Generalized Linear Models

Estimation

Learning Outcomes

- Estimation Procedures
  - Regression Coefficients
  - Dispersion Parameter
- Newton-Raphson Algorithm

Estimating: β

Estimating $\boldsymbol{\beta}$

To obtain the estimates of $\boldsymbol{\beta}$ we can use the maximum log-likelihood approach to obtain $\hat{\boldsymbol{\beta}}$.

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f\left(y_i | \boldsymbol{X}_i; \boldsymbol{\beta}, \phi\right)$$

Maximum Likelihood Approach

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log\{f\,(y_i|\boldsymbol{X}_i;\boldsymbol{\beta},\phi)\}$$

$$\{X_i, Y_i\}_{i=1}^{n}$$

$$Y_i \sim \text{Bernoulli}(p)$$

$$\eta_i = g(p_i) = \beta_0 + \beta_1 X_i$$

$$\eta_i = \ln\left(\frac{p_i}{1-p_i}\right)$$

$$p_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}$$

$$f(y_i) = p_i^{y_i}(1-p_i)^{1-y_i}$$

$$L(p_i) = \prod_{i=1}^{n} p_i^{y_i} (1-p_i)^{1-y_i}$$

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} \left(\frac{e^{n_i}}{1+e^{n_i}}\right)^{y_i} \left(1 - \frac{e^{n_i}}{1+e^{n_i}}\right)^{1-y_i}$$

$$\ell(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i \ln\left(\frac{e^{n_i}}{1+e^{n_i}}\right) + (1-y_i) \ln\left(1 - \frac{e^{n_i}}{1+e^{n_i}}\right)$$

$$\frac{1+e^{n_i}}{1+e^{n_i}} - \frac{e^{n_i}}{1+e^{n_i}}$$

$$\frac{1}{1+e^{n_i}}$$

$$= \sum_{i=1}^{n} y_i \left[\ln(e^{n_i}) - \ln(1+e^{n_i})\right] + (1-y_i) \ln\left(\frac{1}{1+e^{n_i}}\right)$$

$$= \ln(1)^{0} - \ln(1+e^{n_i})$$

$$= \sum_{i=1}^{n} Y_i \eta_i - \cancel{Y_i \ln\left(1+e^{\eta_i}\right)} - \ln\left(\frac{1}{1+e^{\eta_i}}\right) + \cancel{Y_i \ln\left(1+e^{\eta_i}\right)}$$

$$= \sum_{i=1}^{n} Y_i \eta_i - \ln\left(\frac{1}{1+e^{\eta_i}}\right)$$

$$\sum_{i=1}^{n} Y_i\left(\beta_0 + \beta_1 X_i\right) - \ln\left(1+e^{\beta_0+\beta_1 X_i}\right)$$

$$\frac{d\ell}{d\beta_0} = \sum_{i=1}^{n} Y_i - \frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}} = 0$$

$$\bar{Y} = \sum \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}.$$

$$\ln(\bar{Y}) = \ln\left(\sum \frac{e^{u_i}}{1+e^{u_i}}\right)$$

β_0 β_1 do not have closed-form solutions

Numerical Approaches

- Newton-Rhapson Algorithm
- Fisher-Scoring Algorithm
- Nelder-Mead
- BFGS

$\hat{\beta}_0$ $\hat{\beta}_1$ minimize $\ell(\beta_0, \beta_1)$

Estimating: ϕ

Estimating ϕ

Depending on the random variable, the dispersion parameter will need to be estimated to conduct inference procedures. There are 4 methods to estimate the dispersion parameter:

- Maximum Likelihood
- Maximum (Modified) Profile Likelihood Approach
- Mean Deviance Estimator
- Pearson Estimator

Maximum Likelihood Approach

$$\ell(\phi) = \sum_{i=1}^{n} \log\{f\,(y_i|\boldsymbol{X}_i;\boldsymbol{\beta},\phi)\}$$

Bernoulli
Poisson
$\phi = 1$

Maximum (Modified) Profile Likelihood Approach

$$\ell_p(\phi) = \frac{p}{2}\log\phi + \sum_{i=1}^{n}\log\{f\left(y_i|\boldsymbol{X}_i;\hat{\boldsymbol{\beta}},\phi\right)\}$$

Mean Deviance Estimator

$$\tilde{\phi} = \frac{D(y, \hat{\mu})}{n - p}$$

- $D(y, \hat{\mu}) = 2 \sum_{i=1}^{n} \{t(y, y) - t(y, \mu)\}$
- $t(y, \mu) = y\theta - \kappa(\theta)$
- p: number of regression coefficients

Pearson Estimator

$$\bar{\phi} = \frac{\Lambda^2}{n - p}$$

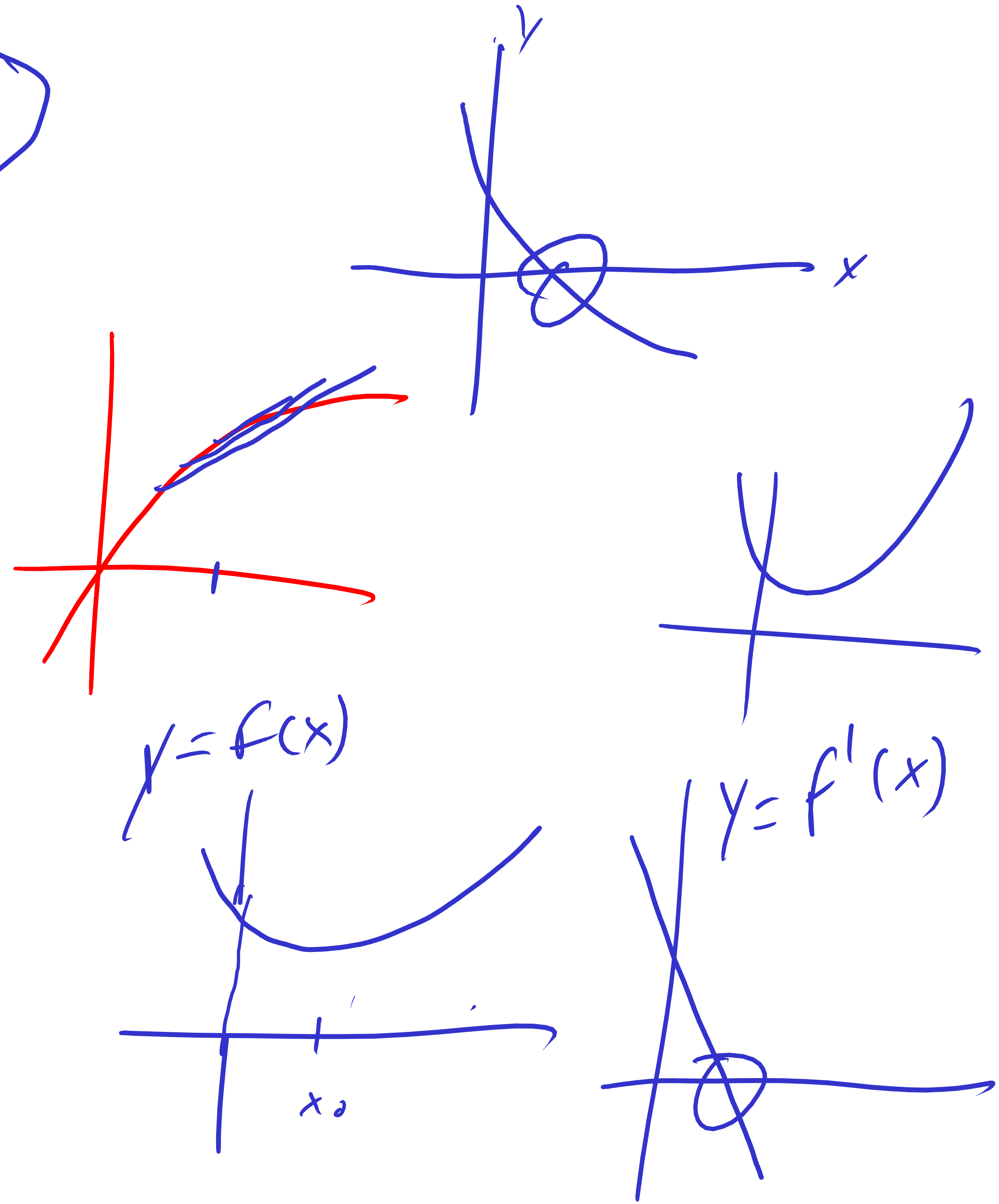
- $\Lambda^2 = \sum_{i=1}^{n} \frac{y_i - \hat{\mu}_i}{V(\hat{\mu}_i)}$
- $\hat{\mu}_i = g^{-1}(\hat{\beta}_0 + \sum_{j=1}^{n} X_{ij}\hat{\beta}_j)$
- $V(\hat{\mu}_i) = \frac{d^2\kappa(\hat{\theta}_i)}{d\theta_i^2}$

Newton-Raphson Algorithm

Numerical Algorithm

In Mathematics and Statistics, numerical algorithms are used to approximate the value of different functions:

- Root Finding:
  - Newton's Method
- Derivatives
  - Secant Step-size
- Integrals
  - Reimman Sums
- Maximization
  - Newton-Raphson



Optimization

Optimization is the techniques used to find the values that maximizes the a function:

$$x_0 = \text{argmax}_x f(x)$$

Newton-Raphson

The Newton-Raphson algorithm is used to estimate the parameters using an iterative algorithm. Given initial estimates, it will update the estimates of the parameters using the Newton step. It will continue iterating and updating the steps until the function converges to the maximum value.

x_{initi}

$\vdots$

x_{old} $$|f(x_{old}) - f(x_{new})| < 10^{-6}$$

x_{new}

Newton-Raphson

$$\beta_j^{(it+1)} = \beta_j^{(it)} - \frac{G_{\beta_j}^{(it)}}{H_{\beta_j}^{(it)}}$$

- $\beta_j^{(it)}$: current estimate of β_j
- $G_{\beta_j}^{(it)} = d\ell(\boldsymbol{\beta})/d\beta_j|_{\beta\_j=\beta_j^{(it)}}$
- $G_{\beta_j}^{(it)} = d^2\ell(\boldsymbol{\beta})/d\beta_j^2|_{\beta_j=\beta_j^{(it)}}$
- $\beta_j^{(it+1)}$: Updated estimate of β_j

Example

Logistic Regression

Let $(Y_i, X_i)_{i=1}^n$ be a data set where $Y_i \overset{iid}{\sim} Bernoulli(p)$. Find the first and second derivative for β_1, when a GLM is fitted to the model.

Poisson Regression

Let $(Y_i, X_i)_{i=1}^n$ be a data set where $Y_i \overset{iid}{\sim} Pois(\lambda)$. Find the first and second derivative for β_0, when a GLM is fitted to the model.

4/27

Del Norte Hall

2555