

# 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:

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

1. where  $\eta$  is the natural parameter
2.  $T$ ,  $a$ , and  $h$  are functions

# The Exponential Family

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

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:

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

which is equivalent to

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

# The Bernoulli distribution

Suppose

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

$$p(x \mid \pi) = \pi^x (1 - \pi)^{1-x} \quad (4)$$

$$= \exp\{x \log(\pi) + (1 - x) \log(1 - \pi)\} \quad (5)$$

$$= \exp\left\{x \log\left(\frac{\pi}{1 - \pi}\right) + \log(1 - \pi)\right\} \quad (6)$$

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

# The Bernoulli distribution

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

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

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

# The Gaussian distribution

Let

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

It follows that

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

$$= \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{1}{2\sigma^2}\mu^2 - \ln \sigma\right\} \quad (10)$$

# The Gaussian distribution

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

$$p(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{1}{2\sigma^2}\mu^2 - \ln \sigma\right\} \quad (12)$$



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



$$T(x) = (x, x^2)^T$$



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



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



# Sufficiency

Sufficiency characterizes what is “essential” in a data set.

In other words, it also tells us what we can throw away or ignore.

The notion of sufficiency goes beyond exponential families, however, there are strong ties to them.

# Sufficiency

A statistic  $T(X)$  is a function of a random variable  $X$ . Let the distribution of  $X$  depend upon the parameter  $\theta$ .

$T(X)$  is sufficient for  $\theta$  if there is no information in  $X$  regarding  $\theta$  beyond that in  $T(X)$ . That is, having observed  $T(X)$ , we can throw away the observed data  $X$  for inferential purposes.

# Bayesian Sufficiency

Let  $\theta$  be a random variable.

Then  $T(X)$  is sufficient for  $\theta$  if the following conditional independence holds:

$\theta$  is independent of  $X$  when we condition on  $T(X)$ .

# Frequentist sufficiency

Let  $\theta$  be an unknown, fixed parameter.

$T(X)$  is sufficient for  $\theta$  if the following holds:

$$p(x \mid T(x), \theta) = p(x \mid T(x))$$

Note: We can use this to find the Neyman factorization theorem, which is quite easy from the Bayesian point of view. It can be utilized under the frequentist one as well.

## Sufficiency and the exponential family

An important feature of the exponential family is that we can find the sufficient statistic from inspection. Specifically  $T(X)$  will be sufficient for  $\eta$ .

# GLMS

Then make the connection to GLMs