STAT 151A Lecture 33

Henry Liev

12 November 2025

Remark 0.1 (Inference for Logistic Regression)

Maximum Likelihood
$y_1, \dots, y_n \overset{iid}{\sim} \text{Bern}(\pi_i)$
$\pi_i = \frac{1}{1+\exp(-\beta x_i - \alpha)}$
When we derive the log-likelihood, take derivaives, 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 $x = \begin{pmatrix} \vec{1} & \vec{x} \end{pmatrix}$
Normal equations $\mathbf{X}\vec{y} = \mathbf{X}\vec{\pi}$ where $\vec{\pi} = \begin{pmatrix} \hat{\pi}_1 \\ \vdots \\ \hat{\pi}_n \end{pmatrix}$
Compare to OLS $\mathbf{X}^\intercal \vec{y} = \mathbf{X}^\intercal \hat{y} = \mathbf{X}^\intercal \mathbf{X}(\mathbf{X}^\intercal \mathbf{X})^{-1}\mathbf{X}^\intercal \hat{y} = \mathbf{X}^\intercal \hat{y}$
Compare to OLS normal equations $\mathbf{X}^\intercal \vec{y} = \mathbf{X}^\intercal \hat{y}$
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 $n \to \infty$, approximately normal
Logistic regression $\to$ as $n \to \infty$ things will be approximately normal $\to$ special case of maximum likelihood inference
Roadmap: Review general inference for iid data
Expand to logistic regression
Assume $y_1, \dots, y_n \overset{iid}{\sim} f(y|\theta)$
How to get inference on $\hat{\theta}_{MLE}$
As $n \to \infty$

Theorem 0.2
If f is smooth then $\hat{\theta}_{MLE} \overset{n\to\infty}{\to} \theta$ (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
Information: $I(\theta) = \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(y|\theta)\right]^2$
Also true that (when f 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 (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) $\hat{\theta}_{MLE} \overset{n\to\infty}{\sim} N(\theta, \frac{1}{nI(\theta)})$

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: $l'(\hat{\theta}_{MLE}) = l'(\theta) + \left(\hat{\theta}_n - \hat{\theta}\right) l''(\theta)$
$0 = l'(\theta) + (\hat{\theta}_n - \theta)l''(\theta)$
$\sqrt{n}(\hat{\theta}_n - \theta) = -\frac{l'(\theta)/\sqrt{n}}{l''(\theta)/n}$
Denominator
$\frac{l''(\theta)}{n} = \frac{1}{n}\sum\frac{\partial}{\partial\theta^2}\log f(y_i|\theta) \to \overset{LLN}{\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(y|\theta)\right]} = -I(\theta)$
Numerator
$\mathbb{E}\left(\frac{l'(\theta)}{\sqrt{n}}\right) = 0$
$\text{Var}\left(\frac{l'(\theta)}{\sqrt{n}}\right) = \frac{1}{n}\sum\left[\frac{\partial}{\partial\theta}\log f(y_i|\theta)\right]^2 \overset{LLN}{\to} I(\theta)$
Put numerator together:
As $n\to\infty$ 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 infromation 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 $\mathbf{X}^\intercal y = \mathbf{X}^\intercal \hat{\pi}$
Theory tells us $\hat{\beta}_{MLE} \overset{approx}{\sim} \mathcal{N}_{p+1}(\beta, I_n(\beta))$
$I_n(\beta) = \sum \mathbb{E}\left[\frac{\partial^2}{\partial \beta^2} \log f(y_i|x_i, \beta)\right]$
$p+1 \times p+1$ matrix (Hessian), variance covariance matrix