Announcements

- Drop 1 HW
- Exams out soon
- Final 3:30–5:30p here

Logistic Regression Pt 2

To perform inference in Logistic Regression, we rely on asymptotic results (i.e. large sample approximation).

[I] Wald test for Indiv. coefficients

Main Idea: When $n \to \infty$, the MLE

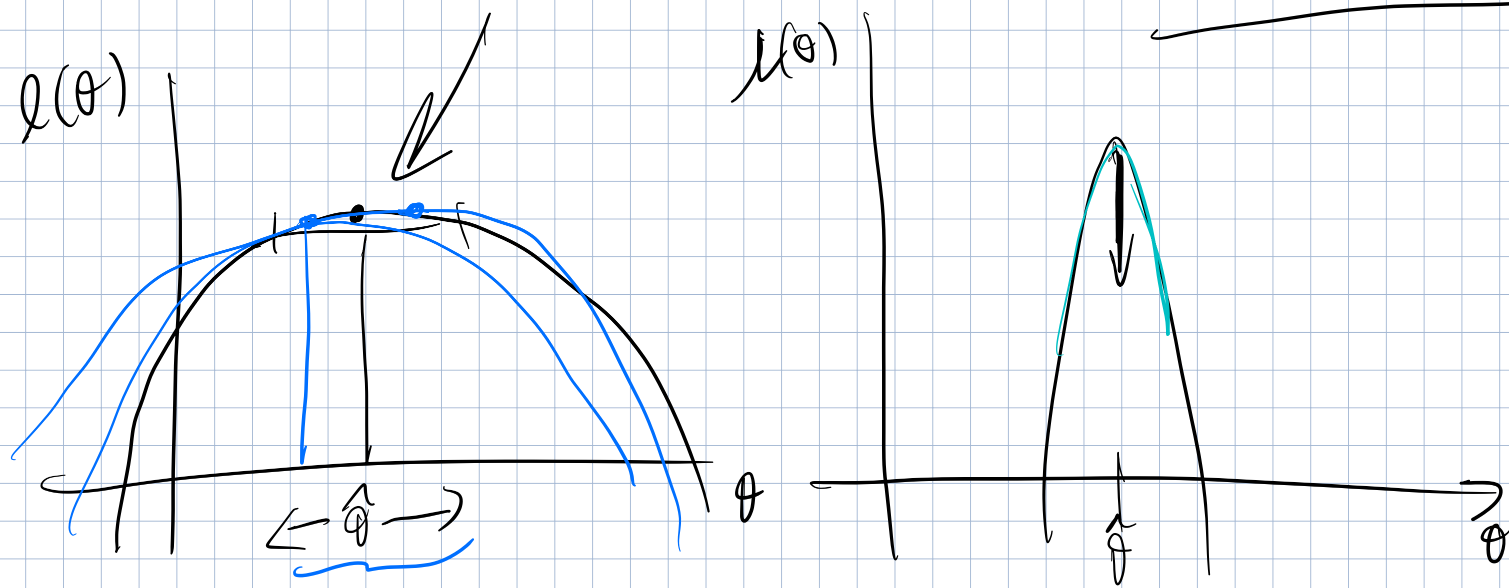
$$\boxed{\hat{\beta}_k \,\dot{\sim}\, N(\beta_k, \underline{\underline{V_k}})}$$

where V_k is defined as follows:

Let G denote the negative Hessian matrix of the loglikelihood:

$$G = -\begin{bmatrix} \frac{\partial^2 l}{\partial \beta_0^2} & \frac{\partial^2 l}{\partial\beta_0\partial\beta_1} & \cdots & \frac{\partial^2 l}{\partial\beta_0\partial\beta_{p-1}} \\ & \boxed{\frac{\partial^2 l}{\partial\beta_1^2}} & & \\ & & \ddots & \\ & & & \frac{\partial^2 l}{\partial\beta_{p-1}^2}\end{bmatrix}$$

Fisher Information



$$Var(\hat\beta) = G^{-1}\Big|_{\beta=\hat\beta} \Rightarrow \underbrace{Var(\hat\beta_k)}_{= V_k} = \left[G^{-1}\Big|_{\beta=\hat\beta}\right]_{k+1,k+1}$$

$$l(\hat{\theta}) \approx l(\theta) + \underline{(\hat{\theta}-\theta)\, l'(\theta)} + \underline{\frac{(\hat{\theta}-\theta)^2}{2!} l''(\theta) + \cdots}$$

$$(\hat{\theta}-\theta) \approx \frac{l(\hat{\theta}) - l(\theta)}{l'(\theta)} \rightarrow \text{CLT type result}$$

(comes from here)

Then we can do hyp. testing like:

$$\frac{\hat{\beta}_k - \beta_k}{SE(\hat{\beta}_k)} \mathrel{\dot\sim} N(0,1)$$

So if I want to test:

$H_0: \beta_k = 0$ vs $H_1: \beta_k \neq 0$

the test stat:

$$Z = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}$$ which rejects H_0 if

$$|Z| > Z^*_{1-\alpha/2}$$

II Deviance & Likelihood Ratio Tests for Reduced & Full Models

If we want to compare two models w/ a different # of predictors, for ex:

$$H_0: \beta_r = \beta_{r+1} = \cdots = \beta_{p-1} = 0$$

$\hookrightarrow$ Reduced: $\text{logit}(\hat{\pi}_i) = \beta_0 + \sum_{j=1}^{r-1} \beta_j X_{ji}$

vs.

H_1: at least one of $\{\beta_j\}_{j=r}^{p-1}$ is not zero

$\hookrightarrow$ Full: $\text{logit}(\hat{\pi}_i) = \beta_0 + \sum_{j=1}^{p-1} \beta_j X_{ji}$

Our test is based on comparing two loglikelihoods: one for the reduced model & one for the full model

Def: Deviance $= -2\ell(\beta)$

For ex:

For the linear regression model:

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_i - \beta_0 - \sum_{j=1}^{p-1} \beta_j x_{ji})^2 / 2\sigma^2} \right]$$

$$= (2\pi\sigma^2)^{-n/2} e^{-\sum_i (y_i - \beta_0 - \sum_{j=1}^{p-1} \beta_j x_{ji})^2 / 2\sigma^2}$$

$$\rightarrow \ell(\beta) = \log \mathcal{L}(\beta) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{\sum_i (y_i - \beta_0 - \sum_{j=1}^{p-1} \beta_j x_{ji})^2}{2\sigma^2}$$

$$\Rightarrow -2\ell(\beta) = \underline{n \log(2\pi\sigma^2)} + \frac{1}{\sigma^2} \sum_i (y_i - \beta_0 - \sum_{j=1}^{p-1} \beta_j x_{ji})^2$$

$\Rightarrow$ Dev $= -2\ell(\beta)$ @ $\beta = \hat{\beta}$ $\Rightarrow$ Dev $= k_1 + k_2 \cdot \underline{\underline{SSE}}$

Think of Deviance as a more general SSE.

To construct a likelihood based test,

define the test stat as:

$$\Delta D = Dev(Reduced) - Dev(Full)$$

Decision Rule:

if $\Delta D > \chi^2_{df, 1-\alpha}$ reject H_0,

$df = P_{Full} - P_{red}$.

If we reject H_0, we have significant evidence that the full model is a much better fit for the data according to likelihood.

Remember for logistic Regression,

$$l = \sum_{i=1}^{n} \left(y_i \log\left(\underline{\pi_i / 1-\pi_i} \right) + \log(1-\pi_i) \right)$$

$\Rightarrow$ Deviance of a fitted model is:

$$-2\ell(\hat{\beta}) = -2\sum_{i=1}^{n}\left(y_i \log\left(\hat{\pi}_i / 1-\hat{\pi}_i\right) + \log(1-\hat{\pi}_i)\right)$$

Reminder:

$$\mathcal{L}(\underset{\sim}{\pi}) = \prod_{i=1}^{n} (\pi_i)^{y_i}(1-\pi_i)^{1-y_i}$$

$$\hookrightarrow \prod_{i=1}^{n} \left(\frac{\pi_i}{1-\pi_i}\right)^{y_i}(1-\pi_i)$$

$\boxed{\text{III}}$ Deviance Residuals

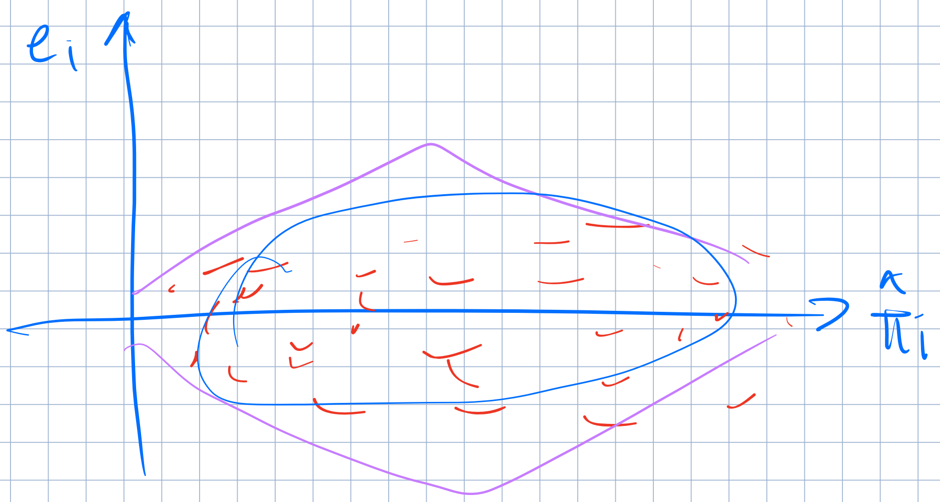
Standard OLS residuals do not immediately apply to logistic regression.

If we try to construct an analog of "residuals" for logistic reg, there are 2 common approaches:

① Pearson Residual:
- similar to studentized residuals

$$e_i = \frac{y_i - \hat{\pi}_i}{\widehat{se}(y_i)} = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}}$$

We would hope to see no pattern in the Pearson residual plot:



↳ interpret just like residual plot in OLS.

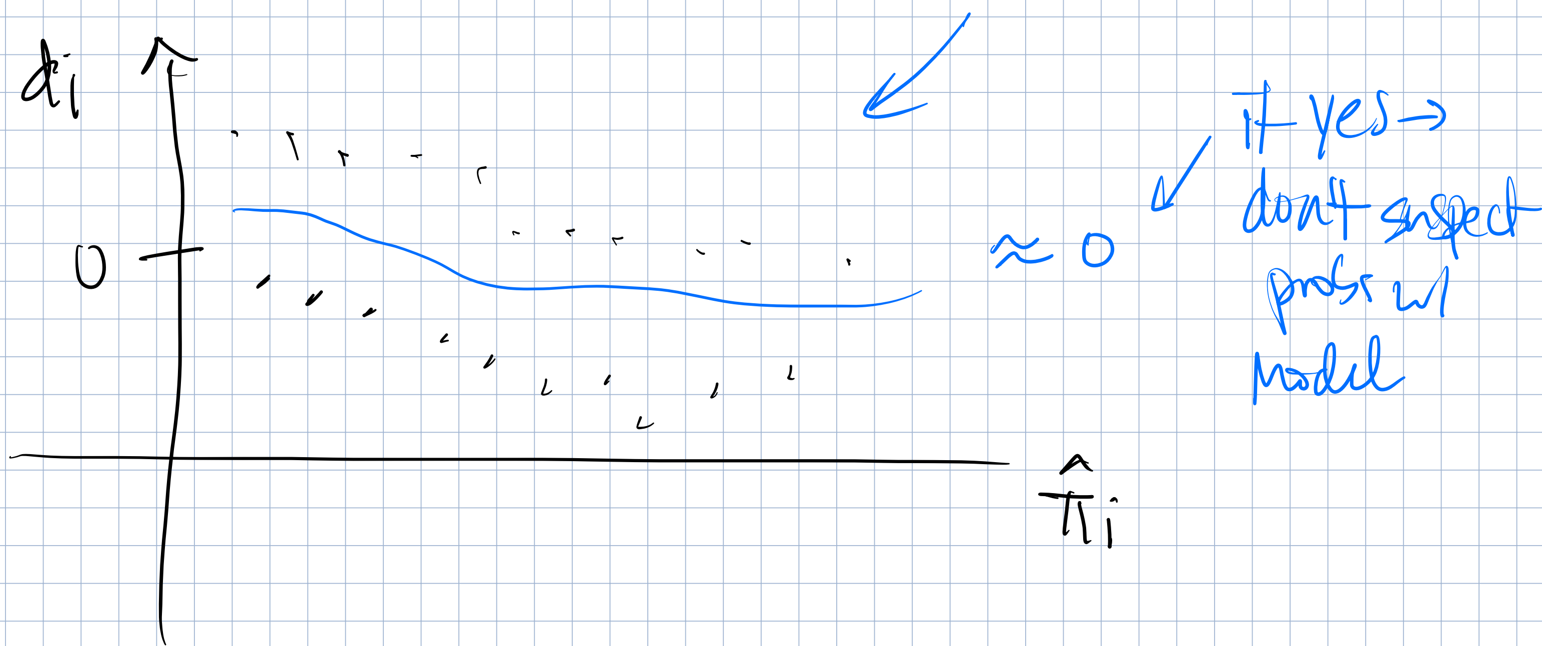
② Deviance Residuals

Define the deviance residuals as:

$$d_i = \begin{cases} \sqrt{-2\left(y_i \log\left(\hat{\pi}_i / 1-\hat{\pi}_i\right) + \log\left(1-\hat{\pi}_i\right)\right)} & \text{if } y_i = 1, \\ -\sqrt{-2\left(y_i \log\left(\hat{\pi}_i / 1-\hat{\pi}_i\right) + \log\left(1-\hat{\pi}_i\right)\right)} & \text{if } y_i = 0 \end{cases}$$

The sqrt of each individual contribution to deviance measures the part of deviance attributable to the i^{th} observation.

We can check the dev. residual scatterplot like this:



Pseudo R^2:

Because there is no OLS principle, the regular R^2 concept doesn't exist for logistic regression to "explain variance."

Some "pseudo R^2" statistics have been proposed to assess goodness-of-fit:

<u>Efron</u>: $$\text{pseudo } R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{\pi}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

<u>McFadden</u>: $$\text{pseudo } R^2 = 1 - \frac{\ell(\text{Full})}{\ell(\text{Null})}$$

<u>Note</u>: AIC/BIC can still be used for model selection b/c they are likelihood based.