Metrika (2000) 52: 237–252

Metrika

Diagnostic smooth tests of fit

Bernhard Klar

Institut für Math. Stochastik, Universität Karlsruhe, Englerstr. 2, D-76128 Karlsruhe, Germany.
email: Bernhard.Klar@math.uni-karlsruhe.de

Received: September 1999

Abstract. Smooth tests are frequently used for testing the goodness of fit of a parametric family of distributions. One reason for the popularity of the smooth tests are the diagnostic properties commonly attributed to them. In recent years, however, it has been realized that these tests are strictly non-diagnostic when used conventionally. The paper examines how the smooth test statistics must be rescaled in order to obtain procedures having diagnostic properties at least for large sample sizes.

Key words: Smooth test, goodness of fit, components, exponential families, bivariate Poisson distribution

1 Introduction

Smooth tests, originally introduced by Jerzy Neyman (1937) to assess the goodness of fit for the uniform distribution, have been extended by several authors to test the hypothesis that an unknown distribution belongs to a parametric family $\{P_\vartheta\}$ (Thomas and Pierce (1979), Bargal and Thomas (1983), Rayner and Best (1986, 1988, 1989, 1995), Koziol (1986, 1987)). Smooth test statistics usually have the form

$$\hat{\Psi}^2_{n,k} = \sum_{j=s+1}^{s+k} \hat{U}^2_{n,j} \quad \text{with} \quad \hat{U}_{n,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j(X_i; \hat{\vartheta}_n). \tag{1}$$

Here, $X_1, X_2, \ldots, X_n$ are independent observations on a random variable X with unknown distribution P_ϑ, $\hat{\vartheta}_n$ is a suitable estimator of the s-dimensional parameter (vector) ϑ, and $\{h_0(\cdot\,;\vartheta) \equiv 1, h_1(\cdot\,;\vartheta), h_2(\cdot\,;\vartheta), \ldots\}$ are orthonormal polynomials with respect to P_ϑ, that is,

$$\int h_j(\cdot\,;\vartheta)h_l(\cdot\,;\vartheta)\,dP_\vartheta = \delta_{jl} \quad (0 \le j, l \le s+k), \tag{2}$$

where δ_{jl} denotes Kronecker's delta. Since the polynomial $h_j(\cdot\,;\vartheta)$ is of degree j ($j \ge 1$), the $\hat{U}_{n,j}$ are called components of degree j. Note, however, that Thomas and Pierce (1979) and Bargal and Thomas (1983) used powers of the distribution function instead of orthonormal functions.

Putting $l = 0$ in (2) yields

$$\int h_j(\cdot\,;\vartheta)\,dP_\vartheta = E_\vartheta[h_j(X;\vartheta)] = 0$$

for $j \ge 1$. This equation describes a relation between the first j moments of X which is valid under P_ϑ.

In case of the Poisson distribution with parameter ϑ, the orthogonal system is given by the Poisson-Charlier polynomials (see Chihara (1978), Section VI.1); in particular,

$$h_1(x,\vartheta) = \frac{x-\vartheta}{\sqrt{\vartheta}}, \quad h_2(x,\vartheta) = \frac{(x-\vartheta)^2 - x}{\sqrt{2}\vartheta}.$$

For $j = 1$, the equation $E_\vartheta[h_j(X;\vartheta)] = 0$ means that the expected value of the Poisson distribution equals ϑ, whereas, for $j = 2$, it stands for the equality of expectation and variance. Since, putting $\hat{\vartheta}_n = \overline{X}_n = n^{-1}\sum_{j=1}^n X_j$,

$$\sum_{i=1}^n h_1(X_i,\hat{\vartheta}_n) = \frac{1}{\sqrt{\hat{\vartheta}_n}}\left(\sum_{i=1}^n X_i - n\overline{X}_n\right) = 0,$$

the first component $\hat{U}_{n,1}$ is zero; hence $s = 1$ in (1). The reason for this is that the usual estimator of ϑ is the method of moments estimator of the mean. Furthermore,

$$\hat{\Psi}^2_{n,1} = \hat{U}^2_{n,2} = \frac{1}{2n}(D_n - n)^2,$$

where $D_n = \sum_{j=1}^n (X_j - \overline{X}_n)^2/\overline{X}_n$ denotes Fisher's index of dispersion. Thus, the smooth test for Poissonity based on the first nonzero component $\hat{U}_{n,2}$ and the dispersion test are equivalent.

The diagnostic properties commonly attributed to the smooth tests are one reason for their popularity: it is assumed that such tests, in case of rejection of a hypothesis H_0, provide some kind of directed diagnosis regarding the kind of departure from H_0 of the underlying distribution. The dispersion test, for example, examines the ratio of variance to expected value, which equals one in case of the Poisson distribution. If the ratio of empirical variance to arithmetic mean is not sufficiently close to one, the hypothesis of Poissonity is rejected (on a given significance level). Furthermore, it is believed that the hypothetical model has been rejected because the variance of the underlying true distribution differs from its expectation. Henze and Klar (1996) showed that the conventional use of the dispersion test does not qualify to this diagnosis since, on one hand, the conclusion refers to a nonparametric population parameter but, on the other hand, it is drawn from the distributional behaviour of the test statistic under the 'narrow' parametric hypothesis of Poissonity. In Henze (1997) and Henze and Klar (1996), it was also shown how to modify this and similar tests to obtain procedures having diagnostic properties at least for large sample sizes. The correct approach is to formulate a nonparametric hypothesis that is tailored to the desired kind of directed diagnosis.

In case of the dispersion test, this hypothesis is the equality of expected value and variance. The rejection of this nonparametric hypothesis on a certain significance level leads to the conclusion that the variance of the underlying distribution differs from its expected value.

It is the aim of this paper to generalize these results to smooth tests of arbitrary order. To this end, the joint limit distribution of components is derived in Section 2 not only under the parametric hypothesis, but also under a suitable nonparametric class of distributions.

Under the parametric hypothesis, and if the maximum likelihood estimation method is used, the usual representation of the asymptotic distribution of components as limit law of score statistics is regained.

In the nonparametric class of distributions, however, the appropriate estimation method is the method of moments, and we show how the components have to be rescaled in order to obtain test statistics which are asymptotically distribution-free within this class.

Section 3 treats smooth tests for certain exponential families for which rescaling is particularly easy. This applies to all cases treated in Rayner and Best (1989).

Since the assertion about the limiting distribution of components in Section 2 remains valid without the assumption that the polynomials h_j are orthogonal, diagnostic smooth tests can be constructed even when no orthogonal system exists. As an example, smooth tests of fit for bivariate Poissonity are discussed in Section 4.

2 The joint limit distribution of components

Let $\{P_\vartheta : \vartheta \in \Theta\}$ be a parametric class of probability measures on the Borel sets of $\mathbb{R}$, where Θ is an open subset of $\mathbb{R}^s$. Suppose that the distributions P_ϑ have common support and that they are distinct, i.e. the mapping $\vartheta \to P_\vartheta$ is one-to-one. Assume further that

$$\int_{-\infty}^{\infty} |x|^k \, dF_\vartheta(x) < \infty \quad (k = 1, 2, \ldots),$$

where $F_\vartheta : \mathbb{R} \to [0, 1]$ denotes the distribution function of P_ϑ. Then, the kth moment is given by

$$\mu_k = \int_{-\infty}^{\infty} x^k \, dF_\vartheta(x) \quad (k = 1, 2, \ldots).$$

If the set $\mathscr{S}(F_\vartheta) = \{x \mid F_\vartheta(x+\varepsilon) - F_\vartheta(x-\varepsilon) > 0 \text{ for each } \varepsilon > 0\}$ of growth points of F_ϑ is infinite, there is a unique system of orthonormal polynomials $\{h_0(\cdot\,;\vartheta) \equiv 1, h_1(\cdot\,;\vartheta), h_2(\cdot\,;\vartheta), \ldots\}$ with respect to P_ϑ (see, e.g., Chihara (1978), Theorem I.3.3 and the corollary after Theorem I.2.2).

Suppose $X, X_1, \ldots, X_n$ are independent observations on a random variable X with (unknown) distribution P. For testing the composite hypothesis

$$H_0 : P \in \{P_\vartheta : \vartheta \in \Theta\},$$

a smooth test of order k pertaining to $\{P_\vartheta\}$ can be used. This test rejects H_0 for large values of

$$\hat{\Psi}^2_{n,k} = \sum_{j=k_0+1}^{k_0+k} \hat{U}^2_{n,j}, \tag{3}$$

where $0 \le k_0 \le s$ and

$$\hat{U}_{n,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j(X_i; \hat{\vartheta}_n).$$

The value of k_0 in (3) depends on the estimation method: it is the largest integer for which $\hat{U}_{n,1} = \cdots = \hat{U}_{n,k_0} = 0$. In many examples, k_0 equals the dimension s of the parameter vector. The estimator $\hat{\vartheta}_n$ of ϑ is assumed to satisfy the following regularity condition under H_0:

(R1) There exists a measurable function $l : \mathbb{R} \times \Theta \to \mathbb{R}^s$ such that

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l(X_i; \vartheta) + o_{P_\vartheta}(1),$$

where $E_\vartheta[l(X;\vartheta)] = 0$ and $E_\vartheta[l^2(X;\vartheta)] < \infty$.

The first result establishes the joint asymptotic distribution of the random vector $\hat{U}_n = (\hat{U}_{n,k_0+1}, \ldots, \hat{U}_{n,k_0+k})'$ under H_0, where the prime denotes transpose. Thereby, $h(x;\vartheta) = (h_{k_0+1}(x;\vartheta), \ldots, h_{k_0+k}(x;\vartheta))'$, and $\nabla_\vartheta h(x;\vartheta)$ is the $(k \times s)$-matrix with entries $\partial h_i(x;\vartheta)/\partial\vartheta_j$, where $\vartheta = (\vartheta_1, \ldots, \vartheta_s)'$.

2.1 Theorem. a) *Suppose that the coefficients $a_{ji}(\vartheta)$ in $h_j(x;\vartheta) = \sum_{i=0}^{j} a_{ji}(\vartheta)x^i$, $k_0 + 1 \le j \le k_0 + k$, have a continuous derivative with respect to ϑ. Let $\hat{\vartheta}_n$ satisfy (R1). Then the limiting distribution of $\hat{U}_n$ under P_ϑ is k-variate normal $N(0, \Sigma)$ with covariance matrix*

$$\Sigma = E_\vartheta[v(X;\vartheta)v(X;\vartheta)'],$$

where $v(x;\vartheta) = h(x;\vartheta) + E_\vartheta[\nabla_\vartheta h(X;\vartheta)]l(x;\vartheta)$.

b) *Additionally, suppose that P_ϑ has a density $f(\cdot\,;\vartheta)$ with respect to a σ-finite measure μ, and that, for fixed $x \in \mathbb{R}$, the derivative of $f(x;\vartheta)$ with respect to ϑ is continuous. Furthermore, assume that, for $k_0 + 1 \le j \le k_0 + k$, the integral of $h_j(\cdot\,;\vartheta)f(\cdot\,;\vartheta)$ with respect to μ may be differentiated under the integral. Then*

$$E_\vartheta[\nabla_\vartheta h(X;\vartheta)] = -C_\vartheta,$$

where C_ϑ is the $(k \times s)$-matrix with entries

$$c_{ij} = E_\vartheta\left[h_i(X;\vartheta)\frac{\partial \log f(X;\vartheta)}{\partial \vartheta_j}\right].$$

c) *If $\hat{\vartheta}_n$ is the maximum likelihood estimator (MLE) of ϑ, then, under the assumptions of b),*

$$\Sigma = I_k - C_\vartheta[I(\vartheta)]^{-1}C'_\vartheta,$$

where I_k denotes the identity matrix of order k, and $I(\vartheta)$ is the Fisher information matrix with entries

$$I_{jk}(\vartheta) = E_\vartheta\left[\left(\frac{\partial}{\partial \vartheta_j}\log f(X;\vartheta)\right) \times \left(\frac{\partial}{\partial \vartheta_k}\log f(X;\vartheta)\right)\right].$$

Proof: Let $U_{n,j}(\gamma) = (1/\sqrt{n})\sum_{i=1}^n h_j(X_i;\gamma)$ so that $\hat{U}_{n,j} = U_{n,j}(\hat{\vartheta}_n)$. By the mean value theorem,

$$U_{n,j}(\hat{\vartheta}_n) = U_{n,j}(\vartheta) + (\hat{\vartheta}_n - \vartheta)'[\nabla_\gamma U_{n,j}(\gamma)|_{\gamma=\vartheta^*}],$$

where ϑ^* is between ϑ and $\hat{\vartheta}_n$. Using Sluzky's lemma and (R1),

$$\begin{aligned} U_{n,j}(\hat{\vartheta}_n) &= U_{n,j}(\vartheta) + \sqrt{n}(\hat{\vartheta}_n - \vartheta)'E[\nabla_\gamma h_j(X;\gamma)|_{\gamma=\vartheta}] + o_{P_\vartheta}(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n [h_j(X_i;\vartheta) + l(X_i;\vartheta)'E[\nabla_\gamma h_j(X;\gamma)|_{\gamma=\vartheta}] + o_{P_\vartheta}(1). \end{aligned}$$

Since $h_j(X;\vartheta)$ and $l(X;\vartheta)$ are centered under P_ϑ, the central limit theorem yields assertion a). Under the standing assumption, b) follows from

$$\begin{aligned} 0 &= \frac{\partial}{\partial \vartheta_j}[E_\vartheta(h_i(X;\vartheta))] = \frac{\partial}{\partial \vartheta_j}\int h_i(x;\vartheta)f(x,\vartheta)\mu(dx) \\ &= \int \frac{\partial(h_i(x;\vartheta)f(x,\vartheta))}{\partial \vartheta_j}\mu(dx) \\ &= \int \frac{\partial h_i(x;\vartheta)}{\partial \vartheta_j}f(x,\vartheta)\mu(dx) + \int h_i(x;\vartheta)\frac{\partial \log f(x;\vartheta)}{\partial \vartheta_j}f(x,\vartheta)\mu(dx) \\ &= E_\vartheta\left[\frac{\partial h_i(X;\vartheta)}{\partial \vartheta_j}\right] + E_\vartheta\left[h_i(X;\vartheta)\frac{\partial \log f(X;\vartheta)}{\partial \vartheta_j}\right]. \end{aligned}$$

If a maximum likelihood estimator satisfies (R1), the function $l(\cdot\,;\vartheta)$ takes the form $l(x;\vartheta) = [I(\vartheta)]^{-1}\cdot\nabla_\vartheta \log f(x;\vartheta)$. Plugging this and the matrix C_ϑ from b) into the covariance matrix Σ, and noting that $E_\vartheta[h(X;\vartheta)h(X;\vartheta)'] = I_k$ by orthonormality of the h_j, assertion c) follows. ■

Remarks:

1. Part a) of Theorem 2.1 is the multivariate version of the assertion in Henze and Klar (1996) concerning a single component. The proof shows that parts a) and b) remain valid without the assumption that the polynomials h_j are orthogonal. Moreover, the X_i may be multivariate random vectors as well (see Section 4).
2. Smooth tests of fit are commonly introduced as score tests which are asymptotically optimal against certain parametric alternatives, similar to the likelihood ratio and the Wald test (Rayner and Best (1989), Chapter 3). This property only holds if the MLE is used. The general distribution theory of score tests then yields the asymptotic distribution given in part c) of Theorem 2.1 (see, e.g., Rayner and Best (1989), p. 80). However, in many cases, for example when testing for normality, it is not possible to define families of smooth alternatives (see Mardia and Kent (1991), p. 356; Kallenberg et al. (1997), p. 45). Hence, this approach is of limited applicability.

The limiting joint normal distribution of the components $\hat{U}_{n,j}$ may be degenerate. In particular, some components may be zero due to the method of estimation. The following lemma shows that this is always the case when using moment estimators.

In the following, we assume that ϑ is a unique function of the moments $\mu_1, \ldots, \mu_s$. Letting $m_l = (1/n)\sum_{i=1}^n X_i^l$ denote the lth empirical moment and writing $\mu_l(\vartheta) = E_\vartheta[X^l], l \geq 1$, the moment estimator $\bar{\vartheta}_n$ is defined by $\mu_l(\bar{\vartheta}_n) = m_l$ for $l = 1, \ldots, s$.

2.2 Lemma. *Let $\bar{\vartheta}_n$ denote the moment estimator of ϑ. If $\bar{\vartheta}_n \in \Theta$*, then

$$\hat{U}_{n,j} = 0, \quad j = 1, \ldots, s.$$

Proof: Write $h_j(x;\vartheta) = \sum_{l=0}^j a_l(\vartheta)x^l$. Since $E_\vartheta[h_j(X;\vartheta)] = 0$ for $j \geq 1$,

$$\sum_{l=0}^{j} a_l(\vartheta)\mu_l(\vartheta) = 0, \quad \vartheta \in \Theta.$$

Assuming $\bar{\vartheta}_n \in \Theta$, it follows that for $j = 1, \ldots, s$

$$\begin{aligned} \frac{1}{\sqrt{n}}\hat{U}_{n,j} &= \frac{1}{n}\sum_{i=1}^{n} h_j(X_i;\bar{\vartheta}_n) \\ &= \sum_{l=0}^{j} a_l(\bar{\vartheta}_n) m_l \\ &= \sum_{l=0}^{j} a_l(\bar{\vartheta}_n)\mu_l(\bar{\vartheta}_n) = 0. \quad \blacksquare \end{aligned}$$

If a smooth test is conducted solely as goodness of fit test for the parametric family P_ϑ, other estimation methods may be used instead of the (non-parametric) method of moments. Then, for example, the first component $\hat{U}_{n,1} = \sqrt{n}(\overline{X}_n - \mu_1(\hat{\vartheta}_n))$ examines whether $\mu_1(\hat{\vartheta}_n)$ is a 'reasonable' estimator of the mean. In general, this will not be the case if $P \notin \{P_\vartheta\}$. A large or small value of $\hat{U}_{n,1}$ indicates that the parametric model may not be appropriate. If the aim is to use the test as a 'diagnostic' procedure, the moment estimator of ϑ has to be used, as will be explained in the following.

The equations $E_\vartheta[h_j(X,\vartheta)]=0$, $j=1,\dots,s$, determine the moments $\mu_1,\dots,\mu_s$, which depend on $\vartheta_1,\dots,\vartheta_s$. Conversely, ϑ is a unique function of $\mu_1,\dots,\mu_s$ by assumption. On the set

$$\mathscr{P} := \left\{P : \int |x|^j P(dx) < \infty, j = 1,\dots,s, \vartheta = \vartheta(\mu_1,\dots,\mu_s) \in \Theta\right\},$$

we define a functional $\delta : \mathscr{P} \to \Theta$ by $\delta(P) = \vartheta(\mu_1,\dots,\mu_s)$ with $\mu_1,\dots,\mu_s$ being the moments of P. Note that

$$E_P[h_j(X,\delta(P))] = 0, \quad j = 1,\dots,s, \tag{4}$$

for each $P \in \mathscr{P}$. To establish convergence in distribution of $\hat{U}_n$ for $P \notin \{P_\vartheta\}$, a suitable requirement is that the estimator $\hat{\vartheta}_n$ of $\vartheta = \delta(P)$ satisfies condition (R1). In general, this will be the case for the moment estimator, but not for an estimator especially tailored to the parametric model P_ϑ. Hence, we assume in the following that the moment estimator $\overline{\vartheta}_n = \overline{\vartheta}_n(\mu_1,\dots,\mu_s)$ is used, which implies $\hat{U}_{n,j} = 0$, $j = 1,\dots,s$, by Lemma 2.2. Consequently, the test statistic $\hat{\Psi}^2_{n,k}$ consists of the components $\hat{U}_{n,s+1},\dots,\hat{U}_{n,s+k}$, i.e. $k_0 = s$. Suppose $\overline{\vartheta}_n$ satisfies (R1). Further, let

$$T_j(P) := \int h_j(\cdot\,;\delta(P))\,dP, \tag{5}$$

and define a nonparametric class of distributions by

$$\mathscr{P}_0 := \{P \in \mathscr{P} : T_{s+1}(P) = \cdots = T_{s+k}(P) = 0, E_P[X^{2(s+k)}] < \infty\}.$$

Obviously, $\mathscr{P}_0$ includes $\{P_\vartheta\}$: the moments $\mu_{s+1},\dots,\mu_{s+k}$ of each distribution in $\mathscr{P}_0$ and each distribution P_ϑ satisfy the same relations. The assertion of Theorem 2.1 a) carries over to the whole family $\mathscr{P}_0$, throughout replacing ϑ by $\delta(P)$ (cf. Section 3 in Henze and Klar (1996)). Note that the covariance matrix Σ in 2.1 a) depends on $P \in \mathscr{P}_0$.

A *directed* smooth test aims at making the diagnosis that, in case of rejection of the hypothesis, there are deviations in the moments $\mu_{s+1},\dots,\mu_{s+k}$ from the corresponding moments of the parametric model. Actually, this means testing the hypothesis $\tilde{H}_0 : P \in \mathscr{P}_0$ against $\tilde{H}_1 : P \notin \mathscr{P}_0$. Such a directed diagnosis, however, can not be achieved by the statistic $\hat{\Psi}^2_{n,k}$ in (3) since Σ in 2.1 a) and hence the asymptotic distribution of $\hat{\Psi}^2_{n,k}$ depends on the underlying $P \in \mathscr{P}_0$. To obtain a statistic which is asymptotically distribution-free over the class $\mathscr{P}_0$, $\hat{\Psi}^2_{n,k}$ has to be rescaled appropriately. The next theorem, which is in the setup and notation of Theorem 2.1, shows that such a rescaling can be done quite easily if $\hat{\Psi}^2_{n,k}$ has a limiting χ^2-distribution under H_0. This applies to all cases treated in Rayner and Best (1989).

2.3 Theorem. a) *Suppose that, under H_0, the covariance matrix Σ in 2.1 is the identitiy matrix, and, hence, $\hat{\Psi}^2_{n,k}$ has a limiting chi-squared distribution with k degrees of freedom under H_0. Further, let the support of $P \in \mathscr{P}_0$ consist of at least $s+k+1$ elements. Then $\Sigma_P = E_P[h(X;\delta(P))h(X;\delta(P))']$ is nonsingular, and $\hat{U}_n'\Sigma_P^{-1}\hat{U}_n$ has a limiting χ^2_k-distribution under P as well.*

b) *If $n \geq s+k+1$ and if the distribution function of P is continuous, then the $(k \times k)$-matrix*

$$\hat{H}_n = \frac{1}{n}\sum_{i=1}^{n}[h(X_i;\hat{\vartheta}_n)h(X_i;\hat{\vartheta}_n)']$$

is nonsingular with probability 1. Moreover, the rescaled statistic

$$\tilde{\Psi}^2_{n,k} := \hat{U}_n'\hat{H}_n^{-1}\hat{U}_n$$

has a limiting χ^2_k-distribution under P.

Proof: By assumption, $\Sigma = I_k$ under H_0. In view of 2.1 a), all entries in $E_\vartheta[\nabla_\vartheta h(X;\vartheta)]$ are zero. Now, $\partial h_r(x;\vartheta)/\partial\vartheta_j$ is a polynomial of degree r with respect to x:

$$\frac{\partial h_r(x;\vartheta)}{\partial\vartheta_j} = \sum_{i=0}^{r}\frac{\partial a_i(\vartheta)}{\partial\vartheta_j}x^i = \sum_{i=0}^{r}c_i(\vartheta)h_i(x;\vartheta).$$

Since $E_\vartheta[h_0(X;\vartheta)] = 1$ and $E_\vartheta[h_i(X;\vartheta)] = 0$ $(i \geq 1)$, we obtain $c_0(\vartheta) = 0$ for each $\vartheta \in \Theta$. In view of (4) and the definition of $\mathscr{P}_0$, this yields $E_P[\nabla_\vartheta h(X;\vartheta)|_{\vartheta=\delta(P)}] = 0$ for each $P \in \mathscr{P}_0$.

Hence, the covariance matrix under P is $\Sigma_P = E_P[h(X;\delta(P))h(X;\delta(P))']$. Using the assumption on the support of P, we obtain for $c_{s+1},\ldots,c_{s+k},\gamma \in \mathbb{R}$

$$P\left(\sum_{j=s+1}^{s+k}c_jh_j(X;\vartheta) = \gamma\right) = P\left(\sum_{j=0}^{s+k}c_j'X^j = \gamma\right) < 1,$$

where $c_{s+1}',\ldots,c_{s+k}' \in \mathbb{R}$ are determined by $c_{s+1},\ldots,c_{s+k}$. Therefore, the matrix Σ_P is nonsingular. Using well-known results, assertion a) follows.

To prove b), note that $\hat{H}_n$ is the empirical version of Σ_P. Hence, the statement about the distribution of $\tilde{\Psi}^2_{n,k}$ follows from a) if $\hat{H}_n$ is nonsingular with probability one.

To this end, let $n \geq s+k+1$ and assume that $X_1,\ldots,X_{s+k+1}$ take different values. Adding $s+1$ rows $(X_1^j,\ldots,X_{s+k+1}^j)$, $j = 0,\ldots,s$, to the $(k \times s+k+1)$-matrix with rows

$$(h_{s+j}(X_1;\hat{\vartheta}_n),\ldots,h_{s+j}(X_{s+k+1};\hat{\vartheta}_n)), \quad j = 1,\ldots,k,$$

yields a matrix with $s+k+1$ rows which can be transformed into a (nonsingular) Vandermonde matrix by elementary row operations. Hence, the original matrix is of rank k. It follows that $\hat{H}_n$ is nonsingular. ■

Remark: If P is a discrete distribution, $\hat{H}_n$ is nonsingular with probability smaller than one. Since the elements of $\hat{H}_n$ converge to the elements of Σ_P almost surely and the determinant of $\hat{H}_n$ depends continuously on these elements, there exists almost surely an integer n_0 such that $\hat{H}_n$ is nonsingular for each $n \geq n_0$. Hence, the assertion about the asymptotic distribution of $\tilde{\Psi}^2_{n,k}$ remains valid for discrete distributions satisfying the assumptions of part a).

3 Smooth tests in exponential families

In this section, we consider a large class of continuous and discrete parametric families of distributions with the property that the pertaining statistics $\hat{\Psi}^2_{n,k}$ have a limiting chi-squared distribution as required in Theorem 2.3. To this end, let P_ϑ be an s-parameter exponential family with density

$$p_\vartheta(x) = C(\vartheta)\exp\left[\sum_{j=1}^{s} \zeta_j(\vartheta)t_j(x)\right]h(x)$$

with respect to some σ-finite measure μ, where $1, t_1(x), \ldots, t_s(x)$ are affinely independent with probability one.

We make the additional assumption that the t_j are polynomials of degree less than or equal to s. Then, we may assume without loss of generality that $t_j(x) = x^j$ for $1 \leq j \leq s$.

Each of the parametric families considered in Rayner and Best (1989) belongs to this class of exponential families: in the continuous case the normal and the exponential distribution (or more generally the Gamma distribution with known shape parameter); in the discrete case the Poisson, the binomial and the geometric distribution.

If the parameter ϑ is one-dimensional, the exponential family is called *linear exponential*. In the discrete case, these distributions are also termed *modified power series distributions* (MPSD), since their probability mass functions have the representation

$$p_\vartheta(x) = \frac{a_x[u(\vartheta)]^x}{\eta(\vartheta)}, \quad a_x \geq 0,$$

where $u(\vartheta)$ is a positive (and usually differentiable) function and η is given by $\eta(\vartheta) = \sum_x a_x[u(\vartheta)]^x$. Besides the discrete distributions listed above, the negative binomial, the logarithmic and the Lagrange distribution are MPSD; moreover, each truncated MSPD like the positive Poisson distribution again is a modified power series distribution (see Johnson, Kotz and Kemp (1992)).

In exponential families, the maximum likelihood equations are

$$E_\vartheta[t_j(X)] = \frac{1}{n}\sum_{i=1}^{n} t_j(x_i), \quad j = 1, \ldots, s,$$

(see, e.g., Lehmann (1983), p. 439). In particular, if $t_j(x) = x^j$, we obtain $E_\vartheta[X^j] = \sum_{i=1}^{n} x_i^j/n$, i.e. the maximum likelihood equations and the equations defining the method of moments estimators coincide. By Lemma 2.2, the first s components are zero. Since

$$\nabla_\vartheta[\log p_\vartheta(x)] = \nabla_\vartheta[\log(C(\vartheta))] + \sum_{j=1}^{s} t_j(x)\nabla_\vartheta[\zeta_j(\vartheta)], \tag{6}$$

$\nabla_\vartheta[\log p_\vartheta(x)]$ is a polynomial with respect to x of degree s; hence

$$E_\vartheta[h_k(X,\vartheta)\nabla_\vartheta \log p_\vartheta(X)] = 0, \quad k > s, \tag{7}$$

where the $h_k(\cdot\,,\vartheta)$ denote the orthogonal polynomials of degree k. In view of 2.1 c), the next result follows.

3.1 Theorem. *Let $\{P_\vartheta\}$ be an s-parameter exponential family satisfying the assumptions of 2.1 c) and assume that the functions $t_j(x)$ are polynomials of degree less than or equal to s. If the maximum likelihood estimator of ϑ is used, $\hat{\Psi}^2_{n,k} = \sum_{j=s+1}^{s+k} \hat{U}^2_{n,j}$ has a limiting chi-squared distribution with k degrees of freedom.*

Consequently, Theorem 2.3 can be used to modify the smooth test of fit in order to have (at least asymptotically) diagnostic properties; if the k moments under consideration differ strongly from the moments of the hypothetical distribution, the modified test rejects the hypothesis for 'sufficiently large' sample size. Here, 'sufficiently large' depends on the underlying distribution which influences the speed of convergence to the limiting distribution.

In general, the necessary modifications are more complicated for parametric families which do not belong to the class described above, as can be seen in Theorem 2.1. Examples are the logistic, the Laplace and the Gumbel distribution or the Gamma distribution with unknown shape parameter, mentioned in Boulerice and Ducharme (1995).

3.2 Example. As an important example, we consider the smooth test of fit for the normal distribution, based on the first two nonzero components. Obviously, the normal distribution satisfies the assumptions of Theorem 3.1. The orthonormal polynomials with respect to the unit normal distribution are the normalized Hermite polynomials; the first four of these are

$$H_1(x) = x, \quad H_2(x) = (x^2 - 1)/\sqrt{2},$$

$$H_3(x) = (x^3 - 3x)/\sqrt{6}, \quad H_4(x) = (x^4 - 6x^2 + 3)/\sqrt{24}.$$

As a consequence, $h_j(x,\mu,\sigma) = H_j((x-\mu)/\sigma)(j \geq 1)$ are orthonormal with respect to the normal distribution $N(\mu,\sigma^2)$; the matrix $\hat{H}_n$ in Theorem 2.3 has the entries

$$\hat{H}_{jk} = \frac{1}{n}\sum_{i=1}^{n} h_j(X_i,\overline{X}_n,\hat{\sigma}_n)h_k(X_i,\overline{X}_n,\hat{\sigma}_n), \quad j,k \in \{3,4\},$$

where $\overline{X}_n$ and $\hat{\sigma}_n$ denote the arithmetic mean and the empirical standard deviation, respectively. In the class

$$\mathscr{P}_0 = \{P \in \mathscr{P} : T_3(P) = T_4(P) = 0, E_P[X^8] < \infty\},$$

the statistic

$$\tilde{\Psi}_{n,2}^2 = \frac{\hat{U}_{n,3}^2 \hat{H}_{44} - 2\hat{U}_{n,3}\hat{U}_{n,4}\hat{H}_{34} + \hat{U}_{n,4}^2 \hat{H}_{33}}{\hat{H}_{33}\hat{H}_{44} - \hat{H}_{34}^2}$$

has a limiting χ_2^2-distribution. The test based on $\tilde{\Psi}_{n,2}^2$ has diagnostic properties. If the skewness and/or the kurtosis of the underlying distribution differ from the corresponding values (= 0 and 3) of the normal distribution, the non-parametric hypothesis $\tilde{H}_0 : P \in \mathscr{P}_0$ will be rejected, at least for sufficiently large sample sizes: if $\alpha \in (0, 1)$ and $\chi_{2,1-\alpha}^2$ denotes the $(1 - \alpha)$-quantile of the χ_2^2-distribution, then

$$\lim_{n\to\infty} P(\tilde{\Psi}_{n,2}^2 > \chi_{2,1-\alpha}^2) = 1. \tag{8}$$

This fact can be seen as follows. If $P \notin \mathscr{P}_0$, a Taylor expansion as in the proof of Theorem 2.1 may be used to derive an asymptotic normal distribution for $\hat{U}_{n,j} - \sqrt{n}Eh_j(X, \overline{X}_n, \hat{\sigma}_n)$, $j = 3, 4$ (assuming $Eh_j^2(X, \overline{X}_n, \hat{\sigma}_n) < \infty$). Since $P \notin \mathscr{P}_0$, $Eh_3(X, \overline{X}_n, \hat{\sigma}_n) \neq 0$ or $Eh_4(X, \overline{X}_n, \hat{\sigma}_n) \neq 0$ and, hence, $|\hat{U}_{n,3}| \to \infty$ or $|\hat{U}_{n,4}| \to \infty$ in probability. A similar reasoning as in the proof of 2.3 b) yields the positive definiteness of $\hat{H}_n$ for large n under P. Hence, the consistency result (8) follows.

To get an impression about the speed of convergence of $\tilde{\Psi}_{n,2}^2$ to the asymptotic distribution if $\tilde{H}_0$ holds, we conducted a simulation study for the nominal level $\alpha = 0.1$ and several distributions from $\mathscr{P}_0$. These distributions were taken to be $N(0, 1)$, mixtures $pN(0, 1) + (1 - p)U(-a, a)$ of a standard normal distribution and a uniform distribution on $(-a, a)$ (denoted by $NU(p, a)$) and mixtures of a t-distribution with v degrees of freedom and a uniform distribution on $(-a, a)$ with equal mixing proportions (denoted by $tU(v, a)$). The parameters (p, a) and (v, a) are determined in such a way that the resulting distributions are in $\mathscr{P}_0$.

Table 1 shows the percentage of 10000 Monte Carlo samples declared significant by the tests based on $\hat{\Psi}_{n,2}^2$ and $\tilde{\Psi}_{n,2}^2$ for the sample sizes $n = 50$, 200, 1000 and 10000. Since the convergence of both statistics to their limit distributions is slow even if the underlying distribution is normal, we used empirical critical values (instead of $\chi_{2,0.9}^2$) to ensure that the statistics maintain the nominal level closely under normality. Note that $\hat{\Psi}_{n,2}^2$ coincides with the statistic of Bowman and Shenton (1975). Therefore, the results of Gastwirth and Owens (1977) can be used to determine the limit law of $\hat{\Psi}_{n,2}^2$. If the underlying distribution P belongs to $\mathscr{P}_0$, one obtains

$$\hat{\Psi}_{n,2}^2 \xrightarrow{\mathscr{D}} \gamma_1 N_1^2 + \gamma_2 N_2^2,$$

where N_1 and N_2 are independent unit normal random variables, and $\xrightarrow{\mathscr{D}}$ denotes weak convergence. The weights are given by

$$\gamma_1 = (E[\tilde{X}^6] - 9)/6, \quad \gamma_2 = (E[\tilde{X}^8] - 12E[\tilde{X}^6] + 99)/24,$$

Table 1. Percentage of 10000 Monte Carlo samples declared significant by the tests based on $\hat{\Psi}^2_{n,2}$ and $\tilde{\Psi}^2_{n,2}$ for several distributions from $\mathscr{P}_0$ ($\alpha = 0.1$)

Distribution	γ_1	γ_2	$\hat{\Psi}^2_{n,2}$ $n = 50$	200	1000	10000	$\tilde{\Psi}^2_{n,2}$ $n = 50$	200	1000	10000
$N(0,1)$	1.0	1.0	10.6	10.2	10.1	10.1	10.2	10.1	10.0	10.2
$NU(0.7, 3.506)$	.59	.36	5.2	2.3	1.2	0.8	8.1	4.7	6.2	9.4
$NU(0.5, 5.331)$	.41	.40	4.8	1.3	0.5	0.4	11.2	5.9	6.7	9.1
$NU(0.7, 1.307)$	1.2	1.4	11.9	15.1	16.3	17.0	15.5	14.9	12.3	10.7
$NU(0.5, 1.258)$	1.4	1.9	14.9	20.3	22.4	25.1	21.9	20.5	14.6	11.6
$NU(0.3, 1.180)$	1.8	3.0	20.0	31.5	35.5	37.3	32.2	28.7	18.9	11.5
$tU(9, 5.892)$	.41	.45	4.8	1.5	0.7	0.5	10.9	6.0	6.3	10.0
$tU(85, 1.306)$	1.5	2.2	14.9	22.9	26.3	28.4	22.9	23.7	16.6	11.1
$tU(20, 1.494)$	1.8	4.2	15.7	28.2	37.0	42.7	26.5	29.3	22.3	13.8

where $\tilde{X} = (X - EX)/(\operatorname{Var} X)^{1/2}$ is the standardized random variable and X is distributed according to P. The values of γ_1 and γ_2 for the different distributions are also given in Table 1 to explain the results for $\hat{\Psi}^2_{n,2}$ for large n.

The results of the simulations are a confirmation of the theoretical findings: For distributions $P \in \mathscr{P}_0$, the empirical size of the test based on $\tilde{\Psi}^2_{n,2}$ approaches its asymptotic value 0.1 for increasing sample sizes, whereas the proportion of the Monte Carlo samples declared significant by the tests based on $\hat{\Psi}^2_{n,2}$ is between 0 and 1. However, it should be noted that the convergence of the empirical size of the test based on $\tilde{\Psi}^2_{n,2}$ is quite slow for long-tailed distributions.

4 Diagnostic smooth tests of fit for bivariate Poissonity

Similarly to the univariate Poisson distribution, the bivariate Poisson distribution is an important discrete distribution in bivariate settings. A detailed description is given in Kocherlakota and Kocherlakota (1992). The bivariate Poisson-distribution is a 3-parameter family with probability mass function

$$f(r, s; \vartheta) = e^{-(\lambda_1^* + \lambda_2^* - \lambda_3)} \sum_{i=0}^{\min(r,s)} \frac{(\lambda_1^* - \lambda_3)^{r-i} (\lambda_2^* - \lambda_3)^{s-i} \lambda_3^i}{(r-i)!(s-i)!i!},$$

where $\vartheta = (\lambda_1^*, \lambda_2^*, \lambda_3)$. Often, the alternative parameterization $\lambda_1 = \lambda_1^* - \lambda_3$ and $\lambda_2 = \lambda_2^* - \lambda_3$ is used.

If the random vector (X, Y) has a bivariate Poisson distribution, X and Y have (univariate) Poisson distributions with parameters λ_1^* and λ_2^*, respectively. The factorial moments $\mu_{[r,s]} = E[X^{[r]} Y^{[s]}]$, where $x^{[r]} = x(x-1)(x-2) \ldots (x - r + 1)$, are given by

$$\mu_{[r,s]} = \lambda_1^{*r} \lambda_2^{*s} \sum_{i=0}^{\min(r,s)} \binom{r}{i} \binom{s}{i} i! \left(\frac{\lambda_3}{\lambda_1^* \lambda_2^*} \right)^i.$$

Hence, $E[X] = \lambda_1^*$, $E[Y] = \lambda_2^*$ and $\operatorname{Cov}(X, Y) = \lambda_3$. Higher non-central moments $\mu'_{r,s} = E[X^r Y^s] (r + s \leq 3)$ are given by

$$\mu'_{2,0} = \lambda_1^{*2} + \lambda_1^*,$$

$$\mu'_{3,0} = \lambda_1^{*3} + 3\lambda_1^{*2} + \lambda_1^*,$$

$$\mu'_{2,1} = \lambda_1^{*2}\lambda_2^* + 2\lambda_1^*\lambda_3 + \lambda_3 + \lambda_1^*\lambda_2^*;$$

similar equations hold for $\mu'_{0,2}, \mu'_{0,3}$ and $\mu'_{1,2}$. Therefore, the corresponding central moments $\mu_{r,s} = E[(X - EX)^r(Y - EY)^s]$ are $\mu_{2,0} = \mu_{3,0} = \lambda_1^*$, $\mu_{0,2} = \mu_{0,3} = \lambda_2^*$ and $\mu_{2,1} = \mu_{1,2} = \lambda_3$.

As in the previous sections, we want to construct smooth tests with diagnostic properties, but it is in general not possible to define a complete system of orthogonal polynomials with respect to a multivariate distribution (an exception is the multivariate normal distribution). However, appropriate bivariate polynomials can be used to detect deviations of the moments of the underlying distribution from the corresponding moments of the bivariate Poisson distribution. The assertion about the limiting distribution of components in Theorem 2.1a) remains valid (see the remark after 2.1), but not the remainder of the theorem since the proof makes use of the orthogonality relations. Instead of restating the general expressions, we consider two examples which are also examined in Rayner and Best (1995).

4.1 Example. In the work of Rayner and Best (1995), smooth tests of fit are derived as score tests with respect to certain alternatives; but again, there is the problem that the 'smooth alternatives' do not exist.

Rayner and Best (1995) use the maximum likelihood method. The MLE of λ_1^* and λ_2^* are given by $\hat{\lambda}_1^* = \overline{X}$ and $\hat{\lambda}_2^* = \overline{Y}$, respectively. The MLE $\hat{\lambda}_3$ of λ_3 differs from the empirical covariance $S_{XY} = n^{-1}\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})$ (i.e. from the corresponding estimator derived by the method of moments) and has to be computed by an iterative technique (see Kocherlakota and Kocherlakota (1992), Section 4.7).

Therefore, a test can be based on $h_1(x, y; \vartheta) = (x - \lambda_1^*)(y - \lambda_2^*) = (x - E[X])(y - (E[Y])$. The appropriate centered component

$$\hat{U}_{n,1} = \frac{1}{\sqrt{n}}\sum_{i=1}^n ((X_i - \hat{\lambda}_1^*)(Y_i - \hat{\lambda}_2^*) - \hat{\lambda}_3)$$

$$= \sqrt{n}(S_{XY} - \hat{\lambda}_3)$$

has a limiting normal distribution with expectation 0 under the hypothesis of bivariate Poissonity. Computing the variance of $\hat{U}_{n,1}$, normalizing and squaring, yields a test statistic which has a limiting χ_1^2-distribution (Rayner and Best (1995), Section 3).

Obviously, the smooth test which rejects the hypothesis of bivariate Poissonity for large values of $\hat{U}_{n,1}^2$ is not a test having diagnostic properties: a significant test result does not necessarily mean that the covariance structure of the data at hand is incompatible with that of a bivariate Poisson distribution for a given significance level (since λ_3 can take any positive value). It rather means that the MLE differs significantly from the nonparametric moment estimator, indicating that the hypothetical parametric model may not be appropriate.

A test which aims at detecting differences in certain higher moments has to use the moment estimator (see the remark after Lemma 2.2). In this case, the test based on $\hat{U}_{n,1}$ makes no sense since the test statistic is always zero.

4.2 Example. By estimating the three parameters of the bivariate Poisson distribution by the method of moments, we can construct a test which is able to detect deviations of the variances of the marginal distributions from those of the hypothetical model. The corresponding test statistic consists of all nonvanishing components of degree less than or equal to two. A reasonable choice are the polynomials $h_1(x, y; \vartheta) = (x - \lambda_1^*)^2$ and $h_2(x, y; \vartheta) = (y - \lambda_2^*)^2$. Since $Eh_1(X, Y; \vartheta)] = \mathrm{Var}(X) = \lambda_1^*$, the first centered component is

$$\hat{U}_{n,1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} ((X_i - \hat{\lambda}_1^*)^2 - \hat{\lambda}_1^*)$$

$$= \sqrt{n}(S_{XX} - \hat{\lambda}_1^*);$$

analogously, the second component is $\hat{U}_{n,2} = \sqrt{n}(S_{YY} - \hat{\lambda}_2^*)$.

It is not difficult to verify that condition (R1) on page 5 is satisfied with $l_1(x, y; \vartheta) = x - \lambda_1^*$, $l_2(x, y; \vartheta) = y - \lambda_2^*$ and $l_3(x, y; \vartheta) = xy - \lambda_3 - (x - \lambda_1^*)\lambda_2^* - (y - \lambda_2^*)\lambda_1^*$. Since $E[\partial h_1/\partial \lambda_1^*] = -1$ and $E[\partial h_1/\partial \lambda_2^*] = E[\partial h_1/\partial \lambda_3] = 0$, the first entry in the covariance matrix Σ in Theorem 2.1 a) is

$$\sigma_{11} = E[((X - \lambda_1^*)^2 - \lambda_1^* - (X - \lambda_1^*))^2]$$

$$= \mu_{4,0} - 2\mu_{3,0} - \lambda_1^{*2} + \lambda_1^*.$$

In a similar way, we get

$$\Sigma = \begin{pmatrix} \mu_{4,0} - 2\mu_{3,0} - \lambda_1^{*2} + \lambda_1^* & \mu_{2,2} - \mu_{2,1} - \mu_{1,2} + \lambda_3 - \lambda_1^* \lambda_2^* \\ * & \mu_{0,4} - 2\mu_{0,3} - \lambda_2^{*2} + \lambda_2^* \end{pmatrix}.$$

Replacing all higher moments in Σ with the corresponding empirical quantities and computing $\hat{U}_n' \hat{\Sigma}^{-1} \hat{U}_n$, yields the following assertion.

4.3 Theorem. *The test statistic $\hat{\Psi}_n^2$ given by the equation*

$$\frac{1}{n}\{(\widehat{\mu_{4,0}} - 2\widehat{\mu_{3,0}} - (\hat{\lambda}_1^*)^2 + \hat{\lambda}_1^*)(\widehat{\mu_{0,4}} - 2\widehat{\mu_{0,3}} - (\hat{\lambda}_2^*)^2 + \hat{\lambda}_2^*)$$

$$- (\widehat{\mu_{2,2}} - \widehat{\mu_{2,1}} - \widehat{\mu_{1,2}} + S_{XY} - \hat{\lambda}_1^* \hat{\lambda}_2^*)^2\} \hat{\Psi}_n^2$$

$$= (\widehat{\mu_{0,4}} - 2\widehat{\mu_{0,3}} - (\hat{\lambda}_2^*)^2 + \hat{\lambda}_2^*)(S_{XX} - \hat{\lambda}_1^*)^2$$

$$- 2(\widehat{\mu_{2,2}} - \widehat{\mu_{2,1}} - \widehat{\mu_{1,2}} + S_{XY} - \hat{\lambda}_1^* \hat{\lambda}_2^*)(S_{XX} - \hat{\lambda}_1^*)(S_{YY} - \hat{\lambda}_2^*)$$

$$+ (\widehat{\mu_{4,0}} - 2\widehat{\mu_{3,0}} - (\hat{\lambda}_1^*)^2 + \hat{\lambda}_1^*)(S_{YY} - \hat{\lambda}_2^*)^2$$

has a limiting chi-squared distribution with two degrees of freedom in the class of all bivariate discrete distributions having the property that the expectation of each marginal distribution equals its variance.

Remark: Using the moments of the bivariate Poisson distribution, the covariance matrix, after simplification, takes the form

$$\Sigma_0 = \begin{pmatrix} \mu_{4,0} - \lambda_1^{*2} - \lambda_1^* & \mu_{2,2} - \lambda_1^* \lambda_2^* - \lambda_3 \\ * & \mu_{0,4} - \lambda_2^{*2} - \lambda_2^* \end{pmatrix}.$$

This is the covariance matrix given in Rayner and Best (1995), Section 4. The corresponding statistic has a limiting χ_2^2-distribution under the hypothesis of bivariate Poissonity, but not within the wider nonparametric class of distributions defined above.

The last example showed how tests with diagnostic properties can be constructed in multivariate settings where no orthogonal system is available. A general objection against the use of moment estimators is the low efficiency compared with other estimators; in case of the bivariate Poisson distribution, the moment estimator S_{XY} of λ_3 is not efficient for large values of the correlation coefficient ρ (see Kocherlakota and Kocherlakota (1992), p. 108). This does not affect the testing procedure; but possibly one would prefer to continue working with an efficient estimator. Yet this argument does not meet the intention of a diagnostic test. The aim of a directed test is to examine whether the data coincide with a certain (simple) model in important characteristics (the first moments). If this is the case, the theoretical model is used even if it is not the 'true' underlying distribution. The term 'efficiency', however, is meaningful only within the parametric model.

Similarly, the statement that diagnostic tests do not always have good power as tests of fit is directed at the parametric model. However, diagnostic tests are not a new goodness of fit statistic for testing the parametric model but they aim to test a different (nonparametric) hypothesis. Therefore, comparisons with the power of goodness of fit tests for the parametric model are of limited meaning.

Acknowledgements. This work is based on a part of the author's doctoral thesis at the University of Karlsruhe, written under the supervision of Professor Norbert Henze, whose guidance is gratefully appreciated.

References

Bargal AI, Thomas DR (1983) Smooth goodness of fit tests for the Weibull distribution with singly censored data. Commun. Statist.-Theory Meth. 12:1431–1447

Boulerice B, Ducharme GR (1995) A note on smooth tests of goodness of fit for location-scale families. Biometrika 82:437–438

Bowman KO, Shenton BR (1975) Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. Biometrika 62:243–250

Chihara TS (1978) An introduction to orthogonal polynomials. New York, Gordon and Breach

Gastwirth JL, Owens MEB (1977) On classical tests of normality. Biometrika 64:135–139

Henze N (1997) Do components of smooth tests of fit have diagnostic properties? Metrika 45:121–130

Henze N, Klar B (1996) Properly rescaled components of smooth tests of fit are diagnostic. Austral. Journ. Statist. 38:61–74

Johnson NL, Kotz S, Kemp AW (1992) Univariate discrete distributions. J. Wiley & Sons, New York

Kallenberg WCM, Ledwina T, Rafajlowicz E (1997) Testing bivariate independence and normality. Sankhyā Ser. A 59:42–59

Kocherlakota S, Kocherlakota K (1992) Bivariate discrete distributions. Marcel Dekker, New York

Koziol JA (1986) Assesing multivariate normality: a compendium. Commun. Statist.-Theory Meth. 15:2763–2783

Koziol JA (1987) An alternative formulation of Neyman's smooth goodness of fit tests under composite alternatives. Metrika 34:17–24

Lehmann EL (1983) Theory of point estimation. J. Wiley & Sons, New York

Mardia KV, Kent JT (1991) Rao score tests for goodness of fit and independence. Biometrika 78:355–363

Neyman J (1937) Smooth test for goodness of fit. Skandinavisk Aktuarietidskrift 20:149–199

Rayner JCW, Best DJ (1986) Neyman-type smooth tests for location-scale families. Biometrika 73:437–446

Rayner JCW, Best DJ (1988) Smooth tests of goodness of fit for regular distributions. Commun. Statist.-Theory Meth. 17:3235–3267

Rayner JCW, Best DJ (1989) Smooth tests of goodness of fit. New York, Oxford University Press

Rayner JCW, Best DJ (1995) Smooth tests for the bivariate Poisson distribution. Austral. J. Statist. 37:233–245

Thomas DR, Pierce DA (1979) Neyman's smooth goodness-of-fit test when the hypothesis is composite. J. Amer. Statist. Ass. 74:441–445