Mathematical Foundations of Infinite-Dimensional Statistical Models

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

XIN Baiying

2024/12/11

2.4.1 Anderson’s Lemma

2.4.1 Anderson's Lemma : Intuition

- Anderson's lemma focuses on **centered Gaussian measures** on **convex and symmetric sets**. Thus first define convex and symmetric sets.

Definition (Convex and Symmetric Set)

A set C in a real vector space is called convex, if $\sum_{i=1}^{n} \lambda_i x_i \in C$ for all $x_i \in C$ and $\lambda_i \in \mathbb{R}$ with $\sum_{i=1}^{n} |\lambda_i| = 1$.

- Convex: Intuitively, a set is convex, then the line segment between any two points in the set is also in the set.
- Symmetric: if $x \in C$, then $-x \in C$.

Example (Balls centered at the origin in Banach spaces)

$$\{x : \|x\| \leq r\}$$

2.4.1 Anderson's Lemma: Intuition

Then intuitevely, for a centered Gaussian measure μ on $\mathbb{R}^n$, if set C is measurable, convex and symmetric, then

$$\mu(C+x) \leq \mu(C) \quad \text{for all } x \in \mathbb{R}^n.$$

- $C+x=\{y+x : y \in C\}$, which is the set obtained by translating C in the direction of x.
- As C is symmetric, the origin must be in C, so $C+x$ can be seen as the set away from the origin by x.

2.4.1 Anderson's Lemma: Intuition

Specifically, assume Y, Z are two independent centered Gaussian random vectors in $\mathbb{R}^n$.

- i.e. Y, Z has 0 mean and covariance matrix C_Y, C_Z respectively.

Further define a combined random vector $X = Y + Z$, with covariance matrix $C_Z = C_X - C_Y$ being non-negative definite.

Then, we have:

$$\Pr\{X \in C\} = \int \Pr\{Y \in C - z\} \mathrm{d}\mathcal{L}_Z(z)$$

- This is because $Y = X - Z$, thus $X \in C \Leftrightarrow Y \in C - z$. Then $\Pr\{Y \in C - z\} = \Pr\{X \in C \mid Z = z\}$. Finally integrate over all possible z.

2.4.1 Anderson’s Lemma: Intuition

(Cont.)

And an inequality:

$$\Pr\{X \in C\} = \int \Pr\{Y \in C - z\} \mathrm{d}\mathcal{L}_Z(z) \le \int \Pr\{Y \in C\} \mathrm{d}\mathcal{L}_Z(z) = \Pr\{Y \in C\}$$

- Intuitively, it can be regarded as Y is a Gaussian random vector 'around' C, and Z is a (symmetric, centered) Gaussian noise added to Y. After convolution, the "effective" mass of $Y + Z$ in C has been reduced. Thus the probability of observing X in C is no more that of Y in C.

2.4.1 Anderson’s Lemma: Brunn-Minkovski Inequality

The modern proof of Anderson’s lemma uses **Brunn-Minkovski** inequality, expressing a **log-concavity** property of some certain functions.

Thus, first introduce some basic concepts w.r.t. **Brunn-Minkovski** inequality for Lebesgue measure in $\mathbb{R}^n$.

Given two sets A, B in vector space, define:

- Minkovski sum: $A + B = \{a + b : a \in A, b \in B\}$.
- λ-dilation: $\lambda A = \{\lambda a : a \in A\}$.

2.4.1 Anderson's Lemma: Brunn-Minkovski Inequality

> † In the following content, m will stand for the Lebesgue measure on $\mathbb{R}$ for any n-dimensional set.

Lemma 2.4.1

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

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

Proof

Generally, it can be proved by the following steps:

1. **Show that $A + B$ is Lebesgue measurable.**
2. **W.L.O.G, assume A, B are compact.**
3. **Translate A, B to $A \subset \{x \leq 0\}, B \subset \{x \geq 0\}$.**
4. **Show that $m(A+B) \geq m(A \cup B) = m(A) + m(B)$.**

2.4.1 Anderson's Lemma: Brunn-Minkovski Inequality

Proof (cont.):

- **Show that $A + B$ is Lebesgue measurable.**
  - As A, B are Borel measurable, then $A \times B$ is Borel measurable in $\mathbb{R}^2$.
  - $A + B$ is the image of a continuous mapping $(x, y) \mapsto x + y$ from $A \times B$ to $\mathbb{R}$.
  - By the property of Borel sets: any continuous mapping of Borel sets are analytic sets; analytic sets on $\mathbb{R}$ are always Lebesgue measurable.
- **W.L.O.G, assume A, B are compact.**
  - m is a Lebesgue measure on $\mathbb{R}$ and thus regular, so we can approximate A, B by compact sets.

2.4.1 Anderson's Lemma: Brunn-Minkovski Inequality

Proof (cont.):

- **Translate** A, B **to** $A \subset \{x \leq 0\}, B \subset \{x \geq 0\}$.
  - Fact: For Lebesgue measure, translation does not change the measure of a set.
  - Re-define $A := A + \{-\sup A\}$, $B := B + \{-\inf B\}$.
  - Then $A \subset \{x \leq 0\}$, $B \subset \{x \geq 0\}, A \cap B = \{0\}$.
- **Show that** $m(A + B) \geq m(A \cup B) = m(A) + m(B)$.
  - $A + B \supseteq A \cup B$, thus $m(A + B) \geq m(A \cup B)$.
  - $m(A \cup B) = m(A) + m(B)$ as $A \cap B = \{0\}$.

$\square$

2.4.1 Anderson's Lemma: Précopa-Leindler Theorem

Précopa-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 $u, v \in \mathbb{R}^n$,

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

Then

$$\int \varphi \mathrm{d}m \geq \left(\int f \mathrm{d}m\right)^\lambda \left(\int g \mathrm{d}m\right)^{1-\lambda} \quad (2.50)$$

- Précopa-Leindler Theorem is a generalization of the classical Hölder inequality.
- Intuitively, (2.49) is a log-concavity property of φ w.r.t. f, g. And Précopa-Leindler shows that such property also holds for the integral perspective.

2.4.1 Anderson's Lemma: Précopa-Leindler Theorem

Précopa-Leindler Theorem (cont.)

Proof:

It can be proved by induction on the number of dimensions n.

For $n = 1$, the inequality is proved from inequality (2.49) with Lemma 2.4.1.

- W.L.O.G, assume $\|f\|_\infty = \|g\|_\infty = 1$.
- Define two sets $\{u : f(u) \geq t\}$ and $\{v : g(v) \geq t\}$. Then $\lambda\{f \geq t\} + (1-\lambda)\{g \geq t\} \subseteq \{\varphi \geq t\}$.
  - By definition, $f^\lambda(u) g^{1-\lambda}(v) \geq t^\lambda t^{1-\lambda} = t$.
  - By (2.49), $\varphi(\lambda u + (1-\lambda)v) \geq f^\lambda(u) g^{1-\lambda}(v) \geq t$.
  - Thus, $\lambda u + (1-\lambda)v \in \{w : \varphi(w) \geq t\}$.

2.4.1 Anderson's Lemma: Précopa-Leindler Theorem

Précopa-Leindler Theorem - Proof (cont. for $n = 1$)

- By Lemma 2.4.1 ($m(A+B) \geq m(A) + m(B)$) and fact $m(\lambda A) = \lambda^n m(A)$:

$$m(\{\varphi \geq t\}) \geq \lambda m(\{f \geq t\}) + (1-\lambda) m(\{g \geq t\})$$

- Integrate the last inequality over t:

$$\int_0^\infty m(\{\varphi \geq t\}) \mathrm{d}t \geq \int_0^\infty \lambda m(\{f \geq t\}) + (1-\lambda) m(\{g \geq t\}) \mathrm{d}t$$

- By definition of measure:

$$\int \varphi \mathrm{d}m \geq \lambda \int f \mathrm{d}m + (1-\lambda) \int g \mathrm{d}m$$

- By concavity of log function ($\lambda a + (1-\lambda) b \geq a^\lambda b^{1-\lambda}$):

$$\int \varphi \mathrm{d}m \geq \left(\int f \mathrm{d}m\right)^\lambda \left(\int g \mathrm{d}m\right)^{1-\lambda} \quad \square$$

2.4.1 Anderson's Lemma: Précopa-Leindler Theorem

Précopa-Leindler Theorem - Proof (cont. for $n-1$ to n)

- Assume that the result holds for $n-1$, now proves it also holds on n.
- In $\mathbb{R}^n$, fix a one dimension's coordinate, say, x_n. Then rewrite x as (x', x_n), where $x' \in \mathbb{R}^{n-1}$. Then re-define $\varphi_{x_n}(x') = \varphi(x', x_n)$, $f_{x_n}(x') = f(x', x_n)$, $g_{x_n}(x') = g(x', x_n)$.
  - Then as x_n is fixed, use the hypothesis on $n-1$ dimensions:

$$\int_{\mathbb{R}^{n-1}} \varphi_{x_n} \mathrm{d}m_{n-1} \geq \left(\int f_{x_n} \mathrm{d}m_{n-1}\right)^{\lambda} \left(\int g_{x_n} \mathrm{d}m_{n-1}\right)^{1-\lambda}$$

- Then integrate over x_n with Fubini Theorem and by induction hypothesis:

$$\int_{\mathbb{R}^n} \varphi \mathrm{d}m \geq \left(\int_{\mathbb{R}^n} f \mathrm{d}m\right)^{\lambda} \left(\int_{\mathbb{R}^n} g \mathrm{d}m\right)^{1-\lambda} \quad \square$$

2.4.1 Anderson's Lemma: Log-concavity of Gauss. Measures in $\mathbb{R}^n$

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 $0 \leq \lambda \leq 1$,

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

- Intuitively, this theorem shows that, for two sets A, B in $\mathbb{R}^n$, if we find a 'average' set $\lambda A + (1 - \lambda)B$ (by convex combination), then the measure of this average set is no less than some 'geometric average' of the measures of A, B.

2.4.1 Anderson's Lemma: Log-concavity of Gauss. Measures in $\mathbb{R}^n$

Proof ($\mu(\lambda A+(1-\lambda)B) \geq \mu(A)^\lambda \mu(B)^{1-\lambda}$)

- Assume μ is supported by a subspace $V \subset \mathbb{R}^n$. And on V, the density of μ is $\phi(x)=c\exp(-|\Gamma x|^2/2)$, where $\Gamma: V \mapsto V$ is defined as $\Gamma=\Sigma^{-1/2}$, where Σ is the covariance matrix of μ; Γ is a strictly positive definite operator.
  - Intuitively, $\phi(x)$ represents the weight of some x in the Gaussian measure.
- Then take logrithm, function $x \mapsto \log\phi(x)=-|\Gamma x|^2/2, x \in V$ has the property of log-concavity:

$$\phi(\lambda u+(1-\lambda)v) \geq \phi^\lambda(u)\phi^{1-\lambda}(v), \quad u,v \in V \quad (2.52)$$

  - Later we will use the **Précopa-Leindler Theorem** to prove the log-concavity of Gaussian measures. And this inequality can be checked that satisfies the condition $*$: $\varphi(\lambda u+(1-\lambda)v) \geq f^\lambda(u)g^{1-\lambda}(v)$.

2.4.1 Anderson's Lemma: Log-concavity of Gauss. Measures in $\mathbb{R}^n$

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

- Consider indicator functions $\mathbb{I}_A, \mathbb{I}_B$ of sets A, B respectively. Then define the density function: $f = \phi\mathbb{I}_{A\cap V}, g = \phi\mathbb{I}_{B\cap V}$. And the density function of $\lambda A + (1-\lambda)B$ is $\varphi = \phi_{\lambda A+(1-\lambda)B} = \phi \cdot \mathbb{I}_{\lambda(A\cap V)+(1-\lambda)(B\cap V)}$.
- Then apply Précopa-Leindler Theorem to give:

$$\int_{\lambda(A\cap V)+(1-\lambda)(B\cap V)} \phi \mathrm{d}m \geq \left(\int_{A\cap V} \phi \mathrm{d}m\right)^\lambda \left(\int_{B\cap V} \phi \mathrm{d}m\right)^{1-\lambda}$$

  where m is the Lebesgue measure on V.
- Given that $\mu(A) = \int_{A\cap V} \phi \mathrm{d}m$, and the same for $\mu(B)$ and $\mu(\lambda A + (1-\lambda)B)$. Then the inequality holds for the Gaussian measure μ on $\mathbb{R}^n$.

$\square$

2.4.1 Anderson's Lemma

Theorem 2.4.4 (Anderson's Lemma)

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

$$\Pr\{X \in C + x\} \leq \Pr\{X \in C\} \quad (2.53)$$

Proof

- Define $\mu = \mathcal{L}(X)$, the Gaussian measure of X.
- Recall (2.51) with $\lambda = \frac{1}{2}$: $\mu(\frac{A+B}{2}) \geq \mu(A)^{1/2}\mu(B)^{1/2}$ for all Borel sets.
    - Define $A = C + x, B = C - x$. As C is symmetric, $\mu(A) = \mu(B)$.
    - Bring in A, B to (2.51):

$$\mu(C) \geq \mu(C + x)^{1/2}\mu(C - x)^{1/2} = \mu(C + x)$$

    - i.e. $\Pr\{X \in C + x\} \leq \Pr\{X \in C\}$.

$\square$

2.4.1 Anderson's Lemma: Infinite-Dimensional Extension

Theorem 2.4.5 (Anderson's Lemma in Infinite-Dimensional Spaces)

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

$$\Pr\{X \in C + x\} \leq \Pr\{X \in C\}$$

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

Proof

- First apply **Hahn-Banach Separation Theorem** and the separability of Banach space to reduce the problem to a countable subset T_C :

  $$C = \cap_{f \in T_C} \{x \in B : |f(x)| \leq 1\}$$

  where $T_C \subset D_C \subset B^*$.
  - This means that, point $x \in B$ belongs to C if and only if for all linear functionals $f \in D_C$, $|f(x)| \leq 1$. And by approximation, it suffices to check only in T_C.

2.4.1 Anderson's Lemma: Infinite-Dimensional Extension

Proof (cont.)

- State that: $\{X \in C\} = \cap_{f \in T_C}\{x \in B : |f(X)| \leq 1\}$ = $\sup_{f \in T_C}\{x \in B : |f(X)| \leq 1\}$ (the last equality holds by property of set operations). Thus,

$$\Pr\{X \in C\} = \Pr\{\sup_{f \in T_C} |f(X)| \leq 1\} = \lim_{n \to \infty} \Pr\{\max_{f \in T_n} |f(X)| \leq 1\}$$

  where T_n is a finite subset of T_C, and $T_n \uparrow T_C$.

- Similarly, for $X + x$:

$$\Pr\{X \in C + x\} = \lim_{n \to \infty} \Pr\{\max_{f \in T_n} |f(X + x)| \leq 1\}$$

- Then by Theorem 2.4.4 (Finite case), for all finite T_n, $\Pr\{\max_{f \in T_n} |f(X + x)| \leq 1\} \leq \Pr\{\max_{f \in T_n} |f(X)| \leq 1\}$.
- By taking limit, $\Pr\{X \in C + x\} \leq \Pr\{X \in C\}. \square$

2.4.1 Anderson's Lemma: Infinite-Dimensional Extension

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

Proof

- Consider a dense countable subset in Banach space B: $\{x_i\}_{i\in\mathbb{N}} \subseteq B$. For each x_i, define a closed ball $C_i = \{x \in B : \|x - x_i\| \leq \epsilon\}$.
  - These C_i covers the whole space B, as $\{x_i\}$ is dense.
  - C_i is closed, convex, symmetric, and thus satisfies the Andr. Lemma.
- Then, by its density: $\Pr\{\|X\| \leq \epsilon\} = \Pr\{\cup_{i=1}^{\infty} C_i\}$
- Given that for each C_i, $\Pr\{X \in C_i\} > 0$, then $\Pr\{\|X\| \leq \epsilon\} = \Pr\{\cup_{i=1}^{\infty} C_i\} > 0$.

$\square$

2.4.1 Anderson's Lemma: Khatri-Sidak Inequality

Collary 2.4.6 (Khatri-Sidak Inequality)

Let $n \geq 2$, and let $g_1, \cdots, g_n$ be jointly normal centered random variables. Then, for all $x \geq 0$,

$$\Pr\{\max_{1 \leq i \leq n} |g_i| \leq x\} \geq \Pr\{|g_1| \leq x\} \Pr\{\max_{2 \leq i \leq n} |g_i| \leq x\}$$

and hence, iterating,

$$\Pr\{\max_{1 \leq i \leq n} |g_i| \leq x\} \geq \prod_{i=1}^{n} \Pr\{|g_i| \leq x\}.$$

2.4.1 Anderson's Lemma: Khatri-Sidak Inequality

Proof ($\Pr\{\max_{1\le i\le n}|g_i|\le x\}\ge \Pr\{|g_1|\le x\}\Pr\{\max_{2\le i\le n}|g_i|\le x\}$)

- Fact: $\Pr\{\max_{1\le i\le n}|g_i|\le x\} = \lim_{t\to\infty}\Pr\{\max_{2\le i\le n}|g_i|\le x, |g_1|\le t\}$.
- Define $f(t) := \Pr(|g_1| < t, (g_2,\cdots,g_n)\in A)$, where A is an arbitrary convex and symmetric subset of $\mathbb{R}^{n-1}$. $g(t) := \Pr(|g_1|\le t)$
- And now consider: $f(t)/g(t) = \Pr((g_2,\cdots,g_n)\in A \mid |g_1|\le t)$.
  - It suffices to show that $f(t)/g(t)$ is monotone decreasing in t:
    - As $t\to\infty$, $\Pr(|g_1|\le t)\to 1$. Then

$$\lim_{t\to\infty}\Pr(\max_{2\le i\le n}|g_i|\le x \mid |g_1|\le t) = \Pr(\max_{2\le i\le n}|g_i|\le x)$$

    - Thus as long as $f(t)/g(t)$ is monotone decreasing,

$$\Pr(\max_{2\le i\le n}|g_i|\le x \mid |g_1|\le t)\ge \Pr(\max_{2\le i\le n}|g_i|\le x).$$

    - And the Khatri-Sidak inequality can be then proved.

2.4.1 Anderson's Lemma: Khatri-Sidak Inequality

Proof (cont.)

Now prove that $f(t)/g(t)$ is monotone decreasing in t:

- Denote ϕ_1 as the density of g_1, and $X = (g_2, \cdots, g_n)$.
- Then we have:

$$\Pr\{X \in A \mid |g_1| \le t\} = \int_{-t}^{t} \Pr\{X \in A \mid g_1 = u\} \mathrm{d}\mathcal{L}_{g_1 | |g_1| \le t}(u) = \int_{-t}^{t} \Pr\{X \in A \mid g_1 = u\} \phi_1(u) \mathrm{d}u / \Pr\{|g_1| \le t\}$$

- Furthermore, there are facts that:
  - $f(t) = \int_{-t}^{t} \Pr\{X \in A \mid g_1 = u\} \phi_1(u) \mathrm{d}u$,
    $f'(t) = 2\Pr\{X \in A \mid g_1 = t\} \phi_1(t)$.
  - $\Pr\{X \in A \mid |g_1| \le t\} \le \Pr\{X \in A | g_1 = t\}$
  - And finally:

$$\begin{aligned} (f/g)'(t) &= 2\varphi_1(t) \Pr\{X \in A \mid g_1 = t\} \Pr\{|g_1| \le t\} - 2 \Pr\{|g_1| \le t, (g_2, \ldots, g_n) \in A\} \varphi_1(t) \\ &= 2\varphi_1(t) \Pr\{|g_1| \le t\} \left[\Pr\{X \in A \mid g_1 = t\} - \Pr\{X \in A \mid |g_1| \le t\}\right] \le 0. \quad \square \end{aligned}$$

2.4.2 Slepian's Lemma and Sudakov's Minorisation

2.4.2 Slepian's Lemma: Identity of Normal Density

Let $f(C, x) = [(2\pi)^n \det C]^{-1/2} \exp(-xC^{-1}x^\top/2)$ be the $\mathcal{N}(0, C)$ density in $\mathbb{R}^n$, where $C = (c_{ij})_{n\times n}$ is a symmetric positive definite matrix, $x = (x_1, \cdots, x_n)$. Then the following identity holds:

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

- The proof of this identity can be done by the inversion formula for characteristic functions of Gaussian measures.

2.4.2 Slepian's Lemma: Theorem 2.4.7

Theorem 2.4.7

Let $X = (X_1, \cdots, X_n)$ and $Y = (Y_1, \cdots, Y_n)$ be centered normal vectors in $\mathbb{R}^n$ s.t. $\mathbb{E}X_i^2 = \mathbb{E}Y_j^2 = 1$ for all i, j. Denote $C_{ij}^1 = \mathbb{E}X_iX_j$, $C_{ij}^0 = \mathbb{E}Y_iY_j$, and $\rho_{ij} = \max\{|C_{ij}^1|, |C_{ij}^0|\}$, $(x)^+ := \max(x, 0)$.

For any $\lambda_i \in \mathbb{R}$, we have:

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

Moreover, for $\mu_i \leq \lambda_i$ and $\nu = \min\{|\lambda_i|, |\mu_i| : i = 1, \ldots, n\}$, we have:

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