Large Sample Hypothesis Testing

Tiandong Wang

Department of Statistics
Texas A&M University

April 19, 2021

Asymptotic Properties of LRT

Asymptotic distribution of $\lambda(\mathbf{X})$ for a simple H_0:

Theorem

For testing $H_0 : \theta = \theta_0$ v.s. $H_1 : \theta \neq \theta_0$, suppose $X_1, \cdots, X_n$ i.i.d. $f(x;\theta)$ (satisfying some regularity conditions). Then under H_0, as $n \to \infty$,

$$-2\log\lambda(\boldsymbol{X}) \xrightarrow{d} \chi_1^2.$$

Hence, reject H_0 iff

$$-2\log\lambda(\mathbf{X}) \geq \chi^2_{1,\alpha}.$$

Example

For $X_1, \cdots, X_n$ iid Poisson(λ), test

$$H_0 : \lambda = \lambda_0 \qquad \text{v.s.} \qquad H_1 : \lambda \neq \lambda_0.$$

Multivariate Case: Wilks' Theorem

Assume that the joint distribution of $X_1, \ldots, X_n$ depends on p unknown parameters and that, under H_0, the joint distribution depends on p_0 unknown parameters. Let $\nu = p - p_0$. Then, under some regularity conditions, when the null hypothesis is true,

$$-2\log\lambda(\mathbf{X}) \xrightarrow{d} \chi^2_\nu,$$

as $n \to \infty$.

- Thus, for large n, the rejection region for a test with approximate significance level α is

$$\{\mathbf{y} : -2\log(\lambda(\mathbf{x})) \geq \chi^2_{\nu,\alpha}\}$$

Remark:

- Wilks' theorem allows us to approximate the "null distribution" of $\lambda(\mathbf{X})$.
- The limiting null distribution of $\lambda(\mathbf{X})$ does not depend on which element of Θ_0 is the true parameter value.
- Asymptotic size α test:

$$\lim_{n\to\infty} \mathbb{P}_\theta(\text{Reject } H_0) = \alpha, \qquad \text{for each } \theta \in \Theta_0.$$

Example: Suppose that $Y_i, i = 1, \ldots, n$, are iid random variables with the probability mass function given by

$$\mathbb{P}(Y = y) = \begin{cases} \theta_j, & \text{if } y = j, j = 1, 2, 3; \\ 0, & \text{otherwise,} \end{cases}$$

where θ_j are unknown parameters s.t. $\sum_j \theta_j = 1$, $\theta_j \geq 0$. Test:

$$H_0 : \theta_1 = \theta_2 = \theta_3 \quad \text{v.s.} \quad H_a : \text{at least one of them is different.}$$

Contingency Tables

A **two-way table** presents categorical data by

- Counting the number of observations that fall into each group for two variables:
  One divided into rows and the other divided into columns.

Example:
Students in grades 4-6 were asked whether good grades, athletic ability, or popularity was most important to them. A two-way table separating the students by grade and by choice of most important factor is shown below:

	Grade			
Goals	4	5	6	Total
Grades	49	50	69	168
Popular	24	36	38	98
Sports	19	22	28	69
Total	92	108	135	335

Goal: Testing the association between the row and column variables in a two-way table.

$$H_0 : \text{No association between the variables.} \quad v.s.$$

$$H_a : \text{Some association does exist.}$$

Solution:

Chi-square Test!

Chi-square Test

Based on a test statistic that measures:

- The divergence of the observed data from the values that would be expected under the H_0 of no association.
- Two-way table: The expected value for each cell in a two-way table is equal to

$$\frac{\text{Row Total} \times \text{Column Total}}{\text{Total number of observations included in the table}}.$$

- Chi-sq test stat:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(\text{Observed}_{ij} - \text{Expected}_{ij})^2}{\text{Expected}_{ij}}.$$

- The test stat is chi-square with $(r-1)(c-1)$ degrees of freedom, where $r = \#$ of rows, $c = \#$ of columns.

Example ctd:

Expected Values

Goals	Grade 4	5	6
Grades	46.1	54.2	67.7
Popular	26.9	31.6	39.5
Sports	18.9	22.2	27.8

- The chi-square statistic

$$\chi^2 = 1.51 \qquad \sim \chi^2_{(3-1)(3-1)}.$$

- So

$$\mathbb{P}(\chi^2_4 \geq 1.51) \approx 0.825,$$

and there is no association between the choice of most important factor and the grade of the student – the difference between observed and expected values under the null hypothesis is negligible.

Remark:

- The usual test for association in contingency tables is a LRT
- Its asymptotic distribution is an example of Wilks' theorem.

For $\sum_{i=1}^{r} a_i = \sum_{j=1}^{c} b_j = 1$, test

$H_0 : \theta_{ij} = a_i b_j \qquad \text{v.s.} \qquad H_1 : \theta_{ij} \neq a_i b_j$ for at least one pair of (i, j).

Likelihood:

$$L(\boldsymbol{\theta}; \mathbf{x}) = C \prod_{i=1}^{r} \prod_{j=1}^{c} \theta_{ij}^{x_{ij}},$$

where the coefficient C is the number of ways that a total of N subjects can be divided in rc groups with x_{ij} in the ij-th group.

Find unrestricted MLE, i.e.

$$\text{Maximize} \quad \log L(\boldsymbol{\theta}; \mathbf{x}) = \log C + \sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij} \log \theta_{ij}$$

$$\text{s.t.} \quad \sum_{i=1}^{r} \sum_{j=1}^{c} \theta_{ij} = 1.$$

Therefore,

$$\widehat{\theta}_{ij} = \frac{X_{ij}}{N}.$$

Find restricted MLE, i.e.

$$\text{Maximize} \quad \log L(\boldsymbol{\theta}; \mathbf{x}) = \log C + \sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij} \log \theta_{ij}$$

$$\text{s.t.} \quad \theta_{ij} = a_i b_j$$

$$\sum_{i=1}^{r} a_i = 1$$

$$\sum_{j=1}^{c} b_j = 1.$$

Therefore,

$$\widehat{\theta}_{ij}^{0} = \frac{R_i}{N} \frac{C_j}{N} =: \frac{E_{ij}}{N},$$

where R_i and C_j are the sum of i-th row and j-th column, respectively.

- By Wilks' theorem: as $N \to \infty$,

$$-2\log\lambda(\mathbf{X}) = 2\sum_{i=1}^{r}\sum_{j=1}^{c} X_{ij}\log(X_{ij}/E_{ij}) \xrightarrow{d} \chi^2_\nu,$$

with $\nu = (r-1)(c-1)$. Why?

- By Taylor expansion,

$$x_{ij}\log\frac{x_{ij}}{e_{ij}} \approx (x_{ij}-e_{ij}) + \frac{1}{2}\frac{(x_{ij}-e_{ij})^2}{e_{ij}}.$$

- Since $\sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}-e_{ij}) = 0$,

$$-2\log\lambda(\mathbf{X}) \approx \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(X_{ij}-E_{ij})^2}{E_{ij}}.$$