# Mathematical Foundations of Infinite-Dimensional Statistical Models:

## 2.2 Isoperimetric Inequality with Applications to Concentration

이종진

Seoul National University

*ga0408@snu.ac.kr*

July 06, 2018

# Table of Contents

# Isoperimetric Inequalities with Applications to Concentration.

- To obtain the best possible concentration inequality with respect to the standard Gaussian measure
  - For Lipschitz functions f about their medians
  - For the supremum norm of a separable Gaussian process X, when $\sup_{t \in T} |X(t)| < \infty$ *a.s.*

1. Prove Isoperimetric inequality on the sphere.
2. Prove Gaussian isoperimetric inequality.
3. Obtain Gaussian concentration inequality.

## 2.2.1 The Isoperimetric Inequality on the Sphere

## Preliminary

- $S^n = \{x \in \mathbb{R}^n : \|x\|^2 = \sum_{i=1}^{n+1} {x_i}^2 = 1\}$, where $x = (x_1, ..., x_{n+1})$;
- p is the north pole, p=(0,...,0,1) (could be an arbitrary point in $S^n$)
- $\mu$ is the uniform probability distribution on $S^n$.
- d is the geodesic distance on $S^n$
- C(x,$\rho$) is a closed cap centered at a point $x \in S^n$ with a radius $\rho$, C(x,$\rho$) $:= \{y : d(x, y) \leq \rho\}$
- $A_\epsilon$ is $\epsilon$-neighbourhood of a set A, $A_\epsilon := \{x : d(x, A) \leq \epsilon\}$, and d(x,A) $= \inf\{d(x, y) : y \in A\}$

# Theorem 2.2.1

Theorem 2.2.1

*Let $A \neq \emptyset$ be a measurable subset of $S^n$, and let $C$ be a cap such that $\mu(C) = \mu(A)$. Then, for all $\epsilon > 0$,*

$$\mu(C_\epsilon) \leq \mu(A_\epsilon) \tag{2.4}$$

The isoperimetric inequality on the sphere states that the caps are the sets of shortest perimeter among all measuralbe sets of a given surface area.

# Proof of Theorem 2.2.1

Proof.

If $\mu(A) = 0$, then C consists of a single point, and (2.4) holds.

If $\mu(A) \neq 0$, It suffices to prove the theorem for A compact.

$(\because)$ There exists $A^m$ compact, $A^m \subset A$, $A^m$ increasing and such that $\mu(A^m) \nearrow \mu(A)$. Let $C^m$ be caps with the same centred as C and with $\mu(C^m) = \mu(A^m)$. since the measuer of a cap is continuous one to one function, we have $\mu(C^m) \nearrow \mu(C)$.

because $A^m$ is compact, $\mu({A^m}_\epsilon) \geq \mu({C^m}_\epsilon)$

$$\therefore \mu(A_\epsilon) \geq \lim \mu({A^m}_\epsilon) \geq \lim \mu({C^m}_\epsilon) \geq \mu(C_\epsilon)$$

# Proof of Theorem 2.2.1

Proof.

We Assume A is compact and $\mu(A) \neq 0$ from now.

The rest of proof is divided into three parts

1. Part 1: Construction and main properties of the symmetrisation operation.($A^*$ is called a symmetrisation of A)
2. Part 2: Defining a certain collection of compact sets $\mathcal{A}$ containing A and proving properties of $\mathcal{A}$
3. Part 3: Completion of the proof of Theorem 2.2.1.

□

# Proof of Theorem 2.2.1, Part 1

- Construct transformation A $\mapsto A^*$ on measurable subsets of the sphere that preserves area $(\mu(A) = \mu(A^*))$, and decrease perimeter $\mu(A^*_\epsilon)) \leq (\mu(A_\epsilon)$.
- H is n-dimensional subspace $(H \subset \mathbb{R}^{n+1})$ that does not contain the north pole p.
- $\sigma = \sigma_H$ be the reflection about H.
  ($\rightarrow$ property : for x,y on the same half-space.)

$$d(x, y) \leq d(x, \sigma(y)) \tag{2.5}$$

- $S_+$ is the open hemisphere that contains p.
- $S_-$ is the other hemisphere.
- $S_0 = S^n \cap H$

# Proof of Theorem 2.2.1, Part 1

- The symmetrisation of A with respect to $\sigma = \sigma_H$, $s_H = A^*$ is defined as

$$
\begin{aligned}
s_H(A) &= A^* \\
&= [A \cap (S_+ \cup S_0)] \cup \{a \in A \cap S_- : \sigma(a) \in A\} \qquad (2.6) \\
&\cup \{\sigma(a) : a \in A \cap S_-, \sigma(a) \notin A\}
\end{aligned}
$$

- The symmetrisation should have two properties
  (i) $\mu(A) = \mu(A^*)$
  (ii) $\mu(A_\epsilon) \leq \mu(A^*_\epsilon)$

## Proof of Theorem 2.2.1, Part 1

(i) $\mu(A) = \mu(A^*)$

Three sets in the defition are disjoint, $\mu\{\sigma(a) : a \in A \cap S_-, \sigma(a) \notin A\} = \mu\{a \in A \cap S_- : \sigma(a) \notin A\}$ implies

$$\mu(A^*) = \mu(A),\ A \in \mathcal{B}(S^{n+1}) \tag{2.7}$$

(ii) $\mu(A_\epsilon) \leq \mu(A^*_\epsilon)$

we prove more, for all $A \in \mathcal{B}(S^{n+1})$ and $\epsilon > 0$, then

$$(A^*)_\epsilon \subset (A_\epsilon)^*, \textit{hence}\ \mu((A^*)_\epsilon) \leq \mu((A_\epsilon)^*) = \mu(A_\epsilon). \tag{2.8}$$

# Proof of Theorem 2.2.1, Part 1

proof of (2.8).

For $x \in (A^*)_\epsilon$, let $y \in A^*$ such that $d(x, y) \leq \epsilon$

1) when x and y lay on different half-spaces,

Using (2.5) and $\sigma$ is involutive isometry we obtain,

$$d(\sigma(x), y) = d(x, \sigma(y)) \leq d(\sigma(x), \sigma(y)) = d(x, y) \leq \epsilon$$

since, $y \in A^*$ implies that $y \in A$ or $\sigma(y) \in A$, in either case $x \in A_\epsilon$ and $\sigma(x) \in A_\epsilon$

$\therefore x \in (A_\epsilon)^*$

# Proof of Theorem 2.2.1, Part 1

proof of (2.8).

2) when x and y are in $S_-$,

The following inequalites still holds,

$$d(\sigma(x), \sigma(y)) = d(x, y) \leq \epsilon$$

since, $y \in A^*$ and $y \in S_-$,

$y \in A$ and $\sigma(y) \in A$, therefore, $x \in A_\epsilon$ and $\sigma(x) \in A_\epsilon$

$\therefore x \in (A_\epsilon)^*$

3) when x and y are in $S_+$,

$y \in A$ or $\sigma(y) \in A$, therefore, $x \in A_\epsilon$ or $\sigma(x) \in A_\epsilon$ and $x \in S_+$

$\therefore x \in (A_\epsilon)^*$

# Proof of Theorem 2.2.1, Part 1

proof of (2.8).

4) when x are in $S_0$,

$y \in A$ implies that $y \in A$ or $\sigma(y) \in A$, therefore $x \in A_\epsilon$ or $\sigma(x) \in A_\epsilon$ and

$x \in S_0$ means $x = \sigma(x)$

$\therefore x \in (A_\epsilon)^*$

5) when y are in $S_0$,

$y \in A$ implies that $y = \sigma(y) \in A$, therefore $x \in A_\epsilon$ and $\sigma(x) \in A_\epsilon$

$\therefore x \in (A_\epsilon)^*$

$$\therefore (A^*)_\epsilon \subset (A_\epsilon)^*$$

□

# Proof of Theorem 2.2.1, Part 2

- ($\mathcal{K}$,h) denotes the set of nonempty compact subsets of $S^n$ equipped with the Hausdorff distance.
- h is Hausdorff distance, $h(A, B) = inf\{\epsilon : A \subset B_\epsilon, B \subset A_\epsilon\}$
- $\mathcal{A}$ is the minimal closed subset of $\mathcal{K}$ that contain A and is preserved by $s_H$ for all n-dimensional subspaces H of $\mathbb{R}^{n+1}$ that do not contain the north pole (K is a closed $\{s_H\}$-invariant cooloction of sets that contain A.)
- (If $A \in \mathcal{A}$, then $s_H(A) \in \mathcal{A}$ for all H with $p \notin H$)
- Proving of the following claim is main purpose of part 2

Claim: If $B \in \mathcal{A}$, then

(a) $\mu(B) = \mu(A)$, and (b) for all $\epsilon > 0$, $\mu(B_\epsilon) \leq \mu(A_\epsilon)$

# Proof of Theorem 2.2.1, Part 2

Proof.

It suffices to show that the collection of closed sets $\mathcal{F}$ satisfying (a) and (b) is preserved by $s_H$ for all H not containing p and is closed subset of $\mathcal{K}$. ($\because$ minimality of A)

1) preserved by $s_H$ for all H not containing p:

For $B \in \mathcal{F}$, $B^* = s_H(B)$

by (2.7), $\mu(B^*) = \mu(B) = \mu(A) \cdots (a)$

by (2.8), $\mu((B^*)_\epsilon) \leq \mu(B_\epsilon) \leq \mu(A_\epsilon) \cdots (b)$

$\therefore B^* = s_H(B) \in \mathcal{F}$

# Proof of Theorem 2.2.1, Part 2

Proof.

2) closed subset of $\mathcal{K}$:

Let $B^n \in \mathcal{F}$ and $h(B^n, B) \to 0$.

Let $\epsilon > 0$ be fixed. Given $\delta > 0$, there exists $n_\delta$ such that $B \subset B^n_\delta$

Then, for all $n \geq n_\delta$, $\mu(B_\epsilon) \leq \mu(B^n_{\delta+\epsilon}) \leq \mu(A_{\delta+\epsilon})$

Letting $\delta \searrow 0$, B satisfies condition (b)

Letting $\epsilon \searrow 0$, $\mu(B) \leq \mu(A)$ and, for all n large enough, $B^n \subset B_\delta$,

then, we get that $\mu(A) = \mu(B^n) \leq \mu(B_\delta)$ letting $\delta \searrow 0$

$\mu(A) \leq \mu(B)$, $\mu(A) = \mu(B)$, B satisfies condition (a).

$\therefore B \in \mathcal{F}$ and $\mathcal{F}$ is a closed subset of $\mathcal{K}$.

□

# Proof of Theorem 2.2.1, Part 3

- Because of the claim about $\mathcal{A}$,

  it suffices to show that if C is the cap centred at p such that

  $\mu(A) = \mu(C)$, then $C \in \mathcal{A}$

  $\rightarrow$ Instead, $\mu(C_\epsilon) \leq \mu(A_\epsilon)$

# Proof of Theorem 2.2.1, Part 3

Proof.

Define $f(B) = \mu(B \cap C), B \in \mathcal{A}$

1) f is upper semicontinuous on $\mathcal{A}$:

For any $B, B^n \in \mathcal{A}$ such that $h(B, B^n) \to 0$ ,

Given $\delta > 0$, for all n large enough, $B^n \subset B_\delta$, and it implies

$B^n \cap C \subset B_\delta \cap C$ (In book, $B^n \cap C \subset (B \cap C_\delta)_\delta$)

Hence, $limsup_n \mu(B^n \cap C) \leq \mu(B_\delta \cap C)$

B and C are closed, if $\delta_n \searrow 0$, then $\cap_n (B_{\delta_n} \cap C) = B \cap C$

$\therefore \limsup_n \mu(B^n \cap C) \leq \mu(B \cap C)$

# Proof of Theorem 2.2.1, Part 3

## Proof.

Since f is upper semicontinuous on $\mathcal{A}$ and $\mathcal{A}$ is compact, f attains its maximum at some $B \in \mathcal{A}$

2) The theorem will be proved if we show that $C \subset B$:

Assume that $C \not\subset B$, then since $\mu(A) = \mu(B) = \mu(C)$.

,and both C and B are closed, $B \setminus C$, $C \setminus B$, have positive $\mu$-measure.

For $x \in B \setminus C, y \in C \setminus B$, H be the subspace of dimension n orthogonal to the vector x-y, and define $s_H$ with respect to $\sigma_H$

Then,

1. $\sigma(y) = x$ (x,y are in $S^n$, H is hyperplane)
2. p is not in H (if p is in H, the reflection of a point in C should be in C)
3. $x \in S_-, y \in S_+$ $(d(y,p) \leq d(x,p) \leq d(\sigma_H(y), p))$

# Proof of Theorem 2.2.1, Part 3

Proof.

By definition of density point, for $\delta > 0$ small enough,

$C(x,\delta) \subset S_-, \sigma(C(x,\delta)) = C(y,\delta) \subset S_+$,

$\mu((B \setminus C) \cap C(x,\delta)) \geq 2\mu(C(x,\delta))/3$, $\mu((C \setminus B) \cap C(y,\delta)) \geq 2\mu(C(y,\delta))/3$

Then, the set $D = ((B \setminus C) \cap C(x,\delta)) \cap \sigma((C \setminus B) \cap C(y,\delta))$ satisfies

$$\mu(D) \geq \mu(C(x,\delta))/3 > 0, D \subset (B \setminus C) \cap S_- \text{ and } \sigma(D) \subset C \setminus B \qquad (2.10)$$

,and (2.10) imply that $\sigma(D) \subset B^* \cap C$ and $\sigma(D) \cap (B \cap C)^* = \emptyset$

$(\because z \in (B \cap C)^*$ implies either $z \in B \cap C$ or $\sigma(z) \in B \cap C)$)

# Proof of Theorem 2.2.1, Part 3

Proof.

$$\mu(B \cap C) = \mu((B \cap C)^*) \leq \mu(B^* \cap C) \tag{2.9}$$

$(\because (B \cap C)^* \subset (B^* \cap C)$

1. $x \in B \cap C \cap (S_+ \cup S_0)$, we obviously have $x \in B^* \cap C$.
2. $x \in B \cap C \cap S_-, \sigma(x) \in B \cap C$, we obviously have $x \in B^* \cap C$.
3. $x = \sigma(z), z \in B \cap C \cap S_-, \sigma(z) \notin B \cap C$,

   $\sigma(z) \in C(\because z \in C)$, then $\sigma(z)$ is not in B and therefore, $x \in B^* \cap C$. )

# Proof of Theorem 2.2.1, Part 3

Proof.

$$\mu(B^* \cap C) \geq \mu((B \cap C)^* \cup \sigma(D)) = \mu((B \cap C)^*) + \mu(D) > \mu((B \cap C)^*),$$

,because $B^* \in \mathcal{A}$, contradicts the fact that f attain it maximum at B.

$\therefore C \subset B$ (In the book, $C \in \mathcal{A}$ )

Then, $C_\epsilon \subset B_\epsilon, \mu(C_\epsilon) \leq \mu(B_\epsilon) \leq \mu(A_\epsilon)$ □

# Lebesgue's density theorem

## Theorem Lebesgue's density

*Let $\mu$ be the Lebesgue measure on the Euclidean space $R^n$ and $A$ be a Lebesgue measurable subset of $R^n$. Define the approximate density of $A$, in a $\epsilon$-neighborhood of a point $x$ in $R^n$ as*

$$d_\epsilon(x) = \left( \frac{\mu(A \cap B_\epsilon(x))}{\mu(B_\epsilon(x))} \right)$$

*,where $B_\epsilon$ denotes the closed ball of radius $\epsilon$ centered at $x$. Lebesgue's density theorem asserts that for almost every point $x$ of $A$, the density is*

$$d(x) = lim_{\epsilon \to 0} d_\epsilon(x) = 1$$

# Proof of Theorem 2.2.1, Part3 : $\mathcal{A}$ is compact

- $\mathcal{A}$ is closed, so it suffices to show that $\mathcal{K}$ is compact.
- Exercise 2.2.8
  - Prove that ($\mathcal{K}$,h), the space of nonempty compact subsets of $S^n$ with the Hausdorff distance, is a compact metric space.
  - Hint:

    1 : Show the map $\mathcal{K} \mapsto C(S^n), A \mapsto d(\cdot, A)$ is an isometry

    2 : Image in $(C(S^n), \|\cdot\|_\infty)$ is compact.

# Proof of Theorem 2.2.1, Part3 : $\mathcal{A}$ is compact

Exercise 2.2.8.

$x \mapsto d(x, A)$ is isometry :

- $\sup\{|d(x,A) - d(x,B)|, x \in S^n\} \leq h(A,B)$, *for* $A, B \in \mathcal{K}$

  $\rightarrow$

  For $\epsilon < \epsilon^* < h(A,B)$, we can take $x \in A - B_\epsilon$(or we can take $x \in B - A_\epsilon$)

  ,and for $\epsilon_n \nearrow \epsilon$, take $x_n \in A - B_{\epsilon_n}$

  then $x = \lim x_n$ is in A($\because$ A is closed) ,$d(x,A) = 0$

  and $d(x_n, B) > \epsilon_n$ $n \rightarrow \infty$ and $d(x,B) \geq \epsilon^*$
- $\sup\{|d(x,A) - d(x,B)|, x \in S^n\} \geq h(A,B)$, *for* $A, B \in \mathcal{K}$

  $\rightarrow$ trivial

□

## Proof of Theorem 2.2.1, Part3 : $\mathcal{A}$ is compact

Exercise 2.2.8.

$\{d(\cdot, A) : A \in \mathcal{K}\}$ is compact :

$\{d(x, A) : x \in S^n, A \in \mathcal{K}\}$ is compact ($\because$ closed and bounded in $\mathbb{R}$)

and $\{d(\cdot, A) : A \in \mathcal{K}\}$ is equi-continuous

By Arzelà–Ascoli theorem, $\{d(\cdot, A) : A \in \mathcal{K}\}$ is compact

$\mathcal{K}$ is compact :

inverse of isometry is continuous fucntion, and continuous function preserves compactness.

$\square$

**Theorem Arzelà–Ascoli**

*Let $X$ be a compact metric(Hausdorff) space, $Y$ be a metric space, $\mathcal{F} \subset Y^X$ be the continuous family of functions. The followings are equivalent:*

- *$\mathcal{F}$ is compact set*
- 
  - *$\mathcal{F}$ be the equi-continuous family of functions.*
  - *for all $x \in X$, $\{f(x) : x \in \mathcal{F}\} \subset Y$ is compact set.*

## 2.2.2 The Gaussian Isoperimetric Inequality for the Standard Gaussian Measure on $\mathbb{R}^{\mathbb{N}}$

## 2.2.2 The Gaussian Isopermetric Inequality for the Standard Gaussian Measure on$\mathbb{R}^{\mathbb{N}}$

- Isoperimetric inequality on the sphere $\rightarrow$ isoperimetric inequality for the probability law $\gamma_n$ of n independente N(0,1)
- By means of Poincare's lemma.
- $\gamma_n$ is the limit of the projection of the uniform distribution on $\sqrt{m}S^{n+m}$ onto $\mathbb{R}^n$ when $m \rightarrow \infty$

# Preliminary

- $g_i, i \in \mathbb{N}$ is a sequence of independent N(0,1) random variables.
- $\gamma_n = \mathcal{L}(g_1, \cdots, g_n)$, standard Gaussian measure on $\mathbb{R}^n$
- $\gamma = \mathcal{L}(\{g_i\}_{i=1}^{\infty})$, standard Gaussian measure on $\mathbb{R}^{\mathbb{N}}$
  (cylindrical $\sigma$-algebra $\mathcal{C}$ of $\mathbb{R}^{\mathbb{N}}$.)
- Euclidean neighbourhoods $A_\epsilon := \{x \in \mathbb{R}^n : d(x, A) \leq \epsilon\} = A + \epsilon O_n$
  d : Euclidean destance, $O_n$ : the closed ball centred at $0 \in \mathbb{R}^n$.

# Gaussian isoperimetric inequality and isoperimetric inequality on the sphere

- Gaussian isoperimetric inequality.

  For $n < \infty$, let $\gamma_n$ be the standard Gaussian measure on $\mathbb{R}^n$, let A be a measurable subset of $\mathbb{R}^n$, and let H be a half-space $H = \{x \in \mathbb{R}^n : \langle x, u \rangle \leq a\}$, u a unit vector, such that $\gamma_n(H) = \gamma_n(A)$ and hence with $a := \Phi^{-1}(\gamma_n(A))$, where $\Phi$ denotes the standard normal distribution function. Then, for all $\epsilon > 0$,

$$\gamma_n(H + \epsilon O_n) \leq \gamma_n(A + \epsilon O_n) \tag{2.12}$$

- (cf. isoperimetric inequality on the sphere)

  Let A $\neq \emptyset$ be a measurable subset of $S^n$, and let C be a cap such that $\mu(C) = \mu(A)$. Then, for all $\epsilon > 0$,

$$\mu(C_\epsilon) \leq \mu(A_\epsilon) \tag{2.4}$$

### Lemma 2.2.2 (Poincare's lemma)

*Let* $\mu_{n+m}$ *be the uniform distribution on* $\sqrt{m}S^{n+m}$*, the sphere of* $\mathbb{R}^{n+m+1}$ *of radius* $\sqrt{m}$ *and centered at the origin. Let* $\pi_m$ *be the orthogonal projection* $\mathbb{R}^{n+m+1} \mapsto \mathbb{R}^n = \{x \in \mathbb{R}^{m+n+1} : x_i = 0, n < i \leq n+m+1\}$ *and let* $\tilde{\pi}_m$ *be the restriction of* $\pi_m$ *to* $\sqrt{m}S^{n+m}$*. Let* $\nu_m = \mu_{n+m} \circ \tilde{\pi}_m^{-1}$ *be the projection onto* $\mathbb{R}^n$ *of* $\mu_{n+m}$*. Then,* $\nu_m$ *has a density* $f_m$ *such that if* $\phi_m$ *is the density of* $\nu_m$*,* $lim_{m \to \infty} f_m(x) = \phi_m(x)$ *for all* $x \in \mathbb{R}^n$ *Therefore,*

$$\nu_n(A) = lim_{m \to \infty} \mu_{n+m}(\tilde{\pi}_m^{-1}(A)) \tag{2.11}$$

*for all Borel sets A of* $\mathbb{R}^n$

# Proof of Poincare's lemma

## Proof.

Set $G_n := (g_1, \ldots, g_n)$ and $G_{m+n+1} := (g_1, \ldots, g_{m+n+1})$

$\mu_{n+m}$ is the law $\sqrt{m} G_{n+m+1} / \left| G_{n+m+1} \right|^{1/2}$.(rotational invariant)

For any measurable set A of $\mathbb{R}^n$,

$$\nu_m(A) = \frac{1}{(2\pi)^{(n+m+1)/2}} \int_{\mathbb{R}^{m+1}} \int_{\tilde{A}(y)} e^{-(|z|^2+|y|^2)/2} dzdy,$$

$$where, z \in \mathbb{R}^n, y \in \mathbb{R}^{m+1}, \tilde{A} = \left\{ z \in \mathcal{R}^n : \sqrt{m/(|z|^2 + |y|^2)}, z \in A \right\}$$

Make the change of variables

$x = \sqrt{m/(|z|^2 + |y|^2)}, (z = |y| \, x / \sqrt{m - |x|^2}, |x| \leq \sqrt{m})$

# Proof of Poincare's lemma

Proof.

$$\nu_m(A) = \frac{1}{(2\pi)^{(n+m+1)/2}}$$

$$\int_A \mathrm{I}(|x| \leq \sqrt{m}) \frac{m}{(m - |x|^2)^n/2 + 1} \int_{\mathbb{R}^{m+1}} |y|^n \, exp\left(-\frac{1}{2}\frac{m\,|y|^2}{m - |x|^2}\right) dydx$$

$$= \frac{E(|G_{m+1}|^n)}{m^{n/2}} \frac{1}{(2\pi)^{n/2}} \int_A \left(1 - \frac{|x|^2}{m}\right)^{(m-1)/2} \mathrm{I}(|x| \leq \sqrt{m}) dx$$

Hence, the density of $\nu_m$ is

$f_m(x) = C_{n,m}(2\pi)^{-n/2}(1 - |x|^2/m)^{(m-1)/2)}(I)(|x|^2 < m), x \in (R)^n$

Clearaly, $(2\pi)^{-n/2}(1 - |x|^2/m)^{(m-1)/2)}(I)(|x|^2 < m) \to (2\pi)^{-n/2}e^{|x|^2/2}$

,for all x as $m \to \infty$

# Proof of Poincare's lemma

Proof.

For $0 \leq a < m$ and $m \geq 2$, we have $1 - a/m \leq e^{-a/2(m-1)}$, it follows that

$(1 - |x|^2/m)^{(m-1)/2)}(I)(|x|^2 < m)$ is dominated by $e^{-|x|^2/4}$

Thus, by dominated convergence theorem, $f_m(x)/C_{n,m} \to (2\pi)^{-n/2} e^{|x|^2/2}$ in $L_1$

(which impies that $C_{n,m}^{-1} \to 1$) □

# Theorem 2.2.3

## Theorem 2.2.3

*For $n < \infty$, let $\gamma_n$ be the standard Gaussian measure of $\mathbb{R}^n$, let $A$ be a measurable subsets of $\mathcal{R}^n$, and let $A$ be a measurable subset of $\mathbb{R}^n$, and let $H$ be a half-space $H = \{x \in \mathbb{R}^n : \langle x, u\rangle \leq a\}$, $u$ a unit vector, such that $\gamma_n(H) = \gamma_n(A)$ and hence with $a := \Phi^{-1}(\gamma_n(A))$, where $\Phi^{-1}$ denots the standard normal distribution function. Then, for all $\epsilon > 0$,*

$$\gamma_n(H + \epsilon O_n) \leq \gamma_n(A + \epsilon O_n), \tag{2.12}$$

*which, by the definition of $a$, is equivalent to*

$$\gamma_n(A + \epsilon O_n) \geq \Phi(\Phi^{-1}(\gamma_n(A)) + \epsilon) \tag{2.13}$$

# Proof of Theorem 2.2.3

Proof.

For $-\sqrt{m} < b < \sqrt{m}$, define a half-space $H_b := \{x \in \mathbb{R}^n : \langle x, u \rangle \leq b\}$, and its pre-image of $\tilde{\pi}$ , $\tilde{\pi}^{-1}(H_b)$ is a nonempty cap.

For $0 < \epsilon < \sqrt{m} - b$, we have

$(\tilde{\pi}^{-1}(H_b))_\epsilon = \tilde{\pi}^{-1}(H_b + \tau(b, \epsilon)O_n) = \tilde{\pi}^{-1}(H_{b+\tau(b,\epsilon)})$, where

$$b + \tau = \sqrt{m} \cos\left(\cos^{-1} \frac{b}{\sqrt{m}} \pm \frac{\epsilon}{\sqrt{m}}\right)$$

which, taking limts in the addition formula for the cosine, immediately gives

$\lim_{m \to \infty} \tau(b, \epsilon) = \epsilon$

# Proof of Theorem 2.2.3

## Proof.

Take $b < a = \Phi^{-1}(\gamma_n(A))$ , so that $H_b \subset H = \{x \in \mathbb{R}^n : \langle x, u\rangle \leq a\}$.

Then, by Poincare's lemma,

$$\lim_m \mu_{n+m}\left(\tilde{\pi}_m^{-1}(A)\right) = \gamma_n(A) > \gamma_n(H_b) = \lim_m \mu_{n+m}\left(\tilde{\pi}_m^{-1}(H_b)\right),$$

That is, for all m large enough, $\mu_{n+m}(\tilde{\pi}_m^{-1}(A)) \geq \mu_{n+m}(\tilde{\pi}_m^{-1}(H_b))$, by the isoperimetric inequality for $\mu_{n+m}$ (Theorem 2.2.1) yields that for each $\epsilon > 0, b + \epsilon < \sqrt{m}$,

$$\mu_{n+m}\left((\tilde{\pi}_m^{-1}(A))_\epsilon\right) \geq \mu_{n+m}\left(\tilde{\pi}_m^{-1}((H_b))_\epsilon\right) = \mu_{n+m}\left(\tilde{\pi}_m^{-1}(H_{b+\tau(b,\epsilon)})\right)$$

By Poincare's lemma again,

$$\gamma_n(A+\epsilon O_n) \geq \limsup_m \mu_{n+m}\left((\tilde{\pi}_m^{-1}(A))_\epsilon\right) \geq \limsup_m \mu_{n+m}\left(\tilde{\pi}_m^{-1}(H_{b+\tau(b,\epsilon)})\right) = \gamma_n(H_{b+\epsilon})$$

Since, this holds for all b<a, it also holds with b replaced by a. □

# Theorem 2.2.4

Theorem 2.2.4

*Let A a Borel set of* $\mathbb{R}^{\mathbb{N}}(A \in \mathcal{C})$*, and* $\gamma$ *be the probability law of* $(g_i : i \in \mathbb{N})$*.*

*Let O denote the unit ball about zero of* $l_2 \subset \mathbb{R}^{\mathbb{N}}$, $O = \{x \in \mathbb{R}^{\mathbb{N}} : \sum_i x_i^2 \leq 1\}$.

*Then, for all* $\epsilon > 0$,

$$\gamma(A + \epsilon O) \geq \Phi(\Phi^{-1}(\gamma(A)) + \epsilon). \tag{2.14}$$

# Proof of Theorem 2.2.4

The proof is indicated in Exercises 2.2.5 throug 2.2.7.

- Exercise 2.2.5
  - Let $\pi_n : \mathbb{R}^{\mathbb{N}} \mapsto \mathbb{R}^n$ be the projection. Then, show that (a) $\gamma_n = \gamma \circ \pi_n^{-1}$, (b) if $K \subset \mathbb{R}^{\mathbb{N}}$ is compact, then $K = \cap_{n=1}^{\infty}(\pi_n(K))$, and (c) K+tO, where O is the closed unit ball of $l_2$, is compact if K is
- Exercise 2.2.6
  - Use Theorem 2.2.3 and Exercise 2.2.5 to prove Threorem 2.2.4 in the particular case where A is a compact set
- Exercise 2.2.7
  - Since $\mathbb{R}^{\mathbb{N}}$ is polish, it follows that $\gamma$ is tight(Proposition 2.1.4). Use this and Exercise 2.2.2 to prove Theorem 2.2.4 for any $A \in \mathcal{C}$

## 2.2.3 Application to Gaussian Concentration

## 2.2.3 Application to Gaussian Concentration

- Translate the isoperimetric inequality into a concentration inequality for funcion $\{g_i\}_{i=1}^n$ about their medians.
- A bound for $\gamma(|f(x) - M| > \epsilon\}$ for all $\epsilon > 0$.
- For Lipschitz functions f about their medians
- For the supremum norm of a separable Gaussian process X, when $\sup_{t \in T} |X(t)| < \infty \; a.s.$

## 2.2.3 Application to Gaussian Concentration

### Definition 2.2.5

*A function $f : \mathbb{R}^{\mathbb{N}} \mapsto \mathbb{R}$ is Lipschitz in the direction of $l_2$, or $l_2$-Lipschitz for short, if it is meaurable and if*

$$\|f\|_{Lip2} := \sup\left\{\frac{|f(x) - f(y)|}{|x-y|} : x, y \in \mathbb{R}^{\mathbb{N}}, x \neq y, x - y \in l_2\right\} < \infty$$

*where $|x-y|$ is the $l_2$ norm of x-y.*

# Theorem 2.2.6

## Theorem Theorem 2.2.6

*If f is an $l_2$-Lipschitz function on* $\mathbb{R}^{\mathbb{N}}$, *and if* $M_f$ *is its median with respect to* $\gamma$, *then,*

$$\begin{aligned}\gamma\{x : f(x) \geq M_f + \epsilon\} \leq (1 - \Phi(\epsilon/\|f\|_{Lip2})),\\ \gamma\{x : f(x) \leq M_f - \epsilon\} \leq (1 - \Phi(\epsilon/\|f\|_{Lip2})),\end{aligned} \tag{2.15}$$

*in particular,*

$$\gamma\{x : |f(x) - M_f| \geq \epsilon\} \leq 2(1 - \Phi(\epsilon/\|f\|_{Lip2})) \leq e^{-\epsilon^2/2\|f\|^2_{Lip2}} \tag{2.16}$$

*for all* $\epsilon > 0$

## Proof of 2.2.6

### Proof.

Let $A^+ = \{x \in \mathbb{R}^{\mathbb{N}} : f(x) \geq M_f\}$ and $A^- = \{x \in \mathbb{R}^{\mathbb{N}} : f(x) \leq M_f\}$.

Then $\gamma(A^+) \geq 1/2, \gamma(A^-) \geq 1/2$.

If $x \in A^+ + \epsilon O$, then there exists $h \in O$ such that $x - \epsilon h \in A^+$,

hence, $f(x - \epsilon h) \geq M_f$ and $f(x) + \epsilon \|f\|_{Lip2} \geq f(x - \epsilon h) \geq M_f$; that is

$A^+ + \epsilon O \subset \{x : f(x) \geq M_f - \epsilon \|f\|_{Lip2}\}$.

Then the Gaussian isoperimetric inequality for $A = A^+$ gives (recall $\Phi^{-1}(1/2) = 0$)

$$\gamma\{f < M_f - \epsilon \|f\|_{Lip2}\} \leq 1 - \gamma(A^+ + \epsilon O) \leq 1 - \Phi(\epsilon),$$

,which is the second inequality in (2.15)

Likewise, $A^- + \epsilon O \subset \{x : f(x) \leq M_f + \epsilon \|f\|_{Lip2}\}$ gives the first inequality in (2.15)

Finally (2.16) follows by $(2(1 - \Phi(x)) \leq e^{-u^2/2}, Exercise 2.2.8)$ □

# Theorem 2.2.7

Theorem 2.2.7 (The Borell-Sudakov-Tsirelson concentraion inequality for Gaussian process)

*Let* $X(t), t \in T$*, be a centred separable Gaussian process such that* $Pr\{\sup_{t\in T} |X(t)| < \infty\} > 0$*, and let* $M$ *be the median of* $\sup_{t\in T} |X(t)|$ *and* $\sigma^2$ *the supremum of the variances* $EX^2(t)$*. Then for all* $\epsilon > 0$,

$$Pr\left\{\sup_{t\in T} |X(t)| > M + u\right\} \leq 1 - \Phi(u/\sigma), Pr\left\{\sup_{t\in T} |X(t)| < M - u\right\} \leq 1 - \Phi(u/\sigma) \tag{2.17}$$

*and hence,*

$$Pr\left\{\left|\sup_{t\in T} |X(t)| - M\right| > u\right\} \leq 2(1 - \Phi(u/\sigma)) \leq e^{-u^2/2\sigma^2} \tag{2.18}$$

## Proof of Theorem 2.2.7

### Proof.

$X(t), t \in T$, be a separable centred Gaussian process such that $Pr\{\sup_{t\in T}|X(t)| < \infty\} > 0$. Then $\sup_{t\in T}|X(t)| = \sup_{t\in T_0}|X(t)| < \infty$ *a.s.*, where $T_0 = \{t_k\}_{k=1}^{\infty}$ is a countable subset of T.

Ortho-normalizing (in $L^2(\mathrm{Pr})$), the jointly normal sequence $\{X(t_k)\}$ yields $X(t_k) = \sum_{i=1}^{k} a_{ki} g_i$, then there exists a version $\tilde{X}$,

$\tilde{X} : \mathbb{R}^{\mathbb{N}} \mapsto \mathbb{R}, \tilde{X}(t_k, x) = \sum_{i=1}^{k} a_{ik} x_i$

Now, define a function $f : \mathbb{R}^{\mathbb{N}} \mapsto \mathbb{R}$ by

$$f(x) = \sup_k \left| \sum_{i=1}^{k} a_{ki} x_i \right|$$

The probability law of f under $\gamma$ is the same as the law of $\sup_{t\in T_0}|X(t)|$, and $\sup_{t\in T}|X(t)|$.

# Proof of Theorem 2.2.7

Proof.

If $h \in O$, th unit ball of $l_2$, by Cauchy-Schwarz,

$$|f(x+h)-f(x)|^2 \leq \sup_k \left|\sum_{i=1}^k a_{ki}h_i\right| \leq \sup_k \left[\sum_{i=1}^k a_{ki}^2 \sum_{i=1}^k h_i^2\right] \leq \sup_k \sum_{i=1}^k a_{ki}^2 = \sup_k EX^2(t_k)$$

Therefore,

$$\|f\|_{Lip2} \leq \sigma^2, \ where\ \sigma^2 := \sigma^2(X) := \sup_{t \in T} EX^2(t)$$

Then, applies Theorem 2.2.6 to the fucntion f. □

If we integrate in (2.18) and let g be a N(0,1) random variable, we obtain,

$$\left| E\sup_{t\in T}|X(t)| - M \right| \leq E\left|\sup_{t\in T}|X(t)| - M\right| \leq \sigma E\,|g| = \sqrt{2/\pi}\sigma, \qquad (2.19)$$

an inequality which is interesting in its own right and which gives, by combining with the same(2.18),

$$Pr\left\{ \left| \sup_{t\in T}|X(t)| - E\sup_{t\in T}|X(t)| \right| > u + \sqrt{2/\pi}\sigma \right\} \leq e^{-u^2/2\sigma^2} \qquad (2.20)$$

which is of the right order for large values of u. We complete this section with simple applications of Theorem 2.2.7 to integrability and moments of the supremum of a Gaussian processes.

## Corollary 2.2.8

*Let $X(t), t \in T$, be a Gaussian process as in Theorem 2.2.7. Let $M$ and $\sigma$ also be as in this theorem, and write $\|X\| := \sup_{t \in T} |X(t)|$ to ease notation. Then there exists $K < \infty$ such that with the same hyperthesis and notation as in the preceding corollary, for all $p \geq 1$,*

$$(E(\|X\|^p)^{1/p} \leq 2E\|X\| + (E\,|g|^p)^{1/p} \leq K\sqrt{p}E\|X\|$$

*for some absolute constant $K$.*

## Proof.

Just integrate inequality (2.18) with respect to $pt^{p-1}dt$ and then use that $M \leq 2E\|X\|$(by Chebyshev) and that $\sigma \leq \sqrt{\pi/2} \sup_{t \in T} E\,|X(t)|$. (and see Exercise 2.1.2) □

# Corollary 2.2.9

Corollary 2.2.9

*Let* $X(t), t \in T$*, be a Gaussian process as in Theorem 2.2.7, and let* $\|X\|$*,* $M$ *and* $\sigma$ *be as in Corollary 2.2.8. Then,*

$$\lim_{u\to\infty} \frac{1}{u^2} \log Pr\{\|X\| > u\} = -\frac{1}{2\sigma^2}$$

*and*

$$Ee^{\lambda\|X\|^2} < \infty \text{ if and only if } \lambda < \frac{1}{\sigma^2}$$

## Proof of Corollary 2.2.9

### Proof.

The first limit follows from the facts that the first inequality in (2.17) can be rewritten as

$$\frac{1}{(u-M)^2}\log Pr\{\|x\| > u\} \leq -\frac{1}{\sigma^2}$$

$Pr\{\|X\| > u\} \geq Pr\{|X(t)| > u\}$ for all $t \in T$ and for a N(0,1) variable g, we do have $u^{-2}\log Pr\{|g| > u/a\} \to -1/2a^2$(by $l'H\hat{o}pital's\ rule$).

For the second statement, just apply the first limit to $Ee^{\lambda\|X\|} = 1 + \int_0^\infty \int_0^{\lambda\|X\|^2} e^v dv d\mathcal{L}(\|X\|)(u) = 1 + \int_0^\infty e^v Pr\{\|X\| > \sqrt{v/\lambda}\}dv$ □