#4 - F2017

Generalized Linear Model (GLM)

Summary

- Linear regression with normality assumption
- Estimation (LSE and MLE)
- Inference on parameter estimators
- Diagnostics

Key terms of GLM

- Exponential family
- Three components of GLM

Reading

- DB Chapter 3
- MN Chapter 2

At the moment, without covariate. Every parameter is constant

Exponential Family

Some books use a different expression: $f(y|\theta,\phi) = c(y,\phi)\exp\left\{\frac{y\theta - a(\theta)}{\phi}\right\}$

See de Jong & Heller (2008), P 35, in "GLM For Insurance Data"

- Easily extend good properties of the normal distribution to the exponential family of distribution.
- Exponential family

  Another form. see Fahrmeir & Tutz, P.21, Table 2.1 $f(y|\theta,\phi,\omega) = f(y|\theta,\phi)$, where $a(\phi) = \phi/\omega$

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

$c(y,\phi) = c(y,\phi,\omega)$

for some specific functions $a(\,)$, $b(\,)$ and $c(\,)$, and where

θ: "Canonical (natural) parameter"

ϕ: "Scale (dispersion) parameter"

  - **Note**: it looks different from the one in your D&B's book
  - This formulation is common in GLM and well-accepted in the literature — it is the so called Canonical Form.
  - Through the course, I will use this formulation

Late, θ may depend on covariates x, the a GLM.

In general, $a(\phi) = \phi/\omega$, ω = a known weight, usually group size. e.g., $Var(\bar{y}) = \sigma^2/m$, $a(\phi) = \sigma^2/m$, $\phi = \sigma^2$, $\omega = m$

Examples from Exponential Family

- **Example 1:** Normal distribution $Y \sim N(\mu, \sigma^2)$.

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

$$= \exp\left[-\frac{y^2-2y\mu+\mu^2}{2\sigma^2} + \frac{1}{2}\log\left(\frac{1}{2\pi\sigma^2}\right)\right]$$

$$\exp\left[\frac{y\cdot\mu - \mu^2/2}{\sigma^2} + \left\{-\frac{y^2}{2\sigma^2} + \frac{1}{2}\log\left(\frac{1}{2\pi\sigma^2}\right)\right\}\right].$$

$$\Rightarrow \theta = \mu, \quad b(\theta) = \mu^2/2, \quad \phi = \sigma^2, \quad \omega = 1, \quad a(\phi) = \phi/\omega = \sigma^2$$

$$c(y, \phi) = -y^2/2\sigma^2 + \frac{1}{2}\log\left(\frac{1}{2\pi\sigma^2}\right) = -\frac{1}{2}\left[\log(2\pi\sigma^2) + y^2/\sigma^2\right].$$

Examples from Exponential Family

- **Example 2:** Binomial distribution $Y \sim \text{Bin}(m, p)$.

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

For this distribution, we have two different GLM forms, since $\text{Bin}(m, p) = \sum_{iid}$ Bernoulli r.v.s, m is a group size:

Form 1 Consider y/m as a proportion, use it as a response variable

$$f(y; \theta, \phi) = \exp\left\{y\log p + (m-y)\log(1-p) + \log\binom{m}{y}\right\}$$

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

$$= \exp\left\{\frac{(y/m)\log\frac{p}{1-p} - \log\frac{1}{1-p}}{1/m} + \log\binom{m}{y}\right\}$$

when $m=1$, $y/m = y$, $y \sim$ Bernoulli $(n=1, p)$

$$\Rightarrow \theta = \log\frac{p}{1-p}, \quad b(\theta) = \log\frac{1}{1-p}, \quad \phi = 1, \quad \omega = m, \quad a(\phi) = \phi/\omega = \frac{1}{m}$$

Form 2. Use y as a response, then $f(y; \theta, \phi) = \exp\left\{\frac{y\log\frac{p}{1-p} - m\log\left(\frac{1}{1-p}\right)}{1} + \log\binom{m}{y}\right\}$

$$\Rightarrow b(\theta) = m\log\left(\frac{1}{1-p}\right), \quad a(\phi) = \phi/\omega = 1, \quad c(y, \phi) = \log\binom{m}{y}$$

Examples from Exponential Family

- **Example 3:** Poisson distribution $Y \sim \text{Poisson}(\lambda)$.

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

$$= \exp[\, y \log\lambda - \lambda - \log(y!)\,]$$

$$= \exp\left[\frac{y\log\lambda - \lambda}{1} + \{-\log(y!)\}\right]$$

$$\Rightarrow \theta = \log\lambda, \quad b(\theta) = \lambda, \quad \phi = 1, \quad \omega = 1, \quad a(\phi) = \phi/\omega = 1$$

$$c(y, \phi) = \log(y!)$$

First, we introduce the two Bartlett identities: For a density $f(y;\theta)$ with a single parameter θ, let $l(\theta) = \log f(y;\theta)$.

$\because \int f(y;\theta)\,dy = 1$

$\therefore \int \frac{\partial f}{\partial\theta}\,dy = 0$

$\int \left\{\left(\frac{\partial f}{\partial\theta}\right)/f\right\} f = 0 \quad (1)$

$\Rightarrow \boxed{E\left(\frac{\partial l(\theta)}{\partial\theta}\right) = 0 \; (i)}$

By (1), $\int \frac{\partial l(\theta)}{\partial\theta} f = 0$, then

$\int \left(\frac{\partial^2 f}{\partial\theta^2}\right) f\,dy + \int \left(\frac{\partial l}{\partial\theta}\right)\left(\frac{\partial f}{\partial\theta}\right) dy = 0$

$\Rightarrow E\left(\frac{\partial^2 l(\theta)}{\partial\theta^2}\right) + E\left\{\left(\frac{\partial l(\theta)}{\partial\theta}\right)\right\}^2 = 0$

$\Rightarrow \boxed{E\left(\frac{\partial^2 l(\theta)}{\partial\theta^2}\right) + Var\left(\frac{\partial l(\theta)}{\partial\theta}\right) = 0} \; (ii)$

Properties of Exponential Family

- Properties: if $Y \sim f(y|\theta, \phi)$ and ϕ is fixed (known), then:

  (i) $\mu \equiv E[Y] = b'(\theta)$. $b''(\theta) = V(\mu)$ is called the variance function.

  (ii) $V(Y) = b''(\theta)a(\phi)$. Not the same as $V(Y) = Var(Y)$

  Proof:

  Use $E[U(\theta)] = 0$ and $E[U(\theta)^2] = -E[\partial^2 l(\theta)/\partial\theta^2]$

  For the canonical form, let $l(\theta) = \log f(y;\theta,\phi)$ Then $l(\theta) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y;\phi)$, $u(\theta) = \frac{\partial l(\theta)}{\partial\theta} = \frac{y - b'(\theta)}{a(\phi)}$, $\frac{\partial^2 l(\theta)}{\partial\theta^2} = -\frac{b''(\theta)}{a(\phi)}$. Use the two Bartlett identities, we have $E(u(\theta)) = 0$, and $E\left(\frac{\partial^2 l(\theta)}{\partial\theta^2}\right) + Var(U(\theta)) = 0$ or $E\left(\frac{\partial^2 l(\theta)}{\partial\theta^2}\right) + E(U^2(\theta)) = 0$

- Note: b'' is often referred as the "variance function"

  It implies $0 = E\left(\frac{\partial l(\theta)}{\partial\theta}\right) = E(u(\theta)) = \frac{E(y) - b'(\theta)}{a(\phi)} \Rightarrow \boxed{\mu = E(Y) = b'(\theta)}$

  Since $E(U^2(\theta)) = E\left[\left(\frac{\partial l(\theta)}{\partial\theta}\right)^2\right] = E\left[\left(\frac{y - b'(\theta)}{a(\phi)}\right)^2\right] = \frac{Var(Y)}{(a(\phi))^2}$

  using $E[U^2(\theta)] = E\left[\frac{\partial^2 l(\theta)}{\partial\theta^2}\right]$, and $\frac{\partial^2 l(\theta)}{\partial\theta^2} = -\frac{b''(\theta)}{a(\phi)}$, we get $\frac{Var(Y)}{(a(\phi))^2} = \frac{b''(\theta)}{a(\phi)} \Rightarrow \boxed{V(Y) = Var(Y) = b''(\theta)\,a(\phi)}$

Examples (Revisit)

- **Example 1:** $Y \sim N(\mu, \sigma^2)$ $\because$ $b(\theta) = \mu^2/2$, $\theta = \mu$, $a(\phi) = \sigma^2/\omega = \sigma^2/1 = \sigma^2$

  $E[Y] = b'(\theta) = b'(\mu) = \mu$ $\because$ $b'(\theta) = (\theta^2/2)' = \theta = \mu$, $b''(\theta) = (\theta)' = 1$

  $V(Y) = b''(\theta)\, a(\phi) = a(\phi) = \sigma^2$ [Here, $\underset{\text{variance function}}{V(\mu)} = Var(Y) = V(Y)$]

- **Example 2:** $Y \sim \text{Bin}(m, p)$. $\Longrightarrow$ We use Y/m as a response variable, then $\theta = \log \frac{p}{1-p}$, $b(\theta) = \log \frac{1}{1-p}$, $a(\phi) = \frac{1}{m}$

  $\Rightarrow p = \frac{e^\theta}{1+e^\theta}$

  $b'(\theta) = \frac{\partial}{\partial p}\left(\log \frac{1}{1-p}\right) \cdot \frac{\partial p}{\partial \theta}$

  $= \frac{1}{1-p} \quad \frac{(1+e^\theta)e^\theta - e^{2\theta}}{(1+e^\theta)^2}$

  $= \frac{1}{1-p} \cdot \frac{e^\theta}{(1+e^\theta)^2}$

  $= \frac{1}{1-p}(1-p)\, p = p$

  $b''(\theta) = \frac{\partial p}{\partial \theta} = p(1-p)$

  $\Rightarrow E\left(\frac{Y}{m}\right) = b'(\theta) \Rightarrow \boxed{E(Y) = m b'(\theta) = mp}$

  $Var\left(\frac{Y}{m}\right) = b''(\theta)\, a(\phi) = p(1-p)/m$

  $\Rightarrow \boxed{V(Y) = Var(Y) = mp(1-p)}$

  $E[Y] = m\, b'(\theta) = mp$

  $V(Y) = mp(1-p)$

- **Example 3:** $Y \sim \text{Poisson}(\lambda)$. (HW)

  $E[Y] = b'(\theta) = b(\theta) = \lambda$

  $V(Y) = b''(\theta)\, a(\phi) = b''(\theta) = \lambda$

Notice: $\theta = \log \lambda$, $b(\theta) = \lambda$

$\Rightarrow$ $b(\theta) = \exp(\theta)$, $b'(\theta) = b''(\theta) = \exp(\theta) = b(\theta) = \lambda$, $a(\phi) = 1$.

Revisit: Why GLM?

- e.g., binary response ($Y =$ death or alive).

  normality assumption fell through.

  $E[Y_i] = x_i^T\beta$ does not hold without constraint.

In situations of the mean and/or variance depending on some covariates X_i

- Need a more general regression framework accounting for response data having a variety of measurement scales.

- Methods for model fitting and inference under this framework.

Three Components of GLM

- **Revisit:** Linear regression model with normality assumption.
  - Random variation: $Y_i \sim N(\mu_i, \sigma^2)$, where $\mu_i = E[Y_i|x_i]$.
  - Linear predictor $\eta_i =$ $x_i^T\beta = \beta_0 + \Sigma_{j=1}^{p} x_{ij}\beta_j$ with p covariates.
  - Link between μ_i and $x_i^T\beta$. In linear regression,

$$\mu_i = x_i^T\beta.$$

* Display this late

[In GLM, what is the change in $\mu_i = g^{-1}(\eta_i)$ for change in x_{jk} holding others constants?

Recall $\eta_i = g(\mu_i)$, then $\mu_i = g^{-1}(\eta_i) = h(\eta_i)$, then

$$\frac{\partial \mu_i}{\partial x_{ik}} = \frac{\partial h(\eta_i)}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial x_{ik}} = \left[\frac{\partial h(\eta_i)}{\partial \eta_i}\right]\beta_k = \propto \beta_k$$

If $g =$ identity link, then $\frac{\partial \mu_i}{\partial x_{ik}} = \beta_k$, interpretation follows that in the linear regression model]

STAT 635-GLM-Lecture Notes 4, Introduction to Generalized Linear Models, Fall 2017

Three Components of GLM

- **Random component:** Y independent r.v. from (exponential) family of distribution with

$$E[Y] = \mu.$$

- **Systematic component** or **linear predictor** p covariates produce a linear predictor η

$$\eta = \beta_0 + \Sigma_{j=1}^{p} x_{ij}\beta_j$$

- **Link function:** the link between the random and systematic components:

$$\eta = g(\mu).$$

  – the link function determined by $\theta = \eta$ is called *Canonical Link*.

Components of GLM (Example)

- **Binary outcome:** $Y \sim$ Bernoulli$(p) \to$ Logistic regression.

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

Here, group size $m=1$, $\theta = \log\frac{p}{1-p}$, $b(\theta) = \log\frac{1}{1-p}$, $a(\phi)=1$

1. Random component $Y \sim$ Bernoulli(p), with $E(Y) = p = \mu$.
2. Linear Predictor: $\eta = \beta_0 + \sum_{j=1}^{p} x_j \beta_j$
3. Link function (Canonical link): $g(p) = \log\left(\frac{p}{1-p}\right)$, called logit link

Desire $g(p) = g(\mu)$. Let $\theta = \eta$, that is $\log\frac{p}{1-p} = \eta$. On the other hand, let $\theta = g(\mu)$ $\mu = E(Y) = p$, that is $\log\frac{p}{1-p} = g(\mu)$, then $g(\mu) = \log\frac{\mu}{1-\mu}$ or $g(p) = \log\frac{p}{1-p}$, a logit link

STAT 635-GLM-Lecture Notes 4, Introduction to Generalized Linear Models, Fall 2017

Components of GLM (Example)

- **Grouped binary data:** e.g., grouped by covariate classes, $Y \sim$ Bin(m,p).
  $f(y|\theta,\phi) = \binom{m}{p} p^y (1-p)^{m-y}$

Consider y/m as the response, m is the group size. Then

1. Random Component Y/m, $Y \sim Bin(m, p)$ with $E(\frac{Y}{m}) = p$
2. Linear predictor: $\eta = \beta_0 + \sum_{j=1}^{p} x_j \beta_j$
3. Link function: Let $\theta = \eta$, then $\log\frac{p}{1-p} = \eta$,
   $\Rightarrow g(p) = \log\frac{p}{1-p}$,
   also a logit link

Interpretation of Parameter

- **Revisit:** linear regression with normal assumption

$$E[Y_i|X_i] = \beta_0 + \beta_1 X_{i1} + \qquad + \beta_p X_{ip}.$$

  – β_k. change in mean response per unit increase of X_{ik} with holding other covariates constant.

- In GLM,

$$\eta_i = X_i^T \beta.$$

  – β_k. change in η_i per unit increase of X_{ik} with holding other covariates constant.

In GLM, what is the change in μ_i for change in x_{ik} holding other Covariates x_{ij}, $j \neq k$? Recall $\eta_i = g(\mu_i)$, then $\mu_i = g^{-1}(\eta_i) = h(\eta_i)$

we obtain $\dfrac{\partial \mu_i}{\partial x_{ik}} = \dfrac{\partial h(\eta_i)}{\partial \eta_i} \dfrac{\partial \eta_i}{\partial x_{ik}} = \left[\dfrac{\partial h(\eta_i)}{\partial \eta_i}\right] \beta_k \underset{prop.}{\propto} \beta_k$.

If $g(\cdot)$ is identity link, then

$\dfrac{\partial \mu_i}{\partial x_{ik}} = \beta_k$, interpretation of β_k follows that in the linear regression model

STAT 635-GLM-Lecture Notes 4, Introduction to Generalized Linear Models, Fall 2017

Logistic Regression

- The logistic regression model, for $Y_i \sim \text{Bernoulli}(p_i)$, can be written as:

In the form of $g(p_i) = \eta_i$

$$\log \frac{p_i}{1-p_i} = X_i^T \beta = \beta_0 + \beta_1 X_{i1} + \qquad + \beta_p X_{ip}.$$

or

$$E[Y_i|X_i] = p_i = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}$$

- Some useful definitions:
  – Odds

$$\frac{p_i}{1-p_i}$$

  – log-odds of the response $Y_i = 1$

$$\log \frac{p_i}{1-p_i}$$

– logit function

$$g(p_i) = \log \frac{p_i}{1 - p_i}.$$

- Interpret β_k for $k = 1, \quad , p.$

$-\beta_k$. change in log-odds per unit increase of x_{ik} with holding other covariates constant

GLMs using the logit link are often called logit models

Other links for probability p in Bernoulli or Binomial distribution

(a) $\eta = \mu$ —— identity link

(b) $\eta = \Phi^{-1}(\mu)$ —— probit link

(c) $\eta = \log(-\log(\mu))$ —— log-log link

(d) $\eta = \log(-\log(1-\mu))$ —— Complementary log-log link

In fact, for any CDF, $F(x)$, let $F(\eta) = p$, then $\eta = F^{-1}(p)$

$g(p) = F^{-1}(p)$ is a link function

Logistic Regression: Example

- Low Birth Weight Study:
  - A case-control study of association between mother's weight (in pounds) at the last menstrual cycle (LWT) and the risk of delivering a low birth weight baby.
  - Data:

$$\text{LOW} = \begin{cases} 0, & \text{Birth weight} \geq 2500g \\ 1, & \text{Birth weight} < 2500g \end{cases}$$

LWT = Mother's Weight (lbs) at the last menstrual period.

AGE = Mother's age in years.

  - Model:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 LWT + \beta_2 AGE.$$

Here, $p = \text{Prob}(\text{Low} = 1) = \text{Prob}(\text{Birth Weight} < 2500g)$

$$* \, p = \frac{\exp(\beta_0+\beta_1 LWT+\beta_2 AGE)}{+ \exp(\beta_0+\beta_1 LWT+\beta_2 AGE)}$$

$$* \, \beta_2 = \log\left(\frac{P(LWT \text{ held Constant}, AGE+1)}{1 - P(LWT \text{ held constant}, AGE+1)}\right)$$

$$- \log\left(\frac{P(LWT \text{ held constant}, AGE)}{1 - P(LWT \text{ held constant}, AGE)}\right)$$

$$= \log \frac{\text{odds (At } LWT \text{ held Constant}, AGE+1)}{\text{odds (At } LWT \text{ held constant}, AGE)}$$

$$= \log(\text{odds ratio})_{(AGE+1) \text{ relative to } (AGE)}$$

Poisson Regression

- **Count data:** $Y \sim$ Poisson$(\lambda) \rightarrow$ Poisson regression.
  - GLM components

$$f(y|\theta,\phi) = \frac{\lambda^y e^{-\lambda}}{y!}$$
$$= \exp\{y \log\lambda - \lambda - \log y!\},$$

thus,

$$\theta = \log\lambda, \;\; b(\theta) = \lambda = e^\theta, \;\; a(\phi \;\; = \phi = 1, \;\; \omega = 1$$

Hence, we have

$$\mu = b'(\theta) = e^\theta = \lambda,$$

and the canonical link is

$$\eta = \theta = g(\mu) = \log(\mu),$$

which is called *log link function.*

Poisson Regression (continued)

- The Poisson regression model can be written as:

$$\log \lambda_i = X_i^T \beta = \beta_0 + \beta_1 X_{i1} + \quad + \beta_p X_{ip}.$$

- Interpret β_k for $k = 1, \quad , p$.

  β_k.

- More about Poisson regression later

  rate;

  - exposure; offset

In general,
$y \sim f(y;\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}$,
$\lambda = E(y)$ is the expected number of deaths in a specified time period, it depends on the population size n and other characteristics of the population.
Let
$E(Y) = \lambda = n r(X^T\beta)$, where $r(X^T\beta)$ is the rate per person-year, and n is the total person years

Poisson Regression: Example

See

- British doctor's smoking and coronary death
  - Data:
    - Age = patient age in years (categorized, e.g., 35-44, 45-54, .)
    - Smoking = $\begin{cases} 0, & \text{non-smoking} \\ 1, & \text{smoking} \end{cases}$
    - Y = death counts from coronary heart disease among male doctors 10 years after survey.
  - Two scenarios:
    - the number of doctors at risk during ith the observation period are the same this is very rare (i.e, For each level of covariates $n_i = n$, $i = 1, \dots, m$

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Smoking}_i,$$

Then $\lambda_i = n_i \exp(x_i^T\beta) = n \exp(x_i^T\beta)$,

$$\log(\lambda_i) = \underbrace{\log(n) + \beta_0}_{\text{Combined into } \beta_0} + \beta_1 Age + \beta_2 Smoking$$

where λ_i = expected death counts. ith

- the total number of person-years of ith observation different

$$\log(\lambda_i) = \log(n_i) + \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Smoking}_i,$$

where n_i = number of doctors at risk during ith observation period.

See BritishDoctorsPoissonReg_F2015.pdf for SAS and R code

- British doctor's data:

	age	smoking	death	personyear
1	35-44	yes	32	52407
2	35-44	no	2	18790
3	45-54	yes	104	43248
4	45-54	no	12	10673
5	55-64	yes	206	28612
6	55-64	no	28	5710
7	65-74	yes	186	12663
8	65-74	no	28	3585
9	75-84	yes	102	5317
10	75-84	no	31	1462

See Table9_123.sas for plots and analysis

# of deaths from coronary heart disease

Total number of person-years during each observation period.

* n_i = # of person-years in the ith observation period

* $\log(n_i)$ is called offset in software

Definition of person-year: One person had a number of years in the follow-up

e.g., One person is followed up for 00 years
=
00 persons are followed up for one year.