Generalized limit theorems for $U-\max$ statistics with kernels defined on a plane

Ekaterina Simarova

SPbSU, EIMI

2021

U-statistics

The functional Θ on a certain class of distributions $\{P\}$ on a measurable space $(X, \mathfrak{A})$ is called *regular* if it can be written in the form

$$\Theta(P) = \int_X \cdots \int_X h(x_1, \ldots, x_m) P(dx_1) \ldots P(dx_m).$$

Symmetric function h is called a *kernel*, and the natural number m is the *degree* of the functional Θ.

U-statistics

U-statistics are unbiased estimates of $\Theta(P)$ which are defined as follows. Let $\xi_1, \xi_2, \dots, \xi_n$ be i.i.d random variables with common distribution P and let $h(x_1, \dots, x_m)$ be the kernel of Θ. Then the U-statistic of degree m is defined as

$$U_n = \binom{n}{m}^{-1} \sum_J h(\xi_{i_1}, \dots, \xi_{i_m}).$$

Here $n \geq m$ and $J = \{(i_1, \dots, i_m) : 1 \leq i_1 < \dots < i_m \leq n\}$ is a set of increasing permutations of indices $i_1, \dots, i_m$.
U-statistics were first introduced and studied in 1946 -1948 by Halmos and Hoeffding.

U − max statistics

We are interested in the following "extremal"counterpart of U-statistics, which was first studied by Lao and Mayer in 2008.

Definition of U − max statistics

$$H_n = \max_J h(\xi_{i_1}, \ldots, \xi_{i_m}).$$

$U - \max$ statistics

We are interested in the following "extremal"counterpart of U-statistics, which was first studied by Lao and Mayer in 2008.

Definition of $U - \max$ statistics

$$H_n = \max_J h(\xi_{i_1}, \ldots, \xi_{i_m}).$$

Examples

1. Largest interpoint distance $\max_{1 \leq i < j \leq n} |\xi_i - \xi_j|$, where $\xi_1, \xi_2, \ldots, \xi_n$ are i.i.d. points in the d-dimensional unit ball $B^d, d \geq 2$.

$U-\max$ statistics

We are interested in the following "extremal"counterpart of U-statistics, which was first studied by Lao and Mayer in 2008.

Definition of $U-\max$ statistics

$$H_n = \max_J h(\xi_{i_1}, \dots, \xi_{i_m}).$$

Examples

1. Largest interpoint distance $\max_{1\le i<j\le n} |\xi_i - \xi_j|$, where $\xi_1, \xi_2, \dots, \xi_n$ are i.i.d. points in the d-dimensional unit ball $B^d, d \geq 2$.
2. Largest perimeter $\max_{1\le i<j<l\le n} \mathrm{peri}(U_i, U_j, U_l)$ and largest area $\max_{1\le i<j<l\le n} \mathrm{area}(U_i, U_j, U_l)$ among all inscribed triangles whose vertices are formed by triplets of points taken from the sample $U_1, \dots, U_n$ of independent and uniformly distributed points on the unit circle S.

Applications

Limiting behavior of the perimeter of inscribed triangle

Let $U_1, U_2, \ldots, U_n$ be independent and uniformly distributed points on the unit circle S and set

$$H_n = \max_{1 \le i < j < l \le n} \mathrm{peri}(U_i, U_j, U_l).$$

Then for each $t > 0$

$$\lim_{n \to \infty} \mathbb{P}\{n^3(3\sqrt{3} - H_n) \le t\} = 1 - \exp\left(-\frac{2t}{9\pi}\right).$$

Previous results

- W. Lao, M. Mayer, *U*-max-statistics, J. Multivariate Anal. 99(2008), 2039–2052.
- M. Mayer, Random Diameters and Other *U*-max-Statistics. Ph.D. Thesis, Bern University, 2008.
- W. Lao, Some weak limit laws for the diameter of random point sets in bounded regions. Ph.D. Thesis, Karlsruhe, 2010.

Previous results

- W. Lao, M. Mayer, *U*-max-statistics, J. Multivariate Anal. 99(2008), 2039–2052.
- M. Mayer, Random Diameters and Other *U*-max-Statistics. Ph.D. Thesis, Bern University, 2008.
- W. Lao, Some weak limit laws for the diameter of random point sets in bounded regions. Ph.D. Thesis, Karlsruhe, 2010.
- E.V.Koroleva, Ya. Yu. Nikitin, *U*-max-statistics and limit theorems for perimeters and areas of random polygons, J. Multivariate Anal. 127(2014), 99–111.
- Ya. Yu. Nikitin, T. A. Polevaya. Limit theorems for areas and perimeters of random inscribed and circumscribed polygons. Zap. Nauchn. Sem. POMI (in Russian), 486(2019), 200–213.

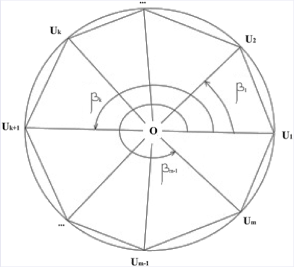
Previous results

- W. Lao, M. Mayer, *U*-max-statistics, J. Multivariate Anal. 99(2008), 2039–2052.
- M. Mayer, Random Diameters and Other *U*-max-Statistics. Ph.D. Thesis, Bern University, 2008.
- W. Lao, Some weak limit laws for the diameter of random point sets in bounded regions. Ph.D. Thesis, Karlsruhe, 2010.
- E.V.Koroleva, Ya. Yu. Nikitin, *U*-max-statistics and limit theorems for perimeters and areas of random polygons, J. Multivariate Anal. 127(2014), 99–111.
- Ya. Yu. Nikitin, T. A. Polevaya. Limit theorems for areas and perimeters of random inscribed and circumscribed polygons. Zap. Nauchn. Sem. POMI (in Russian), 486(2019), 200–213.
- Ya. Yu. Nikitin, E. N. Simarova, Generalized Limit Theorems For U-max Statistics, preprint 2010.04460.
- E. N. Simarova, Extremal random beta polytopes, preprint 2108.10951

Parametrization

Parametrization

Denote by $\beta_i = \angle U_1 O U_{i+1}$.



Consider a function h such that

$$f(U_1, \ldots, U_m) = h(\beta_1, \ldots, \beta_{m-1}).$$

Conditions on kernel f

- Function f is symmetrical;

Conditions on kernel f

- Function f is symmetrical;
- Function h is continuous and can be continuously extended to a function $h : [0, 2\pi]^{m-1} \to \mathbb{R}$.

Conditions on kernel f

- Function f is symmetrical;
- Function h is continuous and can be continuously extended to a function $h : [0, 2\pi]^{m-1} \to \mathbb{R}$.
- Function h attains its maximal value M and this maximum is realized only at a finite number of points $V_1, \dots, V_k \in [0, 2\pi]^{m-1}$. It is assumed that all $V_1, \dots, V_k \in (0, 2\pi)^{m-1}$.

Conditions on kernel f

- Function f is symmetrical;
- Function h is continuous and can be continuously extended to a function $h : [0, 2\pi]^{m-1} \to \mathbb{R}$.
- Function h attains its maximal value M and this maximum is realized only at a finite number of points $V_1, \dots, V_k \in [0, 2\pi]^{m-1}$. It is assumed that all $V_1, \dots, V_k \in (0, 2\pi)^{m-1}$.
- There exists $\delta > 0$ such that function h is three times continuously differentiable in the δ-neighborhood of any maximum point $V_i, \quad i \in \{1, \dots, k\}$.

Conditions on kernel f

- Function f is symmetrical;
- Function h is continuous and can be continuously extended to a function $h : [0, 2\pi]^{m-1} \to \mathbb{R}$.
- Function h attains its maximal value M and this maximum is realized only at a finite number of points $V_1, \dots, V_k \in [0, 2\pi]^{m-1}$. It is assumed that all $V_1, \dots, V_k \in (0, 2\pi)^{m-1}$.
- There exists $\delta > 0$ such that function h is three times continuously differentiable in the δ-neighborhood of any maximum point $V_i, \quad i \in \{1, \dots, k\}$.
- Denote by G_i the Hessian matrix of function h at the point V_i Then

$$\det(G_i) \neq 0 \text{ for all } i \in \{1, \dots, k\}.$$

Conditions on the distribution

- The random points $U_1, \ldots, U_n$ are independently distributed on the unit circle S^1 with the same probability density function $p(x)$.
- The density function p is continuous (therefore, we may think of it as of non-negative continuous 2π-periodic function $p : \mathbb{R} \to \mathbb{R}_+$ such that $\int\limits_0^{2\pi} p(x)\,dx = 1$).
- There exists at least one maximal point of the kernel (which we denote by V_*) such that

$$\int\limits_0^{2\pi} \left[p(x) \prod_{l=1}^{m-1} p(x + V_*^l) \right] dx \neq 0.$$

New results

The general form of limit distribution on a circle

Under previous conditions on f and p the following equality holds for every $t > 0$:

$$\lim_{n\to\infty} \mathbb{P}\{n^{\frac{2m}{m-1}}(M - H_n) \leq t\} = 1 - e^{-t^{\frac{m-1}{2}}\cdot K},$$

where

$$K = \frac{(2\pi)^{\frac{m-1}{2}}}{m!\Gamma\left(\frac{m+1}{2}\right)} \sum_{i=1}^{k} \left(\frac{1}{\sqrt{\det(-G_i)}} \int_0^{2\pi} p(x) \prod_{l=1}^{m-1} p(x + V_i^l)\, dx \right).$$

Special cases

Let $g : [0, 2\pi] \to \mathbb{R}$ be smooth enough and $g''(\frac{2\pi}{m}) \neq 0$. Define symmetric functions f and h as follows:

$$f(U_1, \dots, U_m) = h(\beta_1, \dots, \beta_{m-1}) = \sum_{i=1}^{m} g(\beta_i - \beta_{i-1}),$$

where $0 = \beta_0 \leq \beta_1 \leq \dots \leq \beta_{m-1} \leq \beta_m = 2\pi$.

Theorem

Consider the U-max statistics H_n with kernel f. Suppose that this kernel attains its maximum only at the vertices of a regular m-gon. Then the previous theorem holds and constant K looks as follows:

$$K = \frac{(2\pi)^{\frac{m-1}{2}} \left[\int_0^{2\pi} \prod_{l=0}^{m-1} p(x + \frac{2\pi l}{m})\, dx \right]}{m \left(-g'' \left(\frac{2\pi}{m} \right) \right)^{\frac{m-1}{2}} \Gamma\left(\frac{m+1}{2} \right) \sqrt{m}}.$$

Determinant

Determinant

Recall the definition $h(\beta_1, \dots, \beta_{m-1}) = \sum_{i=1}^{m} g(\beta_i - \beta_{i-1})$.
The Hessian matrix at the point $V^* = (\frac{2\pi}{m}, \frac{4\pi}{m}, \dots, \frac{2(m-1)}{m})$ is

$$G(V^*) = g''\left(\frac{2\pi}{m}\right)\begin{pmatrix} 2 & -1 & 0 & 0 & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 2 \end{pmatrix}$$

Therefore,

$$\det(-G(V^*)) = \left(-g''\left(\frac{2\pi}{m}\right)\right)^{m-1} \cdot m.$$

Generalised perimeter

- Strictly concave functions g :

$$f(U_1,\ldots,U_m)=\sum_{i=1}^{m} g\left(\beta_i-\beta_{i-1}\right)\leq mg\left(\frac{\beta_m-\beta_0}{m}\right)$$
$$=mg\left(\frac{2\pi}{m}\right)$$

Examples

Generalised perimeter

- Strictly concave functions g :

$$f(U_1, \ldots, U_m) = \sum_{i=1}^{m} g\left(\beta_i - \beta_{i-1}\right) \leq mg\left(\frac{\beta_m - \beta_0}{m}\right)$$
$$= mg\left(\frac{2\pi}{m}\right)$$

- Perimeter of an inscribed convex m-gon: $g(x) = 2\sin\left(\frac{x}{2}\right)$.

Examples

Generalised perimeter

- Strictly concave functions g :

$$f(U_1, \ldots, U_m) = \sum_{i=1}^{m} g\left(\beta_i - \beta_{i-1}\right) \leq mg\left(\frac{\beta_m - \beta_0}{m}\right)$$
$$= mg\left(\frac{2\pi}{m}\right)$$

- Perimeter of an inscribed convex m-gon: $g(x) = 2\sin\left(\frac{x}{2}\right)$.
- Area of an inscribed convex m-gon: $g(x) = \frac{1}{2}\sin x$.

Examples

Generalised perimeter

- Generalized perimeter $f(U_1, \ldots, U_m) = \sum_{i=1}^{m} r(|U_i U_{i+1}|)$ when r is strictly concave, increasing function: $g(x) = r(2 \sin \frac{x}{2})$.

Examples

Generalised perimeter

- Generalized perimeter $f(U_1, \ldots, U_m) = \sum_{i=1}^{m} r(|U_i U_{i+1}|)$ when r is strictly concave, increasing function: $g(x) = r(2\sin\frac{x}{2})$.
- Function $r(x) = e^{-ax}x^b(\ln\left(\frac{x}{2}\right))^c$.Denote by $\tau(a,b,c)$ the function which is equal to 1 if function r is strictly concave and increasing, and is equal to -1 if function r is strictly convex and decreasing:

$$\tau(a,b,c) = \begin{cases} (-1)^c, & \text{if } a \geq 0, b \leq 0, c \in \mathbb{N}, \\ -1, & \text{if } a \geq 0, b \leq 0, c = 0, a^2 + b^2 \neq 0, \\ 1, & \text{if } a = 0, 0 < b \leq 1, c = 0. \end{cases}$$

The ideas of the proof

Lao–Mayer theorem

For any real number z define the following functions

- $p_{n,z} = \mathbb{P}\{h(\xi_1, \dots, \xi_m) > z\}$.
- $\lambda_{n,z} = \binom{n}{m} p_{n,z}$.
- $\tau_{n,z}(r) = \frac{\mathbb{P}\{h(\xi_1,\dots,\xi_m)>z,\ h(\xi_{1+m-r},\xi_{2+m-r},\dots,\xi_{2m-r})>z\}}{p_{n,z}}$.

Then the following inequality holds:

$$\left|\mathbb{P}\{H_n \le z\} - e^{-\lambda_{n,z}}\right| \le (1 - e^{-\lambda_{n,z}}) \times$$
$$\times \left[p_{n,z}\left(\binom{n}{m} - \binom{n-m}{m}\right) + \sum_{r=1}^{m-1} \binom{m}{r}\binom{n-m}{m-r}\tau_{n,z}(r)\right].$$

The ideas of the proof

Corollary

Suppose that for some sequence of transformations $z_n : T \to \mathbb{R}$, $T \subset \mathbb{R}$, the following conditions hold:

- $\lim_{n\to\infty} \lambda_{n,z_n(t)} = \lambda_t > 0$.
- $\lim_{n\to\infty} n^{2m-r} p_{n,z} \tau_{n,z}(r) = 0$ for all $r \in \{1, \ldots, m-1\}$.

Then for any $t \in T$

$$\lim_{n\to\infty} \mathbb{P}(H_n \leq z_n(t)) = e^{-\lambda_t}.$$

The limiting relation $\lim_{n\to\infty} \lambda_{n,z_n(t)} = \lambda_t > 0$ implies that $p_{n,z} = O(n^{-m})$. Therefore the right-hand side is equal to

$$O\left(n^{-1} + \sum_{r=1}^{m-1} n^{2m-r} p_{n,z} \tau_{n,z}(r)\right).$$

Proposition 1

Under previous conditions when $\varepsilon \to 0+$ the following relation holds true :

$$\mathbb{P}\{f(U_1, \ldots, U_m) \geq M - \varepsilon\} = \varepsilon^{\frac{m-1}{2}} \cdot m! \cdot K(1 + O(\varepsilon)),$$

where

$$K = \frac{(2\pi)^{\frac{m-1}{2}}}{m!\Gamma\left(\frac{m+1}{2}\right)} \sum_{i=1}^{k} \left(\frac{1}{\sqrt{\det(G_i)}} \int_0^{2\pi} p(x) \prod_{l=1}^{m-1} p(x + V_i^l)\, dx \right).$$

The ideas of the proof

Checking conditions

Consider the sequence of transformations

$$z_n(t) = M - tn^{-\frac{2m}{m-1}} = M - \varepsilon.$$

Then

$$n^m \varepsilon^{\frac{m-1}{2}} = t^{\frac{m-1}{2}}.$$

$$\lim_{n\to\infty} \lambda_{n,z_n(t)} = \lim_{n\to\infty} \frac{n!}{m!(n-m)!} \mathbb{P}\{f(U_1,\dots,U_m) > z_n(t)\} =$$
$$= \frac{1}{m!} \lim_{n\to\infty} \frac{n!}{n^m(n-m)!} n^m \varepsilon^{\frac{m-1}{2}} \varepsilon^{-\frac{m-1}{2}} \mathbb{P}\{H_m > M - \varepsilon\} =$$
$$= \frac{1}{m!} t^{\frac{m-1}{2}} m! K = t^{\frac{m-1}{2}} K =: \lambda_t$$

The ideas of the proof

We obtain that

$$\lim_{n\to\infty} \mathbb{P}\left(H_n \leq z_n(t)\right) = e^{-\lambda_t}$$

for any $t > 0$. Hence,

$$\lim_{n\to\infty} \mathbb{P}\left(H_n \leq M - tn^{-\frac{2m}{m-1}}\right) = e^{-\lambda_t}.$$

Therefore, for any $t > 0$ the following relation is valid:

$$\lim_{n\to\infty} \mathbb{P}\{n^{\frac{2m}{m-1}}(M - H_n) \leq t)\} = 1 - e^{-\lambda_t}.$$

The ideas of the proof

Proposition 2

For each $r \in \{1, \ldots, m-1\}$ we have the following relation:

$$n^{2m-r}\mathbb{P}\{f(U_1, \ldots, U_m) > z_n(t), f(U_{1+m-r}, \ldots, U_{2m-r}) > z_n(t)\} \to 0,$$

when $n \to +\infty$.

Modernization of the conditions on the plane

Conditions on the kernel f

- Denote by $\beta_i = \angle U_1 O U_{i+1}, r_i = \|OU_i\|$. Consider a function h such that

$$f(U_1, \ldots, U_m) = h(\beta_1, \ldots, \beta_{m-1}, r_1, \ldots, r_m);$$

Modernization of the conditions on the plane

Conditions on the kernel f

- Denote by $\beta_i = \angle U_1 O U_{i+1}, r_i = \|OU_i\|$. Consider a function h such that

$$f(U_1, \ldots, U_m) = h(\beta_1, \ldots, \beta_{m-1}, r_1, \ldots, r_m);$$

- h is continuous and can be continuously extended to a function $h : [0, 2\pi]^{n-1} \times [0, 1]^n \to \mathbb{R};$

Modernization of the conditions on the plane

Conditions on the kernel f

- Denote by $\beta_i = \angle U_1 O U_{i+1}, r_i = \|OU_i\|$. Consider a function h such that

$$f(U_1, \ldots, U_m) = h(\beta_1, \ldots, \beta_{m-1}, r_1, \ldots, r_m);$$

- h is continuous and can be continuously extended to a function $h : [0, 2\pi]^{n-1} \times [0, 1]^n \to \mathbb{R}$;
- h attains its maximal value M only at a finite number of points $V_1, \ldots, V_k$ and also we assume that these points satisfy $V_1, \ldots, V_k \in (0, 2\pi)^{m-1} \times \{1\}^m$

Modernization of the conditions on the plane

Conditions on the kernel f

- Denote by $\beta_i = \angle U_1 O U_{i+1}, r_i = \|OU_i\|$. Consider a function h such that

$$f(U_1, \dots, U_m) = h(\beta_1, \dots, \beta_{m-1}, r_1, \dots, r_m);$$

- h is continuous and can be continuously extended to a function $h : [0, 2\pi]^{n-1} \times [0, 1]^n \to \mathbb{R}$;
- h attains its maximal value M only at a finite number of points $V_1, \dots, V_k$ and also we assume that these points satisfy $V_1, \dots, V_k \in (0, 2\pi)^{m-1} \times \{1\}^m$
- for any $i \in \{1, \dots, k\}$, the sub-hessian of h at V_i corresponding to the first $m-1$ arguments is non-degenerate: $\det G_i \neq 0$;

Modernization of the conditions on the plane

Conditions on the kernel f

- Denote by $\beta_i = \angle U_1 O U_{i+1}, r_i = \|OU_i\|$. Consider a function h such that
$$f(U_1, \dots, U_m) = h(\beta_1, \dots, \beta_{m-1}, r_1, \dots, r_m);$$
- h is continuous and can be continuously extended to a function $h : [0, 2\pi]^{n-1} \times [0, 1]^n \to \mathbb{R}$;
- h attains its maximal value M only at a finite number of points $V_1, \dots, V_k$ and also we assume that these points satisfy $V_1, \dots, V_k \in (0, 2\pi)^{m-1} \times \{1\}^m$
- for any $i \in \{1, \dots, k\}$, the sub-hessian of h at V_i corresponding to the first $m-1$ arguments is non-degenerate: $\det G_i \neq 0$;
- for any $i \in \{1, \dots, k\}$ $\frac{\partial h(V_i)}{\partial x_j} \neq 0 \quad \text{for} \quad j = m, \dots, 2m-1.$

Conditions on the density

We consider beta distribution with density

$$p(x) = \frac{\beta + 1}{\pi} \cdot (1 - \|x\|^2)^{\beta} \cdot \mathbf{1}_{\mathbb{B}^2}(x).$$

Distributions in the ball

Limit distribution in the ball

Under the previous conditions on f and p for every $t > 0$ as $n \to \infty$,

$$\mathbb{P}\left[n^{\frac{m}{m(\beta+3/2)-1/2}}(M - H_n) \leq t\right] = \left(1 + O(n^{-\frac{1}{(2\beta+3)m-1}})\right)$$

$$\times\left(1 - \exp\left[-K_m \cdot I[V_1, \ldots, V_k] \cdot t^{m(\beta+3/2)-1/2}\right]\right),$$

where

$$I[V_1, \ldots, V_k] := \sum_{i=1}^{k} \frac{1}{\sqrt{\det(-G_i)}\prod_{j=1}^{m}\left(\frac{\partial h(V_i)}{\partial x_{m-1+j}}\right)^{\beta+1}},$$

$$K_m = \frac{2^{(\beta+1/2)m+1/2}\Gamma\left((\beta+\frac{3}{2})m+\frac{1}{2}\right)(\Gamma(\beta+2))^m}{\pi^{\frac{m-1}{2}} m!\,\Gamma\left(\frac{m+1}{2}\right)\Gamma\left((\beta+\frac{3}{2})m+\beta+\frac{3}{2}\right)}.$$

Distributions in the ball

Proposition 1

Under the previous conditions the following relation holds for $\varepsilon \to 0+$:

$$\mathbb{P}\left[f(U_1, \ldots, U_m) \geq M - \varepsilon\right] =$$
$$m! \cdot K_m \cdot I[V_1, \ldots, V_k] \cdot \varepsilon^{\beta m + m + \frac{m-1}{2}}(1 + O(\varepsilon)),$$

where K_m and $I[V_1, \ldots, V_k]$ are defined earlier.

Examples

Perimeter of convex hull

$$
\begin{aligned}
&h(\beta_1, \ldots, \beta_{m-1}, r_1, \ldots, r_m) \\
&= \sum_{i=1}^{m} \sqrt{r_{i+1}^2 + r_i^2 - 2 r_i r_{i+1} \cos\left(\beta_{i+2} - \beta_{i+1}\right)}, \\
&\frac{\partial h(V_i)}{\partial x_{m-1+j}} = 2 \sin\left(\frac{\pi}{m}\right) \text{ for all } \quad j = m, \ldots, 2m-1, \\
&\det(-G_i) = 2^{1-m} m \left(\sin\left(\frac{\pi}{m}\right)\right)^{m-1}.
\end{aligned}
$$

Examples

Area of convex hull

$$h(\beta_1, \ldots, \beta_{m-1}, r_1, \ldots, r_m) = \sum_{i=1}^{m} \frac{r_i r_{i+1} \sin\left(\beta_{i+2} - \beta_{i+1}\right)}{2},$$

$$\det(-G_i) = 2^{1-m} m \left(\sin\frac{2\pi}{m}\right)^{m-1} \text{for} \quad j = m, \ldots, 2m-1,$$

$$\frac{\partial h(V_i)}{\partial x_{m-1+j}} = \sin\frac{2\pi}{m}.$$

What are the further directions of research?

i) Non-bounded distribution of points;

i) Points on N-spheres, properties of inscribed and cirscumscribed polyhedra;

ii) Points on convex figures and bodies.

The end

Thank you for your attention!