

Lecture 9: GLM Estimation — Deviance and Residuals

MATH3823 Generalised Linear Models

Richard P Mann

MATH3823 Generalised Linear Models

Reading

Course notes: Chapter 4, Sections 4.3–4.5

www.richardpmann.com/MATH3823

Model Comparison: The Key Question

Given:

- A fitted model M with r parameters
- The saturated model \tilde{M} with n parameters

Questions:

- ➊ How well does model M fit the data?
- ➋ How do we compare two nested models?

Answer: The deviance.

Definition of Deviance

Deviance:

$$D = 2\phi \left\{ \ell(\tilde{\boldsymbol{\theta}}; \mathbf{y}) - \ell(\hat{\boldsymbol{\theta}}; \mathbf{y}) \right\}$$

where:

- $\tilde{\boldsymbol{\theta}}$ = MLE under saturated model ($\tilde{\mu}_i = y_i$)
- $\hat{\boldsymbol{\theta}}$ = MLE under fitted model M
- ϕ = scale parameter

Interpretation:

- Measures “distance” from saturated model
- Larger $D =$ worse fit
- $D = 0$ only for the saturated model

Deviance for Specific Distributions

Normal distribution:

$$D = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 = \text{RSS}$$

Poisson distribution:

$$D = 2 \sum_{i=1}^n \left\{ y_i \log \frac{y_i}{\hat{\mu}_i} - (y_i - \hat{\mu}_i) \right\}$$

Binomial distribution:

$$D = 2 \sum_{i=1}^n \left\{ y_i \log \frac{y_i}{\hat{\mu}_i} + (m_i - y_i) \log \frac{m_i - y_i}{m_i - \hat{\mu}_i} \right\}$$

Deviance and Goodness of Fit

Under the null hypothesis that model M is correct:

$$D \stackrel{a}{\sim} \chi_{n-r}^2$$

where r = number of parameters in M .

Goodness-of-fit test:

- Large D relative to $n - r \Rightarrow$ poor fit
- p -value: $\mathbb{P}(\chi_{n-r}^2 > D)$
- Small p -value \Rightarrow reject model M

Caution: χ^2 approximation may be poor for small samples or sparse data.

Comparing Nested Models

For nested models $M_1 \subset M_2$:

Deviance difference:

$$D_1 - D_2 \stackrel{a}{\sim} \chi^2_{r_2 - r_1}$$

under H_0 : M_1 is adequate.

Test procedure:

- ① Fit both models, obtain D_1 and D_2
- ② Calculate $\Delta D = D_1 - D_2$
- ③ Compare to $\chi^2_{r_2 - r_1}$ distribution
- ④ Reject H_0 if p -value is small

When ϕ is Unknown

For Normal distribution: $\phi = \sigma^2$ is unknown.

Estimate from saturated model:

$$\hat{\phi} = \frac{D_{\text{sat}}}{n - r_{\text{sat}}}$$

Use F-test instead of χ^2 :

$$F = \frac{(D_1 - D_2)/(r_2 - r_1)}{\hat{\phi}} \sim F_{r_2 - r_1, n - r_{\text{sat}}}$$

Note: For Poisson and Binomial, $\phi = 1$ is known, so χ^2 test applies.

Types of Residuals

1. Raw (response) residuals:

$$e_i^{\text{raw}} = y_i - \hat{\mu}_i$$

2. Pearson (standardized) residuals:

$$e_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$

3. Deviance residuals:

$$e_i^D = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i}$$

where d_i is the contribution of observation i to deviance.

Pearson Residuals

$$e_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$

Properties:

- Standardized to have approximately unit variance
- Sum of squared Pearson residuals:

$$X^2 = \sum_{i=1}^n (e_i^P)^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

- $X^2 \approx D$ for large samples (both $\sim \chi^2_{n-r}$)

For Poisson: $e_i^P = (y_i - \hat{\mu}_i)/\sqrt{\hat{\mu}_i}$

For Binomial: $e_i^P = (y_i - m_i \hat{p}_i)/\sqrt{m_i \hat{p}_i (1 - \hat{p}_i)}$

Deviance Residuals

$$e_i^D = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i}$$

where $D = \sum_{i=1}^n d_i$.

Properties:

- $\sum_{i=1}^n (e_i^D)^2 = D$ (deviance)
- Often more normally distributed than Pearson residuals
- Preferred for model diagnostics

For Poisson:

$$d_i = 2 \left\{ y_i \log \frac{y_i}{\hat{\mu}_i} - (y_i - \hat{\mu}_i) \right\}$$

Residual Diagnostics

Plot residuals to check:

① Residuals vs. fitted values:

- Should show random scatter
- Patterns suggest model misspecification

② Residuals vs. covariates:

- Check for missed nonlinear effects

③ Q-Q plot of residuals:

- Check approximate normality
- (Less critical for GLMs than for normal models)

Fitting GLMs in R

```
# Fit a Poisson GLM
model <- glm(y ~ x1 + x2, family = poisson)

# Fit a binomial GLM (logistic regression)
model <- glm(cbind(success, failure) ~ x,
             family = binomial)

# Summary with coefficients and deviance
summary(model)

# Analysis of deviance
anova(model, test = "Chisq")

# Compare nested models
anova(model1, model2, test = "Chisq")
```

Extracting Residuals in R

```
# Raw (response) residuals  
residuals(model, type = "response")  
  
# Pearson residuals  
residuals(model, type = "pearson")  
  
# Deviance residuals (default)  
residuals(model, type = "deviance")  
  
# Deviance  
deviance(model)  
  
# Fitted values  
fitted(model)
```

Example: Poisson Regression Output

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 2.3456     0.1234   19.01 <2e-16 ***
x1          0.5678     0.0567   10.01 <2e-16 ***
x2         -0.2345     0.0789   -2.97  0.0030 **
Null deviance: 234.56 on 99 degrees of freedom
Residual deviance: 98.76 on 97 degrees of freedom
AIC: 345.67

Number of Fisher Scoring iterations: 4
```

Key output:

- Residual deviance: 98.76 on 97 df (close to 1:1 ratio = good fit)
- Compare to Null deviance to assess model improvement

Summary

Key points:

- Deviance measures distance from saturated model
- $D \sim \chi^2_{n-r}$ under correct model specification
- Compare nested models via deviance differences
- Pearson residuals: standardized by variance function
- Deviance residuals: sum of squares equals deviance
- Use residual plots for model diagnostics
- R's `glm()` function handles all major GLM families

Next lecture: Logistic regression for binary and binomial data.