Announcements

- Drop lowest HW
- Exams out soon
- Final 3:30-5:30p Here
- Wed OH rescheduled to 3-4p

Logistic Regression pt II

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

I Wald Test for Indiv. Coefficients

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

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

where V_k is defined as follows:

Let G denote the negative Hessian matrix

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

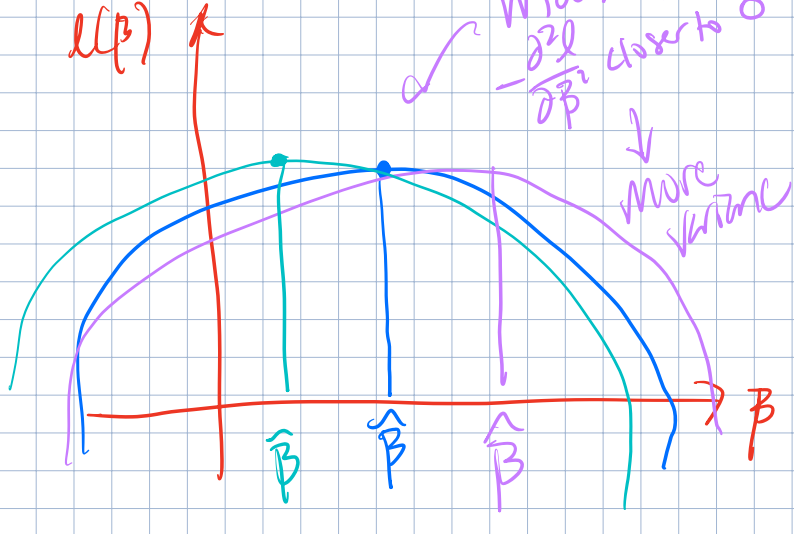
Fisher Information

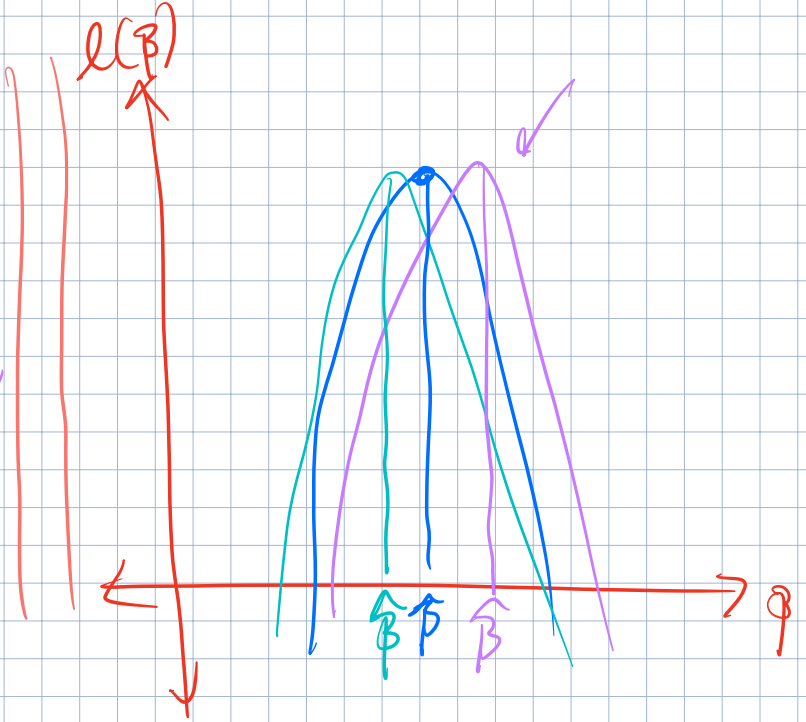
Then the variance of $\hat{\beta}_k$, is

$$\text{Var}(\hat{\beta}) = G^{-1}\Big|_{\beta=\hat{\beta}} \rightarrow \text{Var}(\hat{\beta}_k) = \left[G^{-1}\Big|_{\beta=\hat{\beta}}\right]_{k+1,\,k+1}$$

$(k+1)^{st}$ diag b/c of β_0

Why does this work?





Ok! Now we know that

$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{V_k}} = \frac{\hat{\beta}_k - \beta_k}{SE(\hat{\beta}_k)} \overset{\cdot}{\sim} N(0,1)$$

So an asymptotic test for β_k can be constructed as:

$$H_0: \beta_k = 0 \quad \text{vs} \quad H_1: \beta_k \neq 0$$

The test statistic:

$$Z = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}$$ leads to the decision rule:

If $|Z| > Z^*_{1-\frac{\alpha}{2}}$ then reject H_0.

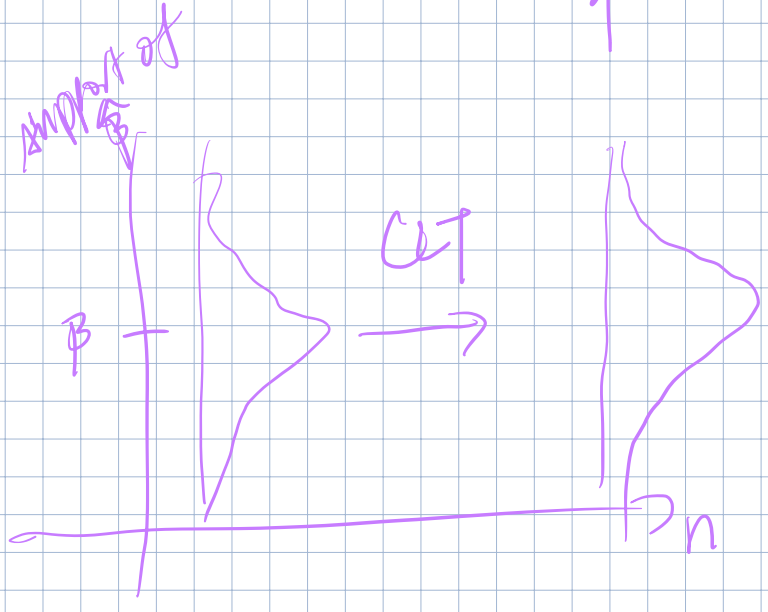
$\Rightarrow$ We think the X_k covariate is predictive of the probability of "success".

$$\ell(\hat{\beta}) \approx \ell(\beta) + (\hat{\beta}-\beta)\ell'(\beta) + \ldots$$

$$(\hat{\beta}-\beta) = \frac{1}{\ell'(\beta)}\left[\ell(\hat{\beta})-\ell(\beta)\right] \leftarrow N(0,1)$$

$$\ell(\beta) = \sum_{i=1}^{n} \log f(y_i|\beta) \xrightarrow{D} N(0, \ldots)$$

iid RV



[II] Deviance & Likelihood Ratio Tests for Reduced & Full Models

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

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

vs

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

$\Updownarrow$

H_0: Reduced model $\Rightarrow \underline{\text{logit}(\hat{\pi}_i)} = \beta_0 + \sum_{j=1}^{r-1} \beta_j x_{ji}$

vs

H_1: Full Model $\Rightarrow \underline{\text{logit}(\hat{\pi}_i)} = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_{ji}$

We will need to construct a test based on log-likelihood → compare log-likelihoods across the fitted models under H_0 & H_1

Deviance

Def: $\text{Dev} = -2\ell(\hat{\beta})$

For ex:

In the OLS regression model, what is deviance?

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

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

$\Rightarrow$

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

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

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

So Deviance is like a generalization of the familiar SSE from OLS.

With this idea, we can consider a statistic based on deviance differences:

$$\Delta D = Dev(\text{reduced}) - Dev(\text{full})$$

Decision Rule:

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

Here $df = p_{Full} - p_{Red} = p - v$

If we reject H_0, we have evidence that the full model is significantly better fit than the reduced model according to likelihood

$\Rightarrow$ include the extra predictors!!

Recall, for logistic regression, the log-likelihood is:

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

$\therefore$ the deviance of a fitted model is just

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

where $\log\left(\hat{\pi}_i / 1-\hat{\pi}_i\right) = \hat{\beta}_0 + \sum_{j=1}^{p-1} \hat{\beta}_j x_{ji}$ ($p-1$ if full, $v-1$ if reduced)

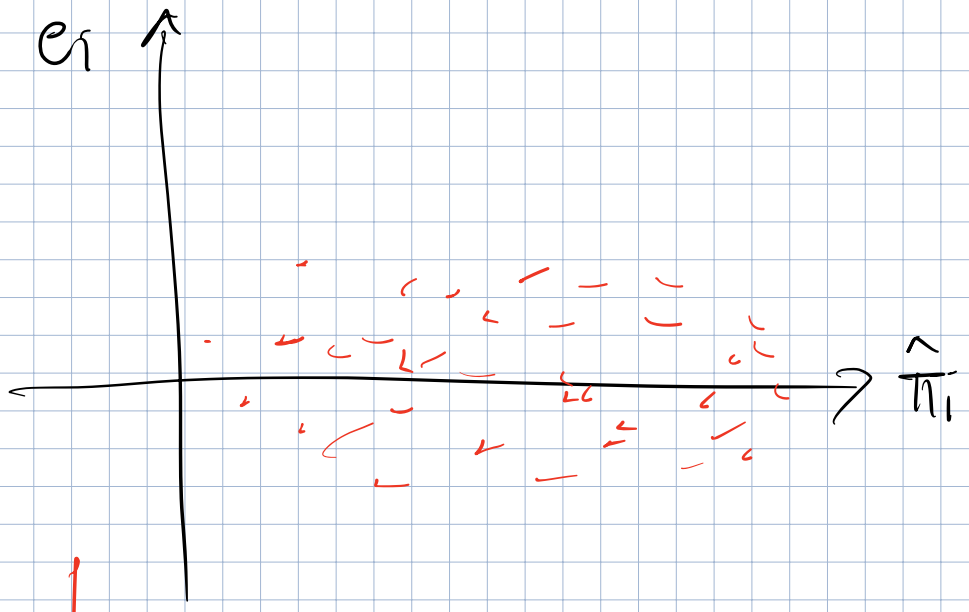
III Deviance Versions of Residuals

① Pearson Residuals:

- feel like standardized residuals in OLS:

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

We could make a "Pearson residual plot" & would hope for no pattern in resids:



↳ Interpret like OLS residuals

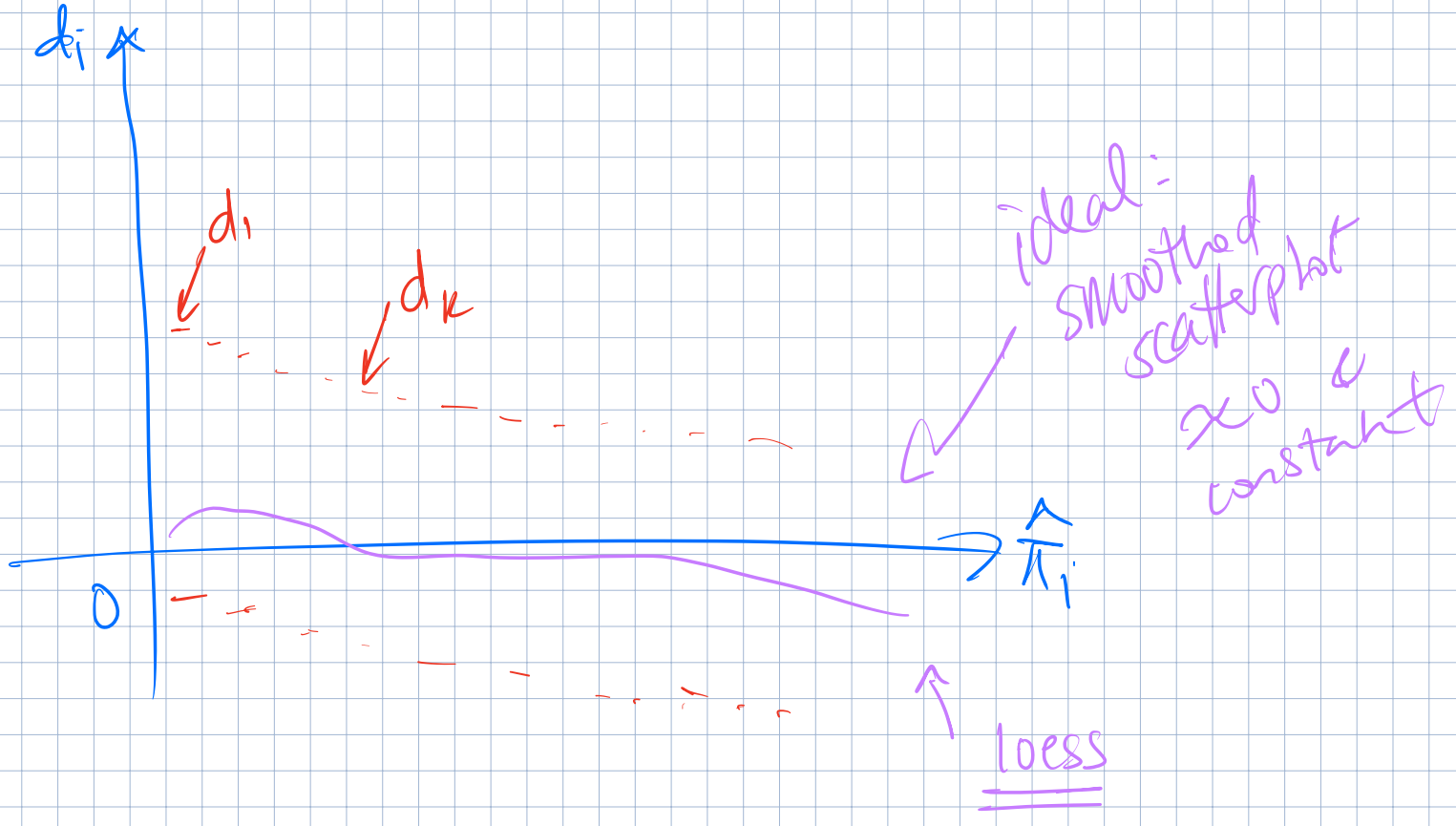
② Deviance Resids

Define 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 square of each deviance residual d_i^2 is basically how much deviance is attributable to the i^{th} obs when considering the deviance of the full sample

For diagnostics, we can look @ scatterplots:



Psendo R^2

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

Some alternative "psendoR^2" stats have been proposed to assess goodness-of-fit:

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

McFadden's $$\text{psendo}R^2 = 1 - \frac{\ell(\text{Full})}{\ell(\text{Null})}$$

Model Selection for Log Reg:

X C_p or Adj R^2

✓ AIC/BIC ok ⇒ b/c they're likelihood based criteria