A Simple Limit Theorem for Exponential Families

David J. Olive*

Southern Illinois University

May 9, 2007

Abstract

In the literature, it is often stated that maximum likelihood estimators and uniformly minimum variance estimators are asymptotically efficient, under regularity conditions. This paper shows that if $X_1, ..., X_n$ are iid from a k–parameter regular exponential family with complete sufficient statistic $\boldsymbol{T}_n$ with $E(\boldsymbol{T}_n) = \boldsymbol{\mu}_T$, then $\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\mu}_T) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{I}(\boldsymbol{\eta}))$ where $\boldsymbol{I}(\boldsymbol{\eta})$ is the information matrix of the natural parameterization of the family. This result avoids the use of complex regularity conditions, and standard results can be obtained by applying the delta method.

KEY WORDS: Complete Minimal Sufficient Statistic, Delta Method, MLE, UMVUE

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

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

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. 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) holds.

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

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

where $h(x)$ and $t_i(x)$ are the same as in Equation (1) and $\boldsymbol{\eta} \in \Omega$.

The one parameter exponential family $f(x|\theta) = h(x)c(\theta) \exp[w(\theta)t(x)]$ has natural parameterization

$$f(x|\eta) = h(x)c^*(\eta) \exp[\eta t(x)].$$

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.

Some care has to be taken with the definitions of Θ and Ω since formulas (1) and (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 $\eta \in \Omega = (-\infty, \infty)$.

Often students are asked to show that a distribution is a k–parameter REF by setting $\eta_i = w_i(\boldsymbol{\theta})$. Then assume that the natural parameter space

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

Finally, show that Ω is a k-dimensional open set by showing that Ω is a cross product of open intervals. For many "brand name" distributions, $\boldsymbol{\eta}$ is a one to one function of $\boldsymbol{\theta}$, and the above map is correct.

Many references suggest that under regularity conditions, large classes of estimators $\hat{\boldsymbol{\theta}}_n$ are asymptotically efficient:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{D} N_k(\mathbf{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta})) \tag{3}$$

where $\boldsymbol{I}(\boldsymbol{\theta})$ is the information matrix for $\boldsymbol{\theta}$. Lehmann (1999, sections 7.4, 7.5) provides a good discussion on constructing estimators such as the "Hodges counterexample" that have "smaller" asymptotic variance than asymptotically efficient estimators.

For asymptotic efficiency in exponential families, Lehmann (1980) and Barndorff–Nielsen (1982) cite Berk (1972). Brown (1986, Problem 5.15.1, p. 172) and McCulloch (1988) have similar results to those in this paper for exponential families with $t_i(\boldsymbol{x}) = x_i$. Portnoy (1977) provides asymptotic theory for unbiased estimators in exponential families, including conditions under which the UMVUE is asymptotically equivalent to the MLE.

In the one parameter setting, Barndorff–Nielsen (1982), Casella and Berger (2002, pp. 472, 515), Cox and Hinkley (1974, p. 286), Lehmann (1983, 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

$$\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{D} N[0, FCRLB_1(\tau(\theta))] \tag{4}$$

where the Fréchet Cramér Rao lower bound 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

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

The **Multivariate Central Limit Theorem** states that if $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ are iid random vectors with $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$, then

$$\sqrt{n}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \xrightarrow{D} N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

where the sample mean

$$\overline{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i.$$

An important special case is the **Central Limit Theorem** where the vectors are random variables. Let $X_1, ..., X_n$ be iid with $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2)$.

The **Multivariate Delta Method** states that if

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

then

$$\sqrt{n}(\boldsymbol{g}(\boldsymbol{T}_n) - \boldsymbol{g}(\boldsymbol{\theta})) \xrightarrow{D} N_d(\boldsymbol{0}, \boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})}\boldsymbol{\Sigma}\boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\theta})})$$

where the $d \times k$ Jacobian matrix of partial derivatives

$$\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} = \begin{bmatrix} \frac{\partial}{\partial \theta_1} g_1(\boldsymbol{\theta}) & \ldots & \frac{\partial}{\partial \theta_k} g_1(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ \frac{\partial}{\partial \theta_1} g_d(\boldsymbol{\theta}) & \ldots & \frac{\partial}{\partial \theta_k} g_d(\boldsymbol{\theta}) \end{bmatrix}.$$

Here the mapping $\boldsymbol{g} : \Re^k \to \Re^d$ needs to be differentiable in a neighborhood of $\boldsymbol{\theta} \in \Re^k$. See Ferguson (1996, p. 45), Lehmann (1999, p. 315), Mardia, Kent and Bibby (1979, p. 52), Sen and Singer (1993, p. 136) or Serfling (1980, p. 122). An important special case is the **Delta Method** where the estimator T_n and parameter θ are real valued. 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.

The following sections will give simple limit theorems for the one parameter and k–parameter exponential families that avoid the complex regularity conditions of Berk (1972).

2 A SIMPLE LIMIT THEOREM FOR A ONE PARAMETER EXPONENTIAL FAMILY

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 uniformly minimum variance unbiased estimator (UMVUE) and generally the maximum likelihood estimator (MLE) of its expectation $\mu_T = E_\theta(T_n) = E_\theta[t(X)]$. Let $\sigma_T^2 = \mathrm{Var}_\theta[\mathrm{t(X)}]$. 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

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

If $\theta = g(\eta)$ and $g'(\eta) \neq 0$, then we will define $I_1(\theta) = I_1(\eta)/[g'(\eta)]^2$. This is a natural definition when using the delta method or when g is one to one and onto. Also Lehmann (1999, p. 468) shows that if $\theta = g(\eta)$, if g' exists and is continuous, and if $g'(\eta) \neq 0$, then $I_1(\theta) = I_1(\eta)/[g'(\eta)]^2$. The simplicity of the following result is rather surprising.

Theorem 1. Let $X_1, ..., X_n$ be iid from a one parameter exponential family with $E(t(X)) = \mu_T \equiv g(\eta)$ and $\text{Var}(T(X)) = \sigma_T^2$.

a) Then

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

where

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

Hence T_n is asymptotically efficient.

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

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

c) Suppose the conditions in b) hold. If $\theta = w^{-1}(\eta)$, $\hat{\theta} = w^{-1}(\hat{\eta})$and $w^{-1'}(\eta) \neq 0$, then

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

d) If the conditions in c) hold and $\tau'(\theta) \neq 0$, then

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

Proof: a) The result follows by the central limit theorem if $\sigma_T^2 = I_1(\eta) = g'(\eta)$. Since

$\log(f(x|\eta)) = \log(h(x)) + \log(c^*(\eta)) + \eta t(x)$,

$$\frac{\partial}{\partial \eta} \log(f(x|\eta)) = \frac{\partial}{\partial \eta} \log(c^*(\eta)) + t(x) = -g(\eta) + t(x).$$

Hence

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

and thus

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

b) By the delta method,

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

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

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

c) By the delta method,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N\left(0, \frac{[w^{-1'}(\eta)]^2}{I_1(\eta)}\right),$$

but

$$\frac{[w^{-1'}(\eta)]^2}{I_1(\eta)} = \frac{1}{I_1(\theta)}.$$

The last equality holds since if $\theta = g(\eta)$ and if $g'(\eta) \neq 0$, then then $I_1(\theta) = I_1(\eta)/[g'(\eta)]^2$. Use $\eta = w(\theta)$ so $\theta = g(\eta) = w^{-1}(\eta)$.

d) The result follows by the delta method. QED

T_n is the UMVUE and of μ_T.

When (as is usually the case) T_n is the MLE of μ_T, the estimators in b), c) and d) will be MLEs by the invariance principle, but will not generally be unbiased. For example, $\hat{\eta} = g^{-1}(T_n)$ is the MLE of η by invariance. 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).

3 A MULTIVARIATE LIMIT THEOREM

Now suppose that $X_1, ..., X_n$ are iid from a k-parameter REF (2). Then the complete minimal sufficient statistic is $\boldsymbol{T}_n = \frac{1}{n}(\sum_{i=1}^n t_1(X_i), ..., \sum_{i=1}^n t_k(X_i))$. 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

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

Several authors, including Barndorff–Nielsen (1982), have noted that the multivariate CLT can be used to show that $\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\mu}_T) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$. The fact that $\boldsymbol{\Sigma} = \boldsymbol{I}(\boldsymbol{\eta})$ appears in Lehmann (1983, p. 127). Also see Cox (1984) and McColloch (1988).

Theorem 2. If $X_1, ..., X_n$ are iid from a k-parameter regular 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

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

To obtain standard results, use the multivariate delta method, assume that both $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are $k \times 1$ vectors, and assume that $\boldsymbol{\eta} = \boldsymbol{g}(\boldsymbol{\theta})$ is a one to one mapping so that the inverse mapping is $\boldsymbol{\theta} = \boldsymbol{g}^{-1}(\boldsymbol{\eta})$. If $\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})}$ is nonsingular, then

$$\boldsymbol{D}^{-1}_{\boldsymbol{g}(\boldsymbol{\theta})} = \boldsymbol{D}_{\boldsymbol{g}^{-1}(\boldsymbol{\eta})} \tag{5}$$

In analysis, the fact that

$$\boldsymbol{D}^{-1}_{\boldsymbol{g}(\boldsymbol{\theta})} = \boldsymbol{D}_{\boldsymbol{g}^{-1}(\boldsymbol{\eta})}$$

is a corollary of the inverse mapping theorem (or of the inverse function theorem). See Apostol (1957, p. 146), Searle (1982, p. 339) and Wade (2004). Also

$$\boldsymbol{I}(\boldsymbol{\eta}) = [\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})}\boldsymbol{I}^{-1}(\boldsymbol{\theta})\boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\theta})}]^{-1} = [\boldsymbol{D}^{-1}_{\boldsymbol{g}(\boldsymbol{\theta})}]^T\boldsymbol{I}(\boldsymbol{\theta})\boldsymbol{D}^{-1}_{\boldsymbol{g}(\boldsymbol{\theta})} = \boldsymbol{D}^T_{\boldsymbol{g}^{-1}(\boldsymbol{\eta})}\boldsymbol{I}(\boldsymbol{\theta})\boldsymbol{D}_{\boldsymbol{g}^{-1}(\boldsymbol{\eta})}. \quad (6)$$

Compare Lehmann (1999, p. 500), Lehmann (1983, p. 127) and Serfling (1980, p. 158).

For example, suppose that $\boldsymbol{\mu}_T$ and $\boldsymbol{\eta}$ are $k \times 1$ vectors, and

$$\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \stackrel{D}{\to} N_k(\boldsymbol{0}, \boldsymbol{I}^{-1}(\boldsymbol{\eta}))$$

where $\boldsymbol{\mu}_T = \boldsymbol{g}(\boldsymbol{\eta})$ and $\boldsymbol{\eta} = \boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)$. Also assume that $\overline{\boldsymbol{T}}_n = \boldsymbol{g}(\hat{\boldsymbol{\eta}})$ and $\hat{\boldsymbol{\eta}} = \boldsymbol{g}^{-1}(\overline{\boldsymbol{T}}_n)$. Then by the multivariate delta method and Theorem 2,

$$\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\mu}_T) = \sqrt{n}(\boldsymbol{g}(\hat{\boldsymbol{\eta}}) - \boldsymbol{g}(\boldsymbol{\eta})) \stackrel{D}{\to} N_k[\boldsymbol{0}, \boldsymbol{I}(\boldsymbol{\eta})] = N_k[\boldsymbol{0}, \boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\eta})}\boldsymbol{I}^{-1}(\boldsymbol{\eta})\boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\eta})}].$$

Hence

$$\boldsymbol{I}(\boldsymbol{\eta}) = \boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\eta})}\boldsymbol{I}^{-1}(\boldsymbol{\eta})\boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\eta})}.$$

Similarly,

$$\sqrt{n}(\boldsymbol{g}^{-1}(\boldsymbol{T}_n) - \boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)) = \sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \stackrel{D}{\to} N_k[\boldsymbol{0}, \boldsymbol{I}^{-1}(\boldsymbol{\eta})] = N_k[\boldsymbol{0}, \boldsymbol{D}_{\boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)}\boldsymbol{I}(\boldsymbol{\eta})\boldsymbol{D}^T_{\boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)}].$$

Thus

$$\boldsymbol{I}^{-1}(\boldsymbol{\eta}) = \boldsymbol{D}_{\boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)}\boldsymbol{I}(\boldsymbol{\eta})\boldsymbol{D}^T_{\boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)} = \boldsymbol{D}_{\boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)}\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\eta})}\boldsymbol{I}^{-1}(\boldsymbol{\eta})\boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\eta})}\boldsymbol{D}^T_{\boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)}$$

as expected by Equation (5). Typically $\hat{\boldsymbol{\theta}}$ is a function of the sufficient statistic $\boldsymbol{T}_n$ and is the unique MLE of $\boldsymbol{\theta}$. Replacing $\boldsymbol{\eta}$ by $\boldsymbol{\theta}$ in the above discussion shows that

$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N_k(\mathbf{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta}))$ is equivalent to $\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\mu}_T) \xrightarrow{D} N_k(\mathbf{0}, \boldsymbol{I}(\boldsymbol{\theta}))$ provided that $\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})}$ is nonsingular.

Theorem 1 can be taught whenever the CLT is taught while Theorem 2 can be taught whenever the multivariate delta method is taught.

4 REFERENCES

Apostol, T.M. (1957), *Mathematical Analysis A Modern Approach to Advanced Calculus,* Addison-Wesley, Reading, MA.

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.

Berk, R.H., (1972), "Consistency and Asymptotic Normality of MLE's for Exponential Models," *The Annals of Mathematical Statistics*, 43, 193-204.

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., Belmont, CA.

Cox, C. (1984), "An Elementary Introduction to Maximum Likelihood Estimations for Multinomial Models: Birch's Theorem and the Delta Method," *The American Statistician*, 38, 283-287.

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

Ferguson, T.S. (1996), *A Course in Large Sample Theory*, Chapman & Hall, NY.

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

Lehmann, E.L. (1980), "Efficient Likelihood Estimators," *The American Statistician*, 34, 233-235.

Lehmann, E.L. (1983), *Theory of Point Estimation,* 1st ed., John Wiley and Sons, NY.

Lehmann, E.L. (1986), *Testing Statistical Hypotheses,* 2nd ed., John Wiley and Sons, NY.

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

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979), *Multivariate Analysis,* Academic Press, London.

McCulloch, R.E. (1988), "Information and the Likelihood Function in Exponential Families, *The American Statistician*, 42, 73-75.

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

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

Searle, S.R. (1982), *Matrix Algebra Useful for Statistics*, John Wiley and Sons, NY.

Sen, P.K., and Singer, J.M. (1993), *Large Sample Methods in Statistics: An Introduction with Applications,* Chapman & Hall, NY.

Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics,* John Wiley and Sons, NY.

Wade, W.R, (2004), *Introduction to Analysis*, 3rd ed., Prentice Hall, Upper Saddle

River, NJ.

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