QUADRATIC STATISTICS IN TESTING PROBLEMS OF LARGE DIMENSION

D.M. CHIBISOV[1]

Steklov Mathematical Institute

We consider testing a simple hypothesis about the mean vector of an N-variate normal distribution against shift alternatives in a Bayesian setting specifying a prior distribution of the mean vector under the alternative. We treat the problem asymptotically, as $N \to \infty$, and state fairly general conditions on the sequence of prior distributions under which the Bayes tests have asymptotically ellipsoidal acceptance regions.

AMS subject classifications: 62F03, 62F05, 62F15.
Keywords and phrases: Hypothesis testing, Bayes tests, multivariate normal distribution, large dimension.

1 Introduction

We consider testing a simple hypothesis about the mean vector of an N-variate normal distribution against shift alternatives in a Bayesian setup specified by a prior distribution of the mean vector under the alternative. Specifically, based on a single observation of an N-variate normal vector with identity covariance matrix we test the hypothesis that it has zero mean vector. We assume that for each N the prior distribution is the product of symmetric univariate distributions, or, in other words, under this prior the mean vector has independent symmetrically distributed components. Furthermore, we require these components to be in a certain sense asymptotically uniformly negligible. The result obtained can be viewed as an asymptotically complete class theorem saying that for this kind of alternatives in large dimension one can restrict oneself to tests with ellipsoidal acceptance regions. At the end of this section we give an example of a prior distribution for which our conditions fail.

The normal shift model of fixed dimension arises in asymptotic hypothesis testing problems about a multivariate parameter, the normal vector under consideration being the limit in distribution of a sequence of (vector-valued) asymptotically sufficient statistics, see, e.g., Roussas (1972), Chapter 6. (The general case of a known positive definite covariance matrix treated therein

[1] Research supported by the INTAS–RFBR grant 95–0099, by DFG-RFBR grant 98-01-04108, and by the RFBR Grant 96-15-96033 for Leading Scientific Scools..

reduces to the case of the identity matrix by a linear transformation.) It is customary for multivariate analysis to use the chi-squared test for this testing problem. However, the underlying property of rotational invariance may often be inadequate, and discarding it we are left with a variety of tests neither of which is intrinsically dominant. Hence we look for a reduction which could be achieved under some natural additional requirements.

It is well known that the shifts of a normal distribution form an exponential family in which case tests with convex acceptance regions constitute an essentially complete class of tests; these are Bayes tests and their weak limits, see, e.g., Roussas (1972), Appendix 4, and references therein. Our study is motivated by nonparametric goodness of fit and signal detection problems, see Ingster (1993), Spokoiny (1998), and references therein. In the former problem the normal shift model again is obtained asymptotically, for large sample size, while in the latter case, when the signal is observed in a white Gaussian noise, it is obtained directly by taking the Fourier coefficients of the observed process with respect to some orthonormal basis on the observation interval. In both these cases it appears natural to consider prior distributions rendering the components of the mean vector independent and symmetrically distributed. (Thus we think, say, of a possible signal as being composed of random and independent harmonics.) For fixed N these assumptions provide a certain reduction (in particular, the acceptance regions become symmetric in each coordinate). However, the treatment of this problem for large dimension (as $N \to \infty$) allows for a substantial reduction to the class of tests of a specific structure, viz., tests with ellipsoidal acceptance regions, provided the priors satisfy a certain uniform negligibility condition. To explain the nature of the result we state in this section a corollary to the main theorem having a more transparent form.

Thus we observe the random vector

$$\mathbf{X} = (X_1, \dots, X_N) \tag{1.1}$$

having normal distribution $N(\mu_N, I_N)$ with $\mu_N = (\mu_{N1}, \dots, \mu_{NN}) \in \mathbf{R^N}$ and I_N the $N \times N$ identity matrix. We test the hypothesis $H_{N0} : \mu_N = \mathbf{0}$ against $H_{N1} : \mu_N \neq \mathbf{0}$. In the Bayesian setup we assume that μ_N under H_{N1} has a prior distribution π_N, which is the product of N coordinate distributions,

$$\pi_N(d\mu_N) = \mathop{\times}_{i=1}^{N} \pi_{Ni}(d\mu_{Ni}), \tag{1.2}$$

so that μ_N is a random vector with independent components having distributions $\pi_{N1}, \dots, \pi_{NN}$. We assume throughout that π_{Ni}, $i = 1, \dots, N$, are symmetric about the origin.

By (1.1) for a given μ_N the distribution of $\mathbf{X}$ has Lebesgue density

$$\varphi_N(\mathbf{x} - \mu_N) = \prod_{i=1}^{N} \varphi(x_i - \mu_{Ni}), \tag{1.3}$$

where $\varphi(\cdot)$ denotes the density of the standard normal distribution. We denote this distribution by $\mathrm{P}_{N,\mu}$ and the corresponding expectation by $\mathrm{E}_{N,\mu}$. In particular, the distribution of $\mathbf{X}$ under H_{N0} has density $\varphi_N(\mathbf{x}) = \prod_{i=1}^N \varphi(x_i)$. This distribution will be denoted by $\mathrm{P}_{N,0}$ and the corresponding expectation by $\mathrm{E}_{N,0}$.

The power of a test with test function $\psi_N(\mathbf{x})$ against a particular alternative μ_N equals

$$\beta_N(\mu_N; \psi_N) = \mathrm{E}_{N,\mu}\psi_N(\mathbf{X}). \tag{1.4}$$

In the Bayesian setup, (1.3) is a conditional density of $\mathbf{X}$ given μ_N, and the marginal distribution of $\mathbf{X}$ has density

$$p_N(\mathbf{x}) = \int \varphi_N(\mathbf{x} - \mu_N)\pi_N(d\mu_N). \tag{1.5}$$

Then the power of the test ψ_N is

$$\beta_N(\pi_N; \psi_N) = \int \psi_N(\mathbf{x})p_N(\mathbf{x})d\mathbf{x} = \int \beta_N(\mu_N; \psi_N)\pi_N(d\mu_N). \tag{1.6}$$

We will refer to $\beta_N(\mu_N; \psi_N)$ given by (1.4) as the power function of the test ψ_N and to $\beta_N(\pi_N; \psi_N)$ given by (1.6) as the average power.

For a preassigned size α, the Bayes test maximizing $\beta_N(\pi_N; \psi_N)$ over size α tests ψ_N rejects H_{N0} for large values of the likelihood ratio (LR)

$$h_N(\mathbf{x}) = \frac{p_N(\mathbf{x})}{\varphi_N(\mathbf{x})}. \tag{1.7}$$

More precisely, the level α Bayes test has critical function

$$\psi_N^h(\mathbf{x}) = \begin{cases} 1, & h_N(\mathbf{x}) > c_N, \\ 0, & h_N(\mathbf{x}) < c_N, \end{cases} \tag{1.8}$$

with c_N and $\psi_N(\mathbf{x})$ on $\{\mathbf{x} : h_N(\mathbf{x}) = c_N\}$ defined so that

$$\mathrm{E}_{N,0}\psi_N(\mathbf{X}) = \int \psi_N(\mathbf{x})\varphi_N(\mathbf{x})d\mathbf{x} = \alpha.$$

The level $\alpha > 0$ will be kept fixed as $N \to \infty$.

In Theorem 2.4 we state conditions on the priors π_N under which the LR h_N is asymptotically approximated by

$$g_N(\mathbf{x}) = \exp\Big[\frac{1}{2}\sum b_{Ni}\left(x_i^2 - 1\right) - \frac{1}{4}B_N\Big], \tag{1.9}$$

where $b_{Ni} \geq 0$ are certain characteristics of π_{Ni} and $B_N = \sum b_{Ni}^2$ (which is assumed to be bounded as $N \to \infty$). Namely, g_N approximates h_N in L_1-norm w.r.t. $\mathrm{P}_{N,0}$, i.e.,

$$\mathrm{E}_{N,0}|h_N - g_N| \to 0 \qquad \text{as} \quad N \to \infty. \tag{1.10}$$

It follows from (1.10) that the test $\psi_N^g(\mathbf{x})$ defined for g_N similarly to (1.8) has asymptotically the same average power as the Bayes test $\psi_N^h(\mathbf{x})$, i.e.,

$$\beta_N(\pi_N; \psi_N^g) - \beta_N(\pi_N; \psi_N^h) \to 0 \qquad \text{as} \quad N \to \infty.$$

To illustrate Theorem 2.4 we state here a special case. Suppose that the distributions π_{Ni} in (1.2) are scale transforms of one and the same distribution π on $\mathbf{R}$ with scale factors $b_{N1}, \dots, b_{NN}$, i.e.,

$$\Pi_{Ni}(\mu) = \Pi(\mu/b_{Ni}), \qquad i = 1, \dots, N,$$

where $\Pi_{Ni}(\mu)$ and $\Pi(\mu)$, $\mu \in \mathbf{R}$, denote the distribution functions corresponding to π_{Ni} and π. Let π and $\{b_{Ni}\}$ satisfy the following conditions:

(Π1) π is symmetric, i.e., $\Pi(\mu) = 1 - \Pi(-\mu)$, $\mu \in \mathbf{R}$;
(Π2) $\int \mu^2 \pi(d\mu) = 1$, $\int \mu^4 \pi(d\mu) < \infty$;
(B1) $b_{Ni} \geq 0$, $b_{N,\max} := \max_{1 \leq i \leq N} b_{Ni} \to 0$ as $N \to \infty$;
(B2) $\sum_{i=1}^N b_{Ni}^4 \to B > 0$ as $N \to \infty$.

Note that the first condition in (Π2) is merely a normalization of π, under which b_{Ni}^2 is the variance of μ_{Ni}.

Corollary 1.1 *Let Conditions* (Π1), (Π2), (B1), (B2) *be fulfilled and let* g_N *be defined by* (1.9) *with* $B_N = B$. *Then* (1.10) *holds.*

Consider the particular case where $b_{N1} = \ldots = b_{NN}$. Obviously, (B1), (B2) are satisfied for $b_{Ni} = (B/N)^{1/4}$, $i = 1, \dots, N$. Then Corollary 1.1 says that, under the independence assumption on the components of μ_N, Conditions (Π1) and (Π2) are sufficient for the Bayes test to be asymptotically chi-squared. It is well known that for any spherically symmetric prior distribution the Bayes test is exactly chi-squared. Under the independence assumption spherical symmetry holds only for π normal. Corollary 1.1 says, however, that Bayes tests become approximately chi-squared for large dimension under arbitrary symmetric π unless π is heavy-tailed (the second condition in (Π2)).

Note that in the setup of Corollary 1.1, when the prior distribution has independent symmetric components differing only by scale factors, (Π2), (B1), and (B2) are exactly conditions for asymptotic normality of

$$\sum \left(\mu_{Ni}^2 - b_{Ni}^2\right).$$

The same is true in the general case (see Remark 2.2).

In the literature Bayes tests in the normal shift model of increasing dimension are used in asymptotically minimax nonparametric hypothesis testing, see Ingster (1993), (1997), and Spokoiny (1998), where further references can be found. In these studies the original problem of signal detection or goodness of fit reduces by a suitable orthogonal decomposition to a testing problem in the normal shift model (possibly, infinite-dimensional). Typically in the minimax setting this is the problem of testing for zero mean against the set of alternatives specified by a "big" ball or ellipsoid in a certain norm, say, l_q-norm, with a "small" ball or ellipsoid in, say, l_p-norm around the origin removed. The problem is treated asymptotically as the size of these domains varies and/or the common variance of the X_i's tends to zero. For some particular prior distributions used in those papers the asymptotically ellipsoidal form of the Bayes tests was established directly. For example, Ingster (1993) uses "Bernoulli priors" specified by symmetric two-point prior distibutions of components. These distibutions obviously satisfy conditions (Π1) and (Π2).

The choice of the prior distribution depends on the shape of the parameter set, specifically, on the degrees p and q of the norms. If the normal shift model originates from, say, a signal detection problem, these degrees are related, qualitatively, to smoothness properties of the least favorable signals and restrictions on their "energy". In this respect Spokoiny (1998) distinguishes four types of alternative sets. Apparently the type of alternatives treated here fits in one of those classes, viz., that of "smooth" signals. Another type of prior distributions used by Ingster (1993) and Spokoiny (1998) for other types of alternatives has three-point component distributions π_{Ni} with masses p_N at points ± 1 (up to scale factors) and mass $1-2p_N$ at 0 with $p_N \to 0$ as $N \to \infty$. Note that the ratio of the fourth moment to the squared variance equals here $1/p_N \to \infty$. For this prior distribution the conditions and the conclusion of Theorem 2.4 fail.

We state the main Theorem 2.4 in Section 2 and give its proof in Section 3. Section 4 contains the proofs of auxiliary results and Corollary 1.1.

2 Main Theorem

Recall that we consider testing the hypothesis $H_0 : \mu_N = \mathbf{0}$ based on the observed N-variate random vector $\mathbf{X} = (X_1, \ldots, X_N)$ with normal distribution $N(\mu_N, I_N)$. Under the alternative μ_N has prior distribution (cf. (1.2))

$$\pi_N(d\mu_N) = \mathop{\times}_{i=1}^{N} \pi_{Ni}(d\mu_{Ni}).$$

Thus under this prior $\{\mu_{Ni}\}$ form a triangular array of r.v.'s independent within each row (for each N) with corresponding distributions π_{Ni}, $i = 1, \ldots, N$.

Assumption (A1). The distributions π_{Ni}, $i = 1, \ldots, N$, $N \in \mathbf{N} = \{1, 2, \ldots\}$, are symmetric, i.e., $\pi_{Ni}(A) = \pi_{Ni}(-A)$ for any Borel set A.

In terms of the corresponding distribution functions this assumption means that $\Pi_{Ni}(\mu) = 1 - \Pi_{Ni}(-\mu)$, $\mu \in \mathbf{R}$ (cf. (Π1) in Section 1).

For $a > 0$, denote

$$\gamma_{Ni}(a) = 1 - \pi_{Ni}([-a, a]) = 2\pi_{Ni}((a, \infty)). \tag{2.1}$$

Assumption (A2). For any $a > 0$,

$$\sum_{i=1}^{N} \gamma_{Ni}(a) \to 0.$$

For a measure Q and a measurable function f (on the same space) we will write

$$Q(f) = \int f(x)Q(dx). \tag{2.2}$$

For $a > 0$, denote by $\pi_{Ni}^{(a)}$ the measure π_{Ni} restricted to the interval $[-a, a]$,

$$\pi_{Ni}^{(a)}(A) = \pi_{Ni}(A \cap [-a, a]). \tag{2.3}$$

Define the corresponding truncated moments as

$$\nu_{k,N,i}(a) = \pi_{Ni}^{(a)}(\mu_{Ni}^{k}), \quad k = 0, 1, 2, \ldots \tag{2.4}$$

Note that due to symmetry of $\pi_{Ni}^{(a)}$ (see (A1) and (2.3)), $\nu_{k,N,i}(a) = 0$ for odd k. Obviously, $\nu_{k,N,i}(a)$ for any even k is a nondecreasing function of a.

Lemma 2.1 *Under Assumptions* (A1), (A2),

$$\sum_{i=1}^{N} \nu_{k,N,i}(a_2) - \sum_{i=1}^{N} \nu_{k,N,i}(a_1) \to 0 \tag{2.5}$$

for any fixed $a_1, a_2 > 0$ and any even $k > 0$.

Proof For $0 < a_1 < a_2$ the left-hand side of (2.5) is nonnegative and bounded by $a_2^k \sum \gamma_{Ni}(a_1)$, which tends to zero by (A2). ∎

Assumption (A3). For any $a > 0$,

$$\limsup_{N\to\infty} \sum_{i=1}^{N} \nu_{4,N,i}(a) < \infty.$$

By Lemma 2.1 the requirement "for any $a > 0$" can be equivalently reduced to the requirement "for some $a > 0$".

Since $\nu_{2,N,i}^2(a) \leq \nu_{4,N,i}(a)$, Assumption (A.3) implies

Corollary 2.2 *Under Assumption* (A3),

$$\limsup_{N\to\infty} B_N(a) < \infty, \tag{2.6}$$

for any $a > 0$, *where*

$$B_N(a) = \sum_{i=1}^{N} \nu_{2,N,i}^2(a). \tag{2.7}$$

Lemma 2.3 *Under Assumptions* (A1)–(A3), *for any fixed* $a_1, a_2 > 0$

$$\Delta_N := B_N(a_2) - B_N(a_1) \to 0 \qquad \text{as} \quad N \to \infty.$$

The proof of this lemma will be given in Section 4.

Theorem 2.4 *Under Assumptions* (A1)–(A3)

$$\mathrm{E}|h_N(\mathbf{X}) - g_N(\mathbf{X}; a)| \to 0 \quad \text{as} \quad N \to \infty \tag{2.8}$$

for any $a > 0$, *where* (*see* (2.4), (2.7))

$$g_n(\mathbf{x}, a) = \exp\Big(\frac{1}{2}\sum_{i=1}^{N} \nu_{2,N,i}(a)(x_i^2 - 1) - \frac{1}{4}B_N(a)\Big). \tag{2.9}$$

Remark 2.1 The relation (2.8) implies, in particular, that the functions $g_N(\cdot, a)$ for different choices of a approach each other in L_1 norm. This can also be verified directly by using Lemmas 2.1 and 2.3.

Remark 2.2 Assumptions (A1)–(A3) imply asymptotic normality of the sequence $\sum_i \mu_{Ni}^2$ with mean $B_N(a)$ and variance $\sum \big(\nu_{4,N,i}(a) - \nu_{2,N,i}^2(a)\big)$ for any $a > 0$, see Loève (1960), Section 22.5. In this respect Corollary 1.1 relates to Theorem 2.4 in the same way as Theorem V.1.2 in Hájek and Šidák (1967) to the general normal convergence theorem in Loève (1960) mentioned above.

3 Proof of Theorem 2.4

Take an $a > 0$. Without loss of generality we will assume that there exists the limit $B(a) := \lim_{N\to\infty} B_N(a)$. (Otherwise assume that (2.8) fails, select a subsequence where the left-hand side of (2.8) stays bounded away from zero and find by (2.6) a further subsequence where $B_N(a)$ converges.) The proof relies on the following one-sided version of Scheffé's Lemma (see Chibisov (1992), Lemma 3.1).

Lemma 3.1 *Let for each $N \in \mathbf{N}$ the random variables $U_N \geq 0$ and $V_N \geq 0$ be defined on a probability space $(\mathcal{X}_N, \mathcal{A}_N, \mathrm{P}_N)$. Assume:* (i) $\mathrm{E}_N U_N \to 1$, $\mathrm{E}_N V_N \to 1$; (ii) *V_N are uniformly integrable w.r.t. P_N, or, equivalently,*

$$\mathrm{E}_N[V_N; A_N] := \int_{A_N} V_N \, d\mathrm{P}_N \to 0 \quad \textit{whenever} \quad \mathrm{P}_N(A_N) \to 0;$$

(iii) $\mathrm{P}_N(U_N < V_N - \varepsilon) \to 0$ *for any $\varepsilon > 0$. Then*

$$\mathrm{E}_N|U_N - V_N| \to 0.$$

We will apply this lemma with $\mathrm{P}_N := \mathrm{P}_{N,0} = N(\mathbf{0}, I_N)$, $U_N := h_N$, and $V_N := g_N(\cdot, a)$. Condition (i) for h_N holds by definition (see (1.5), (1.7)), since $\mathrm{E}_{N,0} h_N = 1$. The following lemma will be used to verify Condition (i) for g_N.

Lemma 3.2 *For any even $k > 0$ and any $a > 0$*

$$\max_{1 \leq i \leq N} \nu_{k,N,i}(a) \to 0.$$

Proof For an arbitrary $\varepsilon > 0$,

$$\limsup_{N \to \infty} \max_{1 \leq i \leq N} \nu_{k,N,i}(a) \leq \varepsilon^k + \limsup_{N \to \infty} \sum_i [\nu_{k,N,i}(a) - \nu_{k,N,i}(\varepsilon)].$$

By Lemma 2.1, the latter term equals 0. Hence the lemma follows. ∎

To check Condition (i) for $g_N = g_N(\cdot, a)$, we use the formula: for a r.v. X with standard normal distribution and any $b < 1$,

$$\mathrm{E} \exp(\frac{1}{2} b X^2) = (1 - b)^{-1/2}.$$

When applied to (2.9), this yields (with dependence on a suppressed)

$$\mathrm{E}_{N,0} g_N = \prod_1^N (1 - \nu_{2,N,i})^{-1/2} \exp(-\frac{1}{2} \nu_{2,N,i} - \frac{1}{4} \nu_{2,N,i}^2).$$

By Taylor we have with some $0 \leq \theta_{Ni} \leq 1$

$$\log \mathrm{E}_{N,0} g_N = \frac{1}{6} \sum_{i=1}^N \frac{\nu_{2,N,i}^3}{(1 - \theta_{Ni} \nu_{2,N,i})^3} \leq \frac{\max_i \nu_{2N,i}}{6(1 - \max_i \nu_{2,N,i})^3} B_N,$$

which tends to zero by (2.6) and Lemma 3.2.

To verify condition (ii), assume the contrary. Then there exist $\varepsilon > 0$, a subsequence $\{N'\} \subset \{N\}$, and sets $A_{N'} \subset \mathbf{R}^{N'}$ with $\mathrm{P}_{N',0}(A_{N'}) \to 0$ such that

$$\mathrm{E}_{N',0}[g_{N'}; A_{N'}] > \varepsilon \qquad \text{for all} \quad N'. \tag{3.1}$$

Recall that we assume the limit $B(a) = \lim_{N'\to\infty} B_{N'}(a)$ to exist. Let $B(a) > 0$. Then

$$(3.2) \qquad (2B(a))^{-1/2} \sum_{i=1}^{N} \nu_{2,N,i}(a)(X_i^2 - 1) \to_d N(0,1)$$

because Lemma 3.2 implies the Lindeberg condition (see Theorem V.1.2 in Hájek and Šidák (1967)). Hence $g_{N'}(\,\cdot\,; a)$ converges in distribution to $g = \exp[\frac{1}{2}Y - \frac{1}{4}B(a)]$, where $Y \sim N(0, 2B(a))$. If $B(a) = 0$, one checks directly that $g_{N'}(\,\cdot\,; a)$ converges in distribution to $g \equiv 1$. In both cases $Eg = 1$. Therefore $g_{N'}$ are uniformly integrable (see Loève (1960), 9.4.e and 11.4.A), which contradicts (3.1).

Thus it remains to prove that

$$(3.3) \qquad \mathrm{P}_{N,0}(h_N < g_N(a) - \varepsilon) \to 0 \quad \text{for any} \quad \varepsilon > 0.$$

If $B(a) = 0$, one can check directly that both h_N and $g_N(a)$ converge to 1 in probability, which implies (3.3). So, assuming $B(a) > 0$ we will prove that

$$(3.4) \qquad \mathrm{P}_{N,0}(\log h_N < \log g_N(a) - \varepsilon) \to 0 \quad \text{for any} \quad \varepsilon > 0.$$

It is readily shown that (3.4) implies (3.3). Indeed, the inequality in (3.4) entails an inequality as in (3.3) unless g_N takes large values, which occurs with a small probability by an argument similar to the one used when checking condition (ii). Thus having shown (3.4) we will have established the conditions of Lemma 3.1, which then implies the theorem.

Now we proceed to the proof of (3.4). By (1.2), (1.3), (1.5), and (1.7),

$$h_N(\mathbf{x}) = \prod_1^N \pi_{Ni}[\exp(x_i\mu_{Ni} - \frac{1}{2}\mu_{Ni}^2)],$$

where $\pi_{Ni}[\ldots]$ means the integral w.r.t. μ_{Ni} as in (2.2). Obviously,

$$(3.5) \qquad h_N(\mathbf{x}) \geq \prod_1^N h_{Ni}(x_i, a),$$

where

$$(3.6) \qquad h_{Ni}(x, a) = \pi_{Ni}^{(a)}[\exp(x\mu_{Ni} - \frac{1}{2}\mu_{Ni}^2)]$$

(see (2.3)).

Now we use the fact that for odd $m \in \mathbf{N}$

$$e^x \geq 1 + \sum_{k=1}^{m} \frac{x^k}{k!} \quad \text{for any} \quad x \in \mathbf{R}.$$

Applying this inequality with $m = 5$ to (3.5) we obtain

$$h_{Ni}(x,a) \geq 1 + \xi_{Ni}(x,a), \tag{3.7}$$

where, using the notation (2.1) and the abbreviation $\nu_k = \nu_{k,N,i}(a)$,

$$\begin{aligned}\xi_{Ni}(x,a) = & -\gamma_{Ni}(a) - \frac{1}{2}\nu_2 + \frac{1}{2}\Big(x^2\nu_2 + \frac{1}{4}\nu_4\Big) - \frac{1}{6}\Big(\frac{3}{2}x^2\nu_4 + \frac{1}{8}\nu_6\Big) \\ & + \frac{1}{24}\Big(x^4\nu_4 + \frac{3}{2}x^2\nu_6 + \frac{1}{16}\nu_8\Big) - \frac{1}{5!}\Big(\frac{5}{2}x^4\nu_6 + \frac{5}{4}x^2\nu_8 + \frac{1}{32}\nu_{10}\Big).\end{aligned} \tag{3.8}$$

To complete the proof, we need the following two lemmas. Their proofs will be given in Section 4.

Lemma 3.3 *Under Assumptions* (A1)–(A3), *for any* $\delta > 0$,

$$\mathrm{P}_{N,0}\big(\min_i \xi_{Ni}(X_i,a) < -\delta\big) \to 0. \tag{3.9}$$

Lemma 3.4 *Under Assumptions* (A1)–(A3)

$$\sum_{i=1}^{N} \xi_{Ni}(X_i,a) - \frac{1}{2}\sum_{i=1}^{N} \nu_{2,N,i}(a)(X_i^2 - 1) \to_{\mathrm{P}_{N,0}} 0. \tag{3.10}$$

By Taylor, for any $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that

$$\log(1+x) \geq x - \frac{1}{2}(1+\varepsilon)x^2 \quad \text{for} \quad x \geq -\delta.$$

Hence by (3.5), (3.6), and (3.7), Lemma 3.3 implies that

$$\mathrm{P}_{N,0}\big(\log h_N(\mathbf{X}) \geq f_N(\mathbf{X})\big) \to 1, \tag{3.11}$$

where

$$f_N(\mathbf{x}) = \sum_{i=1}^{N} \xi_{Ni}(x_i) - \frac{1}{2}(1+\varepsilon)\sum_{i=1}^{N} \xi_{Ni}^2(x_i). \tag{3.12}$$

Lemma 3.4 and (3.2) imply that $\big(2/B_N(a)\big)^{1/2}\sum \xi_{Ni}(X_i)$ is asymptotically normal $N(0,1)$. Therefore

$$\frac{2}{B_N(a)}\sum_{i=1}^{N} \xi_{Ni}^2(X_i) \to_{\mathrm{P}_{N,0}} 1$$

(see Gnedenko and Kolmogorov (1949), Section 28, Theorem 4). Hence comparing (2.9) and (3.12) we obtain by Lemma 3.4

$$f_N(\mathbf{X}) - \log g_N(\mathbf{X},a) + \frac{\varepsilon}{4}B_N(a) \to_{\mathrm{P}_{N,0}} 0. \tag{3.13}$$

Since $\{B_N(a)\}$ is bounded, (3.11) and (3.13) imply (3.4) and hence the theorem.

4 Proofs of auxiliary results

4.1 Proof of Lemma 2.3

Assume $0 < a_1 < a_2$. Then, obviously, $\Delta_N \geq 0$. By (2.7),

$$\begin{aligned}(4.1)\qquad \Delta_N &= \sum_{i=1}^{N}(\nu_{2,N,i}^2(a_2) - \nu_{2,N,i}^2(a_1)) \\ &= \sum_{i=1}^{N}(\nu_{2,N,i}(a_2) - \nu_{2,N,i}(a_1))(\nu_{2,N,i}(a_2) + \nu_{2,N,i}(a_1)).\end{aligned}$$

For each $i = 1, \ldots, N$, when the inequality

$$\nu_{2,N,i}(a_2) - \nu_{2,N,i}(a_1) \leq \varepsilon\nu_{2,N,i}(a_2)$$

holds, we have

$$(4.2)\qquad \nu_{2,N,i}^2(a_2) - \nu_{2,N,i}^2(a_1) \leq 2\varepsilon\nu_{2,N,i}^2(a_2).$$

Otherwise,

$$\nu_{2,N,i}(a_2) + \nu_{2,N,i}(a_1) \leq 2\nu_{2,N,i}(a_2) \leq 2\varepsilon^{-1}(\nu_{2,N,i}(a_2) - \nu_{2,N,i}(a_1)).$$

Hence in this latter case

$$\begin{aligned}(4.3)\qquad \nu_{2,N,i}^2(a_2) - \nu_{2,N,i}^2(a_1) &\leq 2\varepsilon^{-1}(\nu_{2,N,i}(a_2) - \nu_{2,N,i}(a_1))^2 \\ &\leq 2\varepsilon^{-1}(\nu_{4,N,i}(a_2) - \nu_{4,N,i}(a_1)).\end{aligned}$$

Therefore (4.1), (4.2), and (4.3) show that

$$(4.4)\qquad \Delta_N \leq 2\varepsilon\sum_{i=1}^{N}\nu_{2,N,i}^2(a_2) + 2\varepsilon^{-1}\sum_{i=1}^{N}[\nu_{4,N,i}(a_2) - \nu_{4,N,i}(a_1)].$$

The second term in (4.4) tends to 0 as $N \to \infty$ by Lemma 2.1, while the sum in the first term is bounded by Corollary 2.2. Hence $\limsup_N \Delta_N$ can be made arbitrarily small by the choice of ε.

4.2 Proof of Lemma 3.3

Rewrite (3.8) as

$$(4.5)\qquad \xi_{Ni}(a) = -\gamma_{Ni}(a) + \eta_{Ni}^{(1)}(a) + \eta_{Ni}^{(2)}(a) + \eta_{Ni}^{(3)}(a),$$

where $\eta_{Ni}^{(j)}(a) = \eta_{Ni}^{(j)}(X_i, a)$ with

$$(4.6)\qquad \eta_{Ni}^{(1)}(x, a) = \frac{1}{2}(x^2 - 1)\nu_{2,N,i}(a),$$

$$(4.7)\qquad \eta_{Ni}^{(2)}(x, a) = \frac{1}{24}(x^4 - 6x^2 + 3)\nu_{4,N,i}(a),$$

$$(4.8)\qquad \eta_{Ni}^{(3)}(x, a) = \sum c_{jk}x^j\nu_{k,N,i}(a).$$

The sum in (4.8) contains a finite number of terms (actually, six) with even $j \geq 0$ and $k \geq 6$.

For the proof of Lemma 3.3 it suffices to establish the corresponding assertions for each term in the RHS of (4.5). The ones for γ_{Ni}, $\eta_{Ni}^{(1)}$, and $\eta_{Ni}^{(2)}$ follow from Assumption (A2) and Lemma 3.2 (notice that the polynomials in (4.6) and (4.7) are bounded from below). The counterpart of (3.9) for $\eta_{Ni}^{(3)}$ is obtained from the following two lemmas.

Lemma 4.1 *For any $a > 0$ and any even $k > 4$*

$$\sum_{i=1}^{N} \nu_{k,N,i}(a) \to 0. \tag{4.9}$$

Proof By Lemma 2.1, for an arbitrary $\varepsilon > 0$,

$$\begin{aligned}\limsup_{N\to\infty} \sum_{i=1}^{N} \nu_{k,N,i}(a) &= \limsup_{N\to\infty} \sum_{i=1}^{N} \nu_{k,N,i}(\varepsilon) \\ &\leq \varepsilon^{k-4} \limsup_{N\to\infty} \sum_{i=1}^{N} \nu_{4,N,i}(\varepsilon) = C\varepsilon^{k-4}\end{aligned}$$

with $C < \infty$ by Assumption (A3). Hence (4.9) follows. ∎

Lemma 4.2 *Let $Y_1, Y_2, \dots$ be i.i.d. r.v.'s with $\mathrm{E}|Y_1| < \infty$, and let $\{c_{Ni}, i = 1, \dots, N\}$, $N \in \mathbf{N}$, be a triangular array of nonnegative numbers such that*

$$\sum_{i=1}^{N} c_{Ni} \to 0. \tag{4.10}$$

Then $\max_{1 \leq i \leq N} c_{Ni}|Y_i| \to_P 0$.

Proof For an arbitrary $\varepsilon > 0$ we have, using the Markov inequality,

$$\mathrm{P}(\max c_{Ni}|Y_i| > \varepsilon) \leq \sum_{i=1}^{N} \mathrm{P}\Big(|Y_i| > \frac{\varepsilon}{c_{Ni}}\Big) \leq \sum_{i=1}^{N} \frac{c_{Ni}}{\varepsilon} \mathrm{E}|Y_i| \to 0,$$

which proves the lemma. ∎

Now the counterpart of (3.9) for each term of $\eta_{Ni}^{(3)}(X_i, a)$ (see (4.8)) follows by Lemma 4.2, with condition (4.10) for $c_{Ni} := \nu_{k,N,i}(a)$ fulfilled by Lemma 4.1.

4.3 Proof of Lemma 3.4

Comparing (3.10) with (4.5) and taking into account Assumption (A2), we see that it remains to show

$$\Sigma_2 := \sum_{i=1}^{N} \eta_{Ni}^{(2)}(a) \to_{\mathrm{P}_{N,0}} 0 \tag{4.11}$$

and

$$\Sigma_3 := \sum_{i=1}^{N} \eta_{Ni}^{(3)}(a) \to_{\mathrm{P}_{N,0}} 0. \tag{4.12}$$

It is directly verified that $\mathrm{E}(X_1^4 - 6X_1^2 + 3) = 0$, so that $\mathrm{E}\Sigma_2 = 0$; further,

$$\operatorname{var}\Sigma_2 = \mathrm{const} \sum \nu_{4,N,i}^2(a) \le \mathrm{const} \cdot \max_i \nu_{4,N,i}(a) \cdot \sum \nu_{4,N,i}(a) \to 0$$

by Assumption (A3) and Lemma 3.2. This implies equation (4.11). Next, by Lemma 4.1, $\mathrm{E}|\Sigma_3| \to 0$, which proves (4.12).

4.4 Proof of Corollary 1.1

We have (a) to check that (Π1), (Π2), (B1), and (B2) imply Assumptions (A1)–(A3) and (b) to show that the truncated moments $\nu_{2,N,i}(a)$ and the quantity $B_N(a)$ can be asymptotically replaced by b_{Ni}^2 and B respectively.

Assumption (A1) obviously follows from (Π1). The 4th moment assumption in (Π2) implies

$$\gamma_{Ni}(a) = \pi(b_{Ni}|\mu_{Ni}| > a) \le \frac{b_{Ni}^4}{a^4} \int_{|\mu| > a/b_{Ni}} |\mu|^4 \pi(d\mu). \tag{4.13}$$

The last integral tends to zero uniformly in $1 \le i \le N$ by (B1) and (Π2), so (A2) follows from (B2).

To check Assumption (A3), note that

$$\nu_{4,N,i}(a) = b_{Ni}^4 \int_{|\mu| \le b_{Ni} a} \mu^4 d\pi.$$

Hence (A3) follows from (B2) and (Π2).

For (b) we have to show that for any $a > 0$

$$\sum_{i=1}^{N} (\nu_{2,N,i}(a) - b_{Ni}^2)(X_i^2 - 1) \to_{\mathrm{P}_{N,0}} 0 \tag{4.14}$$

and

$$B_N(a) \to B. \tag{4.15}$$

We establish (4.14) by showing that the 2nd moment of the LHS tends to 0, which amounts to

$$\sum_{i=1}^{N}\left(\nu_{2,N,i}(a) - b_{Ni}^2\right)^2 \to 0. \tag{4.16}$$

Observe that

$$\nu_{2,N,i}(a) = b_{Ni}^2 \int_{|\mu|\le a/b_{Ni}} \mu^2 d\pi,$$

hence

$$\left(b_{Ni}^2 - \nu_{2,N,i}(a)\right)^2 = b_{Ni}^4\Big(\int_{|\mu|>a/b_{Ni}} \mu^2 d\pi\Big)^2 \le b_{Ni}^4 \int_{|\mu|>a/b_{Ni}} |\mu|^4 \pi(d\mu).$$

Thus (4.16) is obtained by (B2) and the argument following (4.13).

Now (4.15) follows from (4.16) by the triangle inequality.

It remains to show that under the assumptions of Corollary 1.1 $g_N(\mathbf{x}, a)$ given by (2.16) is approximated in L_1-norm by $g_N(\mathbf{x})$ given by (1.9) with $B_N = B$, i.e.,

$$\mathrm{E}|g_N(\cdot, a) - g_N(\cdot)| \to 0.$$

This follows from Lemma 3.1. Conditions (i) and (ii) of this lemma for $g_N(\cdot, a)$ were established in the proof of Theorem 2.4, condition (i) for g_N is verified in a similar manner, and the two-sided version of condition (iii) follows from (4.14) and (4.15) since they imply that

$$g_N(\mathbf{X}, a) - g_N(\mathbf{X}) \to_{\mathrm{P}_{N,0}} 0.$$

Acknowledgements The author thanks the referees and editors for suggestions that helped to improve the presentation.

REFERENCES

Chibisov, D.M. (1992). Asymptotic optimality of the chi-square test with large number of degrees of freedom within the class of symmetric tests. *Mathematical Methods of Statistics* 1, 55–82.

Gnedenko, B.V. and Kolmogorov, A.N. (1949). *Limit Distributions for Sums of Independent Random Variables.* Gostekhizdat, Moscow–Leningrad (Russian). English translation: Addison–Wesley, Reading, MA, 1954.

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests.* Academia, Prague.

Ingster, Yu.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Mathematical Methods of Statistics* **2**, 85–114, 171–189, 249–268.

Ingster, Yu.I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Mathematical Methods of Statistics* **6**, 47–69.

Loève, M. (1960). *Probability Theory*, 2nd edition. Van Nostrand, Princeton, NJ.

Spokoiny, V.G. (1998). Adaptive and spatially adaptive testing of a nonparametric hypothesis. *Mathematical Methods of Statistics* **7**, 245–273.

Steklov Mathematical Institute
Gubkina str. 8
GSP-1 117966 Moscow
Russia
chibisov@mi.ras.ru