Mathematical background

Cunyi Nan

Summaries

- Measurable space, measurable function
- Probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω sample space, $\mathcal{F}$ event space and $\mathbb{P}$ probability measure
- Random variable $X : \Omega \to \Lambda$
- Induced probability measure

  $\mu_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\})$, here $\mu_X : E \to [0,1]$
- If $\Lambda = \mathbb{R}$, X is univariate r.v.

  If $\Lambda = \mathbb{R}^N$, X is multivariate r.v. or random vector.

  If $\Lambda = \mathbb{R}^N$ and $\exists\ \Pi_X : \mathbb{R}^N \to \mathbb{R}_+$ with $\mu_X(A) = \int_A \Pi_X(A)\mathrm{d}x$, X is absolutely continuous and Π_X is probability density function.

  If $X(\Omega)$ finite or countable, then $p_X(x) := \mathbb{P}(X^{-1}(\{x\}))$ and X is discrete valued.
- If only Π_X and p_X are given: probability distribution.
- If $X : \Omega \to \Lambda$ and $f : \Lambda \to \Xi$, then

  $$\mathbb{E}(f(x)) := \int_\Lambda f(x)\mu_X(\mathrm{d}x)$$

  if it converges absolutely, $\mathbb{E}$ is expectation.

1 Elementary probability theory

Exercise 1.1 (pp.3 Exercise 1.1.2.)**.** If $\mathcal{F}$ is a collection of subsets which is closed under complement and countable unions, then it is closed under countable intersections.

Proof. For $A_1, \dots, A_n, \dots \in \mathcal{F}$, their countable intersections is $\bigcap_{n=1}^{\infty} A_n$.

$$\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} (A_n^c)^c = \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c$$

Since $\mathcal{F}$ is closed under complement and countable unions, by $A_n \in \mathcal{F}$ we have $A_n^c \in \mathcal{F}$ and $\bigcup_{n=1}^{\infty} A_n^c \in \mathcal{F}$. Its complement, which is the equation above, also belongs to $\mathcal{F}$. Hence, it is closed under countable intersections. □

Exercise 1.2 (pp.3 Exercise 1.1.3)**.** If Ω contains N elements, how many elements does its power set contain?

Proof. Obviously the empty set $\varnothing$ is in the power set. For non-empty sets, firstly, if we choose only one element from Ω, there will be N choices i.e. $\binom{N}{1}$. Secondly, if we choose any two elements from Ω, use combinatorics knowledge there will be $\binom{N}{2}$ choices. Continue this until we choose N elements, there's only one choice i.e. $\binom{N}{N}$.

Hence, the power set of Ω contain

$$1 + \binom{N}{1} + \binom{N}{2} + \cdots + \binom{N}{N} = (1+1)^N = 2^N$$

elements. □

Exercise 1.3 (pp.4 Exercise 1.1.6.)**.** If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and if $A, B \in \mathcal{F}$, check that

(1) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(2) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

(3) If $A \subset B$, then $\mathbb{P}(A) \le \mathbb{P}(B)$.

Proof. (1) Since A^c and A are disjoint subsets, by finite additivity $1 = \mathbb{P}(\Omega) = \mathbb{P}(A^c \cup A) = \mathbb{P}(A^c) + \mathbb{P}(A)$, hence $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(2) Let subset $C = A$ and $D = B \setminus (A \cap B)$, then $C \cap D = \varnothing$ and $C \cup D = A \cup B$. By finite additivity, $\mathbb{P}(A \cup B) = \mathbb{P}(C \cup D) = \mathbb{P}(C) + \mathbb{P}(D) = \mathbb{P}(A) + \mathbb{P}(B \setminus (A \cap B))$.

Here $\mathbb{P}(B \setminus (A \cap B)) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$, hence $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

(3) Obviously $B = A \cup (B \setminus A)$ and $A \cap (B \setminus A) = \varnothing$. So $\mathbb{P}(B) - \mathbb{P}(A) = \mathbb{P}(B \setminus A)$ and since $\mathbb{P}$ is probability measure, $\mathbb{P}(B) \ge \mathbb{P}(A)$. □

Lemma 1.4 (pp.4 Lemma 1.1.7.)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and consider* $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{F}$. *If* $A_j \subset A_{j+1}$ *for any* j, *then*

$$\mathbb{P}\left(\bigcup_{j\in\mathbb{N}} A_j\right) = \lim_{j\to\infty} \mathbb{P}(A_j)$$

while if $A_j \supset A_{j+1}$ *for any* j, *then*

$$\mathbb{P}\left(\bigcap_{j\in\mathbb{N}} A_j\right) = \lim_{j\to\infty} \mathbb{P}(A_j)$$

Proof. For $A_j \subset A_{j+1}$ case, let $B_1 = A_1, B_2 = A_2 \setminus A_1, \ldots, B_j = A_j \setminus A_{j-1}$. Then $B_i \cap B_j = \varnothing$ for $i \neq j$ and $\bigcup_{j\in\mathbb{N}} B_j = \bigcup_{j\in\mathbb{N}} A_j$.

By countable additivity,

$$\begin{aligned}\mathbb{P}\left(\bigcup_{j\in\mathbb{N}} A_j\right) = \mathbb{P}\left(\bigcup_{j\in\mathbb{N}} B_j\right) &= \sum_{j\in\mathbb{N}} \mathbb{P}(B_j) = \lim_{n\to\infty} \sum_{j=1}^{n} \mathbb{P}(B_j) \\ &= \lim_{n\to\infty} \sum_{j=1}^{n} (\mathbb{P}(A_j) - \mathbb{P}(A_{j-1})) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{j\to\infty} \mathbb{P}(A_j)\end{aligned}$$

For $A_j \supset A_{j+1}$ case, use De Morgan's law, $\bigcap_j A_j = (\bigcup_j A_j^c)^c$. And $A_j^c \subset A_{j+1}^c$, use the result above to get

$$\mathbb{P}\left(\bigcap_{j\in\mathbb{N}} A_j\right) = \mathbb{P}\left(\left(\bigcup_{j} A_j^c\right)^c\right) = 1 - \mathbb{P}\left(\bigcup_{j} A_j^c\right) = 1 - \lim_{j\to\infty} \mathbb{P}(A_j^c)$$

Hence

$$\mathbb{P}\left(\bigcap_{j\in\mathbb{N}} A_j\right) = \lim_{j\to\infty} \mathbb{P}(A_j)$$

□

Exercise 1.5 (pp.5 Exercise 1.1.13.)**.** (1) Bernoulli distribution: X is $\{0, 1\}$-valued random variable, Bernoulli distribution with parameter $p \in [0, 1]$ is

$$\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p$$

$p_X(0) = \mathbb{P}(X^{-1}(\{0\})) = 1 - p$, while $p_X(1) = \mathbb{P}(X^{-1}(\{1\})) = p$.

(2) Binomial distribution: X is $\{0, 1, \ldots, n\}$-valued random variable, Binomial distribution with parameter $p \in [0, 1]$ is

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \ k \in \{0, 1, \ldots, n\}$$

For $k \in \{0, 1, \ldots, n\}$, $p_X(k) = \mathbb{P}(X^{-1}(\{k\})) = \binom{n}{k} p^k (1-p)^{n-k}$.

(3) Poisson distribution: X is $\mathbb{N}$-valued random variable, Poisson distribution with parameter $\lambda > 0$ is

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \ k \in \mathbb{N}$$

For $k \in \mathbb{N}$, $p_X(k) = \mathbb{P}(X^{-1}(\{k\})) = e^{-\lambda} \frac{\lambda^k}{k!}$.

Exercise 1.6 (pp.6 1.1.15.)**.** Prove cumulative distribution function satisfies:

(1) $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$.

(2) F_X is increasing and right-continuous.

and provide an example of a cumulative distribution function which is not left-continuous.

Proof. (1) Let $\{x_n\}_{n \geq 1}$ be a monotone increasing sequence of real numbers converging to ∞ and let $A_{x_n} = \{\omega \in \Omega : X(\omega) \leq x_n\}$. This means $A_{x_n} \subset A_{x_{n+1}}$ for all $n \in \mathbb{N}$. Since their countable union $\bigcup_{n=1}^{\infty} A_{x_n}$ is whole set Ω, by continuity of probability,

$$\begin{aligned}
\lim_{x \to \infty} F_X(x) &= \lim_{n \to \infty} F_X(x_n) \\
&= \lim_{n \to \infty} \mathbb{P}(A_{x_n}) \\
&= \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_{x_n} \right) \\
&= \mathbb{P}(\Omega) \\
&= 1
\end{aligned}$$

Let $\{y_n\}_{n \geq 1}$ be a monotone decreasing sequence of real numbers converging to $-\infty$ and let $B_{y_n} = \{\omega \in \Omega : X(\omega) \leq y_n\}$. This means $B_{y_{n+1}} \subset B_{y_n}$ for all $n \in \mathbb{N}$. Since their countable intersection $\bigcap_{n=1}^{\infty} B_{y_n}$ is the empty set

$\varnothing$, again by continuity of probability,

$$\begin{aligned}\lim_{x\to-\infty} F_X(x) &= \lim_{n\to\infty} F_X(y_n) \\ &= \lim_{n\to\infty} \mathbb{P}(B_{y_n}) \\ &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_{y_n}\right) \\ &= \mathbb{P}(\varnothing) \\ &= 0\end{aligned}$$

(2) Let $x, y \in \mathbb{R}$ and $x \leq y$. For $\omega \in \Omega$ such that $X(\omega) \leq x$, then $X(\omega) \leq y$. So the event $\{\omega : X(\omega) \leq x\} \subset \{\omega : X(\omega) \leq y\}$. Hence $\mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y)$, and then $F_X(x) \leq F_X(y)$ when $x \leq y$. F_X is increasing.

To show F_X is right-continuous, for every monotone decreasing sequence $\{x_n\}_{n\geq 1}$ tends to $x \in \mathbb{R}$, then $\{(-\infty, x_n]\}_{n\geq 1}$ is also decreasing. Their intersection is $\bigcap_{n=1}^{\infty}(-\infty, x_n] = (-\infty, x]$.

Similar to (1), $F_X(x) = \mathbb{P}(X \leq x) = P_X((-\infty, x]) = \lim_{n\to\infty} P_X((-\infty, x_n]) = \lim_{n\to\infty} \mathbb{P}(X \leq x_n) = \lim_{n\to\infty} F_X(x_n)$.

Hence for every monotone decreasing sequence $\{x_n\}$ tends to x from right side, we have $F_X(x_n) \to F_X(x)$. Since x is arbitrary, then F_X is right continuous.

To give an example not left-continuous, the counterexample

$$\mathbf{1}_{[0,\infty)}(x) = \begin{cases} 0, & -\infty < x < 0 \\ 1, & 0 \leq x < \infty \end{cases}$$

is sufficient. It is right-continuous at $x = 0$ but not left-continuous. □

2 Expectation and variance

Exercise 2.1 (pp.7 Exercise 1.2.5.)**.** For $\sigma > 0$ and $\bar{x} \in \mathbb{R}$ set $\Pi : \mathbb{R} \to \mathbb{R}_+$ by

$$\Pi(x) := \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \bar{x})^2\right)$$

Check that $\int \Pi(x)\mathrm{d}x = 1$. When $X = N(\bar{x}, \sigma^2)$ for the corresponding univariate random variable, called Gaussian random variable. Check that $\mathbb{E}(X) = \bar{x}$, and $\mathrm{Var}(X) = \sigma^2$.

More generally, for $\bar{x} \in \mathbb{R}^N$ and $P \in M_{N\times N}(\mathbb{R})$ with $P > 0$, set $\Pi : \mathbb{R}^N \to \mathbb{R}_+$ with

$$\Pi(x) := \frac{1}{(2\pi)^{\frac{N}{2}} |P|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\bar{x})^T P^{-1}(x-\bar{x})\right)$$

with $|P| := \det(P)$. Check that $\int \Pi(x)\mathrm{d}x = 1$. We write $X = N(\bar{x}, P)$ for the corresponding multivariate random variable, called N-dim Gaussian random variable. Check that $\mathbb{E}(X) = \bar{x}$, and that $P = \mathbb{E}\left((X-\bar{x})(X-\bar{x})^T\right)$.

Proof. Let $y = \frac{x-\bar{x}}{\sqrt{2}\sigma}$, then $x = \sqrt{2}y\sigma + \bar{x}$ and $\mathrm{d}x = \sqrt{2}\sigma\mathrm{d}y$.

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\bar{x})^2\right) \mathrm{d}x = \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} e^{-y^2}\mathrm{d}y = 1$$

by Poisson's integral.

When $X = N(\bar{x}, \sigma^2)$, its expectation is

$$\begin{aligned}
\mathbb{E}(X) &= \int_{\mathbb{R}} x \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\bar{x})^2\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} (\sqrt{2}\sigma y + \bar{x}) e^{-y^2} \cdot \sqrt{2}\sigma \mathrm{d}y \\
&= \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \sqrt{2}\sigma y e^{-y^2}\mathrm{d}y + \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \bar{x} e^{-y^2}\mathrm{d}y \\
&= 0 + \bar{x} \quad \text{(since the first function is odd function)} \\
&= \bar{x}
\end{aligned}$$

To calculate its variance, we can use:

$$\mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2 + (\mathbb{E}(X))^2 - 2\,\mathbb{E}(X)X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

And

$$\begin{aligned}
\mathbb{E}(X^2) &= \int_{\mathbb{R}} x^2 \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\bar{x})^2\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} (\sqrt{2}\sigma y + \bar{x})^2 e^{-y^2} \sqrt{2}\sigma \mathrm{d}y \\
&= \frac{1}{\sqrt{\pi}} \int_{R} \bar{x}^2 e^{-y^2}\mathrm{d}y + \frac{2\sigma^2}{\sqrt{\pi}} \int_{\mathbb{R}} y^2 e^{-y^2}\mathrm{d}y + \frac{2\sqrt{2}\sigma\bar{x}}{\sqrt{\pi}} \int_{\mathbb{R}} y e^{-y^2}\mathrm{d}y \\
&= \bar{x}^2 + \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{2} + 0 \\
&= \bar{x}^2 + \sigma^2
\end{aligned}$$

Hence its variance is

$$\mathrm{Var}(X) = (\bar{x}^2 + \sigma^2) - \bar{x}^2 = \sigma^2$$

For N-dimension situation, let $y = x - \bar{x}$, then

$$\begin{aligned}
\mathbb{E}(X) &= \int_{\mathbb{R}^n} (y + \bar{x}) \frac{1}{(2\pi)^{\frac{N}{2}} |P|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} y^T P^{-1} y\right) \mathrm{d}y \\
&= \int_{\mathbb{R}^n} y \frac{1}{(2\pi)^{\frac{N}{2}} |P|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} y^T P^{-1} y\right) \mathrm{d}y + \int_{\mathbb{R}^n} \bar{x} \frac{1}{(2\pi)^{\frac{N}{2}} |P|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} y^T P^{-1} y\right) \mathrm{d}y \\
&= 0 + \bar{x} \int_{\mathbb{R}^n} \Pi(y) \mathrm{d}y \quad \left(\text{denote } \Pi(y) \text{ as } \frac{1}{(2\pi)^{\frac{N}{2}} |P|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} y^T P^{-1} y\right) \text{ and it's pdf}\right) \\
&= \bar{x}
\end{aligned}$$

By $y = X - x$, $\mathbb{E}((X - \bar{x})(X - \bar{x})^T) = \mathbb{E}(yy^T)$. Since it is symmetric matrix (we can see it from $(\mathbb{E}(yy^T))^T = \mathbb{E}((yy^T)^T) = \mathbb{E}(yy^T)$ or, more detailed in Exercise), we could find an orthogonal matrix Q such that $P = QDQ^T$ where D is a diagonal matrix, $|P| = |D|$. Let $z = Q^T y$,

$$\begin{aligned}
\mathbb{E}(yy^T) &= \int_{\mathbb{R}^N} yy^T \frac{1}{(2\pi)^{\frac{N}{2}} |P|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} y^T P^{-1} y\right) \mathrm{d}y \\
&= \frac{1}{(2\pi)^{\frac{N}{2}} |D|^{\frac{1}{2}}} \int_{\mathbb{R}^N} (Qz)(Qz)^T \exp\left(-\frac{1}{2} z^T D^{-1} z\right) \mathrm{d}z \\
&= \frac{1}{(2\pi)^{\frac{N}{2}} |D|^{\frac{1}{2}}} Q \left(\int_{\mathbb{R}^N} zz^T \exp\left(-\frac{1}{2} z^T D^{-1} z\right) \mathrm{d}z\right) Q^T \\
&= \frac{1}{(2\pi)^{\frac{N}{2}} |D|^{\frac{1}{2}}} (Q (2\pi)^{\frac{N}{2}} |D|^{\frac{1}{2}} D Q^T) \\
&= QDQ^T \\
&= P
\end{aligned}$$

Hence the variance P is $\mathbb{E}((x - \bar{x})(x - \bar{x})^T)$. □

Exercise 2.2 (pp.7 Exercise 1.2.6.)**.** If $X : \Omega \to \mathbb{R}^N$ is absolutely continuous with pdf Π_X and if $\varphi : \mathbb{R}^N \to \mathbb{R}^N$ is bijective and C^∞, show that $Y := \varphi(X) : \Omega \to \mathbb{R}^N$ is a new absolutely continuous random variable, with pdf Π_Y given by $\Pi_Y(y) = \Pi_X(\varphi^{-1}(y))|J_{\varphi^{-1}}(y)|$. Here, $|J_{\varphi^{-1}}(y)|$ denotes the determinant of the Jacobian matrix of φ^{-1}.

Proof. Since φ is bijective and smooth, then its inverse exists. Let $x = \varphi^{-1}(y)$, then $\mathrm{d}x = |J_{\varphi^{-1}}(y)|\mathrm{d}y$. For every $A \in \sigma_B$,

$$\int_A \Pi_X(x)\mathrm{d}x = \int_{\varphi(A)} \Pi_X(\varphi^{-1}(y))|J_{\varphi^{-1}}(y)|\mathrm{d}y = \int_{\varphi(A)} \Pi_Y(y)\mathrm{d}y$$

Since X is absolutely continuous r.v., for every subset $A \in \sigma_B$, its induced probability measure is equal to that of Y. Hence Y is also absolutely continuous r.v. and has pdf $\Pi_Y(y) = \Pi_X(\varphi^{-1}(y))|J_{\varphi^{-1}}(y)|$. □

Comment 2.3. It can be seen as an application of change of variable method in integration.

3 Some useful inequalities

Lemma 3.1 (Markov's inequality)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a non-negative random variable (meaning that $X(\omega) \geq 0$ for all $\omega \in \Omega$). Then for any $a > 0$ the following inequality holds:*

$$\mathbb{P}(X > a) \leq \frac{1}{a}\,\mathbb{E}(X)$$

Proof. For sample $\omega \in \Omega$, the event

$$\mathbf{1}_{\{X>a\}}(\omega) = \begin{cases} 1, & X(\omega) > a \\ 0, & X(\omega) \leq a \end{cases}$$

Then we have

$$a\mathbf{1}_{\{X>a\}}(\omega) = \begin{cases} a, & X(\omega) > a \\ 0, & X(\omega) \leq a \end{cases}$$

Notice it is included in $X(\omega)$ i.e. $a\mathbf{1}_{\{X>a\}}(\omega) \leq X(\omega)$. Take expectation on both sides and recall $\mathbb{E}(\mathbf{1}_{\{X>a\}}) = \mathbb{P}(X > a)$,

$$a\,\mathbb{P}(X > a) \leq \mathbb{E}(X)$$

leading to Markov's inequality. □

Corollary 3.2 (Chebyshev's inequality)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable. Then for any $a > 0$ the following inequalities hold:*

$$\mathbb{P}(|X| > a) \leq \frac{1}{a^2}\,\mathbb{E}(|X|^2)$$

Proof. The proof of Chebyshev's inequality is similar to that of Markov's. The event $a^2\mathbf{1}_{\{|X|\geq a\}}(\omega)$ is included in $|X(\omega)|^2$ by property of indicator function $\mathbf{1}$. After taking expectation we get

$$a^2\,\mathbb{P}(|X|\geq a)\leq \mathbb{E}(|X|^2)$$

leading to the first result. □

Corollary 3.3 (Chernoff's bound)**.** *Let* $(\Omega,\mathcal{F},\mathbb{P})$ *be a probability space, and let* $X:\Omega\to\mathbb{R}$ *be a random variable. For any* $\lambda>0$*,*

$$\mathbb{P}(X>a)\leq e^{-\lambda a}\,\mathbb{E}(e^{\lambda X})$$

Proof. To prove Chernoff's bound, we notice the event $X(\omega)>a$ is equivalent to $e^{\lambda X(\omega)}>e^{\lambda a}$ since the function $e^{\lambda x}$ is monotone increasing. Apply Markov's inequality, then it is proved.

$$\mathbb{P}(X>a)=\mathbb{P}(e^{\lambda X}>e^{\lambda a})\leq \frac{1}{e^{\lambda a}}\,\mathbb{E}(e^{\lambda X})$$

□

Comment 3.4. These inequalities are applied widely in proof of probability theory and stochastic analysis.

Exercise 3.5 (pp.9 Exercise 1.3.2.)**.** Check that the covariance matrix $\mathrm{Cov}(X)$ is symmetric and positive semi-definite, namely it satisfies $a^T\,\mathrm{Cov}(X)a\geq 0$ for any $a\in\mathbb{R}^N$ with $a\neq 0$.

Proof. The covariance matrix is defined as $\mathrm{Cov}(X)=\mathbb{E}((X-\mathbb{E}(X))(X-\mathbb{E}(X))^T)$, we can let $Y=X-\mathbb{E}(X)$ which is also N-dimension vector. It only needs to show $\mathbb{E}(YY^T)$ is symmetric.

For matrix YY^T, it is obviously symmetric. For its (i,j)-entry, it is also a random variable, hence $\mathbb{E}((YY^T)_{i,j})=\mathbb{E}((YY^T)_{j,i})$. And note $(YY^T)_{i,j}$ is a number, hence $\mathbb{E}((YY^T)_{i,j})=(\mathbb{E}((YY^T)_{i,j}))$, leading to $(\mathbb{E}(YY^T))^T=\mathbb{E}(YY^T)$, then it is symmetric matrix.

For $a\neq 0$, from linearity of expectation,

$$a^T\,\mathrm{Cov}(X)a=a^T\,\mathbb{E}(YY^T)a=\mathbb{E}(a^TYY^T)a=\mathbb{E}(a^TYY^Ta)$$

Let $b=a^TY$, the above becomes $\mathbb{E}(bb^T)$. Notice b is univariate r.v., so $bb^T=b^2$, we can know $\mathbb{E}(bb^T)=\mathbb{E}(b^2)\geq 0$. Hence covariance matrix is positive semi-definite. □

4 L^p space

Exercise 4.1 (pp.10 Exercise 1.4.2.)**.** Show that $L^p(\Omega, \mathcal{F}, \mathbb{P})$ are vector spaces, and that $\|\cdot\|_p$ defines a norm on $L^p(\Omega, \mathcal{F}, \mathbb{P})$. Show also that $L^{p_2}(\Omega, \mathcal{F}, \mathbb{P}) \subset L^{p_1}(\Omega, \mathcal{F}, \mathbb{P})$ whenever $p_2 \geq p_1$.

Proof. When $p \in [1, \infty)$, for $f, g \in L^p(\Omega, \mathcal{F}, \mathbb{P})$, we first prove a useful inequality. Since $|x|^p$ is convex on $x \geq 0$, by the definition of convexity,

$$\left|\frac{1}{2}f + \frac{1}{2}g\right|^p \leq \left|\frac{1}{2}|f| + \frac{1}{2}|g|\right|^p \leq \frac{1}{2}|f|^p + \frac{1}{2}|g|^p$$

Multiply 2^p on both sides to get

$$|f+g|^p \leq \frac{1}{2}|2f|^p + \frac{1}{2}|2g|^p = 2^{p-1}(|f|^p + |g|^p)$$

From

$$\int_\Omega |f+g|^p \mathrm{d}\,\mathbb{P} \leq \int_\Omega 2^{p-1}(|f|^p + |g|^p)\mathrm{d}\,\mathbb{P} < \infty$$

we have the result $f + g \in L^p(\Omega, \mathcal{F}, \mathbb{P})$. For every real number $a \in \mathbb{R}$, from $\int_\Omega |af|^p \mathrm{d}\,\mathbb{P} = |a|^p \int_\Omega |f|^p \mathrm{d}\,\mathbb{P} < \infty$ we can know $af \in L^p(\Omega, \mathcal{F}, \mathbb{P})$. Hence $L^p(\Omega, \mathcal{F}, \mathbb{P})$ is vector space. For its norm $\|\cdot\|_p = \left(\int |\cdot|^p \mathrm{d}\,\mathbb{P}\right)^{\frac{1}{p}}$, obviously it is non-negative for every function $f \in L^p$. When $\|f\|_p = 0$, it's equivalent to $f = 0$ a.s. For real number $a \in \mathbb{R}$, $\|af\|_p = |a|\|f\|_p$. And from Minkowski's inequality, $\|f+g\|_p \leq \|f\|_p + \|g\|_p$. Hence $\|\cdot\|_p$ defines a norm on $L^p(\Omega, \mathcal{F}, \mathbb{P})$.

When $p_2 \geq p_1$, suppose $f \in L^{p_2}$, apply Hölder's inequality for $\frac{p_1}{p_2} + \frac{p_2 - p_1}{p_2} = 1$,

$$\int_\Omega |f|^{p_1} \mathrm{d}\,\mathbb{P} \leq \left(\int_\Omega |f|^{p_1 \cdot \frac{p_2}{p_1}} \mathrm{d}\,\mathbb{P}\right)^{\frac{p_1}{p_2}} \cdot \left(\int_\Omega \mathrm{d}\,\mathbb{P}\right)^{\frac{p_2 - p_1}{p_2}} = \|f\|_{p_2}^{p_1} \cdot (\mathbb{P}(\Omega))^{\frac{p_2 - p_1}{p_2}} < \infty$$

Hence $f \in L^{p_1}$, meaning $L^{p_2}(\Omega, \mathcal{F}, \mathbb{P}) \subset L^{p_1}(\Omega, \mathcal{F}, \mathbb{P})$. □

Comment 4.2. This inequality does not hold in the space where measure of Ω is infinity. In finite measure space, the inequality holds for $1 \leq p_1 \leq p_2 \leq \infty$.

Exercise 4.3 (pp.11 Exercise 1.4.4.)**.** Consider a family $\{X_j\}_{j \in \mathbb{N}} \subset L^p(\Omega, \mathcal{F}, \mathbb{P})$ of univariate IID random variable with 0 mean value. Show that the empirical mean $S_N := \frac{1}{N}\sum_{j=1}^N X_j$ converges to 0 in the L^2-sense, or more precisely $\|S_N\|_2 \to 0$ as $N \to \infty$.

Proof. By linearity of expectation and $\mathbb{E}(X_j) = 0$, set $\mathrm{Var}(X_j) = \sigma^2$

$$\mathbb{E}(S_N) = \mathbb{E}\left(\frac{1}{N}\sum_{j=1}^{N} X_j\right) = \frac{1}{N}\sum_{j=1}^{N} \mathbb{E}(X_j) = 0$$

$$\mathrm{Var}(S_N) = \mathrm{Var}\left(\frac{1}{N}\sum_{j=1}^{N} X_j\right) = \frac{1}{N^2}\sum_{j=1}^{N} \mathrm{Var}(X_j) = \frac{1}{N^2} \cdot N\sigma^2 = \frac{\sigma^2}{N}$$

For $\int_\Omega |S_N|^2 \mathrm{d}\,\mathbb{P} = \|S_N\|_2^2$, note it is equal to $\mathbb{E}(S_N^2)$. We already know it is $(\mathbb{E}(S_N))^2 + \mathrm{Var}(S_N)$. Hence

$$\lim_{N\to\infty} \mathbb{E}(S_N^2) = \lim_{N\to\infty} (\mathbb{E}(S_N))^2 + \lim_{N\to\infty} \mathrm{Var}(S_N) = 0 + \lim_{N\to\infty} \frac{\sigma^2}{N} = 0$$

Hence the empirical mean S_N converges to 0 in L^2 sense. □