Revista Colombiana de Estadística
Volumen **23** (2000) No. 1, páginas 35–63

ESTIMACIÓN MÁXIMO VEROSÍMIL DE LOS TAMAÑOS DE DOMINIOS EN PROBLEMAS DE MARCOS MÚLTIPLES CUANDO SE DESCONOCE EL TAMAÑO DE LOS MARCOS

ALVARO. M. MONTENEGRO[1]
DAVID OSPINA[2]

Resumen. Se estudia el problema de obtener estimaciones máximo verosímiles del tamaño de los dominios para el caso de dos marcos traslapados cuando se desconoce el tamaño de los marcos. Se utiliza la técnica de muestreo por áreas y se obtienen estimadores insesgados de varianza mínima para el tamaño de cada uno de los marcos y del dominio traslapado. Se muestran propiedades de consistencia y de normalidad asintótica en el caso en el que las áreas son de igual medida.

Palabras claves:. Marcos múltiples, marcos traslapados, máxima verosimilitud, estimación de parámetros.

1. Introducción

El muestreo en marcos múltiples puede ser definido como un conjunto de varios muestreos (simples) cuyas muestras se combinan para obtener estimaciones de parámetros en la unión de los marcos (Hardley, 1974). Los marcos no necesariamente son de la misma naturaleza. Por ejemplo en aplicaciones de tipo agrícola que forman la mayor parte de la aplicaciones reportadas, se usan dos tipos de marcos definidos como *marco área* y *marco lista.* En tales aplicaciones la unidad de muestreo de un marco lista es un nombre que puede corresponder a un individuo, conjunto de individuos o negocios (Hardley, 1974 y Hill 1977). La unidad reportada para este caso consiste en todas las parcelas de tierra

(1) Profesor Asociado, Departamento de Estadística, Universidad Nacional de Colombia; email: amonte@matematicas.unal.edu.co
(2) Profesor Asociado, Departamento de Estadística, Universidad Nacional de Colombia; email: dospina@matematicas.unal.edu.co .

trabajadas bajo el nombre seleccionado. Por otro lado, la unidad de muestreo del marco área es un segmento de tierra (pequeño bloque de tierra) (Bosecker, 1976). Junto con la delimitación del segmento cada unidad de tierra trabajada constituye una unidad reportada, Hill (1977). Los nombres para cada unidad de tierra trabajada son obtenidos a partir de los datos de la encuesta por cada unidad reportada. Por razones de optimización lo que usualmente se hace es tomar una muestra en el marco área que por lo general tiene una cobertura total de la población y una en el marco lista, por lo general incompleto. Se supone que es posible identificar cuándo una unidad reportada en uno de los marcos pertenece también al otro.

En general se supone que todos los marcos son fracciones traslapadas de la misma población de unidades, de tal manera que las unidades pueden ser clasificados en dominios. Por ejemplo si se tienen dos marcos digamos A y B, estos generan $2^2 - 1 = 3$ dominios definidos de acuerdo a la notación de Hardley (1974) como

Dominio a que contiene solo elementos de A
Dominio b que contiene solo elementos de B
Dominio ab que contiene elementos de A y B.

El problema de estimación en situaciones de marcos múltiples ha sido estudiado por varios autores. Hardley (1962, 1974) desarrolló los fundamentos de la teoría. Lund (Lund, 1968) obtuvo estimaciones del total de una característica en el caso de dos marcos cuando se conoce el tamaño de los dominios; cuando se desconocen tales tamaños utilizó estimadores insesgados de tales tamaños a partir de la muestra y utilizando factores de peso para los dominios llega de nuevo a estimaciones del total de una característica, mejorando en términos de eficiencia y menor complejidad las estimaciones de Hardley. Fuller y Burmeister (1972) por su parte hicieron un completo desarrollo para el caso de dos marcos, presentando estimaciones del tamaño del dominio traslapado más eficientes que las de Hardley.

El problema de estimar el tamaño de los dominios creados por la intersección de múltiples marcos muestrales ha sido estudiado por varios autores usando tres técnicas distintas. King (1960), y Bryant y King (1960) presentaron su aproximación utilizando la técnica ji–cuadrado mínimo modificado, Williams (1957) basó las estimaciones en el método de máxima verosimilitud y Cochran (1965) hizo el desarrollo a partir de la técnica de marcos múltiples. Cochran y Cooke (1967) obtuvieron la varianza de las estimaciones obtenidas de ji–cuadrado y marcos múltiples para muestras grandes. Fuller y Burmeister (1972) obtuvieron estimadores y varianzas cuando se tienen en cuenta las unidades duplicadas y desarrollaron también una forma general de la varianza del estimador máximo verosímil (EMV) cuando se consideran dos marcos y se usa el modelo hipergeométrico. Bryant y King (1960) analizaron el caso de tres marcos pero no compararon varianzas. Ospina (1985), Ospina (1989) desarrolló aproximaciones máximo verosímiles en el caso de m marcos para el tamaño de los dominios haciendo una aproximación con la distribución multinomial para la función de verosimilitud de las observaciones y presentando por primera vez expresiones generales (en este caso asintóticas) para la matriz de covarianzas de los estimadores. Para el cálculo de los estimadores máximo verosímiles Ospina utilizó la técnica de programación goemétrica subrogada basándose en un algoritmo escrito por Cooke (1983).

En la estimación máximo verosímil del tamaño de los dominios en problemas de marcos múltiples se supone que el tamaño de cada marco se conoce. En este trabajo se estudia el caso mas realista en el que se desconoce el tamaño de los marcos y se llega a estimaciones máximo verosímiles para el tamaño de los marcos y de los dominios simultáneamente. cuando se tienen dos marcos y por tanto tres posibles dominios. Se hace uso de la teoría de EMV para el caso de variables aleatorias independientes pero no idénticamente distribuidas para establecer propiedades de consistencia, normalidad y eficiencia asintótica. Como justificación para el desarrollo de este trabajo consideremos dos ejemplos posibles de utilización. Primero, supóngase para el caso de tipo agrícola que se dispone de un marco área pero se desconocen las unidades muestrales total o parcialmente y que se dispone de un marco lista probablemente completo pero muy grande (un directorio) y no se conoce el tamaño pero se tiene una información adicional que puede llevar a una estimación del tamaño. Como segundo ejemplo posible de utilización supongamos que se tienen los directorios de dos empresas telefónicas, que no se dispone del total de abonados en cada caso y se requiere estimar el número de abonados en las dos empresas simultáneamente. En general este trabajo se puede aplicar a todos los problemas de dos marcos traslapados estudiados con anterioridad, para los cuales no sea conocido de antemano su tamaño.

2. Preliminares

La tripla $(\Omega, \mathfrak{F}, \mu)$ denotará un espacio de medida en donde Ω es un conjunto no vacío, $\mathfrak{F}$ una σ -álgebra de subconjuntos de Ω y μ una medida. Si Ω es finito o infinito numerable $\mathfrak{F}$ será $\sigma-$álgebra de todos los subconjuntos de Ω y μ la medida de conteo. Si $\Omega = \mathbf{R}^q$, $\mathfrak{F}$ será $\sigma-$álgebra de los conjuntos de Borel y μ la medida de Lebesgue. Si C es una clase de subconjuntos de Ω y $A \subset \Omega$ denotaremos por $C \cap A$ a la clase $\{B \cap A : B \in C\}$. En este caso $\mathfrak{F}_A$ será la mínima σ -álgebra que contiene a la clase $C \cap A$. En este trabajo no se hará referencia a espacios de medida más generales.

Si f es una función Borel medible definida sobre Ω, entonces $\int_S f d\mu$ representa la integral de f sobre $S \subset \Omega$ con respecto a la medida μ. Si μ es la medida de conteo entonces la integral se reduce a la suma $\sum_{x\in S} f(x)$. Sea X una variable aleatoria que toma valores en una región S con función de probabilidad p_θ y parámetro $\theta \in \Theta \subset \mathbf{R}^p$, entonces μ es la medida de conteo y $\int_S f\ p_\theta\ d\mu$ $=\sum_{x\in S} f(x)p_\theta(x)$.

Nota 1. *La notación $\int_S f d\mu$ es equivalente a $\int_S f(x)d\mu(x)$. Si se omite S en la integral, es decir, si escribe $\int f d\mu$ se asume que la integral se toma sobre todos los valores para los cuales está definida X.*

Si X tiene función de densidad o de probabilidad p_θ $\theta \in \Theta \subset \mathbf{R}^p$, la *matriz de información* del parámetro θ es la matriz de tamaño $p \times p$ dada por

$$I(\theta) = \|I_{ij}(\theta)\| \tag{1}$$

en donde

$$\begin{aligned} I_{ij}(\theta) &= E_\theta\left(\frac{\partial}{\partial\theta_i}\ln p_\theta(X)\cdot\frac{\partial}{\partial\theta_j}\ln p_\theta(X)\right) \\ &= \int\left(\frac{\partial}{\partial\theta_i}\ln p_\theta(X)\cdot\frac{\partial}{\partial\theta_j}\ln p_\theta(X)\right)p_\theta d\mu \end{aligned} \tag{2}$$

Nota 2. *Si se tiene que la derivada parcial con respecto a θ_i del lado izquierdo de*

$$\int p_\theta d\mu(x) = 1 \tag{3}$$

puede ser obtenida bajo el signo integral, entonces

$$E\left(\frac{\partial}{\partial\theta_i}\ln p_\theta(X)\right) = 0 \tag{4}$$

y

$$I_{ij}(\theta) = cov\left(\frac{\partial}{\partial\theta_i}\ln p_\theta(X), \frac{\partial}{\partial\theta_j}\ln p_\theta(X)\right). \tag{5}$$

En este caso $I(\theta)$ es semidefinida positiva. Adicionalmente si, $\partial^2 p_\theta(x)/\partial\theta_i\partial\theta_j$ existen para cada i, j y pueden ser obtenidas diferenciado dos veces bajo el signo integral en (3), entonces

$$I_{ij}(\theta) = -E_\theta\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln p_\theta(X)\right). \tag{6}$$

2.1 Propiedades asintóticas de EMV para el caso de variables independientes no idénticamente distribuidas

En esta sección se introduce la notación utilizada y se presenta la teoría para estimadores máximo verosímiles en el caso de variables aleatorias independientes no idénticamente distribuidas que se usa en la sección . Bradley y Gart (1962) y Hodley (1977) estudiaron el problema, partiendo de condiciones diferentes. En este trabajo es conveniente utilizar las condiciones de Bradley y Gart. Ellos examinaron el método de máxima verosimiltud en situaciones en donde las observaciones no provienen de la misma población, pero que están relacionadas en el sentido de que puede asumirse que tienen algunos parámetros en común, tales poblaciones fueron llamadas poblaciones asociadas.

2.1.1. Notación e hipótesis

Sean $f_i(\mathbf{X}_i, \theta)$ $(i = 1, ..., n)$ funciones de densidad o probabilidad (discretas o continuas) en donde $\mathbf{X}_i$ es un vector aleatorio con valores sobre una región S_i independientemente del vector de parámetros desconocidos $\theta \in \Theta \subset \mathbf{R}^p$. No es necesario que cada f_i dependa de todos los $\theta_1, ..., \theta_p$. Sean $x_{i1}, ..., x_{in_i}$ n_i vectores de observaciones de los vectores $\mathbf{X}_i$ $(i = 1, ..., n)$. La función de verosimilitud para este caso está dada por

$$\phi = \prod_{i=1}^{n} \prod_{\alpha=1}^{n_i} f_i(x_{i\alpha}, \theta) \tag{7}$$

y las ecuaciones normales para la maximización de $\ln \phi$ son

$$\frac{\partial \ln \phi}{\partial \theta_r} = 0 \qquad (r = 1, ...p) \tag{8}$$

Se partirá de las siguientes hipótesis:

M.1. Θ es un intervalo abierto $p-$dimensional, finito, infinito o semi-infinito

M.2. Si $\mathbf{x}_{i1}, ..., \mathbf{x}_{in_i}$ son variables aleatorias independientes e idénticamente distribuidas con función de densidad o de probabilidad $f_i(\mathbf{X}_i, \theta)$ respecto a una medida $\sigma-$finita μ y parámetro $\theta \in \Theta$, se supondrá que las distribuciones $P_i(\theta)$ de las $\mathbf{x}_{ij}$ tienen soporte común. Es decir, el conjunto $A_i = \{\mathbf{x}_i \in S_i : f_i(\mathbf{x}_i, \theta) > 0\}$ es independiente de θ para cada $i = 1, ..., n$. Además se supondrá que las distribuciones $P_i(\theta)$ de las observaciones son diferentes (de otra manera θ no puede ser estimado consistentemente).

M.3. Para casi todo $\mathbf{x}_i \in S_i$ $(i = 1, ..., n)$ y para todo θ en Θ , las derivadas

$$\frac{\partial \ln f_i}{\partial \theta_r}, \quad \frac{\partial^2 \ln f_i}{\partial \theta_r \partial \theta_s} \quad \text{y} \quad \frac{\partial^3 \ln f_i}{\partial \theta_r \partial \theta_s \partial \theta_t}$$

existen para $r, s, t = 1, ..., p; i = 1, ..., n$.

M.4. Para las f_i que son densidades existen funciones $F_{ir}(x_i)$, $F_{irs}(x_i)$ y $H_{irst}(x_i)$ tales que

$$\left|\frac{\partial \ln f_i}{\partial\theta_r}\right| < F_{ir}(x_i), \left|\frac{\partial^2 \ln f_i}{\partial\theta_r\partial\theta_s}\right| < F_{irs}(x_i), \text{ y } \left|\frac{\partial^3 \ln f_i}{\partial\theta_r\partial\theta_s\partial\theta_t}\right| < H_{irst}(x_i)$$

en casi toda parte de S_i con respecto a la medida de Lebesgue y todo $\theta \in \Theta$. Además $F_{ir}(x_i)$, $F_{irs}(x_i)$ son integrables sobre S_i y

$$\int_{S_i} H_{irst}(x_i) f_i dx_i < M_i$$

$(r, s, t = 1, ..., p;\ i = 1, ..., n)$ con M_i constantes positivas. Para las f_i, que son funciones de probabilidad, se tiene la medida de conteo y estas condiciones se traducen en que

$$\sum_{x_i \in S_i} \frac{\partial \ln f_i}{\partial\theta_r} \quad \text{y} \quad \sum_{x_i \in S_i} \frac{\partial^2 \ln f_i}{\partial\theta_r\partial\theta_s}$$

convergen uniformemente para todo $\theta \in \Theta$, y

$$\sum_{x_i \in S_i} H_{irst}(x_i) f_i < M_i$$

Estas hipótesis permiten intercambiar el orden de diferenciación e integración o suma . Si $I^{(i)}(\theta)$ denota la matriz de información correspondiente a $f_i(\mathbf{x}_i, \theta)$, se tiene el siguiente resultado Bradley y Gart (1962), Lehmann (1983), Apostol (1974).

Teorema 3. *Si las condiciones M1 a M4 se cumplen, entonces*

$$E_\theta\left(\frac{\partial \ln f_i}{\partial\theta_r}\right) = \int \left(\frac{\partial \ln f_i}{\partial\theta_r}\right) f_i d\mu = 0 \tag{9}$$

$$\begin{aligned} I_{rs}^{(i)}(\theta) &= E_\theta\left(\frac{\partial \ln f_i}{\partial\theta_r} \cdot \frac{\partial \ln f_i}{\partial\theta_s}\right) \\ &= -E_\theta\left(\frac{\partial^2 \ln f_i}{\partial\theta_r\partial\theta_s}\right) \\ &= cov\left(\frac{\partial \ln f_i}{\partial\theta_r}, \frac{\partial \ln f_i}{\partial\theta_s}\right) \qquad (10) \end{aligned}$$

$(r, s, t = 1, .., p; \quad i = 1, ..., n.)$

Se considerará la siguiente hipótesis adicional.

M.5. Sea $\rho_i = n_i/N$, en donde $N = \sum_{i=1}^{n} n_i$, entonces la matriz definida por

$$I(\theta) = \sum_{i=1}^{n} \rho_i I^{(i)}(\theta) \tag{11}$$

es definida positiva con determinante finito para todo $\theta \in \Theta$.

Nota 4. *Nótese que las ecuaciones (10) implican que $I^{(i)}(\theta)$ es semidefinida positiva. Por otro lado se tiene que si cada una de las matrices de información $I^{(i)}(\theta)$ es definida positiva entonces $I(\theta)$ también lo es. Además:*

$$I_{rs}(\theta) = \sum_{i=1}^{n} \rho_i E_\theta \left(\frac{\partial \ln f_i}{\partial \theta_r} \cdot \frac{\partial \ln f_i}{\partial \theta_s} \right). \tag{12}$$

Si además se satisfacen las condiciones M1 a M4, entonces

$$I_{rs}(\theta) = -\sum_{i=1}^{n} \rho_i E_\theta \left(\frac{\partial^2 \ln f_i}{\partial \theta_r \partial \theta_s} \right) \tag{13}$$

2.1.2 Consistencia

Los siguientes cuatro teoremas fueron demostrados por Bradley y Gart (1962) y son una generalización directa del caso variables aleatorias idénticamente distribuidas.

Teorema 5. *Dada la función de verosimilitud (7) y las hipótesis M.1 a M.5, supóngase que $\hat{\theta}$ es una solución de las ecuaciones normales de verosimiltud (8). Si θ^0 representa el valor verdadero de θ, y $\rho_i = n_i/N$ es constante cuando $n_i \to \infty$, entonces $\hat{\theta}$ es un estimador consistente de θ^0.*

Teorema 6. *Sea $\hat{\theta}$ un estimador consistente de θ^0, que es solución de (8), entonces bajo las hipótesis M.1 a M.5 la matriz*

$$\left(\frac{\partial^2 \ln \phi}{\partial \theta_r \partial \theta_s} \right)_{\theta = \hat{\theta}} \tag{14}$$

es definida negativa con probabilidad que tiende a 1 cuando $n_i = \rho_i N \to \infty$.

Teorema 7. *De todas las posibles soluciones de (8) y bajo los supuestos de las condiciones M.1 a M.5, una y sólo una tiende en probabilidad al vector parámetro verdadero.*

2.1.3 Normalidad Asintótica

Teorema 8. *Bajo las hipótesis M.1 a M.5, si $\hat{\theta}$ es el vector estimador máximo verosímil del vector de parámetros θ^0 entonces $\sqrt{n}\left(\hat{\theta}-\theta^0\right)$ tiene distribución asintótica normal multivariada con media cero y matriz de varianzas covarianzas $\mathbf{I}_0^{-1}$ en donde*

$$\begin{aligned}\mathbf{I}_0 &= \sum_{i=1}^{n} \rho_i E_\theta \left(\frac{\partial \ln f_i}{\partial \theta_r} \cdot \frac{\partial \ln f_i}{\partial \theta_s} \right)_{\theta=\theta^0} \\ &= -\sum_{i=1}^{n} \rho_i E_\theta \left(\frac{\partial^2 \ln f_i}{\partial \theta_r \partial \theta_s} \right)_{\theta=\theta^0} \qquad (15)\end{aligned}$$

3. Estimación máximo verosímil del tamaño del dominio traslapado y el tamaño de los marcos en el caso de dos marcos.

3.1 Notación e hipótesis

Sean $\mathcal{M}_1$ y $\mathcal{M}_2$ dos marcos que se combinan para dar cobertura completa a una población. Entonces pueden existir hasta tres dominios. Al primero pertenecen las unidades que están únicamente en el marco $\mathcal{M}_1$; al segundo dominio pertenececen la unidades que están únicamente en el marco $\mathcal{M}_2$; y al tercer dominio pertenecen las unidades restantes, o sea, las que están en ambos marcos. Sean $(\Omega_1, \mathfrak{F}_1, \mu_1)$ y $(\Omega_2, \mathfrak{F}_2, \mu_2)$ dos espacios de medida en donde los Ω_i son subespacios euclideanos q_i dimensionales, con $\mathfrak{F}_i$ las respectivas σ-algebras de Borel y μ_i las respectivas medidas de Lebesgue. Sean $\mathfrak{C}_1$ y $\mathfrak{C}_2$ clases de subconjuntos de Ω_1 y Ω_2 respectivamente y sean $\mathcal{A}_1$y $\mathcal{A}_2$ subconjuntos de Ω_1 y Ω_2 respectivamente tales que $\mu_1(\mathcal{A}_1)$ y $\mu_2(\mathcal{A}_2)$ son finitas. Se denotará por Ω_{A_i} a la clase $\mathfrak{C}_i \cap \mathcal{A}_i$ $(i = 1, 2)$. $\mathfrak{F}_{A_i}$ es la mínima σ-álgebra que contiene a Ω_{A_i} Entonces $(\Omega_{A_i}, \mathfrak{F}_{A_i}, \mu_i)$ $(i = 1, 2)$ son subespacios de medida finita con μ_i la restrición de la medida de Lebesgue correspondiente.

Se supondrá que $(\Omega_{A_1}, \mathfrak{F}_{A_1}, \mu_1)$ y $(\Omega_{A_2}, \mathfrak{F}_{A_2}, \mu_2)$ están asociados a los marcos $\mathcal{M}_1$ y $\mathcal{M}_2$ respectivamente, en el sentido de que cada unidad del marco $\mathcal{M}_i$ tiene asociada un punto en Ω_{A_i}. Se dirá entonces que la unidad se encuentra ubicada en Ω_{A_i}. Supóngase que existe una clase $\mathcal{T}_1$ de conjuntos disyuntos de la σ-álgebra $\mathfrak{F}_{A_1}$ tal que $\Omega_{A_1} = \cup_{A_{1i} \in \mathcal{T}_1} A_{1i}$. Se usará la notación $a_{1i} = \mu_1(A_{1i})$ y $A_1 = \mu_1(\Omega_{A_1})$. De manera completamente análoga se definen $\mathcal{T}_2$, A_{2j}, a_{2j}, y A_2. Cada elemento de $\mathfrak{F}_{A_i}$ se llamará área.

Sea η_i $i = 1, 2$ dos variables aleatorias definidas como sigue:

A1. Para cada área $t_i \in \mathfrak{F}_{A_i}$, $\eta_i(t_i) = n_i$, en donde n_i es el número de unidades que se encontran ubicadas en el área t_i en un momento determinado. La variable aleatoria η_i puede tomar los valores 0, 1, 2, ...

A2. La probablidad de encontrar asociada exactamente una unidad en un área t_i con medida $h_i = \mu_i(t_i)$ suficientemente pequeña es aproximadamente igual a $\lambda_i h_i$. Es decir, $P[\eta_i(t_i) = 1 \mid h_i = \mu_i(t_i)$ es suficientemente pequeña$] = \lambda_i h_i + o(h_i)$, en donde λ_i es una constante positiva.

A3. La probabilidad de encontrar asociadas más de una unidad en un área de medida muy pequeña es despreciable comparada con la probabilidad de encontrar exáctamente una unidad en la misma área. Es decir, $P[\eta_i(t_i) > 1 \mid h_i = \mu_i(t_i)$ es pequeña$] = o(h_i)$.

A4. Los números de unidades asociadas a áreas disyuntas en cualquier momento son independientes.

Las hipótesis A1 a A4 determinan que las variables η_i tienen distribución de Poisson con parámetro $\lambda_i h_i$. Entonces si t_i es un área de $\mathfrak{F}_{A_i}$ con medida $h_i = \mu_i(t_i)$ se tiene que $P[\eta_i(t_i) = n_i] = e^{-\lambda_i h_i}(\lambda_i h_i)^{n_i}/n_i!$, $n_i = 0, 1, 2, ...$ La cantidad λ_i se denominará densidad del espacio Ω_{A_i}.

3.2 Modelo de muestreo y función de verosimilitud

El propósito de este trabajo es obtener estimaciones máximo verosímiles para los tamaños N_1 y N_2 de los marcos $\mathcal{M}_1$ y $\mathcal{M}_2$ respectivamente y del tamaño θ del dominio traslapado. Se puede observar que si λ_i es la densidad del espacio Ω_{A_i} entonces

$$\lambda_i = N_i/A_i, \quad i = 1, 2 \tag{16}$$

Por tanto, si N_i se desconoce y se encuentra una estimación máximo verosímil $\hat{\lambda}_i$ de λ_i entonces $\hat{N}_i = \hat{\lambda}_i A_i$ es la estimación máximo verosímil de N_i. Además $E[\hat{N}_i] = E[\hat{\lambda}_i]A_i$ y $Var[\hat{N}_i] = Var[\hat{\lambda}_i]A^2$.

El proceso de muestreo consiste en seleccionar aleatoriamente m_1 áreas de $\mathcal{T}_1$, m_2 áreas de $\mathcal{T}_2$ y contar las unidades que se encuentren ubicadas cada una de las áreas. Se supone que existe un mecanismo para identificar cuando una unidad encontrada en un marco también pertenece al otro marco. Sean η_{1i} $i = 1, 2, ...m_1$ las variables aleatorias definidas como el número de unidades encontradas en cada una de las áreas t_{1i}, sean $a_{1i} = \mu_1(t_{1i})$ la medidas del áreas t_{1i}, y sean X_{1i} $i = 1, 2, ..m_1$ las variables aleatorias definidas como el número de unidades encontradas en el área t_{1i} que pertenecen únicamente al marco 1. Bajo la hipótesis adicional de que la pertenencia de una unidad solamente al marco 1 es independiente para todas las unidades de dicho marco y de acuerdo con la ecuación (16), la probabilidad conjunta de encontrar n_{1i} unidades en el área t_{1i} del marco 1, de las cuales x_{1i} pertenecen únicamente al marco 1 está dada por

$$\begin{aligned} P[X_{1i} = x_{1i}, \eta_{1i} = n_{1i} \mid \lambda_1, \theta] &= P[X_{1i} = x_{1i} \mid \eta_{1i} = n_{1i}, \lambda_1, \theta] \\ &\times P[\eta_{1i} = n_{1i}, \mid \lambda_1] \\ &= \binom{n_{1i}}{x_{1i}} \left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1} \right)^{x_{1i}} \left(\frac{\theta}{\lambda_1 A_1} \right)^{n_{1i} - x_{1i}} \\ &\times e^{-\lambda_i a_{1i}} (\lambda_i a_{1i})^{n_{1i}} / n_{1i}! \end{aligned} \tag{17}$$

La probabilidad de la izquierda se justifica teniendo en cuenta que de acuerdo a la hipótesis de independencia del suceso de pertenecer unicamente al marco 1, la variable aleatoria$(X_{1i} \mid \eta_{1i} = n_{1i})$ tiene distribución binomial con parámetros n_{1i} y $p_1 = \left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right)$. La probabilidad del segundo paréntesis se obtiene de las hipótesis A.1 a A.4 de la sección anterior.

La función de probabilidad conjunta del vector aleatorio (X_{1i}, η_{1i}), está dada por

$$\begin{aligned} f_{1i}(X_{1i}, \eta_{1i} \mid \lambda_1, \theta) = \binom{\eta_{1i}}{X_{1i}} \left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1} \right)^{X_{1i}} \left(\frac{\theta}{\lambda_1 A_1} \right)^{\eta_{1i} - X_{1i}} \\ \times e^{-\lambda_1 a_{1i}} (\lambda_1 a_{1i})^{\eta_{1i}} / \eta_{1i}! \end{aligned} \tag{18}$$

en donde $X_{1i} = 0, 1, 2, ...\eta_{1i}$ y $\quad \eta_{1i} = 0, 1, 2, ...$

Si $S = \{(x, n) \mid x = 0, 1, 2, ...n; \quad n = 0, 1, 2...\}$ se verifica que

$$\sum_{(x,n) \in S} f_{1i}(x, n \mid \lambda_1, \theta) = 1 \tag{19}$$

por lo que f_{1i} es una función de probabilidad con valores sobre la región S.

La función de verosimilitud asociada a las observaciones $(x_{1i}, n_{1i}), \quad i = 1, 2, ..m_1$ está dada por

$$\Phi_1((x_{11}, n_{11}), ...(x_{1m_1}, n_{1m_1}) \mid \lambda_1, \theta) = \prod_{i=1}^{m_1} f_{1i}(x_{1i}, n_{1i} \mid \lambda_1, \theta) \tag{20}$$

Sea η_1 la variable aleatoria definida por

$$\eta_1 = \sum_{i=1}^{m_1} \eta_{1i} \tag{21}$$

Entonces η_1 tiene distribución de Poisson con parámetro $\lambda_1 a_1$, en donde $a_1 = \sum_{i=1}^{m_1} a_{1i}$

Sea $(X_1 \mid \eta_1 = \sum_{i=1}^{m_1} \eta_{1i})$ la variable aleatoria definida por

$$(X_1 \mid \eta_1 = \sum_{i=1}^{m_1} \eta_{1i}) = \sum_{i=1}^{m_1} (X_{1i} \mid \eta_{1i} = n_{1i}) \tag{22}$$

La variable $(X_1 \mid \eta_1 = n_1)$ tiene distribución binomial con parámetros $\eta_1 =$

$\sum_{i=1}^{m_1} \eta_{1i}$ y $p_1 = \left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right)$ condicionada a $\eta_1 = n_1$

Si

$$x_1 = \sum_{i=1}^{m_1} x_{1i} \tag{23}$$

entonces la función de verosimilitud Φ_1 de la ecuación (20) se puede escribir como

(24)

$$\Phi_1((x_{11}, n_{11}), ..., (x_{1r_1}, n_{1r_1})|\lambda_1, \theta) {=} K_1 \left(\frac{\lambda_1 A_1 {-} \theta}{\lambda_1 A_1}\right)^{x_1} \left(\frac{\theta}{\lambda_1 A_1}\right)^{n_1 - x_1} e^{-\lambda_1 a_1} \lambda_1^{n_1}$$

en donde

$$K_1 = \prod_{i=1}^{m_1} \binom{n_{1i}}{x_{1i}} {a_{1i}}^{n_{1i}} / n_{1i}!, \qquad n_1 = \sum_{i=1}^{m_1} n_{1i}, \qquad a_1 = \sum_{i=1}^{m_1} a_{1i} \tag{25}$$

Nótese que n_1 es el número total de unidades encontradas en la muestra asociada al espacio Ω_{A_1}, $x_1$el número total de unidades que pertenecen únicamente al marco $\mathcal{M}_1$ y a_1 la medida del área total muestreada

Es claro que los vectores aleatorios (X_{1i}, η_{1i}) $(i = 1, ..., m_1)$ son independientes pero no idénticamente distribuidos, debido a que las áreas t_{1i}, no necesariamente tienen la misma medida. Si la clase $\mathcal{T}_1$ puede ser escogida de tal manera que todas las áreas tengan la misma medida dígase α_1,entonces los vectores (X_{1i}, η_{1i}) $(i = 1, ..., m_1)$ son idénticamente distribuidos con función de probabilidad dada por

$$\begin{aligned} f_1(X_1, \eta_1 \mid \lambda_1, \theta) = & \binom{\eta_1}{X_1} \left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right)^{X_1} \left(\frac{\theta}{\lambda_1 A_1}\right)^{\eta_1 - X_1} \\ & \times e^{-\lambda_1 \alpha_1} (\lambda_1 \alpha_1)^{\eta_1} / \eta_1! \end{aligned} \tag{26}$$

Nota 9. *Obsérvese que en este caso η_1 tiene distribución de Poisson con parámetro $\lambda_1 \alpha_1$, y $(X_1 \mid \eta_1)$ tiene distribución condicional binomial con parámetros η_1 y $p_1 = \left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right)$.*

La función de verosimilitud asociada a las observaciones (x_{1i}, n_{1i}), $i = 1, 2, ..m_1$ está dada en este caso por

$$\Psi_1((x_{11},n_{11}),...(x_{1m_1},n_{1m_1}) \mid \lambda_1,\theta = \prod_{i=1}^{m_1} f_1(n_{1i},x_{1i} \mid \lambda_1,\theta)$$

$$= L_1 \left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right)^{x_1} \left(\frac{\theta}{\lambda_1 A_1}\right)^{n_1 - x_1} e^{-\lambda_1 m_1 \alpha_1} \lambda_1^{n_1} \tag{27}$$

en donde

$$L_1 = \prod_{i=1}^{m_1} \binom{n_{1i}}{x_{1i}} {\alpha_1}^{n_{1i}}/n_{1i}!, \qquad n_1 = \sum_{i=1}^{m_1} n_{1i}, \qquad x_1 = \sum_{i=1}^{m_1} x_{1i} \tag{28}$$

Nota 10. *De manera completamente análoga se hace la construcción para la muestra asociada al espacio* Ω_{A_2}*, cambiando en cada expresión el 1 del primer subíndice por 2. En adelante supondremos definidos los términos* f_{2j} $(j = 1,...,m_2)$, $\Phi_2,$ $K_2,$ $f_2,$ L_2 *y* Ψ_2.

Si las dos muestras son combinadas se tiene que la función de verosimilitud asociada para las observaciones $(x_{1i},n_{1i}), i = 1,2,...m_1,$ $(x_{2j},n_{2j}),$ $j = 1,2,...m_2$ en el caso de áreas de distinta medida está dada por

$$\Phi((x_{11},n_{11}),...(x_{1r_1},n_{1r_1}),(x_{21},n_{21}),...(x_{2r_2},n_{2r_2}) \mid \lambda_1,\lambda_2,\theta) = \Phi_1\Phi_2$$

$$= \prod_{i=1}^{m_1} f_{1i}(n_{1i},x_{1i} \mid \lambda_1,\theta) \prod_{j=1}^{m_2} f_{2i}(n_{2j},x_{2j} \mid \lambda_2,\theta)$$

$$= K_1K_2 \left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right)^{x_1} \left(\frac{\theta}{\lambda_1 A_1}\right)^{n_1 - x_1} e^{-\lambda_1 a_1} \lambda_1^{n_1}$$

$$\times \left(\frac{\lambda_2 A_2 - \theta}{\lambda_2 A_2}\right)^{x_2} \left(\frac{\theta}{\lambda_2 A_2}\right)^{n_2 - x_2} e^{-\lambda_2 a_2} \lambda_2^{n_2}, \tag{29}$$

en donde $(\theta,\lambda_1,\lambda_2) \in \Theta$ y

$$\Theta = \left\{(\theta,\lambda_1,\lambda_2) \in \mathbf{R}^3 : \theta > 0, \quad \lambda_1 A_1 > \theta, \quad \lambda_2 A_2 > \theta\right\} \tag{30}$$

Para el caso de áreas de medida igual en cada espacio la función de verosimilitud está dada por

$$\begin{aligned}\Psi = \Psi_1\Psi_2 &= \prod_{i=1}^{m_1} f_1(n_{1i}, x_{1i} \mid \lambda_1, \theta) \prod_{j=1}^{m_2} f_2(n_{2j}, x_{2j} \mid \lambda_2, \theta) \\ &= L_1 L_2 \left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right)^{x_1} \left(\frac{\theta}{\lambda_1 A_1}\right)^{n_1 - x_1} e^{-\lambda_1 m_1 \alpha_1} \lambda_1^{n_1} \\ &\times \left(\frac{\lambda_2 A_2 - \theta}{\lambda_2 A_2}\right)^{x_2} \left(\frac{\theta}{\lambda_2 A_2}\right)^{n_2 - x_2} e^{-\lambda_2 m_2 \alpha_2} \lambda_2^{n_2} \end{aligned} \tag{31}$$

3.3 Estimación de los parámetros

El logaritmo de la función de verosimitud Φ de (29) esta dado por

$$\begin{aligned}\ln \Phi = K &+ (n_1 - x_1)\ln(\theta) + x_1 \ln(\lambda_1 A_1 - \theta) - \lambda_1 a_1 \\ &+ (n_2 - x_2)\ln(\theta) + x_2 \ln(\lambda_2 A_2 - \theta) - \lambda_2 a_2\end{aligned} \tag{32}$$

y para el caso (31) está dada por

$$\begin{aligned}\ln \Psi = L &+ (n_1 - x_1)\ln(\theta) + x_1 \ln(\lambda_1 A_1 - \theta) - \lambda_1 m_1 \alpha_1 \\ &+ (n_2 - x_2)\ln(\theta) + x_2 \ln(\lambda_2 A_2 - \theta) - \lambda_2 m_2 \alpha_2\end{aligned} \tag{33}$$

en donde K y L son contantes.

Teorema 11. *La función* $\ln \Phi$ *de (32) tiene un único punto crítico*

Prueba. El sistema de ecuaciones normales de verosimilitud para este caso está dado por

$$\frac{\partial \ln \Phi}{\partial \theta} = \frac{n_1 - x_1}{\theta} + \frac{n_2 - x_2}{\theta} - \frac{x_1}{\lambda_1 A_1 - \theta} - \frac{x_2}{\lambda_2 A_2 - \theta} = 0 \tag{34}$$

$$\frac{\partial \ln \Phi}{\partial \lambda_1} = \frac{A_1 x_1}{\lambda_1 A_1 - \theta} - a_1 = 0 \tag{35}$$

$$\frac{\partial \ln \Phi}{\partial \lambda_2} = \frac{A_2 x_2}{\lambda_2 A_2 - \theta} - a_2 = 0 \tag{36}$$

de (35) y (36) se obtiene respectivamente

$$\lambda_1 = \frac{\frac{A_1}{a_1} x_1 + \theta}{A_1} \tag{37}$$

$$\lambda_2 = \frac{\frac{A_2}{a_2} x_2 + \theta}{A_2} \tag{38}$$

y reemplazando (37) y (38) en (34) se obtiene

$$\frac{n_1 - x_1}{\theta} - \frac{a_1}{A_1} + \frac{n_2 - x_2}{\theta} - \frac{a_2}{A_2} = 0 \tag{39}$$

de donde

$$\theta = \frac{(n_1 - x_1) + (n_2 - x_2)}{\frac{a_1}{A_1} + \frac{a_2}{A_2}} \tag{40}$$

□

Nota 12. *Obsérvese que si* $0 < x_1 < n_1$ *y* $0 < x_2 < n_2$ *entonces el punto crı́ico obtenido satisface* $(\theta, \lambda_1, \lambda_2) \in \Theta$.

Teorema 13. *Bajo el supuesto que* $0 < x_1 < n_1$ *y* $0 < x_2 < n_2$ *la matriz de las segundas derivadas parciales de la función de verosimilitud* $\ln \Phi$ *de (32) es definida negativa para todos los valores* $(\theta, \lambda_1, \lambda_2) \in \Theta$*, y por tanto el extremo local del teorema anterior es el único máximo local de* $\ln \Phi$.

Prueba. La matriz de las segundas derivadas parciales de $\ln \Phi$ esta dada por

$$D(\theta, \lambda_1, \lambda_2) = \begin{bmatrix} \frac{\partial^2 \ln \Phi}{\partial \theta^2} & \frac{\partial^2 \ln \Phi}{\partial \theta \partial \lambda_1} & \frac{\partial^2 \ln \Phi}{\partial \theta \partial \lambda_2} \\ \frac{\partial^2 \ln \Phi}{\partial \theta \partial \lambda_1} & \frac{\partial^2 \ln \Phi}{\partial \lambda_1^2} & \frac{\partial^2 \ln \Phi}{\partial \lambda_1 \partial \lambda_2} \\ \frac{\partial^2 \ln \Phi}{\partial \theta \partial \lambda_2} & \frac{\partial^2 \ln \Phi}{\partial \lambda_1 \partial \lambda_2} & \frac{\partial^2 \ln \Phi}{\partial \lambda_2^2} \end{bmatrix} = \tag{41}$$

(42)

$$\begin{bmatrix} -\left(\frac{(n_1-x_1)+(n_2-x_2)}{\theta^2} + \frac{x_1}{(\lambda_1 A_1 - \theta)^2} + \frac{x_2}{(\lambda_2 A_2 - \theta)^2}\right) & \frac{A_1 x_1}{(\lambda_1 A_1 - \theta)^2} & \frac{A_2 x_2}{(\lambda_2 A_2 - \theta)^2} \\ \frac{A_1 x_1}{(\lambda_1 A_1 - \theta)^2} & -\frac{A_1^2 x_1}{(\lambda_1 A_1 - \theta)^2} & 0 \\ \frac{A_2 x_2}{(\lambda_2 A_2 - \theta)^2} & 0 & -\frac{A_2^2 x_2}{(\lambda_2 A_2 - \theta)^2} \end{bmatrix}$$

De acuerdo con el criterio de Sylvester (veáse por ejemplo Krasnov y colaboradores (1976, pág. 13) basta demostrar que

$$D_{11} < 0, \quad \begin{vmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{vmatrix} > 0, \quad \begin{vmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{vmatrix} < 0 \tag{43}$$

Supongamos que $0 < x_1 < n_1$ *y* $0 < x_2 < n_2$ Entonces de (42) se deduce directamente que $D_{11} < 0$. Por otro lado,

$$\begin{aligned} \begin{vmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{vmatrix} &= -\left(\frac{(n_1 - x_1) + (n_2 - x_2)}{\theta^2} + \frac{x_1}{(\lambda_1 A_1 - \theta)^2} + \frac{x_2}{(\lambda_2 A_2 - \theta)^2}\right) \\ &\quad \times \left(-\frac{A_1^2 x_1}{(\lambda_1 A_1 - \theta)^2}\right) - \left(\frac{A_1 x_1}{(\lambda_1 A_1 - \theta)^2}\right)^2 \\ &= \left(\frac{n_1 + n_2 - (x_1 + x_2)}{\theta^2} + \frac{x_2}{(\lambda_2 A_2 - \theta)^2}\right)\left(\frac{A_1 x_1}{(\lambda_1 A_1 - \theta)^2}\right)^2 > 0 \end{aligned}$$

y, finalmente,

$$\begin{aligned}
&\begin{vmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{vmatrix} = D_{31} \begin{vmatrix} D_{12} & D_{13} \\ D_{22} & 0 \end{vmatrix} + D_{33} \begin{vmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{vmatrix} \\
&= \left(\frac{A_2 x_2}{(\lambda_2 A_2 - \theta)^2} \right)^2 \left(\frac{A_1^2 x_1}{(\lambda_1 A_1 - \theta)^2} \right) + \left(-\frac{A_2^2 x_2}{(\lambda_2 A_2 - \theta)^2} \right) \\
&\times \left(\frac{n_1 + n_2 - (x_1 + x_2)}{\theta^2} + \frac{x_2}{(\lambda_2 A_2 - \theta)^2} \right) \left(\frac{A_1^2 x_1}{(\lambda_1 A_1 - \theta)^2} \right) \\
&= \left(-\frac{A_2^2 x_2}{(\lambda_2 A_2 - \theta)^2} \right) \left(\frac{n_1 + n_2 - (x_1 + x_2)}{\theta^2} \right) < 0
\end{aligned}$$

□

De acuerdo con los dos teoremas anteriores, la estimación máximo verosímil del vector de parámetros de $\ln \Phi$ está dada por $(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2)$ en donde

$$\hat{\theta} = \frac{(n_1 - x_1) + (n_2 - x_2)}{\frac{a_1}{A_1} + \frac{a_2}{A_2}} \tag{44}$$

$$\hat{\lambda}_1 = \frac{\frac{A_1}{a_1} x_1 + \hat{\theta}}{A_1} \tag{45}$$

$$\hat{\lambda}_2 = \frac{\frac{A_2}{a_2} x_2 + \hat{\theta}}{A_2} \tag{46}$$

y de la ecuación (16)

$$\hat{N}_1 = A_1 \hat{\lambda}_1 = \frac{A_1}{a_1} x_1 + \hat{\theta} \tag{47}$$

$$\hat{N}_2 = A_2 \hat{\lambda}_2 = \frac{A_2}{a_2} x_2 + \hat{\theta} \tag{48}$$

Nótese que como $0 < x_1 < n_1$ y $0 < x_2 < n_2$ entonces $\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) \in \Theta$.

Análogamente, la estimación máximo verosímil del vector de parámetros de $\ln \Psi$ está dada por $(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2)$ en donde

$$\hat{\theta} = \frac{(n_1 - x_1) + (n_2 - x_2)}{\frac{m_1 \alpha_1}{A_1} + \frac{m_2 \alpha_2}{A_2}} \tag{49}$$

$$\hat{\lambda}_1 = \frac{\frac{A_1}{m_1 \alpha_1} x_1 + \hat{\theta}}{A_1} \tag{50}$$

$$\hat{\lambda}_2 = \frac{\frac{A_2}{m_2 \alpha_2} x_2 + \hat{\theta}}{A_2} \tag{51}$$

$$\hat{N}_1 = \frac{A_1}{m_1\alpha_1}x_1 + \hat{\theta} \tag{52}$$

$$\hat{N}_2 = \frac{A_2}{m_2\alpha_2}x_2 + \hat{\theta} \tag{53}$$

Nota 14. *Para las aplicaciones prácticas deben tomarse como estimadores los enteros mas cercanos a* $\hat{\theta}, \hat{N}_1$ *y* $\hat{N}_2$

3.4. Propiedades de los estimadores

3.4.1. Esperanza de los estimadores

En esta sección se calculan la esperanza y la matriz de varianzas - covarianzas de los estimadores asociados a las estimaciones obtenidas en la seción anterior y se demuestra que todos los estimadores son insesgados

Si X_1 y η_1 son las variables aleatorias definidas en (21) y (22) respectivamente, y X_2 y η_2 son las análogas en el espacio 2 entonces los estimadores de máxima verosimilitud que se obtienen de la sección anterior están dados para el caso de la función de verosimiltud Φ por

$$\hat{\theta} = \frac{(\eta_1 - X_1) + (\eta_2 - X_2)}{\frac{a_1}{A_1} + \frac{a_2}{A_2}} \tag{54}$$

$$\hat{\lambda}_1 = \frac{\frac{A_1}{a_1}X_1 + \hat{\theta}}{A_1} \tag{55}$$

$$\hat{\lambda}_2 = \frac{\frac{A_2}{a_2}X_2 + \hat{\theta}}{A_2} \tag{56}$$

$$\hat{\mathbf{N}}_1 = \frac{A_1}{a_1}X_1 + \hat{\theta} \tag{57}$$

$$\hat{\mathbf{N}}_2 = \frac{A_2}{a_2}X_2 + \hat{\theta} \tag{58}$$

Nótese que para el estimador se usa el mismo símbolo que para la estimación pero resaltado.

Lema 15. *Si* X_1 *y* η_1 *son las variables aleatorias definidas en (21) y (22) entonces para el caso de la función de verosimiltud* Φ *se tiene que*

$$E[X_1] = Var[X_1] = \frac{a_1}{A_1}\left[\lambda_1 A_1 - \theta\right] \tag{59}$$

$$E[\eta_1] = Var[\eta_1] = \lambda_1 a_1 \tag{60}$$

$$E[\eta_1 - X_1] = Var[\eta_1 - X_1] = \frac{a_1}{A_1}\left[\theta\right] \tag{61}$$

$$Cov[\eta_1, X_1] = \frac{a_1}{A_1}\left[\lambda_1 A_1 - \theta\right] \tag{62}$$

$$Cov[\eta_1 - X_1, X_1] = 0 \tag{63}$$

Prueba. La expresión (60) se deduce del hecho de que $\eta_1$tiene distribución de Poisson con parámetro $\lambda_1 a_1$. Por otro lado la variable $(X_1 \mid \eta_1 = n_1)$ tiene distribución binomial con parámetros $\eta 1 = \sum_{i=1}^{m_1} \eta_{1i}$ y $p_1 = \left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right)$, por lo tanto

$$E(X_1 \mid \eta_1) = \eta_1 \left(\frac{\lambda_1 A_1 - theta}{\lambda_1 A_1}\right) \tag{64}$$

de donde

$$E[X_1] = E(E(X_1 \mid \eta_1)) = E[\eta_1]\left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right) = \frac{a_1}{A_1}\left[\lambda_1 A_1 - \theta\right] \tag{65}$$

Por otro lado,

$$Var(X_1 \mid \eta_1) = \eta_1 \left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right)\left(\frac{\theta}{\lambda_1 A_1}\right) \tag{66}$$

Luego

$$E(Var(X_1 \mid \eta_1)) = \lambda_1 a_1 \left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right)\left(\frac{\theta}{\lambda_1 A_1}\right) = \frac{a_1}{A_1}\left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right)\theta \tag{67}$$

Además

$$Var(E(X_1 \mid \eta_1)) = Var(\eta_1)\left(\frac{\lambda_1 A_1 - \theta}{\lambda_1 A_1}\right)^2 = \frac{a_1}{A_1}\frac{(\lambda_1 A_1 - \theta)^2}{\lambda_1 A_1} \tag{68}$$

De las ecuaciones (67) y (68) se obtiene que

$$Var[X_1] = E(Var(X_1 \mid \eta_1)) + Var(E(X_1 \mid \eta_1)) = \frac{a_1}{A_1}\left[\lambda_1 A_1 - \theta\right] \tag{69}$$

De manera análoga si se tiene en cuenta que $(\eta_1 - X_1 \mid \eta_1)$ tiene distribución binomial con parámetros η_1y $q_1 = \left(\frac{\theta}{\lambda_1 A_1}\right)$ se llega a la expresión (61).

Finalmente,

$$Cov[\eta_1, X_1] = \frac{1}{2}\left(Var\left(\eta_1\right) + Var\left(X_1\right) - Var\left(\eta_1 - X_1\right)\right) \tag{70}$$

$$= \frac{1}{2}\left(\lambda_1 a_1 + \frac{a_1}{A_1}\left[\lambda_1 A_1 - \theta\right] - \frac{a_1}{A_1}\left[\theta\right]\right) = \frac{a_1}{A_1}\left[\lambda_1 A_1 - \theta\right] \tag{71}$$

y

$$Cov[\eta_1 - X_1, X_1] = \frac{1}{2}\left(Var\left(\eta_1\right) - Var\left(X_1\right) - Var\left(\eta_1 - X_1\right)\right) = 0 \tag{72}$$

□

Nota 16. *Análogamente se tiene que*

$$E[X_2] = Var[X_2] = \frac{a_2}{A_2}[\lambda_2 A_2 - \theta] \tag{73}$$

$$E[\eta_2] = Var[\eta_2] = \lambda_2 a_2 \tag{74}$$

$$E[\eta_2 - X_2] = Var[\eta_2 - X_2] = \frac{a_2}{A_2}[\theta] \tag{75}$$

$$Cov[\eta_2, X_2] = \frac{a_2}{A_2}[\lambda_2 A_2 - \theta] \tag{76}$$

$$Cov[\eta_2 - X_2, X_2] = 0 \tag{77}$$

Corolario 17.

$$Cov[X_1, \hat{\theta}] = 0 \tag{78}$$

$$Cov[X_2, \hat{\theta}] = 0 \tag{79}$$

Prueba.

$$Cov[X_1, \hat{\theta}] = Cov\left[X_1, \frac{(\eta_1 - X_1) + (\eta_2 - X_2)}{\frac{a_1}{A_1} + \frac{a_2}{A_2}}\right] \tag{80}$$

$$= \frac{A_1 A_2}{a_1 A_2 + a_2 A_1}\left(Cov\left[X_1, \eta_1 - X_1\right] + Cov\left[X_1, \eta_2 - X_2\right]\right) = 0 \tag{81}$$

De la misma forma $Cov[X_2, \hat{\theta}] = 0$. □

3.4.2. Varianza de los estimadores

Teorema 18. *Todos los estimadores de las expresiones (54) a (58) son insesgados.*

Prueba. Todos los resultados se derivan del lema 15. En efecto,

$$E[\hat{\theta}] = E\left[\frac{(\eta_1 - X_1) + (\eta_2 - X_2)}{\frac{a_1}{A_1} + \frac{a_2}{A_2}}\right] = \left[\frac{\theta\frac{a_1}{A_1} + \theta\frac{a_2}{A_2}}{\frac{a_1}{A_1} + \frac{a_2}{A_2}}\right] = \theta \tag{82}$$

$$E[\hat{\lambda}_1] = E\left[\frac{\frac{A_1}{a_1}X_1 + \hat{\theta}}{A_1}\right] = \frac{\frac{A_1}{a_1}\frac{a_1}{A_1}[\lambda_1 A_1 - \theta] + \theta}{A_1} = \lambda_1 \tag{83}$$

La prueba es idéntica para $\hat{\lambda}_{2.}$

$$E[\hat{\mathbf{N}}_1] = E\left[A_1\hat{\lambda}_1\right] = A_1\lambda_1 = N_1 \tag{84}$$

Equivalente para $\hat{\mathbf{N}}_2$ □

Teorema 19. *Las varianzas de los estimadores de las expresiones (54) a (58) están dadas por*

$$Var[\hat{\theta}] = \frac{\theta A_1 A_2}{a_2 A_1 + a_1 A_2} \tag{85}$$

$$Var[\hat{\lambda}_1] = \frac{1}{a_1}\left(\lambda_1 - \frac{\theta a_2}{a_2 A_1 + a_1 A_2}\right) \tag{86}$$

$$Var[\hat{\lambda}_2] = \frac{1}{a_2}\left(\lambda_2 - \frac{\theta a_1}{a_2 A_1 + a_1 A_2}\right) \tag{87}$$

$$Var[\hat{\mathbf{N}}_1] = \frac{A_1^2}{a_1}\left(\lambda_1 - \frac{\theta a_2}{a_2 A_1 + a_1 A_2}\right) \tag{88}$$

$$Var[\hat{\mathbf{N}}_2] = \frac{A_2^2}{a_2}\left(\lambda_2 - \frac{\theta a_1}{a_2 A_1 + a_1 A_2}\right) \tag{89}$$

Prueba.

$$\begin{aligned}
Var[\hat{\theta}] &= Var\left[\frac{(\eta_1 - X_1) + (\eta_2 - X_2)}{\frac{a_1}{A_1} + \frac{a_2}{A_2}}\right] = \frac{Var\left[\eta_1 - X_1\right] + Var\left[\eta_2 - X_2\right]}{\left(\frac{a_1}{A_1} + \frac{a_2}{A_2}\right)^2} \\
&= \frac{\theta \frac{a_1}{A_1} + \theta \frac{a_2}{A_2}}{\left(\frac{a_1}{A_1} + \frac{a_2}{A_2}\right)^2} = \frac{\theta}{\frac{a_1}{A_1} + \frac{a_2}{A_2}} = \frac{\theta A_1 A_2}{a_2 A_1 + a_1 A_2} \\
Var[\hat{\lambda}_1] &= Var\left[\frac{\frac{A_1}{a_1} X_1 + \hat{\theta}}{A_1}\right] = \frac{1}{a_1^2} Var\left[X_1\right] + \frac{1}{A_1^2} Var\left[\hat{\theta}\right] \\
&= \frac{1}{a_1^2}\left(\frac{a_1}{A_1}\left[\lambda_1 A_1 - \theta\right]\right) + \frac{1}{A_1^2}\left(\frac{\theta A_1 A_2}{a_2 A_1 + a_1 A_2}\right) \\
&= \frac{1}{a_1}\left(\lambda_1 - \frac{\theta a_2}{a_2 A_1 + a_1 A_2}\right)
\end{aligned}$$

Similarmente para $Var[\hat{\lambda}_2]$. Las varianzas para $\hat{\mathbf{N}}_1$ y $\hat{\mathbf{N}}_2$ se obtienen de $Var[\hat{\lambda}_1]$ y $Var[\hat{\lambda}_1]$ respectivamente. □

Corolario 20.

$$Cov\left[\hat{\theta}, \hat{\lambda}_1\right] = \frac{\theta A_2}{a_2 A_1 + a_1 A_2} \tag{90}$$

$$Cov\left[\hat{\theta}, \hat{\lambda}_2\right] = \frac{\theta A_1}{a_2 A_1 + a_1 A_2} \tag{91}$$

$$Cov\left[\hat{\lambda}_1, \hat{\lambda}_2\right] = \frac{\theta}{a_2 A_1 + a_1 A_2} \tag{92}$$

Prueba.

$$Cov\left[\hat{\theta},\hat{\lambda}_1\right] = Cov\left[\hat{\theta},\frac{X_1}{a_1}+\frac{\hat{\theta}}{A_1}\right] = \frac{1}{a_1}Cov\left[\hat{\theta},X_1\right]+\frac{1}{A_1}Var\left[\hat{\theta}\right]$$
$$= \frac{\theta A_2}{a_2A_1+a_1A_2}$$

Análogo para $Cov\left[\hat{\theta},\hat{\lambda}_2\right]$.

$$Cov\left[\hat{\lambda}_1,\hat{\lambda}_2\right] = Cov\left[\frac{X_1}{a_1}+\frac{\hat{\theta}}{A_1},\frac{X_2}{a_2}+\frac{\hat{\theta}}{A_2}\right] = \frac{1}{a_1a_2}Cov\left[X_1,X_2\right]$$
$$+\frac{1}{a_1A_2}Cov\left[X_1,\hat{\theta}\right]+\frac{1}{a_2A_1}Cov\left[X_2,\hat{\theta}\right]+\frac{1}{A_2A_1}Var\left[\hat{\theta}\right]$$
$$= \frac{\theta}{a_2A_1+a_1A_2}$$

□

Nota 21. *La matriz de varianzas - covarianzas de los estimadores para la función de verosimilitud* Φ *es*

(93)
$$C_\Phi(\hat{\theta},\hat{\lambda}_1,\hat{\lambda}_2) = \begin{bmatrix} \frac{\theta A_1A_2}{a_2A_1+a_1A_2} & \frac{\theta A_2}{a_2A_1+a_1A_2} & \frac{\theta A_1}{a_2A_1+a_1A_2} \\ \frac{\theta A_2}{a_2A_1+a_1A_2} & \frac{1}{a_1}\left[\lambda_1-\frac{\theta a_2}{a_2A_1+a_1A_2}\right] & \frac{\theta A_1A_2}{a_2A_1+a_1A_2} \\ \frac{\theta A_1}{a_2A_1+a_1A_2} & \frac{\theta A_1A_2}{a_2A_1+a_1A_2} & \frac{1}{a_2}\left[\lambda_2-\frac{\theta a_1}{a_2A_1+a_1A_2}\right] \end{bmatrix}$$

Las expresiones para la esperanza y la varianza de los estimadores obtenidos para la función de verosimiltud Ψ*, se obtienen reemplazando en cada caso* a_1 *por* $m_1\alpha_1$ *y* a_2 *por* $m_2\alpha_2$*, en este caso usamos la notación* $C_\Psi(\hat{\theta},\hat{\lambda}_1,\hat{\lambda}_2)$ *y se tiene*

$$C_\Psi(\hat{\theta},\hat{\lambda}_1,\hat{\lambda}_2) =$$
(94)
$$\begin{bmatrix} \frac{\theta A_1A_2}{m_2\alpha_2A_1+m_1\alpha_1A_2} & \frac{\theta A_2}{m_2\alpha_2A_1+m_1\alpha_1A_2} & \frac{\theta A_1}{m_2\alpha_2A_1+m_1\alpha_1A_2} \\ \frac{\theta A_2}{m_2\alpha_2A_1+m_1\alpha_1A_2} & \frac{\lambda_1}{m_1\alpha_1}-\frac{\frac{\theta m_2\alpha_2}{m_1\alpha_1}}{m_2\alpha_2A_1+m_1\alpha_1A_2} & \frac{\theta A_1A_2}{m_2\alpha_2A_1+m_1\alpha_1A_2} \\ \frac{\theta A_1}{m_2\alpha_2A_1+m_1\alpha_1A_2} & \frac{\theta A_1A_2}{m_2\alpha_2A_1+m_1\alpha_1A_2} & \frac{\lambda_2}{m_2\alpha_2}-\frac{\frac{\theta m_1a_1}{m_2\alpha_2}}{m_2\alpha_2A_1+m_1\alpha_1A_2} \end{bmatrix}$$

Nota 22. *Si en las matrices de covarianzas* $C_\Psi(\hat{\theta},\hat{\lambda}_1,\hat{\lambda}_2)$ *y* $C_\Phi(\hat{\theta},\hat{\lambda}_1,\hat{\lambda}_2)$ *se reemplazan* $\theta,\lambda_1,\lambda_2$ *por* $\hat{\theta},\hat{\lambda}_1,\hat{\lambda}_2$ *respectivamente se obtienen estimadores insesgados para tales matrices. En particular se obtienen estimadores insesgados para las varianzas.*

3.5. Normalidad asintotica de los estimadores para el caso de la función de verosimilitud Ψ

La consistencia de los estimadores máximo verosímiles obtenidos no está garantizada debido a que en el caso de la función de verosimiltud Φ no se obtiene a partir de vectores aleatorios idénticamente distribuidos. Para el caso de la función de verosimiltud Ψ, los vectores aleatorios son idénticamente distribuidos en cada muestra. El hecho de tener dos muestras independientes no permite garantizar directamente la consistencia de los estimadores. En esta sección se demuestra la normalidad asintótica de los estimadores máximo versímiles obtenidos para la función de verosimilitud Ψ definida en la ecuación (31). Se demostrará que las funciones f_1 y f_2 satisfacen las condiciones M.1 a M.5 de la sección . La condiciones M.1 y M.2 son obvias. Para establecer la condición M.3 basta observar que

$$\frac{\partial^3 \ln f_i}{\partial\theta\partial\lambda_1\partial\lambda_2} = 0 \quad (i = 1, 2) \tag{95}$$

Falta verificar las condiciones M.4 y M.5.

Teorema 23. *Las funciones f_{1i} y f_{2j} de la función de verosimiltud Φ definida en la ecuación (29) y las funciones f_1 y f_2 de la función de verosimiltud Ψ definida en la ecuación (31) satisfacen la condición M.4. de la sección .*

Prueba. Según (19) las funciones de probabilidad f_{1i} f_{2j}, f_1 y f_2 están todas definidas en el conjunto $S = \{(x, n) \mid x = 0, 1, 2, ..n; n = 0, 1, 2...\}$. De (95) se sigue que

$$\sum_{(x,n)\in S} \left| \frac{\partial^3 \ln f_i}{\partial\theta\partial\lambda_1\partial\lambda_2} \right| f_i = 0 \tag{96}$$

Análogamente, se tiene que

$$\sum_{(x,n)\in S} \left| \frac{\partial^3 \ln f_{ki}}{\partial\theta\partial\lambda_1\partial\lambda_2} \right| f_{ki} = 0, \;\; k = 1, 2 \tag{97}$$

Se demostrará que todas las sumas de la condición M.4. son cero para f_{1i} f_{2j}, f_1 y f_2. Sea

$$f(x, n; \theta, \lambda) = \binom{n}{x} (p_{\theta\lambda})^x \, (q_{\theta\lambda})^{n-x} \, h(n, \lambda) \tag{98}$$

en donde

$$p_{\theta\lambda} = \frac{\lambda A - \theta}{\lambda A}, \quad q_{\theta\lambda} = \frac{\theta}{\lambda A}, \text{ y } h(n,\lambda) = \frac{e^{-\lambda a}(\lambda a)^n}{n!}, \quad \lambda A > 0.$$

Entonces tomando $h(n,\lambda) = h$

$$\frac{\partial h}{\partial \lambda} = \frac{(n - a\lambda)}{\lambda}h, \;\; \frac{\partial^2 h}{\partial \lambda^2} = \frac{\left(-n + n^2 - 2an\lambda + a^2\lambda^2\right)}{\lambda^2}h$$

y como h es la función de probabilidad de Poisson con parámetro λa entonces se tiene que

$$\sum_{n=0}^{\infty} h = 1, \quad \sum_{n=0}^{\infty} nh = \lambda a \;\text{ y }\; \sum_{n=0}^{\infty} n^2 h = \lambda a\,(\lambda a + 1) \tag{99}$$

Por tanto

$$\sum_{n=0}^{\infty}\sum_{x=0}^{n} \frac{\partial f}{\partial \theta} = \sum_{n=0}^{\infty} h \frac{\partial}{\partial \theta} \sum_{x=0}^{n} \binom{n}{x} (p_{\theta\lambda})^x (q_{\theta\lambda})^{n-x} = \sum_{n=0}^{\infty} h \frac{\partial}{\partial \theta}(1) = 0 \tag{100}$$

$$\sum_{n=0}^{\infty}\sum_{x=0}^{n} \frac{\partial f}{\partial \lambda} = \sum_{n=0}^{\infty} \frac{\partial}{\partial \lambda}\left(h \sum_{x=0}^{n} \binom{n}{x} (p_{\theta\lambda})^x (q_{\theta\lambda})^{n-x}\right) = \sum_{n=0}^{\infty} \frac{\partial h}{\partial \lambda}$$

$$= \sum_{n=0}^{\infty} \frac{(n - a\lambda)}{\lambda} h = \frac{1}{\lambda}\sum_{n=0}^{\infty} nh - a\sum_{n=0}^{\infty} h = 0 \tag{101}$$

de (100) y (101) se sigue que

$$\sum_{n=0}^{\infty}\sum_{x=0}^{n} \frac{\partial^2 f}{\partial \theta \partial \lambda} = \sum_{n=0}^{\infty}\sum_{x=0}^{n} \frac{\partial^2 f}{\partial \theta^2} = 0 \tag{102}$$

y

$$\sum_{n=0}^{\infty}\sum_{x=0}^{n} \frac{\partial^2 f}{\partial \lambda^2} = \sum_{n=0}^{\infty} \frac{\partial^2 h}{\partial \lambda^2} = \sum_{n=0}^{\infty} \frac{\left(-n + n^2 - 2an\lambda + a^2\lambda^2\right)}{\lambda^2} h = 0 \tag{103}$$

□

Finalmente se probará que las funciones f_1 y f_2 satisfacen la condición M.5.

Teorema 24. *Sean* $M = m_1 + m_2$. $\rho_1 = m_1/M$, $\rho_2 = m_2/M$, *y sean* $I^{(i)}(\theta,\lambda_1,\lambda_2)$ $(i = 1,2)$ *las matrices definidas por*

$$I^{(i)}(\theta,\lambda_1,\lambda_2) = -\begin{bmatrix} E\left(\frac{\partial^2 \ln f_i}{\partial\theta\partial\theta}\right) & E\left(\frac{\partial^2 \ln f_i}{\partial\theta\partial\lambda_1}\right) & E\left(\frac{\partial^2 \ln f_i}{\partial\theta\partial\lambda_2}\right) \\ E\left(\frac{\partial^2 \ln f_i}{\partial\theta\partial\lambda_1}\right) & E\left(\frac{\partial^2 \ln f_i}{\partial\lambda_1\partial\lambda_1}\right) & E\left(\frac{\partial^2 \ln f_i}{\partial\lambda_1\partial\lambda_2}\right) \\ E\left(\frac{\partial^2 \ln f_i}{\partial\theta\partial\lambda_2}\right) & E\left(\frac{\partial^2 \ln f_i}{\partial\lambda_1\partial\lambda_2}\right) & E\left(\frac{\partial^2 \ln f_i}{\partial\lambda_2\partial\lambda_2}\right) \end{bmatrix} \tag{104}$$

$i = 1,2$. *Entonces, la matriz*

$$I(\theta,\lambda_1,\lambda_2)=\rho_1 I^{(1)}(\theta,\lambda_1,\lambda_2)+\rho_2 I^{(2)}(\theta,\lambda_1,\lambda_2) \tag{105}$$

es definida positiva con determinante finito para todo $(\theta,\lambda_1,\lambda_2)\in\Theta$.

Prueba.

$$E\left(\frac{\partial^2 \ln f_i}{\partial\theta\partial\theta}\right)=-E\left(\frac{\eta_i-X_i}{\theta^2}+\frac{X_i}{(\lambda_iA_i-\theta)^2}\right)=-\frac{\alpha_i\lambda_i}{\theta\,(\lambda_iA_i-\theta)} \tag{106}$$

$$E\left(\frac{\partial^2 \ln f_i}{\partial\theta\partial\lambda_i}\right)=E\left(\frac{A_iX_i}{(\lambda_iA_i-\theta)^2}\right)=\frac{\alpha_i}{(\lambda_iA_i-\theta)} \tag{107}$$

$$E\left(\frac{\partial^2 \ln f_i}{\partial\theta\partial\lambda_j}\right)=E\left(\frac{\partial^2 \ln f_i}{\partial\lambda_i\partial\lambda_j}\right)=0,\quad i\neq j \tag{108}$$

$$E\left(\frac{\partial^2 \ln f_i}{\partial\lambda_i\partial\lambda i}\right)=-E\left(\frac{A_1^2X_1}{(\lambda_1A_1-\theta)^2}\right)=-\frac{\alpha_iA_i}{(\lambda_iA_i-\theta)} \tag{109}$$

Por tanto:

$$I^{(1)}(\theta,\lambda_1,\lambda_2)=\begin{bmatrix}\frac{\alpha_1\lambda_1}{\theta(\lambda_1A_1-\theta)} & -\frac{\alpha_1}{(\lambda_1A_1-\theta)} & 0\\ -\frac{\alpha_1}{(\lambda_1A_1-\theta)} & \frac{\alpha_1A_1}{(\lambda_1A_1-\theta)} & 0\\ 0 & 0 & 0\end{bmatrix} \tag{110}$$

$$I^{(2)}(\theta,\lambda_1,\lambda_2)=\begin{bmatrix}\frac{\alpha_2\lambda_2}{\theta(\lambda_2A_2-\theta)} & 0 & -\frac{\alpha_2}{(\lambda_2A_2-\theta)}\\ 0 & 0 & 0\\ -\frac{\alpha_2}{(\lambda_2A_2-\theta)} & 0 & \frac{\alpha_2A_2}{(\lambda_2A_2-\theta)}\end{bmatrix} \tag{111}$$

Luego

$$I(\theta,\lambda_1,\lambda_2)=\frac{1}{M}\begin{bmatrix}\frac{m_1\alpha_1\lambda_1}{\theta(\lambda_1A_1-\theta)}+\frac{m_2\alpha_2\lambda_2}{\theta(\lambda_2A_2-\theta)} & -\frac{m_1\alpha_1}{(\lambda_1A_1-\theta)} & -\frac{m_2\alpha_2}{(\lambda_2A_2-\theta)}\\ -\frac{m_1\alpha_1}{(\lambda_1A_1-\theta)} & \frac{m_1\alpha_1A_1}{(\lambda_1A_1-\theta)} & 0\\ -\frac{m_2\alpha_2}{(\lambda_2A_2-\theta)} & 0 & \frac{m_2\alpha_2A_2}{(\lambda_2A_2-\theta)}\end{bmatrix} \tag{112}$$

De acuerdo con el criterio de Sylvester basta demostrar que

$$I_{11}>0,\quad \begin{vmatrix}I_{11} & I_{12}\\ I_{21} & I_{22}\end{vmatrix}>0,\ \text{y}\quad |I|=\begin{vmatrix}I_{11} & I_{12} & I_{13}\\ I_{21} & I_{22} & I_{23}\\ I_{31} & I_{32} & I_{33}\end{vmatrix}>0 \tag{113}$$

Es claro que $I_{11}>0$, porque por hipótesis $\lambda_1A_1-\theta>0$.

$$I_{11}I_{22}-I_{12}^2=\frac{(\rho_1\alpha_1\lambda_1)^2}{\theta\,(\lambda_1A_1-\theta)^2}+\frac{\rho_2\alpha_2\lambda_2}{\theta\,(\lambda_2A_2-\theta)}\cdot\frac{\rho_1\alpha_1A_1}{(\lambda_1A_1-\theta)}+\frac{(\rho_1\alpha_1)^2}{(\lambda_1A_1-\theta)^2}>0$$

Finalmente,

$$
\begin{aligned}
|I| = &-\frac{\rho_2\alpha_2}{(\lambda_2 A_2-\theta)}\begin{vmatrix} I_{12} & I_{13} \\ I_{22} & I_{23}\end{vmatrix} + \frac{\rho_2\alpha_2 A_2}{(\lambda_2 A_2-\theta)}\begin{vmatrix} I_{11} & I_{12} \\ I_{21} & I_{22}\end{vmatrix} \\
&\frac{\rho_2\alpha_2}{(\lambda_2 A_2-\theta)}\left(\frac{\rho_1\alpha_1 A_1}{(\lambda_1 A_1-\theta)}\cdot\frac{\rho_2\alpha_2}{(\lambda_2 A_2-\theta)}\right)+\frac{\rho_2\alpha_2 A_2}{(\lambda_2 A_2-\theta)} \\
&\times\left(\frac{(\rho_1\alpha_1\lambda_1)^2}{\theta\,(\lambda_1 A_1-\theta)^2}+\frac{\rho_2\alpha_2\lambda_2}{\theta\,(\lambda_2 A_2-\theta)}\cdot\frac{\rho_1\alpha_1 A_1}{(\lambda_1 A_1-\theta)}+\frac{(\rho_1\alpha_1)^2}{(\lambda_1 A_1-\theta)^2}\right) \\
&>0
\end{aligned}
$$

□

Nota 25. *Si $M\to\infty$, y $\rho_1 = m_1/M$, $\rho_2 = m_2/M$ permanecen constantes, entonces, $I\to I_0$ en donde*

$$
I_0(\theta,\lambda_1,\lambda_2)=\begin{bmatrix}
\frac{\rho_1\alpha_1\lambda_1}{\theta(\lambda_1A_1-\theta)}+\frac{\rho_2\alpha_2\lambda_2}{\theta(\lambda_2A_2-\theta)} & -\frac{\rho_1\alpha_1}{(\lambda_1A_1-\theta)} & -\frac{\rho_2\alpha_2}{(\lambda_2A_2-\theta)} \\
-\frac{\rho_1\alpha_1}{(\lambda_1A_1-\theta)} & \frac{\rho_1\alpha_1A_1}{(\lambda_1A_1-\theta)} & 0 \\
-\frac{\rho_2\alpha_2}{(\lambda_2A_2-\theta)} & 0 & \frac{\rho_2\alpha_2A_2}{(\lambda_2A_2-\theta)}
\end{bmatrix}
$$

Nota 26. *Antes de establecer el teorema de normalidad asintótica para el vector de estimaciones $\hat{\varphi} = (\hat{\theta},\hat{\lambda}_1,\hat{\lambda}_2)$ es necesario aclarar el significado de la expresión $M\to\infty$. Como m_1 y m_2 son los tamaños de muestra en cada espacio, entonces si $\rho_1 = m_1/M$, $\rho_2 = m_2/M$ permanecen constantes y $M\to\infty$, se tiene que m_1 y m_2 tiende simultáneamente a infinito. Esto implica que el número de áreas S_1 de Ω_{A_1} y el número de áreas S_2 de Ω_{A_2} deben tender ambos a infinito. Es posible tener sucesiones infinitas de áreas a_{1i} en Ω_{A_1} y a_{1j} en Ω_{A_2}, pero en este caso la medida de todas las áreas no puede ser igual, manteniendo el área total finita. Por lo tanto, cuando se requiere que $M\to\infty$ para el caso de áreas iguales se debe suponer que $S_1\to\infty$ y $S_2\to\infty$ simultáneamente, en consecuencia $A_1=\mu(\Omega_{A_1})\to\infty$ y $A_2=\mu(\Omega_{A_2})\to\infty$ simultáneamente y de acuerdo con las hipótesis A.1 a A.4 de la sección $N_1\to\infty$ y $N_2\to\infty$ simultáneamente. Esto significa que el siguiente resultado asintótico es aplicable para muestras grandes para tamaños de marco grandes y medidas de los espacios asociados grandes.*

Teorema 27. *Sea $\hat{\varphi} = (\hat{\theta},\hat{\lambda}_1,\hat{\lambda}_2)$ el vector de estimaciones máximo verosímil del vector de parámetros $\varphi = (\theta,\lambda_1,\lambda_2)$ obtenido para la función de verosimilitud Ψ, entonces$\sqrt{M}\,(\hat{\varphi}-\varphi)$ tiene asintóticamente distribución normal multivariada cuando $M\to\infty$, $m_i=\rho_i M$, $i=1,2$, con media cero y matriz de varianzas-covarianzas $I_0^{-1}(\theta,\lambda_1,\lambda_2)$. En símbolos*

$$
\sqrt{M}\,(\hat{\varphi}-\varphi)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},I_0^{-1}(\theta,\lambda_1,\lambda_2)\right) \tag{114}
$$

En el siguiente lema se exhibe explícitamente la matriz $I_0^{-1}(\theta, \lambda_1, \lambda_2)$.

Teorema 28. *La matriz* $I^{-1}(\theta, \lambda_1, \lambda_2)$*.está dada por*

$$I^{-1}(\theta, \lambda_1, \lambda_2) = \tag{115}$$

$$\begin{bmatrix} \frac{\theta A_1 A_2}{\rho_2\alpha_2 A_1+\rho_1\alpha_1 A_2} & \frac{\theta A_2}{\rho_2\alpha_2 A_1+\rho_1\alpha_1 A_2} & \frac{\theta A_1}{\rho_2\alpha_2 A_1+\rho_1\alpha_1 A_2} \\ \frac{\theta A_2}{\rho_2\alpha_2 A_1+\rho_1\alpha_1 A_2} & \frac{1}{\rho_1\alpha_1}\left[\lambda_1 - \frac{\theta\rho_2\alpha_2}{\rho_2\alpha_2 A_1+\rho_1\alpha_1 A_2}\right] & \frac{\theta A_1 A_2}{\rho_2\alpha_2 A_1+\rho_1\alpha_1 A_2} \\ \frac{\theta A_1}{\rho_2\alpha_2 A_1+\rho_1\alpha_1 A_2} & \frac{\theta A_1 A_2}{\rho_2\alpha_2 A_1+\rho_1\alpha_1 A_2} & \frac{1}{\rho_2\alpha_2}\left[\lambda_2 - \frac{\theta\rho_1 a_1}{\rho_2\alpha_2 A_1+\rho_1\alpha_1 A_2}\right] \end{bmatrix}$$

Prueba. El resultado puede verificarse efectuando el producto $I \cdot I^{-1}$ □

Teorema 29. *Si* $C_\Psi(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2)$ *es la matriz de varizanza - covarianza de los estimadores definida en (94), entonces.*

$$I^{-1}(\theta, \lambda_1, \lambda_2) = M \cdot C_\Psi(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) \tag{116}$$

Prueba. Se deduce de los resultados anteriores. □

Para el caso de la función Φ la matriz de información se construye de la siguiente forma. Para cada función de probabilidad f_{1i} $(i = 1, ...m_1)$ la matriz de infromación asociada al vector parámetro es

$$I_{1i}(\theta, \lambda_1, \lambda_2) = \begin{bmatrix} \frac{a_{1i}\lambda_1}{\theta(\lambda_1 A_1 - \theta)} & -\frac{a_{1i}}{(\lambda_1 A_1 - \theta)} & 0 \\ -\frac{a_{1i}}{(\lambda_1 A_1 - \theta)} & \frac{a_{1i} A_1}{(\lambda_1 A_1 - \theta)} & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{117}$$

Sea

$$I_1(\theta, \lambda_1, \lambda_2) = \sum_{i=1}^{m_1} I_{1i}(\theta, \lambda_1, \lambda_2). \tag{118}$$

Entonces $I_1(\theta, \lambda_1, \lambda_2)$ tiene al misma forma de la matriz $I^{(1)}((\theta, \lambda_1, \lambda_2)$ de la ecuación (110). De la misma forma se construyen las matrices de información para cada función de probabilidad f_{2j} $(j = 1, ...m_2)$ asociadas al vector parámetro. Si tales matrices se notan $I_{2j}(\theta, \lambda_1, \lambda_2)$ sea

$$I_2(\theta, \lambda_1, \lambda_2) = \sum_{i=1}^{m_2} I_{2j}(\theta, \lambda_1, \lambda_2) \tag{119}$$

Entonces $I_2(\theta, \lambda_1, \lambda_2)$ tiene al misma forma de la matriz $I^{(2)}((\theta, \lambda_1, \lambda_2)$ de la ecuación (111). Finalmente, la matriz de información asociada a todas las observaciones está dada por

$$I(\theta, \lambda_1, \lambda_2) = I_1(\theta, \lambda_1, \lambda_2) + I_2(\theta, \lambda_1, \lambda_2) \tag{120}$$

Esta matriz tiene la misma forma de la matriz de la ecuación (112). Basta hacer las identificaciones $a_i = \rho_i \alpha_i$ y se llega la resultado, es decir, para este caso se tiene que

$$I(\theta, \lambda_1, \lambda_2) = \begin{bmatrix} \frac{a_1\lambda_1}{\theta(\lambda_1 A_1-\theta)} + \frac{a_2\lambda_2}{\theta(\lambda_2 A_2-\theta)} & -\frac{a_1}{(\lambda_1 A_1-\theta)} & -\frac{a_2}{(\lambda_2 A_2-\theta)} \\ -\frac{a_1}{(\lambda_1 A_1-\theta)} & \frac{a_1 A_1}{(\lambda_1 A_1-\theta)} & 0 \\ -\frac{a_2}{(\lambda_2 A_2-\theta)} & 0 & \frac{a_2 A_2}{(\lambda_2 A_2-\theta)} \end{bmatrix} \tag{121}$$

De manera similar al caso de la matriz C_Ψ se obtiene el siguiente

Teorema 30. *Si $C_\Phi(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2)$ es la matriz de varizanzas - covarianzas de los estimadores definida en (93), entonces.*

$$I^{-1}(\theta, \lambda_1, \lambda_2) = C_\Phi(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) \tag{122}$$

Nota 31. *Obsérvese que para todos los estimadores (49) a (53) la varianza asintótica coincide con sus varianzas. Nótese que se tiene*

$$\sqrt{M}\left(\hat{\theta}-\theta\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\theta A_1 A_2}{\rho_2\alpha_2 A_1 + \rho_1\alpha_1 A_2}\right) \tag{123}$$

$$\sqrt{M}\left(\hat{\lambda}_1-\lambda\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\rho_1\alpha_1}\left(\lambda_1 - \frac{\theta\rho_2\alpha_2}{\rho_2\alpha_2 A_1 + \rho_1\alpha_1 A_2}\right)\right) \tag{124}$$

$$\sqrt{M}\left(\hat{\lambda}_2-\lambda\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\rho_2\alpha_2}\left(\lambda_2 - \frac{\theta\rho_1 a_1}{\rho_2\alpha_2 A_1 + \rho_1\alpha_1 A_2}\right)\right) \tag{125}$$

$$\sqrt{M}\left(\hat{\mathbf{N}}_1-N_1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{A_1^2}{\rho_1\alpha_1}\left(\lambda_1 - \frac{\theta\rho_2\alpha_2}{\rho_2\alpha_2 A_1 + \rho_1\alpha_1 A_2}\right)\right) \tag{126}$$

$$\sqrt{M}\left(\hat{\mathbf{N}}_2-N_2\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{A_2^2}{\rho_2\alpha_2}\left(\lambda_2 - \frac{\theta\rho_1 a_1}{\rho_2\alpha_2 A_1 + \rho_1\alpha_1 A_2}\right)\right) \tag{127}$$

o equivalentemente

$$\left(\hat{\theta}-\theta, \hat{\mathbf{N}}_1 - N_1, \hat{\mathbf{N}}_2 - N_2\right) \xrightarrow{d} \mathcal{N}\left(0, C_\Psi\right) \tag{128}$$

Además si B es la matriz definida por

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & A_1 & 0 \\ 0 & 0 & A_2 \end{pmatrix} \tag{129}$$

entonces

$$\left(\hat{\theta} - \theta, \hat{\mathbf{N}}_1 - N_1, \hat{\mathbf{N}}_2 - N_2\right) \xrightarrow{d} \mathcal{N}\left(0, BC_\Psi B\right) \tag{130}$$

Teorema 32. *Todos los estimadores son de varianza mínima.*

Prueba. Se sigue del teorema anterior. Veáse por ejemplo (Lehmann, 1983: página 128). □

4. Discusión

En la práctica, para la aplicación de los resultados de este trabajo, se debe verificar que la clases $\mathfrak{C}_1$ y $\mathfrak{C}_2$ de subconjuntos de Ω_1 y Ω_2 son tales que se cumplen las hipótesis A.1 a A.4, lo cual se traduce en que la densidad poblacional en cada una de las clases es aproximadamente similar. Entonces, la clase $\mathfrak{C}_i$ está conformada por subconjuntos de Ω_i para los cuales se satisfacen tales condiciones. El conjunto $\mathcal{A}_i$ puede corresponder a una delimitación externa, por ejemplo, una delimitación geográfica. Un primer camino de generalización de este trabajo consiste en abarcar el caso cuando las hipótesis A.1 a A.4 se cumplen localmente. En este caso se pueden proponer varias clases $\mathfrak{C}_{ik}$ para el espacio i en cada una de las cuales las hipótesis se satisfacen localmente. Para este caso puede admitirse una densidad λ_{ik} en cada clase tal que $N_i = \lambda_{ik} \cdot \mu_i(\Omega_{A_{ik}})$. Con las respectivas identificaciones las funciones de probabilidad asociadas al espacio 1 están dadas por

$$\begin{aligned} f_{1ki}(\eta_{1ki}, X_{1ki} \mid \lambda_{1k}, \theta) \quad &= \quad \binom{\eta_{1ki}}{X_{1ki}} \left(\frac{\lambda_{1k}A_{1k} - \theta}{\lambda_{1k}A_{1k}}\right)^{X_{1i}} \left(\frac{\theta}{\lambda_{1k}A_{1k}}\right)^{\eta_{1ki} - X_{1ki}} \\ &\quad \times e^{-\lambda_{1k}a_{1ki}} (\lambda_{1k}a_{1ki})^{\eta_{1ki}} / \eta_{1ki}! \end{aligned}$$

La normalidad asintótica de los estimadores para el caso de la función de verosimiltud Φ no fue demostrada en este trabajo. Todos los resultados parecen indicar que bajo algunos supuestos adicionales se puede llegar a demostrar. En los trabajos de Bradley y Gart (1962) y Hodley (1977) se proponen algunos supuestos que pueden ayudar en este y en el problema de generalización anterior.

Con respecto a las clases $\mathcal{T}_i$ de subconjuntos de $\mathfrak{F}_{A_i}$, es decir a las áreas que forman la base para el muestreo, no es necesario que sean finitas. Por otro lado, en la expresión de convergencia asintótica $\sqrt{M}\,(\hat{\varphi} - \varphi) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, I_0^{-1}(\theta, \lambda_1, \lambda_2)\right)$ debe recordarse que M es el número total de áreas muestradas en los dos espacios y que se supone que $\rho_i = mi/M$ permanece constante. En consecuencia, como el área total Ω_{A_i} tiene medida finita, para poder aplicar este resultado en la práctica se requiere que las clases $\mathcal{T}_1$ y $\mathcal{T}_2$ tengan cada una un número grande de elementos, y por por tanto, cada uno con medida muy pequeña comparada con la medida de el área Ω_{A_i}. Adicionalmente, este resultado implica que para seleccionar tamaños de muestra, es mejor definir desde el comienzo que las muestras sean proporcionales a dos constantes fijas. Asi, parece que una estrategia razonable para el caso de espacios del mismo tipo consiste en tomar las muestras proporcionales a las medidas de Ω_{A_i}.

Por otro lado, para el caso de áreas iguales, se puede tomar $\alpha_1 = \alpha_2 = 1$, debido a que en este caso los α_i solo participan como constantes de proporcionalidad en el cálculo de $\hat{\lambda}_1$ y $\hat{\lambda}_2$, pero no participan en los estimadores para θ, N_1 y N_2. Observe que si S_1 y S_2 representan el número total de áreas en cada espacio, entonces para cualquier valor de α_1 y α_2 se tiene que

$$\hat{\theta} = \frac{(\eta_1 - X_1) + (\eta_2 - X_2)}{\frac{m_1}{S_1} + \frac{m_2}{S_2}} \tag{131}$$

$$\hat{\lambda}_1 = \frac{1}{\alpha_1}\left[\frac{X_1}{m_1} + \frac{\hat{\theta}}{S_1}\right] \tag{132}$$

$$\hat{\lambda}_2 = \frac{1}{\alpha_2}\left[\frac{X_2}{m_2} + \frac{\hat{\theta}}{S_2}\right] \tag{133}$$

$$\hat{\mathbf{N}}_1 = \frac{S_1}{m_1}X_1 + \hat{\theta} \tag{134}$$

$$\hat{\mathbf{N}}_2 = \frac{S_2}{m_2}X_2 + \hat{\theta} \tag{135}$$

En este caso la matriz de covarianzas toma la forma

$$C_\Psi(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) =$$
$$\begin{bmatrix} \frac{\theta S_1 S_2}{m_2 S_1 + m_1 S_2} & \frac{\theta S_2}{m_2 S_1 + m_1 S_2} & \frac{\theta S_1}{m_2 S_1 + m_1 S_2} \\ \frac{\theta S_2}{m_2 S_1 + m_1 S_2} & \frac{1}{m_1}\left[\lambda_1 - \frac{\theta m_2}{m_2 S_1 + m_1 S_2}\right] & \frac{\theta S_1 S_2}{m_2 S_1 + m_1 S_2} \\ \frac{\theta S_1}{m_2 S_1 + m_1 S_2} & \frac{\theta S_1 S_2}{m_2 S_1 + m_1 S_2} & \frac{1}{m_2}\left[\lambda_2 - \frac{\theta m_1}{m_2 S_1 + m_1 S_2}\right] \end{bmatrix}$$

La generalización del problema a mas de dos marcos puede iniciarse, reemplazando el modelo binomial condicionado por un modelo multinomial condicionado, en donde cada p_i está asociado al número de elementos encontrados que pertenecen a un dominio i para cada área muestrada.

Referencias

[1] T. M. Apostol, *Mathematical Analysis*, Second edition, Ed. Addison Wesley Publishing Company, 1974, USA.

[2] R. R. Bosecker and B. L. Ford, *Multiple frame estimation with stratified overlap domain*, Sample survey research branch research. Division Statistical Reporting Service U.S. Deparment of Agriculture, 1976, Washington

[3] R. A. Bradley and J.J. Gart, *The asymptotic properties of ML estimators when sampling from associated populations*, Biometrika, 49, 1962, 205-214

[4] E. G. Bryant and D. W. King, *Estimation from populations identified by overlapping sampling frames*, Memorias del 120 Congreso Anual de la American Statistical Association, 1960.

[5] R. S. Cochran, *Theory and aplications of multiple frame surveys*, Ph. D. dissertation, 1965, Iowa State University, Ames.

[6] R, S, Cochcran and W. P. Cooke*, The estimation of domain size when sampling frames overlap*, Proceedings of the social science section of The American Statistical Association meeting,1967, Washington D.C.

[7] W. P. Cooke, *Surrogate geometric programing estimation of restricted multinomial proportions*, Communications in Statistics. Sec. Simulation and Computation, **12**,1983, 291-305.

[8] W. A. Fuller and L. F. Burmeister, *Estimators for samples selected from two overlapping frames*, Proceedings of social science section of The American Statistical Asociation, 1972, Montreal.

[9] H. O. Hardley, *Multiple frame surveys,* Proceedings of the social science section of The American Statistical Association meeting,1962, Mineapolis

[10] H. O. Hardley, *Multiple frame metodology and selected aplications*, Sankhaya: The Indian Journal of Statistics, **36**, 1974, 99-118

[11] G. H. Hill, *Associating a reporting unit with a list frame sampling unit in multiple frame sampling*, Ohio and Wisconsin sample survey research branch research. Division Statistical Reporting Service U.S. Deparment of Agriculture, 1977, Washington

[12] B. Hodley, *Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case*, The Annals .of Mathematical Statistics, **42**, 1971, 1977-1991.

[13] D. W. King, *Variance estimators in populations identified by multiple sampling frames*, M.S. Thesis, 1960, University of Wyoming Library, Laramie.

[14] M. L. Krasnov, G. I. Makarenko y A. I. Kiseliov, *Cálculo Variacional*, Ed. Mir, 1976, Moscú.

[15] E. L.Lehman, *Theory of Point Estimation*, Ed. John Wiley & Sons, 1983, New York.

[16] R. E. Lund, *Estimators in multiple frame surveys*, Proceedings of the social science section of the American Statistical Association, 1968, 282-288.

[17] D. Ospina, *Maximum Likelihood aproach on multiple frame surveys*, Ph. D. dissertation, 1985, University of Wyoming Library, Laramie.

[18] D. Ospina, *Una distribución asintótica para estimadores máximo - verosímiles de tamaños de dominios de situaciones de marcos múltiples*, Revista Colombiana de Estadística, **19-20**, 1989, 1-24.

[19] R. E. Williams, *Estimation of Overlapping strata boundaries*, M.S. Thesis, 1957, University of Wyoming Library, Laramie.