

Article

# Linear and Fisher Separability of Random Points in the $d$ -Dimensional Spherical Layer and Inside the $d$ -Dimensional Cube

Sergey Sidorov  and Nikolai Zolotykh \* 

Institute of Information Technologies, Mathematics and Mechanics, Lobachevsky State University, 603950 Nizhni Novgorod, Russia; sergey.sidorov@itmm.unn.ru

\* Correspondence: nikolai.zolotykh@itmm.unn.ru

Received: 21 October 2020; Accepted: 10 November 2020; Published: 12 November 2020



**Abstract:** Stochastic separation theorems play important roles in high-dimensional data analysis and machine learning. It turns out that in high dimensional space, any point of a random set of points can be separated from other points by a hyperplane with high probability, even if the number of points is exponential in terms of dimensions. This and similar facts can be used for constructing correctors for artificial intelligent systems, for determining the intrinsic dimensionality of data and for explaining various natural intelligence phenomena. In this paper, we refine the estimations for the number of points and for the probability in stochastic separation theorems, thereby strengthening some results obtained earlier. We propose the boundaries for linear and Fisher separability, when the points are drawn randomly, independently and uniformly from a  $d$ -dimensional spherical layer and from the cube. These results allow us to better outline the applicability limits of the stochastic separation theorems in applications.

**Keywords:** stochastic separation theorems; random points; 1-convex set; linear separability; Fisher separability; Fisher linear discriminant

## 1. Introduction

It is generally accepted that the modern information world is the world of big data. However, some of the implications of the advent of the big data era remain poorly understood. In his “millennium lecture”, D. L. Donoho [1] described the post-classical world in which the number of features  $d$  is much greater than the sample size  $n$ :  $d \gg n$ . It turns out that many phenomena of the post-classical world are already observed if  $d \gg \log n$ , or, more precisely, when  $ID \gg \log n$ , where  $ID$  is the intrinsic dimensionality of the data [2]. Classical methods of data analysis and machine learning become of little use in such a situation, because usually they require huge amounts of data. Such an unlimited appetite of classical approaches for data is usually considered as a phenomenon of the “curse of dimensionality”. However, the properties  $ID \gg n$  or  $ID \gg \log n$  themselves are neither a curse nor a blessing, and can be beneficial.

One of the “post-classical” phenomena is stochastic separability [3–5]. If the dimensionality of data is high, then under broad assumptions any sample of the data set can be separated from the rest by a hyperplane (or even Fisher discriminant—as a special case) with a probability close to 1 even the number of samples is exponential in terms of dimensions. Thus, high-dimensional datasets exhibit fairly simple geometric properties.

Recently, stochastic separation theorems have been widely used in machine learning for constructing correctors and ensembles of correctors of artificial intelligence systems [6,7], for determining the intrinsic dimensionality of data sets [8,9], for explaining various natural intelligence phenomena, such as grandmother’s neuron [10,11].

In its usual form a stochastic separation theorem is formulated as follows. A random  $n$ -element set in  $\mathbb{R}^d$  is linearly separable with probability  $p > 1 - \vartheta$ , if  $n < ae^{bd}$ . The exact form of the exponential function depends on the probability distribution that determines how the random set is drawn, and on the constant  $\vartheta$  ( $0 < \vartheta < 1$ ). In particular, uniform distributions with different support are considered in [5,12–14]. Wider classes of distributions (including non-i.i.d.) are considered in [7]. Roughly speaking, these classes consist of distributions without sharp peaks in sets with exponentially small volume. Estimates for product distributions in the cube and the standard normal distribution are obtained in [15]. General stochastic separation theorems with optimal bounds for important classes of distributions (log-concave distribution, their convex combinations and product distributions) are proposed in [2].

We note that there are many algorithms for constructing a functional separating a point from all other points in a data set (Fisher linear discriminant, linear programming algorithm, support vector machine, Rosenblatt perceptron, etc.). Among all these methods the computationally cheapest is Fisher discriminant analysis [6]. Other advantages of the Fisher discriminant analysis are its simplicity and the robustness.

The papers [5–7,12] deal with only Fisher separability, whereas [13,14] considered a (more general) linear separability. A comparison of the estimations for linear and Fisher separability allows us to clarify the applicability boundary of these methods, namely, to answer the question of what  $d$  and  $n$  are sufficient in order to use only Fisher separability and so that there is no need to search a more sophisticated linear discriminant.

In [13,14], there were obtained estimates for the cardinality of the set of points that guarantee its linear separability when the points are drawn randomly, independently and uniformly from a  $d$ -dimensional spherical layer and from the unit cube. These results give more accurate estimates than the bounds obtained in [5,12] for Fisher separability.

Our interest in the study of the linear separability in spherical layers is explained, among other reasons, by the possibility of applying our results to determining the intrinsic dimension of data. After applying PCA to the data points for the selection of the major components and subsequent whitening we can map them to a spherical layer of a given thickness. If the intrinsic dimensionality of the initial set of  $n$  points is ID, then we expect that the separability properties of the resulting set of points are similar to the properties of uniformly distributed  $n$  points in dimension  $d$ . In particular, we can use the theoretical estimates for the separation probability to estimate ID (cf. [8,9]).

Here we give even more precise estimations for the number of points in the spherical layer to guarantee their linear separability. We also consider the case of linear separability of random points inside a cube in more detail than it was done in [13]. In particular, we give estimates for the probability of separability of one point. We also report results of computational experiments comparing the theoretical estimations for the probability of the linear and Fisher separabilities with the corresponding experimental frequencies and discuss them.

## 2. Definitions

A point  $X \in \mathbb{R}^d$  is linearly separable from a set  $M \subset \mathbb{R}^d$  if there exists a hyperplane separated  $X$  from  $M$ ; i.e., there exists  $A_X \in \mathbb{R}^d$  such that  $(A_X, X) > (A_X, Y)$  for all  $Y \in M$ .

A point  $X \in \mathbb{R}^d$  is Fisher separable from the set  $M \subset \mathbb{R}^d$  if  $(X, Y) < (X, X)$  for all  $Y \in M$  [6,7].

A set of points  $\{X_1, \dots, X_n\} \subset \mathbb{R}^d$  is called linearly separable [5] or 1-convex [3] if any point  $X_i$  is linearly separable from all other points in the set, or in other words, the set of vertices of their convex hull,  $\text{conv}(X_1, \dots, X_n)$ , coincides with  $\{X_1, \dots, X_n\}$ . The set  $\{X_1, \dots, X_n\}$  is called Fisher separable if  $(X_i, X_j) < (X_i, X_i)$  for all  $i, j$ , such that  $i \neq j$  [6,7].

Fisher separability implies linear separability but not vice versa (even if the set is centered and normalized to unit variance). Thus, if  $M \subset \mathbb{R}^d$  is a random set of points from a certain probability distribution, then the probability that  $M$  is linearly separable is not less than the probability that  $M$  is Fisher separable.

Denote by  $B_d = \{X \in \mathbb{R}^d : \|X\| \leq 1\}$  the  $d$ -dimensional unit ball centered at the origin ( $\|X\|$  means Euclidean norm),  $rB_d$  is the  $d$ -dimensional ball of radius  $r < 1$  centered at the origin and  $Q_d = [0, 1]^d$  is the  $d$ -dimensional unit cube.

Let  $M_n = \{X_1, \dots, X_n\}$  be the set of points chosen randomly, independently, according to the uniform distribution on the  $(1 - r)$ -thick spherical layer  $B_d \setminus rB_d$ , i.e., on the unit ball with spherical cavity of radius  $r$ . Denote by  $P^\circ(d, r, n)$  the probability that  $M_n$  is linearly separable, and by  $P^{\circ F}(d, r, n)$  the probability that  $M_n$  is Fisher separable. Denote by  $P_1^\circ(d, r, n)$  the probability that a random point chosen according to the uniform distribution on  $B_d \setminus rB_d$  is separable from  $M_n$ , and by  $P_1^{\circ F}(d, r, n)$  the probability that a random point is Fisher separable from  $M_n$ .

Now let  $M_n = \{X_1, \dots, X_n\}$  be the set of points chosen randomly, independently, according to the uniform distribution on the cube  $Q_d$ . Let  $P^\square(d, n)$  and  $P^{\square F}(d, n)$  denote the probabilities that  $M_n$  is linearly separable and Fisher separable, respectively. Let  $P_1^\square(d, n)$  and  $P_1^{\square F}(d, n)$  denote the probabilities that a random point chosen according to the uniform distribution on  $Q_d$  is separable and Fisher separable from  $M_n$ , respectively.

### 3. Previous Results

#### 3.1. Random Points in a Spherical Layer

In [5] it was shown (among other results) that for all  $r, \vartheta, n$  and  $d$ , where  $0 < r < 1, 0 < \vartheta < 1, d \in \mathbb{N}$ , if

$$n < \left(\frac{r}{\sqrt{1-r^2}}\right)^d \left(\sqrt{1 + \frac{2\vartheta(1-r^2)^{d/2}}{r^{2d}}} - 1\right), \tag{1}$$

then  $n$  points chosen randomly, independently, according to the uniform distribution on  $B_d \setminus rB_d$  are Fisher separable with a probability greater than  $1 - \vartheta$ , i.e.,  $P^{\circ F}(d, r, n) > 1 - \vartheta$ .

The following statements concerning the Fisher separability of random points in the spherical layer are proved in [12].

- For all  $r$ , where  $0 < r < 1$ , and for any  $d \in \mathbb{N}$

$$P_1^{\circ F}(d, r, n) > (1 - r^d) \left(1 - \frac{(1 - r^2)^{d/2}}{2}\right)^n. \tag{2}$$

- For all  $r, \vartheta$ , where  $0 < r < 1, 0 < \vartheta < 1$ , and for sufficiently large  $d$ , if

$$n < \frac{\vartheta}{(1 - r^2)^{d/2}}, \tag{3}$$

then  $P_1^{\circ F}(d, r, n) > 1 - \vartheta$ .

- For all  $r$ , where  $0 < r < 1$ , and for any  $d \in \mathbb{N}$

$$P^{\circ F}(d, r, n) > \left[ (1 - r^d) \left(1 - (n - 1) \frac{(1 - r^2)^{d/2}}{2}\right) \right]^n. \tag{4}$$

- For all  $r, \vartheta$ , where  $0 < r < 1, 0 < \vartheta < 1$  and for sufficiently large  $d$ , if

$$n < \frac{\sqrt{\vartheta}}{(1 - r^2)^{d/4}}, \tag{5}$$

then  $P^{\circ F}(d, r, n) > 1 - \vartheta$ .

The authors of [5,12] formulate their results for linearly separable sets of points, but in fact in the proofs they used that the sets are only Fisher separable.

Note that all estimates (1)–(5) require  $0 < r < 1$  with strong inequality. This means that they are inapplicable for (maybe the most interesting) case  $r = 0$ , i.e., for the unit ball with no cavities.

A reviewer of the original version of the article drew our attention that for  $r = 0$  better results are obtained in [6,15]. Specifically,

$$P_1^{\circ F}(d, 0, n) \geq 1 - \frac{n}{2^{d+1}}, \tag{6}$$

$$P^{\circ F}(d, 0, n) \geq 1 - \frac{n(n-1)}{2^{d+1}} > 1 - \frac{n^2}{2^{d+1}}, \tag{7}$$

and  $P_1^{\circ F}(d, 0, n) > 1 - \vartheta$  provided that  $n < \vartheta \cdot 2^{d+1}$ . See details in Section 4.4.

The both estimates (1) and (5) are exponentially dependent on  $d$  for fixed  $r, \vartheta$  and the estimate (1) is weaker than (5).

The following results concerning the linear separability of random points in the spherical layer were obtained in [14]:

- For all  $r$ , where  $0 \leq r < 1$ , and for any  $d \in \mathbb{N}$

$$P_1^{\circ}(d, r, n) > 1 - \frac{n}{2^d}. \tag{8}$$

- For all  $r, \vartheta$ , where  $0 \leq r < 1, 0 < \vartheta < 1$ , and for any  $d \in \mathbb{N}$ , if

$$n < \vartheta 2^d, \tag{9}$$

then  $P_1^{\circ}(d, r, n) > 1 - \vartheta$ .

- For all  $r$ , where  $0 \leq r < 1$ , and for any  $d \in \mathbb{N}$

$$P^{\circ}(d, r, n) > 1 - \frac{n(n-1)}{2^d}. \tag{10}$$

- For all  $r, \vartheta$ , where  $0 \leq r < 1, 0 < \vartheta < 1$ , and for any  $d$ , if

$$n < \sqrt{\vartheta 2^d}, \tag{11}$$

then  $P^{\circ}(d, r, n) > 1 - \vartheta$ .

We note that the bounds (8)–(11) do not depend on  $r$ . We remove this drawback in this paper, giving more accurate estimates (see Theorems 1 and 3 and Corollaries 1 and 2).

### 3.2. Random Points Inside a Cube

In [5], a product distribution in the  $Q_d$  is considered. Let the coordinates of a random point  $X = (x_1, \dots, x_d) \in Q_d$  be independent random variables with variances  $\sigma_i^2 > \sigma_0^2 > 0$  ( $i = 1, \dots, d$ ). In [5], it is shown that for all  $\vartheta$  and  $n$ , where  $0 < \vartheta < 1$ , if

$$n < \sqrt{\frac{\vartheta e^{0.5d\sigma_0^4}}{3}}, \tag{12}$$

then  $M_n$  is Fisher separable with a probability greater than  $1 - \vartheta$ . As above, the authors of [5] formulate their result for the linearly separable case, but in fact they used only the Fisher separability.

If all random variables  $x_1, \dots, x_d$  have the uniform distribution on the segment  $[0, 1]$  then  $\sigma_0^2 = \frac{1}{12}$ . Thus, the inequality (12) takes the form

$$n < \sqrt{\frac{\vartheta e^{d/288}}{3}}. \tag{13}$$

We obtain that if  $n$  satisfies (13), then  $P^{\square F}(d, n) > 1 - \vartheta$ .

In [13], it was shown that if we want to guarantee only the linear separability, then the bound (13) can be increased. Namely, if

$$n < \sqrt{\frac{\vartheta c^d}{d+1}}, \quad c = 1.18858,$$

then  $P^{\square}(d, n) > 1 - \vartheta$ . Here we give related estimates including ones for the linear separability of one point (see Theorems 5 and 6 and Corollary 3).

We note that better (and in fact asymptotically optimal) estimates for the Fisher separability in the unit cube are derived in [15]. The papers [13,15] were submitted to the same conference, so these results were derived in parallel and independently. Corollary 7 in [15] states that  $n$  points are Fisher separable with probability greater than  $1 - \vartheta$  provided only that  $n < \sqrt{\vartheta} e^{\gamma d}$  for  $\gamma = 0.23319\dots$  See details in Section 5.

#### 4. Random Points in a Spherical Layer

##### 4.1. The Separability of One Point

The theorem below gives the probability of the linear separability of a random point from a random  $n$ -element set  $M_n = \{X_1, \dots, X_n\}$  in  $B_d \setminus rB_d$ . The proof develops an approach borrowed from [3,16].

The regularized incomplete beta function is defined as  $I_x(a, b) = \frac{B(x; a, b)}{B(a, b)}$ , where

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt, \quad B(x; a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt$$

are beta function and incomplete beta function, respectively (see [17]).

**Theorem 1.** Let  $0 \leq r < 1$ ,  $\alpha = 4r^2(1 - r^2)$ ,  $\beta = 1 - r^2$ ,  $d \in \mathbb{N}$ . Then

(1) for  $0 \leq r \leq \frac{1}{\sqrt{2}}$

$$P_1^\circ(d, r, n) > 1 - n \cdot \frac{1 - 0.5 \left( I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right) + (2r)^d \cdot I_\beta\left(\frac{d+1}{2}, \frac{1}{2}\right) \right)}{2^d(1 - r^d)}; \tag{14}$$

(2) for  $\frac{1}{\sqrt{2}} \leq r < 1$

$$P_1^\circ(d, r, n) > 1 - n \cdot \frac{0.5 \left( I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right) - (2r)^d \cdot I_\beta\left(\frac{d+1}{2}, \frac{1}{2}\right) \right)}{2^d(1 - r^d)}. \tag{15}$$

**Proof.** A random point  $Y$  is linearly separable from  $M_n = \{X_1, \dots, X_n\}$  if and only if  $Y \notin \text{conv}(M_n)$ . Denote this event by  $C$ . Thus,  $P_1^\circ(d, r, n) = P(C)$ . Let us find the upper bound for the probability of the

event  $\bar{C}$ . This event means that the point  $Y$  belongs to the convex hull of  $M_n$ . Since the points in  $M_n$  have the uniform distribution, then the probability of  $\bar{C}$  is

$$P(\bar{C}) = \frac{\text{Vol}(\text{conv}(M_n) \setminus (\text{conv}(M_n) \cap rB_d))}{\text{Vol}(B_d) - \text{Vol}(rB_d)}.$$

First, estimate the numerator of this fraction. We denote by  $S_i$  the ball with center at the origin, with the diameter 1, and the point  $X_i$  lies on this diameter (see Figure 1). Then

$$\text{conv}(M_n) \setminus (\text{conv}(M_n) \cap rB_d) \subseteq \bigcup_{i=1}^n (S_i \setminus (S_i \cap rB_d)) = W$$

and

$$\begin{aligned} \text{Vol}(\text{conv}(M_n) \setminus (\text{conv}(M_n) \cap rB_d)) &\leq \text{Vol}(W) \leq \sum_{i=1}^n \text{Vol}(S_i \setminus (S_i \cap rB_d)) \\ &= \sum_{i=1}^n (\text{Vol}(S_i) - \text{Vol}(S_i \cap rB_d)) = n(\text{Vol}(S_1) - \text{Vol}(S_1 \cap rB_d)) \\ &= n \left( \gamma_d \left( \frac{1}{2} \right)^d - \text{Vol}(S_1 \cap rB_d) \right), \end{aligned}$$

where  $\gamma_d$  is the volume of a ball of radius 1. Hence

$$P(\bar{C}) \leq \frac{n \left( \gamma_d \left( \frac{1}{2} \right)^d - \text{Vol}(S_1 \cap rB_d) \right)}{\gamma_d(1 - r^d)}.$$

Now find  $\text{Vol}(S_1 \cap rB_d)$ . It is obvious that  $\text{Vol}(S_1 \cap rB_d)$  is equal to the sum of the volumes of two spherical caps. We denote by  $\text{Cap}(R, H)$  the volume of a spherical cap of height  $H$  of a ball of radius  $R$ . It is known [18] that

$$\text{Cap}(R, H) = \frac{1}{2} \gamma_d R^d I_{(2RH-H^2)/R^2} \left( \frac{d+1}{2}, \frac{1}{2} \right)$$

if  $0 \leq H \leq R$ .

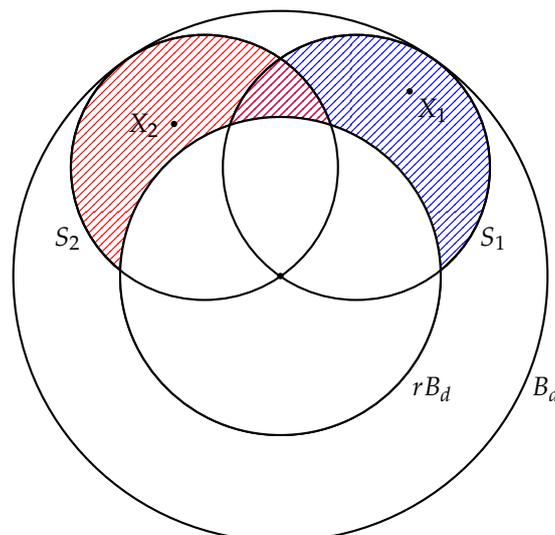


Figure 1. Illustration to the proof of Theorem 1.

Consider two cases:  $0 \leq r \leq \frac{1}{\sqrt{2}}$  and  $\frac{1}{\sqrt{2}} \leq r < 1$  (see Figure 2)

**Case 1** If  $0 \leq r \leq \frac{1}{\sqrt{2}}$ , then the centers of the balls  $S_1, S_2, \dots, S_n$  are inside of the spherical caps of height  $h$  of the ball  $rB_d$  (see the left picture on Figure 2). Therefore, the following equalities are true:

$$r^2 - (r - h)^2 = \left(\frac{1}{2}\right)^2 - \left(r - h - \frac{1}{2}\right)^2,$$

$$r^2 - (r - h)^2 = -(r - h)^2 + (r - h),$$

$$h = r - r^2,$$

$$V_1 = \text{Cap}\left(\frac{1}{2}, r - h\right) = \text{Cap}\left(\frac{1}{2}, r^2\right), \quad V_2 = \text{Cap}(r, h) = \text{Cap}(r, r - r^2).$$

If  $R = \frac{1}{2}, H = r^2$ , then  $(2RH - H^2)/R^2 = 4r^2(1 - r^2) = \alpha$ , hence

$$V_1 = \frac{1}{2} \gamma_d \left(\frac{1}{2}\right)^d I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right).$$

If  $R = r, H = r - r^2$ , then  $(2RH - H^2)/R^2 = 2H/R - (H/R)^2 = 2(1 - r) - (1 - r)^2 = 1 - r^2 = \beta$ , hence

$$V_2 = \frac{1}{2} \gamma_d r^d I_\beta\left(\frac{d+1}{2}, \frac{1}{2}\right).$$

Thus,

$$\text{Vol}(S_1 \cap rB_d) = V_1 + V_2 = \gamma_d \left(\frac{1}{2}\right)^d \left( I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right) + \frac{1}{2} r^d I_\beta\left(\frac{d+1}{2}, \frac{1}{2}\right) \right).$$

Hence

$$\begin{aligned} P(C) = 1 - P(\bar{C}) &\geq 1 - \frac{n \left( \gamma_d \left(\frac{1}{2}\right)^d - \text{Vol}(S_1 \cap rB_d) \right)}{\gamma_d (1 - r^d)} \\ &= 1 - n \cdot \frac{1 - 0.5 \left( I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right) + (2r)^d \cdot I_\beta\left(\frac{d+1}{2}, \frac{1}{2}\right) \right)}{2^d (1 - r^d)}. \end{aligned}$$

**Case 2** If  $\frac{1}{\sqrt{2}} \leq r < 1$ , then the centers of the balls  $S_1, S_2, \dots, S_n$  are outside of the spherical caps of height  $h$  of the ball  $rB_d$  (see the right picture on Figure 2). Therefore, the following equalities are true:

$$r^2 - (r - h)^2 = \left(\frac{1}{2}\right)^2 - \left(r - h - \frac{1}{2}\right)^2,$$

$$r^2 - (r - h)^2 = -(r - h)^2 + (r - h),$$

$$h = r - r^2,$$

$$V_1 = \text{Vol}\left(\frac{1}{2}B_d\right) - \text{Cap}\left(\frac{1}{2}, 1 - (r - h)\right) = \text{Vol}\left(\frac{1}{2}B_d\right) - \text{Cap}\left(\frac{1}{2}, 1 - r^2\right).$$

If  $R = \frac{1}{2}, H = 1 - r^2$ , then  $(2RH - H^2)/R^2 = 4r^2(1 - r^2)$ ; hence,

$$V_1 = \gamma_d \left(\frac{1}{2}\right)^d - \frac{1}{2} \gamma_d \left(\frac{1}{2}\right)^d I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right),$$

where  $\alpha = 4r^2(1 - r^2)$ ,

$$V_2 = \text{Cap}(r, h) = \text{Cap}(r, r - r^2) = \frac{1}{2} \gamma_d r^d I_\beta \left( \frac{d+1}{2}, \frac{1}{2} \right),$$

where  $\beta = 1 - r^2$ . Thus,

$$\text{Vol}(S_1 \cap rB_d) = V_1 + V_2 = \gamma_d \left( \left( \frac{1}{2} \right)^d - \frac{1}{2} \left( \frac{1}{2} \right)^d I_\alpha \left( \frac{d+1}{2}, \frac{1}{2} \right) + \frac{1}{2} r^d I_\beta \left( \frac{d+1}{2}, \frac{1}{2} \right) \right).$$

Hence

$$P(C) = 1 - P(\bar{C}) \geq 1 - \frac{n \left( \gamma_d \left( \frac{1}{2} \right)^d - \text{Vol}(S_1 \cap rB_d) \right)}{\gamma_d (1 - r^d)} = 1 - n \cdot \frac{0.5 \left( I_\alpha \left( \frac{d+1}{2}, \frac{1}{2} \right) - (2r)^d \cdot I_\beta \left( \frac{d+1}{2}, \frac{1}{2} \right) \right)}{2^d (1 - r^d)}.$$

□

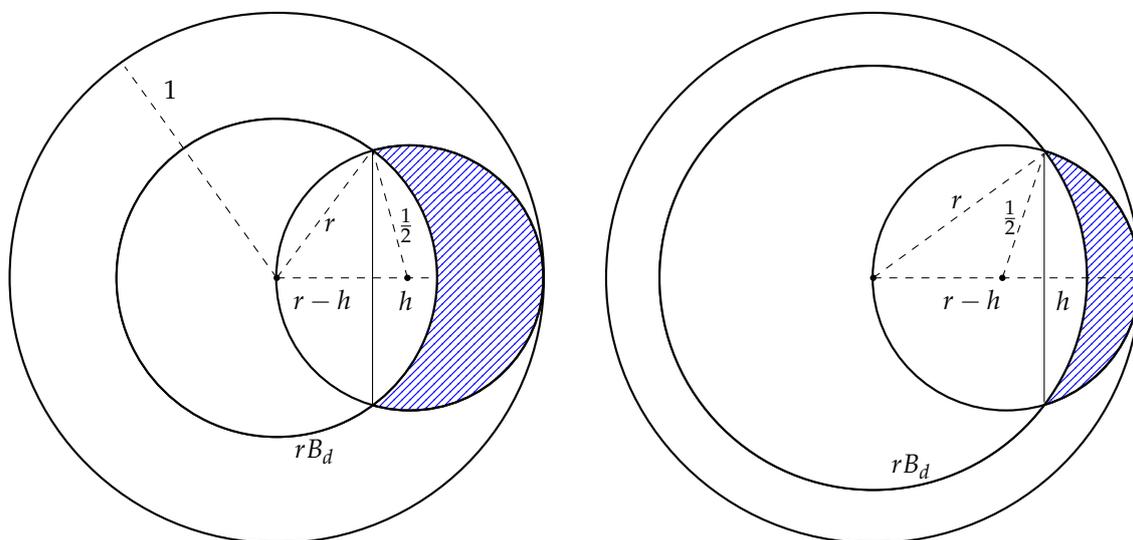


Figure 2. Illustration to the proof of Theorem 1: case 1 (left); case 2 (right).

The estimates (14) and (15) for  $P_1^\circ(d, r, n)$  are monotonically increasing in both  $d$  and  $r$  and decreasing in  $n$ , which corresponds to the behavior of the probability  $P_1^\circ(d, r, n)$  itself (see Figures 3 and 4). On the contrary, the estimate (3) for the probability  $P_1^{\circ F}(d, r, n)$  is nonmonotonic in  $r$  (see Figure 5).

Note that the estimates (14), (15) obtained in Theorem 1 are quite accurate (in the sense that they are close to empirical values), as is illustrated with Figure 4. The experiment also shows that the probabilities  $P_1^\circ(d, r, n)$  and  $P_1^{\circ F}(d, r, n)$  (more precisely, the corresponding frequencies) are quite close to each other, but there is a certain gap between them.

The following corollary gives an estimate for the number of points  $n$  guaranteeing the linear separability of a random point from a random  $n$ -element set  $M_n$  in  $B_d \setminus rB_d$  with probability close to 1.

**Corollary 1.** Let  $0 < \vartheta < 1$ ,  $\alpha = 4r^2(1 - r^2)$ ,  $\beta = 1 - r^2$ ,  $d \in \mathbb{N}$ . If

(1)

$$n < N_1(d, r, \vartheta) = \frac{\vartheta 2^d (1 - r^d)}{1 - 0.5 \left( I_\alpha \left( \frac{d+1}{2}, \frac{1}{2} \right) + (2r)^d \cdot I_\beta \left( \frac{d+1}{2}, \frac{1}{2} \right) \right)}, \quad 0 \leq r \leq \frac{1}{\sqrt{2}}$$

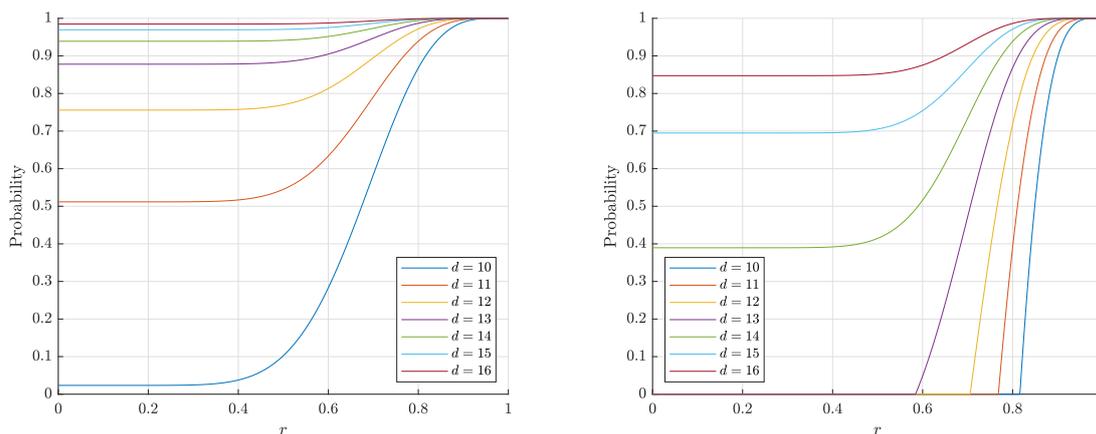
or

(2)

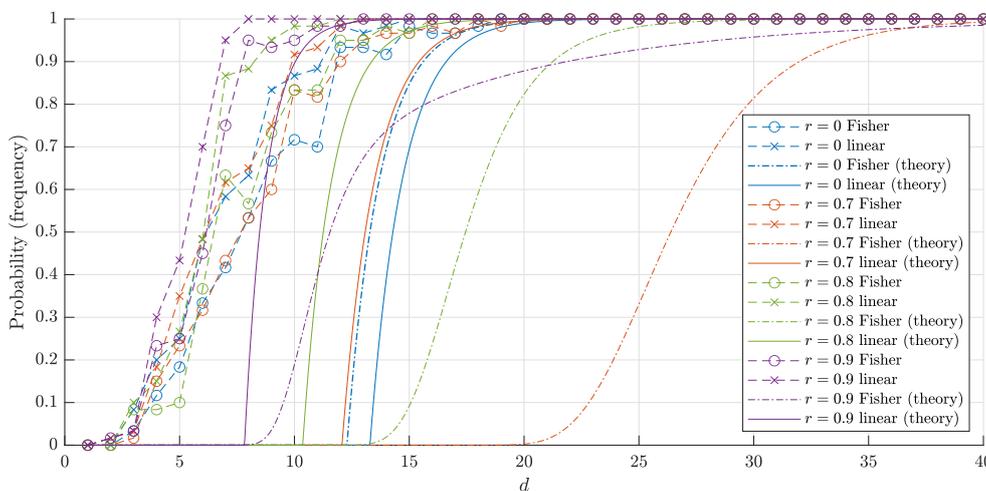
$$n < N_2(d, r, \vartheta) = \frac{\vartheta 2^d (1 - r^d)}{0.5 \left( I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right) - (2r)^d \cdot I_\beta\left(\frac{d+1}{2}, \frac{1}{2}\right) \right)}, \quad \frac{1}{\sqrt{2}} \leq r < 1,$$

then  $P_1^\circ(d, r, n) > 1 - \vartheta$ .

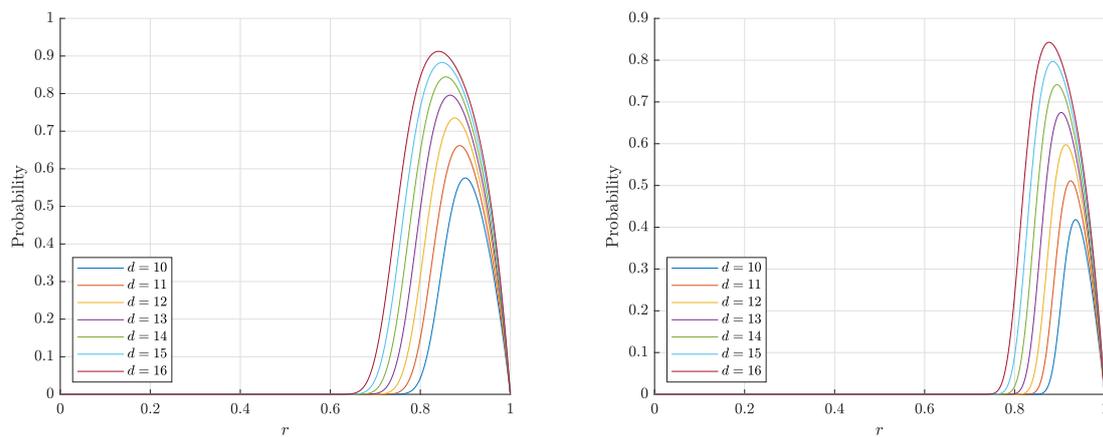
The theorem below establishes asymptotic estimates.



**Figure 3.** The graphs of the right-hand sides of the estimates (14), (15) for the probability  $P_1^\circ(d, r, n)$  that a random point is linear and separable from a set of  $n = 1000$  (left) and  $n = 10,000$  (right) random points in the layer  $B_d \setminus rB_d$ .



**Figure 4.** The graphs of the estimates for the probabilities  $P_1^\circ(d, r, n)$  ( $P_1^F(d, r, n)$ ) that a random point is linearly (and respectively, Fisher) separable from a set of  $n = 10,000$  random points in the layer  $B_d \setminus rB_d$ . The solid lines correspond to the theoretical bounds (14) and (15) for the linear separability. The dash-dotted lines represent the theoretical bounds (2) and (6) for the Fisher separability. The crosses (circles) correspond to the empirical frequencies for linear (and respectively Fisher) separability obtained in 60 trials for each dimension  $d$ .



**Figure 5.** The graphs of the right-hand side of the estimate (3) for the probability  $P_1^{\circ F}(d, r, n)$  that a random point is Fisher separable from a set of  $n = 1000$  (left) and  $n = 10,000$  (right) random points in the layer  $B_d \setminus rB_d$ .

**Theorem 2.**

(1) If  $0 \leq r < \frac{1}{\sqrt{2}}$  then

$$N_1(d, r, \vartheta) \sim \vartheta 2^d.$$

(2) If  $r = \frac{1}{\sqrt{2}}$  then

$$N_1(d, r, \vartheta) = N_2(d, r, \vartheta) \sim \vartheta 2^{d+1}.$$

(3) If  $\frac{1}{\sqrt{2}} < r < 1$  then

$$N_2(d, r, \vartheta) \sim \vartheta \sqrt{2\pi} \cdot \frac{r(2r^2 - 1)}{\sqrt{1 - r^2}} \cdot \sqrt{d + 1} \cdot \left( \frac{1}{r\sqrt{1 - r^2}} \right)^d.$$

**Proof.** The paper [19] gives the following asymptotic expansion for the incomplete beta function

$$B(x; a, b) \sim \frac{x^a}{a} \sum_{k=0}^{\infty} \frac{f_k(b, x)}{a^k} \quad \text{for } 0 \leq x < 1, a \rightarrow \infty$$

and

$$f_k(b, x) = \frac{d^k}{dw^k} \left[ (1 - xe^{-w})^{b-1} \right]_{w=0}.$$

Since  $f_0(b, x) = (1 - x)^{b-1}$  then

$$B(x; a, b) \sim \frac{x^a}{a} (1 - x)^{b-1} + \frac{x^a}{a} \sum_{k=1}^{\infty} \frac{f_k(b, x)}{a^k} \sim \frac{x^a}{a} (1 - x)^{b-1} \quad \text{for } b, x \text{ fixed, } a \rightarrow \infty.$$

Since  $B(a, b) \sim \frac{\Gamma(b)}{a^b}$  for  $b$  fixed and  $a \rightarrow \infty$ , then

$$I_x(a, b) = \frac{B(x; a, b)}{B(a, b)} \sim \frac{x^a (1 - x)^{b-1}}{a^{1-b} \Gamma(b)}$$

for  $b, x$  fixed and  $a \rightarrow \infty$ .

We have  $x = \alpha = 4r^2(1 - r^2)$  or  $x = \beta = 1 - r^2$  and  $a = \frac{d+1}{2}, b = \frac{1}{2}$ ; hence,

$$I_\alpha \left( \frac{d+1}{2}, \frac{1}{2} \right) \sim \frac{\sqrt{2}\alpha^{\frac{d+1}{2}}}{\sqrt{\pi}\sqrt{d+1}\sqrt{1-4r^2+4r^4}} = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{|1-2r^2|} \cdot \frac{\alpha^{\frac{d+1}{2}}}{\sqrt{d+1}}$$

$$(2r)^d I_\beta \left( \frac{d+1}{2}, \frac{1}{2} \right) \sim \frac{(2r)^d \sqrt{2}(\sqrt{1-r^2})^{d+1}}{r\sqrt{\pi}\sqrt{d+1}} = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{2r^2} \cdot \frac{\alpha^{\frac{d+1}{2}}}{\sqrt{d+1}}$$

If  $r = 0$ , then  $\alpha = 0, \beta = 1$ ; hence,  $N_1(d, r, \vartheta) \sim \vartheta 2^d$ .

If  $0 < r < \frac{1}{\sqrt{2}}$ , then  $0 < \alpha < 1$ ; hence,

$$I_\alpha \left( \frac{d+1}{2}, \frac{1}{2} \right) + (2r)^d I_\beta \left( \frac{d+1}{2}, \frac{1}{2} \right) \sim 0$$

and

$$N_1(d, r, \vartheta) \sim \vartheta 2^d.$$

If  $r = \frac{1}{\sqrt{2}}$ , then  $\alpha = 1, \beta = \frac{1}{2}$ ; hence,

$$N_1(d, r, \vartheta) = N_2(d, r, \vartheta) \sim \frac{\vartheta 2^d (1 - r^d)}{0.5 \left( 1 - \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{d+1}} \right)} \sim \vartheta 2^{d+1}.$$

If  $\frac{1}{\sqrt{2}} < r < 1$ , then  $0 < \alpha < 1$ ; hence,

$$I_\alpha \left( \frac{d+1}{2}, \frac{1}{2} \right) - (2r)^d I_\beta \left( \frac{d+1}{2}, \frac{1}{2} \right) \sim \sqrt{\frac{2}{\pi}} \cdot \frac{1}{2r^2(2r^2-1)} \cdot \frac{\alpha^{\frac{d+1}{2}}}{\sqrt{d+1}} = \sqrt{\frac{2}{\pi}} \cdot \frac{\sqrt{1-r^2}}{r(2r^2-1)} \cdot \frac{2^d (r\sqrt{1-r^2})^d}{\sqrt{d+1}}$$

and

$$N_2(d, r, \vartheta) = \frac{\vartheta 2^d (1 - r^d)}{0.5 \left( I_\alpha \left( \frac{d+1}{2}, \frac{1}{2} \right) - (2r)^d \cdot I_\beta \left( \frac{d+1}{2}, \frac{1}{2} \right) \right)} \sim \frac{\vartheta 2^d}{0.5 \sqrt{\frac{2}{\pi}} \cdot \frac{\sqrt{1-r^2}}{r(2r^2-1)} \cdot \frac{2^d (r\sqrt{1-r^2})^d}{\sqrt{d+1}}}$$

$$= \vartheta \sqrt{2\pi} \cdot \frac{r(2r^2-1)}{\sqrt{1-r^2}} \cdot \sqrt{d+1} \cdot \left( \frac{1}{r\sqrt{1-r^2}} \right)^d.$$

□

#### 4.2. Separability of a Set of Points

The theorem below gives the probability of the linear separability of a random  $n$ -element set  $M_n$  in  $B_d \setminus rB_d$ .

**Theorem 3.** Let  $0 \leq r < 1, \alpha = 4r^2(1 - r^2), \beta = 1 - r^2$  and  $d, n \in \mathbb{N}$ . Then

(1) for  $0 \leq r \leq \frac{1}{\sqrt{2}}$

$$P^\circ(d, r, n) > 1 - n(n-1) \cdot \frac{1 - 0.5 \left( I_\alpha \left( \frac{d+1}{2}, \frac{1}{2} \right) + (2r)^d \cdot I_\beta \left( \frac{d+1}{2}, \frac{1}{2} \right) \right)}{2^d (1 - r^d)}; \tag{16}$$

(2) for  $\frac{1}{\sqrt{2}} \leq r < 1$

$$P^\circ(d, r, n) > 1 - n(n-1) \cdot \frac{0.5 \left( I_\alpha \left( \frac{d+1}{2}, \frac{1}{2} \right) - (2r)^d \cdot I_\beta \left( \frac{d+1}{2}, \frac{1}{2} \right) \right)}{2^d (1 - r^d)}. \tag{17}$$

**Proof.** Denote by  $A_n$  the event that  $M_n$  is linearly separable and denote by  $C_i$  the event that  $X_i \notin \text{conv}(M_n \setminus \{X_i\})$  ( $i = 1, \dots, n$ ). Thus,  $P^\circ(d, r, n) = P(A_n)$ . Clearly,  $A_n = C_1 \cap \dots \cap C_n$  and  $P(A_n) = P(C_1 \cap \dots \cap C_n) = 1 - P(\overline{C}_1 \cup \dots \cup \overline{C}_n) \geq 1 - \sum_{i=1}^n P(\overline{C}_i)$ . Let us find an upper bound for the probability of the event  $\overline{C}_i$ . This event means that the point  $X_i$  belongs to the convex hull of the remaining points, i.e.,  $X_i \in \text{conv}(M_n \setminus \{X_i\})$ . In the proof of the previous theorem, it was shown that if  $0 \leq r \leq \frac{1}{\sqrt{2}}$ , then

$$P(\overline{C}_i) \leq (n - 1) \cdot \frac{1 - 0.5 \left( I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right) + (2r)^d \cdot I_\beta\left(\frac{d+1}{2}, \frac{1}{2}\right) \right)}{2^d(1 - r^d)} \quad (i = 1, \dots, n);$$

and if  $\frac{1}{\sqrt{2}} \leq r < 1$ , then

$$P(\overline{C}_i) \leq (n - 1) \cdot \frac{0.5 \left( I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right) - (2r)^d \cdot I_\beta\left(\frac{d+1}{2}, \frac{1}{2}\right) \right)}{2^d(1 - r^d)} \quad (i = 1, \dots, n).$$

Therefore, using the inequality

$$P(A_n) \geq 1 - \sum_{i=1}^n P(\overline{C}_i)$$

we obtain what is required.  $\square$

The graphs of the estimates (16), (17) and corresponding frequencies in 60 trials for  $n = 1000$  and  $n = 10,000$  points are shown in Figures 6 and 7, respectively. The experiment shows that our estimates are quite accurate and close to the corresponding frequencies.

Another important conclusion from the experiment is as follows. Despite the fact that the estimates for both probabilities  $P^{\circ F}(d, r, n)$  and  $P^\circ(d, r, n)$  and corresponding frequencies are close to 1 for sufficiently big  $d$ , the "threshold values" for such a big  $d$  differ greatly. In other words, the blessing of dimensionality when using linear discriminants comes noticeably earlier than if we only use Fisher discriminants. This is achieved at the cost of constructing the usual linear discriminant in comparison with the Fisher one.

The following corollary gives an estimate for the number of points  $n$  guaranteeing the linear separability of a random  $n$ -element set  $M_n$  in  $B_d \setminus rB_d$  with probability close to 1.

**Corollary 2.** Let  $0 < \vartheta < 1$ ,  $\alpha = 4r^2(1 - r^2)$ ,  $\beta = 1 - r^2$ . If

(1)

$$0 \leq r \leq \frac{1}{\sqrt{2}} \quad \text{and} \quad n < \sqrt{N_1(d, r, \vartheta)} = \sqrt{\frac{\vartheta 2^d(1 - r^d)}{1 - 0.5 \left( I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right) + (2r)^d \cdot I_\beta\left(\frac{d+1}{2}, \frac{1}{2}\right) \right)}}$$

or

(2)

$$\frac{1}{\sqrt{2}} \leq r < 1 \quad \text{and} \quad n < \sqrt{N_2(d, r, \vartheta)} = \sqrt{\frac{\vartheta 2^d(1 - r^d)}{0.5 \left( I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right) - (2r)^d \cdot I_\beta\left(\frac{d+1}{2}, \frac{1}{2}\right) \right)'}}$$

then  $P^\circ(d, r, n) > 1 - \vartheta$ .

The theorem below establishes asymptotic estimates for the number of points guaranteeing the linear separability with probability greater than  $1 - \vartheta$ .

**Theorem 4.**

(1) If  $0 \leq r < \frac{1}{\sqrt{2}}$  then

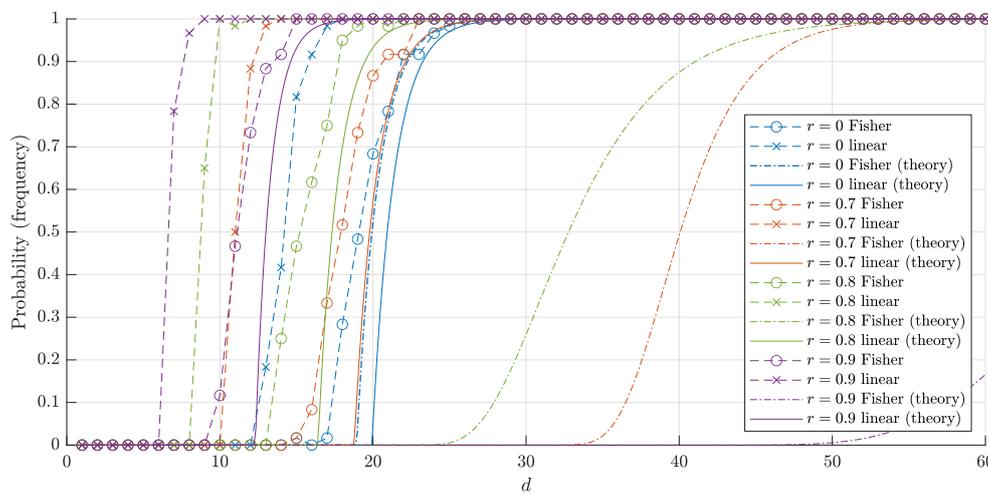
$$\sqrt{N_1(d, r, \vartheta)} \sim \sqrt{\vartheta} 2^{d/2}.$$

(2) If  $r = \frac{1}{\sqrt{2}}$  then

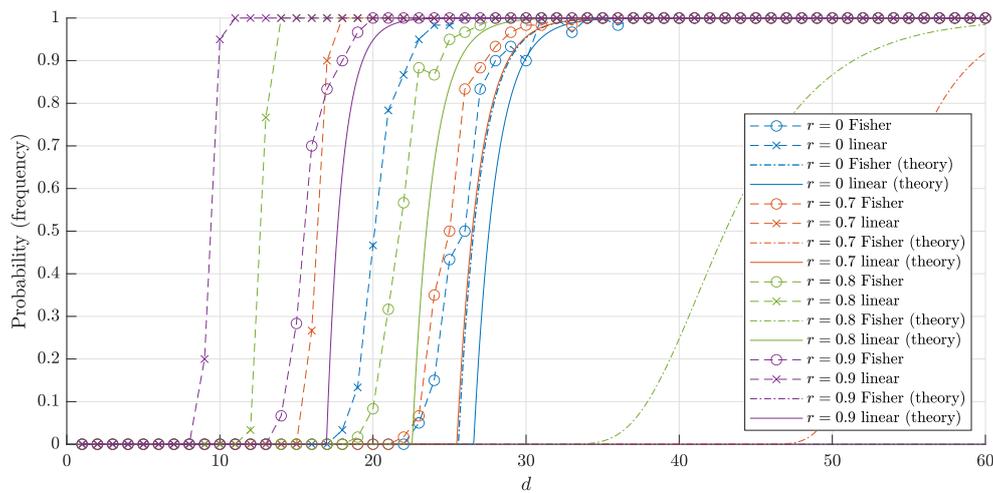
$$\sqrt{N_1(d, r, \vartheta)} = \sqrt{N_2(d, r, \vartheta)} \sim \sqrt{\vartheta} 2^{(d+1)/2}.$$

(3) If  $\frac{1}{\sqrt{2}} < r < 1$  then

$$\sqrt{N_2(d, r, \vartheta)} \sim \sqrt{\vartheta} \sqrt[4]{2\pi} \cdot \frac{\sqrt{r(2r^2 - 1)}}{\sqrt[4]{1 - r^2}} \cdot \sqrt[4]{d + 1} \cdot \left( \frac{1}{r\sqrt{1 - r^2}} \right)^{d/2}.$$



**Figure 6.** The graphs of the estimates for the probabilities  $P^\circ(d, r, n)$  ( $P^{\circ F}(d, r, n)$ ) that a random set of  $n = 1000$  points in  $B_d \setminus rB_d$  is linearly (and respectively Fisher) separable. The solid lines correspond to the theoretical bounds (16) and (17) for the linear separability. The dash-dotted lines represent the theoretical bound (4) and (7) for the Fisher separability. The crosses (circles) correspond to the empirical frequencies for linear (and respectively, Fisher) separability obtained in 60 trials for each dimension  $d$ .



**Figure 7.** The graphs of the estimates for the probabilities  $P^\circ(d, r, n)$  ( $P^{\circ F}(d, r, n)$ ) that a random set of  $n = 10,000$  points in  $B_d \setminus rB_d$  is linearly (and respectively, Fisher) separable. The notation is the same as in Figure 6.

4.3. Comparison of the Results

Let us show that the new estimates (16) and (17) for linear separability tend to be 1 faster than the estimate (4) in [12] for Fisher separability.

**Statement 1.** Let  $0 < r < 1$ ,  $\alpha = 4r^2(1 - r^2)$ ,  $\beta = 1 - r^2$  and  $d, n \in \mathbb{N}$ ,

$$f_1 = n(n - 1) \cdot \frac{1 - 0.5 \left( I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right) + (2r)^d \cdot I_\beta\left(\frac{d+1}{2}, \frac{1}{2}\right) \right)}{2^d(1 - r^d)},$$

$$f_2 = n(n - 1) \cdot \frac{0.5 \left( I_\alpha\left(\frac{d+1}{2}, \frac{1}{2}\right) - (2r)^d \cdot I_\beta\left(\frac{d+1}{2}, \frac{1}{2}\right) \right)}{2^d(1 - r^d)},$$

$$g = 1 - \left[ (1 - r^d) \left( 1 - (n - 1) \frac{(1 - r^2)^{d/2}}{2} \right) \right]^n.$$

For  $r$  and  $n$  fixed

(1) if  $0 < r < \frac{1}{\sqrt{2}}$ , then

$$\frac{g}{f_1} \sim \frac{1}{2} (4 - 4r^2)^{d/2} \rightarrow \infty;$$

(2) if  $r = \frac{1}{\sqrt{2}}$ , then

$$\frac{g}{f_1} = \frac{g}{f_2} \sim \frac{n + 1}{n - 1} \cdot 2^{d/2} \rightarrow \infty;$$

(3) if  $\frac{1}{\sqrt{2}} < r < 1$ , then

$$\frac{g}{f_2} \sim \sqrt{2\pi} \cdot \frac{r(2r^2 - 1)}{(n - 1)\sqrt{1 - r^2}} \cdot \sqrt{d + 1} \cdot \left( \frac{1}{1 - r^2} \right)^{d/2} \rightarrow \infty.$$

**Proof.** If  $0 < r < \frac{1}{\sqrt{2}}$ , then  $g \sim \frac{n(n-1)}{2}(1-r^2)^{d/2}$  and  $f_1 \sim \frac{n(n-1)}{2^d}$  (see the proof of Theorem 2); hence,

$$\frac{g}{f_1} \sim \frac{\frac{n(n-1)}{2}(1-r^2)^{d/2}}{\frac{n(n-1)}{2^d}} = \frac{1}{2}(4-4r^2)^{d/2} \rightarrow \infty, \quad \text{as } 4-4r^2 > 2.$$

If  $r = \frac{1}{\sqrt{2}}$ , then  $g \sim \frac{n(n+1)}{2} \frac{1}{2^{d/2}}$  and  $f_1 = f_2 \sim \frac{n(n-1)}{2^{d+1}}$  (see the proof of Theorem 2); hence,

$$\frac{g}{f_1} = \frac{g}{f_2} \sim \frac{\frac{n(n+1)}{2} \frac{1}{2^{d/2}}}{\frac{n(n-1)}{2^{d+1}}} = \frac{n+1}{n-1} \cdot 2^{d/2} \rightarrow \infty.$$

If  $\frac{1}{\sqrt{2}} < r < 1$ , then  $g \sim nr^d$  and  $f_2 \sim \frac{n(n-1)}{\sqrt{2\pi} \cdot \frac{r(2r^2-1)}{\sqrt{1-r^2}} \cdot \sqrt{d+1} \cdot \left(\frac{1}{r\sqrt{1-r^2}}\right)^d}$  (see the proof of Theorem 2), hence

$$\frac{g}{f_2} \sim \frac{nr^d \sqrt{2\pi} \cdot \frac{r(2r^2-1)}{\sqrt{1-r^2}} \cdot \sqrt{d+1} \cdot \left(\frac{1}{r\sqrt{1-r^2}}\right)^d}{n(n-1)} = \sqrt{2\pi} \cdot \frac{r(2r^2-1)}{(n-1)\sqrt{1-r^2}} \cdot \sqrt{d+1} \cdot \left(\frac{1}{1-r^2}\right)^{d/2} \rightarrow \infty.$$

□

Now let us compare the estimates for the number of points that guarantee the linear and Fisher separabilities of random points in the spherical layer obtained in Corollary 2 and in [12], respectively. The estimate in Corollary 2 for the number of points guaranteeing the linear separability tends to  $\infty$  faster than the estimate (5), guaranteeing the Fisher separability for all  $0 < r < 1$ .

**Statement 2.** Let  $f_1 = \sqrt{N_1(d, r, \vartheta)}$ ,  $f_2 = \sqrt{N_2(d, r, \vartheta)}$ ,  $g = \frac{\sqrt{\vartheta}}{(1-r^2)^{d/4}}$ ,  $0 < r < 1, 0 < \vartheta < 1, d \in \mathbb{N}$ . For  $r$  and  $\vartheta$  fixed

(1) if  $0 < r < \frac{1}{\sqrt{2}}$ , then

$$\frac{f_1}{g} \sim (2\sqrt{1-r^2})^{d/2} \rightarrow \infty;$$

(2) if  $r = \frac{1}{\sqrt{2}}$ , then

$$\frac{f_1}{g} = \frac{f_2}{g} \sim 2^{(d+2)/4} \rightarrow \infty;$$

(3) if  $\frac{1}{\sqrt{2}} < r < 1$ , then  $\frac{f_2}{g} \sim \sqrt{\sqrt{2\pi} \cdot \frac{r(2r^2-1)}{\sqrt{1-r^2}} \cdot (d+1)^{1/4} \cdot \left(\frac{1}{r}\right)^{d/2}} \rightarrow \infty$ .

**Proof.** If  $0 < r < \frac{1}{\sqrt{2}}$  then  $\frac{f_1}{g} \sim \frac{\sqrt{\vartheta 2^d (1-r^2)^{d/4}}}{\sqrt{\vartheta}} = (2\sqrt{1-r^2})^{d/2}$ .

If  $r = \frac{1}{\sqrt{2}}$ , then  $f_1 = f_2 \sim \sqrt{\vartheta 2^{d+1}}$  and  $g = \sqrt{\vartheta 2^{d/4}}$ ; hence,  $\frac{f_1}{g} = \frac{f_2}{g} \sim \frac{\sqrt{\vartheta 2^{d+1}}}{\sqrt{\vartheta 2^{d/4}}} = 2^{(d+2)/4}$ .

If  $\frac{1}{\sqrt{2}} < r < 1$ , then  $f_2 \sim \sqrt{\vartheta \sqrt{2\pi} \cdot \frac{r(2r^2-1)}{\sqrt{1-r^2}} \cdot \sqrt{d+1} \cdot \left(\frac{1}{r\sqrt{1-r^2}}\right)^d}$ ; hence,

$$\begin{aligned} \frac{f_2}{g} &\sim \sqrt{\vartheta \sqrt{2\pi} \cdot \frac{r(2r^2-1)}{\sqrt{1-r^2}} \cdot (d+1)^{1/4} \cdot \left(\frac{1}{r^2(1-r^2)}\right)^{d/4} \frac{(1-r^2)^{d/4}}{\sqrt{\vartheta}}} \\ &= \sqrt{\sqrt{2\pi} \cdot \frac{r(2r^2-1)}{\sqrt{1-r^2}} \cdot (d+1)^{1/4} \cdot \left(\frac{1}{r}\right)^{d/2}}. \end{aligned}$$

□

4.4. A Note about Random Points Inside the Ball ( $r = 0$ )

A reviewer of the original version of the article drew our attention to the fact that for the uniform distribution inside the ball (case  $r = 0$ ), better results are known. Specifically, let  $\bar{p}_{xy}^F$  be the probability that i.i.d. points  $x, y$  inside the ball are not Fisher separable. Let  $I_{xy}$  be the indicator function of this event. Then

$$\bar{p}_{xy}^F = E[I_{xy}] = E[E[I_{xy} | y]] = E[\bar{p}_y],$$

where  $\bar{p}_y$  denotes the probability that  $x$  is not Fisher separable from a given point  $y$ . In [6] (also discussed in [15]), there is a proof that  $E[\bar{p}_y] = 1/2^{d+1}$ . In the notation of our paper, this implies that

$$P_1^{\circ F}(d, 0, n) \geq 1 - \frac{n}{2^{d+1}}, \quad P^{\circ F}(d, 0, n) \geq 1 - \frac{n(n-1)}{2^{d+1}} > 1 - \frac{n^2}{2^{d+1}},$$

and  $P_1^{\circ F}(d, 0, n) > 1 - \vartheta$  provided that  $n < \vartheta \cdot 2^{d+1}$ . This improves the estimate in Theorem 2 for the case  $r = 0$  twice. Note that the same estimate  $n < \vartheta \cdot 2^{d+1}$  was derived for  $r = \frac{1}{\sqrt{2}}$  (see Theorem 2). The reviewer conjectured that estimate  $n < \vartheta \cdot 2^d$  derived in this paper could be improved twice for the whole range  $r \in [0, \frac{1}{\sqrt{2}})$ . The experimental results give support for this hypothesis (see Figures 4–7).

5. Random Points Inside a Cube

Consider a set of points  $M_n = \{X_1, \dots, X_n\}$  choosing randomly, independently and according to the uniform distribution on the  $d$ -dimensional unit cube  $Q_d$ .

**Theorem 5.** Let  $d, n \in \mathbb{N}$ . Then

$$P_1^\square(d, n) > 1 - \frac{n(d+1)}{c^d}, \quad c = 1.18858 \dots \tag{18}$$

**Proof.** A random point  $Y$  is linearly separable from  $M_n = \{X_1, \dots, X_n\}$  if and only if  $Y \notin \text{conv}(M_n)$ . Denote this event by  $C$ . Thus,  $P_1^\square(d, n) = P(C)$ . Let us find the upper bound for the probability of the event  $\bar{C}$ . This event means that the point  $Y$  belongs to the convex hull of  $M_n$ . Since the points in  $M_n$  have the uniform distribution, the probability of  $\bar{C}$  is

$$P(\bar{C}) = \frac{\text{Vol}(\text{conv}(M_n))}{\text{Vol}(Q_d)} = \text{Vol}(\text{conv}(M_n)).$$

In [20] it is proved that the upper bound for the maximal volume of the convex hull of  $k$  points placed in  $Q_d$  is  $\frac{k(d+1)}{c^d}$ , where  $c = 1.18858$ . Thus,  $\text{Vol}(\text{conv}(Y_1, \dots, Y_k)) < \frac{k(d+1)}{c^d}$  so

$$P(\bar{C}) = \text{Vol}(\text{conv}(M_n)) < \frac{n(d+1)}{c^d}.$$

and

$$P_1^\square(d, n) = P(C) = 1 - P(\bar{C}) > 1 - \frac{n(d+1)}{c^d}.$$

□

**Corollary 3.** Let  $0 < \vartheta < 1$ ,

$$n < \frac{\vartheta c^d}{d+1}, \quad c = 1.18858 \dots \tag{19}$$

Then  $P_1^\square(d, n) > 1 - \vartheta$ .

**Theorem 6.** Let  $d, n \in \mathbb{N}$ . Then

$$P^\square(d, n) > 1 - \frac{n(n-1)(d+1)}{c^d}, \quad c = 1.18858. \tag{20}$$

**Proof.** Denote by  $A_n$  the event that  $M_n$  is linearly separable and denote by  $C_i$  the event that  $X_i \notin \text{conv}(M_n \setminus \{X_i\})$  ( $i = 1, \dots, n$ ). Thus,  $P^\square(d, n) = P(A_n)$ . Clearly  $A_n = C_1 \cap \dots \cap C_n$  and  $P(A_n) = P(C_1 \cap \dots \cap C_n) = 1 - P(\bar{C}_1 \cup \dots \cup \bar{C}_n) \geq 1 - \sum_{i=1}^n P(\bar{C}_i)$ . Let us find the upper bound for the probability of the event  $\bar{C}_i$ . This event means that the point  $X_i$  belongs to the convex hull of the remaining points, i.e.,  $X_i \in \text{conv}(M_n \setminus \{X_i\})$ . In the proof of the previous theorem, it was shown that

$$P(\bar{C}_i) \leq \frac{(n-1)(d+1)}{c^d}, \quad c = 1.18858 \quad (i = 1, \dots, n).$$

Hence

$$P(A_n) \geq 1 - \sum_{i=1}^n P(\bar{C}_i) \geq 1 - \frac{n(n-1)(d+1)}{c^d}.$$

□

**Corollary 4.** [13] Let  $0 < \vartheta < 1$ ,

$$n < \sqrt{\frac{\vartheta c^d}{d+1}}, \quad c = 1.18858. \tag{21}$$

Then  $P^\square(d, n) > 1 - \vartheta$ .

We note that the estimate (21) for the number of points guaranteeing the linear separability tends to be  $\infty$  faster than the estimate (13), guaranteeing the Fisher separability because

$$\frac{\sqrt{\frac{\vartheta c^d}{d+1}}}{\sqrt{\frac{\vartheta e^{d/288}}{3}}} = \sqrt{\frac{3}{d+1} \left(\frac{c}{e^{1/288}}\right)^d} \rightarrow \infty, \quad \text{as } d \rightarrow \infty,$$

since  $c/e^{1/288} \approx 1.18446$ .

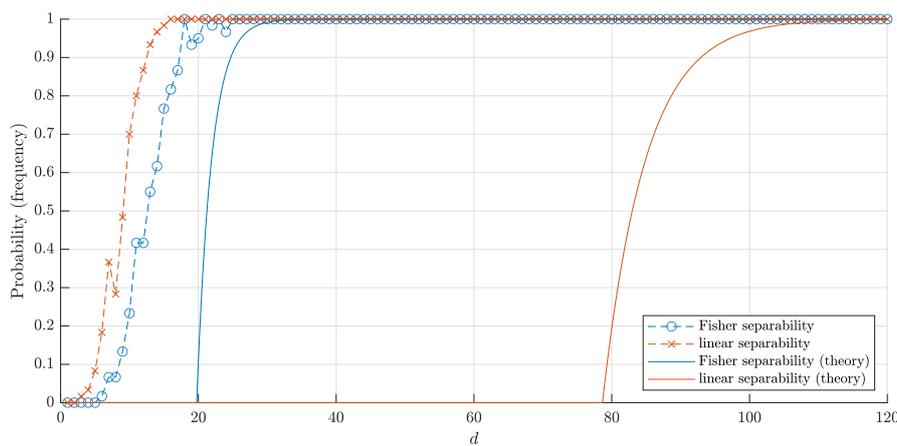
However better (and in fact asymptotically optimal) estimates for the Fisher separability in the unit cube are derived in [15]. Corollary 7 in [15] states that  $n$  points are Fisher separable with probability greater than  $1 - \vartheta$  provided only that  $n < \sqrt{\vartheta} e^{\gamma d}$  for  $\gamma = 0.23319 \dots$ . This can be written as  $n < \sqrt{\vartheta} c^d$  for  $c = e^{2\gamma} = 1.59421 \dots$ . Thus,

$$P_1^{\square F}(d, n) > 1 - \frac{n}{\exp(2\gamma d)} = 1 - \frac{n}{c^d}, \tag{22}$$

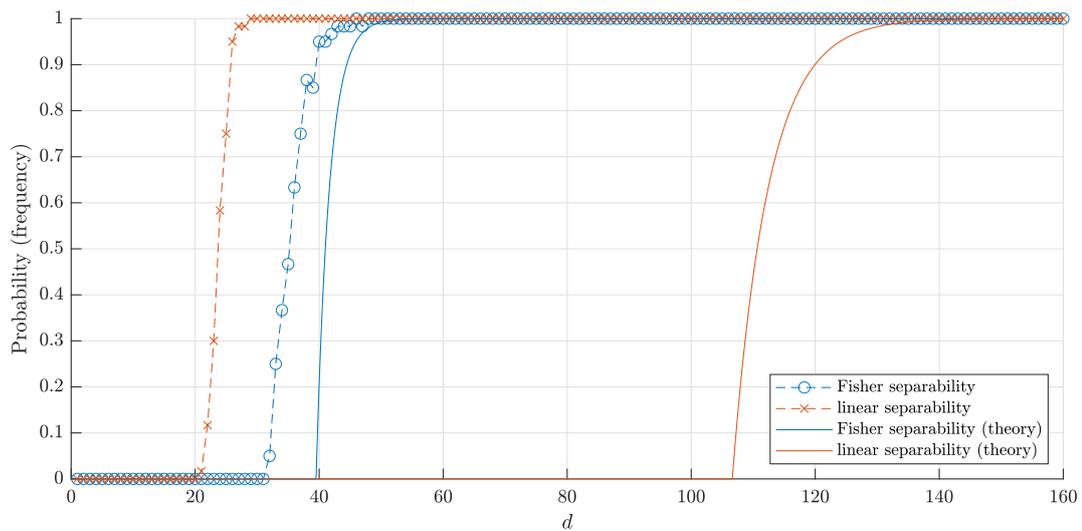
$$P^{\square F}(d, n) > 1 - \frac{n^2}{c^d}. \tag{23}$$

Theorem 6 and Corollary 4 in our paper state the same results with  $c = 1.18858 \dots$ , and for just linear separability instead of Fisher separability. However, [13,15] were submitted to the same conference, so these results were derived in parallel and independently.

The bounds (18) and (20) for the probabilities and corresponding frequencies are presented in Figures 8 and 9.



**Figure 8.** The graphs of the estimate for the probabilities  $P_1^\square(d, n)$  and  $P_1^{\square F}(d, n)$  that a random point is linearly (Fisher) separable from a set of  $n = 10,000$  random points inside the cube  $Q_d$ . The solid red and blue lines correspond to the theoretical bounds (18) and (22) respectively. Red crosses (blue circles) correspond to the empirical frequencies for linear (and respectively, Fisher) separability obtained in 60 trials for each dimension  $d$ .



**Figure 9.** The graphs of the estimates (20) and (23) for the probabilities  $P^\square(d, n)$  and  $P^{\square F}(d, n)$  that a set of  $n = 10,000$  random points inside the unit cube  $Q_d$  is linear and Fisher separable, respectively. The notation is the same as in Figure 8.

### 6. Subsequent Work

In a recent paper [2], explicit and asymptotically optimal estimates of Fisher separation probabilities for spherically invariant distribution (e.g., the standard normal and the uniform distributions) were obtained. Theorem 14 in [2] generalizes the results presented here. Since [2] was submitted to the arxiv later, we did not compare the results of that article with our results.

### 7. Conclusions

In this paper we refined the estimates for the number of points and for the probability in stochastic separation theorems. We gave new bounds for linear separability, when the points are drawn randomly, independently and uniformly from a  $d$ -dimensional spherical layer or from the unit cube. These results

refine some results obtained in [5,12–14] and allow us to better understand the applicability limits of the stochastic separation theorems for high-dimensional data mining and machine learning problems.

The strongest progress was in the estimation for the number of random points in a  $(1 - r)$ -thick spherical layer  $B_d \setminus rB_d$  that are linear separable with high probability. If

$$n \lesssim \sqrt{\vartheta} 2^{d/2}, \quad 0 \leq r < \frac{1}{\sqrt{2}} \quad \text{or} \quad n \lesssim \sqrt{\vartheta} 2^{(d+1)/2}, \quad r = \frac{1}{\sqrt{2}}$$

or

$$n \lesssim \sqrt{\vartheta} \sqrt[4]{2\pi} \cdot \frac{\sqrt{r(2r^2 - 1)}}{\sqrt[4]{1 - r^2}} \cdot \sqrt[4]{d + 1} \cdot \left( \frac{1}{r\sqrt{1 - r^2}} \right)^{d/2}, \quad \frac{1}{\sqrt{2}} < r < 1,$$

then  $n$  i.i.d. random points inside the spherical layer  $B_d \setminus rB_d$  are linear separable with probability at least  $1 - \vartheta$  (the asymptotic inequalities are for  $d \rightarrow \infty$ ).

One of the main results of the experiment comparing linear and Fisher separabilities is as follows. The blessing of dimensionality when using linear discriminants can come noticeably earlier (for smaller values of  $d$ ) than if we only use Fisher discriminants. This is achieved at the cost of constructing the usual linear discriminant in comparison with the Fisher one.

**Author Contributions:** Conceptualization, S.S. and N.Z.; methodology, S.S. and N.Z.; software, S.S. and N.Z.; validation, S.S. and N.Z.; formal analysis, S.S. and N.Z.; investigation, S.S. and N.Z.; resources, S.S. and N.Z.; data curation, S.S. and N.Z.; writing—original draft preparation, S.S. and N.Z.; writing—review and editing, S.S. and N.Z.; visualization, S.S. and N.Z.; supervision, S.S. and N.Z.; project administration, S.S. and N.Z.; funding acquisition, S.S. and N.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work is supported by the Ministry of Science and Higher Education of the Russian Federation (agreement number 075-15-2020-808).

**Acknowledgments:** The authors are grateful to anonymous reviewers for valuable comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Donoho, D.L. High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality. Invited Lecture at Mathematical Challenges of the 21st Century. In Proceedings of the AMS National Meeting, Los Angeles, CA, USA, 6–12 August 2000. Available online: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.329.3392> (accessed on 9 November 2020).
2. Grechuk, B.; Gorban, A.N.; Tyukin, I.Y. General stochastic separation theorems with optimal bounds. *arXiv* **2020**, arXiv:2010.05241.
3. Bárány, I.; Füredi, Z. On the shape of the convex hull of random points. *Probab. Theory Relat. Fields* **1988**, *77*, 231–240. [[CrossRef](#)]
4. Donoho, D.; Tanner, J. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos. Trans. R. Soc. A* **2009**, *367*, 4273–4293. [[CrossRef](#)] [[PubMed](#)]
5. Gorban, A.N.; Tyukin, I.Y. Stochastic separation theorems. *Neural Netw.* **2017**, *94*, 255–259. [[CrossRef](#)] [[PubMed](#)]
6. Gorban, A.N.; Golubkov, A.; Grechuk, B.; Mirkes, E.M.; Tyukin, I.Y. Correction of AI systems by linear discriminants: Probabilistic foundations. *Inf. Sci.* **2018**, *466*, 303–322. [[CrossRef](#)]
7. Gorban, A.N.; Grechuk, B.; Tyukin, I.Y. Augmented artificial intelligence: A conceptual framework. *arXiv* **2018**, arXiv:1802.02172v3.
8. Albergante, L.; Bac, J.; Zinovyev, A. Estimating the effective dimension of large biological datasets using Fisher separability analysis. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019.
9. Bac, J.; Zinovyev, A. Lizard brain: tackling locally low-dimensional yet globally complex organization of multi-dimensional datasets. *Front. Neurobot.* **2020**, *13*, 110. [[CrossRef](#)] [[PubMed](#)]
10. Gorban, A.N.; Makarov, V.A.; Tyukin, I.Y. The unreasonable effectiveness of small neural ensembles in high-dimensional brain. *Phys. Life Rev.* **2019**, *29*, 55–88. [[CrossRef](#)] [[PubMed](#)]

11. Gorban, A.N.; Makarov, V.A.; Tyukin, I.Y. High-Dimensional Brain in a High-Dimensional World: Blessing of Dimensionality. *Entropy* **2020**, *22*, 82. [[CrossRef](#)]
12. Gorban, A.N.; Burton, R.; Romanenko, I.; Tyukin, I.Y. One-trial correction of legacy AI systems and stochastic separation theorems. *Inf. Sci.* **2019**, *484*, 237–254. [[CrossRef](#)]
13. Sidorov, S.V.; Zolotykh, N.Y. On the Linear Separability of Random Points in the  $d$ -dimensional Spherical Layer and in the  $d$ -dimensional Cube. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–4. [[CrossRef](#)]
14. Sidorov, S.V.; Zolotykh, N.Y. Linear and Fisher Separability of Random Points in the  $d$ -dimensional Spherical Layer. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–6. [[CrossRef](#)]
15. Grechuk, B. Practical stochastic separation theorems for product distributions. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8. [[CrossRef](#)]
16. Elekes, G. A geometric inequality and the complexity of computing volume. *Discret. Comput. Geom.* **1986**, *1*, 289–292. [[CrossRef](#)]
17. Paris, R.B. Incomplete beta functions. In *NIST Handbook of Mathematical Functions*; Olver, F.W., Lozier, D.W., Boisvert, R.F., Clark, C.W., Eds.; Cambridge University Press: Cambridge, UK, 2010.
18. Li, S. Concise Formulas for the Area and Volume of a Hyperspherical Cap. *Asian J. Math. Stat.* **2011**, *4*, 66–70. [[CrossRef](#)]
19. López, J.L.; Sesma, J. Asymptotic expansion of the incomplete beta function for large values of the first parameter. *Integral Transform. Spec. Funct.* **1999**, *8*, 233–236 [[CrossRef](#)]
20. Dyer, M.E.; Füredi, Z.; McDiarmid, C. Random points in the  $n$ -cube. *DIMACS Ser. Discret. Math. Theor. Comput. Sci.* **1990**, *1*, 33–38.

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).