北京航空航天大学
经济管理学院
BEIHANG UNIVERSITY
SCHOOL OF ECONOMICS AND MANAGEMENT

Generalized Linear Models

Lecture 6: Generalized Linear Models

Outline

Introduction

- **Model structure** $\rightarrow$ the patterns of interactions and associations.
- **Model parameters** $\rightarrow$ strength of associations.
- **Focus** $\rightarrow$ estimating the model parameters. The basic inference tools (e.g., point estimation, hypothesis testing, and confidence intervals) will be applied to these parameters.

Introduction

When discussing models, we will keep in mind:

- Objective
- Model structure (e.g. variables, formula, equation)
- Model assumptions
- Parameter estimates and interpretation
- Model fit (e.g. tests and statistics)
- Model selection

Linear regression models

- Objective: model the expected value of a continuous variable Y, as a linear function of the continuous predictor X, $E(Y_i) = \beta_0 + \beta_1 X_i$.
- Model structure: $Y_i = \beta_0 + \beta_1 X_i + e_i$.
- Model assumptions: Y is is normally distributed, errors are normally distributed, $e_i \sim N(0, \sigma^2)$, and independent, and X is fixed, and constant variance σ^2.
- Parameter estimates and interpretation: $\hat{\beta}_0$ is estimate of β_0 or the intercept, and $\hat{\beta}_1$ is estimate of the slope, etc. Think about the interpretation of the intercept and the slope.
- Model fit: R^2, residual analysis, F-statistic.
- Model selection: From a plethora of possible predictors, which variables to include?

Generalized linear models (GLM)

- GLM usually refers to conventional linear regression models for a **continuous** response variable given continuous and/or categorical predictors.
- The form is $y_i \sim N(x_i^T\beta, \sigma^2)$, where x_i contains known covariates and β contains the coefficients to be estimated. These models are usually fit by least squares and weighted least squares.
- y_i is assumed to follow an **exponential family distribution** with mean μ_i, which is assumed to be some (often nonlinear) function of $x_i^T\beta$.

GLM assumptions

- The data $Y_1, Y_2, \cdots, Y_n$ are independently distributed, i.e., cases are independent.
- Y_i does NOT need to be normally distributed, but it typically assumes a distribution from an exponential family (e.g. binomial, Poisson)
- GLM does NOT assume a linear relationship between Y and X, but it does assume linear relationship between the transformed response in terms of the link function and the explanatory variables.
- X can be even the power terms or some other nonlinear transformations of the original independent variables.
- Overdispersion (when the observed variance is larger than what the model assumes) maybe present.
- Errors need to be independent but NOT normally distributed.
- It uses MLE rather than OLS.

Outline

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

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

$\theta = \mu \quad \phi = \sigma^2 \quad a(\phi) = \phi \quad b(\theta) = \theta^2/2 \quad c(y,\phi) = -(y^2/\phi + \log(2\pi\phi))/2$

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

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 $g(\mu) = \theta$.

Look back at the exponential family distributions

Table 2.1 *Characteristics of some common univariate distributions in the exponential family*†

	Normal	*Poisson*	*Binomial*	*Gamma*	*Inverse Gaussian*
Notation	$N(\mu, \sigma^2)$	$P(\mu)$	$B(m, \pi)/m$	$G(\mu, \nu)$	$IG(\mu, \sigma^2)$
Range of y	$(-\infty, \infty)$	$0(1)\infty$	$\frac{0(1)m}{m}$	$(0, \infty)$	$(0, \infty)$
Dispersion parameter: ϕ	$\phi = \sigma^2$	1	$1/m$	$\phi = \nu^{-1}$	$\phi = \sigma^2$
Cumulant function: $b(\theta)$	$\theta^2/2$	$\exp(\theta)$	$\log(1 + e^\theta)$	$-\log(-\theta)$	$-(-2\theta)^{1/2}$
$c(y; \phi)$	$-\frac{1}{2}\left(\frac{y^2}{\phi} + \log(2\pi\phi)\right)$	$-\log y!$	$\log \binom{m}{my}$	$\nu \log(\nu y) - \log y$ $- \log \Gamma(\nu)$	$-\frac{1}{2}\left\{\log(2\pi\phi y^3) + \frac{1}{\phi y}\right\}$
$\mu(\theta) = E(Y; \theta)$	θ	$\exp(\theta)$	$e^\theta/(1 + e^\theta)$	$-1/\theta$	$(-2\theta)^{-1/2}$
Canonical link: $\theta(\mu)$	identity	log	logit	reciprocal	$1/\mu^2$
Variance function: $V(\mu)$	1	μ	$\mu(1 - \mu)$	μ^2	μ^3

How to fit GLM? MLE.

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

How to fit GLM? MLE.

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

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

How to fit GLM? MLE.

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

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

$$= \frac{1}{a(\phi)}\sum_{i=1}^{n}[y_i\theta_i - b(\theta_i) + c(y_i, \phi)a(\phi)]$$

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

$$= \frac{1}{a(\phi)}\sum_{i=1}^{n}\left[y_i\frac{\partial \theta_i}{\partial \beta_j} - \frac{\partial b(\theta_i)}{\partial \beta_j}\right]$$

Log Likelihood

Now

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

Therefore

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

Maximum likelihood estimation

Maximum likelihood estimates:

$$\frac{\partial \ell(\boldsymbol{\beta};\mathbf{y})}{\partial \beta_j} = \sum_{i=1}^{n} \left[\frac{y_i - b'(\theta_i)}{V(\mu_i)} \right] \frac{\partial \mu_i}{\partial \beta_j} = 0 \qquad \text{for all } j$$

Iterative Reweighted Least Squares algorithm (IRWLS)

1. Start with an initial guess of $\hat{\boldsymbol{\mu}}^{(0)}$ from which $\hat{\boldsymbol{\eta}}^{(0)}$ can be calculated.
2. At each iteration t, compute the *adjusted dependent variable*

$$\mathbf{z}^{(t)} = \hat{\boldsymbol{\eta}}^{(t)} + (\mathbf{y} - \hat{\boldsymbol{\mu}}^{(t)}) \frac{d\boldsymbol{\eta}}{d\boldsymbol{\mu}} |\hat{\boldsymbol{\mu}}^{(t)}$$

   along with weights

$$\mathbf{W}^{(t)} = \text{diag}\left\{ [V(\hat{\mu}_i^{(t)})]^{-1} \left(\frac{d\mu_i}{d\eta_i} |\hat{\mu}_i^{(t)} \right)^2 \right\}.$$

3. Calculate $\beta^{(t+1)}$ using $\beta^{(t+1)} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{z}$, and hence $\boldsymbol{\mu}^{(t+1)}$.
4. Repeat steps 2 and 3 till convergence.

Reconsideration by Fisher-scoring algorithm

Newton-Raphson method

- To find successively better approximations to the roots of $f(x) = 0$ based on Taylor's expansion.

Newton-Raphson method:

$$\beta^{(t+1)} = \beta^{(t)} + [-l^{''}(\beta^{(t)})]^{-1} \cdot l^{'}(\beta^{(t)})$$

Fisher-scoring algoritm

- Replace the second derivative with its expectation.

Fisher-scoring algoritm:

$$\beta^{(t+1)} = \beta^{(t)} + [-E(l^{''}(\beta^{(t)}))]^{-1} \cdot l^{'}(\beta^{(t)})$$

Fisher-scoring algorithm

$$\beta^{(t+1)} = \beta^{(t)} + [-E(l''(\beta^{(t)}))]^{-1} \cdot l'(\beta^{(t)})$$

Fisher-scoring algorithm

$$\beta^{(t+1)} = \beta^{(t)} + [-E(l''(\beta^{(t)}))]^{-1} \cdot l'(\beta^{(t)})$$

$$\frac{\partial \ell(\beta; \mathbf{y})}{\partial \beta_j} = \sum_{i=1}^{n} \left[\frac{y_i - \mu_i}{V(\mu_i)} \right] \frac{\partial \mu_i}{\partial \beta_j}$$
$$= \sum_{i=1}^{n} \left[\frac{y_i - \mu_i}{V(\mu_i)} \right] \frac{d\mu_i}{d\eta_i} x_j$$

Therefore,

$$l'(\beta) = X^T A(\mathbf{y} - \boldsymbol{\mu}),$$

where

$$A = \text{diag}\left\{ [V(\mu_i)]^{-1} \frac{d\mu_i}{d\eta_i} \right\}.$$

Fisher-scoring algorithm

$$\beta^{(t+1)} = \beta^{(t)} + [-E(l''(\beta^{(t)}))]^{-1} \cdot l'(\beta^{(t)})$$

$$-E\left(\frac{\partial^2 \ell(\beta; \mathbf{y})}{\partial \beta_j \partial \beta_k}\right) = \sum_{i=1}^{n} \frac{1}{V(\mu_i)} \left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij} x_{ik}$$

Therefore,

$$-E(l''(\beta)) = X^T W X,$$

where

$$W = \text{diag}\left\{[V(\mu_i)]^{-1} \left(\frac{d\mu_i}{d\eta_i}\right)^2\right\}.$$

Fisher-scoring algorithm

$$\beta^{(t+1)} = \beta^{(t)} + [-E(l''(\beta^{(t)}))]^{-1} \cdot l'(\beta^{(t)})$$

$$\Downarrow$$

$$\beta^{(t+1)} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}$$

$$\boldsymbol{z} = \boldsymbol{\eta} + (\boldsymbol{y} - \boldsymbol{\mu})\frac{d\eta}{d\mu}$$

Fisher-scoring algorithm

1. Start with an initial guess of $\hat{\boldsymbol{\mu}}^{(0)}$ from which $\hat{\boldsymbol{\eta}}^{(0)}$ can be calculated.
2. At each iteration i, compute the *adjusted dependent variable*

$$\boldsymbol{z}^{(t)} = \hat{\boldsymbol{\eta}}^{(t)} + (\boldsymbol{y} - \hat{\boldsymbol{\mu}}^{(t)})\frac{d\boldsymbol{\eta}}{d\boldsymbol{\mu}}|\hat{\boldsymbol{\mu}}^{(t)}$$

   along with weights

$$\boldsymbol{W}^{(t)} = \text{diag}\left\{[V(\hat{\mu}_i^{(t)})]^{-1}\left(\frac{d\mu_i}{d\eta_i}|\hat{\mu}_i^{(t)}\right)^2\right\}.$$

3. Calculate $\beta^{(t+1)}$ using $\beta^{(t+1)} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}$, and hence $\boldsymbol{\mu}^{(t+1)}$.
4. Repeat steps 2 and 3 till convergence.

Maximum likelihood estimation

- 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

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

family options

```r
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

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

link options

```r
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 or any quasi family

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 df = $\text{df}_\omega - \text{dw}_\Omega$ = 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

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 $-2 \log L$.

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

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

GLM Residuals

Response residuals

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

Pearson residuals

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

Deviance residuals

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

GLM Leverage and Influence

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

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

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

- Leverage values are diagonals of $\boldsymbol{H}$.
- `influence(fit)$hat`

GLM Leverage and Influence

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

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

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

- Leverage values are diagonals of $\boldsymbol{H}$.
- `influence(fit)$hat`

Cooks Distance

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

- `cooks.distance(fit)`

Diagnostics

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

Robust GLM

Maximum likelihood estimates:

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

Robust GLM

Maximum likelihood estimates:

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

Robust Maximum likelihood estimates:

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

where γ_i downweights extreme observations.

- γ is estimated iteratively.
- Implemented in R using `robustbase::glmrob` for binomial and Poisson.

Outline

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 ($\lambda = 0.5$, 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 ($\lambda = 0.5$, df = 2ν)

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; $V(Y) = \mu^2/\nu$

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 ($\lambda = 0.5$, 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; $V(Y) = \mu^2/\nu$

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

When to use Gamma GLM?

- The response has Gamma distribution.
- Standard deviation increases linearly with the response.

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; V = μ^3/λ
- μ = 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; V = μ^3/λ
- μ = 1 is special case (Wald distribution)

- Canonical link is $1/\mu^2$
- Variance increases with μ more rapidly than Gamma.
- As $\lambda \to \infty$, distribution converges to Gaussian.

Tweedie GLM

Tweedie distribution

Exponential family distribution where $V(Y) = a\mu^p$, $a > 0$, $p > 0$.

- 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

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

- 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

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

- 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

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

Compound Poisson-gamma distribution

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

Compound Poisson-gamma distribution

$$
\begin{aligned}
V(Y) &= V_N[E_X(Y \mid N)] + E_N[V_X(Y \mid N)] \\
&= V_N[NE(X)] + E_N[NV(X)] \\
&= V(N)[E(X)]^2 + E(N)V(X) \\
&= E(N)\left([E(X)]^2 + V(X)\right) \\
&= \frac{\mu^{2-p}}{(2-p)\phi}\Big(\phi^2(2-p)^2\mu^{2(p-1)} \\
&\qquad\qquad + \phi^2(2-p)(p-1)\mu^{2(p-1)}\Big) \\
&= \frac{\mu^{2-p+2p-2}\phi^2(2-p)}{(2-p)\phi}\Big((2-p) + (p-1)\Big) \\
&= \phi\mu^p
\end{aligned}
$$

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`