Generalized Linear Models

Objectives:

- Systematic + Random.
- Exponential family.
- Maximum likelihood estimation & inference.

Generalized Linear Models

- Models for independent observations

$$Y_i, \quad i = 1, 2, \ldots, n.$$

- Components of a GLM:
  - ▷ Random component

$$\begin{aligned} Y_i &\sim f(Y_i, \theta_i, \phi) \\ f &\in \text{exponential family} \end{aligned}$$

▷ Systematic component

$$\eta_i = \boldsymbol{X}_i \boldsymbol{\beta}$$

η_i : linear predictor

$\boldsymbol{X}_i$: $(1 \times p)$ covariate vector

$\boldsymbol{\beta}$: $(p \times 1)$ regression coefficient

▷ Link function

$$E(Y_i \mid \boldsymbol{X}_i) = \mu_i$$

$$g(\mu_i) = \boldsymbol{X}_i \boldsymbol{\beta}$$

$g(\cdot)$: link function

Generalized Linear Models

- GLMs generalize the standard linear model:

$$Y_i = \boldsymbol{X}_i \boldsymbol{\beta} + \epsilon_i$$

  - ▷ Random: Normal distribution

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

  - ▷ Systematic: linear combination of covariates

$$\eta_i = \boldsymbol{X}_i \boldsymbol{\beta}$$

  - ▷ Link: identity function

$$\eta_i = \mu_i$$

Generalized Linear Models

- GLMs extend usefully to overdispersed and correlated data:
  - ▷ GEE: marginal models / semi-parametric estimation & inference
  - ▷ GLMM: conditional models / likelihood estimation & inference

Exponential Family

$$(\star) \quad f(y;\theta,\phi) \;=\; \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right]$$

$$\theta \;=\; \text{\underline{canonical} parameter}$$

$$\phi \;=\; \text{fixed (known) scale parameter}$$

Properties: If $Y \sim f(y;\theta,\phi)$ in $(\star)$ then,

$$E(Y) \;=\; \mu \;=\; b'(\theta)$$

$$\text{var}(Y) \;=\; b''(\theta)\cdot a(\phi)$$

Canonical link function: A function $g(\cdot)$ such that:

$$\eta \quad = \quad g(\mu) \quad = \quad \theta \;\; \text{(canonical parameter)}$$

Variance function: A function $V(\cdot)$ such that:

$$\text{var}(Y) \quad = \quad V(\mu) \cdot a(\phi)$$

$$\begin{aligned} \text{Usually} \quad : \quad & a(\phi) = \phi \cdot w \\ & \phi \quad \text{"scale" parameter} \\ & w \quad \text{weight} \end{aligned}$$

Examples of GLMS: logistic regression

$y = s/m$ where s=number of successes / m trials

$$
\begin{aligned}
f(y;\theta,\phi) &= \binom{m}{s}\pi^s(1-\pi)^{m-s} \\
&= \exp\left[\frac{y\cdot\log(\frac{\pi}{1-\pi}) + \log(1-\pi)}{1/m} + \log\binom{m}{s}\right] \\
\Longrightarrow & \\
\theta &= \log(\pi/(1-\pi)) \\
b(\theta) &= -\log(1-\pi) \;=\; \log[1+\exp(\theta)] \\
\mu &= b'(\theta) = \frac{\partial}{\partial\theta}\log[1+\exp(\theta)] \\
&= \exp(\theta)/[1+\exp(\theta)] \;=\; \pi
\end{aligned}
$$

$$
\begin{aligned}
g(\mu) &= \log[\pi/(1-\pi)] = \theta \\
g &: \text{logit,} \quad \text{log-odds function}
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{var}(y) &= \pi(1-\pi) \cdot \frac{1}{m} \\
V(\mu) &= \\
a(\phi) &=
\end{aligned}
$$

- Poisson regression

$$
\begin{aligned}
y &= \text{number of events (count)} \\
\\
f(y;\theta,\phi) &= \lambda^y \exp(-\lambda)/y! \\
&= \exp\left[y \cdot \log(\lambda) - \lambda - \log(y!)\right] \\
\Longrightarrow & \\
\theta &= \log(\lambda) \\
b(\theta) &= \lambda \;=\; \exp(\theta) \\
\mu &= b'(\theta) \;=\; \exp(\theta) \;=\; \lambda \\
\\
g(\mu) &= \theta = \log(\mu) \\
g &: \text{ canonical link is } \log
\end{aligned}
$$

- Poisson regression (continued)

$$\text{var}(y) = \lambda$$

$$V(\mu) =$$

$$a(\phi) =$$

◦ Other examples:

▷ gamma, inverse Gaussian (MN, Table 2.1)

▷ some survival models (MN, Chpt. 13)

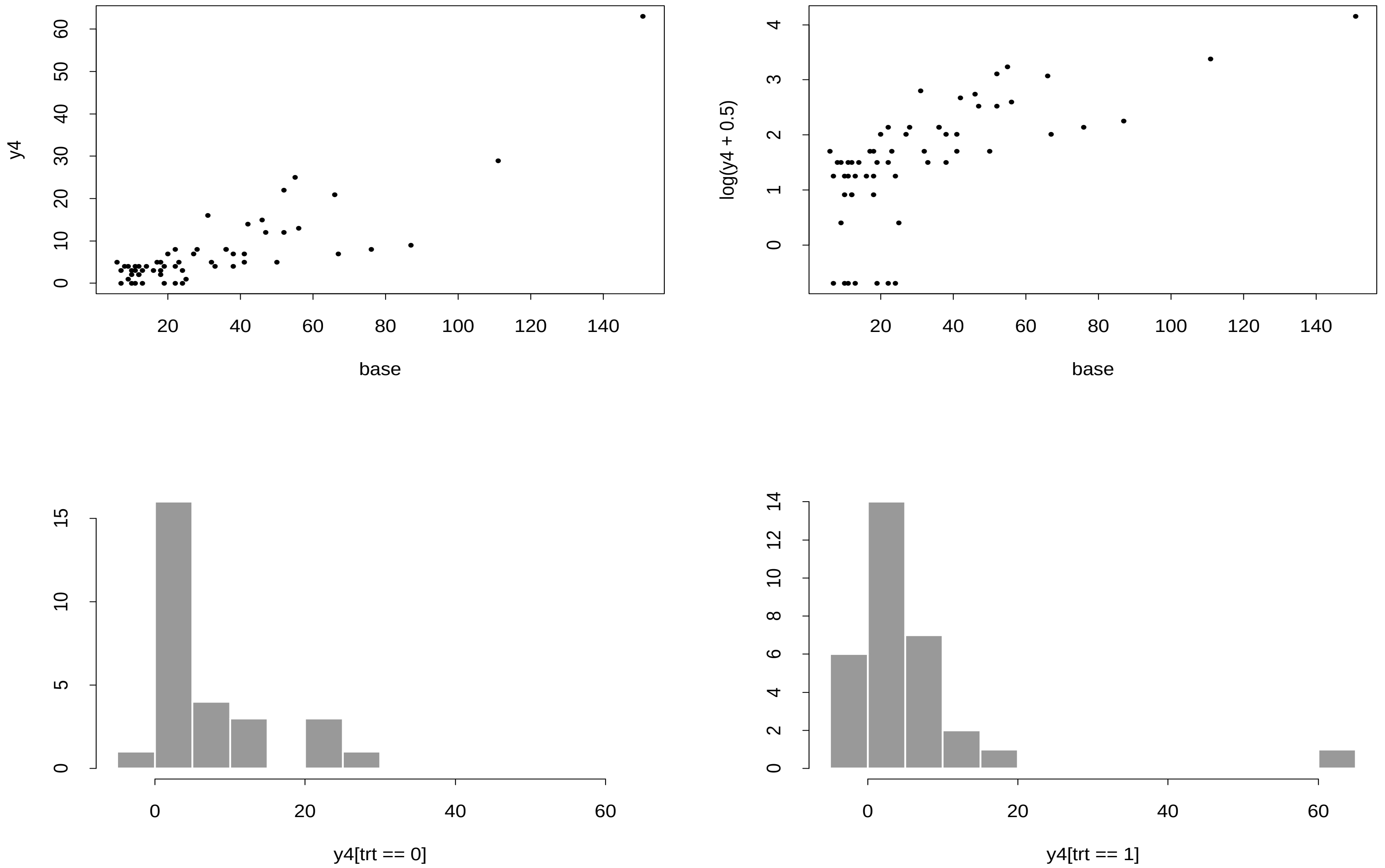
Example: Seizure data (DHLZ ex. 1.6)

- Clinical trial of progabide, evaluating impact on epileptic seizures.
- Data:
  - ▷ age = patient age in years
  - ▷ base = 8-week baseline seizure count (pre-tx)
  - ▷ tx = 0 if assigned placebo; 1 if assigned progabide
  - ▷ Y_1, Y_2, Y_3, Y_4 seizure counts in 4 two-week periods following treatment administration
- Models:
  - ▷ linear model: $Y_4 = \texttt{age} + \texttt{base} + \texttt{tx} + \epsilon$
  - ▷ Poisson GLM: $\log(\mu_4) = \texttt{age} + \texttt{base} + \texttt{tx}$

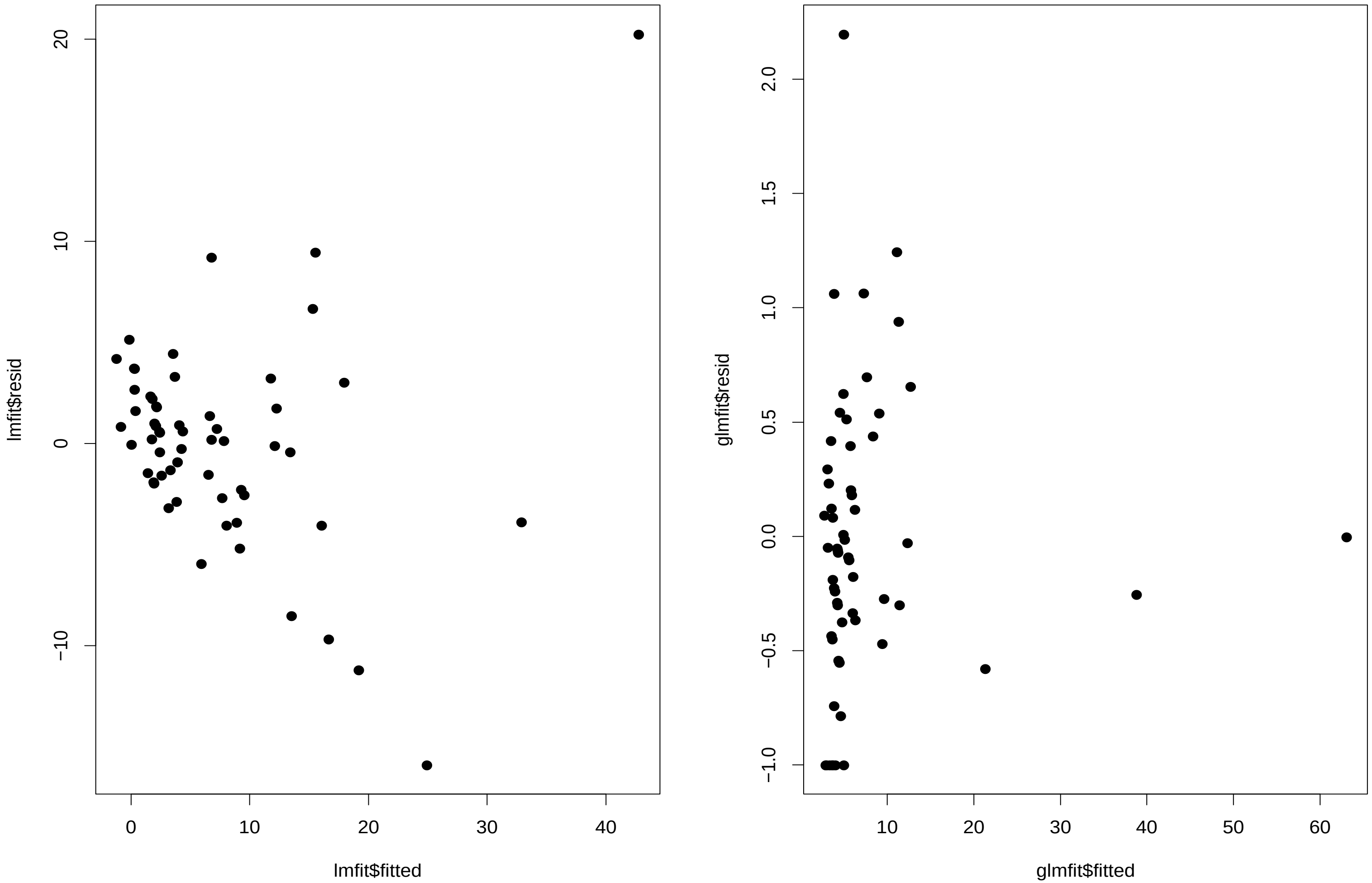
Example: Seizure data (DHLZ ex. 1.6)

	linear regression			Poisson regression		
	est.	s.e.	Z	est.	s.e.	Z
(Int)	-4.97	3.62	-1.37	0.778	0.285	2.73
age	0.12	0.11	1.07	0.014	0.009	1.64
base	0.31	0.03	11.79	0.022	0.001	20.27
tx	-1.36	1.37	-0.99	-0.270	0.102	-2.66

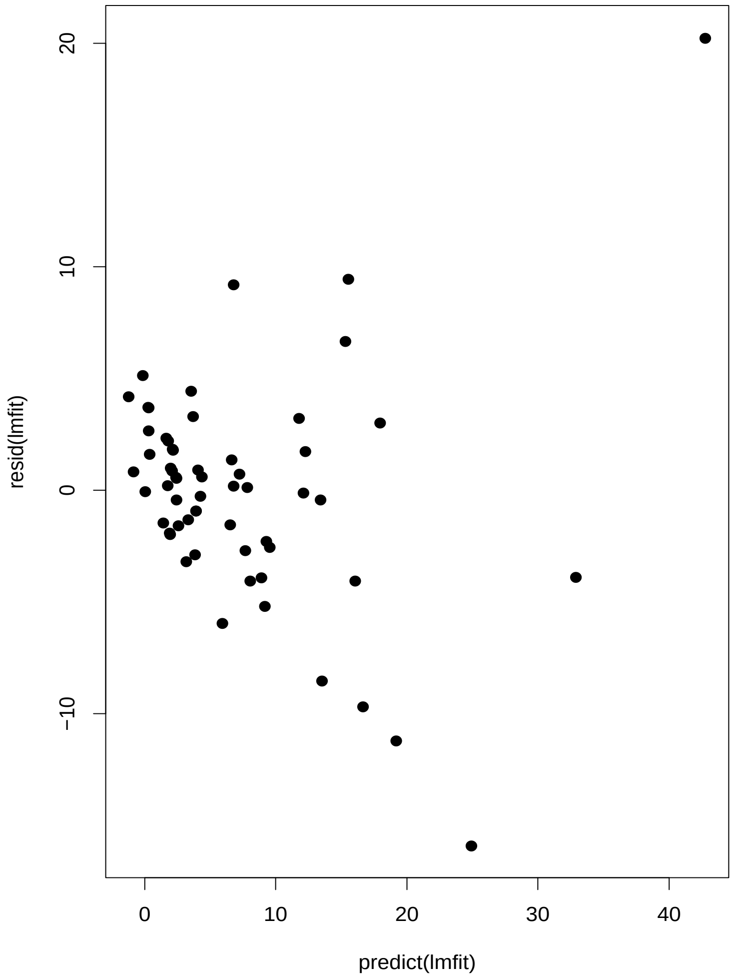
- **Q**: should we use `log(base)` for Poisson regression?
- **Q**: why does inference regarding significance of `TX` differ?

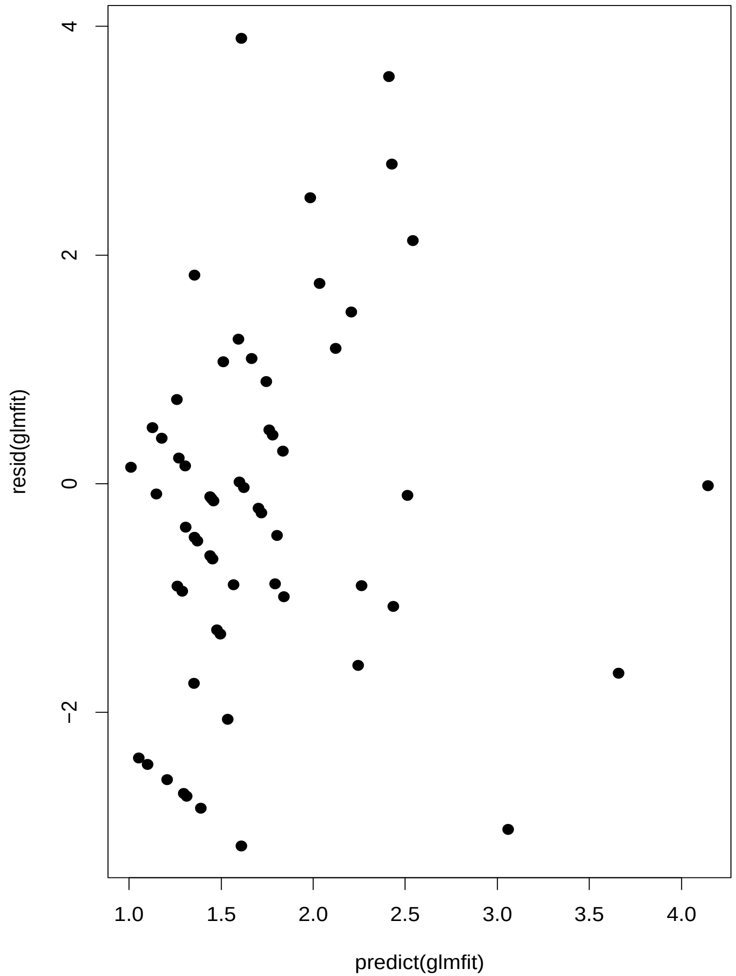

y4
base
log(y4 + 0.5)
base
y4[trt == 0]
y4[trt == 1]

Seizure Residuals vs. Fitted



Seizure Residuals vs. Fitted, using `predict()`





Residual Diagnostics

◦ Used to assess model fit similarly as for linear models

- Q-Q plots for residuals
  (may be hard to interpret for discrete data)
- residual plots:
  - ⋆ vs. fitted values
  - ⋆ vs. omitted covariates
- assessment of systematic departures
- assessment of variance function

Residual Diagnostics

◦ Types of residuals for GLMs:

1. Pearson residual

$$r^P = \frac{y_i - \widehat{\mu}_i}{\sqrt{V(\widehat{\mu}_i)}}$$

$$\sum (r_i^P)^2 = X^2$$

2. Deviance residual (see `resid(fit)`)

$$r_i^D = \text{sign}(y - \widehat{\mu})\sqrt{d_i}$$

$$\sum (r_i^D)^2 = D(y, \widehat{\mu})$$

3. Working residual (see `fit$resid`)

$$r_i^W = (y_i - \widehat{\mu}_i)\frac{\partial \widehat{\eta}_i}{\partial \widehat{\mu}_i} = Z_i - \widehat{\eta}_i$$

Fitting GLMS by Maximum Likelihood

Solve score equations:

$$U_j(\boldsymbol{\beta}) = \frac{\partial}{\partial \beta_j} \log L = 0 \quad j = 1, 2, \ldots, p$$

log-likelihood:

$$\begin{aligned}
\log L &= \sum_{i=1}^{n} \left[\frac{y_i \cdot \theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi) \right] \\
&= \sum \log L_i
\end{aligned}$$

$$\Longrightarrow$$

$$U_j(\boldsymbol{\beta}) = \frac{\partial \log L}{\partial \beta_j} = \sum_i \frac{\partial \log L_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j}$$

$$
\begin{aligned}
\frac{\partial \log L_i}{\partial \theta_i} &= \frac{1}{a_i(\phi)}\,(y_i - b'(\theta_i)) \;=\; \frac{1}{a_i(\phi)}\,(y_i - \mu_i) \\
\frac{\partial \theta_i}{\partial \mu_i} &= \left(\frac{\partial \mu_i}{\partial \theta_i}\right)^{-1} \;=\; 1/V(\mu_i) \\
\frac{\partial \eta_i}{\partial \beta_j} &= X_{ij}
\end{aligned}
$$

Therefore,

$$
U_j(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n}\left(X_{ij}\frac{\partial \mu_i}{\partial \eta_i}\right)\cdot\left[a_i(\phi)\cdot V(\mu_i)\right]^{-1}(Y_i - \mu_i)
$$

GLM Information Matrix

- Either form:

$$[\mathcal{I}_n](j,k) = \text{cov}[U_j(\boldsymbol{\beta}), U_k(\boldsymbol{\beta})]$$

$$= -E\left(\frac{\partial^2 \log L}{\partial \beta_j \partial \beta_k}\right)$$

- Let's consider the second form...

GLM Information Matrix

$$
\begin{aligned}
[\mathcal{I}_n](j,k) &= -E\left[\frac{\partial}{\partial \beta_k} U_j(\boldsymbol{\beta})\right] \\
&= -E\left[\sum_{i=1}^{n} \frac{\partial}{\beta_k}\left\{\left(\frac{\partial \mu_i}{\partial \beta_j}\right) \cdot [a_i(\phi) \cdot V(\mu_i)]^{-1} (Y_i - \mu_i)\right\}\right] \\
&= \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \beta_j}\right) \cdot [a_i(\phi) \cdot V(\mu_i)]^{-1} \left(\frac{\partial \mu_i}{\partial \beta_k}\right)
\end{aligned}
$$

justify

Score and Information

- In vector/matrix form we have:

$$
\begin{aligned}
\boldsymbol{U}(\boldsymbol{\beta}) &= \begin{pmatrix} U_1(\boldsymbol{\beta}) \\ U_2(\boldsymbol{\beta}) \\ \vdots \\ U_p(\boldsymbol{\beta}) \end{pmatrix} \\
\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} &= \begin{pmatrix} \frac{\partial \mu_i}{\partial \beta_1} & \frac{\partial \mu_i}{\partial \beta_2} & \cdots & \frac{\partial \mu_i}{\partial \beta_p} \end{pmatrix} \\
&= \boldsymbol{X}_i \frac{\partial \mu_i}{\partial \eta_i}
\end{aligned}
$$

Score and Information

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right)^T \cdot [a_i(\phi) \cdot V(\mu_i)]^{-1} (Y_i - \mu_i)$$

and

$$\mathcal{I}_n = \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right)^T \cdot [a_i(\phi) \cdot V(\mu_i)]^{-1} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right)$$

Fisher Scoring

Goal: Solve the score equations

$$\boldsymbol{U}(\boldsymbol{\beta}) = \mathbf{0}$$

Iterative estimation is required for most GLMs. The score equations can be solved using Newton-Raphson (uses observed derivative of score) or **Fisher Scoring** which uses the expected derivative of the score (ie. $-\mathcal{I}_n$).

Fisher Scoring

Algorithm:

- Pick an initial value: $\widehat{\boldsymbol{\beta}}^{(0)}$.
- For $j \rightarrow (j+1)$ update $\widehat{\boldsymbol{\beta}}^{(j)}$ via

$$\widehat{\boldsymbol{\beta}}^{(j+1)} = \widehat{\boldsymbol{\beta}}^{(j)} + \left(\widehat{\mathcal{I}}_n^{(j)}\right)^{-1} \boldsymbol{U}(\widehat{\boldsymbol{\beta}}^{(j)})$$

- Evaluate convergence using changes in $\log L$ or $\|\widehat{\boldsymbol{\beta}}^{(j+1)} - \widehat{\boldsymbol{\beta}}^{(j)}\|$.
- Iterate until convergence criterion is satisfied.

Comments on Fisher Scoring:

1. IWLS is equivalent to Fisher Scoring (Biostat 570).
2. Observed and expected information are equivalent for canonical links.
3. Score equations are an example of an **estimating function** (more on that to come!)
4. **Q**: What assumptions make $E[\boldsymbol{U}(\boldsymbol{\beta})] = \mathbf{0}$?
5. **Q**: What is the relationship between $\mathcal{I}_n$ and $\sum \boldsymbol{U}_i \boldsymbol{U}_i^T$?
6. **Q**: What is a 1-step approximation to $\Delta\boldsymbol{\beta}^{(-i)}$?

Inference for GLMs

Review of asymptotic likelihood theory:

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ ---\\ \boldsymbol{\beta}_2 \end{pmatrix} = \begin{matrix} (q \times 1) \\ --- \\ (p - q \times 1) \end{matrix}$$

Goal: Test $\boxed{H_0 \ : \ \boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^0}$

(1) **Likelihood Ratio Test**:

$$2\left[\log L(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2) - \log L(\widehat{\boldsymbol{\beta}}_1^0, \boldsymbol{\beta}_2^0)\right] \sim \chi^2(df = p - q)$$

Inference for GLMs

(2) **Score Test**:

$$\boldsymbol{U}(\boldsymbol{\beta}) = \begin{pmatrix} \boldsymbol{U}_1(\boldsymbol{\beta}_1) \\ --- \\ \boldsymbol{U}_2(\boldsymbol{\beta}_2) \end{pmatrix} = \begin{matrix} (q \times 1) \\ --- \\ (p - q \times 1) \end{matrix}$$

$$\boldsymbol{U}_2(\widehat{\boldsymbol{\beta}}^0)^T \left\{ \text{cov}[\boldsymbol{U}_2(\widehat{\boldsymbol{\beta}}^0)] \right\}^{-1} \boldsymbol{U}_2(\widehat{\boldsymbol{\beta}}^0) \sim \chi^2(df = p - q)$$

(3) **Wald Test**:

$$(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0)^T \left\{ \text{cov}(\widehat{\boldsymbol{\beta}}_2) \right\}^{-1} (\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0) \sim \chi^2(df = p - q)$$

Measures of Discrepancy

There are 2 primary measures:

- deviance
- Pearson's X^2

Deviance: Assume $a_i(\phi) = \phi/m_i$

(eg. normal: ϕ; binomial: $1/m_i$; Poisson: 1)

$$\begin{aligned} \log L(\widehat{\boldsymbol{\beta}}) &= \sum_{i=1}^{n} \log f_i(y_i; \widehat{\theta}_i, \phi) \\ &= \sum_{i} \left\{ \frac{m_i}{\phi} [y_i \widehat{\theta}_i - b(\widehat{\theta}_i)] + c_i(y_i, \phi) \right\} \end{aligned}$$

Now consider $\log L$ as a function of $\widehat{\boldsymbol{\mu}}$, using the relationship $b'(\theta) = \mu$:

$$l(\widehat{\boldsymbol{\mu}}, \phi; \boldsymbol{y}) = \sum_i \left\{ \frac{m_i}{\phi} [y_i \cdot \theta(\widehat{\mu}_i) - b[\theta(\widehat{\mu}_i)]\ ] + c_i(y_i, \phi) \right\}$$

The <u>deviance</u> is:

$$\begin{aligned} D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) &= 2 \cdot \phi \cdot [l(\boldsymbol{y}, \phi; \boldsymbol{y}) - l(\widehat{\boldsymbol{\mu}}, \phi; \boldsymbol{y})] \\ &= 2 \cdot \sum_i m_i \{\ \ y_i \cdot [\theta(y_i) - \theta(\widehat{\mu}_i)] - \\ &\qquad (\ b[\theta(y_i)] - b[\theta(\widehat{\mu}_i)]\ \ )\ \ \} \end{aligned}$$

Deviance

Deviance generalizes the residual sum of squares for linear models:

Model 1		Model 2
$\left(\begin{array}{c} \widehat{\boldsymbol{\beta}}_1 \\ ---- \\ \widehat{\boldsymbol{\beta}}_2 \end{array}\right)$	$\begin{array}{c} (q \times 1) \\ --- \\ (p - q \times 1) \end{array}$	$\left(\begin{array}{c} \widehat{\boldsymbol{\beta}}_1 \\ ---- \\ \boldsymbol{\beta}_2^0 \end{array}\right)$
$\widehat{\boldsymbol{\mu}}_1$		$\widehat{\boldsymbol{\mu}}_2$

Deviance

Linear Model:

$$\frac{SSE(\text{Model 2}) - SSE(\text{Model 1})}{\sigma^2} \sim \chi^2(df = p - q)$$

GLM:

$$\frac{D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}_2) - D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}_1)}{\phi} \sim \chi^2(df = p - q)$$

Examples:

1. Normal: $\log f(y_i; \theta_i, \phi) = -\frac{(y_i - \mu_i)^2}{2\phi} + C$

$$D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = \sum_i (y_i - \widehat{\mu}_i)^2 \ = \ SSE$$

2. Poisson: $\log f(y_i; \theta_i, \phi) = y_i \cdot \log(\mu) - \mu + C$

$$D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = 2 \times \left[\sum_i y_i \cdot \log\left(\frac{y_i}{\widehat{\mu}_i}\right) - (y_i - \widehat{\mu}_i) \right]$$

3. Binomial: $\log f(y_i; \theta_i, \phi) = m_i \left[y_i \cdot \log\left(\frac{\mu}{1-\mu}\right) + \log(1-\mu) \right]$

$$D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = 2 \times \left[\sum_i y_i \cdot \log\left(\frac{y_i}{\widehat{\mu}_i}\right) - (1 - y_i) \cdot \log\left(\frac{1 - y_i}{1 - \widehat{\mu}_i}\right) \right]$$

Pearson's X^2

Assume: $\text{var}(Y_i) = \frac{\phi}{m_i} V(\mu_i)$

Define: $X^2 = \sum_i (y_i - \widehat{\mu}_i)^2 / [V(\widehat{\mu}_i)/m_i]$

Examples:

1. Normal: $X^2 = SSE$
2. Poisson: $X^2 = (y_i - \widehat{\mu}_i)^2 / \widehat{\mu}_i$ (look familiar?)
3. Binomial: $X^2 = (y_i - \widehat{\mu}_i)^2 / [\widehat{\mu}_i (1 - \widehat{\mu}_i)]$

($\star\star$) If the model is correct (mean and variance) then,

$$\frac{X^2}{(n-p)} \approx \phi$$

e.g.

- Normal: $SSE/(n-p) \approx \sigma^2 = \phi$
- Poisson: $X^2/(n-p) \approx 1 = \phi$
- Binomial: $X^2/(n-p) \approx 1 = \phi$

Example: Seizure data (DLZ ex. 1.6)

$$\frac{X^2}{(n-p)} = \frac{136.64}{59-4} = 2.48$$

$$\frac{D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}})}{(n-p)} = \frac{147.02}{59-4} = 2.67$$

Q: Poisson???

Summary:

- GLMs applicable to range of univariate outcomes.
- Systematic variation (regression)
  Random variation (variance function, likelihood)
- Score equations of simple form.
- Inference using:
  likelihood ratios (deviance)
  score statistics
  Wald statistics
- Model checking
  regression structure / variance form $V(\mu)$

References:

McCullagh P., and Nelder J.N. *Generalized Linear Models, Second Edition*, Chapman and Hall, 1989.

Fahrmeir L., and Tutz G. *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer-Verlag, 1991.
(see chapter 2)

McCulloch C.E., and Searle S.R. *Generalized, Linear, and Mixed Models*, Wiley, 2001.
(see chapter 5)

Diggle P., Heagerty P.J., Liang K-Y., Zeger S.L. *Longitudinal Data Analysis, Second Edition*, Oxford, 2002.
(see appendix A)