Estadístico G^2

Otra medida alternativa para la distancia entre $\widehat{\boldsymbol{\pi}}$ y $\mathbf{p}$ muy usada es la **deviance** G^2, que es un estadístico basado en el cociente de verosimilitud.

Si queremos testear

$$\begin{aligned} H_0 &: \text{Modelo restringido } \omega \\ H_1 &: \text{Modelo Saturado } \Omega\,, \end{aligned}$$

el cociente estaría dado por

$$\Lambda = \frac{\max_{\omega} L}{\max_{\Omega} L}$$

Si consideramos $G^2 = -2\log\Lambda$ queda definido el estadístico como

$$\begin{aligned} G^2 &= -2\log\Lambda = 2[l(\mathbf{p},\mathbf{X}) - l(\widehat{\boldsymbol{\pi}},\mathbf{X})] \\ &= 2\left[\sum_{i=1}^{N} X_i \log p_i - \sum_{i=1}^{N} X_i \log \widehat{\pi}_i\right] \end{aligned}$$

$$= 2 \sum_{i=1}^{N} X_i \log \frac{p_i}{\widehat{\pi}_i}$$
$$= 2 \sum_{i=1}^{N} X_i \log \frac{X_i}{n\widehat{\pi}_i}$$

Por lo tanto

$$G^2 = 2 \sum_{i=1}^{N} X_i \log \frac{X_i}{n\widehat{\pi}_i} \,.$$

Probaremos que bajo H_0 la distribución límite de G^2 es también χ^2 con $N - 1 - \#\{$ parámetros bajo $\omega\}$ es decir la misma distribución límite que la del estadístico de Pearson.

Para derivar la distribución asintótica, probaremos que $G^2 - \chi^2 \xrightarrow{p} 0$.

Una ventaja de G^2 es que tiene sentido en modelos más generales.

En el ejemplo de **Identificación Partidaria vs. Sexo**, $G^2 = 7$, que da también un p–valor de 0.03.

Efecto de observar ceros

Si en alguna celda se observa un 0, el estadístico χ^2 puede calcularse sin problemas, siempre que las $\widehat{\pi}$'s sean todas positivas. Sin embargo, el estadístico G^2 tiene problemas, pues si $X_i = 0$, entonces $X_i \log \frac{X_i}{n\widehat{\pi}_i}$ no está definido. Si reescribimos a G^2 como

$$\begin{aligned} G^2 &= -2\log\Lambda = 2\log\frac{L(\mathbf{p},\mathbf{X})}{L(\widehat{\boldsymbol{\pi}},\mathbf{X})} \\ &= 2\log\prod_{i=1}^{N}\left(\frac{X_i/n}{\widehat{\pi}_i}\right)^{X_i} \end{aligned}$$

una celda con un 0 aporta un 1 al producto y por lo tanto puede ser ignorada. Luego podemos calcular a G^2 como

$$2\sum_{i:\,x_i>0} X_i \log\frac{X_i}{n\widehat{\pi}_i}$$

Si alguna $\widehat{\pi}_i$ es 0, los dos estadísticos se *rompen*.

¿Cuán grande debe ser n para tener una buena aproximación?

Sabemos que a medida que n se hace más grande la distribución de χ^2 y de G^2 se aproximan a una distribución límite χ^2, sin embargo nos preguntamos cuán grande es grande.

- Una vieja regla conocida para las binomiales dice que la aproximación χ^2 es buena si $n\widehat{\pi}_i \geq 5$, $i = 1, \ldots, N$.

- Otra regla más permisiva establece que la aproximación χ^2 es buena si a lo sumo el 20 % de las casillas tienen $n\widehat{\pi}_i < 5$, $i = 1, \ldots, N$ y ninguna casilla tiene $n\widehat{\pi}_i < 1$.

- En tablas *sparse* (esparcidas????)(por ejemplo, $n/N < 5$) la aproximación χ^2 es pobre. En realidad, si los datos están distribuidos en la tabla de foma muy desigual, en el sentido de que hay zonas de la tabla que son *sparse*, la aproximación χ^2 también puede ser pobre, aún cuando el n total sea grande.

Hemos probado que los dos estadísticos se aproximan a 0, si el modelo es cierto. Si el modelo no es cierto, ambos crecen, pero la diferencia entre ambos también puede crecer. De manera, que si el modelo tiene un ajuste pobre los dos estadísticos pueden ser grandes y estar lejos uno de otro, i.e., $|\chi^2 - G^2|$ no necesariamente tiende a 0 con n. Aún en esa situación, los correspondientes p–valores pueden estar cerca de 0 y podemos llegar a la misma conclusión a partir de ellos.

Para ser más precisos, consideremos una sucesión de situaciones $\boldsymbol{\pi}_n$ para las cuales la falta de ajuste disminuye con n, es decir trabajaremos con alternativas contiguas. Supongamos que el modelo bajo la hipótesis nula es $\boldsymbol{\pi}$, pero en realidad

$$\boldsymbol{\pi}_n = \mathbf{f}(\theta) + \boldsymbol{\delta}/\sqrt{n}\,,$$

entonces si $\boldsymbol{\delta} = 0$, el modelo es cierto.

Para estas alternativas contiguas, Mitra (1958) demostró que el estadístico de Pearson tiene distribución asintótica χ^2 no central, con $N-1-q$ grados de libertad, con parámetro de no centralidad dado por

$$\lambda = n \sum_{i=1}^{N} \frac{(\pi_{ni} - f_i(\theta))^2}{f_i(\theta)}$$

Notemos que λ tiene la forma del estadístico χ^2 en el que se reemplazó a $\mathbf{p}$ por $\boldsymbol{\pi}_n$ y a $\widehat{\boldsymbol{\pi}}$ por $\mathbf{f}(\boldsymbol{\theta})$. Análogamente, utilizando los mismos reemplazos obtenemos el parámetro de no centralidad de G^2. Haberman (1974) demostró que bajo ciertas condiciones χ^2 y G^2 tienen el mismo parámetro de no centralidad, pero éste no es siempre el caso, (ver Agresti,2002, pag 590).

Residuos de Pearson y deviance

Como ya hemos visto podemos escribir al estadístico de Pearson como

$$\chi^2 = n \sum_{i=1}^{N} e_i^2 \,.$$

A

$$\epsilon_i = \sqrt{n}\,\frac{p_i - \widehat{\pi}_i}{\sqrt{\widehat{\pi}_i}} = \frac{n_i - \widehat{m}_i}{\sqrt{\widehat{m}_i}}$$

se lo conoce como el i–ésimo residuo de Pearson.

Estos residuos se comportan de alguna manera como los residuos estandarizados que conocimos en regresión lineal. Es común que se compare a $|\epsilon_i|$ con 2, indicándose falta de ajuste en la i–ésima casilla si $|\epsilon_i| > 2$. El análisis de estos residuos puede sugerirnos en que sentido los datos se apartan del modelo ajustado.

De la misma forma, la deviance puede interpretarse como una suma de cuadrados de residuos

$$G^2 = \sum_{i=1}^{N} r_i^2$$

donde

$$r_i = \sqrt{\left| 2X_i \log \frac{X_i}{n\widehat{\pi}_i} \right|} \times sgn(X_i - n\widehat{\pi}_i)$$

que se conocen como componentes de la deviance.

Medidas de Asociación

A fin de describir el grado de asociación entre las variables de una tabla de contingencia es frecuente que se usen distintas medidas.

Comenzaremos con tablas de 2×2, como las que siguen

	Y		
X	1	2	Total
1	π_{11}	π_{12}	π_{1+}
2	π_{21}	π_{22}	π_{2+}
Total	π_{+1}	π_{+2}	1

	Y		
X	1	2	Total
1	n_{11}	n_{12}	n_{1+}
2	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	1

Consideremos la siguiente tabla que corresponde a un informe sobre la relación entre el uso de aspirina y el infarto de miocardio realizado por el Physicians Health Study Research Group de Harvard Medical School:

	Infarto de Miocardio		
	si	no	Total
Aspirina	104	10933	11037
Placebo	189	10845	11034

Diferencia de Proporciones o Riesgo Atribuible

Miremos a Y como variable de respuesta y a X como variable explicativa, tal como seríia natural en un muestreo de producto multinomial en que

$$n_{11} \sim Bi(n_{1+}, \frac{\pi_{11}}{\pi_{1+}}) \text{ y } n_{21} \sim Bi(n_{2+}, \frac{\pi_{21}}{\pi_{2+}})$$

independientes.

La diferencia de proporciones se define como

$$\begin{aligned} \delta &= P(Y=1|X=1) - P(Y=1|X=2) \\ &= \frac{\pi_{11}}{\pi_{1+}} - \frac{\pi_{21}}{\pi_{2+}} \\ &= \pi_{1|1} - \pi_{1|2} \end{aligned}$$

Podemos estimar a δ como

$$\begin{aligned} d &= \frac{n_{11}}{n_{1+}} - \frac{n_{21}}{n_{2+}} \\ &= p_{1|1} - p_{1|2} \end{aligned}$$

En el ejemplo de Infarto de Miocardio tenemos

$$d = 104/11037 - 189/11034 = \; 0.0094 \text{ - } 0.0171 = \text{-}0.0077$$

Observemos que

$$\begin{aligned} E(d) &= E(p_{1|1} - p_{1|2}) = \pi_{1|1} - \pi_{1|2} \\ V(d) &= V(p_{1|1} - p_{1|2}) = \frac{\pi_{1|1}(1-\pi_{1|1})}{n_{1+}} + \frac{\pi_{1|2}(1-\pi_{1|2})}{n_{2+}} \end{aligned}$$

siendo la última igualdad cierta por la independencia entre las filas.

Si n_{1+} y n_{2+} son grandes, d es aproximadamente normal, es decir

$$\frac{(p_{1|1} - p_{1|2}) - (\pi_{1|1} - \pi_{1|2})}{\sqrt{\frac{\pi_{1|1}(1-\pi_{1|1})}{n_{1+}} + \frac{\pi_{1|2}(1-\pi_{1|2})}{n_{2+}}}}$$

es aproximadamente $N(0,1)$. Por lo tanto haciendo un plug–in para estimar la varianza podemos obtener un intervalo asintótitco para δ de nivel $1-\alpha$ como

$$d \;\pm\; z_{\alpha/2}\sqrt{\frac{p_{1|1}(1-p_{1|1})}{n_{1+}}+\frac{p_{1|2}(1-p_{1|2})}{n_{2+}}}$$

$$p_{1|1}-p_{1|2} \;\pm\; z_{\alpha/2}\sqrt{\frac{p_{1|1}(1-p_{1|1})}{n_{1+}}+\frac{p_{1|2}(1-p_{1|2})}{n_{2+}}}$$

Riesgo Relativo

Notemos que que la diferencia entre 41 % y 40.1 % es la misma que entre 1 % y 0.1 %. Sin embargo, 1 % es diez veces 0.1 %. Este es un problema de la diferencia de proporciones como medida. Si estamos trabajanos con eventos poco frecuentes $\pi_{1|1}$ y $\pi_{1|2}$ serán muy pequeñas y δ será cercano a 0, aún cuando el efecto sea importante, como en el ejemplo anterior.

Esto es frecuente en epidemiología en donde la prevalencia de ciertas enfermedades es muy baja.

Esto sugiere la conveniencia de considerar una medida relativa como el **riesgo relativo**

$$RR = \frac{P(Y = 1|X = 1)}{P(Y = 1|X = 2)} = \frac{\pi_{11}/\pi_{1+}}{\pi_{21}/\pi_{2+}}$$

El riesgo relativo es una medida no negativa y un riego relativo igual a 1 corresponde a independencia.

El EMV de RR es

$$rr = \frac{n_{11}/n_{1+}}{n_{21}/n_{2+}}$$

En el ejemplo quedaría:

$$rr = \frac{0.0094}{0.0171} = 0.55\,,$$

esto significa que el riesgo de infarto de miocardio en el grupo tratado con aspirina es aproximadamente la mitad que en grupo que recibió placebo.

Dado que podemos aproximar mediante una normal a su logaritmo suele usarse como medida $\log(RR)$, que se estima por $\log(rr) = \log p_{1|1} - \log p_{1|2}$.

Sabemos que

$$\sqrt{n_{i+}}(p_{1|i} - \pi_{1|i}) \xrightarrow{D} N(0, \pi_{1|i}(1 - \pi_{1|i}))\,,$$

luego usando el método Δ obtenemos que

$$\sqrt{n_{i+}}(\log p_{1|i} - \log \pi_{1|i}) \xrightarrow{D} N\left(0, \frac{(1-\pi_{1|i})}{\pi_{1|i}}\right)\,.$$

Por la independencia entre las filas, obtenemos que la varianza asintótica de $\log(rr)$ es

$$V(\log(rr)) \simeq \frac{(1-\pi_{1|1})}{n_{1+}\pi_{1|1}} + \frac{(1-\pi_{1|2})}{n_{2+}\pi_{1|2}}$$

y se puede estimar haciendo un plug–in por

$$\begin{aligned}\widehat{V}(\log(rr)) &\simeq \frac{(1-p_{1|1})}{n_{1+}\,p_{1|1}} + \frac{(1-p_{1|2})}{n_{2+}\,p_{1|2}} \\ &\simeq \frac{1}{n_{11}} - \frac{1}{n_{1+}} + \frac{1}{n_{21}} - \frac{1}{n_{2+}}\end{aligned}$$

Un intervalo de nivel asintótico $1-\alpha$ para $\log(RR)$ es

$$\log(rr) \pm z_{\alpha/2}\sqrt{\widehat{V}(\log(rr))}$$

Como $\log(rr)$ no existe si algún $p_{1|i}=0$ suele usarse

$$\log(\tilde{r}r) = \log\left(\frac{n_{11}+1/2}{n_{1+}+1/2}\right) - \log\left(\frac{n_{21}+1/2}{n_{2+}+1/2}\right)$$

Odds Ratio (Producto Cruzado)

El riesgo relativo es el cociente de dos probabilidades. Podríamos comparar la probabilidad de **si** y de **no** en un mismo estrato. Eso nos lleva a la definición de **odss** o **chance**. El odds de un suceso es

$$odds = \frac{probabilidad}{1 - probabilidad}$$

y toma cualquier valor mayor o igual a 0.

En el ejemplo, tenemos que para el grupo tratado el odds estimado resulta

$$0.0094/(1-0.0094) = 0.0094/0.9906 = 0.0095\,,$$

mientras que para el grupo placebo el odds estimado es

$$0.0171/(1-0.0171) = 0.0171/0.9829 = 0.0174\,.$$

En el grupo que recibió placebo la chance de tener infarto es 0.0174 la de no tener infarto, mientras que en el grupo tratado la chance de infarto es 0.0095 la de no tener infarto.

Dicho de otra manera, la chance de tener infarto respecto de la de no tenerlo en el grupo placebo es aproximadamente el doble que la obtenida en en el grupo tratado.

el grupo tratado, la chance de no

Podríamos comparar los dos odds, por ejemplo considerando su cociente, esto da origen a

$$\begin{aligned}\theta &= \text{odds ratio} = \frac{P(Y=1|X=1)/P(Y=2|X=1)}{P(Y=1|X=2)/P(Y=2|X=2)} \\ &= \frac{\left[\frac{\pi_{11}}{\pi_{1+}}\right] / \left[\frac{\pi_{12}}{\pi_{1+}}\right]}{\left[\frac{\pi_{21}}{\pi_{2+}}\right] / \left[\frac{\pi_{22}}{\pi_{2+}}\right]} \\ &= \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}\end{aligned}$$

Esta medida función de $P(Y|X)$, la inferencia es válida para los tres muestreos vistos.

El EMV es

$$\widehat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

Las propiedades de $\widehat{\theta}$ son fáciles de deducir bajo muestreo multinomial, pero también son válidas con muestreo Poisson o Producto Multinomial en el que los totales marginales por filas o bien por columnas están fijos.

Como con el riesgo relativo podemos deducir un intervalo de nivel asintótico $1-\alpha$ para $\log(\widehat{\theta})$

$$\log(\widehat{\theta}) \pm z_{\alpha/2}\sqrt{\widehat{V}(\log \widehat{\theta})}$$

donde

$$\widehat{V}(\log(\widehat{\theta})) = \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}$$

Notemos además que si intercambiamos los roles de X e Y, obtenemos

$$\theta = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$$

por lo que también puede ser visto como función de $P(X|Y)$, que correspondería a tener n_{+j} fijos. El hecho de que los roles de X e Y puedan ser intercambiados es una propiedad interesante, pues puede ser de gran utilidad pues permite usar estudios restropectivos.