大数定律

续本达

清华大学 工程物理系

2023-10-29 清华

复习

协方差

称 $\mathrm{E}([X-\mathrm{E}(X)][Y-\mathrm{E}(Y)])$ 为随机变量 X,Y 的 协方差 。记为

$$\mathrm{Cov}(X,Y)=\mathrm{E}([X-\mathrm{E}(X)][Y-\mathrm{E}(Y)])=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)$$

若 $\mathrm{Var}(X)>0,\mathrm{Var}(Y)>0$ ，称

$$\mathrm{E}\left(\frac{[X-\mathrm{E}(X)][Y-\mathrm{E}(Y)]}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}\right)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

为随机变量 X,Y 的 相关系数 ，记为

$$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

若 $\rho_{XY}=0$，称随机变量 X,Y 不相关 。

$$\mathrm{Cov}(X,X)=\mathrm{Var}(X)$$

称为 X 的方差

协方差矩阵

设 n 维随机变量 $\vec{X} = (X_1, X_2, \cdots, X_n)$ 每个分量的方差存在，任意两个分量的协方差存在，则称

$$\begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Var}(X_n) \end{bmatrix}$$

$\vec{X}$ 的 协方差矩阵 ， 也称为 方差-协方差 矩阵。记为 $\mathrm{Var}(\vec{X})$。若记 $\vec{X} - \mathrm{E}(\vec{X})$ 为 n 维列向量，则

$$\mathrm{Var}(\vec{X}) = \mathrm{E}[(\vec{X} - \mathrm{E}(\vec{X}))(\vec{X} - \mathrm{E}(\vec{X}))^\intercal]$$

协方差矩阵 对称非负定。

多维正态分布

$$f(\vec{x}) = \frac{1}{2\pi\sqrt{|\Sigma^2|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-2}(\vec{x}-\vec{\mu})^\intercal}$$

其中

$$\Sigma^2 = \mathrm{Var}(\vec{X}), \vec{\mu} = \mathrm{E}(\vec{X}), \vec{x} = (x_1, x_2)$$

高斯过程

- 当 $\vec{x}$ 的维度趋于无穷时，可以描述一个高斯过程。
- 高斯过程可以看成随机的函数，是随机过程的一种。
- 深度神经网络能有效逼近高斯过程。

多维正态分布

$$f(\vec{x}) = \frac{1}{2\pi\sqrt{|\Sigma^2|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-2}(\vec{x}-\vec{\mu})^\intercal}$$

其中

$$\Sigma^2 = \mathrm{Var}(\vec{X}), \vec{\mu} = \mathrm{E}(\vec{X}), \vec{x} = (x_1, x_2)$$

高斯过程

- 当 $\vec{x}$ 的维度趋于无穷时，可以描述一个高斯过程。
- 高斯过程可以看成随机的函数，是随机过程的一种。
- 深度神经网络能有效逼近高斯过程。

大数定律

引子

为何能以某事件发生的频率作为该事件的概率的估计?

- 柯氏概率公理不关心概率解读，这是一个需要证明的命题。

引子

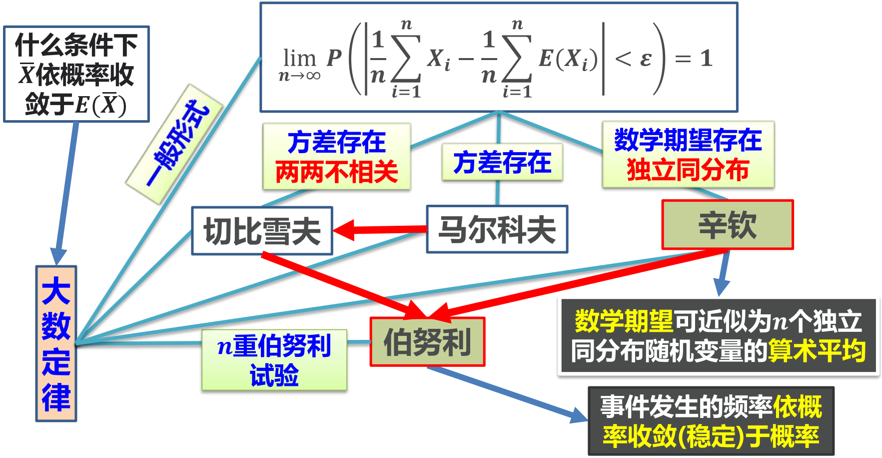
为何能以某事件发生的频率作为该事件的概率的估计?

- 柯氏概率公理不关心概率解读，这是一个需要证明的命题。

引子

为何能以某事件发生的频率作为该事件的概率的估计?

- 柯氏概率公理不关心概率解读，这是一个需要证明的命题。



回顾收敛 $\lim_{n\to\infty} X_n = X$ 的含义

$$\forall \epsilon \exists N (n > N \to |X_n - X| < \epsilon)$$

定义

设 X_n 为一随机变量序列，X 为一随机变量。如果对任意正数 ϵ，有

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) \to 0 \text{ 或 } \lim_{n\to\infty} P(|X_n - X| < \epsilon) \to 1$$

即

$$\forall \epsilon > 0, \epsilon' \exists N \left[n > N \to 1 - P(|X_n - X| < \epsilon) < \epsilon' \right]$$

则称序列 X_n 依概率收敛于 X，记作

$$X_n \xrightarrow{P} X$$

依概率收敛

回顾收敛 $\lim_{n\to\infty} X_n = X$ 的含义

$$\forall\epsilon\exists N(n > N \to |X_n - X| < \epsilon)$$

定义

设 X_n 为一随机变量序列，X 为一随机变量。如果对任意正数 ϵ，有

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) \to 0 \text{ 或 } \lim_{n\to\infty} P(|X_n - X| < \epsilon) \to 1$$

即

$$\forall\epsilon > 0, \epsilon' \exists N \left[n > N \to 1 - P(|X_n - X| < \epsilon) < \epsilon'\right]$$

则称序列 X_n 依概率收敛于 X，记作

$$X_n \xrightarrow{P} X$$

依概率收敛

回顾收敛 $\lim_{n\to\infty} X_n = X$ 的含义

$$\forall\epsilon\exists N(n > N \to |X_n - X| < \epsilon)$$

定义

设 X_n 为一随机变量序列，X 为一随机变量。如果对任意正数 ϵ，有

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) \to 0 \text{ 或 } \lim_{n\to\infty} P(|X_n - X| < \epsilon) \to 1$$

即

$$\forall\epsilon > 0, \epsilon'\exists N\left[n > N \to 1 - P(|X_n - X| < \epsilon) < \epsilon'\right]$$

则称序列 X_n 依概率收敛于 X，记作

$$X_n \xrightarrow{P} X$$

依概率收敛

回顾收敛 $\lim_{n\to\infty} X_n = X$ 的含义

$$\forall\epsilon\exists N(n > N \to |X_n - X| < \epsilon)$$

定义

设 X_n 为一随机变量序列，X 为一随机变量。如果对任意正数 ϵ，有

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) \to 0 \text{ 或 } \lim_{n\to\infty} P(|X_n - X| < \epsilon) \to 1$$

即

$$\forall\epsilon > 0, \epsilon' \exists N \left[n > N \to 1 - P(|X_n - X| < \epsilon) < \epsilon'\right]$$

则称序列 X_n 依概率收敛于 X，记作

$$X_n \xrightarrow{P} X$$

辛钦 Khinchin 大数定律

定理

设 $X_1, X_2, \cdots, X_n$ 是相互独立的随机变量序列，服从同一分布，且具有 $E(X_k) = \mu (k = 1, 2, \cdots, n)$。则对任意正数 ϵ，有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \epsilon\right) = 1$$

定义 $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ ，则

$$\overline{X}_n \xrightarrow{P} \mu$$

数学期望可以由 n 个独立同分布的随机变量的算术平均值近似。

辛钦 Khinchin 大数定律

定理

设 $X_1, X_2, \cdots, X_n$ 是相互独立的随机变量序列，服从同一分布，且具有 $E(X_k) = \mu (k = 1, 2, \cdots, n)$。则对任意正数 ϵ，有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \epsilon\right) = 1$$

定义 $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ ，则

$$\overline{X}_n \xrightarrow{P} \mu$$

数学期望可以由 n 个独立同分布的随机变量的算术平均值近似。

伯努利 Bernoulli 大数定律

设 n_A 是 n 次独立重复试验中事件 A 发生的次数，p 是每次试验 A 发生的概率，

$$\forall \epsilon > 0 \lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| < \epsilon\right) = 1$$

因为 $n_A = X_1 + X_2 + \cdots + X_n$，其中 $X_1, X_2, \cdots, X_n$ 相互独立，且都服从参数为 p 的 0-1 分布，因而 $\mathrm{E}(X_k) = p$。由辛钦大数定律有

$$1 = \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_i X_i - p\right| < \epsilon\right) = \lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| < \epsilon\right)$$

期望与概率的等价性

在概率的频率统计定义中，事件 A 发生的频率 $\frac{n_A}{n}$ 稳定于事件 A 在一次试验中发生的概率是指：
频率 $\frac{n_A}{n}$ 与 p 有较大偏差 $\left|\frac{n_A}{n} - p\right| \geq \epsilon$ 是小概率事件。因而在 n 足够大时，可以用频率近似代替 p。

$$\text{概率公理} \xrightarrow{\text{辛钦大数定律}} \text{平均趋于期望} \xrightarrow{\text{伯努利大数定律}} \text{频率趋于概率}$$

切比雪夫不等式

切比雪夫 Chebyshev 不等式

定理

设随机变量 X 有数学期望 $\mathrm{E}(X) = \mu$ 和方差 $\mathrm{Var}(X) = \sigma^2$，则：

$$\forall \epsilon > 0 P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

仅就连续型随机变量的情况来证明。设 X 的概率密度函数为 $f(x)$，则有

$$P(|X - \mu| \geq \epsilon) = \int_{|x-\mu|\geq\epsilon} f(x)dx$$

$$\leq \int_{|x-\mu|\geq\epsilon} \frac{|x-\mu|^2}{\epsilon^2} f(x)dx$$

$$\leq \frac{1}{\epsilon^2} \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx = \frac{\sigma^2}{\epsilon^2}$$

启示：一个随机变量的方差存在，是很强的条件。

切比雪夫 Chebyshev 不等式

定理

设随机变量 X 有数学期望 $\mathrm{E}(X)=\mu$ 和方差 $\mathrm{Var}(X)=\sigma^2$，则：

$$\forall\epsilon>0 P(|X-\mu|\geq\epsilon)\leq\frac{\sigma^2}{\epsilon^2}.$$

仅就连续型随机变量的情况来证明。设 X 的概率密度函数为 $f(x)$，则有

$$\begin{aligned}P(|X-\mu|\geq\epsilon)&=\int_{|x-\mu|\geq\epsilon}f(x)dx\\&\leq\int_{|x-\mu|\geq\epsilon}\frac{|x-\mu|^2}{\epsilon^2}f(x)dx\\&\leq\frac{1}{\epsilon^2}\int_{-\infty}^{+\infty}(x-\mu)^2f(x)dx=\frac{\sigma^2}{\epsilon^2}\end{aligned}$$

启示：一个随机变量的方差存在，是很强的条件。

切比雪夫 Chebyshev 不等式

定理

设随机变量 X 有数学期望 $\mathrm{E}(X)=\mu$ 和方差 $\mathrm{Var}(X)=\sigma^2$，则：

$$\forall\epsilon>0 P(|X-\mu|\geq\epsilon)\leq\frac{\sigma^2}{\epsilon^2}.$$

仅就连续型随机变量的情况来证明。设 X 的概率密度函数为 $f(x)$，则有

$$\begin{aligned}P(|X-\mu|\geq\epsilon)&=\int_{|x-\mu|\geq\epsilon}f(x)dx\\&\leq\int_{|x-\mu|\geq\epsilon}\frac{|x-\mu|^2}{\epsilon^2}f(x)dx\\&\leq\frac{1}{\epsilon^2}\int_{-\infty}^{+\infty}(x-\mu)^2f(x)dx=\frac{\sigma^2}{\epsilon^2}\end{aligned}$$

启示：一个随机变量的方差存在，是很强的条件。

切比雪夫不等式应用：方差粗估

切比雪夫不等式可以使我们在随机变量 X 的分布未知的情况下，对事件 $|X-\mu| < \epsilon$ 的概率做出估计。

例

设一大批种子中良种占 $\frac{1}{6}$。试用切比雪夫不等式估计在任选的 6000 粒种子中，良种比例与 $\frac{1}{6}$ 比较上下小于 1 % 的概率范围。

设 X 表示 6000 粒种子中的良种数，$X \sim b(6000, \frac{1}{6})$，$\mathrm{E}(X) = 1000$，$\mathrm{Var}(X) = \frac{5000}{6}$

$$P(|X-1000| < 60) = 1 - P(|X-1000| \geq 60) \geq 1-\frac{5000/6}{60^2} = 0.7685$$

切比雪夫不等式应用：方差粗估

切比雪夫不等式可以使我们在随机变量 X 的分布未知的情况下，对事件 $|X-\mu| < \epsilon$ 的概率做出估计。

例

设一大批种子中良种占 $\frac{1}{6}$。试用切比雪夫不等式估计在任选的 6000 粒种子中，良种比例与 $\frac{1}{6}$ 比较上下小于 1 % 的概率范围。

设 X 表示 6000 粒种子中的良种数，$X \sim b(6000, \frac{1}{6})$，$\mathrm{E}(X) = 1000$，$\mathrm{Var}(X) = \frac{5000}{6}$

$$P(|X-1000| < 60) = 1 - P(|X-1000| \geq 60) \geq 1 - \frac{5000/6}{60^2} = 0.7685$$

切比雪夫不等式应用：方差粗估

切比雪夫不等式可以使我们在随机变量 X 的分布未知的情况下，对事件 $|X-\mu| < \epsilon$ 的概率做出估计。

例

设一大批种子中良种占 $\frac{1}{6}$。试用切比雪夫不等式估计在任选的 6000 粒种子中，良种比例与 $\frac{1}{6}$ 比较上下小于 1 % 的概率范围。

设 X 表示 6000 粒种子中的良种数，$X \sim b(6000, \frac{1}{6})$，$\mathrm{E}(X) = 1000$，$\mathrm{Var}(X) = \frac{5000}{6}$

$$P(|X-1000| < 60) = 1 - P(|X-1000| \geq 60) \geq 1 - \frac{5000/6}{60^2} = 0.7685$$

近似方法的比较

二项分布

$$P\left(\left|\frac{X}{6000}-\frac{1}{6}\right| < 0.01\right) = P(940 < X < 1060)$$
$$= \sum_{k=941}^{1059} \binom{6000}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{6000-k} = 0.9607$$

Poisson 分布近似

$$\sum_{k=941}^{1059} \frac{1000^k e^{-1000}}{k!} = 0.9401$$

近似方法的比较

二项分布

$$P\left(\left|\frac{X}{6000}-\frac{1}{6}\right|<0.01\right)=P(940<X<1060)$$
$$=\sum_{k=941}^{1059}\binom{6000}{k}\left(\frac{1}{6}\right)^{k}\left(\frac{5}{6}\right)^{6000-k}=0.9607$$

Poisson 分布近似

$$\sum_{k=941}^{1059}\frac{1000^{k}e^{-1000}}{k!}=0.9401$$

证明大数定律

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

- 条件假定 $\mathrm{Var}(X_k) = \sigma^2$ 存在

$$\mathrm{E}\left[\overline{X}_n\right] = \mu,\ \mathrm{Var}\left[\overline{X}_n\right] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n}\sigma^2$$

- 由切比雪夫不等式

$$P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) \ge 1 - \frac{\sigma^2}{n\epsilon^2}$$
$$\Longrightarrow \lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) = 1$$

证明大数定律

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

- <u>条件假定</u> $\mathrm{Var}(X_k) = \sigma^2$ 存在

$$\mathrm{E}\left[\overline{X}_n\right] = \mu, \quad \mathrm{Var}\left[\overline{X}_n\right] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n}\sigma^2$$

- 由切比雪夫不等式

$$P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) \geq 1 - \frac{\sigma^2}{n\epsilon^2}$$

$$\Longrightarrow \lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) = 1$$

证明大数定律

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

- <u>条件假定</u> $\mathrm{Var}(X_k) = \sigma^2$ 存在

$$\mathrm{E}\left[\overline{X}_n\right] = \mu,\ \mathrm{Var}\left[\overline{X}_n\right] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n}\sigma^2$$

- 由切比雪夫不等式

$$P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) \geq 1 - \frac{\sigma^2}{n\epsilon^2}$$

$$\Longrightarrow \lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) = 1$$

证明大数定律

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

- <u>条件假定</u> $\mathrm{Var}(X_k) = \sigma^2$ 存在

$$\mathrm{E}\left[\overline{X}_n\right] = \mu,\ \mathrm{Var}\left[\overline{X}_n\right] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n}\sigma^2$$

- 由切比雪夫不等式

$$P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) \geq 1 - \frac{\sigma^2}{n\epsilon^2}$$
$$\Longrightarrow \lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) = 1$$

证明大数定律

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

- 条件假定 $\mathrm{Var}(X_k) = \sigma^2$ 存在

$$\mathrm{E}\left[\overline{X}_n\right] = \mu,\ \mathrm{Var}\left[\overline{X}_n\right] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n}\sigma^2$$

- 由切比雪夫不等式

$$P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) \geq 1 - \frac{\sigma^2}{n\epsilon^2}$$

$$\Longrightarrow \lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) = 1$$

证明大数定律

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

- <u>条件假定</u> $\mathrm{Var}(X_k) = \sigma^2$ 存在

$$\mathrm{E}\left[\overline{X}_n\right] = \mu,\ \mathrm{Var}\left[\overline{X}_n\right] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n}\sigma^2$$

- 由切比雪夫不等式

$$P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) \geq 1 - \frac{\sigma^2}{n\epsilon^2}$$
$$\Longrightarrow \lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) = 1$$

特征函数

引子

- 大数定律的证明中，如果方差不存在，无法使用切比雪夫不等式怎么办?
- 特征函数是处理概率论问题的有力工具，其作用在于：
  1. 可将卷积运算化成乘法运算；
  2. 可将求各阶矩的积分运算化成微分运算；
  3. 可将求随机变量序列的极限分布化成一般的函数极限问题；
  4. 可方便处理串级随机变量，应用于核辐射探测。
- 与之类似的概念有矩母函数和生成函数，在此我们专注于特征函数。

定义 (特征函数)

设 X 是一个随机变量，则称 e^{itX} 的数学期望值，即

$$\varphi(t) = \mathrm{E}(e^{itX}), -\infty < t < \infty$$

为随机变量 X 的特征函数。

由于 $|e^{itX}| = 1$，随机变量 X 的特征函数总是存在。

引子

- 大数定律的证明中，如果方差不存在，无法使用切比雪夫不等式怎么办?
- 特征函数是处理概率论问题的有力工具，其作用在于:
  1. 可将卷积运算化成乘法运算;
  2. 可将求各阶矩的积分运算化成微分运算;
  3. 可将求随机变量序列的极限分布化成一般的函数极限问题;
  4. 可方便处理串级随机变量，应用于核辐射探测。
- 与之类似的概念有矩母函数和生成函数，在此我们专注于特征函数。

定义 (特征函数)

设 X 是一个随机变量，则称 $e^{\imath tX}$ 的数学期望值，即

$$\varphi(t) = \mathrm{E}(e^{\imath tX}), -\infty < t < \infty$$

为随机变量 X 的特征函数。

由于 $|e^{\imath tX}| = 1$，随机变量 X 的特征函数总是存在。

引子

- 大数定律的证明中，如果方差不存在，无法使用切比雪夫不等式怎么办？
- 特征函数是处理概率论问题的有力工具，其作用在于：
  1. 可将卷积运算化成乘法运算；
  2. 可将求各阶矩的积分运算化成微分运算；
  3. 可将求随机变量序列的极限分布化成一般的函数极限问题；
  4. 可方便处理串级随机变量，应用于核辐射探测。
- 与之类似的概念有矩母函数和生成函数，在此我们专注于特征函数。

定义 (特征函数)

设 X 是一个随机变量，则称 e^{itX} 的数学期望值，即

$$\varphi(t) = \mathrm{E}(e^{itX}), -\infty < t < \infty$$

为随机变量 X 的特征函数。

由于 $|e^{itX}| = 1$，随机变量 X 的特征函数总是存在。

引子

- 大数定律的证明中，如果方差不存在，无法使用切比雪夫不等式怎么办?
- 特征函数是处理概率论问题的有力工具，其作用在于：
  1. 可将卷积运算化成乘法运算；
  2. 可将求各阶矩的积分运算化成微分运算；
  3. 可将求随机变量序列的极限分布化成一般的函数极限问题；
  4. 可方便处理串级随机变量，应用于核辐射探测。
- 与之类似的概念有矩母函数和生成函数，在此我们专注于特征函数。

定义 (特征函数)

设 X 是一个随机变量，则称 $e^{\imath tX}$ 的数学期望值，即

$$\varphi(t) = \mathrm{E}(e^{\imath tX}), -\infty < t < \infty$$

为随机变量 X 的特征函数。

由于 $|e^{\imath tX}| = 1$，随机变量 X 的特征函数总是存在。

引子

- 大数定律的证明中，如果方差不存在，无法使用切比雪夫不等式怎么办?
- 特征函数是处理概率论问题的有力工具，其作用在于:
  1. 可将卷积运算化成乘法运算;
  2. 可将求各阶矩的积分运算化成微分运算;
  3. 可将求随机变量序列的极限分布化成一般的函数极限问题;
  4. 可方便处理串级随机变量，应用于核辐射探测。
- 与之类似的概念有矩母函数和生成函数，在此我们专注于特征函数。

定义 (特征函数)

设 X 是一个随机变量，则称 $e^{\imath tX}$ 的数学期望值，即

$$\varphi(t) = \mathrm{E}(e^{\imath tX}), -\infty < t < \infty$$

为随机变量 X 的特征函数。

由于 $|e^{\imath tX}| = 1$，随机变量 X 的特征函数总是存在。

计算

连续变量

$$\varphi(t) = \int_{-\infty}^{\infty} e^{\imath tX} f(x) dx$$

是概率密度函数 $f(x)$ 的傅立叶变换。

离散变量

$$\varphi(t) = \sum_{k=0}^{\infty} e^{\imath t x_k} p_k$$

特征函数也是连续函数。

计算

连续变量

$$\varphi(t) = \int_{-\infty}^{\infty} e^{\imath tX} f(x) dx$$

是概率密度函数 $f(x)$ 的傅立叶变换。

离散变量

$$\varphi(t) = \sum_{k=0}^{\infty} e^{\imath t x_k} p_k$$

特征函数也是连续函数。

性质

1. $|\varphi(t)| \leq \varphi(0) = 1$，
2. $\varphi(-t) = \overline{\varphi(t)}$ 复共轭；
3. $\varphi_{aX+b}(t) = e^{\imath bt}\varphi_X(at)$；
4. 若随机变量 X 与 Y 独立，则 $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$；
5. 若 $\mathrm{E}(X^l)$ 存在，则

$$\varphi^{(k)}(0) = \imath^k \mathrm{E}(X^k), 0 \leq k \leq l$$

   因此，可以通过特征函数的导数求随机变量的矩 $\mathrm{E}(X^k) = \frac{\varphi^{(k)}(0)}{\imath^k}$

6. 一致连续

$$\forall\epsilon\exists\delta\forall t(|\varphi(t+\delta)-\varphi(t)|<\epsilon) \text{ 而非 } \forall t\forall\epsilon\exists\delta(|\varphi(t+\delta)-\varphi(t)|<\epsilon)$$

7. 非负定性

$$\forall\vec{z},\vec{t}\left[\sum_{ij}\varphi(t_i-t_j)z_i\bar{z}_j \geq 0\right]$$

性质

1. $|\varphi(t)| \le \varphi(0) = 1$，
2. $\varphi(-t) = \overline{\varphi(t)}$ 复共轭；
3. $\varphi_{aX+b}(t) = e^{\imath bt}\varphi_X(at)$；
4. 若随机变量 X 与 Y 独立，则 $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$；
5. 若 $\mathrm{E}(X^l)$ 存在，则

$$\varphi^{(k)}(0) = \imath^k \mathrm{E}(X^k), 0 \le k \le l$$

因此，可以通过特征函数的导数求随机变量的矩 $\mathrm{E}(X^k) = \frac{\varphi^{(k)}(0)}{\imath^k}$

6. 一致连续

$$\forall\epsilon\exists\delta\forall t(|\varphi(t+\delta)-\varphi(t)|<\epsilon) \text{ 而非 } \forall t\forall\epsilon\exists\delta(|\varphi(t+\delta)-\varphi(t)|<\epsilon)$$

7. 非负定性

$$\forall\vec{z},\vec{t}\left[\sum_{ij}\varphi(t_i-t_j)z_i\bar{z}_j \ge 0\right]$$

性质

1. $|\varphi(t)| \le \varphi(0) = 1$，
2. $\varphi(-t) = \overline{\varphi(t)}$ 复共轭；
3. $\varphi_{aX+b}(t) = e^{\imath bt}\varphi_X(at)$；
4. 若随机变量 X 与 Y 独立，则 $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$；
5. 若 $\mathrm{E}(X^l)$ 存在，则

$$\varphi^{(k)}(0) = \imath^k \mathrm{E}(X^k), 0 \le k \le l$$

因此，可以通过特征函数的导数求随机变量的矩 $\mathrm{E}(X^k) = \frac{\varphi^{(k)}(0)}{\imath^k}$

6. 一致连续

$$\forall\epsilon\exists\delta\forall t(|\varphi(t+\delta)-\varphi(t)|<\epsilon) \text{ 而非 } \forall t\forall\epsilon\exists\delta(|\varphi(t+\delta)-\varphi(t)|<\epsilon)$$

7. 非负定性

$$\forall \vec{z}, \vec{t}\left[\sum_{ij}\varphi(t_i-t_j)z_i\bar{z}_j \ge 0\right]$$

性质

1. $|\varphi(t)| \le \varphi(0) = 1$，
2. $\varphi(-t) = \overline{\varphi(t)}$ 复共轭；
3. $\varphi_{aX+b}(t) = e^{\imath bt}\varphi_X(at)$；
4. 若随机变量 X 与 Y 独立，则 $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$；
5. 若 $\mathrm{E}(X^l)$ 存在，则

$$\varphi^{(k)}(0) = \imath^k \mathrm{E}(X^k), 0 \le k \le l$$

因此，可以通过特征函数的导数求随机变量的矩 $\mathrm{E}(X^k) = \frac{\varphi^{(k)}(0)}{\imath^k}$

6. 一致连续

$$\forall\epsilon\exists\delta\forall t(|\varphi(t+\delta)-\varphi(t)|<\epsilon) \text{ 而非 } \forall t\forall\epsilon\exists\delta(|\varphi(t+\delta)-\varphi(t)|<\epsilon)$$

7. 非负定性

$$\forall\vec{z},\vec{t}\left[\sum_{ij}\varphi(t_i-t_j)z_i\bar{z}_j \ge 0\right]$$

性质

1. $|\varphi(t)| \le \varphi(0) = 1$，
2. $\varphi(-t) = \overline{\varphi(t)}$ 复共轭；
3. $\varphi_{aX+b}(t) = e^{\imath bt}\varphi_X(at)$；
4. 若随机变量 X 与 Y 独立，则 $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$；
5. 若 $\mathrm{E}(X^l)$ 存在，则

$$\varphi^{(k)}(0) = \imath^k \mathrm{E}(X^k), 0 \le k \le l$$

因此，可以通过特征函数的导数求随机变量的矩 $\mathrm{E}(X^k) = \frac{\varphi^{(k)}(0)}{\imath^k}$

6. 一致连续

$$\forall\epsilon\exists\delta\forall t(|\varphi(t+\delta) - \varphi(t)| < \epsilon) \text{ 而非 } \forall t\forall\epsilon\exists\delta(|\varphi(t+\delta) - \varphi(t)| < \epsilon)$$

7. 非负定性

$$\forall\vec{z},\vec{t}\left[\sum_{ij}\varphi(t_i - t_j)z_i\bar{z}_j \ge 0\right]$$

性质

1. $|\varphi(t)| \leq \varphi(0) = 1$，
2. $\varphi(-t) = \overline{\varphi(t)}$ 复共轭；
3. $\varphi_{aX+b}(t) = e^{\imath bt}\varphi_X(at)$；
4. 若随机变量 X 与 Y 独立，则 $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$；
5. 若 $\mathrm{E}(X^l)$ 存在，则

$$\varphi^{(k)}(0) = \imath^k \mathrm{E}(X^k), 0 \leq k \leq l$$

因此，可以通过特征函数的导数求随机变量的矩 $\mathrm{E}(X^k) = \frac{\varphi^{(k)}(0)}{\imath^k}$

6. 一致连续

$$\forall\epsilon\exists\delta\forall t(|\varphi(t+\delta) - \varphi(t)| < \epsilon) \text{ 而非 } \forall t\forall\epsilon\exists\delta(|\varphi(t+\delta) - \varphi(t)| < \epsilon)$$

7. 非负定性

$$\forall\vec{z},\vec{t}\left[\sum_{ij}\varphi(t_i - t_j)z_i\bar{z}_j \geq 0\right]$$

唯一性

1. 逆转公式：设 $F(x)$ 和 $\varphi(t)$ 分别为随机变量 X 的分布函数和特征函数，则对 $F(x)$ 的任意两个连续点 $x_1 < x_2$，

$$F(x_2)-F(x_1) = \lim_{\tau\to\infty} \frac{1}{2\pi}\int_{-\tau}^{\tau} \frac{e^{-\imath t x_1}-e^{-\imath t x_2}}{\imath t}\varphi(t)dt$$

2. 若连续型随机变量 X 的密度函数为 $f(x)$，且 $\int_{-\infty}^{\infty}|\varphi(t)|dt < \infty$，则

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\imath t x}\varphi(t)dt$$

3. 随机变量的分布函数由其特征函数唯一确定。

概言之，傅立叶变换的可逆性和唯一性，保证了（存在特征函数的）随机变量与其特征函数一一对应。

唯一性

1. 逆转公式： 设 $F(x)$ 和 $\varphi(t)$ 分别为随机变量 X 的分布函数和特征函数，则对 $F(x)$ 的任意两个连续点 $x_1 < x_2$，

$$F(x_2) - F(x_1) = \lim_{\tau\to\infty} \frac{1}{2\pi} \int_{-\tau}^{\tau} \frac{e^{-\imath t x_1} - e^{-\imath t x_2}}{\imath t} \varphi(t) dt$$

2. 若连续型随机变量 X 的密度函数为 $f(x)$，且 $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$，则

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\imath t x} \varphi(t) dt$$

3. 随机变量的分布函数由其特征函数唯一确定 。

概言之，傅立叶变换的可逆性和唯一性，保证了（存在特征函数的）随机变量与其特征函数一一对应。

唯一性

❶ 逆转公式： 设 $F(x)$ 和 $\varphi(t)$ 分别为随机变量 X 的分布函数和特征函数，则对 $F(x)$ 的任意两个连续点 $x_1 < x_2$，

$$F(x_2) - F(x_1) = \lim_{\tau \to \infty} \frac{1}{2\pi} \int_{-\tau}^{\tau} \frac{e^{-\imath t x_1} - e^{-\imath t x_2}}{\imath t} \varphi(t) dt$$

❷ 若连续型随机变量 X 的密度函数为 $f(x)$，且 $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$，则

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\imath t x} \varphi(t) dt$$

❸ 随机变量的分布函数由其特征函数唯一确定 。

概言之，傅立叶变换的可逆性和唯一性，保证了（存在特征函数的）随机变量与其特征函数一一对应。

离散分布的特征函数

分布	概率分布$P(X=k)$	特征函数
$(0-1)$分布	$p^k(1-p)^{1-k}$ $k=0,1$	$1-p+pe^{it}$
$b(n,p)$	$C_n^k p^k(1-p)^{n-k}$ $k=0,1,2,\cdots,n$	$\left(1-p+pe^{it}\right)^n$
$\pi(\lambda)$	$\lambda^k e^{-\lambda}/k!$ $k=0,1,2,\cdots$	$e^{\lambda(e^{it}-1)}$
$Ge(p)$	$(1-p)^{k-1}p$ $k=0,1,\cdots$	$p/\left(1-qe^{it}\right)$

连续分布的特征函数

分布	概率密度$f(x)$	期望值
$U(a,b)$	$1/(b-a), \quad a<x<b$	$\dfrac{e^{ibt}-e^{iat}}{it(b-a)}$
$\mathrm{Exp}(\lambda)$	$\lambda e^{-\lambda x}, \quad x>0$	$(1-it/\lambda)^{-1}$
$N(\mu,\sigma^2)$	$\dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty<x<\infty$	$e^{i\mu t-\frac{\sigma^2 t^2}{2}}$
$Ga(\alpha,\lambda)$	$\dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x\geq 0$	$(1-it/\lambda)^{-\alpha}$
$\chi^2(n)$	$\dfrac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)} e^{-x/2}, \quad x>0$	$(1-2it)^{-n/2}$
$Cau(\mu,\lambda)$	$\dfrac{1}{\pi}\dfrac{1}{1+x^2}, \quad -\infty<x<\infty$	$e^{-\lvert t\rvert}$

例：卷王高斯

设随机变量 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$，且 X, Y 独立。其特征函数分别为

$$\varphi_X(t) = e^{\imath\mu_1 t - \frac{\sigma_1^2 t^2}{2}}, \varphi_Y(t) = e^{\imath\mu_2 t - \frac{\sigma_2^2 t^2}{2}}$$

$X + Y$ 的特征函数为

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t) = e^{\imath t(\mu_1+\mu_2) - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$$

这正是 $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ 的特征函数。由特征函数的唯一性

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

例：卷王高斯

设随机变量 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$，且 X, Y 独立。其特征函数分别为

$$\varphi_X(t) = e^{\imath\mu_1 t - \frac{\sigma_1^2 t^2}{2}}, \varphi_Y(t) = e^{\imath\mu_2 t - \frac{\sigma_2^2 t^2}{2}}$$

$X + Y$ 的特征函数为

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t) = e^{\imath t(\mu_1+\mu_2) - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$$

这正是 $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ 的特征函数。由特征函数的唯一性

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

例：卷王高斯

设随机变量 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$，且 X, Y 独立。其特征函数分别为

$$\varphi_X(t) = e^{\imath\mu_1 t - \frac{\sigma_1^2 t^2}{2}}, \varphi_Y(t) = e^{\imath\mu_2 t - \frac{\sigma_2^2 t^2}{2}}$$

$X + Y$ 的特征函数为

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t) = e^{\imath t(\mu_1+\mu_2) - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$$

这正是 $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ 的特征函数。由特征函数的唯一性

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

例：计算期望和方差

随机变量 $X \sim N(\mu, \sigma^2)$，其特征函数为 $\varphi(t) = e^{\imath\mu t - \frac{\sigma^2 t^2}{2}}$

$$\varphi'(t) = (\imath\mu - \sigma^2 t) e^{\imath\mu t - \frac{\sigma^2 t^2}{2}}$$

$$\varphi''(t) = [-\sigma^2 + (\imath\mu - \sigma^2 t)^2] e^{\imath\mu t - \frac{\sigma^2 t^2}{2}}$$

$$\Longrightarrow \quad \mathrm{E}(X) = \frac{\varphi'(0)}{\imath} = \frac{1}{\imath}(\imath\mu)e^0 = \mu$$

$$\mathrm{E}(X^2) = \frac{\varphi''(0)}{\imath^2} = \frac{1}{\imath^2}[-\sigma^2 - \mu^2]e^0 = \sigma^2 + \mu^2$$

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X) = \sigma^2$$

例：计算期望和方差

随机变量 $X \sim N(\mu, \sigma^2)$，其特征函数为 $\varphi(t) = e^{\imath\mu t - \frac{\sigma^2 t^2}{2}}$

$$\varphi'(t) = (\imath\mu - \sigma^2 t) e^{\imath\mu t - \frac{\sigma^2 t^2}{2}}$$

$$\varphi''(t) = [-\sigma^2 + (\imath\mu - \sigma^2 t)^2] e^{\imath\mu t - \frac{\sigma^2 t^2}{2}}$$

$$\Longrightarrow \mathrm{E}(X) = \frac{\varphi'(0)}{\imath} = \frac{1}{\imath}(\imath\mu)e^0 = \mu$$

$$\mathrm{E}(X^2) = \frac{\varphi''(0)}{\imath^2} = \frac{1}{\imath^2}[-\sigma^2 - \mu^2]e^0 = \sigma^2 + \mu^2$$

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X) = \sigma^2$$

例：计算期望和方差

随机变量 $X \sim N(\mu, \sigma^2)$，其特征函数为 $\varphi(t) = e^{\imath\mu t - \frac{\sigma^2 t^2}{2}}$

$$\varphi'(t) = (\imath\mu - \sigma^2 t) e^{\imath\mu t - \frac{\sigma^2 t^2}{2}}$$

$$\varphi''(t) = [-\sigma^2 + (\imath\mu - \sigma^2 t)^2] e^{\imath\mu t - \frac{\sigma^2 t^2}{2}}$$

$$\Longrightarrow \mathrm{E}(X) = \frac{\varphi'(0)}{\imath} = \frac{1}{\imath}(\imath\mu)e^0 = \mu$$

$$\mathrm{E}(X^2) = \frac{\varphi''(0)}{\imath^2} = \frac{1}{\imath^2}[-\sigma^2 - \mu^2]e^0 = \sigma^2 + \mu^2$$

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X) = \sigma^2$$

例：二项分布的极限为泊松分布

设随机变量 $X \sim b(n,p)$，其特征函数为

$$\phi(t) = (1-p + pe^{it})^n$$

令 $\lambda = pn$ 为常数，取 $n \to \infty$ 的极限

$$\phi(t) = (1-p + pe^{it})^n = \left[1 + \frac{\lambda}{n}(1 - e^{it})\right]^n \xrightarrow{n\to\infty} e^{\lambda(1-e^{it})}$$

正是泊松分布的特征函数。

例：二项分布的极限为泊松分布

设随机变量 $X \sim b(n,p)$，其特征函数为

$$\phi(t) = (1{-}p + pe^{\imath t})^n$$

令 $\lambda = pn$ 为常数，取 $n \to \infty$ 的极限

$$\phi(t) = (1{-}p + pe^{\imath t})^n = \left[1 + \frac{\lambda}{n}(1 - e^{\imath t})\right]^n \xrightarrow{n\to\infty} e^{\lambda(1-e^{\imath t})}$$

正是泊松分布的特征函数。

例：二项分布的极限为泊松分布

设随机变量 $X \sim b(n,p)$，其特征函数为

$$\phi(t) = (1-p + pe^{\imath t})^n$$

令 $\lambda = pn$ 为常数，取 $n \to \infty$ 的极限

$$\phi(t) = (1-p + pe^{\imath t})^n = \left[1 + \frac{\lambda}{n}(1 - e^{\imath t})\right]^n \xrightarrow{n\to\infty} e^{\lambda(1-e^{\imath t})}$$

正是泊松分布的特征函数。

例：标准正态分布随机变量的平方和

设 $Y_i \sim N(0,1)$ 那么

$$\varphi_Y(t) = e^{-\frac{t^2}{2}}$$

$Z_i = Y_i^2$ 的特征函数为

$$\begin{aligned}\phi_{Z_i}(t) &= \mathrm{E}(e^{\imath t Y_i^2}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\imath t y^2} e^{-\frac{y^2}{2}}\, dy \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(1/2-\imath t)y^2}\, dy = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\pi}{1/2-\imath t}} = (1-2\imath t)^{-\frac{1}{2}}\end{aligned}$$

正是 $\chi^2(1)$ 的特征函数。

例：标准正态分布随机变量的平方和

设 $Y_i \sim N(0,1)$ 那么

$$\varphi_Y(t) = e^{-\frac{t^2}{2}}$$

$Z_i = Y_i^2$ 的特征函数为

$$\begin{aligned}\phi_{Z_i}(t) = \mathrm{E}(e^{\imath t Y_i^2}) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\imath t y^2} e^{-\frac{y^2}{2}} dy \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(1/2-\imath t)y^2} dy = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{\pi}{1/2-\imath t}} = (1-2\imath t)^{-\frac{1}{2}}\end{aligned}$$

正是 $\chi^2(1)$ 的特征函数。

例：卡方分布的和

$Z = \sum_i Z_i$ 的特征函数为

$$\varphi_Z(t) = (1-2\imath t)^{-\frac{n}{2}}$$

正是 $\chi^2(n)$ 的特征函数。

- 故而，对独立同分布的 $Y_i \sim N(0,1)$

$$\sum_{i=1}^{n} Y_i^2 \sim \chi^2(n)$$

辛钦大数定律的证明

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{P} \mu$$

- 右侧的特征函数是 $e^{\imath t\mu}$
- 设 X_i 的特征函数是 $\varphi(t)$，左侧的特征函数

$$\begin{aligned}
&\left[\varphi\left(\frac{t}{n}\right)\right]^n \\
=&\left[1+\varphi'(0)\frac{t}{n}+o\left(\frac{t}{n}\right)\right]^n \\
=&\left[1+\imath\mu\frac{t}{n}+o\left(\frac{t}{n}\right)\right]^n \\
=&e^{\imath\mu t}, n\to\infty
\end{aligned}$$

- 由特征函数的唯一性，得证

总结

大数定律的意义

1. 具有相同数学期望和方差的独立随机变量序列的算术平均值依概率收敛于数学期望。
2. 当 n 足够大时，算术平均值几乎是一常数。 数学期望 可被 算术均值 近似代替。
3. 伯努利大数定律是辛钦大数定律的特例。