Lecture 2. Exponential Families

Nan Ye

School of Mathematics and Physics
University of Queensland

Recall: definition of GLM

- A GLM has the following structure

$$
\begin{aligned}
\text{(systematic)} \quad & \mathbb{E}(Y \mid \mathbf{x}) = h(\beta^\top \mathbf{x}). \\
\text{(random)} \quad & Y \mid \mathbf{x} \text{ follows an exponential family distribution.}
\end{aligned}
$$

- This is usually separated into three components
  - The linear predictor $\beta^\top \mathbf{x}$.
  - The response function h.
    People often specify the link function $g = h^{-1}$ instead.
  - The exponential family for the conditional distribution of Y given $\mathbf{x}$.

Recall: remarks on exponential families

- It is common!

  normal, Bernoulli, Poisson, and Gamma distributions are exponential families.

- It gives a well-defined model.

  its parameters are determined by the mean $\mu = \mathbb{E}(Y \mid \mathbf{x})$.

- It leads to a unified treatment of many different models.

  linear regression, logistic regression, ...

This Lecture

- One-parameter exponential family: definition and examples.
- Maximum likelihood estimation: score equation, MLE as moment matching.

Exponential Families

One parameter exponential family

An exponential family distribution has a PDF (or PMF for a discrete random variable) of the form

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

- $T(y)$ is called the *natural statistic*.
- $\eta(\theta)$ is called the *natural parameter*.
- ϕ is a *nuisance parameter* treated as known.

Range of y must be independent of θ.

Special forms

- A *natural* exponential family is one parametrized by η, that is, $\theta = \eta$, and the PDF/PMF can be written as

$$f(y \mid \eta, \phi) = \exp\left(\frac{\eta T(y) - A(\eta)}{b(\phi)} + c(y, \phi)\right)$$

- The nuisance parameter can be absent, in which case b and c have the form $b(\phi) = 1$ and $c(y, \phi) = c(y)$.

General form

- In general, $\eta(\theta)$ and $T(y)$ can be vector-valued functions, θ can be a vector, and y can be an arbitrary object (like strings, sequences).
- Of course, $\eta(\theta) T(y)$ need to be replaced by the inner product $\eta(\theta)^\top T(y)$ then.
- In practice, this general form has been heavily used in various domains such as computer vision, natural language processing.
- For this course, we focus on the single-parameter case, but the properties studied here can be easily generalized for the general form.

Examples

Example 1. Gaussian distribution (known σ^2)

- The PDF is

$$f(y \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$
$$= \exp\left(-\frac{y^2}{2\sigma^2} + \frac{\mu y}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \ln\sqrt{2\pi\sigma^2}\right),$$

- This is a natural exponential family with a nuisance parameter σ
  - $\eta(\mu) = \mu$.
  - $T(y) = y$.
  - $A(\mu) = \mu^2/2$.
  - $b(\sigma) = \sigma^2$.
  - $c(y, \sigma) = -\frac{y^2}{2\sigma^2} - \ln\sqrt{2\pi\sigma^2}$.

Example 2. Bernoulli distribution

- The PMF is

$$\begin{aligned} f(y \mid p) &= p^y(1-p)^{1-y} \\ &= \exp\left(y \ln p + (1-y)\ln(1-p)\right) \\ &= \exp\left(y \ln \frac{p}{1-p} + \ln(1-p)\right) \end{aligned}$$

- This is an exponential family without a nuisance parameter
    - $\eta(p) = \ln \frac{p}{1-p}$.
    - $T(y) = y$.
    - $A(p) = -\ln(1-p)$.
    - $c(y) = 0$.

- The natural form is given by

$$f(y \mid \eta) = \frac{e^{y\eta}}{1 + e^{\eta}}.$$

Example 3. Poisson distribution

- The PMF is

$$f(y \mid \lambda) = \frac{\lambda^y}{y!}\exp(-\lambda)$$
$$= \exp(-\lambda + y \ln \lambda - \ln y!)$$

- This is an exponential family without a nuisance parameter
  - $\eta(\lambda) = \ln \lambda$.
  - $T(y) = y$.
  - $A(\lambda) = \lambda$.
  - $c(y) = -\ln y!$.

Example 4. Gamma distribution (fixed shape)

- The PDF, parametrized using the shape k and scale θ, is

$$f(y \mid k, \theta) = \frac{y^{k-1}}{\Gamma(k)\theta^k} e^{-y/\theta}.$$

- An alternative parametrization, sometimes easier to work with, is to use two parameters μ and ν such that μ is the mean, and the variance is μ^2/ν.

- In the (k, θ) parametrization, the mean and variance are $k\theta$ and $k\theta^2$ respectively, thus $k = \nu$ and $\theta = \mu/\nu$, and

$$
\begin{aligned}
f(y \mid \mu, \nu) &= \frac{y^{\nu-1}\nu^{\nu}}{\Gamma(\nu)\mu^{\nu}} e^{-y\nu/\mu} \\
&= \exp\left(-\nu\left(y\frac{1}{\mu} + \ln\mu\right) + (\nu-1)\ln y + \nu\ln\nu - \ln\Gamma(\nu)\right)
\end{aligned}
$$

- This is an exponential family with a nuisance parameter ν
  - $\eta(\mu) = 1/\mu$.
  - $T(y) = y$.
  - $A(\mu) = \ln\mu$
  - $b(\nu) = 1/\nu$.
  - $c(y, \nu) = (\nu - 1)\ln y + \nu\ln\nu - \ln\Gamma(\nu)$.

Example 5. Negative binomial (fixed r)

- The PMF of $N(r, p)$ is

$$f(y \mid p, r) = \binom{y+r-1}{y}(1-p)^r p^y$$
$$= \exp\left(\ln\binom{y+r-1}{y} + r\ln(1-p) + y\ln p\right)$$

- This is an exponential family when r is fixed
  - $\eta(p) = \ln p$.
  - $T(y) = y$.
  - $A(p) = -r\ln(1-p)$.
  - $c(y) = \ln\binom{y+r-1}{y}$.

MLE

Score equation

The MLE θ is the solution to

$$\frac{A'(\theta)}{\eta'(\theta)} = \tilde{T},$$

where $\tilde{T} = \frac{1}{n} \sum_i T(y_i)$.

Derivation

- The log-likelihood of θ given $y_1, \ldots, y_n$ is

$$\ell(\theta) = \sum_i \frac{\eta(\theta) T(y_i) - A(\theta)}{b(\phi)} + c(y_i, \phi).$$

- Set the derivative of the log-likelihood to 0, and rearrange to obtain $A'(\theta)/\eta'(\theta) = \tilde{T}$.

MLE as moment matching

- We will see that $\mathbb{E}(T \mid \theta) = A'(\theta)/\eta'(\theta)$ in next lecture.
- Thus the score equation is $\mathbb{E}(T \mid \theta) = \tilde{T}$.
- MLE is thus matching the first moment of the exponential family to that of the empirical distribution (i.e., a method of moment estimator).

Example. MLE for negative binomial distribution

- Recall: for fixed r, $A(p) = -r \ln(1-p)$, $\eta(p) = \ln p$.
- The score equation is $\frac{rp}{1-p} = \bar{y} = \frac{1}{n}\sum_i y_i$.
- $p = \bar{y}/(\bar{y}+r)$.

What You Need to Know

- Recognise whether a distribution is a one-parameter exponential family.
- Score equation for one-parameter exponential family
- Sufficiency of $\tilde{T}$ for determining MLE.
- MLE as moment matching.