

# Clase 9: El test $\chi^2$

Matías Carrasco

5 de octubre de 2019

## Índice

1. La razón de verosimilitud	1
2. La distribución $\chi^2$	2
3. El Teorema de Wilks	3
4. Test de bondad de ajuste para distribuciones discretas	4
5. Test de independencia para tablas de contingencia	9

## 1. La razón de verosimilitud

En muchas situaciones puede que no sea tan claro cómo elegir el estadístico para el TdH. Esto puede ser porque el modelo de probabilidad adoptado involucra varios parámetros desconocidos, como puede ser la varianza  $\sigma^2$  en el caso de un test sobre la media  $\mu$ , o porque simplemente no intuimos una fórmula clara para el estadístico. El método de *la razón de verosimilitud* sirve en estos casos.

La idea es muy sugestiva. Supongamos que tenemos un muestreo i.i.d.  $X_1, \dots, X_n$  con distribución  $p(x; \theta)$  que depende de un parámetro  $\theta = (\theta_1, \dots, \theta_k)$ . El espacio de parámetros  $P$  está dividido en

$$P = P_0 \cup P_A$$

y que queremos hacer un test

$$\begin{cases} H_0 : \theta \in P_0 \\ H_A : \theta \in P_A \end{cases}$$

Por ejemplo,  $P_0$  podría ser  $\theta \geq 100$  y  $P_1$  podría ser  $\theta < 100$ .

Si bien el interés está la hipótesis nula  $\theta \in P_0$ , puede que la distribución  $p(x; \theta)$  no quede determinada sabiendo que  $\theta \in P_0$ . Es decir, algunas de las coordenadas de estos parámetros pueden ser parámetros desconocidos. Este tipo de parámetros, que son desconocidos pero no son en los que estamos interesados, se llaman *parámetros molestos*.

La idea es ver si las hipótesis de  $P_0$  pueden explicar los datos observados usando la función de verosimilitud

$$L(\theta) = \prod_{i=1}^n p(X_i; \theta).$$

El argumento informal para esto es el siguiente. Para una realización de los datos, determinamos la mejor verosimilitud entre aquellas hipótesis en  $P_0$ , y también la mejor verosimilitud en general. La razón de estas dos verosimilitudes nunca puede exceder la unidad, pero, si es pequeña, será evidencia en contra de la hipótesis nula.

Calculamos la mejor verosimilitud de cada hipótesis en  $P_0$  haciendo

$$L_0 = \sup_{\theta \in P_0} L(\theta),$$

y la mejor verosimilitud en general

$$L_{\text{máx}} = \sup_{\theta} L(\theta).$$

Si  $H_0$  es cierta, es razonable esperar una *razón de verosimilitud*  $q_L = L_0/L_{\text{máx}}$  cercana a 1.

Para probar  $H_0$  contra  $H_A$  podemos usar la región de rechazo  $I_L = \{q_L \leq k\}$ , donde  $k$  es un número real entre 0 y 1. Claramente, la prueba tendrá un nivel de significación  $\alpha$  si  $k$  puede elegirse para satisfacer

$$\sup_{\theta \in P_0} \mathbf{P}(q_L \leq k | \theta) = \alpha.$$

Si  $H_0$  es una hipótesis simple con  $P_0 = \{\theta_0\}$ , tenemos la relación más simple

$$\mathbf{P}(q_L \leq k | \theta_0) = \alpha$$

Para determinar  $k$ , debemos conocer la distribución de la variable aleatoria  $q_L$  bajo  $H_0$ , en donde la muestra aleatoria  $X_1, \dots, X_n$  proviene de la distribución  $p(x; \theta_0)$ .

## 2. La distribución $\chi^2$

Sean  $Z_1, \dots, Z_k$  variables aleatorias independientes con distribución  $N(0, 1)$ . Entonces la suma de sus cuadrados,

$$Q = \sum_{i=1}^k Z_i^2,$$

tiene (por definición) distribución *chi-cuadrado con  $k$  grados de libertad*. Esto se escribe usualmente como  $Q \sim \chi^2(k)$  o  $Q \sim \chi_k^2$ . Geométricamente representa el cuadrado de la distancia al origen del vector  $(Z_1, \dots, Z_k)$  de  $\mathbb{R}^k$ .

La distribución chi-cuadrado depende de un parámetro  $k$  que es un entero positivo que especifica los *grados de libertad* de la distribución

$$k = \text{grados de libertad} = \text{el número de } Z_i\text{'s.}$$

La función de densidad de  $Q$  está dada por:

$$p_Q(x) = \begin{cases} \frac{x^{(k/2-1)} e^{-x/2}}{2^{k/2} \Gamma(\frac{k}{2})}, & x \geq 0; \\ 0, & \text{sino.} \end{cases}$$

en donde  $\Gamma$  es la función Gamma. Pero no la usaremos nunca, ya que disponemos de tablas para calcular probabilidades con la  $\chi^2$ .

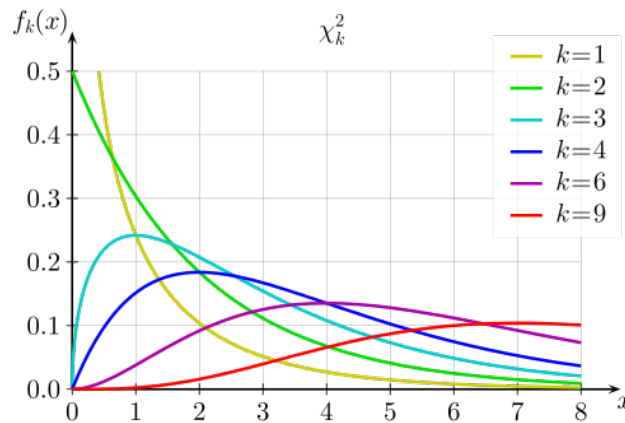


Figura 1: Se muestra las densidades  $\chi_k^2$  para varios valores de  $k$ .

Al final de las notas se muestra una tabla con algunos valores críticos de  $\chi^2$ . Las filas representan diferentes valores de los grados de libertad  $k$  (df por sus siglas en inglés). Las columnas representan los valores críticos para cada probabilidad. Por ejemplo, el valor crítico asociado a la probabilidad 0.05 y 10 grados de libertad es 18.31. Esto quiere decir que  $\mathbf{P}(\chi_{10}^2 \geq 18.31) = 0.05$ . En general, dada la probabilidad  $\alpha \in (0, 1)$ , se denota por  $\chi_k^2(\alpha)$  al valor crítico (de  $k$  grados de libertad) asociado a la probabilidad  $\alpha$ , y esto quiere decir que  $\mathbf{P}(\chi_k^2 \geq \chi_k^2(\alpha)) = \alpha$ . De este modo, en el ejemplo anterior tenemos  $\chi_{10}^2(0.05) = 18.31$ .

### 3. El Teorema de Wilks

Como la función  $-2 \ln x$  es decreciente, la región de rechazo  $\{q_L \leq k\}$  de la razón de verosimilitud puede también escribirse de la forma

$$I = \{-2 \ln q_L \leq c\}$$

para una constante  $c$ . Escribiendo

$$Q_L = -2 \ln q_L = 2(\ln L_{\max} - \ln L_0)$$

la región de rechazo queda  $I = \{Q_L \geq c\}$ . El estadístico  $Q_L$  también se lo suele llamar el *estadístico de la razón de verosimilitud*.

Cuando estamos interesados en hacer el test

$$\begin{cases} H_0 : \theta \in P_0 \\ H_A : \theta \in P_A \end{cases}$$

el estadístico de la razón de verosimilitud se escribe como

$$Q_L = 2 \left( \ell_{\max} - \sup_{\theta \in P_0} \ell(\theta) \right).$$

El resultado fundamental que hace que este estadístico sea muy útil es que su distribución bajo la hipótesis nula se puede aproximar (para muestras razonablemente grandes) por la distribución  $\chi^2$ . Este es el contenido del Teorema de Wilks.

### Teorema de Wilks

Asumiendo que  $H_0$  es cierta,  $Q_L \stackrel{d}{\approx} \chi_p^2$  en donde  $p = \dim P - \dim P_0$ .

## 4. Test de bondad de ajuste para distribuciones discretas

### Hipótesis nula simple

Supongamos que disponemos de  $k$  grupos con  $n_i$  observaciones en el  $i$ -ésimo grupo.

Grupo	1	2	3	4	...	$k$
Frecuencia	$n_1$	$n_2$	$n_3$	$n_4$	...	$n_k$

El total de observaciones es  $n = \sum_{i=1}^k n_i$ . Queremos hacer un test sobre las frecuencias observadas de cada grupo, y por ende el espacio de parámetros es en este caso

$$P = \left\{ \pi = (\pi_1, \dots, \pi_k) : \sum_{i=1}^k \pi_i = 1 \right\}$$

en donde  $\pi_i$  es probabilidad del  $i$ -ésimo grupo.

Nuestro test es para evaluar si el modelo

$$p = (p_1, \dots, p_k), \quad \sum_{i=1}^k p_i = 1$$

explica correctamente las frecuencias observadas  $n_i/n$ . Es decir, queremos hacer el siguiente test

$$\begin{cases} H_0 : \text{las obs. se ajustan al modelo} \\ H_A : \text{las obs. no se ajustan al modelo} \end{cases}$$

De forma equivalente podemos escribir la hipótesis nula como  $H_0 : \pi = p$ , que es una hipótesis simple. Recordar que esto quiere decir que  $H_0$  determina la distribución nula.

La función de verosimilitud es en este caso

$$L(\pi) = \frac{n!}{n_1! \dots n_k!} \prod_{i=1}^k \pi_i^{n_i}$$

Para verificarlo basta pensar en la distribución aleatoria de bolas en celdas:

$$\begin{array}{|c|c|c|c|} \hline n_1 & n_2 & \dots & n_k \\ \hline \pi_1 & \pi_2 & \dots & \pi_k \\ \hline \end{array}$$

Al tomar logaritmo

$$\ell(\pi) = (\text{cte}) + \sum_{i=1}^k n_i \ln \pi_i$$

Para calcular la mejor verosimilitud en  $H_0$  debemos maximizar con la condición  $\sum_{i=1}^k \pi_i = 1$ . Resulta que el máximo se da en  $\hat{\pi}_i = n_i/n$ , por lo que

$$Q_L = 2 \sum_{i=1}^k n_i \left[ \ln \left( \frac{n_i}{n} \right) - \ln p_i \right]$$

Como la  $\dim P = k - 1$  y  $\dim P_0 = 0$ , el Teorema de Wilks nos permite aproximar la distribución nula de  $Q_L$  por una  $\chi^2$  con  $k - 1$  grados de libertad.

### Ejemplo 1

Ana y Beto son los profesores responsables de PyE, y gastaron una fortuna en dados y monedas sesgadas para el curso, por lo que decidieron enviar una factura al Instituto de Matemática para su reembolso. El Instituto sospecha que su informe de gastos (de seis cifras) está trucado, por lo que te llaman para que inspecciones los datos en busca de un fraude. Investigando un poco, recordas que los datos contables se ajustan a la llamada Ley de Benford. Esta establece que la frecuencia relativa del primer dígito (no nulo) de cada entrada debe tener la siguiente distribución:

Dígito $k$ :	1	2	3	4	5	6	7	8	9
Prob. $p(k)$ :	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

Los únicos datos que necesitas son los conteos de todos los primeros dígitos de la factura:

Primer dígito $k$ :	1	2	3	4	5	6	7	8	9
Recuento:	7	13	12	9	9	13	11	10	16

El Instituto no quiere acusar injustamente a Ana y Beto, por lo que te pide que hagas un TdH con un nivel de significación de 0.001. ¿Qué tan bien se ajustan estos datos a la distribución de Benford? ¿Qué recomendación harías al Instituto?

Queremos hacer el test

$$\begin{cases} H_0 : \text{los datos se ajustan a la Ley de Benford} \\ H_A : \text{no } H_0 \end{cases}$$

Para esto debemos:

1. Calcular  $(Q_L)_{\text{obs}}$
2. Hallar el valor crítico  $c$
3. Rechazar  $H_0$  si  $(Q_L)_{\text{obs}} \geq c$ .

Para calcular  $(Q_L)_{\text{obs}}$  lo mas conveniente es hacer una tabla como la siguiente:

$i$	$n_i$	$p_i$	$n_i \ln(n_i/n)$	$n_i \ln p_i$
1	7	0.301	-18.61	-8.40
2	13	0.176	-26.52	-22.58
3	12	0.125	-25.44	-24.95
4	9	0.097	-21.67	-21.00
5	9	0.079	-21.67	-22.84
6	13	0.067	-26.52	-35.14
7	11	0.058	-24.28	-31.32
8	10	0.051	-23.03	-29.76
9	16	0.046	-29.32	-49.27
Suma	100	1	-217.07	-245.27
$(Q_L)_{\text{obs}}$	56.39			

En nuestro caso  $\dim P = k - 1 = 8$  y  $\dim P_0 = 0$ . Entonces  $Q_L \sim \chi^2(8)$ . Si trabajamos al nivel  $\alpha = 0.001$ , de la tabla vemos que el valor crítico es  $c = 26.12$ . Como

$$(Q_L)_{\text{obs}} = 56.39 \geq c = 26.12$$

rechazamos  $H_0$ . Es posible que Ana y Beto hayan hecho un fraude. ■

### El estadístico de Pearson

En la práctica el estadístico  $Q_L$  suele aproximarse por otro, llamado el estadístico de Pearson  $Q_P$ , que por razones históricas fue el primero en estudiar los tests de bondad de ajuste.

La aproximación es la siguiente. Recordar que el desarrollo de Taylor de la función  $x \ln(x/a)$  está dado por

$$x \ln\left(\frac{x}{a}\right) = (x-a) + \frac{(x-a)^2}{2a} - \frac{(x-a)^3}{6a^2} + \dots$$

Por lo tanto, poniendo  $x = n_i$  y  $a = np_i$ , tenemos que

$$n_i \ln\left(\frac{n_i}{np_i}\right) = (n_i - np_i) + \frac{(n_i - np_i)^2}{2np_i} + \dots$$

Luego

$$\sum_{i=1}^k n_i \ln\left(\frac{n_i}{np_i}\right) = n - n \sum_{i=1}^k p_i + \frac{1}{2} \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} + \dots$$

Notar que los dos primeros términos se cancelan. Poniendo

$$E_i = np_i = \text{frecuencia esperada, y } O_i = n_i = \text{frecuencia observada}$$

tenemos que

$$Q_L \approx Q_P = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

En el ejemplo anterior  $Q_P = 62.7$  y  $Q_L = 56.4$ . La aproximación mejora a medida que  $n$  crece.

### Hipótesis nula compuesta

Una de las aplicaciones más importantes de la aproximación  $\chi^2$  al estadístico  $Q_L$  es el test de bondad de ajuste. Veamos como proceder a través de un ejemplo.

#### Ejemplo 2

El conjunto de datos siguiente muestra el número de impactos de bombas aéreas registradas en cada una de las 576 áreas pequeñas de  $1/4\text{km}^2$  en el sur de Londres durante la Segunda Guerra Mundial.

Número de impactos en un área	0	1	2	3	4	5	$\geq 6$
Frecuencia	229	211	93	35	7	1	0

Las transmisiones de propaganda afirmaban que las bombas alemanas podían ser apuntada con precisión. Sin embargo, si este no fuera el caso, los impactos deberían distribuirse aleatoriamente sobre el área y, por lo tanto, deberían poder ajustarse a una distribución de

Poisson. ¿Es este el caso?

Tratemos de responder a esta pregunta, pero primero miremos el caso general.

Supongamos que tenemos  $k$  grupos con  $n_i$  observaciones en el  $i$ -ésimo grupo. Entonces

Grupo	1	2	3	4	...	$k$
Frecuencia	$n_1$	$n_2$	$n_3$	$n_4$	...	$n_k$

en donde  $\sum_{i=1}^k n_i = n$ .

Supongamos además que disponemos de un modelo probabilístico tal que  $\pi_i(\theta)$ , para  $i = 1, 2, \dots, k$ , es la probabilidad del  $i$ -ésimo grupo. Claramente  $\sum_{i=1}^k \pi_i(\theta) = 1$ .

Queremos hacer el siguiente test

$$\begin{cases} H_0 : \text{Las observaciones se ajustan al modelo} \\ H_A : \text{Las observaciones no se ajustan al modelo} \end{cases}$$

Los parámetros para este test son las probabilidades  $\pi_1, \dots, \pi_k$  de cada grupo, sin la restricción de que sean las probabilidades  $\pi_i(\theta)$  del modelo. Entonces el espacio general de parámetros es

$$P = \left\{ \pi = (\pi_1, \dots, \pi_k) : \sum_{i=1}^k \pi_i = 1 \right\}.$$

Luego,  $\dim P = k - 1$ . La hipótesis nula en este caso tiene dimensión igual a la  $\dim(\theta)$ .

La función de verosimilitud es

$$L(\theta) = n! \prod_{i=1}^k \frac{\pi_i(\theta)^{n_i}}{n_i!}$$

y al tomar logaritmo

$$\ell(\theta) = \sum_{i=1}^k n_i \log \pi_i(\theta) + \log n! - \sum_{i=1}^k \log n_i!$$

Sea  $\hat{\theta}$  el estimador de máxima verosimilitud de  $\theta$ , que maximiza  $\ell(\theta)$  siendo solución de  $\ell'(\theta) = 0$ .

La alternativa general es tomar  $\pi_i$  sin restricciones (no tiene porqué seguir el modelo), salvo la condición de normalización  $\sum_{i=1}^k \pi_i = 1$ . Luego, debemos maximizar

$$\ell(\pi) = \sum_{i=1}^k n_i \log \pi_i + \log n! - \sum_{i=1}^k \log n_i! \quad \text{con} \quad g(\pi) = \sum_{i=1}^k \pi_i = 1$$

en donde hemos denotado  $\pi = (\pi_1, \dots, \pi_k)$ . Usando los multiplicadores de Lagrange  $\gamma$ , obtenemos el sistema de  $k$  ecuaciones

$$\frac{\partial \ell}{\partial \pi_i} - \gamma \frac{\partial g}{\partial \pi_i} = 0, \quad 1 \leq i \leq k$$

Es decir

$$\frac{n_i}{\pi_i} - \gamma = 0, \quad 1 \leq i \leq k$$

o lo que es lo mismo

$$n_i - \gamma \pi_i = 0, \quad 1 \leq i \leq k.$$

Sumando en  $i$  vemos que  $\gamma = n$  y  $\hat{\pi}_i = n_i/n$ .

El estadístico de la razón de verosimilitud es entonces

$$\begin{aligned} Q_L &= 2 \left[ \sum_{i=1}^k n_i \log \frac{n_i}{n} - \sum_{i=1}^k n_i \log \pi_i(\hat{\theta}) \right] \\ &= 2 \sum_{i=1}^k n_i \log \left( \frac{n_i}{n \pi_i(\hat{\theta})} \right). \end{aligned}$$

que podemos aproximar (bajo la hipótesis nula) con la distribución  $\chi_p^2$ , en donde  $p = k - \dim \theta$ .

Volvamos a nuestro caso concreto de las bombas en Londres. El primer paso es calcular el estimador de máxima verosimilitud  $\hat{\theta}$ . Recordar que  $\hat{\theta}$  es solución de

$$\ell'(\theta) = \sum_{i=1}^k n_i \frac{\pi_i'(\theta)}{\pi_i(\theta)} = 0.$$

En nuestro caso la distribución de Poisson es

$$\pi_i(\theta) = \frac{\theta^i e^{-\theta}}{i!},$$

por lo que

$$\frac{\pi_i'(\theta)}{\pi_i(\theta)} = \frac{i - \theta}{\theta}.$$

De aquí resulta que  $\hat{\theta}$  debe cumplir

$$0 = \sum_{i=1}^k n_i (i - \hat{\theta}) \Rightarrow \hat{\theta} = \frac{1}{n} \sum_{i=1}^k i \cdot n_i.$$

De los datos de la tabla

$$\hat{\theta} = \frac{535}{576} = 0.928.$$

Usando la f.p.p. de la distribución de Poisson con  $\theta = 0.929$  obtenemos

$i$	0	1	2	3	4	5	$\geq 6$
$\pi_i(\hat{\theta})$	0.3949	0.3669	0.1704	0.0528	0.0123	0.0023	0.0004

El valor observado de  $Q_L$  es entonces

$$(Q_L)_{\text{obs}} = 1.4995.$$

Para determinar la región de rechazo debemos elegir  $c$  para que

$$\mathbf{P}(Q_L \geq c | H_0) = \alpha.$$



Si aproximamos la distribución de  $Q_L$  por la distribución  $\chi_p^2$  con

$$p = k - 1 - \dim \theta = 7 - 1 - 1 = 5.$$

Es decir,  $c = \chi_5^2(\alpha)$ . Para  $\alpha = 0.05$ ,  $\chi_5^2(0.05) = 11.0705$ . Así que no rechazamos  $H_0$ . Es decir, no hay evidencia para afirmar que los impactos de las bombas corresponden a un arma con gran puntería.

También podemos calcular el p-valor. Para esto debemos calcular

$$\mathbf{P}(Q_L \geq (Q_L)_{\text{obs}} | H_0) = \mathbf{P}(\chi_5^2 \geq 1.4995) = 0.913.$$

Claramente no hay la menor evidencia para rechazar  $H_0$ . ■

El test realizado en el ejemplo no es, estrictamente hablando, correcto. La razón de esto es que la distribución  $\chi^2$  que hemos usado para calcular el p-valor es una aproximación, y la calidad de esa aproximación depende del tamaño de la muestra. Afortunadamente hay una regla general que se puede usar en estos casos.

### Regla general para la aproximación $\chi^2$

Se puede aproximar por la distribución  $\chi^2$  siempre que el valor esperado de cada celda en la tabla sea al menos 5. Si el valor esperado de una celda es menor que 5, debe combinarse con otra(s) celda(s) adyacente(s) para obtener un valor adecuado.

### Ejemplo 3

Revisemos entonces los cálculos del ejemplo anterior. Incluyendo los valores esperados en la tabla para los impactos de bombas aéreas, obtenemos la tabla a continuación:

Impactos en el área	0	1	2	3	4	5	$\geq 6$
Frecuencia	229	211	93	35	7	1	0
Frecuencia esperada	227.462	211.334	98.150	30.413	7.027	1.325	0.230

Después de agrupar las últimas tres celdas, obtenemos

Impactos en el área	0	1	2	3	$\geq 4$
Frecuencia	229	211	93	35	8
Frecuencia esperada	227.462	211.334	98.150	30.413	8.582

Calculamos de nuevo el estadístico para obtener  $(Q_L)_{\text{obs}} = 1.101$ , que usando la distribución  $\chi_3^2$  da un p-valor de 0.777. Nuevamente no encontramos evidencia de rechazo de la hipótesis nula. Por lo tanto, no hemos encontrado evidencia de que la bomba pueda ser apuntada con algún grado de precisión. ■

## 5. Test de independencia para tablas de contingencia

Este tipo de tablas en general muestra datos que se obtienen por clasificación cruzada de un número fijo de individuos según dos criterios. Por lo tanto, se muestran como  $n_{ij}$  en una

tabla con  $r$  filas y  $c$  columnas de la siguiente manera:

$$\begin{array}{ccc|c} n_{11} & \cdots & n_{1c} & n_{1.} \\ \vdots & \ddots & \vdots & \vdots \\ n_{r1} & \cdots & n_{rc} & n_{r.} \\ \hline n_{.1} & \cdots & n_{.c} & n \end{array}$$

El objetivo es investigar la independencia de las dos clasificaciones.

### Ejemplo 1

Veamos un ejemplo con un conjunto de datos famoso e histórico. Estos son los datos de Pearson de 1909 sobre el crimen y la bebida. Los datos son los siguientes:

Crimen	Tomador	No tomador
Incendio	50	43
Violación	88	62
Violencia	155	110
Robo	379	300
Acuñaación	18	14
Fraude	63	144

¿Está relacionada la bebida con el crimen?

Supongamos que el  $k$ -ésimo individuo va a parar en la celda  $(X_k, Y_k)$ , para  $k = 1, \dots, n$ , y que los individuos son independientes. Sea

$$\mathbf{P}((X_k, Y_k) = (i, j)) = \theta_{ij}, \quad i = 1, 2, \dots, r; \quad j = 1, 2, \dots, c$$

en donde  $\sum_{i,j} \theta_{ij} = 1$ . Se puede escribir la hipótesis nula de independencia de los clasificadores como

$$H_0 : \theta_{ij} = \phi_i \rho_j, \quad \forall i, j.$$

Dejaremos los detalles como ejercicio, pero veamos como es el procedimiento general.

La función de verosimilitud es

$$L(\theta) = n! \prod_{i,j} \frac{\theta_{ij}^{n_{ij}}}{n_{ij}!}$$

por lo que al tomar logaritmos obtenemos

$$\ell(\theta) = \sum_{i,j} n_{ij} \log \theta_{ij} + \log n! - \sum_{i,j} \log n_{ij}!$$

Bajo  $H_0$  tenemos que  $\theta_{ij} = \phi_i \rho_j$ , por lo que al maximizar en todo  $\phi_i$  y  $\rho_j$  que cumplen

$$\sum_i \phi_i = 1 \quad \text{y} \quad \sum_j \rho_j = 1$$

obtenemos

$$\hat{\phi}_i = \frac{n_{i.}}{n}, \quad \hat{\rho}_j = \frac{n_{.j}}{n}$$

Bajo  $H_A$ , debemos maximizar respecto de  $\theta_{ij}$  bajo la restricción

$$\sum_{ij} \theta_{ij} = 1.$$

El resultado es

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

Finalmente, el estadístico de la razón de verosimilitud es

$$Q_L = 2 \sum_{i=1}^r \sum_{j=1}^c n_{ij} \log \left( \frac{n_{ij}n}{n_i n_j} \right)$$

Para nuestros datos,  $Q_L = 50.52$ . Bajo  $H_0$  el estadístico  $Q_L$  tiene distribución  $\chi_p^2$  en donde

$$\begin{aligned} p &= rc - 1 - (r - 1) - (c - 1) \\ &= rc - r - c + 1 \\ &= (r - 1)(c - 1). \end{aligned}$$

Es decir,  $p = 5$ . Usando la distribución  $\chi^2$  con 5 grados de libertad, el cuantil 0.9999 es 25.75, y podemos rechazar al nivel de significación 0.0001. Hay evidencia abrumadora de que el crimen y la bebida están relacionados. Al menos en Londres, y en 1909. ■

### El estadístico de Pearson

Al igual que para el test de bondad de ajuste, es más común en la práctica utilizar el estadístico de Pearson  $Q_P$ . Con el mismo razonamiento que usamos antes, se puede probar que

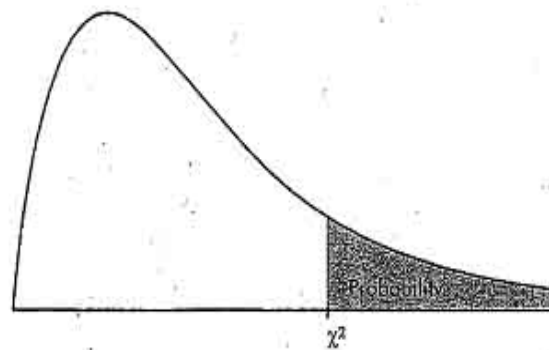
$$Q_L \approx Q_P = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i,j} \frac{\left(n_{ij} - \frac{n_i n_j}{n}\right)^2}{\frac{n_i n_j}{n}}$$

en donde

$$E_{ij} = \frac{n_i n_j}{n} \text{ y } O_{ij} = n_{ij}$$

son la frecuencia esperada y observada respectivamente.

$\chi^2$  CRITICAL VALUES



**TABLE C:  $\chi^2$  CRITICAL VALUES**

df	Tail probability <i>p</i>										
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4