

PONTIFICIA UNIVERSIDAD  
CATÓLICA DEL PERÚ

Escuela de Posgrado



Enlace potencia Cauchy y su reversa en modelos de regresión  
binaria

Tesis para obtener el grado académico de Maestro en Estadística que  
presenta:

**Wilson Edwar Diaz Cajo**

**Asesor:**

**Alex de la Cruz Huayanay**

Lima, 2025


---

## Informe de Similitud

Yo, Alex de la Cruz Huayanay, docente de la Escuela de posgrado de la Pontificia Universidad Católica del Perú, asesor de la tesis titulada *Enlace potencia Cauchy y su reversa en modelos de regresión binaria*, del autor Wilson Edwar Diaz Cajo, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 14%. Así lo consigna el reporte de similitud emitido por el software Turnitin el 03/09/2025.
- He revisado con detalle dicho reporte y confirmo que cada una de las coincidencias detectadas no constituyen plagio alguno.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

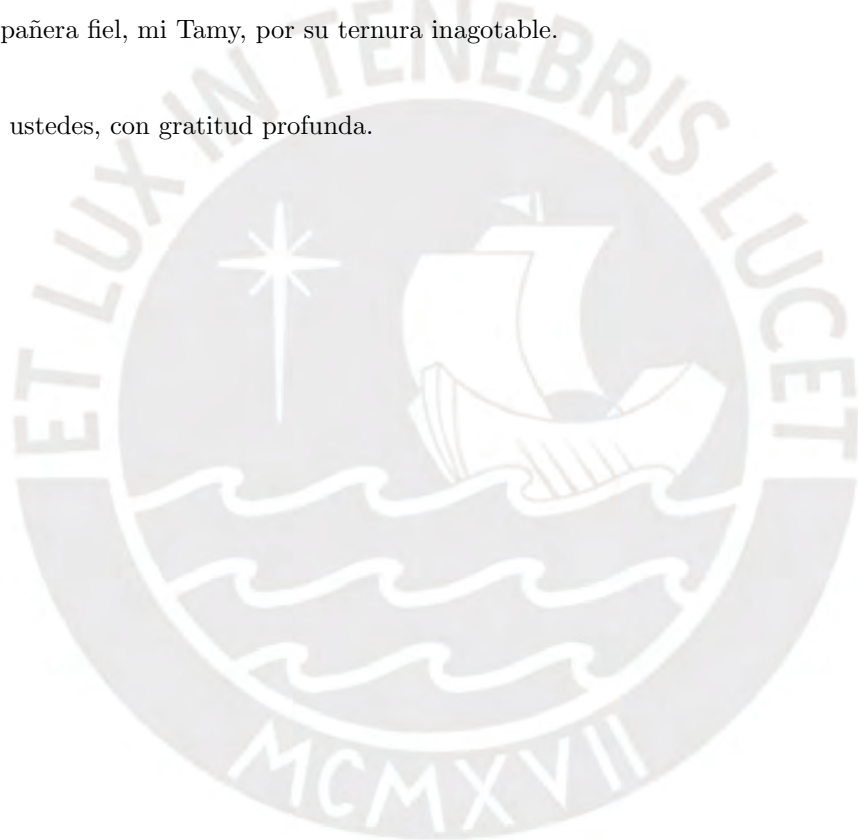
Lima, 03 de setiembre de 2025

<b>Apellidos y nombres del asesor:</b> Alex de la Cruz Huayanay	
DNI: 46121900	<b>Firma:</b> 
ORCID: <a href="https://orcid.org/0000-0003-0746-0803">https://orcid.org/0000-0003-0746-0803</a>	

# Dedicatoria

A mi madre Odulia, por ser mi raíz y mi ejemplo constante de fortaleza. A mi hija Azucena, por dar sentido a cada esfuerzo y llenar mi vida de esperanza. A mi esposa Cesibel, por su paciencia, amor incondicional y apoyo silencioso en cada etapa. A mis hermanos, por su cercanía y respaldo sincero. Y a mi compañera fiel, mi Tamy, por su ternura inagotable.

A todos ustedes, con gratitud profunda.



# Agradecimientos

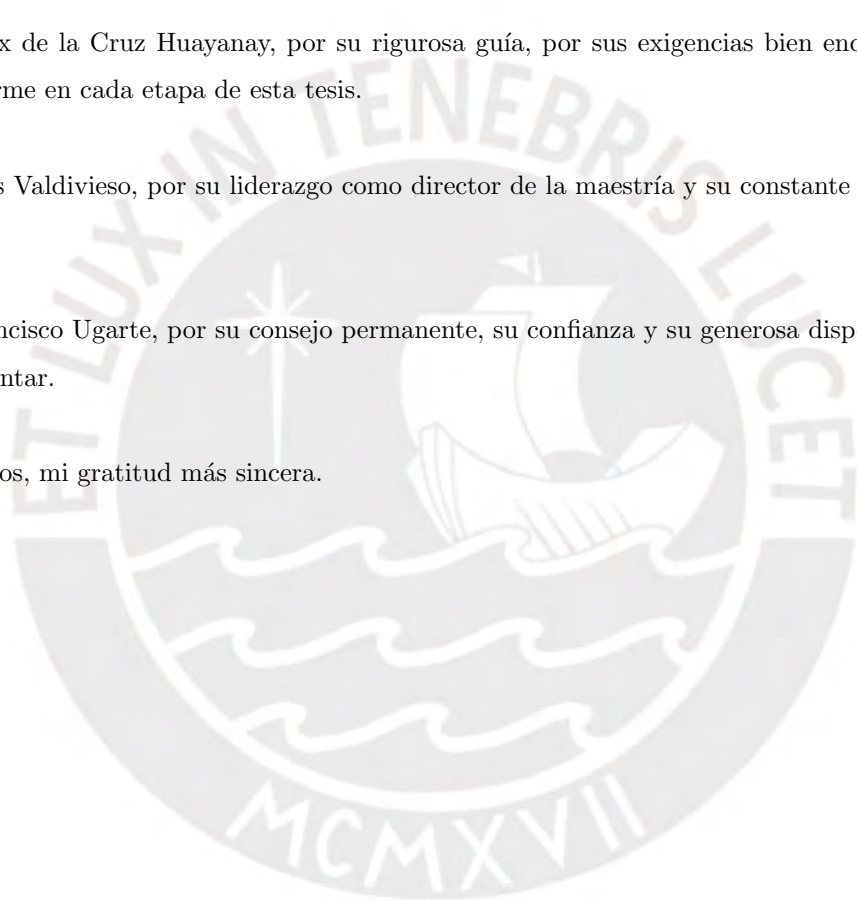
A Dios, por darme la fuerza, el tiempo y la claridad para culminar este trabajo.

Al Dr. Alex de la Cruz Huayanay, por su rigurosa guía, por sus exigencias bien encaminadas y por acompañarme en cada etapa de esta tesis.

Al Dr. Luis Valdivieso, por su liderazgo como director de la maestría y su constante apoyo institucional.

Al Dr. Francisco Ugarte, por su consejo permanente, su confianza y su generosa disposición para escuchar y orientar.

A todos ellos, mi gratitud más sincera.



# Resumen

Esta tesis aborda el estudio y la aplicación de funciones de enlace asimétricas basadas en las distribuciones potencia Cauchy y reversa de potencia Cauchy en modelos de regresión binaria. Dichas funciones incorporan un parámetro de asimetría que permite capturar de manera más precisa el comportamiento de datos desbalanceados, superando las limitaciones de funciones de enlace tradicionales como logit, probit o Cauchy simétrica.

En primer lugar, se realiza una caracterización teórica de ambas distribuciones, examinando propiedades clave como simetría, curtosis, asimetría y distancia de Wasserstein, con el fin de sustentar su adecuación como funciones de enlace. Posteriormente, se desarrollan dos enfoques de inferencia: uno bayesiano, empleando simulación MCMC bajo distintas distribuciones a priori para el parámetro de asimetría; y otro frecuentista, mediante el método de máxima verosimilitud.

Para evaluar el desempeño de los modelos, se lleva a cabo un estudio de simulación en escenarios controlados con diferentes niveles de desbalance. Se analizan métricas como el sesgo, el error cuadrático medio y la forma de la asimetría inducida. Finalmente, se aplica la metodología a un conjunto de datos reales sobre incumplimiento crediticio, comparando el ajuste de los modelos propuestos frente a enlaces clásicos, mediante criterios de información (DIC, WAIC, EAIC, EBIC, AIC, BIC, HQIC) y análisis de residuos.

Los resultados evidencian que las funciones de enlace potencia Cauchy y reversa potencia Cauchy ofrecen mayor flexibilidad y mejor desempeño en contextos con fuerte asimetría o desbalance de clases. Así, esta tesis contribuye tanto al desarrollo teórico de nuevas funciones de enlace como a su validación empírica, ampliando el repertorio de herramientas disponibles para el análisis de datos binarios en estadística aplicada.

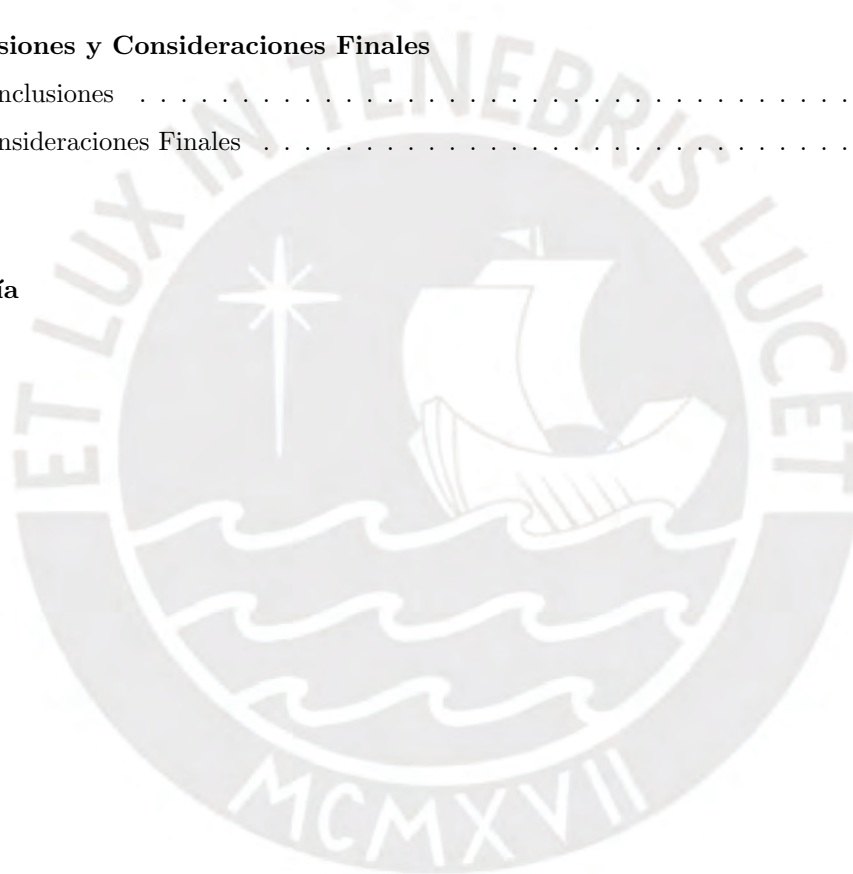
**Palabras clave:** regresión binaria, funciones de enlace asimétricas, distribución potencia Cauchy, inferencia bayesiana, máxima verosimilitud.

# Índice general

Índice de figuras	VIII
Índice de tablas	IX
<b>1. Introducción</b>	<b>1</b>
1.1. Consideraciones preliminares	1
1.2. Objetivos	2
1.3. Organización del trabajo	3
<b>2. Fundamentos teóricos</b>	<b>4</b>
2.1. Simetría y reversibilidad en distribuciones continuas	4
2.1.1. Ejemplos ilustrativos	5
2.2. Distribuciones Potencia y Reversa de Potencia Cauchy	6
2.2.1. Distribución potencia Cauchy	7
2.2.2. Distribución reversa de potencia Cauchy	7
2.2.3. Distribución potencia Cauchy estándar y reversa potencia Cauchy estándar	7
2.2.4. Asimetría	12
2.2.5. Curtosis	15
2.2.6. Diferencia de cuantiles y Distancia de Wasserstein	15
<b>3. Modelo</b>	<b>20</b>
3.1. Regresión Binaria con Función de Enlace Potencia y Reversa Potencia Cauchy	20
3.2. Inferencia vía método de máxima verosimilitud	21
3.3. Inferencia basada en el enfoque Bayesiano	26
3.3.1. Distribución a priori	27
3.3.2. Distribución a posteriori	27
3.4. Criterios de Comparación de Modelos	28
3.4.1. Criterios de Comparación de Modelos para el método de máxima verosimilitud	28
3.4.2. Criterios de Comparación de Modelos bayesianos	29
3.5. Análisis de Residuos	30

---

<b>4. Estudio de Simulación</b>	<b>32</b>
4.1. Análisis de Sensibilidad de la Priori para el Parámetro $\lambda$ . . . . .	32
4.2. Desempeño de la Función de Enlace Potencia Cauchy . . . . .	34
4.2.1. Comparación de curvas estimadas . . . . .	34
4.2.2. Recuperación de parámetros . . . . .	38
<b>5. Aplicación</b>	<b>43</b>
5.1. Descripción del Conjunto de Datos . . . . .	43
5.1.1. Variables Consideradas . . . . .	43
5.2. Análisis de modelos mediante inferencia bayesiana . . . . .	47
5.3. Análisis de modelos mediante inferencia por máxima verosimilitud . . . . .	52
<b>6. Conclusiones y Consideraciones Finales</b>	<b>55</b>
6.1. Conclusiones . . . . .	55
6.2. Consideraciones Finales . . . . .	56
<b>Anexo</b>	<b>57</b>
<b>Bibliografía</b>	<b>60</b>



# Índice de figuras

2.1. fdp Potencia e reversa potencia Cauchy estándar . . . . .	10
2.2. Curva de respuesta para la distribución Cauchy, Potencia Cauchy y distribución reversa de Potencia Cauchy con $\lambda = 3$ (izquierda) y $\lambda = 0,25$ (derecha). . . . .	12
2.3. Coeficiente de asimetría octil $AO(\lambda)$ para las distribuciones potencia Cauchy (PC) y reversa de potencia Cauchy (RPC). . . . .	14
4.1. Curvas promedio y bandas (100 simulaciones) - Verdadero: PC, $\lambda = 0,25$ . La curva verdadera (PC) y la PC ajustada se superponen casi totalmente; los enlaces simétricos se desvían principalmente en las colas. . . . .	34
4.2. Boxplots del estadístico KS - PC vs enlaces simétricos, $\lambda = 0,25$ . . . . .	35
4.3. Curvas promedio y bandas (100 simulaciones) - Verdadero: PC, $\lambda = 2$ . La PC ajustada sigue a la curva verdadera; los simétricos muestran desvíos crecientes en zonas de transición y colas. . . . .	35
4.4. Boxplots del estadístico KS - PC vs enlaces simétricos, $\lambda = 2$ . . . . .	36
4.5. Curvas promedio y bandas (100 simulaciones) - Verdadero: PC, $\lambda = 4$ . La PC ajustada se superpone con la curva verdadera; los enlaces simétricos se separan claramente, en especial <i>Cauchy</i> . . . . .	36
4.6. Boxplots del estadístico KS - PC vs enlaces simétricos, $\lambda = 4$ . . . . .	37
5.1. Distribución de la variable respuesta <i>Default</i> . . . . .	45
5.2. Matriz de correlación entre las variables predictoras. . . . .	45
5.3. Intervalo HPD 95 % para los coeficientes de regresión. . . . .	46
5.4. Gráfico de residuos cuantílicos para el modelo Potencia Cauchy. . . . .	48
5.5. QQ-Plot de residuos para el modelo Potencia Cauchy. . . . .	49
5.6. Gráfico de residuos cuantílicos para el modelo Logístico. . . . .	50
5.7. QQ-Plot de residuos para el modelo Logístico. . . . .	50
5.8. Gráfico de residuos cuantílicos para el modelo Potencia Cauchy (MLE). . . . .	53
5.9. QQ-Plot de residuos para el modelo Potencia Cauchy (MLE). . . . .	54

# Índice de tablas

2.1. Distribuciones y su notación correspondiente . . . . .	7
2.2. Amplitud de asimetría $AO(\lambda)$ para las distribuciones potencia Cauchy (PC) y reversa de potencia Cauchy (RPC), en comparación con funciones simétricas como logit, probit y Cauchy. Se consideran valores de $\lambda$ entre 0.001 y 9999. . . . .	13
2.3. $KO(\lambda)$ para las distribuciones Potencia Cauchy (PC) y Reversa de Potencia Cauchy (RPC), considerando valores entre $\lambda = 0,001$ y $\lambda = 9999$ . . . . .	15
4.1. Análisis de sensibilidad a priori para diferentes distribuciones del parámetro $\lambda$ . . . . .	33
4.2. KS ( $\lambda = 0,25$ ): proporción de rechazos ( $\alpha = 0,05$ ), mediana de $D$ y $D_{0,90}$ en $B = 100$ réplicas. . . . .	35
4.3. KS ( $\lambda = 2$ ): proporción de rechazos ( $\alpha = 0,05$ ), mediana de $D$ y $D_{0,90}$ en $B = 100$ réplicas. . . . .	36
4.4. KS ( $\lambda = 4$ ): proporción de rechazos ( $\alpha = 0,05$ ), mediana de $D$ y $D_{0,90}$ en $B = 100$ réplicas. . . . .	37
4.5. Resultados de $\hat{\beta}_0$ para $\lambda = 2$ . . . . .	39
4.6. Resultados de $\hat{\beta}_1$ para $\lambda = 2$ . . . . .	39
4.7. Resultados de $\hat{\beta}_0$ para $\lambda = 4$ . . . . .	40
4.8. Resultados de $\hat{\beta}_1$ para $\lambda = 4$ . . . . .	40
4.9. Resultados de $\hat{\beta}_0$ para $\lambda = 0,25$ . . . . .	41
4.10. Resultados de $\hat{\beta}_1$ para $\lambda = 0,25$ . . . . .	41
5.1. Estadísticas descriptivas del conjunto de datos. . . . .	44
5.2. Valores VIF de las variables predictoras. . . . .	46
5.3. Criterios de Información para la Comparación de Modelos . . . . .	47
5.4. Parámetros estimados para el modelo Potencia Cauchy. . . . .	47
5.5. Valores de Rhat para los parámetros clave del modelo Potencia Cauchy. . . . .	48
5.6. Criterios de Información (MLE) para Comparación de Modelos . . . . .	52
5.7. Parámetros estimados (MLE) para el modelo Potencia Cauchy. . . . .	53

# Capítulo 1

## Introducción

### 1.1. Consideraciones preliminares

En los modelos de regresión binaria, la variable de respuesta solo puede tomar dos posibles valores, comúnmente denominados *éxito* y *fracaso*, sin pérdida de generalidad. Estos modelos buscan relacionar la probabilidad de éxito con un conjunto de variables explicativas, denominadas covariables. En la práctica, tales modelos son ampliamente utilizados en disciplinas como las ciencias sociales, la educación, la medicina, la psicología, la agronomía y la zoología, entre otras. Para un desarrollo detallado de la regresión binaria, puede consultarse Cox y Snell (1989) y Collett (2002).

Formalmente, se considera el vector  $Y = (Y_1, \dots, Y_n)^\top$ , donde cada  $Y_i$  es una variable aleatoria independiente que toma valores en  $\{0, 1\}$ . Se asume que  $Y_i \sim \text{Bernoulli}(\mu_i)$ , donde  $\mu_i = \Pr(Y_i = 1)$  representa la probabilidad de éxito, mientras que  $\Pr(Y_i = 0) = 1 - \mu_i$  es la probabilidad de fracaso, para  $i = 1, \dots, n$ .

Para vincular la probabilidad  $\mu_i$  con las covariables, se emplean funciones de enlace. Las funciones más comunes son de tipo simétrico, como la logit y la probit. No obstante, diversos estudios han mostrado que estos enlaces pueden ser inadecuados en presencia de datos desbalanceados, es decir, cuando una de las clases domina significativamente a la otra (Bazán et al., 2017a; de la Cruz et al., 2019). Además, se ha observado que las funciones de enlace tradicionales no siempre ofrecen un buen ajuste cuando la tasa de crecimiento de la probabilidad de éxito hacia los extremos (0 o 1) no es simétrica respecto al punto de corte central  $\eta = 0$  (Chen et al., 1999). En otras palabras, estas funciones pueden fallar cuando la probabilidad de una respuesta binaria tiende a 0 y a 1 con velocidades distintas, lo cual es frecuente en contextos donde una clase es mucho más escasa que la otra.

Este tipo de comportamiento es común, por ejemplo, en estudios de predicción de enfermedades raras, fraudes financieros o incumplimiento crediticio, donde los casos positivos representan una fracción muy pequeña del total. En tales situaciones, un modelo mal especificado puede subestimar de forma sistemática la clase minoritaria, generando decisiones inadecuadas con consecuencias prácticas importantes.

---

En este contexto, en los últimos años se han propuesto funciones de enlace asimétricas para mejorar el ajuste en situaciones de desbalance. Particularmente, los trabajos de Bazán et al. (2017a), Lemonte y Bazán (2018) y de la Cruz et al. (2023) han introducido funciones de enlace basadas en distribuciones potencia y reversa de potencia, las cuales incorporan un parámetro adicional que permite capturar la asimetría presente en los datos.

Dichas funciones han mostrado un buen desempeño en estudios de simulación y aplicaciones reales. En particular, dos enlaces destacados son la función potencia Cauchy y la función reversa de potencia Cauchy, caracterizadas por introducir un parámetro de asimetría que otorga mayor flexibilidad al modelo.

No obstante, aún existe una clara carencia en la literatura respecto al análisis formal de las propiedades teóricas de estas funciones de enlace, lo cual limita su comprensión y aplicación práctica. Además, se ha identificado la falta de estudios que exploren el análisis de sensibilidad a priori del parámetro de asimetría bajo un enfoque bayesiano, así como la escasez de desarrollos metodológicos sobre la estimación por máxima verosimilitud e implementación computacional en entornos como R.

En este marco, la presente tesis tiene como propósito abordar estos vacíos mediante el estudio de las propiedades teóricas y prácticas de las funciones de enlace potencia Cauchy y reversa de potencia Cauchy. Se evaluará su comportamiento bajo distintos escenarios de desbalance, y se desarrollarán metodologías de inferencia tanto bayesianas como frecuentistas - incluyendo estimación por máxima verosimilitud - para comparar sus desempeños en estudios de simulación y aplicaciones con datos reales.

## 1.2. Objetivos

El objetivo principal de esta tesis es analizar el desempeño estadístico y las propiedades teóricas de las funciones de enlace potencia Cauchy y reversa de potencia Cauchy, en modelos de regresión binaria con datos desbalanceados, empleando enfoques de inferencia bayesiana y máxima verosimilitud.

Los objetivos específicos son los siguientes:

- Estudiar las propiedades teóricas de las distribuciones potencia Cauchy y reversa de potencia Cauchy en el contexto de regresión binaria.
- Implementar métodos de inferencia bayesiana mediante simulación MCMC, así como estimación por máxima verosimilitud.
- Realizar análisis de sensibilidad a priori del parámetro de asimetría en ambos modelos, explorando diversas distribuciones a priori.
- Diseñar estudios de simulación para evaluar el desempeño de los modelos propuestos en escenarios con datos desbalanceados.
- Aplicar los modelos a un conjunto de datos reales del ámbito financiero relacionados con el incumplimiento crediticio.

- 
- Comparar el desempeño de los modelos propuestos frente a enlaces clásicos (logit, probit y Cauchy) en términos de ajuste, precisión de estimación y sensibilidad a la asimetría.

### 1.3. Organización del trabajo

La tesis se encuentra organizada en seis capítulos, además de un anexo complementario. En el Capítulo 1 se presenta la introducción, que incluye el contexto general del problema, la motivación, los objetivos y la estructura del documento. El Capítulo 2 desarrolla los fundamentos teóricos necesarios: se introducen los conceptos clave, se definen formalmente las distribuciones potencia Cauchy y reversa de potencia Cauchy y se analizan sus propiedades, considerando medidas como la asimetría, la curtosis y la distancia de Wasserstein.

En el Capítulo 3 se construyen los modelos de regresión binaria basados en las funciones de enlace potencia Cauchy y reversa potencia Cauchy. Se detallan los enfoques bayesiano y frecuentista, la formulación estadística, las distribuciones a priori, la verosimilitud y los criterios de comparación de modelos.

El Capítulo 4 presenta un estudio de simulación diseñado para evaluar el comportamiento de los modelos bajo distintos escenarios de desbalance. En una primera parte se realiza la *comparación de modelos* (curvas promedio de  $\hat{\mu}(x)$  con bandas y contraste de Kolmogorov - Smirnov entre las distribuciones de probabilidades estimadas, PC frente a enlaces simétricos); en una segunda parte se aborda la *recuperación de parámetros* mediante sesgo y raíz del error cuadrático medio.

El Capítulo 5 expone la aplicación de la metodología a un conjunto de datos reales del ámbito financiero, relacionados con el incumplimiento crediticio; se comparan los modelos propuestos con enlaces clásicos y se realiza un análisis de residuos. Finalmente, el Capítulo 6 presenta las conclusiones y consideraciones finales, en las que se sintetizan los principales hallazgos, se discuten fortalezas y limitaciones del enfoque y se proponen líneas de investigación futura.

El anexo reúne material complementario, como implementaciones en R, funciones desarrolladas, algoritmos MCMC, especificaciones en Stan y detalles técnicos adicionales que respaldan el desarrollo principal de la tesis.

# Capítulo 2

## Fundamentos teóricos

En este capítulo se introducen formalmente las distribuciones potencia Cauchy (PC) y reversa de potencia Cauchy (RPC), construidas a partir de una distribución base simétrica. Se analizan sus propiedades teóricas, en particular simetría, reversibilidad, asimetría, curtosis y distancia de Wasserstein, con énfasis en la interpretación del parámetro  $\lambda$  como controlador de la asimetría. Estas propiedades fundamentan su aplicación como funciones de enlace en modelos de regresión binaria para datos desbalanceados.

### 2.1. Simetría y reversibilidad en distribuciones continuas

Siguiendo a Bazán et al. (2017a), nos enfocamos en distribuciones continuas para las cuales resulta relevante distinguir entre simetría y reversibilidad. Estas propiedades juegan un papel crucial en la construcción de funciones de enlace asimétricas, especialmente en contextos donde los modelos tradicionales no capturan adecuadamente patrones sesgados en los datos.

**Definición 1. Propiedad de Reversibilidad**

Sea  $S$  una variable aleatoria continua con una determinada distribución de probabilidad denotada por  $S \sim F(\cdot)$ . Decimos que la distribución de  $S$  satisface la propiedad de reversibilidad si la función de distribución acumulada (fda) de  $-S$  es una distribución diferente de la de  $S$  y puede ser escrita como  $-S \sim G(\cdot) \equiv 1 - F(-\cdot)$ . En ese caso, la distribución de  $G(\cdot)$  es llamada distribución reversa de  $F(\cdot)$ .

**Definición 2. Simetría**

Una distribución de probabilidad es llamada simétrica si y solo si existe un valor  $x_0$  tal que  $f(x_0 - d) = f(x_0 + d)$  o  $f(x_0 - d) = -f(x_0 + d)$  para todo número real  $d$ , donde  $f$  es la función de densidad de probabilidad (fdp).

**Proposición 1.** Considere  $S \sim F(\cdot)$  una variable aleatoria continua siguiendo una distribución simétrica en torno al cero ( $x_0 = 0$ ). Entonces  $F(\cdot)$  no satisface la propiedad de reversibilidad. Es decir,  $S$  y  $-S$  tienen la misma distribución y  $F(s) = 1 - F(-s)$ ,  $f(s) = f(-s)$ . Un ejemplo de esto es la distribución Cauchy.

---

Note que, si la simetría en torno del cero ( $x_0 = 0$ ) en una distribución es verificada, esta no presenta reversibilidad. Si no se verifica la simetría, es posible proponer una distribución reversa a ella.

### 2.1.1. Ejemplos ilustrativos

En lo que sigue, cuando se mencione una distribución estándar, se entenderá que corresponde a una distribución con media 0 y desviación estándar 1, salvo que se indique lo contrario. Esta convención facilita la comparación y el análisis en contextos estadísticos específicos.

La distribución normal estándar  $N(0,1)$  es un primer ejemplo de simetría, pues su función de densidad está dada por

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

y satisface que  $f(x) = f(-x)$ . Sin embargo, no es reversible, ya que su función de distribución acumulada  $F(x) = \Phi(x)$  no cumple que  $1 - F(-x) = F(x)$  para todo  $x$ .

Un caso similar ocurre con la distribución logística estándar, cuya función de distribución acumulada está dada por

$$F(x) = \frac{1}{1 + \exp(-x)}, \quad x \in \mathbb{R}.$$

Esta función es simétrica respecto a  $x = 0$ , pues se cumple que  $F(-x) = 1 - F(x)$ . No obstante, la distribución logística tampoco es estrictamente reversible, ya que la condición  $1 - F(-x) = F(x)$  no se satisface para todos los valores de  $x$ .

Por otro lado, la distribución Gumbel para valores máximos se define a través de la función de distribución acumulada

$$F(x) = \exp(-\exp(-x)), \quad x \in \mathbb{R}.$$

A diferencia de las anteriores, esta distribución no es simétrica, pues  $F(-x) \neq 1 - F(x)$ . Además, estudios recientes como Huayanay et al. (2024) han mostrado que sus extensiones potencia Gumbel y reversa potencia Gumbel producen valores constantes para cualquier parámetro  $\lambda$ , lo cual las hace poco útiles en aplicaciones de regresión binaria.

Finalmente, la distribución Cauchy estándar posee densidad

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R},$$

y función de distribución acumulada

$$F(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}.$$

La simetría se verifica ya que  $f(x) = f(-x)$ . No obstante, al igual que en los casos anteriores, no cumple la condición de reversibilidad, pues en general  $1 - F(-x) \neq F(x)$ .

---

## 2.2. Distribuciones Potencia y Reversa de Potencia Cauchy

En el trabajo de Lemonte y Bazán (2018) se describe la construcción de una distribución de potencia basada en considerar una función de distribución acumulada (fda) continua arbitraria y elevarla a una potencia real positiva arbitraria. Así, se propone una nueva función de distribución acumulada (fda) con un parámetro de potencia adicional. En este sentido, se tiene la siguiente definición:

### Definición 3. *Distribución Potencia*

Una variable aleatoria univariada  $T$  tiene distribución de potencia con parámetro de localización  $\mu \in \mathbb{R}$ , parámetro de escala  $\sigma > 0$  y parámetro de forma  $\lambda > 0$ , si su función de distribución acumulada (fda) es de la forma

$$F_P(x | \mu, \sigma, \lambda) = G\left(\frac{x - \mu}{\sigma}\right)^\lambda; \quad x \in \mathbb{R},$$

y su función de densidad de probabilidad (fdp) está dada por:

$$f_P(x | \mu, \sigma, \lambda) = \frac{\lambda}{\sigma} G\left(\frac{x - \mu}{\sigma}\right)^{\lambda-1} g\left(\frac{x - \mu}{\sigma}\right); \quad x \in \mathbb{R},$$

donde  $g(\cdot)$  y  $G(\cdot)$  son, respectivamente, cualquier función de densidad de probabilidad (fdp) y función de distribución acumulada (fda) estándar de una distribución univariada continua con soporte en la recta real ( $\mathbb{R}$ ), esta distribución es llamada de línea de base.

Esta definición se aplica, por ejemplo, cuando la distribución base es la distribución de Cauchy estándar, cuya simetría la hace especialmente adecuada para esta construcción.

### Definición 4. *Distribución Reversa Potencia*

Como fue presentado por Lemonte y Bazán (2018), la función de distribución reversa de potencia (RP) puede ser definida considerando su fda como:

$$F_{RP}(x | \mu, \sigma, \lambda) = 1 - G\left(-\frac{x - \mu}{\sigma}\right)^\lambda; \quad x \in \mathbb{R},$$

donde  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  y  $\lambda > 0$  son parámetros de localización, escala y forma respectivamente, y  $G(\cdot)$  es una fda de la línea de base de una distribución univariada continua en su forma estandarizada.

Al igual que en la definición de la distribución potencia, esta definición es válida cuando  $G(\cdot)$  es, por ejemplo, la función de distribución acumulada estándar de la distribución de Cauchy.

En este trabajo, utilizamos la distribución de Cauchy como la línea de base para construir las distribuciones potencia y reversa de potencia. Esta elección se debe a la simetría inherente de la distribución de Cauchy estándar, lo que facilita su uso en este contexto.

Tabla 2.1: Distribuciones y su notación correspondiente

Tipo	Nombre de la distribución	Notación
Línea de base	Cauchy	C
Potencia	Potencia Cauchy	PC
Reversa de potencia	Reversa de potencia Cauchy	RPC

### 2.2.1. Distribución potencia Cauchy

A partir de la Definición 3, considerando que  $G$  es la función de distribución acumulada estándar Cauchy, tenemos:

La fda de distribución de potencia Cauchy:

$$F_{PC}(x | \mu, \sigma, \lambda) = \left( G \left( \frac{x - \mu}{\sigma} \right) \right)^\lambda$$

La fdp de distribución de potencia Cauchy:

$$f_{PC}(x | \mu, \sigma, \lambda) = \frac{\lambda}{\sigma\pi} \left( G \left( \frac{x - \mu}{\sigma} \right) \right)^{\lambda-1} \left( 1 + \left( \frac{x - \mu}{\sigma} \right)^2 \right)^{-1}$$

### 2.2.2. Distribución reversa de potencia Cauchy

Utilizando la Definición 4, y considerando que  $G$  es la función de distribución acumulada estándar Cauchy, tenemos:

La fda de distribución de reversa potencia Cauchy:

$$F_{RPC}(x | \mu, \sigma, \lambda) = 1 - \left( G \left( -\frac{x - \mu}{\sigma} \right) \right)^\lambda$$

La fdp de distribución de reversa potencia Cauchy:

$$f_{RPC}(x | \mu, \sigma, \lambda) = \frac{\lambda}{\sigma\pi} \left( G \left( -\frac{x - \mu}{\sigma} \right) \right)^{\lambda-1} \left( 1 + \left( \frac{x - \mu}{\sigma} \right)^2 \right)^{-1}$$

### 2.2.3. Distribución potencia Cauchy estándar y reversa potencia Cauchy estándar

A partir de esta sección, presentaremos las definiciones particulares de las distribuciones potencia y reversa de potencia Cauchy en su forma estándar. Estas distribuciones, junto con sus respectivas funciones de densidad y distribución acumulada, serán fundamentales a lo largo de esta tesis, ya que se utilizarán como base en el desarrollo y análisis de los modelos propuestos.

En la Definición 3, la distribución de potencia estándar ocurre cuando  $\mu = 0$  y  $\sigma = 1$ , y su función de distribución acumulada es  $F_P(z) = G(z)^\lambda$ , para  $z \in \mathbb{R}$ , donde  $G$  es la función de distribución acumulada estándar Cauchy.

La fda de distribución de potencia Cauchy estándar:

$$F_{PC}(x | \lambda) = (G(x))^\lambda$$

La fdp de distribución de potencia Cauchy estándar:

$$f_{PC}(x | \lambda) = \frac{\lambda}{\pi} (G(x))^{\lambda-1} (1+x^2)^{-1}$$

De la misma manera, las distribuciones reversa de potencia estándar con una distribución de base simétrica  $G$  y su función de distribución acumulada, pueden ser introducidas a partir de la Definición 4, considerando  $\mu = 0$  y  $\sigma = 1$ . Tenemos:

La fda de distribución de reversa potencia Cauchy estándar:

$$F_{RPC}(x | \lambda) = 1 - (G(-x))^\lambda$$

La fdp de distribución de reversa potencia Cauchy estándar:

$$f_{RPC}(x | \lambda) = \frac{\lambda}{\pi} (G(-x))^{\lambda-1} (1+x^2)^{-1}$$

**Proposición 2.** Para las distribuciones potencia Cauchy en su versión estándar, se sigue que:

$$F_{PC}(-x | \lambda) = \left[ 1 - \left( \frac{1}{2} + \frac{1}{\pi} \arctan(x) \right) \right]^\lambda; \quad x \in \mathbb{R}.$$

**Demostración.** Note que se cumple que  $G(-z) = 1 - G(z)$ , donde:

$$G(z) = \frac{1}{2} + \frac{1}{\pi} \arctan(z),$$

así obtenemos lo deseado. □

**Proposición 3.**  $F_{RPC}(\cdot)$  satisface la propiedad de reversibilidad, ya que:

$$F_{RPC}(x | \lambda) + F_{PC}(-x | \lambda) = 1$$

donde  $F_{PC}(\cdot)$  y  $F_{RPC}(\cdot)$  son las funciones de distribución acumulada de la distribución potencia Cauchy estándar y la distribución reversa de potencia Cauchy estándar, respectivamente.

**Demostración.** Veamos:

$$F_{PC}(-x | \lambda) = \left( \frac{1}{\pi} \arctan(-x) + \frac{1}{2} \right)^\lambda$$

y

$$F_{RPC}(x | \lambda) = 1 - \left[ \frac{1}{\pi} \arctan(-x) + \frac{1}{2} \right]^\lambda$$

Sumando ambas ecuaciones, se sigue que:

$$F_{RPC}(x | \lambda) + F_P(-x | \lambda) = 1.$$

□

**Proposición 4.** Para todo  $\lambda \neq 1$ , las funciones de distribución potencia Cauchy y reversa de potencia Cauchy estándar no son simétricas, esto es,  $f_{PC}(x | \lambda) \neq f_{PC}(-x | \lambda)$  y  $f_{RPC}(x | \lambda) \neq f_{RPC}(-x | \lambda)$ .

**Demostración.** La función de densidad de probabilidad potencia Cauchy estándar es:

$$f_{PC}(x | \lambda) = \frac{\lambda}{\pi} \left( \frac{1}{\pi} \arctan(t) + \frac{1}{2} \right)^{\lambda-1} (1+x^2)^{-1}$$

Así que, si evaluamos esta función en el punto  $-x$ , obtenemos:

$$f_{PC}(-x | \lambda) = \frac{\lambda}{\pi} \left( \frac{1}{\pi} \arctan(-x) + \frac{1}{2} \right)^{\lambda-1} (1+x^2)^{-1} \neq f_{PC}(x | \lambda).$$

De manera similar, para la función de densidad de probabilidad reversa de potencia Cauchy estándar, tenemos:

$$f_{RPC}(x | \lambda) = \frac{\lambda}{\pi} \left[ \frac{1}{\pi} \arctan(-x) + \frac{1}{2} \right]^{\lambda-1} (1+x^2)^{-1}$$

Si evaluamos esta función en el punto  $-t$ , obtenemos:

$$f_{RPC}(-x | \lambda) = \frac{\lambda}{\pi} \left[ \frac{1}{\pi} \arctan(x) + \frac{1}{2} \right]^{\lambda-1} (1+x^2)^{-1} \neq f_{RPC}(x | \lambda).$$

Por lo tanto, podemos concluir que tanto las distribuciones potencia como reversa de potencia no son simétricas en torno a cero ( $x_0 = 0$ ).

□

Observemos que para  $\lambda = 1$  si hay simetría, pues tenemos que  $f_{PC}(x) = f_{RPC}(x) = \frac{\lambda}{\pi} (1+x^2)^{-1}$ .

En la siguiente figura se muestra la curva de las funciones de densidad de probabilidad (fdp) para las distribuciones de potencia estándar y reversa de potencia Cauchy estándar con diferentes valores de  $\lambda$ .

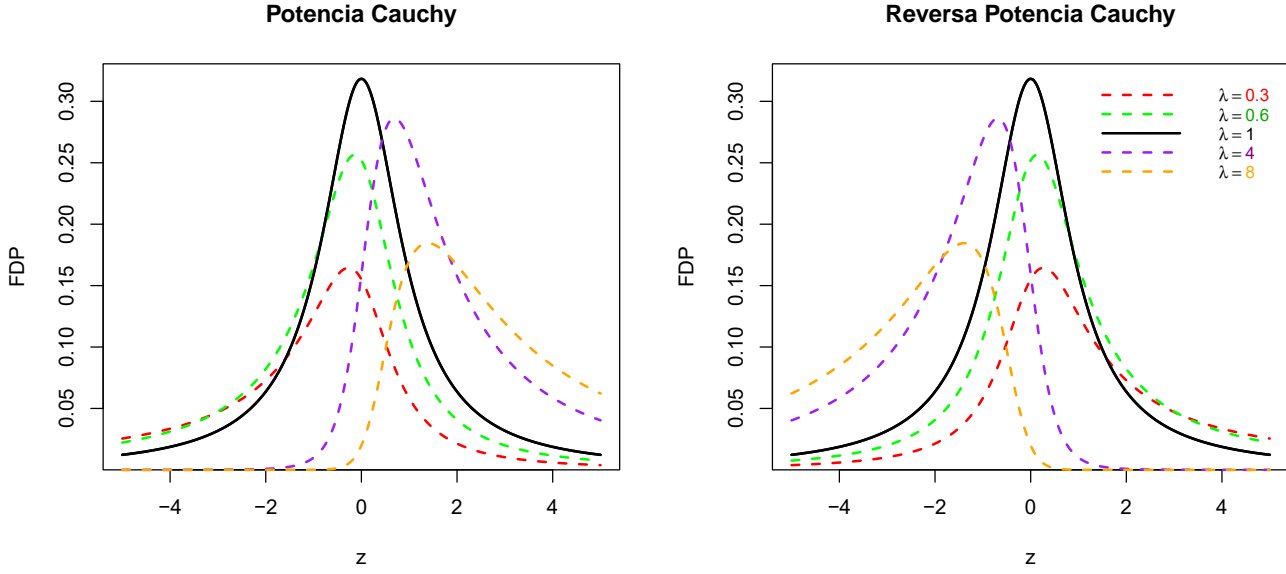


Figura 2.1: fdp Potencia e reversa potencia Cauchy estándar

La importancia de las distribuciones de potencia y reversa de potencia Cauchy radica principalmente en que el parámetro adicional  $\lambda$  introduce el grado de asimetría. Por lo tanto, podemos tener un alto grado de asimetría (tanto positiva como negativa) para las distribuciones de potencia y reversa de potencia Cauchy, dependiendo de los valores del parámetro de forma adicional  $\lambda$ .

En particular, cuando  $\lambda = 1$ , tenemos simetría.

**Observación 1.** *Las distribución de potencia Cauchy está sesgada hacia la derecha (asimetría positiva) si  $\lambda > 1$  y hacia la izquierda (asimetría negativa) si  $0 < \lambda < 1$ . Las distribución reversa de potencia está sesgada hacia la izquierda (asimetría negativa) si  $\lambda > 1$  y hacia la derecha (asimetría positiva) si  $0 < \lambda < 1$ .*

En el contexto de las funciones de enlace potencia y reversa de potencia Cauchy, es importante generar valores de estas distribuciones para realizar simulaciones y estudiar su comportamiento. Para ello, podemos utilizar el método de la inversa, que nos permite obtener valores simulados a partir de la distribución uniforme estándar. En particular, las funciones cuantílicas de las distribuciones potencia y reversa de potencia Cauchy nos permiten generar estos valores de forma eficiente. A continuación, se presenta una proposición que describe cómo generar valores de estas distribuciones utilizando el método inverso.

**Proposición 5.** *Sea  $U \sim \text{Uniforme}(0, 1)$ , entonces:*

- $X = Q_{PC}(U) = \tan\left(\pi\left(U^{1/\lambda} - \frac{1}{2}\right)\right)$  sigue una distribución potencia Cauchy.
- $X = Q_{RPC}(U) = -\tan\left(\pi\left((1-U)^{1/\lambda} - \frac{1}{2}\right)\right)$  sigue una distribución reversa de potencia Cauchy.

*Donde  $Q_{PC}(U)$  y  $Q_{RPC}(U)$  son los valores de los cuantiles generados por las funciones de distribución acumulada potencia Cauchy y reversa de potencia Cauchy, respectivamente.*

---

**Demostración.**

**Para la distribución potencia Cauchy:**

$$F_{PC}(x) = \left( \frac{1}{2} + \frac{1}{\pi} \arctan(x) \right)^\lambda$$

Queremos encontrar  $X$  tal que  $F_{PC}(X) = U$ , es decir:

$$\left( \frac{1}{2} + \frac{1}{\pi} \arctan(X) \right)^\lambda = U$$

Aplicamos la inversa:

$$\frac{1}{2} + \frac{1}{\pi} \arctan(X) = U^{1/\lambda}$$

Despejamos  $X$ :

$$\arctan(X) = \pi \left( U^{1/\lambda} - \frac{1}{2} \right)$$

Por lo tanto:

$$X = \tan \left( \pi \left( U^{1/\lambda} - \frac{1}{2} \right) \right)$$

**Para la distribución reversa de potencia Cauchy:**

$$F_{RPC}(x) = 1 - \left( \frac{1}{2} + \frac{1}{\pi} \arctan(-x) \right)^\lambda$$

Queremos encontrar  $X$  tal que  $F_{RPC}(X) = U$ , es decir:

$$1 - \left( \frac{1}{2} + \frac{1}{\pi} \arctan(-X) \right)^\lambda = U$$

Lo que implica:

$$\left( \frac{1}{2} + \frac{1}{\pi} \arctan(-X) \right)^\lambda = 1 - U$$

Aplicamos la inversa:

$$\frac{1}{2} + \frac{1}{\pi} \arctan(-X) = (1 - U)^{1/\lambda}$$

Despejamos  $X$ :

$$\arctan(-X) = \pi \left( (1 - U)^{1/\lambda} - \frac{1}{2} \right)$$

Finalmente:

$$X = -\tan \left( \pi \left( (1 - U)^{1/\lambda} - \frac{1}{2} \right) \right)$$

□

Siguiendo lo planteado por Bazán et al. (2017a), cuando  $\lambda < 1$  (o  $\lambda > 1$ ), la curva de la distribución potencia tiende a estar por encima (o por debajo) de la curva de la distribución base dentro de cierto rango de valores de  $\eta$ . De manera similar, para cada valor de  $\lambda$ , la curva de la distribución reversa de potencia actúa como un reflejo de la distribución potencia. Así, cuando  $\lambda < 1$  (o  $\lambda > 1$ ), la curva

de la distribución reversa generalmente se encuentra por debajo (o por encima) de la curva de la distribución base. En la figura 2 se ilustra un ejemplo utilizando la distribución Cauchy como base, con valores de  $\lambda = \{0,25, 3\}$ .

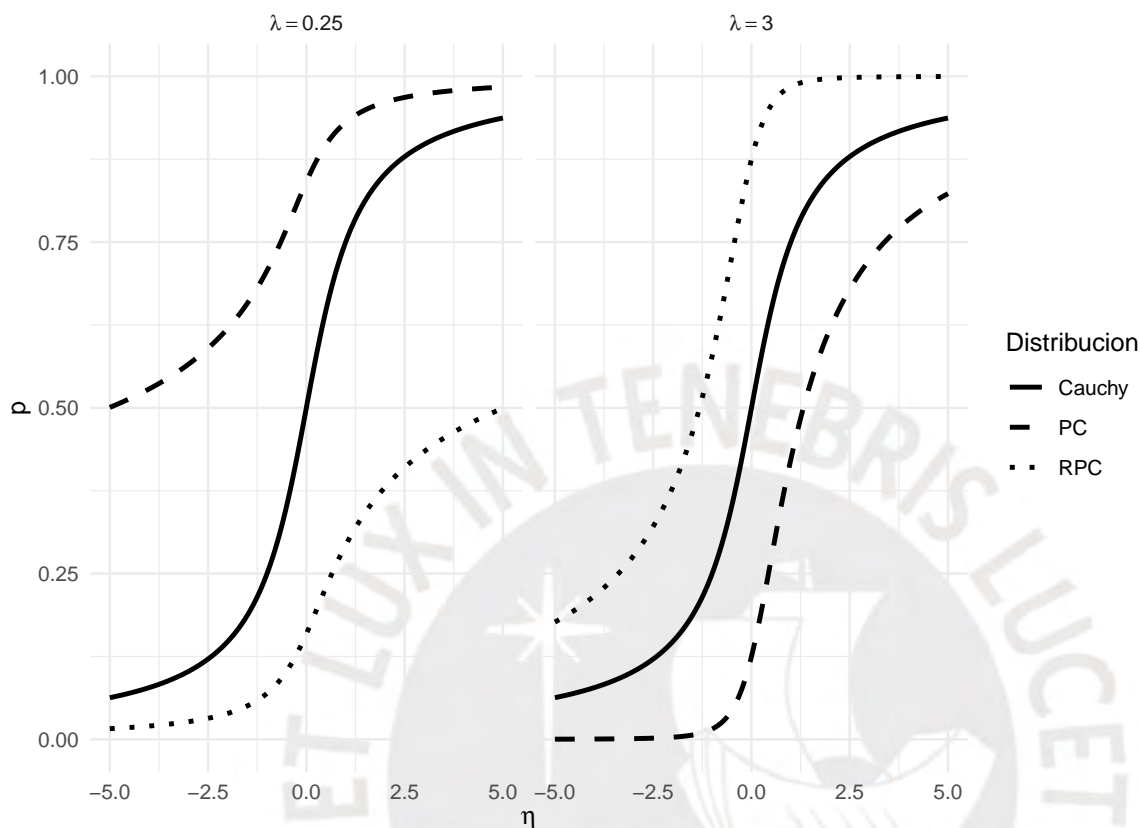


Figura 2.2: Curva de respuesta para la distribución Cauchy, Potencia Cauchy y distribución reversa de Potencia Cauchy con  $\lambda = 3$  (izquierda) y  $\lambda = 0,25$  (derecha).

#### 2.2.4. Asimetría

Las distribuciones potencia Cauchy (PC) y reversa de potencia Cauchy (RPC) afectan tanto la asimetría como la curtosis de las distribuciones base. Por ello, en esta sección analizamos cómo se comportan estas medidas para las distribuciones mencionadas. Estudios previos, como los de de la Cruz et al. (2019) y Chumbimune (2017), ya han explorado este tipo de distribuciones. Sin embargo, calcular medidas tradicionales de asimetría y curtosis, como el coeficiente de Pearson, resulta complicado debido a que las distribuciones potencia y reversa de potencia no tienen una forma analítica cerrada para los momentos. Un ejemplo de esta dificultad se discute en Gupta y Gupta (2008). Para evitar estos inconvenientes, una opción viable es utilizar medidas alternativas basadas en cuantiles.

Una manera efectiva de medir la asimetría es usando el coeficiente de asimetría octil, propuesto por Brys et al. (2004), que se expresa como:

$$AO = \frac{(O_7 - O_4) - (O_4 - O_1)}{O_7 - O_1}$$

donde  $O_a$  representa el  $a$ -ésimo octil. En el caso de las distribuciones potencia y reversa de potencia Cauchy, la asimetría depende del parámetro  $\lambda$  y se puede expresar mediante la siguiente fórmula:

$$AO(\lambda) = \frac{Q(0,875, \lambda) - 2Q(0,5, \lambda) + Q(0,125, \lambda)}{Q(0,875, \lambda) - Q(0,125, \lambda)}$$

Aquí,  $Q(p, \lambda)$  es la función cuantil correspondiente, con  $0 < p < 1$  y  $\lambda > 0$ , lo que garantiza que  $0 < AO < 1$ .

Las funciones cuantílicas para las distribuciones potencia Cauchy (PC) y reversa de potencia Cauchy (RPC) se define de la siguiente manera:

$$Q_u^{PC}(\mu, \lambda) = \tan\left(\pi\left(\mu^{1/\lambda} - 0,5\right)\right)$$

$$Q_u^{RPC}(\mu, \lambda) = -\tan\left(\pi\left((1 - \mu)^{1/\lambda} - 0,5\right)\right)$$

donde  $Q(\mu, \lambda)$  es la función cuantil ajustada según el parámetro de potencia  $\lambda$ , con  $\mu$  representando la probabilidad acumulada. La relación entre estas funciones cuantílicas implica que, a medida que  $\lambda$  varía, el comportamiento de la asimetría en ambas distribuciones cambia de manera simétrica pero opuesta.

Utilizando las fórmulas cuantílicas presentadas previamente, se calcularon los valores de  $AO(\lambda)$  para las distribuciones potencia y reversa de potencia Cauchy y distintos valores de  $\lambda$ , que se muestran en la siguiente tabla. También se calcula la amplitud de las diferencias en el coeficiente de asimetría, denotada por  $r$ , que representa la diferencia entre el valor máximo y mínimo de la asimetría según el valor de  $\lambda$ .

Tabla 2.2: Amplitud de asimetría  $AO(\lambda)$  para las distribuciones potencia Cauchy (PC) y reversa de potencia Cauchy (RPC), en comparación con funciones simétricas como logit, probit y Cauchy. Se consideran valores de  $\lambda$  entre 0.001 y 9999.

Distribución	Mínimo	Máximo	$\lambda < 1$	$\lambda \geq 1$	$r^*$
Logística	0	0	(0; 0)	[0; 0]	0
Normal estándar	0	0	(0; 0)	[0; 0]	0
Cauchy	0	0	(0; 0)	[0; 0]	0
PC	-1.0000	0.7255	-1,0000; 0,0000	[0,0000; 0,7255)	1.7255
RPC	-0.7255	1.0000	(0,0000; 1,0000)	[-0,7255; 0,0000)	1.7255

Es importante notar que las funciones de enlace basadas en las distribuciones potencia Cauchy (PC) y reversa de potencia Cauchy (RPC) presentan una amplitud considerable de asimetría, la cual varía en función del parámetro  $\lambda$ . A diferencia de las funciones simétricas tradicionales como logit, probit y Cauchy cuyas medidas de asimetría son nulas, las funciones PC y RPC permiten una representación más flexible de relaciones no lineales en presencia de sesgos estructurales.

Como se muestra en la Tabla 2.2, esta amplitud alcanza hasta  $r^* = 1,7255$ , reflejando su capacidad para adaptarse a distintos grados de asimetría. Este comportamiento es especialmente útil en contextos donde los datos presentan fuerte desbalance o colas largas, condiciones en las que los

modelos convencionales resultan insuficientes.

Las distribuciones logística, normal estándar y Cauchy presentan un coeficiente de asimetría octil igual a 0 porque son distribuciones simétricas en su forma estándar. Sus funciones de densidad acumulada y cuantílicas cuando se considera  $\lambda = 1$ .

Ahora veamos el siguiente gráfico

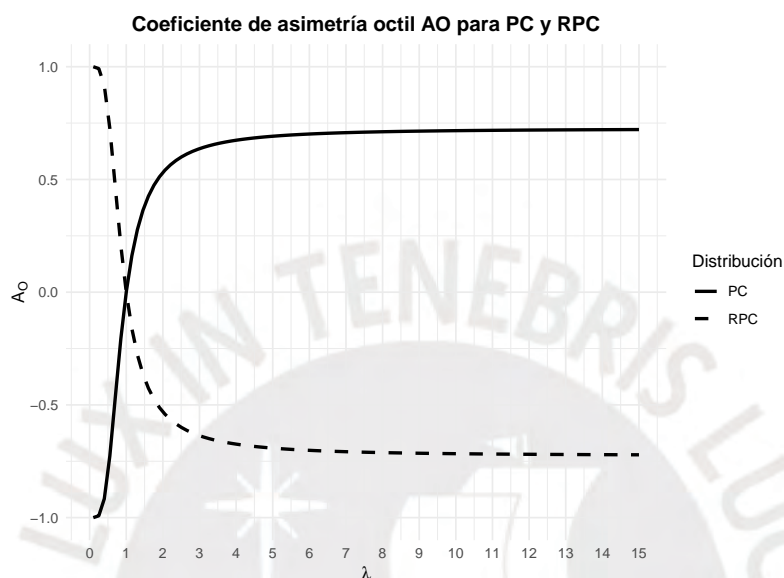


Figura 2.3: Coeficiente de asimetría octil  $AO(\lambda)$  para las distribuciones potencia Cauchy (PC) y reversa de potencia Cauchy (RPC).

En esta figura se presenta el comportamiento del coeficiente de asimetría octil  $AO(\lambda)$  para las distribuciones potencia Cauchy (PC) y reversa de potencia Cauchy (RPC). A partir de la gráfica se pueden extraer varias conclusiones.

En primer lugar, cuando  $\lambda = 1$ , ambas distribuciones son simétricas, ya que el coeficiente de asimetría octil toma el valor de 0. Esto confirma que en ese caso ambas coinciden con la simetría de la distribución Cauchy en su forma estándar.

Cuando  $\lambda > 1$ , la distribución potencia Cauchy muestra una asimetría positiva creciente, lo que indica una inclinación hacia la derecha, mientras que la reversa de potencia Cauchy exhibe una asimetría negativa creciente, inclinándose hacia la izquierda. Por el contrario, para valores de  $\lambda < 1$ , la potencia Cauchy presenta asimetría negativa (hacia la izquierda), mientras que la reversa de potencia Cauchy muestra asimetría positiva (hacia la derecha).

A partir de valores aproximados de  $\lambda = 3$ , ambas distribuciones estabilizan su asimetría, de modo que un incremento adicional en  $\lambda$  no produce cambios significativos en la magnitud de dicha asimetría. En general, el coeficiente  $AO(\lambda)$  permite representar diferentes grados y direcciones de asimetría en las distribuciones potencia Cauchy y reversa de potencia Cauchy, lo que resulta relevante para modelar datos desbalanceados o con estructura asimétrica, ya que el parámetro  $\lambda$  controla directamente tanto la magnitud como el sentido de la asimetría.

Además, considerando la relación entre ambas distribuciones, se cumple la identidad

$$AO_{RPC}(\lambda) = -AO_{PC}(\lambda),$$

donde  $AO_{PC}(\lambda)$  y  $AO_{RPC}(\lambda)$  representan los coeficientes de asimetría octil de la distribución potencia Cauchy y de la reversa de potencia Cauchy, respectivamente.

### 2.2.5. Curtosis

Al igual que con la medida de asimetría, empleamos el coeficiente de curtosis basado en cuantiles, propuesto por Moors (1988). Este coeficiente se define como:

$$KO = \frac{(O_7 - O_5) + (O_3 - O_1)}{O_6 - O_2}$$

Sin embargo, dado que estos valores no se encuentran en una escala adecuada, utilizamos la curtosis de la distribución Normal estándar, que es 1.233, para redimensionar esta medida. Por lo tanto, la siguiente medida de curtosis es:

$$KO(\lambda) = \frac{100}{1,233} \times \left( \frac{Q(0,875, \lambda) - Q(0,625, \lambda) + Q(0,375, \lambda) - Q(0,125, \lambda)}{Q(0,75, \lambda) - Q(0,25, \lambda)} - 1,233 \right)$$

donde  $Q(p, \lambda)$  es la función cuantil de las distribuciones potencia Cauchy(PC) y reversa de potencia Cauchy(RPC). A continuación, presentamos los valores de  $KO(\lambda)$  para estas distribuciones con diferentes valores de  $\lambda$ :

Tabla 2.3:  $KO(\lambda)$  para las distribuciones Potencia Cauchy (PC) y Reversa de Potencia Cauchy (RPC), considerando valores entre  $\lambda = 0,001$  y  $\lambda = 9999$ .

Distribución	Mínimo	Máximo	$0 < \lambda < 1$	$\lambda \geq 1$	$r^*$
PC	56.5886	37527838.9177	(62.2085; 37527838.9177)	(56.5886; 73.7012)	37527782.3291
RPC	-73.7012	-56.5886	(-37527838.9177; -62.2085)	(-73.7012; -56.5886)	37527782.3291

Los valores de  $KO(\lambda)$  varían significativamente según el parámetro  $\lambda$ , y se puede observar que tanto la **potencia Cauchy** como la **reversa de potencia Cauchy** presentan una gran amplitud de curtosis. Además, cuando  $\lambda = 1$ , ambas distribuciones alcanzan la simetría y la curtosis de la distribución base (Cauchy).

### 2.2.6. Diferencia de cuantiles y Distancia de Wasserstein

#### Definición 5. Diferencia de cuantiles

Podemos definir la **diferencia de cuantiles** considerando dos probabilidades diferentes  $p_1$  y  $p_2$ , respectivamente, para el primer y segundo caso, de la siguiente forma:

$$D((p_1, \lambda_1); (p_2, \lambda_2)) = Q(p_1, \lambda_1) - Q(p_2, \lambda_2)$$

donde  $Q(p, \lambda)$  es la función cuantil de las distribuciones **potencia Cauchy (PC)** o **reversa de potencia Cauchy (RPC)**, con  $\lambda > 0$  el parámetro de forma.

**Definición 6. Distancia de Wasserstein**

La **distancia de Wasserstein**, que mide la diferencia entre los cuantiles y está dada por:

$$W_m(f_1, f_2)(p) = \int_0^1 |F_1^{-1}(p) - F_2^{-1}(p)|^m dp = \int_0^1 |D((p, \lambda_1); (p, \lambda_2))|^m dp,$$

donde  $F^{-1}$  es la inversa de la función de distribución acumulada (fda). Para distribuciones unidimensionales, la distancia de Wasserstein  $W_m$  es simplemente la distancia  $L_m$  entre variables aleatorias simuladas de una distribución uniforme en el intervalo  $[0, 1]$ .

**Proposición 6.** Sea  $X_1 \sim f_1$  con  $f_1$  siendo la distribución potencia Cauchy  $PC(\lambda_1)$ , y  $X_2 \sim f_2$  con  $f_2$  siendo la distribución potencia Cauchy  $PC(\lambda_2)$ . Para las probabilidades  $p_1$  y  $p_2$ , respectivamente, la **diferencia de cuantiles** entre ellos está dada por:

$$D((p_1, \lambda_1); (p_2, \lambda_2)) = \tan\left(\frac{\pi}{2}(p_2^{1/\lambda_2} - 0,5)\right) - \tan\left(\frac{\pi}{2}(p_1^{1/\lambda_1} - 0,5)\right)$$

**Demostración.** La diferencia de cuantiles para las distribuciones potencia Cauchy se calcula a partir de las funciones cuantílicas correspondientes. La función cuantil de la distribución potencia Cauchy está dada por:

$$Q_{PC}(p, \lambda) = \tan\left(\frac{\pi}{2}(p^{1/\lambda} - 0,5)\right)$$

Así que, la diferencia de cuantiles para  $p_1$  y  $p_2$ , evaluada en  $\lambda_1$  y  $\lambda_2$ , se expresa como:

$$D((p_1, \lambda_1); (p_2, \lambda_2)) = Q_{PC}(p_1, \lambda_1) - Q_{PC}(p_2, \lambda_2)$$

Sustituyendo las expresiones de las funciones cuantílicas obtenemos:

$$D((p_1, \lambda_1); (p_2, \lambda_2)) = \tan\left(\frac{\pi}{2}(p_1^{1/\lambda_1} - 0,5)\right) - \tan\left(\frac{\pi}{2}(p_2^{1/\lambda_2} - 0,5)\right)$$

Por lo tanto, hemos demostrado que la diferencia de cuantiles entre las distribuciones potencia Cauchy con parámetros de forma  $\lambda_1$  y  $\lambda_2$  está dada por:

$$D((p_1, \lambda_1); (p_2, \lambda_2)) = \tan\left(\frac{\pi}{2}(p_2^{1/\lambda_2} - 0,5)\right) - \tan\left(\frac{\pi}{2}(p_1^{1/\lambda_1} - 0,5)\right)$$

□

Cuando  $p_1 = p_2 = p$ , la diferencia de cuantiles entre las distribuciones  $PC(\lambda_1)$  y  $PC(\lambda_2)$  está dada por:

$$D((p, \lambda_1); (p, \lambda_2)) = \tan\left(\frac{\pi}{2}(p^{1/\lambda_1} - 0,5)\right) - \tan\left(\frac{\pi}{2}(p^{1/\lambda_2} - 0,5)\right),$$

Análogamente, se demuestra lo mismo para la distribución de la reversa potencia Cauchy.

**Proposición 7.** Sea  $X_1 \sim f_1$  con  $f_1$  siendo la distribución reversa de potencia Cauchy  $RPC(\lambda_1)$ , y  $X_2 \sim f_2$  con  $f_2$  siendo la distribución reversa de potencia Cauchy  $RPC(\lambda_2)$ . Para las probabilidades  $p_1$  y  $p_2$ , respectivamente, la **diferencia de cuantiles** entre ellas está dada por:

$$D((p_1, \lambda_1); (p_2, \lambda_2)) = \tan\left(\frac{\pi}{2}\left((1-p_2)^{1/\lambda_2} - \frac{1}{2}\right)\right) - \tan\left(\frac{\pi}{2}\left((1-p_1)^{1/\lambda_1} - \frac{1}{2}\right)\right)$$

**Demostración.** La diferencia de cuantiles para las distribuciones reversa de potencia Cauchy se calcula a partir de sus funciones cuantílicas. La función cuantil de la distribución reversa de potencia Cauchy está dada por:

$$Q_{RPC}(p, \lambda) = -\tan\left(\frac{\pi}{2}\left((1-p)^{1/\lambda} - \frac{1}{2}\right)\right)$$

Así que, la diferencia de cuantiles para  $p_1$  y  $p_2$ , evaluada en  $\lambda_1$  y  $\lambda_2$ , se expresa como:

$$D((p_1, \lambda_1); (p_2, \lambda_2)) = Q_{RPC}(p_1, \lambda_1) - Q_{RPC}(p_2, \lambda_2)$$

Sustituyendo las expresiones explícitas de las funciones cuantílicas obtenemos:

$$D((p_1, \lambda_1); (p_2, \lambda_2)) = -\tan\left(\frac{\pi}{2}\left((1-p_1)^{1/\lambda_1} - \frac{1}{2}\right)\right) + \tan\left(\frac{\pi}{2}\left((1-p_2)^{1/\lambda_2} - \frac{1}{2}\right)\right)$$

Reordenando los términos, la forma final es:

$$D((p_1, \lambda_1); (p_2, \lambda_2)) = \tan\left(\frac{\pi}{2}\left((1-p_2)^{1/\lambda_2} - \frac{1}{2}\right)\right) - \tan\left(\frac{\pi}{2}\left((1-p_1)^{1/\lambda_1} - \frac{1}{2}\right)\right)$$

□

Cuando  $p_1 = p_2 = p$ , la diferencia de cuantiles entre las distribuciones  $RPC(\lambda_1)$  y  $RPC(\lambda_2)$  está dada por:

$$D((p, \lambda_1); (p, \lambda_2)) = \tan\left(\frac{\pi}{2}\left((1-p)^{1/\lambda_2} - \frac{1}{2}\right)\right) - \tan\left(\frac{\pi}{2}\left((1-p)^{1/\lambda_1} - \frac{1}{2}\right)\right)$$

Para evaluar la viabilidad de la distribución potencia Cauchy como función de enlace en regresión binaria, estamos interesados en conocer la distancia entre los cuantiles para diferentes valores de  $\lambda$ . Para ello, consideramos la **distancia de Wasserstein**.

**Proposición 8.** Sea  $X_1 \sim f_1$  donde  $f_1$  sigue una distribución potencia Cauchy  $PC(\lambda_1)$ , y  $X_2 \sim f_2$  donde  $f_2$  sigue una distribución potencia Cauchy  $PC(\lambda_2)$ . Para la misma probabilidad  $p$ , la **distancia de Wasserstein** está dada por:

$$W_m(f_1, f_2)(p) = \int_0^1 \left| \tan\left(\frac{\pi}{2}(p^{1/\lambda_1} - 0,5)\right) - \tan\left(\frac{\pi}{2}(p^{1/\lambda_2} - 0,5)\right) \right|^m dp.$$

**Demostración.** Para calcular la distancia de Wasserstein entre las distribuciones potencia Cauchy,

utilizamos sus funciones cuantil. La función cuantil de la distribución potencia Cauchy está dada por:

$$Q_{PC}(p, \lambda) = \tan\left(\frac{\pi}{2}(p^{1/\lambda} - 0,5)\right).$$

Por lo tanto, la diferencia de cuantiles para  $p$  entre  $f_1$  y  $f_2$  es:

$$D((p, \lambda_1); (p, \lambda_2)) = \tan\left(\frac{\pi}{2}(p^{1/\lambda_1} - 0,5)\right) - \tan\left(\frac{\pi}{2}(p^{1/\lambda_2} - 0,5)\right).$$

La distancia de Wasserstein  $W_m(f_1, f_2)(p)$  se calcula integrando esta diferencia de cuantiles elevada a la potencia  $m$  sobre el intervalo unitario  $[0, 1]$ :

$$W_m(f_1, f_2)(p) = \int_0^1 |D((p, \lambda_1); (p, \lambda_2))|^m dp.$$

Por lo tanto, la distancia de Wasserstein entre las distribuciones potencia Cauchy depende de los valores de  $\lambda_1$  y  $\lambda_2$  y es una función de la diferencia de cuantiles.  $\square$

Para el caso de la distribución potencia Cauchy, se demuestra que la distancia de Wasserstein es la misma que en las distribuciones potencia Cauchy.

**Proposición 9.** Sea  $X_1 \sim f_1$  donde  $f_1$  sigue una distribución reversa de potencia Cauchy  $RPC(\lambda_1)$ , y  $X_2 \sim f_2$  donde  $f_2$  sigue una distribución reversa de potencia Cauchy  $RPC(\lambda_2)$ . Para la misma probabilidad  $p \in (0, 1)$ , la **distancia de Wasserstein** de orden  $m$  entre ambas distribuciones también está dada por:

$$W_m(f_1, f_2)(p) = \int_0^1 \left| \tan\left(\frac{\pi}{2}(p^{1/\lambda_1} - 0,5)\right) - \tan\left(\frac{\pi}{2}(p^{1/\lambda_2} - 0,5)\right) \right|^m dp.$$

**Demostración.** La distancia de Wasserstein de orden  $m$  entre las distribuciones  $RPC(\lambda_1)$  y  $RPC(\lambda_2)$ , con funciones de distribución acumulada  $F_1$  y  $F_2$  respectivamente, se define como:

$$W_m(f_1, f_2) = \int_0^1 |F_1^{-1}(p) - F_2^{-1}(p)|^m dp.$$

La función cuantil de la distribución reversa de potencia Cauchy está dada por:

$$Q_{RPC}(p, \lambda) = -\tan\left(\frac{\pi}{2}\left((1-p)^{1/\lambda} - \frac{1}{2}\right)\right),$$

por lo tanto, tenemos:

$$F_1^{-1}(p) = Q_{RPC}(p, \lambda_1) = -\tan\left(\frac{\pi}{2}\left((1-p)^{1/\lambda_1} - \frac{1}{2}\right)\right),$$

$$F_2^{-1}(p) = Q_{RPC}(p, \lambda_2) = -\tan\left(\frac{\pi}{2}\left((1-p)^{1/\lambda_2} - \frac{1}{2}\right)\right).$$

Sustituyendo estas expresiones en la definición de la distancia de Wasserstein, obtenemos:

$$W_m(f_1, f_2) = \int_0^1 \left| -\tan\left(\frac{\pi}{2}\left((1-p)^{1/\lambda_1} - \frac{1}{2}\right)\right) + \tan\left(\frac{\pi}{2}\left((1-p)^{1/\lambda_2} - \frac{1}{2}\right)\right) \right|^m dp.$$

Ahora, aplicamos el cambio de variable  $u = 1 - p$ , con lo cual cuando  $p = 0$ ,  $u = 1$ , y cuando  $p = 1$ ,  $u = 0$ . Por tanto, la integral se convierte en:

$$W_m(f_1, f_2) = \int_1^0 \left| -\tan\left(\frac{\pi}{2}\left(u^{1/\lambda_1} - \frac{1}{2}\right)\right) + \tan\left(\frac{\pi}{2}\left(u^{1/\lambda_2} - \frac{1}{2}\right)\right) \right|^m (-du).$$

Revirtiendo los límites, obtenemos:

$$W_m(f_1, f_2) = \int_0^1 \left| \tan\left(\frac{\pi}{2}\left(u^{1/\lambda_2} - \frac{1}{2}\right)\right) - \tan\left(\frac{\pi}{2}\left(u^{1/\lambda_1} - \frac{1}{2}\right)\right) \right|^m du.$$

Como  $u$  es solo una variable muda de integración, podemos renombrarla nuevamente como  $p$ , y se concluye que:

$$W_m(f_1, f_2) = \int_0^1 \left| \tan\left(\frac{\pi}{2}\left(p^{1/\lambda_2} - 0,5\right)\right) - \tan\left(\frac{\pi}{2}\left(p^{1/\lambda_1} - 0,5\right)\right) \right|^m dp.$$

Por tanto, se concluye que la distancia de Wasserstein entre dos distribuciones reversa de potencia Cauchy tiene la misma expresión que la obtenida para las distribuciones potencia Cauchy.  $\square$

En este capítulo, hemos analizado las distribuciones potencia Cauchy y reversa potencia Cauchy, resaltando aspectos clave como su simetría, reversibilidad y cómo el parámetro  $\lambda$  controla la asimetría. Estas no son solo propiedades teóricas; tienen una utilidad práctica importante.

Cuando usamos estas distribuciones como funciones de enlace en modelos de regresión binaria, su comportamiento, ya sea simétrico o asimétrico, está directamente relacionado con las formas de las distribuciones acumuladas que describimos antes. Esto es especialmente valioso para trabajar con datos desbalanceados, ya que  $\lambda$  nos permite ajustar la asimetría de manera flexible y modelar patrones complejos.

En los próximos capítulo, veremos cómo estas distribuciones se aplican como funciones de enlace en la regresión binaria y cómo sus propiedades se convierten en herramientas prácticas para analizar datos reales.

# Capítulo 3

## Modelo

Luego de haber establecido las bases teóricas de las funciones de enlace estudiadas, este capítulo presenta su incorporación formal en modelos de regresión binaria, considerando tanto el enfoque bayesiano como el frecuentista, con el propósito de evaluar su aplicabilidad y desempeño.

### 3.1. Regresión Binaria con Función de Enlace Potencia y Reversa Potencia Cauchy

Sea  $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$  un vector  $n \times 1$  de variables aleatorias respuesta independientes. Para cada observación  $i = 1, \dots, n$ , se define el vector de covariables  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$  y un vector de coeficientes de regresión  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ . La probabilidad de éxito se denota como  $\mu_i = \mathbb{P}(Y_i = 1)$ , mientras que la probabilidad de fracaso es  $\mathbb{P}(Y_i = 0) = 1 - \mu_i$ .

En el modelo propuesto, la probabilidad de éxito  $\mu_i$  está determinada por una función de enlace basada en la distribución potencia Cauchy o en la reversa de potencia Cauchy, de manera que

$$\begin{aligned} Y_i | \boldsymbol{\beta}, \lambda &\sim \text{Bernoulli}(\mu_i), \quad i = 1, \dots, n, \\ \mu_i &= F_\lambda(\eta_i), \\ \eta_i &= \mathbf{x}_i^\top \boldsymbol{\beta}, \end{aligned} \tag{3.1}$$

donde  $F_\lambda(\cdot)$  representa la función de distribución acumulada correspondiente a la distribución potencia Cauchy o reversa de potencia Cauchy, y  $\lambda$  es el parámetro que controla la asimetría.

En el marco de los Modelos Lineales Generalizados (MLG), la función inversa  $F^{-1}(\cdot)$  recibe el nombre de función de enlace. Esta función relaciona la media de la variable respuesta  $\mathbb{E}[Y_i]$  con la combinación lineal de las covariables  $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ . De esta forma, la elección de la función de enlace determina cómo los cambios en las covariables influyen en la respuesta esperada del modelo. Cuando la distribución  $F(\cdot)$  es simétrica en torno a cero, como ocurre en los casos logit, probit o Cauchy, la función de enlace resultante también es simétrica. En contraste, si  $F(\cdot)$  es asimétrica, la función de enlace lo será igualmente, lo que puede resultar ventajoso en la modelación de datos desbalanceados.

En particular, para las distribuciones potencia Cauchy y reversa de potencia Cauchy, las propiedades de las funciones de enlace dependen directamente del valor del parámetro  $\lambda$ . Cuando  $\lambda = 1$ , ambas funciones de enlace son simétricas, reproduciendo el comportamiento de la Cauchy estándar. En cambio, cuando  $\lambda \neq 1$ , las funciones se vuelven asimétricas, lo que permite capturar patrones de desbalance en los datos. Esta flexibilidad convierte al parámetro  $\lambda$  en un mecanismo de ajuste del grado y la dirección de la asimetría, ampliando la aplicabilidad de estos enlaces en problemas de clasificación binaria.

En este contexto, el modelo de regresión binaria propuesto involucra dos tipos de parámetros desconocidos. Por un lado,  $\beta$  constituye un vector de parámetros estructurales propios de las observaciones, independiente de la elección del modelo. Por otro,  $\lambda$  es un parámetro asociado a la función de enlace que regula la asimetría. La estimación de ambos parámetros se realizará mediante dos enfoques: la inferencia basada en el método de máxima verosimilitud y la inferencia desde la perspectiva bayesiana.

### 3.2. Inferencia vía método de máxima verosimilitud

De manera general, para hacer inferencias sobre  $\beta$ , se puede usar un método de estimación de parámetros llamado Método de Máxima Verosimilitud (MV).

**Definición 7. Estimador de máxima verosimilitud**

Considere  $n$  observaciones  $x_1, \dots, x_n$  iid y la verosimilitud de  $\beta$  dada por

$$\mathcal{L}_n(\beta) = \prod_{i=1}^n f(x_i; \beta).$$

Luego, la función de log-verosimilitud es definida como

$$\ell_n(\beta) = \log \mathcal{L}_n(\beta).$$

Un resultado inmediato es

$$\ell_n(\beta) = \sum_{i=1}^n \log f(x_i; \beta).$$

Decimos que  $\hat{\beta}$  es un estimador de máxima verosimilitud (EMV) para  $\beta$  si

$$\mathcal{L}_n(\hat{\beta}) = \sup_{\beta_0 \in B} \mathcal{L}_n(\beta_0) \quad \text{o equivalentemente} \quad \ell_n(\hat{\beta}) = \sup_{\beta_0 \in B} \ell_n(\beta_0).$$

Otra forma de definir un EMV es a través de la siguiente expresión:

$$\hat{\beta} = \arg \max_{\beta_0 \in B} \ell_n(\beta_0).$$

**Definición 8. Función score**

La función score es definida por la siguiente expresión:

$$U = U(\beta) = \frac{\partial \ell_n(\beta)}{\partial \beta}.$$

Algunas propiedades de esta función son las siguientes.

1.  $E(U) = 0$
2.  $\text{Var}(U) = E(U^2) = E\left(-\frac{\partial^2 \ell_n(\beta)}{\partial \beta^2}\right)$ , si  $\beta$  es un escalar.
3.  $\text{Cov}(U) = E(UU^\top) = E\left(-\frac{\partial^2 \ell_n(\beta)}{\partial \beta \partial \beta^\top}\right)$ , si  $\beta$  es un vector.

En algunos casos, podemos obtener  $\hat{\beta}$  de forma explícita resolviendo:

$$U = 0.$$

En otros casos, se requiere un procedimiento numérico para lo cual usamos la matriz de información de Fisher.

**Definición 9. Matriz de información de Fisher.**

La información de Fisher se define por:

$$\mathcal{I} = \text{Cov}(U).$$

Bajo condiciones de regularidad, se tiene que:

$$\mathcal{I} = \mathcal{I}(\beta) = E(UU^\top) = E\left(\frac{\partial \ell_n(\beta)}{\partial \beta} \frac{\partial \ell_n(\beta)^\top}{\partial \beta}\right) = E\left(-\frac{\partial^2 \ell_n(\beta)}{\partial \beta \partial \beta^\top}\right).$$

Usando una expansión de Taylor para  $U$  alrededor de  $\hat{\beta}$  hasta primer orden y tomando  $U = 0$ , obtenemos:

$$U - \hat{U} = (\beta - \hat{\beta})U'.$$

Sustituyendo  $U'$  por su valor esperado, es decir, por  $\mathcal{I}$ :

$$U - \hat{U} = -\mathcal{I}(\beta - \hat{\beta}),$$

$$\mathcal{I}^{-1}U = (\hat{\beta} - \beta),$$

$$\hat{\beta} = \beta + \mathcal{I}^{-1}U.$$

$$\beta^{(j+1)} = \beta^{(j)} + (\mathcal{I}^{-1})^{(j)} U^{(j)}.$$

La solución para encontrar un EMV es un proceso iterativo hasta que

$$\|\beta^{(j+1)} - \beta^{(j)}\| < \varepsilon.$$

---

## Aplicando la inferencia vía método de máxima verosimilitud para el modelo en estudio

Para nuestro caso la función de verosimilitud asociada al modelo está dada por:

$$L(\beta, \lambda | y, X) = \prod_{i=1}^n \left[ \left( \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \right)^\lambda \right]^{y_i} \left[ 1 - \left( \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \right)^\lambda \right]^{1-y_i},$$

donde  $\eta_i = x_i^T \beta$ , para la fda de la distribución potencia Cauchy, y

$$L(\beta, \lambda | y, X) = \prod_{i=1}^n \left[ 1 - \left( \frac{1}{\pi} \arctan(-\eta_i) + \frac{1}{2} \right)^\lambda \right]^{y_i} \left[ \left( \frac{1}{\pi} \arctan(-\eta_i) + \frac{1}{2} \right)^\lambda \right]^{1-y_i},$$

donde  $\eta_i = x_i^T \beta$ , para la fda de la distribución reversa de potencia Cauchy.

Consideremos el método de máxima verosimilitud (ML) para estimar conjuntamente los parámetros  $\beta = (\beta_1, \dots, \beta_p)^\top$  y  $\lambda$ . Sea  $\mathbf{y} = (y_1, \dots, y_n)^\top$  el vector de respuestas binarias observadas. La función de log-verosimilitud para el modelo con la función de enlace Potencia Cauchy es

$$\ell(\beta, \lambda | y, X) = \sum_{i=1}^n y_i \log \left[ \left( \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \right)^\lambda \right] + \sum_{i=1}^n (1 - y_i) \log \left[ 1 - \left( \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \right)^\lambda \right].$$

Mientras que la función de log-verosimilitud para el modelo con función de enlace de reversa potencia Cauchy es

$$\ell(\beta, \lambda | y, X) = \sum_{i=1}^n y_i \log \left[ 1 - \left( \frac{1}{\pi} \arctan(-\eta_i) + \frac{1}{2} \right)^\lambda \right] + \sum_{i=1}^n (1 - y_i) \log \left[ \left( \frac{1}{\pi} \arctan(-\eta_i) + \frac{1}{2} \right)^\lambda \right].$$

Donde  $X = (x_1, \dots, x_n)^\top$  es una matriz de orden  $n \times p$  con rango completo. La función de score, obtenida al diferenciar la función de log-verosimilitud con respecto a los parámetros desconocidos, es  $U(\beta, \lambda) = (U_\beta(\beta, \lambda)^\top, U_\lambda(\beta, \lambda)^\top)^\top$ , y

$$U_\beta(\beta, \lambda) = X^\top W(y - \mu), \quad U_\lambda(\beta, \lambda) = s^\top (y - \mu),$$

donde  $W = \text{diag}(w_1, \dots, w_n)$ ,  $\mu = (\mu_1, \dots, \mu_n)^\top$  y  $s = (s_1, \dots, s_n)^\top$ .

Para el modelo de Potencia Cauchy:

$$\mu_i = \left( \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \right)^\lambda, \quad w_i = \frac{\lambda \left( \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \right)^{\lambda-1} \frac{1}{\pi(1+\eta_i^2)}}{\mu_i(1-\mu_i)}, \quad s_i = \lambda^{-1} \frac{\log(\mu_i)}{1-\mu_i}.$$

Para el modelo de Reversa Potencia Cauchy:

$$\mu_i = 1 - \left( \frac{1}{\pi} \arctan(-\eta_i) + \frac{1}{2} \right)^\lambda, \quad w_i = \frac{\lambda \left( \frac{1}{\pi} \arctan(-\eta_i) + \frac{1}{2} \right)^{\lambda-1} \frac{-1}{\pi(1+\eta_i^2)}}{\mu_i(1-\mu_i)}, \quad s_i = -\lambda^{-1} \frac{\log(1-\mu_i)}{\mu_i}.$$

La matriz esperada de información de Fisher de orden  $(p+1) \times (p+1)$  toma la forma

$$K(\boldsymbol{\beta}, \lambda) = \begin{bmatrix} X^\top V X & X^\top t \\ t^\top X & \text{tr}(D) \end{bmatrix},$$

donde  $V = \text{diag}(v_1, \dots, v_n)$ ,  $D = \text{diag}(d_1, \dots, d_n)$ ,  $t = (t_1, \dots, t_n)^\top$  y  $\text{tr}(\cdot)$  denota el operador traza.

Para el modelo de regresión con enlace potencia Cauchy, tenemos:

$$v_i = \frac{\left[ \lambda \left( \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \right)^{\lambda-1} \frac{1}{\pi(1+\eta_i^2)} \right]^2}{[\mu_i(1-\mu_i)]}, \quad d_i = \lambda^{-2} \mu_i \frac{[\log(\mu_i)]^2}{(1-\mu_i)},$$

$$t_i = \left( \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \right)^{\lambda-1} \frac{1}{\pi(1+\eta_i^2)} \frac{\log(\mu_i)}{(1-\mu_i)}, \quad \mu_i = \left( \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \right)^\lambda.$$

Para el modelo de regresión con enlace de reversa potencia Cauchy, se tiene:

$$v_i = \frac{\left[ \lambda \left( \frac{1}{\pi} \arctan(-\eta_i) + \frac{1}{2} \right)^{\lambda-1} \frac{1}{\pi(1+\eta_i^2)} \right]^2}{\mu_i(1-\mu_i)}, \quad d_i = \lambda^{-2} (1-\mu_i) \frac{[\log(1-\mu_i)]^2}{\mu_i},$$

$$t_i = -\lambda^{-1} \left( \frac{1}{\pi} \arctan(-\eta_i) + \frac{1}{2} \right)^{\lambda-1} \frac{1}{\pi(1+\eta_i^2)} \frac{\log(1-\mu_i)}{\mu_i}, \quad \mu_i = 1 - \left( \frac{1}{\pi} \arctan(-\eta_i) + \frac{1}{2} \right)^\lambda.$$

Cabe destacar que los parámetros  $\boldsymbol{\beta}$  y  $\lambda$  no son ortogonales en el sentido de Cox y Reid (1987).

Sean  $P$  y  $M$  las matrices de dimensión  $(n+1) \times (p+1)$  y  $(n+1) \times (n+1)$ , respectivamente, dadas por:

$$P = \begin{bmatrix} X & 0_{n,1} \\ 0_{1,p} & 1 \end{bmatrix}, \quad M = \begin{bmatrix} V & t \\ t^\top & \text{tr}\{D\} \end{bmatrix}.$$

Donde  $0_{l,c}$  una matriz de ceros de dimensión  $l \times c$ . Entonces, la matriz de información de Fisher esperada de dimensión  $(p+1) \times (p+1)$  y su inversa pueden expresarse como

$$K(\boldsymbol{\beta}, \lambda) = P^\top M P, \quad \text{y} \quad K(\boldsymbol{\beta}, \lambda)^{-1} = (P^\top M P)^{-1},$$

respectivamente. Utilizando expresiones estándar para la inversa de matrices particionadas (por ejemplo, Rao (1973), p. 33), obtenemos

$$K(\boldsymbol{\beta}, \lambda)^{-1} = \begin{bmatrix} K^{\beta\beta} & K^{\beta\lambda} \\ K^{\lambda\beta} & K^{\lambda\lambda} \end{bmatrix},$$

donde

$$K^{\beta\beta} = (X^\top V X)^{-1} \left[ I_p + \frac{X^\top t t^\top X (X^\top V X)^{-1}}{\gamma} \right],$$

$$K^{\beta\lambda} = (K^{\lambda\beta})^\top = -\frac{(X^\top V X)^{-1} X^\top t}{\gamma}, \quad K^{\lambda\lambda} = \gamma^{-1},$$

y

$$\gamma = \text{tr}\{D\} - t^\top X (X^\top V X)^{-1} X^\top t.$$

Aquí,  $I_p$  es la matriz identidad de dimensión  $p \times p$ .

Los estimadores de máxima verosimilitud (ML)  $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)^\top$  y  $\hat{\lambda}$  de  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$  y  $\lambda$ , respectivamente, pueden obtenerse resolviendo simultáneamente el sistema de ecuaciones no lineales

$$U_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}, \hat{\lambda}) = 0_{p,1} \quad \text{y} \quad U_{\lambda}(\hat{\boldsymbol{\beta}}, \hat{\lambda}) = 0.$$

No existe una expresión en forma cerrada para los estimadores ML  $\hat{\boldsymbol{\beta}}$  y  $\hat{\lambda}$ , por lo que su cálculo debe realizarse numéricamente mediante un algoritmo de optimización no lineal. Por ejemplo, se pueden aplicar técnicas iterativas como Newton-Raphson (o los métodos de Gauss-Newton y Cuasi-Newton) para resolver estas ecuaciones y obtener  $\hat{\boldsymbol{\beta}}$  y  $\hat{\lambda}$  numéricamente.

Por otro lado, también se puede usar el método de puntuación de Fisher para estimar

$$\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \lambda)^\top$$

resolviendo iterativamente la ecuación

$$\boldsymbol{\theta}^{(m+1)} = \left( P^\top M^{(m)} P \right)^{-1} P^\top M^{(m)} z^{(m)},$$

(4)

donde  $m = 0, 1, \dots$  es el contador de iteraciones, y

$$z = P\boldsymbol{\theta} + P \left( P^\top M P \right)^{-1} \begin{pmatrix} X^\top W(y - \mu) \\ s^\top (y - \mu) \end{pmatrix}.$$

Cada ciclo, a través del esquema iterativo (4), consiste en un algoritmo iterativo de mínimos cuadrados ponderados para optimizar la función de log-verosimilitud. Nótese que  $z$  actúa como una variable respuesta modificada en el esquema iterativo (4). Utilizando la ecuación (4) y cualquier software (por ejemplo, MAPLE, MATLAB, OX, R, SAS) con una rutina de regresión lineal ponderada,

se puede calcular el estimador ML  $\hat{\theta} = (\hat{\beta}^\top, \hat{\lambda})^\top$  de  $\theta = (\beta^\top, \lambda)^\top$  iterativamente.

Las iteraciones continúan hasta que se alcanza la convergencia (se debe definir un criterio de parada). Cuando  $n$  es grande y bajo algunas condiciones regulares débiles, tenemos que

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} \stackrel{a}{\sim} \mathcal{N}_{p+1} \left( \begin{bmatrix} \beta \\ \lambda \end{bmatrix}, K(\beta, \lambda)^{-1} \right),$$

donde “ $\stackrel{a}{\sim}$ ” significa distribuido aproximadamente. La distribución normal asintótica anterior puede utilizarse para construir intervalos de confianza aproximados y regiones de confianza para los parámetros. Sea  $\beta_r$  (para  $r = 1, \dots, p$ ) el  $r$ -ésimo componente de  $\beta$ . Para  $0 < \alpha < 1/2$ , los intervalos de confianza asintóticos son

$$\hat{\beta}_r \pm \Phi^{-1}(1 - \alpha/2) \text{se}(\hat{\beta}_r) \quad \text{y} \quad \hat{\lambda} \pm \Phi^{-1}(1 - \alpha/2) \text{se}(\hat{\lambda}),$$

ambos con un nivel de confianza asintótico de  $100(1 - \alpha)\%$ . Aquí,  $\text{se}(\cdot)$  es la raíz cuadrada del elemento diagonal de  $K(\hat{\beta}, \hat{\lambda})^{-1}$  correspondiente a cada parámetro (es decir, el error estándar asintótico), y  $\Phi^{-1}(\cdot)$  es la función cuantil normal estándar.

A continuación, se presenta la formulación del modelo bajo un enfoque Bayesiano.

### 3.3. Inferencia basada en el enfoque Bayesiano

Bajo un enfoque bayesiano, se asume que el vector de parámetros desconocidos es aleatorio y posee una distribución a priori. En términos generales, el modelo bayesiano para la regresión binaria con enlace potencia o reversa de potencia Cauchy es:

$$\begin{aligned} Y_i | \beta, \lambda &\sim \text{Bernoulli}(u_i), \quad i = 1, \dots, n, \\ u_i &= F_\lambda(\eta_i), \\ \eta_i &= \mathbf{x}_i^\top \beta, \\ (\beta, \lambda) &\sim \pi(\beta, \lambda). \end{aligned} \tag{3.2}$$

donde  $F_\lambda(\cdot)$  es la función de distribución acumulada asociada al enlace Potencia Cauchy o Reversa Potencia Cauchy.

Como los parámetros son estimados conjuntamente, lo que se define como estimación no condicional, necesitamos definir una distribución a priori conjunta para estos parámetros.

La función de verosimilitud asociada al modelo es la misma que en el enfoque anterior:

$$L(\beta, \lambda | y, X) = \prod_{i=1}^n \left[ \left( \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \right)^\lambda \right]^{y_i} \left[ 1 - \left( \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \right)^\lambda \right]^{1-y_i},$$

donde  $\eta = \mathbf{x}_i^T \beta$ , para la fda de la distribución potencia Cauchy, y

$$L(\beta, \lambda | y, X) = \prod_{i=1}^n \left[ 1 - \left( \frac{1}{\pi} \arctan(-\eta_i) + \frac{1}{2} \right)^\lambda \right]^{y_i} \left[ \left( \frac{1}{\pi} \arctan(-\eta_i) + \frac{1}{2} \right)^\lambda \right]^{1-y_i},$$

donde  $\eta_i = x_i^\top \beta$ , para la fda de la distribución reversa de potencia Cauchy.

### 3.3.1. Distribución a priori

Según Bazán et al. (2014) es conveniente hacer una reparametrización para el parámetro de forma  $\lambda$ , definiendo  $\delta = \log(\lambda)$  de modo que  $\delta \in \mathbb{R}$ . Luego, se especifica una distribución a priori para  $\beta$  y  $\delta$ , y dado que ambos parámetros son diferentes, se asume independencia entre ellos. Por lo tanto, la distribución a priori conjunta es dada por:

$$\pi(\beta, \delta) = \pi_1(\beta)\pi_2(\delta). \quad (3.3)$$

Como considerado en Bazán et al. (2014), la distribución a priori del vector de coeficientes de regresión  $\beta$  asume una función de distribución Normal, de tal manera que

$$\beta_j \sim \mathcal{N}(\mu_{\beta_j}, \sigma_{\beta_j}^2) \quad j = 1, \dots, p. \quad (3.4)$$

Como se tiene ignorancia a priori sobre  $\beta$ , se emplean hiperparámetros vagos  $\mu_{\beta_j} = 0$  y  $\sigma_{\beta_j}^2 = 10^2$ , de forma análoga a Anyosa (2017). Además, sea  $\delta = \log \lambda$  con a priori  $\delta \sim \mathcal{U}(-2, 2)$ ; así,  $\lambda \in [e^{-2}, e^2]$ , intervalo donde valores fuera son poco probables (Bazán et al., 2017b). Con estas especificaciones, el modelo queda:

$$\begin{aligned} Y_i | \beta, \lambda &\sim \text{Bernoulli}(u_i), \quad i = 1, \dots, n, \\ u_i &= F_\lambda(\eta_i), \\ \eta_i &= \mathbf{x}_i^\top \beta, \\ \beta_j &\sim \mathcal{N}(0, 10^2), \quad j = 1, \dots, p, \\ \delta &\sim \mathcal{U}(-2, 2), \quad \lambda = \exp(\delta). \end{aligned} \quad (3.5)$$

donde  $F_\lambda(\cdot)$  es la fda asociada al enlace Potencia Cauchy o Reversa Potencia Cauchy.

### 3.3.2. Distribución a posteriori

La densidad conjunta de la distribución a posteriori de los modelos de regresión binaria con función de enlace potencia o reversa de potencia Cauchy, tiene la siguiente forma:

$$\begin{aligned} \pi(\beta, \delta | y, X) &\propto L(\beta, \delta | y, X)\pi(\beta)\pi(\delta) \propto \\ &\prod_{i=1}^n [F_\lambda(x_i^\top \beta)]^{y_i} [1 - F_\lambda(x_i^\top \beta)]^{1-y_i} \prod_{j=1}^p \frac{1}{\sqrt{2(10^2)}} \exp\left(-\frac{\beta_j^2}{2(10^2)}\right) \\ &\propto \prod_{i=1}^n [F_\lambda(x_i^\top \beta)]^{y_i} [1 - F_\lambda(x_i^\top \beta)]^{1-y_i} \prod_{j=1}^p \exp\left(-\frac{\beta_j^2}{2(10^2)}\right). \end{aligned} \quad (3.6)$$

---

Donde  $F_\lambda(\cdot)$  es la fda de la distribución potencia o reversa de potencia Cauchy.

Sin embargo, esta distribución a posteriori no pertenece a una familia de distribución conocida y no tiene una forma cerrada, lo que hace que esta distribución sea analíticamente intratable. Una alternativa es el uso de métodos de simulación para obtener la muestra de la distribución a posteriori. En particular, se pueden utilizar los métodos de Monte Carlo vía Cadenas de Markov (Markov Chain Monte Carlo: MCMC).

En este trabajo, utilizamos el algoritmo No-U-Turn Sampler (NUTS), desarrollado por Hoffman y Gelman (2014), el cual es una extensión del algoritmo Monte Carlo Hamiltoniano (HMC). Este método permite a la cadena de Markov explorar eficientemente la distribución a posteriori objetivo, superando las limitaciones de otros algoritmos MCMC como Metropolis-Hastings y Gibbs. NUTS ajusta dinámicamente el tamaño del salto y el número de iteraciones en el espacio de parámetros, evitando "vueltas en U" garantizando una rápida convergencia con baja autocorrelación entre muestras (Neal, 2012). La teoría necesaria para este tema lo veremos en el Anexo 1.

En este trabajo, implementaremos nuestras simulaciones utilizando el lenguaje de programación estadístico R. Para la estimación bayesiana, emplearemos el algoritmo NUTS, integrado en el software Stan.

### 3.4. Criterios de Comparación de Modelos

#### 3.4.1. Criterios de Comparación de Modelos para el método de máxima verosimilitud

En el contexto de la inferencia basada en el método de máxima verosimilitud, la selección de modelos se realiza a través de criterios que balancean el ajuste del modelo y su complejidad. En este estudio, se utilizaron tres criterios clásicos: AIC, BIC y HQIC, que se describen a continuación.

1. **Criterio de Información de Akaike (AIC).** El AIC busca un equilibrio entre el ajuste del modelo y el número de parámetros, penalizando la complejidad para evitar el sobreajuste. Se define como

$$\text{AIC} = -2 \cdot \ell(\hat{\theta}) + 2k,$$

donde  $\ell(\hat{\theta})$  es el valor de la log-verosimilitud evaluada en el estimador de máxima verosimilitud  $\hat{\theta}$ , y  $k$  es el número de parámetros estimados. El modelo que minimiza el AIC es considerado el más adecuado dentro del conjunto evaluado.

2. **Criterio de Información Bayesiano (BIC).** También conocido como el criterio de Schwarz, el BIC penaliza más fuertemente la complejidad del modelo, especialmente cuando el tamaño

---

de muestra es grande. Se define como

$$\text{BIC} = -2 \cdot \ell(\hat{\theta}) + k \cdot \log(n),$$

donde  $n$  representa el tamaño muestral y  $k$  es el número de parámetros estimados. El BIC favorece modelos más parsimoniosos y se prefiere aquel con menor valor de este criterio.

3. **Criterio de Información de Hannan-Quinn (HQIC).** El HQIC es una alternativa que penaliza la complejidad del modelo de manera intermedia entre AIC y BIC. Se define como

$$\text{HQIC} = -2 \cdot \ell(\hat{\theta}) + 2k \cdot \log(\log(n)).$$

Al igual que los criterios anteriores, el modelo con menor valor de HQIC es considerado superior bajo este criterio.

### 3.4.2. Criterios de Comparación de Modelos bayesianos

En el contexto de la selección de modelos bayesianos, es esencial evaluar el ajuste de los modelos usando diferentes criterios que consideren tanto la calidad del ajuste como la complejidad del modelo. A continuación, se presentan los criterios utilizados en este estudio: EAIC, EBIC, DIC y WAIC.

1. **Criterio de Información de Akaike Ampliado (EAIC).** El EAIC es una extensión del criterio de Akaike tradicional, adaptado para modelos bayesianos. Se basa en la media a posteriori de la desviación del modelo y se calcula como

$$\text{EAIC} = \bar{D} + 2P,$$

donde  $\bar{D} = \frac{1}{M} \sum_{m=1}^M D(\theta^{(m)})$  corresponde a la media a posteriori de la desviación, con  $D(\theta^{(m)}) = -2 \sum_{i=1}^n \log f(y_i | \theta^{(m)})$ , y  $\theta^{(m)}$  es la  $m$ -ésima realización de la distribución a posteriori obtenida mediante MCMC. El término  $P$  representa el número de parámetros del modelo. El EAIC selecciona el modelo que minimiza su valor, favoreciendo un equilibrio entre ajuste y simplicidad.

2. **Criterio de Información Bayesiano Ampliado (EBIC).** El EBIC es una extensión bayesiana del criterio de información de Schwarz. Se define como

$$\text{EBIC} = \bar{D} + P \log(n),$$

donde  $\bar{D}$  es la media a posteriori de la desviación,  $n$  es el tamaño de la muestra y  $P$  es el número de parámetros desconocidos del modelo. El EBIC introduce una penalización más fuerte para modelos con un mayor número de parámetros y con un tamaño muestral más grande, lo que lo hace más conservador en la selección de modelos.

3. **Criterio de Información de Deviance (DIC).** El DIC se considera una generalización

bayesiana del criterio de Akaike. Se calcula como

$$\text{DIC} = \bar{D} + \hat{\rho}_D = 2\bar{D} - \hat{D},$$

donde  $\bar{D}$  es la media a posteriori de la desviación y  $\hat{D}$  corresponde a la desviación evaluada en la media de la distribución a posteriori de los parámetros,

$$\hat{D} = D \left( \frac{1}{M} \sum_{m=1}^M \theta^{(m)} \right).$$

El término  $\hat{\rho}_D = \bar{D} - \hat{D}$  representa el número efectivo de parámetros del modelo. El DIC favorece modelos con un buen ajuste, penalizando aquellos con mayor complejidad.

4. **Criterio de Información Ampliamente Aplicable (WAIC).** El WAIC, propuesto por watanabe (2010), es un enfoque completamente bayesiano para la comparación de modelos. Se define como

$$\text{WAIC} = -2 (\text{lppd} - p_{\text{WAIC}}),$$

donde la lppd (log predictive pointwise density) se calcula como

$$\text{lppd} = \sum_{i=1}^n \log \left( \frac{1}{M} \sum_{m=1}^M p(y_i | \theta^{(m)}) \right),$$

y el número efectivo de parámetros,  $p_{\text{WAIC}}$ , se estima mediante

$$p_{\text{WAIC}} = 2 \sum_{i=1}^n \left( \log \left( \frac{1}{M} \sum_{m=1}^M p(y_i | \theta^{(m)}) \right) - \frac{1}{M} \sum_{m=1}^M \log(p(y_i | \theta^{(m)})) \right).$$

El WAIC tiene la ventaja de utilizar toda la distribución a posteriori, lo que lo hace más preciso para la selección de modelos en comparación con criterios que se basan en estimaciones puntuales.

### 3.5. Análisis de Residuos

En el modelo lineal generalizado, el residuo de desviación para datos de respuesta binaria adopta la forma:

$$r_i^d = (y_i - \hat{\mu}_i) \left[ 2y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) + 2(1 - y_i) \log \left( \frac{1 - y_i}{1 - \hat{\mu}_i} \right) \right]^{1/2}, \quad i = 1, \dots, n.$$

En particular, para los modelos de regresión binaria con funciones de enlace potencia Cauchy y reversa de potencia Cauchy, las probabilidades estimadas se definen de la siguiente manera. En el caso del modelo con función de enlace potencia Cauchy, se tiene

$$\hat{\mu}_i = \left( \frac{1}{\pi} \arctan(\hat{\eta}_i) + \frac{1}{2} \right)^\lambda.$$

---

Por otro lado, para el modelo con función de enlace reversa de potencia Cauchy, se utiliza

$$\hat{\mu}_i = 1 - \left( \frac{1}{\pi} \arctan(-\hat{\eta}_i) + \frac{1}{2} \right)^\lambda,$$

donde  $\hat{\eta}_i = x_i^\top \hat{\beta}$  es el predictor lineal estimado para la observación  $i$ .

Con el objetivo de evaluar la adecuación del modelo, Lemonte y Bazán (2018) recomiendan el uso de *residuos cuantílicos aleatorizados normalizados*, propuestos por Dunn y Smyth (1996). Estos residuos están diseñados para que, bajo un modelo correctamente especificado, sigan una distribución normal estándar. La forma general de estos residuos es

$$r_{q,i} = \Phi^{-1}(u_i), \quad i = 1, \dots, n,$$

donde  $\Phi^{-1}(\cdot)$  denota la función cuantil de la distribución normal estándar. Para respuestas binarias,  $u_i$  es un valor aleatorio proveniente de una distribución uniforme en el intervalo

$$[I_{1-\hat{\mu}_i}(2 - y_i, y_i), I_{1-\hat{\mu}_i}(1 - y_i, y_i + 1)], \quad i = 1, \dots, n,$$

donde  $I_x(a, b)$  representa la función beta incompleta regularizada.

De esta forma, los residuos cuantílicos aleatorizados normalizados se adaptan tanto al modelo con función de enlace potencia Cauchy como al de reversa de potencia Cauchy mediante las respectivas definiciones de  $\hat{\mu}_i$ , sin requerir suposiciones adicionales. Si el modelo está correctamente especificado, estos residuos deberían seguir una distribución normal estándar. Los valores de  $\hat{\mu}_i$  y  $\hat{\eta}_i$  pueden ser obtenidos a partir de los estimadores por máxima verosimilitud o de las medias a posteriori bajo el enfoque bayesiano, actuando como estimadores *plug-in*.

## Capítulo 4

# Estudio de Simulación

El objetivo principal de este capítulo es evaluar el desempeño del modelo estudiado, es decir, regresión binaria con función de enlace potencia Cauchy, cuya flexibilidad está determinada por el parámetro  $\lambda$ . La simulación permitirá comparar con otros modelos, examinar de manera controlada la recuperación de parámetros y la calidad de los ajustes.

Los datos se generaron a partir de un modelo de regresión binaria con una sola covariable continua  $x$ , siguiendo la estructura

$$Y_i | \beta, \lambda \sim \text{Bernoulli}(u_i), \quad \forall i = 1, \dots, n,$$

donde

$$u_i = \left( \frac{1}{\pi} \arctan(\eta_i) + \frac{1}{2} \right)^\lambda, \quad \eta_i = \beta_0 + \beta_1 x_i.$$

El diseño contempló tres tamaños muestrales ( $n = 1000, n = 2000$  y  $n = 4000$ ) y tres valores del parámetro ( $\lambda = 2, \lambda = 4, \lambda = 0,25$ ). Para cada combinación de tamaño de muestra y valor de  $\lambda$  se realizaron 100 réplicas, lo cual permite obtener conclusiones estables y comparables entre escenarios.

Inicialmente, dado que  $\lambda$  determina la forma del enlace potencia Cauchy, se evaluó la sensibilidad de las estimaciones ante diferentes elecciones de distribuciones a priori propuestos en la literatura. Este análisis busca identificar cómo la información previa influye en la recuperación de parámetros y en la estabilidad de los resultados de la regresión.

En las siguientes secciones se presentan los resultados obtenidos con los enlaces logit, probit, Cauchy y potencia Cauchy. La evaluación del desempeño incluyó la comparación de curvas estimadas frente a la función verdadera, el cálculo del estadístico de Kolmogorov - Smirnov (KS) sobre las probabilidades predichas y el análisis de la capacidad de recuperación de parámetros.

### 4.1. Análisis de Sensibilidad de la Priori para el Parámetro $\lambda$

Se realizó un análisis de sensibilidad para explorar el uso de tres diferentes distribuciones a priori para el parámetro  $\lambda$ , considerando los modelos de enlace propuestos: potencia Cauchy y reversa de

potencia Cauchy. Las distribuciones a priori evaluadas fueron:  $\log(\lambda) \sim U(-2, 2)$ , propuesta por Bazán et al. (2014);  $\log(\lambda) \sim N(0, 1,5)$ , como sugiere Wang et al. (2010); y  $\lambda \sim \text{Cauchy}(0, 2,5)$ , utilizada en modelos bayesianos de acuerdo con Gelman et al. (2008).

En esta simulación, los parámetros verdaderos fueron fijados en  $\beta_0 = -0,5$ ,  $\beta_1 = 1,5$  y  $\lambda = 4$ , lo que genera un desbalance aproximado del 11.76%. Los parámetros de control más importantes para la generación de los datos y la simulación de Monte Carlo (MCMC) fueron un tamaño de muestra  $N = 1000$ , un número de réplicas  $M = 100$ , y una configuración de 3000 iteraciones en cada ejecución del MCMC, de las cuales 1000 se usaron como calentamiento (*warmup*). Además, se emplearon dos cadenas y un espaciado (*thinning*) de 2.

Con estas condiciones, se obtuvieron los siguientes resultados.

Tabla 4.1: Análisis de sensibilidad a priori para diferentes distribuciones del parámetro  $\lambda$

<b>A Priori</b>	<b>Parámetro</b>	<b>Media a posteriori</b>	<b>SD</b>	<b>Sesgo</b>	<b>RECM</b>
Uniforme	$\beta_0 = -0,5$	-0.507	0.342	-0.007	0.340
	$\beta_1 = 1,5$	1.524	0.099	0.024	0.101
	$\lambda = 4$	4.068	0.765	0.068	0.764
Cauchy	$\beta_0 = -0,5$	-0.543	0.463	-0.043	0.463
	$\beta_1 = 1,5$	1.673	0.113	0.173	0.206
	$\lambda = 4$	4.582	1.790	0.582	1.874
Normal Logarítmica	$\beta_0 = -0,5$	-1.182	0.304	-0.682	0.746
	$\beta_1 = 1,5$	1.813	0.105	0.313	0.330
	$\lambda = 4$	2.679	0.427	-1.321	1.388

Basándose en los resultados presentados en la tabla, se puede analizar el desempeño de cada priori en términos del sesgo y de la raíz del error cuadrático medio (RECM) para los parámetros  $\beta_0$ ,  $\beta_1$  y  $\lambda$ .

En cuanto al sesgo, la priori Uniforme muestra el valor más bajo para  $\beta_0$ , seguida por la Cauchy y, en último lugar, la Normal Logarítmica. Para el parámetro  $\beta_1$ , la Uniforme también presenta el menor sesgo, lo que indica una mejor recuperación del valor verdadero frente a las otras dos prioris. En el caso de  $\lambda$ , el sesgo más bajo corresponde a la Normal Logarítmica, aunque su valor sigue siendo relativamente alto en comparación con los sesgos observados en  $\beta_0$  y  $\beta_1$ .

Si se considera el RECM, la priori Uniforme tiene el valor más bajo para  $\beta_0$ , lo que sugiere un mejor ajuste. La priori Cauchy presenta un RECM mayor, mientras que la Normal Logarítmica alcanza el más alto, evidenciando mayor variabilidad en la estimación de este parámetro. Para  $\beta_1$ , la Uniforme vuelve a ser la que logra el menor RECM, confirmando su buen desempeño. En el caso de  $\lambda$ , el valor más bajo de RECM se observa en la Normal Logarítmica, seguida de la Uniforme, mientras que la Cauchy registra el peor resultado.

En conjunto, estos hallazgos permiten concluir que la priori Uniforme se destaca en la estimación de  $\beta_0$  y  $\beta_1$ , al presentar menor sesgo y menor RECM, lo que la convierte en la más confiable para la recuperación de los parámetros estructurales del modelo. Aunque la Normal Logarítmica muestra un desempeño más favorable para  $\lambda$ , su comportamiento en los parámetros  $\beta_0$  y  $\beta_1$  es menos satisfactorio. Por lo tanto, considerando tanto el sesgo como el RECM, la priori Uniforme puede considerarse la

más adecuada para el modelo con función de enlace potencia Cauchy, ya que ofrece un mejor balance en la estimación de los parámetros principales.

## 4.2. Desempeño de la Función de Enlace Potencia Cauchy

### 4.2.1. Comparación de curvas estimadas

Se compara la función de enlace Potencia Cauchy (PC) frente a los enlaces simétricos *logit*, *probit* y *Cauchy*. Los datos se generaron bajo el modelo verdadero PC con  $n = 4000$  y  $B = 100$  réplicas, fijando  $\beta_0 = -0,5$ ,  $\beta_1 = 1,5$  y  $x \sim \mathcal{N}(0, 1)$ . Se evaluaron tres escenarios para el parámetro de asimetría:  $\lambda = 0,25$ ,  $\lambda = 2$  y  $\lambda = 4$ . En cada réplica se ajustaron los cuatro modelos (PC con  $\lambda$  igual al verdadero y los tres enlaces simétricos) y se obtuvieron, por un lado, las *curvas promedio* de  $\hat{\mu}(x)$  con bandas punto-a-punto (2.5% - 97.5%) y, por otro, la *comparación de distribuciones* de las probabilidades estimadas  $\hat{\mu}$  mediante la prueba de Kolmogorov - Smirnov (KS), contrastando PC contra cada enlace simétrico.

La hipótesis nula del contraste KS en cada réplica es  $H_0 : F_{\hat{\mu}_{PC}} = F_{\hat{\mu}_{link}}$  frente a  $H_1 : F_{\hat{\mu}_{PC}} \neq F_{\hat{\mu}_{link}}$ , con nivel  $\alpha = 0,05$ . Se resume, para cada par, la proporción de rechazos, la mediana del estadístico  $D$  y su percentil 90 ( $D_{0,90}$ ) a lo largo de las  $B$  réplicas.

#### Escenario $\lambda = 0,25$

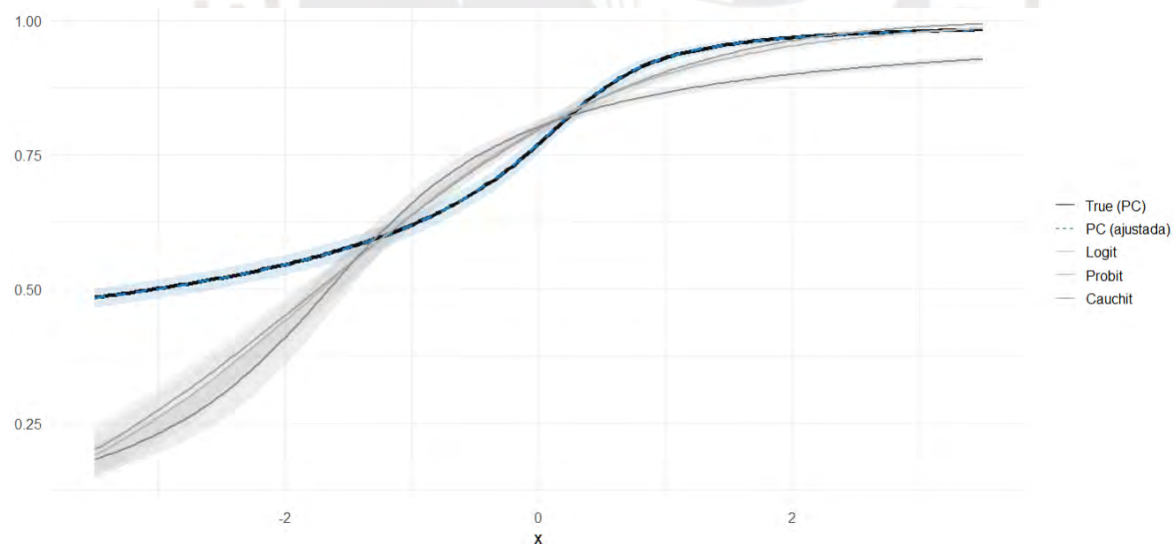


Figura 4.1: Curvas promedio y bandas (100 simulaciones) - Verdadero: PC,  $\lambda = 0,25$ . La curva verdadera (PC) y la PC ajustada se superponen casi totalmente; los enlaces simétricos se desvían principalmente en las colas.

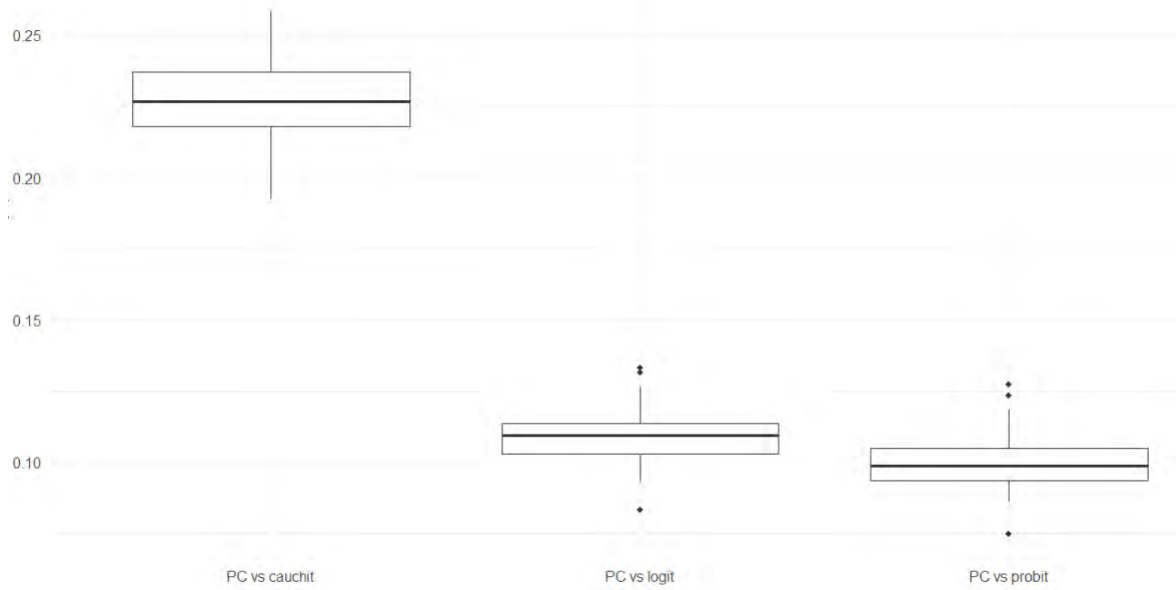


Figura 4.2: Boxplots del estadístico KS - PC vs enlaces simétricos,  $\lambda = 0,25$ .

Tabla 4.2: KS ( $\lambda = 0,25$ ): proporción de rechazos ( $\alpha = 0,05$ ), mediana de  $D$  y  $D_{0,90}$  en  $B = 100$  réplicas.

Par	Prop. rechazo	D mediana	$D_{0,90}$
PC vs Cauchit	1.00	0.227	0.246
PC vs Logit	1.00	0.109	0.119
PC vs Probit	1.00	0.099	0.111

**Escenario  $\lambda = 2$**

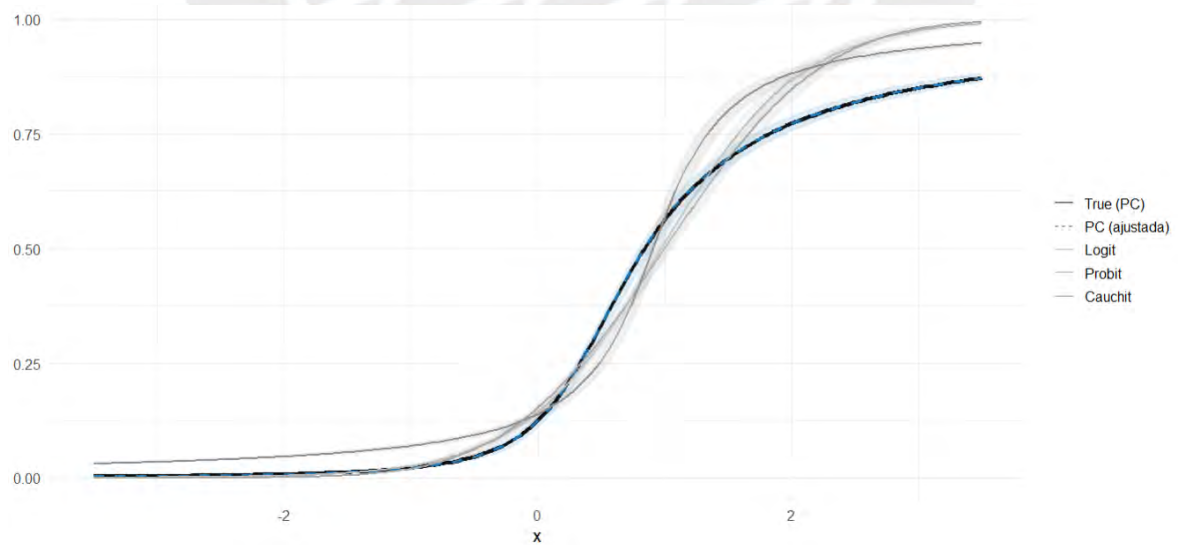


Figura 4.3: Curvas promedio y bandas (100 simulaciones) - Verdadero: PC,  $\lambda = 2$ . La PC ajustada sigue a la curva verdadera; los simétricos muestran desvíos crecientes en zonas de transición y colas.

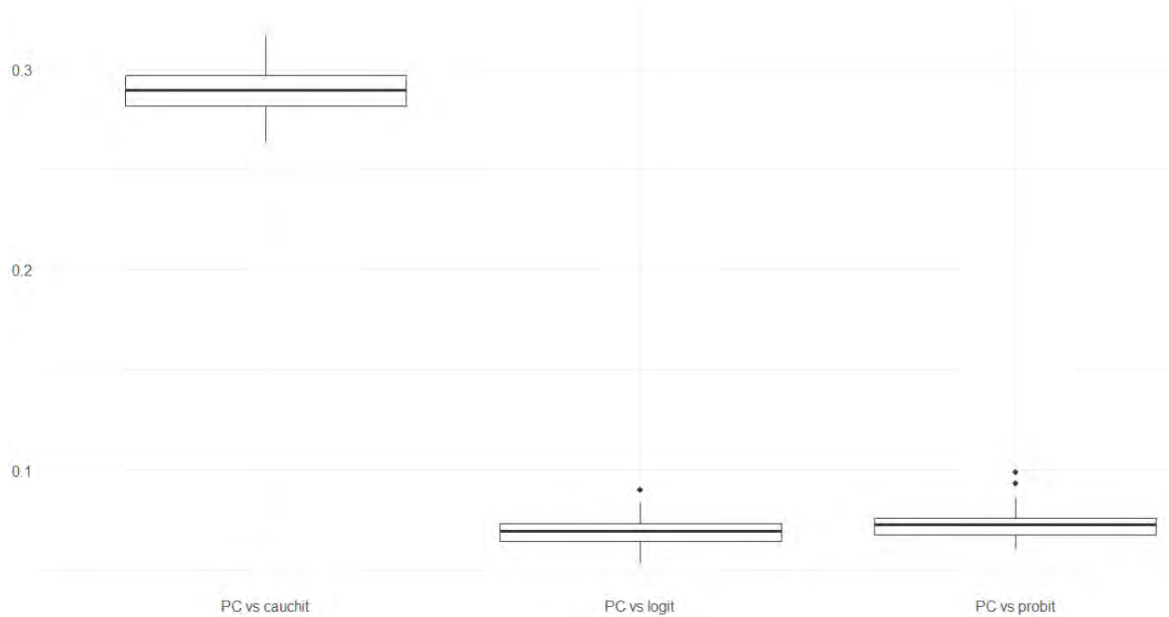


Figura 4.4: Boxplots del estadístico KS - PC vs enlaces simétricos,  $\lambda = 2$ .

Tabla 4.3: KS ( $\lambda = 2$ ): proporción de rechazos ( $\alpha = 0,05$ ), mediana de  $D$  y  $D_{0,90}$  en  $B = 100$  réplicas.

Par	Prop. rechazo	D mediana	$D_{0,90}$
PC vs Cauchit	1.00	0.290	0.305
PC vs Logit	1.00	0.070	0.078
PC vs Probit	1.00	0.073	0.079

#### Escenario $\lambda = 4$

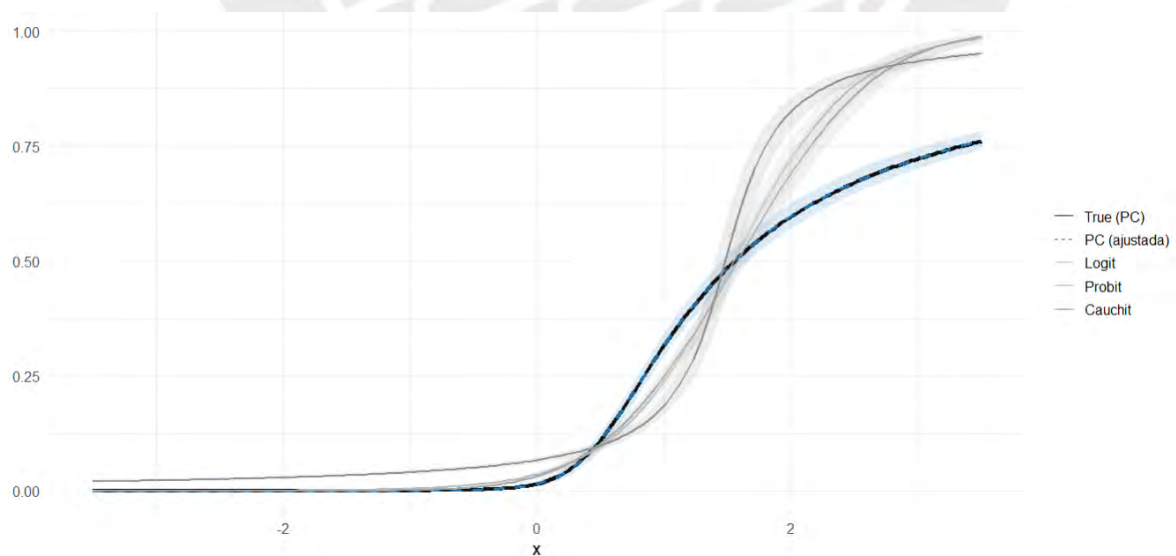


Figura 4.5: Curvas promedio y bandas (100 simulaciones) - Verdadero: PC,  $\lambda = 4$ . La PC ajustada se superpone con la curva verdadera; los enlaces simétricos se separan claramente, en especial *Cauchy*.

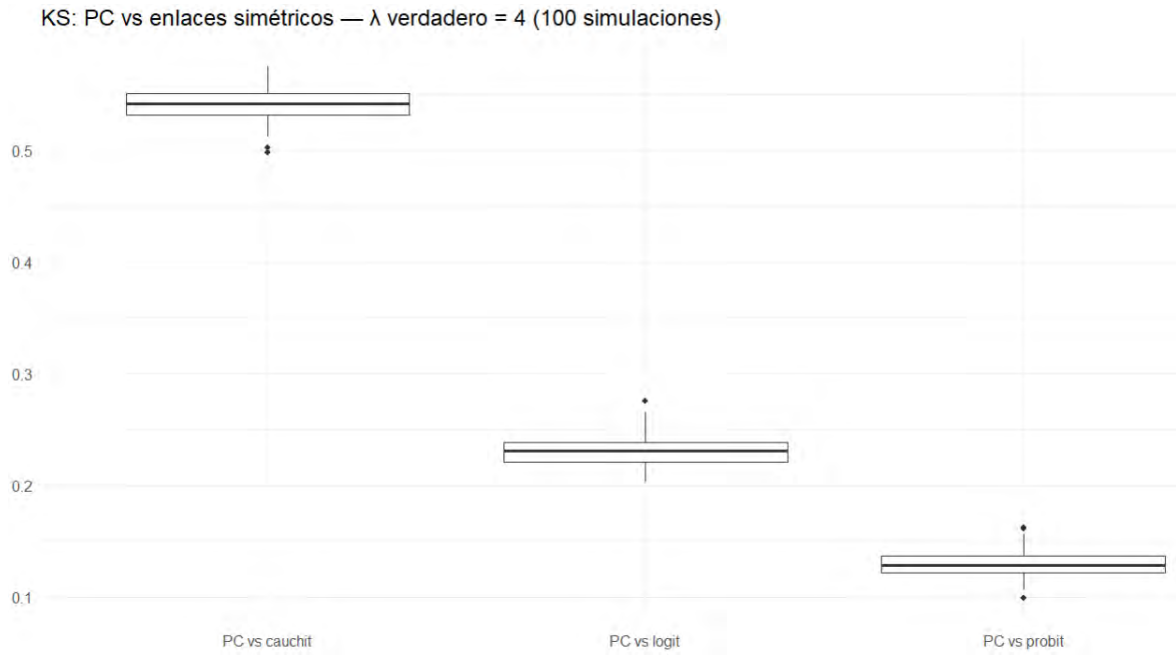


Figura 4.6: Boxplots del estadístico KS - PC vs enlaces simétricos,  $\lambda = 4$ .

Tabla 4.4: KS ( $\lambda = 4$ ): proporción de rechazos ( $\alpha = 0,05$ ), mediana de  $D$  y  $D_{0,90}$  en  $B = 100$  réplicas.

Par	Prop. rechazo	D mediana	$D_{0,90}$
PC vs Cauchit	1.00	0.542	0.560
PC vs Logit	1.00	0.231	0.253
PC vs Probit	1.00	0.128	0.146

En los tres escenarios evaluados ( $n = 4000$ ,  $B = 100$ ,  $\lambda \in \{0,25, 2, 4\}$ ), la curva verdadera (PC) y la PC ajustada prácticamente se superponen a lo largo de todo el rango de  $x$ , lo que confirma que el procedimiento de estimación recupera la forma correcta de la probabilidad media cuando el enlace verdadero es asimétrico. Los enlaces simétricos (logit, probit y Cauchy) se apartan de la referencia sobre todo en las colas, y esta separación se intensifica a medida que crece  $\lambda$ .

La comparación de *distribuciones* de probabilidades estimadas mediante la prueba de Kolmogorov - Smirnov (KS) muestra diferencias sistemáticas entre PC y cada enlace simétrico: la proporción de rechazos es 1,00 en todos los pares y para los tres valores de  $\lambda$ . La magnitud típica de la discrepancia, medida por la mediana del estadístico  $D$ , aumenta con  $\lambda$  y sigue un patrón consistente: el mayor desacople se observa frente a *Cauchy*, luego *logit*, y *probit* es el que más se aproxima a PC, aunque sigue siendo estadísticamente distinto. En particular, para  $\lambda = 0,25$  se obtiene  $D_{\text{med}} \approx 0,227$  (PC vs Cauchy), 0,109 (PC vs logit) y 0,099 (PC vs probit); para  $\lambda = 2$ ,  $D_{\text{med}} \approx 0,290$ , 0,070 y 0,073, respectivamente; y para  $\lambda = 4$ ,  $D_{\text{med}} \approx 0,542$ , 0,231 y 0,128. Los boxplots de  $D$  por par de modelos reflejan estas diferencias en nivel y su variabilidad entre réplicas.

En conclusión. Cuando los datos se generan con un enlace Potencia Cauchy, el modelo PC reproduce con precisión la relación entre  $x$  y la probabilidad de éxito, mientras que los enlaces simétricos

---

presentan discrepancias de forma que se traducen en diferencias en la *distribución* de  $\hat{\mu}$  detectables por KS. Estas diferencias son especialmente relevantes frente a *Cauchy* y se vuelven más pronunciadas para valores grandes de  $\lambda$ . En escenarios desbalanceados con asimetría marcada, la función de enlace Potencia Cauchy ofrece un ajuste superior a los enlaces simétricos tradicionales.

#### 4.2.2. Recuperación de parámetros

Este estudio de simulación se desarrolló para evaluar el desempeño de diferentes funciones de enlace en modelos de regresión binaria con datos desbalanceados. Se compararon cinco funciones de enlace: logit, probit, Cauchy, potencia Cauchy y reversa de potencia Cauchy, con el objetivo de analizar su comportamiento en términos de sesgo, raíz del error cuadrático medio (RECM) y asimetría.

Las funciones de enlace logit, probit y Cauchy son simétricas y se usan comúnmente en la literatura, mientras que las funciones de enlace potencia Cauchy y reversa de potencia Cauchy son asimétricas y permiten ajustar diferentes niveles de desbalance en los datos mediante el parámetro de forma  $\lambda$ .

El estudio considera tres tamaños de muestra ( $n = 1000$ ,  $n = 2000$  y  $n = 4000$ ), y tres valores de  $\lambda$  ( $\lambda = 2$ ,  $\lambda = 4$ ,  $\lambda = 0,25$ ), aplicados únicamente a las funciones de enlace de potencia Cauchy y reversa de potencia Cauchy. Para cada combinación de tamaño de muestra y valor de  $\lambda$ , se realizaron 100 réplicas.

Los parámetros verdaderos en el análisis fueron  $\beta_0 = -0,5$  y  $\beta_1 = 1,5$ , mientras que la covariable se generó siguiendo una distribución normal estándar ( $x \sim N(0,1)$ ). Los datos desbalanceados se construyeron ajustando la probabilidad de éxito en cada observación, lo que permitió observar el comportamiento de las estimaciones en distintos niveles de desbalance, reflejados en la asimetría de las distribuciones resultantes.

Para el cálculo de la asimetría en las funciones de enlace simétricas (logit, probit y Cauchy) se fijó  $\lambda = 1$ , ya que estas distribuciones no dependen de dicho parámetro en su forma estándar.

El desempeño de los modelos se evaluó considerando varias métricas. En primer lugar, se analizó la estimación de los coeficientes  $\beta$  a partir del promedio obtenido en las réplicas. En segundo lugar, se calculó el sesgo, entendido como la diferencia promedio entre las estimaciones y el valor verdadero de los parámetros. En tercer lugar, se utilizó la raíz del error cuadrático medio (RECM) para medir la precisión de las estimaciones. Finalmente, se evaluó la asimetría como medida de la desviación respecto a la simetría en la distribución de las estimaciones.

A continuación, se presentan seis tablas que muestran los resultados de las simulaciones para las estimaciones de  $\beta_0$  y  $\beta_1$ , con las funciones de enlace evaluadas en los tres tamaños de muestra y los distintos valores de  $\lambda$ . En el caso de las funciones de enlace simétricas (logit, probit y Cauchy), se utilizó un  $\lambda = 1$  para el cálculo de la asimetría, dado que estas distribuciones no dependen de este parámetro.

Tabla 4.5: Resultados de  $\hat{\beta}_0$  para  $\lambda = 2$

Modelo	$n$	$\hat{\beta}_0$	Sesgo	RECM	$\hat{\lambda}$	Asimetría
Logit	1000	-1.789	-1.289	1.291	1.0	0.0
	2000	-1.503	-1.003	1.004	1.0	0.0
	4000	-1.398	-0.898	0.898	1.0	0.0
Probit	1000	-0.990	-0.490	0.504	1.0	0.0
	2000	-1.191	-0.691	0.693	1.0	0.0
	4000	-0.987	-0.487	0.493	1.0	0.0
Cauchy	1000	-2.010	-1.510	1.527	1.0	0.0
	2000	-1.728	-1.228	1.248	1.0	0.0
	4000	-1.514	-1.014	1.019	1.0	0.0
Pot. Cauchy	1000	-0.499	0.001	0.075	1.981	0.181
	2000	-0.502	-0.002	0.029	1.988	0.183
	4000	-0.501	-0.001	0.020	1.981	0.181
Rev. Pot. Cauchy	1000	-0.533	-0.033	0.507	0.744	0.564
	2000	-0.494	0.006	0.454	0.736	0.572
	4000	-0.526	-0.026	0.212	0.742	0.568

Tabla 4.6: Resultados de  $\hat{\beta}_1$  para  $\lambda = 2$

Modelo	$n$	$\hat{\beta}_1$	Sesgo	RECM	$\hat{\lambda}$	Asimetría
Logit	1000	1.576	0.076	0.127	1.0	0.0
	2000	1.100	-0.400	0.408	1.0	0.0
	4000	0.995	-0.505	0.508	1.0	0.0
Probit	1000	0.979	-0.521	0.555	1.0	0.0
	2000	1.006	-0.494	0.512	1.0	0.0
	4000	1.097	-0.403	0.410	1.0	0.0
Cauchy	1000	2.174	0.674	0.696	1.0	0.0
	2000	1.911	0.411	0.431	1.0	0.0
	4000	1.804	0.304	0.327	1.0	0.0
Pot. Cauchy	1000	1.505	0.005	0.107	1.981	0.181
	2000	1.498	-0.002	0.056	1.988	0.183
	4000	1.502	0.002	0.036	1.981	0.181
Rev. Pot. Cauchy	1000	1.563	0.063	0.516	0.744	0.564
	2000	1.398	-0.102	0.416	0.736	0.572
	4000	1.503	0.003	0.293	0.742	0.568

Tabla 4.7: Resultados de  $\hat{\beta}_0$  para  $\lambda = 4$

Modelo	$n$	$\hat{\beta}_0$	Sesgo	RECM	$\hat{\lambda}$	Asimetría
Logit	1000	-1.789	-1.289	1.291	1.0	0.0
	2000	-1.503	-1.003	1.004	1.0	0.0
	4000	-1.398	-0.898	0.898	1.0	0.0
Probit	1000	-0.990	-0.490	0.504	1.0	0.0
	2000	-1.191	-0.691	0.693	1.0	0.0
	4000	-0.987	-0.487	0.493	1.0	0.0
Cauchy	1000	-2.010	-1.510	1.527	1.0	0.0
	2000	-1.728	-1.228	1.248	1.0	0.0
	4000	-1.514	-1.014	1.019	1.0	0.0
Pot. Cauchy	1000	-0.454	0.046	0.369	4.005	0.181
	2000	-0.525	-0.025	0.390	3.980	0.183
	4000	-0.482	0.018	0.196	4.024	0.181
Rev. Pot. Cauchy	1000	-0.389	0.111	0.779	0.489	0.564
	2000	-0.590	-0.090	0.939	0.496	0.572
	4000	-0.530	-0.030	0.420	0.506	0.568

Tabla 4.8: Resultados de  $\hat{\beta}_1$  para  $\lambda = 4$

Modelo	$n$	$\hat{\beta}_1$	Sesgo	RECM	$\hat{\lambda}$	Asimetría
Logit	1000	1.576	0.076	0.127	1.0	0.0
	2000	1.100	-0.400	0.408	1.0	0.0
	4000	0.995	-0.505	0.508	1.0	0.0
Probit	1000	0.979	-0.521	0.555	1.0	0.0
	2000	1.006	-0.494	0.512	1.0	0.0
	4000	1.097	-0.403	0.410	1.0	0.0
Cauchy	1000	2.174	0.674	0.696	1.0	0.0
	2000	1.911	0.411	0.431	1.0	0.0
	4000	1.804	0.304	0.327	1.0	0.0
Pot. Cauchy	1000	1.523	0.023	0.405	4.005	0.181
	2000	1.479	-0.021	0.430	3.980	0.183
	4000	1.524	0.024	0.181	4.024	0.181
Rev. Pot. Cauchy	1000	1.506	0.006	0.596	0.489	0.564
	2000	1.448	-0.052	0.532	0.496	0.572
	4000	1.503	0.003	0.472	0.506	0.568

Tabla 4.9: Resultados de  $\hat{\beta}_0$  para  $\lambda = 0,25$

Modelo	$n$	$\hat{\beta}_0$	Sesgo	RECM	$\hat{\lambda}$	Asimetría
Logit	1000	-0.789	-0.289	0.291	1.0	0.0
	2000	-0.503	-0.003	0.004	1.0	0.0
	4000	-0.398	-0.098	0.098	1.0	0.0
Probit	1000	-0.490	-0.240	0.245	1.0	0.0
	2000	-0.191	-0.091	0.093	1.0	0.0
	4000	-0.187	-0.087	0.093	1.0	0.0
Cauchy	1000	-1.010	-0.510	0.527	1.0	0.0
	2000	-0.728	-0.228	0.248	1.0	0.0
	4000	-0.514	-0.014	0.019	1.0	0.0
Pot. Cauchy	1000	-0.505	-0.005	0.052	0.275	0.217
	2000	-0.500	0.000	0.038	0.270	0.220
	4000	-0.504	-0.004	0.025	0.279	0.225
Rev. Pot. Cauchy	1000	-0.482	-0.018	0.128	4.227	0.980
	2000	-0.494	-0.006	0.061	4.220	0.978
	4000	-0.497	-0.003	0.056	4.278	0.985

Tabla 4.10: Resultados de  $\hat{\beta}_1$  para  $\lambda = 0,25$

Modelo	$n$	$\hat{\beta}_1$	Sesgo	RECM	$\hat{\lambda}$	Asimetría
Logit	1000	1.576	0.076	0.127	1.0	0.0
	2000	1.100	-0.400	0.408	1.0	0.0
	4000	0.995	-0.505	0.508	1.0	0.0
Probit	1000	0.979	-0.521	0.555	1.0	0.0
	2000	1.006	-0.494	0.512	1.0	0.0
	4000	1.097	-0.403	0.410	1.0	0.0
Cauchy	1000	2.174	0.674	0.696	1.0	0.0
	2000	1.911	0.411	0.431	1.0	0.0
	4000	1.804	0.304	0.327	1.0	0.0
Pot. Cauchy	1000	1.498	-0.002	0.104	0.275	0.217
	2000	1.489	-0.011	0.074	0.270	0.220
	4000	1.501	0.001	0.042	0.279	0.225
Rev. Pot. Cauchy	1000	1.505	0.005	0.106	4.227	0.980
	2000	1.499	-0.001	0.082	4.220	0.978
	4000	1.493	-0.007	0.070	4.278	0.985

Basándonos en los resultados presentados en las tablas, se puede evaluar el desempeño de las funciones de enlace en términos del sesgo, la raíz del error cuadrático medio (RECM) y la asimetría para los parámetros  $\beta_0$ ,  $\beta_1$  y  $\lambda$ .

En lo referente al sesgo, se observa que en  $\beta_0$  las funciones de enlace simétricas (logit, probit y Cauchy) exhiben valores considerablemente superiores en comparación con las funciones de enlace potencia Cauchy y reversa de potencia Cauchy, diferencia que resulta especialmente notoria en tamaños de muestra más pequeños. Para  $\beta_1$ , el comportamiento es similar, ya que las funciones de enlace potencia Cauchy y reversa de potencia Cauchy muestran un sesgo más reducido, lo que sugiere estimaciones más acertadas en contextos con datos desbalanceados. En el caso de  $\lambda$ , que aplica úni-

---

camente a las funciones de enlace potencia Cauchy y reversa de potencia Cauchy, el sesgo varía con el valor de este parámetro, aunque la potencia Cauchy tiende a mostrar un desempeño más estable.

En cuanto al RECM, se aprecia que para  $\beta_0$  los valores más bajos corresponden nuevamente a las funciones de enlace potencia Cauchy y reversa de potencia Cauchy, lo que refleja una mayor precisión en la estimación. Para  $\beta_1$ , ambas funciones también mantienen un mejor desempeño, confirmando su capacidad de manejar de forma más efectiva el desbalance en los datos. Con respecto a  $\lambda$ , los resultados indican que el RECM tiende a ser más bajo en la función potencia Cauchy, sobre todo cuando  $\lambda$  toma valores pequeños como 0.25. Sin embargo, la función reversa de potencia Cauchy también presenta un desempeño satisfactorio, particularmente en tamaños de muestra mayores.

Respecto a la asimetría, se advierte que, al estar definida en función de las estimaciones de  $\lambda$ , las funciones potencia Cauchy y reversa de potencia Cauchy presentan valores significativos de asimetría, lo cual evidencia su capacidad para capturar las características propias de datos desbalanceados. Este comportamiento es consistente con lo esperado y resulta útil en aplicaciones donde la distribución de la variable respuesta no es simétrica. Por el contrario, en las funciones simétricas como logit, probit y Cauchy, la asimetría es nula, lo que limita su capacidad para adaptarse a tales escenarios.

Finalmente, integrando los resultados de la comparación de modelos y de la recuperación de parámetros, se confirma que cuando el mecanismo generador es asimétrico (Potencia Cauchy), el modelo PC reproduce con alta fidelidad la relación entre  $x$  y la probabilidad de éxito a lo largo de todo el rango, mientras que los enlaces simétricos (*logit*, *probit* y *cauchy*) se apartan sobre todo en las colas. Esta diferencia de forma se refleja asimismo en la *distribución* de las probabilidades estimadas: las pruebas de Kolmogorov - Smirnov rechazan sistemáticamente la igualdad entre PC y cada enlace simétrico, con distancias  $D$  típicas que crecen al aumentar  $\lambda$  y con el mayor desacople frente a *cauchy*. En términos de recuperación de parámetros, PC ofrece sesgos y RECM marcadamente menores para  $\beta_0$  y  $\beta_1$  en todos los tamaños muestrales y niveles de asimetría considerados, mientras que los enlaces simétricos mantienen sesgos residuales que se atenúan más lentamente con  $n$ . En conjunto, la evidencia apoya el uso de Potencia Cauchy como especificación de referencia en escenarios desbalanceados con asimetría, quedando los enlaces simétricos como comparadores de sensibilidad; cuando se requiera modelar asimetrías de signo opuesto, la reversa Potencia Cauchy constituye una alternativa análoga.

# Capítulo 5

## Aplicación

### 5.1. Descripción del Conjunto de Datos

En este capítulo se presenta una aplicación práctica de la metodología propuesta para el pronóstico de incumplimiento crediticio. La base de datos utilizada proviene de una entidad financiera dedicada a la otorgación de créditos y servicios de financiamiento. Inicialmente, la base contaba con 31,567 observaciones y 16 variables. Luego de un proceso de limpieza y transformación, se obtuvo un conjunto de datos final compuesto por 21,018 observaciones y 10 variables.

#### 5.1.1. Variables Consideradas

La variable de interés es *Default*, que clasifica a los clientes en dos categorías: 1 cuando el cliente presenta incumplimiento de pagos (cliente malo) y 0 cuando no lo presenta (cliente bueno). Este problema se caracteriza por un **desbalance significativo** en la variable respuesta: un 81.31 % corresponde a clientes buenos (*Default* = 0) y solo un 18.69 % a clientes malos (*Default* = 1).

Por otro lado, las variables predictoras utilizadas describen el comportamiento crediticio de los clientes y se detallan a continuación:

- *AntigüedadCDK\_Hist*: Máximo número de meses de antigüedad del cliente en la institución.
- *Max\_MontoDesembolsado\_Hist*: Monto máximo desembolsado históricamente al cliente (en soles).
- *Max\_Tasa\_Anuual*: Tasa anual de interés máxima registrada para el cliente.
- *Max\_NumeroCuotas*: Número máximo de cuotas otorgadas en un crédito.
- *MaxDiasAtraso\_12m*: Número máximo de días de atraso en los últimos 12 meses.
- *TieneTC\_12m*: Indicador binario que señala si el cliente tuvo una tarjeta de crédito en los últimos 12 meses (1 = Sí, 0 = No).
- *MaxNumeroEntidades\_12m*: Número máximo de entidades financieras con las que el cliente interactuó en los últimos 12 meses.

- *REDUCCIONU12M*: Número de meses en los que el cliente redujo su deuda en los últimos 12 meses.
- *MaximoSaldoUlt12M*: Saldo máximo de deuda registrado en los últimos 12 meses (en soles).

La siguiente tabla presenta un resumen estadístico de las variables utilizadas en el análisis, incluyendo los valores mínimos, máximos, medias, desviaciones estándar, asimetría y curtosis.

Tabla 5.1: Estadísticas descriptivas del conjunto de datos.

Variable	Mín	Media	Máx	Asimetría	Curtosis	Desv.
<i>Default</i>	0.00	0.19	1.00	1.61	0.58	0.39
<i>AntiguedadCDK_Hist</i>	0.00	37.11	133.00	0.89	-0.09	31.06
<i>Max_MontoDesembolsado_Hist</i>	0.50	11.27	100.00	3.12	10.96	16.82
<i>Max_Tasa_Anuual</i>	2.29	37.95	83.00	0.09	0.73	16.27
<i>Max_NumeroCuotas</i>	1.00	16.86	84.00	1.36	2.68	9.92
<i>MaxDiasAtraso_12m</i>	1.00	1.41	3.00	1.46	0.22	0.78
<i>TieneTC_12m</i>	0.00	0.09	1.00	2.87	6.26	0.29
<i>MaxNumeroEntidades_12m</i>	1.00	2.00	6.00	1.02	0.76	1.04
<i>REDUCCIONU12M</i>	0.00	7.09	12.00	-0.69	-0.66	3.40
<i>MaximoSaldoUlt12M</i>	0.07	27.74	783.78	4.19	26.53	46.79

Como se observa, algunas variables como *Max\_MontoDesembolsado\_Hist* y *MaximoSaldoUlt12M* presentan una alta asimetría y curtosis, lo que indica la presencia de valores extremos. Adicionalmente, la variable *Default* muestra un desbalance importante, lo que justifica la necesidad de utilizar funciones de enlace asimétricas, como la potencia Cauchy y su reversa, para modelar la relación entre las variables predictoras y la variable de interés.

La siguiente figura muestra la distribución de la variable respuesta *Default*. Se confirma el desbalance de clases, con una proporción del 18.69% de clientes malos (*Default* = 1).

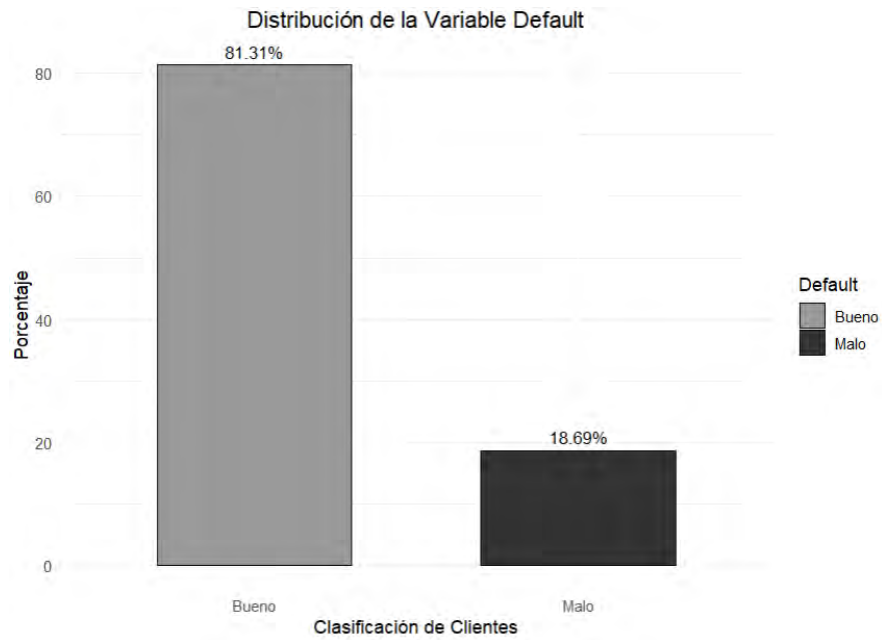


Figura 5.1: Distribución de la variable respuesta *Default*.

La siguiente figura presenta la matriz de correlación entre las variables predictoras. Se observa que no existen correlaciones extremadamente altas entre las variables.

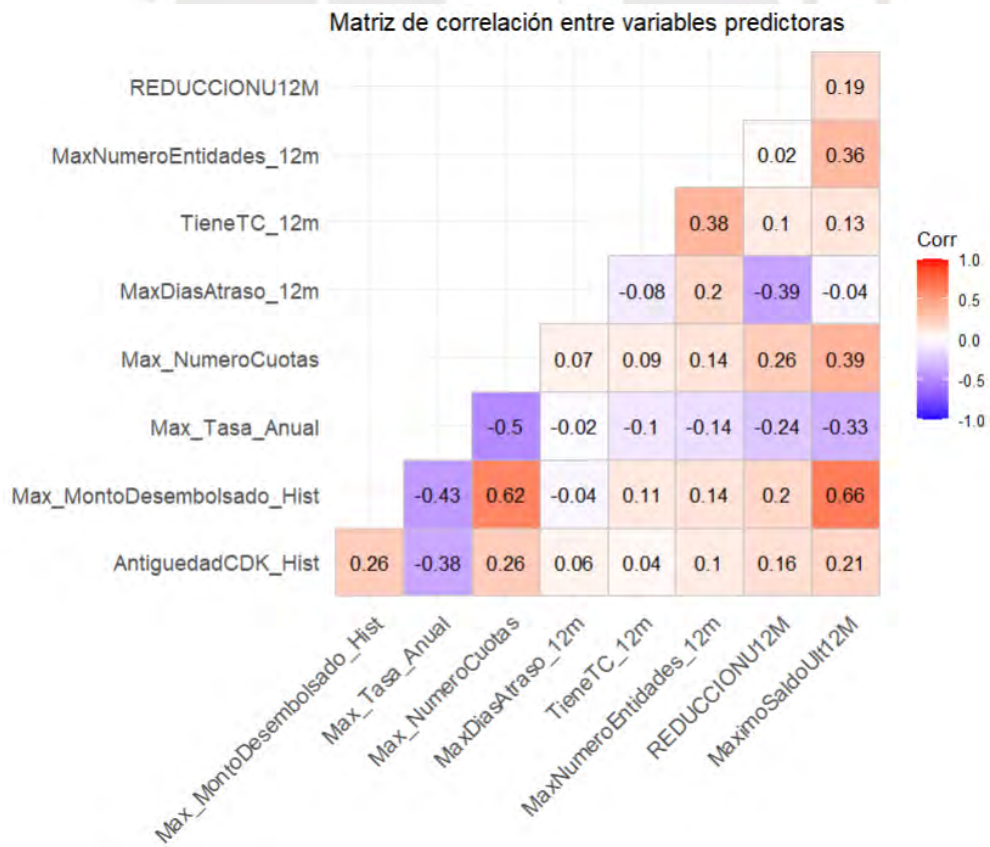


Figura 5.2: Matriz de correlación entre las variables predictoras.

La siguiente tabla muestra los valores VIF calculados para las variables predictoras. Valores de

VIF inferiores a 5 indican que no existen problemas significativos de multicolinealidad.

Tabla 5.2: Valores VIF de las variables predictoras.

Variable	VIF
<i>AntiguedadCDK_Hist</i>	1.196
<i>Max_Tasa_Anual</i>	1.501
<i>Max_NumeroCuotas</i>	1.512
<i>MaxDiasAtraso_12m</i>	1.334
<i>TieneTC_12m</i>	1.210
<i>MaxNumeroEntidades_12m</i>	1.432
<i>REDUCCIONU12M</i>	1.347
<i>MaximoSaldoUlt12M</i>	1.407

La siguiente figura muestra los intervalos HPD (Highest Posterior Density) al 95 % para los coeficientes de regresión asociados a las variables predictoras.

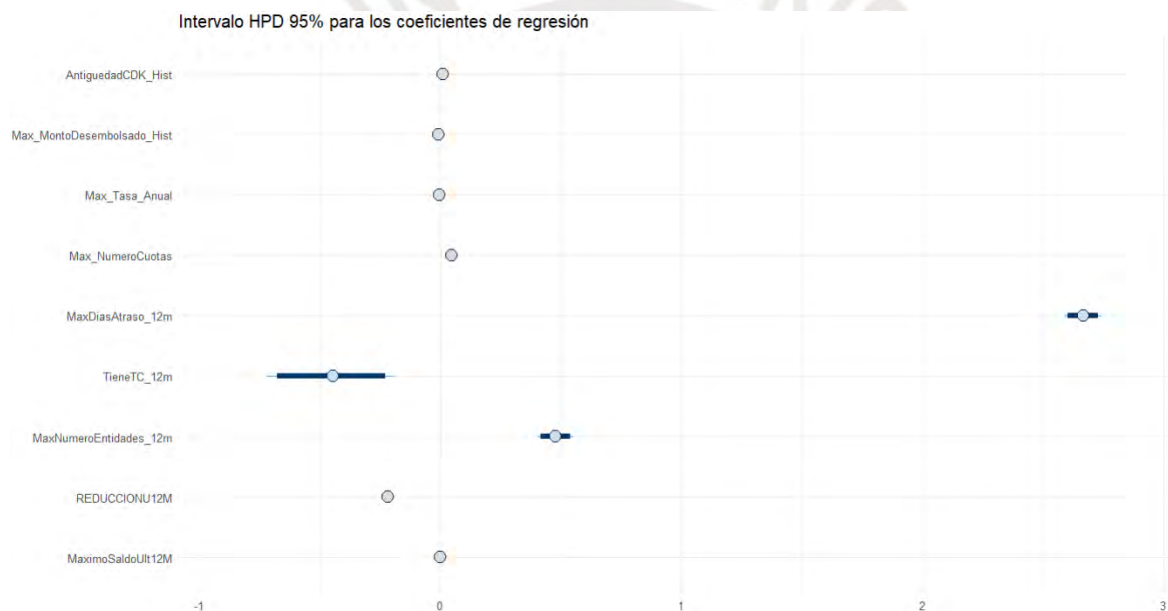


Figura 5.3: Intervalo HPD 95 % para los coeficientes de regresión.

Este gráfico confirma la relevancia de todas las variables incluidas en el modelo, ya que los intervalos HPD no incluyen el valor cero de manera consistente en la mayoría de los predictores.

El análisis preliminar evidenció tres aspectos fundamentales. En primer lugar, se constató la presencia de un desbalance considerable en la variable respuesta. En segundo lugar, no se detectaron problemas de multicolinealidad entre las variables predictoras. Finalmente, se confirmó que todas las variables predictoras resultan relevantes y, por tanto, se mantienen en el modelo debido a su aporte significativo a la precisión de las estimaciones.

## 5.2. Análisis de modelos mediante inferencia bayesiana

En esta sección se evaluaron cinco modelos de regresión binaria utilizando distintas funciones de enlace: logit, probit, Cauchy, Potencia Cauchy y Reversa Potencia Cauchy. Los modelos se ajustaron bajo un enfoque de inferencia bayesiana, con el objetivo de estimar la probabilidad de incumplimiento crediticio en función de un conjunto de covariables. Las estimaciones se realizaron mediante muestreo MCMC a partir de modelos especificados en el lenguaje `Stan`, y se utilizaron criterios de información adaptados al marco bayesiano (DIC, EAIC, EBIC y WAIC) para comparar el desempeño de los modelos, tal como se detalló en el capítulo metodológico.

La siguiente tabla muestra los resultados de los criterios de información para cada modelo evaluado.

Tabla 5.3: Criterios de Información para la Comparación de Modelos

Modelo	DIC	EAIC	EBIC	WAIC
Logístico	7908.622	7916.708	7988.286	7908.677
Probit	8205.053	8212.268	8283.846	8205.159
Cauchy	7712.636	7716.705	7788.283	7712.772
Potencia Cauchy	<b>7420.38</b>	<b>7428.443</b>	<b>7507.974</b>	<b>7420.479</b>
Reversa Potencia Cauchy	7469.457	7475.673	7555.205	7469.661

El modelo con función de enlace Potencia Cauchy presenta los menores valores en todos los criterios de comparación y en segundo lugar el modelo con función de enlace Reversa potencia Cauchy, lo que indica un mejor balance entre el ajuste a los datos y la penalización por complejidad. Estos resultados destacan la capacidad de los modelos Potencia Cauchy y Reversa potencia Cauchy para capturar patrones en los datos desbalanceados con mayor precisión que los otros modelos.

### Presentación del Mejor Modelo

A continuación, se presentan los coeficientes estimados para el mejor modelo ajustado, junto con sus desviaciones estándar e intervalos creíbles del 95 %:

Tabla 5.4: Parámetros estimados para el modelo Potencia Cauchy.

Parámetro	Estimado	Desviación Estándar	2.5 %	97.5 %
$\beta_1$	-0.0083	0.0017	-0.0116	-0.0050
$\beta_2$	-0.0357	0.0055	-0.0468	-0.0247
$\beta_3$	-0.0931	0.0040	-0.1012	-0.0855
$\beta_4$	-0.0120	0.0060	-0.0244	-0.0008
$\beta_5$	3.3335	0.1275	3.0867	3.5990
$\beta_6$	-0.6250	0.2282	-1.0876	-0.1689
$\beta_7$	0.1347	0.0521	0.0302	0.2333
$\beta_8$	-0.4329	0.0238	-0.4807	-0.3872
$\beta_9$	0.0048	0.0016	0.0014	0.0076
$\log(\lambda)$	0.4328	0.0264	0.3802	0.4862

Para garantizar la calidad del ajuste del modelo, se evaluaron los valores de *Rhat* para los parámetros estimados. Un valor de *Rhat* cercano a 1 indica una buena convergencia:

Tabla 5.5: Valores de *Rhat* para los parámetros clave del modelo Potencia Cauchy.

Parámetro	Rhat
$\beta_1$	1.0029
$\beta_2$	1.0026
$\beta_3$	1.0003
$\beta_4$	0.9991
$\beta_5$	1.0004
$\beta_6$	1.0012
$\beta_7$	0.9997
$\beta_8$	1.0019
$\beta_9$	0.9998
$\log(\lambda)$	1.0002

Todos los valores de *Rhat* están dentro del rango aceptable, indicando una buena convergencia del modelo.

#### Diagnóstico del mejor Modelo

El diagnóstico del modelo se realiza mediante el análisis de residuos, evaluando si cumplen con las distribuciones esperadas. A continuación, se presentan los gráficos de residuos cuantílicos y el QQ-Plot:

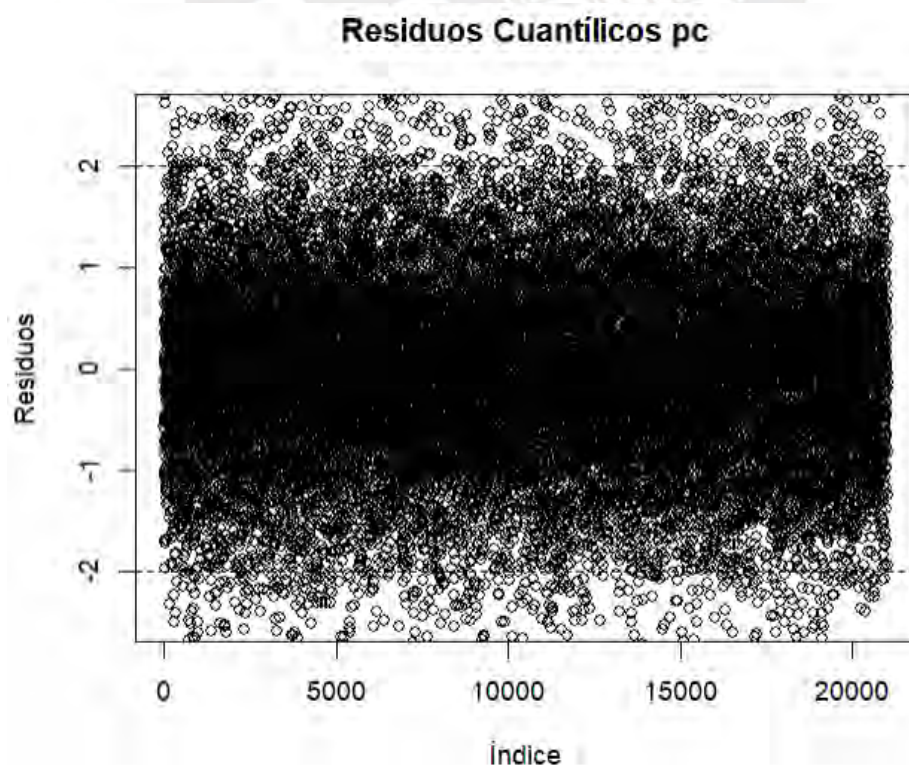


Figura 5.4: Gráfico de residuos cuantílicos para el modelo Potencia Cauchy.

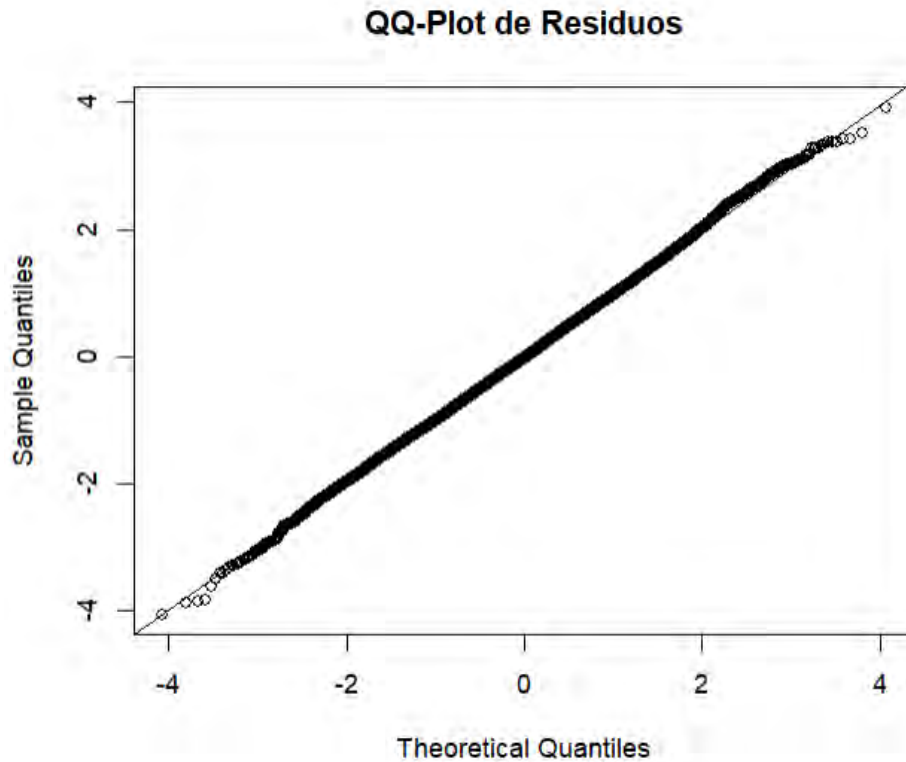


Figura 5.5: QQ-Plot de residuos para el modelo Potencia Cauchy.

El gráfico de residuos cuantílicos (ver la figura 5.4) muestra que los residuos se encuentran mayoritariamente dentro del rango esperado de  $[-2, 2]$ , lo que indica que no hay evidencia significativa de falta de ajuste en el modelo. Además, la distribución uniforme de los residuos alrededor de cero sugiere que el modelo captura adecuadamente la relación entre las covariables y la respuesta, sin patrones aparentes que indiquen problemas de especificación.

El QQ-Plot (ver la figura 5.5) evalúa si los residuos siguen una distribución normal estándar, como es esperado en este tipo de análisis. Los puntos se alinean de manera adecuada con la línea de referencia, particularmente en los cuartiles centrales, indicando que los residuos son aproximadamente normales. Algunas desviaciones leves se observan en los extremos, pero no son lo suficientemente significativas como para cuestionar el ajuste general del modelo.

**A modo de comparación.** Luego de analizar los residuos para el modelo Potencia Cauchy, se realizó un análisis similar para el modelo Logístico, con la intención de comparar el ajuste de ambos modelos y evaluar cuál describe de manera más adecuada la relación entre las covariables y la variable de interés.

A continuación, se presentan los gráficos de residuos cuantílicos y QQ-Plot correspondientes al modelo Logístico:

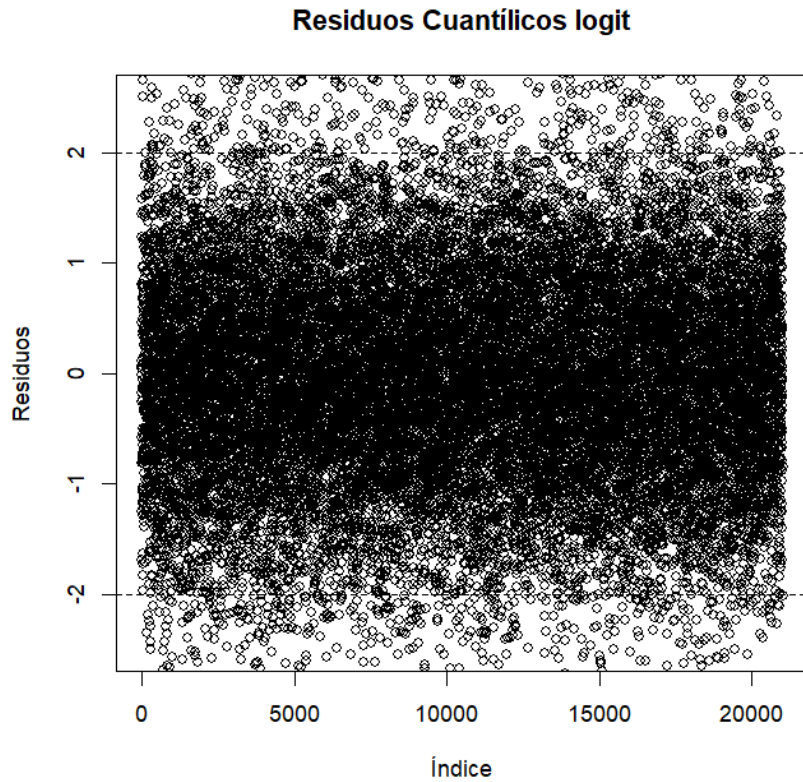


Figura 5.6: Gráfico de residuos cuantílicos para el modelo Logístico.

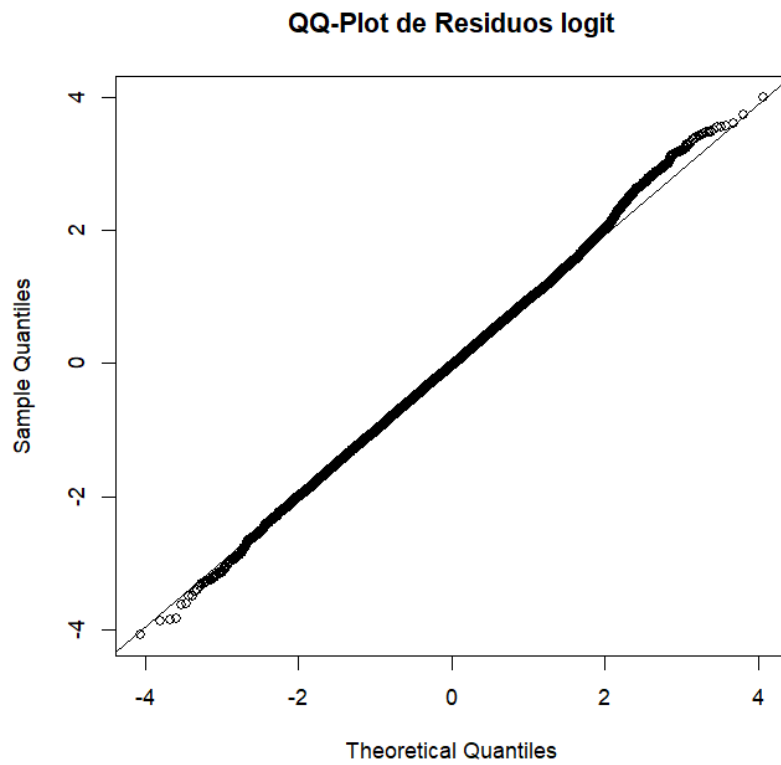


Figura 5.7: QQ-Plot de residuos para el modelo Logístico.

Este análisis de comparación es relevante porque un modelo adecuado debe presentar residuos

---

bien distribuidos alrededor de cero (en el gráfico cuantílico) y alinearse con la normal teórica (en el QQ-Plot).

Los gráficos del modelo Logístico (ver la figura 5.6 y la figura 5.7) permiten identificar posibles patrones o desviaciones que no se observan en el modelo Potencia Cauchy (ver la figura 5.4 y la figura 5.5), justificando así la elección final del modelo más robusto para datos desbalanceados.

### **Comparación de los Gráficos de Residuos**

En los gráficos de residuos cuantílicos, el modelo Potencia Cauchy (ver la figura 5.4) muestra residuos más concentrados y distribuidos uniformemente dentro del rango esperado de  $[-2, 2]$ , mientras que en el modelo Logístico (ver 5.6) se observa una mayor dispersión, particularmente en las regiones extremas.

Por otro lado, en los QQ-Plots, el modelo Potencia Cauchy (ver la figura 5.5) presenta una mejor alineación con la línea de referencia, especialmente en los cuartiles centrales, lo que indica una distribución más cercana a la normal teórica. En el modelo Logístico (ver la figura 5.7), aunque los residuos también siguen un patrón aproximadamente normal, se observan desviaciones más pronunciadas en los extremos.

Según lo presentado tenemos que, bajo el enfoque bayesiano, el modelo con función de enlace Potencia Cauchy fue seleccionado como el mejor ajustado para los datos analizados, al presentar los valores más bajos en los criterios de información EAIC, EBIC y WAIC. Este modelo destaca por su capacidad para manejar datos desbalanceados, introduciendo asimetría en la distribución de probabilidades a través del parámetro  $\lambda$ , estimado en  $\exp(0,4328) \approx 1,541$ .

Los coeficientes estimados indican que ciertas covariables tienen un impacto significativo en la probabilidad de incumplimiento crediticio. En particular, los coeficientes  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  y  $\beta_4$  presentan valores negativos, lo que sugiere que un incremento en las variables asociadas reduce la probabilidad de incumplimiento. Por otro lado, el coeficiente  $\beta_5$  muestra el mayor impacto positivo, indicando una relación directa y significativa con la probabilidad de incumplimiento.

El análisis de residuos refuerza la superioridad del modelo Potencia Cauchy. Los gráficos de residuos cuantílicos muestran una distribución uniforme alrededor de cero, sin patrones evidentes que indiquen falta de ajuste, mientras que el QQ-Plot confirma que los residuos se alinean adecuadamente con la normal teórica, salvo ligeras desviaciones en los extremos.

En contraste, los residuos del modelo Logístico presentan una mayor dispersión en los gráficos cuantílicos, especialmente en las regiones extremas, y el QQ-Plot exhibe desviaciones más notorias respecto a la línea de referencia. Estos resultados sugieren que el modelo Logístico no captura con la misma precisión las características asimétricas y desbalanceadas de los datos.

En conclusión, el modelo Potencia Cauchy ofrece un ajuste más robusto y preciso, destacándose por su capacidad para modelar datos desbalanceados y mejorar la predicción del incumplimiento crediticio.

En la siguiente sección, se presenta un análisis complementario donde los mismos modelos son evaluados bajo el enfoque de máxima verosimilitud. Esto permitirá comparar los resultados obtenidos

por ambos métodos de inferencia, destacando la robustez y consistencia de las estimaciones, así como la sensibilidad frente a los métodos utilizados.

### 5.3. Análisis de modelos mediante inferencia por máxima verosimilitud

En esta sección se evaluaron nuevamente los cinco modelos de regresión binaria con funciones de enlace logit, probit, Cauchy, Potencia Cauchy y Reversa Potencia Cauchy, pero esta vez bajo el enfoque de inferencia frecuentista, utilizando el método de máxima verosimilitud (MLE). El objetivo fue contrastar los resultados obtenidos con el enfoque bayesiano, y verificar la robustez de las estimaciones obtenidas para predecir el incumplimiento crediticio.

Para cada modelo, se definió y maximizó la función de log-verosimilitud correspondiente mediante el algoritmo BFGS. La comparación entre modelos se realizó utilizando los criterios de información clásicos: AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) y HQIC (Hannan–Quinn Information Criterion), descritos previamente en el capítulo metodológico.

A continuación, se presentan los resultados de los criterios de información para cada uno de los modelos:

Tabla 5.6: Criterios de Información (MLE) para Comparación de Modelos

Modelo	AIC	BIC	HQIC
Logístico	7907.636	7979.214	7930.997
Probit	8203.306	8274.884	8226.668
Cauchy	7707.781	7779.359	7731.143
Potencia Cauchy	<b>7418.420</b>	<b>7497.951</b>	<b>7444.378</b>
Reversa Potencia Cauchy	7465.530	7545.061	7491.487

El modelo con función de enlace Potencia Cauchy presenta los valores más bajos en los tres criterios de información (AIC, BIC y HQIC), lo que lo posiciona como el mejor modelo bajo el enfoque de máxima verosimilitud. Le sigue el modelo Reversa Potencia Cauchy. Esto sugiere que los modelos con funciones de enlace flexibles permiten un mejor ajuste a los datos desbalanceados, manteniendo un equilibrio adecuado entre calidad de ajuste y penalización por complejidad.

#### Presentación del Mejor Modelo (MLE)

A continuación, se presentan los coeficientes estimados para el modelo Potencia Cauchy ajustado por máxima verosimilitud, así como sus errores estándar e intervalos de confianza del 95 %:

Tabla 5.7: Parámetros estimados (MLE) para el modelo Potencia Cauchy.

Parámetro	Estimado	Error Estándar	IC 95 %
$\beta_1$	-0.0080	0.0013	[-0.0113, -0.0048]
$\beta_2$	-0.0352	0.0054	[-0.0459, -0.0246]
$\beta_3$	-0.0918	0.0040	[-0.0993, -0.0842]
$\beta_4$	-0.0112	0.0059	[-0.0227, 0.0004]
$\beta_5$	3.2826	0.1215	[3.0397, 3.5255]
$\beta_6$	-0.6121	0.2269	[-1.0595, -0.1648]
$\beta_7$	0.1372	0.0504	[0.0384, 0.2361]
$\beta_8$	-0.4259	0.0239	[-0.4722, -0.3796]
$\beta_9$	0.0049	0.0015	[0.0020, 0.0077]
$\lambda$	1.5546	0.0414	[1.4734, 1.6359]

### Diagnóstico del Modelo

Al igual que en la inferencia bayesiana, se evaluó la calidad del ajuste del modelo mediante análisis de residuos. Se calcularon residuos cuantílicos y gráficos QQ para examinar la normalidad de los residuos estandarizados. Los resultados obtenidos son coherentes con los del enfoque bayesiano, reforzando la calidad del modelo Potencia Cauchy para describir patrones en los datos.

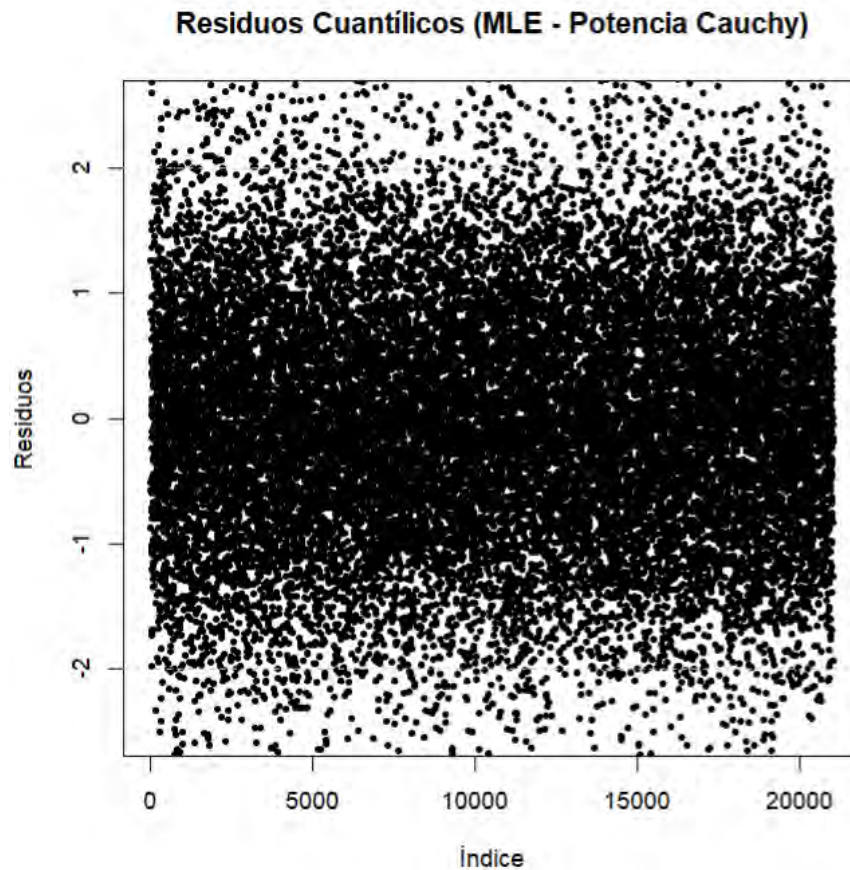


Figura 5.8: Gráfico de residuos cuantílicos para el modelo Potencia Cauchy (MLE).

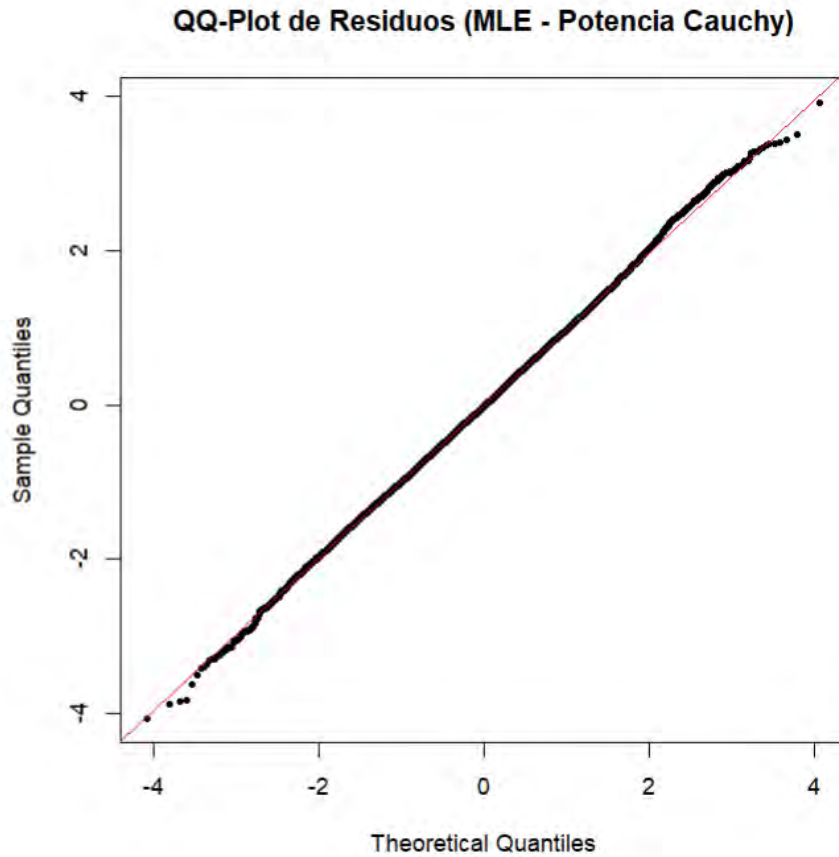


Figura 5.9: QQ-Plot de residuos para el modelo Potencia Cauchy (MLE).

En conclusión, los resultados obtenidos por máxima verosimilitud coinciden con los obtenidos bajo el enfoque bayesiano: el modelo Potencia Cauchy es consistentemente el que mejor explica la variable de respuesta, seguido por el modelo Reversa Potencia Cauchy. Esto refuerza la robustez de las funciones de enlace basadas en transformaciones de la distribución Cauchy, especialmente en contextos con datos desbalanceados y alta asimetría.

Finalmente, se destaca que los coeficientes estimados y los diagnósticos de ajuste obtenidos por ambos métodos son coherentes entre sí, lo cual fortalece la validez de las conclusiones y proporciona evidencia robusta para la elección del modelo Potencia Cauchy en la predicción del incumplimiento crediticio.

## Capítulo 6

# Conclusiones y Consideraciones

## Finales

### 6.1. Conclusiones

Esta tesis ha propuesto y evaluado el uso de funciones de enlace asimétricas basadas en las distribuciones potencia Cauchy y reversa potencia Cauchy en el contexto de los modelos de regresión binaria. A través de una estrategia metodológica que combinó enfoques bayesianos y frecuentistas - específicamente la inferencia bayesiana mediante MCMC y la estimación por máxima verosimilitud (MLE)- se exploró su desempeño en escenarios con fuerte desbalance de clases.

Los hallazgos derivados de los estudios de simulación y de la aplicación empírica permiten concluir que estas funciones de enlace asimétricas ofrecen mejoras sustanciales en comparación con los enlaces simétricos tradicionales como el logit, el probit y el Cauchy. En los estudios de simulación, los modelos con funciones de enlace potencia Cauchy y reversa potencia Cauchy presentaron menores niveles de sesgo y error cuadrático medio (RECM) en la estimación de los parámetros  $\beta_0$ ,  $\beta_1$  y  $\lambda$ , especialmente bajo condiciones de desbalance extremo. Asimismo, el análisis de sensibilidad respecto a diferentes distribuciones a priori para el parámetro de asimetría  $\lambda$  reveló que la distribución uniforme produjo estimaciones más estables y menos sesgadas. En la aplicación al problema de predicción de incumplimiento crediticio, el modelo con enlace potencia Cauchy obtuvo el mejor ajuste según múltiples criterios de información, tanto bayesianos (EAIC, EBIC, WAIC) como frecuentistas (AIC, BIC, HQIC), superando a los modelos con funciones simétricas. Finalmente, la incorporación del enfoque por máxima verosimilitud permitió validar los resultados obtenidos bajo el enfoque bayesiano, demostrando la robustez y consistencia de las estimaciones, así como la superioridad del modelo con enlace potencia Cauchy en ambos marcos.

Estos resultados evidencian el potencial de las funciones de enlace asimétricas como herramientas eficaces para modelar relaciones no lineales y desequilibradas entre variables explicativas y una respuesta binaria.

---

## 6.2. Consideraciones Finales

Los resultados de esta investigación destacan tanto el aporte práctico como teórico de las funciones de enlace potencia Cauchy y reversa potencia Cauchy. Su capacidad para capturar asimetría mediante un parámetro ajustable  $\lambda$  representa una ventaja significativa en el modelamiento de datos reales, particularmente en dominios donde la distribución de la respuesta no es equilibrada, como en aplicaciones financieras.

No obstante, también se reconocen ciertos desafíos y limitaciones. La interpretación del parámetro  $\lambda$  y su impacto sobre la curvatura de la función de enlace aún requiere una formalización teórica más accesible para facilitar su uso generalizado. Además, la implementación computacional, especialmente en entornos bayesianos con MCMC, implica mayores tiempos de procesamiento en comparación con modelos clásicos; sin embargo, este aspecto puede optimizarse mediante algoritmos eficientes como NUTS.

La inclusión del enfoque frecuentista mediante MLE amplía la aplicabilidad de estas funciones de enlace a contextos donde el enfoque bayesiano no sea viable o preferido. La consistencia de los resultados obtenidos desde ambos paradigmas refuerza la confiabilidad de las conclusiones.

En cuanto a las líneas futuras de investigación, se sugiere analizar con mayor profundidad la interpretación geométrica y estadística del parámetro  $\lambda$  en aplicaciones concretas, así como extender el uso de estas funciones a modelos más complejos, incluyendo regresión multinomial, modelos jerárquicos o series de tiempo. También resulta de interés investigar nuevas funciones de enlace derivadas de otras distribuciones asimétricas, ampliando el repertorio de herramientas disponibles, y evaluar su comportamiento en presencia de colinealidad alta entre covariables o en escenarios de alta dimensionalidad.

En suma, esta tesis establece una base sólida desde los planos teórico, metodológico y aplicado para el uso de funciones de enlace asimétricas en regresión binaria, y abre nuevas perspectivas para investigaciones futuras en estadística aplicada y modelamiento probabilístico.

# Anexo

Los métodos MCMC son un conjunto de técnicas de simulación estocástica para obtener muestras de una distribución comúnmente desconocida. Estos métodos se basan en la construcción de cadenas de Markov cuya distribución de interés es la distribución invariante (estacionaria). En el contexto de inferencia Bayesiana, la distribución estacionaria es la distribución a posteriori.

Entre los algoritmos MCMC más conocidos están el algoritmo de Metropolis-Hastings, el algoritmo de rechazo adaptativo y el algoritmo mostrador de Gibbs. Más detalles sobre estos algoritmos pueden ser vistos en Gamerman y Lopes (2006) y Roberts y Smith (1994).

En este trabajo, consideramos el algoritmo de No-U-Turn Sampler desarrollado por Hoffman y Gelman (2014), una extensión para el eficiente algoritmo MCMC denominado Monte Carlo Hamiltoniano (HMC). Este algoritmo permite que la cadena de Markov explore la distribución objetivo con mucho más eficiencia que otros métodos MCMC ampliamente conocidos como el muestreo de Metropolis-Hastings y Gibbs (implementado por ejemplo en JAGS), usando dinámica hamiltoniana en vez de una distribución de probabilidad y haciendo que la simulación converja más rápidamente para la distribución objetivo con baja autocorrelación (Neal, 2012). Este algoritmo está implementado en el software Stan.

El algoritmo HMC fue propuesto por Duane et al. (1987) en la simulación dinámica de sistemas moleculares, denominado Monte Carlo Híbrido. Posteriormente, Neal (1996) introdujo el método en redes neuronales Bayesianas. Neal (2012) presenta propiedades importantes de este método y demuestra su convergencia para alguna distribución de interés, por ejemplo, la distribución a posteriori de los parámetros.

El Monte Carlo Hamiltoniano está basado en el concepto de dinámica hamiltoniana, la cual está definida en términos de la localización del objeto  $q$  asociada a una energía potencial  $U(q)$  y su momento  $k$  asociado a una energía cinética  $K(k)$  en algún tiempo  $t$ . La energía total del sistema, llamada Hamiltoniana  $H(q; k)$ , se define como:

$$H(q; k) = U(q) + K(k);$$

donde  $q$  y  $k$  son vectores de dimensión  $d$ . Así, la dinámica hamiltoniana está definida por el sistema de ecuaciones diferenciales:

$$\frac{dq}{dt} = \frac{\partial H}{\partial k} = \nabla_k K(k) \quad \text{y} \quad \frac{dk}{dt} = -\frac{\partial H}{\partial q} = -\nabla_q U(q).$$

Si tenemos una posición inicial y un momento inicial en el tiempo  $t_0$ , es decir,  $(q(t_0); k(t_0))$ , entonces podemos prever la localización y el momento del objeto en cualquier momento futuro  $t = t_0 + T$  simulando dinámicas por una duración de tiempo  $T$ .

Para relacionar  $H(q; k)$  con la distribución objetivo  $p(q | y; X)$ , usamos un concepto de la mecánica estadística conocido como distribución canónica. En este caso, la distribución a posteriori de los parámetros del modelo es el foco de interés y, por lo tanto, estos parámetros asumirán el papel de la posición  $q$ . La distribución a posteriori se puede ver como una distribución canónica (con  $T = 1$ ) usando una función de energía potencial definida de la siguiente manera:

$$U(q) = -\log(L(q | y; X) \cdot p(q));$$

donde  $p(q)$  es la densidad a priori y  $L(q | y; X)$  es la función de verosimilitud. Note que la función de densidad conjunta de  $(q; k)$  está definida por la exponencial negativa del Hamiltoniano, ya que la función  $K(k)$  tiene la forma del núcleo de una distribución Normal multivariada  $N_d(0; M)$ . Así, tenemos que:

$$h(q; k) \propto \exp\{-H(q; k)\} \propto L(q | y; X) \cdot p(q) \cdot \exp(-k^\top M^{-1}k).$$

Para los modelos de potencia y reversa de potencia, los parámetros a posteriori se definen como  $q = (b; d)^\top$ , un vector de dimensión  $d = p + 1$ . Por lo tanto,  $U(q)$  tiene la siguiente forma:

$$U(b; d) = -\sum_{i=1}^n y_i \log(F_d(x_i^\top b)) - \sum_{i=1}^n (1 - y_i) \log(1 - F_d(x_i^\top b)) - \sum_{j=1}^p \frac{b_j^2}{2(10^2)};$$

Donde  $F_d(\cdot)$  es la fda de la distribución potencia o reversa de potencia Cauchy.

Las derivadas parciales  $\nabla_q U(q)$ , conocidas como vector de puntuación, se definen como:

$$\begin{pmatrix} \frac{\partial U(b; d)}{\partial b_1} \\ \vdots \\ \frac{\partial U(b; d)}{\partial b_p} \\ \frac{\partial U(b; d)}{\partial d} \end{pmatrix}^\top.$$

Por otro lado, el vector de momento  $k$  se puede obtener generando valores de una distribución Normal multivariada con media 0 y matriz de varianzas y covarianzas  $M = \text{diag}\{m_1, \dots, m_d\}$  de modo que:

$$K(k) = \frac{1}{2} k^\top M^{-1} k = \sum_{i=1}^d \frac{k_i^2}{2m_i}.$$

Con las especificaciones anteriores, el algoritmo HMC en su forma simple, tomando  $M = I$  como una matriz identidad  $d \times d$ , se define por:

- Proporcione una posición inicial  $q(0)$  y los valores de  $e$ ,  $N$  y  $M$ .
- Inicie un contador  $i = 1; \dots; N$  (tamaño de la cadena)

- 
- Genere  $k^* \sim \text{Nd}(0; M)$  y  $u \sim U(0; 1)$
  - Haga  $\{q^*, k^*\} = \{q(i-1), k^*\}$  y  $H_0 = H\{q^*, k^*\}$
  - Para  $j = 1$  hasta  $M$ , aplique el método Leapfrog:
    - $k^* = k^* + (e/2)\nabla_q U(q(i-1))$
    - $q(i-1) = q(i-1) + ek^*$
    - $k^* = k^* + (e/2)\nabla_q U(q(i-1))$
  - Cuando  $j = M$ , haga  $\{q^L, k^L\} = \{q(i-1), k^*\}$  y  $H_1 = H\{q^L, k^L\}$
  - Calcule la aceptación:

$$h\{q^L, k^L\}; \{q^*, k^*\} = \min\{\exp(H_0 - H_1), 1\} = a$$

- Haga  $q_i = \begin{cases} q^L, & \text{con probabilidad } a \\ q^*, & \text{en caso contrario} \end{cases}$

El rendimiento del HMC depende críticamente de la elección de valores adecuados para  $e$  y  $M$ . Una mala elección de estos puede llevar a una alta tasa de rechazo o un tiempo de procesamiento computacional muy alto. Ajustar estos parámetros para cualquier problema específico requiere cierto conocimiento y, generalmente, una o más ejecuciones preliminares.

El No-U-Turn Sampler (NUTS) es una extensión del HMC que elimina la necesidad de especificar un valor fijo de  $M$  y define  $e$  en función del algoritmo de media dual de Nesterov (2009). Así, el NUTS puede ejecutarse sin ningún ajuste manual, y las muestras generadas son al menos tan buenas como las del HMC (Neal, 2012).

La revisión del método MCH y el rendimiento del algoritmo NUTS se pueden ver en el trabajo de Anyosa (2017).

# Bibliografía

- Anyosa, S. A. C. (2017). *Regressão binária usando ligações potência e reversa de potência* [Tesis de maestría, Universidade Federal de São Carlos]. Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP.
- Bazán, J. L., Romeo, J. S., y Rodrigues, J. (2014). A Bayesian approach for the flexible skew-normal class of distributions. *Statistical Papers*, 55(1), 45–65.
- Bazán, J. L., Lemes, R., y Bonat, W. H. (2017). Skewed link functions in binary regression models. *Communications in Statistics - Theory and Methods*, 46(11), 5537–5552.
- Bazán, J., Torres-Avilés, F., Suzuki, A., y Louzada, F. (2017). Power and reversal power links for binary regressions: An application for motor insurance policyholders. *Applied Stochastic Models in Business and Industry*, 33(1), 22–34.
- Brys, G., Hubert, M., & Struyf, A. (2004). A robust measure of skewness. *Journal of Computational and Graphical Statistics*, 13(4), 996–1017.
- Chen, M.-H., Dey, D. K., y Shao, Q.-M. (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, 94(448), 1172–1186.
- Chumbimune, W. (2017). *Distribuciones reversa de potencia: propiedades y aplicaciones*. Tesis de licenciatura, Pontificia Universidad Católica del Perú.
- Collett, D. (2002). *Modelling Binary Data* (2nd ed.). Chapman & Hall/CRC.
- Cox, D. R., y Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 49(1), 1–39.
- Cox, D. R., y Snell, E. J. (1989). *Analysis of Binary Data* (2nd ed.). Chapman & Hall.
- de la Cruz, A., Bazán, J. L., Cancho, V. G., y Dey, D. K. (2019). Performance of asymmetric links and correction methods for imbalanced data in binary regression. *Journal of Statistical Computation and Simulation*, 89(9), 1694–1714.
- de la Cruz, A., Bazán, J. L., y Diniz, C. A. R. (2023). Longitudinal binary response models using alternative links for medical data. *Brazilian Journal of Probability and Statistics*, 37(2), 365–392.

- 
- Dunn, P. K., y Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3), 236–244.
- Gupta, R. D., y Gupta, P. L. (2008). A generalized exponential distribution: Estimation and applications. *Journal of Statistical Computation and Simulation*, 78(11), 1021–1031.
- Huayanay, R., Bazán, J. L., y De la Cruz, R. (2024). Performance of evaluation metrics for classification in imbalanced data. *Revista Colombiana de Estadística*, 47(1), 89–108. <https://doi.org/10.15446/rce.v47n1.102423>
- Lemonte, A. J., y Bazán, J. L. (2018). New links for binary regression: an application to coca cultivation in Peru. *TEST*, 27(3), 597–617.
- Moors, J. J. A. (1988). A quantile alternative for kurtosis. *The Statistician*, 37(1), 25–32.
- Paal, B. Van der. (2014). A comparison of different methods for modelling rare events data. Tesis de maestría, Ghent University.
- Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). Wiley.
- Stan Development Team. (2016). *Stan modeling language user's guide and reference manual*. Technical Report.

