Lecture 24: Likelihood ratio tests

Ciaran Evans

Likelihood ratio test

Let $X_1, ..., X_n$ be a sample from a distribution with parameter $\theta \in \mathbb{R}^d$. We wish to test $H_0 : \theta \in \Theta_0$ vs. $H_A : \theta \in \Theta_1$.

The **likelihood ratio test** (LRT) rejects H_0 when

$$\frac{\sup\limits_{\theta \in \Theta_1} L(\theta|\mathbf{X})}{\sup\limits_{\theta \in \Theta_0} L(\theta|\mathbf{X})} > k,$$

where k is chosen such that $\sup\limits_{\theta \in \Theta_0} \beta_{LR}(\theta) \leq \alpha$.

Example: linear regression with normal data

Suppose we observe $(X_1, Y_1), \ldots, (X_n, Y_n)$, where $Y_i = \beta^T X_i + \varepsilon_i$ and $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Partition $\beta = (\beta_{(1)}, \beta_{(2)})^T$. We wish to test $H_0 : \beta_{(2)} = 0$ vs. $H_A : \beta_{(2)} \neq 0$.

Full model (H_A): $Y_i = \beta^T X_i + \varepsilon_i$

Reduced model (H_0): $Y_i = \beta_{(1)}^T X_{i(1)} + \varepsilon_i$

Test statistic: $F = \dfrac{(SSE_{reduced} - SSE_{full})/q}{SSE_{full}/(n-p)}$

parameters tested

parameters in full model

Under H_0: $F \sim F_{q, n-p}$

$SSE_{full} = \sum_i (Y_i - \hat{\beta}_{full}^T X_i)^2$ $\quad$ $SSE_{reduced} = \sum_i (Y_i - \hat{\beta}_{(1)red}^T X_{i(1)})^2$

$\hat{\beta}_{full} = (X^T X)^{-1} X^T Y$ $\quad$ $\hat{\beta}_{(1)red.} = (X_{(1)}^T X_{(1)})^{-1} X_{(1)}^T Y$

MLE: $\hat{\sigma}^2_{full} = \frac{1}{n} SSE_{full}$ $\quad$ $\hat{\sigma}^2_{reduced} = \frac{1}{n} SSE_{reduced}$

Example: linear regression with normal data

Suppose we observe $(X_1, Y_1), \ldots, (X_n, Y_n)$, where $Y_i = \beta^T X_i + \varepsilon_i$ and $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Partition $\beta = (\beta_{(1)}, \beta_{(2)})^T$. We wish to test $H_0 : \beta_{(2)} = 0$ vs. $H_A : \beta_{(2)} \neq 0$.

LRT: rejects if
$$\frac{\sup_{\beta, \sigma^2} L(\beta, \sigma^2 \mid X, Y)}{\sup_{\substack{\beta, \sigma^2 \\ \beta_{(2)} = 0}} L(\beta, \sigma^2 \mid X, Y)} > k$$

$$\Rightarrow \quad \frac{L(\hat{\beta}_{full}, \hat{\sigma}^2_{full} \mid X, Y)}{L(\hat{\beta}_{reduced}, \hat{\sigma}^2_{reduced} \mid X, Y)} > k$$

$$\Leftrightarrow \quad \log L(\hat{\beta}_{full}, \hat{\sigma}^2_{full} \mid X, Y) - \log L(\hat{\beta}_{red}, \hat{\sigma}^2_{red} \mid X, Y) > \log(k)$$

$$\log L(\hat{\beta}, \hat{\sigma}^2) = \log\left(\left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\right)^n \exp\left\{-\frac{1}{2\hat{\sigma}^2} \sum_i (Y_i - \hat{\beta}_i^T X_i)^2\right\}\right)$$

Example: linear regression with normal data

$SSE = n\hat{\sigma}^2$

Suppose we observe $(X_1, Y_1), \ldots, (X_n, Y_n)$, where $Y_i = \beta^T X_i + \varepsilon_i$ and $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Partition $\beta = (\beta_{(1)}, \beta_{(2)})^T$. We wish to test $H_0 : \beta_{(2)} = 0$ vs. $H_A : \beta_{(2)} \neq 0$.

$$\log L(\hat{\beta}, \hat{\sigma}^2) = \log\left(\left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\right)^n \exp\left\{-\frac{1}{2\hat{\sigma}^2}\sum_i (Y_i - \hat{\beta}_i^T X_i)^2\right\}\right)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} SSE$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}^2) - \frac{n}{2}$$

$$\log L(\hat{\beta}_{full}, \hat{\sigma}^2_{full}) - \log L(\hat{\beta}_{red}, \hat{\sigma}^2_{red}) = \frac{n}{2}\log(\hat{\sigma}^2_{red}) - \frac{n}{2}\log(\hat{\sigma}^2_{full})$$

$\Rightarrow$ LRT rejects H_0 if $\frac{n}{2}\log\left(\frac{\hat{\sigma}^2_{red}}{\hat{\sigma}^2_{full}}\right) > \log(k)$

$$\Leftrightarrow \frac{n}{2}\log\left(\frac{SSE_{red}}{SSE_{full}}\right) > \log(k)$$

$$\Leftrightarrow \frac{(SSE_{red} - SSE_{full})/q}{SSE_{full}/(n-p)} > \underbrace{\left(\exp\left(\frac{2\log k}{n}\right) - 1\right)\left(\frac{n-p}{q}\right)}_{\text{choose to be upper } \alpha \text{ quantile } F_{q,n-p}}$$

Asymptotics of the LRT

Suppose we observe iid data $X_1, ..., X_n$ from a distribution with parameter $\theta \in \mathbb{R}$, and we wish to test $H_0 : \theta = \theta_0$ vs. $H_A : \theta \neq \theta_0$.

Theorem: Under H_0,

$$\underbrace{2 \log \left(\frac{L(\widehat{\theta}_{MLE} | \mathbf{X})}{L(\theta_0 | \mathbf{X})} \right)}_{\geq 0} \xrightarrow{d} \chi^2_1$$

b/c $L(\hat{\theta}_{MLE} | x) \geq L(\theta | x)$

$\forall \theta$

Generalization to higher dimensions

Suppose we observe iid data $X_1, ..., X_n$ with parameter $\theta \in \mathbb{R}^d$. Partition $\theta = (\theta_{(1)}, \theta_{(2)})^T$, with $\theta_{(2)} \in \mathbb{R}^q$. We wish to test

$$H_0 : \theta_{(2)} = \mathbf{0} \qquad H_A : \theta_{(2)} \neq \mathbf{0}$$

Theorem: Under H_0,

$$2 \log \left(\frac{\sup_\theta L(\theta|\mathbf{X})}{\sup\limits_{\theta:\theta_{(2)}=0} L(\theta|\mathbf{X})} \right) \xrightarrow{d} \chi^2_q$$