

# Modelo Lineal Generalizado

1er. Cuatrimestre de 2005

El objetivo de esta materia será estudiar la relación entre dos o más variables, que podrán ser tanto continuas como categóricas.

En algunos casos diferenciaremos entre variable de respuesta y variables explicativas, mientras que en otros, simplemente, nos interesará estudiar la asociación entre las variables presentes sin hacer esta distinción.

A diferencia de lo que tratamos en Modelo Lineal, la variable de respuesta podrá ser categórica.

Abordaremos tres grandes temas:

- Tablas de Contingencia

- Tablas de Contingencia
- Modelo Lineal Generalizado

- Tablas de Contingencia
- Modelo Lineal Generalizado
- Modelos Log-lineales

- Tablas de Contingencia
- Modelo Lineal Generalizado
- Modelos Log-lineales

Veremos algunos ejemplos que introduzcan estos temas.

Consideremos el caso en el una muestra de 980 norteamericanos fue clasificada de acuerdo con el sexo y su identificación político-partidaria. En esta situación nos interesa estudiar si hay asociación o no entre las variables categóricas **Sexo** y **Identificación partidaria**.

<b>S:</b> Sexo	<b>C:</b> Identificación partidaria			Total
	Demócrata	Independiente	Republicano	
Mujer	279	73	225	577
Hombre	165	47	191	403
Total	444	120	416	980

Cuadro 1: General Social Survey, 1991

Esta es una **tablas de contingencia** bastante sencilla.

Para responder este tipo de preguntas, en primera instancia, veremos los test de independencia o de homogeneidad basados en la distribución  $\chi^2$  que fueron introducidos por Pearson.

Sin embargo, estos tests, como muchos otros, tienen algunas limitaciones. Una de ellas es que si bien nos indican cuanta evidencia hay de asociación entre las variables de interés, no nos dicen nada sobre la naturaleza de esta relación.

Para comprender más profundamente la asociación entre variables nos ayudarán los **modelos log-lineales** y los **modelos lineales generalizados**, siendo estos últimos una generalización que incluye modelo lineal habitual.



Edad	Educación	Más Hijos?	Uso de Anticonceptivos		Total
			No	Si	
< 25	Baja	Si	53	6	59
		No	10	4	14
	Alta	Si	212	52	264
		No	50	10	60
25-29	Baja	Si	60	14	74
		No	19	10	29
	Alta	Si	155	54	209
		No	65	27	92
30-39	Baja	Si	112	33	145
		No	77	80	157
	Alta	Si	118	46	164
		No	68	78	146
40-49	Baja	Si	35	6	41
		No	46	48	94
	Alta	Si	8	8	16
		No	12	31	43
Total			1100	507	1607

Cuadro 2: Encuesta de Fertilidad Fiji, 1975

## Ejemplo: Datos de Anticoncepción

Los siguientes datos tomados de Little (1978) corresponden a 1607 mujeres casadas y fértiles entrevistadas por la Encuesta de Fertilidad Fiji de 1975, clasificadas por edad, nivel de educación, deseo de tener más hijos y el uso de anticonceptivos.

El objetivo es describir cómo el uso de métodos anticonceptivos varía según la *edad*, el *nivel de educación* y el *deseo de tener más hijos*.

En este ejemplo consideramos a *Anticoncepción* como variable de respuesta y a las demás como predictoras. En este caso, todas las predictoras son tratadas como variables categóricas, sin embargo el modelo que presentaremos permite introducir variables independientes continuas y discretas.

Por ejemplo, una pregunta que sería interesante responder es si la asociación entre educación y anticoncepción es afectada por el hecho de que mujeres con un nivel de educación más elevado prefieren familias más chicas que las mujeres con niveles de educación inferior.

En esta situación podríamos intentar modelar la variable de respuesta en función de las explicativas.

Recordemos que en el modelo lineal habitual, si  $Y$  es nuestra variable de respuesta y  $x$  la explicativa podemos formular el modelo como:

$$E(Y) = \beta_0 + \beta_1 x \quad (1)$$

Si, como en el ejemplo, nuestra variable de respuesta es binomial, entonces  $E(Y) = \pi$ , por lo tanto la generalización inmediata de (1) sería:

$$E(Y) = \pi = \pi(x) = \beta_0 + \beta_1 x \quad (2)$$

Sin embargo, (2) no parece ser adecuado, pues  $\beta_0 + \beta_1 x$  podría tomar valores fuera del intervalo  $(0, 1)$ .

Un problema evidente de este modelo es que las probabilidades  $\Pi_i$  son acotadas, mientras que las  $\mathbf{x}'_i \boldsymbol{\beta}$  pueden tomar cualquier valor real. Si bien esto podría controlarse imponiendo complicadas restricciones a los coeficientes, esta solución no resulta muy natural.

Una solución sencilla es *transformar* la probabilidad mediante una función que mapee el intervalo  $(0, 1)$  sobre la recta real y luego modelar esta transformación como una función lineal de las variables independientes.

Una elección muy frecuente es:

$$\text{logit}(\Pi) = \log \left[ \frac{\Pi}{1 - \Pi} \right] = \beta_0 + \beta_1 x_1$$

que da origen al modelo de regresión logística. Esta es sólo una elección posible y veremos más adelante porque es una elección razonable.

Un punto a destacar es que en este modelo es natural la heteroscedasticidad, pues  $V(Y) = \pi(1 - \pi)$ , que será función de  $x$ .

El modelo definido es un caso del **modelo lineal generalizado** con respuesta binomial y función de enlace logit.

Si bien los coeficientes  $\beta$  tienen una interpretación similar a la que tienen en el modelo lineal, debemos tener en cuenta que el miembro de la derecha es un logit y no una media, por lo que deberemos precisar cuál es su significado en este caso.

Estos temas los desarrollaremos en el contexto más general del **modelo lineal generalizado**.

Este modelo es una extensión que comprende al modelo lineal que aplicamos cuando el supuesto de normalidad es razonable y que abarca también el caso de una respuesta Poisson, Binomial Negativa, Gamma, Exponencial, etc.

Una vez establecido el modelo que queremos ajustar deberemos estimar los parámetros, hallar intervalos de confianza para los mismos, evaluar la bondad del ajuste y es probable que nos interese realizar algún test que involucre a los parámetros. También tendremos que evaluar la influencia de las observaciones en la determinación de los valores estimados.

En nuestro último ejemplo consideramos de nuevo el caso de una tabla de contingencia. Supongamos  $F$  es la variable que identificamos en las filas y  $C$  la de las columnas y que nos interesa estudiar la asociación de las variables categóricas.

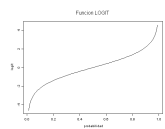
	$C$					
$F$	1	2	...	$j$	...	$J$
1	.	.		...		.
2	.	.		...		.
.	.	.		...		.
$i$	.	.		$\pi_{ij}$		.
.	.	.		...		.
$I$	.	.		...		.

Cuadro 3: Tabla Genérica

Sabemos que si  $F$  y  $C$  son independientes, las  $\pi_{ij}$  se pueden escribir en términos de las marginales como

$$\pi_{ij} = \pi_i^F \pi_j^C .$$

Qué ocurre en el caso general cuando no suponemos independencia?



Si pensamos en los valores esperados,  $m_{ij} = n\pi_{ij}$  también podremos expresar a  $m_{ij}$  usando un modelo multiplicativo:

$$m_{ij} = \tau \tau_i^F \tau_j^C \tau_{ij}^{FC}, \quad (3)$$

donde, como en ANOVA, los  $\tau$  deberán satisfacer ciertas restricciones.

Si tomamos logaritmo en (3) queda:

$$\begin{aligned} \log m_{ij} &= \log \tau + \log \tau_i^F + \log \tau_j^C + \log \tau_{ij}^{FC} \\ \log m_{ij} &= \mu + \mu_i^F + \mu_j^C + \mu_{ij}^{FC} \end{aligned}$$

que resulta un modelo aditivo, al que estamos más acostumbrados.

Este tipo de modelos recibe el nombre de **log-lineal**. Una diferencia con el modelo lineal habitual es que aquí las dos variables tienen un rol simétrico. El investigador deducirá una asociación entre las variables interpretando los parámetros. Esta tarea puede ser más o menos compleja si la cantidad de parámetros es muy elevada, como ocurre cuando aumenta el número de variables en el problema.



## Bibliografía:

- Agresti, A. (1990). *Categorical Data Analysis*. Wiley, New York.
- Christensen, R. (1997). *Log-linear Models and Logistic Regression*. 2da. Edición. New York: Springer Verlag.
- Cordeiro, G. M. (1986). *Modelos Lineares Generalizados*. VII SINAPE, Campinas S.P.
- Dobson, A. (2001). *An Introduction to Generalized Linear Models*. 2da. Edición. Londres: Chapman and Hall.
- Lindsey, J. (1997). *Applying Generalized Linear Models*. New York: Springer Verlag .
- Mc. Cullagh y Nelder, J. A. (1989). *Generalized Linear Models*. 2da. Edición. Londres: Chapman and Hall.
- Santner, T. y Duffy, D. (1989). *The Statistical Analysis of Discrete Data*. New York: Springer Verlag.
- Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*. New York: Wiley.

## Tablas de Contingencia

En la primera parte del curso nos ocuparemos de estudiar la relación entre 2 ó 3 variables categóricas. Introduciremos parámetros que describen la asociación entre variables categóricas y luego haremos inferencia sobre estos parámetro

Sean  $X$  e  $Y$  dos variables categóricas de respuesta, de manera que  $X$  tiene  $I$  niveles e  $Y$  tiene  $J$  niveles posibles. Cuando clasificamos sujetos de acuerdo a las dos variables tenemos  $IJ$  combinaciones posibles. Las respuestas  $(X, Y)$  de un individuo elegido a la azar tiene una distribución de probabilidad y representamos esta distribución en un rectángulo que tiene  $I$  filas para  $X$  y  $J$  columnas para  $Y$ .

Las casillas de la tabla representan los  $IJ$  resultados posibles. La probabilidad de que  $(X, Y)$  caiga en la casilla que corresponde a la fila  $i$  y columna  $j$  será  $\pi_{ij}$ . Cuando las celdas contienen la frecuencia de cada resultado  $ij$  tenemos una **tabla de contingencia**, término que introdujo Pearson en 1904. También suele llamársela **tabla de clasificación cruzada**. Una tabla de contingencia

$X$	$Y$					
	1	2	...	$j$	...	$J$
1	.	.		...		.
2	.	.		...		.
.	.	.		...		.
.	.	.		...		.
$i$	.	.		...		.
.	.	.		...		.
.	.	.		...		.
$I$	.	.		...		.

Cuadro 4: Tabla Genérica

con  $I$  filas y  $J$  columnas se dice una tabla de  $I \times J$ .

La distribución de probabilidad  $\pi_{ij}$  es la distribución conjunta de  $X$  e  $Y$ , mientras que las marginales de ambas variables las obtendremos sumando filas

y columnas respectivamente:  $\pi_{i+}$  y  $\pi_{+j}$ , donde

$$\pi_{i+} = \sum_{j=1}^J \pi_{ij} \quad \pi_{+j} = \sum_{i=1}^I \pi_{ij}$$

En muchos casos una de las variables, digamos  $Y$  es una variable de respuesta y la otra,  $X$ , es una variable explicativa. En general, es de interés estudiar cómo cambia la distribución de  $Y$  cuando pasamos de un nivel de  $X$  a otro. Dado que un sujeto está clasificado en la fila  $i$  de  $X$ ,  $\pi_{j|i}$  es la probabilidad de que clasifique en la columna  $j$  de  $Y$ , es decir  $\{\pi_{1|i}, \dots, \pi_{J|i}\}$  es la **probabilidad condicional** de  $Y$  dado que  $X = i$ . En términos de las probabilidades definidas, tenemos que

$$\pi_{j|i} = \frac{\pi_{ij}}{\pi_{i+}} \quad \forall i, j.$$

Diremos que  $X$  e  $Y$  son **independientes** si

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

y cuando vale la independencia

$$\pi_{j|i} = \frac{\pi_{ij}}{\pi_{i+}} = \frac{\pi_{i+}\pi_{+j}}{\pi_{i+}} = \pi_{+j} \quad \forall i, j.$$

En una tabla de  $2 \times 2$  tendríamos:

Filas	Columnas		
	1	2	Total
1	$\pi_{11}$ $(\pi_{1 1})$	$\pi_{12}$ $(\pi_{2 1})$	$\pi_{1+}$ $(1)$
2	$\pi_{21}$ $(\pi_{1 2})$	$\pi_{22}$ $(\pi_{2 2})$	$\pi_{2+}$ $(1)$
Total	$\pi_{+1}$	$\pi_{+2}$	1

Cuadro 5: Tabla de  $2 \times 2$

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

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

En el **caso muestral** la información también suele disponerse sobre una tabla como la que sigue:

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

Cuadro 6: Tabla Genérica

## Ejemplo

Comencemos por considerar un ejemplo de los más sencillos de tablas de contingencia en el que tenemos dos variables.

	<b>C: Cree en la vida después de la muerte</b>		Total
	Si	No	
<b>S:Sexo</b>			
Mujer	435	147	582
Hombre	375	134	509
Total	810	281	1091

Cuadro 7: General Social Survey, 1991

En general tendremos:

**Cuando hay sólo dos variables la hipótesis que suele ser de interés es si las variables son independientes.**

Podemos escribir esta hipótesis como

$$H_o : \Pi_{ij} = \Pi_{i+} \Pi_{+j} \quad \forall i \forall j$$

Para resolver este problema necesitamos estimar las probabilidades: en el

<b>S</b>	<b>C</b>		Total
	Si	No	
Mujer	$n_{11}$	$n_{12}$	$n_{1+}$
Hombre	$n_{21}$	$n_{22}$	$n_{2+}$
Total	$n_{+1}$	$n_{+2}$	$n_{++} = n$

Cuadro 8: Tabla de  $2 \times 2$

enfoque tradicional estimamos las probabilidades bajo el supuesto de independencia y las comparamos los valores observados con los valores esperados mediante un estadístico  $\chi^2$  de Pearson cuya distribución asintótica necesitaremos estudiar.

Podríamos estimar por máxima verosimilitud las probabilidades y realizar un test de cociente de verosimilitud, para esto necesitamos asumir una distribución subyacente.

Si cada una de las  $n$  observaciones es clasificada en forma independiente en una de las  $I \times J$  celdas de la tabla con probabilidad  $\Pi_{ij}$ , entonces el vector aleatorio que representa el número de individuos clasificados en la celda  $(i, j)$ ,  $\mathbf{n}$ , tiene distribución multinomial. La frecuencias esperadas en cada casilla son



$\mu_{ij} = n\pi_{ij}$ . Para  $I = J = 2$  sería

$$P(\mathbf{n} = (n_{11}^*, \dots, n_{22}^*)) = \frac{n!}{n_{11}^*! n_{12}^*! n_{21}^*! n_{22}^*!} \Pi_{11}^{n_{11}^*} \Pi_{12}^{n_{12}^*} \Pi_{21}^{n_{21}^*} \Pi_{22}^{n_{22}^*}.$$

Salvo constantes, el log-likelihood en el caso general queda:

$$\ell = \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log \Pi_{ij}$$

La maximización de  $\ell$  debe contemplar que  $\sum_{i=1}^I \sum_{j=1}^J \Pi_{ij} = 1$ .

Encontraremos que el EMV es

$$\hat{\Pi}_{ij} = \frac{n_{ij}}{n}$$

Bajo el supuesto de independencia tendremos que  $\Pi_{ij} = \Pi_{i+} \Pi_{+j}$  si calculamos el EMV bajo esta hipótesis obtenemos

$$\hat{\Pi}_{i+} = \frac{n_{i+}}{n} \quad \text{y} \quad \hat{\Pi}_{+j} = \frac{n_{+j}}{n}$$

Combinando estos estimadores las probabilidades estimadas resultan

$$\widehat{\Pi}_{ij} = \widehat{\Pi}_{i+} \widehat{\Pi}_{+j}$$

y las frecuencias esperadas estimadas

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

que es el resultado conocido.

El ajuste de las frecuencias estimadas puede realizarse a través del estadístico de Pearson o a la deviance que como veremos tienen distribución asintóticamente

con

$(I - 1)(J - 1)$  grados de libertad.

## Tipos de Muestreo

Dada una tabla de contingencia hay varios esquemas de muestreo que pueden conducir a los datos tal como los hemos observado y que influirán en el modelo de probabilidad que usaremos. En este caso tenemos dos factores  $S$  y  $C$  cada uno con dos niveles. En general, tendremos un factor *fila* con  $I$  niveles y un factor *columna* con  $J$  niveles que corresponde a una tabla de  $I \times J$ .

El número total de celdas es  $N = I \times J$ . Los totales marginales muestrales son

$$\begin{aligned}n_{i+} &= \sum_{j=1}^J n_{ij} && \text{total fila} \\n_{+j} &= \sum_{i=1}^I n_{ij} && \text{total columna} \\n_{++} &= n = \sum_{i=1}^I \sum_{j=1}^J n_{ij} && \text{gran total}\end{aligned}$$

Usaremos una notación similar a la anterior, por ejemplo,  $p_{ij}$  será la distribu-

ción conjunta muestral de  $(X, Y)$  definida por

$$p_{ij} = \frac{n_{ij}}{n}.$$

Las condicionales quedarán definidas análogamente, por ejemplo,

$$p_{j|i} = \frac{p_{ij}}{p_{i+}} = \frac{n_{ij}}{n_{i+}} \quad \forall i, j.$$

En el ejemplo anterior asumimos que todos los datos han sido recolectados muestreando 1091 individuos que fueron clasificados de acuerdo con el **sexo** y la **creencia en la vida después de la muerte**. Vemos las dos variables como respuesta y nos interesa su distribución conjunta.

En los experimentos que responden a este esquema de muestreo, seleccionamos una muestra de  $n$  individuos de una población y registramos los valores  $(X, Y)$  para cada individuo. La distribución conjunta de  $\{n_{ij}\}$  es multinomial de parámetros  $n$  y  $\boldsymbol{\pi} = \{\pi_{ij}\}$ :  $M(n, \boldsymbol{\pi})$ , donde

$$\pi_{ij} = P(X = i, Y = j).$$

En este caso el gran total  $n$  es conocido y fijo. A veces, se expresa los parámetros como medias de las celdas:

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

Esto se conoce como **muestreo multinomial**.

Dado que la distribución multinomial aparecerá con frecuencia en el análisis de datos categóricos, repasaremos algunas de sus propiedades.

## Distribución Multinomial

Supongamos que realizamos  $n$  ensayos independientes, de manera que cada ensayo puede resultar en uno de los eventos  $E_1, \dots, E_N$  (los  $E_j$ 's son mutuamente excluyentes y exhaustivos). En cada ensayo, el evento  $E_j$  puede ocurrir con probabilidad  $\pi_j$  y por lo tanto  $\pi_1 + \dots + \pi_N = 1$ .

Si llamamos

$$X_j = \text{número de veces que el evento } E_j \text{ ocurre,}$$

entonces

$$X_1 + \dots + X_N = n$$

y la distribución de  $\mathbf{X} = (X_1, \dots, X_N)'$  es multinomial de parámetros  $n$  y  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)'$ , es decir

$$\mathbf{X} = (X_1, \dots, X_N)' \sim M(n, \pi_1, \dots, \pi_N)$$

$$\mathbf{X} = (X_1, \dots, X_N)' \sim M(n, \boldsymbol{\Pi}).$$

De manera que:

$$\left\{ \begin{array}{ll} P(X_1 = x_1, \dots, X_N = x_N) = \frac{n!}{x_1! \dots x_N!} \pi_1^{x_1} \dots \pi_N^{x_N} & \text{si } \sum_{i=1}^N x_i = n \\ 0 & \text{caso contrario (c.c.)} \end{array} \right.$$

Como en el caso de la distribución binomial, algunas propiedades más elementales de esta distribución, como el cálculo de esperanza o matriz de covarianza, se deducen fácilmente pensando a  $\mathbf{X}$  como una suma de vectores de 0's y 1's. Más precisamente, podemos escribir

$$\mathbf{X} = \mathbf{Y}_1 + \dots + \mathbf{Y}_n$$

donde las  $\mathbf{Y}_i$  son independientes y cada una es  $M(1, \pi_1, \dots, \pi_N) = M(1, \mathbf{\Pi})$ .

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

$\mathbf{Y}_i$  es un vector con un 1 en la coordenada  $j$  si  $E_j$  ocurrió en el  $i$ -ésimo ensayo y en el resto de las posiciones 0's.

Los elementos de  $\mathbf{Y}_i$  son Bernoulli correlacionadas.

## Esperanza y Varianza

Por ejemplo, supongamos que  $N = 2$  y que  $\mathbf{Y} \sim M(1, \pi_1, \pi_2)$ . Los resultados posibles son

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ con probabilidad } \pi_1$$
$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ con probabilidad } \pi_2 = 1 - \pi_1$$

La media de  $\mathbf{Y}$  es

$$E(\mathbf{Y}) = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}$$

El segundo momento de  $\mathbf{Y}$  es:

$$E(\mathbf{Y}\mathbf{Y}') = E \begin{pmatrix} Y_1^2 & Y_1Y_2 \\ Y_1Y_2 & Y_2^2 \end{pmatrix}$$



$$= \begin{pmatrix} \pi_1 & 0 \\ 0 & \pi_2 \end{pmatrix}$$

Por lo tanto, la matriz de covarianza de  $\mathbf{Y}$  es:

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{Y}} &= E(\mathbf{Y}\mathbf{Y}') - E(\mathbf{Y})E(\mathbf{Y}') \\ &= \begin{pmatrix} \pi_1 & 0 \\ 0 & \pi_2 \end{pmatrix} - \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} (\pi_1 \quad , \quad \pi_2) \\ &= \begin{pmatrix} \pi_1(1 - \pi_1) & \pi_1\pi_2 \\ \pi_1\pi_2 & \pi_2(1 - \pi_2) \end{pmatrix} \end{aligned}$$

En general, se puede demostrar que si hay  $N$  resultados posibles e  $\mathbf{Y} \sim M(1, \pi_1, \dots, \pi_N)$

$$E(\mathbf{Y}) = \boldsymbol{\Pi} = (\pi_1, \dots, \pi_N)'$$

$$\Sigma_{\mathbf{Y}} = \Delta(\mathbf{\Pi}) - \mathbf{\Pi}\mathbf{\Pi}'$$

Luego, si  $\mathbf{X} \sim M(n, \pi_1, \dots, \pi_N)$

$$\begin{aligned} E(\mathbf{X}) &= n\mathbf{\Pi} \\ \Sigma_{\mathbf{X}} &= \Delta(n\mathbf{\Pi}) - n\mathbf{\Pi}\mathbf{\Pi}' \end{aligned}$$

Probaremos que si  $\pi_j \neq 0, \forall j$

$$rg(\Sigma_{\mathbf{X}}) = N - 1.$$

Este es un resultado intuitivo, ya que las variables  $X_1, \dots, X_N$  son linealmente dependientes. Si alguna  $\pi_j = 0$ , entonces el rango disminuye ( $N - 1 - \#\{\pi_j = 0\}$ ), lo que también es esperable, si pensamos que estamos imponiendo nuevas restricciones, en tanto por cada  $\pi_j = 0$  imponemos que el  $X_j$  correspondiente también se anule.

## Otras Propiedades

Enumeraremos algunas propiedades de la distribución multinomial que nos resultarán útiles, algunas serán ejercicio de la práctica.

1. El espacio paramétrico natural de la distribución multinomial es el simplex, hipertetraedro en  $R^N$  definido por

$$\mathcal{S} = \{\boldsymbol{\pi} : \pi_j, \pi_1 + \dots + \pi_N = 1\}.$$

El simplex puede verse como un espacio  $N - 1$ - dimensional en  $R^N$ . Por ejemplo, en  $R^3$  sería el tetraedro con vértices  $(1, 0, 0)$ ,  $(0, 1, 0)$  y  $(0, 0, 1)$  que muestra el gráfico

Por otro lado, como  $\pi_N = 1 - \sum_{j=1}^{N-1} \pi_j$ , podemos pensar en un espacio en  $R^{N-1}$ . Cuando  $N = 3$ , en  $R^2$  podríamos ver el espacio paramétrico como el triángulo de vértices  $(1,0)$  y  $(0,1)$  como se muestra en la Figura 2.

2. Si  $\mathbf{X} = (X_1, \dots, X_N)' \sim M(n, \pi_1, \dots, \pi_N)$ , entonces

$$\begin{aligned} X_j &\sim Bi(n, \pi_j) \\ Cov(X_i, X_j) &= -n \pi_i \pi_j \quad i \neq j, \end{aligned}$$

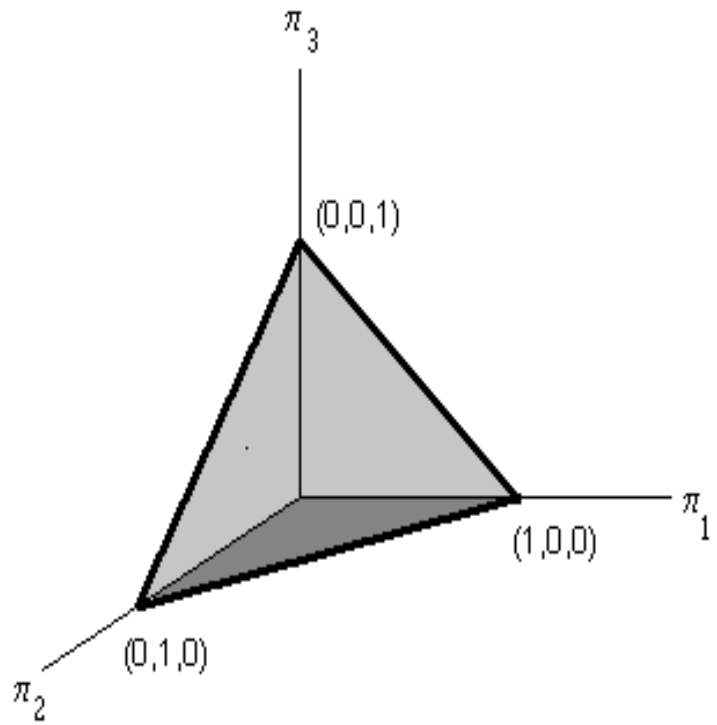


Figura 1: Simplex en  $R^3$

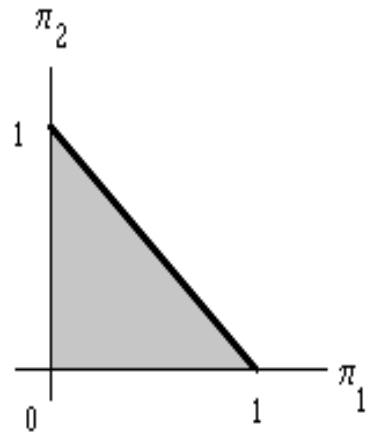


Figura 2: Simplex en  $R^2$

es decir las  $X_i$  están negativamente correlacionadas.

3. Si  $\mathbf{X} = (X_1, \dots, X_N)' \sim M(n, \pi_1, \dots, \pi_N)$ , entonces

$$\mathbf{X}^* = (X_1 + X_2, X_3, \dots, X_N)' \sim M(n, \pi_1 + \pi_2, \pi_3, \dots, \pi_N)$$

Es decir, si se colapsa una multinomial sumando celdas la distribución sigue siendo multinomial.

4. Sea  $\mathbf{X} = (X_1, \dots, X_N)' \sim M(n, \pi_1, \dots, \pi_N)$ . Consideremos la distribución condicional de

$$\mathbf{X} \mid \begin{array}{l} X_1 + X_2 = z \\ X_3 + \dots + X_N = n - z \end{array}$$

Los vectores  $(X_1, X_2)'$  y  $(X_3, \dots, X_N)'$  son condicionalmente independientes y multinomiales:

$$(X_1, X_2)' \sim M\left(z, \frac{\pi_1}{\pi_1 + \pi_2}, \frac{\pi_2}{\pi_1 + \pi_2}\right)$$

$$(X_3, \dots, X_N)' \sim M\left(n - z, \frac{\pi_3}{\pi_3 + \dots + \pi_N}, \dots, \frac{\pi_N}{\pi_3 + \dots + \pi_N}\right)$$

5. Si  $X_1, \dots, X_N$  son variables independientes tales que  $X_j \sim \mathcal{P}(\lambda_j)$ , entonces

$$(X_1, \dots, X_N)' |_{\sum_{j=1}^N X_j = n} \sim M(n, \pi_1, \dots, \pi_N)$$

donde

$$\pi_j = \frac{\lambda_j}{\lambda_1 + \dots + \lambda_N}$$

Por lo tanto, la distribución de  $X_1, \dots, X_N$  puede ser factorizada en el producto de

$$n = \sum_{j=1}^N X_j \sim \mathcal{P}\left(\sum_{j=1}^N \lambda_j\right)$$

y

$$(X_1, \dots, X_N)' |_{n=n^*} \sim M(n^*, \pi_1, \dots, \pi_N)$$

Esto será especialmente útil a la hora de calcular la función de verosimilitud bajo ciertas condiciones.

En nuestro ejemplo hemos hablado de la función de verosimilitud.

¿Cómo sería en el caso general de una multinomial cualquiera?

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

Luego la función de verosimilitud será:

$$L = L(n_1, \dots, n_N, \pi_1, \dots, \pi_N) = \frac{n!}{\prod_{i=1}^N n_i!} \prod_{i=1}^N \pi_i^{n_i} \quad \text{donde } \sum_{i=1}^I \pi_i = 1$$

Como el  $\ln$  es una función estrictamente creciente y  $n! / \prod_{i=1}^N n_i!$  es constante, hallar el máximo de  $L$  equivale a hallar el máximo de

$$l = \ln L = \ln \left( \frac{n!}{\prod_{i=1}^N n_i!} \right) + \sum_{i=1}^N n_i \ln \pi_i \quad \text{donde } \sum_{i=1}^N \pi_i = 1$$



Como  $\sum_{i=1}^N \pi_i = 1$ , entonces  $\pi_N = 1 - \sum_{i=1}^{N-1} \pi_i$ , por lo tanto buscamos el máximo de

$$l = \ln L = cte + \sum_{i=1}^{N-1} n_i \ln \pi_i + n_N \ln\left(1 - \sum_{i=1}^{N-1} \pi_i\right)$$

Para buscar los puntos críticos planteamos:

$$\frac{\partial l}{\partial \pi_i} = \frac{n_i}{\pi_i} - \frac{n_N}{1 - \sum_{i=1}^{N-1} \pi_i} = \frac{n_i}{\pi_i} - \frac{n_N}{\pi_N} = 0$$

Luego,

$$\frac{n_i}{n_N} = \frac{\pi_i}{\pi_N} \Rightarrow \frac{n}{n_N} = \frac{n_N}{\pi_N} \Rightarrow \hat{\pi}_N = \frac{n_N}{n}$$

$$\Rightarrow \hat{\pi}_i = \frac{n_i \hat{\pi}_N}{n_N} = \frac{n_i}{n}$$

Por lo tanto, tal como es de esperar

$$\widehat{\pi}_i = \frac{n_i}{n} = p_i \quad i = 1, \dots, N$$

## Muestreo Poisson

Otra posibilidad es tratar a los 4 datos de la tabla del ejemplo como si viniesen de realizaciones independientes de v.a. con distribución Poisson.

Un proceso que sustentaría este modelo es que en cada celda los individuos llegan aleatoriamente al lugar donde se los clasifica. En este caso  $n$  no está pre-fijado y todos los valores de la tabla son aleatorios.

En el **muestreo de Poisson** tenemos que

$$n_{ij} \sim \mathcal{P}(m_{ij}) \quad \begin{array}{l} i = 1, \dots, I \\ j = 1, \dots, J \end{array}$$

independientes. En este esquema el gran total no está fijo, sino que es aleatorio.

**Ejemplo:**

Supongamos que se realiza en un control de velocidad durante una hora. Para ello se cuenta con un radar que registra la velocidad de cada auto que pasa por el puesto de observación. Supongamos que de cada auto que pasa se registra la velocidad y la marca del auto. Así se obtienen

$X$  = marca del auto (1 = Ford, 2 = Fiat, 3 = Chevrolet, 4 = Otros)

$Y$  = si el auto excede el límite de velocidad (1 = Si, 0 = No).

Es claro, que  $n \sim \mathcal{P}(m_{++})$ .

Si tenemos un muestreo de Poisson, la distribución de los  $n_{ij}$  condicional a que  $n$  está fijo en un valor, digamos  $n^*$ , ya no es más Poisson, más aún ya no son más independientes.

La distribución condicional de los  $n_{ij}$  dado que  $n = n^*$  es multinomial. Para simplificar la notación, indicaremos  $\{n_1, \dots, n_N\}$ , entonces si  $\sum_{i=1}^N k_i = n^*$

$$P(n_1 = k_1, \dots, n_N = k_N \mid \sum_{i=1}^N n_i = n^*) \\ = \frac{P(n_1 = k_1, \dots, n_N = k_N)}{P(n = n^*)}$$

$$\begin{aligned}
&= \prod_{i=1}^N \frac{e^{-m_i} m_i^{k_i}}{k_i!} \frac{n^*!}{e^{-\sum_{i=1}^N m_i} \left\{ \sum_{i=1}^N m_i^{k_i} \right\}^{n^*}} \\
&= \left\{ \prod_{i=1}^N m_i^{k_i} \right\} \frac{n^*!}{\prod_{i=1}^N k_i!} \frac{1}{\left\{ \sum_{i=1}^N m_i \right\}^{n^*}} \\
&= \frac{n^*!}{\prod_{i=1}^N k_i!} \prod_{i=1}^N \pi_i^{k_i}
\end{aligned}$$

donde  $\pi_i = \frac{m_i}{\sum_{j=1}^N m_j}$  o volviendo a la notación original:

$$\pi_{ij} = \frac{m_{ij}}{m_{++}}$$

Entonces:

$$(n_1, \dots, n_N) |_{n=n^*} \sim M(n^*, \pi_1, \dots, \pi_N).$$

Esto nos servirá a la hora de plantear el la verosimilitud cuando deseemos

estimar.

En efecto, podemos factorizar la verosimilitud como el producto del likelihood de la Poisson  $n$  ( $n \sim \mathcal{P}(m_{++})$ ) y el likelihood de una multinomial correspondiente a  $\{n_{ij}\}$  dado  $n$ , con parámetros

$$\pi_{ij} = \frac{m_{ij}}{m_{++}}$$

El total  $n$  no da información acerca de las  $\pi_{ij}$ .

Es interesante observar, que desde el punto de vista de la verosimilitud, la inferencia sobre  $\boldsymbol{\pi}$  es la misma si  $n$  es considerado fijo o aleatorio.

## Muestreo Multinomial Independiente

Otra alternativa podría ser que sea razonable asumir que se tomó una muestra de 582 mujeres y otra muestra independiente de 509 hombres a los que se clasificó según su creencia. En este caso nos centramos en la distribución condicional de la creencia dado cada nivel de sexo.

En este esquema los totales por fila están fijos y tenemos  $n_1$  individuos de sexo femenino y  $n_2$  individuos de sexo masculino.

Si  $\Pi_i$  es la probabilidad de que el individuo crea en la vida después de la muerte para el nivel  $i$  de sexo tendremos:

$$\Pi_i = P(C = 1 | S = i) = \frac{\Pi_{i1}}{\Pi_{i+}}$$

Las variables de interés serán:  $Y_{i1} \sim Bi(n_i, \Pi_i)$ ,  $i = 1, 2$ , que cuentan el número de individuos que sí creen en cada sexo.

La distribución conjunta de  $(Y_{11}, Y_{21})$  es

$$P((Y_{11}, Y_{21}) = (y_{11}, y_{21})) = \frac{n_1!}{y_{11}!y_{12}!} \Pi_1^{y_{11}} (1 - \Pi_1)^{y_{12}} \frac{n_2!}{y_{21}!y_{22}!} \Pi_2^{y_{21}} (1 - \Pi_2)^{y_{22}}$$

La hipótesis de interés es la de **homogeneidad**, es decir que la probabilidad de creencia es la misma en los dos sexos:

$$H_0 : \Pi_1 = \Pi_2$$

Notemos que bajo independencia  $\Pi_{ij} = \Pi_{i+}\Pi_{+j}$ , entonces la probabilidad condicional

$$\Pi_{j|i} = \frac{\Pi_{ij}}{\Pi_{i+}} = \Pi_{+j}$$

es decir no depende de la fila  $i$ , con lo cual homogeneidad e independencia son equivalentes.

Supongamos que decidimos de antemano que vamos a muestrear  $n_{i+}$  individuos con  $X = i$  ( $i = 1, \dots, I$ ) y que para cada uno de ellos registramos el valor de  $Y$ .

En este esquema cada fila de tabla  $(n_{i1}, n_{i1}, \dots, n_{iJ})'$  es multinomial con probabilidades

$$\pi_{j|i} = \frac{\pi_{ij}}{\pi_{i+}}$$

y las filas son independientes.

Este tipo de muestreo es razonable de aplicar cuando los datos provienen de un muestreo aleatorio estratificado (estratos definidos por  $X$ ) o en un experimento en el  $X$  = grupo de tratamiento.

También es adecuado cuando no tenemos totales por filas fijos, pero estamos interesados en  $P(Y|X)$  y no en  $P(X)$ , lo que corresponde a que  $Y$  es el resultado de interés y no deseamos modelar a  $X$ .

Por lo que vimos, si los datos tienen distribución multinomial, el likelihood se puede factorizar en dos factores: el que corresponde al total de las filas y lo que corresponden a las multinomiales de las filas.

La importancia de estos resultados es que el análisis que hagamos es independiente del esquema de muestreo y depende de los parámetros de interés.



## Resumiendo

1. Si lo que nos interesa son funciones de las  $\Pi_{j|i}$  podemos tratar los datos como si viniesen de un producto-multinomial
2. Si el total está fijo podemos tratar a las columnas y a las filas como respuestas y testear independencia a través de un modelo multinomial
3. Podríamos tratar a todas las casillas como aleatorias y ajustar usando una Poisson.

## 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 verosimilitud  $L(\theta, y)$  es simplemente  $f(y, \theta)$  mirada como función de  $\theta$  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  $\Theta$ .

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

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

que está definido salvo una constante aditiva.

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

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

y la log-verosimilitud

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

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  $\phi$ ,  $\widehat{\phi}$ , se calcula como

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

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

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

La derivada primera de  $l(\theta)$  respecto de  $\theta$  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,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$ , el vector de score es

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

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

$$E(l'(\boldsymbol{\theta}))|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \int l'(\boldsymbol{\theta}_0) f(y, \boldsymbol{\theta}_0) dy = \int \frac{f'(y, \boldsymbol{\theta}_0)}{f(y, \boldsymbol{\theta}_0)} f(y, \boldsymbol{\theta}_0) dy = 0$$

La varianza de los score  $u(\boldsymbol{\theta})$  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

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

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

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

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

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

## Estimación y Tests de Bondad de Ajuste

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

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

Cuadro 9: 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  $\pi_{ij}$  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 invariancia, el estimador de máxima verosimilitud de  $\pi_{ij}$  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:

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

## Test de Bondad de Ajuste

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

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

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

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

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

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

entonces

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

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

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

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

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

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  $\chi_1^2$ .

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  $\mathbf{X}_n = (X_{1n}, \dots, X_{kn})'$  una sucesión de vectores aleatorios y  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)' \in \mathfrak{R}^k$ .

Si  $\forall \boldsymbol{\lambda} \in \mathfrak{R}^k$

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

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

## Teorema Central del Límite Multivariado (TCLM)

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

Si  $\bar{\mathbf{U}}_n = (\bar{U}_{1n}, \dots, \bar{U}_{kn})'$  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 anterior debemos estudiar la distribución de  $\boldsymbol{\lambda}'\bar{\mathbf{U}}_n$ .

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

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

Por el TCL univariado, tenemos que

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

es decir,

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

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

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

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

Consideremos el vector  $\mathbf{Y}_i \sim M(1, \pi_1, \dots, \pi_N)$  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  $j$ :

$$\mathbf{Y}_i = (0, \dots, \underset{j}{1}, \dots, 0)$$

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

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

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

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

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  $\pi_i$ 's están relacionados,  $\Sigma = \Delta(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'$  no es invertible.

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

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

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

Calculando efectivamente la forma cuadrática que estamos considerando, veremos que

$$n(\tilde{\mathbf{p}} - \tilde{\boldsymbol{\pi}}_0)' \tilde{\Sigma}_0^{-1} (\tilde{\mathbf{p}} - \tilde{\boldsymbol{\pi}}_0) = n \sum_{j=1}^N \frac{(p_j - \pi_{j0})^2}{\pi_{j0}}$$

### **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  $\chi^2$  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 frecuencia 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  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$ , como muestran los siguientes ejemplos.

## Test de Independencia

### Independencia en una tabla de $2 \times 2$

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

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

Cuadro 10:

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

Para hallar los estimadores de máxima verosimilitud de  $\alpha$  y  $\beta$  tenemos que maximizar:

$$L = L(X_{11}, X_{12}, X_{21}, X_{22}, \alpha, \beta) =$$

$$= \frac{n!}{X_{11}!X_{12}!X_{21}!X_{22}!}(\alpha\beta)^{X_{11}}(\alpha(1-\beta))^{X_{12}}((1-\alpha)\beta)^{X_{21}}((1-\alpha)(1-\beta))^{X_{22}}$$

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

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

De (2) resulta:

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

por lo tanto

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

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 $2 \times 2$ 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.

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

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

	Febrero	
Enero	1	0
1	$\pi_{11}$	$\pi_{12}$
0	$\pi_{21}$	$\pi_{22}$

Cuadro 11:

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:

$$\hat{\alpha} = \frac{X_{11}}{n}.$$

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



## 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,  $\hat{\theta}$  es solución de las ecuaciones de "score":

$$\frac{\partial}{\partial \theta_j} l(\boldsymbol{\pi}(\theta), X) = 0, \quad j = 1, \dots, q. \quad (4)$$

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

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

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  $\hat{\theta}$  en términos de  $p$  y luego aplicando el método  $\Delta$ .

Esto nos permitirá derivar la distribución del estadístico  $\chi^2$  en casos bastante generales.

Para que el EMV de  $\theta$  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  $\theta_r$  tal que  $\boldsymbol{\pi}(\theta_r) \rightarrow \boldsymbol{\pi}(\theta_0)$  a medida que  $r \rightarrow \infty$ , es decir que no hay valores  $\theta$  lejanos a  $\theta_0$  que den prácticamente las mismas probabilidades que  $\boldsymbol{\pi}(\theta_0)$ .

Es decir:

$\forall \delta > 0$ , existe  $\epsilon > 0$  tal que si  $\|\theta - \theta_0\| > \delta$  entonces  $\|\pi(\theta) - \pi(\theta_0)\| > \epsilon$ .

Esta condición implica una más débil:

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

Bajo la condición fuerte de identificabilidad y continuidad de las funciones  $\pi_i(\theta)$ , se puede demostrar que el EMV de  $\theta$  existe y que converge a  $\theta_0$  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

$$\frac{\partial L}{\partial \theta_j} = 0, \quad j = 1, \dots, q. \quad (5)$$

Por último, si  $\pi(\theta) \neq \pi(\beta)$  si  $\theta \neq \beta$ , las funciones  $\pi_i(\theta)$  tienen derivadas parciales de primer orden continuas en  $\theta_0$  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  $\theta_0$  no es singular, entonces existe una raíz consistente de (5) 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  $P(x_1, \dots, x_n, \theta)$  es una función de probabilidad de las variables aleatorias  $X_1, \dots, X_n$ . Sea  $Z_n = (z_n^1, \dots, z_n^q)'$  el vector de derivadas

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

y el vector de sesgos

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

siendo  $T_n = (T_n^1, \dots, T_n^q)'$  es un estimador consistente de  $(\theta_1, \dots, \theta_q)'$ .

Diremos que  $T_n$  es un estimador eficiente si

$$\sqrt{n}\|B_n - \mathbf{D}Z_n\| \xrightarrow{p} 0 \tag{6}$$

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

- ¿Qué diría esta definición en el caso univariado?

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

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

$T_n$ , estimador consistente de  $\theta$ , será eficiente si

$$\sqrt{n}(T_n - \theta - \delta(\theta)Z_n) \xrightarrow{p} 0 \quad (7)$$

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  $X_1, \dots, X_n$  una sucesión de v. a. i.i.d. con densidad  $p(x, \theta)$ , siendo  $\theta \in \mathfrak{R}$ . Además supongamos que

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

Entonces, la condición (7) implica que la distribución asintótica de  $\sqrt{n}(T_n - \theta)$  es normal:

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

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

Por (7),  $\sqrt{n}(T_n - \theta)$  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 (6) 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  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)'$ , donde  $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}) = \pi(\theta_1, \dots, \theta_q)'$ . 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  $\chi^2$  correspondiente.

El resultado que probaremos es el siguiente:

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

- a)  $\widehat{\boldsymbol{\theta}}$  es un estimador eficiente de  $\boldsymbol{\theta}$  en el sentido de (6)
- b) Cada  $\pi_i(\boldsymbol{\theta})$  admite derivadas parciales continuas de 1<sup>er</sup> orden respecto de  $\theta_j$ ,  $j = 1, \dots, q$ ,  $i = 1, \dots, N$ .
- c) La matriz  $\mathbf{M} = \left\{ \frac{\partial \pi_r}{\partial \theta_s} \right\}$  de  $N \times q$  calculada en los verdaderos  $\boldsymbol{\theta}$  tiene rango  $q$ .

Luego, tenemos que

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



Comenzaremos por probar el siguiente resultado auxiliar.

**Lema:** Supongamos que  $\theta_0$ , 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<sup>er</sup> orden continuas en  $\theta_0$ .
- c) La matriz  $\{i_{rs}\}$  es no singular en  $\theta_0$ , 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  $\frac{\partial l}{\partial \theta_i} = 0 \quad i = 1, \dots, q$ , ecuación de verosimilitud y

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

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  $\frac{\partial l}{\partial \theta_i} = 0$  es eficiente y su distribución asintótica es normal  $q$ -variada, siendo

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

donde

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

## Esquema de demostración

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

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

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

$$\sum_{i=1}^N \sqrt{n} \frac{p_i - \pi_i(\theta_0)}{\bar{\pi}_i} \frac{\partial \pi_i}{\partial \bar{\theta}_r} = \sum_{i=1}^N \sum_{s=1}^q \sqrt{n} (\bar{\theta}_s - \theta_{0s}) \frac{\partial \pi_i}{\partial \bar{\theta}_r} \frac{\partial \pi_i}{\partial \tilde{\theta}_s} \frac{1}{\bar{\pi}_i} \quad (8)$$

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

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

reemplazando a  $\bar{\boldsymbol{\pi}}$  por  $\boldsymbol{\pi}_0$  en (8) podemos reescribir (8) :

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

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

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

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  $\mathfrak{R}$  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  $\Delta$  generaliza este resultado para una función de vector aleatorio.

**Lema 2: Método  $\Delta$  una función de vector aleatorio.**

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

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

Sea  $g$  una función tal que  $g : \mathfrak{R}^N \longrightarrow \mathfrak{R}$  diferenciable. Luego si

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

entonces

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

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

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

donde

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

Aplicando todos los resultados anteriores obtenemos que:

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

Notemos que el estadístico  $\chi^2$  puede escribirse como

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

donde

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

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

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

**Teorema:** Sea  $Y$  un vector con distribución  $N(\boldsymbol{\nu}, \Sigma)$ . Una condición necesaria y suficiente para que  $(Y - \boldsymbol{\nu})' \mathbf{C} (Y - \boldsymbol{\nu})$  tenga distribución  $\chi^2$  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  $\Sigma$  es no singular la condición se simplifica a  $\mathbf{C} \Sigma \mathbf{C} = \mathbf{C}$ ). (Rao, 1965, p. 150)

Como hemos visto,  $\chi^2 = \mathbf{e}' \mathbf{e}$ , luego aplicaremos el resultado de Rao con  $\boldsymbol{\nu} = 0$ ,  $\mathbf{C} = I$ ,  $\Sigma = I - \boldsymbol{\pi}(\theta_0)^{1/2} \boldsymbol{\pi}'(\theta_0)^{1/2} - A(A'A)^{-1}A'$ , por lo que resultará que

$$\mathbf{e}' \mathbf{e} \xrightarrow{\mathcal{D}} \chi_{N-1-q}^2$$



## 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

<b>S:</b> Sexo	<b>C:</b> Identificación partidaria			Total
	Demócrata	Independiente	Republicano	
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  $\chi^2$  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 probabilidades dependen de una cantidad menor de parámetros desconocidos,  $\theta$ .

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:

	<b>Segunda Infección</b>	
	Si	No
<b>Primera Infección</b>		
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 infección era igual que la probabilidad 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:

	<b>Segunda Infección</b>		
	Si	No	Total
<b>Primera Infección</b>			
Si	$\pi^2$	$\pi(1 - \pi)$	$\pi$
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

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

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

El estadístico de Pearson da  $\chi^2 = 19.7$  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.

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

Cuadro 15:

## Estadístico $G^2$

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

Si queremos testear

$H_0$  : Modelo restringido  $\omega$

$H_1$  : Modelo Saturado  $\Omega$ ,

el cociente estaría dado por

$$\Lambda = \frac{\text{máx}_\omega L}{\text{máx}_\Omega L}$$

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

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

Probaremos que bajo  $H_0$  la distribución límite de  $G^2$  es también  $\chi^2$  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  $\chi^2$  puede calcularse sin problemas, siempre que las  $\hat{\pi}$ 's sean todas positivas. Sin embargo, el estadístico  $G^2$  tiene problemas, pues si  $X_i = 0$ , entonces  $X_i \log \frac{X_i}{n\hat{\pi}_i}$  no está definido. Si reescribimos a  $G^2$  como

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

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

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

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



## Cuán grande debe ser $n$ para tener una buena aproximación?

Sabemos que a medida que  $n$  crece la distribución de  $\chi^2$  y de  $G^2$  se aproximan a una distribución límite  $\chi^2$ , sin embargo nos preguntamos cuán grande es grande.

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

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

- En tablas *sparse* (esparcidas????)(por ejemplo,  $n/N < 5$ ) la aproximación  $\chi^2$  es pobre. En realidad, si los datos están distribuidos en la tabla de forma muy desigual, en el sentido de que hay zonas de la tabla que son *sparse*, la aproximación  $\chi^2$  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

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

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

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

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

Observemos que  $\lambda$  tiene la forma del estadístico  $\chi^2$  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  $\chi^2$  y  $G^2$  tienen el mismo parámetro de no centralidad, pero éste no es siempre el caso.

## Residuos de Pearson y deviance

Como ya hemos visto podemos escribir al estadístico de Pearson como

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

donde

$$e_i = \sqrt{n} \frac{p_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i}}$$

se 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  $|e_i|$  con 2, indicándose falta de ajuste en la  $i$ -ésima casilla si  $|e_i| > 2$ . 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 cuadra-

dos de residuos

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

donde

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

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  $2 \times 2$ , como las que siguen

X	Y		Total	X	Y		Total
	1	2			1	2	
1	$\pi_{11}$	$\pi_{12}$	$\pi_{1+}$	1	$n_{11}$	$n_{12}$	$n_{1+}$
2	$\pi_{21}$	$\pi_{22}$	$\pi_{2+}$	2	$n_{21}$	$n_{22}$	$n_{2+}$
Total	$\pi_{+1}$	$\pi_{+2}$	1	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		Total
	si	no	
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 sería 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

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

Dado que  $\delta$  es función de los parámetros de  $P(Y|X)$ , los estimadores de máxima verosimilitud serán los mismos bajo los tres tipos de muestreo que hemos visto.

Podemos estimar a  $\delta$  como

$$d = \frac{n_{11}}{n_{1+}} - \frac{n_{21}}{n_{2+}}$$

$$= p_{1|1} - p_{1|2}$$

En el ejemplo de Infarto de Miocardio tenemos

$$d = 104/11037 - 189/11034 = 0.0094 - 0.0171 = -0.0077$$

Observemos que

$$E(d) = E(p_{1|1} - p_{1|2}) = \pi_{1|1} - \pi_{1|2}$$

$$V(d) = V(p_{1|1} - p_{1|2}) = \frac{\pi_{1|1}(1 - \pi_{1|1})}{n_{1+}} + \frac{\pi_{1|2}(1 - \pi_{1|2})}{n_{2+}}$$

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 asintótico para  $\delta$  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. Si estamos trabajamos con eventos poco frecuentes  $\pi_{1|1}$  y  $\pi_{1|2}$  serán muy pequeñas y  $\delta$  será cercano a 0, aún cuando el efecto sea importante, como en el ejemplo. 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 riesgo relativo igual a 1 corresponde a independencia.

$RR$  esta medida es sólo función de  $P(Y|X)$ , por lo tanto la inferencia que hagamos sobre  $RR$  será la misma para los tres muestreos que hemos considerado. La comparación en la otra respuesta da otro riesgo relativo.

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 una aproximación normal a su logaritmo es buena 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  $\Delta$  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 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  $\pi_{1|i} = 0$  suele usarse

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

## 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 **odds** o **chance**. El odds de un suceso es

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

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 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, en el grupo placebo la chance de no tener infarto es 57.47 veces la de infarto, mientras que en el grupo tratado, la chance de no tener infarto es 105.26 veces la de infarto.

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

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

Como antes observamos que al ser esta medida función de  $P(Y|X)$ , la inferencia es válida para los tres muestreos vistos.

El EMV es

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

Las propiedades de  $\hat{\theta}$  son fáciles 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  $\log(\hat{\theta})$

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

donde

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

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.