g. Another justification for $\chi^2$

i. Each of the summands $X_j$ is approximately $\mathcal{N}(m\pi_j, m\pi_j(1 - \pi_j))$

ii. Then $(X_j - m\pi_j)^2/[m\pi_j(1 - \pi_j))] \sim \chi_1^2$ approximately

iii. Sum should be $\sim \chi_k^2$

- Except that summands are dependent
- Compensate by Dropping $1 - \pi_j$ from denominator and adjusting d.f.

F: 13.6

13. Tests of Equality of Proportions

a. $X_j \sim \text{Bin}(\pi_j, m_j)$ for $j = 1, \ldots, k$, all independent

b. $H_0 : \pi_1 = \ldots = \pi_k$ vs. $H_A : \pi_i \neq \pi_j$ for some $i, j$.

c. Approach 1: Generalized Likelihood Ratio

i. $L(\pi_1, \ldots, \pi_k) = \prod_{j=1}^{k} \left[ (m_j^{X_j})^{X_j}(1 - \pi_j)^{m_j-X_j} \right]$

ii. $\hat{\pi}_j = (\sum_{j=1}^{k} X_j)/(\sum_{j=1}^{k} m_j)$, $\tilde{\pi}_j = X_j/m_j$

iii. $\Lambda = \frac{\prod_{j=1}^{k} \left[ (m_j^{X_j})^{X_j}(1 - \hat{\pi}_j)^{m_j-X_j} \right]}{\prod_{j=1}^{k} \left[ (m_j^{X_j})^{X_j}(1 - \tilde{\pi}_j)^{m_j-X_j} \right]} = \frac{\prod_{j=1}^{k} \left[ (m_j^{X_j})^{X_j}(1 - \hat{\pi}_j)^{m_j-X_j} \right]}{\prod_{j=1}^{k} \left[ (m_j^{X_j})^{X_j}(1 - \tilde{\pi}_j)^{m_j-X_j} \right]}$

iv. No simplification possible
v. Take \(-2 \log(A)\)

\[
\begin{align*}
&= 2 \left( \sum_{j=1}^{k} [X_j \{\log(\tilde{\pi}_j) - \log(\hat{\pi}_j)\}] \\
&\quad + (m_j - X_j) \{\log(1 - \tilde{\pi}_j) - \log(1 - \hat{\pi}_j)\} \right) \\
&= 2 \left( \sum_{j=1}^{k} [X_j \{\log(X_j/m_j) - \log(\hat{\pi}_j)\}] \\
&\quad + (m_j - X_j) \{\log(1 - X_j/m_j) - \log(1 - \hat{\pi}_j)\} \right)
\end{align*}
\]

vi. Approximate by \(\chi^2\)

d. Approach 2: Approximate \(-2 \log(A)\) using Taylor series as before:

\[
\begin{align*}
&\approx \sum_{j=1}^{k} \frac{(X_j - m_j \hat{\pi})^2}{m_j \hat{\pi}} + \sum_{j=1}^{k} \frac{(m_j - X_j - m_j (1 - \hat{\pi}))^2}{m_j (1 - \hat{\pi})} \\
&= \sum_{j=1}^{k} \frac{(X_j - m_j \hat{\pi})^2}{m_j \hat{\pi} (1 - \hat{\pi})}
\end{align*}
\]

and compare to \(\chi^2_{k-1}\)

e. Call \(m_j \hat{\pi}_j\) and \(m_j (1 - \hat{\pi}_j)\) the expected number of successes and failures resp.

f. Describe statistic as sum of observed minus expected, squared, divided by expected, for successes and failures, and all \(k\)
g. Degrees of freedom are $k - 1$ rather than $k$ because $\hat{\pi}$ is estimated from data

i. Hence summands are not independent

F: 13.7

14. Tests of Equality of Proportions – Multinomial Case

a. $X_j \sim M(\pi_j, m_j)$ for $j = 1, \ldots, k$, all independent

b. $H_0: \pi_1 = \ldots = \pi_k$ vs. $H_A: \pi_l \neq \pi_j$ for some $l, j$.

c. Generalized Likelihood Ratio

i. $L(\pi_1, \ldots, \pi_k) = \prod_{j=1}^{k} \prod_{i=1}^{I} \pi^{X_{ji}}_{ji}$

- Multinomial coefficients omitted since they cancel

ii. $\hat{\pi}_{ij} = (\sum_{l=1}^{k} X_{li})/(\sum_{l=1}^{k} m_l)$, $\tilde{\pi}_{ji} = X_{ji}/m_j$

iii. $\Lambda = \prod_{j=1}^{k} \prod_{l=1}^{J} \hat{\pi}_{jl}^{X_{jl}} / \prod_{j=1}^{k} \prod_{l=1}^{J} \tilde{\pi}_{jl}^{X_{jl}}$

iv. No simplification possible

v. Take $-2 \log(\Lambda) = 2 \sum_{j=1}^{k} \sum_{i=1}^{I} X_{ji} \{ \log(\tilde{\pi}_{ji}) - \log(\hat{\pi}_{ji}) \}$

vi. Approximate by $\chi^2$

- Degrees of freedom are $(k - 1)(J - 1)$

vii. Test statistic and null distribution are same if you swap rows and columns