Exponential Families

This handout expands on section 3.4 in (Casella and Berger, 2002) CB who define a **family** of pdf's or pmf's $\{f(x|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\ \}$ to be an exponential family if

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp[\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)] \tag{1}$$

for $x \in \mathcal{X}$ where $c(\boldsymbol{\theta}) \geq 0$, $h(x) \geq 0$ **does not depend on** $\boldsymbol{\theta}$ and $t_i(x) : \mathcal{X} \to \Re$ **does not depend on** $\boldsymbol{\theta}$**.** (The value x may be a scalar or a vector.)

The family given by (1) is a k-**parameter exponential family** if k is the smallest integer where (1) holds. (If the t_i and w_i are linearly independent, then k is as small as possible.)

Because of the exponential function, many other parameterizations are possible. If $h(x) = g(x)I_{\mathcal{X}}(x)$, then usually $c(\boldsymbol{\theta})$ and $g(x)$ are positive, so another parameterization is $f(x|\boldsymbol{\theta}) = \exp[\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x) + d(\boldsymbol{\theta}) + S(x)]I_{\mathcal{X}}(x)$ where $S(x) = \log(g(x))$, $d(\boldsymbol{\theta}) = \log(c(\boldsymbol{\theta}))$, and the support $\mathcal{X}$ does not depend on $\boldsymbol{\theta}$.

The parameterization that uses the **natural parameter** $\boldsymbol{\eta}$ is especially useful for theory. The **natural parameterization for an exponential family** is

$$f(x|\boldsymbol{\eta}) = h(x)c^*(\boldsymbol{\eta}) \exp[\sum_{i=1}^{k} \eta_i t_i(x)] \tag{2}$$

where $h(x)$ and $t_i(x)$ are the same as in equation (1) and $\boldsymbol{\eta} \in \Omega$. Notice that this parameterization is not unique. If $b > 0$, then $b\eta_i$ and $t_i(x)/b$ would also work.

Let $\tilde{\Omega}$ be the set where the integral of the kernel function is finite:

$$\tilde{\Omega} = \{\boldsymbol{\eta} = (\eta_1, ..., \eta_k) : \frac{1}{c^*(\boldsymbol{\eta})} \equiv \int_{-\infty}^{\infty} h(x) \exp[\sum_{i=1}^{k} \eta_i t_i(x)]dx < \infty\}.$$

(Replace the integral by a sum for a pmf.) An interesting fact is that $\tilde{\Omega}$ is a convex set.

The next important idea is that of a **regular exponential family** (and of a full exponential family).

Condition E1: the natural parameter space $\Omega = \tilde{\Omega}$.
Condition E2: assume that in the natural parameterization, neither the η_i nor the t_i satisfy a linearity constraint.
Condition E3: Ω is a k-dimensional open set.
If conditions E1), E2) and E3) hold then the family is a **regular exponential family** (REF). If conditions E1) and E2) hold then the family is *full.* For a one parameter exponential family, a one dimensional rectangle is just an interval, and the only type of function of one variable that satisfies a linearity constraint is a constant function. Notice that every REF is full.

Some care has to be taken with the definitions of Θ and Ω since formulas (1) and (2) need to to hold for every $\boldsymbol{\theta} \in \Theta$ and for every $\boldsymbol{\eta} \in \Omega$. For a continuous random variable or vector, the pdf needs to exist. Hence all degenerate distributions need to be deleted from Θ and Ω. For continuous and discrete distributions, the natural parameter needs to exist (and often does not exist for discrete degenerate distributions). As a rule of thumb, remove values from Θ that cause the pmf to have the form 0^0. For example, for the binomial(n,p) distribution with n known, the natural parameter $\eta = \log(p/(1-p))$. Hence instead of using $\Theta = [0,1]$, use $p \in \Theta = (0,1)$, so that $\eta \in \Omega = (-\infty, \infty)$.

For the χ^2_p distribution, the parameter space $\Theta = \{1, 2, ...\}$, ie the set of positive integers. This distribution is the gamma$(\alpha = p/2, \beta = 2)$ distribution. Hence the natural parameter $\eta = p$ for $p > 0$ and $\Omega = (0, \infty)$. This is the only distribution in Casella and Berger where this type of problem occurs.

Sometimes a restriction is placed on the parameter space Ω resulting in a new parameter space Ω_C where Ω_C does not contain a k-dimensional rectangle. For example, the $N(\theta, \theta^2)$ distribution is a 2-parameter exponential family with $\eta_1 = -1/(2\theta^2)$ and $\eta_2 = 1/\theta$. Thus $\Omega_C = \{(\eta_1, \eta_2)|, -\infty < \eta_1 < 0, -\infty < \eta_2 < \infty, \eta_2 \neq 0\}$. The graph of this parameter space is a quadratic and cannot contain a 2-dimensional rectangle and $\dim(\Theta) = 1 < 2 = \dim(\Omega_C)$.

The beta, binomial, gamma, geometric, multinomial, multivariate normal, normal, and Poisson distributions are regular exponential families (after using suitable care in defining the parameter space and for the multinomial in defining $\mathcal{X}$). The inverse Gaussian distribution may be the only named distribution that is a full but not regular exponential family.

Usually an exponential family has **three important parameterizations**. The parameterization with $\boldsymbol{\theta}$ is how the family is usually written. The parameterization (1) is used to demonstrate that the family is an exponential family and the natural parameterization (2) is used to show that the family is a REF (or full) so that nice theorems hold.

Example. Let $f(x|\mu, \sigma)$ be the $N(\mu, \sigma^2)$ family of pdf's. Then $\boldsymbol{\theta} = (\mu, \sigma)$ where $-\infty < \mu < \infty$ and $\sigma > 0$. Recall that μ is the mean and σ is the SD of the distribution. CB p. 113-114 shows that this family is a 2-parameter exponential family with $w_1(\boldsymbol{\theta}) = 1/\sigma^2$ if $\sigma > 0$ and $w_2(\boldsymbol{\theta}) = \mu/\sigma^2$ if $\sigma > 0$. Hence $\eta_1 = 1/\sigma^2$ and $\eta_2 = \mu/\sigma^2$ if $\sigma > 0$. Plotting η_1 on the horizontal axis and η_2 on the vertical axis yields the right half plane which certainly contains a 2-dimensional rectangle. Hence the family is a REF.

The one-parameter exponential family occurs often enough to be of interest. A one-dimensional rectangle is just a nonempty interval and the family has the form $f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp[w(\boldsymbol{\theta})t(x)]$. Special cases include the binomial(n,p), Poisson(λ), negative binomial(r,p) with r known, the geometric(p), the $N(\mu, \sigma^2)$ with μ known, the $N(\mu, \sigma^2)$ with σ known, the exponential(λ), the gamma(α, β) with α known, and the gamma(α, β) with β known families.

In the definition of the exponential family is almost unchanged if $\boldsymbol{X}$ is a random vector instead of a random variable. Just replace the scalar x by the vector $\boldsymbol{x}$.

If the t_i or η_i satisfy a linearity constraint, then the number of terms in the exponent of equation (1) can be reduced. As an example, suppose that $X_1, ..., X_k$ follow the multinomial$_k(m, p_1, ..., p_k)$ distribution. The $\sum X_i = m$ and $\sum p_i = 1$. Hence

$$f(x_1, ..., x_k) = m! \prod_{i=1}^{k} \frac{p_i^{x_i}}{x!} = \exp[m \log(p_k) + x_1 \log(p_1/p_k) + ... + x_{k-1} \log(p_{k-1}/p_k] h(\boldsymbol{x})$$

which is a $k-1$ dimensional exponential family.

Result 1: CB, p. 217. If $X_1, ..., X_n$ are iid where X_1 is from an exponential family with pdf or pmf $f(x|\boldsymbol{\theta})$, then the joint pdf or pmf is given by

$$f(x_1, ..., x_n|\boldsymbol{\theta}) = (\prod_{j=1}^{n} h(x_j))[c(\boldsymbol{\theta})]^n \exp[w_1(\boldsymbol{\theta}) \sum_{j=1}^{n} t_1(x_j) + ... + w_k(\boldsymbol{\theta}) \sum_{j=1}^{n} t_k(x_j)]$$

which is a k-parameter exponential family and $\boldsymbol{t} = (\sum_{j=1}^n t_1(X_j), ..., \sum_{j=1}^n t_k(X_j))$ is a sufficient statistic for $(X_1, ..., X_n)$. Also $\boldsymbol{t}$ has an exponential family distribution.

Result 2 (see CB, p. 288). This sufficient statistic is a complete minimal sufficient if X_1 is from a REF.

Result 3: CB 1st edition, p. 125. For an exponential family, derivatives of all orders may be exchanged

$$\frac{\partial^j}{\partial \eta_i^j} \int_{-\infty}^{\infty} f(x|\boldsymbol{\eta}) dx = \int_{-\infty}^{\infty} \frac{\partial^j}{\partial \eta_i^j} f(x|\boldsymbol{\eta}) dx.$$

Result 4: CB, p. 341-342. CRLB is basically achieved only if the family is a 1 parameter REF.

Result 5: Schervish, 1995, p. 418. The MLE's have a Gaussian limiting distribution if X_1 is from a REF.

Result 6: Lehmann, 1986, p. 80. Under weak regularity conditions, the existence of UMP tests against one sided alternatives for all sample sizes and one α implies that the underlying distribution of the data is from an exponential family.

Result 7: CB p. 406, problem 8.27. A one parameter exponential family satisfies the MLR property.

Result 8: Barndorff–Nielsen, 1982. Suppose that the natural parameterization of the k-parameter exponential family is used so that Ω is a k-dimensional convex set (usually an open interval or cross product of open intervals). Then the log likelihood function log $L(\boldsymbol{\eta})$ is a strictly concave function of $\boldsymbol{\eta}$. Hence if $\hat{\boldsymbol{\eta}}$ is a critical point of log L($\boldsymbol{\eta}$) and if $\hat{\boldsymbol{\eta}} \in \Omega$ then $\hat{\boldsymbol{\eta}}$ is the unique MLE of $\boldsymbol{\eta}$. (The Hessian matrix of 2nd derivatives does not need to be checked.)

Note: with discrete distributions, there is a positive probability that $\hat{\boldsymbol{\eta}}$ is not in Ω. In this case the MLE does not exist. If $\boldsymbol{t}$ is the complete sufficient statistic and C is the closed convex hull of the support of $\boldsymbol{t}$, then the MLE exists iff $\boldsymbol{t} \in int\ C$ where $int\ C$ is the interior of C. An example is the Poisson distribution. The MLE does not exist if $\sum_{i=1}^n X_i = 0$.

A **curved exponential family** is a k-parameter exponential family where the dimension of the vector θ is $d < k$. The following example is a 2-parameter exponential family with $d = 1$.

The following example is taken from Cox and Hinckley (1974, p. 31) and illustrates why Ω should be used instead of Θ. Let $X_1, ..., X_n$ be iid $N(\mu, \gamma_o^2 \mu^2)$ random variables where $\gamma_o^2 > 0$ is **known** and $\mu > 0$. Then $f(x; \mu)$ is a two parameter exponential family with $\Theta = (0, \infty)$ (which contains a one dimensional rectangle), and $(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ is a minimal sufficient statistic. However

$$E_\mu[\frac{n + \gamma_o^2}{1 + \gamma_o^2} \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2] = 0$$

for all μ so the minimal sufficient statistic is not complete.

REFERENCES

Barndorff-Nielsen, O. (1982), "Exponential Families," in *Encylopedia of Statistical Sciences,* eds. Kotz, S. and Johnson, N.L., John Wiley and Sons, NY, 587-596.

Casella, G., and Berger, R.L. (2002), *Statistical Inference,* 2nd ed. (1990, 1st ed.) , Duxbury Press, Belmont, CA.

Cox, D.R., and Hinckley, D.V. (1974), *Theoretical Statistics,* Chapman and Hall, London.

Lehmann, E.L. (1986), *Testing Statistical Hypotheses,* John Wiley and Sons, NY.

Olive, D.J. (2004), *Using Exponential Families in an Inference Course*, unpublishable document available from (http://www.math.siu.edu/olive/infer.htm).

Schervish, M.J. (1995), *Theory of Statistics,* Springer-Verlag, NY.