Mathematical Foundations of Infinite-Dimensional Statistical Models:

2.4 Anderson's Lemma, Comparison and Sudakov's Lower Bound

presented by Boyoung Kim

Seoul National University

August 05, 2018

Table of Contents

Anderson's Lemma, Comparison and Sudakov's Lower Bound

- Obtain a lower bound for $E \sup_t X(t)$

  in terms of the metrix entropy of the (pseudo-)metric space (T, d_X), where X is a Gaussian process and $d_X^2(s,t) = E(X(t) - X(s))^2$.

- Anderson's Inequality
  - It is regarding the probability, relative to a certered Gaussian measure on $\mathbb{R}^n$, of a convex symmetric set and its translates,
  - It is related to the fact that centered Gaussian measures on $\mathbb{R}^n$ are log-concave.

- Slepian's lemma : comparing the distributions of the suprema of the processes.

2.4.1 Anderson's Lemma

Lemma 2.4.1

- A set C in vector space is convex and symmetric if $\sum_{i=1}^{n} \lambda_i x_i \in C$ whenever $x_i \in C$ and $\lambda_i \in \mathbb{R}$ satisfy $\sum_{i=1}^{n} |\lambda_i| = 1, n < \infty$.
- Given two sets A and B in vector space, their Minkowski addition is $A + B = \{x + y : x \in A, y \in B\}$ and λA is defined as $\lambda A = \{\lambda x : x \in A\}$. In this subsection, m will stand for Lebesgue measure on $\mathbb{R}$ for any n.

Lemma 2.4.1

Let A and B be Borel measurable sets in $\mathbb{R}$. Then

$$m(A + B) \geq m(A) + m(B).$$

Theorem 2.4.2 (Precopa-Leindler theorem)

Let f, g, φ be Lebesgue measurable functions on $\mathbb{R}^n$ taking values in $[0, \infty]$ and satisfying, for some $0 < \lambda < 1$ and all $u, v \in \mathbb{R}^n$,

$$\varphi(\lambda u + (1-\lambda)v) \geq f^{\lambda}(u) g^{1-\lambda}(v). \tag{2.49}$$

Then

$$\int \varphi \, dm \geq \Big(\int f \, dm \Big)^{\lambda} \Big(\int g \, dm \Big)^{1-\lambda}. \tag{2.50}$$

Proof.

The proof is by induction on the dimension n.

For $n = 1$, theorem 2.4.1 is used. □

Log-concavity of Gaussian measures in $\mathbb{R}^n$

Centered Gaussian measures on $\mathbb{R}^n$ are log concave.

Theorem 2.4.3 (Log-concavity of Gaussian measures in $\mathbb{R}^n$)

Let μ be a centered Gaussian measure on $\mathbb{R}^n$. Then, for any Borel sets A, B in $\mathbb{R}^n$ and $0 \leq \lambda \leq 1$, we have

$$\mu(\lambda A + (1-\lambda)B) \geq (\mu(A))^{\lambda}(\mu(B))^{1-\lambda}. \tag{2.51}$$

Theorem 2.4.4 (Anderson's lemma)

Let $X = (g_1, \ldots, g_n)$ *be a centred jointly normal vector in* $\mathbb{R}^n$*, and let* C *be a measurable convex symmetric set of* $\mathbb{R}^n$*. Then, for all* $x \in \mathbb{R}^n$*,*

$$Pr\{X + x \in C\} \leq Pr\{X \in C\}. \tag{2.53}$$

Proof.

Let $\mu = L(X)$. Let $A = C + x$, $B = C - x$, and $\lambda = 1/2$ in (2.51) and by summetry of μ and symmetry of C, $\mu(A) = \mu(B)$. So we obtain
$\mu(C) \geq \mu(C + x)$. □

Theorem 2.4.5

Theorem 2.4.4 Extends to infinite dimensions.

Theorem 2.4.5

Let B be a separable Banach space, let X be a B-valued centred Gaussian random variable and let C be a closed, convex, symmetric subset of B. Then, for all $x \in B$,

$$Pr\{X + x \in C\} \leq Pr\{X \in C\}. \tag{2.53}$$

In particular, $Pr\{\|X\| \leq \epsilon\} > 0$, for all $\epsilon > 0$.

Proof.

By Hahn-Banach separation theorem, ... For the last claim, apply the first part to closed balls $C_i = \{x : \|x - x_i\| \leq \epsilon\}$ for x_i countable dense subset of B. □

2.4.2 Slepian's Lemma and Sudakov's Minorisation

2.4.2 Slepian's Lemma and Sudakov's Minorisation

- Useful identity regarding derivatives of the multidimensional normal density.
- Let $f(C,x) = ((2\pi)^n detC)^{-1/2} e^{-xC^{-1}x^T/2}$ be the $N(0, C)$ density in $\mathbb{R}^n$, where $C = (C_{ij})$ is an $n \times n$ symmetric positive definite matrix $x = (x_1, \ldots, x_n)$. Then

$$\frac{\partial f(C,x)}{\partial C_{ij}} = \frac{\partial^2 f(C,x)}{\partial x_i x_j} = \frac{\partial^2 f(C,x)}{\partial x_j x_i}, 1 \leq i < j \leq n. \tag{2.54}$$

Theorem 2.4.7

Let $X = (X_1, \ldots, X_n)$ *and* $Y = (Y_1, \ldots, Y_n)$ *be centred normal vectors in* $\mathbb{R}^n$ *such that* $EX_i^2 = EY_j^2 = 1, 1 \le i, j \le n$*. Set, for each* $1 \le i < j \le n$*,* $C_{ij}^1 = E(X_iX_j), C_{ij}^0 = E(Y_iY_j)$ *and* $\rho_{ij} = \max\{|C_{ij}^0|, |C_{ij}^1|\}$*. Then for any* $\lambda_i \in \mathbb{R}$*,*

$$Pr\bigcap_{i=1}^{n}\{X_i \le \lambda_i\} - Pr\bigcap_{i=1}^{n}\{Y_i \le \lambda_i\} \le \frac{1}{2\pi}\sum_{1\le i<j\le n}(C_{ij}^1 - C_{ij}^0)^+ \frac{1}{(1-\rho_{ij}^2)^{1/2}} exp\left(-\frac{(\lambda_i^2+\lambda_j^2)/2}{1+\rho_{ij}}\right). \tag{2.55}$$

Moreover, if $\mu_i \le \lambda_i$ *and* $\nu = min\{|\lambda_i|, |\lambda_i| : i = 1, \ldots, n\}$*, then*

$$\left|Pr\bigcap_{i=1}^{n}\{\mu_i \le X_i \le \lambda_i\} - Pr\bigcap_{i=1}^{n}\{\mu_i \le Y_i \le \lambda_i\}\right| \le \frac{2}{\pi}\sum_{1\le i<j\le n}|C_{ij}^1 - C_{ij}^0| \frac{1}{(1-\rho_{ij}^2)^{1/2}} exp\left(-\frac{\nu^2}{1+\rho_{ij}}\right). \tag{2.56}$$

Slepian's lemma

It allows comparing the distributions of the suprema of $X(t)$ and $Y(t)$ if the covariance of one of the processes dominates the other.

Theorem 2.4.8 (Slepian's lemma)

Let $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_n)$ be centred jointly normal vectors in $\mathbb{R}^n$ such that

$$E(X_i X_j) \le E(Y_i Y_j) \text{ and } EX_i^2 = EY_i^2 \text{ for } , 1 \le i, j \le n. \tag{2.58}$$

Then, for any $\lambda_i \in \mathbb{R}$, $i \le n$,

$$Pr\left(\bigcap_{i=1}^{n} \{Y_i > \lambda_i\}\right) \le Pr\left(\bigcap_{i=1}^{n} \{X_i > \lambda_i\}\right), \tag{2.59}$$

and therefore,

$$E \max_{i \le n} Y_i \le E \max_{i \le n} X_i \tag{2.60}$$

Remark 2.4.9

Comparison of expected values of the maximum of absolute values.

Remark

For X_i symmetric, for any $i_0 \in \{1, \ldots, n\}$,

$$E \max_{i \le n} X_i \le E \max_{i \le n} |X_i| \le E|X_{i_0}| + E \max_{i,j} |X_i - X_j| \le E|X_{i_0}| + 2E \max_{i \le n} X_i.$$

Corollary 2.4.10

It is easier because it does not require $EX_i^2 = EY_i^2, i \leq n$.

Corollary 2.4.10

Let $X = (X_1, \ldots, X_n)$ *and* $Y = (Y_1, \ldots, Y_n)$ *be two centred, jointly normal vectors in* $\mathbb{R}^n$*, and assume that*

$$E(Y_i - Y_j)^2 \leq E(X_i - X_j)^2, i, j \in \{1, \ldots, n\}.$$

Then

$$E \max_{i \leq n} Y_i \leq 2E \max_{i \leq n} X_i.$$

Lemma 2.4.11

Finally, we will apply the comparison results to obtain a lower bound for $E\sup_t X(t)$ where X is a Gaussian process.

The entropy lower bound will follow from the following evaluation of the maximum of a finite number of independent normal variables.

Lemma 2.4.11

Let $g_i, i \in \mathbb{N}$, be independent standard normal random variables. Then

a. $\lim_{n\to\infty} \frac{E\max_{i\le n}|g_i|}{\sqrt{2\log n}} = 1$, *and*

b. *There exists $K < \infty$ such that, for all $n > 1$,*

$$K^{-1}(\log n)^{1/2} \le E\max_{i\le n} g_i \le E\max_{i\le n}|g_i| \le K(\log n)^{1/2}.$$

Sudakov's lower bound

Recall that given a (pseudo-) metric space (T, d), $N(T, d, \epsilon)$ denotes the ϵ-covering number of (T, d) and the logarithm of the covering number of (T, d) is known as its metric entropy.

Theorem 2.4.12 (Sudakov's lower bound)

There exists $K < \infty$ such that if $X(t), t \in T$, is a centred Gaussian process and $d_x(s, t) = (E(X(t) - X(s))^2)^{1/2}$ denotes the associated pseudo-metric on T, then, for all $\epsilon > 0$,

$$\epsilon \sqrt{log \ N(T, d_x, \epsilon)} \leq K \sup_{S \subset T, S \ finite} E \max_{t \in S} X(t).$$

Corollary 2.4.13 (Sudakov's theorem)

Let $X(t), t \in T$, be a centred Gaussian process, and let d_X be the associated pseudo-distance. If $\liminf_{\epsilon \downarrow 0} \epsilon \sqrt{logN(T, d_X, \epsilon)} = \infty$, *then* $\sup_{t \in T} |X(t)| = \infty$ *a.s., so X is not sample bounded.*

This corollary shows that if a centered Gaussian process X is sample bounded, then the covering numbers $N(T, d_X, \epsilon)$ are all finite.

Corollary 2.4.14

Stronger version :

Corollary 2.4.14

Let $X(t), t \in T$, be a sample continuous centred Gaussian process. Then,

$$\lim_{\epsilon \to 0} \epsilon \sqrt{logN(T, d_X, \epsilon)} = 0$$

.

Comparison with the upper bound in Theorem 2.3.6.

Lower bound for $E\sup_{t\in T}|X(t)|$ in Theorem 2.4.12 and the upper bound in Theorem 2.3.6 for X a centred Gaussian process with $X(t_0)=0$ a.s. for some $t_0 \in T$.

Note that if $logN(T, d_X, 1/\tau)$ is bounded above and below by a constant times a regularly varing function of τ, then both bounds combine to give that there exists $K < \infty$ such that

$$\frac{1}{K}\sigma_X\sqrt{logN(T, d_x, \sigma_X)} \leq E\sup_{t\in T}|X(t)| \leq K\sigma_X\sqrt{logN(T, d_x, \sigma_X)} \qquad (2.61)$$

(Appendix) Theorem 2.3.6

Theorem 2.3.6

Let (T, d) be a pseudo-metric space, and let $X(t), t \in T$ be a stochastic process sub-Gaussian with respect to the pseudo-distance d, that is, one whose increments satisfy condition (2.36). Then, for all finite subsets $S \subseteq T$ and points $t_0 \in T$, the following inequalities hold:

$$E \max_{t \in S} |X(t)| \leq E|X(t_0)| + 4\sqrt{2} \int_0^{D/2} \sqrt{log 2N(T, d, \epsilon)} d\epsilon, \tag{2.37}$$

where D is the diameter of (T, d), and

$$E \max_{s,t \in S, d(s,t) \leq \delta} |X(t) - X(s)| \leq (16\sqrt{2} + 2) \int_0^{\delta} \sqrt{log 2N(T, d, \epsilon)} d\epsilon, \tag{2.38}$$

for all $\delta > 0$, where the integrals are taken to be 0 if $D = 0$.

(Appendix) Definition 2.2.1, 2.2.3

Definition 2.2.3

확률과정 $X_t(t \in T)$ *가 표본유계 (sample bounded)* $\Leftrightarrow$ *어떤* X *의 version* $\tilde{X}$ *의 sample path 들이 거의 모든* ω *에 대해 고르게 유계, 즉* $\sup_{t \in T} |\tilde{X}| < \infty$*a.s..*

(T, d) *가 (유사) 거리공간일 때* X *가 표본연속 (sample continuous)* $\Leftrightarrow$ *어떤* X *의 version* $\tilde{X}$*의 sample path 들이 거의 모든* ω *에 대해 유계이고 고르게 연속.*

Definition 2.2.1

지표집합 T*에 대한 확률과정* X, Y *가 각각의 유한차원 분포가 같으면 각각을 서로에 대한 version 이라 한다.*