Mathematical Foundations of Infinite-Dimensional Statistical Models

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

XIN Baiying

2024/12/18

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:

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

- 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:

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

2.4.2 Slepian's Lemma: Theorem 2.4.7

Proof of (2.55): $P(\bigcap\{X_i \leq \lambda_i\}) - P(\bigcap\{Y_i \leq \lambda_i\}) \leq \frac{1}{2\pi}\sum(C^1_{ij} - C^0_{ij})^+ \frac{\exp(-\frac{\lambda_i^2+\lambda_j^2}{2(1+\rho_{ij})})}{(1-\rho_{ij}^2)^{1/2}}$

First we can make two assumptions to simplify the proof:

1. Covariance matrix of X and Y (C^1 and C^0) are invertible.

If necessary, we can redefine X and Y by adding a small standard Gaussian noise to make C^1 and C^0 invertible: $X_\epsilon = (1-\epsilon^2)^{1/2}X + \epsilon G$, $Y_\epsilon = (1-\epsilon^2)^{1/2}Y + \epsilon G$。

- Here G is the Standard Gaussian white noise independent of X, Y, making X_ϵ and Y_ϵ have invertible covariance matrices. And its invertibility guarantees the existence of the density function of X_ϵ and Y_ϵ.
- As $\epsilon \to 0$, $X_\epsilon \to X$ and $Y_\epsilon \to Y$ in distribution, i.e. X_ϵ, Y_ϵ can be used to approximate X, Y.

2.X and Y are independent.

- As the whole theory does not concern the joint distribution of X and Y.

2.4.2 Slepian's Lemma: Theorem 2.4.7

Proof of (2.55): $P(\bigcap\{X_i \le \lambda_i\}) - P(\bigcap\{Y_i \le \lambda_i\}) \le \frac{1}{2\pi}\sum(C_{ij}^1 - C_{ij}^0)^+ \frac{\exp(-\frac{\lambda_i^2+\lambda_j^2}{2(1+\rho_{ij})})}{(1-\rho_{ij}^2)^{1/2}}$

Under the assumptions, we can define a path connecting X and Y:

$$X(t) = t^{1/2}X + (1-t)^{1/2}Y, \quad t \in [0,1]$$

- $X(0) = Y$, $X(1) = X$. $X(t)$ connects X and Y smoothly in $\mathbb{R}^n$ by tuning t.
- $C^t = \mathrm{Cov}(X(t)) = tC^1 + (1-t)C^0$.
  - C^t is a positive definite matrix for all $t \in [0,1]$ (as a convex combination of positive definite matrices C^1 and C^0).

2.4.2 Slepian's Lemma: Theorem 2.4.7

Proof of (2.55): $P(\bigcap\{X_i \le \lambda_i\}) - P(\bigcap\{Y_i \le \lambda_i\}) \le \frac{1}{2\pi}\sum(C_{ij}^1 - C_{ij}^0)^+ \frac{\exp(-\frac{\lambda_i^2+\lambda_j^2}{2(1+\rho_{ij})})}{(1-\rho_{ij}^2)^{1/2}}$

Correspondingly, define the density function of $X(t)$ as f_t, then

$$F_{X_t}(t) = \int_{-\infty}^{\lambda_1} \cdots \int_{-\infty}^{\lambda_n} f_t(x)\mathrm{d}x, \quad (2.57)$$

which can be seen to be in $C([0,1])$.

- $F(0) = \Pr(Y_1 \le \lambda_1, \cdots, Y_n \le \lambda_n) = \Pr(\bigcap\{Y_i \le \lambda_i\})$. Similarly, so is $F(1)$.

Thus for (2.55), $LHS = F(1) - F(0)$.

And by **Fundamental Theorem of Calculus**:

$$LHS = F(1) - F(0) = \int_0^1 F'(t)\mathrm{d}t$$

2.4.2 Slepian's Lemma: Theorem 2.4.7

Proof of (2.55): $P(\bigcap\{X_i \le \lambda_i\}) - P(\bigcap\{Y_i \le \lambda_i\}) \le \frac{1}{2\pi}\sum(C^1_{ij} - C^0_{ij})^+ \frac{\exp(-\frac{\lambda_i^2+\lambda_j^2}{2(1+\rho_{ij})})}{(1-\rho_{ij}^2)^{1/2}}$.

Further derivation of $F'(t)$:

- As $F(t) = \int_{-\infty}^{\lambda_1}\cdots\int_{-\infty}^{\lambda_n} f_t(x)\mathrm{d}x$, and given the fact that integration and differentiation can be exchanged, we have: $F'(t) = \int_{-\infty}^{\lambda_1}\cdots\int_{-\infty}^{\lambda_n}\frac{\partial f_t}{\partial t}(x)\mathrm{d}x$.
- Consider $\frac{\partial f_t}{\partial t}$ by the Chain Rule: $\frac{\partial f_t}{\partial t} = \sum_{1\le i<j\le n}\frac{\partial f_t}{\partial C_{ij}}\frac{\partial C_{ij}}{\partial t}$ (As f_t is the density function of $X(t)$, and thus of course depends on C^t).
  - Plus, as $C^t = tC^1 + (1-t)C^0$, we have $\frac{\partial C_{ij}}{\partial t} = C^1_{ij} - C^0_{ij}$. Moreover, by (2.54) we have shown that: $\frac{\partial f_t}{\partial C_{ij}} = \frac{\partial^2 f_t}{\partial x_i x_j} = \frac{\partial^2 f_t}{\partial x_j x_i}$. Thus, $\frac{\partial f_t}{\partial t} = \sum_{1\le i<j\le n}\frac{\partial^2 f_t}{\partial x_j x_i}(C^1_{ij} - C^0_{ij})$
- Bring this back to $F'(t)$, we have:

$$F'(t) = \int_{-\infty}^{\lambda_1}\cdots\int_{-\infty}^{\lambda_n}\sum_{1\le i<j\le n}\frac{\partial^2 f_t}{\partial x_j x_i}(C^1_{ij} - C^0_{ij})\mathrm{d}x = \sum_{1\le i<j\le n}(C^1_{ij} - C^0_{ij})\cdot\overbrace{\int_{-\infty}^{\lambda_1}\cdots\int_{-\infty}^{\lambda_n}\frac{\partial^2 f_t}{\partial x_j x_i}\mathrm{d}x}^{\mathcal{I}}$$

2.4.2 Slepian's Lemma: Theorem 2.4.7

Proof of (2.55): $P(\bigcap\{X_i \le \lambda_i\}) - P(\bigcap\{Y_i \le \lambda_i\}) \le \frac{1}{2\pi}\sum(C^1_{ij} - C^0_{ij})^+ \frac{\exp(-\frac{\lambda_i^2+\lambda_j^2}{2(1+\rho_{ij})})}{(1-\rho_{ij}^2)^{1/2}}$.

Recall what we have derived so far:

- Under two assumptions, we can define a new random vector $X(t) = t^{1/2}X + (1-t)^{1/2}Y$, with covariance matrix $C^t = tC^1 + (1-t)C^0$, and density function f_t.
- It then can be shown that, LHS of (2.55) can be written as $LHS = \int_0^1 F'(t)\mathrm{d}t$, and $F'(t) = \sum(C^1_{ij} - C^0_{ij}) \cdot \int_{-\infty\cdots-\infty}^{\lambda_1\cdots\lambda_n} \frac{\partial^2 f_t}{\partial x_j x_i}\mathrm{d}x$

Now the key is to calculate $\mathcal{I} \triangleq \int_{-\infty}^{\lambda_1}\cdots\int_{-\infty}^{\lambda_n} \frac{\partial^2 f_t}{\partial x_j x_i}\mathrm{d}x$:

$$\mathcal{I} \le \int_{\mathbb{R}^{n-2}} \mathrm{d}x_k \cdot \int_{-\infty}^{\lambda_i}\int_{-\infty}^{\lambda_j} \frac{\partial^2 f_t}{\partial x_j x_i}\mathrm{d}x_i \mathrm{d}x_j = \int_{\mathbb{R}^{n-2}} f_t(x_k, \lambda_i, \lambda_j)\mathrm{d}x_k.$$

where $x_k \in \mathbb{R}^{n-2}$ denotes the rest of the variables in x except x_i and x_j.

2.4.2 Slepian's Lemma: Theorem 2.4.7

Proof of (2.55): $P(\bigcap\{X_i \leq \lambda_i\}) - P(\bigcap\{Y_i \leq \lambda_i\}) \leq \frac{1}{2\pi}\sum(C^1_{ij} - C^0_{ij})^+ \frac{\exp(-\frac{\lambda_i^2+\lambda_j^2}{2(1+\rho_{ij})})}{(1-\rho_{ij}^2)^{1/2}}$.

Observe the last inequation: $I \leq \int_{\mathbb{R}^{n-2}} f_t(x_k, \lambda_i, \lambda_j)\mathrm{d}x_k$

- Note that $x_k \in \mathbb{R}^{n-2}$, and f_t is the density function of $X(t) \sim \mathcal{N}_n(0, C^t)$.
- Subvector $(X_i(t), X_j(t))^\top$ is still a Gaussian vector $\mathcal{N}_2(0, \begin{bmatrix} 1 & C^t_{ij} \\ C^t_{ij} & 1 \end{bmatrix})$, with density function $f_t(x_i, x_j) = \frac{1}{2\pi(1-C^t_{ij})^{1/2}}\exp\left(-\frac{x_i^2+x_j^2-2C^t_{ij}x_ix_j}{2(1-C^t_{ij})}\right)$ (simply by Gaussian's pdf).
- Joint pdf f_{ij} in $\mathbb{R}^2$ space can also be regarded as the integral of f_t in $\mathbb{R}^n$ space over $x_k \in \mathbb{R}^{n-2}$: $f_{ij}(x_i, x_j) = \int_{\mathbb{R}^{n-2}} f_t(x_k, x_i, x_j)\mathrm{d}x_k$, which is exactly the last integral in the above equation, with $x_i = \lambda_i, x_j = \lambda_j$.

Thus, we can further derive that:

$$\mathcal{I} \leq \frac{1}{2\pi(1-(C^t_{ij})^2)^{1/2}}\exp\left(-\frac{\lambda_i^2+\lambda_j^2-2C^t_{ij}\lambda_i\lambda_j}{2(1-(C^t_{ij})^2)}\right)$$

2.4.2 Slepian's Lemma: Theorem 2.4.7

Proof of (2.55): $P(\bigcap\{X_i \leq \lambda_i\}) - P(\bigcap\{Y_i \leq \lambda_i\}) \leq \frac{1}{2\pi}\sum(C_{ij}^1 - C_{ij}^0)^+ \frac{\exp(-\frac{\lambda_i^2+\lambda_j^2}{2(1+\rho_{ij})})}{(1-\rho_{ij}^2)^{1/2}}$.

Furthermore, $\mathcal{I} \leq \frac{1}{2\pi(1-(C_{ij}^t)^2)^{1/2}} \exp\left(-\frac{\lambda_i^2+\lambda_j^2-2C_{ij}^t\lambda_i\lambda_j}{2(1-(C_{ij}^t)^2)}\right) \leq \cdots :$

- $\cdots \leq \frac{1}{2\pi(1-(|C_{ij}^t|)^2)^{1/2}} \exp\left(-\frac{\lambda_i^2+\lambda_j^2-2|C_{ij}^t|\lambda_i\lambda_j}{2(1-(|C_{ij}^t|)^2)}\right)$ (as $|C_{ij}^t|$ is a more loose bound).
- $\leq \frac{1}{2\pi(1-\rho_{ij}^2)^{1/2}} \exp\left(-\frac{\lambda_i^2+\lambda_j^2-2\rho_{ij}\lambda_i\lambda_j}{2(1-\rho_{ij}^2)}\right)$ (by definition, $\rho_{ij} = \max\{|C_{ij}^1|, |C_{ij}^0|\}$).
- $\leq \frac{1}{2\pi(1-\rho_{ij}^2)^{1/2}} \exp\left(-\frac{\lambda_i^2+\lambda_j^2}{1+\rho_{ij}}\right)$ (as for function with form: $f(u) = \frac{a^2-2abu+b^2}{1-u}, u \in [0,\infty)$, the minimum is attained at $u=0$)

Hence, given $F'(t) = \sum(C_{ij}^1 - C_{ij}^0) \cdot \mathcal{I}$, we can derive that:

$$F'(t) \leq \sum_{1\leq i<j\leq n} (C_{ij}^1 - C_{ij}^0)^+ \cdot \frac{1}{2\pi(1-\rho_{ij}^2)^{1/2}} \exp\left(-\frac{\lambda_i^2+\lambda_j^2}{1+\rho_{ij}}\right)$$

□

2.4.2 Slepian's Lemma: Theorem 2.4.7

Proof of (2.56):

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

Define $\tilde{F}(t) = \int_{\mu_1}^{\lambda_1}\cdots\int_{\mu_n}^{\lambda_n} f_t(x)\mathrm{d}x$, then f_t is the same density function of $X(t)$ as before. The only difference from $F(t)$ is the integration interval, which is now $\mu_i \le x_i \le \lambda_i$.

Similarly, we can derive that:

$$\tilde{F}'(t) = \sum_{1\le i<j\le n}(C_{ij}^1 - C_{ij}^0)\cdot\int_{\mu_1}^{\lambda_1}\cdots\int_{\mu_n}^{\lambda_n}\frac{\partial^2 f_t}{\partial x_j x_i}\mathrm{d}x$$

Then by the similar procedure as before, we can derive that:

$$= |\tilde{F}'(t)| \le \frac{4}{2\pi}\sum_{1\le i<j\le n}|C_{ij}^1 - C_{ij}^0|\cdot\frac{1}{(1-\rho_{ij}^2)^{1/2}}\exp\left(-\frac{\nu^2}{1+\rho_{ij}}\right)$$

which yields (2.56) by integrating over $t \in [0,1]$.

□

2.4.2 Slepian's Lemma: Theorem 2.4.8

Theorem 2.4.8: (Slepian's Lemma)
Let $X = (X_1, \cdots, X_n)$ and $Y = (Y_1, \cdots, Y_n)$ be centered jointly Gaussian vectors in $\mathbb{R}^n$ s.t.

$$\mathbb{E}(X_iX_j) \leq \mathbb{E}(Y_iY_j),\ \mathbb{E}(X_i^2) = \mathbb{E}(Y_i^2), \quad \forall 1 \leq i, j \leq n. \quad (2.58)$$

- *i.e. X and Y have the same var, and the cov of X is less than that of Y.*

Then, for all $\lambda_i \in \mathbb{R}, i \leq n$,

$$\Pr\left(\bigcup_{i=1}^{n}\{Y_i > \lambda_i\}\right) \leq \Pr\left(\bigcup_{i=1}^{n}\{X_i > \lambda_i\}\right) \quad (2.59)$$

- *i.e. At least exists one i, s.t. $Y_i > \lambda_i$ has a lower probability than $X_i > \lambda_i$.*

and therefore,

$$\mathbb{E}\left[\max_{1\leq i\leq n} Y_i\right] \leq \mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] \quad (2.60)$$

- *i.e. The expectation of the maximum of Y is lower than that of X.*

2.4.2 Slepian's Lemma: Theorem 2.4.8

In general, if the variables turn to be more correlated, the probability of extreme events will be lower.

Proof of (2.59)

- It can be shown that it satisfies the conditions of Theorem 2.4.7, i.e. $\forall\lambda$, $\Pr(\bigcap_{i=1}^n\{X_i \le \lambda_i\}) - \Pr(\bigcap_{i=1}^n\{Y_i \le \lambda_i\}) \le 0$, which is equivalent to (2.59).

Proof of (2.60)

- The expectation can be expressed as: $\mathbb{E}[\max_{1\le i\le n} Y_i] = \int_0^\infty \Pr(\max_{1\le i\le n} Y_i > t)\mathrm{d}t = \int_0^\infty \Pr(\bigcup_{i=1}^n\{Y_i > t\})\mathrm{d}t$. Thus by (2.59), we can finish the proof.

$\square$

2.4.2 Slepian's Lemma: Remark 2.4.9

Remark 2.4.9

For symmetric random vectors X_i (i.e. X_i has the same distribution as $-X_i$) and for any $i_0 \in \{1, \cdots, n\}$, the inequation (2.59) can be strengthened to:

$$\mathbb{E}[\max_{i\leq n} X_i] \leq \mathbb{E}[\max_{i\leq n} |X_i|] \leq \mathbb{E}|X_{i0}| + \mathbb{E}[\max_{i\leq n} |X_i - X_j|] \overset{(*)}{\leq} \mathbb{E}|X_{i_0}| + 2\mathbb{E}[\max_{i\leq n} X_i] \quad (2.61)$$

Here, $(*)$ holds as:

- By $|X_i - X_j| \leq |X_i| + |X_j|$, we have $\max_{i,j} |X_i - X_j| \leq \max_i |X_i| + \max_i |-X_i| = 2\max_i |X_i|$. Then the inequality of expectation follows.
    - The idea is simple: the max absolute difference should not exceed the twice of the max absolute value.

2.4.2 Slepian's Lemma: Corollary 2.4.10

Corollary 2.4.10

Let $X = (X_1, \cdots, X_n)$ and $Y = (Y_1, \cdots, Y_n)$ be centered jointly Gaussian vectors in $\mathbb{R}^n$, and assume that:

$$\mathbb{E}(Y_i - Y_j)^2 \leq \mathbb{E}(X_i - X_j)^2, \quad \forall i, j \in \{1, \cdots, n\}$$

Then

$$\mathbb{E}[\max_{i \leq n} Y_i] \leq 2\, \mathbb{E}[\max_{i \leq n} X_i]$$

- This corollary is sometimes easier to apply as it does not require $\mathbb{E}(X_i^2) = \mathbb{E}(Y_i^2)$.

Proof

- **W.L.O.G., first simplify the problem as follows**:
  Redefine $X_i := X_i - X_1$ and $Y_i := Y_i - Y_1$ and assume $X_1 = Y_1 = 0$. Then condition $\mathbb{E}(Y_i - Y_j)^2 \leq \mathbb{E}(X_i - X_j)^2$ can be reduced to $\mathbb{E}Y_i^2 \leq \mathbb{E}X_i^2$.

2.4.2 Slepian's Lemma: Corollary 2.4.10

Proof (cont.)

- **For convenience, define new variables $\tilde{X}_i, \tilde{Y}_i$ as follows:**
    - $\tilde{X}_i = X_i + \sqrt{\sigma_X^2 + \mathbb{E}Y_i^2 - \mathbb{E}X_i^2} \cdot g, \quad \tilde{Y}_i = Y_i + \sigma_X g, \quad i = 1, \cdots, n$
        - $\sigma_X^2 = \max_{i \le n} \mathbb{E}X_i^2$.
        - g is a standard Gaussian random variable independent of X_i, Y_i. It can be regarded as a noise term. It keeps the property of Gaussianity but also makes the analysis more flexible.
    - Here check the property of $\tilde{X}_i, \tilde{Y}_i$:
        - $\mathbb{E}\tilde{X}_i^2 = \mathbb{E}X_i^2 + \sigma_X^2 + \mathbb{E}Y_i^2 - \mathbb{E}X_i^2 = \sigma_X^2 + \mathbb{E}Y_i^2$.
        - $\mathbb{E}\tilde{Y}_i^2 = \mathbb{E}Y_i^2 + \sigma_X^2$.
        - $\mathbb{E}(\tilde{Y}_i - \tilde{Y}_j)^2 = \mathbb{E}(Y_i - Y_j)^2 \le \mathbb{E}(X_i - X_j)^2 = \mathbb{E}(\tilde{X}_i - \tilde{X}_j)^2$.

2.4.2 Slepian's Lemma: Corollary 2.4.10

Proof (cont.)

- **Apply Slepian's Lemma:**
  - We have just checked that $\mathbb{E}(\tilde{Y}_i - \tilde{Y}_j)^2 \leq \mathbb{E}(\tilde{X}_i - \tilde{X}_j)^2$, which satisfies the condition of Slepian's Lemma.
  - Then by Slepian's Lemma, we have $\mathbb{E}[\max_{i\leq n} \tilde{Y}_i] \leq \mathbb{E}[\max_{i\leq n} \tilde{X}_i]$.
    - $\mathbb{E}[\max_{i\leq n} \tilde{Y}_i] = \mathbb{E}[\max_{i\leq n} Y_i + \sigma_X g] = \mathbb{E}[\max_{i\leq n} Y_i] + \sigma_X \mathbb{E}g = \mathbb{E}[\max_{i\leq n} Y_i]$.
    - $\mathbb{E}[\max_{i\leq n} \tilde{X}_i] = \mathbb{E}[\max_{i\leq n} X_i + \max \sqrt{\cdot} g] \leq \mathbb{E}[\max_{i\leq n} X_i] + \mathbb{E}[\max \sqrt{\cdot} g]$
      - For the second term, as $\mathbb{E}Y_i^2 - \mathbb{E}X_i^2 \leq 0$, we have $\sqrt{\sigma_X^2 + \mathbb{E}Y_i^2 - \mathbb{E}X_i^2} \leq \sigma_X$. Thus $\mathbb{E}[\max \sqrt{\cdot} g] \leq \sigma_X \mathbb{E}[\max g] = \sigma_X \mathbb{E}g^+$.
      - Combine the results: $\mathbb{E}[\max_{i\leq n} \tilde{X}_i] \leq \mathbb{E}[\max_{i\leq n} X_i] + \sigma_X \mathbb{E}g^+$
  - So far, $\mathbb{E}[\max_{i\leq n} Y_i] \leq \mathbb{E}[\max_{i\leq n} X_i] + \sigma_X \mathbb{E}g^+ = \mathbb{E}[\max_{i\leq n} X_i] + \sigma_X \frac{1}{\sqrt{2\pi}}$ $(\star)$
    - As $\mathbb{E}g^+ = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t = \frac{1}{\sqrt{2\pi}}$.

2.4.2 Slepian's Lemma: Corollary 2.4.10

Proof (cont.)

- **Apply Remark 2.4.9**:
    - Moreover, $\sigma_X \triangleq \max \sqrt{\mathbb{E}X_i^2} \overset{(*)}{=} \sqrt{\frac{\pi}{2}} \max \mathbb{E}|X_i| \overset{(\dagger)}{\leq} 2\sqrt{\frac{\pi}{2}}\mathbb{E}[\max_{i\leq n} X_i]$
        - $(*)$ can be directly derived from normal distribution's moments.
        - $(\dagger)$ is due to Remark 2.4.9 and let $i_0 = 1$.
    - Bring back to $(\star)$, we have
      $\mathbb{E}[\max_{i\leq n} Y_i] \leq \mathbb{E}[\max_{i\leq n} X_i] + 2\sqrt{\frac{\pi}{2}}\mathbb{E}[\max_{i\leq n} X_i] = 2\mathbb{E}[\max_{i\leq n} X_i]$.

$\square$

> Note: In fact, constant 2 inequality is suboptimal: it can be improved to 1.

2.4.2 Sudakov's Lower Bound: Lemma 2.4.11

In the last part of this section, we will focus on Gaussian processes and metric entropy.

- Assume X is a Gaussian process defined on T, and we can define a metric (or distance) by: $d_X(s,t) = \mathbb{E}(X(t) - X(s))^2$ (it can be regarded as a MSE between $X(t)$ and $X(s)$).
- Then we can define metric entropy of the space (T, d_X) as: $\mathtt{N}(\epsilon, T, d_X)$, which is the minimal number of balls of radius ϵ needed to cover the space T.

2.4.2 Sudakov's Lower Bound: Lemma 2.4.11

Lemma 2.4.11

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

1. $\lim_{n\to\infty} \frac{\mathbb{E}[\max_{i\le n}|g_i|]}{\sqrt{2\log n}} = 1.$
2. There exists $K < \infty$ s.t. for all $n > 1$,

$$K^{-1}\sqrt{2\log n} \le \mathbb{E}[\max_{i\le n} g_i] \le \mathbb{E}[\max_{i\le n}|g_i|] \le K\sqrt{2\log n}.$$

Intuitively,

- The first part shows that the expectation of the maximum of standard Gaussian random variables grows as $\mathcal{O}(\sqrt{2\log n})$.
- The second part gives a more precise bound for the expectation of the maximum of standard Gaussian random variables.

2.4.2 Sudakov's Lower Bound: Lemma 2.4.11

Proof of 2.4.11- a: $\lim_{n\to\infty} \frac{\mathbb{E}[\max_{i\le n}|g_i|]}{\sqrt{2\log n}} = 1$

$$\mathbb{E}[\max_{i\le n}|g_i|] \overset{(1)}{=} \int_0^\infty \Pr(\max_{i\le n}|g_i| > t)\mathrm{d}t \overset{(2)}{\le} \delta + n\int_\delta^\infty \Pr(|g| > t)\mathrm{d}t \overset{(3)}{=} \delta + n\sqrt{\frac{2}{\pi}}\int_\delta^\infty \exp\left(-\frac{u^2}{2}\right)\int_\delta^u \mathrm{d}t\mathrm{d}u$$
$$\overset{(4)}{\le} \delta + n\sqrt{\frac{2}{\pi}}\exp\left(-\frac{\delta^2}{2}\right) - n\sqrt{\frac{2}{\pi}}\frac{\delta^2}{\delta^2+1}\exp\left(-\frac{\delta^2}{2}\right) \overset{(5)}{=} \delta + n\sqrt{\frac{2}{\pi}}\frac{1}{\delta^2+1}\exp\left(-\frac{\delta^2}{2}\right)$$

- (1): By properties of expectation.
- (2): $\int_0^\infty \mathrm{P}(\max\cdot)\mathrm{d}t = \int_0^\delta \mathrm{P}(\max\cdot)\mathrm{d}t + \int_\delta^\infty \mathrm{P}(\max\cdot)\mathrm{d}t \le \int_0^\delta 1\mathrm{d}t + n\int_\delta^\infty \Pr(|g_i| > t)\mathrm{d}t$. For some $\delta > 0$.
- (3): As $g_i \sim \mathcal{N}(0,1)$, $\Pr(|g| > t) = \sqrt{\frac{2}{\pi}}\int_t^\infty \exp\left(-\frac{u^2}{2}\right)\mathrm{d}u$. Plus, $\int_\delta^\infty\int_t^\infty f(u)\mathrm{d}t\mathrm{d}u = \int_\delta^\infty f(u)\int_\delta^u \mathrm{d}t\mathrm{d}u$.
- (4): Continue from (3), as $\int_\delta^u \mathrm{d}t = u - \delta$, $(3) = \delta + n\sqrt{\frac{2}{\pi}}\int_\delta^\infty \exp\left(-\frac{u^2}{2}\right)\mathrm{d}u - n\sqrt{\frac{2}{\pi}}\int_\delta^\infty u\exp\left(-\frac{u^2}{2}\right)\mathrm{d}u$, and the last term can be approximated by integration by parts.
- (5): By simplification.

2.4.2 Sudakov's Lower Bound: Lemma 2.4.11

Proof of 2.4.11- a: $\lim_{n\to\infty} \frac{\mathbb{E}[\max_{i\le n}|g_i|]}{\sqrt{2\log n}} = 1$

So far, $\mathbb{E}[\max_{i\le n}|g_i|] \le \delta + n\sqrt{\frac{2}{\pi}}\frac{1}{\delta^2+1}\exp\left(-\frac{\delta^2}{2}\right)$.

- Here, set $\delta = \sqrt{2\log n}$, then this upper bound can be simplified as: $\mathbb{E}[\max_{i\le n}|g_i|] \le \sqrt{2\log n} + n\sqrt{\frac{2}{\pi}}\frac{\exp(-\log n)}{2\log n+1} = \sqrt{2\log n} + \sqrt{\frac{2}{\pi}}\frac{1}{(2\log n+1)}$.
- Thus, $\lim_{n\to\infty}\sup \frac{\mathbb{E}[\max_{i\le n}|g_i|]}{\sqrt{2\log n}} \le \frac{\sqrt{2\log n}+\sqrt{\frac{2}{\pi}}\frac{1}{(2\log n+1)}}{\sqrt{2\log n}} = 1.$ (♠)

2.4.2 Sudakov's Lower Bound: Lemma 2.4.11

Proof of 2.4.11- a: $\lim_{n\to\infty} \frac{\mathbb{E}[\max_{i\le n}|g_i|]}{\sqrt{2\log n}} = 1$

On the other hand,

$$\Pr(|g|>t) = 2\Pr(g>t) = 2\int_t^\infty \frac{1}{\sqrt{2\pi}}\exp(-u^2/2) \overset{(\star)}{\geq} \sqrt{\frac{2}{\pi}}\exp(-t^2/2)\frac{t}{t^2+1}$$

- $(\star)$: This can be checked by intergration by parts.

Now for $t \leq \sqrt{(2-\delta)\log n}$, (for $0<\delta<2$), we have:

$$\Pr(|g|>t) \geq \sqrt{\frac{2}{\pi}}\frac{\sqrt{(2-\delta)\log n}}{(2-\delta)\log n+1}n^{-(2-\delta)/2} := \frac{c(n,\delta)}{n}$$

Then consider the tail probability of $\max_{i\le n}|g_i|$, we have:

$$\Pr(\max_{i\le n}|g_i|>t) \geq 1-(1-\Pr(|g|>t))^n \geq 1-(1-c(n,\delta)/n)^n \geq 1-\exp(-c(n,\delta))$$

2.4.2 Sudakov's Lower Bound: Lemma 2.4.11

Proof of 2.4.11- a: $\lim_{n\to\infty} \frac{\mathbb{E}[\max_{i\le n}|g_i|]}{\sqrt{2\log n}} = 1$, b^*

Then consider the tail expectation of $\max_{i\le n}|g_i|$, we have:

$$\begin{aligned}\mathbb{E}[\max_{i\le n}|g_i|] &\overset{(1)}{=} \int_0^{\sqrt{(2-\delta)\log n}} \Pr(\max_{i\le n}|g_i|>t)\mathrm{d}t \overset{(2)}{\ge} \int_0^{\sqrt{(2-\delta)\log n}} (1-\exp(-c(n,\delta)))\mathrm{d}t \\ &= \sqrt{(2-\delta)\log n}\,(1-\exp(-c(n,\delta)))\end{aligned}$$

which yields that

$$\liminf_{n\to\infty} \frac{\mathbb{E}[\max_{i\le n}|g_i|]}{\sqrt{(2-\delta)\log n}} \ge \liminf_{n\to\infty} \frac{\sqrt{(2-\delta)\log n}\,(1-\exp(-c(n,\delta)))}{\sqrt{(2-\delta)\log n}} = 1, \;\; \forall 0<\delta<2. \quad (\clubsuit)$$

Letting $\delta\to 0$, together with ($\spadesuit$) and ($\clubsuit$), we can derive that $\lim_{n\to\infty} \frac{\mathbb{E}[\max_{i\le n}|g_i|]}{\sqrt{2\log n}} = 1$, which finishes the proof.

□

(b) can be then derived from (a) directly as a consequence using Remark 2.4.9. □

2.4.2 Sudakov's Lower Bound: Lemma 2.4.12

Before **Sudakov's Lower Bound**, first recall some concepts of **metric entropy**.

- Given a metric or pseudo-metric space (T, d), $\mathtt{N}(\epsilon, T, d)$ denotes the ϵ-covering number of T, and that the packing numbers, denoted as $\mathtt{D}(T, d, \epsilon)$, are comparable to the covering numbers. Concretely, $\mathtt{N}(\epsilon, T, d) \leq \mathtt{D}(T, d, \epsilon)$.
  - **Metric Space** (T, d): Given a set T and a metric $d : T \times T \to \mathbb{R}_{\geq 0}$, which satisfies: 1. $d(x, y) \geq 0, d(x, y) = 0 \Leftrightarrow x = y$; 2. $d(x, y) = d(y, x)$; 3. $d(x, y) \leq d(x, z) + d(z, y)$.
  - **Pseudo-metric Space** (T, d): It is a loosened version of metric space, which allows $d(x, y) = 0$ even if $x \neq y$.
  - ϵ-**Covering Number**: It is the minimal number of balls of radius ϵ needed to cover the space T. It indicates the complexity of the space.
  - ϵ-**Packing Number**: It is the maximal number of disjoint balls of radius ϵ that can be packed into the space T. Ciove

2.4.2 Sudakov's Lower Bound: Theorem 2.4.12

Theorem 2.4.12 (Sudakov's Lower Bound)

There exists a constant $K < \infty$ s.t. if $X(t), t \in T$, is a centered Gaussian process and $d_X(s,t) = \sqrt{\mathbb{E}(X(t) - X(s))^2}$ denotes the associated pseudo-metric on T, then for all $\epsilon > 0$:

$$\epsilon\sqrt{\log \mathtt{N}(\epsilon, T, d_X)} \leq K \sup_{S_{\text{finite}} \subset T} \mathbb{E}\left[\max_{t \in S_{\text{finite}}} X(t)\right]$$

Intuitively,

- The theorem gives a lower bound on the metric entropy of the space (T, d_X) in terms of the expectation of the maximum of the Gaussian process $X(t)$.
- LHS indicates the complexity of the space by the covering number.
- For RHS, as T may be complex, we only need to consider the finite subset of T to calculate the expectation of the maximum of $X(t)$. For different subset S_{finite}, the expectation also varies; thus we need to take the supremum for the 'worst' case.

2.4.2 Sudakov's Lower Bound: Theorem 2.4.12

Proof

- Let N be any finite number not exceeding $\mathtt{N}(\epsilon, T, d_X)$ (which may or may not be finite). Since $\mathtt{D}(T, d_X, \epsilon) \geq \mathtt{N}(\epsilon, T, d_X)$, $\mathtt{D} \geq \mathtt{N} \geq N$.
- Thus, we can always find N points in T, denoted as $S = \{t_1, \cdots, t_N\}$, s.t. $d_X(t_i, t_j) \geq \epsilon, \forall 1 \leq i \neq j \leq N$.
    - Intuitively, these points are 'far' from each other so that they cannot be covered by a ball of radius ϵ.
- Introduce $g_i, i \leq N$ be i.i.d standard Gaussian random variables, and set $X^*(t_i) = \epsilon g_i/2, \forall i \leq N$.
    - Here, $\epsilon/2$ is a factor to ensure the pseudo-metric to be consistent.
        - $\mathbb{E}[X^*(t_i) - X^*(t_j)]^2 = \mathbb{E}[\epsilon(g_i - g_j)/2]^2 = \epsilon^2/2 \leq \epsilon^2 \leq d_X(t_i, t_j)^2$ †.

2.4.2 Sudakov's Lower Bound: Theorem 2.4.12

Proof (cont.)

- Now that we have constructed two Gaussian vectors: $\mathbf{X} = [X(t_1), \cdots, X(t_N)]$ and $\mathbf{X}^* = [X^*(t_1), \cdots, X^*(t_N)]$.
  - By Corollary 2.4.10, since $\mathbb{E}[X^*(t_i)X^*(t_j)] \leq \mathbb{E}[X(t_i)X(t_j)]$ (as is shown in †), we have $\mathbb{E}[\max_{i\leq N} X^*(t_i)] \leq 2\mathbb{E}[\max_{i\leq N} X(t_i)]$.
  - Recall that $X^*(t_i) = \epsilon g_i/2$, then $\mathbb{E}[\max_{i\leq N} X^*(t_i)] = \frac{\epsilon}{2}\mathbb{E}[\max_{i\leq N} g_i]$.
- Further consider Lemma 2.4.11, for such g_i's, we have: $K^{-1}\sqrt{2\log N} \leq \mathbb{E}[\max_{i\leq N} g_i] \leq K\sqrt{2\log N}$, i.e. $\mathbb{E}[\max_{i\leq N} g_i] \sim \sqrt{2\log N}$.
- Thus, given $\mathbb{E}[\max_{i\leq N} X^*(t_i)] \leq 2\mathbb{E}[\max_{i\leq N} X(t_i)]$, we have $\frac{\epsilon}{2}\sqrt{2\log N} \leq 2\mathbb{E}[\max_{i\leq N} X(t_i)]$, i.e. $\epsilon\sqrt{\log N} \leq K\,\mathbb{E}[\max_{i\leq N} X(t_i)]$.
- Finally, as N is arbitrary, we can take the supremum over all finite subsets S_{finite} of T to derive the theorem: $\epsilon\sqrt{\log \mathtt{N}(\epsilon, T, d_X)} \leq K \sup_{S_{\text{finite}}\subset T} \mathbb{E}\left[\max_{t\in S_{\text{finite}}} X(t)\right]$.

$\square$

2.4.2 Sudakov's Lower Bound: Corollary 2.4.13 (Sudakov's Theorem)

Corollary 2.4.13 (Sudakov's Theorem)

Let $X(t), t \in T$ be a centred Gaussian process, let d_X be the associated pseudo-distance. If $\liminf_{\epsilon \downarrow 0} \epsilon \sqrt{\log \mathtt{N}(\epsilon, T, d_X)} = \infty$, then $\sup_{t \in T} |X(t)| = \infty$ almost surely, i.e. X is not sample bounded.

Intuitively,

- As $\mathtt{N}$ measures the complexity of the space, if the covering number (complexity) grows too fast as ϵ decreases to 0, then the maximum of $X(t)$ will be almost surely impossible to control in a finite range, i.e. $\sup_{t \in T} |X(t)| = \infty$ almost surely.
- Specifically, $\liminf_{\epsilon \downarrow 0} \epsilon \sqrt{\log \mathtt{N}(\epsilon, T, d_X)}$ indicates that we are considering a sufficiently small $\epsilon > 0$; $\epsilon \cdot \sqrt{\log \mathtt{N}(\epsilon)}$ combines the decreasing rate of ϵ and the increasing rate of $\mathtt{N}(\epsilon)$; $\liminf$ ensures that though the convergence may not be strict, such lower bound of trending to infinity is sufficient to guarantee the unboundedness of $X(t)$.

2.4.2 Sudakov's Lower Bound: Corollary 2.4.13 (Sudakov's Theorem)

Proof

- According to Theorem 2.4.12 (Sudakov's LB), we have $\epsilon\sqrt{\log \mathtt{N}(\epsilon, T, d_X)} \le K \sup_{S_{\text{finite}} \subset T} \mathbb{E}\left[\max_{t\in S} X(t)\right]$. By assumption $\liminf_{\epsilon\downarrow 0} \epsilon\sqrt{\log \mathtt{N}(\epsilon, T, d_X)} = \infty$, it indicates that $\mathbb{E}\left[\max_{t\in S} X(t)\right]$ must be unbounded.
- Thus, we can construct a sequence of finite subsets $S_n \subset T$ s.t. $\mathbb{E}\sup_{t\in S_n}|X(t)| \nearrow \infty$ ($\nearrow$ denotes non-decreasing convergence).
  - Here, S_n is a sequence of increasing finite subsets of T, formally, $S_1 \subset S_2 \subset \cdots, \bigcup_{n\in\mathbb{N}} S_n = T$.
- By **Monotone Convergence Theorem**, it guarantees $\mathbb{E}\sup_{t\in\cup S_n}|X(t)| = \lim_{n\to\infty}\mathbb{E}\sup_{t\in S_n}|X(t)| = \infty$. And as $\mathbb{E}[\sup_{t\in S_n}|X(t)|] \to \infty$, $\mathbb{E}[\sup_{t\in\cup S_n}|X(t)|] = \infty$ almost surely.
- As $\bigcup_{n=1}^{\infty} S_n$ is countable, and Gaussian process X is separable on countable set, we apply Theorem 2.1.20(b) $\Pr\{\sup_{t\in\cup S_n}|X(t)| < \infty\} = 0$, thus $\sup_{t\in T}|X(t)| = \infty$ almost surely.

2.4.2 Sudakov's Lower Bound: Corollary 2.4.14

By **Sudakov's Theorem**, if a centred Gaussian process is sample bounded (i.e. $\sup_{t \in T} |X(t)| < \infty$ almost surely), then the covering numbers $\mathtt{N}(\epsilon, T, d_X) < \infty$ for all $\epsilon > 0$, i.e. the covering number is finite, and the metric space (T, d_X) is not only separable but also totally bounded.

Furthermore, if X is sample continuous, then a stronger result holds as **Corollary 2.4.14**:

- **Sample Continuity**: $\Pr(\forall t_0 \in T, \lim_{t \to t_0} X(t) = X(t_0)) = 1.$

Corollary 2.4.14

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

$$\lim_{\epsilon \to 0} \epsilon \sqrt{\log \mathtt{N}(\epsilon, T, d_X)} = 0$$

2.4.2 Sudakov's Lower Bound: Corollary 2.4.14

Proof

- Consider local increments $|X(t) - X(s)|$:
  - As X is sample continuous, the sample paths of X is uniformly continuous and bounded, and thus X is sample bounded (by **Theorem 2.1.10**), i.e. $\mathbb{E}[\sup_{t\in T}|X(t)|] < \infty$.
  - Furthermore, for arbitrary $\delta > 0$, since $\sup_{d_X(s,t)<\delta}|X(t) - X(s)| \le 2\sup_{t\in T}|X(t)|$, we have $\mathbb{E}[\sup_{d_X(s,t)<\delta}|X(t) - X(s)|] < \infty$.
  - Define $\eta(\delta) := \mathbb{E}[\sup_{d_X(s,t)<\delta}|X(t) - X(s)|]$, then by **Dominate Converge Theorem**, $\eta(\delta) \to 0$ as $\delta \to 0$.
    - *It means that: if $d_X(s,t)$ is sufficiently small, the increment $|X(t) - X(s)|$ is also expected to be trivial.*

2.4.2 Sudakov's Lower Bound: Corollary 2.4.14

Proof (cont.)

- As X is sample continuous, it also implies that (T, d_X) is totally bounded, i.e. for any $\delta > 0, \exists A_{\text{finite}} \subset T$, s.t. A is δ-dense in T.
  - A is δ-dense in T means: $\forall t \in T, \exists s \in A_{\text{finite}}$, s.t. $d_X(t, s) < \delta$.
    - *It means that the points in A is 'dense' enough, such that for any points in T, we can always find a point in A that is close in enough (no further than δ).*
    - It means that we can partition space T into balls of radius δ centered at points in $s \in A_{\text{finite}}$. And here, each ball represents a subset $T_s \subset T, (T_s = \{t \in T : d_X(s, t) < \delta\})$, with the radius no larger than δ.
  - For each T_s, consider the process $Y_t = X_t - X_s, t \in T_s$.
    - As T_s is smaller than δ, then by **Sudakov's Theorem**, T_s has an ϵ-dense subset $B_s \subset T_s$, whose cardinality satisfies: $\epsilon\sqrt{\log \operatorname{Card}(B_s)} \leq K\eta(\delta) \diamond$.

2.4.2 Sudakov's Lower Bound: Corollary 2.4.14

Proof (cont.)

Then we can derive that:

$$\epsilon\sqrt{\log \mathtt{N}(\epsilon, T, d_X)} \overset{(1)}{\leq} \epsilon\sqrt{\log \mathrm{Card}(\bigcup_{s\in A} B_s)} \overset{(2)}{\leq} \epsilon\sqrt{\log[\mathrm{Card}(\mathrm{A}) \times \max_{s\in A} \mathrm{Card}(B_s)]}$$
$$\overset{(3)}{\leq} \epsilon\sqrt{\log \mathrm{Card}(\mathrm{A}) + \frac{K^2\eta^2(\delta)}{\epsilon^2}} \overset{(4)}{\leq} \epsilon\sqrt{\log \mathrm{Card}(\mathrm{A})} + K\eta(\delta)$$

- (1): As $B = \cup B_s$, each point in T_s can be covered by a ball of radius ϵ in B_s, thus by definition, the cardinality of B is the upper bound of the covering number.
- (2): By property of cardinality.
- (3): By $\diamond : \log \mathrm{Card}(B_s) \leq \frac{K^2\eta^2(\delta)}{\epsilon^2}$.
- (4): By square root inequality.

2.4.2 Sudakov's Lower Bound: Corollary 2.4.14

Proof (cont.)

So far:

$$\epsilon\sqrt{\log \mathtt{N}(\epsilon, T, d_X)} \leq \epsilon\sqrt{\log \mathrm{Card}(\mathrm{A})} + K\eta(\delta)$$

Thus, for all $\delta > 0$,

$$\limsup_{\epsilon\to 0} \epsilon\sqrt{\log \mathtt{N}(\epsilon, T, d_X)} \leq K\eta(\delta)$$

which then proves the corollary by letting $\lim_{\delta\to 0} \eta(\delta) = 0$.

$\square$

2.4.2 Sudakov's Lower Bound: Summary

Finally, combining **Theorem 2.4.12** and **Theorem 2.3.6**, here gives a two-sided bound for $\mathbb{E}[\max_{i\le n} X_i]$:

Assume $X(t), t \in T$ is a centred Gaussian process, $d_X(s,t)$ is the associated pseudo-metric on T, and $\mathtt{N}(\epsilon, T, d_X)$ is the covering number of the space (T, d_X), $\sigma_X^2 = \max \mathbb{E} X_i^2$, $D = \sup_{s,t\in T} d_X(s,t)$ as the diameter of the space. Then the expectation of the maximum of the Gaussian process $X(t)$ satisfies:

$$\frac{1}{K}\sigma_X\sqrt{\log \mathtt{N}(T, d_X, \sigma_X)} \le \mathbb{E}\sup_{t\in T}|X(t)| \le K\sigma_X\sqrt{\log \mathtt{N}(T, d_X, \sigma_X)} \quad (2.61)$$

where $K > 0$ is a constant independent of T, d_X.

■

Thanks