


# Desempeño estadístico de un indicador para selección de modelos bivariados de tipo cópula

Julian David Peñaranda Ruiz

Universidad del Valle
Facultad de Ingeniería, Programa Académico de Estadística
Santiago de Cali Colombia
2024 .




# Desempeño estadístico de un indicador para selección de modelos bivariados de tipo cópula

Julian David Peñaranda Ruiz

Trabajo de grado

Director: Ph.D. José Rafael Tovar Cuevas
Codirector: Ph.D. Miguel Angel Marmolejo Lasprilla

Universidad del Valle
Facultad de Ingeniería, Programa Académico de Estadística
Santiago de Cali, Colombia

# Agradecimientos

Este documento representa años de dedicación y esfuerzo en mi carrera. Mi experiencia académica no habría sido la misma sin las personas que me han acompañado a lo largo de mi vida universitaria. Es posible que algunas personas no sean mencionadas en este documento, pero cada una de ellas ha sido fundamental para mi crecimiento profesional y personal.

En primer lugar, deseo expresar mi más profundo agradecimiento a mis tutores de trabajo de grado. Al profesor Rafael Tovar, quien, como director, brindó orientaciones claras y resolvió numerosas dudas sobre los temas abordados en este trabajo. Asimismo, al profesor Miguel Marmolejo, cuyo curso de seminario fue clave para alcanzar los objetivos de esta investigación. Gracias a su pasión por la enseñanza y su forma excepcional de explicar, logré comprender gran parte de la teoría subyacente en este trabajo. También agradezco al cuerpo docente de la carrera, quienes con sus enseñanzas teóricas, prácticas y humanas contribuyeron a mi formación como profesional.

A mis compañeros de estudio y trabajo, Andrés Palomino, Natalia Pérez e Isaac Sánchez, quienes fueron un pilar fundamental a lo largo de mi carrera, ayudándome a comprender con mayor profundidad varios de los temas estudiados. Aprecio también a las amistades que surgieron durante este camino, aquellas personas con quienes compartí momentos inolvidables, en especial a Brandon Paniquita, Andrés Dorado, Kimberly Sarasti, Andrés Olave y Mauricio Ramírez. Gracias a ellos, la experiencia universitaria fue mucho más enriquecedora y amena.

Estoy infinitamente agradecido con mi familia. A mi madre, por su cariño incondicional y su comprensión en los momentos difíciles. A mi padre, quien siempre expresó su orgullo y me motivó a seguir adelante. De manera especial, a mi abuela, cuyo apoyo inquebrantable fue crucial para superar los retos que implicaban mis viajes a la universidad desde un municipio distinto al de la sede principal.

A mi novia, Angee Grijalba, quien fue un pilar emocional y psicológico durante los desafíos académicos y personales. Su apoyo constante y su ayuda en la preparación de mi sustentación fueron determinantes para que todo saliera bien.

Por encima de todo, agradezco a mi hijo, Isaac Peñaranda Gómez, quien me dio la oportunidad de ser padre. Aunque llegó a mi vida en una etapa temprana, su presencia ha sido una experiencia maravillosa. Gracias a él, he crecido como persona, adquirido madurez y aprendido a gestionar mis responsabilidades. No puedo sentirme más orgulloso del niño que es hoy en día: bien portado, comprensivo y lleno de curiosidad. Deseo con todo mi corazón que su vida esté llena de felicidad y éxito, porque sin duda es un niño brillante y especial.

A todos ustedes, gracias.

# Resumen

En este documento se profundizó en el análisis computacional del indicador de bondad de ajuste para funciones cópulas ($I$) bivariadas propuesto por Tovar Cuevas, Portilla Yela, & Achcar (2019), el cual se conforma por la suma de diferencias al cuadrado entre la medida de concordancia empírica y teórica desarrollada por Ledwina (2014). La principal iniciativa de esta investigación fue la busqueda de una prueba libre de distribución para la bondad de ajuste entre conjuntos de datos bivariados y funciones cópulas, por lo cual fueron usados distintos métodos de simulación para recrear los casos de estudio propuesto en Tovar Cuevas et al. (2019) y algunos de sus resultados. A partir de la réplica del indicador $I(model[j]) = \sqrt{\sum_{i=1}^{n}(q_n^j(u_i, v_i) - q_\theta^j(u_i, v_i))^2}$, donde $j$ es la función cópula j-ésima evaluada en el indicador (Tovar Cuevas et al., 2019). El resultado más inmediato del procedimiento mátematico corresponde al cuadrado del indicador ($I^2$), por lo cual se generaron distintas discusiones acerca de la función de distribución de su forma al cuadraro $I^2$ con la cual se esperaba concluir acerca de su raíz $I$. Después de que se hicieran los ajustes a las distintas funciones de distribución candidatas, funciones las cuales tenían comportamientos similiares a la distribución muestral del indicador, en otras palabras, positivas y asimétricas hacia la derecha. Algunas de estas funciones fueron la Gamma de dos y tres parámetros las cuales por las pruebas de Anderson-Darling realizadas a las distribuciones muestrales del indicador para cada caso de estudio, sugirieron que el indicador posiblemente se ajustase a una función Gamma con parámetro de no centralidad en la totalidad de los casos de estudio.

La sugerencia para una prueba de hipótesis fue dada en este documento, donde se propuso una prueba haciendo uso de la misma muestra y simulaciones del indicador (Tovar Cuevas et al., 2019) a partir de la hipótesis nula y la estimación del parámetro resultante de la muestra observada. La propuesta fue hecha con base en otros artículos realizacionados con pruebas de bondad de ajuste de funciones cópulas, especialmente fue propuesto con la investigación realizada para otro indicador desarrollado por Huang & Prokhorov (2014). La prueba propuesta en este documento, generó buenos resultados dentro de la misma función cópula, en otras palabras, si se usa correctamente con un conjunto de datos determinados es poco probable que si los datos provienen de la misma función cópula la prueba rechace la hipótesis nula. Por otro lado, si se usa con funciones cópula distintas, depende de la medición que realice la función cópula, si ambas modelan dependencias lineales débiles o modelan lo mismo, la prueba tiende a no ser confiable en muestras pequeñas. Los resultados son semejantes entre la forma al cuadrado del indicador $I^2$ y su forma escalada $I$ debido a que por propiedades de la probabilidad $P(X^2 \geq x^2) = P(X \geq x)$ si y solo si $x$ es creciente y positiva.



***Palabras clave:***

Bondad de ajuste, Estimador, Prueba de hipótesis, Indicador, Medida de Ledwina, Funciones cópula, Dependencia entre dos variables

# Abstract

*In this document, the computational analysis of the goodness-of-fit indicator for bivariate copula functions ($I$) proposed by Tovar Cuevas et al. (2019) was explored in depth. This indicator consists of the sum of squared differences between the empirical and theoretical concordance measures developed by Ledwina (2014). The main initiative of this research was the search for a distribution-free test for the goodness of fit between bivariate data sets and copula functions. To achieve this, various simulation methods were used to recreate the case studies proposed in Tovar Cuevas et al. (2019) and some of their results. The replication of the indicator* $I(model[j]) = \sqrt{\sum_{i=1}^{n}(q_n^j(u_i, v_i) - q_\theta^j(u_i, v_i))^2}$*, where $j$ is the $j$-th copula function evaluated in the indicator (Tovar Cuevas et al., 2019), was carried out. The most immediate result of the mathematical procedure corresponds to the square of the indicator ($I^2$), which led to various discussions regarding the distribution function of its squared form $I^2$, from which conclusions about its root $I$ were expected. After adjustments were made to the candidate distribution functions—functions that exhibited behaviors similar to the sample distribution of the indicator, that is, positive and right-skewed—a few of these functions, such as the two- and three-parameter Gamma distributions, were tested. The Anderson-Darling tests applied to the sample distributions of the indicator for each case study suggested that the indicator might fit a Gamma function with a non-centrality parameter across all the case studies.*

*A suggestion to hypothesis testing was give in this document. The proposed test uses the same sample and simulations of the indicator (Tovar Cuevas et al., 2019), based on the null hypothesis and the estimation of the resulting parameter from the observed sample. This proposal was inspired by other studies related to goodness-of-fit tests for copula functions, particularly drawing from research on another indicator developed by Huang & Prokhorov (2014). The test proposed in this document produced good results within the same copula function. In other words, if used correctly with a given data set, it is unlikely that the test will reject the null hypothesis if the data originates from the same copula function. On the other hand, if used with different copula functions, the test's reliability depends on how the copula functions measure dependence. If both copulas model weak linear dependencies or the same type of dependence, the test tends to be unreliable in small samples. The results are similar between the squared form of the indicator $I^2$ and its scaled form $I$, due to the probability property $P(X^2 \geq x^2) = P(X \geq x)$ if and only if $x$ is increasing and positive.*



***Keywords:***

*Goodness-of-fit, Estimator, Hypothesis test, Indicator, Ledwina measure, Copula functions, Dependence between two variables*

# Índice

**Agradecimientos** **3**

**Resumen** **4**

**1. Introducción** **12**
1.1. Plantamiento del problema . . . 13
1.2. Pregunta del investigación . . . 14
1.3. Justifación . . . 14
1.4. Objetivo General . . . 15
1.4.1. Objetivos Específicos . . . 15
1.5. Antecendentes . . . 15
1.5.1. Pruebas alternativas (Genest & Rémillard, 2008) y (Genest, Rémillard, & Beaudoin, 2009): . . . 16
1.5.2. Coeficiente de concordancia de Ledwina (Ledwina, 2014), (Swanepoel & Allison, 2013): . . . 16
1.5.3. Método para seleccionar funciones cópulas (Portilla Yela & Tovar Cuevas, 2018; Tovar Cuevas et al., 2019): . . . 16

**2. Marco teórico** **18**
2.1. Probabílidad y estadística . . . 18
2.1.1. Convergencia en distribución . . . 18
2.1.2. Transformada integral de probabílidad . . . 18
2.1.3. Bootstrap . . . 19
2.1.4. Prueba de hipótesis . . . 20
2.1.5. Prueba de Anderson-Darling . . . 21
2.2. Funciones cópula . . . 22
2.2.1. Teorema de Sklar . . . 22
2.2.2. Método de simulación haciendo uso de la transformada inversa. (Mai & Scherer, 2013, p. 23) . . . 23
2.2.3. Método de simulación haciendo uso de la función de distribución de la función característica Arquimediana. (Mai & Scherer, 2013, p. 61) . . . 24
2.2.4. Función de distribución empírica (cópula empírica) . . . 24
2.2.5. Coeficiente de concordancia: $\tau$ de Kendall . . . 25
2.2.6. Estimación de parámetros para funciones cópula . . . 25
2.2.7. Funciones cópula de estudio. . . . 26
2.2.8. Coeficiente de concordancia: $\boldsymbol{q}$ de Ledwina . . . 31
2.3. Métodos de selección de funciones cópula . . . 33
2.3.1. Propuesta de Tovar, Portilla y Achcar . . . 33

**3. Metodología** **35**
3.1. Esquema de estudio . . . 35
3.2. Generación de réplicas del indicador . . . 38
3.2.1. Réplica del estudio principal (Tovar Cuevas et al., 2019) . . . 38

3.2.2. Generación del indicador . . . 38
3.2.3. Ajuste del indicador a una posible función de distribución . . . 40
3.3. Calidad del estimador $I^2$ como medida de la bondad de ajuste de los datos a una función cópula . . . 40
3.3.1. Sesgo en las estimaciones . . . 42
3.3.2. Error cuadrático medio en las estimaciones . . . 43
3.3.3. Robustez . . . 43
3.4. Prueba de hipótesis . . . 44
3.4.1. Confianza de la prueba . . . 47
3.4.2. Potencia de la prueba . . . 49

**4. Resultados** **52**
4.1. Análisis de la distribución del indicador . . . 52
4.1.1. Función cópula Clayton: Análisis exploratorio $m = 1000$ réplicas de $I^2$. . . . 52
4.1.2. Función cópula FGM: Análisis exploratorio $m = 1000$ réplicas de $I^2$. . . . 54
4.1.3. Función cópula Frank: Análisis exploratorio $m = 1000$ réplicas de $I^2$. . . . 56
4.1.4. Función cópula Gumbel-Barnett: Análisis exploratorio $m = 1000$ réplicas de $I^2$. . . . 57
4.1.5. Función cópula Gumbel-Hougaard: Análisis exploratorio $m = 1000$ réplicas de $I^2$. . . . 59
4.1.6. Prueba de Anderson-Darling: Validación del análisis exploratorio de las simulaciones. . . . 61
4.2. Evaluación de la calidad de $I^2$ como estimador de la bondad de ajuste entre conjuntos de datos y funciones cópulas. . . . 62
4.2.1. Función cópula Clayton: Evaluación del sesgo y el error cuadrático medio para evaluar la eficiencia y la consistencia en la estimación. . . . 62
4.2.2. Función cópula FGM: Evaluación del sesgo y el error cuadrático medio para evaluar la eficiencia y la consistencia en la estimación. 63
4.2.3. Función cópula Frank: Evaluación del sesgo y el error cuadrático medio para evaluar la eficiencia y la consistencia en la estimación. 64
4.2.4. Función Cópula Gumbel-Barnett: Evaluación del sesgo y el error cuadrático medio para evaluar la eficiencia y la consistencia en la estimación. . . . 65
4.2.5. Función cópula Gumbel-Hougaard: Evaluación del sesgo y el error cuadrático medio para evaluar la eficiencia y la consistencia en la estimación. . . . 66
4.3. Confianza y potencia de la prueba de hipótesis basada en el indicador $I^2$ 67
4.3.1. Función cópula Clayton: confianza y potencia de la prueba de hipótesis. . . . 68
4.3.2. Función cópula FGM: confianza y potencia de la prueba de hipótesis. . . . 69

4.3.3. Función cópula Frank: confianza y potencia de la prueba de hipótesis. . . . 70
4.3.4. Función cópula Gumbel-Barnett: confianza y potencia de la prueba de hipótesis. . . . 70
4.3.5. Función cópula Gumbel-Hougaard: confianza y potencia de la prueba de hipótesis. . . . 71

**5. Conclusiones y recomendaciones** **74**
5.1. Conclusiones . . . 74
5.2. Recomendaciones . . . 75

**6. Referencias** **77**

**A. Anexos** **79**
A.1. Demostraciones . . . 79
A.1.1. Aproximación de la esperanza de una función cópula . . . 79
A.2. Propiedades y distribuciones asintóticas . . . 81
A.2.1. Coeficiente de concordancia de Ledwina . . . 81
A.2.2. Indicador de bondad de ajuste (Tovar Cuevas et al., 2019) . . . 82
A.3. Estimaciones de las funciones candidatas para la distribución del indicador $I^2$ . . . 83
A.3.1. Robustez del indicador $I^2$ . . . 85
A.4. Pseudo código . . . 87
A.4.1. Generación de la prueba y réplicas de $I^2$ a partir de la muestra. 87

# Índice de figuras

1. Diagrama de dispersión de la función copula Clayton. Fuente: Elaboración propia . . . 27
2. Diagrama de dispersión de la función copula FGM. Fuente: Elaboración propia . . . 28
3. Diagrama de dispersión de la función copula Frank. Fuente: Elaboración propia . . . 29
4. Diagrama de dispersión de la función copula Gumbel-Barnett. Fuente: Elaboración propia . . . 30
5. Diagrama de dispersión de la función copula Gumbel-Hougaard. Fuente: Elaboración propia . . . 31
6. Metodología de estudio. Fuente: Elaboración propia . . . 37
7. Proceso de remuestreo, donde la función densidad superior es la muestral y las 4 inferiores son muestras bootstrap de la misma. Fuente: Elaboración propia . . . 42
8. Cálculo de la confianza de la prueba haciendo uso de técnicas de aleatorización y remuestreo (muestra de m réplicas del indicador $I^2$ para una función cópula FGM con tamaño de muestra 100 y parámetro de dependencia $\theta = 0.2$). Fuente: Elaboración propia . . . 48
9. Cálculo de la potencia de la prueba haciendo uso de técnicas de aleatorización y remuestreo (muestra de m réplicas del indicador $I^2$ para una función cópula FGM con tamaño de muestra 100 y parámetro de dependencia $\theta = 0.2$ con $\tau_\theta = 0.04...$, frente a una muestra aleatoria de tamaño de muestra 100 y parámetro de dependecia $\theta = 0.093$ con $\tau_\theta = 0.04...$ ). Fuente: Elaboración propia . . . 50
10. Distribución del indicador $I^2$: Clayton dependencia débil. Fuente: Elaboración propia . . . 53
11. Distribución del indicador $I^2$: Clayton dependencia moderada. Fuente: Elaboración propia . . . 53
12. Distribución del indicador $I^2$: Clayton dependencia fuerte. Fuente: Elaboración propia . . . 54
13. Distribución del indicador $I^2$: FGM dependencia débil. Fuente: Elaboración propia . . . 54
14. Distribución del indicador $I^2$: FGM dependencia moderada. Fuente: Elaboración propia . . . 55
15. Distribución del indicador $I^2$: FGM dependencia fuerte. Fuente: Elaboración propia . . . 55
16. Distribución del indicador $I^2$: Frank dependencia débil. Fuente: Elaboración propia . . . 56
17. Distribución del indicador $I^2$: Frank dependencia moderada. Fuente: Elaboración propia . . . 56
18. Distribución del indicador $I^2$: Frank dependencia Fuerte. Fuente: Elaboración propia . . . 57

19. Distribución del indicador $I^2$: Gumbel-Barnett dependencia débil. Fuente: Elaboración propia . . . . . 58
20. Distribución del indicador $I^2$: Gumbel-Barnett dependencia moderada. Fuente: Elaboración propia . . . . . 58
21. Distribución del indicador $I^2$: Gumbel-Barnett dependencia fuerte. Fuente: Elaboración propia . . . . . 59
22. Distribución del indicador $I^2$: Gumbel-Hougaard dependencia débil. Fuente: Elaboración propia . . . . . 59
23. Distribución del indicador $I^2$: Gumbel-Hougaard dependencia moderada. Fuente: Elaboración propia . . . . . 60
24. Distribución del indicador $I^2$: Gumbel-Hougaard dependencia fuerte. Fuente: Elaboración propia . . . . . 60
25. Estimaciones del error cuadrático medio y del sesgo del indicador $I^2$ para la función cópula Clayton. Fuente: Elaboración propia . . . . . 63
26. Estimaciones del error cuadrático medio y del sesgo del indicador $I^2$ para la función cópula FGM. Fuente: Elaboración propia . . . . . 64
27. Estimaciones del error cuadrático medio y del sesgo del indicador $I^2$ para la función cópula Frank. Fuente: Elaboración propia . . . . . 65
28. Estimaciones del error cuadrático medio y del sesgo del indicador $I^2$ para la función cópula Gumbel-Barnett. Fuente: Elaboración propia . 66
29. Estimaciones del error cuadrático medio y del sesgo del indicador $I$ función cópula Gumbel-Hougaard. Fuente: Elaboración propia . . . . . 67
30. Representación del cuadrado unitaro. Fuente: Elaboración propia . . . 79
31. Teorema del trapecio en $\mathbb{R}^2$ con marginal en $u$. Fuente: Elaboración propia 80
32. Teorema del trapecio en $\mathbb{R}^2$ con marginal en $v$. Fuente: Elaboración propia 80
33. Teorema del trapecio en $\mathbb{R}^2$ en $u$ y $v$. Fuente: Elaboración propia . . . 81
34. Diferencia promedio entre indicadores contaminados con una función cópula Gumble-Hougaard para la función cópula Clayton, con niveles de contaminación $\delta = 0.01, 0.05, 0.1, 0.15, 0.2$. Fuente: Elaboración propia 85
35. Diferencia promedio entre indicadores contaminados con una función cópula Frank para la función cópula FGM, con niveles de contaminación $\delta = 0.01, 0.05, 0.1, 0.15, 0.2$. Fuente: Elaboración propia . . . . . 85
36. Diferencia promedio entre indicadores contaminados con una función cópula Gumbel-Hougaard para la función cópula Frank, con niveles de contaminación $\delta = 0.01, 0.05, 0.1, 0.15, 0.2$. Fuente: Elaboración propia 86
37. Diferencia promedio entre indicadores contaminados con una función cópula Frank para la función cópula Gumbel-Barnett, con niveles de contaminación $\delta = 0.01, 0.05, 0.1, 0.15, 0.2$. Fuente: Elaboración propia 86
38. Diferencia promedio entre indicadores contaminados con una función cópula Gumbel-Hougaard para la función cópula Clayton, con niveles de contaminación $\delta = 0.01, 0.05, 0.1, 0.15, 0.2$. Fuente: Elaboración propia 87

# Lista de tablas

1. Prueba de Anderson-Darling. Ya que el nivel de significancia se fija como $\alpha = 0.05$, tome todo valor menor a este como un no se distribuye bajo esta función y todo valor mayor a este como que se distribuye bajo la función. . . . . 61
2. Confianza y potencia de la prueba de hipótesis cuando $H_0 : (U, V) \sim Clayton_\theta$, $H_a : (U, V) \not\sim Clayton_\theta$ . . . . . 68
3. Confianza y potencia de la prueba de hipótesis cuando $H_0 : (U, V) \sim FGM_\theta$, $H_a : (U, V) \not\sim FGM_\theta$ . . . . . 69
4. Confianza y potencia de la prueba de hipótesis cuando $H_0 : (U, V) \sim Frank_\theta$, $H_a : (U, V) \not\sim Frank_\theta$ . . . . . 70
5. Confianza y potencia de la prueba de hipótesis cuando $H_0 : (U, V) \sim Gumbel - Barnett_\theta$, $H_a : (U, V) \not\sim Gumbel - Barnett_\theta$ . . . . . 71
6. Confianza y potencia de la prueba de hipótesis cuando $H_0 : (U, V) \sim Gumbel - Hougaard_\theta$, $H_a : (U, V) \not\sim Gumbel - Hougaard_\theta$ . . . . . 73
7. Valor estimado de los parámetros de las distribuciones candidatas . . . 84

# 1. Introducción

En diversas áreas del conocimiento, como la medicina, la economía, la ingeniería y la ecología, es común trabajar con conjuntos de datos que contienen errores aleatorios, ya sea debido a las herramientas de medición, efectos naturales o errores humanos. Estos datos pueden presentar diversas propiedades, como independencia, correlación o autocorrelación entre eventos. Una estructura de dependencia en los datos se refiere a un comportamiento común entre dos o más variables de medición, y puede modelarse utilizando diversas estructuras estadísticas. Entre las técnicas estadísticas más utilizadas para abordar estos fenómenos se encuentran los modelos de riesgo, los coeficientes de correlación, los coeficientes de concordancia y las funciones cópula (Nelsen, 2006), entre otras.

Este trabajo se enfocó en comprender un poco más las funciones cópula bivarias las cuales han sido ampliamente estudiadas en la literatura (Genest et al., 2009; Ledwina, 2014; Tovar Cuevas et al., 2019). Las funciones cópula bivariadas se utilizan para vincular las funciones de distribució de 2 variables aleatorias las cuales provienen de marginales $F_X(x), F_Y(y)$, esto se hace por medio de una relación directa entre la función de distribución conjunta evaluada en las inversas de las marginales y la función cópula (Van Vliet, 2023). Esta relación esta definida bajo el teorema de **Sklar**, pues, $C_{\boldsymbol{\theta}}(u,v) = F_{\boldsymbol{\theta}}(F_X^{-1}(u), F_Y^{-1}(v))$, donde la transformación permite ver estructuras de dependencias entre las variables aleatorias $X$ y $Y$ por medio de probabilidades condicionales, donde $\frac{\partial C_{\boldsymbol{\theta}(u,v)}}{\partial u} = F_{V|U}(v \mid u = u_0, \boldsymbol{\theta})$ o $\frac{\partial C_{\boldsymbol{\theta}(u,v)}}{\partial v} = F_{U|V}(u \mid v = v_0, \boldsymbol{\theta})$ (Mai & Scherer, 2013), aunque no es directo, es necesario recordar que $U$ y $V$ se extraen de la transformada integral de probabilidad de $X$ y $Y$ (Durante & Sempi, 2015).

Las funciones cópulas, al igual que las funciones de distribución, cuentan con diferentes formas de medir su bondad de ajuste frente a conjuntos de datos. Estas formas se encuentran documentados en varios trabajos como Genest et al. (2009); Huang & Prokhorov (2014); Tovar Cuevas et al. (2019). El indicador propuesto en el artículo de Tovar Cuevas et al. (2019) es un estadístico de bondad de ajuste diseñado para funciones cópulas bivariadas, que depende de dos funciones marginales de un conjunto de datos, una función cópula y la función de distribución conjunta empírica de las marginales. Este indicador se define en términos de la medida de concordancia propuesta por Ledwina (2014), que describe la relación de crecimiento o decrecimiento conjunto entre las variables aleatorias. El estadístico de Tovar Cuevas et al. (2019) se plantea como la distancia euclidiana entre la medida de concordancia muestral y teórica de los datos: $I_j = \sqrt{\sum_{i=1}^{n}(q_n^j(F_X(x_i), F_Y(y_i)) - q_\theta^j(F_X(x_i), F_Y(y_i)))^2}$ (Tovar Cuevas et al., 2019, p.73), donde $q_n^j(\cdot,\cdot)$ es la función de concordancia muestral y $q_\theta^j(\cdot,\cdot)$ es la concordancia teórica asociada a la función cópula j-ésima.

En el estudio realizado por Tovar Cuevas, Portilla Yela, & Achcar (2019) no se profundizó en una distribución la cual permitiera realizar pruebas de bondad de ajuste en el función del indicador $I_j$, sino que se realizó un estudio empírico a través del análisis conjunto de varias funciones cópula, donde se cuantificó el éxito de seleccionar bien

la función cópula simulada frente a otras cinco funciones cópula. Esto haciendo uso del mín$\{I_j\}$ $j \in \{1, \ldots, 5\}$ donde j corresponde a la función cópula j-ésima evaluada en el conjunto de datos. La propuesta en este trabajo fue estudiar el indicador como un estimador de bondad de ajuste, por lo cual se le estimó una posible función de distribución de probabilidad por medio de métodos empíricos (histogramas, pruebas de bondad de ajuste y simulación), también se le midió sus cualidades como estimador de bondad de ajuste (sesgo, error cuadrático medio, consistencia y robustez) y se propuso una prueba de hipótesis la cual fuera réplicable e independiente de la función cópula que se deseara estudiar (con su cálculo respectivo de confianza y potencia de la prueba).

Al no encontrarse una función de distribución que se ajustara adecuadamente al comportamiento del indicador, y dado que las estimaciones de los parámetros no eran similares entre las distintas funciones cópulas y sus niveles de dependencia, se optó por realizar una prueba basada en aleatorización y remuestreo (Bootstrap no paramétrico). Este enfoque se fundamenta en estudios previos realizados por Genest et al. (2009) y Huang & Prokhorov (2014). Como resultado general, cuando las funciones cópulas capturan comportamientos distintos en las marginales, como por ejemplo, la modelación de colas pesadas frente a relaciones lineales, la prueba de bondad de ajuste propuesta en este trabajo demuestra buena potencia para distinguir entre funciones cópulas.

## 1.1. Plantamiento del problema

La problemática de la especificidad del modelo ha sido un tema recurrente en la estadística, afectando diversas ramas como la inferencia clásica, la inferencia bayesiana, el modelado lineal, el modelado no lineal y las series temporales, entre otras. Estas áreas cuentan con un conjunto de pruebas formales y gráficas para evaluar si un modelo se ajusta correctamente a los datos o si, por el contrario, presenta un mal ajuste. Las pruebas más comunes para medir la bondad de ajuste entre un conjunto de datos y una función de distribución están la prueba de Kolmogorov-Smirnov y la prueba de Anderson-Darling (Sheskin, 2003).

Las funciones cópulas modelan la estructura de dependencia entre conjuntos de variables. Una medida de dependencia puede entenderse como la influencia que una función marginal ejerce sobre la otra en el caso bivariado. Este fenómeno se observa teóricamente cuando se expresa la función de densidad conjunta en términos de la función cópula: $f_{X,Y}(x, y) = f_X(x) f_Y(y) c_\theta(F_X(x), F_Y(y))$. El estudio de la dependencia entre dos variables aleatorias ha sido ampliamente investigado. En investigaciones más recientes, como el trabajo de Ledwina (2014), se introduce una medida de concordancia $q_\theta = \frac{C_\theta(u,v)-uv}{w}$, donde $uv$ es la cópula independiente y $w = \sqrt{u(1-u)v(1-v)}$, que representa la varianza de la estimación de la función cópula $V[C_n(u, v)] = \frac{u(1-u)v(1-v)}{n}$, utilizando los resultados de Swanepoel & Allison (2013).

En estudios recientes, Tovar Cuevas et al. (2019) propone un indicador de bondad de ajuste para la selección de modelos bivariados de tipo cópula, basado en la medida de concordancia planteada por Ledwina (2014) y en el uso del estadístico de Cramér-Von Mises (Sheskin, 2003). Este estudio cuantificó el éxito al clasificar correctamente los modelos de tipo cópula (confianza del indicador) y lo comparó con el trabajo previo realizado por Genest, Quessy, & Rémillard (2006). Para ello, se realizaron comparaciones entre cinco funciones cópulas (Clayton, Frank, Gumbel-Hougaard, Farlie-Gumbel-Morgenstern, Gumbel-Barnett), estimando el parámetro mediante métodos bayesianos y clásicos, utilizando el $\tau$ de Kendall para estimar el parámetro de la función cópula evaluada.

El estudio realizado por Tovar Cuevas, Portilla Yela, & Achcar (2019) fue basado en críterios empíricos, por lo cual, solo se cuantificó la probabilidad de seleccionar correctamente el modelo si este era puesto a prueba con otros cinco. Sin embargo, en dicho estudio no se consideró asignar una función de distribución al indicador ni evaluar la calidad de este como un estimador de bondad de ajuste. Esto no hace posible realizar una prueba de hipótesis para bondad de ajuste de datos a funciones cópula bivariadas, sino que limita a establecer modelos candidatos para poder realizar el críterio empírico donde se encuentre el modelo que cumpla que $I_j$ sea el mínimo entre todas las demás funciones cópula.

## 1.2. Pregunta del investigación

- Partiendo de la teoría de la estimación, ¿cuáles son las propiedades estadísticas del indicador de bonda de ajuste construido por Tovar Cuevas, Portilla Yela, & Achcar (2019), cuando este es visto como un estimador?

## 1.3. Justifación

La selección del modelo correcto en la industria juega un papel fundamental debido a la necesidad de mejorar la precisión en las estimaciones y predicciones. Cometer errores en la especificación de modelos estadísticos puede generar sesgos y varianzas elevadas en las estimaciones, lo que conlleva problemas de inexactitud e imprecisión, además de malas predicciones.

Las funciones cópula, al ser relativamente recientes en comparación con otros campos estadísticos, han despertado gran interés tanto en sectores industriales como académicos. Un ejemplo es el sector financiero (Bouye, Durlleman, Nikeghbali, Riboulet, & Roncalli, 2000), donde las funciones cópula se han integrado en múltiples ramas de la estadística, como la inferencia estadística aplicada a estructuras de datos multivariados (Pappadà, Durante, & Torelli, 2018), el análisis de series temporales para modelar errores heterocedásticos (Zhang, Zhou, & Lin, 2021), y el reconocimiento de imágenes mediante métodos de machine learning (Tsanakas & Zhu, 2022), buscando una mejor estandarización de imágenes. Estos campos requieren una correcta especificación del modelo, por lo que varios autores han propuesto pruebas para funciones cópula en espacios bivariados (Genest et al., 2009).

El trabajo de Genest et al. (2006) fue uno de los primeros en abordar la selección de funciones cópula utilizando el estadístico de Cramer-Von Mises (Sheskin, 2003), realizando una comparativa de la precisión con la que el estadístico de bondad de ajuste acertaba el modelo cópula empírico. Por su parte, Ledwina (2014) propusieron una medida de concordancia en el intervalo $[-1, 1]$, similar a otras medidas estadísticas como el coeficiente de correlación de Pearson, el coeficiente de Spearman y el $\tau$ de Kendall, siendo estas últimas también medidas de concordancia (Nelsen, 2006). El indicador propuesto por Tovar Cuevas et al. (2019) utiliza el estadístico de concordancia desarrollado por Ledwina (2014), basado en el estadístico de Cramer-Von Mises, el cual es ampliamente utilizado en estudios de bondad de ajuste para funciones cópula (Genest et al., 2009). La medida de concordancia propuesta por Ledwina (2014) otorga mayor visibilidad a las colas de la distribución, dado que el denominador incluye términos como $w = \sqrt{uv(1-u)(1-v)}$, los cuales amplifican la medida cuando $u$ o $v \to 0$ o $u$ o $v \to 1$. El estudio de este estimador es crucial, ya que si presenta sesgo o varianza elevada, la estimación resultaría inexacta o imprecisa, lo que generaría una incorrecta especificación en los resultados y las inferencias.

## 1.4. Objetivo General

Desarrollar un estudio de simulación que permita evaluar el desempeño estadístico del indicador de bondad de ajuste construido por Tovar Cuevas et al. (2019) para selección de modelos cópula.

### 1.4.1. Objetivos Específicos

- Establecer de forma empírica la distribución muestral del indicador asumiendolo como un estimador.
- Evaluar de forma empírica el desempeño estadístico del indicador como un estimador de la bondad de ajuste de los datos a la función cópula.
- Construir un procedimiento de prueba de hipótesis para evaluar de forma empírica la bondad de ajuste.

## 1.5. Antecendentes

En esta sección se presentan los antecedentes de este trabajo de grado, resumiendo una colección de artículos publicados previamente por varios autores. Los dos primeros se refieren a investigaciones sobre pruebas de bondad de ajuste para modelos cópula, abordadas de manera general. El segundo trabajo analiza la potencia de varias pruebas recopiladas para funciones cópula. En segundo lugar, se incluyen dos trabajos que validan el indicador utilizado en este estudio y analizan algunas de sus propiedades. Por último, los dos artículos finales sustentan la lógica de este trabajo y algunos de los procedimientos empleados.

### 1.5.1. Pruebas alternativas (Genest & Rémillard, 2008) y (Genest et al., 2009):

En Genest & Rémillard (2008) se describe un procedimiento de ajuste adecuado para modelos basados en la transformada integral de probabilidad o transformada inversa. En este artículo, proponen una prueba de bondad de ajuste para funciones cópulas de la familia Arquimediana, debido a que estas poseen una función arquimediana característica, la cual se utiliza para comparar los coeficientes de concordancia del $\tau$ de Kendall empírico y teórico. En sus resultados, se encuentra la comparativa empírica entre los estadísticos de bondad de ajuste propuestos, uno utilizando la medida de Cramér-von Mises y otro la de Kolmogorov-Smirnov. Por otro lado, en Genest et al. (2009) se realizó un estudio de potencia referente a las distintas pruebas de bondad de ajuste para modelos bivariados de tipo cópula. Los resultados muestran que el comportamiento de las pruebas no es estable y que estas mejoran su potencia con tamaños de muestra muy grandes, esto debido a que todas son extraídas de métodos bootstrap.

### 1.5.2. Coeficiente de concordancia de Ledwina (Ledwina, 2014), (Swanepoel & Allison, 2013):

En Ledwina (2014) se presenta el estudio base para el coeficiente de concordancia utilizado en este trabajo, el cual establece una relación de orden basada en dos funciones a trozos idénticas para pares de variables uniformes. Al igual que muchos coeficientes de concordancia, incluye una prueba de hipótesis para medir la independencia entre las variables, basada en la normalidad si y solo si $C_n(u, v) \to uv$, es decir, cuando la cópula empírica es muy cercana a la cópula de independencia. La normalidad de los resultados está respaldada por el estudio de Swanepoel & Allison (2013), en el cual se detalla el valor esperado de la estimación de la cópula empírica y su varianza. Se concluye que si la cópula empírica corresponde a la independencia, entonces su esperanza es $E[C_n] = uv$ y su varianza es $V[C_n] = \frac{1}{n-1}uv(1-u)(1-v)$, resultados que utiliza Ledwina para construir su estadístico de prueba $q_n = \frac{C_n(u,v)-uv}{\sqrt{uv(1-u)(1-v)}}$.

### 1.5.3. Método para seleccionar funciones cópulas (Portilla Yela & Tovar Cuevas, 2018; Tovar Cuevas et al., 2019):

La cópula Gumbel-Barnett fue definida en Gumbel (1960), donde se introdujo una función exponencial con marginales exponenciales, lo que alude a una posible generalización de la función exponencial n-variada. Esta distribución también se encuentra registrada en Nelsen (2006), donde, se muestra como un tipo de función de distribución con pseudo-inversa y nombrada como **Gumbel-Barnett**, por otro lado, en Portilla Yela & Tovar Cuevas (2018), se propusieron distintos métodos para la estimación del parámetro de la Gumbel-Barnett, como métodos clásicos y bayesianos.

Los resultados del artículo de Tovar Cuevas et al. (2019) son fundamentales en este trabajo, ya que proporcionan parte de la metodología implementada, los casos de simulación, los niveles de dependencia y las familias de funciones cópulas a estudiar.

Los resultados revelan que el indicador $I(\text{model}[j])$ no siempre es el mejor para evaluar la bondad de ajuste entre funciones cópulas y que su confianza aumenta conforme el tamaño de muestra crece, lo cual es consistente con los estudios revisados en la Sección 1.5.1.

# 2. Marco teórico

Esta sección del documento se compone de herramientas y definiciones las cuales se pueden tomar como necesarias y desconocidas para el lector, también, se toma en consideración que algunos conceptos de probabílidad y estadística como la definición de muestra aleatoria, los momentos, rangos y demás son conocidas por el lector de este texto.

## 2.1. Probabílidad y estadística

Los conceptos de probabilidad y estadísta que se darán a continuación están estrechamente ligados con el uso que se le dará para la construcción y compresión de los objetivos basados en las funciones cópulas. Por un lado, conceptos como la convergencia en probabilidad y la transformada intregral tienen un peso netamente probabilístico. Por otro lado, las formas de estimación de parámetros (como el remuestreo) y las pruebas de bondad de ajuste para funciones de distribución muestral univariado poseen un enfoque estadístico.

### 2.1.1. Convergencia en distribución

En (Casella & Berger, 2001, p. 235) se encuentra la definición de la convergencia en distribución. Dada una secuencia de variables aleatorias $X_1, X_2, \ldots$ . $\boldsymbol{X}_n$ convergen en distribución a una variable aleatoria $\boldsymbol{X}$ si se cumple que

$$\lim_{n\to\infty} F_{\boldsymbol{X_n}}(\boldsymbol{x}) = F_{\boldsymbol{X}}(\boldsymbol{x}),$$

para todo punto de $\boldsymbol{x}$ donde $F_{\boldsymbol{X}}(\boldsymbol{x})$ es continua.

### 2.1.2. Transformada integral de probabílidad

La transformada integral de probabilidad, se define en (Durante & Sempi, 2015, p. 7) como la función inversa de la función de una distribución, en otras palabras. Dado $X$ una variable aleatoria sobre el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$[1] con una función de distribución asociada $F_X(\cdot)$ y la inversa de la función de distribución $F_X^{-1}(\cdot)$

$$F_X(F_X^{-1}(u)) = F_X(inf\{x \in \mathbb{R} : F_X(x) \geq u\}) = F_X(\tilde{x}) = u.$$

Esta propiedad atribuye a que cualquier función de distribución acumulada, pueda ser vista como una variable aleatoria uniforme en el espacio [0,1], pues, en las Ecuaciones 1 y 2 respectivamente, se muestra las relaciones de transición entre rangos de variables.

$$F_X(x) = \mathbb{P}(X \leq x) = u \tag{1}$$

$$F_X^{-1}(u) = \mathbb{P}(U \leq F_X(x)) = x \tag{2}$$

[1] Se recuerda que $\Omega$ es el espacio de eventos, $\mathcal{F}$ es la sigma-álgebra más pequeña asociada a las particiones de $\Omega$ o el espacio Borel y $\mathcal{P}$ son las probabilidades de las particiones de $\mathcal{F}$, con $\Omega \subseteq \mathcal{F}$ y $\mathbb{P} : \mathbb{R} \to [0, 1]$.

Además, la transformada integral ó la transformada inversa, es ampliamente usada en el campo de la estadística, pues, se usa principalmente para la generación de las variables aleatorias en los métodos de simulación, a partir de la inversa de la función de distribución de $X$ y una variable aleatoria $U \sim Unif(0,1)$. Con lo cual, se es capaz de producir conjuntos de datos aleatorios en el rango de $X$, haciendo uso de la Ecuación 2, en otras palabras, $F_X^{-1}(u)$.

La transformada integral o transformada inversa, también, es usada en el campo de las funciones cópula, pues, estas son definidas en el espacio $[0,1]^p$, donde p, es el orden del espacio. En $\mathbb{R}^2$ pertenecen al espacio $[0,1]^2$ por lo cual es usada en vectores aleatorios bidimensionales donde cada uno de sus ejes posee una función de distribución acumulada uniforme, en pocas palabras, es invertible por medio del teorema de la transformada integral de probabílidad.

Dentro de los métodos estadísticos, existe una alternativa no paramétrica para el teórema de la transformada inversa, para esta se asocia la variable uniforme al resultado de la función de distribución empírica acumulada o por medio de los rangos de las variables aleatorias. Sea $X$ una variable con una función de distribución empírica acumulada.

$$u_i = F_n(X_i) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}(x_j \leq X_i) \quad x_j \in \mathbb{R}.$$

O su alternativa por rangos, con $R_i$ representando el rango del valor i-ésimo de la muestra.

$$u_i = \frac{R_i}{n}.$$

En ambas casos, tanto haciendo uso de la función de distribución empírica acumulada como en el uso de rangos, se asume que $u_i \sim Unif(0,1)$. En algunos textos, por ejemplo en Ledwina (2014) se sugiere otra aproximación por rangos, en la cual, en vez de usar $n$ como denominador, se usa $n+1$

$$u_i = \frac{R_i}{n+1}.$$

### 2.1.3. Bootstrap

Los estimadores Bootstrap son muy populares dentro de la rama de la estadística computacional, además, de que son versátiles cuando la distribución de las variables aleatorias poseen estructuras complejas o no conocidas. En el libro (Casella & Berger, 2001, p. 478), un ejemplo muy conocido es la estimación del error cuádratico medio o la varianza de un estimador de un parámetro

$$Var_B^*(\hat{\theta}) = \frac{1}{B-1}\sum_{i=1}^{B}(\hat{\theta}_i^* - \hat{\theta}^*)^2.$$

Donde $\hat{\theta}^*$ es una estimación del parámetro real y $\hat{\theta}_i^*$, son estimaciones por remuestreo del parámetro, se espera que por relaciones $Var_B^*(\hat{\theta}) \overset{B\to\infty}{\to} Var^*(\hat{\theta}) \overset{n\to\infty}{\to} Var(\hat{\theta})$.

### 2.1.4. Prueba de hipótesis

Una prueba de hipótesis es un método estadísto para medir la no verasidad de algún supuesto sobre la muestra observada. Estas formalmente se definen en Casella & Berger (2001), con distintos métodos para calculara una prueba de hipótesis como la **LRT** o razón de verosimilitud, la **Test más potente**, entre otros. Estas pruebas mencionadas, son pruebas que corresponde a los Test de muestras grandes, donde se usa la misma muestra para realizar las pruebas de hipótesis.

Una prueba de hipótesis se define normalmente, con la intención de rechaza la hipótesis nula, pues estas se plantean sobre supuestos en la muestra y los datos permiten rechazar estos supuestos en caso tal de no ser correctos. Otra aclaración sobre las pruebas de hipótesis es que estas se definen bajo parámetros y bajo signos de igualdad, en otras palabras, $H_0 : \mu = 0$ y $H_a : \mu > 0$, donde $\mu$ puede ser el parámetro de localización de una función de distribución normal.

Las pruebas de hipótesis al igual que los estimadores, son un recurso estadístico el cual debe de ser evaluado en cuanto a su sensibilidad (confianza) y su espeficidad (potencia). La primera, corresponde a que dado un nivel de significancia preestablecido ¿cuál es la probabilidad de no rechazar la hipótesis nula si esta es cierta? La segunda, corresponde a que dado un nivel de significancia preestablecido nuevalemnte ¿cuál es la probabilidad de rechazar la hipótesis nula si esta es falsa?

***Pruebas de hipótesis empíricas***

Una prueba de hipótesis también puede ser realizada sin necesidad de conocer la distribución real del estimador, esto, a través de distintos métodos de remuestreo los cuales se basan en la técnica del uso recurrente del Booststrap (Sección 2.1.3). Un ejemplo a esto, es que en caso tal de querer evaluar que la media de una muestra aleatoria con función de distribución desconocida ($X \sim F$) es igual a 0, o algún otro valor, se pueden realizar una serie de remuestreos con repetición ó réplicas del valor de la media

$$\bar{\boldsymbol{X}}^{(\boldsymbol{k})} = \{\bar{X}^{(1)}, \ldots, \bar{X}^{(k)}\},$$

donde $k$ corresponde al número de veces que se realiza el remuestreo sobre la muestra. La media al ser un estimador que converge débilmente a una $N(\mu, \frac{\sigma}{\sqrt{N}})$, su prueba puede ser análizada asintóticamente por la misma distribución

$$\hat{\mu} = \frac{1}{k}\sum_{i=1}^{k} \bar{X}^{(k)},$$

$$\hat{\sigma} = \frac{1}{k-1}\sum_{i=1}^{k} \left(\bar{X}^{(k)} - \hat{\mu}\right)^2,$$

$$Z_B = \frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{n}}.$$

Donde $Z_B$ converge débilmente a una distribución $N(0,1)$, con valor-p $2 * P(Z \geq Z_B)$ (Casella & Berger, 2001).

***Confianza de pruebas empíricas***

La forma para calcular la confianza de las pruebas empíricas es tomar un coeficiente de significancia fijo para la prueba y luego, ver por críterios empíricos si el estadístico de prueba ($T_c$) se ajusta a la región de confianza ($R_A$). Dado esto se calcula

$$Confianza(\alpha) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{1}(T_c^{(i)} \in R_A) \ ,$$

donde $k$ es la cantidad de remuestreos del estadístico de prueba.

***Potencia de pruebas empíricas***

Para la potencia se tiene el mismo críterio (Huang & Prokhorov, 2014), donde, a partir de un nivel $\alpha$ de significancia preestablecida se calcula cuantas veces el estadístico de prueba ($T_c$) pertenece a la región de rechazo ($R_C$). Dado lo anterior, la potencia empírica se calcula

$$Potencia(\alpha) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{1}(T_c^{(i)} \in R_C).$$

En muchos documentos se toma la confianza como $1 - \alpha$ y la potencia como $1 - \beta$ (Casella & Berger, 2001), sin embargo, al realizarse por medio de estudios empíricos la notación se vuelve rebundante, pues, se calculan estimadores de $\beta$ y $\alpha$ por el desconocimiento de las distribuciones reales de los estadísticos de prueba.

### 2.1.5. Prueba de Anderson-Darling

La prueba de Anderson-Darling es una prueba estadística utilizada para evaluar la bondad de ajuste de un conjunto de datos a una distribución teórica. Esta prueba es una extensión de la prueba de Kolmogorov-Smirnov y se enfoca en las colas de la distribución, lo que la hace particularmente eficaz para detectar discrepancias en las colas, a diferencia de otras pruebas que pueden ser más sensibles a la parte central de la distribución.

El estadístico de Anderson-Darling, denotado como $A^2$, se calcula de la siguiente manera:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} \left( (2i-1) \left( \ln(F(X_{(i)})) + \ln(1 - F(X_{(n+1-i)})) \right) \right)$$

donde:

- $n$ es el número de observaciones,
- $F$ es la función de distribución acumulada teórica,
- $X_{(i)}$ son los estadísticos de orden de las observaciones.

El valor del estadístico $A^2$ se compara con un umbral crítico que depende del nivel de significancia elegido. Si el valor de $A^2$ excede este umbral, se rechaza la hipótesis nula, que establece que los datos provienen de la distribución especificada.

La prueba de Anderson-Darling se ha convertido en una herramienta ampliamente utilizada en estadística para la validación de modelos y ajustes de distribución (Sheskin, 2003).

## 2.2. Funciones cópula

La función cópula fue propuesta por Sklar en 1959, como se menciona en el artículo de Van der Vaart y otros (Van Vliet, 2023, p.4), donde se revisa el teorema de Sklar. Según Sklar, la función cópula se define de la siguiente manera: $C(0, 0, 0, \ldots, 0) = 0$ y $C(1, 1, \ldots, u, 1) = u$. Estas definiciones se relacionan con las distribuciones marginales y la función de distribución conjunta de un vector aleatorio.

En estadística, la función de distribución conjunta se suele denotar como $F_{\boldsymbol{X}}(\cdot)$. Sin embargo, Sklar la llama $H_{\boldsymbol{X}}(\cdot)$ en (Van Vliet, 2023, p. 4). Esta notación también se encuentra en otros trabajos sobre cópulas y sus medidas de concordancia, como en Nelsen (2006), Genest et al. (2009), y Ledwina (2014).

Una cópula se asocia más bien con la estructura de relación de la función de distribución conjunta, pues, **parcialmente miden probabilidades condicionales entre variables**. Existen distintas distribuciones conjuntas que poseen las mismas funciones de distribuciones marginales, además, desde una función de distribución conjunta se pueden hallar sus funciones de distribución marginales; sin embargo, las funciones de distribución marginales por sí solas no pueden reconstruir directamente la función de distribución conjunta, a menos que las variables sean independientes entre sí. A continuación, se presentarán definiciones centradas únicamente en funciones bivariadas.

En términos simples, una función cópula describe la estructura de dependencia que conecta las funciones de distribución marginales de un vector aleatorio. La función cópula transmite dos tipos de información (Llorens, 2017):

- El comportamiento de cada variable de forma aislada: $F_{X_i}(x_i)$.
- El comportamiento conjunto de las variables, que refleja la estructura de dependencia entre ellas, la cual es modelada por la función cópula.

### 2.2.1. Teorema de Sklar

La definición de una cópula bivariada se encuentra en (Nelsen, 2006, p. 21)

Dado una función de distribución conjunta $F_{X,Y}(x, y)$ con marginales $F_X(\cdot)$ y $F_Y(\cdot)$.

Entonces existe una función cópula $C(\cdot,\cdot)$ tal que para toda pareja $x, y$ en $\mathbb{R}^2$ se cumple

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

Además, si $F_X(\cdot)$ y $F_Y(\cdot)$ son continúos, entonces, existe una única función cópula asociada.

De la definición anterior existe un recíproco, este se encuentra en (Nelsen, 2006, p. 22), expresa que toda cópula, puede, puede ser escritos en terminos de la transformada integral Sección 2.1.2 y su función de distribución conjunta, esto, en muchos textos se encuentra como la transformada integral n variada, o una generalización de la misma. La cópula, entonces, se puede expresar de la forma

$$C_\theta(u,v) = F_{X,Y}(x,y) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v)); \qquad F_X(x) = u \qquad F_Y(y) = v,$$

donde se fácilita el proceso para hallar la estimación de una función de distribución conjunta dada las marginales uniformes o sea, $U, V \sim Unif([0,1]^2)$. Se vuelve a recalcar, que en caso tal de ser $X$ y $Y$ variables aleatorias con una función de distribución continúa esta posee una única cópula asociada, sin embargo, en caso tal de ser discreta o mixta, existen infinidad de cópulas las cuales se pueden asociar a la función de distribución conjunta de los datos. Por último, se recuerda que en caso tal de ser $X$ inpendiente a $Y$, entonces, la cópula asociada a la independencia se define en la Ecuación

$$\Pi(u,v) = u * v = F_X(x) * F_Y(y).$$

### 2.2.2. Método de simulación haciendo uso de la transformada inversa. (Mai & Scherer, 2013, p. 23)

En el libro de Mai & Scherer (2013) se encuentran varios métodos de simulación para funciones cópula, entre estos está el de la transformada inversa. Este consiste en que a partir de la función cópula ($C_\theta(u,v)$) y el hecho de que las derivadas parciales corresponden a las funciones de distribución condicionales de $F_{U|V}(u \mid v = v_0, \theta)$ y de $F_{V|U}(v \mid u = u = 0, \theta)$, se simule una de las dos a partir de la otra. Este proceso se realiza a partir de la creación de una nueva variable uniforme $W \sim Unif(0,1)$ la cual hace referencia a los posibles valores que toma la función en su derivada parcial. Por ende, se pueden obtener muestras simuladas de una función cópula con el conocimiento de una de las dos marginales y el resultado de la primera derivada parcial. El procedimiento se encuentra en (Mai & Scherer, 2013, p. 23)

1. Simule un valor de $W \sim Unif(0,1)$ que corresponde a la primera derivada de la función cópula y simule un valor de $Z_1 \sim Unif(0,1)$ que corresponde al valor que toma una de las marginales en una primera muestra.
2. Calcule $F_{V|U}(v \mid u = z_1, \theta) = \frac{\partial}{\partial u}\big|_{u=z_1} C_\theta(u,v), \quad v \in [0,1]$.
3. Calcule la inversa de $F_{V|U}(v \mid u = z_1, \theta)$, donde $F^{-1}_{V|U} = \text{ínf}\{u > 0 : F_{V|U}(v \mid u = z_1, \theta) \leq c\}, \quad c \in [0,1]$.

4. Compute a $Z_2 = F^{-1}_{V|U}(w \mid u = z_1, \theta)$
5. Regrese las variables aleatorias calculadas, las cuales se distibuyen como la función cópula deseada, $(Z_1, Z_2) \sim C_\theta$.

### 2.2.3. Método de simulación haciendo uso de la función de distribución de la función característica Arquimediana. (Mai & Scherer, 2013, p. 61)

Dada una función $\varphi(t)$ continua y estrictamente decreciente que va de $[0,1] \to [0,\infty]$, tal que $\varphi(1) = 0$, esta posee como transformada inversa de Laplace a $M$ que es una variable aleatoria positiva, esto debido a que la función $\varphi(t)$ es parte de los momentos de la función de distribución de $M$. Para generar variables aleatorias de la familia arquimediana, es necesaria una variable aleatoria exponencial, tal que $E \sim Exp(1)$ y $U_j = \varphi(E_j/M)$. Haciendo uso de este método no es necesario atribuirle una función de distribución inversa, no obstante, se genera la dificultad para conocer la transformada inversa de Laplace asociada a la función arquimediana de la cópula. En los casos de estudio (Tovar Cuevas et al., 2019) la función cópula Gumbel-Hougaard fue simulada bajo este método debido a la complejidad para calcular la inversa de su primera derivada parcial. El algoritmo redactado en (Mai & Scherer, 2013, p. 6) está dado por

1. Genere un valor aleatorio $E_1 \sim Exp(1)$ y $E_2 \sim Exp(1)$.
2. Genere un valor aleatorio positivo de $M \sim F_M$ el cual tenga como transformada de Laplace a $\varphi$ (la función arquimediana de la cópula).
3. Regrese $U_1 = \varphi(E_1/M)$ y $U_2 = \varphi(E_2/M)$.

Además de este método, existen otros respecto al uso de la función arquimediana de la cópula, en caso de desear conocerlos lea Mai & Scherer (2013) pagina 62.

### 2.2.4. Función de distribución empírica (cópula empírica)

La cópula empírica, es la estimación de la función cópula de forma no paramétrica, pues, esta no necesita un parámetro conocido o una distribución conocida. Si se habla de las funciones de distribución comunes, la cópula empírica es la representación de la función de distribución conjunta acumulada $F_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(x_i \leq x, y_i \leq y)$ que se usa en la estadística clásica, sin embargo, en este caso se usa para funciones cópulas [2].

En Nelsen (2006) se encuentra documentada la función de la cópula empírica, esto mediante algortimos y expresiones computacionales. Si se desea expresar de forma más concreta o teórica, se encuentran los artículos de Genest et al. (2009); Ledwina (2014); Tovar Cuevas et al. (2019) como un sustento a garantizar una expresión mátematica de la misma para un vector aleatorio bivariado.

[2] Al igual que $F(F_X^{-1}(u), F_Y^{-1}(v)) = C(u,v)$, la relación entre la cópula empírica y la función de distribución conjunta es $F_n(F_X^{-1}(u), F_Y^{-1}(v)) = C_n(u,v)$

Dado $(X_1, Y_1), \ldots, (X_n, Y_n)$ una muestra de vectores aleatorios de una función de distribución $F(\cdot, \cdot)$. Además, sean $R_i$ y $S_i$ los rangos de $X_i$ y $Y_i$ respectivamente para $i = \{1, \ldots, n\}$, la cópula empírica se define en Ledwina (2014) como

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

En Genest et al. (2009) se menciona, que a pesar de no ser precisamente una función cópula, esta es llamada como cópula empírica. Esta definición ó forma de mencionarla es importante en el estudio debido a que en Tovar Cuevas et al. (2019) es usada para la creación del indiciador de bondad de ajuste para funciones cópulas bivariadas.

### 2.2.5. Coeficiente de concordancia: $\tau$ de Kendall

El $\tau$ de Kendall es una medida de concordancia, la cual estima la proporción de los pares concordantes y discordantes de la distribución de los datos, una concordancia se puede definir como el empate entre el orden de la muestra y la relación de crecimiento o decrecimiento de la misma. En Nelsen (2006) se define de la forma

Dados $(X_1, Y_1)$ y $(X_2, Y_2)$ vectores aleatorios independientes e identicamente distribuidos, para cada distribución conjunta $F_{X,Y}(\cdot, \cdot)$. Entonces, el $\tau$ de Kendall se puede ver en

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

$$\tau_\theta = 4 \int \int_{I^2} C_\theta(u, v) dC_\theta(u, v) - 1 \tag{3}$$

también, puede ser expresado como $4E(C_\theta(u, v)) - 1$. El $\tau$ de Kendall es una medida de concordancia muy asociada a los mómentos muestrales de la cópula, pues, está relacionado en gran parte con el primer momento de la función cópula de X y Y.

Su medición no paramétrica (Nelsen, 2006) está dada por

$$\tau_n = \frac{2}{n(n-1)} \sum_{i<j}^{n} sgn(x_i - x_j) sgn(y_i - y_j)$$

Donde *sgn* corresponde a la función señal, la cual para el caso es -1 si es discordante, 0 si es igual, 1 si es parcordante.

### 2.2.6. Estimación de parámetros para funciones cópula

La estimación de los parámetros de una función cópula se encuentra ligada al $\tau$ de Kendall (funciones de los momentos) y a la función de máxima verosimilitud de la

función cópula. La función de máxima verosimilitud, se construye a partir de la función de densidad de la cópula, haciendo uso de la doble derivada parcial para el caso bivariado y la n-ésima derivada parcial para el caso n-variado, debido a que en este documento solo se trabajará con funciones cópulas bivariadas, la densidad bivariada se define [3]

$$c_\theta(u,v) = \frac{\partial^2}{\partial u \partial v} C_\theta(u,v) \quad u,v \in [0,1]^2 \tag{4}$$

La Ecuación 4 es conocida en Nelsen (2006) como la densidad de la cópula bivariada, al igual que una función de densidad común posee la carácterística de que su máxima verosimilitud no sea necesariamente concava, por lo cual, es mejor recurrir a la transformación usando la log-verosimilitud. La cual, es una transformación uno a uno, monotona, acotada y continua, permitiendo garantizar la máximización de forma análitica y númerica de la función, buscando el parámetro que genere mayor similitud entre la muestra observada y su estimación. La expresión de la máxima verosimilitud de la función cópula se ve definida por

$$\hat{\theta} = \underset{\boldsymbol{\theta}\in\Theta}{\max} \mathcal{L}_n(\theta \mid u,v) = \underset{\boldsymbol{\theta}\in\Theta}{\max} \left\{ \sum_{i=1}^{n} log\left[c_\theta(u_i,v_i)\right] \right\}. \tag{5}$$

En la Ecuación 5 se puede ver una idea más común respecto a funciones de distribución tradicionales, en esta se busca el parámetro que máximice la suma de los logaritmos de las densidades de la función cópula bivariada.

Otro método muy famoso es el del $\tau$ inverso de Kendall (o momentos), pues, este guarda una relación directa con la esperanza de la función cópula

$$g(\theta) = \tau_\theta = 4E[C_\theta(u,v)] - 1,$$

Cuando $(U,V)$ provienen de una muestra de un vector aleatorio, la estimación se hace mediante el uso del $\tau$ de Kendall no paramétrica

$$\hat{\theta} = g^{-1}(\tau_n).$$

### 2.2.7. Funciones cópula de estudio.

En el estudio realizado por Tovar Cuevas et al. (2019) se análizaron un total de 5 funciones cópula, con distintos niveles de dependencia, distintos tamaños de muestra y la estimación de sus parámetros de distintas formas.

***Función cópula Clayton:***

La función cópula Clayton ó la función de distribución Pareto bivariada, es una función que gráficamente posee una forma distintiva (de embudo) Figura 1, esta función cópula,

[3] Note que la densidad bajo independencia es 1, dado que $C(u,v) = uv$ y $\frac{\partial^2}{\partial u \partial v} C(u,v) = 1$

esta diseñada para modelar dependencias fuertes en colas inferiores, pues, su coeficiente de concordancia de Kendall $\tau \in [-1, 1]$, además, añadir que cuando $\theta \to 0$, se encuentra indenpendencia

$$C_\theta(u, v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}} \quad \mathbb{1}_{[0,1]^2}(u, v) \quad \theta \in [-1, \infty) - \{0\}.$$

En la Figura 1 se encuentran los distintos niveles de dependencias fijados previamente en Tovar Cuevas et al. (2019), donde $\theta = \{0.5, 2.0, 8.0\}$ son los valores que toma el parámetro para la dependencia débil, moderado y fuerte respectivamente.

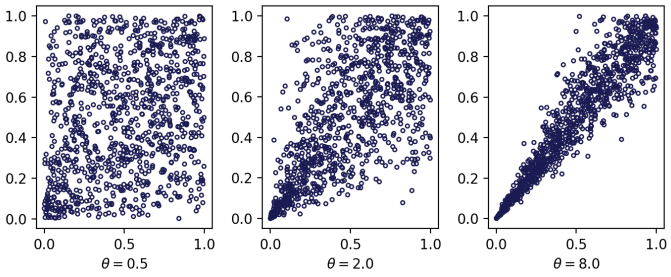


**Figura 1**: Diagrama de dispersión de la función copula Clayton. Fuente: Elaboración propia

La función de densidad de la función cópula Clayton esta definida en Nelsen (2006)

$$c_\theta(u, v) = (\theta + 1) \cdot (uv)^{-\theta-1} \cdot \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}-2}, \quad \mathbb{1}_{[0,1]^2}(u, v), \quad \theta \in [-1, \infty) - \{0\}.$$

El $\tau$ de Kendall de la función cópula Clayton también se encuentra definido en (Nelsen, 2006), este tiene una expresión cerrada

$$\tau_\theta = \frac{\theta}{\theta + 2} \quad \theta \in [-1, \infty) - \{0\},$$

los valores de esta medida se encuentran en el espacio [-1,1], donde 1 o -1 corresponden a valores de dependencia total.

***Función cópula Farley-Gumbel-Morgenstern (FGM):***

La función cópula FGM, se usa para modelar dependencias débiles, pues, la medida de concordancia de Kendall $\tau \in [-0.222, 0.222]$, además, a pesar de que el comportamiento del vector aleatorio parezca ser independiente, en algunas ocasiones no necesariamente lo es. Un ejemplo a esto, es cuando se toma una muestra de dos sustancias a un mismo individuo.

La estructura de la función cópula FGM bivariada se visualiza en la Figura 2 en esta,

se pueden ver los niveles de dependencia previamente fijados en Tovar Cuevas et al. (2019), donde $\theta = \{0.2, 0.5, 0.8\}$ corresponden a los valores que toma el parámetro para la dependencia débil, moderada y alta en el dominio positivo de la función cópula.

La función cópula FGM definida como.

$$C_\theta(u, v) = uv[1 + \theta(1 - u)(1 - v)] \;\; \mathbb{1}_{[0,1]^2}(u, v), \quad \theta \in [-1, 1].$$

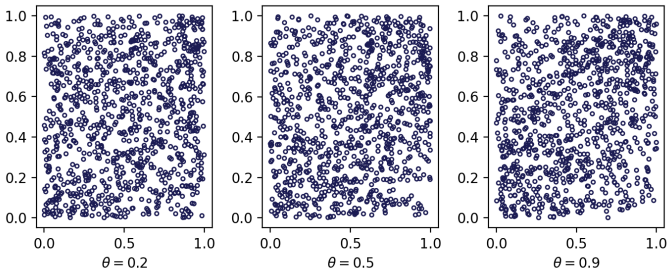


**Figura 2**: Diagrama de dispersión de la función copula FGM. Fuente: Elaboración propia

Su función de densidad se expresa

$$c_\theta(u, v) = 1 + \theta(1 - 2u)(1 - 2v), \quad \mathbb{1}_{[0,1]^2}(u, v) \;\; \theta \in [-1, 1],$$

y la forma de calcular su $\tau$ de Kendall es cerrada, por lo cual se puede expresar como

$$\tau = \frac{2\theta}{9}, \quad \theta \in [-1, 1].$$

los valores de esta medida de concordancia se encuentran en el espacio $[-0.222..., 0.222...]$ lo cual es un rango reducido del espacio $[-1, 1]$

***Función cópula Frank:***

La función cópula Frank es apropiada para modelar tendencias lineales débiles, su función de distribución de probabilidad acumulada se define en la Ecuación **??** y su estructura es visible en la Figura 3, cuando $\theta \to 0$ se asume independencia.

En la Figura 3 se puede visualizar la relación lineal entre las variables aleatorias $u$ y $v$, se puede ver que a medida que el parámetro de dependencia aumenta, está se vuelve más cercana a una linea recta. Los valores para $\theta = \{1.86, 5.73, 18.19\}$ son valores previamente definidos en Tovar Cuevas et al. (2019), los cuales, corresponden a dependencia débil, moderada y fuerte. La medida de concordancia de Kendall $\tau \in [-1, 1]$

por lo cual, las concordancias de está cópulas tienen el máximo rango posible.

$$C_\theta(u,v) = -\frac{1}{\theta} log\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right) \quad \mathbb{1}_{[0,1]^2}(u,v), \quad \theta \in (-\infty,\infty) - \{0\}$$

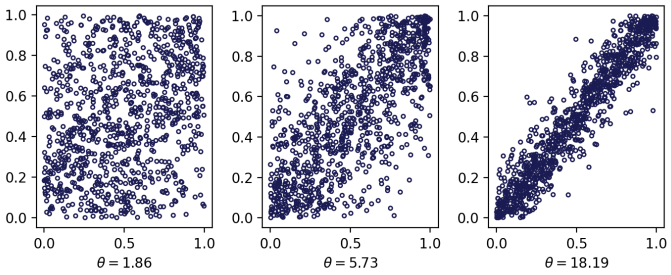


**Figura 3**: Diagrama de dispersión de la función copula Frank. Fuente: Elaboración propia

Su función de densidad está dada por

$$c(u,v;\theta) = \frac{\theta e^{-\theta(u+v)}(e^{-\theta}-1)}{((e^{-\theta u}-1)(e^{-\theta v}-1)+e^{-\theta}-1)^2} \quad \mathbb{1}_{[0,1]^2}(u,v), \quad \theta \in (-\infty,\infty) - \{0\},$$

y su $\tau$ de Kendall se calcula por medio de la expresión

$$\tau = 1 - \frac{4}{\theta}\left(1 - \frac{1}{\theta}\int_0^\theta \frac{t}{e^t - 1}\, dt\right), \quad \theta \in (-\infty,\infty) - \{0\}.$$

Al igual que la función cópula Clayton, esta medida de concordancia se encuentra definida en el espacio $[-1,1]$.

***Función cópula Gumbel-Barnett:***

La función cópula Gumbel-Barnett, nace principalmente de una distribución con marginales exponenciales, realizada en Gumbel (1960), en esté artículo revisan a detalle algunos conceptos como los momentos o la correlación de esta distribución, además, en Diaz & Cuadras (2022) extienden el concepto de la familia Gumbel-Barnett, en donde, la relacionan familia FGM y otras familias. El coeficiente de concordancia del $\tau$ de Kendall de está cópula no tiene forma analitica cerrada, por ende, por medio del uso de la regla del trapecio y el teorema del valor medio en la derivada se encontraron las siguientes cotas $\tau \in (-0.361, 0]$ [4], donde, la inferior corresponde cuando $\theta \to 1$, lo cual señala que la Gumbel-Barnett modela dependencias inversas entre variables

[4]Este resultado fue corroborado tanto en **R** como en **Python**

aleatorias bivariadas.

En la Figura 4 se encuentra la estructura visual de la función cópula, esta se define bajo la Ecuación **??**.Los del parámetro $\theta = \{0.2, 0.5, 0.8\}$ corresponden a los niveles bajo, moderado y alto de dependencia de la función cópula.

$$C_\theta(u,v) = u + v - 1 + (1-u)(1-v)e^{-\theta ln(1-u)ln(1-v)} \quad \mathbb{1}_{[0,1)^2}(u,v), \quad \theta \in [0,1]$$

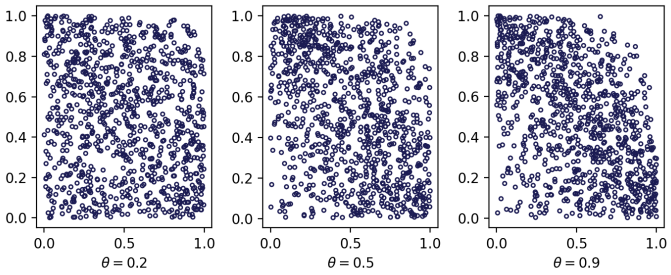


**Figura 4**: Diagrama de dispersión de la función copula Gumbel-Barnett. Fuente: Elaboración propia

Con función de densidad

$$c_\theta(u,v) = e^{-\theta \ln(1-u)\ln(1-v)}(1 - \theta\ln(1-u))(1 - \theta\ln(1-v)) \quad \mathbb{1}_{[0,1)^2}(u,v), \quad \theta \in [0,1].$$

Esta función cópula no posee un valor exacto para su función del $\tau$ de Kendall, sin embargo, este toma valores apróximados en el rango $(-0.361, 0]$ por lo cual, se evidencia que modela dependencias débiles negativas.

***Función cópula Gumbel-Hougaard:***

La función cópula Gumbel-Hougaard se usa para modelar las dependencias fuertes en las colas de una función de ditribución. Su parámetro de pendencia $\theta$ toma valores mayores o iguales a uno, cuando este $\theta \to \infty$ se logra la dependencia perfecta. Su medida de concordancia del $\tau$ de Kendall se encuentra en el espacio [0,1]. Su función se observa en la Ecuación **??** y su diagrama de dispersión se visualiza en la Figura 5.

En la Figura 5 se visualiza que a medida que aumenta la dependencia, la cópula adopta una forma de rombo con mayor concentración en la punta superior derecha, los valores para el parámetro $\theta = \{1.25, 2, 5\}$ reflejan los distintos niveles de dependencia extraídos de Tovar Cuevas et al. (2019).

$$C_\theta(u,v) = e^{-\left[(-log(u))^\theta + (-log(v))^\theta\right]^{\frac{1}{\theta}}} \quad \mathbb{1}_{(0,1]}(u,v), \quad \theta \in [1,\infty)$$

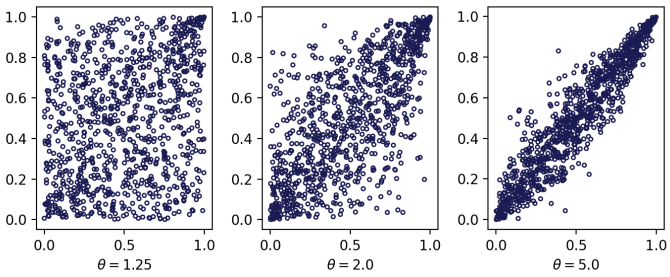


**Figura 5**: Diagrama de dispersión de la función copula Gumbel-Hougaard. Fuente: Elaboración propia

Su función de densidad se expresa como

$$c(u,v;\theta) = e^{-w}\frac{(-\log u)^{\theta-1}}{u}\left(\frac{1}{\theta}-1\right)w^{\frac{2}{\theta}-3}\frac{(-\log v)^{\theta-1}}{v} \quad \mathbb{1}_{(0,1]}(u,v), \quad \theta \in [1,\infty),$$

donde $w = -\left[(-log(u))^{\theta} + (-log(v))^{\theta}\right]^{\frac{1}{\theta}}$ y su función del $\tau$ de Kendall se calcula por la expresión

$$\tau = 1 - \frac{1}{\theta}, \quad \theta \in [1,\infty).$$

Esta medida del $\tau$ de Kendall muestra que esta función cópula solo modela dependencias positivas entre pares de variables aleatorias.

### 2.2.8. Coeficiente de concordancia: *q* de Ledwina

La medida de concordancia desarrollada por Ledwina (2014) es en sí misma una medida de dependencia en cuádrantes (Nelsen, 2006), estás a diferencia de las medidas de concordancia comunes, poseen la partícularidad de medir la dependencia en pares de puntos determinados en los datos. Está medida cumple al igual que cualquier medida de correlación entre variables aleatorias, la propiedad de que se encuentra en el espacio [-1,1]. La medida de concordancia se define

$$q_\theta(u,v) = \frac{C_\theta(u,v) - \Pi(u,v)}{\sqrt{\Pi(u,v)(1-v)(1-u)}},$$

donde $C_\theta$ corresponde a la cópula de donde provienen los datos, $\Pi$ corresponde a la función cópula independiencia y el denominador corresponde a una constante de normalización.

La medida de concordancia asociada a la función cópula es $q_\theta$, además, esta medida cumple la propiedad de estar en el rango $-1 \leq q(\cdot,\cdot) \leq 1$, esta en el texto de Ledwina, se considera una correlación de Pearson convencional, esto entre la función cópula empírica, la cópula de independencia y la constante de normalización

$\sqrt{uv(1-u)(1-v)}$, esto solo ocurre si $C_n(u,v) \overset{n\to\infty}{\to} uv$.

La medida de concordancia $q_\theta$ nace de la correción entre dos funciones deterministas las cuales miden dependencia punto a punto, la función se puede ver a continuación

$$\phi_z(r) = -\sqrt{\frac{1-z}{z}}\mathbb{1}_{[0,z]}(r) + \sqrt{\frac{z}{1-z}}\mathbb{1}_{(z,1]}(r). \tag{6}$$

Donde en la Ecuación 6 se puede ver la función con la cual se mide la dependencia de dos valores, un $z$ fijo y un $r$ el cual es dado, en este caso, $r$ se convierte en una sucesión de variables aleatorias las cuales se rigen por una función de distribución **uniforme** en el espacio $[0,1]$. En consecuencia, $E(\phi_u(U))$ se calcula usando la función de densidad de la variable uniforme $U$. La correlación entre dos variables se define como

$$Corr(\phi_u(U), \phi_v(V)) = \frac{E(\phi_u(U)\phi_v(V)) - E(\phi_u(U))E(\phi_v(V))}{Var(\phi_u(U))Var(\phi_v(V))} = q_\theta(u,v),$$

con lo cual,

$$E(\phi_z(Z)) = -\sqrt{\frac{1-z}{z}}P(0 \le Z \le z) + \sqrt{\frac{z}{1-z}}P(z < Z \le 1)$$

$$= -\sqrt{\frac{1-z}{z}}z + \sqrt{\frac{z}{1-z}}(1-z)$$

$$= -\sqrt{z(1-z)} + \sqrt{(1-z)z} = 0.$$

Luego,

$$E(\phi_z(Z)^2) = E\left[\left(-\sqrt{\frac{1-z}{z}}\boldsymbol{I}_{[0,z]}(Z) + \sqrt{\frac{z}{1-z}}\boldsymbol{I}_{(z,1]}(Z)\right)^2\right]$$

$$= \frac{(1-z)}{z}P(0 \le Z \le z) + 0 + \frac{z}{1-z}P(z < Z \le 1)$$

$$= (1-z) + z = 1 = Var(\phi_z(Z)).$$

Y por último,

$$E[\phi_u(U)\phi_v(V)] = \sqrt{\frac{(1-u)(1-v)}{uv}}C_\theta(u,v) + \sqrt{\frac{(1-u)v}{u(1-v)}}[u - C_\theta(u,v)]$$

$$-\sqrt{\frac{u(1-v)}{(1-u)v}}[v - C_\theta(u,v)] + \sqrt{\frac{uv}{(1-u)(1-v)}}[1 + C_\theta(u,v) - v - u]$$

$$= C_\theta(u,v)\frac{1}{\sqrt{u(1-u)v(1-v)}} - \frac{uv}{\sqrt{u(1-u)v(1-v)}} - \frac{uv}{\sqrt{u(1-u)v(1-v)}} + \frac{uv}{\sqrt{u(1-u)v(1-v)}}$$

$$= \frac{C_\theta(u,v) - uv}{\sqrt{u(1-u)v(1-v)}}$$

$$E(\phi_u(U)\phi_v(V)) = \frac{C_\theta(u,v) - \Pi(u,v)}{\sqrt{u(1-u)v(1-v)}}.$$

La autora menciona que ante la alternativa empírica $q_n(u,v) = \frac{C_n(u,v)-uv}{\sqrt{u(1-u)v(1-v)}}$ haciendo uso de los estudios realizados por Swanepoel & Allison (2013), si se asume independiencia entre $u$ y $v$ la distribución del estimador $\sqrt{n}q_n(u,v) \sim N(0,1)$ debido a que $\sqrt{\frac{u(1-u)v(1-v)}{n}}$ corresponde a la varianza muestral de la cópula empírica (para más detalle revisar Anexo A.2.1).

## 2.3. Métodos de selección de funciones cópula

Existen gran cantidad de métodos para seleccionar funciones cópula, algunos de estos adaptan métodos tradicionales al entorno de las variables aleatorias con estructuras de dependencia, en Genest et al. (2009) se realizó un estudio de potencia entre varios métodos. En algunos de estos se usaron coeficientes de concordancia como el $\tau$ de Kendall y en otros se usaron estadísticos basados en la estructura de Cramer-Von Mises, Kolmogorov-Smirnov y Anderson-Darling. Esto es debido a que dichas pruebas son usadas con valores en el espacio unitario (las funciones de distribución de las variables). En estudios más recientes, se encuentran las investigaciones realizadas por Ledwina & Wyłupek (2014), estudio el cual fue usado por Tovar Cuevas et al. (2019) para implementar un indicador de bondad de ajuste entre conjuntos de datos bivariados y funciones cópula.

### 2.3.1. Propuesta de Tovar, Portilla y Achcar

La prueba de estudio en este trabajo se encuentra en Tovar Cuevas et al. (2019), donde se adaptó el coeficiente de concordancia de Ledwina expresado en la Sección 2.2.8, esto por medio de los coeficientes empíricos (Ecuación 7) y teóricos (Ecuación 8)

$$q_n(u,v) = \frac{C_n(u,v) - uv}{\sqrt{uv(1-u)(1-v)}}, \tag{7}$$

$$q_\theta(u,v) = \frac{C_\theta(u,v) - uv}{\sqrt{uv(1-u)(1-v)}}. \tag{8}$$

Con lo cual se plantea un estadístico de Cramer-Von Mises el cual busca calcular la distancia euclidiana entre las dos medidas. De recuerda que esta al ser una distancia al cuadrado, no es necesario ningún orden, pues, $(a-b)^2 = (b-a)^2$ por propiedades de los exponentes pares. Este estadístico se planteó en el artículo como $I(model[j])$, este se expresa de mejor forma en la Ecuación 9

$$I(model[j]) = \sqrt{\sum_{i=1}^{n} \left(q_\theta(u_i, v_i) - q_n(u_i, v_i)\right)^2} \quad j = 1, 2, \ldots, k \quad i = 1, \ldots, n. \tag{9}$$

Sin embargo, la adaptación de $j$ corresponde a la evaluación del mismo para cada uno de las funciones cópulas estudiadas, en el caso de la investigación fueron 5, no obstante, queriendo adaptarlo a una forma más clásica del estadístico de Cramer-Von Mises. Se toma $I = \sqrt{\sum_{i=1}^{n} \left(q_\theta(u, v) - q_n(u, v)\right)^2}$. Buscando generalizar el concepto, pues en muchos textos de la literatura se usa la forma cuadrada de las diferencias como un indicador de bondad de ajuste (Genest et al., 2009) con $S^2$ ó variaciones de la notación, por lo mismo, en la Ecuación 10 se expresa el indicador del estudio en terminos más comunes y comparables.

$$I^2 = \sum_{i=1}^{n} \left(q_\theta(u_i, v_i) - q_n(u_i, v_i)\right)^2. \tag{10}$$

En la Ecuación 10 se pueden encontrar las medidas de dependencias establecidad en Ledwina (2014), está expresión puede ser simplificada aún más si se toma a concideración que ambas tienen como componente interno a la cópula de independencia,

$$I^2 = \sum_{i=1}^{n} \left( \frac{C_\theta(u_i, v_i) - u_i v_i}{\sqrt{u_i v_i (1-u_i)(1-v_i)}} - \frac{C_n(u_i, v_i) - u_i v_i}{\sqrt{u_i v_i (1-u_i)(1-v_i)}} \right)^2.$$

Donde esta puede ser reescrita totalmente de la forma expresada en la Ecuación 11

$$I^2 = \sum_{i=1}^{n} \left( \frac{C_\theta(u_i, v_i) - C_n(u_i, v_i)}{\sqrt{u_i v_i (1-u_i)(1-v_i)}} \right)^2. \tag{11}$$

Dando así resultado a $I^2$ como una función en base a las distancias de la cópula empírica y la cópula teórica (Ecuación 11), además, posee una constante la cual está en función de las marginales uniformes de cada una de las variables aleatorias, este puede ser un factor el cual genere disminución o inflación de varianza, por lo cual es necesario estudiar también si influye positiva o negativamente en el análisis futuro de este trabajo (Anexo A.2.2).

# 3. Metodología

En esta sección se describen las metodologías implementadas para alcanzar los objetivos planteados en este trabajo de grado. El software de simulación utilizado fue Python, debido a su agilidad en algunos subprocesos (como el uso de múltiples hilos) en comparación con R. Este último se empleó para realizar comparaciones de cálculos, junto con Wolfram Mathematica, buscando mitigar errores en las simulaciones o en la resolución de ecuaciones no lineales. Casi todo el trabajo se llevó a cabo con las librerías de Scipy, Pandas, Numpy, Findiff y Numdifftools, que fueron esenciales para manejar funciones matemáticas, gestión de datos y operaciones entre vectores. Estas herramientas fueron útiles en casos que requerían integración, derivación con respecto a la cópula o alguna función relacionada.

Toda la información de este trabajo está disponible en GitHub. En este repositorio también se encuentran módulos creados a medida para cumplir con los objetivos, incluyendo funciones como la doble derivada parcial de la función cópula y el cálculo aproximado de la integral $E[C_\theta(u,v)] = \int_{\mathbb{I}^2} C_\theta(u,v)\frac{\partial^2}{\partial u \partial v}C_\theta(u,v)\,dudv$ mediante la regla del trapecio y sumas de Riemann (Anexo A.1.1), con el fin de calcular el $\tau$ de Kendall.

## 3.1. Esquema de estudio

En la Figura 6 se puede visualizar el flujo de trabajo general del estudio. Este se presenta de manera global, con la intención de profundizar en secciones posteriores. En cuanto a los aspectos visuales, el rectángulo rojo corresponde al inicio de la metodología, donde se definen los casos de estudio, los cuales contienen el número de réplicas del indicador, los tamaños de muestra a trabajar, los niveles de dependencia de las funciones cópula y las familias de funciones cópula a trabajar.

En la Figura 6 los rectángulos de color amarillo corresponden a los procesos a que se realizaron para el cumplimiento del primer objetivo, el rectángulo que se encuentra en la parte superior corresponde a la generación de réplicas del indicador $I^2$ las cuales se realizaron por medio de la simulación de varios vectores aleatorios de $(U,V)$. Por otro lado, el que se encuentra en una zona más abajo, corresponde a la evaluación de la distribución muestral de las réplicas del indicador $I^2$ frente a funciones de distribución unidimensionales conocidas, este proceso se realizó por medio de la prueba de Anderson-Darling.

El rectángulo de color azul (Figura 6) corresponde al proceso por el cual se desarrolló el segundo objetivo, este tenía como finalidad cuantificar las propiedades de insesgamiento, variabilidad, estabilidad y robustez del indicador $I^2$ propuesto por Tovar Cuevas et al. (2019) siendo visto como un estimador de bondad de ajuste. La finalidad de este objetivo fue ver cuantas réplicas del indicador son necesarias para obtener una buena aproximación de la prueba de hipótesis basada en este.

En la Figura 6 también se pueden ver tres rectángulos de color verde, estos corresponden al desarrollo del tercer objetivo. En esta etapa se trabajó lo correspondiende a la prueba de hipótesis empírica, la región de rechazo, la confianza de la prueba y la potencia de la misma. En esta parte se estableció la hipótesis en torno a la muestra de un vector aleatorio $(U, V)$, donde está busca ver si los datos se ajustan a la función cópula propuesta en la hipótesis nula, mientras que, en la hipótesis alternativa se expresa lo contrario. Respecto al indicador el ajuste perfecto se da cuando $I^2 \to 0$ y cuando $I^2 \to \infty$ se define como que no existe ajuste entre la función cópula propuesta y la muestra del vector aleatorio.

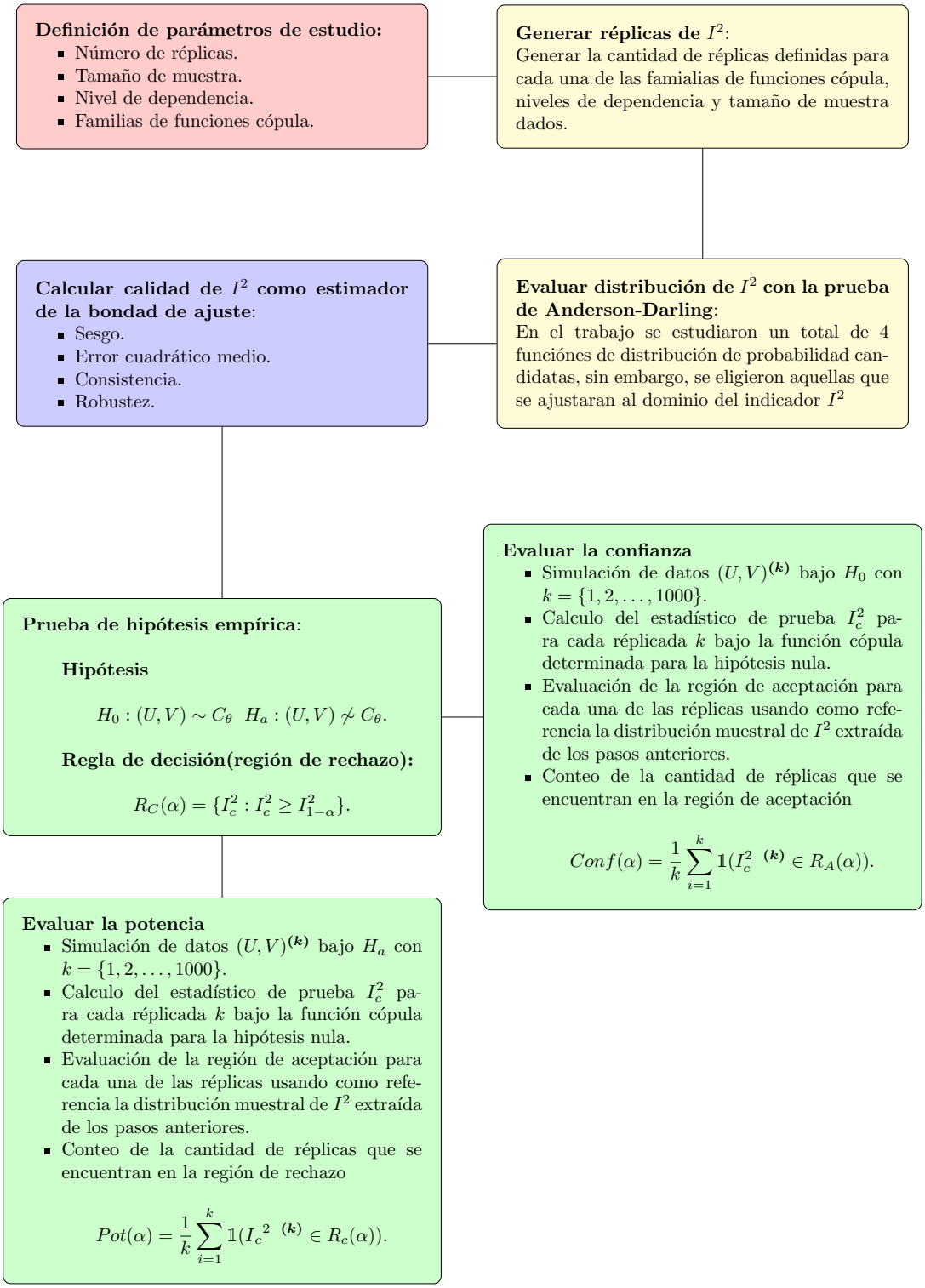


**Figura 6**: Metodología de estudio. Fuente: Elaboración propia

## 3.2. Generación de réplicas del indicador

En esta sección se buscó abordar el primer proceso del esquema de estudio presentado en la Figura 6, realizando un análisis más profundo de los pasos a seguir para alcanzar el primer objetivo de este trabajo. Dado que este documento se basa en el indicador propuesto por Tovar Cuevas, Portilla Yela, & Achcar (2019), el punto más relevante es la réplica del indicador $I(model[j]) = \sqrt{\sum_{i=1}^{n}(q_n^j(u_i, v_i) - q_\theta^j(u_i, v_u))^2}$ y su adaptación a la finalidad de este estudio ($I^2$). Primero, se profundizó en la parte interna de la raíz, con la cual se buscaba encontrar una distribución de probabilidad que se ajustara al indicador $I^2$, para luego llegar a conclusiones sobre $I$. Este detalle fue expresado previamente en la Sección 2.3.1 del marco teórico.

### 3.2.1. Réplica del estudio principal (Tovar Cuevas et al., 2019)

Los conceptos a tener presentes para la réplica del estudio son la generación de las variables aleatorias con estructura de dependencia, expresadas en Tovar Cuevas et al. (2019), y el coeficiente de concordancia de Ledwina, vista en la Sección 2.2.8.

Para la simulación de variables aleatorias con estructuras de dependencia existen varios métodos. Sin embargo, en este caso se mencionan dos, los cuales son:

- El método de la transformada inversa (Sección 2.2.2).
- El método de la función característica de la función arquimediana (Sección 2.2.3).

El primer método es una forma generalizada para la simulación de funciones cópula, debido a su facilidad para ser replicado, el otro es un método más enfocado en las funciones cópula de la familia arquimediana. Estos métodos son necesarios para la réplica del estudio, ya que no todas las funciones cópula son fácilmente simulables por el método de la transformada inversa, la cual necesita de una función cópula invertible de forma cerrada en su primera derivada.

### 3.2.2. Generación del indicador

Debido a que el indicador desarrollado en Tovar Cuevas et al. (2019) no posee una distribución conocida, es necesaria la generación de un número grande de réplicas del mismo para poder crear una distribución muestral que sea válida como una función asintótica. La generación del indicador se realizó con los insumos de varias muestras aleatorias de variables dependientes, donde sus métodos de generación fueron explicadas previamente. Por otro lado, se necesita la función cópula empírica (función de distribución empírica de los datos simulados, Sección 2.2.4) y el coeficiente de concordancia de Ledwina (Sección 2.2.8).

Para la generación de variables aleatorias con estructura de dependencia, **primero** fue necesario definir el tamaño de muestra ($n$), para este estudio se tuvieron un total de 5 niveles $\{30, 50, 100, 500, 1000\}$. **Segundo**, se precisó definir una función cópula fija para cada réplica del indicador. Para las funciones cópulas, se definieron 5 familias de cópulas de estudio: Clayton, FGM, Frank, Gumbel-Barnett, Gumbel-Hougaard. **Por último**, se fijaron niveles de dependencia distintos fuerte, moderado, débil, la mayoría

de estos casos de estudio fueron extradaídos de Tovar Cuevas et al. (2019) debido a que el estudio trató de abarcar la confianza del indicador en cada uno de los casos.

Con lo anterior, los pasos a seguir están definidos. **Primero**, se generó un vector aleatorio tal que $(U, V) \sim C_\theta$ de tamaño $n$ con base en una función cópula y un parámetro de dependencia $\theta$:

$$(U, V) \sim C_\theta.$$

**Segundo**, se estimó la función de distribución empírica o cópula empírica del vector aleatorio (Sección 2.2.4):

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(u_i \leq u, v_i \leq v).$$

**Tercero**, se evaluó el coeficiente de Ledwina (Sección 2.2.8) tanto empírico como teórico, en otras palabras, se calculan $q_n(u, v) = \frac{C_n(u,v)-uv}{\sqrt{uv(1-u)(1-v)}}$ y $q_\theta(u, v) = \frac{C_\theta(u,v)-uv}{\sqrt{uv(1-u)(1-v)}}$, respectivamente. **Cuarto**, se calculó el indicador de Tovar (Sección 2.3.1) de la forma:

$$I^2 = \sum_{i=1}^{n} (q_n(u_i, v_i) - q(u_i, v_i))^2.$$

Por último, se replicaron $m$ veces los pasos del primero al cuarto para hallar la distribución asintótica del indicador, un ejemplo a esto es la generación de 100 muestras de una función cópula Clayton con $\theta = 8$ con un total de 10 réplicas. Los pasos algorítmicos del procedimiento se encuentran a continuación.

1. Inicie $n$, $m$ y $\theta$
2. Genere una muestra de tamaño n del vector aleatorio $(U, V) \sim C_\theta$.
   - El vector $(U, V)$ representa dos variables aleatorias dependientes que tienen un comportamiento dado por una función cópula $C_\theta$.
   - Se requiere que $n$ sea un número entero positivo el cual define el tamaño de la muestra generada.
3. Calcule $C_n(u, v)$ y $C_\theta(u, v)$.
   - $C_n(u, v)$ es la cópula empírica, calculada a partir de los datos simulados. Se utiliza para estimar la dependencia entre las variables.
   - $C_\theta(u, v)$ es la cópula teórica, que se define a partir del modelo de cópula seleccionado y el parámetro de dependencia $\theta$.
4. Calcule $\Pi = u * v$ y $w = \sqrt{u * (1 - u) * v * (1 - v)}$.
   - $\Pi$ es el producto de las marginales de dos variables aleatorias, que se utiliza como un valor de referencia para medir la dependencia.
   - $w$ es una medida de la variabilidad conjunta de las variables $U$ y $V$, que se necesita para normalizar los coeficientes de concordancia.
5. Calcule $q_n(u, v) = \frac{C_n(u,v)-\Pi(u,v)}{w}$ y $q_\theta(u, v) = \frac{C_\theta(u,v)-\Pi(u,v)}{w}$ usando los valores previamente calculados.

- $q_n(u, v)$ es el coeficiente de concordancia empírico, que mide la diferencia entre la cópula empírica y la esperada bajo independencia.
- $q_\theta(u, v)$ es el coeficiente de concordancia teórico, que mide la diferencia entre la cópula teórica y la esperada bajo independencia.

6. Genere $I^2 = \sum_{i=1}^{n} \left(q_n(u_i, v_i) - q_\theta(u_i, v_i)\right)^2$.
   - $I^2$ es una medida que cuantifica la variación total entre los coeficientes de concordancia empírico y teórico a través de todas las muestras generadas.
7. Repita $m$ veces los pasos 2 hasta el 7 y regrese $\boldsymbol{I^2} := \{I_1^2, \cdots, I_m^2\}$ ó $\boldsymbol{I} := \{I_1, \cdots, I_m\}$.
   - $\boldsymbol{I^2}$ contiene todos los valores del indicador al cuadrado obtenidos en las $m$ repeticiones.
   - $\boldsymbol{I}$ son los valores del indicador.

### 3.2.3. Ajuste del indicador a una posible función de distribución

Después de la evaluación del indicador para $m$ réplicas, se consideraron diversas **distribuciones candidatas** propuestas por el paquete **distfit**, el cual busca minimizar la suma de cuadrados de una regresión lineal entre la distribución empírica del indicador y una distribución teórica. En otras palabras, se realiza internamente una comparación cuantil-cuantil. Para este estudio, se seleccionaron las **tres mejores distribuciones** que cumplían con la condición de tener la mínima suma de cuadrados y que, además, se ajustaran al rango del indicador ($I^2 \in [0, \infty]$).

Posteriormente, se utilizó la prueba formal de Anderson-Darling para identificar cuáles de las **funciones de distribución candidatas** eran las más adecuadas en términos de bondad de ajuste respecto a la distribución del indicador. Esta prueba se realizó un total de $k$ veces con lo cual se buscaba consistencia en los resultados de la prueba, esto haciendo uso de $k$ remuestreos de las $m$ réplicas de $I^2$,

$$I^{2\ (\mathbf{k})} = \{I_1^{2\ \ k}, \ldots, I_m^{2k}\ \},$$

en donde $I^{2\ (\mathbf{k})}$ denota el remuestreo k-ésimo de $I^2$ y se realizó la prueba de Anderson-Darling a cada una de estas, donde

$$H_0 : I^2 \sim f(\boldsymbol{\lambda}) \quad H_a : I^2 \not\sim f(\boldsymbol{\lambda}),$$

donde $\boldsymbol{\lambda}$ corresponde al vector de parámetros de la posible función de distribución candidata, por último, se le calculó el promedio del *valor-p* obtenido en todas las pruebas.

## 3.3. Calidad del estimador $\boldsymbol{I^2}$ como medida de la bondad de ajuste de los datos a una función cópula

Dadas las réplicas extraídas en el primer objetivo se buscó analizar el comportamiento del indicador $I^2$ (Tovar Cuevas et al., 2019) como una medida estimada de la bondad de ajuste de conjuntos de datos a una función cópula. En esta Sección se definen

todos los procesos necesarios para cumplir el segundo objetivo de este trabajo. La calidad del estimador juega un papel importante en el análisis del indicador, ya que se evalúa cuán precisa es la estimación de la bondad de ajuste bajo la hipótesis nula de la prueba, en este caso $I^2 \to 0$. En términos generales, cuando se hace referencia a la calidad de la estimación, en la líteratura normalmente se calcula el sesgo del indicador respecto a su valor esperado y al error cuadrático medio entre el valor esperado del indicador y su estimación (Portilla Yela & Tovar Cuevas, 2018). En ambos casos, se espera que, si el indicador realmente es una buena medida de la bondad de ajuste de los datos a una función cópula, tanto el sesgo como el error cuadrático medio tiendan a cero a medida que el tamaño de la muestra aumente.

Haciendo uso de técnicas de remuestreo y aleatorización (Bootstrap), se busca replicar la metodología implementada en Portilla Yela & Tovar Cuevas (2018) para medir la calidad del indicador como un estimador de la bondad de ajuste entre conjuntos de datos y funciones cópula. En este caso de estudio, el estimador de un buen ajuste es el promedio del indicador $I^2$ ante el mayor tamaño de muestra propuesto en este trabajo ($n = 1000$), ya que esta representa el valor esperado de la bondad de ajuste, es decir, $E[I]$ si se conociera su distribución real. Tome en encuenta que cuando se haga referencia a $k$ será tomado como el número de remuestreos booststrap realizados a la distribución ya generada del indicador y cuando se refiera a $\boldsymbol{k}$ se denominará como el conjunto de todos los remuestreos.

Entonces, se define el valor real para $I^2$ como el promedio de $I^2$ cuando el tamaño de muestra es $n = 1000$ esto para cada nivel de dependencia y se compara contra varias estimaciones de este valor para los distintos tamaños de muestra y sus distintos parámetros de dependencia, este proceso se realiza por cada parámetro de dependencia evaluado en la función cópula.

$$\bar{I}^{2\ (\boldsymbol{k})} = \frac{1}{m}\sum_{i=1}^{m} I_i^{2\ (\boldsymbol{k})}, \tag{12}$$

$$\hat{E}[\bar{I}^2] = \frac{1}{k}\sum_{i=1}^{k} \bar{I}^{2\ (\boldsymbol{k})}, \tag{13}$$

$$\hat{V}[\bar{I}^2] = \frac{1}{k-1}\sum_{i=1}^{k}\left(\bar{I}^{2\ (\boldsymbol{k})} - \hat{E}[\bar{I}^2]\right)^2. \tag{14}$$

Para la estimación del error cuadrático medio y del sesgo, se realizaron una secuencia de pasos. **Primero**, se calcula la estimación del valor esperado del indicador. **Segundo**, se obtiene $k$ muestras booststrap de la distribución empírica del indicador (extraer muestras de m réplicas para cada remuestreo). En este caso, se elegieron un total de 1000 remuestreos, sin embargo, este valor fue seleccionado a conveniencia.

Por último, se calcula el promedio del vector de muestras Bootstrap $I^{2\ *k}$, este proceso se repite varias veces con tal de conseguir consistencia en los resultados.

En la Figura 7 se puede visualizar un ejemplo de la distribución del indicador, donde se extraen $k = 4$ muestras bootstrap (a forma de ejemplo). En la parte superior de la gráfica se presenta la distribución del indicador resultante de $m = 1000$ réplicas de su proceso generador (Sección 3.2.2), mientras que en la parte inferior de la gráfica se almacenan las densidades de las 4 muestras bootstrap de la distribución empírica del indicador. Se puede observar que el patrón de comportamiento muestral no varía significativamente pues las distribuciones de los remuestreos siguen siendo continuas y asimétricas hacia la derecha.

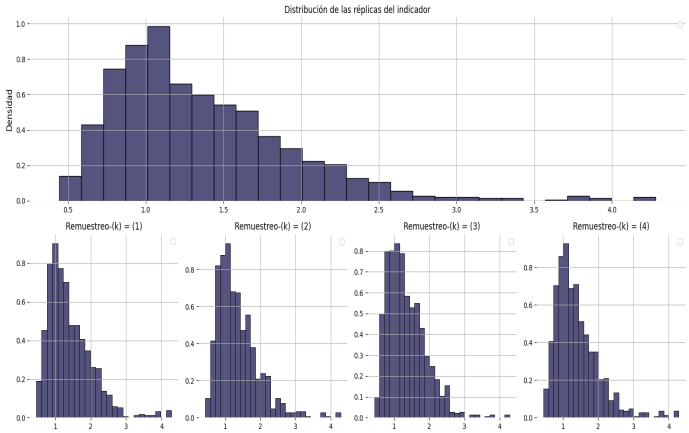


**Figura 7**: Proceso de remuestreo, donde la función densidad superior es la muestral y las 4 inferiores son muestras bootstrap de la misma. Fuente: Elaboración propia

### 3.3.1. Sesgo en las estimaciones

Con base a el estudio realizado por Portilla Yela & Tovar Cuevas (2018). se utilizó el sesgo y el error cuadrático medio como medidas para conocer la calidad de estimación de la bondad de ajuste que proporciona el indicado $I^2$. Esto haciendo uso del promedio de varios remuestreos booststrap realizados sobre la distribución muestral de $m$ réplicas del indicador $I^2$. Esta expresión se encuentra en la Ecuación 13

$$Bias(\bar{I}^2) = \bar{I}^2 - \hat{E}[\bar{I}^2]. \tag{15}$$

Lo que se espera del sesgo, para asumir que la estimación es buena, es que este sea cercano a cero, es decir, $Bias(\hat{\bar{I}}^2) \to 0, \quad n \to \infty, \quad m \to \infty$. En otras palabras, se asume que a medida que el tamaño de muestra aumente y el número de réplicas támbien lo haga, lo que se espera es que $\bar{I}^2$ tienda a $\hat{E}[\bar{I}^2]$ y este a su vez tienda a el verdadero valor de $I^2$.

### 3.3.2. Error cuadrático medio en las estimaciones

El error cuadrático medio se compone del sesgo y la varianza de la estimación (Ecuación 14). Debido a que se desconoce la distribución real $I^2$, no es posible conocer de forma exacta si este es un estimador de mínima varianza, por lo cual se usó booststrap para medir el error cuadrático medio de las estimaciones de $I^2$. Esto también permite conocer la estabilidad de la prueba de hipótesis.

$$MSE(\bar{I}^2) = \hat{V}[\bar{I}^2] + Bias(\bar{I}^2)^2. \tag{16}$$

Cuando el sesgo es nulo y la varianza es mínima, entonces se entiende que los estimadores del indicador $I^2$ son aquellos de mejor calidad.

1. Genere varios remuestreos de la distribución muestral del indicador $I^2$ la cual previamente fue generada a partir de m réplicas de muestras de tamaño n de un vector aleatorio $(U, V) \sim C_\theta$.
2. Calcule el promedio del k-ésimo remuestreo $\bar{I}^{2\ (\boldsymbol{k})} = \frac{1}{m}\sum_{i=1}^{m} I_i^{2\ (\boldsymbol{k})}$.
3. Calcule $\hat{E}[\bar{I}^2] = \frac{1}{k}\sum_{i=1}^{k} \bar{I}^{2\ (\boldsymbol{k})}$.
4. Calcule $\hat{V}[\bar{I}^2 =] = \frac{1}{k-1}\sum_{i=1}^{k}\left(\bar{I}^{2\ (\boldsymbol{k})} - \hat{E}[\bar{I}^2]\right)^2$.
5. Calcule el sesgo de estimación con $Bias(\bar{I}^2) = \bar{I}^2 - \hat{E}[\bar{I}^2]$.
6. Calcule el error cuadrático del estimador con $MSE(\bar{I}^2) = \hat{V}[\bar{I}^2] + Bias(\bar{I}^2)^2$

### 3.3.3. Robustez

La robustez de un indicador se mide en que tan capaz es de no verse afectado por datos atípicos o muestras externas (Casella & Berger, 2001), por lo cual se proponen una serie de mixturas entre funciones cópula, donde, se desea si el indicador es robusto ante muestras atípicas o ruidos de otras funciones cópulas, con lo cual se propone un porcentaje de contaminación $\delta = \{0.01, 0.05, 0.1, 0.15, 0.2\}$ los cuales permitan conocer visualmente que tan bueno es el indicador si se encuentra con datos atípicos.

| Función cópula original | Función cópula de ruido |
|---|---|
| Clayton | Gumbel-Hougaard |
| Gumbel-Barnett | Frank |
| Frank | Gumbel-Hougaard |
| Gumbel-Hougaard | Clayton |
| FGM | Frank |

El proceso de la mixtura se hizo seleccionando un valor aleatorio entre $\{1, 2, 3, \ldots, n\}$ e insertando un valor de ruido a ese elemento en la muestra original, esto se replicó con el porcentaje del ruido que se deseaba introducir en la muestra original de $(U, V)$. Luego, se calculó el indicador $I^{*2}$ haciendo uso de la muestra contaminada. Se realizó esto por cada réplica de $(U, V)$ que se calculada de $I^2$ y se crean un grupo de indicadores

contaminados a distintos porcentajes de ruido (Wilcox, 2012). Por último, se calculó

$$D(\delta) = \frac{1}{m}\sum_{i=1}^{m}\frac{|I^{*2}_{i(\delta)} - I^2_i|}{I^2_i}.$$

para evaluar la proporción de cambio entre el valor real del indicador al cuadrado y el valor del mismo frente a los datos contaminados, lo esperado de este indicador es que no sea robusto pues, $s^2$ en la literatura (Casella & Berger, 2001) es un estimador no robusto, debido a que, cualquier perturbación en la muestra genera grandes cambios en la suma de diferencias al cuadrado.

## 3.4. Prueba de hipótesis

El tercer objetivo de este trabajo es definir un lineamiento a seguir para llevar a cabo una prueba de hipótesis sobre la distribución muestral del indicador $I^2$. Previamente, se buscó una distribución que pudiese ajustarse al comportamiento aleatorio del indicador, sin embargo, esto resulta poco probable para muestras pequeñas ($n \leq 30$). Con el fin de establecer una prueba se propone un procedimiento similar al realizado por Huang & Prokhorov (2014).

En esta sección, se plantea una prueba de hipótesis simple utilizando métodos de remuestreo sobre la distribución empírica del indicador. Esta prueba se compone de dos elementos importantes. **En primer lugar**, se utilizó una distribución muestral bajo la hipótesis nula, que establezca que **los datos se ajustan a una función cópula**. **En segundo lugar**, se empleó un estadístico de contraste el cual aseguró la hipótesis alternativa, la cual sostiene que **los datos no se ajustan a la función cópula**.

***Transformación de la muestra***

Dado que la prueba se planteó para vectores de variables aleatorias bidimensionales, se tomó una muestra de un vector aleatorio $(\boldsymbol{X}, \boldsymbol{Y})$, al cual se le realizó una transformación al espacio $(0,1)^2$ (vector aleatorio $(\boldsymbol{U}, \boldsymbol{V})$). Dicha transformación se construye utilizando el teorema de la transformación integral de probabilidad (Sección 2.1.2) o por medio de los rangos de $X$ y $Y$ por separado (marginales empíricas).

Como ejemplo para la variable aleatoria $X$, se utilizó el conjunto de rangos $R$, donde $R_1, R_2, \ldots, R_n$ son los rangos respectivos para $X_1, X_2, \ldots, X_n$. Siguiendo la metodología de autores como Ledwina (2014), se adoptó la transformación $U_i = \frac{R_i}{n+1}$ $i = 1, \ldots, n$. **Nota:** En el cálculo de la Confianza$(1-\alpha)$ y en la Potencia$(1-\beta$ no se usó esta transformación debido a que los datos provenían directamente de la función cópula a evaluar.

***Asignación de la función cópula***

La asignación de la función cópula se realizó bajo el supuesto inicial de que $(U, V) \sim C_\theta$:

$$H_0 : (U, V) \sim C_\theta \quad \text{y} \quad H_1 : (U, V) \not\sim C_\theta.$$

Con este supuesto, se procede a calcular la estimación del parámetro de la función cópula ($\theta$).

***Criterio del $\tau$ de Kendall inverso***

La estimación de $\theta$ se lleva a cabo mediante el método de los momentos de la función cópula (Sección 2.2.6). Esta forma de estimación permite realizar una aproximación del parámetro real utilizando el coeficiente de concordancia de Kendall:

$$\tau_\theta = 4E[C_\theta(u, v)] - 1$$

Sin embargo, al no conocer ni el parámetro ni el coeficiente real, se emplean métodos no paramétricos y métodos numéricos recursivos. El parámetro estimado es aquel que satisface la condición de que la diferencia entre el tau de Kendall no paramétrico $\tau_0$ y el paramétrico $\tau_\theta$ sea nula:

$$\hat{\theta} : \tau_n - 4E[C_\theta(u, v)] + 1 = 0$$

Otra condición adicional es que $\tau_0$ no puede ser mayor al valor máximo del $\tau_\theta$ de Kendall ni menor al mínimo del coeficiente de concordancia de Kendall para esa función cópula en particular. Una forma de convertir esta estimación en una más flexible es asumir que existe un error aleatorio tal que $|\tau_0 - \text{mín}\{\tau_\theta\}| \leq \varepsilon$ o $|\tau_0 - \text{máx}\{\tau_\theta\}| \leq \varepsilon$. En otras palabras, existe un margen de error entre el coeficiente real y el muestral. Esto también asegura la imposibilidad de comparar una muestra con una función cópula que sea muy distinta a la propuesta.

Se usó la función de densidad de la función cópula, además, se exploraron alternativas no analíticas como la estimación de la segundad derivada parcial por medio del teorema del valor medio, de la forma

$$c_\theta(u, v) \approx \lim_{h \to 0} \frac{C_\theta(u + h, v + h) - C_\theta(u - h, v + h) - C_\theta(u + h, v - h) + C_\theta(u - h, v - h)}{4h^2},$$

la cual es una solución más simple que el cálculo de algunas funciones de densidad de funciones cópula no estudiadas en este trabajo e incluso más simple que algunas de las funciones de densidad de algunas funciones cópula estudiadas como la Gumbel-Barnett y la Gumbel-Hougaard.

Por otro lado, la integral correspondiente a $E[C_\theta(u,v)]$ fue calculada por medio de la regla del trapecio de la forma que se adapte al caso de estudio bivariado [5]

$$\begin{aligned}E[C_\theta(u,v)] \approx \frac{1}{4n^2}\,[&C_\theta(u_1,v_1)c_\theta(u_1,v_1) + C_\theta(u_n,v_1)c_\theta(u_n,v_1)+\\ &2\sum_{i=2}^{n-1} C_\theta(u_i,v_1)c_\theta(u_i,v_1) + 2\sum_{i=2}^{n-1} C_\theta(u_i,v_n)c_\theta(u_i,v_n)+\\ &4\sum_{j=2}^{n-1}\sum_{i=2}^{n-1} C_\theta(u_i,v_j)c_\theta(u_i,v_j)+\\ &2\sum_{j=2}^{n-1} C_\theta(u_1,v_j)c_\theta(u_1,v_j) + 2\sum_{j=2}^{n-1} C_\theta(u_n,v_j)c_\theta(u_n,v_j)+\\ &C_\theta(u_1,v_n)c_\theta(u_1,v_n) + C_\theta(u_n,v_n)c_\theta(u_n,v_n)]\,.\end{aligned}$$

Después de realizar la estimación del parámetro $\hat{\theta}$, se asume este como semilla numérica para el uso del método de máxima verosimilitud haciendo uso del método número **Nelder-Mead** (Sección 2.2.6). Este método busca estar lo más cercano al parámetro real de la muestra, utilizando la función de densidad de la función cópula $c_\theta(u,v)$ [6]:

$$\tilde{\theta} = arg\max_{\boldsymbol{\theta}\in\Theta}\sum_{i=1}^{n}\log(c_\theta(u_i,v_i)).$$

***Valor-p y región de confianza***

Para el valor-p de la prueba de hipótesis, dada la función cópula supuesta y la estimación del parámetro $\tilde{\theta}$, se realizaron simulaciones de tamaño n de la función cópula junto al parámetro de dependencia. Con estas simulaciones se calculan $m = 1000$ réplicas del estimador, para después de obtener la distribución muestral del indicador bajo la hipótesis nula $I^2$, se calcula el indicador estadístico de contraste $I_c^2$ [7]. Posteriormente, se calcula empíricamente $P(I^2 \geq I_c^2)$ y este valor se compara con el nivel de significancia de la prueba.

La región de aceptación se define con base a lo anterior, ya que al definir un nivel de significancia $\alpha$, se puede determinar si el valor calculado pertenece o no a un conjunto de valores de aceptación

$$R_A(\alpha) = \left\{I_c^2 : I_c^2 < I_{1-\alpha}^2\right\},$$

[5] La regla del trapecio para una dimensión se define como $\int_{\mathbb{R}} f(x)\,dx \approx \frac{1}{2n}\left(f(x_1) + 2\sum_{i=2}^{n-1} f(x_i) + f(x_n)\right)$, donde $n$ es el tamaño de la muestra aleatoria. Con el uso de esta fórmula fue posible realizar la aproximación del $\tau$ de Kendall de la función Gumbel-Barnett (Anexo A.1.1).

[6] El métdo número implementado fue **Nelder-Mead** el cual usa la función **fmin** del paquete de python scipy

[7] El subíndice $c$ se refiere a calculado y no a la cópula.

esto basandose en la literatura (Genest et al., 2009; Huang & Prokhorov, 2014). El algorítmo a usar en una muestra real:

1. Inicie una matriz de datos $X, Y$.
2. Transforme las variables $X, Y$ al espacio $0, 1$ haciendo uso de los rangos de las variables, denotando $u_i = \frac{R_i}{n+1}$ y $v_i = \frac{S_i}{n+1}$.
3. Calcule $\tau_n = \frac{2}{n(n-1)} \sum_{i<j} sgn(x_i - x_j) sgn(y_i - y_j)$ como el estadístico no paramétrico del $\tau$ de Kendall.
4. Inicie $C_\theta$ como la función cópula esperada.
5. Estime $\theta$ tal que $\tau_0 - \tau_\theta = 0$, o sea, $\tau_0 - (4E(C_\theta(u, v)) - 1) = 0$ [8] y almacénelo en $\hat{\theta}$.
6. Calcule $\tilde{\theta}$ por máxima verosimilitud de forma iterativa (haciendo uso de métodos conocidos como Netwon-Raphson o métodos implementados en paquetes estadísticos como Nelder-Mead del paquete scipy) usando $\hat{\theta}$ como semilla.
7. Inicie $m$ como número de repeticiones para la generación del estadístico.
8. Genere $I^2$ usando $m$ simulaciones de $(U^*, V^*) \sim C_{\tilde{\theta}}$ a partir de la estimación de la muestra inicial.
9. Calcule $I_c^2 = \sum_{i=1}^{n} \left( \frac{C_n(u_i, v_i) - C_{\tilde{\theta}}(u_i, v_i)}{\sqrt{u_i v_i (1-u_i)(1-v_i)}} \right)^2$ usando los valores observados. Inicie $\alpha$ como nivel de significancia.
10. Verifique si $I_c^2$ pertenece a la región de aceptación $R_A(\alpha) = \{I_c^2 : I_c^2 < I_{1-\alpha}^2\}$.
11. Calcule $\frac{1}{m} \sum_{i=1}^{m} \mathbb{1}(I_i^2 \geq I_c^2)$ y almacénelo en $Valor - p$.

### 3.4.1. Confianza de la prueba

La confianza de la prueba se basa en la pregunta fundamental: ¿cuántas veces no rechazará la función cópula, suponiendo que esta es la verdadera, si se realiza esta prueba en múltiples ocasiones? La respuesta a esta incógnita se obtiene mediante la construcción de una región de aceptación, que se define para cada submuestra del indicador y el indicador calculado para cada muestra $(U, V) \sim C_\theta$ bajo la hipótesis nula.

De manera similar a los estimadores Bootstrap, se toman remuestreos de la distribución empírica del estimador. Estas muestras se denominan $I^{2\ (\boldsymbol{k})}$, donde el superíndice $\boldsymbol{k}$ indica la $k$-ésimo evaluación de la prueba de hipótesis:

$$H_0 : (U, V) \sim C_\theta \quad \text{y} \quad H_1 : (U, V) \not\sim C_\theta$$

***Cálculo de la confianza de la prueba***

Para realizar este cálculo, se usaron las $m$ réplicas logradas en el primer objetivo, con las cuales se usaron para medir el cuantil $1 - \alpha$ para $I^2$ en cada uno de los casos de estudio, se calculó la confianza asumiendo $\theta$ conocido, pues, este parámetro también pudo haber sido obtenido bajo el método del $\tau$ de Kendall inverso y/o el método de máxima verosimilitud de la función cópula ($\tilde{\theta}$).

$$I_{1-\alpha}^2 = Q_{1-\alpha}(I^2).$$

[8] Esto se puede realizar usando las funciones de densidad de la función cópula o por el teorema del valor medio, además, del uso de la integral si esta es cerrada o usando la regla del trapecio.

Como contraste a este cuantil, se generaron $k$ vectores aleatorios $(U,V)^{(\boldsymbol{k})} \sim C_{\tilde{\theta}}$ haciendo uso del parámetro de dependencia inicial o de su estimación $\tilde{\theta}$ y la función cópula previamente establecida. Estos vectores se utilizan para calcular diversos estadísticos de contraste, que se agrupan como:

$$\boldsymbol{I}_c^{2\ (\boldsymbol{k})} = \{I_c^{2\ (\boldsymbol{1})}, I_c^{2\ (\boldsymbol{2})}, I_c^{2\ (\boldsymbol{3})}, \ldots, I_c^{2\ (\boldsymbol{k})}\},$$

donde, cada uno de los estadísticos calculados en el remuestro k-ésimo esta dado por

$$I_c^{2\ (k)} = \sum_{i=1}^{n} \frac{C_n(u_i^{(k)}, v_i^{(k)}) - C_\theta(u_i^{(k)}, v_i^{(k)})}{\sqrt{\Pi(u_i^{(k)}, v_i^{(k)})(1-u_i^{(k)})(1-v_i^{(k)})}}.$$

Este proceso de remuestreo permite evaluar la calidad de la prueba de hipótesis bajo el conocimiento de la hipótesis nula.

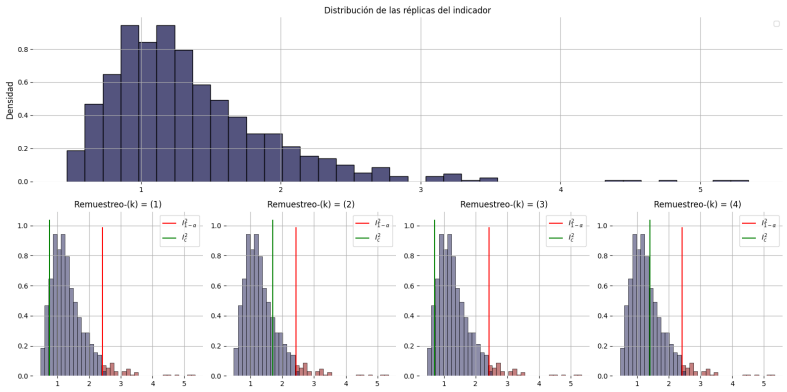


**Figura 8**: Cálculo de la confianza de la prueba haciendo uso de técnicas de aleatorización y remuestreo (muestra de m réplicas del indicador $I^2$ para una función cópula FGM con tamaño de muestra 100 y parámetro de dependencia $\theta = 0.2$).
Fuente: Elaboración propia

En la Figura 8 se visualiza la evaluación de la confianza de la prueba de hipótesis a partir de la distribución muestral de m réplicas del indicador $I^2$. En la gráfica se visualizan dos líneas, una roja que corresponde al valor máximo que permite la región de aceptación antes de determinar que un valor calculado no pertence a la función cópula y una línea verde, la cual permite ver el valor calculado bajo la función cópula esperada. El cálculo de la confianza se evalúa haciendo uso de varias comparativas entre el indicador calculado y los remuestreos con $k = 1000$ de la muestra de tamaño $m$ del indicador $I^2$.

$$Confianza_{(\alpha)} = P(\boldsymbol{I}_c^{2\ (\boldsymbol{k})} \in RA_{(\alpha)} \mid \boldsymbol{I}_c^{2\ (\boldsymbol{k})} \in H_0)$$

1. Inicie los parámetros $n$, $k$, $m$ y $\alpha$, $\theta$.
2. Dada la función cópula $C_\theta$ y el tamaño de muestra, genere una muestra de tamaño $m$ del indicador $I^2$ y considerelo como la distribución poblacional del estadístico.
3. A partir de las $m$ réplicas de $I^2$ calcule el cuantil $1-\alpha$.
4. Genere una muestra de tamaño $k$ del indicador $I^2$, donde $k = m$, haciendo uso de esta muestra como estadístico de contraste, cuando los datos provienen de la función cópula que esta bajo $H_0$ [9].
5. Calcule $\frac{1}{k}\sum_{i=1}^{k} \mathbb{1}(I_c^{2\ (\boldsymbol{k})} < I_{1-\alpha}^2)$ y almacénelo como la confianza de la prueba.
6. Para obtener mayor consistencia de los resultados, repita el procedimiento varias veces.

### 3.4.2. Potencia de la prueba

La metodología implementada para evaluar la potencia de la prueba es similar a la de la confianza. La diferencia de este proceso se encuentra en que el estadístico de contraste se genera para la hipótesis alternatica ($I_c^2$). En este caso, dado que no se dispone de una solución teórica o analitica para la función de distribución del indicador ($I^2$), **se opta por realizar simulaciones cuyo objetivo sea comprar las funciones cópula que tengan un rango absoluto en los coeficientes** del $\tau$ de Kendall o sea, que el $\tau$ de Kendall sea el mismo en valor absoluto.

Uno de estos casos involucra las funciones cópula FGM y Gumbel-Barnett ambas con rangos de dependencia similares, la primera con un $\tau$ de Kendall entre $[-0.222\ldots, 0.222\ldots]$, donde su parte negativa se encuentra dentro del rango del $\tau$ de Kendall de la función cópula Gumbel-Barnett. Como en este estudio los parámetros de la FGM son solo positivos, para poderla comprar con la Gumbel-Barnett se tomó el rango negativo también, pues, en terminos absolutos corresponden al mismo coeficiente de concordancia. Esto también se aplicará a otras funciones cópula como en el caso de la Clayton o la Frank frente a la Gumbel-Barnett.

[9] El índice $k$ no indica una potencia, sino que se refiere a la $k$-ésima muestra.

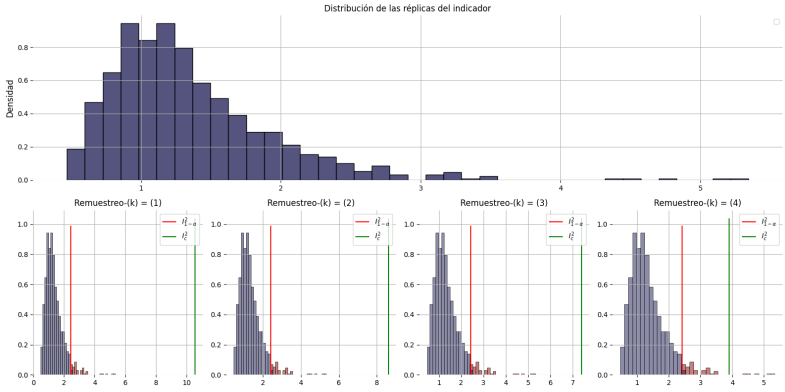


**Figura 9**: Cálculo de la potencia de la prueba haciendo uso de técnicas de aleatorización y remuestreo (muestra de m réplicas del indicador $I^2$ para una función cópula FGM con tamaño de muestra 100 y parámetro de dependencia $\theta = 0.2$ con $\tau_\theta = 0.04...$, frente a una muestra aleatoria de tamaño de muestra 100 y parámetro de dependecia $\theta = 0.093$ con $\tau_\theta = 0.04...$ ). Fuente: Elaboración propia

En la Figura 9 se puede visualizar el comportamiento de la prueba bajo la hipótesis alternativa. En este caso, la línea verde se encuentra alejada de la distribución muestral del indicador y del límite superior (línea roja). Esto contrasta con lo que ocurre en la Figura 8, donde la línea verde se sitúa por atras de la línea del límite superior lo que significa que el valor-p o la probabilidad de que el valor pertenezca la distribución del indicador $I^2$ bajo las mismas condiciones es alta. Este comportamiento debe persistir si la prueba posee una buena potencia. Su ecuación se presenta a continuación:

$$Potencia_\beta = P\left(I_c^2 \in R_C(\beta) \mid H_0 \text{ es falsa}\right)$$

en esta ocasión $\beta$ hace referencia al error tipo II de la prueba de hipótesis. A continuación, se presenta el algoritmo correspondiente a esta parte:

1. Inicie los parámetros $n$, $k$, $m$ y $\beta$, $\theta$.
2. Si $\tau_{H_a} \subseteq \tau_{H_0}$, estime $\theta_{H_a}$ a partir de la relación que existe entre el $\tau_{H_0}$ y $\theta_{H_a}$ haciendo uso del $\tau$ inverso de Kendall para la hipótesis alternativa. Si no está contenido, detenga el proceso.
3. Genere una muestra de tamaño $m$ $I^2$ bajo la hipótesis nula.
4. A partir de las $m$ réplicas de $I^2$ calcule el cuantil $1 - \beta$.
5. Genere $k$ muestras de tamaño $n$ de $(U^{(\boldsymbol{k})}, V^{(\boldsymbol{k})}) \sim C_{\tilde{\theta}\ H_a}$.
6. Calcule $I_c^2$ usando $\sum_i^n (q_n(u_i^{(\boldsymbol{k})}, v_i^{(\boldsymbol{k})}) - q_{\theta\, H_0}(u_i^{(\boldsymbol{k})}, v_i^{(\boldsymbol{k})}))$.
7. Calcule $\frac{1}{k}\sum_{i=1}^k \mathbb{1}(I_c^{2\ (\boldsymbol{k})} \geq I_{1-\beta})$ y almacénelo como la potencia de la prueba.

Como consideración, procedimiento para lograr la potencia y la confianza debe de ser repetido un total de $r$ veces, esto para obtener consistencia en los resultados de la prueba. Dada la metodología previa, se puede entonces entrar a hablar de los resultados de este trabajo, pues, el uso de cada uno de lo métodos de esta Sección son

fundamentales para poder obtener resultados coherentes frente al estudio previamente realizado en Tovar Cuevas et al. (2019).

# 4. Resultados

Dentro de está Sección se encuentran los resultados de las investigaciones realizadas para cada objetivo, recordando que estos parten desde el problema de si el estadístico posee una distribución que se ajuste correctamente a el o no, para luego, tomarlo como punto de partida para los demás objetivos de este trabajo, se adelanta que en cuestiones de distribución, se encontró una posible distribución que se ajuste de forma empírica a la distribución muestral del indicador $I^2$, además, se encontraron hallazgos como que la distribución siempre es asimétrica hacia la derecha o que la mayoría de los valores se encuentran entre 0 y 5 (por supuesto, esto bajo algunas cópulas).

Para esté trabajo se tomaron en cuenta 5 familias cópula (Clayton, FGM, Frank, Gumbel-Barnett, Gumbel-Hougaard) con distintos niveles de dependencias marcados a detalle en Tovar Cuevas et al. (2019), los cuales, se clasificaron como (Alto, Moderado y Bajo), para 5 tamaños de muestra distintos 30, 50, 100, 500, 1000 con la finalidad de realizar un estudio más detallado del indicador ante muestras pequeñas y grandes.

## 4.1. Análisis de la distribución del indicador

Dada la cantidad de posibles combianciones entre funciones cópula, dependencias y tamaños de muestra. Se optó por usar métodos de visualización más compactos del proceso, donde, se puede explicar de forma gráfica o tabulada cada uno de los comportamientos de los distintos métodos de ajuste. Para este objetivo fueron usadas las Secciones 3.2, 3.2.3 con el fin de determinar el comportamiento del indicador $I^2$. La función cópula Gumbel-Barnett generó varias complicaciones en el desarrollo de este trabajo, pues, al no tener un intervalo para su coeficiente del $\tau$ de Kendall, este debió ser previamente estimado para realizar algunas comparatis. El estudio de la función de distribución para $I^2$ se divide en dos partes, un primer vistazo exploratorio, donde se encuentran un grupo de posibles funciones de distribución y la segunda, la evaluación de la distribución que mejor se adapte a la estructura matématica del indicador. **Nota:** las funciones de distribución son sugeriadas por el paquete **distfit** basandose en el uso de la suma de los cuadrados del error.

### 4.1.1. Función cópula Clayton: Análisis exploratorio $m = 1000$ réplicas de $I^2$.

El comportamiento del indicador $I^2$, se encuentra determinado por una distribución positiva, asimétrica hacia la derecha (Figura 10). En la Figura 10 se visualiza el comportamiento de la distribución del indicador, donde, al lado izquierdo, se ve su distribución asintótica $n = 1000$ y $m = 1000$, también, se observa que la función Gamma es una de las funciones que se ajustan bien a estos datos, además, de otras funciones como la Log-Normal, la Generalizada de valores extremos, la Beta de cuatro parámetros.

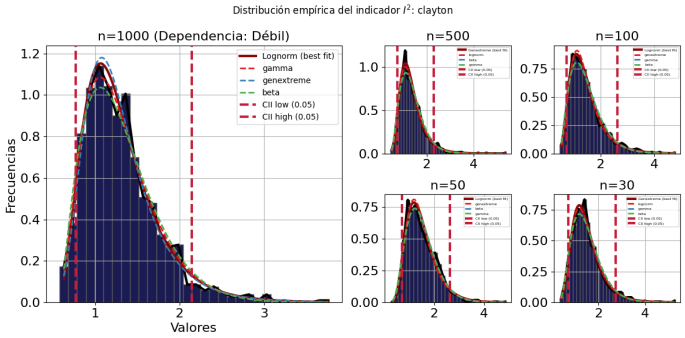


**Figura 10**: Distribución del indicador $I^2$: Clayton dependencia débil. Fuente: Elaboración propia

En la Figura 11 se visualiza el comportamiento de la distribución del indicador para una dependencia moderada de la Clayton; la asimétria de está distribución es menos marcada que en el caso débil, sin embargo, las distribuciones recomendadas nuevamente por el paquete estadístico **distfit** son las funciones de distribución Gamma y Beta de cuatro parámetros, Log-Normal, Generalizada de valores extremos.

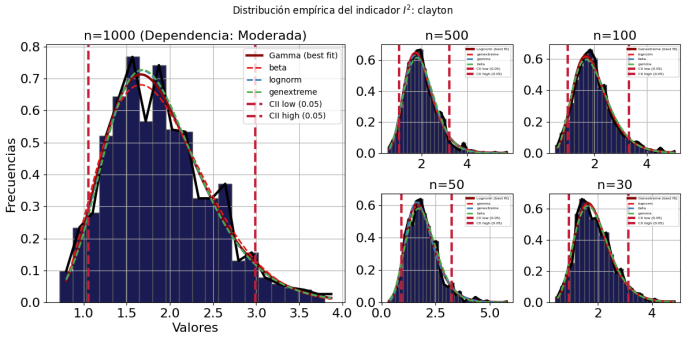


**Figura 11**: Distribución del indicador $I^2$: Clayton dependencia moderada. Fuente: Elaboración propia

Para la dependencia fuerte de la Clayton, se encuentra que la función de distribución Gamma también es una buena candidata para modelar el comportamiento del indicador. Todo esto visible en la Figura 12. Se puede ver, que a pesar de pertenecer a la misma familia de funciones cópula y que el paquete sugiere distribuciones parecidas, el comportamiento de la distribución no es exactamente el mismo, pues a medida que se fortalece el parámetro de dependencia, la distribución toma una forma más dispersa sobre el eje $x$.

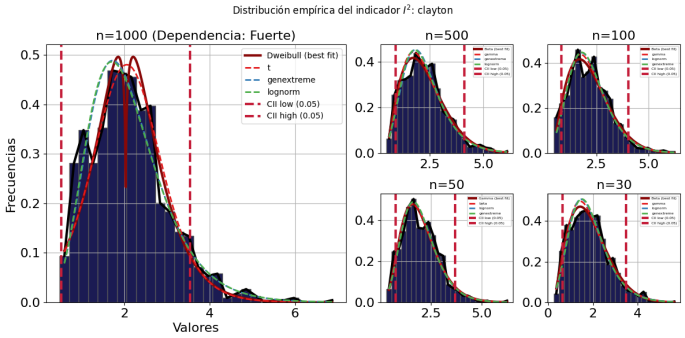


**Figura 12**: Distribución del indicador $I^2$: Clayton dependencia fuerte. Fuente: Elaboración propia

### 4.1.2. Función cópula FGM: Análisis exploratorio $m = 1000$ réplicas de $I^2$.

En la Figura 13 se visualiza la distribución del indicador de prueba cuando la función cópula FGM está modelando relaciones débiles, en este caso, de $\theta = 0.2$, se ve que el comportamiento es muy similar al de la Clayton con dependencia débil $\theta = 2$ (Figura 10), pues, sus valores, tienen asimétria hacia la derecha y el valor más grande de la distribución es muy cercano a 3. Esto para el caso asintótico $m = 1000$ y $n = 1000$. Se visualiza nuevamente que las candidatas a distribuciones son las funciones Gamma, Generalizada de valores extremos, t-Student y la Log-Normal.

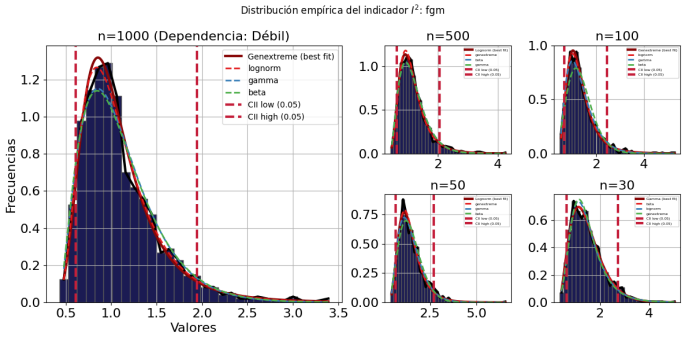


**Figura 13**: Distribución del indicador $I^2$: FGM dependencia débil. Fuente: Elaboración propia

Cuando $\theta = 0.5$ se ve que el estadístico de prueba posee una distribución (Figura 14) muy similar frente a la FGM de dependencia débil (Figura 13), esto puede dar un indicio de que la distribución del indicador es sensible a la distribución, sin embargo,

no a la estimación de los parámetros. Las distribuciones recomendadas nuevamente son, la función Gamma, la función Log-Normal, la función Beta de cuatro parámetros y la Generalizada de valores extremos.

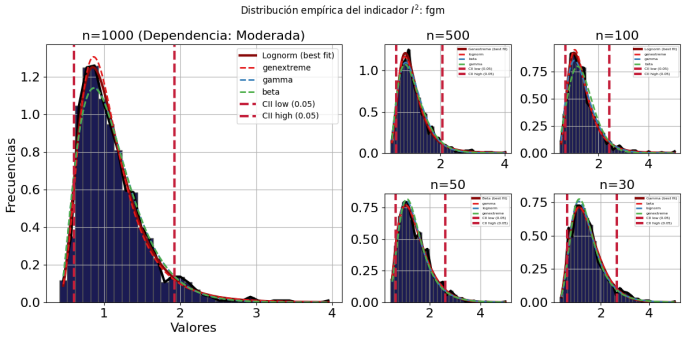


**Figura 14**: Distribución del indicador $I^2$: FGM dependencia moderada. Fuente: Elaboración propia

En la Figura 15 se observa que el cambio en la distribución es casi imperceptible por no decir que no es visible, lo cual, puede reflejar problemas de potencia en la prueba no parámetrica de la FGM para detectar alguna mala específicidad en los parámetros de la cópula. La función de distribución Gamma y otras funciones son candidatas a modelar el comportamiento del indicador de prueba.

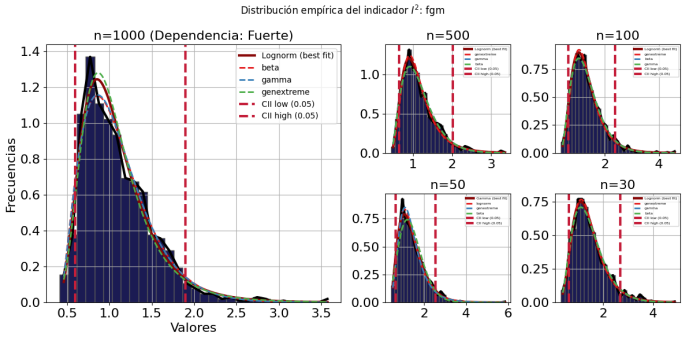


**Figura 15**: Distribución del indicador $I^2$: FGM dependencia fuerte. Fuente: Elaboración propia

### 4.1.3. Función cópula Frank: Análisis exploratorio $m = 1000$ réplicas de $I^2$.

En la Figura 16 se visualiza el comportamiento del indicador para la función cópula Frank, se observa que se comporta similar al indicador $I^2$ que genera la función cópula FGM en todos sus niveles de dependencia, visible en la Figura 15. Las funciones de distribución sugeridas son la función Gamma, la Generalizada de valores extremos, la Beta de cuatro parámetros y la Log-Normal.

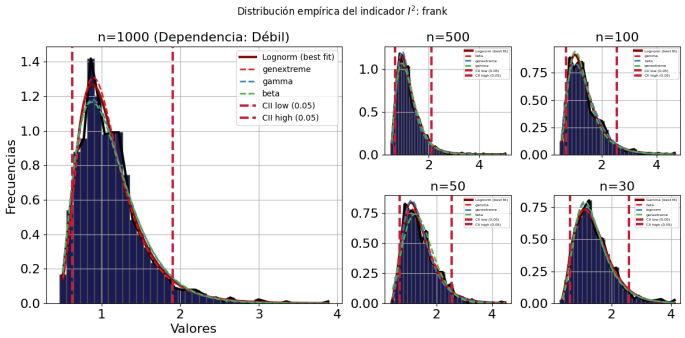


**Figura 16**: Distribución del indicador $I^2$: Frank dependencia débil. Fuente: Elaboración propia

En la Figura 17 se observa que si existe diferencia respecto a su dependencia débil (Figura 16), sin embargo, las distribuciones sugeridas son las mismas, pero, el orden es distinto en cuanto al ranking de ajuste $RSS$, siendo la Gamma la posible función de distribución que mejor modele el comportamiento del indicador.

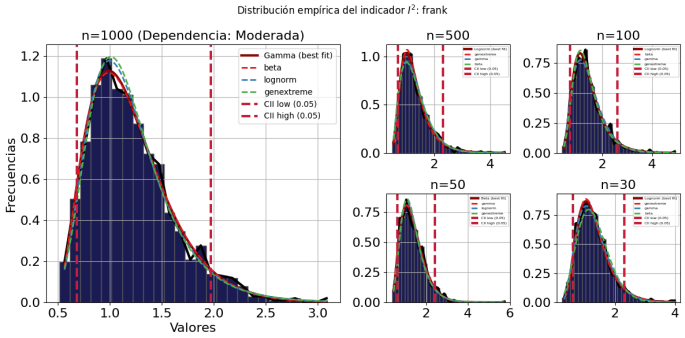


**Figura 17**: Distribución del indicador $I^2$: Frank dependencia moderada. Fuente: Elaboración propia

En la Figura 18 se puede visualizar el comportamiento del indicador para la función cópula Frank cuando esta posee una dependencia fuerte, su comportamiento a medida que aumenta su dependencia es menos asimétrico hacia la derecha, lo cual puede generar dificultad para ajustarle un modelo completamente asimétrico. Nuevamente la función de distribución Gamma aparece sugerida en el paquete utilizado (**distfit**), al igual que la función Beta de cuatro parámetros, Generalizada de valores extremos y la Log-Normal.

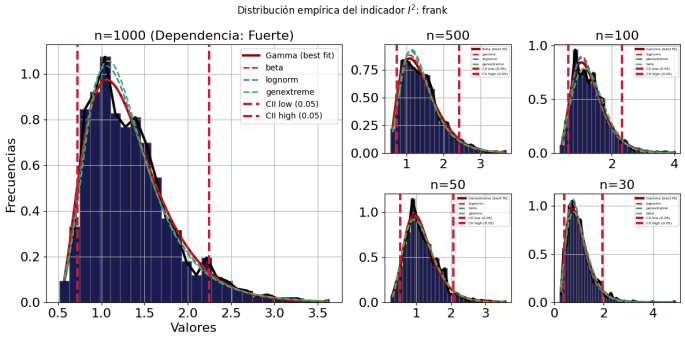


**Figura 18**: Distribución del indicador $I^2$: Frank dependencia Fuerte. Fuente: Elaboración propia

### 4.1.4. Función cópula Gumbel-Barnett: Análisis exploratorio $m = 1000$ réplicas de $I^2$.

En la Figura 19 se puede ver como la tendencia de la distribución es muy similar a las previas, además, se ve que casi siempre esté estadístico tiende a tener valores entre 0.5 a 2. Las distribuciones son las mismas, sin embargo, el orden otorgado por el indicador $RSS$ es un poco distinto en comparativa a las funciones cópula estudiadas previamente. La Gamma mostró ser una buena candidata para modelar el comportamiento de la función muestral del indicador $I^2$.

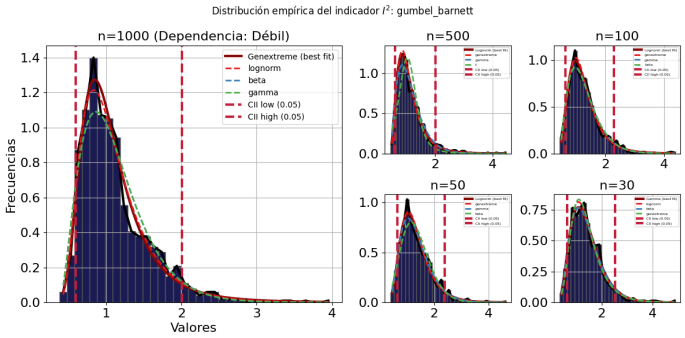


**Figura 19**: Distribución del indicador $I^2$: Gumbel-Barnett dependencia débil.
Fuente: Elaboración propia

En la Figura 20 se encuentra el comportamiento del indicador cuando $\theta = 0.5$, se ve que este se comporta similar a la misma cuando la dependencia es débil. Las distribuciones sugeridas son la Generalizada de valores extremos, la Log-Normal, la Beta de cuatro parámetros y la Gamma, que son las que se ha observado en este estudio.

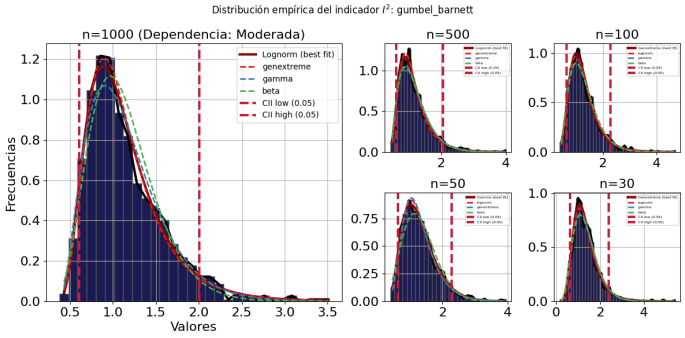


**Figura 20**: Distribución del indicador $I^2$: Gumbel-Barnett dependencia moderada.
Fuente: Elaboración propia

En la Figura 21 se observa el comportamiento de la distribución del indicador cuando la dependencia de la Función cópula Gumbel-Barnett es alta. Se puede notar un desplazamiento hacia la derecha, sin embargo, las funciones sugeridas para este estadístico son las mismas para modelar el comportamiento distribucional de $I^2$.

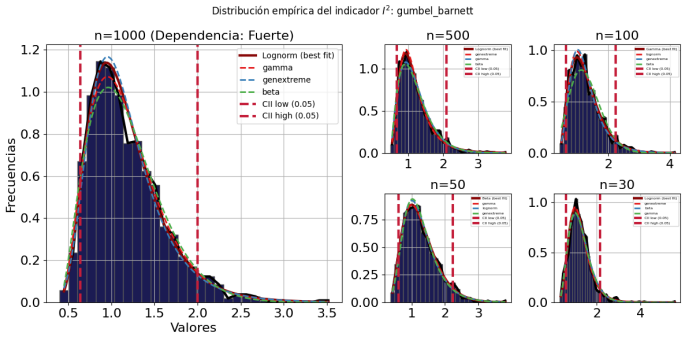


**Figura 21**: Distribución del indicador $I^2$: Gumbel-Barnett dependencia fuerte.
Fuente: Elaboración propia

### 4.1.5. Función cópula Gumbel-Hougaard: Análisis exploratorio $m = 1000$ réplicas de $I^2$.

En la Figura 22 se puede observar un comportamiento diferente a cuanto las demás cópulas análizadas previamente, se puede observar que la distribución es un poco menos asimétrica hacia la izquierda. No obstante, las funciones sugeridas fueron la Gamma y otras funciones como la Log-Normal, la Beta y la Generalizada de valores extremos.

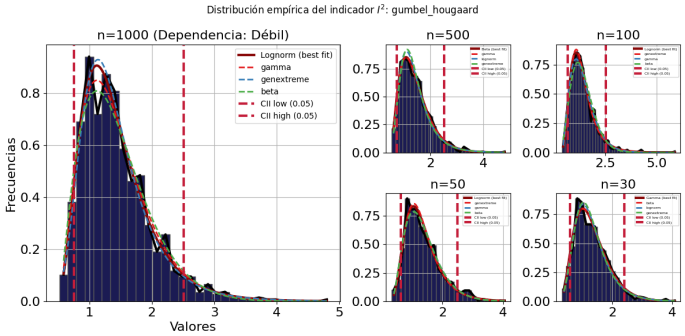


**Figura 22**: Distribución del indicador $I^2$: Gumbel-Hougaard dependencia débil.
Fuente: Elaboración propia

En la Figura 23 se visualiza que en vez de esta crecer hacia la derecha como las demás distribuciones del indicador, esta decrece por lo menos para el caso donde la dependencia es moderada ($\theta = 2$), esto puede deberse a menor independencia en la cópula frente al parámetro de dependencia débil ($\theta = 1.25$). Las funciones sugeridas para modelar su comportamiento son las funciones Gamma y Beta, Log-Normal y $I^2$.

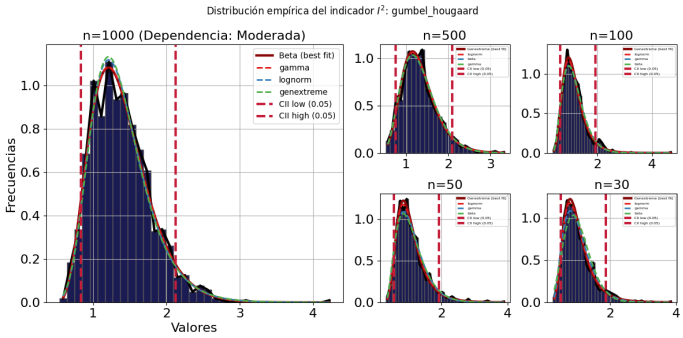


**Figura 23**: Distribución del indicador $I^2$: Gumbel-Hougaard dependencia moderada. Fuente: Elaboración propia

En la Figura 24 se encuentra la distribución del último estadístico de este análisis previo antes de usar la prueba de Anderson-Darling para medir el mejor ajuste. La distribución del indicador, en muchos casos tiende a ser simétrica a medida que aumenta la medida de dependencia entre las variables aleatorias y se llega a lo asintótico. Las funciones sugeridas en este caso son Gamma y otro grupo de funciones como la GEV, la Log-Normal y la Beta, sin embargo, las últimas no se adaptan al dominio del indicador.

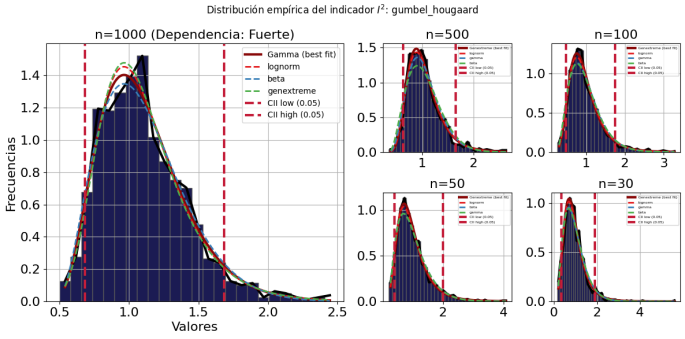


**Figura 24**: Distribución del indicador $I^2$: Gumbel-Hougaard dependencia fuerte. Fuente: Elaboración propia

Las estimaciones de cada una de las distribuciones previas se encuentran en el Anexo A.3 de este documento, las cuales fueron calculadas por medio del método de momentos.

### 4.1.6. Prueba de Anderson-Darling: Validación del análisis exploratorio de las simulaciones.

La prueba de Anderson-Darling (Sección 2.1.5), garantiza una buena estimación para conocer si un conjunto de datos pertenecen o no a una distribución. En este caso, al tener una gran cantidad de combinaciones entre variables y posibles funciones, se decidió realizar una Tabla de pruebas de bondad de ajuste (Tabla 1) donde se puede ver los resultados de las 75 pruebas de Anderson-Darling que se hicieron bajo el supuesto de que el indicador tomaba una función de distribución positiva y asimétrica hacia la derecha. En la Tabla 1 se puede visualizar que en ningún caso se ajustó a una Gamma de dos parámetros. Por otro lado, la función Gamma de tres parámetros se ajustó en todos los casos, sin embargo, el ajuste de parámetros es muy variable entre funciones cópula y entre sus niveles de dependencia, por lo cual, la estimación de la función Gamma depende en alguna medida de la información de la función cópula (Anexo A.3). **Nota:** no todas las funciones de distribución sugeridas fueron usadas para evaluar, esto es debido a que el indicador resultante es una suma de difencias al cuadrado por lo cual, no cabe a priori una razón por la cual se deba de usar la función de distribución Log-Normal, la función de distribución Beta de cuatro parámetros, ni la función Generalizada de valores extremos la cual tiene su dominio marcado en todos los reales.

**Tabla 1**: Prueba de Anderson-Darling. Ya que el nivel de significancia se fija como $\alpha = 0.05$, tome todo valor menor a este como un no se distribuye bajo esta función y todo valor mayor a este como que se distribuye bajo la función.

| | Dependencia | Débil | | Moderada | | Fuerte | |
|---|---|---|---|---|---|---|---|
| | Distribución | Gamma $(\zeta, \nu)$ | Gamma $(\zeta, \eta, \nu)$ | Gamma $(\zeta, \nu)$ | Gamma $(\zeta, \eta, \nu)$ | Gamma $(\zeta, \nu)$ | Gamma $(\zeta, \eta, \nu)$ |
| Función Cópula | n | | | | | | |
| Clayton | 30 | 0.001 | 0.188 | 0.001 | 0.214 | 0.001 | 0.225 |
| | 50 | 0.001 | 0.157 | 0.001 | 0.215 | 0.001 | 0.199 |
| | 100 | 0.001 | 0.204 | 0.001 | 0.211 | 0.001 | 0.213 |
| | 500 | 0.001 | 0.179 | 0.001 | 0.209 | 0.001 | 0.220 |
| | 1000 | 0.001 | 0.238 | 0.001 | 0.221 | 0.001 | 0.170 |
| FGM | 30 | 0.001 | 0.243 | 0.001 | 0.232 | 0.001 | 0.224 |
| | 50 | 0.001 | 0.229 | 0.001 | 0.243 | 0.001 | 0.244 |
| | 100 | 0.001 | 0.222 | 0.001 | 0.215 | 0.001 | 0.238 |
| | 500 | 0.001 | 0.201 | 0.001 | 0.220 | 0.001 | 0.218 |
| | 1000 | 0.001 | 0.190 | 0.001 | 0.188 | 0.001 | 0.228 |
| Frank | 30 | 0.001 | 0.237 | 0.001 | 0.220 | 0.001 | 0.240 |
| | 50 | 0.001 | 0.237 | 0.001 | 0.243 | 0.001 | 0.230 |
| | 100 | 0.001 | 0.222 | 0.001 | 0.220 | 0.001 | 0.240 |
| | 500 | 0.001 | 0.223 | 0.001 | 0.210 | 0.001 | 0.243 |
| | 1000 | 0.001 | 0.216 | 0.001 | 0.243 | 0.001 | 0.240 |
| Gumbel-Barnett | 30 | 0.001 | 0.228 | 0.001 | 0.125 | 0.001 | 0.210 |
| | 50 | 0.001 | 0.230 | 0.001 | 0.236 | 0.001 | 0.239 |
| | 100 | 0.001 | 0.222 | 0.001 | 0.209 | 0.001 | 0.240 |
| | 500 | 0.001 | 0.168 | 0.001 | 0.174 | 0.001 | 0.205 |
| | 1000 | 0.001 | 0.138 | 0.001 | 0.203 | 0.001 | 0.233 |
| Gumbel-Hougaard | 30 | 0.001 | 0.231 | 0.001 | 0.180 | 0.001 | 0.194 |
| | 50 | 0.001 | 0.222 | 0.001 | 0.191 | 0.001 | 0.220 |
| | 100 | 0.001 | 0.233 | 0.001 | 0.198 | 0.001 | 0.160 |
| | 500 | 0.001 | 0.247 | 0.001 | 0.220 | 0.001 | 0.118 |
| | 1000 | 0.001 | 0.242 | 0.001 | 0.214 | 0.001 | 0.230 |

## 4.2. Evaluación de la calidad de $I^2$ como estimador de la bondad de ajuste entre conjuntos de datos y funciones cópulas.

La evaluación de la calidad de $I^2$ como indicador de bondad de ajuste para conjuntos de datos frente a funciones cópulas, es necesaria para asegurarse que las estimaciones que se obtengan serán representativas para la hipótesis nula o al valor esperado de la distribución real del indicador $I^2$. La metodología con la cual se obtuvieron estos resultados se encuentra expresada en Sección 3.3, en la cual se realizaron distintas técnicas de estimación por remuestreo basado en estudios previos (Portilla Yela & Tovar Cuevas, 2018). Los resultados generales muestran que el error cuadrático medio y el sesgo tienden a ser más pequeños a medida que el tamaño de muestra $n$ y el número de réplicas $m$ del indicador crecen.

Las estimaciones de $I^2$ son poco robustas, esto debido a que el indicador es conformado a partir de la suma de errores al cuadrado entre una función cópula y una función cópula empírica, por lo cual, si uno de los valores es atípico, el estimador será fácilmente desplazado hacia la derecha (ver Anexo A.3.1 en el cual se visualizan los niveles de contaminación 0.01,0.05, 0.1, 0.15, 0.20 y como crece la diferencia absoluta del indicador a medida que aumenta el tamaño de muestra).

### 4.2.1. Función cópula Clayton: Evaluación del sesgo y el error cuadrático medio para evaluar la eficiencia y la consistencia en la estimación.

En la Figura 25 se puede visualizar en el lado superior el sesgo del promedio del indicador, el cual, se asume como el valor que se espera que tome el indicador $I^2$ para la función cópula Clayton. Se puede ver, que la distribución del indicador $I^2$ es tendiende a un valor particular a medida que el tamaño de muestra $n$ crece y la cantidad de réplicas $m$ sirve como un estabilizador para controlar la variabilidad de las muestras del indicador. Se separó la comparativa según la dependencia con la cual se mide el indicador $I^2$ pues, su comportamientos puede depender de los parámetros de la función cópula. Se puede observar que el estimador es insesgado para casos donde $n$ tiende a infinito, por otro lado, en muestras pequeñas se presentan variaciones, esto por la aleatoriedad del proceso, la cual fue controlada en su mayoría por varias repeticiones de la simulación.

En la Figura 25 también se puede ver al lado inferior el error cuadrático medio del indicador $I^2$ bajo la función cópula Clayton, esta medida también depende del tamaño de muestra $n$, la cantidad de réplicas $m$ y el nivel de dependencia $\theta$. La dependencia parece ser significativa para el comportamiento del indicador $I^2$, pues, a medida que este aumenta, el error cuadrático medio y el sesgo tienden a ser más altos cuando el tamaño de la muestra es pequeño.

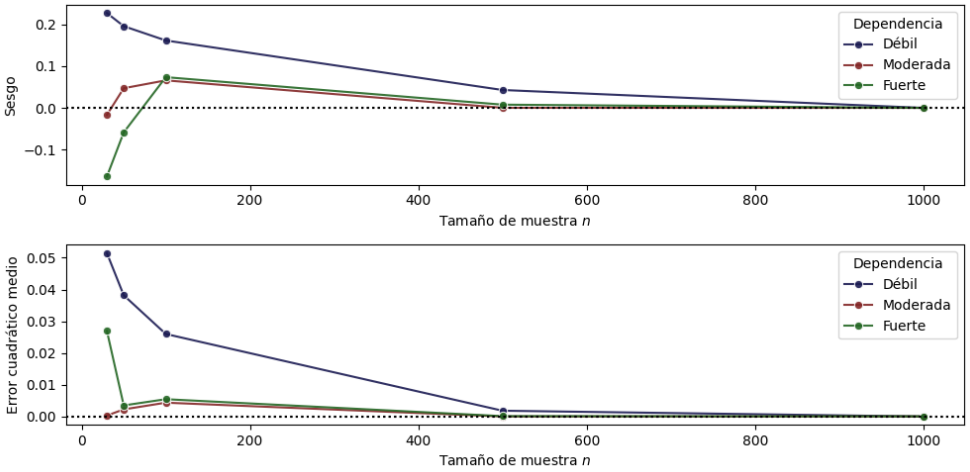


**Figura 25**: Estimaciones del error cuadrático medio y del sesgo del indicador $I^2$ para la función cópula Clayton. Fuente: Elaboración propia

### 4.2.2. Función cópula FGM: Evaluación del sesgo y el error cuadrático medio para evaluar la eficiencia y la consistencia en la estimación.

En la Figura 26 se puede visualizar el sesgo del indicador $I^2$ al lado superior del gráfico, en este, se observa que a diferencia de la función cópula Clayton vista en la Figura 25 el parámetro de la función cópula FGM no genera un cambio distintivo en el comportamiento de los gráficos de tendencia del sesgo. Sin embargo, se puede ver que el tamaño de la muestra $n$ vuelve a ser un factor fundamental para ver la evolución del indicador hacia un valor puntual, en este caso, se que el valor real del estimador se calculó como el promedio de las réplicas generadas bajo un tamaño de muestra lo suficientemente grande. **Nota:** Función cópula FGM tiene un comportamiento partícular con el indicador $I^2$, pues esta al modelar dependencias muy débiles y contener la expresión de la función cópula independencia interna en ella de forma explícita ($C_\theta(u,v) = uv + \theta uv(1-u)(1-v)$), toma valores muy cercanos a cero en la expresión de $q_\theta(u,v) = \frac{C_\theta(u,v) - \Pi(u,v)}{\sqrt{uv(1-u)(1-v)}}$.

En la Figura 26 se puede observar el error cuadrático medio del indicador en el lado inferior del gráfico $I^2$. Se puede ver que este tiene modificaciones muy leves a medida que el tamaño de muestra $n$ aumenta o el nivel de dependencia se acerca a uno, por lo cual, si se toma $I^2$ como un estimador para la bondad de ajuste de $(U,V)$ frente a una función cópula, para la función cópula FGM su calidad en la prueba de hipótesis es independiente de la cantidad de réplicas $m$ que se hagan previo a la evaluación de la prueba de hipótesis.

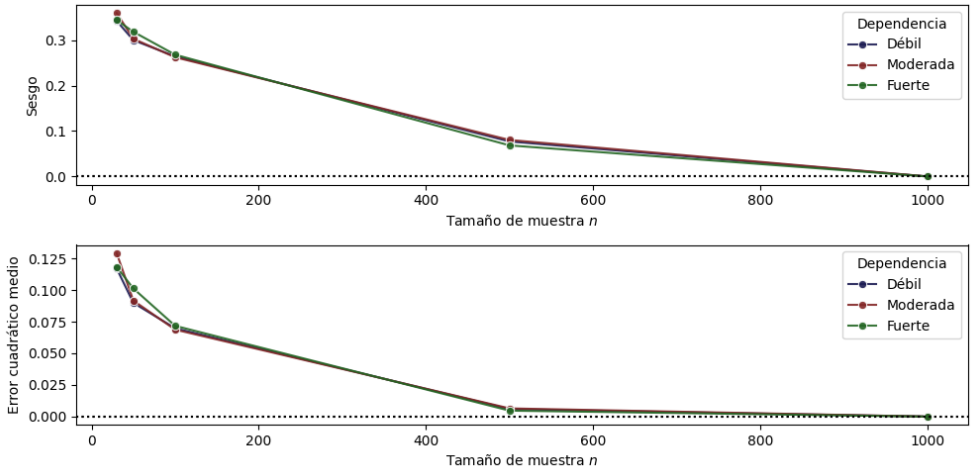


**Figura 26**: Estimaciones del error cuadrático medio y del sesgo del indicador $I^2$ para la función cópula FGM. Fuente: Elaboración propia

### 4.2.3. Función cópula Frank: Evaluación del sesgo y el error cuadrático medio para evaluar la eficiencia y la consistencia en la estimación.

En la Figura 27 se puede observar el sesgo de la estimación al lado superior de la gráfica, este en su mayoría dependende del tamaño de muestra $n$, el nivel de depencia $\theta$, pues las primeras para la función cópula Frank generan una mayor estabilidad en su estimación puntual ya que el valor del sesgo del indicador $I^2$ disminuye y es tendiente a cero cuando $n$ tiende a infinito.

El error cuadrático medio también se puede ver en la Figura 27, específicamente en la parte inferior del gráfico. El error cuadrático medio tiene un comportamiento decreciente únicamente cuando el tamaño de muestra $n$ previos a calcular $I^2$ aumente. Por otro lado, se visualiza que el sesgo y el error cuadrático medio se ven afectos a medida que el parámetro de dependencia $\theta$ crece. Lo que sugiere que los errores entre $q_n(u, v)$ y $q_\theta(u, v)$ son más variables a medida que $\theta \to \infty$.

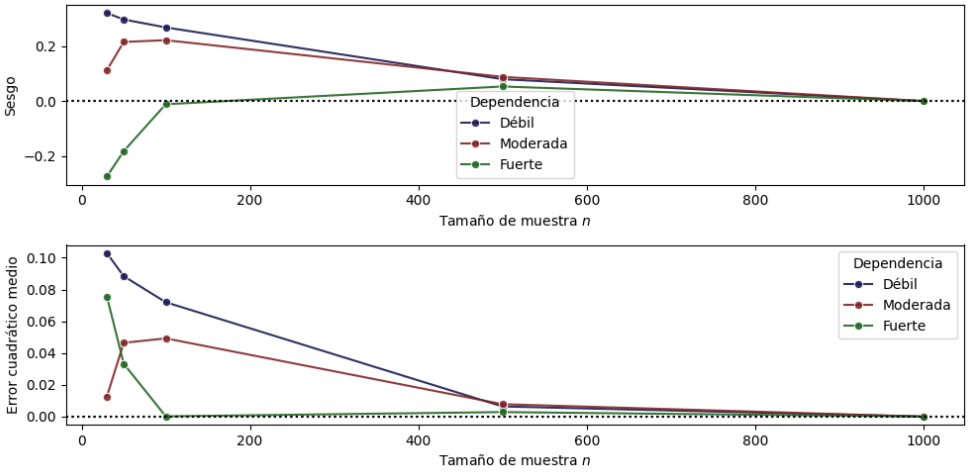


**Figura 27**: Estimaciones del error cuadrático medio y del sesgo del indicador $I^2$ para la función cópula Frank. Fuente: Elaboración propia

### 4.2.4. Función Cópula Gumbel-Barnett: Evaluación del sesgo y el error cuadrático medio para evaluar la eficiencia y la consistencia en la estimación.

En la Figura 28 se pueden ver las medidas del sesgo y el error cuadrático medio del promedio del indicador $I^2$. En el lado superior del gráfico, se encuentra el sesgo del promedio del indicador $I^2$, el cual tiende a cero de forma estrictamente positiva a diferencia de las Figuras 25, 26 y 27. Esta función cópula tiene una tendencia y estabilidad pronunciada frente a sus parámetros de dependencia $\theta$ a medida que estos se acercan a uno (dependencia fuerte), el sesgo genera una tendencia más pronunciada a cero a medida que el tamaño de muestra $n$ de $(U, V)$ aumenta.

En la Figura 28 también se evidencia el comportamiento decreciente del error cuadrático medio en el lado inferior de la gráfica. Se puede ver que a medida que aumenta el nivel de dependencia $\theta$, la brecha que existe entre los errores para cada tamaño de muestra tiende a disminuir más rápido, lo cual fue visto con anterioridad en las funciones cópulas ya análizadas.

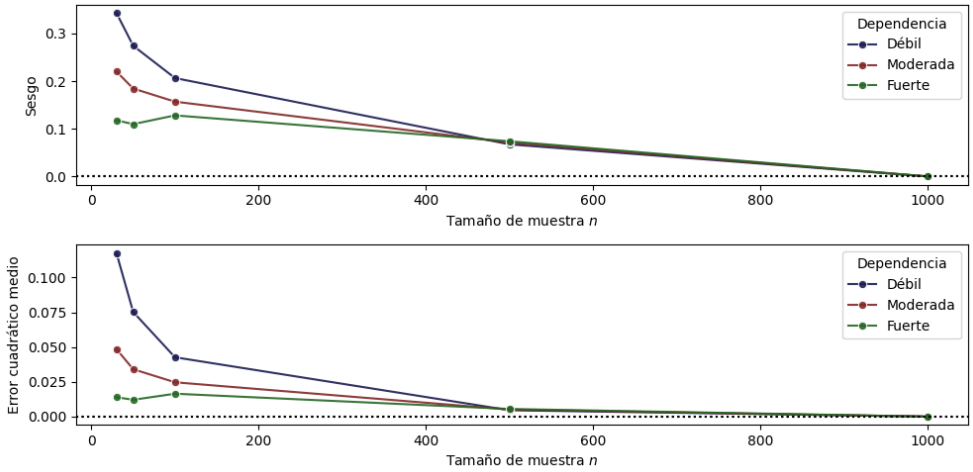


**Figura 28**: Estimaciones del error cuadrático medio y del sesgo del indicador $I^2$ para la función cópula Gumbel-Barnett. Fuente: Elaboración propia

### 4.2.5. Función cópula Gumbel-Hougaard: Evaluación del sesgo y el error cuadrático medio para evaluar la eficiencia y la consistencia en la estimación.

El sesgo de la función cópula Gumbel-Hougaard para el valor promedio de $I^2$ (valor esperado empípirico) se encuentra en la Figura 29, la cual muestra al lado superior que el indicador $I^2$ para la función cópula Gumbel-Hougaard disminuye su sesgo a medida que el tamaño de muestra $n$ de $(U, V)$ aumenta, por otro lado, el parámetro de dependencia $\theta$ genera un comportamiento distintivo para el sesgo, pues, a diferencia de las demás fucniones, no genera convergencia más rápida al valor espero de $I^2$.

A medida que el tamaño de la muestra $n$ se evidencia un comportamiento decreciente en el error cuadrático medio en la parte inferior de la Figura 29. Este comportamiento es visible en las demás funciones cópula previamente analizadas.

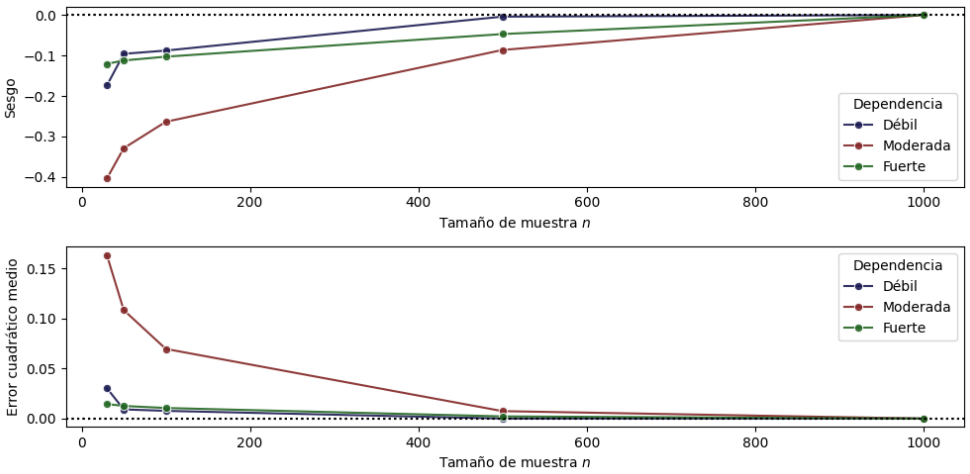


**Figura 29**: Estimaciones del error cuadrático medio y del sesgo del indicador $I$ función cópula Gumbel-Hougaard. Fuente: Elaboración propia

Se encontró en el estudio que la calidad de la estimación de $I^2$ como estimador de bondad de ajuste depende principalmente del tamaño de la muestra y su tendencia a ser muy grande, además, la cantidad de réplicas del indicador al cuadrado genera mayor estabilidad en los resultados, esto es acompañado de la literatura (Genest et al., 2009; Huang & Prokhorov, 2014) donde se usan una cantidad de réplicas $m$ grandes dependiendo al tamaño de la muestra que se va a utilizar. En algunas funciones cópula como la Gumbel-Barnett y la función cópula Gumbel-Hougaard es necesario de parámetros de dependencia grandes dentro de la función para obtener consistencia en las estimaciones según lo visto en la Figura 28.

## 4.3. Confianza y potencia de la prueba de hipótesis basada en el indicador $I^2$

Las pruebas de hipótesis (planteadas en la Sección 3.4), al igual que las estimaciones tienen dos medidas para calcular su calidad. **La primera**, la confianza, es una medida que representa que tan buena es la prueba cuando los datos si provienen de la distribución propuesta en la hipótesis nula (solo si se mide la bondad de ajuste), está medida tiene un valor complementario, el cual se denomina **error tipo I($\alpha$)** y este cuantifica que tan probable es que la prueba **rechace la hipótesis nula cuando esta es cierta**, en términos de bondad de ajuste corresponde a que tan probable es que la prueba me sugiera que los datos no se ajustan a una función copula dado que estos provienen de la función copula. **La segunda**, la potencia corresponde a la proporción de veces que la prueba sugiere que los datos no se distribuyen bajo la función cópula dado que las muestras obtenidas viene de una función cópula distinta, está medida también tiene su complemento, el cual corresponde al **error tipo II ($\beta$)** y este indica que tan probable es que **no se rechace** la prueba de hipótesis para una

función cópula dado que los datos vienen de otra (tome de referencia una función cópula Clayton frente a una función cópula Frank).

Para realizar comparativas entre distintas funciones cópulas **no se puede tomar de referencia el parámetro** de cada una, sino, **sus coeficientes de concordancia** los cuales permiten ver si sus niveles de dependencia son equivalentes, en este caso, para realizar las comparativas primero se fijó un valor de $\theta$ (dado en los casos de estudio) para luego calcular su $\tau_\theta$ de Kendall y por último calcular el **parámetro de dependencia** de la función copula alternativa para poder compararar ambas funciones cópula en las mismas condiciones.

### 4.3.1. Función cópula Clayton: confianza y potencia de la prueba de hipótesis.

**Tabla 2**: Confianza y potencia de la prueba de hipótesis cuando $H_0 : (U, V) \sim Clayton_\theta$, $H_a : (U, V) \not\sim Clayton_\theta$

| $H_0$ | Dependencia | n | Potencia $(1-\beta)$ FGM $\tau_\theta \in [-2/9, 2/9]$ | Frank $\tau_\theta \in (-\infty, \infty)$ | Gumbel-Barnett $\tau_\theta \in (-0.361, 0]$ | Gumbel-Hougaard $\tau_\theta \in [0, \infty)$ | Confianza $(1-\alpha)$ Clayton $\tau_\theta \in (-\infty, \infty)$ |
|---|---|---|---|---|---|---|---|
| Clayton | Débil $\tau_{\theta\approx 0.5,-0.333} = \pm 0.2$ | 30 | 0.06 | 0.08 | 0.66 | 0.13 | 0.96 |
| | | 50 | 0.12 | 0.12 | 0.93 | 0.27 | 0.95 |
| | | 100 | 0.19 | 0.22 | 1.00 | 0.54 | 0.95 |
| | | 500 | 0.91 | 0.90 | 1.00 | 1.00 | 0.95 |
| | | 1000 | 1.00 | 1.00 | 1.00 | 1.00 | 0.94 |
| | Moderado $\tau_{\theta\approx 2,-0.666} = \pm 0.5$ | 30 | | 0.21 | | 0.32 | 0.95 |
| | | 50 | | 0.31 | | 0.63 | 0.95 |
| | | 100 | | 0.64 | | 0.97 | 0.95 |
| | | 500 | | 1.00 | | 1.00 | 0.94 |
| | | 1000 | | 1.00 | | 1.00 | 0.94 |
| | Fuerte $\tau_{\theta\approx 8,-0.889} = \pm 0.8$ | 30 | | 0.07 | | 0.05 | 0.95 |
| | | 50 | | 0.14 | | 0.17 | 0.95 |
| | | 100 | | 0.38 | | 0.79 | 0.95 |
| | | 500 | | 1.00 | | 1.00 | 0.93 |
| | | 1000 | | 1.00 | | 1.00 | 0.93 |

En la Tabla 2 se pueden observar la potencia y la confianza de la prueba de hipótesis. Esto a partir del calculo de varias simulaciones de la prueba de hipótesis para ver la proporción de rechazos de los datos cuando sus valores no vienen de la distribución propuesta para $H_0$ (potencia) y su proporción de no rechazo cuando los datos provienen de la distribución tomada para $H_0$ (confianza). **Nota: Nota:** Aunque $\theta$ no toma valores negativos en los casos de estudio, los valores de referencia en la Tabla 2 se establecieron en función del coeficiente de concordancia $\tau_\theta$, el cual permite analizar las dependencias del vector aleatorio. Por ello, para todas las funciones cópula consideradas en este estudio, se tomaron tanto los valores positivos como negativos del coeficiente, cuando estos existen. .

La potencia estimada de la prueba de hipótesis para la función cópula Clayton dado que $(U, V) \sim FGM(\theta)$ aumenta progresivamente a medida que el tamaño de muestra $n$ crece. No obstante, en el caso de la función cópula Frank esta aumenta de forma

acelerada a medida que el parametro de dependencia aumente $\theta = 8$ y $\tau_\theta = 0.8$ y el tamaño de muestra $n$ es mayor a 100. Este comportamiento se repite por cada función cópula bajo el título de la Potencia, lo que desea cuantificar esto es cuantas veces la hipótesis nula fue rechazada dado que los datos provienen de una función cópula distinta con el mismo coeficiente de concordancia (en el caso de la Gumbel-Barnett, donde $n = 50$, $\theta = 0.5$ y $\alpha = 0.05$, la función FGM apenás tiene un 6 % de probabilidad de ser rechazada).

Por otro lado, la confianza $(1 - \alpha)$ evalua cuantas veces la función cópula puede ser no rechazada dado que si cumple la hipótesis nula, en este caso, la *Clayton* se evalúa con los mismos parámetros ya dados $\theta = \{0.5, 2, 8\}$, además los valores de la confianza oscilan alrededor del nivel de significancia $\alpha$, lo cual sugiere que es necesaria una cantidad mayor de réplicas $m > 1000$ para encontrar una convergencia exacta.

### 4.3.2. Función cópula FGM: confianza y potencia de la prueba de hipótesis.

**Tabla 3**: Confianza y potencia de la prueba de hipótesis cuando $H_0 : (U, V) \sim FGM_\theta$, $H_a : (U, V) \not\sim FGM_\theta$

| $H_0$ | Dependencia | n | Potencia $(1-\beta)$ Clayton $\tau_\theta \in (-\infty, \infty)$ | Frank $\tau_\theta \in (-\infty, \infty)$ | Gumbel-Barnett $\tau_\theta \in (-0.361, 0]$ | Gumbel-Hougaard $\tau_\theta \in [0, \infty)$ | Confianza $(1-\alpha)$ FGM $\tau_\theta \in [-2/9, 2/9]$ |
|---|---|---|---|---|---|---|---|
| FGM | Débil $\tau_{\theta=\pm 0.2} \approx \pm 0.044$ | 30 | 0.04 | 0.05 | 0.07 | 0.06 | 0.95 |
| | | 50 | 0.05 | 0.05 | 0.09 | 0.06 | 0.95 |
| | | 100 | 0.05 | 0.05 | 0.16 | 0.09 | 0.95 |
| | | 500 | 0.06 | 0.05 | 0.69 | 0.13 | 0.95 |
| | | 1000 | 0.09 | 0.05 | 0.95 | 0.20 | 0.95 |
| | Moderado $\tau_{\theta=\pm 0.5} \approx \pm 0.11$ | 30 | 0.04 | 0.04 | 0.24 | 0.06 | 0.95 |
| | | 50 | 0.05 | 0.06 | 0.43 | 0.11 | 0.95 |
| | | 100 | 0.06 | 0.06 | 0.79 | 0.17 | 0.95 |
| | | 500 | 0.29 | 0.06 | 1.00 | 0.48 | 0.95 |
| | | 1000 | 0.69 | 0.06 | 1.00 | 0.75 | 0.96 |
| | Fuerte $\tau_{\theta=\pm 0.9} = \pm 0.2$ | 30 | 0.04 | 0.05 | 0.74 | 0.09 | 0.96 |
| | | 50 | 0.06 | 0.07 | 0.94 | 0.16 | 0.96 |
| | | 100 | 0.12 | 0.07 | 1.00 | 0.30 | 0.96 |
| | | 500 | 0.93 | 0.08 | 1.00 | 0.90 | 0.95 |
| | | 1000 | 1.00 | 0.10 | 1.00 | 1.00 | 0.95 |

Al igual que con la función cópula Clayton, la Tabla 3 contiene la información sobre la confianza y la potencia de la prueba. En este caso, se resaltan dos casos partículares de la función cópula FGM, la cual tiene un comportamiento débil en la potencia con la función cópula Frank y la función cópula Gumbel-Barnett, las cuales también modelan comportamientos muy débiles cuando su coeficiente de concordancia de Kendall se evalua en {0.044, 0.11, 0.2}, sin embargo, si se puede determinar que la potencia de la función cópula FGM frente a las funciones cópula Gumbel-Barnett y Frank es muy baja para tamaños de muestra pequeños para la primera e incapaz de detectar diferencias entre la función cópula Frank y la FGM.

Por el lado de su confianza, la función cópula FGM toma oscilaciones de valores cercanos al nivel de sigficancia, lo cual es lo esperado para una cantidad 1000 réplicas y 1000 remuestreos.

### 4.3.3. Función cópula Frank: confianza y potencia de la prueba de hipótesis.

En la Tabla 4 se encuentran almacenados los resultados la confianza de la prueba de hipótesis y de su potencia. En este caso, se puede ver que la función cópula Frank, determina parcialmente bien cuando sus datos no provienen de la misma función cópula, esto a excepción de las funciones cópula FGM y la cópula Gumbel-Hougaard las cuales mostraron resultados en la potencia de la prueba de hipótesis muy bajos cuando la hipótesis nula es que los datos se distribuyen Frank con dependencia débil (hasta en tamaños de muestra $n = 1000$ para la función cópula FGM).

Por otro lado, la confianza tiene el mismo comportamiento visto en las anteriores funciones cópula, lo cual dice que en el 95 % de los casos, si la hipótesis nula es verdadera esta no va ser rechazada, lo cual es positivo en los casos donde se está seguro de la función cópula de los datos y tan solo se desea validar por medios estadísticos.

**Tabla 4**: Confianza y potencia de la prueba de hipótesis cuando $H_0 : (U, V) \sim Frank_\theta$, $H_a : (U, V) \not\sim Frank_\theta$

| $H_0$ | Dependencia | n | Potencia $(1-\beta)$ Clayton $\tau_\theta \in (-\infty, \infty)$ | FGM $\tau_\theta \in [-2/9, 2/9]$ | Gumbel-Barnett $\tau_\theta \in (-0.361, 0]$ | Gumbel-Hougaard $\tau_\theta \in [0, \infty)$ | Confianza $(1-\alpha)$ Frank $\tau_\theta \in (-\infty, \infty)$ |
|---|---|---|---|---|---|---|---|
| Frank | Débil $\tau_{\theta=\pm 1.86} \approx \pm 0.2$ | 30 | 0.03 | 0.03 | 0.72 | 0.07 | 0.96 |
| | | 50 | 0.05 | 0.05 | 0.95 | 0.15 | 0.96 |
| | | 100 | 0.08 | 0.05 | 1.00 | 0.25 | 0.96 |
| | | 500 | 0.87 | 0.05 | 1.00 | 0.86 | 0.95 |
| | | 1000 | 1.00 | 0.05 | 1.00 | 0.99 | 0.94 |
| | Moderado $\tau_{\theta=\pm 5.73} \approx \pm 0.5$ | 30 | 0.04 | | | 0.07 | 0.96 |
| | | 50 | 0.10 | | | 0.20 | 0.96 |
| | | 100 | 0.63 | | | 0.61 | 0.95 |
| | | 500 | 1.00 | | | 1.00 | 0.95 |
| | | 1000 | 1.00 | | | 1.00 | 0.95 |
| | Fuerte $\tau_{\theta=\pm 18.19} \approx \pm 0.8$ | 30 | 0.02 | | | 0.03 | 0.96 |
| | | 50 | 0.05 | | | 0.04 | 0.96 |
| | | 100 | 0.40 | | | 0.22 | 0.96 |
| | | 500 | 1.00 | | | 1.00 | 0.93 |
| | | 1000 | 1.00 | | | 1.00 | 0.93 |

### 4.3.4. Función cópula Gumbel-Barnett: confianza y potencia de la prueba de hipótesis.

En la Tabla 5 se encuentran los resultados de la prueba de hipótesis siendo simulada esta un total de $k = 1000$ para la función cópula Gumbel-Barnett para distintos niveles de concordancia. En la tabla, se ve que no tiene una comparativa con la función cópula Gumbel-Hougaard, esto es debido a que esta función no cumple con el rango de concordancia de la función cópula Gumbel-Barnett, pues la primera modela dependencias positivas y la segunda modela dependencias negativas. Al igual que con

la función cópula Frank, la potencia es baja ante la función cópula FGM (con un máximo registrado de 52 %) y ante la función cópula Frank (con un máximo registrado de 25 %) lo cual indica que la mayoría de las veces que los datos sean extraídos de una función cópula FGM, la prueba no rechazará correctamente.

Por otro lado su confianza al igual que las demás funciones cópula es oscilante entre los valores cercanos al verdadero valor esperado de la confianza de la prueba, el cual es determinado por el nivel de significancia $\alpha$ preestablecido.

**Tabla 5**: Confianza y potencia de la prueba de hipótesis cuando $H_0 : (U, V) \sim Gumbel - Barnett_\theta$, $H_a : (U, V) \not\sim Gumbel - Barnett_\theta$

| $H_0$ | Dependencia | n | Potencia $(1-\beta)$ Clayton $\tau_\theta \in (-\infty, \infty)$ | FGM $\tau_\theta \in [-2/9, 2/9]$ | Frank $\tau_\theta \in (-\infty, \infty)$ | Gumbel-Hougaard $\tau_\theta \in [0, \infty)$ | Confianza $(1-\alpha)$ Gumbel-Barnett $\tau_\theta \in (-0.361, 0]$ |
|---|---|---|---|---|---|---|---|
| Gumbel-Barnett | Débil $\tau_{\theta=0.2} \approx -0.09$ | 30 | 0.07 | 0.06 | 0.06 | | 0.96 |
| | | 50 | 0.12 | 0.07 | 0.09 | | 0.96 |
| | | 100 | 0.13 | 0.07 | 0.09 | | 0.96 |
| | | 500 | 0.55 | 0.11 | 0.11 | | 0.95 |
| | | 1000 | 0.94 | 0.22 | 0.23 | | 0.94 |
| | Moderado $\tau_{\theta=0.5} \approx -0.21$ | 30 | 0.13 | 0.07 | 0.07 | | 0.95 |
| | | 50 | 0.22 | 0.07 | 0.10 | | 0.95 |
| | | 100 | 0.35 | 0.07 | 0.10 | | 0.95 |
| | | 500 | 1.00 | 0.15 | 0.15 | | 0.96 |
| | | 1000 | 1.00 | 0.47 | 0.43 | | 0.96 |
| | Fuerte $\tau_{\theta=0.9} \approx -0.333$ | 30 | 0.21 | | 0.06 | | 0.95 |
| | | 50 | 0.29 | | 0.06 | | 0.95 |
| | | 100 | 0.61 | | 0.06 | | 0.95 |
| | | 500 | 1.00 | | 0.11 | | 0.95 |
| | | 1000 | 1.00 | | 0.27 | | 0.95 |

En la Tabla 5 se puede observar que la confianza (1-$\alpha$) se mantiene consistente a medida que el tamaño de muestra $n$ aumenta, pues este tiende al nivel de significancia $\alpha$. Por otro lado, la potencia (1-$\beta$) disminuye conforme al aumento en el tamaño de muestra o se mantiene oscilante entre valores puntuales, significa, se puede observar que la potencia es baja para las funciones cópula Frank y FGM, pues, la prueba a penas las distingue con tamaños de muestra mayores a 100, lo cual no es un buen indicativo para la prueba.
Con respecto la confianza, se ve un ligero decrecimiento con respecto al nivel de dependencia y al tamaño de la muestra, siendo este poco variable y muy cercana a la confianza real $1 - \alpha = 0.95$.

### 4.3.5. Función cópula Gumbel-Hougaard: confianza y potencia de la prueba de hipótesis.

En la Tabla 6 se encuentran los resultados de la potencia del indicador $I^2$ cuando este evalua la bondad de ajuste entre datos simulados y la función cópula. Se puede ver que ante datos simulados de una función cópula FGM y una función cópula Frank la potencia es baja en niveles de concordancia bajos (esto para la función cópula Gumbel-Hougaard (Tovar Cuevas et al., 2019)), en otras palabras hay una posibilidad del 34 % de cometer un error tipo II con la función cópula Frank en tamaños de muestra grandes y niveles de dependencia bajos, lo cual no es un buen indicativo en

la potencia cuando se desean mostras que los datos se distribuyen como una función cópula Gumbel-Hougaard.

Por otro lado, la confianza oscila un poco por encima del valor $1 - \alpha$ de confianza preestablecido para la prueba. Lo cual es visto en algunas de las funciones cópula anteriores, esto indica que la prueba no rechaza la hipótesis nula casi el 95 % de las veces que esta es evaluada.

**Tabla 6**: Confianza y potencia de la prueba de hipótesis cuando $H_0 : (U, V) \sim Gumbel - Hougaard_\theta$, $H_a : (U, V) \not\sim Gumbel - Hougaard_\theta$

| $H_0$ | Dependencia | n | Potencia $(1-\beta)$ Clayton $\tau_\theta \in (-\infty, \infty)$ | FGM $\tau_\theta \in (-2/9, 2/9)$ | Frank $\tau_\theta \in (-\infty, \infty)$ | Gumbel-Barnett $\tau_\theta \in (-0.361, 0]$ | Confianza $(1-\alpha)$ Gumbel-Hougaard $\tau_\theta \in [0, \infty)$ |
|---|---|---|---|---|---|---|---|
| Gumbel-Hougaard | Débil $\tau_{\theta=1.25} = 0.2$ | 30 | 0.05 | 0.02 | 0.02 | | 0.95 |
| | | 50 | 0.05 | 0.02 | 0.02 | | 0.95 |
| | | 100 | 0.14 | 0.02 | 0.03 | | 0.95 |
| | | 500 | 0.98 | 0.24 | 0.19 | | 0.94 |
| | | 1000 | 1.00 | 0.76 | 0.66 | | 0.94 |
| | Moderado $\tau_{\theta=2} = 0.5$ | 30 | 0.17 | | 0.08 | | 0.93 |
| | | 50 | 0.44 | | 0.10 | | 0.95 |
| | | 100 | 0.94 | | 0.24 | | 0.95 |
| | | 500 | 1.00 | | 1.00 | | 0.95 |
| | | 1000 | 1.00 | | 1.00 | | 0.96 |
| | Fuerte $\tau_{\theta=5} = 0.8$ | 30 | 0.07 | | 0.11 | | 0.95 |
| | | 50 | 0.22 | | 0.20 | | 0.95 |
| | | 100 | 0.77 | | 0.43 | | 0.95 |
| | | 500 | 1.00 | | 1.00 | | 0.95 |
| | | 1000 | 1.00 | | 1.00 | | 0.96 |

# 5. Conclusiones y recomendaciones

El estudio del indicador propuesto en Tovar Cuevas et al. (2019), y el coeficiente de concordancia desarrollado en Ledwina (2014) es un gran acercamiento a la teoría detrás de las funciones cópula, pues, resumen propiedad de los modelos que miden las dependencias entre variables aleatorias.

La prueba de hipótesis propuesta en este trabajo, culmina con la evaluación de los errores tipo, donde, en muchos casos son bastante explicitos a la hora de hablar de la buena o mala calidad de la prueba de bondad de ajuste basada en el indicador desarrollado en Tovar Cuevas et al. (2019).

## 5.1. Conclusiones

Las conclusiones más importantes del estudio son que para el primer objetivo, se encontró una posible función de distribución que se ajusta al comportamiento de la distribución de $I^2$, con una tasa **total** de aceptación por parte de la prueba de Anderson-Darling, sin embargo no fue análizada a fondo debido a que sus estimaciones eran variadas según la función cópula, el nivel de dependencia $\theta$ y el tamaño de muestra $n$ de $(U, V)$. Por lo cual, no se permite extrapolar para cualquier conjunto de funciones cópula. En este caso tan solo fueron estudiadas las cinco funciones cópula propuesta a investigar en Tovar Cuevas et al. (2019) (Clayton, FGM, Frank, Gumbel-Barnett, Gumbel-Hougaard), sin embargo, se reconoce que existen infinitas o un grupo muy grande de ellas por lo cual se expresa que con el conjunto visto no es posible concluir en general para todas las funciones cópula, pues, el estadístico de prueba de Anderson-Darling rechazó la hipótesis de modelar la distribución muestral del indicador bajo una distribución Gamma de dos parámetros y para la distribución Gamma no la rechazó pero sus estimaciones eran muy variables entre funciones cópulas un ejemplo a esto es la función cópula Clayton con dependencia moderada para la función de distribución Gamma con tres parámetros ($\zeta$, $\eta$, $\nu$ ) con diferentes valores para $n = 50$ (6.1, 0.1, 0.3) y $n = 1000$ (7.6, 0.2, 0.2) o frente a la función cópula Frank con el mismo nivel de dependencia (Tovar Cuevas et al., 2019) y $n = 1000$ ( 3.9, 0.4, 0.2).

El segundo objetivo dió indicios de que el indicador puede poseer comportamientos muy distintos entre una función cópula u otra, donde, a partir de lo mismo se puede presentar una mayor variabilidad en los datos y en las estimaciones de $I^2$. Pues en algunos de los casos de estudio, se encontró que la varianza se mínimizaba cuando el parámetro era grande (funciones cópula Clayton, Frank, FGM, Gumbel-Barnett) y en otros que la varianza crecía en dependencias moderadas (caso de la función cópula Gumbel-Hougaard), sin embargo, esto disminuía a medida que la muestra $n$ aumentaba. Además, se indentificó que el tamaño de la muestra pertenece posiblemente a un valor clave para definir el comportamiento de $I^2$ ya que este tenía un sesgo y un error cuadrático medio más pequeño a medida que este aumenta, esto también permite conocer la consistencia del parámetro, pues, es visible que tiende a un valor puntual a medida que el tamaño de la muestra aumenta, además de que es poco robusto pues, cualquier pertubación en la muestra genera grandes cambios en su

estimación, pues, en el Anexo A.3.1 se puede visualizar que se encuentran variaciones leves y elevadas a medida que la muestra tiene un nivel de contaminación entre $\{0.01, 0.05, 0.1, 0.15, 0.2\}$ generando un error de más del 50 % entre la estimación real y la estimación contaminada.

La prueba de hipótesis afirmó los resultados conseguidos en Tovar Cuevas et al. (2019), pues, en su análisis de confianza se determinó que el indicador en pocas ocasiones rechaza la prueba si la hipótesis nula es verdadera, esto también se puede ver con el segundo objetivo ya que en este se encontró que la estimación bajo $H_0$ tiene un sesgo el cual tiende a cero. Sin embargo, la prueba de hipótesis posee dificultades a la hora de determinar que los datos no pertenecen a una función cópula en particular, casos como la función cópula FGM frente a la función cópula Frank (apenas un 10 % de probabilidad de rechazo cuando la hipótesis nula es que los datos se distribuyen FGM cuando provienen de una función cópula Frank), demuestran que hay funciones cópula las cuales no son fácilmente distinguibles para una prueba basada en el indicador $I^2$. Esto a pesar de ser negativo, concuerda con otros estudios como el realizado por Huang & Prokhorov (2014) (con una potencia en dependencias bajas de las funciones cópula Normal, Clayton y Gumbel del 2.5 %, 4.2 % y 3.5 % respectivamente frente a la función cópula Frank) y los realizados por (Genest et al., 2009), donde se evicencia que la función cópula Frank se ajusta muy bien a las dependencias débiles, moderada y fuertes de naturaleza líneal. En otras palabras, si las funciones cópula conceptualmente se usan para medir las mismas estructuras de dependencia, entonces, la prueba de bondad de ajuste no será potente para distinguir entre ellas. Frente a la potencia general de la prueba de hipótesis es particularmente baja $1 - \beta < 80\,\%$ para muestras menores a $n = 500$ y niveles de dependencia bajos, esto también se puede deber al grupo de estudio o que en algunos casos las funciones cópula estudiadas modelan comportamientos similares en los datos.

## 5.2. Recomendaciones

En este estudio, surgen algunas recomendaciones para el desarrollo de futuros trabajos. A pesar de que la prueba es sencilla computacionalmente hablando, tiene un exigencia alta en procesamiento, por lo cual, es necesario que se pueda contar con herramientas de computo potentes para un análisis más detallado del indicador bajo más grupos de funciones cópula. Fuera de la simulación del indicador (Tovar Cuevas et al., 2019), si se desea más exactitud en el cálculo teórico del $\tau$ de Kendall ó de la expresión $E[C_\theta(u, v)]$, se necesita una grilla de valores más grande, en este estudio se trabajó con una grilla de valores con un tamaño máximo de 100.000.000 de datos, sin embargo, solo se puedieron estimar los tres primeros decimales de la $E[C_\theta(u, v)]$ de la función cópula Gumbel-Barnett (Portilla Yela & Tovar Cuevas, 2018). Por ende, se sugieren que ante la necesidad de resultados más exactos se necesitan herramientas de computo más potentes.

Por el lado de la prueba de hipótesis, en este trabajo se usaron métodos de aleatorización y remuestreo para calcular la potencia y confianza de la prueba de hipótesis

sugerida en la Sección 3.4. Es posibilidad para un trabajo futuro, comparar la metodología de prueba de hipótesis propuesto en este documento frente a un método de permutación en la misma muestra, en cuyo caso la permutación fuera mejor, se generarían resultados más rápidos y computacionalmente hablando, en el caso contrario, sería un buen ejercicio académico.

En cuanto a recomendaciones teóricas. Aunque el indicador propuesto por Tovar Cuevas et al. (2019) es una buena aproximación a una prueba de bondad de ajuste para funciones cópula, una posible desventaja se encuentra en el denominador, la cual posiblemente no calcule de forma óptima la varianza de la función cópula empírica (Ver Anexo A.2.2), donde, basadose en los resultados de Swanepoel & Allison (2013) se generó el estadístico de prueba de Ledwina (2014) el cual evalua la independencia en los cuadrantes ($L_n(u,v) = \sqrt{n}q_n(u,v)$), el cual posee una distribución normal con media 0 y varianza 1 cuando $u$ y $v$ son independientes (Ver Anexo A.2.1). Por lo cual, el estadísto de Tovar Cuevas et al. (2019) es un estadístico libre de distribución y con una función de distribución $\sqrt{\frac{\chi^2_n}{n}}$ si y solo si $u$ y $v$ son independientes, esto fue visible en los resultados del estudio cuando se midieron las dependencias débiles de las funciones cópula (Sección 4.1), pues, estos tendían a tener una media igual a uno ($E[I^2] = E\left[\frac{\chi^2_n}{n}\right] = \frac{n}{n} = 1$) por lo cual, el indicador si tendía a tener un comportamiento $\frac{\chi_n}{\sqrt{n}}$ cuando la colección de datos era casi independiente.

Una forma de generalizar el indicador, es tomar la expresión de la varianza de la cópula empírica desde el trabajo de Swanepoel & Allison (2013), la cual es una expresión bastante extensa y compleja en algunas funciones cópula, con lo cual, el indicador de bondad de ajuste para funciones cópula se pueden expresar como

$$I^2 = \sum_{i=1}^{n} \left( \frac{C_n(u_i, v_i) - C_\theta(u_i, v_i)}{\sqrt{V[C_n(u_i, v_i)]}} \right)^2 .$$

El cual, si es un estadístico de bondad de ajuste libre de distribución (para más detalle revisar (Swanepoel & Allison, 2013, p. 1738)).

# 6. Referencias

Bouye, E., Durlleman, V., Nikeghbali, A., Riboulet, G., Roncalli, T. (2000). *Copulas for finance.*

Casella, G., & Berger, R.L. (2001). *Statistical inference.* Cengage Learning.

Diaz, W., & Cuadras, C.M. (2022). An extension of the gumbel–barnett family of copulas. *Metrika*, *85*(7), 913–926, https://doi.org/doi.org/10.1007/s00184-022-00859-0

Durante, F., & Sempi, C. (2015). *Principles of copula theory.* CRC press.

Genest, C., Quessy, J.-F., Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, *33*(2), 337–366, https://doi.org/doi.org/10.1111/j.1467-9469.2006.00470.x

Genest, C., & Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'ihp probabilités et statistiques* (Vol. 44, pp. 1096–1127).

Genest, C., Rémillard, B., Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, *44*(2), 199-213, https://doi.org/https://doi.org/10.1016/j.insmatheco.2007.10.005

Gumbel, E.J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, *55*(292), 698–707, https://doi.org/doi.org/10.2307/2281591

Huang, W., & Prokhorov, A. (2014). A goodness-of-fit test for copulas. *Econometric Reviews*, *33*(7), 751–771, https://doi.org/doi.org/10.1080/07474938.2012.690692

Ledwina, T. (2014). *Dependence function for bivariate cdf's.*

Ledwina, T., & Wyłupek, G. (2014). Validation of positive quadrant dependence. *Insurance: Mathematics and Economics*, *56*, 38–47, https://doi.org/10.1016/j.insmatheco.2014.02.008

Llorens, L.L. (2017). *Teoría de cópulas: introducción y aplicaciones a solvencia ii.* Fundación MAPFRE.

Mai, J.-F., & Scherer, M. (2013). *Simulating copulas: Stochastic models, sampling algorithms and applications* (Vol. 81) (n.º 2).

Nelsen, R.B. (2006). *An introduction to copulas*. Springer.

Pappadà, R., Durante, F., Torelli, N. (2018). A graphical tool for copula selection based on tail dependence. *Classification,(big) data analysis and statistical learning* (pp. 211–218).

Portilla Yela, J., & Tovar Cuevas, J.R. (2018). Estimating the gumbel-barnett copula parameter of dependence. *Revista Colombiana de Estadística*, *41*(1), 53–73, https://doi.org/doi.org/10.15446/rce.v41n1.64900

Sheskin, D.J. (2003). *Handbook of parametric and nonparametric statistical procedures*. Chapman and hall/CRC.

Swanepoel, J., & Allison, J. (2013). Some new results on the empirical copula estimator with applications. *Statistics & Probability Letters*, *83*(7), 1731–1739, https://doi.org/https://doi.org/10.1016/j.spl.2013.03.027

Tovar Cuevas, J.R., Portilla Yela, J., Achcar, J.A. (2019). A method to select bivariate copula functions. *Revista Colombiana de Estadística*, *42*(1), 61 – 80, https://doi.org/doi.org/10.15446/rce.v42n1.71780

Tsanakas, A., & Zhu, R. (2022). Selecting bivariate copula models using image recognition. *ASTIN Bulletin: The Journal of the IAA*, *52*(3), 707–734, https://doi.org/doi.org/10.1017/asb.2022.12

Van Vliet, B. (2023). *Abe sklar's fonctions de repartition a n dimensions et leurs marges.*

Wilcox, R.R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed. ed.). Amsterdam ;: Academic Press.

Zhang, S., Zhou, Q.M., Lin, H. (2021). Goodness-of-fit test of copula functions for semi-parametric univariate time series models. *Statistical Papers*, *62*, 1697–1721, https://doi.org/10.1007/s00362-019-01153-4

# A. Anexos

## A.1. Demostraciones

### A.1.1. Aproximación de la esperanza de una función cópula

En la metodología, se expresó una aproximación para la esperanza de una función cópula $E[c_\theta(u,v)]$, esta basandose en la regla del trapecio. La regla del trapecio se expresa de la forma

$$\int_a^b f(x)dx = F(b) - F(a) \approx \frac{1}{2n}\left(f(x_1) + 2\sum_{i=2}^{n-1} f(x_i) + f(x_n)\right),$$

donde n es la cantidad de divisiones entre a y b. Entonces, su definición para dos dimensiones se puede visualizar de la siguiente forma:

Imagine un cuadrado en el plano (representado en la Figura 30), los cuales, cada pixel corresponde a un intervalo entre dos puntos arbitrarios en las variables independientes de la función cópula.

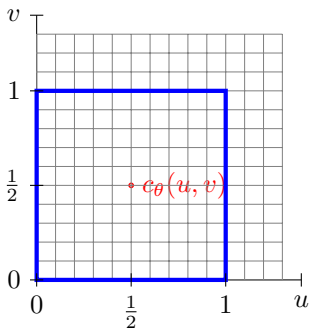


**Figura 30**: Representación del cuadrado unitaro. Fuente: Elaboración propia

En la Figura 30 se puede visualizar las margenes de los dominios de las variables de una función cópula bivariada, donde, este se representa en forma plana. Por otro lado, si se calcula la integral de la función en el espacio $[0,1]^2$ solo en el eje $u$. Se observa en la Figura 31 la marginalización del teorema del trapecio

$$\int_0^1 C_\theta(u,v)c_\theta(u,v)du \approx \frac{1}{2n}\left(C(u_1,v)c(u_1,v) + 2\sum_{i=2}^{n-1} C(u_i,v)c(u_i,v) + C(u_n,v)c(u_n,v)\right). \tag{17}$$

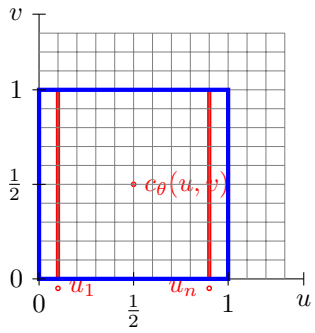


**Figura 31**: Teorema del trapecio en $\mathbb{R}^2$ con marginal en $u$. Fuente: Elaboración propia

Las líneas de color rojo en la Figura 31, se pueden tomar como la marginales evaluadas en $C(u_1, v)$ y $C(u_n, v)$, donde, v no varia. La integral en el intervalo $[u_1, u_n]$ con respecto a la altura (la cual corresponde a la función cópula $c_\theta(u, v)$) da como resultado la aproximación de la Ecuación 17.

Por otro lado, si se realiza el mismo proceso de marginalización con respecto a v se consigue el mismo resultado (visible en la Figura 32).

$$\int_0^1 C_\theta(u,v)c_\theta(u,v)dv \approx \frac{1}{2n}\left(C(u,v_1)c(u,v_1) + 2\sum_{j=2}^{n-1} C(u,v_j)c(u,v_j) + C(u,v_n)c(u,v_n)\right). \tag{18}$$

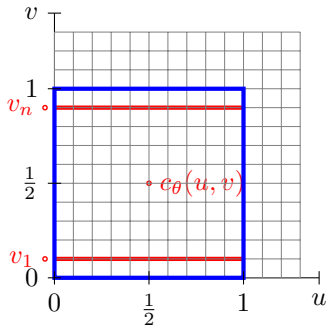


**Figura 32**: Teorema del trapecio en $\mathbb{R}^2$ con marginal en $v$. Fuente: Elaboración propia

El resultado del área de la Figura 32 sobre la línea del eje v con respecto a su altura (el valor que toma la función cópula cuando $u$ es fija y ve se mueve), es correspodiente a la Ecuación 18

Por último, la expresión en dos dimensiones que se encuentra dada en la metodología, corresponde al análisis de la Figura 33, la cual da como resultado la Ecuación 19, pues, los puntos que hacen referencia a la intersección de la recta; mientras que, los puntos contenidos en la línea simple, corresponde a los valores marginales de la integral, además, los puntos contenidos en el área roja de la Figura 33 son correspondientes a la doble sumatoria de la ecuación.

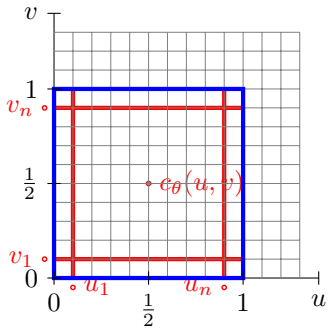


**Figura 33**: Teorema del trapecio en $\mathbb{R}^2$ en $u$ y $v$. Fuente: Elaboración propia

El volumen del área contenida entre los cuatro segmentos de recta(incluyendo a los mismos) visibles en la Figura 33, corresponde al resultado apróximado es

$$\begin{aligned} E[C_\theta(u,v)] \approx \tfrac{1}{4n^2}\,[&C_\theta(u_1,v_1)c_\theta(u_1,v_1) + C_\theta(u_n,v_1)c_\theta(u_n,v_1) + \\ &2\textstyle\sum_{i=2}^{n-1} C_\theta(u_i,v_1)c_\theta(u_i,v_1) + 2\sum_{i=2}^{n-1} C_\theta(u_i,v_n)c_\theta(u_i,v_n) + \\ &4\textstyle\sum_{j=2}^{n-1}\sum_{i=2}^{n-1} C_\theta(u_i,v_j)c_\theta(u_i,v_j) + \\ &2\textstyle\sum_{j=2}^{n-1} C_\theta(u_1,v_j)c_\theta(u_1,v_j) + 2\sum_{j=2}^{n-1} C_\theta(u_n,v_j)c_\theta(u_n,v_j) + \\ &C_\theta(u_1,v_n)c_\theta(u_1,v_n) + C_\theta(u_n,v_n)c_\theta(u_n,v_n)]\,, \end{aligned} \tag{19}$$

con la cual computacionalmente se pueden conseguir valores muy cercanos al valor real de la $E[c_\theta(u,v)]$ la cual es util para coeficientes como el $\tau$ de Kendall y el $\rho$ de Spearman.

Algunas condiciones especiales, es que a medida que la grilla o el ancho del píxel disminuye, los valores evaluados en el área azul se apróximan más a los valores obtenida del área roja.

## A.2. Propiedades y distribuciones asintóticas

### A.2.1. Coeficiente de concordancia de Ledwina

El estadístico de prueba del coeficiente de concordancia de Ledwina, expuesto en Ledwina (2014) posee una distribución asíntotica tal que $\sqrt{n}q_n(u,v) \overset{n\to\infty}{\to} N(0,1)$, esto debido a los resultados extraídos de la investigación realizada en Swanepoel & Allison (2013), donde, la varianza y la esperanza del estimador de la función cópula

$C_n(u,v)$, cuando $u$ y $v$ provienen de la función cópula de independencia se expresa en la Ecuación 20 y 21 respectivamente

$$E[C_n(u,v)] = uv \tag{20}$$

$$V[C_n(u,v)] = \frac{uv(1-u)(1-v)}{n} \tag{21}$$

La expresiones expuestas en las Ecuaciones 20 y 21 se cumplen para todo caso donde $(U,V) \sim \Pi$, sin embargo, la varianza solo cumple esa propiedad cuando $u_i = \frac{R_i}{n+1}$ y $v_i = \frac{S_i}{n+1}$. En otras palabras, cuando los rangos son divididos entre n+1 observaciones. Propiedad que cumple el estadístico de contraste propuesto en Ledwina (2014). Por lo cual, es prudente afirmar que, $\sqrt{n}$ veces la diferencia entre el estadístico de contraste y el estadístico esperado se distribuye también normal con media 0 y varianza 1, esto es visible a continuación

$$\begin{aligned}
\sqrt{n}(q_n(u,v) - q_\theta(u,v)) &= \sqrt{n}\left(\frac{C_n(u,v) - \Pi(u,v)}{\sqrt{uv(1-u)(1-v)}} - \frac{C_\theta(u,v) - \Pi(u,v)}{\sqrt{uv(1-u)(1-v)}}\right) \\
&= \sqrt{n}\left(\frac{C_n(u,v) - C_\theta(u,v)}{\sqrt{uv(1-u)(1-v)}}\right) \\
&= \sqrt{n}\frac{C_n(u,v) - \Pi(u,v)}{\sqrt{uv(1-u)(1-v)}}; \;\; C_\theta(u,v) \to \Pi(u,v) \\
&\Rightarrow \sqrt{n}q_n(u,v) = \sqrt{n}(q_n(u,v) - q_\theta(u,v)) \Leftrightarrow C_\theta(u,v) \to \Pi(u,v) \\
&\Rightarrow \sqrt{n}(q_n(u,v) - q_\theta(u,v)) \sim N(0,1).
\end{aligned}$$

Este resultado, será de utilidad en la siguiente Anexo.

### A.2.2. Indicador de bondad de ajuste (Tovar Cuevas et al., 2019)

Con el resultado de la Sección A.2.1, se puede determinar la distribución del estadístico propuesto por Tovar Cuevas, Portilla Yela, & Achcar (2019), con lo cual, si se sabe que el vector aleatorio es independiente, entonces, $\sqrt{n}$ veces la diferencia entre la estimación del coeficiente de Ledwina y el valor esperado tiene una distribución normal estandar. Por lo mismo, se puede concluir que el indicador propuesto en Tovar Cuevas et al. (2019), tiene una distribución $\chi^2$ si $u, v \sim \Pi(u,v)$, pues

$$nI^2 = n\sum_{i=1}^{n}\left(q_n(u_i,v_i) - q_\theta(u_i,v_i)\right)^2 = n\sum_{i=1}^{m}\left(N(0,1)\right)^2 \sim \chi^2_n.$$

Con lo cual, el indicador modificado $nI^2$ presentado en Tovar Cuevas et al. (2019) presenta una distribución **asintótica** $\chi^2_n$ cuando se desea medir la independencia entre los datos.

## A.3. Estimaciones de las funciones candidatas para la distribución del indicador $I^2$

Las estimaciones para cada una de las distribuciones Gamma de dos parámetros (Ecuación 22) y Gamma de tres parámetros (Ecuación 23) se encuentra a continuación (La estimación de la Gamma de tres parámetros se realizó por máxima verosimilitud).

$$\boldsymbol{\lambda} = (\zeta, \nu), \qquad \zeta = \frac{\bar{y}^2}{S_y^2}, \qquad \nu = \frac{\bar{y}}{S_y^2}, \tag{22}$$

$$\boldsymbol{\lambda} = (\zeta, \eta, \nu) = \{\boldsymbol{\lambda} : \underset{\boldsymbol{\lambda}\in\Lambda}{\text{máx}}\{\mathcal{L}(\boldsymbol{\lambda} \mid I_1^2, \ldots, I_n^2)\}\}. \tag{23}$$

En las ecuaciones anteriores $\zeta$ corresponde a un parámetro de forma en las distribuciones, $\nu$ a un parámetro de escala y $\eta$ corresponde a un parámetro de localización para las variables aleatorias.

En la Tabla 7 se puede observar la estimación de los parámetros del indicador $I^2$ visto en 4.1 por cada distribución candidata esto con la finalidad de visualizar un resumen de la estimación. Además, algunas conclusiones tempranas que se pueden extraer es que no existe forma de generalizar los parámetros entre funciones cópulas y dentro de la misma función cópula, pues, los resultados pueden variar dependiendo a los niveles de dependencia de la función cópula, el tamaño de la muestra o a la función cópula elegida.

**Tabla 7**: Valor estimado de los parámetros de las distribuciones candidatas

| | | | Gamma triparamétrica | | | Gamma biparamétrica | |
|---|---|---|---|---|---|---|---|
| | | | $\zeta$ | $\eta$ | $\nu$ | $\zeta$ | $\nu$ |
| Clayton | Débil | 30 | 2.8 | 0.4 | 0.4 | 5.0 | 3.3 |
| | | 50 | 3.3 | 0.4 | 0.3 | 5.8 | 3.9 |
| | | 100 | 3.5 | 0.4 | 0.3 | 6.3 | 4.3 |
| | | 500 | 3.3 | 0.5 | 0.3 | 7.4 | 5.5 |
| | | 1000 | 3.7 | 0.5 | 0.2 | 7.8 | 6.0 |
| | Moderada | 30 | 8.8 | -0.2 | 0.2 | 7.4 | 3.9 |
| | | 50 | 6.1 | 0.1 | 0.3 | 6.8 | 3.5 |
| | | 100 | 6.2 | 0.2 | 0.3 | 7.4 | 3.8 |
| | | 500 | 6.3 | 0.3 | 0.3 | 8.7 | 4.6 |
| | | 1000 | 7.6 | 0.2 | 0.2 | 9.3 | 4.9 |
| | Fuerte | 30 | 5.2 | -0.2 | 0.4 | 4.6 | 2.4 |
| | | 50 | 4.6 | 0.0 | 0.4 | 4.9 | 2.4 |
| | | 100 | 5.3 | 0.0 | 0.4 | 5.4 | 2.5 |
| | | 500 | 4.3 | 0.2 | 0.4 | 5.3 | 2.5 |
| | | 1000 | 3.6 | 0.3 | 0.5 | 5.1 | 2.4 |
| FGM | Débil | 30 | 3.5 | 0.2 | 0.3 | 5.0 | 3.5 |
| | | 50 | 3.1 | 0.3 | 0.3 | 5.2 | 3.7 |
| | | 100 | 2.8 | 0.4 | 0.3 | 5.5 | 4.0 |
| | | 500 | 2.9 | 0.4 | 0.3 | 6.3 | 5.4 |
| | | 1000 | 3.7 | 0.3 | 0.2 | 6.3 | 5.7 |
| | Moderada | 30 | 3.1 | 0.3 | 0.4 | 5.1 | 3.5 |
| | | 50 | 3.2 | 0.3 | 0.3 | 5.2 | 3.8 |
| | | 100 | 2.9 | 0.4 | 0.3 | 5.6 | 4.1 |
| | | 500 | 3.1 | 0.4 | 0.3 | 6.4 | 5.5 |
| | | 1000 | 3.5 | 0.3 | 0.2 | 6.2 | 5.7 |
| | Fuerte | 30 | 3.6 | 0.2 | 0.3 | 4.9 | 3.4 |
| | | 50 | 3.3 | 0.3 | 0.3 | 5.4 | 3.8 |
| | | 100 | 2.8 | 0.4 | 0.3 | 5.4 | 4.0 |
| | | 500 | 3.3 | 0.4 | 0.2 | 6.7 | 5.8 |
| | | 1000 | 4.1 | 0.3 | 0.2 | 6.7 | 6.2 |
| Frank | Débil | 30 | 3.4 | 0.3 | 0.3 | 5.0 | 3.5 |
| | | 50 | 3.3 | 0.3 | 0.3 | 5.5 | 3.9 |
| | | 100 | 2.8 | 0.4 | 0.4 | 5.2 | 3.8 |
| | | 500 | 3.0 | 0.4 | 0.3 | 6.6 | 5.5 |
| | | 1000 | 3.7 | 0.3 | 0.2 | 6.5 | 5.8 |
| | Moderada | 30 | 4.6 | 0.1 | 0.3 | 5.0 | 3.8 |
| | | 50 | 3.6 | 0.3 | 0.3 | 5.7 | 4.1 |
| | | 100 | 3.5 | 0.4 | 0.3 | 6.1 | 4.3 |
| | | 500 | 3.6 | 0.4 | 0.2 | 7.5 | 5.8 |
| | | 1000 | 3.9 | 0.4 | 0.2 | 8.1 | 6.7 |
| | Fuerte | 30 | 3.4 | 0.1 | 0.3 | 4.1 | 4.0 |
| | | 50 | 3.3 | 0.2 | 0.3 | 5.0 | 4.5 |
| | | 100 | 3.4 | 0.3 | 0.3 | 6.1 | 4.7 |
| | | 500 | 3.3 | 0.4 | 0.3 | 7.0 | 5.1 |
| | | 1000 | 3.1 | 0.5 | 0.3 | 7.2 | 5.5 |
| Gumbel-Barnett | Débil | 30 | 3.3 | 0.3 | 0.3 | 5.0 | 3.5 |
| | | 50 | 3.5 | 0.3 | 0.3 | 5.3 | 3.9 |
| | | 100 | 2.7 | 0.4 | 0.3 | 5.5 | 4.3 |
| | | 500 | 2.6 | 0.4 | 0.3 | 6.6 | 5.8 |
| | | 1000 | 3.3 | 0.4 | 0.2 | 6.5 | 6.0 |
| | Moderada | 30 | 4.1 | 0.2 | 0.3 | 5.5 | 4.2 |
| | | 50 | 4.0 | 0.2 | 0.3 | 5.7 | 4.4 |
| | | 100 | 2.9 | 0.4 | 0.3 | 5.4 | 4.2 |
| | | 500 | 2.8 | 0.5 | 0.3 | 6.9 | 5.9 |
| | | 1000 | 3.1 | 0.4 | 0.2 | 6.9 | 6.2 |
| | Fuerte | 30 | 3.8 | 0.3 | 0.3 | 5.7 | 4.5 |
| | | 50 | 3.6 | 0.3 | 0.3 | 5.9 | 4.7 |
| | | 100 | 2.8 | 0.4 | 0.3 | 5.7 | 4.5 |
| | | 500 | 3.7 | 0.4 | 0.2 | 7.5 | 6.2 |
| | | 1000 | 3.3 | 0.4 | 0.2 | 7.5 | 6.6 |
| Gumbel-Hougaard | Débil | 30 | 4.1 | 0.2 | 0.3 | 5.5 | 4.3 |
| | | 50 | 4.0 | 0.2 | 0.3 | 5.5 | 4.0 |
| | | 100 | 3.3 | 0.3 | 0.3 | 5.4 | 4.0 |
| | | 500 | 3.6 | 0.4 | 0.3 | 6.5 | 4.5 |
| | | 1000 | 3.1 | 0.4 | 0.3 | 5.7 | 3.9 |
| | Moderada | 30 | 3.9 | 0.2 | 0.2 | 5.2 | 5.2 |
| | | 50 | 3.3 | 0.3 | 0.2 | 5.7 | 5.3 |
| | | 100 | 3.9 | 0.4 | 0.2 | 7.7 | 6.7 |
| | | 500 | 5.5 | 0.3 | 0.2 | 9.4 | 7.1 |
| | | 1000 | 4.5 | 0.4 | 0.2 | 9.2 | 6.6 |
| | Fuerte | 30 | 2.8 | 0.1 | 0.3 | 3.3 | 3.4 |
| | | 50 | 3.1 | 0.1 | 0.3 | 4.0 | 4.1 |
| | | 100 | 4.1 | 0.2 | 0.2 | 6.0 | 6.1 |
| | | 500 | 4.4 | 0.4 | 0.2 | 9.5 | 9.1 |
| | | 1000 | 5.9 | 0.3 | 0.1 | 11.8 | 10.8 |

### A.3.1. Robustez del indicador $I^2$

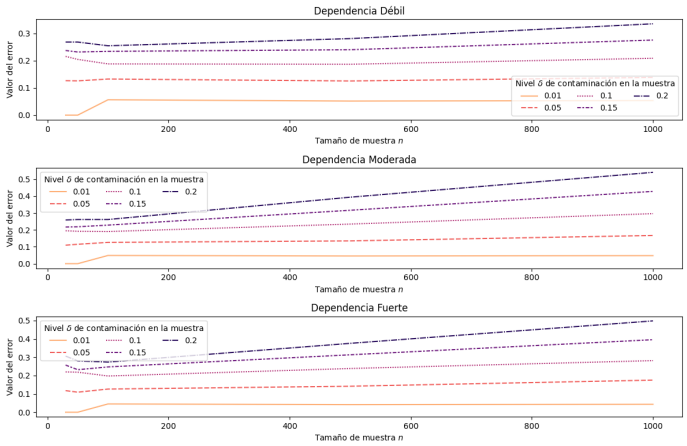


**Figura 34**: Diferencia promedio entre indicadores contaminados con una función cópula Gumble-Hougaard para la función cópula Clayton, con niveles de contaminación $\delta = 0.01, 0.05, 0.1, 0.15, 0.2$. Fuente: Elaboración propia

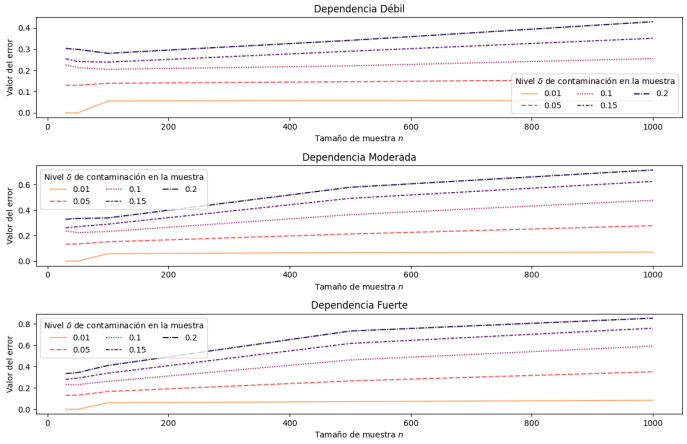


**Figura 35**: Diferencia promedio entre indicadores contaminados con una función cópula Frank para la función cópula FGM, con niveles de contaminación $\delta = 0.01, 0.05, 0.1, 0.15, 0.2$. Fuente: Elaboración propia

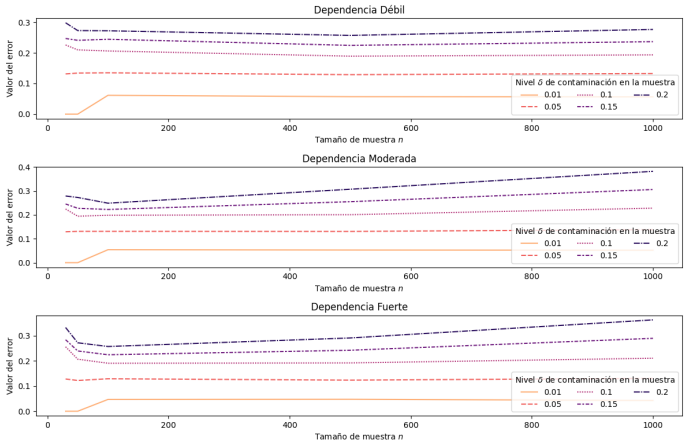


**Figura 36**: Diferencia promedio entre indicadores contaminados con una función cópula Gumbel-Hougaard para la función cópula Frank, con niveles de contaminación $\delta = 0.01, 0.05, 0.1, 0.15, 0.2$. Fuente: Elaboración propia

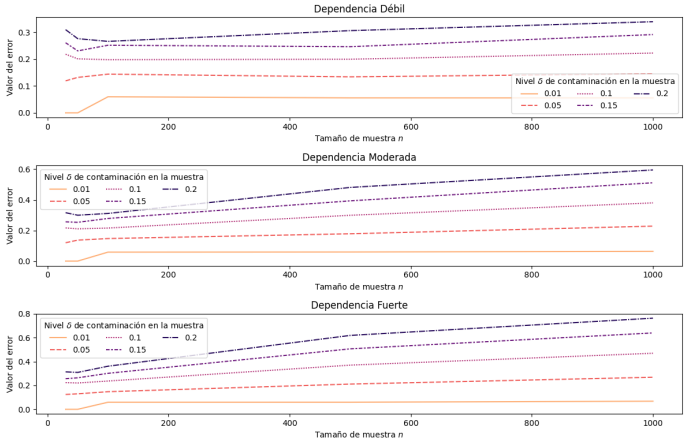


**Figura 37**: Diferencia promedio entre indicadores contaminados con una función cópula Frank para la función cópula Gumbel-Barnett, con niveles de contaminación $\delta = 0.01, 0.05, 0.1, 0.15, 0.2$. Fuente: Elaboración propia

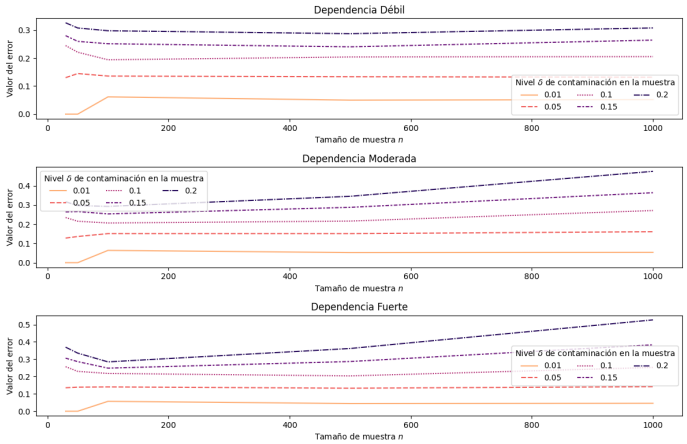


**Figura 38**: Diferencia promedio entre indicadores contaminados con una función cópula Gumbel-Hougaard para la función cópula Clayton, con niveles de contaminación $\delta = 0.01, 0.05, 0.1, 0.15, 0.2$. Fuente: Elaboración propia

## A.4. Pseudo código

En este Anexo, se encuentra almacenada la información del pseudo código usado para generar las muestras del indicador $I^2$ y su algoritmo para la extracción de la prueba a partir de una muestra dada.

### A.4.1. Generación de la prueba y réplicas de $I^2$ a partir de la muestra.

1. Inicie la muestra de datos bivariados $(X, Y)$ y tome $n$ como el tamaño de la muestra.
2. Inicie la función cópula que desea evaluar $C_\theta$.
3. Inicie la cantidad de réplicas del indicador bajo $H_0$ y almacenelo en $m$.
4. Transforme el vector aleatorio de $\mathbb{R}^2 \to [0,1]^2$ haciendo uso del método de la transformada inversa (si conoce la distribución de las marginales) ó haciendo uso de los rangos ($R$ para $X$ y $S$ para $Y$) de los datos, donde $U = \frac{R}{n+1}$ y $V = \frac{S}{n+1}$.
5. Calcule el $\tau$ de Kendall empírico, $\tau_n = \frac{2}{n(n-1)} \sum_{i<j}^{n} sgn(x_i - x_j) sgn(y_i - y_j)$
6. Haciendo uso de los métodos de estimación para funciones cópula, use el $\tau$ de Kendall inverso ($\theta_0$ tal que $\tau_n - 4E[C(u,v)] + 1 = 0$) y en caso de desear más precisión use el valor obtenido como semilla para realizar el método de máxima-veromilitud de la expresión $\hat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}(\boldsymbol{\theta_0} \mid U, V)$ y almacenelo en $\hat{\theta}$.
7. Calcule la función de distribución empírica de los datos $Cn(u,v)$.
8. Calcule los datos evaluados en la función cópula $C_{\hat{\theta}}(u,v)$.
9. Calcule la independencia de los datos $\Pi(u,v) = u \times v$.
10. Calcule el coeficiente de concordancia empírico de Ledwina $q_n(u,v) = \frac{C_n(u,v) - \Pi(u,v)}{\sqrt{\Pi(u,v) \times (1-u) \times (1-v)}}$.

11. Calcule el coeficiente de concordancia teórico de Ledwina $q_{\hat{\theta}}(u,v) = \frac{C_{\hat{\theta}}(u,v)-\Pi(u,v)}{\sqrt{\Pi(u,v)\times(1-u)\times(1-v)}}$.
12. Calcule el estadístico de contraste a partir de la muestra $I_c^2 = \sum_{i=1}^{n}\left(q_n(u*,v*) - q_{\hat{\theta}}(u*,v*)\right)^2$.
13. Genere un vector vacio $\boldsymbol{I}^2$
14. Para i en 1 hasta $m$:
    - $U*, V* \sim C_{\hat{\theta}}$ de longitud $n$.
    - $Cn(u*,v*)$.
    - $\Pi(u*,v*) = u* \times v*$.
    - $q_n(u*,v*) = \frac{C_n(u*,v*)-\Pi(u*,v*)}{\sqrt{\Pi(u*,v*)\times(1-u*)\times(1-v*)}}$.
    - $q_{\hat{\theta}}(u*,v*) = \frac{C_{\hat{\theta}}(u*,v*)-\Pi(u*,v*)}{\sqrt{\Pi(u*,v*)\times(1-u*)\times(1-v*)}}$.
    - $I^2 = \sum_{i=1}^{n}\left(q_n(u*,v*) - q_{\hat{\theta}}(u*,v*)\right)^2$.
    - Añada a $\boldsymbol{I}^2$ el valor de $I^2$
15. Calcule el valor-p empírico con $\frac{1}{m}\sum_{i=1}^{m}\mathbb{1}\{\boldsymbol{I}_i^2 \geq I_c^2\}$ y regrese este valor e $I_c^2$.

La documentación del trabajo en terminos de código se encuentra en Github en donde se pueden encontrar funciones hechas a medida para el trabajo, además, de la forma en la que se calcularon medidas como la función cópula empírica a través de operaciones matriciales.