REGULARITY CONDITIONS, ASYMPTOTICS, AND THE EXPONENTIAL CLASS

by

BU-805-M January, 1983

Walter W. Piegorsch

Biometrics Unit, Cornell University, Ithaca, New York

Abstract

The regularity conditions for the consistency, efficiency, and asymptotic Normality of the maximum likelihood estimate (MLE) are quite complicated; their explicit reference in an early course on mathematical statistics is rare. In this paper, an example showing the inability of a probability density to satisfy these conditions — even though the MLE is asymptotically Normal — is considered. A simplified set of conditions for the asymptotic Normality of the MLE is derived when the underlying density function is a member of the one-parameter, natural exponential family. These conditions involve functions of the exponential factorization, and of the distribution's low-order expectations. They are, therefore, simple enough to warrant explicit reference in a post-calculus statistics course. Tabular factorizations of some exponential class densities are included.

Paper No. BU-805-M in the Biometrics Unit.

1. INTRODUCTION

Post-calculus courses in theoretical statistics sometimes consider the optimality of a method of estimation by examining the large-sample properties of the estimator produced. A well-known case is that of maximum likelihood (ML) estimation. Under certain regularity conditions on the underlying density function, the ML estimator is consistent, efficient, and asymptotically Normal (Norden, 1972, sec. 5, 1973, sec. 6). It is unusual, however, to see these regularity conditions explicitly stated; indeed, many are complicated enough to perhaps be left unspecified for a first course in mathematical statistics.

It was Cramér (1946, p. 500) who first put forth a set of sufficient conditions for the asymptotic Normality of the ML estimator. Since then, the investigation and reconstruction of these regularity conditions have received a great deal of attention in the literature (Daniels, 1961; Weiss, 1963, 1966; LeCam, 1970; Lehmann, 1980). Norden's review (1972) on ML estimation provides a particularly extensive bibliography.

Many authors considering Cramér's conditions argue that, because of their complexity, they do not always apply to a particular density. There are numerous examples of inconsistent ML estimators (Bahadur, 1958; Ferguson, 1982). Others note that the conditions may fail to hold even though the ML estimator is asymptotically Normal (Kulldorf, 1957).

Even when the underlying density is a member of the rather well-behaved exponential class, there are examples where Cramér's conditions are not satisfied. In section 2 we consider such an example, along with a series of conditions – proposed by Kulldorf (1957) – which are simpler and easier to apply. In section 3, Kulldorf's conditions are used to develop a simplified set of conditions for exponential class densities. These new conditions assure the asymptotic Normality of the ML estimator. Also, they are simple enough to state and verify in many cases, so as to make their explicit reference in a post-calculus statistics course worthwhile.

2. EXAMPLE

Denote the density function as $f \equiv f(x;\theta)$, where $\theta \in \Theta$ is the parameter under consideration. Cramér's conditions can be summarized as follows (Norden, 1973):

C.1. $\partial^i \log f/\partial\theta^i$ exist $\forall\ \theta$ and almost all x (i = 1, 2, 3);

C.2. $\int_{-\infty}^{\infty} \frac{\partial^i f}{\partial\theta^i}\, dx = 0 \quad \forall\ \theta \quad (i = 1, 2)$;

C.3. $E[\partial^2 \log f(X;\theta)/\partial\theta^2]$ exists, and is negative, $\forall\ \theta$;

C.4. There is some H(x) such that $|\partial^3 \log f/\partial\theta^3| < H(x)$ and $E[H(X)]$ exists $\forall\ \theta$.

When all four conditions are satisfied, Cramér (1946, p. 500) shows that the ML estimator of θ, $\hat{\theta}_n$, exists, and that the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ is $N\big(0, 1/I(\theta)\big)$. $I(\theta) = E[(\partial \log f(X;\theta)/d\theta)^2]$ is the Fisher information value.

Kulldorf (1957) gives an example where the ML estimator has such an asymptotic distribution, yet the density does not satisfy condition 4: the $N(0,\theta)$ random variable. A perhaps simpler example with a discrete density is the Poisson (θ) distribution.

<u>Example</u>:
$$f(x;\theta) = \frac{e^{-\theta}\theta^x}{x!} I_{\{0,1,\ldots\}}(x) \qquad (2.1)$$

with $\Theta = (0,\infty)$. It is easy to show that conditions 1 through 3 are satisfied for this case. The first three partial derivatives of $\log f$ are $(x/\theta) - 1$, $-x/\theta^2$, and $2x/\theta^3$. They all exist as necessary. Also, since

$$\frac{\partial f(x;\theta)}{\partial\theta} = f(x;\theta)(\frac{x}{\theta} - 1) \tag{2.2}$$

and

$$\frac{\partial^2 f(x;\theta)}{\partial\theta^2} = f(x;\theta)[1 - x(\theta^{-2} + 2\theta^{-1}) + (x/\theta)^2] , \tag{2.3}$$

it is easy to show that the expectations in C.2 are both identically equal to zero. Lastly, $E[-X/\theta^2] = -\theta^{-1}$ is always negative for $\theta > 0$.

However, the value $|\partial^3 \log f/\partial\theta^3| = 2x\theta^{-3}$ is not bounded in the open interval $0 < \theta < \infty$. Thus, condition 4 is not satisfied.

Still, it is well-known that the ML estimator here is $\hat{\theta}_n = \bar{X}_n$. The central limit theorem gives the asymptotic distribution of $\sqrt{n}(\bar{X}_n - \theta)$ as $N(0,\theta)$ here, since $E[X] = \theta$. Also, because $I(\theta) = var(X)/\theta^2 = \theta^{-1}$, we see that the ML estimator has the desired asymptotic properties, even though condition 4 is violated.

Kulldorf (1957) has proposed a substitute for C.4:

C.5. There is some $g(\theta)$ which is positive and twice differentiable $\forall\ \theta$, and a function $H(x)$ such that, for each θ,

$$\left|\frac{\partial^2}{\partial\theta^2}\left[g(\theta)\ \frac{\partial \log f}{\partial\theta}\right]\right| < H(x)$$

and

$$E(H[X]) < \infty .$$

He then shows that conditions 1, 2, 3, and 5 are sufficient for the asymptotic Normality which we seek. In the Poisson example, we see that condition 5 is easily satisfied by $g(\theta) = \theta$, and by taking $H(x)$ as any positive constant. Thus Kulldorf's theorem substantiates the results of our example.

3. SIMPLIFIED CONDITIONS FOR THE EXPONENTIAL CLASS

The literature on estimation in the exponential class is both large and diversified (Davidson and Solomon, 1974; Hwang, 1982), including works on ML estimation (Berk, 1972; Crain, 1976; Nordberg, 1980). This one-parameter class of functions involves densities of the form

$$f(x;\theta) = a(\theta)b(x)\exp\{c(\theta)t(x)\} \ . \qquad (3.1)$$

The appendix lists factorizations of some common one-parameter densities in the form of (3.1). A slightly simpler and easier to deal with factorization involves the natural parameterization of f:

$$f(x;\tau) = \gamma(\tau)b(x)\exp\{\tau t(x)\} \qquad (3.2)$$

where $\tau \equiv \tau(\theta)$ is the natural parameter of X. The natural parameter space — the subset of $\mathbb{R}$ for which (3.2) represents a density — is denoted by $\mathcal{T}$ (Ferguson, 1967, p. 128). Some examples of random variables in this τ-based factorization appear in Table 1.

TABLE 1: Examples of Naturally Parameterized Exponential Class Densities.

Distribution	$\tau \equiv \tau(\theta)$	$\mathcal{T}$	$\gamma(\tau)$	$b(x)$	$t(x)$
Binomial (n,θ)	$\log \frac{\theta}{1-\theta}$	$(-\infty,\infty)$	$(1+e^{\tau})^{-n}$	$\binom{n}{x}I_{\{0,1,\cdots,n\}}(x)$	x
Geometric	$\log(1-\theta)$	$(-\infty,0)$	$1-e^{\tau}$	$I_{\{0,1,\cdots\}}(x)$	x
Poisson	$\log\theta$	$(-\infty,\infty)$	$\exp\{-e^{\tau}\}$	$\frac{1}{x!}\, I_{\{0,1,\cdots\}}(x)$	x
Exponential	θ	$(0,\infty)$	τ	$I_{(0,\infty)}(x)$	$-x$
Beta $(\theta,1)$	$\theta - 1$	$(-1,\infty)$	$\tau+1$	$I_{(0,1)}(x)$	$\log x$
Normal $(0,\theta)$	$-1/2\theta$	$(-\infty,0)$	$(-2\tau)^{\frac{1}{2}}$	$(2\pi)^{-\frac{1}{2}}$	x^2
Normal $(\theta,1)$	θ	$(-\infty,\infty)$	$\exp\{-\pi/2\}$	$[2\pi\exp(x^2)]^{-\frac{1}{2}}$	x

Random variables with such density functions possess a wide variety of optimal properties (see Ferguson, 1967, or Lindgren, 1976). For example, Ferguson (1967, p. 129) suggests that derivatives of all orders may be passed within the integral in expressions of the form $\frac{\partial}{\partial\tau} E[\phi(X)]$. This is critical in the following lemmas.

<u>Lemma 3.1</u>: $E[t(X)] = -\gamma'(\tau)/\gamma(\tau)$.

<u>Proof</u>: Since $f(x;\tau)$ is a density, we know

$$\int \gamma(\tau)b(x)e^{\tau t(x)}dx = 1 \ ,$$

so that

$$\frac{\partial}{\partial\tau}\int \gamma(\tau)b(x)e^{\tau t(x)}dx = 0 \tag{3.3}$$

(the integrals are taken over the entire support of X). From the result in Ferguson (1967) mentioned above, (3.3) gives

$$\begin{aligned}\int b(x)e^{\tau t(x)}&[\gamma'(\tau) + \gamma(\tau)t(x)]dx \\ &= \frac{\gamma'(\tau)}{\gamma(\tau)}\left[\int f(x;\tau)dx\right] + \int t(x)f(x;\tau)dx \\ &= \frac{\gamma'(\tau)}{\gamma(\tau)} + E[t(X)] = 0 \ .\end{aligned}$$

Thus the lemma is proven. Lemma 3.1 can be extended to include

<u>Lemma 3.2</u>: $\frac{\partial}{\partial\tau} E[t^n(X)] = E[t^{n+1}(X)] - E[t(X)]E[t^n(X)]$.

<u>Proof</u>: As in Lemma 3.1,

$$\begin{aligned}\frac{\partial}{\partial\tau}\int t^n(x)&\gamma(\tau)b(x)e^{\tau t(x)}dx \\ &= \int t^n(x)b(x)e^{\tau t(x)}[\gamma'(\tau) + \gamma(\tau)t(x)]dx \\ &= \frac{\gamma'(\tau)}{\gamma(\tau)} E[t^n(X)] + E[t^{n+1}(X)] \ .\end{aligned}$$

From Lemma 3.1, this is

$$\frac{\partial}{\partial\tau} E[t^n(X)] = E[t^{n+1}(X)] - E[t(X)]E[t^n(X)] .$$

<u>Corollary</u>: $\frac{\partial}{\partial\tau} E[t(X)] = \text{var}[t(X)]$.

We are now ready to consider the simplified conditions for the natural exponential class.

<u>Theorem</u>: For a density of the form (3.2), the asymptotic distribution of $\sqrt{n}(\hat{\tau}_n - \tau)$ is $N\left(0, 1/I(\tau)\right)$ when the following conditions are satisfied:

(a) The quantities $E[t(X)]$, $\text{var}[t(X)]$, and $\frac{\partial}{\partial\tau}\text{var}[t(X)]$ exist $\forall\ \tau$.

(b) $\gamma(\tau) > 0$ and $\gamma'''(\tau)$ exists $\forall\ \tau$.

<u>Proof</u>: Consider each of Kulldorf's conditions separately:

C.1. $\partial \log f/\partial\tau = t(x) + \gamma'(\tau)/\gamma(\tau)$ which, by Lemma 3.1, is simply $t(x) - E[t(X)]$.

$\partial^2 \log f/\partial\tau^2 = -\frac{\partial}{\partial\tau} E[t(X)]$, since $t(x)$ is independent of τ (and, we should note, exists for all x). From the corollary to Lemma 3.2, this is simply $-\text{var}[t(X)]$.

Thus when condition (a) is satisfied, so is C.1.

C.2. $$\partial f/\partial\tau = \gamma(\tau)b(x)t(x)e^{\tau t(x)} + \gamma'(\tau)b(x)e^{\tau t(x)}$$
$$= \gamma(\tau)b(x)e^{\tau t(x)}\left[t(x) + \left(\gamma'(\tau)/\gamma(\tau)\right)\right], \text{ so}$$

$$\int_{-\infty}^{\infty} \frac{\partial f}{\partial\tau}\, dx = E[t(X)] + \frac{\gamma'(\tau)}{\gamma(\tau)} \equiv 0 \quad \text{from Lemma 3.1.}$$

Similarly, it is easy to show that

$$\int_{-\infty}^{\infty} \frac{\partial^2 f}{\partial\tau^2}\, dx = \frac{\gamma''(\tau)}{\gamma(\tau)} - 2E^2[t(X)] + E[t^2(X)] . \qquad (3.4)$$

Now, as in Lemma 3.1, we can say that

$$0 = \frac{\partial^2}{\partial\tau^2} \int \gamma(\tau) b(x) e^{\tau t(x)} dx$$

$$= \frac{\partial}{\partial\tau}\left[\frac{\gamma'(\tau)}{\gamma(\tau)} + E[t(X)]\right]$$

$$= \frac{\gamma''(\tau)}{\gamma(\tau)} - \left[\frac{\gamma'(\tau)}{\gamma(\tau)}\right]^2 + \text{var}[t(X)]$$

$$= \frac{\gamma''(\tau)}{\gamma(\tau)} - E^2[t(X)] + E[t^2(X)] - E^2[t(X)] ,$$

so that

$$\frac{\gamma''(\tau)}{\gamma(\tau)} = 2E^2[t(X)] - E[t^2(X)] .$$

Then, from (3.4), we have

$$\int_{-\infty}^{\infty} \frac{\partial^2 f}{\partial\tau^2}\, dx \equiv 0 .$$

Thus C.2 is satisfied.

C.3. Since $\partial^2 \log f/\partial\tau^2 = -\text{var}[t(X)]$ is a negative constant with respect to X, $E[\partial^2 \log f(X;\tau)/\partial\tau^2]$ is always negative.

C.5. If we let $g(\tau) = \gamma(\tau)$, then condition (b) assures us of a properly defined choice for g.

To see that the condition is satisfied, take

$$\left| \frac{\partial^2}{\partial\tau^2}\left(g\, \frac{\partial \log f}{\partial\tau}\right) \right| = \left| t(x)\gamma''(\tau) + \gamma'''(\tau)\right| . \tag{3.5}$$

From the triangle inequality, this is less than or equal to $|t(x)\gamma''(\tau)| + |\gamma'''(\tau)|$. Therefore, we could set

$$H(x) = |t(x)\gamma''(\tau)| + |\gamma'''(\tau)| + 1 .$$

Its expectation exists $\forall\ \tau$ when $\gamma'''(\tau)$ and $E[t(X)]$ exist $\forall\ \tau$. Both of these requirements are satisfied under conditions (a) and (b), thus C.5 is satisfied.

Since these four conditions have been met, Kulldorf's theorem assures us of the result.

It is relatively easy, and in some cases a pedagogic use of mathematical and statistical methods, to verify the conditions of the theorem for many natural exponential models. All of the densities presented in Table 1 satisfy the two conditions; as an example, let's return to the Poisson distribution. The natural parameter is $\tau = \log\theta$. Its ML estimator, by the ML invariance property, is $\hat{\tau}_n = \log \bar{X}_n$. Now, $t(x) = x$ so that $E[X] = \text{var}[X] = e^{\tau}$ satisfies condition (a) quickly ($\tau \in R$). Also, $\gamma(\tau) = \exp(-e^{\tau})$ is certainly positive and thrice differentiable on R. Thus we can conclude that $\sqrt{n}(\hat{\tau}_n - \tau)$ has an asymptotic Normal distribution.

It should be noted that condition (b) can be modified in those unusual cases where $\gamma(\tau)$ does not satisfy it. The choice of $g(\tau) = \gamma(\tau)$ in Kulldorf's condition 5 was done for convenience, i.e., since $\partial \log f/\partial\tau = \gamma'(\tau)/\gamma(\tau)$, the product $g \cdot \partial \log f/\partial\tau$ simplifies to $\gamma'(\tau)$. Other possible choices of g include $g(\tau) = \gamma(\tau)/\gamma'(\tau)$ or $g(\tau) = \gamma(\tau)\gamma'(\tau)$, depending on the structure of $\gamma(\tau)$ and $\gamma'(\tau)$.

ACKNOWLEDGMENT

I wish to thank Dr. George Casella for his very helpful comments and suggestions for the preparation of this manuscript.

REFERENCES

Bahadur, R. R. (1958). Examples of inconsistency of maximum likelihood estimates. Sankhyā 20, 207-210.

Berk, R. H. (1972). Consistency and asymptotic Normality of MLE's for exponential models. Ann. Math. Statist. 43, 193-204.

Crain, B. R. (1976). Exponential models, maximum likelihood estimation, and the Haar condition. J. Amer. Statist. Assoc. 71, 737-740.

Cramér, Harald (1946). Mathematical Methods of Statistics. Princeton: Princeton Univ. Press.

Daniels, H. E. (1961). The asymptotic efficiency of the maximum likelihood estimator. Proc. 4th Berkeley Symp. Math. Statist. Prob. 1, 151-163.

Davidson, R. R. and Solomon, D. L. (1974). Moment-type estimation in the exponential family. Commun. Statist. 3, 1101-1108.

Ferguson, T. S. (1967). Mathematical Statistics: A Decision Theoretic Approach. New York: Academic Press.

Ferguson, T. S. (1982). An inconsistent maximum likelihood estimate. J. Amer. Statist. Assoc. 77, 831-834.

Hwang, J. T. (1982). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial case. Ann. Statist. 10, 857-867.

Kulldorf, G. (1957). On the conditions for consistency and asymptotic efficiency of maximum likelihood estimates. Skand. Aktuarietidskrift 40, 129-144.

LeCam, L. (1970). On the assumptions used to prove asymptotic Normality of maximum likelihood estimates. Ann. Math. Statist. 41, 802-828.

Lehmann, E. L. (1980). Efficient likelihood estimators. Amer. Statist. 34, 233-235.

Lindgren, B. W. (1976). Statistical Theory, 3rd ed. New York: Macmillan.

Nordberg, L. (1980). Asymptotic Normality of maximum likelihood estimators based on independent, unequally distributed observations in exponential family models. Scand. J. Statist. 7, 27-32.

Norden, R. H. (1972). A survey of maximum likelihood estimation. Int. Statist. Rev. 40, 329-354.

Norden, R. H. (1973). A survey of maximum likelihood estimation II. Int. Statist. Rev. 41, 39-58.

Weiss, L. (1963). The relative maxima of the likelihood function. Skand. Aktuarietidskrift 46, 162-166.

Weiss, L. (1966). The relative maxima of the likelihood function II. Skand. Aktuarietidskrift 49, 119-121.

APPENDIX: Some one-parameter exponential families in their θ-based factorizations.

Distribution	$f(x;\theta)$	$a(\theta)$	$b(x)$	$c(\theta)$	$t(x)$
Binomial (n,θ)	$\binom{n}{x}\theta^x(1-\theta)^{n-x}I_{\{0,1,\cdots,n\}}(x)$	$(1-\theta)^n$	$\binom{n}{x}I_{\{0,1,\cdots,n\}}(x)$	$\ln\frac{\theta}{1-\theta}$	x
Geometric	$\theta(1-\theta)^x I_{\{0,1,\cdots\}}(x)$	θ	$I_{\{0,1,\cdots\}}(x)$	$\ln(1-\theta)$	x
Neg. Binomial (r,θ)	$\binom{r+x-1}{x}\theta^r(1-\theta)^x I_{\{0,1,\cdots\}}(x)$	θ^r	$\binom{r+x-1}{x}I_{\{0,1,\cdots\}}(x)$	$\ln(1-\theta)$	x
Poisson	$e^{-\theta}\frac{\theta^x}{x!}I_{\{0,1,\cdots\}}(x)$	$e^{-\theta}$	$\frac{1}{x!}I_{\{0,1,\cdots\}}(x)$	$\ln\theta$	x
Trunc. Poisson	$\frac{e^{-\theta}\theta^x}{(1-e^{-\theta})x!}I_{\{1,2,\cdots\}}(x)$	$1/(e^{\theta}-1)$	$\frac{1}{x!}I_{\{1,2,\cdots\}}(x)$	$\ln\theta$	x
Logarithmic	$-\frac{(1-\theta)^x}{x\ln\theta}I_{\{1,2,\cdots\}}(x)$	$1/\ln\theta$	$-\frac{1}{x}I_{\{1,2,\cdots\}}(x)$	$\ln(1-\theta)$	x
Exponential	$\theta e^{-\theta}I_{(0,\infty)}(x)$	θ	$I_{(0,\infty)}(x)$	$-\theta$	x
Gamma (n,θ)	$\frac{\theta^n x^{n-1}e^{-\theta x}}{(n-1)!}I_{(0,\infty)}(x)$	θ^n	$\frac{x^{n-1}}{(n-1)!}I_{(0,\infty)}(x)$	$-\theta$	x^{n-1}
Beta $(\theta,1)$	$\theta x^{\theta-1}I_{(0,1)}(x)$	θ	$I_{(0,1)}(x)$	$\theta-1$	$\ln x$
Maxwell	$x^2\sqrt{2/\pi\theta^3}\exp\{-x^2/2\theta\}I_{(0,\infty)}(x)$	$\theta^{-3/2}$	$x^3\sqrt{2/\pi}\,I_{(0,\infty)}(x)$	$-1/2\theta$	x^2
Rayleigh	$\frac{x}{\theta}\exp\{-x^2/2\theta\}I_{(0,\infty)}(x)$	$1/\theta$	$xI_{(0,\infty)}(x)$	$-1/2\theta$	x^2
Pareto (x_0,θ)	$\frac{\theta x_0^{\theta}}{x^{\theta+1}}I_{(x_0,\infty)}(x)$	θx_0^{θ}	$I_{(x_0,\infty)}(x)$	$-(\theta+1)$	$\ln x$
Normal $(0,\theta)$	$(2\pi\theta)^{-\frac{1}{2}}\exp\{-x^2/2\theta\}$	$\theta^{-\frac{1}{2}}$	$(2\pi)^{-\frac{1}{2}}$	$-1/2\theta$	x^2
Normal $(\theta,1)$	$(2\pi)^{-\frac{1}{2}}\exp\{-(x-\theta)^2/2\}$	$\exp\{-\theta/2\}$	$(2\pi e^{x^2})^{-\frac{1}{2}}$	θ	x
Normal (θ,θ)	$(2\pi\theta)^{-\frac{1}{2}}\exp\{-(x-\theta)^2/2\theta\}$	$\exp\{-\theta/2\}/\sqrt{\theta}$	$e^x/\sqrt{2\pi}$	$-1/2\theta$	x^2
Lognormal (θ,θ)	$\frac{\exp\{-(\ln x-\theta)^2/2\theta\}}{x\sqrt{2\pi\theta}}I_{(0,\infty)}(x)$	$\exp\{-\theta/2\}/\sqrt{\theta}$	$(2\pi)^{-\frac{1}{2}}I_{(0,\infty)}(x)$	$-1/2\theta$	$\ln^2 x$

Source: Adapted, in part, from Lindgren (1976, Ch. 5).