

Motivation

In this poster we will discuss how concentration manifests in the n -dimensional analogues of cubes and spheres.

This topic falls under a more general area of study referred to as Concentration of Measures which was put forward by V. Milman whilst studying the asymptotic geometry of Banach Spaces. [1].

Applications

The study has given rise to Concentration Inequalities which have found applications ranging from probability theory to statistical physics as demonstrated by S. Chatterjee and P. Dey's paper [2].

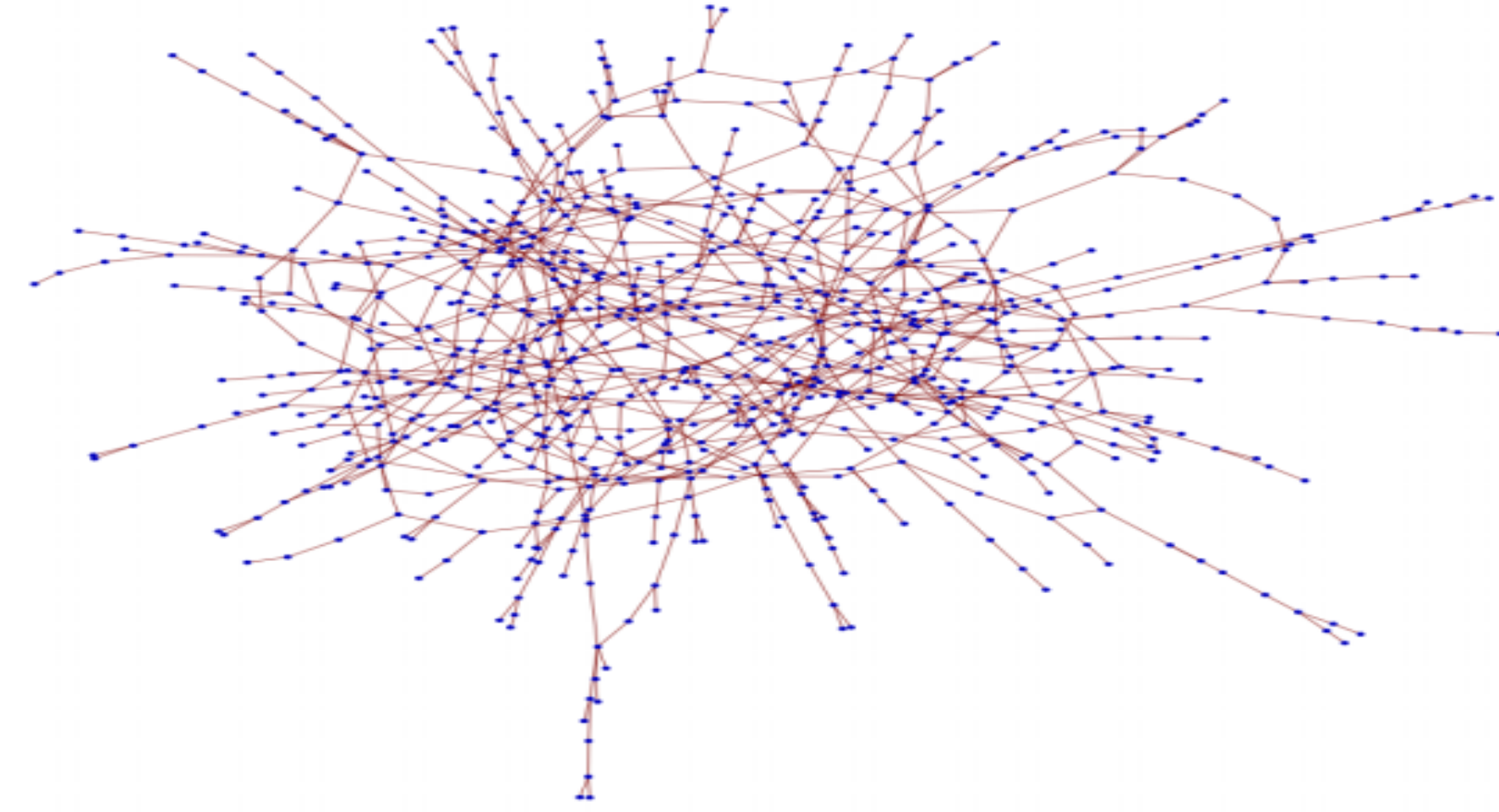


Figure 1: An Erdős-Rényi random graph [3]. Concentration inequalities can be used to study the number of triangles on such graphs.

Definitions

Uniformly distributed - A random variable (X_1, \dots, X_n) on $[-1, 1]^n$ is uniformly distributed if for a Borel measurable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\mathbb{E}(f(X_1, \dots, X_n)) = \frac{1}{2^n} \int_{-1}^1 \dots \int_{-1}^1 f(x_1, \dots, x_n) dx_1 \dots dx_n$$

Volume - Given $A \subseteq \mathbb{R}^n$

$$\text{Vol}(A) := \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} 1_A(x) dx_1 \dots dx_n$$

n -Cube - For notational ease we define

$$\mathbb{C}_n := [-1, 1]^n := [-1, 1] \times \dots \times [-1, 1]$$

n -Sphere -

$$\mathbb{S}^n(r) := \{x \in \mathbb{R}^{n+1} : \|x\| = r\}$$

Concentration of Volume in an n -Cube

Theorem - Almost all the volume of the n -Cube $[-1, 1]^n$ concentrates on the $(n-1)$ -Sphere with radius $\sqrt{n/3}$. Formally we want to show that:

$$\forall \varepsilon > 0, \text{Vol}(U_{n,\varepsilon} \cap [-1, 1]^n) \rightarrow \text{Vol}([-1, 1]^n)$$

$$U_{n,\varepsilon} := \left\{x : (1 - \varepsilon)\sqrt{n/3} < \|x\| < (1 + \varepsilon)\sqrt{n/3}\right\}$$

The first key observation is noticing the relationship between the volume of the ε -band around the boundary and the probability that a random point chosen uniformly from the cube falls inside the ε -band.

$$\begin{aligned} \text{Vol}(U_{n,\varepsilon} \cap [-1, 1]^n) / \text{Vol}([-1, 1]^n) &= \text{Vol}(U_{n,\varepsilon} \cap [-1, 1]^n) / 2^n \\ &= \frac{1}{2^n} \int_{-1}^1 \dots \int_{-1}^1 1_{U_{n,\varepsilon}}(x_1, \dots, x_n) dx_1 \dots dx_n = \mathbb{E}(1_{U_{n,\varepsilon}}(X_1, \dots, X_n)) = \mathbb{P}((X_1, \dots, X_n) \in U_{n,\varepsilon}) \end{aligned}$$

The next key observation is that if (X_1, \dots, X_n) is uniformly distributed on $[-1, 1]^n$ then $\{X_i\}$ are i.i.d. uniform distributions on $[-1, 1]$. This allows us to invoke the Weak Law of Large Numbers:

Let $S_n := X_1^2 + \dots + X_n^2$ (and by direct computation $\mathbb{E}(X_i^2) = 1/3$). Note: $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$

$$\left| \sqrt{\frac{3S_n}{n}} - 1 \right| = \left| \frac{\frac{3S_n}{n} - 1}{\sqrt{\frac{3S_n}{n} + 1}} \right| \leq \left| \frac{3S_n}{n} - 1 \right| < \varepsilon \Rightarrow \mathbb{P}\left(\left| \frac{3S_n}{n} - 1 \right| < \varepsilon\right) \leq \mathbb{P}\left(\left| \sqrt{\frac{3S_n}{n}} - 1 \right| < \varepsilon\right) \leq 1$$

By the Weak Law of Large Numbers the LHS converges to 1 and so by the Squeeze Theorem the middle converges to 1. Now observe that as $\varepsilon > 0$ was arbitrary that

$$\forall \varepsilon > 0, \mathbb{P}\left(\left| \sqrt{\frac{3S_n}{n}} - 1 \right| < \varepsilon\right) = \frac{\text{Vol}(U_{n,\varepsilon} \cap [-1, 1]^n)}{\text{Vol}([-1, 1]^n)} \rightarrow 1 \quad \square$$

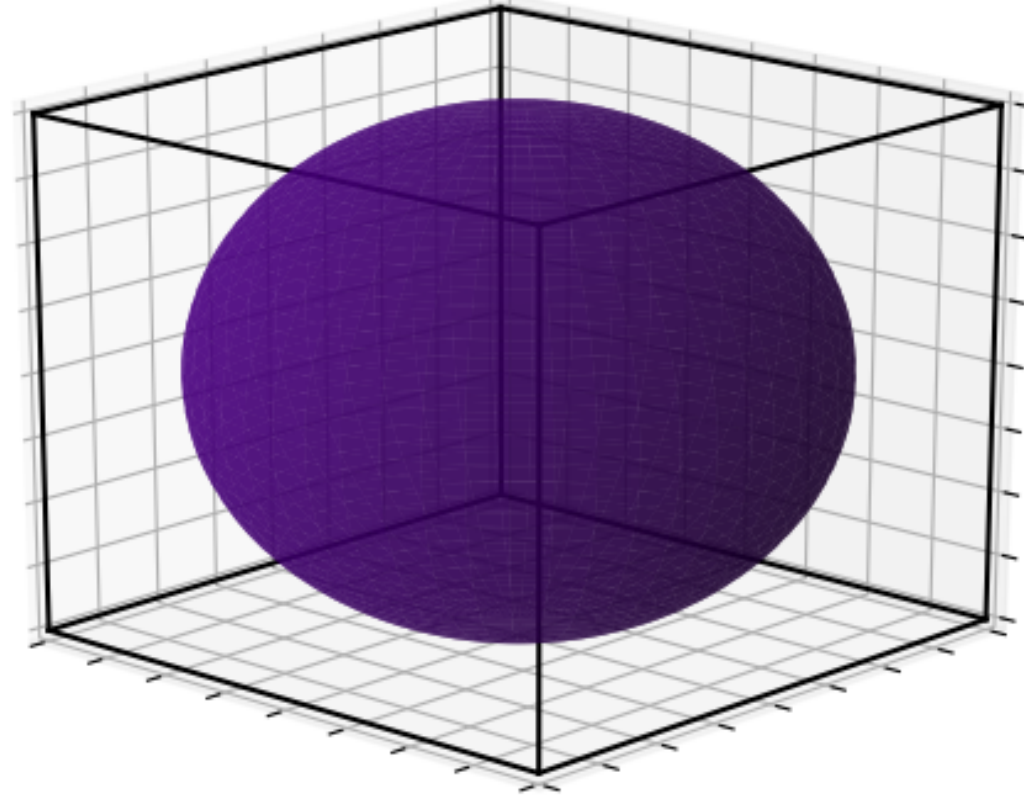


Figure 2: Visualisation for $n = 3$

Concentration of Area in an n -Sphere

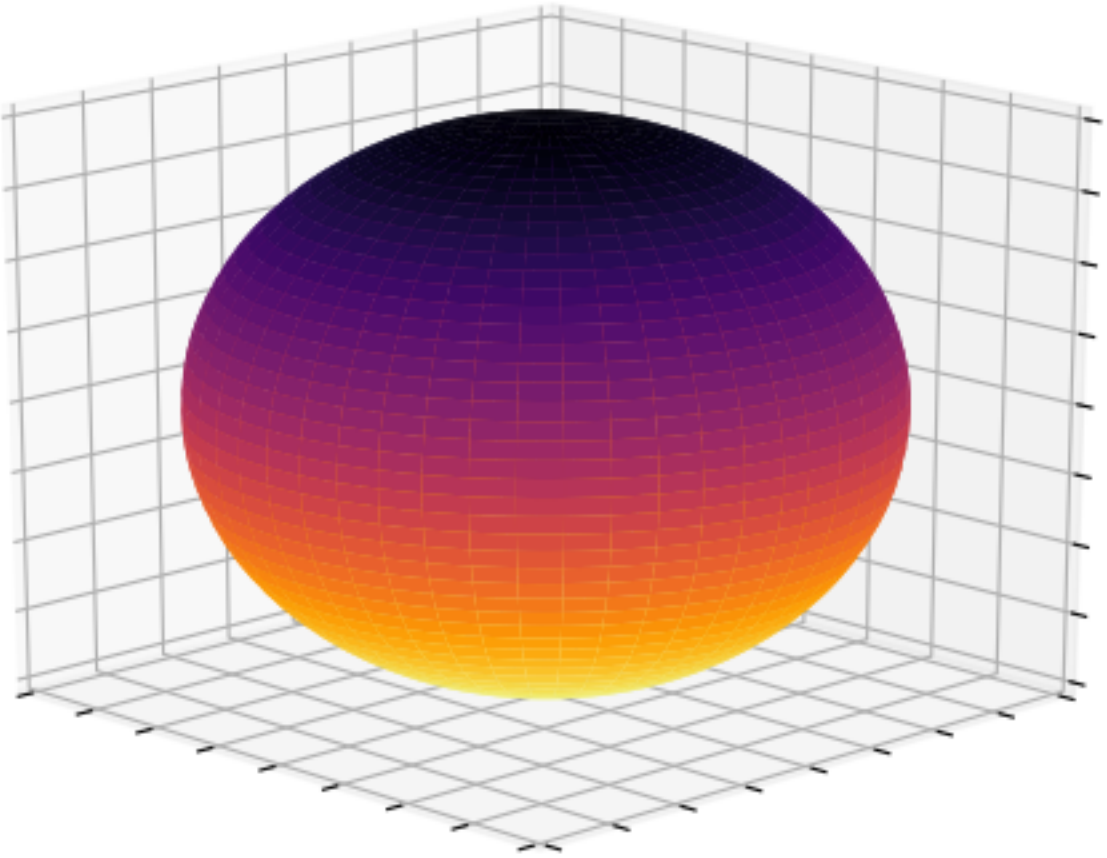


Figure 3: These are purely high dimensional phenomena and cannot be visualised.

Surprisingly, concentration occurs even when $\{X_i\}$ are not entirely independent. Points on the n -Sphere have $n+1$ coordinates and one is completely determined by the others as their norm must be 1. We will show that in an n -Sphere almost all the area concentrates by the equators.

Theorem - Let $A \subseteq \mathbb{S}^{n-1}$ be a measurable set with $\mathbb{P}(A) \geq 1/2$, and let $A_\varepsilon := \{x \in \mathbb{S}^{n-1} : \exists a \in A : \|x - a\| \leq \varepsilon\}$ then it follows that

$$\forall \varepsilon > 0, \mathbb{P}(A) \geq 1 - 2e^{-\varepsilon^2 n/4}$$

In other words, concentration manifests around any region that occupies at least half of the sphere for sufficiently high dimensions.

Corollary - If we consider A as the northern hemisphere and the southern hemisphere (both occupy half the area) then as concentration occurs at both then concentration must occur at their intersection: the equator.

The content of this section can be attributed to Ch. 14 of J. Matousek's Lectures on Discrete Geometry [4].

Simulation

By uniformly sampling points from the n -Cube, displaying their magnitudes into a histogram and superimposing vertical lines representing the boundaries of the ε -bands we can observe the concentration phenomenon.

Note that the figure below deceptively looks like a normal distribution and one might be misled by the central limit theorem as our coordinates are i.i.d. But we actually have the distribution of the square

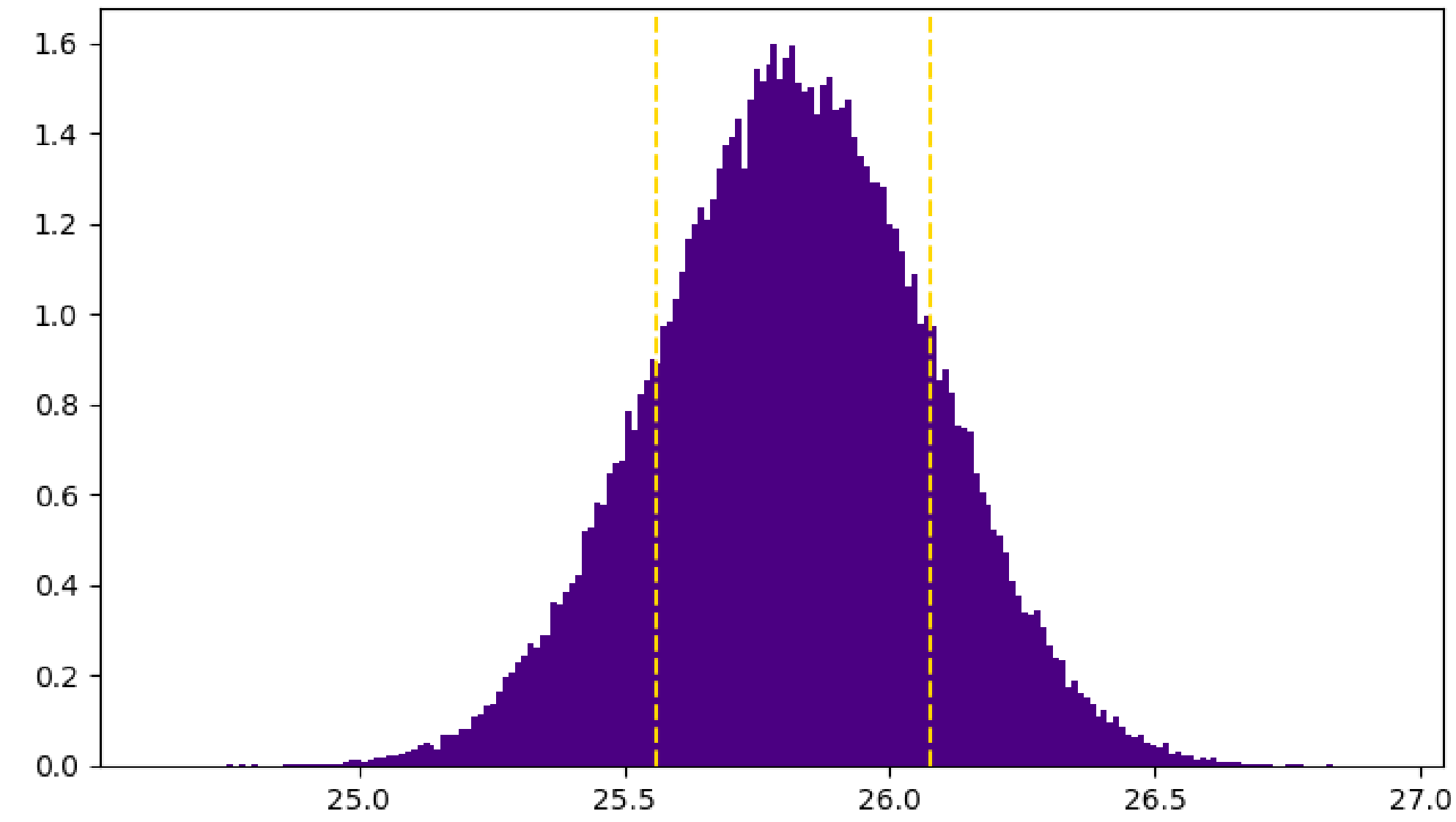


Figure 4: For $n = 2000$ and $\varepsilon = 0.01$ we compute that 68% of points fall inside the ε -band.

To see the code used to generate the above and relevant further resources use the following QR Code to my GitHub page.



References

- [1] M. Ledoux. Introduction. In: *The concentration of measure phenomenon*. Mathematical Surveys and Monographs no.89, American Mathematical Society; 2001. p. vii.
- [2] S. Chatterjee, P. Dey. Abstract. In: *Application of Stein's Method for Concentration Inequalities*. *The Annals of Probability* 2010; Vol. 38, No.6, 2443 - 2485. Available from: DOI: 10.1214/10-AOP542
- [3] O. Narayan, I. Saniee, G. Tucci. Figure 4. In: *Lack of Spectral Gap and Hyperbolicity in Asymptotic Erdős-Rényi Random Graphs*; 2010. p. 8.
- [4] J. Matousek. Ch 14. Measure Concentration and Almost Spherical Sections. In: *Lectures on Discrete Geometry*. Vol. 212. Springer Science Business Media; 2013. p. 299-302