
86	1	Gaussian	0.92	0
87	19	Survival BB7	5	3.11
88	1	Gaussian	0.91	0
89	19	Survival BB7	5	3.4
90	1	Gaussian	0.92	0
91	1	Gaussian	0.91	0
92	1	Gaussian	0.91	0
93	1	Gaussian	0.91	0
94	19	Survival BB7	5	3.61
95	1	Gaussian	0.91	0
96	19	Survival BB7	5	2.99
97	1	Gaussian	0.91	0
98	1	Gaussian	0.92	0
99	19	Survival BB7	5	3.2
100	19	Survival BB7	5	3.17

Apéndice C

Marginales log-normal

Tabla C.1: Pares de variables aleatorias marginales log-normal $n = 50$, $r = 20$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	6	Joe	4.27	0
2	5	Frank	4.1	0
3	6	Joe	4.1	0
4	6	Joe	3.94	0
5	5	Frank	4.97	0
6	5	Frank	4.68	0
7	5	Frank	4.54	0
8	6	Joe	4.56	0
9	19	Survival BB7	1.03	2.85
10	5	Frank	4.72	0
11	6	Joe	4.34	0
12	6	Joe	4.6	0
13	5	Frank	3.97	0
14	5	Frank	3.95	0
15	6	Joe	4.46	0
16	6	Joe	3.69	0
17	5	Frank	4.04	0
18	5	Frank	3.53	0
19	5	Frank	5.88	0
20	9	BB7	4.31	0.13

Tabla C.2: Pares de variables aleatorias marginales log-normal $n = 50$, $r = 40$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	6	Joe	3.7	0
2	5	Frank	4.74	0
3	6	Joe	4.13	0
4	5	Frank	3.99	0
5	6	Joe	4.39	0
6	5	Frank	3.48	0
7	5	Frank	4.6	0
8	6	Joe	4.45	0
9	6	Joe	3.83	0
10	6	Joe	4.29	0
11	6	Joe	4.66	0
12	6	Joe	3.95	0
13	5	Frank	5.07	0
14	6	Joe	3.76	0
15	6	Joe	4.98	0
16	5	Frank	4.59	0
17	6	Joe	5.04	0
18	6	Joe	3.72	0
19	5	Frank	4.5	0
20	5	Frank	3.85	0
21	5	Frank	4.53	0
22	5	Frank	5.05	0
23	19	Survival BB7	1.02	2.93
24	5	Frank	4.87	0
25	6	Joe	3.92	0
26	5	Frank	4.76	0
27	6	Joe	4.24	0
28	5	Frank	4.58	0
29	5	Frank	4.49	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
30	6	Joe	4.81	0
31	5	Frank	3.77	0
32	6	Joe	4.27	0
33	5	Frank	5.32	0
34	5	Frank	3.71	0
35	6	Joe	4.17	0
36	5	Frank	4.42	0
37	6	Joe	4.36	0
38	5	Frank	4.7	0
39	6	Joe	3.89	0
40	6	Joe	4.02	0

Tabla C.3: Pares de variables aleatorias marginales log-normal $n = 50$, $r = 100$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	6	Joe	4.57	0
2	6	Joe	3.59	0
3	5	Frank	5.03	0
4	5	Frank	4.22	0
5	6	Joe	4.71	0
6	5	Frank	4.52	0
7	5	Frank	4.71	0
8	6	Joe	4.21	0
9	6	Joe	4.44	0
10	6	Joe	3.78	0
11	5	Frank	4.59	0
12	6	Joe	3.93	0
13	5	Frank	3.94	0
14	5	Frank	5.11	0
15	6	Joe	4.5	0
16	6	Joe	4.18	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
17	5	Frank	3.86	0
18	6	Joe	4.06	0
19	5	Frank	5.12	0
20	5	Frank	4.56	0
21	6	Joe	3.79	0
22	5	Frank	4.3	0
23	5	Frank	4.44	0
24	9	BB7	3.96	0.09
25	19	Survival BB7	1.02	3.14
26	6	Joe	4.03	0
27	6	Joe	3.78	0
28	5	Frank	4.63	0
29	19	Survival BB7	1.02	3.59
30	6	Joe	4.38	0
31	5	Frank	4.55	0
32	5	Frank	5.63	0
33	6	Joe	4.6	0
34	6	Joe	4.36	0
35	6	Joe	4.41	0
36	5	Frank	5.22	0
37	6	Joe	4.16	0
38	6	Joe	4.46	0
39	6	Joe	4.67	0
40	6	Joe	3.75	0
41	5	Frank	4.84	0
42	5	Frank	5.02	0
43	6	Joe	3.85	0
44	5	Frank	4.57	0
45	6	Joe	4.63	0
46	5	Frank	4.38	0
47	6	Joe	3.86	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
48	6	Joe	5.03	0
49	5	Frank	4.77	0
50	6	Joe	4.02	0
51	5	Frank	4.45	0
52	5	Frank	3.99	0
53	5	Frank	4.93	0
54	6	Joe	4.11	0
55	19	Survival BB7	1.03	3.5
56	6	Joe	4.68	0
57	5	Frank	4.46	0
58	6	Joe	4.53	0
59	6	Joe	3.88	0
60	5	Frank	4.44	0
61	6	Joe	3.77	0
62	6	Joe	4.26	0
63	6	Joe	4.16	0
64	6	Joe	4.6	0
65	5	Frank	4.2	0
66	5	Frank	4.4	0
67	6	Joe	4.56	0
68	6	Joe	4.07	0
69	5	Frank	4.37	0
70	19	Survival BB7	1.04	3.53
71	5	Frank	4.4	0
72	5	Frank	4.25	0
73	6	Joe	4.46	0
74	5	Frank	4.12	0
75	6	Joe	3.82	0
76	5	Frank	3.79	0
77	5	Frank	4.37	0
78	6	Joe	3.72	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
79	19	Survival BB7	1.04	3.63
80	5	Frank	4.72	0
81	5	Frank	4.16	0
82	6	Joe	4.85	0
83	5	Frank	4.56	0
84	5	Frank	5	0
85	5	Frank	4.46	0
86	5	Frank	5	0
87	6	Joe	4.39	0
88	6	Joe	4.37	0
89	6	Joe	4.17	0
90	5	Frank	4.46	0
91	5	Frank	3.54	0
92	5	Frank	3.76	0
93	6	Joe	3.79	0
94	5	Frank	4.91	0
95	6	Joe	3.56	0
96	19	Survival BB7	1.03	3.57
97	6	Joe	5.24	0
98	5	Frank	4.7	0
99	5	Frank	4.48	0
100	5	Frank	5.06	0

Tabla C.4: Pares de variables aleatorias marginales log-normal $n = 100$, $r = 20$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	6	Joe	4.21	0
2	19	Survival BB7	1.02	3.27
3	6	Joe	4.5	0
4	5	Frank	5.2	0
5	5	Frank	4.49	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
6	6	Joe	4.17	0
7	5	Frank	4.51	0
8	6	Joe	4.18	0
9	6	Joe	4.41	0
10	5	Frank	4.29	0
11	5	Frank	4.38	0
12	5	Frank	4.74	0
13	5	Frank	4.4	0
14	5	Frank	4.99	0
15	6	Joe	4.1	0
16	6	Joe	3.92	0
17	5	Frank	4.45	0
18	5	Frank	4.86	0
19	19	Survival BB7	1.03	3.68
20	5	Frank	4.48	0

Tabla C.5: Pares de variables aleatorias marginales log-normal $n = 100$, $r = 40$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	6	Joe	4.22	0
2	5	Frank	3.97	0
3	5	Frank	5.19	0
4	6	Joe	4.14	0
5	6	Joe	4.54	0
6	5	Frank	4.3	0
7	19	Survival BB7	1.03	3.55
8	6	Joe	4.06	0
9	19	Survival BB7	1.01	3.22
10	6	Joe	4.02	0
11	6	Joe	4.13	0
12	6	Joe	4.27	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
13	5	Frank	4.95	0
14	5	Frank	5.48	0
15	19	Survival BB7	1.02	2.85
16	6	Joe	4.65	0
17	5	Frank	4.21	0
18	5	Frank	4.91	0
19	6	Joe	3.96	0
20	6	Joe	3.82	0
21	5	Frank	4.48	0
22	6	Joe	4.15	0
23	6	Joe	4.27	0
24	6	Joe	4.07	0
25	6	Joe	4.28	0
26	6	Joe	3.95	0
27	5	Frank	4.51	0
28	6	Joe	3.92	0
29	6	Joe	4.31	0
30	6	Joe	4.17	0
31	5	Frank	4.72	0
32	5	Frank	4.69	0
33	6	Joe	4.39	0
34	6	Joe	4.49	0
35	6	Joe	4.32	0
36	6	Joe	4.02	0
37	5	Frank	5.08	0
38	5	Frank	4.39	0
39	5	Frank	4.37	0
40	6	Joe	3.95	0

Tabla C.6: Pares de variables aleatorias marginales log-normal $n = 100$, $r = 100$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	5	Frank	4.43	0
2	5	Frank	4.17	0
3	5	Frank	4.87	0
4	6	Joe	4.2	0
5	5	Frank	4.52	0
6	6	Joe	3.61	0
7	5	Frank	4.91	0
8	5	Frank	5.03	0
9	6	Joe	3.74	0
10	5	Frank	4.33	0
11	5	Frank	4.54	0
12	6	Joe	4.15	0
13	5	Frank	3.72	0
14	5	Frank	3.95	0
15	5	Frank	4.18	0
16	9	BB7	3.74	0.05
17	5	Frank	4.4	0
18	19	Survival BB7	1.02	3.34
19	5	Frank	4.81	0
20	5	Frank	4.85	0
21	5	Frank	4.93	0
22	5	Frank	4.97	0
23	5	Frank	4.59	0
24	6	Joe	4.49	0
25	5	Frank	4.48	0
26	6	Joe	4.11	0
27	6	Joe	3.96	0
28	19	Survival BB7	1.01	2.72
29	6	Joe	4.3	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
30	6	Joe	4.24	0
31	5	Frank	4.7	0
32	6	Joe	4.04	0
33	5	Frank	3.89	0
34	5	Frank	4.49	0
35	5	Frank	4.34	0
36	5	Frank	4.53	0
37	19	Survival BB7	0.39	0
38	5	Frank	4.44	0
39	5	Frank	4.38	0
40	5	Frank	4.66	0
41	6	Joe	4.86	0
42	5	Frank	4.63	0
43	5	Frank	4.91	0
44	6	Joe	4.24	0
45	6	Joe	4.51	0
46	6	Joe	4.11	0
47	6	Joe	4.83	0
48	5	Frank	4.64	0
49	19	Survival BB7	1.02	3.47
50	6	Joe	4.88	0
51	6	Joe	4.12	0
52	5	Frank	4.59	0
53	5	Frank	4.99	0
54	6	Joe	3.71	0
55	5	Frank	3.94	0
56	5	Frank	5.14	0
57	6	Joe	3.85	0
58	5	Frank	4.81	0
59	6	Joe	3.76	0
60	6	Joe	4.18	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
61	6	Joe	4.43	0
62	6	Joe	4.06	0
63	5	Frank	4.88	0
64	5	Frank	4.55	0
65	19	Survival BB7	1.02	2.95
66	5	Frank	4.6	0
67	6	Joe	4.17	0
68	5	Frank	3.51	0
69	5	Frank	4.86	0
70	6	Joe	3.98	0
71	19	Survival BB7	1.01	2.95
72	5	Frank	4.25	0
73	5	Frank	4.93	0
74	6	Joe	4.08	0
75	5	Frank	4.4	0
76	5	Frank	4.87	0
77	6	Joe	3.99	0
78	5	Frank	5.05	0
79	5	Frank	4.63	0
80	6	Joe	4.26	0
81	5	Frank	4.45	0
82	5	Frank	4.5	0
83	19	Survival BB7	1.03	3.37
84	6	Joe	4.07	0
85	6	Joe	4.35	0
86	5	Frank	4.27	0
87	6	Joe	4.19	0
88	6	Joe	4.03	0
89	5	Frank	4.99	0
90	5	Frank	3.88	0
91	19	Survival BB7	1.01	3.05

Pares	Familia	Tipo de cópula	θ_1	θ_1
92	5	Frank	3.98	0
93	5	Frank	4.31	0
94	6	Joe	4.43	0
95	5	Frank	4.24	0
96	6	Joe	3.78	0
97	5	Frank	4.05	0
98	5	Frank	4.9	0
99	6	Joe	4.17	0
100	5	Frank	4.76	0

Tabla C.7: Pares de variables aleatorias marginales log-normal $n = 500$, $r = 20$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	5	Frank	4.39	0
2	5	Frank	5.02	0
3	5	Frank	4.87	0
4	5	Frank	5.12	0
5	5	Frank	4.77	0
6	5	Frank	4.67	0
7	19	Survival BB7	1.01	3.15
8	19	Survival BB7	1	3.11
9	6	Joe	4.01	0
10	6	Joe	3.99	0
11	5	Frank	5.05	0
12	5	Frank	4.99	0
13	19	Survival BB7	1.01	3.24
14	6	Joe	3.88	0
15	5	Frank	4.85	0
16	6	Joe	3.89	0
17	6	Joe	4.01	0
18	6	Joe	3.98	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
19	5	Frank	4.81	0
20	5	Frank	4.54	0

Tabla C.8: Pares de variables aleatorias marginales log-normal $n = 500$, $r = 40$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	5	Frank	4.73	0
2	5	Frank	4.67	0
3	6	Joe	4	0
4	6	Joe	3.93	0
5	5	Frank	4.84	0
6	5	Frank	4.74	0
7	6	Joe	3.94	0
8	5	Frank	4.98	0
9	5	Frank	4.78	0
10	5	Frank	4.76	0
11	6	Joe	4	0
12	6	Joe	3.81	0
13	5	Frank	4.8	0
14	6	Joe	4.28	0
15	6	Joe	3.89	0
16	6	Joe	3.97	0
17	5	Frank	4.72	0
18	5	Frank	4.67	0
19	6	Joe	4.02	0
20	9	BB7	3.95	0.03
21	5	Frank	4.66	0
22	5	Frank	4.78	0
23	6	Joe	4.06	0
24	6	Joe	3.94	0
25	5	Frank	4.72	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
26	19	Survival BB7	1.01	2.94
27	5	Frank	4.76	0
28	6	Joe	4.08	0
29	5	Frank	4.64	0
30	6	Joe	4.04	0
31	6	Joe	3.94	0
32	9	BB7	4.01	0.02
33	6	Joe	3.92	0
34	19	Survival BB7	1.01	3.2
35	6	Joe	4.07	0
36	6	Joe	3.81	0
37	6	Joe	3.9	0
38	6	Joe	4.03	0
39	5	Frank	4.92	0
40	5	Frank	4.83	0

Tabla C.9: Pares de variables aleatorias marginales log-normal $n = 500$, $r = 100$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	5	Frank	5.25	0
2	6	Joe	4.02	0
3	5	Frank	4.72	0
4	5	Frank	5	0
5	19	Survival BB7	1.01	3.25
6	6	Joe	4.16	0
7	5	Frank	4.59	0
8	5	Frank	4.69	0
9	5	Frank	4.91	0
10	5	Frank	4.92	0
11	6	Joe	4.06	0
12	6	Joe	3.87	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
13	5	Frank	4.26	0
14	5	Frank	4.88	0
15	6	Joe	3.95	0
16	6	Joe	3.99	0
17	5	Frank	4.81	0
18	6	Joe	3.92	0
19	5	Frank	4.69	0
20	6	Joe	3.86	0
21	5	Frank	5.11	0
22	5	Frank	4.98	0
23	5	Frank	4.79	0
24	5	Frank	4.88	0
25	6	Joe	4.13	0
26	6	Joe	4.07	0
27	6	Joe	3.82	0
28	5	Frank	4.77	0
29	6	Joe	3.83	0
30	6	Joe	4.04	0
31	6	Joe	4.13	0
32	6	Joe	3.82	0
33	5	Frank	4.84	0
34	6	Joe	4	0
35	6	Joe	4.04	0
36	6	Joe	3.91	0
37	5	Frank	4.81	0
38	6	Joe	4.05	0
39	5	Frank	4.88	0
40	6	Joe	4.09	0
41	6	Joe	3.98	0
42	5	Frank	4.88	0
43	5	Frank	4.82	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
44	5	Frank	4.76	0
45	6	Joe	3.75	0
46	5	Frank	4.93	0
47	5	Frank	4.77	0
48	6	Joe	4.01	0
49	5	Frank	4.87	0
50	5	Frank	4.82	0
51	5	Frank	4.54	0
52	6	Joe	4.23	0
53	5	Frank	4.95	0
54	5	Frank	4.98	0
55	6	Joe	4.33	0
56	6	Joe	4.15	0
57	6	Joe	4.05	0
58	5	Frank	4.65	0
59	5	Frank	4.75	0
60	5	Frank	4.87	0
61	5	Frank	4.83	0
62	6	Joe	3.97	0
63	5	Frank	4.59	0
64	5	Frank	4.64	0
65	6	Joe	3.96	0
66	5	Frank	4.8	0
67	5	Frank	4.78	0
68	5	Frank	4.57	0
69	19	Survival BB7	1	3.16
70	6	Joe	3.92	0
71	5	Frank	4.93	0
72	6	Joe	4.42	0
73	6	Joe	4.08	0
74	5	Frank	4.49	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
75	6	Joe	3.99	0
76	5	Frank	4.94	0
77	5	Frank	4.53	0
78	5	Frank	4.99	0
79	5	Frank	4.79	0
80	6	Joe	4.02	0
81	6	Joe	3.89	0
82	6	Joe	4.03	0
83	6	Joe	4.08	0
84	6	Joe	4.1	0
85	6	Joe	3.93	0
86	5	Frank	4.95	0
87	6	Joe	4.19	0
88	6	Joe	4.18	0
89	5	Frank	4.94	0
90	5	Frank	4.98	0
91	5	Frank	4.9	0
92	6	Joe	3.94	0
93	6	Joe	3.99	0
94	5	Frank	4.83	0
95	5	Frank	4.94	0
96	5	Frank	4.83	0
97	5	Frank	4.91	0
98	5	Frank	4.82	0
99	5	Frank	4.8	0
100	6	Joe	4.01	0

Apéndice D

Marginales t-student

Tabla D.1: Pares de variables aleatorias marginales t-student $n = 50$, $r = 20$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	19	Survival BB7	5	3.49
2	1	Gaussian	0.89	0
3	14	Survival Gumbel	4.59	0
4	1	Gaussian	0.9	0
5	1	Gaussian	0.9	0
6	19	Survival BB7	5	4.26
7	1	Gaussian	0.93	0
8	1	Gaussian	0.89	0
9	1	Gaussian	0.9	0
10	19	Survival BB7	5	5.22
11	1	Gaussian	0.91	0
12	1	Gaussian	0.93	0
13	19	Survival BB7	5	3.81
14	1	Gaussian	0.9	0
15	19	Survival BB7	5	3.9
16	1	Gaussian	0.91	0
17	19	Survival BB7	5	5.09
18	1	Gaussian	0.89	0
19	14	Survival Gumbel	5.1	0
20	1	Gaussian	0.91	0

Tabla D.2: Pares de variables aleatorias marginales t-studen $n = 50$, $r = 40$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	1	Gaussian	0.91	0
2	9	BB7	2.39	6
3	14	Survival Gumbel	4.99	0
4	1	Gaussian	0.91	0
5	1	Gaussian	0.92	0
6	19	Survival BB7	5	3.12
7	1	Gaussian	0.92	0
8	19	Survival BB7	5	3.61
9	1	Gaussian	0.89	0
10	19	Survival BB7	5	4
11	1	Gaussian	0.92	0
12	1	Gaussian	0.91	0
13	1	Gaussian	0.89	0
14	1	Gaussian	0.94	0
15	1	Gaussian	0.87	0
16	1	Gaussian	0.91	0
17	1	Gaussian	0.92	0
18	1	Gaussian	0.89	0
19	19	Survival BB7	5	3.91
20	1	Gaussian	0.89	0
21	9	BB7	2.45	6
22	19	Survival BB7	5	3.63
23	1	Gaussian	0.9	0
24	1	Gaussian	0.92	0
25	19	Survival BB7	5	3.37
26	19	Survival BB7	5	3.52
27	1	Gaussian	0.91	0
28	19	Survival BB7	5	4.61
29	1	Gaussian	0.89	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
30	1	Gaussian	0.91	0
31	1	Gaussian	0.92	0
32	19	Survival BB7	5	2.81
33	1	Gaussian	0.91	0
34	1	Gaussian	0.91	0
35	19	Survival BB7	5	3.32
36	1	Gaussian	0.89	0
37	1	Gaussian	0.94	0
38	19	Survival BB7	5	3.28
39	19	Survival BB7	5	3.24
40	1	Gaussian	0.91	0

Tabla D.3: Pares de variables aleatorias marginales t-studen $n = 50$, $r = 100$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	1	Gaussian	0.92	0
2	19	Survival BB7	5	6
3	1	Gaussian	0.91	0
4	1	Gaussian	0.9	0
5	19	Survival BB7	5	4.62
6	1	Gaussian	0.88	0
7	1	Gaussian	0.93	0
8	1	Gaussian	0.89	0
9	19	Survival BB7	5	3.71
10	1	Gaussian	0.89	0
11	1	Gaussian	0.9	0
12	19	Survival BB7	5	3.12
13	1	Gaussian	0.9	0
14	1	Gaussian	0.9	0
15	1	Gaussian	0.92	0
16	1	Gaussian	0.89	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
17	19	Survival BB7	5	4.61
18	19	Survival BB7	5	3.47
19	19	Survival BB7	5	3.16
20	1	Gaussian	0.91	0
21	1	Gaussian	0.93	0
22	1	Gaussian	0.92	0
23	14	Survival Gumbel	5.26	0
24	19	Survival BB7	5	5.06
25	19	Survival BB7	5	3.21
26	1	Gaussian	0.92	0
27	19	Survival BB7	5	5.76
28	19	Survival BB7	5	3.6
29	14	Survival Gumbel	5.45	0
30	1	Gaussian	0.9	0
31	1	Gaussian	0.91	0
32	1	Gaussian	0.89	0
33	14	Survival Gumbel	4.89	0
34	9	BB7	2.34	6
35	1	Gaussian	0.93	0
36	1	Gaussian	0.89	0
37	19	Survival BB7	5	3.22
38	1	Gaussian	0.92	0
39	14	Survival Gumbel	4.53	0
40	1	Gaussian	0.9	0
41	1	Gaussian	0.91	0
42	1	Gaussian	0.91	0
43	19	Survival BB7	5	3.7
44	19	Survival BB7	5	3.12
45	1	Gaussian	0.92	0
46	1	Gaussian	0.93	0
47	1	Gaussian	0.88	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
48	1	Gaussian	0.94	0
49	1	Gaussian	0.91	0
50	1	Gaussian	0.9	0
51	1	Gaussian	0.9	0
52	19	Survival BB7	5	5.05
53	1	Gaussian	0.93	0
54	1	Gaussian	0.9	0
55	19	Survival BB7	5	4.39
56	1	Gaussian	0.91	0
57	19	Survival BB7	5	4.43
58	19	Survival BB7	5	3.45
59	1	Gaussian	0.89	0
60	14	Survival Gumbel	5.42	0
61	1	Gaussian	0.94	0
62	1	Gaussian	0.91	0
63	1	Gaussian	0.91	0
64	1	Gaussian	0.9	0
65	1	Gaussian	0.89	0
66	1	Gaussian	0.91	0
67	1	Gaussian	0.88	0
68	1	Gaussian	0.94	0
69	1	Gaussian	0.93	0
70	1	Gaussian	0.91	0
71	1	Gaussian	0.92	0
72	1	Gaussian	0.92	0
73	9	BB7	2.37	6
74	1	Gaussian	0.94	0
75	19	Survival BB7	5	2.96
76	1	Gaussian	0.91	0
77	19	Survival BB7	5	3.98
78	1	Gaussian	0.9	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
79	1	Gaussian	0.89	0
80	1	Gaussian	0.92	0
81	1	Gaussian	0.92	0
82	1	Gaussian	0.92	0
83	19	Survival BB7	5	3.8
84	19	Survival BB7	5	2.56
85	19	Survival BB7	5	3.84
86	19	Survival BB7	5	4.33
87	19	Survival BB7	5	3.5
88	1	Gaussian	0.89	0
89	1	Gaussian	0.92	0
90	14	Survival Gumbel	4.93	0
91	19	Survival BB7	5	3.6
92	1	Gaussian	0.94	0
93	19	Survival BB7	5	3.89
94	1	Gaussian	0.92	0
95	1	Gaussian	0.9	0
96	1	Gaussian	0.9	0
97	1	Gaussian	0.91	0
98	1	Gaussian	0.93	0
99	19	Survival BB7	5	3.08
100	14	Survival Gumbel	4.9	0

Tabla D.4: Pares de variables aleatorias marginales t-studen $n = 100$, $r = 20$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	19	Survival BB7	5	3.7
2	1	Gaussian	0.91	0
3	1	Gaussian	0.9	0
4	1	Gaussian	0.91	0
5	1	Gaussian	0.9	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
6	1	Gaussian	0.91	0
7	1	Gaussian	0.91	0
8	1	Gaussian	0.92	0
9	1	Gaussian	0.91	0
10	1	Gaussian	0.91	0
11	1	Gaussian	0.93	0
12	1	Gaussian	0.91	0
13	19	Survival BB7	5	3.15
14	1	Gaussian	0.89	0
15	19	Survival BB7	5	3.24
16	19	Survival BB7	5	3.28
17	1	Gaussian	0.9	0
18	1	Gaussian	0.91	0
19	1	Gaussian	0.9	0
20	19	Survival BB7	5	3.51

Tabla D.5: Pares de variables aleatorias marginales t-studen $n = 100$, $r = 40$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	19	Survival BB7	5	3.19
2	19	Survival BB7	5	4.15
3	19	Survival BB7	5	3.04
4	19	Survival BB7	5	3.62
5	19	Survival BB7	5	4.38
6	1	Gaussian	0.92	0
7	19	Survival BB7	5	3.95
8	1	Gaussian	0.91	0
9	19	Survival BB7	5	2.94
10	19	Survival BB7	5	3.23
11	19	Survival BB7	5	3.34
12	1	Gaussian	0.91	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
13	1	Gaussian	0.92	0
14	1	Gaussian	0.91	0
15	19	Survival BB7	5	3.47
16	1	Gaussian	0.91	0
17	19	Survival BB7	5	3.94
18	19	Survival BB7	5	3.8
19	1	Gaussian	0.91	0
20	19	Survival BB7	5	3.76
21	1	Gaussian	0.91	0
22	1	Gaussian	0.91	0
23	19	Survival BB7	5	3.25
24	1	Gaussian	2.99	2.16
25	1	Gaussian	0.91	0
26	1	Gaussian	0.9	0
27	19	Survival BB7	5	2.91
28	19	Survival BB7	5	3.39
29	19	Survival BB7	5	3.31
30	1	Gaussian	0.92	0
31	19	Survival BB7	5	3.74
32	19	Survival BB7	5	3.36
33	1	Gaussian	0.91	0
34	1	Gaussian	0.91	0
35	19	Survival BB7	5	3.17
36	19	Survival BB7	5	3.31
37	1	Gaussian	0.91	0
38	19	Survival BB7	5	3.26
39	19	Survival BB7	5	3.99
40	1	Gaussian	0.9	0

Tabla D.6: Pares de variables aleatorias marginales t-studen $n = 100$, $r = 100$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	19	Survival BB7	5	3.03
2	1	Gaussian	0.9	0
3	1	Gaussian	0.9	0
4	1	Gaussian	0.91	0
5	19	Survival BB7	5	3.63
6	1	Gaussian	0.91	0
7	19	Survival BB7	5	3.06
8	1	Gaussian	0.92	0
9	1	Gaussian	0.91	0
10	19	Survival BB7	5	3.21
11	19	Survival BB7	5	2.86
12	19	Survival BB7	5	3.69
13	19	Survival BB7	5	3.43
14	1	Gaussian	0.91	0
15	9	BB7	2.5	6
16	19	Survival BB7	5	4.03
17	19	Survival BB7	5	3.82
18	19	Survival BB7	5	3.52
19	19	Survival BB7	5	3.99
20	19	Survival BB7	5	3.01
21	1	Gaussian	0.9	0
22	19	Survival BB7	5	3.95
23	1	Gaussian	0.89	0
24	19	Survival BB7	5	3.43
25	1	Gaussian	0.9	0
26	1	Gaussian	0.9	0
27	1	Gaussian	0.91	0
28	1	Gaussian	0.89	0
29	19	Survival BB7	5	3.05

Pares	Familia	Tipo de cópula	θ_1	θ_1
30	1	Gaussian	0.91	0
31	19	Survival BB7	5	3.61
32	1	Gaussian	0.91	0
33	19	Survival BB7	5	4.42
34	1	Gaussian	0.9	0
35	19	Survival BB7	5	3.43
36	19	Survival BB7	5	3.56
37	19	Survival BB7	5	3.07
38	19	Survival BB7	5	3.94
39	19	Survival BB7	5	2.7
40	19	Survival BB7	5	3.98
41	19	Survival BB7	5	3.59
42	1	Gaussian	0.91	0
43	19	Survival BB7	5	3.33
44	19	Survival BB7	5	3.77
45	19	Survival BB7	5	3.26
46	1	Gaussian	0.9	0
47	19	Survival BB7	5	3.29
48	19	Survival BB7	5	2.85
49	1	Gaussian	0.92	0
50	1	Gaussian	0.9	0
51	19	Survival BB7	5	4.04
52	19	Survival BB7	5	4.27
53	19	Survival BB7	5	3.91
54	19	Survival BB7	5	3.98
55	19	Survival BB7	5	2.94
56	19	Survival BB7	5	3.64
57	1	Gaussian	0.91	0
58	19	Survival BB7	5	3.04
59	19	Survival BB7	5	2.93
60	19	Survival BB7	5	3.58

Pares	Familia	Tipo de cópula	θ_1	θ_1
61	1	Gaussian	0.91	0
62	1	Gaussian	0.92	0
63	1	Gaussian	0.9	0
64	1	Gaussian	0.9	0
65	19	Survival BB7	5	4.75
66	1	Gaussian	0.91	0
67	1	Gaussian	0.91	0
68	1	Gaussian	0.91	0
69	1	Gaussian	0.91	0
70	1	Gaussian	0.91	0
71	19	Survival BB7	5	2.91
72	19	Survival BB7	5	3.61
73	19	Survival BB7	5	3.01
74	1	Gaussian	0.91	0
75	19	Survival BB7	5	3.42
76	1	Gaussian	0.92	0
77	19	Survival BB7	5	3.33
78	1	Gaussian	0.92	0
79	19	Survival BB7	5	3.48
80	19	Survival BB7	5	3.17
81	19	Survival BB7	5	3.56
82	19	Survival BB7	5	3.45
83	1	Gaussian	0.91	0
84	1	Gaussian	0.91	0
85	19	Survival BB7	5	3.4
86	19	Survival BB7	5	3.66
87	19	Survival BB7	5	3.67
88	19	Survival BB7	5	3.36
89	19	Survival BB7	5	3.01
90	1	Gaussian	0.91	0
91	19	Survival BB7	5	3.36

Pares	Familia	Tipo de cópula	θ_1	θ_1
92	19	Survival BB7	5	2.95
93	1	Gaussian	0.92	0
94	1	Gaussian	0.91	0
95	19	Survival BB7	5	3.24
96	19	Survival BB7	5	3.17
97	19	Survival BB7	5	3.59
98	19	Survival BB7	5	3.94
99	1	Gaussian	0.91	0
100	19	Survival BB7	5	3.4

Tabla D.7: Pares de variables aleatorias marginales t-studen $n = 500$, $r = 20$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	19	Survival BB7	5	3.23
2	19	Survival BB7	5	3.2
3	19	Survival BB7	5	3.28
4	19	Survival BB7	5	3.09
5	19	Survival BB7	5	3.07
6	19	Survival BB7	5	3.03
7	1	Gaussian	0.91	0
8	19	Survival BB7	5	3.46
9	19	Survival BB7	5	3.78
10	19	Survival BB7	5	3.02
11	19	Survival BB7	5	3.21
12	1	Gaussian	0.91	0
13	1	Gaussian	0.91	0
14	1	Gaussian	0.9	0
15	19	Survival BB7	5	3.01
16	19	Survival BB7	5	3.53
17	1	Gaussian	0.91	0
18	19	Survival BB7	5	3.31

Pares	Familia	Tipo de cópula	θ_1	θ_1
19	1	Gaussian	0.92	0
20	1	Gaussian	0.9	0

Tabla D.8: Pares de variables aleatorias marginales t-studen $n = 500$, $r = 40$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	19	Survival BB7	5	3.26
2	1	Gaussian	0.91	0
3	19	Survival BB7	5	3.23
4	19	Survival BB7	5	2.92
5	1	Gaussian	0.91	0
6	19	Survival BB7	5	3.13
7	1	Gaussian	0.91	0
8	1	Gaussian	0.91	0
9	19	Survival BB7	5	3.3
10	1	Gaussian	0.91	0
11	19	Survival BB7	5	3.21
12	19	Survival BB7	5	3.44
13	1	Gaussian	0.91	0
14	19	Survival BB7	5	3.18
15	1	Gaussian	0.91	0
16	19	Survival BB7	5	3.03
17	1	Gaussian	0.9	0
18	19	Survival BB7	5	3.2
19	1	Gaussian	0.9	0
20	19	Survival BB7	5	3.47
21	19	Survival BB7	5	3.31
22	1	Gaussian	0.91	0
23	1	Gaussian	0.91	0
24	1	Gaussian	0.91	0
25	1	Gaussian	0.91	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
26	1	Gaussian	0.91	0
27	1	Gaussian	0.91	0
28	19	Survival BB7	5	2.92
29	19	Survival BB7	5	3.19
30	1	Gaussian	0.91	0
31	1	Gaussian	0.91	0
32	1	Gaussian	0.91	0
33	1	Gaussian	0.91	0
34	19	Survival BB7	5	3.05
35	1	Gaussian	0.91	0
36	1	Gaussian	0.91	0
37	1	Gaussian	0.91	0
38	1	Gaussian	0.91	0
39	1	Gaussian	0.91	0
40	19	Survival BB7	5	3.11

Tabla D.9: Pares de variables aleatorias marginales t-studen $n = 500$, $r = 100$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	19	Survival BB7	5	3
2	1	Gaussian	0.91	0
3	1	Gaussian	0.91	0
4	19	Survival BB7	5	3.6
5	1	Gaussian	0.91	0
6	1	Gaussian	0.91	0
7	1	Gaussian	0.92	0
8	1	Gaussian	0.91	0
9	1	Gaussian	0.92	0
10	19	Survival BB7	5	3.18
11	1	Gaussian	0.91	0
12	19	Survival BB7	5	3.1

Pares	Familia	Tipo de cópula	θ_1	θ_1
13	1	Gaussian	0.91	0
14	1	Gaussian	0.91	0
15	1	Gaussian	0.91	0
16	19	Survival BB7	5	3.07
17	1	Gaussian	0.91	0
18	19	Survival BB7	5	3.4
19	19	Survival BB7	5	3.24
20	19	Survival BB7	5	3.15
21	1	Gaussian	0.91	0
22	19	Survival BB7	5	3.07
23	19	Survival BB7	5	3.21
24	19	Survival BB7	5	3.37
25	1	Gaussian	0.91	0
26	1	Gaussian	0.91	0
27	19	Survival BB7	5	3.14
28	1	Gaussian	0.91	0
29	1	Gaussian	0.92	0
30	19	Survival BB7	5	3.51
31	19	Survival BB7	5	3.23
32	1	Gaussian	0.92	0
33	1	Gaussian	0.91	0
34	19	Survival BB7	5	3.33
35	19	Survival BB7	5	3.23
36	1	Gaussian	0.91	0
37	19	Survival BB7	5	3.29
38	19	Survival BB7	5	3.36
39	1	Gaussian	0.92	0
40	1	Gaussian	0.92	0
41	1	Gaussian	0.91	0
42	1	Gaussian	0.92	0
43	1	Gaussian	0.91	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
44	1	Gaussian	0.92	0
45	19	Survival BB7	5	3.52
46	19	Survival BB7	5	3.18
47	19	Survival BB7	5	3.27
48	1	Gaussian	0.92	0
49	1	Gaussian	0.91	0
50	1	Gaussian	0.91	0
51	1	Gaussian	0.91	0
52	1	Gaussian	0.91	0
53	1	Gaussian	0.91	0
54	19	Survival BB7	5	3.32
55	1	Gaussian	0.91	0
56	19	Survival BB7	5	3
57	19	Survival BB7	5	2.89
58	1	Gaussian	0.91	0
59	19	Survival BB7	5	3.16
60	19	Survival BB7	5	3.14
61	1	Gaussian	0.91	0
62	19	Survival BB7	5	3.17
63	1	Gaussian	0.91	0
64	1	Gaussian	0.91	0
65	19	Survival BB7	5	3.13
66	19	Survival BB7	5	3.24
67	19	Survival BB7	5	3.05
68	1	Gaussian	0.91	0
69	1	Gaussian	0.91	0
70	1	Gaussian	0.92	0
71	1	Gaussian	0.91	0
72	1	Gaussian	0.91	0
73	1	Gaussian	0.9	0
74	19	Survival BB7	5	3.03

Pares	Familia	Tipo de cópula	θ_1	θ_1
75	19	Survival BB7	5	3.2
76	1	Gaussian	0.92	0
77	19	Survival BB7	5	3.23
78	19	Survival BB7	5	3.26
79	1	Gaussian	0.91	0
80	19	Survival BB7	5	3.03
81	19	Survival BB7	5	3.55
82	19	Survival BB7	5	3.24
83	1	Gaussian	0.91	0
84	1	Gaussian	0.91	0
85	19	Survival BB7	5	3.21
86	19	Survival BB7	5	3.17
87	1	Gaussian	0.91	0
88	19	Survival BB7	5	3.23
89	19	Survival BB7	5	2.95
90	1	Gaussian	0.91	0
91	19	Survival BB7	5	3.14
92	1	Gaussian	0.92	0
93	1	Gaussian	0.91	0
94	1	Gaussian	0.91	0
95	19	Survival BB7	5	3.16
96	1	Gaussian	0.91	0
97	19	Survival BB7	5	3.59
98	19	Survival BB7	5	3.57
99	1	Gaussian	0.91	0
100	1	Gaussian	0.91	0

Apéndice E

Marginales weibull

Tabla E.1: Pares de variables aleatorias marginales weibull $n = 50$, $r = 20$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	5	Frank	3.25	0
2	0	Independence	0	0
3	2	t	0.65	30
4	1	Gaussian	0.49	0
5	9	BB7	3.37	0
6	0	Independence	0	0
7	2	t	0.74	30
8	9	BB7	3.8	0
9	2	t	0.61	18.2
10	2	t	0.54	30
11	9	BB7	3.24	0
12	2	t	0.67	2
13	2	t	0.38	30
14	0	Independence	0	0
15	9	BB7	3.3	0
16	9	BB7	2.85	0
17	9	BB7	3.95	0
18	2	t	0.48	30
19	6	Joe	3.23	0
20	9	BB7	3.1	0

Tabla E.2: Pares de variables aleatorias marginales weibull $n = 50$, $r = 40$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	9	BB7	4.37	0
2	10	BB8	4.41	1
3	6	Joe	3.76	0
4	1	Gaussian	0.5	0
5	19	Survival BB7	1.5	2.5
6	6	Joe	3.34	0
7	9	BB7	3.29	0
8	6	Joe	3.39	0
9	1	Gaussian	0.36	0
10	6	Joe	4.13	0
11	6	Joe	3.53	0
12	6	Joe	3.22	0
13	6	Joe	3.78	0
14	6	Joe	3.06	0
15	1	Gaussian	0.61	0
16	6	Joe	3.78	0
17	5	Frank	3.71	0
18	19	Survival BB7	1.83	3.5
19	19	Survival BB7	1.53	3.81
20	9	BB7	3.45	0
21	6	Joe	3.56	0
22	2	t	0.57	30
23	6	Joe	3.36	0
24	6	Joe	3.54	0
25	5	Frank	3.65	0
26	10	BB8	3.55	1
27	234	Tawn2_270	-2.52	0.37
28	9	BB7	3.05	0.03
29	5	Frank	3.05	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
30	19	Survival BB7	1.35	2.61
31	19	Survival BB7	1.34	3.28
32	2	t	0.55	30
33	5	Frank	4.25	0
34	1	Gaussian	0.49	0
35	0	Independence	0	0
36	1	Gaussian	0.45	0
37	9	BB7	3.28	0.1
38	6	Joe	3.46	0
39	9	BB7	3.43	0
40	9	BB7	3.56	0

Tabla E.3: Pares de variables aleatorias marginales weibull $n = 50$, $r = 100$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	6	Joe	2.98	0
2	2	t	0.59	30
3	2	t	0.46	30
4	2	t	0.45	2
5	124	Tawn1_90	-20	0.04
6	9	BB7	3.68	0
7	10	BB8	3.51	1
8	9	BB7	4.64	0
9	6	Joe	3.22	0
10	6	Joe	3.21	0
11	234	Tawn2_270	-20	0.05
12	19	Survival BB7	1.2	2.97
13	2	t	0.54	30
14	6	Joe	2.82	0
15	2	t	0.37	30
16	10	BB8	2.65	1

Pares	Familia	Tipo de cópula	θ_1	θ_1
17	9	BB7	3.96	0
18	6	Joe	3.4	0
19	1	Gaussian	0.75	0
20	6	Joe	2.95	0
21	6	Joe	3.44	0
22	6	Joe	3.07	0
23	10	BB8	3.4	1
24	6	Joe	3.26	0
25	5	Frank	2.92	0
26	9	BB7	2.96	0
27	2	t	0.54	4.03
28	9	BB7	3.34	0
29	2	t	0.61	30
30	6	Joe	3.84	0
31	0	Independence	0	0
32	0	Independence	0	0
33	9	BB7	3.83	0
34	124	Tawn1_90	-20	0.03
35	6	Joe	3.54	0
36	10	BB8	2.99	1
37	9	BB7	3.44	0
38	124	Tawn1_90	-20	0.03
39	5	Frank	3.1	0
40	6	Joe	3.66	0
41	1	Gaussian	0.64	0
42	124	Tawn1_90	-20	0.05
43	234	Tawn2_270	-20	0.03
44	9	BB7	4.45	0
45	9	BB7	2.96	0
46	6	Joe	3.84	0
47	6	Joe	3.16	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
48	9	BB7	3.39	0
49	2	t	0.6	7.08
50	6	Joe	3.26	0
51	124	Tawn1_90	-18.49	0.01
52	6	Joe	4.35	0
53	5	Frank	3.41	0
54	2	t	0.52	30
55	10	BB8	2.77	1
56	2	t	0.49	30
57	5	Frank	3.38	0
58	6	Joe	2.79	0
59	6	Joe	3.7	0
60	6	Joe	3.24	0
61	6	Joe	3.6	0
62	0	Independence	0	0
63	9	BB7	3.38	0
64	10	BB8	2.7	1
65	6	Joe	3.13	0
66	234	Tawn2_270	-2.82	0.19
67	6	Joe	3.25	0
68	124	Tawn1_90	-8.06	0.03
69	6	Joe	3.99	0
70	9	BB7	3.42	0
71	6	Joe	3.49	0
72	6	Joe	3.49	0
73	124	Tawn1_90	-3.32	0.11
74	2	t	0.5	12.9
75	9	BB7	3.44	0
76	1	Gaussian	0.54	0
77	1	Gaussian	0.52	0
78	1	Gaussian	0.49	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
79	1	Gaussian	0.46	0
80	9	BB7	3.19	0
81	6	Joe	3.64	0
82	234	type2_270	-7.59	0.07
83	5	Frank	3.85	0
84	5	Frank	3.3	0
85	5	Frank	3.69	0
86	6	Joe	3.41	0
87	9	BB7	3.51	0.08
88	5	Frank	3.91	0
89	6	Joe	2.7	0
90	5	Frank	3.83	0
91	6	Joe	3.35	0
92	9	BB7	3.1	0
93	2	t	0.62	30
94	2	t	0.54	30
95	5	Frank	6.93	0
96	9	BB7	2.91	0
97	2	t	0.54	11.79
98	2	t	0.56	30
99	9	BB7	3.09	0
100	5	Frank	3.06	0

Tabla E.4: Pares de variables aleatorias marginales weibull $n = 100$, $r = 20$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	2	t	0.53	30
2	5	Frank	3.55	0
3	6	Joe	3.25	0
4	5	Frank	2.94	0
5	10	BB8	3.38	1

Pares	Familia	Tipo de cópula	θ_1	θ_1
6	5	Frank	2.38	0
7	19	Survival BB7	1.24	3.29
8	6	Joe	3.08	0
9	5	Frank	4.09	0
10	5	Frank	3.5	0
11	5	Frank	4.79	0
12	5	Frank	3.85	0
13	6	Joe	3.45	0
14	9	BB7	3.56	0
15	10	BB8	3.1	1
16	5	Frank	3.65	0
17	5	Frank	3.11	0
18	6	Joe	3.32	0
19	9	BB7	3.11	0
20	6	Joe	3.29	0

Tabla E.5: Pares de variables aleatorias marginales weibull $n = 100$, $r = 40$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	2	t	0.4	30
2	9	BB7	3.78	0.08
3	5	Frank	2.62	0
4	10	BB8	3.67	1
5	1	Gaussian	0.51	0
6	6	Joe	3.63	0
7	6	Joe	3.48	0
8	9	BB7	3.5	0
9	6	Joe	3.49	0
10	9	BB7	3.2	0.08
11	6	Joe	3.26	0
12	5	Frank	4.45	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
13	2	t	0.41	30
14	10	BB8	3.27	1
15	2	t	0.59	5.43
16	10	BB8	2.99	1
17	5	Frank	4.04	0
18	2	t	0.58	4.02
19	2	t	0.58	5.57
20	6	Joe	3.3	0
21	10	BB8	3.57	1
22	6	Joe	3.54	0
23	2	t	0.59	5.45
24	10	BB8	3.73	1
25	6	Joe	3.31	0
26	5	Frank	4.55	0
27	10	BB8	3.45	1
28	5	Frank	3.46	0
29	6	Joe	2.94	0
30	10	BB8	3.29	1
31	9	BB7	2.89	0
32	2	t	0.55	30
33	6	Joe	3.1	0
34	5	Frank	3.04	0
35	5	Frank	3.24	0
36	6	Joe	2.96	0
37	6	Joe	3.15	0
38	2	t	0.5	30
39	10	BB8	3.76	1
40	6	Joe	3.66	0

Tabla E.6: Pares de variables aleatorias marginales weibull $n = 100$, $r = 100$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	1	Gaussian	0.64	0
2	2	t	0.52	30
3	1	Gaussian	0.56	0
4	10	BB8	3.76	1
5	6	Joe	3.24	0
6	6	Joe	3.27	0
7	1	Gaussian	0.34	0
8	5	Frank	2.91	0
9	2	t	0.45	30
10	5	Frank	3.86	0
11	5	Frank	3.7	0
12	6	Joe	3.84	0
13	5	Frank	3.04	0
14	10	BB8	3.59	1
15	9	BB7	3.24	0
16	5	Frank	3.35	0
17	2	t	0.47	30
18	5	Frank	3.62	0
19	9	BB7	3.05	0
20	5	Frank	3.9	0
21	6	Joe	3.01	0
22	9	BB7	3.34	0
23	5	Frank	3.93	0
24	10	BB8	3.23	1
25	9	BB7	3.12	0
26	9	BB7	3.09	0.17
27	5	Frank	4.33	0
28	6	Joe	3.08	0
29	5	Frank	4.19	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
30	5	Frank	3.62	0
31	5	Frank	4.68	0
32	6	Joe	3.73	0
33	5	Frank	3	0
34	5	Frank	3.82	0
35	234	Tawn2_270	-2.57	0.24
36	5	Frank	3.08	0
37	5	Frank	3.9	0
38	2	t	0.61	20.04
39	9	BB7	3.14	0
40	10	BB8	3.09	1
41	10	BB8	3.28	1
42	5	Frank	4.03	0
43	9	BB7	2.95	0
44	2	t	0.43	30
45	0	Independence	0	0
46	5	Frank	3.54	0
47	5	Frank	3.83	0
48	6	Joe	3.05	0
49	2	t	0.55	30
50	5	Frank	3.15	0
51	2	t	0.56	5.36
52	5	Frank	3.98	0
53	6	Joe	3.54	0
54	6	Joe	3.09	0
55	5	Frank	3.5	0
56	10	BB8	3.57	1
57	1	Gaussian	0.51	0
58	5	Frank	3.77	0
59	1	Gaussian	0.62	0
60	5	Frank	3.17	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
61	6	Joe	3.06	0
62	5	Frank	4.62	0
63	5	Frank	4.14	0
64	5	Frank	3.67	0
65	2	t	0.47	30
66	9	BB7	2.87	0
67	2	t	0.49	2.55
68	10	BB8	3.37	1
69	10	BB8	3.34	1
70	5	Frank	3.66	0
71	2	t	0.49	30
72	234	Tawn2_270	-4.01	0.11
73	6	Joe	3.17	0
74	6	Joe	3.55	0
75	6	Joe	3.44	0
76	6	Joe	3.56	0
77	5	Frank	3.38	0
78	6	Joe	3.28	0
79	6	Joe	2.93	0
80	5	Frank	3.41	0
81	6	Joe	3.28	0
82	2	t	0.54	30
83	1	Gaussian	0.56	0
84	6	Joe	3.76	0
85	6	Joe	3.03	0
86	5	Frank	3.74	0
87	5	Frank	4.19	0
88	5	Frank	4.04	0
89	10	BB8	3.6	1
90	6	Joe	3.51	0
91	5	Frank	3.27	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
92	5	Frank	4.69	0
93	6	Joe	3.37	0
94	6	Joe	3.48	0
95	10	BB8	3.27	1
96	6	Joe	2.97	0
97	2	t	0.4	30
98	2	t	0.49	30
99	6	Joe	2.92	0
100	5	Frank	3.6	0

Tabla E.7: Pares de variables aleatorias marginales weibull $n = 500$, $r = 20$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	5	Frank	4.14	0
2	5	Frank	4.19	0
3	5	Frank	4.33	0
4	10	BB8	3.08	1
5	5	Frank	4.01	0
6	6	Joe	3.09	0
7	5	Frank	4.14	0
8	10	BB8	3.05	1
9	5	Frank	3.99	0
10	5	Frank	3.75	0
11	6	Joe	3.13	0
12	5	Frank	3.55	0
13	5	Frank	3.74	0
14	5	Frank	4.16	0
15	5	Frank	3.68	0
16	5	Frank	4.09	0
17	5	Frank	4.3	0
18	6	Joe	3.06	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
19	5	Frank	4.12	0
20	10	BB8	3.13	1

Tabla E.8: Pares de variables aleatorias marginales weibull $n = 500$, $r = 40$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	6	Joe	3.07	0
2	10	BB8	3.41	1
3	5	Frank	3.67	0
4	6	Joe	3.24	0
5	5	Frank	4.09	0
6	5	Frank	3.55	0
7	10	BB8	3.2	1
8	5	Frank	3.56	0
9	6	Joe	3.19	0
10	6	Joe	3.13	0
11	5	Frank	4.16	0
12	5	Frank	3.8	0
13	5	Frank	4.45	0
14	5	Frank	3.97	0
15	5	Frank	4.26	0
16	5	Frank	3.96	0
17	5	Frank	3.8	0
18	6	Joe	3.2	0
19	5	Frank	4.25	0
20	10	BB8	3.12	1
21	5	Frank	3.93	0
22	6	Joe	3.03	0
23	5	Frank	3.75	0
24	5	Frank	4.04	0
25	6	Joe	3.16	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
26	10	BB8	3.41	1
27	5	Frank	3.91	0
28	10	BB8	2.94	1
29	6	Joe	3.12	0
30	5	Frank	3.96	0
31	5	Frank	3.5	0
32	5	Frank	3.99	0
33	5	Frank	3.93	0
34	5	Frank	3.98	0
35	5	Frank	4.04	0
36	5	Frank	3.99	0
37	5	Frank	3.39	0
38	5	Frank	3.7	0
39	5	Frank	4.05	0
40	5	Frank	3.93	0

Tabla E.9: Pares de variables aleatorias marginales weibull $n = 500$, $r = 100$

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	5	Frank	4.36	0
2	6	Joe	3.07	0
3	6	Joe	3.16	0
4	5	Frank	4.04	0
5	5	Frank	3.87	0
6	6	Joe	3.35	0
7	2	t	0.55	30
8	5	Frank	3.87	0
9	10	BB8	3.01	1
10	10	BB8	3.28	1
11	5	Frank	4.22	0
12	5	Frank	3.99	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
13	5	Frank	4.46	0
14	6	Joe	3.07	0
15	6	Joe	3.07	0
16	6	Joe	3.02	0
17	10	BB8	3.03	1
18	5	Frank	4.27	0
19	5	Frank	3.8	0
20	6	Joe	2.98	0
21	5	Frank	3.45	0
22	10	BB8	3.5	1
23	9	BB7	3.34	0
24	5	Frank	4.35	0
25	6	Joe	3.19	0
26	10	BB8	3.2	1
27	6	Joe	3.26	0
28	5	Frank	3.91	0
29	10	BB8	3.26	1
30	5	Frank	3.16	0
31	6	Joe	3.14	0
32	5	Frank	4.08	0
33	10	BB8	3.26	1
34	6	Joe	3.01	0
35	5	Frank	4.1	0
36	6	Joe	3.36	0
37	5	Frank	4.53	0
38	5	Frank	4.11	0
39	6	Joe	3.03	0
40	6	Joe	3.05	0
41	5	Frank	4.32	0
42	10	BB8	3.09	1
43	10	BB8	3.01	1

Pares	Familia	Tipo de cópula	θ_1	θ_1
44	5	Frank	3.37	0
45	10	BB8	3.06	1
46	10	BB8	3.1	1
47	5	Frank	4.42	0
48	5	Frank	4.46	0
49	6	Joe	3.05	0
50	6	Joe	3	0
51	5	Frank	3.56	0
52	10	BB8	3.09	1
53	6	Joe	3.15	0
54	6	Joe	3.27	0
55	5	Frank	3.8	0
56	10	BB8	3.08	1
57	5	Frank	4.06	0
58	5	Frank	4.28	0
59	6	Joe	3.39	0
60	5	Frank	4.35	0
61	5	Frank	4.06	0
62	5	Frank	4.15	0
63	5	Frank	4.16	0
64	5	Frank	3.39	0
65	6	Joe	3.2	0
66	5	Frank	4.06	0
67	10	BB8	3.16	1
68	6	Joe	3.32	0
69	6	Joe	3.01	0
70	6	Joe	3.1	0
71	5	Frank	4.01	0
72	6	Joe	3.02	0
73	5	Frank	3.96	0
74	5	Frank	4.22	0

Pares	Familia	Tipo de cópula	θ_1	θ_1
75	5	Frank	3.82	0
76	10	BB8	3.27	1
77	6	Joe	3.4	0
78	6	Joe	3.02	0
79	6	Joe	3.19	0
80	5	Frank	4.01	0
81	6	Joe	3.06	0
82	2	t	0.49	30
83	5	Frank	4.3	0
84	6	Joe	3	0
85	5	Frank	3.77	0
86	6	Joe	3.15	0
87	10	BB8	3.01	1
88	5	Frank	3.48	0
89	10	BB8	3.11	1
90	10	BB8	3.11	1
91	10	BB8	3.16	1
92	5	Frank	3.83	0
93	5	Frank	4.07	0
94	6	Joe	3.08	0
95	5	Frank	4.02	0
96	5	Frank	3.61	0
97	6	Joe	3.13	0
98	5	Frank	3.75	0
99	5	Frank	3.95	0
100	5	Frank	4.2	0

Apéndice F

Simulación para datos biomédicos

Tabla F.1: aplicación datos biomédicos

Pares	Familia	Tipo de cópula	θ_1	θ_1
1	204	Tawn2	2.06	0.06
2	5	Frank	1.18	0
3	104	Tawn	5.86	0.06
4	26	J90	-1.29	0
5	134	Tawn270	-20	0.01
6	114	Tawn180	2.04	0.08
7	234	Tawn2_270	-20	0.02
8	214	Tawn2_180	3.46	0.2
9	204	Tawn2	2.43	0.09
10	26	J90	-1.73	0
11	134	Tawn270	-1.59	0.08
12	26	J90	-1.28	0
14	204	Tawn2	9.06	0.01
15	26	J90	-1.35	0
16	134	Tawn270	-1.57	0.15
17	26	J90	-1.33	0
18	26	J90	-1.31	0
19	36	J270	-1.12	0

20	114	Tawn180	1.85	0.08
21	26	J90	-1.3	0
22	26	J90	-1.37	0
23	114	Tawn180	2.34	0.1
24	114	Tawn180	3.34	0.07
25	214	Tawn2_180	5.29	0.11
26	5	Frank	0.97	0
27	214	Tawn2_180	20	0.07
28	104	Tawn	16.07	0.03
29	234	Tawn2_270	-1.74	0.22
30	26	J90	-1.43	0
31	204	Tawn2	2.53	0.02
32	26	J90	-1.3	0
33	36	J270	-1.13	0
34	214	Tawn2_180	20	0.1
35	26	J90	-1.47	0
36	26	J90	-1.47	0
37	114	Tawn180	1.92	0.1
38	26	J90	-1.35	0
39	204	Tawn2	20	0.01
40	134	Tawn270	-20	0.01
41	214	Tawn2_180	7.64	0.13
42	36	J270	-1.26	0
43	114	Tawn180	3.49	0.08
44	26	J90	-1.31	0
45	26	J90	-1.31	0
46	6	Joe	1.24	0
47	214	Tawn2_180	16.24	0.1
48	204	Tawn2	20	0.01
49	134	Tawn270	-6.96	0.03
50	26	J90	-1.58	0

Bibliografía

- Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4(none):40 – 79.
- Cintas, R. (2007). Teoría de cópulas y control de riesgo financiero.
- Datapeaker (2022). Técnicas de validación cruzada en aprendizaje automático para un mejor modelo;[consultado el 24 de febrero de 2023]. <https://datapeaker.com/big-data/tecnicas-de-validacion-cruzada-en-aprendizaje-automatico-para-un-mejor-modelo/>.
- Fenech, M., Chang, W. P., Kirsch-Volders, M., Holland, N., Bonassi, S., and Zeiger, E. (2003). Humn project: detailed description of the scoring criteria for the cytokinesis-block micronucleus assay using isolated human lymphocyte cultures. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 534(1-2):65–75.
- Fisher, N. and Switzer, P. (1985). Chi-plots para evaluar la dependencia. *Biometrika*, (2):253–265.
- Gallardo, B. (2010). Teoría de cópulas y aplicaciones en simulación de riesgos financieros y en ingeniería civil. *Doctoral dissertation, Master Thesis, Estadística Aplicada, Universidad de Granada, Granada, España*.
- Genest, C. and Boies, J. (2003). Detecting dependence with kendall plots. *The American Statistician*, (4):275–284.
- Gómez, W. (2017). Análisis de frecuencia hidrológico multivariado para eventos extremos mediante funciones cópula arquimedianas. casos de estudio: cuenca baja del río tunjuelo y región de la mojana (colombia). *Departamento de Ingeniería Civil y Agrícola*.
- Jiménez, L. A. (2018). Criterio de akaike para la selección de modelos con transformaciones. *Universidad Santo Tomás*, pages 10–15.

- Kozak, A. and Kozak, R. (2003). Does cross validation provide additional information in the evaluation of regression models?. *Canadian Journal of Forest Research*, (33(4)):976–987.
- Morales, M. (2013). Aproximación al análisis de correlación de valores extremos entre tipos de cambio de países emergentes.
- Moreno, D. (2012). Método para elegir una cópula arquimediana óptima. *Departamento de Estadística*.
- Nelsen, R. (2006). *An introduction to copulas*. Springer, 2 edition.
- Ortego, M. (2014). Estimación bayesiana de cópulas extremas en procesos de poisson. *Doctoral dissertation, Ph. D. thesis, Universitat Politècnica de Catalunya*.
- Pastor-Sierra, K., Espitia-Pérez, L., Espitia-Pérez, P., Peñata-Taborda, A., Brango, H., Galeano-Páez, C., Bru-Cordero, O. E., Palma-Parra, M., Díaz, S. M., Trillos, C., Briceño, L., Álvaro J. Idrovo, Miranda-Pacheco, J., Téllez, E., Jiménez-Vidal, L., Coneo-Pretelt, A., Álvarez, A. H., Arteaga-Arroyo, G., Ricardo-Caldera, D., Salcedo-Arteaga, S., Porras-Ramírez, A., and Varona-Urbe, M. (2023). Micronuclei frequency and exposure to chemical mixtures in three colombian mining populations. *Science of The Total Environment*, 901:165789.
- Pérez-Planells, L., Delegido, J., Rivera-Caicedo, J. P., and Verrelst, J. (2015). Análisis de métodos de validación cruzada para la obtención robusta de parámetros biofísicos. *Revista de teledetección*, (44):55–65.
- SCIENCE, D. (2020). Validación cruzada k-fold;[consultado el 24 de febrero de 2023]. <https://datascience.eu/es/aprendizaje-automatico/validacion-cruzada-de-k-fold/>.
- Sepúlveda, W. and Garzón, L. (2017). Modelación conjunta de variables meteorológicas medidas en la estación la sirena en el valle del cauca a través de funciones cópula. *Universidad del Valle*.
- Services, A. W. (2022). Validación cruzada;[consultado el 19 de marzo de 2022]. <https://docs.aws.amazon.com/es-es/machine-learning/latest/dg/cross-validation.html>.