Lecture 25: Likelihood ratio tests

Ciaran Evans

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 (and assuming required regularity conditions),

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

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 (and assuming required regularity conditions),

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

Earthquake data

Data from the 2015 Gorkha earthquake on 211774 buildings, with variables including:

- `Damage`: whether the building sustained any damage (1) or not (0)
- Age: the age of the building (in years)
- `Surface`: a categorical variable recording the surface condition of the land around the building. There are three different levels: n, o, and t

Likelihood ratio tests

```
m1 <- glm(Damage ~ Age*Surface, data = earthquake,
          family = binomial)
summary(m1)$coefficients
```

```
##                    Estimate  Std. Error     z value       P
## (Intercept)    1.411099267 0.032512137  43.4022302  0.000
## Age            0.059786157 0.002099615  28.4748245 2.401
## Surfaceo       0.061461279 0.072860676   0.8435453  3.98
## Surfacet      -0.474024473 0.034382357 -13.7868520  3.058
## Age:Surfaceo   0.002807968 0.005087768   0.5519056  5.810
## Age:Surfacet   0.008163407 0.002230082   3.6605868  2.516
```

We want to test whether the relationship between Age and Damage is the same for all three surface conditions. What hypotheses do we test?

Likelihood ratio tests

Full model:

```
m1 <- glm(Damage ~ Age*Surface, data = earthquake,
          family = binomial)
```

Reduced model:

```
m2 <- glm(Damage ~ Age + Surface, data = earthquake,
          family = binomial)
```

Comparing deviances

```
m1 <- glm(Damage ~ Age*Surface, data = earthquake,
          family = binomial)
m1$deviance
```

```
## [1] 139150.5
```

```
m2 <- glm(Damage ~ Age + Surface, data = earthquake,
          family = binomial)
m2$deviance
```

```
## [1] 139164.4
```

Comparing deviances

```
m1 <- glm(Damage ~ Age*Surface, data = earthquake,
          family = binomial)

m2 <- glm(Damage ~ Age + Surface, data = earthquake,
          family = binomial)

pchisq(m2$deviance - m1$deviance,
       m2$df.residual - m1$df.residual,
       lower.tail = F)
```

```
## [1] 0.0009433955
```