Modelo Lineal Generalizado

1er. Cuatrimestre de 2010

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}

Cuadro 1: Tabla de $I \times J$

Sea n_{ij} el número de individuos que tienen $P(X = i, Y = j)$, de manera que

$$n = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$$

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 por la propiedad de invariancia de los estimadores de máxima verosimilitud, bajo independencia el estimador de máxima verosimilitud de π_{ij} sería:

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

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:

$$\begin{aligned}
\widehat{m}_{i+} &= \sum_{j=1}^{J} \frac{n_{i+}n_{+j}}{n} = n_{i+} \\
\widehat{m}_{+j} &= \sum_{i=1}^{I} \frac{n_{i+}n_{+j}}{n} = n_{+j}
\end{aligned}$$

Test de Bondad de Ajuste

Pearson (1900) presentó un test que sirve para evaluar si una distribución multinomial tiene ciertas probabilidades π_{ij0} propuestas.

Como antes, *vectorizaremos* la tabla notando $\{n_1, \ldots, n_N\}$ a las casillas y $\{\pi_1, \ldots, \pi_N\}$ las probabilidades correspondiente y $n = \sum_{i=1}^{N} n_i$.

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

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

entonces

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

Si además agregásemos la restricción $\sum_{i=1}^{N} n_i = n$, sería natural perder un grado de libertad y que la distribución asintótica del estadístico resultase χ^2_{N-1}.Por todo esto, la regla de decisión sería

Rechazamos H_0 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

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

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 algunos resultados teóricos.

Proposición 1:

Sean $\mathbf{X}_n = (X_{1n}, \ldots, X_{kn})'$ una sucesión de v.a. 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 $1 \leq i \leq k$ $\bar{U}_{in} = \frac{1}{n} \sum_{j=1}^{n} U_{ij}$, entonces

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

Según la Proposición 1 debemos estudiar la distribución de $\boldsymbol{\lambda}'\bar{\mathbf{U}}_n$.

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

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

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

Proposición 2:

Si $\mathbf{Z}_n \xrightarrow{\mathcal{D}} \mathbf{Z}$ y g es una función continua, entonces

$$g(\mathbf{Z}_n) \xrightarrow{\mathcal{D}} g(\mathbf{Z}) \,.$$

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,

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

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 $\mathbf{T}_n = (T_n^1, \ldots, T_n^N)$ 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

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

Ahora estudiaremos la distribución asintótica de $(\frac{n_1}{n}, \ldots, \frac{n_{N-1}}{n})$ cuando

$$(n_1, \ldots, n_N)' \sim M(n, \pi_1, \ldots, \pi_N) \qquad \sum_{i=1}^{N} \pi_i = 1.$$

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:

$$\mathbf{Y}_i = (0, \ldots, \underset{j}{\overset{\downarrow}{1}}, \ldots, 0)' \qquad 1 \leq i \leq n$$

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

$$\begin{aligned} E(\mathbf{Y}_i) &= \boldsymbol{\pi} \\ \Sigma_{\mathbf{Y}_i} &= \Delta(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' \end{aligned}$$

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

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

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, luego aplicando el T.C.L. multivariado a $\mathbf{T}\mathbf{p}$

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

donde $\Delta(\tilde{\boldsymbol{\pi}}) - \tilde{\boldsymbol{\pi}}\tilde{\boldsymbol{\pi}}'$ 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 $\mathbf{x}'\mathbf{D}^{-1}\mathbf{x}$ es una función continua de $\mathbf{x}$ por la proposición 2 tenemos que

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

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 $m_2 = 2005.75$.

Luego, si queremos testear la hipótesis nula

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

el estadístico χ^2 es:

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

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 frecuencia suponemos que $\boldsymbol{\pi}$ yace en un subconjunto de menor dimensión de $\mathcal{S}$.

Por jemplo, los elementos de $\boldsymbol{\pi}$ podrían estar 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.

Luego, 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:

	$B=1$	$B=2$	
$A=1$	X_{11}	X_{12}	α
$A=2$	X_{21}	X_{22}	$1-\alpha$
	β	$1-\beta$	

Cuadro 2:

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

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

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

Este es un modelo restringido que depende del parámetro

$$\boldsymbol{\theta} = (\alpha, \beta)',$$

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:

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

por lo tanto

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

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

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

En el caso general, es decir en las tablas de contingencia de $I \times J$ con muestreo multinomial puede ser de interés testear la hipótesis de independencia, es decir:

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

es decir, la hipótesis nula depende de ciertos parámetros.

Por esto si bien para testear esta hipótesis usaremos un test de tipo Pearson, antes será necesario estudiar la distribución asintótica de dicho estadístico bajo estas condiciones.

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}

Cuadro 3:

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, $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$ podría expresarse como:

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

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

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 5e)

Supongamos que $\mathbf{X} \sim M(n, \boldsymbol{\pi}(\boldsymbol{\theta}_0))$, donde $\boldsymbol{\pi}(\boldsymbol{\theta}_0) = (\pi_1(\boldsymbol{\theta}_0), \ldots, \pi_N(\boldsymbol{\theta}_0))$ y $\theta_0 \in \Re^q$.

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, $\widehat{\boldsymbol{\theta}}$ es solución de las ecuaciones de "score":

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

Demostraremos que bajo condiciones de regularidad del EMV $\widehat{\boldsymbol{\theta}}$ es asintóticamente normal, es decir

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N_q(0, (\mathbf{A}'\mathbf{A})^{-1})$$

donde

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

Este resultado lo deduciremos expresando a $\widehat{\boldsymbol{\theta}}$ en términos de las frecuencias relativas, es decir $\mathbf{p}$, 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 $\boldsymbol{\theta}$ exista, es necesaria una **condición de identificabilidad fuerte**:

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

$$\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\delta} \sum_{i=1}^{N} \pi_i(\boldsymbol{\theta}_0) \log \frac{\pi_i(\boldsymbol{\theta}_0)}{\pi_i(\boldsymbol{\theta})} \geq \epsilon$$

Esta condición implica que fuera de la bola $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta$ no hay ninguna sucesión de puntos θ_r tal que $\boldsymbol{\pi}(\theta_r) \to \boldsymbol{\pi}(\boldsymbol{\theta}_0)$ a medida que $r \to \infty$, es decir que no hay valores θ lejanos a $\boldsymbol{\theta}_0$ que den practicamente las mismas probabilidades que $\boldsymbol{\pi}(\boldsymbol{\theta}_0)$.

Es decir:

$$\forall\delta > 0, \text{ exite } \epsilon > 0 \text{ tal que si } \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > \delta \text{ entonces } \|\boldsymbol{\pi}(\boldsymbol{\theta}) - \boldsymbol{\pi}(\boldsymbol{\theta}_0)\| > \epsilon\,.$$

Bajo la condición fuerte de identificabilidad y continuidad de las funciones $\pi_i(\boldsymbol{\theta})$, se puede demostrar que el EMV de $\boldsymbol{\theta}$ existe y que converge a $\boldsymbol{\theta}_0$ con probabilidad 1.

Más aún, si las funciones $\pi_i(\boldsymbol{\theta})$ 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 $\boldsymbol{\pi}(\boldsymbol{\theta}) \neq \boldsymbol{\pi}(\boldsymbol{\beta})$ si $\boldsymbol{\theta} \neq \boldsymbol{\beta}$, las funciones $\pi_i(\boldsymbol{\theta})$ tienen derivadas parciales de primer orden continuas en $\boldsymbol{\theta}_0$ y la matriz de información $\mathbf{I}$

$$\mathbf{I}_{rs} = \sum_{i=1}^{N} \frac{1}{\pi_i} \frac{\partial \pi_i}{\partial \theta_r} \frac{\partial \pi_i}{\partial \theta_s}$$

evaluada en $\boldsymbol{\theta}_0$ no es singular, entonces existe una raíz consistente de (2) que puede no ser un EMV y que tiene distribución asintótica normal.

Deduciremos un resultado análogo al que ya vimos para el caso no paramétrico cuando $\widehat{\boldsymbol{\pi}} = \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})$ 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 $\pi_1(\boldsymbol{\theta}), \ldots, \pi_N(\boldsymbol{\theta})$ que involucran a q parámetros $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$. Además, supongamos que:

- a) θ_0 es un punto interior del espacio paramétrico.
- b) $\pi_i(\theta_0) > 0 \forall i$.
- c) Cada $\boldsymbol{\pi}(\boldsymbol{\theta})$ admite derivadas parciales continuas de 1^{er} orden en un entrono de θ_0.
- d) La matriz $\mathbf{M} = \{\frac{\partial \pi_r(\boldsymbol{\theta})}{\partial \theta_s}\}$ de $N \times q$ evaluada en $\boldsymbol{\theta}_0$ 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 1: Supongamos que $\boldsymbol{\theta}_0$, valor verdadero del parámetro, es un punto interior del espacio paramétrico, $\pi_i(\boldsymbol{\theta}_0) > 0 \quad \forall i$ y que además se cumplen las siguientes condiciones:

- a) $\pi_i(\boldsymbol{\theta}) \neq \pi_i(\boldsymbol{\beta})$ para algún i si $\theta \neq \beta$ (condición de identificabilidad).
- b) $\pi_i(\theta)$ admite derivadas parciales de 1^{er} orden continuas en $\boldsymbol{\theta}_0$.
- c) La matriz $\{I_{rs}\}$ es no singular en $\boldsymbol{\theta}_0$, donde

$$I_{rs} = \sum_{j=1}^{N} \frac{1}{\pi_j(\boldsymbol{\theta}_0)} \frac{\partial}{\partial \theta_r} \pi_j(\boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_s} \pi_j(\boldsymbol{\theta}_0)$$

Luego, existe una raíz consistente $\bar{\theta}$ de $\dfrac{\partial l}{\partial \theta_i} = 0 \quad i = 1, \ldots, q$, ecuación de verosimilitud y tomando $\widehat{\theta} = \bar{\theta}$ tenemos que su distribución asintótica es normal q–variada

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

donde

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

Esquema de demostración

- (1) Lemita: Si $\sum_{i=1} a_i \geq \sum_{i=1} b_i$, donde $a_i > 0$ y $b_i > 0$, entonces

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

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

$$\operatorname*{ínf}_{\|\theta-\theta_0\|=\delta} S(\theta) > \epsilon > 0$$

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

$$\sum_{i=1}^{N} \sqrt{n}\, \frac{p_i - \pi_i(\boldsymbol{\theta}_0)}{\pi_i(\bar{\boldsymbol{\theta}})} \, \frac{\partial \pi_i(\bar{\boldsymbol{\theta}})}{\partial \theta_r} \; = \; \sum_{i=1}^{N} \sum_{s=1}^{q} \sqrt{n}\, (\bar{\theta}_s - \theta_{0s}) \, \frac{\partial \pi_i(\bar{\boldsymbol{\theta}})}{\partial \theta_r} \frac{\partial \pi_i(\tilde{\boldsymbol{\theta}})}{\partial \theta_s} \frac{1}{\pi_i(\bar{\boldsymbol{\theta}})} \tag{3}$$

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

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

reemplazando a $\bar{\boldsymbol{\pi}} = \boldsymbol{\pi}(\bar{\boldsymbol{\theta}})$ por su límite $\boldsymbol{\pi}_0 = \boldsymbol{\pi}(\boldsymbol{\theta}_0)$ en (3) podemos reescribir dicha ecuación como:

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

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

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

Dado que

$$\sqrt{n}(\mathbf{p} - \boldsymbol{\pi}(\boldsymbol{\theta}_0)) \xrightarrow{\mathcal{D}} N_q(0, \Delta(\boldsymbol{\pi}_0) - \boldsymbol{\pi}_0\boldsymbol{\pi}_0')$$

deduciremos que

$$\sqrt{n}(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N_q(0, (\mathbf{A}'\mathbf{A})^{-1})\,.$$

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

Aplicando todos los resultados anteriores obtenemos que:

$$\sqrt{n}(\boldsymbol{\pi}(\bar{\boldsymbol{\theta}}) - \boldsymbol{\pi}(\boldsymbol{\theta}_0)) \xrightarrow{D} N(0, \frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}(\mathbf{A}'\mathbf{A})^{-1}\frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta}_0)'}{\partial \boldsymbol{\theta}})$$

Notemos que el estadístico χ^2 puede escribirse como

$$\chi^2 = n\|\mathbf{e}\|^2$$

donde

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

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

Recordemos que

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

Por lo tanto

$$\sqrt{n}\begin{pmatrix}\mathbf{p}-\boldsymbol{\pi}_0\\ \bar{\boldsymbol{\pi}}-\boldsymbol{\pi}_0\end{pmatrix} = \begin{pmatrix}\mathbf{I}\\ \mathbf{D}\end{pmatrix}\sqrt{n}(\mathbf{p}-\boldsymbol{\pi}_0)+o_p(1)$$

de donde

$$\sqrt{n}\begin{pmatrix}\mathbf{p}-\boldsymbol{\pi}_0\\ \bar{\boldsymbol{\pi}}-\boldsymbol{\pi}_0\end{pmatrix} \xrightarrow{D} N(0,\Sigma^*)$$

donde

$$\Sigma^* = \begin{pmatrix}\Delta(\boldsymbol{\pi}_0)-\boldsymbol{\pi}_0\boldsymbol{\pi}_0' & (\Delta(\boldsymbol{\pi}_0)-\boldsymbol{\pi}_0\boldsymbol{\pi}_0')\mathbf{D}'\\ \mathbf{D}(\Delta(\boldsymbol{\pi}_0)-\boldsymbol{\pi}_0\boldsymbol{\pi}_0) & \mathbf{D}(\Delta(\boldsymbol{\pi}_0)-\boldsymbol{\pi}_0\boldsymbol{\pi}_0')\mathbf{D}'\end{pmatrix}$$

Luego, deduciremos que

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

Para usar el método Δ probaremos que:

$$\frac{\partial e_i}{\partial p_i} = \bar{\pi}_i^{-1/2}$$

$$\begin{aligned}\frac{\partial e_i}{\partial \bar{\pi}_i} &= -\frac{1}{2}\,\frac{p_i+\bar{\pi}_i}{\bar{\pi}_i^{\,3/2}}\\ \frac{\partial e_i}{\partial p_j} &= \frac{\partial e_i}{\partial \bar{\pi}_j} = 0\end{aligned}$$

De manera que

$$\begin{aligned}\frac{\partial \mathbf{e}}{\partial \mathbf{p}} &= \Delta(\bar{\pi}_i^{\,-1/2})\\ \frac{\partial \mathbf{e}}{\partial \bar{\boldsymbol{\pi}}} &= -\frac{1}{2}\,(\Delta(\mathbf{p})+\Delta(\bar{\boldsymbol{\pi}}))\Delta(\bar{\boldsymbol{\pi}}^{-3/2})\end{aligned}$$

Teorema: Sea $\mathbf{Y}$ un vector con distribución $N(\nu, \Sigma)$. Una condición necesaria y suficiente para que $(\mathbf{Y}-\nu)'\mathbf{C}(\mathbf{Y}-\nu)$ tenga distribución χ^2 es que $\Sigma\mathbf{C}\Sigma\mathbf{C}\Sigma = \Sigma\mathbf{C}\Sigma$, donde los grados de libertad serán el rango de $\mathbf{C}\Sigma$ (si Σ es no singular la condición se simplifica a $\mathbf{C}\Sigma\mathbf{C} = \mathbf{C}$). (Rao, 1965, p. 150)

Como hemos visto, $\chi^2 = \sqrt{n}\mathbf{e}'\sqrt{n}\mathbf{e}$, luego aplicaremos el resultado de Rao con $\nu = 0$, $\mathbf{C} = I$, $\Sigma = I - \boldsymbol{\pi}(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{\pi}'(\boldsymbol{\theta}_0)^{1/2} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$), por lo que resultará que

$$\chi^2 = \sqrt{n}\mathbf{e}'\sqrt{n}\mathbf{e} \xrightarrow{\mathcal{D}} \chi^2_{N-1-q}$$

Volviendo al Test de Independencia:

En una tabla de $I \times J$ con muestreo multinomial, la hipótesis nula de independencia equivale a

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

Usando el estadístico de Pearson tendríamos

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \widehat{m}_{ij})^2}{\widehat{m}_{ij}}$$

Respecto a los grados de libertad, estos están determinados por la cantidad de casillas y de parámetros, que en este caso serán

$$I * J - 1 - (I - 1) - (J - 1) = (I - 1) * (J - 1)\,.$$

En el siguiente ejemplo se clasifica a los individuos según su **Identificación Partidaria** y el **Género**. En la siguiente tabla tenemos los valores observados y en rojo los valores predichos bajo el modelo de independencia

	C: Identificación partidaria			
S:Género	Demócrata	Independiente	Republicano	Total
Fem	279	73	225	577
	(261.4)	(70.7)	(244.9)	
Masc	165	47	191	403
	(182.6)	(49.3)	(171.1)	
Total	444	120	416	980

Cuadro 4: Datos de la General Social Survey, 1991.

El valor del estadístico test de χ^2 es 7.01, con $IJ-1-(I-1)-(J-1)=(I-1)(J-1)=(2-1)(3-1)=2$ grados de libertad. El p–valor correspondiente es 0.03, de manera que a los valores habituales se rechazaría la hipótesis de independencia, indicando que el género y la identificación partidaria estarían asociados.

Otro Ejemplo

Este es otro ejemplo en que las probablidades dependen de una cantidad menor de parámetros desconocidos, θ.

Una muestra de 156 terneros nacidos en Florida fueron clasificados de acuerdo a que hayan contraído neumonía dentro de los 60 días de haber nacido. Los terneros que contrayeron neumonía fueron a su vez clasificados según se hayan infectado o no a los 15 días de haberse curado. La Tabla muestra los datos recolectados:

	Segunda Infección	
	Si	No
Primera Infección		
Si	30	63
No	0	63

Cuadro 5:

Es claro que los terneros que no tuvieron una primera infección no pudieron reinfectarse, es por ello que ninguna observación puede verse en la casilla 21 y en consecuencia en la tabla $n_{21} = 0$. Esto es lo que se conoce como un **cero estructural**. El objetivo en este estudio era testear si la probabilidad de una primera infeción era igual que la proabilidad de una segunda infección, dado que el ternero había contraído una primera infección.

Es decir, la hipótesis a testear es

$$H_0 : \pi_{11} + \pi_{12} = \frac{\pi_{11}}{\pi_{11} + \pi_{12}}$$

o equivalentemente $\pi_{11} = (\pi_{11} + \pi_{12})^2$. De manera que si llamamos $\pi = \pi_{11} + \pi_{12}$ a la probabilidad de infección primaria, el modelo bajo H_0 corresponde a una **trinomial** como muestra la siguiente tabla:

En este caso el likelihood resulta

$$(\pi^2)^{n_{11}}(\pi(1-\pi))^{n_{12}}(1-\pi)^{n_{22}},$$

Primera Infección	**Segunda Infección** Si	No	Total
Si	π^2	$\pi(1-\pi)$	π
No	$-$	$1-\pi$	$1-\pi$

Cuadro 6:

el log–likelihood queda

$$n_{11}\log(\pi^2) + n_{12}\log(\pi - \pi^2) + n_{22}\log(1-\pi)\,,$$

Derivando e igualando a 0 resulta

$$\widehat{\pi} = \frac{2n_{11} + n_{12}}{2n_{11} + 2n_{12} + n_{22}}$$

En la siguiente tabla se muestran en rojo (2do. renglón) los valores esperados bajo H_0

	Segunda Infección	
Primera Infección	Si	No
Si	30	63
	(38.1)	(39)
No	0	63
	(–)	(78.9)

Cuadro 7:

El estadístico de Pearson da χ^2 =19.7 con un total de (3-1)-1=1 grados de libertad. Dado que el p–valor es 0.00001 hay una fuerte evidencia contra H_0. Si miramos la tabla encontramos que muchos más terneros contraen una primera infección y no la segunda de lo que el modelo bajo H_0 predice. Con esto los investigadores concluyeron que la primera infección tiene un efecto inmunizador.