

Suppose the linear model on p. 166N holds, where both  $\beta$  and  $\tau$  are unknown.

Suppose also that

$$\beta|\tau \sim N(\mu_0, \tau^{-1}\Sigma_0)$$

and

$$\tau \sim \text{gamma}\left(\frac{\delta_0}{2}, \frac{\gamma_0}{2}\right).$$

Then

$$\beta|\mathbf{y} \sim t_{n+\delta_0}\left(\tilde{\beta}, \tilde{s}^2\left(\mathbf{X}^T\mathbf{X} + \Sigma_0^{-1}\right)^{-1}\right),$$

where

$$\tilde{\beta} = \Lambda\mu_0 + (\mathbf{I} - \Lambda)\hat{\beta},$$

$$\Lambda = \left(\mathbf{X}^T\mathbf{X} + \Sigma_0^{-1}\right)^{-1}\Sigma_0^{-1},$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

and

$$\tilde{s}^2 = \frac{1}{(n + \delta_0)} \left[ \mathbf{y}^T (\mathbf{I} - \mathbf{M}) \mathbf{y} + (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}_0)^T (\boldsymbol{\Lambda}^T \mathbf{X}^T \mathbf{X}) (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}_0) + \gamma_0 \right].$$

Also,

$$\tau | \mathbf{y} \sim \text{gamma} \left( \frac{n + \delta_0}{2}, \frac{(n + \delta_0) \tilde{s}^2}{2} \right).$$

### *Generalized Linear Models*

Generalized linear models are natural extensions of classical linear models. Special cases of GLM:

- Linear regression
- ANOVA models
- Logit and probit models

- Log-linear models
- Models for survival data

Aspects of linear model:

1. Random component:  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$ , where  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ . Note that each component of  $\mathbf{Y}$  has same variance.
2. Systematic component: The **covariate** is the systematic component of the model. For the  $i$ th observation let

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}, \quad i = 1, \dots, n.$$

We call  $\eta_i$  the **linear predictor**. So, the relationship between the **mean** of  $Y_i$  and  $\eta_i$  is

$$E(Y_i | \mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{x}_i^T \boldsymbol{\beta} = \eta_i, \quad i = 1, \dots, n.$$

Generalized linear models involve **two extensions** of the Gaussian linear model.

1. The distribution of  $\mathbf{Y}$  is in the **exponential family**.
2. The relationship between  $\mu_i = E(Y_i|\mathbf{x}_i, \boldsymbol{\beta})$  and  $\eta_i$  is more general:

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}.$$

The function  $g$  is called the  **$\mu$ -link function** (or simply the **link function**) and relates the mean of  $Y_i$  to the linear predictor  $\eta_i$ .

The random variable  $Y$  has a distribution in the *exponential family* with *canonical parameter*  $\theta$  and *dispersion parameter*  $\phi$  if  $Y$  has pdf (or pmf)

$$f(y|\theta, \phi) = \exp \{ [y\theta - b(\theta)] / a(\phi) + c(y, \phi) \}.$$

Without loss of generality, we assume that  $a(\phi) \equiv \phi$ . Note that

$$\exp \left[ \frac{b(\theta)}{\phi} \right] = \int_{-\infty}^{\infty} \exp [y\theta/\phi + c(y, \phi)] dy.$$

When using the exponential family as a model for data, we will assume that  $b$  and  $c$  are **known** functions.

The quantity  $\theta$  is called the *canonical parameter*.

Excellent reference for GLM:

[McCullagh and Nelder, 1989, Chapman and Hall](#)

Distributions included in the exponential family:

- normal
- gamma
- binomial
- Poisson
- beta
- multinomial
- inverse Gaussian

Consider the normal distribution

$$\begin{aligned} f(y|\mu, \sigma^2) &= (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] \\ &= \exp\left[\frac{(y\mu - \mu^2/2)}{\sigma^2} - \frac{1}{2}\left(\frac{y^2}{\sigma^2} + \log(2\pi\sigma^2)\right)\right]. \end{aligned}$$

Therefore,

$$\theta = \mu, \quad \phi = \sigma^2, \quad b(\theta) = \frac{\theta^2}{2}$$

and

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

Suppose now that

$$Y \sim \text{binomial}(n, p).$$

Then

$$\begin{aligned} f(y|n, p) &= \binom{n}{y} p^y (1-p)^{n-y} \\ &= \exp \left[ y \log p + (n-y) \log(1-p) \right. \\ &\quad \left. + \log \binom{n}{y} \right] \\ &= \exp \left[ y \log \left( \frac{p}{1-p} \right) \right. \\ &\quad \left. - n \log \left( \frac{1}{1-p} \right) + \log \binom{n}{y} \right] \\ &= \exp \left\{ n \left[ \frac{y}{n} \log \left( \frac{p}{1-p} \right) \right. \right. \\ &\quad \left. \left. - \log \left( \frac{1}{1-p} \right) \right] + \log \binom{n}{y} \right\}. \end{aligned}$$

So, in the binomial case

$$\theta = \log \left( \frac{p}{1-p} \right), \quad \phi = \frac{1}{n},$$

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

and

$$c(y, \phi) = \log \binom{n}{y}.$$

The transformation

$$\log \left( \frac{p}{1-p} \right)$$

is called the **logit transformation**.

Let  $Y$  have distribution in the exponential family. Then

$$\begin{aligned}\ell(\theta, \phi) &= \log f(y|\theta, \phi) \\ &= \frac{y\theta - b(\theta)}{\phi} + c(y, \phi).\end{aligned}$$

The following two identities hold for the exponential family:

$$E \left[ \frac{\partial \ell(\theta, \phi)}{\partial \theta} \right] = 0$$

and

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

This leads to expressions for the mean and variance of  $Y$ .

It's easy to check that

$$\frac{\partial \ell(\theta, \phi)}{\partial \theta} = \frac{y - b'(\theta)}{\phi}$$

and

$$\frac{\partial^2 \ell(\theta, \phi)}{\partial \theta^2} = \frac{-b''(\theta)}{\phi}.$$

Using the two identities on the previous page, then, we have

$$E[Y|\theta, \phi] = b'(\theta)$$

and

$$\text{Var}(Y|\theta, \phi) = \phi b''(\theta).$$

In a generalized linear model, we assume that  $Y_i$  has distribution in the exponential family with parameter  $\theta = \theta_i$  that depends on the linear predictor  $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ .

It is assumed that

$$\theta_i = \theta(\eta_i), \quad i = 1, \dots, n, \quad (1)$$

where  $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$  and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ .

The function  $\theta$  in (1) is called the  $\theta$ -link and is often more convenient to use than the  $\mu$ -link. The  $\theta$ -link is a 1-1 function of the  $\mu$ -link.

Once  $\theta(\eta_i)$  is given, one can write the likelihood function as a function of  $(\boldsymbol{\beta}, \phi)$ .

When  $\theta_i = \eta_i$ , we say that we have a **canonical link**. The function  $\theta$  in (1) can be any monotonic function.

### Example

Suppose  $Y_i \sim \text{bin}(1, p_i)$ ,  $i = 1, \dots, n$ , where  $Y_1, \dots, Y_n$  are independent.

$$\begin{aligned} f(y|p_i) &= \exp \left[ y \log \left( \frac{p_i}{1-p_i} \right) - \log \left( \frac{1}{1-p_i} \right) \right] \\ &= \exp \left[ y\theta_i - \log \left( 1 + e^{\theta_i} \right) \right]. \end{aligned}$$

If we use the canonical link, then  $\theta_i = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ . The likelihood function is then

$$\begin{aligned} f(\mathbf{y}|\boldsymbol{\beta}) &= \prod_{i=1}^n f(y_i|\boldsymbol{\beta}) \\ &= \prod_{i=1}^n \exp \left[ y_i \mathbf{x}_i^T \boldsymbol{\beta} - \log \left( 1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}} \right) \right] \\ &= \exp \left[ \sum_{i=1}^n \left( y_i \mathbf{x}_i^T \boldsymbol{\beta} - \log \left( 1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}} \right) \right) \right]. \end{aligned}$$

For this model, the relationship between  $\theta_i$  and  $\mu_i$  is

$$\theta_i = \log \left( \frac{\mu_i}{1 - \mu_i} \right),$$

where

$$\mu_i = E[Y_i|p_i] = p_i.$$

Therefore

$$\mu_i = \frac{e^{\theta_i}}{1 + e^{\theta_i}}.$$

Consider now a probit model, which has  $\mu$ -link

$$\Phi^{-1}(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}.$$

So,  $g \equiv \Phi^{-1}$  is the  $\mu$ -link.

We can find the  $\theta$ -link as follows.

$$\Phi^{-1}(\mu_i) = \eta_i \quad \Leftrightarrow$$

$$\mu_i = \Phi(\eta_i) \quad \Leftrightarrow$$

$$\frac{e^{\theta_i}}{1 + e^{\theta_i}} = \Phi(\eta_i) \quad \Leftrightarrow$$

$$\theta_i = \log \left( \frac{\Phi(\eta_i)}{1 - \Phi(\eta_i)} \right).$$

So, the  $\theta$ -link is

$$\theta(\eta) = \log \left( \frac{\Phi(\eta)}{1 - \Phi(\eta)} \right).$$

The likelihood function for the probit model is

$$\exp \left\{ \sum_{i=1}^n \left[ y_i \log \left( \frac{\Phi(\mathbf{x}_i^T \boldsymbol{\beta})}{1 - \Phi(\mathbf{x}_i^T \boldsymbol{\beta})} \right) - \log \left( \frac{1}{1 - \Phi(\mathbf{x}_i^T \boldsymbol{\beta})} \right) \right] \right\}.$$