APPLICATIONES MATHEMATICAE
36,4 (2009), pp. 375–395

Tadeusz Inglot and Alicja Janic (Wrocław)

HOW POWERFUL ARE DATA DRIVEN SCORE TESTS FOR UNIFORMITY

Abstract. We construct a new class of data driven tests for uniformity, which have greater average power than existing ones for finite samples. Using a simulation study, we show that these tests as well as some "optimal maximum test" attain an average power close to the optimal Bayes test. Finally, we prove that, in the middle range of the power function, the loss in average power of the "optimal maximum test" with respect to the Neyman–Pearson tests, constructed separately for each alternative, in the Gaussian shift problem can be measured by the Shannon entropy of a prior distribution. This explains similar behaviour of the average power of our data driven tests.

1. Introduction. Nonparametric tests play an important role in statistical inference. Usually the main difficulty in constructing good nonparametric tests is connected with the infinite-dimensionality of the set of alternatives. It is well known that, for a fixed sample size, increasing the dimension of a ball of alternatives results in a low power of any test outside some subspace (see e.g. Janssen, 2000). In recent years the very promising idea of data driven score tests has been developed by Bickel and Ritov (1992), Eubank and LaRiccia (1992), Ledwina (1994), Kallenberg and Ledwina (1995), Fan (1996), Kallenberg and Ledwina (1997), Kallenberg and Ledwina (1999), Janic-Wróblewska and Ledwina (2000), Inglot and Janic-Wróblewska (2003), Claeskens and Hjort (2004), Fromont and Laurent (2006), Langovoy (2008) and Wyłupek (2008), to mention only some of the articles published. Data driven tests are two-step constructions. In the first step a model (from a

2000 *Mathematics Subject Classification*: 62G10, 62B10, 62C10.
Key words and phrases: testing uniformity, data driven score test, selection rule, optimal Bayes test, maximum test, Gaussian shift problem, Shannon entropy, Monte Carlo simulations.

DOI: 10.4064/am36-4-1

given list) is chosen by some selection rule and in the second step a (good) testing procedure is applied using the model selected. The most popular selection rules are Schwarz's BIC and Akaike's AIC. Some indications for the choice of BIC and AIC in such problems are discussed in Inglot and Ledwina (2006a). Focusing on the problem of testing uniformity and BIC type selection rules, starting with the paper of Ledwina (1994), through Inglot and Ledwina (1996), Inglot, Kallenberg and Ledwina (1998) up to Inglot and Ledwina (2001), several asymptotic optimality properties of data driven tests have been shown. However, it is still unclear how much improvement in power is possible for finite samples.

For simplicity, in this article we restrict ourselves to the problem of testing uniformity. Our aim is threefold.

The first (and main) aim is to propose a new class of data driven tests which are more flexible and have greater average power for finite samples than existing ones. The idea of our construction comes from a paper of Inglot and Ledwina (2006a) and generalizes the approach given there. In that paper a selection rule was built from BIC and AIC type rules by some thresholding procedure which led to a clear improvement in power. The advantages of this construction were confirmed for other testing problems such as testing in regression models (Inglot and Ledwina, 2006b) or testing in the k-sample problem (Wyłupek, 2008). The threshold was based on the magnitude of the maximal empirical Fourier coefficient. For "mixed" alternatives, which do not have one dominating Fourier coefficient, such a solution is unsatisfactory. Our new solution resolves this problem by deriving thresholds which are sensitive for both "simple" ("pure") and "mixed" alternatives.

The second aim is to show using simulations that, for a finite set of orthogonal alternatives, the tests proposed have average power almost as great as the optimal Bayes test. The empirical average powers of these new tests are also compared with those of the Neyman–Pearson tests against single alternatives which correspond to priors degenerating to one-point distributions. It can be observed that the maximal loss in average power of our tests with respect to the Neyman–Pearson test is close to the Shannon entropy of the prior distribution.

Finally, to explain this phenomenon, we consider in the Appendix an optimal maximum test based on weighted empirical Fourier coefficients with weights chosen for a given prior distribution. We prove that, in the middle range of the power function, with a finite set of orthogonal alternatives the loss in average power for this optimal maximum test in the two-sided Gaussian shift problem is measured by the Shannon entropy of the prior distribution. The connection between the average empirical power and the entropy of the prior distribution for moderate and large sample sizes can be explained by the facts that the optimal Bayes test and our data driven tests both attain average power close to the optimal maximum test and the empirical Fourier coefficients have an approximately normal distribution, leading in this way to the limiting Gaussian shift problem.

The paper is organized as follows. In Section 2 we construct selection rules and corresponding test statistics. Moreover, we state Proposition 1 and Theorem 2, which establish the consistency of the new selection rules and asymptotic null distribution of the test statistics. In Section 3 we report the results of some simulation experiments. Section 4 contains the proof of Proposition 1. In the Appendix we define the optimal Bayes test (implemented in Section 3) and study the power behaviour of the above mentioned optimal maximum test in the two-sided Gaussian shift problem.

2. Selection rules and test statistics. Let $\underline{X} = (X_1, \ldots, X_n)$ be a sample from an absolutely continuous distribution P on the interval $[0,1]$ with density p. The null hypothesis H_0 asserts that $p = p_0$, where $p_0(x) = 1$ for all $x \in [0,1]$. Throughout this section, P_0 will denote the uniform distribution over $[0,1]$, and E_0 the expectation with respect to P_0.

Let $b_1, b_2, \ldots$ be an orthonormal system of bounded functions in $L_2[0,1]$ such that $E_0 b_j(X_1) = 0$. Embed p_0 into a k-dimensional exponential family $\mathcal{P}_k$ of densities given by

$$p_k(x,\theta) = p_0(x) c_k(\theta) \exp\Big\{\sum_{j=1}^{k} \theta_j b_j(x)\Big\}, \tag{1}$$

where $\theta = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k$ and $c_k(\theta)$ is a normalizing factor. Testing H_0 within $\mathcal{P}_k$ is equivalent to testing that $\theta = 0$. The score statistic for this reduced problem takes the form

$$N_k = \sum_{j=1}^{k} n\hat{b}_j^2, \quad \text{where} \quad \hat{b}_j = \frac{1}{n}\sum_{i=1}^{n} b_j(X_i), \quad j = 1, \ldots, k. \tag{2}$$

The choice of the dimension k of the family $\mathcal{P}_k$ is crucial for the behaviour of the goodness of fit test based on N_k. So, data based selection of a proper dimension is desirable. Two selection rules: simplified AIC and simplified BIC are often applied. In our testing problem they can be defined as follows. Let $d(n) \geq 1$ be the maximal dimension of the model we allow. Then simplified AIC is given by

$$A1 = \min\{1 \leq k \leq d(n) : N_k - 2k \geq N_j - 2j,\ j = 1, \ldots, d(n)\}$$

and simplified BIC by

$$S1 = \min\{1 \leq k \leq d(n) : N_k - k\log n \geq N_j - j\log n,\ j = 1, \ldots, d(n)\}.$$

Recall that the original AIC (Akaike, 1974) and BIC (Schwarz, 1978) are based on maximized loglikelihood for (1) and under local alternatives are asymptotically equivalent to $A1$ and $S1$, respectively. The resulting tests based on N_{A1} and N_{S1} are examples of data driven score tests. Both tests have nice optimality properties e.g. in the sense of asymptotically vanishing shortcoming (Inglot and Ledwina, 2001, Kallenberg, 2002). However, their behaviour for small and moderate sample sizes is often quite different. This is a consequence of different penalties applied in both selection rules. In particular, small Akaike penalty results in inconsistency of the criterion and in large pertaining critical values (see Table 1 in Section 3). In contrast, Schwarz penalty leads to a consistent selection rule. As a consequence, for small sample sizes, the corresponding critical values are relatively small, and the powers for alternatives well described by few terms of the Fourier expansion in the system $\{b_j\}$ are relatively high. On the other hand, large Schwarz penalty causes oversmoothing under small sample sizes. Hence, if one tries to detect distributions with sharp peaks or high frequency oscillations, the power of N_{S1} is often much smaller than that of N_{A1} (cf. discussion in Inglot and Ledwina, 2006a, and further references therein).

To combine the advantages of both selection rules described above, Inglot and Ledwina (2006a) proposed a new selection rule ($T1$ in their paper) which balances between $A1$ and $S1$, assigning Akaike's penalty when the greatest squared empirical Fourier coefficient is too large, and Schwarz's penalty otherwise. As a result, the data driven test based on the statistic N_{T1} attains, roughly speaking, the power close to the maximum of the powers of two competing tests based on N_{A1} and N_{S1}. The idea of constructing $T1$ is developed below to obtain more flexible and sensitive selection rules L. Namely, instead of considering only the greatest empirical Fourier coefficient we shall take into account a few largest squared empirical Fourier coefficients to decide which penalty to apply.

To this end, for each sample size n choose a natural number $D = D_n$ with $1 \le D_n \ll d(n)$. For each $j = 1, \ldots, D_n$ let $Y_{j,n}$ be the number of those squared and normalized empirical Fourier coefficients $n\hat{b}_1^2, \ldots, n\hat{b}_{d(n)}^2$ which are greater than some threshold $c_{j,n}^2$. Consider the event

$$W_n = \bigcup_{j=1}^{D_n} \{Y_{j,n} \ge j\}. \tag{3}$$

Next, take a small positive mass $\delta = \delta_n$, with $\delta_n \to 0$ as $n \to \infty$, and choose $c_{1,n}^2 > c_{2,n}^2 > \cdots > c_{D_n,n}^2$ in such a way that $P_0(W_n) = \delta_n$. Let $\mathbf{1}_B$ be the indicator of the event B and B^c the complement of B, and consider the balanced penalty

$$\pi(j,n) = j\log n \cdot \mathbf{1}_{W_n^c} + 2j \cdot \mathbf{1}_{W_n}. \tag{4}$$

Now, the corresponding selection rule is defined by

$$L = \min\{1 \le k \le d(n) : N_k - \pi(k,n) \ge N_j - \pi(j,n),\ j = 1,\ldots,d(n)\}.$$

Finally, the new data driven test statistic (in fact, a class of statistics depending on the choice of D and δ) is set to be N_L.

Obviously, $S1 \le L \le A1$ a.s. for $n \ge 8$, and consequently $N_{S1} \le N_L \le N_{A1}$ a.s. Hence, N_L preserves all asymptotic optimality properties possessed by both N_{A1} and N_{S1}. It is intuitively clear that enlarging D_n we obtain tests which are more sensitive for alternatives having several meaningful Fourier coefficients in the expansion with respect to the system $\{b_j\}$.

To make the construction work in practice it is enough to ensure the relation $P_0(W_n) = \delta_n$ up to some approximation. To do this observe that for large n's the random vector $\sqrt{n}\,\hat{b} = (\sqrt{n}\,\hat{b}_1,\ldots,\sqrt{n}\,\hat{b}_{d(n)})$ has, under P_0, a distribution close to that for the standard normal vector $(Z_1,\ldots,Z_{d(n)})$. Consequently, for each j, $Y_{j,n}$ has, under P_0, approximately binomial distribution with parameters $d(n)$ and $\mathbf{P}(|Z_1| \ge c_{j,n}) = 2[1-\Phi(c_{j,n})]$, where Φ denotes the standard normal distribution function. Using this approximation, we can write

$$P_0(Y_{j,n} \ge j) \simeq P_0(Y_{j,n} = j) \simeq \binom{d(n)}{j}[2(1-\Phi(c_{j,n}))]^j.$$

We have omitted the factor $[2\Phi(c_{j,n})-1]^{d(n)-j}$ because $\Phi(c_{j,n})$ is so close to 1 that the condition $P_0(W_n) = \delta_n$ could be satisfied. Now, using (3) and taking $P_0(Y_{j,n} \ge j) \simeq \delta_n D_n^{-1}$ leads to $\binom{d(n)}{j}[2(1-\Phi(c_{j,n}))]^j \simeq \delta_n D_n^{-1}$. Finally, we propose to take thresholds $c_{j,n}^2$ given by the last formula, i.e. satisfying the equality

$$1-\Phi(c_{j,n}) = \frac{1}{2}\left(\delta_n D_n^{-1}\left[\binom{d(n)}{j}\right]^{-1}\right)^{1/j}, \quad j = 1,\ldots,D_n. \tag{5}$$

With $d(n)$, D_n and δ_n chosen as described above, and $c_{j,n}^2$ calculated from (5), formulas (3) and (4) define a penalty $\pi(k,n)$ for our selection rule L.

The selection rule $T1$ of Inglot and Ledwina (2006a) is a special case of the above construction. It corresponds to $D_n = 1$, $d(n) = 12$ and $\delta_n \simeq 0.0106$ for $n = 100$. Then $c_{1,n} \simeq 3.245$ and the relation $c_{1,n}^2 = c\log n$ leads to $c = 2.4$, as was proposed in Inglot and Ledwina (2006a).

To preserve sufficient stability of N_L under P_0 it is desirable that the selection rule L should be consistent. Below we give conditions under which this holds. Define $\max_{1\le j\le k}\sup_x |b_j(x)| = B_k$. Then we have the following proposition.

Proposition 1. *Suppose* $1 \le D_n < d(n) < n$ *and* $0 < \delta_n < 1$ *are such that*

$$(6) \qquad \delta_n \to 0, \quad \limsup_{n\to\infty} \frac{D_n \log[\sqrt{D_n}\log(d(n)/\delta_n)]}{\log(1/\delta_n)} < 1$$

and

$$(7) \qquad B^2_{d(n)} D_n \log(1/\delta_n) = O(n^\gamma)$$

for some $\gamma < 1$. *Then for* $c^2_{j,n}$*'s given by* (5) *we have* $P_0(W_n) \to 0$ *as* $n \to \infty$. *Consequently*, $P_0(L = S1) \to 1$ *as* $n \to \infty$.

Condition (6) is mild. For example, if $\delta_n = C_1 n^{-\tau}$ with positive C_1 and τ then (6) holds for any $d(n)$ and $D_n \le C_2(\log n)/\log\log n$ with $C_2 < 2\tau/3$.

Now, using a condition guaranteeing consistency of $S1$ (cf. Theorem 3.2 in Kallenberg and Ledwina, 1995, and the assumption in Theorem 7.9 in Inglot and Ledwina, 1996) we can establish the asymptotic distribution of N_L.

Theorem 2. *Suppose that, in addition to* (6) *and* (7), *we have* $B^2_{d(n)} d^2(n) = o(n/\log n)$. *Then* $P_0(S1 > 1) \to 0$. *Consequently*, $P_0(L > 1) \to 0$ *and* $N_L \xrightarrow{\mathcal{D}} \chi^2_1$ *under* P_0, *where* χ^2_1 *denotes the central chi-square statistic with one degree of freedom.*

Of course, the critical values of the tests based on N_L lie between those of the tests based on N_{A1} and N_{S1} (cf. Table 1 in Section 3). Proposition 1 shows that adjusting $d(n)$, δ_n and D_n appropriately, we keep the critical values rather close to those for the test based on N_{S1}. Recall that for moderate sample sizes the critical values for the test based on N_{S1} are essentially larger than the asymptotic ones. Moreover, these values slowly approach the asymptotic ones. Obviously, the same facts remain true for the new tests based on N_L.

3. Simulation study. To make our notation more precise we shall write in this section $N_L(D,\delta)$ rather than N_L, omitting simultaneously the subscript n. Obviously, for fixed sample size n the choice of δ essentially influences the performance of $N_L(D,\delta)$, both under the null and alternative hypotheses. The empirical critical values of $N_L(D,\delta)$ for $n = 100$ change smoothly as δ increases, from 5.586 of N_{S1} which corresponds to $\delta = 0$ (i.e. $N_{S1} = N_L(D,0)$) to 15.684 of N_{A1} which corresponds to $\delta = 1$ (and practically to $\delta \ge 0.5$, i.e. $N_{A1} \simeq N_L(D,0.5)$). For illustration see Table 1, where influence of increasing D on critical values of $N_L(D,\delta)$ is also shown.

Table 1. The behaviour of simulated critical values of $N_L(D,\delta)$ according to switching parameters D and δ. The Legendre basis, $n=100$, $d(n)=12$, $\alpha=0.05$, 30 000 MC.

D	$\delta=0$	$\delta=.01$	$\delta=.03$	$\delta=.05$	$\delta=.09$	$\delta=.5$
1	5.586	5.993	6.836	7.908	10.667	14.962
2	5.586	5.993	6.850	7.731	10.634	14.985
3	5.586	5.972	6.747	7.650	10.311	14.911
6	5.586	5.957	6.511	7.187	9.026	14.725

It can be observed that the simulated critical values slightly decrease with an increase of D. Our simulation experience reported in Tables 3–7 prompts us to recommend, for moderate sample sizes and $\alpha=0.05$, the choice $\delta=0.03$ to 0.05 and $D=2$ or $D=3$. For such choices, the corresponding critical values for some selected sample sizes are presented in Table 2.

Table 2. Simulated critical values of $N_L(D,\delta)$ for different sample sizes and switching parameters $\delta=0.03$, 0.05 and $D=2,3$. The Legendre basis, $d(n)=12$, $\alpha=0.05$, 30 000 MC.

δ	D	$n=25$	$n=50$	$n=100$	$n=200$	$n=400$
0.03	2	8.151	7.201	6.850	6.421	5.902
	3	8.181	7.174	6.747	6.361	5.873
0.05	2	9.097	8.265	7.731	7.360	6.631
	3	9.016	8.244	7.650	7.107	6.609

For other significance levels α a reasonable choice for δ seems to be between $\alpha/2$ and α.

Our primary goal in this section is to compare, in terms of average power, the performance of the new tests $N_L(D,\delta)$ with the two-sided optimal Bayes test described in the Appendix and given by the formula (A.2). We restrict ourselves to the case $n=100$, $\alpha=0,05$ and take $d(n)=K=12$. We consider the Legendre basis. Some more simulations not presented here yield the same picture for the cosine basis.

Now, for $j=1,\dots,12$ consider the alternatives $p_{12}(x,\pm 0.25e_j)$, where $e_1,\dots,e_{12}$ is the standard basis in Euclidean space $\mathbb{R}^{12}$ and $p_k(x,\theta)$ denotes the density from the exponential family given by (1). Let $T_n^*=T^*$ be the two-sided optimal Bayes test given as in (A.2) in the Appendix defined by the above 24 densities under the uniform prior distribution. We want to compare the power behaviour of $N_L(D,\delta)$ and T^* for the alternatives $p_{12}(x,\pm 0.25e_j)$, $j=1,\dots,12$. To show the whole picture we also include the one-sided Neyman–Pearson test denoted by NP (constructed for each alternative separately), the maximum test $M = \max_{1 \le j \le 12}\{n\hat{b}_j^2\}$ and Neyman smooth test $N_{12} = \sum_{j=1}^{12} n\hat{b}_j^2$, where $\hat{b}_j$'s denote the empirical Fourier coefficients (cf. (2)) with respect to the Legendre basis (b_j). The results are shown in Table 3.

Table 3. Comparison of powers and average powers $1-\bar{\beta}$ (in %) of $N_L(D,\delta)$, T^*, NP, M and N_{12}. The Legendre basis, $n = 100$, $\alpha = 0.05$, $d(n) = 12$, 10 000 MC, alternatives $p_{12}(x, \theta_j)$, uniform prior.

θ_j	NP	$T^\star$	M	N_{S1} ($\delta = 0$)	$N_L(1,\delta)$ (N_{T1}) .01	$N_L(1,\delta)$ .05	$N_L(1,\delta)$ (N_{A1}) 0.5	$N_L(2,\delta)$ .03	$N_L(2,\delta)$ .05	$N_L(3,\delta)$ .03	$N_L(3,\delta)$ .05	$N_L(6,\delta)$ .05	$N_L(6,\delta)$ .09	N_{12}
$.25e_1$	80	38	36	57	53	37	18	46	40	47	40	44	33	28
$.25e_2$	81	48	46	68	66	53	32	61	56	62	56	59	49	42
$.25e_3$	80	39	37	38	38	42	25	40	41	41	42	41	41	30
$.25e_4$	80	46	44	30	34	47	37	40	45	39	43	39	43	39
$.25e_5$	82	40	38	14	26	40	32	32	37	30	34	29	34	30
$.25e_6$	82	45	44	13	31	45	43	39	44	37	41	37	42	39
$.25e_7$	80	40	38	07	26	36	37	31	35	29	33	29	34	30
$.25e_8$	83	45	43	08	30	40	45	36	40	35	38	35	39	38
$.25e_9$	81	40	37	06	24	30	35	28	31	27	30	27	30	30
$.25e_{10}$	83	45	43	07	27	34	38	33	35	32	34	32	35	37
$.25e_{11}$	79	39	37	06	21	25	27	25	26	24	26	25	26	30
$.25e_{12}$	82	46	42	06	24	28	29	28	29	28	29	28	30	37
$-.25e_1$	80	39	36	57	53	38	18	46	40	47	40	44	32	28
$-.25e_2$	79	28	27	54	51	32	10	43	35	44	35	40	26	14
$-.25e_3$	79	39	37	38	38	41	24	39	40	40	41	40	40	29
$-.25e_4$	79	30	29	15	20	31	19	24	28	22	25	21	25	19
$-.25e_5$	80	39	37	14	26	40	31	32	37	30	34	30	34	31
$-.25e_6$	79	32	31	07	20	32	26	24	29	22	26	22	26	21
$-.25e_7$	81	39	38	07	26	37	37	31	35	29	33	29	33	29
$-.25e_8$	78	33	32	06	20	27	29	24	26	22	25	22	25	22
$-.25e_9$	80	40	38	06	24	31	36	29	31	28	31	28	31	30
$-.25e_{10}$	79	34	33	05	18	22	23	21	22	20	22	20	22	23
$-.25e_{11}$	81	40	38	06	22	26	27	25	26	25	26	25	26	31
$-.25e_{12}$	79	35	33	06	15	17	17	17	18	17	18	17	17	25
$1-\bar{\beta}$	80.3	39.1	37.2	20.0	30.5	34.6	29.0	33.1	34.4	32.4	33.4	31.8	32.2	29.7

Note that $\delta = 0$ corresponds to the test N_{S1}, $\delta = 0.5$ practically to the test N_{A1}, while $\delta \simeq 0.01$, $D = 1$ to the test N_{T1} considered by Inglot and

Ledwina (2006a). The power of the NP test is almost constant but both T^* and M have some fluctuations around the average power depending on whether an "odd" or "even" alternative occurs. It can be observed that the loss in average power for T^* and M with respect to the case when full information about the alternative is available equals ca. 41% and 43%, respectively. This agrees quite well with the approximation derived in Theorem A.2 of the Appendix (in our case the entropy of the prior distribution is $\log_2 12$ plus 1 bit for two-sided test, $\rho = 2.5$ and consequently $(1 + \log_2 12)0.221/\rho \simeq 40.5\%$). On the other hand, the loss in average power for $N_L(D, \delta)$ with respect to T^* in the middle range of the power function is about 5% and does not change significantly for reasonable choices of D and δ. The extreme tests N_{S1} and N_{A1} are perceptibly weaker. Moreover, N_{A1} is a little worse than the nonadaptive Neyman smooth test N_{12}. For reasonable choices of D and δ the test based on $N_L(D, \delta)$ preserves super sensitivity for the first two axes, which is an interesting and welcome smoothing property of the test based on N_{S1}.

We have also compared our test with a recently proposed new test by Fromont and Laurent (2006). It turns out that for an analogous set of 24 alternatives built on the cosine basis their test attains average power 27.5% while the test $N_L(D, \delta)$ based on the cosine basis with $D = 1$, $\delta = 0.05$ and $D = 3$, $\delta = 0.03$ gives average powers 34.2% and 31.1%, respectively. Complete results are shown in Table 4.

Table 4. Comparison of powers and average powers $1 - \bar{\beta}$ (in %) of Fromont and Laurent test (FL), $N_L^* = N_L(1, 0.05)$ and $N_L^{**} = N_L(3, 0.03)$. Alternatives $p_{12}(x, \theta_j)$, $\theta_j = \pm 0.25 e_j$, built on the cosine basis, uniform prior. $n = 100$, $d(n) = 12$, $\alpha = 0.05$, 10 000 MC.

		j												
test		1	2	3	4	5	6	7	8	9	10	11	12	$1 - \bar{\beta}$
FL	$+.25e_j$	36	53	30	39	25	31	21	25	18	21	13	18	
	$-.25e_j$	35	54	31	38	25	32	20	25	17	21	14	17	27.5
N_L^*	$+.25e_j$	43	49	41	40	39	38	34	33	28	26	22	21	
	$-.25e_j$	42	48	41	38	38	37	35	31	29	25	23	19	34.2
N_L^{**}	$+.25e_j$	51	56	40	30	28	28	28	27	25	24	21	21	
	$-.25e_j$	49	55	40	27	28	27	28	26	25	23	22	18	31.1

To show how our tests perform for different sample sizes we compare three cases $n = 25$, $n = 100$ and $n = 400$, modifying appropriately the distance of the same alternatives from the null distribution. Table 5 shows that for smaller n the power of N_L is closer (in average) to the power of T^*.

Table 5. Comparison of powers and average powers $1-\bar{\beta}$ (in %) of $N_L^{**} = N_L(3, 0.03)$, T^*, M and N_{12} for different sample sizes. The Legendre basis, $\alpha = 0.05$, 10 000 MC, alternatives $p_{12}(x, \theta_j)$, uniform prior. $n = 25, 100, 400$. $d(25) = 9$, $d(100) = d(400) = 12$.

$n=25$					$n=100$					$n=400$				
θ_j	$T^\star$	M	N_L^{**}	N_9	θ_j	$T^\star$	M	N_L^{**}	N_{12}	θ_j	$T^\star$	M	N_L^{**}	N_{12}
$.5e_1$	41	39	37	31	$.25e_1$	38	36	47	28	$.125e_1$	38	38	54	31
$.5e_2$	59	56	62	54	$.25e_2$	48	46	62	42	$.125e_2$	43	42	57	38
$.5e_3$	41	42	50	35	$.25e_3$	39	37	41	30	$.125e_3$	38	38	29	31
$.5e_4$	56	55	58	51	$.25e_4$	46	44	39	39	$.125e_4$	43	41	32	36
$.5e_5$	43	43	36	36	$.25e_5$	40	38	30	30	$.125e_5$	39	39	28	32
$.5e_6$	55	54	47	50	$.25e_6$	45	44	37	39	$.125e_6$	42	41	30	36
$.5e_7$	44	44	33	37	$.25e_7$	40	38	29	30	$.125e_7$	39	38	27	31
$.5e_8$	56	54	44	50	$.25e_8$	45	43	35	38	$.125e_8$	42	41	30	35
$.5e_9$	46	44	30	38	$.25e_9$	40	37	27	30	$.125e_9$	40	38	26	31
$.5e_{10}$					$.25e_{10}$	45	43	32	37	$.125e_{10}$	42	40	27	35
$.5e_{11}$					$.25e_{11}$	39	37	24	30	$.125e_{11}$	40	39	23	32
$.5e_{12}$					$.25e_{12}$	46	42	28	37	$.125e_{12}$	42	40	24	34
$-.5e_1$	41	40	48	32	$-.25e_1$	39	36	47	28	$-.125e_1$	38	38	54	31
$-.5e_2$	21	21	22	10	$-.25e_2$	28	27	44	14	$-.125e_2$	34	33	50	23
$-.5e_3$	43	42	49	35	$-.25e_3$	39	37	40	29	$-.125e_3$	39	38	29	31
$-.5e_4$	27	28	25	17	$-.25e_4$	30	29	22	19	$-.125e_4$	36	34	23	26
$-.5e_5$	43	44	37	37	$-.25e_5$	39	37	30	31	$-.125e_5$	39	37	27	31
$-.5e_6$	30	30	19	19	$-.25e_6$	32	31	22	21	$-.125e_6$	35	34	23	26
$-.5e_7$	45	44	34	37	$-.25e_7$	39	38	29	29	$-.125e_7$	39	38	27	31
$-.5e_8$	32	32	19	22	$-.25e_8$	33	32	22	22	$-.125e_8$	35	35	23	27
$-.5e_9$	47	45	31	38	$-.25e_9$	40	38	28	30	$-.125e_9$	39	39	26	32
$-.5e_{10}$					$-.25e_{10}$	34	33	20	23	$-.125e_{10}$	36	35	21	27
$-.5e_{11}$					$-.25e_{11}$	40	38	25	31	$-.125e_{11}$	39	37	23	31
$-.5e_{12}$					$-.25e_{12}$	35	33	17	25	$-.125e_{12}$	36	36	19	29
$1-\bar{\beta}$	42.8	42.1	37.2	34.9		39.1	37.2	32.4	29.7		38.9	38.0	30.5	31.1

Table 3 presents an artificial situation. So, we also want to show the behaviour of $N_L(D, \delta)$ in more realistic situations, when alternatives have two or more meaningful Fourier coefficients. Although estimates obtained in the Appendix do not cover such cases, we include in Table 6 powers of the corresponding optimal Bayes test for comparison. First, we consider equal Fourier coefficients on two axes such that the L_2-distance of each alternative density from p_0 is approximately the same as before. We restrict ourselves to alternatives with two positive coefficients on the first six axes, resulting in 15 different alternatives. By T^{**} we denote the two-sided optimal Bayes test given by (A.2) constructed for the set of these 15 alternatives plus 45 alternatives obtained by changing signs and under the uniform prior. We also add the NP test and N_{12} for better comparison. The results are shown in Table 6.

Table 6. Comparison of powers and average powers (in %) of $N_L(D,\delta)$, T^{**}, NP and N_{12}. The Legendre basis, $n = 100$, $\alpha = 0.05$, $d(n) = 12$, 10 000 MC, alternatives $p_6(x,\theta)$, uniform prior.

θ						NP	T^{**}	N_{S1}	N_{T1}	$N_L(D,0.05)$ D 1	2	3	N_{A1}	N_{12}
.167	.167	0	0	0	0	80	39	58	56	45	48	48	29	39
.167	0	.167	0	0	0	77	32	38	36	29	33	33	23	29
.167	0	0	.167	0	0	77	37	33	32	30	32	32	28	33
.167	0	0	0	.167	0	76	33	29	30	27	29	28	25	28
.167	0	0	0	0	.167	77	33	29	31	28	30	30	29	31
0	.167	.167	0	0	0	81	37	51	50	44	46	47	32	38
0	.167	0	.167	0	0	84	57	54	53	50	53	53	46	50
0	.167	0	0	.167	0	81	49	41	42	39	41	41	36	39
0	.167	0	0	0	.167	82	50	42	44	43	45	45	42	45
0	0	.167	.167	0	0	80	39	32	33	37	38	38	34	37
0	0	.167	0	.167	0	79	37	23	25	30	32	32	31	32
0	0	.167	0	0	.167	81	47	26	30	37	39	39	38	38
0	0	0	.167	.167	0	79	35	20	24	34	35	35	37	36
0	0	0	.167	0	.167	82	52	24	31	43	45	44	49	48
0	0	0	0	.167	.167	80	35	14	22	34	35	34	38	35
average power						79.7	40.8	34.3	35.9	36.7	38.7	38.6	34.5	37.2

Nice behaviour of $N_L(2,0.05)$ for the alternatives from Table 6 is not surprising since we have just disturbed P_0 exactly on two axes. However, other statistics $N_L(D,\delta)$ provide tests only slightly worse.

Table 7 shows the performance of $N_L(D,\delta)$ when there are three meaningful Fourier coefficients.

As could be expected, the test based on $N_L(3,0.05)$ attains the best average power for alternatives from Table 7. Still, other choices of D and δ give only slightly weaker tests.

Finally, we compare in Table 8 the average powers attained by four of our tests with typical choices of D and δ with the average powers of the optimal Bayes test under some particular prior distributions.

Table 7. Comparison of powers and average powers (in %) of $N_L(D,\delta)$ and N_{12}. The Legendre basis, $n = 100$, $\alpha = 0.05$, $d(n) = 12$, 10 000 MC, alternatives $p_6(x,\theta)$, uniform prior.

θ						N_{S1}	N_{T1}	$N_L(D, 0.05)$ D 1	2	3	N_{A1}	N_{12}
.165	.11	.11	0	0	0	52	50	42	44	45	33	40
.11	.165	.11	0	0	0	55	54	46	48	49	35	42
.11	.11	.165	0	0	0	47	46	42	44	45	34	40
0	0	0	.165	.11	.11	23	27	37	39	39	44	44
0	0	0	.11	.165	.11	20	25	35	37	37	43	41
0	0	0	.11	.11	.165	20	26	36	38	38	45	43
.165	0	.11	.11	0	0	39	37	32	35	36	32	36
.165	0	0	.11	.11	0	34	33	30	33	33	32	35
.165	0	0	0	.11	.11	32	33	29	32	33	32	35
0	.165	.11	.11	0	0	52	52	46	48	49	40	46
0	.165	0	.11	.11	0	47	46	41	44	45	40	44
0	.165	0	0	.11	.11	41	42	38	41	42	39	41
0	0	.165	.11	.11	0	30	31	36	37	38	37	39
0	0	.165	0	.11	.11	27	29	35	37	38	39	39
0	0	.165	.11	0	.11	31	33	38	40	41	40	42
.11	.11	0	.165	0	0	42	41	40	42	43	38	43
.11	.11	0	0	.165	0	33	34	33	35	35	33	35
.11	.11	0	0	0	.165	32	36	36	38	38	36	39
average power						36.5	37.5	37.3	39.6	40.2	37.3	40.2

Table 8. Comparison of average powers (in %) of N_{S1}, N_{T1}, $N_L^* = N_L(1, 0.05)$, $N_L^{**} = N_L(3, 0.03)$ and the optimal Bayes test T. The Legendre basis, $n = 100$, $\alpha = 0.05$, $d(n) = 12$, 10 000 MC, alternatives $p_6(x, +0.25e_j)$.

prior distribution w_j						average power N_{S1}	N_{T1}	N_L^*	N_L^{**}	T
0.250	0.250	0.250	0.250	0	0	48.1	47.9	44.8	47.3	55.2
0.500	0.250	0.125	0.125	0	0	54.0	52.0	42.6	49.0	56.7
0.1667	0.1667	0.1667	0.1667	0.1666	0.1666	36.7	41.3	44.0	42.7	51.0
0.500	0.250	0.125	0.065	0.030	0.030	53.0	47.7	42.6	48.7	55.0
0.250	0.250	0.200	0.150	0.100	0.050	45.4	46.6	44.2	46.1	52.1
0	0.3334	0	0.3333	0	0.3333	37.0	43.7	48.0	52.7	61.5

From Table 8 it is seen that the test N_{S1} behaves very well if a prior distribution is concentrated on the first 2–4 axes. Otherwise, N_L performs better.

Let us finish this section by some practical recommendations for an implementation of the tests based on $N_L(D,\delta)$ for $n=100$ and $\alpha=0.05$. For a given prior distribution W on orthogonal directions b_j, order them according to decreasing values of w_j. If W is practically concentrated on at most four axes then use N_L with $\delta=0$, i.e. N_{S1}. Otherwise, use N_L with $D=2$ or 3 and with δ between 0.01–0.05 depending on how much mass W distributes on the first few axes. For other sample sizes and significance levels these recommendations should be appropriately modified.

4. Proof of Proposition 1. From (5) it follows that, for each n, $c_{j,n}$'s decrease when j increases and $1-\Phi(c_{j,n})\le \delta_n^{1/D_n}$ for every $j\le D_n$. Since (6) implies $D_n/\log(1/\delta_n)\to 0$, we infer that $c_{j,n}\to\infty$ for all j's. Moreover, applying the inequality $1-\Phi(x)\le \exp\{-x^2/2\}$ for $x\ge 1$ we deduce from (5) for all j's and n sufficiently large that

$$\frac{c_{j,n}^2}{2}\le\frac{c_{1,n}^2}{2}\le\log\frac{2D_nd(n)}{\delta_n}. \tag{8}$$

On the other hand, the inequality

$$1-\Phi(x)\ge\exp\left\{-\frac{x^2}{2}-\frac{1}{2}\log\frac{x^2}{2}-\frac{1}{2}\log 8\pi\right\}\qquad\text{for } x\ge 2$$

together with (5) gives

$$\begin{aligned}&j\left(\frac{c_{j,n}^2}{2}+\frac{1}{2}\log\frac{c_{j,n}^2}{2}+\frac{1}{2}\log 2\pi\right)\\&\qquad\ge\log\frac{1}{\delta_n}+\log D_n+\log\binom{d(n)}{j}\ge\log\frac{1}{\delta_n}+\log\binom{d(n)}{j}\end{aligned} \tag{9}$$

for all j's and n sufficiently large.

Now, for each $j=1,\dots,D_n$, let $\mathcal{A}_j$ denote the family of all subsets of $\{1,\dots,d(n)\}$ of size j. Then from (3) and the definition of $Y_{j,n}$ we can write

$$P_0(W_n)\le\sum_{j=1}^{D_n}P_0(Y_{j,n}\ge j)\le\sum_{j=1}^{D_n}\sum_{A\in\mathcal{A}_j}P_0(|\sqrt{n}\,\hat{b}|_A^2\ge jc_{j,n}^2), \tag{10}$$

where for $A\subset\{1,\dots,d(n)\}$ and $v\in\mathbb{R}^{d(n)}$ we have set $|v|_A^2=\sum_{i\in A}v_i^2$. By the orthonormality of the system $\{b_j\}$ it follows that, under P_0, the random vector $\sqrt{n}\,\hat{b}$ has mean 0 and unit covariance matrix. This and the uniform boundedness of the functions b_j allow us to apply Prokhorov's inequality (Prokhorov, 1973) to estimate the right-hand side of (10). So, for sufficiently large n's we have

$$P_0(W_n)\le\sum_{j=1}^{D_n}\frac{C}{\Gamma(j/2)}\exp\{-\Delta_{jn}\}, \tag{11}$$

where

$$(12)\qquad \Delta_{jn} = \frac{jc_{j,n}^2}{2}(1-\eta_{j,n}) - \frac{j-1}{2}\log\frac{jc_{j,n}^2}{2} - \log\binom{d(n)}{j}$$

with $\eta_{j,n}^2 \le B_{d(n)}^2 jc_{j,n}^2 n^{-1} \le 2B_{d(n)}^2 D_n n^{-1}\log(2D_n d(n)/\delta_n) = \eta_n^2$, C is an absolute constant and $\Gamma(\cdot)$ is the Euler gamma function. Observe that by (7) it follows that $\eta_n \to 0$, which justifies the application of Prokhorov's inequality. We estimate the exponent (12) as follows:

$$(13)\qquad \Delta_{jn} \ge \left[\frac{jc_{j,n}^2}{2} + \frac{j}{2}\log\frac{c_{j,n}^2}{2} + \frac{j}{2}\log 2\pi - \log\binom{d(n)}{j}\right](1-\eta_n)$$
$$-D_n\log\frac{c_{j,n}^2}{2} - \frac{D_n}{2}\log 2\pi D_n - \eta_n D_n \log d(n).$$

Observe that the last term in (13) is $o(\log(1/\delta_n))$. Indeed, by the form of η_n we can write

$$(14)\qquad \eta_n^2 D_n^2 \log^2 d(n) = \frac{2B_{d(n)}^2 D_n \log(1/\delta_n)}{n^\gamma}\,\frac{\log^2 d(n)\log(2D_n d(n)/\delta_n)}{n^{1-\gamma}\log(1/\delta_n)}\,D_n^2.$$

Now, by (7) the first factor in (14) is bounded, the second tends to zero (since $\gamma < 1$, $D_n < d(n) < n$ and $\delta_n \to 0$) while D_n^2 is $o(\log^2(1/\delta_n))$ by the assumption (6). Applying (8) and (9) and omitting expressions of order $o(\log(1/\delta_n))$ in the middle terms of (13) we obtain

$$(15)\qquad \Delta_{jn} \ge \log(1/\delta_n)\left(1 - \frac{D_n\log[\sqrt{D_n}\log(d(n)/\delta_n)]}{\log(1/\delta_n)} + o(1)\right).$$

By (6) it follows that the right-hand side of (15) tends to infinity as $n\to\infty$. As $\sum_{j=1}^{\infty}(\Gamma(j/2))^{-1} < \infty$, the assertion of Proposition 1 follows from (11). ∎

Appendix. In this section we collect some auxiliary considerations. First, we define the optimal Bayes test for a finite set of alternatives which is implemented in Section 3. We also show its relation to the limiting Gaussian shift problem. Finally, we study an optimal maximum test for the two-sided Gaussian shift problem and estimate its power. This estimate displays the connection of the loss in average power for this maximum test (with respect to the Neyman–Pearson test) with the entropy of a prior distribution. In this indirect way we explain a phenomenon observed in our simulations in Section 3.

Optimal Bayes tests. Let $\underline{X} = (X_1,\ldots,X_n)$ be a sample from distribution P on a space $\mathcal{X}$. Suppose we want to test the hypothesis $H_0 : P = P_0$, where P_0 has density p_0 with respect to some σ-finite measure μ on $\mathcal{X}$. Let $\mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\}$ be a family of possible alternatives ($P_\vartheta \ne P_0$ for every

$\vartheta \in \Theta$), where for each $\vartheta \in \Theta$ the distribution P_ϑ has density p_ϑ. Let W be a prior distribution on Θ (endowed with some σ-field). Denote by $p_{n,\vartheta}(\underline{x})$ and $p_{n,0}(\underline{x})$ the likelihood functions corresponding to P_ϑ and P_0, respectively. It is well known (see e.g. Clarke and Barron, 1990, p. 460, or Janssen, 2003, Sec. 2.4) that the optimal Bayes test (i.e. ensuring the smallest average probability of the second kind error $\overline{\beta}$) of H_0 against $H_1 : P \in \mathcal{P}$ is given by the statistic

$$\int_\Theta \frac{p_{n,\vartheta}(\underline{X})}{p_{n,0}(\underline{X})} W(d\vartheta). \tag{A.1}$$

Now, let $\mathcal{P} = \{P_1, \ldots, P_K\}$ be a finite family of alternatives and let $W = (w_1, \ldots, w_K)$ be a prior distribution on $\mathcal{P}$. Then the test statistic in (A.1) takes the form

$$T_n^* = \sum_{j=1}^K w_j \frac{p_{n,j}(\underline{X})}{p_{n,0}(\underline{X})}. \tag{A.2}$$

Here we have set $p_j = dP_j/d\mu$, $j = 1, \ldots, K$, and $p_{n,j}(\underline{X}) = \prod_{i=1}^n p_j(X_i)$.

Suppose the alternatives $P_j = P_j^{(n)}$ approach P_0 at the rate $n^{-1/2}$ under fixed K. Namely, assume that for some $\rho > 0$ and every $j = 0, 1, \ldots, K$,

$$E_{P_j} \log \frac{p_r^{(n)}(X_1)}{p_0(X_1)} = -\frac{\rho^2}{2n} + \delta_{jr}\frac{\rho^2}{n} + o\Big(\frac{1}{n}\Big), \qquad r = 1, \ldots, K, \tag{A.3}$$

and

$$\operatorname{Cov}_{P_j}\bigg(\log \frac{p_1^{(n)}(X_1)}{p_0(X_1)}, \ldots, \log \frac{p_K^{(n)}(X_1)}{p_0(X_1)}\bigg) = \frac{\rho^2}{n} I + o\Big(\frac{1}{n}\Big), \tag{A.4}$$

where I is the identity matrix and δ_{jr} is the Kronecker delta. For example, conditions (A.3) and (A.4) hold when $p_j^{(n)} = p_0(1 + \rho n^{-1/2} b_j)$ with bounded functions b_j satisfying $\int p_0 b_j \, d\mu = 0$ and $\int p_0 b_j b_r \, d\mu = \delta_{jr}$. A straightforward application of the central limit theorem leads to the following proposition.

PROPOSITION A.1. *Assume* (A.3) *and* (A.4). *Then for* T_n^* *given by* (A.2) *and for any prior distribution* W *we have*

$$T_n^* \xrightarrow{\mathcal{D}} \sum_{j=1}^K w_j \exp\{\rho Z_j - \rho^2/2\} \qquad \text{under } P_0$$

and

$$T_n^* \xrightarrow{\mathcal{D}} \sum_{j \neq r}^K w_j \exp\{\rho Z_j - \rho^2/2\} + w_r \exp\{\rho Z_r + \rho^2/2\} \qquad \text{under } P_r^{(n)},$$

where $Z_1, \ldots, Z_K$ *are i.i.d. standard normal random variables.*

Hence, for alternatives satisfying (A.3) and (A.4) the optimal Bayes test based on T_n^* is asymptotically equivalent to the optimal Bayes test in the Gaussian shift problem. In this last problem, P_0 is the standard normal distribution in R^K, $P_j^{(n)}$ has normal distribution $N(\rho n^{-1/2}e_j, I)$, where e_j denotes the unit vector on the jth axis, and $\mathcal{P} = \{P_1^{(n)}, \ldots, P_K^{(n)}\}$ is a fixed set of alternatives for given n. Below, we describe a two-sided version of this limiting testing problem which corresponds to comparisons made in Section 3.

Optimal Bayes test for the two-sided Gaussian shift problem. As before, let P_0 be the standard normal distribution in $\mathbb{R}^K$, $K \geq 1$, and $P_{j\pm}^{(n)}$, $j = 1, \ldots, K$, be normal $N(\pm\rho n^{-1/2}e_j, I)$ distributions in $\mathbb{R}^K$ with $\rho > 0$ fixed and known and e_j as above. We want to test

$$\text{(A.5)} \qquad H_0 : P = P_0 \quad \text{against} \quad P \in \mathcal{P} = \{P_{1+}^{(n)}, P_{1-}^{(n)}, \ldots, P_{K+}^{(n)}, P_{K-}^{(n)}\}.$$

If $W = (w_{1+}, w_{1-}, \ldots, w_{K+}, w_{K-})$ is a prior distribution with $w_{j+} = w_{j-} = \frac{1}{2}w_j$ on the actual set of alternatives $\mathcal{P}$ then the statistic of the optimal Bayes test takes by (A.2) the form

$$\text{(A.6)} \qquad \mathcal{T}_n^* = \sum_{j=1}^{K} w_j \exp\{-\rho^2/2\}\cosh(\rho\sqrt{n}\,|\overline{X}_j|),$$

where $\overline{X} = (\overline{X}_1, \ldots, \overline{X}_K)$ is a vector of sample means. Since, under P_0, $\sqrt{n}\,\overline{X}_j = Z_j$, $j = 1, \ldots, K$, are independent standard normal random variables, the critical value t_α of this test satisfies the relation

$$\mathbf{P}\Big(\sum_{j=1}^{K} w_j \exp\{-\rho^2/2\}\cosh(\rho Z_j) \geq t_\alpha\Big) = \alpha.$$

Optimal maximum test for the two-sided Gaussian shift problem. Consider again the testing problem as in (A.5). First observe that for testing H_0 against two equiprobable alternatives $P_{j+}^{(n)}, P_{j-}^{(n)}$ the statistic of the optimal Bayes test (two-sided Neyman–Pearson test) has the form $\sqrt{n}\,|\overline{X}_j|$ (cf. (A.6)). We shall reject H_0 in (A.5) if at least one of the "partial" tests $\sqrt{n}\,|\overline{X}_j| \geq c_j$, $j = 1, \ldots, K$, will reject it. We shall use different "partial" critical values $c_1, \ldots, c_K$ according to different prior probabilities $w_1, \ldots, w_K$. We have to choose them so as to maintain a given significance level α. Since, under P_0, $\sqrt{n}\overline{X}_j = Z_j$, $j = 1, \ldots, K$, are independent standard normal random variables this leads to the relation

$$\text{(A.7)} \qquad \prod_{j=1}^{K} (2\Phi(c_j) - 1) = 1 - \alpha.$$

By (A.7) the probability of the second kind error under the alternatives $P_{j+}^{(n)}, P_{j-}^{(n)}$ can be written as

$$\beta_j = \prod_{r\neq j}(2\Phi(c_r)-1)(\Phi(c_j-\rho)-\Phi(-c_j-\rho)) = (1-\alpha)\,\frac{\Phi(c_j-\rho)-\Phi(-c_j-\rho)}{2\Phi(c_j)-1}.$$

We want to choose c_j's in an optimal way to minimize the average second kind error

$$\bar{\beta} = \sum_{j=1}^{K}\beta_j w_j = (1-\alpha)\sum_{j=1}^{K} w_j\,\frac{\Phi(c_j-\rho)-\Phi(-c_j-\rho)}{2\Phi(c_j)-1} \tag{A.8}$$

under a given prior distribution W and under the constraint (A.7). Differentiating the expression $\bar{\beta}-\lambda(\prod_{j=1}^{K}(2\Phi(c_j)-1)-1+\alpha)$ with respect to consecutive c_j's and equating them to 0 we get

$$\frac{\phi(c_j-\rho)+\phi(c_j+\rho)}{2\phi(c_j)} - \frac{\Phi(c_j-\rho)-\Phi(-c_j-\rho)}{2\Phi(c_j)-1} = \frac{\lambda}{w_j}, \quad j=1,\ldots,K, \tag{A.9}$$

where ϕ denotes the density of the standard normal distribution. As $2\Phi(c_j)-1$ is close to 1 while $\phi(c_j)$ is close to 0, the second term in (A.9) is small in comparison to the first one. So, omitting it as well as the term $\phi(c_j+\rho)$, which is much smaller than $\phi(c_j-\rho)$, we obtain

$$c_j \simeq C + \frac{1}{\rho}\ln\frac{1}{w_j}, \qquad j=1,\ldots,K,$$

for some constant C.

Finally, for our testing problem (A.5) we consider the test statistic

$$\mathcal{M}_n = \max_{1\le j\le K}\frac{\sqrt{n}\,|\bar{X}_j|}{c_j}, \tag{A.10}$$

where $c_j = C + \frac{1}{\rho}\ln\frac{1}{w_j}$, i.e. the c_j are close to the optimal choice with the constant C, depending on α, ρ and W, uniquely determined by (A.7). The test rejects H_0 when $\mathcal{M}_n \ge 1$. So, the average second kind error has the form (A.8) with the above c_j's.

In the theorem below we need to apply a linear approximation of the function $\Phi(x)$. The maximal slope of $\Phi(x)$ equals $(2\pi)^{-1/2}$. However, the points $c_j-\rho$ in (A.8) oscillate in some interval arround 0. So, a kind of "average" slope would be more adequate for linear approximation of $\Phi(x)$. Let us take $s_0 = 4/(5\sqrt{2\pi})$ as the "average" slope and consider the two tangent lines of $\Phi(x)$ corresponding to this slope. Then we get the following estimates:

$$\text{(A.11)}\qquad \begin{aligned} \frac{4}{5\sqrt{2\pi}}\,x + A \le \Phi(x), &\qquad x \le x_0,\\ \Phi(x) \le \frac{4}{5\sqrt{2\pi}}\,x + 1 - A, &\qquad x \ge -x_0, \end{aligned}$$

where

$$A = \Phi\Bigl(-\sqrt{\ln\frac{25}{16}}\Bigr) + \frac{4}{5\sqrt{2\pi}}\sqrt{\ln\frac{25}{16}}$$

and the positive number x_0 is the unique solution of the equation $\Phi(x_0) = \frac{4}{5\sqrt{2\pi}}x_0 + A$. Note that $x_0 \simeq 1.44$ with $x_0 > 1.44$. The difference between the two sides of (A.11) is quite small and equals $1 - 2A \simeq 0.07$. Using (A.11) we can now estimate the average second kind error $\overline{\beta}$.

THEOREM A.2. *Assume the significance level satisfies $\alpha \le 0.1$ and ρ, α are chosen so that*

$$\text{(A.12)}\qquad 0 \le \rho - \Phi^{-1}\Bigl(1 - \frac{\alpha}{2}\Bigr) \le x_0.$$

Moreover, suppose a prior distribution W is sufficiently regular, i.e.

$$\text{(A.13)}\qquad \begin{aligned} &\sum_{j\,:\,x_0 < c_j - \rho \le 2} w_j = a_1 \le 0.2, \qquad \sum_{j\,:\,c_j - \rho > 2} w_j = a_2 \le 0.02,\\ &\min_{1 \le j \le K} w_j \ge 0.0001. \end{aligned}$$

Then the average second kind error $\overline{\beta}$, given by (A.8), *of the optimal maximum test based on the statistic $\mathcal{M}_n$ satisfies*

$$\text{(A.14)}\qquad \overline{\beta} \le \frac{4\ln 2}{5\sqrt{2\pi}\,\rho}\,H(W) + \frac{4(C-\rho)}{5\sqrt{2\pi}} + 1 - A$$

and

$$\text{(A.15)}\qquad \overline{\beta} \ge (1-\alpha)\Bigl[\frac{4\ln 2}{5\sqrt{2\pi}\rho}\,H(W) + 0.98\,\frac{4(C-\rho)}{5\sqrt{2\pi}} + A - B\Bigr],$$

where $H(W) = -\sum_{j=1}^{K} w_j \log_2 w_j$ is the Shannon entropy of W and $B = B' + B''$ with B', B'' defined in (A.17) and (A.18), respectively.

Proof. From (A.7) it follows that $c_j \ge \Phi^{-1}(1 - \alpha/2)$ for every j. Hence from (A.12) we have $c_j - \rho \ge -x_0$. So, applying the upper estimate in (A.11) to (A.8) and using again (A.7) we get

$$\overline{\beta} \le \sum_{j=1}^{K} w_j \Phi(c_j - \rho) \le \frac{4\ln 2}{5\sqrt{2\pi}\rho}\,H(W) + \frac{4(C-\rho)}{5\sqrt{2\pi}} + 1 - A,$$

which is exactly (A.14).

For the proof of (A.15) we have from (A.8)

$$\bar{\beta} \geq (1-\alpha)\sum_{j=1}^{K} w_j \Phi(c_j-\rho) - (1-\alpha)\sum_{j=1}^{K} w_j \Phi(-c_j-\rho). \tag{A.16}$$

The inequalities $c_j \geq \Phi^{-1}(1-\alpha/2)$ (cf. (A.7)) and $\rho \geq \Phi^{-1}(1-\alpha/2)$ (cf. (A.12)) together prove that $\Phi(-c_j-\rho)$ can be bounded by

$$\Phi(-2\Phi^{-1}(1-\alpha/2)) \leq \Phi(-2\Phi^{-1}(0.95)) = B' \tag{A.17}$$

due to the assumption $\alpha \leq 0.1$. By the lower estimate in (A.11) the first term in (A.16) is greater than or equal to

$$(1-\alpha)\left[\frac{4\ln 2}{5\sqrt{2\pi}\rho}H(W) + \frac{4(C-\rho)}{5\sqrt{2\pi}} + A - R\right],$$

where

$$R = \sum_{j\,:\,c_j-\rho>x_0} w_j\left[\frac{4}{5\sqrt{2\pi}}(c_j-\rho) + A - \Phi(c_j-\rho)\right].$$

Using the definition of c_j's, the inequality $\rho \geq \Phi^{-1}(1-\alpha/2)$ (cf. (A.12)) and the assumption $w_j \geq 0.0001$ (cf. (A.13)) we get

$$\begin{aligned} R &\leq \sum_{x_0<c_j-\rho\leq 2} w_j\left[\frac{8}{5\sqrt{2\pi}} + A - \Phi(x_0)\right] \\ &\quad + \sum_{j\,:\,c_j-\rho>2} w_j\left[\frac{16\ln 10}{5\sqrt{2\pi}\rho} + \frac{4(C-\rho)}{5\sqrt{2\pi}} + A - \Phi(2)\right] \\ &\leq a_1\left[\frac{8}{5\sqrt{2\pi}} + A - \Phi(x_0)\right] \\ &\quad + a_2\left[\frac{16\ln 10}{5\sqrt{2\pi}\,\Phi^{-1}(0.95)} + \frac{4(C-\rho)}{5\sqrt{2\pi}} + A - \Phi(2)\right] \\ &\leq a_2\,\frac{4(C-\rho)}{5\sqrt{2\pi}} + B''. \end{aligned} \tag{A.18}$$

Inserting (A.17) and (A.18) into (A.16) we obtain (A.15), thus finishing the proof. ∎

Inequalities (A.14) and (A.15) can be interpreted to say that for prior distributions satisfying (A.13) the loss of power for one bit of entropy of W is approximately $4(\ln 2)/(5\sqrt{2\pi}\rho) \approx 0.221/\rho$. Such a phenomenon can be observed in Table 3 and holds true approximately also for the optimal Bayes test and the tests $N_L(D,\delta)$ (see our comment on Table 3 in Section 3).

Remark A.3. Observe that, under the assumption of Theorem A.2, the estimates (A.14) and (A.15) are sharp. This follows since $C \leq c_j$,

$j = 1, \ldots, K$, and by (A.7) and (A.12) we have $C - \rho \le \Phi^{-1}(1 - \alpha/(2K)) - \Phi^{-1}(1 - \alpha/2)$, which for typical K and α is not much greater than 1 while $B \simeq 0.062$ and $1 - 2A + B + \alpha(\mathrm{A} - \mathrm{B}) \simeq 0.17$.

REMARK A.4. The assumption (A.12) means that the power of the two-sided Neyman–Pearson test of H_0 against $P_{j\pm}^{(n)}$ is approximately in the interval $[0.5, 0.925]$ while (A.13) is a kind of restriction on the magnitude of the entropy of the prior distribution W. Assumption (A.13) cannot be omitted and for a "wild" prior distribution W inequality (A.15) may not be true. To see this, consider $\rho = 2.5$, $\alpha = 0.05$, $K = 5001$ and $w_1 = 0.5$ while $w_2 = \cdots = w_{5001} = 0.0001$. Then $C \simeq 1.686$ and $\bar{\beta} \simeq 0.622$ from (A.8). However, (A.15) gives $\bar{\beta} \ge 0.742$, which is not true. On the other hand, (A.13) is not too restrictive. For example, if $\alpha = 0.05$, $\rho = 2$ and under the uniform prior distribution, (A.13) holds for relatively large $K \le 88$. The regularity assumption (A.13) can be replaced by another one. Our choice is, certainly, subjective and indicates rather what kind of restrictions are needed to get estimates similar to (A.14) and (A.15).

Acknowledgements. We are grateful to the referee for useful remarks and corrections which improved the presentation of the results. We also thank T. Ledwina for reading the manuscript and many constructive comments.

References

H. Akaike (1974), *A new look at statistical model identification*, IEEE Trans. Automat. Control 19, 716–723.

P. J. Bickel and Y. Ritov (1992), *Testing for goodness of fit: A new approach*, in: Nonparametric Statistics and Related Topics, A. Saleh (ed.), Amsterdam, North-Holland, 51–57.

G. Claeskens and N. L. Hjort (2004), *Goodness of fit via non-parametric likelihood ratios*, Scand. J. Statist. 31, 487–512.

B. S. Clarke and A. R. Barron (1990), *Information-theoretic asymptotics of Bayes methods*, IEEE Trans. Inform. Theory 36, 453–471.

R. L. Eubank and V. N. La Riccia (1992), *Asymptotic comparison of Cramér–von Mises and nonparametric function estimation techniques for testing goodness-of-fit*, Ann. Statist. 20, 2071–2086.

J. Fan (1996), *Test of significance based on wavelet thresholding and Neyman's truncation*, J. Amer. Statist. Assoc. 91, 674–688.

M. Fromont and B. Laurent (2006), *Adaptive goodness-of-fit tests in a density model*, Ann. Statist. 34, 680–720.

T. Inglot and A. Janic-Wróblewska (2003), *Data driven chi-square test for uniformity with unequal cells*, J. Statist. Comput. Simul. 73, 545–561.

T. Inglot, W. C. M. Kallenberg and T. Ledwina (1998), *Vanishing shortcoming of data driven Neyman's test*, in: Asymptotic Methods in Probability and Statistics. A volume in Honour of Miklós Csörgő, B. Szyszkowicz (ed.), Elsevier, Amsterdam, 811–829.

T. Inglot and T. Ledwina (1996), *Asymptotic optimality of data-driven Neyman's tests for uniformity*, Ann. Statist. 24, 1982–2019.

T. Inglot and T. Ledwina (2001), *Intermediate approach to comparison of some goodness-of-fit tests*, Ann. Inst. Statist. Math. 53, 810–834.

T. Inglot and T. Ledwina (2006a), *Towards data driven selection of a penalty function for data driven Neyman tests*, Linear Algebra Appl. 417, 124–133.

T. Inglot and T. Ledwina (2006b), *Data driven score tests for homoscedstic linear regression model: the construction and simulations*, in: Prague Stochastics 2006, M. Hušková and M. Janžura (eds.), Matfyzpress, Prague, 124–137.

A. Janic-Wróblewska and T. Ledwina (2000), *Data driven-rank test for two-sample problem*, Scand. J. Statist. 27, 281–297.

A. Janssen (2000), *Global power functions of goodness of fit tests*, Ann. Statist. 28, 239–253.

A. Janssen (2003), *Which power of goodness of fit tests can really be expected: intermediate versus contiguous alternatives*, Statistics and Decisions 21, 301–325.

W. C. M. Kallenberg (2002), *The penalty in data driven Neyman's tests*, Math. Methods Statist. 11, 323–340.

W. C. M. Kallenberg and T. Ledwina (1995), *Consistency and Monte Carlo simulation of a data driven version of smooth goodness of fit tests*, Ann. Statist. 23, 1594–1608.

W. C. M. Kallenberg and T. Ledwina (1997), *Data-driven smooth tests when the hypothesis is composite*, J. Amer. Statist. Assoc. 92, 1094–1104.

W. C. M. Kallenberg and T. Ledwina (1999), *Data driven rank tests for independence*, ibid. 94, 285–301.

M. A. Langovoy (2008), *Data-driven efficient score tests for deconvolution hypotheses*, Inverse Problems 24, 1–17.

T. Ledwina (1994), *Data driven version of Neyman's smooth test of fit*, J. Amer. Statist. Assoc. 89, 1000–1005.

A. V. Prokhorov (1973), *On sums of random vectors*, Theory Probab. Appl. 18, 186–188; Russian original: Teor. Veroyat. Primen. 18, 193–195.

G. Schwarz (1978), *Estimating the dimension of a model*, Ann. Statist. 6, 461–464.

G. Wyłupek (2008), *Data driven k-sample tests*, Technometrics, submitted.

Tadeusz Inglot
Institute of Mathematics
Polish Academy of Sciences
and
Institute of Mathematics and Informatics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: Tadeusz.Inglot@pwr.wroc.pl

Alicja Janic
Institute of Mathematics and Informatics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: Alicja.Janic@pwr.wroc.pl

Received on 23.10.2008;
revised version on 6.2.2009 (1970)