Lecture 07: Exponential Families

Mathematical Statistics I, MATH 60061/70061

Tuesday September 21, 2021

Reference: Casella & Berger, 3.4

Exponential families

A family of PDFs or PMFs indexed by $\boldsymbol{\theta}$ is called an **exponential family** if it can be expressed as

$$f(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta},$$

where $\boldsymbol{\Theta}$ is the set of all values of $\boldsymbol{\theta}$ (parameter space), $h(x) \geq 0$ and $t_1(x), \ldots, t_k(x)$ are real-valued functions of observation x (not depending on $\boldsymbol{\theta}$), and $c(\boldsymbol{\theta}) \geq 0$ and $w_1(\boldsymbol{\theta}), ..., w_k(\boldsymbol{\theta})$ are functions of the possibly vector-valued $\boldsymbol{\theta}$ (not depending on x).

Note that the expression for f may not be unique.

Exponential families

Many common families introduced in the previous lectures are exponential families. These include

- Continuous families: Normal, Gamma, Beta
- Discrete families: Binomial, Poisson, Negative Binomial

To verify that a family of PDFs or PMFs is an exponential family, we must identify the functions $h(x)$, $c(\boldsymbol{\theta})$, $w_i(\boldsymbol{\theta})$, and $t_i(x)$ and show that the family has the given form

$$f(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right).$$

Binomial exponential family

Let n be a positive integer and consider the $\mathrm{Bin}(n,p)$ family with $0<p<1$. Then the PMF for this family, for $x=0,\ldots,n$ and $0<p<1$, is

$$\begin{aligned} f(x \mid p) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \binom{n}{x} (1-p)^n \left(\frac{p}{1-p}\right)^x \\ &= \binom{n}{x} (1-p)^n \exp\left(\log\left(\frac{p}{1-p} x\right)\right). \end{aligned}$$

Binomial exponential family

Define

$$h(x) = \begin{cases} \binom{n}{x} & x = 0, 1, \dots, n \\ 0 & \text{otherwise}, \end{cases} \qquad t_1(x) = x,$$

$$c(p) = (1-p)^n, \quad w_1(p) = \log\left(\frac{p}{1-p}\right), \quad 0 < p < 1.$$

Then we have

$$\begin{aligned} f(x \mid p) &= \binom{n}{x}(1-p)^n \exp\left(\log\left(\frac{p}{1-p}x\right)\right) \\ &= h(x)c(p)\exp[w_1(p)t_1(x)]. \end{aligned}$$

So, the Binomial family with $0 < p < 1$ and a fixed n is an exponential family ($k = 1$).

Moments of an exponential family

Exponential families have many nice properties. The following result is a useful calculational shortcut for moments of an exponential family.

If X is a random variable with PDF or PMF from an exponential family and $w_i(\boldsymbol{\theta})$'s are differentiable functions, then

$$E\left(\sum_{i=1}^{k}\frac{\partial w_i(\boldsymbol{\theta})}{\partial\theta_j}t_i(X)\right) = -\frac{\partial\log c(\boldsymbol{\theta})}{\partial\theta_j},$$

$$\mathrm{Var}\left(\sum_{i=1}^{k}\frac{\partial w_i(\boldsymbol{\theta})}{\partial\theta_j}t_i(X)\right) = -\frac{\partial^2\log c(\boldsymbol{\theta})}{\partial\theta_j^2} - E\left(\sum_{i=1}^{k}\frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial\theta_j^2}t_i(X)\right),$$

where θ_j is the jth component of $\boldsymbol{\theta}$.

Proof

From the exponential family expression for f,

$$\log f(X \mid \boldsymbol{\theta}) = \log h(X) + \log c(\boldsymbol{\theta}) + \sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(X)$$

Differentiating this expression leads to

$$\frac{\partial \log f(X \mid \boldsymbol{\theta})}{\partial \theta_j} = \frac{\partial \log c(\boldsymbol{\theta})}{\partial \theta_j} + \sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)$$

Taking expectation, we obtain

$$E\left(\frac{\partial \log f(X \mid \boldsymbol{\theta})}{\partial \theta_j}\right) = \frac{\partial \log c(\boldsymbol{\theta})}{\partial \theta_j} + E\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right)$$

If $f(x \mid \boldsymbol{\theta})$ is a PDF (the proof for PMF is similar), then the left side of the previous expression is

$$
\begin{aligned}
E\left(\frac{\partial \log f(X \mid \boldsymbol{\theta})}{\partial \theta_j}\right) &= \int_{-\infty}^{\infty} \frac{\partial \log f(x \mid \boldsymbol{\theta})}{\partial \theta_j} f(x \mid \boldsymbol{\theta}) dx \\
&= \int_{-\infty}^{\infty} \frac{\partial f(x \mid \boldsymbol{\theta})}{\partial \theta_j} dx \\
&= \frac{\partial}{\partial \theta_j} \int_{-\infty}^{\infty} f(x \mid \boldsymbol{\theta}) dx \\
&= \frac{\partial 1}{\partial \theta_j} = 0
\end{aligned}
$$

We interchanged the differentiation and integration, which is justified under the exponential family assumption.

This proves the first result.

Note that

$$\frac{\partial^2 \log f(X \mid \boldsymbol{\theta})}{\partial \theta_j^2} = \frac{\partial}{\partial \theta_j}\left[\frac{\frac{\partial f(X \mid \boldsymbol{\theta})}{\partial \theta_j}}{f(X \mid \boldsymbol{\theta})}\right] = \frac{\frac{\partial^2 f(X \mid \boldsymbol{\theta})}{\partial \theta_j^2}}{f(X \mid \boldsymbol{\theta})} - \left[\frac{\frac{\partial f(X \mid \boldsymbol{\theta})}{\partial \theta_j}}{f(X \mid \boldsymbol{\theta})}\right]^2$$

Then

$$\begin{aligned}
E\left(\frac{\partial^2 \log f(X \mid \boldsymbol{\theta})}{\partial \theta_j^2}\right) &= \int_{-\infty}^{\infty}\left\{\frac{\frac{\partial^2 f(x \mid \boldsymbol{\theta})}{\partial \theta_j^2}}{f(x \mid \boldsymbol{\theta})} - \left[\frac{\frac{\partial f(x \mid \boldsymbol{\theta})}{\partial \theta_j}}{f(x \mid \boldsymbol{\theta})}\right]^2\right\} f(x \mid \boldsymbol{\theta}) dx \\
&= \int_{-\infty}^{\infty} \frac{\partial^2 f(x \mid \boldsymbol{\theta})}{\partial \theta_j^2} dx - \int_{-\infty}^{\infty}\left[\frac{\partial \log f(x \mid \boldsymbol{\theta})}{\partial \theta_j}\right]^2 f(x \mid \boldsymbol{\theta}) dx \\
&= -\int_{-\infty}^{\infty}\left[\frac{\partial \log c(\boldsymbol{\theta})}{\partial \theta_j} + \sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(x)\right]^2 f(x \mid \boldsymbol{\theta}) dx \\
&= -\mathrm{Var}\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right).
\end{aligned}$$

The last equality follows from the first result.

Then the second result follows from

$$\frac{\partial^2 \log f(X \mid \boldsymbol{\theta})}{\partial \theta_j^2} = \frac{\partial^2 \log c(\boldsymbol{\theta})}{\partial \theta_j^2} + \sum_{i=1}^{k} \frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial \theta_j^2} t_i(X)$$

Binomial mean and variance

$$
\begin{aligned}
f(x \mid p) &= \binom{n}{x}(1-p)^n \exp\left(\log\left(\frac{p}{1-p}x\right)\right) \\
&= h(x)c(p)\exp[w_1(p)t_1(x)].
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dp}w_1(p) &= \frac{d}{dp}\log\frac{p}{1-p} = \frac{1}{p(1-p)} \\
\frac{d}{dp}\log c(p) &= \frac{d}{dp}n\log(1-p) = \frac{-n}{1-p},
\end{aligned}
$$

so we have

$$
E\left(\frac{1}{p(1-p)}X\right) = \frac{n}{1-p} \Rightarrow E(X) = np.
$$

The variance identity works in a similar manner.

Normal exponential family

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\boldsymbol{\theta} = (\mu, \sigma)$

$$
\begin{aligned}
f(x \mid \mu, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right).
\end{aligned}
$$

Define

$$h(x) = 1 \text{ for all } x$$

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

$$w_1(\boldsymbol{\theta}) = 1/\sigma^2, \quad w_2(\boldsymbol{\theta}) = \mu/\sigma^2$$

$$t_1(x) = -x^2/2, \quad t_2(x) = x$$

So, this Normal family is an exponential family with $k = 2$.

Normal mean and variance

With

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

$$w_1(\boldsymbol{\theta}) = 1/\sigma^2,\ w_2(\boldsymbol{\theta}) = \mu/\sigma^2,\ t_1(x) = -x^2/2,\ t_2(x) = x$$

we obtain $E(X) = \mu$ from equation

$$-\frac{\partial \log c(\boldsymbol{\theta})}{\partial \mu} = \frac{\mu}{\sigma^2} = E\left(\sum_{i=1}^{2} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \mu} t_i(X)\right) = E\left(\frac{X}{\sigma^2}\right)$$

Also,

$$-\frac{\partial \log c(\boldsymbol{\theta})}{\partial \sigma} = \frac{\mu^2}{\sigma^3} + \frac{1}{\sigma} = E\left(\sum_{i=1}^{2} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \sigma} t_i(X)\right) = E\left(\frac{X^2}{\sigma^3} - \frac{2\mu X}{\sigma^3}\right)$$

Using $E(X) = \mu$, we obtain from this equation that $\mathrm{Var}(X) = \sigma^2$.

Natural exponential family

If $\eta_i = w_i(\boldsymbol{\theta})$, $i = 1, \ldots, k$, and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)$, the form of the exponential family becomes

$$f(x \mid \boldsymbol{\eta}) = h(x) c^*(\boldsymbol{\eta}) \exp\left(\sum_{i=1}^{k} \eta_i t_i(x) \right),$$

where $\boldsymbol{\eta}$ is called the natural parameter.

- The set of $\boldsymbol{\eta}$'s for which $f(x \mid \boldsymbol{\eta})$ is a well-defined PDF or PMF is called the **natural parameter space**.
- The natural parameter space is *convex*, among other useful mathematical properties.

Full or curved exponential families

In the exponential family representation,

$$f(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta},$$

if the dimension of $\boldsymbol{\theta}$ is k, then the family is a **full exponential family**. If the dimension of $\boldsymbol{\theta}$ is less than k, the family is a **curved exponential family**.

- An example of a full exponential family is $\mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$.
- An example of a curved exponential family is $\mathcal{N}(\mu, \mu^2)$, $\mu \in \mathbb{R}$.