Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The CWI is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

CWI Tract 48

Asymptotics for generalized chi-square goodness-of-fit tests

F.C. Drost

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

1980 Mathematics Subject Classification: Primary: 62E20, 62F05. Secondary: 60F05, 62E25.
ISBN 90 6196 348 6
NUGI-code: 811

Printed in the Netherlands

PREFACE

This monograph considers the behaviour of various types of chi-square goodness-of-fit test statistics. The first chapter gives a review of recent literature on the subject.

Chapter II investigates the influence of the number of classes k in the presence of a location-scale nuisance parameter. When $k \to \infty$, we prove the asymptotic normality of the Moore-Spruill (1975) class of χ^2 statistics (extending Morris' (1975) theorem for the Pearson statistic when $k \to \infty$ and no nuisance parameters are present). Criteria are developed whether to choose a large or a small number of classes. A theoretical explanation was still lacking for simulations showing that the Rao-Robson-Nikulin test dominates other commonly used χ^2 tests. The present limit theorem implies that, when $k \to \infty$, the Rao-Robson-Nikulin test is better in the sense of Pitman efficiency.

The choice of the location-scale estimator is the subject of the third chapter. Non-robust estimation (i.e. the estimator is not $\sqrt{n}$-consistent under local alternatives) is best: under non-robust estimation general EDF tests, including generalized χ^2 tests, are consistent while the asymptotic local power remains bounded away from one in more classical situations of e.g. ML or robust estimation under heavy-tailed alternatives. Complementary results are given for $\sqrt{n}$-consistent estimators having a relatively large bias or variance under local alternatives. A simulation study illustrates the theoretical results of the second and the third chapter.

The last chapter deals with power approximations for the Cressie-Read (1984) class of χ^2 statistics when no nuisance parameters are present. Although classical (moment-corrected) χ^2 approximations work reasonably well under the hypothesis, non-central χ^2 power approximations are inadequate for moderate sample sizes. A non-local Taylor expansion of the test statistic yields a new approximation based on a weighted sum of independent non-central χ^2 distributions; the distribution error is of order $O(n^{-\frac{1}{2}})$ uniformly in alternatives and levels. Exact power computations for n=20,50 show that the new approximation is very good and is greatly superior to traditional ones.

It is a pleasure to me to express my deep gratitude to prof.dr. J. Oosterhoff and dr. W.C.M. Kallenberg for their constant encouragement and stimulating advice. Also I like to thank prof. D.S. Moore for the stimulating discussions during his visit at the Vrije Universiteit. Although I cannot mention them all, I want to thank everyone who contributed to the pleasant working conditions at the Vrije Universiteit.

I thank Mr. K. Snel for his excellent typing of the manuscript and Mr. D. Zwarst and his technical staff of the Centrum voor Wiskunde en Informatica for the reproduction.

Feike C. Drost

CONTENTS

CHAPTER I

INTRODUCTION

I.1. GOODNESS-OF-FIT TESTS FOR SIMPLE HYPOTHESES

Let $Y_1,\ldots,Y_n$ be n independent identically distributed real-valued random variables with distribution function F^Y and consider the general testing problem $F^Y = H$ for a given distribution H. A well-known omnibus goodness-of-fit test is the classical Pearson chi-square test (cf. Pearson (1900)). The original problem is reduced to a multinomial setting partitioning the range of the Y_j's into k cells $I_{k1},\ldots,I_{kk}$. Let $p_{ki}(0) = P_H(Y_j \in I_{ki})$ denote the cell-probabilities under H and let $\bar{N}_{ki} = \#\{j; Y_j \in I_{ki}\}$ denote the cell counts $(i = 1,\ldots,k)$, then Pearson's chi-square test rejects for large values of

$$\text{(I.1.1)} \qquad \bar{P}_n = \sum_{i=1}^{k} (\bar{N}_{ki} - np_{ki}(0))^2/(np_{ki}(0)).$$

A common competing statistic is based on the likelihood ratio for grouped observations:

$$\overline{LR}_n = 2 \sum_{i=1}^{k} \bar{N}_{ki} \log (\bar{N}_{ki}/(np_{ki}(0))).$$

Cressie and Read (1984) have systematized the theory of multinomial goodness-of-fit tests by considering the class of statistics

$$\text{(I.1.2)} \qquad \overline{CR}_n(\lambda) = \frac{2n}{\lambda(\lambda+1)} \sum_{i=1}^{k} p_{ki}(0)\left\{\left(\frac{\bar{N}_{ki}}{np_{ki}(0)}\right)^{\lambda+1} - 1\right\} \quad (\lambda \in \mathbb{R}\setminus\{-1,0\});$$

for $\lambda = 0$ and $\lambda = -1$ $\overline{CR}_n(\lambda)$ is defined by continuity in λ. This class includes the likelihood ratio statistic $(\lambda = 0)$, the modified likelihood ratio statistic $(\lambda = -1)$ and the statistics of Pearson $(\lambda = 1)$, Freeman-Tukey $(\lambda = -\frac{1}{2})$ and Neyman $(\lambda = -2)$.

Little is known about the exact distributions for small and moderate sample

sizes. As $n \to \infty$ and k is fixed, however, the Cressie-Read statistics have limiting chi-square distributions with k-1 degrees of freedom under H (cf. Cressie and Read (1984)). Esséen (1945) has shown that this approximation is very accurate if $\lambda \approx 1$ (cf. also Yarnold (1972), Larntz (1978), Cressie and Read (1984)). The large sample theory is less satisfactory under alternatives. Local and nonlocal theories lead to different conclusions (cf. Cressie and Read (1984)). Moreover, exact power computations and simulations are not easily explained by these (non)local theories (cf. e.g. West and Kempthorne (1971), Cressie and Read (1984), Kallenberg et al. (1985), Quine and Robinson (1985), Kallenberg (1985)).
In practice the number of classes is taken larger if n is larger: $k = k_n \to \infty$ as $n \to \infty$ (cf. Mann and Wald (1942)). Tumanyan (1956) and Steck (1957) proved the asymptotic normality of Pearson's chi-square statistic under H when $k \to \infty$, $k = \mathcal{O}(n)$. Using a different technique Morris (1975) extended these results to local alternatives for $\overline{P}_n$ and $\overline{LR}_n$. The asymptotically optimal choice of k is investigated in Kallenberg et al. (1985); they obtained simple criteria (based on the information function) for $\overline{P}_n$ and $\overline{LR}_n$. These results generalize to the Cressie-Read tests.

Next consider omnibus goodness-of-fit tests which are based on the raw data instead of on grouped observations. Well-known examples are the Kolmogorov-Smirnov test and the Cramér-von Mises test. They belong to the general class of EDF statistics

$$\overline{T}_n = T(n^{\frac{1}{2}}(F_n(\cdot) - H(\cdot))),$$

which are functionals of the difference between the empirical distribution function F_n of $Y_1,\ldots,Y_n$ and the hypothesized distribution function H. For fixed k the Cressie-Read statistics appear as a special example. In principle the limiting distributions of EDF statistics are known, they are functionals of Brownian bridges (cf. Billingsley (1968)). In practice they are seldom useful.

I.2. COMPOSITE HYPOTHESES

The situation sketched above is rather simple. In actual problems one often encounters p-dimensional nuisance parameters θ and one wants to test the composite null-hypothesis

$$\text{(I.2.1)} \qquad H_0 : F^Y \in \{H^*(\cdot;\theta);\theta \in \Theta \subset \mathbb{R}^p\},$$

where $H^*(\cdot;\theta)$ is a given distribution for each θ.

Several modifications of $\overline{P}_n$ have been proposed. Let $\overline{\theta}_n^{ML} = \overline{\theta}_n^{ML}(\overline{N}_{k1},\ldots,\overline{N}_{kk})$ be the maximum likelihood estimator of θ based on grouped observations and estimate the unknown cell probabilities by the probabilities of $I_{k1},\ldots,I_{kk}$ under $H^*(\cdot;\overline{\theta}_n^{ML})$. Replacing unknown by estimated probabilities in (I.1.1), one obtains the Pearson-Fisher statistic (cf. Fisher (1924)). In a similar way one obtains the statistic of Chernoff and Lehmann (1954), who estimated θ by the maximum likelihood estimator $\hat{\theta}_n^{ML} = \hat{\theta}_n^{ML}(Y_1,\ldots,Y_n)$ under the model H_0. A drawback of these methods is that cells are supposed to be fixed while the distributions vary with θ. This results in widely different cell probabilities. Therefore, when the original observations are available, partition the support of $H^*(\cdot;\theta)$ into k θ-dependent classes $I_{ki}^*(\theta)$ with probabilities $p_{ki}(0) > 0$ $(i = 1,\ldots,k)$ independent of θ. Let

$$(I.2.2) \qquad N_{ki}(\theta) = \#\{j; Y_j \in I_{ki}^*(\theta)\}$$

denote the number of observations in the i-th cell and define the random k-vector $V_k(\theta)$ by its components

$$(I.2.3) \qquad V_{ki}(\theta) = (N_{ki}(\theta) - np_{ki}(0))/(np_{ki}(0))^{\frac{1}{2}} \qquad (i = 1,\ldots,k).$$

Let $\hat{\theta}_n$ be some estimator of θ, then Roy (1956) and Watson (1957,1958) proposed a Pearson type test based on random cells:

$$(I.2.4) \qquad WR_n = \|V_k(\hat{\theta}_n)\|^2$$

($\|\cdot\|$ denotes Euclidian distance). Similarly the Cressie-Read class is generalized to the random-cell situation:

$$(I.2.5) \qquad CR_n(\lambda) = \frac{2n}{\lambda(\lambda+1)} \sum_{i=1}^{k} p_{ki}(0)\left\{\left(\frac{N_{ki}(\hat{\theta}_n)}{np_{ki}(0)}\right)^{\lambda+1} - 1\right\}.$$

Imposing regularity conditions on $\hat{\theta}_n$, the limiting distributions of $WR_n = CR_n(1)$ under H_0 and local alternatives are weighted sums of k independent (noncentral) chi-square variables. The same holds true for the Cressie-Read class because Taylor expansion of $CR_n(\lambda)$ shows that $CR_n(\lambda) - WR_n$ converges to zero in probability (cf. Cressie and Read (1984)). The limiting null distribution of the Chernoff-Lehmann statistic is known to be of this type too but the limiting null distribution of the Pearson-

Fisher statistic is an ordinary chi-square.

To avoid awkward limiting distributions Nikulin (1973) and Rao and Robson (1974) proposed to use a quadratic form

$$(I.2.6) \qquad RRN_n = V_k(\hat{\theta}_n)' \Sigma_k^- V_k(\hat{\theta}_n),$$

where Σ_k^- is a generalized inverse of the asymptotic covariance matrix of $V_k(\hat{\theta}_n)$ under H_0 (cf. (II.1.11)). The Dzhaparidze-Nikulin statistic projects $V_k(\hat{\theta}_n)$ on a suitable linear subspace of $\mathbb{R}^k$ such that all perturbations due to the replacement of θ by $\hat{\theta}_n$ are removed (cf. Dzhaparidze and Nikulin (1974)):

$$(I.2.7) \qquad DN_n = V_k(\hat{\theta}_n)'[I_k - B_k(B_k'B_k)^{-1}B_k']V_k(\hat{\theta}_n)$$

(B_k is defined in (II.1.9)). As opposed to the Cressie-Read class the latter two statistics have limiting (noncentral) chi-square distributions. Moore and Spruill (1975) developed the theory of general quadratic forms:

$$(I.2.8) \qquad MS_n = V_k(\hat{\theta}_n)' \Gamma_k V_k(\hat{\theta}_n),$$

where Γ_k is an arbitrary $(k \times k)$-matrix. They obtained the limiting distributions of MS_n both under H_0 and local alternatives.
For fixed k Spruill (1976) showed that if WR_n and RRN_n are compared by means of the approximate Bahadur slope, RRN_n is uniformly at least as efficient as WR_n. The analogous result for Pitman efficiencies is not true; in several cases local theory implies that WR_n and DN_n are preferable to RRN_n (cf. Drost (1987), also cf. Moore (1977), LeCam et al. (1984)). The simulation studies of Rao and Robson (1974) are better explained by the nonlocal theory than by the local theory; RRN_n generally dominates WR_n.

In the presence of nuisance parameters a natural extension of general EDF statistics is proposed by Chernoff and Lehmann (1954), Neuhaus (1976) and Csörgő and Révész (1981 a). Estimate the unknown distribution function $H^*(\cdot;\theta)$ by $H^*(\cdot;\hat{\theta}_n)$ and apply the classical EDF functionals T to the difference $n^{1/2}(F_n(\cdot) - H^*(\cdot;\hat{\theta}_n))$:

$$(I.2.9) \qquad \tilde{T}_n = T(n^{1/2}(F_n(\cdot) - H^*(\cdot;\hat{\theta}_n))).$$

Durbin (1973) suggested a different approach. Let $\zeta : \mathbb{R} \times \Theta \to \mathbb{R}$ be a transformation such that $\zeta(Y_i;\theta)$ is distributed as H under H_0 if θ is true and replace the original observation Y_i by $Z_i = \zeta(Y_i;\hat{\theta}_n)$ $(i = 1,\ldots,n)$. Then, with $\hat{F}_n$ the empirical distribution function of $Z_1,\ldots,Z_n$, the classical EDF functionals T are applied to $n^{\frac{1}{2}}(\hat{F}_n(\cdot) - H(\cdot))$:

$$(I.2.10) \qquad T_n = T(n^{\frac{1}{2}}(\hat{F}_n(\cdot) - H(\cdot))).$$

Under regularity conditions, including the asymptotic normality of $\hat{\theta}_n$ under H_0 and local alternatives the limiting distributions are functionals of Brownian bridges with parameters generally depending on H and the type of estimator (cf. e.g. Csörgö and Révész (1981 a)). Stephens (1974) suggested useful approximations of these limiting distributions.

To estimate θ it is common practice to employ a maximum likelihood estimator of θ under H_0 or an asymptotically equivalent estimator. Little is known about the influence of the method of estimation on the power of tests. This is not surprising: the asymptotic distributions (under alternatives) are quite complicated for most goodness-of-fit tests in common use and hence, a comparison of (asymptotic) powers under different estimators may not shed much light on the problem.

Quite another statistic is proposed by Csörgö and Révész (1981 b) when the nuisance parameter is a location-scale parameter. Their test is based on spacings between order statistics and is independent of θ. Again the limiting distribution under H_0 is a complex functional of a Brownian bridge; under fixed alternatives the test is consistent.

I.3. OUTLINE OF RESULTS

In the remaining chapters we restrict attention to a location-scale nuisance parameter

$$\theta = (\mu,\sigma)' \text{ and}$$
$$\Theta = \mathbb{R} \times (0,\infty)$$

and consider the testing problem

$$(I.3.1) \qquad H_0 : F^Y \in \mathcal{H} = \{H^*(\cdot;\theta) = H(\tfrac{\cdot-\mu}{\sigma}); \mu \in \mathbb{R}, \sigma > 0\}.$$

In Chapter II the results of Tumanyan (1956), Steck (1957) and Morris (1975) are extended to a subclass of the Moore-Spruill class, including the classical extensions WR_n, RRN_n and DN_n of $\overline{P}_n$. When k tends slowly to infinity these statistics have normal limiting distributions

under H_0 and under local mixture alternatives (Theorem II.2.2). The parameters of the normal distributions are the leading terms of the expectations and variances of the statistics for fixed k. Bickel and Rosenblatt (1973) obtained a similar result for the Watson-Roy statistic when the Fisher information is finite. In the proof of Theorem II.2.2 we rewrite the statistics under consideration as the sum of $\overline{P}_n$ and some remainder terms that are small in probability. Then the desired result follows from Morris (1975). (Without proof several authors claimed that results for fixed k are easily extended to the case where k is of order $\sqrt{n}$. But they seem to have overlooked the problems arising from bounding the remainder terms due to the growing dimension of $V_k(\theta)$.)
As an important consequence of Theorem II.2.2 we show that when $k \to \infty$ the Rao-Robson-Nikulin test generally dominates the statistics DN_n and WR_n in the sense of Pitman (Corollary II.3.1). This partly explains the conflicting results for fixed k between simulation studies of Rao and Robson (1974) and local theory (cf. Moore (1977), LeCam et al. (1984), Drost (1986)).
Theorem II.2.2 implies also that the criterion whether to keep k bounded or to let $k \to \infty$ for the classical Pearson statistic (cf. Kallenberg et al. (1985)) extends to the location-scale nuisance parameter case.

The effect of the estimation procedure on the asymptotic local power is investigated in Chapter III. We show by very crude methods that other estimators than the usual maximum likelihood estimators may lead to a large increase of power for certain interesting classes of alternatives.
First we consider strongly non-robust estimation, i.e. situations where $\hat{\theta}_n$ behaves well under H_0 but is not $\sqrt{n}$-consistent under local alternatives. Then general EDF statistics, including $CR_n(\lambda)$ and MS_n, are consistent (Theorem III.2.1), while the asymptotic local power is bounded away from 1 in more classical situations where $\hat{\theta}_n$ is $\sqrt{n}$-consistent (cf. Durbin (1973), Moore and Spruill (1975), Cressie and Read (1984)).
In the second part of this chapter we assume that $\hat{\theta}_n$ is $\sqrt{n}$-consistent. Then, for special chi-square type tests, we prove oncemore that it is preferable to use non-robust estimators, i.e. estimators which have a relatively large bias or variance under local alternatives (cf. Section III.3).
Theoretical results and numerical evidence suggest that non-robust estimators are best. Robust estimation leads often to a substantial loss in

power for interesting ranges of alternatives while the gain in special directions is relatively small. This explains also simulations of Stephens (1974), who pointed at the high power of several goodness-of-fit tests when parameters are estimated, compared to the same tests with known parameters.

Finally the last chapter deals with power approximations for tests of the Cressie-Read class when no nuisance parameters are present. Although classical (moment-corrected) chi-square approximations work reasonably well under the null-hypothesis for $CR_n(\lambda)$ (cf. Larntz (1978), Cressie and Read (1984)), (moment-corrected) non-central chi-square power approximations are inadequate for moderate sample sizes (cf. Figure IV.3.1). When the power is high a simple more or less accurate approximation is a normal one (cf. Broffitt and Randles (1977)). It gives a crude impression of the power as a function of λ. Quite often, however, the errors are ten percent or more (cf. Figure IV.3.2). In Section IV.2 we present a new approximation based on a weighted sum of independent noncentral chi-square distributions. In Theorem IV.2.1 we show that the accuracy is of order $O(n^{-\frac{1}{2}})$ uniformly in alternatives (local or nonlocal) and levels.
Exact power computations for n = 20 and n = 50 show that the approximations are very good especially in the range $0 \le \lambda \le 2$ (cf. Section IV.3).

CHAPTER II

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

II.1. PRELIMINARIES

II.1.1. Local alternatives

Let $Y_1,\ldots,Y_n$ be i.i.d. real-valued absolutely continuous random variables with distribution function F^Y and consider chi-square type tests (I.2.8) for the testing problem (I.3.1) when $k \to \infty$. It seems out of the question to obtain useful asymptotic properties for the whole class of Moore-Spruill statistics when $k \to \infty$ because the choice of Γ_k depends on k and the class of all $(k\times k)$-matrices is too large when $k \to \infty$. Estimating the nuisance parameter θ disturbs the simple covariance matrix of $V_k(\theta)$ when θ is known. This motivates the following subclass of the Moore-Spruill class:

$$\text{(II.1.1)} \qquad X_n^2 = V_k(\hat{\theta}_n)'[I_k - D_kD_k' + D_k\Lambda(k)D_k']V_k(\hat{\theta}_n),$$

where I_k is the $(k\times k)$-identity matrix, D_k the orthonormal matrix defined in the line preceding (II.1.12) and $\Lambda(k)$ a symmetric nonnegative definite matrix characterizing X_n^2. This subclass contains the classical random-cell generalizations WR_n, RRN_n and DN_n of $\overline{P}_n$. The matrix $I_k - D_kD_k'$ projects $V_k(\hat{\theta}_n)$ on the orthogonal complement of the column space of D_k (col. (D_k)) and removes all noise due to $\hat{\theta}_n$ (cf. Dzhaparidze and Nikulin (1974)). The second part $D_k\Lambda(k)D_k'$ permits a large degree of freedom in directions sensitive to the estimator $\hat{\theta}_n$ (cf. Rao and Robson (1974), Hsuan (1974), McCulloch (1985)).

To study the behaviour of X_n^2 for a broad class of alternatives, let G be any given alternative and consider the contamination family of location-scale distributions

$$\text{(II.1.2)} \qquad \mathcal{G}_{\eta_n} = \{G^*_{\eta_n}(\cdot;\theta) = G_{\eta_n}\left(\frac{\cdot-\mu}{\sigma}\right) = (1-\eta_n)H\left(\frac{\cdot-\mu}{\sigma}\right) + \eta_n G\left(\frac{\cdot-\mu}{\sigma}\right); \mu\in\mathbb{R}, \sigma>0\},$$

where $\eta_n \to 0$ as $n \to \infty$. A common choice of η_n is of type $\eta_n = n^{-\frac{1}{2}}\gamma + o(n^{-\frac{1}{2}})$ for some fixed $\gamma > 0$. When k is bounded this rate results in an asymptotic local power bounded away from one at the local alternative hypothesis

(II.1.3) $\quad H_{1n} : F^Y \in \mathcal{G}_{\eta_n}$

(cf. Moore and Spruill (1975)). We consider general $\eta_n \to 0$ when $k \to \infty$.

II.1.2. Assumptions on the distributions.

Denote the gradient with respect to θ by ∇_θ (∇'_θ transposes ∇_θ), let $E_\eta\{v(Y)\}$ denote the expectation of $v(Y)$ with respect to G_η, let the symbols o, o_p, O and O_p have a componentwise interpretation if they are used for vectors or matrices and put $\theta_0 = (0,1)'$.

We denote the null (alternative) distribution H (G) by its equivalent G_0 (G_1) to unify notation in the remainder of this manuscript. Denote the densities corresponding to $G^*_\eta(x;\theta)$ $(G_\eta(x))$ by $g^*_\eta(x;\theta)$ $(g_\eta(x))$ and assume the following regularity conditions

$$
\text{C.II.1} \quad \begin{cases}
\text{-a} & (\forall\, x,y \in \mathbb{R})\ |g_0(x) - g_0(y)| \le L_0|x-y| \quad \text{for some } L_0 < \infty \\
\text{-b} & \lim\limits_{|x|\to\infty} x g_0(x) = 0 \\
\text{-c} & E_0\{\|\nabla_\theta \log g^*_0(Y;\theta)|_{\theta=\theta_0}\|^2\} < \infty \\
\text{-d} & G_1 \text{ is differentiable} \\
\text{-e} & M_1 = \sup\limits_{x\in\mathbb{R}} g_1(x) < \infty
\end{cases}
$$

The Lipschitz continuity of g_0 implies absolute continuity. Let $g_0^{(1)}$ be a derivative of g_0 with respect to Lebesgue measure; so C.II.1-c is properly defined and implies the finite existence of the Fisher-information matrix $J_\theta = \sigma^2 J$, where

(II.1.4) $$J = E_0[\nabla_\theta \log g^*_0(Y;\theta)|_{\theta=\theta_0}\ \nabla'_\theta \log g^*_0(Y;\theta)|_{\theta=\theta_0}].$$

The conditions C.II.1-a,e imply that g_η is bounded

(II.1.5) $$M_\eta = \sup_{x\in\mathbb{R}} g_\eta(x) < \infty \quad (0 \le \eta \le 1).$$

II.1.3. Assumptions on the estimator $\hat{\theta}_n$.

Suppose $\hat{\theta}_n$ is location-scale equivariant and admits the pointwise representation

$$\text{(II.1.6)} \qquad n^{\frac{1}{2}}(\hat{\theta}_n - \theta) = n^{-\frac{1}{2}}\sigma \sum_{j=1}^{n} h\left(\frac{Y_j-\mu}{\sigma}\right) + \sigma Q_n\left(\frac{Y_1-\mu}{\sigma},\ldots,\frac{Y_n-\mu}{\sigma}\right),$$

where $h = (h_1,h_2)' : \mathbb{R} \to \mathbb{R}^2$ is the vector-valued influence function and $Q_n = (Q_{1n},Q_{2n})' : \mathbb{R}^n \to \mathbb{R}^2$ the remainder. The influence function h often coincides with the influence curve (cf. Huber (1981) or Serfling (1980)). Assume

$$\text{C.II.2} \qquad \begin{cases} \text{-a} & E_0\{h(Y)\} = 0 \\ \text{-b} & E_0[h(Y)h(Y)'] = A^{-1} \\ \text{-c} & E_1\{h(Y)\} = 0 \text{ if } n^{\frac{1}{2}}\eta_n \text{ is unbounded} \\ \text{-d} & E_1\{\|h(Y)\|^2\} < \infty \\ \text{-e} & Q_n(Y_1,\ldots,Y_n) = O_p(1) \text{ under } G_0 \text{ and } G_{\eta_n} \end{cases}$$

where A is a finite nonsingular matrix. Sufficient for (II.1.6) to hold under $H_{1n}(n^{\frac{1}{2}}\eta_n = O(1))$ is that (II.1.6) holds under H_0 and $\int (g_1/g_0 - 1)^2 dG_0 < \infty$, because under the latter condition the joint distributions of $Y_1,\ldots,Y_n$ under H_{1n} and under H_0 are contiguous (cf. Oosterhoff and van Zwet (1979) and Oosterhoff (1985)). However, the integral condition is not satisfied for heavy-tailed alternatives as shown in Kallenberg et al. (1985). Of course (II.1.6) and C.II.2 may still hold under H_{1n} for particular estimators in such cases too. Condition C.II.2 implies that $\hat{\theta}_n$ is $\sqrt{n}$-consistent under H_0 and H_{1n} (if $Q_n = o_p(1)$ it even implies the asymptotic normality of $\hat{\theta}_n$)

$$\text{(II.1.7)} \qquad n^{\frac{1}{2}}(\hat{\theta}_n - \theta) = O_p(1) \quad \text{under } H_0 \text{ and } H_{1n}.$$

In Chapter III we shall consider the case $n^{\frac{1}{2}}\|\hat{\theta}_n - \theta\| \to \infty$ under H_{1n}. Under regularity conditions Bickel (1982) (cf. also Hájek (1972)) showed that the maximum likelihood estimator $\hat{\theta}_n^{ML}$ admits the representation (II.1.6) with $Q_n = o_p(1)$ and

$$\text{(II.1.8)} \qquad h^{ML}(x) = J^{-1}\nabla_\theta \log g_0^*(x;\theta)|_{\theta=\theta_0} = -J^{-1}(g_0^{(1)}(x)/g_0(x), 1+xg_0^{(1)}(x)/g_0(x))',$$

implying A = J.

II.1.4. Definitions and notations.

A natural choice for the cell-boundaries in the location-scale model is of type $\mu + a\sigma$. Let $\{a_k\}_{k\geq 2}$ be a sequence of (k+1)-vectors with components $-\infty = a_{k0} < \ldots < a_{kk} = \infty$ inducing a partition of $\mathbb{R}$ into k disjoint intervals (for each θ)

$$I^*_{ki}(\theta) = (\mu + a_{ki-1}\sigma, \mu + a_{ki}\sigma] \qquad (i = 1,\ldots,k).$$

Let I_{ki} be the interval

$$I_{ki} = I^*_{ki}(\theta_0) = (a_{ki-1}, a_{ki}]$$

and put

$$p^*_{ki}(\eta,\theta^*,\theta) = \int_{I^*_{ki}(\theta^*)} dG^*_\eta(x;\theta)$$

the probability of $I^*_{ki}(\theta^*)$ under $G^*_\eta(x;\theta)$ $(i=1,\ldots,k)$. The cell-probabilities are independent of $\theta = \theta^*$

$$p_{ki}(\eta) \equiv p^*_{ki}(\eta,\theta,\theta) = \int_{I_{ki}} dG_\eta(x) \qquad (i=1,\ldots,k).$$

Define the (k×2)-matrices $B_k = [B_{1k}, B_{2k}]$ and $C_k = [C_{1k}, C_{2k}]$ by their i-th rows

$$\begin{aligned} B_{ki} &= p_{ki}^{-\frac{1}{2}}(0)\nabla'_\theta p^*_{ki}(0,\theta_0,\theta)\big|_{\theta=\theta_0} \\ &= p_{ki}^{-\frac{1}{2}}(0)[g_0(a_{ki-1}) - g_0(a_{ki}), a_{ki-1}g_0(a_{ki-1}) - a_{ki}g_0(a_{ki})] \qquad \text{(II.1.9)} \\ C_{ki} &= (p_{ki}^{-\frac{1}{2}}(0) \int_{I_{ki}} h(x)'dG_0(x))\cdot A \end{aligned}$$

$(i=1,\ldots,k)$. Direct calculation shows that the rows of B_k can also be written as

$$B_{ki} = p_{ki}^{-\frac{1}{2}}(0) \int_{I_{ki}} \nabla'_\theta \log g^*_0(x;\theta)\big|_{\theta=\theta_0} dG_0(x) \qquad (i=1,\ldots,k).$$

Note that $C_k = B_k$ when using $\hat\theta_n^{ML}$ (with h^{ML} given by (II.1.8)). Using Lemma A of Kallenberg et al. (1985), the Cauchy-Schwarz inequality and the finite existence of J and A it is easily seen that in the general case

$$\text{(II.1.10)} \qquad B'_kB_k = \mathcal{O}(1) \quad \text{and} \quad C'_kC_k = \mathcal{O}(1) \quad \text{as } k \to \infty.$$

Straightforward calculation shows that the asymptotic covariance matrix under H_0 of the non-vanishing part of $V_k(\hat{\theta}_n)$ is given by (cf. Remark II.4.1)

$$\text{(II.1.11)} \qquad \Sigma_k = I_k - q_k q_k' + (B_k - C_k)A^{-1}(B_k - C_k)' - C_k A^{-1} C_k',$$

where $q_k = (p_{k1}^{\frac{1}{2}}(0), \ldots, p_{kk}^{\frac{1}{2}}(0))'$. Σ_k does not depend upon the location-scale parameter. Note that $\|q_k\|^2 = 1$, $q_k' B_k = q_k' C_k = 0$. Let D_k be a matrix with orthonormal columns such that col. (D_k) = col. $([B_k, C_k])$ and let $\Psi(k)$ be the nonnegative definite matrix

$$\text{(II.1.12)} \qquad \Psi(k) = D_k' \, \Sigma_k \, D_k.$$

Then, substituting (II.1.12) in the RHS of (II.1.13),

$$\text{(II.1.13)} \qquad \Sigma_k = I_k - q_k q_k' - D_k D_k' + D_k \Psi(k) D_k'.$$

Define the k-vector $d_k(\eta_n)$ by its components

$$\begin{aligned} d_{ki}(\eta_n) &= n^{\frac{1}{2}}(p_{ki}(\eta_n) - p_{ki}(0))/p_{ki}^{\frac{1}{2}}(0) \\ &= n^{\frac{1}{2}}\eta_n(p_{ki}(1) - p_{ki}(0))/p_{ki}^{\frac{1}{2}}(0) \qquad (i = 1, \ldots, k). \end{aligned}$$

Finally define $\Delta_k^*(\eta_n)$, the noncentrality parameter $\Delta_k(\eta_n)$, the location parameter $m_k(\eta_n)$ and the variance parameter $s_k^2(\eta_n)$ by

$$\text{(II.1.14)} \qquad \begin{aligned} \Delta_k^*(\eta_n) &= \|\Lambda(k)^{\frac{1}{2}} D_k'(d_k(\eta_n) - B_k n^{\frac{1}{2}}\eta_n E_1\{h(Y)\})\|^2 \\ \Delta_k(\eta_n) &= \|[I_k - D_k D_k'](d_k(\eta_n) - B_k n^{\frac{1}{2}}\eta_n E_1\{h(Y)\})\|^2 + \Delta_k^*(\eta_n) \\ m_k(\eta_n) &= k + \Delta_k(\eta_n) \text{ and} \\ s_k^2(\eta_n) &= 2k + 4\Delta_k(\eta_n). \end{aligned}$$

The last two parameters are the leading terms of the expectation and the variance of X_n^2. Note that $\Delta_k(0) = 0$, thus $m_k(0) = k$ and $s_k^2(0) = 2k$.

II.1.5. Assumptions on the rate of k.

Let $k = k(n)$ be a particular choice of the number of cells and assume (under H_0 put $\eta_n = 0$)

$$\text{C.II.3}\quad\begin{cases} \text{-a} & k \to \infty \text{ and } \eta_n \to 0 \text{ as } n \to \infty \\ \text{-b} & \lim\limits_{k\to\infty} \max\limits_{1\le i\le k} p_{ki}(0) = 0 \\ \text{-c} & \lim\limits_{n\to\infty} \max\limits_{1\le i\le k} d^2_{ki}(\eta_n)/(k + \|d_k(\eta_n)\|^2) = 0 \\ \text{-d} & (n^{-\frac{1}{2}} + \eta_n)(1 + k^{-\frac{1}{2}}\lambda_k) \log^{\frac{3}{2}} k \sum\limits_{i=1}^{k} p^{-1}_{ki}(0) = o(1) \end{cases}$$

where λ_k is the maximum eigenvalue of $\Lambda(k)$. Condition C.II.3-b puts a more or less natural restriction on the cell-width as $k \to \infty$ and implies

$$\lim_{n\to\infty} \max_{1\le i\le k} p_{ki}(\eta_n) = 0$$

Condition C.II.3-c is the *uan* (uniformly asymptotically negligible) condition of Morris (1975). A better grasp of C.II.3-d is obtained by looking at the special case $\eta_n = n^{-\frac{1}{2}}\gamma + o(n^{-\frac{1}{2}})$ and $p_{ki}(0) = k^{-1}$ $(i = 1,\ldots,k)$. Then C.II.3-d reduces to

$$n^{-\frac{1}{2}}(1 + k^{-\frac{1}{2}}\lambda_k)k^2 \log^{\frac{3}{2}} k = o(1).$$

For the test statistics with $\lambda_k = O(k^{\frac{1}{2}})$ this implies that the maximum number of cells is slightly less then $n^{\frac{1}{4}}$; examples are WR_n and DN_n (cf. Section II.3). In other important examples where $\lambda_k \simeq k$ this bound reduces to $n^{\frac{1}{5}}$ (cf. RRN_n in Examples II.3.3 - II.3.5). We end this subsection with some technical conditions (under H_0 put $\eta_n = 0$)

$$\text{C.II.4}\quad\begin{cases} \text{-a} & Q_n(Y_1,\ldots,Y_n) = o_p(k^{\frac{1}{4}}/(1 + \lambda_k)^{\frac{1}{2}}) \text{ under } G_{\eta_n} \\ \text{-b} & \text{tr.}(\Lambda(k)^{\frac{1}{2}}\Psi(k)\Lambda(k)^{\frac{1}{2}}) = o(k^{\frac{1}{2}}) \\ \text{-c} & D'_k d_k(\eta_n) = o(s_k(\eta_n)) \\ \text{-d} & (n^{-\frac{1}{2}} + \eta_n) \max\{a^4_{k1}, a^4_{kk-1}\} \log^{-\frac{3}{2}} k = O(1) \end{cases}$$

The curious condition C.II.4-a is often implied by C.II.3-d because the remainder term Q_n is usually of order $O_p(n^{-\frac{1}{4}})$ (Serfling (1980), Ch.2 proves this for estimators based on quantiles; for regular estimators one even expects $Q_n = O_p(n^{-\frac{1}{2}})$).

In the Appendix II.5 it is shown that $D'_k d_k(\eta_n)/(n^{\frac{1}{2}}\eta_n)$ is a kind of Riemann-Stieltjes sum approximating an integral. Under very restrictive conditions this approximation is quite accurate, but even if the conditions of

Lemma II.5.1 are not satisfied the accuracy is often of order $o(1)$, implying $D_k' d_k(\eta_n) = O(n^{\frac{1}{2}}\eta_n)$. The last condition puts a restriction on the extreme cell-probabilities. Often the tail-behaviour of g_0 bounds the a_{ki}'s by log k, in which case C.II.3-d implies C.II.4-d.

II.1.6. Complementary remarks.

Consider statistics which do not depend on h through C_k and suppose $\lambda_k = o(k^{\frac{1}{2}})$. Then the representation (II.1.6) is not necessary to obtain Theorem II.2.2 but (II.1.7) suffices. Taking, however, h = 0 and $A = A^{-1} = 0$ (reducing a lot of terms to zero in this section) we can incorporate these cases in the framework of (II.1.6). Thus, assuming $\lambda_k = o(k^{\frac{1}{2}})$ and (II.1.7), we omit the conditions C.II.2 and C.II.4-a,b (to delete C.II.4-b use $\mathrm{tr}.\Lambda(k) = O(\lambda_k) = o(k^{\frac{1}{2}})$ and derive from (II.1.10) that $\mathrm{tr}.\Psi(k) = O(1)$).

Finally note that the conditions C.II.3-b,c and C.II.4-c can be replaced by

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

if $\eta_n = n^{-\frac{1}{2}}\gamma + o(n^{-\frac{1}{2}})$ and if all cells are equiprobable (use $\|d_k(\eta_n)\|^2 = o(k)$).

II.2. MAIN RESULTS

In this section the limiting null and alternative distributions of the test statistic X_n^2 are given for the testing problem H_0 versus H_{1n}. In the proofs it is sufficient to restrict attention to the special choice $\theta_0 = (0,1)'$ of θ because the distribution of X_n^2 is invariant with respect to θ.

<u>PROPOSITION II.2.1</u>. (Morris (1975)) *Assume* C.II.3, *then*

$$(\overline{P}_n - k)/(2k)^{\frac{1}{2}} \to_{d_0} N(0,1) \text{ and}$$
$$(\text{II.2.1}) \qquad (\overline{P}_n - (k + \|d_k(\eta_n)\|^2))/(2k + 4\|d_k(\eta_n)\|^2)^{\frac{1}{2}} \to_{d_{1n}} N(0,1).$$

<u>PROOF</u>. Under H_{1n}. Let μ_k and σ_{ik}^2 be the expressions appearing in (5.5) and (5.7) of Morris (1975). Straightforward calculation shows $\mu_k = k + \|d_k(\eta_n)\|^2 + o(1)$ and $\Sigma_{i=1}^k \sigma_{ik}^2 = (2k + 4\|d_k(\eta_n)\|^2)(1 + o(1))$. The conditions of Theorem 5.1 of Morris (1975) are directly implied by our

conditions. Application of this theorem yields the desired result. □

REMARK II.2.1. The conditions of Proposition II.2.1 can be relaxed. For more detailed results about $\overline{P}_n$ we refer to Morris (1975) and Kallenberg et al. (1985). □

THEOREM II.2.2. *Consider the statistics* X_n^2 *for testing* H_0 *against the family of alternatives* (II.1.2) *determined by* G_1. *Assume* C.II.1 - C.II.4, *then*

(II.2.2a) $(X_n^2 - k)/(2k)^{\frac{1}{2}} \to_{d_0} N(0,1)$,

(II.2.2b) $(X_n^2 - m_k(\eta_n))/s_k(\eta_n) \to_{d_{1n}} N(0,1)$ *if* $\limsup_{n\to\infty} \Delta_k^*(\eta_n)/s_k(\eta_n) < \infty$ *and*

(II.2.2c) $(X_n^2 - k)/(2k)^{\frac{1}{2}} \to_{P_{1n}} \infty$ *if* $\lim_{n\to\infty} \Delta_k^*(\eta_n)/s_k(\eta_n) = \infty$.

PROOF. cf. Section II.4. □

REMARK II.2.2. Obviously Theorem II.2.2 continues to hold if θ is either a location or a scale parameter. The proof requires some slight modifications in notation. □

In the remainder of this section we state some corollaries concerning the number of classes, the relative efficiency of test statistics of type (II.1.1) and the choice of estimators.

It is common practice to choose the parameter η_n such that the asymptotic local power is bounded away from α and 1. So the additional condition on $\Delta_k^*(\eta_n)$ to obtain limiting normal distributions in Theorem II.2.2 is quite natural, since otherwise there exists a subsequence of $\{X_n^2\}$ for which the power tends to one. This is further elaborated in Corollary II.2.3; the ratio of the noncentrality parameter $\Delta_k(\eta_n)$ and the square root of k determines the asymptotic power

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

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

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

COROLLARY II.2.3. *Assume* C.II.1 - C.II.4, *then*

$$\text{(II.2.3)} \qquad \lim_{n\to\infty} \beta_\alpha(X_n^2,n,\eta_n) = \begin{cases} \alpha \\ \\ 1 \end{cases} \quad \textit{iff} \ \lim_{n\to\infty} \Delta_k(\eta_n)/k^{\frac{1}{2}} = \begin{cases} 0 \\ \\ \infty \end{cases}$$

PROOF. The critical values of the test X_n^2 satisfy

$$c_k = k + (2k)^{\frac{1}{2}}\xi_\alpha + o(k^{\frac{1}{2}}),$$

where $\xi_\alpha = \Phi^{-1}(1-\alpha)$ denotes the upper α-point of the standard normal distribution function Φ. Because every subsequence of $\Delta_k^*(\eta_n)/s_k(\eta_n)$ has a further sequence with a limit (finite or infinite) we assume without loss of generality that the sequence $\Delta_k^*(\eta_n)/s_k(\eta_n)$ has a limit. If $\lim_{n\to\infty} \Delta_k^*(\eta_n)/s_k(\eta_n) < \infty$ apply

$$\beta_\alpha(X_n^2,n,\eta_n) = P_{1n}(X_n^2 > c_k) =$$

$$= P_{1n}((X_n^2 - m_k(\eta_n))/s_k(\eta_n) > -\Delta_k(\eta_n)/s_k(\eta_n) + \xi_\alpha(2k)^{\frac{1}{2}}/s_k(\eta_n) + o(1))$$

$$= \begin{cases} \alpha \\ \\ 1 \end{cases} \text{iff } \Delta_k(\eta_n)/s_k(\eta_n) \to \begin{cases} 0 \\ \\ \infty \end{cases}$$

(using (II.2.2b)) and otherwise apply

$$\beta_\alpha(X_n^2,n,\eta_n) = P_{1n}(X_n^2 > c_k) = P_{1n}((X_n^2 - k)/(2k)^{\frac{1}{2}} > \xi_\alpha + o(1)) \to 1$$

(using (II.2.2c)). Combination of these two results yields (II.2.3). □

REMARK II.2.3. Assume C.II.1 - C.II.4, then Corollary II.2.3 implies

$$\lim_{n\to\infty} \Delta_k(n^{-\frac{1}{2}})/k^{\frac{1}{2}} = \begin{cases} 0 \\ \\ \infty \end{cases} \Rightarrow \begin{matrix} \text{asymptotic local power} \\ \\ \text{of } X_n^2 \text{ highest for} \end{matrix} \begin{cases} \text{bounded } k \\ \\ k \to \infty \end{cases}$$

because the choice $\eta_n = n^{-\frac{1}{2}}\gamma + o(n^{-\frac{1}{2}})$ results in an asymptotic local power between α and 1 for bounded k. □

To evaluate the relative efficiency of test statistics of type (II.1.1) we introduce some more notation. Let $S_k^{(i)}$ be a statistic of type (II.1.1) induced by the matrix $\Lambda(k)^{(i)}$ and the estimator $\hat{\theta}_n^{(i)}$ $(i = 1,2)$ and define the sequence $n_1(n)$

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

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 II.2.4 it is shown that the relative performance of $S_k^{(1)}$ and $S_k^{(2)}$ only depends upon the ratio of their noncentrality parameters $\Delta_k^{(1)}(\eta_n)$ and $\Delta_k^{(2)}(\eta_n)$.

<u>COROLLARY II.2.4</u>. *Consider a nonincreasing sequence* $\{\eta_n\}$ *such that* $n\eta_n^2$ *varies slowly as* $n \to \infty$. *For* $S_k^{(1)}$ *assume* $\Delta_k^{(1)}(\eta_n)/k^{\frac{1}{2}} \to c_1 > 0$ *and assume* C.II.1 - C.II.4 *under* H_0 *and* H_{1n} *(with contamination factor* η_n*). For* $S_k^{(2)}$ *assume* $\Delta_k^{(2)}(\eta_n)/k^{\frac{1}{2}} \to c_2 < \infty$ *and assume* C.II.1 - C.II.4 *under* H_0 *and* H_{1n} *(for all contamination factors* $\tau_n = \eta_{m^{-1}(n)}$ *with* $m(n) \to \infty$ *as* $n \to \infty$ *and such that* $\limsup_{n\to\infty} m(n)/n < \delta^* + c_1/c_2$ *for some small fixed* $\delta^* > 0$).
Then

$$\text{(II.2.4)} \qquad e_p(S^{(1)},S^{(2)}) = c_1/c_2.$$

<u>PROOF</u>. Let $a_n \approx b_n$ have the interpretation: $(\forall\, \varepsilon > 0)\ (\exists\, n_0)(\forall\, n > n_0)$ $|a_n - b_n| < \varepsilon$. Note that $\beta_\alpha(S_k^{(1)},n,\eta_n) \to \Phi(c_1 - \xi_\alpha) > \alpha$. First assume $0 < c_1, c_2 < \infty$ and let $\{m = m(n)\}$ be a sequence such that $m \le (-\delta + c_1/c_2)n = \bar{n}$ for some $0 < \delta < c_1/c_2$. If m remains bounded

$$\beta_\alpha(S^{(2)}_{k(m)},m,\eta_n) \approx \alpha < \Phi(c_1 - \xi_\alpha) \approx \beta_\alpha(S_k^{(1)},n,\eta_n),$$

otherwise

$$\beta_\alpha(S^{(2)}_{k(m)},m,\eta_n) \approx \Phi\left(\frac{\eta_n^2}{\eta_m^2}\,\frac{\Delta_{k(m)}(\eta_m)}{k^{\frac{1}{2}}(m)} - \xi_\alpha\right)$$

$$\le \Phi\left(\frac{\bar{n}}{n}\,\frac{n\eta_n^2}{\bar{n}\eta_{\bar{n}}^2}\,\frac{\Delta_{k(m)}(\eta_m)}{k^{\frac{1}{2}}(m)} - \xi_\alpha\right) \approx \Phi((-\delta + c_1/c_2)c_2 - \xi_\alpha)$$

$$< \Phi(c_1 - \xi_\alpha) \approx \beta_\alpha(S_k^{(1)},n,\eta_n).$$

Similarly one proves for $m(n) = (\delta + c_1/c_2)n \quad (0 < \delta < \delta^*)$

$$\beta_\alpha(S^{(2)}_{k(m)},m,\eta_n) \approx \Phi((\delta + c_1/c_2)c_2 - \xi_\alpha)$$

$$> \Phi(c_1 - \xi_\alpha) \approx \beta_\alpha(S_k^{(1)},n,\eta_n)$$

and thus $e_p(S^{(1)},S^{(2)}) = c_1/c_2$. Next consider sequences $\{m = m(n)\}$ such that $m \le Mn = \bar{n}$ for some $M \in \mathbb{R}$. Then, if $c_1 = \infty$,

$$\beta_\alpha(S^{(2)}_{k(m)},m,\eta_n) < 1 \approx \beta_\alpha(S^{(1)}_k,n,\eta_n)$$

and, if $c_2 = 0$,

$$\beta_\alpha(S^{(2)}_{k(m)},m,\eta_n) \approx \alpha < \Phi(c_1 - \xi_\alpha) \approx \beta_\alpha(S^{(1)}_k,n,\eta_n).$$

Hence $e_p(S^{(1)},S^{(2)}) = \infty$. □

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

The noncentrality parameter of X^2_n is heavily influenced by the choice of the estimator via the influence function h. The relative performance of X^2_n for different kinds of estimators can be calculated from Corollary II.2.4, cf. subsection III.3.2 for a more detailed discussion.

II.3. CLASSICAL EXTENSIONS OF $\overline{P}_n$ AND EXAMPLES

II.3.1. Asymptotic distributions.

In this section we investigate several properties of the statistics WR_n, RRN_n and DN_n (cf. Section I.2). The Rao-Robson-Nikulin statistic is not precisely defined in (I.2.6) 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. If $k \to \infty$ the choice of Σ_k^- is more delicate (cf. Example II.3.2). Therefore we restrict attention to the Moore-Penrose generalized inverse Σ_k^+ of Σ_k; from now on we assume that

$$\text{(II.3.1)} \qquad RRN_n = V_k(\hat{\theta}_n)' \, \Sigma_k^+ \, V_k(\hat{\theta}_n).$$

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

$$\text{(II.3.2)} \qquad \widetilde{DN}_n = V_k(\hat{\theta}_n)'[I_k - D_k D_k']V_k(\hat{\theta}_n),$$

which projects $V_k(\hat{\theta}_n)$ on the linear subspace of $\mathbb{R}^k$ orthogonal to col.(D_k). Note that $\widetilde{DN}_n = DN_n$ when using the maximum likelihood estimator $\hat{\theta}^{ML}_n$ and the influence function h^{ML} of Bickel (1982) (cf. (II.1.8)). In Table II.3.1

the limiting null-distributions of the statistics WR_n, RRN_n, DN_n and $\widetilde{DN}_n$ are given when k is fixed. Note that these statistics belong to the class of statistics (II.1.1).

Table II.3.1. Classical extensions of $\overline{P}_n$.

X_n^2	$\Lambda(k)$	limiting null-distribution
WR_n	$I_{r(D_k)}$	$\chi^2_{k-1-r(D_k)} + \sum_{i=1}^{r(D_k)} \lambda_{ki}\chi^2_{1i}$ *1)
RRN_n	$\Psi(k)^+$	$\chi^2_{r(\Sigma_k)}$ *2)
DN_n	$I_{r(D_k)} - z(z'z)^{-1}z'$, *3)	$\chi^2_{k-1-r(B_k)}$
$\widetilde{DN}_n$	0	$\chi^2_{k-1-r(D_k)}$

*1) λ_{ki} $(i=1,\dots,r(D_k))$ are the eigenvalues of $\Psi(k)$

*2) $r(\Sigma_k) = k-1-r(D_k)+r(\Psi(k))$

*3) z is a $(r(D_k)\times r(B_k))$-matrix such that $\text{col.}(D_k z) = \text{col.}(B_k)$

Let $k \to \infty$ as $n \to \infty$. Then the conditions C.II.4-a,b are trivially satisfied for the statistics WR_n, DN_n and $\widetilde{DN}_n$ because $\lambda_k \le 1 = o(k^{\frac{1}{2}})$. Taking $h = 0$, $A = A^{-1} = 0$ condition C.II.2 is also automatically fulfilled for the Watson-Roy statistic and the Dzhaparidze-Nikulin statistic (cf. subsection II.1.6). This choice is not possible for the modified Dzhaparidze-Nikulin statistic because we essentially use a non-null function h in the definition of D_k (C_k). Note also that C.II.4-b is satisfied for RRN_n.

The Watson-Roy test also appears in density estimation theory. As a particular case Bickel and Rosenblatt (1973) obtained limiting distributions of this type when the Fisher information is finite. In Example II.3.1 these results are derived from Theorem II.2.2.

<u>EXAMPLE II.3.1</u>. Put $p_{ki}(0) = k^{-1}$ $(i=1,\dots,k)$, $\eta_n = n^{-\frac{1}{2}}k^{\frac{1}{4}}\gamma + o(n^{-\frac{1}{2}}k^{\frac{1}{4}})$ and assume C.II.1, C.II.3-c, C.II.4-c,d, (II.1.7) and $n^{-\frac{1}{2}}k^{\frac{9}{4}}\log^{\frac{3}{2}} k \to 0$ as $k \to \infty$ and $n \to \infty$. Then

$$(WR_n - k)/(2k)^{\frac{1}{2}} \to_{d_0} N(0,1) \text{ and}$$

$$(WR_n - k)/(2k)^{\frac{1}{2}} \to_{d_{1n}} N\left(\frac{\gamma^2}{\sqrt{2}} I(g_1,2),1\right),$$

where $I(g_1,2)$ is the Fisher information (cf. (II.3.7)).

PROOF. Lemma A1 of Kallenberg et al. (1985) shows $\|d_k(\eta_n)\|^2/(n\eta_n^2) \to I(g_1,2)$. Application of Theorem II.2.2 yields the statements. □

For such distant alternatives as considered in Example II.3.1 the asymptotic local power is bounded away from 1 if the Fisher information is finite. Indeed, in view of Remark II.2.3, we recommend bounded k in our testing problem (I.3.1).

The second example shows that special sequences for the generalized inverse can have a disastrous effect on RRN_n. This is due to the fact that one has to choose a generalized inverse for every k.

EXAMPLE II.3.2. Consider the Laplace null-hypothesis with unknown location

$$H_0 : f^Y \in \{g_0(x) = \tfrac{1}{2}\exp(-|x-\mu|);\mu \in \mathbb{R}\}.$$

Put $p_{ki}(0) = k^{-1}$ $(i=1,\ldots,k)$, let $\hat{\mu}_n = \text{med}(Y_1,\ldots,Y_n)$ and let $k = k(n)$ be a sequence of even numbers tending to infinity such that $k^2 \log^{\frac{3}{2}} k = o(n^{\frac{1}{2}})$. Then the limiting null-distributions of WR_n and $DN_n = RRN_n$ (with the Moore-Penrose generalized inverse) are asymptotically normal with parameters k and 2k.

In general the asymptotic distribution of RRN_n depends upon the choice of Σ_k^-.

PROOF. Under H_0 the median satisfies condition C.II.2 with $h^{ML}(x) = \text{sgn}(x)$ and $Q_n = O_p(n^{-\frac{1}{4}})$ (cf. Serfling (1980), Ch. 2) and, because $B_k'B_k = J = 1$, we obtain $\Sigma_k = \Sigma_k^+ = I_k - q_k q_k' - q_k^*(q_k^*)'$, with $q_k^* = (-k^{-\frac{1}{2}},\ldots,-k^{-\frac{1}{2}},k^{-\frac{1}{2}},\ldots,k^{-\frac{1}{2}})'$. C.II.1 - C.II.4 are easily verified using $|a_{k1}| = a_{kk-1} = \log(k/2)$. Theorem II.2.2 yields the first statement.

To prove that the asymptotic distribution of RRN_n depends upon the choice of Σ_k^- we consider the generalized inverse

$$\Sigma_k^- = I_k + knq_k^*(q_k^*)'$$

of Σ_k. Then

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

Thus $(RRN_n^* - k)/(2k)^{\frac{1}{2}}$ does not converge to a normal distribution. □

Note that the distribution of RRN_n does not depend on the choice of Σ_k^- if $r(\Sigma_k) = k-1$. Although $r(\Sigma_k) < k-1$ seems rather pathological it quite often occurs when using estimators based on quantiles, e.g. the median or the interquartile range.

In Example II.3.3 we consider a normal null-hypothesis against a Cauchy alternative. Although $n^{\frac{1}{2}}(\overline{Y}_n - \mu) = O_p(1)$ does not hold at the fixed alternative, it is satisfied under the sequence of local alternatives.

EXAMPLE II.3.3. Consider the testing problem of a normal null-hypothesis with unknown location

$$H_0 : F^Y \in \{\Phi(x-\mu); \mu \in \mathbb{R}\}$$

against the local alternative

$$H_{1n} : F^Y \in \{(1-\eta_n)\Phi(x-\mu) + \eta_n(\tfrac{1}{2} + \tfrac{1}{\pi} \arctan(x-\mu)); \mu \in \mathbb{R}\}.$$

Put $p_{ki}(0) = k^{-1}$ $(i = 1,\dots,k)$, $\eta_n = n^{-\frac{1}{2}}\gamma + o(n^{-\frac{1}{2}})$, let $\hat{\mu}_n = n^{-1} \Sigma_{j=1}^n Y_j$ and let the sequence $k = k(n) \to \infty$ such that $k^{\frac{5}{2}} \log^{\frac{5}{2}} k = o(n^{\frac{1}{2}})$. Then $\lim_{n\to\infty} \beta_\alpha(WR_n,n,\eta_n) = \lim_{n\to\infty} \beta_\alpha(DN_n,n,\eta_n) = \lim_{n\to\infty} \beta_\alpha(RRN_n,n,\eta_n) = 1$.

PROOF. First we derive the limiting distributions of WR_n, RRN_n and DN_n under H_0. The mean estimator fulfills C.II.2 with $h^{ML}(x) = x$. Furthermore note that $|a_{k1}| = a_{kk-1} \le (2 \log k)^{\frac{1}{2}}$ and $\lambda_k^{RRN} = 1/(1 - B_k'B_k) \le$ $\le 2k \log k$. From Theorem II.2.2 it follows that WR_n, RRN_n and DN_n are asymptotically normal with parameters k and 2k. With the previous choice of h condition C.II.2 is violated under local alternatives. Note, however, that $n^{\frac{1}{2}}(\hat{\mu}_n - \mu) \to_{d_{1n}} N(0,1) + \gamma\, Cauchy = O_p(1)$. Application of Remark II.4.2 yields the limiting distributions of WR_n and DN_n under local alternatives. Corollary II.2.3, $RRN_n \ge \widetilde{DN}_n = DN_n$, $(B_k'B_k)^{-\frac{1}{2}}B_k'd_k(\eta_n) = 0$ and $\Delta_k^{WR}(\eta_n) = \Delta_k^{DN}(\eta_n) \ge n^{\frac{1}{2}}\eta_n(p_{k1}(1) - 1/k)^2 k \ge ck/\log k$ (for some $c > 0$) imply the desired statements. □

II.3.2. Pitman efficiencies.

Let $\hat{\theta}_n$ be a particular choice for the estimator and assume that the same influence function h is used for each of the statistics under consideration. McCulloch (1985) proved that if one uses $\hat{\theta}_n^{ML}$ the Rao-Robson-Nikulin statistic is the sum of the Dzhaparidze-Nikulin statistic and the positive statistic for testing normality which is proposed by Hsuan (1974). This result generalizes to

$$(\text{II.3.3}) \qquad \widetilde{DN}_n \le DN_n \le WR_n; \quad \widetilde{DN}_n \le RRN_n.$$

Similar relations are true for the corresponding noncentrality parameters.

COROLLARY II.3.1. *Assume the conditions of Corollary* II.2.4 *for the statistics* $\widetilde{DN}_n$, DN_n, WR_n *and* RRN_n *and suppose* $D_k' d_k(\eta_n) = o(k^{\frac{1}{4}})$, *then*

$$(\text{II.3.4}) \qquad e_p(\widetilde{DN},DN) = e_p(\widetilde{DN},WR) = 1 \quad \textit{and}$$

$$(\text{II.3.5}) \qquad e_p(\widetilde{DN},RRN) = \lim_{n\to\infty} \|d_k(\eta_n)\|^2/\Delta_k^{RRN}(\eta_n) \le 1.$$

PROOF. The noncentrality parameters of WR_n, DN_n and $\widetilde{DN}_n$ are equal to $\|d_k(\eta_n)\|^2 + o(k^{\frac{1}{2}})$. Application of Corollary II.2.4 yields (II.3.4) and (II.3.5). □

Hence, under mild conditions the Rao-Robson-Nikulin test turns out to be the best one of the classical generalizations of $\overline{P}_n$ if k tends slowly to infinity (cf. Section III.4 for numerical illustrations).

The tests (II.1.1) can also be used for the simple testing problem G_0 versus G_{η_n}, inserting the estimator $\hat{\theta}_n$ for the known value of the location-scale parameter. In this manner X_n^2 becomes a competitor for the classical Pearson chi-square statistic. The limiting distributions of X_n^2 for simple hypotheses are identical to the limiting distributions under composite hypotheses because the testing problem is invariant with respect to the location-scale parameter. Thus, assuming the conditions of Corollary II.3.1,

$$(\text{II.3.6}) \qquad e_p(\overline{P},RRN) \le 1 = e_p(\overline{P},\widetilde{DN}) = e_p(\overline{P},DN) = e_p(\overline{P},WR).$$

Finally we investigate the behaviour of the ratio $\Delta_k(\eta_n)/(k^{\frac{1}{2}}n\eta_n^2)$ for the test statistics (I.2.4), (I.2.7), (II.3.1) and (II.3.2); it plays an essential role in our discussions. For fixed k the parameter $\Delta_k(\eta_n)$ heavily depends upon the choice of the estimator $\hat{\theta}_n$ and the matrix $\Lambda(k)$. When $k \to \infty$ Proposition II.3.2 shows under very severe conditions that the behaviour of $\Delta_k(\eta_n)/(n\eta_n^2 k^{\frac{1}{2}})$ is only determined by the information function (cf. Vajda (1973))

$$\text{(II.3.7)} \qquad I(g_1,r) = E_0\{|g_1(Y)/g_0(Y) - 1|^r\} \qquad (1 \le r < \infty)$$

for values near $r = \frac{4}{3}$ and not on $\hat{\theta}_n$. In this very regular case the criterion to keep k bounded or to let $k \to \infty$ only depends on the ratio of the densities g_1 and g_0 (cf. Remark II.2.3).

PROPOSITION II.3.2.

i) *Let* X_n^2 *be a statistic of type* (II.1.1) *with noncentrality parameter* $\Delta_k(\eta_n)$. *Assume*

- $I(g_1,\frac{4}{3}+\rho) = \infty$ *for some* $\rho < 0$
- $\forall\ \varepsilon > 0 \quad q(x) = g_1(G_0^{-1}(x))/g_0(G_0^{-1}(x))$ *is bounded on* $[\varepsilon,1-\varepsilon]$
- *if* $q(x)$ *is not bounded in a neighborhood of* 0 (1), *then* $q(x^{-1})$ ($q(1-x^{-1})$) *varies regularly at* ∞
- $\|D_k' d_k(\eta_n)\| = o(\|d_k(\eta_n)\|)$
- $\limsup_{k\to\infty} \max_{1\le i\le k} kp_{ki}(0) < \infty$

Then

$$\text{(II.3.8)} \qquad \lim_{n\to\infty} \Delta_k(\eta_n)/(k^{\frac{1}{2}} n\eta_n^2) = \infty.$$

ii) *Let* X_n^2 *be one of the statistics* (I.2.4), (I.2.7), (II.3.1) *and* (II.3.2) *with noncentrality parameter* $\Delta_k(\eta_n)$. *Assume*

- $I(g_1,\frac{4}{3}+\rho) < \infty$ *for some* $\rho > 0$
- *for sufficiently small* ε *the components of* $h(G_0^{-1}(x))$ *are Lipschitz continuous of order* $\frac{1}{4}$ *on* $[\varepsilon,1-\varepsilon]$ *and monotone on the intervals* $(0,\varepsilon)$ *and* $(1-\varepsilon,1)$
- $h(G_0^{-1}(x)) = O(x^{-\delta})$ *and* $h(G_0^{-1}(1-x)) = O(x^{-\delta})$ *as* $x \downarrow 0$ *for some* $0 \le \delta < \frac{1}{4}$
- $\lambda_k = o(k^{\frac{1}{2}+\frac{2+6\rho}{4+3\rho}-2\delta})$ *as* $k \to \infty$
- $\liminf_{k\to\infty} \min_{1\le i\le k} kp_{ki}(0) > 0$

Then

$$\text{(II.3.9)} \qquad \lim_{n\to\infty} \Delta_k(\eta_n)/(k^{\frac{1}{2}} n\eta_n^2) = 0.$$

PROOF. cf. Section II.4. □

In Example II.3.4 we consider the testing problem of an exponential null-hypothesis against the contamination of two exponential densities. It is shown that the Pitman efficiency of RRN_n with respect to DN_n and WR_n is strictly greater than one.

EXAMPLE II.3.4. Consider the testing problem of an exponential null-hypothesis with unknown scale

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

against the contamination of two exponential densities

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

Put $p_{ki}(0) = k^{-1}$ $(i = 1,\dots,k)$ and $\eta_n = n^{-\frac{1}{2}}\gamma + o(n^{-\frac{1}{2}})$, let $\hat{\sigma}_n = n^{-1}\Sigma_{j=1}^n Y_j$ and let the sequence $k = k(n) \to \infty$ such that $k^{\frac{5}{2}} \log^{\frac{3}{2}} k = o(n^{\frac{1}{2}})$. Then $e_p(RRN,WR) = e_p(RRN,DN) > 1$.

PROOF. Take $h(x) = h^{ML}(x) = x-1$ and note that $0 \le a_{k1} \le a_{kk-1} = \log k$, $(B_k'B_k)^{-\frac{1}{2}} |B_k'd_k(\eta_n)| \le 2\gamma \log k$ and that $\lambda_k/k = k^{-1}/(1 - B_k'B_k)$ is bounded away from zero and infinity. Now the conditions C.II.1 - C.II.4 are easily verified. By tedious algebra one can show $\Delta_k^{DN}(\eta_n) = \Delta_k^{WR}(\eta_n) + O(\log^2 k) = O(k^{\frac{1}{2}})$ and $\liminf_{n\to\infty} (\Delta_k^{RRN}(\eta_n) - \Delta_k^{DN}(\eta_n))/k^{\frac{1}{2}} > 0$. Application of Corollary II.3.1 yields the desired result. □

Proposition II.3.2 is not applicable to Example II.3.4 because the parameters are such that $I(g_1, \frac{4}{3} + \rho) = \infty$ for all $\rho > 0$. If, however, we choose an alternative G_1 such that $I(g_1, \frac{4}{3} + \rho) < \infty$ for some $\rho > 0$, the conditions of Proposition II.3.2 are satisfied because $\lambda_k = O(k)$ and $h_2^{ML}(F_0^{-1}(1-x)) = -\log x - 1 = o(x^{-\delta})$ for all $\delta > 0$.
Note that the contamination of two different distributions from the same location-scale family does not lie in that family. This is an unpleasant feature of the local family (II.1.2): Although H_0 and H_1 coincide Example II.3.4 shows that the power at local alternatives H_{1n} can be appreciable.

The final example shows that, for fixed k, one cannot order the classical extensions of $\overline{P}_n$. The order which holds for $k \to \infty$ may even be reversed.

<u>EXAMPLE II.3.5</u>. Consider the testing problem of a normal null-hypothesis with unknown location versus a regular symmetric alternative. Put $p_{ki}(0) = k^{-1}$ $(i = 1,\ldots,k)$, $\eta_n = n^{-\frac{1}{2}}\gamma + o(n^{-\frac{1}{2}})$, and let $\hat{\mu}_n = n^{-1}\,\Sigma_{j=1}^n\, Y_j$. Then, for fixed k,

$$e_p(DN,WR) > 1$$

$$e_p(WR,RRN) > 1$$

while, for $k \to \infty$ and $k^{\frac{5}{2}}\log^{\frac{5}{2}} k = o(n^{\frac{1}{2}})$,

$$e_p(DN,WR) = e_p(WR,RRN) = 1.$$

<u>PROOF</u>. cf. Example 5.1 of Drost (1987). □

II.4. PROOFS

II.4.1. Proof of Theorem II.2.2.

The proof is based on three lemmas. The first one rewrites $V_k(\hat{\theta}_n)$ as the sum of the classical Pearson term and two remainder terms. The last two lemmas investigate the influence of the error terms. Proofs are only given under H_{1n} and $\theta = \theta_0 = (0,1)'$ (the proof under H_0 is obtained by substituting $\eta_n = 0$). Note that (II.1.7) implies that $\hat{\mu}_n \to_P 0$ and $\hat{\sigma}_n \to_P 1$ if θ_0 is true. Throughout this section we assume without further references the conditions C.II.1 - C.II.4.
Introduce the notations $(a_{ij})_{ij}$ for the matrix with (i,j)-th entry a_{ij}, $(b_i)_i$ for the vector with elements b_i and let δ_{ij} be the Kronecker symbol. The relevant values for the indices i and j are derived from the text. Let $G_\eta(U) = \int_U dG_\eta(x)$ and let $g(x)\Big|_a^b = g(b) - g(a)$; $g(x)\Big|_a^b - \Big|_c^d$ is defined in a similar manner.

<u>LEMMA II.4.1</u>. (Moore and Spruill (1975))

$$\text{(II.4.1)} \qquad V_k(\hat{\theta}_n) = V_k(\theta_0) - B_k n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) + R_k,$$

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

$$R_{1ki} = p_{ki}^{-\frac{1}{2}}(0)n^{\frac{1}{2}}\{F_n(x) - G_{\eta_n}(x)\}\Big|_{\hat{\mu}_n + a_{ki-1}\hat{\sigma}_n}^{\hat{\mu}_n + a_{ki}\hat{\sigma}_n} - \Big|_{a_{ki-1}}^{a_{ki}} \quad \textit{and}$$

$$R_{2ki} = B_{ki}n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) + p_{ki}^{-\frac{1}{2}}(0)n^{\frac{1}{2}}(p_{ki}^*(\eta_n,\hat{\theta}_n,\theta_0) - p_{ki}(\eta_n))$$

$$(i = 1,\ldots,k).$$

PROOF. Direct calculation. □

REMARK II.4.1. Relation (II.1.6) implies $V_k(\hat{\theta}_n) = V_k(\theta_0) - B_k n^{-\frac{1}{2}} \Sigma_{j=1}^n h(Y_j) + - B_k Q_n(Y_1,\ldots,Y_n) + R_k$ and the covariance matrix Σ_k (cf. (II.1.11)) is obtained by evaluating the expectation

$$E_0[\{V_k(\theta_0) - B_k n^{-\frac{1}{2}} \sum_{j=1}^{n} h(Y_j)\}\{V_k(\theta_0) - B_k n^{-\frac{1}{2}} \sum_{j=1}^{n} h(Y_j)\}']. \quad \square$$

LEMMA II.4.2.

$$\|R_k\|^2 = o_p(1) \textit{ under } H_0 \textit{ and } H_{1n} \textit{ and}$$

(II.4.2)

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

PROOF. Under H_{1n}. The conclusions are implied by similar statements about R_{1k} and R_{2k}.

A. $\|R_{1k}\|^2 = o_p(1)$ and $\|\Lambda(k)^{\frac{1}{2}}D_k'R_{1k}\|^2 = o_p(k^{\frac{1}{2}})$.

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

Let B(n,p) be a random variable having a binomial distribution with parameters n and p. Let c be some sufficiently large constant. We use the following inequalities (cf. Bahadur (1966), Hoeffding (1963)).

(II.4.3) $P(|\text{Bin }(n,p) - np| \geq nt) \leq c \cdot \exp(-2nt^2)$ uniformly in $p \in (0,1)$ and $t > 0$,

(II.4.4) $P(|\text{Bin }(n,p) - np| \geq t) \leq c \cdot \exp(-\frac{1}{2}t^2/(np+t))$ uniformly in $p \in (0,1)$ and $t > 0$,

(II.4.5) $P(\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \geq t) \leq c \cdot \exp(-2nt^2)$ uniformly in $t > 0$ and distribution F.

Let $\varepsilon > 0$. Because of (II.1.7) there exists N_ε such that for all n

$$P_{1n}(n^{\frac{1}{2}}\|\hat{\theta}_n - \theta_0\| > N_\varepsilon) \le \varepsilon/4.$$

Let n_0 be a sufficiently large integer. Define intervals

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

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

$$G_{\eta_n}(J_{ni}) = 2n^{-\frac{1}{2}} \log^{\frac{1}{2}} k \quad (i = 1,\dots,k-1). \tag{II.4.6}$$

Note that C.II.3-d implies that the RHS of (II.4.6) tends to zero. The construction is possible because of (II.1.5) and C.II.4-d. Note

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

Let $\delta > 0$. Because of C.II.3-d, for $n \ge n_0$,

$$256\, n^{-\frac{1}{2}} \log^{\frac{3}{2}} k \sum_{i=1}^{k} p_{ki}^{-1}(0) \le \delta \text{ and}$$
$$256\, n^{-\frac{1}{2}}k^{-\frac{1}{2}}\lambda_k \log^{\frac{3}{2}} k \sum_{i=1}^{k} p_{ki}^{-1}(0) \le \delta.$$

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

$$\begin{aligned}
&P_{1n}\Bigg(\sum_{i=1}^{k} p_{ki}^{-1}(0)n\Bigg(\{F_n(x) - G_{\eta_n}(x)\}\Bigg|_{a_{ki}}^{\hat{\mu}_n + a_{ki}\hat{\sigma}_n} - \Bigg|_{a_{ki-1}}^{\hat{\mu}_n + a_{ki-1}\hat{\sigma}_n}\Bigg)^2 \ge\\
&\qquad\qquad \ge 256\, n^{-\frac{1}{2}} \log^{\frac{3}{2}} k \sum_{i=1}^{k} p_{ki}^{-1}(0)\Bigg)\\
&\le P_{1n}(\exists\, i \in \{1,\dots,k-1\}\ \ \hat{\mu}_n + a_{ki}\hat{\sigma}_n \notin J_{ni}) +\\
&+ P_{1n}\Bigg(\forall\, i \in \{1,\dots,k-1\}\ \ \hat{\mu}_n + a_{ki}\hat{\sigma}_n \in J_{ni};\\
&\qquad \exists\, i \in \{1,\dots,k-1\}\ \ \Bigg|\{F_n(x) - G_{\eta_n}(x)\}\Bigg|_{a_{ki}}^{\hat{\mu}_n + a_{ki}\hat{\sigma}_n}\Bigg| \ge 8n^{-\frac{3}{4}} \log^{\frac{3}{4}} k\Bigg)
\end{aligned}$$

$$\le \frac{\varepsilon}{2} + \sum_{i=1}^{k-1} P_{1n}(\sup_{U \in J_{ni}} |F_n(U) - G_{\eta_n}(U)| \ge 8n^{-\frac{3}{4}} \log^{\frac{3}{4}} k), \tag{II.4.7}$$

where $\mathcal{J}_{ni} = \{U \subset J_{ni};\ U$ is an interval$\}$. To prove that the second term of (II.4.7) is bounded by $\varepsilon/2$ define the conditional probability

$$\pi_i(j) = P_{1n}(\sup_{U\in\mathcal{J}_{ni}} |F_n(U) - G_{\eta_n}(U)| \geq 8n^{-\frac{3}{4}} \log^{\frac{3}{4}} k \,|\, F_n(J_{ni}) = j/n)$$
$$(i=1,\dots,k-1;\ j=0,\dots,n).$$

Then

$$\text{(II.4.8)}\quad \begin{aligned} &P_{1n}(\sup_{U\in\mathcal{J}_{ni}} |F_n(U) - G_{\eta_n}(U)| \geq 8n^{-\frac{3}{4}} \log^{\frac{3}{4}} k) \\ &= \Big(\sum_{j\leq n^{\frac{1}{2}} \log^{\frac{1}{2}} k} + \sum_{j> n^{\frac{1}{2}} \log^{\frac{1}{2}} k}\Big) \pi_i(j) P_{1n}(F_n(J_{ni}) = j/n). \end{aligned}$$

For $n \geq n_0$ it is easily seen that the first sum of the RHS is bounded by $\varepsilon/(4k)$ using (II.4.3), (II.4.6), $\pi_i(j) \leq 1$ and

$$\begin{aligned} &P_{1n}(F_n(J_{ni}) \leq n^{-\frac{1}{2}} \log^{\frac{1}{2}} k) \\ &\leq P_{1n}(n\,|F_n(J_{ni}) - G_{\eta_n}(J_{ni})| \geq n^{\frac{1}{2}} \log^{\frac{1}{2}} k) \\ &\leq c\cdot \exp(-2 \log k) \leq \varepsilon/(4k). \end{aligned}$$

Next we show that the second sum of (II.4.8) is also bounded by $\varepsilon/(4k)$. Note that for $j \neq 0$ conditionally given $F_n(J_{ni}) = j/n$

$$\begin{aligned} \sup_{U\in\mathcal{J}_{ni}} |F_n(U) - G_{\eta_n}(U)| &\leq G_{\eta_n}(J_{ni}) \left\{ \sup_{U\in\mathcal{J}_{ni}} \left| \frac{F_n(U)}{G_{\eta_n}(J_{ni})} - \frac{F_n(U)}{F_n(J_{ni})} \right| + \right. \\ &\left. \quad + \sup_{U\in\mathcal{J}_{ni}} \left| \frac{F_n(U)}{F_n(J_{ni})} - \frac{G_{\eta_n}(U)}{G_{\eta_n}(J_{ni})} \right| \right\} \\ &\leq |F_n(J_{ni}) - G_{\eta_n}(J_{ni})| + G_{\eta_n}(J_{ni}) \sup_{U\in\mathcal{J}_{ni}} |\tilde{F}_j(U) - \tilde{G}_{\eta_n}(U)|, \end{aligned}$$

where $\tilde{G}_{\eta_n}$ 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. Define

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

$$\pi_{2i}(j) = P_{1n}(\sup_{U\in\mathcal{J}_{ni}} |\tilde{F}_j(U) - \tilde{G}_{\eta_n}(U)| \geq 2n^{-\frac{1}{4}} \log^{\frac{1}{4}} k)$$
$$(i=1,\dots,k-1;\ j=1,\dots,n).$$

Using (II.4.4), (II.4.5) and $\pi_i(j) \le \pi_{1i}(j) + \pi_{2i}(j)$, for $n \ge n_0$ the second sum of (II.4.8) is bounded by

$$\begin{aligned}
&\sum_{j > n^{\frac{1}{2}} \log^{\frac{1}{2}} k} (\pi_{1i}(j) + \pi_{2i}(j)) P_{1n}(F_n(J_{ni}) = j/n) \\
&\le P_{1n}(n\,|F_n(J_{ni}) - G_{\eta_n}(J_{ni})| \ge 4n^{\frac{1}{4}} \log^{\frac{3}{4}} k) + \\
&+ \sum_{j > n^{\frac{1}{2}} \log^{\frac{1}{2}} k} P_{1n}(\sup_{x \in \mathbb{R}} |\tilde{F}_j(x) - \tilde{G}_{\eta_n}(x)| \ge n^{-\frac{1}{4}} \log^{\frac{1}{4}} k) P_{1n}(F_n(J_{ni}) = j/n) \\
&\le c \cdot \exp(-4 \log k / (1 + 2n^{-\frac{1}{4}} \log^{\frac{1}{4}} k)) + \\
&+ \sum_{j > n^{\frac{1}{2}} \log^{\frac{1}{2}} k} c \cdot \exp(-2jn^{-\frac{1}{2}} \log^{\frac{1}{2}} k) P_{1n}(F_n(J_{ni}) = j/n) \\
&\le 2c \cdot \exp(-2 \log k) \le \varepsilon/(4k).
\end{aligned}$$

Thus from (II.4.7) and (II.4.8) we obtain $P_{1n}(\|R_{1k}\|^2 \ge \delta) \le \varepsilon$ and $P_{1n}(\|\Lambda(k)^{\frac{1}{2}} D_k' R_{1k}\|^2 \ge \delta k^{\frac{1}{2}}) \le \varepsilon$.

B. $\quad \|R_{2k}\|^2 = o_p(1)$ and $\|\Lambda(k)^{\frac{1}{2}} D_k' R_{2k}\|^2 = o_p(k^{\frac{1}{2}})$

Observe that for $i = 1, \dots, k$

$$\begin{aligned}
&|n^{-\frac{1}{2}} p_{ki}^{\frac{1}{2}}(0) R_{2ki}| = |p_{ki}^{\frac{1}{2}}(0) B_{ki}(\hat{\theta}_n - \theta_0) - p_{ki}(\eta_n) + p_{ki}^*(\eta_n, \hat{\theta}_n, \theta_0)| \\
&= \left| p_{ki}^{\frac{1}{2}}(0) B_{ki}(\hat{\theta}_n - \theta_0) + G_{\eta_n}(x) \Big|_{a_{ki-1} + \hat{\mu}_n + (\hat{\sigma}_n - 1) a_{ki-1}}^{a_{ki} + \hat{\mu}_n + (\hat{\sigma}_n - 1) a_{ki}} - \Big|_{a_{ki-1}}^{a_{ki}} \right| \\
&= \Big| -g_0(x)\{\hat{\mu}_n + (\hat{\sigma}_n - 1)x\} \Big|_{a_{ki-1}}^{a_{ki}} + \\
&+ g_{\eta_n}(x^*(x))\{\hat{\mu}_n + (\hat{\sigma}_n - 1)x\} \Big|_{a_{ki-1}}^{a_{ki}} \Big|
\end{aligned}$$

where $x^*(x)$ is a random point between x and $x + \hat{\mu}_n + (\hat{\sigma}_n - 1)x$

$$\begin{aligned}
&\le \max_{\{a_{k1}, \dots, a_{kk-1}\}} 2\{|\hat{\mu}_n| + |\hat{\sigma}_n - 1|\,|x|\} \cdot \\
&\qquad \cdot \{L_0 |x^*(x) - x| + \eta_n\, g_1(x^*(x)) - g_0(x^*(x))|\} \\
&\le 2L_0 \max_{\{a_{k1}, a_{kk-1}\}} \{|\hat{\mu}_n| + |\hat{\sigma}_n - 1|\,|x|\}^2 + \\
&+ 2\eta_n (M_0 + M_1) \max_{\{a_{k1}, a_{kk-1}\}} \{|\hat{\mu}_n| + |\hat{\sigma}_n - 1|\,|x|\} \\
&\le c \cdot \max(a_{k1}^2, a_{kk-1}^2) \|\hat{\theta}_n - \theta_0\| (\|\hat{\theta}_n - \theta_0\| + \eta_n).
\end{aligned}$$

Thus $||R_{2k}||^2 \le 2c^2 n||\hat{\theta}_n - \theta_0||^2(||\hat{\theta}_n - \theta_0||^2 + \eta_n^2) \max\{a_{k1}^4, a_{kk-1}^4\} \sum_{i=1}^{k} p_{ki}^{-1}(0)$. Using $||\Lambda(k)^{\frac{1}{2}} D_k' R_{2k}||^2 \le \lambda_k ||R_{2k}||^2$, B is implied by (II.1.7), C.II.3-d and C.II.4-d. □

LEMMA II.4.3.

i) $$||D_k' V_k(\theta_0)||^2 = O_p(1) \textit{ under } H_0 \textit{ and}$$

(II.4.9)

$$||D_k'(V_k(\theta_0) - d_k(\eta_n))||^2 = O_p(1) \textit{ under } H_{1n}.$$

ii) $$||\Lambda(k)^{\frac{1}{2}} D_k'(V_k(\theta_0) - B_k n^{-\frac{1}{2}} \sum_{j=1}^{n} h(Y_j))||^2 = o_p(k^{\frac{1}{2}}) \textit{ under } H_0 \textit{ and}$$

(II.4.10)

$$||\Lambda(k)^{\frac{1}{2}} D_k'(V_k(\theta_0) - d_k(\eta_n) - B_k n^{-\frac{1}{2}} \sum_{j=1}^{n} \{h(Y_j) - E_{\eta_n} h(Y_j)\}||^2 = o_p(s_k(\eta_n))$$

under H_{1n}.

PROOF. Under H_{1n}. First we evaluate the expectations

$$E_{\eta_n}[(V_k(\theta_0) - d_k(\eta_n))(V_k(\theta_0) - d_k(\eta_n))']$$
$$= I_k + n^{-\frac{1}{2}}(\delta_{ij} p_{ki}^{-\frac{1}{2}}(0) d_{ki}(\eta_n))_{ij} - (q_k + n^{-\frac{1}{2}} d_k(\eta_n))(q_k + n^{-\frac{1}{2}} d_k(\eta_n))',$$
$$E_{\eta_n}[B_k n^{-1} \sum_{j=1}^{n} \{h(Y_j) - E_{\eta_n} h(Y_j)\} \sum_{j=1}^{n} \{h(Y_j) - E_{\eta_n} h(Y_j)\}' B_k']$$
$$= (1 - \eta_n) B_k A^{-1} B_k' + \eta_n B_k E_1[h(Y)h(Y)'] B_k' +$$
$$- \eta_n^2 B_k E_1\{h(Y)\} E_1\{h(Y)'\} B_k' \quad \text{and}$$
$$E_{\eta_n}[B_k n^{-\frac{1}{2}} \sum_{j=1}^{n} \{h(Y_j) - E_{\eta_n} h(Y_j)\}(V_k(\theta_0) - d_k(\eta_n))']$$
$$= B_k A^{-1} C_k' + \eta_n B_k (p_{ki}^{-\frac{1}{2}}(0) \int_{I_{ki}} h(y)' dG_1(y) - G_0(y))_i' +$$
$$- \eta_n B_k E_1\{h(Y)\}(q_k + n^{-\frac{1}{2}} d_k(\eta_n))'.$$

Proof of part i): note that (using C.II.3-d)

$$E_{\eta_n}\{||D_k'(V_k(\theta_0) - d_k(\eta_n))||^2\}$$
$$\le \text{tr.} D_k'[I_k + n^{-\frac{1}{2}}(\delta_{ij} p_{ki}^{-\frac{1}{2}}(0) d_{ki}(\eta_n))_{ij}] D_k$$
$$\le (1 + \max_{1 \le i \le k} n^{-\frac{1}{2}} p_{ki}^{-\frac{1}{2}}(0) |d_{ki}(\eta_n)|) \text{tr.} D_k' D_k$$
$$= O(1).$$

The proof is completed using the implication

$$(\text{II.4.11}) \qquad E\{\|X_k\|^2\} = O(1) \text{ uniform in } k \Rightarrow \|X_k\|^2 = O_p(1) \text{ uniform in } k.$$

Proof of part ii): let $\lambda(U)$ denote the maximum eigenvalue of the matrix U (as before λ_k is the maximum eigenvalue of $\Lambda(k)$), then

$$E_{\eta_n}\{\|\Lambda(k)^{\frac{1}{2}}D_k'(V_k(\theta_0) - d_k(\eta_n) - B_k n^{-\frac{1}{2}} \sum_{j=1}^{n} \{h(Y_j) - E_{\eta_n}h(Y_j)\})\|^2\}$$

$$\leq \text{tr.}\,\Lambda(k)D_k'[I_k + n^{-\frac{1}{2}}(\delta_{ij}p_{ki}^{-\frac{1}{2}}(0)d_{ki}(\eta_n))_{ij} + B_kA^{-1}B_k' +$$

$$+ \eta_n B_k E_1[h(Y)h(Y)']B_k' - B_kA^{-1}C_k' - C_kA^{-1}B_k' +$$

$$- \eta_n B_k(p_{ki}^{-\frac{1}{2}}(0) \int_{I_{ki}} h(y)'\,dG_1(y) - G_0(y))_i' +$$

$$- \eta_n (p_{ki}^{-\frac{1}{2}}(0) \int_{I_{ki}} h(y)'\,dG_1(y) - G_0(y))_i B_k' +$$

$$+ \eta_n B_k E_1\{h(Y)\}(q_k + n^{-\frac{1}{2}}d_k(\eta_n))' +$$

$$+ \eta_n (q_k + n^{-\frac{1}{2}}d_k(\eta_n))E_1\{h(Y)'\}B_k']D_k$$

$$\leq \text{tr.}\,\Lambda(k)^{\frac{1}{2}}\Psi(k)\Lambda(k)^{\frac{1}{2}} + \eta_n\lambda_k\,\text{tr.}\,D_k'D_k \max_{1\leq i\leq k} p_{ki}^{-1}(0)\;|p_{ki}(1) - p_{ki}(0)| +$$

$$+ \eta_n\lambda_k\lambda(B_k'B_k)E_1\{\|h(Y)\|^2\} +$$

$$+ 2\eta_n\lambda_k\Big\{\text{tr.}\,B_k'B_k \sum_{i=1}^{k} p_{ki}^{-1}(0) \int_{I_{ki}} h(y)'dG_1(y) - G_0(y)$$

$$\int_{I_{ki}} h(y)dG_1(y) - G_0(y)\Big\}^{\frac{1}{2}} +$$

$$+ 2\eta_n^2\lambda_k\{\lambda(B_k'B_k)\|E_1\{h(Y)\}\|^2 \sum_{i=1}^{k} p_{ki}^{-1}(0)(p_{ki}(1) - p_{ki}(0))^2\}^{\frac{1}{2}}$$

$$= o(s_k(\eta_n)) \quad \text{using C.II.3-d and C.II.4-b.}$$

The proof is completed using (II.4.11). □

<u>PROOF of Theorem II.2.2.</u> Under H_{1n}. The column space of $[B_k, C_k]$ is the kernel of the projection matrix $I_k - D_kD_k'$, thus (using Lemma II.4.1)

$$X_n^2 = V_k(\hat\theta_n)'[I_k - D_kD_k' + D_k\Lambda(k)D_k']V_k(\hat\theta_n)$$

$$(\text{II.4.12}) \qquad = (V_k(\theta_0) + R_k)'\,[I - D_kD_k'](V_k(\theta_0) + R_k) +$$

$$+ (V_k(\theta_0) - d_k(\eta_n) - B_k n^{-\frac{1}{2}} \sum_{j=1}^{n} \{h(Y_j) - E_{\eta_n}h(Y_j)\} +$$

$$+ d_k(\eta_n) - B_k n^{\frac{1}{2}} E_{\eta_n}\{h(Y)\} + R_k - B_k Q_n(Y_1,\ldots,Y_n))' D_k \Lambda(k) D_k' \cdot$$

$$\cdot (V_k(\theta_0) - d_k(\eta_n) - B_k n^{-\frac{1}{2}} \sum_{j=1}^{n} \{h(Y_j) - E_{\eta_n} h(Y_j)\} +$$

$$+ d_k(\eta_n) - B_k n^{\frac{1}{2}} E_{\eta_n}\{h(Y)\} + R_k - B_k Q_n(Y_1,\ldots,Y_n)).$$

Proposition II.2.1 implies $\overline{P}_n = O_p(k + \|d_k(\eta_n)\|^2) = O_p(s_k^2(\eta_n))$, thus the first term of the RHS can be rewritten as

$$\|V_k(\theta_0)\|^2 - \|D_k'(V_k(\theta_0) - d_k(\eta_n))\|^2 - \|D_k' d_k(\eta_n)\|^2 +$$

$$- 2(V_k(\theta_0) - d_k(\eta_n))' D_k D_k' d_k(\eta_n) + \|[I_k - D_k D_k']^{\frac{1}{2}} R_k\|^2 +$$

$$+ 2V_k(\theta_0)'[I_k - D_k D_k'] R_k$$

$$= AN(k + \|[I_k - D_k D_k']^{\frac{1}{2}} d_k(\eta_n)\|^2, 2k + 4\|d_k(\eta_n)\|^2) + o_p(s_k(\eta_n))$$

(use the previous lemmas, C.II.4-c and the Cauchy-Schwarz inequality for the cross-terms). Similarly we treat the second term of (II.4.12)

$$\|\Lambda(k)^{\frac{1}{2}} D_k'(V_k(\theta_0) - d_k(\eta_n) - B_k n^{-\frac{1}{2}} \sum_{j=1}^{n} \{h(Y_j) - E_{\eta_n} h(Y_j)\})\|^2 +$$

$$+ \|\Lambda(k)^{\frac{1}{2}} D_k'(R_k - B_k Q_n(Y_1,\ldots,Y_n))\|^2 + \Delta_k^*(\eta_n) + \text{cross-terms}$$

$$= \Delta_k^*(\eta_n) + o_p(s_k(\eta_n)) + o_p((\Delta_k^*(\eta_n) s_k(\eta_n))^{\frac{1}{2}})$$

(also use C.II.4-a). Formulas (II.2.2a) - (II.2.2c) are directly implied by

$$X_n^2 = AN(m_k(\eta_n), s_k^2(\eta_n) - \Delta_k^*(\eta_n)) + o_p(s_k(\eta_n)) + o_p((\Delta_k^*(\eta_n) s_k(\eta_n))^{\frac{1}{2}})$$

$$X_n^2 = k + \Delta_k(\eta_n) + O_p(s_k(\eta_n)) + o_p((\Delta_k^*(\eta_n) s_k(\eta_n))^{\frac{1}{2}}). \quad \square$$

<u>REMARK II.4.2.</u> Omit the conditions C.II.2-c,d,e under local alternatives but assume that (II.1.7) holds. Then, if $\lambda_k = o(s_k(\eta_n))$, some minor modifications in the final part of the proof show that Theorem II.2.2 is still applicable (replace $E_1\{h(Y)\}$ by zero). $\square$

II.4.2. Proof of Proposition II.3.2.

If $\Delta_k^*(\eta_n)/(k^{\frac{1}{2}} n \eta_n^2) \to 0$ under the conditions of part ii) the proposition follows directly from Propositions 4.2 and 4.4 of Kallenberg et al. (1985) because $\|[I_k - D_k D_k']^{\frac{1}{2}} d_k(\eta_n)\|^2 \le \Delta_k(\eta_n) \le \|d_k(\eta_n)\|^2 + \Delta_k^*(\eta_n)$. The proof of $\Delta_k^*(\eta_n)/(k^{\frac{1}{2}} n \eta_n^2) \to 0$ is only required for the Rao-Robson-Nikulin statistic. First we give some algebraic relations

$$B'B - B'BAB'B \text{ nonnegative definite} \Rightarrow I - BAB' \text{ nonnegative definite and}$$

$$A,B \text{ nonnegative definite and symmetric} \Rightarrow A - A(A+B)^{+}A \text{ nonnegative definite,}$$

which follow from the spectral decomposition of symmetric matrices. These are used for the matrices $(B_k - C_k)A^{-1}(B_k - C_k)'$ and $I_k - C_kA^{-1}C_k'$. To prove that the last matrix is nonnegative definite we show $A - C_k'C_k \geq 0$. Define for each $v \in \mathbb{R}^2$ $g_v(y) = v'Ah(y)$, then

$$\begin{aligned}
& v'(A - C_k'C_k)v \\
&= v'A\Big(\int h(y)h(y)'\,dG_0(y) + \\
&\quad - \sum_{i=1}^{k} p_{ki}^{-1}(0) \int_{I_{ki}} h(y)dG_0(y) \int_{I_{ki}} h(y)'\,dG_0(y)\Big)Av \\
&= \sum_{i=1}^{k} \Big\{\int_{I_{ki}} g_v^2(y)dG_0(y) - p_{ki}^{-1}(0)\Big\{\int_{I_{ki}} 1\cdot g_v(y)dG_0(y)\Big\}^2\Big\} \geq 0.
\end{aligned}$$

Thus, with $\Lambda(k) = \Psi(k)^{+} = (D_k'\Sigma_k D_k)^{+}$,

$$\begin{aligned}
\Delta_k^*(\eta_n) &= \|\Lambda(k)^{\frac{1}{2}}D_k'(d_k(\eta_n) - B_kn^{\frac{1}{2}}E_{\eta_n}\{h(Y)\})\|^2 \\
&= \|\Lambda(k)^{\frac{1}{2}}[D_k'D_k - D_k'C_kA^{-1}C_k'D_k + D_k'C_kA^{-1}C_k'D_k]\,\cdot \\
&\qquad \cdot D_k'(d_k(\eta_n) - B_kn^{\frac{1}{2}}E_{\eta_n}\{h(Y)\})\|^2 \\
&\leq 2\|\Lambda(k)^{\frac{1}{2}}[D_k'D_k - D_k'C_kA^{-1}C_k'D_k]D_k'(d_k(\eta_n) - B_kn^{\frac{1}{2}}E_{\eta_n}\{h(Y)\})\|^2 + \\
&\quad + 2\|\Lambda(k)^{\frac{1}{2}}D_k'C_kA^{-1}C_k'D_kD_k'(d_k(\eta_n) - B_kn^{\frac{1}{2}}E_{\eta_n}\{h(y)\})\|^2 \\
&\leq 2\|[D_k'D_k - D_k'C_kA^{-1}C_k'D_k]^{\frac{1}{2}}D_k'(d_k(\eta_n) - B_kn^{\frac{1}{2}}E_{\eta_n}\{h(Y)\})\|^2 \\
&\quad + 2\|\Lambda(k)^{\frac{1}{2}}D_k'C_kA^{-1}\Big(C_k'd_k(\eta_n) - An^{\frac{1}{2}}E_{\eta_n}\{h(Y)\} + \\
&\qquad + (A - C_k'C_k)n^{\frac{1}{2}}E_{\eta_n}\{h(Y)\} - C_k'(B_k - C_k)n^{\frac{1}{2}}E_{\eta_n}\{h(Y)\}\Big)\|^2 \\
&\leq 2\|d_k(\eta_n) - B_kn^{\frac{1}{2}}E_{\eta_n}\{h(Y)\}\|^2 + \\
&\quad + 6\|\Lambda(k)^{\frac{1}{2}}D_k'C_kA^{-1}(C_k'A^{-1}(C_k'd_k(\eta_n) - An^{\frac{1}{2}}E_{\eta_n}\{h(Y)\})\|^2 +
\end{aligned}$$

$$
\begin{aligned}
&+ 6\|\Lambda(k)^{\frac12} D_k' C_k A^{-1}(A - C_k'C_k) n^{\frac12} E_{\eta_n}\{h(Y)\}\|^2 + \\
&+ 6\|\Lambda(k)^{\frac12}[D_k'D_k - D_k'C_kA^{-1}C_k'D_k - D_k'D_k]D_k'(B_k - C_k)n^{\frac12}E_{\eta_n}\{h(Y)\}\|^2 \\
&\le 4\|d_k(\eta_n)\|^2 + 4\|B_k n^{\frac12} E_{\eta_n}\{h(Y)\}\|^2 + \\
&+ 6\lambda_k\|D_k'C_kA^{-1}(C_k'd_k(\eta_n) - An^{\frac12}E_{\eta_n}\{h(Y)\})\|^2 + \\
&+ 6\|\Lambda(k)^{\frac12}[D_k'D_k - D_k'C_kA^{-1}C_k'D_k]D_k'C_kn^{\frac12}E_{\eta_n}\{h(Y)\}\|^2 + \\
&+ 12\|\Lambda(k)^{\frac12}D_k'(B_k - C_k)A^{-\frac12}A^{\frac12}n^{\frac12}E_{\eta_n}\{h(Y)\}\|^2 + \\
&+ 12\|\Lambda(k)^{\frac12}[D_k'D_k - D_k'C_kA^{-1}C_k'D_k]D_k'(B_k - C_k)n^{\frac12}E_{\eta_n}\{h(Y)\}\|^2 \\
&\le 4\|d_k(\eta_n)\|^2 + 4\|B_kn^{\frac12}E_{\eta_n}\{h(Y)\}\|^2 + \\
&+ 6\lambda_k\|A^{-\frac12}(C_k'd_k(\eta_n) - An^{\frac12}E_{\eta_n}\{h(Y)\})\|^2 + \\
&+ 6\|[D_k'D_k - D_k'C_kA^{-1}C_k'D_k]^{\frac12}D_k'C_kn^{\frac12}E_{\eta_n}\{h(Y)\}\|^2 + \\
&+ 12\|A^{\frac12}n^{\frac12}E_{\eta_n}\{h(Y)\}\|^2 + \\
&+ 12\|[D_k'D_k - D_k'C_kA^{-1}C_k'D_k]^{\frac12}D_k'(B_k - C_k)n^{\frac12}E_{\eta_n}\{h(Y)\}\|^2 \\
&\le 4\|d_k(\eta_n)\|^2 + n\eta_n^2\|E_1\{h(Y)\}\|^2\Big(4\lambda(B_k'B_k) + 6\lambda(C_k'C_k) + \\
&\qquad\qquad + 12\lambda(A) + 12\lambda((B_k - C_k)'(B_k - C_k))\Big) + \\
&+ 6\lambda_k n\eta_n^2\Big\|A^{\frac12}\sum_{i=1}^{k} p_{ki}^{-1}(0)\int_{I_{ki}} g_1(y)dy\int_{I_{ki}} h(y)g_0(y)dy + \\
&\qquad\qquad - \int_{I_{ki}} h(y)g_1(y)dy\Big\|^2 \\
&= o(k^{\frac12}n\eta_n^2),
\end{aligned}
$$

using Proposition 4.2 of Kallenberg et al. (1985), (II.1.10), C.II.2-c and Lemma II.5.1. □

II.5. APPENDIX

<u>LEMMA II.5.1</u>. *Let* q *and* r *be nonnegative measurable functions on* (0,a), a > 0 *satisfying the following conditions*

$$\int_0^a q^{\frac43+\rho}(x)dx < \infty \text{ for some } \rho > 0$$

r is a monotone function, bounded on (a^*,a) *for each* $0 < a^* < a$ *and* $r(x) = O(x^{-\delta})$ *for some* $0 \le \delta < \frac{1}{4}$ *as* $x \downarrow 0$.

Then, writing $U_{ki} = \left(\frac{i-1}{k}, \frac{i}{k} \wedge a\right)$, *as* $k \to \infty$

$$\text{(II.5.1)} \qquad \int_0^a r(x)q(x)dx - \sum_{i=1}^{[ak]+1} k \int_{U_{ki}} r(x)dx \int_{U_{ki}} q(x)dx = O\left(k^{-\frac{1+3\rho}{4+3\rho}+\delta}\right).$$

PROOF. Put $\omega = \int_0^a r(x)q(x)dx$ and note that $\omega < \infty$. Let

$$\omega_k = \sum_{i=2}^{[ak]+1} k \int_{U_{ki}} r(x)dx \int_{U_{ki}} q(x)dx \qquad (k \ge 1).$$

First consider the case that r is nonincreasing. By Hölder's inequality and the inequality $(a-b)^c \le a^c - b^c$ if $a > b > 0$ and $c \ge 1$

$$\omega_k \ge \sum_{i\ge2} r\left(\frac{i}{k}\right) \int_{U_{ki}} q(x)dx$$

$$\ge \sum_{i\ge2} \int_{U_{ki}} r(x)q(x)dx - \sum_{i\ge2} \left\{r\left(\frac{i-1}{k}\right) - r\left(\frac{i}{k}\right)\right\} \int_{U_{ki}} q(x)dx$$

$$\ge \omega - \int_{U_{k1}} r(x)q(x)dx - \sum_{i\ge2} \left\{r\left(\frac{i-1}{k}\right) - r\left(\frac{i}{k}\right)\right\} \int_{U_{ki}} q(x)dx$$

$$\ge \omega - \left[\int_{U_{k1}} \{r(x)\}^{\frac{4+3\rho}{1+3\rho}} dx\right]^{\frac{1+3\rho}{4+3\rho}} \left[\int_{U_{k1}} \{q(x)\}^{\frac{4+3\rho}{3}} dx\right]^{\frac{3}{4+3\rho}} +$$

$$-\left[\sum_{i\ge2} \left\{r\left(\frac{i-1}{k}\right) - r\left(\frac{i}{k}\right)\right\}^{\frac{4+3\rho}{1+3\rho}}\right]^{\frac{1+3\rho}{4+3\rho}} \left[\sum_{i\ge2} \left(\int_{U_{ki}} q(x)dx\right)^{\frac{4+3\rho}{3}}\right]^{\frac{3}{4+3\rho}}$$

$$\ge \omega - \left[\int_{U_{k1}} x^{-\delta\frac{4+3\rho}{1+3\rho}} dx\right]^{\frac{1+3\rho}{4+3\rho}} \left[\int_0^a q^{\frac{4}{3}+\rho}(x)dx\right]^{\frac{3}{4+3\rho}} O(1) +$$

$$-\left[\sum_{i\ge2} \left\{r\left(\frac{i-1}{k}\right)\right\}^{\frac{4+3\rho}{1+3\rho}} - \left\{r\left(\frac{i}{k}\right)\right\}^{\frac{4+3\rho}{1+3\rho}}\right]^{\frac{1+3\rho}{4+3\rho}} k^{-\frac{1+3\rho}{4+3\rho}}$$

$$\left[\sum_{i\ge2} k^{-1} \left(k \int_{U_{ki}} q(x)dx\right)^{\frac{4+3\rho}{3}}\right]^{\frac{3}{4+3\rho}}$$

$$= \omega + O\left(k^{-\frac{1+3\rho}{4+3\rho}+\delta}\right)$$

(to bound the last factor within brackets we applied Lemma A 1 of Kallenberg et al. (1985)). By the same line of argument

$\omega_k \leq \omega + O\left(k^{-\frac{1+3\rho}{4+3\rho}+\delta}\right)$. Since $k \int_{U_{k1}} r(x)dx \int_{U_{k1}} q(x)dx = O\left(k^{-\frac{1+3\rho}{4+3\rho}+\delta}\right)$ (by Hölder's inequality), (II.5.1) follows.

Now suppose r is nondecreasing, hence bounded on (0,a). Proceeding as before we obtain $\omega - \omega_k = O\left(k^{-\frac{1+3\rho}{4+3\rho}}\right)$ and the proof is complete. □

REMARK II.5.1. If $r(x) = O(x^{-\delta})$ for all $\delta > 0$, then Lemma II.5.1 guarantees that the LHS of (II.5.1) is of order $o(k^{-\frac{1}{4}-\varepsilon})$ for some $\varepsilon > 0$. One easily generalizes Lemma II.5.1 to partitions where the intervals have variable length. One obtains the same bound if k·minimal length is bounded away from zero. □

Observe that the sum in the LHS of (II.5.1) is a kind of Riemann-Stieltjes sum approximating the integral $\int_0^a r(x)q(x)dx$. Under strong regularity conditions Lemma II.5.1 shows that the precision is of order $O\left(k^{-\frac{1+3\rho}{4+3\rho}+\delta}\right)$. E.g. with $q(x) = g_1(G_0^{-1}(x))/g_0(G_0^{-1}(x))$ and $r(x) = Ah(G_0^{-1}(x))$,

$$C_k'd_k(\eta_n)/(n^{\frac{1}{2}}\eta_n) = \sum_{i=1}^{k} p_{ki}^{-1}(0) \int_{U_{ki}} q(x)dx \int_{U_{ki}} r(x)dx$$

approximates the integral $\int_0^1 q(x)r(x)dx = AE_1\{h(Y)\}$. Even if the conditions of Lemma II.5.1 are not satisfied the precision of this approximation is often of the order $o(1)$, implying $C_k'd_k = O(1)$. In the same manner $B_k'd_k = O(1)$ and $D_k'd_k = O(1)$.

CHAPTER III

THE POWER OF EDF TESTS OF FIT UNDER NON-ROBUST ESTIMATION OF LOCATION-SCALE NUISANCE PARAMETERS

III.1. PRELIMINARIES

III.1.1. Assumptions.

It is well-known that under regularity conditions, including the asymptotic normality of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$ under H_0, the estimated empirical process converges to a Brownian bridge B_θ which may depend on θ (cf. Durbin (1973), Neuhaus (1976), Csörgő and Révész (1981a)):

$$\begin{aligned} &n^{\frac{1}{2}}(\hat{F}_n(\cdot) - G_0(\cdot)) \\ &= n^{\frac{1}{2}}(F_n(\hat{\mu}_n + \cdot\hat{\sigma}_n) - G_0^*(\hat{\mu}_n + \cdot\hat{\sigma}_n;\hat{\theta}_n)) \to_{d_0} B_\theta \quad \text{under } H_0. \end{aligned}$$

In principle this leads to the asymptotic null-distributions of EDF statistics. Of course the only EDF statistics of practical interest are those which have a limit distribution independent of θ under H_0. If the estimator $\hat{\theta}_n$ of θ is location-scale equivariant, i.e. (cf. (II.1.6))

C.III.1 $$n^{\frac{1}{2}}(\hat{\theta}_n - \theta) = \sigma Q_n\left(\frac{Y_1-\mu}{\sigma}, \dots, \frac{Y_n-\mu}{\sigma}\right),$$

then the exact distribution of $\hat{F}_n$ and hence of T_n is invariant under θ. The limiting distribution of T_n is often a bounded functional of $B = B_\theta$, thus T_n is bounded in probability (uniformly in θ). (The statistic $\tilde{T}_n$ often satisfies this relation too.) Therefore equivariant estimators are often used. In this chapter we do not bother about the derivation of the limiting distributions under H_0, but assume $T_n = O_p(1)$ under H_0 and $\tilde{T}_n = O_p(1)$ under H_0 uniformly in θ.

Traditionally one assumes that the estimator $\hat{\theta}_n$ is $\sqrt{n}$-consistent both under H_0 and local alternatives in order to obtain the limiting distributions. Note, however, although the sample mean and the sample variance are optimal estimators in a normal null-hypothesis model, they need not to be

$\sqrt{n}$-consistent for particular heavy-tailed alternatives. In this chapter we consider such strongly non-robust estimators:

C.III.2 $\quad n^{\frac{1}{2}}\|\hat{\theta}_n - \theta\| \to_p \infty$ under H_{1n},

where H_{1n} is a sequence of local alternatives

$$H_{1n} : F^Y \in \{(1-\eta_n)G_0^*(\cdot;\theta) + \eta_n G_1^*(\cdot;\theta); \mu \in \mathbb{R}, \sigma > 0\}$$

such that $n^{\frac{1}{2}}\eta_n = O(1)$.

Estimators based on sample quantiles cannot satisfy C.III.2 because their influence curves are bounded.

III.1.2. EDF tests.

Our consistency Theorem III.2.1 applies to the following EDF test statistics:

the Kolmogorov-Smirnov statistic

(III.1.1) $\quad KS_n = \max\{KS_n^+, KS_n^-\}$,

where

$$KS_n^+ = n^{\frac{1}{2}} \sup_{y\in\mathbb{R}} \{\hat{F}_n(y) - G_0(y)\} \quad \text{and}$$

$$KS_n^- = -n^{\frac{1}{2}} \inf_{y\in\mathbb{R}} \{\hat{F}_n(y) - G_0(y)\},$$

the Kuiper statistic

(III.1.2) $\quad K_n = KS_n^+ + KS_n^-$,

the Cramér-von Mises statistic

(III.1.3) $\quad CM_n = n \int \{\hat{F}_n(y) - G_0(y)\}^2 dG_0(y)$,

the Anderson-Darling statistic

(III.1.4) $\quad AD_n = n \int \{\hat{F}_n(y) - G_0(y)\}^2 d \log\{G_0(y)/(1-G_0(y))\}$,

generalized χ^2-statistics based on random cells

(III.1.5) $\quad MS_n = V_k(\hat{\theta}_n)' \Gamma_k V_k(\hat{\theta}_n)$,

generalized χ^2-statistics based on fixed cells

(III.1.6) $\quad \widetilde{MS}_n = \widetilde{V}_k(\hat{\theta}_n)'\,\Gamma_k \widetilde{V}_k(\hat{\theta}_n)$,

Cressie-Read statistics based on random cells

(III.1.7) $$CR_n(\lambda) = \frac{2n}{\lambda(\lambda+1)} \sum_{i=1}^{k} p_{ki}(0)\left\{\left(\frac{N_{ki}(\hat{\theta}_n)}{np_{ki}(0)}\right)^{\lambda+1} - 1\right\}$$

and Cressie-Read statistics based on fixed cells

(III.1.8) $$\widetilde{CR}_n(\lambda) = \frac{2n}{\lambda(\lambda+1)} \sum_{i=1}^{k} p_{ki}(0,\theta_0,\hat{\theta}_n)\left\{\left(\frac{\overline{N}_{ki}}{np_{ki}(0,\theta_0,\hat{\theta}_n)}\right)^{\lambda+1} - 1\right\} .$$

Here $\theta_0 = (0,1)'$, $\overline{N}_{ki} = \#\{j; Y_j \in I_{ki}\}$, $\widetilde{V}_{ki}(\hat{\theta}_n) = (\overline{N}_{ki} - np_{ki}(0,\theta_0,\hat{\theta}_n))/(np_{ki}(0,\theta_0,\hat{\theta}_n))^{\frac{1}{2}}$ $(i = 1,\ldots,k)$ and Γ_k is a positive definite $(k \times k)$-matrix (the other symbols are defined in Sections I.2 and II.1.4).

The test statistics (III.1.6) and (III.1.8) belong to the class defined in (I.2.9), the other ones to the class defined in (I.2.10). The statistics X_n^2 defined in (II.1.1) are of the form (III.1.5) if $\Lambda(k) - I$ is a nonnegative definite matrix. This condition excludes the (modified) Dzhaparidze-Nikulin statistic. Using estimators based on quantiles it excludes also RRN_n in special cases (cf. Example II.3.2), but quite generally $\Psi(k)^+ - I$ is nonnegative definite (e.g. if $r(\Sigma_k) = k-1$).

III.2. STRONGLY NON-ROBUST ESTIMATION

III.2.1. A consistency theorem for EDF statistics.

Throughout this section we assume C.III.1. and suppose that the EDF statistics are bounded in probability under H_0. In view of the discussion in Section III.1 these assumptions are satisfied in practical applications. In combination with the strongly non-robustness property C.III.2 and a harmless regularity condition this leads to the following consistency theorem.

THEOREM III.2.1. *Assume* C.III.1 *and* C.III.2.

i) *Suppose* G_0 *is differentiable on a non-empty open set* S *such that* $(\forall y \in S)\ g_0(y) > c > 0$. *If the test statistics* (III.1.1) - (III.1.4) *are bounded in probability under* H_0, *then the power of the corresponding tests tends to one under local alternatives* H_{1n}.

ii) *Suppose there exist at least two boundary points* a_{ki} *and* a_{kj}

($i \neq j \in \{1,\dots,k-1\}$) *such that* G_0 *is continuously differentiable with positive derivative on small open balls around* a_{ki} *and* a_{kj}. *If the test statistics* (III.1.5) *and* (III.1.7) *are bounded in probability under* H_0, *then the power of the corresponding tests tends to one under local alternatives* H_{1n}.

iii) *Suppose* G_0 *is continuously differentiable on* $\mathbb{R}$ *and suppose* g_0 *is positive on* $\mathbb{R}$. *If the test statistics* (III.1.6) *and* (III.1.8) *are bounded in probability under* H_0, *then the power of the corresponding tests tends to one under local alternatives* H_{1n} *(but not uniformly in* θ*)*.

PROOF. cf. Section III.5. □

A slight modification of the proof shows that Theorem III.2.1 also applies to the one-sided tests based on KS_n^+ (KS_n^-) provided that S contains positive and negative numbers and that C.III.2 is replaced by $n^{\frac{1}{2}}|\hat{\sigma}_n - \sigma| \to_{P_{1n}} \infty$ and $n^{\frac{1}{2}}(\hat{\mu}_n - \mu) = O_p(1)$ under H_{1n} or by $n^{\frac{1}{2}}(\hat{\mu}_n - \mu) \to_{P_{1n}} \infty$ ($n^{\frac{1}{2}}(\hat{\mu}_n - \mu) \to_{P_{1n}} -\infty$). The result for KS_n^+ also holds for the peak statistic $\max_{1\le i\le k} V_{ki}(\hat{\theta}_n)$ introduced in Dijkstra et al. (1984) in a special case. Examples show that Theorem III.2.1 is not necessarily true for MS_n and $\widetilde{MS}_n$ if Γ_k has an eigenvalue zero (e.g. the Dzhaparidze-Nikulin statistic). Similarly Theorem III.2.1 is not true for the Watson statistic:

$$W_n = CM_n - n[\int \{\hat{F}_n(y) - G_0(y)\}dG_0(y)]^2.$$

Both terms of the RHS tend to infinity under the conditions of Theorem III.2.1 but they may kill each other.

In practice the number of classes on which the statistics (III.1.5) - (III.1.8) are based is often taken larger if n is larger; $k = k_n \to \infty$ as $n \to \infty$. In this setup EDF tests are no longer bounded in probability under H_0. Strengthening the non-robustness condition C.III.2, another version of Theorem III.2.1 is still in force.

COROLLARY III.2.2. *Let* $k = k_n \to \infty$ *as* $n \to \infty$, *assume* $k^{-\frac{1}{2}}n^{\frac{1}{2}}\|\hat{\theta}_n - \theta\| \to_{P_{1n}} \infty$, C.III.1 *and suppose that the eigenvalues of* Γ_k ($k \in \mathbb{N}$) *are bounded away from* 0.

i) *Suppose, for each* k, *there exist at least two boundary points* a_{ki} *and* a_{kj} ($i \neq j \in \{1,\dots,k-1\}$) *such that* G_0 *is continuously differentiable with positive derivative on small open balls (with fixed radius) around* a_{ki} *and* a_{kj} *and such that* $\max\{|a_{ki}|,|a_{kj}|\} < M < \infty$ *and* $|a_{ki} - a_{kj}| > \delta > 0$. *If the test statistics* (III.1.5) *and* (III.1.7) *are of order* $O_p(k)$ *under*

H_0, *then the power of the corresponding tests tends to one under local alternatives* H_{1n}.

ii) *Suppose* G_0 *is continuously differentiable on* $\mathbb{R}$ *and suppose* g_0 *is positive on* $\mathbb{R}$. *Assume that, for each* k, *there exist at least two boundary points* a_{ki} *and* a_{kj} $(i \neq j \in \{1,\ldots,k-1\})$ *such that* $(\forall\ k)$ $\max\{|a_{ki}|,|a_{kj}|\} < M < \infty$ *and* $|a_{ki} - a_{kj}| > \delta > 0$. *If the test statistics* (III.1.6) *and* (III.1.8) *are of order* $O_p(k)$ *under* H_0 *uniformly in* θ, *then the power of the corresponding tests tends to one under local alternatives* H_{1n}.

PROOF. cf. Section III.5. □

Compare the results of Theorem III.2.1 and Corollary III.2.2 with more classical situations. Assuming $\hat{\theta}_n$ to be $\sqrt{n}$-consistent and to satisfy some additional regularity conditions under local alternatives, the EDF statistics are also bounded in probability under H_{1n}. Hence the power is bounded away from one. Non-robust estimation thus leads to a substantial gain of power!

Of course it is clear that one cannot choose estimators that satisfy C.III.2 for the whole range of alternatives. Theorem III.2.1 and Corollary III.2.2 do not give any information about the possible loss in directions where $\hat{\theta}_n$ is $\sqrt{n}$-consistent under H_{1n}. Simulations, however, show that losses are comparatively small (cf. Section III.4).
Since sample moments are very sensitive to heavy tails, EDF tests based on sample moment type estimators are expected to perform well against heavy tailed alternatives (not only for a normal null-hypothesis!). Medians, trimmed means, interquartile ranges or median absolute deviations are less attractive estimators from this point of view.

Stephens (1974), pointing at the high power of several goodness-of-fit tests when parameters are estimated compared to the same tests with known parameters, merely commented that the precise location and scale are relatively unimportant when fitting data. For EDF tests the effect on non-robust estimation is a more satisfactory explanation, since in the examples considered by Stephens the sample mean and sample variance are used as estimators. The tests discussed by Witting (1959) and Bofinger (1973), however, for which both the cell boundaries and the estimators are based on sample quantiles, will not enjoy the increased power due to non-robust estimation.

III.2.2. Extension to Neyman smooth tests.

Results similar to Theorem III.2.1 hold for quite different classes of goodness-of-fit tests. Consider the Neyman smooth tests based on the test statistics

(III.2.1) $\quad N_n = U_n(\hat{\theta}_n)'\Gamma_k U_n(\hat{\theta}_n)$,

where $U_n(\theta)$ is a k-vector with components

$$U_{ni}(\theta) = n^{-\frac{1}{2}} \sum_{j=1}^{n} \left\{G_0^i\left(\frac{Y_j-\mu}{\sigma}\right) - (i+1)^{-1}\right\} \qquad (i=1,\ldots,k)$$

and Γ_k is a positive definite $(k\times k)$-matrix (cf. Thomas and Pierce (1979) for a recent discussion). A natural choice for Γ_k is the inverse of the asymptotic covariance matrix of $U_n(\hat{\theta}_n)$ under H_0.

PROPOSITION III.2.3. *Assume* C.III.1, C.III.2, $n^{\frac{1}{2}}(\hat{\theta}_n - \theta) = O_p(1)$ *under* H_0 *and*

(III.2.2) $\quad \min\{|\hat{\mu}_n - \mu|, |\hat{\sigma}_n - \sigma|\} \to_{P_{1n}} 0.$

Suppose G_0 *is symmetric about zero, twice continuously differentiable with bounded derivatives* g_0 *and* g_0' *and bounded* $yg_0(y)$ *and* $y^2g_0'(y)$ *(on* $\mathbb{R}$*). Then the power of the Neyman smooth test* N_n*, with* $k \geq 2$*, tends to one under* H_{1n}.

PROOF. cf. Section III.5. □

REMARK III.2.1. In view of the preceding results the condition (III.2.2) is somewhat startling. Note, however, that a lot of estimators are consistent under local alternatives although they are not $\sqrt{n}$-consistent. □

From the proofs it is obvious that all previous results continue to hold if θ is one-dimensional, i.e. in pure location or in pure scale families. In this case the condition (III.2.2) in Proposition III.2.3 can be suppressed. In pure location models this proposition also holds for $k = 1$. Note that for pure scale families on $\mathbb{R}^+$ g_0 need only be positive and continuous on $\mathbb{R}^+$ in Theorem III.2.1 iii).

III.3. COMPLEMENTARY RESULTS FOR NON-ROBUST ESTIMATION

The non-robustness condition C.III.2 imposed on the estimator $\hat{\theta}_n$ seems to be very strong. For instance if H_0 specifies a normal location family and G_1 is a standard Cauchy alternative, the sample mean $\overline{Y}_n$ does not yet satisfy C.III.2 under local contamination families with contamination factor $\eta_n = n^{-\frac{1}{2}}\gamma + o(n^{-\frac{1}{2}})$, because $n^{\frac{1}{2}}(\overline{Y}_n - \mu)$ still has a (non-normal) limit distribution. Hence Theorem III.2.1 only applies to rather extreme classes of alternatives. However, to describe small sample size behaviour extreme classes of alternatives may be of interest.

In this section we show that the good power properties of RRN_n under non-robust estimation also holds true for broader, less extreme classes of alternatives. Although this is hard to prove we believe nevertheless that this extends to general EDF statistics. Numerical evidence in a couple of examples supports this view (cf. Section III.4).

Let $\eta_n = n^{-\frac{1}{2}}\gamma + o(n^{-\frac{1}{2}})$, consider local families of type (II.1.2) and assume C.II.1 and C.II.2 (with $Q_n = o_p(1)$ both under G_0 and G_{η_n}). Then, under H_{1n}, RRN_n converges in distribution to a noncentral chi-square distribution with $r(\Sigma_k)$ degrees of freedom (cf. Moore and Spruill (1975)). In the remainder of this section we assume $r(\Sigma_k) = k-1$; hence

(III.3.1) $$RRN_n \to_{d_{1n}} \chi^2_{k-1}(\Delta_k(\eta_n)),$$

where $\Delta_k(\eta_n)$ is defined in (II.1.14). One can show by straightforward calculation that Σ_k^+ is given by

$$\Sigma_k^+ = I_k - q_k q_k' - (B_k - C_k)R^{-1}(B_k - C_k)' +$$
$$+ (I_k - (B_k - C_k)R^{-1}(B_k - C_k)')C_k S^{-1} C_k'(I_k - (B_k - C_k)R^{-1}(B_k - C_k)'),$$

where

$$R = A + (B_k - C_k)'(B_k - C_k) \quad \text{and}$$

$$S = A - C_k'C_k + C_k'(B_k - C_k)R^{-1}(B_k - C_k)'C_k.$$

III.3.1. Bias effects.

Consider the simple situation that $\theta = \mu$. To analyse the effect of the estimation procedure we compare the maximum likelihood estimator $\hat{\theta}_n^{ML}$ with some arbitrary estimator $\hat{\theta}_n$. Suppose that both estimators allow the representation (II.1.6) and that the limiting distributions under H_{1n} are given by (III.3.1) (under H_0 $\Delta_k(\eta_n) = 0$). For the influence function h^{ML} of $\hat{\theta}_n^{ML}$ we assume $E_1\{h^{ML}(Y)\} = B_k' d_k(\eta_n) = 0$. In location families this holds true e.g. if both the densities g_0 and g_1 and the partition $I_{k1}, \ldots, I_{kk}$

are symmetric about zero. This is not uncommon.
To compare the respective powers it is sufficient to analyse $\Delta_k(\eta_n) - \Delta_k^{ML}(\eta_n)$. Clearly, if $\Delta_k(\eta_n) > \Delta_k^{ML}(\eta_n)$, the Rao-Robson-Nikulin test based on $\hat{\theta}_n$ will be asymptotically at least as efficient than the same test based on $\hat{\theta}_n^{ML}$. Tedious algebra shows (A is now a scalar)

$$\begin{aligned} \Delta_k(\eta_n) - \Delta_k^{ML}(\eta_n) &= \Delta_k(\eta_n) - \|d_k(\eta_n)\|^2 \\ &= \gamma^2 B_k' B_k (E_1\{h(Y)\} - S_1(h))^2 / (1 - 2C_k' A^{-1} B_k + B'_k A^{-1} B_k), \end{aligned} \tag{III.3.2}$$

where

$$S_1(h) = \gamma^{-1} A^{-1} C_k' d_k(\eta_n) = \sum_{i=1}^{k} p_{ki}(1) p_{ki}^{-1}(0) \int_{I_{ki}} h(y)\, dG_0(y)$$

is a kind of Riemann sum corresponding to $E_1\{h(Y)\}$.
By assumption the denominator is positive. Hence the maximum likelihood estimator is the worst choice for $\hat{\theta}_n$. Note, however, that in symmetric location models $E_1\{h(Y)\} - S_1(h) = 0$ if h is antisymmetric; although the choice of an asymmetric influence function improves the power this is quite unnatural.
Note that $E_1\{h(Y)\} - S_1(h) = 0$ if h is constant on the invervals $I_{k1},\ldots,I_{kk}$. Conversely, since $S_1(h)$ is bounded by $\frac{1}{2} \max_{1\le i\le k} p_{ki}^{-1}(0) E_0\{|h(Y)|\}$, it follows that $|E_1\{h(Y)\} - S_1(h)|$ is large for those alternatives for which $|E_1\{h(Y)\}|$ is large, i.e. for which $\hat{\theta}_n$ has a large bias. Of course, one can also compare two arbitrary estimators; in the RHS of (III.3.2) one obtains the difference of two complex expressions. A comparison is difficult because it is not the bias $E_1\{h(Y)\}$ but the difference $E_1\{h(Y)\} - S_1(h)$ which plays a crucial role.

Imposing similar conditions in scale or in location-scale models one may derive similar expressions (cf. Drost et al. (1985)). The conditions $E_1\{h^{ML}(Y)\} = B_k' d_k(\eta_n) = 0$ are, however, rather restrictive in these cases.

III.3.2. Variance effects.

In the local contamination family C.II.2 implies

$$\mathrm{cov}_{\eta_n}\{h(Y)\} \to A^{-1} = \mathrm{cov}_0\{h(Y)\},$$

i.e. the asymptotic variance of $\hat{\theta}_n$ under H_0 and under H_{1n} is the same. In view of the previous results it is, however, not unlikely that a much larger variance of $\hat{\theta}_n$ under alternatives than under H_0 might increase the power of the Rao-Robson-Nikulin test. To model this situation, we slightly modify the family of local alternatives:

(III.3.3) $\quad \{G^{*}_{\tau,n}(\cdot;\theta) = (1-\tau_n)G_0(\frac{\cdot-\mu}{\sigma}) + \tau_n G_{1n}(\frac{\cdot-\mu}{\sigma});\mu \in \mathbb{R},\sigma > 0\},$

where $\tau_n = n^{-\frac{1}{2}}\gamma + o(n^{-\frac{1}{2}})$, G_{1n} has density g_{1n} and satisfies, for given partition $I_{k1},\ldots,I_{kk}$,

(III.3.4) $\quad G_{1n}(a_{ki}) = G_1(a_{ki}) \qquad (i=1,\ldots,k-1).$

We briefly denote the local alternatives (III.3.3) by H^{*}_{1n} and assume that the representation (II.1.6) of $\hat{\theta}_n$ continues to hold under H^{*}_{1n} with

$$\text{C.III.3} \quad \begin{cases} \text{-a} \quad E_{\tau,n}\{h(Y)\} = 0 \quad (n \in \mathbb{N}) \\ \text{-b} \quad \lim_{n\to\infty} \operatorname{cov}_{\tau,n}\{h(Y)\} = A^{-1} + A_2 \\ \text{-c} \quad \lim_{n\to\infty} n^{-\frac{1}{2}} \int \|h(y)\| dG_{1n}(y) = 0 \\ \text{-d} \quad Q_n(Y_1,\ldots,Y_n) = o_p(1) \text{ under } G_{\tau,n} \end{cases}$$

where A_2 is a finite (2×2)-matrix. For A_2 not to vanish the tails of the distribution determined by G_{1n} must strongly depend on τ_n. The trick of condition (III.3.4) is that the multinomial part of the distribution of RRN_n behaves exactly as in the local family H_{1n}. Note that (III.3.3) forces A_2 to be nonnegative definite.

We first show that Theorem 4.2 of Moore and Spruill (1975) remains valid in the present situation.

LEMMA III.3.1. *Consider the family of alternatives* (III.3.3) *and an estimator* $\hat{\theta}_n$ *of* θ. *Assume* C.II.1, C.II.2 (*under* H_0 *with* $Q_n = o_p(1)$) *and* C.III.3, *let* $B_k' d_k(\tau_n) = C_k' d_k(\tau_n) = 0$, *let* $r(\Sigma_k) = k-1$ *and* $k \geq 4$. *Then*

(III.3.5) $\quad RRN_n \to_{d^{*}_{1n}} \chi^2_{k-3}(\|d_k(\tau_n)\|^2) + (1+\rho_1)\chi^2_1 + (1+\rho_2)\chi^2_1,$

where ρ_1 *and* ρ_2 *are the two eigenvalues of* $A_2^{\frac{1}{2}} B_k' \Sigma_k^{+} B_k A_2^{\frac{1}{2}}$.

PROOF. Since the distribution of RRN_n is invariant under θ, take $\theta = \theta_0 = (0,1)'$. Obviously Lemma II.4.1 remains valid under H^{*}_{1n} with $R_k = o_p(1)$ (cf. also Moore and Spruill (1975), Theorem 4.1). Asymptotically the covariance matrix of the principle part of $V_k(\hat{\theta}_n)$ under H^{*}_{1n} equals

$$\Sigma_{\tau,n} = \Sigma_k + B_k A_2 B_k',$$

where we have used C.III.3-a,b and Remark II.4.1. The central limit theorem implies

$$V_k(\hat\theta_n) \to_{d^*_{1n}} N(d_k(\tau_n), \Sigma_{\tau,n})$$

since, for each linear combination of the components of $V_k(\hat\theta_n)$, the Lindeberg condition holds.

The conditions $B_k' d_k(\tau_n) = C_k' d_k(\tau_n) = 0$ imply $\Sigma_{\tau,n}\, d_k(\tau_n) = \Sigma_k\, d_k(\tau_n) = d_k(\tau_n)$. Let P_k be an orthonormal matrix such that $P_k d_k(\tau_n) = (\|d_k(\tau_n)\|, 0, \ldots, 0)'$ and $P_k (\Sigma_k^+)^{\frac{1}{2}} \Sigma_{\tau,n} (\Sigma_k^+)^{\frac{1}{2}} P_k' = I_k - P_k q_k q_k' P_k' + P_k (\Sigma_k^+)^{\frac{1}{2}} B_k A_2 B_k' (\Sigma_k^+)^{\frac{1}{2}} P_k' =$
$= \mathrm{diag}(1, \ldots, 1, 1+\rho_1, 1+\rho_2, 0)$. Hence

$$\begin{aligned} RRN_n &= \|(\Sigma_k^+)^{\frac{1}{2}} V_k(\hat\theta_k)\|^2 \\ &= \|P_k (\Sigma_k^+)^{\frac{1}{2}} V_k(\hat\theta_n)\|^2 \to_{d^*_{1n}} \chi^2_{k-3}(\|d_k(\tau_n)\|^2) + (1+\rho_1)\chi_1^2 + (1+\rho_2)\chi_1^2 . \end{aligned}$$

This can also be derived from Corollary 2.2 in Dik and de Gunst (1985). □

__PROPOSITION III.3.2.__ *Let* $\hat\theta_n^{(i)}$ *satisfy the conditions of Lemma* III.3.1 *and let* $\rho_1^{(i)}$ *and* $\rho_2^{(i)}$ *be the corresponding eigenvalues* (i = 1,2). *If* $\rho_1^{(1)}$ *and* $\rho_2^{(1)}$ *are both larger (smaller) than* $\rho_1^{(2)}$ *and* $\rho_2^{(2)}$, *then the Rao-Robson-Nikulin test based on* $\hat\theta_n^{(1)}$ *is asymptotically more (less) powerful than the same test based on* $\hat\theta_n^{(2)}$ *against the local alternatives* (III.3.3).

__PROOF__. Immediate from Lemma III.3.1 and the fact that both test statistics are asymptotically distributed as χ^2_{k-1} under H_0. □

Roughly speaking, the proposition states that if A_2 is large relative to Σ_k, the test RRN_n has a large asymptotic power. This will generally be true if the variance of $\hat\theta_n$ is much larger under H^*_{1n} than under H_0. As noted before, the condition $B_k' d_k(\tau_n) = C_k' d_k(\tau_n) = 0$ is severely restrictive in (location-) scale families, but is often satisfied in symmetric location families.

__REMARK III.3.1__. Suppose the matrix $\Sigma_k^{(1)}$ associated with $\hat\theta_n^{(1)}$ satisfies $r(\Sigma_k^{(1)}) = r \le k-2$ (instead of $r = k-1$). Then the first term on the right in (III.3.5) has r-2 df's. Since in this case the asymptotic null distribution of RRN_n also has r-2 df's, we still have the implication $\rho_1^{(1)} > \rho_1^{(2)}$, $\rho_2^{(1)} > \rho_2^{(2)} \Rightarrow$ the asymptotic power of RRN_n is higher with $\hat\theta_n^{(1)}$ than with $\hat\theta_n^{(2)}$, but the implication for the reverse inequalities is not necessarily true, depending on how large the differences $\rho_i^{(2)} - \rho_i^{(1)}$ are. □

Results similar to those in subsections III.3.1 and III.3.2 cannot be obtained for the Watson-Roy test, or in general for MS_n and $CR_n(\lambda)$, since the asymptotic null-distribution depends on the choice of the estimator. Nevertheless, in view of the close relationship between RRN_n, MS_n and $CR_n(\lambda)$ it is quite likely that materially the same properties hold for these classes of tests.

Since influence functions of almost all estimators of location and scale are bounded on compact sets, light-tailed alternatives with heavy centers will not distort the random intervals as much as heavy-tailed alternatives. Hence it is more rewarding to choose estimators appropriate for heavy-tailed alternatives than for light-tailed alternatives.

III.4. NUMERICAL EXAMPLES

The asymptotic theory of the previous sections suggest the following *rule of thumb*:

> *To achieve a high power of EDF tests against a class of alternatives, non-robust estimators are best.*

Moment estimators are expected to perform well against heavy-tailed alternatives. One expects a similar result for light-tailed alternatives: high powers for estimators with influence curves concentrated in the center of the alternative density. However, since the densities considered here are bounded, the latter effect is probably small.

The power of various EDF tests is compared for a couple of well-known estimators $\hat{\theta}_n$ of θ. The estimators, based on a sample $Y_1,\ldots,Y_n$ are: M_n = sample median, $\overline{Y}_{n,\cdot 1}$ = trimmed mean with 10% trimming on both sides, $\overline{Y}_n$ = sample mean, Mad_n = median of $|Y_j - M_n|$ $(j=1,\ldots,n)$, $D_n = n^{-1}\Sigma_{j=1}^{n} |Y_j - \overline{Y}_n|$, S_n^2 = sample variance and I_n = interquartile range. The influence curves are given in Table II.4.1 (ζ_α denotes the α point of the distribution G_0).

Table III.4.1. Influence functions

estimator	influence function	conditions for C.II.2 under H_0
estimator for μ		
M_n	$\{g_0(\zeta_{.5})\}^{-1}\{\frac{1}{2} - 1_{(-\infty,\zeta_{.5})}(y)\}$	$g_0(\zeta_{.5}) > 0,\ \zeta_{.5} = 0$
$\overline{Y}_{n,\cdot 1}$	$\begin{cases} \frac{5}{4}\zeta_{.9}\,\mathrm{sgn}(y) & \lvert y\rvert > \zeta_{.9} \\ \frac{5}{4}y & \lvert y\rvert < \zeta_{.9} \end{cases}$	g_0 symmetric and $g_0(\zeta_{.9}) > 0$
$\overline{Y}_n$	y	$E_0\{Y\} = 0,\ E_0\{Y^2\} < \infty$
estimator for σ		
$\mathrm{Mad}_n/\zeta_{.75}$	$\{2g_0(\zeta_{.75})\zeta_{.75}\}^{-1} \cdot \{\frac{1}{2} - 1_{(-\zeta_{.75},\zeta_{.75})}(y)\}$	g_0 symmetric and $g_0(\zeta_{.75}) > 0$
$D_n/E_0\lvert Y\rvert$	$\{E_0\lvert Y\rvert\}^{-1}\lvert y\rvert - 1$	$E_0\{Y\} = 0,\ E_0\{Y^2\} < \infty$
S_n	$\frac{1}{2}\{(y-r)^2 - 1\}$	$E_0\{Y\} = r,\ \mathrm{var}_0\{Y\} = 1,\ E_0\{Y^4\} < \infty$
$\overline{Y}_n$	$y - 1$	$E_0\{Y\} = 1,\ E_0\{Y^2\} < \infty$
$I_n/\log 3$	$\begin{cases} 0 & y < \zeta_{.25} \\ -\frac{4}{3}(\log 3)^{-1} & \zeta_{.25} < y < \zeta_{.75} \\ \frac{8}{3}(\log 3)^{-1} & y > \zeta_{.75} \end{cases}$	g_0 exponential

In location-scale models the influence functions are obtained by combining the corresponding influence functions.

We consider four different null hypotheses and simulate for each of them the power at a heavy-tailed, a light-tailed and a skew alternative.

I. H_0: normal location family $N(\mu,1)$, $\mu \in \mathbb{R}$.

Alternatives: Cauchy ($\sigma = \frac{1}{2}$) density $2\pi^{-1}(1+4y^2)^{-1}$ heavy-tailed

normal ($\sigma = \frac{3}{4}$) density $\frac{4}{3}\phi(\frac{4}{3}y)$ light-tailed

Gumbel density $e^{-y}\exp(-e^{-y})$ skew

Estimators of μ: M_n, $\overline{Y}_{n,\cdot 1}$ and $\overline{Y}_n$.

II. H_0: normal location-scale family $N(\mu,\sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$.

Alternatives: Cauchy density $\pi^{-1}(1+y^2)^{-1}$ heavy-tailed

normal mixture density $\frac{1}{2}\{\phi(y+\frac{3}{2}) + \phi(y-\frac{3}{2})\}$

Gumbel density $e^{-y}\exp(-e^{-y})$ skew

Estimators of (μ,σ): $(M_n, \mathrm{Mad}_n/\Phi^{-1}(\frac{3}{4}))$, $(\overline{Y}_n, \mathrm{Mad}_n/\Phi^{-1}(\frac{3}{4}))$,
$(\overline{Y}_n, (\frac{\pi}{2})^{\frac{1}{2}} D_n)$ and $(\overline{Y}_n, S_n)$.

III. H_0: Laplace location family, density $\frac{1}{2}\exp(-|y-\mu|)$, $\mu \in \mathbb{R}$.

Alternatives: Cauchy ($\sigma = \frac{1}{2}$) density $2\pi^{-1}(1+4y^2)^{-1}$ heavy-tailed

normal ($\sigma = \frac{3}{4}$) density $\frac{4}{3}\phi(\frac{4}{3}y)$ light-tailed

Gumbel density $e^{-y}\exp(-e^{-y})$ skew

Estimators of μ: M_n, $\overline{Y}_{n,\cdot 1}$ and $\overline{Y}_n$.

IV. H_0: exponential scale family, density $\sigma^{-1}e^{-(y/\sigma)}1_{(0,\infty)}(y)$, $\sigma > 0$.

Alternatives: gamma ($\alpha = \frac{3}{2}$) $(\Gamma(\frac{3}{2}))^{-1}y^{\frac{1}{2}}e^{-y}1_{(0,\infty)}(y)$ heavy-tailed

gamma ($\alpha = \frac{3}{4}$) $(\Gamma(\frac{3}{4}))^{-1}y^{-\frac{1}{4}}e^{-y}1_{(0,\infty)}(y)$ light-tailed

halfnormal $2\phi(y)1_{(0,\infty)}(y)$ light-tailed

Estimators of σ: $I_n/\log 3$, $\overline{Y}_n$ and S_n.

Comment: According to the rule of thumb, the tests based on the estimators should be increasingly (decreasingly) powerful against heavy-tailed and skew (light-tailed) alternatives. The third problem is included, not for its intrinsic importance, but to demonstrate that efficient estimation under H_0 is not necessarily effective.

We verify the assumptions of the previous sections. The conditions C.II.1 and C.III.1 are trivially satisfied in all cases and C.II.2 is also fulfilled with the following exception: the estimators involving $\overline{Y}_n$ with respect to Cauchy alternatives.

For symmetric null-distributions and alternatives (ignoring Cauchy alternatives but e.g. including Laplace heavy-tailed alternatives) the first two location estimators listed satisfy assumption C.III.3; note that $A_2 = 0$ for M_n, $\overline{Y}_{n,\cdot 1}$ and $(M_n, \mathrm{Mad}_n/\Phi^{-1}(\frac{3}{4}))$. The only example satisfying the non-robustness condition C.III.2 is given in case II: the Cauchy alternative with estimator $(\overline{Y}_n,S_n)$: a very high power is thus expected. The sample mean $\overline{Y}_n$ alone does not satisfy the non-robustness condition at the local Cauchy shift alternatives and hence our theory does not predict high power in this case. However, to embed a sample of moderate size from a Cauchy distribution in a local contamination family, this family must have an even heavier-tailed distribution as an endpoint. For such a contamination family the non-robustness condition will be satisfied. Of course such a family will not often exist, but the argument explains nevertheless that the effect of heavy-tailed alternatives extends much further than the non-robustness condition suggests.

Monte Carlo experiments have been run to estimate the true power of several goodness-of-fit tests for sample size n = 50 and nominal level α = .05. For the chi-square type tests LR_n, WR_n and RRN_n equiprobable cells (under H_0) are employed with k = 4,5,6,7,8,9,10,12 and 15. Note that $r(\Sigma_k) < k-1$ when using estimators involving $\hat{\mu}_n = M_n$ and k even: $r(\Sigma_k) = k-3$ if $\hat{\theta}_n = (M_n,Mad_n)$ and k is a multiple of 4, otherwise $r(\Sigma_k) = k-2$. Since $V_k(\hat{\theta}_n) \in \text{col.}\ (\Sigma_k)$ for these cases the Rao-Robson-Nikulin statistic is invariant under the choice of Σ_k^-. In fact we used a more simple form than the Moore-Penrose generalized inverse Σ_k^+.

Critical values are estimated for each H_0 based on 20000 samples and compared to the asymptotic critical values from χ^2 distributions for LR_n, WR_n and RRN_n and the critical values reported in Stephens (1974). Quite surprisingly the obtained values were close to the asymptotic values for moment estimators while, for robust estimators, the asymptotic values often lead to errors of more than .005 in size.

For the EDF tests the power at the alternatives in the testing problems I - IV was simulated using 10000 samples in each case, thus reducing simulation error to less than .01 with confidence .95. For each test and estimator the same samples were employed for better comparison.

The simulated powers are displayed in Figures III.4.1 - III.4.4. In most cases the power differences can be well explained by the rule of thumb. In problem II the very low power of the test based on $(M_n,cMad_n)$ stands out; the predicted increase of power of the tests based on consecutively $(\overline{Y}_n, cMad_n)$, $(\overline{Y}_n, cD_n)$ and $(\overline{Y}_n, S_n)$ against heavy-tailed alternatives is rather small, suggesting that estimation of μ has more effect on the power than estimation of σ. In problem III the test based on the median (the asymptotically efficient estimator) does not perform well, in agreement with the rule of thumb. For case IV the picture is less clear. Although for such a family Theorem III.2.1 still applies, the assumptions of Section III.3 are far from satisfied due to the lack of symmetry. With skew distributions the relative robustness of different estimators is less clear.

Full location-scale families generally do not satisfy the assumptions of Section III.3 either, since the scale component violates the symmetry conditions. This may not be so serious if the effect of estimating location is more important than the effect of estimating scale and if for fixed scale the family is symmetric (cf. problem II).

Instead of using simulation, the true power of the Rao-Robson-Nikulin test can also be approximated by the theoretical asymptotic χ^2 power computed from (III.3.1), replacing η_n by 1. Another approximation is suggested by (III.3.5), taking $\tau_n = 1$ and determining ρ_1 and ρ_2 from

$$A_2 = \mathrm{cov}_1\{h(Y)\} - \mathrm{cov}_0\{h(Y)\}$$

for those cases where the covariance matrices are finite. Note that the approximations coincide whenever $A_2 = 0$. Both approximations, however, turn out to be unreliable; errors between .1 and .25 or larger are quite common (for $n = 50$).

Apart from the numerical inaccuracy of the approximation, there is also a more theoretical problem. The true power of the tests does not depend on the values of the location-scale parameters, but the asymptotic power in the local models H_{1n} and H^*_{1n} does depend on the choice of the alternative G_1 (or $G_{1,n,k}$) and hence on the choice of the location-scale parameters defining G_1 (or $G_{1,n,k}$). In symmetric pure location models it is natural to fix G_1 so that points of symmetry under G_0 and G_1 coincide, but in other models the choice is not unambiguous and yet determines the asymptotic power. As an extreme case, when testing the normality hypothesis of problem II with 50 observations, the χ^2 approximation (III.3.1) yields considerable power at fixed normal (!) alternatives if one takes

Figure III.4.1. Normal location null-hypothesis. Powers under three alternatives for several EDF tests and three estimators: ·················· = M_n; ---------- = $\overline{Y}_{n,.1}$; ———— = $\overline{Y}_n$.

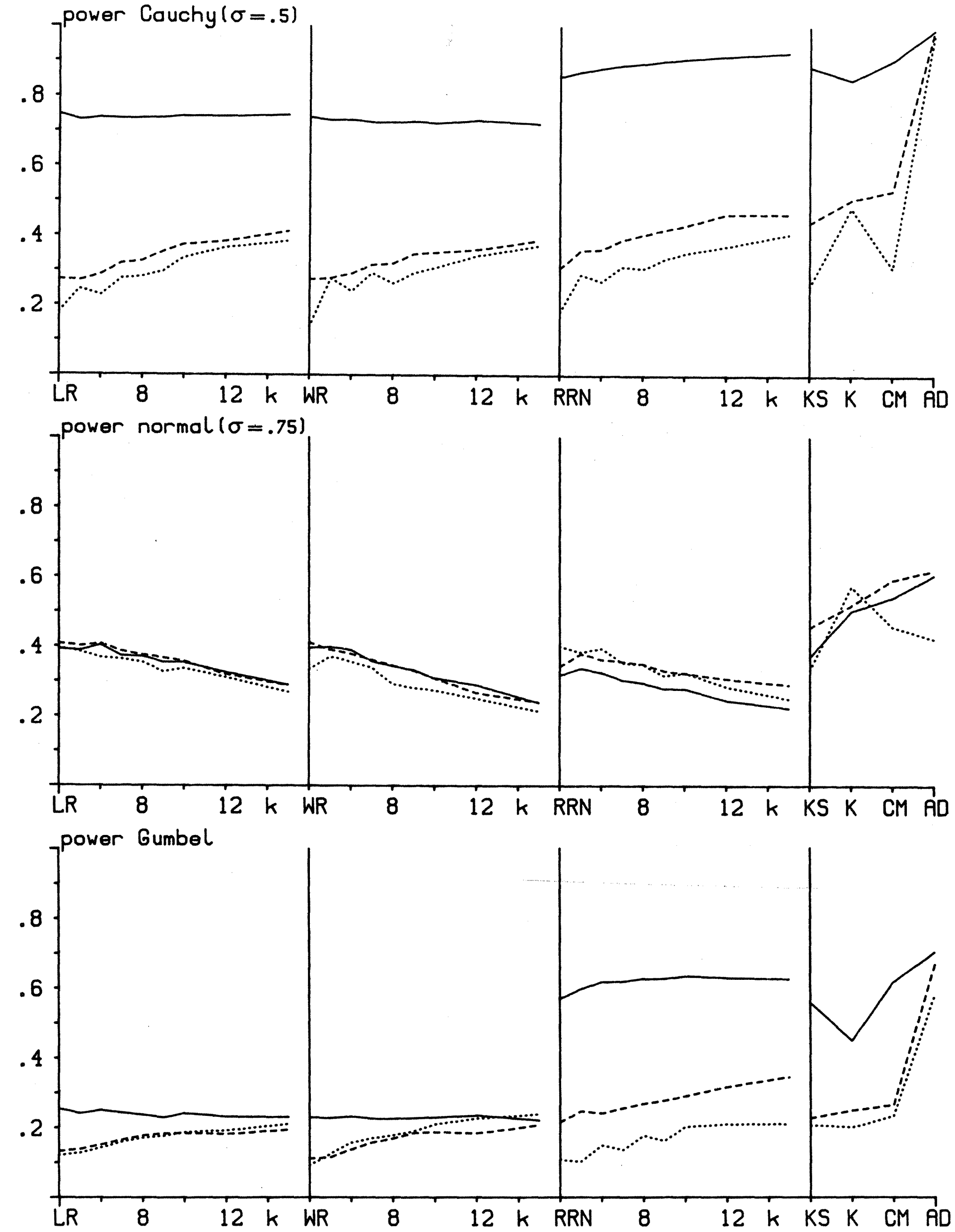


Figure III.4.2. Normal location-scale null-hypothesis. Powers under three alternatives for several EDF tests and four estimators: ·········· = (M_n, Mad_n); - - - - - - = $(\overline{Y}_n, Mad_n)$; —·—·— = $(\overline{Y}_n, (\pi/2)^{\frac{1}{2}} D_n)$; ———— = $(\overline{Y}_n, S_n)$.

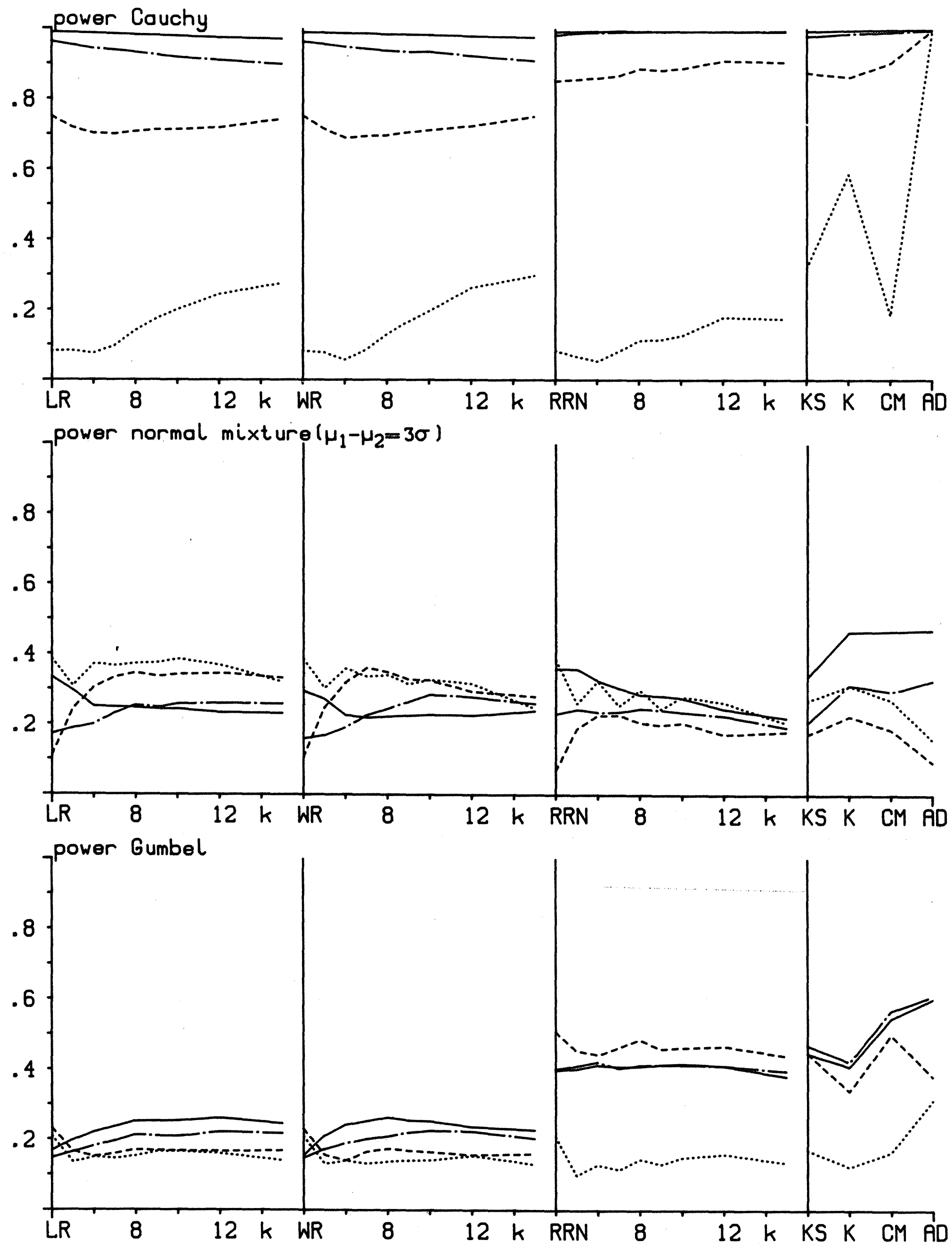


Figure III.4.3. Laplace location null-hypothesis. Powers under three alternatives for several EDF tests and three estimators: .... = M_n; ----------- = $\overline{Y}_{n,.1}$; ———— = $\overline{Y}_n$.

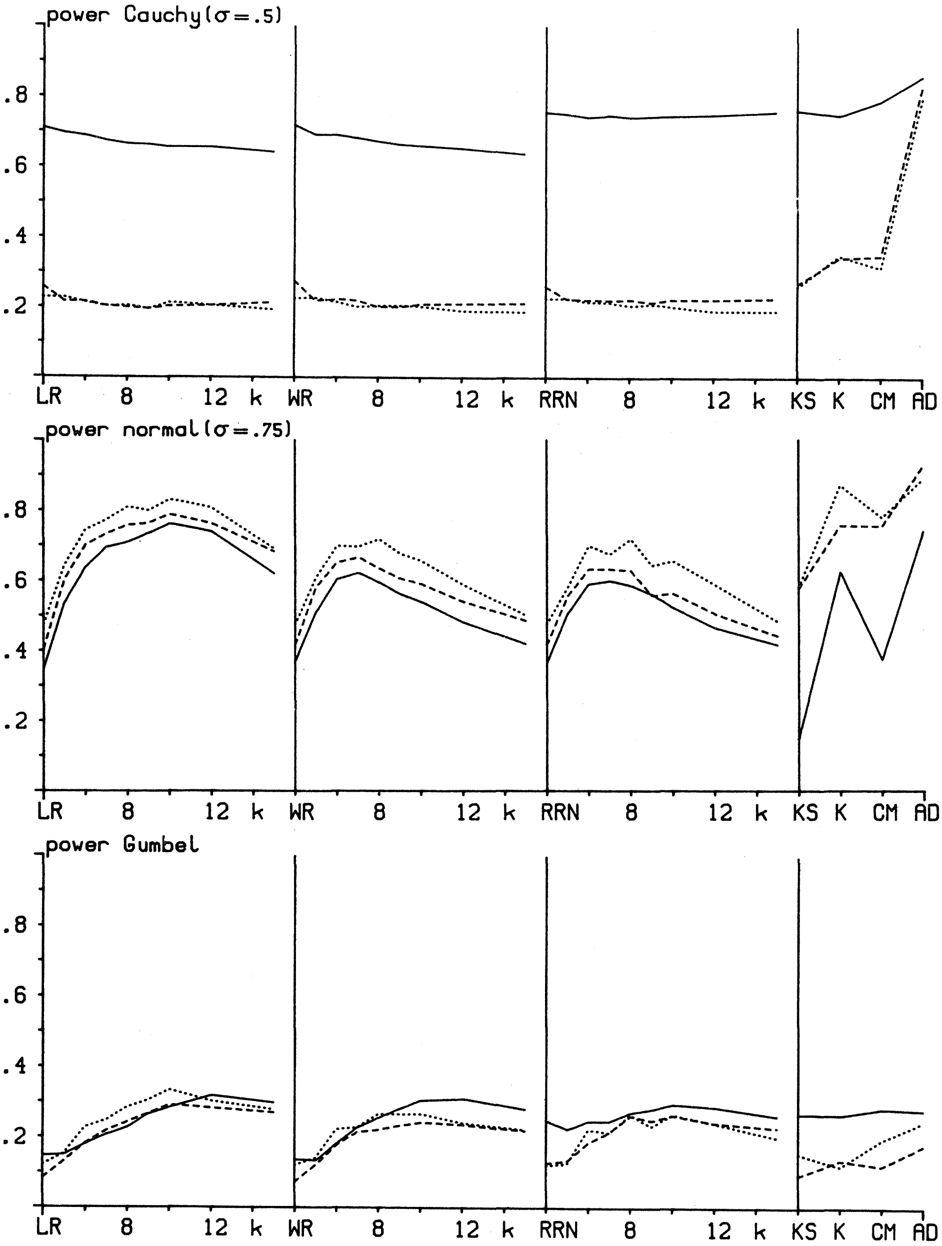
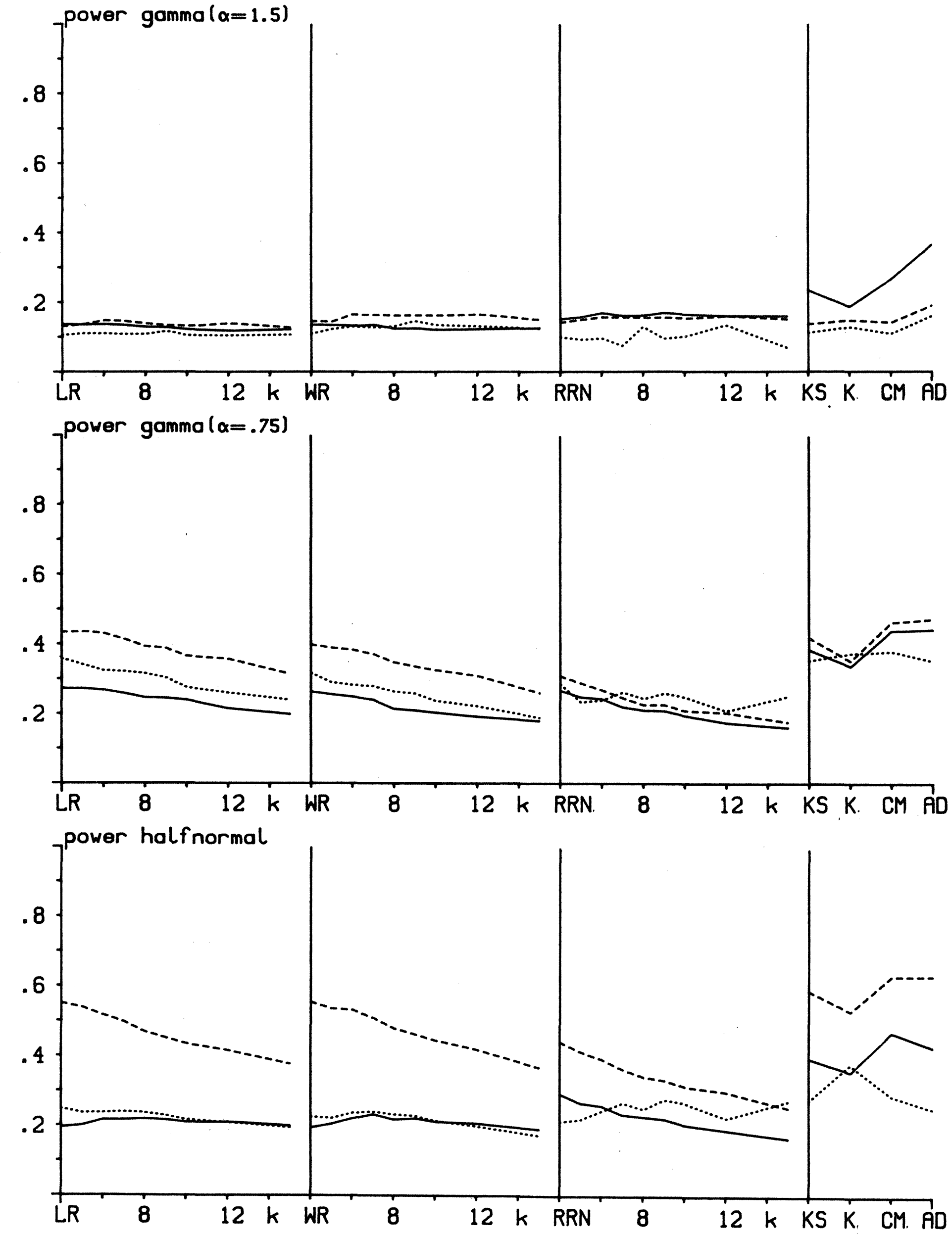


Figure III.4.4. Exponential scale null-hypothesis. Powers under three alternatives for several EDF tests and three estimators: .... = $I_n/\log 3$; ———— = $\overline{Y}_n$; - - - - - = S_n.



G_0 to be $N(0,1)$ and G_1 to be $N(0,\sigma^2)$ with σ^2 much larger than one (cf. also Example II.3.4 for problem IV). The reason, of course, is that the corresponding local family of alternatives consists of mixtures of normals which are non-normal. A sound asymptotic approximation to the small-sample power of chi-square type goodness-of-fit tests in the presence of location-scale parameters still seems to be lacking.

The simulations also illustrate the theory of Chapter II. For the Watson-Roy test and the Rao-Robson-Nikulin test we expect an increasing (decreasing) power as a function of k for heavy (light) tailed alternatives. Comparison of Figures III.4.1 - III.4.4 with Figure 1 of Kallenberg et al. (1985) indicates that the effect of the choice of k is relatively small in the presence of nuisance parameters. Note also the higher power of the Rao-Robson-Nikulin test compared to the Watson-Roy and the likelihood ratio test. The most powerful test statistic, however, seems to be the Anderson-Darling statistic (cf. also Green and Hegazy (1976)).

III.5. PROOFS

III.5.1. Proof of Theorem III.2.1.

Introduce the error term

$$E_n(y) = \frac{\hat{\mu}_n - \mu}{\sigma} + y\frac{\hat{\sigma}_n - \sigma}{\sigma} \qquad (y \in \mathbb{R}).$$

In the proof of part i) and ii) it is sufficient to restrict attention to the special choice $\theta_0 = (0,1)'$ of θ because the distribution of T_n is invariant with respect to θ.

i) Take $x_1 \neq x_2 \in S$. Observe that C.III.2 implies

$$P_{1n}(\|\hat{\theta}_n - \theta_0\| > (\max\{(x_1+x_2)^2,(x_1-x_2)^2\}+4)^{\frac{1}{2}}\xi_n/|x_1-x_2|) \to 1$$

for a sequence $\{\xi_n\}$ such that $n^{\frac{1}{2}}\xi_n \to \infty$ and $\xi_n \to 0$. Calculation of the intersection points of the lines $E_n(x_1) = \xi_n$, $E_n(x_2) = \xi_n$, $-E_n(x_1) = \xi_n$ and $-E_n(x_2) = \xi_n$ in the $(\hat{\mu}_n, \hat{\sigma}_n - 1)$-plane shows

$$\{\max\{|E_n(x_1)|,|E_n(x_2)|\} > \xi_n\}$$

$$\supset \left\{|\hat{\mu}_n| > \xi_n \max\left\{\frac{|x_1+x_2|}{|x_1-x_2|},1\right\}\right\} \cup \left\{|\hat{\sigma}_n - 1| > 2\xi_n/|x_1-x_2|\right\}$$

$$\supset \{\|\hat{\theta}_n - \theta_0\| > (\max\{(x_1+x_2)^2,(x_1-x_2)^2\}+4)^{\frac{1}{2}}\xi_n/|x_1-x_2|\}.$$

Hence

$$\text{(III.5.1)} \qquad P_{1n}(\max\{|E_n(x_1)|,|E_n(x_2)|\} > \xi_n) \to 1.$$

(We consider two different arguments x_1 and x_2 because $P_{1n}(|E_n(x_1)| > \xi_n)$ does not necessarily tend to one as $n \to \infty$.) Put $b_n = \frac{1}{2}cn^{\frac{1}{2}}\xi_n$ and note that $b_n \to \infty$ as $n \to \infty$. Observe that

$$\begin{aligned}
&\{n^{\frac{1}{2}}|\hat{F}_n(x_1) - G_0(x_1)| > b_n\} \\
&= \{n^{\frac{1}{2}}|F_n(\hat{\mu}_n + x_1\hat{\sigma}_n) - G_0(\hat{\mu}_n + x_1\hat{\sigma}_n) + G_0(x_1 + E_n(x_1)) - G_0(x_1)| > b_n\} \\
&\supset \{n^{\frac{1}{2}} \sup_{y\in\mathbb{R}} |F_n(y) - G_0(y)| < \tfrac{1}{2}b_n\} \cap \{|E_n(x_1)| > \xi_n\} \cap \\
&\qquad\qquad \cap \{n^{\frac{1}{2}}|G_0(x_1 + E_n(x_1)) - G_0(x_1)| > \tfrac{3}{2}b_n\} \\
&= \{n^{\frac{1}{2}} \sup_{y\in\mathbb{R}} |F_n(y) - G_0(y)| < \tfrac{1}{2}b_n\} \cap \{|E_n(x_1)| > \xi_n\}
\end{aligned}$$

for n sufficiently large. A similar relation holds true for x_2. Hence, using the weak convergence of $n^{\frac{1}{2}}(F_n(\cdot) - G_0(\cdot))$ under H_{1n} (cf. Shorack (1979)),

$$\text{(III.5.2)} \qquad P_{1n}(n^{\frac{1}{2}} \max\{|\hat{F}_n(x_1) - G_0(x_1)|,|\hat{F}_n(x_2) - G_0(x_2)|\} \ge b_n) \to 1.$$

In a similar way one proves for sufficiently small ε

$$\text{(III.5.3)} \qquad \begin{aligned} P_{1n}(n^{\frac{1}{2}} \inf_{\|(\varepsilon_1,\varepsilon_2)\|<\varepsilon} \max\{&|\hat{F}_n(x_1+\varepsilon_1) - G_0(x_1+\varepsilon_1)|, \\ &|\hat{F}_n(x_2+\varepsilon_2) - G_0(x_2+\varepsilon_2)|\} \ge b_n) \to 1. \end{aligned}$$

This immediately implies that the test statistics (III.1.1) - (III.1.4) tend to infinity in probability under H_{1n}. Noting that the critical values of the tests are bounded above, the proof is complete.

ii) Grouping of observations decreases the Cressie-Read statistics. Pool the observations twice into two classes, once with cells $(-\infty,a_{ki}],(a_{ki},\infty)$ and once with cells $(-\infty,a_{kj}],(a_{kj},\infty)$ (for notational simplicity take $i = 1$ and $j = k-1$). Then

$$\begin{aligned} CR_n(\lambda) \ge \frac{2n}{\lambda(\lambda+1)} \max\Bigg\{ &\frac{N_{k1}^{\lambda+1}(\hat{\theta}_n)}{n^{\lambda+1}p_{k1}^{\lambda}(0)} + \frac{(n-N_{k1}(\hat{\theta}_n))^{\lambda+1}}{n^{\lambda+1}(1-p_{k1}(0))^{\lambda}} - 1, \\ &\frac{N_{kk}^{\lambda+1}(\hat{\theta}_n)}{n^{\lambda+1}p_{kk}^{\lambda}(0)} + \frac{(n-N_{kk}(\hat{\theta}_n))^{\lambda+1}}{n^{\lambda+1}(1-p_{kk}(0))^{\lambda}} -1 \Bigg\}. \end{aligned}$$

Taylor expansion of the last expression and bounding its second derivative with respect to $N_{ki}(\hat{\theta}_n)/n$ ($i = 1$ or k) yields the further inequality

$$CR_n(\lambda) \geq n^{-1} \max\{(N_{k1}(\hat{\theta}_n) - np_{k1}(0))^2, (N_{kk}(\hat{\theta}_n) - np_{kk}(0))^2\}$$

$$= n \max\{|\hat{F}_n(a_{k1}) - G_0(a_{k1})|^2, |\hat{F}_n(a_{kk-1}) - G_0(a_{kk-1})|^2\}.$$

Let $\mathcal{B}(a_{k1})$ and $\mathcal{B}(a_{kk-1})$ be small open balls with centers a_{k1} and a_{kk-1}. Let $K \subset \mathcal{B}(a_{k1}) \cup \mathcal{B}(a_{kk-1})$ be a compact set such that a_{k1} and a_{kk-1} are interior points of K. Choose $c = \frac{1}{2} \min_{y \in K} g_0(y)$ and put $S = \{y \in \mathcal{B}(a_{k1}) \cup \mathcal{B}(a_{kk-1});$ $g_0(y) > c\}$, then application of (III.5.2) with $x_1 = a_{k1}$ and $x_2 = a_{kk-1}$ yields $CR_n(\lambda) \to_{P_{1n}} \infty$. Finally

$$MS_n \geq \gamma_1 V_k(\hat{\theta}_n)' V_k(\hat{\theta}_n) = \gamma_1 CR_n(1),$$

where $\gamma_1 > 0$ is the smallest eigenvalue of Γ_k, completes the proof of part ii).

iii) As in part ii) it is sufficient to show that

$$n^{-1} \max\{(\overline{N}_{k1} - np_{k1}(0,\theta_0,\hat{\theta}_n))^2, (\overline{N}_{kk} - np_{kk}(0,\theta_0,\hat{\theta}_n))^2\} \to_{P_{1n}} \infty.$$

The distribution of $\tilde{T}_n$ is not invariant with respect to θ and hence the proof has to be given for arbitrary θ. Fix θ. Replacing a_{k1} and a_{kk-1} by $(a_{k1} - \mu)/\sigma$ and $(a_{kk-1} - \mu)/\sigma$, construct $c = c(\theta)$ and $S = S(\theta)$ as in part ii). Choosing $\xi_n = \xi_n(\theta)$ as in part i) implies (cf. (III.5.1))

$$P_{1n}\left(\max\left\{\left|E_n\left(\frac{a_{k1} - \mu}{\sigma}\right)\right|, \left|E_n\left(\frac{a_{kk-1} - \mu}{\sigma}\right)\right|\right\} > \xi_n\right) \to 1.$$

Define

$$\overline{E}_n(y) = \frac{y - \hat{\mu}_n}{\hat{\sigma}_n} - \frac{y - \mu}{\sigma} = -\sigma E_n\left(\frac{y-\mu}{\sigma}\right)/\hat{\sigma}_n \qquad (y \in \mathbb{R})$$

and let $\{\Delta_n\}$ be a sequence such that $\Delta_n \to \infty$ and $n^{\frac{1}{2}}\xi_n/\Delta_n \to \infty$ as $n \to \infty$. Then

$$\begin{aligned}
&P_{1n}(\max\{|\overline{E}_n(a_{k1})|, |\overline{E}_n(a_{kk-1})|\} > \xi_n/\Delta_n) \\
&= P_{1n}\left(\sigma \max\left\{\left|E_n\left(\frac{a_{k1}-\mu}{\sigma}\right)\right|, \left|E_n\left(\frac{a_{kk-1}-\mu}{\sigma}\right)\right|\right\}\Big/\hat{\sigma}_n > \xi_n/\Delta_n; \hat{\sigma}_n/\sigma \leq \Delta_n\right) \\
&+ P_{1n}\left(\max\left\{|\overline{E}_n(a_{k1})|, \left|\overline{E}_n(a_{k1}) - \frac{a_{kk-1}-a_{k1}}{\sigma}\left(1 - \frac{\sigma}{\hat{\sigma}_n}\right)\right|\right\} > \xi_n/\Delta_n;\right. \\
&\qquad\qquad \left.\hat{\sigma}_n/\sigma > \Delta_n\right) \\
&\to 1.
\end{aligned}$$

Put $b_n = b_n(\theta) = \frac{1}{2}cn^{\frac{1}{2}}\xi_n/\Delta_n$ ($b_n \to \infty$ as $n \to \infty$) and note that

$$\{n^{\frac{1}{2}}|F_n(a_{k1}) - G_0^*(a_{k1};\hat{\theta}_n)| > b_n\}$$
$$= \left\{n^{\frac{1}{2}}\left|F_n(a_{k1}) - G_0\left(\frac{a_{k1}-\mu}{\sigma}\right) - \left(G_0\left(\frac{a_{k1}-\mu}{\sigma} + \bar{E}_n(a_{k1})\right) - G_0\left(\frac{a_{k1}-\mu}{\sigma}\right)\right)\right| > b_n\right\}$$
$$\supset \left\{n^{\frac{1}{2}} \sup_{y\in\mathbb{R}}\left|F_n(y) - G_0\left(\frac{y-\mu}{\sigma}\right)\right| < \tfrac{1}{2}b_n\right\} \cap \left\{|\bar{E}_n(a_{k1})| > \xi_n/\Delta_n\right\}$$
$$\cap \left\{n^{\frac{1}{2}}\left|G_0\left(\frac{a_{k1}-\mu}{\sigma} + \bar{E}_n(a_{k1})\right) - G_0\left(\frac{a_{k1}-\mu}{\sigma}\right)\right| > \tfrac{3}{2}b_n\right\}$$
$$= \left\{n^{\frac{1}{2}} \sup_{y\in\mathbb{R}}\left|F_n(y) - G_0\left(\frac{y-\mu}{\sigma}\right)\right| < \tfrac{1}{2}b_n\right\} \cap \left\{|\bar{E}_n(a_{k1})| > \xi_n/\Delta_n\right\}$$

for sufficiently large n. Proceeding as in part i) yields the desired result. □

<u>REMARK III.5.1</u>. From the proof of Theorem III.2.1 one observes that this theorem applies also to other EDF tests. It is sufficient to require that (III.5.2) or (III.5.3) implies $\mathcal{T}_n \to_{P_{1n}} \infty$. A similar result holds true for $\tilde{\mathcal{T}}_n$. □

III.5.2. Proof of Corollary III.2.2.

Similarly to the proof of Theorem III.2.1 ii),iii) one may construct c (c(θ)) and S (S(θ)) since for each k there exist boundary points $a_{ki} \neq a_{kj}$ such that $(\forall k)\ \max\{|a_{ki}|,|a_{kj}|\} < M < \infty$. Then proceeding as in the first part of the proof of Theorem III.2.1, choose ξ_n ($\xi_n(\theta)$) such that $\xi_n \to 0$, $n^{\frac{1}{2}}\xi_n/k^{\frac{1}{2}} \to \infty$ and (III.5.1) holds true. Put $b_n^* = b_n/k^{\frac{1}{2}}$, then $b_n^* \to \infty$ and for each of the tests (III.1.5) - (III.1.8)

$$P_{1n}(\mathcal{T}_n \geq (b_n^*)^2k) \to 1 \quad \text{and}$$
$$P_{1n}(\tilde{\mathcal{T}}_n \geq (b_n^*)^2k) \to 1.$$

The proof of the corollary is complete since the critical values of the tests are of order k. □

III.5.3. Proof of Proposition III.2.3.

It is sufficient to restrict the proof to $\theta = \theta_0 = (0,1)'$. Write

$$U_{ni}(\hat{\theta}_n) = n^{-\frac{1}{2}}\sum_{j=1}^{n}\left\{G_0^i\left(\frac{Y_j-\hat{\mu}_n}{\hat{\sigma}_n}\right) - G_0^i(Y_j)\right\} + n^{-\frac{1}{2}}\sum_{j=1}^{n}\{G_0^i(Y_j) - (i+1)^{-1}\}.$$

Application of the mean value theorem and the central limit theorem yields $U_{ni}(\hat{\theta}_n) = O_p(1)$ $(i=1,\dots,k)$ and hence $N_n = O_p(1)$ under H_0.

To show that $N_n \to_{P_{1n}} \infty$ define the random variables

$$\xi_n = \begin{cases} \min\{|\hat{\mu}_n|, |\hat{\sigma}_n - 1|\} & \text{if } \min\{|\hat{\mu}_n|, |\hat{\sigma}_n - 1|\} > n^{-\frac{1}{2}} \log n \\ \max\{|\hat{\mu}_n|, |\hat{\sigma}_n - 1|\} & \text{if } \max\{|\hat{\mu}_n|, |\hat{\sigma}_n - 1|\} < n^{-\frac{1}{2}} \log n \\ n^{-\frac{1}{2}} \log n & \text{elsewhere} \end{cases}$$

Note that by C.III.2 and (III.2.2) $\xi_n \to_{P_{n1}} 0$ and $n^{\frac{1}{2}}\xi_n \to_{P_{1n}} \infty$. Let $S_n(\varepsilon)$ be the event such that

$$|n^{-\frac{1}{2}} \sum_{j=1}^{n} \{G_0^i(Y_j) - (i+1)^{-1}\}| \le n^{\frac{1}{2}}\xi_n \varepsilon \qquad (i=1,2)$$

$$|n^{-1} \sum_{j=1}^{n} g_0(Y_j) - E_0\{g_0(Y)\}| \le \varepsilon$$

$$|n^{-1} \sum_{j=1}^{n} Y_j g_0(Y_j)| \le \varepsilon$$

$$|2n^{-1} \sum_{j=1}^{n} g_0(Y_j)(G_0(Y_j) - \tfrac{1}{2})| \le \varepsilon$$

$$|2n^{-1} \sum_{j=1}^{n} Y_j g_0(Y_j)(G_0(Y_j) - \tfrac{1}{2}) - 2E_0\{Yg_0(Y)(G_0(Y) - \tfrac{1}{2})\}| \le \varepsilon$$

$$\xi_n \le \max\left\{\tfrac{1}{2}, \varepsilon/\left[8 \max_{y\in\mathbb{R}} \lambda\left(\nabla_\theta, \nabla_\theta G_0\left(\frac{y-\mu}{\sigma}\right)\Big|_{\theta=\theta_0}\right)\right],\right.$$

$$\left.\varepsilon/\left[8 \max_{y\in\mathbb{R}} \lambda\left(\nabla_\theta, \nabla_\theta\left(G_0\left(\frac{y-\mu}{\sigma}\right) - \tfrac{1}{2}\right)^2\Big|_{\theta=\theta_0}\right)\right]\right\},$$

where $\lambda(A)$ denotes the maximum eigenvalue of A. Note that for all $\varepsilon > 0$ $P_{1n}(S_n(\varepsilon)) \to 1$ as $n \to \infty$. Put $\varepsilon = \frac{1}{5}\min\{E_0\{g_0(Y)\}, E_0\{Yg_0(Y)(G_0(Y) - \frac{1}{2})\}\} > 0$ and let $\gamma_1 > 0$ be the smallest eigenvalue of Γ_k. Then $N_n \ge \gamma_1 \|U_n(\hat{\theta}_n)\|^2$ implies

$$P_{1n}(N_n \ge \gamma_1 \varepsilon^2 n \xi_n^2)$$

$$\text{(III.5.4)} \qquad \ge P_{1n}(U_{n1}^2(\hat{\theta}_n) \ge \varepsilon^2 n\xi_n^2, |\hat{\sigma}_n - 1| \le |\hat{\mu}_n|, S_n(\varepsilon)) +$$

$$+ P_{1n}((U_{n1}(\hat{\theta}_n) - U_{n2}(\hat{\theta}_n))^2 \ge 2\varepsilon^2 n\xi_n^2, |\hat{\mu}_n| < |\hat{\sigma}_n - 1|, S_n(\varepsilon)).$$

Suppose $(Y_1,\dots,Y_n) \in S_n(\varepsilon)$ and $|\hat{\sigma}_n - 1| \le \hat{\mu}_n$, then $|\hat{\sigma}_n - 1| \le \xi_n \le \hat{\mu}_n$ and the error in the Taylor expansion given below is uniformly less than $\xi_n\varepsilon$. It follows under H_{1n}

$$
\begin{aligned}
U_{n1}(\hat{\theta}_n) &\le n^{-\frac{1}{2}} \sum_{j=1}^{n} \{G_0\left(\frac{Y_j-\hat{\mu}_n}{\hat{\sigma}_n}\right) - G_0(Y_j)\} + n^{\frac{1}{2}}\xi_n \varepsilon \\
&\le n^{-\frac{1}{2}} \sum_{j=1}^{n} \{G_0\left(\frac{Y_j-\xi_n}{\hat{\sigma}_n}\right) - G_0(Y_j)\} + n^{\frac{1}{2}}\xi_n \varepsilon \\
&\le -n^{\frac{1}{2}}\xi_n n^{-1} \sum_{j=1}^{n} g_0(Y_j) - n^{\frac{1}{2}}(\hat{\sigma}_n - 1) n^{-1} \sum_{j=1}^{n} Y_j g_0(Y_j) + 2n^{\frac{1}{2}}\xi_n \varepsilon \\
&\le -n^{\frac{1}{2}}\xi_n (E_0\{g_0(Y)\} - 4\varepsilon) \le -n^{\frac{1}{2}}\xi_n \varepsilon.
\end{aligned}
$$

Similarly $(Y_1,\dots,Y_n) \in S_n(\varepsilon)$ and $|\hat{\sigma}_n - 1| \le -\hat{\mu}_n$ implies

$$
U_{n1}(\hat{\theta}_n) \ge n^{\frac{1}{2}}\xi_n \varepsilon
$$

and hence the first probability in the RHS of (III.5.4) is equal to $P_{1n}(|\hat{\sigma}_n - 1| \le |\hat{\mu}_n|, S_n(\varepsilon))$. Next suppose $(Y_1,\dots,Y_n) \in S_n(\varepsilon)$ and $|\hat{\mu}_n| < \hat{\sigma}_n - 1$, then $|\hat{\mu}_n| \le \xi_n \le \hat{\sigma}_n - 1$ and the error in the Taylor expansion given below is uniformly less than $\xi_n \varepsilon$. In this case

$$
\begin{aligned}
U_{n2}(\hat{\theta}_n) - U_{n1}(\hat{\theta}_n) &\le n^{-\frac{1}{2}} \sum_{j=1}^{n} \{(G_0\left(\frac{Y_j-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \tfrac{1}{2})^2 - (G_0(Y_j) - \tfrac{1}{2})^2\} + \\
&\qquad + 2n^{\frac{1}{2}}\xi_n \varepsilon \\
&\le n^{-\frac{1}{2}} \sum_{j=1}^{n} \{(G_0\left(\frac{Y_j-\hat{\mu}_n}{1+\xi_n}\right) - \tfrac{1}{2})^2 - (G_0(Y_j) - \tfrac{1}{2})^2\} + 2n^{\frac{1}{2}}\xi_n \varepsilon \\
&\le -n^{\frac{1}{2}}\hat{\mu}_n n^{-1}\, 2 \sum_{j=1}^{n} g_0(Y_j)(G_0(Y_j) - \tfrac{1}{2}) + \\
&\qquad -n^{\frac{1}{2}}\xi_n n^{-1}\, 2 \sum_{j=1}^{n} Y_j g_0(Y_j)(G_0(Y_j) - \tfrac{1}{2}) + 3n^{\frac{1}{2}}\xi_n \varepsilon \\
&\le -n^{\frac{1}{2}}\xi_n (2E_0\{Y g_0(y)(G_0(Y) - \tfrac{1}{2})\} - 5\varepsilon) \le -n^{\frac{1}{2}}\xi_n 5\varepsilon.
\end{aligned}
$$

Similarly $(Y_1,\dots,Y_n) \in S_n(\varepsilon)$ and $|\hat{\mu}_n| < 1 - \hat{\sigma}_n$ implies

$$
U_{n2}(\hat{\theta}_n) - U_{n1}(\hat{\theta}_n) \ge n^{\frac{1}{2}}\xi_n 5\varepsilon
$$

and hence the second probability in the RHS of (III.5.4) is equal to $P_{1n}(|\hat{\mu}_n| < |\hat{\sigma}_n - 1|, S_n(\varepsilon))$. Thus

$$
P_{1n}(N_n \ge \gamma_1 \varepsilon^2 n \xi_n^2) \ge P(S_n(\varepsilon)) \to 1
$$

and the proof is complete. $\square$

CHAPTER IV

POWER APPROXIMATIONS TO MULTINOMIAL TESTS OF FIT

IV.1. INTRODUCTION

In this chapter we return to the simple hypothesis

$$H_0 : F^Y = G_0$$

for some specified distribution function G_0 and consider the class of Cressie-Read statistics $\overline{CR}_n(\lambda)$. For this class the null hypothesis reduces to the following multinomial testing problem

$$\text{(IV.1.1)} \qquad H_0 : p^Y = p,$$

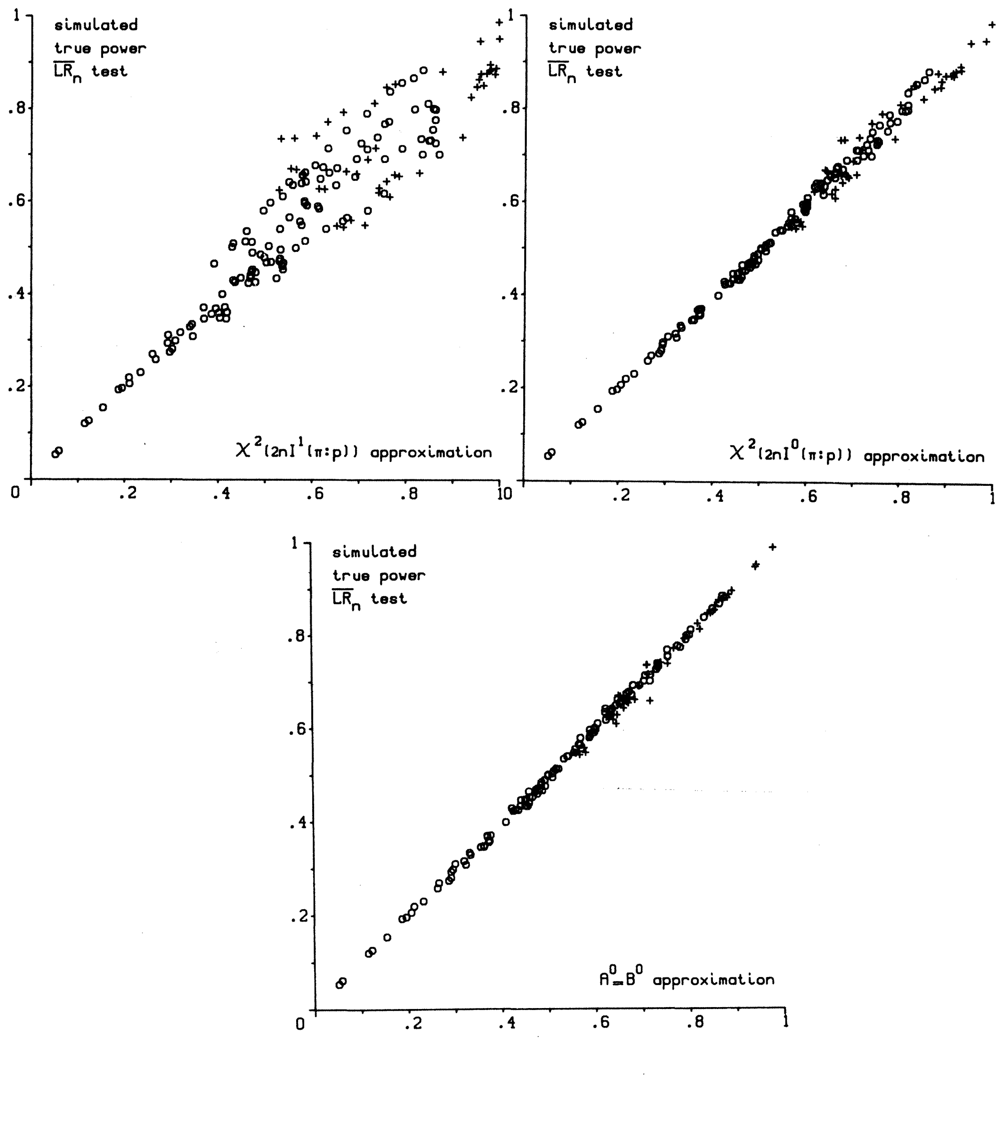
where $p = (p_1,\ldots,p_k)$ is the vector of cell-probabilities under H_0. Since classical asymptotics for multinomial goodness-of-fit tests fail to adequately describe the finite-sample truth, we present instead a new large-sample approximation to the distribution of the statistics $\overline{CR}_n(\lambda)$. The new approximation follows the structure of the statistics $\overline{CR}_n(\lambda)$ more closely, avoiding local expansion of the alternative cell probabilities. Hence the approximation is valid for an almost unrestricted range of alternative distributions G of the observations $Y_1,\ldots,Y_n$ and a fixed number k of cells. This method is computationally feasible, gives excellent agreement with exact computations and Monte Carlo results, and enables us to extend the qualitative insights obtained in earlier work.

The classical approximation to the limiting alternative distribution of $\overline{CR}_n(\lambda)$ for k fixed is based on sequences of local alternatives G_n with cell probabilities π_{in} such that

$$\text{(IV.1.2)} \qquad \delta = \lim_{n\to\infty} n \sum_{i=1}^{k} (\pi_{in} - p_i)^2/p_i$$

exists and is finite. Then the limiting distribution of any $\overline{CR}_n(\lambda)$ under G_n is noncentral chi-square, $\chi^2_{k-1}(\delta)$. The distribution $\chi^2_{k-1}(\delta)$ gives an

Figure IV.1.1. Three approximations to the power of the $\overline{LR}_n$ test, compared to simulated true powers for n=100 and level α=.05. For the A^0 approximation see (IV.2.8). The testing problems and data are taken from Kallenberg et al. (1985), Fig 2. Equal (0) and unequal (+) null probabilities are considered.



adequate approximation to the power of $\overline{P}_n$ ($\lambda = 1$) against an alternative G_n. The approximation is less good for $\overline{LR}_n$ (λ=0), as Figure 2 of Kallenberg et al. (1985) demonstrates, and further calculations show it to be very poor for more extreme values of λ.

The classical limiting distributions under the null hypothesis and sequences of local alternatives follow from a Taylor series expansion of the statistic $\overline{CR}_n(\lambda)$ to second order terms. All $\overline{CR}_n(\lambda)$ are asymptotically equivalent. To illustrate the insensitivity of this approach, note that the same Taylor series expansion shows that one can also write the noncentrality parameter as

(IV.1.3) $$\delta = \lim_{n\to\infty} 2nI^{\lambda}(\pi_n : p)$$

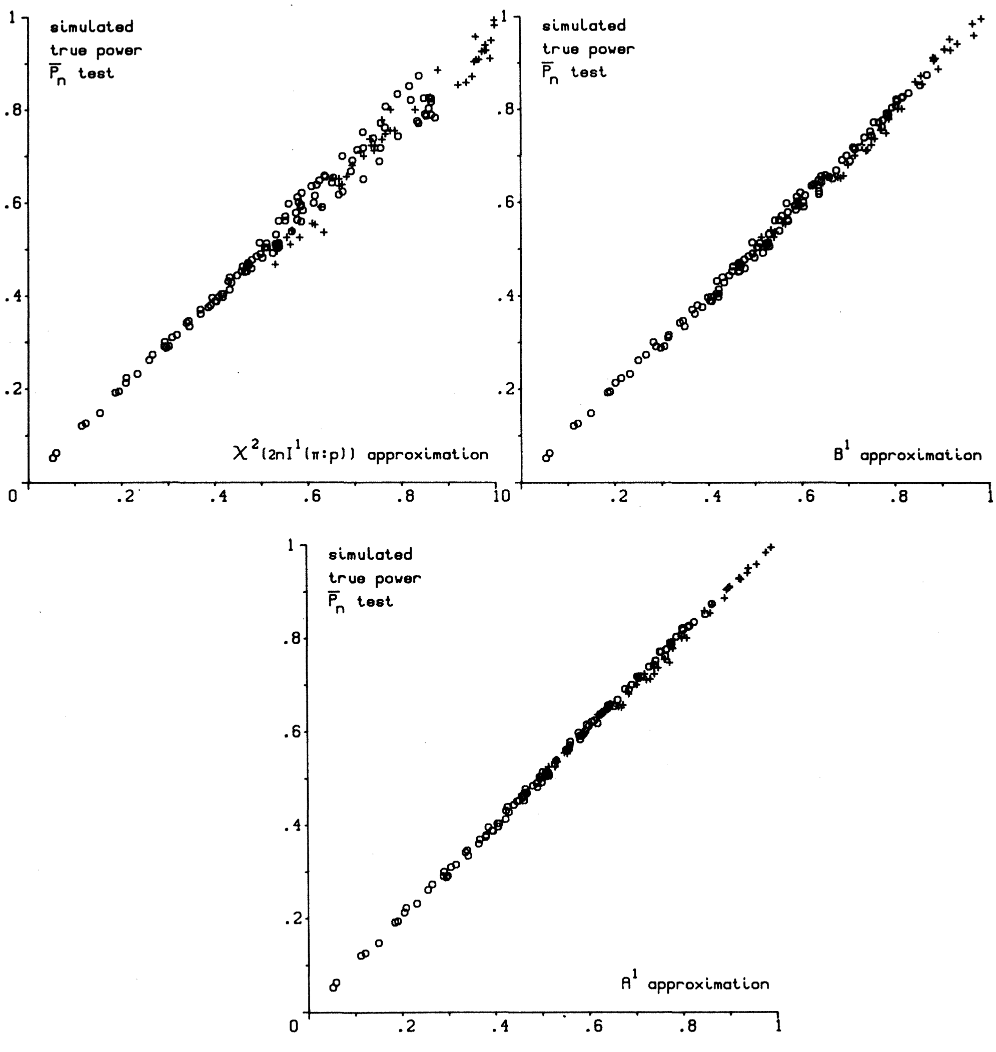
for arbitrary λ, where

$$I^{\lambda}(q:p) = \frac{1}{\lambda(\lambda+1)} \sum_{i=1}^{k} p_i\{(q_i/p_i)^{\lambda+1} - 1\}$$

is a directed divergence between the discrete distributions p and q on k points. A discussion of the role of $2I^{\lambda}(\pi:p)$ as a measure of lack of fit is given in Moore (1984).
For finite n, this alternative choice of δ with $\lambda = 0$ does in fact improve the numerical accuracy of the noncentral chi-square approximation to the power of $\overline{LR}_n$, as Figure IV.1.1 shows. But for λ far from 1, the chi-square approximation to the power is very poor, just as is the case for the chi-square approximation to the null distribution.

We propose the use of a large-sample approximation A^{λ} based on a Taylor series expansion of $\overline{CR}_n(\lambda)$ under an arbitrary sequence of alternatives G_n. The expansion is essentially the same as in the classical approximation, but does not rely on the local character of G_n. It reduces to the usual null hypothesis theory when $G_1 = G_0$. But the first order terms are not negligible under fixed $G_1 \neq G_0$. Hence the approximating distributions are more complex, generally linear combinations of noncentral chi-squares plus a constant. The approximating distribution is simplest when $\lambda = 0$, taking the form $\chi^2_{k-1}(\delta^0) + \xi^0$, where the constant ξ^0 and noncentrality parameter δ^0 are functions of the π_i and p_i given by (IV.2.9). Figure IV.1.1 shows that this approximation to the distribution of $\overline{LR}_n$ is superior to the classical result based on (IV.1.2) or the alternative expression

Figure IV.1.2. Three approximations to the power of the $\overline{P}_n$ test, compared to simulated true powers for n=100 and level $\alpha = .05$. For the A^1 approximation see (IV.2.7), for the B^1 approximation see (IV.2.10). Same data as in Figure IV.1.1.



(IV.1.3). Our approximations are easily implemented via suitable software, and are analytically much simpler than such competing methods as Edgeworth-type expansions of the distribution functions.

A second approach is to look for the "locally best" noncentral chi-square approximation. In the series expansion of $\overline{CR}_n(\lambda)$ only the coefficients of the quadratic term are expanded locally to obtain an approximation B^λ which is still as close as possible to $\overline{CR}_n(\lambda)$ but with the pleasing feature that it has a simple distribution of the form $\bar{r}_\lambda \chi^2_{k-1}(\delta^{(\lambda)}) + \xi^{(\lambda)}$, which is given in (IV.2.10). For $\overline{LR}_n$ this approximation coincides with the A^λ approximation described above. In Figure IV.1.2 both approximations are compared with the classical $\chi^2_{k-1}(\delta)$ approximation for $\overline{P}_n$. Although the locally best noncentral χ^2 approximation cannot be expected to do as well as the A^λ approximation, it performs a lot better than the $\chi^2_{k-1}(\delta)$ approximations with δ from (IV.1.2) and (IV.1.3).

Similar expansions of $\overline{CR}_n(\lambda)$ under fixed G_1 and k are easily obtained for the case of testing fit to a parametric family $\{G_0^*(\cdot;\theta);\theta \in \Theta \subset \mathbb{R}^p\}$, where the parameter θ is estimated by $\hat{\theta}_n(Y_1,\ldots,Y_n)$. Yet we do not know whether a result like Theorem IV.2.1 holds true for these cases too. The distribution theory of the resulting expressions is complex and needs second order expansions of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$. We confine the present investigation to the case of completely specified G_0.

IV.2. APPROXIMATIONS

IV.2.1 Expansions and approximations.

For an alternative G_n let $\pi_n = (\pi_{1n},\ldots,\pi_{kn})$ be the vector of cell probabilities. The π_{in} are assumed to be bounded away from zero. The distribution of the vector $\bar{N}_k$ of cell counts under G_n will often be denoted by P_{π_n}; E_{π_n} and var_{π_n} have a similar interpretation. Put

$$\text{(IV.2.1)} \qquad \begin{aligned} Y_{in} &= (\bar{N}_{ki} - n\pi_{in})/(n\pi_{in})^{\frac{1}{2}}, \\ r_{in} &= \pi_{in}/p_i \qquad (i=1,\ldots,k) \end{aligned}$$

and consider the following Taylor expansion of $\overline{CR}_n(\lambda)$ under G_n, for $\lambda \neq -1,0$,

$$\overline{CR}_n(\lambda) = 2nI^{\lambda}(\bar{N}_k/n:p)$$

$$= \frac{2}{\lambda(1+\lambda)}\left[\sum_{i=1}^{k} r_{in}^{\lambda} n\pi_{in}\{1+(n\pi_{in})^{-\frac{1}{2}}Y_{in}\}^{\lambda+1} - n\right]$$

$$(IV.2.2) \qquad = \frac{2}{\lambda(1+\lambda)}\left[\sum_{i=1}^{k} r_{in}^{\lambda} n\pi_{in}\{1+(\lambda+1)(n\pi_{in})^{-\frac{1}{2}}Y_{in} + \tfrac{1}{2}(\lambda+1)\lambda(n\pi_{in})^{-1}Y_{in}^2 + \right.$$

$$\left. + \tfrac{1}{6}(\lambda+1)\lambda(\lambda-1)(n\pi_{in})^{-\frac{3}{2}}Y_{in}^3 + \ldots\} - n\right]$$

$$= A^{\lambda}(Y_n) + O_p(n^{-\frac{1}{2}}),$$

where

$$A^{\lambda}(Y_n) = \sum_{i=1}^{k} r_{in}^{\lambda}\{Y_{in} + \lambda^{-1}(n\pi_{in})^{\frac{1}{2}}\}^2 + 2nI^{\lambda}(\pi_n:p) - n\lambda^{-2}\sum_{i=1}^{k}\pi_{in}r_{in}^{\lambda}$$

$$(IV.2.3) \qquad = \sum_{i=1}^{k} r_{in}^{\lambda}\{Y_{in} + \lambda^{-1}(n\pi_{in})^{\frac{1}{2}}(1-r_{in}^{-\lambda})\}^2 + 2nI^{\lambda}(\pi_n:p) +$$

$$- n\lambda^{-2}\sum_{i=1}^{k}\pi_{in}(1-r_{in}^{-\lambda})^2 r_{in}^{\lambda}$$

$$= B^{\lambda}(Y_n) + \sum_{i=1}^{k}(r_{in}^{\lambda} - \bar{r}_{n\lambda})Y_{in}^2$$

with $\bar{r}_{n\lambda} = \sum_{i=1}^{k}(1-\pi_{in})r_{in}^{\lambda}/(k-1)$ and

$$(IV.2.4) \qquad B^{\lambda}(Y_n) = \bar{r}_{n\lambda}\sum_{i=1}^{k}\{Y_{in} + (\lambda\bar{r}_{n\lambda})^{-1}(n\pi_{in})^{\frac{1}{2}}(r_{in}^{\lambda}-1)\}^2 + 2nI^{\lambda}(\pi_n:p) +$$

$$-n\lambda^{-2}\bar{r}_{n\lambda}^{-1}\sum_{i=1}^{k}\pi_{in}(r_{in}^{\lambda}-1)^2.$$

The linear part of $\overline{CR}_n(\lambda)$ (and A^{λ}) is given by

$$(IV.2.5) \qquad L^{\lambda}(Y_n) = 2n^{\frac{1}{2}}\lambda^{-1}\sum_{i=1}^{k}\pi_{in}^{\frac{1}{2}}(r_{in}^{\lambda}-1)Y_{in} + 2nI^{\lambda}(\pi_n:p).$$

The second expression for A^{λ} is useful because with this form the expansion (IV.2.2) extends to $\lambda = 0$ or -1 by taking appropriate limits in both $\overline{CR}_n(\lambda)$ and A^{λ}. Note that $A^{\lambda} = \overline{CR}_n(\lambda)$ if $\lambda = 1$ and $A^{\lambda} = B^{\lambda}$ if $\lambda = 0$. The quadratic B^{λ} is introduced because its asymptotic distribution is simple due to equal coefficients of the quadratic terms; among such forms it is closest to A^{λ} (and thus to $\overline{CR}_n(\lambda)$) in the sense that $E_{\pi_n}\{A^{\lambda}(Y_n) - B^{\lambda}(Y_n)\} = 0$.

Inserting $\bar{r}_{n\lambda} = 1$ in (IV.2.4) we get a slightly simpler form of B^{λ}, denoted by B_1^{λ}, which is sometimes useful too.

The leading parts $A^{\lambda}(Y_n)$, $B^{\lambda}(Y_n)$ (or $B_1^{\lambda}(Y_n)$) and $L^{\lambda}(Y_n)$ of the expansion of $\overline{CR}_n(\lambda)$ can be used to construct approximations to the distribution of $\overline{CR}_n(\lambda)$. Let U_n have a multivariate normal distribution

$$U_n = (U_{1n},\dots,U_{kn})' \sim N_k(0, I_k - \pi_n^{\frac{1}{2}}\pi_n^{\frac{1}{2}\,'})$$

where $\pi_n^{\frac{1}{2}} = (\pi_{1n}^{\frac{1}{2}},\dots,\pi_{kn}^{\frac{1}{2}})'$. Since Y_n is asymptotically distributed as U_n (under G_n), replace Y_n by U_n in A^λ, B^λ and L^λ and consider the approximations $A^\lambda(U_n)$, $B^\lambda(U_n)$ and $L^\lambda(U_n)$. The construction suggest that $A^\lambda(U_n)$ is the best and $L^\lambda(U_n)$ the least reliable approximation.

To derive their distributions, we employ the following notation. Let Q_n be the diagonal matrix

$$Q_n = \begin{pmatrix} r_{1n}^{\frac{1}{2}\lambda} & & \\ & \ddots & \\ & & r_{kn}^{\frac{1}{2}\lambda} \end{pmatrix},$$

and $\mu_n = (\mu_{1n},\dots,\mu_{kn})'$ the vector

$$\mu_n = n^{\frac{1}{2}\lambda - 1}(\pi_{1n}^{\frac{1}{2}}(1 - r_{1n}^{-\lambda}),\dots,\pi_{kn}^{\frac{1}{2}}(1 - r_{kn}^{-\lambda}))'.$$

Let $\theta_{1n},\dots,\theta_{kn}$ be the eigenvalues and S_n the $(k \times k)$ orthonormal matrix of eigenvectors of the matrix $Q_n(I_k - \pi_n^{\frac{1}{2}}\pi_n^{\frac{1}{2}\,'})Q_n$:

$$\text{(IV.2.6)} \qquad Q_n(I_k - \pi_n^{\frac{1}{2}}\pi_n^{\frac{1}{2}\,'})Q_n S_n = S_n \begin{pmatrix} \theta_{1n} & & \\ & \ddots & \\ & & \theta_{kn} \end{pmatrix}.$$

Let $T_n = \mathrm{diag}(\theta_{1n},\dots,\theta_{kn})$ and write

$$\omega_n = (\omega_{1n},\dots,\omega_{kn})' = S_n' Q_n \mu_n.$$

$Z_1, Z_2, \dots$ are independent standard normal random variables.

Apply the orthogonal transformation S_n' to $A^\lambda(U_n)$. Then

$$\begin{aligned} A^\lambda(U_n) &= \|Q_n(U_n + \mu_n)\|^2 + 2nI^\lambda(\pi_n : p) - \|Q_n\mu_n\|^2 \\ &= \|S_n'Q_n(U_n + \mu_n)\|^2 + 2nI^\lambda(\pi_n : p) - \|S_n'Q_n\mu_n\|^2 \\ &\sim \sum_{\theta_{in}\neq 0} \theta_{in}(Z_i + \omega_{in}/\theta_{in}^{\frac{1}{2}})^2 + \sum_{\theta_{in}=0} \omega_{in}^2 + 2nI^\lambda(\pi_n : p) - \sum_{i=1}^{k} \omega_{in}^2 \end{aligned}$$

since $S_n'Q_n(U_n + \mu_n) \sim N_k(\omega_n, S_n'Q_n(I_k - \pi_n^{\frac{1}{2}}\pi_n^{\frac{1}{2}\,'})Q_nS_n) = N_k(\omega_n, T_n)$. As one of the θ_{in} vanishes, assume $\theta_{kn} = 0$. It follows that $A^\lambda(U_n)$ is distributed as

$$\sum_{i=1}^{k-1} \theta_{in}(Z_i + \omega_{in}/\theta_{in}^{\frac{1}{2}})^2 + 2nI^{\lambda}(\pi_n:p) - \sum_{i=1}^{k-1} \omega_{in}^2$$

or, in a more suggestive notation, as

$$\text{(IV.2.7)} \qquad \sum_{i=1}^{k-1} \theta_{in}\chi_{1i}^2(\omega_{in}^2/\theta_{in}) + 2nI^{\lambda}(\pi_n:p) - \sum_{i=1}^{k-1} \omega_{in}^2.$$

This can also be derived from Corollary 2.2 in Dik and de Gunst (1985).

In the particular case $\lambda = 0$, the A^{λ} approximation to the distribution of $\overline{LR}_n$ reduces to the shifted noncentral χ^2 distribution

$$\text{(IV.2.8)} \qquad \chi_{k-1}^2(\delta_n^0) + \xi_n^0,$$

where

$$\text{(IV.2.9)} \qquad \begin{aligned} \delta_n^0 &= n \sum_{i=1}^{k} \pi_{in}(\log r_{in})^2 - n\left\{\sum_{i=1}^{k} \pi_{in} \log r_{in}\right\}^2 \\ \xi_n^0 &= 2n \sum_{i=1}^{k} \pi_{in} \log r_{in} - \delta_n^0 . \end{aligned}$$

In all other cases we get linear combinations of noncentral χ_1^2 distributions. The approximation for $\overline{P}_n$ is no exception; it is just (IV.2.7) with $\lambda = 1$.

A similar orthogonal transformation shows that $B^{\lambda}(U_n)$ is distributed as

$$\text{(IV.2.10)} \qquad \bar{r}_{n\lambda}\chi_{k-1}^2(\delta_n^{(\lambda)}) + \xi_n^{(\lambda)},$$

where

$$\text{(IV.2.11)} \qquad \begin{aligned} \delta_n^{(\lambda)} &= n(\lambda\bar{r}_{n\lambda})^{-2}\left[\sum_{i=1}^{k} \pi_{in}(r_{in}^{\lambda} - 1)^2 - \left\{\sum_{i=1}^{k} \pi_{in}(r_{in}^{\lambda} - 1)\right\}^2\right] \\ \xi_n^{(\lambda)} &= 2nI^{\lambda}(\pi_n:p) - \bar{r}_{n\lambda}\delta_n^{(\lambda)} . \end{aligned}$$

In the sense explained above B^{λ} is the "best noncentral χ^2 approximation" to the distribution of $\overline{CR}_n(\lambda)$. For $\lambda = 0$ it coincides with A^{λ}. The distribution of $B_1^{\lambda}(U_n)$ follows by inserting $\bar{r}_{n\lambda} = 1$ in (IV.2.10) and (IV.2.11).

The more classical approximations $\chi_{k-1}^2(\delta_n)$ with $\delta_n = 2nI^1(\pi_n:p)$ or $\delta_n = 2nI^{\lambda}(\pi_n:p)$ are in fact approximations to B_1^{λ} for $\max_i|\pi_{in}-p_i| \to 0$ as $n \to \infty$. In an expansion of π_{in} around p_i the parameters of the $B_1^{\lambda}(U_n)$ distribution can be written as

$$\begin{aligned} \delta_{n1}^{(\lambda)} &= 2nI^1(\pi_n:p) + n\lambda \sum_{i=1}^{k} (\pi_{in} - p_i)^3/p_i^2 + O(n\max_i |\pi_{in} - p_i|^4) \\ \xi_{n1}^{(\lambda)} &= -\tfrac{1}{3}n(2\lambda+1) \sum_{i=1}^{k} (\pi_{in} - p_i)^3/p_i^2 + O(n\max_i |\pi_{in} - p_i|^4). \end{aligned}$$

The $\chi^2_{k-1}(\delta_n)$ distributions arise if $\delta^{(\lambda)}_{n1}$ is approximated by δ_n and $\xi^{(\lambda)}_{n1}$ by 0. The advantage of A^λ and B^λ, that no expansions of π_{in} around p_i are made, is thus lost.

Obviously the linear approximation $L^\lambda(U_n)$ is normally distributed with expectation $2nI^\lambda(\pi_n:p)$ and variance

$$4n\lambda^{-2}\{\sum_{i=1}^{k} \pi_{in}(r^\lambda_{in}-1)^2 - (\sum_{i=1}^{k} \pi_i(r^\lambda_{in}-1))^2\}.$$

IV.2.2. Asymptotic error bounds.

As a counterpart to the stochastic expansions in subsection IV.2.1 we have the following theorem on the distribution error of the expansions. See Cox and Reid (1986) for related work.

THEOREM IV.2.1. *Let* $k \geq 3$ *and* $\lambda \in \mathbb{R}$. *Let* $0 < \varepsilon < \frac{1}{k}$ *and* $\Pi_\varepsilon = \{\pi \in \mathbb{R}^k;\ \min_i \pi_i \geq \varepsilon,\ \Sigma\, \pi_i = 1\}$. *Let* $\{s_n\}$, $s_n > 0$, *be a nondecreasing sequence, let* $\Pi(s_n) = \{\pi \in \Pi_\varepsilon;\ \max_i |\pi_i - p_i| \leq s_n/n^{\frac{1}{2}}\}$ *and let* $\Pi^*(s_n) = \Pi_\varepsilon \setminus \Pi(s_n)$.

i) *As* $n \to \infty$

$$\text{(IV.2.12)} \qquad \sup_{\pi\in\Pi_\varepsilon}\ \sup_{c>0} |P_\pi(\overline{CR}_n(\lambda) > c) - P(A^\lambda(U_n) > c)| = O(n^{-\frac{1}{2}}).$$

ii) *If* $s_n/n^{\frac{1}{2}} \to 0$, *then as* $n \to \infty$

$$\text{(IV.2.13)} \qquad \sup_{\pi\in\Pi(s_n)}\ \sup_{c>0} |P_\pi(\overline{CR}_n(\lambda) > c) - P(B^\lambda(U_n) > c)| = O(s_n n^{-\frac{1}{2}}).$$

iii) *If* $s_n/n^{\frac{1}{2}} \to 0$ *and* $\delta_{n\lambda} = 2nI^\lambda(\pi:p)$, *then as* $n \to \infty$

$$\text{(IV.2.14)} \qquad \sup_{\pi\in\Pi(s_n)}\ \sup_{c>0} |P_\pi(\overline{CR}_n(\lambda) > c) - P(\chi^2_{k-1}(\delta_{n\lambda}) > c)| = O(s_n n^{-\frac{1}{2}}).$$

iv) *If* $s_n/n^{\frac{1}{4}} \to 0$ *and* $\delta_n = 2nI^1(\pi:p)$, $\lambda \neq 1$, *then as* $n \to \infty$

$$\text{(IV.2.15)} \qquad \sup_{\pi\in\Pi(s_n)}\ \sup_{c>0} |P_\pi(\overline{CR}_n(\lambda) > c) - P(\chi^2_{k-1}(\delta_n) > c)| = O(s_n^2 n^{-\frac{1}{2}}).$$

v) *If* $s_n \to \infty$ *and* $s_n/n^{\frac{1}{2}} < 1$, *then as* $n \to \infty$

$$\text{(IV.2.16)} \qquad \sup_{\pi\in\Pi^*(s_n)}\ \sup_{c>0} |P_\pi(\overline{CR}_n(\lambda) > c) - P(L^\lambda(U_n) > c)| = O(s_n^{-1}).$$

The error bounds in iv) *and* v), *when larger than* $O(n^{-\frac{1}{2}})$, *are sharp.*

PROOF. cf. Section IV.4. □

REMARK IV.2.1. The theorem also holds for $k = 2$ if c is restricted to $\gamma < c < \infty$ for any fixed $\gamma > 0$, i.e. for levels bounded away from one. The bound in ii) remains valid if we further simplify $B^\lambda(U_n)$ by taking $\bar{r}_{n\lambda} = 1$ everywhere. □

REMARK IV.2.2. The error bound in v) also holds for normal approximations based on moments of $\overline{CR}_n(\lambda)$ or $A^\lambda(U_n)$ and remains sharp. □

REMARK IV.2.3. We have no proof that the bounds in i) - iii) are sharp. □

The strength of the theorem lies in the uniformity of the error bounds. The error of the A^λ approximation is at most $Cn^{-\frac{1}{2}}$ for *all* alternatives $\pi_n \in \Pi_\varepsilon$ (and all significance levels α_n). Hence the A^λ approximation is satisfactory both from a local and a nonlocal point of view. Taking $\pi_{in} = p_i$ $(i = 1,\dots,k)$, $A^\lambda(U_n) = \|U_n\|^2$ and hence under the null hypothesis the approximation is equivalent to the classical χ^2_{k-1} null approximation.

The B^λ and the $\chi^2_{k-1}(\delta_n)$ approximations have the same error bound $O(n^{-\frac{1}{2}})$ as A^λ for contiguous alternatives (with $\max_i |\pi_{in} - p_i| = O(n^{-\frac{1}{2}})$). For more distant alternatives they do not perform as well and they may not be any good at all for fixed alternatives. The classical approximation does even worse. For alternatives with $\max_i |\pi_{in} - p_i| \approx n^{-\frac{1}{4}}$ the error bound in iv) approaches $O(1)$, whereas the bounds in ii) and iii) are still $O(n^{-\frac{1}{4}})$. Examples show that for appropriate c and $\max_i |\pi_{in} - p_i| > n^{-\frac{1}{4}}$ the actual approximation error of $\chi^2_{k-1}(\delta_n)$ with $\delta_n = 2nI^1(\pi_n{:}p)$ can indeed increase to one for any $\lambda \neq 1$!

Of course these results are not surprising. For contiguous alternatives A^λ, B^λ and $\chi^2_{k-1}(\delta_n)$ do equally well because expanding π_{in} around p_i the three statistics are asymptotically equivalent up to $O_p(n^{-\frac{1}{2}})$. For more distant alternatives the use of local expansions leads to a loss and hence A^λ is superior.

The linear approximation L^λ shows quite different behaviour. It works quite well for fixed alternatives (comparable to A^λ), but breaks down completely for contiguous alternatives. Broffitt and Randles (1977) used the asymptotic normality of the Pearson statistic under fixed alternatives to propose a normal approximation to large powers of the $\overline{P}_n$ test. For large k ($k \to \infty$ as $n \to \infty$) the normal approximation also improves locally, see Morris (1975), but here we restrict ourselves to moderate values of k.

IV.2.3. Moments and moment adjustments.

The fit of the approximations to the true distribution of $\overline{CR}_n(\lambda)$ can also be judged from the similarity of the moments. Under fixed alternatives we find by elementary calculations

$$E_\pi\{\overline{CR}_n(\lambda)\} = E\{B^\lambda(U_n)\} + n^{-1}(\lambda-1)\Big\{\tfrac{1}{3}\Sigma r_i^\lambda(1-3\pi_i+2\pi_i^2)/\pi_i + \\ + \tfrac{1}{4}(\lambda-2)\Sigma r_i^\lambda(1-\pi_i)^2/\pi_i\Big\} + O(n^{-2}),\ \lambda > -1$$

$$E\{A^\lambda(U_n)\} = E\{B^\lambda(U_n)\}$$

$$E\{B^\lambda(U_n)\} = 2nI^\lambda(\pi:p) + \bar{r}_\lambda(k-1)$$

$$E\{B_1^\lambda(U_n)\} = 2nI^\lambda(\pi:p) + k - 1$$

$$E\{\chi^2_{k-1}(2nI^\lambda(\pi:p))\} = 2nI^\lambda(\pi:p) + k - 1$$

$$E\{\chi^2_{k-1}(2nI^1(\pi:p))\} = 2nI^1(\pi:p) + k - 1$$

$$E\{L^\lambda(U_n)\} = 2nI^\lambda(\pi:p)$$

and

$$\mathrm{var}_\pi\{\overline{CR}_n(\lambda)\} = \mathrm{var}\{A^\lambda(U_n)\} - 4(1+\lambda^{-1})\{\Sigma\ \pi_i r_i^{2\lambda} - (\Sigma\ \pi_i r_i^\lambda)^2\} + \\ + 4\{\Sigma\ r_i^{2\lambda} - \Sigma\ r_i^\lambda\ \Sigma\pi_i r_i^\lambda\} + O(n^{-1}),\ \lambda > -1$$

$$\mathrm{var}\{A^\lambda(U_n)\} = \mathrm{var}\{B^\lambda(U_n)\} - 2\{\Sigma\ \pi_i r_i^{2\lambda} - (\Sigma\ \pi_i r_i^\lambda)^2\} + \\ + 2\{\Sigma\ (1-\pi_i)r_i^{2\lambda} - \bar{r}_\lambda^2(k-1)\}$$

$$\mathrm{var}\{B^\lambda(U_n)\} = \mathrm{var}\{L^\lambda(U_n)\} + 2\bar{r}_\lambda^2(k-1)$$

$$\mathrm{var}\{B_1^\lambda(U_n)\} = \mathrm{var}\ \{L^\lambda(U_n)\} + 2(k-1)$$

$$\mathrm{var}\{\chi^2_{k-1}(2nI^\lambda(\pi:p))\} = 8nI^\lambda(\pi:p) + 2(k-1)$$

$$\mathrm{var}\{\chi^2_{k-1}(2nI^1(\pi:p))\} = 8nI^1(\pi:p) + 2(k-1)$$

$$\mathrm{var}\{L^\lambda(U_n)\} = 4n\lambda^{-2}\{\Sigma\ \pi_i r_i^{2\lambda} - (\Sigma\ \pi_i r_i^\lambda)^2\}$$

The first moment of A^λ and B^λ fits best to the moment of $\overline{CR}_n(\lambda)$. The fit deteriorates if we go to B_1^λ and the other approximations.

Of course the statistics can be adjusted by a linear transformation to get the "right" first two moments (of $\overline{CR}_n(\lambda)$). Such transformations are quite common for the Pearson statistic and the likelihood ratio statistic under the null hypothesis.

Cressie and Read (1985) calculate the first and second moment of $\overline{CR}_n(\lambda)$ under contiguous alternatives and suggest a moment adjusted $\chi^2_{k-1}(\delta_n)$ approximation, but they do not pursue this subject.

Broffitt and Randles (1977) use the exact moments of $\overline{P}_n$ under fixed alternatives to construct a normal approximation to the power of the $\overline{P}_n$ test. This is nothing else than the moment adjusted L^λ approximation. The order of magnitude of the error bound in Theorem IV.2.1v) is not changed. More numerical details are reported in Section IV.3.

IV.2.4. Asymptotic efficiencies.

A comparison of the (asymptotic) efficiency of the different $\overline{CR}_n(\lambda)$'s is not easily based on (IV.2.7) since this expression lacks transparency. From (IV.2.2) - (IV.2.4)

$$\overline{CR}_n(\lambda)/n = 2I^\lambda(\pi_n:p) + O_p(n^{-\frac{1}{2}}),$$

suggesting that $2I^\lambda(\pi_n:p)$ may be a good measure of the relative power of the tests in the Cressie-Read class. This coincides with the approach by approximate Bahadur efficiencies, since the approximate slope of the statistic $\overline{CR}_n(\lambda)$ at π equals $2I^\lambda(\pi:p)$. Of course this is also in accordance with the $\chi^2_{k-1}(2nI^\lambda(\pi:p))$ approximation of the power of the test based on $\overline{CR}_n(\lambda)$ (cf. also Moore (1984)).

The discrepancy measure $2I^\lambda(\pi_n:p)$ is still rather unwieldy. For alternatives with $\pi_{in} = p_i + o(1)$ $(i = 1,\dots,k)$ a Taylor expansion is informative:

$$2I^\lambda(\pi_n:p) = \sum_{i=1}^{k}(\pi_{in}-p_i)^2/p_i + \tfrac{1}{3}(\lambda-1)\sum_{i=1}^{k}\left(\frac{\pi_{in}}{p_i}-1\right)(\pi_{in}-p_i)^2/p_i + \frac{1}{12}(\lambda-1)(\lambda-2)\sum_{i=1}^{k}(\pi_{in}-p_i)^4/p_i^3 + \dots \quad \text{(IV.2.17)}$$

The first two terms of the expansion suggest an increase of power with λ for those alternatives π_n for which the more important contributions

$(\pi_{in} - p_i)^2/p_i$ to the leading term $2I^1(\pi_n:p)$ have positive $\pi_{in} - p_i$, and a decrease in the opposite case. In other words: large values of λ seem a good choice for sharply peaked likelihood ratios π_{in}/p_i; small values of λ seem preferable for deep dips of π_{in}/p_i. Cressie and Read (1984) make the same recommendations, based on numerical evidence.

If $p_1 = \ldots = p_k$ and the alternatives are anti-symmetric, i.e. for each i there is an index j such that $\pi_{in} - p_i = -(\pi_{jn} - p_j)$, the second term of the expansion (IV.2.17) vanishes and the third term may be of interest. This term suggest a local minimum of the power between $\lambda = 1$ and $\lambda = 2$.

IV.3. NUMERICAL RESULTS

To investigate the small-sample performance of the A^λ and B^λ approximations relative to the other approximations discussed in Section IV.2 we present some new numerical work and survey other work in the same area. Our conclusions are summarized in subsection IV.3.3.

IV.3.1. Power computations.

For sample sizes n = 20 and n = 50 and level α = .05 the true power of the tests in the Cressie-Read class is computed in a number of examples by direct enumeration. Randomization was used to get exact size .05. The corresponding critical values were also used as a starting point for the various power approximations, to put all approximations on an equal footing.

For small n and $\lambda \leq -1$ the critical value increases in giant steps to $+\infty$ as $\alpha \to 0$ (cf. Read (1984)). In fact, for $\lambda \leq -1$ the null probability $P_p(\overline{CR}_n(\lambda) = \infty) = P_p(\min_i N_i = 0)$ is considerable if $\min_i E_p\{N_i\}$ is small, e.g. for n = 20, k = 5 or n = 50, k = 10. If the critical value turned out to be $+\infty$, we discarded values $\lambda \leq -1$. (Since computations are in steps of .25 for λ, the smallest λ reported in such cases is -.75.)

The true power of the tests is an irregular discontinuous function of λ if n is small (this is somewhat obscured in our graphs due to linear interpolation). The reason is, that as λ varies sample points $(n_1,\ldots,n_k)$ in the rejection region are replaced by others with different probabilities under alternatives.

In Figure IV.3.1 the true power of the $\overline{CR}_n(\lambda)$ tests is compared to the traditional $\chi^2_{k-1}(2nI^1)$ approximation, the A^λ approximation (IV.2.7) and the

Figure IV.3.1. Three approximations to the power of the test $\overline{CR}_n(\lambda)$ $(-2 \le \lambda \le 4)$, compared to exact powers for level $\alpha = .05$ in a couple of examples with varying k,n,p and π: ———— = exact power; ---------- = A^λ approximation; ··········· = B^λ approximation; ——— – – = $\chi^2(2nI^1(\pi:p))$ approximation.

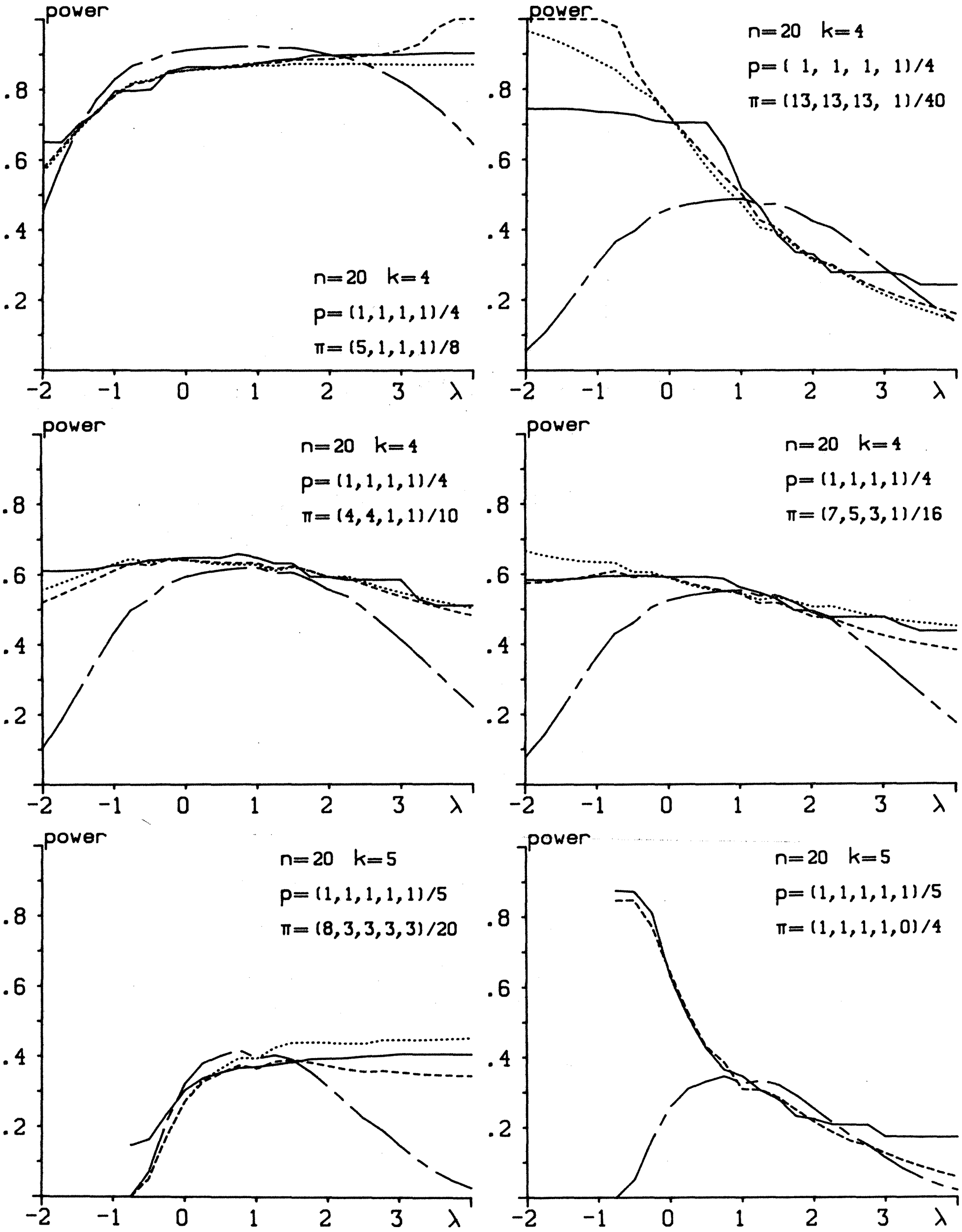


Figure IV.3.1 (continued).

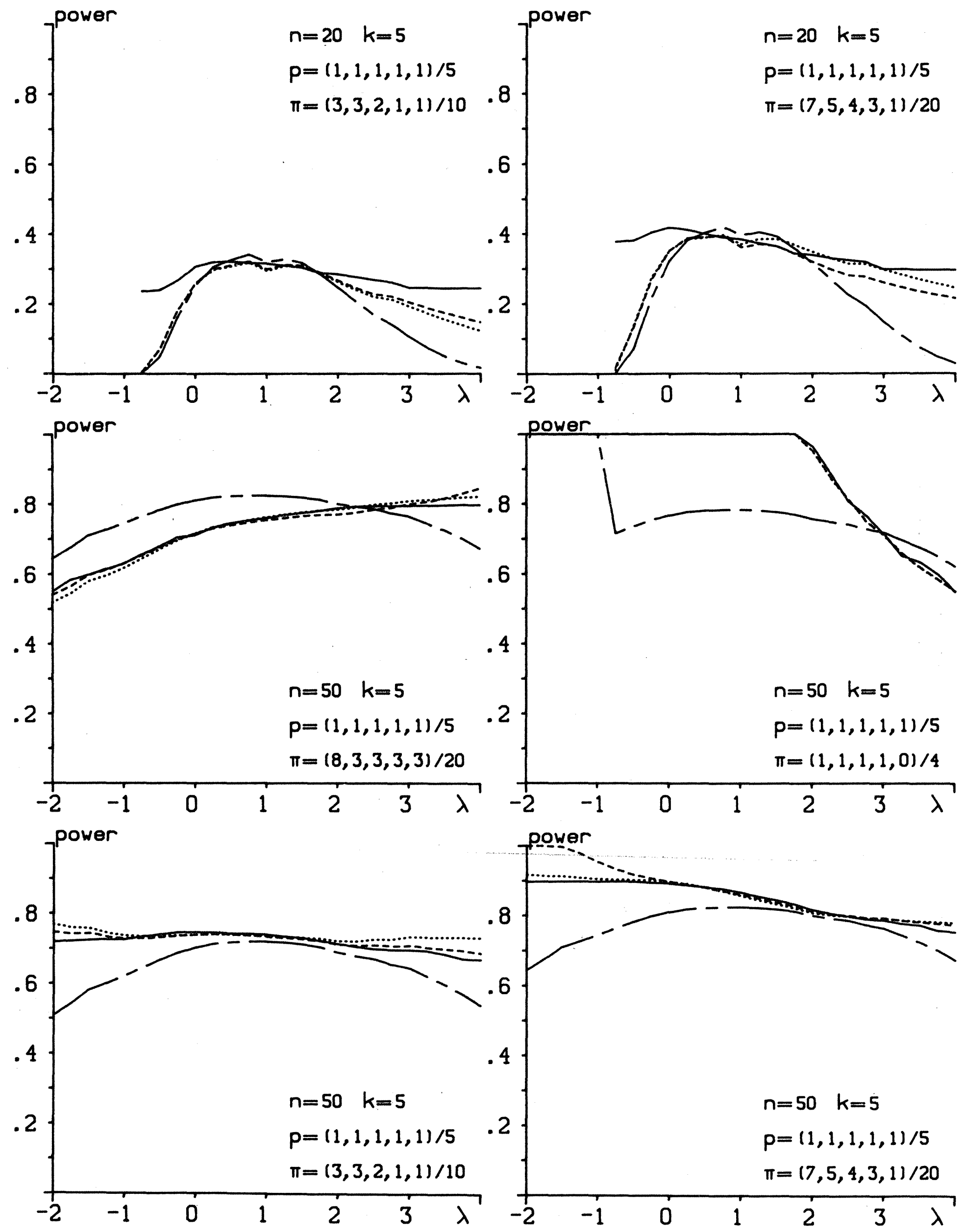
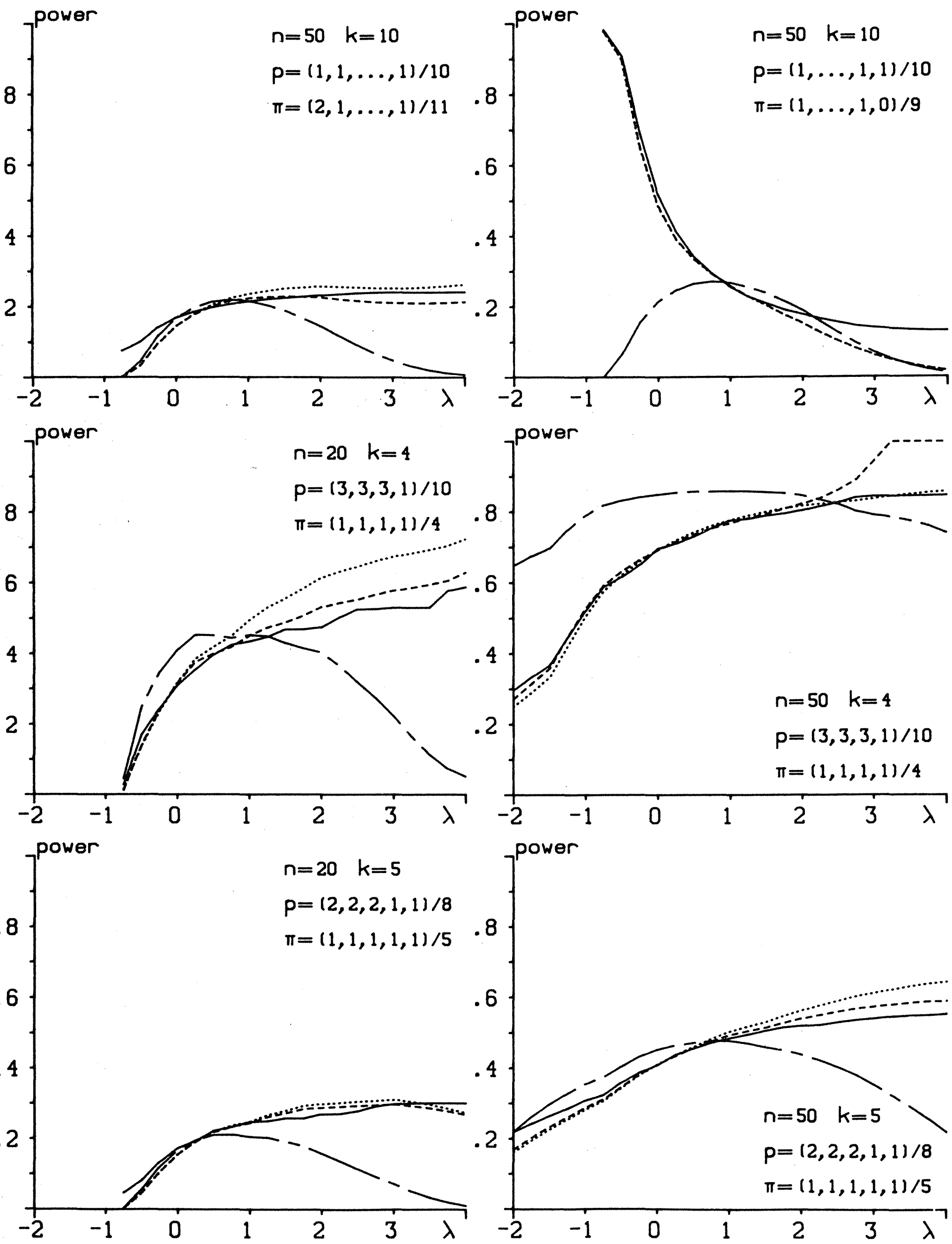


Figure IV.3.1 (continued).



B^{λ} approximation (IV.2.10). Based on these graphs and numerous other examples not reported here, we draw the following conclusions:

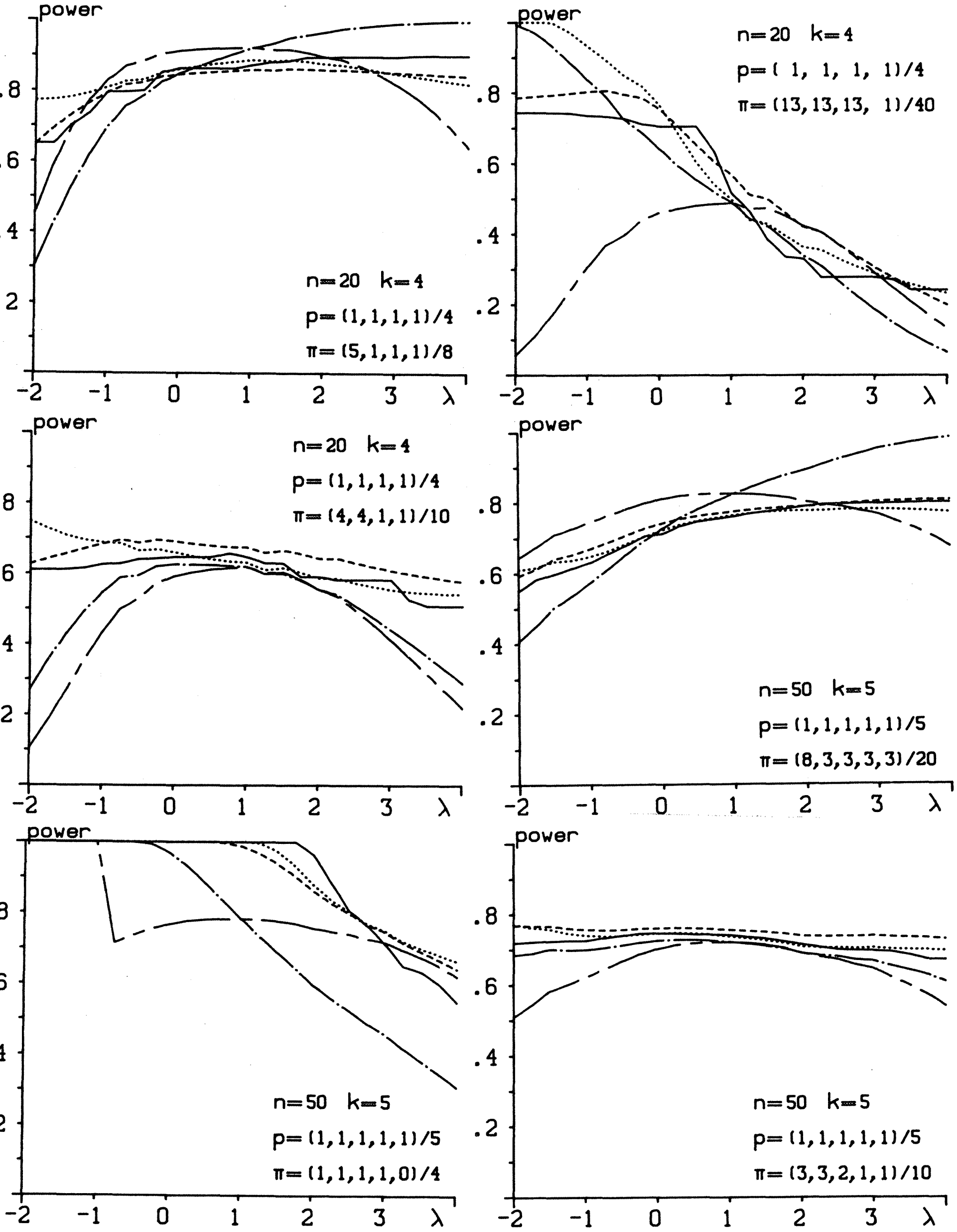
- the $\chi^2_{k-1}(2nI^1)$ approximation is tolerable but not quite satisfactory for $\lambda = 1$ (the Pearson test); the approximation error often exceeds .04, even for $n = 100$. For values of λ away from 1 the approximation breaks down completely;
- the A^{λ} approximation is superior to the other approximations and is quite accurate for a wide range of λ values; the approximation error is usually smaller than .03 ($n = 20$, $0 \le \lambda \le 2\frac{1}{2}$) or .015 ($n = 50$, $-\frac{1}{2} \le \lambda \le 3$); for the $\overline{P}_n$ test the approximation is excellent;
- the B^{λ} approximation is satisfactory but not as accurate as the A^{λ} approximation (except near $\lambda = 0$ where they coincide);
- if one or more of the expected cell counts under the hypothesis is small, approximations are often inaccurate for $\lambda \le 0$, but probabilities $\pi_i = 0$ (under G_n) do not invalidate the A^{λ} approximation;
- for alternatives where one π_i/p_i is very large, the power increases with λ; for alternatives with one very small value of π_i/p_i, the power decreases, and for alternatives with only moderate values of π_i/p_i, the highest power is usually achieved for λ between -1 and $\frac{1}{4}$.

These conclusions confirm the simulation results displayed in Figures IV.1.1 and IV.1.2 and agree with the theory developed in Section IV.2. Somewhat surprisingly, the A^{λ} and B^{λ} approximations under alternatives are satisfactory for a broader range of λ values than the χ^2 null hypothesis approximation (cf. Cressie and Read (1984)), although A^{λ} and B^{λ} reduce to this approximation under H_0.

We also comment on some other approximations discussed in Section IV.2. The $\chi^2_{k-1}(2nI^{\lambda})$ approximation is an improvement on the $\chi^2_{k-1}(2nI^1)$ approximation for $\lambda \neq 1$, but falls short of the B^{λ} approximation. Adjusting the first moments, using the nonlocal terms of $E\{\overline{CR}_n(\lambda)\}$, reported in subsection IV.2.3, the adjusted $\chi^2_{k-1}(\delta_n)$ approximations for $\delta_n = 2nI^{\lambda}$ and $\delta_n = 2nI^1$ almost coincide. For small sample sizes these approximations turn out to be serious competitors to B^{λ}. This can be theoretically justified, the moment corrected approximations have the same error bound as B^{λ}.

A few typical cases are displayed in Figure IV.3.2. The B_1^{λ} approximation is about as effective as B^{λ} for $0 \leqq \lambda \leqq 1$, but slightly less

<u>Figure IV.3.2</u>. Four approximations to the power of the test $\overline{CR}_n(\lambda)$ ($-2 \le \lambda \le 4$), compared to exact powers for level $\alpha = .05$ in a couple of examples with varying k,n,p and π: ——————— = exact power; ·················· = moment corr. $\chi^2(2nI^1(\pi:p))$ approx.; ——— — — = $\chi^2(2nI^1(\pi:p))$ approx.; —·———·— = $\chi^2(2nI^\lambda(\pi:p))$ approx.; --------- = $N(E_0\{A^\lambda\}, \mathrm{var}_0\{A^\lambda\})$ approx.



accurate outside this interval.

The linear approximation L^{λ}, cf. (IV.2.5), is disastrous. Adjusting its moments by using a normal approximation with the moments of $\overline{CR}_n(\lambda)$ is an improvement, but the normal approximation is still better when the A^{λ} moments are employed (see subsection IV.2.3). It then competes with $\chi^2_{k-1}(2nI^{\lambda})$. Yet, errors of .06 (n = 20) or .04 (n = 50) are quite common in the range $-\frac{1}{2} < \lambda < 2\frac{1}{2}$. See Figure IV.3.2.

A few words on computational aspects are in order. The noncentral χ^2 approximations do not present any difficulties since library routines are widely available. To compute the A^{λ} approximation (IV.2.7) the eigenvalues and eigenvectors in equation (IV.2.6) are first determined. One can then proceed in different ways. We have employed a procedure described by Kotz, Johnson and Boyd (1967) which expresses the cdf of (IV.2.7) as a weighted sum of cdf's of central χ^2 distributions with positive and negative weights. Sheil and O'Muircheartaigh (1977) give a similar procedure with positive weights only. A different approach is described in Davies (1980), cf. also Farebrother (1984).

IV.3.2. Related work.

The A^{λ} statistic for $\lambda = 1$ first appears in Patnaik (1949, p. 219), he refers to a Pearson Type III distribution and finds the approximation too cumbersome. In his opinion the $\chi^2_{k-1}(2nI^1)$ approximation is adequate for practical purposes (when $\overline{P}_n$ is used).

Slakter (1968) simulated the power of the Pearson test in a large number of cases and compared it to the $\chi^2_{k-1}(2nI^1)$ approximation. His main conclusion is that the χ^2 approximation overestimates the true power quite a bit for small sample sizes. Haber (1980) remarks that Slakter's conclusion is too pessimistic; he only finds that χ^2 approximations close to one are suspect.

In West and Kempthorne (1971) and Kallenberg et al. (1985) the powers of the $\overline{P}_n$ test and the $\overline{LR}_n$ test are compared. West and Kempthorne (1971) find that the "$\overline{P}_n$ test seems to have more sensitivity against alternatives for which one component of π is relatively large, whereas the $\overline{LR}_n$ test appears better against alternatives where one component is relatively small". This exactly confirms our own results and is also in agreement with the numerical results in Cressie and Read (1984) and in Kallenberg et al. (1985). West and Kempthorne (1971) also note that, for fixed alternatives, the differences between the sensitivities of both tests are diminished if n increases. This can be explained by a local expansion of the noncentrality $2nI^0(\pi:p)$. However, even for very large n there are alternatives for which the power differences are large since $2n|I^1(\pi_n:p) - I^0(\pi_n:p)|$ can be made large for any n.

IV.3.3. Discussion

Global asymptotic expansions of test statistics are a useful tool for constructing satisfactory power approximations. The importance of a global approach is illustrated not only by the striking accuracy of the A^λ approximation, but also by the fact that the strictly nonlocal normal approximation (with A^λ moments) is a serious competitor to the strictly local $\chi^2_{k-1}(\delta_n)$ approximation. The asymptotic error bounds give a good idea of the relative accuracy of the various approximations in small samples. Moment adjustments often improve approximations.

Our recommendations on the use of power approximations are as follows:

i) do not use the $\chi^2_{k-1}(2nI^1)$ approximation for other $\overline{CR}_n(\lambda)$ tests than $\overline{P}_n$;

ii) for accurate work always use the A^λ approximation (or exact computation);

iii) for many practical purposes B^λ is a good and simple approximation for $-1 < \lambda < 3$. For $0 \le \lambda \le 1$ the B_1^λ approximation is equally effective and slightly simpler;

iv) for quick and dirty work the unorthodox normal approximation based on the moments of A^λ (see subsection IV.2.3) deserves consideration.

Since global expansions are not very transparent, local expansions are still useful to get insight in qualitative power properties of the tests. Local theory and extensive numerical work by several authors lead to the following recommendations concerning the choice of λ:

a) if alternatives with one or two large values of π_i/p_i are of special interest, use the Pearson test ($\lambda = 1$) or $\overline{CR}_n(\lambda)$ with $\lambda = 2$;

b) if alternatives with one or two small values of π_i/p_i are of special interest, use the likelihood ratio test ($\lambda = 0$) or the Freeman-Tukey test ($\lambda = -\frac{1}{2}$);

c) in other cases the $\overline{P}_n$ and $\overline{LR}_n$ test are competative; the $\overline{LR}_n$ test is perhaps more powerful for a somewhat broader range of alternatives, but the null hypothesis distribution of $\overline{P}_n$ is easier to control (Cressie and Read (1984) advise $\lambda = \frac{2}{3}$).

There seems to be no sound reason to choose extreme values of λ; distributions under hypothesis and alternative are hard to approximate and the power gain is uncertain.

IV.4. PROOFS

Before proving Theorem IV.2.1 we derive some preliminary results. Let $Z_1, Z_2, \ldots$ be i.i.d. standard normal variables with cdf Φ and density ϕ. Repeatedly we use the order relation

$$(\text{IV}.4.1) \qquad P(|Z_1| > \log n) = o(n^{-\frac{1}{2}}),$$

which continues to hold if Z_1 is replaced by a standardized binomial Bin(n,p) variable, uniformly in p bounded away from 0 and 1. We begin with a crucial lemma, which is in the same spirit as Theorem 1 in Cox and Reid (1986). Although many constants and sets in the sequel depend on n, subscripts n are suppressed to simplify notation.

<u>LEMMA IV.4.1</u>. *Let* a_i, b_i $(i=1,\ldots,m)$ *and* c *be real numbers,* $m \geq 2$, *let* $P_i(\cdot)$ $(i=1,\ldots,m)$ *be polynomials of fixed degree* $q \geq 0$ *and let* $a_0 > 0$ *and* $d_0 > 0$ *be fixed. Uniformly for* $a_i > a_0$, $b_i \in \mathbb{R}$, $c > 0$ *and the coefficients of the* P_i *bounded by* d_0,

$$(\text{IV}.4.2) \qquad P\left(\sum_{i=1}^{m} a_i(Z_i - b_i)^2 + n^{-\frac{1}{2}} \sum_{i=1}^{m} P_i(|Z_i|) \leq c\right) = P\left(\sum_{i=1}^{m} a_i(Z_i - b_i)^2 \leq c\right) + O(n^{-\frac{1}{2}}).$$

The relation continues to hold for $m = 1$ *if* $c > \gamma$ *for some fixed* $\gamma > 0$.

<u>PROOF</u>. Assume without loss of generality $b_i \geq 0$ $(i=1,\ldots,m)$. We systematically treat the case $m \geq 2$ and only comment on points where arguments for $m = 1$ are different. Put

$$c_0 = 4^{q+1} m d_0 n^{-\frac{1}{2}} \min\{\log^q n, \max\{1, b_1^q, \ldots, b_m^q\}\}.$$

We consider three cases.

i) Let $q = 0$. The statistic in the LHS of (IV.4.2) is now of the form $\Sigma\, a_i(Z_i - b_i)^2 + n^{-\frac{1}{2}}d$ with $|d| \le d_0$ and (IV.4.2) follows from the mean value theorem if the density of $\Sigma\, a_i(Z_i - b_i)^2$ is bounded. For $m = 2$ it is easily derived from a convolution representation that the density is smaller than $\frac{1}{2}(a_1 a_2)^{-\frac{1}{2}}$. An induction argument shows that for $m \ge 3$ the density is bounded by the same bound. For $m = 1$ the density is trivially bounded for arguments away from zero.

ii) Let $c \ge c_0$ and $q \ge 1$. It is sufficient to prove the lemma for the special cases $P_i(|z_i|) = d_0\, \Sigma_{j=1}^{q}\, |z_i|^j$ $(i = 1,\dots,m)$ and $P_i(|z_i|) = -d_0\Sigma_{j=1}^{q}\, |z_i|^j$ $(i = 1,\dots,m)$ (the constant terms have been dealt wtih in i)). Let $T(z) = \Sigma\, a_i(z_i - b_i)^2 + n^{-\frac{1}{2}}\, \Sigma\, P_i(|z_i|)$ $(z = (z_1,\dots,z_m)' \in \mathbb{R}^m)$ and $T = T(Z_1,\dots,Z_m)$. Write $T = T_+$ for the first special case and $T = T_-$ for the second special case. Put

$$h(z) = m \min\{\log^q n,\ \max\{\max_i |z_i|,\ \max_i |z_i|^q\}\} \quad (z \in \mathbb{R}^m),$$

$$S = \{z \in \mathbb{R}^m;\ \Sigma\, a_i(z_i - b_i)^2 = c\} \quad \text{and}$$

$$t_z = 2^{q+1} d_0 n^{-\frac{1}{2}} c^{-1} h(z).$$

We first show that $0 \le t_z \le \frac{1}{2}$ for $z \in S$ and large n. Note that

$$\begin{aligned}\max_i |z_i| &\le \max_i |z_i - b_i| + \max_i b_i\\ &\le a_0^{-\frac{1}{2}}\{\Sigma\, a_i(z_i - b_i)^2\}^{\frac{1}{2}} + \max_i b_i\\ &= a_0^{-\frac{1}{2}} c^{\frac{1}{2}} + \max_i b_i .\end{aligned}$$

Hence, if $\max_i b_i \ge a_0^{-\frac{1}{2}} c^{\frac{1}{2}}$, then in view of $c \ge c_0$

$$t_z \le 2^{q+1} d_0 n^{-\frac{1}{2}} c^{-1} m \min\{\log^q n, \max\{2 \max_i b_i, 2^q \max_i b_i^q\}\} \le \tfrac{1}{2}.$$

Conversely, if $\max_i b_i \le a_0^{-\frac{1}{2}} c^{\frac{1}{2}}$, then for large n

$$t_z \le 2^{q+1} d_0 n^{-\frac{1}{2}} c^{-1} m \max_i |z_i| \le 2^{q+2} d_0 m (a_0 n c)^{-\frac{1}{2}} < \tfrac{1}{2}$$

if $\max_i |z_i| \le 1$ or if $q = 1$, and

$$t_z \le 2^{q+1} d_0 n^{-\frac{1}{2}} c^{-1} m \max_i z_i^2 \log^{q-2} n \le 2^{q+3} a_0^{-1} m d_0 n^{-\frac{1}{2}} \log^{q-2} n < \tfrac{1}{2}$$

otherwise. Hence $t_z \le \frac{1}{2}$. Let

$$A_+ = \{y \in \mathbb{R}^m;\ y = z - t(z-b), t_z < t \le 1, z \in S, \max_i |y_i| < \log n\}$$

Suppose $y \in A_+$. Then

$$\begin{aligned}
&T_+(y)\\
&= (1-t)^2 \, \Sigma\, a_i(z_i - b_i)^2 + n^{-\frac{1}{2}} d_0 \sum_{j=1}^{q} \min\{m \log^j n, \sum_{i=1}^{m} |z_i - t(z_i - b_i)|^j\}\\
&\le (1-t)^2 c + n^{-\frac{1}{2}} d_0 \sum_{j=1}^{q} \min\{m \log^j n, 2^{j-1} m \max_i |z_i|^j +\\
&\qquad\qquad + 2^{j-1} t^j m \max_i |z_i - b_i|^j\}\\
&\le (1-t)c + n^{-\frac{1}{2}} d_0 2^q m \min\{\log^q n, \max\{\max_i |z_i|, \max_i |z_i|^q\}\} +\\
&\qquad + n^{-\frac{1}{2}} d_0 \sum_{j=1}^{q} \min\{m \log^j n, 2^j t^j m \max_i |z_i - b_i|^j\}\\
&\le c - tc + \tfrac{1}{2} t_z c + n^{-\frac{1}{2}} d_0 m \sum_{j=2}^{q} \min\{\log^{j-2} n, (2a_0^{-\frac{1}{2}} c^{\frac{1}{2}})^{j-2}\} 4a_0^{-1} c\\
&\qquad + 2n^{-\frac{1}{2}} d_0 m (a_0 c)^{-\frac{1}{2}} c\\
&\le c,
\end{aligned}$$

where the last inequality holds for large n. Hence $T_+(y) \le c$ if $y \in A_+$. Next let

$$A_- = \{y \in \mathbb{R}^m;\ y = z + t(z-b), t_z < t < \infty, z \in S, \max_i |y_i| < \log n\}.$$

Suppose $y \in A_-$. Then by a similar argument, for large n,

$$\begin{aligned}
&T_-(y)\\
&= (1+t)^2 \, \Sigma\, a_i(z_i - b_i)^2 - n^{-\frac{1}{2}} d_0 \sum_{j=1}^{q} \min\{m \log^j n, \sum_{i=1}^{m} |z_i + t(z_i - b_i)|^j\}\\
&\ge c + 2tc - \tfrac{1}{2} t_z c - n^{-\frac{1}{2}} d_0 m \sum_{j=1}^{q} \min\{\log^j n, (2ta_0^{-\frac{1}{2}} c^{\frac{1}{2}})^j\}\\
&> c
\end{aligned}$$

(to obtain the last inequality consider small and large values of t seperately). Hence $T_-(y) > c$ if $y \in A_-$ and if n is large enough. Put

$$D = \{y \in \mathbb{R}^m:\ y = z + t(z-b), -t_z \le t \le t_z, z \in S\}.$$

The preceding arguments show that the symmetric difference of the sets $\{z : T(z) \le c\}$ and $\{z : \Sigma\, a_i(z_i - b_i)^2 \le c\}$ is contained in the strip D around the surface S of an m-dimensional ellipsoid and possible in the distant set

$\{z\,;\max_i|z_i| > \log n\}$ which has probability $o(n^{-\frac{1}{2}})$. The width of the strip is larger for points z of S far from the origin than for points close to the origin. The probability of the event $(Z_1,\ldots,Z_m) \in D$ is small precisely because the strip is very narrow in the region of large density. Obviously

$$|P(T \le c) - P(\sum_{i=1}^m a_i(Z_i - b_i)^2 \le c)| \le \int_D\!\!\ldots\!\int \prod_{i=1}^m \phi(y_i)dy_1\ldots dy_m + o(n^{-\frac{1}{2}}).$$

For large n we have on $D \cap \{y\,;\max_i|y_i| \le \log n\}$

$$\prod_{i=1}^m \phi(y_i) = \prod_{i=1}^m \phi(z_i + t(z_i - b_i)) = \prod_{i=1}^m \phi(z_i)(1 + o(1)) \quad (z \in S).$$

This follows from $t^2\,\Sigma(z_i - b_i)^2 \le t_z^2 a_0^{-1}c = o(1)$ and, using the Cauchy-Schwarz inequality, from $|t\,\Sigma z_i(z_i - b_i)| \le t_z(a_0^{-1}c\,\Sigma z_i^2)^{\frac{1}{2}} = o(1)$ since $\Sigma z_i^2 \le 2\,\Sigma y_i^2 + 2t^2\,\Sigma(z_i - b_i)^2 \le 2(m+1)\log^2 n$ for large n.

Hence it suffices to prove that uniformly

$$\text{(IV.4.3)} \qquad \int_D\!\!\ldots\!\int \prod_{i=1}^m \phi(z_i)dy_1\ldots dy_m = O(n^{-\frac{1}{2}}).$$

Substituting $v_i = (y_i - b_i)a_i^{\frac{1}{2}}$ $(i = 1,\ldots,m)$ the LHS of (IV.4.3) equals

$$\text{(IV.4.4)} \qquad \prod_{i=1}^m a_i^{-\frac{1}{2}} \int_{D^*}\!\!\ldots\!\int \prod_{i=1}^m \phi(a_i^{-\frac{1}{2}}w_i + b_i)dv_1\ldots dv_m,$$

where

$$D^* = \{v \in \mathbb{R}^m; v = (1+t)w, -t_w \le t \le t_w, w \in S^*\}$$

$$S^* = \{w \in \mathbb{R}^m; \sum_{i=1}^m w_i^2 = c\}$$

$$t_w = 2^{q+1}d_0 n^{-\frac{1}{2}}c^{-1}h^*(w) \quad \text{with } h^*(w) = h(a_1^{-\frac{1}{2}}w_1 + b_1,\ldots,a_m^{-\frac{1}{2}}w_m + b_m).$$

Introduce polar coordinates $r, \theta_1,\ldots,\theta_{m-1}$; let $v_1 = r\cos\theta_1$, $v_2 = r\sin\theta_1\cos\theta_2,\ldots,v_{m-1} = r\sin\theta_1\ldots\sin\theta_{m-2}\cos\theta_{m-1}$, $v_m = r\sin\theta_1\ldots\sin\theta_{m-2}\sin\theta_{m-1}$.
Write $g(\theta) = (\sin\theta_1)^{m-2}(\sin\theta_2)^{m-3}\ldots\sin\theta_{m-2}$, $r_1 = (1 - t_w)c^{\frac{1}{2}}$ and $r_2 = (1 + t_w)c^{\frac{1}{2}}$. Noting that $w \in S^*$ only depends on $\theta_1,\ldots,\theta_{m-1}$ (and not on r) and that $r_2^m - r_1^m \le (2c^{\frac{1}{2}})^m t_w$ (since $0 \le t_w \le \frac{1}{2}$), (IV.4.4) equals

$$\prod_{i=1}^m a_i^{-\frac{1}{2}} \int_0^\pi\!\!\ldots\!\int_0^\pi\!\int_0^{2\pi}\!\int_{r_1}^{r_2} r^{m-1}g(\theta) \prod_{i=1}^m \phi(a_i^{-\frac{1}{2}}w_i + b_i)dr d\theta_{m-1}\ldots d\theta_1$$

$$\text{(IV.4.5)} \qquad \le \gamma_0 m^{-1}(\Pi\, a_i^{-\frac{1}{2}})n^{-\frac{1}{2}}c^{\frac{1}{2}m-1}\int_0^\pi\!\!\ldots\!\int_0^\pi\!\int_0^{2\pi} h^*(w)g(\theta)\,\Pi\,\phi(a_i^{-\frac{1}{2}}w_i + b_i)d\theta_{m-1}\ldots d\theta_1$$

$$\le \gamma_0\gamma_1(\Pi\, a_i^{-\frac{1}{2}})n^{-\frac{1}{2}}c^{\frac{1}{2}m-1}\int_0^\pi\!\!\ldots\!\int_0^\pi\!\int_0^{2\pi} g(\theta)\,\Pi\,\phi(\tfrac{1}{2}(a_i^{-\frac{1}{2}}w_i + b_i))d\theta_{m-1}\ldots d\theta_1$$

where $\gamma_0 = 2^{q+1+m} d_0$. The last inequality follows from $\max_i |z_i|^q \, \Pi \, e^{-\frac{3}{8} z_i^2} \le$ $\le \max_i |z_i|^q \exp(-\frac{3}{8} \max_i z_i^2) \le \gamma_1$, for all $z \in \mathbb{R}^m$. The integral in the RHS of (IV.4.5) is certainly bounded and hence (IV.4.3) is proved if c is bounded. Suppose $c > 1$. Note that $c^{-1} \le \frac{1}{2} m^{-1} \{(1+c^{-1})^m - (1-c^{-1})^m\}$. Put $\rho_1 = (1-c^{-1}) c^{\frac{1}{2}}$ and $\rho_2 = (1+c^{-1}) c^{\frac{1}{2}}$, then the RHS of (IV.4.5) is smaller than

$$\frac{1}{2} \gamma_0 \gamma_1 (\Pi \, a_i^{-\frac{1}{2}}) n^{-\frac{1}{2}} \int_0^\pi \ldots \int_0^\pi \int_0^{2\pi} \int_{\rho_1}^{\rho_2} r^{m-1} g(\theta) \, \Pi \, \phi(\tfrac{1}{2}(a_i^{-\frac{1}{2}} w_i + b_i)) \, dr \, d\theta_{m-1} \ldots d\theta_1$$

$$\le \frac{1}{2} \gamma_0 \gamma_1 n^{-\frac{1}{2}} \exp((8a_0)^{-1}) \int_{D_0} \ldots \int \Pi \, \phi(2^{-\frac{3}{2}} y_i) \, dy_1 \ldots dy_m$$

where

$$D_0 = \{y \in \mathbb{R}^m ; \; y = z + t(z-b), -c^{-1} \le t \le c^{-1}, z \in S\}$$

and where we have used the inequality (in the reverse order) $\Sigma \, y_i^2 = \Sigma \, (z_i + t(z_i - b_i))^2 \le 2 \Sigma \, z_i^2 + 2c^{-2} \Sigma \, (z_i - b_i)^2 \le 2 \Sigma \, z_i^2 + 2a_0^{-1}$, valid for $y \in D_0$. Since the last integral is bounded, (IV.4.3) is proved in this case too.

For $m = 1$ (IV.4.3) is replaced by

$$\Phi(z_0 + t_{z_0} \tilde{c}) - \Phi(z_0 - t_{z_0} \tilde{c}) + \Phi(z_{00} + t_{z_{00}} \tilde{c}) - \Phi(z_{00} - t_{z_{00}} \tilde{c}) = O(n^{-\frac{1}{2}})$$

by the mean value theorem, where $z_0 = b_1 + (c/a_1)^{\frac{1}{2}}$, $z_{00} = b_1 - (c/a_1)^{\frac{1}{2}}$ and $\tilde{c} = (c/a_0)^{\frac{1}{2}}$ and where c is assumed bounded away from zero.

iii) Let $c < c_0$. We first show that uniformly

$$\text{(IV.4.6)} \qquad P(\Sigma \, a_i (Z_i - b_i)^2 \le c_0) = O(n^{-\frac{1}{2}})$$

The noncentral chi-square density (at x) of $(Z_1 - b_1)^2 + Z_2^2$ is given by

$$(4\pi)^{-1} \int_0^x y^{-\frac{1}{2}} (x-y)^{-\frac{1}{2}} \{1 + \exp(-2b_1 y^{\frac{1}{2}})\} \exp\{-\tfrac{1}{2}(y^{\frac{1}{2}} - b_1)^2 - \tfrac{1}{2}(x-y)\} \, dy$$

$$\le (2\pi)^{-1} \exp(-\tfrac{1}{2} x - \tfrac{1}{2} b_1^2 + b_1 x^{\frac{1}{2}}) \int_0^x y^{-\frac{1}{2}} (x-y)^{-\frac{1}{2}} dy$$

$$= (\tfrac{1}{2} \pi)^{\frac{1}{2}} \phi(x^{\frac{1}{2}} - b_1).$$

Assuming $b_1 = \max_i b_i$, it follows that

$$P(\Sigma \, a_i (Z_i - b_i)^2 \le c_0)$$

$$\le P((Z_1 - b_1)^2 + Z_2^2 \le a_0^{-1} c_0)$$

$$\le (\tfrac{1}{2}\pi)^{\frac{1}{2}} \int_0^{c_0/a_0} \phi(x^{\frac{1}{2}} - b_1)dx$$
$$\le (\tfrac{1}{2}\pi)^{\frac{1}{2}} a_0^{-1} c_0 \max_{0\le\theta\le 1} \phi(\theta a_0^{-\frac{1}{2}} c_0^{\frac{1}{2}} - b_1).$$

It is easily verified that the RHS is of order $O(n^{-\frac{1}{2}})$, proving (IV.4.6). For $m = 1$ and c bounded away from zero this case does not occur.

Now suppose the lemma is proved for $c = c_0$ (see i) and ii)). Then for $c < c_0$

$$P(T \le c) \le P(T \le c_0) = P(\Sigma\, a_i(Z_i - b_i)^2 \le c_0) + O(n^{-\frac{1}{2}}) = O(n^{-\frac{1}{2}}).$$

Since also $P(\Sigma\, a_i(Z_i - b_i)^2 \le c) = O(n^{-\frac{1}{2}})$, (IV.4.2) immediately follows.

This completes the proof of Lemma IV.4.1. □

REMARK IV.4.1. There is nothing sacred about the integers $n \in \mathbb{N}$ in (IV.4.2). They can be replaced by any $s_n > 0$ such that $s_n \to \infty$ as $n \to \infty$. □

COROLLARY IV.4.2. *Let* $U_1,\dots,U_k$ *be jointly* $N_k(0, I_k - \pi^{\frac{1}{2}}\pi^{\frac{1}{2}\prime})$ *distributed with* $\pi \in \Pi_\varepsilon$. *Replacing* $Z_1,\dots,Z_m$ *in Lemma* IV.4.1 *by* $U_1,\dots,U_k$, *the lemma remains valid in the sense that* (IV.4.2) *holds for* $k \ge 3$ *(and also for* $k = 2$ *if* $c > \gamma > 0$).

PROOF. By an orthogonal transformation similar to that in Section IV.2

$$\sum_{i=1}^{k} a_i(U_i - b_i)^2 \sim \sum_{i=1}^{k-1} \alpha_i(Z_i - \beta_i)^2 + \beta_0^2$$

and

$$\sum_{i=1}^{k} P_i(|U_i|) = \sum_{i=1}^{k}\sum_{j=0}^{q} d_{ij}|U_i|^j \sim \sum_{i=1}^{k}\sum_{j=0}^{q} d_{ij}\left|\sum_{s=1}^{k-1} g_{is}Z_s\right|^j$$

where the α_i are positive and bounded away from zero, the g_{is} (and d_{ij}) are bounded and $\sim$ denotes "is distributed as". Since for $1 \le j \le q$

$$\left|\sum_{s=1}^{k-1} g_{is}Z_s\right|^j \le \sum_{s=1}^{k-1} h_{is}^{(j)}|Z_s|^j,$$

where the $h_{is}^{(j)}$ are again bounded, the desired result follows from Lemma IV.4.1. □

LEMMA IV.4.3. *Let* a_i, d_i $(i = 1,\dots,m)$ *and* c *be real numbers,* $m \ge 1$, *and let* $a_0 > 0$ *and* $d_0 > 0$ *be fixed. Then, uniformly for* $\max_i |a_i| > a_0$, $|d_i| < d_0$ *and* $c > 0$,

$$P\left(\sum_{i=1}^{m} a_iZ_i + n^{-\frac{1}{2}} \sum_{i=1}^{m} d_iZ_i^2 < c\right) = P\left(\sum_{i=1}^{m} a_iZ_i < c\right) + O(n^{-\frac{1}{2}}).$$

The error bound is sharp; Remark IV.4.1 *again applies.*

PROOF. It is sufficient to prove that

$$\text{(IV.4.7)} \qquad P(\Sigma\, a_i Z_i \pm n^{-\frac{1}{2}} d_0 \Sigma Z_i^2 < c) = P(\Sigma\, a_i Z_i < c) + O(n^{-\frac{1}{2}}).$$

Consider an orhtogonal transformation $\tilde{Z} = \Psi Z$ where Ψ is an $m \times m$ orthonormal matrix with first row $||a||^{-1}(a_1,\ldots,a_m)$. Then

$$P(\Sigma\, a_i Z_i \pm n^{-\frac{1}{2}} d_0 \Sigma Z_i^2 < c) = P(||a||\tilde{Z}_1 \pm n^{-\frac{1}{2}} d_0 \Sigma \tilde{Z}_i^2 < c).$$

By direct calculation, using $||a|| \geq a_0$, uniformly

$$P(||a||\tilde{Z}_1 \pm n^{-\frac{1}{2}} d_0 \tilde{Z}_1^2 < c) = P(||a||\tilde{Z}_1 < c) + O(n^{-\frac{1}{2}}).$$

Hence, by a convolution argument,

$$P(||a||\tilde{Z}_1 \pm n^{-\frac{1}{2}} d_0 \Sigma \tilde{Z}_i^2 < c) = P(||a||\tilde{Z}_1 \pm n^{-\frac{1}{2}} d_0 \chi^2_{m-1} < c) + O(n^{-\frac{1}{2}}),$$

where $\tilde{Z}_1$ and χ^2_{m-1} are independent. A conditioning argument immediately shows that the RHS equals $P(||a||\tilde{Z}_1 < c) + O(n^{-\frac{1}{2}})$ and (IV.4.7) is proved.

That the error bound is sharp follows by direct calculation for $m = 1$ and hence in general. □

LEMMA IV.4.4. *Let* $f_m(x;\delta)$ *denote the density of the noncentral* $\chi^2_m(\delta)$ *distribution,* $m \geq 1$. *Then*

$$f_m(x;\delta) \leq \begin{cases} C_m x^{(m-2)/2} \exp\{-\frac{1}{2}(x^{\frac{1}{2}}-\delta^{\frac{1}{2}})^2\} & \text{for } x > 0,\ \delta > 0 \\ C_m x^{-\frac{1}{2}} \exp\{-\frac{1}{2}(x^{\frac{1}{2}}-\delta^{\frac{1}{2}})^2\} & \text{for } 0 < x < 4\delta \\ C_m x^{-\frac{1}{2}} (x/\delta)^{(m-1)/4} \exp\{-\frac{1}{2}(x^{\frac{1}{2}}-\delta^{\frac{1}{2}})^2\} & \text{for } 4m^2 < \delta < x \end{cases}$$

where the positive constants C_m *do not depend on* x *or* δ. *Conversely,*

$$f_m(x;\delta) > C_m \delta^{-\frac{1}{2}} \text{ for } |x^{\frac{1}{2}} - \delta^{\frac{1}{2}}| < 1,\ \delta > 4.$$

PROOF. All statements are trivial for $m = 1$; so assume $m \geq 2$. Let $v = v(x,\delta)$ be a real-valued function satisfying $0 \leq v \leq \frac{1}{2}x$, and let $b_m = (2^m\pi)^{-\frac{1}{2}}\Gamma((m-1)/2)^{-1}$. Then

$$
\begin{aligned}
f_m(x;\delta) &= \int_0^x f_{m-1}(x-y;0)\ f_1(y;\delta)\ dy \\
&\leq b_m \int_0^x y^{-\frac{1}{2}}(x-y)^{(m-3)/2}\exp(-\tfrac{1}{2}\delta - \tfrac{1}{2}x + \delta^{\frac{1}{2}}y^{\frac{1}{2}})\ dy \\
&\leq b_m\ \exp\{-\tfrac{1}{2}(x^{\frac{1}{2}}-\delta^{\frac{1}{2}})^2\}[\ \exp\{-\delta^{\frac{1}{2}}(x^{\frac{1}{2}}-(x-v)^{\frac{1}{2}})\} \int_0^{x-v} y^{-\frac{1}{2}}(x-y)^{(m-3)/2}dy \\
&\qquad + \int_{x-v}^{x} y^{-\frac{1}{2}}(x-y)^{(m-3)/2}\exp\{-\delta^{\frac{1}{2}}(x^{\frac{1}{2}}-y^{\frac{1}{2}})\}dy] \\
&\leq b_m\ \exp\{-\tfrac{1}{2}(x^{\frac{1}{2}}-\delta^{\frac{1}{2}})^2\}[\ \exp\{-\tfrac{1}{2}(\delta/x)^{\frac{1}{2}}v\} \int_0^{x} y^{-\frac{1}{2}}(x-y)^{(m-3)/2}dy \\
&\qquad + (x-v)^{-\frac{1}{2}} \int_0^{v} z^{(m-3)/2}\exp\{-\tfrac{1}{2}(\delta/x)^{\frac{1}{2}}z\}\ dz] \\
&\leq b_m\ \exp\{-\tfrac{1}{2}(x^{\frac{1}{2}}-\delta^{\frac{1}{2}})^2\}[\ \exp\{-\tfrac{1}{2}(\delta/x)^{\frac{1}{2}}v\}\ x^{(m-2)/2}B(\tfrac{1}{2},\ \tfrac{1}{2}(m-1)) \\
&\qquad + 2^{\frac{1}{2}}x^{-\frac{1}{2}}(x/\delta)^{(m-1)/4} \int_0^{(\delta/x)^{\frac{1}{2}}v} w^{(m-3)/2}e^{-\frac{1}{2}w}\ dw]\ .
\end{aligned}
$$

The first inequality follows by taking $v \equiv 0$. The second inequality follows from the first one if $x < 4m^2$; otherwise take $v(x,\delta) \equiv 2(m-1)\log x$ (bound the last integral by $2^{(m-1)/2}\Gamma(\frac{1}{2}(m-1))$). The third inequality is obtained by taking $v(x,\delta) \equiv (m-1)(x/\delta)^{\frac{1}{2}}\log x$ (bound the last integral as before).

To prove the reverse inequality, assume $\delta > 4$, let $|x^{\frac{1}{2}}-\delta^{\frac{1}{2}}| < 1$ and observe that

$$
\begin{aligned}
f_m(x;\delta) &> \tfrac{1}{2}\ b_m \int_0^x y^{-\frac{1}{2}}(x-y)^{(m-3)/2}\exp(-\tfrac{1}{2}\delta - \tfrac{1}{2}x + \delta^{\frac{1}{2}}y^{\frac{1}{2}})\ dy \\
&\geq \tfrac{1}{2}\ b_m\ \exp\{-\tfrac{1}{2}(x^{\frac{1}{2}}-\delta^{\frac{1}{2}})^2\}\ x^{-\frac{1}{2}} \int_{x-1}^{x} (x-y)^{(m-3)/2}\exp\{-\delta^{\frac{1}{2}}(x^{\frac{1}{2}}-y^{\frac{1}{2}})\}\ dy \\
&\geq \tfrac{1}{2}\ b_m\ \exp\{-\tfrac{1}{2}(x^{\frac{1}{2}}-\delta^{\frac{1}{2}})^2 - \delta^{\frac{1}{2}}x^{-\frac{1}{2}}\}\ (\delta^{\frac{1}{2}}+1)^{-1}2(m-1)^{-1} \\
&\geq \tfrac{1}{2}\ (m-1)^{-1}b_m\ e^{-5/2}\delta^{-\frac{1}{2}}.\ \square
\end{aligned}
$$

We are now prepared to prove our main theorem.

PROOF of Theorem IV.2.1.

i) Let E_n denote the set $\{y \in \mathbb{R}^k;\ \max_i |y_i| < \log n\}$. In view of (IV.4.1) $P_\pi(Y_n \in E_n) = 1 - o(n^{-\frac{1}{2}})$ uniformly for $\pi \in \Pi_\varepsilon$. By Corollary 17.2 in Bhattacharya and Ranga Rao (1976)

$$\sup_B |P_\pi(Y_n \in B) - P(U \in B)| = O(n^{-\frac{1}{2}})$$

where the supremum is taken over all Borel measurable convex sets $B \subset \mathbb{R}^k$ and where $U = (U_1,\ldots,U_k)'$ is distributed as in Corollary IV.4.2. The error bound is uniform in $\pi \in \Pi_\varepsilon$. Consider $\overline{CR}_n(\lambda)$ as a function of Y_n, $\overline{CR}_n(\lambda) = R^\lambda(Y_n)$. Since $\overline{CR}_n(\lambda)$ is a convex function of Y_n on E_n, and E_n itself is a convex set, it follows that

$$\sup |P_\pi(\overline{CR}_n(\lambda) < c, Y_n \in E_n) - P(R^\lambda(U) < c, U \in E_n)| = O(n^{-\frac{1}{2}})$$

or

$$\sup |P_\pi(\overline{CR}_n(\lambda) < c) - P(R^\lambda(U) < c)| = O(n^{-\frac{1}{2}}), \tag{IV.4.8}$$

where the supremum is taken over $\pi \in \Pi_\varepsilon$ and $c > 0$.

Conditionally on $Y_n \in E_n$, the terms in the expansion (IV.2.2) beyond the third power of Y_{in} are uniformly bounded by $\varepsilon_n = d_\lambda n^{-1} \log^4 n$, where d_λ is a suitable positive constant. This remains true after replacing Y_n by U. By Corollary IV.4.2 (with $q = 3$)

$$P(A^\lambda(U) + \tfrac{1}{3}n^{-\frac{1}{2}}(\lambda-1)\,\Sigma\, r_i^\lambda \pi_i^{-\frac{1}{2}} U_i^3 \pm \varepsilon_n < c) = P(A^\lambda(U) < c) + O(n^{-\frac{1}{2}}).$$

Combining this result with (IV.4.8), i) is established.

ii) By (IV.2.3)

$$A^\lambda(U) = B^\lambda(U) + \Sigma\, (r_i^\lambda - \bar{r}_\lambda) U_i^2$$

where $\bar{r}_\lambda \to 1$ and $r_i^\lambda - \bar{r}_\lambda = O(\max_i |\pi_i - p_i|)$ as $\max_i |\pi_i - p_i| \to 0$. By Corollary IV.4.2 (with $q = 2$)

$$P(A^\lambda(U) < c) = P(B^\lambda(U) < c) + O(\max_i |\pi_i - p_i|).$$

The desired result now follows from part i). Note that this argument remains valid if one takes $\bar{r}_\lambda = 1$.

(iii) Let

$$\tilde{r}_\lambda = [\Sigma\, \pi_i (r_i^\lambda - 1)^2 - \{\Sigma\, \pi_i (r_i^\lambda - 1)\}^2]^{\frac{1}{2}} \{2\lambda^2 I^\lambda(\pi:p)\}^{-\frac{1}{2}}. \tag{IV.4.9}$$

Define $\tilde{B}^\lambda$ by (IV.2.4) with $\tilde{r}_\lambda$ replacing $\bar{r}_\lambda$. Obviously

$$A^\lambda(U) = \tilde{B}^\lambda(U) + \Sigma(r_i^\lambda - \tilde{r}_\lambda)U_i^2,$$

where $\tilde{r}_\lambda = 1+O(\max_i|\pi_i-p_i|)$ and $r_i^\lambda-\tilde{r}_\lambda = O(\max_i|\pi_i-p_i|)$. Similarly to (IV.2.10) and (IV.2.11) $\tilde{B}^\lambda(U)$ is distributed as

$$\tilde{r}_\lambda\ \chi^2_{k-1}(\delta_{n\lambda}) + (1-\tilde{r}_\lambda)\delta_{n\lambda}.$$

By Corollary IV.4.2 (with q = 2)

$$P_\pi(A^\lambda(U) < c) = P(\tilde{B}^\lambda(U) < c) + O(\max_i|\pi_i-p_i|).$$

Moreover, in the notation of Lemma IV.4.4,

$$\text{(IV.4.10)}\qquad P_\pi(\tilde{B}^\lambda(U) < c) = P(\tilde{r}_\lambda\chi^2_{k-1}(\delta_{n\lambda}) + (1-\tilde{r}_\lambda)\delta_{n\lambda} < c)$$

$$= P(\chi^2_{k-1}(\delta_{n\lambda}) < c + (c-\delta_{n\lambda})(1-\tilde{r}_\lambda)/\tilde{r}_\lambda)$$

$$= P(\chi^2_{k-1}(\delta_{n\lambda}) < c) + (c-\delta_{n\lambda})(1-\tilde{r}_\lambda)\tilde{r}_\lambda^{-1}\ f_{k-1}(\theta_n;\delta_{n\lambda})$$

where $\theta_n = c + t_n(c-\delta_{n\lambda})(1-\tilde{r}_\lambda)/\tilde{r}_\lambda$, $0 \le t_n \le 1$. The first part of Lemma IV.4.4 implies

$$f_{k-1}(\theta_n;\delta_{n\lambda}) \le \begin{cases} C\ \theta_n^{-\frac{1}{2}}\exp(-\frac{1}{2}(\theta_n^{\frac{1}{2}}-\delta_{n\lambda}^{\frac{1}{2}})^2) \text{ if } |\theta_n^{\frac{1}{2}}-\delta_{n\lambda}^{\frac{1}{2}}| < \frac{1}{2}\delta_{n\lambda}^{\frac{1}{2}}. \\ C\ \exp\{-\frac{1}{4}(\theta_n^{\frac{1}{2}}-\delta_{n\lambda}^{\frac{1}{2}})^2\} \quad \text{otherwise.} \end{cases}$$

Since $c - \delta_{n\lambda} = (\theta_n-\delta_{n\lambda})(1+o(1))$, it follows that

$$(c-\delta_{n\lambda})f_{k-1}(\theta_n;\delta_{n\lambda}) = O(1).$$

Hence the last term in the RHS of (IV.4.10) is of order $O(s_n n^{-\frac{1}{2}})$ and the desired result follows from part (i).

(iv) Define $\tilde{r}_{\lambda 1}$ as $\tilde{r}_\lambda$ in (IV.4.9) with $I^\lambda(\pi:p)$ replaced by $I^1(\pi:p)$ and define $\tilde{B}_1^\lambda(U)$ as $\tilde{B}^\lambda(U)$ with $\tilde{r}_\lambda$ replaced by $\tilde{r}_{\lambda 1}$. Again

$$A^\lambda(U) = \tilde{B}_1^\lambda(U) + \Sigma(r_i^\lambda - \tilde{r}_{\lambda 1})U_i^2$$

and

$$\tilde{B}_1^\lambda(U) \sim \tilde{r}_{\lambda 1}\ \chi^2_{k-1}(\delta_n) + (1-\tilde{r}_{\lambda 1})\delta_n + \delta_{n\lambda} - \delta_n$$

with $\tilde{r}_{\lambda 1} = 1+O(\max_i|\pi_i-p_i|)$ and $r_i^\lambda-\tilde{r}_{\lambda 1} = O(\max_i|\pi_i-p_i|)$.

Proceeding as in (iii)

$$P(\tilde{r}_{\lambda 1}\ \chi^2_{k-1}(\delta_n) + (1-\tilde{r}_{\lambda 1})\delta_n + \delta_{n\lambda} - \delta_n < c)$$

$$= P(\chi^2_{k-1}(\delta_n) + \delta_{n\lambda}-\delta_n < c) + O(s_n n^{-\frac{1}{2}}).$$

By a local expansion

$$\text{(IV.4.11)} \qquad \delta_{n\lambda}-\delta_n = \tfrac{1}{3}(\lambda-1)n\ \Sigma\ (\pi_i-p_i)^3/p_i^2 + O(n\ \max_i|\pi_i-p_i|^4),$$

implying $(\delta_{n\lambda}-\delta_n)/\delta_n^{\frac{1}{2}} = O(s_n^2\ n^{-\frac{1}{2}})$. Since by Lemma IV.4.4 $f_{k-1}(x;\delta) \leq C\ \delta^{-\frac{1}{2}}$ (all $x > 0$), another application of the mean value theorem yields

$$P_\pi(\tilde{B}_1^\lambda(U) < c) - P(\chi^2_{k-1}(\delta_n) < c) = O(s_n^2\ n^{-\frac{1}{2}}).$$

The desired result follows again from part (i).

To prove that the bound in (iv) is sharp, it suffices to show that for given $\{s_n\}$

$$P(\chi^2_{k-1}(\delta_n) + \delta_{n\lambda}-\delta_n < \delta_n) - P(\chi^2_{k-1}(\delta_n) < \delta_n) > \varepsilon\ s_n^2\ n^{-\frac{1}{2}}$$

for some $\varepsilon > 0$ and appropriate $\pi_1,\ldots,\pi_k$. First note that $\pi_1,\ldots,\pi_k$ exist such that both δ_n/s_n^2 and $|\delta_{n\lambda}-\delta_n|/(s_n^3\ n^{-\frac{1}{2}})$ are bounded away from 0 and ∞, cf. (IV.4.11). Since

$$|\delta_n - (\delta_{n\lambda}-\delta_n)|^{\frac{1}{2}}-\delta_n^{\frac{1}{2}} = o(1) \quad \text{as } s_n^2\ n^{-\frac{1}{2}} \to 0,$$

the second part of Lemma IV.4.4 and the mean value theorem imply the above inequality.

(v) By (IV.2.3), (IV.2.5) and (IV.2.7)

$$\text{(IV.4.12)} \quad P(A^{\lambda}(U) < c) = P(L^{\lambda}(U) + \Sigma\, r_i^{\lambda}\, U_i^2 < c)$$

$$= P(2 \sum_{i=1}^{k-1} \theta_{in}^{\frac{1}{2}} \omega_{in} Z_i + 2nI^{\lambda}(\pi{:}p) + \sum_{i=1}^{k-1} \theta_{in} Z_i^2 < c).$$

The θ_{in} $(i=1,\ldots,k-1)$ are bounded away from 0 and ∞ and the first k-1 components of $\omega_n = S_n'Q_n\mu_n$ (see subsection IV.2.1) satisfy $\max_i \omega_{in}^2 > \varepsilon_1 n \max_i (r_i^{\lambda/2} - r_i^{-\lambda/2})^2/\lambda^2$ for some $\varepsilon_1 > 0$. Hence, dividing both members in the last event of (IV.4.12) by $n^{\frac{1}{2}}\max_i |r_i^{\lambda/2} - r_i^{-\lambda/2}|/|\lambda|$, Lemma IV.4.3 implies that the RHS of (IV.4.12) equals

$$P(2 \sum_{i=1}^{k-1} \theta_{in}^{\frac{1}{2}} \omega_{in} Z_i + 2nI^{\lambda}(\pi{:}p) < c) + O((n^{\frac{1}{2}}\max_i |\pi_i - p_i|)^{-1})$$

where the remainder term is $O(s_n^{-1})$. The desired result now follows from part (i). That the bound is sharp is an easy exercise. □

REFERENCES

[1] BAHADUR, R.R. (1966), "A note on quantiles in large samples", *Ann. Math. Statist.* 37, 577-580.

[2] BHATTACHARYA, R.N. and RANGA RAO, R. (1976), *Normal Approximation and Asymptotic Expansions*, Wiley, New York.

[3] BICKEL, P.J. (1982), "On adaptive estimation", *Ann. Statist.* 10, 647-671.

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

[5] BILLINGSLEY, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

[6] BOFINGER, E. (1973), "Goodness-of-fit test using sample quantiles", *J.R. Statist. Soc., Ser. B* 35, 277-284.

[7] BROFFITT, J.D. and RANDLES, R.H. (1977), "A power approximation for the chi-square goodness-of-fit test: simple hypothesis case", *J. Amer. Statist. Assoc.* 72, 604-607.

[8] CHERNOFF, H. and LEHMANN, E.L. (1954), "The use of maximum likelihood estimators in χ^2 tests for goodness of fit", *Ann. Math. Statist.* 25, 579-586.

[9] COX, D.R. and REID, N. (1986), "Approximations to noncentral distributions", *Canad. J. Statist.*, to appear.

[10] CRESSIE, N. and READ, T.R.C. (1984), "Multinomial goodness-of-fit tests", *J.R. Statist. Soc., Ser B* 46, 440-464.

[11] CSÖRGÖ, M. and RÉVÉSZ, P. (1981a), *Strong Approximations in Probability and Statistics*, Academic Press, New York.

[12] CSÖRGÖ, M. and RÉVÉSZ, P. (1981b), "Quantile processes and sums of weighted spacings for composite goodness-of-fit", in CSÖRGÖ, M. et al. (ed's), *Statistics and related topics*, North Holland, Amsterdam.

[13] DAHIYA, R.C. and GURLAND, J. (1972), "Pearson chi-squared test of fit with random intervals", *Biometrika* 59, 147-153.

[14] DAVIES, R.B. (1980), "Algorithm AS 155. The distribution of a linear combination of χ^2 random variables", *Applied Statistics* 29, 366-369.

[15] DIJKSTRA, J.B., RIETJENS, T.J.M. and STEUTEL, F.W. (1984), "A simple test for uniformity", *Statistica Neerlandica* 38, 33-44.

[16] DIK, J.J. and DE GUNST, M.C.M. (1985), "The distribution of general quadratic forms in normal variables", *Statistica Neerlandica* 39, 14-26.

[17] DROST, F.C. (1987), "A limit theorem for some modified chi-square statistics when the number of classes increases", in: BAUER, P. et al. (ed's), *Mathematical Statistics and Probability Theory, Vol. B*, D. Reidel Publishing Company, Dordrecht, 43 - 58.

[18] DROST, F.C., KALLENBERG, W.C.M. and OOSTERHOFF, J. (1985), "Estimating nuisance parameters in random-cell chi-square type goodness-of-fit tests", Technical report 303, Dept. of Mathematics and Computer Science, Free University, Amsterdam.

[19] DURBIN, J. (1973), "Weak convergence of the sample distribution function when parameters are estimated", *Ann. Statist.* 1, 279-290.

[20] DZHAPARIDZE, K.O. and NIKULIN, M.S. (1974), "On a modification of the standard statistics of Pearson", *Theor. Probability Appl.* 19, 851-853.

[21] ESSÉEN, C.G. (1945), "Fourier analysis of distribution functions", *Acta Math.* 77, 1-125.

[22] FAREBROTHER, R.W. (1984), "A remark on algorithms AS 106, AS 153 and AS 155. The distribution of a linear combination of χ^2 random variables", *Applied Statistics* 33, 366-369.

[23] FISHER, R.A. (1924), "The conditions under which χ^2 measures the discrepancy between observation and hypothesis", *J.R. Statist. Soc.* 87, 442-450.

[24] GREEN, J.R. and HEGAZY, Y.A.S. (1976), "Powerful modified-EDF goodness-of-fit tests", *J. Americ. Statist. Assoc.* 71, 204-209.

[25] HABER, M. (1980), "A comparative simulation study of the small sample powers of several goodness of fit tests", *J. Statist. Computation and Simulation* 11, 241-251.

[26] HÁJEK, J. (1972), "Local asymptotic minimax and admissibility in estimation", *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* 1, 175-194, University of California Press.

[27] HOEFFDING, W. (1963), "Probability inequalities for sums of bounded random variables", *J. Amer. Statist. Assoc.* 58, 13-30.

[28] HSUAN, T.A. (1974), "Some chi-square statistics for goodness-of-fit tests", Ph.D. thesis, Cornell University, Ithaca, New York.

[29] HUBER, P.J. (1981), *Robust Statistics*, Wiley, New York.

[30] KALLENBERG, W.C.M. (1985), "On moderate and large deviations in multinomial distributions", *Ann. Statist.* 13, 1554-1580.

[31] 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.

[32] KOTZ, S., JOHNSON, N.L. and BOYD, D.W. (1967a), "Series representations of distributions of quadratic forms in normal variables. I. Central case", *Ann. Math. Statist.* 38, 823-837.

[33] KOTZ, S., JOHNSON, N.L. and BOYD, D.W. (1967b), "Series representations of distributions of quadratic forms in normal variables. II. Non-central case", *Ann. Math. Statist.* 38, 838-848.

[34] LARNTZ, K. (1978), "Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics", *J. Amer. Statist. Assoc.* 73, 253-263.

[35] LECAM, L., MAHAN, C.M. and SINGH, A. (1983), "An extension of a theorem of H. Chernoff and E.L. Lehmann", in: RIZVI, M.H., RUSTAGI, J.S. and SIEGMUND, D. (ed's), *Recent Advances in Statistics*, Academic Press, New York, 303-337.

[36] MANN, H.B. and WALD, A. (1942), "On the choice of the number of intervals in the application of the chi-square test", *Ann. Math. Statist.* 13, 306-317.

[37] McCULLOCH, C.E. (1985), "Relationships among some chi-square goodness of fit statistics", *Commun. Statist. - Theor. Meth.* 14, 593-603.

[38] 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.

[39] MOORE, D.S. (1984), "Measures of lack of fit from tests of chi-squared type", *J. Statist. Planning and Inf.* 10, 151-166.

[40] 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.

[41] MORRIS, C. (1975), "Central limit theorems for multinomial sums", *Ann. Statist.* 3, 165-188.

[42] NEUHAUS, G. (1976), "Weak convergence under contiguous alternatives of the empirical process when parameters are estimated: the D_k approach", in GAENSSLER, P. and RÉVÉSZ, P. (ed's), *Empirical Distributions and Processes*, Springer Verlag, 68-82.

[43] NIKULIN, M.S. (1973), "Chi-square test for continuous distributions with shift and scale parameters", *Theor. of Probability Appl.* 3, 599-616.

[44] OOSTERHOFF, J. (1985), "The choice of cells in chi-square tests", *Statistica Neerlandica* 39, 115-128.

[45] OOSTERHOFF, J. and VAN ZWET, W.R. (1979), "A note on contiguity and Hellinger distance", in: JURECKOVÁ, J. (ed.), *Contributions to Statistics, J. Hájek Memorial Volume*, Academia, Prague, 157-166.

[46] PATNAIK, P.B. (1949), "The non-central χ^2- and F-distributions and their applications", *Biometrika* 36, 202-232.

[47] PEARSON, K. (1900), "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably to be supposed to have arisen from random sampling", *Philosophical Magazine, Ser* 5 50, 157-175. Reprinted in K. Pearson (1956), *Karl Pearson's Early Statistical Papers*, Cambridge University Press, 339-357.

[48] QUINE, M.P. and ROBINSON, J. (1985), "Efficiencies of chi-square and likelihood ratio goodness-of-fit tests", *Ann. Statist.* 13, 727-742.

[49] 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.

[50] READ, T.R.C. (1984), "Small-sample comparisons for the power divergence goodness-of-fit statistics", *J. Amer. Statist. Assoc.* 63, 929-935.

[51] ROY, A.R. (1956), "On χ^2 statistics with variable intervals", Technical report No. 1, Standford Univ. Dept. of Statistics.

[52] RUYMGAART, F.H. (1974), "Asymptotic normality of nonparametric tests for independence", *Ann. Statist.* 2, 892-910.

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

[54] SHEIL, J. and O' MUIRCHEARTAIGH, I. (1977), "Algorithm AS 106. The distribution of non-negative quadratic forms in normal variables", *Applied Statistics* 26, 92-98.

[55] SHIRAHATA, S. (1976), "Tests for independence in finite contingency tables", *Ann. Statist.* 4, 542-553.

[56] SHORACK, G.R. (1979), "The weighted empirical process of row independent random variables with arbitrary distribution functions", *Statistica Neerlandica* 33, 169-189.

[57] SLAKTER, M.J. (1968), "Accuracy of an approximation to the power of the chi-square goodness of fit tests with small but equal expected frequencies", *J. Amer. Statist. Assoc.* 63, 912-918.

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

[59] STECK, G.P. (1957), "Limit theorems for conditional distributions", *Univ. California Publ. Statist.* 2 (12), 237-284.

[60] STEPHENS, M.A. (1974), "EDF statistics for goodness of fit and some comparisons", *J. Amer. Statist. Assoc.* 69, 730-737.

[61] THOMAS, D.R. and PIERCE, D.A. (1979), "Neyman's smooth goodness-of-fit test when the hypothesis is composite", *J. Amer. Statist. Assoc.* 74, 441-445.

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

[63] VAJDA, I. (1973), "χ^{α}-divergence and generalized Fisher information", *Trans. 6-th Prague Conf. on Inf. Theory, Random Proc., Stat. Dec. Functions* 1971, Academia, Prague, 872-886.

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

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

[66] WEST, E.N. and KEMPTHORNE, O. (1971), "A comparison of the chi and the likelihood ratio tests for composite alternatives", *J. of Computation and Simulation* 1, 1-33.

[67] WITTING, H. (1959), "Über einen χ^2-Test, dessen Klassen durch geordnete Stichprobenfunktionen festgelegt werden", *Arch. Math.* 10, 468-479.

[68] YARNOLD, J.K. (1972), "Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set", *Ann. Statist.* 43, 1566-1580.

INDEX

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print; replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservation of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statement.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Emperical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallman spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *Go-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewal theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequences.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.

2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.

3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.

4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.

5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.

6 A.P.W. Böhm. *Dataflow computation.* 1984.

7 A. Blokhuis. *Few-distance sets.* 1984.

8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.

9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.

10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.

11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.

12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.

13 D. Grune. *On the design of ALEPH.* 1985.

14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.

15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.

16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.

17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.

18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.

19 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 1: Philosophy, framework, computer science.* 1986.

20 B.F. Schriever. *Order dependence.* 1986.

21 D.P. van der Vecht. *Inequalities for stopped Brownian motion.* 1986.

22 J.C.S.P. van der Woude. *Topological dynamix.* 1986.

23 A.F. Monna. *Methods, concepts and ideas in mathematics: aspects of an evolution.* 1986.

24 J.C.M. Baeten. *Filters and ultrafilters over definable subsets of admissible ordinals.* 1986.

25 A.W.J. Kolen. *Tree network and planar rectilinear location theory.* 1986.

26 A.H. Veen. *The misconstrued semicolon: Reconciling imperative languages and dataflow machines.* 1986.

27 A.J.M. van Engelen. *Homogeneous zero-dimensional absolute Borel sets.* 1986.

28 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 2: Applications to natural language.* 1986.

29 H.L. Trentelman. *Almost invariant subspaces and high gain feedback.* 1986.

30 A.G. de Kok. *Production-inventory control models: approximations and algorithms.* 1987.

31 E.E.M. van Berkum. *Optimal paired comparison designs for factorial experiments.* 1987.

32 J.H.J. Einmahl. *Multivariate empirical processes.* 1987.

33 O.J. Vrieze. *Stochastic games with finite state and action spaces.* 1987.

34 P.H.M. Kersten. *Infinitesimal symmetries: a computational approach.* 1987.

35 M.L. Eaton. *Lectures on topics in probability inequalities.* 1987.

36 A.H.P. van der Burgh, R.M.M. Mattheij (eds.). *Proceedings of the first international conference on industrial and applied mathematics (ICIAM 87).* 1987.

37 L. Stougie. *Design and analysis of algorithms for stochastic integer programming.* 1987.

38 J.B.G. Frenk. *On Banach algebras, renewal measures and regenerative processes.* 1987.

39 H.J.M. Peters, O.J. Vrieze (eds.). *Surveys in game theory and related topics.* 1987.

40 J.L. Geluk, L. de Haan. *Regular variation, extensions and Tauberian theorems.* 1987.

41 Sape J. Mullender (ed.). *The Amoeba distributed operating system: Selected papers 1984-1987.* 1987.

42 P.R.J. Asveld, A. Nijholt (eds.). *Essays on concepts, formalisms, and tools.* 1987.

43 H.L. Bodlaender. *Distributed computing: structure and complexity.* 1987.

44 A.W. van der Vaart. *Statistical estimation in large parameter spaces.* 1988.

45 S.A. van de Geer. *Regression analysis and empirical processes.* 1988.

46 S.P. Spekreijse. *Multigrid solution of the steady Euler equations.* 1988.

47 J.B. Dijkstra. *Analysis of means in some non-standard situations.* 1988.

48 F.C. Drost. *Asymptotics for generalized chi-square goodness-of-fit tests.* 1988.