ELSEVIER

Asymptotic optimality of new adaptive test in regression model

Tadeusz Inglot [a], Teresa Ledwina [b],*

[a] *Polish Academy of Sciences and Wrocław University of Technology, Institute of Mathematics and Informatics, Wrocław University of Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland*
[b] *Polish Academy of Sciences, Institute of Mathematics, ul. Kopernika 18, 51-617 Wrocław, Poland*

Abstract

This paper presents some results on asymptotic optimality of the test procedure introduced recently by Baraud et al. [Y. Baraud, S. Huet, B. Laurent, Adaptive tests of linear hypotheses by model selection, Ann. Statist. 31 (2003) 225–251]. The optimality relies on some comparison of ability of the new statistic to discriminate between null hypothesis and convergent alternatives with capability of respective Neyman–Pearson test.
© 2005 Elsevier SAS. All rights reserved.

Résumé

Cet article présente des résultats d'optimalité asymptotique pour la procédure de test introduite récemment par Baraud et al. [Y. Baraud, S. Huet, B. Laurent, Adaptive tests of linear hypotheses by model selection, Ann. Statist. 31 (2003) 225–251]. L'optimalité repose sur la comparaison entre la capacité à distinguer l'hypothèse nulle d'alternatives convergentes de cette nouvelle procédure et celle du test de Neyman–Pearson correspondant.
© 2005 Elsevier SAS. All rights reserved.

1. Introduction

For testing an adequacy of some regression models, various tests, based on empirical Fourier coefficients, were proposed. Several examples can be found in Eubank [4], Fan and Huang [6] and Baraud et al. [1], e.g.

The paper of Baraud et al. [1] (BHL in what follows) provided a new solution in the area. The new test is a combination of tests for some selected subproblems. The purpose of the present contribution is to study some optimal asymptotic property of this new procedure.

There are many ways to define asymptotic optimality of tests. Two main streams are: asymptotic efficiency and minimaxity.

---

* Corresponding author. Fax: +48 71 3481098.
 E-mail address: ledwina@impan.pan.wroc.pl (T. Ledwina).

By asymptotic efficiency we mean that the optimal is the most efficient with respect to some asymptotic relative efficiency measure. We shall discuss briefly some aspects of the use of four known efficiency measures. Our notations and formulations shall be strictly related to those used in Section 1 of Nikitin [18]. We refer there for more detailed definitions and great amount of related information.

Consider some abstract testing problem

$$H : \theta \in \Theta_0 \subset \Theta$$

against

$$A : \theta \in \Theta_1 = \Theta \setminus \Theta_0.$$

Let n be a sample size and suppose we have two sequences of statistics $\{T_n\}$ and $\{V_n\}$, large values of them to be significant. Let α be given level of significance and let β be a given power. Assume $0 < \alpha < \beta < 1$ and denote by $N_T(\alpha, \beta, \theta)$ the minimal sample size necessary for the test, based on $\{T_n\}$, being on the level α to have power not less than β at the point $\theta \in \Theta_1$. $N_V(\alpha, \beta, \theta)$ is defined in the same way. The relative efficiency of $\{V_n\}$ with respect to $\{T_n\}$ is given by

$$e_{V,T}(\alpha, \beta, \theta) = N_T(\alpha, \beta, \theta)/N_V(\alpha, \beta, \theta).$$

Three classic notions of the asymptotic relative efficiency (ARE) of the sequence $\{V_n\}$ with respect to $\{T_n\}$ are as follows.

If for $0 < \alpha < \beta < 1$ and $\theta \to \theta_0 \in \partial\Theta_0$ (in a certain topology on Θ) there exists the limit

$$e_{V,T}^{\mathrm{P}}(\alpha, \beta, \theta_0) = \lim_{\theta \to \theta_0} e_{V,T}(\alpha, \beta, \theta), \qquad (1.1)$$

it is called the Pitman ARE.

If for $\beta \in (0, 1)$ and $\theta \in \Theta_1$ there exists the limit

$$e_{V,T}^{\mathrm{B}}(\beta, \theta) = \lim_{\alpha \downarrow 0} e_{V,T}(\alpha, \beta, \theta), \qquad (1.2)$$

it is called the Bahadur ARE.

If for $\alpha \in (0, 1)$ and $\theta \in \Theta_1$ there exists the limit

$$e_{V,T}^{\mathrm{HL}}(\alpha, \theta) = \lim_{\beta \uparrow 1} e_{V,T}(\alpha, \beta, \theta), \qquad (1.3)$$

it is called the Hodges–Lehmann ARE.

Though these measures are very useful for comparisons of various classic statistics, there are many situations when they fail to work.

In particular, the Pitman ARE is applicable in practice to asymptotically normal statistics T_n and V_n and for sequences of alternatives $\{\theta_n\}$ approaching $\theta_0 \in \partial\Theta_0$ at the rate $1/\sqrt{n}$. However, many statistics considered in contemporary literature have asymptotic power α for alternatives converging at the rate $1/\sqrt{n}$. Therefore, slower rates are required to get nontrivial and meaningful results. Needless to say that nonnormal asymptotic distributions are common, as well. Therefore, the use of Pitman efficiency is excluded in considerable number of important cases. On the other hand, the standard way of calculating the Bahadur ARE goes through a derivation of exact slopes. For this purpose large deviation asymptotics, under H_0, of $\{T_n\}$ and $\{V_n\}$ has to be established. Obviously, the asymptotics should be nontrivial. Again, it turns out that for some complicated statistics this requirement is not fulfilled. An important example is provided in Inglot [7]. Finally, note that an application of Hodges–Lehmann efficiency to several goodness-of-fit statistics leads to completely useless, from statistical point of view, conclusions (cf. Nikitin [17,18]). Briefly speaking, this efficiency is not able to discriminate statistics based on L_2-, sup- and other seminorms of the empirical process and the Neyman–Pearson statistic under known alternative. This is in obvious conflict with outcomes of many other efficiency measures as well as existing evidence on power behavior of such tests. For completeness note that similar effect is observed when one considers fixed α, large class of alternatives convergent to the null set and looks at the rate of convergence of the power β to 1. For some illustration see Inglot and Ledwina [11].

There is a growing evidence that the fourth ARE, which we shall discuss now, is much widely applicable and gives results which are consistent, to a reasonable extent, with finite sample power behavior of various, both classic and

modern, solutions. This efficiency was introduced by Kallenberg in 1983 as an intermediate approach between (1.1) and (1.2). Following the above scheme we can describe roughly this ARE as follows.

If for $\beta \in (0, 1)$, $\theta \to \theta_0 \in \partial \Theta_0$ (in a certain topology on Θ) and $\alpha \to 0$ at the controlled rate there exists the limit

$$e_{V,T}^{K}(\beta, \theta_0) = \lim_{\alpha \downarrow 0, \; \theta \to \theta_0} e_{V,T}(\alpha, \beta, \theta). \tag{1.4}$$

it is called the Kallenberg ARE.

Formally, the definition requires slightly modified quantities $N_T(\alpha, \beta, \theta)$ and $N_V(\alpha, \beta, \theta)$, but we shall not go here into such details. Note that the characteristic feature of this approach is that $\theta \to \theta_0$ slower than in (1.1) and $\alpha \to 0$ slower than in (1.2). It makes possible to apply this efficiency in problems when $1/\sqrt{n}$ rates are too restrictive. Moreover, slower rate of convergence of $\alpha \to 0$ allows one to replace the requirement on nondegenerate large deviations by less restrictive postulate on existence of nondegenerate moderate deviations. This further extends applicability of (1.4). Another appealing feature of the Kallenberg ARE is that simultaneously it allows to compare also procedures for which (1.1) and (1.2) work. Typically, the three ARE results then coincide. So, the range of statistics which can be compared by (1.4) is considerably large. Closing, note that (1.4) was named by the author as intermediate efficiency. For short, we shall call alternatives converging to some null parameter slower than $1/\sqrt{n}$ as intermediate ones.

In view of the discussed above increasing complexity of tests proposed recent years, sequences of intermediate alternatives were naturally incorporated to the minimax approach, as well. Last years, main stream of investigation in this area focused on the derivation of rates of consistent minimax distinguishability in relation to degree of smoothness of alternative sequences. Great interest in this area was stimulated by long series of Ingster's papers. However, recent results of Inglot and Ledwina [11] indicate that it is hard to find a reliable interpretation of this approach.

In view of the above, we have decided to show optimality of the BHL procedure in the spirit of the Kallenberg approach. Before presenting our solution, let us mention that BHL procedure is nonstandard in many aspects. First point is that the basic variant of this solution which we shall investigate, gives the test only approximately on the level α. Next point is, that the test statistic strongly depends on the given significance level. According to our knowledge, such case was not a subject of existing results on efficiency calculations. In particular, it is not clear how to calculate the Kallenberg efficiency in such case. Being faced with this situation, we have decided to elaborate at once more fine approach, which is closely related to Kallenberg's setting but concerns intermediate comparison of asymptotic power of the given procedure and the best one, in case the alternative would be known. More precisely, we shall show that for suitable sequences of levels $\{\alpha_n\}$ tending to 0 and related sequences of alternatives $\{\theta_n\}$ tending to the null one, both procedures can simultaneously be on the same level α_n and achieve the same nondegenerate asymptotic power under θ_n, as $n \to \infty$. We supplemented the above by including also the case of fixed alternatives. In this way we have proved that, under large class of convergent alternatives and some fixed ones, the BHL construction is asymptotically as informative as the Neyman–Pearson statistic, in case the alternative was known. In other words, the BHL solution efficiently adapts to the data at hand.

Our approach is discussed in more details in Subsection 2.2. The discussion is preceded by brief presentation of the variant of BHL solution which we consider. Technical tools that are needed to prove the optimality results are collected in Subsection 2.3. Subsection 2.4 contains formal statement of the optimality results and some related discussion. Section 3 contains some new auxiliary results on tails and quantiles of chi-square distribution. Also derivations of some technical tools needed to get the optimality are presented there. Appendix collects proofs of auxiliary results.

2. Optimality of BHL construction

2.1. The BHL construction

We consider a simplest set-up which is consistent with the framework of BHL. Namely, we assume that the observations have the structure

$$y_i = c + r(t_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where c is unknown constant, $r(t)$ is unknown function, $t_i = (2i - 1)/2n$, $i = 1, \ldots, n$, while $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $N(0, 1)$. Throughout we assume $r(t) \in L_2[0, 1]$ and $\int_0^1 r(t)\,dt = 0$. The null hypothesis asserts

H_0: $r \equiv 0$.

To present BHL test of H_0 we need several auxiliary notations.

Introduce the cosine system

$$b_k(t) = \sqrt{2}\cos(\pi k t), \quad k \geqslant 1, \ t \in [0, 1],$$

and the empirical Fourier coefficients

$$\hat{\phi}_k = \frac{1}{n}\sum_{i=1}^{n} y_i b_k(t_i), \quad k \geqslant 1. \tag{2.2}$$

Note that $\sum_{i=1}^{n} b_k(t_i) = 0$ for $k = 1, \ldots, n-1$ and therefore, for such k's, the distribution of $\hat{\phi}_k$ does not depend on c. Hence, in what follows we set $c = 0$.

Let χ_k^2 denotes a central chi-square distributed random variable with k degrees of freedom. Let $q_k(w)$ be $(1 - w)$-quantile of χ_k^2, i.e.

$$P\big(\chi_k^2 \geqslant q_k(w)\big) = w. \tag{2.3}$$

Throughout α is the prescribed significance level. We consider the test statistic (6) in BHL in case the procedure P2 on p. 228, ibidem, is used. In our application $\mathcal{M} = \{1, \ldots, l(n)\}$, for some $l(n)$, $l(n) < n$, $\|\Pi_m y\|_m^2 = \sum_{k=1}^{m}(\hat{\phi}_k)^2$ and $\alpha_m = \alpha/|\mathcal{M}| = \alpha/l(n)$. Therefore, rescaled by n, the formula (6) reads as follows

$$\widehat{T}_\alpha = \max_{1 \leqslant j \leqslant l(n)} \left\{ \sum_{k=1}^{j}(\sqrt{n}\hat{\phi}_k)^2 - q_j(w_n) \right\}, \quad w_n = \frac{\alpha}{l(n)}. \tag{2.4}$$

H_0 is rejected if $\widehat{T}_\alpha > 0$. With this choice of w_n, $\widehat{T}_\alpha$ is approximately on the level α. The null hypothesis is rejected if at least one of smooth tests, forming $\widehat{T}_\alpha$, rejects respective submodel.

BHL described the set of alternatives for which $\widehat{T}_\alpha$ has finite sample power at least β, $\beta \in (\alpha, 1)$. Their result can be interpreted as follows: under fixed alternative and fixed n, the power of $\widehat{T}_\alpha$ is comparable to the power of the best of smooth tests forming $\widehat{T}_\alpha$ (cf. BHL, p. 231). As mentioned in Section 1, our approach is asymptotic and allows some comparison of $\widehat{T}_\alpha$ with the best existing solution, in case the alternative was known. Moreover, since the optimality notion which we use was applied earlier to some other tests, one can do some further comparisons. Cf. Remark 2 of Section 2.4 for related discussion. Another advantage of our approach is that it can be applied to non-Gaussian models as well. For some evidence see Kallenberg [16] and Inglot and Ledwina [9], e.g.

2.2. Further discussion of the optimality notion

Our approach has roots in the notion of vanishing shortcoming which was initiated by Oosterhoff [19]. This notion and some simplified variants of it were further elaborated in Oosterhoff and van Zwet [20], Kallenberg [15], Inglot et al. [8], Inglot and Ledwina [9,10], Kallenberg [16], Ducharme and Ledwina [3], among others. Here we shall restrict mainly our attention to the core part of the notion of vanishing shortcoming, similarly as treated recently in Ducharme and Ledwina [3], e.g. Since the terminology is rich, to avoid misunderstanding, we shall tell here simply on optimality.

We consider independent random variables (y_i, ϵ_i), $i = 1, \ldots, n$, $y_i = r(t_i) + \epsilon_i$, $\epsilon_i \sim N(0, 1)$, the hypothesis $H_0: r \equiv 0$ and auxiliary alternative

$$A_0: \quad r(t) = r_n(t) = \vartheta_n g(t),$$

where

$$g \in \mathcal{F} = \left\{ f: \sup_{t \in [0,1]} |f(t)| < \infty, \ \int_0^1 f(t)\,dt = 0, \ \int_0^1 f^2(t)\,dt = 1 \right\},$$

while $\vartheta_n \in R$ is a constant or $\vartheta_n \to 0$ as $n \to \infty$.

It is known that vanishing shortcoming can be expected only under fixed or intermediate alternatives (cf. Kallenberg [15]). Therefore, we shall restrict attention to the case $n\vartheta_n^2 \to \infty$, only. Similarly as in Ducharme and Led-

wina [3], we shall state our results only for some simple subclass of the above described alternatives. Namely, we shall consider alternatives of the form

$$A_0^*: \quad r_n(t) = \frac{\rho}{n^\xi} f(t), \quad f(t) = \sum_{k=1}^{d} \phi_k b_k(t), \quad \rho > 0, \quad \sum_{k=1}^{d} \phi_k^2 = 1, \tag{2.5}$$

where $d = d(f)$ is fixed but otherwise arbitrary, $\phi_d \neq 0$ and $\phi_1, \ldots, \phi_d$ are Fourier coefficients of f.

This simple case has the advantage that many technicalities can be skipped. Obviously, the simplification causes that some aspects are less seen or not seen at all. However, our main purpose is to discuss the core of the approach, its interpretation and possible further developments. Therefore, it seems it is sufficient to elaborate this basic case, only.

On the other hand, since the Gaussian case, we consider here, is relatively simple, some bigger flexibility in rates of convergence in (2.5), than those allowed in Ducharme and Ledwina [3] and some earlier developments, can be achieved. It turns out that, in the present case, the whole range $[0, 1/2)$ for ξ can be covered (cf. Remark 1 of Section 2.4). Therefore we shall consider

$$\rho_n = \sqrt{n}\left(\rho n^{-\xi}\right) \quad \text{and} \quad \xi \in [0, 1/2). \tag{2.6}$$

Denote for a moment by NP the Neyman–Pearson statistic for testing H_0 against A_0^*. In view of the peculiarities of the BHL construction, which we discussed in the final part of Section 1, in contrast to some previous developments, we shall phrase here the optimality criterion as a property of the test statistics rather than tests, it-selves. It seems that this new formulation allows also for more handy interpretation of previous developments.

Roughly speaking, the optimality of $\widehat{T}_\alpha$ means that, for large class of levels α_n, $\alpha_n \to 0$ as $n \to \infty$, one can construct right-handed critical regions based on NP and $\widehat{T}_\alpha$ such that both regions have the same exact size α_n for n sufficiently large, and the same asymptotic power $\beta \in (0, 1)$. This means that, for large n, the statistic $\widehat{T}_\alpha$ recovers the unknown alternative A_0^* as precisely as the best statistic in case A_0^* was known. Another words, from the point of view of controlling the precision of inference in terms of exact error of the first kind and asymptotic, under intermediate alternatives, error $1 - \beta \in (0, 1)$ of the second kind, $\widehat{T}_\alpha$ is as informative as NP.

To control the asymptotic power and exact size of tests based on $\widehat{T}_\alpha$ and NP we shall use the results collected below.

2.3. Technical tools needed to derive the optimality

Denote by P_n the distribution of the sample induced by (2.5) and (2.6) and set P_0 for the distribution of the sample under H_0.

Theorem 1. *Assume that $l(n) \to \infty$ and $l(n) = \mathrm{o}(\rho_n)$ as $n \to \infty$. Then under* (2.5) *and* (2.6)

$$\lim_{n \to \infty} P_n\left(\frac{\widehat{T}_\alpha - \rho_n^2}{2\rho_n} \leqslant x\right) = \boldsymbol{\Phi}(x), \quad x \in R,$$

where $\boldsymbol{\Phi}$ is cdf of $N(0, 1)$ random variable.

Theorem 2. *Let $\{x_n\}$ be a real sequence such that $nx_n^2 \to \infty$. Assume that $nx_n^2 \geqslant l(n)$ for sufficiently large n. Then*

$$P_0\left(\widehat{T}_\alpha \geqslant nx_n^2\right) \leqslant \exp\left\{-\frac{1}{2}nx_n^2 + \frac{1}{2}l(n)\log nx_n^2 + \mathrm{O}(1)\right\}.$$

Assume now that we know the alternative P_n (cf. (2.5) and (2.6)). The logarithm of the Neyman–Pearson statistic for H_0 against P_n has the form

$$\mathcal{L}_n = \rho_n V_n - \frac{1}{2}\rho_n^2 c_n(f),$$

where

$$V_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_i f(t_i) \quad \text{and} \quad c_n(f) = \frac{1}{n} \sum_{i=1}^{n} f^2(t_i). \tag{2.7}$$

We have $E_{P_n}\mathcal{L}_n = \frac{1}{2}\rho_n^2 c_n(f)$ and $\mathrm{Var}_{P_n}\mathcal{L}_n = \rho_n^2 c_n(f)$. Hence

$$P_n\left(\frac{\mathcal{L}_n - E_{P_n}\mathcal{L}_n}{\sqrt{\mathrm{Var}_{P_n}\mathcal{L}_n}} \leqslant x\right) = P_n\left(\frac{V_n - \rho_n c_n(f)}{[c_n(f)]^{1/2}} \leqslant x\right) = \boldsymbol{\Phi}(x), \quad x \in R.$$

Therefore, in what follows, V_n will be our benchmark. Recall that $\int_0^1 f(t)\,dt = 0$ and $\int_0^1 f^2(t)\,dt = 1$. The form of f implies also that f is bounded and satisfies Lipschitz condition. This implies $|c_n(f) - 1| \leqslant c/n$ for some constant c depending only on f. So, by (2.6) we have

$$\lim_{n\to\infty} P_n(V_n - \rho_n \leqslant x) = \boldsymbol{\Phi}(x).$$

Since $E_{P_0}V_n = 0$ and $\mathrm{Var}_{P_0}V_n = c_n(f)$ then the standard bounds for tails of $\boldsymbol{\Phi}$ and the above imply

Theorem 3. *Let $\{x_n\}$ be a sequence of positive numbers such that $nx_n^2 \to \infty$. Then*

$$P_0(V_n \geqslant \sqrt{n}x_n) = 1 - \boldsymbol{\Phi}\left(\sqrt{n}x_n / \{c_n(f)\}^{1/2}\right) = \exp\left\{-\frac{1}{2}nx_n^2 - \frac{1}{2}\log nx_n^2 + \mathrm{O}(1)\right\}. \tag{2.8}$$

Theorems 1 and 2 will be proved in Section 3. Now, we shall present our basic result.

2.4. Optimality result, some conclusions and remarks

For $k \in R$ set

$$\mathcal{C}_{n,k}^{(1)} = \left\{V_n - \rho n^{1/2-\xi} \geqslant k\right\}, \qquad \mathcal{C}_{n,k}^{(2)} = \left\{\frac{\widehat{T}_\alpha - \rho^2 n^{1-2\xi}}{2\rho n^{1/2-\xi}} \geqslant k\right\}$$

and

$$\alpha_n = P_0\left(\mathcal{C}_{n,k}^{(2)}\right).$$

Theorem 4. *Assume P_n obeys (2.5) for some $\rho > 0$, $\xi \in [0, 1/2)$ and arbitrary f having finite expansion in the system $b_1, b_2, \dots$. Suppose $l(n) \to \infty$ and $l(n) = \mathrm{o}(n^{1/2-\xi}/\log n)$. Then, for any $k \in R$, there exists a real sequence $\{o_n\}$, $o_n \to 0$ as $n \to \infty$, such that for critical regions $\mathcal{C}_{n,k}^{(2)}$ and $\mathcal{C}_{n,k+o_n}^{(1)}$ and all n sufficiently large it holds that*

$$\alpha_n = P_0\left(\mathcal{C}_{n,k}^{(2)}\right) = P_0\left(\mathcal{C}_{n,k+o_n}^{(1)}\right) = \exp\left\{-\frac{1}{2}\rho^2 n^{1-2\xi} - k\rho n^{1/2-\xi} + \mathrm{o}\left(n^{1/2-\xi}\right)\right\}$$

and simultaneously

$$\lim_{n\to\infty} P_n\left(\mathcal{C}_{n,k}^{(2)}\right) = \lim_{n\to\infty} P_n\left(\mathcal{C}_{n,k+o_n}^{(1)}\right) = 1 - \boldsymbol{\Phi}(k).$$

Theorem 4 can be proved in the same way as Theorem 4.1 in Ducharme and Ledwina [3], e.g. Therefore we omit a proof.

Remark 1. Due to the Gaussian model, we consider, no asymptotics is needed to expand $P_0(V_n \geqslant \sqrt{n}x_n)$ and only pure Gaussian tails are involved. This is a reason that in this case we have got the result for any $\xi \in [0, 1/2)$, in contrast to some previous results holding for $\xi \in (1/4, 1/2)$, only (cf. Theorem 3.3(3) in Inglot and Ledwina [10] and Theorem 4.1 in Ducharme and Ledwina [3], e.g.).

There are possible some reformulations and extensions of Theorem 4 in the spirit of Theorems 5.1 and 5.2 of Inglot et al. [8] (cf. also Theorem 4.6 and 4.7 in Kallenberg [16]). To state them let us denote by $\beta_n^+(\gamma, P)$ the power under P of the level γ test with critical region $\{V_n - \rho n^{1/2-\xi} \geqslant k_{1,\gamma}\}$. Similarly set $\beta_n(\gamma, P)$ to be the power under P of the level γ test with critical region

$$\left\{\frac{\widehat{T}_\alpha - \rho^2 n^{1-2\xi}}{2\rho n^{1/2-\xi}} \geqslant k_{2,\gamma}\right\}.$$

Put

$$R_n(\gamma, P) = \beta_n^+(\gamma, P) - \beta_n(\gamma, P). \tag{2.9}$$

The quantity (2.9) is called shortcoming.

Theorem 5. *Let $\{t_n\}$ be a sequence of real numbers and set*

$$\alpha_n = P_0\left(\frac{\widehat{T}_\alpha - \rho^2 n^{1-2\xi}}{2\rho n^{1/2-\xi}} \geqslant t_n\right). \tag{2.10}$$

Suppose

$$-\infty < \liminf_{n\to\infty} t_n \leqslant \limsup_{n\to\infty} t_n < \infty.$$

Then, under the assumptions of Theorem 4, we have

$$\lim_{n\to\infty} R_n(\alpha_n, P_n) = 0.$$

Theorem 5 shows that the shortcoming vanishes in the case of arbitrary α_n, $\alpha_n \to 0$, such that the asymptotic power of the tests under consideration stays away 0 and 1.

The next modification covers the case of arbitrary asymptotic power.

Theorem 6. *Under the assumptions of Theorem 4 and arbitrary $\alpha_n \to 0$, given by (2.10), it holds*

$$\lim_{n\to\infty} R_n(\alpha_n, P_n) = 0. \tag{2.11}$$

It is intuitive that almost the same argument proves Theorems 4 and 5. For some evidence see proof of Theorem 4.6 in Kallenberg [16]. Theorem 6 can be proved exactly as Theorem 5.2 in Inglot et al. [8].

The property (2.11) is called vanishing shortcoming.

Remark 2. Theorem 4 and its derivatives show that the optimality of $\widehat{T}_\alpha$ provides an evidence that the construction, evaluated from the point of view of the theory of testing, well exploits the information contained in the sample.

Classic goodness of fit statistics are, as a rule, not optimal in the sense we consider. Illustrative considerations contained in Inglot and Ledwina [10] nicely show that the approach clearly exhibits well known weaknesses of such constructions. Moreover, the approach makes it possible to express in quantitative and nicely interpretable way the degree to which the optimality is violated (cf. Theorem 3.3(2) in Inglot and Ledwina [10]). Also in the present problem one could indicate some tests which are not optimal. Natural candidates are counterparts of classic solutions. For some evidence see Eubank [4]. Recent results of Inglot and Ledwina [12] are the premise for such a claim in the case of Eubank and Hart [5] and Fan and Huang [6] solutions, as well.

It is also worth noticing that it is not excluded that there are many optimal solutions in a given problem. An example of such situation is uniformity testing and the class of data driven statistics (cf. Kallenberg [16]). It is expected that similar conclusion shall take place for counterparts of these constructions in the problem considered in the present paper. For example, the statistic I_λ considered in Eubank [4] is an example of such counterpart, which can be proved to be optimal for testing H_0 against A_0^*.

It is also clear that under fixed alternative and fixed α finite sample power behavior of optimal constructions are not necessarily the same. Therefore some further analysis and discussion, taking into account some desirable features in a possible application, is welcome to choose the best solution. For some comments on such aspects of optimal data driven statistics see Inglot and Ledwina [13].

On the other hand, the exhibited existence of a whole class of optimal statistics is a good motivation for elaborating more subtle variant of the notion of optimality which would allow for ordering them.

3. Proofs of results of Section 2.3

3.1. Some auxiliary results

We start with some bounds for tails and quantiles of χ_k^2 – a random variable with central chi-square distribution with k degrees of freedom.

Set

$$
c(k) = \frac{2}{\sqrt{k}\,\Gamma(k/2)} \left(\frac{k}{2e} \right)^{k/2}, \qquad \mathcal{E}_k(x) = \exp\left\{ -\frac{x}{2} + \frac{k}{2} + \frac{k-2}{2}\log\frac{x}{k} - \frac{1}{2}\log k \right\},
$$
$$
\mathcal{G}_k(x) = \frac{x}{\sqrt{k}} \mathcal{E}_k(x) = \exp\left\{ -\frac{x}{2} + \frac{k}{2}\log\frac{ex}{k} \right\}.
\tag{3.1}
$$

Lemma 1. *For $k \geqslant 2$ and $x > 0$ it holds*

$$
P\left(\chi_k^2 \geqslant x\right) \geqslant c(k)\mathcal{E}_k(x) \geqslant \frac{1}{2}\mathcal{E}_k(x)
\tag{3.2}
$$

while for $k \geqslant 2$ and $x > k - 2$ we have

$$
P\left(\chi_k^2 \geqslant x\right) \leqslant c(k)\left[\frac{x}{x-k+2} \right]\mathcal{E}_k(x) \leqslant \frac{1}{\sqrt{2\pi}}\left[\frac{x}{x-k+2} \right]\mathcal{E}_k(x).
\tag{3.3}
$$

Hence, for $k \geqslant 2$ and $x > k - 2$ it follows

$$
\frac{1}{2}\left[\frac{\sqrt{k}}{x} \right]\mathcal{G}_k(x) \leqslant P\left(\chi_k^2 \geqslant x\right) \leqslant \frac{1}{\sqrt{2\pi}}\left[\frac{\sqrt{k}}{x-k+2} \right]\mathcal{G}_k(x).
\tag{3.4}
$$

Next lemma presents a bound which, for large x's, is less sharp than (3.4) but sometimes is more convenient in applications.

Lemma 2. *For any $k \geqslant 1$ and $x \geqslant k$ we have*

$$
P\left(\chi_k^2 \geqslant x\right) \leqslant \mathcal{G}_k(x).
\tag{3.5}
$$

The form of $\widehat{T}_\alpha$ (cf. (2.4)) causes also a need to deal with the quantiles $q_k(\cdot)$. Our collection of results on this quantity starts with an upper bound derived by Birgé [2] (cf. his (8.34) or (15) in BHL).

For all $k \geqslant 1$ and all $w \in (0, 1)$, $\quad q_k(w) \leqslant k + 2\sqrt{k\log(1/w)} + 2\log(1/w)$. $\tag{3.6}$

Moreover, we have

Lemma 3.

(1) *For any $w \in (0, 1)$ $q_1(w) \leqslant 1 + 2\log(1/w) - \log(1 + 2\log(1/w))$.* $\tag{3.7}$
(2) *For any $w \in (0, 1/2)$ $q_1(w) \geqslant \log(1/2w)$.* $\tag{3.8}$
(3) *For any $w \in (0, 1)$ $q_2(w) = 2\log(1/w)$.* $\tag{3.9}$
(4) *For any $k \geqslant 3$ and all $w \leqslant 1/k$ $q_k(w) \geqslant k + 2\log(1/w) - 3$.* $\tag{3.10}$
   A refined variant of (4) is as follows.
(5) *For any $k \geqslant 3$ and all $w \leqslant e^{-k/3}$*

$$
q_k(w) \geqslant k + 2\log(1/w) + \frac{1}{3}\sqrt{k}\log\log(1/w) - Q, \quad Q = 3.1.
\tag{3.11}
$$

(6) *For any $k \geqslant 32$ and all $w \in (e^{-k/3}, 1/k)$ $q_k(w) \geqslant k + 2\log(1/w) + \frac{1}{4}\sqrt{k\log(1/w)}$.* $\tag{3.12}$

Proofs of Lemmas 1 and 3 are given in Appendix A. An easy proof of Lemma 2 can be found in Inglot and Ledwina [12].

Now we shall present some simple useful facts on exact, approximate and empirical Fourier coefficients. Recall that the distribution of y_i's under null model was denoted by P_0. Consider now the alternatives defined by

$$r(t) = r_n(t) = \frac{\rho}{n^\xi} f(t), \quad \rho > 0, \quad \xi \in [0, 1/2], \quad \int_0^1 f(t)\,dt = 0, \quad \int_0^1 f^2(t)\,dt = 1. \tag{3.13}$$

Introduce

$$\phi_k = \int_0^1 f(t)b_k(t)\,dt, \quad \phi_{kn} = \frac{1}{n}\sum_{i=1}^n f(t_i)b_k(t_i), \quad \hat{\phi}_k = \frac{1}{n}\sum_{i=1}^n y_i b_k(t_i) \tag{3.14}$$

and set

$$\tilde{\phi} = (\phi_{1n}, \ldots, \phi_{n-1\,n}), \qquad \hat{\phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_{n-1}).$$

Lemma 4 ((Eubank [4])). *Assume that $f' \in L_2[0, 1]$ and $\int_0^1 f(t)\,dt = 0$. Then there is a constant A independent of k and n such that*

$$|\phi_{kn} - \phi_k| \leqslant A/n, \quad 1 \leqslant k \leqslant n - 1. \tag{3.15}$$

Consequently, for any integer $1 \leqslant j \leqslant n - 1$

$$\left| \sum_{k=1}^j \phi_{kn}^2 - \sum_{k=1}^j \phi_k^2 \right| = O\left(\frac{\sqrt{j}}{n}\right). \tag{3.16}$$

Another helpful observation is that for some i.i.d. $Z_1, \ldots, Z_{n-1}$, $Z_i \sim N(0, 1)$ and $Z = (Z_1, \ldots, Z_{n-1})$, it holds

$$\sqrt{n}\hat{\phi} \overset{D}{=} Z \quad \text{under } H_0,$$
$$\sqrt{n}\hat{\phi} \overset{D}{=} Z + \rho_n\tilde{\phi}, \quad \rho_n = \rho n^{1/2-\xi} \quad \text{under (3.13)}. \tag{3.17}$$

Finally for $v = (v_1, v_2, \ldots) \in \mathbb{R}^\infty$ we set

$$|v|_j = \left(\sum_{k=1}^j v_k^2 \right)^{1/2} \quad \text{for } j \geqslant 1 \quad \text{and} \quad |v|_{jk}^2 = |v|_k^2 - |v|_j^2 \quad \text{for } 1 \leqslant j < k. \tag{3.18}$$

The same convention shall be applied to the components of $Z = (Z_1, \ldots, Z_{n-1})$ and $\sqrt{n}\hat{\phi} = \sqrt{n}(\hat{\phi}_1, \ldots, \hat{\phi}_{n-1})$.

Now we shall rewrite the test statistic $\widehat{T}_\alpha$ in an equivalent form. For this purpose set

$$\tilde{m} = \min\{m\colon 1 \leqslant m \leqslant l(n),\ \big|\sqrt{n}\hat{\phi}\big|_m^2 - q_m(w_n) \geqslant \big|\sqrt{n}\hat{\phi}\big|_j^2 - q_j(w_n),\ j = 1, \ldots, l(n)\}. \tag{3.19}$$

With this notation

$$\widehat{T}_\alpha = \big|\sqrt{n}\hat{\phi}\big|_{\tilde{m}}^2 - q_{\tilde{m}}(w_n). \tag{3.20}$$

3.2. Proof of Theorem 1

Since $l(n) \to \infty$, by (2.4) and (3.17), we infer that

$$P_n\left(\frac{\widehat{T}_\alpha - \rho_n^2}{2\rho_n} \leqslant x \right) \leqslant P_0\left(\frac{|Z + \rho_n\tilde{\phi}|_d^2 - q_d(w_n) - \rho_n^2}{2\rho_n} \leqslant x \right)$$
$$= P_0\left(\frac{1}{2\rho_n}|Z|_d^2 + \sum_{k=1}^d \phi_{kn} Z_k + \frac{\rho_n}{2}\left(\sum_{k=1}^d \phi_{kn}^2 - 1 \right) \leqslant x + \frac{1}{2\rho_n} q_d(w_n) \right).$$

By (3.16) and properties of f in (2.5), we have $|\tilde{\phi}|_d^2 - 1 = O(1/n)$. The upper bound (3.6) and the assumption $l(n) = o(\rho_n)$ imply that $q_d(w_n)/\rho_n \to 0$. Since $(\sum_{k=1}^d \phi_{kn} Z_k)/|\tilde{\phi}|_d \sim N(0, 1)$, the above implies

$$\limsup_{n \to \infty} P_n\left(\frac{\widehat{T}_\alpha - \rho_n^2}{2\rho_n} \leqslant x \right) \leqslant \boldsymbol{\Phi}(x), \quad x \in R.$$

On the other hand, since $q_{\tilde{m}}(w_n) \geqslant 0$ and $\tilde{m} \leqslant l(n)$ we get

$$P_n\left(\frac{\widehat{T}_\alpha - \rho_n^2}{2\rho_n} \leqslant x \right) \geqslant P_0\left(\frac{|Z + \rho_n \tilde{\phi}|_{l(n)}^2 - \rho_n^2}{2\rho_n} \leqslant x \right).$$

By (3.5) and similar argument as above, the proof is concluded. $\quad\square$

3.3. Proof of Theorem 2

Exploiting (3.20), (3.17), (3.8), (3.9) and (3.10) we obtain for n sufficiently large

$$
\begin{aligned}
P_0\big(\widehat{T}_\alpha \geqslant nx_n^2\big) &= \sum_{k=1}^{l(n)} P_0\big(\widehat{T}_\alpha \geqslant nx_n^2, \tilde{m} = k\big) \\
&\leqslant \sum_{k=1}^{l(n)} P_0\big(|\sqrt{n}\hat{\phi}|_k^2 - q_k(w_n) \geqslant nx_n^2\big) \leqslant P_0\big(\chi_1^2 \geqslant nx_n^2 + \log(1/2w_n)\big) \\
&\quad + P_0\big(\chi_2^2 \geqslant nx_n^2 + \log(1/w_n)\big) + \sum_{k=3}^{l(n)} P_0\big(\chi_k^2 \geqslant nx_n^2 + \log(1/w_n) + k - 3\big).
\end{aligned}
\tag{3.21}
$$

The first term of (3.21) can be majorized using standard bounds for normal tails. The second term can be calculated explicitly. Omitting $k - 3$ in the third term and exploiting the assumption $nx_n^2 \geqslant l(n)$ an application of (3.5) concludes the proof. $\quad\square$

Acknowledgements

Appendix A. Proofs of auxiliary lemmas

Proof of Lemma 1. Set

$$I_k(x) = \int_x^\infty u^{(k/2)-1} e^{-u/2} \, du.$$

Then $I_k(x) = [2^{k/2}\Gamma(k/2)]P(\chi_k^2 \geqslant x)$. Integration by parts yields for any $k \geqslant 1$ and any $x > 0$

$$I_k(x) = 2x^{k/2-1} e^{-x/2} + (k-2)\int_x^\infty \frac{1}{u} u^{k/2-1} e^{-u/2} \, du.$$

Hence, for $k \geqslant 2$ and $x > 0$ we get

$$P\big(\chi_k^2 \geqslant x\big) \geqslant \frac{2}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}. \tag{A.1}$$

Now, observe that for $k \geqslant 2$ we have $I_k(x) \leqslant 2x^{k/2-1} e^{-x/2} + [(k-2)/x]I_k(x)$. Hence, for $x > k - 2$, we obtain

$$P\big(\chi_k^2 \geqslant x\big) \leqslant \frac{2}{2^{k/2}\Gamma(k/2)} \left[\frac{x}{x-k+2} \right] x^{k/2-1} e^{-x/2}. \tag{A.2}$$

Using the notations (3.1), (A.1) and (A.2) can be equivalently stated as

$$P(\chi_k^2 \geqslant x) \geqslant c(k)\mathcal{E}_k(x) \quad \text{and} \quad P(\chi_k^2 \geqslant x) \leqslant c(k)\left[\frac{x}{x-k+2}\right]\mathcal{E}_k(x). \tag{A.3}$$

In view of (A.3), the proof of (3.2)–(3.4) shall be concluded, provided that we shall show that for each $k \geqslant 2$ it holds $c(k) \in (1/2, 1/\sqrt{2\pi}]$. For this purpose observe that for any natural k we have

$$\frac{c(k+2)}{c(k)} = \mathrm{e}^{-1}\left(\frac{k+2}{k}\right)^{(k+1)/2}$$

while for $k \geqslant 2$ it holds $2[(k+2)/k]^{(k+1)/2} \geqslant \mathrm{e}$. The last statement is an easy consequence of series expansion of $\log(1+x)$ for $x = 2/k$. Observing that $c(3) > c(2) > 1/2$ we infer that $c(k) > 1/2$. On the other hand, by Stirling's formula, $\lim_{s\to\infty} c(2s) = \lim_{s\to\infty} c(2s-1) = 1/\sqrt{2\pi}$. This and the monotonicity of related subsequences, conclude the proof. $\quad\square$

Proof of Lemma 3. Set $q_1^0(w) = 1 + 2\log(1/w) - \log(1 + 2\log(1/w))$. To prove (1) it is enough to show that $w \geqslant P(\chi_1^2 \geqslant q_1^0(w))$. Since $P(\chi_1^2 \geqslant q_1^0(w)) = 2(1 - \boldsymbol{\Phi}(\sqrt{q_1^0(w)}))$, due to $1 - \boldsymbol{\Phi}(x) \leqslant (1/\sqrt{2\pi}x)\exp\{-x^2/2\}$, the problem reduces to the question if $\sqrt{2/\pi\mathrm{e}}\sqrt{u/(u - \log u)} \leqslant 1$, where $u = 1 + 2\log(1/w)$. Since the function $\sqrt{u/(u - \log u)}$ has the maximal value equal to $\sqrt{\mathrm{e}/(\mathrm{e}-1)}$, the conclusion follows.

Using the bound $\boldsymbol{\Phi}(x) \leqslant \frac{1}{2}(1 + \sqrt{1 - \exp\{-x^2\}})$ (cf. (13.48) in Johnson et al. [14]) (2) follows by the similar argument.

Since χ_2^2 has the exponential distribution, (3) is obvious.

To prove (4) set $q_k^*(w) = k + 2\log(1/w) - 3$. By (3.2) it is enough to show that $\frac{1}{2}\mathcal{E}_k(q_k^*(w)) \geqslant w$. The last holds if $f_1(t) \geqslant 0$, where $f_1(t) = (k-2)\log(1 + (2t-3)/k) + 3 - \log k - \log 4$, $t = \log(1/w)$, with $t \geqslant \log k$, by the assumption. Since f_1 is increasing, it holds $f_1(t) \geqslant f_1(\log k)$. The rest of the proof follows by an elementary checking that $f(\log k) \geqslant 0$ for $k \geqslant 3$.

The proof of (5) is similar to that of (4). We set $q_k^*(w) = k + 2\log(1/w) + \frac{1}{3}\sqrt{k}\log\log(1/w) - Q$ and, using (3.2), we reduce the problem to checking if for $t = \log(1/w)$, $t \geqslant k/3$, the function

$$f_2(t) = (k-2)\log\left\{1 + \left(2t + \frac{1}{3}\sqrt{k}\log t - Q\right)\Big/k\right\} - \frac{1}{3}\sqrt{k}\log t + Q - \log 4k$$

is nonnegative. Since $k \geqslant 3$, the function f_2 is increasing on $(1, \infty)$. Therefore it is enough to verify that $f_2(k/3) \geqslant 0$. We have $(k-2)^{-1}f_2(k/3) = g_1(k) + g_2(k) + g_3(k)$, where

$$g_1(k) = \log\left(\frac{5}{3} + \left\{\left[\frac{\sqrt{k}}{3}\right]\log\left(\frac{k}{3}\right) - Q\right\}\Big/k\right), \quad g_2(k) = -\left[\sqrt{k}\log\left(\frac{k}{3}\right)\right]\Big/3(k-2), \quad g_3(k) = \frac{Q - \log 4k}{k-2}.$$

Both functions $g_2(k)$ and $g_3(k)$ attain minimal value at $k = 13$. These values equal -0.0771 and -0.1602, respectively (up to four decimal digits). Therefore $(k-2)^{-1}f_2(k/3) \geqslant \log(5/3 - Q/k) - 0.2373$, which is positive for $k \geqslant 8$. For $k = 3, \ldots, 7$ one can verify by straightforward calculation that $\sum_{i=1}^3 g_i(k) > 0$.

To prove (6) set $q_k^*(w) = k + 2\log(1/w) + \frac{1}{4}\sqrt{k\log(1/w)}$ and $t = \log(1/w)$ again. By (3.2), we reduce the problem to checking that $f_3(t) \geqslant 0$, where $f_3(t) = (k-2)\log\{1 + (2t + \frac{1}{4}\sqrt{kt})/k\} - \frac{1}{4}\sqrt{kt} - \log 4k$. Since $\log k \leqslant t \leqslant k/3$ we can apply the inequality $\log(1+y) \geqslant y - y^2/2$, $y \in (0, 1)$. Since $\frac{1}{2}\sqrt{t/k} \leqslant 1/\sqrt{12}$ this yields $f_3(t) \geqslant [(k-2)/k]\varphi_k(t) - \log 4k - 1/\sqrt{12}$, where $\varphi_k(t) = \frac{63}{32}t - \frac{1}{2\sqrt{k}}t^{3/2} - \frac{2}{k}t^2$. Since φ_k is increasing on $[\log k, k/3]$ we get $f_3(t) \geqslant [(k-2)/k]\varphi_k(\log k) - \log 4k - \frac{1}{\sqrt{12}}$. As $[\log y]/y$ is decreasing for $y > \mathrm{e}$, the above yields $f_3(t) \geqslant g(k)$, for $k \geqslant 32$, with

$$g(k) = \left[\frac{k-2}{k}\right][\log k]\left(\frac{63}{32} - \frac{1}{2}\sqrt{\frac{\log 32}{32}} - 2\frac{\log 32}{32}\right) - \log 4k - 1/\sqrt{12}.$$

Since $g(k)$ is increasing we have $g(k) \geqslant g(32) > 0$ for $k \geqslant 32$. This concludes the proof. $\quad\square$

References

[1] Y. Baraud, S. Huet, B. Laurent, Adaptive tests of linear hypotheses by model selection, Ann. Statist. 31 (2003) 225–251.

[2] L. Birgé, An alternative point of view on Lepski's method, in: State of the Art in Probability and Statistics, Leiden 1999, IMS, Beachwood, OH, 2001, pp. 113–133.

[3] G.R. Ducharme, T. Ledwina, Efficient and adaptive nonparametric test for the two-sample problem, Ann. Statist. 31 (2003) 2036–2058.

[4] R.L. Eubank, Testing for no effect by cosine series methods, Scand. J. Statist. 27 (2000) 747–763.

[5] R.L. Eubank, J.D. Hart, Testing goodness of fit in regression via order selection criteria, Ann. Statist. 20 (1992) 1412–1425.

[6] J. Fan, L.-S. Huang, Goodness-of-fit tests for parametric regression models, J. Amer. Statist. Assoc. 96 (2001) 640–652.

[7] T. Inglot, On large deviation theorem for data driven Neyman's statistic, Statist. Probab. Lett. 47 (2000) 487–509.

[8] T. Inglot, W.C.M. Kallenberg, T. Ledwina, Vanishing shortcoming of data driven Neyman's tests, in: Asymptotic Methods in Probability and Statistics, A Volume to Honour Miklós Csörgő, Elsevier, Amsterdam, 1998, pp. 811–829.

[9] T. Inglot, T. Ledwina, Asymptotic optimality of data driven smooth tests for location-scale family, Sankhyā Ser. A 63 (2001) 41–71.

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

[11] T. Inglot, T. Ledwina, On consistent minimax distinguishability and intermediate efficiency of Cramér–von Mises test, J. Statist. Plann. Inference 124 (2004) 453–474.

[12] T. Inglot, T. Ledwina, Intermediate efficiency of some max-type statistics, J. Statist. Plann. Inference (2006), in press.

[13] T. Inglot, T. Ledwina, Towards data driven selection of penalty function for data driven Neyman tests, Linear Algebra Appl. (2005), in press.

[14] N.L. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, vol. 1, Wiley Interscience, New York, 1994.

[15] W.C.M. Kallenberg, Asymptotic Optimality of Likelihood Ratio Tests in Exponential Families, Math. Centre Tracts, vol. 77, CWI, Amsterdam, 1978.

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

[17] Ya.Yu. Nikitin, On the Hodges–Lehmann asymptotic efficiency of nonparametric tests of goodness-of-fit and homogeneity, Theory Probab. Appl. 32 (1987) 77–88.

[18] Ya.Yu. Nikitin, Asymptotic Efficiency of Nonparametric Tests, Cambridge University Press, New York, 1995.

[19] J. Oosterhoff, Combination of One-Sided Statistical Tests, Mathematical Centre, Amsterdam, 1969.

[20] J. Oosterhoff, W.R. van Zwet, The likelihood ratio test for the multinomial distribution, in: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. I, 1972, pp. 31–50.

Further reading

[21] W.C.M. Kallenberg, Intermediate efficiency, theory and examples, Ann. Statist. 11 (1983) 170–182.