STAT 528 - Advanced Regression Analysis II

Exponential family theory

Daniel J. Eck
Department of Statistics
University of Illinois

Agenda for today

- Course software and GitHub
- Go over basics of exponential family theory

Exponential family

An *exponential family of distributions* is a parametric statistical model having log likelihood that takes the form

$$l(\theta) = \langle y, \theta \rangle - c(\theta), \tag{1}$$

where y is a vector statistic and θ is a vector parameter, and

- $\langle y, \theta \rangle$ is the usual inner product,
- $c(\theta)$ is the cumulant function.

This uses the convention that terms that do not contain the parameter vector can be dropped from a log likelihood; otherwise additional terms also appear in (1).

When the log likelihood can be expressed as (1) we say that y is the **canonical statistic** and θ is the **canonical parameter**.

Example: Binomial distribution

Let $X \sim \text{Binomial}(n,p)$ where $0 < p < 1$. We can write the log probability mass function for X

$$
\begin{aligned}
l(p) &= \log\left(\binom{n}{x}\right) + x\log(p) + (n-x)\log(1-p) \\
&\propto \log\left(\binom{n}{x}\right) + x\log(p) + (n-x)\log(1-p)
\end{aligned}
$$

in exponential family form

$$l(\theta) = \langle y, \theta \rangle - c(\theta).$$

Densities

Let w represent the full data, then the densities have the form

$$f_\theta(w) = h(w) \exp\left(\langle Y(w), \theta \rangle - c(\theta)\right) \tag{2}$$

and the word "density'' here can refer to a PMF, PDF, or to a density with respect to a positive measure.

The $h(w)$ arises from any term not containing the parameter that is dropped in going from log densities to log likelihood as we saw on the previous slide.

The function h has to be nonnegative, and any point w such that $h(w) = 0$ is not in the support of any distribution in the family.

Example: Binomial distribution

Let $X \sim \text{Binomial}(n,p)$ where $0 < p < 1$. We can write the probability mass function for X

$$f_p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

as an exponential family density

$$f_\theta(w) = h(w) \exp\left(\langle Y(w), \theta \rangle - c(\theta)\right).$$

Example: Normal distribution

Let $W \sim N(\mu, \sigma^2)$. Then we can write

$$f_{\mu,\sigma^2}(w) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(w-\mu)^2}{2\sigma^2}\right)$$

as an exponential family density

$$f_\theta(w) = h(w) \exp\left(\langle Y(w), \theta \rangle - c(\theta)\right),$$

where

$$c(\theta) = \frac{1}{2}\left(\frac{\theta_1^2}{2\theta_2} - \log(2\theta_2)\right).$$

Cumulant functions

Being a density, (2) must sum, integrate, or sum-integrate to one. Hence,

$$\begin{aligned} 1 &= \int f_\theta(w) dw \\ &= \int \exp\left(\langle Y(w), \theta \rangle - c(\theta)\right) h(w) dw \\ &= \exp\left(-c(\theta)\right) \int \exp\left(\langle Y(w), \theta \rangle\right) h(w) dw. \end{aligned}$$

Rearranging the above implies that

$$c(\theta) = \log\left(\int \exp\left(\langle Y(w), \theta \rangle\right) h(w) dw \right).$$

The cumulant function is the log Laplace transformation corresponding to the *generating measure* given by

$$\lambda(dw) = h(w)dw$$

when the random variable is continuous. Under this formulation

$$c(\theta) = \log\left(\int \exp\left(\langle Y(w), \theta\rangle\right)\lambda(dw)\right).$$

In our log likelihood based definition of the exponential family (1), the dropped terms which do not appear in the log likelihood are incorporated into the counting measure (discrete distributions) or Lebesgue measure (continuous distributions).