The Annals of Statistics
1988, Vol. 16, No. 1, 433–449

ON THE ASYMPTOTIC DISTRIBUTION OF QUADRATIC FORMS IN UNIFORM ORDER STATISTICS[1]

BY PETER GUTTORP AND RICHARD A. LOCKHART

University of Washington and Simon Fraser University

The asymptotic distribution of quadratic forms in uniform order statistics is studied under contiguous alternatives. Using minimal conditions on the sequence of forms, the limiting distribution is shown to be the convolution of a sum of weighted noncentral chi-squares and a normal variate. The results give approximate distribution theory even when no limit exists. As an example, high-order spacings statistics are shown to have trivial asymptotic power unless the order of the spacings grows linearly with the sample size. The results are derived from a modification of an invariance principle for quadratic forms due to Rotar, which we prove by martingale central limit methods.

1. Introduction and statement of results. Suppose $U_1 < \cdots < U_n$ are the order statistics for a sample of size n from a distribution on $[0,1]$. Many tests of the hypothesis that the distribution is uniform over $[0,1]$ are based on statistics of the form (at least asymptotically)

$$T = \sum M_{ij}(U_i - i/(n+1))(U_j - j/(n+1)).$$

Examples include the Cramér–von Mises statistic, Watson's U^2-statistic, Greenwood's statistic and Cressie's overlapping spacings statistics. The large-sample distribution of such statistics has been studied by many authors. For empirical process approaches, see Durbin (1973). For an approach based on U-statistics, see Gregory (1977). Generally, such work imposes structure on the matrix $\mathbf{M}$ of the type $nM_{ij} = M(i/n, j/n)$, where $M(s,t)$ is a continuous kernel on $[0,1]^2$. A notable exception is de Wet and Venter (1973), Section 4.

In this work we obtain the asymptotic distribution of T on sequences of contiguous alternatives for a nearly arbitrary sequence of matrices $\mathbf{M}$ subject to the condition that no spacing $U_i - U_{i-1}$ contributes significantly to T in the limit. Our alternatives will have densities of the form $1 + \delta\eta(x)/n^{1/2}$ where $\int\eta^2 = 1$; these are essentially the alternatives considered by Gregory (1977) for a different class of statistics.

Our results may be summarised as follows. The statistic T is the sum of two quadratic forms in the spacings. The first of these forms (which we refer to as the diagonal part of T) is a linear combination of the squares of the centered spacings, $U_i - U_{i-1} - 1/(n+1)$, and will have a limiting normal distribution

Received July 1986; revised May 1987.

[1]Partially supported by NSERC, NSF and the U.S. Environmental Protection Agency through a cooperative agreement with SIMS.

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

Key words and phrases. Goodness of fit, martingale central limit theorem, asymptotic relative efficiency.

under our conditions. The second quadratic form (the off-diagonal part of T) is the part of T due to cross-products of such centered spacings. This form will have a limiting distribution, which is the convolution of a sum of weighted chi-squares (central under the null hypothesis) and a normal variate. The two quadratic forms are not independent but the normal component of the off-diagonal part is asymptotically independent of the diagonal part. The mean of the limiting normal distribution of the diagonal part and the noncentrality parameters of the off-diagonal part respond to the alternative, but the normal component of the off-diagonal part does not. As a result, only the diagonal part and the sum of weighted chi-squares piece of the off-diagonal form contribute to the asymptotic power of the statistic.

We will also see that the diagonal part of T tends to have poor power. It cannot have an asymptotic correlation with the log-likelihood ratio of more than $2^{-1/2}$. In addition, many statistics put equal weight on all the squared spacings; in this case the diagonal part has the same limiting distribution under the null and alternative hypotheses.

In Section 2 we state our main results and discuss their ramifications. A distributional approximation for T is given, which avoids the need to embed $\mathbf{M}$ in a sequence of matrices.

In Section 3 we present an application to overlapping m-spacings statistics such as those of Cressie (1976, 1979). We obtain an expression for the limiting behaviour on our alternatives of such statistics, which holds uniformly in m. As a consequence we show that m must grow proportionally to n to obtain nontrivial asymptotic power against the alternatives considered here.

Our results here overlap those of Hall (1986). Hall has obtained the limiting distribution of the statistics we consider for certain sequences of smooth alternatives. These alternatives approach the null at such a rate that nontrivial limiting power is obtained. When $m = o(n)$ these sequences are more distant than the contiguous alternatives considered here. Thus Hall's results are more general than ours although our results place fewer restrictions on the alternative densities. When m grows proportionally to n, Hall shows convergence on alternatives of our sort (but with smoothness assumptions) to the integral of the square of a Gaussian process. Our techniques provide a more explicit form of the limiting distribution on a wider family of alternative distributions.

We include in Section 3 a brief discussion of other statistics of the form T, concentrating on the Cramér–von Mises statistic to illustrate possible applications.

The results in Section 2 are derived from a modification of an invariance principle for quadratic forms due to Rotar (1973, 1975, 1979). Our results in this area are of some independent interest; we state them in Section 4. Other potential applications are also discussed in this section.

We give proofs (and some minor related results) in Section 5.

2. Limiting distributions for the test statistic. Suppose that $\mathbf{X}^{\mathrm{T}} = (X_1, \dots, X_{n+1})$ is a vector of iid centered exponentials [i.e., X_i has density

$\exp(-(x+1))$ for $x > -1$] and let $\bar{X} = \Sigma_1^{n+1} X_i/(n+1)$. If we define

$$U_i = \frac{i}{n+1} + \frac{\left(\Sigma_{j=1}^{i} X_j - i\bar{X}\right)}{(n+1)(1+\bar{X})},$$

then $(U_1, \dots, U_n)$ have the joint distribution of uniform order statistics.

Let $\mathbf{A}$ be the $n \times n+1$ matrix $A_{ij} = 1(i \ge j)/(n+1)$, $\mathbf{I}$ be the $n+1 \times n+1$ identity matrix and $\mathbf{J}$ be the $n+1 \times n+1$ matrix $J_{ij} = 1/(n+1)$. Set $\mathbf{M}^* = (\mathbf{I} - \mathbf{J})\mathbf{A}^{\mathrm{T}}\mathbf{M}\mathbf{A}(\mathbf{I} - \mathbf{J})$. Substituting for U_i in T, we get

$$T = \frac{\mathbf{X}^{\mathrm{T}}\mathbf{M}^*\mathbf{X}}{(1+\bar{X})^2}.$$

Define

$$s^{*2} = 2\sum_{i \ne j} M_{ij}^{*2}$$

and

$$t^{*2} = 8\sum M_{ii}^{*2} - 4\left(\sum M_{ii}^*\right)^2/(n+1).$$

Let $a_i = M_{ii}^*/t^*$ and $Q_{ij} = M_{ij}^*/s^*$ for $i \ne j$. Then $\mathbf{Q}$ is a symmetric $n+1 \times n+1$ matrix with $2\,\mathrm{tr}(\mathbf{Q}^2) = 1$ and $Q_{ii} = 0$, whereas $\mathbf{a}$ is an $n+1$ vector with $8\Sigma a_i^2 - 4(n+1)\bar{a}^2 = 1$. If $s^* = 0$ let $\mathbf{Q}$ be any symmetric matrix with $2\,\mathrm{tr}(\mathbf{Q}^2) = 1$. If $t^* = 0$ let $\mathbf{a}$ be any vector with $8\Sigma a_i^2 - 4(n+1)\bar{a}^2 = 1$. Now define

$$T^* = s\sum Q_{ij}X_iX_j + t\sum\left(a_i\left(X_i^2 - 1\right) - 2\bar{a}X_i\right),$$

where $s = s^*/(s^{*2} + t^{*2})^{1/2}$ and $t = t^*/(s^{*2} + t^{*2})^{1/2}$. In Section 5 we show that

$$(T - E(T))/\mathrm{Var}^{1/2}(T) - T^* \to 0,$$

in probability as $n \to \infty$. Note that $E(T^*) = 0$ and $\mathrm{Var}(T^*) = 1$. We use P, E and Var to denote probability, expectation and variance under the uniform null hypothesis. Our assumptions will be cast in terms of s, t, $\mathbf{Q}$ and $\mathbf{a}$.

We will be interested in alternative densities $1 + \delta\eta(x)/n^{1/2}$, where we will assume that $\int \eta^2 = 1$. Under these conditions the log-likelihood ratio is asymptotically equivalent (see Section 5 for an outline of a proof), under P, to

$$\delta\mathbf{h}^{\mathrm{T}}\mathbf{X} - \delta^2/2,$$

where $\mathbf{h}$ is a vector with ith entry

$$h_i = (n+1)^{-1/2}\int \eta(x)1(i - 1 < (n+1)x < i)\,dx.$$

Let P_n and E_n denote probability and expectation for samples of size n under these alternatives.

In general, objects with Roman letter names depend on n, whereas objects with Greek letter names do not. The dependence on n is suppressed wherever possible.

To state the results let $l_1, \ldots, l_{n+1}$ be the eigenvalues of $\mathbf{Q}$ arranged in order of decreasing absolute value and let $\mathbf{e}_1, \ldots, \mathbf{e}_{n+1}$ be the corresponding orthonormal eigenvectors. Let $(Z_n;\ n = 0, 1, \ldots)$ be independent standard normal variates. Put $r_i = 2\mathbf{e}_i^{\mathrm{T}}(\mathbf{a} - \bar{a}\mathbf{1})$, where $\mathbf{1}$ is a vector of 1's. Let $s_i = \mathbf{e}_i^{\mathrm{T}}\mathbf{h}$ and $u = 2\mathbf{h}^{\mathrm{T}}(\mathbf{a} - \bar{a}\mathbf{1})$. Let $r_0^2 = 1 - \sum_1^{n+1} r_i^2 = 4\sum_1^{n+1} a_i^2$. Define

$$T_1 = s\sum_1^{n+1} l_i\left[(Z_i + \delta s_i)^2 - 1\right] + t\left[\sum_0^{n+1} r_i Z_i + \delta u\right].$$

THEOREM 1. *If, as $n \to \infty$,*

$$s\max\left\{\sum_j Q_{ij}^2;\ 1 \le i \le n+1\right\} + t\max\{|a_i|;\ 1 \le i \le n+1\} \to 0,$$

then

$$\sup_x \left\{|P_n([T - E(T)]/\mathrm{Var}^{1/2}(T) \le x) - P(T_1 \le x)|\right\} \to 0.$$

The distribution function of T_1 is numerically computable. If $s = 0$, then $t = 1$ and T_1 is simply a $N(\delta u, 1)$ random variable. If $s \ne 0$ and $l_i \ne 0$ put $m_i = r_i t/(2 l_i s)$; otherwise put $m_i = 0$. Let $r_0^{*2} = 1 - 4\sum_1^{n+1} l_i^2 m_i^2$ and $u^* = u - 2s\sum_1^{n+1} l_i m_i s_i/t$. (If all l_i are nonzero, then $u^* = u$ and $r_0^* = r_0$). Then T_1 has the same distribution as

$$s\sum_1^{n+1} l_i\left[(Z_i + m_i + \delta s_i)^2 - (1 + m_i^2)\right] + t\left[r_0^* Z_0 + \delta u^*\right].$$

The distribution function of the latter quantity may be calculated numerically by Fourier inversion of the characteristic function; see Imhof (1961) and Durbin, Knott and Taylor (1975).

The theorem provides distributional approximations for T, which may be calculated without embedding the matrix $\mathbf{M}$ in a sequence of matrices. In Section 3 we illustrate the importance of this point in the context of overlapping spacings statistics.

To get insight into the power of T, it is useful to consider special cases where T_1 has a limiting distribution. In practice, the eigenvalues of $\mathbf{Q}$ often converge. Suppose there are real numbers σ, τ, v, λ_i, ρ_i and σ_i for $i = 1, 2, \ldots$, such that the Roman letter quantities of the theorem all converge to their Greek letter counterparts. In general, it is possible that $2\sum_1^\infty \lambda_i^2 < 1 = 2\sum_1^{n+1} l_i^2$. Let $\lambda_0 = (1 - 2\sum_1^\infty \lambda_i^2)^{1/2}$. Similarly, define $\rho_0 = (1 - \sum_1^\infty \rho_i^2)^{1/2}$ (so that it is possible that ρ_0 is not the limit of r_0). Then our result becomes the following.

THEOREM 2. *Under P_n, as $n \to \infty$,*

$$(T - E(T))\mathrm{Var}^{1/2}(T) \Rightarrow \sigma\left[\sum_1^\infty \lambda_i\left\{(Z_i + \delta\sigma_i)^2 - 1\right\} + \lambda_0 Z_0\right] + \tau\left[\sum_0^\infty \rho_i Z_i + \delta\nu\right].$$

Here and throughout the paper the symbol $\Rightarrow$ denotes convergence in distribution.

We will examine two common cases. In the first case $(n+1)Q_{ij} = \chi(i/(n+1), j/(n+1))$ for some continuous kernel $\chi(s,t)$ on $[0,1]^2$ with $2\int\int\chi^2 = 1$. In this case the eigenvalues l_i (ordered by decreasing absolute value) converge to the eigenvalues (similarly ordered) λ_i of the integral operator χ. Since $2\int\int\chi^2 = 1$ we have $2\Sigma\lambda_i^2 = 1$, so that $\lambda_0 = 0$. Then $\mathbf{e}_i^{\mathrm{T}}\mathbf{h} \to \int\phi_i(s)\eta(s)\,ds = \sigma_i$, where $(\phi_n;\ n = 1,2,\dots)$ are the orthonormal eigenfunctions corresponding to $(\lambda_n;\ n = 1,2,\dots)$. Assume, in addition, that $a_i = \alpha(i/(n+1))/(n+1)$ for some continuous function α on $[0,1]$ with $8\int\alpha^2 - 4(\int\alpha)^2 = 1$. We then have $\rho_i = 2(\int\phi_i\alpha - \int\phi_i\int\alpha)$ for all $i \geq 1$, $\nu = 2\int\alpha\eta$ and $\rho_0 = 4\int\alpha^2$.

The conditions on χ and α can be weakened considerably; what is really needed is sums to be approximated by integrals. For simplicity, the discussion in the next paragraph assumes that η is continuous.

When $\sigma = 0$ the result should be written as

$$(T - E(T))\mathrm{Var}^{1/2}(T) \Rightarrow Z_0 + \delta\nu. \tag{2.1}$$

This is the special case of statistics dominated by terms of the form $(U_{i+1} - U_i)^2$; see the discussion of spacings statistics in Section 3 for examples. Subject to the condition that $8\int\alpha^2 - 4(\int\alpha)^2 = 1$, the quantity $|\nu|$ is maximised by $\alpha(s) = 2^{-3/2}\eta(s)$ in which case $\nu = 2^{-1/2}$. Under P_n, the log-likelihood ratio statistic converges in distribution to $Z_0 + \delta$. Thus, for equal sample sizes, a statistic with limiting distribution (2.1) requires a departure from the null roughly $2^{1/2}$ bigger to achieve the same power as the likelihood ratio test; alternatively, the Pitman efficiency of the most powerful test of this form is 0.5.

When $\tau = 0$ our limiting approximation is

$$\sum\lambda_i\left[(Z_i + \delta\sigma_i)^2 - 1\right] + \lambda_0 Z_0.$$

This quantity has mean $\delta^2\Sigma_1^\infty\lambda_i\sigma_i^2$ and variance 1. The power of the test will then be large when the mean is large. Since $\Sigma\sigma_i^2 = 1$ the mean is maximised in absolute value by $\sigma_1 = 1$ and $\lambda_1^2 = \frac{1}{2}$. This requires $\eta = \phi_1$ and all eigenvalues other than the first to be 0. The corresponding statistic is simply the two-sided equal-tailed likelihood ratio test against the family of alternatives indexed by δ.

The second case we want to consider is when all eigenvalues of $\mathbf{Q}$ converge to 0, or equivalently, $\mathrm{tr}(\mathbf{Q}^4) \to 0$. Then $(T - E(T))/\mathrm{Var}^{1/2}(T)$ is well approximated in distribution (under P_n) by $t(Z_0 + \delta u) + sZ_1$. Note that all the power comes from the term $t\delta u$. The off-diagonal part of T contributes nothing to the power of the test.

In summary then, good power for a statistic of the form T will be obtained only when the limiting distribution under the null hypothesis is a sum of weighted central chi-squares. Such a test will have good power against departures in the direction of those eigenvectors of $\mathbf{Q}$ corresponding to large eigenvalues. Attempts to focus on too many directions at once give poor power against all alternatives.

3. Examples.

3.1. *Spacings statistics.* Greenwood's (1946) statistic is $\Sigma_0^n(U_{i+1} - U_i)^2$ where $U_0 = 0$ and $U_{n+1} = 1$. Cressie (1976, 1979) studies the overlapping m-spacings generalisations

$$C_m = \sum_0^n (U_{i+m} - U_i)^2,$$

where $U_{n+1+k} = 1 + U_k$, and

$$C_m^* = \sum_0^{n+1-m} (U_{i+m} - U_i)^2.$$

Del Pino (1979) restricts the sum to a set of i such that the m-spacings involved do not overlap; we deal here with Cressie's more powerful overlapping versions. The statistic C_m was developed for use on the circle, whereas C_m^* was intended for use on the unit interval. The work of Čibisov (1961) shows that both C_m and C_m^* have asymptotic relative efficiency 0 against our alternatives when m is fixed. Cressie extends this to the case where m grows with n sufficiently slowly that $m^3/n \to 0$. For a more rigorous discussion, see Beirlant and van Zuijlen (1984). Hall (1986) extends the conclusion to $m = o(n)$ for a somewhat more restricted class of alternatives.

Guttorp and Lockhart (1986) study the limiting behaviour of C_m under the null hypothesis for general sequences of values of m. If $2m \le n + 1$, then, in the notation of Section 2, we have $M_{ij}^* = [(m - |i-j|_n)_+/m^2] - 1/(n+1)$, where $|i-j|_n = \min(|i-j|, n+1-|i-j|)$ and $x_+ = \max(x, 0)$. In this case the vector $\mathbf{a} - \bar{a}\mathbf{1} = \mathbf{0}$. We consider separately three cases: $m = 1$, $m > 1$ but $m = o(n)$ and $m/n \to \gamma \in (0, \frac{1}{2}]$.

When $m = 1$ we find $s = (2n+1)^{-1/2}$ and $(C_1 - E(C_1))/\mathrm{Var}^{1/2}(C_1) \Rightarrow Z_0$ under P_n.

When $m > 1$ but $m = o(n)$, Guttorp and Lockhart (1986) show that $\mathrm{tr}(\mathbf{Q}^4) \to 0$. Following the discussion in Section 2, we see that $(C_m - E(C_m))/\mathrm{Var}^{1/2}(C_m) \Rightarrow Z_0$ under P_n.

When $m/n \to \gamma > 0$ we find $t \to 0$. Following the discussion leading to Theorem 2, we are led to a kernel whose spectrum may be found as in Guttorp and Lockhart (1986). The eigenvalues of $\mathbf{Q}$ converge to the set $\lambda_{2k}(\gamma) = \lambda_{2k-1}(\gamma) = \zeta[1 - \cos(2\pi k\gamma)]/k^2$, where $\zeta^2 = 1/(4\pi^4\gamma^3(4/3 - 2\gamma))$. The corresponding eigenfunctions are $\phi_{2k}(t) = 2^{1/2}\cos(2\pi kt)$ and $\phi_{2k-1}(t) = 2^{1/2}\sin(2\pi kt)$ so that the σ_k are simply the Fourier coefficients in the expansion of η. (The kernel also has an eigenvalue of 0 with corresponding eigenfunction 1.) Thus in this case we have under P_n that

$$(C_m - E(C_m))/\mathrm{Var}^{1/2}(C_m) \Rightarrow \sum_1^\infty \lambda_k(\gamma)(Z_k + \delta\sigma_k)^2 - 1\big).$$

As $\gamma \to 0$ this limiting distribution converges to Z_0. Our conclusions may be summarised in the following theorem. In the statement c depends on m and n

via $c = c(m, n) = m/(n+1)$.

Theorem 3. *As* $n \to \infty$,

$$\sup\left\{\left|P_n\left[(C_m - E(C_m))/\mathrm{Var}^{1/2}(C_m) \le x\right]\right.\right.$$
$$\left.\left. - P\left[\sum_1^\infty \lambda_k(c)\left((Z_k + \delta\sigma_k)^2 - 1\right) \le x\right]\right|\right\} \to 0,$$

where the sup *extends over all real x and all* $m \le (n+1)/2$.

Suppose one is considering the use of C_m with $m = 10$ and $n = 100$. The limiting distribution of C_m is normal if m is treated as fixed at 10 or as $m = n^{1/2}$; if m is treated as $n/10$, then the limiting distribution is (a quite skewed) sum of weighted chi-squares. Theorem 3 allows us to avoid thinking about how big m would be if n were different.

The result in Theorem 3 holds with C_m^* replacing C_m, but we are unable to calculate the limiting eigenvalues and eigenfunctions explicitly.

In Guttorp and Lockhart (1986), Monte Carlo studies show that the approximation provided under the null hypothesis by Theorem 3 is good for moderate n (40 or so). It is improved by taking $c(m, n)$ to be the value of c such that $(C_m - E(C_m))/\mathrm{Var}^{1/2}(C_m)$ and $\Sigma_1^\infty \lambda_k(c)(Z_k^2 - 1)$ have the same skewness; the required formulas are in Guttorp and Lockhart.

Theorem 3 provides some insight on the power of C_m and C_m^* and complements the results of Hall (1986). Theorem 3 shows that C_m and C_m^* have 0 ARE against our contiguous alternatives unless m/n is bounded away from 0. Hall (1986) has obtained nontrivial asymptotic power for a class of statistics including C_m and C_m^* against more distant alternatives when $m = o(n)$; his results establish the efficiency assertion stated previously for our alternatives in the case of smooth η.

When $m/n \to \gamma > 0$ Hall establishes weak convergence of C_m to the integral of the square of a Gaussian process. The distribution of this integral is precisely that given in our Theorem 3; Hall does not give this explicit form. Hall notes that rational values of γ lead to different general conclusions about power than irrational values. Theorem 3 sheds some light on this phenomenon. If γ is rational, there will be an infinite number of values of k for which $\lambda_{2k}(\gamma) = \lambda_{2k-1}(\gamma) = 0$, namely, those k for which $k\gamma$ is an integer; if γ is irrational, no such values of k exist. Any η that has nonzero Fourier coefficients only for such k will have a limiting distribution on the alternative the same as on the null; for such alternatives the power is asymptotically equal to the level. If $\gamma = p/q$ in lowest terms, then any η that is periodic with period $1/q$ will have this property; if q is at all large this means a rapidly oscillating alternative density.

The worst case arises when $q = 2$ or $\gamma = \frac{1}{2}$. In this case the statistic C_m will have poor power against alternative density whose shape on $[0, \frac{1}{2})$ is duplicated on $[\frac{1}{2}, 1)$; such a density will have at least two modes. On a more positive note, the statistic should have good power against unimodal symmetric alternatives for which σ_1 and σ_2 will likely be large. It seems likely that the latter sort of alternative is more important in practice. Nevertheless, the existence of alternatives against which the test has trivial asymptotic power is disquieting for an omnibus test of fit.

It should be noted that the asymptotic distributions vary continuously both as a function of γ and as a function of η. Thus the apparent distinction between rational and irrational γ is not really a sharp one. Theorem 3 enables one to calculate approximate powers without figuring out whether $\lim_{n\to\infty} m/n$ is rational or not.

We do not know of any contiguous alternatives for which the overlapping spacings statistics with $m = o(n)$ have nontrivial asymptotic power. On the other hand, we are not certain that contiguous alternatives are the only ones of interest when considering omnibus tests of fit. For example, Cheng and Stephens (1987) have considered alternative densities with a rectangular bump of height not converging to 1 with a width converging to 0. They obtain nontrivial asymptotic power for spacings statistics with fixed m and show that EDF tests do not better even though the alternatives are not contiguous. Note that these alternatives are not within the scope of Hall's work since they are not smooth.

Cressie (1979) considers alternative densities $1 + \eta(x)/n^{1/4}$, where η is a stepfunction. Under these alternatives he obtains nontrivial asymptotic power, whereas the Kolmogorov–Smirnov test and other EDF tests have power converging to 1. It follows that these alternatives are not contiguous to the uniform null hypothesis. When $n^{1/4}$ is replaced by $n^{1/2}$ the alternatives become contiguous and the remarks above apply.

Thus, as far as contiguous alternatives are concerned, the statistics C_m and C_m^* will have poor power properties unless m is chosen to grow linearly with n.

Hall (1984) gives a detailed study of generalisations of C_m under a wide range of alternatives with m fixed.

3.2. *Other statistics.* The results of Section 2 can be applied to EDF statistics such as the Cramér–von Mises statistic W^2, which is asymptotically equivalent to the statistic $\Sigma_1^n(U_i - i/(n+1))^2$. In this case we may apply the discussion following Theorem 2. Extensive algebra shows that $\sigma = 1$. The kernel χ is proportional to $\rho(s,t) = (s^2+t^2)/2 - \max(s,t) + 1/3$. Differentiate the eigenvalue equation

$$\lambda\phi(s) = \int_0^1 \phi(t)\rho(s,t)\,dt$$

three times to see that the eigenfunctions are sines and cosines. Put $s = 1$ and $s = 0$ in the first derivative to see that the eigenvalues of χ and $\lambda_k = (90)^{1/2}/(\pi k)^2$, with corresponding eigenfunctions $2^{1/2}\sin(\pi k s)$. (In addition, 0 is an eigenvalue with eigenfunction 1.) The quantities σ_i are the Fourier coefficients of η. Čibisov (1965) shows that under our sequence of alternatives, the empirical process converges to a Brownian bridge $B_0(t)$ plus a drift $\mu(t) = \delta\int_0^t\eta(s)\,ds$. The statistic W^2 thus converges in distribution to $\int_0^1(B_0(t)+\mu(t))^2\,dt$. This leads to the same limiting distribution as that given in Theorem 2.

4. An invariance principle for quadratic forms in independent variates. In Section 2 we noted that T is asymptotically equivalent to T^*, a linear combination of a quadratic form and two linear forms in independent mean 0 variance 1 random variables. The distribution of T under our alternatives is deduced using contiguity techniques from the limiting joint distribution of T^* and the log-likelihood ratio, itself asymptotically equivalent to another linear form in independent variates. Thus we are led to the study of the joint distribution of quadratic and linear forms in independent variates.

In this section we give asymptotic approximations for such joint distributions. The section concludes with applications of the results to statistics other than those of Section 2.

The theorems of this section have a substantial history. See Whittle (1964), Varberg (1966), de Wet and Venter (1973) and Rotar'(1973, 1975, 1979) among others.

4.1. *Results.* Suppose that for each n $(G_k;\ k = 1,\dots,m)$ is a family of independent σ-fields. (We continue the convention that Roman letters depend on n; the dependence is suppressed wherever possible.) Assume $\mathbf{X}, \mathbf{Y}_1,\dots,\mathbf{Y}_\pi$ are random m-vectors. Assume $E(X_i) = E(Y_{ji}) = 0$ and $\mathrm{Var}(X_i) = \mathrm{Var}(Y_{ji}) = 1$ for $i = 1,\dots,m$ and $j = 1,\dots,\pi$. Assume $X_i, Y_{ji} \in G_i$ for all i, j. (Y_{ji} is the ith entry of $\mathbf{Y}_j$.)

Let $\mathbf{Q}$ be an $m \times m$ symmetric matrix with $Q_{ii} = 0$ and $2\,\mathrm{tr}(\mathbf{Q}^2) = 1$. Let $(l_i, \mathbf{e}_i;\ i = 1,\dots,m)$ be the eigenvalues and corresponding orthonormal eigenvectors of $\mathbf{Q}$. Let $\mathbf{b}_1,\dots,\mathbf{b}_\pi$ be unit length m-vectors.

Put $s_{ijk} = E(Y_{ik}Y_{jk})$. Define a $\pi + 1 \times \pi + 1$ matrix $\mathbf{C}$ by $C_{ii} = 1$ for $i = 0,\dots,\pi, C_{0i} = C_{i0} = 0$ for $i = 1,\dots,\pi$, and otherwise $C_{ij} = \sum_k b_{ik}b_{jk}s_{ijk}$. Let H be the joint distribution function of $(\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X}, \mathbf{b}_1^{\mathrm{T}}\mathbf{Y}_1,\dots,\mathbf{b}_\pi^{\mathrm{T}}\mathbf{Y}_\pi)$ and H^* be the multivariate normal distribution with mean 0 and variance–covariance matrix $\mathbf{C}$. Let Δ metrise weak convergence of distributions on $\mathbb{R}^{\pi+1}$.

THEOREM 4. *Assume the family $(X_i, Y_{ji}, 1 \le i \le m, 1 \le j \le \pi, n \ge 1)$ is uniformly square integrable. If, as $n \to \infty$,*

$$\mathrm{tr}(\mathbf{Q}^4) \to 0$$

and

$$\max\{|b_{ji}|;\ 1 \le i \le m, 1 \le j \le \pi\} \to 0,$$

then

$$\Delta(H, H^*) \to 0.$$

This result can be used to deduce distributional approximations for $\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X}$ without the condition $\mathrm{tr}(\mathbf{Q}^4) \to 0$. In $\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X}$ the coefficient of X_i is a mean 0 random variable with variance $4\sum_j Q_{ij}^2$. Let $\beta(\mathbf{Q}) = \max\{\sum_j Q_{ij}^2;\ 1 \le i \le m\}$.

THEOREM 5. *Suppose $\{X_i;\ 1 \le i \le m,\ n \ge 1\}$ is uniformly square integrable. Assume $\beta(\mathbf{Q}) \to 0$ as $n \to \infty$. If l_i converges to λ_i for each i and $\sum\lambda_i^2 =$*

$(1-\lambda_0^2)/2$, *then*

$$\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X} \Rightarrow \lambda_0 Z_0 + \sum_1^\infty \lambda_i (Z_i^2 - 1).$$

When $\lambda_0 = 0$ Theorem 5 is a neater version, with better conditions, of the main result of de Wet and Venter (1973), who need four finite moments, identical distributions and a variety of technical conditions on $\mathbf{Q}$.

COROLLARY 1 (Rotar, 1975). *If* $\{X_i;\ 1 \le i \le m,\ n \ge 1\}$ *is uniformly square integrable and* $\beta(\mathbf{Q}) \to 0$ *as* $n \to \infty$, *then*

$$\sup_x \left| P(\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X} \le x) - P\left(\sum Q_{ij} Z_i Z_j \le x\right) \right| \to 0.$$

Of course,

$$P\left(\sum Q_{ij} Z_i Z_j \le x\right) = P\left(\sum l_i Z_i^2 \le x\right) = P\left(\sum l_i [Z_i^2 - 1] \le x\right).$$

This corollary is an invariance principle in asserting that the limiting distribution of $\mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X}$ is the same for a large class of distributions of $\mathbf{X}$. The limit may then be calculated for a special $\mathbf{X}$, namely, one with independent, standard normal entries.

The corollary is not as general as the work of Rotar (1979). Rotar does multilinear forms and uses a weaker negligibility condition. On the other hand, his results seem to require a structure assumption on the covariances between the entries of $\mathbf{X}$ and those of the $\mathbf{Y}_j$. We have not been able to deduce our Theorem 4 from his results. Moreover, we are able to sharpen Theorem 4 to obtain mixing convergence; see Theorem 6 in Section 5.

It is important to observe that in many applications the X_i have identical distributions for all i and n. In this case the uniform square integrability assumption is trivially satisfied.

4.2. *Tests for exponentiality.* Suppose $Y_1 < \cdots < Y_n$ are the order statistics from the exponential distribution with mean θ. Many tests of fit to the exponential distribution are based on some quadratic form in the Y_i divided by an estimate of scale. Since the normalised spacings, $D_i = (n-i+1)(Y_i - Y_{i-1})$ where $Y_0 = 0$, are independent exponential variables with mean θ, the theory of Section 4.1 can be applied to study the power of such tests against contiguous alternatives.

Lockhart (1985) and McLaren and Lockhart (1987) consider tests based on the correlation coefficient between the Y_i and the quantities $m_i = \sum_{n-i+1}^m 1/j$, which are the expectations of the Y_i computed under $\theta = 1$. Lockhart uses a special case of Theorem 4 to deduce the asymptotic normality under the null hypothesis of the test statistic. McLaren and Lockhart use that result to show that such tests have 0 Pitman ARE against some natural alternative sequences. The present work shows that this conclusion will extend to any sequence of alternatives such that the log-likelihood ratio is an asymptotically normally distributed linear combination of the D_i. This requirement places restrictions on the tails of the alternative density. Some of the work of McLaren and Lockhart may help to determine what restrictions are necessary.

4.3. *Other applications.* Whittle (1964) studies convergence in distribution of quadratic forms with a view toward applications in time series. He considers statistics $\sum_1^n a_{i-j} X_i X_j$, where the X_i are iid with mean 0 and a finite moment of order $4 + \varepsilon$. If $\sum_{-\infty}^{\infty} a_i^2 < \infty$, he obtains asymptotic normality. The summability hypothesis can be used to show that $\mathrm{tr}(\mathbf{Q}^4) \to 0$ for the appropriate matrices $\mathbf{Q}$. Using the results of Section 4.1, we see that if $a_0 = 0$ we need the X_i to be independent, mean 0, variance 1 and uniformly square integrable. If $a_0 \neq 0$ and the X_i are identically distributed we need only four finite moments. We have not been able to see whether the summability hypothesis is necessary.

The results of Section 4.1 may be expected to have applications in the study of regression problems with nonnormal errors. Although we have not pursued such questions, it seems likely that the most interesting possibilities are connected with the situation where the number of parameters is large. One might study the problem with the number of parameters tending to infinity along with the sample size.

5. Proofs and related results.

The equivalence of T^* and T. By Basu's theorem, $T = \mathbf{X}^{\mathrm{T}}\mathbf{M}^*\mathbf{X}/(1 + \overline{X})^2$ is independent of $(1 + \overline{X})^2$. Thus

$$ET^k = \frac{E(\mathbf{X}^{\mathrm{T}}\mathbf{M}^*\mathbf{X})^k}{E(1 + \overline{X})^{2k}},$$

and, putting $W = \mathbf{X}^{\mathrm{T}}\mathbf{M}^*\mathbf{X} - E(T)(1 + \overline{X})^2$, $\mathrm{Var}\, T = \mathrm{Var}\, W E(1 + \overline{X})^4$. Then

$$\frac{T - E(T)}{\mathrm{Var}^{1/2} T} - \frac{W}{\mathrm{Var}^{1/2} W} = \frac{T - E(T)}{\mathrm{Var}^{1/2} T}\left[1 - \left(\frac{\mathrm{Var}\, T}{\mathrm{Var}\, W}\right)^{1/2} (1 + \overline{X})^2\right] \to 0,$$

in probability. Write $W = W_1 + W_2$, where

$$W_1 = \sum_{i \neq j} M_{ij}^* X_i X_j + \sum \left(M_{ii}^*\left[X_i^2 - 1\right] - 2M_{ii}^* \overline{X}\right),$$

so $W_2 = \mathrm{tr}(\mathbf{M}^*)(1 + 2\overline{X} - (n + 1)\overline{X}^2)/(n + 2)$. Then $T^* = W_1/\mathrm{Var}^{1/2} W_1$. Compute

$$\begin{aligned}\mathrm{Var}\, W_1 &= 2\sum_{i \neq j} M_{ij}^{*2} + 8\sum M_{ii}^{*2} - 4\,\mathrm{tr}^2(\mathbf{M}^*)/(n + 1)\\ &\geq 4\,\mathrm{tr}^2(\mathbf{M}^*)/(n + 1)\end{aligned}$$

and

$$\mathrm{Var}\, W_2 = 2\,\mathrm{tr}^2(\mathbf{M}^*)/[(n + 1)(n + 2)].$$

Since $\mathrm{Var}\, W_2/\mathrm{Var}\, W_1 \to 0$, it follows that $W/\mathrm{Var}^{1/2} W - T^* \to 0$ in probability.

Asymptotic forms of the log-likelihood ratio. The log-likelihood ratio is $\sum \log(1 + \delta\eta(U_i)/n^{1/2})$. Standard Taylor expansion techniques show this is asymptotically equivalent to $\delta\Sigma\eta(U_i)/n^{1/2} - \delta^2/2$. Fix ε positive. There is a twice continuously differentiable η^* such that $\int(\eta - \eta^*)^2 < \varepsilon^3/(18\delta^2)$ and $\int\eta^* = 0$. Use Chebyshev's inequality to see that $P(|\delta\Sigma(\eta(U_i) - \eta^*(U_i))|/n^{1/2} > \varepsilon/3) < \varepsilon/2$. A Taylor expansion shows that $\Sigma\eta^*(U_i)/n^{1/2} + \mathbf{h}^{*\mathrm{T}}\mathbf{X} \to 0$ in probability, where $\mathbf{h}^*$ is defined from η^* as $\mathbf{h}$ is defined from η. By the Cauchy–Schwarz inequality $(h_i^* - h_i)^2$ is bounded by

$$\int(\eta - \eta^*)^2(x)1(i - 1 < (n + 1)x < i)\,dx$$

so that $P(|\delta(\mathbf{h}^* - \mathbf{h})^{\mathrm{T}}\mathbf{X}| > \varepsilon/3) < \varepsilon/2$. Thus, for n sufficiently large, $P(|\Sigma \log(1 + \delta\eta(U_i))/n^{1/2} + \mathbf{h}^{\mathrm{T}}\mathbf{X} + \delta^2/2| > \varepsilon) < \varepsilon$. Since ε is arbitrary we are done.

Notation and matrix inequalities. We will need several vector and matrix norms. For vectors we will need the usual L_p-norms $|\cdot|_p$ for $p = 1, 2$ and ∞. For matrices we will use throughout the fact that $\mathrm{tr}^{1/2}(\mathbf{Q}^2)$, $\beta^{1/2}(\mathbf{Q}) = \max^{1/2}\{\Sigma_j Q_{ij}^2;\ 1 \le i \le m\}$ and $\rho(\mathbf{Q}) = \sup\{|\mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x}|/\mathbf{x}^{\mathrm{T}}\mathbf{x};\ \mathbf{x} \ne \mathbf{0}\} = |l_1|$ are norms. We need the inequalities

$$\mathrm{tr}(\mathbf{Q}^4) \le \rho^2(\mathbf{Q})\mathrm{tr}(\mathbf{Q}^2), \tag{5.1}$$

$$\rho^2(\mathbf{Q}) \le \mathrm{tr}(\mathbf{Q}^2), \tag{5.2}$$

$$\beta^2(\mathbf{Q}) \le \mathrm{tr}(\mathbf{Q}^4). \tag{5.3}$$

PROOF OF THEOREM 4. Define

$$\tau(\mathbf{Q}) = \sum_{i=1}^{m-3}\sum_{j=i+1}^{m-2}\sum_{k=j+1}^{m-1}\sum_{l=k+1}^{m} Q_{ik}Q_{kj}Q_{jl}Q_{li}.$$

If p is a permutation of $\{1,\dots,m\}$ define $(p\mathbf{Q})_{ij} = Q_{pi,pj}$. We will prove later that the condition $\mathrm{tr}(\mathbf{Q}^4) \to 0$ is equivalent to the two conditions

$$\beta(\mathbf{Q}) \to 0 \tag{5.4}$$

and

(5.5) there is a sequence of permutations p such that $\tau(p\mathbf{Q}) \to 0$.

We will need the notation $G(k, n)$ for G_k to indicate the dependence on n.

THEOREM 6. *Theorem 4 remains true if the condition* $\mathrm{tr}(\mathbf{Q}^4) \to 0$ *is replaced by* (5.4) *and* (5.5). *If, in addition* $G(p_n i, n) \subset G(p_{n+1}i, n + 1)$ *for* $i = 1,\dots,m$, *then the convergence is mixing.*

See Hall and Heyde (1980), page 64, for a definition of mixing convergence.

For the sequence p of permutations in (5.5) let $F_k = \bigvee_1^k G_{pj}$. The Cramér–Wold device reduces the problem to the case $\pi = 1$. Define

$$S_{k,0} = \sum_{i=1}^{k}\sum_{j=1}^{k} Q_{pi,pj}X_{pi}X_{pj}$$

and

$$S_{k,1} = \sum_{j=1}^{k} b_{pj} Y_{pj}.$$

Fix r_0 and r_1 such that $r_0^2 + r_1^2 = 1$. Define $S_k = r_0 S_{k,0} + r_1 S_{k,1}$. Then $(S_k, F_k; k = 1,\dots, m)$ is a mean 0 square integrable martingale. Let $D_k = S_k - S_{k-1}$. Theorem 6 will follow from the martingale central limit theorem, Corollary 3.1 (and ensuing remarks) of Hall and Heyde (1980), page 58, provided we show that

$$V(\varepsilon) = \sum_{2}^{m} E\big(D_k^2 1(|D_k| > \varepsilon) | F_{k-1}\big) \to 0, \quad \text{in probability for each } \varepsilon > 0, \tag{5.6}$$

and

$$V = V(0) = \sum_{2}^{m} E\big(D_k^2 | F_{k-1}\big) \to 1, \quad \text{in probability.} \tag{5.7}$$

If, in addition, the σ-fields G_k satisfy the nesting condition of the theorem, then $F_{k,n}$ will be a sub-σ-field of $F_{k,n+1}$. In this case the convergence will be mixing; see Corollary 3.2 of Hall and Heyde.

For notational convenience, we now assume that each p is the identity permutation. Let $D_{k,i} = S_{k,i} - S_{k-1,i}$ and define $V_i(\varepsilon)$ by (5.6) with $D_{k,i}$ replacing D_k. In view of the inequality

$$(x+y)^2 1(|x+y| > \varepsilon) \le 4\big(x^2 1(|x| > \varepsilon/2) + y^2 1(|y| > \varepsilon/2)\big),$$

condition (5.6) will follow from $V_i(\varepsilon) \to 0$ in probability for $i = 0, 1$. For $V_1(\varepsilon)$ this is a simple consequence of uniform square integrability and the condition on $\mathbf{b}$.

Let $\Phi_k(t) = E[X_k^2 1(|X_k| > t)]$ and $\Phi(t) = \sup_{k,n} \Phi_k(t)$. Then

$$V_0(\varepsilon) = 4\sum_k \left[\sum_{j=1}^{k-1} Q_{kj} X_j\right]^2 \Phi_k\left(\varepsilon \Big/ \left[2\left|\sum_{j=1}^{k-1} Q_{kj} X_j\right|\right]\right) \le V_0(0)\Phi(\varepsilon/M),$$

where $M = 2\max\{|\sum_{j=1}^{k-1} Q_{kj} X_j|;\ 1 \le k \le m\}$. Since $E[V_0(0)] = 1$, it suffices to show that $M \to 0$ is probability; this follows from the next lemma.

LEMMA 1. *Suppose $(Y(t);\ t \in T)$ is a family of uniformly square integrable independent random variables. Let A be the set of all a: $T \to \mathbb{R}$ such that $a(t) = 0$ for all but finitely many t and $\sum_t a^2(t) = 1$. Then the family $\{\sum a(t) Y(t);\ a \in A\}$ is uniformly square integrable.*

The lemma is proved below. Now put $t_k^2 = \sum_{j=1}^{k} Q_{kj}^2$ and $W_k = (\sum_{j=1}^{k} Q_{kj} X_j)/t_k$. Then

$$\begin{aligned} P(M > \varepsilon) &\le \sum P(|W_k| > \varepsilon/t_k) \\ &\le \sum_k t_k^2 E\big[W_k^2 1(|W_k| > \varepsilon/t_k)\big]/\varepsilon^2 \\ &\le \sum_k t_k^2 E\big[W_k^2 1\big(|W_k| > \varepsilon/\beta^{1/2}(\mathbf{Q})\big)\big]/\varepsilon^2. \end{aligned}$$

The lemma and (5.3) imply (5.6).

To check (5.7) we will show that $V_i = V_i(0)$ tends to 1 in probability for $i = 0, 1$ and $V_{12} = \Sigma E(D_{k,1}D_{k,2}|F_{k-1}) \to 0$ in probability.

First, $V_{12} = 2r_0r_1\Sigma_k b_k \Sigma_{i=1}^{k-1} Q_{ki}X_i$ has mean 0 and variance proportional to $\Sigma_{i=1}^{m-1}(\Sigma_{k=i+1}^{m} Q_{ki}b_k)^2 = \mathbf{b}^T\mathbf{R}\mathbf{b}$, where $R_{ij} = \Sigma_{k>i,j} Q_{ki}Q_{kj}$. Since $\mathbf{b}$ is a unit vector it suffices to check that $\rho(\mathbf{R}) \to 0$. But from (5.2)

$$\rho^2(\mathbf{R}) \le \operatorname{tr}(\mathbf{R}^2) = 4\tau(\mathbf{Q}) + 2\sum_{i<j<k} Q_{ik}^2Q_{jk}^2 + 2\sum_{k<i<j} Q_{ik}^2Q_{jk}^2 + \sum_{i<k} Q_{ik}^4.$$

We may write $V_0 = 8R_1 + R_2$, where

$$R_1 = \sum_{i=1}^{m-2}\sum_{j=i+1}^{m-1}\sum_{k=j+1}^{m} Q_{ki}Q_{kj}X_iX_j$$

and

$$R_2 = 4\sum_{i=1}^{m-1}\sum_{j=i+1}^{m} Q_{ij}^2X_i^2.$$

Then R_1 has mean 0 and variance $2\tau(\mathbf{Q}) + \Sigma_{i<j<k}Q_{ik}^2Q_{jk}^2$, which converges to 0 by the assumptions on β and τ.

The following triangular array version of the law of large numbers shows that $R_2 \to 1$ in probability and that $V_1 \to 1$ in probability.

LEMMA 2. *Suppose, for each n, that $\mathbf{a}_n$ is a vector in $\mathbb{R}^m$ with $|\mathbf{a}_n|_1 = 1$. Suppose $(Y_n;\ n \ge 1)$ is a sequence of independent uniformly integrable mean 0 random variables. If $|\mathbf{a}_n|_\infty \to 0$, then $\Sigma_i a_{in}Y_i \to 0$ in probability.*

The lemma is proved by truncation as usual. The lemma may be applied to R_2 in view of (5.3). This completes the proof of Theorem 6. The earlier variance calculation for R_1, the identity

$$\sum_p \tau(p\mathbf{Q})/m! = \left(\operatorname{tr}(\mathbf{Q}^4) - 2\sum_{ijk} Q_{ij}^2Q_{ik}^2 + \sum_{ij} Q_{ij}^4\right)/24$$

and (5.3) show that $\operatorname{tr}(\mathbf{Q}^4) \to 0$ implies (5.4) and (5.5). □

We remark that our experience with examples is that it is usually no harder to prove $\tau(\mathbf{Q}) \to 0$ than to prove $\operatorname{tr}(\mathbf{Q}^4) \to 0$. We do not know whether the two are equivalent in the presence of the negligibility condition $\beta(\mathbf{Q}) \to 0$.

In many cases all the entries in $\mathbf{Q}$ are positive. In this case $\tau(\mathbf{Q}) < \operatorname{tr}(\mathbf{Q}^4)$ so that the introduction of the permutations p is unnecessary; this often makes it easier to check the nesting conditions leading to mixing convergence.

PROOF OF THEOREM 5. We give the proof only for the case where infinitely many of the limiting eigenvalues are nonzero. The case of finitely many nonzero limiting eigenvalues is simpler.

Again $(\mathbf{e}_k;\ 1 \le k \le m)$ are orthonormal eigenvectors corresponding to the eigenvalues $(l_k;\ 1 \le k \le m)$. Let $U_k = \Sigma_{j=1}^m e_{k,j}X_j$. Put $W_k = U_k^2 - \Sigma_i e_{k,i}^2X_i^2$.

Lemma 2 shows that $U_k^2 - W_k$ tends to 0 in probability for each k provided we show that $|\mathbf{e}_k|_\infty \to 0$ for each k. This follows from the inequality

$$|l_k e_{k,j}| = \left|\sum_i Q_{ji} e_{k,i}\right| \le \beta^{1/2}(\mathbf{Q}). \tag{5.8}$$

For each fixed L, $T_2 = \Sigma_{L+1}^m l_i W_i$ is a quadratic form in $(X_1, \dots, X_m)$. Let $\mathbf{Q}_L$ be the matrix of this quadratic form; $\mathbf{Q}_L$ has all diagonal entries equal to 0. We claim there is a sequence $L = L_n$ tending to infinity so slowly that

$$\operatorname{tr}(\mathbf{Q}_L^2) \to (1 - \lambda_0^2)/2 \tag{5.9}$$

and that for any such sequence

$$\operatorname{tr}(\mathbf{Q}_L^4) \to 0. \tag{5.10}$$

For this sequence L_n and any fixed K, Theorem 4 shows that

$$(T_2/\lambda_0, U_1, \dots, U_K) \Rightarrow (Z_0, \dots, Z_K).$$

Let H_K be the distribution function of $\Sigma_1^K l_i W_i + T_2$, and let H_K^* be the distribution of $\Sigma_1^K \lambda_i (Z_i^2 - 1) + \lambda_0 Z_0$. If Δ metrises weak convergence on $\mathbb{R}$, then

$$\Delta(H_K, H_K^*) \to 0, \tag{5.11}$$

for some sequence $K = K_n$ tending to infinity sufficiently slowly.

Now $\operatorname{Var}(W_k) = 2(1 - \Sigma_i e_{k,i}^4)$ and, for $j \ne k$, $\operatorname{Cov}(W_j, W_k) = -2\Sigma_i e_{k,i}^2 e_{j,i}^2$. Summing separately over positive and negative eigenvalues and using the general inequality $\operatorname{Var}(Y + Y^*) \le 2(\operatorname{Var}(Y) + \operatorname{Var}(Y^*))$, we see that

$$\operatorname{Var}\left(\sum_{K+1}^{L} l_k W_k\right) \le 2 \sum_{K+1}^{L} l_k^2. \tag{5.12}$$

Now choose a sequence K_n tending to infinity so slowly that (5.11) continues to hold. For this sequence the upper bound in (5.12) tends to 0. Theorem 5 follows from the identity $\mathbf{X}^T\mathbf{Q}\mathbf{X} = \Sigma l_k W_k$.

To check (5.9) and (5.10) fix L and write $\mathbf{Q}_L = \mathbf{Q}_L^* - \mathbf{D}$, where $\mathbf{Q}_L^* = \Sigma_{L+1}^m l_i \mathbf{e}_i \mathbf{e}_i^T$ and $\mathbf{D}$ is a diagonal matrix with $D_{jj} = \Sigma_1^L l_i e_{i,j}^2$. Now

$$\frac{1}{2} - \sum_1^L l_i^2 = \operatorname{tr}(\mathbf{Q}_L{}^{*2}) = \operatorname{tr}(\mathbf{Q}_L^2) + \operatorname{tr}(\mathbf{D}^2).$$

Use (5.8) to see that $\operatorname{tr}(\mathbf{D}^2) \to 0$. Thus for each fixed L we have $\operatorname{tr}(\mathbf{Q}_L^2) \to \frac{1}{2} - \Sigma_1^L \lambda_i^2$. Furthermore, by (5.1),

$$\operatorname{tr}(\mathbf{Q}_L^4) \le \operatorname{tr}(\mathbf{Q}_L^2)\rho^2(\mathbf{Q}_L^* - \mathbf{D}) \le 2(l_{L+1}^2 + \rho^2(\mathbf{D}))\operatorname{tr}(\mathbf{Q}_L^2).$$

It is now easily seen that choosing L_n tending to infinity sufficiently slowly, we can have (5.9) and (5.10). □

Proof of Corollary 1. From any counterexample sequence a subsequence may be extracted, which is a counterexample and for which the eigenvalues converge. Apply Theorem 2 to obtain a contradiction. □

PROOF OF LEMMA 1. If the conclusion is not true, there is an $\varepsilon > 0$ and a counterexample sequence $Z_n = \sum_{i=1}^m a_{ni} Y_{ni}$ with $E(Z_n^2 1(|Z_n| > n)) > \varepsilon$ for all n. There is no loss in assuming that $|a_{n1}| \geq \cdots \geq |a_{nm}|$. Since $\mathrm{Var}(Z_n) = 1$ there is a subsequence along which $m \to \mu$ and $Z_n \Rightarrow Z$. When $\mu = \infty$, extract a further subsequence along which $Y_{ni} \Rightarrow Y_{\infty i}$ and $a_{ni} \to \alpha_i$. Define $\alpha_0^2 = 1 - \sum_1^\infty \alpha_i^2$. Use the Lindeberg central limit theorem and sequences L and K as in the proof of Theorem 5 to prove that $Z_n \Rightarrow \sum_0^\infty \alpha_i Y_{\infty i}$, where $Y_{\infty 0}$ is standard normal and the $Y_{\infty i}$ are taken to be independent. This limit has variance 1, but Z has variance at most $1 - \varepsilon$. The case of finite μ is easier. □

PROOF OF THEOREMS 1 AND 2. From any counterexample sequence for Theorem 1, a subsequence may be extracted, which is a counterexample and for which there are real numbers λ_i, σ_i, ρ_i, σ, τ and v such that all of the following convergences hold: $s \to \sigma$, $t \to \tau$ and $u \to v$, and for each $i \geq 1$, $l_i \to \lambda_i$, $s_i \to \sigma_i$ and $r_i \to \rho_i$. Thus Theorem 1 reduces to Theorem 2.

We prove Theorem 2 only in the case σ, τ and λ_i are positive for all $i \geq 0$. Let $V_1 = \sum(a_i(X_i^2 - 1) - 2\bar{a}X_i)$ and $V_2 = \sum h_i X_i$. With U_k, W_k, T_2 and L_n chosen as in the proof of Theorem 5 we have, from Theorem 4, under P,

$$(T_2/\lambda_0, U_1, \dots, U_K, V_1, V_2)$$
$$\Rightarrow \left(Z_0, \dots, Z_K, \sum_1^K \rho_i Z_i + \rho_{0K} Z_{-1}, \sum_1^K \sigma_i Z_i + \sigma_{0K} Z_{-1} + \sigma_{0K}^* Z_{-2}\right),$$

for each K. Under P_n we have [see Hájek and Šidák (1967), page 208]

$$(T_2/\lambda_0, U_1, \dots, U_K, V_1) \Rightarrow \left(Z_0, Z_1 + \delta\sigma_1, \dots, Z_K + \delta\sigma_K, \sum_1^K \rho_i Z_i + \rho_{0K} Z_{-1} + \delta v\right),$$

for each K. Let H_K be the distribution function of

$$s\left(\sum_1^K l_i W_i + T_2\right) + t\sum\left(a_i\left(X_i^2 - 1\right) - 2\bar{a}X_i\right)$$

and let H_K^* be the distribution of

$$\sigma\left(\sum_1^K \lambda_i\left((Z_i + \delta\sigma_i)^2 - 1\right) + \lambda_0 Z_0\right) + \tau\left(\sum_1^K \rho_i Z_i + \rho_{0K} Z_{-1} + \delta v\right).$$

As in the proof of Theorem 5 there is then a sequence $K = K_n$ tending to infinity so slowly that $\Delta(H_K, H_K^*) \to 0$. Use (5.12) to finish the proof. □

Acknowledgments. This work was done while the first author was visiting the University of British Columbia. It was revised while the second author was visiting the University of Waterloo. We would like to thank Michael Stephens for many helpful discussions and a referee for several useful suggestions, particularly for drawing the work of Hall on spacings to our attention.

REFERENCES

BEIRLANT, J. and VAN ZUIJLEN, M. C. A. (1984). Asymptotic theory for statistics based on m-th order overlapping uniform spacings, where m increases with the sample size N. Report 8434, Dept. Math., Catholic Univ., Nijmegen, The Netherlands.

CHENG, R. and STEPHENS, M. A. (1987). Overlapping spacings statistics. Technical Report, Dept. Math. Statist., Simon Fraser Univ.

ČIBISOV, D. M. (1961). On the tests of fit based on sample spacings. *Theory Probab. Appl.* **6** 325–329.

ČIBISOV, D. M. (1965). An investigation of the asymptotic power of the tests of fit. *Theory Probab. Appl.* **10** 421–437.

CRESSIE, N. (1976). On the logarithms of high order spacings. *Biometrika* **63** 343–355.

CRESSIE, N. (1979). An optimal statistic based on higher order gaps. *Biometrika* **66** 619–627.

DEL PINO, G. E. (1979). On the asymptotic distribution of k-spacings with applications to goodness-of-fit tests. *Ann. Statist.* **7** 1058–1065.

DE WET, T. and VENTER, J. H. (1973). Asymptotic distributions for quadratic forms with applications to tests of fit. *Ann. Statist.* **1** 380–387.

DURBIN, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. SIAM, Philadelphia.

DURBIN, J., KNOTT, M. and TAYLOR, C. C. (1975). Components of Cramér–von Mises statistics. II. *J. Roy. Statist. Soc. Ser. B* **37** 216–237.

GREENWOOD, M. (1946). The statistical study of infectious disease. *J. Roy. Statist. Soc. Ser. A* **109** 85–110.

GREGORY, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* **5** 110–123.

GUTTORP, P. and LOCKHART, R. (1986). On the asymptotic distribution of high order spacings statistics. Technical Report, Dept. Math. Statist., Simon Fraser Univ.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic, New York.

HALL, P. (1984). Limit theorems for sums of general functions of m-spacings. *Math. Proc. Cambridge Philos. Soc.* **96** 517–532.

HALL, P. (1986). On powerful distributional tests based on sample spacings. *J. Multivariate Anal.* **19** 201–224.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic, New York.

IMHOF, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* **48** 419–426.

LOCKHART, R. A. (1985). The asymptotic distribution of the correlation coefficient in testing fit to the exponential distribution. *Canad. J. Statist.* **13** 253–256.

MCLAREN, C. G. and LOCKHART, R. A. (1987). On the asymptotic efficiency of correlation tests of fit. *Canad. J. Statist.* **15** 159–167.

ROTAR, V. I. (1973). Some limit theorems for polynomials of second degree. *Theory Probab. Appl.* **18** 499–507.

ROTAR, V. I. (1975). On the distribution of a quadratic form in many random variables. *Theory Probab. Appl.* **20** 880–882.

ROTAR, V. I. (1979). Limit theorems for polylinear forms. *J. Multivariate Anal.* **9** 511–530.

VARBERG, D. E. (1966). Convergence of quadratic forms in independent random variables. *Ann. Math. Statist.* **37** 567–575.

WHITTLE, P. (1964). On the convergence to normality of quadratic forms in independent variables. *Theory Probab. Appl.* **9** 103–108.

DEPARTMENT OF STATISTICS GN-22
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195

DEPARTMENT OF MATHEMATICS AND STATISTICS
SIMON FRASER UNIVERSITY
BURNABY, BRITISH COLUMBIA V5A 1S6
CANADA