

UNIVERSIDAD DE COSTA RICA

ESCUELA DE MATEMÁTICA

DEPARTAMENTO DE MATEMÁTICA PURA Y CIENCIAS ACTUARIALES
DISTRIBUCIÓN DE PÉRDIDAS

Proyecto Distribución de Pérdidas

BITÁCORAS

Grupo 05

Realizado por

Daniel Núñez - B85667

Andrés González - B83413



UNIVERSIDAD DE
COSTA RICA

EMat

Escuela de
Matemática

Índice general

Índice general	I
1 Bitácora 1	1
1. Planificación	1
Idea principal	1
Base de datos	2
Principios y/o teorías	3
Fichas bibliográficas	4
2. Escritura	6
2 Anexos	7
1. Uve de Gowin	7
Referencias	8

Bitácora 1

1. Planificación

Idea principal

La idea inicial que se planteó va de la mano con calcular la pérdida esperada de un Banco a la hora de otorgar créditos basado en el perfil de la persona. Bajo este contexto se pueden encontrar dos tipos de riesgos:

- 1) El primer riesgo que se distingue es aquel en el que la persona presente una probabilidad alta de que no pague el crédito y que aún así el banco lo apruebe.
- 2) El segundo haciendo alusión al riesgo de que una persona tenga una probabilidad de pagar de vuelta el crédito pero el banco lo niegue, lo que resulta en una pérdida para la entidad bancaria.

Cuatro formas diferentes de reenplantearse la idea anterior son:

- 1) ¿Cuál sería la pérdida esperada de un banco a la hora de determinar la elegibilidad de una persona para un crédito?
- 2) ¿Por qué es importante conocer las probabilidades de pago o de impago de una persona solicitante?
- 3) ¿Cuáles son los factores que más importan para determinar la probabilidad de impago de un individuo?
- 4) ¿Cómo se relacionan las pérdidas con un buen sistema de elegibilidad para créditos?

Y las posibles respuestas de estas preguntas serían:

- 1) La pérdida esperada del banco del banco se puede estimar utilizando los datos de una persona y con esto se elige para crédito o no (Lógica). Para esto se recopilan datos de individuos optando por crédito de manera confidencial (Ética). Lo que facilitará al banco la toma de decisión de otorgar un crédito minimizando sus pérdidas (Emocional).
- 2) Es importante conocer las probabilidades de pago o impago de una persona pues esto determina si el banco otorga un crédito o no (Lógica). Sin embargo se promueve realizar un estudio aparte para cada individuo pues la decisión final de otorgar el crédito debería ser dado por un experto en el tema y no solo en los resultados del modelo (Ética). Esto le sirve al banco o institución financiera para conocer clientes potenciales para obtener ganancias (Emocional).

- 3) Los factores que más afectan la probabilidad de impago de un individuo son los determinados por el modelo realizado (Lógica). Esto es de gran utilidad para el banco pues se pueden realizar campañas especializadas en la obtención de clientes que cumplan dichos factores (Emocional). Asimismo se recomienda tomar estos resultados como ayuda a la toma de decisiones no como algo definitivo (Ética).
- 4) Las pérdidas esperadas por impago de créditos están directamente relacionadas con un sistema robusto de elegibilidad (Lógica), pues si se otorgan créditos a personas con alta probabilidad de pago y se niegan a los que no, minimizaría las pérdidas por impago de la institución (Emocional). Siempre siguiendo los estatutos y leyes para el otorgamiento y denegación de créditos (Ética).

Base de datos

Para efectos de la investigación se va a trabajar con una base de datos de un banco alemán que contiene información de personas solicitantes de un crédito y con base en esta información se determina su elegibilidad. Los datos fueron recuperados de kaggle y originalmente fueron obtenidos de Penn State Eberly College of Science. Los datos son públicos y son de libre acceso, sin embargo, por motivos de confidencialidad el nombre del banco nunca se menciona. Vale la pena mencionar que los datos no están delimitados temporalmente.

La población de estudio se define como las personas que solicitaron un crédito en esta entidad bancaria ubicada en Alemania mientras que la unidad estadística se define como la persona solicitante de un crédito en el banco alemán estudiado. La muestra para el desarrollo de dicha investigación consta de 1000 individuos. Asimismo, la base cuenta con 21 variables de interés donde en su columna matriz se encuentra la variable binaria de elegibilidad, que toma el valor de 1 si fue elegible y el valor de 0 si no lo fue.

Ahora se procede a dar una leve explicación de cada una de las variables que conforman esta base:

- Elegibilidad: toma el valor de 1 si fue elegible y el valor de 0 si no lo fue.
- Account Balance: variable categórica que toma el valor de 1 si la persona no cuenta con ninguna cuenta en el banco, el valor de 2 si no tiene un balance pendiente con el banco y el valor de 3 si sí tiene un balance pendiente.
- Duration: la duración en meses del crédito solicitado
- Payment Status of Previous Credit: variable categórica que toma el valor de 1 si el individuo presenta problema con el pago del crédito anterior, el valor de 2 si ya lo pagó y el valor de 3 si no tiene problemas con el crédito anterior
- Purpose: variable categórica que toma el valor de 1 si es para un auto nuevo, 2 si es para un auto de segunda mano, 3 si es para una casa/apartamento y 4 si es para cualquier otra cosa.
- Credit Amount: el monto del crédito solicitado en “Deutsche Mark” (DM), que es la unidad monetaria usada en la base.
- Saving/Stock value: toma el valor de 1 si no tiene nada de ahorros o de stock, de 2 si el valor es menor a los 100 DM, 3 si se está en el intervalo [100, 500[DM , 4 si está en [500, 1000[DM y 5 si está arriba de los 1000 DM.
- Length of current employment: variable categórica que toma el valor de 1 si es desempleado, de 2 si tiene menos de año, de 3 si es de 1 a 4 años, de 4 si es de 4 a 7 años y de 5 si es mayor a 7 años.
- Sex/marital status: toma el valor de 1 si es un hombre divorciado, el de 2 si es hombre soltero, el de 3 si es hombre casado/viudo y el de 4 si es mujer

- **Guarantors:** variable binaria que toma el valor de 1 si la persona tiene un fiador y de 0 si no lo tiene.
- **Duration in current address:** variable categórica que determina cuánto tiempo lleva la persona viviendo en la última dirección registrada. Toma el valor de 1 si es menos de un año, el de 2 si lleva entre uno y 4 años, el de 3 si lleva entre 4 y 7 años y el de 4 si lleva más de 7 años.
- **Most valuable asset:** toma el valor de 1 si no tiene ninguno, el de 2 si es un carro, el de 3 si es un seguro de vida y el de 4 si son bienes raíces.
- **Edad:** edad en años
- **Tarjetas de créditos:** toma el valor de 1 si el aplicante tiene tarjetas con otros bancos, el valor de 2 si tiene tarjetas de créditos con empresas y el de 3 si no tiene nada.
- **Type of department:** toma el valor de 1 si no paga renta/hipoteca, el de 2 en caso de pague renta o hipoteca y el de 3 si es dueño de la vivienda/apartamento.
- **No. of credits at this bank:** toma el valor de 1 si solo tiene 1, el de 2 si tiene entre 2 y 3 créditos, el de 3 si tiene entre 4 y 5 créditos y el de 4 tiene más de 6 créditos.
- **Occupation:** 1 en caso de que sea desempleado o no calificado, 2 en caso de que sea un residente permanente no calificado, 3 en caso de que sea una persona calificada y 4 en caso de que sea un ejecutivo/a.
- **No. of dependents:** número de personas que mantiene. Toma el valor de 1 si son más de 3 y de 2 si son entre 0 y 2 personas.
- **Foreign worker:** variable binaria que toma el valor de 0 en caso de que sea un trabajador extranjero y 1 en caso de que no lo sea.

Como se pudo observar, la base cuenta con una cantidad considerable de variables de interés, donde leyendo la descripción de cada una se puede entender e intuir el posible impacto que tengan en la elegibilidad de las personas. Sin embargo, se tratará de reducir la cantidad de variables haciendo un análisis exploratorio de los datos de tal manera que se pueda eliminar variables que estén correlacionados entre sí. Para llevar a cabo dicho proceso, se utilizará el modelo de cópulas para eliminar o unificar este tipo de variables y se utilizarán modelos predictivos como la regresión logística y Random Forest para determinar la elegibilidad de las personas.

Principios y/o teorías

Para el desarrollo de esta investigación se planea utilizar el modelo de cópulas para relacionar los riesgos mencionados ya que se está buscando correlacionar las pérdidas asociadas a ambos. Para esto resulta útil implementar un modelo de cópulas puesto que como mencionan Escarela y Hernández, estos modelos ayudan a obtener una distribución conjunta a partir de varias funciones de distribución asociadas a variables aleatorias con cierta relación. Es decir, con este método se construye una distribución multivariada a partir de las distribuciones univariadas de las variables respectivas. Este tipo de modelo también resulta en una forma de estructurar la dependencia de estas parejas de variables aleatorias en distribuciones conjuntas (Escalera & Hernández, 2009).

Como este trabajo se basa en estimar las pérdidas que puede incurrir la institución financiera considerando los riesgos que conlleva el impago de créditos y así determinar la elegibilidad de posibles clientes, se va a recurrir a la teoría de valores extremos. Esta teoría tiene muchas aplicaciones tanto en ámbitos financieros como en otros, puesto que generalmente se utiliza para modelar eventos catastróficos, con pocas ocurrencias como pueden ser desastres naturales y con más relación a seguros,

estos se utilizan para modelar riesgos que no son tan frecuentes y que si suceden, representan una suma considerable de dinero. Este tipo de teoría se basa en estudiar las distribuciones después de un cierto valor o umbral establecido, por lo que generalmente se basa más en el estudio de las colas de la distribución y las pérdidas que se pueden incurrir por montos que exceden dicho umbral (Smith, 2009).

Fichas bibliográficas

- **Título:** The Grouped t-copula with an Application to Credit Risk
 - **Autor(es):** Stéphane Daul, Enrico De Giorgi, Filip Lindskog & Alexander McNeil
 - **Año:** 2003
 - **Nombre del tema:** Cópulas
 - **Forma de organizarlo:**
 - Cronológico:** 2003
 - Metodológico:** Grouped t-Copulas
 - Temático:** Métodos para explicar la dependencia de covariables
 - Teoría:** Uso de cópulas en el riesgo crediticio
 - **Resumen en una oración:**
 - **Argumento central:** Lo esencial de este artículo es la justificación del uso en particular de la t-Cópulas agrupadas en problemas de riesgo crediticio.
 - **Problemas con el tema:** Los parámetros de los grados de libertad calculados por lo autores fueron considerablemente altos, sin embargo, aún así siguen dando incrementos importantes en el aumento del riesgo comparado con el modelo con cópula Gaussiana.
 - **Resumen en un párrafo:** Se extiende el concepto de t-Cópula para generar una nueva t-Cópula agrupada que describe de manera más efectiva la dependencia que existe entre factores de riesgo en el ámbito financiero. En el artículo también se explica el proceso de cálculo de los nuevos parámetros de las t-Cópulas agrupadas. Básicamente la idea que hay de fondo es juntar los diferentes factores de riesgo en esta t-Cópula agrupada donde las variables aleatorias dentro de esta cópula tienen una t-cópula con diferentes grados de libertad. Como resultado, esto da una estructura de dependencia más flexible que se ajusta mejor para modelos de riesgos que tengan múltiples factores. A modo de conclusión, utilizan el value-at-risk para comparar el modelo usando t-Cópulas con el modelo de t-Cópulas agrupadas y este segundo siempre arroja resultados con medidas de riesgo mayores, por lo que se tiene un modelo más conservador y preciso.
-
- **Título:** Application of Vine Copulas to Credit Portfolio Risk Modeling
 - **Autor(es):** Marco Geidosch & Matthias Fischer
 - **Año:** 2016

- **Nombre del tema:** Cópulas
- **Forma de organizarlo:**
 - Cronológico:** 2015
 - Metodológico:** Vine Copulas
 - Temático:** Métodos para explicar la dependencia de covariables
 - Teoría:** Uso de cópulas para modelar problemas relacionados con riesgo
- **Resumen en una oración:** Dado lo poco flexibles que pueden llegar a ser los modelos convencionales de cópulas se propone el uso de Vine Cópulas para modelar problemas multivariados relacionados con riesgo
- **Argumento central:** Mostrar la superioridad de las Vine Cópulas sobre las cópulas convencionales para la modelación del riesgo
- **Problemas con el tema:** La flexibilidad que presentan las Vine Cópulas viene con la desventaja de que hay una abundante número de estructuras de las que se puede escoger y a priori no se sabe cuál es la que mejor describe los datos.
- **Resumen en un párrafo:** El uso de cópulas convencionales presenta dos limitantes, la primera es que cuando el número de dimensiones es mayor que dos, el número de familias de cópulas aplicables se limita al caso elíptico y arquimediano. La segunda limitante es que solo dependencias simétricas y simples pueden ser modeladas y eso está lejos de ser considerado un escenario real. De ahí nace la necesidad de usar Vine Cópulas. Las Vines Cópulas se usan para describir cópulas de varias variables usando como base cópulas bivariadas o "pair-copulas". La idea que existe de fondo con los Vines Cópulas es que distribuciones multivariadas se pueden descomponer secuencialmente en descomposiciones bivariadas y para llevar acabo dicho proceso se hace uso de los D-vine y los R-vine. Lo que se hace es tomar la densidad multivariada de la distribución y se transforma en funciones de densidad de las cópulas bivariadas.

-
- **Título:** Extreme value theory as a risk management tool
 - **Autor(es):** Paul Embrechts, Sidney I. Resnick y Gennady Samorodnitsky
 - **Año:** 1999
 - **Nombre del tema:** Teoría de valores extremos
 - **Forma de organizarlo:**
 - Cronológico:** 1999
 - Metodológico:** Distribuciones con valores extremos
 - Temático:** Teoría de valores extremos y aplicaciones
 - Teoría:** Teoría de valores extremos
 - **Resumen en una oración:** Teoría de valores extremos con su fundamento teórico y las distribuciones usuales para esta metodología junto con sus aplicaciones.
 - **Argumento central:** Base teórica y posibles aplicaciones a la teoría de valores extremos en manejo de riesgo.

- **Problemas con el tema:** El autor no menciona problemas.
- **Resumen en un párrafo:** Se presentan ejemplos de como eventos catastróficos costosos de predecir como terremotos, incendios, entre otros, que supusieron un gasto enorme no presupuestado por parte de las compañías aseguradoras. Es por esto que se empezaron a realizar estudios y métodos de prever escenarios como estos para estar preparado a asumir estas pérdidas, como el VaR. Estos eventos extremos en el mundo de las finanzas se pueden ver como pérdidas masivas en industria o caídas considerables en mercados bursátiles, para los cuales la teoría de valores extremos provee metodologías robustas y con fundamento para cuantificar estos eventos y sus consecuencias. Es por esto que se explican diferentes herramientas que provee la teoría de valores extremos en el ámbito del manejo de riesgo.

2. Escritura

La pregunta a desarrollar durante la investigación es, ¿Cuál sería la pérdida esperada de un banco a la hora de determinar la elegibilidad de una persona para un crédito? por lo que se planea primeramente realizar un modelo de clasificación para determinar si una persona es elegible a crédito o no. Tomando en cuenta los dos riesgos asociados de una mala clasificación. Una vez con esto, y con los datos escogidos realizar un modelo de estimación de pérdidas combinando ambos riesgos utilizando un modelo basado en cópulas y a su vez realizar un análisis tomando en cuenta la teoría de valores extremos para aquellos riesgos que supongan una pérdida mayor a un cierto umbral.

En cuanto a lo estudiado hasta el momento, se tiene una base de datos con información sobre la elegibilidad de personas a un crédito bancario. Lo primero que se piensa implementar es un modelo de clasificación como regresión logística o bosques aleatorios (Random forests) y de esta manera asociar el riesgo de clasificar un individuo como no elegible cuando este si debería ser elegible y viceversa. Esto debido a que los riesgos mencionados representarían pérdidas al banco, por lo que para encontrar una distribución de pérdidas que contemple ambos se debe de alguna manera encontrar una distribución conjunta.

Para resolver este problema de encontrar la distribución conjunta se ha investigado sobre los modelos de cópulas que ayudan a correlacionar estas distribuciones individuales univariadas en una distribución multivariada, junto con las relaciones y dependencias de estas variables entre sí. Sin embargo, la mayoría de distribuciones de cópulas están hechas para tomar en cuenta solo dos variables y se menciona que para más variables estas distribuciones presentan ciertas limitaciones. Es por esto que se ha indagado en otro tipos de modelos como el de vine-copulas que funciona para más de dos variables. Sin embargo, no es la única metodología de cópulas estudiadas, puesto que también se investigó los modelos de t-copulas y t-copulas agrupadas que permiten más flexibilidad a la hora de estructurar la dependencia entre las variables. Aun así, todavía no se tiene cual de los modelos implementar aunque se está considerando realizar varias para poder realizar una comparación con los resultados obtenidos por los distintos modelos y tener conclusiones más completas.

Como ya se mencionó, se planean realizar modelos para encontrar la distribución de pérdidas del banco conjunta, pero el otro tema a considerar es como ajustar esta distribución a posibles pérdidas altas que se pueden dar por otorgar un crédito a individuos que no van a incurrir en el pago. Es por esto, que se planea utilizar la teoría de valores extremos y distintas metodologías de esta teoría para poder realizar el análisis de las pérdidas esperadas en estos casos extremos. Por lo tanto este trabajo se piensa estructurar en tres grandes pasos, el primero el de realizar un modelo de elegibilidad de crédito, después encontrar una distribución de pérdidas conjunta para el banco y finalmente complementar esta estimación de pérdidas en caso de valores extremos.

Anexos

1. Uve de Gowin

Teorias/Principios/Metodologias:

Modelo de cópulas puesto que estos modelos ayudan a obtener una distribución conjunta a partir de varias funciones de distribución asociadas a variables aleatorias con cierta relación.

Teoría de valores extremos, el cual se basa en estudiar las distribuciones después de un cierto valor o umbral establecido

Conceptos:

- **Crédito:** operación de financiación donde un 'acreedor', presta una cierta cifra monetaria a un 'deudor', quien garantiza al acreedor que retornará esta cantidad solicitada más una cantidad adicional, llamada 'intereses'.
- **Pérdida esperada:** valor esperado de pérdida por riesgo crediticio en un horizonte de tiempo determinado
- **Elegibilidad:** Cualidad de la persona que puede ser elegida para algo

¿Cuál sería la pérdida esperada de un banco a la hora de determinar la elegibilidad de una persona para un crédito?

Pregunta de Investigación

Afirmaciones y resultados:

Registros: Los datos fueron extraídos de Kaggle y la base de datos consta de 1000 observaciones y 21 variables diferentes donde cada fila describe las características de elegibilidad de una persona

El objeto de estudio serán los diferentes factores que influyen sobre la elegibilidad de una persona para un crédito en un banco alemán

Referencias

- Cediel, Y. F. (2016). Regular vine cópulas: Una aplicación al cálculo de valor al riesgo. *Revista de Investigación En Modelos Financieros*, 2, 30–64.
- Daul, S., De Giorgi, E. G., Lindskog, F., & McNeil, A. (2003). The grouped t-copula with an application to credit risk. *Available at SSRN 1358956*.
- Embrechts, P., Resnick, S. I., & Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2), 30–41.
- Escarela, G., & Hernández, A. (2009). Modelado de parejas aleatorias usando cópulas. *Revista Colombiana de Estadística*, 32(1), 33–58.
- Geidosch, M., & Fischer, M. (2016). Application of vine copulas to credit portfolio risk modeling. *Journal of Risk and Financial Management*, 9(2), 4.
- Smith, R. (2009). Extreme value theory. *Department of Statistics and Operations Research, University of North Carolina*.