Likelihood ratio tests

Asymptotics of the LRT

Suppose we observe iid data $X_1, \ldots, X_n$ and want to test

$H_0: \theta = \theta_0$ vs. $H_A: \theta \neq \theta_0$. $(\theta \in \mathbb{R})$

Under H_0,

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

Proof: (To make notation easier, let $\ell(\theta) = \log L(\theta | X)$)

① Using Taylor expansions,

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

② $-\frac{1}{n}\ell''(\hat{\theta}) \xrightarrow{P} ?$, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} ?$

③ Apply Slutsky's & continuous mapping theorem

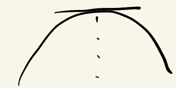
Proof $\quad (\hat{\theta} = \text{MLE})$

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

(2nd-order Taylor expansion around $\hat{\theta}$)

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

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



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

$$-\frac{1}{n}\ell''(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2}{\partial\theta^2}\log f(X_i \mid \theta)\right) \xrightarrow[\text{WLLN}]{P} -E\left[\frac{\partial^2}{\partial\theta^2}\log f(X_i \mid \theta)\right] = I_1(\theta)$$

Under H_0, consistency of MLE $\Rightarrow \hat{\theta} \xrightarrow{P} \theta_0$

$$\Rightarrow -\frac{1}{n}\ell''(\hat{\theta}) \xrightarrow{P} I_1(\theta_0)$$

Under H_0, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I_1^{-1}(\theta_0)) = I_1^{-\frac{1}{2}}(\theta_0)N(0,1)$

(asymptotic normality of MLE)

③ CMT: $$\left(\sqrt{n}(\hat{\theta} - \theta_0)\right)^2 \xrightarrow{d} I_1^{-1}(\theta_0)\,\chi_1^2$$

Slutsky's: $$-\frac{1}{n}\ell''(\hat{\theta})\left(\sqrt{n}(\hat{\theta} - \theta_0)\right)^2 \xrightarrow{d} I_1(\theta_0)\,I_1^{-1}(\theta_0)\,\chi_1^2 = \chi_1^2$$

//

Generalization to higher dimensions

Suppose we observe iid data $X_1, \ldots, X_n$ with parameter $\theta \in \mathbb{R}^d$

Partition $\theta = (\theta_{(1)}, \theta_{(2)})^T$, with $\theta_{(2)} \in \mathbb{R}^q$, $q \leq d$

we want to test $H_0: \theta_{(2)} = \theta_{(2)0}$ vs. $H_A: \theta_{(2)} \neq \theta_{(2)0}$

Under H_0,

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

q: # parameters tested

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

Surface n: Slope on Age = 0.060

surface o: Slope on Age = 0.060 + 0.003

Likelihood ratio tests

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

```
...
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.411099   0.032512  43.402  < 2e-16 ***
## Age            0.059786   0.002100  28.475  < 2e-16 ***
## Surfaceo       0.061461   0.072861   0.844 0.398924
## Surfacet      -0.474024   0.034382 -13.787  < 2e-16 ***
## Age:Surfaceo   0.002808   0.005088   0.552 0.581013
## Age:Surfacet   0.008163   0.002230   3.661 0.000252 ***
##
##     Null deviance: 153536  on 211773  degrees of freedom
## Residual deviance: 139150  on 211768  degrees of freedom
...
```

$H_0: \beta_4 = \beta_5 = 0$ H_A: at least one of $\beta_4, \beta_5 \neq 0$

> 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)
```

LRT: rejects when $\underbrace{2 \log \left(\frac{L(\hat{\beta}_{full})}{L(\hat{\beta}_{reduced})} \right)}_{2\ell(\hat{\beta}_{full}) - 2\ell(\hat{\beta}_{reduced})}$ is large

Likelihood ratio tests

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

```
...
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.411099   0.032512  43.402  < 2e-16 ***
## Age            0.059786   0.002100  28.475  < 2e-16 ***
## Surfaceo       0.061461   0.072861   0.844 0.398924
## Surfacet      -0.474024   0.034382 -13.787  < 2e-16 ***
## Age:Surfaceo   0.002808   0.005088   0.552 0.581013
## Age:Surfacet   0.008163   0.002230   3.661 0.000252 ***
##
##     Null deviance: 153536  on 211773  degrees of freedom
## Residual deviance: 139150  on 211768  degrees of freedom
...
```

What information replaces R^2 and R^2_{adj} in the GLM output?

deviance!

Deviance

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

$$2\ell(\textbf{saturated model}) - 2\ell(\widehat{\beta})$$

$$2\ell(\hat{\beta}) = 2\sum_{i=1}^{n} \log\left(\hat{p}_i^{\,Y_i}(1-\hat{p}_i)^{1-Y_i}\right) \qquad \hat{p}_i = \frac{e^{\hat{\beta}^T x_i}}{1+e^{\hat{\beta}^T x_i}}$$

Saturated model: perfectly fits the observed data

$$\hat{p}_i = Y_i \quad \Rightarrow \quad 2\ell(\text{saturated}) = 2\sum_{i=1}^{n} \log\left(Y_i^{Y_i}(1-Y_i)^{1-Y_i}\right) = 2n\log(1) = 0$$

$\Rightarrow$ for binary logistic regression, deviance $= -2\ell(\hat{\beta})$

Residual and null deviance

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

```
...
##     Null deviance: 153536  on 211773  degrees of freedom
## Residual deviance: 139150  on 211768  degrees of freedom
...
```

$\overbrace{211773}^{n-1}$, $\underbrace{211768}_{n-p}$

linear reg. analogue: SSTotal ($\sum_i (Y_i - \bar{Y})^2$)

$n = \#\text{obs.}$, $p = \#\text{parameters}$

Residual deviance: deviance for the fitted model

$$-2\ell(\hat{\beta}) = 139150$$

Null deviance: (residual) deviance for fitted intercept-only model

i.e. $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0$

LRT : $2\ell(\hat\beta_{full}) - 2\ell(\hat\beta_{reduced}) \approx \chi^2_q$ q = # parameters tested

= deviance reduced − deviance$_{full}$ $= df_{red.} - df_{full}$

Comparing deviances

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

full model

```
...
##     Null deviance: 153536  on 211773  degrees of freedom
## Residual deviance: 139150  on 211768  degrees of freedom
...
```

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

reduced model

```
...
##     Null deviance: 153536  on 211773  degrees of freedom
## Residual deviance: 139164  on 211770  degrees of freedom
...
```

LRT : 139164 − 139150 = 14

Calculate p-value by comparing χ^2_2

> How should I use this output to calculate a test statistic?

Comparing deviances

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

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

test stat

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

df

```
## [1] 0.0009433954
```

Summary: LRT for logistic regression

$Y_i \sim \text{Bernoulli}(p_i)$

<u>Full model:</u> $\log\left(\frac{p_i}{1-p_i}\right) = \beta^T X_i$

$\beta = \begin{bmatrix} \beta_{(1)} \\ \beta_{(2)} \end{bmatrix}$, $\beta_{(2)} \in \mathbb{R}^q$ $H_0: \beta_{(2)} = 0$ $H_A: \beta_{(2)} \neq 0$

<u>Reduced model:</u> $\log\left(\frac{p_i}{1-p_i}\right) = \beta_{(1)}^T X_{i(1)}$

① Fit full and reduced models, calculate (residual) deviances

② Test statistic: $G = 2\ell(\hat{\beta}_{full}) - 2\ell(\hat{\beta}_{reduced}) = \text{deviance}_{reduced} - \text{deviance}_{full}$

③ Under H_0, $G \sim \chi^2_q \leftarrow$ # parameters tested

$= df_{reduced} - df_{full}$