Homework 1 *Categorical Data Analysis, S. S. Mukherjee, Fall 2019*

Topics: Goodness-of-fit tests, tests of independence Due on August 15, 2019

Name of student: SUBHRAJYOTY ROY
Roll number: MB1911

1. **The Cressie-Read power divergence and multinomial goodness-of-fit tests** [15 points]

The Cressie-Read power divergence between two pmf's $p = (p_i)_{i\in S}$ and $q = (q_i)_{i\in S}$ on the same support S is defined as

$$I_\lambda(p||q) = \frac{1}{\lambda(\lambda+1)} \sum_{i\in S} p_i \left[\left(\frac{p_i}{q_i}\right)^\lambda - 1\right],$$

where $\lambda \in \mathbb{R} \setminus \{0, -1\}$.

(a) Show that

$$2I_1(p, q) = \sum_{i\in S} \frac{(p_i - q_i)^2}{q_i} = \sum_{i\in S} \left(\frac{p_i}{q_i} - 1\right)^2 q_i = D_{\chi^2}(p||q),$$

Pearson's χ^2-divergence between p and q.

(b) Show that

$$\lim_{\lambda\to 0} I_\lambda(p||q) = \sum_{i\in S} p_i \log\left(\frac{p_i}{q_i}\right) = D_{\mathrm{KL}}(p||q),$$

the Kullback-Liebler divergence between p and q.

(c) What is $\lim_{\lambda\to -1} I_\lambda(p||q)$?

Define I_λ at $\lambda = 0, -1$ by these limiting values. Now suppose $(n_1, \ldots, n_k) \sim \text{Multinomial}(n; p = (p_i)_{i=1}^k)$. Let $\hat{p}_i = \frac{n_i}{n}$ be the sample proportions.

(d) Show that $2nI_1(\hat{p}||p_0)$ is nothing but Pearson's χ^2 statistic for testing $H_0 : p = p_0$.

(e) Show that $2nI_0(\hat{p}||p_0)$ is nothing but the likelihood ratio (LR) statistic for testing $H_0 : p = p_0$.

(f) Show, by establishing a CLT for $(n_i)_{i=1}^{k-1}$, that $2nI_1(\hat{p}||p_0) \xrightarrow[H_0]{d} \chi^2_{k-1}$.

(g) Let $X_i = \frac{\hat{p}_i}{p_i} - 1$. Express $2nI_\lambda(\hat{p}||p_0)$ in terms of X_i and then show, by relating $2nI_\lambda(\hat{p}||p_0)$ to $2nI_1(\hat{p}||p_0)$ via a Taylor expansion, that $2nI_\lambda(\hat{p}||p_0) \xrightarrow[H_0]{d} \chi^2_{k-1}$.

Solution.

(a) In the definition of $I_\lambda(p||q)$, putting $\lambda = 1$, we obtain the following;

$$\begin{aligned}
I_1(p||q) &= \frac{1}{2}\sum_{i\in S} p_i \left[\left(\frac{p_i}{q_i}\right) - 1\right] \\
&= \frac{1}{2}\sum_{i\in S} q_i \times \frac{p_i}{q_i}\left[\left(\frac{p_i}{q_i}\right) - 1\right] \\
&= \frac{1}{2}\sum_{i\in S} q_i \times \left[\left[\left(\frac{p_i}{q_i}\right) - 1\right]^2 + \left[\left(\frac{p_i}{q_i}\right) - 1\right]\right] \\
&= \frac{1}{2}\sum_{i\in S} q_i \left(\frac{p_i}{q_i} - 1\right)^2 + \frac{1}{2}\sum_{i\in S}(p_i - q_i) \\
&= \frac{1}{2}\sum_{i\in S} q_i \left(\frac{p_i}{q_i} - 1\right)^2
\end{aligned}$$

where the last line follows from the fact that $(p_i)_{i\in S}$ and $(q_i)_{i\in S}$ both being p.m.f. on the same support S, $\sum_{i\in S} p_i = \sum_{i\in S} q_i = 1$. Finally, note that, we can rewrite it as;

$$\begin{aligned}
2I_1\left(p||q\right) &= \sum_{i\in S} q_i \left(\frac{p_i}{q_i} - 1\right)^2 \\
&= \sum_{i\in S} q_i \left(\frac{p_i - q_i}{q_i}\right)^2 \\
&= \sum_{i\in S} \frac{(p_i - q_i)^2}{q_i} \\
&= D_{\chi^2}\left(p||q\right)
\end{aligned}$$

(b) Consider the following standard limit result from analysis.

$$\lim_{x\to 0} \frac{a^x - 1}{x} = \ln a \qquad \forall a > 0$$

Now, we consider the limit in question;

$$\begin{aligned}
\lim_{\lambda\to 0} I_\lambda\left(p||q\right) &= \lim_{\lambda\to 0} \frac{1}{\lambda(\lambda+1)} \sum_{i\in S} p_i \left[\left(\frac{p_i}{q_i}\right)^\lambda - 1\right] \\
&= \lim_{\lambda\to 0} \frac{1}{\lambda+1} \times \lim_{\lambda\to 0} \sum_{i\in S} \frac{p_i}{\lambda} \left[\left(\frac{p_i}{q_i}\right)^\lambda - 1\right] \\
&= 1 \times \sum_{i\in S} \left[\lim_{\lambda\to 0} \frac{p_i}{\lambda} \left[\left(\frac{p_i}{q_i}\right)^\lambda - 1\right]\right], \\
&\qquad \text{provided we can interchange the sum and the limit} \\
&= \sum_{i\in S} p_i \log\left(\frac{p_i}{q_i}\right) \\
&= D_{\mathrm{KL}}\left(p||q\right)
\end{aligned}$$

If we assume that the support S is a finite set, then the interchangeability of the finite sum over elements of S and the limit $\lambda \to 0$ is justified.

(c) We apply the same standard limit results from analysis.

$$
\begin{aligned}
\lim_{\lambda \to -1} I_\lambda (p||q) &= \lim_{\lambda \to -1} \frac{1}{\lambda(\lambda+1)} \sum_{i\in S} p_i \left[\left(\frac{p_i}{q_i}\right)^\lambda - 1 \right] \\
&= \lim_{\lambda \to -1} \frac{1}{\lambda} \times \lim_{\lambda \to -1} \sum_{i\in S} p_i \frac{1}{(\lambda+1)} \left[\left(\frac{p_i}{q_i}\right)^\lambda - 1 \right] \\
&= (-1) \times \lim_{\lambda \to -1} \sum_{i\in S} q_i \frac{p_i}{q_i} \frac{1}{(\lambda+1)} \left[\left(\frac{p_i}{q_i}\right)^\lambda - 1 \right] \\
&= (-1) \times \lim_{\lambda \to -1} \sum_{i\in S} q_i \frac{1}{(\lambda+1)} \left[\left(\frac{p_i}{q_i}\right)^{\lambda+1} - \frac{p_i}{q_i} \right] \\
&= (-1) \times \lim_{\lambda \to -1} \sum_{i\in S} q_i \frac{1}{(\lambda+1)} \left[\left[\left(\frac{p_i}{q_i}\right)^{\lambda+1} - 1 \right] - \left[\frac{p_i}{q_i} - 1 \right] \right] \\
&= (-1) \times \lim_{\lambda \to -1} \sum_{i\in S} q_i \frac{1}{(\lambda+1)} \left[\left(\frac{p_i}{q_i}\right)^{\lambda+1} - 1 \right] + \lim_{\lambda \to -1} \frac{1}{(\lambda+1)} \sum_{i\in S} (p_i - q_i) \\
&= (-1) \times \sum_{i\in S} \left[\lim_{\lambda \to -1} \frac{q_i}{(\lambda+1)} \left[\left(\frac{p_i}{q_i}\right)^{\lambda+1} - 1 \right] \right], \\
&\quad \text{provided we can interchange the sum and the limit} \\
&= (-1) \times \sum_{i\in S} q_i \log\left(\frac{p_i}{q_i}\right) \\
&= \sum_{i\in S} q_i \log\left(\frac{q_i}{p_i}\right)
\end{aligned}
$$

If we assume that the support S is a finite set, then the interchangeability of the finite sum over elements of S and the limit $\lambda \to (-1)$ is justified.

(d) Consider the setup $(n_1, n_2, \dots n_k) \sim$ Multinomial $\left(n; p = (p_i)_{i=1}^k\right)$, and we want to test the hypothesis $H_0 : p = p_0$, where we assume $p_0 = (p_{0,i})_{i=1}^k$ is known.
The expected frequency of i-th cell is given by; $np_{0,i}$ and the observed frequency of i-th cell is given by n_i. Therefore, to test H_0, Pearson's χ^2 statistic is given as;

$$
\begin{aligned}
\chi^2_{obs} &= \sum_{i=1}^k \frac{(n_i - np_{0,i})^2}{np_{0,i}} \\
&= \sum_{i=1}^k \frac{\left(n\left(\frac{n_i}{n} - p_{0,i}\right)\right)^2}{np_{0,i}} \\
&= \sum_{i=1}^k n \frac{\left(\frac{n_i}{n} - p_{0,i}\right)^2}{p_{0,i}} \\
&= 2nI_1 (\hat{p}||p_0)
\end{aligned}
$$

where the last line follows from part (a).

(e) Consider the multinomial setup described in part (d). To test the same hypothesis $H_0 : p = p_0$, the Likelihood Ratio statistic would be as follows;

$$
\text{LR} = \frac{\prod_{i=1}^k p_{0,i}^{n_i}}{\max_{p=(p_1,p_2,\dots p_k)} \prod_{i=1}^k p_i^{n_i}}
$$

Note that, the maximum of the denominator happens when $p_i = \hat{p}_i$, the sample proportion (as this is the m.l.e. of p_i under multinomial setup). Therefore, the likelihood ratio reduces to;

$$\begin{aligned}
\text{LR} &= \frac{\prod_{i=1}^k p_{0,i}^{n_i}}{\prod_{i=1}^k \hat{p}_i^{n_i}} \\
-2\log \text{LR} &= -2\log\left[\prod_{i=1}^k \left(\frac{p_{0,i}}{\hat{p}_i}\right)^{n_i}\right] \\
&= -2\sum_{i=1}^k n_i \log\left(\frac{p_{0,i}}{\hat{p}_i}\right) \\
&= 2\sum_{i=1}^k n\hat{p}_i \log\left(\frac{\hat{p}_i}{p_{0,i}}\right) \\
&= 2nI_0\left(\hat{p}||p_0\right)
\end{aligned}$$

where the last line follows from part (b).

(f) Note that, under H_0, the cell frequencies $n_i \sim \text{Multinomial}(n; p_0)$ for all $i = 1, 2, \dots k$. Consider the i.i.d. random variables X_{ij}, where $i = 1, 2, \dots n$ and $j = 1, 2, \dots k$, each of which follow $\text{Multinomial}(1, p_0)$, which represents whether i-th multinomial trial results in an observation corresponding to j-th cell. In that case, $n_j = \sum_{i=1}^n X_{ij}$ for all $j = 1, 2, \dots k$. Note that, under H_0;

$$\begin{aligned}
E\left[X_{ij}\right] &= p_{0,j} \\
Var\left[X_{ij}\right] &= p_{0,j}\left(1 - p_{0,j}\right) \\
Cov\left[X_{ij,il}\right] &= -p_{0,j}p_{0,l}
\end{aligned}$$

Consider the vector

$$\mathbf{X}_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{i(k-1)} \end{pmatrix}$$

Note that, this random vector has mean given by;

$$\mathbf{p}_0 = \begin{pmatrix} p_{0,1} \\ p_{0,2} \\ \vdots \\ p_{0,(k-1)} \end{pmatrix}$$

and variance-covariance matrix given by;

$$\begin{aligned}
\Sigma &= \begin{pmatrix} p_{0,1}\left(1 - p_{0,1}\right) & -p_{0,1}p_{0,2} & \cdots & -p_{0,1}p_{0,(k-1)} \\ -p_{0,2}p_{0,1} & p_{0,2}\left(1 - p_{0,2}\right) & \cdots & -p_{0,2}p_{0,(k-1)} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{0,(k-1)}p_{0,1} & -p_{0,1}p_{0,(k-1)} & \cdots & p_{0,(k-1)}\left(1 - p_{0,(k-1)}\right) \end{pmatrix} \\
&= \text{diag}\left(p_{0,1}, p_{0,2}, \dots p_{0,(k-1)}\right) - \mathbf{p}_0\mathbf{p}_0^\intercal
\end{aligned}$$

where diag$(\cdot)$ denotes a diagonal matrix of its arguments. Denote the above diagonal matrix by D. Now, since $\Sigma = D - \mathbf{p}_0\mathbf{p}_0^\intercal$, therefore, its inverse is given by;

$$\Sigma^{-1} = D^{-1} + \frac{\left(D^{-1}\mathbf{p}_0\mathbf{p}_0^\intercal D^{-1}\right)}{1 - \mathbf{p}_0^\intercal D^{-1}\mathbf{p}_0}$$

Note that, $1-\mathbf{p}_0^\intercal D^{-1}\mathbf{p}_0 = \sum_{j=1}^k p_{0,j} - \sum_{j=1}^{k-1} p_{0,j} = p_{0,k}$. Also, $D^{-1}\mathbf{p}_0 = (1,1,\dots 1)^\intercal$. Hence, $(D^{-1}\mathbf{p}_0\mathbf{p}_0^\intercal D^{-1})$ is a matrix of order $(k-1)\times(k-1)$ with all elements equal to 1. Therefore,

$$\Sigma^{-1} = D^{-1} + \frac{1}{p_{0,k}}\mathbf{1}\mathbf{1}^\intercal$$

Note that, each of the $\mathbf{X}_i$'s are independent and identically distributed random variables, hence, by applying Central Limit Theorem, we obtain;

$$\sqrt{n}\Sigma^{-1/2}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i - \mathbf{p}_0\right) \xrightarrow[H_0]{d} N_{k-1}\left(0, I_{(k-1)}\right)$$

Now;

$$\begin{aligned} 2nI_1\left(\hat{p}||p_0\right) &= n\sum_{j=1}^k \frac{(\hat{p}_i - p_{0,i})^2}{p_{0,i}} \\ &= n\left[\sum_{j=1}^{k-1}\frac{(\hat{p}_i - p_{0,i})^2}{p_{0,i}} + \frac{(\hat{p}_k - p_{0,k})^2}{p_{0,k}}\right] \\ &= n\left[\sum_{j=1}^{k-1}\frac{(\hat{p}_i - p_{0,i})^2}{p_{0,i}} + \frac{\left(\sum_{j=1}^{k-1}(\hat{p}_j - p_{0,j})\right)^2}{p_{0,k}}\right] \end{aligned}$$

where the last line follows from the fact that $\sum_{j=1}^k(\hat{p}_j - p_{0,j}) = 0$, as both of them constitutes a p.m.f. As $\hat{p}_j = \frac{n_j}{n} = \sum_{i=1}^n \frac{X_{ij}}{n}$, it follows that;

$$\begin{aligned} 2nI_1\left(\hat{p}||p_0\right) &= n\times\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i - \mathbf{p}_0\right)^\intercal \Sigma^{-1}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i - \mathbf{p}_0\right) \\ &= \left[\sqrt{n}\Sigma^{-1/2}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i - \mathbf{p}_0\right)\right]^\intercal\left[\sqrt{n}\Sigma^{-1/2}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i - \mathbf{p}_0\right)\right] \\ &\xrightarrow[H_0]{d} \chi^2_{k-1}, \text{ as desired.} \end{aligned}$$

(g) Let, $X_i = \frac{\hat{p}_i}{p_{0,i}} - 1$. Consider the form of power divergence;

$$\begin{aligned} 2nI_\lambda\left(\hat{p}||p_0\right) &= \frac{2n}{\lambda(\lambda+1)}\sum_{i=1}^k \hat{p}_i\left[\left(\frac{\hat{p}_i}{p_{0,i}}\right)^\lambda - 1\right] \\ &= \frac{2n}{\lambda(\lambda+1)}\sum_{i=1}^k p_{0,i}\left[\left(\frac{\hat{p}_i}{p_{0,i}}\right)^{\lambda+1} - 1 - \left(\frac{\hat{p}_i}{p_{0,i}}\right) + 1\right] \\ &= \frac{2n}{\lambda(\lambda+1)}\sum_{i=1}^k p_{0,i}\left[\left(\frac{\hat{p}_i}{p_{0,i}}\right)^{\lambda+1} - 1\right] - \frac{2n}{\lambda(\lambda+1)}\sum_{i=1}^k(\hat{p}_i - p_{0,i}) \\ &= \frac{2n}{\lambda(\lambda+1)}\sum_{i=1}^k p_{0,i}\left[\left(\frac{\hat{p}_i}{p_{0,i}}\right)^{\lambda+1} - 1\right], \text{since the additional part is 0} \\ &= \frac{2n}{\lambda(\lambda+1)}\sum_{i=1}^k p_{0,i}\left[(X_i+1)^{\lambda+1} - 1\right] \end{aligned}$$

Note that, under H_0, each $X_i = 0$. Applying Taylor expansion on $g(x) = \left((x+1)^{\lambda+1} - 1\right)$ about $x = 0$, we obtain;

$$
\begin{aligned}
g(x) &= g(0) + xg'(0) + \frac{x^2}{2}g''(0) + O_p(x^3) \\
&= 0 + (\lambda+1)x + \frac{\lambda(\lambda+1)}{2}x^2 + O_p(x^3)
\end{aligned}
$$

Therefore, we have;

$$
\begin{aligned}
2nI_\lambda\left(\hat{p}||p_0\right) &= \frac{2n}{\lambda(\lambda+1)}\sum_{i=1}^{k} p_{0,i}\left[(X_i+1)^{\lambda+1} - 1\right] \\
&= \frac{2n}{\lambda(\lambda+1)}\sum_{i=1}^{k} p_{0,i}\left[(\lambda+1)X_i + \frac{\lambda(\lambda+1)}{2}X_i^2 + O_p(X_i^3)\right], \text{ by using Taylor expansion} \\
&= n\sum_{i=1}^{k} p_{0,i}\left[X_i^2 + O_p(X_i^3)\right], \text{ since } \sum_{i=1}^{k} p_{0,i}X_i = 0 \\
&= n\sum_{i=1}^{k} p_{0,i}\left(\frac{\hat{p}_i}{p_{0,i}} - 1\right)^2 + \sum_{i=1}^{k}(np_{0,i})O_p(X_i^3) \\
&= 2nI_1\left(\hat{p}||p_0\right) + \sum_{i=1}^{k}(np_{0,i})O_p(X_i^3), \text{ using part (a)} \\
&= 2nI_1\left(\hat{p}||p_0\right) + \sum_{i=1}^{k}(np_{0,i})O_p(\frac{1}{n\sqrt{n}})
\end{aligned}
$$

where the last line follows from the fact that, $X_i = \left(\frac{\hat{p}_i}{p_{0,i}} - 1\right) = \frac{(n_i - np_{0,i})}{np_{0,i}}$, whereas the quantity, $\frac{(n_i - np_{0,i})}{\sqrt{np_{0,i}}}$ has an asymptotic normal distribution under H_0, therefore, bounded in probability. Hence, $O_p(X_i) = O_p(\frac{1}{\sqrt{np_{0,i}}})$. Therefore, under H_0;

$$2nI_\lambda\left(\hat{p}||p_0\right) \xrightarrow{P} 2nI_1\left(\hat{p}||p_0\right)$$

Hence, using the result of part (f) and Slutsky's theorem, we get;

$$2nI_\lambda\left(\hat{p}||p_0\right) \xrightarrow[H_0]{d} \chi^2_{k-1}, \text{ as desired}$$

2. **Testing conditional independence** [5]

In a 3-dimensional multinomial table, obtain MLE's of the p_{ijk}'s under the null hypothesis

$$H_0 : \text{row} \perp\!\!\!\perp \text{column} \mid \text{layer}.$$

What are the limiting distributions of the likelihood ratio and the χ^2 tests for this hypothesis? Test if 'Sex' is independent of 'Support_Abortion' given 'Status' in the `Abortion` data in the **R** package `vcdExtra`.

Solution.

Note that, under $H_0 : \text{row} \perp\!\!\!\perp \text{column} \mid \text{layer}$, we have;

$$p_{ijk} = p_{\cdot\cdot k}p_{ij|k} = p_{\cdot\cdot k}p_{i|k}p_{j|k}$$

Now, under H_0, we obtain the likelihood as follows;

$$
\begin{aligned}
&\mathcal{L}(p|n_{ijk}) \propto \prod_{i,j,k} (p_{ijk})^{n_{ijk}} \\
\Rightarrow\ &\mathcal{L}(p|n_{ijk}) \propto \prod_{i,j,k} \left(p_{\cdot\cdot k} p_{i|k} p_{j|k}\right)^{n_{ijk}} \\
\Rightarrow\ &\ell(p|n_{ijk}) = \text{constant} + \sum_{i,j,k} n_{ijk} \left(\log p_{\cdot\cdot k} + \log p_{i|k} + \log p_{j|k}\right), \text{ taking logarithm to both sides} \\
\Rightarrow\ &\ell(p|n_{ijk}) = \text{constant} + \sum_{k} n_{\cdot\cdot k} \log p_{\cdot\cdot k} + \sum_{i,k} n_{i\cdot k} \log p_{i|k} + \sum_{j,k} n_{\cdot j,k} \log p_{j|k}
\end{aligned}
$$

To maximize this likelihood with respect to $p_{\cdot\cdot k}$, $p_{i|k}$ and $p_{j|k}$, we differentiate the log likelihood with respect to those variables and set them equal to 0. Differentiating with respect to $p_{\cdot\cdot k}$, we obtain;

$$
\begin{aligned}
&\frac{\partial \ell}{\partial p_{\cdot\cdot k}} = \frac{n_{\cdot\cdot k}}{p_{\cdot\cdot k}} - \frac{n - \sum_{k'=1}^{L-1} n_{\cdot\cdot k'}}{1 - \sum_{k=1}^{L-1} p_{\cdot\cdot k'}}, \text{ where L is the number of layers} \\
\Rightarrow\ &\frac{\partial \ell}{\partial p_{\cdot\cdot k}} = \frac{n_{\cdot\cdot k}}{p_{\cdot\cdot k}} - \frac{n_{\cdot\cdot L}}{p_{\cdot\cdot L}}
\end{aligned}
$$

Setting the above equal to 0 means, we have $\frac{n_{\cdot\cdot k}}{p_{\cdot\cdot k}} = \lambda$, a constant for any $k = 1, 2, \ldots L$. Clearly, we would have $\lambda = \frac{n_{\cdot\cdot 1} + n_{\cdot\cdot 2} + \ldots n_{\cdot\cdot L}}{p_{\cdot\cdot 1} + p_{\cdot\cdot 2} + \ldots p_{\cdot\cdot L}} = \frac{n}{1}$. Therefore, we have the m.l.e. $\hat{p}_{\cdot\cdot k} = \frac{n_{\cdot\cdot k}}{n}$.
In a similar way, we would have the other m.l.e. as;

$$
\begin{aligned}
\hat{p}_{i|k} &= \frac{n_{i\cdot k}}{\sum_{i,k} n_{i\cdot k}} = \frac{n_{i\cdot k}}{n_{\cdot\cdot k}} \\
\hat{p}_{j|k} &= \frac{n_{\cdot jk}}{\sum_{j,k} n_{\cdot jk}} = \frac{n_{\cdot jk}}{n_{\cdot\cdot k}}
\end{aligned}
$$

Now, we consider **Likelihood Ratio** for testing H_0.

$$
\lambda = \frac{\sup_{H_0} \prod_{i,j,k} (p_{ijk})^{n_{ijk}}}{\sup \prod_{i,j,k} (p_{ijk})^{n_{ijk}}} = \frac{\sup \prod_{i,j,k} \left(p_{\cdot\cdot k} p_{i|k} p_{j|k}\right)^{n_{ijk}}}{\prod_{i,j,k} (\hat{p}_{ijk})^{n_{ijk}}} = \prod_{i,j,k} \left(\frac{\hat{p}_{\cdot\cdot k} \hat{p}_{i|k} \hat{p}_{j|k}}{\hat{p}_{ijk}}\right)^{n_{ijk}}
$$

Observe that,

$$
\hat{p}_{\cdot\cdot k} \hat{p}_{i|k} \hat{p}_{j|k} = \frac{n_{\cdot\cdot k}}{n} \frac{n_{i\cdot k}}{n_{\cdot\cdot k}} \frac{n_{\cdot jk}}{n_{\cdot\cdot k}} = \frac{n_{i\cdot k} n_{\cdot jk}}{n_{\cdot\cdot k} n}
$$

Hence,

$$
\frac{\hat{p}_{\cdot\cdot k} \hat{p}_{i|k} \hat{p}_{j|k}}{\hat{p}_{ijk}} = \frac{n_{i\cdot k} n_{\cdot jk}}{n_{\cdot\cdot k} n} \times \frac{n}{n_{ijk}} = \frac{n_{i\cdot k} n_{\cdot jk}}{n_{\cdot\cdot k} n_{ijk}}
$$

Therefore,

$$
\begin{aligned}
&\log \lambda = \sum_{i,j,k} n_{ijk} \log \left(\frac{n_{i\cdot k} n_{\cdot jk}}{n_{\cdot\cdot k} n_{ijk}}\right) \\
\Rightarrow\ &-2 \log \lambda = \sum_{i,j,k} n_{ijk} \log n_{ijk} + \sum_{k} n_{\cdot\cdot k} \log n_{\cdot\cdot k} - \sum_{i,k} n_{i\cdot k} \log n_{i\cdot k} - \sum_{j,k} n_{\cdot jk} \log n_{\cdot jk}
\end{aligned}
$$

Applying Wilk's theorem, we get that, the above quantity $-2\log\lambda$ asymptotically follows a χ^2 distribution with degrees of freedom given by; df = number of free parameters under full model$-$number of free parameters $(RCL-1)-((R-1)L+(C-1)L+(L-1))$, where R, C, L are the number of rows, columns and layers in the contingency table. Note that, under null hypothesis, the free parameters are $p_{i|k}, p_{j|k}$ and $p_{\cdot\cdot k}$, which respectively are $(R-1)L, (C-1)L$ and $(L-1)$ in numbers. Therefore, finally, we have;

$$df = (RCL-1)-(RL-L+CL-L+L-1) = (RCL-RL-CL+L) = L(R-1)(C-1)$$

Therefore, the limiting distribution of likelihood ratio statistic under H_0 would be a central χ^2 distribution with $(R-1)(C-1)L$ as degrees of freedom.

Now, considering **Pearsonian chi-sqaured test statistic**, we would have expected frequency of (i,j,k)-th cell entry under H_0 as;

$$\hat{n}_{ijk} = n\hat{p}_{ijk} = n\hat{p}_{\cdot\cdot k}\hat{p}_{i|k}\hat{p}_{j|k} = \frac{n_{i\cdot k}n_{\cdot jk}}{n_{\cdot\cdot k}}$$

Hence, Pearson's chi-sqaured test statistic would be;

$$\chi^2_{observed} = \sum_{ijk}\frac{\left(n_{ijk}-\dfrac{n_{i\cdot k}n_{\cdot jk}}{n_{\cdot\cdot k}}\right)^2}{\dfrac{n_{i\cdot k}n_{\cdot jk}}{n_{\cdot\cdot k}}} = \sum_{ijk}\frac{(n_{ijk}n_{\cdot\cdot k}-n_{i\cdot k}n_{\cdot jk})^2}{n_{\cdot\cdot k}n_{i\cdot k}n_{\cdot jk}}$$

which also asymptotically follows a central χ^2 distribution with degrees of freedom equal to; $df = (RCL-1)-$ number of free parameters which are estimated $= (RCL-1)-((R-1)L+(C-1)L+(L-1)) = (R-1)(C-1)L$ as before.

The final part of the exercise asks to write **R** code which tests the hypothesis H_0 on **Abortion** data from **vcdExtra** package. Hence, firstly we load the required package and then load the data.

```
library(vcdExtra)
data("Abortion")
ftable(Abortion)
```

```
, , Support_Abortion = Yes

Status
Sex       Lo  Hi
Female 171 138
Male   152 167

, , Support_Abortion = No

Status
Sex       Lo  Hi
Female  79 112
Male   148 133
```

Note that, current data is in a format of 2x2 table with *Sex* and *Status* as row and column, while *Support_Abortion* as layer. However, we need to restructure the data in a way so that *Sex* and *Support_Abortion* to be row and column variable respectively, while *Status* is layer variable.

```r
Abortion2 =aperm(Abortion, c(1,3,2))
dimnames(Abortion2)
```

```
$Sex
[1] "Female" "Male"

$Support_Abortion
[1] "Yes" "No"

$Status
[1] "Lo" "Hi"
```

Now, to test the null hypothesis H_0 : row $\perp\!\!\!\perp$ column | layer we use **Cochran-Mantel-Haenszel Chi-Squared Test**;

```r
mantelhaen.test(Abortion2)
```

```
Mantel-Haenszel chi-squared test with continuity correction

data:  Abortion2
Mantel-Haenszel X-squared = 7.9435, df = 1, p-value = 0.004826
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
1.117674 1.808322
sample estimates:
common odds ratio
1.421659
```

Note that, we get an p-value of 0.004826, which is extremely lower than the significance level of $\alpha = 0.05$. Hence, we reject the null hypothesis that the variables *Sex* and *Support_Abortion* are independent given the layered variable *Status* in the light of *Abortion* data.

Thank You!