Chi-squared approximation for the power divergence family of statistics

Zubov V. N. Ulyanov V. V.

Abstract

In this paper we consider the convergence of the *power divergence* family of statistics $\{T_\lambda(\boldsymbol{Y}), \lambda \in \mathbb{R}\}$ constructed from the multinomial distribution of degree k, to chi-squared distribution with $k-1$ degrees of freedom. We show that

$$\Pr(T_\lambda(\boldsymbol{Y}) < c) = G_{k-1}(c) + O(n^{-1+\frac{1}{k}})$$

where $G_r(c)$ is the cumulative distribution function of a chi-squared variable with r degrees of freedom. In the proof we utilize known number theory results on the approximation of the number of integer points in a given set by its volume. Namely, E. Hlawka's theorem (1950) about the number of above-mentioned points in a convex set with a closed smooth boundary.

Key words: *E. Hlawka's theorem, weak convergence, Gaussian curvature, manifold, approximation by chi-squared distribution, power divergence family of statistics .*

Contents

1 Introduction and the main result **3**
1.1 Introduction 3
1.2 The main result 6

2 Reduction of the term J_2 **8**
2.1 Some ancillary facts from differential geometry 8
2.2 Preliminary lemmas 9
2.3 The transformation of J_2 representation into a simplified form 18

3 Applicability of Hlawka's theorem to the sequence $B^\lambda(n)$ **21**
3.1 Convexity of B^λ 21
3.2 Sufficient conditions for the applicabilityof Hlawka's theorem . 23
3.3 Fulfillment of sufficient conditions for the sets $B^\lambda(n)$ 25

4 Summarizing the point **41**

A Proof of root convergence estimate **42**

B Obtaining a lower estimate for the functional $T_\lambda(\boldsymbol{x})$ **43**

C Acknowledgement **48**

1 Introduction and the main result

1.1 Introduction

Consider a vector $\boldsymbol{Y} = (Y_1, \ldots, Y_k)^T$ with multinomial distribution $M_k(n, \pi)$, i. e.

$$\Pr(Y_1 = n_1, \ldots, Y_k = n_k) = \begin{cases} n! \prod_{j=1}^{k} (\pi_j^{n_j}/n_j!), & n_j = 0, 1, \ldots, n \ (j = 1, \ldots, k) \\ & \text{and} \sum_{j=1}^{k} n_j = n, \\ 0, & \text{otherwise,} \end{cases}$$

where $\pi = (\pi_1, \ldots, \pi_k)^T, \pi_j > 0, \sum_{j=1}^{k} \pi_j = 1$. From this point on, we will assume the validity of the hypothesis $H_0 \colon \pi = \boldsymbol{p}$. The covariance matrix of the vector $\boldsymbol{Y}$ is known to equal $\Omega = (\delta_i^j \, p_i - p_i p_j) \in \mathbb{R}^{(k-1)\times(k-1)}$. The main object of the current study is the power divergence family of statistics:

$$t_\lambda(\boldsymbol{Y}) = \frac{2}{\lambda(\lambda+1)} \sum_{j=1}^{k} Y_j \left[\left(\frac{Y_j}{np_j} \right)^\lambda - 1 \right], \ \lambda \in \mathbb{R},$$

R e m a r k 1. When $\lambda = 0, -1$, this notation should be understood as a result of passage to the limit.

R e m a r k 2. These statistics were first introduced in [9] and [10] being denoted by $2nI^\lambda(\boldsymbol{Y})$. Putting $\lambda = 1, \lambda = -1/2$ and $\lambda = 0$ we can obtain the chi-squared statistic, the Freeman-Tukey statistic, and the log-likelihood ratio statistic respectively.

We consider transformation

$$X_j = (Y_j - np_j)/\sqrt{n}, \ j = 1, \ldots, k, \ r = k-1, \ \boldsymbol{X} = (X_1, \ldots, X_r)^T.$$

Herein the vector $\boldsymbol{X}$ is the vector whose components are reduced to the lattice,

$$L = \{\boldsymbol{x} = (x_1, \ldots, x_r)^T; \, \boldsymbol{x} = \frac{\boldsymbol{m} - n\boldsymbol{p}}{\sqrt{n}}, \, \boldsymbol{p} = (p_1, \ldots, p_r)^T, \, \boldsymbol{m} = (n_1, \ldots, n_r)^T\},$$

where n_j are non-negative integers.

R e m a r k 3. The statistic $t_\lambda(\boldsymbol{Y})$ can be expressed as a function of $\boldsymbol{X}$ in the form

$$T_\lambda(\boldsymbol{x}) = \frac{2n}{\lambda(\lambda+1)} \left[\sum_{j=1}^{k} p_j \left(\left(1 + \frac{x_j}{\sqrt{n} p_j} \right)^{\lambda+1} - 1 \right) \right], \tag{1}$$

and then, via the Taylor's expansion, transformed to the form

$$T_\lambda(\boldsymbol{x}) = \sum_{i=1}^{k} \left(\frac{x_i^2}{p_i} + \frac{(\lambda-1)x_i^3}{3\sqrt{n}p_i^2} + \frac{(\lambda-1)(\lambda-2)x_i^4}{12p_i^3 n} + O\left(n^{-3/2}\right) \right). \tag{2}$$

We call a set $B \subset \mathbb{R}^r$ *extended convex set*, if for for all $l = \overline{1, r}$ it can be expressed in the form:

$$B = \{\boldsymbol{x} = (x_1, \dots, x_r)^T \colon \lambda_l(x^*) < x_l < \theta_l(x^*) \text{ and} \\ x^* = (x_1, \dots, x_{l-1}, x_{l+1}, \dots, x_r)^T \in B_l\},$$

where B_l is some subset of $\mathbb{R}^{r-1}$ and $\lambda_l(x^*), \theta_l(x^*)$ are continuous functions on $\mathbb{R}^{r-1}$. Additionally, we introduce the following notation

$$[h(\boldsymbol{x})]_{\lambda_l(x^*)}^{\theta_l(x^*)} = h(x_1, \dots, x_{l-1}, \theta_l(x^*), x_{l+1}, \dots, x_r) \\ - h(x_1, \dots, x_{l-1}, \lambda_l(x^*), x_{l+1}, \dots, x_r).$$

It is a known fact that the distributions of all statistics in the family converge to chi-squared distribution with $k-1$ degrees of freedom (see e.g. [9], p. 443). However, more intriguing is the problem of the estimation of the rate of convergence to the limiting distribution.

For any bounded extended convex set B J. Yarnold in [1] obtained an asymptotic expansion, which in [6] was converted to

$$\Pr(\boldsymbol{X} \in B) = J_1 + J_2 + O(n^{-1}). \tag{3}$$

with

$$J_1 = \int \cdots \int_B \phi(\boldsymbol{x}) \left\{ 1 + \frac{1}{\sqrt{n}} h_1(\boldsymbol{x}) + \frac{1}{n} h_2(\boldsymbol{x}) \right\} dx, \text{ where}$$

$$h_1(\boldsymbol{x}) = -\frac{1}{2} \sum_{j=1}^{k} \frac{x_j}{p_j} + \frac{1}{6} \sum_{j=1}^{k} x_j \left(\frac{x_j}{p_j} \right)^2,$$

$$h_2(\boldsymbol{x}) = \frac{1}{2} h_1(\boldsymbol{x})^2 + \frac{1}{12} \left(1 - \sum_{j=1}^{k} \frac{1}{p_j} \right) + \frac{1}{4} \sum_{j=1}^{k} \left(\frac{x_j}{p_j} \right)^2 - \frac{1}{12} \sum_{j=1}^{k} x_j \left(\frac{x_j}{p_j} \right)^3;$$

$$J_2 = -\frac{1}{\sqrt{n}} \sum\nolimits_{l=1}^{r} n^{-(r-l)/2} \sum\nolimits_{x_{l+1} \in L_{l+1}} \cdots \sum\nolimits_{x_r \in L_r} \\ \left[\int \cdots \int_{B_l} [S_1(\sqrt{n} x_l + n p_l) \phi(\boldsymbol{x})]_{\lambda_l(x^*)}^{\theta_l(x^*)} \, dx_1, \cdots, dx_{l-1} \right]; \tag{4}$$

$$L_j = \{\boldsymbol{x} \colon x_j = \frac{n_j - n p_j}{\sqrt{n}}, \ n_j \text{ and } p_j \text{ defined as before}\};$$

$$S_1(x) = x - \lfloor x \rfloor - 1/2, \ \lfloor x \rfloor \text{ is the integer part } x;$$

$$\phi(\boldsymbol{x}) = \frac{1}{(2\pi)^{r/2} |\Omega|^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{x}^T \Omega^{-1} \boldsymbol{x} \right).$$

R e m a r k 4. In [1] Yarnold showed that $J_2 = O(n^{-1/2})$.

R e m a r k 5. Using elementary transformations it can be easily shown that the determinant of the matrix Ω equals $\prod_{i=1}^{k} p_i$.

Yarnold also examined this expansion for the most known power divergence statistic, which is the chi-squared statistic. Define B^λ as $\{\boldsymbol{x} \mid T_\lambda(\boldsymbol{x}) < c\}$. It is easy to show that B^1 is an ellipsoid, which is a particular case of a bounded extended convex set. J. Yarnold managed to simplify the item (4) in this simple case and converted the expansion (3) to

$$\Pr(\boldsymbol{X} \in B^1) = G_r(c) + (N^1 - n^{r/2} V^1)\, e^{-c/2} \Big/ \left((2\pi n)^r \prod_{j=1}^{k} p_j\right)^{1/2} + O(n^{-1}), \tag{5}$$

where $G_r(c)$ is the chi-squared distribution function with r degrees of freedom; N^1 is the number of points of the lattice L in B^1; V^1 is the volume of B^1. Using the result of Esseen [8], he obtained an estimate of the second item in (5) in the form $O(n^{-(k-1)/k})$.

M. Shiotani and Y. Fujikoshi in [6] showed that, when $\lambda = 0, \lambda = -1/2$, we have

$$\begin{gathered} J_1 = G_r(c) + O(n^{-1}) \\ J_2 = (N^\lambda - n^{r/2} V^\lambda)\, e^{-c/2} \Big/ \left((2\pi n)^r \prod_{j=1}^{k} p_j\right)^{1/2} + o(1), \qquad (6) \\ V^\lambda = V^1 + O(n^{-1}). \end{gathered}$$

These results were expanded by T. Read to the case $\lambda \in \mathbb{R}$. In particular Theorem 3.1 in [10] implies

$$\Pr\left(T_\lambda < c\right) = \Pr\left(\chi_r^2 < c\right) + J_2 + O\left(n^{-1}\right). \tag{7}$$

This reduces the problem to the estimation of the order of J_2.

It is worth mentioning that papers [6] and [10] do not estimate the residual in (6). Consequently, it was impossible to construct estimates of the rate of convergence of statistics T_λ to the limiting distribution, grounded on the simple representation for J_2 initially suggested by J. Yarnold.

In this paper for any power divergence statistic we eliminated lapses of papers [6] and [10] pinpointed in the previous paragraph. Then we constructed an estimate for J_2 based on the fundamental number theory result of E. Hlawka [13].

The paper is divided into two parts. In the first one (section 2) we discuss the possibility to reduce J_2 and to convert it to the form (6). At that we accentuate correct estimation of the error of such transformation. In the second part (section 3) we investigate the applicability of the afore-mentioned theorem from number theory to the set B^λ.

1.2 The main result

In lemmas 13, 5, and 2 it is shown that $B^\lambda = \{\boldsymbol{x} \mid T_\lambda(\boldsymbol{x}) < c\}$ is a bounded extended-convex (strictly convex) set. As it has been already mentioned, in accordance with the results of J. Yarnold

$$J_2 = O\left(n^{-\frac{1}{2}}\right). \tag{8}$$

For the specific case of $r = 2$ this estimate has been considerably refined in [12]:

$$J_2 = O\left(n^{-\frac{50}{73}} (\log n)^{\frac{315}{146}}\right), r = 2.$$

For future reference we state the theorem, which pivoted the results of [12].

Theorem 1 (M. N. Huxley, 1993)**.** *Let D be a two-dimensional convex set with area A, bounded by a simple closed curve C, divided into a finite number of pieces each of those being 3 times continuously differentiable in the following sense. Namely, on each piece C_i the radius of curvature ρ is positive (and not infinite), continuous, and continuously differentiable with respect to the angle of contingence ψ. Then in a set that is obtained from D by translation and linear expansion of order M, the number of integer points equals*

$$N = AM^2 + O\left(IM^K (\log M)^\Lambda\right)$$
$$K = \frac{46}{73},\ \Lambda = \frac{315}{146},$$

where I is a number depending only on the properties of the curve C, but on the parameters M or A.

Proof. See [4], as well as [3]. □

In this paper we generalize the estimates of [12] to any dimension. We utilize proposition 9 of [13].

Theorem 2 (E. Hlawka, 1950)**.** *Let D be a compact convex set in $\mathbb{R}^m$ with the origin as its inner point. We denote the volume of this set by A. Assume that the boundary of this set is an $m-1$-dimensional surface of class $\mathbb{C}^\infty$, the Gaussian curvature being non-zero and finite everywhere on the surface.*

Also assume that a specially defined ≪canonical≫ map from the unit sphere to D is one-one and belongs to the class $\mathbb{C}^\infty$. Then in the set that is obtained from the initial one by translation along an arbitrary vector and by linear expansion with the factor M the number of integer points is

$$N = AM^m + O\left(IM^{m-2+\frac{2}{m+1}}\right)$$

where the constant I is a number dependent only on the properties of the curve C, but on the parameters M or A.

Proof. see [13], p.25-28. □

R e m a r k 6. Providing that $m = 2$, the statement of theorem 2 is weaker than the result of Huxley.

The above theorem is applicable in the current paper with $M = \sqrt{n}$. Therefore, for any fixed λ we have to deal not with a single set, but rather with a sequence of sets $B^\lambda(n)$ converging in some sense to the limiting set B_1 when $n \to \infty$. The type of this convergence will be elaborated in the sequel. At present it is worth noting that the constant I in our case, generally speaking, is dependent on n. Only having ascertained the fulfillment of the inequality

$$|I(n)| \leqslant C_0,$$

where C_0 is an absolute constant, we are able to apply theorem 2 without a change of the overall order of the error with respect to n. This statement will be proven in a separate lemma.

In the paper we prove the following important estimate of J_2 in the space of any fixed dimension $r \geqslant 3$.

Theorem 3. *For the term J_2 from decomposition* (7) *the following estimate holds*

$$J_2 = O\left(n^{-1+\frac{1}{r+1}}\right), \ r \geqslant 3, \tag{9}$$

Corollary. *For the statistic $T_\lambda(\boldsymbol{x})$ denoted by formula* (1) *it holds that*

$$\Pr(T_\lambda(\boldsymbol{x}) < c) = G_r(c) + O\left(n^{-1+\frac{1}{r+1}}\right), \ r \geqslant 3.$$

R e m a r k 7. In the case of Karl Pearson chi-squared statistics, i.e. when $\lambda = 1$, using result of Götze for ellipsoids (see [5]) and applying Yarnold's arguments from [1] one can show (see [2]) that

$$\Pr(T_1(\boldsymbol{x}) < c) = G_r(c) + O\left(n^{-1}\right), \ \text{for} \quad r \geqslant 5.$$

2 Reduction of the term J_2 to a simplified form

Let N^λ be the number of lattice points of

$$L = \{\boldsymbol{x} : x_j = \frac{1}{\sqrt{n}}(m_j - np_j),\ m_j \in \mathbb{Z},\ j = \overline{1,r}\}, \tag{10}$$

in B^λ, i. e. $N^\lambda = \#(L \cap B^\lambda)$, and V^λ is the volume of B^λ.

Theorem 4. *The item J_2 can be expressed in the form*

$$J_2 = dn^{-\frac{r}{2}}(N^\lambda - n^{\frac{r}{2}}V^\lambda) + O(n^{-1}), \tag{11}$$

where

$$d = \frac{1}{e^{\frac{c}{2}}\sqrt{(2\pi)^r \prod_{j=1}^k p_j}}.$$

Before we present the proof for this theorem, let us prove some ancillary statements.

2.1 Some ancillary facts from differential geometry

Let us first recall some definitions from a course on differential geometry.

Definition 1. *r-dimensional manifold* is defined as the set

$$M = \bigcup_i U_i \text{ where}$$

$$U_i \overset{\phi_i}{\leftrightarrow} V_i \subset \mathbb{R}^r\ \forall i.$$

ϕ_i is a one-one and continuous mapping (homeomorphism).

$U_i \subset U$ is called a *map*, and the set of U_i is called a *map atlas.* For instance, the circle $x^2+y^2 = 1$ is a one-dimensional manifold in $\mathbb{R}^2$: $U_{1,2} = \{x \mid sign(x) = \pm 1\}$, $U_{3,4} = \{y \mid sign(y) = \pm 1\}$, $\phi_{1,2}^{-1} = (y, \pm\sqrt{1-y^2})$, $\phi_{3,4}^{-1} = (x, \pm\sqrt{1-x^2})$.

The coordinate system to which ϕ_i map subsets U_i is called *local.* Functions that determine the transformation of coordinates while moving from one local coordinate system in $\mathbb{R}^r$ to another are called *transition functions.*

Definition 2. r-dimensional manifold M is said to have the smoothness class of $\mathbb{C}^m$ if all transition functions belong to this class. In particular, in the case when all transition functions are infinitely differentiable it is said that $M \in \mathbb{C}^\infty$.

For a more detailed overview of above-mentioned and related notions the reader should consult chapter 3 of [11].

From the implicit function theorem it follows that the set $M = \{T(\boldsymbol{x}) = c\}$ is an $n-1$-dimensional manifold in $\mathbb{R}^n$ if in the vicinity of each point $P \in M$ we can find a smooth implicit dependence of some coordinate on the rest $n-1$ coordinates.

Theorem 5. *Let $f : \mathbb{R}^n \to \mathbb{R}^1$ be a mapping of class C^∞, $M_c = \{\boldsymbol{x} : \ f(\boldsymbol{x}) = c\}$. If the gradient of f is non-zero at each point on the set M_c, then M_c is a smooth $(n-1)$-dimensional manifold of class C^∞.*

Proof. See [11], ch. 3, §3, theorem 2. □

R e m a r k. The assumptions of the theorem are still met if the mapping f is given on the set $Q \subset \mathbb{R}^n$ where $Q \supset M_c$.

R e m a r k. As it will be shown in lemma 5 the set $B^\lambda = \{\boldsymbol{x} \mid T_\lambda(\boldsymbol{x}) < c\}$ satisfies theorem 5. Thus we can study it with the help of powerful differential geometry instruments.

2.2 Preliminary lemmas

Lemma 1. *There exist such positive coefficients $a_1(\lambda, \boldsymbol{p}), a_2(\lambda, \boldsymbol{p}), \ldots, a_k(\lambda, \boldsymbol{p})$ and positive numbers $c_1, c_2, \ldots, c_k$ that*

$$T_\lambda(x) \geqslant a_1\, x_1^2 + a_2\, x_2^2 + \cdots + a_k\, x_k^2 - c_1 - c_2 - \cdots - c_k.$$

Proof. Included in appendix B. □

Lemma 2. *The set $B^\lambda = \{\boldsymbol{x} \mid T_\lambda(\boldsymbol{x}) < c\}$ is bounded.*

Proof. According to lemma 1 we obtain for any $\lambda \in \mathbb{R}$:

$$a_1\, x_1^2 + a_2\, x_2^2 + \cdots + a_k\, x_k^2 < c + c_1 + c_2 + \cdots + c_k.$$

Hence,

$$\forall i \ |x_i| \leqslant \sqrt{\frac{c + \sum_{l=1}^{k} c_l}{a_i(\lambda, \boldsymbol{p})}}.$$

□

Lemma 3. *Let Ω^{-1} be an inverse matrix to the covariance matrix of $\boldsymbol{Y}$, and let the range of coordinates x_i be bounded. Then the statistic $T_\lambda(\boldsymbol{x})$ can be expressed as a ≪quadratic form≫ i. e.*

$$T_\lambda(\boldsymbol{x}) = (\Omega^{-1}(n, x)x, x) \ \textit{where}$$
$$\Omega_{ij}^{-1}(n, x) = \Omega_{ij}^{-1} + O(n^{-\frac{1}{2}}) \ \textit{uniformly in } x.$$

Proof. With the help of Taylor's expansion we can obtain a schema that is analogous to (2), to within $O\left(n^{-\frac{1}{2}}\right)$ in each item. Since the range of each coordinate x_i is bounded, we can assume the estimate of this error to be independent from x. Since $x_k = -(x_1 + \ldots + x_r)$, we obtain

$$
\begin{aligned}
T_\lambda(\boldsymbol{x}) &= \sum_{i=1}^{k} x_i^2 \left(\frac{1}{p_i} + O\left(\frac{1}{\sqrt{n}}\right)\right) = \\
&= \sum_{i=1}^{r} x_i^2 \left(\frac{1}{p_i} + \frac{1}{p_{r+1}} + O\left(\frac{1}{\sqrt{n}}\right)\right) + 2\sum_{i<j} x_i x_j \left(\frac{1}{p_{r+1}} + O\left(\frac{1}{\sqrt{n}}\right)\right).
\end{aligned}
$$

It remains to note that

$$
\Omega_{ij}^{-1} = \begin{cases} \frac{1}{p_i} + \frac{1}{p_{r+1}} & \text{when } i = j, \\ \frac{1}{p_{r+1}} & \text{when } i \neq j. \end{cases}
$$

□

We will extract just one of the coordinates from equations defining the sets B^λ and B^1. Without compromising generality we will further assume that x_1 is such a coordinate.

Definition 3. Let us name the section of B^λ *maximum section* with respect to x_1 (*maximum* with respect to direction $\boldsymbol{e}$) if the result of an orthogonal projection of this section to the plane $x_1 = \mathit{const}$ (to a plane that is orthogonal to the vector $\boldsymbol{e}$) seen as an $r-1$-dimensional set is congruent to the projection of the whole set to the same plane. At that, obviously, the projection of the maximum section of the set B^λ is congruent to the set B_1^λ, and for B^1 it is congruent to the set B_1^1.

Lemma 4. *Let $S = \{\boldsymbol{x} \mid T(\boldsymbol{x}) = c\}$ be a smooth $n-1$-dimensional manifold in $\mathbb{R}^n$ and $\boldsymbol{e}$ is a certain direction. Then the maximum section with respect to $\boldsymbol{e}$ can be obtained from the necessary constraint*

$$
\frac{\partial T(\boldsymbol{x})}{\partial e} = 0.
$$

If the necessary constraint holds, the sufficient condition for the existence of (not necessarily single) maximum section would be the simultaneous fulfilment of the constraints below at any given point P on the section's boundary.

1. $T(\boldsymbol{x}) = c$,
2. $\partial^2 T(\boldsymbol{x}) / \partial e^2 > 0$,

3. minimum of $T(\boldsymbol{x})$ on the line $\boldsymbol{x} = P + \boldsymbol{e}t$ is global with respect to t.

Proof. N e c e s s i t y. From definition 3 it follows that the maximum section is defined by the points on the intersection of the projected set with the family of projecting lines, which are aligned with a directing vector $\boldsymbol{e}$. To obtain the boundary of the maximum section Q it is necessary to extract those lines of the family that intersect the set only in boundary points. Knowing that each such line has the form $\boldsymbol{x} = \boldsymbol{x_0} + \boldsymbol{e}t$ we can set the task in terms of the minimization of $T_\lambda(\boldsymbol{x})$ on the line.

It is known that the directional derivative at a point P can be calculated as per the formula

$$\frac{\partial T(\boldsymbol{x})}{\partial e} = \left.\frac{\partial T(\boldsymbol{x}(t))}{\partial t}\right|_{t=0}$$

where $\boldsymbol{x}(t)$ is any parameterized space curve that is expressed in the following form in the vicinity of a point $P = \boldsymbol{x}(0)$

$$\boldsymbol{x(t)} = \boldsymbol{x(0)} + \boldsymbol{e}t + o(t).$$

Obviously, for any point on the surface S there exists a corresponding projective line. Since as stipulated above such a line will intersect the set only on its boundary, in the vicinity of P on the line it holds that $T(\boldsymbol{x}) \geqslant c$. Therefore, the function $T(\boldsymbol{x}(0) + \boldsymbol{e}t)$ reaches its minimum when $t = 0$ (not necessarily strict minimum). Hence,

$$0 = \left.\frac{d\,T(\boldsymbol{x}(t))}{d\,t}\right|_{t=0} = \sum_{i=1}^{n} \frac{\partial T(\boldsymbol{x}(t))}{\partial x_i} \times \frac{d\,x_i(t)}{d\,t} = \operatorname{grad} T(\boldsymbol{x})\,\boldsymbol{e} = \frac{\partial T(\boldsymbol{x})}{\partial e}.$$

S u f f i c i e n c y.

The fulfillment of the first condition is obvious. The second condition, together with the necessary one, becomes sufficient for the existence of a local minimum of the function $T(\boldsymbol{x}(0) + \boldsymbol{e}t)$. Indeed, by direct calculations it is possible to show that

$$\left.\frac{d^2\,T(\boldsymbol{x}(t))}{d\,t^2}\right|_{t=0} = \left.\sum_{i,j} \frac{\partial^2 T(\boldsymbol{x})}{\partial x_i\,x_j} x_i'(t) x_j'(t)\right|_{t=0} =$$

$$= (\operatorname{grad}(\operatorname{grad} T(\boldsymbol{x}),\, \boldsymbol{e}),\, \boldsymbol{e}) = \left.\frac{\partial^2 T(\boldsymbol{x})}{\partial \boldsymbol{e}^2}\right|_{\boldsymbol{x}=\boldsymbol{x}(0)}.$$

If, in addition to the aforesaid, at the point P the third condition holds, then the corresponding projective line touches S not only in the infinitesimal vicinity of P, but also globally, i. e. the point belongs to the maximum section.

□

Lemma 5. *In the space* $\mathbb{R}^r$ *the set*

$$T_\lambda(\boldsymbol{x}) = c \tag{12}$$

is an $(r-1)$*-dimensional manifold (surface) of class* C^∞*.*

Proof. The idea of the proof is due to Zh.Assylbekov. The function $T_\lambda(\boldsymbol{x})$ is defined on the set (13), which is infinitely increasing when n approaches infinity:

$$Q = \{\boldsymbol{x} : \ x_j > -\sqrt{n}p_j, \ j = \overline{1,r}, \ x_1 + ... + x_r < \sqrt{n}p_{r+1}\}. \tag{13}$$

Coupled with the boundedness of B^λ, we obtain that beginning with some fixed N the set (13) fully incorporates the surface (12). Further, we know that the function $T_\lambda(\boldsymbol{x})$ is infinitely differentiable as a superposition of infinitely differentiable functions. Let us show that the gradient of this function does not equal zero everywhere on the surface (12). Assume there exists a point $\boldsymbol{x}^0$ on (12) such that

$$\operatorname{grad}\left[T_\lambda(\boldsymbol{x}^0)\right] = 0 \ \Rightarrow \ \frac{\partial(T_\lambda)}{\partial x_j}(\boldsymbol{x}^{\mathbf{0}}) = 0, \ j = \overline{1,r}$$
$$\Leftrightarrow \ \frac{x_j^0}{\sqrt{n}p_j} = -\frac{x_1^0 + \cdots + x_r^0}{\sqrt{n}p_{r+1}}, \ j = \overline{1,r}.$$

We can rewrite the last r equations in the form:

$$\underbrace{\begin{pmatrix} \frac{1}{\sqrt{n}p_1} + \frac{1}{\sqrt{n}p_{r+1}} & \frac{1}{\sqrt{n}p_{r+1}} & \cdots & \frac{1}{\sqrt{n}p_{r+1}} \\ \frac{1}{\sqrt{n}p_{r+1}} & \frac{1}{\sqrt{n}p_2} + \frac{1}{\sqrt{n}p_{r+1}} & \cdots & \frac{1}{\sqrt{n}p_{r+1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{n}p_{r+1}} & \frac{1}{\sqrt{n}p_{r+1}} & \cdots & \frac{1}{\sqrt{n}p_r} + \frac{1}{\sqrt{n}p_{r+1}} \end{pmatrix}}_{C} \begin{pmatrix} x_1^0 \\ x_2^0 \\ \vdots \\ x_r^0 \end{pmatrix} = 0.$$

To within a constant the matrix of this system is the inverse for the covariance matrix Ω. The inverse exists due to remark 5. Consequently, only the vector $\boldsymbol{x}^0 = (0, \ldots, 0)'$ can serve as a solution. But, on the other hand, this point does not belong to the surface since

$$T_\lambda(\boldsymbol{x}^0) = T_\lambda(0, \ldots, 0) = 0 < c.$$

Summarizing we have

$$\operatorname{grad}\left[T_\lambda(\boldsymbol{x})\right] \neq 0 \tag{14}$$

on the whole surface (12).

Applying theorem 5 to the map T_λ we obtain the statement of the current lemma.

End of proof. □

Now let us define the maximum section of the set B^λ in the direction of the axis Ox_1 from the condition $\frac{\partial T_\lambda}{\partial x_1} = 0$. It determines a plane in an r-dimensional space. For $T_1(\boldsymbol{x})$:

$$T_1(\boldsymbol{x}) = \sum_{i=1}^{r} \frac{x_i^2}{p_i} + \frac{(x_1 + \cdots + x_r)^2}{p_{r+1}}, \quad \frac{\partial T_1}{\partial x_1} = \frac{2x_1}{p_1} + \frac{2}{p_{r+1}}(x_1 + \cdots + x_r) = 0.$$

Similarly for $T_\lambda(\boldsymbol{x})$:

$$T_\lambda(\boldsymbol{x}) = \frac{2n}{\lambda(\lambda+1)} \left(\sum_{i=1}^{r} p_i \left[\left(1 + \frac{x_i}{\sqrt{n}p_i}\right)^{\lambda+1} - 1 \right] + p_{r+1} \left[\left(1 - \frac{x_1 + \cdots + x_r}{\sqrt{n}p_{r+1}}\right)^{\lambda+1} - 1 \right] \right) =$$
$$= \frac{2n}{\lambda(\lambda+1)} \left(-1 + \sum_{i=1}^{r} p_i \left(1 + \frac{x_i}{\sqrt{n}p_i}\right)^{\lambda+1} + p_{r+1} \left(1 - \frac{x_1 + \cdots + x_r}{\sqrt{n}p_{r+1}}\right)^{\lambda+1} \right)$$

$$\frac{\partial T_\lambda(\boldsymbol{x})}{\partial x_1} = \frac{2n}{\lambda} \left(\frac{\left(1 + \frac{x_1}{\sqrt{n}p_1}\right)^{\lambda}}{\sqrt{n}} - \frac{\left(1 - \frac{x_1 + \cdots + x_r}{\sqrt{n}p_{r+1}}\right)^{\lambda}}{\sqrt{n}} \right) = 0,$$

from whence we obtain a condition

$$\left(1 + \frac{x_1}{\sqrt{n}p_1}\right)^{\lambda} = \left(1 - \frac{x_1 + \cdots + x_r}{\sqrt{n}p_{r+1}}\right)^{\lambda},$$

which, accounting for the non-negativeness of the expressions in the power base, gives

$$\frac{x_1}{p_1} + \frac{x_1 + \cdots + x_r}{p_{r+1}} = 0, \tag{15}$$

i. e. the same plane as in the case of the chi-squared statistic.

R e m a r k 8. When $\lambda = 0, -1$, this plane is obtained via proceeding to the limit with regard to λ.

Since $\partial^2 T(\boldsymbol{x}) / \partial x_1^2 > 0$ holds everywhere, from lemma 4 at the intersection of plane (15) with the manifold ∂B^λ we have a single maximum section. We now find the intersection of this plane with the prelimiting and limiting sets. For B^1 we get

$$x_1 = -\frac{p_1}{p_1 + p_{r+1}}(x_2 + \cdots + x_r),$$

$$\frac{p_1}{(p_1 + p_{r+1})^2}(x_2 + \cdots + x_r)^2 + \sum_{i=2}^{r} \frac{x_i^2}{p_i} + \frac{(x_2 + \cdots + x_r)^2 \left(1 - \frac{p_1}{p_1 + p_{r+1}}\right)^2}{p_{r+1}} = c, \text{ or}$$

$$(x_2 + \cdots + x_r)^2 \left(\frac{p_1}{(p_1 + p_{r+1})^2} + \frac{p_{r+1}}{(p_1 + p_{r+1})^2} \right) + \sum_{i=2}^{r} \frac{x_i^2}{p_i} = c,$$

$$\frac{1}{p_1 + p_{r+1}}(x_2 + \cdots + x_r)^2 + \sum_{i=2}^{r} \frac{x_i^2}{p_i} = c. \tag{16}$$

Remark 9. We would have obtained the same result if we had extracted the first coordinate from the equation $T_1(\boldsymbol{x}) = c$, which in this case turns into a quadratic one. The domain of the unrestrained variables $(x_2, \ldots, x_r)$ in this case is determined by the non-negativeness of the discriminant and coincides with the interior of the domain defined by the quadratic form (16).

For the set B_λ the projection of the maximum section to the plane $x_1 = 0$, in accordance with the aforesaid, can be expressed in the form

$$\frac{c\lambda(\lambda+1)}{2n} = -1 + \sum_{i=2}^{r} p_i \left(1 + \frac{x_i}{\sqrt{n} p_i}\right)^{\lambda+1} + p_1 \left(1 - \frac{x_2 + \cdots + x_r}{\sqrt{n}(p_1 + p_{r+1})}\right)^{\lambda+1} +$$
$$+ p_{r+1} \left(1 - \frac{x_2 + \cdots + x_r - \frac{p_1}{p_1 + p_{r+1}}(x_2 + \cdots + x_r)}{\sqrt{n} p_{r+1}}\right)^{\lambda+1}.$$

Interestingly, we could express the fourth item in the right-hand side of the last equality in the same form as the third, and then add the two. We thus obtain the following for the prelimiting set:

$$\frac{c\lambda(\lambda+1)}{2n} = -1 + \sum_{i=2}^{r} p_i \left(1 + \frac{x_i}{\sqrt{n} p_i}\right)^{\lambda+1} + (p_1 + p_{r+1}) \left(1 - \frac{x_2 + \cdots + x_r}{\sqrt{n}(p_1 + p_{r+1})}\right)^{\lambda+1} \tag{17}$$

It luminously holds that

Proposition 1. *The equation* (17) *can be expressed in the form* $T_\lambda(x') = c$*, where* $x' = (x_2, \ldots, x_r)$*,* $p'_2 = p_2, \ldots p'_r = p_r$*,* $p'_{r+1} = p_1 + p_{r+1}$*. The corresponding limiting equation will obviously be* $T_1(x_2, \ldots, x_r) = c$*, with the same set of probabilities as above.*

This corollary means that the projection of the maximum section of the set B^λ to the $r-1$-dimensional space of variables is the same set B^λ, but which has a different set of probabilities and independent variables, as well as is one point "less dimensional".

Now the time has come to introduce a complementary notation

Notation.

$$\tilde{B}_1^1\colon \frac{1}{p_1+p_{r+1}}(x_2+\ldots+x_r)^2+\sum_{i=2}^{r}\frac{x_i^2}{p_i}<c-\frac{a}{\sqrt{n}},$$

where a is a constant. Analogously, we define $\tilde{B}_l^1,\ l\geqslant 2$.

Lemma 6.

$$V_{\tilde{B}_1^1}=V_{B_1^1}-\frac{a(r-1)}{2c\sqrt{n}}V_{B_1^1}+O\left(\frac{1}{n}\right) \tag{18}$$

Proof. Obviously, the mapping

$$y_i=\sqrt{\frac{c}{c-\frac{a}{\sqrt{n}}}}\,x_i,\ i=\overline{2,r}$$

converts the set $\tilde{B}_1^1$ into the following set

$$\sum_{i=2}^{r}\frac{y_i^2}{p_i}+\frac{(y_2+\ldots+y_r)^2}{p_1+p_{r+1}}<c.$$

At that the Jacobian of the map comes in the form

$$J=\begin{vmatrix}\sqrt{\frac{c-\frac{a}{\sqrt{n}}}{c}} & 0 & \cdots & 0 & 0\\ 0 & \sqrt{\frac{c-\frac{a}{\sqrt{n}}}{c}} & \cdots & 0 & 0\\ \multicolumn{5}{c}{\dotfill}\\ 0 & 0 & \cdots & \sqrt{\frac{c-\frac{a}{\sqrt{n}}}{c}} & 0\\ 0 & 0 & \cdots & 0 & \sqrt{\frac{c-\frac{a}{\sqrt{n}}}{c}}\end{vmatrix}=\left(\frac{c-\frac{a}{\sqrt{n}}}{c}\right)^{\frac{r-1}{2}}=$$

$$=\left(1-\frac{a}{\sqrt{n}c}\right)^{\frac{r-1}{2}}=1-\frac{a(r-1)}{2\sqrt{n}c}+O\left(\frac{1}{n}\right)$$

Now what the lemma states follows from the representation of volume as an integral with regard to variables $(x_2,\ldots,x_r)$ and the rule of the change of variables in an integral. $\square$

Lemma 7.

$$V_{B_1^\lambda} = V_{B_1^1}\left[1 + O\left(\frac{1}{n}\right)\right]$$

Proof. In [9] it is shown that

$$V_{B^\lambda} = V_{B^1}\cdot[1 + \frac{c}{24(k+1)n}\{(\lambda-1)^2[5S - 3k^2 - 6k + 4] -$$
$$- 3(\lambda-1)(\lambda-2)[S - 2k + 1]\}] + O\left(n^{-\frac{3}{2}}\right), S = \sum_{j=1}^{k} p_j^{-1}.$$

Nevertheless, from proposition 1 we know that the set B^λ is self-similar in the sense that the projection of its maximum section with respect to any coordinate equals the set B^λ taken with the dimension of one unit lower. Therefore

$$V_{B_1^\lambda} = V_{B_1^1}\cdot[1 + \frac{c}{24(r+1)n}\{(\lambda-1)^2[5S - 3r^2 - 6r + 4] -$$
$$- 3(\lambda-1)(\lambda-2)[S - 2r + 1]\}] + O\left(n^{-\frac{3}{2}}\right), S = \sum_{j=2}^{r} \frac{1}{p'_j}.$$

□

Lemma 8. *There exists such a constant $a = a(\lambda, \boldsymbol{p}, c)$, that beginning with some n_0*

$$\tilde{B}_1^1 \subset B_1^\lambda \text{ and } V\left(B_1^\lambda \setminus \tilde{B}_1^1\right) = \frac{a(r-1)}{2\sqrt{n}c}\cdot V_{B_1^1} + O\left(\frac{1}{n}\right)$$

Proof. We choose the constant a in the way that the set $\tilde{B}_1^1$ is a subset of B_1^λ. Let $(x_2, \dots, x_r)$ belong to $\tilde{B}_1^1$, and $\boldsymbol{p}' = (p_2, p_3, \dots, p_r, p_1 + p_{r+1})$. Then

$$T_1^{\boldsymbol{p}'}(x_2, \dots, x_r) < c - \frac{a}{\sqrt{n}}, \tag{19}$$

where $T_1^{\boldsymbol{p}'}$ is the statistic T_1 taken for the set of probabilities $\boldsymbol{p}'$ and variables $(x_2, \dots, x_r)$. On the other hand,

$$T_\lambda^{\boldsymbol{p}'}(x_2, \dots, x_r) = T_1^{\boldsymbol{p}'}(x_2, \dots, x_r) + \sum_{i=2}^{k} \frac{(\lambda-1)x_i^3}{3\sqrt{n}(p'_i)^2} +$$
$$+ \sum_{i=2}^{k} \frac{(\lambda-1)(\lambda-2)x_i^4}{12n(p'_i)^3} + O\left(n^{-\frac{3}{2}}\right)$$

Substituting inequality (19) we obtain a new inequality of the form

$$T_\lambda^{\boldsymbol{p}'}(x_2,\dots,x_r) < c - \frac{a}{\sqrt{n}} + \sum_{i=2}^{k} \frac{(\lambda-1)x_i^3}{3\sqrt{n}p_i^2} + O\left(\frac{1}{n}\right) \leqslant$$
$$\leqslant [\ x_i \text{ uniformly bounded }] \leqslant$$
$$\leqslant c - \frac{a}{\sqrt{n}} + \frac{c_2(\lambda,\boldsymbol{p},c)}{\sqrt{n}} + O\left(\frac{1}{n}\right) < [a = c_2+1] < c,\ \forall n \geqslant N(\lambda,\boldsymbol{p},c)$$

So we can assert that $\tilde{B}_1^1 \subset B_1^\lambda$. Then

$$V\left(B_1^\lambda \setminus \tilde{B}_1^1\right) = V_{B_1^\lambda} - V_{\tilde{B}_1^1} = [\text{ lemma 6 and 7 }] =$$
$$= \frac{a(r-1)}{2c\sqrt{n}} V_{B_1^1} + O\left(\frac{1}{n}\right)$$

□

Now let us estimate the number of lattice points in the difference of theese two sets (in the space of dimensionality $r-1$).

$$N_{B_1^\lambda\setminus\tilde{B}_1^1} = N_{B_1^\lambda} - N_{\tilde{B}_1^1} =$$
$$= n^{\frac{r-1}{2}} \cdot \left(V_{B_1^\lambda} - V_{\tilde{B}_1^1}\right) + \alpha_n = O\left(n^{\frac{r-2}{2}}\right) + o(n^{\frac{r-2}{2}}).$$

The last equality is proven in the following lemma

Lemma 9.

$$\alpha_n = o(n^{\frac{r-2}{2}})$$

Proof. For $r=3$ the estimate of the error α_n follows from Huxley's theorem, and for greater r it follows from Hlawka's theorem. Indeed, the applicability of these theorems to $\tilde{B}_1$ is obvious, and for B^λ it follows

1. from [12] (when $r=3$), and

$$\alpha_n = O\left(n^{\frac{23}{73}}\right) = o(\sqrt{n});$$

2. from statement 2 proven in the second part of the present paper (for any $r>3$), and

$$\alpha_n = O\left(n^{\frac{r-2}{2}-\frac{1}{2}+\frac{1}{r}}\right).$$

□

In view of the aforesaid, we obtain a summary lemma

Lemma 10.

$$N_{B_1^\lambda\setminus\tilde{B}_1^1} = O\left(n^{\frac{r-2}{2}}\right) \tag{20}$$

2.3 The transformation of J_2 representation into a simplified form

We will prove theorem 4, if we express J_2 in the form (11). Consider one item of the embracing sum with respect to l in representation (4).

$$n^{-\frac{(r-l+1)}{2}} \sum\nolimits_{x_{l+1}\in L_{l+1}} \cdots \sum\nolimits_{x_r \in L_r} \left[\int \cdots \int \chi_{B_l^\lambda}(\boldsymbol{x})[S_1(\sqrt{n}x_l + np_l)\phi(\boldsymbol{x})]_{\lambda_l(x^*)}^{\theta_l(x^*)}\, dx_1, \cdots, dx_{l-1} \right] \tag{21}$$

Having expanded the indicator function into a sum of indicator functions $\chi_{B_l^\lambda \cap \tilde{B}_l^1} + \chi_{B_l^\lambda \setminus \tilde{B}_l^1}$, we will split it into two parts. Three cases are possible for the part that comprises the indicator over the difference of sets.

1. $l = 1$. The expression (21) consists only of the following sums:

$$n^{-\frac{r}{2}} \sum_{x_2 \in L_2} \cdots \sum_{x_r \in L_r} \chi_{B_1^\lambda \setminus \tilde{B}_1^1}(\boldsymbol{x})[S_1(\sqrt{n}x_1 + np_1)\phi(\boldsymbol{x})]_{\lambda_1(x^*)}^{\theta_1(x^*)}$$

   It has the order $O\left(\frac{1}{n}\right)$ because the number of lattice points in the difference $B_1^\lambda \setminus \tilde{B}_1^1$, according to lemma 10, has the order $O\left(n^{\frac{r-2}{2}}\right)$.

2. $l = r$. The integration is carried out over the set $B_r^\lambda \setminus \tilde{B}_r^1$ with the Lebesgue measure $O\left(n^{-\frac{1}{2}}\right)$, which, together with the coefficient $n^{-\frac{(r-l+1)}{2}}$, results in the final order of $O\left(\frac{1}{n}\right)$.

3. General case: $l = t,\ 1 < t < r$. Here not only summation but also integration has to be carried out. After the integration with respect to variables $x_1, \ldots, x_t$ follows the summation of the value $O(1)$ over the lattice with respect to coordinates $x_{t+1}, \ldots, x_r$. In this summation only those points of the lattice are taken that belong to $B_{x_1,\ldots,x_t}^\lambda \setminus \tilde{B}_{x_1,\ldots,x_t}^1$. Due to the property of self-similarity (see proposition 1), we can sequentially fix the coordinates $x_1, \ldots, x_t$ and prove that the two obtained sets have the same structure as their predecessors. Consequently, in line with lemma 10 the number of points of a corresponding lattice of dimension $r - t$ in the difference set equals $O\left(n^{\frac{r-t-1}{2}}\right)$. Providing for the coefficient before the item we obtain a part in J_2 of the order $O\left(\frac{1}{n}\right)$.

Let us further address to the other item. We transform the expression

$$[S_1(\sqrt{n}x_l + np_l)\phi(\boldsymbol{x})]_{\lambda_l(x^*)}^{\theta_l(x^*)} \tag{22}$$

in the following way

$$S_1(\sqrt{n}\theta_l(x^*) + np_l)(\phi(\theta_l(x^*), x^*) - \phi(\theta(x^*), x^*)) + d[S_1(\sqrt{n}\theta_l(x^*) + np_l)-$$
$$-S_1(\sqrt{n}\lambda_l(x^*) + np_l)] - S_1(\sqrt{n}\lambda_l(x^*) + np_l)(\phi(\lambda_l(x^*), x^*) - \phi(\lambda(x^*), x^*)),$$

where $\theta(x^*)$ and $\lambda(x^*)$ are analogues of $\theta_l(x^*)$ and $\lambda_l(x^*)$ for B^1. At that $d = \phi(\theta(x^*), x^*) = \phi(\lambda(x^*), x^*)$.

R e m a r k 10. Applying Lagrange's theorem to the function ϕ in the first and the hindmost part of the expression obtained, we could reduce the problem of the estimation of their orders to the estimation of the rate of uniform convergence of the roots $\theta_l(x^*)$ and $\lambda_l(x^*)$ (on the set $B_l^\lambda \cap \tilde{B}_l^1$). Indeed, the following holds

Theorem 6. *On the set $B_l^\lambda \cap \tilde{B}_l^1$ the following uniform estimates hold:*

$$|\theta_l(x^*) - \theta(x^*)| \leqslant \frac{C}{n^{\frac{1}{4}}},\ |\lambda_l(x^*) - \lambda(x^*)| \leqslant \frac{C}{n^{\frac{1}{4}}}.$$

Proof. See appendix. □

However, if this is the case, then the residual $O\left(n^{-\frac{1}{4}}\right)$ results in the error of $O\left(n^{-\frac{3}{4}}\right)$ in the aggregate representation for J_2 after the summation over all lattice points, belonging to $B_l^\lambda \cap \tilde{B}_l^1 = \tilde{B}_l^1$. This error turns out to be the leading term in J_2 for all $r \geqslant 3$ if we use Hlawka's result, which is not helpful for proving the main theorem of the paper.

Let us change our method and utilize remark 3.

Theorem 7. *Expression* (22) *will take the form*

$$d[S_1(\sqrt{n}\theta_l(x^*) + np_l) - S_1(\sqrt{n}\lambda_l(x^*) + np_l)] + O\left(n^{-\frac{1}{2}}\right).$$

Proof. We have

$$(\Omega^{-1}x, x) = \Omega_{11}^{-1}x_1^2 + 2\sum_{j=2}^{r}\Omega_{1j}^{-1}x_1x_j + \sum_{i=2}^{r}\sum_{j=1}^{r}\Omega_{ij}^{-1}x_ix_j.$$

$$|\phi(\theta_l(x^*), x^*) - \phi(\theta(x^*), x^*)| = \frac{1}{(2\pi)^{\frac{r}{2}}|\Omega|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\sum_{i=2}^{r}\sum_{j=1}^{r}\Omega_{ij}^{-1}x_i x_j} \times$$
$$\times \left| e^{-\frac{1}{2}(\Omega_{11}^{-1}\theta_l(x^*)^2 + 2\sum_{j=2}^{r}\Omega_{1j}^{-1}\theta_l(x^*)x_j)} - e^{-\frac{1}{2}(\Omega_{11}^{-1}\theta(x^*)^2 + 2\sum_{j=2}^{r}\Omega_{1j}^{-1}\theta(x^*)x_j)} \right| \leqslant$$
$$\leqslant [\text{ Lipschitz property of the exponent }] \leqslant$$
$$\leqslant L\left(\Omega_{11}^{-1}(\theta_l^2(x^*) - \theta^2(x^*)) + 2\sum_{j=2}^{r}\Omega_{ij}^{-1}(\theta_l(x^*) - \theta(x^*))x_j\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

It is possible to estimate the last expression using the relationship derived from lemma 2,

$$(\Omega^{-1}x, x)|_{(\theta(x^*), x^*)} = c, \ (\Omega^{-1}(n, x)x, x)|_{(\theta_l(x^*), x^*)} = c,$$
$$\Omega_{ij}^{-1}(n, x) - \Omega_{ij}^{-1} = O(n^{-\frac{1}{2}}).$$

□

If we summate the error obtained through the theorem over the lattice points in the set $B_l^\lambda \cap \tilde{B}_l^1$ (integrate in the appropriate case) and multiply by a corresponding coefficient, we will obtain $O(n^{-1})$ in the aggregate representation for J_2.

Now it can be seen that the principal part of J_2 is a sum-integral of the form

$$n^{-\frac{(r-l+1)}{2}} \sum\nolimits_{x_{l+1}\in L_{l+1}} \cdots \sum\nolimits_{x_r \in L_r}$$
$$\left[\int \cdots \int \chi_{B_l^\lambda} \cap \chi_{\tilde{B}_l^1}(\boldsymbol{x})[d \cdot S_1(\sqrt{n}x_l + np_l)]_{\lambda_l(x^*)}^{\theta_l(x^*)} \, dx_1, \cdots, dx_{l-1}\right]$$

Rewriting it as a difference through the use of indicators $\chi_{B_l^\lambda}$ and $\chi_{B_l^\lambda \setminus \tilde{B}_l^1}$ and attributing the sum-integral over the difference of sets to the error, we have

$$n^{-\frac{(r-l+1)}{2}} \sum\nolimits_{x_{l+1}\in L_{l+1}} \cdots \sum\nolimits_{x_r \in L_r}$$
$$\left[\int \cdots \int \chi_{B_l^\lambda}(\boldsymbol{x})[d \cdot S_1(\sqrt{n}x_l + np_l)]_{\lambda_l(x^*)}^{\theta_l(x^*)} \, dx_1, \cdots, dx_{l-1}\right] + O(n^{-1})$$

Finally, we apply the reasoning on p. 1571-1572, [1] for the chi-squared statistic to the principal part of the last expression and obtain the item J_2 in the form

$$J_2 = (N^\lambda - n^{\frac{r}{2}}V^\lambda)\, e^{-\frac{c}{2}} / \left((2\pi n)^r \prod\nolimits_{j=1}^{k} p_j\right)^{\frac{1}{2}} + O\left(n^{-1}\right). \tag{23}$$

Thus, we obtain the simplified version of J_2. End of the first part.

3 Applicability of Hlawka's theorem to the sequence $B^\lambda(n)$

On the next step we aim at the estimation of

$$N^\lambda - n^{\frac{r}{2}} V^\lambda,$$

taken from (23). To do this we investigate geometric properties of the set B^λ.

3.1 Convexity of B^λ

Lemma 11. *Let a function $f(x)$ be defined and have two derivatives on a convex set Q. Then the function is strictly convex on Q if the second differential $d^2 f$ of this function at all points Q is a positively defined quadratic form.*

Proof. See [7], chapter 14, §7, lemma 2. □

Lemma 12. *The function $T_\lambda(\boldsymbol{x})$ defined by formula* (2) *is strictly convex on the set*

$$Q = \{\boldsymbol{x} : \ x_j > -\sqrt{n} p_j, \ j = \overline{1, r}, \ x_1 + ... + x_r < \sqrt{n} p_{r+1}\}. \tag{24}$$

Proof. The idea of the proof is due to Zh.Assylbekov. The set Q is convex since it is an open r-dimensional pyramid. We compute second-order partial derivatives of the function $T_\lambda(\boldsymbol{x})$:

$$\frac{\partial^2 (T_\lambda)}{\partial x_i^2} = \frac{2}{p_i} \left(1 + \frac{x_i}{\sqrt{n} p_i}\right)^{\lambda - 1} + \frac{2}{p_{r+1}} \left(1 - \frac{x_1 + \cdots + x_r}{\sqrt{n} p_{r+1}}\right)^{\lambda - 1}, \ i = \overline{1, r}, \tag{25}$$

$$\frac{\partial^2 (T_\lambda)}{\partial x_i \partial x_j} = \frac{2}{p_{r+1}} \left(1 - \frac{x_1 + \cdots + x_r}{\sqrt{n} p_{r+1}}\right)^{\lambda - 1}, \ i \neq j. \tag{26}$$

All the above-mentioned derivatives are continuous on Q, that's why the function $T_\lambda(\boldsymbol{x})$ is two times differentiable on Q. Due to lemma 11 the statement of the current lemma will be proven if we show that $d^2(T_\lambda)$ is a positively defined quadratic form. To do this it is sufficient to prove that leading principal minors Δ_l, $l = \overline{1, r}$ of the matrix $A = \left(\frac{\partial^2 (T_\lambda)}{\partial x_i \partial x_j}\right)$ are positive and use Sylvester's criterion. We then make use of induction with respect to l:

1. $l = 1$.

$$\Delta_1 = \frac{\partial^2 (T_\lambda)}{\partial x_1^2} > 0 \tag{27}$$

   due to (25) and (13).

2. Let $\Delta_{l-1} > 0$. We denote

$$a_i = \frac{2}{p_i}\left(1 + \frac{x_i}{\sqrt{n}p_i}\right)^{\lambda-1}, \; i = \overline{1, r}, \tag{28}$$

$$b = \frac{2}{p_{r+1}}\left(1 - \frac{x_1 + \dots + x_r}{\sqrt{n}p_{r+1}}\right)^{\lambda-1}. \tag{29}$$

Observe that $a_i > 0$, $b > 0$ due to (13). It follows from (25) and (26) that the matrix A can be rewritten in the form:

$$\begin{pmatrix} a_1 + b & b & \dots & b \\ b & a_2 + b & \dots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \dots & a_r + b \end{pmatrix}, \tag{30}$$

and, consequently,

$$\Delta_l = \begin{vmatrix} a_1 + b & b & \dots & b \\ b & a_2 + b & \dots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \dots & a_l + b \end{vmatrix} = \{ \text{ from the property of}$$

$$\text{a determinant}\} = \underbrace{\begin{vmatrix} a_1 + b & b & \dots & b \\ b & a_2 + b & \dots & b \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_l \end{vmatrix}}_{A_l} + \underbrace{\begin{vmatrix} a_1 + b & b & \dots & b \\ b & a_2 + b & \dots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \dots & b \end{vmatrix}}_{B_l}. \tag{31}$$

Conducting the decomposition of the determinant A_l with respect to the last row, we obtain:

$$A_l = \Delta_{l-1} a_l > 0 \tag{32}$$

due to the induction assumption. Subtracting from the first $(l-1)$ rows the lth row, we obtain in the determinant B_l:

$$B_l = \begin{vmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \dots & b \end{vmatrix} = a_1 a_2 \dots a_{l-1} b > 0. \tag{33}$$

From (31), (32), and (33) we infer that

$$\Delta_l > 0.$$

End of proof. $\square$

Lemma 13. *The set B^λ is strictly convex.*

Proof. The idea of the proof is due to Zh.Assylbekov. See also Lemma 5 in [12].

We fix some $\boldsymbol{x_1} \in B^\lambda$, $\boldsymbol{x_2} \in B^\lambda$, $t \in [0, 1]$. Then $T_\lambda(\boldsymbol{x_1}) < c$, $T_\lambda(\boldsymbol{x_2}) < c$. Due to lemma 12 the function $T_\lambda(\boldsymbol{x})$ is strictly convex on Q. Therefore,

$$T_\lambda(\boldsymbol{x_1} + t(\boldsymbol{x_2} - \boldsymbol{x_1})) < T_\lambda(\boldsymbol{x_1}) + t(T_\lambda(\boldsymbol{x_2}) - T_\lambda(\boldsymbol{x_1})) = \\ = (1-t)T_\lambda(\boldsymbol{x_1}) + tT_\lambda(\boldsymbol{x_2}) < (1-t)c + tc = c.$$

Consequently, $\boldsymbol{x_1}+t(\boldsymbol{x_2}-\boldsymbol{x_1}) \in B^\lambda$. Repeating this reasoning for an arbitrary pair of points $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ taken from the boundary of the set B^λ (i.e. for such $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ that $T_\lambda(\boldsymbol{x_1}) = c$, $T_\lambda(\boldsymbol{x_2}) = c$), we prove the strict convexity of the set. End of proof. $\square$

3.2 Sufficient conditions for the applicability of Hlawka's theorem

Recall that N^λ is the number of lattice points in L that fall into the set B^λ. Since the lattice L has a step equal to $n^{-\frac{1}{2}}$, we can regard N^λ as the number of integer points in the set derived from the set B^λ by a linear extension with the factor $\sqrt{n}$. Thus, in terms of theorem 2 we can consider the linear factor $M = \sqrt{n}$.

For a start we will show that the condition on the «canonical» mapping can be excluded from those conditions of theorem 2 that require our consideration. The mapping is from $\mathbb{R}^r$ to $\mathbb{R}^r$, and it maps each vector u on the unit sphere to a vector $\boldsymbol{x}(u) \in B^\lambda(n)$ such that the unit normal to the surface at this point equals u. Obviously, the vector $\boldsymbol{x}(u)$ defined in such a way is equal to the support vector of the set $B^\lambda(c)$ in the direction u. We can assume that all the set is parameterized by points of a unit sphere. At that the mapping inverse to the «canonical» mapping moves the radius-vector of any point on the surface into the normal vector to the surface at this point.

Since $B^\lambda(c)$ is a strictly convex set (lemma 11), the «canonical» mapping is one-one. Moreover, the set $B^\lambda(c)$ is implicitly defined by a function of class $\mathbb{C}^\infty$ and, consequently, can be regarded as a level surface . Hence, it is possible to define a normal at a point on the surface via a normalized gradient of the function $T^\lambda(\boldsymbol{x})$, which in accordance with the aforesaid does not equal zero and is infinitely smooth. As a result the inverse and initial ≪canonical≫ mappings are infinitely differentiable in our case.

The following lemma states the requirements that should be satisfied in order to get rid of the dependence on n in the result of theorem 2.

Lemma 14. *Assume that the conditions of theorem 2 are satisfied for $B(n)$, and, moreover,*

1. *at every point of the boundary of the set its Gaussian curvature $K_n(u)$ is located within limits that are independent from n, u and uniformly separated from zero with regard to these parameters:*

$$0 < K_0 \leqslant K_n(u) \leqslant K_1,$$

2. *for any u on the unit sphere the support function $H_n(u)$ of the set $B(n)$ is uniformly bounded with respect to n and uniformly separated from 0, i. e.*

$$H_1 \geqslant H_n(u) \geqslant H_0 > 0, \ |u| = 1.$$

3. *Partial derivatives of $H_n(u)$ of any order have a uniform upper bound with respect to n.*

Then

$$\left|N - n^{\frac{r}{2}}\,V\right| \leqslant c \cdot n^{\frac{r}{2}-1+\frac{1}{r+1}}, \tag{34}$$

where the constant c does not depend on n.

Proof. The proof almost verbatim reiterates the reasoning in the proof of proposition 9 of [13]. However, we have to ensure that residual constants will be bounded uniformly in n. To achieve it we consistently trace estimates in Satz 1-9. Some short remarks on this process are given below.

Satz 1. Does not involve any residual terms.

Satz 2 (Hilfssatz 1). In the proof of Satz 2 Hlawka introduces additional parametrization of the unit sphere E_m by points of another unit sphere E_{m-1}:

$$u_1 = \cos v, u_j = \sin v \cdot a_j (j \geqslant 2), \ \sum_{j=2}^{m} a_j^2 = 1, \ \boldsymbol{x} = \boldsymbol{x}(v).$$

At that all the derivatives of functions u_j with respect to v are bounded. In place of functions f and g being used in Hilfssatz 1 the functions $f_n(v) = x_1(v), \ g_n(v) = K_n(v) \cdot \sin^{m-2} v \cdot \cos v, \ a = 0, \ b = \pi$ are taken. From estimates (14) – (17) in Satz 2 and the reasoning that immediately follows we can conclude that the estimates

$$f_n''(a) \leqslant -\rho_1 < 0, \; f_n''(b) \geqslant \rho_1 > 0,$$
$$\min_{[a+c_1, b-c_1]} |f_n'(x)| = C_1 > 0, \; \max_{[a,b]} |f_n'''(x)| = C_2(n) \leqslant C_2$$

are uniform in n. Let us go on to Hilfssatz 1. First, note the constant C from Hillfsatz 1 can be regarded as uniformly bounded. Moreover, since $K_n(u)$ is the sum of all $m-1$-dimensional minors of the gessian of the support function $H_n(u)$, this curvature, together with its derivatives of all orders, will be uniformly bounded in n. Consequently, the same will hold for $g_n(v)$. That's why $O(e^{-j})$ in (6), Hilfssatz 1 can be deemed independent from n. Tracing the whole proof throughout Hilfssatz 1 makes sure that the order of errors is nowhere dependent on n. Then we trace the order of errors in Satz 2 in the same way.

Satz 3, Satz 4. All the errors can be regarded as independent from n providing the requirements of the theorem are fulfilled.

Satz 5. In equality (3) constants C_1 and C_2 are uniformly bounded in n.

Satz 6 - Satz 8. These sections prepare the ground for conclusions narrated in Satz 9. We just continue to trace the order of errors.

Satz 9. Hlawka uses the results of previous sections of his paper. He manipulates with the residuals, which as proven before are not dependent on n. As a result we obtain inequalities (9), which are translated into the following equality

$$\Phi(y,t) = Vt^{\frac{m}{2}} + O\left(t^{\frac{m(m-1)}{2(m+1)}}\right), \tag{35}$$

where V is the m-dimensional volume of the set $B(n)$, $\sqrt{t}$ is the order of linear expansion ($M = \sqrt{n}$), y is the transition vector of the set with respect to the origin, $\Phi(y,t)$ — the number of integer points in the set obtained by the linear expansion and the transition. Putting $m = r$, $t = n$ we obtain the sought after equality (34).

□

3.3 Fulfillment of sufficient conditions for the sets $B^\lambda(n)$

We investigate the fulfillment of lemma-14 requirements for the sets $B^\lambda(n)$. First, we look at B^1 as the limit of the above-mentioned sets and at the same time the simplest member of the family $B^\lambda(n)$.

Lemma 15. *The Gaussian curvature of a unit sphere in a multidimensional space equals one at each point of its surface.*

Proof. Since the Gaussian curvature is an invariant of weight 0 (such known absolute invariant) with respect to the group of smooth non-degenerate coordinate transformations, it is sufficient to consider a unit sphere in the multidimensional spherical system of coordinates. We have

$$\begin{aligned} x_1 &= r\cos\theta_1, \\ x_2 &= r\sin\theta_1\cos\theta_2, \\ x_3 &= r\sin\theta_1\sin\theta_2\cos\theta_2, \\ &\dots \\ x_{r-1} &= r\sin\theta_1\sin\theta_2\dots\sin\theta_{r-2}\cos\theta_{r-1}, \\ x_r &= r\sin\theta_1\sin\theta_2\dots\sin\theta_{r-2}\sin\theta_{r-1} \end{aligned} \tag{36}$$

where $\theta_i \in [0,\pi), i = \overline{1, r-2}$, and $\theta_{r-1} \in [-\pi,\pi)$. In the new coordinate system the sphere has a simple representation of $r = 1$; such value of the spherical radius determines an unequivocal parametrization of the sphere, which is obtained from coordinate transformation equations.

The Gaussian curvature at any fixed point u can be calculated via a ratio of the determinants corresponding to the second and the first forms of the surface given on the tangent space. Utilizing a simple calculation it can be proven that the first- form matrix $I \in \mathbb{R}^{(r-1)\times(r-1)}$ for the unit sphere consisting of elements

$$I_{ij} = (\frac{\partial \boldsymbol{x}}{\partial\theta_i}, \frac{\partial \boldsymbol{x}}{\partial\theta_j}),$$

equals the identity matrix. That is to say, $I_{ij} = \delta_i^j$. Further, the second form $II \in \mathbb{R}^{(r-1)\times(r-1)}$ consists of elements

$$II_{ij} = (\frac{\partial^2 \boldsymbol{x}}{\partial\theta_i\partial\theta_j}, \boldsymbol{n}) = -(\frac{\partial \boldsymbol{x}}{\partial\theta_i}, \frac{\partial \boldsymbol{n}}{\partial\theta_j})$$

where $\boldsymbol{n}$ is an «outward-looking» unit normal vector to the surface at the given point. The obvious geometrical fact that $\boldsymbol{x}$ is orthogonal to each vector $\frac{\partial \boldsymbol{x}}{\partial\theta_i}$ can be verified by calculating the corresponding scalar product in terms of the spherical parametrization $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{\theta})$. Since for the unit sphere $||\boldsymbol{x}|| = 1$, we conclude that the system of vectors

$$\left\{ \frac{\partial \boldsymbol{x}}{\partial\theta_i}\ {}_{i=\overline{1,r-1}}, \boldsymbol{x} \right\}$$

is orthonormal. Therefore, the normal

$$\boldsymbol{n} = \frac{[\frac{\partial \boldsymbol{x}}{\partial\theta_1} \times \frac{\partial \boldsymbol{x}}{\partial\theta_2} \times \dots \times \frac{\partial \boldsymbol{x}}{\partial\theta_{r-1}}]}{||[\frac{\partial \boldsymbol{x}}{\partial\theta_1} \times \frac{\partial \boldsymbol{x}}{\partial\theta_2} \times \dots \times \frac{\partial \boldsymbol{x}}{\partial\theta_{r-1}}]||} \tag{37}$$

should be equal to the spherical radius-vector $\boldsymbol{x}$ in modulus. In order to determine whether these vectors have the same orientation we consider the determinant $D(\left\{\frac{\partial \boldsymbol{x}}{\partial \theta_i}\ i=\overline{1,r-1}, \boldsymbol{x}\right\})$. Since any orthonormal system in the multidimensional space belongs to one and only one of two classes of equivalence (divided by the sign of the determinant of this system), an «outward-looking» unit normal by definition represents the very normal vector that complements the subsystem $\frac{\partial \boldsymbol{x}}{\partial \theta_i}\ i=\overline{1,r-1}$ to a «right» basis; a positive sign of D will mean that $\boldsymbol{n}=\boldsymbol{x}$, and a negative sign will result in $\boldsymbol{n}=-\boldsymbol{x}$. We thus have

$$D(\frac{\partial \boldsymbol{x}}{\partial \theta_1}, \frac{\partial \boldsymbol{x}}{\partial \theta_2}, \ldots, \frac{\partial \boldsymbol{x}}{\partial \theta_{r-1}}, \boldsymbol{x}) = (-1)^{r-1} \times D(\boldsymbol{x}, \frac{\partial \boldsymbol{x}}{\partial \theta_1}, \frac{\partial \boldsymbol{x}}{\partial \theta_2}, \ldots, \frac{\partial \boldsymbol{x}}{\partial \theta_{r-1}}) =$$
$$= \frac{(-1)^{r-1}}{r^{r-1}} \times J \text{ where } J = \frac{D(x_1, x_2, \ldots, x_r)}{D(r, \theta_1, \theta_2, \ldots, \theta_{r-1})}. \quad (38)$$

It is known that the Jacobian J of the transition from a spherical coordinate system to a rectangular coordinate system equals

$$J = r^{r-1} \prod_{k=1}^{r-1} (\sin \theta_i)^{r-1-k} > 0.$$

Hence, $\boldsymbol{n} = (-1)^{r-1} \times \boldsymbol{x}$. Substituting into the formula for II_{ij} we obtain

$$II_{ij} = (-1)^r I_{ij},\ II = (-1)^r I,\ |II| = (-1)^{r(r-1)} |I|.$$

Keeping in mind that the product of two consecutive natural numbers is always divisible by two, the Gaussian curvature should equal 1.

End of proof. □

Lemma 16. *The Gaussian curvature of the set B_1 is uniformly separated from* 0.

Proof. The proof is predicated on the fact that B_1 is an image of the orthogonal transformation (rotation) of an ellipsoid. For the Gaussian curvature is not changed under orthogonal transformations of the surface, it is sufficient to prove that the curvature is uniformly separated from 0 for a «standard» ellipsoid

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \cdots + \frac{x_r^2}{a_r^2} = 1.$$

Note that such an ellipsoid can be obtained via a linear transformation of the unit sphere. The transformation matrix

$$A = diag(a_1, a_2, \ldots, a_r),\ \boldsymbol{y} = A\boldsymbol{x},\ \boldsymbol{x} \in S_1(0),$$

is obviously non-degenerate.

We then look how the first form changes under a non-degenerate linear transformation of coordinates. We first notice that the matrix I is a Gram matrix and therefore can be expressed in the form

$$I = BB^T, B \text{ is the matrix of vector coordinates } \partial \boldsymbol{x}/\,\partial\theta_i \text{ written in rows.}$$

After the transformation of A these vectors for a given point become $\partial A\boldsymbol{x}/\,\partial\theta_i = A\,\partial\boldsymbol{x}/\,\partial\theta_i$. Hence, we get

$$I' = (BA^T)(BA^T)^T = BA^TAB^T.$$

Since $B \in \mathbb{R}^{(r-1)\times r}$, we are generally not able to decompose the determinant of the matrix I' into the product of determinants A and B. However, for a unit sphere $(\frac{\partial \boldsymbol{x}}{\partial\theta_i}, \frac{\partial \boldsymbol{x}}{\partial\theta_j}) = \delta_i^j$, and that's why for the ellipsoid $I'_{ij} = A^TAI_{ij} = a_ia_j\delta_i^j$ and $|I'| = |A|^2|I| = |A|^2$.

In order to comprehend how the elements of the second form are changed we in the first place define the notion of a normal in a multidimensional space. In the space $\mathbb{R}^m$ we use the symbol $[x_1 \times x_2 \times \ldots \times x_k]$ to denote an $m-k$-valent tensor $\boldsymbol{z}$, the coordinates of which in the simplest case are given by formulae

$$\boldsymbol{z}_{i_1i_2\ldots i_{m-k}} = \sum_{j_1j_2\cdots j_k} \delta_{j_1j_2\ldots j_ki_1i_2\ldots i_{m-k}} x_1^{j_1}x_2^{j_2}\ldots x_k^{j_k}. \tag{39}$$

Herein the symbol $\delta_{f_1f_2\ldots f_s}$ denotes ±1, depending on the evenness of the number of transpositions necessary to move the permutation of $f_1, f_2, \ldots, f_s$ into the natural permutation. It denotes 0 if the set $f_1, f_2, \ldots, f_s$ is either not a permutation or can not be moved into the natural permutation $1, 2, \ldots, s$. The reader should consult book [**?**] as a detailed source on tensor calculus and vector products. The operations with the generalized δ are explored in book [**?**].

Put $m = r,\ k = r-1$. Then from formula (39) we obtain

$$\boldsymbol{z}_i = \sum_{j_1j_2\cdots j_{r-1}} \delta_{j_1j_2\cdots j_{r-1}i} x_1^{j_1}x_2^{j_2}\ldots x_{r-1}^{j_{r-1}} = \\ = \sum_{j_1j_2\cdots j_{r-1}} (-1)^{\sigma(j_1,j_2,\ldots,j_{r-1})}\delta_{\tilde{j}_1\tilde{j}_2\cdots\tilde{j}_{r-1}i} x_1^{j_1}x_2^{j_2}\ldots x_{r-1}^{j_{r-1}}$$

where $\tilde{j}_1, \tilde{j}_2, \ldots, \tilde{j}_{r-1}$ are ordered ascendingly. Noting that the coordinates j_l are changed on the set $\{1, 2, \ldots, r\} \setminus \{i\}$ we come to the symbolical determinant (denote the j-th coordinate of the vector x_i by x_i^j) of the form

$$z = (-1)^{r-1} \begin{vmatrix} \boldsymbol{e_1} & \boldsymbol{e_2} & \dots & \boldsymbol{e_r} \\ x_1^1 & x_1^2 & \cdots & x_1^r \\ x_2^1 & x_2^2 & \cdots & x_2^r \\ \dots & \dots & \dots & \dots \\ x_{r-1}^1 & x_{r-1}^2 & \cdots & x_{r-1}^r \end{vmatrix} \tag{40}$$

Applying this formula to the vectors $\boldsymbol{y} = A\boldsymbol{x}$ and accounting for the diagonal property of A, we get

$$z_i' = \prod_{k \neq i} a_k \times z_i, \quad \forall i. \tag{41}$$

Since $\frac{\partial A\boldsymbol{x}}{\partial\theta_i\partial\theta_j} = A\frac{\partial\boldsymbol{x}}{\partial\theta_i\partial\theta_j}$, for the elements of the second form of the transformed set we have

$$II_{ij}' = (A\frac{\partial\boldsymbol{x}}{\partial\theta_i\partial\theta_j}, \frac{\boldsymbol{N'}}{|\boldsymbol{N'}|}) = \prod_{k=1}^{r} a_k \times II_{ij} \times \frac{|\boldsymbol{N}|}{|\boldsymbol{N'}|}. \tag{42}$$

Herein we use $\boldsymbol{N}, \boldsymbol{N'}$ to denote vector products of tangent vectors to the initial and transformed surfaces respectively. Further, for an arbitrary smooth $r-1$-dimensional manifold we prove the following equality:

$$|\boldsymbol{N}|^2 = \det(I). \tag{43}$$

To accomplish this we make use of the index notation for sums which is explained in detail, for instance, in [**?**]. From the definition of the vector product

$$\begin{aligned} |\boldsymbol{N}|^2 = &\begin{vmatrix} x_1^2 & \cdots & x_1^r \\ \dots & \dots & \dots \\ x_{r-1}^2 & \cdots & x_{r-1}^r \end{vmatrix}^2 + \cdots + \begin{vmatrix} x_1^1 & \cdots & x_1^{r-1} \\ \dots & \dots & \dots \\ x_{r-1}^1 & \cdots & x_{r-1}^{r-1} \end{vmatrix}^2 = \\ &= \sum_{i=1}^{r-1} (\delta_{i_1 i_2 \dots i_{r-1} i} \times x_1^{i_1} x_2^{i_2} \dots x_{r-1}^{i_{r-1}})^2 = \\ &= \delta_{i_1 i_2 \dots i_{r-1} i} \times \delta_{j_1 j_2 \dots j_{r-1} i} \times x_1^{i_1} x_1^{j_1} x_2^{i_2} x_2^{j_2} \dots x_{r-1}^{i_{r-1}} x_{r-1}^{j_{r-1}} \end{aligned}$$

Then

$$\delta_{i_1 i_2 \dots i_{r-1} i} \times \delta_{j_1 j_2 \dots j_{r-1} i} = \delta_{i_1 i_2 \dots i_{r-1} i}^{j_1 j_2 \dots j_{r-1} i} = \delta_{i_1 i_2 \dots i_{r-1}}^{j_1 j_2 \dots j_{r-1}}.$$

Consider the set of indices $(l_1, l_2, \dots, l_{r-1})$ obtained from $(1, 2, \dots, r-1)$ by number of transpositions needed to move the permutation $(j_1, j_2, \dots, j_{r-1})$

into the permutation $(i_1, i_2, \ldots, i_{r-1})$. When the selection $\boldsymbol{i}$ is fixed, we can transpose the multipliers $x_1^{j_1}$, $x_2^{j_2}$, ..., $x_{r-1}^{j_{r-1}}$ in each summand in the way that the selection of their lower indices would spell as $\boldsymbol{l}$. For one summand we obtain

$$\delta_{i_1 i_2 \ldots i_{r-1}}^{j_1 j_2 \ldots j_{r-1}} \times x_1^{i_1} x_{l_1}^{i_1} \ldots x_{r-1}^{i_{r-1}} x_{l_{r-1}}^{i_{r-1}}$$

The whole sum can be expressed in the index notation:

$$\delta_{l_1 l_2 \ldots l_{r-1}} \times x_1^{i_1} x_{l_1}^{i_1} \ldots x_{r-1}^{i_{r-1}} x_{l_{r-1}}^{i_{r-1}} = \delta_{l_1 l_2 \ldots l_{r-1}} \times (x_1, x_{l_1})(x_2, x_{l_2}) \ldots (x_{r-1}, x_{l_{r-1}}) = \\ = \det I$$

Thus, statement (43) is proven.

We now compare (42) and (43). We get

$$II_{ij} = |A| \times II_{ij} \times \frac{\sqrt{|I|}}{\sqrt{|I'|}}.$$

Hence,

$$K' = \frac{|II'|}{|I'|} = \frac{|A|^{r-1} |II| |I|^{\frac{r-1}{2}}}{|I'|^{\frac{r+1}{2}}} = \\ = K|A|^{r-1} \left(\frac{|I|}{|I'|} \right)^{\frac{r+1}{2}} = K \frac{|A|^{r-1}}{(|A|^2)^{\frac{r+1}{2}}} = \frac{K}{|A|^2} = \frac{1}{\prod_{i=1}^{r} a_i^2} > 0$$

□

Lemma 17. *Assume that the manifold B has an unequivocal smooth parametrization in the spherical m-dimensional system of coordinates*

$$\boldsymbol{x} = \boldsymbol{r}(\boldsymbol{\theta}) = (r(\boldsymbol{\theta}), \theta_1, \ldots, \theta_{m-1}).$$

Then its first form can be written in the following way

$$I = \begin{pmatrix} r^2 + \frac{\partial^2 r}{\partial \theta_1^2} & \frac{\partial r}{\partial \theta_1} \frac{\partial r}{\partial \theta_2} & \cdots & \frac{\partial r}{\partial \theta_1} \frac{\partial r}{\partial \theta_{m-1}} \\ \frac{\partial r}{\partial \theta_2} \frac{\partial r}{\partial \theta_1} & r^2 + \frac{\partial^2 r}{\partial \theta_2^2} & \cdots & \frac{\partial r}{\partial \theta_2} \frac{\partial r}{\partial \theta_{m-1}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial r}{\partial \theta_{m-1}} \frac{\partial r}{\partial \theta_1} & \frac{\partial r}{\partial \theta_{m-1}} \frac{\partial r}{\partial \theta_2} & \cdots & r^2 + \frac{\partial^2 r}{\partial \theta_{m-1}^2} \end{pmatrix} \tag{44}$$

Proof. This lemma can be proven by calculating vectors $\partial \boldsymbol{r} / \partial \theta_i$ in the multidimensional spherical system of coordinates and then calculating their pairwise scalar products. □

Lemma 18. *The sets $B^\lambda(n)$ have an unequivocal smooth parametrization in the space $\mathbb{R}^r$.*

Proof. The proof is grounded on the implicit function theorem and the fact that there exists a constant s independent from n such that the derivative of T_λ with respect to the polar radius fulfills the inequality

$$\frac{\partial T_\lambda(\boldsymbol{x}(\boldsymbol{\theta}))}{\partial r} \geqslant s$$

Indeed, expressing the statistic T_λ in the spherical coordinate system and introducing the aliases

$$a_{r+1} = \frac{1}{p_{r+1}},\ b_i = \frac{1}{p_i},\ a_i = b_i + a_{r+1},\ i = \overline{1, r},$$

we have

$$\frac{\partial T_\lambda}{\partial r} = 2r(\theta) \cdot f(\theta_1\theta_2 \ldots \theta_{r-1}) + O\left(\frac{1}{\sqrt{n}}\right)$$

where

$$f(\theta_1\theta_2 \ldots \theta_{r-1}) = a_1 \cos^2\theta_1 + a_2 \sin^2\theta_1 \cos^2\theta_2 + \cdots \\ + a_r \sin^2\theta_1 \sin^2\theta_2 \ldots \sin^2\theta_{r-2} \sin^2\theta_{r-1} + \frac{2}{p_{r+1}} \sum_{\substack{i<j \\ i,j=1}}^{r} g_{ij}$$

$$g_{i,\ l+i} = \begin{cases} i = 1,\ l = 1 \\ \qquad \cos\theta_1 \sin\theta_1 \cos\theta_2, \\ i = 1,\ 1 \leqslant 1 < l < r-1 \\ \qquad \cos\theta_1 \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{1+l-1} \cos\theta_{l+1}, \\ i = 1,\ l = r-1 \\ \qquad \cos\theta_1 \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{l-1} \sin\theta_l, \\ i > 1,\ l = r-i \\ \qquad \sin^2\theta_1 \sin^2\theta_2 \ldots \sin^2\theta_{i-1} (\cos\theta_i \sin\theta_i) \sin\theta_{i+1} \ldots \sin\theta_{i+l-2} \sin\theta_{i+l-1}, \\ i > 1,\ 1 \leqslant l < r-i \\ \qquad \sin^2\theta_1 \sin^2\theta_2 \ldots \sin^2\theta_{i-1} (\cos\theta_i \sin\theta_i) \sin\theta_{i+1} \ldots \sin\theta_{i+l-1} \cos\theta_{i+l}. \end{cases}$$

Note that for the sets $B^\lambda(n)$ the radius-vector r is uniformly separated from 0, and we only have to construct a lower estimate for f.

Let us turn to the double argument with respect to the angle θ_1. Then

$$f(\theta_1\theta_2\dots\theta_{r-1}) = a_1\left(\frac{1+\cos 2\theta_1}{2}\right) + a_2\left(\frac{1-\cos 2\theta_1}{2}\right)\cos^2\theta_2 + \dots$$
$$+ a_r\left(\frac{1-\cos 2\theta_1}{2}\right)\sin^2\theta_2\dots\sin^2\theta_{r-2}\sin^2\theta_{r-1} +$$
$$+ a_{r+1}\,(1-\cos 2\theta_1)\times\sum_{\substack{i=2\\ \overline{l=1,\ r-i}}}^{r-1} g'_{i,\ i+l} + a_{r+1}\sin 2\theta_1\times\sum_{l=1}^{r-1} g''_{1,\ 1+l}.$$

Herein

$$g'_{i,\ i+l} = \begin{cases} l < r-i & \sin^2\theta_2\dots\sin^2\theta_{i-1}(\sin\theta_i\cos\theta_i)\sin\theta_{i+1}\dots\sin\theta_{i+l-1}\sin\theta_{i+l},\\ l = r-i & \sin^2\theta_2\dots\sin^2\theta_{i-1}(\sin\theta_i\cos\theta_i)\sin\theta_{i+1}\dots\sin\theta_{i+l-2}\cos\theta_{i+l-1},\end{cases} \tag{45}$$

$$g''_{1,\ l+1} = \begin{cases} l=1 & \cos\theta_2,\\ 1<l<r-1 & \sin\theta_2\dots\sin\theta_l\cos\theta_{l+1},\\ l=r-1 & \sin\theta_2\dots\sin\theta_{l-1}\sin\theta_l.\end{cases} \tag{46}$$

System (46) defines a unit sphere in the space of dimension $r-1$. It means that all the equalities that are valid for the coordinates of such a sphere also hold for the quantities $g''_{1,1+l}$ (further denoted by x'_{l+1}). In particular, $(x'_2)^2 + (x'_3)^2 + \dots + (x'_r)^2 = 1$.

After regrouping and applying $a_i = b_i + a_{r+1}$ the function f can be expressed as

$$f = \frac{a_{r+1}}{2}\left(1+\cos^2\theta_2+\dots+\sin^2\theta_2\sin^2\theta_3\dots\sin^2\theta_{r-2}\sin^2\theta_{r-1}\right) + \frac{b_1}{2} +$$
$$+\left(\frac{b_2}{2}\cos^2\theta_2 + \frac{b_3}{2}\sin^2\theta_2\cos^2\theta_3 + \dots + \frac{b_r}{2}\sin^2\theta_2\sin^2\theta_{r-2}\sin^2\theta_{r-1}\right) +$$
$$+ a_{r+1}\sum_{i=2}^{r-1}\sum_{l=1}^{r-i} g'_{i,\ i+l} +$$
$$+\cos 2\theta_1[\frac{a_{r+1}}{2}(1-\cos^2\theta_2-\dots-\sin^2\theta_2\sin^2\theta_3\dots\sin^2\theta_{r-2}\cos^2\theta_{r-1} -$$
$$-2\sum_{i=2}^{r-1}\sum_{l=1}^{r-i} g'_{i,\ i+l}) + \frac{b_1}{2} - \left(\frac{b_2}{2}\cos^2\theta_2 + \dots + \frac{b_r}{2}\sin^2\theta_2\dots\sin\theta_{r-2}\cos^2\theta_{r-1}\right)] +$$
$$+\sin 2\theta_1\left(a_{r+1}\sum_{l=1}^{r-1} g''_{1,1+l}\right).$$

In the sequel we use the notation

$$\Delta = x_2' + x_3' + \dots + x_r',$$
$$B = b_2 \cos^2\theta_2 + b_3 \sin^2\theta_2 \cos^2\theta_3 + \ldots + b_r \sin^2\theta_2 \sin^2\theta_3 \cdots \sin^2\theta_{r-2} \sin^2\theta_{r-1}.$$

We note that from the definition of B it follows that

$$0 < min(b_1, b_2, \ldots, b_r) \leqslant B \leqslant max(b_1, b_2, \ldots, b_r).$$

To construct the final uniform in n lower estimate for the partial derivative of the initial statistic we express the function f in terms of all the notational elements described so far:

$$\frac{a_{r+1}}{2}(1+\Delta^2) + \frac{b_1+B}{2} + \cos 2\theta_1 (\frac{a_{r+1}}{2}(1-\Delta^2) + \frac{b_1 - B}{2}) + \sin 2\theta_1 \Delta a_{r+1} \geqslant$$
$$\geqslant \frac{a_{r+1}}{2}(1+\Delta^2) + \frac{b_1+B}{2} -$$
$$-\sqrt{\left(\frac{a_{r+1}}{2}\right)^2 (1-\Delta^2)^2 + \frac{(b_1-B)^2}{4} + \frac{a_{r+1}}{2}(1-\Delta^2)\left(\frac{b_1-B}{2}\right) + \Delta^2 a_{r+1}^2} >$$
$$> \frac{a_{r+1}}{2}(1+\Delta^2) + \frac{b_1+B}{2} -$$
$$-\sqrt{\left(\frac{a_{r+1}}{2}(1+\Delta^2)\right)^2 + \left(\frac{b_1+B}{2}\right)^2 + \frac{a_{r+1}}{2}(1+\Delta^2)\frac{b_1+B}{2}} \geqslant s > 0.$$

We observe that in the last inequality the lower boundary is independent from n and $\boldsymbol{\theta}$. □

Denote the scalar radius-vector of B^λ by $r_n(\boldsymbol{\theta})$, and the radius-vector of B^1 by $r(\boldsymbol{\theta})$.

Lemma 19. *There exists a uniform (in $\boldsymbol{\theta}$) convergence $r_n(\boldsymbol{\theta}) \rightrightarrows r(\boldsymbol{\theta})$ and an analogous uniform convergence for partial derivatives of any order.*

Proof. Without loss of generality we discuss only the three-dimensional case when $r = 2$. In this case instead of the vector of parameters $\boldsymbol{\theta}$ we have only one parameter to be named t. The proof for higher dimensions mirrors the one below.

Let $r_n(t)$ be the polar radius of the set B^λ, and $r(t)$ be the polar radius of the set B^1. Then it can be proven that

$$|r_n(t) - r(t)| \leqslant \frac{C}{\sqrt{n}} \tag{47}$$

Indeed, we have

$$T_\lambda(r_n(t),t) = c = T_1(r(t),t),\ T_\lambda(r,t) = T_1(r,t) + O\left(\frac{1}{\sqrt{n}}\right)$$

At that the error in the second equality is uniform in n due to the limitedness of the domain of coordinates. We have

$$T_1(r_n(t),t) - T_\lambda(r_n(t),t) = O\left(\frac{1}{\sqrt{n}}\right).$$

Hence, we can obtain a uniform estimate of the form:

$$|T_1(r_n(t),t) - T_1(r(t),t)| = |T_1(r_n(t),t) - T_\lambda(r_n(t),t)| \leqslant \frac{C}{\sqrt{n}}.$$

On the other hand,

$$\begin{aligned} T_1(r_n(t),t) - T_1(r(t),t) = {} & \frac{(r_n(t)\cos t)^2}{p_1} + \frac{(r_n(t)\sin t)^2}{p_2} + \frac{(r_n(t)(\cos t + \sin t))^2}{p_3} - \\ & - \left[\frac{(r(t)\cos t)^2}{p_1} + \frac{(r(t)\sin t)^2}{p_2} + \frac{(r(t)(\cos t + \sin t))^2}{p_3}\right] = \\ & = \left[\cos^2 t\left(\frac{1}{p_1} + \frac{1}{p_3}\right) + \sin^2 t\left(\frac{1}{p_2} + \frac{1}{p_3}\right) + \frac{\sin 2t}{p_3}\right](r_n^2(t) - r^2(t)) \end{aligned}$$

From the previous lemma we know that the first multiplier is uniformly lower-bounded (let us denote this multiplier by E and the corresponding lower bound by E_0). We have

$$|r_n(t) - r(t)| \leqslant \frac{C/\sqrt{n}}{E(r_n(t) + r(t))} \leqslant \frac{C/\sqrt{n}}{E_0 \cdot r(t)} \leqslant \frac{C'}{\sqrt{n}}.$$

The last transition follows from the trivial non-negativeness of $r_n(t)$ and the existence of a uniform lower bound for $r(t)$.

Thus, (47) is proven.

We know that the derivatives of solutions $r_n(t)$, $r(t)$ are expressed in terms of the derivatives of an implicit function with respect to its arguments t and $r(t)$. At that in the denominator we will notice the first derivative with respect to r of the functionals $T_\lambda(r,t), T_1(r,t)$ to some power, for instance,

$$r_n'(t) = -\frac{\frac{\partial T_\lambda(r_n(t),t)}{\partial t}}{\frac{\partial T_\lambda(r_n(t),t)}{\partial r}},\ r'(t) = -\frac{\frac{\partial T_1(r(t),t)}{\partial t}}{\frac{\partial T_1(r(t),t)}{\partial r}}$$

From what was proven in the previous lemma

$$\exists N\colon \forall n \geqslant N \ \frac{\partial T_1(r(t),t)}{\partial r} \geqslant s > 0, \ \frac{\partial T_\lambda(r_n(t),t)}{\partial r} \geqslant s > 0.$$

In that very lemma it was virtually shown that

$$\frac{\partial T_\lambda(r(t),t)}{\partial r} = \frac{\partial T_1(r(t),t)}{\partial r} + O\left(\frac{1}{\sqrt{n}}\right).$$

Similarly, we can obtain the same for the derivatives with respect to t:

$$\frac{\partial T_\lambda(r(t),t)}{\partial t} = \frac{\partial T_1(r(t),t)}{\partial t} + O\left(\frac{1}{\sqrt{n}}\right).$$

So it can be easily seen that

$$\frac{\frac{\partial T_\lambda(r(t),t)}{\partial t}}{\frac{\partial T_\lambda(r(t),t)}{\partial r}} = \frac{\frac{\partial T_1(r(t),t)}{\partial t}}{\frac{\partial T_1(r(t),t)}{\partial r}} + O\left(\frac{1}{\sqrt{n}}\right).$$

Let us work out the difference $r'(t) - r_n'(t)$:

$$\frac{\frac{\partial T_\lambda(r_n(t),t)}{\partial t}}{\frac{\partial T_\lambda(r_n(t),t)}{\partial r}} - \frac{\frac{\partial T_1(r(t),t)}{\partial t}}{\frac{\partial T_1(r(t),t)}{\partial r}} =$$

$$= \left(\frac{\frac{\partial T_\lambda(r_n(t),t)}{\partial t}}{\frac{\partial T_\lambda(r_n(t),t)}{\partial r}} - \frac{\frac{\partial T_\lambda(r(t),t)}{\partial t}}{\frac{\partial T_\lambda(r(t),t)}{\partial r}}\right) + \left(\frac{\frac{\partial T_\lambda(r(t),t)}{\partial t}}{\frac{\partial T_\lambda(r(t),t)}{\partial r}} - \frac{\frac{\partial T_1(r(t),t)}{\partial t}}{\frac{\partial T_1(r(t),t)}{\partial r}}\right)$$

Adding up the smoothness of the functions $T_\lambda(r,t),\ T_1(r,t)$ with respect to the combination of their arguments, the boundedness of the domain for (r,t), and Lagrange's theorem, we reach the inequality

$$|r_n'(t) - r'(t)| \leqslant M \cdot |r_n(t) - r(t)| + O\left(\frac{1}{\sqrt{n}}\right),$$

which entails the uniform convergence of the first derivatives of the polar radius. We can prove the uniform convergence for higher-order derivatives in absolutely the same way. □

Corollary 1. *There exists a uniform convergence in* $\boldsymbol{\theta}$ *of the Gaussian curvature of* B^λ *to the Gaussian curvature of* B^1.

Proof. The statement follows from formulae (40), (44), formulae for first- and second-form coefficients and the fact that a determinant is a sum of products of its elements. The structure of the Gaussian curvature in terms of the spherical system of coordinates preserves the uniform convergence originating from the uniform convergence of corresponding radius-vectors. □

Corollary 2. *The Gaussian curvature of the sequence $B^\lambda(n)$ is uniformly bounded and uniformly separated from zero mirroring the behavior of the Gaussian curvature of the limiting ellipsoid B^1.*

In what follows we utilize the Hausdorff metric that measures the distance between two sets A and B. Recall the definition:

$$haus(A,B) = r \Leftrightarrow \begin{cases} A \subset B + S_r(0), \\ B \subset A + S_r(0). \end{cases}$$

Lemma 20. *The support functions $H_n(\psi)$ of the manifolds $B^\lambda(n)$ are uniformly bounded and uniformly separated from 0 on a unit sphere $|\psi| = 1$.*

Proof. The proof consists of proving two substatements.

1. The sequence $B^\lambda(n)$ converges in the Hausdorff metric to the limiting ellipsoid B^1, and there exists a positive constant d such that $\mathrm{haus}(B^\lambda(n), B^1) \leqslant \frac{d}{\sqrt{n}}$.
2. For the support function H of the manifold B^1 it holds that

$$H_1 \geqslant H(\psi) \geqslant H_0 > 0, \ |\psi| = 1.$$

Providing these substatements are proven, we can take advantage of the inequality widely known in optimal control theory

$$|H_A(\psi) - H_B(\psi)| \leqslant |\psi| \times \mathrm{haus}(A,B). \tag{48}$$

We will thereby obtain the uniform in n estimate for the difference of the support functions as needed.

The statement of point 2 becomes obvious if we take into account that the ellipsoid B^1 includes 0 as its inner point. In this case we are able to find a ball $S_r(0)$ fully incorporated into B^1. That is,

$$H(\psi) \geqslant H_{S_r(0)}(\psi) = r|\psi| = r > 0$$

on a unit sphere. On the other hand, the upper estimate follows from the boundedness of the ellipsoid, i. e. the possibility to insert it into a ball of some fixed radius.

Let us prove point 1. Conducting the reasoning similar to the one in lemma 8 we obtain that there exist such constants a_1 and a_2 independent from n that the sets

$$\tilde{B}^1\colon \frac{(x_1+\ldots+x_r)^2}{p_{r+1}}+\sum_{i=1}^{r}\frac{x_i^2}{p_i}<c-\frac{a_1}{\sqrt{n}}, \tag{49}$$

$$\hat{B}^1\colon \frac{(x_1+\ldots+x_r)^2}{p_{r+1}}+\sum_{i=1}^{r}\frac{x_i^2}{p_i}<c+\frac{a_2}{\sqrt{n}} \tag{50}$$

are in the following relationships with each other

$$\tilde{B}^1\subset B^\lambda\subset\hat{B}^1. \tag{51}$$

On the other hand, there exist such $d>0$ that

$$\hat{B}^1\subset B^1+S_{\frac{d}{\sqrt{n}}}(0),\ B^1\subset\tilde{B}_1+S_{\frac{d}{\sqrt{n}}}(0). \tag{52}$$

Consider, for instance, the first of these relationships. In the sequel we use the following matrix rule for sets: $\boldsymbol{A}B=\{y=\boldsymbol{A}x\mid x\in B\}$. We have

$$\hat{B}^1=\sqrt{\frac{c+\frac{a_2}{\sqrt{n}}}{c}}\boldsymbol{E}B^1=B^1+\frac{a_2}{2c\sqrt{n}}\boldsymbol{E}B_1+O\left(\frac{1}{n}\right)\boldsymbol{E}B^1.$$

Since B^1 is bounded, there exists such $b>0$ that

$$\hat{B}^1\subset B^1+\frac{a_2b}{2c\sqrt{n}}S_1(0). \tag{53}$$

It remains to require the fulfillment of the inequality on d:

$$\frac{d}{\sqrt{n}}\geqslant\frac{a_2b}{2c\sqrt{n}}.$$

Under this requirement the right part (53) will be embedded into the right part of (52), which is what we strive to prove.

In summary for some constant d simultaneously

$$B^\lambda\subset\hat{B}^1\subset B^1+S_{\frac{d}{\sqrt{n}}}(0),\ B^1\subset\tilde{B}^1\subset B^\lambda+S_{\frac{d}{\sqrt{n}}}(0),$$

which proves point 1 completely. □

We have ascertained the first two requirements of lemma 14. Now let us check the requirement regarding partial derivatives of the support function $H_n(\psi)$.

Lemma 21. *All partial derivatives of the function $H_n(\psi)$ are uniformly in n upper bounded.*

Proof. The uniform boundedness of first-order partial derivatives follows from the boundedness of the set $B^\lambda(c)$ and the equalities

$$\frac{\partial H_n(\psi)}{\partial \psi_i} = x_i(\psi)$$

where $x_i(\psi)$ is the i-th component of the image of the special mapping from a unit sphere to $B^\lambda(c)$ suggested above. The equalities hold due to general convex body theory and are proven, for instance, in [14], p. 58.

Derivatives of second and higher orders of the function $H_n(\psi)$ can be therefore considered derivatives of the components of the vector $\boldsymbol{x}$. From optimal control theory it is known that the vector $\boldsymbol{x}(\psi)$ represents a solution of the following optimization problem

$$\sum_{i=1}^{r} x_i \psi_i \to max, \tag{54}$$

$$T_\lambda(\boldsymbol{x}) = c. \tag{55}$$

We use Lagrange's method to seek conditional extrema with fixed ψ and λ. Everywhere in what follows we will assume that $\lambda \neq 0$. For the case $\lambda = 0$ the reasoning is similar. We have

$$L = \sum_{i=1}^{r} x_i \psi_i + \beta(T_\lambda - c),$$

$$\frac{\partial L}{\partial x_i} = \psi_i + \beta \cdot \frac{\partial T_\lambda}{\partial x_i} = 0, \; \frac{\partial L}{\partial \beta} = T_\lambda - c = 0$$

Hence we obtain a system of $r+1$ non-linear equations with respect to the dependent variables $x_1, \ldots, x_r, \beta$ and independent variables $\psi_1, \ldots, \psi_r$.

$$\begin{cases} F_1(x_1, \ldots, x_r, \beta, \psi_1, \ldots, \psi_r) = 0, \\ F_2(x_1, \ldots, x_r, \beta, \psi_1, \ldots, \psi_r) = 0, \\ \ldots \\ F_r(x_1, \ldots, x_r, \beta, \psi_1, \ldots, \psi_r) = 0, \\ T_\lambda(x_1, \ldots, x_r) - c = 0 \end{cases} \tag{56}$$

Herein

$$F_i = \left(1 + \frac{x_i}{\sqrt{n} p_i}\right)^\lambda - \left(1 - \frac{x_1 + \cdots + x_r}{\sqrt{n} p_{r+1}}\right)^\lambda + \frac{\psi_i \lambda}{2\beta\sqrt{n}}. \tag{57}$$

It is clearly seen that all functions of the system, together with all their partial derivatives, are infinitely differentiable on the set $B^\lambda(c)$. Without loss of generality we consider partial derivatives of the dependent variables with respect to ψ_1. To obtain them we differentiate all equations of the system with respect to ψ_1 and in what follows we will use these equations to simplify the reasoning and summary results. Denoting

$$a = \frac{1}{p_{r+1}}\left(1 - \frac{x_1 + x_2 + \ldots + x_r}{\sqrt{n}p_{r+1}}\right)^{\lambda-1}, \; b_i = \frac{1}{p_i}\left(1 + \frac{x_i}{\sqrt{n}p_i}\right)^{\lambda-1}, \; i = \overline{1,r}, \tag{58}$$

$$c_i = \left(1 + \frac{x_i}{\sqrt{n}p_i}\right)^{\lambda} - \left(1 - \frac{x_1 + x_2 + \ldots + x_r}{\sqrt{n}p_{r+1}}\right)^{\lambda}, \; t = \frac{1}{2\beta}, \tag{59}$$

taking into account

$$c_i = -\frac{\psi_i \lambda}{2\sqrt{n}\beta}, \; i = \overline{1,r},$$

and cancelling the common multiplier $\frac{-t\lambda}{\sqrt{n}}$ out of the last (differentiated) equation, we obtain a system of linear equations over

$$\boldsymbol{y} = (\frac{\partial x_1}{\partial \psi_1}, \frac{\partial x_2}{\partial \psi_1}, \ldots, \frac{\partial x_r}{\partial \psi_1}, \frac{\partial t}{\partial \psi_1})^T$$

of the following form

$$\begin{bmatrix} a+b_1 & a & \cdots & a & \psi_1 \\ a & a+b_2 & \cdots & a & \psi_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a & a & \cdots & a & \psi_r \\ \psi_1 & \psi_2 & \cdots & \psi_r & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial x_1}{\partial \psi_1} \\ \frac{\partial x_2}{\partial \psi_1} \\ \vdots \\ \frac{\partial x_r}{\partial \psi_1} \\ \frac{\partial t}{\partial \psi_1} \end{bmatrix} = \begin{bmatrix} -t \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}. \tag{60}$$

Each component of a solution to this system is a quotient of the determinant of the matrix that is derived by substituting the right column into columns of the coefficient matrix and the coefficient matrix determinant (name them J' and J respectively). Higher-order partial derivative components are obtained by differentiating the equations of system (60). Final formulae would be more complex, but similar in structure to the simplest case of the ratio $\frac{J'}{J}$. Namely, we get a fraction with a polynomial over J, J' and their derivatives in the numerator and with a power of J in the denominator.

The determinant of the coefficient matrix can be calculated by decomposing it into a sum of determinants.

$$-J(1,2,\ldots,r) =$$
$$= \begin{vmatrix} a & a & \cdots & a & -\psi_1 \\ a & a+b_2 & \cdots & a & -\psi_2 \\ \multicolumn{5}{c}{\dotfill} \\ a & a & \cdots & a+b_r & -\psi_r \\ \psi_1 & \psi_2 & \cdots & \psi_r & 0 \end{vmatrix} + \begin{vmatrix} b_1 & a & \cdots & a & -\psi_1 \\ 0 & a+b_2 & \cdots & a & -\psi_2 \\ \multicolumn{5}{c}{\dotfill} \\ 0 & a & \cdots & a+b_r & -\psi_r \\ 0 & \psi_2 & \cdots & \psi_r & 0 \end{vmatrix}.$$

From this recurrent relationship

$$-J(1,2,\ldots,r) = -\left(b_1 J(2,3,\ldots,r) + a\prod_{i=2}^{r} b_i \left(\frac{\psi_1^2}{a} + \sum_{j=2}^{r} \frac{(\psi_1-\psi_j)^2}{b_j}\right)\right), \tag{61}$$

$$J(r) = \begin{vmatrix} a+b_r & -\psi_r \\ \psi_r & 0 \end{vmatrix} \tag{62}$$

we get

$$-J(1,2,\ldots,r) = \sum_{i=1}^{r} \psi_i^2 \prod_{k\neq i} b_k + a\cdot \sum_{1\leqslant l<m\leqslant r} (\psi_l-\psi_m)^2 \cdot \prod_{\substack{k\neq l\\ k\neq m}} b_k.$$

We know that

$$a \overset{n\to\infty}{\rightarrow} \frac{1}{p_{r+1}},\ b_i \overset{n\to\infty}{\rightarrow} \frac{1}{p_i},\ \sum_{i=1}^{r} \psi_i^2 = 1. \tag{63}$$

Consequently, $|J|$ is uniformly in n separated from 0.

The determinant J' can in turn be expressed in the form (the right part is inserted into the j-th column):

$$J' = (-t)(-1)^{1+j}\cdot \begin{vmatrix} a & a+b_2 & a & \cdots & a & a & \psi_2 \\ a & a & a+b_3 & \cdots & a & a & \psi_3 \\ \multicolumn{7}{c}{\dotfill} \\ a & a & a & \cdots & a & a+b_r & \psi_r \\ \psi_1 & \cdots & \psi_{j-1} & \psi_{j+1} & \cdots & \psi_r & 0 \end{vmatrix} \tag{64}$$

Note that due to (63) the determinant in the right part of the last equality is uniformly bounded in n and $\boldsymbol{\psi}$.

To finalize the proof of the uniform boundedness of the partial derivatives we equate functions F_i from (57) to zero according to system (56):

$$t = -\frac{\sqrt{n}c_i}{\lambda\psi_i},\ \forall i = \overline{1,r}. \tag{65}$$

Further, we slice the compact set $K : \sum_{i=1}^{r} \psi_i^2 = 1$ in the way described below. Consider some infinitesimal $\epsilon_1 > 0$ and the vicinity $U(\epsilon_1, \boldsymbol{\psi}) = \{\boldsymbol{\psi} \in K \mid |\psi_1| \leqslant \epsilon_1\}$. Put $S_1 = U^C(\epsilon_1, \boldsymbol{\psi})$. Within the set $S_1^C = U(\epsilon_1, \boldsymbol{\psi})$ we consider another vicinity $U(\epsilon_2, \boldsymbol{\psi}) = \{\boldsymbol{\psi} \in K \mid |\psi_2| \leqslant \epsilon_2\}$ and denote by S_2 the intersection $S_1^C \cap U^C(\epsilon_2, \boldsymbol{\psi})$. Continuing this process we can construct the sets $S_1, S_2, \ldots, S_r$, the process being finite because at least one component of the vector $\boldsymbol{\psi}$ is not close to zero. Since the union of all S_i covers the unit sphere K, it is sufficient to validate the uniform boundedness of the partial derivatives on each S_i, and then unite the results.

Since on S_i the inequality $|\psi_i| \geqslant \epsilon_i$ holds uniformly in n, we are able to make use of the i-th equality in (65) and formula (59) in order to obtain

$$|J'| \leqslant \frac{\sqrt{n} c_i}{\lambda \psi_i} C_1 \leqslant \frac{C_2}{\epsilon_i}.$$

From this and the inequality $|J| \geqslant C_3 > 0$, proved above follows the statement of the lemma. $\square$

From what was proven above we can formulate the following summary statement.

Proposition 2. *All the conditions of lemma 14 are fulfilled for the sequence of sets $B^\lambda(n)$.*

This is to wrap up the second part of the current paper. We now can go on to proving the main result encapsulated in theorem 3.

4 Summarizing the point

From corollary 2 to lemma 19, lemma 20, and lemma 21 it follows that we can apply lemma 14 to the sets $B^\lambda(n)$. Substituting (34) into (23) we obtain estimate (9).

A Proof of root convergence estimate

Theorem 8. *On the set* $B_l^\lambda(c) \cap \tilde{B}_l^1(c)$ *(mentioned in remark 10) the following uniform estimates hold:*

$$|\theta_l(x^*) - \theta(x^*)| \leqslant \frac{C}{n^{\frac{1}{4}}}, \ |\lambda_l(x^*) - \lambda(x^*)| \leqslant \frac{C}{n^{\frac{1}{4}}}.$$

Proof. Without loss of generality consider the root $\theta_l(x^*)$ with $l = 1$ denoted by $\theta_n(x^*)$ in what follows. We have $T_\lambda(\boldsymbol{x}) = T_1(\boldsymbol{x}) + O\left(\frac{1}{\sqrt{n}}\right)$; therefore $T_1(\theta_n(x^*), x^*) + O\left(\frac{1}{\sqrt{n}}\right) = c$. Since

$$\theta(x^*) = \frac{-\frac{1}{p_k} \cdot (x_2 + \ldots + x_r) + \sqrt{D}}{\frac{1}{p_1} + \frac{1}{p_k}}, \ D = c - \sum_{i=2}^{r} \frac{x_i^2}{p_i} - \frac{(x_2 + \ldots + x_r)^2}{p_1 + p_k},$$

for $\theta_n(x^*)$ we obtain a similar expression:

$$\theta_n(x^*) = \frac{-\frac{1}{p_k} \cdot (x_2 + \ldots + x_r) + \sqrt{D_n}}{\frac{1}{p_1} + \frac{1}{p_k}}, \ D_n = D + O\left(\frac{1}{\sqrt{n}}\right).$$

Then it not too difficult to measure the root difference

$$|\theta_n(x^*) - \theta(x^*)| = \left(\frac{1}{p_1} + \frac{1}{p_k}\right) |\sqrt{D_n} - \sqrt{D}| \leqslant \frac{\frac{C}{\sqrt{n}}}{\sqrt{D_n} + \sqrt{D}} =$$

$$= \frac{\frac{C}{\sqrt{n}}}{\sqrt{c - \sum_{i=2}^r \frac{x_i^2}{p_i} - \frac{(x_2 + \ldots + x_r)^2}{p_1 + p_k}} + \sqrt{c - \sum_{i=2}^r \frac{x_i^2}{p_i} - \frac{(x_2 + \ldots + x_r)^2}{p_1 + p_k} + O\left(\frac{1}{\sqrt{n}}\right)}} \leqslant$$

$$\leqslant [\text{ we reside on the set } \tilde{B}_1^1(c), \text{ see definition in section 2}] \leqslant$$

$$\leqslant \frac{\frac{C}{\sqrt{n}}}{\sqrt{\frac{a}{\sqrt{n}}} + \sqrt{\frac{a}{\sqrt{n}} + O\left(\frac{1}{\sqrt{n}}\right)}} \leqslant \frac{\frac{C}{\sqrt{n}}}{\sqrt{\frac{a}{\sqrt{n}}}} = O\left(n^{-\frac{1}{4}}\right).$$

End of proof. □

B Obtaining a lower estimate for the functional $T_\lambda(\boldsymbol{x})$

Lemma 22. *There exist positive coefficients* $a_1(\lambda, \boldsymbol{p}), a_2(\lambda, \boldsymbol{p}), \ldots, a_k(\lambda, \boldsymbol{p})$ *and positive constants* $c_1, c_2, \ldots, c_k$ *such that*

$$T_\lambda(\boldsymbol{x}) \geqslant a_1\, x_1^2 + a_2\, x_2^2 + \cdots + a_k\, x_k^2 - c_1 - c_2 - \cdots - c_k.$$

Proof. From remark 1 we have

$$T_\lambda(\boldsymbol{x}) = \frac{2n}{\lambda(\lambda+1)} \left[\sum_{j=1}^{k} p_j \left(\left(1 + \frac{x_j}{\sqrt{n p_j}}\right)^{\lambda+1} - 1 \right) \right], \lambda \notin \{0, 1\}; \tag{66}$$

$$T_0(\boldsymbol{x}) = 2n \sum_{j=1}^{k} p_j \left(1 + \frac{x_j}{\sqrt{n p_j}}\right) \ln \left(1 + \frac{x_j}{\sqrt{n p_j}}\right); \tag{67}$$

$$T_{-1}(\boldsymbol{x}) = -2n \sum_{j=1}^{k} p_j \ln \left(1 + \frac{x_j}{\sqrt{n p_j}}\right). \tag{68}$$

Case $\lambda \notin \{-1, 0\}$. Let us consider a fuction of one variable

$$f(\boldsymbol{x}) = \frac{2np}{\lambda(\lambda+1)} \left(\left(1 + \frac{x}{\sqrt{np}}\right)^{\lambda+1} - 1 \right) \tag{69}$$

It is defined for all $x \in D = (-\sqrt{n}p, \sqrt{n}(1-p))$, which follows from inequalities

$$\forall i = \overline{1,k}\ x_i > -\sqrt{n}p_i,\ x_1 + x_2 + \cdots + x_k < \sqrt{n}p_k,$$

leading to inequalities

$$\forall i = \overline{1,k}\ x_i < \sqrt{n}(1 - p_i).$$

At that $f(0) = 0$, and

$$f'(x) = \frac{2\sqrt{n}}{\lambda} \left(1 + \frac{x}{\sqrt{np}}\right)^{\lambda},\ \operatorname{sign}(f'(x)) = \operatorname{sign}(\lambda); \tag{70}$$

$$f''(x) = \frac{2}{p} \left(1 + \frac{x}{\sqrt{np}}\right)^{\lambda-1},\ f''(x) > 0\ \forall \lambda; \tag{71}$$

$$\lim_{x \to -\sqrt{n}p} f(x) = \begin{cases} -\frac{2np}{\lambda(\lambda+1)}, & \lambda > -1, \\ +\infty, & \lambda < -1; \end{cases} \tag{72}$$

$$\lim_{x \to \sqrt{n}(1-p)} f(x) = \left(\frac{1}{p^{\lambda+1}} - 1\right) \frac{2np}{\lambda(\lambda+1)}. \tag{73}$$

We seek a function of the form

$$\Psi(x) = ax^2 + bx - c\colon c > 0,\ a = a(\lambda, \boldsymbol{p}) > 0,\ b = b(\lambda, n), \tag{74}$$

$$f(x) \geqslant \Psi(x),\ \forall x \in (-\sqrt{n}p, \sqrt{n}(1-p)),\ \forall n. \tag{75}$$

From the analysis for $f(x)$ it follows that the search for $\Psi(x)$ splits into two cases: in the first one where $\lambda > 0$ the initial function strictly increases on the whole segment whereas in the second one where $\lambda < 0$ it strictly decreases in the same manner. In the sequel, for simplicity we will be dealing with both cases simultaneously

Given the strict convexity of $f(x)$ we can formulate sufficient conditions for the function $\Psi(x)$ in the form:

$$\text{when } \lambda > 0 \begin{cases} \text{a)} & \Psi(0) = -c,\ c > 0 \\ \text{b)} & \Psi(-\sqrt{n}p) \leqslant \frac{-2np}{\lambda(\lambda+1)} \\ \text{c)} & 0 < \Psi'(x) \leqslant f'(x),\ x \in \mathbb{R}^+ \cap D,\ \forall n \\ \text{d)} & \Psi''(x) > 0; \end{cases}$$

$$\text{when } \lambda < 0 \begin{cases} \text{a)} & \Psi(0) = -c,\ c > 0 \\ \text{b)} & \Psi(\sqrt{n}(1-p)) \leqslant \frac{-2np}{\lambda(\lambda+1)} \left(1 - \frac{1}{p^{\lambda+1}}\right) \\ \text{c)} & 0 > \Psi'(x) \geqslant f'(x),\ x \in \mathbb{R}^- \cap D,\ \forall n \\ \text{d)} & \Psi''(x) > 0. \end{cases}$$

We conduct further argumentation as follows. Inequalities $\Psi'(x) > 0, \Psi''(x) > 0$ define a convex quadratic trinomial, whose apex is located to the left of 0 when $\lambda > 0$ and to the right of 0 when $\lambda < 0$. From type-a) conditions we can determine the constant term of the quadratic trinomial. It is an infinitesimal number. Since the decomposition of $f'(x)$ in the Taylor series when $\lambda \neq 0$ has the form

$$f'(x) = \frac{2\sqrt{n}}{\lambda} + \frac{2x}{p} + \frac{(\lambda-1)x^2}{\sqrt{n}p^2} + \cdots,$$

from c) we can conclude that $b(n) \leqslant \frac{2\sqrt{n}}{\lambda}$ when $\lambda > 0$ ($b(n) \geqslant \frac{2\sqrt{n}}{\lambda}$ when $\lambda < 0$). Condition c) can in turn be rewritten as

$$\operatorname{sign}(\lambda)\phi(x) = \operatorname{sign}(\lambda)(f'(x) - \psi'(x)) \geqslant 0,\ \forall x \in \mathbb{R}^{\operatorname{sign}(\lambda)} \cap D,$$

and for its fulfillment we stipulate that

$$\phi(0) \geqslant 0,\ \phi'(x) = (f''(x) - \psi''(x)) \geqslant 0\ \forall x \in \mathbb{R}^{\operatorname{sign}(\lambda)} \cap D.$$

The first of the conditions is already fulfilled due to previously imposed restrictions on $b(n)$. The second one is equivalent to the non-negativeness of the minimum of $\phi'(x)$ on the given domain $x \in \mathbb{R}^{\mathrm{sign}(\lambda)} \cap D$. At that in case $\lambda < 0$ we consider the minimum on the whole D for simplicity. We obtain the following conditions on $a(\lambda, \boldsymbol{p})$:

$$\text{when } \lambda \geqslant 1 \ a \leqslant \frac{1}{p}, \ \text{when } \lambda < 1 \ a \leqslant \frac{1}{p^\lambda}.$$

Let us see what condition b) gives us. We have the inequality on $b(n)$:

$$b(n) \geqslant a\sqrt{n}p + \frac{2\sqrt{n}}{\lambda(\lambda+1)} - \frac{c}{\sqrt{np}}, \ \text{when } \lambda > 0$$

$$\begin{aligned} b(n) \leqslant \frac{c}{\sqrt{n}(1-p)} - a\sqrt{n}(1-p) - \frac{2\sqrt{n}p}{\lambda(\lambda+1)(1-p)}\left(1 - \frac{1}{p^{\lambda+1}}\right) = \\ = \frac{c}{\sqrt{n}(1-p)} - a\sqrt{n}(1-p) + \frac{2\sqrt{n}(1-p^{\lambda+1})}{\lambda(\lambda+1)(1-p)}\frac{1}{p^\lambda}, \ \text{when } \lambda < 0, \end{aligned}$$

and can require neglecting the minute c that

$$b(n) \geqslant \frac{2\sqrt{n}}{\lambda}\left(\frac{1}{\lambda+1} + \frac{ap\lambda}{2}\right), \ \text{when } \lambda > 0$$

$$b(n) \leqslant \frac{2\sqrt{n}}{\lambda}\left(\frac{1-p^{\lambda+1}}{(\lambda+1)(1-p)p^\lambda} - \frac{a\lambda(1-p)}{2}\right), \ \text{when } \lambda < 0.$$

Comparing in each of the two cases lower and upper estimates for $b(n)$, we can come to a conclusion that the necessary $b(n)$ exists if and only if the following additional inequalities hold $a(\lambda, \boldsymbol{p})$:

$$a \leqslant \frac{2}{p\lambda}\left(1 - \frac{1}{\lambda+1}\right) = \frac{2}{p(\lambda+1)}, \ \text{when } \lambda > 0;$$

$$a \leqslant \frac{2}{\lambda(\lambda+1)p^\lambda(1-p)^2}(1 - p^\lambda - \lambda p^\lambda + \lambda p^{\lambda+1}) = \theta(p,\lambda), \ \text{when } \lambda < 0.$$

Now we can summarize and give the inequalities sufficient for the fulfillment of conditions a)-c) for each fixed λ.

$$\text{for } \lambda > 1 \ ,$$

$$a \leqslant \frac{1}{p}, \ a \leqslant \frac{2}{p(\lambda+1)}, \ b(n) \leqslant \frac{2\sqrt{n}}{\lambda}, \ b(n) \geqslant \frac{2\sqrt{n}}{\lambda}\left(\frac{1}{\lambda+1} + \frac{ap\lambda}{2}\right);$$

$$\text{for } \lambda \in (0,1] \ ,$$

$$a \leqslant \frac{1}{p^\lambda}, \ a \leqslant \frac{2}{p(\lambda+1)}, b(n) \leqslant \frac{2\sqrt{n}}{\lambda}, \ b(n) \geqslant \frac{2\sqrt{n}}{\lambda}\left(\frac{1}{\lambda+1} + \frac{ap\lambda}{2}\right);$$

$$\text{for } \lambda < 0 \ ,$$

$$a \leqslant \frac{1}{p^\lambda}, \ a \leqslant \theta(p,\lambda), \ b(n) \geqslant \frac{2\sqrt{n}}{\lambda}, \ b(n) \leqslant \frac{2\sqrt{n}}{\lambda}\left(\frac{1-p^{\lambda+1}}{(\lambda+1)(1-p)p^\lambda} - \frac{a\lambda(1-p)}{2}\right).$$

We put $b(n) = \frac{2\sqrt{n}}{\lambda}$. For a per each λ we choose the stronger of the two inequalities and set a equal to the value that turns this inequality into an equality. Obviously, when $\lambda \in (0,1)$, we have $\frac{2}{p(\lambda+1)} > \frac{1}{p^\lambda}$. Moreover, $\theta(p,\lambda) > \frac{1}{p^\lambda}$ when $\lambda < 0$. Indeed, the comparison of the last two values is equivalent to the comparison of 0 with $Q(p) = 2(1 - p^\lambda - \lambda p^\lambda + \lambda p^{\lambda+1}) - \lambda(\lambda+1)(1-p)^2$. At that

$$\lim_{p\to 0} Q(p) = \begin{cases} -\infty, & \lambda \in (-1,0), \\ \infty, & \lambda < -1 \end{cases} \tag{76}$$

$$Q(1) = 0,\ Q'(p) = 2\lambda(\lambda+1)(1-p)(1-p^\lambda), \tag{77}$$

so that $Q(p)$ is positive when $\lambda < -1$ and negative when $\lambda \in (-1,0)$. Ultimately, we obtain

$$a(\lambda,p) = \begin{cases} \frac{2}{p(\lambda+1)}, & \lambda > 1, \\ \frac{1}{p^\lambda}, & \lambda \in (0,1], \\ \frac{1}{p^\lambda}, & \lambda < 0; \end{cases} \tag{78}$$

$$b(\lambda,n) = \frac{2\sqrt{n}}{\lambda}; \tag{79}$$

$$\Psi(x) = a(\lambda,p)x^2 + b(\lambda,n)x - c. \tag{80}$$

Case $\lambda = -1$. In this case

$$f(x) = -2np\ \ln\left(1 + \frac{x}{\sqrt{np}}\right),\ f(0) = 0$$
$$f'(x) = \frac{-2\sqrt{n}}{1 + \frac{x}{\sqrt{np}}} < 0,$$
$$f''(x) = \frac{2}{p\left(1 + \frac{x}{\sqrt{np}}\right)^2} > 0,$$

It follows that we can act in accordance with the algorithm for $\lambda < 0$. At that $b(\lambda,n) = -2\sqrt{n}$, and

$$a = \min(\lim_{\lambda\to-1} \theta(p,\lambda), \lim_{\lambda\to-1} \frac{1}{p^\lambda}) = \min(\frac{2}{1-p}\left(1 + \frac{p\ln p}{1-p}\right), p) = p.$$

Case $\lambda = 0$. In this case

$$f(x) = 2np\left(1+\frac{x}{\sqrt{np}}\right)\ln\left(1+\frac{x}{\sqrt{np}}\right),\ f(0)=0,\ \lim_{x\to-\sqrt{np}} f(x) = 0$$

$$f'(x) = 2\sqrt{n}\left(\ln\left(1+\frac{x}{\sqrt{np}}\right)+1\right),$$

$$f''(x) = \frac{2}{p+\frac{x}{\sqrt{n}}} > 0,$$

$$f(\sqrt{n}(1-p)) = 2n\ln\frac{1}{p}.$$

The function $f(x)$, unlike in previously considered situations, has an interior extremum (miminum) on D at the point $x = \sqrt{np}(\frac{1}{e}-1)$, which is equal to $\frac{-2np}{e}$. Due to the decomposition of the derivative $f'(x)$ we can take $b(n) = 2\sqrt{n}$. Let us consider the difference

$$Z(x) = f(x) - \Psi(x).$$

With given $b(n)$

$$Z'(x) = 2\sqrt{n}\ln\left(1+\frac{x}{\sqrt{np}}\right) - 2ax,\ Z'(0) = 0$$

$$Z''(x) = 2(\frac{1}{p+\frac{x}{\sqrt{n}}} - a) \geqslant 2-2a \text{ on } D.$$

We require that $a \leqslant 1$. Then the second derivative $Z''(x)$ will be positive when $x \in (-\sqrt{np}, \sqrt{n}(1-p))$ making the first one increase on this interval. Consequently, the extremum $(0, Z(0))$ will function as the single minimum. In that $Z(0) = c > 0$, $Z(x) > 0$ holds on D. Making $a = 1$ we obtain

$$\Psi(x) = x^2 + 2\sqrt{n}x - c.$$

Summary $\lambda \in \mathbb{R}$.

$$a(\lambda,p) = \begin{cases} \frac{2}{p(\lambda+1)}, & \lambda\in[1,\infty), \\ \frac{1}{p^\lambda}, & \lambda\in(-\infty,1); \end{cases} \tag{81}$$

$$b(\lambda,n) = \begin{cases} \frac{2\sqrt{n}}{\lambda}, & \lambda\neq 0 \\ 2\sqrt{n}, & \lambda = 0. \end{cases} \tag{82}$$

$$\Psi(x) = a(\lambda,p)x^2 + b(\lambda,n)x - c. \tag{83}$$

We recall that the above-mentioned lower-bound polynomials are found for each pair (x_i, p_i), $i = \overline{1,k}$. To obtain an aggregated below estimate it remains to sum over i taking into account that the dependence on n is encapsulated in the coefficient $b(\lambda, n)$, which is not dependent on i. Therefore

$$\sum_i b(\lambda, n) x_i = b(\lambda, n) \sum_i x_i = 0.$$

End of proof for the lemma. $\square$

C Acknowledgement

The second author is supported by RFBR grants, No. 08-01-00567, No. 08-01-90252 and No. 08-01-91205.

The authors are grateful to Prof.F.Götze and Mr Zh.Assylbekov for useful and inspiring discussions.

References

[1] J. K. Yarnold, Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set. The Annals of Mathematical Statistics, **43** (1972), 1566–1580.

[2] V. V. Ulyanov, F. Götze, On approximations for a distribution of a sum of lattice random vectors and its statistical applications, Hiroshima Statistical Research Group. Technical Report, March (2003), 10 pp.

[3] M. N. Huxley, Exponential sums and lattice points II. Proceedings of London Mathematical Society, **66** (1993), 279–301.

[4] M. N. Huxley, Exponential sums and lattice points III. Proceedings of London Mathematical Society, **87** (2003), 591-609.

[5] F. Götze, Lattice point problems and values of quadratic forms, Inventiones mathematicae, **157** 2004, 195–226.

[6] M. Siotani, Y. Fujikoshi, Asymptotic approximations for the distributions of multinomial goodness-of-fit statistics, Hiroshima Mathematical Journal, **14** (1984), 115–124.

[7] V. A. Ilyin, E. G. Posdnyak, Mathematical Analysis Foundations (in Russian), Part I, pub. FIZMATLIT, Moscow, 2002.

[8] C. G. Esseen, Fourier analysis of distribution functions, Acta Mathematica, **77** (1945), 1–125.

[9] N. A. C. Cressie, T. R. C. Read, Multinomial goodness-of-fit tests, Journal of the Royal Statistical Society, Series B, **46** (1984), 440–464.

[10] T. R. C. Read, Closer asymptotic approximations for the distributions of the power divergence goodness-of-fit statistics., The Annals of Mathematical Statistics, Part A, **36** (1984), 59–69.

[11] A. S. Mishenko, A. T. Fomenko, Differential geometry and topology course. Pub. Factorial Press, Moscow, 2000.

[12] Zh. A. Assylbekov, V. V. Ulyanov, V. N. Zubov. On approximation of goodness-of-fit statistics for discrete three dimensional data. Hiroshima Statistical Research Group. Technical Report 08–09, 2008, 26 pp. (Submitted to Theory of Probability and Its Applications.)

[13] E. Hlawka Über integrale auf konvexen körpern I. Mh. Math **54** (1950), 1–36.

[14] T. Bonnesen, W. Fenchel, Theorie der konvexen körper, Springer Verlag, Berlin, 1934.