Goodness of Fit Tests
Mathematics 47: Lecture 32

Dan Sloughter

Furman University

May 5, 2006
Testing multinomial parameters

- Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a finite distribution with $k$ possible outcomes.
Testing multinomial parameters

- Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a finite distribution with $k$ possible outcomes.
- For $i = 1, 2, \ldots, k$, let $p_i =$ probability of outcome $i$. 

$$p_i = \text{probability of outcome } i.$$
Testing multinomial parameters

- Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a finite distribution with $k$ possible outcomes.
- For $i = 1, 2, \ldots, k$, let
  \[ p_i = \text{probability of outcome } i. \]
- Let $\mathbf{p} = (p_1, p_2, \ldots, p_k)$. 
Testing multinomial parameters

- Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a finite distribution with $k$ possible outcomes.
- For $i = 1, 2, \ldots, k$, let
  \[ p_i = \text{probability of outcome } i. \]
- Let $\mathbf{p} = (p_1, p_2, \ldots, p_k)$.
- Given some $\Pi = (\pi_1, \pi_2, \ldots, \pi_k)$, where $0 \leq \pi_i \leq 1$ for $i = 1, 2, \ldots, k$ and $\sum_{i=1}^{k} \pi_i = 1$, suppose we wish to test
  \[ H_0 : \mathbf{p} = \Pi \]
  \[ H_A : \mathbf{p} \neq \Pi. \]
If, for \( i = 1, 2, \ldots, k \), we let

\[ Y_i = \text{number of outcomes of type } i, \]

then the likelihood function is

\[ L(p_1, p_2, \ldots, p_k) = p_1^{Y_1} p_2^{Y_2} \cdots p_k^{Y_k}. \]
Testing multinomial parameters (cont’d)

▶ If, for \( i = 1, 2, \ldots, k \), we let

\[
Y_i = \text{number of outcomes of type } i,
\]

then the likelihood function is

\[
L(p_1, p_2, \ldots, p_k) = p_1^{Y_1} p_2^{Y_2} \cdots p_k^{Y_k}.
\]

▶ We have seen previously that the maximum likelihood estimator for \( p_i \) is

\[
\hat{p}_i = \frac{Y_i}{n},
\]

so the generalized likelihood ratio is

\[
\Lambda = \frac{L(\pi_1, \pi_2, \ldots, \pi_k)}{L(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_k)} = \prod_{i=1}^{k} \left( \frac{\pi_i}{\hat{p}_i} \right)^{Y_i}.
\]
Testing multinomial parameters (cont’d)

Hence

\[-2 \log(\Lambda) = -2 \sum_{i=1}^{k} Y_i \log \left( \frac{\pi_i}{\hat{p}_i} \right)\]

\[= 2 \sum_{i=1}^{k} Y_i \log \left( \frac{\hat{p}_i}{\pi_i} \right)\]

\[= 2 \sum_{i=1}^{k} Y_i \log \left( \frac{Y_i}{n\pi_i} \right) .\]
Testing multinomial parameters (cont’d)

-2 \log(\Lambda) = -2 \sum_{i=1}^{k} Y_i \log \left( \frac{\pi_i}{\hat{p}_i} \right)

= 2 \sum_{i=1}^{k} Y_i \log \left( \frac{\hat{p}_i}{\pi_i} \right)

= 2 \sum_{i=1}^{k} Y_i \log \left( \frac{Y_i}{n\pi_i} \right).

- Note: when \( H_0 \) is true, \( n\pi_i = E[Y_i] \), the expected number of observations of outcome \( i \).
Testing multinomial parameters (cont’d)

Hence

\[-2 \log(\Lambda) = -2 \sum_{i=1}^{k} Y_i \log \left( \frac{\pi_i}{\hat{p}_i} \right)\]

\[= 2 \sum_{i=1}^{k} Y_i \log \left( \frac{\hat{p}_i}{\pi_i} \right)\]

\[= 2 \sum_{i=1}^{k} Y_i \log \left( \frac{Y_i}{n\pi_i} \right).\]

Note: when \(H_0\) is true, \(n\pi_i = E[Y_i]\), the expected number of observations of outcome \(i\).

Hence we may remember the final expression for \(-2 \log(\Lambda)\) as

\[2 \sum (\text{Observed}) \log \left( \frac{\text{Observed}}{\text{Expected}} \right),\]

where the sum extends over all possible outcomes.
Testing multinomial parameters (cont’d)

Now for large $n$, $-2 \log(\Lambda)$ is approximately $\chi^2(k - 1)$ ($k - 1$, not $k$, because there are only $k - 1$ parameters since $\sum_{i=1}^{k} p_i = 1$).

Note: A conservative rule of thumb is to assume the approximation to the chi-square distribution is reasonable as long as $n \pi_i \geq 5$ for $i = 1, 2, \ldots, k$. 

Dan Slaughter (Furman University)
Now for large $n$, $-2 \log(\Lambda)$ is approximately $\chi^2(k - 1)$ ($k - 1$, not $k$, because there are only $k - 1$ parameters since $\sum_{i=1}^{k} p_i = 1$).

Since the generalized likelihood ratio test rejects $H_0$ for small values of $\Lambda$, we reject $H_0$ for large values of $-2 \log(\Lambda)$, with, for an observed value $\lambda$ of $\Lambda$, $p$-value equal to $P(U \geq -2 \log(\lambda))$, where $U$ is $\chi^2(k - 1)$.
Now for large $n$, $-2 \log(\Lambda)$ is approximately $\chi^2(k-1)$ ($k-1$, not $k$, because there are only $k-1$ parameters since $\sum_{i=1}^{k} p_i = 1$).

Since the generalized likelihood ratio test rejects $H_0$ for small values of $\Lambda$, we reject $H_0$ for large values of $-2 \log(\Lambda)$, with, for an observed value $\lambda$ of $\Lambda$, $p$-value equal to $P(U \geq -2 \log(\lambda))$, where $U$ is $\chi^2(k-1)$.

Note: A conservative rule of thumb is to assume the approximation to the chi-square distribution is reasonable as long as $n\pi_i \geq 5$ for $i = 1, 2, \ldots, k$. 
Example

In one of his genetic experiments, Gregor Mendel crossed 556 smooth-yellow male peas with wrinkled-green female peas. For each of the 556 crossings, there are four possible outcomes: smooth-yellow, smooth-green, wrinkled-yellow, and wrinkled-green.

Let \( p_1 \) = probability of smooth-yellow, \( p_2 \) = probability of smooth-green, \( p_3 \) = probability of wrinkled-yellow, and \( p_4 \) = probability of wrinkled-green.

Mendel's theory predicted that \( p_1 = \frac{9}{16} \), \( p_2 = \frac{3}{16} \), \( p_3 = \frac{3}{16} \), and \( p_4 = \frac{1}{16} \).
In one of his genetic experiments, Gregor Mendel crossed 556 smooth-yellow male peas with wrinkled-green female peas. For each of the 556 crossings, there are four possible outcomes: smooth-yellow, smooth-green, wrinkled-yellow, and wrinkled-green.
Example

- In one of his genetic experiments, Gregor Mendel crossed 556 smooth-yellow male peas with wrinkled-green female peas. For each of the 556 crossings, there are four possible outcomes: smooth-yellow, smooth-green, wrinkled-yellow, and wrinkled-green.

- Let

\[
p_1 = \text{probability of smooth-yellow},
\]
\[
p_2 = \text{probability of smooth-green},
\]
\[
p_3 = \text{probability of wrinkled-yellow},
\]
\[
p_4 = \text{probability of wrinkled-green}.
\]
Example

- In one of his genetic experiments, Gregor Mendel crossed 556 smooth-yellow male peas with wrinkled-green female peas. For each of the 556 crossings, there are four possible outcomes: smooth-yellow, smooth-green, wrinkled-yellow, and wrinkled-green.

- Let

\[ p_1 = \text{probability of smooth-yellow}, \]
\[ p_2 = \text{probability of smooth-green}, \]
\[ p_3 = \text{probability of wrinkled-yellow}, \]
\[ p_4 = \text{probability of wrinkled-green}. \]

- Mendel’s theory predicted that \( p_1 = \frac{9}{16}, p_2 = \frac{3}{16}, p_3 = \frac{3}{16}, \) and \( p_4 = \frac{1}{16}. \)
Example

The following table contains Mendel's data, along with the expected frequency of each outcome under $H_0$:

<table>
<thead>
<tr>
<th>Type</th>
<th>Observed</th>
<th>Frequency</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smooth-yellow</td>
<td>315</td>
<td></td>
<td>312.75</td>
</tr>
<tr>
<td>Smooth-green</td>
<td>108</td>
<td></td>
<td>104.25</td>
</tr>
<tr>
<td>Wrinkled-yellow</td>
<td>101</td>
<td></td>
<td>104.25</td>
</tr>
<tr>
<td>Wrinkled-green</td>
<td>32</td>
<td></td>
<td>34.75</td>
</tr>
<tr>
<td>Total</td>
<td>556</td>
<td></td>
<td>556.00</td>
</tr>
</tbody>
</table>

$H_A$: $p \neq (\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16})$.
Example

- Hence we want to test

\[ H_0 : \mathbf{p} = \left( \frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16} \right) \]

\[ H_A : \mathbf{p} \neq \left( \frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16} \right) . \]
Example

- Hence we want to test

\[ H_0 : p = \left( \frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16} \right) \]

\[ H_A : p \neq \left( \frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16} \right) . \]

- The following table contains Mendel’s data, along with the expected frequency of each outcome under \( H_0 \):

<table>
<thead>
<tr>
<th>Type</th>
<th>Observed Frequency</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smooth-yellow</td>
<td>315</td>
<td>312.75</td>
</tr>
<tr>
<td>Smooth-green</td>
<td>108</td>
<td>104.25</td>
</tr>
<tr>
<td>Wrinkled-yellow</td>
<td>101</td>
<td>104.25</td>
</tr>
<tr>
<td>Wrinkled-green</td>
<td>32</td>
<td>34.75</td>
</tr>
<tr>
<td>Total</td>
<td>556</td>
<td>556.00</td>
</tr>
</tbody>
</table>
We may now compute 
\[-2 \log(\lambda) = 2 \sum_{i=1}^{4} y_i \log \left( \frac{y_i}{\pi_i} \right) = 0.4754.\]

If \(U\) is \(\chi^2(3)\), then the \(p\)-value for this test is 
\[P(U \geq 0.4754) = 0.9242617,\]
giving us no evidence for rejecting the null hypothesis.

Indeed, these data provide such a good fit to Mendel's data that it has been suspected that his assistants manipulated the data to fit the hypothesis.
Example

» We may now compute

\[-2 \log(\lambda) = 2 \sum_{i=1}^{4} y_i \log \left( \frac{y_i}{n\pi_i} \right) = 0.4754.\]
Example

- We may now compute

\[-2 \log(\lambda) = 2 \sum_{i=1}^{4} y_i \log \left( \frac{y_i}{n\pi_i} \right) = 0.4754.\]

- If $U$ is $\chi^2(3)$, then the $p$-value for this test is $P(U \geq 0.4754) = 0.9242617$, giving us no evidence for rejecting the null hypothesis.
Example

- We may now compute

\[-2 \log(\lambda) = 2 \sum_{i=1}^{4} y_i \log \left( \frac{y_i}{n\pi_i} \right) = 0.4754.\]

- If $U$ is $\chi^2(3)$, then the $p$-value for this test is $P(U \geq 0.4754) = 0.9242617$, giving us no evidence for rejecting the null hypothesis.

- Indeed, these data provide such a good fit to Mendel’s data that it has been suspected that his assistants manipulated the data to fit the hypothesis.
Pearson’s chi-square statistic

Let

\[ f(x) = x \log \left( \frac{x}{a} \right). \]
Pearson’s chi-square statistic

Let

\[ f(x) = x \log \left( \frac{x}{a} \right). \]

Then

\[ f'(x) = x \cdot \frac{1}{x} \cdot \frac{1}{a} + \log \left( \frac{x}{a} \right) = 1 + \log \left( \frac{x}{a} \right), \]

and

\[ f''(x) = \frac{1}{x}. \]
Pearson’s chi-square statistic

Let

\[ f(x) = x \log \left( \frac{x}{a} \right). \]

Then

\[ f'(x) = x \cdot \frac{1}{x} \cdot \frac{1}{a} + \log \left( \frac{x}{a} \right) = 1 + \log \left( \frac{x}{a} \right) \]

and

\[ f''(x) = \frac{1}{x}. \]

So \( f(a) = 0, \) \( f'(a) = 1, \) and \( f''(a) = \frac{1}{a}. \)
Pearson’s chi-square statistic

Let

\[ f(x) = x \log \left( \frac{x}{a} \right). \]

Then

\[ f'(x) = x \cdot \frac{1}{x} \cdot \frac{1}{a} + \log \left( \frac{x}{a} \right) = 1 + \log \left( \frac{x}{a} \right) \]

and

\[ f''(x) = \frac{1}{x}. \]

So \( f(a) = 0, \) \( f'(a) = 1, \) and \( f''(a) = \frac{1}{a}. \)

Hence, for \( x \) close to \( a, \)

\[ f(x) \approx (x - a) + \frac{1}{2a}(x - a)^2. \]
Pearson’s chi-square statistic

Thus, when $H_0$ is true and $n$ is large,

$$y_i \log \left( \frac{y_i}{n\pi_i} \right) \approx (y_i - n\pi_i) + \frac{1}{2n\pi_i} (y_i - n\pi_i)^2.$$
Pearson’s chi-square statistic

Thus, when $H_0$ is true and $n$ is large,

$$y_i \log \left( \frac{y_i}{n \pi_i} \right) \approx (y_i - n \pi_i) + \frac{1}{2n \pi_i} (y_i - n \pi_i)^2.$$ 

It follows that

$$-2 \log(\Lambda) = 2 \sum_{i=1}^{k} Y_i \log \left( \frac{Y_i}{n \pi_i} \right)$$

$$\approx 2 \sum_{i=1}^{k} (Y_i - n \pi_i) + \sum_{i=1}^{k} \frac{(Y_i - n \pi_i)^2}{n \pi_i}$$

$$= 2(n - n) + \sum_{i=1}^{k} \frac{(Y_i - n \pi_i)^2}{n \pi_i}$$

$$= \sum_{i=1}^{k} \frac{(Y_i - n \pi_i)^2}{n \pi_i}.$$
We call

\[ Q = \sum_{i=1}^{k} \frac{(Y_i - n\pi_i)^2}{n\pi_i} \]

*Pearson’s chi-square statistic.*
Pearson’s chi-square statistic

- We call

\[ Q = \sum_{i=1}^{k} \frac{(Y_i - n\pi_i)^2}{n\pi_i} \]

*Pearson’s chi-square statistic.*

- We may remember this formula as

\[ \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \]

where the sum extends over all possible outcomes.
Pearson’s chi-square statistic

- We call

\[ Q = \sum_{i=1}^{k} \frac{(Y_i - n\pi_i)^2}{n\pi_i} \]

*Pearson’s chi-square statistic.*

- We may remember this formula as

\[ \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}, \]

where the sum extends over all possible outcomes.

- We may use either \(-2\log(\Lambda)\) or \(Q\) to perform the *chi-square goodness of fit test.*
Example
Example

For Mendel’s data, we have

\[ q = \sum_{i=1}^{4} \frac{(y_i - n\pi_i)^2}{n\pi_i} = 0.4700, \]

differing only slightly from the value of \(-2 \log(\lambda)\) computed above.
Example

- For Mendel’s data, we have

\[ q = \sum_{i=1}^{4} \frac{(y_i - n\pi_i)^2}{n\pi_i} = 0.4700, \]

differing only slightly from the value of \(-2 \log(\lambda)\) computed above.

- Note: If the observed frequencies are in the vector \(x\) and the hypothesized probabilities are in the vector \(\pi\), then the \(R\) command

\[
> \text{chisq.test}(x, \text{p} = \pi)
\]

performs the goodness of fit test using Pearson’s chi-square statistic.