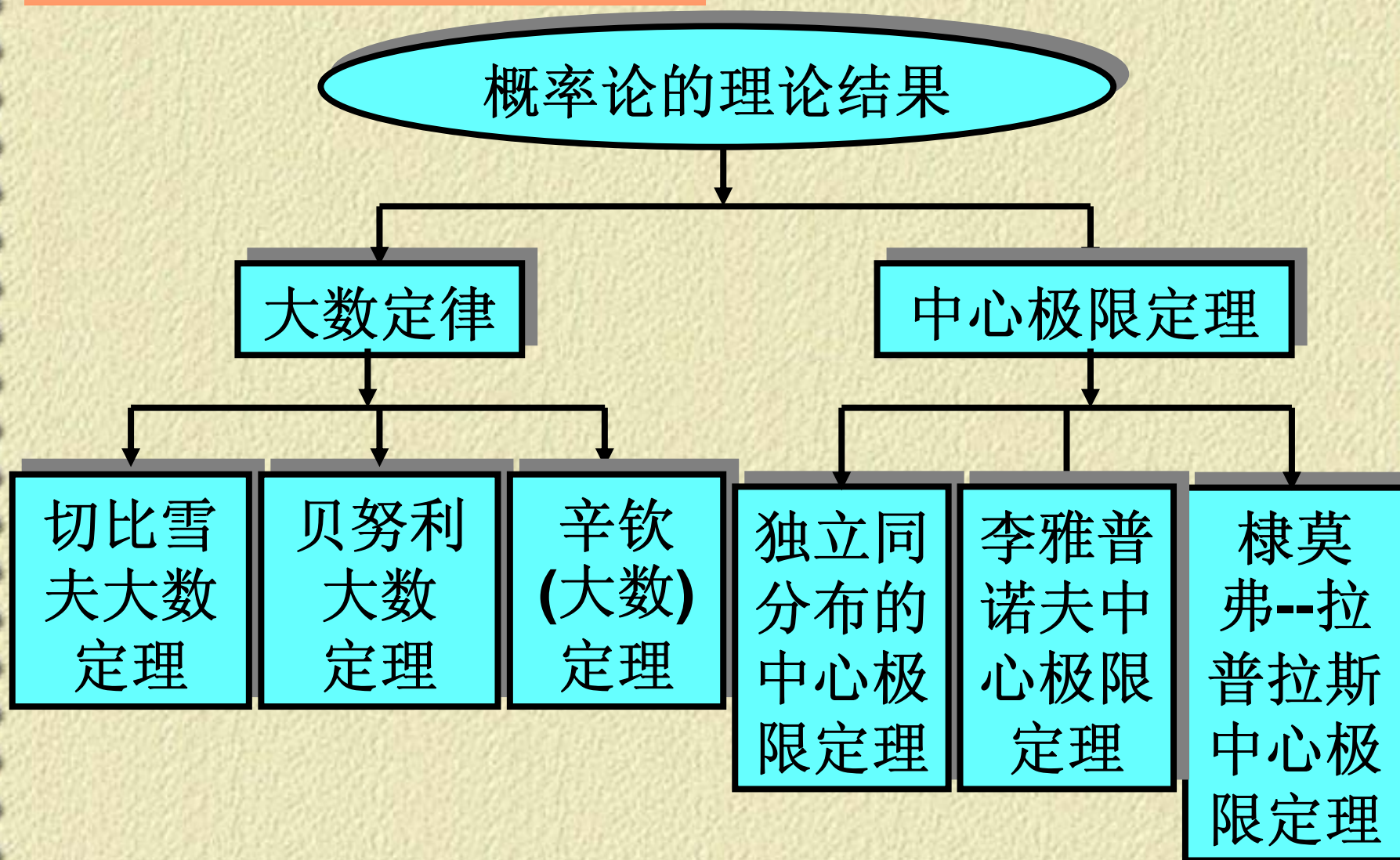


第五章知识结构图



第五章 大数定律及中心极限定理

大数定律和中心极限定理是概率论中的比较深入与重要的理论结果。本章主要讨论如下两个问题：

1. 在一定条件下, 一系列随机变量的算术平均值（按某种意义）收敛于所希望的平均值的定理称为“**大数定律**”。

——“概率是频率的稳定值”提供了理论依据，它以严格的数学形式表达了随机现象最根本的性质之一：**平均结果的稳定性**。它是随机现象统计规律的具体表现，也成为数理统计的理论基础。

2. 在一定条件下，大量的相互独立的随机变量之和的概率分布近似于正态分布的定理称为“中心极限定理”。

中心极限定理它以严格的数学形式证明了随机现象另一个最根本的性质之一：如果一个量是由大量相互独立的随机因素的影响所造成，而每一个别因素在总影响中所起的作用不大，则这种量一般都服从或近似服从正态分布。它是随机现象统计规律的具体表现，也成为数理统计的理论基础。而正态分布又是应用最广的分布。

第一节 大数定律

大数定律的客观背景

大量的随机现象中平均结果的稳定性:



大量抛掷硬币
正面出现频率



生产过程中的
废品率



字母使用频率

.....

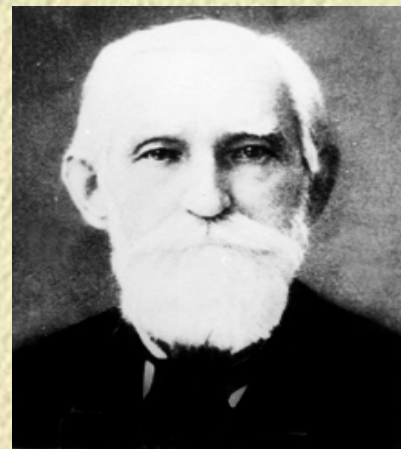
一. 切比雪夫大数定律

定理1（切比雪夫大数定律）

设 X_1, X_2, \dots 是相互独立的随机变量序列，它们都有有限的方差，并且方差有共同的上界，即 $D(X_i) \leq K, i=1, 2, \dots$ ，则对任意的 $\varepsilon > 0$ ：

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i) \right| < \varepsilon \right\} = 1$$

注：切比雪夫大数定律表明：独立随机变量序列 $\{X_n\}$ ，如果方差有共同的上界，则：



切比雪夫, П. Л.

切比雪夫

算术平均值 $\frac{1}{n} \sum_{i=1}^n X_i$ 与其数学期望 $\frac{1}{n} \sum_{i=1}^n E(X_i)$ 的偏差

是很小的，其概率接近于1。即当 n 充分大时，

$\frac{1}{n} \sum_{i=1}^n X_i$ 取值接近于其数学期望的概率接近于 1

切比雪夫大数定律给出了平均值稳定性的科学描述

定理2（切比雪夫大数定律的特殊情况）

亦称为独立同分布的大数定律

设 X_1, X_2, \dots 是相互独立的随机变量序列，它们具有相同的数学期望与方差： $E(X_k) = \mu$ ， $D(X_k) = \sigma^2$ $k = 1, 2, \dots$ ，作前 n 个随机变量的算术平均值：

$\overline{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ ，则对任意的 $\varepsilon > 0$ 有：

$$\lim_{n \rightarrow \infty} P \left\{ \left| \overline{X}_n - \mu \right| < \varepsilon \right\} = \lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| < \varepsilon \right\} = 1$$

证明： $\because E\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n E(X_k) = \frac{1}{n} \cdot n\mu = \mu$

$$D\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} \sum_{k=1}^n D(X_k) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{1}{n} \sigma^2$$

由切比雪夫不等式可得：

$$(P(X - E(X) \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2})$$

$$P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| < \varepsilon \right\} \geq 1 - \frac{\sigma^2/n}{\varepsilon^2}$$

令： $n \rightarrow \infty$ ，并注意到概率 ≤ 1 ，所以得：

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| < \varepsilon \right\} = 1$$

注：▲ 定理2 是定理1中当数学期望和方差给定具体值时的特殊情况。它表明当 n 很大时，随机变量 X_1, X_2, \dots, X_n 的算术平均值在概率意义下接近于数学期望 $E(X_k) = \mu$

其作用：在数理统计中如不知道 μ ，则可用其算术平均值来近似代替，则定理2就提供了理论依据。

▲ 称序列 $Y_1, Y_2, \dots, Y_n, \dots$ 依概率收敛于常数 a

指的是：对任意正数 ε ， $\lim_{n \rightarrow \infty} P(|Y_n - a| < \varepsilon) = 1$

记为： $Y_n \xrightarrow{P} a$

由此，定理2 的结论可叙述为：序列 $\overline{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$
依概率收敛于常数 μ

▲ 依概率收敛的序列具有如下性质：

设 $X_n \xrightarrow{P} a$ ， $Y_n \xrightarrow{P} b$ ，又设函数 $g(x, y)$ 在点 (a, b) 处连续，则有：

$$g(X_n, Y_n) \xrightarrow{P} g(a, b)$$

[证明]：（略）见教材P 146 页下方的小字体

二. 贝努利大数定理

定理3. (贝努利定理)

设随机变量 $X_1, X_2, \dots, X_n, \dots$
相互独立, 且同时服从以 p 为参数
的(0-1)分布。则对任意正数 ε 有:

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{n_A}{n} - p \right| < \varepsilon \right\} = 1$$

或

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{n_A}{n} - p \right| \geq \varepsilon \right\} = 0$$



雅各布第一·伯努利

贝努利

其中: $n_A = X_1 + X_2 + \cdots X_n$

p 是事件 A 在每次试验中发生的概率。

n 次独立
重复试验
中事件 A
发生的次
数

证明: $\because X_k$ 服从 $(0-1)$ 分布

$$\therefore E(X_k) = p \quad k = 1, 2, \dots$$

令:
$$\frac{\sum_{k=1}^n X_k}{n} = \frac{n_A}{n}$$

则由定理2可得:

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{n_A}{n} - \mu \right| < \varepsilon \right\} = \lim_{n \rightarrow \infty} P \left\{ \left| \frac{n_A}{n} - p \right| < \varepsilon \right\} = 1$$

- 注: ▲ 定理3 表明: 当 n 很大时, 事件 A 发生的频率 $\frac{n_A}{n}$ 接近于事件 A 发生的概率 P , 即证明了频率的稳定性。从而, 当 n (试验次数) 很大时可以用事件发生的频率来近似代替事件的概率。
- ▲ 称事件 A 发生的频率 依概率收敛于事件 A 的概率 P 。
- ▲ 贝努里大数定律 提供了通过试验来确定事件概率的方法。

问题 定理2, 定理3 均为独立同分布的大数定律, 并且均要求随机变量的期望与方差都存在。如果 没有 随机变量的 方差存在 的条件, 那么大数定律的结论是否仍成立?

三. 辛钦定理

定理4. (辛钦定理)

设 $X_1, X_2, \dots, X_n, \dots$ 相互独立, 并且服从同一分布, 且具有数学期望:

$$E(X_k) = \mu \quad (k = 1, 2, \dots)$$

则对任意的正数 ε 有:

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| < \varepsilon \right\} = 1$$



辛钦

注：辛钦大数定律为寻找随机变量的期望值提供了一条实际可行的途径。

总的：

大数定律从各个角度描述了样本的算术平均值及频率的稳定性。

也为人们习惯上经常采用的用样本的算术平均值去代替或估计其平均值；用频率去代替或估计其“概率”提供了理论上的依据。

