Chapter 8. Generalized linear model (I) 일반화선형모형

8.1 Data and Statistical Models 자료와 통계적 모형의 분류

(1) Types of data

data : numerical (quantitative) and categorical (qualitative)

- numerical : continuous and discrete
- categorical : nominal and ordinal (명목형/순서형)

(2) Types of categorical data

- response : binary (dichotomous) and polytomous (polychotomous)
- categorical covariate : factor and level

(3) Types of models

- categorical response : logistic and log-linear
- continuous response : multiple linear regression and ANOVA

8.2 Exponential Family 지수분포군

(1) Definition

Let Y be r.v. with pdf

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

then Y belongs to an *exponential family* with *natural* (*canonical*) parameter $\boxed{\theta}$ if ϕ is known. Also, we assume that $a(\ \cdot\ ), b(\ \cdot\ ), c(\ \cdot,\ \cdot\ )$ are known functions.

→ unknown

(i) $N(\mu, \sigma^2)$ case

$$\text{pdf: } f(y;\theta,\,\phi) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} = \exp\left[-\frac{1}{2\sigma^2}(y^2-2\mu y+\mu^2) - \frac{1}{2}\log(2\pi\sigma^2)\right]$$

$\mu, \underbrace{\sigma^2}_{\text{known}}$

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

so that ① $\boxed{\theta = \mu}$, ④ $\boxed{\phi = \sigma^2}$ and

$$③\ \boxed{a(\phi) = \phi},\ ②\ \boxed{b(\theta) = \frac{\theta^2}{2}},\ ⑤\ \boxed{c(y,\,\phi) = -\frac{1}{2}\left\{\frac{y^2}{\sigma^2} + log(2\pi\sigma^2)\right\}}$$

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

(ii) $P(\lambda)$ case

$$\text{pdf: } f(y;\theta,\,\phi) = \lambda^y e^{-\lambda}/y!, \quad y = 0,\ 1,\ 2,\ \cdots$$

$$= \exp(\underbrace{ylog\,\lambda - \lambda}_{1} - log\,y!),\ y = 0, 1, 2, \cdots$$

so that ① $\boxed{\theta = log\lambda}$, ④ $\boxed{\phi = 1}$ and

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

$$③\ \boxed{a(\phi) = 1},\ ②\ \boxed{b(\theta) = \overset{\lambda}{\cancel{e^\theta}}},\ ⑤\ \boxed{c(y,\,\phi) = -log\,y!}$$

(iii) $B(n, \pi)$ case

$0 < \pi < 1$

pdf: $$f(y;\theta,\phi) = \binom{n}{y}\pi^y(1-\pi)^{n-y}, \quad y = 0, 1, \cdots, n$$

$= \exp\left[y \cdot \log\pi + (n-y)\log(1-\pi) + \log\binom{n}{y}\right]$

$= \exp\left[y \cdot \log\pi - y\log(1-\pi) + n\log(1-\pi) + \log\binom{n}{y}\right]$

$= \exp\left[y\log\frac{\pi}{1-\pi} + n\log(1-\pi) + \log\binom{n}{y}\right]$

$$= \exp\left\{\frac{\frac{1}{n}ylog\left(\frac{\pi}{1-\pi}\right) + log(1-\pi)}{1/n} + log\binom{n}{y}\right\}$$

so that $\boxed{\theta = log\left(\frac{\pi}{1-\pi}\right)}$, $\boxed{\phi = 1/n}$ and $\quad n = \frac{1}{\phi}$

$e^\theta = \frac{\pi}{1-\pi}$

$\pi = e^\theta - e^\theta\pi$

$(1+e^\theta)\pi = e^\theta$

$\pi = \frac{1}{1+e^\theta}$

$1-\pi = \frac{\cancel{1}+e^\theta-\cancel{1}}{1+e^\theta} = \frac{e^\theta}{1+e^\theta}$

$= -\log(1-\pi) = -\log\left(\frac{e^\theta}{1+e^\theta}\right) = \log(1+e^\theta)$

$\boxed{a(\phi) = \phi}$, $\boxed{b(\theta) = log(1+e^\theta)}$, $\boxed{c(y,\phi) = log\binom{1/\phi}{y}}$

※ Exponential Family = $\{N(\mu,\sigma^2), P(\lambda), B(n,\pi), \Gamma(\alpha,\beta), IG, \cdots\}$
IG: Inverse Gaussian

$U(0,\theta) \notin$ Exponential Family

(2) Properties

Let

likehood $\quad L(\theta;y) = f(y;\theta)$ ← pdf

log-likelihood $\quad l(\theta;\phi,y) = logf(y;\theta,\phi)$

be the log-likelihood function, then we have the following theorem, called

Bartlett identity.

Theorem 8.1

$$E\left(\frac{\partial l}{\partial\theta}\right) = 0, \;\; E\left(\frac{\partial^2 l}{\partial\theta^2}\right) + E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right] = 0$$

(Proof) $= \int \frac{\partial l(\theta)}{\partial\theta}\cdot f(y;\theta)dy = \int \frac{\partial \log f(y;\theta)}{\partial\theta} f(y;\theta)dy = \int \frac{\frac{\partial}{\partial\theta}f(y;\theta)}{\cancel{f(y;\theta)}}\cdot\cancel{f(y;\theta)}dy$

$$E\left(\frac{\partial l}{\partial\theta}\right) = \int\frac{1}{f(y;\theta)}\left(\frac{\partial}{\partial\theta}\right)f(y;\theta)dy = \int\frac{\partial}{\partial\theta}f(y;\theta)dy = \frac{\partial}{\partial\theta}\int f(y;\theta)dy = 0$$