Austral. J. Statist. **38**(1), 1996, 61–74

PROPERLY RESCALED COMPONENTS OF SMOOTH TESTS OF FIT ARE DIAGNOSTIC

NORBERT HENZE[1] AND BERNHARD KLAR[1]

Universität Karlsruhe

Summary

The paper illustrates a typical pitfall associated with a conventional interpretation of components of smooth tests of fit, for example the first non-zero components of the smooth tests for Poissonity, exponentiality, normality and the geometric distribution. In order to achieve a directed diagnosis concerning the kind of departure from a hypothesised model, an appropriate rescaling of components is necessary.

Key words: Goodness of fit; smooth tests; components; Poisson distribution; geometric distribution; exponential distribution; normal distribution.

1. Introduction

Smooth goodness of fit tests, introduced by Neyman (1937), are commonly regarded as a compromise between omnibus tests and focused procedures each having high power in one dimension of the (infinite dimensional) parameter space.

To test the hypothesis H_0 that an unknown distribution P over (the Borel sets of) $\mathbb{R}$ belongs to a specified 'regular' class $\{P_\vartheta : \vartheta \in \Theta\}$ of distributions indexed by an s-dimensional parameter ϑ, we use smooth test statistics that usually have the form

$$\widehat{\Psi}^2_{nk} = \sum_{j=s+1}^{s+k} \widehat{U}^2_{nj},$$

where

$$\widehat{U}_{nj} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j(X_i\,; \hat{\vartheta}_n). \tag{1.1}$$

Here, $X_1, \ldots, X_n$ is an independent and identically distributed (i.i.d.) sample from P, $\hat{\vartheta}_n = \hat{\vartheta}_n(X_1, \ldots, X_n)$ is the maximum likelihood estimator of ϑ under the parametric model, and $h_0(\cdot, \vartheta) \equiv 1$, $h_1(\cdot; \vartheta), \ldots, h_{s+k}(\cdot; \vartheta)$ is some set of orthonormal polynomials with respect to P_ϑ, i.e. we have

$$\int h_j(\cdot; \vartheta) h_\ell(\cdot; \vartheta)\, dP_\vartheta = \delta_{j\ell} \qquad (0 \le j, \ell \le s+k), \tag{1.2}$$

Received August 1995; revised March 1996; accepted March 1996.

[1]Institut für Mathematische Stochastik, Universität Karlsruhe, Englerstr. 2, D–76128 Karlsruhe, Germany.

Acknowledgments: The authors thank the referees for their constructive comments.

($\vartheta \in \Theta$), where $\delta_{j\ell}$ is the Kronecker delta function (see Rayner & Best, 1989).

H_0 is rejected at large values of $\widehat{\Psi}^2_{nk}$. In many cases, such as when testing for normality or testing for exponentiality, the limiting distribution as $n \to \infty$ of $\widehat{\Psi}^2_{nk}$ under P_ϑ is χ^2_k (see, however, Boulerice & Ducharme, 1995).

It is conventional statistical wisdom that, in case of rejection of H_0, the so-called *components* $\widehat{U}_{n,s+1}, \dots, \widehat{U}_{n,s+k}$ of $\widehat{\Psi}^2_{nk}$ should provide some kind of 'directed diagnosis' regarding the type of departure of P from H_0. In this respect, the first nonzero component $\widehat{U}_{n,s+1}$ plays an important role. For example, when testing for normality with both parameters unknown, we have $s = 2$, and $\widehat{U}_{n3}$ is a standardised version of the usual sample skewness coefficient, typically denoted by $\sqrt{b_1}$.

In recent years, however, attitudes towards the diagnostic capabilities of the skewness coefficient in particular and of components of smooth tests of fit in general have changed drastically (Horswell & Looney, 1992, 1993; Henze, 1994, 1996a; Rayner, Best & Mathews, 1995).

Generalising the findings of Rayner *et al.* (1995) for the skewness coefficient, we show that some widely used tests, each based on the first nonzero component of a smooth test of fit, are strictly non-diagnostic when used conventionally. To achieve the desired 'directed diagnosis', an appropriate rescaling has to be done.

The paper is organised as follows. In Section 2 we consider in detail the typical diagnostic pitfall associated with Fisher's dispersion index which, up to a linear transformation, is the first nonzero component of the smooth test for Poissonity. Section 3 gives some general results. Smooth tests for the geometric, the exponential and the normal distributions are treated briefly in Section 4. The final section presents some concluding discussion.

We use standard notation, such as $\xrightarrow{\mathcal{D}}$ to denote convergence in distribution, $\sim$ to mean 'has the same distribution as', and $\mathcal{N}(\mu, \sigma^2)$ is the normal distribution with expectation μ and variance σ^2.

2. Testing for Poissonity

Let $X_1, \dots, X_n, \dots$ be a sequence of i.i.d. nonnegative integer valued random variables with common unknown distribution P. Many formal goodness of fit tests have been suggested for testing the hypothesis of Poissonity,

$$H_0: \quad P \text{ is } P_0(\lambda) \text{ for some } \lambda > 0,$$

(for some recent proposals see, e.g. Epps (1995), Henze (1996b) and references therein). Apart from the celebrated χ^2 test with its well-known deficiencies (in particular, the possibility of manipulation of p-values since cell selection is usually done after a close inspection of the data), a common procedure for checking the adequacy of a Poisson model is to calculate Fisher's index of dispersion,

$$D_n = \frac{\sum_{j=1}^n (X_j - \bar{X}_n)^2}{\bar{X}_n},$$

where $\bar{X}_n = n^{-1}\sum_{j=1}^{n} X_j$, and to reject H_0 for large or small values of D_n.

It is commonly believed that, when H_0 is rejected, the statistic D_n provides a 'directed diagnosis' regarding the kind of departure of P from H_0 in the sense that large (small) values of D_n indicate that the underlying distribution has a variance greater (smaller) than the mean (see, e.g. Epps, 1995 p.1463). Note that

$$\widehat{U}_{n2} = \frac{1}{\sqrt{2n}}(D_n - n),$$

a 'standardised version' of D_n, is the first nonzero component of Neyman's smooth test of fit for the Poisson distribution based on the Poisson–Charlier orthonormal polynomials (see Rayner & Best, 1989 Section 6.4). The second Poisson–Charlier polynomial is given by

$$h_2(x;\vartheta) = \frac{1}{\vartheta\sqrt{2}}\big((x-\vartheta)^2 - x\big). \tag{2.1}$$

Rayner & Best (1989 p.91) argue that 'the components of this statistic (the smooth test based on the Poisson–Charlier orthonormal polynomials) ... may be interpreted as identifying deviations of the data from the Poisson moments.'

To illustrate the 'diagnostic pitfall' associated with $\widehat{U}_{n2}$, let $\mathcal{P}$ denote the set of all non-degenerate probability distributions of $\{0, 1, \ldots\}$ having finite fourth moment. Define the functional $\gamma : \mathcal{P} \to \mathbb{R}$ by

$$\gamma(P) = \frac{1}{\sqrt{2}}\left(\frac{\operatorname{var} X}{\mathrm{E}X} - 1\right),$$

where X has distribution P.

Writing $\hat{P}_n$ for the empirical measure of n i.i.d. copies $X_1, \ldots, X_n$ of X, where $X \sim P$ and $P \in \mathcal{P}$, we have

$$\gamma(\hat{P}_n) = \frac{1}{\sqrt{n}}\hat{U}_{n2}.$$

Moreover, $\gamma(\hat{P}_n) \to \gamma(P)$ almost surely as $n \to \infty$ by the strong law of large numbers. Thus, $n^{-1/2}\hat{U}_{n2}$ is a strongly consistent estimator of the population parameter $\gamma(P)$. Note that

$$\gamma(P)\left\{\begin{matrix} > \\ = \\ < \end{matrix}\right\} 0 \quad \Longleftrightarrow \quad \operatorname{var} X \left\{\begin{matrix} > \\ = \\ < \end{matrix}\right\} \mathrm{E}X,$$

and that $\gamma(P) = 0$ if P has a Poisson distribution. It is tempting to say that a 'too large' ('too small') observed value of $\gamma(\hat{P}_n)$ is an indication of $\gamma(P) > 0$ ($\gamma(P) < 0$) and thus of overdispersion (underdispersion). This conclusion, however, is not justified for the simple reason that, on one hand being 'too large' or 'too small' is judged from the asymptotic behaviour of $\gamma(\hat{P}_n)$ *under the hypothesis of Poissonity*. On the other hand, the conclusion refers to the *nonparametric population parameter* $\gamma(P)$, which (this is the important point) may be zero even if the underlying distribution is not of Poisson type. To make the argument precise, let $\mathcal{P}_0$ denote the set of all distributions P in $\mathcal{P}$ such that $\gamma(P) = 0$. We then have the following result.

Theorem 2.1. *Let $X_1, \ldots, X_n, \ldots$ be i.i.d. with distribution P, where $P \in \mathcal{P}_0$. Then*

$$\widehat{U}_{n2} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(P)) \qquad (n \to \infty),$$

where

$$\sigma^2(P) = \frac{1}{2\mu^2}(\mu_4 - 2\mu_3 - 2\mu_2\mu + \mu_2 + \mu^2) = \int h_2^2(\cdot\,;\mu)\, dP, \tag{2.2}$$

and $\mu = \mathrm{E}X$, $\mu_j = \mathrm{E}(X-\mu)^j$ $(j = 2, 3, 4)$, where $X \sim P$.

Proof. Since $\mu = \mu_2$ if $P \in \mathcal{P}_0$, straightforward algebra shows that $\widehat{U}_{n2}$ equals

$$\frac{1}{\bar{X}_n\sqrt{2}} \cdot \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [X_j^2 - \mu_2 - \mu^2 - (1+2\mu)(X_j - \mu)] - \frac{n(\bar{X}_n - \mu)^2}{\bar{X}_n\sqrt{2n}},$$

and the assertion follows from the Central Limit Theorem and Slutsky's Lemma.

Corollary 2.2. *If P is $P_0(\lambda)$ for some $\lambda > 0$, then*

$$\widehat{U}_{n2} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{as } n \to \infty.$$

Proof. If $X \sim P_0(\lambda)$, then $\mu = \mu_2 = \mu_3 = \lambda$, $\mu_4 = \lambda + 3\lambda^2$ and $\sigma^2(P) = 1$.

Consequently, an asymptotically level α test of H_0 based on $\widehat{U}_{n2}$ rejects the hypothesis of Poissonity if $|\widehat{U}_{n2}| \geq z_{\alpha/2}$, where $z_{\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. Theorem 2.1 implies that the asymptotic power of this test against a fixed alternative distribution P from $\mathcal{P}_0$ is equal to $2\big(1 - \Phi(\sigma(P)^{-1} z_{\alpha/2})\big)$, where Φ denotes the standard normal distribution function. Since

$$\inf_{P \in \mathcal{P}_0} \sigma^2(P) = 0 \qquad \text{and} \qquad \sup_{P \in \mathcal{P}_0} \sigma^2(P) = \infty$$

(see the examples below), it follows that a 'too large' observed value of $|\widehat{U}_{n2}|$ may result not only from a population P having $\gamma(P) \neq 0$ but also from an underlying population P which is neither overdispersed nor underdispersed, but has a large value of $\sigma^2(P)$.

More generally, it can be shown that, for $P \in \mathcal{P}$, $\widehat{U}_{n2} - \sqrt{n}\gamma \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_*^2(P))$ as $n \to \infty$, where

$$\sigma_*^2(P) = \frac{1}{2\mu^2}\left(\mu_4 - 2\mu_3 + \mu_2(1-\mu_2) + \gamma\left(\mu_2\gamma - 2(\mu_3 - \mu_2)\right)\right),$$

and $\gamma = \gamma(P)$ for short. Consequently, the mean drift $\sqrt{n}\gamma$ for a distribution P in $\mathcal{P}\backslash\mathcal{P}_0$ may, at least for sample sizes occurring in practice, be overcompensated by a large variance $\sigma_*^2(P)$. This means that a large (small) observed value of $\widehat{U}_{n2}$ may result from a distribution P having a negative (positive) value of γ, but a large value of $\sigma_*^2(P)$. To obtain an asymptotically level α test of $\tilde{\mathrm{H}}_0$: $P \in \mathcal{P}_0$ against $\tilde{\mathrm{H}}_1$: $P \in \mathcal{P}\backslash\mathcal{P}_0$, which is the aim of a directional diagnosis, $\widehat{U}_{n2}$ must be rescaled. From Theorem 2.1 we have the following result.

Corollary 2.3. *Let*

$$\widetilde{U}_{n2} = \frac{\widehat{U}_{n2}}{\left(\int h_2^2(\cdot;\bar{X}_n)\, d\hat{P}_n\right)^{1/2}}.$$

The test which rejects $\tilde{\mathrm{H}}_0$ *if* $|\widetilde{U}_{n2}| \geq z_{\alpha/2}$ *is an asymptotically level* α*-test of* $\tilde{\mathrm{H}}_0$ *against* $\tilde{\mathrm{H}}_1$.

Note that $\widetilde{U}_{n2} = \widehat{U}_{n2}/\sigma(\hat{P}_n)$, where

$$\sigma^2(\hat{P}_n) = \int h_2^2(\cdot;\bar{X}_n)\, d\hat{P}_n = (2\bar{X}_n^2 n)^{-1} \sum_{j=1}^{n} \left((X_j - \bar{X}_n)^2 - X_j\right)^2,$$

and $\sigma^2(P)$ is given in (2.2). It is easily seen that the test based on $\widetilde{U}_{n2}$ is consistent against each fixed alternative distribution from $\tilde{\mathrm{H}}_1$.

Example 2.1. Let $Q_{kp} = \frac{1}{2}p\,\delta_0 + (1-p)\delta_k + \frac{1}{2}p\,\delta_{2k}$ $(0 < p < 1)$, be the distribution putting mass $\frac{1}{2}p$ on 0 and $2k$, and mass $1-p$ on k, where $k \in \mathbb{N}$. Writing X for a random variable having distribution Q_{kp}, it follows that $\mathrm{E}X = k$ and, for $j \geq 2$,

$$\mu_j = \tfrac{1}{2}p(-k)^j + \tfrac{1}{2}pk^j = \begin{cases} pk^j, & \text{if } j \text{ is even,} \\ 0 & \text{otherwise.} \end{cases}$$

Since the condition $\mu_2 = \mu$ is equivalent to $p = 1/k$, we see that $\{Q_{k,1/k} : k = 2, 3, \ldots\}$ is a subclass of $\mathcal{P}_0$. From $\mu_4 = pk^4 = k^3$ we obtain

$$\sigma^2(Q_{k,1/k}) = \frac{\mu_4 - 2\mu_3 + \mu(1-\mu)}{2\mu^2} = \frac{k^3 + k(1-k)}{2k^2},$$

and thus $\sup_{P \in \mathcal{P}_0} \sigma^2(P) = \infty$.

Example 2.2. Let

$$\widetilde{Q}_{k,\ell,p} = \tfrac{1}{2}p\,\delta_{\ell-k} + (1-p)\delta_\ell + \tfrac{1}{2}p\,\delta_{\ell+k} \qquad (0<p<1;\ k,\ell\in\mathbb{N},\ k<\ell)$$

be the distribution putting mass $\frac{1}{2}p$ on $\ell-k$ and $\ell+k$, and mass $1-p$ on ℓ. In this case we have $\mu=\ell$ and, for $j\ge 2$,

$$\mu_j = \begin{cases} pk^j & \text{if } j \text{ is even,} \\ 0 & \text{otherwise.} \end{cases}$$

Fixing $k\ge 2$ and putting $\ell_k = k^2-1$, $p_k = (k^2-1)/k^2$, it follows that $\mu_2=\ell=\mu$. Thus, $\{\widetilde{Q}_{k,\ell_k,p_k} : k\ge 2\}$ is a subclass of $\mathcal{P}_0$. Since an easy calculation shows that

$$\sigma^2(\tilde{Q}_{k,\ell_k,p_k}) = \frac{1}{k^2-1},$$

we obtain $\inf_{P\in\mathcal{P}_0}\sigma^2(P)=0$.

To enhance the theoretical findings concerning the behaviour of $\widehat{U}_{n2}$ and $\widetilde{U}_{n2}$, we conducted a small simulation study for the nominal level $\alpha=0.05$ and several distributions from $\mathcal{P}_0$. These distributions were taken to be $P_0(\lambda)$ with $\lambda=2$ and $\lambda=5$, the discrete uniform distribution $\mathcal{U}(4)$ over $\{0,1,2,3,4\}$, a mixture of the two Binomial distributions $\mathcal{B}(6,\frac{2}{3})$ and $\mathcal{B}(12,\frac{2}{3})$ with equal mixing proportions (denoted by $\mathcal{MB}(6,\frac{2}{3};12,\frac{2}{3})$). The last two distributions are mixtures $q\mathcal{B}(m,p)+(1-q)\mathcal{L}(\xi)$ of a Logarithmic distribution with $\xi=0.8$ ($\xi=0.9$) and a Binomial distribution with $m=6$ and $p=-\xi/\big(m(1-\xi)\log(1-\xi)\big)$ (this condition ensures equality of expectations of the two components). The parameter q is determined by the additional requirement 'mean = variance'. After some algebra, one obtains

$$q = \frac{1+mp-(1-\xi)^{-1}}{1-p+mp-(1-\xi)^{-1}}.$$

The resulting mixtures are denoted by $\mathcal{BL}(m,\xi)$.

For the distribution $\mathcal{U}(4)$, we have $\mu=\mu_2=2$, $\mu_3=0$, $\mu_4=6.8$ and thus $\sigma^2\big(\mathcal{U}(4)\big)=0.6$. In this case, the asymptotic power of the test based on $\widehat{U}_{n2}$ (calculated from (2.2)) is given by

$$2\big(1-\Phi(1.96/\sqrt{0.6})\big)\approx 0.011.$$

For the mixture $\mathcal{MB}(6,\frac{2}{3};12,\frac{2}{3})$, the corresponding values are $\mu=\mu_2=6$, $\mu_3=10/3$, $\mu_4=674/9$ and $\sigma^2\big(\mathcal{MB}(6,\frac{2}{3};12,\frac{2}{3})\big)=0.53$ resulting in an asymptotic power of

$$2\big(1-\Phi(1.96/\sqrt{0.53})\big)\approx 0.007.$$

TABLE 1

Empirical power of the tests based on $\widehat{U}_{n2}$ and $\widetilde{U}_{n2}$ for various distributions from $\mathcal{P}_0$ ($\alpha = 0.05$, $n = 50$ and $n = 200$, 5000 Monte Carlo samples)

	$P_0(2)$	$P_0(5)$	$\mathcal{U}(4)$	$\mathcal{MB}(6, \frac{2}{3}; 12, \frac{2}{3})$	$\mathcal{BL}(6, 0.8)$	$\mathcal{BL}(6, 0.9)$
$\widehat{U}_{50,2}$	0.046	0.047	0.012	0.010	0.196	0.527
$\widehat{U}_{200,2}$	0.048	0.044	0.013	0.008	0.352	0.685
∞	0.05	0.05	0.011	0.007	0.56	0.76
$\widetilde{U}_{50,2}$	0.084	0.081	0.050	0.065	0.088	0.455
$\widetilde{U}_{200,2}$	0.062	0.063	0.054	0.058	0.052	0.339

For the two mixtures of a Binomial distribution and a Logarithmic distribution, we obtain $\sigma^2(\mathcal{BL}(6, 0.8)) \approx 11.12$ and $\sigma^2(\mathcal{BL}(6, 0.9) \approx 42.9$, resulting in asymptotic powers of 0.56 and 0.76 respectively.

In Table 1, the rows $\widehat{U}_{50,2}$ and $\widehat{U}_{200,2}$ show the empirical power, for the six distributions of $\mathcal{P}_0$ described above, of the test rejecting H_0 if $|\widehat{U}_{n2}| \geq z_{\alpha/2}$ for the sample sizes $n = 50$ and $n = 200$ respectively. The nominal level α was taken to be 0.05, and there were 5000 Monte Carlo replications. The corresponding results for the test rejecting $\widetilde{\mathrm{H}}_0$ if $|\widetilde{U}_{n2}| \geq z_{\alpha/2}$ are given in the rows $\widetilde{U}_{50,2}$ and $\widetilde{U}_{200,2}$. The asymptotic theoretical power of the test based on $\widehat{U}_{n2}$ is shown in the row denoted ∞.

Recall that the two-sided test based on $\widehat{U}_{n2}$ (which aims at testing Poissonity against non-Poissonity) is inconsistent against each of the four alternative distributions chosen in Table 1. On the other hand, a two-sided test based on $\widetilde{U}_{n2}$ aims at testing the much broader 'nonparametric' hypothesis $\widetilde{\mathrm{H}}_0 : P \in \mathcal{P}_0$ against the 'directed diagnosis' alternative $\widetilde{\mathrm{H}}_1 : P \in \mathcal{P} \backslash \mathcal{P}_0$. Since all six distributions figuring in Table 1 belong to $\mathcal{P}_0$, the last two lines in Table 1 give the empirical *size* of this test for six distributions from $\mathcal{P}_0$ and sample sizes $n = 50$ and $n = 200$.

Although the convergence of the empirical size of the test based on $\widetilde{U}_{n2}$ to its asymptotic value 0.05 is rather slow for the distribution $\mathcal{BL}(6, 0.9)$ (for $n = 2000$ and $n = 10\,000$ the empirical sizes are 0.108 and 0.052 respectively), the results clearly illustrate the theoretical findings.

3. General Results

Guided by the discussion of a 'correct' interpretation of the first nonzero component of the smooth test of fit for Poissonity, we now consider a component $\widehat{U}_{nj}$ as given in (1.1) in the general setting of Section 1. We assume that the underlying distribution P has a finite moment of order $2 \cdot j$, so that $\mathrm{E}h_j^2(X; \vartheta) < \infty$. Let $\delta(P)$ be some functional, defined on the set $\mathcal{P}$ of such distributions and taking values in Θ, with the property that $\delta(P_\vartheta) = \vartheta \quad (\vartheta \in \Theta)$. Of course,

$\delta(P) = \mathrm{E}(X)$ for the Poisson case treated in the previous section. Furthermore, let the parameter set Θ be an open interval in $\mathbb{R}$, and let $\hat{\vartheta}_n = \hat{\vartheta}_n(X_1, \ldots, X_n)$ be any estimator for $\vartheta = \delta(P)$ such that, under P, the sequence $\sqrt{n}(\hat{\vartheta}_n - \vartheta)$ is bounded in probability. Finally, let $h_j(\cdot; \vartheta)$ be a smooth function of ϑ so that, with $\vartheta = \delta(P)$, the Taylor expansion

$$\begin{aligned}\widehat{U}_{nj} - \sqrt{n}\mathrm{E}h_j(X;\vartheta) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[h_j(X_i;\hat{\vartheta}_n) - \mathrm{E}h_j(X;\vartheta)] \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[h_j(X_i;\vartheta) - \mathrm{E}h_j(X;\vartheta)] \\ &\quad + \mathrm{E}\Big[\frac{\partial}{\partial\vartheta}h_j(X;\vartheta)\Big|_{\vartheta=\delta(P)}\Big]\sqrt{n}(\hat{\vartheta}_n - \vartheta) + o_P(1)\end{aligned}$$

holds. Putting

$$T_j(P) = \int h_j(\cdot;\delta(P))\,dP,$$

we have $T_j(P_\vartheta) = 0 \quad (\vartheta \in \Theta)$, so that the functional T_j vanishes over the hypothesised parametric family $\{P_\vartheta : \vartheta \in \Theta\}$. Now, under the standing moment conditions, $\mathcal{P}_0 = \{P : T_j(P) = 0\}$ defines a nonparametric family of distributions containing $\{P_\vartheta : \vartheta \in \Theta\}$.

A directed diagnosis with respect to $\widehat{U}_{nj}$ aims at testing the nonparametric hypothesis $\widetilde{\mathrm{H}}_0 : \; P \in \mathcal{P}_0$ against the alternative $\widetilde{\mathrm{H}}_1 : \; P \in \mathcal{P}\backslash\mathcal{P}_0$. To derive an asymptotically level α test of $\widetilde{\mathrm{H}}_0$ against $\widetilde{\mathrm{H}}_1$ based on $\widehat{U}_{nj}$ we use the Taylor expansion given above and distinguish two cases. The first case is that

$$\mathrm{E}\Big[\frac{\partial}{\partial\vartheta}h_j(X;\vartheta)\Big|_{\vartheta=\delta(P)}\Big] = 0 \quad \text{for each } P \in \mathcal{P}_0. \tag{3.1}$$

This condition typically holds for the first nonzero component h_{s+1}. Using (2.1), (3.1) is readily verified for the Poisson case, and it also holds for each of the examples given in Section 4. Note that, regardless of further assumptions on the estimator $\hat{\vartheta}_n$, (3.1) implies

$$\widehat{U}_{nj} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_j(X_i;\vartheta) + o_P(1),$$

and thus

$$\widehat{U}_{nj} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \mathrm{E}h_j^2(X;\delta(P))\Big)$$

as $n \to \infty$ if P belongs to $\mathcal{P}_0$. To obtain an asymptotically disribution-free test of $\mathcal{P}_0$ and thus a directed diagnosis concerning the value $T_j(P)$, one has to divide

$\widehat{U}_{nj}$ by the consistent estimator

$$\Big(\int h_j^2(\cdot;\delta(\hat{P}_n))\,d\hat{P}_n\Big)^{1/2} = \Big(\frac{1}{n}\sum_{i=1}^{n} h_j^2(X_i;\delta(\hat{P}_n))\Big)^{1/2}$$

of the standard deviation $\big(\mathrm{E}h_j^2(X;\delta(P))\big)^{1/2}$.

For higher components, condition (3.1) does not hold in general. In fact, $\mathrm{E}\big[\partial/\partial\vartheta\, h_j(X;\vartheta)\big|_{\vartheta=\delta(P)}\big]$ typically vanishes only for some distributions P in $\mathcal{P}_0$. A simple example is given by the third Poisson–Charlier polynomial

$$h_3(x;\vartheta) = (6\vartheta^3)^{-1/2}\big((x-\vartheta)^3 - 3x(x-\vartheta) + 2x\big)$$

in the Poisson case. In this case, $T_3(P) = 0$ if and only if

$$2 - 3\frac{\mathrm{E}[X(X-1)]}{\vartheta^2} + \frac{\mathrm{E}[X(X-1)(X-2)]}{\vartheta^3} = 0,$$

and

$$\mathrm{E}\Big[\frac{\partial}{\partial\vartheta}h_3(X;\vartheta)\Big|_{\vartheta=\delta(P)}\Big] = 0 \quad\Longleftrightarrow\quad \frac{\mathrm{E}[X(X-1)]}{\vartheta^2} = \frac{\mathrm{E}[X(X-1)(X-2)]}{\vartheta^3}.$$

To obtain the asymptotic distribution of $\widehat{U}_{nj}$ for $P \in \mathcal{P}_0$ if (3.1) does not hold, we need some additional information on the estimator $\hat{\vartheta}_n$. If $\sqrt{n}(\hat{\vartheta}_n - \vartheta)$ allows for a standard representation of the form

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell(X_i;\delta(P)) + o_P(1), \tag{3.2}$$

where $\mathrm{E}\ell(X;\delta(P)) = 0$ and $\mathrm{E}\ell^2(X;\delta(P)) < \infty$, the Taylor expansion given above yields

$$\widehat{U}_{nj} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2(P))$$

as $n\to\infty$ under $P \in \mathcal{P}_0$, where

$$\sigma^2(P) = \mathrm{E}\Big(h_j(X;\vartheta) + \mathrm{E}\Big[\frac{\partial}{\partial\vartheta}h_j(X;\vartheta)\Big]\ell(X;\vartheta)\Big)^2$$

and $\vartheta = \delta(P)$. As before, $\widehat{U}_{nj}$ has to be divided by $\sigma(\hat{P}_n)$ in order to obtain an asymptotically distribution-free test of $\mathcal{P}_0$.

Of course, the Taylor expansion may also be used to derive an asymptotic normal distribution for $\widehat{U}_{nj} - \sqrt{n}\,\mathrm{E}\,h_j(X;\delta(P))$ if $P \notin \mathcal{P}_0$, provided that $\sqrt{n}(\hat{\vartheta}_n - \vartheta)$ has a standard representation of the type (3.2).

4. Further Examples

4.1. The Geometric Distribution

Let $G(p)$ be the geometric distribution with probability mass function qp^x ($x = 0, 1, \ldots$), where $0 < p < 1$ and $q = 1 - p$. For testing the hypothesis

$$\mathrm{H}_0: \quad P \text{ is } G(p) \text{ for some } p \in (0,1),$$

based on an i.i.d. sample $X_1, \ldots, X_n$ with unknown distribution P over $\{0, 1, \ldots\}$, the first nonzero component of Neyman's smooth test of fit for the geometric distribution, based on the second Meixner orthogonal polynomial $h_2(x, \vartheta) = \left(2\vartheta(1+\vartheta)\right)^{-1}\left(x(x-1) - 4\vartheta x + 2\vartheta^2\right)$, is given by

$$\widehat{U}_{n2} = \frac{1}{2\bar{X}_n(1+\bar{X}_n)} \cdot \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(X_j(X_j - 1) - 4\bar{X}_n X_j + 2\bar{X}_n^2\right)$$

(see Best & Rayner, 1989). Here, $\vartheta = \mathrm{E}X = \delta(P)$.

Retaining the notation $\mathcal{P}$ and $\hat{P}_n$ from Section 3, we have

$$T_2(\hat{P}_n) = \frac{1}{\sqrt{n}} \widehat{U}_{n2},$$

where

$$T_2(P) = \tfrac{1}{2}\left(\frac{\operatorname{var} X}{\mathrm{E}X(1+\mathrm{E}X)} - 1\right) = \int h_2(\cdot; \delta(P))\, dP,$$

and X is a random variable having distribution P.

Since $\mathrm{E}X = q/p$ and $\operatorname{var} X = q/p^2$ if $X \sim G(p)$, we see that $T_2(P) = 0$ if the distribution of P is geometric. However, similar to the Poisson case, the problem for a directed diagnosis based on $\widehat{U}_{n2}$ is that the equation $T_2(P) = 0$ does not characterise the class of geometric distributions. To make the argument precise, let $\mathcal{P}_0$ be the set of all distributions P in $\mathcal{P}$ such that $T_2(P) = 0$. Retaining the notation μ and μ_j from Theorem 2.1, we have the following result.

Theorem 4.1. *Let $X_1, \ldots, X_n, \ldots$ be i.i.d. with distribution P, where $P \in \mathcal{P}_0$. Then*

$$\widehat{U}_{n2} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(P)),$$

where

$$\sigma^2(P) = \frac{\mu_4 - 2(1+2\mu)\mu_3 + \mu_2(2\mu^2 + \mu + 1) + \mu^2(1+\mu)^2}{4\mu^2(1+\mu)^2}. \tag{4.1}$$

Proof. Since $P \in \mathcal{P}_0$ entails $\mu_2 = \mu(1+\mu)$, some calculations yield

$$\widehat{U}_{n2} = \frac{1}{2\bar{X}_n(1+\bar{X}_n)} \cdot \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Y_j - \frac{\sqrt{n}(\bar{X}_n - \mu)^2}{\bar{X}_n(1+\bar{X}_n)},$$

where $Y_j = X_j(X_j - 1) - 2\mu^2 - 4\mu(X_j - \mu)$ $(j = 1, \ldots, n)$ are centred random variables. Since $\operatorname{var} Y_1$ is the numerator figuring in (4.1), the assertion follows.

If the distribution P is geometric, we have $\mu_4 = \mu(1+\mu)[1 + 9\mu(1+\mu)]$ and $\mu_3 = \mu(1+\mu)(1+2\mu)$ which implies $\sigma^2(P) = 1$. Consequently, the asympotic H_0-distribution of $\widehat{U}_{n2}$ as $n \to \infty$ is $\mathcal{N}(0,1)$ irrespective of p, and an asymptotically level α test of H_0 based on $\widehat{U}_{n2}$ rejects H_0 if $|\widehat{U}_{n2}| \geq z_{\alpha/2}$.

However, a tempting diagnosis, such as 'the huge positive $\widehat{U}_{n2}$ value indicates the observed distribution has a higher variance than expected for a geometric distribution' (Best & Rayner, 1989 p.310) is not valid in view of Theorem 4.1. The underlying distribution P may share the relation $\mu_2 = \mu(1 + \mu)$ with the geometric distribution (i.e. we have $T_2(P) = 0$), but there may be a large value of $\sigma^2(P)$ leading to a large absolute value of $\widehat{U}_{n2}$ and thus to rejection of H_0 (but with a wrong directed diagnosis). On the other hand, we may have $T_2(P) = 0$ and a small value of $\sigma^2(P)$ leading to a small absolute value of $\widehat{U}_{n2}$ and thus to false acceptance of H_0.

As before a possible remedy with regard to desired 'directed diagnosis' is to estimate the asymptotic variance $\sigma^2(P)$ by $\sigma^2(\hat{P}_n) = \int h_2^2(\cdot; \bar{X}_n)\, d\hat{P}_n$ and use the rescaled statistic $\widetilde{U}_{n2} = \widehat{U}_{n2}/\sigma(\hat{P}_n)$. By Theorem 4.1, rejection of the hypothesis $\widetilde{\mathrm{H}}_0$: $P \in \mathcal{Q}_0$ if $|\widetilde{U}_{n2}| \geq z_{\alpha/2}$ defines an asymptotically level α and consistent test for '$T_2(P) = 0$' against '$T_2(P) \neq 0$'.

4.2. The Exponential Distribution

As a further example, we consider the problem of testing the hypothesis of exponentiality,

$$\mathrm{H}_0: \quad P \text{ is } \exp(\lambda) \text{ for some } \lambda > 0,$$

based on an i.i.d. sample $X_1, \ldots, X_n$ of non-negative random variables having distribution P. The first nonzero component of Neyman's smooth test of H_0, based on the second Laguerre polynomial $h_2(x, \vartheta) = \frac{1}{2}(2 - 4\vartheta x + \vartheta^2 x^2)$, where $\vartheta = (\mathrm{E}X)^{-1} = \delta(P)$, is given by

$$\widehat{U}_{n2} = \frac{1}{\sqrt{n}}\left(\frac{1}{2}\sum_{j=1}^{n}\left(\frac{X_j}{\bar{X}_n}\right)^2 - n\right),$$

(Rayner & Best, 1989 Section 6.3; Koziol, 1987 p.22). Note that, up to one-to-one transformations, $\widehat{U}_{n2}$ coincides with Greenwood's (1946) statistic

$$G_n = \sum_{j=1}^{n}\left(\frac{X_j}{X_1 + \cdots + X_n}\right)^2,$$

and with the modified Shapiro–Wilk statistic

$$W_n^* = \left[1 + \frac{n+1}{n}\sum_{j=1}^{n}\left(\frac{X_j}{\bar{X}_n} - 1\right)^2\right]^{-1}$$

studied by Stephens (1978). Assuming appropriate moment conditions, it follows that $T_2(\hat{P}_n) = n^{-1/2}\widehat{U}_{n2}$, where

$$T_2(P) = \tfrac{1}{2}\Big(\frac{\operatorname{var} X}{(\mathrm{E}X)^2} - 1\Big).$$

Note that $T_2(P) = 0$ if P is $\exp(\lambda)$ for some $\lambda > 0$.

By analogy with Theorem 2.1 and Theorem 4.1, we have the following result. The easy proof is omitted.

Theorem 4.2. *If the distribution P is nondegenerate with finite fourth moment and $T_2(P) = 0$, then*

$$\widehat{U}_{n2} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(P)) \qquad (n \to \infty),$$

where

$$\sigma^2(P) = \frac{\mu_4 - 4\mu\mu_3 + 2\mu_2\mu^2 + \mu^4}{4\mu^4} = \int h_2^2(\cdot; \delta(P))\, dP.$$

In particular, under H_0, $\widehat{U}_{n2} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

Of course, the pitfall associated with a conventional interpretation of a large observed value of $|\widehat{U}_{n2}|$, *mutatis mutandi*, is the same as before. To obtain an asymptotically distribution-free consistent test for $T_2(P) = 0$ against $T_2(P) \neq 0$, one has to use the rescaled statistic $\widetilde{U}_{n2} = \widehat{U}_{n2}/\sigma(\hat{P}_n)$. Under the conditions of Theorem 4.2, the asymptotic distribution of $\widetilde{U}_{n2}$ is $\mathcal{N}(0,1)$.

4.3. The Normal Distribution

Writing $\hat{\sigma}_n^2$ for the sample variance of $X_1, \ldots, X_n$, the first nonzero component of the smooth test for the normal distribution is given by the normalised sample skewness

$$\widehat{U}_{n3} = \sqrt{\frac{n}{6}} \cdot \frac{1}{n} \sum_{j=1}^{n} \Big(\frac{X_j - \bar{X}_n}{\hat{\sigma}_n}\Big)^3 = \sqrt{\frac{n}{6}} \cdot \sqrt{b_1}$$

(see, e.g. Rayner *et al.*, 1995).

In this case, we have $\vartheta = (\mu, \sigma) = \big(\mathrm{E}X, (\operatorname{var} X)^{1/2}\big) = \delta(P)$,

$$h_3(x;\vartheta) = \frac{1}{\sqrt{6}}\Big(\Big(\frac{x-\mu}{\sigma}\Big)^3 - 3\frac{x-\mu}{\sigma}\Big),$$

and $T_3(\hat{P}_n) = n^{-1/2}\widehat{U}_{n3}$, where

$$T_3(P) = \int h_3(\cdot; \delta(P))\, dP = \frac{\mu_3}{\sqrt{6}\sigma^3}.$$

It is well known (e.g. Gastwirth & Owens, 1977; Rayner *et al.*, 1995) that for any nondegenerate P with $T_3(P) = 0$ (i.e. $\mu_3 = 0$) and finite sixth moment, we have $\widehat{U}_{n3} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(P))$, where

$$\sigma^2(P) = \int h_3^2(\cdot; \delta(P))\, dP = \frac{1}{6}\mu_2^{-3}(\mu_6 - 6\mu_2\mu_4 + 9\mu_2^3).$$

To achieve a directed diagnosis with respect to population skewness, $\widehat{U}_{n3}$ has to be rescaled by $\sigma(\hat{P}_n)$.

5. Concluding Discussion

We have seen that any component $\widehat{U}_{nj}$ of any smooth test of fit is strictly non-diagnostic when used conventionally. The reason for this is that a 'directed diagnosis' aims at testing the nonparametric hypothesis that an associated population parameter $T_j(P)$ is zero. However, this diagnosis is based solely on the distributional behaviour of the component (test statistic) for a small parametric subclass of all distributions P satisfying $T_j(P) = 0$.

The directed diagnosis is achieved, at least asymptotically (i.e. for large sample sizes), by rescaling $\widehat{U}_{nj}$ by $\sigma(\hat{P}_n)$, where $\sigma^2(P)$ is the variance of the limiting centred normal distribution for $\widehat{U}_{nj}$ which holds if $T_j(P) = 0$. Note that $\sigma^2(\hat{P}_n)$ (as an integral of a positive function with respect to $\hat{P}_n$) is positive so that negative variances do not occur.

Since the speed of convergence to the standard normal distribution for a rescaled component may depend heavily on the underlying distribution (see Table 1), diagnostic tests based on rescaled components should not be used for small samples. In this respect, a large scale simulation study would be helpful.

Regarding power of tests based on components, note that the (conventional) level α test for H_0: $P \in \{P_\vartheta : \vartheta \in \Theta\}$ based on a component $\widehat{U}_{nj}$ is inconsistent against a fixed alternative distribution from the nonparametric class $\mathcal{P}_0$ of distributions P satisfying $T_j(P) = 0$. Moreover, it is easily seen that this test has a limiting asymptotic power between α and 1 against a sequence of contiguous alternatives to $\{P_\vartheta : \vartheta \in \Theta\}$ from $\mathcal{P} \backslash \mathcal{P}_0$ (see also Rayner *et al.*, 1995, who speak of 'skewed' contiguous alternatives in the case of testing for normality).

The test for $\widetilde{\mathrm{H}}_0$: $P \in \mathcal{P}_0$ against $\widetilde{\mathrm{H}}_1$: $P \in \mathcal{P} \backslash \mathcal{P}_0$ based on the rescaled component $\widetilde{U}_{nj}$ rejects $\widetilde{\mathrm{H}}_0$ at level α if $|\widetilde{U}_{nj}| \geq z_{\alpha/2}$. This test has limiting constant size α over $\widetilde{\mathrm{H}}_0$, is consistent against each fixed alternative distribution from $\widetilde{\mathrm{H}}_1$ and has limiting power between α and 1 against a sequence of contiguous alternatives from $\mathcal{P} \backslash \mathcal{P}_0$.

References

BEST, D.J. & RAYNER, J.C.W. (1989). Goodness of fit for the geometric distribution. *Biom. J.* **31**, 307–311.

BOULERICE, B. & DUCHARME, G.R. (1995). A note on smooth tests of goodness of fit for location–scale families. *Biometrika* **82**, 437–438.

EPPS, T.W. (1995). A test of fit for lattice distributions. *Comm. Statist. A — Theory Methods* **24**, 1455–1479.

GASTWIRTH, J.L. & OWENS, M.E.B. (1977). On classical tests of normality. *Biometrika* **64**, 135–139.

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

HENZE, N. (1994). Tests auf Normalverteilung. (In German; English summary.) *Allg. Statist. Archiv* **78**, 293–317.

— (1996a). Do components of smooth tests of fit have diagnostic properties? *Metrika* **43** (to appear).

— (1996b). Empirical distribution function goodness of fit tests for discrete models. *Canad. J. Statist.* **24**, 1–13.

HORSWELL, R.L. & LOONEY, S.W. (1992). A comparison of tests for multivariate normality that are based on measures of multivariate skewness and kurtosis. *J. Statist. Comp. Simul.* **42**, 21–38.

— & — (1993). Diagnostic limitations of skewness coefficients in assessing departures from univariate and multivariate normality. *Comm. Statist. B — Simulation Comput.* **22**, 437–459.

KOZIOL, J.A. (1987). An alternative formulation of Neyman's smooth goodness of fit tests under composite alternatives. *Metrika* **34**, 17–24.

NEYMAN, J. (1937). Smooth tests for goodness of fit. *Skand. Aktuarietidskrift* **20**, 150–199.

RAYNER, J.C.W. & BEST, D.J. (1989). *Smooth Tests of Goodness of Fit.* New York: Oxford University Press.

—, — & MATHEWS, K.L. (1995). Interpreting the skewness coefficient. *Comm. Statist. A — Theory Methods* **24**, 593–600.

STEPHENS, M.A. (1978). EDF statistics for goodness-of-fit and some comparisons. *J. Amer. Statist. Assoc.* **69**, 730–737.