



Servy, Elsa

Marí, Gonzalo

Wojdyla, Daniel

Instituto de Investigaciones Teóricas y Aplicadas en Estadística, Escuela de Estadística

ESTIMACIÓN DE PROBABILIDADES DE TRANSICIÓN BAJO DISEÑOS MUESTRALES COMPLEJOS*

1. INTRODUCCIÓN

La existencia de desempleo o desocupación en el mercado de trabajo es de notable importancia, ya que eso indica que una parte de la población no participa, al menos como trabajadores, del esfuerzo productivo ni de la distribución del producto social. Esto significa un despilfarro de recursos para el país y una penuria para los desocupados.

En Argentina, los Censos y la Muestra Permanente de Hogares (EPH) permiten medir la magnitud de la desocupación. Los datos provenientes de la EPH permiten estudios dinámicos y comparaciones espaciales, pues la encuesta se viene realizando desde 1970, en dos ocasiones por año y alcanza 28 aglomerados del país.

Este estudio está dirigido al aspecto dinámico de la condición de actividad a través del tiempo. En particular, toma en cuenta la tasa de desempleo condicional al estado ocupacional anterior. Por lo tanto requiere información de dos ondas consecutivas de la EPH, ya que incluye el estado ocupacional en la onda anterior como variable explicativa.

La EPH utiliza una muestra de tipo estratificado, con etapas múltiples de selección. El esquema de rotación de la encuesta impone que en cada onda salga de ella una submuestra, constituida por el 25% de los hogares, los que son reemplazados por un número equivalente de hogares elegidos en forma independiente de modo que la muestra, después de cuatro ondas, es sustituida en su totalidad.

Para la Escuela de Estadística la problemática de la evaluación de las probabilidades de transición desde un estado ocupacional a otro no es un tema nuevo. Para los años 96 y 97 se calcularon las probabilidades de transición entre los estados ocupados, desocupados e inactivos y se realizó un test de estacionariedad de esas transiciones a través de las 4 ondas de la EPH. Con datos de cuatro ondas, las transiciones son sólo tres. El resultado del test rechazó la estacionariedad (Servy, Hachuel, Boggio, Cuesta (1999)).

Se realizaron también estimaciones y tests respecto de las probabilidades de transición, ajustándolas por variables socio-económicas y demográficas (Servy, Hachuel, Boggio, Cuesta (1999) y Servy, Hachuel, Boggio (2000)).

* Trabajo realizado en el marco del proyecto de investigación "Modelos referidos al fenómeno del desempleo a partir de la Encuesta Permanente de Hogares: Alcance y Limitaciones". PICT 02-09897. Agencia de Promoción de la Investigación Científica y Tecnológica

En algunos de los métodos utilizados se tomó en cuenta el carácter rotativo de la muestra haciendo uso de un método que es una adaptación de la metodología de Grizzle, Starmer y Koch.

Sin embargo, en esos análisis no se tomó en cuenta la naturaleza compleja de la muestra de la EPH, es decir, el plan de diseño complejo con múltiples etapas que incluye estratos, así como radios y hogares como unidades sucesivas de muestreo.

El propósito de este trabajo es evaluar mediante un estudio de simulación hasta qué punto se distorsionan las propiedades de los estimadores clásicos - creados bajo el supuesto de que la muestra es aleatoria simple - cuando se los aplica a una muestra de estructura compleja

2. MÉTODO CLÁSICO PARA ESTIMAR LAS PROBABILIDADES DE TRANSICIÓN

2.1. Se considera el método clásico propuesto por Anderson y Goodman (1957). Supóngase que el espacio de estados de una cadena de Markov tiene k elementos $\{1, 2, \dots, k\}$, y que los tiempos de observación de la cadena son T , indexados por el subíndice t ($t=1, 2, \dots, T$). Sea $n_i(0)$ el número de unidades en el estado i -ésimo ($i=1, 2, \dots, k$) en el tiempo inicial $t=0$. Sea ε_i ($i=1, 2, \dots, k$) la probabilidad asociada con $n_i(0)$. Se puede pensar a los $n_i(0)$'s como variables aleatorias distribuidas multinomialmente con probabilidades ε_i . Supóngase que se observan N unidades experimentales, donde $N = \sum_{i=1}^k n_i(0)$ de modo que una observación de una unidad experimental consiste de una sucesión de k eventos observados por ella en cada uno de los tiempos $t=0, 1, \dots, T$. Sean ellos (x_0, x_1, \dots, x_T) . De esa manera, la distribución marginal del conjunto $\{n_i(0), i=1, 2, \dots, k\}$ es,

$$\frac{N!}{\prod_{i=1}^k n_i(0)!} \prod_{i=1}^k \varepsilon_i^{n_i(0)} \quad (2.1.1)$$

Cada observación corresponde a una realización de una cadena de Markov. Dado un estado inicial x_0 , existen k^T posibles sucesiones. Sea $n_{ij}(t)$ el número de unidades que se encuentran en el estado i en el tiempo $t-1$ y en el estado j en el tiempo t . El conjunto $\{n_{ij}(t), i, j=1, \dots, k; t=1, \dots, T\}$ de cardinalidad $k^2 T$ es un conjunto de estadísticas suficientes para las sucesiones observadas. Sean,

$$\begin{aligned} n_{\bullet j}(t) &= \sum_{i=1}^k n_{ij}(t) = n_j(t), \\ n_{i\bullet}(t) &= \sum_{j=1}^k n_{ij}(t) = n_i(t-1). \end{aligned} \quad (2.1.2)$$

Sea $p_{ij}(t-1, t)$ la probabilidad condicional de estar en el estado j en el tiempo t , dado que estaba en el estado i en el tiempo $(t-1)$. Para un valor de t fijo, la matriz de transición es,

$$P(t-1, t) = \begin{bmatrix} p_{11}(t-1, t) & \cdots & p_{1k}(t-1, t) \\ \vdots & \ddots & \vdots \\ p_{k1}(t-1, t) & \cdots & p_{kk}(t-1, t) \end{bmatrix}. \quad (2.1.3)$$

Para un estado fijo i , las $\{n_{ij}(t), \dots, n_{ik}(t)\}$ o sea las frecuencias de estar en el estado i -ésimo en el momento $t-1$ y en el estado j -ésimo en el estado t , tienen condicionalmente a $n_i(t-1)$ una distribución multinomial,

$$\frac{n_i(t-1)!}{\prod_{j=1}^k n_{ij}(t)!} \prod_{j=1}^k p_{ij}(t-1, t)^{n_{ij}(t)}. \quad (2.1.4)$$

Tratándose de una cadena de Markov, donde la distribución de $\{n_{ij}(t), i, j=1, \dots, k\}$ es independiente (condicionalmente) a $n_i(t-1)$, de los resultados previos de la cadena, $P(n_i(0), n_{ij}(t), i, j=1, \dots, k; t=1, \dots, T)$ es

$$\frac{N!}{\prod_{i=1}^k n_i(0)!} \prod_{i=1}^k \varepsilon_i^{n_i(0)} \prod_{t=1}^T \left\{ \prod_{i=1}^k \frac{n_i(t-1)!}{\prod_{j=1}^k n_{ij}(t)!} \prod_{j=1}^k p_{ij}(t-1, t)^{n_{ij}(t)} \right\}. \quad (2.1.5)$$

La función de verosimilitud de $\{\varepsilon_i\}$ y $p_{ij}(t-1, t)$ es un producto de dos funciones, una que involucra a los ε 's y otra a los p_{ij} 's, o sea,

$$L = \left[\prod_{i=1}^k \varepsilon_i^{n_i(0)} \right] \times \left[\prod_{t=1}^T \prod_{i=1}^k \prod_{j=1}^k p_{ij}(t-1, t)^{n_{ij}(t)} \right] \quad (2.1.6)$$

Las estimaciones de máxima verosimilitud son,

$$\hat{\varepsilon}_i = n_i(0) \quad (2.1.7)$$

$$\hat{p}_{ij}(t-1, t) = \frac{n_{ij}(t)}{n_i(t-1)}. \quad (2.1.8)$$

Estas estimaciones son las mismas que las que serían obtenidas si para cada i y t tuviésemos $n_i(t-1)$ observaciones provenientes de una distribución multinomial con probabilidades $p_{ij}(t-1, t)$ y frecuencias $n_{ij}(t)$.

Para una cadena de Markov con probabilidades de transición estacionarias, un resultado más fuerte sobre suficiencia es que el conjunto $n_{ij} = \sum_{t=1}^T n_{ij}(t)$ forma un conjunto de estadísticas suficientes.

En este caso, $p_{ij}(t-1, t) = p_{ij}$ para $t=1, \dots, T$ y el segundo factor de la función de verosimilitud se reduce a

$$\prod_{i=1}^k \prod_{j=1}^k p_{ij}^{n_{ij}(t)}, \quad (2.1.9)$$

y las estimaciones de las probabilidades de transición son

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i\bullet}} = \frac{\sum_{t=1}^T n_{ij}(t)}{\sum_{t=1}^T n_i(t-1)} \quad (2.1.10)$$

Las estimaciones de máxima verosimilitud de las probabilidades iniciales se mantienen invariables.

Podemos describir esas estimaciones de la manera siguiente. En la estructura más general, cuando las probabilidades de transición no son estacionarias, las transiciones observadas pueden ser representadas en términos de un conjunto de estadísticas suficientes $\{n_{ij}(t)\}$ y pueden ser tratadas como si fuesen las frecuencias observadas $x_{ij}=n_{ij}(t)$ de una tabla de contingencia $k \times k \times T$, donde el total para cada nivel de T es igual al tamaño de la muestra. Cuando las probabilidades de transición son estacionarias, se puede reducir la tabla $k \times k \times T$ a una tabla $k \times k$ con frecuencias $n_{ij}=x_{ij}=\sum_{t=1}^T x_{ijt}$.

2.2. La suposición de estacionariedad, o sea, que las probabilidades de transición son independientes del tiempo, puede ser sometida a test. La hipótesis de estacionariedad puede escribirse, para i fijo

$$H_0 : p_{ij}(t) = p_{ij} \quad t = 1, 2, \dots, T \quad (2.2.1)$$

Bajo la hipótesis alternativa, las estimaciones de las probabilidades de transición en el tiempo t son,

$$\hat{p}_{ij} = \frac{n_{ij}(t)}{n_i(t-1)} \quad (2.2.2)$$

Para un i dado, el conjunto $\{p_{ij}(t), i, j=1, \dots, k; t=1, \dots, T\}$ tiene la misma distribución asintótica que los estimadores de las probabilidades multinomiales $p_{ij}(t)$ de T muestras independientes.

Sea la siguiente una representación tabular de las estimaciones de las probabilidades de transición calculadas según (2.2.2)

Tabla 2.2.1

| t | j | | | |
|---|-------------------|-------------------|---|-------------------|
| | 1 | 2 | ↑ | k |
| 1 | $\hat{p}_{i1}(1)$ | $\hat{p}_{i2}(1)$ | ↑ | $\hat{p}_{ik}(1)$ |
| 2 | $\hat{p}_{i1}(2)$ | $\hat{p}_{i2}(2)$ | ↑ | $\hat{p}_{ik}(2)$ |
| → | → | → | | → |
| T | $\hat{p}_{i1}(T)$ | $\hat{p}_{i2}(T)$ | ↑ | $\hat{p}_{ik}(T)$ |

Si la hipótesis de estacionariedad se cumple, todas las estimaciones de una misma columna son estimaciones de un parámetro común p_{ij} . Según la hipótesis de estacionariedad, un estimador de p_{ij} es \hat{p}_{ij} como en (2.1.10).

La estadística Chi Cuadrado para el test es

$$\phi_i = \sum_{t=1}^T \sum_{j=1}^k \frac{n_i(t-1) [\hat{\rho}_{ij}(t) - \hat{\rho}_{ij}]^2}{\hat{\rho}_{ij}} \quad (2.2.3)$$

Bajo la hipótesis nula, ϕ_i tiene una distribución χ^2 con $(k-1)(T-1)$ grados de libertad.

O bien puede usarse el criterio de Neyman-Pearson

$$\lambda_i = -2 \sum_{t=1}^T \sum_{j=1}^k n_{ij}(t) \log \left(\frac{\hat{\rho}_{ij}}{\hat{\rho}_{ij}(t)} \right) \quad (2.2.4)$$

que tiene una distribución asintótica χ^2 con $(k-1)(T-1)$ grados de libertad.

3. ESTUDIO DE SIMULACIÓN

3.1 ALGORITMO USADO EN LAS SIMULACIONES

Los datos se generan como matrices de k columnas y tres filas. Las columnas representan los k individuos del mismo conglomerado, observados en tres períodos sucesivos

$$\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ x_{13} & x_{23} & \cdots & x_{k3} \end{pmatrix} \quad (3.1.1)$$

Si los conglomerados se asocian con los hogares de la EPH, cada vector/columna representa a un miembro del hogar y cada fila a una onda de la EPH.

Las variables x_{ij} ($i=1, \dots, k$, $j=1, 2, 3$) toman los valores 0 y 1 (que designan en la aplicación considerada los estados: ocupado y desocupado). Los vectores-columna (que representan los individuos del conglomerado) están correlacionados entre sí de la manera descrita en Servy, Hachuel, Wojdyla (1998). En la misma se describe un modelo (denominado MMCC) para generar conglomerados de variables categóricas bivariadas. A los efectos de la aplicación que se describe más adelante, el modelo y el algoritmo de simulación se extendieron a tres dimensiones.

De acuerdo con el algoritmo, se pueden generar conglomerados de vectores tridimensionales tales que, la distribución general de vectores que ignora su pertenecía a diferentes conglomerados, puede ser fijada por el usuario. Es decir, es posible determinar las probabilidades de los eventos

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\} \quad (3.1.2)$$

en la población combinada, que ignora los conglomerados.

En este estudio, las probabilidades $p(x_1, x_2, x_3)$ se fijaron de forma tal de que ellas coincidan con las de una cadena estacionaria de Markov de primer orden



$$p(x_1, x_2, x_3) = p(x_1) p(x_2 | x_1) p(x_3 | x_2) \quad (3.1.3)$$

Las probabilidades $p(x_3 | x_2)$ se fijaron iguales a las correspondientes a $p(x_2 | x_1)$

3.2 PLAN DE SIMULACIÓN

- Los parámetros utilizados en el estudio fueron los siguientes:

Para la población combinada la distribución inicial es

$$P(x_1) = \begin{cases} 0.8 & \text{si } x_1 = 0 \\ 0.2 & \text{si } x_1 = 1 \end{cases} \quad (3.2.1)$$

Las probabilidades condicionales $P(x_2/x_1)$ se exhiben en las tablas siguientes

| x_1 | x_2 | |
|-------|-------|------|
| | 0 | 1 |
| 0 | 0.50 | 0.50 |
| 1 | 0.50 | 0.50 |

| x_1 | x_2 | |
|-------|-------|------|
| | 0 | 1 |
| 0 | 0.55 | 0.45 |
| 1 | 0.45 | 0.55 |

| x_1 | x_2 | |
|-------|-------|------|
| | 0 | 1 |
| 0 | 0.60 | 0.40 |
| 1 | 0.40 | 0.60 |

| x_1 | x_2 | |
|-------|-------|------|
| | 0 | 1 |
| 0 | 0.65 | 0.35 |
| 1 | 0.35 | 0.65 |

| x_1 | x_2 | |
|-------|-------|------|
| | 0 | 1 |
| 0 | 0.70 | 0.30 |
| 1 | 0.30 | 0.70 |

(3.2.2)

Las mismas tablas se usaron para fijar las probabilidades $P(x_2/x_1)$. Luego, las probabilidades de transición fueron establecidas como homogéneas a través de las ondas.

- Las correlaciones intra-conglomerado se tomaron en tres categorías: "baja", "media" y "alta".
- Los tamaños de las muestras de conglomerados fueron fijadas así: 15, 30, 70, 100, 500 y 1000.
- Los tamaños de los conglomerados, k , se tomaron entre 2 y 5.

Dentro del campo de variación de los parámetros, sólo se generaron datos de aquellos casos que son "viables" según el modelo MMCC. Para cada escenario "viable" se generaron 1000 muestras.

3.3 MÉTODO DE EVALUACIÓN DE LA ESTADÍSTICA DE ANDERSON Y GOODMAN

Para cada uno de los diversos escenarios (distintas matrices de transición y distintas correlaciones entre sujetos del mismo conglomerado), considerando seis tamaños muestrales diferentes ($n=15, 30, 70, 100, 500, 1000$) y cuatro tamaños de conglomerados ($k=2,3,4,5$) se realizaron 1000 repeticiones del experimento que consiste en la generación

de una muestra de n conglomerados trivariados de tamaño k .

Con cada muestra se calculó $\hat{p}(0|0)$, un estimador de $p(x_2|x_1)=p(x_3|x_2)=p(0|0)$, a partir de la fórmula (2.1.10) considerando que en todos los escenarios se supuso que las probabilidades de transición eran homogéneas a través del tiempo u onda. El estimador de la i -ésima muestra se designa $\hat{p}_i(0|0)$.

Con las 1000 repeticiones realizadas para cada escenario se calcularon las siguientes estadísticas

$$\bar{\hat{p}}(0|0) = \frac{1}{1000} \sum_{i=1}^{1000} \hat{p}_i(0|0); \frac{(\bar{\hat{p}}(0|0) - p(0|0))}{p(0|0)} * 100 \quad (3.3.1)$$

donde $p(0|0)$ es el verdadero valor fijado en (3.2.2). La segunda expresión en (3.3.1) es el porcentaje del sesgo relativo de la media.

También se calculó,

$$ECM = \frac{1}{1000} \sum_{i=1}^{1000} (\bar{\hat{p}}(0|0) - p(0|0))^2 \quad (3.3.2)$$

A partir de (3.3.1) y (3.3.2) se obtuvo un estimador de la variancia real de $\hat{p}_i(0|0)$

$$\text{var}_r = ECM - (\bar{\hat{p}}(0|0) - p(0|0))^2 \quad (3.3.3)$$

Si se considera que la muestra de nk individuos es aleatoria simple (mas), ella provee una estimación de la variancia del estimador $p_i(0|0)$ de la forma

$$\frac{\hat{p}_i(0|0)(1 - \hat{p}_i(0|0))}{nk} \quad (3.3.4)$$

Usando las 1000 repeticiones, se calculó el valor promedio de esa variancia,

$$\text{var}_{mas} = \frac{1}{1000} \sum_{i=1}^{1000} \frac{\hat{p}_i(0|0)(1 - \hat{p}_i(0|0))}{nk} \quad (3.3.5)$$

Este valor estima el valor esperado de la variancia que un analista habría calculado si hubiese creído que la muestra era de tipo simple al azar. Como un indicador del error que puede cometerse al ignorar el verdadero diseño muestral, se tomó el porcentaje del sesgo relativo,

$$\frac{\text{var}_{mas} - \text{var}_r}{\text{var}_r} * 100 \quad (3.3.6)$$

Los indicadores (3.3.1) y (3.3.6) se presentan en las tablas 1, 2, 3, 4, 5, 6 del Apéndice.

4. ANÁLISIS DE LOS RESULTADOS

Las medias, sobre tamaños de conglomerados y valores de $p(0|0)$, de los valores absolutos de los porcentajes del sesgo relativo (en la estimación de $p(0|0)$) se exhiben en la tabla 4.1.

El sesgo relativo es pequeño, y disminuye en valor absoluto cuando el tamaño de la muestra aumenta. El signo es variable, y hay una leve tendencia a aumentar en valor absoluto cuando las correlaciones entre los individuos del mismo conglomerado aumentan.

Tabla 4.1 – Porcentajes del sesgo relativo (en valor absoluto) de la estimación $\hat{\rho}_i(0|0)$ según tamaño de muestra de conglomerados y correlación intra-conglomerado

| Correlación | Tamaño de Muestra | | | | | |
|-------------|-------------------|-------|-------|-------|-------|-------|
| | 15 | 30 | 70 | 100 | 500 | 1000 |
| Baja | 0.306 | 0.367 | 0.146 | 0.117 | 0.084 | 0.043 |
| Media | 0.387 | 0.462 | 0.211 | 0.120 | 0.039 | 0.068 |
| Alta | 0.534 | 0.435 | 0.229 | 0.189 | 0.100 | 0.092 |

Las medias de los valores absolutos de los porcentajes del sesgo relativo de la variancia que supone una muestra simple al azar respecto a la que toma en cuenta el verdadero diseño, se presentan en la tabla 4.2.

Las simulaciones muestran que la verdadera variancia del estimador se subestima cuando se ignora el carácter complejo de la muestra. Esta subestimación aumenta considerablemente cuando el tamaño de la correlación intra-clase sube, y no disminuye cuando se incrementa el tamaño muestral.

Tabla 4.2 – Porcentajes del sesgo relativo (en valor absoluto) de la estimación de la variancia del estimador $\hat{\rho}_i(0|0)$ cuando se ignora la estructura compleja de la muestra

| Correlación | Tamaño de Muestra | | | | | |
|-------------|-------------------|-------|-------|-------|-------|-------|
| | 15 | 30 | 70 | 100 | 500 | 1000 |
| Baja | 24.66 | 22.60 | 21.86 | 23.26 | 24.09 | 22.69 |
| Media | 49.74 | 52.88 | 48.66 | 50.49 | 51.30 | 49.18 |
| Alta | 65.05 | 59.11 | 64.49 | 54.71 | 64.60 | 64.59 |

5. COMENTARIOS Y CONCLUSIONES

El presente estudio muestra que el estimador de Anderson y Goodman para las probabilidades de transición entre los estados de una cadena de Markov estacionaria con un espacio de estados constituido por dos eventos (asimilables a ocupado y desocupado en nuestra aplicación), produce estimaciones puntuales próximas a las verdaderas, aún cuando la muestra de conglomerados es pequeña. El signo de la diferencia puede ser positivo o negativo y para muestras de 500 o más conglomerados la diferencia relativa se torna despreciable.

Por el contrario, la variancia del estimador en cuestión es subestimada notablemente si se utiliza para su estimación la fórmula que corresponde a un diseño de muestra aleatorio simple.



El sesgo aumenta notablemente cuando los conglomerados de la muestra incrementan su correlación intraclase, y ello es independiente del tamaño de la muestra.

Se piensa continuar esta investigación para determinar las distorsiones que pueden presentar los errores de tipos I y II en el test de estacionariedad propuesto por Anderson y Goodman, bajo muestreo complejo. También se realizarán estudios sobre los estimadores clásicos que incluyen variables explicativas.

6. BIBLIOGRAFÍA

Anderson, T.W. y Goodman, L.A. (1957). Statistical Inference about Markov Chains. *Annals of Mathematical Statistics*, 28, 89-100.

Servy, E., Hachuel, L, Boggio, G., Cuesta, C. (1999). Modelos de Transición para el estudio del evento ocupado-desocupado en el Gran Buenos Aires. *Actas del CLATSE IV*, Mendoza, Argentina

Servy, E., Hachuel, L, Boggio, G., Cuesta, C. (2000). Adaptación de la Metodología de Grizzle, Starmer y Koch a muestras de panel rotativas. *Actas de la SAE*.

Servy, E., Hachuel, L, Wojdyla, D. (1998). *Análisis de Tablas de Contingencia para Muestras de Tamaño Complejo*. Informe Técnico INDEC - Escuela de Estadística.



APÉNDICE

Tabla 1. Porcentajes de Sesgos relativos (en valores absolutos) del estimador de la probabilidad de transición y del estimador de su variancia, para un tamaño de muestra igual a 15.

| | | correlaciones intra-conglomerado | | | | | | |
|----|---|----------------------------------|------|-------|------|--------|------|-------|
| | | baja | | media | | alta | | |
| | | sesrel | | var | | sesrel | | |
| n | k | p(0 0) | | | | | | |
| 15 | 2 | 0.5 | 0.58 | 22.91 | 0.02 | 38.46 | 0.39 | 55.14 |
| | | 0.55 | 0.28 | 20.75 | 0.11 | 45.47 | 0.42 | 58.40 |
| | | 0.6 | 0.49 | 20.20 | 1.04 | 40.64 | 0.72 | 53.90 |
| | | 0.65 | 0.21 | 21.77 | 0.01 | 38.63 | 0.41 | 52.89 |
| | | 0.7 | | | 0.23 | 48.46 | | |
| | 3 | 0.5 | 0.14 | 21.33 | 0.05 | 50.65 | 1.50 | 66.75 |
| | | 0.55 | 0.36 | 32.16 | 0.33 | 52.97 | 0.36 | 67.26 |
| | | 0.6 | 0.10 | 37.02 | 1.25 | 51.66 | 0.59 | 65.88 |
| | | 0.65 | | | | | 0.44 | 61.87 |
| | 4 | 0.5 | 0.21 | 16.28 | 0.75 | 56.90 | 1.04 | 72.73 |
| | | 0.55 | 0.76 | 19.38 | 0.07 | 50.91 | 0.01 | 68.44 |
| | | 0.6 | | | 0.77 | 64.10 | 0.43 | 73.17 |
| | 5 | 0.5 | 0.05 | 25.63 | 0.07 | 53.45 | 0.29 | 77.19 |
| | | 0.55 | 0.19 | 33.79 | 0.32 | 54.38 | 0.34 | 72.23 |

n: tamaño de la muestra
k: tamaño de los conglomerados



Tabla 2. Porcentajes de Sesgos relativos (en valores absolutos) del estimador de la probabilidad de transición y del estimador de su variancia, para un tamaño de muestra igual a 30.

| | | correlaciones intra-conglomerado | | | | | | |
|----|---|----------------------------------|------|--------|------|--------|------|-------|
| | | baja | | media | | alta | | |
| | | sesrel | var | sesrel | var | sesrel | var | |
| n | k | p(0 0) | | | | | | |
| 30 | 2 | 0.5 | 0.18 | 23.75 | 0.71 | 40.72 | 0.37 | 54.00 |
| | | 0.55 | 0.65 | 21.83 | 0.44 | 42.13 | 0.43 | 55.42 |
| | | 0.6 | 0.46 | 17.80 | 0.36 | 41.77 | 0.43 | 51.26 |
| | | 0.65 | 0.28 | 17.04 | 0.65 | 38.70 | 0.62 | 56.16 |
| | | 0.7 | | | 0.21 | 45.35 | | |
| | 3 | 0.5 | 0.62 | 15.88 | 0.93 | 46.09 | 0.64 | 65.07 |
| | | 0.55 | 0.58 | 31.01 | 0.22 | 47.76 | 1.06 | 65.07 |
| | | 0.6 | 0.24 | 37.47 | 0.48 | 55.51 | 0.31 | 66.23 |
| | | 0.65 | | | | | 1.08 | 64.10 |
| | 4 | 0.5 | 0.52 | 12.90 | 0.31 | 53.63 | 0.17 | 73.51 |
| | | 0.55 | 0.42 | 23.12 | 0.50 | 49.00 | 0.08 | 66.72 |
| | | 0.6 | | | 0.14 | 64.77 | 0.01 | 71.46 |
| | 5 | 0.5 | 0.03 | 15.27 | 0.42 | 52.30 | 0.05 | 75.94 |
| | | 0.55 | 0.06 | 32.58 | 0.64 | 54.16 | 0.40 | 68.63 |

n: tamaño de la muestra
k: tamaño de los conglomerados



Tabla 3. Porcentajes de Sesgos relativos (en valores absolutos) del estimador de la probabilidad de transición y del estimador de su variancia, para un tamaño de muestra igual a 70.

| | | correlaciones intra-conglomerado | | | | | | |
|----|---|----------------------------------|------|--------|------|--------|------|-------|
| | | baja | | media | | alta | | |
| | | sesrel | | sesrel | | sesrel | | |
| | | var | | var | | var | | |
| n | k | p(0 0) | | | | | | |
| | | % | | | | | | |
| 70 | 2 | 0.5 | 0.16 | 16.68 | 0.05 | 39.13 | 0.28 | 57.39 |
| | | 0.55 | 0.11 | 14.53 | 0.19 | 41.25 | 0.13 | 54.59 |
| | | 0.6 | 0.06 | 15.92 | 0.16 | 40.50 | 0.33 | 52.75 |
| | | 0.65 | 0.27 | 15.57 | 0.19 | 40.75 | 0.22 | 50.98 |
| | | 0.7 | | | 0.16 | 45.48 | | |
| | 3 | 0.5 | 0.11 | 23.72 | 0.20 | 45.77 | 0.28 | 66.66 |
| | | 0.55 | 0.11 | 31.90 | 0.29 | 50.30 | 0.01 | 67.18 |
| | | 0.6 | 0.16 | 34.35 | 0.06 | 46.55 | 0.69 | 66.82 |
| | | 0.65 | | | | | 0.30 | 64.32 |
| | 4 | 0.5 | 0.39 | 21.83 | 0.20 | 50.22 | 0.01 | 73.19 |
| | | 0.55 | 0.07 | 16.98 | 0.33 | 55.27 | 0.22 | 66.14 |
| | | 0.6 | | | 0.50 | 64.68 | 0.13 | 71.70 |
| | 5 | 0.5 | 0.01 | 19.99 | 0.40 | 55.79 | 0.26 | 76.61 |
| | | 0.55 | 0.16 | 28.99 | 0.01 | 56.83 | 0.12 | 70.00 |

n: tamaño de la muestra
k: tamaño de los conglomerados



Tabla 4. Porcentajes de Sesgos relativos (en valores absolutos) del estimador de la probabilidad de transición y del estimador de su variancia, para un tamaño de muestra igual a 100.

| | | correlaciones intra-conglomerado | | | | | | |
|-----|---|----------------------------------|------|--------|------|--------|------|-------|
| | | baja | | media | | alta | | |
| | | sesrel | | sesrel | | sesrel | | |
| | | var | | var | | var | | |
| n | k | p(0 0) | | | | | | |
| | | % | | | | | | |
| 100 | 2 | 0.5 | 0.22 | 20.88 | 0.14 | 42.48 | 0.02 | 56.67 |
| | | 0.55 | 0.10 | 21.97 | 0.26 | 39.83 | 0.45 | 50.19 |
| | | 0.6 | 0.28 | 16.14 | 0.03 | 37.28 | 0.13 | 53.13 |
| | | 0.65 | 0.27 | 13.47 | 0.09 | 39.80 | 0.16 | 56.26 |
| | | 0.7 | | | 0.03 | 50.76 | | |
| | 3 | 0.5 | 0.24 | 21.60 | 0.00 | 51.34 | 0.06 | 67.01 |
| | | 0.55 | 0.13 | 31.77 | 0.46 | 50.21 | 0.07 | 65.91 |
| | | 0.6 | 0.06 | 34.84 | 0.08 | 52.72 | 0.18 | 65.16 |
| | | 0.65 | | | | | 0.14 | 64.42 |
| | 4 | 0.5 | 0.07 | 27.53 | 0.01 | 54.52 | 0.46 | 71.41 |
| | | 0.55 | 0.29 | 15.02 | 0.16 | 56.29 | 0.28 | 65.32 |
| | | 0.6 | | | 0.07 | 64.60 | 0.00 | 68.30 |
| | 5 | 0.5 | 0.04 | 25.63 | 0.04 | 58.61 | 0.44 | 74.77 |
| | | 0.55 | 0.25 | 27.00 | 0.20 | 57.86 | 0.07 | 71.53 |

n: tamaño de la muestra
k: tamaño de los conglomerados



Tabla 5. Porcentajes de Sesgos relativos (en valores absolutos) del estimador de la probabilidad de transición y del estimador de su variancia, para un tamaño de muestra igual a 500.

| | | correlaciones intra-conglomerado | | | | | | |
|-----|---|----------------------------------|------|--------|------|--------|------|-------|
| | | baja | | media | | alta | | |
| | | sesrel | var | sesrel | var | sesrel | var | |
| n | k | p(0 0) | | | | | | |
| 500 | 2 | 0.5 | 0.05 | 23.21 | 0.02 | 39.84 | 0.09 | 53.87 |
| | | 0.55 | 0.12 | 23.56 | 0.03 | 61.08 | 0.23 | 54.49 |
| | | 0.6 | 0.14 | 27.24 | 0.02 | 37.38 | 0.18 | 58.05 |
| | | 0.65 | 0.01 | 25.04 | 0.05 | 42.72 | 0.09 | 54.05 |
| | | 0.7 | | | 0.00 | 49.85 | | |
| | 3 | 0.5 | 0.00 | 22.02 | 0.06 | 49.47 | 0.10 | 66.83 |
| | | 0.55 | 0.16 | 28.42 | 0.02 | 50.47 | 0.07 | 67.33 |
| | | 0.6 | 0.17 | 39.04 | 0.02 | 52.59 | 0.06 | 65.20 |
| | | 0.65 | | | | | 0.18 | 61.12 |
| | 4 | 0.5 | 0.02 | 20.26 | 0.00 | 54.08 | 0.11 | 73.80 |
| | | 0.55 | 0.09 | 10.62 | 0.18 | 54.19 | 0.16 | 67.53 |
| | | 0.6 | | | 0.06 | 63.38 | 0.03 | 70.48 |
| | 5 | 0.5 | 0.11 | 15.42 | 0.02 | 56.54 | 0.00 | 75.50 |
| | | 0.55 | 0.05 | 30.12 | 0.03 | 55.30 | 0.00 | 71.57 |

n: tamaño de la muestra
k: tamaño de los conglomerados



Tabla 6. Porcentajes de Sesgos relativos (en valores absolutos) del estimador de la probabilidad de transición y del estimador de su variancia, para un tamaño de muestra igual a 1000.

| | | correlaciones intra-conglomerado | | | | | | |
|------|---|----------------------------------|------|--------|------|--------|------|-------|
| | | baja | | media | | alta | | |
| | | sesrel | | sesrel | | sesrel | | |
| | | var | | var | | var | | |
| n | k | P(0 0) | | | | | | |
| | | % | | | | | | |
| 1000 | 2 | 0.5 | 0.06 | 15.64 | 0.07 | 37.25 | 0.21 | 54.85 |
| | | 0.55 | 0.07 | 22.79 | 0.02 | 37.43 | 0.05 | 53.89 |
| | | 0.6 | 0.03 | 14.50 | 0.06 | 39.17 | 0.15 | 53.21 |
| | | 0.65 | 0.02 | 18.77 | 0.10 | 40.97 | 0.05 | 55.23 |
| | | 0.7 | | | 0.02 | 50.83 | | |
| | 3 | 0.5 | 0.06 | 19.83 | 0.06 | 48.70 | 0.00 | 64.18 |
| | | 0.55 | 0.04 | 28.84 | 0.01 | 51.99 | 0.15 | 65.86 |
| | | 0.6 | 0.07 | 32.70 | 0.10 | 56.73 | 0.15 | 66.44 |
| | | 0.65 | | | | | 0.10 | 63.07 |
| | 4 | 0.5 | 0.03 | 18.28 | 0.12 | 50.36 | 0.15 | 73.46 |
| | | 0.55 | 0.04 | 23.58 | 0.08 | 49.45 | 0.03 | 68.07 |
| | | 0.6 | | | 0.11 | 64.73 | 0.04 | 72.65 |
| | 5 | 0.5 | 0.04 | 22.09 | 0.07 | 56.53 | 0.08 | 76.17 |
| | | 0.55 | 0.01 | 32.44 | 0.07 | 55.24 | 0.03 | 72.53 |

n: tamaño de la muestra
k: tamaño de los conglomerados