STAT 528 - Advanced Regression Analysis II

Exponential family theory (part 3)

Daniel J. Eck
Department of Statistics
University of Illinois

Last time

- mean value parameters
- maximum likelihood estimation (MLE)
- asymptotics of MLE
- finite sample concentration of MLE

Learning Objectives Today

- generalized linear models (GLMs)
- different parameterizations

Canonical linear submodels: intro to GLMs

We now motivate generalized linear models (GLMs) within the context of exponential theory.

A canonical affine submodel of an exponential family is a submodel having parameterization

$$\theta = a + M\beta$$

where:

- $\theta \in \mathbb{R}^n$ is the canonical parameter vector
- $\beta \in \mathbb{R}^p$ is the canonical parameter vector for the submodel
- $a \in \mathbb{R}^n$ is a known offset vector
- $M \in \mathbb{R}^{n \times p}$ is a known *model matrix*.

In most applications the offset vector is not used giving parameterization

$$\theta = M\beta,$$

in which case we say the submodel is a *canonical linear submodel*.

We will restrict attention to the canonical linear submodel in this class.

The canonical linear submodel log likelihood is given by

$$
\begin{aligned}
l(\theta) &= \langle y, \theta \rangle - c(\theta) \\
&= \langle y, M\beta \rangle - c(M\beta) \\
&= \langle M'y, \beta \rangle - c_\beta(\beta),
\end{aligned}
\tag{1}
$$

and we see that we again have an exponential family with

- canonical statistic $M'y$
- cumulant function $\beta \mapsto c_\beta(\beta) = c(M\beta)$
- submodel canonical parameter vector β

Let's step back a bit.

In applications n denotes the sample size. $\theta \in \mathbb{R}^n$ is an arbitrary vector specifying one parameter for individual. This is a saturated model.

For a model to be useful, we need dimension reduction

$$\theta = M\beta.$$

In other words, $\theta \in \mathrm{span}(M)$.

Full regular submodels

If the originally given full canonical parameter space was Θ, then the full submodel canonical parameter space is

$$B = \{\beta : M\beta \in \Theta\}.$$

Thus a canonical linear submodel gives us a new exponential family, with lower-dimensional canonical parameter and statistic.

The submodel exponential family is full regular if the original exponential family was full regular.

Parameterizations

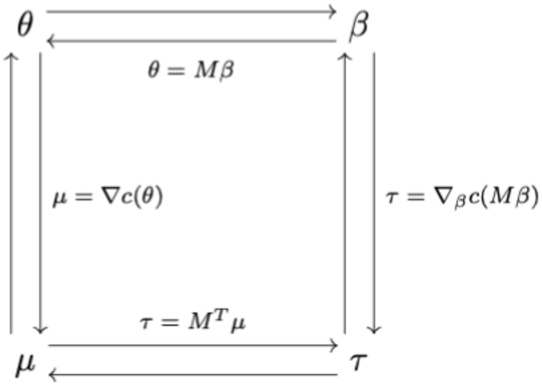
Now we have four parameters:

- the saturated model canonical and mean value parameters θ and μ
- the canonical linear submodel canonical and mean value parameters β and $\tau = M^T \mu$.

The observed equals expected property for the submodel is

$$\hat{\tau} = M^T \hat{\mu} = M^T y. \tag{2}$$

A depiction of the transformations necessary to change between parameterizations.



Note: $\mu \to M^T \mu$ is usually not one-to-one (when $n > p$ and M is full column rank).

Hence we cannot determine $\hat{\theta}$ and $\hat{\beta}$ from them either.

The only way to determine the MLEs is to maximize the log likelihood (1) to obtain $\hat{\beta}$ and then

- $\hat{\theta} = M\hat{\beta}$
- $\hat{\mu} = \nabla c(\hat{\theta})$
- $\hat{\tau} = M^T \hat{\mu}$.

GLMs and link functions

Recall that the saturated model canonical parameter vector θ is *linked* to the saturated model mean value parameter vector through the change-of-parameter mappings $g(\theta)$.

We can reparameterize $\theta = M\beta$ and write

$$\mu = \mathrm{E}_\theta(Y) = g(M\beta)$$

which implies that we can write

$$g^{-1}\left(\mathrm{E}_\theta(Y)\right) = M\beta.$$

Therefore, a linear function of the canonical submodel parameter vector is linked to the mean of the exponential family through the inverse change-of-parameter mapping g^{-1}.

This is the basis of exponential family generalized linear models with link function g^{-1}.

Note that most treatments of GLMs will present g^{-1} as the link function. Instead we motivated a change of parameters mapping from canonical to mean value parameters.

Example: logistic regression

Let M have rows x_i', $i = 1, \ldots, n$. Then $\theta_i = x_i'\beta$ and

$$x_i'\beta = \log\left(\frac{p_i}{1 - p_i}\right) \qquad \text{and} \qquad p_i = \left(\frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)}\right),$$

where

$$\begin{aligned}
&\sum_{i=1}^{n} [y_i \log(p_i) - (1 - y_i) \log(p_i)] = \sum_{i=1}^{n} \left[y_i x_i'\beta - \log(1 + \exp(x_i'\beta))\right] \\
&= \sum_{i=1}^{n} \left[\langle y_i, x_i'\beta\rangle - \log(1 + \exp(x_i'\beta))\right] \\
&= \langle y, \theta\rangle - \sum_{i=1}^{n} \log(1 + \exp(\theta_i)) \\
&= \langle y, \theta\rangle - c(\theta).
\end{aligned}$$

Alternatively,

$$\begin{aligned}
&\sum_{i=1}^{n} \left[\langle y_i, x_i'\beta\rangle - \log(1+\exp(x_i'\beta))\right] \\
&= \langle \sum_{i=1}^{n} y_i x_i, \beta\rangle - \sum_{i=1}^{n} \log(1+\exp(x_i'\beta)) \\
&= \langle M'y, \beta\rangle - c_\beta(\beta)
\end{aligned}$$