Asymptotic Tests

Theorem. ***Wilks' Theorem*** *Under regularity conditions, the null distribution of* $2 \log \Lambda$ *tends to chi squared distribution with d.f. equal to* $\dim \Theta - \dim \Theta_0$ *(the dimensinon of free parameters) as the sample size goes to infinity. Let* $X_1, \cdots, X_n \sim f_\theta$ *where* $\theta = (\theta_1, \cdots, \theta_p) \in \Theta$ *is a vector of parameters, and we wish to test the null hypothesis*

$$\mathcal{H}_0 : \theta_1 = \theta_1^0, \theta_2 = \theta_2^0, \cdot, \theta_r = \theta_r^0 (1 \le r \le p)$$

against unrestricted alternative for GLRT. Let $\Lambda(\underline{X})$ *be the resultant GLR statitic. Then under some regularity condition to those required for asymptotic normality of the MLE,*

$$2 \log \Lambda(\underline{X}) \xrightarrow[\mathcal{H}_0]{\mathcal{D}} \chi_r^2$$

where r *is the number of paramters constrained by* $\mathcal{H}_0$. *The proof consists of using asymptotic property of MLE and Taylor expansion over the parameter vector* θ

Remark. In case where the exact distribution of likelihood ratio corresponding to specific hypotheses is difficult to determine, we use Wilks' theorem. It states that for large sample, test statistic $2 \log \Lambda$ for a nested model will be asymptotically chi-squared distributed with d.f. equal to difference in dimensionality of Θ and Θ_0, when $\mathcal{H}_0$ holds true.

Example. Test equality of Poisson means Let $X_1, \cdots, X_m \overset{i.i.d.}{\sim} Pois(\lambda_X)$ and $Y_1, \cdots, Y_n \overset{i.i.d.}{\sim} Pois(\lambda_Y)$ be two independent samples, suppose we want to test $\mathcal{H}_0 : \lambda_X = \lambda_Y$ vs. $\mathcal{H}_1 : \lambda_X \ne \lambda_Y$ First we find unrestricted MLE

$$\hat{\lambda}_X = \overline{X} \qquad \hat{\lambda}_Y = \overline{Y}$$

Plug into likelihood function

$$\mathcal{L}(\hat{\lambda}_X, \hat{\lambda}_Y) = e^{-m\overline{X}} \frac{(\overline{X})^{m\overline{X}}}{\prod_{i=1}^m X_i!} \times e^{-n\overline{Y}} \frac{(\overline{Y})^{n\overline{Y}}}{\prod_{i=1}^n Y_i!}$$

For unrestricted likelihood under $\mathcal{H}_0 : \lambda_X = \lambda_Y = \lambda_0$ we find restricted MLE

$$\hat{\lambda}_0 = \frac{m\overline{X} + n\overline{Y}}{m + n}$$

Plug into likelihood function

$$\mathcal{L}(\hat{\lambda}_0) = e^{-m\overline{X} - n\overline{Y}} \frac{\left(\frac{m\overline{X} + n\overline{Y}}{m+n}\right)^{m\overline{X} + n\overline{Y}}}{\prod_{i=1}^m X_i! \prod_{i=1}^n Y_i!}$$

We then find GLRT and use Wilks' theorem to find rejection region

$$\mathcal{C} = \left\{ 2 \left(m\overline{X} \log \overline{Y} + n\overline{Y} \log \overline{Y} - (m\overline{X} + n\overline{Y}) \log \left(\frac{m\overline{X} + n\overline{Y}}{m+n}\right) \right) \ge \chi_{1,1-\alpha}^2 \right\}$$

Definition. *GLRT by parametric bootstrap* *We evaluate test directly by taking large number of simulations*

1. *Sample data under $\mathcal{H}_0$: two random Poisson samples with same λ with size m and n*

2. *Evaluate test statistic $2\log\Lambda$ at simulated data*

3. *repeat for very large number of times*

4. *empirical p-value given by*

$$p - value = \frac{\# \text{ of samples for which } 2\log\Lambda \geq 4.58 \text{(the observed data)}}{N}$$

Note p-value converges in probability to true p-value as $N \to \infty$ regardless size of sizes of data and regularity assumptions required for Wilk's Theorem; hence it is awesome for small sample sizes

Example. Motivation Dice game. If dice is fair we have $X \sim Unif(1, \cdots, 6)$ where X is the outcome of rolling die once. Denote $p_j = P(X = j), j = 1, \cdots, 6$ we test

$$\begin{cases} \mathcal{H}_0 : p_1 = \cdots = p_6 = \dfrac{1}{6} \\ \mathcal{H}_1 : otherwise \end{cases}$$

at level α

Definition. *Goodness of Fit Test* *Suppose $X_1, \cdots, X_n$ is a random sample from a discrete distribution with k possible values $s_1, \cdots, s_k$, with corresponding probabilities $p_1, \cdots, p_k$ i.e. $p_j = \mathbb{P}(X = s_j)$. Now we dnote $O_j = \#\{i : X_i = s_j\}$ (the observed jth cell count), then we have likelihood*

$$\mathcal{L}(p_1, \cdots, p_k) = \prod_{i=1}^{k} p_i^{O_i} \qquad l(p_1, \cdots, p_k) = \sum_{i=1}^{n} O_i \log p_i$$

We try to find MLE of $(p_1, \cdots, p_k)$ using constrained optimization (i.e. sum of p_i equates to 1)

$$\max_{p_1, \cdots, p_k} \left\{ \sum_{i=1}^{n} O_i \log p_i \quad s.t. \quad \sum_{i=1}^{k} p_i = 1 \right\}$$

By method of Lagrange Multipliers

1. *Lagrangian function*

$$\mathfrak{L}(p_1, \cdots, p_k, \lambda) = \sum_{i=1}^{n} O_i \log p_i + \lambda(\sum_{i=1}^{k} p_i - 1)$$

2. *Take its gradient and solve for zero*

$$\frac{\partial \mathfrak{L}}{\partial p_j} = \frac{O_j}{p_j} + \lambda = 0 \quad \Rightarrow \quad \hat{p}_j = -\frac{O_j}{\lambda}, j = 1, \cdots, k$$

$$\frac{\partial \mathfrak{L}}{\partial \lambda} = \sum_{i=1}^{k} p - i - 1 = 0 \qquad \sum_{j=1}^{k} \hat{}_j = -\frac{1}{\lambda} \sum_{j=1}^{k} O_j = 1 \quad \Rightarrow \quad \hat{p}_j = \frac{O_j}{n}$$

3. *Now we want to test* $\mathcal{H}_0 : p_1 = p_1^0, \cdots, p_k = p_k^0$ *vs unrestricted alternative. Find the generlized likelihood ratio*

$$\Lambda = \prod_{i=1}^{k} \left(\frac{O_i}{np_i^0} \right)^{O_i} = \prod_{i=1}^{k} \left(\frac{O_i}{\mathbb{E}_i} \right)^{O_i}$$

where $\mathbb{E}_j := np_i^0$ *is the expected jth cell count under* $\mathcal{H}_0$

4. *Use Wilk's Theorem*

$$2 \log \Lambda = 2 \sum_{i=1}^{k} O_i \log \left(\frac{O_i}{\mathbb{E}_i} \right) \xrightarrow[\mathcal{H}_0]{\mathcal{D}} \chi_{k-1}^2$$

Take 2nd order Taylor expansion of $\log O_i$ *and simplify the test statistic*

$$2 \sum_{i=1}^{k} O_i \log \left(\frac{O_i}{\mathbb{E}_i} \right) = \sum_{i=1}^{k} \frac{(O_i - \mathbb{E}_i)^2}{\mathbb{E}_i} - \sum_{i=1}^{k} \frac{(O_i - \mathbb{E}_i)^3}{\mathbb{E}_i^2}$$

where the second term is negligible in probability as count of each cell goes to infinity. (this is a necessary condition for Pearson χ^2 *test to work) We have*

$$\mathcal{X}^2 := \sum_{i=1}^{k} \frac{(O_i - \mathbb{E}_i)^2}{\mathbb{E}_i} \xrightarrow[\mathcal{H}_0]{\mathcal{D}} \chi_{k-1}^2$$

5. **Pearson's** χ^2 **test of goodness of fit** *at level* α *is given by*

$$\mathcal{C} = \{\mathcal{X}^2 \geq \chi_{k-1,1-\alpha}^2\}$$

Definition. **Pearson's** χ^2 **Test** *A statistical test applied to set of categorical data to evaluate how likely it is that any observed difference between the sets arose by chance, suitable for unpaired data from large samples. It tests a null hypothesis stating that frequency distribution of certain evenets observed in a sample is consistent with a particular theoretical distribution. THere are 3 types of comparisons*

1. **test of goodness of fit** *establishes whether an observed freuency distribution differs from a theoretical distribution*

2. ***test of homogeneity*** *compares distributino of counts for two or more groups using the same categorical variable.*

3. ***test of independence*** *assess whether unpaired observations on 2 variables are independent of each other.*

The procedure is as follows

1. *Calculate* $\mathcal{X}^2 := \sum_{i=1}^{k} \frac{(O_i - \mathbb{E}_i)^2}{\mathbb{E}_i} =,$ *which resembles a normalized sum of squared deviation between observed and theoretical frequencies*

2. *Determine degrees of freedom, df, of that statistic.*

   - *For a **test of goodness-of-fit**, this is number of categories reduced by number of parameters of the fitted distribution.*
   - *For **test of homogeneity**, df $= (r-1)(c-1)$ where r is number of categories and columns c is number of independent groups.*
   - *For **test of independence**, df $= (r-1)(c-1)$ where r is number of categories in one varialbe and c is number of categories in another variable.*

3. *Select level of significance.*

4. *Compare $\mathcal{X}^2$ to critical value from $\chi_d f, 1 - \alpha$ (one-sided only since test is only in one direction)*

5. *Reject or Accept*

Example. Test whether observation are random variables whose distribution belong to a given family of distribution, the *theoretical frequencies* are calculated using a distribution from that family fitted in some standard way. The reduction in the degrees of freedom is calculated as $p = s + 1$ where s is the number of co-variates used in fitting the distribution, i.e. $p = 3$ for Normal distribution and $p = 2$ for poisson distribution. Hence $df = n - p$ where n is the number of categories. Unlike Student's t test, degree of freedom is not based on number of observations. The number of observations does not influence the number of degree of freedom.

Definition. ***Degree of freedom in Goodness of Fit Test*** *Degree of freedom is the number of parameter restricted by $\mathcal{H}_0$,which is equivalent to the number of free parameter - number of free parameter under $\mathcal{H}_0$. In general for tests under which the parameter for null distribution is not known and we construct test using asymptotics, then*

$$d.f. = \ Number \ of \ Cells \ - \ Number \ of \ estimated \ parameter \ -1$$

1. *Test hypothesis such that data follows Poisson, under which λ is not known; And the counts are split into k cells, then*

$$d.f. = k - 1 - 1 = k - 2$$

4

2. *In the case of Normal distribution with unknown μ and λ*

$$d.f. = k - 2 - 1 = k - 3$$

Definition. *Categorical Variable Test of Independence*

Example. There are two categorical variables, the vaccination status and health outcome. We want to test $\mathcal{H}_0$: vaccination status and health outcome are independent vs $\mathcal{H}_1$: they are not independent. We denote O for observed contingency table and E for expected cell counts under $\mathcal{H}_0$. We estimate marginal distribution of two variables using fixed table margin and law of independence

$$\mathbb{P}(X, Y) = \mathbb{P}(X)\mathbb{P}(Y)$$

After computing the expected contingency table E. We use

$$\mathcal{X}^2 = \sum_{i=1}^{\#rows} \sum_{j=1}^{\#cols} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \xrightarrow[\mathcal{H}_0]{\mathcal{D}} \chi^2_{(r-1)(c-1), 1-\alpha}$$

There are $r \cdot c - 1$ free parameter under Θ and $(r - 1) = (c - 1)$ free parameter under Θ_0 hence we have

$$d.f. = (r \cdot c - 1) - (r - 1 + c - 1) = (r - c)(c - 1)$$