第二章 广义线性模型的指数族分布

Generalized linear models

- Suppose we observe data $\{y_i\}$, which is a realization of a set of independent RVs $\{Y_i\}$.
- A **generalized linear model** (GLM) has the following components.
  - A **random** component,

    $$Y_i \sim f(y_i|\theta_i, \phi),$$

    where $f(\cdot)$ is a pdf/pmf for an exponential family.
  - A **link function** $g(\cdot)$ which satisfies

    $$\eta_i = g(\mu_i),$$

    where $\mu_i = E(Y_i)$.
  - A **systematic** or linear predictor component,

    $$x^T\beta$$

The exponential family

- A family of pdfs or pmfs is called an **exponential** family if

$$f(y|\boldsymbol{\theta}) = a(\boldsymbol{\theta})b(y)\exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(y)\right), y \in A,$$

  where we assume
  - A does not depend on some k-dim vector, $\boldsymbol{\theta}$.
  - $a(\boldsymbol{\theta}) > 0$ is a real valued function which does not depend on y.
  - $w_i(\boldsymbol{\theta})(i = 1, ..., k)$ are real valued functions which do not depend on y.
  - $b(y) \geq 0$ is a real valued function which does not depend on $\boldsymbol{\theta}$.
  - $t_i(y)(i = 1, ..., k)$ are real valued functions which do not depend on $\boldsymbol{\theta}$.

The canonical form

- Instead of $\boldsymbol{\theta}$ being the parameters we can let $\eta_i = w_i(\boldsymbol{\theta})(i = 1, ..., n)$ be the **canonical** or **natural** parameters and the pdf or pmf becomes

$$f(y|\boldsymbol{\eta}) = a^*(\boldsymbol{\eta})b(y)\exp\left(\sum_{i=1}^{k}\eta_i t_i(y)\right), y \in A.$$

- This equational form is **not** unique.

Examples of exponential families

- All these distributions are exponential families:
  - Normal
  - Gamma
  - Beta
  - Inverse Gaussian
  - Binomial
  - Poisson
  - Negative Binomial

Examples

Normal distribution:

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

Binomial distribution:

$$f(y|p) = \binom{m}{y} p^y (1-p)^{m-y}$$

A counter-example

- $f(y|\theta) = \theta^{-1} exp(1 - (y/\theta))$, $0 < \theta < y < \infty$ is not an exponential family.

Sufficient statistics for $\boldsymbol{\eta}$

- Suppose we have a set of independent RVs $\{Y_i\}$, with $Y_i \sim f(y_i|\boldsymbol{\eta})$, (in the canonical form).
- Then

$$
\begin{aligned}
&\prod_{i=1}^{n} f(y_i|\boldsymbol{\eta}) \\
&= \prod_{i=1}^{n} \left[a(y_i) b^*(\boldsymbol{\eta}) \exp\left(\sum_{j=1}^{k} \eta_j t_j(y_i) \right) \right] \\
&= \left(\prod_{i=1}^{n} a(y_i) \right) (b^*(\boldsymbol{\eta}))^n \exp\left(\sum_{j=1}^{k} \eta_j \sum_{i=1}^{n} t_j(y_i) \right).
\end{aligned}
$$

- Hence $\{t_j(\boldsymbol{y}) = \sum_{i=1}^{n} t_j(y_i) : j = 1, ..., n\}$ is a set of sufficient statistics for η, where $\boldsymbol{y} = (y_1, ..., y_n)$.

Likelihood for exponential families

- From the previous slide we see that the likelihood function is

$$
\begin{aligned}
L(\boldsymbol{\eta}|\mathbf{y}) &= \prod_{i=1}^{n} f(y_i|\boldsymbol{\eta}) \\
&= \left(\prod_{i=1}^{n} a(y_i)\right) (b^*(\boldsymbol{\eta}))^n \exp\left(\sum_{j=1}^{k} \eta_j t_j(\mathbf{y})\right).
\end{aligned}
$$

  and thus the log-likelihood is

$$
l(\boldsymbol{\eta}|\mathbf{y}) = \sum_{i=1}^{n} \log a(y_i) + n \log b^*(\boldsymbol{\eta}) + \sum_{j=1}^{k} \eta_j t_j(\mathbf{y}).
$$

The GLM setup

- We often use a specific notation for the exponential family due to Aitkin, et al (1989).
- Suppose $\{Y_i : i = 1, ..., n\}$ are a set of independent RVs with pdf/pmf

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

  - $\boldsymbol{\theta}$ is the **canonical** parameter.
  - ϕ is the **scale** parameter.
- Also assume there is a fixed function $g(\cdot)$ such that $\eta_i = g(\mu_i) = \mathbf{x}_i^T\boldsymbol{\beta}$ where $\mu_i = E(Y_i|X_i)$.
- We will show that θ_i and μ_i are related.

Is f an exponential family?

- If ϕ is fixed (known) then $f(\cdot)$ is an exponential family.
- It may not be an exponential family if ϕ is unknown.

Linear models as a GLM

- Let

$$a(\phi) = \phi,$$

$$b(\theta_i) = \theta_i^2/2,$$

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

- Then

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

  which is the normal density.

- We obtain the linear model if we let

$$\eta_i = \theta_i = \mu_i = E(Y_i) = \boldsymbol{x}_i^T\boldsymbol{\beta}\text{ , and } \phi = \sigma^2.$$

Mean and variance of Y_i

- Suppose that $Y_i \sim f(y_i|\theta_i, \phi)$. Then we will show that

$$E(Y_i) = b^{'}(\theta_i);$$

$$var(Y_i) = b^{''}(\theta_i)a(\phi),$$

where we define

$$b^{'}(\theta_i) = \frac{\partial}{\partial\theta_i} b(\theta_i);$$

$$b^{''}(\theta_i) = \frac{\partial^2}{\partial\theta_i^2} b(\theta_i).$$

- First we check for $Y_i \sim N(\mu_i, \sigma^2)$:

The likelihood function for Y_i

- The likelihood function for Y_i is

$$
\begin{aligned}
L(\theta_i|y_i) &= f(y_i|\theta_i,\phi) \\
&= \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right).
\end{aligned}
$$

- The log likelihood function is thus

$$
l(\theta_i|y_i) = \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi).
$$

- Taking the derivative with respect to μ_i we have

$$
\frac{\partial}{\partial\theta_i} l(\theta_i|y_i) = \frac{1}{a(\phi)}[y_i - b^{'}(\theta_i)].
$$

The likelihood function for Y_i

- Using the identity $E\left[\frac{\partial}{\partial\theta_i} l(\theta_i|y_i)\right] = 0$ if follows that

$$0 = \frac{1}{a(\phi)}[E(Y_i) - b^{'}(\theta_i)].$$

and for $a(\phi) > 0$

$$E(Y_i) = b^{'}(\theta_i)$$

The second derivative of the log-likelihood

- Now
$$\frac{\partial^2}{\partial \theta_i^2} l(\theta_i|y_i) = -\frac{b^{''}(\theta_i)}{a(\phi)}$$
- Using the identity,
$$-E\left[\frac{\partial^2}{\partial \theta_i^2} l(\theta_i|y_i)\right] = E\left[\frac{\partial}{\partial \theta_i} l(\theta_i|y_i)\right]^2$$

  we obtain
$$\frac{b^{''}(\theta_i)}{a(\phi)} = \frac{E[Y_i - b^{'}(\theta_i)]^2}{a(\phi)^2}$$
$$i.e., \quad b^{''}(\theta_i) = \frac{E[Y_i - E(Y_i)]^2}{a(\phi)}$$
$$i.e., \quad var(Y_i) = b^{''}(\theta_i) a(\phi)$$
- Thus variance of Y_i depends on the canonical parameters θ_i ('the mean'), and the scaling parameter, ϕ

Example: Binomial

- Suppose $Y_i (i = 1, ..., n)$ are independent binomial , $B(n, p_i) RV_s$.
- Then

$$f(y_i|p_i) = \binom{n}{y_i} p_i^{y_i} (1 - p_i)^{n - y_i}$$
$$= \left(\frac{p_i}{1 - p_i}\right)^{y_i} \binom{n}{y_i} (1 - p_i)^n$$
$$= \exp\left(y_i \log\left(\frac{p_i}{1 - p_i}\right) + n_i \log(1 - p_i) + \log\binom{n}{y_i}\right)$$

- Let

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

so that

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

Example: Binomial

- Since

$$\log(1-p_i) = \log\left(\frac{1}{1+e^{\theta_i}}\right),$$

we have

$$f(y_i|p_i,\phi) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right).$$

where we let

$$a(\phi) = \phi,$$
$$\phi = 1,$$
$$b(\theta_i) = n\log(1+e^{\theta_i}),$$
$$c(y_i,\phi) = \log\binom{n}{y_i}.$$

Example: Checking the moments

- The two derivatives of $b(\theta_i)$ are

$$b^{'}(\theta_i) = \frac{ne^{\theta_i}}{1+e^{\theta_i}} = np_i$$

$$\begin{aligned} b^{''}(\theta_i) &= \frac{n\left[e^{\theta_i}(1+e^{\theta_i}) - e^{2\theta_i})\right]}{(1+e^{\theta_i})^2} \\ &= \frac{ne^{\theta_i}}{(1+e^{\theta_i})^2} \\ &= np_i(1-p_i) \end{aligned}$$

- Check:

$$E(Y_i) = b^{'}(\theta_i) = np_i$$

$$var(Y_i) = b^{''}(\theta_i)a(\phi) = np_i(1-p_i)$$

- In this case the variance function, $V(\mu_i)$, is

$$V(\mu_i) = \frac{\mu_i(n-\mu_i)}{n}$$

Example: Poisson

- Suppose $Y_i (i = 1, ..., n)$ are independent Poisson, $Po(\lambda_i) RV_s$.
- Then

$$f(y_i|\lambda_i, \phi) = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!}$$
$$= \exp(y_i \log(\lambda_i) - \lambda_i - \log(y_i!)),$$

- Let $\theta_i = \log(\lambda_i)$. Then $\lambda_i = e^{\theta_i}$ and

$$f(y_i|\lambda_i, \phi) = \exp(y_i\theta_i - e^{\theta_i} - \log(y_i!)),$$
$$= \exp(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)),$$

where we let

$$a(\phi) = \phi,$$
$$\phi = 1,$$
$$b(\theta_i) = e^{\theta_i},$$
$$c(y_i, \phi) = \log(y_i!).$$

Example: Poisson

- The two derivative of $b(\theta_i)$ are

$$b^{'}(\theta_i) = e^{\theta_i},$$
$$b^{''}(\theta_i) = e^{\theta_i},$$

- The variance function, $V(\mu_i) = \mu_i$.

Other distributions

- Gamma, Inverse Gaussian etc.

The link function

- The **link function** is a function $g(\cdot)$ such that

$$\eta_i = g(\mu_i),$$

where $\mu_i = E(Y_i)$.

- We have already showed that $\mu_i = b^{'}(\theta_i)$.
- Thus, $g(\cdot)$ **links** the mean, μ_i or the canonical parameter, θ_i, to η_i.
- We normally assume that $g(\cdot)$ is a bijective, continuous and differentiable function.

-Why?

The canonical link

- Definition: The **canonical link** is the function $g(\cdot)$ such that $\eta_i = \theta_i$.
- This implies that

$$\eta_i = g(b^{'}(\theta_i)) = \theta_i.$$

Thus $g(b^{'}(\cdot))$ must be the identity function.

- Example:

1.**Normal**:$\eta_i = \theta_i = \mu_i$.

This is the **identity link** function.

2.**Binominal**,$B(m_i, p_i)$ divided by m_i:

$\eta_i = \theta_i = \ln(\frac{p_i}{1-p_i}) = \ln(\frac{\mu_i}{1-\mu_i})$.

This is the **logit link** function (log odds ratio).

3.**Possion**:$\eta_i = \theta_i = \ln(\lambda_i) = \ln(\mu_i)$.

This is the **log link** function.