Econ 719, N. M. Kiefer

Exponential Families

The model function for the exponential family model M is given by

$$p(x;\theta) = \exp\{\theta \bullet t(x) - \kappa(\theta) - h(x)\}$$

θ is the canonical parameter and $t(x)$ the canonical statistic. Here x is a continuous or discrete, possibly vector valued random variable. The function $h(x)$ is here written in the exponent; sometimes it is brought outside as a multiplicative factor in an obviously equivalent representation. Regarded as a function of x for fixed θ, this is a density, usually with respect to Lebesgue or counting measure. The factor $h(x)$ can be absorbed in the density. The function $\kappa(\theta)$ is the log of the normalizing constant. Note that any exponential model has many exponential representations, for example $s = a + Bt$ and $\gamma = c + B^{-1}\theta$ also work; further the vectors θ and t can be "padded out" with extra elements. The smallest integer k for which the model has the exponential representation with θ and t having dimension k is the order of the model, and the representation (not unique) is minimal. In a minimal representation, t is a minimal sufficient statistic.

Let $\Theta = \{\theta \mid \int \exp\{\theta \bullet t - h(x)\}d\mu < \infty\}$, a convex subset of R^k. To see convexity, write $\gamma(\theta) = \exp\{\kappa(\theta)\} = \int \exp\{\theta \bullet t - h(x)\}d\mu$. Then for θ and θ' in Θ and $0 < \lambda < 1$,

$$\gamma(\lambda\theta + (1-\lambda)\theta') = \int \exp\{\theta \bullet t - h(x)\}^{\lambda} \exp\{\theta' \bullet t - h(x)\}^{1-\lambda} d\mu$$

$$< \lambda \int \exp\{\theta \bullet t - h(x)\}\, d\mu + (1-l)\int \exp\{\theta' \bullet t - h(x)\}\, d\mu$$

$$= \lambda\, \gamma(\theta) + (1-\lambda)\, \gamma(\theta') \qquad < \infty$$

The full exponential model is the model with parameter space Θ and model function (wrt the measure μ)

$$p(x;\theta) = \exp\{\theta \bullet t(x) - \kappa(\theta) - h(x)\}$$

where $\kappa(\theta) = \ln\int \exp\{\theta \bullet t - h(x)\}d\mu$.

If Θ is open, the model is regular. In many cases we will consider the restricted parameter space with $\theta = \theta(\omega)$, ω in Ω, a subset of R^d, with $d<k$. The model is then a (k,d) exponential model.

Note that the moment generating function

$$M(s) = E\exp\{t \bullet s\} = \int \exp\{t \bullet (\theta+s) - \kappa(\theta+s) + (\kappa(\theta+s) - \kappa(\theta)) - h(x)\} d\mu$$

$$= \exp\{\kappa(\theta+s) - \kappa(\theta)\}$$

and hence the cumulant generating function $K(s) = \ln M(s) = \kappa(\theta+s) - \kappa(\theta)$. Thus the m^{th} cumulant is given by

$$\kappa_{i1\ldots im}(\theta) = \partial^m \kappa(\theta)/\partial\theta^{i1}\ldots\partial\theta^{im}$$

$$= K s_{i1},\ldots s_{im}(s)\,|_{s=0}$$

A useful reparametrization is given by the mean value mapping $\tau(\theta) = Et$ defined on Θ (precisely, on intΘ). Note that $\tau(\theta) = D_\theta \kappa(\theta)$. Let C be the closed convex hull of the support of t. Then $\tau(\Theta) \subset \text{int}C$. The model is called steep if $|\tau(\theta)| \to \infty$ as θ goes to a boundary point of closΘ. Regular models are steep. A core exponential model is steep and full.

Define the Legendre transform of a real differentiable function f on $U \subset R^k$ by

$$f^*(x) = x \bullet \partial f/\partial x - f(x).$$

When f is regular on U, $f^{**} = f$. This transformation is useful for going back and forth between the canonical and mean-value parametrizations.

For a core exponential model,

1) $\tau(\Theta) = \text{int}C$,
2) $\kappa(\theta)$ is strictly convex.
3) κ^* is strictly convex and satisfies $\kappa^*(t) = \sup_\theta\{\theta \bullet t - \kappa(\theta)\}$, the maximized value of the log likelihood function at the statistic t.

4) κ and κ^* are smooth and $\partial\kappa/\partial\theta = \tau$, $\partial\kappa^*/\partial\tau = \theta$,

$\partial^2\kappa/\partial\theta\partial\theta' = \Sigma$,

$\partial^2\kappa^*/\partial\tau\partial\tau' = \Sigma^{-1}$

where $\Sigma = \mathrm{Var}(t)$ and $\tau = \tau(\theta)$.

5) The MLE exists iff it is in intC.

Then θ^, the MLE, is the unique solution to Et = T, where T is the realized value of the statistic t and E is E_θ.

Exponential families have been widely studied and many popular models fall into this category. Any statistical model admitting a fixed dimension sufficient statistic is an exponential model. This property is extremely useful and can be extended to the claim that models admitting "approximately sufficient" statistics are "approximately exponential." Finally, note that observed and expected information are the same and are in fact $\partial^2\kappa/\partial\theta\partial\theta'$.

Example 1: The exponential distribution $p(x,\gamma) = \gamma^{-1}\exp\{-\gamma x\}$ is an exponential family model. Write $p(x,\gamma) = \exp\{-\gamma x - \ln\gamma\}$; then $\theta = -\gamma$, $t = x$ (or Σx) and $\kappa(\theta) = \ln(-\theta)$. Hence the cumulant sequence is $-\theta^{-1}$, θ^{-2}, $-2\theta^{-3}$,etc.

Example 2: The normal distribution $p(x,\mu,\sigma) = (2\pi\sigma^2)^{-1/2}\exp\{-(x-\mu)^2/2\sigma^2\}$. Write this as $p(x,\mu,\sigma) = \exp\{-1/2\ln(2\pi\sigma^2) - x^2/2\sigma^2 + x\mu/2\sigma^2 - \mu/2\sigma^2\}$

$= \exp\{\theta\bullet t(x) - \kappa(\theta)\}$

with $\theta = (\mu/2\sigma^2,-1/2\sigma^2)$ and $t = (x, x^2)$ and $\kappa(\theta) = \theta_1^2/\theta_2 + 1/2\ln(-\theta_2) - c$. The mean value parametrization has $\tau_1 = -\theta_1/\theta_2$ and $\tau_2 = (\theta_1/\theta_2)^2 - 1/(2\theta_2)$

Example 3: Binomial (n, p). $p(x|p) = (n!/(x!(n-x)!)p^x(1-p)^{(1-x)}$

Example 4: The Gamma distribution with scale parameter 1, $p(x|\alpha) = \Gamma(\alpha)^{-1}x^{\alpha-1}e^{-x}$.

Example 5: The Inverse Normal $p(x|\mu,\lambda) = (\lambda/2\pi)^{1/2}x^{-3/2}\exp\{-\lambda(x-\mu)^2/(2\mu^2x)\}$

(reparametrize to α,λ where $\alpha = \lambda/\mu^2$.)

Curved Exponential Families

Example: Normal with variance equal to mean-squared.

Approximation by CLT – asymptotic distributions

Example: The binomial parameter estimator p^ has an asymptotic normal distribution with mean p and variance p(1-p)/n.