

STAT 151A Lecture 33

Henry Liev

12 November 2025

Remark 0.1 (Inference for Logistic Regression)

Maximum Likelihood

$$y_1, \dots, y_n \stackrel{iid}{\sim} \text{Bern}(\pi_i)$$

$$\pi_i = \frac{1}{1 + \exp(-\alpha - \beta x_i)}$$

When we derive the log-likelihood, take derivatives, and set to zero, we get normal equations

$$\sum y_i = \sum \frac{1}{1 + \exp(-\hat{\alpha} - \hat{\beta} x_i)} = \sum \hat{\pi}_i$$

$$\sum y_i x_i = \sum \frac{1}{1 + \exp(-\hat{\alpha} - \hat{\beta} x_i)} x_i = \sum \hat{\pi}_i x_i$$

In matrix form: let $x = (\vec{1} \quad \vec{x})$

$$\text{Normal equations } \mathbf{X}\vec{y} = \mathbf{X}\vec{\pi} \text{ where } \vec{\pi} = \begin{pmatrix} \hat{\pi}_1 \\ \vdots \\ \hat{\pi}_n \end{pmatrix}$$

Compare to OLS $\mathbf{X}^\top \vec{y} = \mathbf{X}^\top \hat{y} = \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \hat{y} = \mathbf{X}^\top \hat{y}$

Compare to OLS normal equations $\mathbf{X}^\top \vec{y} = \mathbf{X}^\top \hat{y}$

What about confidence intervals and hypothesis tests for $\hat{\beta}$?

OLS Case: NLM: if ε_i are normal iid, homoskedastic, then everything is normal

HLM/Bootstrap: Can't be sure they are always normal in finite samples but as $n \rightarrow \infty$, approximately normal

Logistic regression \rightarrow as $n \rightarrow \infty$ things will be approximately normal \rightarrow special case of maximum likelihood inference

Roadmap: Review general inference for iid data

Expand to logistic regression

Assume $y_1, \dots, y_n \stackrel{iid}{\sim} f(y|\theta)$

How to get inference on $\hat{\theta}_{MLE}$

As $n \rightarrow \infty$

Theorem 0.2 (Theorem A)

If f is smooth then $\hat{\theta}_{MLE} \xrightarrow{n \rightarrow \infty} \theta$ (Consistency)

Proof. Depends on Law of Large Numbers applied to log-likelihood

■

Remark 0.3

But what about the $\text{Var}(\hat{\theta}_{MLE})$ and its distribution?

Useful tools

$$\text{Information: } I(\theta) = \mathbb{E} \left[\frac{\partial}{\partial \theta} \log f(y|\theta) \right]^2$$

$$\text{Also true that (when } f \text{ is smooth) } I(\theta) = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta^2} f(y|\theta) \right]$$

Lemma 0.4

When f is smooth $\mathbb{E} \left[\frac{\partial}{\partial \theta} \log f(y|\theta) \right] = 0$

Proof. $\int \frac{\partial}{\partial \theta} \log f(y|\theta) f(y|\theta) dy$

$$\int \frac{\frac{\partial}{\partial \theta} f(y|\theta)}{f(y|\theta)} f(y|\theta) dy$$

$$\int \frac{\partial}{\partial \theta} f(y|\theta) dy = \frac{\partial}{\partial \theta} \int f(y|\theta) dy = \frac{\partial}{\partial \theta} 1 = 0$$

■

Theorem 0.5 (Theorem B: Asymptotic Distribution of $\hat{\theta}_{MLE}$)

When f is smooth, θ is not an extreme value, and support of f does not depend on θ

$$\sqrt{nI(\theta)}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} N(0, 1)$$

In other words (roughly) $\hat{\theta}_{MLE} \xrightarrow{n \rightarrow \infty} N(\theta, \frac{1}{nI(\theta)})$

Proof. How did we get $\hat{\theta}_{MLE}$? We maximized the log likelihood: $\hat{l}(\hat{\theta}_{MLE})$

At $\hat{\theta}_{MLE} : l'(\hat{\theta}_{MLE}) = 0$

Taylor expansion: $l'(\hat{\theta}_{MLE}) = l'(\theta) + (\hat{\theta}_{MLE} - \theta)l''(\theta)$

$$0 = l'(\theta) + (\hat{\theta}_{MLE} - \theta)l''(\theta)$$

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) = -\frac{l'(\theta)/\sqrt{n}}{l''(\theta)/n}$$

Denominator

$$\frac{l''(\theta)}{n} = \frac{1}{n} \sum \frac{\partial}{\partial \theta^2} \log f(y_i|\theta) \xrightarrow{LLN} \mathbb{E} \left[\frac{\partial^2}{\partial \theta^2} \log f(y|\theta) \right] = -I(\theta)$$

Numerator

$$\mathbb{E} \left(\frac{l'(\theta)}{\sqrt{n}} \right) = 0$$

$$\text{Var} \left(\frac{l'(\theta)}{\sqrt{n}} \right) = \frac{1}{n} \sum \left[\frac{\partial}{\partial \theta} \log f(y_i|\theta) \right]^2 \xrightarrow{LLN} I(\theta)$$

Put numerator together:

As $n \rightarrow \infty$ numerator will be approx $N(0, I(\theta))$ by CLT

Denominator: $I(\theta)$ right hand side is $\frac{\sqrt{I(\theta)}N(0, 1)}{I(\theta)} \sim \sqrt{n}(\hat{\theta}_{MLE} - \theta), N(0, 1) \sim \sqrt{nI(\theta)}(\hat{\theta}_{MLE} - \theta)$

■

Remark. $I(\theta)$ is the information for one observation

$I_n(\theta)$ is the information for n observations

Non-iid case $I_i(\theta)$ for each point and overall information $\sum I_i(\theta) = I_n(\theta)$

In non-iid (?) case we write

$$\sqrt{I_n(\theta)}(\hat{\theta}_{MLE} - \theta) \sim N(0, 1)$$

Remark 0.6 (Back to Logistic Regression)

We get $\hat{\beta}_{MLE}$ by solving $\mathbf{X}^\top \mathbf{y} = \mathbf{X}^\top \hat{\pi}$

Theory tells us $\hat{\beta}_{MLE} \stackrel{approx}{\sim} \mathcal{N}_{p+1}(\beta, I_n(\beta))$

$$I_n(\beta) = \sum \mathbb{E} \left[\frac{\partial^2}{\partial \beta^2} \log f(y_i | x_i, \beta) \right]$$

$p + 1 \times p + 1$ matrix (Hessian), variance covariance matrix