STAT 770 Oct. 12 Lectures
GLMs – Estimating Equations without Likelihoods

Reading and Topics for this lecture: Chapter 4, Secs. 4.5, 4.7

(1) Recap of GLM Estimating Equations (with scalar θ)

(2) Score Equation & Asymptotic Variance

(3) Deviances and Log LR's

(4) Other Model Examples: Noncanonical, Dispersion

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

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

Recap of GLM from Likelihood

Start from Y_i scalar outcomes $Y_i \sim f(y, \theta_i) = \exp\left(\theta_i\, y - c(\theta_i)\right) h(y)$

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

θ **natural parameter**: book expresses it as function $Q(\theta)$

$E(Y_i \,|\, X_i) = c'(\theta_i) = \mu_i$, Var$(Y_i \,|\, X_i) = c''(\theta_i)$

$$\boxed{\text{Var}(Y_i \,|\, X_i) = v(\mu_i)\ , \quad v = c'' \circ (c')^{-1}}$$

Link function: g monotonic, smooth, known, $g(\mu_i) = \eta_i$

GLM relationships $\beta \mapsto \eta_i \mapsto \mu_i \mapsto \theta_i = (c')^{-1} \circ g^{-1}(X_i^{tr}\beta)$

Score Equation and Information Matrix

Score Equation: $\nabla_\beta \log L(\beta) = \sum_{i=1}^n X_i \frac{Y_i - \mu_i}{g'(\mu_i)\, v(\mu_i)} = \mathbf{0}$

$\mu_i = g^{-1}(X_i^{tr}\beta)$, and if link is **canonical**, $g'(\mu_i)\, v(\mu_i) \equiv 1$

Suppose $\hat{\beta}$ MLE solves GLM score equation. Theory (later in lecture) shows it is consistent: under P_β, as $n \to \infty$, $\hat{\beta} \xrightarrow{P} \beta$

$$0 = \nabla \log L(\hat{\beta}) \approx \nabla \log L(\beta) + \sum_{i=1}^n \nabla_\beta^{\otimes 2} \log f(Y_i, \theta_i(\beta))\, (\hat{\beta} - \beta)$$

So * $$\sqrt{n}\,(\hat{\beta} - \beta) \overset{\mathcal{D}}{\approx} \mathcal{I}^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \frac{Y_i - \mu_i}{g'(\mu_i)\, v(\mu_i)} \right)$$

where $$\mathcal{I} = -\frac{1}{n} E\Big[\sum_{i=1}^n X_i \nabla_\beta^{tr} \frac{Y_i - \mu_i}{g'(\mu_i)\, v(\mu_i)} \Big]$$

Information Matrix Versus Observed Information

Observed Information $J \; = \; -\frac{1}{n}\left[\sum_{i=1}^n X_i \, \nabla_\beta^{tr} \frac{Y_i - \mu_i}{g'(\mu_i)\, v(\mu_i)}\right]\Big|_{\beta=\widehat{\beta}}$

Fisher Information $\mathcal{I} \; = \; -\frac{1}{n}\, E\left[\sum_{i=1}^n X_i \, \nabla_\beta^{tr} \frac{Y_i - \mu_i}{g'(\mu_i)\, v(\mu_i)}\right]$

$$= \frac{-1}{n}\, E\left[\sum_{i=1}^n X_i \,(Y_i - \mu_i)\, \nabla_\beta^{tr} \frac{1}{g'(\mu_i)\, v(\mu_i)}\right] + \frac{1}{n}\, E\left[\sum_{i=1}^n X_i \frac{\nabla \mu_i}{g'(\mu_i)\, v(\mu_i)}\right]$$

Use $E(Y_i \,|\, X_i) = \mu_i$ and $\mu_i = g^{-1}(X_i^{tr}\,\beta)$: then

$\mathcal{I} \; = \; \frac{1}{n} \sum_{i=1}^n X_i\, X_i^{tr} \left[(g'(\mu_i))^2\, v(\mu_i)\right]^{-1}$ for $\{X_i\}$ fixed

Observed information has additional (smaller-order) term involving $Y_i - \mu_i$ [only] when the link is non-canonical

Drawing Conclusions from the Variance Expressions

General case (under regularity conditions): $\mu_i = g^{-1}(X_i^{tr}\beta)$, and

$\widehat{\beta}$ solves $\sum_{i=1}^{n} X_i \frac{Y_i - \mu_i}{g'(\mu_i)\, v(\mu_i)} = 0$, $\qquad \widehat{\mu}_i = g^{-1}(X_i^{tr}\,\widehat{\beta})$

$$\widehat{\beta} - \beta \overset{\mathcal{D}}{\approx} \mathcal{N}\Big(\mathbf{0},\ \Big[\sum_{i=1}^{n} X_i\, X_i^{tr} \left((g'(\widehat{\mu}_i))^2\, v(\widehat{\mu}_i)\right)^{-1}\Big]^{-1}\Big)$$

$$\text{Variance matrix} \quad = \quad \text{Information}^{-1} \; = \; \left(\mathbf{X}^{tr}\, W\, \mathbf{X}\right)^{-1}$$

$\mathbf{X}_{n\times p}$ has i'th row X_i , $\qquad W_{n\times n} = \mathtt{diag}\Big(\big[g'(\widehat{\mu}_i))^2\, v(\widehat{\mu}_i)\big]^{-1}\Big)$

With canonical link: $J = \mathcal{I}\Big|_{\beta=\widehat{\beta}}$ and $\qquad W = \mathtt{diag}(v(\widehat{\mu}_i))$

Deviances and Likelihood Ratio Statistics

deviance *between* models with dim$(\beta) = q$ or p is

$$\boxed{2 \cdot \Big(\log(L(\widehat{\beta}^{(p)}) - \log(L(\widehat{\beta}^{(q)})\Big)}$$

Wilks' Theorem says (under H_0 that $\beta^{(q)}$ is correct) $\sim \chi^2_{p-q}$

(residual) Deviance: *between* current β and Saturated model

Null Model: GLM with intercept only, $q = 1$

Saturated model: GLM with $\widehat{\mu}_i \equiv Y_i$ (examples to follow)

Null Deviance: Deviance between Null and Saturated model

(incremental) Deviance: Analysis of Deviance table line for terms (q+1):p (single compionent or factor) is deviance between $\beta^{(q)}$ and $\beta^{(p)}$ models

R Script with Links, Deviances, Data Structures

(i) Links are arguments to `family` within `glm`

(ii) Null, incremental and residual deviances within `anova` tables

(iii) Residual Deviances useful for *categorical* variables X_i that may include row, column factors and interactions

(iv) Data may be at unit obs. level or aggregated over common X_i categories

(v) Exhibit variances using `$cov.unscaled` component of `glm` fitted object

Other GLMs and Extensions

Non-canonical Link Examples:

(I). Binomial outcome Y_i, $\mu_i \in (0, 1)$, link g^{-1} any dist'n function other than `plogis`, e.g. $g^{-1} = \Phi$ for probit model.

(II). Poisson outcome Y_i, link g^{-1} any monotone map on $\mathbb{R}$, can be identity, e.g. for linear model, if $X_i'\beta$ all positive

Models with Overdispersion: will cover this extension next time

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

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

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

and $\frac{1}{n} \sum_{i=1}^{n} \nabla\{Q(Y_i, X_i, \beta)\}^{tr} \approx E_{\beta_0}\Big(\nabla_\beta \{Q(Y_1, X_1, \beta)\}^{tr}\Big) = A(\beta)$

$M(\beta)$ is $\mathbf{0}$ at $\beta = \beta_0$, and conclude (via *empirical process theory*) that with prob. $\to 1$ there is solution in $B_\epsilon(\beta_0)$.

Nonsingularity of $A(\beta)$ near β_0 implies root of $M(\cdot)$ locally unique: if $\beta^*, \tilde{\beta}$ are solutions in $B_\epsilon(\beta_0)$, then $\mathbf{0} = M(\beta^*) - M(\tilde{\beta})$

$$\approx \Big[E_\beta\Big(\nabla_\beta \{Q(Y_1, X_1, \beta_0)\}^{tr}\Big)\Big]^{tr} (\beta^* - \tilde{\beta}) + o\Big(\|\beta^* - \tilde{\beta}\|\Big)$$