Module 8: Generalized linear regression

Jianhui Gao

July 21, 2025

Outline

In this module, we will review generalized linear regression.

Logistic regression

- Each response is binary: $y_i = 1, 0$
- Explanatory variables x_i^T as usual
- Model

$$P\left(y_i = 1 \mid x_i\right) = \frac{\exp\left(x_i^\top \beta\right)}{1 + \exp\left(x_i^\top \beta\right)}$$

Generalized linear models (GLMs)

- Generalized Linear Models extend the classical set-up to allow for a wider range of distributions
- GLMs have three pieces
  1. random component: $y_i \sim$ some distribution with $E\left[y_i|\mathbf{x}_i\right] = \mu_i$
  2. systematic component: $\mathbf{x}_i^T \beta$
  3. The link function that links the random and systematic components $g\left(u_i\right) = \mathbf{x}_i^T \beta$
- Distributions of y_i comes from exponential family.

Exponential family

The random variable Y belongs to the exponential family of distributions if its support does not depend upon any unknown parameters and its density or probability mass function takes the form

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

where θ is the canonical parameter (related to the mean), ϕ is a dispersion parameter (often related to the variance), $b(\theta)$ is the cumulant function (which helps derive the mean and variance), and $c(y, \phi)$ is a normalization term ensuring the density integrates (or sums) to 1.

Example 1 Gaussian distribution

The Gaussian (Normal) distribution can be written in exponential family form as:

$$f(y \mid \mu, \sigma^2) = \exp\left(\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right)$$

where:

$$\begin{aligned}
\theta &= \mu && \text{(natural parameter)} \\
\phi &= \sigma^2 && \text{(dispersion parameter)} \\
b(\theta) &= \frac{\theta^2}{2} && \text{(cumulant function)} \\
c(y, \phi) &= -\frac{y^2}{2\phi} - \frac{1}{2}\log(2\pi\phi) && \text{(normalization term)}
\end{aligned}$$

Example 2 Poisson distribution

The Poisson distribution with parameter λ (rate parameter) can be written in exponential family form as:

$$f(y \mid \lambda) = \exp\left(y \log \lambda - \lambda - \log(y!)\right)$$

where:

$$\begin{aligned}
\theta &= \log \lambda && \text{(natural parameter)} \\
\phi &= 1 && \text{(dispersion parameter)} \\
b(\theta) &= e^{\theta} && \text{(cumulant function)} \\
c(y, \phi) &= -\log(y!) && \text{(normalization term)}
\end{aligned}$$

Example 3 Binomial distribution

The Binomial distribution with parameters n (number of trials) and p (success probability) can be written in exponential family form as:

$$f(y \mid p) = \exp\left(y\log\left(\frac{p}{1-p}\right) + n\log(1-p) + \log\binom{n}{y}\right)$$

where:

$$\begin{aligned}
\theta &= \log\left(\frac{p}{1-p}\right) && \text{(natural parameter, log-odds)} \\
\phi &= 1 && \text{(dispersion parameter)} \\
b(\theta) &= n\log(1+e^{\theta}) && \text{(cumulant function)} \\
c(y,\phi) &= \log\binom{n}{y} && \text{(normalization term)}
\end{aligned}$$

MGF of exponential family

The MGF is given by:

$$
\begin{aligned}
M_Y(t) &= \mathbb{E}[e^{tY}] \\
&= \int e^{ty} \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right) dy \\
&= \exp\left(-\frac{b(\theta)}{\phi}\right) \int \exp\left(y\frac{\theta + \phi t}{a(\phi)} + c(y,\phi)\right) dy \\
&= \exp\left(-\frac{b(\theta)}{\phi}\right) \exp\left(\frac{b(\theta + \phi t)}{a(\phi)}\right) \int f(y \mid \theta + \phi t, \phi) dy \\
&= \exp\left(\frac{b(\theta + \phi t) - b(\theta)}{a(\phi)}\right)
\end{aligned}
$$

where the integral equals 1 because $f(y \mid \theta + \phi t, \phi)$ is a valid density/pmf.

Mean of the exponential family

Using the MGF:

$$M'_Y(t) = \frac{d}{dt}\exp\left(\frac{b(\theta+\phi t)-b(\theta)}{a(\phi)}\right) = b'(\theta+\phi t)\exp\left(\frac{b(\theta+\phi t)-b(\theta)}{a(\phi)}\right)$$

Evaluating at $t = 0$:

$$\mathbb{E}[Y] = M'_Y(0) = b'(\theta)$$

Variance of the exponential family

The second derivative of the MGF:

$$M''_Y(t) = \left[a(\phi)b''(\theta + \phi t) + (b'(\theta + \phi t))^2\right] \exp\left(\frac{b(\theta + \phi t) - b(\theta)}{a(\phi)}\right)$$

Evaluating at $t = 0$:

$$\mathbb{E}[Y^2] = M''_Y(0) = a(\phi)b''(\theta) + (b'(\theta))^2$$

Thus:

$$\text{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = a(\phi)b''(\theta)$$

Link function

The second element of the generalization is that instead of modeling the mean, as before, we will introduce a one-to-one continuous differentiable transformation $g(\mu_i)$ of the mean $\mu_i = E[y_i]$ and model that

$$(\mu_i) = \eta_i = {x_i}^\top \beta$$

where η_i is the linear predictor. The function $g(\cdot)$ is called the link function.

Link function

Since $g(\cdot)$ is a one-to-one transformation, we can invert it to get the mean:

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(x_i^\top \beta)$$

Note that we do not transform the response variable y_i itself, but rather the mean of the response variable.

Link function in R

"glm" has several options for family:

binomial (link = "logit")
gaussian(link = "identity")
Gamma(link = "inverse")
inverse.gaussian(link = $"1/\mathrm{mu}^{\wedge}2"$)
poisson(link = "log")
quasi (link = "identity", variance = "constant")
quasibinomial (link = "logit")
quasipoisson(link = "log")

MLE

An important practical feature of generalized linear models is that they can all be fit to data using the same algorithm, a form of iteratively re-weighted least squares (IRLS).

1. Initialize $\hat{\mu}_i^{(0)} = y_i + \epsilon$, $\eta_i^{(0)} = g(\hat{\mu}_i^{(0)})$
2. While not converged:
    - Working response: $z_i = \eta_i^{(k)} + (y_i - \hat{\mu}_i^{(k)}) \left(\frac{d\eta}{d\mu}\right)$
    - Weights: $w_i = \left[V(\hat{\mu}_i^{(k)}) \left(\frac{d\mu}{d\eta}\right)^2 \right]^{-1}$
    - Update: $\beta^{(k+1)} = (X^\top WX)^{-1} X^\top Wz$
    - Update $\eta_i^{(k+1)}$, $\hat{\mu}_i^{(k+1)}$

Gaussian Special Case of IRLS

For linear regression ($Y \sim N(\mu, \sigma^2)$):

- **Link**: Identity $g(\mu) = \mu$
- **Variance**: $V(\mu) = 1$ (constant)
- **Weights**: $w_i = 1$ (equal weighting)
- **Working response**: $z_i = y_i$ (original data)

IRLS reduces to ordinary least squares:

$$\beta^{(k+1)} = (X^\top X)^{-1} X^\top y \quad \text{(single iteration)}$$

Key observations:

- Link derivative $\frac{d\mu}{d\eta} = 1$
- No reweighting needed (homoscedasticity)
- Exact solution in one step

Asymptotics

The asymptotic distribution of the MLE $\hat{\beta}$ is given by:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, (X^\top W X)^{-1}\phi)$$

In the case of the Gaussian distribution, ϕ is the variance σ^2, W is the identity matrix, and the covariance matrix simplifies to: $(X^\top X)^{-1}\sigma^2$.