Generalized Linear Models

Estimation

$$E(Y) = g^{-1}(\beta_0 + \beta_1 X)$$

Linear Regression

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (X^T X)^{-1} X^T Y$$

Normal

Binary

Binomial — $g(p) = \ln\left(\frac{p}{1-p}\right) \leftarrow$ logit

Poisson $g(\lambda) = \log(\lambda)$

$\hat{\beta}_0 = ?$ $\hat{\beta}_1 = ?$

Learning Outcomes

- Estimation Procedures
  - Regression Coefficients $\theta = g(\beta)$
  - Dispersion Parameter ϕ
- Newton-Raphson Algorithm

Scientific Computing

Estimating: β

Estimating $\boldsymbol{\beta}$

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

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

$$f(y_i | X_i)$$

Maximum Likelihood Approach

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

Bird

X_i: wing span

Y_i: $\begin{cases} 1 & \text{is brown} \\ 0 & \end{cases}$

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

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

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

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

$$\log\left(\frac{p}{1-p}\right) = \overbrace{\beta_0 + \beta_1 X_i}^{\eta_i}$$

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

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

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

$$\ell(\beta_0,\beta_1) = \sum \ln\left\{\left(\frac{e^{\eta_i}}{1+e^{\eta_i}}\right)^{y_i} \left(1-\frac{e^{\eta_i}}{1+e^{\eta_i}}\right)^{1-y_i}\right\}$$

$$= \sum y_i \left[\ln(e^{\eta_i}) - \ln(1+e^{\eta_i})\right] +$$

$$q_i = 1+e^{\eta_i}$$

$$(1-y_i)\ln\left(1-\frac{e^{\eta_i}}{1+e^{\eta_i}}\right) \qquad 1-\frac{e^{\eta_i}}{1+e^{\eta_i}} = \frac{1}{1+e^{\eta_i}}$$

$$= \sum y_i \eta_i - y_i \ln(q_i) + (1-y_i)\left[-\ln(q_i)\right]$$

$$= \sum y_i \eta_i \cancel{- y_i \ln(q_i)} - \ln(q_i) \cancel{+ y_i \ln(q_i)}$$

$$= \sum y_i \eta_i - \ln(q_i)$$

$$\ell(\beta_0,\beta_1) = \sum \ln\left\{ \left(\frac{e^{\eta_i}}{1+e^{\eta_i}}\right)^{y_i} \left(1-\frac{e^{\eta_i}}{1+e^{\eta_i}}\right)^{1-y_i} \right\}$$

$$= \sum y_i(\beta_0 + \beta_1 x_i) - \ln\left(1+e^{\beta_0+\beta_1 x_i}\right) = \ell(\beta_0,\beta_1)$$

$$\frac{d\,\ell(\beta_0,\beta_1)}{d\beta_1} = \sum y_i x_i - \frac{x_i e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$$

Numerical Approaches

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

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

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

Maximum (Modified) Profile Likelihood Approach

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

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

$\hat{\sigma}^2$

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

Linear Regression

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

Newton-Raphson Algorithm

Numerical Algorithm

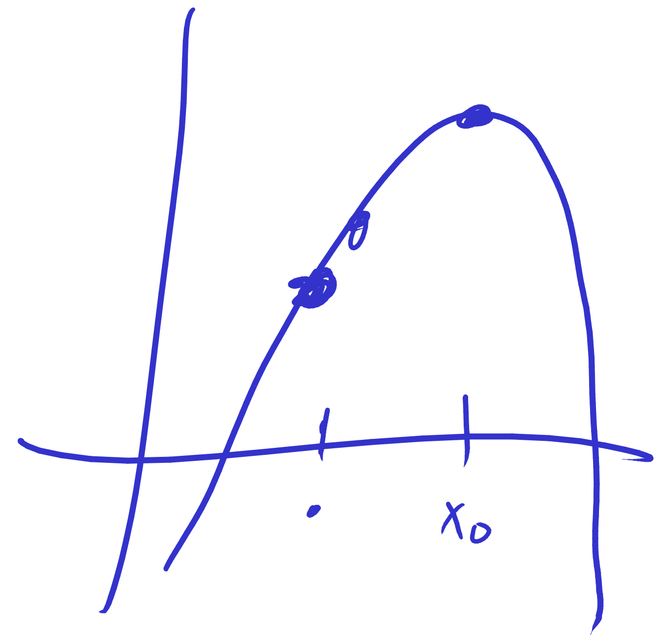
In Mathematics and Statistics, numerical algorithms are used to approximate the value of different functions:

- 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:

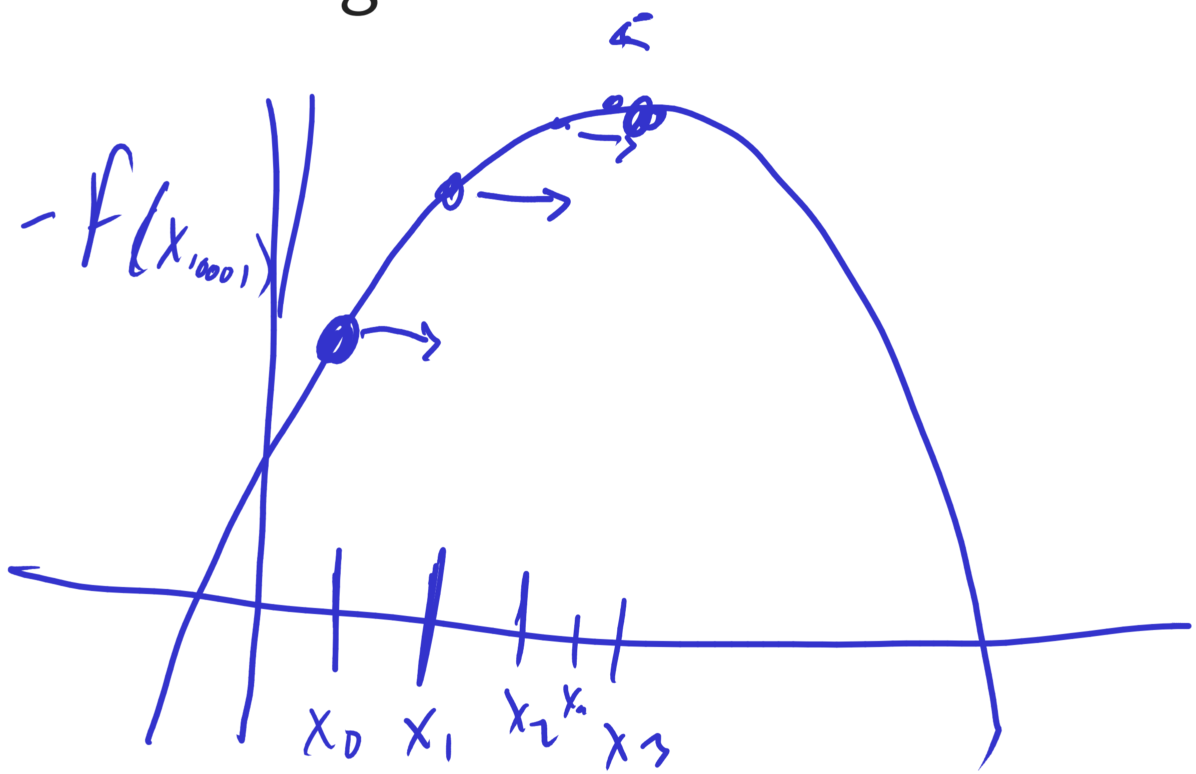
$$x_0 = \text{argmax}_x f(x)$$



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.

$$10^{-5} \geq \left| f(x_{1000}) - f(x_{1001}) \right|$$



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
- $G_{\beta_j}^{(it)} = d\ell(\boldsymbol{\beta})/d\beta_j|_{\beta_{\cancel{\_}}j=\beta_j^{(it)}}$
- $\overset{H}{\cancel{G}}{}_{\beta_j}^{(it)} = d^2\ell(\boldsymbol{\beta})/d\beta_j^2|_{\beta_j=\beta_j^{(it)}}$
- $\beta_j^{(it+1)}$: 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.

$\eta_i = \beta_0 + \beta_1 x_i$

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

$$\ell(\beta_0, \beta_1) = \sum Y_i \eta_i - \ln(q_i)$$

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

$$\frac{d\ell}{d\beta_1} = \sum Y_i x_i - x_i m_i$$

$q_i = 1 - m_i$

$$\frac{d^2\ell}{d\beta_1^2} = \sum -x_i^2 m_i q_i$$

$$\frac{d\, m_i}{d\beta_1} = \frac{e^{n_i}}{1+e^{n_i}} = \frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}} \qquad (f\cdot g)' = f'g + fg'$$

$$e^{\beta_0+\beta_1 X_i}\left(1+e^{\beta_0+\beta_1 X_i}\right)^{-1}$$

$$X_i e^{\beta_0+\beta_1 X_i}\left(1+e^{\beta_0+\beta_1 X_i}\right)^{-1} - e^{\beta_0+\beta_1 X_i}\left(1+e^{\beta_0+\beta_1 X_i}\right)^{-2} e^{\beta_0+\beta_1 X_i} X_i$$

$$X_i\left[\frac{e^{n_i}}{1+e^{n_i}} - \frac{e^{n_i}e^{n_i}}{(1+e^{n_i})^2}\right]$$

$$X_i \left[\frac{e^{n_i}(1+e^{n_i})}{(1+e^{n_i})^2} - \frac{e^{n_i}\, e^{n_i}}{(1+e^{n_i})^2} \right]$$

$$X_i \left[\frac{e^{n_i}(1+e^{n_i}) - (e^{n_i})^2}{(1+e^{n_i})^2} \right]$$

$$X_i \left[\frac{e^{n_i} \cancel{+(e^{n_i})^2} - (e^{n_i})^2}{(1+e^{n_i})^2} \right]$$

$$x_i \frac{e^{h_i}}{(1+e^{h_i})^2}$$

$$q_i = \frac{1}{1+e^{h_i}}$$

$$m_i = \frac{e^{h_i}}{1+e^{h_i}}$$

$$x_i \frac{e^{h_i}}{1+e^{h_i}} \cdot \frac{1}{1+e^{h_i}}$$

$$\sum x_i^2 m_i q_i$$

$$x_i \quad m_i \quad q_i$$

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.

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{Y_i}}{Y_i!}$$

$$f_{Y_i} = \frac{e^{-\lambda} \lambda^{Y_i}}{Y_i!}$$

$$\eta_i = \ln(\lambda) = \beta_0 + \beta_1 X_i$$

$$L(\beta) = \prod_{i=1}^{n} \frac{e^{-e^{\eta_i}} e^{\eta_i Y_i}}{X_i!}$$

$$\lambda_i = e^{\eta_i}$$

$$\ell(\beta) = \sum \ln\left(\frac{e^{-e^{\eta_i}} e^{\eta_i y_i}}{y_i!}\right)$$

$$= \sum \ln\left(e^{-e^{\eta_i}}\right) + \ln\left(e^{\eta_i y_i}\right) - \ln(y_i!)$$

$$\ell(\beta) = \sum -e^{\eta_i} + y_i \eta_i - \ln(y_i!)$$

$$\eta_i = \beta_0 + \beta_1 x_i$$

$$\frac{d\ell}{d\beta_0} = \sum -e^{\eta_i} + y_i$$

$$\frac{d^2\ell}{d\beta_0^2} = \sum -e^{\eta_i}$$

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

$$L(\beta) = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^\alpha} x_i^{\alpha-1} e^{-x_i/\beta}$$

$$\ell(\beta) = \sum -\ln(\Gamma(\alpha)) - \alpha\ln(\beta) + (\alpha-1)\ln(x_i) - x_i/\beta$$

$$\frac{d\ell}{d\beta} = \sum \frac{-\alpha}{\beta} + \frac{x_i}{\beta^2}$$

$$\frac{-n\alpha}{\beta} + \frac{\sum X_i}{\beta^2} = 0$$

$$\beta^2 \frac{\sum X_i}{\beta^2} = \frac{n\alpha}{\cancel{\beta}}\beta^2$$

$$\sum X_i = n\alpha\beta$$

$$\frac{\sum X_i}{n\alpha} = \hat{\beta}$$

$$\frac{\bar{X}}{\alpha}$$