

# PROYECTO FINAL DE ESTADISTICA MATEMATICA

Armando Lara, Dario Quishpe , Jorge Arguello

2024-02-29

## Contents

<b>1</b>	<b>Introducción</b>	<b>1</b>
<b>2</b>	<b>Objetivos</b>	<b>1</b>
<b>3</b>	<b>Metodología</b>	<b>1</b>
3.1	Recolección de los datos . . . . .	1
3.2	Construcción del Indicador de solvencia . . . . .	2
3.3	Algoritmo Bootstrap . . . . .	2
3.4	Estimación de parámetros mediante el Algoritmo EM . . . . .	3
3.5	Prueba de distribución de mixturas normales . . . . .	3
3.6	Simulación de datos mediante los parámetros obtenidos . . . . .	3
3.7	Comparación con Montecarlo y Bootstrap . . . . .	3
<b>4</b>	<b>Experimentación</b>	<b>4</b>
<b>5</b>	<b>Conslusiones y recomendaciones</b>	<b>7</b>
<b>6</b>	<b>Anexos</b>	<b>7</b>
<b>7</b>	<b>Referencias</b>	<b>7</b>

## 1 Introducción

Para una coperativa de ahorro y crédito es fundamental el análisis de La solvencia financiera de sus socios pues es un factor crítico al evaluar las distintas solicitudes de créditos permitiendo tomar decisiones sobre las mismas.Estás decisiones tienen impacto directo en los diferentes indicadores que son de gran relevancia en la institución para su monitoreo y supervisión. Dentro de estos se encuentran indicadores de mora por productos crediticios, porcentaje de colocación,riesgo crediticio ,etc.Está medida de solvencia nos proporciona una visión de la capacidad que tienen los socios para cumplir con sus obligaciones financieras, ya que en caso de un indicador de solvencia saludable sugiere una menor probabilidad de incumplimiento en el pago

de deudas, así como también permite a la cooperativa tomar decisiones de crédito informadas y personalizadas, dando oportunidad de adaptar los términos del crédito u ofrecer productos convenientes según la solvencia individual.

La cooperativa de ahorro y crédito denominada “X”, en vista de la importancia de la información otorgada por este indicador, optó por hacer el acompañamiento de técnicas estadísticas a las decisiones de los analistas sobre los créditos. El presente proyecto intenta poner como un punto de partida el estudio de este indicador en un conjunto de socios.

## 2 Objetivos

- Generar estimaciones mediante la aplicación de las técnicas de remuestreo para realizar inferencia sobre el indicador de solvencia, con la finalidad de estimar el sesgo, la media, la varianza, un intervalo de confianza, así como realizar un contraste de una hipótesis de acuerdo a valores considerados por el jefe del área según su experiencia.
- Realizar comparaciones entre los valores estimados obtenidos del modelo paramétrico (Montecarlo) y del modelo no paramétrico (Bootstrap)
- Obtener conclusiones que permitan tomar decisiones mediante los valores estimados con las técnicas implementadas y compararlas con los valores considerados por el jefe del área resultantes de su experiencia.

## 3 Metodología

### 3.1 Recolección de los datos

Mediante la colaboración del jefe del área de fábrica de crédito de la cooperativa “X” se obtuvo una base de datos de solicitudes de crédito con **660 registros** con corte al mes de **Agosto del 2023**, la cual contiene **35 columnas** con información sobre variables relevantes, dentro de estas las que son de nuestro principal interés son las siguientes

- Total Patrimonio Neto: Esta ofrece la imagen de la salud financiera de una persona, es un resumen de lo que se posee (bienes), menos lo que se debe a otros (pasivos).
- Activos: Un activo es una propiedad o capital propiedad de una persona o compañía que tiene un valor económico

Por cuestiones de confidencialidad de los clientes de dicha cooperativa, se omitió las variables con respecto a información personal, tomando únicamente para nuestro interés en este trabajo sólo las dos variables para generar el indicador.

### 3.2 Construcción del Indicador de solvencia

Dentro de una cooperativa, es importante la representación de la capacidad financiera mediante un indicador, el patrimonio neto representa la cantidad de flujo con la que está a disposición la cooperativa para hacer frente a situaciones futuras. Por lo tanto, dividir el patrimonio neto por los activos totales proporciona una medida de la capacidad de un socio para cubrir sus obligaciones con sus activos.

Por lo tanto, se construye el indicador de solvencia como:

$$I = \text{Patrimonio neto} / \text{Activos}$$

### 3.3 Algoritmo Bootstrap

En general, para el remuestreo bootstrap, seguiremos estos pasos:

1. Para cada  $i = 1, \dots, n$ , generar  $L_i^*$  a partir de  $F_n$ .
2. Obtener  $L^* = (L_1^*, \dots, L_n^*)$ .
3. Repetir  $B$  veces los pasos 1 y 2 para obtener réplicas  $L^*(1), \dots, L^*(B)$ .
4. Usar estas réplicas bootstrap para aproximar la distribución de remuestreo de  $R$ .

Ahora, si consideramos  $\hat{\theta} = T(L)$ , consideramos  $F$  la distribución conocida y definimos el estadístico

$$R(L, F) = \hat{\theta} - \theta$$

Vamos a aproximar el sesgo y la varianza:

$$\begin{aligned} \text{Sesgo}(\hat{\theta}) &= E(\hat{\theta} - \theta) = E(R) \\ \text{Var}(\theta) &= \text{Var}(\hat{\theta} - \theta) \end{aligned}$$

1. Para cada  $i = 1, \dots, n$ , generar  $L_i^*$  a partir de  $\hat{F}$  y obtener  $L^* = (L_1^*, \dots, L_n^*)$ .
2. Calcular  $R^* = R(L^*, F) = \hat{\theta}^* - \hat{\theta}$ .
3. Repetir  $B$  veces los pasos 1 y 2 para obtener las réplicas  $R^*(1), \dots, R^*(B)$ .
4. Usar las réplicas bootstrap para aproximar las características de interés:

$$\begin{aligned} \text{Sesgo}^*(\hat{\theta}^*) &= \frac{1}{B} \sum_{b=1}^B R^*(b) \\ \text{Var}^*(\hat{\theta}^*) &= \frac{1}{B} \sum_{b=1}^B (R^*(b) - \bar{R}^*)^2 \end{aligned}$$

### 3.4 Estimación de parámetros mediante el Algoritmo EM

Mediante el algoritmo EM Expectation-Maximization, se procederá a estimar los parámetros del indicador de solvencia de la supoblación. Este algoritmo permitirá identificar los parámetros óptimos de la distribución (en un principio desconocida) del indicador de solvencia. Este algoritmo nos será de gran utilidad, pues una vez estimados los parámetros se los usará para hacer inferencia e identificar su distribución.

### 3.5 Prueba de distribución de mixturas normales

Una vez que se obtenga los parámetros como la media, la varianza y los pesos se aplicará un test estadístico apropiado para comprobar si la distribución del indicador es una Mixtura de distribuciones normales, dicho test fue realizado por Priscila Guayasamín para la clasificación de cooperativas por segmentos [1]. En el caso en que el test rechace la hipótesis de mixturas de distribuciones normales, se procederá a realizar una transformación de nuestros datos para lograr adaptarlos lo mejor posible al desarrollo del proyecto.

### 3.6 Simulación de datos mediante los parámetros obtenidos

Si mediante el test adaptado en [1] se obtiene que no hay suficiente evidencia estadística para rechazar el hecho de que la distribución del indicador es una mixtura de normales, se procederá a realizar la simulación de nuevos datos obtenidos de dichas simulaciones con el fin de poder construir el indicador de solvencia, en esta simulación se incluirá los parámetros obtenidos por el algoritmo EM y con una distribución de mixtura de normales. Esto servirá para realizar las estimaciones intervalos de confianza del indicador así como contrastes de hipótesis.

### 3.7 Comparación con Montecarlo y Bootstrap

Una vez obtenidas las estimaciones de los parámetros mediante el algoritmo EM y la distribución de mixturas normales, se procede a comparar los resultados utilizando técnicas de Montecarlo y Bootstrap. Esta comparación abarca tanto las estimaciones de los parámetros como sus intervalos de confianza al 95%, así como el sesgo y el contraste de hipótesis para la media (proporciones de rechazo).

Para aplicar Montecarlo en la estimación de parámetros se realiza lo siguiente:

**Generación de datos aleatorios:** Generamos un conjunto de datos aleatorios para las variables patrimonio neto y activos. En nuestro caso una mixtura de normales, con los parámetros que se obtuvo del algoritmo EM.

**Muestras:** Realizamos múltiples muestras aleatorias de tamaño  $n$  de estas variables generado en el paso anterior. Estas muestras deben tomarse con reemplazo, es decir, cada observación puede ser seleccionada más de una vez en la muestra.

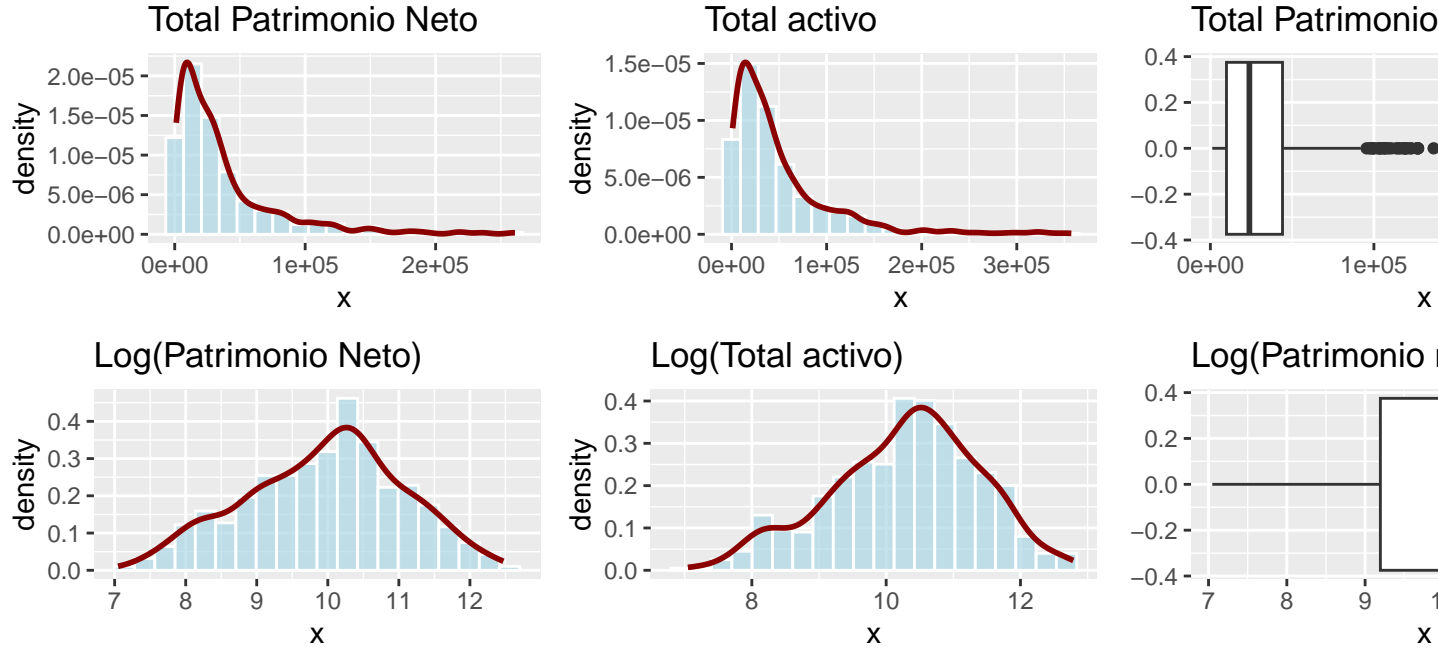
**Construcción del indicador:** Luego, para cada muestra generada de las variables, construimos un indicador y de ellas guardamos sus medias y varianzas obtenidas en un vector.

**Cálculo de estimaciones finales:** Calculamos la media de los estimadores obtenidos en el indicador de cada iteración, esto se hace promediando las medias y varianzas de las muestras generadas para el indicador anteriores.

Finalmente, comparamos los resultados obtenidos con cada metodología y su respectiva discusión, detallando así las ventajas y desventajas obtenidas con cada una.

## 4 Experimentación

En primer lugar se requiere ajustar el indicador de solvencia muestral a una distribución multivariante. Para ello se realiza el test AD modificado en [1] obteniendo para los datos originales obteniendo un **valor-p** de  $5e+7$ , prácticamente cero, por lo cual rechazamos la hipótesis nula es decir los datos no provienen de una distribución de mixturas normales. En vista de esto se generan varias transformaciones sobre las variables que se requieren para generar el indicador con lo cual se llegó a la conclusión de que la más conveniente y óptimo para nuestros datos es una transformación **Logaritmo** como se ve a continuación



Realizando esta transformación obtenemos un **valor-p** de 0.4748 lo cual nos indica que no hay suficiente evidencia estadística para rechazar la hipótesis nula, por tanto los datos provienen de una distribución de mixturas normales.

Ya con esto se procede a ejecutar el algoritmo EM con el cual se encuentra los parámetros de la distribución normal multivariante. Las medias y matrices de varianza y covarianza de la subpoblación son las siguientes

## number of iterations= 80

$$\mu_1 = [9.850069; 10.47885]$$

$$\mu_2 = [10.032998; 10.12234]$$

$$\Sigma_1 = \begin{bmatrix} 1.025812 & 0.8741010 \\ 0.874101 & 0.8984864 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} 1.352622 & 1.3783079 \\ 1.378308 & 1.4139173 \end{bmatrix}$$

y los pesos correspondientes son

$$\lambda_1 = 0.4434142; \lambda_2 = 0.5565858$$

los resultados de este ajuste lo podemos ver a continuación

También podemos observar la gráfica correspondiente a los datos originales transformados

Del procedimiento anterior encontramos una distribución exacta (mixtura de normales multivariante) que describe la distribución entre el patrimonio neto transformado y los activos transformados. Ahora mediante Montecarlo simulamos los datos requeridos para la construcción del indicador de solvencia.

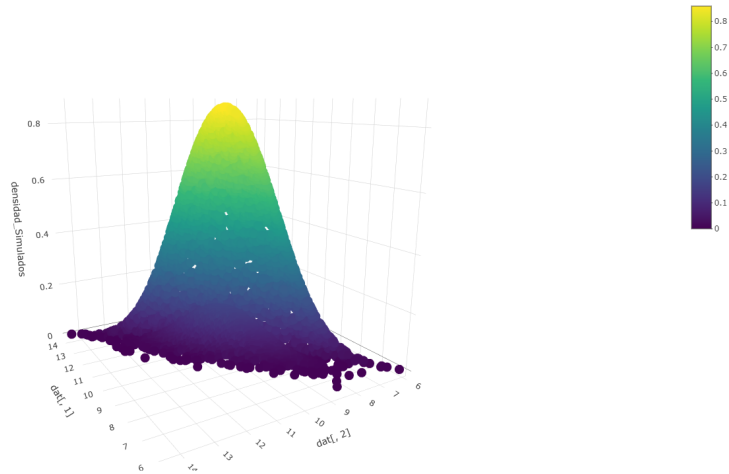


Figure 1: Densidad Simulados

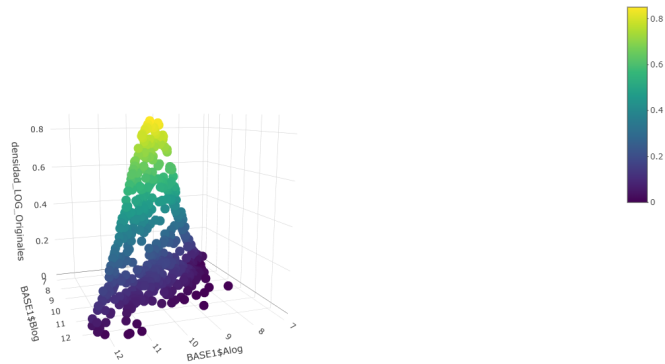
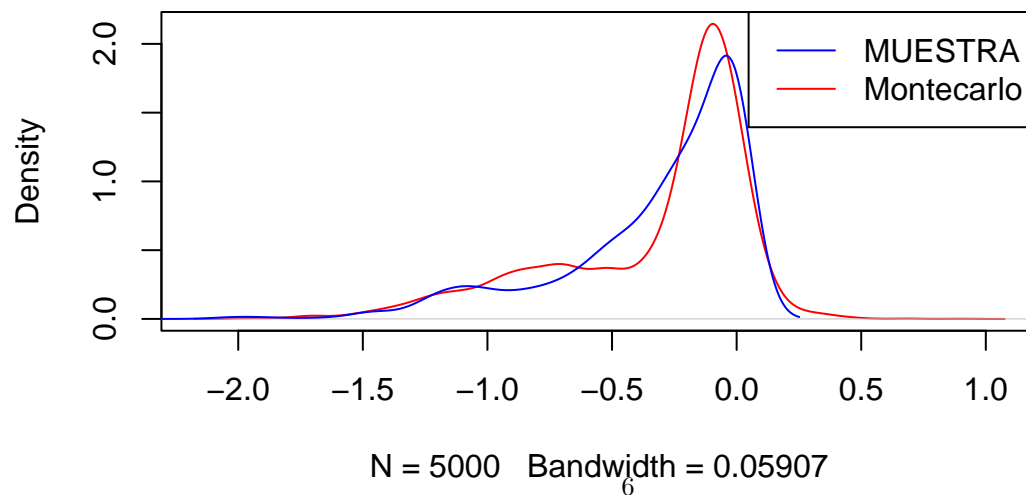


Figure 2: Densidad Originales transformados.

## Distribución del indicador de Solvencia



Finalmente presentamos los resultados obtenidos de las simulaciones realizadas

Table 1: **Estimaciones por Montecarlo y Bootstrap (Datos Transformados)**

Método/Medida	Media	Varianza	Sd	LINF	LimSUP
Montecarlo	-0.3285	0.1554	0.3942	-0.3396859	-0.3179281
Bootstrap	-0.328538	0.235	0.221	-0.3589944	-0.2990529

Para el entendimiento de los resultados retransformamos los valores estimados

Table 2: **Estimaciones por Montecarlo y Bootstrap (Datos Retransformados)**

Método	Media	Var	Sd	LINF	LSUP
Montecarlo	0.719	1.1681	1.483255	0.71326	0.72658
Bootstrap	0.766	0.05147959	0.2268183	0.7491227	0.7837504

Podemos observar que el metodo por Montecarlo es más angosto que bootstrap , pero en cambio este último presenta sesgos bastante pequeños como se muestra a continuación.

Table 3: **Estimaciones por Montecarlo y Bootstrap (Datos Retransformados)**

Método	SESGO MEDIA	SESGO Varianza	SESGO Sd
Montecarlo	-0.0464	1.1166	0.1671
Bootstrap	-5.237828e-05	8.694168e-05	0.0002643589

Para la prueba de hipótesis el jefe según su experiencia segura que el indice de solvencia en promedio es igual 72% . Por lo cual nuestra prueba de hipótesis viene dada por

$$H_0 : \theta = 0.72$$

$$H_a : \theta \neq 0.72$$

Donde \$ \$ representa la media poblacional de nuestro estimador.

##

## Proporción de rechazos al 1%= 0.008

##

## Proporción de rechazos al 5%= 0.049

##

## Proporción de rechazos al 10%= 0.1

Esta salida proporciona información sobre la tasa de rechazo bajo diferentes niveles de significancia, lo que es útil para evaluar el poder estadístico del procedimiento.

## 5 Conclusiones y recomendaciones

## 6 Anexos

## 7 Referencias

[1] Guayasamín, P. (17 de septiembre de 2023). Simulación Montecarlo y Bootstrap: Cooperativas deterioradas en el indicador de liquidez diciembre 2022 - enero 2023.

[2] Fernández-Casal, R., Cao, R., & Costa, J. (2023). Técnicas de Simulación y Remuestreo (github). Recuperado de [<https://rubenfcasal.github.io/simbook/>]

[3] Cao, R., & Fernández-Casal, R. (2021). Técnicas de Remuestreo (github). Recuperado de [[https://rubenfcasal.github.io/book\\_remuestreo/](https://rubenfcasal.github.io/book_remuestreo/)]