

Un modelo lineal mixto adaptado a una cadena de Markov con espacio de estados ordinal. Aplicación a datos sobre promedios académicos de estudiantes

A Linear Mixed Model Adapted to a Markov Chain with Ordinal State Space. Application to Data about Grade Point Average (GPA)

ROGER JESÚS TOVAR^{1,a}, JUAN CARLOS SALAZAR^{2,b}

¹DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA, FACULTAD DE CIENCIAS BÁSICAS E INGENIERÍAS, UNIVERSIDAD DE CÓRDOBA, MONTERÍA, COLOMBIA

²ESCUELA DE ESTADÍSTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE COLOMBIA, MEDELLÍN, COLOMBIA

Resumen

En este artículo se estudia una metodología para estimar los efectos de las covariables usando un modelo lineal mixto con intercepto aleatorio y respuesta policótoma categórica ordinal, bajo distintas especificaciones distribucionales de dicho efecto aleatorio. Esta metodología constituye una extensión de la propuesta hecha por Salazar et al. (2007), en la medida que en este último trabajo se presentan resultados obtenidos con un modelo donde la respuesta es nominal.

Específicamente, se considera una cadena de Markov de $k + 2$ estados con dos estados absorbentes que compiten entre sí y k estados transitorios. Con este modelo se obtiene la función de verosimilitud de los datos. Luego, por medio de un estudio de simulación se evalúa el efecto sobre las estimaciones bajo distintas formas distribucionales para el efecto aleatorio. La maximización de la función de verosimilitud se lleva a cabo numéricamente utilizando el método de la cuadratura de Gauss en asocio con el algoritmo de Newton-Raphson. Finalmente, se ilustra la metodología usando datos sobre los promedios acumulados de estudiantes de la Universidad Nacional de Colombia, sede Medellín, recolectados entre 2005 y 2007.

Palabras clave: datos longitudinales, cadenas de Markov, regresión logística, cuadratura de Gauss.

Abstract

^aProfesor asistente. E-mail: rjesust@gmail.com

^bProfesor asociado. E-mail: jcsalaza@unalmed.edu.co

In this paper we study methodology to estimate the effects of covariates using a linear mixed model with random intercept and polytomous ordinal categorical response, under different distributional specifications of this random intercept. This methodology represents an extension of the one proposed in Salazar et al. (2007), where it is presented results obtained using a model where the response is treated in a nominal scale. Specifically, it is considered a Markov chain with $k + 2$ states with two absorbing and k transient states. The likelihood function for the data is derived. Under this model and using a simulation study we assess the effects on the estimates under different distributional specifications for the random intercept. The likelihood function is maximized using the Gauss quadrature method in conjunction with the Newton-Raphson algorithm. Finally, we illustrated the methodology using data about the Grade Point Average (GPA) from students of the Universidad Nacional de Colombia, at Medellín, collected from 2005 to 2007.

Key words: Longitudinal data, Markov chains, Logistic regression, Gaussian quadrature.

1. Introducción

Los datos longitudinales son muy comunes en investigación y usualmente surgen de medir una o varias respuestas, junto con características propias de cada sujeto o unidad experimental repetidamente a través del tiempo. Algunas de estas respuestas pueden pensarse como realizaciones aleatorias de un cierto proceso estocástico con espacio de estados discreto. Es de interés para el investigador estimar el efecto de las covariables sobre estas respuestas y hacerlo teniendo en cuenta las posibles fuentes de variabilidad. Para modelar esta variabilidad, la clase de modelos lineales mixtos ha demostrado ser de gran utilidad; Hedeker & Gibbons (1994), Aitkin & Alfó (1998), Hedeker et al. (2000), Ten Have et al. (2003), Gao (2004), Salazar et al. (2007), entre otros, muestran cómo estos modelos se pueden usar para modelar respuestas categóricas longitudinales en escala nominal.

En este artículo se propone un modelo de regresión logística policótoma con efectos aleatorios para respuesta categórica ordinal, considerando un modelo de estados múltiples de Markov. Por medio de un estudio de simulación se evalúan las bondades del método. Estas simulaciones se realizaron con base en un modelo de k estados transitorios que se comunican entre sí y dos estados absorbentes que compiten entre ellos (estados $k + 1$ y $k + 2$). El ejemplo de aplicación se basa en un modelo similar con $k = 3$ estados transitorios y dos estados absorbentes.

Este modelo constituye una extensión de la propuesta hecha por Salazar et al. (2007), cuando se considera que el espacio de estados es de tipo ordinal. Este enfoque permite establecer una conexión entre procesos estocásticos y modelos lineales y por lo tanto constituye uno de los aportes más importantes de este artículo.

Para estimar los efectos de las covariables incluidas en el modelo se usa la función de verosimilitud, la cual tiene en cuenta información que ha sido recolectada de manera longitudinal y la naturaleza ordinal de la respuesta. La maximización de la

función de verosimilitud permite evaluar el efecto de las covariables en las diferentes transiciones; esto se puede lograr por medio de un modelo de regresión logística policótoma con efectos aleatorios. Para evaluar la integral que aparece dentro de la función de verosimilitud se utiliza el método de la cuadratura de Gauss conjuntamente con el algoritmo de Newton-Raphson (Abramowitz & Stegun 1972); esta combinación ha mostrado buenos resultados en estudios anteriores (Salazar 2004). Se decidió usar la cuadratura de Gauss para evaluar la integral de la función de verosimilitud ya que este método se recomienda usualmente en la literatura para evaluar integrales donde el integrando tiene formas complicadas (Skrondal & Rabeth-Hesketh 2004).

Este artículo está organizado de la siguiente forma: en la sección 2 se presenta una expresión para el cálculo de las probabilidades de transición; en la sección 3 se describe el modelo estadístico propuesto; en la sección 4 se construye la función de verosimilitud; la sección 5 contiene una descripción y los resultados de un estudio de simulación; la sección 6 incluye la ilustración del método usando datos reales; y finalmente, en la sección 7 se discuten los méritos del método.

2. Probabilidades de transición en el modelo de estados múltiples

Considere una cadena de Markov $\{Y(t) : t \in \{1, \dots, T\}\}$. Se asume que el espacio de estados para esta cadena tiene $k + 2$ estados, donde los primeros k se comunican entre sí y los estados $k + 1$ y $k + 2$ son estados competitivos absorbentes. Sea $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{iT_i})$, el vector que representa los estados observados por el i -ésimo sujeto en T_i ocasiones distintas. Suponga que para un individuo particular se observa $y_i = (y_{i1}, y_{i2}, \dots, y_{iT_i})$, y denote por $f(y_i)$ la función de distribución para el vector aleatorio y_i , entonces,

$$\begin{aligned} f(y_i) &= f(y_{i1}, y_{i2}, \dots, y_{iT_i}) \\ &= f(y_{i1})f(y_{i2} | y_{i1})f(y_{i3} | y_{i1}, y_{i2}) \cdots f(y_{iT_i} | y_{i1}, y_{i2}, \dots, y_{iT_i-1}) \end{aligned}$$

Como no es posible especificar $f(y_{i1})$, se condiciona respecto a la probabilidad inicial $f(y_{i1})$ (Diggle et al. 2001). Así, la distribución conjunta del vector aleatorio y_i , condicionada en y_{i1} puede ser escrita como:

$$f(y_i | y_{i1}) = f(y_{i2} | y_{i1})f(y_{i3} | y_{i1}, y_{i2}) \cdots f(y_{iT_i} | y_{i1}, y_{i2}, \dots, y_{iT_i-1})$$

si se asume que se satisface la propiedad de Markov, se tiene que,

$$f(y_i | y_{i1}) = f(y_{i2} | y_{i1})f(y_{i3} | y_{i2}) \cdots f(y_{iT_i} | y_{iT_i-1})$$

Esta simplificación permite reducir la complejidad en la función de verosimilitud y facilita el proceso de evaluación y optimización.

Se denota la probabilidad de transición del estado a al estado b para el i -ésimo sujeto como:

$$\begin{aligned} p_{ab}^{(i)} &= f(b | a) \\ &= P(Y_i(l) = b | Y_i(l-1) = a) \end{aligned}$$

donde a se refiere al estado ocupado por el i -ésimo sujeto en el tiempo t_l . Por ejemplo $p_{ab}^{(i)}$ es la transición del estado a al estado b durante el intervalo de tiempo (t_{l-1}, t_l) ; así, si $a, b \in \{1, \dots, k+2\}$, se tiene que $p_{ab}^{(i)} = P(Y_i(l) = b | Y_i(l-1) = a)$.

Si se considera la naturaleza ordinal del espacio de estados, es de interés hallar las probabilidades acumuladas hasta la categoría b de la variable de respuesta. Sea $\gamma_{ab}^{(i)}$ la probabilidad acumulada hasta el estado b para el i -ésimo sujeto, dado que en la última ocasión el i -ésimo sujeto se encontraba en el estado a ; entonces, si $a, b \in \{1, 2, \dots, k+2\}$, se tiene que

$$\gamma_{ab}^{(i)} = \sum_{j=1}^b p_{aj}^{(i)} = \sum_{j=1}^b P(Y_i(l) = j | Y_i(l-1) = a) = P(Y_i(l) \leq b | Y_i(l-1) = a)$$

Ahora, sea $a \in \{1, \dots, k\}$ y $b = 1$, entonces,

$$\begin{aligned} p_{a1}^{(i)} &= P(Y_i(l) = 1 | Y_i(l-1) = a) = P(Y_i(l) \leq 1 | Y_i(l-1) = a) \\ &= \gamma_{a1}^{(i)} \end{aligned}$$

Sea $a \in \{1, \dots, k\}$ y $b \in \{2, \dots, k+2\}$, entonces,

$$\begin{aligned} p_{ab}^{(i)} &= P(Y_i(l) = b | Y_i(l-1) = a) \\ &= P(Y_i(l) \leq b | Y_i(l-1) = a) - P(Y_i(l) \leq b-1 | Y_i(l-1) = a) \\ &= \gamma_{ab}^{(i)} - \gamma_{ab-1}^{(i)} \end{aligned}$$

es decir,

$$p_{ab}^{(i)} = \begin{cases} \gamma_{ab}^{(i)}, & \text{si } b = 1 \text{ y } a = 1, \dots, k; \\ \gamma_{ab}^{(i)} - \gamma_{ab-1}^{(i)}, & \text{si } b = 2, \dots, k+2 \text{ y } a = 1, \dots, k \end{cases} \quad (1)$$

son las probabilidades de transición cuando se tiene en cuenta la naturaleza ordinal de los estados.

3. Modelo estadístico

Se asume que la trayectoria de respuestas de cada sujeto está afectada por efectos fijos (covariables de interés) y por un efecto aleatorio que es específico a cada participante. Así, las probabilidades de transición p_{ab} , descritas en la sección anterior, deben estar influenciadas por estos efectos fijos y aleatorios. Para conectar estas probabilidades con estos efectos fijos y aleatorios se asume un modelo de

regresión logística policótoma con efecto aleatorio y respuesta ordinal. La elección de este modelo está motivada por el hecho de que un modelo con efecto aleatorio explica la variabilidad extra que un modelo de solo efectos fijos no explica. Así, para $a \in \{1, 2, \dots, k\}$ y $b \in \{1, 2, \dots, k+2\}$, se propone un modelo estadístico de la forma:

$$\text{logit} \left[\gamma_{ab}^{(i)}(\theta \mid x_i, b_i) \right] = \alpha_b + \beta_{ab}^T x_i + b_i^T z_i \quad (2)$$

donde x_i es un vector $p \times 1$ de covariables para el sujeto i , b_i es un vector $r \times 1$ de efectos aleatorios, z_i es vector de diseño para los r efectos aleatorios y $\theta = (\alpha_b, \beta_{ab})$ es un vector de parámetros desconocidos.

La formulación de este modelo en términos de funciones logit permite encontrar una expresión cerrada para las probabilidades de transición. A partir de la ecuación (2) se puede demostrar que, para cualquier $a \in \{1, 2, \dots, k\}$, las probabilidades de transición, $p_{ab}^{(i)}(\theta \mid x_i, b_i)$, se pueden escribir como:

$$\begin{cases} \frac{1}{1+\exp[-(\alpha_b + \beta_{ab}^T x_i + b_i^T z_i)]}, & b = 1; \\ \frac{1}{1+\exp[-(\alpha_b + \beta_{ab}^T x_i + b_i^T z_i)]} - \frac{1}{1+\exp[-(\alpha_{b-1} + \beta_{ab-1}^T x_i + b_i^T z_i)]}, & b = 2, \dots, k+1; \\ \frac{1}{1+\exp(\alpha_{b-1} + \beta_{ab-1}^T x_i + b_i^T z_i)}, & b = k+2 \end{cases} \quad (3)$$

Si el estado a es absorbente, entonces

$$p_{ab}^{(i)} = \begin{cases} 1, & \text{si } b = a; \\ 0, & \text{si } b \neq a \end{cases} \quad (4)$$

Nótese que esta última probabilidad no depende del vector θ o de los efectos aleatorios.

Las probabilidades definidas en (3) y (4) satisfacen la condición: $\sum_{b=1}^{k+2} p_{ab} = 1$ para $a \in \{1, \dots, k+2\}$.

4. La función de verosimilitud

La estrategia de estimación de los parámetros del modelo se basa en la función de verosimilitud. La contribución del sujeto i a la verosimilitud puede expresarse por medio del siguiente elemento de probabilidad:

$$\prod_{l=1}^{T_i} p_{y_{l-1}y_l}^{(i)}(\theta \mid x_i), \quad i = 1, 2, \dots, m$$

donde T_i es el número de veces que la persona ha sido evaluada (número de visitas) y m es el número de sujetos bajo estudio.

Asumiendo que la variabilidad en las respuestas de cada participante se explica no solo por los efectos fijos, sino también por efectos aleatorios, y si cada función de

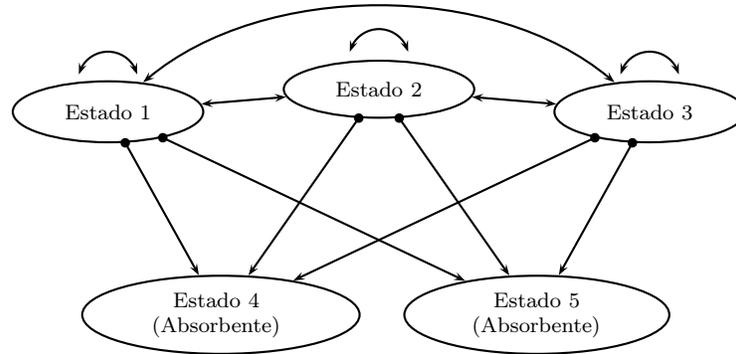


FIGURA 1: Modelo de Markov de 5 estados con estados 4 y 5 absorbentes y competitivos.

probabilidad de transición comparte un efecto aleatorio común b_i , la verosimilitud marginal para el i -ésimo sujeto es:

$$\int_{\Omega} \prod_{l=1}^{T_i} p_{y_{l-1}y_l}^{(i)}(\theta | x_i, b_i) dF(b_i)$$

donde Ω es el soporte para la distribución del vector aleatorio b_i y F su función de distribución acumulada (CDF). Así, la función de verosimilitud se expresa como:

$$L(\theta | x) = \prod_{i=1}^m \int_{\Omega} \prod_{l=1}^{T_i} p_{y_{l-1}y_l}^{(i)}(\theta | x_i, b_i) dF(b_i) \quad (5)$$

Aquí, $\theta = (\alpha, \beta)$ es el vector de parámetros desconocidos que representan los efectos del intercepto y las covariables respectivamente.

La maximización de esta función de verosimilitud produce las estimaciones para los efectos de las covariables. Sin embargo, la presencia de la integral dificulta su evaluación y es por esta razón que se requiere de un método numérico para su tratamiento. Para este fin, se optó por el método de la cuadratura de Gauss.

5. Estudio de simulación

Con el fin de explorar los alcances y limitaciones de la metodología, se lleva a cabo un estudio de simulación. Se considera un modelo de cinco estados con dos estados absorbentes competitivos y tres estados transitorios que se comunican entre sí. En la figura 1 se ilustra gráficamente la situación.

5.1. Metodología

Se asume que las probabilidades de transición siguen un modelo de regresión logística policótoma con un intercepto aleatorio y respuesta policótoma categó-

rica ordinal. En este modelo se incorporan dos covariables, una dicótoma y otra continua.

Se explora cómo la forma distribucional para el término del efecto aleatorio afecta la estimaciones de los efectos de las covariables por máxima verosimilitud, para esto se asumen distintas formas distribucionales del término aleatorio: normal(0,1), uniforme(0,1), Cauchy(0,1), logística(0,1) y doble exponencial(0,1).

Para cada una de estas distribuciones se generan 1000 muestras aleatorias con tamaños $m = 100, 200, 400$ y 800 , y número máximo de visitas $l = 3, 4$ y 5 . Las optimizaciones de las verosimilitudes se hacen con el algoritmo de Newton-Raphson en asociación con la cuadratura de Gauss. Los gradientes y hessianas también se obtienen numéricamente para estimar las varianzas de las estimaciones. Todas las simulaciones se realizan con el paquete SAS Institute Inc. (1989).

5.2. Resultados del estudio de simulación

En las tablas 1 a 3 se muestran las estimaciones promedio para los parámetros de referencia y sus respectivos errores estándar promedio (e.e.), asumiendo cinco formas distribucionales para el término aleatorio con un número máximo de visitas de 3, 4 y 5, respectivamente.

En la tabla 1, en la columna etiquetada Parámetro de referencia, los primeros cuatro valores corresponden a interceptos relacionados con estados distintos al llamado estado base que en este caso es el estado 5. En esta misma tabla se puede ver que, cuando el número de visitas es igual a 3, y la distribución para el efecto aleatorio es normal, los valores para el parámetro de referencia son sobrestimados en ambas covariables, dicótoma y continua; pero en la medida en que se aumenta el tamaño de la muestra, las estimaciones se aproximan más al parámetro de referencia.

Cuando la distribución para el efecto aleatorio se asume uniforme y se considera la covariable continua, se obtienen estimaciones más cercanas al parámetro de referencia que las obtenidas con la distribución normal, debido a que en general las estimaciones para la covariable dicótoma no presentan errores estándar tan altos. Con la distribución Cauchy, los valores de referencia son sobrestimados, y a medida que se aumenta el tamaño de la muestra, la discrepancia entre los valores estimados y los valores de referencia no parece disminuir significativamente; esto ocurre en ambas covariables. En particular, la distribución Cauchy presenta las estimaciones más alejadas respecto al parámetro de referencia. Cuando se asumen distribuciones para el efecto aleatorio logística y doble exponencial, las estimaciones de los valores de referencia para ambas covariables están por encima de los valores objetivo y, al aumentar el tamaño de la muestra, las estimaciones no mejoran considerablemente.

Resultados similares para número máximo de visitas igual a 4 se muestran en la tabla 2.

Aquí se observa que cuando el efecto aleatorio se asume normal, los valores de referencias se siguen sobrestimando y los errores estándar de las estimaciones siguen siendo muy altos; en el caso uniforme, las estimaciones se alejan más respecto a los parámetros de referencia en comparación con 3 visitas y los errores estándar

TABLA 1: Estimación de los parámetros de referencia (error estándar), número máximo de visitas = 3.

		Número máximo de visitas 3			
Distribución asumida	Parámetro de referencia	$n = 100$ Media (e.e.)	$n = 200$ Media (e.e.)	$n = 400$ Media (e.e.)	$n = 800$ Media (e.e.)
Normal	-0.809	-0.902(0.244)	-0.884(0.244)	-0.875(0.120)	-0.865(0.085)
	0.982	1.088(0.252)	1.088(0.251)	1.081(0.124)	1.084(0.088)
	1.825	2.131(0.295)	2.150(0.295)	2.117(0.145)	2.118(0.102)
	2.491	2.797(0.349)	2.814(0.350)	2.747(0.170)	2.747(0.119)
	Dicót. 0.117	0.135(0.049)	0.133(0.049)	0.129(0.024)	0.129(0.017)
	Cont. 0.785	0.847(0.450)	0.829(0.444)	0.832(0.219)	0.834(0.154)
Uniforme	-0.809	-0.801(0.208)	-0.777(0.145)	-0.773(0.102)	-0.771(0.072)
	0.982	1.157(0.190)	1.152(0.133)	1.151(0.094)	1.144(0.066)
	1.825	2.194(0.234)	2.179(0.164)	2.175(0.115)	2.165(0.081)
	2.491	2.872(0.293)	2.828(0.203)	2.814(0.142)	2.796(0.100)
	Dicót. 0.117	0.125(0.038)	0.122(0.026)	0.122(0.018)	0.122(0.013)
	Cont. 0.785	0.791(0.336)	0.782(0.233)	0.772(0.164)	0.780(0.115)
Cauchy	-0.809	-1.128(0.309)	-1.100(0.215)	-1.097(0.152)	-1.097(0.107)
	0.982	1.080(0.306)	1.098(0.214)	1.089(0.151)	1.082(0.106)
	1.825	2.274(0.357)	2.291(0.251)	2.268(0.176)	2.263(0.124)
	2.491	3.036(0.419)	3.029(0.293)	3.008(0.205)	2.994(0.145)
	Dicót. 0.117	0.149(0.062)	0.148(0.043)	0.148(0.030)	0.148(0.021)
	Cont. 0.785	0.992(0.568)	0.952(0.395)	0.955(0.277)	0.978(0.195)
Logística	-0.809	-0.994(0.322)	-0.969(0.226)	-0.966(0.159)	-0.962(0.112)
	0.982	0.873(0.323)	0.892(0.227)	0.886(0.160)	0.883(0.113)
	1.825	1.883(0.347)	1.862(0.244)	1.856(0.172)	1.853(0.121)
	2.491	2.418(0.374)	2.443(0.263)	2.432(0.185)	2.431(0.131)
	Dicót. 0.117	0.134(0.065)	0.129(0.045)	0.128(0.032)	0.128(0.022)
	Cont. 0.785	0.856(0.603)	0.815(0.419)	0.822(0.294)	0.830(0.208)
Laplace	-0.809	-0.947(0.274)	-0.935(0.192)	-0.917(0.135)	-0.923(0.095)
	0.982	0.999(0.275)	0.992(0.193)	1.009(0.136)	0.995(0.096)
	1.825	2.024(0.313)	1.995(0.220)	2.009(0.155)	1.999(0.109)
	2.491	2.627(0.358)	2.602(0.249)	2.612(0.176)	2.598(0.124)
	Dicót. 0.117	0.135(0.055)	0.129(0.038)	0.128(0.027)	0.128(0.019)
	Cont. 0.785	0.832(0.508)	0.863(0.351)	0.825(0.246)	0.831(0.174)

son altos. Cuando se asume distribución Cauchy para el término de efecto aleatorio, las estimaciones siguen siendo las más alejadas de los valores de referencia con desviaciones estándar promedio muy altas. Para la distribución logística, se obtienen estimaciones muy cercanas al valor de referencia en la covariable dicótoma, y en la covariable continua las estimaciones no muestran buenas aproximaciones. Finalmente, las estimaciones siguen sobrestimando los valores de referencia y los errores estándar siguen siendo altos cuando la distribución es doble exponencial.

La tabla 3 muestra los resultados de la simulación cuando se tienen 5 visitas. Aquí se puede ver que, con la distribución normal, las estimaciones sobrestiman los efectos de las covariables y para la covariable dicótoma son más cercanas al valor de referencia que las estimaciones con la covariable continua. Cuando la distribución del efecto aleatorio es uniforme, las estimaciones de los valores de referencia son muy estables en ambas covariables, aunque se sobrestiman los efectos

TABLA 2: Estimación de los parámetros de referencia (error estándar), número máximo de visitas = 4.

		Número máximo de visitas 4			
Distribución asumida	Parámetro de referencia	$n = 100$ Media (e.e.)	$n = 200$ Media (e.e.)	$n = 400$ Media (e.e.)	$n = 800$ Media (e.e.)
Normal	-0.809	-0.925(0.230)	-0.933(0.161)	-0.933(0.114)	-0.923(0.080)
	0.982	0.933(0.234)	0.922(0.164)	0.919(0.116)	0.920(0.082)
	1.825	1.885(0.265)	1.872(0.185)	1.871(0.131)	1.865(0.092)
	2.491	2.551(0.307)	2.518(0.214)	2.511(0.150)	2.499(0.106)
	Dicót. 0.117	0.127(0.046)	0.129(0.032)	0.128(0.022)	0.128(0.016)
	Cont. 0.785	0.824(0.421)	0.850(0.294)	0.836(0.206)	0.835(0.145)
Uniforme	-0.809	-0.823(0.191)	-0.819(0.134)	-0.807(0.094)	-0.803(0.066)
	0.982	1.061(0.173)	1.063(0.122)	1.073(0.086)	1.068(0.060)
	1.825	2.032(0.206)	2.023(0.145)	2.025(0.102)	2.024(0.072)
	2.491	2.692(0.253)	2.678(0.178)	2.679(0.125)	2.672(0.088)
	Dicót. 0.117	0.125(0.034)	0.124(0.024)	0.124(0.017)	0.123(0.012)
	Cont. 0.785	0.818(0.303)	0.795(0.211)	0.795(0.148)	0.796(0.104)
Cauchy	-0.809	-1.085(0.280)	-1.066(0.195)	-1.066(0.138)	-1.066(0.097)
	0.982	0.908(0.276)	0.918(0.193)	0.913(0.136)	0.910(0.096)
	1.825	1.941(0.309)	1.958(0.217)	1.949(0.153)	1.942(0.108)
	2.491	2.684(0.357)	2.676(0.250)	2.663(0.175)	2.654(0.124)
	Dicót. 0.117	0.137(0.056)	0.136(0.039)	0.135(0.027)	0.135(0.019)
	Cont. 0.785	0.905(0.508)	0.905(0.354)	0.893(0.250)	0.903(0.175)
Logística	-0.809	-1.038(0.307)	-1.018(0.216)	-1.027(0.152)	-1.027(0.107)
	0.982	0.645(0.305)	0.649(0.215)	0.639(0.152)	0.639(0.107)
	1.825	1.507(0.321)	1.497(0.226)	1.492(0.159)	1.494(0.112)
	2.491	2.106(0.342)	2.076(0.240)	2.068(0.169)	2.066(0.119)
	Dicót. 0.117	0.113(0.061)	0.115(0.043)	0.117(0.030)	0.115(0.021)
	Cont. 0.785	0.823(0.574)	0.787(0.400)	0.792(0.280)	0.800(0.198)
Laplace	-0.809	-0.959(0.255)	-0.946(0.179)	-0.947(0.126)	-0.917(0.135)
	0.982	0.833(0.253)	0.837(0.178)	0.829(0.125)	1.009(0.136)
	1.825	1.733(0.279)	1.740(0.196)	1.728(0.138)	2.009(0.155)
	2.491	2.353(0.314)	2.336(0.220)	2.323(0.154)	2.612(0.176)
	Dicót. 0.117	0.124(0.051)	0.122(0.035)	0.121(0.025)	0.128(0.027)
	Cont. 0.785	0.815(0.468)	0.792(0.323)	0.795(0.227)	0.825(0.244)

de las mismas. Con la distribución Cauchy se siguen sobrestimando los valores del parámetro de referencia y en la variable dicótoma las estimaciones son estables a medida que se aumenta el tamaño de la muestra pero muy alejadas del parámetro de referencia. Para las distribuciones logística y doble exponencial, los valores de referencia pasan de ser sobrestimados a ser subestimados, cuando se aumenta el número máximo de visitas de 3 a 5.

En general, se puede ver que la distribución uniforme muestra las estimaciones más cercanas a los valores de referencia y tiene los errores estándar más bajos, mientras que la Cauchy presenta las estimaciones más alejadas. La distribución normal sobrestima los valores de referencia y los errores estándar son altos. Las distribuciones logística y doble exponencial sobrestiman los valores de referencia cuando el número de visitas es 3 y las subestiman cuando el número de visitas es 5, lo que muestra que no son muy estables.

TABLA 3: Estimación de los parámetros de referencia (error estándar), número máximo de visitas = 5.

		Número máximo de visitas 5			
Distribución asumida	Parámetro de referencia	$n = 100$ Media (e.e.)	$n = 200$ Media (e.e.)	$n = 400$ Media (e.e.)	$n = 800$ Media (e.e.)
Normal	-0.809	-0.964(0.222)	-0.960(0.156)	-0.952(0.110)	-0.954(0.078)
	0.982	0.832(0.224)	0.827(0.158)	0.832(0.111)	0.832(0.078)
	1.825	1.728(0.248)	1.732(0.175)	1.734(0.123)	1.732(0.087)
	2.491	2.390(0.285)	2.378(0.199)	2.379(0.141)	2.375(0.099)
	Dicót. 0.117	0.125(0.044)	0.124(0.031)	0.125(0.021)	0.124(0.015)
	Cont. 0.785	0.856(0.405)	0.840(0.282)	0.836(0.198)	0.829(0.140)
Uniforme	-0.809	-0.841(0.181)	-0.826(0.127)	-0.826(0.089)	-0.827(0.063)
	0.982	1.026(0.164)	1.024(0.115)	1.024(0.081)	1.022(0.057)
	1.825	1.953(0.192)	1.941(0.134)	1.950(0.095)	1.943(0.067)
	2.491	2.621(0.235)	2.602(0.164)	2.611(0.116)	2.597(0.081)
	Dicót. 0.117	0.126(0.032)	0.124(0.022)	0.124(0.016)	0.124(0.011)
	Cont. 0.785	0.801(0.283)	0.800(0.196)	0.801(0.137)	0.802(0.097)
Cauchy	-0.809	-1.051(0.262)	-1.055(0.184)	-1.046(0.130)	-1.042(0.091)
	0.982	0.837(0.259)	0.815(0.181)	0.831(0.128)	0.830(0.090)
	1.825	1.807(0.285)	1.786(0.200)	1.796(0.141)	1.791(0.100)
	2.491	2.529(0.326)	2.500(0.228)	2.497(0.161)	2.489(0.113)
	Dicót. 0.117	0.126(0.051)	0.125(0.036)	0.125(0.025)	0.124(0.018)
	Cont. 0.785	0.874(0.473)	0.883(0.331)	0.864(0.232)	0.865(0.164)
Logística	-0.809	-1.062(0.299)	-1.049(0.210)	-1.048(0.148)	-1.047(0.104)
	0.982	0.517(0.296)	0.523(0.208)	0.522(0.147)	0.520(0.104)
	1.825	1.316(0.308)	1.315(0.216)	1.319(0.153)	1.311(0.108)
	2.491	1.904(0.325)	1.883(0.228)	1.886(0.161)	1.879(0.113)
	Dicót. 0.117	0.101(0.059)	0.102(0.041)	0.102(0.029)	0.101(0.020)
	Cont. 0.785	0.754(0.553)	0.767(0.388)	0.766(0.271)	0.775(0.192)
Laplace	-0.809	-0.963(0.243)	-0.949(0.171)	-0.956(0.120)	-0.952(0.085)
	0.982	0.748(0.240)	0.749(0.169)	0.744(0.119)	0.745(0.084)
	1.825	1.590(0.261)	1.592(0.183)	1.586(0.129)	1.587(0.091)
	2.491	2.205(0.290)	2.188(0.203)	2.174(0.143)	2.176(0.101)
	Dicót. 0.117	0.115(0.047)	0.111(0.033)	0.114(0.023)	0.114(0.016)
	Cont. 0.785	0.797(0.440)	0.778(0.307)	0.788(0.215)	0.780(0.152)

Cuando la covariable es de tipo dicótoma, se obtienen mejores aproximaciones que cuando la covariable considerada en el modelo es de tipo continua indistintamente de la distribución considerada bajo estudio. También se observa que cuando se aumenta el tamaño de la muestra de 100 a 400, se logra reducir el error estándar de las estimaciones a la mitad, y aumentar el número de visitas también reduce el error de las estimaciones.

Con base en estos resultados, se recomiendan tamaños de muestra mínimo de 400 y número de visitas mínimo de 4, para obtener estimaciones razonables. Una descripción más detallada de los resultados del estudio de simulación se puede ver en Tovar (2008).

6. Ilustración del método

Para ilustrar el método se consideran datos longitudinales de una cohorte de 980 estudiantes seleccionados aleatoriamente de la población total de estudiantes que se encontraban matriculados en el periodo 01 de 2005 en la Universidad Nacional de Colombia, sede Medellín.

Los datos consisten en promedios acumulados por semestre de los estudiantes durante cinco periodos académicos (01 de 2005 hasta 01 de 2007). El número de estudiantes en la muestra se determinó usando (6) (Brown & Prescott 1999):

$$n = \frac{2(z_{1-\alpha/2} + z_{\beta})^2 \sigma^2 \{1 + (l-1)\rho\}}{l\Delta^2} \quad (6)$$

donde: l es el número de medidas repetidas para cada estudiante; σ^2 es la variación entre los estudiantes para la variable de respuesta (promedios académicos); ρ es la correlación entre observaciones sobre el mismo estudiante; α es el nivel de significancia fijado de antemano; β es la potencia deseada de la prueba y Δ es una diferencia mínima a ser detectada en la variable de respuesta.

En este ejemplo, $\sigma^2 = 0.2435$, $\rho = 0.7674$; estos dos valores fueron calculados a partir de la base completa usando una rutina implementada en el *PROC MIXED* de SAS (Littell et al. 1996). Para una significancia $\alpha = 0.01$ y una potencia $\beta = 0.95$, con $l = 5$ mediciones (visitas) para cada estudiante, se tiene $n = 490$ personas a seleccionar en dos grupos, hombres y mujeres, para detectar una diferencia mínima de $\Delta = 0.12$ en la variable de respuesta.

Cada estudiante, en cada una de las cinco mediciones, es clasificado en una de cinco categorías mutuamente excluyentes, definidas como: Excelente Promedio, en los cuales se encuentran los estudiantes con promedio académico superior o igual a 4.0 y conservan la calidad de estudiante; Regular o Buen Promedio, estudiantes con promedio académico superior o igual a 3.0 pero menor a 4.0 y conservan la calidad de estudiante; Mal Promedio, estudiantes con promedio académico inferior a 3.0 y conservan la calidad de estudiante; Pérdida de Calidad I, en la cual se encuentran los estudiantes que perdieron la calidad de estudiante por bajo rendimiento, se retiraron, abandonaron la universidad o fueron expulsados por alguna medida disciplinaria u otra razón (excepto que hayan finalizado su ciclo de estudios); y Pérdida de Calidad II, en la cual se encuentran los estudiantes que perdieron la calidad de estudiante por finalización del ciclo de estudio.

Las covariables de interés que se tienen en cuenta para la construcción del modelo son:

- Género: una variable dicótoma que representa el género del estudiante (1 = hombre y 0 = mujer).
- TipoCol: una variable dicótoma que representa el tipo de colegio del cual es egresado un estudiante en el nivel secundaria (1 = privado y 0 = otro caso).
- Residen: una variable dicótoma que representa lugar de residencia del estudiante (1 = urbano y 0 = otro caso).

- Estrato: una variable policótoma que representa el estrato socioeconómico al cual pertenece el individuo (1 = bajo, 2 = medio y 3 = alto).
- Edad: edad en años del estudiante.

El número de transiciones entre los estados considerados durante el periodo de seguimiento se resume en la tabla 4.

Aquí se muestra que son muy pocos los estudiantes que, estando en el estado mal promedio, transitan en la próxima medición al estado excelente promedio (solo 14 estudiantes). Un comportamiento similar ocurre en la transición de excelente a mal promedio (11 estudiantes). Por otro lado, es muy poco común que un estudiante que esté en el estado mal promedio logre culminar satisfactoriamente su ciclo de estudios; esto se puede ver en la transición del estado mal promedio a pérdida de calidad II (un estudiante).

TABLA 4: Datos sobre promedios acumulados. Frecuencia y porcentaje de transiciones entre estados. Fuente: Centro de Cómputo y Telecomunicaciones, Universidad Nacional de Colombia, sede Medellín. Frecuencia(porcentaje).

Estado previo del promedio	Estado actual de promedio					Total
	Excelente	Regular o bueno	Malo	Pérdida de calidad I	Pérdida de calidad II	
Excelente	393(63.6)	165(26.7)	11(1.8)	20(3.2)	29(4.7)	618(100)
Regular o bueno	217(15.1)	869(60.6)	263(18.3)	70(4.9)	16(1.1)	1435(100)
Malo	14(1.6)	239(27.0)	285(32.2)	345(39.0)	1(0.1)	884(100)

El modelo que se considera en este ejemplo consiste en cinco estados mutuamente excluyentes en los cuales se puede clasificar a un estudiante (figura 2); tres estados transitorios: excelente promedio (estado 1), regular o buen promedio (estado 2) y mal promedio (estado 3); y dos estados absorbentes: pérdida de calidad I (estado 4) y pérdida de calidad II (estado 5).

En este modelo, el estado pérdida de la calidad I es considerado como un estado absorbente aunque en la práctica puede ocurrir que una persona que ha perdido la calidad de estudiante puede regresar a la universidad; sin embargo, y para ser consecuentes con el método propuesto, en este análisis se considera que el estudiante no regresa a la universidad después de haber perdido la calidad de estudiante, y por tanto esta información posterior a la primera pérdida de la calidad de estudiante no es tenida en cuenta en el modelo.

Para calcular la matriz de transición usando los parámetros estimados, se incorpora el estado previo del estudiante en el modelo por medio de variables indicadoras: $PrevioExc$ igual a 1 si el estado previo es excelente promedio y 0 en otro caso; $PrevioRegBuen$ igual a 1 si el estado previo es regular o buen promedio y 0 en otro caso.

Note que cuando los dos $PrevioExc$ y $PrevioRegBuen$ son iguales a cero, el estado previo de los estudiantes es mal promedio.

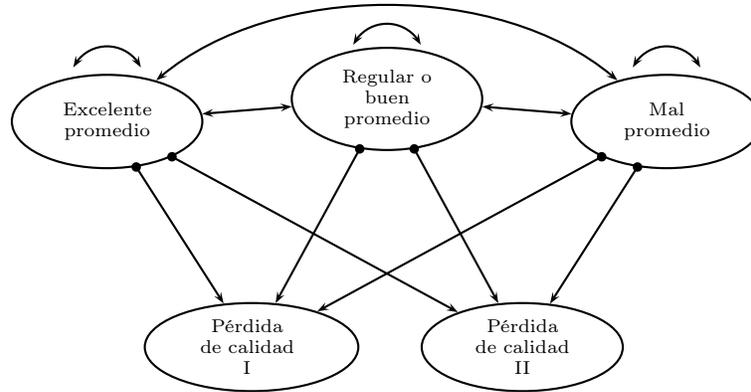


FIGURA 2: Modelo de estados múltiples, con 3 estados transitorios y dos estados absorbentes para el promedio académico de un estudiante.

Para esta aplicación se asume que el efecto aleatorio se distribuye como una $N(0, \sigma^2)$, y el log de la verosimilitud, $\ell(\beta, \sigma^2 | Z)$, se maximiza usando el método de la cuadratura de Gauss con 32 pesos gaussianos, combinado con el algoritmo de Newton-Raphson.

Usando un procedimiento de selección hacia atrás, se excluyeron del modelo las covariables Estrato y Residen. Los parámetros estimados que se obtienen al ajustar el modelo con las covariables Género, TipoCol, Edad, PrevioExc y PrevioRegBuen, se muestran en la tabla 5.

TABLA 5: Estimaciones por máxima verosimilitud para los parámetros del modelo logit.

Parámetro	Estimado	Error estándar	Estadístico de Wald	Valor P
Intercepto1	-2.7158	0.3795	51.2120	< 0.0001
Intercepto2	0.2834	0.3742	0.5736	0.4488
Intercepto3	1.7768	0.3755	22.3902	< 0.0001
Intercepto4	4.6432	0.4035	132.4182	< 0.0001
Género	-0.2781	0.1018	7.4629	0.0063
TipoCol	0.2778	0.1015	7.4908	0.0062
Edad	-0.0443	0.0173	6.5572	0.0104
PrevioExc	3.4231	0.1490	527.7967	< 0.0001
PrevioRegBuen	1.6673	0.1009	273.0519	< 0.0001
Log likelihood	-3419.69			< 0.0001

Usando estas estimaciones para los coeficientes de las covariables consideradas en el modelo, es posible calcular razones de *odds* como se muestran en la tabla 6, por ejemplo: el logit estimado de estar en la categoría regular o buen promedio para un hombre (Género=1), de 19 años de edad (Edad=19), egresado de un colegio privado (TipoCol=1) cuando el estado previo es mal promedio (PrevioExc=0 y

PrevioReg=0), denotado $\text{logit}_{32,H}$ es:

$$\begin{aligned}\text{logit}_{32,H} &= 0.2834 - 0.2781(1) + 0.2778(1) - 0.0443(19) + 3.4231(0) + 1.6673(0) \\ &= -0.5586\end{aligned}$$

También, el logit estimado de estar en la categoría regular o buen promedio para una mujer (Género=0), de 19 años de edad, egresada de un colegio privado (Tipo-Col=1) cuando el estado previo es mal promedio (PrevioExc=0 y PrevioReg=0), denotado $\text{logit}_{32,M}$ es:

$$\begin{aligned}\text{logit}_{32,H} &= 0.2834 - 0.2781(0) + 0.2778(1) - 0.0443(19) + 3.4231(0) + 1.6673(0) \\ &= -0.2805\end{aligned}$$

por lo tanto, la razón de *odds* de estar en la categoría regular o buen promedio, que compara hombres *vs.* mujeres es:

$$\text{OR}_{H\text{vs.}M} = \frac{\exp(-0.5586)}{\exp(-0.2805)} = 0.76$$

esto significa que el riesgo de pasar a la categoría regular o buen promedio para un hombre de 19 años que ha estudiado en colegio privado, es 0.76 veces el riesgo de pasar a la categoría regular o buen promedio de una mujer de la misma edad y en el tipo de colegio, cuando su promedio académico en el semestre anterior es mal promedio.

Note que

$$\text{OR}_{H\text{vs.}M} = \frac{\exp(-0.5586)}{\exp(-0.2805)} = \exp[-0.5586 - (-0.2805)] = \exp(-0.2781)$$

TABLA 6: Razón de *odds* (OR) e intervalo de confianza al 95 % para la razón de *odds*, para los factores de riesgo.

Comparación	OR	IC al 95 %
Hombres <i>vs.</i> Mujeres	0.76	(0.62,0.92)
Colegio privado <i>vs.</i> Colegio no privado	1.32	(1.08,1.61)
Edad	0.97	(0.92,0.99)
Exce <i>vs.</i> No Exce.	30.66	(22.90,41.06)
Regular o Bueno <i>vs.</i> No Regular o Bueno	5.30	(4.35,6.46)

La tabla 7 muestra las matrices de transición típicas para valores arbitrarios de las covariables. Estas matrices reflejan el efecto del género sobre el promedio acumulado de los estudiantes. Al lado izquierdo se muestra la matriz de transición de un estudiante con género masculino, de 20 años de edad graduado en colegio privado, mientras que al lado derecho se encuentra la matriz de transición para un estudiante de género femenino, con 20 años de edad y graduada en colegio privado.

TABLA 7: Intercepto aleatorio. Panel izquierdo: matriz de transición para un hombre (Género=1), de 20 años de edad graduado en colegio privado (TipoCol=1), con $b_i = 0.08356253$. Panel derecho: matriz de transición para una mujer (Género=0), de 20 años de edad graduada en colegio privado (TipoCol=0), con $b_i = 0.08356253$.

0.476	0.472	0.040	0.011	0.001	0.546	0.415	0.031	0.009	0.001
0.136	0.623	0.174	0.062	0.004	0.172	0.633	0.143	0.048	0.003
0.029	0.344	0.353	0.253	0.021	0.038	0.402	0.338	0.200	0.023
0	0	0	1	0	0	0	0	1	0
0	0	0	0	1	0	0	0	0	1

En la entrada [3, 4] de ambas matrices se puede ver que, si en el semestre anterior un hombre de 20 años de edad, graduado en colegio privado, tuvo mal promedio (estado previo=3), el riesgo de perder la calidad de estudiante por bajo rendimiento, abandonar la universidad o ser expulsado es 25.3% (panel izquierdo en la tabla 7), mientras que para una mujer de la misma edad y mismo tipo de colegio, que ha tenido mal promedio en el semestre anterior, el riesgo de perder la calidad de estudiante por bajo rendimiento, abandonar la universidad o ser expulsado es 20.0% (panel derecho en la tabla 7); luego los hombres tienen $25.3/20.0 = 1.265$ veces más riesgo de perder la calidad de estudiante por bajo rendimiento, abandonar la universidad o ser expulsado que una mujer, cuando el promedio en el semestre anterior es regular; esto es, el riesgo para los hombres es mayor en un 26.5%.

Las figuras 3 y 4 muestran las probabilidades estimadas de perder la calidad de estudiante por bajo rendimiento, abandonar la universidad o ser expulsados, como una función de la edad de los estudiantes (en personas del mismo tipo de colegio), para género masculino y femenino respectivamente, dependiendo del estado previo.

Se observa que cuando el estado previo es mal promedio acumulado, la probabilidad de transitar al estado pérdida de calidad I es mayor que cuando el estado previo es regular o bueno, o excelente promedio acumulado (en ambos géneros). También se muestra que, independientemente del género, los estudiantes con menor riesgo de transitar al estado pérdida de calidad I son los que se encuentran en la categoría excelente promedio acumulado.

Finalmente, el ajuste del modelo con intercepto aleatorio se justifica comparando el log verosimilitud con el modelo sin efecto aleatorio; $-2(\ell_{aleatorio} - \ell_{noaleatorio}) = -2(-3419.69 + 3344.46) = 150.46$, $P < 0.0001$; sin embargo, el ajuste de ambos modelos (con efecto aleatorio y sin efecto aleatorio) es adecuado.

7. Conclusiones

En este artículo se discutió el problema de estimación de parámetros en una cadena de Markov con dos estados absorbentes que compiten entre sí, asumiendo un modelo de regresión logística policótoma con respuesta ordinal e intercepto aleatorio, bajo distintas formas distribucionales de dicho término aleatorio.

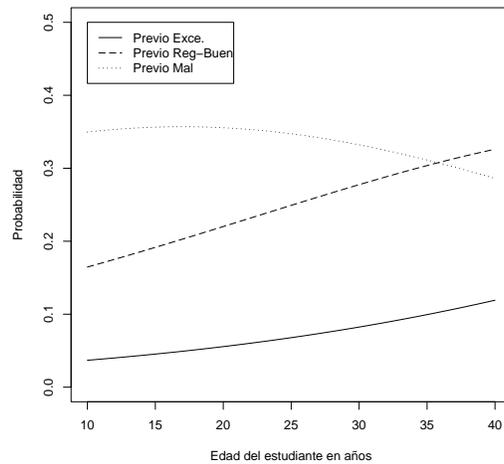


FIGURA 3: Probabilidad estimada de estar en la categoría: pérdida de la calidad I como una función de la edad. Género masculino.

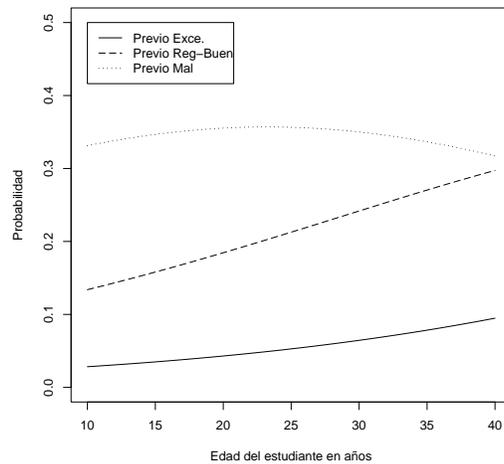


FIGURA 4: Probabilidad estimada de estar en la categoría: pérdida de la calidad I como una función de la edad. Género femenino.

Se mostró cómo la teoría de cadenas de Markov ayuda a reducir la complejidad de la función de verosimilitud propuesta. El efecto de los diferentes supuestos de la forma distribucional del efecto aleatorio sobre el vector de parámetros desconocidos es evaluado por medio de un estudio de simulación. En él se encuentran algunas condiciones de tamaño de muestra y número de visitas necesarios para

obtener estimaciones razonables de los coeficientes de las covariables consideradas en el modelo, cuando el proceso es visto como un modelo de regresión logística policótoma con respuesta ordinal.

La estimación de estos parámetros nos permite calcular las probabilidades de transición de un estado a otro, como se ilustra en el ejemplo de aplicación.

Una de las ventajas más relevantes de este método es que tiene en cuenta la naturaleza ordinal de la variable de respuesta, y se pueden incorporar variables de tipo categórico y continuas en el modelo, sin embargo la incorporación del efecto aleatorio aumenta los errores estándar de las estimaciones.

De acuerdo con los resultados del estudio de simulación, este método es robusto a cambios en la distribución para el término de efecto aleatorio.

Este método también es útil para analizar datos categóricos provenientes de estudios longitudinales y permite hacer inferencia para los predictores por medio de razones de *odds*.

Por medio de este modelo se puede estudiar el papel de los predictores sobre la transición de un sujeto específico y cómo la matriz de probabilidades de transición de un paso puede ser obtenida.

Una de las desventajas del método es que, al introducir nuevas covariables, se dificultan las estimaciones de los parámetros de referencia.

En el estudio de simulación, el método muestra buenas aproximaciones cuando se considera que la distribución del término aleatorio es normal y uniforme. Cuando se considera que la distribución del término aleatorio es Cauchy estimaciones muy por encima de los valores de referencia y cuando se asume logística o Laplace (doble exponencial), el método produce sobrestimaciones y subestimaciones. Además, cuando las covariables son dicótomas, las estimaciones son mejores que cuando son continuas.

[Recibido: julio de 2008 — Aceptado: septiembre de 2009]

Referencias

- Abramowitz, M. & Stegun, I. A. (1972), *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Dover Publications, INC, New York, United States.
- Aitkin, M. A. & Alfó, M. (1998), 'Regression Models for Binary Longitudinal Responses', *Statistics and Computing* **8**, 289–307.
- Brown, H. & Prescott, R. (1999), *Applied Mixed Models in Medicine*, John Wiley & Sons, New York, United States.
- Diggle, P., Heagerty, P., Liang, K. & Zeger, S. (2001), *Analysis of Longitudinal Data*, 2 edn, Oxford University Press, New York, United States.
- Gao, S. (2004), 'A shared effect parameter approach for longitudinal dementia data with non-ignorable missing data', *Statistics in Medicine* **23**, 211–219.

- Hedeker, D. & Gibbons, R. D. (1994), 'A Random-Effects Ordinal Regression Models for Multilevel Analysis', *Biometrics* **50**(4), 933–944.
- Hedeker, D., Siddiqui, O. & Hu, F. B. (2000), 'Random Effects Regression Analysis of Correlated Grouped Time Survival Data', *Statistical Methods in Medical Research* **9**(2), 161–179.
- Littell, R. C., Milliken, G. A., Stroup, W. W. & Wolfinger, R. D. (1996), *SAS System for Mixed Models*, SAS Institute Inc.
- Salazar, J. C. (2004), Multi-State Markov Models for Longitudinal Data, Tesis de doctorado, Departamento de Estadística, Universidad de Kentucky, Lexington, United States.
- Salazar, J. C., Schmitt, F. A., Yu, L., Mendiondo, M. S. & Kryscio, R. J. (2007), 'Shared Random Effects Analysis of Multi-State Markov Models: Application to a Longitudinal Study of Transitions to Dementia', *Statistics in Medicine* **26**, 568–580.
- SAS Institute Inc. (1989), *SAS/IML Software: usage and reference, versión 6*, 1 edn, Cary, NC: SAS Institute Inc.
- Skrondal, A. & Rabeth-Hesketh, S. (2004), *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*, Chapman & Hall/CRC, Boca Raton, United States.
- Ten Have, T. R., Miller, M., Reboussin, B. A. & James, M. K. (2003), 'Mixed Effects Logistic Regression Models for Longitudinal Ordinal Functional Response Data with Multiple-Cause Drop-Out from the Longitudinal Study of Aging', *Biometrics* **56**(1), 279–287.
- Tovar, R. J. (2008), Efecto en las estimaciones bajo distintas formas distribucionales del término aleatorio en un modelo lineal mixto adaptado a una cadena de Markov con espacio de estados ordinal, Tesis de maestría, Escuela de Estadística, Universidad Nacional de Colombia, Medellín, Colombia.