New Instability Results for High Dimensional Nearest Neighbor Search

Chris Giannella[a]

[a] *Dept. of Computer Science, New Mexico State Univ., Las Cruces NM, cgiannel@acm.org*

Abstract

Consider a dataset of $n(d)$ points generated independently from $\mathbb{R}^d$ according to a common p.d.f. f_d with $support(f_d) = [0,1]^d$ and $sup\{f_d([0,1]^d)\}$ growing sub-exponentially in d. We prove that: (i) if $n(d)$ grows sub-exponentially in d, then, for any query point $\vec{q}^d \in [0,1]^d$ and any $\epsilon > 0$, the ratio of the distance between any two dataset points and $\vec{q}^d$ is less that $1+\epsilon$ with probability $\to 1$ as $d \to \infty$; (ii) if $n(d) > [4(1+\epsilon)]^d$ for large d, then for all $\vec{q}^d \in [0,1]^d$ (except a small subset) and any $\epsilon > 0$, the distance ratio is less than $1+\epsilon$ with limiting probability strictly bounded away from one. Moreover, we provide preliminary results along the lines of (i) when $f_d = N(\vec{\mu}_d, \Sigma_d)$.

Key words: information retrieval, curse of dimensionality

1. Introduction

Nearest neighbor search on high-dimensional data is a difficult (and well-studied) problem, in part, because many commonly used distance functions can exhibit greatly different behavior in low versus high-dimensional spaces – a phenomenon often referred to as the "curse of dimensionality". In an effort to rigorously analyze this phenomenon, Beyer *et al.* [3] defined a nearest neighbor query with respect to a reference query point $\vec{q}^d \in \mathbb{R}^d$ as *unstable* if all of the points in the dataset are nearly the same distance from $\vec{q}^d$. In this event, the query can be thought meaningless since there is little reason to return any one point over another (see figure 2 in [3]). Beyer *et al.* (then later others [4], [11]) established sufficient conditions on the data generation distributions and dataset sizes under which the probability of query instability approaches one as $d \to \infty$. Such conditions provide useful insight into how the curse can be mitigated or must be tolerated as unavoidable. We develop a new set of sufficient conditions which improve upon the current ones – see sub-sections 1.2 and 1.3 for a description of our contributions and their relationship to the literature.

1.1. Notations and Definitions

Given $n(.) : \mathbb{N} \to \mathbb{N}$, we represent a d-dimensional, size $n(d)$ dataset with i.i.d. random vectors $\vec{Y}_1, \ldots, \vec{Y}_{n(d)}$ having common p.d.f. f_d. Let $support(f_d)$ denote the topological closure of $\{\vec{y} \in \mathbb{R}^d : f_d(\vec{y}) > 0\}$. Given posative real number p, the distance between a pair of points $\vec{z}, \vec{w} \in \mathbb{R}^d$ is defined as: $||\vec{z} - \vec{w}||_p = \left[\sum_{j=1}^{d} |z_j - w_j|^p\right]^{1/p}$. Given $\epsilon > 0$, the probability of a nearest neighbor query $\vec{q}^d \in support(f_d)$ being unstable is $P_{d,n(.),\vec{q}^d}$ $= Pr\left[\max_{i=1}^{n(d)}\{||\vec{Y}_i - \vec{q}^d||_p\} \leq (1+\epsilon)\min_{i=1}^{n(d)}\{||\vec{Y}_i - \vec{q}^d||_p\}\right]$.

The space of all possible query point sequences is $\prod_{d=1}^{\infty} support(f_d)$. We say that data distribution sequence $\{f_d : d = 1, 2, \cdots\}$ and dataset size function $n(.)$ admit nearest neighbor instability if for any $\epsilon > 0$ and any query point sequence $\{\vec{q}^d\} \in \prod_{d=1}^{\infty} support(f_d)$, it is the case that $lim_{d\to\infty} P_{d,n(.),\vec{q}^d} = 1$. We say that $\{f_d\}$ and $n(.)$ strongly fail to admit nearest neighbor instability if there exists $\zeta < 1$ and a "large" $\mathcal{Q} \subseteq \prod_{d=1}^{\infty} support(f_d)$, such that for any $\epsilon > 0$ and for any $\{\vec{q}^d\} \in \mathcal{Q}$, it is the case that $lim_{d\to\infty} P_{d,n(.),\vec{q}^d} < \zeta$. Let $\mathcal{Q}^d$ denote the d^{th} component of $\mathcal{Q}$. We say that $\mathcal{Q}$ is "large" if for any $0 \leq \omega < 1$, it is the case that, $lim_{d\to\infty} \frac{\omega^d Volume(support(f_d))}{Volume(\mathcal{Q}^d)} = 0$. Note, if $support(f_d) = [0,1]^d$, this last condition is equivalent to $lim_{d\to\infty} \frac{Volume([0,\omega]^d)}{Volume(\mathcal{Q}^d)} = 0$.

A function $g : \mathbb{N} \to \mathbb{N}$ is said to grow sub-exponentially if $lim_{d\to\infty} \frac{log(g(d))}{d} = 0$. A sequence of functions, $f_d : \mathbb{R}^d \to \mathbb{R}$; $d = 1, 2, \ldots$, is said to be bounded above sub-exponentially if, for all d, $sup\{f_d(\mathbb{R}^d)\} \leq g(d)$.

1.2. Our Contributions

For any $\{f_d\}$ bounded above sub-exponentially and $support(f_d) = [0,1]^d$, we prove the following: (i) if $n(.)$ grows sub-exponentially, then nearest neighbor instability is admitted; (ii) if $n(d) > [4(1+\epsilon)]^d$ for large d, then (with $p \geq 1$) instability strongly fails to be admitted. Moreover, we describe preliminary results toward establishing sufficient conditions under which $\{N(\vec{\mu}_d, \Sigma_d)\}$ admits instability.

1.3. Related Work

Beyer *et al.* [3] established sufficient conditions upon $n(.)$ and $\{f_d\}$ for the admission of nearest neighbor instability. They proved[1] that instability is admitted if $n(.)$ is constant and $\{f_d\}$ satisfies: $lim_{d\to\infty} Var\left[\frac{||\vec{Y}_1-\vec{q}^d||_p}{E[||\vec{Y}_1-\vec{q}^d||_p]}\right] = 0$, for any $\{\vec{q}^d\}$ (the *relative variance* goes to zero). Pestov [11], proved[2] that (Corollary 5.5) instability is admitted (except for a small set of query point sequences) if $n(.)$ is sub-exponentially growing and $\{f_d\}$ satisfies three conditions, most notably, $\{f_d\}$ forms a *normal Levy family* as defined with respect to the "concentration of measure" phenomena. Francois *et al.* [4] proved that instability is admitted (with $\{\vec{q}^d\} = \{\vec{0}\}$) if $n(.)$ is constant and each distribution in $\{f_d\}$ has i.i.d. attributes with mean and variance not dependent on d.

Our contributions significantly advance the above results as follows. Our sufficient conditions allow $n(.)$ to grow with d (unlike Beyer *et al.* and Francois *et al.*), are quite broad (unlike Francois *et al.* who require the data distributions to have i.i.d. attributes), and are easy to interpret (unlike Beyer *et al.* or Pestov *et al.* which leave open the question of which data distribution sequences satisfy the relative variance condition or normal Levy condition, respectively). Moreover, we provide results showing that the sub-exponential growth assumption on $n(.)$ is strongly necessary: if $n(.)$ grows exponentially, then instability fails to be admitted for a large space of query point sequences. Finally, we provide preliminary results toward establishing sufficient conditions for $\{N(\vec{\mu}_d, \Sigma_d)\}$. To our knowledge, the sufficient conditions for this distribution sequence remain unknown.

Aggarwal *et al.* [2] considered distance functions with p a positive integer and proved that, for constant $n(.) = N$ and data distributions with i.i.d. attributes supported on $(0,1)$, $C_p \le lim_{d\to\infty} \frac{E\left[max_{i=1}^{N}||\vec{Y}_i||_p - min_{i=1}^{N}||\vec{Y}_i||_p\right]}{d^{1/p-1/2}} \le (N-1)C_p$, with C_p a constant not dependent on d. They argued that high-dimensional nearest neighbor behavior is sharply different for each of the following three types of distance functions: $p = 1$, $p = 2$, and $p \ge 3$. However, unlike our contributions, they do not provide sufficient conditions on instability and they make the restrictive i.i.d. data attribute assumption. Hsu and Chen [7] proved[3] that, for constant $n(.)$, the relative variance condition of Beyer is a *necessary*

[1] They considered any non-negative distance function and did not restrict query points to reside in $support(f_d)$.

[2] He considered any metric distance function.

[3] They consider any non-negative distance function.

as well as a sufficient condition for instability admission. They go on to develop a basis for empirically testing whether instability is exhibited.

Shaft and Ramakrishnan [12] considered the related problem of analytically quantifying the inherent limits of nearest-neighbor indexing on high-dimensional data. They proved that, under conditions related to those in Beyer *et al.*, the performance of a broad class of index structures approaches that of linear scan as $d \to \infty$. In the stochastic geometry literature, Zanger [13] studied the behavior of a general class of clustering functions as $d \to \infty$ and established a connection to the concentration of measure phenomenon. A more broadly studied problem in this literature is the behavior of nearest neighbor structures as the *dataset size* goes to infinity and d remains constant. For example, Penrose [10] considered data generated i.i.d. from a continuous p.d.f. with compact support (and "smooth" boundary) and showed that, as $N \to \infty$, the distance of any point to its k nearest neighbor converges, almost surely, to a constant not dependent on N.

A vast literature exists on the development of data structures and algorithms for nearest neighbor search, for brevity, see the discussion and citations in [7].

2. Instability Results

First we develop a lower-bound on $P_{d,n(.),\vec{q}^d}$ making no assumptions on $\{f_d\}$ or $n(.)$. Define $\delta(\epsilon, p) = [(1+\epsilon)^p - 1]/[(1+\epsilon)^p + 1]$ and let $\gamma \geq 0$. If for all $1 \leq i \leq n(d)$, $\left| ||\vec{Y}_i - \vec{q}^d||_p^p - \gamma \right| \leq \gamma\delta(\epsilon, p)$, then $\max_{i=1}^{n(d)}\{||\vec{Y}_i - \vec{q}^d||_p^p\} \leq \min_{i=1}^{n(d)}\{||\vec{Y}_i - \vec{q}^d||_p^p\}\frac{[1+\delta(\epsilon,p)]}{[1-\delta(\epsilon,p)]} = \min_{i=1}^{n(d)}\{||\vec{Y}_i - \vec{q}^d||_p^p\}(1+\epsilon)^p$. Thus, $\max_{i=1}^{n(d)}\{||\vec{Y}_i - \vec{q}^d||_p\} \leq \min_{i=1}^{n(d)}\{||\vec{Y}_i - \vec{q}^d||_p\}(1+\epsilon)$. Using this and the fact that $\vec{Y}_1, \ldots \vec{Y}_{n(d)}$ are i.i.d.,

$$\begin{aligned} P_{d,n(.),\vec{q}^d} &\geq Pr\left[\forall i, \left| ||\vec{Y}_i - \vec{q}^d||_p^p - \gamma \right| \leq \gamma\delta(\epsilon, p)\right] \\ &= \left(1 - Pr\left[\left| ||\vec{Y}_1 - \vec{q}^d||_p^p - \gamma \right| > \gamma\delta(\epsilon, p)\right]\right)^{n(d)}. \end{aligned} \tag{1}$$

Our results are reduced to upper-bounding the probability that a sum of random variables, $||\vec{Y}_1 - \vec{q}^d||_p^p$, deviates significantly from a fixed value γ. To our knowledge, developing a useful bound in the most general case is not possible. To get around this problem, we show how our assumptions on $\{f_d\}$ and $n(.)$ allow the dependences between the components of $(\vec{Y}_1 - \vec{q}^d)$ to be broken, and thus, open the door to applying standard concentration bounds (*e.g.* Hoeffding) on the r.h.s. of (1).

Assume $\{f_d\}$ is bounded above sub-exponentially and $support(f_d) = [0,1]^d$. Let $U_1, \ldots, U_d$ be i.i.d. and distributed uniformly on $[0,1]$. Let S denote $\{\vec{y} \in [0,1]^d : \left| ||\vec{y} - \vec{q}^d||_p^p - \gamma \right| > \gamma\delta(\epsilon,p)\}$. There exists sub-exponentially growing function $\beta(.)$ such that,

$$\begin{aligned}
Pr\left[\left| ||\vec{Y}_1 - \vec{q}^d||_p^p - \gamma \right| > \gamma\delta(\epsilon,p)\right] &= \int_{\vec{y}\in S} f_d(\vec{y})\partial\vec{y} \\
&\leq \beta(d)\int_{\vec{y}\in S}\partial\vec{y} \\
&= \beta(d)Pr\left[\left|\sum_{j=1}^{d}|U_j - q_j^d|^p - \gamma\right| > \gamma\delta(\epsilon,p)\right] \\
&\leq \beta(d)2exp\left(\frac{-2\delta(\epsilon,p)^2\left[\frac{d}{(p+1)2^p}\right]^2}{d}\right).
\end{aligned}$$

The first equality and inequality follow from the fact that $support(f_d) = [0,1]^d$ and f_d is bounded above sub-exponentially, respectively. The second inequality follows from Theorem 2 of Hoeffding [6].[4] Plugging this bound into the r.h.s. of inequality (1) yields an expression which goes to one as $d \to \infty$, due to the sub-exponential growth assumptions on $n(.)$ and $\beta(.)$.

3. Dataset Size Assumption

Now we relax the assumption that $n(.)$ grows sub-exponentially while still assuming that $\{f_d\}$ is bounded above sub-exponentially and $support(f_d) = [0,1]^d$. Suppose that, for large d, $n(d) > [4(1+\epsilon)]^d$. We further assume that $p \geq 1$. Our goal in this section is to show that $\{f_d\}$ and $n(.)$ strongly fail to admit instability.

[4] With $\gamma = \sum_{j=1}^d E[|U_j - q_j^d|^p]$, $X_j = |U_j - q_j^d|^p$, and $t = \left(\delta(\epsilon,p)\sum_{j=1}^d E[|U_j - q_j^d|^p]\right)/d$. Clearly $t > 0$. Also, since $support(U_j) = [0,1]$ and $\vec{q}^d \in support(f_d) = [0,1]^d$, then $0 \leq |U_j - q_j^d|^p \leq 1$. Finally, $E[|U_j - q_j^d|^p] = [(q_j^d)^{p+1} + (1-q_j^d)^{p+1}]/(p+1)$ which, for $0 \leq q_j^d \leq 1$, obtains its minimum of $1/(p+1)2^p$ at $q_j^d = 1/2$.

Fix $99/100 < \zeta < 1$ and define $\mathcal{Q}^d$ as $\{\vec{q}^d \in [0,1]^d : Pr[\max_{i=1}^{n(d)} ||\vec{q}^d - \vec{Y}_i||_p - (1+\epsilon)\min_{i=1}^{n(d)} ||\vec{q}^d - \vec{Y}_i||_p \geq 0] \geq 1-\zeta\}$ and $\mathcal{Q} = \Pi_{d=1}^{\infty} Q^d$. Clearly, for any $\{\vec{q}^d\} \in \mathcal{Q}$, $lim_{d\to\infty} P_{d,n(.),\vec{q}^d} \leq \zeta$. Hence, all that remains is to show that $\mathcal{Q}$ is large, *i.e.* for any $0 \leq \omega < 1$, $lim_{d\to\infty} \frac{Volume([0,\omega]^d)}{Volume(\mathcal{Q}^d)} = 0$.

Let $\vec{Y}$ be distributed as f_d and be independent of $\vec{Y}_1, \ldots, \vec{Y}_{n(d)}$. Define random variables $D_{min} = \min_{i=1}^{n(d)}\{||\vec{Y} - \vec{Y}_i||_p\}$, $D_{max} = \max_{i=1}^{n(d)}\{||\vec{Y} - \vec{Y}_i||_p\}$. Such random variables (or related ones) have received considerable study in the stochastic geometry literature. Using one such study [9], we prove, in Appendix A, the following two inequalities with Z denoting $D_{max} - (1+\epsilon)D_{min}$:

$$lim_{d\to\infty} \frac{E[Z]}{d^{1/p}} \geq \frac{1}{100} \text{ and } Volume(\mathcal{Q}^d) \geq \left[\frac{1}{\zeta\beta(d)}\right]\left[\frac{E[Z]}{d^{1/p}} + \zeta - 1\right]. \tag{2}$$

For any $0 \leq \omega < 1$, inequalities (2) as well as the assumptions that $99/100 < \zeta < 1$ and $\beta(d)$ grows sub-exponentially imply that $lim_{d\to\infty} \frac{Volume([0,\omega]^d)}{Volume(\mathcal{Q}^d)} = 0$, as needed.

4. Multi-Variate Gaussian Distributions – Preliminary Results

We provide preliminary results concerning instability admission over an important class of distributions that do not satisfy our assumptions above: $\{N(\vec{\mu}_d, \Sigma_d)\}$. The following simple strategy yields a sufficient condition in the case that: $\vec{q}^d = 0$, $\vec{\mu}_d = 0$, $p = 2$, and the number of eigenvalues of Σ_d which do not go to zero grows faster than $n(.)$. Using the eigenvalue decomposition of Σ_d, it can be shown that

$$\begin{aligned} & Pr\left[\left|||\vec{Y}_1||_2^2 - E[||\vec{Y}_1||_2^2]\right| > E[||\vec{Y}_1||_2^2]\delta(\epsilon,2)\right] \\ = \quad & Pr\left[\left|\sum_{j=1}^{d} W_j^2 - E\left[\sum_{j=1}^{d} W_j^2\right]\right| > E\left[\sum_{j=1}^{d} W_j^2\right]\delta(\epsilon,2)\right], \end{aligned}$$

where the $W's$ are independent and distributed as $N(0,\lambda_j^2)$ with λ_j the j^{th} largest eigenvalue of Σ_d. Chebyshev's inequality shows that the r.h.s. of the equation above is bounded above by

$$\left[\frac{2}{\delta(\epsilon,2)}\right]\left[\frac{\sum_{j=1}^{d}\lambda_j^4}{\sum_{j=1}^{d}\lambda_j^4 + 2\sum_{1\leq\ell\neq k\leq d}\lambda_\ell^2\lambda_k^2}\right].$$

Plugging this bound into the r.h.s. of inequality (1), with $\gamma = E[||\vec{Y}_1||_2^2]$, our assumptions above on $n(.)$ and the $\lambda' s$ imply that $lim_{d\to\infty} P_{d,n(.),\vec{q}^d} = 1$.

Extending the above strategy to $\vec{q}^d, \vec{\mu}_d \neq 0$ and larger growth rates for $n(.)$ seems possible utilizing more complex properties of weighted, non-central chi-square distributions. However, extending beyond $p = 2$ seems difficult as only the 2-norm is preserved by orthogonal transformations. Also, extending beyond multi-variate Gaussian data distributions seems difficult owing to the fact that independence of the $W's$ depends upon the Gaussian assumption.

A. Appendix: Some Proofs

First we prove the left inequality in (2): $lim_{d\to\infty} \frac{E[Z]}{d^{1/p}} \geq \frac{1}{100}$, where $Z = D_{max} - (1+\epsilon)D_{min} = \max_{i=1}^{n(d)}\{||\vec{Y} - \vec{Y}_i||_p\} - (1+\epsilon)\min_{i=1}^{n(d)}\{||\vec{Y} - \vec{Y}_i||_p\}$.

Theorems 1.1 and 1.2 of [9] produce an upper-bound on $E[D_{min}]$ and a lower-bound on $E[D_{max}]$, respectively. These combine to yield[5]

$$\begin{aligned}\frac{E[Z]}{d^{1/p}} \quad &\geq \quad \frac{\Gamma(n(d)+1/d)\Gamma(n(d)+1)}{d^{1/p}\Gamma(n(d))\Gamma(n(d)+1+1/d)3^{1/2}2^{1/d}e^{(1/2d)}||f_d||_2^{2/d}V_{d,p}^{1/d}} \\ &- \quad \frac{2(1+\epsilon)}{d^{1/p}(n(d)+1)^{1/d}V_{d,p}^{1/d}} - o\left(\frac{1+\epsilon}{d^{1/p}(n(d)+1)^{1/d}}\right).\end{aligned}$$

Thus,[6]

$$lim_{d\to\infty}\frac{E[Z]}{d^{1/p}} \geq lim_{d\to\infty}\left(\frac{1}{3^{1/2}d^{1/p}V_{d,p}^{1/d}} - \frac{1}{2d^{1/p}V_{d,p}^{1/d}}\right).$$

From [8] (using the fact that $p \geq 1$) and Stirling's approximation[7] of $\Gamma(.)$ (6.1.3.7 in [1]), $lim_{d\to\infty} d^{1/p}V_{d,p}^{1/d} \leq 2(ep)^{1/p}$. Hence, the above limit is bounded below by (1/100), as desired.

[5] $V_{d,p}$ denotes the volume of the unit-ball in $\mathbb{R}^d$ with respect to the p-norm. $\Gamma(.)$ denotes the standard gamma function.

[6] $lim_{d\to\infty}||f_d||_2^{2/d} \leq 1$ since $support(f_d) = [0,1]^d$ and sequence $\{f_d\}$ is bounded above sub-exponentially. Also, the ratio of the $\Gamma()'s$ approaches one because of the equality $\Gamma(z+1) = z\Gamma(z)$ for any $z \in \mathbb{R}$. Finally, $lim_{d\to\infty}(n(d)+1)^{1/d} \geq 4(1+\epsilon)$ since, by assumption, $n(d) > [4(1+\epsilon)]^d$ for large d.

[7] For large z, $\Gamma(z) \approx exp(-z)z^{z-1/2}(2\pi)^{1/2}$.

Now we prove the right inequality in (2): $Volume(\mathcal{Q}^d) \geq \left[\frac{1}{\zeta\beta(d)}\right]\left[\frac{E[Z]}{d^{1/p}}+\zeta-1\right]$, where $\mathcal{Q}^d = \{\vec{q}^d \in [0,1]^d : Pr[\max_{i=1}^{n(d)} ||\vec{q}^d - \vec{Y}_i||_p - (1+\epsilon)\min_{i=1}^{n(d)} ||\vec{q}^d - \vec{Y}_i||_p \geq 0] \geq 1-\zeta\}$ and $99/100 < \zeta < 1$.

Let f_Z and $f_{Z|\vec{Y}}$ denote the p.d.f of Z and the conditional p.d.f of Z given $\vec{Y}$, respectively. Since $support(f_d) = [0,1]^d$, then $support(f_Z) \subseteq [0, d^{1/p}]$, thus, $E[Z] = \int_{z=0}^{d^{1/p}} z f_Z(z)\partial z \leq d^{1/p}\int_{z=0}^{d^{1/p}} f_Z(z)\partial z = d^{1/p}\int_{z=0}^{d^{1/p}}\int_{\vec{q}^d\in[0,1]^d} f_{Z|\vec{Y}}(z|\vec{q}^d) f_d(\vec{q}^d)\partial\vec{q}^d\partial z = d^{1/p}\int_{\vec{q}^d\in[0,1]^d}\int_{z=0}^{d^{1/p}} f_{Z|\vec{Y}}(z|\vec{q}^d) f_d(\vec{q}^d)\partial z\partial\vec{q}^d$. Hence,

$$\begin{aligned}
\frac{E[Z]}{d^{1/p}} &\leq \int_{\vec{q}^d\in[0,1]^d} f_d(\vec{q}^d)\left[\int_{z=0}^{d^{1/p}} f_{Z|\vec{Y}}(z|\vec{q}^d)\partial z\right]\partial\vec{q}^d \\
&= \int_{\vec{q}^d\in[0,1]^d} f_d(\vec{q}^d) Pr\left[\max_{i=1}^{n(d)}\{||\vec{q}^d - \vec{Y}_i||_p\} - (1+\epsilon)\min_{i=1}^{n(d)}\{||\vec{q}^d - \vec{Y}_i||_p\} \geq 0\right]\partial\vec{q}^d \\
&= \int_{\vec{q}^d\in\mathcal{Q}^d} f_d(\vec{q}^d) Pr\left[\cdots\right]\partial\vec{q}^d + \int_{\vec{q}^d\in([0,1]^d\setminus\mathcal{Q}^d)} f_d(\vec{q}^d) Pr\left[\cdots\right]\partial\vec{q}^d \\
&\leq \int_{\vec{q}^d\in\mathcal{Q}^d} f_d(\vec{q}^d)\partial\vec{q}^d + (1-\zeta)\int_{\vec{q}^d\in([0,1]^d\setminus\mathcal{Q}^d)} f_d(\vec{q}^d)\partial\vec{q}^d \\
&= Pr[\vec{Y}\in\mathcal{Q}^d] + (1-\zeta)Pr[\vec{Y}\in([0,1]^d\setminus\mathcal{Q}^d)] \\
&= \zeta Pr[\vec{Y}\in\mathcal{Q}^d] + 1 - \zeta \\
&\leq \zeta\beta(d) Volume(\mathcal{Q}^d) + 1 - \zeta.
\end{aligned}$$

The second inequality follows from the definition of $\mathcal{Q}^d$ and the last inequality follow from the assumption that f_d is bounded above sub-exponentially. The desired inequality follows.

References

[1] Abramowitz M. and Stegun I. (editors), "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables", *National Bureau of Standards, Applied Mathematics Series* **55**, 1964.

[2] Aggarwal C., Hinneburg A., and Keim D., "On the Surprising Behavior of Distance Metrics in High Dimensional Space", *Lecture Notes in Computer Science*, **1973**, Springer-Verlag, 2001, 420-434.

[3] Beyer K., Goldstein J., Ramakrishnan R., and Shaft U., "When Is 'Nearest Neighbor' Meaningful?", *Lecture Notes in Computer Science*, **1540**, Springer-Verlag, 1998, 217-235.

[4] Francois D., Wertz V., and Verleysen M., "The Concentration of Fractional Distances", *IEEE Trans. on Know. and Data Eng.*, **19(7)**, 2007, 873-886.

[5] Hinneburg A., Aggarwal C., and Keim D., "What is the Nearest Neighbor in High Dimensional Spaces?", *Proc. VLDB Conf.*, 2000, 506-515.

[6] Hoeffding W., "Probability Inequalities for Sums of Bounded Random Variables", *J. Amer. Stat. Assoc.*, **58(301)**, 1963, 13-30.

[7] Hsu C.-M. and Chen M.-S., "On the Design and Applicability of Distance Functions in High-Dimensional Data Space", *IEEE Tran. on Know. and Data Eng.*, **21(4)**, 2009, 523-536.

[8] Huang Z. and He B., "Volume of the Unit Ball in a n-Dimensional Normed Space and Its Asymptotic Properties", *J. Shanghai Univ.*, **12(2)**, 2008, 107-109.

[9] Liitiainen E., Lendasse A., and Corona F., "Bounds on the Mean Power-Weighted Nearest Neighbor Distance", *Proc. of the Royal Soc. A*, **464**, 2008, 2293-2301.

[10] Penrose M., "A Strong Law for the Largest Nearest-Neighbor Link Between Random Points", *J. London Math. Soc.*, **60(2)**, 1999, 951-960.

[11] Pestov V., "On the Geometry of Similarity Search: Dimensionality Curse and Concentration of Measure", *Info. Proc. Let.*, **73(1-2)**, 2000, 47-51.

[12] Shaft U. and Ramakrishnan R., "Theory of Nearest Neighbors Indexability", *ACM Trans. on Database Sys.*, **31(3)**, 2006, 814-838.

[13] Zanger D., "Concentration of Measure and Cluster Analysis", *Stat. & Prob. Let.*, **65**, 2003, 65-70.