Lecture 31: Nov 8

Last time

- Location Scale Families
- Exponential Family

Today

- Exponential Family

Exponential Families A family of pdfs or pmfs with vector parameter $\boldsymbol{\theta}$ is called an *exponential family* if it can be expressed as

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta})t_j(x)\right), \quad x \in S \subset \mathbb{R} \tag{1}$$

where S is not defined in terms of $\boldsymbol{\theta}$, $h(x)$, $c(\boldsymbol{\theta}) \geqslant 0$ and the functions are just functions of the parameters specified; i.e. h is free of $\boldsymbol{\theta}$, $c(\boldsymbol{\theta})$ is free of x, etc...

Examples:

- One-dimensional: Exponential, Poisson
- Two-dimensional: Gaussian

Exponential family parameterizations are unique except for multiplying constant factors.

Example: Gaussian Let $f(x|\mu,\sigma^2)$ be the $n(\mu,\sigma^2)$ family of pdfs, where $\boldsymbol{\theta} = (\mu,\sigma)$. Then

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

Thus

$$\begin{aligned} h(x) &= \tfrac{1}{\sqrt{2\pi}} & c(\mu,\sigma) &= \tfrac{1}{\sigma}\exp\left(-\tfrac{\mu^2}{2\sigma^2}\right) \\ w_1(\mu,\sigma) &= -\tfrac{1}{2\sigma^2} & w_2(\mu,\sigma) &= \tfrac{\mu}{\sigma^2} \\ t_1(x) &= x^2 & t_2(x) &= x \end{aligned}$$

The parameter space is $(\mu,\sigma^2) \in \mathbb{R} \times (0,\infty)$.

Example: Binomial Let $f(x|p)$ be the $binomial(n,p)$, $0 < p < 1$ family of pmfs.

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

Thus,

$$\begin{array}{ll} h(x) = \binom{n}{x}, \quad x = 0, \ldots, n & w_1(p) = \log\left(\frac{p}{1-p}\right) \\ c(p) = (1-p)^n, 0 < p < 1 & t_1(x) = x \end{array}$$

Note that this works when p is considered the parameter, while n is fixed. Also, p cannot be 0 or 1. Otherwise, the range changes.

More examples The following distributions belong to Exponential families:

- Continuous: exponential, Gaussian, gamma, beta, χ^2
- Discrete: Poisson, geometric, binomial (fixed # trials), negative binomial (fixed # successes)

The following distributions not exponential families:

- Continuous: t, F, unifrom E.g.:$X \sim U(0, \theta)$

$$f_X(x) = \theta^{-1} 1(0 < x < \theta)$$

- Discrete: uniform, hypergeometric

Theorem If X is a random variable with pdf or pmf of the form 1, then

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

Although these equations may look formidable, when applied to specific cases they can work out quite nicely. Their advantage is that we can replace integration or summation by differentiation, which is often more straightforward.

Example (Normal exponential family) Let $f(x|\mu, \sigma^2)$ be the $N(\mu, \sigma^2)$ family of pdfs, where $\boldsymbol{\theta} = (\mu, \sigma), -\infty < \mu < \infty, \sigma > 0$. Then

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

Define

$$\theta_1 = \tfrac{1}{\sigma^2} > 0, \quad \theta_2 = \tfrac{\mu}{\sigma^2} \in \mathbb{R}$$

Then

$$f_X(x) = \frac{\sqrt{\theta_1}}{\sqrt{2\pi}} \exp\left(-\frac{\theta_2^2}{2\theta_1}\right) \exp\left(-\theta_1 \frac{x^2}{2} + \theta_2 x\right)$$

and

$$
\begin{array}{ll}
h(x) = 1 \text{ for all } x; & \\
c(\boldsymbol{\theta}) = c(\theta_1, \theta_2) = \exp\left(-\frac{\theta_2^2}{2\theta_1}\right), & (\theta_1, \theta_2) \in (0, \infty) \times \mathbb{R} \\
w_1(\boldsymbol{\theta}) = \theta_1 & t_1(x) = -x^2/2 \\
w_2(\boldsymbol{\theta}) = \theta_2 & t_2(x) = x
\end{array}
$$

Therefore, by the above theorem

$$
\begin{aligned}
E(X) &= -\frac{\partial}{\partial \theta_2} \log c(\boldsymbol{\theta}) = \frac{\theta_2}{\theta_1} = \mu \\
Var(X) &= -\frac{\partial^2}{\partial \theta_2^2} \log c(\boldsymbol{\theta}) = -\frac{1}{\theta_1} = \sigma^2
\end{aligned} \tag{2}
$$