Stat 151a Lecture Notes 23

Chae Yeon Lee

April 20, 2023

1 Overview

Today:

- More Generalized Linear Models

Relevant Readings:

- Fox 15.1 - 15.3 (Skip 15.2.2, 15.3.4)

2 Brief Recap

- If $\vec{y}$ domain is not $\mathbb{R}$ (like if it is binary), we might prefer to assume a different parametric distribution on $\vec{y}$ rather than that $\vec{y}$ is linear and normal.
- GLMs are a framework that allow this
- Now we cover their estimates and inferences.

3 Generalized Linear Models

GLM is a framework to allow $\vec{y}$ to follow various parametric distribution.

- **Bernoulli:** $\vec{y}$ is binary
- **Poisson, Negative Binomial:** $\vec{y} \in \mathbb{Z}^+$ (count data)
- **Exponential, Gamma:** $\vec{y} \in \mathbb{R}^+$ (continuous data)
- **Multinomial:** $\vec{y}$ is a count of occurences of K different outcomes.

The GLM "recipe":

1. Specify the conditional distribution of $y_i|\vec{x}_i^T$ (not necessarily normal; e.g. $y_i \sim \text{Ber}(\pi_i)$)

2. Consider the parameters of the distribution in 1, their relation to $\mathbb{E}[y_i|\vec{x}_i^T]$, and their domain.

   e.g. $y_i|\vec{x}_i^T \sim \text{Bernoulli}(\pi_i)$. The mean of Bernoulli RV is π_i. Note that $\pi \in (0,1)$.

3. Find a <u>link</u> function g that transforms $\mathbb{E}[y_i|\vec{x}_i^T]$ into the range $(-\infty,\infty)$.

   e.g. $y_i \sim \text{Ber}(\pi_i)$. Let g be the <u>logit</u> function. $g(\pi_i) = log(\frac{\pi_i}{1-\pi_i})$

   Then $g : (0,1) \to (-\infty,\infty)$ as desired. Here, we are inputting π_i but π_i is what we are actually estimating.

4. $g(\pi_i)$ can be modelled linearly:

$$g(\pi_i) = \vec{x}_i^T \vec{\beta}$$

And, we define the <u>inverse link function</u>, g^{-1}. e.g. $y_i|\vec{x}_i^T \sim \text{Bernoulli}(\pi_i)$

$$g(\pi_i) = log(\frac{\pi_i}{1-\pi_i}) = \vec{x}_i^T \vec{\beta}$$

$$\Rightarrow \frac{\pi_i}{1-\pi_i} = exp(\vec{x}_i^T \beta$$

$$\Rightarrow \pi_i = exp(\vec{x}_i^T \vec{\beta}) - \pi_i exp(\vec{x}_i^T \vec{\beta})$$

$$\Rightarrow \pi_i = \frac{exp(\vec{x}_i^T \vec{\beta})}{1 + exp(\vec{x}_i^T \vec{\beta})}$$

(Explicit or inverse logit function)

$$\Rightarrow \frac{1}{1 + exp(-\vec{x}_i^T \vec{\beta})}$$

$$:= g^{-1}(\vec{x}_i^T \vec{\beta})$$

Notes:

- A whole variety of Link function could make π_i to $(-\infty,\infty)$
- The logit function is the canonical link; The **canonical link functions** are based on the likelihood function and are the most popular. Canononical link functions can be derived from the likelihood functions.
- if $y_i|\vec{x}_i^T \sim \text{Poisson}(\lambda_i)$, $\lambda_i \in (0,\infty)$.
- Logit link is the framework for logistic regression.

Table 15.2 Canonical Link, Response Range, and Conditional Variance Function for Exponential Families

Family	*Canonical Link*	*Range of Y_i*	$V(Y_i \mid \eta_i)$
Gaussian	Identity	$(-\infty, +\infty)$	ϕ
Binomial	Logit	$\frac{0, 1, ..., n_i}{n_i}$	$\frac{\mu_i(1-\mu_i)}{n_i}$
Poisson	Log	0,1,2,...	μ_i
Gamma	Inverse	$(0,\infty)$	$\phi\mu_i^2$
Inverse-Gaussian	Inverse-square	$(0,\infty)$	$\phi\mu_i^3$

NOTE: ϕ is the dispersion parameter, η_i is the linear predictor, and μ_i is the expectation of Y_i (the response). In the binomial family, n_i is the number of trials.

3.0.1 GLM VS OLS

Contrast the GLM framework from running OLS on some transformed outcome (via function g)

- OLS: $\mathbb{E}[g(y_i)] = \vec{x}_i{}^T\vec{\beta}$
- GLM: $g(\mathbb{E}[y_i]) = \vec{x}_i{}^T\vec{\beta}$ since $\mathbb{E}[y_i] = g^{-1}(\vec{x}_i{}^T\vec{\beta})$
- Thus, OLS and GLM are not equal if g is nonlinear.
- The variance of y_i is now often tied to its mean, so it can fluctuate. For OLS, we assumed that the variance is constant. Now, it is not.

  E.g. **Ber(π_i) variance** is $\pi_i(1-\pi_i)$, **Pois(λ_i) variance** is λ_i.
- Some distributions – such as Normal, Negative Binomial, and Gamma – require that we estimate a separate dispersion parameter, ϕ. For normal GLM, $\phi = \sigma_\epsilon^2$. Negative Binomial is equal to Poisson when $\phi = 1$.
- Sometimes, a non-canonical link is a better fit to the data. For example, we might think that the event probability (π_i) is asymmetric in X.

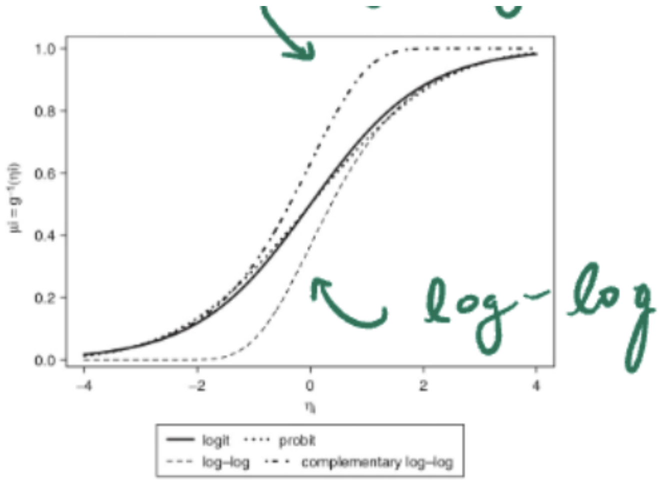

Figure 15.1 Logit, probit, log-log, and complementary log-log links for binomial data. The variances of the normal and logistic distributions have been equated to facilitate the comparison of the logit and probit links [by graphing the cumulative distribution function of $N(0, \pi^2/3)$ for the probit link].

Table 15.1 Some Common Link Functions and Their Inverses

Link	$\eta_i = g(\mu_i)$	$\mu_i = g^{-1}(\eta_i)$
Identity	μ_i	η_i
Log	$\log_e \mu_i$	e^{η_i}
Inverse	μ_i^{-1}	η_i^{-1}
Inverse-square	μ_i^{-2}	$\eta_i^{-1/2}$
Square-root	$\sqrt{\mu_i}$	η_i^2
Logit	$\log_e \dfrac{\mu_i}{1-\mu_i}$	$\dfrac{1}{1+e^{-\eta_i}}$
Probit	$\Phi^{-1}(\mu_i)$	$\Phi(\eta_i)$
Log-log	$-\log_e[-\log_e(\mu_i)]$	$\exp[-\exp(-\eta_i)]$
Complementary log-log	$\log_e[-\log_e(1-\mu_i)]$	$1-\exp[-\exp(\eta_i)]$

NOTE: μ_i is the expected value of the response, η_i is the linear predictor, and $\Phi(\cdot)$ is the cumulative distribution function of the standard-normal distribution.

Figure 15.1 Annotation: x axis is $\vec{x}_i{}^T\hat{B}$ and y axis predicted probability associated with that $\vec{x}_i{}^T\hat{B}$.

3.1 Estimation

The data model for the Bernoulli set up is now $y_i \sim \text{Ber}(g^{-1}(\vec{x}_i{}^T\vec{\beta}))$

- Bad News is that we this no longer corresponds to a least squares problem so we don't have a closed form solution for $\hat{\beta}$
- Good news is that we can still write the Likelihood $L(\hat{\beta}, \vec{y}, X)$. $log(L(\hat{\beta}, \vec{y}, X))$ is convex for many GLM distributions, so it is straightforward to solve via common numerical methods.
- If dispersion (ϕ) parameter must be estimated, we can do alternating optimization.
  1. Init $t = 0, \hat{\phi}_0$ while not converged:
  2. $\hat{\beta}_{t+1} = \hat{\phi}_t$
  3. $\hat{\phi}_{t+1} = \hat{\phi}|\hat{\beta}_{t+1}$ (can be ϕ_{MLE} or ϕ_{MOM})
  4. $t = t+1$

3.2 Prediction

Once we optimize the GLM for $\hat{\beta}$, prediction is straight forward.
$\hat{y}_i = \mathbb{E}[y_i|\vec{x}_i{}^T] = g^{-1}(\vec{x_i^T}\hat{\beta})$
e.g. logistic regression $\hat{y}_i = \frac{1}{1+exp(-\hat{x}_i{}^T\hat{\beta})}$

3.3 Inference

How do we assess model performance?

1. Residual Deviance (Dm):

$$D_m := 2(logL_s - logL_m)$$

   L_m : Likelihood of our model L_s : Likehood of saturated model. (i.e. model with one parameter for each observation.)

2. Example for logistic regression. For logistic regression, $D_m = -2logL_m$. We will prove this: In saturated model, we can fit the training logits perfectly: (i.e. if $y_i = 1$, then $y^{-1}(\hat{x}_i{}^Y\hat{\beta}_{MLE}) \approx 1$)

Fact: If $\hat{\beta}_i$ is a MLE estimate, then $\frac{\hat{B_i - B_i}}{\hat{SE}(\hat{\beta}_i)}$ is asymptotically normal.