STAT 770 Oct. 7 Lectures
GLMs – Estimating Equations without Likelihoods

Reading and Topics for this lecture: Chapter 4, Secs. 4.4-4.5

(1) Recap of GLM Ingredients (with scalar θ)

(2) Score Equation for GLM using Likelihood

(3) GLM Score Eq'n – Recalling Logistic & Poisson Regression

(4) Using GLM Estimating Eq'n without the Likelihood!

(5) More on fitting `R` models using `glm`

(6) Fitting Logistic & Poisson Regressions

(7) Deviances vs Log LR's

Natural Exponential Families, Scalar θ, $T(y) = y$

probability mass function $\qquad f(y,\theta) \;=\; h(y)\,\exp\Big(\theta y - c(\theta)\Big)$

(1) $$0 = \frac{d}{d\theta}\sum_y f(y,\theta) \;=\; \sum_y \left(y - c'(\theta)\right) f(y,\theta)$$

Therefore $\qquad 0 \;=\; E_\theta\Big(Y - c'(\theta)\Big) \quad \Rightarrow \quad \boxed{c'(\theta) = E_\theta(Y))}$

(2) $$0 \;=\; \frac{d}{d\theta}\sum_y \left(y - c'(\theta)\right) f(y,\theta) \;=\; \sum_y \Big[(y - c'(\theta))^2 - c''(\theta)\Big] f(y,\theta)$$

Therefore $\qquad 0 \;=\; E_\theta\Big[(Y - c'(\theta))^2 - c''(\theta)\Big] \quad \Rightarrow \quad \boxed{c''(\theta) = \mathrm{Var}_\theta(Y))}$

Ingredients and Terminology for GLMs

Y_i outcomes 'satisfying' $Y_i \sim f(y, \theta_i) = \exp\left(\theta_i' y - c(\theta_i)\right) h(y)$

X_i (vector) regressor variables entering model via $\eta_i = \beta' X_i$

μ_i conditional expectation of Y_i given X_i

θ_i monotonically related to $\mu_i = c'(\theta_i)$ through model

$g(\mu_i) = \eta_i$ **link function** g monotonic, smooth

GLM contains relationships $\beta \mapsto \eta_i \mapsto \mu_i \mapsto \theta_i$

specifying likelihood $L(\beta) = \prod_{i=1}^{n} f(Y_i, \theta_i)$

Deriving the GLM Score Equation

$$\eta_i \;=\; X_i'\beta, \quad \mu_i \;=\; g^{-1}(\eta_i), \quad E_{\theta_i}(Y_i \,|\, X_i) \;=\; c'(\theta_i) \;=\; \mu_i$$

$$\nabla_\beta \log f(Y_i,\, \theta_i) \;=\; (\nabla_\beta\, \eta_i) \left(\frac{d\mu_i}{d\eta_i}\,\frac{d\theta_i}{d\mu_i}\right) \frac{\partial}{\partial\theta_i}\Big[\theta_i\, Y_i - c(\theta_i)\Big]$$

$$=\; X_i \left(\frac{d\eta_i}{d\mu_i}\right)^{-1} \left(\frac{d\mu_i}{d\theta_i}\right)^{-1} \Big[Y_i - c'(\theta_i)\Big]$$

$$=\; X_i \Big(g'(\mu_i)\,\cdot\, c''(\theta_i)\Big)^{-1} \Big[Y_i - c'(\theta_i)\Big] \;=\; X_i\, \frac{Y_i - \mu_i}{g'(\mu_i)\,\mathrm{Var}_{\theta_i}(Y_i)}$$

$$\nabla_\beta \log L(\beta) \;=\; \sum_{i=1}^{n} X_i\, \frac{Y_i - \mu_i}{g'(\mu_i)\, v(\mu_i)}$$

where $v(\mu_i) \;\equiv\; c''(\theta_i) \;=\; \mathrm{Var}_{\theta_i}(Y_i), \qquad \mu_i = c'(\theta_i).$

Notes on GLM Score Equation

(I) In some, models, with **canonical link function** $g(\eta)$, the choice satisfies $\eta_i = \theta_i$. This is true in Logistic Regression $\log(p/(1-p)) = \theta = \beta^{tr} X$, and also in Poisson Regression $\log(\lambda) = \theta = \beta^{tr} X$. It is not true in Probit Regression where Y_i is binary, so $\theta = \log(p/(1-p))$, but $\mu_i = p = \Phi(\eta_i)$.

(II) When the link is canonical, the terms $\frac{d\mu_i}{d\eta_i} \frac{d\theta_i}{d\mu_i}$ in the 2nd line of derivation on the previous slide disappear, and $g'(\mu_i)\, c''(\theta_i) \equiv 1$.

The Score Equation becomes $\sum_{i=1}^n X_i \left(Y_i - c'(\beta^{tr} X_i) \right) = 0$.

(III) The Score Equation (to be solved for MLE) does not depend on the model except through its mean μ_i and its variance $v(\mu_i)$ expressed as a function of the mean.

Logistic Regression Estimating Equation

Logistic Regression is a model for binary Y_i given X_i :

$$\text{logit}\, P(Y_i = 1 \,|\, X_i) \;=\; X_i'\beta \;, \qquad \text{logit}(x) = \log\left(\frac{x}{1-x}\right) \quad \text{log-odds}$$

$$\boxed{P(Y_i = 1 \,|\, X_i,\, \beta) \;=\; e^{\beta' X_i}/(1 + e^{\beta' X_i})} \qquad \texttt{plogis}(x) = \frac{e^x}{1+e^x}$$

Data $\{(Y_i, X_i)\}_{i=1}^n$, $\qquad L(\beta) \;=\; \prod_{i=1}^n \left[\left(\frac{e^{\beta' X_i}}{1+e^{\beta' X_i}} \right)^{Y_i} \left(\frac{1}{1+e^{\beta' X_i}} \right)^{1-Y_i} \right]$

logLik $\qquad \log L(\beta) \;=\; \sum_{i=1}^n \left[Y_i\, \beta' X_i \;-\; \log\left(1 + e^{\beta' X_i}\right) \right]$

Equation: $\qquad \boxed{\nabla \log L(\beta) \;=\; \sum_{i=1}^n X_i \left[Y_i - \frac{e^{\beta' X_i}}{1+e^{\beta' X_i}} \right] \;=\; \mathbf{0}}$

Compare least-squares estimating equation !

Poisson Regression Estimating Equation

Poisson Regression: model for Poisson counts Y_i given X_i :

log $\left\{E(Y_i \mid X_i)\right\} = X_i'\beta$, **Poisson rate** $\lambda_i = e^{\beta' X_i}$ for Y_i

$$\boxed{P(Y_i = k \mid X_i,\, \beta) \;=\; e^{-\lambda_i}\,\lambda_i^k /\, k! \;=\; \texttt{dpois}(k, \lambda_i)}$$

Data $\{(Y_i, X_i)\}_{i=1}^n$, $L(\beta) \;= \prod_{i=1}^n \left[\text{exp}\left(-\, e^{\beta' X_i}\right) e^{Y_i\,\beta' X_i}\right]$

logLik $\log L(\beta) \;=\; \sum_{i=1}^n \left[Y_i\,\beta' X_i \;-\; e^{\beta' X_i}\right]$

Equation: $$\boxed{\nabla \log L(\beta) \;=\; \sum_{i=1}^n X_i \left[Y_i \;-\; e^{\beta' X_i}\right] \;=\; \mathbf{0}}$$

Estimating Equation for GLM

Now work backwards. **Assume Y_i conditionally independent given X_i with (conditional) means μ_i satisfying $\mu_i = g(X_i' \beta)$, with β the same for $1 \le i \le n$ and g known.**

Assume (for now) that Var$(Y_i \,|\, X_i) = v(\mu_i)$, with $v(\cdot)$ known.

Idea: estimate β as the solution of

$$\sum_{i=1}^{n} X_i \, \frac{Y_i \, - \, g^{-1}(X_i^{tr}\beta)}{g'(g^{-1}(X_i^{tr}\beta))\, v(g^{-1}(X_i^{tr}\beta))} \; = \; \mathbf{0}$$

by writing $\mu_i = g^{-1}(X_i^{tr}\beta)$.

Recall g^{-1} = logit for Logistic Regression, log for Poisson.

General Idea of Estimating Equations

Suppose *iid* (Y_i, X_i) satisfy $E_{\beta_0}\Big[Q(Y_i, X_i, \beta_0)\Big] \equiv 0$ in model P_β with parameter $\beta = \beta_0$ (and maybe other nuisance parameters).

Law of Large Numbers implies

$$n^{-1} \sum_{i=1}^{n} Q(Y_i, X_i, \beta) \xrightarrow{P_\beta} 0$$

Assume regularity conditions (smoothness and moments) on Q

as in MLE-theory special case $Q(Y_i, X_i, \beta) = \nabla_\beta \log f(Y_i \,|\, X_i, \beta))$:

Q continuously differentiable in β, with $E_\beta\Big(\nabla_\beta \{Q(Y_1, X_1, \beta)\}^{tr}\Big)$ **nonsingular**

Further Steps in Estimating Equation Theory

View $n^{-1} \sum_{i=1}^{n} Q(Y_i, X_i, \beta)$ as random function for $\beta \in B_\epsilon(\beta_0)$

Under regularity conditions, uniformly close to $E_{\beta_0}\Big(Q(Y_1, X_1, \beta)\Big)$

This function is $\mathbf{0}$ at $\beta = \beta_0$, and nonsingularity of gradient implies there exists solution within $B_\epsilon(\beta_0)$.

This can be used also to show that there exist locally consistent estimating equation solutions in P_{β_0} probability. Will take this further next time to develop asymptotic normality and information-like matrices to invert for asymptotic variance.