Imperial College London

Lecture 10: Likelihood Ratio Tests

Statistical Modelling I

Dr. Riccardo Passeggeri

Outline

Introduction

Motivation

- Idea behind the maximum likelihood estimator: parameter with the highest likelihood is "best". Can this idea be used to create a test?
- More precisely, consider the hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{against} \quad H_1 : \theta \in \Theta_1 := \Theta \setminus \Theta_0$$

- Main idea: compare the maximised likelihood L under H_0 ($\sup_{\theta\in\Theta_0} L(\theta)$) to the unrestricted maximum likelihood ($\sup_{\theta\in\Theta} L(\theta)$). If the latter is (much?) larger then $\sup_{\theta\in\Theta_1} L(\theta) >> \sup_{\theta\in\Theta_0} L(\theta)$, casting doubt on H_0.

Example: $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$

For MLEs, we have seen the (approx.) pivotal quantity $\sqrt{nI_f(\hat{\theta})}(\hat{\theta} - \theta_0)$

Definition: Likelihood Ratio Test Statistic

Definition

Suppose we observe the data y. The **likelihood ratio test statistic** is

$$t(y) = \frac{\sup_{\theta \in \Theta} L(\theta; y)}{\sup_{\theta \in \Theta_0} L(\theta; y)} = \frac{\text{max. lik. under } H_0 + H_1}{\text{max. lik. under } H_0}$$

- If $t(y)$ is "large" this will indicate support for H_1, so reject H_0 when

  $$t(y) \geq k,$$

  where k is chosen to make

  $$\sup_{\theta \in \Theta_0} P_\theta(t(Y) \geq k) = (\text{or } \leq)\alpha$$

  (e.g. $\alpha = 0.05$).
- The choice of k ensures that we get a test to the level α.

Examples

Example 1: $X \sim$Binomial(n, θ), $\theta \in (0, 1) = \Theta$

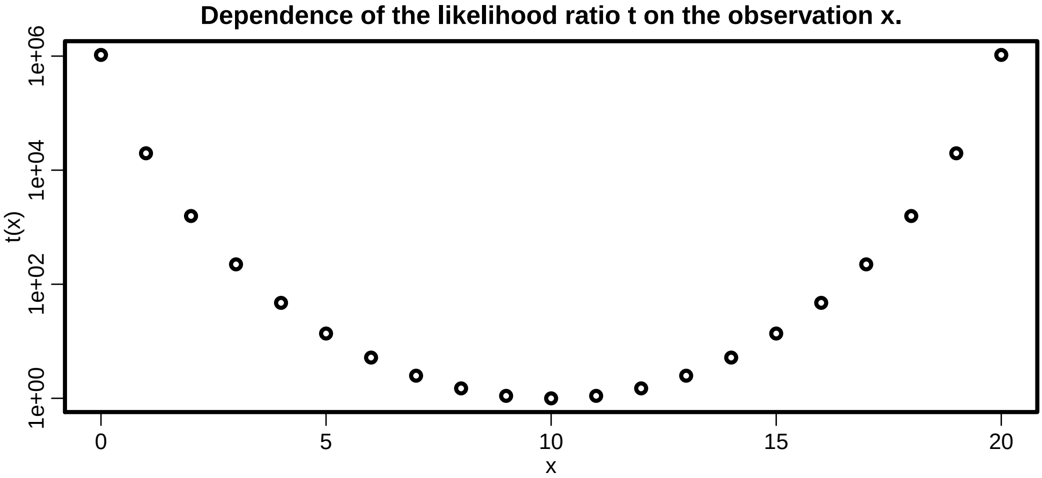
$$H_0 : \theta = 0.5 \quad \text{v.s.} \quad H_1 : \theta \neq 0.5$$

- Here, $\Theta_0 = \{0.5\}$, $\Theta_1 = (0, 0.5) \cup (0.5, 1)$
- The likelihood is

$$L : \Theta \to \mathbb{R},\ \theta \mapsto P_\theta(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}.$$

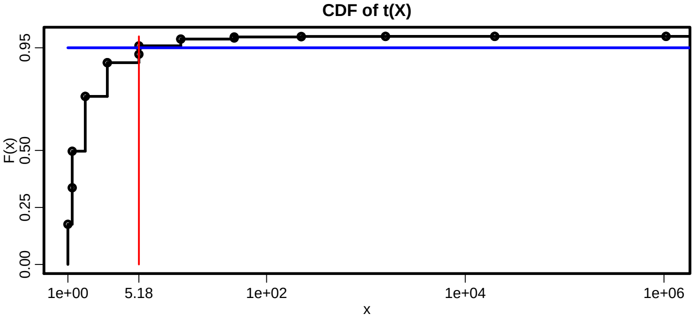
- The LRT statistic is

$$t(x) = \frac{\sup_{\theta \in \Theta} L(\theta)}{\sup_{\theta \in \Theta_0} L(\theta)} = \frac{L(\frac{x}{n})}{L(0.5)} = \frac{\left(\frac{x}{n}\right)^x \left(1 - \frac{x}{n}\right)^{n-x}}{0.5^n}$$



The above is for n=20. Note the log−scale on the y−axis.

- To construct the test we need the distribution of t under H_0



- In this case rejecting if $t > 5.19$ leads to a test with level 5%

Example 2: $X_i \sim \text{Binomial}(n, \theta_i), i = 1, 2$ independent

$$H_0 : \theta_1 = \theta_2 \text{ v.s. } H_1 : \theta_1 \neq \theta_2$$

The LRT statistic is

$$t = \frac{L(x_1/n, x_2/n)}{L(\frac{x_1+x_2}{2n}, \frac{x_1+x_2}{2n})}$$

The distribution of t under H_0 is not easy to obtain as for each value of $(\theta_1, \theta_2) \in \Theta_0$ the distribution of t may be different.

Example 3: m factories producing light bulbs. Are all factories producing bulbs of the same quality?

Observation: life-length of n light bulbs from each factory; Y_{ij} = life-length of bulb j from factory i.

Model: Y_{ij} indep. Exp(λ_i), i=1,...,m; j=1,...n, $\lambda_i > 0$ unknown, $i = 1, \dots, m$.

$$H_0 : \lambda_1 = \cdots = \lambda_m \quad \text{v.s.} \quad H_1 : \text{ not } H_0$$

Under $H_0 + H_1$ (using $\theta^t = (\lambda_1, \ldots, \lambda_m)$):

$$L(\theta) = \prod_{i=1}^{m}\prod_{j=1}^{n} \lambda_i \exp(-\lambda_i Y_{ij}) = \prod_{i=1}^{m} \lambda_i^n e^{-\lambda_i \sum_j y_{ij}}$$

$\implies$ m likelihoods of iid Exponential(λ_i) observations, leading to the MLE

$$\hat{\lambda}_i = \frac{1}{\bar{y}_i} \text{ where } \bar{y}_i = \frac{1}{n}\sum_j y_{ij}.$$

Under H_0 (setting $\lambda := \lambda_1 = \cdots = \lambda_m$):

$$L(\theta; y) = \lambda^{mn} e^{-\lambda \sum_{i,j} y_{ij}}$$

$\implies$ the likelihood for iid Exponential(λ) observations
MLE is

$$\hat{\lambda} = \frac{1}{\bar{y}}$$

LRT statistic:

$$t(y) = \frac{\sup_{\theta\in\Theta} L(\theta; y)}{\sup_{H_0} L(\theta; y)} = \frac{\frac{e^{-mn}}{(\prod_i \bar{y}_i)^n}}{\frac{e^{-mn}}{\bar{y}^{mn}}} = \frac{\bar{y}^{mn}}{(\prod_i \bar{y}_i)^n}$$

Remarks

- To construct a test we would need to know the distr. of $t(Y)$ under H_0. (Not easy)
- Even if it were known - the distribution of $t(Y)$ may depend on λ and hence, choosing k according to $\sup_{\lambda>0} P_\lambda(t(Y) \geq k) = \alpha$ may not be easy.

A Large Sample Result

The LRT statistic is asymptotically χ^2_r

Theorem

Let $Y_1, \ldots, Y_n$ be a random sample and denote $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$. Under mild regularity conditions

$$2 \log t(\mathbf{Y}_n) \xrightarrow{d} \chi^2_r \quad (n \to \infty)$$

under H_0, where $r = \#$independent restrictions on θ needed to define H_0.

Alternative way to derive the degrees of freedom r:

$$r = \#\text{ of independent parameters under full model} - \#\text{ of independent parameters under } H_0$$

Simplifying the Examples

- $X \sim$Binomial(n, θ), $\theta \in (0,1) = \Theta$ with $H_0 : \theta = 0.5$ v.s. $H_1 : \theta \neq 0.5$: r=1
- $X_i \sim$ Binomial$(n, \theta_i), i = 1, 2$ indep., $\theta \in (0,1)^2$ with $H_0 : \theta_1 = \theta_2$ v.s. $H_1 : \theta_1 \neq \theta_2$: r=1
- "light bulbs": $r = m - 1$

Proof of Asymptotic Distribution

Outline of Proof

Theorem

Let $Y_1, ..., Y_n$ be a random sample and denote $\mathbf{Y}_n = (Y_1, ..., Y_n)$. Under certain regularity conditions (in particular H_0 must be "nested" in H_1, i.e. Θ_0 is a lower-dimensional subspace/subset of Θ),

$$2 \log t(\mathbf{Y}_n) \xrightarrow{d} \chi^2_r \quad (n \to \infty)$$

under H_0, where $r = \#$independent restrictions on θ needed to define H_0.

1. *Taylor expansion* of $\ell(\theta) := \log L(\theta)$
2. *Slutsky's lemma*, *continuous mapping theorem*, *MLE theorem*, and *WLLN*.
3. NB: for clarity, I will sketch the univariate case (see main notes for $\Theta \subset \mathbb{R}^d$)

Next lecture

We consider linear models, which is one of the most common classes of statistical models.