Comenzaremos por probar el siguiente resultado auxiliar.

Lema: 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

$$i_{rs} = \sum_j \frac{1}{\pi_j} \frac{\partial \pi_j(\boldsymbol{\theta}_0)}{\partial \theta_r} \frac{\partial \pi_j(\boldsymbol{\theta}_0)}{\partial \theta_s}$$

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

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

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

$$Z_j = \sum_{i=1}^{N} \frac{p_i}{\pi_i(\theta_0)} \frac{\partial \pi_i(\boldsymbol{\theta}_0)}{\partial \theta_j}$$

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

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

Esquema de demostración

- (1) Lemita: Si $\sum_{i=1} a_i$ y $\sum_{i=1} b_i$ son series convergentes, donde $a_i > 0$ y $b_i > 0$, tal que $\sum_{i=1} a_i \geq \sum_{i=1} b_i$, 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 $\tilde{\theta}_s \in (\bar{\theta}_s, \theta_{0s})$

$$\sum_{i=1}^{N} \sqrt{n}\, \frac{p_i - \pi_i(\boldsymbol{\theta}_0)}{\pi_i(\bar{\boldsymbol{\theta}})}\, \frac{\partial \pi_i(\bar{\boldsymbol{\theta}})}{\partial \theta_r} \;=\; \sum_{i=1}^{N} \sum_{s=1}^{q} \sqrt{n}\, (\bar{\theta}_s - \theta_{0s})\, \frac{\partial \pi_i(\bar{\boldsymbol{\theta}})}{\partial \theta_r} \frac{\partial \pi_i(\tilde{\boldsymbol{\theta}})}{\partial \theta_s} \frac{1}{\pi_i(\bar{\boldsymbol{\theta}})} \tag{8}$$

- (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 (8) podemos reescribir dicha ecuación como:

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

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

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

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 $\boldsymbol{\pi}(\bar{\boldsymbol{\theta}})$.

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

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

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

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

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:

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

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

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

Volvamos al ejemplo de la **filiación partidaria** que vimos en la primera clase. En la siguiente tabla tenemos los valores observados y en rojo los valores predichos bajo el modelo de independencia

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

Cuadro 12: 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 sexo 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 en si se reinfectaron o no a los 15 días de haberse curado. La Tabla muestra los datos recolectados:

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

Cuadro 13:

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:

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

Cuadro 14:

En este caso el likelihood resulta

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

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 15:

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.