

现代信号处理: Homework 5

Due on Jan. 30, 2026

赵钊

学号 12541018

要求: latex

DDL: 2026/1/4 下午 23:59 分前提交 pdf 电子版

电子版以 "homework5-姓名-学号" 形式发送到 12332186@mail.sustech.edu.cn 邮箱

Problem 1

Given observations $x[n]$ for $n = 0, 1, \dots, N-1$, where the samples are i.i.d. and distributed according to $U[\theta_1, \theta_2]$, find a sufficient statistic for $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$.

解答:

设观测数据 x_0, x_1, \dots, x_{N-1} 独立同分布于均匀分布 $U[\theta_1, \theta_2]$, 其中 $\theta_1 < \theta_2$ 。

均匀分布的概率密度函数为:

$$p(x_n | \theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1} \cdot I_{[\theta_1, \theta_2]}(x_n),$$

其中 $I_{[\theta_1, \theta_2]}(\cdot)$ 为区间示性函数。

样本的联合概率密度为:

$$p(\mathbf{x} | \theta_1, \theta_2) = \prod_{n=0}^{N-1} \frac{1}{\theta_2 - \theta_1} \cdot I_{[\theta_1, \theta_2]}(x_n) = \frac{1}{(\theta_2 - \theta_1)^N} \cdot \prod_{n=0}^{N-1} I_{[\theta_1, \theta_2]}(x_n).$$

注意到:

$$\prod_{n=0}^{N-1} I_{[\theta_1, \theta_2]}(x_n) = I\left\{\min_n x_n \geq \theta_1\right\} \cdot I\left\{\max_n x_n \leq \theta_2\right\}.$$

因此,

$$p(\mathbf{x} | \theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^N} \cdot I\left\{\min_n x_n \geq \theta_1\right\} \cdot I\left\{\max_n x_n \leq \theta_2\right\}.$$

根据 Neyman-Fisher 分解定理, 取

$$T(\mathbf{x}) = \left(\min_{0 \leq n \leq N-1} x_n, \max_{0 \leq n \leq N-1} x_n \right),$$

则联合密度可写为 $p(\mathbf{x} | \theta_1, \theta_2) = g(T(\mathbf{x}) | \theta_1, \theta_2) \cdot h(\mathbf{x})$, 其中

$$g(T(\mathbf{x}) | \theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^N} \cdot I\left\{\min_n x_n \geq \theta_1\right\} \cdot I\left\{\max_n x_n \leq \theta_2\right\}, \quad h(\mathbf{x}) = 1.$$

由于 $h(\mathbf{x})$ 不依赖于 θ_1, θ_2 , 因此 $T(\mathbf{x})$ 是 $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ 的充分统计量。

$$T(\mathbf{x}) = \left(\min_{0 \leq n \leq N-1} x[n], \max_{0 \leq n \leq N-1} x[n] \right)$$

Problem 2

For $n = 0, 1, \dots, N-1$, suppose $x[n] = Ar^n + w[n]$, where A is an unknown parameter, r is an unknown constant, and $w[n]$ is white noise with zero mean and variance σ^2 . Find the BLUE of A and its minimum variance. Does the minimum variance tend to zero as $N \rightarrow \infty$?

解答:

观测模型为 $x[n] = Ar^n + w[n]$, 其中 $n = 0, 1, \dots, N-1$, A 是待估计参数, r 是已知常数 (若 r 未知则模型非线性, 无法直接使用线性 BLUE 公式), $w[n]$ 是零均值白噪声, 方差为 σ^2 , 即 $E[w[n]] = 0$, $E[w[n]w[m]] = \sigma^2\delta_{nm}$.

将模型写为向量形式:

$$\mathbf{x} = A\mathbf{h} + \mathbf{w},$$

其中 $\mathbf{x} = [x[0], x[1], \dots, x[N-1]]^T$, $\mathbf{h} = [1, r, r^2, \dots, r^{N-1}]^T$, $\mathbf{w} \sim (\mathbf{0}, \sigma^2\mathbf{I}_N)$.

根据 Gauss-Markov 定理, A 的 BLUE 为:

$$\hat{A}_{\text{BLUE}} = (\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})^{-1} \mathbf{h}^T \mathbf{C}^{-1} \mathbf{x},$$

其中 $\mathbf{C} = \sigma^2 \mathbf{I}$, 故 $\mathbf{C}^{-1} = \frac{1}{\sigma^2} \mathbf{I}$. 代入得:

$$\hat{A}_{\text{BLUE}} = \left(\mathbf{h}^T \frac{1}{\sigma^2} \mathbf{I} \mathbf{h} \right)^{-1} \mathbf{h}^T \frac{1}{\sigma^2} \mathbf{x} = \frac{\sigma^2}{\mathbf{h}^T \mathbf{h}} \cdot \frac{1}{\sigma^2} \mathbf{h}^T \mathbf{x} = \frac{\mathbf{h}^T \mathbf{x}}{\mathbf{h}^T \mathbf{h}}.$$

即:

$$\hat{A}_{\text{BLUE}} = \frac{\sum_{n=0}^{N-1} r^n x[n]}{\sum_{n=0}^{N-1} r^{2n}}.$$

最小方差为:

$$\text{Var}(\hat{A}_{\text{BLUE}}) = (\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})^{-1} = \frac{\sigma^2}{\mathbf{h}^T \mathbf{h}} = \frac{\sigma^2}{\sum_{n=0}^{N-1} r^{2n}}.$$

因此:

$$\text{Var}(\hat{A}_{\text{BLUE}}) = \frac{\sigma^2}{\sum_{n=0}^{N-1} r^{2n}}.$$

计算 $\sum_{n=0}^{N-1} r^{2n}$: 若 $|r| \neq 1$, $\sum_{n=0}^{N-1} r^{2n} = \frac{1-r^{2N}}{1-r^2}$; 若 $|r| = 1$, $\sum_{n=0}^{N-1} r^{2n} = N$.

当 $N \rightarrow \infty$ 时: 若 $|r| < 1$, $\sum_{n=0}^{N-1} r^{2n} \rightarrow \frac{1}{1-r^2}$, 方差 $\rightarrow \sigma^2(1-r^2) > 0$; 若 $|r| = 1$, 方差 $= \frac{\sigma^2}{N} \rightarrow 0$;

若 $|r| > 1$, $\sum_{n=0}^{N-1} r^{2n} \sim \frac{r^{2N}}{r^2-1}$, 方差 $\sim \sigma^2 \frac{r^2-1}{r^{2N}} \rightarrow 0$.

综上, 当 $N \rightarrow \infty$ 时, 最小方差在 $|r| \geq 1$ 时趋于零, 在 $|r| < 1$ 时不趋于零。

Problem 3

The observed i.i.d. samples $\{x[0], x[1], \dots, x[N-1]\}$ follow the distributions below:

a. Laplace:

$$p(x[n]; \mu) = \frac{1}{2} \exp[-|x[n] - \mu|]$$

b. Gaussian:

$$p(x[n]; \mu) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x[n] - \mu)^2\right]$$

Find the BLUE of the mean μ for both cases. Also, explain the MVU estimator of μ .

解答:

1、高斯分布

观测模型: $x[n] = \mu + w[n]$, 其中 $w[n] \sim \mathcal{N}(0, 1)$ 独立同分布。

噪声协方差矩阵为 $\mathbf{C} = \mathbf{I}$. 根据 Gauss-Markov 定理, μ 的 BLUE 为:

$$\hat{\mu}_{\text{BLUE}} = (\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{C}^{-1} \mathbf{x} = (\mathbf{1}^T \mathbf{I} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{x} = \frac{1}{N} \sum_{n=0}^{N-1} x[n].$$

由于高斯分布是指数族, 样本均值 $\bar{x} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$ 是充分完备统计量。根据 Lehmann-Scheffé 定理, 它是 MVU (最小方差无偏估计), 方差为 $\frac{1}{N}$, 达到 Cramér-Rao 下界。

因此对于高斯分布: BLUE = MVU = \bar{x} 。

2、拉普拉斯分布

观测模型: $x[n] = \mu + w[n]$, 其中 $w[n]$ 服从拉普拉斯分布 $p(w) = \frac{1}{2}e^{-|w|}$ 。

拉普拉斯分布的方差为 2, 故噪声协方差矩阵为 $\mathbf{C} = 2\mathbf{I}$ 。BLUE 为:

$$\hat{\mu}_{\text{BLUE}} = (\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{C}^{-1} \mathbf{x} = \left(\mathbf{1}^T \frac{1}{2} \mathbf{I} \mathbf{1} \right)^{-1} \mathbf{1}^T \frac{1}{2} \mathbf{I} \mathbf{x} = \frac{2}{N} \cdot \frac{1}{2} \sum_{n=0}^{N-1} x[n] = \frac{1}{N} \sum_{n=0}^{N-1} x[n].$$

因此拉普拉斯分布的 BLUE 也是样本均值。

然而, 对于拉普拉斯分布, μ 的 MVU 估计是样本中位数。原因如下: 拉普拉斯分布下, 样本中位数是最大似然估计, 且是充分统计量。样本均值方差为 $\frac{2}{N}$, 样本中位数方差约为 $\frac{1}{N}$ (渐近), 后者更小。样本中位数是无偏的且达到 Fisher 信息量下界, 因此是 MVU。

Problem 4

Assume that the observed signal is $x[n] = As[n] + w[n]$, for $n = 0, 1, \dots, N-1$, where $w[n]$ is noise with zero mean and covariance matrix \mathbf{C} , and $s[n]$ is a known signal. The amplitude A is the parameter to be estimated. Find the BLUE of A . Discuss what happens if the characteristic vector of \mathbf{C} is $\mathbf{s} = [s[0] \ s[1] \ \dots \ s[N-1]]^T$. Also, find the minimum variance.

解答:

根据 Gauss-Markov 定理, A 的 BLUE 为:

$$\hat{A}_{\text{BLUE}} = (\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s})^{-1} \mathbf{s}^T \mathbf{C}^{-1} \mathbf{x} = \frac{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}$$

$$\text{Var}(\hat{A}_{\text{BLUE}}) = (\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s})^{-1} = \frac{1}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}$$

假设 \mathbf{s} 是 \mathbf{C} 的一个特征向量, 对应特征值为 λ , 即:

$$\mathbf{C}\mathbf{s} = \lambda\mathbf{s} \quad \Rightarrow \quad \mathbf{C}^{-1}\mathbf{s} = \frac{1}{\lambda}\mathbf{s}.$$

代入 BLUE 公式:

$$\hat{A}_{\text{BLUE}} = \frac{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}} = \frac{\mathbf{s}^T \left(\frac{1}{\lambda} \mathbf{s} \right) \mathbf{x}}{\mathbf{s}^T \left(\frac{1}{\lambda} \mathbf{s} \right) \mathbf{s}} = \frac{\frac{1}{\lambda} \mathbf{s}^T \mathbf{x}}{\frac{1}{\lambda} \mathbf{s}^T \mathbf{s}} = \frac{\mathbf{s}^T \mathbf{x}}{\mathbf{s}^T \mathbf{s}}.$$

此时 BLUE 退化为普通最小二乘估计 (相当于噪声为白噪声时的结果)。

方差为:

$$\text{Var}(\hat{A}_{\text{BLUE}}) = \frac{1}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}} = \frac{1}{\mathbf{s}^T \left(\frac{1}{\lambda} \mathbf{s} \right)} = \frac{\lambda}{\mathbf{s}^T \mathbf{s}}.$$

当信号向量 \mathbf{s} 恰为噪声协方差矩阵 \mathbf{C} 的特征向量时, BLUE 简化为最小二乘估计, 无需进行预白化处理。此时估计器的方差与特征值 λ 成正比, λ 越大 (该方向噪声功率越大), 估计方差越大。

Problem 5

Prove the linearity property of the BLUE with respect to linear transformations of θ . Specifically, if we wish to estimate

$$\alpha = \mathbf{B}\theta + \mathbf{b},$$

where \mathbf{B} is a known $p \times p$ invertible matrix and \mathbf{b} is a known $p \times 1$ vector, prove that the BLUE of α is given by

$$\hat{\alpha} = \mathbf{B}\hat{\theta} + \mathbf{b},$$

where $\hat{\theta}$ is the BLUE of θ . Assume the data model $\mathbf{x} = \mathbf{H}\theta + \mathbf{w}$, where $E(\mathbf{w}) = \mathbf{0}$ and $E(\mathbf{w}\mathbf{w}^T) = \mathbf{C}$. Hint: Substitute θ for α in the data model.

解答:

考虑线性变换:

$$\alpha = \mathbf{B}\theta + \mathbf{b},$$

其中 \mathbf{B} 是 $p \times p$ 可逆矩阵, \mathbf{b} 是 $p \times 1$ 已知向量。

线性性与无偏性:

由于 $\hat{\theta}$ 是 \mathbf{x} 的线性函数 ($\hat{\theta} = \mathbf{G}\mathbf{x}$, 其中 $\mathbf{G} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}$),

$$\hat{\alpha} = \mathbf{B}\hat{\theta} + \mathbf{b} = \mathbf{B}\mathbf{G}\mathbf{x} + \mathbf{b}$$

也是 \mathbf{x} 的线性函数 (加常数项 \mathbf{b})。

无偏性:

$$E[\hat{\alpha}] = \mathbf{B}E[\hat{\theta}] + \mathbf{b} = \mathbf{B}\theta + \mathbf{b} = \alpha.$$

最小方差性:

设 $\tilde{\alpha} = \mathbf{A}\mathbf{x} + \mathbf{b}$ 是 α 的任意线性无偏估计, 则无偏性要求:

$$E[\tilde{\alpha}] = \mathbf{A}\mathbf{H}\theta + \mathbf{b} = \mathbf{B}\theta + \mathbf{b} \quad \forall \theta,$$

即 $\mathbf{A}\mathbf{H} = \mathbf{B}$ 。

$\tilde{\alpha}$ 的协方差为:

$$\text{Cov}(\tilde{\alpha}) = \mathbf{A}\mathbf{C}\mathbf{A}^T.$$

考虑 $\hat{\theta}$ 的协方差:

$$\text{Cov}(\hat{\theta}) = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}.$$

由于 $\hat{\theta}$ 是 θ 的 BLUE, 对于任意满足 $\mathbf{D}\mathbf{H} = \mathbf{I}$ 的矩阵 \mathbf{D} , 有

$$\text{Cov}(\mathbf{D}\mathbf{x}) \succeq \text{Cov}(\hat{\theta}),$$

其中 \succeq 表示矩阵半正定序。

令 $\mathbf{A}_0 = \mathbf{B}\mathbf{G}$, 其中 $\mathbf{G} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}$ 。则 $\mathbf{A}_0\mathbf{H} = \mathbf{B}\mathbf{G}\mathbf{H} = \mathbf{B}$, 且

$$\hat{\alpha} = \mathbf{A}_0\mathbf{x} + \mathbf{b}.$$

对于任意满足 $\mathbf{A}\mathbf{H} = \mathbf{B}$ 的 \mathbf{A} , 考虑:

$$\mathbf{A} = \mathbf{A}_0 + \mathbf{E},$$

则 $\mathbf{E}\mathbf{H} = \mathbf{0}$ 。

计算协方差差:

$$\begin{aligned} \text{Cov}(\tilde{\alpha}) - \text{Cov}(\hat{\alpha}) &= (\mathbf{A}_0 + \mathbf{E})\mathbf{C}(\mathbf{A}_0 + \mathbf{E})^T - \mathbf{A}_0\mathbf{C}\mathbf{A}_0^T \\ &= \mathbf{A}_0\mathbf{C}\mathbf{E}^T + \mathbf{E}\mathbf{C}\mathbf{A}_0^T + \mathbf{E}\mathbf{C}\mathbf{E}^T. \end{aligned}$$

由于 $\mathbf{E}\mathbf{H} = \mathbf{0}$, 且 $\mathbf{A}_0 = \mathbf{B}(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}$, 可以验证 $\mathbf{A}_0\mathbf{C}\mathbf{E}^T = \mathbf{0}$, $\mathbf{E}\mathbf{C}\mathbf{A}_0^T = \mathbf{0}$. 因此:

$$\text{Cov}(\tilde{\hat{\alpha}}) - \text{Cov}(\hat{\alpha}) = \mathbf{E}\mathbf{C}\mathbf{E}^T \succeq \mathbf{0}.$$

这表明 $\hat{\alpha}$ 的协方差矩阵最小。

因此, $\hat{\alpha} = \mathbf{B}\hat{\theta} + \mathbf{b}$ 是 $\alpha = \mathbf{B}\theta + \mathbf{b}$ 的 BLUE。

Problem 6

For the general linear model

$$\mathbf{x} = \mathbf{H}\theta + \mathbf{s} + \mathbf{w},$$

where \mathbf{s} is a known $N \times 1$ vector, $E(\mathbf{w}) = \mathbf{0}$, and $E(\mathbf{w}\mathbf{w}^T) = \mathbf{C}$, find the BLUE of θ .

解答: 考虑线性模型:

$$\mathbf{x} = \mathbf{H}\theta + \mathbf{s} + \mathbf{w},$$

其中:

- $\mathbf{x} \in \mathbb{R}^N$: 观测向量
- $\mathbf{H} \in \mathbb{R}^{N \times p}$: 已知设计矩阵
- $\theta \in \mathbb{R}^p$: 待估计参数向量
- $\mathbf{s} \in \mathbb{R}^N$: 已知偏移向量
- $\mathbf{w} \in \mathbb{R}^N$: 噪声向量, 满足 $E[\mathbf{w}] = \mathbf{0}$, $\text{Cov}(\mathbf{w}) = \mathbf{C}$

将已知偏移量 \mathbf{s} 移到左边:

$$\mathbf{x} - \mathbf{s} = \mathbf{H}\theta + \mathbf{w}.$$

定义新观测向量:

$$\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{s}.$$

则模型简化为标准形式:

$$\tilde{\mathbf{x}} = \mathbf{H}\theta + \mathbf{w}, \quad E[\mathbf{w}] = \mathbf{0}, \quad \text{Cov}(\mathbf{w}) = \mathbf{C}.$$

根据 Gauss-Markov 定理, θ 的 BLUE 为:

$$\hat{\theta}_{\text{BLUE}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\tilde{\mathbf{x}}.$$

代入 $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{s}$, 得:

$$\hat{\theta}_{\text{BLUE}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\mathbf{x} - \mathbf{s}).$$

Problem 7

We observe N i.i.d. samples from the following PDFs:

a. Gaussian:

$$p(x; \mu) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x - \mu)^2\right]$$

b. Exponential:

$$p(x; \lambda) = \begin{cases} \lambda \exp(-\lambda x) & x > 0 \\ 0 & x < 0 \end{cases}$$

In each case, find the MLE of the unknown parameter and verify that it indeed maximizes the likelihood function. Is the estimator meaningful?

解答:

高斯分布:

观测 N 个独立同分布样本 $x_1, \dots, x_N \sim \mathcal{N}(\mu, 1)$ 。

似然函数:

$$L(\mu) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(x_i - \mu)^2 \right] = (2\pi)^{-N/2} \exp \left[-\frac{1}{2} \sum_{i=1}^N (x_i - \mu)^2 \right].$$

对数似然函数:

$$\ell(\mu) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N (x_i - \mu)^2.$$

求导:

$$\frac{d\ell}{d\mu} = \sum_{i=1}^N (x_i - \mu) = \sum_{i=1}^N x_i - N\mu.$$

令导数为零:

$$\sum_{i=1}^N x_i - N\mu = 0 \quad \Rightarrow \quad \hat{\mu}_{\text{MLE}} = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}.$$

二阶导数验证:

$$\frac{d^2\ell}{d\mu^2} = -N < 0,$$

因此 $\hat{\mu}_{\text{MLE}} = \bar{x}$ 确实最大化似然函数。

指数分布:

观测 N 个独立同分布样本 $x_1, \dots, x_N \sim \text{Exp}(\lambda)$, 概率密度函数:

$$p(x; \lambda) = \lambda e^{-\lambda x}, \quad x > 0.$$

似然函数:

$$L(\lambda) = \prod_{i=1}^N \lambda e^{-\lambda x_i} = \lambda^N \exp \left(-\lambda \sum_{i=1}^N x_i \right).$$

对数似然函数:

$$\ell(\lambda) = N \ln \lambda - \lambda \sum_{i=1}^N x_i.$$

求导:

$$\frac{d\ell}{d\lambda} = \frac{N}{\lambda} - \sum_{i=1}^N x_i.$$

令导数为零:

$$\frac{N}{\lambda} - \sum_{i=1}^N x_i = 0 \quad \Rightarrow \quad \hat{\lambda}_{\text{MLE}} = \frac{N}{\sum_{i=1}^N x_i} = \frac{1}{\bar{x}}.$$

二阶导数验证:

$$\frac{d^2\ell}{d\lambda^2} = -\frac{N}{\lambda^2} < 0,$$

因此 $\hat{\lambda}_{\text{MLE}} = 1/\bar{x}$ 确实最大化似然函数。

Problem 8

The following is the formal definition of a consistent estimator: If for any given $\epsilon > 0$, it satisfies

$$\lim_{N \rightarrow \infty} \Pr \left\{ |\hat{\theta} - \theta| > \epsilon \right\} = 0,$$

then the estimator $\hat{\theta}$ is consistent.

Prove that for the problem of estimating a DC level A in white Gaussian noise with known variance, the sample mean is a consistent estimator. Hint: Use Chebyshev's inequality.

解答:

$$E[\hat{A}] = E \left[\frac{1}{N} \sum_{n=1}^N (A + w[n]) \right] = \frac{1}{N} \sum_{n=1}^N A + \frac{1}{N} \sum_{n=1}^N E[w[n]] = A + 0 = A.$$

由于 $w[n]$ 独立同分布:

$$\text{Var}(\hat{A}) = \text{Var} \left(\frac{1}{N} \sum_{n=1}^N x[n] \right) = \frac{1}{N^2} \sum_{n=1}^N \text{Var}(x[n]) = \frac{1}{N^2} \cdot N\sigma^2 = \frac{\sigma^2}{N}.$$

切比雪夫不等式: 对任意随机变量 X 满足 $E[X] = \mu$, $\text{Var}(X) = \sigma_X^2$, 有

$$\Pr(|X - \mu| \geq \epsilon) \leq \frac{\sigma_X^2}{\epsilon^2}, \quad \forall \epsilon > 0.$$

令 $X = \hat{A}$, 则 $\mu = A$, $\sigma_X^2 = \frac{\sigma^2}{N}$, 于是:

$$\Pr(|\hat{A} - A| > \epsilon) \leq \frac{\sigma^2}{N\epsilon^2}.$$

当 $N \rightarrow \infty$ 时:

$$\lim_{N \rightarrow \infty} \Pr(|\hat{A} - A| > \epsilon) \leq \lim_{N \rightarrow \infty} \frac{\sigma^2}{N\epsilon^2} = 0.$$

由于概率非负, 必有:

$$\lim_{N \rightarrow \infty} \Pr(|\hat{A} - A| > \epsilon) = 0.$$

根据一致估计量的定义 (对于任意 $\epsilon > 0$, $\lim_{N \rightarrow \infty} \Pr(|\hat{\theta} - \theta| > \epsilon) = 0$), 样本均值 $\hat{A} = \frac{1}{N} \sum_{n=1}^N x[n]$ 是 A 的一致估计量。