Exponential Families of Distributions

BS2 Statistical Inference, Lecture 5
Michaelmas Term 2004

Steffen Lauritzen, University of Oxford; October 26, 2004

Curved Exponential Families

A family $\mathcal{F} = \{f(\cdot;\theta), \theta \in \Theta\}$ is said to be (curved) *exponential* if the densities have the form

$$f(x;\theta) = b(x)e^{a(\theta)^\top t(x) - c(\theta)},$$

where $b(x)$ is known, $t(x)^\top = (t_1(x), \ldots, t_k(x))$ is a vector of known real-valued functions, $a(\theta) = (a_1(\theta), \ldots, a_k(\theta))$ are twice continuously differentiable functions of $\theta \in \Theta$, and Θ is an open and connected subset of $\mathcal{R}^d$ with $d \leq k$. We also assume that the Jacobian

$$J(\theta) = \{j_{rs}(\theta)\} = \left\{ \frac{\partial a_r(\theta)}{\partial \theta_s} \right\}$$

has everywhere full rank d.

Since the density integrates to one we must have

$$c(\theta) = \log \int b(x) e^{a(\theta)^\top t(x)}\, dx,$$

or a similar expression with the integral replaced by a sum, when X is discrete.

The representation of the family is said to be *minimal* if $(1, t_1, \ldots, t_k)$ are linearly independent, i.e. if no linear combination exists so that

$$\alpha_0 + \alpha_1 t_1(X) + \cdots + \alpha_k t_k(X) = 0.$$

Then the *dimension* of the family is equal to d, the rank of $J(\theta)$.

In the following we always assume that families are minimally represented.

Example

Most interesting families of distributions are (curved) exponential families. Consider for example $X \sim \mathcal{N}(\mu, \sigma^2)$ where μ and $\sigma^2 > 0$ are both unknown.

If we rewrite the density as

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2} + x\frac{\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log\sigma^2}$$

we clearly see the structure of an exponential family with $t_1(x) = x^2$, $t_2(x) = x$,

$$a_1(\mu, \sigma^2) = \frac{-1}{2\sigma^2}, \quad a_2(\mu, \sigma^2) = \frac{\mu}{\sigma^2},$$

and

$$c(\mu, \sigma^2) = \frac{\mu^2}{2\sigma^2} + \frac{1}{2}\log\sigma^2.$$

Stability under repeated sampling

If $X = (X_1, \ldots, X_n)$ is an i.i.d. sample from an exponential family of individual distributions, the joint distribution is also exponential and of *the same dimension*

$$
\begin{aligned}
f(x;\theta) &= \prod_i \left\{ b(x_i) e^{a(\theta)^\top t(x_i) - c(\theta)} \right\} \\
&= \left\{ \prod_i b(x_i) \right\} e^{a(\theta)^\top \sum_i t(x_i) - nc(\theta)}.
\end{aligned}
$$

In particular the functions $a_r(\theta)$ are *unchanged*, the statistics $t_r(X) = \sum_i t_r(X_i)$ are *obtained by simple summation,* and $c_n(\theta) = nc(\theta)$.

Sufficiency in exponential families

Clearly, from the factorization theorem it follows that the statistic $T = t(X)$ is sufficient. Indeed it can be shown that *if the family is minimally represented, T is minimal sufficient*.

If $d = k$ we say the exponential family is *linear*.

If the family is linear, $T = t(X)$ is complete and sufficient.

The proof of the latter (important) result involves complex function theory, in particular analytic extension, and is outside the scope of this course.

Canonical parametrizations

Things become particularly simple when $a_r(\theta) = \theta_r$. The parametrization is then said to be *canonical* and the family is said to be a *canonical* family.

By differentiation under the integral sign in this expression, we get *for a canonical family* that

$$\mathbf{E}_\theta\{t_r(X)\} = \frac{\partial}{\partial\theta_r}c(\theta), \quad \mathrm{Cov}_\theta\{t_r(X), t_s(X)\} = \frac{\partial^2}{\partial\theta_r\theta_s}c(\theta).$$

Note that the assumptions already made imply that *differentiation and integration can indeed be interchanged.*

Example

Consider the family of Gamma distributions with

$$f(x;\alpha,\beta) \quad = \quad \frac{\beta^{\alpha}x^{\alpha-1}}{\Gamma(\alpha)}e^{-\beta x} = \frac{1}{x}e^{\alpha\log x-\beta x+\alpha\log\beta-\log\Gamma(\alpha)}.$$

If $\alpha > 0$ and $\beta > 0$ are unknown, this is clearly a linear and canonical exponential family with $b(x) = 1/x$ so

$$\mathbf{E}(X) = -\mathbf{E}(-X) = -\frac{\partial}{\partial\beta}\{-\alpha\log\beta + \log\Gamma(\alpha)\} = \alpha/\beta$$

and

$$\mathbf{V}(X) = \mathbf{V}(-X) = \frac{\partial^2}{\partial\beta^2}\{-\alpha\log\beta + \log\Gamma(\alpha)\} = \alpha/\beta^2.$$

We also get

$$\mathbf{E}(\log X) = \frac{\partial}{\partial\alpha}\left\{-\alpha\log\beta + \log\Gamma(\alpha)\right\} = \psi(\alpha) - \log\beta$$

and similarly

$$\begin{aligned}\mathbf{V}(\log X) &= \mathbf{V}(-\log X)\\ &= \frac{\partial^2}{\partial\alpha^2}\left\{-\alpha\log\beta + \log\Gamma(\alpha)\right\} = \psi'(\alpha)\end{aligned}$$

where $\psi(\alpha) = D\log\Gamma(\alpha)$ is the *Digamma* function and $\psi'(\alpha) = D^2\log\Gamma(\alpha)$ the *Trigamma* function.

Note that

$$\psi(1) = -\gamma = -0.5772156649\ldots$$

where γ is Euler's constant.

Score and Fisher information

For curved exponential families, *the required regularity conditions are fulfilled* and for a linear exponential family we get the (multivariate) score and Fisher information as

$$\begin{aligned} S_r(\theta) &= \frac{\partial}{\partial\theta_r}\log f(X;\theta) \\ &= t_r(X) - \frac{\partial}{\partial\theta_r}c(\theta) = t_r(X) - \mathbf{E}_\theta\{t_r(X)\} \end{aligned}$$

and

$$\begin{aligned} i_{rs}(\theta) &= \mathrm{Cov}\{S_r(\theta), S_s(\theta)\} \\ &= \mathbf{E}\left\{-\frac{\partial^2}{\partial\theta_r\theta_s}\log f(X;\theta)\right\} = \frac{\partial^2}{\partial\theta_r\theta_s}c(\theta). \end{aligned}$$