Revista Colombiana de Estadística
Nºs 19-20, junio-diciembre 1989

UNA DISTRIBUCION ASINTOTICA PARA ESTIMADORES MAXIMO-VEROSIMILES DE TAMAÑOS DE DOMINIOS DE SITUACIONES DE MARCOS MULTIPLES

David Ospina Botero

Profesor Asociado
Universidad Nacional de Col.

Abstract. The likelihood of domain sizes in overlapped frame situations is approximated with a multinomial. The variance-covariance matrix for the joint distributions of MLE estimators is obtained based on the asymptotic behavior of these MLE's. The algorithm presented by Cooke (Comm. Statist. B-Simulation Comput. 12(1983) 291-305) is applied to simulate some estimates of domain sizes and variances of those estimates.

Resumen. La función de verosimilitud de los tamaños de dominios en el caso de marcos superpuestos se aproxima con una multinomial. La matriz de varianza-covarianza para las distribuciones conjuntas de los estimadores máximo-verosímiles se obtiene en base al comportamiento asintótico que éstos tienen. El algoritmo presentado por Cooke (Comm. Statist. B-simulation Compt. 12(1982) 291-305) se aplica para simular algunas estimaciones de tamaños de dominios y varianzas de esas estimaciones.

Palabras y frases claves:

Distribución Binomial,

estimación Ji-cuadrado mínimo,

Distribución multinomial,
marcos muestrales traslapados,
programación geométrica subrogada,
matriz de varianza-covarianza.

1. Introducción.

La estimación del número de unidades diferentes en los dominios creados por la intersección de marcos muestrales se ha estudiado usando tres aproximaciones distintas: Ji-Cuadrado Mínimo Modificado (Bryant y King, 1960), Máxima-verosimilitud (Williams, 1957) y Marcos Múltiples (Cochran, 1967). La varianza para muestras grandes de las estimaciones para la Ji-cuadrado y los Marcos Múltiples fueron obtenidas por Cochran y Cooke (1975). Fuller y Burmeister (1972) usaron el Método de Marcos Múltiples para derivar estimadores y varianzas de los estimadores de tamaños de dominios cuando se tienen en cuenta las unidades duplicadas. También obtuvieron una forma general para la varianza del estimador Máximo-Verosímil (EMV) cuando se consideran dos marcos y se usa el modelo hipergeométrico. El artículo de Bryant y King analizó el caso de tres marcos pero no comparó varianzas. Cochran y Toro (1973) aplicaron CONVEX, un algoritmo de Harthley y Hocking (1963) que fué programado por LaMotte y Oxspring (1971), para obtener las estimaciones máximo-verosímiles, en los casos de dos y tres marcos. La Programación Geométrica (Alldredge y Amstrong, (1974) y la Programación Geométrica Subrogada (PGS) (Cooke, 1974) son técnicas que se han utilizado para resolver el mismo problema de una manera eficiente. Cooke (1983) presentó una solución general para las estimaciones máximo-verosimiles de proporciones multinomiales sujetas a un número dado de restricciones de igualdades o desigualdades lineales. Este fué un paso importante debido a que las estimaciones máximo-verosímiles de los tamaños de dominios se pueden obtener sin importar el número de marcos que se intersectan. Sin embargo el problema de estimar la matriz general de varianza-covarianza no ha sido tratado. La ausencia de una forma cerrada para la densidad conjunta de los estimadores de tamaños de dominios hace la solución de este problema un poco difícil. El mayor número de marcos reportado para analizar es de tres por Bryant y King (1960) usando el Criterio de la Ji-Cuadrado Mínima Modificada, y por Cochran, usando la Aproximación de Marcos Múltiples.

En este artículo se presenta una aproximación general al problema de la estimación con base en la derivación de una función de verosimilitud aproximada para los tamaños de los dominios.

2. Función de verosimilitud aproximada para los estimadores de tamaños de dominios cuando existen marcos múltiples.

2.1. Caso de dos marcos.

Cuando dos marcos se combinan para proporcionar una cobertura completa de la población, pueden existir hasta tres dominios. Unas unidades únicamente en el marco 1, otras en el marco 2 y las restantes en ambos marcos. El proceso de muestreo consiste en seleccionar n_1 unidades al azar de N_1 en el marco 1 y n_2 unidades al azar de N_2 en el marco 2. De las n_i unidades del marco i, x_i están en el marco i únicamente y $n_i - x_i$ en ambos marcos. La función de verosimilitud exacta para el número de unidades muestrales en cada dominio es hipergeométrica cuando el muestreo se lleva a cabo sin reemplazamiento. Sin embargo, para muestras grandes de poblaciones grandes la hipergeométrica se aproxima bien a la distribución binomial. Para aplicar estrictamente la distribución binomial (y posteriormente la multinomial) uno debe identificar las unidades que provienen del dominio traslapado de los marcos. Esta situación se asocia en general con marcos representados por listas cuyos elementos son personas. Cuando los elementos seleccionados son entrevistados o responden a cuestionarios enviados por correo uno está en condiciones de preguntar ciertas características que los clasificarán en el dominio correspondiente.

La función de verosimilitud para el número de unidades, x_i, provenientes del marco i y que no pertenecen al dominio traslapado es aproximada por

$$f(x_i, n_i, \theta) = \binom{n_i}{x_i}\left(\frac{N_i-\theta}{N_i}\right)^{x_i}\left(\frac{\theta}{N_i}\right)^{n_i-x_i} \tag{2.1.1}$$

$$x_i = 0,1,\ldots,n_i; \quad 0 \leqslant \theta \leqslant N_i, \quad i = 1,2$$

donde

N_i = el número de unidades en el marco i

θ = el número de unidades en el dominio traslapado

n_i = tamaño de muestra del marco i.

Debido a que el muestreo realizado en un marco es independiente del muestreo realizado en el otro, la función de verosimilitud combinada es el producto de las funciones de verosimilitud individuales

$$L_1(\theta) = f(x_1, n_1; \theta) f(x_2, n_2; \theta) \tag{2.1.2}$$

El siguiente teorema debe considerarse para la varianza de la distribución muestral de los EMV de θ, θ^*:

TEOREMA 2.1. *Si la función de verosimilitud de* θ *está dada por*

$$L_1(\theta) = f(x_1,n_1;\theta)f(x_2,n_2;\theta)$$

donde $f(x_1,n_1;\theta)$ *y* $f(x_2,n_2;\theta)$ *son distribuciones binomiales con parámetros* n_1,θ *y* n_2,θ *respectivamente, entonces la cota inferior de Cramer-Rao para la varianza de* θ^*, $V[\theta^*]$, *donde* θ^* *es el EMV de* θ, *viene dada por*

$$V[\theta^*] \geqslant \frac{\theta(N_1-\theta)(N_2-\theta)}{n_1(N_2-\theta)+n_2(N_1-\theta)} \qquad (2.1.3)$$

Demostración. Si la cota inferior de Cramer Rao para $V[\theta^*]$ se escribe en términos de $L_1(\theta)$ como

$$V[\theta^*] \geqslant \frac{1}{-\Sigma\{\frac{\partial^2}{\partial\theta^2}\log L_1(\theta)\}} \qquad (2.1.4)$$

la demostración es un simple tema de cálculo elemental y operaciones algebraicas.

2.2. Aproximación de la función de verosimilitud.

En esta sección la función de verosimilitud $L_1(\theta)$, (2.1.2) se aproximará a la distribución de probabilidad multinomial con parámetros n $(= n_1+n_2)$ y θ. Mediante esta aproximación se podrán usar algunos resultados teóricos importantes concernientes a la distribución asintótica de EMV. Por otra parte el uso de la distribución multinomial hace posible una generalización simple del proceso a más de dos marcos muestrales.

La forma general de la distribución multinomial para el caso de dos marcos es

$$f(x_1,x_2;n,\theta) = \binom{n}{x_1,x_2} p_1^{x_1} p_2^{x_2} (1-p_1-p_2)^{n-x_1-x_2} \qquad (2.2.1)$$

donde:

$n = n_1+n_2$ (tamaño de muestra total)

x_i = Número observado de unidades en la muestra que pertenecen al dominio no traslapado del marco i.

p_i = Probabilidad de que cualquier unidad pertenezca al dominio no traslapado del marco i.

= P_r [Unidad se tomó del marco i] P_r [Unidad pertenece al dominio no traslapado i / Unidad se tomó del marco i]

$$= \frac{n_i}{n} \cdot \frac{(N_i - \theta)}{N_i} \qquad (2.2.2)$$

Por tanto (2.2.1) puede escribirse

$$f(x_1,x_2;n,\theta) = \binom{n}{x_1,x_2} \left[\frac{n_1(N_1-\theta)}{nN_1}\right]^{x_1} \left[\frac{n_2(N_2-\theta)}{nN_2}\right]^{x_2} \left[\frac{\theta}{n}\left(\frac{n_1}{N_1}+\frac{n_2}{N_2}\right)\right]^{n-x_1-x_2} \qquad (2.2.3)$$

y la función de verosimilitud puede expresarse como:

$$L_2(\theta) = K_2(N_1-\theta)^{x_1}(N_2-\theta)^{x_2}\,\theta^{n-x_1-x_2} \qquad (2.2.4)$$

donde K_2 no depende de θ.

La estimación máximo-verosímil de θ, θ^*, se obtiene solucionando la ecuación cuadrática:

$$n\theta^{*2} - [n(N_1+N_2)-x_1N_1-x_2N_2]\theta^* + N_1N_2(n-x_1-x_2) = 0 \qquad (2.2.5)$$

Puede mostrarse que las soluciones a (2.2.5) son ambas reales y que la menor de estas raíces cae en el intervalo $0 \leqslant \theta \leqslant \min(N_1,N_2)$. Por tanto esta raíz se toma como la estimación máximo-verosímil de θ.

La estimación obtenida usando (2.1.2) será la misma que la obtenida usando (2.2.4) ya que los términos que involucran θ son los mismos en ambos casos. La varianza para nuestras grandes es nuevamente

$$V[\theta^*] \geqslant \frac{\theta(N_1-\theta)(N_2-\theta)}{n_1(N_2-\theta)+n_2(N_1-\theta)}$$

Usando (2.2.4) en vez de (2.1.2) para obtener el EMV de θ y su varianza asintótica nos permite generalizar el proceso a más de 2 marcos gracias a la bien definida estructura de la distribución multinomial.

2,3. Caso general: *M* marcos.

Para M marcos, el número máximo de dominios está dado por

$$T = \binom{M}{1} + \binom{M}{2} + \ldots + \binom{M}{M} = 2^M - 1 \qquad (2.3.1)$$

Ya que los primeros M tamaños de dominios pueden estimarse una vez los otros se conocen, el número máximo de tamaños de dominio (parámetros) a estimarse es

$$S = T - M = 2^M - M - 1. \tag{2.3.2}$$

Notación. Los subíndices para los valores observados y parámetros identifican los marcos que se traslapan en cada caso.

θ_i = Número de elementos que pertenecen al marco i pero no pertenecen a ningún traslape entre marcos (tamaño de dominio no traslapado en marco i).

θ_{ij} = Número de elementos que pertenecen al traslape entre marcos i y j pero no pertenecen a traslapes de orden superior (tamaño de dominio traslapado ij).

$\vdots$

$\theta_{ij\cdots R}$ = Número de elementos que pertenecen al traslape entre los marcos $i, j, \ldots,$ y R pero no pertenecen a tralapes de orden superior (tamaño de dominio traslapado $ij\ldots R$).

$\vdots$

$\theta_{123\ldots M}$ = Número de elementos que pertenecen al traslape entre todos los marcos (tamaño de dominio traslapado $123\ldots M$).

where $i, j, \ldots, R$ y M son enteros positivos con

$$i < j < \ldots < R < \ldots < M \tag{2.3.3}$$

x_i = Número observado de unidades que pertenecen al dominio i

x_{ij} = Número observado de unidades que pertenecen al dominio ij

$\vdots$

$x_{ij\ldots R}$ = Número observado de unidades que pertenecen al dominio $ij\ldots R$.

$\vdots$

$x_{123\ldots M}$ = Número observado de unidades que pertenecen al dominio $123\ldots M$

(2.3.4)

El número de unidades en el marco i se denotará por $N_i, i = 1,2,\ldots,M$.

El tamaño muestral del marco i es n_i y las fracciones muestrales son

$$h_i = \frac{n_i}{N_i} \tag{2.3.5}$$

El tamaño de la muestra total es

$$\begin{aligned} n &= n_1 + n_2 + \ldots + n_M \\ &= \sum_{i=1}^{M} n_i \\ &= \sum_{i=1}^{M} x_i + \sum_{i<j}^{M} x_{ij} + \ldots + x_{123\ldots M} \end{aligned} \tag{2.3.6}$$

Los S tamaños de dominios a estimar son los correspondientes a la intersección de dos marcos como mínimo. Por tanto

$$\underline{\theta}_I = (\theta_{12}, \theta_{13}, \ldots, \theta_{1M}, \theta_{123}, \theta_{124}, \ldots, \theta_{123\ldots M})^T \tag{2.3.7}$$

la función de verosinilitud general puede aproximarse por

$$L(\underline{\theta}_I) = K \prod_{i=1}^{M} p_i^{x_i} \prod_{i<j}^{M} p_{ij}^{x_{ij}} \ldots \prod_{i<j<\ldots<R}^{M} p_{ij\ldots R}^{x_{12\ldots R}} \ldots p_{123\ldots M}^{x_{123\ldots M}} \tag{2.3.8}$$

donde

$$K = \frac{n!}{\left(\prod_{i=1}^{M} x_i!\right)\left(\prod_{i<j}^{M} x_{ij}!\right)\ldots\left(x_{123\ldots M}!\right)} \,. \tag{2.3.9}$$

y

$$\sum_{i=1}^{M} p_i + \sum_{i<j}^{M} p_{ij} \cdots \sum_{i<j<\dots<R}^{M} p_{ij\dots R} + p_{123\dots M} = 1 \qquad (2.3.10)$$

Siguiendo el mismo procedimiento usado en la Sección 2.2 los p son:

$$p_i = \frac{n_i}{n}\cdot\frac{\theta_i}{N_i} = \frac{h_i}{n}\cdot\theta_i$$

$$p_{ij} = \frac{n_i}{n}\cdot\frac{\theta_{ij}}{N_i} + \frac{n_j}{n}\cdot\frac{\theta_{ij}}{N_j}$$

$$= \frac{(h_i+h_j)}{n}\cdot\theta_{ij}$$

$$\vdots$$

$$p_{ij\dots R} = \frac{n_i}{n}\cdot\frac{\theta_{ij..R}}{N_i} + \frac{n_j}{n}\cdot\frac{\theta_{ij..R}}{N_j} +\dots+ \frac{n_R}{n}\cdot\frac{\theta_{ij..R}}{N_R}$$

$$= \frac{(h_i+h_j+\dots+h_R)}{n}\cdot\theta_{ij\dots R}$$

$$\vdots$$

$$p_{123\dots M} = \frac{\left(\sum_{i=1}^{M} h_i\right)}{n}\cdot\theta_{123\dots M} \qquad (2.3.11)$$

Por tanto:

$$L(\underline{\theta}_I) = K \prod_{i=1}^{M} \left[\frac{h_i\theta_i}{n}\right]^{x_i} \prod_{i<j}^{M} \left[\frac{(h_i+h_j)\theta_{ij}}{n}\right]^{x_{ij}}$$

$$\dots \prod_{i<j<\dots<R}^{M} \left[\frac{(h_i+h_j+\dots+h_R)\theta_{ij\dots R}}{n}\right]^{x_{ij\dots R}}$$

$$\dots \left[\frac{\sum_{i=1}^{M} h_i\cdot\theta_{123\dots M}}{n}\right]^{x_{123\dots M}} \qquad (2.3.12)$$

donde K se ha definido por (2.3.9).

3. Distribución asintótica de los estimadores maximo-verosimiles de los tamaños de dominios.

La función de verosimilitud dada por (2.3.12) puede escribirse

$$L(\underline{\theta}_I) = K_1 \prod_{i=1}^{T} [f_i(\underline{\theta}_I)]^{x_i} \tag{3.1}$$

$$(x_{M+1} = x_{12}, x_{M+2} = x_{13}, \ldots, x_T = x_{123\ldots M})$$

donde

$$\underline{f}(\underline{\theta}_I) = \left[\frac{h_1}{n}\theta_1, \frac{h_2}{n}\theta_2, \ldots, \frac{h_M}{n}\theta_M, \frac{(h_1+h_2)}{n}\theta_{M+1}, \ldots, \frac{\sum_{i=1}^{M} h_i}{n}\theta_T\right]$$

$$(\theta_{M+1} = \theta_{12}, \theta_{M+2} = \theta_{13}, \ldots, \theta_T = \theta_{123\ldots M})$$

El EMV de $\underline{\theta}_I$, $\underline{\theta}_I^*$, puede escribirse como

$$\underline{\theta}_I^* = (\theta_{M+1}^*, \theta_{M+2}^*, \ldots, \theta_T^*)^T, \tag{3.2}$$

θ_i^* siendo el EMV de θ_i, $i = M+1, M+2, \ldots, T$.

La distribución asintótica de $\underline{\theta}_I^*$, está dada por

$$L[\sqrt{n}(\underline{\theta}_I^* - \underline{\theta}_I)] \to N[\underline{0}, (A^T A)^{-1}], \tag{3.3}$$

$A^T A$ siendo la matriz de información de Fisher (Bishop, Fienberg y Holland, 1975).

En base al supuesto de muestras grandes, (3.3) puede escribirse como

$$\sqrt{n}(\underline{\theta}_I^* - \underline{\theta}_I) \sim N[0, (A^T A)^{-1}] \tag{3.4}$$

El elemento (i,j) de la matriz A^TA es

$$= \sum_{k=1}^{T} \frac{\partial \log[f_k(\underline{\theta}_I)]}{\partial \theta_i} \cdot \frac{\partial \log[f_k(\underline{\theta}_I)]}{\partial \theta_j} \cdot f_k(\underline{\theta}_I) \qquad (3.5)$$

La matriz de varianza-covarianza asintótica de $\underline{\theta}^*$ puede identificarse ya como

$$\text{COV}(\underline{\theta}_I^*) = \frac{1}{n}(A^TA)^{-1} \qquad (3.6)$$

Un estimador de (3.6) es

$$\text{COV}^*(\underline{\theta}_I^*) = \frac{1}{n}[A(\underline{\theta}_I^*)^T A(\underline{\theta}_I^*)]^{-1},$$ (Bishop, Fienberg y Holland, 1975).

donde $[A(\underline{\theta}_I^*)^T A(\underline{\theta}_I^*)]$ es una matriz cuyos elementos (i,j) se calculan de acuerdo con (3.5) pero reemplazando $\underline{\theta}_I$ por su EMV $\underline{\theta}_I^*$.

Para determinar la matriz de varianza-covarianza de $\underline{\theta}_I^*$, se hace

$$B = n(A^TA)$$

Por tanto:

$$\text{COV}[\theta_I^*] = B^{-1}.$$

Los elementos (i,j) de B se definen como

$$= \sum_{k=1}^{T} \frac{\partial \log(g_k(\underline{\theta}_I))}{\partial \theta_i} \cdot \frac{\partial \log(g_k(\underline{\theta}_I))}{\partial \theta_j} \cdot g_k(\underline{\theta}_I) \qquad (3.8)$$

donde

$$g_k(\underline{\theta}_I) = n f_k(\underline{\theta}_I), \quad k = 1,2,\ldots,T \qquad (3.9)$$

Los elementos de la diagonal de la matriz B, que se asocian con las varianzas de los dominios traslapados están dados por

$$\frac{1}{\theta_{ij..R}}\left\{\frac{h_i(\theta_i+\theta_{ij..R})}{\theta_i}+\frac{h_j(\theta_j+\theta_{ij..R})}{\theta_j}+\ldots+\frac{h_R(\theta_R+\theta_{ij..R})}{\theta_R}\right\} \tag{3.10}$$

donde el traslape se ha generado por los marcos $i,j,\ldots,R$. Los elementos restantes, no pertenecientes a la diagonal, se representan por el término general

$$\sum_i \frac{h_i}{\theta_i} \tag{3.11}$$

donde i toma los valores correspondientes a los marcos comunes a los dominios traslapados en cada caso.

4. Proceso de simulación.

Con el fin de ilustrar los resultados teóricos se llevaron a caso simulaciones para los casos de dos y tres marcos. El proceso de simulación consistió de varias etapas:

(1) Se generaron muestras aleatorias para las distribuciones Binomial (dos marcos) y Multinomial (tres marcos) haciendo uso de las subrutinas GGBN (Binomial Random Deviate Generator) y GGMTN (Multinomial Random Deviate Generator) del IMSL (Institute of Mathematical Statistics Library, 1980).

(2) Se registró el número de unidades para cada dominio en la muestra combinada.

(3) Se escribieron subrutinas para obtener las estimaciones máximo-verosímiles en los casos de dos y tres marcos. Estas subrutinas están basadas en los algoritmos de la Programación Geométrica Subrogada (PGS) (Cooke, 1983). El método PGS proporciona una solución general en términos de los multiplicado-

res subrogados para la estimación máximo-verosímil de proporciones multinomiales con restricciones.

(4) Una vez se generaron todas las estimaciones, se analizaron las distribuciones "empíricas" de estas estimaciones. Un test dado por Ryan y Joiner (1982) que es esencialmente equivalente al test de Shapiro-Wilk (1965) se utilizó para probar el supuesto de normalidad en cada caso para niveles de significancia diferentes. El único tamaño de dominio que fue necesario estimar en el caso de dos marcos fue θ_{12}, número de unidades en el traslape de los dos marcos, ya que θ_1 y θ_2 se pueden estimar por sustracción una vez que se conoce θ_{12}. En el caso de tres marcos se necesitó estimar cuatro tamaños de dominios θ_{12}, θ_{13}, θ_{23} y θ_{123}. Las estimaciones de los tamaños de dominios θ_1,θ_2 y θ_3 (no traslapados) se obtuvieron por sustracción una vez los tamaños de los dominios traslapados se han estimado.

Para ser consistentes con la aproximación multinomial, se tomaron tamaños de marcos suficientemente grandes (el mínimo tamaño de marco considerado fué 10.000). Los tamaños de los dominios tralapados se consideran como mínimo de un 5% pero no mayor del 90% de ninguno de los marcos.

Con el fin de considerar afijaciones proporcionales y no proporcionales, se tomaron tamaños de muestras diferentes para cada simulación. El menor tamaño de muestra considerado fue de 50 para lograr una aproximación respetable al supuesto de muestras grandes. Se tomaron 75 muestras de cada marco con el objeto de poder utilizar las tablas presentadas por Ryan y Joiner (1987) para decidir con respecto a la aproximación normal para la distribución de los EMV de los tamaños de dominios traslapados en cada caso. El valor observado de la estadístico R_p (coeficiente de correlación del cuadro de probabilidad) se compara con los valores teóricos dados por las tablas para diferentes valores de α (nivel de significancia). Si él caía por debajo del valor crítico apropiado, el supuesto de normalidad quedaba en duda. Una vez la normalidad se verificaba, el número de muestras se incrementaba a 1000 para ilustrar la convergencia del estadístico de prueba R_p a 1. Una "correlación límite" de 1 equivale al supuesto de normalidad (Ryan y Joiner). Los valores críticos aproximados de R_p para M (número de muestras) = 75 son .9865 si α = .10, .9835 si α = .05 y .9757 si α = .01. Los resultados fueron altamente satisfactorios ya que solamente en 9 de los 49 casos considerados el valor observado del estadístico fue ligeramente menor que el valor crítico para α = .10. Sin embargo, cuando el número de muestras se incrementó a 1000 el R_p observado fue muy cercano a 1 (el valor mínimo fue .994). Estos resultados concuerdan con la conclusión de normalidad. Algunos resultados se muestran en las tablas 4.1 a 4.7. Las tablas dan la media de las estimaciones máximo-verosímiles para los tamaños de dominio, la media de θ^*_{12}, y para la varianza de los EMV de los tamaños de dominio $\hat{\sigma}^2_{\theta^*}$. También se dan los sesgos relativos, definidos en términos de la diferencia entre la media de las estimaciones y el parámetro real dividida por los valores reales, para los θ^* y las $\hat{\sigma}^2_{\theta^*}$. Un signo "*" en la última columna indica que la hipótesis de normalidad se rechazó al nivel del 10%.

*

TABLA 4.1.

Dos marcos - marcos y tamaños de muestras iguales.

$N_1 = N_2 = 10000; \quad n_1 = n_2 = 100; \quad M = 75$

valor verdadero θ_{12}	media de θ^*_{12}	sesgo relativo	$\sigma^2_{\theta^*_{12}}$	media de $\hat{\sigma}^2_{\theta^*_{12}}$	sesgo relativo	R_p	α		
							.10	.05	.01
1000	1000	.0000	45000	44757	-.0054	.990	-	-	-
3000	3028	.0093	105000	105217	.0021	.994	-	-	-
5000	4971	-.0058	125000	124410	-.0047	.997	-	-	-
7000	7029	.0041	105000	103942	-.0101	.996	-	-	-
9000	9031	.0034	45000	43549	-.0322	.998	-	-	-

TABLA 4.2.

Dos marcos - marcos desiguales - afijación proporcional

$N_1 = 10000$; $N_2 = 20000$; $n_1 = 50$; $n_2 = 100$; $M = 75$

valor verdadero θ^*_{12}	media de θ^*_{12}	sesgo relativo	$\sigma^2_{\theta^*_{12}}$	media de $\hat{\sigma}^2_{\theta^*_{12}}$	sesgo relativo	R_p	α		
							.10	.05	.01
10000	979	-.0210	92432	89887	-.0275	.990	-	-	-
3000	2975	-.0083	230323	226812	-.0152	.994	-	-	-
5000	4914	-.0172	300000	295472	-.0151	.994	-	-	-
7000	6848	-.0217	287368	287319	-.0002	.995	-	-	-
9000	8878	-.0136	152308	160291	.0524	.983	-	-	-

TABLA 4.3.

Dos marcos - marcos desiguales - afijación no-propocional

$N_1 = 10000$; $N_2 = 20000$; $n_1 = 300$; $n_2 = 200$; $M = 75$

valor verdadero θ_{12}	media de θ^*_{12}	sesgo relativo	$\sigma^2_{\theta^*_{12}}$	media de $\hat{\sigma}^2_{\theta^*_{12}}$	sesgo relativo	R_p	α .10	.05	.01
1000	1054	.0530	26000	27906	.0733	.995	-	-	-
3000	3095	.0317	62160	64646	.0400	.994	-	-	-
5000	5113	.0226	76087	77397	.0172	.986	*	-	-
7000	7105	.0150	66000	65382	-.0094	.996	-	-	-
9000	9059	.0066	29368	27888	-.0504	.994	-	-	-

TABLA 4.4.

Tres marcos - marcos y tamaños de muestras iguales.

$N_1 = N_2 = N_3 = 10000$; $n_1 = n_2 = n_3 = 100$; $M = 75$

tamaños de dominio verdaderos.	media de θ^*_{12}	sesgo relativo	$\sigma^2_{\theta^*_{12}}$	media de $\hat{\sigma}^2_{\theta^*_{12}}$	sesgo relativo	R_p	α		
							.10	.05	.01
$\theta^{\circ}_{12}=1000$	1022	.0220	44444	45039	.0134	.986	*	-	-
$\theta^{\circ}_{13}=1000$	1004	.0040	44444	44379	-.0015	.992	-	-	-
$\theta^{\circ}_{23}=1000$	980	-.0200	44444	43433	-.0227	.995	-	-	-
$\theta^{\circ}_{123}=1000$	1011	.0110	30000	30184	.0061	.994	-	-	-

TABLA 4.5.

Tres marcos - marcos desiguales -
Afijación proporcional

$N_1 = 10000, \quad N_2 = N_3 = 20000; \; n_1 = 100; \; n_2 = n_3 = 200;$

$M = 75$

tamaños de dominio verdaderos.	media de θ^*_{12}	sesgo relativo	$\sigma^2_{\theta^*_{12}}$	media de $\hat{\sigma}^2_{\theta^*_{12}}$	sesgo relativo	R_p	α		
							.10	.05	.01
$\theta^\circ_{12} = 2000$	2012	.0060	79187	79049	-.0017	.991	-	-	-
$\theta^\circ_{13} = 4000$	3976	-.0060	124449	123506	-.0076	.990	-	-	-
$\theta^\circ_{23} = 3000$	2961	-.0130	123626	121820	-.0146	.995	-	-	-
$\theta^\circ_{123} = 1000$	1013	.0130	30920	31211	.0094	.989	-	-	-

TABLA 4.6.

Tres marcos - marcos desiguales
Afijación no-proporcional.

$N_1 = 10000,\ N_2 = 20000,\ N_3 = 50000;\ n_1 = 300,\ n_2 = 100.$

$n_3 = 200;\quad M = 75$

tamaños de dominio verdaderos.	media de θ^*_{12}	sesgo relativo	$\sigma^2_{\theta^*_{12}}$	media de $\hat{\sigma}^2_{\theta^*_{12}}$	sesgo relativo	R_p	α		
							.10	.05	.01
$\theta^\circ_{12} = 2000$	1999	-.0005	45030	44851	-.0040	.996	-	-	-
$\theta^\circ_{13} = 4000$	4029	.0073	67436	67378	-.0009	.994	-	-	-
$\theta^\circ_{23} = 8000$	7989	-.0014	520683	516885	-.0073	.998	-	-	-
$\theta^\circ_{123} = 3000$	2996	-.0013	55452	55236	-.0039	.993	-	-	-

TABLA 4.7.

Tres marcos - marcos desiguales - afijación proporcional.

$N_1 = 10000$, $N_2 = 20000$, $N_3 = 50000$; $n_1 = 100$, $n_2 = 200$

$n_3 = 500$; $M = 75$

tamaños de dominio verdaderos.	media de θ^*_{12}	sesgo relativo	$\sigma^2_{\theta^*_{12}}$	media de $\hat{\sigma}^2_{\theta^*_{12}}$	sesgo relativo	R_p	α .10	.05	.01
$\theta^\circ_{12} = 2000$	2013	.0065	75642	75434	-.0027	.986	*	-	-
$\theta^\circ_{13} = 4000$	3960	-.0100	111933	111084	-.0076	.992	-	-	-
$\theta^\circ_{23} = 8000$	7968	-.0040	249412	248247	-.0047	.992	-	-	-
$\theta^\circ_{123} = 3000$	2994	-.0020	76933	76598	-.0044	.998	-	-	-

BIBLIOGRAFIA

- Alldredge, J.R. y D.W. Armstrong (1974). "Maximum likelihood estimation for the multinomial distribution using geometric programming, *Technometrics*, *16*, 585-587.
- Bishop, M.M., S.E. Fienberg y P.W. Holland (1975). *Discrete multiviariate Analysis: Theory and Practice*, The MIT Press, Cambridge, Massachusetts, 441-446, 457-480, 502-513.
- Bryant, E.G. y D.W. King (1960). "Estimation from populations identified by overlapping Sampling frames", artículo presentado en el 120º Congreso Anual de la American Statistical Association, Palo Alto, California, Estados Unidos.
- Cochran, R.S. (1967). "The estimation of domain sizes when sampling frames are interlockings", Proceeding of the Social Statistic Section of the American Statistical Association Meetings, Washington, D.C., Estados Unidos.
- ———————— y W.P. Cooke (1975). "The estimation of domain sizes when sampling frames overlap", College of Commerce and Industry Research Paper Series Nº 93, University of Wyoming, Laramie, Wyoming, Estados Unidos.
- ———————— y T. Toro (1973). "Maximum likelihood estimates of overlap sizes, an application of CONVEX", College of Commerce and Industry Research Paper Series Nº 8, University of Wyoming, Laramie, Wyoming, Estados Unidos.
- Cooke, W.P. (1974). "Properties of geometric programming estimators of domain sizes", artículo presentado en el 134º Congreso Anual de la Amercian Statistical Asso-

ciation, St. Louis, Missouri, Estados Unidos.

- Cooke, W.P. (1983). "Surrogate geometric programming estimates of restricted multinomial proportions", *Communications in Statistics*, *12*, 291-305.
- Fuller, W.A. y L.F. Busmeister (1972). "Estimators for samples selected from two overlapping frames", Proceedings of the Social Statistics Section of the American Statistical Association, 245-249.
- Hartley, H.O. y R.R. Hocking (1963). "Convex programming by Tangential approximation", *Management Sciences*, *9*, 600-612.
- IMSL, Institute of Mathematical Statistics Library (1980), Houston, Texas, Estados Unidos.
- LaMotte, L.R. y R.H. Oxspring (1971). "A computer program for the Hartley-Hocking convex programming algorithms", programa de computador no publicado, Texas A&M University, College Station, Texas, Estados Unidos.
- Ryan, T.A., Jr. y B.L. Joiner. "Normal probability plots and tests for normality", Reporte Técnico. Statistics Department, Pennsylvania State University, Estados Unidos.
- Shapiro, S.S. y M.B. Wilk (1965). "An analysis of variance test for normality (complete samples)", *Biometrika*, *52*, 591.
- Williams, R.E. (1957). "Estimation of overlapping strata boundaries", Tesis de Maestría, University of Wyoming, Laramie, Wyoming, Estados Unidos.

*