

Análisis del supuesto de independencia en el aseguramiento en salud

¿Son la frecuencia y la severidad independientes?

Santiago Mosquera Daza

**Trabajo de Grado para optar por el título de
Magíster en Economía**

**Director del trabajo de Grado:
Norman Danilo Maldonado Vargas, PhD.**

**Universidad Icesi
Facultad de Ciencias Administrativas y Económicas
Santiago de Cali, Junio de 2020**

Resumen

El proceso de tarificación para seguros en salud puede ser abordado desde diferentes modelos. Uno de ellos es el Modelo de Frecuencia-Severidad (FSM), que pese a plantearse desde diferentes enfoques, su practicidad y facilidad técnica hicieron de uno recurrente a un costo: El supuesto de independencia entre la Frecuencia y la Severidad. No obstante, hay dos principales razones para invalidar dicho supuesto bajo este tipo de aseguramiento: Los factores que inciden en dichos componentes y el carácter injustificable en el ejercicio del aseguramiento en sí. Por estas razones, este trabajo se embarca en el contraste de dicha independencia, en el que los resultados no sólo la inhabilitaron, sino que también mostraron parte de la naturaleza de dicha relación en el contexto de un *Plan Voluntario en Salud* de una EPS.

Palabras Claves: *Aseguramiento en salud, Frecuencia, Severidad, Independencia, Cópula.*

Abstract

The rate-making process in health insurance can be addressed from different models. One of them is the Frequency-Severity Model (FSM) whose standard version is widely used and suggested in insurance pricing textbooks, this because of its practicability and usability thanks to the independence assumption between the pure premium components: Frequency and Severity. Nevertheless, no free lunch since two reasons invalidate this assumption: the interlinked factors that affect the aforementioned elements and the health insurance nature which makes it untenable. In this document, the independence hypothesis is rejected in the Private Health Insurance (PHI) context, as well as a description of the dependence structure is given.

Key words: *Health insurance, Frequency, Severity, Independence, Copula*

Especialmente agradecimiento al equipo PROESA por el acompañamiento durante el programa.

Santiago M.D.

Tabla de contenido

1	Introducción	3
2	Marco Conceptual	6
3	Métodos	13
3.1	Datos	13
3.2	Metodología	16
3.2.1	Asociación lineal	17
3.2.2	Asociación de Rango	18
3.2.3	Cópula	19
4	Resultados	23
4.1	Asociación lineal y Asociación de rango	23
4.2	Cópula	26
4.3	Discusión	31
5	Conclusiones	34
Apéndice A	Vocabulario	35
Apéndice B	Cópulas y dependencia	37
Apéndice C	Cópulas estimadas	38
Apéndice D	Propiedades deseables de una medida de asociación	40

1 Introducción

Un sistema de salud es un «conjunto de relaciones estructuradas» entre poblaciones e instituciones como consecuencia de la interacción entre las necesidades sociales (públicas y privadas) en salud y la «respuesta social organizada» institucional. Las funciones elementales de este sistema son el financiamiento, la prestación de servicios en salud, la modulación y la articulación. El buen funcionamiento de estos elementos y su equilibrio sistémico se conocen como *Pluralismo Estructural*, un modelo bajo el cual los sistemas de salud latinoamericanos tomaron como referente a través de distintos medios de convergencia institucional (Londoño & Frenk, 1997).

Para el caso colombiano, dicha convergencia significó la transición de una administración centralizada de los fondos de seguros en salud a un mercado con competencia regulada, cuyos instrumentos fueron, principalmente, la Ley 100 de 1993 y la Ley 60 de 1993 (Castaño & Zambrano, 2006; Rodríguez-Monguió & Infante, 2004). Este modelo buscó la financiación para una cobertura universal por medio de: una seguridad social ampliada; la prestación pluralista de diferente naturaleza; la existencia de una entidad que, alejándose de la prestación directa, se encargara de la gestión del sistema (labor conocida como modulación); y, por último, la administración del pago de la cobertura mencionada a través de Organizaciones para la Articulación de Servicios de Salud (OASS) (Londoño & Frenk, 1997).

Como resultado, se consolidó el SGSSS (Sistema General de Seguridad Social en Salud), compuesto por: la población colombiana, cuya afiliación obligatoria es una de las labores de las EPS (Empresas Promotoras en Salud) que, junto a la red de IPS (Instituciones Prestadoras de Salud), ofrecen un paquete de servicios conocido como el Plan de Beneficios en Salud (PBS), el cual está sujeto a los regímenes de aseguramiento disponibles: el régimen contributivo (RC), que cubre a los individuos con capacidad de pago con vínculo laboral o pensional; y el régimen subsidiado (RS), que ampara a quienes no cuentan con dicha capacidad y quedan por fuera del RC. La operación del primer régimen está sujeta a las cotizaciones de los afiliados, mientras que la del segundo régimen se basa en una naturaleza mixta de contribuciones cruzadas de los afiliados del RC y fuentes fiscales (Guerrero, Gallego, Becerril-Montekio & Vázquez, 2011).

Las fuentes de financiamiento del SGSSS son: las contribuciones obligatorias de los afiliados, los impuestos generales, las contribuciones patronales, los fondos de petróleo y SOAT (Seguro Obligatorio de Accidentes de Tránsito), las cuotas de recuperación del régimen subsidiado, y los copagos y cuotas moderadoras (Guerrero y col., 2011). La administración de dichos fondos es lo que posibilita el «Proceso de Financiamiento-Prestación» (Londoño & Frenk, 1997) y está a cargo de las EPS, quienes incurren en la prevención y gestión del riesgo (Paredes Cubillos, 2015), entre otras funciones encontradas en el artículo 14 de la Ley 1122 de 2007. La provisión de los fondos destinados a estas instituciones es transferida desde el Gobierno Nacional bajo el UPC (Unidad de Pago por Capitación), el cual es un giro de recursos por cada uno de los afiliados registrados y validados (Congreso de la República de Colombia, 2011) con la cual se busca la financiación del PBS.

Las EPS que operan bajo RC pueden ofrecer, además del PBS, un paquete de servicios

conocido como Plan Voluntario en Salud (PVS), el cual es un seguro financiado por los agentes que demandan coberturas asistenciales paralelas a lo ofrecido por el PBS y cuya definición se encuentra en el artículo 37 de la Ley 1438 de 2011 (Congreso de la República de Colombia, 2011). Una particularidad de este tipo de aseguramiento es su fuente de financiación que no proviene de las cotizaciones obligatorias o subsidios mencionados previamente, sino de las tarifas o primas comerciales, las cuales son las cantidades que los individuos asegurados pagan a las EPS por el cubrimiento de los seguros (Werner y col., 2016), cuyos valores son reguladas por la Superintendencia Nacional de Salud.

El cálculo de dichas primas comerciales, conocido como tarificación, se hace sobre la prima pura (o costo del riesgo), la cual es el monto de compensación que paga la aseguradora por afiliado en exposición o expuesto, entendiéndose a exposición como «(...) *la unidad básica de riesgo que subyace en la prima del seguro (...)*» y que sigue tres criterios: Proporcionalidad directa a las pérdidas esperadas; practicidad; y con precedencia histórica (Werner y col., 2016). Para efectos del documento, la exposición es el tiempo de afiliación del individuo, pues guarda una relación directa con la cantidad de peticiones que hace el afiliado al asegurador por la indemnización estipulada en el contrato de aseguramiento por cada servicio de salud que el individuo consume, donde dichas solicitudes reciben el nombre de reclamaciones (Werner y col., 2016).

El proceso de tarificación en cuestión puede llevarse a través de diversos métodos, donde algunos se apoyan en un modelo base que asume el supuesto de independencia entre los elementos que componen a la prima pura, lo que conlleva al planteamiento del costo del riesgo en términos del producto de sus elementos. El primer componente es la frecuencia, que es la cantidad de reclamaciones por afiliado expuesto (Pitacco, 2014). El segundo elemento es la severidad, la cual es el costo promedio de las reclamaciones (Werner y col., 2016) calculada como el monto de reclamaciones dividido por la cantidad de reclamaciones.

Sin embargo, a pesar de que el supuesto de independencia puede ser un punto de partida en manuales de tarificación, incluyendo en salud (Pitacco, 2014), se ha asumido *a priori* (Shi & Guszczka, 2016), por lo que es considerado restrictivo e inválido en la literatura actuarial (Czado, Kastenmeier, Brechmann & Min, 2012; Garrido, Genest & Schulz, 2016; Krämer, Brechmann, Silvestrini & Czado, 2013), aunque represente una conveniencia para el cálculo (Garrido y col., 2016; Klugman, Panjer & Willmot, 2012; G. Y. Lee & Shi, 2019). Además, ignorar las dependencias entre la frecuencia y la severidad conlleva a subestimaciones en las estadísticas de las pérdidas o en su valor agregado (Cummins & Wiltbank, 1983; Krämer y col., 2013), puesto que la prima pura es el mayor componente dentro de la prima comercial (Shi, Feng & Ivantsova, 2015), lo que tiene como consecuencia la incorrecta gestión del riesgo y su efecto sobre la tarificación (Garrido y col., 2016).

En otras palabras, la subestimación del costo esperado, el cual es la base para la tarificación del plan comercial que ofrece la EPS, conlleva a la subvaloración de las tarifas, lo que se traduce en menores ingresos que, siendo insuficientes para cubrir los costos reales, ocasionan beneficios negativos. Además, un menor valor en las primas implica una subvaloración en la reserva de primas, la cual es una proporción de la prima aún no devengada por la aseguradora

en salud. De modo que el contraste de la hipótesis de independencia y su consecuente rechazo servirían como base para la aplicación de métodos de tarificación que contemple la asociación entre la frecuencia y la severidad.

De conformidad con lo anterior, el objetivo de esta investigación es el contraste del supuesto de independencia en aseguramiento privado en salud y el estudio de la estructura de la relación entre los elementos de la prima pura. El mencionado contraste y el posterior análisis de la estructura de la asociación se han llevado a cabo en diferentes tipos de aseguramiento, pero hasta el momento no se ha realizado en aseguramiento privado en salud pese a la relevancia del asunto. Además, se contextualiza en un sistema de salud donde el aseguramiento privado coexiste con el aseguramiento público y obligatorio como el caso Colombiano, lo que representa el inicio de investigaciones futuras para estimar el costo de asumir el supuesto y la sensibilidad de la estructura de la dependencia producto de la incorporación de variables dentro del modelamiento.

Con esto dicho, lo que resta del documento se organiza de la siguiente forma: la segunda sección contiene el Marco Conceptual, el cual cubre brevemente la relación del tema con la Teoría de la Ruina, además de estudiar los modelos de pérdida agregada, el aseguramiento, el supuesto y los estudios de la dependencia. La tercera sección describe la metodología, la cual ha sido dividida en tres partes de acuerdo al tipo de medida de asociación y el empleo de la cópula. Los resultados, al igual que la metodología, se dividen en los mencionados grupos, los cuales se muestran en la cuarta sección y cuya discusión se brinda en la siguiente sección. En la sexta sección se encuentran las conclusiones y, finalmente, en las tres últimas secciones se introducen los apéndices.

2 Marco Conceptual

Metodología de búsqueda

Considerando que el marco conceptual del presente documento lo componen la economía de la salud, la estadística y las ciencias actuariales, la construcción del estado del arte se basó en lo propuesto por Card (2012) cuyos motores de búsqueda fueron [Google](#), [ScienceDirect](#), [Taylor & Francis Online](#) y [ArXiv](#) bajo las siguientes sentencias:

- independence AND actuarial AND frequency AND severity
- health AND insurance AND independence AND frequency AND severity
- ratemaking AND health AND independence AND frequency AND severity
- Independence AND frequency AND severity
- (independence OR dependence) AND actuarial AND frequency AND severity

La literatura que no fue obtenida bajo las especificaciones anteriores fue transversal a las referencias consultadas. Esta incluye la literatura encontrada en la bibliografía de los documentos hallados (*Backward Research*), como también la que referenciaba el material consultado (*Forward Research*). La revisión se realizó bajo un enfoque de investigación narrativa que sirviera de base para el análisis a nivel secundario sobre el supuesto de independencia (Card, 2012). De manera análoga, la revisión excluyó las referencias que estuvieran por fuera del alcance del presente documento, como las que discutían los mejores ajustes de modelo, el mejor enfoque estadístico, entre otros. En aras de ilustrar esta afirmación, no se contemplaron trabajos como el de Gschlöbl y Czado (2007), que pese a mostrar que existe cierta dependencia entre la frecuencia y severidad, el objetivo del paper es modelar dicho par de variables bajo un enfoque bayesiano con la incorporación de dependencias espaciales. Una vez cubierto lo anterior, el análisis del marco conceptual inicia con la introducción del contexto, el cual comprende una breve revisión a la Teoría del Riesgo y el *Modelo de Riesgo Colectivo*. Después se procede al planteamiento del supuesto y finalmente se incorpora el conjunto de estudios directos e indirectos que lo han abordado.

Análisis

La evaluación del riesgo dentro de un portafolio constituye el cimiento para el ejercicio del aseguramiento (Klugman y col., 2012). El riesgo, entendido como la combinación de la reclamación y su probabilidad de ocurrencia (Duncan, 2011), puede ser definido como la incertidumbre proveniente de una pérdida financiera (Marcinko & Hetico, 2006) que, en materia de este documento, se deriva de la provisión de un servicio en salud a un individuo específico. De esta manera, el riesgo en salud se define como la «*consecuencia financiera de cualquier alteración permanente o temporal de las condiciones de salud del individuo, lo cual puede resultar en costos en salud o en un impacto negativo sobre el ingreso [del individuo] debido a la reducción en la capacidad de pérdida laboral, o ambos*» (Pitacco, 2014)¹.

El análisis de las pérdidas constituye un papel preponderante cuando de ellas se obtiene la distribución de probabilidad, lo que permite conformar una función de las pérdidas agregadas cuyo modelamiento es el objetivo central de la Teoría del Riesgo (*Ruin Theory* o *Risk Theory*) (Cummins & Wiltbank, 1983). Los componentes de dicha distribución son, a su vez, funciones de distribución de los elementos empleados para calcular las probabilidades de pérdidas y, de esta forma, garantizar una mejor toma de decisiones para las aseguradoras (Cummins & Wiltbank, 1983; Shi y col., 2015). El primer elemento es la frecuencia, definida como la cantidad de reclamaciones por expuesto, donde este último es el asegurado que está en exposición, entendiéndose ésta como la unidad básica de riesgo y tratada en el documento como el tiempo de afiliación del individuo. El número de expuestos es la suma de los afiliados ponderada por su exposición o tiempo de afiliación (Werner y col., 2016). El segundo elemento es la severidad, la cual es el costo promedio de las reclamaciones que resulta de dividir su costo total entre su cantidad (Werner y col., 2016).

Con esto dicho, abordar la literatura en torno al supuesto de independencia entre la frecuencia y la severidad requiere, no sólo un antecedente de los diferentes enfoques actuariales y estadísticos que lo hayan tratado de manera indirecta o no, sino un contexto en el cual tomar *a priori* dicho supuesto conlleve también a asumir sus implicaciones. Antes de proceder, entiéndase dependencia como la relación que existe entre dos características de las unidades estudiadas, la cual puede ser determinística, en la que conocer el nivel de una de las variables es suficiente para saber la magnitud de cambio en la otra; o estadística, donde el conocimiento de dicho nivel genera diferencias en la distribución del otro atributo (Wermuth & Cox, 2014). En lo que respecta al presente documento, la asociación entre la frecuencia y la severidad puede deberse a sucesos que contribuyan en los procesos de la frecuencia (factores que aumenten el número de reclamaciones por expuesto), como aquellos que se dan bajo un tipo específico de asegurado (controlando por sus características), y su consecuente severidad para cualquiera de los sucesos mencionados (Cummins & Wiltbank, 1983).

Ahora bien, el antecedente mencionado parte del *Modelo de Riesgo Colectivo* y sus derivaciones que, bajo un escenario de tarificación para el aseguramiento, conlleven a una toma de decisiones en los métodos empleados. La primera decisión sobre la cual se puede abordar

¹Para una revisión extensa del riesgo dentro del aseguramiento se sugiere remitirse a (Dionne & Harrington, 2014)

la literatura parte de la separabilidad de los componentes mencionados, es decir, modelar el costo de la pérdida o modelar por separado la frecuencia y la severidad. De no realizarse dicha división, la literatura explorada aborda el modelamiento de la suma de las reclamaciones a través de un proceso compuesto, como el Modelo Tweedie, en el que la frecuencia tiene una distribución Poisson y la cantidad agregada de las reclamaciones obedece a una distribución gamma, lo que permite modelar la prima pura como un todo (Jorgensen, 1987; Shi & Guszcz, 2016).

Por otro lado, si la decisión parte de la separación de los elementos de la prima pura, se introduce el Modelo de Dos Partes o Modelo de Frecuencia-Severidad (FSM), que en su versión estándar expresa a la prima pura como el producto de las esperanzas matemáticas de la frecuencia y la severidad (Pitacco, 2014), cuyo enfoque univariado es referenciado en este escrito como FSM-EUI, lo cual es el FSM bajo Enfoque Univariado de Independencia. Cabe decir que la separabilidad tiene beneficios en la práctica, es decir, situaciones donde el FSM ofrece ventajas sobre el Modelo Tweedie. La primer ventaja es el análisis de los factores que afectan a la prima pura por medio de distintos canales, es decir, variables que de manera individual impactan a la frecuencia o la severidad pueden ser identificadas y determinadas previo al modelamiento de la prima pura, lo cual no podría realizarse a través del Modelo Tweedie. La segunda ventaja es la estabilidad del FSM, es decir, menor varianza aleatoria, lo cual puede reducirse al separar las fuentes de la variabilidad que pueden provenir de una variable que sólo afecta a una de las dos partes de la prima pura pero que se considera en el modelo conjunto del costo de la pérdida, lo cual también puede conllevar a un sobreajuste. Por último, el Modelo Tweedie asume implícitamente que la frecuencia y la severidad se mueven en la misma dirección, lo cual es una supuesto al que se puede prescindir al modelar la prima pura por medio de la separabilidad de sus componentes (Goldburd, Khare & Tevet, 2008).

Del enfoque univariado mencionado se deriva el Modelo de Pérdida Agregada (MPA), el cual tiene tres elementos principales: El individuo expuesto, la frecuencia y la severidad. Siguiendo la notación de Pitacco (2014), el MPA a nivel de expuesto se plantea de la siguiente forma:

- Sea N la frecuencia o *claim frequency*, con $N \in \mathbb{Z}^* \mid \mathbb{Z}^* = \{0\} \cup \mathbb{Z}^+$
- Sea X_j el monto de la j -ésima reclamación del asegurado o el valor del gasto médico con $X_j \in \mathbb{R}^* \mid \mathbb{R}^* = \{0\} \cup \mathbb{R}^+$
- Sean Y_j el monto que la aseguradora en salud paga por la j -ésima reclamación del asegurado, también llamado *claim severity*. Nótese que según las condiciones de la póliza o plan de salud, $Y_j \leq X_j$.
- Sea S el monto agregado de las reclamaciones que satisface las propiedades de distribuciones compuestas entonces

$$S = \begin{cases} 0 & \text{si } N = 0 \\ \sum_{j \in N} Y_j & \text{si } N > 0 \end{cases}$$

El proceso de tarificación implica hallar la prima equivalente, para la cual ha de conocerse $\Pi = \mathbb{E}[S]$ que requiere de los siguientes supuestos:

1. Las variables aleatorias $\{X_j\}$ son independientes de N . Este es el supuesto discutido a lo largo del presente escrito.
2. $\{X_j\}$ es i.i.d y su valor esperado puede ser $\mathbb{E}[X_1]$.
3. Si $Y_j = \phi(X_j)$, donde $\phi(\cdot)$ es una forma funcional, entonces $\{Y_j\}$ es i.i.d y su valor esperado puede ser $\mathbb{E}[Y_1]$.

Si se procede con el Teorema de la Partición teniendo a n como cualquier posible valor que pueda tomar la variable aleatoria N (Farhadi, 2016; Grimmett & Welsh, 2014)², entonces

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}\left[\sum_{j \in N} Y_j\right] \\ \mathbb{E}[S] &= \sum_{n \in N} \mathbb{E}\left[\sum_{j \in N} Y_j \mid N = n\right] \mathbb{P}(N = n)\end{aligned}$$

De esta forma, al aplicar el supuesto de independencia, resulta la ecuación (1)

$$\mathbb{E}[S] = \sum_{n \in N} \mathbb{E}\left[\sum_{j \in N} Y_j\right] \mathbb{P}(N = n) \quad (1)$$

lo que permite seguir con el desarrollo para obtener la prima pura $\mathbb{E}[S]$:

Aplicando Supuesto 3

$$\mathbb{E}[S] = \sum_{n \in N} \mathbb{E}[nY_1] \mathbb{P}(N = n)$$

$$\mathbb{E}[S] = \mathbb{E}[Y_1] \left(\sum_{n \in N} n\mathbb{P}(N = n)\right)$$

Aplicando definición de expectativa

$$\mathbb{E}[S] = \mathbb{E}[Y_1] \mathbb{E}[N]$$

Por otro lado, si se decidiese por no asumir el supuesto de independencia, se contemplan dos grupos de enfoque: Condicional y Multivariado (Shi & Guszczka, 2016). El primero es un refinamiento al FSM estándar, el cual modela la severidad en términos de la frecuencia, por lo que no sólo deja de lado el supuesto, sino también se especifica bajo distribuciones condicionales, de aquí que sea referenciado como FSM-EUC (FSM bajo Enfoque Univariado

² $\mathbb{E}(A) = \sum_i \mathbb{E}(A|B_i)\mathbb{P}(B_i)$

Condicional) cuya formulación se observa en la ecuación (2). Entretanto, el segundo grupo de enfoque se diferencia de los anteriores porque basa sus métodos en el Análisis Multivariado, entre los cuales existe un refinamiento adicional del FSM bajo el enfoque multivariado (FSM-EM) (Shi & Guszczka, 2016).

$$f(Y, N) = f(N)f(Y | N) \quad (2)$$

Cabe decir que las diferentes aristas descritas son relevantes en la medida en que estas metodologías sirvan como medios para confirmar la existencia de una tradición en lo referente a mantener dicho supuesto. Es decir, existe una rigidez en la transición del EUI a cualquiera de las especificaciones EUC o EM, resistencia que puede provenir de criterios basados en la bondad del EUI, los cuales contemplan sus ventajas por encima de las otorgadas por los enfoques EUC o EM; como también de apreciaciones que resaltan los inconvenientes que tienen los modelos con EUC o EM. Estas apreciaciones pueden observarse en los estudios que sugieren que el supuesto es conveniente en la práctica (Garrido y col., 2016), puesto que los métodos EUC y EM, relajando el supuesto, conllevan a un incremento en la carga computacional (como las estimaciones por máxima verosimilitud) o a estimaciones de parámetros que son difíciles de obtener, como el modelamiento con cópulas (G. Y. Lee & Shi, 2019; W. Lee, Park & Ahn, 2018).

No obstante, hay diferentes percepciones sobre las implicaciones de asumir el supuesto por su conveniencia, dado que los beneficios provenientes de la separabilidad del Modelamiento de Dos Partes o Modelo de Frecuencia-Severidad implican implícitamente el supuesto de independencia (Garrido y col., 2016). Sin embargo, estos beneficios aluden a la posibilidad de analizar por separado los efectos de la frecuencia y la severidad sobre las pérdidas agregadas, lo cual no insinúa que las diferentes realizaciones aleatorias de los montos o valores de las reclamaciones no se relacionen con la cantidad de éstas (Goldburd y col., 2008), es decir, el supuesto de independencia (Pitacco, 2014).

En otros términos, el beneficio de la separabilidad alude estrictamente al FSM estándar o FSM-EUI, de la cual G. Y. Lee y Shi (2019) discutieron como una propiedad que facilita la inferencia estadística para los dos procesos, aunque esto no implica necesariamente que el FSM no pueda acudir a la incorporación de enfoques condicionales o modelamientos conjuntos, como se vio en la ecuación (2). Sin embargo, pese al beneficio de la practicidad, el supuesto ha sido asumido *a priori* dentro del modelo estándar de reclamaciones agregadas para el *Modelo de Riesgo Colectivo*, como se pudo apreciar en la ecuación (1) (Garrido y col., 2016; Shi y col., 2015).

Esta injustificación pone en relieve unos planteamientos restrictivos sobre el modelamiento de dicha agregación de pérdidas, puesto que se circunscribe al producto de las esperanzas matemáticas de la frecuencia y la severidad³ sin contemplar la posible relación que exista entre ellas y sus factores (Krämer y col., 2013). Como consecuencia, se obtienen estimaciones inexactas, lo que repercute en las pérdidas agregadas modeladas bajo el *Modelo de Riesgo*

³ $E[S] = \mathbb{E}[Y_1] \mathbb{E}[N]$

Colectivo , lo cual es un resultado que fue cubierto por Cummins y Wiltbank (1983), quienes advirtieron que ignorar las posibles dependencias entre la frecuencia y la severidad tendría como consecuencia la subestimación en las estadísticas de la pérdida, lo que a su vez impactaría negativamente sobre la toma de decisiones. Posteriormente fueron Krämer y col. (2013) los que evaluaron dicha observación para el mercado de aseguramiento de autos cuyos hallazgos mostraron que, asumiendo *a priori* el supuesto, la pérdida total del portafolio del seguro resulta subestimada, lo que además impacta de manera adversa sobre las políticas de precio.

Ahora bien, hasta el momento se ha analizado las consecuencias de dicho supuesto y su contexto, mas no se ha abordado el contraste de la independencia dentro de la literatura, la cual puede realizarse por medio de tres grupos, no necesariamente excluyentes entre sí: El primero comprende la literatura que emplea medidas de asociación. Los dos últimos grupos son propuestos por Shi y col. (2015), quienes dando razones para prescindir del supuesto, plantearon dos estrategias: el enfoque condicional, que evalúa la significancia del coeficiente de la frecuencia como covariable del modelo de la severidad ; y el enfoque multivariado que cuantifica el parámetro de asociación vía cópula.

En el primer grupo están Krämer y col. (2013), quienes encontraron que una asociación monótona negativa entre la frecuencia y la severidad implica una sobreestimación de las pérdidas agregadas. Asimismo, Kholifah, Lestari y Devila (2019), quienes obtuvieron distintos τ de Kendall para diferentes modelos de Cópulas, las cuales son funciones de distribución multivariadas compuestas por funciones de distribución marginales uniformes (Díaz Monroy & Morales Rivera, 2012; Nelsen, 2006), en donde la asociación resultante es positiva bajo el contexto de aseguramiento de autos. Bajo esta misma línea está el trabajo de Czado y col. (2012), quienes empleando una regresión basada en cópulas, encontraron una pequeña dependencia significativa y positiva entre la frecuencia y la severidad dentro del portafolio de seguros de auto.

Por otra parte, Garrido y col. (2016) introdujeron la frecuencia como una covariable en el modelo del valor de las reclamaciones, en el que hallaron un efecto de dependencia pequeño y una asociación observada negativa y débil entre la cantidad de reclamaciones y el valor de éstas para seguros contra colisión. Esta dependencia negativa difiere del estimado por G. Y. Lee y Shi (2019) para seguros de responsabilidad en automóviles, cuyos resultados arrojaron una asociación positiva que pone en relieve el impacto de la naturaleza de la póliza de aseguramiento sobre el signo de la asociación. Del mismo modo G. Y. Lee y Shi (2019), en aras de mejorar la predicción de modelos de calificación de seguros de autos, plantearon un modelo multivariado condicional del FSM para datos longitudinales en los cuales determinaron tres tipos de cópulas gaussianas que capturaron las dependencias temporales en la frecuencia y la severidad, como también la dependencia contemporánea entre los mencionados componentes. Asimismo Shi y col. (2015), quienes empleando sus enfoques condicional y multivariado, mostraron la existencia de dependencia en aseguramiento de autos, con lo que justificaron recurrir a modelos que incorporen la asociación entre los componentes trabajados en el *Modelo de Riesgo Colectivo* .

Por su parte, dentro del aseguramiento en salud, Edward W Frees, Gao y Rosenberg

(2011) llevaron a cabo modelos de predicción para la frecuencia y la severidad bajo dos ámbitos: hospitalario y ambulatorio; sus hallazgos mostraron la existencia de una dependencia medida a través del coeficiente del predictor de frecuencia dentro del modelamiento de gastos de las reclamaciones, el cual fue sólo significativo para el ámbito ambulatorio. Cabe resaltar que la revisión no consideró las investigaciones que han abordado las relaciones de dependencia de frecuencia y severidad en términos de diferentes categorías, de las cuales algunas aplicaciones en aseguramiento en salud son: Carriere (1997), quien realizó un contraste de independencia para la frecuencia y uno para la severidad comparando las reclamaciones de medicamentos y las de otra naturaleza; Zimmer y Trivedi (2006) quienes investigaron con cópulas la interdependencia entre el aseguramiento en salud y la demanda de cuidado en salud en parejas casadas; y Erhardt y Czado (2012), quienes desarrollaron un *Modelo de Riesgo Colectivo* que incorpora las dependencias en términos de distintos tipos de cubrimiento para Planes Voluntarios en Salud, a saber: ambulatorio, hospitalario y tratamientos dentales, entre otros.

Lo anterior permite deducir que los estudios en torno a la relación de la frecuencia y la severidad en aseguramiento en salud han estado condicionados a distintos factores, entre los cuales están: la naturaleza de la reclamación, su vínculo con la demanda bajo vínculos del reclamante y, finalmente, los ámbitos de la prestación en salud. Sin embargo, el contraste del supuesto, independientemente de los factores mencionados, no ha sido realizado bajo el contexto de aseguramiento privado en salud, pese a las razones que hacen del supuesto injustificable.

3 Métodos

El conjunto de procedimientos empleados en esta sección se codificó en el lenguaje de programación funcional R, del cual se tomaron y se modificaron algunas funciones de los siguientes paquetes (*packages*): `data.table`, `tidyverse`, `scales`, `ggpubr`, `boot`, `psych`, `latticeExtra`, `MASS`, `AER`, `fitdistrplus`, `actuar`, `ismev`, `VGAM`, `copula`, `VineCopula`, `VC2copula` y `grDevices`.

Asimismo, se contemplaron dos pares de frecuencia y severidad. El primer par tomó a la severidad como el monto total de las reclamaciones a nivel de individuo dentro del período de aseguramiento, que siendo conocido como *Claim Amount* en la literatura actuarial (Garrido y col., 2016; Pitacco, 2014), se referenció en el documento como Severidad 1. En el segundo par se definió a la severidad como el monto promedio por reclamación a nivel de individuo en el período de aseguramiento, llamado en la literatura actuarial como *Average Claim Amount* (Cummins, 1991; Krämer y col., 2013; Shi & Guszczka, 2016) y tratado en el estudio como Severidad 2. Para ambas parejas, la definición de frecuencia no cambió, con lo cual los dos pares fueron Frecuencia-Severidad 1 y Frecuencia-Severidad 2. Con esto mencionado, la presente sección inicia con la descripción de los datos y, posteriormente, se introduce la metodología.

3.1 Datos

La información utilizada fue provista por una de las EPS con las que PROESA tiene convenio para fines investigativos. Su formato de tipo relacional implicó el uso de la integración de R con SQL y sólo se abarcó el año 2018; en parte por ser el último año de información detallada a nivel de prestación de servicio en salud del individuo, como también por tratarse de datos de uno de los planes complementarios de la EPS cuya periodicidad de pago es anual.

En lo que concierne a las reclamaciones, se tomaron a las facturas como unidad de conteo para las reclamaciones agregadas. Esto porque la factura proviene de la interacción entre la EPS y la IPS, el cual es un vínculo que da cuenta de las prestaciones de servicios en salud que la IPS ofrece al afiliado cuyo cubrimiento le compete a la EPS. Dado que el análisis se da a nivel de afiliado para un sólo año, se contaron las facturas cuyo año de radicación correspondiera al año de prestación del servicio en salud. Para la evaluación, se contó con la información de 35,744 afiliados de quienes se tenían datos demográficos y administrativos. Sin embargo, por cuestiones del objetivo del presente documento, se requirieron únicamente la cantidad y el monto de las reclamaciones, como también el número de individuos afiliados y expuestos. La información mostró indicios de una asociación positiva entre los dos pares de frecuencia y severidad, como se puede apreciar en el cuadro 1 que también evidencia la dispersión de los datos, donde cada afiliado tuvo 7 reclamaciones en promedio, además de un costo medio por reclamación igual a \$174,848.1 y un costo promedio por afiliado de \$1,778,474.4.

Cuadro 1: Descripción de la Frecuencia y Severidad en el año 2,018

Medida	Severidad 1	Severidad 2	Frecuencia
Mínimo	\$0.0	\$0.0	0
Decil 1	\$0.0	\$0.0	0
Decil 2	\$21,998.2	\$16,851.0	1
Decil 3	\$111,641.7	\$40,878.6	2
Decil 4	\$267,849.9	\$63,693.4	3
Decil 5	\$485,588.0	\$91,044.6	5
Decil 6	\$797,836.9	\$119,319.3	7
Decil 7	\$1,248,452.8	\$152,566.5	9
Decil 8	\$2,014,710.6	\$202,435.8	12
Decil 9	\$3,913,251.7	\$319,188.6	17
Máximo	\$202,230,530.0	\$73,888,350.0	180
Promedio	\$1,778,474.4	\$174,848.1	7

Fuente: Elaboración propia

La asociación también puede detallarse en los gráficos *Marginal Plot* de las figuras 1 y 2. Este tipo de gráficos permiten obtener información de los datos desde tres frentes: El primero alude a la fortaleza de la relación que tienen las variables evaluadas que, como podrá notarse en la figura 1, se aprecia una asociación positiva entre la frecuencia y la severidad 1, mientras que dicha relación no es clara en la figura 2. El segundo frente que aporta este tipo de gráficos es la información sobre la dispersión de los datos que, al tomar como referencia el punto rojo que muestra el promedio de las variables, se espera por la varianza de este tipo de datos, además de la presencia de valores extremos y de datos atípicos. Por último, los histogramas que se ubican en el eje superior y el eje lateral derecho proveen información sobre la distribución de la frecuencia y la severidad, de las cuales se puede percibir la asimetría con sesgo positivo y el comportamiento leptocúrtico de la información para las variables tratadas.

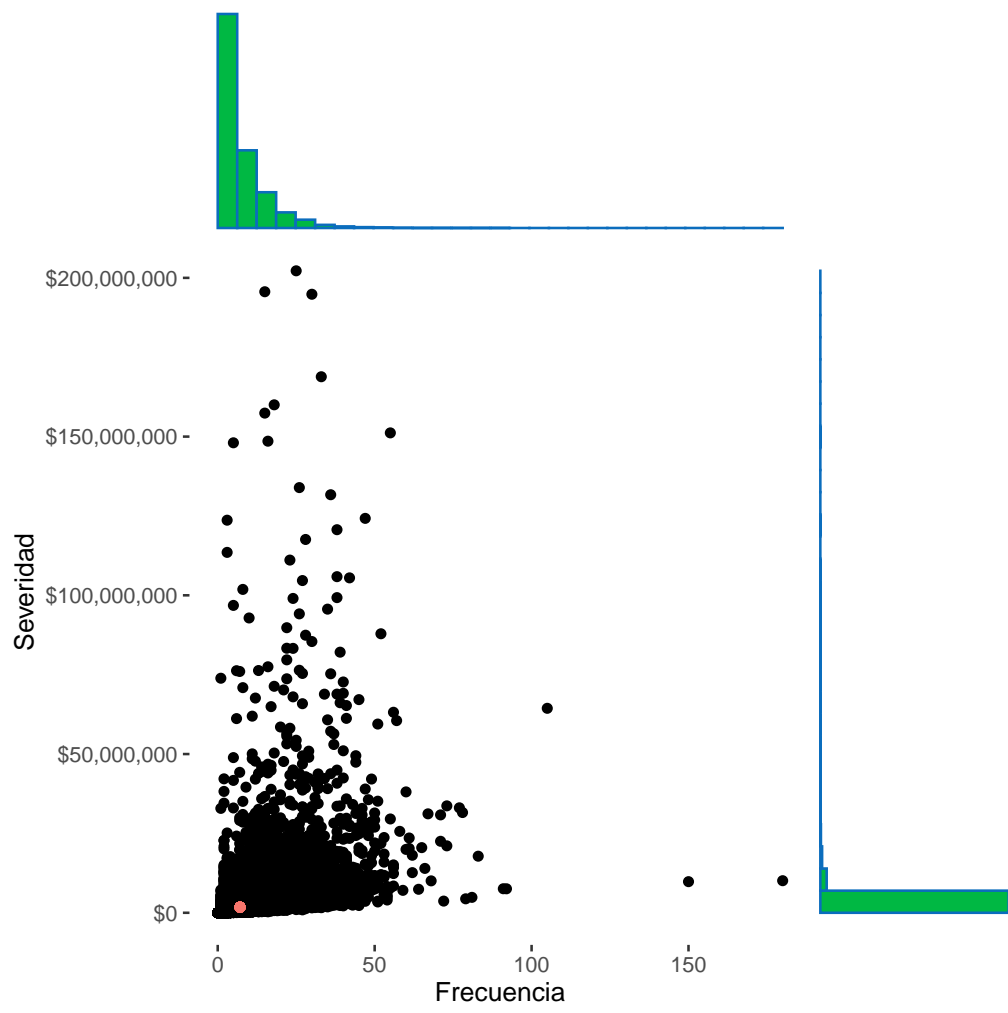


Figura 1: *Marginal Plot* de Frecuencia-Severidad 1

Fuente: Elaboración propia

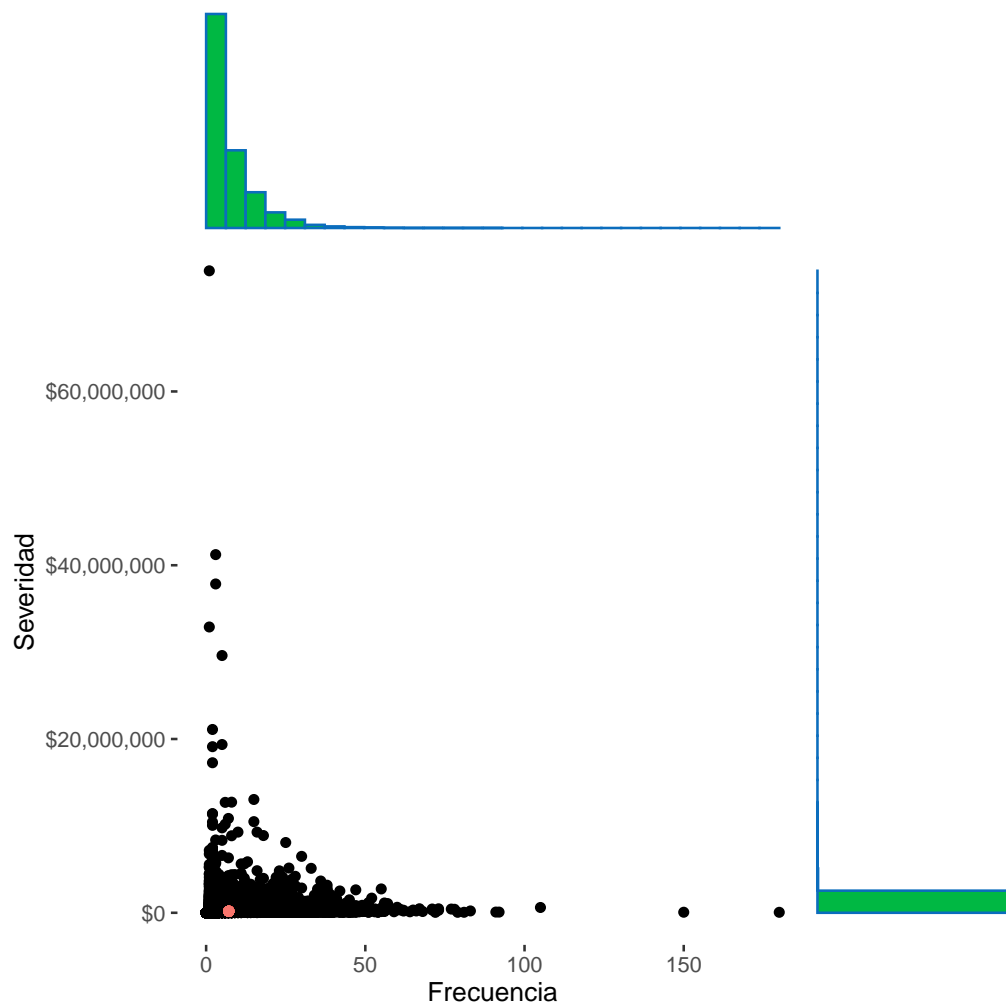


Figura 2: *Marginal Plot* de Frecuencia-Severidad 2

Fuente: Elaboración propia

3.2 Metodología

En la presente subsección se detalla el conjunto de procedimientos empleados que inició con la aplicación de medidas de dependencia para el contraste del supuesto. Para este fin se consideraron el coeficiente de correlación de Pearson y algunos coeficientes de asociación de rango. A partir de sus resultados, se procedió a la aplicación de Cópulas como método de soporte que también permitió el estudio de la estructura de la dependencia resultante entre ambas variables. En virtud de la prueba del supuesto y su consecuente análisis, cabe decir que la independencia conlleva a la nulidad de las medidas de asociación, razón por la que toda prueba consistió en negar el antecedente por medio del consecuente (*modus tollendo tollens*), es decir, encontrar una medida de dependencia distinta de cero. De ahí que la hipótesis nula

(H_o) se definiera como $H_o : \delta = 0$ con δ como medida de dependencia.

A continuación, se describe con mayor precisión lo mencionado bajo las siguientes fases: En una primera etapa, se discute la medida de dependencia de correlación lineal. Posteriormente se introducen las medidas de asociación de rango que, corrigiendo las inconvenientes de la correlación lineal, presentan a su vez un problema de inconsistencia que da entrada a la tercera fase, en la que se discuten las cópulas.

3.2.1 Asociación lineal

La medida de asociación lineal utilizada fue la correlación de Pearson (r_P), la cual también es llamada *Pearson Product-Moment Correlation* y resulta de la división de la covarianza ($\text{Cov}[\cdot]$) de la frecuencia (X) y la severidad (Y) entre la raíz del producto de sus respectivas varianzas ($\sqrt{\text{Var}[X]\text{Var}[Y]}$), como lo muestra la ecuación 3

$$r_P(X, Y) = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}} \quad (3)$$

El coeficiente resultante, r_P , cuyos valores oscilan entre -1 y 1, mide la independencia marginal bivariada entre variables con asociación estadística no lineal cuando $r_P \neq 0$ (Wermuth & Cox, 2014), aunque ésta es fuente de confusión pese a ser una «medida canónica» para la distribución normal multivariada, dado que las distribuciones de las variables empleadas en finanzas y aseguramiento no suelen pertenecer a dicha clase. Asimismo, la medida requiere una matriz de covarianza definida (observar numerador en (3)), lo que usualmente no se cumple en este tipo de aseguramiento puesto que los modelos de pérdidas pueden tener varianza infinita. Por otro lado, un valor nulo del *Pearson Product-Moment Correlation* no implica independencia, además de que la medida en cuestión no es invariante ante transformaciones crecientes no lineales en las variables⁴. Por último, la medida de correlación lineal sólo cumple con dos de las propiedades deseables de una medida de asociación discutidas en el apéndice D, las cuales son la simetría ($r_P(X, Y) = r_P(Y, X)$) y la normalización ($r_P(X, Y) \in [-1, 1]$) (Embrechts, McNeil & Straumann, 1998).

Sin embargo, aun con sus limitaciones y restricciones, el empleo del *Pearson Product-Moment Correlation* resulta conveniente en esta investigación puesto que el interés es contrastar la independencia y no la dependencia, es decir, es suficiente con estimar un valor no nulo en el *Pearson Product-Moment Correlation* para rechazar la hipótesis de independencia entre la Frecuencia y la Severidad. Por tal motivo, se llevaron a cabo de manera conjunta dos pruebas que buscaban justificar si, dada la naturaleza de las propias variables, el *Pearson Product-Moment Correlation* podría dar una señal, no sólo de dependencia, sino de la estructura de la asociación resultante: La primera fue una prueba gráfica de linealidad a través de gráficos de residuo parcial; y la segunda fue una prueba de distribución normal bivariada a través del Test de Shapiro-Wilk Bivariado para probar si la frecuencia y la severidad seguían

⁴Por ejemplo, una transformación de tipo $\arcsin(V)$ con V como X o Y

una distribución multivariada esférica o elíptica, en las cuales el *Pearson Product-Moment Correlation* es una medida natural de dependencia (Kumar, 2010).

3.2.2 Asociación de Rango

Dados los inconvenientes de la correlación de Pearson, se utilizaron tres medidas de dependencia alternativas⁵ basadas en la correlación de rango (*rank correlation*) o sujetas a la concordancia y discordancia estadística (Nelsen, 2006). Entiéndase como correlación de rango la que se deriva del rango (R_{ik}) de la k -ésima variable con i -ésima observación, X_{ik} , donde $i = 1, \dots, n$, tal que el ordenamiento obtenido a lo largo de X_{1k}, \dots, X_{nk} , implique que la observación más pequeña tenga un rango igual a 1 (Rémillard, 2010).

Por su parte, defínase la concordancia estadística como la característica bajo la cual dos pares de observaciones $\{i, j\}$ ($i \neq j$) de frecuencia y severidad (x_i, y_i) y (x_j, y_j) cumplen si $(x_i < x_j) \wedge (y_i < y_j) | (x_i > x_j) \wedge (y_i > y_j)$ ó $(x_i - x_j)(y_i - y_j) > 0$, es decir, asociaciones entre la frecuencia (X) y la severidad (Y) cuando una de ellas tiende a la misma dirección conforme la otra toma valores más grandes o más pequeños. Su contraparte, la discordancia, alude al cumplimiento de $(x_i < x_j) \wedge (y_i > y_j) | (x_i > x_j) \wedge (y_i < y_j)$ ó $(x_i - x_j)(y_i - y_j) < 0$ (Nelsen, 2006). Cabe decir que las medidas de asociación de rango tienen como ventaja la invarianza ante transformaciones estrictamente crecientes (Joe, 1997), lo cual es un atributo deseable en el contraste de hipótesis y en estimaciones estadísticas (Kumar, 2010).

Ahora bien, la primera de las medidas fue el Rho de Spearman (ρ), el cual se obtuvo de la correlación lineal de las Funciones de Distribución Marginal (FDM) $F_1(\cdot)$ y $F_2(\cdot)$ de X y Y , respectivamente. Su formulación es $\rho = r_P(F_1(X), F_2(Y))$ (Joe, 1997) y se considera como un *Pearson Product-Moment Correlation* en términos de «... la proporción de variabilidad explicada» (Puka, 2010), es decir, refleja la asociación monótona entre las variables implicadas (Modarres, 2010). No obstante, el Rho de Spearman es únicamente invariante ante transformaciones monótonas bajo cambios en escala o en localización (Modarres, 2010). La segunda medida de asociación de rango fue el Tau de Kendall (τ) que mide qué tan fuerte es la relación entre variables en nivel ordinal (Puka, 2010) y se calcula, de manera no paramétrica, como la diferencia entre las probabilidades de concordancia y discordancia (Nelsen, 2006). Una de las ventajas que otorga esta medida sobre ρ es el ajuste con datos atados (*tied data*)⁶, como los utilizados para el desarrollo de esta investigación. La tercera y última de estas medidas es el coeficiente de correlación medial o Beta de Blomqvist (β), la cual es una modificación al τ cuyas probabilidades de concordancia y discordancia se hacen con respecto a un vector aleatorio de las variables X (Frecuencia) y Y (Severidad) y sus respectivas medianas \tilde{x} y \tilde{y} (Nelsen, 2006).

De este modo, sean $(X_1, Y_1), (X_2, Y_2), (X_1, Y_3)$ tres vectores aleatorios iid de las variables

⁵Para una revisión general de las características deseadas de una medida de dependencia, remítase al apéndice D

⁶Conjunto de datos de dimensión d $\{\mathbf{X}_i\}_{i=1}^n$ en cuya muestra aleatoria univariada $\{X_{ij}\}_{i=1}^n$ se dé $X_i = X_k$ para algún $i \neq k$ (Hofert, Kojadinovic, Machler & Yan, 2018)

X (Frecuencia) y Y (Severidad), las versiones poblacionales de las medidas anteriores son, respectivamente, las ecuaciones (4), (5) y (6).

$$\rho = 3 (\mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) < 0]) \quad (4)$$

$$\tau = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0] \quad (5)$$

$$\beta = \mathbb{P}[(X - \tilde{x})(Y - \tilde{y}) > 0] - \mathbb{P}[(X - \tilde{x})(Y - \tilde{y}) < 0] \quad (6)$$

3.2.3 Cópula

Si bien las medidas de asociación de rango brindaron ciertas ventajas sobre la correlación, tienen un problema de inconsistencia. En otros términos, el poder de las pruebas de independencia no tiende a 1 conforme el tamaño de la muestra converge al tamaño poblacional (Rémillard, 2010). Dado este antecedente, se usaron estadísticos basados en la función de distribución empírica conjunta \hat{H} (Rémillard, 2010), la cual proviene de $H(x, y) = \mathbb{P}[X \leq x, Y \leq y]$ ⁷, donde F y G son las FDM de X y Y (Nelsen, 2006).

En vista de que la dependencia entre la frecuencia y la severidad conllevaría a una especificación particular de $H(\cdot)$, se introdujo la cópula como un elemento que diera cuenta de la estructura de dicha dependencia (Embrechts y col., 1998). Con esto dicho, defínase la cópula como una función de distribución multivariante de marginales univariadas uniformes (Embrechts y col., 1998) cuyo modelamiento, realizado por separado de las marginales (Shi y col., 2015), permite describir la estructura mencionada. Por ejemplo, si la frecuencia y la severidad hubiesen sido independientes bajo este tipo de aseguramiento, la especificación de su cópula recibiría el nombre de Cópula de Independencia (C_{Π}), la cual sería igual al producto de sus marginales (Edward W. Frees & Valdez, 1998). El rol de las cópulas en la relación que existe entre las funciones de distribución multivariante y las marginales es otorgado por el Teorema de Sklar (Nelsen, 2006), el cual es el siguiente:

Teorema 1 (Teorema de Sklar). *Sea H una función de distribución conjunta con marginales F y G . Entonces existe una cópula C tal que para todo x, y en $\bar{\mathbf{R}}$,*

$$H(x, y) = C(F(x), G(y)) \quad (7)$$

Si F y G son continuas, entonces C es única; de otra manera, C es únicamente determinada sobre $\text{Ran}F \times \text{Ran}G$. Por el contrario, si C es una cópula y F y G son funciones de distribución, entonces la función H definida por (7) es una función de distribución conjunta con marginales F y G (Nelsen, 2006, Teorema 2.3.3).

Por consiguiente, si existe un vector $\mathbf{U} = (u, v)$, con $\mathbf{U} \sim U(0, 1)$, y las funciones de cuantiles, $(F^{-1}(x), G^{-1}(y))$, entonces se pueden definir $u = F(x)$ y $v = G(y)$ tal que H se

⁷Su función de supervivencia es $\bar{H}(x, y) = 1 - F(x) - G(y) + H(x, y)$

reescriba como en (8), lo que muestra que la Cópula es una función de variables uniformemente distribuidas. Y como el Teorema 1 brinda una especificación de H en términos de la cópula⁸, las ecuaciones (4), (5) y (6) pueden reescribirse, respectivamente, como (9), (10) y (11), donde $\mathbf{I}^2 = [0, 1] \times [0, 1]$.

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)) \quad (8)$$

$$\rho = 12 \int \int_{\mathbf{I}^2} C(u, v) dudv - 3 \quad (9)$$

$$\tau = 4 \int \int_{\mathbf{I}^2} C(u, v) dC(u, v) - 1 \quad (10)$$

$$\beta = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1 \quad (11)$$

No obstante, las medidas anteriores no consideraron la dependencia en los valores extremos, aspecto relevante bajo el tipo de aseguramiento trabajado, donde los valores de las distribuciones de la frecuencia y la severidad pueden estar ampliamente desviadas con respecto al valor esperado (Joe, 1997). Por este motivo, se introdujo la dependencia de cola (*tail dependence*), la cual Joe (1997) y Nelsen (2006) relacionaron con el nivel de dependencia en las colas de los cuadrantes superiores e inferiores de la distribución bivariada sobre \mathbf{I}^2 , y cuya medida proviene de la evaluación de la distribución bivariada en el límite de sus extremos 1 y 0, por lo que resulta en dos variantes, a saber: la de dependencia de cola superior, λ_U (ecuación (12)), y la de dependencia de cola inferior, λ_L (ecuación (13)), donde ambas toman el valor de cero cuando no existe este tipo de dependencia.

$$\lambda_U = \lim_{u \rightarrow 1} \frac{\bar{C}(u, u)}{1 - u} \quad (12)$$

$$\lambda_L = \lim_{u \rightarrow 0} \frac{C(u, u)}{u} \quad (13)$$

Por otro lado, en lo que concierne a la estimación de las cópulas, se consideraron diferentes especificaciones puesto que las marginales trabajadas son mixtas (la frecuencia sigue una distribución discreta y la severidad una distribución continua), lo que implica que la cópula

⁸Las medidas de asociación de rango en términos de H (Nelsen, 2006) son

$$\rho = 12 \int \int \bar{H} dF dG$$

$$\tau = 4 \int H dH - 1$$

$$\beta = 4 \int \bar{H} d\bar{H} - 1$$

C es «únicamente determinada en $RanF \times RanG$ » (Teorema 1). No obstante, como lo sugirieron Trivedi y Zimmer (2005), la no unicidad de la representación de la cópula es « *un asunto teórico que necesita ser confrontado con pruebas analíticas, pero que no inhibe las aplicaciones empíricas*». En otros términos, para que sea posible la identificación de una única cópula para la frecuencia y la severidad se requiere saber *a priori* la forma de su distribución conjunta, lo cual, al no ser posible en la práctica, conlleva a que investigadores se enfrenten al modelamiento y selección de una cópula adecuada que logre dar cuenta de la estructura de la dependencia sin «(...) *sacrificar las propiedades atractivas de las marginales*» (Trivedi & Zimmer, 2005). En consecuencia, los métodos disponibles para la selección de la cópula fueron:

1. Estimación Paramétrica de las FDM: Parte del supuesto de que la función de distribución conjunta H , siendo desconocida, pertenece a una familia paramétrica de funciones de distribución, para luego estimar la cópula desconocida C_{θ_0} por medio de:
 - 1.1 El estimador por máxima verosimilitud (*MLE: Maximum Likelihood Estimator*), cuya carga computacional conlleva al empleo del método [1.2] (Hofert y col., 2018).
 - 1.2 El estimador de dos etapas conocido como IFME (*Inference Functions for Margins Estimator*), que consiste en la estimación de los vectores de parámetros desconocidos de las marginales en la primera etapa, y posteriormente en la estimación del vector del parámetro de la cópula desconocida en la segunda etapa. No obstante, presenta una pérdida de eficiencia con respecto al MLE a costa de la ventaja computacional (Hofert y col., 2018).
2. Estimación No Paramétrica de las FDM: Dado que la estimación anterior puede conllevar a errores de especificación de las marginales, la estimación no paramétrica es una alternativa basada en funciones de distribución empíricas reescaladas a partir del empleo de pseudo-observaciones (método de reescala con base al rango). Con esto resulta el siguiente par:
 - 2.1 El estimador de método de momentos (*MME: Method-of-Moments Estimator*), el cual reemplaza los momentos de las variables aleatorias por los momentos de la cópula a través de τ o ρ (Hofert y col., 2018).
 - 2.2 El estimador por máxima pseudo-verosimilitud (*MPL: Maximum Pseudo-Likelihood Estimator*), el cual consiste en reemplazar la muestra inobservable de U por las pseudo-observaciones y así llevar a cabo el algoritmo de MLE (Hofert y col., 2018).

Para efectos de este documento, se tomaron dos de las cuatro alternativas anteriores: El IFME y el MPL. En lo referido al IFME, la primera etapa consistió en la elección de la distribución teórica que mejor se ajustara a los datos y cuya lista de opciones fueron tomadas de Tse (2009) o de las propias características de las variables. De esta manera, en la Frecuencia

se tuvieron en cuenta las distribuciones Geométrica, Binomial Negativa y Poisson; y en la Severidad (tanto Severidad 1 como Severidad 2) se consideraron las distribuciones Exponencial, Loglogística, Pareto, Gamma y Gumbel. Posteriormente, la segunda etapa consistió en la búsqueda de la mejor cópula de acuerdo al criterio de información de Akaike (AIC), o a través de las pruebas de bondad de ajuste (GoF) con base en los métodos de White y de Kendall, que pueden ser encontrados en Nagler y col. (2019), y cuya estimación se realizó por medio de MLE. Por su parte, la segunda alternativa (MPLE) se realizó bajo enfoque no paramétrico que, a diferencia del anterior, tomó como insumo la base de datos en pseudo-observaciones como primera etapa y, posteriormente en la segunda etapa, se llevó a cabo el MLE.

Una vez determinadas las diferentes especificaciones de las cópulas, se procedió a calcular las medidas de asociación de rango y de dependencia de cola con base a las ecuaciones (9), (10), (11), (12) y (13), las cuales sirvieron como punto de comparación con las inicialmente calculadas. Lo anterior puso en relieve que las medidas están sujetas a los parámetros de las Cópulas, y éstas a su vez a su familia. Se sugiere remitirse al apéndice B para una mayor discusión sobre las cópulas y la dependencia.

4 Resultados

En esta sección se discuten los resultados obtenidos de los tres grupos de procedimientos descritos en la sección 3. En el desarrollo del análisis se contempló la influencia de los *outliers*, definiéndolos como todo dato por encima del percentil 99 de la distribución de Severidad 1, pues es donde estaban ubicados los valores más extremos.

En suma, los resultados mostraron la presencia de dependencia entre la frecuencia y la severidad para el tipo de aseguramiento trabajado. Es decir, se rechazó la hipótesis de una medida de dependencia igual a cero, por lo que el supuesto de la independencia no es válido para el FSM-EUI.

En cuanto a los resultados generales por grupo de procedimiento, los dos pares de análisis de frecuencia y severidad no siguieron una distribución de la familia elíptica, además de no estar relacionados de manera lineal, aunque esto no implicó la no idoneidad del *Pearson Product-Moment Correlation* que fue distinto de cero. Esta situación fue similar con las medidas de asociación de rango que, junto a sus respectivas especificaciones en términos de la función de distribución conjunta, rechazaron la hipótesis nula de independencia. Además, con la introducción de las cópulas se encontró la presencia de una dependencia de cola y un patrón en la estructura de la dependencia.

Las generalidades anteriores se discuten a continuación con mayor precisión, resaltando la agrupación de los dos primeros grupos de métodos del esquema propuesto en la sección anterior:

4.1 Asociación lineal y Asociación de rango

Iniciando por la correlación lineal, en el cuadro 3 se encuentra el valor de r_P para la pareja Frecuencia-Severidad 1 que fue igual a 0.6733, valor que es superior a 0.4482, que corresponde al cálculo sin *outliers*. Por su parte, la correlación del par Frecuencia-Severidad 2 fue 0.1615 con *outliers* y 0.0934 sin ellos, lo que puso en relieve la sensibilidad del *Pearson Product-Moment Correlation* ante los datos atípicos. Cabe recordar que esta correlación es una medida de dependencia toda vez que la frecuencia y la severidad tengan una relación lineal y sigan una distribución normal bivariada. Sin embargo, en la prueba de linealidad a través de gráficos de residuo parcial (*Partial Residual Plot*) para los dos pares de análisis, se observó la no linealidad entre la frecuencia y la severidad (en ambas definiciones), lo cual se puede detallar en las diferencias entre las líneas de los residuos (azul) y las líneas de los componentes (rosa) en los gráficos 3 y 4, los cuales están basados en los datos sin los *outliers* y cuyos resultados fueron robustos con respecto a toda la base de datos. Lo anterior es respaldado por el test de Shapiro-Wilk bivariado, que sugirió el rechazo de la hipótesis de una distribución normal bivariada de la frecuencia y la severidad, lo cual puede observarse en el cuadro 2 donde los resultados fueron robustos ante *outliers* y los valores p fueron inferior a 0.001.

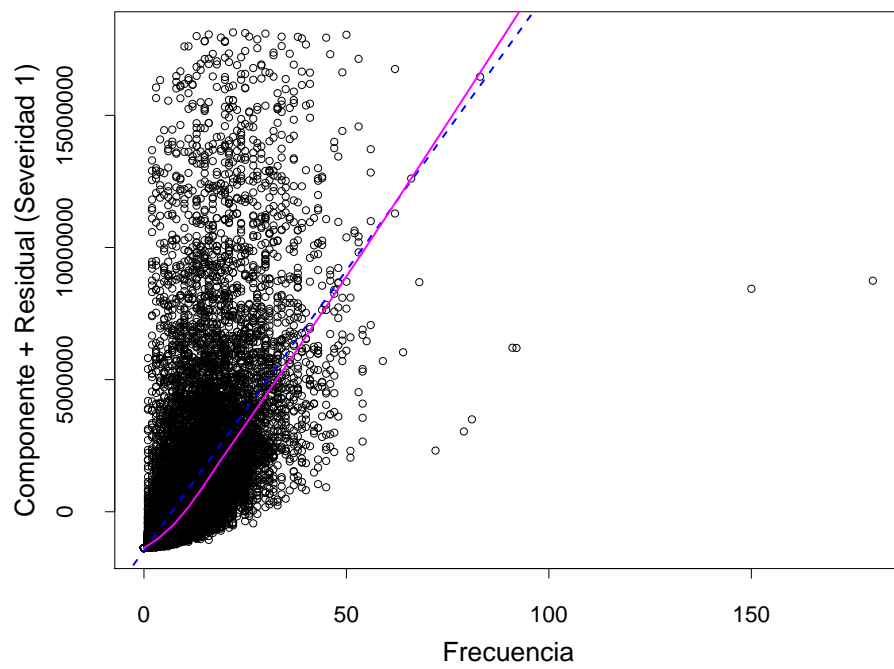


Figura 3: Residual Partial Plot para *Frecuencia-Severidad 1*

Fuente: Elaboración propia

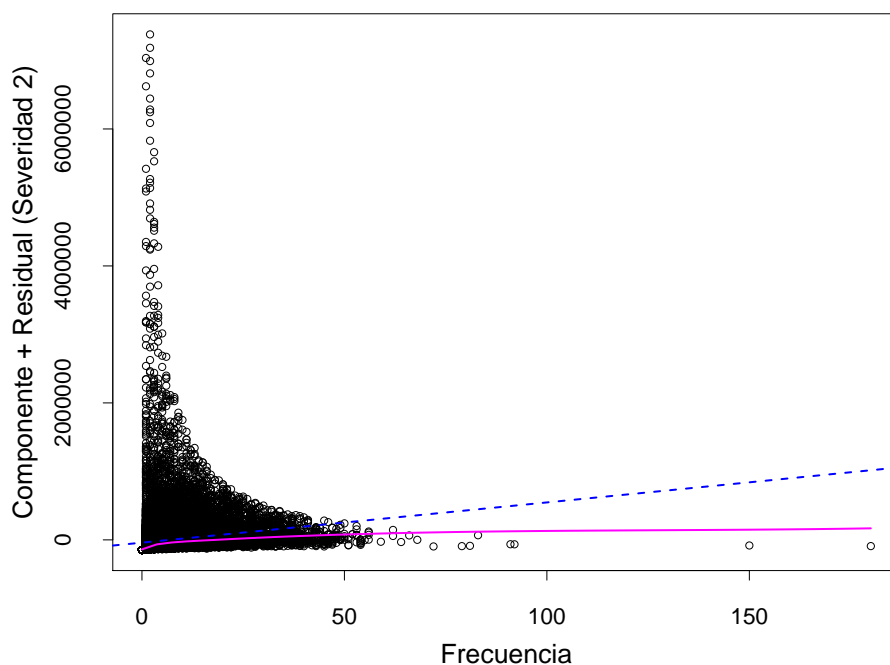


Figura 4: Residual Partial Plot para *Frecuencia-Severidad 2*

Fuente: Elaboración propia

Cuadro 2: Resultados del Test de Shapiro-Wilk Bivariado

Datos	Pareja	Estadístico Shapiro-Wilk	Valor p
Completa	Frecuencia-Severidad 1	0.2184	<0.001
Completa	Frecuencia-Severidad 2	0.2340	<0.001
Sin Outliers	Frecuencia-Severidad 1	0.5595	<0.001
Sin Outliers	Frecuencia-Severidad 2	0.3694	<0.001

Fuente: Elaboración propia

Por ende, lo anterior evidenció que la medida de correlación lineal no es adecuada para medir la independencia entre la frecuencia y la severidad, puesto que no tienen: (1) una relación lineal y (2) una distribución normal bivariada. No obstante, los valores de la medida distintos de cero fueron suficientes para deducir la existencia de una asociación, como se mencionó en la sección 3⁹. Estos resultados fueron respaldados por las otras medidas de asociación disponibles en el cuadro 3, las cuales, siendo más robustas ante los valores atípicos,

⁹Debido al *modus tollendo tollens* que implica lo siguiente: Si X y Y son independientes, entonces su correlación es cero. Por ende, negar el consecuente conlleva a negar el antecedente.

evidenciaron el alto grado de dependencia entre la Frecuencia y la Severidad 1, ya sea con ρ , que mide la asociación entre los rangos, o τ , que muestra el grado de monotonicidad en la relación. No obstante, la intensidad de dichas asociaciones se vieron menguadas al analizar el par que toma a la severidad como el costo promedio por reclamación a nivel de afiliado (Severidad 2). Como conclusión, las medidas de asociación lineal y de rango dieron valores que rechazaron la H_o de una independencia entre la frecuencia y la severidad.

Cuadro 3: Medidas de asociación

Outliers	Frecuencia-Severidad 1			Frecuencia-Severidad 2		
	r_P	ρ	τ	r_P	ρ	τ
Sí	0.4482	0.9102	0.7732	0.0934	0.6707	0.5115
No	0.6733	0.9099	0.7726	0.1615	0.665	0.5059

Resultados con $p < 0.001$

Fuente: Elaboración propia

4.2 Cópula

Una vez corroborada la existencia de una dependencia entre las variables, se procedió a la especificación de la estructura de dicha asociación y al consecuente cálculo de las medidas de dependencia en términos de cópula, donde cabe recordar que se consideraron el τ y β , además de las medidas de dependencia de cola λ_L y λ_U . Se prescindió del ρ , puesto que alguna de sus versiones en cópulas resultan difíciles de estimar; aunque sus resultados fueron capturados por el β .

Como se mencionó en el apartado 3.2.3, la primera de las alternativas para la especificación de dicha estructura partió del enfoque paramétrico bajo el método de IFME, del cual se eligieron las mejores distribuciones teóricas que se ajustaran a los datos siguiendo el criterio de información de Akaike (AIC). Dichas distribuciones fueron la Binomial Negativa, para la Frecuencia; Gamma, para la Severidad 1; y Log-Logística, para la Severidad 2. En aras de ilustrar estos resultados, las figuras 5, 6 y 7 mostraron los ajustes de las distribuciones candidatas para cada la Frecuencia, la Severidad 1 y la Severidad 2, respectivamente, donde la mayor cercanía de los diferentes tipos de líneas a la senda de la Función de Distribución Acumulada (FDA) empírica de la variable evaluada, que es de color negro, mostraría la distribución teórica con mejor ajuste.

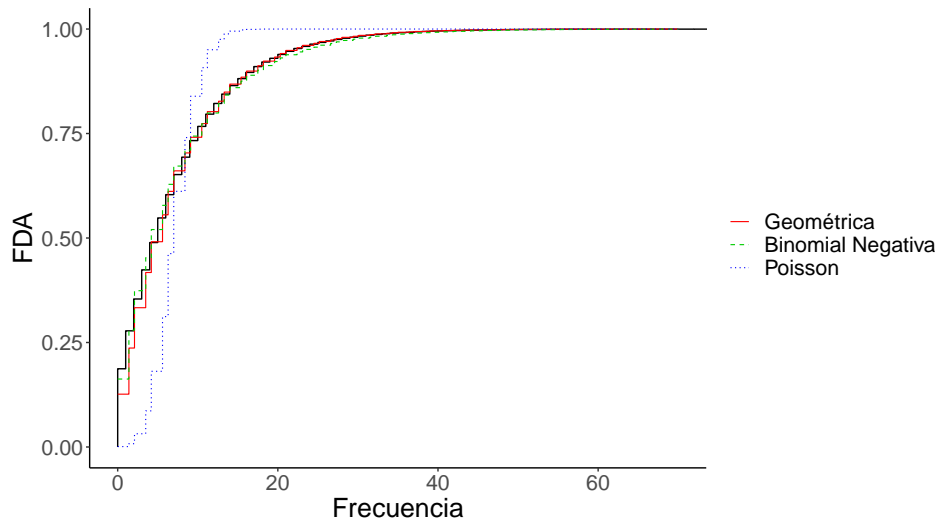


Figura 5: FDA y distribuciones teóricas para la Frecuencia

Fuente: Elaboración propia

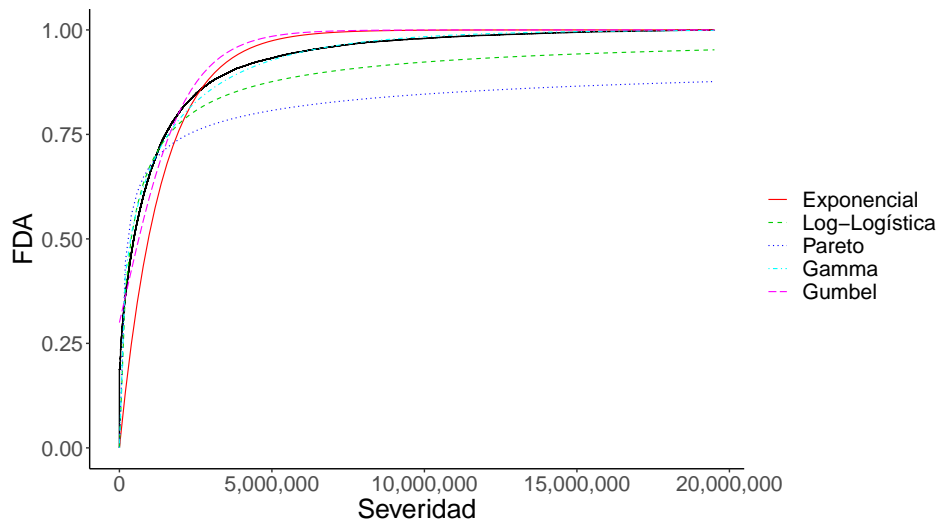


Figura 6: FDA y distribuciones teóricas para la Severidad 1

Fuente: Elaboración propia

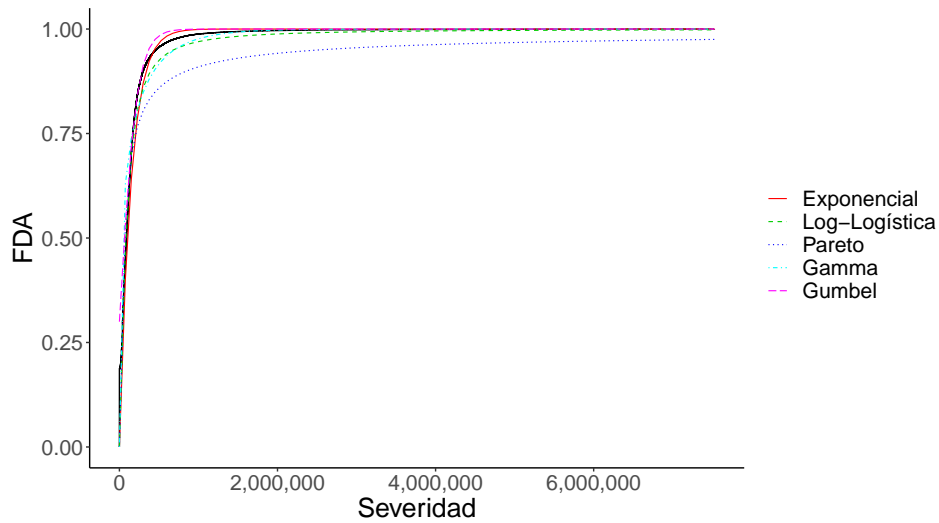


Figura 7: FDA y distribuciones teóricas para la Severidad 2

Fuente: Elaboración propia

Una vez obtenidas las distribuciones teóricas con mejor ajuste, se procedió a la estimación por MLE, cuyos resultados, como las del enfoque no paramétrico, se encuentran disponibles en el cuadro 4. En lo que respecta a las cópulas seleccionadas¹⁰, se apreció la prevalencia de la Cópula Joe para cada una de las parejas bajo ambos enfoques. Una característica de esta cópula es la dependencia de cola superior, lo que implica que, de entrada, garantizó un valor positivo en λ_U , lo que visualmente se puede apreciar en la acumulación en la esquina superior derecha de figura 8 que ilustra el contorno de la cópula Joe obtenida de manera no paramétrica para la pareja Frecuencia-Severidad 1 bajo el método de selección de Kendall y cuyos ejes son las pseudo-observaciones de la frecuencia (u_1) y la severidad (u_2). Adicionalmente, los valores de los parámetros θ de la cópula Joe fueron superiores a 1, lo que indica su lejanía a la Cópula de Independencia (C_I), que sería el resultado en caso de que las variables analizadas hubiesen sido independientes.

Cuadro 4: Cópulas, Parámetros y Medidas de Dependencia

Enfoque	Grupo	Método de Selección	Cópula	τ	β	λ_L	λ_U
Paramétrico	Frecuencia-Severidad 1	AIC	Frank($\theta = 10.892$)	0.687	0.747	0	0
		Kendall	Joe($\theta = 3.232$)	0.544	0.563	0	0.761
	Frecuencia-Severidad 2	AIC	Survival BB8($\theta_1 = 6, \theta_2 = 0.575$)	0.426	0.472	0	0
		Kendall	Joe($\theta = 1.812$)	0.31	0.306	0	0.534
No Paramétrico	Frecuencia-Severidad 1	AIC	Survival BB6($\theta_1 = 6, \theta_2 = 1.321$)	0.79	0.819	0.909	0
		Kendall	Joe($\theta = 7.571$)	0.773	0.809	0	0.904
	Frecuencia-Severidad 2	AIC	Survival Joe($\theta = 3.848$)	0.602	0.627	0.803	0
		Kendall	Joe($\theta = 2.902$)	0.506	0.52	0	0.73

Fuente: Elaboración propia

¹⁰Sus características se encuentran en el apéndice C

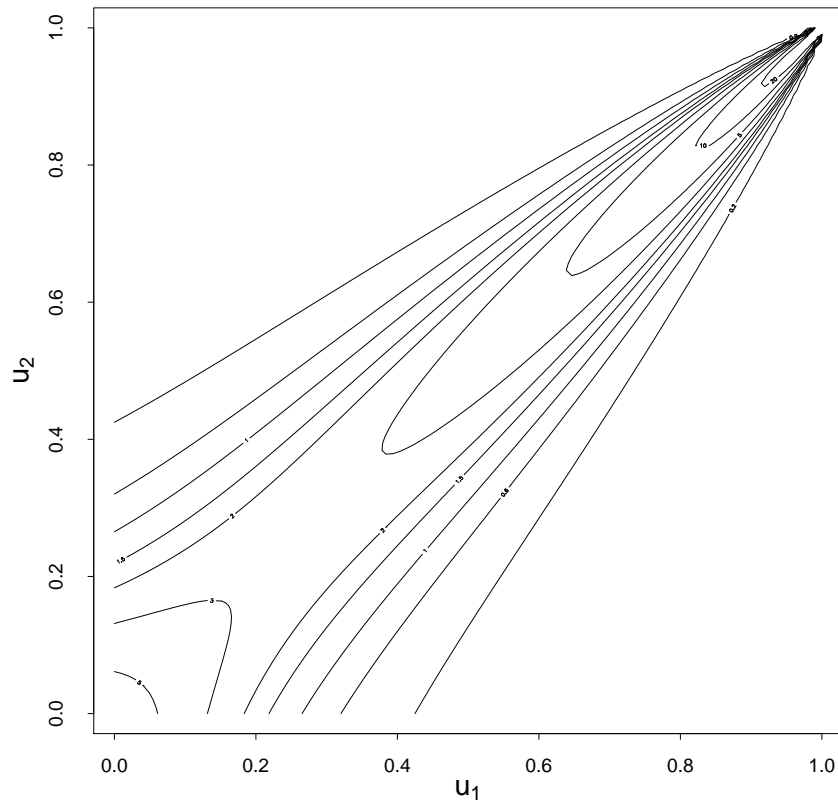


Figura 8: Gráfico de Contorno de Joe($\theta=7,57$)

Fuente: Elaboración propia

Por otro lado, la existencia de la cópula de supervivencia de Joe (*Survival Joe*) indicó la presencia de una dependencia en la cola inferior, lo que se tradujo en un valor positivo en λ_L . Esta consecuencia también fue proveniente de la cópula de supervivencia *Survival BB6*, cuyo primer parámetro señaló qué tan fuerte es la concordancia en su planteamiento convencional. Aunque lo anterior no se evidenció en la cópula biparamétrica de supervivencia *Survival BB8* para la pareja Frecuencia-Severidad 2 bajo enfoque paramétrico. Visualmente se puede apreciar la dependencia de cola inferior en la figura 9, la cual representa la cópula seleccionada para la pareja Frecuencia - Severidad 1 bajo AIC en enfoque no paramétrico.

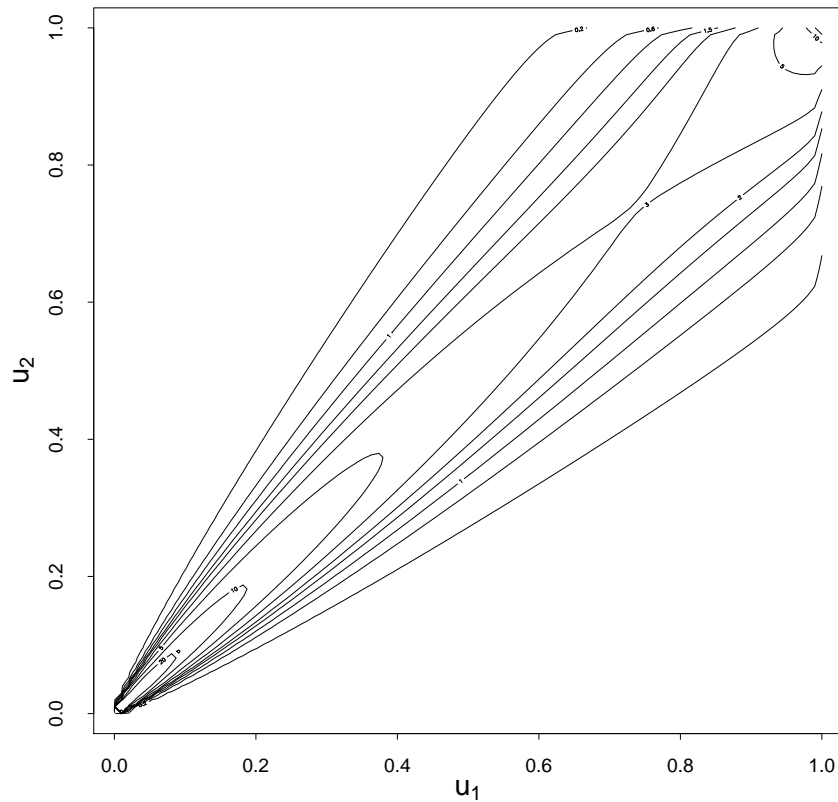


Figura 9: Gráfico de Contorno de *Survival BB6* ($\theta_1=6, \theta_2=1.321$)

Fuente: Elaboración propia

En lo que concierne a las medidas de asociación, se percata que el β fue una medida aproximada al ρ , como lo sugirió Nelsen (2006), aunque sólo se pudo apreciar bajo el enfoque no paramétrico indistintamente de la pareja analizada. Este mismo fenómeno pudo notarse en el τ resultante del cuadro 4, el cual se acercó a sus respectivos valores en el cuadro 3 para ambas parejas. Esto implicó que los inconvenientes resaltados por Hofert y col. (2018), en lo que respecta a los problemas de especificación incorrecta de las marginales, tuvieron como resultado las subestimaciones de las medidas de dependencia. Adicionalmente, el signo positivo de las medidas de asociación y la existencia de dependencia en los respectivos valores extremos no sólo fue observado analíticamente, como lo sugirió el gráfico de dispersión de las variables en pseudo-observaciones (Figura 10).

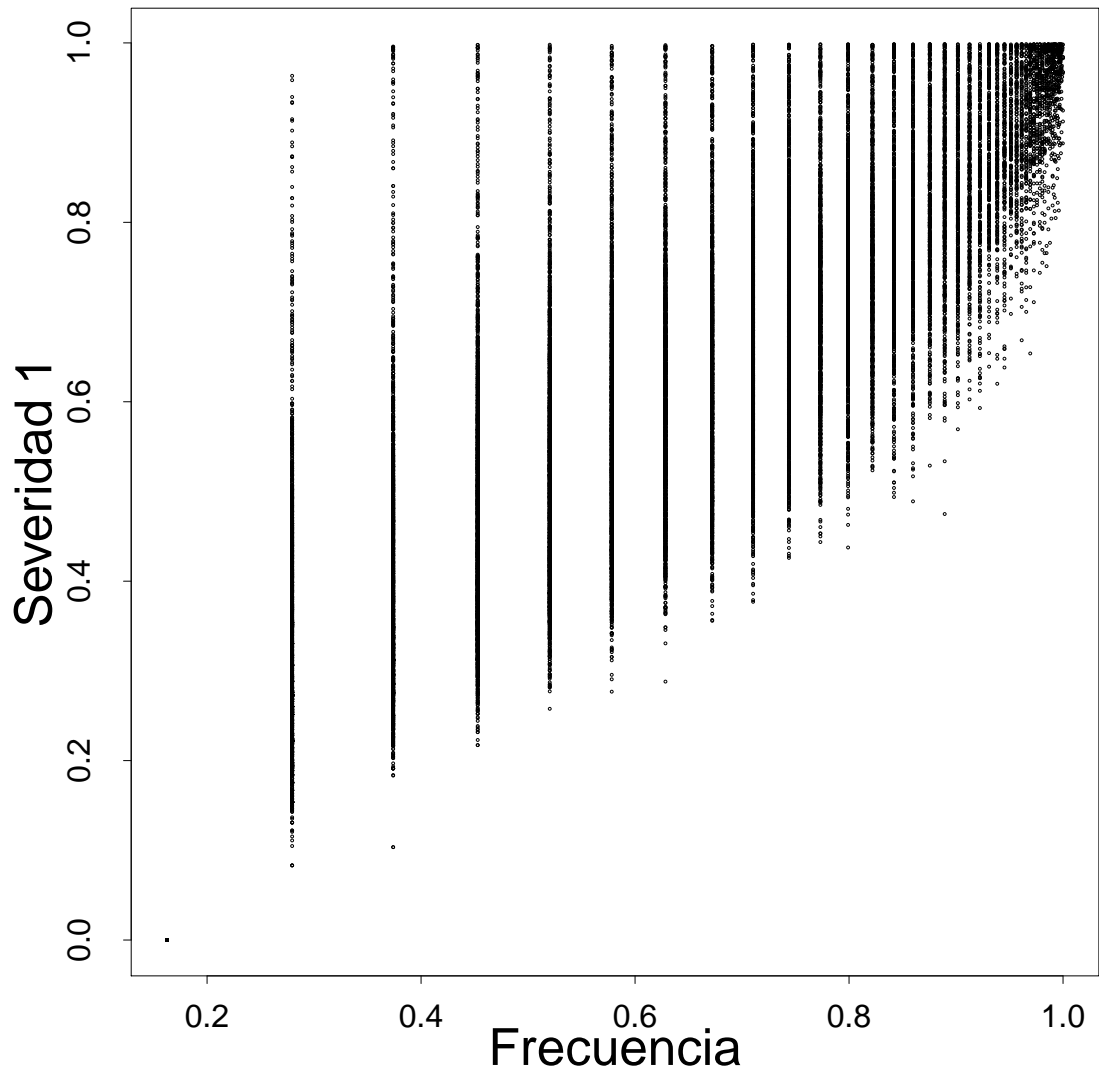


Figura 10: Gráfico de dispersión de pseudo-observaciones para la pareja *Frecuencia-Severidad 1*

Fuente: Elaboración propia

4.3 Discusión

Los signos positivos de las medidas de dependencia estimadas mostraron la posible subestimación de las pérdidas agregadas que son incurridas por la aseguradora cuando el supuesto de independencia asumido no se cumple (Krämer y col., 2013). La subestimación en las pérdidas agregadas puede tener repercusión sobre las reservas técnicas, las cuales son recursos destinados para el cubrimiento de obligaciones actuales o contingentes de la EPS, por lo que funciona

como instrumento de determinación de la aptitud de la aseguradora en su operación y cuyo marco legal lo brinda el decreto 2702 de 2014 (Gobierno Nacional, 2014; Superintendencia Nacional de Salud, 2014) . Estas reservas pueden contener obligaciones pendientes y conocidas, pendientes no conocidas o de otro tipo; obligaciones que al basarse en costos subestimados, pueden conllevar a la subvaloración de las reservas técnicas, lo que impactaría en la solvencia o en la condición financiera de la aseguradora, como lo sería el caso de reservas de primas al ser una función de las primas estimadas.

En aras de ilustrar la subestimación del costo de las pérdidas se estimó la prima pura y se agregó a nivel del *pool* bajo tres diferentes modelos¹¹. El primero de ellos es el Modelo Tweedie que, al ser un proceso compuesto de Poisson, no separa la prima pura en sus componentes (Jorgensen, 1987) y cuya prima estimada se encuentra en la columna **Modelo Tweedie** del cuadro 5. El segundo de ellos es la versión separada del proceso compuesto mencionado, es decir, un Modelo de Dos Partes con distribuciones gamma para la severidad y Poisson para la frecuencia cuya prima estimada está en la columna **FSM (Gamma-Poisson)**. Por último, dado que en este documento la distribución Binomial Negativa fue elegida elegida para la frecuencia de acuerdo con el criterio de información de Akaike (AIC), el tercer modelo es uno de Frecuencia y Severidad que contempla la posibilidad de sobredispersión, la cual es un fenómeno en el que la varianza suele ser superior a la media y que puede resultar de ajustar a la frecuencia con una distribución Poisson (Goldburd y col., 2008)¹²; los resultados de este último modelo se encuentran en la columna **FSM (Gamma - Binomial Neg)** del cuadro 5. Los resultados evidencian una subestimación de los costos totales, cuyos valores representan cerca del 75 % de los costos observados que equivalen a \$63,569,789,935 y cuyas diferencias en nivelen están entre \$14,920,143,020 (Modelo FSM (Gamma-Poisson)) y \$15,841,899,602 (Modelo FSM (Gamma-Binomial Negativa))¹³.

Cuadro 5: Ejemplos de subestimación de primas puras

Descripción	Observado	Modelo Tweedie	FSM (Gamma-Poisson)	FSM (Gamma-Binomial Neg)
Costos totales	\$63,569,789,935	\$48,431,757,514	\$48,649,646,915	\$47,727,890,333
Diferencia con lo observado	\$0	\$15,138,032,421	\$14,920,143,020	\$15,841,899,602
Proporción con lo observado	1	0.7619	0.7653	0.7508

Fuente: Elaboración propia

Por otro lado, la presencia de dependencia en las colas evidenció la existencia de una asociación en los valores extremos en materia de aseguramiento para Planes Voluntarios en Salud, los cuales se relacionan con escenarios de aseguramiento para eventos en salud con Alta Frecuencia y Alta Severidad (AFAS) o Baja Frecuencia y Baja Severidad (BFBS). El

¹¹Se controló por sexo y por género para todos los modelos.

¹²Lo que hace la distribución Binomial Negativa es tratar a la media de la distribución Poisson (usualmente representada como λ) como una variable aleatoria que se distribuye con gamma (Goldburd y col., 2008)

¹³Este documento es el inicio de una investigación en la que se busca estimar el monto de subestimación por medio de la introducción cópulas dentro del análisis de regresión. Este procedimiento se basaría en lo hecho por Krämer y col. (2013) quienes concluyeron que las pérdidas totales del **FSM (Gamma-Poisson)** equivalen al 93.4 % de las resultantes de una regresión basada en una Cópula Clayton en un contexto de aseguramiento de autos alemán

primero de ellos puede comprender: Fractura de cadera, Accidente Cerebrovascular (ACV), Infarto Agudo de Miocardio (IAM), cáncer de mama; cáncer cervical; reemplazo articular; trasplantes y entre otros. Y el segundo de ellos puede referirse a las cataratas, entre otros¹⁴. De aquí que sea importante mencionar que la estructura de la dependencia resultante está sujeta a la naturaleza esperada de los datos y al contexto sistémico en el que se desenvuelve este tipo de aseguramiento privado producto del mecanismo de pago, además de la existencia de un aseguramiento público paralelo.

Cabe decir que la posibilidad de modelar la dependencia a través de cópulas fijó una pauta metodológica para la investigación futura sobre el monto de subestimación de la pérdida para la aseguradora en salud a causa del incumplimiento del supuesto de independencia. En consecuencia, los resultados que contemplen dicha asociación puede aplicarse en métodos de tarificación de primas de riesgo en salud cuya precisión en los cálculos posibilitarán una correcta gestión del riesgo, que no sólo se traduce en ganancias para la aseguradora, sino en sostenibilidad para el aseguramiento al tener individuos más saludables (Gonzalo, 2015). Igualmente, reconociendo que una parte de la metodología empleó un enfoque paramétrico para la estimación de las cópulas, no se consideraron controles en los ajustes de las distribuciones iniciales. La razón por la cual no se llevó a cabo reside en que dicho procedimiento prepararía el inicio del cálculo del monto perdido a causa de asumir incorrectamente el supuesto, lo que implicaría el uso de cópulas condicionales sobre las cuales los criterios de selección están por fuera del paradigma frecuentista, lo cual fue un límite impuesto al inicio del documento al no considerar la literatura con un paradigma bayesiano. Finalmente, los resultados se circunscriben a aseguramiento en salud con Planes Voluntarios en Salud, específicamente los planes complementarios, lo que deja abierto el contraste de dicho supuesto para otros tipos de planes.

¹⁴Se agradece la colaboración de César Augusto Guevara Cuellar y Danny Moreano por las amables sugerencias.

5 Conclusiones

El objetivo central del documento fue el contraste del supuesto de independencia entre la frecuencia y la severidad en aseguramiento en salud, como también el estudio de la estructura de la dependencia resultante tomando como contexto el relacionado con Planes Voluntarios en Salud de una EPS colombiana para el año 2018.

Los resultados rechazaron la hipótesis de la independencia mencionada, los cuales fueron robustos ante datos atípicos y ante dos pares de análisis de frecuencia y severidad. El análisis del contraste se dividió en tres grupos. En el primero de ellos, se encontró que la correlación lineal es diferente de cero, aunque dicha medida tuviese inconvenientes debido al incumplimiento de linealidad entre la frecuencia y la severidad y la no distribución normal bivariada en este tipo de variables. En el segundo grupo de análisis, las medidas de asociación de rango presentaron resultados más robustos, corrigiendo los inconvenientes de la correlación lineal.

En cuanto al tercer grupo, se obtuvieron sólidos resultados con la introducción de las cópulas y las medidas de dependencia de cola, los cuales mostraron la presencia de una asociación en los valores extremos de la distribución bivariada de la Frecuencia y la Severidad. Este tipo de dependencia puede dar una pauta para políticas internas de las aseguradoras, puesto que dicho patrón en los valores extremos pueden provenir de pacientes de alto costo o de siniestros cuya naturaleza pueda indicar una mayor probabilidad de ocurrencia.

Asimismo, los resultados mostraron que la relación entre la frecuencia y la severidad es positiva, lo que indicaría que las primas podrían estar subestimadas si se emplea el supuesto en cuestión. Por esta razón, el documento ofreció una revelencia práctica, puesto que el contraste del supuesto tiene una repercusión sobre la gestión del riesgo por parte de las aseguradoras, pues al considerar la relación entre frecuencia y severidad antes de predecir la prima pura, aporta precisión sobre la medida, lo que mejora la planeación de las aseguradoras. Por ende, otorga ventajas en el mercado estratégico y ofrece un sustento para una mejor toma de decisiones, reconociendo que una correcta gestión del riesgo se traduce, no sólo en ganancias para la aseguradora (o no siniestralidad), sino en su sostenibilidad (Gonzalo, 2015).

Por último, el documento ha expuesto un nicho para el desarrollo de investigaciones sobre nuevos elementos diferenciadores entre el riesgo en salud y el riesgo financiero desde el marco conceptual tratado, puesto que la existencia y el reconocimiento de los distintos factores incidentes permiten explorar la naturaleza de sus dependencias.

Apéndice A Vocabulario

- **Aseguramiento en salud** (*Health Insurance*): De acuerdo con Pitacco (2014), se refiere al conjunto de productos financieros que, sujetos a un siniestro proveniente de una enfermedad o un accidente, proveen beneficios en aras de cubrir los gastos o pérdidas de ingresos asociados. Estos gastos son los incurridos ante la prestación de servicios médicos bajo los ámbitos de hospitalización, urgencias, ambulatorio o domiciliario. Por su parte, la pérdida de ingresos puede ser parcial o total, permanente o no, como alguna invalidez.

Este tipo de aseguramiento se clasifica en dos grupos según la naturaleza del siniestro y el tiempo de provisión de los beneficios; estos son: De vida (*life insurance*) y de No vida (*non-life health insurance*). De estos se dividen diferentes tipos como los seguros de enfermedad, de accidente, de protección del ingreso, de enfermedad crónica o aseguramiento de cuidados en salud de largo plazo, entre otros (Pitacco, 2014).

Para efectos de este documento, y sujeto al contexto colombiano, los beneficios que se contemplan son monetarios, los cuales recurren a los beneficios de reembolso (*Reimbursement benefits*) (Pitacco, 2014) que asumen los costos en salud (total o parcialmente) y se aplica cuando la aseguradora es intermediaria dentro de la provisión de salud al estilo de Londoño y Frenk (1997). Para una revisión más general, se recomienda leer el capítulo 3 de Pitacco (2014).

- **Gestión del riesgo:** Al ser el punto central de la función de una aseguradora en salud, se proponen tres definiciones bajo tres enfoques: Procedimental Organizacional y Procedimental en Salud. La primera se obtiene de Pitacco (2014), quien ve a la gestión del riesgo como un proceso que consta de los siguientes pasos:
 1. Determinación de objetivo: El beneficio esperado, la participación dentro del mercado, etc.
 2. Identificación del riesgo: Se separan las causas y los componentes del riesgo, donde el primero alude a la presencia de morbilidad, invalidez, longevidad, gastos, etc; mientras que el segundo se refiere a desviaciones sistemáticas o fluctuaciones aleatorias. Es similar a la identificación de la naturaleza del riesgo si es controlable o no desde el punto de vista financiero.
 3. Evaluación del riesgo: Modelamiento estocástico del riesgo.
 4. Evaluación de impacto: Análisis que relaciona los efectos del modelamiento del riesgo en términos de los resultados de interés. Por ejemplo, sobre los objetivos determinados en el primer paso.
 5. Análisis de acciones: Análisis de costo-beneficio para conformación de un conjunto de acciones.
 6. Elección de acciones: Elección dentro del conjunto de acciones.
 7. Monitoreo: Seguimiento sobre los resultados y las bases estadísticas.

8. Diseño de producto: Este es el punto de conexión entre el monitoreo y la identificación del riesgo, sobre el cual se hacen constantes ajustes en vía de los objetivos del punto inicial.

El segundo enfoque, no alejado al de Pitacco (2014) es propuesto por Paredes Cubillos (2015), quien define a la gestión del riesgo como la identificación y el análisis de situaciones que impacten sobre el giro de los negocios de la firma y, posteriormente, la planeación y la ejecución de procesos sujetos a eventos esperados de siniestralidad.

Apéndice B Cópulas y dependencia

La forma en que la cópula se relaciona con la dependencia tiene que ver con lo planteado por Nelsen (2006) y Heilpern (2014) a través de las cotas Frechet-Hoeffding, las cuales permiten entender cómo las medidas de dependencia y de asociación se relacionan entre sí y las cópulas. Más precisamente, sea u_1 y u_2 las f.d.m uniformes, y las cópulas bivariadas: $C(\mathbf{u}) = C(u_1, u_2)$, $W(\mathbf{u}) = \max(u_1 + u_2 - 1, 0)$ y $M(\mathbf{u}) = \min(u_1, u_2)$. Entonces, las cotas Frechet-Hoeffding se definen como

$$W(\mathbf{u}) \leq C(\mathbf{u}) \leq M(\mathbf{u})$$

Por ende, si $F(\mathbf{x})$ es una función cópula general:

$$F(\mathbf{x}) = \begin{cases} F(\mathbf{x}) := M(\mathbf{u}) \Rightarrow \text{comonotónica} \\ F(\mathbf{x}) := W(\mathbf{u}) \Rightarrow \text{contramonotónica} \end{cases}$$

Lo cual significa que teniendo dos variables. x_1 y x_2 (frecuencia y severidad), una cópula comonotónica muestra que las variables son estricta y positivamente dependientes. Por su parte, una cópula contramonotónica sugiere que las variables son estricta y negativaemnte dependientes.

Adicional a lo anterior, la manera en que la concordancia estadística se relaciona con dicha dependencia es ofrecida por Díaz Monroy y Morales Rivera (2012), quienes sugirieron que dicho concepto implica que los componentes de X con distribución F_X tome valores simultáneos con una variable Y con distribución F_Y , cuando ambas distribuciones pertenecen a la misma familia de distribuciones p -variantes. En términos del presente documento, habrá concordancia siempre que sea más probable que incrementos (o decrementos) en la severidad se den de manera simultánea antes incrementos (o decrementos) en la frecuencia.

Apéndice C Cópulas estimadas

Sea C una cópula con variables aleatorias uniformes $\mathbf{u} = \{u, v\}$ y $\bar{\mathbf{u}} = \{1-u, 1-v\}$ entonces, de acuerdo con Nelsen (2006):

- Cópula de Supervivencia (*Survival Copula*)

$$\hat{C}(\mathbf{u}) = u + v - 1 + C(1 - u, 1 - v) \quad (14)$$

- Cópula de independencia

$$\Pi(\mathbf{u}) = uv \quad (15)$$

Lo anterior implica que $\tau = \rho = \beta = 0$ (Edward W. Frees & Valdez, 1998).

- Propiedades generales: Con $\{u_1 \leq u_2, v_1 \leq v_2\}$

$$\begin{aligned} C(u, 0) &= 0 = C(0, v) \\ C(u, 1) &= v \text{ y } C(1, v) = v \\ C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) &\geq 0 \end{aligned}$$

Con esto dicho, las cópulas trabajadas tienen las siguientes características (Joe, 1997):

- **Frank**

$$C_{\text{Frank}} = C(\mathbf{u}; \theta) = -\theta^{-1} \log \left(\frac{[\eta - (1 - e^{-\theta u})(1 - e^{-\theta v})]}{\eta} \right) \quad (16)$$

Con $\eta = 1 - e^{-\theta}$ y $\theta \in (-\infty, \infty)$

Esta familia de cópulas tiene como característica los siguientes aspectos:

$$\begin{aligned} \lim_{\theta \rightarrow \infty} C_{\text{Frank}} &= M(\mathbf{u}) \\ \lim_{\theta \rightarrow 0} C_{\text{Frank}} &= \Pi(\mathbf{u}) \\ \lim_{\theta \rightarrow -\infty} C_{\text{Frank}} &= W(\mathbf{u}) \end{aligned}$$

- **Joe**

$$C_{\text{Joe}} = C(\mathbf{u}; \theta) = 1 - ((1 - u)^\theta + (1 - v)^\theta - ((1 - u)(1 - v))^\theta)^{1/\theta} \quad (17)$$

$$\begin{aligned} \lim_{\theta \rightarrow \infty} C_{\text{Joe}} &= M(\mathbf{u}) \\ C_{\text{Joe}}(\mathbf{u}; \theta = 1) &= \Pi(\mathbf{u}) \end{aligned}$$

- **Survival Joe:** Al aplicar sobre la ecuación 17 lo sugerido por 14, queda lo siguiente:

$$\hat{C}_{\text{SurvivalJoe}} = C(\mathbf{u}; \theta) = u + v - (u^\theta + v^\theta - (uv)^\theta)^{1/\theta} \quad (18)$$

$$\begin{aligned} \lim_{\theta \rightarrow \infty} \hat{C}_{\text{SurvivalJoe}} &= W(\mathbf{u}) \\ \hat{C}_{\text{SurvivalJoe}}(\mathbf{u}; \theta = 1) &= \Pi(\mathbf{u}) \end{aligned}$$

- **Survival BB6:** Esta pertenece a la familia de cópulas con dos parámetros.

$$\hat{C}_{\text{SurvivalBB6}} = C(\mathbf{u}; \theta_1, \theta_2) = u + v - (1 - \Phi)^{1/\theta_1} \quad (19)$$

Con

$$\Phi = \exp \left[- \left((-\log(1 - u^{\theta_1}))^{\theta_2} + (-\log(1 - v^{\theta_1}))^{\theta_2} \right)^{\frac{1}{\theta_1}} \right]$$

$$\theta_1, \theta_2 \geq 1$$

$$\lim_{\theta_1 \rightarrow \infty} \hat{C}_{\text{SurvivalBB6}} = \lim_{\theta_2 \rightarrow \infty} \hat{C}_{\text{SurvivalBB6}} = W(\mathbf{u})$$

- **Survival BB8**

$$\hat{C}_{\text{SurvivalBB8}} = C(\mathbf{u}; \theta_1, \theta_2) = u + v - 1 + \frac{1}{\theta_2} \left[1 - \left(1 - \left(\frac{\Phi}{1 - (1 - \theta_2)^{\theta_1}} \right)^{\frac{1}{\theta_1}} \right) \right] \quad (20)$$

con

$$\Phi = [1 - (1 - \theta_2(1 - u))^{\theta_1}] [1 - (1 - \theta_2(1 - v))^{\theta_1}]$$

$$\lim_{\theta_1 \rightarrow 1} \hat{C}_{\text{SurvivalBB8}} = \lim_{\theta_2 \rightarrow 0} \hat{C}_{\text{SurvivalBB8}} = \Pi(\mathbf{u})$$

Apéndice D Propiedades deseables de una medida de asociación

Sea $\delta(x_1, x_2)$ la medida de dependencia entre dos variables tal que cumpla, según Heilpern (2014), con:

- Buena definición: Invariabilidad con respecto a distribuciones marginales propias: $\Rightarrow \delta(\mathbf{x}) = \delta(C)$
- Invarianza con respecto a las permutaciones.
- Normalización: Si ambas variables son independientes, entonces $\delta(\mathbf{x}) = 0$. Si son comonotónicos, entonces $\delta(\mathbf{x}) = 1$. Si son contramonotónicos, entonces $\delta(\mathbf{x}) = -1$
- Monotonicidad
- Comportamiento bajo transformación: Las transformaciones monotónicas no le afectan.
- Continuidad.

Un aporte adicional lo hacen Embrechts y col. (1998) de la siguiente forma:

P1 Simetría $\delta(X, Y) = \delta(Y, X)$.

P2 Normalización $-1 \leq \delta(X, Y) \leq 1$.

P3 $\delta(X, Y) = 1 \iff X, Y$ son comonotónicos

$\delta(X, Y) = -1 \iff X, Y$ son contracomonotónicos

P4 Para la transformación estrictamente monotónica $T : \mathbb{R} \rightarrow \mathbb{R}$ sobre el rango de X :

$$\delta(T(X), Y) = \begin{cases} \delta(X, Y) & T \text{ creciente} \\ -\delta(X, Y) & T \text{ decreciente} \end{cases}$$

P5 $\delta(X, Y) = 0 \iff X, Y$ son independientes. Aunque ninguna medida puede cumplir P4 y P5 a la vez.

De acuerdo con lo anterior, y retomando los numerales $\{1, 2, 3, 6, 7, 8\}$ del **Teorema 3** de Embrechts y col. (1998) :

- $\rho(X, Y) = \rho(Y, X), \tau(X, Y) = \tau(Y, X)$
- Si X y Y son independientes, entonces $\rho(X, Y) = \tau(Y, X) = 0$
- $-1 \leq \rho(X, Y) = \tau(Y, X) \leq 1$

- Para cualquier transformación $T : \mathbb{R} \rightarrow \mathbb{R}$ estrictamente monotónica sobre el rango de X , entonces satisface P4.
- $\rho(X, Y) = \tau(X, Y) = 1 \iff C = M(\mathbf{u}) \iff Y = T(X)$ con T creciente
- $\rho(X, Y) = \tau(X, Y) = -1 \iff C = W(\mathbf{u}) \iff Y = T(X)$ con T decreciente

Referencias

- Card, N. (2012). *Applied Meta-Analysis for Social Sciences Research*. The Guilford Press.
- Carriere, J. F. (1997). Testing independence in bivariate distributions of claim frequencies and severities. *Insurance: Mathematics and Economics*, 21(1), 81-89. doi:[10.1016/S0167-6687\(97\)88952-4](https://doi.org/10.1016/S0167-6687(97)88952-4)
- Castaño, R. & Zambrano, A. (2006). Biased selection within the social health insurance market in Colombia. *Health Policy*, 79(2-3), 313-324. doi:[10.1016/j.healthpol.2006.01.010](https://doi.org/10.1016/j.healthpol.2006.01.010)
- Congreso de la República de Colombia. (2011). Ley 1438 del 2011.
- Cummins, D. (1991). Statistical and Financial Models of Insurance Pricing and the Insurance Firm. *The Journal of Risk and Insurance*, 58(2), 261. doi:[10.2307/253237](https://doi.org/10.2307/253237)
- Cummins, D. & Wiltbank, L. (1983). Estimating the Total Claims Distribution Using Multivariate Frequency and Severity Distributions. *The Journal of Risk and Insurance*, 50(3), 377. doi:[10.2307/252434](https://doi.org/10.2307/252434)
- Czado, C., Kastenmeier, R., Brechmann, E. C. & Min, A. (2012). A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal*, (4), 278-305. doi:[10.1080/03461238.2010.546147](https://doi.org/10.1080/03461238.2010.546147)
- Díaz Monroy, L. G. & Morales Rivera, M. A. (2012). *Estadística multivariada: inferencia y métodos*.
- Dionne, G. & Harrington, S. E. (2014). *Insurance and insurance markets*. doi:[10.1016/B978-0-444-53685-3.00005-2](https://doi.org/10.1016/B978-0-444-53685-3.00005-2)
- Duncan, I. (2011). Healthcare Risk Adjustment and Predictive Modeling.
- Embrechts, P., McNeil, A. & Straumann, D. (1998). Correlation and Dependency in Risk Management: Properties and Pitfalls. *Review of Economics and Statistics*.
- Erhardt, V. & Czado, C. (2012). Modeling dependent yearly claim totals including zero claims in private health insurance. *Scandinavian Actuarial Journal*, (2), 106-129. doi:[10.1080/03461238.2010.489762](https://doi.org/10.1080/03461238.2010.489762)
- Farhadi, H. (2016). Aggregate Loss Models.
- Frees, E. W. [Edward W.] & Valdez, E. A. (1998). Understanding Relationships Using Copulas. *North American Actuarial Journal*, 2(1), 1-25. doi:[10.1080/10920277.1998.10595667](https://doi.org/10.1080/10920277.1998.10595667)
- Frees, E. W. [Edward W], Gao, J. & Rosenberg, M. A. (2011). Predicting the Frequency and Amount of Health Care Expenditures. *North American Actuarial Journal*, 15(3), 377-392.
- Garrido, J., Genest, C. & Schulz, J. (2016). Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, 70, 205-215. doi:[10.1016/j.insmathco.2016.06.006](https://doi.org/10.1016/j.insmathco.2016.06.006)
- Gobierno Nacional. (2014). Decreto 2702 de 2014.
- Goldburd, M., Khare, A. & Tevet, D. (2008). *Generalized linear models for insurance data*. doi:[10.1017/CBO9780511755408](https://doi.org/10.1017/CBO9780511755408)
- Gonzalo, A. (2015). Claves para un aseguramiento en salud exitoso. *Conexión*, 24-31.
- Grimmett, G. & Welsh, D. (2014). *Generalized linear models for insurance data*.

- Gschlöbl, S. & Czado, C. (2007). Spatial modelling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal*, 2007(3), 202-225. doi:[10.1080/03461230701414764](https://doi.org/10.1080/03461230701414764)
- Guerrero, R., Gallego, A. I., Becerril-Montekio, V. & Vázquez, J. (2011). Sistema de salud de Colombia. *Salud Pública de México*, 53(2), 144-155.
- Heilpern, S. (2014). Multivariate Measures of Dependence based on Copulas. *Mathematical Economics*, 10(17).
- Hofert, M., Kojadinovic, I., Machler, M. & Yan, J. (2018). *Elements of Copula Modeling with R*.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*.
- Jorgensen, B. (1987). Exponential Dispersion Models. *Journal of the Royal Statistical Society*, 49(2), 127-162.
- Kholifah, A. R. U., Lestari, D. & Devila, S. (2019). Premium calculation using marginal generalized linear model combined with copula. *AIP Conference Proceedings*, 2168(November). doi:[10.1063/1.5132462](https://doi.org/10.1063/1.5132462)
- Klugman, S. A., Panjer, H. H. & Willmot, G. E. (2012). *Loss models: from data to decisions*, 3rd ed.
- Krämer, N., Brechmann, E. C., Silvestrini, D. & Czado, C. (2013). Total loss estimation using copula-based regression models. *Insurance: Mathematics and Economics*, 53(3), 829-839. doi:[10.1016/j.insmatheco.2013.09.003](https://doi.org/10.1016/j.insmatheco.2013.09.003). arXiv: [1209.5356](https://arxiv.org/abs/1209.5356)
- Kumar, P. (2010). Copulas: Distribution Functions and Simulation. En M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 309-312). doi:[10.1007/978-3-642-04898-2_191](https://doi.org/10.1007/978-3-642-04898-2_191)
- Lee, G. Y. & Shi, P. (2019). A dependent frequency–severity approach to modeling longitudinal insurance claims. *Insurance: Mathematics and Economics*, 87, 115-129. doi:[10.1016/j.insmatheco.2019.04.004](https://doi.org/10.1016/j.insmatheco.2019.04.004)
- Lee, W., Park, S. C. & Ahn, J. Y. (2018). Investigating dependence between frequency and severity via simple generalized linear models. *Journal of the Korean Statistical Society*, 48(1), 13-28. doi:[10.1016/j.jkss.2018.07.003](https://doi.org/10.1016/j.jkss.2018.07.003)
- Londoño, J. L. & Frenk, J. (1997). Structured pluralism: Towards an innovative model for health system reform in Latin America. *Health Policy*, 41(1), 1-36. doi:[10.1016/S0168-8510\(97\)00010-9](https://doi.org/10.1016/S0168-8510(97)00010-9)
- Marcinko, D. E. & Hetico, H. R. (2006). Dictionary of Health Insurance and Managed Care, 372.
- Modarres, R. (2010). Measures of Dependence. En M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 804-805).
- Nagler, T., Schepsmeier, U., Brechmann, E. C., Graeler, B., Erhardt, T., Almeida, C., ... Vatter, T. (2019). Statistical Inference of Vine Copulas.
- Nelsen, R. (2006). *An Introduction to Copulas*. doi:<https://doi.org/10.1007/0-387-28678-0>
- Paredes Cubillos, N. (2015). ¿ES NECESARIO REPENSAR EL CÁLCULO DE LA UPC? *Conexión2*, 12-17.
- Pitacco, E. (2014). *Health Insurance*. doi:[10.1007/978-3-319-12235-9](https://doi.org/10.1007/978-3-319-12235-9)

- Puka, L. (2010). Kendall's Tau. En M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 713-715).
- Rémillard, B. (2010). Tests of Independence. En M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 1598-1601).
- Rodriguez-Monguió, R. & Infante, A. (2004). Universal health care for Colombians 10 years after Law 100: Challenges and opportunities. *Health Policy*, 68(2), 129-142. doi:[10.1016/j.healthpol.2003.10.004](https://doi.org/10.1016/j.healthpol.2003.10.004)
- Shi, P., Feng, X. & Ivantsova, A. (2015). Dependent frequency-severity modeling of insurance claims. *Insurance: Mathematics and Economics*, 64, 417-428. doi:[10.1016/j.insmatheco.2015.07.006](https://doi.org/10.1016/j.insmatheco.2015.07.006)
- Shi, P. & Guszczka, J. (2016). Frameworks for General Insurance Ratemaking: Beyond the Generalized Linear Model. En *Predictive modeling applications in actuarial science volume ii: Case studies in insurance* (pp. 100-124).
- Superintendencia Nacional de Salud. (2014). Resolución 4175 de 29 de diciembre de 2014.
- Trivedi, P. K. & Zimmer, D. M. (2005). *Copula modeling: An introduction for practitioners*. doi:[10.1561/0800000005](https://doi.org/10.1561/0800000005)
- Tse, Y.-K. (2009). *Nonlife Actuarial Models Theory, Methods and Evaluation*.
- Wermuth, N. & Cox, D. R. (2014). Statistical Dependence and Independence. *Wiley StatsRef: Statistics Reference Online*, (1980), 4260-4264. doi:[10.1002/9781118445112.stat05968](https://doi.org/10.1002/9781118445112.stat05968)
- Werner, G., Modlin, C., Gannon, A., Guven, S., Gennett, C., Kucera, J., . . . Otto, D. (2016). Basic Ratemaking. (October), 1-423.
- Zimmer, D. M. & Trivedi, P. K. (2006). Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *Journal of Business and Economic Statistics*, 24(1), 63-76. doi:[10.1198/073500105000000153](https://doi.org/10.1198/073500105000000153)