



ETC3580: Advanced Statistical Modelling

Week 6: Generalized Linear Models

Outline

1 Exponential family distributions

2 Generalized Linear Models

3 GLM Diagnostics

4 Additional distributions

Exponential family distributions

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

- θ is canonical parameter for location
- ϕ is dispersion parameter for scale
- a , b and c are functions.

Exponential family distributions

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

- θ is canonical parameter for location
- ϕ is dispersion parameter for scale
- a , b and c are functions.

Example: Normal

$$f(y|\theta, \phi) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(y - \mu)^2}{2\sigma^2} \right]$$

$$\theta = \mu \quad \phi = \sigma^2 \quad a(\phi) = \phi \quad b(\theta) = \theta^2/2$$

$$c(y, \phi) = -(y^2/\phi + \log(2\pi\phi))/2$$

Exponential family distributions

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

- θ is canonical parameter for location
- ϕ is dispersion parameter for scale
- a , b and c are functions.

Example: Poisson

$$f(y|\theta, \phi) = e^{-\mu} \mu^y / y!$$

What are θ , ϕ , a , b and c ?

Exponential family distributions

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

- θ is canonical parameter for location
- ϕ is dispersion parameter for scale
- a , b and c are functions.

Example: Binomial

$$f(y|\theta, \phi) = \binom{m}{y} p^y (1-p)^{m-y}$$

What are θ , ϕ , a , b and c ?

Exponential family distributions

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

- θ is canonical parameter for location
- ϕ is dispersion parameter for scale
- a , b and c are functions.

Examples: Normal, Poisson, Binomial,
gamma, inverse Gaussian

Moments

- 1 Mean: $b'(\theta)$
- 2 Variance: $b''(\theta)a(\phi)$

Outline

1 Exponential family distributions

2 Generalized Linear Models

3 GLM Diagnostics

4 Additional distributions

Generalized Linear Models

A GLM consists of three components:

- 1 Distribution (from the exponential family of distributions)
- 2 Linear predictors
- 3 Link function

Link functions

- The predictors are assumed to affect the response through a linear relationship.
- The link function g “links” the mean to the linear predictors.

$$g(\mu) = \beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q$$

Link functions

- The predictors are assumed to affect the response through a linear relationship.
- The link function g “links” the mean to the linear predictors.

$$g(\mu) = \beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q$$

- g must be monotone, continuous and differentiable.
- g must map the space of μ to \mathbb{R} .
- Canonical link has $g(\mu) = \theta$.

Link functions

Family	Canonical link	Variance
Normal	μ	1
Poisson	$\log \mu$	μ
Binomial	$\log(\mu/(1 - \mu))$	$\mu(1 - \mu)$
Gamma	$1/\mu$	μ^2
Inverse Gaussian	$1/\mu^2$	μ^3

- Canonical link means $\mathbf{X}'\mathbf{y}$ is *sufficient*.
- Makes estimation easy and inference exact.

Log Likelihood

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

Log Likelihood

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

$$\begin{aligned} \ell(\beta; \mathbf{y}) = \log L(\beta; \mathbf{y}) &= \sum_{i=1}^n \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right] \\ &= \frac{1}{a(\phi)} \sum_{i=1}^n [y_i \theta_i - b(\theta_i) + c(y_i, \phi) a(\phi)] \end{aligned}$$

Log Likelihood

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

$$\begin{aligned}\ell(\beta; \mathbf{y}) = \log L(\beta; \mathbf{y}) &= \sum_{i=1}^n \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right] \\ &= \frac{1}{a(\phi)} \sum_{i=1}^n [y_i \theta_i - b(\theta_i) + c(y_i, \phi) a(\phi)]\end{aligned}$$

$$\begin{aligned}\frac{\partial \ell(\beta; \mathbf{y})}{\partial \beta_j} &= \frac{1}{a(\phi)} \sum_{i=1}^n \frac{\partial}{\partial \beta_j} [y_i \theta_i - b(\theta_i)] \\ &= \frac{1}{a(\phi)} \sum_{i=1}^n \left[y_i \frac{\partial \theta_i}{\partial \beta_j} - \frac{\partial b(\theta_i)}{\partial \beta_j} \right]\end{aligned}$$

Log Likelihood

Now

$$\frac{\partial b(\theta)}{\partial \beta_j} = b'(\theta) \frac{\partial \theta}{\partial \beta_j} \quad \text{and} \quad \frac{\partial \theta}{\partial \beta_j} = \frac{\partial \theta}{\partial \mu} \frac{\partial \mu}{\partial \beta_j} = \frac{1}{b''(\theta)} \frac{\partial \mu}{\partial \beta_j}$$

Therefore

$$\begin{aligned} \frac{\partial \ell(\beta; \mathbf{y})}{\partial \beta_j} &= \frac{1}{a(\phi)} \sum_{i=1}^n \left[y_i \frac{1}{b''(\theta_i)} \frac{\partial \mu_i}{\partial \beta_j} - \frac{b'(\theta_i)}{b''(\theta_i)} \frac{\partial \mu_i}{\partial \beta_j} \right] \\ &= \frac{1}{a(\phi)} \sum_{i=1}^n \left[\frac{y_i - b'(\theta_i)}{b''(\theta_i)} \right] \frac{\partial \mu_i}{\partial \beta_j} \\ &= \sum_{i=1}^n \left[\frac{y_i - \mu_i}{V(\mu_i)} \right] \frac{\partial \mu_i}{\partial \beta_j} \end{aligned}$$

Maximum likelihood estimation

Maximum likelihood estimates:

$$\sum_{i=1}^n \left[\frac{y_i - b'(\theta_i)}{V(\mu_i)} \right] \frac{\partial \mu_i}{\partial \beta_j} = 0 \quad \text{for all } j$$

Maximum likelihood estimation

Maximum likelihood estimates:

$$\sum_{i=1}^n \left[\frac{y_i - b'(\theta_i)}{V(\mu_i)} \right] \frac{\partial \mu_i}{\partial \beta_j} = 0 \quad \text{for all } j$$

Same equations as for weighted least squares with known $V(\mu)$:

$$\text{Minimize} \quad \sum_{i=1}^n \left[\frac{(y_i - \mu_i)^2}{V(\mu_i)} \right]$$

Maximum likelihood estimation

Maximum likelihood estimates:

$$\sum_{i=1}^n \left[\frac{y_i - b'(\theta_i)}{V(\mu_i)} \right] \frac{\partial \mu_i}{\partial \beta_j} = 0 \quad \text{for all } j$$

Same equations as for weighted least squares with known $V(\mu)$:

$$\text{Minimize} \quad \sum_{i=1}^n \left[\frac{(y_i - \mu_i)^2}{V(\mu_i)} \right]$$

- When $V(\mu)$ is constant, OLS = MLE
- We can guess μ , and use iterated WLS to solve.
- Distribution not used, only $g(\mu)$ and $V(\mu)$.
- So same equations work for quasi-likelihood

Implementation in R

```
fit <- glm(y ~ x1 + x2, family, data)
```

family options

```
binomial(link = "logit")
```

```
gaussian(link = "identity")
```

```
Gamma(link = "inverse")
```

```
inverse.gaussian(link = "1/mu^2")
```

```
poisson(link = "log")
```

```
quasi(link = "identity", variance = "constant")
```

```
quasibinomial(link = "logit")
```

```
quasipoisson(link = "log")
```

Implementation in R

```
fit <- glm(y ~ x1 + x2, family, data)
```

link options

identity

log

inverse

logit

probit

cauchit

cloglog

sqrt

$1/\mu^2$

power

Question

What is the difference between these?

```
lm(y ~ x)
```

and

```
glm(y ~ x)
```

Question

What is the difference between these?

```
lm(log(y) ~ x)
```

and

```
glm(y ~ x, family=gaussian(link='log'))
```

Deviances

GLM

Deviance: $D = -2 \log L$

Gaussian

$$\sum (y_i - \hat{\mu}_i)^2$$

Poisson

$$2 \sum \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right]$$

Binomial

$$2 \sum \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) + (m - y_i) \log \left(\frac{m - y_i}{m - \hat{\mu}_i} \right) \right]$$

Gamma

$$2 \sum \left[-\log \left(\frac{y_i}{\hat{\mu}_i} \right) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right]$$

Inverse Gaussian

$$\sum \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2 y_i}$$

Hypothesis tests

Goodness-of-fit test

- Does data fit assumed distribution?
- Deviance has χ^2 distribution with $df = n - \#$ estimated parameters
- Only works for large n and for distributions with no dispersion parameter.
- Does not work for binary GLM, Gaussian LM, or any quasi family
- Binary does not need checking
- For Gaussian, look at residuals
- For quasi family, it depends ...

Hypothesis tests

Comparing nested models

Large model Ω ; small model ω

- Change in Deviance ($D_\omega - D_\Omega$) equivalent to log-ratio test and has χ^2 distribution with $\text{df} = \text{df}_\omega - \text{df}_\Omega =$ difference in number of parameters
- For quasi-likelihood, use an F approximation instead (exact for Gaussian).

$$F = \frac{D_\omega - D_\Omega}{\hat{\phi}(\text{df}_\omega - \text{df}_\Omega)} \quad \text{where} \quad \hat{\phi} = \frac{1}{n - p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\mu_i)}$$

Outline

1 Exponential family distributions

2 Generalized Linear Models

3 GLM Diagnostics

4 Additional distributions

GLM Residuals

Response residuals: Observation - estimate

$$e_i = y_i - \hat{\mu}_i$$

GLM Residuals

Response residuals: Observation – estimate

$$e_i = y_i - \hat{\mu}_i$$

Pearson residuals: Standardized

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$

GLM Residuals

Response residuals: Observation – estimate

$$e_i = y_i - \hat{\mu}_i$$

Pearson residuals: Standardized

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$

Deviance residuals: Signed root contribution to $-2 \log L$.

$$-2 \log L = \sum \delta_i$$

$$d_i = \text{sign}(y_i - \hat{\mu}_i) \sqrt{\delta_i}$$

GLM Residuals

Response residuals

```
augment(fit, type.residuals='response')
```

Pearson residuals

```
augment(fit, type.residuals='pearson')
```

Deviance residuals

```
augment(fit, type.residuals='deviance')
```

```
augment(fit)
```

GLM Leverage and Influence

IRWLS algorithm used for estimation means that we can easily define the hat matrix:

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}^{1/2}$$

where \mathbf{W} is diagonal with values $\frac{1}{V(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$.

- Leverage values are diagonals of \mathbf{H} .
- `augment(fit) %>% select(.hat)`

GLM Leverage and Influence

IRWLS algorithm used for estimation means that we can easily define the hat matrix:

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}^{1/2}$$

where \mathbf{W} is diagonal with values $\frac{1}{v(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$.

- Leverage values are diagonals of \mathbf{H} .

- `augment(fit) %>% select(.hat)`

Cooks Distance

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' (\mathbf{X}' \mathbf{W} \mathbf{X}) (\hat{\beta}_{(i)} - \hat{\beta})}{p \hat{\phi}}$$

- `augment(fit) %>% select(.cooksd)`

Case checking

- Outliers (large residuals)
- High leverage points (large effect on estimates)

Model checking

- Heteroskedasticity
- Linearity
- Distribution

Diagnostics

Case checking

- Why do we have outliers? Perhaps omit them?
- Reduce leverage through transforming predictors

Model checking

- Heteroskedasticity: perhaps use weights?
- Linearity:
 - transform predictors
 - add quadratic or other transformed variable
 - use nonparametric regressor (later)
- Distribution:
 - allow for overdispersion using quasilikelihood
 - zero-inflated
 - often fixing hetero and linearity will fix distribution

Outline

1 Exponential family distributions

2 Generalized Linear Models

3 GLM Diagnostics

4 Additional distributions

Gamma GLM

Defined on \mathbb{R}^+

Gamma distribution

$$f(y) = \frac{1}{\Gamma(\nu)} \lambda^\nu y^{\nu-1} e^{-\lambda y}$$

- ν describes shape; λ describes scale
- χ^2 is special case ($\lambda = 0.5$, $\text{df} = 2\nu$)

Gamma GLM

Defined on \mathbb{R}^+

Gamma distribution

$$f(y) = \frac{1}{\Gamma(\nu)} \lambda^\nu y^{\nu-1} e^{-\lambda y}$$

- ν describes shape; λ describes scale
- χ^2 is special case ($\lambda = 0.5$, $\text{df} = 2\nu$)

Reparameterize with $\lambda = \nu/\mu$:

$$f(y) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^\nu y^{\nu-1} e^{-y\nu/\mu}$$

- ν describes shape
- μ is mean; $\text{Var}(Y) = \mu^2/\nu$

Gamma GLM

Defined on \mathbb{R}^+

Gamma distribution

$$f(y) = \frac{1}{\Gamma(\nu)} \lambda^\nu y^{\nu-1} e^{-\lambda y}$$

- ν describes shape; λ describes scale
- χ^2 is special case ($\lambda = 0.5$, $\text{df} = 2\nu$)

Reparameterize with $\lambda = \nu/\mu$:

$$f(y) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^\nu y^{\nu-1} e^{-y\nu/\mu}$$

- ν describes shape
- μ is mean; $\text{Var}(Y) = \mu^2/\nu$

```
glm(y ~ x1+x2, family=Gamma(link='log'))
```

- Canonical link is inverse. Better to use log.
- When variance small, it is very similar to Gaussian model with logged response.
- Inference sensitive to distributional mis-specification

Inverse Gaussian GLM

Defined on \mathbb{R}^+

Inverse Gaussian distribution

$$f(y) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp \left[-\lambda(y - \mu)^2 / 2\mu^2 y \right]$$

- μ = mean; $\text{Var} = \mu^3 / \lambda$
- $\mu = 1$ is special case (Wald distribution)

Inverse Gaussian GLM

Defined on \mathbb{R}^+

Inverse Gaussian distribution

$$f(y) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp \left[-\lambda(y - \mu)^2 / 2\mu^2 y \right]$$

- μ = mean; $\text{Var} = \mu^3 / \lambda$
- $\mu = 1$ is special case (Wald distribution)
- Canonical link is $1/\mu^2$
- Variance increases with μ more rapidly than Gamma.
- As $\lambda \rightarrow \infty$, distribution converges to Gaussian.
- First derived by Schrödinger

Tweedie distribution

Exponential family distribution where $\text{Var}(Y) = a\mu^p$, $a > 0, p > 0$.

- normal distribution, $p = 0$
- Poisson distribution, $p = 1$
- compound Poisson–gamma distribution, $1 < p < 2$
- gamma distribution, $p = 2$
- positive stable distributions, $2 < p < 3$
- inverse Gaussian distribution, $p = 3$
- positive stable distributions, $p > 3$
- extreme stable distributions, $p = \infty$

For $0 < p < 1$ no Tweedie model exists.

Compound Poisson-gamma distribution

$$Y = \sum_{i=1}^N X_i, \quad N \sim \text{Poisson}, \quad X_i \sim \text{Gamma}$$

- Continuous on $[0, \infty]$ with a spike at 0.
(e.g., rainfall, insurance payouts.)
- Tweedie distribution with $1 < p < 2$.
- Poisson mean: $\mu^{2-p}/[(2-p)\phi]$.
- Gamma parameters: $\nu = (2-p)/(p-1)$,
 $\lambda = 1/[\phi(p-1)\mu^{p-1}]$

Compound Poisson-gamma distribution

$$Y = \sum_{i=1}^N X_i, \quad N \sim \text{Poisson}, \quad X_i \sim \text{Gamma}$$

- Continuous on $[0, \infty]$ with a spike at 0.
(e.g., rainfall, insurance payouts.)
- Tweedie distribution with $1 < p < 2$.
- Poisson mean: $\mu^{2-p}/[(2-p)\phi]$.
- Gamma parameters: $\nu = (2-p)/(p-1)$,
 $\lambda = 1/[\phi(p-1)\mu^{p-1}]$

Show mean = μ .

Show var = $\phi\mu^p$

Compound Poisson-gamma distribution

$$E(N) = \text{Var}(N) = \frac{\mu^{2-p}}{(2-p)\phi}$$

$$E(X) = \frac{(2-p)}{(p-1)} \left(\phi(p-1)\mu^{p-1} \right) = \phi(2-p)\mu^{p-1}$$

$$\begin{aligned}\text{Var}(X) &= \frac{(2-p)}{(p-1)} \left(\phi^2(p-1)^2\mu^{2(p-1)} \right) \\ &= \phi^2(2-p)(p-1)\mu^{2(p-1)}\end{aligned}$$

Compound Poisson-gamma distribution

$$\begin{aligned} E(Y) &= E_N[Y \mid N] \\ &= E_N[NE(X)] \\ &= E(N)E(X) \\ &= \frac{\mu^{2-p}}{(2-p)\phi} [\phi(2-p)\mu^{p-1}] \\ &= \mu \end{aligned}$$

Compound Poisson-gamma distribution

$$\begin{aligned}\text{Var}(Y) &= \text{Var}_N[\text{E}_X(Y \mid N)] + \text{E}_N[\text{Var}_X(Y \mid N)] \\&= \text{Var}_N[N\text{E}(X)] + \text{E}_N[N\text{Var}(X)] \\&= \text{Var}(N)[\text{E}(X)]^2 + \text{E}(N)\text{Var}(X) \\&= \text{E}(N) \left([\text{E}(X)]^2 + \text{Var}(X) \right) \\&= \frac{\mu^{2-p}}{(2-p)\phi} \left(\phi^2(2-p)^2\mu^{2(p-1)} \right. \\&\quad \left. + \phi^2(2-p)(p-1)\mu^{2(p-1)} \right) \\&= \frac{\mu^{2-p+2p-2}\phi^2(2-p)}{(2-p)\phi} \left((2-p) + (p-1) \right) \\&= \phi\mu^p\end{aligned}$$

Compound Poisson-gamma GLM

```
mgcv::gam(y ~ x1 + x2,  
           family=tw(link="log"))
```

- Estimates p assuming it is in $(1, 2)$.

Tweedie GLMs

- No R function for general Tweedie GLM.
- When using R, user needs to choose
 - $p = 0$ (Gaussian)
`lm` or `glm`
 - $p = 1$ (Poisson)
`glm(family=poisson)`
 - $1 < p < 2$ (Compound Poisson gamma)
`mgcv::gam(family=tw)`
 - $p = 2$ (Gamma)
`glm(family=Gamma)`
 - $p = 3$ (Inverse Gaussian)
`glm(family=inverse.gaussian`