Modelo estadístico Bayesiano para datos binarios desbalanceados

Alex de la Cruz[1], Jorge Bazán[2]

[1] aldehu@usp.br

[1,2]Universidad de São Paulo, Brasil.

Lima, 05 de octubre

Contenido

1. Introducción

El modelo matemático que intenta explicar una variable a partir de una función de otras variables es llamado modelos de regresión.

1. Introducción

El modelo matemático que intenta explicar una variable a partir de una función de otras variables es llamado modelos de regresión.

Modelos Lineales Generalizados (MCCULLAGH; NELDER, 1989). Considere una variable respuesta, un conjunto de covariables y coeficientes de regresión, respectivamente dadas por:

1. Introducción

El modelo matemático que intenta explicar una variable a partir de una función de otras variables es llamado modelos de regresión.

Modelos Lineales Generalizados (MCCULLAGH; NELDER, 1989). Considere una variable respuesta, un conjunto de covariables y coeficientes de regresión, respectivamente dadas por:

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad \text{y} \quad \boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$$

. Con esto

$$\underset{\mu}{\boldsymbol{Y}} \xleftarrow{\text{explicado por}} \underset{f(\cdot)}{\boldsymbol{X}}$$

2. Modelos de regresión Binaria

Nos interesa estudiar la relación de un conjunto de covariables y una variable respuesta dicotómica.

2. Modelos de regresión Binaria

Nos interesa estudiar la relación de un conjunto de covariables y una variable respuesta dicotómica.
Sea:

- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ vector de respuestas con valores 1 o 0.

2. Modelos de regresión Binaria

Nos interesa estudiar la relación de un conjunto de covariables y una variable respuesta dicotómica.

Sea:

- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ vector de respuestas con valores 1 o 0.
- $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})^\top$ vector de p covariables.

2. Modelos de regresión Binaria

Nos interesa estudiar la relación de un conjunto de covariables y una variable respuesta dicotómica.

Sea:

- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ vector de respuestas con valores 1 o 0.
- $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})^\top$ vector de p covariables.

2. Modelos de regresión Binaria

Nos interesa estudiar la relación de un conjunto de covariables y una variable respuesta dicotómica.
Sea:

- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ vector de respuestas con valores 1 o 0.
- $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})^\top$ vector de p covariables.
- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ coeficientes de regresión.

Modelos de regresión Binaria

2. Modelos de regresión Binaria

Nos interesa estudiar la relación de un conjunto de covariables y una variable respuesta dicotómica.
Sea:

- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ vector de respuestas con valores 1 o 0.
- $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})^\top$ vector de p covariables.
- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ coeficientes de regresión.

Modelos de regresión Binaria

$$\begin{aligned} Y_i &\sim \text{Bernoullí}\,(u_i) \ \text{ independientes } \quad \forall\ i = 1, \ldots, n \\ u_i &= P(Y_i = 1) = F\,(\eta_i) \\ \eta_i &= \boldsymbol{x_i}^T \boldsymbol{\beta} = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} \end{aligned} \tag{1}$$

2. Modelos de regresión Binaria

Función de enlace

2. Modelos de regresión Binaria

Función de enlace

Relaciona η_i y μ_i

$$\begin{gathered}\eta_i = g\left(\mu_i\right) \\ u_i = P(Y_i = 1) \\ \eta_i = \beta_1 x_{i1} + \ldots + \beta_p x_{ip} \\ i = 1, \ldots, n\end{gathered} \tag{2}$$

2. Modelos de regresión Binaria

Función de enlace

Relaciona η_i y μ_i

$$\begin{gathered} \eta_i = g\left(\mu_i\right) \\ u_i = P(Y_i = 1) \\ \eta_i = \beta_1 x_{i1} + \ldots + \beta_p x_{ip} \\ i = 1, \ldots, n \end{gathered} \tag{2}$$

Ejemplo: modelo de regresión logística

$$\eta = \log\left(\frac{\mu}{1-\mu}\right)$$

$$\mu = \frac{\exp\left(\beta_1 x_1 + \ldots + \beta_p x_p\right)}{1 + \exp\left(\beta_1 x_1 + \ldots + \beta_p x_p\right)}$$

2. Modelos de regresión Binaria

Curvas de probabilidad acumulada.

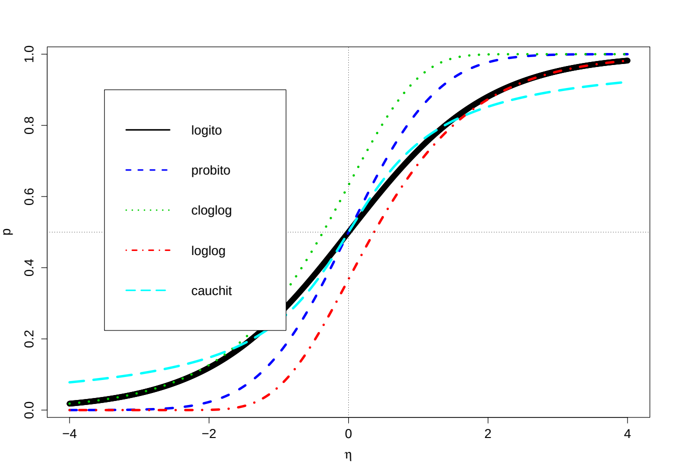

Figura 1: Enlaces comunes

3. Regresión binaria potencia e reversa de potencia

3. Regresión binaria potencia e reversa de potencia

Está basado en las distribuciones potencia y reversa de potencia de Lemonte e Bazán (2017).

3. Regresión binaria potencia e reversa de potencia

Está basado en las distribuciones potencia y reversa de potencia de Lemonte e Bazán (2017).

Distribución potencia y reversa de potencia

Una variable aleatoria univariada T tiene distribución potencia si su fda es dada por:

$$F_p(t \mid \mu, \sigma) = G\left(\frac{t-\mu}{\sigma}\right)^{\lambda}, \quad t \in \mathbb{R}, \tag{3}$$

3. Regresión binaria potencia e reversa de potencia

Está basado en las distribuciones potencia y reversa de potencia de Lemonte e Bazán (2017).

Distribución potencia y reversa de potencia

Una variable aleatoria univariada T tiene distribución potencia si su fda es dada por:

$$F_p(t \mid \mu, \sigma) = G\left(\frac{t-\mu}{\sigma}\right)^{\lambda}, \quad t \in \mathbb{R}, \tag{3}$$

y su correspondiente función reversa de potencia

$$F_{rp}(t \mid \mu, \sigma) = 1 - G\left(-\left(\frac{t-\mu}{\sigma}\right)\right)^{\lambda}, \quad t \in \mathbb{R} \tag{4}$$

3. Regresión binaria potencia e reversa de potencia

Así tenemos dos tipos de función de enlace

3. Regresión binaria potencia e reversa de potencia

Así tenemos dos tipos de función de enlace

Potencia

$$\mu_i = F_p(\eta_i) = G(\eta_i)^{\lambda}, \quad \eta_i = F_p^{-1}(\mu_i) \tag{5}$$

3. Regresión binaria potencia e reversa de potencia

Así tenemos dos tipos de función de enlace

Potencia

$$\mu_i = F_p(\eta_i) = G(\eta_i)^\lambda, \quad \eta_i = F_p^{-1}(\mu_i) \tag{5}$$

Reversa de potencia

$$\mu_i = F_{rp}(\eta_i) = 1 - G(-\eta_i)^\lambda, \quad \eta_i = F_{rp}^{-1}(\mu_i) \tag{6}$$

en que $\eta_i = \mathbf{x_i}^\top \boldsymbol{\beta} = \beta_1 x_{i1} + \ldots + \beta_k x_{ip}$ es el i-ésimo predictor lineal, $i = 1, \ldots, n$

3. Regresión binaria potencia e reversa de potencia

Tabela 1: Modelos para regresión binaria usando enlace de potencia e reversa de potencia

Modelo	Notação	μ	η
Potência normal	**PN**	$[\Phi(\eta)]^{\lambda}$	$\eta = \Phi^{-1}\left(\mu^{1/\lambda}\right)$
Reversa de potência normal	RPN	$1 - [\Phi(-\eta)]^{\lambda}$	$\eta = -\Phi^{-1}\left((1-\mu)^{1/\lambda}\right)$
Potência logística	**PL**	$\left[\frac{1}{1+\exp(-\eta)}\right]^{\lambda}$	$\eta = -\log\left(\mu^{-1/\lambda} - 1\right)$
Reversa de potência logística	RPL	$1 - \left[\frac{1}{1+\exp(\eta)}\right]^{\lambda}$	$\eta = \log\left((1-\mu)^{-1/\lambda} - 1\right)$
Potência Laplace	**PLAPLACE**	$\left[\frac{1}{2} + \frac{1}{2}\text{sign}(\eta)\left\{1 - e^{-\lvert\eta\rvert}\right\}\right]^{\lambda}$	$\eta = {G_1}^{-1}\left(\mu^{1/\lambda}\right)$
Reversa de potência Laplace	RPLAPLACE	$1 - \left[\frac{1}{2} - \frac{1}{2}\text{sign}(\eta)\left\{1 - e^{-\lvert\eta\rvert}\right\}\right]^{\lambda}$	$\eta = -{G_1}^{-1}\left((1-\mu)^{1/\lambda}\right)$
Potência Cauchy	**PC**	$\left[\frac{1}{\pi}\arctan(\eta) + \frac{1}{2}\right]^{\lambda}$	$\eta = \tan\left[\left(\mu^{1/\lambda} - \frac{1}{2}\right)\pi\right]$
Reversa de potência Cauchy	RPC	$1 - \left[\frac{1}{\pi}\arctan(-\eta) + \frac{1}{2}\right]^{\lambda}$	$\eta = -\tan\left[\left((1-\mu)^{1/\lambda} - \frac{1}{2}\right)\pi\right]$

${G_1}^{-1}(\cdot)$ es la inversa de la FDA de una distribución Laplace

3. Regresión binaria potencia e reversa de potencia

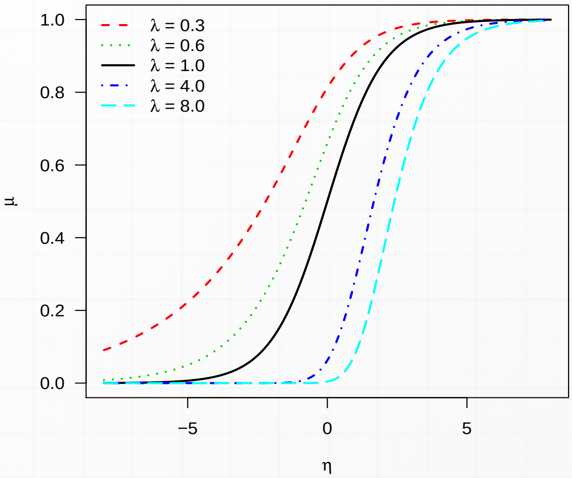

Figura 2: Enlace potencia

4. Inferencia Bayesiana de los parámetros

El modelo Bayesiano para regresión binaria con enlaces potencia y reversa de potencia es dado por:

4. Inferencia Bayesiana de los parámetros

El modelo Bayesiano para regresión binaria con enlaces potencia y reversa de potencia es dado por:

$$\begin{aligned} Y_i \mid \beta, \lambda &\sim \text{Bernoulli}\left(u_i\right) \quad \forall\ i = 1, ..., n \\ u_i &= F_\lambda\left(\eta_i\right) \\ \eta_i &= \boldsymbol{x_i}^T \boldsymbol{\beta} \\ (\boldsymbol{\beta}, \lambda) &\sim \pi\left(\boldsymbol{\beta}, \lambda\right) \end{aligned} \tag{7}$$

4. Inferencia Bayesiana de los parámetros

El modelo Bayesiano para regresión binaria con enlaces potencia y reversa de potencia es dado por:

$$\begin{aligned} Y_i \mid \beta, \lambda &\sim \text{Bernoulli}\left(u_i\right) \quad \forall\ i = 1, ..., n \\ u_i &= F_\lambda\left(\eta_i\right) \\ \eta_i &= \boldsymbol{x_i}^T \boldsymbol{\beta} \\ (\boldsymbol{\beta}, \lambda) &\sim \pi\left(\boldsymbol{\beta}, \lambda\right) \end{aligned} \tag{7}$$

donde $F_\lambda(\cdot)$ es una FDA de una distribución potencia o reversa de potencia y $\pi\left(\boldsymbol{\beta}, \lambda\right)$ es la distribución conjunta *a priori* de los parámetros.

4. Inferencia Bayesiana de los parámetros

El modelo Bayesiano para regresión binaria con enlaces potencia y reversa de potencia es dado por:

$$
\begin{aligned}
Y_i \mid \beta, \lambda &\sim \text{Bernoulli}\left(u_i\right) \quad \forall\ i = 1, ..., n \\
u_i &= F_\lambda\left(\eta_i\right) \\
\eta_i &= \mathbf{x_i}^T \boldsymbol{\beta} \\
(\boldsymbol{\beta}, \lambda) &\sim \pi\left(\boldsymbol{\beta}, \lambda\right)
\end{aligned} \tag{7}
$$

donde $F_\lambda(\cdot)$ es una FDA de una distribución potencia o reversa de potencia y $\pi\left(\boldsymbol{\beta}, \lambda\right)$ es la distribución conjunta *a priori* de los parámetros.

$$
L(\boldsymbol{\beta}, \lambda \mid \boldsymbol{y}, \boldsymbol{X}) = \prod_{i=1}^{n} \left[F_\lambda\left(\eta_i\right)\right]^{y_i} \left[1 - F_\lambda\left(\eta_i\right)\right]^{1-y_i} \tag{8}
$$

4. Inferencia Bayesiana de los parámetros

Con $\delta = \log(\lambda)$, $\delta \in \mathbb{R}$ asumimos independencia entre $\boldsymbol{\beta}$ y δ, luego

4. Inferencia Bayesiana de los parámetros

Con $\delta = \log(\lambda)$, $\delta \in \mathbb{R}$ asumimos independencia entre $\boldsymbol{\beta}$ y δ, luego

$$\pi(\boldsymbol{\beta}, \delta) = \pi_1(\boldsymbol{\beta})\, \pi_2(\delta) \tag{9}$$

Consideramos $\beta_j \sim N\left(0, 10^2\right)$ para $j = 1, ..., p$ y $\delta \sim U(-2, 2)$.

4. Inferencia Bayesiana de los parámetros

consecuentemente, la fdp *a posteriori* es de la forma

$$\pi(\boldsymbol{\beta}, \delta \mid \boldsymbol{y}, \boldsymbol{X}) \propto L\left(\boldsymbol{\beta}, \delta \mid \boldsymbol{y}, \boldsymbol{X}\right) \pi\left(\boldsymbol{\beta}\right) \pi\left(\delta\right) \tag{10}$$

$$\pi(\boldsymbol{\beta}, \delta \mid \boldsymbol{y}, \boldsymbol{X}) \propto \prod_{i=1}^{n} \left[F_{\delta}\left(\boldsymbol{x_i}^T \boldsymbol{\beta}\right)\right]^{y_i} \left[1 - F_{\delta}\left(\boldsymbol{x_i}^T \boldsymbol{\beta}\right)\right]^{1-y_i} \prod_{j=1}^{p} \exp\left\{-\frac{\beta_j^2}{2(10^2)}\right\}$$

4. Inferencia Bayesiana de los parámetros

consecuentemente, la fdp *a posteriori* es de la forma

$$\pi(\boldsymbol{\beta}, \delta \mid \boldsymbol{y}, \boldsymbol{X}) \propto L\left(\boldsymbol{\beta}, \delta \mid \boldsymbol{y}, \boldsymbol{X}\right) \pi\left(\boldsymbol{\beta}\right) \pi\left(\delta\right) \tag{10}$$

$$\pi(\boldsymbol{\beta}, \delta \mid \boldsymbol{y}, \boldsymbol{X}) \propto \prod_{i=1}^{n} \left[F_{\delta}\left(\boldsymbol{x_i}^T \boldsymbol{\beta}\right)\right]^{y_i} \left[1 - F_{\delta}\left(\boldsymbol{x_i}^T \boldsymbol{\beta}\right)\right]^{1-y_i} \prod_{j=1}^{p} \exp\left\{-\frac{\beta_j^2}{2(10^2)}\right\}$$

Una forma de hacer muestreo de esta distribución es a través de MCMC y en Python se puede usar a través del paquete Pystan el algoritmo No-U-Turn Sampler de Monte Carlo Hamiltoniano (HMC).

5. Implementación en Python

5. Implementación en Python

- **Instalación de** `Anaconda`

ANACONDA

https://www.anaconda.com/download/

5. Implementación en Python

- **Instalación de** Anaconda

ANACONDA https://www.anaconda.com/download/

- **Instalación de** package Pystan

```
conda install -c conda-forge pystan
```

5. Implementación en Python

- **Instalación de** `Anaconda`

ANACONDA https://www.anaconda.com/download/

- **Instalación de** `package Pystan`

```
conda install -c conda-forge pystan
```

- **Code**

```
import numpy as np
import pandas as pd
import pystan
dat = pd.read_csv('data.csv')
data_frame = dict(dat)
```

5. Implementación en Python

- **Instalación de Anaconda**

ANACONDA https://www.anaconda.com/download/

- **Instalación de package Pystan**

```
conda install -c conda-forge pystan
```

- **Code**

```
import numpy as np
import pandas as pd
import pystan
dat = pd.read_csv('data.csv')
data_frame = dict(dat)
```

```
data = {}
data['y'] = Y
data['X'] = X
data['k'] = data['X'].shape[1]
data['n'] = data['X'].shape[0]
```

5. Implementación en Python

```
model_code = """
data{
  int<lower=0> k;
  int<lower=0> n;
  int<lower=0, upper=1> y[n];
  matrix[n,k] X;
}
parameters{
  vector[k] beta;
  real loglambda;
}
transformed parameters  {
  real lambda;
  vector[n] p;
  vector[n] prob;
  lambda = exp(loglambda);
  for (i in 1:n) {
  prob[i] = pow(logistic_cdf(X[i]*beta, 0, 1),lambda);
  }
}
model {
  beta ~ normal(0.0,100);
  loglambda ~ uniform(-2,2);
  y ~ bernoulli(prob);
}"""
```

5. Implementación en Python

```python
npar=data['k']
#Modelo Power logistica
chains = 1
iter = 4000
warmup = 2000
thin = 2
seed = 11113     #151141
mod_l = pystan.StanModel(file="model_code")
fit_l = mod_l.sampling(
    data=data,
    chains=chains,
    iter=iter,
    warmup=warmup,
    thin=thin,
    seed=seed)
#summary
s_l = fit_l.summary()
```

6. Aplicación

Fue analizado datos relacionados a la calidad del vino, relacionado a 11 atributos químicos, aquí, considerados como covariables:

6. Aplicación

Fue analizado datos relacionados a la calidad del vino, relacionado a 11 atributos químicos, aquí, considerados como covariables: *acidez fija, acidez volátil, ácido cítrico, azúcar residual, cloruros, dióxido de azufre libre, dióxido de azufre total, densidad, pH, sulfatos* e *alcohol*.

6. Aplicación

Fue analizado datos relacionados a la calidad del vino, relacionado a 11 atributos químicos, aquí, considerados como covariables: *acidez fija, acidez volátil, ácido cítrico, azúcar residual, cloruros, dióxido de azufre libre, dióxido de azufre total, densidad, pH, sulfatos* e *alcohol*.
Los datos contienen un total de 4898 observaciones (Ajuste=3429, validación=1469). La calidad de vino varia de cero (muy mala) a 10 (excelente), i.e., $Z = \{0, \ldots, 10\}$. Nos interesa calidad de vino superior a 6.

6. Aplicación

La proporción de unos fue de 20%, consecuentemente son dados desbalanceados.

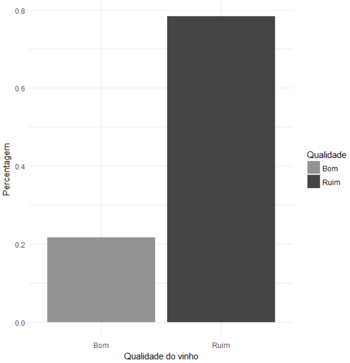

Figura 3: Calidad de vino

6. Aplicación

La proporción de unos fue de 20%, consecuentemente son dados desbalanceados.

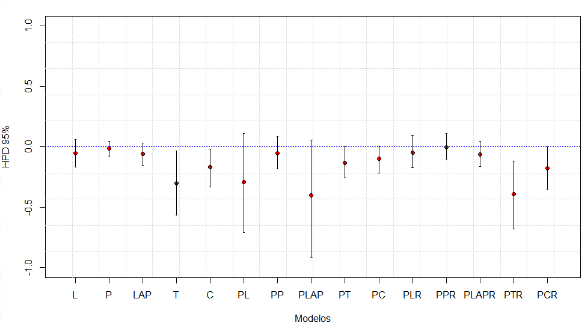

Figura 4: Intervalo HPD 95% para el coeficiente asociado a *dióxido de enxofre total*

6. Aplicación

Los códigos para todos los modelos fueron implementados en el lenguaje de `Stan` a través de `Python` usando `Pystan` (TEAM, 2017).

6. Aplicación

Los códigos para todos los modelos fueron implementados en el lenguaje de `Stan` a través de `Python` usando `Pystan` (TEAM, 2017).

La comparación de los modelos son realizadas usando los criterios DIC, EAIC, EBIC, WAIC y LOO. (GELMAN; HWANG; VEHTARI, 2014)

6. Aplicación

Modelo	DIC	EAIC	EBIC	IC	WAIC	LOO
Logística	4192.66	4204.99	4276.46	4202.32	4193.11	4198.49
Normal	4194.11	4205.73	4277.19	4204.49	4193.76	4193.73
Laplace	4197.72	4208.58	4280.05	4208.85	4197.54	4197.52
Cauchy	4217.10	4230.19	4301.65	4226.01	4218.80	4218.79
P-Logística	4221.42	4209.47	4280.93	4255.37	4201.16	4201.32
P-Normal	4190.28	4203.77	4275.23	4198.80	4191.58	4191.57
P-Laplace	4237.96	4227.21	4298.68	4270.70	4221.20	4221.62
P-Cauchy	4184.29	4194.52	4265.98	4196.06	4183.86	4183.84
RP-Logística	4195.76	4202.89	4274.35	4210.64	4192.84	4192.83
RP-Normal	4199.95	4209.80	4281.26	4212.10	4198.38	4190.38
RP-Laplace	4187.14	4197.600	4269.062	4198.689	4186.84	4186.82
RP-Cauchy	4180.83	4191.67	4263.13	4191.99	4181.40	4181.38

Tabela 2: Criterios de comparación de modelos para dados de calidad de vino.

6. Aplicación

Tabela 3: Estimación de los parámetros con el modelo RP-Cauchy para datos de calidad de vino

Atributo	Parâmetro	Estimativa	Desvio padrão	HPD (95%)
Intercepto	β_1	-0.895	0.161	(-1.202, -0.579)
Acidez fija	β_2	0.381	0.080	(0.228, 0.534)
Acidez volátil	β_3	-0.306	0.050	(-0.397, -0.207)
Azúcar residual	β_4	1.110	0.185	(0.746, 1.470)
Cloruros	β_5	-0.533	0.103	(-0.734, -0.342)
Dióxido de azufre libre	β_6	0.170	0.049	(0.077, 0.264)
Densidad	β_7	-1.524	0.293	(-2.126, -0.980)
pH	β_8	0.496	0.078	(0.350, 0.656)
Sulfatos	β_9	0.221	0.044	(0.142, 0.314)
Alcohol	β_{10}	0.339	0.126	(0.114, 0.606)
Parâmetro de assim.	λ	1.729	0.100	(1.536, 1.926)

6. Aplicación

Además consideramos la comparación de algunas medidas predictivas (FAWCETT, 2006)

Tabela 4: Medidas predictivas para regresión binaria con enlace Logit e potencia Cauchy

Modelo	Valor		Observado 0	Observado 1	ACC	TPR	TNR	AUC	CSI
Logístico	Previsto	**0**	1086	268	0.774	0.160	0.944	0.79	0.042
		1	64	51					
R-potencia Cauchy	Previsto	**0**	1168	170	0.855	0.341	0.964	0.81	0.068
		1	43	88					

ACC: Acuracia, TPR: Sensibilidad, TNR: Especificidad, AUC: Área debajo de la curva ROC, CSI: critical success index

$$ACC = \frac{VP + VN}{VP + FP + VN + FN}, \quad TPR = \frac{VP}{VP + FN}, \quad TNR = \frac{VN}{VN + FP}$$

$$CSI = \frac{VP}{VP + FP + FN}$$

6. Aplicación

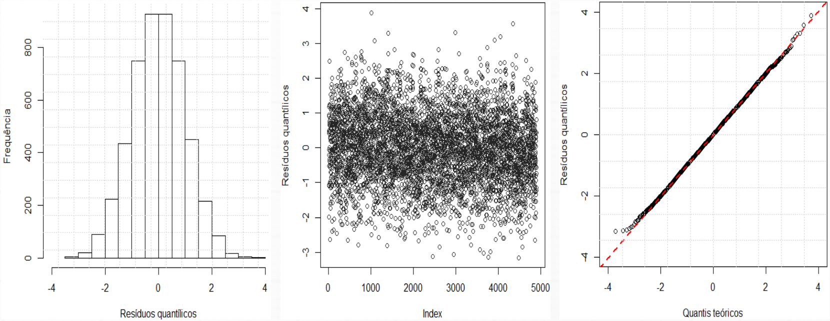

Figura 5: Residuos cuantílicos para a regresión binaria reversa de potencia Cauchy para calidad de vino

6. Aplicación

Portanto o modelo final escolhido tem a seguinte forma:

$$Y_i \sim \text{Bernoulli}\left(\hat{u}_i\right) \quad \forall\ i = 1, \ldots, 4898$$

$$\hat{u}_i = 1 - \left[\frac{1}{2} + \frac{1}{\pi}\text{arctan}\left(-\hat{\eta}_i\right)\right]^{1.73}$$

em que

$\hat{\eta}_i = -0.895 + 0.38x_2 - 0.31x_3 + 1.11x_4 - 0.53x_5 + 0.17x_6 - 1.52x_7 + 0.50x_8 + 0.22x_9 + 0.34x_{10}$

x_2=acidez fija, x_3=acidez volátil, x_4=azúcar residual, x_5=cloruros, x_6=dióxido de azufre libre, x_7=densidad, x_8=pH, x_9=Sulfatos e x_{10}=alcohol

6. Aplicación

- Los modelos estadísticos binarios potencia y reversa de potencia para datos desbalanceados, es una alternativa para problemas de clasificación.
- En la aplicación se mostró que los modelos usuales incluso la regresión logística no muestran un buen desempeño cuando los datos son desbalanceados. En este caso, el modelo reversa de potencia Cauchy presentó el mejor ajuste.
- Una ventaja de utilizar este tipo de modelos es la fácil interpretación de los parámetros asociados a los atributos o covariables.

7. Principales referencias

FAWCETT, T. An introduction to roc analysis. *Pattern recognition letters*, Elsevier, v. 27, n. 8, p. 861–874, 2006.

GELMAN, A.; HWANG, J.; VEHTARI, A. Understanding predictive information criteria for bayesian models. *Statistics and Computing*, Kluwer Academic Publishers, Hingham, MA, USA, v. 24, n. 6, p. 997–1016, nov. 2014. ISSN 0960-3174. Disponível em: ⟨http://dx.doi.org/10.1007/s11222-013-9416-2⟩.

LEMONTE, A. J.; BAZÁN, J. L. New links for binary regression: an application to coca cultivation in peru. *TEST*, Sep 2017. ISSN 1863-8260. Disponível em: ⟨https://doi.org/10.1007/s11749-017-0563-1⟩.

MCCULLAGH, P.; NELDER, J. *Generalized Linear Models, Second Edition*. [S.l.]: Taylor & Francis, 1989. (Chapman & Hall/CRC Monographs on Statistics & Applied Probability). ISBN 9780412317606.

TEAM, S. D. *PyStan: the Python interface to Stan, Version 2.16.0.0*. 2017. Disponível em: ⟨http://mc-stan.org⟩.