The Annals of Statistics
1975, Vol. 3, No. 1, 223–227

ASYMPTOTIC NORMALITY OF THE POSTERIOR DISTRIBUTION FOR EXPONENTIAL MODELS

BY BRADFORD R. CRAIN AND RONNIE L. MORGAN

University of Oklahoma and West Chester State College

Let $f(x)$ be a pdf of exponential form with respect to the measure μ. Suppose a prior pdf π has been placed on the natural parameter space Ω, where π is a density (with respect to m-dimensional Lebesgue measure) which is both positive and continuous at τ^*, the true but unknown parameter value. Using basic properties of exponential families and certain associated convex functions, it is shown that the posterior pdf tends to the multivariate normal.

1. Introduction. The asymptotic normality of the posterior distribution has been discussed widely in the literature. Lindley [9] addresses the problem intuitively with a heuristic approach. Le Cam [8] shows that under very general conditions, the scaled posterior distribution converges to the normal distribution for almost all sample sequences. Walker [10] and Dawid [5] establish the limiting normality of posterior distributions in the presence of a fair list of regularity conditions on the density and parameter space. Johnson [6] does likewise but derives an asymptotic expansion for the posterior distribution involving the standard normal cdf as leading term. Johnson [7] treats the special case of the one-parameter exponential family in a later edition.

In this paper we use a technique suggested by Buehler [4], plus some well-known properties of exponential families and convex sets and functions, but no special regularity conditions, to show that the density of the scaled posterior distribution tends with increasing sample size to the multivariate normal density, and hence convergence in distribution will follow via Scheffé's theorem.

2. Assumptions and notation. Let $f(x \mid \tau)$ be a pdf over the measurable space $\mathscr{X}$ with respect to the measure μ. Assume an exponential model with natural parametrization so that the pdf $f(x \mid \tau)$ is given by

$$f(x \mid \tau) = \exp[\tau'\phi(x) - \psi(\tau)] . \tag{2.1}$$

Here $\phi(x) = (\phi_1(x), \phi_2(x), \cdots, \phi_m(x))'$, $x \in \mathscr{X}$ and the parameter $\tau = (\tau_1, \tau_2, \cdots, \tau_m)'$ varies in the natural parameter space $\Omega \subseteq R^m$, where

$$\Omega = \{\tau : \textstyle\int \exp[\tau'\phi(x)]\, d\mu(x) < \infty\} .$$

The normalizing function $\psi(\tau)$ equals $\log \int \exp[\tau'\phi(x)]\, d\mu(x)$, with the integration

Received March 1973; revised January 1974.

AMS 1970 *subject classifications.* Primary 62E20; Secondary 41A60, 60F05.

Key words and phrases. Posterior distribution, asymptotic normality, exponential models, Bayesian methods.

over $\mathscr{X}$. We assume without loss of generality that Ω has nonempty interior, and that the components of $\phi(\cdot)$ do not lie on a flat $[\mu]$. This latter condition means if for some $c \in R^m$ and $c_0 \in R$, $c'\phi(x) = c_0[\mu]$, then both c and c_0 vanish. Finally, assume that $\tau^* = (\tau_1^*, \tau_2^*, \cdots, \tau_m^*)'$, the true but unknown value of τ, lies in the interior Ω^0 of Ω.

3. Properties of exponential families. We list here for the convenience of the reader some basic properties of exponential families, a recent exposition of which is to be found in Berk [3].

P.1. The function $\psi(\tau)$ is convex over the convex set Ω.

P.2. $\psi(\tau)$ is lower semi-continuous.

Let X be a random variable with pdf $f(x \mid \tau)$. Since the functions 1, $\phi_1(x)$, $\phi_2(x), \cdots, \phi_m(x)$ are assumed linearly independent $[\mu]$, the random variable $\phi(X)$ will not be distributed over any flat F in R^m, which means for any flat $F \subseteqq R^m$, $\Pr(\phi(X) \notin F) > 0$.

P.3. The following statements are equivalent.

(i) $\phi(X)$ is not distributed on a flat.

(ii) $\psi(\tau)$ is strictly convex on Ω.

(iii) $\tau \to f(\cdot \mid \tau)$ is 1–1 on Ω.

P.4. $\psi(\tau)$ is differentiable, even analytic, on Ω^0.

P.5. $E_\tau(\phi(X)) = \dot{\psi}(\tau)$, $\tau \in \Omega^0$, where $\dot{\psi}(\tau) = (\partial\psi(\tau)/\partial\tau_1, \cdots, \partial\psi(\tau)/\partial\tau_m)'$.

P.6. $\ddot{\psi}(\tau) = (\partial^2\psi(\tau)/\partial\tau_i\,\partial\tau_j)$ is the variance-covariance matrix $(m \times m)$ of $\phi(X)$.

P.7. The mapping $\dot{\psi}(\cdot)$ is 1–1.

P.8. The set $\dot{\psi}(\Omega^0)$ is open.

4. Asymptotic normality of the posterior distribution τ. Let $X_1, X_2, \cdots, X_n$ be a random sample from $f(x \mid \tau)$, where $\tau = \tau^*$ is the actual but unknown value of τ. The likelihood function is

$$L(\tau \mid X_1, \cdots, X_n) = \prod_{j=1}^n \exp\{\tau'\phi(X_j) - \psi(\tau)\} \\ = \exp\{n[\tau'\bar{\phi}_n - \psi(\tau)]\}, \tag{4.1}$$

where $\bar{\phi}_n = n^{-1} \sum_{j=1}^n \phi(X_j)$. Define $\tau_n^* = (\tau_{n1}^*, \tau_{n2}^*, \cdots, \tau_{nm}^*)'$ to be the (unique) maximum likelihood estimator of τ^*. τ_n^* will exist and be in Ω^0 for n sufficiently large. Let $\theta = (\theta_1, \theta_2, \cdots, \theta_m)'$ be defined by $\theta = \dot{\psi}(\tau^*)$. Then $\theta = E_{\tau^*}\phi(X)$. By the likelihood equation (4.1), τ_n^* satisfies $\dot{\psi}(\tau_n^*) = \bar{\phi}_n$. Using the inverse function theorem (see Apostle [1]), $\dot{\psi}(\cdot)$ will have a local inverse which is continuous. Since $\bar{\phi}_n \to \theta$ a.s. as $n \to \infty$, $\tau_n^* \to \tau^*$ with probability one.

Now let $\pi(\tau)$ be a prior pdf over Ω with respect to m-dimensional Lebesgue measure which is continuous and positive at $\tau = \tau^*$. The posterior density of τ, given the observations $X_1, X_2, \cdots, X_n$, is $p_n(\tau) = \pi(\tau)\exp\{n[\tau'\bar{\phi}_n - \psi(\tau)]\}/\int_\Omega \pi(\tau)\exp\{n[\tau'\bar{\phi}_n - \psi(\tau)]\,d\tau$. The centered and scaled variate $Z = n^{\frac{1}{2}}(\tau - \tau_n^*)$ will have a density which tends pointwise in the limit to the multivariate normal with mean zero and covariance matrix $[\ddot{\psi}(\tau^*)]^{-1}$ (note that $\ddot{\psi}(\tau)$

and $[\ddot{\psi}(\tau)]^{-1}$ are positive definite for every $\tau \in \Omega^0$). The density for Z, $g_n(z)$, is then

$$g_n(z) = \frac{\pi(z_n + \tau_n^*)\exp\{n[(z_n + \tau_n^*)'\phi_n - \psi(z_n + \tau_n^*)]\}}{n^{m/2}\int_\Omega \pi(\tau)\exp\{n[\tau'\phi_n - \psi(\tau)]\}\,d\tau}, \tag{4.2}$$

where $z_n = zn^{-\frac{1}{2}}$. Let $k_n(z)$ be defined by

$$k_n(z) = (z + \tau_n^*)'\phi_n - \psi(z + \tau_n^*). \tag{4.3}$$

Then $g_n(z)$ can be written

$$g_n(z) = \frac{\pi(z_n + \tau_n^*)\exp\{n[k_n(z_n) - k_n(0)]\}}{n^{m/2}\int_\Omega \pi(\tau)\exp\{n[\tau'\phi_n - \psi(\tau) - k_n(0)]\}\,d\tau}. \tag{4.4}$$

By the Taylor series expansion of $k_n(z)$ about $z = 0$,

$$\begin{aligned} k_n(z_n) - k_n(0) &= z_n'[\phi_n - \dot{\psi}(\tau_n^*)] - z_n'\ddot{\psi}(\xi_n + \tau_n^*)z_n/2 \\ &= -(2n)^{-1}z'\ddot{\psi}(\xi_n + \tau_n^*)z, \end{aligned} \tag{4.5}$$

where ξ_n lies on the line segment between $z_n = zn^{-\frac{1}{2}}$ and the origin. Since $\tau_n^* \to \tau^*$ with probability one, it follows that $P[\forall z, \ddot{\psi}(\xi_n + \tau_n^*) \to \ddot{\psi}(\tau^*)] = 1$. Thus with probability one, the numerator of (4.4) tends pointwise to $\pi(\tau^*) \times \exp\{-z'\ddot{\psi}(\tau^*)z/2\}$.

Turning to the denominator of $g_n(z)$ in (4.4), we note that Berk [2] has shown that the posterior probability of any neighborhood, say U, of τ^* goes to one, with probability one, as $n \to \infty$. Then for almost all sample sequences,

$$\lim_{n\to\infty} \int_U \pi(\tau)\exp\{n[\tau'\phi_n - \psi(\tau)]\}\,d\tau / \int_\Omega \pi(\tau)\exp\{n[\tau'\phi_n - \psi(\tau)]\}\,d\tau = 1. \tag{4.6}$$

Since $\tau_n^* \to \tau^*$ almost surely, it follows that for any $\delta > 0$, there exists a neighborhood U of τ^* such that eventually $U \subseteqq \{\tau : |\tau - \tau_n^*| < \delta\}$. Thus in the denominator of $g_n(z)$, we may replace the region of integration Ω by $\{\tau : |\tau - \tau_n^*| < \delta\}$. Noting that $|\tau = \tau_n^*| < \delta$ corresponds to $|z_n| < \delta$, we have that in ratio, the denominator of $g_n(z)$ in (4.4) is asymptotically,

$$\begin{aligned} &\int_{|z_n|<\delta} \pi(z_n + \tau_n^*)\exp\{n[k_n(z_n) - k_n(0)]\}\,dz \\ &\quad = \int_{|z|<\delta n^{\frac{1}{2}}} \pi(z_n + \tau_n^*)\exp\{-z'\ddot{\psi}(\xi_n + \tau_n^*)z/2\}\,dz, \end{aligned} \tag{4.7}$$

where in the last integrand (4.5) was incorporated.

We define the following modulus. Let

$$\rho(\tau) = \inf_{|z|=1} z'\ddot{\psi}(\tau)z. \tag{4.8}$$

Clearly $z'\ddot{\psi}(\tau)z \geqq \rho(\tau)z'z$. The compactness of the unit shell $\{z : |z| = 1\}$ implies that $\rho(\cdot)$ is continuous. In particular, since $\ddot{\psi}(\tau^*)$ is positive definite, $\rho(\tau^*) > 0$ and therefore $\rho(\cdot) > 0$ on a neighborhood of τ^*. By choosing δ sufficiently small, we then have $r = \inf\{\rho(\tau) : |\tau - \tau^*| < 2\delta\} > 0$. It follows that the integrand in (4.7) is with probability one, eventually bounded above by $h(z)$, where

$$h(z) = [\sup_{|\tau-\tau^*|<2\delta} \pi(\tau)]\exp\{-rz'z/2\} \tag{4.9}$$

is an integrable function. Since pointwise, the integrand in (4.7) approaches $\pi(\tau^*)\exp\{-z'\ddot{\psi}(\tau^*)z/2\}$, it follows from the Lebesgue dominated convergence theorem that with probability one, (4.7) and hence the denominator of $g_n(z)$ approaches $\pi(\tau^*)\int\exp\{-z'\ddot{\psi}(\tau^*)z/2\}\,dz$. Thus we have proven the following.

THEOREM 1. *With probability one,* $g_n(z)\to N(z\,|\,0,[\ddot{\psi}(\tau^*)]^{-1})$, *the multivariate normal* pdf *with zero mean and covariance matrix* $[\ddot{\psi}(\tau^*)]^{-1}$.

Convergence in distribution of $Z = n^{\frac{1}{2}}(\tau-\tau_n{}^*)$ to the $N(0,[\ddot{\psi}(\tau^*)]^{-1})$ distribution follows immediately from Scheffé's theorem. It also follows that $g_n(z)$ tends to the $N(0,[\ddot{\psi}(\tau^*)]^{-1})$ pdf in L_1-norm.

5. Extensions. A similar result holds for the more general exponential model $f(x\,|\,\theta)=\exp\{\alpha(\theta)'\phi(x)-\psi(\alpha(\theta))\}$ where θ ranges in a parameter set Θ. Assuming that Θ^0 is not empty, that the true parameter value $\theta^*\in\Theta^0$, and that $\alpha(\cdot)$ is a diffeomorphism betwen Θ and $\alpha(\Theta)\subseteqq\Omega$, the preceding discussion can be modified to prove the following result:

THEOREM 2. *Let* $\pi(\theta)$ *be a prior* pdf (wrt *Lebesgue measure*) *over the set* Θ *which is positive and continuous at* θ^*, *let* $\theta_n{}^*$ *be the maximum likelihood estimate of* θ. *Then* $g_n(z)$, *the* pdf *of* $Z=n^{\frac{1}{2}}(\theta-\theta_n{}^*)$, *tends pointwise with increasing sample size to the multivariate normal density*

$$\frac{\pi(\theta^*)\exp\{-z'A'(\theta^*)\ddot{\psi}(\alpha(\theta^*))A(\theta^*)z/2\}}{\int\pi(\theta^*)\exp\{-z'A'(\theta^*)\ddot{\psi}(\alpha(\theta^*))A(\theta^*)z/2\}\,dz},$$

where $A(\theta)=(\partial\alpha_i(\theta)/\partial\theta_j)$.

Acknowledgment. The authors wish to thank Professor H. D. Brunk for bringing this problem to their attention, and Professor Robert H. Berk for greatly simplifying the original version of this article.

REFERENCES

[1] APOSTLE, T. M. (1957). *Mathematical Analysis.* Addison-Wesley, Reading.

[2] BERK, R. H. (1970). Consistency a posteriori. *Ann. Math. Statist.* **41** 894–906.

[3] BERK, R. H. (1972). Consistency and asymptotic normality of MLE's for exponential models. *Ann. Math. Statist.* **43** 193–204.

[4] BUEHLER, R. J. (1965). The limit of the *n*th power of a density. *Ann. Math. Statist.* **36** 1878–1882.

[5] DAWID, A. P. (1970). On the limiting normality of posterior distributions. *Proc. Cambridge Philos. Soc.* **67** 625–633.

[6] JOHNSON, R. A. (1967 a). Asymptotic expansions associated with the *n*th power of a density. *Ann. Math. Statist.* **38** 1266–1272.

[7] JOHNSON, R. A. (1967 b). An asymptotic expansion for posterior distributions. *Ann. Math. Statist.* **38** 1899–1907.

[8] LECAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. California Publ. Statist.* **1** 277–330.

[9] LINDLEY, D. V. (1965). *Introduction to Probability and Statistics, Part 2. Inference.* Cambridge Univ. Press.

[10] WALKER, A. M. (1969). On the asymptotic behavior of posterior distributions. *J. Roy. Statist. Soc. Ser. B* **31** 80–88.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OKLAHOMA
601 ELM AVENUE, ROOM 423
NORMAN, OKLAHOMA 73069

DEPARTMENT OF MATHEMATICS
WEST CHESTER STATE COLLEGE
WEST CHESTER, PENNSYLVANIA 19380