2nd Bartlett identity

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

$$\frac{\partial^2 l(\theta)}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left(\frac{\frac{\partial}{\partial\theta}f(y;\theta)}{f(y;\theta)}\right) = \frac{\frac{\partial^2}{\partial\theta^2}f(y;\theta)}{f(y;\theta)} - \left\{\frac{\frac{\partial}{\partial\theta}f(y;\theta)}{f(y;\theta)}\right\}^2$$

$$\boxed{E\left(\frac{\partial^2 l(\theta)}{\partial\theta^2}\right)} = \int \frac{\partial^2}{\partial\theta^2} f(y;\theta)dy - E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right] = \frac{\partial^2}{\partial\theta^2}\underbrace{\int f(y;\theta)dy}_{=1} - E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right] = \boxed{-E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right]}$$

which completes the proof.

We call $\boxed{U = \frac{\partial l}{\partial\theta}\ \text{score function}}$ (스코어함수) and

$$-E\left(\frac{\partial^2 l}{\partial\theta^2}\right) = E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right]$$

[적용예] is called *Fisher information number*. Now, we apply the Bartlett identity in Theorem 8.1 to the exponential family. The log-likelihood function is

$$\boxed{l(\theta; y) = \frac{\{y\theta - b(\theta)\}}{a(\phi)} + c(y, \phi)}$$

and we have

$$\boxed{\frac{\partial l}{\partial\theta} = \frac{\{y - b'(\theta)\}}{a(\phi)}, \quad \frac{\partial^2 l}{\partial\theta^2} = -\frac{b''(\theta)}{a(\phi)}}$$

hence, we have $E\left[\frac{\partial l(\theta)}{\partial\theta}\right] = E\left[\frac{y - b'(\theta)}{a(\phi)}\right] = 0 \Rightarrow E(y) = b'(\theta)$

$$\boxed{E(Y) = \mu = b'(\theta).}$$

(e.g.) $Y \sim N(\mu, \sigma^2)$: $b(\theta) = \frac{\theta^2}{2}$, $\theta = \mu \Rightarrow b'(\theta) = \theta = \mu = E(Y)$

$Y \sim P(\lambda)$: $b(\theta) = e^\theta$, $\theta = \log\lambda \Rightarrow b'(\theta) = e^\theta = \lambda = E(Y)$

$Y \sim B(n, \pi)$: $b(\theta) = \log(1 + e^\theta)$, $\theta = \log\left(\frac{\pi}{1-\pi}\right) \Rightarrow b'(\theta) = \frac{e^\theta}{1+e^\theta} = \pi = E(Y) \leftarrow Y \sim B(1, \pi)$

Fisher Information $= E\left[\left(\frac{\partial \ell(\theta)}{\partial\theta}\right)^2\right] = E\left[\left(\frac{y-b'(\theta)}{a(\phi)}\right)^2\right] = \frac{1}{a^2(\phi)}\boxed{E\left[\{y-b'(\theta)\}^2\right]} = \frac{1}{a^2(\phi)}\boxed{Var(y)}$ =

Fisher Information $= -E\left[\boxed{\frac{\partial^2 \ell(\theta)}{\partial\theta^2}}\right] = E\left[\frac{b''(\theta)}{a(\phi)}\right] = \frac{b''(\theta)}{a(\phi)}$

$= -\frac{b''(\theta)}{a(\phi)}$

$\frac{1}{a^2(\phi)} Var(y) = \frac{b''(\theta)}{a(\phi)}$

Also, from $-\frac{b''(\theta)}{a(\phi)} + \frac{Var(Y)}{a^2(\phi)} = 0$, we have

$$\boxed{Var(Y) = b''(\theta)a(\phi).}$$

We call $b''(\theta)$ *variance function*, and ϕ *dispersion parameter*. Further, we can express θ in terms of $\mu = E(Y)$, and we denote the variance function as $V(\mu)$.

8.3 Construction of GLMs 일반화 선형모형의 구성

The *GLM (Generalized Linear Models)*, suggested by Nelder and Wedderburn (1972), consists of 3 parts;

반응변수 1. $Y_1, \cdots, Y_n$ are independent and belongs to an exponential family.

선형예측치 2. $\boxed{\eta = \sum_{j=0}^{p-1} \boxed{X_j}\beta_j}$ is called *linear predictor*, where $X_0 \equiv 1$.

설명변수

연결함수 (link function) 3. There exists a function g, called a *link function*, s.t. $\boxed{g(\mu_i) = \eta_i,}$ where η_i is the linear predictor and $\mu_i = E(Y_i)$. Also, g is assumed to be monotone and differentiable 미분가능한 단조함수

The classical multiple linear model can be regarded as a special case of GLM with

$$\mu_i = \sum_{j=0}^{p-1} x_{ij}\,\beta_j = \eta_i$$

$\overset{1}{\overset{\shortparallel}{\boxed{x_0}}}\beta_0 + x_1\beta_1 + \cdots + x_{p-1}\beta_{p-1}$

$= \beta_0 + x_1\beta_1 + \cdots + x_{p-1}\beta_{p-1}$

$\eta = g(\mu)$

with identity link function.

For the binomial response, 3 popular link functions are as follows;

1. logit : $\eta = log\{\mu/(1-\mu)\}, \;\; 0 < \mu < 1$
2. probit : $\eta = \Phi^{-1}(\mu), \;\; \Phi(\cdot)$
3. complementary log-log : $\eta = log\{-log(1-\mu)\}$

A link function satisfying $\theta = \eta$ is called *natural* (*canonical*) *link*, and
자연연계함수
natural link functions for each distribution is as follows;

$Y \sim N(\mu, \sigma^2)$, $\theta = \mu \Rightarrow \eta = \mu$ (g: identity)

normal $: \eta = \mu$ $\rightarrow E(Y) = \sum_{j=0}^{p-1} \beta_j X_j$

Poisson $: \eta = log\mu$ $Y \sim P(\lambda)$, $\theta = \log\lambda$, $E(Y) = \lambda$ $\Rightarrow \eta = \log\lambda$ (g: log function)

binomial $: \eta = log\{\mu/(1-\mu)\}$ $Y \sim B(n, \pi)$, $\theta = \log(\frac{\pi}{1-\pi})$, $E(Y) = \pi$

$\Rightarrow \eta = \log(\frac{\pi}{1-\pi})$ (g: logit function)

$Y \sim \Gamma(\alpha, \beta) \Rightarrow \eta = \mu^{-1}$ gamma $: \eta = \mu^{-1}$
(g: inverse function)

inverse Gaussian $: \eta = \mu^{-2}$ $\boxed{Y \sim IG(\mu, \sigma^2)} \Rightarrow \eta = \mu^{-2}$
$\frac{1}{Y} \sim N(\mu, \sigma^2)$

Also, properties of exponential families are summarized in the followings;

distribution	Normal	Poisson	Binomial	Gamma	Inverse Gaussian
notation	$N(\mu, \sigma^2)$	$P(\mu)$	$B(m, \pi)$	$G(\mu, \nu)$	$IG(\mu, \sigma^2)$
dispersion	σ^2	1	$1/m$	$s\nu^{-1}$	σ^2
$b(\theta)$	$\theta^2/2$	e^θ	$log(1+e^\theta)$	$-log(-\theta)$	$-(-2\theta)^{1/2}$
$\mu(\theta)$	θ	e^θ	$e^\theta/(1+e^\theta)$	$-1/\theta$	$(-2\theta)^-1/2$
natural link	identity	log	logit	inverse	$1/\mu^2$
$V(\mu)$	1	μ	$\mu(1-\mu)$	μ^2	μ^3