The Annals of Statistics
1988, Vol. 16, No. 1, 356–366

ASYMPTOTIC BEHAVIOR OF LIKELIHOOD METHODS FOR EXPONENTIAL FAMILIES WHEN THE NUMBER OF PARAMETERS TENDS TO INFINITY[1]

BY STEPHEN PORTNOY

University of Illinois

Consider a sample of size n from a regular exponential family in p_n dimensions. Let $\hat{\theta}_n$ denote the maximum likelihood estimator, and consider the case where p_n tends to infinity with n and where $\{\theta_n\}$ is a sequence of parameter values in R^{p_n}. Moment conditions are provided under which $\|\hat{\theta}_n - \theta_n\| = O_p(\sqrt{p_n/n})$ and $\|\hat{\theta}_n - \theta_n - \overline{X}_n\| = O_p\ (p_n/n)$, where $\overline{X}_n$ is the sample mean. The latter result provides normal approximation results when $p_n^2/n \to 0$. It is shown by example that even for a single coordinate of $(\hat{\theta}_n - \theta_n)$, $p_n^2/n \to 0$ may be needed for normal approximation. However, if $p_n^{3/2}/n \to 0$, the likelihood ratio test statistic Λ for a simple hypothesis has a chi-square approximation in the sense that $(-2\log\Lambda - p_n)/\sqrt{2p_n} \to_D \mathcal{N}(0,1)$.

1. Introduction. Most statistical procedures depend heavily on asymptotic methods either for inferential purposes or for justification, particularly in nonnormal cases. These methods generally rely on the central limit theorem in the parameter space and provide good approximations for remarkably small sample sizes when the dimension of the parameter space is small. However, in all but the simplest problems, some models with a relatively large number of parameters are considered, particularly when the sample size is large. Thus, asymptotic results which permit the number of parameters p to grow with the sample size n are needed. For example, in a linear regression problem with five independent variables, a sample size $n = 100$ might be considered adequate. However, consideration of a quadratic model requires $p = 21$ parameters; for p^2/n to be small (a condition often required for adequate approximation), n must be much larger than 500 observations.

The fundamental question for applied statistics is how large can p be (compared to n) so that asymptotic distributional approximations for maximum likelihood estimators and likelihood ratio tests may be accepted as reliable. The basic conclusion presented here for exponential families is that asymptotic approximations are trustworthy generally if $p^{3/2}/n$ is small [and EA_n in condition (3.10) is small], but otherwise may be in substantial error, particularly if p^2/n is not small. The results here apply directly to such applications as contingency tables (or more general multinomial situations), nonlinear problems in multivariate normal situations, Markov chains and so forth, although I believe that the basic conclusion should hold for arbitrary smooth distributional situa-

Received October 1986; revised April 1987.

[1]Research partially supported by NSF Grants MCS-83-01834 and DMS-85-03785.

AMS 1980 *subject classifications.* Primary 62E20; secondary 60F05, 62F10.

Key words and phrases. Asymptotics, maximum likelihood, central limit theorem, exponential family.

tions. In particular, if p^2/n is not small (or if $p^{3/2}/n$ is not small and one wants to be safe) distributions of maximum likelihood estimators or likelihood ratio test statistics must be found using alternative methods, for example, Monte Carlo simulations or more extensive asymptotic expansions (which at present are not generally available unless p is very small compared to n). One positive result presented here is that maximum likelihood estimators will tend to be (asymptotically) consistent if $p/n \to 0$, but the important message here is a word of caution about applying standard asymptotic approximations when p is not small.

The general framework for the results assumes a sequence of problems with observation spaces $\mathscr{X}_n$, parameter spaces $\Theta_n \subset R^{p_n}$ with $p_n \to \infty$ and distributions $Q_\theta^{(n)}$ for $\theta_n \in \Theta_n$. Since Θ_n differ, the true parameter value cannot be fixed, but a sequence $\{\theta_n \in \Theta_n\colon n = 1, 2, \dots\}$ must be considered and strong laws are precluded. However, versions of weak laws can often provide conditions under which estimators $\{\hat{\theta}_n\}$ satisfy $\|\hat{\theta}_n - \theta_n\| \to_P 0$ (for some appropriate norm on Θ_n). Similarly, normal approximation results (particularly in the classical row-wise independent, triangular array situation) will often provide a normal approximation for functions of $(\hat{\theta}_n - \theta_n)$ or a uniform normal approximation for $S_n^{-1/2}(\hat{\theta}_n - \theta_n)$ in R^{p_n} [where $S_n = \operatorname{Cov}(\hat{\theta}_n)$] in the sense that

$$\left|P\left\{S_n^{-1/2}\left(\hat{\theta}_n - \theta_n\right) \in A_n\right\} - \Phi(A_n)\right| \to 0$$

uniformly over appropriate sets $A_n \subset R^{p_n}$ (with Φ the standard normal distribution in R^{p_n}). Some examples of such results are the following.

For using the chi-square test for testing a simple hypothesis for a multinomial distribution with p_n cells, results of Morris (1975) and Koehler and Larntz (1982), among others, provide conditions under which chi-square or normal approximations can be used when $p_n \to \infty$. For the most general situations, these results appear to require $p_n^2/n \to 0$, but in many cases $p_n^{3/2}/n \to 0$ may suffice. For example, Theorem 5.2 of Morris [(1975), page 183] gives a normal approximation for the difference of the likelihood ratio test statistic and a bias term B_n. Let $\{\theta_{nj}\colon j = 1, \dots, p_n\}$ denote the cell probabilities. If $\{\theta_{nj}\}$ lie away from the boundary of the simplex in the sense that $\sum_{j=1}^{p_n}(1/(p_n^2\theta_{nj}))$ converges to a nonzero constant, then $B_n = O(p_n^{3/2}/n)$, and so $p_n^{3/2}/n \to 0$ suffices.

Some related results obtaining consistency and asymptotic normality in regression settings are presented in Huber (1973), Yohai and Moronna (1979), Ringland (1983) and Portnoy (1984, 1985, 1986a, b and 1987).

The specific framework here takes $X_1, \dots, X_n$ i.i.d. according to a regular exponential family distribution $P_\theta^{(n)}$ in R^p with canonical form density

$$P_\theta^{(n)}(x) = \exp\{-\psi_n(\theta) + \theta' x\}, \qquad \theta \in \Theta_n, \tag{1.1}$$

with respect to some sigma-finite dominating measure, where Θ_n is the natural parameter space (assumed here to be a nonempty set). From now on, we will generally suppress the subscript n on the dimension p and will let $\|\cdot\|$ denote the usual Euclidean norm in R^p. As is well known, the maximum likelihood estimator $\hat{\Theta}_n$ satisfies

$$\psi_n'(\hat{\theta}_n) = \overline{X}, \tag{1.2}$$

where $\overline{X} \equiv \overline{X}_n = \sum_{i=1}^n X_i$ and ψ_n' denotes the gradient of ψ_n. When necessary, the coordinates of X_i will be denoted using a second subscript $j = 1, \dots, p$ (X_{ij}).

The results here depend on asymptotic (Taylor series) expansions and on central limit theorems. The basic expansions are given in Section 2 and hold under reasonable conditions since derivatives of $\psi(\theta)$ can be expressed in terms of expectations. These expansions should hold for general regular families, but the conditions would be extremely unwieldy and artificial. Expansions and asymptotic normality results are given in Section 3. For general approximation in R^p, a central limit theorem for $\overline{X}$ is given in Portnoy (1986b) (which shows that $p^2/n \to 0$ is essentially necessary and sufficient for general normality). However, an asymptotic approximation for the likelihood ratio test statistic for testing a simple hypothesis is also given which holds if $p^{3/2}/n \to 0$. This result is based on a normal approximation for $(\|\overline{X} - EX\|^2 - p)/\sqrt{2p}$ which is derived in Section 4 from a martingale central limit theorem and is of substantial interest on its own.

2. Consistency. The asymptotic results require expansions for the logarithm of the likelihood function, which in this case only involves $\psi(\theta)$. Since $\psi(\theta)$ is a cumulant generating function, its derivatives can be expressed in terms of moments, thus giving the following straightforward Taylor series expansions.

PROPOSITION 2.1. *Let $\psi_n'(\theta)$ denote the gradient and $\psi_n''(\theta)$ denote the Hessian of ψ_n. For each n and any θ and θ_0 in Θ_n, the following three expansions hold for some $\tilde{\theta}$ between θ and θ_0 in Θ_n:*

$$\begin{aligned}\psi_n(\theta) = {} & \psi_n(\theta_0) + (\theta - \theta_0)'\psi_n'(\theta_0) + \tfrac{1}{2}(\theta - \theta_0)'\psi_n''(\theta_0)(\theta - \theta_0) \\ & + \tfrac{1}{6}E_{\theta_0}((\theta - \theta_0)'U)^3 \\ & + \tfrac{1}{24}\left\{E_{\tilde{\theta}}((\theta - \theta_0)'U)^4 - 3\left[E_{\tilde{\theta}}((\theta - \theta_0)'U)^2\right]^2\right\},\end{aligned} \tag{2.1}$$

$$\begin{aligned}\psi_n'(\theta) = {} & \psi_n'(\theta_0) + (\theta - \theta_0)'\psi_n''(\theta_0) + \tfrac{1}{2}E_{\theta_0}((\theta - \theta_0)'U)^2 U \\ & + \tfrac{1}{6}E_{\tilde{\theta}}((\theta - \theta_0)'U)^3 U \\ & - \tfrac{1}{2}E_{\tilde{\theta}}((\theta - \theta_0)'U)^2 E_{\tilde{\theta}}((\theta - \theta_0)'U)U,\end{aligned} \tag{2.2}$$

$$\psi_n'(\theta) = \psi_n'(\theta_0) + (\theta - \theta_0)'\psi_n''(\theta_0) + \tfrac{1}{2}E_{\tilde{\theta}}((\theta - \theta_0)'U)^2 U, \tag{2.3}$$

where $E_\theta f(U)$ denotes the expectation of $f(V - E_\theta V)$ with $V \sim P_\theta$.

THEOREM 2.1. *Let $\underline{\lambda}(\theta)$ be the minimum eigenvalue of $\psi''(\theta)$ and let $\overline{\lambda}(\theta)$ be the maximum eigenvalue. For $\theta_n \in \Theta_n$, define B_n and b_n by*

$$B_n = \overline{\lambda}^{1/2}(\theta_n)/\underline{\lambda}(\theta_n), \qquad b_n = \sqrt{p/n}\,B_n.$$

Let $\theta_n \in \Theta_n$ satisfy $B_n = O(1)$ and

$$\sup\left\{|E_\theta(a'U)^3|: \|\theta - \theta_n\| \le (1.2)b_n, \|a_n\| = 1\right\} \le (0.1)\sqrt{n/p}\,\underline{\lambda}^2(\theta_n)/\bar{\lambda}^{1/2}(\theta_n). \tag{2.4}$$

Then the maximum likelihood estimators $\{\hat{\theta}_n\}$ are norm consistent in the sense that

$$\|\hat{\theta}_n - \theta_n\| = O_p\left(\sqrt{p/n}\right). \tag{2.5}$$

PROOF. Let $F(\theta) = \psi'(\theta) - \bar{X}$. The maximum likelihood estimators uniquely satisfy $F(\hat{\theta}) = 0$. By Theorem 6.3.4 of Ortega and Rheinboldt [(1970), page 163], if $(\theta - \theta_n)'F(\theta) \ge 0$ for all $\theta \in \Theta_n$ satisfying $\|\theta - \theta_n\| = (1.2)b_n$, then there is a root of $F(\theta) = 0$ in $\|\theta - \theta_n\| \le (1.2)b_n$, that is, (2.5) would follow. By (2.3), we have [with $\mu = E_{\theta_n}X = \psi'(\theta_n)$],

$$\begin{aligned}(\theta - \theta_n)'F(\theta) &= (\theta - \theta_n)'\psi_n'(\theta_n) + (\theta - \theta_n)'\psi_n''(\theta_n)(\theta - \theta_n)\\ &\quad + \tfrac{1}{2}E_{\tilde{\theta}}((\theta - \theta_n)'U)^3 - (\theta - \theta_n)'\bar{X}\\ &= -(\theta - \theta_n)'(\bar{X} - \mu) + (\theta - \theta_n)'\psi_n''(\theta_n)(\theta - \theta_n)\\ &\quad + \tfrac{1}{2}E_{\tilde{\theta}}((\theta - \theta_n)'U)^3.\end{aligned} \tag{2.6}$$

From Theorem 4.1 [formula (4.11)], $\{(\bar{X} - \mu)'(\psi_n''(\theta_n))^{-1}(\bar{X} - \mu)\}^{1/2} = \sqrt{p/n}$ $(1 + o_p(1))$. Hence, using (2.4), for $\|\theta - \theta_n\| = (1.2)b_n$ with probability tending to 1,

$$\begin{aligned}(\theta - \theta_n)'F(\theta) &\ge \underline{\lambda}(\theta_n)\|\theta - \theta_n\|^2 - (1.05)\sqrt{p/n}\,\bar{\lambda}^{1/2}(\theta_n)\|\theta - \theta_n\|\\ &\quad - (0.1)\sqrt{n/p}\,\underline{\lambda}^2(\theta_n)/\underline{\lambda}^{1/2}(\theta_n)\|\theta - \theta_n\|^3\\ &= (p/n)\bar{\lambda}(\theta_n)/\underline{\lambda}(\theta_n)\left((1.2)^2 - (1.05)(1.2) - 0.1(1.2)^3\right)\\ &> 0.\end{aligned} \tag{2.7}$$

Hence, the result follows by the previously cited theorem of Ortega and Rheinboldt. □

REMARK. Since the parameter space is not fixed, there is no way to obtain consistency at every sequence $\theta_n \in \Theta_n$ in general. However, in practice, the true parameter value would tend not to be very extreme. Since the right-hand side of (2.4) would generally tend to infinity (if $p/n \to 0$), one could expect (2.4) to hold for the parameter values of interest.

3. Asymptotic normality. In this section, we will fix the sequence $\{\theta_n\}$ and consider the asymptotic distribution of $\sqrt{n}(\psi''(\theta_n))^{1/2}(\hat{\theta}_n - \theta_n)$ by relating it to that of $\sqrt{n}(\psi''(\theta_n))^{-1/2}(\bar{X}_n - \mu_n)$. Since affine transformations take exponential families into exponential families, without loss of generality the notation can be considerably simplified by taking

$$\mu_n = EX = \psi'(\theta_n) = 0, \qquad \operatorname{Cov}(X) = \psi''(\theta_n) = I. \tag{3.1}$$

The first result bounds $\|(\hat{\theta}_n - \theta_n) - \bar{X}\|$ so that the asymptotic distribution of $(\hat{\theta}_n - \theta_n)$ can be derived from central limit theorem results for $\bar{X}$.

THEOREM 3.1. *Suppose* (2.5) *holds so that* $\|\hat{\theta}_n - \theta_n\|^2 \le cp/n$ *in probability and suppose that for some constant* B,

$$\sup\{E_\theta|a'U|^4 : \|a\| = 1,\ \|\theta - \theta_n\|^2 \le cp/n\} \le B, \tag{3.2}$$
$$\text{where } U = V - E_\theta V \text{ with } V \sim P_\theta.$$

Note that by Hölder's inequality, (3.2) *implies*

$$\sup\{E_\theta|a'U|^{k_1}|b'U|^{k_2} : \|a\| = 1,\ \|b\| = 1,\ \|\theta - \theta_n\|^2 \le cp/n\} \le B \tag{3.3}$$

for k_1 *and* k_2 *positive integers with* $k_1 + k_2 \le 4$. *Then, if* $p/n \to 0$,

$$\|(\hat{\theta} - \theta_n) - \bar{X}\| = O_p(p/n).$$

PROOF. Since $\psi'(\hat{\theta}_n) = \bar{X}$, (3.1) and expansion (2.2) yield

$$(\hat{\theta}_n - \theta_n) = \bar{X} - \tfrac{1}{2}E_{\theta_n}((\theta - \theta_n)'U)^2 U + \Delta, \tag{3.4}$$

where Δ is the last two terms of (2.2). From (3.3),

$$\begin{aligned} |(\hat{\theta}_n - \theta_n)'\Delta| &\le \tfrac{1}{6}E_{\tilde{\theta}}((\hat{\theta}_n - \theta_n)'U)^4 + \tfrac{1}{2}\{E_{\tilde{\theta}}((\hat{\theta}_n - \theta_n)'U)^2\}^2 \\ &\le B\|\hat{\theta}_n - \theta_n\|^4 = O_p(p^2/n^2), \end{aligned}$$

(since $\|\tilde{\theta} - \theta_n\|^2 \le \|\hat{\theta}_n - \theta_n\|^2 \le cp/n$ in probability). Similarly, $|\bar{X}'\Delta| = O_p(p^2/n^2)$; hence,

$$\|\hat{\theta}_n - \theta_n\|^2 = (\hat{\theta}_n - \theta_n)'X - \tfrac{1}{2}E_{\theta_n}((\hat{\theta}_n - \theta_n)'U)^3 + O_p(p^2/n^2), \tag{3.5}$$

$$(\hat{\theta}_n - \theta_n)'\bar{X} = \|\bar{X}\|^2 - \tfrac{1}{2}E_{\theta_n}((\hat{\theta}_n - \theta_n)'U)^2(\bar{X}'U) + O_p(p^2/n^2). \tag{3.6}$$

Subtracting (3.6) from (3.5),

$$\begin{aligned} \|(\hat{\theta}_n - \theta_n) - \bar{X}\|^2 &= -\tfrac{1}{2}E_{\theta_n}(U'(\hat{\theta}_n - \theta_n))^2 U'((\hat{\theta}_n - \theta_n) - \bar{X}) + O_p(p^2/n^2) \\ &\le \tfrac{1}{2}\{E_{\theta_n}(U'(\hat{\theta}_n - \theta_n))^4\}^{1/2}\{E_{\theta_n}(U'((\hat{\theta}_n - \theta_n) - \bar{X}))^2\}^{1/2} \\ &\quad + O_p(p^2/n^2) \\ &\le \|(\hat{\theta}_n - \theta_n) - \bar{X}\|A_n + B_n \end{aligned} \tag{3.7}$$

by condition (3.3), where $A_n = O_p(p/n)$ and $B_n = O_p(p^2/n^2)$. Solving this inequality yields the result. □

Theorem 3.1 immediately provides asymptotic approximations for many functions of interest. In particular, assume $f\colon R^p \to R$ is such that $\|f'(x)\|$ is uniformly bounded (in x and p). Then from Theorem 3.1,

$$f(\sqrt{n}(\hat{\theta}_n - \theta_n)) = f(\sqrt{n}\bar{X}) + O_p(p/\sqrt{n}). \tag{3.8}$$

Portnoy (1986b) shows that if $p^2/n \to 0$, $f(\sqrt{n}\,\overline{X})$ can be approximated by assuming $\sqrt{n}\,\overline{X} \sim \mathscr{N}_p(0, I)$ (and that no faster rate will work in complete generality). Thus, normal approximation will tend to hold if $p^2/n \to 0$. It is sometimes possible to obtain a better sufficient condition for normal approximation by carefully considering error terms in (3.4). For example, if $\|a_n\| = 1$, $\sqrt{n}\,a_n'\overline{X}_n \to \mathscr{N}(0,1)$ as long as $n \to \infty$ with no condition on p. Proposition 3.1 gives an error term for $a_n'(\hat{\theta}_n - \theta_n)$ which will often be of smaller order than p^2/n. It follows directly from (3.4) and Theorems 2.1 and 3.1.

PROPOSITION 3.1. *If* $\|a_n\| = 1$ *and* (3.2) *holds, then*

$$\sqrt{n}\,a_n'\big(\hat{\theta}_n - \theta_n\big) = \sqrt{n}\,a_n'\overline{X}_n - \frac{\sqrt{n}}{2}E_{\theta_n}\big(\overline{X}_n'U\big)^2(a_n'U) + O_p\bigg(\frac{p^{3/2}}{n}\bigg). \tag{3.9}$$

Now let $A_n = A_n(\overline{X}) = \sqrt{n}\,E_{\theta_n}(U'\overline{X})^2(a_n'U)$. Then straightforward calculations (see the Appendix) show that

$$EA_n = \frac{1}{\sqrt{n}}E_{\theta_n}\|U\|^2(a_n'U), \qquad \operatorname{Var} A_n = O\bigg(\frac{p}{n}\bigg). \tag{3.10}$$

Thus, $\sqrt{n}\,a_n'(\hat{\theta}_n - \theta_n)$ will be asymptotically normal if $p^{3/2}/n \to 0$ and $EA_n \to 0$. This later convergence will often hold, but the following example shows that EA_n may actually be of order $p/\sqrt{n}$.

Let Y be a random variable (bounded) with $EY = 0$, $EY^2 = 1$ and $EY^3 = 1$. Let $U = (Y, YZ_1, \dots, YZ_{p-1})$, where $(Z_1, \dots, Z_{p-1})' \sim \mathscr{N}_{p-1}(0, I)$. Let $a_n = (1, 0, 0, \dots, 0)'$. Then $EU_j^2U_1 = EY^3Z_j^2 = 1$ so that

$$EA_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{p} EU_j^2U_1 = \frac{p}{\sqrt{n}}.$$

Note that the conditions for Theorem 3.1 and Proposition 3.1 hold for this example.

Last, we show that the likelihood ratio test statistic for testing a simple hypothesis can be approximated using the normal approximation for χ_p^2 as long as $p^{3/2}/n \to 0$.

THEOREM 3.2. *Let* Λ *be the likelihood ratio test statistic for testing* H_0: $\theta = \theta_n$ *versus* H_1: $\theta \neq \theta_n$. *Assume the conditions for Theorem* 3.1. *Then, if* $p^{3/2}/n \to 0$,

$$(-2\log\Lambda - p)/\sqrt{2p} = \big(n\|\overline{X}\|^2 - p\big)/\sqrt{2p} + o_p(1). \tag{3.11}$$

Thus, the left side of (3.11) *is approximately* $\mathscr{N}(0,1)$ *by Theorem* 4.1.

PROOF. By definition of Λ and (2.1),

$$\begin{aligned} -2\log\Lambda &= 2n\big\{\big(\hat{\theta}_n - \theta_n\big)'\overline{X} - \big(\psi\big(\hat{\theta}_n\big) - \psi(\theta_n)\big)\big\} \\ &= n\bigg\{\|\overline{X}\|^2 - \|\overline{X} - \big(\hat{\theta}_n - \theta_n\big)\|^2 + \frac{1}{6}E_{\theta_n}\big(\big(\hat{\theta}_n - \theta_n\big)'U\big)^3\bigg\} + O_p\bigg(\frac{p^2}{n}\bigg), \end{aligned} \tag{3.12}$$

where (3.2) is used as in the proof of Theorem 3.1.

Now, using (3.2) again (with E denoting E_{θ_n}),

$$E\big((\hat{\theta}_n-\theta)'U\big)^3 = E(\bar{X}'U)^3 + 3E(X'U)^2(\hat{\theta}_n-\theta_n-\bar{X})'U \tag{3.13}$$
$$+3E(X'U)\big((\hat{\theta}_n-\theta_n-\bar{X})'U\big)^2 + E\big((\hat{\theta}_n-\theta_n-\bar{X})'U\big)^3$$
$$= E(\bar{X}'U)^3 + O_p\left(\left(\frac{p}{n}\right)^2\right) + O_p\left(\left(\frac{p}{n}\right)^{5/2}\right) + O_p\left(\left(\frac{p}{n}\right)^3\right).$$

From Proposition A.2, $E^{\bar{X}}(E\bar{X}'U)^3)^2 = O(p^{5/2}/n^3 + p^3/n^4)$ (where $E^{\bar{X}}$ denotes expectation over $\bar{X}$). Hence,

$$E(\bar{X}'U)^3 = O_p(p^{5/4}/n^{3/2} + p^{3/2}/n^2). \tag{3.14}$$

The result follows by combining (3.12)–(3.14). □

Note that [since $\psi''(\theta_n)$ is assumed to equal I] it is possible to test H_0 using the statistic $\|\hat{\theta}_n - \theta_n\|^2$. As in the proof of Theorem 3.2, it can be shown that

$$\big(n\|\hat{\theta}_n-\theta_n\|^2 - p\big)/\sqrt{2p} = \big(n\|X\|^2 - p\big)/\sqrt{2p} + o_p(1),$$

if $p^{3/2}/n \to 0$; hence, the same normal approximation holds.

4. Asymptotic behavior of $\|\bar{X}\|^2$. Let $X_1,\ldots,X_n$ be i.i.d. vectors in R^p with $EX_1 = 0$, $\mathrm{Cov}(X_1) = I$ and $EX_{1j}^6 \le B < +\infty$ (for $j=1,\ldots,p$), where the basic "triangular array" situation is being considered but the coordinate index (j) will be suppressed for convenience. The following result shows that the asymptotic behavior of $\|\bar{X}_n\|^2$ follows from a martingale central limit theorem.

THEOREM 4.1. *Define $T_n \in R^p$ and S_n by*

$$T_n = \sum_{i=1}^{n} X_i, \qquad S_n = T_n'T_n - np = \|T_n\|^2 - np. \tag{4.1}$$

Then, if $p/n \to 0$,

$$\big(n\|\bar{X}\|^2 - p\big)/\sqrt{2p} = S_n/\big(n\sqrt{2p}\big) \to_D \mathcal{N}(0,1). \tag{4.2}$$

PROOF. Let $\mathscr{F}_n$ be the sigma-field $\mathscr{S}(X_1,\ldots,X_n) = \mathscr{S}(T_1,\ldots,T_n)$. That S_n is an $\mathscr{F}_n$-martingale follows directly from

$$S_n = (T_{n-1}+X_n)'(T_{n-1}+X_n) - np \tag{4.3}$$
$$= \|T_{n-1}\|^2 - (n-1)p + 2X_n'T_{n-1} + \|X_n\|^2 - p.$$

Now consider the martingale difference sequence

$$D_n = S_n - S_{n-1} = 2X_n'T_{n-1} + \|X_n\|^2 - p. \tag{4.4}$$

Define $\sigma_i^2 = ED_i^2$ and $s_n^2 = \sum_{i=1}^n \sigma_i^2$.

To apply the martingale central limit theorem of Chow and Teicher [(1978), equation (8), page 316], it suffices to show that if $p/n \to 0$, then both

$$\sum_{i=1}^{n} E|D_i|^3/s_n^3 \to 0 \tag{4.5}$$

and

$$\sum_{i=1}^{n} E|E(D_i^2|\mathscr{F}_{i-1}) - \sigma_i^2|/s_n^2 \to 0. \tag{4.6}$$

First note that

$$\begin{aligned}\sigma_i^2 = ED_i^2 &= 4ET_{i-1}'X_iX_i'T_{i-1} \\ &\quad + 4EX_i'T_{i-1}(\|X_i\|^2 - p) + E(\|X_i\|^2 - p)^2 \\ &= 4E\|T_{i-1}\|^2 + 0 + O(p^2) \\ &= 4(i-1)p + O(p^2)\end{aligned} \tag{4.7}$$

since

$$E\|X_i\|^{2k} \le p^k B \quad \text{for } k \le 3 \ (\text{since } EX_{1j}^6 \le B). \tag{4.8}$$

Therefore,

$$s_n^2 = \sum_{i=1}^{n} \sigma_i^2 = 2n(n-1)p + O(np^2) = 2n^2p\left(1 + O\left(\frac{p}{n}\right)\right). \tag{4.9}$$

To prove (4.5), by (4.4)

$$E|D_i|^3 < 8\left(E(X_i'T_{i-1})^6\right)^{1/2} + 8p^3.$$

From the Appendix (A.17) for some constant c,

$$E(X_i'T_{i-1})^6 \le c(i^3p^3 + i^2p^5 + ip^6).$$

Hence, for some c' and c'',

$$\begin{aligned}\sum_{i=1}^{n} E|D_i|^3 &\le c' \sum_{i=1}^{n} (i^{3/2}p^{3/2} + ip^{5/2} + \sqrt{i}\,p^3 + p^3) \\ &\le c''(n^{5/2}p^{3/2} + n^2p^{5/2} + n^{3/2}p^3).\end{aligned}$$

Therefore, from (4.9), if $p/n \to 0$,

$$\sum_{i=1}^{n} E|D_i|^3/s_n^3 \le c'''(1/\sqrt{n} + p/n + p^{3/2}/n^{3/2}) \to 0$$

and (4.5) follows.

For (4.6), as in (4.7),

$$E[D_i^2|\mathscr{F}_{i-1}] = 4\|T_{i-1}\|^2 + 4T_{i-1}EX_i(\|X_i\|^2 - p) + O(p^2).$$

Hence, from (4.7),

$$\begin{aligned}E|E[D_i^2|\mathscr{F}_{i-1}] - \sigma_i^2| &\le \left\{E\left(E[D_i^2|\mathscr{F}_{i-1}] - \sigma_i^2\right)^2\right\}^{1/2} \\ &\le c\left\{E\left(\|T_{i-1}\|^2 - (i-1)p\right)^2 + \left(EX_i(\|X_i\|^2 - p)\right)\right. \\ &\qquad \left.\times ET_{i-1}'T_{i-1}\left(EX_i'(\|X_i\|^2 - p)\right) + O(p^4)\right\}^{1/2}.\end{aligned} \tag{4.10}$$

But $ET'_{i-1}T_{i-1} = iI$ and [from (4.8)]

$$\|EX_i(\|X_i\|^2 - p)\|^2 = E\|X_i\|^2(\|X_i\|^2 - p)^2 = O(p^3).$$

Therefore, from (4.10),

$$E|E[D_i^2|\mathscr{F}_{i-1}] - \sigma_i^2| \le c'\{2i^2p + ip^3 + p^4\}^{1/2}.$$

Hence, using (4.9),

$$\sum_{i=1}^{n} E|E[D_i^2|\mathscr{F}_{i-1}] - \sigma_i^2|/s_n^2 \le c'''(n^2p^{1/2} + n^{3/2}p^{3/2} + np^2)/(n^2p) \to 0$$

if $p/n \to 0$ and (4.6) follows. Thus, from the Chow–Teicher theorem,

$$S_n/s_n = S_n/(n\sqrt{2p})(1 + o(1)) \to_D \mathscr{N}(0,1). \qquad \square$$

Last, note that (4.2) immediately yields the result that

$$\|\overline{X}\|^2 = p/n + O_p(\sqrt{p}/n) = (p/n)(1 + o_p(1)). \tag{4.11}$$

APPENDIX

Let $X_1, \ldots, X_n$ be i.i.d. in R^p with $EX = 0$ and $\mathrm{Cov}(X) = I$. The computations here all involve expectations of the form

$$E(a'\overline{X})^k = \frac{1}{n^k}\sum_{i_1} \cdots \sum_{i_k}\sum_{j_1} \cdots \sum_{j_k} a_{j_1} \cdots a_{j_k} EX_{i_1j_1} \cdots X_{i_kj_k}.$$

Since $EX = 0$, subscripts i_ν and i_μ must be equal at least in pairs. Furthermore, since $\mathrm{Cov}(X) = I$, for a pair of equal subscripts which differ from all other i-subscripts, the double sum over j_ν and j_μ reduces to a single sum over $j_\mu = j_\nu = j$. This argument is used in each of the following results.

PROPOSITION A.1. *Define for* $\|a\| = 1$ *and* $U \sim X$ *(independently),*

$$A_n = \sqrt{n}\,E_0(U'\overline{X})^2(a'U), \tag{A.1}$$

where E_0 *denotes expectation over the distribution of* U. *Suppose, for some constant* B,

$$\sup\{E_0|a'U|^3 : \|a\| = 1\} \le B. \tag{A.2}$$

Then

$$EA_n = \frac{1}{\sqrt{n}}E_0\|U\|^2(a'U) \quad \textit{and} \quad \mathrm{Var}\,A_n \le \frac{2pB^2}{n}\left(2 + \frac{p}{n}\right). \tag{A.3}$$

PROOF. As previously noted, since $EX_i = 0$, $\mathrm{Cov}(X_i) = I$,

$$EA_n = \frac{\sqrt{n}}{n^2}E_0\sum_{i_1}\sum_{i_2}\sum_{j_1}\sum_{j_2} U_{j_1}U_{j_2}(a'U)EX_{i_1j_1}X_{i_2j_2}$$

$$= \frac{\sqrt{n}}{n^2}\sum_i E_0 \sum_j U_j^2(a'U) = \frac{1}{\sqrt{n}}E_0\|U\|^2(a'U).$$

Note that by (A.2) and Hölder's inequality,

$$(\text{A.4})\quad \left|E_0\|U\|^2(a'U)\right| \le \sum_j E_0 U_j^2|a'U| \le p\left(E_0|U_1|^3\right)^{2/3}\left(E_0|a'U|^3\right)^{1/3} \le pB.$$

Similarly, with $V \sim U$ independently,

$$\begin{aligned} EA_n^2 &= \frac{1}{n^3}E_0\sum\cdots\sum U_{j_1}U_{j_2}V_{j_1}V_{j_2}(a'U)(a'V)EX_{i_1j_1}X_{i_2j_2}X_{i_3j_3}X_{i_4j_4} \\ &= \frac{1}{n^3}E_0\sum_{i\neq l}\sum\sum_j\sum_k U_j^2V_k^2(a'U)(b'V) \qquad (i_1 = i_2,\, i_3 = i_4) \\ (\text{A.5})\qquad &\quad + \frac{2}{n^3}\sum_{i\neq l}\sum\sum_j\sum_k U_jV_jU_kV_k(a'U)(a'V) \qquad \begin{pmatrix} i_1 = i_3, & i_2 = i_4 \\ i_1 = i_4, & i_2 = i_3 \end{pmatrix} \\ &\quad + \frac{1}{n^3}E_0\sum_i\sum_{j_1\cdots j_4}\cdots\sum U_{j_1}U_{j_2}V_{j_3}V_{j_4}(a'U)(a'V) \qquad (i_1 = i_2 = i_3 = i_4). \end{aligned}$$

Subtracting $(EA_n)^2$ from the first term in (A.5) gives

$$(\text{A.6})\qquad (1/n - (n-1)/n^2)\left(E_0\|U\|^2(a'U)\right)^2 \le (Bp)^2/n^2$$

by (A.4). Using (A.2) for the expectation over V, the second term in (A.5) is bounded by

$$(\text{A.7})\qquad \begin{aligned} (2(n-1)/n^2)|E_0(U'V)^2(a'U)(a'V)| &\le 2B|E_0\|U\|^2\|a\|(a'U)|/n \\ &\le 2B^2p/n, \end{aligned}$$

again by (A.4). Similarly, if $e = (1,\dots,1)'$, the last term in (A.5) is bounded by

$$(\text{A.8})\qquad \frac{1}{n^2}\left(E_0\left(\sum_j U_j\right)^2(a'U)\right)^2 = \frac{1}{n^2}\left(E_0(e'U)^2(a'U)\right)^2 \le \frac{(pB)^2}{n^2}.$$

Thus, the bound on $\operatorname{Var} A_n$ follows from (A.5)–(A.8). □

PROPOSITION A.2. *Under the hypotheses of Proposition* A.1,

$$(\text{A.9})\qquad E\left(E_0(U'\overline{X})^3\right)^2 \le c\left(p^{5/2}/n^3 + p^3/n^4\right)$$

for some constant c if $EX_{ij}^6 \le B$ for all $j = 1,\dots,p$.

PROOF. Expanding as in (A.5) and summing,

$$\begin{aligned} &E\left(E_0(U'\overline{X})^3\right)^2 = \\ (\text{A.10})\qquad &(c/n^3)\left(E_0\|U\|^2\|V\|^2U'V + E_0|U'V|^3\right) \\ (\text{A.11})\qquad &\quad + (c/n^4)EE_0|U'X|^3|V'X|^3 \\ (\text{A.12})\qquad &\quad + (c/n^4)E\left(E_0\|U\|^2|U'X|\,|V'X|^3 + E_0|U'V|(U'X)^2(V'X)^2\right) \\ (\text{A.13})\qquad &\quad + (c/n^5)EE_0|U'X|^3|V'X|^3. \end{aligned}$$

For (A.10), from (A.4) (for expectation over V) and Hölder's inequality,

$$E_0\|U\|^2\|V\|^2|U'V| \le pBE\|U\|^3 \le B'p^{5/2} \tag{A.14}$$

and from (A.2),

$$E_0|U'V|^3 \le BE_0\|U\|^3 \le cp^{3/2}. \tag{A.15}$$

Therefore, (A.10) contributes the first term in (A.9). From (A.15), (A.11) is bounded by $(c/n^4)E\|X\|^6 \le cp^3/n^4$, which gives the second term in (A.9). Similarly [using (A.2) and (A.4)], terms (A.12) and (A.13) are also bounded by the second term in (A.9). □

Last, note that Proposition A.3, which is required for Theorem 4.1, does not need condition (A.2). The proof is similar to the preceding proof.

PROPOSITION A.3. *Let* $T_{n-1} = \sum_{i=1}^{n-1} X_i$ *and suppose*

$$EX_{ij}^6 \le B \quad \text{for } j = 1, \dots, p. \tag{A.16}$$

Then for some constant c,

$$E(X_n'T_{n-1})^6 \le c(n^3p^3 + n^2p^5 + np^6). \tag{A.17}$$

REFERENCES

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory*. Springer, New York.

HUBER, P. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

KOEHLER, K. and LARNTZ, K. (1980). Goodness-of-fit statistics for sparse multinomials. *J. Amer. Statist. Assoc.* **75** 336–344.

MORRIS, C. (1975). Central limit theorems for multinomial sums. *Ann. Statist.* **3** 165–188.

ORTEGA, J. M. and RHEINBOLDT, W. C. (1970). *Iterative Solution of Nonlinear Equations in Several Variables*. Academic, New York.

PORTNOY, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large, I. Consistency. *Ann. Statist.* **12** 1298–1309.

PORTNOY, S. (1985). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large, II. Asymptotic normality. *Ann. Statist.* **13** 1403–1417.

PORTNOY, S. (1986a). Asymptotic behavior of the empiric distribution of M-estimated residuals from a regression model with many parameters. *Ann. Statist.* **14** 1152–1170.

PORTNOY, S. (1986b). On the central limit theorem in R^p when $p \to \infty$. *Probab. Theory Related Fields* **73** 571–583.

PORTNOY, S. (1987). A central limit theorem applicable to robust regression estimators. *J. Multivariate Anal.* **22** 24–50.

RINGLAND, J. (1983). Robust multiple comparisons. *J. Amer. Statist. Assoc.* **78** 145–151.

YOHAI, V. J. and MARONNA, R. A. (1979). Asymptotic behavior of M-estimators for the linear model. *Ann. Statist.* **7** 258–268.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
CHAMPAIGN, ILLINOIS 61820