The Multinomial Model*

STA 312: Fall 2022

1 Binomial Distribution

The Bernoulli Distribution

- Simple probability model: Toss a coin with $P(\text{Head}) = \pi$, one time. Let Y equal the number of heads.
- Probability (mass) function of Y:

$$P(y) = \begin{cases} \pi^y(1-\pi)^{1-y} & \text{for } y = 0 \text{ or } 1 \\ 0 & \text{Otherwise} \end{cases}$$

- An *indicator random variable* equals one if some event happens, and zero if it does not happen.
  - 1=Female, 0=Male
  - 1=Lived, 0=Died
  - 1=Passed, 0=Failed
- Indicators are usually assumed to have a Bernoulli distribution.

The Binomial Distribution

- Simple probability model: Toss a coin with $P(\text{Head}) = \pi$. Toss it n times. Let Y equal the number of heads.
- Probability (mass) function of X:

$$P(y) = \begin{cases} \binom{n}{y}\pi^y(1-\pi)^{n-y} & \text{for } y = 0, 1, \ldots, n \\ 0 & \text{Otherwise} \end{cases}$$

- The Bernoulli is a special case of the Binomial, with $n = 1$.

*See last page for copyright information.

Why does $P(y) = \binom{n}{y}\pi^y(1-\pi)^{n-y}$
For the Binomial Distribution?

Toss a coin n times with $P(\text{Head}) = \pi$, and let Y equal the number of heads. Why does $P(Y = y) = \binom{n}{y}\pi^y(1-\pi)^{n-y}$?

- The sample space is the set of all strings of n letters composed of H and T.
- By the Multiplication Principle, there are 2^n elements.
- If two different strings have y heads (and $n - y$ tails), they have the same probability.
- For example, $P\{HHTH\} = P\{THHH\} = \pi^3(1-\pi)$ by independence.
- Count the number of ways that y positions out of n can be chosen to have the symbol H.
- n choose y is $\binom{n}{y} = \frac{n!}{y!(n-y)!}$.
- So $P(Y = y) = \binom{n}{y}\pi^y(1-\pi)^{n-y}$ ■

2 Multinomial Distribution

Multinomial coefficient
For c categories

From n objects, number of ways to choose

- n_1 of type 1
- n_2 of type 2 $\quad\vdots$
- 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 $M(n, \boldsymbol{\pi})$*, where* $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_c)$

- 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

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

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

$$\begin{aligned} P(n_1, \ldots, n_{c-1}) &= \frac{n!}{n_1! \ \cdots \ n_{c-1}!(n - \sum_{j=1}^{c-1} n_j)!} \\ &\quad \times \pi_1^{n_1} \cdots \pi_{c-1}^{n_{c-1}} (1 - \sum_{j=1}^{c-1} \pi_j)^{n - \sum_{j=1}^{c-1} n_j} \end{aligned}$$

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.*

$$\sum_{n_{c-1}=0}^{n-\sum_{j=1}^{c-2} n_j} \frac{n!}{n_1! \ \ldots \ n_{c-1}!(n - \sum_{j=1}^{c-1} n_j)!} \pi_1^{n_1} \cdots \pi_{c-1}^{n_{c-1}} (1 - \sum_{j=1}^{c-1} \pi_j)^{n - \sum_{j=1}^{c-1} n_j} \times \frac{(n - \sum_{j=1}^{c-2} n_j)!}{(n - \sum_{j=1}^{c-2} n_j)!}$$

$$\begin{aligned} &= \frac{n!}{n_1! \ \ldots \ n_{c-2}!(n - \sum_{j=1}^{c-2} n_j)!} \pi_1^{n_1} \cdots \pi_{c-2}^{n_{c-2}} \\ &\quad \times \sum_{n_{c-1}=0}^{n-\sum_{j=1}^{c-2} n_j} \frac{(n - \sum_{j=1}^{c-2} n_j)!}{n_{c-1}!(n - \sum_{j=1}^{c-2} n_j - n_{c-1})!} \pi_{c-1}^{n_{c-1}} (1 - \sum_{j=1}^{c-2} \pi_j - \pi_{c-1})^{n - \sum_{j=1}^{c-2} n_j - n_{c-1}} \\ &= \frac{n!}{n_1! \ \ldots \ n_{c-2}!(n - \sum_{j=1}^{c-2} n_j)!} \pi_1^{n_1} \cdots \pi_{c-2}^{n_{c-2}} (1 - \sum_{j=1}^{c-2} \pi_j)^{n - \sum_{j=1}^{c-2} n_j}, \end{aligned}$$

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 $1, \ldots, c$ 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?

$$\binom{30}{15\ 10\ 5} 0.60^{15} 0.30^{10} 0.10^{5} = \frac{4933527642332542053801}{381469726562500000000000} \approx 0.0129$$

What is the probability that exactly 5 are unemployed?

$$\binom{30}{5\ 25} 0.10^{5} 0.90^{25} = \frac{51152385317007572507646551997}{500000000000000000000000000000} \approx 0.1023$$

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?

$$\binom{30}{15\ 10\ 5} 0.60^{15} 0.30^{10} 0.10^{5} = \frac{4933527642332542053801}{381469726562500000000000} \approx 0.0129$$

```
> dmultinom(c(15,10,5), prob=c(.6, .3, .1))
[1] 0.01293295
```

3 Estimation

Hypothetical data file

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

Case	Job	Y_1	Y_2	Y_3
1	1	1	0	0
2	3	0	0	1
3	2	0	1	0
4	1	1	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	2	0	1	0
Total		$\sum_{i=1}^n y_{i,1}$	$\sum_{i=1}^n y_{i,2}$	$\sum_{i=1}^n y_{i,3}$

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 $M(1, \boldsymbol{\pi})$, so $E(Y_{i,j}) = \pi_j$.
- Column totals n_j count the number of times each category occurs: Joint distribution is $M(n, \boldsymbol{\pi})$
- 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

Job Category	Frequency	Percent
Employed in field	106	53
Employed outside field	74	37
Unemployed	20	10
Total	200	100.0

Likelihood function for the multinomial

$$
\begin{aligned}
\ell(\boldsymbol{\pi}) &= \prod_{i=1}^{n} Pr\{Y_{i,1} = y_{i,1}, Y_{i,2} = y_{i,2}, \ldots, Y_{i,c} = y_{i,c}|\boldsymbol{\pi}\} \\
&= \prod_{i=1}^{n} \pi_1^{y_{i,1}} \pi_2^{y_{i,2}} \cdots \pi_c^{y_{i,c}} \\
&= \pi_1^{\sum_{i=1}^n y_{i,1}} \pi_2^{\sum_{i=1}^n y_{i,2}} \cdots \pi_c^{\sum_{i=1}^n y_{i,c}} \\
&= \pi_1^{n_1} \pi_2^{n_2} \cdots \pi_c^{n_c}
\end{aligned}
$$

- Product of n probability mass functions, each $M(1, \boldsymbol{\pi})$
- 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

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

- 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**

$$
\begin{aligned}
\ell(\boldsymbol{\pi}) &= \pi_1^{n_1} \cdots \pi_c^{n_c} \\
&= \pi_1^{n_1} \cdots \pi_{c-1}^{n_{c-1}} \left(1 - \sum_{j=1}^{c-1} \pi_j\right)^{n - \sum_{j=1}^{c-1} n_j} \\
\log \ell(\boldsymbol{\pi}) &= \sum_{j=1}^{c-1} n_j \log \pi_j + \left(n - \sum_{j=1}^{c-1} n_j\right) \log \left(1 - \sum_{j=1}^{c-1} \pi_j\right)
\end{aligned}
$$

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

Set all the partial derivatives to zero and solve
For π_j, j = 1 . . . , c-1

$$
\widehat{\pi}_j = \frac{n_j}{n} = p_j = \frac{\sum_{i=1}^{n} y_{i,j}}{n} = \overline{y}_j
$$

So the MLE is the sample proportion, which is also a sample mean.

In matrix terms: $\widehat{\boldsymbol{\pi}} = \mathbf{p} = \overline{\mathbf{Y}}_n$

$$
\begin{pmatrix} \widehat{\pi}_1 \\ \vdots \\ \widehat{\pi}_{c-1} \end{pmatrix} = \begin{pmatrix} p_1 \\ \vdots \\ p_{c-1} \end{pmatrix} = \begin{pmatrix} \overline{Y}_1 \\ \vdots \\ \overline{Y}_{c-1} \end{pmatrix}
$$

Remarks:

- Multivariate Law of Large Numbers says $\mathbf{p} \stackrel{p}{\rightarrow} \boldsymbol{\pi}$
- Multivariate Central Limit Theorem says that $\overline{\mathbf{Y}}_n$ is approximately multivariate normal for large n.
- Because $n_j \sim B(n, \pi_j)$, $\frac{n_j}{n} = \overline{Y}_j = p_j$ is approximately $N\left(\pi_j, \frac{\pi_j(1-\pi_j)}{n}\right)$.

- Approximate π_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 π_j
95% confidence interval for true proportion unemployed

Job Category	Frequency	Percent
Employed in field	106	53
Employed outside field	74	37
Unemployed	20	10
Total	200	100.0

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

So a confidence interval for π_3 is

$$\begin{aligned} p_3 \pm 1.96\sqrt{\frac{p_3(1-p_3)}{n}} &= 0.10 \pm 1.96\sqrt{\frac{0.10(1-0.10)}{200}} \\ &= 0.10 \pm 0.042 \\ &= (0.058, 0.142) \end{aligned}$$

4 Hypothesis tests

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

$$\begin{array}{l} Y_1, \ldots, Y_n \stackrel{i.i.d.}{\sim} F_\beta,\ \beta \in \mathcal{B}, \\ H_0 : \beta \in \mathcal{B}_0 \text{ v.s. } H_1 : \beta \in \mathcal{B}_1 = \mathcal{B} \cap \mathcal{B}_0^c \end{array}$$

Test Statistic:

$$G^2 = -2\log\left(\frac{\max_{\beta\in\mathcal{B}_0} \ell(\beta)}{\max_{\beta\in\mathcal{B}} \ell(\beta)}\right) = -2\log\left(\frac{\ell(\widehat{\beta}_0)}{\ell(\widehat{\beta})}\right)$$

What to do
And how to think about it

$$G^2 = -2\log\left(\frac{\max_{\beta\in\mathcal{B}_0} \ell(\beta)}{\max_{\beta\in\mathcal{B}} \ell(\beta)}\right) = -2\log\left(\frac{\ell(\widehat{\beta}_0)}{\ell(\widehat{\beta})}\right)$$

- Maximize the likelihood over the whole parameter space. You already did this to calculate the MLE, denoted by $\widehat{\beta}$. Evaluate the likelihood there. That's the denominator.
- Maximize the likelihood over just the parameter values where H_0 is true – that is, over $\mathcal{B}_0$. This yields a restricted MLE, denoted by $\widehat{\beta}_0$. Evaluate the likelihood there. That's the numerator.
- The numerator cannot be larger, because $\mathcal{B}_0 \subset \mathcal{B}$.
- 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.

Distribution of G^2 under H_0

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 $df = 6$

Example

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.

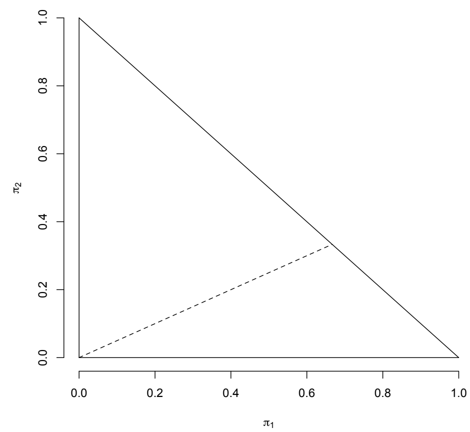
Some detailed questions

To guide us through the problem

- What is the model?
$$Y_1, \ldots, Y_n \stackrel{i.i.d.}{\sim} M(1, (\pi_1, \pi_2, \pi_3))$$
- What is the null hypothesis, in symbols?
$$H_0 : \pi_1 = 2\pi_2$$
- What are the degrees of freedom for this test?

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

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



$$
\begin{aligned}
\mathcal{B} &= \{(\pi_1, \pi_2) : 0 < \pi_1 < 1, 0 < \pi_2 < 1, \pi_1 + \pi_2 < 1\} \\
\mathcal{B}_0 &= \{(\pi_1, \pi_2) : 0 < \pi_1 < 1, 0 < \pi_2 < 1, \pi_1 = 2\pi_2\}
\end{aligned}
$$

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.

$$
\begin{aligned}
\mathbf{p} &= (\frac{n_1}{n}, \frac{n_2}{n}, \frac{n_3}{n}) \\
&= (\frac{106}{200}, \frac{74}{200}, \frac{20}{200}) \\
&= (0.53, 0.37, 0.10)
\end{aligned}
$$

Derive the restricted MLE

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

$$
\begin{aligned}
& \frac{\partial}{\partial \pi}\left(n_1 \log(2\pi) + n_2 \log \pi + n_3 \log(1-3\pi)\right) \\
= \; & \frac{n_1}{\pi} + \frac{n_2}{\pi} + \frac{n_3}{1-3\pi}(-3) \stackrel{\text{set}}{=} 0 \\
\Rightarrow \; & \frac{n_1+n_2}{\pi} = \frac{3n_3}{1-3\pi} \\
\Rightarrow \; & (n_1+n_2)(1-3\pi) = 3\pi n_3 \\
\Rightarrow \; & n_1+n_2 = 3\pi(n_1+n_2+n_3) = 3\pi n \\
\Rightarrow \; & \pi = \frac{n_1+n_2}{3n}
\end{aligned}
$$

So $\widehat{\boldsymbol{\pi}} = \left(\frac{2(n_1+n_2)}{3n}, \frac{n_1+n_2}{3n}, \frac{n_3}{n}\right)$. From now on, $\widehat{\boldsymbol{\pi}}$ means the *restricted* MLE.

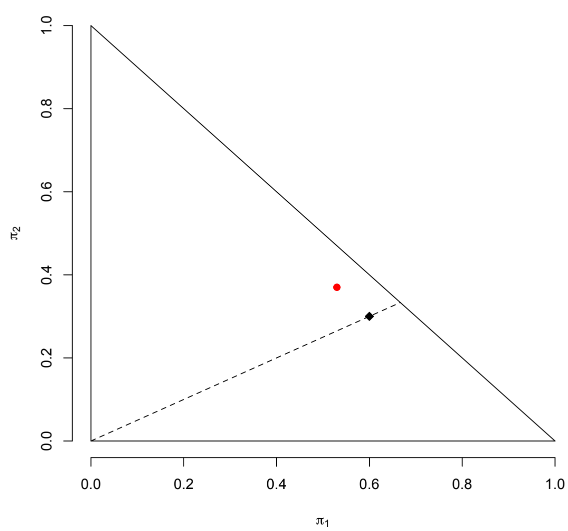
Give the restricted MLE in numeric form

The answer is a vector of 3 numbers

$$
\begin{aligned}
\widehat{\boldsymbol{\pi}} &= \left(\frac{2(n_1+n_2)}{3n}, \frac{n_1+n_2}{3n}, \frac{n_3}{n}\right) \\
&= \left(\frac{2(106+74)}{600}, \frac{106+74}{600}, \frac{20}{200}\right) \\
&= (0.6, 0.3, 0.1)
\end{aligned}
$$

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.

$$
\begin{aligned}
G^2 &= -2 \log \frac{\widehat{\pi}_1^{n_1} \widehat{\pi}_2^{n_2} p_3^{n_3}}{p_1^{n_1} p_2^{n_2} p_3^{n_3}} \\
&= -2 \left(\log \left[\frac{\widehat{\pi}_1}{p_1} \right]^{n_1} + \log \left[\frac{\widehat{\pi}_2}{p_2} \right]^{n_2} \right) \\
&= -2 \left(n_1 \log \frac{\widehat{\pi}_1}{p_1} + n_2 \log \frac{\widehat{\pi}_2}{p_2} \right) \\
&= -2 \left(106 \log \frac{0.60}{0.53} + 74 \log \frac{0.30}{0.37} \right) \\
&= 4.739
\end{aligned}
$$

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

```
> 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.

	Related	Unrelated	Unemployed
Observed	106	74	20
Expected	120	60	20
Residual	-14	14	0

Expected frequency is $E(n_j) = \mu_j = n\pi_j$. Estimated expected frequency is $\widehat{\mu}_j = n\widehat{\pi}_j$

Write G^2 in terms of observed and expected frequencies
For a general hypothesis about a multinomial

$$
\begin{aligned}
G^2 &= -2\log\left(\frac{\ell_0}{\ell_1}\right) \\
&= -2\log\left(\frac{\prod_{j=1}^{c}\widehat{\pi}_j^{n_j}}{\prod_{j=1}^{c}p_j^{n_j}}\right) \\
&= -2\log\prod_{j=1}^{k}\left(\frac{\widehat{\pi}_j}{p_j}\right)^{n_j} = 2\sum_{j=1}^{c}-\log\left(\frac{\widehat{\pi}_j}{p_j}\right)^{x_j} \\
&= 2\sum_{j=1}^{c}n_j\log\left(\frac{\widehat{\pi}_j}{p_j}\right)^{-1} = 2\sum_{j=1}^{c}n_j\log\left(\frac{p_j}{\widehat{\pi}_j}\right) \\
&= 2\sum_{j=1}^{c}n_j\log\left(\frac{n_j}{n\widehat{\pi}_j}\right) = 2\sum_{j=1}^{c}n_j\log\left(\frac{n_j}{\widehat{\mu}_j}\right)
\end{aligned}
$$

Likelihood ratio test for the multinomial
Jobs data

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

```
> 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

$$X^2 = \sum_{j=1}^{c}\frac{(n_j-\widehat{\mu}_j)^2}{\widehat{\mu}_j}$$

where $\widehat{\mu}_j = n\widehat{\pi}_j$

- 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

Observed	106	74	20
Expected	120	60	20

$$
\begin{aligned}
X^2 &= \sum_{j=1}^{c} \frac{(n_j - \widehat{\mu}_j)^2}{\widehat{\mu}_j} \\
&= \frac{(106-120)^2}{120} + \frac{(74-60)^2}{60} + 0 \\
&= 4.9 \quad \text{(Compare } G^2 = 4.74\text{)}
\end{aligned}
$$

Two chi-squared test statistics

There are plenty more.

$$
G^2 = 2\sum_{j=1}^{c} n_j \log\left(\frac{n_j}{\widehat{\mu}_j}\right) \qquad\qquad X^2 = \sum_{j=1}^{c} \frac{(n_j - \widehat{\mu}_j)^2}{\widehat{\mu}_j}
$$

- Both compare observed to expected frequencies.
- By expected we mean *estimated* expected: $\widehat{\mu}_j = n\widehat{\pi}_j$.
- 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:

1	2	3	4	5	6
72	39	54	44	44	47

- 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 π_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 $\mathcal{B}_0$? 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 $\widehat{\boldsymbol{\pi}}$. Your final answer is a set of 6 numbers.
- Give the estimated expected frequencies $(\widehat{\mu}_1, \ldots, \widehat{\mu}_6)$.
- 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.