Using Exponential Families in an Inference Course

David J. Olive[*]

Southern Illinois University

March 29, 2008

Abstract

Many statistics and math departments offer a one semester course in statistical inference that covers minimal and complete sufficient statistics, maximum likelihood estimators, uniform minimum variance estimators, uniformly most powerful tests and large sample theory. Using the theory of exponential families can greatly simplify the teaching of these and other topics, and the goal of this paper is to present some of this theory in a manner that is accessible to graduate students.

KEY WORDS: MLE's, Sufficient Statistics, UMVUE's, UMP Tests

[*] David J. Olive is Associate Professor, Department of Mathematics, Mailcode 4408, Southern Illinois University, Carbondale, IL 62901-4408, USA. Email: dolive@math.siu.edu

1 INTRODUCTION

A one semester Master's level course in statistical inference typically covers minimal and complete sufficient statistics, maximum likelihood estimators (MLE), uniform minimum variance estimators (UMVUE) and the Fréchet Cramér Rao lower bound (FCRLB), uniformly most powerful (UMP) tests and large sample theory. Such topics can be greatly simplified by using the theory of exponential families, but the texts for this type of course tend to devote only a few pages to these families. The goal of this paper is to present some of this theory in a manner that is accessible to students.

Often a "brand name distribution" such as the normal distribution will have three useful parameterizations: the *usual parameterization* with parameter space Θ_U is simply the formula for the probability distribution or mass function (pdf or pmf, respectively) given when the distribution is first defined. The *k-parameter exponential family parameterization* with parameter space Θ, given in Equation (1.1) below, provides a simple way to determine if the distribution is an exponential family while the *natural parameterization* with parameter space Ω, given in Equation (1.2) below, is used for *theory* that requires a complete sufficient statistic.

A *family* of pdf's or pmf's $\{f(x|\boldsymbol{\theta}) : \boldsymbol{\theta} = (\theta_1, ..., \theta_j) \in \Theta \}$ is an exponential family if

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp[\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)] \tag{1.1}$$

for $x \in \mathcal{X}$ where $c(\boldsymbol{\theta}) \geq 0$, and $h(x) \geq 0$. The functions c, h, t_i, and w_i are real valued functions. The parameter $\boldsymbol{\theta}$ can be a scalar, and x can be vector valued. In the definition, it is crucial that $c, w_1, ..., w_k$ do not depend on x and that $h, t_1, ..., t_k$ do not depend on

$\boldsymbol{\theta}$. The support of the distribution is $\mathcal{X}$ and the parameter space is Θ. The family given is a **k-parameter exponential family** if k is the smallest integer where (1.1) holds.

The parameterization is not unique since, for example, w_i could be multiplied by a nonzero constant b if t_i is divided by b. 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 $\mathcal{X}$ does not depend on $\boldsymbol{\theta}$. Thus the uniform$(0,\theta)$ family is not an exponential family since the support $(0, \theta)$ depends on θ.

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

$$f(x|\boldsymbol{\eta}) = h(x)c^*(\boldsymbol{\eta}) \exp[\sum_{i=1}^k \eta_i t_i(x)] \tag{1.2}$$

where $h(x)$ and $t_i(x)$ are the same as in Equation (1.1) and $\boldsymbol{\eta} \in \Omega$. Again, the parameterization is not unique. If $b \neq 0$, then $b\eta_i$ and $t_i(x)/b$ would also work.

The next important idea is that of a regular exponential family (and of a full exponential family). Let $d_i(y)$ denote $t_i(x)$, $w_i(\boldsymbol{\theta})$ or η_i. A *linearity constraint* is satisfied by $d_1(y), ..., d_k(y)$ if $\sum_{i=1}^k a_i d_i(y) = c$ for some constants a_i and c and for all y in the sample or parameter space where not all of the $a_i = 0$. If $\sum_{i=1}^k a_i d_i(y) = c$ for all y only if $a_1 = \cdots = a_k = 0$, then the $d_i(y)$ do not satisfy a linearity constraint. See Johanson (1979, p. 3). In linear algebra, we would say that the $d_i(y)$ are *linearly independent* if they do not satisfy a linearity constraint. 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.

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. For a one parameter exponential family, the open set is usually an open interval, and the only type of function of one variable that satisfies a linearity constraint is a constant function.

Some care has to be taken with the definitions of Θ and Ω since formulas (1.1) and (1.2) need 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 $\eta \in \Omega = (-\infty, \infty)$.

These conditions have some redundancy. If Ω contains a k-dimensional rectangle, no η_i is completely determined by the remaining $\eta_j's$. In particular, the η_i cannot satisfy a linearity constraint. If the η_i do satisfy a linearity constraint, then the η_i lie on a hyperplane of dimension at most k, and such a surface cannot contain a k-dimensional rectangle. For example, if $k = 2$, a line cannot contain an open box. If $k = 2$ and $\eta_2 = \eta_1^2$, then the parameter space does not contain a 2-dimensional rectangle, although η_1 and η_2 do not satisfy a linearity constraint.

Example 1. 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 standard deviation (SD) of the distribution. The usual parameterization is

$$f(x|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-(x-\mu)^2}{2\sigma^2}) I_{\Re}(x)$$

where $\Re = (-\infty, \infty)$ and the indicator $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ otherwise. Since

$$f(x|\mu, \sigma) = \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-\mu^2}{2\sigma^2})}_{c(\mu,\sigma)\geq 0} \exp(\underbrace{\frac{-1}{2\sigma^2}}_{w_1(\boldsymbol{\theta})} \underbrace{x^2}_{t_1(x)} + \underbrace{\frac{\mu}{\sigma^2}}_{w_2(\boldsymbol{\theta})} \underbrace{x}_{t_2(x)}) \underbrace{I_{\Re}(x)}_{h(x)\geq 0},$$

this family is a 2-parameter exponential family. Hence $\eta_1 = -0.5/\sigma^2$ 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 left half plane which certainly contains a 2-dimensional rectangle. Since t_1 and t_2 lie on a quadratic rather than a line, the family is a REF. Notice that if $X_1, ..., X_n$ are iid $N(\mu, \sigma^2)$ random variables, then the joint pdf

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = f(x_1, ..., x_n|\mu, \sigma) = \underbrace{[\frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-\mu^2}{2\sigma^2})]^n}_{C(\mu,\sigma)\geq 0} \exp(\underbrace{\frac{-1}{2\sigma^2}}_{w_1(\boldsymbol{\theta})} \underbrace{\sum_{i=1}^n x_i^2}_{T_1(\boldsymbol{x})} + \underbrace{\frac{\mu}{\sigma^2}}_{w_2(\boldsymbol{\theta})} \underbrace{\sum_{i=1}^n x_i}_{T_2(\boldsymbol{x})}) \underbrace{1}_{h(\boldsymbol{x})\geq 0},$$

and is thus a 2-parameter REF.

Some one parameter REF's include the $N(\mu, \sigma^2)$ family with either μ or $\sigma^2 > 0$ known, the gamma(α, β) family with either $\alpha > 0$ or $\beta > 0$ known, the beta(α, β) family with either $\alpha > 0$ or $\beta > 0$ known, the exponential family, the Poisson family, and the Rayleigh family. See Casella and Berger (2002, p. 131). The binomial(n, p) family with n known and $0 < p < 1$ is a REF as is the negative binomial family with r known and $0 < p < 1$. The geometric(p) and Pareto(α, β) families with α known are also one parameter REF's. The gamma(α, β) and beta(α, β) families are 2-parameter REF's. The inverse Gaussian distribution is full but not regular. The two parameter Cauchy distribution is not an exponential family because its pdf cannot be put into the form of Equation (1.1).

If the t_i or η_i satisfy a linearity constraint, then the number of terms in the exponent of Equation (1.1) can be reduced. As an example, suppose that $X_1, ..., X_j$ follow the multinomial$_j(n, p_1, ..., p_j)$ distribution which has $\dim(\Theta) = j$. Then $\sum_{i=1}^j X_i = n$ and $\sum_{i=1}^j p_i = 1$. Let $h(\boldsymbol{x}) = n!/(\prod_{i=1}^j x_i!)$. Then

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

which is a $j-1$ dimensional REF. See Lehmann (1983, p. 28). Similarly, let $\boldsymbol{\mu}$ be a $1 \times j$ row vector and let $\boldsymbol{\Sigma}$ be a $j \times j$ positive definite matrix. Then the usual parameterization of the multivariate normal MVN$_j(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution has $\dim(\Theta) = j + j^2$ but is a $j + j(j+1)/2$ parameter REF. These are important examples of REF's where $\dim(\Theta) > \dim(\Omega)$.

The natural parameterization can result in a family that is much larger than the family defined by the usual parameterization. See the definition of $\Omega = \tilde{\Omega}$ on p. 4.

Casella and Berger (2002, p. 114) remarks that

$$\{\boldsymbol{\eta} : \boldsymbol{\eta} = (w_1(\boldsymbol{\theta}), ..., w_k(\boldsymbol{\theta}))|\boldsymbol{\theta} \in \Theta\} \subseteq \Omega,$$

but often Ω is a strictly larger set. An example is the χ^2_p distribution. This distribution is not a REF since the set of integers is not a convex subset of the real line. Nevertheless, the natural parameterization is the gamma$(\alpha, \beta = 2)$ family which is a REF. Note that this family has uncountably many members while the χ^2_p family does not.

It may be a good idea to inform students that they will only be presented problems where they can simply set $\eta_i = w_i(\boldsymbol{\theta})$ for families other than the χ^2_p distribution. Assume that $\dim(\Theta) = k = \dim(\Omega)$. Assume that in the usual parameterization Θ_U is as big as possible (replace the integral by a sum for a pmf):

$$\Theta_U = \{\boldsymbol{\theta} \in \Re^k : \int f(x|\boldsymbol{\theta})dx = 1\},$$

and let

$$\Theta = \{\boldsymbol{\theta} \in \Theta_U : w_1(\boldsymbol{\theta}), ..., w_k(\boldsymbol{\theta}) \text{ are defined }\}.$$

Then assume that the natural parameter space

$$\Omega = \{(\eta_1, ..., \eta_k) : \eta_i = w_i(\boldsymbol{\theta}) \text{ for } \boldsymbol{\theta} \in \Theta\}.$$

In other words, simply define $\eta_i = w_i(\boldsymbol{\theta})$. For many common distributions, $\boldsymbol{\eta}$ is a one to one function of $\boldsymbol{\theta}$, and the above map is correct.

Example 2. The binomial(n, p) pmf is

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x} = \underbrace{\binom{n}{x}}_{h(x)\geq 0} \underbrace{(1-p)^n}_{c(p)\geq 0} \exp[\underbrace{\log(\frac{p}{1-p})}_{w(p)} \underbrace{x}_{t(x)}]$$

for $x = 0, 1, \ldots, n$ where $\Theta_U = [0, 1]$. Since $\eta = \log(p/(1-p))$ is undefined for $p = 0$ and $p = 1$, $\Theta = (0, 1)$.

A **curved exponential family** is a k-parameter exponential family where the elements of $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$ are completely determined by $d < k$ of the elements. This family is neither full nor regular since it places a restriction on the parameter space Ω resulting in a new parameter space $\tilde{\Omega}$ where $\tilde{\Omega}$ 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 $\tilde{\Omega} = \{(\eta_1, \eta_2)|, -\infty < \eta_1 < 0, -\infty < \eta_2 < \infty, \eta_2 \neq 0\}$. The graph of this parameter space is a quadratic and cannot contain a 2-dimensional rectangle.

The following sections show that exponential families can be used to simplify the theory of sufficiency, MLE's, UMVUE's, and UMP tests.

2 Exponential Families and Sufficiency.

Finding sufficient, minimal sufficient, and complete sufficient statistics is often simple for REF's. A statistic $\boldsymbol{T}(X_1, ..., X_n)$ is a *sufficient statistic* for $\boldsymbol{\theta}$ if the conditional distribution of $(X_1, ..., X_n)$ given $\boldsymbol{T}$ does not depend on $\boldsymbol{\theta}$. A sufficient statistic $\boldsymbol{T}(\boldsymbol{X})$ is a *minimal sufficient statistic* if for any other sufficient statistic $\boldsymbol{T}'(\boldsymbol{X})$, $\boldsymbol{T}(\boldsymbol{X})$ is a function of $\boldsymbol{T}'(\boldsymbol{X})$. Suppose that a *statistic* $\boldsymbol{T}(\boldsymbol{X})$ has a pmf or pdf $f(\boldsymbol{t}|\boldsymbol{\theta})$. Then $\boldsymbol{T}(\boldsymbol{X})$ is a *complete statistic* if $E_{\boldsymbol{\theta}}[g(\boldsymbol{T})] = 0$ for all $\boldsymbol{\theta}$ implies that $P_{\boldsymbol{\theta}}[g(\boldsymbol{T}(\boldsymbol{X})) = 0] = 1$ for all $\boldsymbol{\theta}$.

There are several important facts concerning such statistics. First, a one to one function of a sufficient, minimal sufficient, or complete sufficient statistic is sufficient, minimal sufficient, or complete sufficient respectively. *Bahadur's Theorem* states that a finite dimensional complete sufficient statistic is also minimal sufficient (Bahadur 1958 and Lehmann and Scheffé 1950). Assume that a *sample* $\boldsymbol{X} = (X_1, ..., X_n)$ consists on n independent random variables. Three very important theorems for sufficient statistics follow.

Factorization Theorem: Let $f(\boldsymbol{x}|\boldsymbol{\theta})$ denote the pdf or pmf of a sample $\boldsymbol{X}$. A statistic $\boldsymbol{T}(\boldsymbol{X})$ is a sufficient statistic for $\boldsymbol{\theta}$ if for all sample points $\boldsymbol{x}$ and for all parameter points $\boldsymbol{\theta}$,

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = g(\boldsymbol{T}(\boldsymbol{x})|\boldsymbol{\theta})h(\boldsymbol{x})$$

where both g and h are nonnegative functions.

Lehmann-Scheffé Theorem for Minimal Sufficient Statistics: Let $f(\boldsymbol{x}|\boldsymbol{\theta})$ be the pmf or pdf of a sample $\boldsymbol{X}$. Let $c_{\boldsymbol{x},\boldsymbol{y}}$ be a constant. Suppose there exists a function $\boldsymbol{T}(\boldsymbol{x})$ such that for any two sample points $\boldsymbol{x}$ and $\boldsymbol{y}$, the ratio $R_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{\theta}) = f(\boldsymbol{x}|\boldsymbol{\theta})/f(\boldsymbol{y}|\boldsymbol{\theta}) = c_{\boldsymbol{x},\boldsymbol{y}}$ for all $\boldsymbol{\theta}$ in Θ iff $\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{T}(\boldsymbol{y})$. Then $\boldsymbol{T}(\boldsymbol{X})$ is a minimal sufficient statistic for $\boldsymbol{\theta}$.

Sufficiency, Minimal Sufficiency, and Completeness of Exponential Families: Suppose that $X_1, ..., X_n$ are iid from an exponential family with the natural parameterization given by Equation (1.2) so that the joint pdf or pmf is given by

$$f(x_1, ..., x_n|\boldsymbol{\eta}) = (\prod_{j=1}^{n} h(x_j))[c^*(\boldsymbol{\eta})]^n \exp[\eta_1 \sum_{j=1}^{n} t_1(x_j) + \cdots + \eta_k \sum_{j=1}^{n} t_k(x_j)]$$

which is a k-parameter exponential family. Then $\boldsymbol{T}(\boldsymbol{X}) = (\sum_{j=1}^n t_1(X_j), ..., \sum_{j=1}^n t_k(X_j))$is

a) a sufficient statistic for $\boldsymbol{\eta}$,

b) a minimal sufficient statistic for $\boldsymbol{\eta}$ if $\eta_1, ..., \eta_k$ do not satisfy a linearity constraint,

c) a complete sufficient statistic for $\boldsymbol{\eta}$ if Ω contains a k-dimensional rectangle.

Proof: a) Use the factorization theorem.

b) The ratio

$$\frac{f(\boldsymbol{x}|\boldsymbol{\eta})}{f(\boldsymbol{y}|\boldsymbol{\eta})} = \frac{\prod_{j=1}^n h(x_j)}{\prod_{j=1}^n h(y_j)} \exp[\sum_{i=1}^k \eta_i (T_i(\boldsymbol{x}) - T_i(\boldsymbol{y}))]$$

is equal to a constant with respect to $\boldsymbol{\eta}$ iff

$$\sum_{i=1}^k \eta_i [T_i(\boldsymbol{x}) - T_i(\boldsymbol{y})] = \sum_{i=1}^k \eta_i a_i = d$$

for all η_i where d is some constant and where $a_i = T_i(\boldsymbol{x}) - T_i(\boldsymbol{y})$ and $T_i(\boldsymbol{x}) = \sum_{j=1}^n t_i(x_j)$. Since the η_i do not satisfy a linearity constraint, $\sum_{i=1}^k \eta_i a_i = d$ iff all of the $a_i = 0$. Hence

$$\boldsymbol{T}(\boldsymbol{X}) = (T_1(\boldsymbol{X}), ..., T_k(\boldsymbol{X}))$$

is a minimal sufficient statistic by the Lehmann-Scheffé theorem.

c) See Lehmann (1986, p. 142).

Remarks. In the Lehmann-Scheffé Theorem, for R to be constant as a function of $\boldsymbol{\theta}$, define $0/0 = c_{\boldsymbol{x},\boldsymbol{y}}$. In a), k does not need to be as small as possible. Presenting the various parameterizations of the exponential family immediately before presenting sufficiency will help prepare students for the factorization theorem. Typically $\eta_i = w_i(\boldsymbol{\theta})$, and $\boldsymbol{T}$ is also a sufficient, minimal sufficient, or complete sufficient statistic for $\boldsymbol{\theta}$ if the appropriate conditions from the above theorem hold. (For completeness, also check that $\boldsymbol{\eta}$ is a one to one map of $\boldsymbol{\theta}$.) The proof of part b) expands on remarks given in Johanson (1979, p. 3) and Lehmann (1983, p. 44) but is typically not given in inference texts. The theorem gives a particularly simple way to find complete sufficient statistics for one parameter exponential families and for any family that is known to be REF. If it is known that the distribution is regular, find the exponential family parameterization given by Equation (1.1) or (1.2). These parameterizations give $t_1(x), ..., t_k(x)$. Then

$\boldsymbol{T}(\boldsymbol{X}) = (\sum_{j=1}^n t_1(X_j), ..., \sum_{j=1}^n t_k(X_j))$. In Example 1, $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is a complete sufficient statistic for (μ, σ^2). The one to one functions $(\overline{X}, S^2)$ and $(\overline{X}, S)$ are also complete sufficient where $\overline{X}$ is the sample mean and S is the sample standard deviation. In Example 2, $\sum_{i=1}^n t(X_i) = \sum_{i=1}^n X_i$ is complete sufficient statistic for p. Other techniques for showing whether a statistic is minimal sufficient are illustrated in Sampson and Spencer (1976).

Example 3, Cox and Hinckley (1974, p. 31). 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. This example illustrates that the rectangle needs to be contained in Ω rather than Θ. As a rule of thumb, a k-parameter minimal sufficient statistic for a d-dimensional parameter where $d < k$ will not be complete. In this example $d = 1 < 2 = k$.

Example 4. The theory does not say that any sufficient statistic from a REF is complete. Let X be a random variable from a normal $N(0, \sigma^2)$ distribution with $\sigma^2 > 0$. This family is a REF with complete minimal sufficient statistic X^2. The data X is also a sufficient statistic, but X is not a function of X^2. Hence X is not minimal sufficient and (by Bahadur's theorem) not complete.

Example 5. In testing theory, a single sample is often created by combining two samples of iid data. Let $X_1, ..., X_n$ be iid exponential(β) and $Y_1, ..., Y_m$ iid exponential($\beta/2$).

If the two samples are independent, then the joint pdf $f(\boldsymbol{x}, \boldsymbol{y}|\beta)$ belongs to a regular one parameter exponential family with complete sufficient statistic $T = \sum_{i=1}^n X_i + 2\sum_{i=1}^m Y_i$.

3 Exponential Families and MLE's.

The following definitions are used in the theory of MLE's. Let $f(\boldsymbol{x}|\boldsymbol{\theta})$ be the pmf or pdf of a sample $\boldsymbol{X}$. If $\boldsymbol{X} = \boldsymbol{x}$ is observed, then the *likelihood function* $L(\boldsymbol{\theta}) = f(\boldsymbol{x}|\boldsymbol{\theta})$. For each sample point $\boldsymbol{x} = (x_1, ..., x_n)$, let $\hat{\boldsymbol{\theta}}(\boldsymbol{x}) \in \Theta$ be the parameter value at which $L(\boldsymbol{\theta}|\boldsymbol{x})$ attains its maximum as a function of $\boldsymbol{\theta}$ with $\boldsymbol{x}$ held fixed. Then the maximum likelihood estimator (**MLE**) of the parameter $\boldsymbol{\theta}$ based on the sample $\boldsymbol{X}$ is $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$.

Existence and Limiting Distribution of the MLE (Barndorff–Nielsen 1982): Suppose that the natural parameterization of the k-parameter REF is used so that Ω is an open 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!) The MLE's have a Gaussian limiting distribution if the family is a REF (also see Schervish, 1995, p. 418).

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 $\sum_{i=1}^n X_i = 0$ if $\Theta = (0, \infty)$.

Remarks: For 1–parameter exponential families, check that the critical point is a global maximum using standard calculus techniques such as calculating the second derivative of the log likelihood $\log L(\boldsymbol{\theta}|\boldsymbol{x})$. Casella and Berger (2002, p. 317) give a very useful result for a scalar valued parameter: if $K(\theta)$ is a continuous function defined on an interval with endpoints $a < b$ (not necessarily finite), differentiable on (a, b), and if the **critical point is unique**, then the critical point is a *global maximum* if it is a local maximum (since otherwise there would be a local minimum and then the critical point would not be unique). These techniques should be used since verifying that the family is regular is often more difficult than using calculus. Also, often the MLE is desired for a parameter space Θ_U which is not an open set (e.g. for $\Theta_U = [0, 1]$ instead of $\Theta = (0, 1)$). For k-parameter exponential families with $k > 1$, it is usually easier to verify that the family is regular than to calculate the Hessian matrix.

The **Invariance Principle** is also important: if $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$, then $h(\hat{\boldsymbol{\theta}})$ is the MLE of $h(\boldsymbol{\theta})$. Many texts refer to Zehna (1966) for a proof of the invariance principle, but a compelling alternative proof that uses a genuine likelihood (unlike Zehna's pseudo-likelihood) is given in Berk (1967).

Another useful fact is that if the MLE is unique, then the MLE is a function of the minimal sufficient statistic. See Levy (1985) and Moore (1971). This fact is useful since exponential families tend to have a tractable log likelihood and an easily found minimal sufficient statistic. In Example 5, the MLE for β is $\hat{\beta} = T/(n + m)$.

4 Exponential Families, UMVUE's and the FCRLB.

The following notation will be useful. Let $\tau(\boldsymbol{\theta})$ be a real valued function of $\boldsymbol{\theta}$, and let $W \equiv W(X_1, ..., X_n)$ be an estimator of $\tau(\boldsymbol{\theta})$. The *bias* of the estimator W for $\tau(\boldsymbol{\theta})$ is $\text{Bias}_W(\tau(\boldsymbol{\theta})) = E_{\boldsymbol{\theta}} W - \tau(\boldsymbol{\theta})$. The *mean squared error* (MSE) of an estimator W for $\tau(\boldsymbol{\theta})$ is

$$MSE_W(\tau(\boldsymbol{\theta})) = E_{\boldsymbol{\theta}}[(W - \tau(\boldsymbol{\theta}))^2] = Var_{\boldsymbol{\theta}}(W) + [Bias_W(\tau(\boldsymbol{\theta}))]^2.$$

W is an *unbiased estimator* of $\tau(\boldsymbol{\theta})$ if $E_\theta W = \tau(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$, and W_U is the uniformly minimum variance unbiased estimator (UMVUE) of $\tau(\boldsymbol{\theta})$ if W_U is an unbiased estimator of $\tau(\boldsymbol{\theta})$ and if $\text{Var}_{\boldsymbol{\theta}} W_U \leq Var_{\boldsymbol{\theta}} W$ for all $\boldsymbol{\theta}$ where W is any other unbiased estimator of $\tau(\boldsymbol{\theta})$. The following theorem is extremely useful since it is often easy to find a complete sufficient statistic for a one parameter REF.

Lehmann-Scheffé Theorem for UMVUE's: If $T(\boldsymbol{X})$ is a complete sufficient statistic for θ, then $g(T(\boldsymbol{X}))$ *is the UMVUE of its expectation.* In particular, if $S(\boldsymbol{X})$ is *any unbiased estimator* of $\tau(\theta)$, then $W_U \equiv E[S(\boldsymbol{X})|T(\boldsymbol{X})]$ is the UMVUE of $\tau(\theta)$.

The following facts can be useful for computing the conditional expectation (Rao-Blackwellization). Suppose $X_1, ..., X_n$ are iid with finite expectation.

a) Then $E[X_1 | \sum_{i=1}^n X_i = y] = y/n$.

b) If the X_i are iid Poisson(λ), then $(X_1 | \sum_{i=1}^n X_i = y) \sim$ binomial$(y, 1/n)$.

c) If the X_i are iid Bernoulli(p), then $(X_1 | \sum_{i=1}^n X_i = y) \sim$ Bernoulli(y/n).

d) If the X_i are iid N(μ, σ^2), then $(X_1 | \sum_{i=1}^n X_i = y) \sim N[y/n, \sigma^2(1 - 1/n)]$.

Often students will be asked to compute a lower bound on the variance of unbiased estimators when θ is a scalar. The *information number* or *Fisher Information* is

$$I_n(\theta) = E_\theta[(\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(X_i|\theta))^2].$$

If X comes from an exponential family, then

$$I_1(\theta) = E_\theta[(\frac{\partial}{\partial\theta}\log f(X|\theta))^2] = -E_\theta[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)].$$

If the derivative and integral operators can be interchanged, and if $X_1, ..., X_n$ are iid, then $I_n(\theta) = nI_1(\theta)$.

Lemma, Casella and Berger (1990, p. 312): If X comes from an exponential family, then the derivative and integral operators can be interchanged:

$$\frac{d}{d\theta}\int\dots\int g(\boldsymbol{x})f(\boldsymbol{x}|\theta)d\boldsymbol{x} = \int\dots\int g(\boldsymbol{x})\frac{\partial}{\partial\theta}f(\boldsymbol{x}|\theta)d\boldsymbol{x}$$

for any function $g(\boldsymbol{x})$ with $E_\theta|g(\boldsymbol{X})| < \infty$.

Fréchet Cramér Rao Lower Bound or Information Inequality: Let $X_1, ..., X_n$ be independent with joint pdf or pmf $f(\boldsymbol{x}|\theta)$ that satisfies equation (4.1). Let $W(X_1, ..., X_n)$ be any estimator of $\tau(\theta) \equiv E_\theta W(\boldsymbol{X})$. Then

$$Var_\theta W(\boldsymbol{X}) \geq \frac{[\frac{d}{d\theta}E_\theta W(\boldsymbol{X})]^2}{E_\theta[(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\theta))^2]} = \frac{[\tau'(\theta)]^2}{I_n(\theta)}.$$

The quantity $\dfrac{[\tau'(\theta)]^2}{I_n(\theta)}$ = FCRLB($\tau(\theta)$) is the **Fréchet Cramér Rao lower bound** (FCRLB) for the variance of unbiased estimators of $\tau(\theta)$.

Many inference tests suggest that a UMVUE can be found by determining whether an unbiased estimator W has a variance equal to the FCRLB. This method rarely works since typically equality holds only if

1) the data come from a one parameter REF with complete sufficient statistic T, and

2) $W = a + bT$ is a linear function of T.

The FCRLB inequality will typically be strict for nonlinear functions of T if the data is from a one parameter REF. If T is complete, $g(T)$ is the UMVUE of its expectation, and determining that T is the complete sufficient statistic from a one parameter REF is simpler than computing $Var_\theta W$ and FCRLB($\tau(\theta)$). If the family is not an exponential family, the FCRLB may **not be a lower bound** on the variance of unbiased estimators of $\tau(\theta)$. For a more precise statement of when the FCRLB is achieved and for some counterexamples, see Wijsman (1973) and Joshi (1976). Although the FCRLB is not very useful for finding UMVUE's, it is useful for finding the asymptotic variances of UMVUE's and MLE's. See Portnoy (1977). Karakostas (1985) has useful references for UMVUE's.

5 Exponential Families, the Neyman Pearson Lemma, and UMP tests.

The following concepts are useful in testing theory. The *power function* of a hypothesis test of H_o vs H_A is $\beta(\theta) = P_\theta(H_o \text{ is rejected})$ for $\theta \in \Theta$. Let $0 \le \alpha \le 1$. Then a test with power function $\beta(\theta)$ is a *level α test* if

$$\sup_{\theta \in \Theta_o} \beta(\theta) \le \alpha.$$

Consider all level α tests of $H_o : \theta \in \Theta_o$ vs $H_A : \theta \in \Theta_A$. A *uniformly most powerful* (UMP) level α test is a test with power function $\beta_{UMP}(\theta)$ such that $\beta_{UMP}(\theta) \ge \beta'(\theta)$ for every $\theta \in \Theta_A$ where β' is a power function for any level α test of H_o vs H_A. The following three theorems can be used to find UMP tests.

One Sided UMP Tests for Exponential Families (Mood, Graybill, and Boes 1974, p. 424 and Bickel and Doksum 1977, p. 199): Let $X_1, ..., X_n$ be a sample with a joint pdf or pmf from a one parameter exponential family where $w(\theta)$ is strictly increasing and $T(\boldsymbol{x})$ is the complete sufficient statistic. Alternatively, let $X_1, ..., X_n$ be iid with pdf or pmf

$$f(x|\theta) = h(x)c(\theta)\exp[w(\theta)t(x)]$$

from a one parameter exponential family where θ is real and $w(\theta)$ is strictly increasing. Here $T(\boldsymbol{x}) = \sum_{i=1}^n t(x_i)$. Then the UMP test for $H_o : \theta \leq \theta_o$ vs $H_A : \theta > \theta_o$ rejects H_o if $T(\boldsymbol{x}) > k$ and rejects H_o with probability γ if $T(\boldsymbol{x}) = k$ where $\alpha = P_{\theta_o}(T > k) + \gamma P_{\theta_o}(T = k)$. The UMP test for $H_o : \theta \geq \theta_o$ vs $H_A : \theta < \theta_o$ rejects H_o if $T(\boldsymbol{x}) < k$ and rejects H_o with probability γ if $T(\boldsymbol{x}) = k$ where $\alpha = P_{\theta_o}(T < k) + \gamma P_{\theta_o}(T = k)$.

Remarks: As a mnemonic, note that the *inequality used in the rejection region is the same as the inequality in the alternative hypothesis.* Usually $\gamma = 0$ if f is a pdf. Suppose that the parameterization is

$$f(x|\theta) = h(x)c(\theta)\exp[\tilde{w}(\theta)\tilde{t}(x)]$$

where $\tilde{w}(\theta)$ is strictly decreasing. Then set $w(\theta) = -\tilde{w}(\theta)$ and $t(x) = -\tilde{t}(x)$.

The Neyman Pearson Lemma: Consider testing $H_o : \theta = \theta_o$ vs $H_A : \theta = \theta_A$ where the pdf or pmf corresponding to θ_i is $f(\boldsymbol{x}|\theta_i)$ for $i = o, A$. Suppose the test rejects H_o if $f(\boldsymbol{x}|\theta_A) > kf(\boldsymbol{x}|\theta_o)$, and rejects H_o with probability γ if $f(\boldsymbol{x}|\theta_A) = kf(\boldsymbol{x}|\theta_o)$ for some $k \geq 0$. If

$$\alpha = \beta(\theta_o) = P_{\theta_o}[f(\boldsymbol{x}|\theta_A) > kf(\boldsymbol{x}|\theta_o)] + \gamma P_{\theta_o}[f(\boldsymbol{x}|\theta_A) = kf(\boldsymbol{x}|\theta_o)],$$

then this test is an UMP level α test.

One Sided UMP Tests via the Neyman Pearson Lemma: Suppose that the hypotheses are of the form $H_o : \theta \leq \theta_o$ vs $H_A : \theta > \theta_o$ or $H_o : \theta \geq \theta_o$ vs $H_A : \theta < \theta_o$, or that the inequality in H_o is replaced by equality. Also assume that

$$\sup_{\theta \in \Theta_o} \beta(\theta) = \beta(\theta_o).$$

Pick $\theta_A \in \Theta_A$ and use the Neyman Pearson lemma to find the UMP test for $K_o : \theta = \theta_o$ vs $K_A : \theta = \theta_A$. Then the UMP test rejects K_o if $f(\boldsymbol{x}|\theta_A) > kf(\boldsymbol{x}|\theta_o)$, and rejects K_o with probability γ if $f(\boldsymbol{x}|\theta_A) = kf(\boldsymbol{x}|\theta_o)$ for some $k \geq 0$ where $\alpha = \beta(\theta_o)$. This test is also the UMP level α test for $H_o : \theta \in \Theta_o$ vs $H_A : \theta \in \Theta_A$ if k *does not depend on the value of* θ_A.

The result for exponential families is simpler than using the Neyman Pearson lemma since the test statistic T will have a distribution from an exponential family. See Casella and Berger (2002, p. 217). This result often enables students to find the cutoff value k. To find a UMP test via the Neyman Pearson lemma, students need to check that the cutoff value k does not depend on $\theta_A \in \Theta_A$ and usually they need to transform the test statistic to put the test in *useful form.* With exponential families, the transformed test statistic is often T.

6 Limit Theorems For Exponential Families.

Barndorff–Nielsen (1982), Casella and Berger (2002, pp. 472, 515), Cox and Hinkley (1974, p. 286), Lehmann and Casella (1998, Section 6.3), Schervish (1995, p. 418), and many others suggest that under regularity conditions if $X_1, ..., X_n$ are iid from a one parameter regular exponential family, and if $\hat{\theta}$ is the MLE of θ, then

$$\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{D} N[0, FCRLB_1(\tau(\theta))] \tag{6.1}$$

where the Fréchet Cramér Rao lower bound based on a sample of size one for $\tau(\theta)$ is

$$FCRLB_1(\tau(\theta)) = \frac{[\tau'(\theta)]^2}{I_1(\theta)}$$

and the Fisher information based on a sample of size one is

$$I_1(\theta) = -E_\theta[\frac{\partial^2}{\partial\theta^2}\log(f(x|\theta))].$$

Recall the **Central Limit Theorem:** Let $X_1, ..., X_n$ be iid with $EX = \mu$ and $Var(X) = \sigma^2$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2).$$

Hence

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) = \sqrt{n}\left(\frac{\sum_{i=1}^n X_i - n\mu}{n\sigma}\right) \xrightarrow{D} N(0, 1).$$

Recall the **Delta Method:** Suppose that $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2)$. Then

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2[g'(\theta)]^2)$$

if $g'(\theta) \neq 0$ exists.

Cox and Hinkley (1974, p. 286) observe that in a one parameter regular exponential family, $T_n = \frac{1}{n}\sum_{i=1}^n t(X_i)$ is the UMVUE and generally the MLE of its expectation $\mu_T = E_\theta(T_n) = E_\theta[t(X)]$. Let $\sigma_T^2 = Var_\theta[t(X)]$. These values can be found by using the distribution of $t(X)$ or by using the Casella and Berger (2002, pp. 112, 133) formulas

$$\mu_T = \frac{-c'(\theta)}{c(\theta)w'(\theta)} = \frac{-\partial}{\partial \eta}\log(c^*(\eta)),$$

and

$$\sigma_T^2 = \frac{\frac{-\partial^2}{\partial \theta^2}\log(c(\theta)) - [w''(\theta)]\mu_T}{[w'(\theta)]^2} = \frac{-\partial^2}{\partial \eta^2}\log(c^*(\eta)).$$

The simplicity of the following result is rather surprising.

Theorem. Let $X_1, ..., X_n$ be iid from a one parameter exponential family with $E(t(X)) = \mu_T \equiv g(\eta)$ and $Var(T(X)) = \sigma_T^2$.

a) Then

$$\sqrt{n}[T_n - \mu_T] \xrightarrow{D} N(0, I_1(\eta))$$

where

$$I_1(\eta) = \sigma_T^2 = g'(\eta) = \frac{[g'(\eta)]^2}{I_1(\eta)}.$$

b) If $\eta = g^{-1}(\mu_T)$, $\hat{\eta} = g^{-1}(T_n)$, $g^{-1'}(\mu_T) \neq 0$ exists, and $\tau'(\eta) \neq 0$ exists, then

$$\sqrt{n}[\tau(\hat{\eta}) - \tau(\eta)] \xrightarrow{D} N\left(0, \frac{[\tau'(\eta)]^2}{I_1(\eta)}\right).$$

Proof: a) The result follows by the central limit theorem if $\sigma_T^2 = I_1(\eta) = g'(\eta)$. Since $f(x|\eta) = h(x)c^*(\eta)\exp[\eta t(x)]$,

$$\frac{\partial}{\partial \eta}\log(f(x|\eta)) = \frac{c^{*'}(\eta)}{c^*(\eta)} + t(x) = -g(\eta) + t(x).$$

Hence

$$\frac{\partial^2}{\partial \eta^2}\log(f(x|\eta)) = \frac{c^{*'}(\eta)c^{*''}(\eta) - [c^{*'}(\eta)]^2}{[c^*(\eta)]^2} = -g'(\eta),$$

and thus

$$I_1(\eta) = \frac{-c^{*''}(\eta)}{c^*(\eta)} + \left[\frac{c^{*'}(\eta)}{c^*(\eta)}\right]^2 = \frac{-\partial^2}{\partial \eta^2}\log(c^*(\eta)) = \sigma_T^2 = g'(\eta).$$

b) By the Delta Method,

$$\sqrt{n}(\hat{\eta} - \eta) = \sqrt{n}[g^{-1}(T_n) - g^{-1}(\mu_T)] \xrightarrow{D} N(0, \sigma_T^2 [g^{-1'}(\mu_T)]^2),$$

but

$$g^{-1'}(\mu_T) = \frac{1}{g'(g^{-1}(\mu_T))} = \frac{1}{g'(\eta)}.$$

Hence

$$\sigma_T^2 [g^{-1'}(\mu_T)]^2 = \frac{[g'(\eta)]^2}{I_1(\eta)} \frac{1}{[g'(\eta)]^2} = \frac{1}{I_1(\eta)}.$$

So

$$\sqrt{n}(\hat{\eta} - \eta) \xrightarrow{D} N\left(0, \frac{1}{I_1(\eta)}\right),$$

and the result follows by the Delta Method. QED

When (as is usually the case) T_n is the MLE of μ_T, $\hat{\eta}$ is the MLE of η by the invariance principle. If $\eta = w(\theta)$ and $\theta = w^{-1}(\eta)$, a limit theorem for $\hat{\theta}$ can also be obtained.

A similar result holds for k-parameter exponential families. Let $\boldsymbol{T}_n = (\sum_{i=1}^n t_1(X_i), ..., \sum_{i=1}^n t_k(X_i))$ and let $\boldsymbol{\mu}_T = (E(t_1(X), ..., E(t_k(X)))$. From Lehmann (1986, p. 66) and Lehmann (1999, pp. 497, 499), for $\boldsymbol{\eta} \in \Omega$,

$$E(T_i(X)) = \frac{-\partial}{\partial \eta_i} \log(c^*(\boldsymbol{\eta})),$$

and

$$Cov(T_i(X), T_j(X)) \equiv \sigma_{i,j} = \frac{-\partial^2}{\partial \eta_i \partial \eta_j} \log(c^*(\boldsymbol{\eta})),$$

and the Information matrix

$$\boldsymbol{I}(\boldsymbol{\eta}) = [\boldsymbol{I}_{i,j}]$$

where

$$\boldsymbol{I}_{i,j} = E\left[\frac{\partial}{\partial \eta_i} \log(f(x|\boldsymbol{\eta})) \frac{\partial}{\partial \eta_i} \log(f(x|\boldsymbol{\eta}))\right] = -E\left[\frac{\partial^2}{\partial \eta_i \partial \eta_j} \log(f(x|\boldsymbol{\eta}))\right].$$

Theorem. If $X_1, ..., X_n$ are iid from a regular k-parameter exponential family, then

$$\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\mu}_T) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{I}(\boldsymbol{\eta})).$$

Proof. By the multivariate central limit theorem,

$$\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\mu}_T) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma} = [\sigma_{i,j}]$. Hence the result follows if $\sigma_{i,j} = \boldsymbol{I}_{i,j}$. Since

$$\log(f(x|\boldsymbol{\eta})) = \log(h(x)) + \log(c^*(\boldsymbol{\eta})) + \sum_{l=1}^{k} \eta_l T_l(x),$$

$$\frac{\partial}{\partial \eta_i} \log(f(x|\boldsymbol{\eta})) = \frac{\partial}{\partial \eta_i} \log(c^*(\boldsymbol{\eta})) + T_i(X).$$

Hence

$$-\boldsymbol{I}_{i,j} = \frac{\partial^2}{\partial \eta_i \partial \eta_j} \log(f(x|\boldsymbol{\eta})) = \frac{\partial^2}{\partial \eta_i \partial \eta_j} \log(c^*(\boldsymbol{\eta})) = -\sigma_{i,j}.$$

7 Conclusions

One of the most important uses of exponential families is that the theory often provides two methods for doing inference. For example, minimal sufficient statistics can be found with either the Lehmann-Scheffé theorem or by finding $\boldsymbol{T}$ from the exponential family parameterization. Similarly, if $X_1, ..., X_n$ are iid from a one parameter REF with complete sufficient statistic $T(\boldsymbol{X})$, then one sided UMP tests can be found by using the Neyman Pearson lemma or by using exponential family theory.

It should be emphasized that the complete sufficient statistic from a regular exponential family is well behaved, even if the exponential family is not (see Casella and Berger 2002, pp. 112, 133). The one sided stable distribution with index 1/2 is an interesting example. See Lehmann (1999, p. 76) and Besbeas and Morgan (2004). An interesting subclass of exponential families is given by Rahman and Gupta (1993).

The main focus of this paper is using exponential families for a one semester course, and many important topics that might be covered by a two semester course have been omitted. For example, exponential families are useful for finding conjugate priors when Bayesian statistics are covered. History and references for additional topics can be found in Lehmann (1983, p. 70), Brown (1986) and Barndorff-Nielsen (1978, 1982).

8 References

Bahadur, R.R. (1958), "Examples of Inconsistency of Maximum Likelihood Estimators," *Sankhya,* 20, 207-210.

Barndorff-Nielsen, O. (1978), *Information and Exponential Families in Statistical Theory*, John Wiley and Sons, NY.

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

Berk, R. (1967), "Review 1922 of 'Invariance of Maximum Likelihood Estimators' by Peter W. Zehna," *Mathematical Reviews,* 33, 342-343.

Besbeas, P. and Morgan, B.J.T. (2004), "Efficient and Robust Estimation for the One-sided Stable Distribution of Index 1/2," *Statistics and Probability Letters,* 66, 251-257.

Bickel, P.J., and Doksum, K.A. (2000), *Mathematical Statistics: Basic Ideas and Selected Topics,* Vol. 1., 2nd ed., Upper Saddle River, NJ.

Brown, L.D. (1986), *Fundamentals of Statistical Exponential Families with Applications* *in Statistical Decision Theory*, Institute of Mathematical Statistics Lecture Notes – Monograph Series, IMS Haywood, CA.

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

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

Johanson, S. (1979), *Introduction to the Theory of Regular Exponential Families,* Institute of Mathematical Statistics, University of Copenhagen, Copenhagen, Denmark.

Joshi, V.M. (1976), "On the Attainment of the Cramér-Rao Lower Bound, *The Annals of Statistics,* 4, 998-1002.

Karakostas, K.X. (1985), "On Minimum Variance Estimators," *The American Statistician,* 39, 303-305.

Lehmann, E.L. (1983), *Theory of Point Estimation,* John Wiley and Sons, NY.

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

Lehmann, E.L. (1999), *Elements of Large–Sample Theory*, Springer-Verlag, NY.

Lehmann, E.L., and Scheffé, H. (1950), "Completeness, Similar Regions, and Unbiased Estimation," *Sankhya,* 10, 305-340.

Levy, M.S. (1985), "A Note on Nonunique MLE's and Sufficient Statistics", *The American Statistician,* 39, 66.

Mood, A.M., and Graybill, F.A., and Boes, D.C. (1974), *Introduction to the Theory of Statistics,* 3rd ed., McGraw-Hill, NY.

Moore, D.S. (1971), "Maximum Likelihood and Sufficient Statistics," *The American Mathematical Monthly,* 78, 50-52.

Portnoy, S. (1977), "Asymptotic Efficiency of Minimum Variance Unbiased Estimators," *The Annals of Statistics,* 5, 522-529.

Rahman, M.S., and Gupta, R.P. (1993), "Family of Transformed Chi-Square Distributions," *Communications in Statistics: Theory and Methods,* 22, 135-146.

Sampson, A., and Spencer, B. (1976), "Sufficiency, Minimal Sufficiency, and the Lack Thereof," *The American Statistician,* 30, 34-35.

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

Wijsman, R.A. (1973), "On the Attainment of the Cramér-Rao Lower Bound, *The Annals of Statistics,* 1, 538-542.

Zehna, P.W. (1966), "Invariance of Maximum Likelihood Estimators," *Annals of Mathematical Statistics,* 37, 744.