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 de Nacional de Colombia
Sede-Medellín

Escuela de Estadística

Contenido

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 ceros y/o unos, son comunes en estudios clínicos, financieros, económicos y sociales.

2. Los modelos de regresión mixto para datos proporcionales inflados con ceros 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 ceros y/o unos, son comunes en estudios clínicos, financieros, económicos y sociales.

2. Los modelos de regresión mixto para datos proporcionales inflados con ceros 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 un modelo de regresión mixto para datos proporcionales inflados en cero y/o uno usando máxima verosimilitud.

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 dado por:

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

Distribuciones

Distribución ZOIP (Zeros Ones Inflated Proporcional)

La media y varianza de Y, están dadas por:

$$E(y) = p_1 + (1 - p_0 - p_1)E^*(y) \tag{2}$$

$$\begin{aligned} Var(y) = {} & p_1(1 - p1) + \\ & + (1 - p_0 - p_1)\left[Var^*(y) + (p_0 + p_1)[E^*(y)]^2 - 2E^*(y)p_1\right] \end{aligned} \tag{3}$$

Modelo de regresión ZOIP con efectos fijos

Sea $y_1, y_2, \ldots, y_n$ variables aleatorias independientes tal que cada y_i, para $i = 1, \ldots, n$, tiene función de densidad de probabilidad dado por la distribución ZOIPcon 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:

$$\begin{aligned} h_1(\mu_i) &= \mathbf{x}_{i1}^\top \boldsymbol{\beta}_1, \\ h_2(\sigma_i) &= \mathbf{x}_{i2}^\top \boldsymbol{\beta}_2, \\ h_3(p_{0i}) &= \mathbf{x}_{i3}^\top \boldsymbol{\beta}_3, \\ h_4(p_{1i}) &= \mathbf{x}_{i4}^\top \boldsymbol{\beta}_4 \end{aligned} \tag{4}$$

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 de observaciones independientes, es de la forma:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} g(\mathbf{y}_i; \mu_i, \sigma_i, p_{0i}, p_{1i}) = \prod_{i=1}^{n} f_0(\mathbf{y}_i; p_{0i}) \prod_{i=1}^{n} f_1(\mathbf{y}_i; p_{1i}) \prod_{i=1}^{n} f_2(\mathbf{y}_i; \mu_i, \sigma_i) \tag{5}$$

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

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

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:

$$\begin{aligned} y_{ij}|\gamma_{1i},\gamma_{2i} &\overset{\text{ind}}{\sim} ZOIP(\mu_{ij},\sigma_{ij},p_{0ij},p_{1ij}), \\ h_1(\mu_{ij}) &= \mathbf{x}_{ij1}^{\top}\boldsymbol{\beta}_1+\gamma_{1i}, \\ h_2(\sigma_{ij}) &= \mathbf{x}_{ij2}^{\top}\boldsymbol{\beta}_2+\gamma_{2i}, \\ h_3(p_{0ij}) &= \mathbf{x}_{ij3}^{\top}\boldsymbol{\beta}_3, \\ h_4(p_{1ij}) &= \mathbf{x}_{ij4}^{\top}\boldsymbol{\beta}_4, \\ \gamma_{1i} &\overset{\text{i.i.d}}{\sim} N(0,\lambda_1^2), \\ \gamma_{2i} &\overset{\text{i.i.d}}{\sim} N(0,\lambda_2^2), \end{aligned} \tag{6}$$

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

Estimación vía máxima verosimilitud y AGHQ

El vector de parámetros para el modelo (6) 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 funcion de log-verosimilitud $\ell(\boldsymbol{\theta})$ esta 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.

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

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

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

Ajuste distribución ZOIP

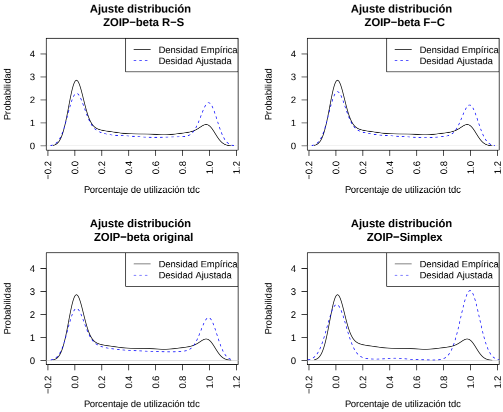


Figura: Ajuste de diferentes distribuciones y parametrizaciones ZOIP al porcentaje de utilización de una tdc.

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.

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

$$\begin{aligned} h_1(\mu_i) &= \beta_{10} + \beta_{11}x_{1i} + \beta_{12}x_{2i} + \beta_{13}x_{3i}, \\ h_2(\sigma_i) &= \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i} + \beta_{23}x_{3i}, \\ h_3(p_{0i}) &= \beta_{30} + \beta_{31}x_{1i} + \beta_{32}x_{2i} + \beta_{33}x_{3i}, \\ h_4(p_{1i}) &= \beta_{40} + \beta_{41}x_{1i} + \beta_{42}x_{2i} + \beta_{43}x_{3i}, \end{aligned}$$

y_i es el porcentaje de utilizacion 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.

$$
\begin{aligned}
y_i &\sim ZOIP(\mu_i, \sigma_i, p_{0i}, p_{1i}), \quad i = 1, 2, \ldots, 9206 \\
h_1(\mu_i) &= \beta_{10} + \beta_{11}x_{1i} + \beta_{12}x_{2i} + \beta_{13}x_{3i}, \\
h_2(\sigma_i) &= \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i} + \beta_{23}x_{3i}, \\
h_3(p_{0i}) &= \beta_{30} + \beta_{31}x_{1i} + \beta_{32}x_{2i} + \beta_{33}x_{3i}, \\
h_4(p_{1i}) &= \beta_{40} + \beta_{41}x_{1i} + \beta_{42}x_{2i} + \beta_{43}x_{3i},
\end{aligned}
$$

y_i es el porcentaje de utilizacion 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.

$$
\begin{aligned}
y_i &\sim ZOIP(\mu_i, \sigma_i, p_{0i}, p_{1i}), \quad i = 1, 2, \ldots, 9206 \\
h_1(\mu_i) &= \beta_{10} + \beta_{11}x_{1i} + \beta_{12}x_{2i} + \beta_{13}x_{3i}, \\
h_2(\sigma_i) &= \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i} + \beta_{23}x_{3i}, \\
h_3(p_{0i}) &= \beta_{30} + \beta_{31}x_{1i} + \beta_{32}x_{2i} + \beta_{33}x_{3i}, \\
h_4(p_{1i}) &= \beta_{40} + \beta_{41}x_{1i} + \beta_{42}x_{2i} + \beta_{43}x_{3i},
\end{aligned}
$$

y_i es el porcentaje de utilizacion 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.

$$
\begin{aligned}
y_i \sim ZOIP(\mu_i, \sigma_i, p_{0i}, p_{1i}), \quad i = 1, 2, \ldots, 9206 \\
h_1(\mu_i) &= \beta_{10} + \beta_{11}x_{1i} + \beta_{12}x_{2i} + \beta_{13}x_{3i}, \\
h_2(\sigma_i) &= \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{2i} + \beta_{23}x_{3i}, \\
h_3(p_{0i}) &= \beta_{30} + \beta_{31}x_{1i} + \beta_{32}x_{2i} + \beta_{33}x_{3i}, \\
h_4(p_{1i}) &= \beta_{40} + \beta_{41}x_{1i} + \beta_{42}x_{2i} + \beta_{43}x_{3i},
\end{aligned}
$$

y_i es el porcentaje de utilizacion 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.

Aplicación modelo de regresión ZOIP

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	
	σ	$\hat{\beta}_{20}$	0.822	0.038	$< 2{,}2e^{-16}$	
		$\hat{\beta}_{21}$	-0.197	0.078	0.0114	
		$\hat{\beta}_{22}$	-0.006	0.002	0.0013	
		$\hat{\beta}_{23}$	-0.003	0.007	0.6741	
	p_0	$\hat{\beta}_{30}$	-1.496	0.101	$< 2{,}2e^{-16}$	
		$\hat{\beta}_{31}$	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	
	p_1	$\hat{\beta}_{40}$	-1.480	0.095	$< 2{,}2e^{-16}$	
		$\hat{\beta}_{41}$	-0.630	0.254	0.0132	
		$\hat{\beta}_{42}$	0.011	0.006	0.0666	
		$\hat{\beta}_{43}$	-0.069	0.022	0.0022	

Cuadro: Parámetros regresores estimados de un modelo de regresión ZOIP-beta con parametrización Rigby y Stasinopoulos (2005) en el porcentaje de utilización de una tdc. Con un criterio de convergencia de $1e^{-10}$

Aplicación modelo de regresión ZOIP mixto

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.

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

con $\gamma_{1i} \sim N(0,\lambda_1^2)$ y $\gamma_{2i} \sim N(0,\lambda_2^2)$.
y_{ij} es el porcentaje de utilización de la j-ésima tdc perteneciente a la i-ésima ciudad, $i = 1,2,\ldots,10$ y $j = 1,2,\ldots,15$.
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: Estimación de los efectos fijos y los componentes de varianza λ_1 y $\lambda 2$ 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:

$$
\begin{aligned}
y_{ij}|\gamma_{1i},\gamma_{2i} &\overset{\text{ind}}{\sim} ZOIP(\mu_{ij},\sigma_{ij},p_0,p_1),\\
h_1(\mu_{ij}) &= -1{,}13+\gamma_{1i}+0{,}33x_{1ij},\\
h_2(\sigma_{ij}) &= 0{,}33+\gamma_{2i}+0{,}14x_{1ij},\\
h_3(p_0) &= 0{,}23,\\
h_4(p_1) &= 0{,}07,
\end{aligned}
$$

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 [4].

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 [4].

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 [4].

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

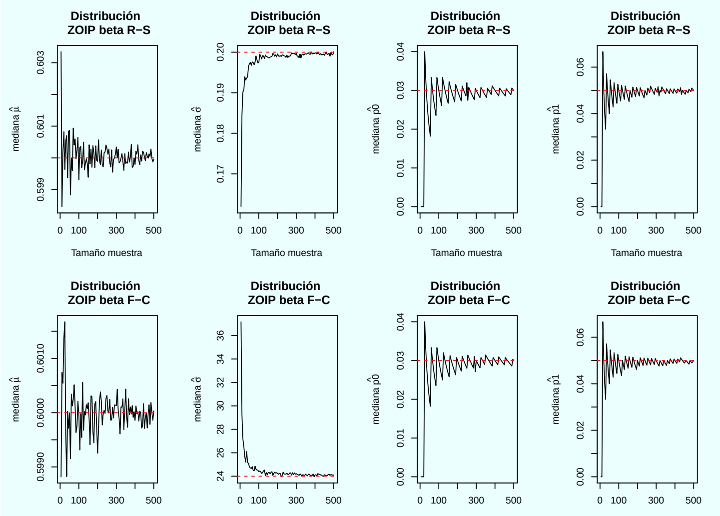
Escenario 1: Distribución ZOIP-beta($\mu = 0{,}6, \sigma = 0{,}2, p_0 = 0{,}03, p_1 = 0{,}05$) para el caso de la parametrización de [4].

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].

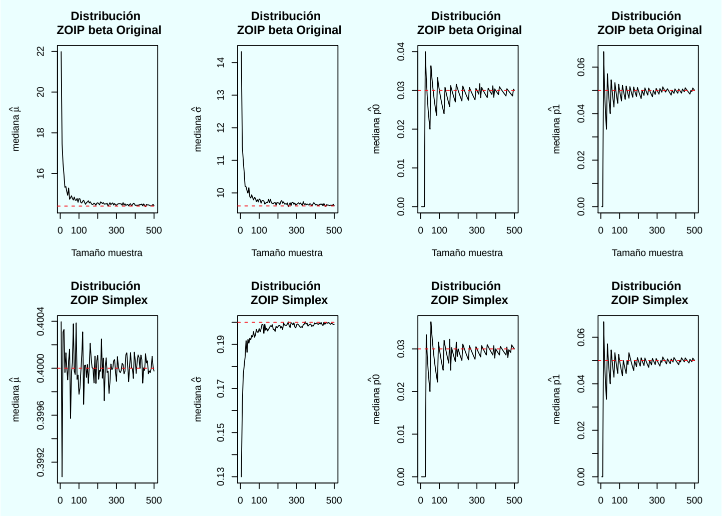
Escenario 3: Distribución ZOIP-beta($\mu = 14{,}4, \sigma = 9{,}6, p_0 = 0{,}03, p_1 = 0{,}05$) en la parametrización original.

Escenario 4: Distribución ZOIP-simplex($\mu = 0{,}4, \sigma = 0{,}2, p_0 = 0{,}03, p_1 = 0{,}05$).

Estudio de simulación: Distribución ZOIP



Estudio de simulación: Distribución ZOIP



Estudio de simulación: Distribución ZOIP

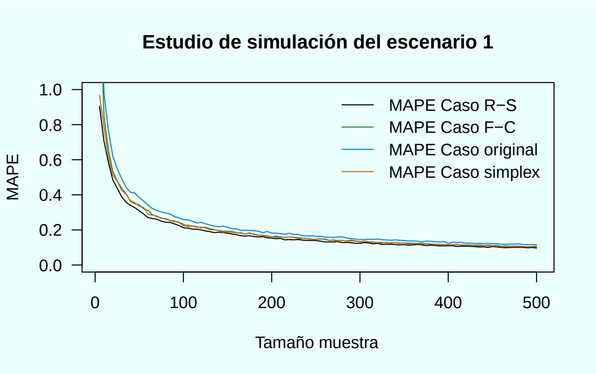


Figura: Mape (Error porcentual absoluto medio) para los escenarios de simulación y para distintos valores de n.

Referencias

S. Ferrari and F. Cribari-Neto.
Beta regression for modelling rates and proportions.
Journal of applied statistics, 31(7):799–815, 2004.

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.

B. Rigby and M. Stasinopoulos.
Generalized additive models for location, scale and shape.
Applied Statistical, 54(3):507–554, 2005.

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:

$$f(y;p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1} \quad ; \quad 0 < y < 1 \quad , \tag{7}$$

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:

$$E(y) = \frac{p}{p+q} \tag{8}$$

$$Var(y) = \frac{pq}{(p+q)^2(p+q+1)} \tag{9}$$

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:

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \quad ; \quad 0 < y < 1 \tag{10}$$

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

$$E(y) = \mu \tag{11}$$

$$Var(y) = \frac{\mu(1-\mu)}{1+\phi} \tag{12}$$

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:

$$f(y;\mu,\sigma) = B(\mu,\sigma) y^{\mu((1-\sigma^2)/\sigma^2)-1}(1-y)^{(1-\mu)((1-\sigma^2)/\sigma^2)-1} \quad ; 0 < y < 1 \quad , \tag{13}$$

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:

$$E(y) = \mu \tag{14}$$

$$Var(y) = \sigma^2\mu(1-\mu) \tag{15}$$

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:

$$f(y;\mu,\sigma^2) = \left\{2\pi\sigma^2[y(1-y)]^3\right\}^{-1/2} \exp\left\{-\frac{y(1-y)\mu^2(1-\mu)^2}{2\sigma^2(y-\mu)^2}\right\} \qquad (16)$$

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

Distribución simplex.

Además el valor esperado y la varianza están dadas por:

$$E(y) = \mu \tag{17}$$

$$Var(y) = \mu(1-\mu) - \frac{1}{\sqrt{2\sigma^2}} \exp\left\{\frac{1}{2\sigma^2\mu^2(1-\mu)^2}\right\} \Gamma\left\{\frac{1}{2}, \frac{1}{2\sigma^2\mu^2(1-\mu)^2}\right\} \tag{18}$$

donde $\Gamma(a,b)$ está dado por la función$-\Gamma$ incompleta definido como $\Gamma(a,b) = \int_b^{\infty} t^{a-1} b^t dt$. ver más en [5].

Volver

Cuadratura de Gauss-Hermite adaptativa.

Se define Como:

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

El conjunto de los n puntos de la cuadratura $\mathbf{P} = \{p_1, p_2, \ldots, p_n\}$ 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 $\mathbf{W} = \{w_1, w_2, \ldots, w_n\}$ 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 $\Re^q$, 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:

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

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

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

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

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^* estan 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