

# Lecture Note - 05: Exponential Family, Generalize Linear Model (GLM)

Dihui Lai

March 5, 2020

## Contents

<b>1</b>	<b>Exponential Family</b>	<b>2</b>
1.1	Probability Density Function . . . . .	2
1.2	Special Case: Gaussian Distribution . . . . .	2
1.3	Special Case: Bernoulli Distribution . . . . .	3
<b>2</b>	<b>Moments of Exponential Family</b>	<b>3</b>
2.1	Mean . . . . .	3
2.2	Variance . . . . .	4
<b>3</b>	<b>Generalized Linear Model (GLM)</b>	<b>5</b>
3.1	Link Function . . . . .	5
3.2	Canonical Link Function of Bernouli Distributioin . . . . .	5
3.3	Canonical Link Function of Gaussian Distribution . . . . .	6

# 1 Exponential Family

## 1.1 Probability Density Function

The probability density function (PDF) of an exponential family can be written in a unified format

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

There are two parameters in the function:  $\theta$  and  $\phi$ . Probability distributions like normal, Poisson or binomial can all be written in this format when we choose the right functions  $a(\cdot)$ ,  $b(\cdot)$ , and  $c(y, \cdot)$  and parameter  $\theta$ ,  $\phi$ .

## 1.2 Special Case: Gaussian Distribution

For example, if we set  $\theta$  as the mean and  $\phi$  as the standard deviation:

$$\theta = \mu$$

$$\phi = \sigma$$

Make the three functions  $a$ ,  $b$ ,  $c$ , to be the followings:

$$a(\phi) = \phi^2 = \sigma^2$$

$$b(\theta) = \frac{\theta^2}{2} = \frac{\mu^2}{2}$$

$$c(y, \phi) = -\frac{y^2}{2\phi^2} - \log(\sqrt{2\pi}\phi) = -\frac{y^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma)$$

Substitute the functions into  $f(y)$ , we have

$$\begin{aligned} f(y) &= \exp \left( \frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{y^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma) \right) \\ &= \exp \left( \frac{2y\mu - \frac{\mu^2}{2} - y^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma) \right) \\ &= \exp \left( \frac{-(y - \mu)^2}{2\sigma^2} \right) \exp(-\log(\sqrt{2\pi}\sigma)) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp \left( \frac{-(y - \mu)^2}{2\sigma^2} \right) \end{aligned}$$

This is exactly the Gaussian distribution with mean  $\mu$  and standard deviation  $\sigma$

## 1.3 Special Case: Bernoulli Distribution

Alternatively, if we set the parameters and the functions to be the followings

$$\begin{aligned}\theta &= \log(p) - \log(1 - p) \\ a(\phi) &= 1 \\ b(\theta) &= \log(1 + e^\theta) = -\log(1 - p) \\ c(y, \phi) &= 0\end{aligned}$$

We have the PDF function as

$$\begin{aligned}f(y) &= \exp\left(\frac{y(\log(p) - \log(1 - p)) - (-\log(1 - p))}{1} + 0\right) \\ &= \exp(\log(p^y) + \log(1 - p)^{(-y)} + \log(1 - p)^1) \\ &= \exp(\log(p^y) + \log(1 - p)^{(1-y)}) \\ &= p^y(1 - p)^{(1-y)} \\ &= \begin{cases} p, & y = 1 \\ 1 - p, & y = 0 \end{cases}\end{aligned}$$

, which is exactly the probability mass of a Bernoulli distribution.

## 2 Moments of Exponential Family

### 2.1 Mean

If a random variable  $Y$ 's PDF belongs to the exponential family. We have the mean and variance of  $Y$  immediately as

$$E(Y) = \frac{d}{d\theta}b(\theta) \tag{1}$$

$$Var(Y) = a(\phi) \frac{d^2}{d\theta^2}b(\theta) \tag{2}$$

To derive the equations above, we can simply use the condition that  $f(y)$  has to satisfy

$$\int_{-\infty}^{\infty} f(y) dy = 1$$

i.e.

$$\int_{-\infty}^{\infty} \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right) dy = 1 \tag{3}$$

Take the derivative against  $\theta$  on both side of the equation, we have

$$\frac{d}{d\theta} \int_{-\infty}^{\infty} \exp \left( \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right) dy = 0$$

Without proof, we assume the derivative and integration are interchangeable. We then get

$$\int_{-\infty}^{\infty} \frac{d}{d\theta} \exp \left( \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right) dy = 0$$

The the derivative against the exponential function, we have

$$\int_{-\infty}^{\infty} (y - b'(\theta)) \exp \left( \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right) dy = 0$$

Rearrange the equation, we have

$$\int_{-\infty}^{\infty} y \exp \left( \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right) dy = b'(\theta) \int_{-\infty}^{\infty} \exp \left( \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right) dy$$

This is pretty nice. Because the left hand side is simply the definition of the expectation of  $y$ . On the right side, we can simplify the term using equation (3). We therefore end up having

$$E(Y) = \int_{-\infty}^{\infty} y \exp \left( \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right) dy = b'(\theta) 1 = b'(\theta)$$

## 2.2 Variance

To get the variance of  $Y$ , we can use the same trick of taking derivative but against equation (1) or its equivalence

$$\int_{-\infty}^{\infty} y \exp \left( \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right) dy = b'(\theta)$$

i.e.

$$\frac{d}{d\theta} \int_{-\infty}^{\infty} y \exp \left( \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right) dy = b''(\theta)$$

Again, assuming the derivative and integration is interchangeable, we have

$$\int_{-\infty}^{\infty} \frac{y(y - b'(\theta))}{a(\phi)} \exp \left( \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right) dy = b''(\theta)$$

We therefore have

$$E(Y^2) - (b'(\theta))^2 = a(\phi)b''(\theta)$$

or equivalently

$$Var(Y) = E(Y^2) - E(Y)^2 = a(\phi)b''(\theta)$$

### 3 Generalized Linear Model (GLM)

In logistic regression, the model assume the observed outcomes or values of the target variable follow binomial distribution with parameter  $p$ . The parameter  $p$  or the expectation of the distribution is modeled as a logistic function of a series of predictors  $\vec{x}$  i.e.  $p = \frac{1}{1+\exp(-\vec{\beta}\cdot\vec{x})}$ .

#### 3.1 Link Function

In general, our target variable could follow different types of distribution e.g. Poisson, Gaussian etc. The generalized linear model tries can be used when the target variable  $y$  follows a distribution that belongs to the exponential family.

The expectation of  $Y$  is modeled using a link function  $g^{-1}(\cdot)$

$$E(Y) = b'(\theta) = g(\vec{x} \cdot \vec{\beta})$$

For simplicity, we denote  $\eta = \vec{x} \cdot \vec{\beta}$  and thus

$$E(Y) = b'(\theta) = g(\eta) \tag{4}$$

If we want to estimate the parameter  $\theta$ , we can solve the equation (4) by taking the inverse of the function  $b'(\cdot)$ . If we choose the inverse link function  $g(\cdot)$  to be the same as  $b'(\cdot)$

$$g(\cdot) = b'(\cdot) \tag{5}$$

we get a model with nice property where  $\eta = \theta$ . A link function chosen this way is called a Canonical link function.

#### 3.2 Canonical Link Function of Bernouli Distributioin

Use Bernouli distribution as an example, we have  $b'(\theta) = \frac{e^\theta}{1+e^\theta}$ . The inverse canonical link function is

$$g(\eta) = \frac{e^\eta}{1 + e^\eta} = \frac{1}{1 + e^{-\eta}}$$

. As we known,

$$p = b'(\theta) = \frac{1}{1 + e^{-\eta}}$$

. The equation looks a lot like the probability model in the logistic regression model. In fact, as we will learn in the following section, GLM in the special caes of Bernouli distribution is equivalent to logistic regression.

### 3.3 Canonical Link Function of Gaussian Distribution

In the case of a Gaussian distribution,  $b'(\theta) = \theta$ , which is an identity function. Choose the link function as the identify function, we have

$$g(\eta) = \eta$$

Since we have  $\mu = \theta$ , the canonical link function of a Gaussian distribution models the expectation of the target variable as of the following

$$\mu = \eta = \vec{x} \cdot \vec{\beta}$$

This looks a lot like the linear regression model, except that we don not know how GLM estimates the error function yet. In the following sections, we are goingn to discuss how to use MLE to optimize the a GLM.