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

  \[ Y_i \sim f(Y_i, \theta_i, \phi) \]

  \[ f \in \text{exponential family} \]
Systematic component

\[ \eta_i = X_i \beta \]

\( \eta_i \) : linear predictor

\( X_i \) : \((1 \times p)\) covariate vector

\( \beta \) : \((p \times 1)\) regression coefficient

Link function

\[ E(Y_i \mid X_i) = \mu_i \]

\[ g(\mu_i) = X_i \beta \]

\( g(\cdot) \) : link function
Generalized Linear Models

- GLMs generalize the standard linear model:

\[ Y_i = X_i \beta + \epsilon_i \]

▷ Random: Normal distribution

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

▷ Systematic: linear combination of covariates

\[ \eta_i = X_i \beta \]

▷ 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

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

\[
\begin{align*}
\theta & = \text{canonical parameter} \\
\phi & = \text{fixed (known) scale parameter}
\end{align*}
\]

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

\[
\begin{align*}
E(Y) & = \mu = b'(\theta) \\
\text{var}(Y) & = b''(\theta) \cdot a(\phi)
\end{align*}
\]
**Canonical link function:** A function $g(\cdot)$ such that:

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

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

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

Usually : $a(\phi) = \phi \cdot w$

$\phi$ "scale" parameter

$w$ weight

Heagerty, Bio/Stat 571
Examples of GLMS: logistic regression

\[ y = \frac{s}{m} \quad \text{where} \quad s = \text{number of successes} / m \text{ trials} \]

\[
f(y; \theta, \phi) = \binom{m}{s} \pi^s (1 - \pi)^{m-s} \]

\[
= \exp \left[ \frac{y \cdot \log \left( \frac{\pi}{1-\pi} \right) + \log(1 - \pi)}{1/m} + \log \left( \binom{m}{s} \right) \right]
\]

\[
\implies \quad \theta = \log \left( \frac{\pi}{1-\pi} \right) \]

\[
b(\theta) = -\log(1 - \pi) = \log[1 + \exp(\theta)]
\]

\[
\mu = \frac{\partial}{\partial \theta} \log[1 + \exp(\theta)] = \exp(\theta) / [1 + \exp(\theta)] = \pi
\]
\[ g(\mu) = \log[\pi/(1 - \pi)] = \theta \]

\( g : \) logit, log-odds function

\[ \text{var}(y) = \pi(1 - \pi) \cdot \frac{1}{m} \]

\[ V(\mu) = \]

\[ a(\phi) = \]
Poisson regression

\[ y = \text{number of events (count)} \]

\[
f(y; \theta, \phi) = \lambda^y \exp(-\lambda)/y!
= \exp \left[ y \cdot \log(\lambda) - \lambda - \log(y!) \right]
\]

\[ \theta = \log(\lambda) \]

\[ b(\theta) = \lambda = \exp(\theta) \]

\[ \mu = b'(\theta) = \exp(\theta) = \lambda \]

\[ g(\mu) = \theta = \log(\mu) \]

\[ g : \text{canonical link is log} \]
• 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:
  - $\Delta$ age = patient age in years
  - $\Delta$ base = 8-week baseline seizure count (pre-tx)
  - $\Delta$ 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**: $Y_4 = \text{age} + \text{base} + \text{tx} + \epsilon$
  - **Poisson GLM**: $\log(\mu_4) = \text{age} + \text{base} + \text{tx}$
### Example: Seizure data (DHLZ ex. 1.6)

<table>
<thead>
<tr>
<th></th>
<th>linear regression</th>
<th>Poisson regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est.</td>
<td>s.e.</td>
</tr>
<tr>
<td>(Int)</td>
<td>-4.97</td>
<td>3.62</td>
</tr>
<tr>
<td>age</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td>base</td>
<td>0.31</td>
<td>0.03</td>
</tr>
<tr>
<td>tx</td>
<td>-1.36</td>
<td>1.37</td>
</tr>
</tbody>
</table>

- **Q**: should we use \( \log(\text{base}) \) for Poisson regression?
- **Q**: why does inference regarding significance of TX differ?
Seizure Residuals vs. Fitted
Seizure Residuals vs. Fitted, using \texttt{predict()}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{seizure_residuals.png}
\end{figure}
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

     \[ r^P_i = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} \]

     \[ \sum (r^P_i)^2 = X^2 \]

  2. Deviance residual (see \texttt{resid(fit)})

     \[ r^D_i = \text{sign}(y - \hat{\mu}) \sqrt{d_i} \]

     \[ \sum (r^D_i)^2 = D(y, \hat{\mu}) \]

  3. Working residual (see \texttt{fit$resid}):

     \[ r^W_i = (y_i - \hat{\mu}_i) \frac{\partial \hat{\eta}_i}{\partial \hat{\mu}_i} = Z_i - \hat{\eta}_i \]
Fitting GLMS by Maximum Likelihood

Solve score equations:

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

log-likelihood:

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

\[
= \sum \log L_i
\]

\[ \Rightarrow \]

\[ U_j(\beta) = \frac{\partial \log L}{\partial \beta_j} = \sum_i \frac{\partial \log L_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j} \]
\[
\frac{\partial \log L_i}{\partial \theta_i} = \frac{1}{a_i(\phi)} (y_i - b'(\theta_i)) = \frac{1}{a_i(\phi)} (y_i - \mu_i)
\]

\[
\frac{\partial \theta_i}{\partial \mu_i} = \left( \frac{\partial \mu_i}{\partial \theta_i} \right)^{-1} = 1/V(\mu_i)
\]

\[
\frac{\partial \eta_i}{\partial \beta_j} = X_{ij}
\]

Therefore,

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

- Either form:

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

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

- Let's consider the second form...
GLM Information Matrix

\[ [I_n](j, k) = -E \left[ \frac{\partial}{\partial \beta_k} U_j(\beta) \right] \]

\[ = -E \left[ \sum_{i=1}^{n} \frac{\partial}{\beta_k} \left\{ \left( \frac{\partial \mu_i}{\partial \beta_j} \right) \cdot [a_i(\phi) \cdot V(\mu_i)]^{-1} (Y_i - \mu_i) \right\} \right] \]

\[ = \sum_{i=1}^{n} \left( \frac{\partial \mu_i}{\partial \beta_j} \right) \cdot [a_i(\phi) \cdot V(\mu_i)]^{-1} \left( \frac{\partial \mu_i}{\partial \beta_k} \right) \]

justify
Score and Information

- In vector/matrix form we have:

\[
U(\beta) = \begin{pmatrix}
U_1(\beta) \\
U_2(\beta) \\
\vdots \\
U_p(\beta)
\end{pmatrix}
\]

\[
\frac{\partial \mu_i}{\partial \beta} = \begin{pmatrix}
\frac{\partial \mu_i}{\partial \beta_1} & \frac{\partial \mu_i}{\partial \beta_2} & \cdots & \frac{\partial \mu_i}{\partial \beta_p}
\end{pmatrix}
\]

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

and

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

Goal: Solve the score equations

\[ U(\beta) = 0 \]

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. \(-I_n\)).
Fisher Scoring

Algorithm:

- Pick an initial value: \( \hat{\beta}^{(0)} \).

- For \( j \to (j + 1) \) update \( \hat{\beta}^{(j)} \) via

\[
\hat{\beta}^{(j+1)} = \hat{\beta}^{(j)} + \left( \hat{I}_n^{(j)} \right)^{-1} U(\hat{\beta}^{(j)})
\]

- Evaluate convergence using changes in \( \log L \) or \( \| \hat{\beta}^{(j+1)} - \hat{\beta}^{(j)} \| \).

- 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 $E[U(\beta)] = 0$?
5. Q: What is the relationship between $\mathcal{I}_n$ and $\sum U_i U_i^T$?
6. Q: What is a 1-step approximation to $\Delta \beta^{(-i)}$?
Inference for GLMs

Review of asymptotic likelihood theory:

\[ \beta = \begin{pmatrix} \beta_1 \\ - \\ - \\ - \\ \beta_2 \end{pmatrix} = \begin{pmatrix} (q \times 1) \\ (p - q \times 1) \end{pmatrix} \]

Goal: Test \( H_0 : \beta_2 = \beta_2^0 \)

(1) Likelihood Ratio Test:

\[ 2 \left[ \log L(\hat{\beta}_1, \hat{\beta}_2) - \log L(\hat{\beta}_1^0, \hat{\beta}_2^0) \right] \sim \chi^2(df = p - q) \]
Inference for GLMs

(2) Score Test:

\[
U(\beta) = \begin{pmatrix}
U_1(\beta_1) \\
\vdots \\
U_2(\beta_2)
\end{pmatrix} = \begin{pmatrix}
(q \times 1) \\
\vdots \\
(p - q \times 1)
\end{pmatrix}
\]

\[
U_2(\hat{\beta}^0)^T \left\{ \text{cov}[U_2(\hat{\beta}^0)] \right\}^{-1} U_2(\hat{\beta}^0) \sim \chi^2(df = p - q)
\]

(3) Wald Test:

\[
(\hat{\beta}_2 - \beta_2^0)^T \left\{ \text{cov}(\hat{\beta}_2) \right\}^{-1} (\hat{\beta}_2 - \beta_2^0) \sim \chi^2(df = p - q)
\]
Measures of Discrepancy

There are 2 primary measures:
- deviance
- Pearson’s $X^2$

Deviance: Assume $a_i(\phi) = \phi/m_i$
(eg. normal: $\phi$; binomial: $1/m_i$; Poisson: 1)

$$
\log L(\hat{\beta}) = \sum_{i=1}^{n} \log f_i(y_i; \hat{\theta}_i, \phi)
= \sum_{i} \left\{ \frac{m_i}{\phi} [y_i\hat{\theta}_i - b(\hat{\theta}_i)] + c_i(y_i, \phi) \right\}
$$
Now consider $\log L$ as a function of $\hat{\mu}$, using the relationship $b'(\theta) = \mu$:

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

The deviance is:

$$D(y, \hat{\mu}) = 2 \cdot \phi \cdot [l(y, \phi; y) - l(\hat{\mu}, \phi; y)]$$

$$= 2 \cdot \sum_i m_i \left\{ y_i \cdot [\theta(y_i) - \theta(\hat{\mu}_i)] - (b[\theta(y_i)] - b[\theta(\hat{\mu}_i)]) \right\}$$
Deviance generalizes the residual sum of squares for linear models:

\[
\begin{align*}
\text{Model 1} & : & \left( \begin{array}{c} \hat{\beta}_1 \\ - - - \\ \hat{\beta}_2 \end{array} \right) & \quad \begin{array}{c} (q \times 1) \\ \text{Model 2} & : & \left( \begin{array}{c} \hat{\beta}_1 \\ - - - \\ 0 \end{array} \right) \end{array} & \quad \begin{array}{c} (p - q \times 1) \\ \beta^0_2 \end{array} \\
\hat{\mu}_1 & & & \hat{\mu}_2
\end{align*}
\]
Deviance

**Linear Model:**

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

**GLM:**

\[
\frac{D(y, \hat{\mu}_2) - D(y, \hat{\mu}_1)}{\phi} \sim \chi^2(df = p - q)
\]
Examples:

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

   \[
   D(y, \hat{\mu}) = \sum_i (y_i - \hat{\mu}_i)^2 = \text{SSE}
   \]

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

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

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

   \[
   D(y, \hat{\mu}) = 2 \times \left[ \sum_i y_i \cdot \log \left( \frac{y_i}{\hat{\mu}_i} \right) - (1 - y_i) \cdot \log \left( \frac{1 - y_i}{1 - \hat{\mu}_i} \right) \right]
   \]
Pearson’s $X^2$

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

**Define:**

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

**Examples:**

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

(⋆⋆) If the model is correct (mean and variance) then,

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

- Normal: \( \frac{SSE}{(n - p)} \approx \sigma^2 = \phi \)

- Poisson: \( \frac{X^2}{(n - p)} \approx 1 = \phi \)

- Binomial: \( \frac{X^2}{(n - p)} \approx 1 = \phi \)
Example: Seizure data (DLZ ex. 1.6)

\[
\frac{X^2}{n - p} = \frac{136.64}{59 - 4} = 2.48
\]

\[
\frac{D(y, \hat{\mu})}{n - p} = \frac{147.02}{59 - 4} = 2.67
\]

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 $V(\mu)$

(see chapter 2)

(see chapter 5)

(see appendix A)