χ^2 **test for multinomial data** (part 1)

Lecture 16b (STAT 24400 F24)

The multinomial distribution (definition)

The multinomial distribution is a generalization of the binomial:

- We have $m \geq 2$ categories $\longleftarrow$ for a binomial, $m = 2$ — success & failure
- Each category i has probability $p_i \geq 0$, with $p_1 + \cdots + p_m = 1$ $\longleftarrow$ for a binomial, the prob's are written as p & $1 - p$
- Draw n observations, which are $\perp\!\!\!\perp$ and each obey these probabilities, & count X_i = total # falling into category i, for $i = 1, \ldots, m$

The multinomial distribution (one-way example)

Example:

Probabilities

Category 1	Category 2	Category 3	Category 4	Category 5	Category 6
p_1	p_2	p_3	p_4	p_5	p_6

Observed counts

Category 1	Category 2	Category 3	Category 4	Category 5	Category 6
X_1	X_2	X_3	X_4	X_5	X_6

Note: X_i's are counts, not individual observations, $X_1 + \cdots + X_m = n$ (here $m = 6$).

The multinomial distribution (two-way example)

In the $m = 6$ example, the data may have a two-way structure:

Probabilities

	Col. 1	Col. 2	Col. 3
Row 1	p_1	p_2	p_3
Row 2	p_4	p_5	p_6

Observed counts

	Col. 1	Col. 2	Col. 3
Row 1	X_1	X_2	X_3
Row 2	X_4	X_5	X_6

It may be convenient to use different labeling to reflect the structure, e.g.:

	Col. 1	Col. 2	Col. 3
Row 1	p_{11}	p_{12}	p_{13}
Row 2	p_{21}	p_{22}	p_{23}

	Col. 1	Col. 2	Col. 3
Row 1	X_{11}	X_{12}	X_{13}
Row 2	X_{21}	X_{22}	X_{23}

The multinomial distribution (labeling)

Example:

Probabilities

Category 1	Category 2	Category 3	Category 4
p_1	p_2	p_3	p_4

Observed counts

Category 1	...
X_1	...

It may be convenient to use different labeling for the categories, e.g.:

0 hits	1 hits	2 hits	3 hits
p_0	p_1	p_2	p_3

$\leftarrow$ # of bullseye hits with 3 darts thrown

or non-numerical labelling, e.g.:

Blood type O	Blood type A	Blood type B	Blood type AB
p_O	p_A	p_B	p_{AB}

The multinomial distribution (notations)

- Careful with the notation:

  X_i is *not* the ith data point. It's the total # in category i

  $\rightsquigarrow$ the X_i's are not $\perp\!\!\!\perp$ (must satisfy $X_1 + \cdots + X_m = n$)

- Other common notation:

  O_i instead of X_i (O stands for "observed"), or
  N_i instead of X_i

Hypotheses for multinomial data (examples)

Many common questions arise with multinomial data,
which can be framed as hypotheses about parameters $(p_1, \ldots, p_m)$.

Some typical questions for a two-way table:

	Undergrads	Grad students	Faculty
Prefer morning	p_{11}	p_{12}	p_{13}
Prefer afternoon	p_{21}	p_{22}	p_{23}

- Are time preferences the same for each subpopulation?

  $\rightsquigarrow$ test if $\dfrac{p_{11}}{p_{21}} = \dfrac{p_{12}}{p_{22}} = \dfrac{p_{13}}{p_{23}}$

- Are the two choices equally popular for faculty?

  $\rightsquigarrow$ test if $p_{13} = p_{23}$ (Caution: This is not testing $p_{13} = p_{23} = 0.5$.)

Hypotheses for multinomial data (typical questions)

Some typical questions for a one-way tables.

- Example

Blood type O	Blood type A	Blood type B	Blood type AB
p_O	p_A	p_B	p_{AB}

  - Are all blood types equally likely?
    $\rightsquigarrow$ test if $p_O = p_A = p_B = p_{AB}$
  - Is it true that type A is twice as common as type AB?
    $\rightsquigarrow$ test if $p_A = 2p_{AB}$

- Example

0 hits	1 hits	2 hits	3 hits
p_0	p_1	p_2	p_3

  - Is the data consistent with a Binomial distribution?
    $\rightsquigarrow$ Test if, for some $p \in (0, 1)$,
    $p_i = \binom{3}{i} p^i (1-p)^{3-i}$ for each $i = 0, 1, 2, 3$.

Hypotheses for multinomial data (General setting)

Formulation in a general setting:

Define the *probability simplex* (a subset of $\mathbb{R}^m$):

$$\Delta_m = \{(p_1, \ldots, p_m) \in \mathbb{R}^m : p_i \geq 0 \text{ for all } i,\ p_1 + \cdots + p_m = 1\}$$

We will learn to run tests of the form

$$H_0 : (p_1, \ldots, p_m) \in \Omega_0 \quad \text{vs} \quad H_1 : (p_1, \ldots, p_m) \in \underbrace{\Delta_m \backslash \Omega_0}_{\text{this means: in } \Delta_m \text{ but not in } \Omega_0}$$

where Ω_0 is defined by one or more equality constraints.

Examples

- Testing if $p_1 = \cdots = p_m \quad \rightsquigarrow \quad \Omega_0 = \{(p_1, \ldots, p_m) \in \Delta_m : p_1 = \cdots = p_m\}$
- Testing if $p_1 = 2p_2 \quad \rightsquigarrow \quad \Omega_0 = \{(p_1, \ldots, p_m) \in \Delta_m : p_1 = 2p_2\}$

Hypotheses for multinomial data (other cases)

Not all questions can be framed with a test of this form, e.g.,

- Test inequalities, e.g., test $H_0 : p_1 \leq p_2$ vs $H_1 : p_1 > p_2$
- Test $H_0 : p_1 = p_2 = p_3 = p_4$ vs $H_1 : p_1 = p_2 \neq p_3 = p_4$

(These types of tests are not covered in this course)

Calculating the MLE for multinomial data (without constraints)

- Without constraints, i.e., parameter space $(p_1, \ldots, p_m) \in \Delta_m$:

  $$\text{Likelihood} = L(p_1, \cdots, p_m | X_1, \cdots, X_m) = \frac{n!}{X_1! \cdot \ldots \cdot X_m!}\, p_1^{X_1} \cdot \ldots \cdot p_m^{X_m}$$

  The MLE (without constraints on $\Delta_m = \Omega_0 \cup \Omega_1$)

  $$(\hat{p}_1, \cdots, \hat{p}_m) = \underset{(p_1, \ldots, p_m) \in \Delta_m}{\text{argmax}} \frac{n!}{X_1! \cdot \ldots \cdot X_m!}\, p_1^{X_1} \cdot \ldots \cdot p_m^{X_m}$$

  maximizes the likelihood at

  $$\hat{p}_1 = \frac{X_1}{n}\ ,\ \ldots\ ,\ \hat{p}_m = \frac{X_m}{n}$$

  i.e., for each i, $\hat{p}_i$ is the observed fraction of the sample (of size n) that falls into category i

  Note This is consistent with the binomial case of $m = 2$.

Calculating the MLE for multinomial data (with constraints)

- With constraints, i.e., parameter space $(p_1, \ldots, p_m) \in \Omega_0$ under H_0,

  The MLE
  $$\underset{(p_1, \ldots, p_m) \in \Omega_0}{\text{argmax}} \frac{n!}{X_1! \cdot \ldots \cdot X_m!}\, p_1^{X_1} \cdot \ldots \cdot p_m^{X_m}$$

  The derivation of the MLE would depend on the specific structure of Ω_0.

  General strategy:

  - Find the dimension of Ω_0 (how many free parameters?)
  - Rewrite $(p_1, \ldots, p_m)$ as a function of the free parameters.
  - Set each derivative of the log-likelihood to zero, then solve.
  - Translate back to the original model parameters (the p_i's).

Example

The data:

Category 1	Category 2	Category 3	Category 4	Category 5	Total
X_1	X_2	X_3	X_4	X_5	n

The multinomial model:

Category 1	Category 2	Category 3	Category 4	Category 5
p_1	p_2	p_3	p_4	p_5

Suppose we want to test H_0: $p_1 = p_2$ & $p_3 = p_4$

Reparameterize:

$$\begin{cases} p_1 = p_2 = p \\ p_3 = p_4 = q \\ p_5 = 1 - 2p - 2q \end{cases} \qquad \rightsquigarrow \qquad \text{dimension}(\Omega_0) = 2$$

Example (cont.)

$$\text{Likelihood} = \frac{n!}{X_1! \cdot \ldots \cdot X_5!}\ p_1^{X_1} p_2^{X_2} p_3^{X_3} p_4^{X_4} p_5^{X_5}$$

$$\text{(under } H_0 \to\text{)} \qquad = \frac{n!}{X_1! \cdot \ldots \cdot X_5!}\ p^{X_1} p^{X_2} q^{X_3} q^{X_4} (1 - 2p - 2q)^{X_5}$$

$$\text{Log lik.} = \begin{pmatrix} \text{terms that don't} \\ \text{depend on } p \text{ or } q \end{pmatrix} + (X_1 + X_2)\log(p) + (X_3 + X_4)\log(q) + X_5 \log(1 - 2p - 2q)$$

Taking derivatives w.r.t. the free parameters p and q,

$$\frac{\partial}{\partial p}(\text{Log lik.}) = \frac{X_1 + X_2}{p} - \frac{2X_5}{1 - 2p - 2q}$$

$$\frac{\partial}{\partial q}(\text{Log lik.}) = \frac{X_3 + X_4}{q} - \frac{2X_5}{1 - 2p - 2q}$$

Example (cont.)

Set both derivatives to zero, obtain 2 equations with 2 unknowns p and q.
Solve (exercise), also use $X_1 + X_2 + X_3 + X_4 + X_5 = n$,

$$\rightsquigarrow \quad \hat{p} = \frac{X_1 + X_2}{2n}, \qquad \hat{q} = \frac{X_3 + X_4}{2n}$$

Translate back to the original model parameters, the MLE for Ω_0 under H_0:

$$\hat{p}_1 = \frac{X_1 + X_2}{2n}, \qquad \hat{p}_2 = \frac{X_1 + X_2}{2n}$$

$$\hat{p}_3 = \frac{X_3 + X_4}{2n}, \qquad \hat{p}_4 = \frac{X_3 + X_4}{2n}$$

$$\hat{p}_5 = \frac{X_5}{n} \qquad (\leftarrow \text{ use } \textstyle\sum_j \hat{p}_j = 1, \sum_i X_i = n)$$

Generalized LRT

To run a generalized LRT, we calculate

$$\Lambda = \frac{\max_{(p_1,\ldots,p_m)\in\Omega_0} \frac{n!}{X_1!\cdot\ldots\cdot X_m!}\ p_1^{X_1} \cdot \ldots \cdot p_m^{X_m}}{\max_{(p_1,\ldots,p_m)\in\Delta_m} \frac{n!}{X_1!\cdot\ldots\cdot X_m!}\ p_1^{X_1} \cdot \ldots \cdot p_m^{X_m}} \quad \begin{matrix} \leftarrow \text{ best likelihood under } H_0 \\ \leftarrow \text{ best likelihood under } H_0 \text{ or } H_1 \end{matrix}$$

$$= \frac{\prod_{i=1}^m \hat{p}_i^{X_i}}{\prod_{i=1}^m \left(\frac{X_i}{n}\right)^{X_i}} \quad \begin{matrix} \leftarrow (\hat{p}_1, \ldots, \hat{p}_m) \text{ is the MLE in } \Omega_0 \\ \leftarrow (\frac{X_1}{n}, \ldots, \frac{X_m}{n}) \text{ is the MLE in } \Delta_m \end{matrix}$$

To test H_0 / calculate p-value —
compare $-2\log(\Lambda)$ to $\chi^2_{d-d_0}$ distrib. (its approximate null distrib.)

Calculating the degrees of freedom:

- d_0 = dimension of Ω_0 (how many free parameters?)
- d = dimension of $\Delta_m = m - 1$ (not m, since $p_1 + \cdots + p_m = 1$)

Pearson's χ^2 test

A different test — Pearson's χ^2 test

- For each cell $i = 1, \dots, m$, calculate the expected count, according to the MLE for H_0:
  $$\text{Expected count in cell } i \; = n \cdot \hat{p}_i$$
- Calculate the discrepancy between observed & expected count in each cell, and add it up:
  $$X^2 = \sum_{i=1}^{m} \frac{\left(X_i - n \cdot \hat{p}_i\right)^2}{n \cdot \hat{p}_i}$$
  $\leftarrow$ squared because difference may be positive or negative
  $\leftarrow$ a large difference is more unusual if expected count is low
- To test H_0 / calculate p-value —
  compare X^2 to $\chi^2_{d-d_0}$ distrib. (its approximate null distrib.)

Pearson's χ^2 test

The statistic is sometimes written as

observed count (i.e., X_i)

expected count (i.e., $n \cdot \hat{p}_i$)

$$X^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i}$$

Example (cont.)

The observed counts (X_i's):

Category 1	Category 2	Category 3	Category 4	Category 5	Total
10	15	30	20	50	125

We want to test H_0: $p_1 = p_2$ & $p_3 = p_4$

- d_0 = dimension of $\Omega_0 = 2$ (we reparameterized with p & q)
- d = dimension of $\Delta_5 = 5 - 1 = 4$

Example (cont.)

Plug in the observed data into the MLE:

- MLE under H_0:
  $$\hat{p}_1 = \hat{p}_2 = \frac{10+15}{2 \cdot 125} = 0.1, \; \hat{p}_3 = \hat{p}_4 = \frac{30+20}{2 \cdot 125} = 0.2, \; \hat{p}_5 = \frac{50}{125} = 0.4$$
- MLE under $H_0 \cup H_1$:

$$\hat{p}_1 = \frac{10}{125} = 0.08, \; \hat{p}_2 = \frac{15}{125} = 0.12, \; \hat{p}_3 = \frac{30}{125} = 0.24, \; \hat{p}_4 = \frac{20}{125} = 0.16, \; \hat{p}_5 = \frac{50}{125} = 0.4$$

Example (cont.)

Generalized likelihood ratio test:

$$\Lambda = \frac{\max_{(p_1,\dots,p_m)\in\Omega_0} \frac{n!}{X_1!\cdot\ldots\cdot X_m!}\, p_1^{X_1}\cdot\ldots\cdot p_m^{X_m}}{\max_{(p_1,\dots,p_m)\in\Delta_m} \frac{n!}{X_1!\cdot\ldots\cdot X_m!}\, p_1^{X_1}\cdot\ldots\cdot p_m^{X_m}}$$

$$= \frac{\frac{125!}{10!15!30!20!50!}\cdot 0.1^{10}\cdot 0.1^{15}\cdot 0.2^{30}\cdot 0.2^{20}\cdot 0.4^{50}}{\frac{125!}{10!15!30!20!50!}\cdot 0.08^{10}\cdot 0.12^{15}\cdot 0.24^{30}\cdot 0.16^{20}\cdot 0.4^{50}} = 0.22087$$

$$-2\log(\Lambda) = 3.0203$$

$$\text{p-value } = \mathbb{P}(\chi^2_{df=4-2} \geq 3.0203) = 1 - F_{\chi^2_2}(3.0203) = 0.2209$$

$\Rightarrow$ Do not reject H_0: $p_1 = p_2$ & $p_3 = p_4$

Example (cont.)

Pearson's χ^2 test:

$$X^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{m} \frac{\left(X_i - n\cdot\hat{p}_i\right)^2}{n\cdot\hat{p}_i}$$

$$= \frac{(10-125\cdot 0.1)^2}{125\cdot 0.1} + \frac{(15-125\cdot 0.1)^2}{125\cdot 0.1} + \frac{(30-125\cdot 0.2)^2}{125\cdot 0.2} + \frac{(20-125\cdot 0.2)^2}{125\cdot 0.2} + \frac{(50-125\cdot 0.4)^2}{125\cdot 0.4} = 3$$

$$\text{p-value } = 1 - F_{\chi^2_2}(3) = 0.2231$$

$\Rightarrow$ same conclusion as the generalized LRT.

Comparing the two tests

- Asymptotically, the two tests are equivalent, because $X^2 \approx -2\log(\Lambda)$
- The approx. null distrib. is $\chi^2_{d-d_0}$ for both tests
- For a finite sample size we may get somewhat different answers
  (i.e., $X^2 \neq -2\log(\Lambda)$)
- And, we may get somewhat different Type I errors
  (i.e., null distrib.'s are not exactly $\chi^2_{d-d_0}$, and may not be the same)
- More common to use Pearson's χ^2 test
- It is not valid to run both tests & choose the better p-value—
  this is an instance of multiple testing

Appendix - multinomial coefficents

The coefficient of the multinomial probability

$$\frac{n!}{X_1!\cdot\ldots\cdot X_m!}\, p_1^{X_1}\cdot\ldots\cdot p_m^{X_m} = \binom{n}{X_1!,\ \cdots\ ,\ X_m!}\, p_1^{X_1}\cdot\ldots\cdot p_m^{X_m}$$

is the # of ways to put n objects into m categories ("sorting into groups", lecture 1a)

Derivations

- Case $m = 2, n_2 = n - n_1$ (binomial): $\binom{n}{n_1,\, n-n_1} = \frac{n}{n_1!(n-n_1)!} = \binom{n}{n_1}$

  which is putting n items into two categories of sizes n_1 and $n_2 = n - n_1$, respectively.
- Case $m = 3$: first splitting n items into subgroups of n_1 and $n - n_1$,
  then further splitting $n - n_1$ into two groups of n_2 and $n_3 = n - n_1 - n_2$.
  Thus the number of ways of putting n objects into groups of sizes n_1, n_2, n_3 is

  $$\binom{n}{n_1}\binom{n-n_1}{n_2} = \frac{n!}{n_1!(n-n_1)!}\cdot\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} = \frac{n!}{n_1!n_2!(n-n_1-n_2)!} = \binom{n}{n_1,\, n_2,\, n_3}$$
- and so on. More formally, mathematical induction can be used to prove for general m.