

Concentration of measure.

Michel Talagrand

September 21, 2026

Dedicated to the memory of F. Guerra.

I will explain the basic form of concentration of measure

I will explain the basic form of concentration of measure

The fundamental case of Gaussian measure

I will explain the basic form of concentration of measure

The fundamental case of Gaussian measure

Its application to the S.K. model:

I will explain the basic form of concentration of measure

The fundamental case of Gaussian measure

Its application to the S.K. model:

The fluctuations of the energy of the ground state are of order $\sqrt{N}$ (whereas the energy of the ground state is of order N).

I will explain the basic form of concentration of measure

The fundamental case of Gaussian measure

Its application to the S.K. model:

The fluctuations of the energy of the ground state are of order $\sqrt{N}$ (whereas the energy of the ground state is of order N).

My own contribution to concentration of measure: The Taxi Driver inequality.

I will explain the basic form of concentration of measure

The fundamental case of Gaussian measure

Its application to the S.K. model:

The fluctuations of the energy of the ground state are of order $\sqrt{N}$ (whereas the energy of the ground state is of order N).

My own contribution to concentration of measure: The Taxi Driver inequality.

Application: Same conclusion as above for the version of the SK model where the random interactions are random ± 1 rather than Gaussian.

Basic idea of Concentration of measure

For many probability measures on a space of large dimension the following occurs: For any set A such that $P(A) \geq 1/2$ most of the points of the space are close to A .

Basic idea of Concentration of measure

For many probability measures on a space of large dimension the following occurs: For any set A such that $P(A) \geq 1/2$ most of the points of the space are close to A .

As a consequence, well behaved functions have a tendency to be almost constant.

Basic idea of Concentration of measure

For many probability measures on a space of large dimension the following occurs: For any set A such that $P(A) \geq 1/2$ most of the points of the space are close to A .

As a consequence, well behaved functions have a tendency to be almost constant.

Distance of a point to a set: (X, d) metric space. Given a point $x \in X$ and a subset A of X

$$d(x, A) = \inf\{d(x, a) \mid a \in A\}.$$

Basic idea of Concentration of measure

For many probability measures on a space of large dimension the following occurs: For any set A such that $P(A) \geq 1/2$ most of the points of the space are close to A .

As a consequence, well behaved functions have a tendency to be almost constant.

Distance of a point to a set: (X, d) metric space. Given a point $x \in X$ and a subset A of X

$$d(x, A) = \inf\{d(x, a) \mid a \in A\}.$$

“A point x is close to A if it is close to at least a point of A .”

Basic idea of Concentration of measure

For many probability measures on a space of large dimension the following occurs: For any set A such that $P(A) \geq 1/2$ most of the points of the space are close to A .

As a consequence, well behaved functions have a tendency to be almost constant.

Distance of a point to a set: (X, d) metric space. Given a point $x \in X$ and a subset A of X

$$d(x, A) = \inf\{d(x, a) \mid a \in A\}.$$

“A point x is close to A if it is close to at least a point of A .”

Consider moreover a probability P on X . Gromov-Milman of the formulation of concentration of measure: for each subset $A \subset X$ such $P(A) \geq 1/2$ then $P(d(x, A) \geq t)$ is small when t is large.

Consequences in term of Lipschitz functions.

A function f on a metric space (X, d) is Lipschitz if

$$x, y \in X \Rightarrow |f(x) - f(y)| \leq d(x, y).$$

Consequences in term of Lipschitz functions.

A function f on a metric space (X, d) is Lipschitz if

$$x, y \in X \Rightarrow |f(x) - f(y)| \leq d(x, y).$$

That is a good notion of “well behaved function”.

Consequences in term of Lipschitz functions.

A function f on a metric space (X, d) is Lipschitz if

$$x, y \in X \Rightarrow |f(x) - f(y)| \leq d(x, y).$$

That is a good notion of “well behaved function”.

When there is concentration of measure, it is rare that Lipschitz functions take values very different from their median.

Consequences in term of Lipschitz functions.

A function f on a metric space (X, d) is Lipschitz if

$$x, y \in X \Rightarrow |f(x) - f(y)| \leq d(x, y).$$

That is a good notion of “well behaved function”.

When there is concentration of measure, it is rare that Lipschitz functions take values very different from their median. Let M be a median of f :

$$P(f \geq M) \geq 1/2; P(f \leq M) \geq 1/2.$$

Then

$$f(x) \geq M + t \Rightarrow d(x, \{f \leq M\}) \geq t$$

and for large t this set has a small measure.

And conversely

When the Lipschitz functions are almost constant, there is concentration of measure.

And conversely

When the Lipschitz functions are almost constant, there is concentration of measure.

To see this consider a set A . The function $f(x) = d(x, A)$ is Lipschitz.

And conversely

When the Lipschitz functions are almost constant, there is concentration of measure.

To see this consider a set A . The function $f(x) = d(x, A)$ is Lipschitz.

If $P(A) \geq 1/2$ the median M of f is 0. Then $P(d(x, A) \geq t) = P(f \geq M + t)$.

Fundamental exemple: The Gaussian measure γ_N

γ_1 the Gaussian measure on $\mathbb{R}$, of density $c \exp(-x^2/2)$.

Fundamental exemple: The Gaussian measure γ_N

γ_1 the Gaussian measure on $\mathbb{R}$, of density $c \exp(-x^2/2)$. here $c = 1/\sqrt{2\pi}$ is the normalization factor.

Fundamental example: The Gaussian measure γ_N

γ_1 the Gaussian measure on $\mathbb{R}$, of density $c \exp(-x^2/2)$. here $c = 1/\sqrt{2\pi}$ is the normalization factor.

γ_N the canonical Gaussian measure on $\mathbb{R}^N$ of density $c^N \exp(-\|x\|^2/2)$ at the point $x \in \mathbb{R}^N$, where $\|x\|$ is the Euclidean norm, $\|x\|^2 = \sum_{j \leq N} x_j^2$.

Fundamental example: The Gaussian measure γ_N

γ_1 the Gaussian measure on $\mathbb{R}$, of density $c \exp(-x^2/2)$. here $c = 1/\sqrt{2\pi}$ is the normalization factor.

γ_N the canonical Gaussian measure on $\mathbb{R}^N$ of density $c^N \exp(-\|x\|^2/2)$ at the point $x \in \mathbb{R}^N$, where $\|x\|$ is the Euclidean norm, $\|x\|^2 = \sum_{j \leq N} x_j^2$.

γ_N is invariant by rotations, it is “round” .

Fundamental exemple: The Gaussian measure γ_N

γ_1 the Gaussian measure on $\mathbb{R}$, of density $c \exp(-x^2/2)$. here $c = 1/\sqrt{2\pi}$ is the normalization factor.

γ_N the canonical Gaussian measure on $\mathbb{R}^N$ of density $c^N \exp(-\|x\|^2/2)$ at the point $x \in \mathbb{R}^N$, where $\|x\|$ is the Euclidean norm, $\|x\|^2 = \sum_{j \leq N} x_j^2$.

γ_N is invariant by rotations, it is “round” .

γ_N is a product measure

$$c^N \exp(-\|x\|^2) = \prod_{j \leq N} c \exp(-x_j^2/2)$$

et donc $\gamma_N = \gamma_1^{\otimes N}$. In a sense, γ_N “is a cube”. The coordinate functions $x = (x_j)_{j \leq N} \mapsto x_i$ are independent of law γ_1 .

The Gaussian isoperimetric inequality.

Given $a > 0$, $t > 0$, the largest possible value of $\gamma_N(d(x, A) \geq t)$ given that $\gamma_n(A) = a$ is obtained when A is a half space.

The Gaussian isoperimetric inequality.

Given $a > 0$, $t > 0$, the largest possible value of $\gamma_N(d(x, A) \geq t)$ given that $\gamma_n(A) = a$ is obtained when A is a half space.

This is in fact much more precise than concentration of measure.

The Gaussian isoperimetric inequality.

Given $a > 0$, $t > 0$, the largest possible value of $\gamma_N(d(x, A) \geq t)$ given that $\gamma_n(A) = a$ is obtained when A is a half space.

This is in fact much more precise than concentration of measure.

In one dimension everything is easy....

The Gaussian isoperimetric inequality.

Given $a > 0$, $t > 0$, the largest possible value of $\gamma_N(d(x, A) \geq t)$ given that $\gamma_N(A) = a$ is obtained when A is a half space.

This is in fact much more precise than concentration of measure.

In one dimension everything is easy....

Not caring about the best constants, let us try to remember:

$$\gamma_N(A) \geq 1/2 \Rightarrow \gamma_N(d(x, A) \geq t) \leq \exp(-t^2/2).$$

Application to the supremum of Gaussian processes

A Gaussian process is a family $(X_t)_{t \in T}$ of random variables which are centered and jointly Gaussian.

Application to the supremum of Gaussian processes

A Gaussian process is a family $(X_t)_{t \in T}$ of random variables which are centered and jointly Gaussian.

Let $\rho = \sup_{t \in T} (EX_t^2)^{1/2}$.

Application to the supremum of Gaussian processes

A Gaussian process is a family $(X_t)_{t \in T}$ of random variables which are centered and jointly Gaussian.

Let $\rho = \sup_{t \in T} (\mathbb{E} X_t^2)^{1/2}$.

The fluctuations of $\sup_{t \in T} X_t$ are of order at most ρ .

Application to the supremum of Gaussian processes

A Gaussian process is a family $(X_t)_{t \in T}$ of random variables which are centered and jointly Gaussian.

Let $\rho = \sup_{t \in T} (EX_t^2)^{1/2}$.

The fluctuations of $\sup_{t \in T} X_t$ are of order at most ρ .

Independent of the size of $\sup_{t \in T} X_t$ (which can be typically much bigger).

Application to the supremum of Gaussian processes

A Gaussian process is a family $(X_t)_{t \in T}$ of random variables which are centered and jointly Gaussian.

Let $\rho = \sup_{t \in T} (\mathbb{E} X_t^2)^{1/2}$.

The fluctuations of $\sup_{t \in T} X_t$ are of order at most ρ .

Independent of the size of $\sup_{t \in T} X_t$ (which can be typically much bigger).

Consider $g_{i,j}$ independent Gaussian r.v.s. For $\sigma = (\sigma_i)_{i \leq N} \in \{-1, 1\}^N$ let

$$X_\sigma = \frac{1}{\sqrt{N}} \sum_{i < j \leq N} g_{i,j} \sigma_i \sigma_j.$$

Application to the supremum of Gaussian processes

A Gaussian process is a family $(X_t)_{t \in T}$ of random variables which are centered and jointly Gaussian.

Let $\rho = \sup_{t \in T} (\mathbb{E} X_t^2)^{1/2}$.

The fluctuations of $\sup_{t \in T} X_t$ are of order at most ρ .

Independent of the size of $\sup_{t \in T} X_t$ (which can be typically much bigger).

Consider $g_{i,j}$ independent Gaussian r.v.s. For $\sigma = (\sigma_i)_{i \leq N} \in \{-1, 1\}^N$ let

$$X_\sigma = \frac{1}{\sqrt{N}} \sum_{i < j \leq N} g_{i,j} \sigma_i \sigma_j.$$

Then ρ is of order $\sqrt{N}$ and $\sup_\sigma X_\sigma$ is of order N .

Application to the supremum of Gaussian processes

A Gaussian process is a family $(X_t)_{t \in T}$ of random variables which are centered and jointly Gaussian.

Let $\rho = \sup_{t \in T} (\mathbb{E} X_t^2)^{1/2}$.

The fluctuations of $\sup_{t \in T} X_t$ are of order at most ρ .

Independent of the size of $\sup_{t \in T} X_t$ (which can be typically much bigger).

Consider $g_{i,j}$ independent Gaussian r.v.s. For $\sigma = (\sigma_i)_{i \leq N} \in \{-1, 1\}^N$ let

$$X_\sigma = \frac{1}{\sqrt{N}} \sum_{i < j \leq N} g_{i,j} \sigma_i \sigma_j.$$

Then ρ is of order $\sqrt{N}$ and $\sup_\sigma X_\sigma$ is of order N . **Much stronger** than “self-averaging”.

The concept of independent r.v.s is fundamental in probability theory. Given probability spaces $(X_i, P_i)_{i \leq N}$ there exists a probability P on the product $X = \prod_{i \leq N} X_i$ such that for any sets $A_i \subset X_i$

$$P\left(\prod_{i \leq N} A_i\right) = \prod_{i \leq N} P_i(A_i).$$

Product spaces

The concept of independent r.v.s is fundamental in probability theory. Given probability spaces $(X_i, P_i)_{i \leq N}$ there exists a probability P on the product $X = \prod_{i \leq N} X_i$ such that for any sets $A_i \subset X_i$

$$P\left(\prod_{i \leq N} A_i\right) = \prod_{i \leq N} P_i(A_i).$$

In this setting, large dimension=large value of N .

Product spaces

The concept of independent r.v.s is fundamental in probability theory. Given probability spaces $(X_i, P_i)_{i \leq N}$ there exists a probability P on the product $X = \prod_{i \leq N} X_i$ such that for any sets $A_i \subset X_i$

$$P\left(\prod_{i \leq N} A_i\right) = \prod_{i \leq N} P_i(A_i).$$

In this setting, large dimension=large value of N .

If this abstraction bothers you, think of $X = [0, 1]^N$ provided with the volume measure.

Product spaces

The concept of independent r.v.s is fundamental in probability theory. Given probability spaces $(X_i, P_i)_{i \leq N}$ there exists a probability P on the product $X = \prod_{i \leq N} X_i$ such that for any sets $A_i \subset X_i$

$$P\left(\prod_{i \leq N} A_i\right) = \prod_{i \leq N} P_i(A_i).$$

In this setting, large dimension=large value of N .

If this abstraction bothers you, think of $X = [0, 1]^N$ provided with the volume measure. $X_i = [0, 1]$

Hamming distance

The Hamming distance of two points in a product of spaces is the number of coordinates where these two points are different.

Hamming distance

The Hamming distance of two points in a product of spaces is the number of coordinates where these two points are different.

$$x = (x_i)_{i \leq N}, y = (y_i)_{i \leq N}, d(x, y) = \text{card} \{i \leq N, x_i \neq y_i\}.$$

Theorem (The Mac Diarmid inequality)

Consider a function f on $X = \prod X_i$ which is Lipschitz for the Hamming distance. Then for $t > 0$ we have

$$P(|f - Ef| \geq t) \leq 2 \exp(-2t^2/N).$$

The fluctuations of f around Ef (or the median of f) are at most of order. $\sqrt{N}$.

Hamming distance

The Hamming distance of two points in a product of spaces is the number of coordinates where these two points are different.

$$x = (x_i)_{i \leq N}, y = (y_i)_{i \leq N}, d(x, y) = \text{card} \{i \leq N, x_i \neq y_i\}.$$

Theorem (The Mac Diarmid inequality)

Consider a function f on $X = \prod X_i$ which is Lipschitz for the Hamming distance. Then for $t > 0$ we have

$$P(|f - Ef| \geq t) \leq 2 \exp(-2t^2/N).$$

The fluctuations of f around Ef (or the median of f) are at most of order. $\sqrt{N}$.

If $P(A) \geq 1/2$ most of the points of X are within a distance $\sqrt{N}$ of A .

Test on a fundamental example.

We denote by $\mathcal{N}$ the collection of subsets of $\{1, \dots, N\}$.

Test on a fundamental example.

We denote by $\mathcal{N}$ the collection of subsets of $\{1, \dots, N\}$. So $\mathcal{N}$ is $\{0, 1\}^N$.

Let us fix a number $0 < p \leq 1/2$. We construct a random subset $J \in \mathcal{N}$ by deciding whether each $i \leq N$ is in J at random with probability p (and independently over the values of i). The law of J is a product measure μ_p on $\{0, 1\}^N$.

Test on a fundamental example.

We denote by $\mathcal{N}$ the collection of subsets of $\{1, \dots, N\}$. So $\mathcal{N}$ is $\{0, 1\}^N$.

Let us fix a number $0 < p \leq 1/2$. We construct a random subset $J \in \mathcal{N}$ by deciding whether each $i \leq N$ is in J at random with probability p (and independently over the values of i). The law of J is a product measure μ_p on $\{0, 1\}^N$. The Hamming distance is given by $d(J, K) = \text{card}[(K \setminus J) \cup (J \setminus K)]$.

Test on a fundamental example.

We denote by $\mathcal{N}$ the collection of subsets of $\{1, \dots, N\}$. So $\mathcal{N}$ is $\{0, 1\}^N$.

Let us fix a number $0 < p \leq 1/2$. We construct a random subset $J \in \mathcal{N}$ by deciding whether each $i \leq N$ is in J at random with probability p (and independently over the values of i). The law of J is a product measure μ_p on $\{0, 1\}^N$. The Hamming distance is given by $d(J, K) = \text{card}[(K \setminus J) \cup (J \setminus K)]$.

The r.v $\text{card } J$ is the sum of N independent r.v.s. By the CLT its distribution is close to $Np + \sqrt{Np(1-p)}g$ where g is standard normal.

Test on a fundamental example.

We denote by $\mathcal{N}$ the collection of subsets of $\{1, \dots, N\}$. So $\mathcal{N}$ is $\{0, 1\}^N$.

Let us fix a number $0 < p \leq 1/2$. We construct a random subset $J \in \mathcal{N}$ by deciding whether each $i \leq N$ is in J at random with probability p (and independently over the values of i). The law of J is a product measure μ_p on $\{0, 1\}^N$. The Hamming distance is given by $d(J, K) = \text{card}[(K \setminus J) \cup (J \setminus K)]$.

The r.v $\text{card } J$ is the sum of N independent r.v.s. By the CLT its distribution is close to $Np + \sqrt{Np(1-p)}g$ where g is standard normal.

Assume Np integer. Consider

$$A = \{J \in \mathcal{N}; \text{card } J \leq Np\}.$$

Then $P(A) \simeq 1/2$.

Continuation of the Test

Consider $J \subset \mathcal{N}$ with $\text{card } J \geq Np$. For such J consider $K \subset J$ with $\text{card } K = Np$. Then $K \in A$ and $d(K, J) = \text{card } J - Np$.

Continuation of the Test

Consider $J \subset \mathcal{N}$ with $\text{card } J \geq Np$. For such J consider $K \subset J$ with $\text{card } K = Np$. Then $K \in A$ and $d(K, J) = \text{card } J - Np$.

For the typical J , $\text{card } J = Np + u\sqrt{Np(1-p)}$ for u of order 1. Consider the case $u > 0$ for otherwise $J \in A$. So $d(J, A) = u\sqrt{Np(1-p)}$ is indeed (crudely) of order $\sqrt{N}$.

Continuation of the Test

Consider $J \subset \mathcal{N}$ with $\text{card } J \geq Np$. For such J consider $K \subset J$ with $\text{card } K = Np$. Then $K \in A$ and $d(K, J) = \text{card } J - Np$.

For the typical J , $\text{card } J = Np + u\sqrt{Np(1-p)}$ for u of order 1. Consider the case $u > 0$ for otherwise $J \in A$. So $d(J, A) = u\sqrt{Np(1-p)}$ is indeed (crudely) of order $\sqrt{N}$.

By much more is true. **Each** $K \subset J$ set $\text{card } K = Np$ approaches J . “ K captures all the points of J except $\text{card } J - Np$ of them”.

Continuation of the Test

Consider $J \subset \mathcal{N}$ with $\text{card } J \geq Np$. For such J consider $K \subset J$ with $\text{card } K = Np$. Then $K \in A$ and $d(K, J) = \text{card } J - Np$.

For the typical J , $\text{card } J = Np + u\sqrt{Np(1-p)}$ for u of order 1. Consider the case $u > 0$ for otherwise $J \in A$. So $d(J, A) = u\sqrt{Np(1-p)}$ is indeed (crudely) of order $\sqrt{N}$.

By much more is true. **Each** $K \subset J$ set $\text{card } K = Np$ approaches J . “ K captures all the points of J except $\text{card } J - Np$ of them”.

A point i of J is in average missed

$$\alpha_i := \frac{\text{card } J - Np}{\text{card } J}$$

times. This number is about $u/\sqrt{Np}$.

Continuation of the Test

Consider $J \subset \mathcal{N}$ with $\text{card } J \geq Np$. For such J consider $K \subset J$ with $\text{card } K = Np$. Then $K \in A$ and $d(K, J) = \text{card } J - Np$.

For the typical J , $\text{card } J = Np + u\sqrt{Np(1-p)}$ for u of order 1. Consider the case $u > 0$ for otherwise $J \in A$. So $d(J, A) = u\sqrt{Np(1-p)}$ is indeed (crudely) of order $\sqrt{N}$.

By much more is true. **Each** $K \subset J$ set $\text{card } K = Np$ approaches J . “ K captures all the points of J except $\text{card } J - Np$ of them”.

A point i of J is in average missed

$$\alpha_j := \frac{\text{card } J - Np}{\text{card } J}$$

times. This number is about $u/\sqrt{Np}$. And J has about Np points. Thus $\sum_{i \in J} \alpha_i^2$ is about u^2 , **independent** de p et N .

Continuation of the Test

Consider $J \subset \mathcal{N}$ with $\text{card } J \geq Np$. For such J consider $K \subset J$ with $\text{card } K = Np$. Then $K \in A$ and $d(K, J) = \text{card } J - Np$.

For the typical J , $\text{card } J = Np + u\sqrt{Np(1-p)}$ for u of order 1. Consider the case $u > 0$ for otherwise $J \in A$. So $d(J, A) = u\sqrt{Np(1-p)}$ is indeed (crudely) of order $\sqrt{N}$.

By much more is true. **Each** $K \subset J$ set $\text{card } K = Np$ approaches J . “ K captures all the points of J except $\text{card } J - Np$ of them”.

A point i of J is in average missed

$$\alpha_i := \frac{\text{card } J - Np}{\text{card } J}$$

times. This number is about $u/\sqrt{Np}$. And J has about Np points. Thus $\sum_{i \in J} \alpha_i^2$ is about u^2 , **independent** de p et N .

Crucial step

We still consider probability spaces (X_i, P_i) , their product $X = \prod_{i \leq N} X_i$ provided with the product probability P . For a subset A of X et $x \in X$, we define a new notion of “distance” between A and x , a quantity $\varphi(x, A) \geq 0$.

Crucial step

We still consider probability spaces (X_i, P_i) , their product $X = \prod_{i \leq N} X_i$ provided with the product probability P . For a subset A of X et $x \in X$, we define a new notion of “distance” between A and x , a quantity $\varphi(x, A) \geq 0$.

There are two equivalent definitions, the equivalence being non trivial.

Crucial step

We still consider probability spaces (X_i, P_i) , their product $X = \prod_{i \leq N} X_i$ provided with the product probability P . For a subset A of X et $x \in X$, we define a new notion of “distance” between A and x , a quantity $\varphi(x, A) \geq 0$.

There are two equivalent definitions, the equivalence being non trivial. So consider $x \in X$.

Crucial step

We still consider probability spaces (X_i, P_i) , their product $X = \prod_{i \leq N} X_i$ provided with the product probability P . For a subset A of X et $x \in X$, we define a new notion of “distance” between A and x , a quantity $\varphi(x, A) \geq 0$.

There are two equivalent definitions, the equivalence being non trivial. So consider $x \in X$.

Definition 1: For $y \in A$ consider $z = z(x, y) \in \mathbb{R}^N$ such $z_i = 0$ si $x_i = y_i$ and $z_i = 1$ otherwise

Crucial step

We still consider probability spaces (X_i, P_i) , their product $X = \prod_{i \leq N} X_i$ provided with the product probability P . For a subset A of X and $x \in X$, we define a new notion of “distance” between A and x , a quantity $\varphi(x, A) \geq 0$.

There are two equivalent definitions, the equivalence being non trivial. So consider $x \in X$.

Definition 1: For $y \in A$ consider $z = z(x, y) \in \mathbb{R}^N$ such $z_i = 0$ si $x_i = y_i$ and $z_i = 1$ otherwise Consider the convex hull C_x of all the points $z(x, y)$ when y varies in A .

Crucial step

We still consider probability spaces (X_i, P_i) , their product $X = \prod_{i \leq N} X_i$ provided with the product probability P . For a subset A of X and $x \in X$, we define a new notion of “distance” between A and x , a quantity $\varphi(x, A) \geq 0$.

There are two equivalent definitions, the equivalence being non trivial. So consider $x \in X$.

Definition 1: For $y \in A$ consider $z = z(x, y) \in \mathbb{R}^N$ such $z_i = 0$ si $x_i = y_i$ and $z_i = 1$ otherwise Consider the convex hull C_x of all the points $z(x, y)$ when y varies in A . Define

$\varphi(x, A) = d(0, C_x)$, when d is the Euclidean distance.

Definition 2.

$$\varphi(x, A) = \sup \inf_{y \in A} \sum_{i: y_i \neq x_i} \alpha_i,$$

where the supremum is over the sequences (α_i) such that $\sum_i \alpha_i^2 \leq 1$.

Definition 2.

$$\varphi(x, A) = \sup_{y \in A} \inf_{i; y_i \neq x_i} \sum \alpha_i,$$

where the supremum is over the sequences (α_i) such that $\sum_i \alpha_i^2 \leq 1$.

Observation: the choice $\alpha_i = 1/\sqrt{N}$ gives $d(x, A) \leq \sqrt{N}\varphi(x, A)$.

Definition 2.

$$\varphi(x, A) = \sup_{y \in A} \inf_{i; y_i \neq x_i} \sum \alpha_i,$$

where the supremum is over the sequences (α_i) such that $\sum_i \alpha_i^2 \leq 1$.

Observation: the choice $\alpha_i = 1/\sqrt{N}$ gives $d(x, A) \leq \sqrt{N}\varphi(x, A)$.

Theorem (The Taxi driver inequality)

It holds

$$\int \exp\left(\frac{1}{4}\varphi(x, A)^2\right) dP(x) \leq \frac{1}{P(A)}.$$

No more N !

Easy proof by induction over N . What was hard was to find the correct formulation.

Easy proof by induction over N . What was hard was to find the correct formulation.

Consequently:

$$P(\varphi(x, A) \geq t)P(A) \leq \exp(-t^2/4),$$

Easy proof by induction over N . What was hard was to find the correct formulation.

Consequently:

$$P(\varphi(x, A) \geq t)P(A) \leq \exp(-t^2/4),$$

and since $d(x, A) \leq \sqrt{N}\varphi(x, A)$ we have

$$P(d(x, A) \geq t\sqrt{N}) \leq \frac{1}{P(A)} \exp(-t^2/4).$$

Most of the points of X are at distance at most of order $\sqrt{N}$ of A .

Theorem

Let f be a **convex** Lipschitz function on $\mathbb{R}^N$ and P be a product probability on $[-1, 1]^N \subset \mathbb{R}^N$. Then for $a \in \mathbb{R}$ and $t > 0$ we have

$$P(f \leq a)P(f \geq a + t) \leq \exp(-t^2/16).$$

Theorem

Let f be a **convex** Lipschitz function on $\mathbb{R}^N$ and P be a product probability on $[-1, 1]^N \subset \mathbb{R}^N$. Then for $a \in \mathbb{R}$ and $t > 0$ we have

$$P(f \leq a)P(f \geq a + t) \leq \exp(-t^2/16).$$

This is a precise version of “the fluctuations of f are of order 1”.

Theorem

Let f be a **convex** Lipschitz function on $\mathbb{R}^N$ and P be a product probability on $[-1, 1]^N \subset \mathbb{R}^N$. Then for $a \in \mathbb{R}$ and $t > 0$ we have

$$P(f \leq a)P(f \geq a + t) \leq \exp(-t^2/16).$$

This is a precise version of “the fluctuations of f are of order 1”. In a flowery language, a *convex* Lipschitz function behaves as well for P as a Lipschitz function behaves for γ_N .