The Multinomial Model\footnote{See last slide for copyright information.}

STA 312: Fall 2012
Overview

1. Multinomial Coefficients
2. Multinomial Distribution
3. Estimation
4. Hypothesis tests
5. Power
**Multinomial coefficient**

*For c categories*

From $n$ objects, number of ways to choose

- $n_1$ of type 1
- $n_2$ of type 2
- $\vdots$
- $n_c$ of type $c$

$$\binom{n}{n_1 \cdots n_c} = \frac{n!}{n_1! \cdots n_c!}$$
Example of a multinomial coefficient
A counting problem

Of 30 graduating students, how many ways are there for 15 to be employed in a job related to their field of study, 10 to be employed in a job unrelated to their field of study, and 5 unemployed?

\[
\binom{30}{15 \ 10 \ 5} = 465, 817, 912, 560
\]
Multinomial Distribution
Denote by $M(n, \pi)$, where $\pi = (\pi_1, \ldots, \pi_c)$

- Statistical experiment with $c$ outcomes
- Repeated independently $n$ times
- $Pr(\text{Outcome } j) = \pi_j$, $j = 1, \ldots, c$
- Number of times outcome $j$ occurs is $n_j$, $j = 1, \ldots, c$
- An integer-valued multivariate distribution

$$P(n_1, \ldots, n_c) = \binom{n}{n_1 \ldots n_c} \pi_1^{n_1} \cdots \pi_c^{n_c},$$

where $0 \leq n_j \leq n$, $\sum_{j=1}^c n_j = n$, $0 < \pi_j < 1$, and $\sum_{j=1}^c \pi_j = 1$. 
There are actually $c - 1$ variables and $c - 1$ parameters
In the multinomial with $c$ categories

\[
P(n_1, \ldots, n_{c-1}) = \frac{n!}{n_1! \cdots n_{c-1}!(n - \sum_{j=1}^{c-1} n_j)!} \\
\times \pi_1^{n_1} \cdots \pi_{c-1}^{n_{c-1}} (1 - \sum_{j=1}^{c-1} \pi_j)^{n - \sum_{j=1}^{c-1} n_j}
\]
Marginals of the multinomial are multinomial too
Add over \( n_{c-1} \), which goes from zero to whatever is left over from the other counts.

\[
\begin{align*}
\sum_{n_{c-1}=0}^{n-\sum_{j=1}^{c-2} n_j} \frac{n!}{n_1! \cdots n_{c-1}!(n - \sum_{j=1}^{c-1} n_j)!} \pi_1 \cdots \pi_{c-1} (1 - \sum_{j=1}^{c-1} \pi_j)^{n-\sum_{j=1}^{c-1} n_j} \times \frac{(n - \sum_{j=1}^{c-2} n_j)!}{(n - \sum_{j=1}^{c-2} n_j)!} \\
\times \sum_{n_{c-1}=0}^{n-\sum_{j=1}^{c-2} n_j} \frac{(n - \sum_{j=1}^{c-2} n_j)!}{n_{c-1}!(n - \sum_{j=1}^{c-2} n_j - n_{c-1})!} \pi_{c-1} (1 - \sum_{j=1}^{c-2} \pi_j - \pi_{c-1})^{n-\sum_{j=1}^{c-2} n_j - n_{c-1}} \\
= \frac{n!}{n_1! \cdots n_{c-2}!(n - \sum_{j=1}^{c-2} n_j)!} \pi_1 \cdots \pi_{c-2} (1 - \sum_{j=1}^{c-2} \pi_j)^{n-\sum_{j=1}^{c-2} n_j},
\end{align*}
\]

where the last equality follows from the Binomial Theorem. It’s multinomial with \( c - 1 \) categories.
Observe
You are responsible for these *implications* of the last slide.

- Adding over $n_{c-1}$ throws it into the last (“leftover”) category.
- Labels 1, ..., $c$ are arbitrary, so this means you can combine any 2 categories and the result is still multinomial.
- $c$ is arbitrary, so you can keep doing it and combine any number of categories.
- When only two categories are left, the result is binomial
  - $E(n_j) = n\pi_j = \mu_j$, $Var(n_j) = n\pi_j(1 - \pi_j)$
Sample problem
Recent university graduates

- Probability of job related to field of study = 0.60
- Probability of job unrelated to field of study = 0.30
- Probability of no job = 0.10

Of 30 randomly chosen students, what is probability that 15 are employed in a job related to their field of study, 10 are employed in a job unrelated to their field of study, and 5 are unemployed?

\[
\binom{30}{15, 10, 5} \times 0.60^{15} \times 0.30^{10} \times 0.10^5 = \frac{4933527642332542053801}{38146972656250000000000000} \approx 0.0129
\]

What is the probability that exactly 5 are unemployed?

\[
\binom{30}{5, 25} \times 0.10^5 \times 0.90^{25} = \frac{51152385317007572507646551997}{500000000000000000000000000000} \approx 0.1023
\]
Conditional probabilities are multinomial too

- Given that a student finds a job, what is the probability that the job will be in the student’s field of study?

\[ P(\text{Field} | \text{Job}) = \frac{P(\text{Field, Job})}{P(\text{Job})} = \frac{0.60}{0.90} = \frac{2}{3} \]

- Suppose we choose 50 students at random from those who found jobs. What is the probability that exactly \( y \) of them will be employed in their field of study, for \( y = 0, \ldots, 50 \)?

\[ P(y | \text{Job}) = \binom{50}{y} \left( \frac{2}{3} \right)^y \left( 1 - \frac{2}{3} \right)^{50-y} \]
Calculating multinomial probabilities with R

Of 30 randomly chosen students, what is probability that 15 are employed in a job related to their field of study, 10 are employed in a job unrelated to their field of study, and 5 are unemployed?

\[
\binom{30}{15\ 10\ 5} 0.60^{15} 0.30^{10} 0.10^5 = \frac{4933527642332542053801}{3814697265625000000000000000} \approx 0.0129
\]

> dmultinom(c(15,10,5), prob=c(.6, .3, .1))
[1] 0.01293295
### Hypothetical data file

Let $Y_{i,j}$ be indicators for category membership, $i = 1, \ldots, n$ and $j = 1, \ldots, c$

<table>
<thead>
<tr>
<th>Case</th>
<th>Job</th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>$n$</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Total: $\sum_{i=1}^{n} y_{i,1} \quad \sum_{i=1}^{n} y_{i,2} \quad \sum_{i=1}^{n} y_{i,3}$

Note that

- A real data file will almost never have the redundant variables $Y_1$, $Y_2$ and $Y_3$.
- $\sum_{i=1}^{n} y_{i,j} = n_j$
Lessons from the data file

- Cases ($n$ of them) are independent $M(1, \pi)$, so $E(Y_{i,j}) = \pi_j$.
- Column totals $n_j$ count the number of times each category occurs: Joint distribution is $M(n, \pi)$
- If you make a frequency table (frequency distribution)
  - The $n_j$ counts are the cell frequencies!
  - They are random variables, and now we know their joint distribution.
  - Each individual (marginal) table frequency is $B(n, \pi_j)$.
  - Expected value of cell frequency $j$ is $E(n_j) = n\pi_j = \mu_j$
- Tables of 2 and or more dimensions present no problems; form combination variables.
Example of a frequency table
For the Jobs data

<table>
<thead>
<tr>
<th>Job Category</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employed in field</td>
<td>106</td>
<td>53</td>
</tr>
<tr>
<td>Employed outside field</td>
<td>74</td>
<td>37</td>
</tr>
<tr>
<td>Unemployed</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>100.0</td>
</tr>
</tbody>
</table>
Likelihood function for the multinomial

\[ \ell(\pi) = \prod_{i=1}^{n} Pr\{Y_{i,1} = y_{i,1}, Y_{i,2} = y_{i,2}, \ldots, Y_{i,c} = y_{i,c} | \pi\} \]

\[ = \prod_{i=1}^{n} \pi_{1}^{y_{i,1}} \pi_{2}^{y_{i,2}} \cdots \pi_{c}^{y_{i,c}} \]

\[ = \prod_{i=1}^{n} \frac{\sum_{i=1}^{n} y_{i,1}}{\pi_{1}} \frac{\sum_{i=1}^{n} y_{i,2}}{\pi_{2}} \cdots \frac{\sum_{i=1}^{n} y_{i,c}}{\pi_{c}} \]

\[ = \pi_{1}^{n_{1}} \pi_{2}^{n_{2}} \cdots \pi_{c}^{n_{c}} \]

- Product of \( n \) probability mass functions, each \( M(1, \pi) \)
- Depends upon the sample data only through the vector of \( c \) frequency counts: \( (n_{1}, \ldots, n_{c}) \)
All you need is the frequency table

\[ \ell(\pi) = \pi_1^{n_1} \pi_2^{n_2} \cdots \pi_c^{n_c} \]

- Likelihood function depends upon the sample data only through the frequency counts.
- By the factorization theorem, \((n_1, \ldots, n_c)\) is a sufficient statistic.
- All the information about the parameter in the sample data is contained in the sufficient statistic.
- So everything the sample data could tell you about \((\pi_1, \ldots, \pi_c)\) is given by in \((n_1, \ldots, n_c)\).
- You don’t need the raw data.
Log likelihood: $c - 1$ parameters

\[
\ell(\pi) = \pi_1^{n_1} \cdots \pi_c^{n_c} \\
= \pi_1^{n_1} \cdots \pi_{c-1}^{n_{c-1}} \left(1 - \sum_{j=1}^{c-1} \pi_j\right)^{n - \sum_{j=1}^{c-1} n_j}
\]

\[
\log \ell(\pi) = \sum_{j=1}^{c-1} n_j \log \pi_j + \left(n - \sum_{j=1}^{c-1} n_j\right) \log \left(1 - \sum_{j=1}^{c-1} \pi_j\right)
\]

\[
\frac{\partial \log \ell}{\partial \pi_j} = \frac{n_j}{\pi_j} - \frac{n - \sum_{k=1}^{c-1} n_k}{1 - \sum_{k=1}^{c-1} \pi_k}, \text{ for } j = 1, \ldots, c - 1
\]

Set all the partial derivatives to zero and solve
For \( \pi_j, j = 1 \ldots, c-1 \)

\[
\hat{\pi}_j = \frac{n_j}{n} = p_j = \frac{\sum_{i=1}^{n} y_{i,j}}{n} = \bar{y}_j
\]

So the MLE is the sample proportion, which is also a sample mean.
In matrix terms: \( \hat{\pi} = p = Y_n \)

\[
\begin{pmatrix}
\hat{\pi}_1 \\
\vdots \\
\hat{\pi}_{c-1}
\end{pmatrix} = \begin{pmatrix}
p_1 \\
\vdots \\
p_{c-1}
\end{pmatrix} = \begin{pmatrix}
Y_1 \\
\vdots \\
Y_{c-1}
\end{pmatrix}
\]

Remarks:

- Multivariate Law of Large Numbers says \( p \xrightarrow{p} \pi \)
- Multivariate Central Limit Theorem says that \( Y_n \) is approximately multivariate normal for large \( n \).
- Because \( n_j \sim B(n, \pi_j) \), \( \frac{n_j}{n} = \bar{Y}_j = p_j \) is approximately \( N \left( \pi_j, \frac{\pi_j(1-\pi_j)}{n} \right) \).
- Approximate \( \pi_j \) with \( p_j \) in the variance if necessary.
- Can be used in confidence intervals and tests about a single parameter.
- We have been using \( c - 1 \) categories only for technical convenience.
Confidence interval for a single parameter $\pi_j$

95% confidence interval for true proportion unemployed

<table>
<thead>
<tr>
<th>Job Category</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employed in field</td>
<td>106</td>
<td>53</td>
</tr>
<tr>
<td>Employed outside field</td>
<td>74</td>
<td>37</td>
</tr>
<tr>
<td>Unemployed</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>100.0</td>
</tr>
</tbody>
</table>

$p_3 \approx N \left( \pi_3, \frac{\pi_3(1 - \pi_3)}{n} \right)$

So a confidence interval for $\pi_3$ is

$$p_3 \pm 1.96 \sqrt{\frac{p_3(1 - p_3)}{n}} = 0.10 \pm 1.96 \sqrt{\frac{0.10(1 - 0.10)}{200}} = 0.10 \pm 0.042 = (0.058, 0.242)$$
For general tests on multinomial data

We will use mostly
- Pearson chi-squared tests
- Large-sample likelihood ratio tests

There are other possibilities, including
- Wald tests
- Score tests

All these are large-sample chi-squared tests, justified as $n \to \infty$
**Likelihood ratio tests**

**In general**

**Setup**

\[ Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} F_\beta, \beta \in \mathcal{B}, \]

\[ H_0 : \beta \in \mathcal{B}_0 \text{ v.s. } H_1 : \beta \in \mathcal{B}_1 = \mathcal{B} \cap \mathcal{B}_0^c \]

**Test Statistic:**

\[
G^2 = -2 \log \left( \frac{\max_{\beta \in \mathcal{B}_0} \ell(\beta)}{\max_{\beta \in \mathcal{B}} \ell(\beta)} \right) = -2 \log \left( \frac{\ell_0}{\ell_1} \right)
\]
What to do
And how to think about it

\[ G^2 = -2 \log \left( \frac{\max_{\beta \in B_0} \ell(\beta)}{\max_{\beta \in B} \ell(\beta)} \right) = -2 \log \left( \frac{\ell_0}{\ell_1} \right) \]

- Maximize the likelihood over the whole parameter space. You already did this to calculate the MLE. Evaluate the likelihood there. That’s the denominator.
- Maximize the likelihood over just the parameter values where \( H_0 \) is true – that is, over \( B_0 \). This yields a restricted MLE. Evaluate the likelihood there. That’s the numerator.
- The numerator cannot be larger, because \( B_0 \subset B \).
- If the numerator is a lot less than the denominator, the null hypothesis is unbelievable, and
  - The ratio is close to zero
  - The log of the ratio is a big negative number
  - \(-2\) times the log is a big positive number
  - Reject \( H_0 \) when \( G^2 \) is large enough.
Given some technical conditions,

- $G^2$ has an approximate chi-squared distribution under $H_0$ for large $n$.
- Degrees of freedom equal number of (non-redundant) equalities specified by $H_0$.
- Reject $H_0$ when $G^2$ is larger than the chi-squared critical value.
Counting degrees of freedom

- Express $H_0$ as a set of linear combinations of the parameters, set equal to constants (usually zeros for regression problems).

- Degrees of freedom = number of non-redundant (linearly independent) linear combinations.

Suppose $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_7)$, with

$H_0 : \beta_1 = \beta_2, \beta_6 = \beta_7, \frac{1}{3} (\beta_1 + \beta_2 + \beta_3) = \frac{1}{3} (\beta_4 + \beta_5 + \beta_6)$

Then $df = 3$: Count the equals signs.

But if $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \frac{1}{7}$, then $df = 6$
University administrators recognize that the percentage of students who are unemployed after graduation will vary depending upon economic conditions, but they claim that still, about twice as many students will be employed in a job related to their field of study, compared to those who get an unrelated job. To test this hypothesis, they select a random sample of 200 students from the most recent class, and observe 106 employed in a job related to their field of study, 74 employed in a job unrelated to their field of study, and 20 unemployed. Test the hypothesis using a large-sample likelihood ratio test and the usual 0.05 significance level. State your conclusions in symbols and words.
Some detailed questions
To guide us through the problem

• What is the model?

\[ Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} M(1, (\pi_1, \pi_2, \pi_3)) \]

• What is the null hypothesis, in symbols?

\[ H_0 : \pi_1 = 2\pi_2 \]

• What are the degrees of freedom for this test?

1
What is the parameter space $\mathcal{B}$?
What is the restricted parameter space $\mathcal{B}_0$?

\[
\mathcal{B} = \{ (\pi_1, \pi_2) : 0 < \pi_1 < 1, 0 < \pi_2 < 1, \pi_1 + \pi_2 < 1 \}
\]
\[
\mathcal{B}_0 = \{ (\pi_1, \pi_2) : 0 < \pi_1 < 1, 0 < \pi_2 < 1, \pi_1 = 2\pi_2 \}.
\]
What is the unrestricted MLE?

Give the answer in both symbolic and numerical form. Just write it down. There is no need to show any work.

\[ \mathbf{p} = \left( \frac{n_1}{n}, \frac{n_2}{n}, \frac{n_3}{n} \right) \]

\[ = \left( \frac{106}{200}, \frac{74}{200}, \frac{20}{200} \right) \]

\[ = (0.53, 0.37, 0.10) \]
Derive the restricted MLE

Your answer is a symbolic expression. It’s a vector. Show your work.

\[ \frac{\partial}{\partial \pi} (n_1 \log(2\pi) + n_2 \log \pi + n_3 \log(1 - 3\pi)) \]

\[ = \frac{n_1}{\pi} + \frac{n_2}{\pi} + \frac{n_3}{1 - 3\pi}(-3) \quad \text{set} \quad 0 \]

\[ \Rightarrow \frac{n_1 + n_2}{\pi} = \frac{3n_3}{1 - 3\pi} \]

\[ \Rightarrow (n_1 + n_2)(1 - 3\pi) = 3\pi n_3 \]

\[ \Rightarrow n_1 + n_2 = 3\pi(n_1 + n_2 + n_3) = 3\pi n \]

\[ \Rightarrow \pi = \frac{n_1 + n_2}{3n} \]

So \( \hat{\pi} = \left( \frac{2(n_1 + n_2)}{3n}, \frac{n_1 + n_2}{3n}, \frac{n_3}{n} \right) \). From now on, \( \hat{\pi} \) means the restricted MLE.
Give the restricted MLE in numeric form
The answer is a vector of 3 numbers

\[ \hat{\pi} = \left( \frac{2(n_1 + n_2)}{3n}, \frac{n_1 + n_2}{3n}, \frac{n_3}{n} \right) \]

\[ = \left( \frac{2(106 + 74)}{600}, \frac{106 + 74}{600}, \frac{20}{200} \right) \]

\[ = (0.6, 0.3, 0.1) \]
Show the restricted and unrestricted MLEs
Restricted is black diamond, unrestricted is red circle
Calculate $G^2$. Show your work.
The answer is a number.

$$G^2 = -2 \log \frac{\hat{\pi}_1^{n_1} \hat{\pi}_2^{n_2} p_3^{n_3}}{p_1^{n_1} p_2^{n_2} p_3^{n_3}}$$

$$= -2 \left( \log \left[ \frac{\hat{\pi}_1}{p_1} \right]^{n_1} + \log \left[ \frac{\hat{\pi}_2}{p_2} \right]^{n_2} \right)$$

$$= -2 \left( n_1 \log \frac{\hat{\pi}_1}{p_1} + n_2 \log \frac{\hat{\pi}_2}{p_2} \right)$$

$$= -2 \left( 106 \log \frac{0.60}{0.53} + 74 \log \frac{0.30}{0.37} \right)$$

$$= 4.739$$
Do the calculation with R.
Display the critical value and $p$-value as well.

```r
> G2 <- -2*(106*log(60/53)+74*log(30/37)); G2
[1] 4.739477
> qchisq(0.95,df=1) # Critical value
[1] 3.841459
> pval <- 1-pchisq(G2,df=1); pval
[1] 0.02947803
```
State your conclusions

- **In symbols**: Reject $H_0 : \pi_1 = 2\pi_2$ at $\alpha = 0.05$.
- **In words**: More graduates appear to be employed in jobs unrelated to their fields of study than predicted.

The statement in words can be justified by comparing observed frequencies to those expected under $H_0$.

<table>
<thead>
<tr>
<th>Related</th>
<th>Unrelated</th>
<th>Unemployed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>106</td>
<td>74</td>
</tr>
<tr>
<td>Expected</td>
<td>120</td>
<td>60</td>
</tr>
<tr>
<td>Residual</td>
<td>-14</td>
<td>14</td>
</tr>
</tbody>
</table>

Expected frequency is $E(n_j) = \mu_j = n\pi_j$. Estimated expected frequency is $\hat{\mu}_j = n\hat{\pi}_j$. 
Write $G^2$ in terms of observed and expected frequencies

For a general hypothesis about a multinomial

\[
G^2 = -2 \log \left( \frac{\ell_0}{\ell_1} \right)
\]

\[
= -2 \log \left( \frac{\prod_{j=1}^{c} \hat{\pi}_j^{n_j}}{\prod_{j=1}^{c} p_j^{n_j}} \right)
\]

\[
= -2 \log \prod_{j=1}^{k} \left( \frac{\hat{\pi}_j}{p_j} \right)^{n_j} = 2 \sum_{j=1}^{c} - \log \left( \frac{\hat{\pi}_j}{p_j} \right) x_j
\]

\[
= 2 \sum_{j=1}^{c} n_j \log \left( \frac{\hat{\pi}_j}{p_j} \right)^{-1} = 2 \sum_{j=1}^{c} n_j \log \left( \frac{p_j}{\hat{\pi}_j} \right)
\]

\[
= 2 \sum_{j=1}^{c} n_j \log \left( \frac{n_j}{n\hat{\pi}_j} \right) = 2 \sum_{j=1}^{c} n_j \log \left( \frac{n_j}{\hat{\mu}_j} \right)
\]
Likelihood ratio test for the multinomial Jobs data

\[ G^2 = 2 \sum_{j=1}^{c} n_j \log \left( \frac{n_j}{n \hat{\pi}_j} \right) = 2 \sum_{j=1}^{c} n_j \log \left( \frac{n_j}{\hat{\mu}_j} \right) \]

> freq = c(106,74,20); n = sum(freq)
> pihat = c(0.6,0.3,0.1); muhat = n*pihat
> G2 = 2 * sum(freq*log(freq/muhat)); G2
[1] 4.739477
Pearson’s chi-squared test
Comparing observed and expected frequencies

\[ X^2 = \sum_{j=1}^{c} \frac{(n_j - \hat{\mu}_j)^2}{\hat{\mu}_j} \]

where \( \hat{\mu}_j = n\hat{\pi}_j \)

- A large value means the observed frequencies are far from what is expected given \( H_0 \).
- A large value makes \( H_0 \) less believable.
- Distributed approximately as chi-squared for large \( n \) if \( H_0 \) is true.
Pearson Chi-squared on the jobs data

<table>
<thead>
<tr>
<th>Observed</th>
<th>106</th>
<th>74</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected</td>
<td>120</td>
<td>60</td>
<td>20</td>
</tr>
</tbody>
</table>

\[
X^2 = \sum_{j=1}^{c} \frac{(n_j - \hat{\mu}_j)^2}{\hat{\mu}_j}
\]

\[
= \frac{(106 - 120)^2}{120} + \frac{(74 - 60)^2}{60} + 0
\]

\[
= 4.9 \quad \text{(Compare } G^2 = 4.74\text{)}
\]
Two chi-squared test statistics

There are plenty more.

\[ G^2 = 2 \sum_{j=1}^{c} n_j \log \left( \frac{n_j}{\hat{\mu}_j} \right) \quad X^2 = \sum_{j=1}^{c} \frac{(n_j - \hat{\mu}_j)^2}{\hat{\mu}_j} \]

- Both compare observed to expected frequencies.
- By expected we mean estimated expected: \( \hat{\mu}_j = n\hat{\pi}_j \).
- Both equal zero when all observed frequencies equal the corresponding expected frequencies.
- Both have approximate chi-squared distributions with the same \( df \) when \( H_0 \) is true, for large \( n \).
- Values are close for large \( n \) when \( H_0 \) is true.
- Both go to infinity when \( H_0 \) is false.
- \( X^2 \) works better for smaller samples.
- \( X^2 \) is specific to multinomial data; \( G^2 \) is more general.
Rules of thumb

- Small expected frequencies can create trouble by inflating the test statistic.
- $G^2$ is okay if all (estimated) expected frequencies are at least 5.
- $X^2$ is okay if all (estimated) expected frequencies are at least 1.
One more example: Is a die fair?

Roll the die 300 times and observe these frequencies:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>39</td>
<td>54</td>
<td>44</td>
<td>44</td>
<td>47</td>
</tr>
</tbody>
</table>

- State a reasonable model for these data.
- Without any derivation, estimate the probability of rolling a 1. Your answer is a number.
- Give an approximate 95% confidence interval for the probability of rolling a 1. Your answer is a set of two numbers.
- What is the null hypothesis corresponding to the main question, in symbols?
- What is the parameter space $\mathcal{B}$?
- What is the restricted parameter space $\mathcal{B}_0$?
- What are the degrees of freedom? The answer is a number.
- What is the critical value of the test statistic at $\alpha = 0.05$? The answer is a number.
Questions continued

- What are the expected frequencies under $H_0$? Give 6 numbers.
- Carry out the likelihood ratio test.
  - What is the value of the test statistic? Your answer is a number. Show some work.
  - Do you reject $H_0$ at $\alpha = 0.05$? Answer Yes or No.
  - Using R, calculate the $p$-value.
  - Do the data provide convincing evidence against the null hypothesis?
- Carry out Pearson test. Answer the same questions you did for the likelihood ratio test.
More questions
To help with the plain language conclusion

- Does the confidence interval for $\pi_1$ allow you to reject $H_0 : \pi_1 = \frac{1}{6}$ at $\alpha = 0.05$? Answer Yes or No.
- In plain language, what do you conclude from the test corresponding to the confidence interval? (You need not actually carry out the test.)
- Is there evidence that the chances of getting 2 through 6 are unequal? This question requires its own slide.
Is there evidence that the chances of getting 2 through 6 are unequal?

- What is the null hypothesis?
- What is the restricted parameter space $B_0$? It’s convenient to make the first category the residual category.
- Write the likelihood function for the restricted model. How many free parameters are there in this model?
- Obtain the restricted MLE $\hat{\pi}$. Your final answer is a set of 6 numbers.
- Give the estimated expected frequencies ($\hat{\mu}_1, \ldots, \hat{\mu}_6$).
- Calculate the likelihood ratio test statistic. Your answer is a number.
Questions continued

- What are the degrees of freedom of the test? The answer is a number.
- What is the critical value of the test statistic at $\alpha = 0.05$? The answer is a number.
- Do you reject $H_0$ at $\alpha = 0.05$? Answer Yes or No.
- In plain language, what (if anything) do you conclude from the test.

- In plain language, what are your overall conclusion about this die?

For most statistical analyses, your final conclusions should be regarded as hypotheses that need to be tested on a new set of data.
If $X \sim N(\mu, \sigma^2)$, then

- $Z = \left( \frac{X-\mu}{\sigma} \right)^2 \sim \chi^2(1)$
- $Y = \frac{X^2}{\sigma^2}$ is said to have a non-central chi-squared distribution with degrees of freedom one and non-centrality parameter $\lambda = \frac{\mu^2}{\sigma^2}$.
- Write $Y \sim \chi^2(1, \lambda)$
Facts about the non-central chi-squared distribution
With one $df$

$$Y \sim \chi^2(1, \lambda), \text{ where } \lambda \geq 0$$

- $Pr\{Y > 0\} = 1$, of course.
- If $\lambda = 0$, the non-central chi-squared reduces to the ordinary central chi-squared.
- The distribution is “stochastically increasing” in $\lambda$, meaning that if $Y_1 \sim \chi^2(1, \lambda_1)$ and $Y_2 \sim \chi^2(1, \lambda_2)$ with $\lambda_1 > \lambda_2$, then $Pr\{Y_1 > y\} > Pr\{Y_2 > y\}$ for any $y > 0$.
- $\lim_{\lambda \to \infty} Pr\{Y > y\} = 1$
- There are efficient algorithms for calculating non-central chi-squared probabilities. R’s `pchisq` function does it.
An example  
Back to the coffee taste test

\[ Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} B(1, \pi) \]

\[ H_0 : \pi = \pi_0 = \frac{1}{2} \]

Reject \( H_0 \) if \( |Z_2| = \left| \frac{\sqrt{n}(p-\pi_0)}{\sqrt{p(1-p)}} \right| > z_{\alpha/2} \)

Suppose that in the population, 60% of consumers would prefer the new blend. If we test 100 consumers, what is the probability of obtaining results that are statistically significant?

That is, if \( \pi = 0.60 \), what is the power with \( n = 100 \)?
Recall that if $X \sim N(\mu, \sigma^2)$, then $\frac{X^2}{\sigma^2} \sim \chi^2(1, \frac{\mu^2}{\sigma^2})$.

Reject $H_0$ if

$$|Z_2| = \left| \frac{\sqrt{n}(p - \pi_0)}{\sqrt{p(1-p)}} \right| > z_{\alpha/2} \iff Z_2^2 > z_{\alpha/2}^2 = \chi^2_\alpha(1)$$

For large $n$, $X = p - \pi_0$ is approximately normal, with $\mu = \pi - \pi_0$ and $\sigma^2 = \frac{\pi(1-\pi)}{n}$. So,

$$Z_2^2 = \frac{(p - \pi_0)^2}{p(1-p)/n} \approx \frac{(p - \pi_0)^2}{\pi(1-\pi)/n} = \frac{X^2}{\sigma^2}$$

approx $\chi^2 \left( 1, n \frac{(\pi - \pi_0)^2}{\pi(1-\pi)} \right)$
We have found that

The Wald chi-squared test statistic of \( H_0 : \pi = \pi_0 \)

\[
Z_2^2 = \frac{n(p - \pi_0)^2}{p(1 - p)}
\]

has an asymptotic non-central chi-squared distribution with \( df = 1 \) and non-centrality parameter

\[
\lambda = n \frac{(\pi - \pi_0)^2}{\pi(1 - \pi)}
\]

Notice the similarity, and also that

- If \( \pi = \pi_0 \), then \( \lambda = 0 \) and \( Z_2^2 \) has a central chi-squared distribution.
- The probability of exceeding any critical value (power) can be made as large as desired by making \( \lambda \) bigger.
- There are 2 ways to make \( \lambda \) bigger.
Power calculation with R

For \( n = 100, \pi_0 = 0.50 \) and \( \pi = 0.60 \)

```r
> # Power for Wald chisquare test of H0: pi=pi0
> n=100; pi0=0.50; pi=0.60
> lambda = n * (pi-pi0)^2 / (pi*(1-pi))
> critval = qchisq(0.95,1)
> power = 1-pchisq(critval,1,lambda); power
[1] 0.5324209
```
Let $X_1, \ldots, X_n$ be independent $N(\mu_i, \sigma^2_i)$. Then

$$Y = \sum_{i=1}^{n} \frac{X_i^2}{\sigma^2_i} \sim \chi^2(n, \lambda), \text{ where } \lambda = \sum_{i=1}^{n} \frac{\mu_i^2}{\sigma^2_i}$$

- Density is a bit messy
- Reduces to central chi-squared when $\lambda = 0$.
- Generalizes to $Y \sim \chi^2(\nu, \lambda)$, where $\nu > 0$ as well as $\lambda > 0$
- Stochastically increasing in $\lambda$, meaning $Pr\{Y > y\}$ can be increased by increasing $\lambda$.
- $\lim_{\lambda \to \infty} Pr\{Y > y\} = 1$
- Probabilities are easy to calculate numerically.
Non-centrality parameters for Pearson and likelihood Ratio chi-squared tests
Re-writing the test statistics in terms of $p_j$ ...

\[ X^2 = n \sum_{j=1}^{c} \frac{(p_j - \hat{\pi}_j)^2}{\hat{\pi}_j} \]
\[ \lambda = n \sum_{j=1}^{c} \frac{[\pi_j - \pi_j(M)]^2}{\pi_j(M)} \]
\[ G^2 = 2n \sum_{j=1}^{c} p_j \log \left( \frac{p_j}{\hat{\pi}_j} \right) \]
\[ \lambda = 2n \sum_{j=1}^{c} \pi_j \log \left( \frac{\pi_j}{\pi_j(M)} \right) , \]

- Where $\hat{\pi}_j \to \pi_j(M)$ as $n \to \infty$ under the alternative hypothesis.
- That is, $\pi_j(M)$ is the large-sample target of the restricted MLE.
- The technical meaning is convergence in probability, or $\hat{\pi}_j \xrightarrow{p} \pi_j(M)$
- $\pi_j(M)$ is a function of $\pi_1, \ldots, \pi_c$. 
For the fair (?) die example

Suppose we want to test whether the die is fair, but it is not. The true \( \pi = \left( \frac{2}{13}, \frac{2}{13}, \frac{3}{13}, \frac{2}{13}, \frac{2}{13}, \frac{2}{13} \right) \). What is the power of the Pearson chi-squared test for \( n = 300 \)?

Because \( B_0 = \{(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})\} \), it’s easy to see \( \pi_j(M) = \frac{1}{6} \).

\[
\lambda = n \sum_{j=1}^{c} \frac{[\pi_j - \pi_j(M)]^2}{\pi_j(M)}
\]

\[
= 300 \left( \frac{\left( \frac{3}{13} - \frac{1}{6} \right)^2}{\frac{1}{6}} + 5 \left( \frac{\frac{2}{13} - \frac{1}{6}}{\frac{1}{6}} \right)^2 \right)
\]

\[
= 8.87574
\]
Calculate power with R

```r
> piM = 1/6 + numeric(6); piM
[1] 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667
> pi = c(2,2,3,2,2,2)/13; pi
[1] 0.1538462 0.1538462 0.2307692 0.1538462 0.1538462 0.1538462
> lambda = 300 * sum( (pi-piM)^2/piM ); lambda
[1] 8.87574
> critval = qchisq(0.95,5); critval
[1] 11.0705
> power = 1-pchisq(critval,5,lambda); power
[1] 0.6165159
```
Power of 0.62 is not too impressive. How many times would you have to roll the die to detect this degree of unfairness with probability 0.90?

```r
> piM = 1/6 + numeric(6)
> pi = c(2,2,3,2,2,2)/13
> critval = qchisq(0.95,5)
>
> n = 0; power = 0.05
> while(power < 0.90)
+ { n = n+1
+   lambda = n * sum( (pi-piM)^2/piM )
+   power = power = 1-pchisq(critval,5,lambda)
+ }
> n; power
[1] 557
[1] 0.9001972
```
Sometimes, finding $\pi(M)$ is more challenging

Recall the Jobs example, with $H_0 : \pi_1 = 2\pi_2$

$$\hat{\pi} = \left( \frac{2(n_1 + n_2)}{3n}, \frac{n_1 + n_2}{3n}, \frac{n_3}{n} \right)$$

$$= \left( \frac{2}{3} \left( \frac{n_1}{n} + \frac{n_2}{n} \right), \frac{1}{3} \left( \frac{n_1}{n} + \frac{n_2}{n} \right), \frac{n_3}{n} \right)$$

$$= \left( \frac{2}{3} (p_1 + p_2), \frac{1}{3} (p_1 + p_2), p_3 \right)$$

$$\xrightarrow{p} \left( \frac{2}{3} (\pi_1 + \pi_2), \frac{1}{3} (\pi_1 + \pi_2), \pi_3 \right) = \pi(M)$$

By the Law of Large Numbers.
This slide show was prepared by Jerry Brunner, Department of Statistics, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. The \LaTeX{} source code is available from the course website:
http://www.utstat.toronto.edu/~brunner/oldclass/312f12