Lecture 26: Likelihood ratio tests

Asymptotics of the LRT

Suppose we observe iid data $X_1, \dots, X_n$ and we want to test $H_0: \theta = \theta_0$ vs. $H_A: \theta \neq \theta_0$ $(\theta \in \mathbb{R})$

Under H_0,

$$\underbrace{2\ell(\hat{\theta}_{MLE} \mid X) - 2\ell(\theta_0 \mid X)}_{= 2 \log \left(\frac{L(\hat{\theta} \mid X)}{L(\theta_0 \mid X)} \right)} \xrightarrow{d} \chi^2_1$$

Proof sketch:

1. Taylor expansion: $2\ell(\hat{\theta}) - 2\ell(\theta_0) \approx -\ell''(\hat{\theta})(\hat{\theta} - \theta_0)^2$

2. $-\frac{1}{n}\ell''(\hat{\theta}) \xrightarrow{P} \mathcal{I}_1(\theta_0)$ (under H_0), $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}_1^{-1}(\theta_0))$

3. Apply Slutsky's & continuous mapping theorem

Proof : $(\hat{\theta} = \text{MLE})$

① $$\ell(\theta_0) \approx \ell(\hat{\theta}) + \underbrace{\ell'(\hat{\theta})}_{=0}(\theta_0 - \hat{\theta}) + \frac{\ell''(\hat{\theta})(\theta_0 - \hat{\theta})^2}{2}$$ (2nd order Taylor expansion)

$$2\ell(\theta_0) \approx 2\ell(\hat{\theta}) + \ell''(\hat{\theta})(\theta_0 - \hat{\theta})^2$$

$$\Rightarrow 2\ell(\hat{\theta}) - 2\ell(\theta_0) \approx -\ell''(\hat{\theta})(\hat{\theta} - \theta_0)^2$$

② know (previous proof of asymptotic normality) that (under H_0)

$$-\frac{1}{n}\ell''(\hat{\theta}) \xrightarrow{p} \mathcal{I}_1(\theta_0) \quad \text{(WLLN)}$$

$$\sqrt{n}(\hat{\theta} - \theta_0)^2 \xrightarrow{d} N(0, \mathcal{I}_1^{-1}(\theta_0))$$

$$-\ell''(\hat{\theta})(\hat{\theta} - \theta_0)^2 = \left(-\frac{1}{n}\ell''(\hat{\theta})\right)\left(\sqrt{n}(\hat{\theta} - \theta_0)\right)^2$$

③ $$= \Big(\underbrace{-\frac{1}{\sqrt{n}}\ell''(\hat{\theta})^{\frac{1}{2}}}_{\xrightarrow{p} \mathcal{I}_1(\theta_0)^{\frac{1}{2}} \text{ (CMT)}} \cdot \underbrace{\sqrt{n}(\hat{\theta} - \theta_0)}_{\xrightarrow{d} N(0, \mathcal{I}_1^{-1}(\theta_0))}\Big)^2 \xrightarrow{d} N(0,1)$$ (Slutsky's)

(CMT)

$$\Rightarrow -\ell''(\hat{\theta})(\hat{\theta} - \theta_0)^2 \xrightarrow{d} \chi^2_1$$ //

Generalization to higher dimensions

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

We want to test $H_0: \theta_{(2)} = \theta_{(2)0}$

$H_A: \theta_{(2)} \neq \theta_{(2)0}$

Under H_0,

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

$\uparrow$ # parameters tested

Intuition: $\hat{\theta}_{(2)} - \theta_{(2)0} \approx N(\dots) \in \mathbb{R}^q$

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

if Surface = O

$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = (\beta_0 + \beta_2) + (\beta_1 + \beta_4)\, Age$

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

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.411099267	0.032512137	43.4022302	0.000000e+00
Age	0.059786157	0.002099615	28.4748245	2.401973e-178
Surfaceo	0.061461279	0.072860676	0.8435453	3.989236e-01
Surfacet	-0.474024473	0.034382357	-13.7868520	3.058165e-43
Age:Surfaceo	0.002807968	0.005087768	0.5519056	5.810130e-01
Age:Surfacet	0.008163407	0.002230082	3.6605868	2.516383e-04

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

$H_0: \beta_4 = \beta_5 = 0$

$H_A:$ at least one of $\beta_4, \beta_5 \neq 0$

$\beta_0 + \beta_1 Age + \beta_2 SurfaceO + \beta_3 SurfaceT$

$+ \beta_4 Age \cdot SurfaceO + \beta_5 Age\, SurfaceT$

when $Age = 0$, Surface = N (not O, not T)

if Surface = N: $\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \beta_0 + \beta_1 Age$

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)
```

LRT: reject when $\underbrace{2\log\left(\frac{L(\hat{\beta}_{full})}{L(\hat{\beta}_{reduced})}\right)}_{\approx \chi^2_2}$ is large

$\approx \chi^2_2$ ← testing $\beta_4 = \beta_5 = 0$

Comparing deviances

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

← (similar to SSE)

```
[1] 139150.5
```

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

```
[1] 139164.4
```

Test statistic: $\text{deviance}_{\text{reduced}} - \text{deviance}_{\text{full}} \sim \chi^2_2$

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] 0.0009433954
```

(linear regression: saturated model $\hat{\mu}_i = Y_i$)

Deviance

Definition: The *deviance* of a fitted model with parameter estimates $\hat{\beta}$ is given by

$$2\ell(\text{saturated model}) - 2\ell(\hat{\beta})$$

$\text{deviance}_{\text{reduced}} - \text{deviance}_{\text{full}}$:

$$\cancel{2\ell(\text{saturated})} - 2\ell(\hat{\beta}_{\text{reduce}}) - \cancel{2\ell(\text{saturated})} + 2\ell(\hat{\beta}_{\text{full}})$$

$$= 2\ell(\hat{\beta}_{\text{full}}) - 2\ell(\hat{\beta}_{\text{reduced}})$$

Saturated model: model perfectly fits observed data

$$\hat{p}_i = Y_i$$

Binomial: $2\ell(\text{saturated}) = 2\sum_i \log\left(Y_i^{Y_i}(1-Y_i)^{1-Y_i}\right)$

$$= 2n\log(1) = 0$$

Summary: LRT for logistic regression