Statistical Inference for Categorical Data:

1. Conditional and Unconditional tests
2. Examples

Homework:
Problems: 1.6, 1.8, 1.30
1.5.1 Estimation of Multinomial parameters

Data: \( n \) observations (\( n \) is fixed)

\( y_1, \ldots, y_c \) - frequencies \( \sum_{j=1}^{c} y_j = n \)

\( C = \) the number of Categories

The multinomial pdf is proportional to the kernel

\[
L(T^n) = \prod_{i=1}^{c} \pi_j^{y_j}, \quad \sum_{j=1}^{c} \pi_j = 1, \quad \pi_j > 0
\]

Here \( \pi_j \) is the proportion of the category \( j \), \( j = 1, 2, \ldots, c \)

\[
L(T^n) = \ln L(T^n) = \sum_{j=1}^{c} y_j \ln \pi_j
\]

Let us show that MLE of \( \pi \) is \( \hat{\pi} \)

with \( \hat{\pi}_j = \frac{y_j}{n} \), \( j = 1, 2, \ldots, c \).
Note that $\pi_c = 1 - \pi_1 - \pi_2 - \ldots - \pi_{c-1}$ and

$$\frac{\partial \pi_c}{\partial \pi_j} = -1, \quad \text{for } j = 1, 2, \ldots, c-1$$

So that

$$\frac{\partial \ln(\pi_c)}{\partial \pi_j} = \frac{\partial \ln(\pi_c)}{\partial \pi_c} \cdot \frac{\partial \pi_c}{\partial \pi_j} = \frac{1}{\pi_c} \cdot (-1)$$

and

$$\frac{\partial L(\pi)}{\partial \pi_j} = \frac{c}{\pi_j} \sum_{j=1}^{c} v_j \ln \frac{\pi_j}{\pi_c} =$$

$$= \frac{v_j}{\pi_j} + \frac{1}{\pi_j} \sum_{j=1}^{c} v_j \ln \frac{\pi_j}{\pi_c} = \frac{v_j}{\pi_j} - \frac{v_c}{\pi_c}$$

i.e. $\frac{\partial L(\pi)}{\partial \pi_j} = \frac{v_j}{\pi_j} - \frac{v_c}{\pi_c}$

On the other hand

$$\sum_{j=1}^{c} \frac{v_j}{\pi_j} = 1 \quad \text{i.e.} \quad \sum_{j=1}^{c} \pi_c \cdot \frac{v_j}{\pi_c} = 1$$

or $\frac{\pi_c}{\pi_c} \cdot \frac{v_c}{\pi_c} = 1$ or $\frac{\pi_c}{\pi_c} = \frac{v_c}{\pi_c}$
and \[ \hat{\pi}_j = \frac{\hat{\nu}_j}{n}, \quad \hat{\nu}_j = \frac{\nu_j}{n} \quad \text{for} \quad j = 1, 2, \ldots, c-1 \]

Finally: \[ \hat{\pi}_j = \frac{\nu_j}{n}, \quad \text{for} \quad j = 1, 2, \ldots, c \]

MLE(\(\hat{\pi}_j\)) are the sample proportions for categories 1, 2, \ldots, c

15.2 Pearson Statistic

Consider \( H_0: \hat{\pi}_j = \pi_{j0}, \quad j = 1, 2, \ldots, c \)

where \( \sum_{j=1}^{c} \pi_{j0} = 1 \)

Pearson Test:

\[
\chi^2 = \sum_{j=1}^{c} \frac{(\nu_j - n \pi_{j0})^2}{n \pi_{j0}} \quad n \pi_{j0} \quad \text{as} \quad n \to \infty
\]

This test compares the observed frequencies \( \nu_j \) with expected ones: \( \pi_{j0} = \frac{n \pi_{j0}}{n} \)

\[ p\text{-value} = P\left( \chi^2_{c-1} \geq \chi^2_{obs} \right) \]

Remark. Here the number \( c \) is fixed
15.5 Likelihood-Ratio

Chi-squared Test

Consider again testing $H_0: \pi_j = \pi_{0j}, \ j = 1, 2, \ldots, c$

Let us consider the likelihood ratio

$$\Lambda = \frac{\prod (\pi_{0j})^{y_j}}{\prod \pi_j^{y_j}} = \frac{L_0}{L_1}$$

with $\pi_j = \frac{y_j}{n}, \ j = 1, 2, \ldots, c$

Denote

$$G^2 = -2 \ln \Lambda = -2 \sum_{j=1}^{c} y_j \ln \left( \frac{\pi_{0j}}{\pi_j} \right)$$

$$= 2 \sum_{j=1}^{c} y_j \ln \frac{\pi_j}{\pi_{0j}}$$

or

$$G^2 = 2 \sum_{j=1}^{c} \frac{y_j}{n} \ln \frac{\pi_j}{\pi_{0j}} \approx \chi^2$$

Since $\dim (H_0) = 0$ under $H_0$, the parameters are specified.

Under $(H_0 \cup H_a)$: $\dim (H_0 \cup H_a) = c - 1$
Use Testing with Estimated Expected frequencies

Assume that $\pi_j = \pi_j(\theta)$ are functions of a smaller set of parameters $\theta = (\theta_1, \ldots, \theta_p)$ 
$\dim(\theta) = p$. 

Construct MLE of $\theta : \hat{\theta}$. 

Then 
$$\sum_{j=1}^c \frac{[(2_j - n \cdot \pi_j(\hat{\theta})]^2}{n \cdot \pi_j(\hat{\theta})} \cdot \frac{n}{c-1-p}$$

Remark: The distribution of $\chi^2$, as $n \to \infty$, converge quickly to $\chi^2_{c-1-p}$ than that of distribution of LR test 

$$G^2 = -2 \ln \Lambda$$, as $n \to \infty$. 


**STAT 555  Spring 2005  L3**

**Ch. 1**

Problem 1.7  \( n = 20 \)  \( i \) trial  \( \pi \) new drug is better

Let

\[ \pi = P(\text{new drug is better}) = P(S) = P(X = 1) \]

\[ 1 - \pi = P(\text{new drug is not better}) = P(F) = P(X = 0) \]

\[ X_i = 1 \]

\[ X_i = 0 \]

\[ x \]

\[ f(x) = \pi \cdot (1 - \pi) \]

\[ x = 0, 1 \]

Let \( Y = \sum_{i=1}^{20} X_i \)

Likelihood function

\[ L(\pi) = \prod_{i=1}^{20} f(X_i) = \prod_{i=1}^{20} \pi \cdot (1 - \pi) \]

\[ \sum_{i=1}^{20} X_i \]

\[ n - \sum_{i=1}^{20} X_i \]

\[ \pi \]

\[ (1 - \pi) \]

\[ Y \cdot (1 - \pi)^{n - Y} \]

Let us test \( H_0 : \pi = \pi_0 \) vs \( H_0 : \pi \neq \pi_0 \)

where \( \pi_0 = 0.5 \)
After experiment \( Y = 20 \)

(a): MLE: \( \hat{\pi} = \frac{20}{20} = 1.0 \)

\[ l(\pi) = \pi (1-\pi) \]

\[ \text{argmax } l(\pi) = 1.0 = \hat{\pi} \]

MLE of \( \pi \) is \( \hat{\pi} = 1.0 \)

Remark: MLE of \( \pi \) in general form: \( \hat{\pi} = \frac{Y}{n} \)

(b): Wald Test: MLE - \( \pi_0 \)

\[ T_W = \frac{\hat{\pi} - \pi_0}{\sqrt{\text{var}(\hat{\pi})}} \]

but \( \text{var}(\hat{\pi}) = \frac{\hat{\pi}(1-\hat{\pi})}{n} \) so that

\[ \text{var}(\hat{\pi}) = \frac{\hat{\pi}(1-\hat{\pi})}{20} = \frac{1(1-1)}{20} = 0 \]

\[ T_W, \text{ critical} = \frac{1.0 - 0.5}{\sqrt{1.0}} = \infty \]

Wald 95% CI \( \hat{\pi} \):

\[ \hat{\pi} \pm 1.96 \sqrt{\frac{1.0}{20}} \]

i.e. CI \( \pi \): (1.0; 1.0)
(c): Conduct a score test, report the p-value
Construct a 95% score CI.

The score test:

\[ T_0 = \frac{\hat{\lambda}_n(\hat{\pi}_0)}{\sqrt{\hat{\gamma}(\hat{\pi}_0)}}, \text{ where } \hat{\lambda}_n(\pi) = \frac{\partial \tilde{L}(\pi)}{\partial \pi} \]

but \[ L(\pi) = \ln L(\pi) = (\sum x_i) \ln \pi + (n - \sum x_i) \ln (1-\pi) \]

\[ = Y \ln \pi + (n - Y) \ln (1-\pi) \]

\[ \hat{\lambda}_n(\pi) = \frac{\partial L'(\pi)}{\partial \pi} = \frac{Y}{\tilde{\pi}} - \frac{n-Y}{1-\tilde{\pi}} = \frac{Y - n \tilde{\pi} - n \pi + n \tilde{\pi}}{\tilde{\pi}(1-\tilde{\pi})} = \frac{Y - n \tilde{\pi}}{\tilde{\pi}(1-\tilde{\pi})} \]

\[ \hat{\pi}_n(\pi) = \frac{\hat{\pi}_n(\hat{\pi}_0)}{\hat{\pi}_0 (1-\hat{\pi}_0)} \]

\[ \hat{\gamma}(\pi) = -\tilde{E} \left[ \frac{\hat{\pi}^2 L'(\pi)}{\hat{\pi} + \hat{\pi}^2} \right] = -\tilde{E} \left[ \hat{\pi} - \frac{\tilde{\pi} - \tilde{\pi}^2}{1 - \tilde{\pi}} \right] \]

\[ = \frac{1}{\pi^2} E(Y) + \frac{1}{(1-\pi)^2} E(n-Y) = \frac{1}{\pi^2} \frac{n \pi}{\pi^2} + \frac{1}{(1-\pi)^2} \frac{n(1-\pi)}{(1-\pi)^2} = \frac{n}{\pi^2} + \frac{n}{1-\pi} = \frac{n}{\pi(1-\pi)} \]
\[ i.e. \quad i(\pi_0) = \frac{\pi + \pi}{n_0 + 1 - \pi_0} = \frac{\pi}{n_0(1 - \pi_0)} \]

\[ T = \frac{Y - n \pi_0}{n_0(1 - \pi_0) \cdot \sqrt{\frac{n}{n_0(1 - \pi_0)}}} = \frac{Y - n \pi_0}{\sqrt{n(\pi_0)(1 - \pi_0)}} \]

\[ T \approx \frac{1 - \pi}{\sqrt{(0.5)(0.5)}} \]

\[ T_{\text{obs}} = \sqrt{20} = 4.47 \]

\[ H_0: \pi = \pi_0 \]

\[ H_1: \pi \neq \pi_0 \]

\[ p\text{-value} = 2 \times P(N(0,1) \geq \frac{T}{\sqrt{20}}) = 2 \times P(N(0,1) \geq 4.47) < 0.001 \]

the score CI₀!

the endpoints: \[ \pi \pm \frac{T_{\text{obs}}}{\sqrt{20(1 - \pi_0)}} = 1.96 \]

\[ 95\% \ CI_{\pi} : (0.839; 1.0) \]

-18-
Ch. 1 Conditional Wald Test $T_1$:

Problem 1.18: Let $Y_1 \sim \text{Poisson} (\mu_1)$ $Y_2 \sim \text{Poisson} (\mu_2)$. $Y_1$ and $Y_2$ are independent r.v.'s.

To test $H_0: \mu_1 = \mu_2$ given the sample $(Y_1, Y_2)$.

Approach 1: $Y_1 = \lambda_{Y_1 + Y_2} = \lambda \sim \text{Bin}(\lambda, n, \pi)$

where $\pi = \frac{\mu_1}{\mu_1 + \mu_2}$

The binomial test is based on $T_1$.

$H_0: \mu_1 = \mu_2$ is equivalent to $H_0: \pi = \frac{1}{2}$ for binomial parameter.

Large sample Wald Test:

$T_1 = n \cdot \frac{1}{\hat{\pi}}$ with $\hat{\pi} = \frac{Y_1}{Y_1 + Y_2}$ $n \sim \text{N}(0, 1)$

$\sqrt{n \cdot \hat{\pi} (1 - \hat{\pi})}$

given $n = Y_1 + Y_2$. 