

Modelo de regresión mixto para datos proporcionales inflados con ceros y/o unos y creación del paquete **ZOIP** en R

Juan Camilo Díaz Zapata
Tutor: Freddy Hernández Barajas

Universidad Nacional de Colombia
Sede-Medellín

Escuela de Estadística

Contenido

- 1 Introducción
- 2 Metodología
 - Distribución ZOIP (Zeros Ones Inflated Proporcional)
 - Modelo de regresión ZOIP con efectos fijos
 - Modelo de regresión ZOIP con efectos mixtos
- 3 Paquete ZOIP de R
- 4 Aplicación a datos reales
- 5 Estudio de simulación
- 6 Conclusiones y trabajos futuros
- 7 Referencias

Motivación

1. Variables de interés como porcentajes, proporciones y tasas inflados con ceros y/o unos, son comunes en estudios clínicos, financieros, económicos y sociales.

Motivación

1. Variables de interés como porcentajes, proporciones y tasas inflados con zeros y/o unos, son comunes en estudios clínicos, financieros, económicos y sociales.
2. Los modelos de regresión mixto para datos proporcionales inflados con zeros y/o unos, son útiles para determinar el comportamiento de una variable proporcional a partir de variables consideradas como efectos fijos y aleatorios.

Motivación

1. Variables de interés como porcentajes, proporciones y tasas inflados con zeros y/o unos, son comunes en estudios clínicos, financieros, económicos y sociales.
2. Los modelos de regresión mixto para datos proporcionales inflados con zeros y/o unos, son útiles para determinar el comportamiento de una variable proporcional a partir de variables consideradas como efectos fijos y aleatorios.
3. La integración de los modelos estadísticos en un paquete del sistema computacional R permite que la utilización de los modelos estadísticos a problemas aplicados sea de gran facilidad para la comunidad estadística.

Estado del arte



Planteamiento del problema

1. Cómo se podría realizar un modelo de regresión mixto para datos proporcionales inflados con ceros y/o unos, sobre las distribuciones beta o simplex, donde la estimación de sus parámetros sea vía máxima verosimilitud.

Planteamiento del problema

1. Cómo se podría realizar un modelo de regresión mixto para datos proporcionales inflados con ceros y/o unos, sobre las distribuciones beta o simplex, donde la estimación de sus parámetros sea vía máxima verosimilitud.
2. No existe un paquete en R que reúna las principales distribuciones y modelos de regresión de efectos fijos y mixtos para modelar los datos proporcionales inflados con ceros y/o unos vía máxima verosimilitud y la cuadratura de Gauss-Hermite adaptativa.

Planteamiento del problema

1. Cómo se podría realizar un modelo de regresión mixto para datos proporcionales inflados con ceros y/o unos, sobre las distribuciones beta o simplex, donde la estimación de sus parámetros sea vía máxima verosimilitud.
2. No existe un paquete en R que reúna las principales distribuciones y modelos de regresión de efectos fijos y mixtos para modelar los datos proporcionales inflados con ceros y/o unos vía máxima verosimilitud y la cuadratura de Gauss-Hermite adaptativa.

Objetivo general

Objetivo general

Proponer una metodología para estimación de los parámetros de un modelo de regresión mixto para datos proporcionales inflados con ceros y/o unos.

Distribución ZOIP (Zeros Ones Inflated Proporcional)

Si la variable aleatoria Y tiene distribución ZOIP con parámetros μ , σ , p_0 y p_1 , se denotará como $Y \sim \text{ZOIP}(\mu, \sigma, p_0, p_1)$, la función de densidad de probabilidad está dada por:

$$g(y; \mu, \sigma, p_0, p_1) = \begin{cases} p_0 & \text{si } y = 0, \\ p_1 & \text{si } y = 1, \\ (1 - p_0 - p_1)f(y; \mu, \sigma) & \text{si } y \in (0, 1) \end{cases}$$

Distribución ZOIP (Zeros Ones Inflated Proporcional)

La media y varianza de Y , están dadas por:

$$E(y) = p_1 + (1 - p_0 - p_1)E^*(y)$$

$$\begin{aligned} Var(y) = & p_1(1 - p_1) + \\ & + (1 - p_0 - p_1) [Var^*(y) + (p_0 + p_1)[E^*(y)]^2 - 2E^*(y)p_1] \end{aligned}$$

Modelo de regresión ZOIP con efectos fijos

Sea y_1, y_2, \dots, y_n variables aleatorias independientes tal que cada y_i , para $i = 1, \dots, n$, tiene función de densidad de probabilidad dado por la distribución ZOIP con parámetros $\mu = \mu_i$, $\sigma = \sigma_i$, $p_0 = p_{0i}$, y $p_1 = p_{1i}$. Se asume que μ_i , σ_i , p_{0i} y p_{1i} se definen como:

$$\begin{aligned}
 h_1(\mu_i) &= \mathbf{x}_{i1}^\top \boldsymbol{\beta}_1, \\
 h_2(\sigma_i) &= \mathbf{x}_{i2}^\top \boldsymbol{\beta}_2, \\
 h_3(p_{0i}) &= \mathbf{x}_{i3}^\top \boldsymbol{\beta}_3, \\
 h_4(p_{1i}) &= \mathbf{x}_{i4}^\top \boldsymbol{\beta}_4
 \end{aligned} \tag{1}$$

Estimación vía máxima verosimilitud

La función de verosimilitud para $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\beta}_3^\top, \boldsymbol{\beta}_4^\top)^\top$, basado en una muestra \mathbf{y}_i de tamaño n de observaciones independientes, es de la forma:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n g(\mathbf{y}_i; \mu_i, \sigma_i, p_{0i}, p_{1i}) = \prod_{i=1}^n f_0(\mathbf{y}_i; p_{0i}) \prod_{i=1}^n f_1(\mathbf{y}_i; p_{1i}) \prod_{i=1}^n f_2(\mathbf{y}_i; \mu_i, \sigma_i)$$

donde las funciones $f_0(\mathbf{y}_i; p_{0i})$ y $f_1(\mathbf{y}_i; p_{1i})$ son explicadas por:

$$f_j(\mathbf{y}_i; p_{ji}) = p_{ji}^{S_j(y_i)} (1 - p_{ji})^{1 - S_j(y_i)} \quad ; \quad j = 0, 1$$

Con

$$S_j(y_i) = \begin{cases} 1 & \text{si } y_i = j \\ 0 & \text{si } y_i \neq j \end{cases} \quad ; \quad j = 0, 1$$

Estimación vía máxima verosimilitud

$$\ell(\boldsymbol{\theta}) = \ell_0(\boldsymbol{\beta}_3) + \ell_1(\boldsymbol{\beta}_4) + \ell_2(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$$

$$\ell_0(\boldsymbol{\beta}_3) = \sum_{i=1}^n p_{0i}^{S_0(y_i)} (1 - p_{0i})^{1-S_0(y_i)}$$

$$\ell_1(\boldsymbol{\beta}_4) = \sum_{i=1}^n p_{1i}^{S_1(y_i)} (1 - p_{1i})^{1-S_1(y_i)}$$

$$\ell_2(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \sum_{i=1: y_i \in (0,1)}^n f_3(y_i; \mu_i, \sigma_i)$$

Con $p_{0i} = h_3^{-1}(\mathbf{x}_{i3}^\top \boldsymbol{\beta}_3)$, $p_{1i} = h_4^{-1}(\mathbf{x}_{i4}^\top \boldsymbol{\beta}_4)$, $\mu_i = h_1^{-1}(\mathbf{x}_{i1}^\top \boldsymbol{\beta}_1)$ y $\sigma_i = h_2^{-1}(\mathbf{x}_{i2}^\top \boldsymbol{\beta}_2)$ como se definió en (1).

Modelo de regresión ZOIP con efectos mixtos

1. Los efectos aleatorios sobre un modelo de regresión ocurren cuando los datos dependen o se puede ver afectados por una variable netamente identificativa.

Modelo de regresión ZOIP con efectos mixtos

1. Los efectos aleatorios sobre un modelo de regresión ocurren cuando los datos dependen o se puede ver afectados por una variable netamente identificativa.
2. Un modelo de regresión con efectos mixtos es útil porque se contempla de manera correcta el problema, ya que hay correlación entre las observaciones.

Modelo de regresión ZOIP con efectos mixtos

1. Los efectos aleatorios sobre un modelo de regresión ocurren cuando los datos dependen o se puede ver afectados por una variable netamente identificativa.
2. Un modelo de regresión con efectos mixtos es útil porque se contempla de manera correcta el problema, ya que hay correlación entre las observaciones.
3. Un modelo de regresión con efectos mixtos es útil por la forma de determinar los parámetros de un efecto aleatorio.

Modelo de regresión ZOIP con efectos mixtos

Sea y_{ij} la j -ésima medida del i -ésimo grupo, una formulación matemática para el modelo es la siguiente:

$$\begin{aligned}
 y_{ij} | \gamma_{1i}, \gamma_{2i} &\stackrel{\text{ind}}{\sim} \text{ZOIP}(\mu_{ij}, \sigma_{ij}, p_{0ij}, p_{1ij}), \\
 h_1(\mu_{ij}) &= \mathbf{x}_{ij1}^\top \boldsymbol{\beta}_1 + \gamma_{1i}, \\
 h_2(\sigma_{ij}) &= \mathbf{x}_{ij2}^\top \boldsymbol{\beta}_2 + \gamma_{2i}, \\
 h_3(p_{0ij}) &= \mathbf{x}_{ij3}^\top \boldsymbol{\beta}_3, \\
 h_4(p_{1ij}) &= \mathbf{x}_{ij4}^\top \boldsymbol{\beta}_4, \\
 \gamma_{1i} &\stackrel{\text{i.i.d}}{\sim} N(0, \lambda_1^2), \\
 \gamma_{2i} &\stackrel{\text{i.i.d}}{\sim} N(0, \lambda_2^2),
 \end{aligned} \tag{2}$$

con $i = 1, 2, \dots, N$ y $j = 1, 2, \dots, n_i$.

Estimación vía máxima verosimilitud y AGHQ

El vector de parámetros para el modelo (2) es $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\beta}_3^\top, \boldsymbol{\beta}_4^\top, \lambda_1, \lambda_2)^\top$ y de esa forma la función de log-verosimilitud $\ell(\boldsymbol{\theta})$ está dado por:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^N \log \left[\int_{\mathbb{R}^2} \prod_{j=1}^{n_i} f_y(y_{ij} | \gamma_{1i}, \gamma_{2i}; \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4) \cdot f(\gamma_{1i} | \lambda_1) f(\gamma_{2i} | \lambda_2) d\gamma_{1i} d\gamma_{2i} \right]$$

Estimación vía máxima verosimilitud y AGHQ

Para encontrar el punto máximo de la función de log-verosimilitud anterior, es necesario realizar una aproximación por medio de la cuadratura de Gauss-Hermite adaptativa con o sin *pruning*, que permita encontrar dicho máximo.

$$\ell(\boldsymbol{\theta}) \approx \sum_{i=1}^N \log \left[\sum_{k_1=1}^{Q_1} \sum_{k_2=1}^{Q_2} \prod_{j=1}^{n_i} f(y_{ij} | \sqrt{2}\lambda_1 z_{k_1}, \sqrt{2}\lambda_2 z_{k_2}; \beta_1, \beta_2, \beta_3, \beta_4) \cdot \frac{w_{k_1} w_{k_2}}{\pi} \right]$$

AGHQ

Instalación del Paquete **ZOIP** de R

Para acceder a la última versión del paquete **ZOIP** se encuentra ubicada en GitHub, el cual es un alojamiento de repositorios Git, para obtener dicha versión es necesario ejecutar el siguiente código que instala el paquete **devtools**, que es necesario para descargar el paquete **ZOIP**.

```
#Para instalar la version disponible en el CRAN
install.packages("ZOIP")

#Para instalar la version en desarrollo desde GitHub
if (!require("devtools")) install.packages("devtools")
devtools::install_github("jucdiaz/ZOIP", force = TRUE)
library(ZOIP) # Carga el paquete
```

Funciones en el paquete ZOIP

El paquete **ZOIP** de R es útil para ajustar distribución ZOIP, modelos de regresión de efectos fijos y mixtos para datos proporcionales inflados con ceros y/o unos. Estas son las principales funciones:

dZOIP

pZOIP

qZOIP

rZOIP

RM.ZOIP

RMM.ZOIP



Función RM.ZOIP

```
RM.ZOIP(formula.mu, formula.sigma = ~1, formula.p0 = ~1,  
        formula.p1 = ~1, data, link = c("identity", "identity",  
        "identity", "identity"), family = "R-S",  
optimizer = "nlminb")
```

Ejemplo:

```
link <- c("logit", "logit", "identity", "logit")  
mod<-RM.ZOIP(formula.mu = y_i ~ x1, formula.sigma = ~x1 + x2,  
             formula.p0 = ~1, formula.p1 = ~x2, data = data,  
             link = link, family = "R-S")  
summary(mod)
```


Función RMM.ZOIP

```
RMM.ZOIP(formula.mu, formula.sigma = ~1, formula.p0 = ~1,  
          formula.p1 = ~1, data, formula.random,  
          link = c("identity", "identity", "identity", "identity"),  
          family = "R-S", optimizer = "nlminb",  
          n.points = 11, pruning = TRUE)
```

Ejemplo:

```
link <- c("logit", "logit", "identity", "identity")  
mod<-RMM.ZOIP(formula.mu = Y ~ log(Days), formula.sigma = ~log(Days),  
              formula.p0 = ~1, formula.p1 = ~1, data = base,  
              formula.random = ~1 | subject, link = link)  
summary(mod)
```

Ajuste distribución ZOIP

Ajuste de una distribución ZOIP a datos reales sobre la utilización de una tarjeta de crédito (tdc) en un banco de Colombia.

Ajuste distribución ZOIP

Ajuste de una distribución ZOIP a datos reales sobre la utilización de una tarjeta de crédito (tdc) en un banco de Colombia.

```
RM.ZOIP(data = porc_tdc, family = "R-S")  
RM.ZOIP(data = porc_tdc, family = "F-C")  
RM.ZOIP(data = porc_tdc, family = "Original")  
RM.ZOIP(data = porc_tdc, family = "Simplex")
```

Ajuste distribución ZOIP

Familia	Parámetro	Estimación	Error estándar	Valor P	Log-Verosimilitud
R-S	μ	0.4040	0.0037	$< 2.2e^{-16}$	5854.067
	σ	0.6601	0.0027	$< 2.2e^{-16}$	
	p_0	0.2219	0.0043	$< 2.2e^{-16}$	
	p_1	0.0695	0.0027	$< 2.2e^{-16}$	
F-C	μ	0.4040	0.0037	$< 2.2e^{-16}$	5854.067
	σ	0.4040	0.0037	$< 2.2e^{-16}$	
	p_0	0.2219	0.0043	$< 2.2e^{-16}$	
	p_1	0.0695	0.0027	$< 2.2e^{-16}$	
original	μ	0.5233	0.0080	$< 2.2e^{-16}$	5854.067
	σ	0.7719	0.0130	$< 2.2e^{-16}$	
	p_0	0.2219	0.0043	$< 2.2e^{-16}$	
	p_1	0.0695	0.0027	$< 2.2e^{-16}$	
simplex	μ	0.5741	0.0010	$< 2.2e^{-16}$	54425.63
	σ	4885.44	18.2430	$< 2.2e^{-16}$	
	p_0	0.1497	0.0032	$< 2.2e^{-16}$	
	p_1	0.0090	0.0004	$< 2.2e^{-16}$	

Cuadro 1: Ajuste de diferentes distribuciones ZOIP en el porcentaje de utilización de una tdc, en un banco colombiano. Fuente: Cortesía del banco.

Ajuste distribución ZOIP

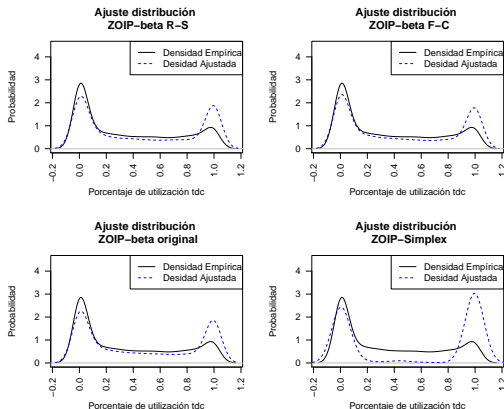


Figura 1: Ajuste de diferentes distribuciones y parametrizaciones ZOIP al porcentaje de utilización de una tdc.

Aplicación modelo de regresión ZOIP

Se ajusto un modelo de regresión ZOIP mediante el paquete **ZOIP**, que permitió explicar el comportamiento utilización de una tdc mediante unas variables dadas.

$$y_i \sim ZOIP(\mu_i, \sigma_i, p_{0i}, p_{1i}), \quad i = 1, 2, \dots, 9206$$

$$h_1(\mu_i) = \beta_{10} + \beta_{11}x_{1i} + \beta_{12}x_{2i} + \beta_{13}x_{3i},$$

$$h_2(\sigma_i) = \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i} + \beta_{23}x_{3i},$$

$$h_3(p_{0i}) = \beta_{30} + \beta_{31}x_{1i} + \beta_{32}x_{2i} + \beta_{33}x_{3i},$$

$$h_4(p_{1i}) = \beta_{40} + \beta_{41}x_{1i} + \beta_{42}x_{2i} + \beta_{43}x_{3i},$$

y_i es el porcentaje de utilización de la i -ésima tdc.

Aplicación modelo de regresión ZOIP

Se ajusto un modelo de regresión ZOIP mediante el paquete **ZOIP**, que permitió explicar el comportamiento utilización de una tdc mediante unas variables dadas.

$$y_i \sim ZOIP(\mu_i, \sigma_i, p_{0i}, p_{1i}), \quad i = 1, 2, \dots, 9206$$

$$h_1(\mu_i) = \beta_{10} + \beta_{11}x_{1i} + \beta_{12}x_{2i} + \beta_{13}x_{3i},$$

$$h_2(\sigma_i) = \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i} + \beta_{23}x_{3i},$$

$$h_3(p_{0i}) = \beta_{30} + \beta_{31}x_{1i} + \beta_{32}x_{2i} + \beta_{33}x_{3i},$$

$$h_4(p_{1i}) = \beta_{40} + \beta_{41}x_{1i} + \beta_{42}x_{2i} + \beta_{43}x_{3i},$$

y_i es el porcentaje de utilización de la i -ésima tdc.

x_{1i} es el valor del **score** del i -ésimo individuo asociada a la tarjeta de crédito.

Aplicación modelo de regresión ZOIP

Se ajusto un modelo de regresión ZOIP mediante el paquete **ZOIP**, que permitió explicar el comportamiento utilización de una tdc mediante unas variables dadas.

$$y_i \sim ZOIP(\mu_i, \sigma_i, p_{0i}, p_{1i}), \quad i = 1, 2, \dots, 9206$$

$$h_1(\mu_i) = \beta_{10} + \beta_{11}x_{1i} + \beta_{12}x_{2i} + \beta_{13}x_{3i},$$

$$h_2(\sigma_i) = \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i} + \beta_{23}x_{3i},$$

$$h_3(p_{0i}) = \beta_{30} + \beta_{31}x_{1i} + \beta_{32}x_{2i} + \beta_{33}x_{3i},$$

$$h_4(p_{1i}) = \beta_{40} + \beta_{41}x_{1i} + \beta_{42}x_{2i} + \beta_{43}x_{3i},$$

y_i es el porcentaje de utilización de la i -ésima tdc.

x_{1i} es el valor del **score** del i -ésimo individuo asociada a la tarjeta de crédito.

x_{2i} es valor del **promedio de cuotas** al que difiere sus compras de la i -ésima tarjeta de crédito.

Aplicación modelo de regresión ZOIP

Se ajusto un modelo de regresión ZOIP mediante el paquete **ZOIP**, que permitió explicar el comportamiento utilización de una tdc mediante unas variables dadas.

$$y_i \sim ZOIP(\mu_i, \sigma_i, p_{0i}, p_{1i}), \quad i = 1, 2, \dots, 9206$$

$$h_1(\mu_i) = \beta_{10} + \beta_{11}x_{1i} + \beta_{12}x_{2i} + \beta_{13}x_{3i},$$

$$h_2(\sigma_i) = \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i} + \beta_{23}x_{3i},$$

$$h_3(p_{0i}) = \beta_{30} + \beta_{31}x_{1i} + \beta_{32}x_{2i} + \beta_{33}x_{3i},$$

$$h_4(p_{1i}) = \beta_{40} + \beta_{41}x_{1i} + \beta_{42}x_{2i} + \beta_{43}x_{3i},$$

y_i es el porcentaje de utilización de la i -ésima tdc.

x_{1i} es el valor del **score** del i -ésimo individuo asociada a la tarjeta de crédito.

x_{2i} es valor del **promedio de cuotas** al que difiere sus compras de la i -ésima tarjeta de crédito.

x_{3i} es el valor del **cupo otorgado** de la i -ésima tarjeta de crédito.

Familia	Parámetro	β 's	Estimación	Error estándar	Valor P	Log-Verosimilitud
R-S	μ	$\hat{\beta}_{10}$	-0.046	0.050	0.3618	-5414.738
		$\hat{\beta}_{11}$	-0.354	0.107	0.0009	
		$\hat{\beta}_{12}$	0.022	0.002	$< 2.2e^{-16}$	
		$\hat{\beta}_{13}$	-0.025	0.009	0.0074	
		$\hat{\beta}_{20}$	0.822	0.038	$< 2.2e^{-16}$	
	σ	$\hat{\beta}_{21}$	-0.197	0.078	0.0114	
		$\hat{\beta}_{22}$	-0.006	0.002	0.0013	
		$\hat{\beta}_{23}$	-0.003	0.007	0.6741	
		$\hat{\beta}_{30}$	-1.496	0.101	$< 2.2e^{-16}$	
	p_0	$\hat{\beta}_{31}$	0.724	0.185	$8.87e^{-5}$	
		$\hat{\beta}_{32}$	-0.153	0.009	$< 2.2e^{-16}$	
		$\hat{\beta}_{33}$	0.002	0.015	0.1243	
		$\hat{\beta}_{40}$	-1.480	0.095	$< 2.2e^{-16}$	
	p_1	$\hat{\beta}_{41}$	-0.630	0.254	0.0132	
		$\hat{\beta}_{42}$	0.011	0.006	0.0666	
		$\hat{\beta}_{43}$	-0.069	0.022	0.0022	

A set of small navigation icons typically found in Beamer presentations, including symbols for back, forward, search, and other slide controls.

Se planteó un modelo de regresión ZOIP-beta con intercepto aleatorio en el parámetro de la media y la varianza, dado por la variable *ciudad* y un efecto fijo en la media y la varianza dado por la variable *total mora*, para la variable respuesta porcentaje de utilización de una tdc.

$$\begin{aligned} y_{ij} | \gamma_{1i}, \gamma_{2i} &\stackrel{\text{ind}}{\sim} \text{ZOIP}(\mu_{ij}, \sigma_{ij}, p_0, p_1), \\ h_1(\mu_{ij}) &= \beta_{10} + \gamma_{1i} + \beta_{11}x_{1ij}, \\ h_2(\sigma_{ij}) &= \beta_{20} + \gamma_{2i} + \beta_{21}x_{1ij}, \\ h_3(p_0) &= \beta_{30}, \\ h_4(p_1) &= \beta_{40}, \end{aligned}$$

y_{ij} es el porcentaje de utilización de la j -ésima tdc perteneciente a la i -ésima ciudad, $i = 1, 2, \dots, 10$ y $j = 1, 2, \dots, 15$.

x_{1ij} : es el valor del tiempo en mora en meses de la j -ésima tdc asociada a la i -ésima ciudad.

Aplicación modelo de regresión ZOIP mixto

Parámetro	β 's	Estimación	Error estándar	Valor P
μ	$\hat{\beta}_{10}$	-1.13	0.24	$5.4e^{-6}$
	$\hat{\beta}_{11}$	0.33	0.13	0.008
	$\hat{\lambda}_1$	0.51	0.304	0.093
σ	$\hat{\beta}_{20}$	0.33	0.20	0.095
	$\hat{\beta}_{21}$	0.14	0.09	0.157
	$\hat{\lambda}_2$	0.40	0.31	0.199
p_0	$\hat{\beta}_{30}$	0.23	0.03	$3.4e^{-11}$
p_1	$\hat{\beta}_{40}$	0.07	0.02	0.0011

Cuadro 3: Estimación de los efectos fijos y los componentes de varianza λ_1 y λ_2 del modelo de regresión ZOIP mixto para el porcentaje utilización de una tdc.

Aplicación modelo de regresión ZOIP mixto

El modelo propuesto se puede reescribir con los parámetros estimados así:

$$\begin{aligned} y_{ij} | \gamma_{1i}, \gamma_{2i} &\stackrel{\text{ind}}{\sim} ZOIP(\mu_{ij}, \sigma_{ij}, p_0, p_1), \\ h_1(\mu_{ij}) &= -1.13 + \gamma_{1i} + 0.33x_{1ij}, \\ h_2(\sigma_{ij}) &= 0.33 + \gamma_{2i} + 0.14x_{1ij}, \\ h_3(p_0) &= 0.23, \\ h_4(p_1) &= 0.07, \end{aligned}$$

donde $\gamma_{1i} \sim N(0, 0.51^2)$ y $\gamma_{2i} \sim N(0, 0.40^2)$.

Objetivos del estudio de simulación

1. Determinar el desempeño del proceso de estimación de los parámetros.

Objetivos del estudio de simulación

1. Determinar el desempeño del proceso de estimación de los parámetros.
2. Observar la funcionalidad de las funciones del paquete **ZOIP**.

Objetivos del estudio de simulación

1. Determinar el desempeño del proceso de estimación de los parámetros.
2. Observar la funcionalidad de las funciones del paquete **ZOIP**.
3. Determinar los mejores escenarios de simulación para la estimación de los parámetros de acuerdo con el modelo propuesto.

Objetivos del estudio de simulación

1. Determinar el desempeño del proceso de estimación de los parámetros.
2. Observar la funcionalidad de las funciones del paquete **ZOIP**.
3. Determinar los mejores escenarios de simulación para la estimación de los parámetros de acuerdo con el modelo propuesto.
4. Conocer el tamaño muestral mínimo para obtener una estimación razonablemente buena de los parámetros de los modelos propuestos.

Estudio de simulación: Distribución ZOIP

Escenario 1: Distribución ZOIP-beta($\mu = 0.6, \sigma = 0.2, p_0 = 0.03, p_1 = 0.05$) para el caso de la parametrización de [9].

Estudio de simulación: Distribución ZOIP

Escenario 1: Distribución ZOIP-beta($\mu = 0.6, \sigma = 0.2, p_0 = 0.03, p_1 = 0.05$) para el caso de la parametrización de [9].

Escenario 2: Distribución ZOIP-beta($\mu = 0.6, \sigma = 24, p_0 = 0.03, p_1 = 0.05$) para el caso de la parametrización de [1].

Estudio de simulación: Distribución ZOIP

Escenario 1: Distribución ZOIP-beta($\mu = 0.6, \sigma = 0.2, p_0 = 0.03, p_1 = 0.05$) para el caso de la parametrización de [9].

Escenario 2: Distribución ZOIP-beta($\mu = 0.6, \sigma = 24, p_0 = 0.03, p_1 = 0.05$) para el caso de la parametrización de [1].

Escenario 3: Distribución ZOIP-beta($\mu = 14.4, \sigma = 9.6, p_0 = 0.03, p_1 = 0.05$) en la parametrización original.

Estudio de simulación: Distribución ZOIP

Escenario 1: Distribución ZOIP-beta($\mu = 0.6, \sigma = 0.2, p_0 = 0.03, p_1 = 0.05$) para el caso de la parametrización de [9].

Escenario 2: Distribución ZOIP-beta($\mu = 0.6, \sigma = 24, p_0 = 0.03, p_1 = 0.05$) para el caso de la parametrización de [1].

Escenario 3: Distribución ZOIP-beta($\mu = 14.4, \sigma = 9.6, p_0 = 0.03, p_1 = 0.05$) en la parametrización original.

Escenario 4: Distribución ZOIP-simplex($\mu = 0.4, \sigma = 0.2, p_0 = 0.03, p_1 = 0.05$).

Estudio de simulación: Distribución ZOIP

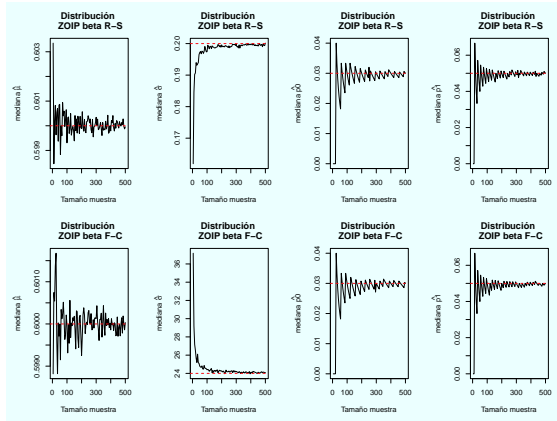


Figura 2: Mediana de los parámetros estimados en el escenario 1 y 2 para distintos valores de n , las líneas rojas representan el verdadero valor del parámetro.

Estudio de simulación: Distribución ZOIP

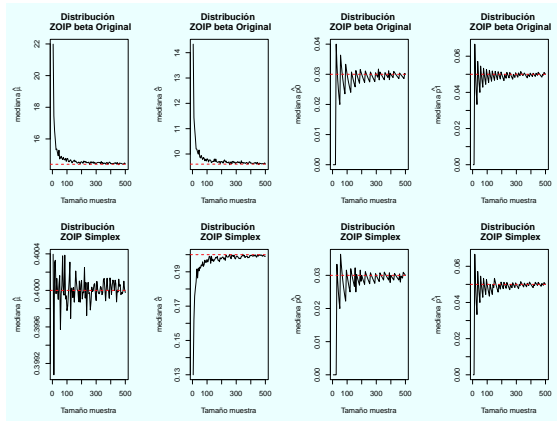


Figura 3: Mediana de los parámetros estimados en el escenario 3 y 4 para distintos valores de n , las líneas rojas representan el verdadero valor del parámetro.

Estudio de simulación: Distribución ZOIP

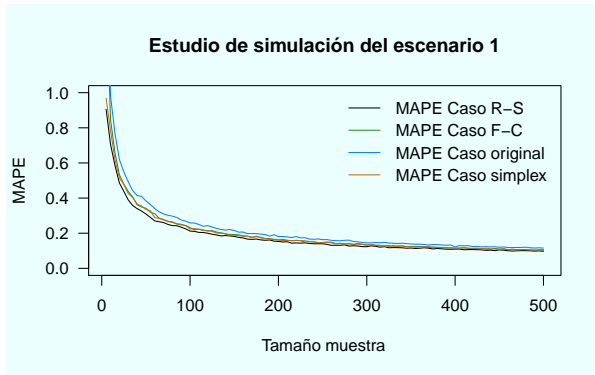


Figura 4: Mape (Error porcentual absoluto medio) $\sum_{i=1}^n (|y_i - \hat{y}_i|/y_i)/n$ para los escenarios de simulación y para distintos valores de n .

Estudio de simulación: Modelo regresión ZOIP

Estructura del modelo de regresión ZOIP con efectos fijos, asociada a los estudios de simulación.

$$\begin{aligned}y_i &\sim ZOIP(\mu_i, \sigma_i, p_{0i}, p_{1i}), \\h_1(\mu_i) &= \beta_{10} + \beta_{11}x_{1i}, \\h_2(\sigma_i) &= \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i}, \\h_3(p_{0i}) &= \beta_{30}, \\h_4(p_{1i}) &= \beta_{40} + \beta_{41}x_{2i}\end{aligned}$$

Escenario 1, 2, 3: μ : $\beta_{10} = 0.2$, $\beta_{11} = -1$ - σ : $\beta_{20} = 0.3$, $\beta_{21} = 3$, $\beta_{22} = 0.9$ - p_0 : $\beta_{30} = 0.07$ - p_1 : $\beta_{40} = 0.02$, $\beta_{41} = -4$ con la parametrización de [1], original y simplex.

Estudio de simulación: Modelo regresión ZOIP

Estructura del modelo de regresión ZOIP con efectos fijos, asociada a los estudios de simulación.

$$\begin{aligned}y_i &\sim ZOIP(\mu_i, \sigma_i, p_{0i}, p_{1i}), \\h_1(\mu_i) &= \beta_{10} + \beta_{11}x_{1i}, \\h_2(\sigma_i) &= \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i}, \\h_3(p_{0i}) &= \beta_{30}, \\h_4(p_{1i}) &= \beta_{40} + \beta_{41}x_{2i}\end{aligned}$$

Escenario 1, 2, 3: μ : $\beta_{10} = 0.2$, $\beta_{11} = -1$ - σ : $\beta_{20} = 0.3$, $\beta_{21} = 3$, $\beta_{22} = 0.9$ - p_0 : $\beta_{30} = 0.07$ - p_1 : $\beta_{40} = 0.02$, $\beta_{41} = -4$ con la parametrización de [1], original y simplex.

Escenario 4: μ : $\beta_{10} = 0.2$, $\beta_{11} = -1$ - σ : $\beta_{20} = 0.3$, $\beta_{21} = -2$, $\beta_{22} = -4$ - p_0 : $\beta_{30} = 0.07$ - p_1 : $\beta_{40} = 0.02$, $\beta_{41} = -4$ con la parametrización de [9].

Estudio de simulación: Modelo regresión ZOIP

Estructura del modelo de regresión ZOIP con efectos fijos, asociada a los estudios de simulación.

$$\begin{aligned}y_i &\sim ZOIP(\mu_i, \sigma_i, p_{0i}, p_{1i}), \\h_1(\mu_i) &= \beta_{10} + \beta_{11}x_{1i}, \\h_2(\sigma_i) &= \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i}, \\h_3(p_{0i}) &= \beta_{30}, \\h_4(p_{1i}) &= \beta_{40} + \beta_{41}x_{2i}\end{aligned}$$

Escenario 1, 2, 3: μ : $\beta_{10} = 0.2$, $\beta_{11} = -1$ - σ : $\beta_{20} = 0.3$, $\beta_{21} = 3$, $\beta_{22} = 0.9$ - p_0 : $\beta_{30} = 0.07$ - p_1 : $\beta_{40} = 0.02$, $\beta_{41} = -4$ con la parametrización de [1], original y simplex.

Escenario 4: μ : $\beta_{10} = 0.2$, $\beta_{11} = -1$ - σ : $\beta_{20} = 0.3$, $\beta_{21} = -2$, $\beta_{22} = -4$ - p_0 : $\beta_{30} = 0.07$ - p_1 : $\beta_{40} = 0.02$, $\beta_{41} = -4$ con la parametrización de [9].

Estudio de simulación: Modelo regresión ZOIP

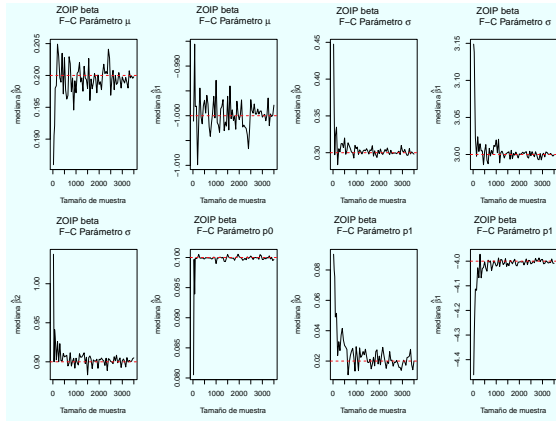


Figura 5: Simulación de un modelo de regresión ZOIP-beta para la parametrización F-C con diferentes valores de n .

Estudio de simulación: Modelo regresión ZOIP

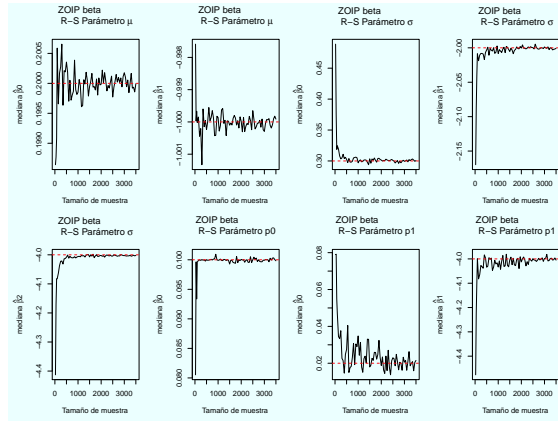


Figura 6: Simulación de un modelo de regresión ZOIP-beta para la parametrización R-S con diferentes valores de n .

Estudio de simulación: Modelo regresión ZOIP

Familia	μ		σ			p_0	p_1	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_0$	$\hat{\beta}_1$
R-S	1.25	0.32	1.45	2.55	1.38	4.86	383.09	4.88
F-C	14.22	3.96	22.21	2.9	10.14	4.86	91.21	4.88
original	22.34	8.03	22.55	3.62	8.69	4.84	90.58	4.96
simplex	13.93	5.89	24.49	3.11	11.01	4.85	91.15	4.81

Cuadro 4: Mediana del MAPE (Error porcentual absoluto medio) en porcentaje para los diferentes Parámetros en las diferentes parametrizaciones.

Estudio de simulación: Modelo regresión ZOIP mixto

El modelo de regresión ZOIP mixto en el cual se basa el estudio de simulación se muestra a continuación:

$$\begin{aligned}y_{ij}|\gamma_{1i}, \gamma_{2i} &\stackrel{\text{ind}}{\sim} \text{ZOIP}(\mu_{ij}, \sigma_{ij}, p_0, p_1), \\h_1(\mu_{ij}) &= -1.13 + \gamma_{1i} + 0.33x_{1ij}, \\h_2(\sigma_{ij}) &= 0.33 + \gamma_{2i} + 0.14x_{1ij}, \\h_3(p_0) &= 0.23, \\h_4(p_1) &= 0.07,\end{aligned}$$

donde $\gamma_{1i} \sim N(0, 0.51^2)$ y $\gamma_{2i} \sim N(0, 0.40^2)$.

Estudio de simulación: Modelo regresión ZOIP mixto

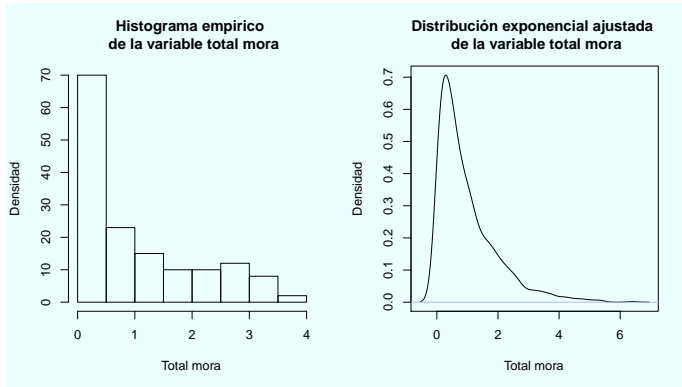


Figura 8: Ajuste de la distribución exponencial a la variable *total mora*

Estudio de simulación: Modelo regresión ZOIP mixto

Parámetro	β 's	Con <i>pruning</i>			Sin <i>pruning</i>		
		$n_i = 5$	$n_i = 20$	$n_i = 50$	$n_i = 5$	$n_i = 20$	$n_i = 50$
μ	$\beta_{10} = -1.13$	-1.137	-1.110	-1.076	-1.128	-1.120	-1.080
	$\beta_{11} = 0.33$	0.321	0.327	0.326	0.331	0.330	0.327
	$\lambda_1 = 0.51$	0.879	0.576	0.507	0.882	0.568	0.498
σ	$\beta_{20} = 0.33$	0.445	0.380	0.336	0.452	0.377	0.345
	$\beta_{21} = 0.14$	0.072	0.118	0.132	0.066	0.121	0.133
	$\lambda_2 = 0.4$	0.728	0.450	0.396	0.727	0.456	0.398
p_0	$\beta_{30} = 0.23$	0.220	0.230	0.230	0.220	0.230	0.230
p_1	$\beta_{40} = 0.07$	0.060	0.070	0.070	0.060	0.070	0.072
Med tiempo(Seg)		115.72	130.85	140.58	61.69	163.36	218.68
Med num. iter		22	30	34	22	30	34

Cuadro 5: Mediana de los parámetros estimados en el modelo ZOIP mixto para tres tamaños de muestra y con la estrategia de “con y sin *pruning*” y para todos los valores de Q .

Estudio de simulación: Modelo regresión ZOIP mixto

Parámetro	β 's	Con <i>pruning</i>			Sin <i>pruning</i>		
		$Q = 3$	$Q = 10$	$Q = 20$	$Q = 3$	$Q = 10$	$Q = 20$
μ	$\beta_{10} = -1.13$	-1.122	-1.069	-1.120	-1.117	-1.076	-1.129
	$\beta_{11} = 0.33$	0.323	0.322	0.333	0.334	0.322	0.329
	$\lambda_1 = 0.51$	0.632	0.626	0.634	0.629	0.616	0.623
σ	$\beta_{20} = 0.33$	0.400	0.365	0.366	0.382	0.379	0.373
	$\beta_{21} = 0.14$	0.117	0.119	0.120	0.123	0.117	0.121
	$\lambda_2 = 0.4$	0.490	0.487	0.491	0.501	0.482	0.486
p_0	$\beta_{30} = 0.23$	0.228	0.228	0.226	0.226	0.226	0.226
p_1	$\beta_{40} = 0.07$	0.068	0.070	0.070	0.070	0.068	0.070
Med tiempo(Seg)		75.295	128.28	271.825	74.295	162.35	367.545
Med num. iter		30	29	29	29	29	29

Cuadro 6: Mediana de los parámetros estimados en el modelo ZOIP mixto para tres diferentes números de puntos de la cuadratura de Gauss-Hermite y con la estrategia de “con y sin *pruning*” y para todos los tamaños de muestra.

Estudio de simulación: Modelo regresión ZOIP mixto

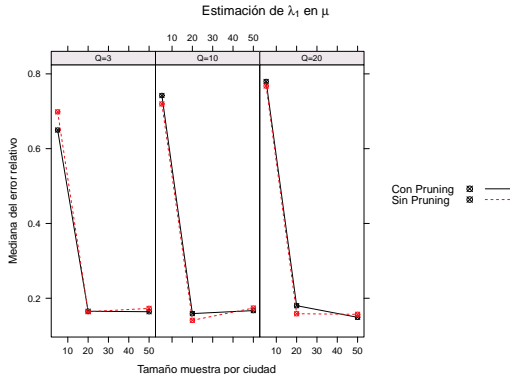


Figura 9: Mediana del error relativo para la estimación del parámetro λ_1 desviación estándar del intercepto aleatorio asociado a la μ , variando el tamaño de muestra, el número de puntos de la cuadratura y la utilización de *pruning*.

Estudio de simulación: Modelo regresión ZOIP mixto

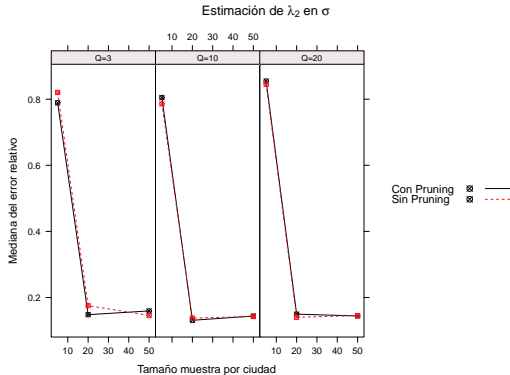


Figura 10: Mediana del error relativo para la estimación del parámetro λ_2 desviación estándar del intercepto aleatorio asociado a σ , variando el tamaño de muestra, el número de puntos de la cuadratura y la utilización de *pruning*.

Estudio de simulación: Modelo regresión ZOIP mixto

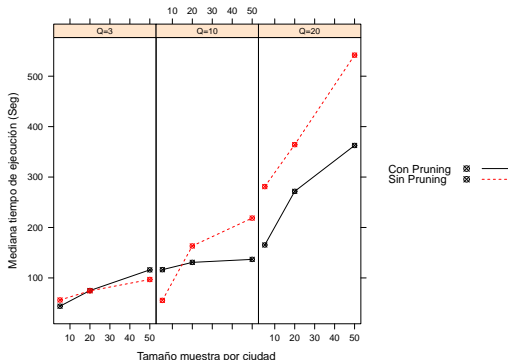


Figura 11: Mediana del tiempo de ejecución del modelo de regresión ZOIP mixto, bajo la función de `RMM.ZOIP` del paquete **ZOIP** de R.

Estudio de simulación: Conclusiones

1. Los factores que más influyen sobre la estimación es el tamaño de muestra de cada uno de los grupos, es decir n_i , ya que este factor hace que el error relativo de la estimación de todos los parámetros se vea reducido considerablemente cuando se aumenta.
2. El hecho de utilizar la metodología *pruning* hace que los valores de las estimaciones de los parámetros del modelo no cambien, pero sí que el tiempo de ejecución se vea reducido en un 50 %.
3. El efecto del número de puntos de la cuadratura de Gauss-Hermite no influye demasiado en la estimación de los parámetros de efectos fijos y aleatorios.

Conclusiones

1. Los modelos de regresión y distribución ZOIP propuestos son muy flexibles, ya que permite obtener diferentes opciones de modelado en una sola función.
2. Los modelos de regresión y distribución ZOIP permiten ajustarse a datos que se encuentran inflados solo con ceros o solo con unos, incluso a datos proporcionales no inflados.
3. Sobre la aplicación a datos reales se nota la utilidad, facilidad y convergencia de los modelos propuestos para explicar una problemática real de una compañía.
4. El paquete **ZOIP** se implementan diferentes funciones, con el objetivo de que sean utilizados los modelos de regresión ZOIP por un usuario. Dicho paquete muestra convergencia de sus funciones desarrollados, por medio de los estudios de simulación.

Resultados

- ✓ Publicación del paquete **ZOIP** en el CRAN de R.
- ✓ Artículo científico en una revista nacional indexada a Colciencias, Universidad Santo Tomas.
- ✓ Presentación de la distribución ZOIP en el simposio internacional de estadística 2017.
- ✓ Presentación del paquete **ZOIP** en el LatinR 2018, a realizarse el próximo mes de septiembre en Buenos Aires - Argentina.

Trabajos futuros

- Incluir un análisis de selección de variables en los modelos de regresión ZOIP fijos y mixtos.
- Incluir interceptos aleatorios en los parámetros asociados a la inflación de la distribución ZOIP, en el modelo de regresión ZOIP mixto propuesto en este trabajo.
- Incluir pendientes aleatorias en los diferentes parámetros de la regresión ZOIP.
- Incluir más funciones de método S3 en el paquete **ZOIP**, tales como **plot**, **predict** AIC, BIC, entre otras.
- Realizar mejoras continuas del paquete **ZOIP**, a medida que se vayan adquiriendo retroalimentaciones de los usuarios que utilizan el paquete.
- Extender los estudios de simulación del modelo de regresión ZOIP mixto, cuando se tienen muestras no balanceadas y demás parametrizaciones de la distribución ZOIP-beta y la distribución ZOIP-simplex.

Referencias I



S. Ferrari and F. Cribari-Neto.

Beta regression for modelling rates and proportions.

Journal of applied statistics, 31(7):799–815, 2004.



D. Galvis and V. Lachos.

Bayesian analysis of regression models for proportional data in the presence of zeros and ones.

PhD thesis, Estatal University of Campinas, 2014.



B. Jørgensen.

The theory of dispersion models.

Computational statistics and Data analysis, 76, 1997.



B. Jørgensen and Barndorff-Nielsen.

Some parametric models on the simplex.

Journal of multivariate analysis, 39(1):106–116, 1991.

Referencias II



R. Ospina and S. Ferrari.

Inflated beta distributions online: doi:10.1007/s00362-008-0125-4.
Statistical papers, 2010.



R. Ospina and S. Ferrari.

A general class of zero-or-one inflated beta regression models.
Computational Statistics and Data Analysis, 56(6):1609–1623, 2012.



Z. Qiu, P. Song, and M. Tan.

Simplex mixed-effects models for longitudinal proportional data.
Scandinavian Journal of Statistics, 35(4):577–596, 2008.



R Core Team.

R: A Language and Environment for Statistical Computing.
R Foundation for Statistical Computing, Vienna, Austria, 2018.

Referencias III



B. Rigby and M. Stasinopoulos.

Generalized additive models for location, scale and shape.

Applied Statistical, 54(3):507–554, 2005.



B. Rigby and M. Stasinopoulos.

Instructions on how to use the gamlss package in r.

Computational statistics and Data analysis, 2:194–195, 2008.



O. Usuga and V. Giampaoli.

Modelos de regressão beta com efeitos aleatórios normais e não normais para dados longitudinais.

PhD thesis, Universidade de São Paulo, 2013.



A. Zeileis, F. Cribari-Neto, and B. Grün.

Beta regression in r.

Journal of statistical software, 34(2):1–24, 2010.

Referencias IV



P. Zhang, Z. Qiu, and C. Shi.

simplexreg: An r package for regression analysis of proportional data using the simplex distribution.

Journal of Statistical Software, 71(11), 2016.

Muchas Gracias...

Distribución beta original.

La función de densidad de probabilidad de la distribución beta con parámetros p y q es denotada por $Be(p, q)$ y es dada por:

$$f(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1} \quad ; \quad 0 < y < 1 \quad , \quad (3)$$

donde los parámetros $p > 0$, $q > 0$ y $\Gamma(\cdot)$ es la función gamma. El valor esperado y la varianza de y están dadas por:

$$E(y) = \frac{p}{p+q} \quad (4)$$

$$Var(y) = \frac{pq}{(p+q)^2(p+q+1)} \quad (5)$$

Volver

Distribución beta parametrización Ferrari y Cribari-Neto(2004).

La función de densidad de probabilidad de la distribución beta con parámetros μ y ϕ denotado por $Be(\mu, \phi)$ está dada por:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1} \quad ; \quad 0 < y < 1 \quad (6)$$

donde $0 < \mu < 1$ y $\phi > 0$. El valor esperado y la varianza de y están dados por:

$$E(y) = \mu \quad (7)$$

$$Var(y) = \frac{\mu(1-\mu)}{1+\phi} \quad (8)$$

Distribución beta parametrización Rigby y Stasinopoulos(2008).

La función de densidad de probabilidad de la distribución beta con parámetros μ y σ denotado como $Be(\mu, \sigma)$ y está dada por:

$$f(y; \mu, \sigma) = B(\mu, \sigma) y^{\mu((1-\sigma^2)/\sigma^2)-1} (1-y)^{(1-\mu)((1-\sigma^2)/\sigma^2)-1} \quad ; 0 < y < 1, \quad (9)$$

$$\text{donde } B(\mu, \sigma) = \frac{\Gamma((1-\sigma^2)/\sigma^2)}{\Gamma(\mu((1-\sigma^2)/\sigma^2))\Gamma((1-\mu)((1-\sigma^2)/\sigma^2))},$$

donde $0 < \mu < 1$ y $0 < \sigma < 1$. La media y la varianza de y están dadas por:

$$E(y) = \mu \quad (10)$$

$$Var(y) = \sigma^2 \mu(1 - \mu) \quad (11)$$

Distribución simplex.

la distribución simplex es introducida por Barndorff-Nielsen y Jørgensen(1991) es un caso particular de los modelos de dispersión, cuya función de densidad de probabilidad depende de los parámetros μ y Σ^2 denotado por $S(\mu, \sigma^2)$ dado por:

$$f(y; \mu, \sigma^2) = \{2\pi\sigma^2[y(1-y)]^3\}^{-1/2} \exp\left\{-\frac{y(1-y)\mu^2(1-\mu)^2}{2\sigma^2(y-\mu)^2}\right\} \quad (12)$$

donde $0 < \mu < 1$ y $\sigma > 0$.

Distribución simplex.

Además el valor esperado y la varianza están dadas por:

$$E(y) = \mu \quad (13)$$

$$Var(y) = \mu(1 - \mu) - \frac{1}{\sqrt{2\sigma^2}} \exp \left\{ \frac{1}{2\sigma^2\mu^2(1 - \mu)^2} \right\} \Gamma \left\{ \frac{1}{2}, \frac{1}{2\sigma^2\mu^2(1 - \mu)^2} \right\} \quad (14)$$

donde $\Gamma(a, b)$ está dado por la función Γ incompleta definido como $\Gamma(a, b) = \int_b^\infty t^{a-1} b^t dt$. ver más en [13].

Volver

Cuadratura de Gauss-Hermite adaptativa.

Se define Como:

$$\int_{\mathbb{R}} g(x) dx \approx \sum_{i=1}^n g(p_i) \exp(p_i^2) w_i.$$

El conjunto de los n puntos de la cuadratura $\mathbf{P} = \{p_1, p_2, \dots, p_n\}$ corresponde a las raíces del polinomio de Hermite dado por:

$$H_n(x) = (-1)^n e^{-x^2} \frac{d^n}{dx^n} e^{-x^2},$$

con pesos asociados $\mathbf{W} = \{w_1, w_2, \dots, w_n\}$ dados por

$$w_i = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(x_i)]^2}.$$

Cuadratura de Gauss-Hermite adaptativa.

La cuadratura de Gauss-Hermite adaptativa (AGHQ) es propuesta por Liu y Pierce (1994); Pinheiro y Bates (1995), es básicamente una transformación de los puntos asociados a la cuadratura, centrando y extendiendo alrededor del máximo valor de \hat{x} de la función $\log(g(x))$. La transformación de los puntos de la cuadratura p_i definido como p_i^* , está dado por $p_i^* = \sqrt{2\hat{\sigma}}p_i + \hat{x}$ donde:

$$\hat{\sigma}^2 = \left[-\frac{d^2}{dx^2} \log(g(x)) \Big|_{x=\hat{x}} \right]^{-1}.$$

Así, la aproximación de la integral de $g(x)$ sobre \Re está dado por:

$$\int_{\Re} g(x) dx \approx \sqrt{2\hat{\sigma}} \sum_{i=1}^n g(p_i^*) \exp(p_i^2) w_i.$$

Cuadratura de Gauss-Hermite adaptativa.

Si extendemos la AGHQ a una integral de dimensión q de la función $g(x)$ sobre \mathbb{R}^q , en este caso, con una cuadratura de n puntos, \mathbf{Z} está basado en el producto cartesiano de \mathbf{P} , y los pesos de la cuadratura de \mathbf{A} está basado similarmente en el producto Kronecker, denotado por \otimes , los pesos originales \mathbf{W} , es dado:

$$\mathbf{Z} = \underbrace{P \times \dots \times P}_{q \text{ veces}} = P^q,$$

$$\mathbf{A} = \underbrace{W \otimes \dots \otimes W}_{q \text{ veces}}.$$

Así, la expresión para la integral aproximada de $g(x)$ sobre \mathbb{R}^q está dado por:

$$\int_{\mathbb{R}^q} g(x) dx \approx |\hat{Q}|^{1/2} 2^{q/2} \sum_{i=1}^{n^q} g(z_i^*) \exp(z_i^\top z_i) a_i,$$

Cuadratura de Gauss-Hermite adaptativa.

donde z_i y a_i corresponde a los elementos de \mathbf{Z} y \mathbf{A} , respectivamente. Los nuevos puntos de la cuadratura z_i^* están centrados en el máximo de \hat{x} del $\log(g(x))$ y está dado por $z_i^* = \hat{x} + \sqrt{2}\hat{Q}^{1/2}z_i$, donde $\hat{Q}^{1/2}$ corresponde a la descomposición de Cholesky de la curvatura de la matriz \hat{Q} , que se encuentra dado por:

$$\hat{Q} = \left[-\frac{d^2}{dx^2} \log(g(x)) \Big|_{x=\hat{x}} \right]^{-1}.$$

Cuadratura de Gauss-Hermite adaptativa con Pruning.

La cuadratura de Gauss-Hermite adaptativa con pruning consiste en eliminar puntos de la cuadratura, tales que el peso a_i asociado al punto es menor que un valor de referencia dado por θ , que está dado por:

$$\theta = \frac{w_{[1]} w_{[\frac{n+1}{2}]}}{n^{q-1}}.$$

Volver