Distributional inference

Albers, Casper Johannes

Publication date:
2003

Chapter 4

A goodness of fit test, smoother than smooth

> *'There are two points which must be borne in mind when deciding upon the order of the test to be applied in any particular case. This, of course, is a question entirely beyond the limits of statistical theory which concerns itself only with the properties of different tests. When these properties are elucidated it is a problem for the practical statistician to choose the test, the properties of which correspond most closely to the circumstances of the problem he considers. My personal feeling is that in most practical cases, there will be no need to go beyond the fourth order test. But this is only and* opinion *and not any mathematical result.'*
>
> J. NEYMAN[1]

The theory of the previous chapter will be used to construct a test for the null hypothesis H_0: $f = \psi$. For such a test an estimate $\hat{f}$ of f is compared with ψ. The total variation distance between $\hat{f}$ and ψ is used as test statistic. As $\hat{f}$ we will use $f_n^{(m)}$ defined in Chapter 3. The choice of the smoothing parameter m is discussed and the distribution of the test statistic under H_0 is studied theoretically by using asymptotic theory, and numerically by performing simulation experiments to specify critical values. Relations with other proposals from the literature are discussed. Most results of this chapter also appear in ALBERS AND SCHAAFSMA (2003b).

4.1 Goodness of fit testing

After KARL PEARSON's breakthrough paper about the χ^2 test was published in 1900, many improvements were made and modifications were suggested. NEYMAN (1937), for example, considered continuous analogues of Pearson's problem. We concentrate the attention on such analogue.

Problem
Given are the outcomes $x_{[1]} < x_{[2]} < \ldots < x_{[n]}$ of an independent random sample $X_1, \ldots, X_n$ from a probability distribution on $\mathbb{R}$ with a 'smooth' density f, 'not unlike a given density ψ'. *Required* is a statement about the truth or falsity of the hypothesis of equality: $H_0 : f = \psi$.

[1]'Smooth' test for goodness of fit, *Skandinavisk Aktuarietidskrift*, **20**, 1937.

The statistician who has to solve this problem may be appalled by the abundance of proposals. PEARSON's test (1900) requires a classification of the data, the choice of the number of classes in particular, as well as the class borders. NEYMAN's smooth test (1937) provides the freedom (but also the task) to specify an orthonormal basis for an L_2 space, e.g. the normalized Legendre polynomials. KOLMOGOROV's test (1933)[2] is yet another possibility. The freedom to choose has a dark side, which we like to illustrate as follows. An investigator of human growth used Kolmogorov's test to see whether a normal distribution could explain his data. The test did not lead to rejection of the hypothesis. Subsequent analyses were made under normality assumptions. The investigator was criticized because an application of Pearson's χ^2 test displayed the existence of significant deviations from normality. This illustration by means of an example from practice can, of course, be supported by theoretical power considerations from a Neyman-Pearson perspective. Maximum shortcomings of level-α tests are equal to $1 - \alpha$ for a variety of nonparametric alternatives. In spite of this discomforting background, we like to study the problem because of its relevance. The present-day statistician may choose a computational approach by using some estimate of f, e.g. a kernel estimate. He will then compare the estimate and the postulated density ψ. Our method fits within this approach: we concentrate the attention on the idea to reject H_0 if the area $||\hat{f} - \psi||_1$ between the graph of ψ and that of a special nonparametric density estimate $\hat{f}$ is too large. This estimate $\hat{f}$ will be constructed by adapting the theory of Chapter 3 to the present situation, see Section 4.2. The null distribution of the test statistic $||\hat{f} - \psi||_1$ is studied to determine P-values and tables of critical values. The approach can be embedded in the general theory of goodness of fit tests which includes the testing of parametric models, e.g. that where ψ is a normal density with unspecified parameters, and the testing of nonparametric models, e.g. that f is a concave function on some finite interval or that $\log f$ is a concave function (strong unimodality), see, e.g. DÜMBGEN AND SPOKOINY (2001).

Another alternative to 'standard goodness of fit testing theory' is to use *pre-test procedures*, which may be described as follows

> 'Suppose we want to test a given hypothesis, but we doubt whether we may assume a restricted, but possibly incorrect model, or have to resort to a larger and thus less precise model. To settle this issue, we perform a preliminary test on the adequacy of the restricted model. If this test fails to reject, then we feel free to stick to the restricted model and use a test specifically suitable for that model. Otherwise, we use an alternative main test which is more general and appropriate for the large model, but less powerful than the first test when the restricted model holds.'

(BOON, 1999, see also W. ALBERS ET AL. ,2000, 2001). We shall not go into further details and restrict ourselves to 'standard' goodness of fit problems.

[2] In literature also referred to as the Kolmogorov-Smirnov test, following SMIRNOV's extension (1939)of Kolmogorov's test to the two-sample problem.

The estimate $\hat{f}$ to be used in our approach depends on the sample size n and on the degree m of a polynomial to be specified. That is why the notation $\hat{f} = f_n^{(m)}$ is used, together with $t_n^{(m)} = ||f_n^{(m)} - \psi||_1$ for the outcome of the test statistic $T_n^{(m)}$. The P-value $\mathrm{P}_0(T_n^{(m)} \geq t_n^{(m)}) = \alpha(x)$ will be used as degree of belief in H_0. Here P_0 refers to the distribution of $T_n^{(m)}$ under H_0. The complement $\varphi(x) = 1 - \alpha(x)$ of $\alpha(x)$ is the degree of belief in the falsity of H_0. Salomé et al. (1999) suggests that the degree of belief interpretation is questionable. If, however, H_0 is rejected for $\alpha(x)$ smaller than some nominal level, then one is acting according to the Neyman-Pearson theory. In practice this Neyman-Pearson approach is quite natural: subsequent analyses are based on the assumption that $f \equiv \psi$ if H_0 is maintained; if H_0 is rejected, then some estimate, in our case $f_n^{(m)}$, will be used. This in contrast to using, e.g., Pearson's test which has the drawback that 'if the null hypothesis is rejected then there is no alternative distribution indicated.' (Rayner and Best, 1989, p. 33).

Similar to Section 3.2, the probability transform $x_i \to u_i = \Psi(x_i)$ is applied to the original observations giving us

$$u_{[0]} = 0, \qquad u_{[i]} = \Psi(x_{[i]}) \quad (i = 1, \ldots, n), \qquad u_{[n+1]} = 1,$$

as the basis of the analysis. Note that $\Psi(X_i)$ has distribution function $G = F \circ \Psi^{-1}$, quantile function $B = G^{-1} = \Psi \circ F^{-1}$, density function $g(u) = f(\Psi^{-1}(u))/\psi(u)$, quantile density $b(p) = B'(p)$, etcetera. The null hypothesis H_0: $f = \psi$ is equivalent to H_0: $g \equiv 1$ and to H_0: $b \equiv 1$.

The notations $u_i = \Psi(x_i)$ and, of course, $u_{[i]} = \Psi(x_{[i]})$ are used, instead of $y_i = \Psi(x_i)$ of the previous chapter, because

1. To emphasize that under the null hypothesis the u_i constitute a sample from the standard uniform distribution;
2. To emphasize the different context we are in now.

The example

Let the data $x_{[1]}, \ldots, x_{[20]}$ be given as:

3.89,	7.44,	8.65,	9.40,	10.00,	11.27,	11.52,
14.23,	15.52,	15.63,	16.39,	17.33,	18.37,	21.12,
21.76,	22.54,	23.29,	23.36,	24.17,	24.57.	

The information is provided that the underlying density f ($f(x) = (5^{-4}x + 50^{-1})\mathbf{1}_{[0,25]}$) is such that the support $\{x; f(x) > 0\} = (0, 25)$. We wish to test H_0: $f = \psi$ where $\psi(x) = .04$ is the density of the uniform distribution on the support indicated. Using $\Psi(x) = x/25$ as notation for the distribution function of this null distribution, it is natural to apply the probability transform where x_i is replaced by

$$u_i = \Psi(x_i) = \tfrac{1}{25} x_i.$$

The true distribution of $U_i = \Psi(X_i)$ has distribution function $G = F \circ \Psi^{-1}$ where $G(u) = F(25u)$ and

$$g(u) = \frac{f(\Psi^{-1}(u))}{\psi(\Psi^{-1}(u))} = 25 f(25u).$$

We concentrate the attention on the formulations H_0: $g \equiv 1$ and especially H_0: $b \equiv 1$. If a simple alternative is considered, e.g. H_1: $g(u) = 2u$, then we can apply the Neyman-Pearson Fundamental Lemma. For this special alternative H_1 we reject H_0 if $\prod_{i=1}^n u_i$ is sufficiently large, or, equivalently, if $-2\sum \log(u_i)$ is sufficiently small. It is well known that the distribution of $-2\sum \log(U_i)$ is χ^2_{2n} if H_0 is true. The P-value

$$\mathrm{P}\left(\chi^2_{2n} \leq -2\sum \log(u_i)\right) = \mathrm{P}\left(\chi^2_{40} \leq 21.62\right)$$

thus obtained, for the example, is equal to .0078. Hence H_0 is rejected at all levels of significance $\alpha \geq .0078$.

In practice we do not know which simple alternative to choose and we are in need of an omnibus test. Using the density estimate $f_n^{(m)}$, with smoothing parameter $m = 1$, we obtain $||f_{20}^{(1)} - \psi||_1 = |\bar{u} - \frac{1}{2}| = .141$ for the example mentioned, see Figure 4.2 (page 85, left) and Section 4.3. With this choice of m the distribution of the test statistic under H_0 is obvious. $U_1, \ldots, U_n$ being an independent random sample from $\mathcal{U}(0,1)$ and using a normal approximation we obtain the P-value

$$\mathrm{P}\left(\left|\tfrac{1}{n}\sum U_i - \tfrac{1}{2}\right| > .141\right) \approx 2\Phi\left(-.141\sqrt{20 \cdot 12}\right) = .029.$$

Note that Karl Pearson's test requires the specification of the number of cells such that the χ^2_k distribution applies. If we take $k = 1$, we arrive at the two-sided sign test which, for our data, provides $P = .263$. If we take $k = 2$, then we have to work with the exact null distribution of Pearson's statistic. Computations provided $P = .14$. About the possibility of a further increase of k, KALLENBERG ET AL. (1985) wrote

> 'In a classical paper by MANN AND WALD (1942), a rule is given to let k increase with n roughly at the rate $n^{2/5}$ when using intervals with equal probability under H_0. More recent numerical work, however, has shown that for particular alternatives, a small fixed value of k often gives much better power (cf. BEST AND RAYNER, 1981)'.

Besides the choice of the number of classes k, the class borders $[l_1, r_1), \ldots, [l_k, r_k)$ have to be specified. KENDALL AND STUART (1973) suggest the borders are chosen such that one obtains the 'equiprobable χ^2 test' with the probabilities $\mathrm{P}(l_i \leq X < r_i) = k^{-1}$, $(i = 1, \ldots, k)$ under H_0, because this test is locally unbiased, in case of a simple alternative (cf. BEST AND RAYNER, 1981). Furthermore, 'when not only the frequencies ν_i but also the original observations X_i are available, reduction of the data through grouping results in tests that tend to be less efficient than those based on the Kolmogorov or related statistics' (LEHMANN, 1986, p. 480).

4.2 Specifications

To test H_0: $f = \psi$, consider the area

$$||f_n^{(m)} - \psi||_1 = ||g_n^{(m)} - 1||_1 = ||b_n^{(m)} - 1||_1$$

between the graphs of ψ and $f_n^{(m)}$. Note that the first equality follows from the fact that the L_1-norm corresponds to the total variation norm which is invariant under bimeasurable bijections (cf. DEVROYE, 1987 and DUNFORD AND SCHWARTZ, 1957) . The second equality can be established by noting that, in general, $||b-1||_1$ is equal to

$$\begin{aligned} ||b-1||_1 &= \int_0^1 |B'(p)-1| \, \mathrm{d}p \\ &= \int_0^1 \left|(G^{-1})'(p)-1\right| \, \mathrm{d}p \\ &= \int_0^1 \left|\frac{1}{g(G^{-1}(p))}-1\right| \, \mathrm{d}p \\ &= \int_0^1 \left|\frac{1}{g(u)}-1\right| \, \mathrm{d}G(u) \\ &= ||g-1||_1. \end{aligned}$$

The definition of $B_n^{(m)}$ (and, hence, of $b_n^{(m)}$, $g_n^{(m)}$, $f_n^{(m)}$, etcetera) was given in Section 3.6. Recall that the U-statistic $B_n^{(m)}$ can be written as the L-statistic

$$B_n^{(m)}(p) = p^{m+1} + \sum_{j=1}^{m} \binom{m+1}{j} p^j (1-p)^{m+1-j} \sum_{i=j}^{n-m+j} \frac{\binom{i-1}{j-1}\binom{n-i}{m-j}}{\binom{n}{m}} u_{[i]}.$$

Differentiation provides $b_n^{(m)}$. This is a convex combination of the densities of the Beta$(i+1, m+1-i)$ distributions ($i = 0, \ldots, m$), which, of course, are polynomials of degree m. Note that the $\hat{g} = g_n^{(m)}$ and $\hat{f} = f_n^{(m)}$ obtained are genuine probability density functions: they are nonnegative everywhere and integrate up to one. In the density estimation case it was suggested that m should be proportional to $\sqrt{n}$. In the present context of testing H_0: $b = 1$ some further smoothing is indicated. We recommend a choice of m not larger than 4 or 5 if an omnibus test is required, see the end of Section 4.1.

Though we are primarily interested in using $T_n^{(m)}$ with outcome

$$t_n^{(m)} = ||f_n^{(m)} - \psi||_1 = ||b_n^{(m)} - 1||_1$$

as test statistic, one can discuss some other test statistics as well, e.g. the Kolmogorov distance with outcome

$$\tilde{t}_n^{(m)} = ||F_n^{(m)} - \Psi||_\infty = ||B_n^{(m)} - p||_\infty.$$

A comparison between tests based on different dissimilarity coefficients is not performed though it might be of interest.

4.3 The extreme case $m = 1$

Ignoring the degenerate case $m = 0$ where the smoothing is so strong that $B_n^{(0)}(p) = p$ does not depend on the data, we start with $m = 1$ where

$$B_n^{(1)}(p) = 2\bar{u}p(1-p) + p^2 = (1-\bar{u})p^2 + \bar{u}(2p - p^2)$$

is a convex combination of the quantile function $2p - p^2$ of the Beta$(1, \frac{1}{2})$ distribution and the quantile function p^2 of the Beta$(\frac{1}{2}, 1)$ distribution. This, of course, does not mean that the inverse $G_n^{(1)}$ of $B_n^{(1)}$ is a convex combination as well. Note that for $\bar{u} = \frac{1}{2}$ the uniform distribution appears.

Theoretical intermezzo

It is of some theoretical interest to consider the quantile functions $B_\theta(p) = (1-\theta)p^2 + \theta(2p - p^2)$ for arbitrary $\theta \in [0, 1]$. Here $B_n^{(1)}(p)$ corresponds to $B_\theta(p)$ if $\theta = \bar{u}$. An elementary analysis provides

$$G_\theta(u) = \begin{cases} (\theta - \sqrt{\theta^2 - (2\theta - 1)u})/(2\theta - 1) & \text{if } \theta \neq 1/2 \\ u & \text{if } \theta = 1/2 \end{cases}$$

with density

$$g_\theta(u) = \frac{1}{2\sqrt{\theta^2 + (1-2\theta)u}} \qquad (0 < u < 1)$$

(for $\theta = 0$ the Beta$(1, \frac{1}{2})$ distribution is obtained, for $\theta = 1$ it is Beta$(\frac{1}{2}, 1)$).

It is possible to extend this family $\{g_\theta \mid \theta \in [0, 1]\}$ of densities by allowing negative values of θ, as well as values of $\theta > 1$. This extension, however, serves no practical purpose because we are interested in the testing of H_0: $g = 1$ and, hence, in the densities 'not too far from $g_{1/2}$'. If X_θ is a random variable with density function g_θ, then (for arbitrary $\theta \in \mathbb{R}$)

$$\begin{aligned} \mathbf{E}\, X_\theta &= \int_0^1 u g_\theta(u)\, \mathrm{d}u \\ &= \int_0^1 B_\theta(p)\, \mathrm{d}G_\theta(B_\theta(p)) \\ &= \int_0^1 B_\theta(p)\, \mathrm{d}p \\ &= \tfrac{1}{3}\theta + \tfrac{1}{3}. \end{aligned}$$

In a parametric approach to the testing of H_0: $g \equiv 1$ the attention might be concentrated on level-α tests which are 'optimal' if g belongs to some parametric family $\{g_\theta \mid \theta \in \Theta\}$. The densities just considered to test H_0: $g \equiv 1$ or, equivalently, H_0: $\theta = \frac{1}{2}$, constitute a possibility but not as convenient as the exponential densities

$$\tilde{g}_\theta \propto \mathrm{e}^{\theta u - \chi(u)},$$

because in the latter case a uniformly most powerful unbiased size-α test exists for H_0: $\theta = 0$ whereas in the earlier case only locally most powerful tests exist, and a less satisfactory criterion has to be chosen. However, the choice of this one-parameter exponential family is a convenient but subjective ‘rationalization’: if $t : [0,1] \to \mathbb{R}$ is any increasing function of u, e.g. $t(u) = \log u$, then another exponential family $\{\tilde{g}_\theta | \theta \in \Theta\}$ is obtained by specifying $\log\left(\tilde{g}_\theta(u)\right) = c(\theta) + \theta t(u)$. The uniformly most powerful unbiased size-α test is then based on $n^{-1}\sum t(u_i)$.

The definition of g_θ was based on the idea that $b_\theta(p) = (1-\theta)2p + \theta(2-2p)$ is a linear function of p if $0 \le \theta \le 1$. It is of similar interest to study the assumption $g \in \{\bar{g}_\theta | \theta \in \Theta\}$ where $\bar{g}_\theta(u) = 2\theta u + (1-\theta)(2-2u)$ and, hence, $\bar{b}_\theta = g_\theta$. If one now wants to test H_0: $\theta = \frac{1}{2}$ versus A: $\theta \ne \frac{1}{2}$ then, again, $|\bar{u} - \frac{1}{2}|$ is obtained as the basis of a locally MP unbiased but not uniformly MP test. It is interesting, from a theoretical viewpoint, to study this testing problem in detail (there will exist a most stringent size-α test, this can be compared with the most stringent somewhere MP invariant size-α test, etcetera). This, however, serves almost no practical purpose. (*End of intermezzo.*)

In Section 4.2 the test statistics $T_n^{(m)}$ and $\tilde{T}_n^{(m)}$ were defined. For $m = 1$ we have

$$\begin{aligned}
t_n^{(1)} &= ||b_n^{(1)} - 1||_1 \\
&= \int_0^1 |2(1-\bar{u})p + \bar{u}(2-2p) - 1| \, \mathrm{d}p \\
&= |2\bar{u} - 1| \int_0^1 |1 - 2p| \, \mathrm{d}p \\
&= |\bar{u} - \tfrac{1}{2}|, \quad \text{and} \\
\tilde{t}_n^{(1)} &= \sup_p \left| B_n^{(1)}(p) - p \right| \\
&= \sup_p |2\bar{u} - 1| p(1-p) \\
&= \tfrac{1}{2}|\bar{u} - \tfrac{1}{2}|.
\end{aligned}$$

Conclusion

If one agrees upon $m = 1$, then there is consensus about using the deviation of $\bar{u}$ from $\frac{1}{2}$ as test statistic. The corresponding P-value is approximately given by

$$\mathrm{P}(\chi_1^2 \le 12(\bar{u} - \tfrac{1}{2})^2) = 2\Phi(-\sqrt{12n}|\bar{u} - \tfrac{1}{2}|).$$

This test corresponds to that of Neyman (1937) if a polynomial of degree 1 is used. The test rejects H_0 if the deviation of $\bar{u}$ from $\frac{1}{2}$ is sufficiently large. The test has no power if the true density g is such that $\int ug(u) \, \mathrm{d}u = \frac{1}{2}$. For all densities with this expectation different from $\frac{1}{2}$, the power of the level-α-test is governed by the asymptotic normality of $\bar{U}$ around this expectation.

The example revisited

See Section 4.1 and note that the density $f_{20}^{(1)}$ in Figure 4.2 (left) corresponds to

$g_n^{(1)}(u) = g_\theta(u)$ with $\theta = \bar{u} = .641$ (see the beginning of this section and note that $f_{20}^{(1)}(x)$ is equal to $g_{20}^{(1)}(\Psi(x))\psi(x) = \frac{1}{25} g_{20}^{(1)}(\frac{x}{25})$). The P-value

$$\mathrm{P}_1 = 2\Phi(-\sqrt{12n}|\bar{u} - \tfrac{1}{2}|) = .0290,$$

was already mentioned in Section 4.1.

4.4 The case $m = 2$

The exact equivalence with a Neyman smooth test vanishes if $m = 2$ because then

$$\begin{aligned} B_n^{(2)}(p) &= p^3 + \frac{3p(1-p)}{\binom{n}{2}} \sum_{i=1}^{n} (n - i + p(2i - n - 1)) u_{[i]} \\ &= p^3 + 3p(1-p)\bar{u} + 3p(1-p)(p - \tfrac{1}{2})g \\ &= p + 3p(1-p)\varepsilon + 3p(1-p)(p - \tfrac{1}{2})\delta \end{aligned}$$

where

$$\begin{aligned} \bar{u} &= \tfrac{1}{n} \sum_{i=1}^{n} u_i && (U\text{-statistic}) \\ &= \tfrac{1}{n} \sum_{i=1}^{n} u_{[i]} && (L\text{-statistic}) \\ &= \tfrac{1}{2} + \varepsilon, && \text{and} \end{aligned}$$

$$\begin{aligned} g &= \tfrac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} |u_i - u_j| && (U\text{-statistic}) \\ &= \tfrac{2}{n(n-1)} \sum_{i=1}^{n} (2i - n - 1) u_{[i]} && (L\text{-statistic}) \\ &= \tfrac{1}{3} + \delta \end{aligned}$$

are respectively the sample mean and Gini's mean difference. We introduced $\varepsilon = \bar{u} - \frac{1}{2}$ and $\delta = g - \frac{1}{3}$ because, under H_0, $\mathbf{E}\ \bar{U} = \frac{1}{2}$ and $\mathbf{E}\ G = \mathbf{E}\ |U_i - U_j| = \frac{1}{3}$. Many years ago there was a discussion between Fisher, Edgeworth and Gini about the appropriateness of measures of spread like the sample standard deviation s and Gini's mean difference g. It is interesting to note that in case $m = 2$ the statistics $\bar{u}$ and g (or ε and δ) appear if one expresses $b_n^{(2)}$ as a linear combination of the polynomials 1, p, and p^2.

Note: it would be of some theoretical interest to study the distributions on $[0, 1]$ with quantile functions $\{B_\theta | \theta \in \Theta\}$ where B_θ is obtained from the expression of $B_n^{(2)}$ by replacing (ε, δ) by $\theta \in (\theta_1, \theta_2)$. As a similar study for $m = 1$ suggests (see the theoretical intermezzo in Section 4.3) no results of practical interest are expected, and that is why this study is not performed here. To continue, we need the following result:

Theorem 4.1 *Under* H_0*:* $f = \psi$ *we have*

$$\mathcal{L}\, n^{1/2} \begin{pmatrix} \varepsilon \\ \delta \end{pmatrix} \longrightarrow \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma^2 & \rho \\ \rho & \tau^2 \end{bmatrix} \right),$$

with $\sigma^2 = \frac{1}{12}$*,* $\tau^2 = \frac{1}{45}$*, and* $\rho = 0$ *and exact equalities for* $\mathrm{Var}\,(\varepsilon) = \frac{1}{n}\sigma^2$ *and* $\mathrm{Cov}\,(\varepsilon, \delta) = 0$.

For the proof see Appendix A. It follows from this theorem that under H_0

$$\mathcal{L}\, n\left(12\varepsilon^2 + 45\delta^2\right) \rightarrow \chi_2^2 = \mathrm{Gamma}(1, \tfrac{1}{2}),$$

and that, hence, using any positive multiple of $12\varepsilon^2 + 45\delta^2$ as test statistic, the approximate P-value

$$\mathrm{P}_2^{(1)} = \mathrm{P}(\chi_2^2 \leq n(12\varepsilon^2 + 45\delta^2)) = \exp(-n(3\varepsilon^2 + 11.25\delta^2))$$

is obtained. Though it is convenient to use $12\varepsilon^2 + 45\delta^2$ as test statistic, we prefer to concentrate the attention on the test statistic $T_n^{(2)}$ with outcome

$$t_n^{(2)} = ||b_n^{(2)} - 1||_1 = 3\int_0^1 \left|-3\delta p^2 + (3\delta - 2\varepsilon)p + (\varepsilon - \tfrac{1}{2}\delta)\right| \,\mathrm{d}p.$$

In practice, the computer will be used to obtain $t_n^{(2)}$. To increase the understanding, an approximation is obtained by using

$$t_n^{(2)} = \begin{cases} 3^{-1/2}|\delta| & \text{if } \varepsilon = 0 \\ \frac{3}{2}|\varepsilon| & \text{if } \delta = 0 \\ \frac{16|\varepsilon|}{9} = \frac{3|\varepsilon|}{2} + \frac{5|\delta|}{36} & \text{if } \delta = 2\varepsilon \\ \frac{8|\varepsilon|}{3\sqrt{3}} = \frac{3|\varepsilon|}{2} + \frac{(16\sqrt{3}-27)|\delta|}{12} & \text{if } \delta = \frac{2\varepsilon}{3} \end{cases}$$

to suggest the approximation

$$t_n^{(2)} \approx \sqrt{\tfrac{9}{4}\varepsilon^2 + \tfrac{1}{3}\delta^2}$$

which, actually, is an upper bound, see Figure 4.1. We conclude that in this statistic much more weight is assigned to ε than in the χ_2^2-statistic derived earlier because

$$\tfrac{9}{4}\varepsilon^2 + \tfrac{1}{3}\delta^2 = \tfrac{1}{3}(\tfrac{27}{4}\varepsilon^2 + \delta^2)$$

while

$$12\varepsilon^2 + 45\delta^2 = 45(\tfrac{12}{45}\varepsilon^2 + \delta^2).$$

and $\frac{27}{4} \gg \frac{12}{45}$.

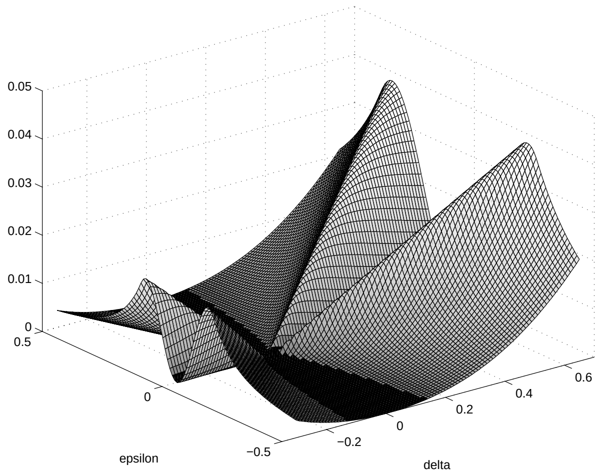

Figure 4.1: Difference $\sqrt{\frac{9}{4}\varepsilon^2 + \frac{1}{3}\delta^2} - t_n^{(2)}$*, for all theoretical possibilities of* (ε, δ)*.*

Note that the difference between its approximate value and $t_n^{(2)}$ is always nonnegative and that the approximation is very good if (ε, δ) is in the neighborhood of its expected value $(0, 0)$ under H_0.

The computation of the P-value $\mathrm{P}_0(T_n^{(2)} \geq t_n^{(2)})$ will be done on basis of an extensive numerical simulation as well as on the basis of the approximation where $\frac{9}{4}\varepsilon^2 + \frac{1}{3}\delta^2$ is used as test statistic and, for the sake of convenience, its distribution under H_0 is approximated by that of $c_n\chi_1^2$ with $c_n = (\frac{9}{4}\frac{1}{12} + \frac{1}{3}\frac{1}{45})n^{-1} = .195n^{-1}$ such that the (asymptotic) expectations coincide. The χ_1^2 distribution is used for reasons of simplicity and because $\frac{9}{4}\varepsilon^2$ is the dominant term. This provides the approximate P-value

$$\mathrm{P}_2^{(2)} = \mathrm{P}(\chi_1^2 \leq c_n^{-1}(\tfrac{9}{4}\varepsilon^2 + \tfrac{1}{3}\delta^2)) = 2\Phi\left(-\sqrt{c_n^{-1}(\tfrac{9}{4}\varepsilon^2 + \tfrac{1}{3}\delta^2)}\right).$$

The exact distribution of $T_n^{(2)}$ under H_0 is simulated as explained in Appendix B. Table B.1 provides critical values for $t_{n,\alpha}^{(2)}$. A numerical elaboration is presented in the following section.

The example, revisited
With respect to the example of Section 4.1 we have $\bar{u} = .64$, $g = .30$, $\varepsilon = .141$, $\delta = -.038$. The χ_2^2-test provides the approximate P-value

$$\mathrm{P}_2^{(1)} = .0231.$$

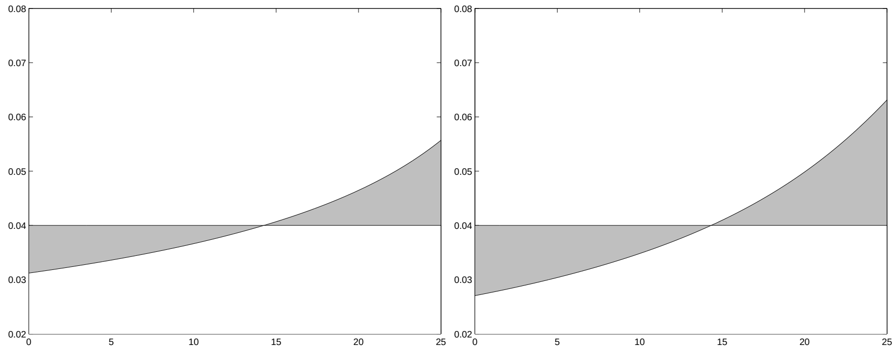

Figure 4.2: Density estimates (curves) of the data in the example with $m = 1$ (left) and $m = 2$ (right) and the uniform null-distribution. The test areas indicated are .141 ($m = 1$) and .212 ($m = 2$).

The test based on $\frac{9}{4}\varepsilon^2 + \frac{1}{3}\delta^2$ provides

$$\mathrm{P}_2^{(2)} = 2\Phi\left(-\sqrt{\tfrac{20}{.195}(\tfrac{9}{4}.141^2 + \tfrac{1}{3}.038^2)}\right) = .0313,$$

while the approximation $(\frac{9}{4}\varepsilon^2 + \frac{1}{3}\delta^2)^{1/2} = .212$ is very close to the shaded area $t_n^{(2)} = .212$ in Figure 4.2 (right). Note that $\varepsilon = .141$ is more than twice as large as its standard deviation $(\frac{1}{20} \times \frac{1}{12})^{1/2} = .065$ whereas $\delta = -.038$ (in absolute value) is comparable to its standard deviation $(\frac{1}{20} \times \frac{1}{45})^{1/2} = .033$.

Using Table B.1 in the appendix, it follows from $t_n^{(2)} = .212$ that the corresponding P-value $\mathrm{P}_2^{(3)}$ is slightly larger that .025. Note that $t_{20,.025}^{(2)} = .216$ while $t_{20,.05}^{(2)} = .190$; a linear interpolation provides

$$\mathrm{P}_2^{(3)} \approx .029.$$

4.5 The general case $m \geq 2$

In principle, the results of the previous two sections can be generalized to $m = 3, 4, \ldots$. In Section 4.4 exact representations, and approximations, were given in terms of $\varepsilon = \bar{u} - \frac{1}{2}$ and $\delta = g - \frac{1}{3}$, but simulations were needed to study the exact distribution of the test statistic $T_n^{(2)}$ under H_0. In the case $m = 3$ it is possible to define a special weighted sum h of $\sum i^2 u_{[i]}$, $\sum i u_{[i]}$, and $\sum u_{[i]}$, and deviation $\eta = h - \mathbf{E}\ h$, such that $\frac{1}{n}(\varepsilon, \delta, \eta)$ has a trivariate asymptotically normal distribution, probably without correlations, with $b_n^{(3)}$ a function of $\varepsilon, \delta, \eta$. To be specific,

$$\begin{aligned} B_n^{(3)} &= p^4 + 2\binom{n}{3}^{-1} p(1-p) \times \Bigg[(5p^2 - 5p + 1)\sum_{i=0}^{n+1} i^2 u_{[i]} \\ &\quad +(-5(n+1)p^2 + (7n+1)p - 2n + 1)\sum_{i=0}^{n+1} i u_{[i]} \\ &\quad + ((n^2 + 2n + 2)p^2 - n(2n-1)p + n(n-1))\sum_{i=0}^{n+1} u_{[i]}\Bigg]. \end{aligned}$$

Clearly, the results are too complicated to be useful. Anyway, the distribution of the test statistic has to be studied via simulation and we suggest to rely on the simulation results in Tables B.2 and B.3, regardless the complicatedness of $B_n^{(3)}$. Therefore, this lack of concreteness is a nuisance, but it does not affect the general line of thought.

Theoretical considerations
In Section 4.2 the symmetrized $B_n^{(m)}(p)$ (U-statistic) has been rewritten as an L-statistic. It was noted that the statistic $T_n^{(m)}$ we are interested in has as its outcome

$$t_n^{(m)} = ||f_n^{(m)} - \psi||_1 = ||b_n^{(m)} - 1||_1$$

This implies that stochastic theory about the distribution of $b_n^{(m)}(p) - 1$ is needed, especially under H_0: $b \equiv 1$. In principle, theory on U- and L-statistics is applicable, for p fixed. The results of Section 3.3 suggest that the distribution of $m^{-1/4}n^{1/2}T_n^{(m)}$ is 'fairly stable' under H_0. This can be used to predict results of the simulation experiments.

For $m = 1$ it is known from Section 4.3 that the distribution of $m^{-1/4}n^{1/2}T_n^{(1)} = n^{1/2}|\bar{u} - \frac{1}{2}|$ is, under H_0, approximately (asymptotically) like $\frac{1}{\sqrt{12}}\chi = |\mathcal{N}(0, \frac{1}{12})|$. For $m = 2$ it is known from Section 4.4 that, under H_0, the distribution of $m^{-1/4}n^{1/2}T_n^{(2)} \approx 2^{-1/4}n^{1/2}(\frac{9}{4}\varepsilon^2 + \frac{1}{3}\delta^2)^{1/2}$ is not much different from that of $\sqrt{T}$ where T has the negative-exponential distribution. These approximations provide confidence in the simulation results reported in Appendix B.

Numerical elaboration
Figure 4.3 shows for $m = 2$, 3, 4 that the relation between the average (logarithm of) $t_{n,\alpha}^{(m)}$ and the (logarithm of the) sample size is approximately linear, with slopes independent of the choice of m and α. Postulating a common slope β, this slope can be estimated, as well as the intercepts. After that, approximate P-values for can be derived using these values. However, since the intercepts depend on both m and α in a complicated way, Tables B.1, B.2, and B.3 cannot be dispensed with.

The choice of m
The tables provided in Appendix B (and the obvious result for $m = 1$) enable us to test H_0 on the basis of $T_n^{(m)}$ $(m = 1, 2, 3, 4)$. In principle a choice of m has to be made

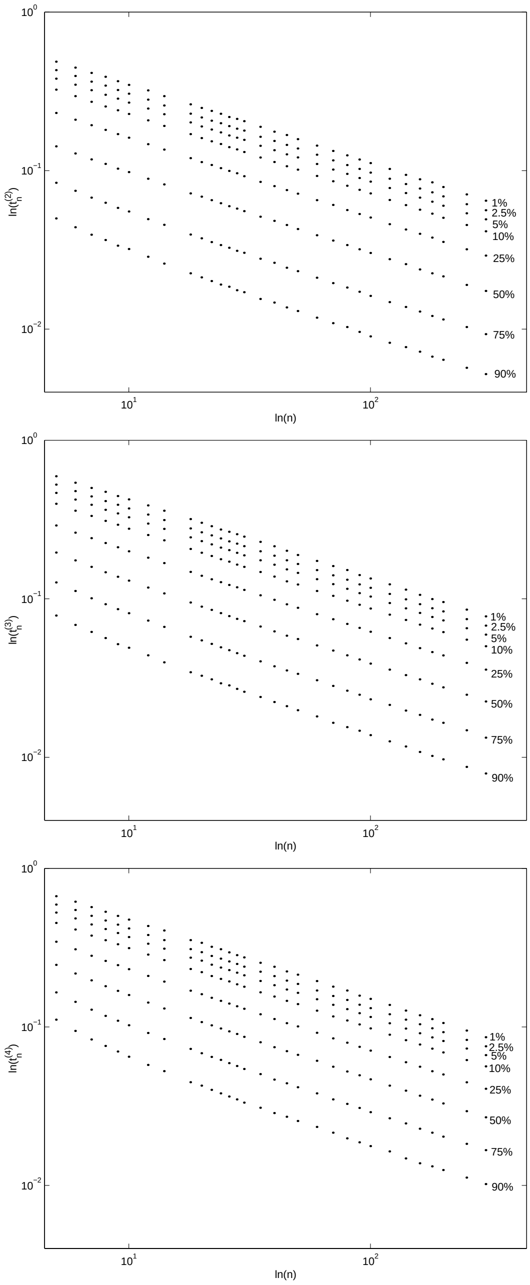

Figure 4.3: Display of simulation results, each dot represents an average $\log\left(t_{n,\alpha}^{(m)}\right)$ *(see Appendix B) for a specific choice of* $\log n$ *and* α*, with respectively* $m = 2$*, 3, 4.*

if one wants to comply with the Neyman-Pearson approach. At the end of Section 4.1 reference has been made to suggestions in the literature that (for similar problems) a small value of m is indicated. For the sample size $n = 20$ of our example, perhaps $m = 1$ or 2 is appropriate. For $n = 100$ the choice $m = 3$ or 4 may be attractive. We content ourselves with such suggestions because, in practice, the statistician will not worry too much about the effect of applying various values of m. In our opinion it is reasonable to discuss the deviations $\varepsilon = \bar{u} - \frac{1}{3}$ and $\delta = g - \frac{1}{3}$ as well as $t_n^{(1)}$, $t_n^{(2)}$, and, perhaps, if $m = 1$ and $m = 2$ do not lead to rejection of H_0, $t_n^{(3)}$ and $t_n^{(4)}$ as well. Further rationalization will not be effective.

4.6 Relation with Neyman's smooth tests

The quantile density estimate $b_n^{(m)}(p) = \frac{\partial}{\partial p} B_n^{(m)}(p)$ is a *positive* polynomial function on $[0, 1]$ of degree m. In fact, $b_n^{(m)}$ is a *convex* combination (with positive weights expressible as L-statistics) of the densities of the $\mathrm{Beta}(i + 1, m + 1 - i)$ distributions $(i = 0, \ldots, m)$ which, of course, are polynomials of degree m. This representation is very fortunate, because it implies that the $b_n^{(m)}$ and, hence, the density estimates $g_n^{(m)}$ and $f_n^{(m)}$ are genuine probability densities.

Nevertheless, it may be appealing to the mathematician to discuss an alternative basis of $L_2[0, 1]$, e.g. that of orthogonal polynomials or of other functions (e.g. trigonometric polynomials). This is of particular interest if we are discussing the density $g = G'$ of $U_1 = \Psi(X_1)$ (with $G = F \circ \Psi^{-1} = B^{-1}$) and the way it 'should be estimated'.

The discussion about the choice of basis is included, partly because it is related to the specification of NEYMAN's smooth tests (1937). For the terminology and motivation behind this method, we refer to RAYNER AND BEST (1989). Let $\varphi_0 \equiv 1, \varphi_1, \varphi_2, \ldots$ be any system of linearly independent squared integrable functions on $[0, 1]$. The Gram-Schmidt orthogonalization process provides the orthonormal basis $\psi_0, \psi_1, \psi_2, \ldots$ of (a subspace of) $L_2[0, 1]$

$$\varphi_0 \equiv \psi_0 \equiv 1, \quad \varphi_1 = \frac{\psi_1 - (\psi_1, \psi_0)\psi_0}{||\psi_1 - (\psi_1, \psi_0)\psi_0||}, \quad \ldots$$

If a function $h \in L_2[0, 1]$ (a quantile density or a probability density or some other function) can be written as a linear combination of $\varphi_0, \ldots, \varphi_k$ then it can equally well be written as a linear combination of $\psi_0, \ldots, \psi_k$. Now a variety of related paradigms can be discussed.

Paradigm 1
Focussing on the quantile densities, and starting from the estimate $b_n^{(m)}$ with $m = k$, we can consider $\varphi_h(p) = p^h$ $(h = 0, \ldots, k)$ and determine the weights $w_{n,h}$ $(h = 0, \ldots, k)$ such that $b_n^{(k)}(p) = \sum_{h=0}^{k} w_{n,h} p^h$. Section 4.3 shows that, in case $k = 1$,

$$w_{n,0} = 2\bar{u}, \qquad w_{n,1} = 2(1 - 2\bar{u}).$$

Section 4.4 provides for $k = 2$ that

$$w_{n,0} = 1 + 3(\varepsilon - \tfrac{1}{2}\delta), \qquad w_{n,1} = -6\varepsilon + \tfrac{9}{4}\delta, \qquad w_{n,2} = -9\delta.$$

For higher m expressions can be obtained as well using the Gram-Schmidt process. The formulas are too complicated to be given. Under H_0: $b \equiv 1$, the ideal weight $w^*_{n,0} = 1$ and the other weights are 0. The deviations from the ideal weights are 2ε and -42ε in the case $k = 1$; $-3(\varepsilon - \frac{1}{2}\delta)$, $-6\varepsilon + \frac{9}{4}\delta$, and -9δ in the case $k = 2$, etcetera. The first deviation is a linear combination of the other ones. Hence it is reasonable to use $w_{n,1}, \dots, w_{n,k}$ as the basis of a test. This can be a χ^2_k test, but there are arguments to pay more attention to $w_{n,1}$ than to $w_{n,2}$, etcetera. That is what has been done in Section 4.5.

Paradigm 2

Focussing on the probability densities, Neyman (1937) provided a unified approach by choosing a number k and corresponding basis functions $\varphi_0, \dots, \varphi_k$, preferably the corresponding orthonormal ones $\psi_0, \dots, \psi_k$. Important examples are that where $\varphi_h(y) = y^h$ $(h = 0, \dots, k)$ and that where $\varphi_1, \dots, \varphi_k$ are the indicator functions of the subintervals $(p_0, p_0 + p_1], (p_0 + p_1, p_0 + p_1 + p_2], \dots, (1 - p_k, 1]$ of $[0, 1]$ to test H_0: $g \equiv 1$, the last approach providing Pearson's χ^2_k test, with approximate P-value

$$\mathrm{P}\left(\chi^2_k \geq \sum_{j=0}^{k} \frac{(n_j - np_j)^2}{np_j}\right)$$

where n_j is the number of observations in cell j. Regarding the choice of the number of components k in Neyman's test, Rayner and Best (1989) states that '$k \leq 4$ will usually suffice'. See Inglot et al. (1990, 1994), and Kallenberg et al. (1985) for extensive analyses in this respect.

Given such system $\varphi_0, \dots, \varphi_k$ and being interested in the testing of H_0: $g \equiv 1$, we can consider the random variables $(\varphi_0(U_i) \equiv 1), \varphi_1(U_i), \dots, \varphi_k(U_i)$ and their expectations and (co)variances under H_0. We have

$$\begin{aligned} \mu_j &= \mathbf{E}_0 \varphi_j(U_i) = \int_0^1 \varphi_j(u)\,\mathrm{d}u \\ \sigma_{jh} &= \mathrm{Cov}_0(\varphi_j(U_i), \varphi_h(U_i)) = \int_0^1 (\varphi_j(u) - \mu_j)(\varphi_h(u) - \mu_h)\,\mathrm{d}u \end{aligned}$$

$(j, h = 1, \dots, k)$. The random variables $U_1, \dots, U_k$ actually observed have density g not necessarily equal to that of $\mathcal{U}(0, 1)$. The true expectations $\mathbf{E}\ \varphi_j(U_i)$ may deviate from their values μ_j under H_0. Unbiased estimates are provided by the sample means $T_j = n^{-1} \sum_{i=1}^{n} \varphi_j(U_i)$. The Multivariate Central Limit Theorem implies that $(T_1, \dots, T_k)'$ is asymptotically normally distributed, under H_0, with expectation $\mu = (\mu_1, \dots, \mu_k)$ and covariance matrix $\frac{1}{n}\Sigma = (\frac{1}{n}\sigma_{jh})$. (Note that this covariance matrix is equal to the inverse of Fisher's information matrix.) Hence

$$n(T - \mu)'\Sigma^{-1}(T - \mu) \sim \chi^2_k$$

and an approximate P-value is given by $\mathrm{P}\left(\chi_k^2 \geq n(t-\mu)'\Sigma^{-1}(t-\mu)\right)$.

In the situation of classified data considered by him, Karl Pearson regarded this P-value as a 'fairly reasonable criterion of the probability that the deviations $t_j - \mu_j$ can be supposed to have arisen from random sampling', i.e. of the probability than H_0 is true. This formulation allows some modification of P. In fact, it suggests that a recalibration might be considered. In this respect SALOMÉ ET AL. (1999) is of interest: if P is very small, say 10^{-4}, then it would not be wise to bet on the basis of the odds ratio $(1-p)/p$ that H_0 is false, there is a tendency to display 'overconfidence'. If one uses the orthonormal system $\psi_0, \ldots, \psi_k$ instead of $\varphi_0, \ldots, \varphi_k$ then one has

$$\begin{aligned} \mu_j &= \mathrm{E}_0\, \psi_j(U_i) = 0 \\ \sigma_{jh} &= \mathrm{Cov}_0\, (\psi_j(U_i), \psi_h(U_i)) = (\psi_j, \psi_h) = \mathbf{1}_{j=h} \end{aligned}$$

One will then obtain the P-value

$$\mathrm{P}\left(\chi_k^2 \geq \sum_{j=1}^{k} \frac{1}{n}\left(\sum_{i=1}^{n} \psi_j(U_i)\right)^2\right)$$

which, of course, is in *exact* agreement with the expression derived before.

As it is natural to expect that the 'deviations' $\frac{1}{n}\sum_{i=1}^{n} \phi_j(U_i)$ are of decreasing importance if j increases, it is also natural to replace the test statistic used by some weighted sum

$$\sum_{j=1}^{k} \frac{w_{n,j}}{n}\left(\sum_{i=1}^{n} \psi_j(U_i)\right)^2$$

the weights $w_{n,1} > w_{n,2} > \ldots$ being decreasing. It has been established in SCHAAFSMA AND STEERNEMAN (1981) that this may lead to a qualitative improvement. The choice of weights, however, is a delicate issue. That is why we keep contenting ourselves with the discussion presented at the end of Section 4.5.

Paradigm 3

Returning to the problem of estimating the probability density function, Neyman played with the idea that any density g on $[0,1]$ can be approximated by its projection on the $k+1$ dimensional subspace spanned by $\varphi_0, \ldots, \varphi_k$ or, equivalently, by $\psi_0, \ldots, \psi_k$. His line of thought was, more or less, that

$$\hat{g} = 1 + \sum_{j=1}^{k}\left(n^{-1}\sum_{i=1}^{n} \psi_j(u_i)\right)\psi_j$$

is a nice estimate of g, especially because it is an unbiased estimate of g if g itself corresponds to its projection $1 + \sum_{j=1}^{k}(g, \psi_j)\psi_j$ on the $k+1$ dimensional subspace indicated. A serious drawback of these estimates is that negative values may arise.

This difficulty does not appear if we use the estimates of the form $b_n^{(m)}$, neither will it appear if we follow, e.g., the *maximum entropy* approach of JAYNES (1996) which, in the present context, is as follows (for a short discussion on the method of maximum entropy, see Section 3.8).

Suppose that we have estimates

$$\hat{\mu}_j = n^{-1} \sum_{i=1}^{n} \psi_j(u_i)$$

of the expectations $\mu_j = \mathbf{E}\ \psi_j(U_i)$ and that we are interested in the true density g of U_i $(i = 1, \ldots, n)$. Our estimate $\hat{g}$ of g 'should' satisfy the restrictions $\int_0^1 \psi_j(u)g(u)\ \mathrm{d}u = \hat{\mu}_j,\ (j = 1, \ldots, k)$ and be such that the Shannon entropy

$$-\int_0^1 g(u) \log(g(u))\mathrm{d}u$$

is maximum. The solution to this optimalization problem is, somewhat surprisingly, that $\hat{g} = g_{\hat{\theta}}$ where

$$g_\theta(u) = \exp\left(\theta_1\psi_1(u) + \ldots + \theta_k\psi_k(u) - c(\theta)\right)$$

defines an exponential family and $\hat{\theta}$ is the maximum likelihood estimate of θ. The P-value thus obtained for testing H_0: $g \equiv 1$, or 'equivalently' for testing H_0: $\theta = 0_k$, is given by

$$\mathrm{P}\left(\chi_k^2 \geq \tfrac{1}{n}\hat{\theta}'\hat{\theta}\right).$$

Conclusion

Though the arguments are varying, most methods discussed have in common that H_0: $g \equiv 1$ is tested by choosing k linearly independent functions $\varphi_1, \ldots, \varphi_k : [0,1] \to \mathbb{R}$ in addition to $\varphi_0 \equiv 1$, considering the corresponding orthonormal basis $\psi_0 \equiv 1, \psi_1, \ldots, \psi_k$, and by applying the χ_k^2 test indicated. Many authors have discussed the choice of the number k. KARL PEARSON (1900) himself stated

> 'Thus, if we take a very great number of groups our test becomes illusory. We must confine our attention in calculating P to a finite number of groups, and this is undoubtedly what happens in actual statistics. The number k of degrees of freedom will rarely exceed 30, often not greater than 12.'

Later generations of statisticians dealing with Neyman's smooth tests have made recommendations that k should not be larger than 2 or 3 (see the citation by Kallenberg in Section 4.1 and the remark and references earlier in this section).

The discussion here is about whether or not some additional basis vectors, say $e_{k+1}, \ldots, e_{k+l}$ should be involved in the analysis. If they are involved then they get equal rights and the χ_k^2 test is replaced by a χ_{k+l}^2 test. Authors like SCHAAFSMA AND STEERNEMAN (1981) indicated that a weighted combination, with more weight on ψ_i than on ψ_{i+1} is appropriate if considerable deviations ψ_j are less likely if j is larger.

4.7 Relations with other goodness of fit tests

We are fascinated by the total-variation (or L_1) distance $||f-\psi||_1$ and the Kolmogorov distance $||F-\Psi||_\infty$ because the total-variation distance is invariant under bijective mappings while the Kolmogorov distance is invariant under monotonous transformations. Under certain monotonicity assumptions we have that $||f-\psi||_1 = 2||F-\Psi||_\infty$. We always have $||f-\psi||_1 \le 2||F-\Psi||_\infty$ (see, e.g., LOÈVE, 1955). Both distances are such that they do not change if the distribution functions $G = F \circ \Psi^{-1}$ are replaced by the corresponding quantile functions. Our test statistic $\tilde{T}_n^{(m)} = ||F_n^{(m)} - \Psi||_\infty$ has been obtained by replacing the unknown true quantile function B in $||F-\Psi||_\infty = ||B-1||_\infty$ by the corresponding estimate $B_n^{(m)}$ which is a continuous and increasing analogue of the empirical quantile function. KOLMOGOROV's test (1933) is based on $||\hat{B}-1||_\infty$ where $\hat{B}$ is the empirical quantile function. As the true quantile function is smooth, the estimates $B_n^{(m)}$ will be closer to the truth, on the average, than the discontinuous functions $\hat{B}$ on which they are based. It is also reasonable to expect that the power properties of the tests based on $||f_n^{(m)} - \psi||_1$ and $||F_n^{(m)} - \Psi||_\infty$ are better than those based on Kolmogorov's test. Much will depend, however, on the alternative hypotheses to be considered and on the choice of m to be made. A possibility for further research is to perform, by using extensive simulation studies, a concrete power comparison between the $||f_n^{(m)} - \psi||_1$-statistics $(m = 1,\ 2,\ 3)$ in the line of the power comparison in INGLOT ET AL. (1994).

A more delicate issue is as follows. If one accepts that the context asks for a test statistic of the form $||\hat{f}-\psi||_1$ then the question arises which nonparametric density estimate $\hat{f}$ one should use. In DE BRUIN ET AL. (1999) it was made very clear that the estimator f_n studied there is 'not unreasonable though some further improvement is possible'. Such improvement can be achieved

1. By using $f_n^{(m)}$ instead of $f_n = f_n^{(n)}$; or
2. By using a kernel estimator k_n, preferably with the bandwidth determined such that the method is optimal for estimating ψ itself. (Note that ψ is given.)

The comparison between the tests based on the statistics $||f_n^{(m)} - \psi||_1$ and $||k_n - \psi||_1$ will depend on a large number of arguments, e.g. the choice of m and of the basic kernel, the alternative hypotheses for which power comparisons are made, etcetera. *An argument in favor of* $||f_n^{(m)} - \psi||_1 = ||h_n^{(m)} - 1||_1$ *is that the distribution of the test statistic under* H_0 *does not depend on* ψ (critical values can be found in Appendix B). For the test statistics $||k_n - \psi||_1$ additional simulation studies would be needed.