The Annals of Statistics
1989, Vol. 17, No. 3, 1285–1300

GENERALIZED CHI-SQUARE GOODNESS-OF-FIT TESTS FOR LOCATION-SCALE MODELS WHEN THE NUMBER OF CLASSES TENDS TO INFINITY

BY F. C. DROST

University of Tilburg

In this paper we consider generalized chi-square goodness-of-fit tests based on increasingly finer partitions (as the sample size increases) for models with location-scale nuisance parameters. The asymptotic distributions are derived both under the null hypothesis and under local alternatives, obtained by taking contamination families of densities between the null hypothesis and fixed alternative hypotheses. If the number of random cells increases to infinity, the Rao–Robson–Nikulin test statistic is shown to be superior to the Watson–Roy and Dzhaparidze–Nikulin statistics. Conditions are derived under which it is optimal to let the number of classes tend to infinity.

1. Introduction and summary. Let $Y_1, \ldots, Y_n$ be i.i.d. real valued random variables with an absolutely continuous distribution F^Y. A classical goodness-of-fit problem is to test the composite null hypothesis that F^Y belongs to the location-scale family induced by F,

$$H_0: F^Y \in \mathscr{F}_0 = \left\{ F^*(\cdot\,; \vartheta) = F\left(\frac{\cdot - \mu}{\sigma}\right); \mu \in \mathbb{R}, \sigma > 0 \right\}, \tag{1.1}$$

where $\vartheta = (\mu, \sigma)'$ denotes the unknown location-scale parameter.

Well-known omnibus goodness-of-fit tests for (1.1) are chi-square type tests. Suppose that the real line is partitioned into k cells

$$I_{ki}^*(\vartheta) = (\mu + a_{ki-1}\sigma, \mu + a_{ki}\sigma], \qquad i = 1, \ldots, k,$$

where $-\infty = a_{k0} < \cdots < a_{kk} = \infty$. Denote the number of observations in the ith cell by

$$N_{ki}(\vartheta) = \#\{j; Y_j \in I_{ki}^*(\vartheta)\}, \qquad i = 1, \ldots, k,$$

and denote the probability of the ith cell under $F^*(\cdot\,; \vartheta)$ by

$$p_{ki} = \int_{I_{ki}^*(\vartheta)} dF^*(y; \vartheta) = \int_{(a_{ki-1}, a_{ki}]} dF(y), \qquad i = 1, \ldots, k.$$

It is customary to use equiprobable cells. Hence we choose the constants a_{ki} such that

$$p_{ki} = 1/k, \qquad i = 1, \ldots, k.$$

Received June 1986; revised August 1988.

AMS 1980 *subject classifications.* Primary 62E20, 62F05; secondary 62F10, 62F20.

Key words and phrases. Generalized chi-square tests, Rao–Robson–Nikulin statistic, Watson–Roy statistic, Dzhaparidze–Nikulin statistic, location-scale model, nuisance parameters, number of classes, goodness-of-fit.

Define the random k-vector $V_k(\vartheta)$ by its components

$$V_{ki}(\vartheta) = (N_{ki}(\vartheta) - n/k)/(n/k)^{1/2}, \qquad i = 1, \dots, k.$$

Let $\hat{\vartheta}_n$ be some estimator of the unknown nuisance parameter ϑ. In the present framework it is natural to consider chi-square statistics based on random cells. For fixed k, Moore and Spruill (1975) derived the limiting distributions of quadratic forms in $V_k(\hat{\vartheta}_n)$,

$$(1.2) \qquad MS_n = V_k(\hat{\vartheta}_n)'\Gamma_k V_k(\hat{\vartheta}_n),$$

where Γ_k is a symmetric nonnegative definite $k \times k$ matrix. Classical examples in the Moore–Spruill class are [cf. Roy (1956), Watson (1957, 1958), Nikulin (1973), Rao and Robson (1974) and Dzhaparidze and Nikulin (1974)]:

1. The Watson–Roy statistic $WR_n = \|V_k(\hat{\vartheta}_n)\|^2 = V_k(\hat{\vartheta}_n)'V_k(\hat{\vartheta}_n)$.
2. The Rao–Robson–Nikulin statistic $RRN_n = V_k(\hat{\vartheta}_n)'\Sigma_k^- V_k(\hat{\vartheta}_n)$.
3. The Dzhaparidze-Nikulin statistic

$$DN_n = V_k(\hat{\vartheta}_n)'\left[I_k - B_k(B_k'B_k)^{-1}B_k'\right]V_k(\hat{\vartheta}_n),$$

where I_k is the $k \times k$ identity matrix, B_k and Σ_k are defined in (2.4) and (2.7).

The Watson–Roy statistic is the Pearson statistic based on random cells. Asymptotically WR_n is distributed as a weighted sum of chi-square variables. Nikulin (1973) and Rao and Robson (1974) apply Wald's method to the random k-vector $V_k(\hat{\vartheta}_n)$. Dzhaparidze and Nikulin (1974) project $V_k(\hat{\vartheta}_n)$ on the orthogonal complement of the column space of B_k [col(B_k)], removing perturbations due to the estimator $\hat{\vartheta}_n$ (cf. Lemma 5.1). These two statistics have limiting chi-square distributions.

To study the behaviour of MS_n for a broad class of alternatives, let G_1 be any given alternative and consider the contamination family of location-scale distributions

$$\mathscr{G}_{1n} = \left\{G_{1n}^*(\cdot\,;\vartheta) = G_{1n}\left(\frac{\cdot - \mu}{\sigma}\right) = (1-\eta_n)F\left(\frac{\cdot - \mu}{\sigma}\right) + \eta_n G_1\left(\frac{\cdot - \mu}{\sigma}\right); \mu \in \mathbb{R}, \sigma > 0\right\},$$

where $\eta_n \to 0$ as $n \to \infty$. Assume that

$$\eta_n = n^{-1/2}\gamma + o(n^{-1/2}) \quad \text{as } n \to \infty$$

for some fixed $\gamma > 0$. For the local alternative hypothesis

$$(1.3) \qquad H_{1n}: F^Y \in \mathscr{G}_{1n}$$

this rate results in an asymptotic local power (ALP) between α and 1 when k is fixed.

Tumanyan (1956), Steck (1957) and Morris (1975) proved the asymptotic normality of the classical Pearson statistic when $k \to \infty$ and no nuisance parameters are present. Although several authors claimed (without proof) that the limiting distribution results for MS_n of k fixed are easily extended to the case of an increasing number of classes, nontrivial problems arise in the remainder terms due to the growing dimension of V_k. It seems impossible to obtain useful asymptotic properties for the whole Moore–Spruill class when $k \to \infty$, because Γ_k depends on k and the class of all $k \times k$ matrices is too large when $k \to \infty$ (cf. Example 4.1). Moreover, $V_k(\hat{\vartheta}_n)$ has a more complicated covariance matrix than $V_k(\vartheta)$. This motivates the following subclass of the Moore–Spruill class:

$$(1.4) \qquad X_n^2 = V_k(\hat{\vartheta}_n)'[I_k - D_k D_k' + D_k \Lambda(k) D_k'] V_k(\hat{\vartheta}_n),$$

where D_k is the matrix with orthonormal columns defined in the line preceding (2.8) and $\Lambda(k)$ is an arbitrary symmetric nonnegative definite matrix. The matrix $I_k - D_k D_k'$ projects $V_n(\hat{\vartheta}_n)$ in such a way that noise due to $\hat{\vartheta}_n$ is removed [similar to the approach of Dzhaparidze and Nikulin (1974); cf. Lemma 5.1]. The second part $D_k \Lambda(k) D_k'$ creates a large degree of freedom in directions sensitive to the estimator $\hat{\vartheta}_n$ [cf. Rao and Robson (1974), Hsuan (1974) and McCulloch (1985)].

In the presence of a location-scale nuisance parameter the statistics X_n^2 have normal limiting distributions when $k \to \infty$ both under H_0 and H_{1n} (Theorem 3.1). This is our main result which permits some interesting conclusions.

It is well-known that, for fixed k, the Rao–Robson–Nikulin test is uniformly at least as efficient as the Watson–Roy test in the sense of approximate Bahadur slopes [cf. Spruill (1976)]. The analogous result for Pitman efficiencies is not true [cf. Drost (1987a), Moore (1977) and Le Cam, Mahan and Singh (1983)]. If $k \to \infty$, however, RRN_n is at least as efficient in the sense of Pitman as WR_n and DN_n (Corollary 4.1):

$$(1.5) \qquad e_p(RRN, WR) = e_p(RRN, DN) \geq 1.$$

This supports the conclusions of the simulation study of Rao and Robson (1974) where RRN_n is found to be substantially more powerful than WR_n. See also Section 4 for a small simulation study.

Kallenberg, Oosterhoff and Schriever (1985) developed a simple criterion for keeping the number of classes k bounded (or not) for the classical Pearson statistic. We obtain a similar criterion for X_n^2. Let δ_k denote the noncentrality parameter of X_n^2. Then (cf. Remark 3.2)

$$(1.6) \qquad \lim_{k \to \infty} \delta_k / k^{1/2} = \begin{cases} 0 \\ \infty \end{cases} \Rightarrow \text{ALP of } X_n^2 \text{ highest for } \begin{cases} \text{small } k \\ k \to \infty \end{cases}.$$

Since $\delta_k / k^{1/2}$ tends to infinity for heavy-tailed alternatives we recommend a large number of classes if one is mainly interested in heavy-tailed alternatives. A small number of classes is proposed for light-tailed alternatives since then $\delta_k / k^{1/2} \to 0$ as $k \to \infty$.

For fixed k the limiting distributions of X_n^2 depend on the estimator $\hat{\vartheta}_n$. If $k \to \infty$, however, and if the largest eigenvalue of $\Lambda(k)$ is $o(k^{1/2})$ the dependence of $\hat{\vartheta}_n$ disappears. Examples are WR_n and DN_n (generally the limiting distribution of RRN_n depends upon the estimator when $k \to \infty$). Hence the efficiency comparisons in (1.5) remain valid even if different estimators are used in the test statistics WR_n, RRN_n and DN_n.

The Watson–Roy statistic also appears in density estimation theory. Under slightly different conditions Bickel and Rosenblatt (1973) obtained the limiting distributions for WR_n under H_0 and light-tailed alternatives when $k \to \infty$.

2. Assumptions and notation.

2.1. *Assumptions on the distributions.* Denote the gradient with respect to ϑ by ∇_ϑ (∇'_ϑ transposes ∇_ϑ), let $E_{1n}\{v(Y)\}$ denote the expectation of $v(Y)$ with respect to G_{1n} (E_0 and E_1 denote expectations with respect to F and G_1), let the symbols o, o_p, O and O_p have a componentwise interpretation if they are used for vectors or matrices and put $\vartheta_0 = (0,1)'$. Denote the densities corresponding to $F^*(\cdot\,;\vartheta)$ and $F(\cdot)$ $[G^*_{1n}(\cdot\,;\vartheta)$ and $G_{1n}(\cdot)]$ by $f^*(\cdot\,;\vartheta)$ and $f(\cdot)$ $[g^*_{1n}(\cdot\,;\vartheta)$ and $g_{1n}(\cdot)]$ and assume the following regularity conditions:

(C1) (a) $(\forall x, y \in \mathbb{R})|f(x) - f(y)| \le L|x-y|$,

(b) $\lim_{|y|\to\infty} yf(y) = 0$,

(c) $E_0\{\|\nabla_\vartheta \log f^*(Y;\vartheta)|_{\vartheta=\vartheta_0}\|^2\} < \infty$,

(d) G_1 is differentiable with derivative g_1,

(e) $M_1 = \sup_{y\in\mathbb{R}} g_1(y) < \infty$.

Note that (C1a) implies the existence of a derivative $f'(\cdot)$ of $f(\cdot)$ with respect to F a.e. Condition (C1c) implies the finite existence of the Fisher information matrix $J_\vartheta = \sigma^2 J$, where

$$J = E_0\left[\nabla_\vartheta \log f^*(Y;\vartheta)|_{\vartheta=\vartheta_0}\nabla'_\vartheta \log f^*(Y;\vartheta)|_{\vartheta=\vartheta_0}\right].$$

2.2. *Assumptions on the estimator* $\hat{\vartheta}_n$. Assume that $\hat{\vartheta}_n$ is location-scale equivariant and admits the pointwise representation

$$(2.1)\quad n^{1/2}(\hat{\vartheta}_n - \vartheta) = n^{-1/2}\sigma \sum_{j=1}^{n} h\left(\frac{Y_j-\mu}{\sigma}\right) + \sigma Q_n\left(\frac{Y_1-\mu}{\sigma},\dots,\frac{Y_n-\mu}{\sigma}\right),$$

where $h\colon \mathbb{R} \to \mathbb{R}^2$ is the vector-valued influence function and $Q_n\colon \mathbb{R}^n \to \mathbb{R}^2$ is the remainder. Assume

(C2) (a) $E_0\{h(Y)\} = 0$,

(b) $E_0[h(Y)h(Y)'] = A^{-1}$,

(c) $E_1\{\|h(Y)\|^2\} < \infty$,

(d) $Q_n(Y_1,\dots,Y_n) = O_p(1)$ under F and under $\{G_{1n}\}$ as $n \to \infty$, where A is a finite nonsingular matrix. Condition (C2) implies that $\hat{\vartheta}_n$ is $\sqrt{n}$-consistent under H_0 and H_{1n}:

$$n^{1/2}\big(\hat{\vartheta}_n - \vartheta\big) = O_p(1) \quad \text{under } H_0 \text{ and } H_{1n}. \tag{2.2}$$

Under regularity conditions Bickel (1982) showed that the maximum likelihood estimator $\hat{\vartheta}_n^{\mathrm{ML}}$ admits the representation (2.1) with $Q_n = o_p(1)$ and

$$\begin{aligned} h^{\mathrm{ML}}(y) &= J^{-1}\nabla_{\vartheta} \log f^*(y; \vartheta)|_{\vartheta=\vartheta_0} \\ &= -J^{-1}\big(f'(y)/f(y), 1 + yf'(y)/f(y)\big)', \end{aligned} \tag{2.3}$$

implying $A = J$.

2.3. *Further definitions and notation.* Let I_{ki} be the interval

$$I_{ki} = I_{ki}^*(\vartheta_0) = (a_{ki-1}, a_{ki}], \qquad i = 1, \dots, k.$$

Define the $k \times 2$ matrices B_k and C_k by their ith rows

$$\begin{aligned} B_{ki} &= k^{1/2} \int_{I_{ki}} \nabla_{\vartheta}' \log f^*(y; \vartheta)|_{\vartheta=\vartheta_0}\, dF(y) \\ &= k^{1/2}\big[f(a_{ki-1}) - f(a_{ki}), a_{ki-1}f(a_{ki-1}) - a_{ki}f(a_{ki})\big], \end{aligned} \tag{2.4}$$

$$C_{ki} = \left\{k^{1/2} \int_{I_{ki}} h(y)'\, dF(y)\right\} \cdot A, \tag{2.5}$$

$i = 1, \dots, k$. Note that $C_k = B_k$ when using $\hat{\vartheta}_n^{\mathrm{ML}}$ [with h from (2.3)]. Using Lemma A of Kallenberg, Oosterhoff and Schriever (1985) and the Cauchy–Schwarz inequality, the assumptions imply in the general case

$$B_k'B_k = O(1) \quad \text{and} \quad C_k'C_k = O(1) \quad \text{as } k \to \infty. \tag{2.6}$$

Straightforward calculation shows that the asymptotic covariance matrix under H_0 of the nonvanishing part of $V_k(\hat{\vartheta}_n)$ is given by [cf. Moore and Spruill (1975)]

$$\Sigma_k = I_k - q_kq_k' + (B_k - C_k)A^{-1}(B_k - C_k)' - C_kA^{-1}C_k', \tag{2.7}$$

where $q_k = (k^{-1/2}, \dots, k^{-1/2})'$. This matrix does not depend upon the location-scale parameter. Note that $\|q_k\|^2 = 1$ and $q_k'B_k = q_k'C_k = q_k'V_k(\vartheta) = 0$. Let D_k be a matrix with orthonormal columns such that $\mathrm{col}(D_k) = \mathrm{col}([B_k, C_k])$ and let $\Psi(k)$ be the nonnegative definite matrix

$$\Psi(k) = D_k'\Sigma_kD_k. \tag{2.8}$$

Then

$$\Sigma_k = I_k - q_kq_k' - D_kD_k' + D_k\Psi(k)D_k' \tag{2.9}$$

as may be seen by substituting (2.8) in the RHS of (2.9). Note that the Moore–Penrose generalized inverse Σ_k^+ of Σ_k is given by [cf. Rao and Mitra (1971), Chapter 2]

$$\Sigma_k^+ = I_k - q_kq_k' - D_kD_k' + D_k\Psi(k)^+D_k'.$$

Hence, the classical generalizations WR_n, RRN_n and DN_n of the Pearson statistic belong to the class (1.4) with $\Lambda(k) = I_{r(D_k)}$, $\Psi(k)^+$ and 0, respectively.

Let π_{ki} be the ith cell probability under $G_1^*(\cdot;\vartheta)$,

$$\pi_{ki} = \int_{I_{ki}^*(\vartheta)} dG_1^*(y;\vartheta) = \int_{I_{ki}} dG_1(y), \qquad i = 1,\dots,k,$$

and define the k-vector d_k by its components

$$d_{ki} = \gamma k^{1/2}\left(\pi_{ki} - k^{-1}\right), \qquad i = 1,\dots,k.$$

Finally define δ_k^*, the noncentrality parameter δ_k, the location parameter m_k and the variance parameter s_k^2 by

$$\delta_k^* = \|\Lambda(k)^{1/2} D_k'\left(d_k - \gamma B_k E_1\{h(Y)\}\right)\|^2,$$

$$\delta_k = \|[I_k - D_k D_k']d_k\|^2 + \delta_k^*,$$

$$m_k = k + \delta_k,$$

$$s_k^2 = 2k + 4\delta_k.$$

The last two terms are the leading parts of the expectation and the variance of X_n^2. Note that, under H_0, $\delta_k = 0$ and thus $m_k = k$ and $s_k^2 = 2k$.

2.4. *Assumptions on the rate of k.* Let $k = k(n)$ be a particular choice of the number of cells and assume

(C3) (a) $k \to \infty$ as $n \to \infty$,

(b) $\lim_{k\to\infty} \max_{1\le i\le k} \pi_{ki} = 0$,

(c) $n^{-1/2}(1 + k^{-1/2}\lambda_k)k^2 \log^{3/2} k = o(1)$ as $n \to \infty$,

where λ_k is the maximum eigenvalue of $\Lambda(k)$. For test statistics with $\lambda_k = O(k^{1/2})$ condition (C3c) implies that the maximum number of classes is slightly less than $n^{1/4}$; examples are WR_n and DN_n. We end this section with some more technical conditions involving k:

(C4) (a) $Q_n(Y_1,\dots,Y_n) = o_p(k^{1/4}/(1+\lambda_k)^{1/2})$ under F and under $\{G_{1n}\}$ as $n \to \infty$,

(b) $\mathrm{tr}(\Lambda(k)^{1/2}\Psi(k)\Lambda(k)^{1/2}) = o(k^{1/2})$ as $k \to \infty$,

(c) $n^{-1/2}\max\{a_{k1}^4, a_{kk-1}^4\}\log^{-3/2} k = O(1)$ as $n \to \infty$.

The curious condition (C4a) is often implied by (C3c) because Q_n is usually of order $O_p(n^{-1/4})$ [Serfling (1980), Chapter 2, proves this for estimators based on quantiles; for regular estimators one even expects $Q_n = O_p(n^{-1/2})$].

2.5. *The role of the function h.* Consider statistics which do not depend on h through C_k and suppose $\lambda_k = o(k^{1/2})$ (examples are WR_n and DN_n). Then the representation (2.1) is not necessary to obtain Theorem 3.1 but (2.2) suffices.

Taking, however, $h = 0$ and $A = A^{-1} = 0$, we can incorporate these cases in the framework of (2.1). Thus assuming $\lambda_k = o(k^{1/2})$ and (2.2), we omit the conditions (C2) and (C4a, b) [to delete (C4b) use $\operatorname{tr}(\Lambda(k)) = O(\lambda_k) = o(k^{1/2})$ and derive from (2.6) that $\operatorname{tr}(\Psi(k)) = O(1)$]. In these cases the limiting distributions do not depend on the estimator. All effects are incorporated in the negligible remainder term Q_n.

3. Main results. In this section the limiting null and alternative distributions of the test statistics X_n^2 for the testing problem H_0 versus H_{1n} are given. All proofs appear in Section 5.

THEOREM 3.1. *Consider the statistics X_n^2 for testing H_0 against the family of alternatives* (1.3) *determined by G_1. Assume* (C1)–(C4). *Then*

$$(X_n^2 - k)/(2k)^{1/2} \to_{d_0} N(0,1), \tag{3.1a}$$

$$(X_n^2 - m_k)/s_k \to_{d_{1n}} N(0,1) \quad \text{if} \quad \limsup_{k\to\infty} \delta_k^*/s_k < \infty, \tag{3.1b}$$

$$(X_n^2 - k)/(2k)^{1/2} \to_{p_{1n}} \infty \quad \text{if} \quad \lim_{k\to\infty} \delta_k^*/s_k = \infty. \tag{3.1c}$$

REMARK 3.1. Obviously Theorem 3.1 continues to hold if ϑ is either a location or a scale parameter.

In the remainder of this section we state some corollaries concerning the number of classes and the relative efficiency of test statistics of type (1.4).

It is common practice to choose the local alternatives such that the ALP is bounded away from α and 1. So the additional condition on δ_k^* in (3.1b) is quite natural, since otherwise there exists a subsequence of $\{X_n^2\}$ for which the power tends to 1. This is further elaborated in Corollary 3.2. Asymptotically the ratio of the noncentrality parameter δ_k and the square root of k determines the power

$$\beta_\alpha(X_n^2, n, H_{1n}) = P_{1n}(X_n^2(Y_1, \dots, Y_n) > c_k)$$

of the test X_n^2, where the critical values c_k are given by

$$c_k = \inf\{c;\ P_0(X_n^2(Y_1, \dots, Y_n) > c) \le \alpha\}.$$

COROLLARY 3.2. *Assume* (C1)–(C4). *Then*

$$\lim_{n\to\infty} \beta_\alpha(X_n^2, n, H_{1n}) = \begin{cases}\alpha\\ 1\end{cases} \quad \text{iff} \quad \lim_{k\to\infty} \delta_k/k^{1/2} = \begin{cases}0\\ \infty\end{cases}. \tag{3.2}$$

REMARK 3.2. The ALP is between α and 1 for bounded k. If $k \to \infty$ the ALP is smaller (α) if $\delta_k/k^{1/2} \to 0$ and higher (1) if $\delta_k/k^{1/2} \to \infty$. Thus under appropriate conditions relation (1.6) is implied by Corollary 3.2.

To evaluate the relative efficiency of test statistics of type (1.4) we introduce some more notation. Let $S_k^{(i)}$ be a statistic of type (1.4) induced by the matrix

$\Lambda(k)^{(i)}$ and the estimator $\hat{\vartheta}_n^{(i)}$ $(i = 1, 2)$. Define the sequence $n_1(n)$,

$$n_1(n) = \min\left\{n_1; \beta_\alpha\left(S_k^{(1)}, n, H_{1n}\right) - \beta_\alpha\left(S_{k(n_1)}^{(2)}, n_1, H_{1n}\right) \le 0\right\}.$$

The Pitman efficiency of $S_k^{(1)}$ with respect to $S_k^{(2)}$ is defined by

$$e_p(S^{(1)}, S^{(2)}) = \lim_{n\to\infty} n_1(n)/n$$

provided that this limit exists. In Corollary 3.3 it is shown that the relative performance of $S_k^{(1)}$ and $S_k^{(2)}$ only depends upon the ratio of their noncentrality parameters $\delta_k^{(1)}$ and $\delta_k^{(2)}$.

COROLLARY 3.3. *Assume* $\delta_k^{(1)}/(2k)^{1/2} \to c_1$, $0 < c_1 \le \infty$, *and* $\delta_k^{(2)}/(2k)^{1/2} \to c_2$, $0 \le c_2 < \infty$. *Suppose that Theorem* 3.1 *holds both for* $S_k^{(1)}$ *and* $S_k^{(2)}$ *for all sequences of contamination factors* $\eta_n = n^{-1/2}\gamma + o(n^{-1/2})$, $0 \le \gamma < \infty$. *Then*

$$e_p(S^{(1)}, S^{(2)}) = c_1/c_2. \tag{3.3}$$

REMARK 3.3. This corollary resembles Theorem 5.1 of Shirahata (1976); the dependence upon k in relation (5.4) of Shirahata (1976) disappears when $k \to \infty$.

REMARK 3.4. The scope of Corollary 3.3 seems rather limited because $\lim_{k\to\infty} \delta_k/k^{1/2} = 0$ or ∞ in "most" situations. Then the powers of X_n^2 tend to α or 1 under H_{1n} (Corollary 3.2). This suggests that the rate $\eta_n = n^{-1/2}\gamma + o(n^{-1/2})$ is not automatically appropriate when $k \to \infty$. If more generally one chooses η_n such that $n\eta_n^2\delta_{k(n)}/k(n)^{1/2}$ has a positive finite limit [for a given sequence $k(n)$], Theorem 3.1 can be extended to cover such alternatives [cf. Drost (1987b)] and Corollary 3.3 is still applicable.

4. Applications. We now investigate the implications of the previous results for the statistics WR_n, RRN_n and DN_n. The Rao–Robson–Nikulin statistic is not precisely defined in Section 1 because we did not specify the generalized inverse of Σ_k. Although the exact distribution of RRN_n depends upon the choice of Σ_k^- in several examples where $r(\Sigma_k) < k - 1$, the limiting null distribution of RRN_n is generally independent of this choice when k is fixed. Example 4.1 shows that the choice of Σ_k^- is more delicate if $k \to \infty$.

EXAMPLE 4.1. Consider the Laplace null hypothesis with unknown location

$$H_0: f^Y \in \left\{\tfrac{1}{2}\exp(-|\cdot - \mu|); \mu \in \mathbb{R}\right\},$$

let $\hat{\mu}_n = \mathrm{med}(Y_1, \dots, Y_n)$ and let $k = k(n)$ be a sequence of even numbers tending to infinity such that $k^2 \log^{3/2} k = o(n^{1/2})$. The null hypothesis conditions of Theorem 3.1 are easily verified, implying that WR_n and $DN_n = RRN_n$ (with the Moore–Penrose generalized inverse) are asymptotically normal with parameters k and $2k$. To show that the asymptotic distribution of RRN_n generally depends upon the choice of the generalized inverse of Σ_k consider the generalized inverse

$$\Sigma_k^- = I_k + c_n n q_k^* q_k^{*\prime},$$

with (c_n) an arbitrary sequence and $q_k^* = (-k^{-1/2}, \dots, -k^{-1/2}, k^{-1/2}, \dots, k^{-1/2})$. Then

$$RRN_n = V_k(\hat{\mu}_n)'\Sigma_k^- V_k(\hat{\mu}_n) = WR_n + \begin{cases} 0 & \text{if } n \begin{cases}\text{even}\\ \text{odd}\end{cases} \\ c_n & \end{cases}$$

and $(RRN_n - k)/(2k)^{1/2}$ converges to a standard normal distribution only if $c_n \to 0$ as $n \to \infty$. In general the standardization depends upon c_n and hence upon the generalized inverse.

This example also shows that one cannot expect to obtain a useful limit theorem for the whole class of Moore–Spruill test statistics when $k \to \infty$.

To avoid pathological behaviour we restrict attention to the Moore–Penrose generalized inverse Σ_k^+ of Σ_k; from now on we assume that

$$RRN_n = V_k(\hat{\vartheta}_n)'\Sigma_k^+ V_k(\hat{\vartheta}_n).$$

To make comparisons of WR_n, RRN_n and DN_n more transparent define the modified Dzhaparidze–Nikulin statistic

$$\tilde{D}\tilde{N}_n = V_k(\hat{\vartheta}_n)'[I_k - D_k D_k']V_k(\hat{\vartheta}_n),$$

which projects $V_k(\hat{\vartheta}_n)$ on the linear subspace of $\mathbb{R}^k$ orthogonal to $\operatorname{col}(D_k)$. Note that $\tilde{D}\tilde{N}_n = DN_n$ when using the maximum likelihood estimator $\hat{\vartheta}_n^{\mathrm{ML}}$ and the influence function h^{ML} of Bickel (1982) [cf. (2.3)].

McCulloch (1985) proved that if one uses $\hat{\vartheta}_n^{\mathrm{ML}}$ the Rao–Robson–Nikulin statistic is the sum of the Dzhaparidze–Nikulin statistic and the positive statistic for testing normality proposed by Hsuan (1974). Let $\hat{\vartheta}_n$ be a more general choice of the estimator and assume that the same function h is used in the definition of $\tilde{D}\tilde{N}_n$ and RRN_n. Then we have the inequalities

$$\tilde{D}\tilde{N}_n \le DN_n \le WR_n \quad \text{and} \quad \tilde{D}\tilde{N}_n \le RRN_n.$$

Similar relations are true for the corresponding noncentrality parameters.

Corollary 4.1. *Assume* $\|d_k\|^2/k^{1/2} \to c$, $0 < c < \infty$. *Suppose that the conditions* (C1)–(C3) *and* (C4a, c) *are fulfilled for* $\tilde{D}\tilde{N}_n$, DN_n, WR_n *and* RRN_n *and assume* $D_k' d_k = o(k^{1/4})$. *Then*

$$e_p(\tilde{D}\tilde{N}, DN) = e_p(\tilde{D}\tilde{N}, WR) = 1 \tag{4.1a}$$

and

$$e_p(RRN, \tilde{D}\tilde{N}) = \lim_{k\to\infty} \delta_k^{RRN}/\|d_k\|^2 \ge 1. \tag{4.1b}$$

Proof. The noncentrality parameters of WR_n, DN_n and $\tilde{D}\tilde{N}_n$ are equal to $\|d_k\|^2 + o(k^{1/2})$. Note also that condition (C4b) is fulfilled for RRN_n. Application of Corollary 3.3 yields the desired results. □

Tedious calculations show that the inequality in (4.1b) is strict in several examples where the alternatives are not light-tailed, e.g., consider the testing problem of an exponential null hypothesis with unknown scale against the contamination of two exponential densities,

$$H_0: F^Y \in \{1 - \exp(-\cdot/\sigma); \sigma > 0\}$$

versus

$$H_{1n}: F^Y \in \left\{(1 - n^{-1/2})(1 - \exp(-\cdot/\sigma)) + n^{-1/2}(1 - \exp(-\cdot/4\sigma)); \sigma > 0\right\}.$$

Then $e_p(RRN, \tilde{D}\tilde{N}) = e_p(RRN, DN) = e_p(RRN, WR) > 1$. Hence under mild conditions the Rao–Robson–Nikulin test turns out to be the best of the classical generalizations of the Pearson test if k tends slowly to infinity.

The theory is illustrated with a simulation study for a normal location and a normal location-scale hypothesis with maximum likelihood estimators. Simulations were performed on a CDC-Cyber using the FORTRAN programming language, the STATAL random number generator and the statistical package STAR. In Figure 4.1 simulated powers (based on 10,000 replications of a sample size $n = 100$) are given for three kinds of alternatives: heavy-tailed, light-tailed

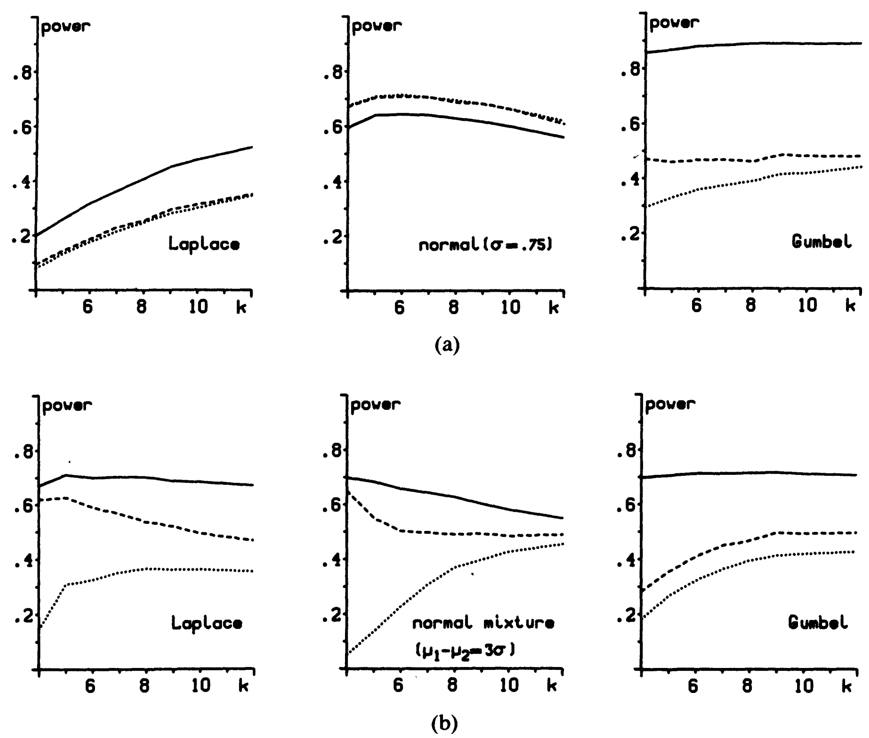

FIG. 4.1. —= RRN_n, ---= WR_n *and* ··· = D_n. (*a*) *Normal location null hypothesis*, $\hat{\mu}_n = \overline{Y}_n$. (*b*) *Normal location-scale null hypothesis*, $(\hat{\mu}_n, \hat{\sigma}_n) = (\overline{Y}_n, S_n)$.

and skew [other cases including nonmaximum likelihood estimators are reported in Drost (1987b)]. Generally the Rao–Robson–Nikulin test dominates WR_n and $DN_n = \tilde{D}\tilde{N}_n$. The influence of the choice of k is less strong than in Kallenberg, Oosterhoff and Schriever (1985). At the heavy-tailed Laplace alternative in the location-scale model it is even opposite to our expectations for WR_n. In other cases, however, a large (small) number of classes is preferable for heavy- (light-) tailed alternatives [cf. also Drost (1987b)].

5. Proofs. Throughout this section we assume without further references the conditions (C1)–(C4). The proof of Theorem 3.1 is based on three lemmas. The first one rewrites $V_k(\hat{\vartheta}_n)$ as the sum of $V_k(\vartheta_0)$ and two remainder terms. The last two lemmas investigate the influence of the error terms. Application of Theorem 5.1 of Morris (1975) yields the desired result. Proofs are only given under H_{1n} and $\vartheta = \vartheta_0 = (0,1)'$. Note that (2.2) implies that $\hat{\mu}_n \to_p 0$ and $\hat{\sigma}_n \to_p 1$ if ϑ_0 is true.

Introduce the notation $g(y)|_a^b = g(b) - g(a)$; $g(y)|_a^b - |_c^d$ is similarly defined. Let $G_{1n}(U) = \int_U dG_{1n}(y)$ and let F_n be the empirical distribution function of $Y_1, \dots, Y_n$.

LEMMA 5.1 [Moore and Spruill (1975)].

$$V_k(\hat{\vartheta}_n) = V_k(\vartheta_0) - B_k n^{1/2}(\hat{\vartheta}_n - \vartheta_0) + R_k, \tag{5.1}$$

where $R_k = R_{1k} + R_{2k}$ *are random k-vectors with components*

$$R_{1ki} = p_{ki}^{-1/2} n^{1/2} \{F_n(y) - G_{1n}(y)\}\Big|_{\hat{\mu}_n + a_{ki-1}\hat{\sigma}_n}^{\hat{\mu}_n + a_{ki}\hat{\sigma}_n} - \Big|_{a_{ki-1}}^{a_{ki}},$$

$$R_{2ki} = B_{ki} n^{1/2}(\hat{\vartheta}_n - \vartheta_0) + p_{ki}^{-1/2} n^{1/2} G_{1n}(y)\Big|_{\hat{\mu}_n + a_{ki-1}\hat{\sigma}_n}^{\hat{\mu}_n + a_{ki}\hat{\sigma}_n} - \Big|_{a_{ki-1}}^{a_{ki}}, \qquad i = 1, \dots, k.$$

PROOF. Direct calculation. □

LEMMA 5.2.

$$\|R_k\|^2 = o_p(1) \quad \textit{under } H_0 \textit{ and } H_{1n}, \tag{5.2a}$$

$$\|\Lambda(k)^{1/2} D_k' R_k\|^2 = o_p(k^{1/2}) \quad \textit{under } H_0 \textit{ and } H_{1n}. \tag{5.2b}$$

PROOF. The conclusions are implied by similar statements about R_{1k} [part (a) of the proof] and R_{2k} [part (b) of the proof].

(a) The proof is based on a modification of Ruymgaart (1974). His theorem is not directly applicable because for $k \to \infty$ the mean of k random variables is not necessarily tight.

Let $\varepsilon > 0$. Because of (2.2) there exists N_ε such that for all n,

$$P_{1n}\left(n^{1/2}\|\hat{\vartheta}_n - \vartheta_0\| > N_\varepsilon\right) \le \varepsilon/4.$$

Let n_0 be a sufficiently large integer. Define intervals

$$J_{ni} = \{y \in \mathbb{R};\, |y - a_{ki}| \le c_{ni} + (1 + |a_{ki}|)n^{-1/2}N_\varepsilon\}, \qquad i = 1,\dots,k-1,$$

where the constants $c_{ni} \ge 0$ are chosen such that for $n \ge n_0$,

$$G_{1n}(J_{ni}) = 2n^{-1/2}\log^{1/2} k, \qquad i = 1,\dots,k-1. \tag{5.3}$$

The construction is possible because of (C1) and (C4d). Note

$$\begin{aligned} &P_{1n}(\exists\, i \in \{1,\dots,k-1\}\hat{\mu}_n + a_{ki}\hat{\sigma}_n \notin J_{ni}) \\ &\quad = P_{1n}(\exists\, i \in \{1,\dots,k-1\}|\hat{\mu}_n + a_{ki}(\hat{\sigma}_n - 1)| > c_{ni} + (1 + |a_{ki}|)n^{-1/2}N_\varepsilon) \\ &\quad \le P_{1n}(n^{1/2}|\hat{\mu}_n| > N_\varepsilon) + P_{1n}(n^{1/2}|\hat{\sigma}_n - 1| > N_\varepsilon) \le \varepsilon/2. \end{aligned}$$

Let $\delta > 0$. Then for $n \ge n_0$ condition (C3c) implies

$$256 n^{-1/2}(1 + k^{-1/2}\lambda_k)k^2 \log^{3/2} k \le \delta.$$

Because $\|\Lambda(k)^{1/2}D_k' R_{1k}\|^2 \le \lambda_k \|R_{1k}\|^2$ the probabilities $P_{1n}(\|R_{1k}\|^2 \ge \delta)$ and $P_{1n}(\|\Lambda(k)^{1/2}D_k' R_{1k}\|^2 \ge \delta k^{1/2})$ are both bounded by

$$\begin{aligned} &P_{1n}\left(kn\left(\{F_n(y) - G_{1n}(y)\}\Big|_{a_{ki}}^{\hat{\mu}_n + a_{ki}\hat{\sigma}_n} - \Big|_{a_{ki-1}}^{\hat{\mu}_n + a_{ki-1}\hat{\sigma}_n}\right)^2 \ge 256 n^{-1/2}k^2\log^{3/2} k\right) \\ &\quad \le P_{1n}(\exists\, i \in \{1,\dots,k-1\}\hat{\mu}_n + a_{ki}\hat{\sigma}_n \notin J_{ni}) \\ &\qquad + P_{1n}\Big(\forall\, i \in \{1,\dots,k-1\}\hat{\mu}_n + a_{ki}\hat{\sigma}_n \in J_{ni}; \\ &\qquad\qquad \exists\, i \in \{1,\dots,k-1\}|\{F_n(y) - G_{1n}(y)\}\Big|_{a_{ki}}^{\hat{\mu}_n + a_{ki}\hat{\sigma}_n}| \\ &\qquad\qquad \ge 8n^{-3/4}\log^{3/4} k\Big) \end{aligned}$$

$$\le \varepsilon/2 + \sum_{i=1}^{k-1} P_{1n}\left(\sup_{U \in \mathscr{J}_{ni}} |F_n(U) - G_{1n}(U)| \ge 8n^{-3/4}\log^{3/4} k\right), \tag{5.4}$$

where $\mathscr{J}_{ni} = \{U \subset J_{ni};\, U \text{ is an interval}\}$. To prove that the second term of (5.4) is bounded by $\varepsilon/2$ define the conditional probability

$$\pi_i(j) = P_{1n}\left(\sup_{U \in \mathscr{J}_{ni}} |F_n(U) - G_{1n}(U)| \ge 8n^{-3/4}\log^{3/4} k \,\Big|\, F_n(J_{ni}) = j/n\right),$$

$i = 1,\dots,k-1$; $j = 0,\dots,n$. Then rewrite the second term of (5.4) as

$$\left(\sum_{j \le n^{1/2}\log^{1/2} k} + \sum_{j > n^{1/2}\log^{1/2} k}\right)\pi_i(j) P_{1n}(F_n(J_{ni}) = j/n). \tag{5.5}$$

For $n \ge n_0$ it is easily seen that the first sum of (5.5) is bounded by $\varepsilon/(4k)$ using (5.3), $\pi_i(j) \le 1$ and

$$\begin{aligned} P_{1n}(F_n(J_{ni}) \le n^{-1/2}\log^{1/2} k) &\le P_{1n}(n|F_n(J_{ni}) - G_{1n}(J_{ni})| \ge n^{1/2}\log^{1/2} k) \\ &\le C\exp(-2\log k) \le \varepsilon/(4k). \end{aligned}$$

Next we show that the second sum of (5.5) is also bounded by $\varepsilon/(4k)$. Note that for $j \neq 0$ conditionally given $F_n(J_{ni}) = j/n$,

$$\sup_{U \in \mathscr{J}_{ni}} |F_n(U) - G_{1n}(U)| \leq |F_n(J_{ni}) - G_{1n}(J_{ni})| + G_{1n}(J_{ni}) \sup_{U \in \mathscr{J}_{ni}} |\tilde{F}_j(U) - \tilde{G}_{1n}(U)|,$$

where $\tilde{G}_{1n}$ is the conditional distribution of Y_1 under H_{1n} given J_{ni} and $\tilde{F}_j$ is the corresponding empirical distribution function based on j observations [cf. Ruymgaart (1974), page 902]. Define

$$\pi_{1i}(j) = P_{1n}\big(|F_n(J_{ni}) - G_{1n}(J_{ni})| \geq 4n^{-3/4}\log^{3/4} k \big| F_n(J_{ni}) = j/n\big),$$

$$\pi_{2i}(j) = P_{1n}\Big(\sup_{U \in \mathscr{J}_{ni}} |\tilde{F}_j(U) - \tilde{G}_{1n}(U)| \geq 2n^{-1/4}\log^{1/4} k\Big),$$

$i = 1, \dots, k-1$; $j = 1, \dots, n$. Using $\pi_i(j) \leq \pi_{1i}(j) + \pi_{2i}(j)$ and Bernstein's inequality [cf. Bahadur (1966), (12)], for $n \geq n_0$ the second sum of (5.5) is bounded by

$$\begin{aligned}
&\sum_{j > n^{1/2}\log^{1/2} k} (\pi_{1i}(j) + \pi_{2i}(j)) P_{1n}(F_n(J_{ni}) = j/n) \\
&\quad \leq P_{1n}\big(n|F_n(J_{ni}) - G_{1n}(J_{ni})| \geq 4n^{1/4}\log^{3/4} k\big) \\
&\qquad + \sum_{j > n^{1/2}\log^{1/2} k} P_{1n}\Big(\sup_{y \in \mathbb{R}} |\tilde{F}_j(y) - \tilde{G}_{1n}(y)| \geq n^{-1/4}\log^{1/4} k\Big) \\
&\qquad\qquad \times P_{1n}(F_n(J_{ni}) = j/n) \\
&\quad \leq C\exp\big(-4\log k/(1 + 2n^{-1/4}\log^{1/4} k)\big) \\
&\qquad + \sum_{j > n^{1/2}\log^{1/2} k} C\exp(-2jn^{-1/2}\log^{1/2} k) P_{1n}(F_n(J_{ni}) = j/n) \\
&\quad \leq 2C\exp(-2\log k) \leq \varepsilon/(4k).
\end{aligned}$$

Thus from (5.4) and (5.5) we obtain the desired result for R_{1k}.

(b) A Taylor expansion of R_{2ki} shows

$$|R_{2ki}| \leq Ck^{1/2}\max(a_{k1}^2, a_{kk-1}^2) n^{1/2} \|\hat{\vartheta}_n - \vartheta_0\| \big(\|\hat{\vartheta}_n - \vartheta_0\| + n^{-1/2}\big).$$

Using $\|\Lambda(k)^{1/2} D_k' R_{2k}\|^2 \leq \lambda_k \|R_{2k}\|^2$, (b) is implied by (2.2), (C3c) and (C4c). □

LEMMA 5.3.

$$\|D_k' V_k(\vartheta_0)\|^2 = O_p(1) \quad \textit{under } H_0 \tag{5.6a}$$

and

$$\|D_k'(V_k(\vartheta_0) - d_k)\|^2 = O_p(1) \quad \textit{under } H_{1n}. \tag{5.6b}$$

$$\Big\|\Lambda(k)^{1/2} D_k' \Big(V_k(\vartheta_0) - B_k n^{-1/2} \sum_{j=1}^{n} h(Y_j)\Big)\Big\|^2 = o_p(k^{1/2}) \quad \textit{under } H_0 \tag{5.7a}$$

and

$$(5.7\mathrm{b})\quad \left\| \Lambda(k)^{1/2} D_k' \left(V_k(\vartheta_0) - d_k - B_k n^{-1/2} \sum_{j=1}^{n} \{h(Y_j) - E_{1n} h(Y_j)\} \right) \right\|^2 = o_p(s_k) \quad \textit{under } H_{1n}.$$

PROOF. Evaluate the expectations of the LHS of (5.6) and (5.7) and show that they are of the indicated order of magnitude. The proof is completed using the implication

$$E\{\|X_k\|^2\} = O(1) \quad \text{uniform in} \quad k \Rightarrow \|X_k\|^2 = O_p(1) \quad \text{uniform in } k. \qquad \square$$

PROOF OF THEOREM 3.1. Note that Theorem 5.1 of Morris (1975) implies $\|V_k(\vartheta_0)\|^2 = AN(k + \|d_k\|^2,\ 2k + 4\|d_k\|^2) = O_p(k + \|d_k\|^2) = O_p(s_k^2)$ and also note that col($[B_k, C_k]$) is the kernel of the projection matrix $I_k - D_k D_k'$. Thus the first part of X_n^2 can be rewritten as

$$\begin{aligned}
\|[I_k - D_k D_k'] V_k(\hat{\vartheta}_n)\|^2 &= \|[I_k - D_k D_k'](V_k(\vartheta_0) + R_k)\|^2 \\
&= \|V_k(\theta_0)\|^2 - \|D_k'(V_k(\theta_0) - d_k)\|^2 - \|D_k' d_k\|^2 \\
&\quad - 2(V_k(\vartheta_0) - d_k)' D_k D_k' d_k + \|[I_k - D_k D_k'] R_k\|^2 \\
&\quad + 2 V_k(\vartheta_0)' [I_k - D_k D_k'] R_k \\
&= AN\left(k + \|[I_k - D_k D_k'] d_k\|^2, 2k + 4\|d_k\|^2\right) + o_p(s_k)
\end{aligned}$$

[use the previous lemmas, $D_k' d_k = o(s_k)$ and the Cauchy–Schwarz inequality for the cross-terms]. Similarly we treat the second term of X_n^2,

$$\begin{aligned}
&\|\Lambda(k)^{1/2} D_k' V_k(\hat{\vartheta}_n)\|^2 \\
&\quad = \left\| \Lambda(k)^{1/2} D_k' \left(V_k(\vartheta_0) - d_k - B_k n^{-1/2} \sum_{j=1}^{n} \{h(Y_j) - E_{1n} h(Y_j)\} \right) \right\|^2 \\
&\qquad + \|\Lambda(k)^{1/2} D_k' (R_k - B_k Q_n(Y_1, \dots, Y_n))\|^2 + \delta_k^* + \text{cross-terms} \\
&\quad = \delta_k^* + o_p(s_k) + o_p\left((\delta_k^* s_k)^{1/2}\right)
\end{aligned}$$

[also use (C4a)]. The theorem follows. $\square$

PROOF OF COROLLARY 3.2. The critical values of the test X_n^2 satisfy $c_k = k + (2k)^{1/2}\zeta_\alpha + o(k^{1/2})$, where $\zeta_\alpha = \Phi^{-1}(1 - \alpha)$ denotes the upper α-point of the standard normal distribution function Φ. Because every subsequence of δ_k^*/s_k has a further sequence with a limit (finite or infinite) we assume without loss of generality that the sequence δ_k^*/s_k has a limit. If $\lim_{k\to\infty} \delta_k^*/s_k < \infty$ apply

$$\beta_\alpha(X_n^2, n, H_{1n}) = P_{1n}\big((X_n^2 - m_k)/s_k > -\delta_k/s_k + \zeta_\alpha(2k)^{1/2}/s_k + o(1)\big)$$
$$\to \begin{cases}\alpha \\ 1\end{cases} \text{iff } \delta_k/s_k \to \begin{cases}0 \\ \infty\end{cases}$$

[using (3.1b)] and otherwise apply

$$\beta_\alpha(X_n^2, n, H_{1n}) = P_{1n}\big((X_n^2 - k)/(2k)^{1/2} > \zeta_\alpha + o(1)\big) \to 1$$

[using (3.1c)]. Combination of these two results yields (3.2). □

PROOF OF COROLLARY 3.3. Let $a_n \approx b_n$ have the interpretation ($\forall\, \varepsilon > 0$, $\exists\, n_0$, $\forall\, n > n_0$) $|a_n - b_n| < \varepsilon$. Suppose $0 < c_1, c_2 < \infty$ and let $m = m(n)$ be a sequence such that $m \le (-\delta + c_1/c_2)n$ for some $0 < \delta < c_1/c_2$. If m remains bounded,

$$\beta_\alpha(S_{k(m)}^{(2)}, m, H_{1n}) \approx \alpha < \Phi(c_1 - \zeta_\alpha) \approx \beta_\alpha(S_k^{(1)}, n, H_{1n});$$

otherwise

$$\beta_\alpha(S_{k(m)}^{(2)}, m, H_{1n}) \approx \Phi\left(\frac{m}{n}\frac{\delta_{k(m)}^{(2)}}{(2k(m))^{1/2}} - \zeta_\alpha\right) \approx \Phi\left(\frac{m}{n}c_2 - \zeta_\alpha\right)$$
$$\le \Phi((-\delta + c_1/c_2)c_2 - \zeta_\alpha) < \Phi(c_1 - \zeta_\alpha) \approx \beta_\alpha(S_k^{(1)}, n, H_{1n}).$$

Similarly one proves for $m = (\delta + c_1/c_2)n$ ($\delta > 0$),

$$\beta_\alpha(S_{k(m)}^{(2)}, m, H_{1n}) \approx \Phi((\delta + c_1/c_2)c_2 - \zeta_\alpha) > \Phi(c_1 - \zeta_\alpha) \approx \beta_\alpha(S_k^{(1)}, n, H_{1n})$$

and thus $e_p(S^{(1)}, S^{(2)}) = c_1/c_2$. If $c_1 = \infty$ or $c_2 = 0$, consider the sequences $m = Mn$ ($M > 0$) and proceed in a similar way to obtain $e_p(S^{(1)}, S^{(2)}) = \infty$. □

Acknowledgments. The author wishes to express his gratitude to J. Oosterhoff and W. C. M. Kallenberg for their constant encouragement and valuable suggestions. The author would like to thank an Associate Editor and two referees for their helpful comments.

REFERENCES

BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.

DROST, F. C. (1987a). A limit theorem for some modified chi-square statistics when the number of classes increases. In *Mathematical Statistics and Probability Theory* (P. Bauer, F. Konecny and W. Wertz, eds.) **B** 43–58. Reidel, Dordrecht.

DROST, F. C. (1987b). Asymptotics for generalized chi-square goodness-of-fit tests. Dissertation, Free University, Amsterdam.

DZHAPARIDZE, K. O. and NIKULIN, M. S. (1974). On a modification of the standard statistics of Pearson. *Theory Probab. Appl.* **19** 851–853.

HSUAN, T. A. (1974). Some chi-square statistics for goodness-of-fit tests. Dissertation, Cornell Univ.

KALLENBERG, W. C. M., OOSTERHOFF, J. and SCHRIEVER, B. F. (1985). The number of classes in chi-squared goodness-of-fit tests. *J. Amer. Statist. Assoc.* **80** 959–968.

LE CAM, L., MAHAN, C. M. and SINGH, A. (1983). An extension of a theorem of H. Chernoff and E. L. Lehmann. In *Recent Advances in Statistics* (M. H. Rizvi, J. S. Rustagi and D. Siegmund, eds.) 303–337. Academic, New York.

MCCULLOCH, C. E. (1985). Relationships among some chi-square goodness of fit statistics. *Comm. Statist. A—Theory Methods* **14** 593–603.

MOORE, D. S. (1977). Generalized inverses, Wald's method and the construction of chi-squared tests of fit. *J. Amer. Statist. Assoc.* **72** 131–137.

MOORE, D. S. and SPRUILL, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.* **3** 599–616.

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

NIKULIN, M. S. (1973). Chi-square test for continuous distributions with shift and scale parameters. *Theory Probab. Appl.* **18** 559–568.

RAO, C. R. and MITRA, S. K. (1971). *Generalized Inverse of Matrices and Its Applications*. Wiley, New York.

RAO, K. C. and ROBSON, D. S. (1974). A chi-square statistic for goodness of fit tests within the exponential family. *Comm. Statist.* **3** 1139–1153.

ROY, A. R. (1956). On χ^2 statistics with variable intervals. Technical Report No. 1, Dept. Statistics, Stanford Univ.

RUYMGAART, F. H. (1974). Asymptotic normality of nonparametric tests for independence. *Ann. Statist.* **2** 892–910.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

SHIRAHATA, S. (1976). Tests for independence in finite contingency tables. *Ann. Statist.* **4** 542–553.

SPRUILL, M. C. (1976). A comparison of chi-square goodness-of-fit tests based on approximate Bahadur slope. *Ann. Statist.* **4** 409–412.

STECK, G. P. (1957). Limit theorems for conditional distributions. *Univ. California Publ. Statist.* **2** 237–284.

TUMANYAN, S. H. (1956). On the asymptotic distribution of the chi-square criterion. *Dokl. Akad. Nauk. SSSR* **94** 1011–1012.

WATSON, G. S. (1957). The χ^2 goodness-of-fit test for normal distributions. *Biometrika* **44** 336–348.

WATSON, G. S. (1958). On chi-square goodness-of-fit tests for continuous distributions. *J. Roy. Statist. Soc. Ser. B* **20** 44–61.

ECONOMETRIC INSTITUTE
UNIVERSITY OF TILBURG
P.O. BOX 90153
TILBURG
THE NETHERLANDS