Modelo Lineal Generalizado

Propiedades de los Estimadores de Máxima Verosimilitud

Recordemos que si la variable aleatoria Y tiene función de densidad o probabilidad puntual $f(y, \theta)$, la verosimiltud $L(\theta, y)$ es simplemente $f(y, \theta)$ mirada como función de θ con y fijo.

La función de probabilidad o densidad es definida sobre el soporte $y \in \mathcal{Y}$, mientras que la verosimilitud es definida sobre un espacio paramétrico Θ.

En muchos casoS trabajamos con el logaritmo de la verosimilitud (log-likelihood)

$$l(\theta, y) = \log L(\theta, y)$$

que está definido salvo una cosntante aditiva.

En general, tendremos una muestra aleatoria $Y_1, \ldots, Y_n$ con distribución $f(y, \theta)$, de manera que la verosimilitud será:

$$L(\theta) = \prod_{i=1}^{n} f(y_i, \theta)$$

y la log-versosimilitud

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(y_i, \theta)$$

Una propiedad útil de los EMV es la de **invariancia** que dice que si g es una función con inversa g^{-1}, de manera que $\phi = g(\theta)$ implica que $\theta = g^{-1}(\phi)$, entonces el EMV de ϕ, $\widehat{\phi}$, se calcula como

$$\widehat{\phi} = g(\widehat{\theta})$$

siendo $\widehat{\theta}$ el EMV de θ.

Como ya sabemos, podemos maximizar $L(\theta)$ o bien maximizar $l(\theta)$. En problemas regulares, el EMV puede hallarse igualando a 0 las derivadas pimeras de $l(\theta)$ respecto de θ.

La derivada primera de $l(\theta)$ respecto de θ se llama score. En el caso univariado tenemos:

$$l'(\theta) = \sum_{i=1}^{n} u_i(\theta)$$

donde

$$u_i(\theta) = \frac{\partial}{\partial\theta} \log f(y_i, \theta)$$

Si tenemos q parámetros, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$, el vector de score es

$$l'(\theta) = \begin{bmatrix} \frac{\partial l}{\partial\theta_1} \\ \frac{\partial l}{\partial\theta_2} \\ . \\ . \\ \frac{\partial l}{\partial\theta_q} \end{bmatrix} = \begin{bmatrix} \Sigma_{i=1}^{n} \frac{\partial}{\partial\theta_1} log f(y_i, \theta) \\ \Sigma_{i=1}^{n} \frac{\partial}{\partial\theta_2} log f(y_i, \theta) \\ . \\ . \\ \Sigma_{i=1}^{n} \frac{\partial}{\partial\theta_q} log f(y_i, \theta) \end{bmatrix}$$

Una propiedad bien conocida del score es que su esperanza es nula:

$$E(l'(\theta))|_{\theta=\theta_0} = \int l'((\theta_0))f(y, \theta_0)dy = \int \frac{f'(y, \theta_0)}{f(y, \theta_0)} f(y, \theta_0)dy = 0$$

La varianza de los score $u(\theta)$ es conocida como la **información de Fisher**. En el caso univariado, la información de Fisher es:

$$\begin{aligned} i(\theta) &= V(u(\theta)) = V(l'(\theta)) \\ &= E(u^2(\theta)) \end{aligned}$$

Recordemos que

$$i(\theta) = E(-l''(\theta)) = -E(\frac{\partial^2}{\partial\theta^2} log f(y,\theta))$$

En el caso multivariado, $i(\theta)$ es una matriz de $q \times q$ tal que:

$$\{i(\theta)\}_{ij} = -E(\frac{\partial^2}{\partial\theta_i\partial\theta_j} log f(y,\theta))$$

En Estadística se probó que, bajo condiciones de regularidad, los EMV son asintóticamente normales, de manera que

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, i^{-1}(\theta))$$

Estimación y Tests de Bondad de Ajuste

Supongamos que tenemos un muestreo multinomial y obtenemos la tabla (X, Y) en n individuos.

	$Y=1$	$Y=2$		$Y=J$
$X=1$	n_{11}	n_{12}		n_{1J}
$X=2$	n_{21}	n_{22}		n_{2J}
.	.	.	. .	.
.	.	.	. .	.
.	.	.	. .	.
$X=I$	n_{I1}	n_{I2}		n_{IJ}

Sea n_{ij} el número de individuos que tienen $P(X=i, Y=j)$, de manera que $n = \Sigma_{i=1}^{I} \Sigma_{j=1}^{J} n_{ij}$.

Por lo que ya vimos los estimadores de máxima verosimilitud de π_{ij} son

$$\widehat{\pi}_{ij} = \frac{n_{ij}}{n} \quad \forall i, j \, .$$

Si las dos variables categóricas fueran independientes, tendríamos

$$\pi_{ij} = \pi_i \pi_j \quad \forall i, j \, ,$$

luego como veremos por invarianza, el estimador de máxima verosimilitud de π_{ij} sería bajo independencia:

$$\widehat{\pi}_{ij} = \widehat{\pi}_{i+} \widehat{\pi}_{+j} = \frac{n_{i+} n_{+j}}{n^2} \quad \forall i, j \, .$$

Dado que $n_{ij} \sim Bi(n, \pi_{ij})$,

$$m_{ij} = E(n_{ij}) = n\pi_{ij} \, .$$

Bajo el supuesto de independencia, el EMV es

$$\widehat{m}_{ij} = n\widehat{\pi}_{ij} = \frac{n_{i+} n_{+j}}{n}$$

Estos estimadores tienen la propiedad de tener las mismas marginales que la tabla:

$$\widehat{m}_{i+} \; = \; \sum_{j=1}^{J} \frac{n_{i+} n_{+j}}{n} = n_{i+}$$

$$\widehat{m}_{+j} = \sum_{i=1}^{I} \frac{n_{i+} n_{+j}}{n} = n_{+j}$$

Test de Bondad de Ajuste

Veremos un test presentado por Pearson (1900) que sirve para evaluar si una distribución multinomial tiene ciertas probabilidades π_{ij} propuestas.

Para simplificar la notación, como antes, indicaremos $\{n_1, \ldots, n_N\}$ las observaciones de cada casilla, con $n = \sum_{i=1}^{N} n_i$ y siendo $\{\pi_1, \ldots, \pi_N\}$ las probabilidades de cada celda.

Supongamos que las hipótesis a testear son

$$H_0 : \pi_i = \pi_{i0}, \sum_{j=1}^{N} \pi_{i0} = \sum_{j=1}^{N} \pi_i = 1 \quad H_1 : \exists i : \pi_i \neq \pi_{i0}$$

Pearson propuso el siguiente estadístico:

$$\chi^2 = \sum_{j=1}^{N} \frac{(n_i - m_{i0})^2}{m_{i0}} \quad \text{donde } m_{i0} = n\pi_{i0}$$

La idea intuitiva es que comparamos el valor **observado** (n_i) con el valor **esperado** (m_{i0}) bajo H_0, suele decirse :

$$\frac{(observado - esperado)^2}{esperado}.$$

Intuitivamente rechazaremos H_0 cuando esto sea muy grande. ¿Cuán grande?

El argumento heurístico que dio Pearson es el siguiente: si $n_1, \ldots, n_N$ fueran v.a. independientes tales que $n_i \sim \mathcal{P}(m_i)$, bajo ciertas condiciones

$$\frac{n_i - m_i}{\sqrt{m_i}} \overset{a}{\sim} N(0,1)$$

entonces

$$\sum_{i=1}^{N} \left[\frac{n_i - m_i}{\sqrt{m_i}}\right]^2 \overset{a}{\sim} \chi^2_N$$

Si además agregamos la restricción $\Sigma_{i=1}^{N} n_i = n$, es natural que perdamos un grado de libertad y que la distribución asintótica del estadístico resulte χ^2_{N-1}.Por todo esto, la regla de decisión será

$$\text{Rechazamos } H_0 \text{ si } \chi^2 > \chi^2_{N-1,\alpha}$$

En el caso en que $N = 2$, el estadístico queda

$$n\frac{(\widehat{p}-\pi_0)^2}{\pi_0} + n\frac{(\widehat{p}-\pi_0)^2}{1-\pi_0} = n\frac{(\widehat{p}-\pi_0)^2}{\pi_0(1-\pi_0)} = \left[\frac{\widehat{p}-\pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}\right]^2$$

que es el cuadrado del test habitual para testear

$$H_0 : \pi_0 = \pi \quad H_1 : \pi_0 \neq \pi$$

que tiene distribución asintótica normal y en consecuencia, su cuadrado lo compararíamos con una χ^2_1.

Veamos la justificación teórica de este test. Comenzaremos por presentar el Teorema Central del Límite Multivariado, que resulta del caso univariado aplicando la siguiente

Proposición: Sean $\mathbf{X}_n = (X_{1n}, \ldots, X_{kn})'$ una sucesión de vectores aleatorios y $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)' \in \Re^k$.
Si $\forall \boldsymbol{\lambda} \in \Re^k$

$$\boldsymbol{\lambda}'\mathbf{X}_n = \lambda_1 X_{1n} + \ldots + \lambda_k X_{kn} \xrightarrow{\mathcal{D}} \lambda_1 X_1 + \ldots + \lambda_k X_k \,,$$

donde $\mathbf{X} = (X_1, \ldots, X_k)' \sim \mathcal{F}$, entonces la distribución límite de $\mathbf{X}_n = (X_{1n}, \ldots, X_{kn})'$ existe y es $\mathcal{F}$.

Teorema Central del Límite Multivariado (TCLM)

Sea $\mathbf{U}_n = (U_{1n}, \ldots, U_{kn})'$ una sucesión de vectores aleatorios tales que $E(\mathbf{U}_n) = \boldsymbol{\mu}$ y $\Sigma_{\mathbf{U}_n} = \Sigma$, $n = 1, 2, \ldots$

Si $\bar{\mathbf{U}}_n = (\bar{U}_{1n}, \ldots, \bar{U}_{kn})'$ es el vector de promedios, donde para cada $1 \leq i \leq k$ $\bar{U}_{in} = \dfrac{1}{n} \sum_{j=1}^{n} U_{ij}$, entonces

$$\sqrt{n}(\bar{\mathbf{U}}_n - \boldsymbol{\mu}) \xrightarrow{\mathcal{D}} N_k(0, \Sigma) \,.$$

Según la proposición anterior debemos estudiar la distribución de $\boldsymbol{\lambda}'\bar{\mathbf{U}}_n$.

$$
\begin{aligned}
\boldsymbol{\lambda}'\bar{\mathbf{U}}_n &= \lambda_1 \bar{U}_{1n} + \ldots + \lambda_k \bar{U}_{kn} \\
&= \lambda_1 \frac{\Sigma_{j=1}^n U_{1j}}{n} + \ldots + \lambda_k \frac{\Sigma_{j=1}^n U_{kj}}{n} \\
&= \frac{1}{n} \sum_{j=1}^n \boldsymbol{\lambda}'\mathbf{U}_j \\
&= \frac{1}{n} \sum_{j=1}^n W_j \\
&= \bar{W}
\end{aligned}
$$

donde $E(W_i) = \boldsymbol{\lambda}'\boldsymbol{\mu}$, $Var(W_i) = \boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda}$.

Por el TCL univariado, tenemos que

$$\sqrt{n}(\bar{W}_n - \boldsymbol{\lambda}'\boldsymbol{\mu}) \xrightarrow{\mathcal{D}} N_k(0, \boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda})\,,$$

es decir,

$$\sqrt{n}(\boldsymbol{\lambda}'\bar{\mathbf{U}}_n - \boldsymbol{\lambda}'\boldsymbol{\mu}) \xrightarrow{\mathcal{D}} N_k(0, \boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda})\,,$$

que corresponde a la distribución de $\boldsymbol{\lambda}'\mathbf{U}$, con $\mathbf{U} \sim N_k(0, \Sigma)$ por lo que

$$\sqrt{n}(\bar{\mathbf{U}}_n - \boldsymbol{\mu}) \xrightarrow{\mathcal{D}} N_k(0, \Sigma)\,.$$

Ahora estudiaremos la distribución asintótica de $(X_1, \ldots, X_{N-1})$ cuando $(X_1, \ldots, X_N)' \sim M(n, \pi_1, \ldots, \pi_N)$, $\sum\limits_{i=1}^{N} \pi_i$. Llamemos $\mathbf{p} = (p_1, \ldots, p_N)'$, $p_i = \dfrac{n_i}{n}$.

Consideremos el vector $\mathbf{Y}_i \sim M(1, \pi_1, \ldots, \pi_N)$ que ya definimos con todas sus componentes iguales a 0 y un único 1 en la coordenada j-ésima si en la i-ésima observación ocurrió la categoría j:

$$\mathbf{Y}_i = (0, \ldots, \underset{j}{\overset{\downarrow}{1}}, \ldots, 0)$$

Recordemos que si $\mathbf{Y}_i \sim M(1, \pi_1, \ldots, \pi_N)$

$$\begin{aligned} E(\mathbf{Y}_i) &= \mathbf{\Pi} \\ \mathbf{\Sigma}_{\mathbf{Y}_i} &= \Delta(\mathbf{\Pi}) - \mathbf{\Pi}\mathbf{\Pi}' \end{aligned}$$

Podemos escribir al vector $\mathbf{p}$ en términos de los $\mathbf{Y}_i$:

$$\mathbf{p} = \sum_{i=1}^{n} \frac{\mathbf{Y}_i}{n} = (\bar{Y}_1, \ldots, \bar{Y}_N)$$

entonces por el T.C.L multivariado sabemos que

$$\sqrt{n}(\mathbf{p} - \boldsymbol{\pi}) \xrightarrow{\mathcal{D}} N_N(0, \Delta(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}')\,.$$

Ya hemos visto que como los π_i's están relacionados, $\Sigma = \Delta(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'$ no es invertible.

Definamos $\tilde{\mathbf{p}} = (p_1, \ldots, p_{N-1})'$ y $\tilde{\boldsymbol{\pi}} = (\pi_1, \ldots, \pi_{N-1})'$.

Notemos que $\tilde{\mathbf{p}} = \mathbf{T}\mathbf{p}$, siendo $\mathbf{T}$ es una transformación lineal adecuada, luego aplicando el T.C.L. multivariado a $\mathbf{T}\mathbf{p}$

$$\sqrt{n}(\tilde{\mathbf{p}} - \tilde{\boldsymbol{\pi}}) \xrightarrow{\mathcal{D}} N_{N-1}(0, \Delta(\tilde{\boldsymbol{\pi}}) - \tilde{\boldsymbol{\pi}}\tilde{\boldsymbol{\pi}}')\,.$$

Ahora, $\Delta(\tilde{\boldsymbol{\pi}}) - \tilde{\boldsymbol{\pi}}\tilde{\boldsymbol{\pi}}')$ sí es invertible

Esto quiere decir que bajo H_0

$$\sqrt{n}(\tilde{\mathbf{p}} - \tilde{\boldsymbol{\pi}}_0) \xrightarrow{\mathcal{D}} N_{N-1}(0, \tilde{\Sigma}_0)\,.$$

donde $\tilde{\Sigma}_0 = \Delta(\tilde{\boldsymbol{\pi}}_0) - \tilde{\boldsymbol{\pi}}_0\tilde{\boldsymbol{\pi}}_0'$. Por lo tanto, como ya sabemos

$$n(\tilde{\mathbf{p}} - \tilde{\boldsymbol{\pi}}_0')\,\tilde{\Sigma}_0^{-1}\,(\tilde{\mathbf{p}} - \tilde{\boldsymbol{\pi}}_0) \xrightarrow{\mathcal{D}} \chi^2_{N-1}$$

Calculando efectivamente la forma cuadrática que estamos considerando, veremos que

$$n(\tilde{\mathbf{p}} - \tilde{\boldsymbol{\pi}}_0')\,\tilde{\Sigma}_0^{-1}\,(\tilde{\mathbf{p}} - \tilde{\boldsymbol{\pi}}_0) = n\sum_{j=1}^{N} \frac{(p_i - \pi_{i0})^2}{\pi_{i0}}$$

Ejemplo: Leyes de Mendel

El test de Pearson fue usado para testear las leyes de herencia de la teoría de Mendel. Mendel cruzó arvejas de cepa amarilla con arvejas de cepa verde puras y predijo que la segunda generación de híbridos serían un 75 % amarillas y un 25 % verdes, siendo las amarillas las de carácter dominante.

En un experimento de $n = 8023$ semillas, resultaron $n_1 = 6022$ amarillas y $n_2 = 2001$ verdes. Las frecuencias relativas esperadas eran $\pi_1 = 0{,}75$ y $\pi_2 = 0{,}25$, por lo tanto $m_1 = 6017{,}25$ y $m_2 = 20005{,}75$.

Luego, si queremos testear la hipótesis nula

$$H_0 : \pi_1 = 0.75, \;\; \pi_2 = 0.25$$

el estadístico χ^2 es:

$$\begin{aligned} \chi^2 &= \frac{(n_1 - 6017.25)^2}{6017.25} + \frac{(n_2 - 2005.75)^2}{2005.75} \\ &= 0.015 \end{aligned}$$

con un **p-valor=0.88**, lo que **no contradice la teoría de Mendel**.

Cuando $\boldsymbol{\pi}$ puede yacer en cualquier lugar de $\mathcal{S}$ decimos que el modelo es saturado. Este modelo tiene $N-1$ parámetros. Sin embargo, con freciencia supondremos que $\boldsymbol{\pi}$ yace en un subconjunto de menor dimensión de $\mathcal{S}$. Supondremos que los elementos de $\boldsymbol{\pi}$ están determinados por q parámetros desconocidos $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$, como muestran los siguientes ejemplos.

Test de Independencia

Independencia en una tabla de 2×2

Supongamos que $\mathbf{X} = (X_{11}, X_{12}, X_{21}, X_{22})'$ es el vector de frecuencias de una tabla de 2×2:

	$B=1$	$B=2$
$A=1$	X_{11}	X_{12}
$A=2$	X_{21}	X_{22}

De manera que X_{ij} es el número de individuos para los cuales $(A=i, B=j)$. Si A

y B no están relacionados, entonces en todas las casillas valdrá:

$$\pi_{ij} = P(A = i, B = j) = P(A = i)P(B = j)$$

Llamemos $\alpha = P(A = i)$ y $\beta = P(B = j)$, luego

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} \\ \pi_{12} \\ \pi_{21} \\ \pi_{22} \end{bmatrix} = \begin{bmatrix} \alpha\beta \\ \alpha(1-\beta) \\ (1-\alpha)\beta \\ (1-\alpha)(1-\beta) \end{bmatrix}$$

Este es un modelo restringido que depende del parámetro

$$\theta = (\alpha, \beta)\,,$$

donde $0 \leq \alpha \leq 1, \quad 0 \leq \beta \leq 1.$

Para hallar los estimadores de máxima verosimilitud de α y β tenemos que maximizar:

$$\begin{aligned} L &= L(X_{11}, X_{12}, X_{21}, X_{22}, \alpha, \beta) = \\ &= \frac{n!}{X_{11}!X_{12}!X_{21}!X_{22}!}(\alpha\beta)^{X_{11}}(\alpha(1-\beta))^{X_{12}}((1-\alpha)\beta)^{X_{21}}((1-\alpha)(1-\beta))^{X_{22}} \end{aligned}$$

Después de tomar logaritmo, obtenemos:

$$\begin{aligned} l = \ln(L) = cte \; & + \; X_{11}\ln(\alpha\beta) + X_{12}\ln(\alpha(1-\beta)) \\ & + \; X_{21}\ln((1-\alpha)\beta) + X_{22}\ln((1-\alpha)(1-\beta)) \end{aligned}$$

Después de derivar e igualar a 0, queda:

$$\begin{aligned} (1) \; : \; \frac{\partial l}{\partial \alpha} &= \frac{X_{11}+X_{12}}{\alpha} - \frac{X_{21}+X_{22}}{1-\alpha} = 0 \\ (2) \; : \; \frac{\partial l}{\partial \beta} &= \frac{X_{11}+X_{21}}{\beta} - \frac{X_{12}+X_{22}}{1-\beta} = 0 \end{aligned}$$

De (1) resulta:

$$(1-\alpha)(X_{11}+X_{12}) - \alpha(X_{21}+X_{22}) = 0\,,$$

por lo tanto

$$\widehat{\alpha} = \frac{X_{11}+X_{12}}{n}\,.$$

De (2) resulta:

$$(1-\beta)(X_{11}+X_{21})-\beta(X_{12}+X_{22})=0\,,$$

por lo tanto

$$\widehat{\beta}=\frac{X_{11}+X_{21}}{n}\,.$$

En el caso general de una tabla de $I \times J$, el modelo sería $\pi_{ij} = \pi_{i+}\pi_{+j}$.

Test de Independencia

Vimos que en las tablas de contingencia con muestreo multinomial puede ser de interés testear la hipótesis de independencia, es decir:

$$H_0 : \pi_{ij} = \pi_{i+}\pi_{+j} \quad \forall i, j$$

La hipótesis nula depende de ciertos parámetros.

Por esto si bien para testear esta hipótesis usaremos un test de tipo Pearson, antes será necesario probar algunos resultados.

Otro ejemplo es el de las tablas simétricas.

Ejemplo: Tabla de 2×2 con simetría

Consideremos $\mathbf{X} = (X_{11}, X_{12}, X_{21}, X_{22})'$ como en el ejemplo anterior, pero supongamos) que ahora A y B representan dos características medidas en dos oportunidades distintas. Por ejemplo, A podría ser la respuesta a

A : **¿Apoya usted la gestión de gobierno?**

medida en el mes de enero (1=Si, 0=No) y B la misma pregunta hecha tres meses después.

	Febrero	
Enero	1	0
1	π_{11}	π_{12}
0	π_{21}	π_{22}

En este tipo de esquema el interés del investigador es detectar un cambio en el tiempo. Si no hubiera ningún cambio, la probabilidad de "Si"en enero

$$P(A = 1) = \pi_{11} + \pi_{12}$$

sería igual a la probabilidad de "Si"tres meses después

$$P(B = 1) = \pi_{11} + \pi_{21}\,.$$

Observemos $P(A = 1) = P(B = 1)$ si y sólo si $\pi_{12} = \pi_{21}$, que se conoce como la condición de **simetría**. Bajo simetría, $\boldsymbol{\pi} = \boldsymbol{\pi}(\theta)$ podría expresarse como:

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} \\ \pi_{12} \\ \pi_{21} \\ \pi_{22} \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \\ \beta \\ 1 - \alpha - 2\beta \end{bmatrix},$$

con $\theta = (\alpha, \beta)'$.

Será un ejercicio de la práctica probar que los EMV bajo este modelo son:

$$\widehat{\alpha} = \frac{X_{11}}{n}\,.$$

$$\widehat{\beta} = \frac{X_{12} + X_{21}}{2n}\,.$$

Caso Paramétrico General (Rao, Capítulo 5)

Aún cuando en los dos ejemplos anteriores los EMV tienen una forma cerrada, en otros modelos más complicados, los EMV deben ser computados en forma iterativa.

En general, $\widehat{\theta}$ es solución de las ecuaciones de ”score”:

$$\frac{\partial}{\partial \theta_j}\, l(\boldsymbol{\pi}(\theta), X) = 0\,, \quad j = 1, \ldots, q\,. \tag{1}$$

Bajo condiciones de regularidad del EMV $\widehat{\theta}$ que precisaremos, demostraremos que

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N_q(0, (\mathbf{A}'\mathbf{A})^{-1})$$

donde

$$\mathbf{A} = \Delta(\boldsymbol{\pi}_0)^{-1/2} \left(\frac{\partial \boldsymbol{\pi}}{\partial \theta}\right)_{\theta=\theta_0}\,.$$

Este resultado lo deduciremos expresando a $\widehat{\theta}$ en términos de p y luego aplicando el método Δ.

Esto nos permitirá derivar la distribución del estadístico χ^2 en casos bastante generales.

Para que el EMV de θ exista, es necesaria una **condición de identificabilidad fuerte**:

Dado $\delta > 0$, existe $\epsilon > 0$ tal que

$$\inf_{\|\theta-\theta_0\|>\delta} \sum_{i=1}^{N} \pi_i(\theta_0) \log \frac{\pi_i(\theta_0)}{\pi_i(\theta)} \geq \epsilon$$

Esta condición implica que fuera de la bola $\|\theta - \theta_0\| \leq \delta$ no hay ninguna sucesión de puntos θ_r tal que $\boldsymbol{\pi}(\theta_r) \to \boldsymbol{\pi}(\theta_0)$ a medida que $r \to \infty$, es decir que no hay valores θ lejanos a θ_0 que den practicamente las mismas probabilidades que $\boldsymbol{\pi}(\theta_0)$.

Es decir:

$$\forall \delta > 0, \text{ exite } \epsilon > 0 \text{ tal que si } \|\theta - \theta_0\| > \delta \text{ entonces } \|\boldsymbol{\pi}(\theta) - \boldsymbol{\pi}(\theta_0)\| > \epsilon\,.$$

Esta condición implica una más débil:

$$\boldsymbol{\pi}(\theta) \neq \boldsymbol{\pi}(\beta) \text{ si } \theta \neq \beta\,.$$

Bajo la condición fuerte de identificabilidad y continuidad de las funciones $\pi_i(\theta)$, se puede demostrar que el EMV de θ existe y que converge a θ_0 con probabilidad 1.

Más aún, si las funciones $\pi_i(\theta)$ tienen derivadas parciales de primer orden, se puede demostrar que el EMV es solución de

$$\frac{\partial L}{\partial \theta_j} = 0\,, \quad j = 1, \ldots, q\,. \tag{2}$$

Por último, si $\boldsymbol{\pi}(\theta) \neq \boldsymbol{\pi}(\beta)$ si $\theta \neq \beta$, las funciones $\pi_i(\theta)$ tienen derivadas parciales de primer orden continuas en θ_0 y si la matriz de información $\mathcal{I}$

$$\mathcal{I}_{rs} = \sum_{i=1}^{N} \frac{1}{\pi_i} \frac{\partial \pi_i}{\partial \theta_r} \frac{\partial \pi_i}{\partial \tilde{\theta}_s}$$

evaluada en θ_0 no es singular, entonces existe una raíz consistente de (2) que puede no ser un EMV, pero que es eficiente en el sentido que definiremos y que tiene distribución asintótica normal

En este contexto hablaremos de eficiencia asintótica en el siguiente sentido.

Definición: Sea $P(x_1, \ldots, x_n, \theta)$ es una función de probabilidad de las variables aleatorias $X_1, \ldots, X_n$. Sea $Z_n = (z_n^1, \ldots, z_n^q)'$ el vector de derivadas

$$z_n^i = \frac{1}{n} \frac{\partial \ln P}{\partial \theta_i} = \frac{1}{n} \frac{\partial l}{\partial \theta_i} \quad i = 1, \ldots, q$$

y el vector de sesgos

$$B_n = (T_n - \theta)' = (T_n^1 - \theta_1, \ldots, T_n^q - \theta_q)'$$

siendo $T_n = (T_n^1, \ldots, T_n^q)'$ es un estimador consistente de $(\theta_1, \ldots, \theta_q)'$. Diremos que T_n es un estimador eficiente si

$$\sqrt{n} \| B_n - \mathbf{D} Z_n \| \xrightarrow{p} 0 \tag{3}$$

o a.s., donde $\mathbf{D}$ es una matriz constante que puede depender de θ.

- ¿Qué diría esta definición en el caso univariado?

  Supongamos que $p(x_1, \ldots, x_n, \theta)$ es la densidad conjunta de las variables aleatorias $X_1, \ldots, X_n$ y llamemos

$$Z_n = \frac{1}{n} \frac{\partial \log p(x_1, \ldots, x_n, \theta)}{\partial \theta}$$

  T_n, estimador consistente de θ, será eficiente si

$$\sqrt{n}(T_n - \theta - \delta(\theta) Z_n) \xrightarrow{p} 0 \qquad (4)$$

  en probabilidad o con probabilidad 1, donde $\delta(\theta)$ no involucra a las observaciones.

- ¿Qué interés tiene esta definición?

  Tenemos, por ejemplo, el siguiente resultado a partir de esta definición.

  Sean $X_1, \ldots, X_n$ una sucesión de v. a. i.i.d. con densidad $p(x, \theta)$, siendo$\theta \in \Re$. Además supongamos que

$$\int p'(x, \theta) dx = 0 \qquad \int \frac{p'^2}{p} dx = i(\theta) > 0$$

Entonces, la condición (4) implica que la distribución asintótica de $\sqrt{n}(T_n - \theta)$ es normal:

De hecho, por el T.C.L. la distribución asintótica de

$$\sqrt{n} z_n = \frac{1}{\sqrt{n}} \left(\frac{p'(X_1, \theta)}{p(X_1, \theta)} + \ldots + \frac{p'(X_n, \theta)}{p(X_n, \theta)} \right) \xrightarrow{p} N(0, i(\theta))$$

Por (4), $\sqrt{n}(T_n - \theta)$ tiene la misma distribución que $\sqrt{n}\delta(\theta) Z_n$, de donde deducimos la normalidad asintótica de $\sqrt{n}(T_n - \theta)$.

- Caso Multivariado Bajo condiciones similares a las enunciadas en el ítem anterior, la distribución asintótica de $\sqrt{n}\mathbf{Z}_n$ es normal q-variada con media 0 y matriz de covarianza $\mathcal{I} = \{i_{rs}\}$, matriz de información.

  Luego, si $\|\mathbf{D}\| \neq 0$, la condición (3) implica que $\sqrt{n}(T_n - \theta)$ es también normal con distribución normal q–variada y con matriz de covarianza $\mathbf{D}\mathcal{I}\mathbf{D}$.

Supondremos que las casillas tienen distribución multinomial con probabilidad $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)'$, donde $\boldsymbol{\pi} = \boldsymbol{\pi}(\theta) = \pi(\theta_1, \ldots, \theta_q)'$. Deduciremos un resultado análogo al que ya vimos para el caso no paramétrico cuando $\widehat{\boldsymbol{\pi}} = \boldsymbol{\pi}(\widehat{\theta})$ y calcularemos grados de libertad de la χ^2 correspondiente.

El resultado que probaremos es el siguiente:

Teorema: Supongamos que las probabilidades de las casillas son $\pi_1(\theta), \ldots, \pi_N(\theta)$ que involucran a q parámetros $\theta = (\theta_1, \ldots, \theta_q)'$. Además, supongamos que:

- a) $\widehat{\theta}$ es un estimador eficiente de θ en el sentido de (3)
- b) Cada $\pi_i(\theta)$ admite derivadas parciales continuas de 1^{er} orden respecto de θ_j, $j = 1, \ldots, q$, $i = 1, \ldots, N$.
- c) La matriz $\mathbf{M} = \{\dfrac{\partial \pi_r}{\partial \theta_s}\}$ de $N \times q$ calculada en los verdaderos θ tiene rango q.

Luego, tenemos que

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \widehat{m}_i)^2}{\widehat{m}_i} = \sum_{i=1}^{N} \frac{(n_i - n\widehat{\pi}_i)^2}{n\widehat{\pi}_i} \quad \xrightarrow{\mathcal{D}} \quad \chi^2_{N-1-q}$$

Comenzaremos por probar el siguiente resultado auxiliar.

Lema: Supongamos que θ_0, valor verdadero del parámetro, es un punto interior del espacio paramétrico, $\pi_i(\theta_0) > 0 \quad \forall i$ y que además se cumplen las siguientes condiciones:

- a) $\pi_i(\theta) \neq \pi_i(\beta)$ para algún i si $\theta \neq \beta$ (condición de identificabilidad).
- b) $\pi_i(\theta)$ admite derivadas parciales de 1^{er} orden continuas en θ_0.
- c) La matriz $\{i_{rs}\}$ es no singular en θ_0, donde

$$i_{rs} = \sum_j \frac{1}{\pi_j} \frac{\partial \pi_j}{\partial \theta_r} \frac{\partial \pi_j}{\partial \theta_s}$$

Luego, existe una raíz consistente $\tilde{\theta}$ de $\dfrac{\partial l}{\partial \theta_i} = 0 \quad i = 1, \ldots, q$, ecuación de verosimilitud y

$$\sqrt{n}\,|\tilde{\theta}_r - \theta_{0r} - i^{r1} Z_1 - \ldots - i^{rq} Z_q| \xrightarrow{p} 0 \quad j = 1, \ldots, q$$

donde $i^{rs} = \{\mathcal{I}^{-1}\}_{rs}$

$$Z_j = \sum_{i=1}^{N} \frac{p_i}{\pi_i(\theta_0)} \frac{\partial \pi_i}{\partial \theta_{0j}}$$

Es decir, el estimador derivado de la ecuación $\dfrac{\partial l}{\partial \theta_i} = 0$ es eficiente y su distribución asintótica es normal q–variada, siendo

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N_q(0, (\mathbf{A}'\mathbf{A})^{-1})$$

donde

$$\mathbf{A} = \Delta(\boldsymbol{\pi}_0)^{-1/2} \left(\frac{\partial \boldsymbol{\pi}}{\partial \theta}\right)_{\theta=\theta_0}.$$

Esquema de demostración

- (1) Lemita: Si $\Sigma_{i=1}\, a_i$ y $\Sigma_{i=1}\, b_i$ son series convergentes, donde $a_i > 0$ y $b_i > 0$, tal que $\Sigma_{i=1}\, a_i \geq \Sigma_{i=1}\, b_i$, entonces

$$\sum_{i=1}^{N} a_i \log \frac{b_i}{a_i} \leq 0 \qquad \text{si } \sum_{i=1} a_i \geq \sum_{i=1} b_i$$

y la igualdad se alcanza si $a_i = b_i \forall i$.

- (2) Consideremos la función

$$S(\theta) = \sum_{i=1}^{N} \pi_i(\theta_0) \log \frac{\pi_i(\theta_0)}{\pi_i(\theta)}$$

sobre la bola

$$\|\theta - \theta_0\| \leq \delta$$

- (3) Con esto probamos que

$$\inf_{\|\theta-\theta_0\|=\delta} S(\theta) > \epsilon > 0$$

- (4) A partir de (3) usando un argumento de continuidad, vemos que si $\|\theta - \theta_0\| = \delta$

$$\sum_{i=1}^{N} p_i \log \pi_i(\theta_0) > \sum_{i=1}^{N} p_i \log \pi_i(\theta)$$

por lo tanto el máximo se alcanza en $\bar{\theta}$, punto interior de $\|\theta - \theta_0\| \leq \delta$.

- 5) Probamos que para $\tilde{\theta}_s \in (\bar{\theta}_s, \theta_{0s})$

$$\sum_{i=1}^{N} \sqrt{n}\, \frac{p_i - \pi_i(\theta_0)}{\bar{\pi}_i} \, \frac{\partial \pi_i}{\partial \bar{\theta}_r} \; = \; \sum_{i=1}^{N} \sum_{s=1}^{q} \sqrt{n}\, (\bar{\theta}_s - \theta_{0s}) \, \frac{\partial \pi_i}{\partial \bar{\theta}_r} \frac{\partial \pi_i}{\partial \tilde{\theta}_s} \frac{1}{\bar{\pi}_i} \tag{5}$$

- (6) Finalmente, si definimos $\mathbf{A} \in \Re^{N\times q}$

$$\begin{aligned} \{\mathbf{A}\} \; &= \; a_{ij} = \pi_{0i}^{-1/2} \, \frac{\partial \pi_i(\theta)}{\partial \theta_{0j}} \\ &= \; \Delta(\boldsymbol{\pi}_0^{-1/2}) \, \frac{\partial \boldsymbol{\pi}}{\partial \theta_0} \end{aligned}$$

reemplazando a $\bar{\boldsymbol{\pi}}$ por $\boldsymbol{\pi}_0$ en (5) podemos reescribir (5) :

$$\mathbf{A}' \, \Delta(\boldsymbol{\pi}_0^{-1/2}) \sqrt{n} (\mathbf{p} - \boldsymbol{\pi}_0) \overset{(a)}{=} (\mathbf{A}'\mathbf{A}) \, \sqrt{n} (\bar{\theta} - \theta_0)$$

Como $(\mathbf{A}'\mathbf{A})$ es invertible por hipótesis, tenemos que

$$\begin{aligned} \sqrt{n}(\bar{\theta} - \theta_0) \; &\overset{(a)}{=} \; (\mathbf{A}'\mathbf{A})^{-1} \, \mathbf{A}' \;\; \Delta(\boldsymbol{\pi}_0^{-1/2}) \sqrt{n} (\mathbf{p} - \boldsymbol{\pi}_0) \\ &\overset{(a)}{=} \; \sqrt{n} \mathbf{D} Z_n \end{aligned}$$

De donde deduciremos que

$$\sqrt{n}(\bar{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N_q(0, (\mathbf{A}'\mathbf{A})^{-1})\,.$$

En realidad, necesitamos algo más, ya que nos interesa la distribución de $\boldsymbol{\pi}(\theta)$. En $\Re$ tenemos que si

$$\sqrt{n}(X_n - \mu) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

entonces bajo condiciones de suavidad de la función g, entonces

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{\mathcal{D}} N(0, \sigma^2(g'(\mu))^2)$$

El siguiente lema, conocido como Método Δ generaliza este resultado para una función de vector aleatorio.

Lema 2: **Método Δ una función de vector aleatorio.**

Supongamos que $T_n = (T_n^1, \ldots, T_n^N)$ es una sucesión de vectores aleatorios tal que

$$\sqrt{n}((T_n^1, \ldots, T_n^N) - (\theta_1, \ldots, \theta_N)) \xrightarrow{\mathcal{D}} N(0, \Sigma)$$

Sea g una función tal que $g : \Re^N \longrightarrow \Re$ diferenciable. Luego si

$$\phi = (\phi_i) = \frac{\partial g}{\partial t_i}|_{t=\theta}\ ,$$

entonces

$$\sqrt{n}(g(T_n^1, \ldots, T_n^N) - g(\theta_1, \ldots, \theta_N)) \xrightarrow{\mathcal{D}} N(0, \phi' \Sigma \phi)$$

Análogamente, si en lugar de un campo escalar tenemos un campo vectorial, es decir $g : \Re^N \longrightarrow \Re^q$, donde cada componente g_i es diferenciable como en el lema anterior, obtenemos

$$\sqrt{n}(g(T_n^1, \ldots, T_n^N) - g(\theta_1, \ldots, \theta_N)) \xrightarrow{\mathcal{D}} N_q(0, G\Sigma G')$$

donde

$$G_{ij} = \frac{\partial g_i}{\partial t_j}|_{t=\theta}$$

Aplicando todos los resultados anteriores obtenemos que:

$$\sqrt{n}(\boldsymbol{\pi}(\bar{\theta}) - \boldsymbol{\pi}(\theta_0)) \xrightarrow{D} N(0, \frac{\partial \boldsymbol{\pi}}{\partial \theta_0}(A'A)^{-1}\frac{\partial \boldsymbol{\pi}'}{\partial \theta_0})$$

Notemos que el estadístico χ^2 puede escribirse como

$$\chi^2 = \mathbf{e}'\mathbf{e}$$

donde

$$\mathbf{e}' = \left(\sqrt{n}\,\frac{p_1 - \pi_1(\bar{\theta})}{\sqrt{\pi_1(\bar{\theta})}}, \ldots, \sqrt{n}\,\frac{p_N - \pi_N(\bar{\theta})}{\sqrt{\pi_N(\bar{\theta})}}\right)$$

Para derivar la distribución asintótica de χ^2 necesitaremos la conjunta de $(\mathbf{p}, \boldsymbol{\pi}(\bar{\theta}))$ y deduciremos que

$$\mathbf{e} \xrightarrow{\mathcal{D}} N(0, I - \boldsymbol{\pi}(\theta_0)^{1/2}\boldsymbol{\pi}'(\theta_0)^{1/2} - A(A'A)^{-1}A')$$