Module 9: Exponential Families and Generalized Linear Regression

Rebecca C. Steorts

The Exponential Family

The probability density in the exponential family takes the following form:

$$p(x \mid \eta) = h(x) \exp\{\eta^T T(x) - A(\eta), \} \tag{1}$$

1. where η is the natural parameter
2. T, a, and h are functions

The Exponential Family

$$p(x \mid \eta) = h(x) \exp\{\eta^T T(x) - A(\eta)\} \tag{2}$$

1. The form of $h(x)$ is not of fundamental importance
2. $$A(\eta) = log \int h(x) \exp\{\eta^T T(x)\} \, dx$$

The Exponential Family

It is also common to write the exponential family in the following way:

$$p(x \mid \eta) = \frac{1}{Z(\eta)} h(x) \exp\{\eta^T T(x)\}, \tag{3}$$

which is equivalent to

$$A(\eta) = \log Z(\eta).$$

The Bernoulli distribution

Suppose

$$X \mid \pi \sim \text{Bernoulli}(\pi).$$

$$\begin{aligned} p(x \mid \pi) &= \pi^x (1-\pi)^{1-x} && (4) \\ &= \exp\{x \log(\pi) + (1-x)\log(1-\pi)\} && (5) \\ &= \exp\{x \log(\frac{\pi}{1-\pi}) + \log(1-\pi)\} && (6) \end{aligned}$$

What is the trick: Take the exponential of the log of the original distribution!

The Bernoulli distribution

$$p(x \mid \eta) = h(x) \exp\{\eta^T T(x) - A(\eta)\} \tag{7}$$

$$p(x \mid \pi) = \exp\{x \log(\frac{\pi}{1-\pi}) + \log(1-\pi)\} \tag{8}$$

- $\eta = \log \frac{\pi}{1-\pi} \implies \eta = \frac{e^\eta}{1+e^\eta} = \frac{1}{1+e^{-\eta}}$
- $T(x) = x$
- $A(\eta) = -log(1-\pi) = \log(1+e^\eta)$
- $h(x) = 1$

The Gaussian distribution

Let

$$X \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$$

It follows that

$$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{\frac{-1}{2\sigma^2}(x-\mu)^2\} \tag{9}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{1}{2\sigma^2}\mu^2 - \ln\sigma\} \tag{10}$$

Sufficiency

One nice property of exponential families is that T(X) will be sufficient for η.

Generalized Linear Models

We now turn to problems involving pairs of variables (X,Y), where both variables are assumed to be observed.

We summarize the structural component of the models as:

$$\mu = E[Y \mid X] = f(\theta^T x).$$

Example: In linear regression, $f(\cdot)$ is the identity function.

GLMS

There are two choices for GLMS

1. There is the choice of the exponential family distribution.
2. There is the choice of the response function $f(\cdot)$, which is often called the link function.

We will look at this for an example of logistic regression, where the response variable is the Bernoulli distribution.

Exercise: The Gaussian distribution

$$p(x \mid \eta) = h(x) \exp\{\eta^T T(x) - A(\eta)\} \tag{11}$$

$$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \exp\{\frac{\mu}{\sigma^2} x - \frac{1}{2\sigma^2} x^2 - \frac{1}{2\sigma^2} \mu^2 - \ln \sigma\} \tag{12}$$

$$\eta = (\frac{\mu}{\sigma^2}, \frac{-1}{2\sigma^2})^T \tag{13}$$

$$T(x) = (x, x^2)^T \tag{14}$$

$$A(\eta) = \frac{\mu^2}{2\sigma^2} + \ln \sigma = \frac{-\eta_1^2}{4\eta_2} - \frac{-1}{2} \ln(-2\eta_2) \tag{15}$$

$$h(x) = \frac{1}{\sqrt{2\pi}} \tag{16}$$

Remark: The univariate Gaussian is a two-parameter distribution, where its sufficient statistics will be a vector.