第五章
大数定律及中心极限定理

王笑尘

北京邮电大学网络空间安全学院

wxiaochen@bupt.edu.cn

第五章 知识结构图

- 概率论的理论结果
  - 大数定律
    - 切比雪夫大数定律
    - 伯努利大数定律
    - 辛钦(大数)定律
  - 中心极限定理
    - 独立同分布的中心极限定理
    - 李雅普诺夫中心极限定理
    - 棣莫弗—拉普拉斯中心极限定理

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比较深入与重要的理论结果。本章主要讨论如下两个问题：

- 在一定条件下,一列随机变量的算术平均值（按某种意义）收敛于这些项的均值的定理称为“大数定律”。
- 大数定律为概率论存在的基础 —“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。

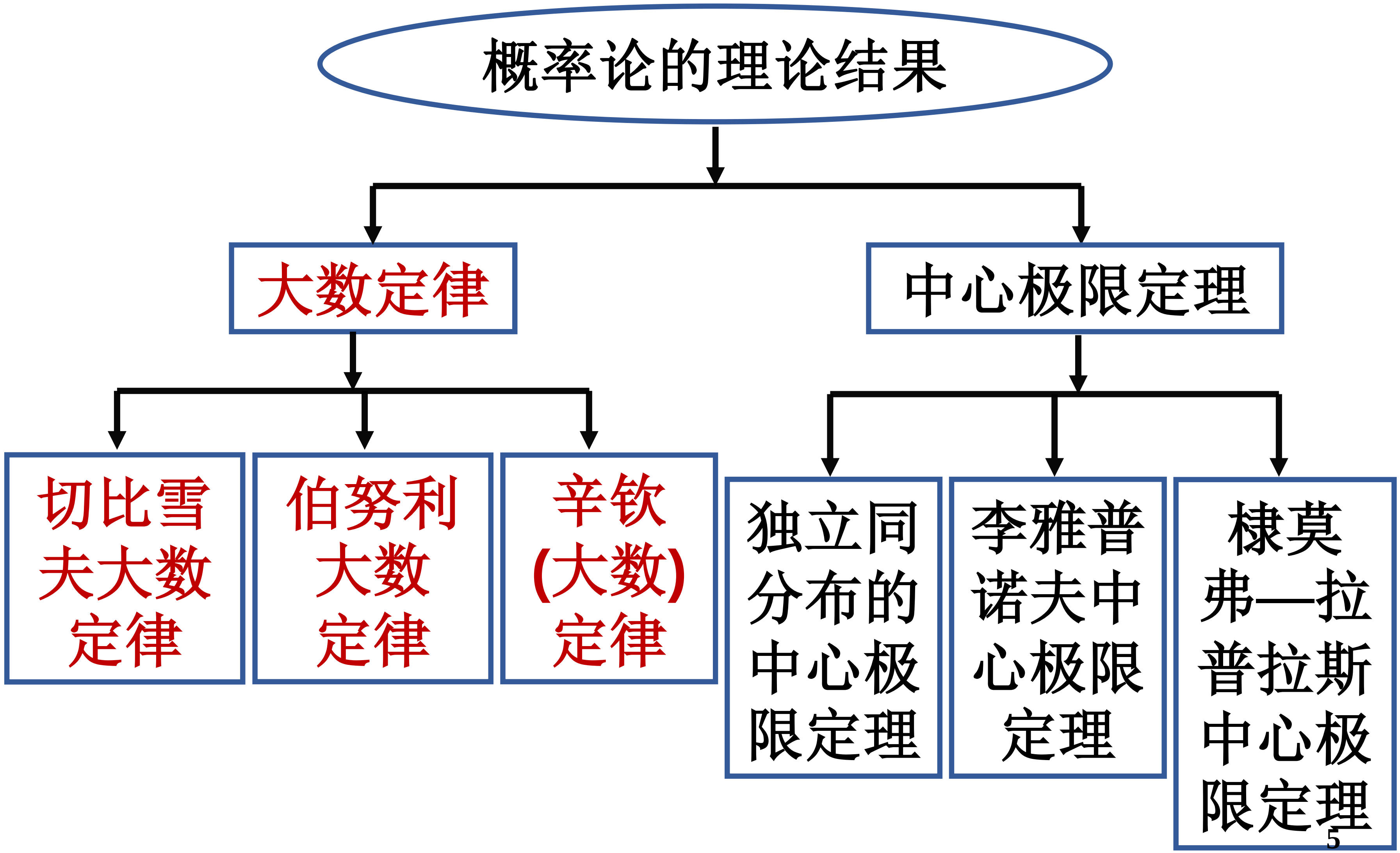
第五章 大数定律及中心极限定理

- 在一定条件下，大量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



概率论的理论结果
大数定律
中心极限定理
切比雪夫大数定律
伯努利大数定律
辛钦(大数)定律
独立同分布的中心极限定理
李雅普诺夫中心极限定理
棣莫弗—拉普拉斯中心极限定理

第一节 大数定律

大数定律的客观背景

大量随机现象中**平均结果的稳定性**：

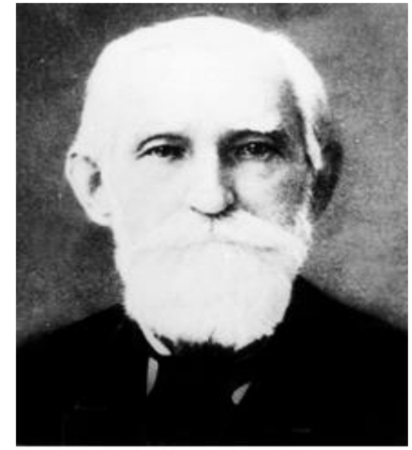
大量抛掷硬币正面出现频率

生产过程中的废品率

字母使用频率

……

一. 切比雪夫大数定律



切比雪夫，П. Л.

切比雪夫

定理1（切比雪夫大数定律）

设 $X_1, X_2, \ldots$ 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 $D(X_i) \le K$，　$i=1, 2, \ldots$，则对任意的$\varepsilon > 0$，

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

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：算术平均值收敛于对应期望值的概率接近于1。

- 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，两者近似的概率接近于1。
- 即当 n 充分大时，$\frac{1}{n}\sum_{i=1}^{n}X_i$ 差不多不再是随机的了，其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**）

设 $X_1, X_2, \ldots$ 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**：$E(X_k)=\mu$， $D(X_k)=\sigma^2$

$k=1,2,\cdots$，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 $k=1,2,\cdots$，有：

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

证明:

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

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

由切比雪夫不等式可得: $P(|X-\mu|<\varepsilon)\geq 1-\frac{\sigma^2}{\varepsilon^2}$

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

令: $n\to\infty$, 并注意到概率 ≤ 1, 所以得:

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

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

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 $X_1, X_2, \cdots X_n$ 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 $Y_1, Y_2, \cdots Y_n, \cdots$ 依概率收敛于常数 a

指的是：对任意正数 ε 有 $\lim\limits_{n\to\infty} P(|Y_n - a| < \varepsilon) = 1$

记为：$Y_n \xrightarrow{P} a$

▲ 由此，定理2的结论可叙述为：序列 $\overline{X_n} = \dfrac{1}{n}\sum\limits_{k=1}^{n} X_k$ 依概率收敛于常数 μ

定理2

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

▲ 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：

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

二. 伯努利大数定律

定理3. (伯努利定理)

设随机变量 $X_1, X_2, \cdots X_n, \cdots\cdots$ 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：

雅各布第一・伯努利

伯努利

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1 \quad \text{或} \quad \lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right\} = 0$$

其中：$n_A = X_1 + X_2 + \cdots X_n$

n 次独立重复试验中事件A发生的次数

p是事件 A 在每次试验中发生的概率。

证明: $\because \quad X_k$ 服从 $(0-1)$ 分布

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

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

则由定理2（算术平均值接近其数学期望）可得:

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

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。

▲ 伯努利大数定律提供了通过试验来确定事件概率的方法。

总结:

- 定理1（要求独立、方差有限）、定理2（定理1的特殊情况—独立同分布，方差、期望都存在）给出了n很大时，算数平均值接近其数学期望的结论;
- 定理3（独立同分布，方差、期望都存在），给出了n很大时，频率接近其概率的结论。

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. 辛钦大数定律

定理4. （辛钦大数定理）

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


辛钦

$$E(X_k) = \mu \qquad (k = 1, 2\cdots)$$

则对任意的正数 ε 有:

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

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。

总结：

- 相互独立的随机变量，期望存在的条件下，若方差有限（定理1），或同分布（定理4），则当n很大时，算数平均值接近于数学期望。
- 独立同分布的(0-1)随机变量，方差与期望都存在（定理3），当n很大时，频率接近于概率。
- 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 。
- 也为用样本的算术平均值去代替或估计其“数学期望”；用频率去代替或估计其“概率”提供了理论上的依据。

第二节 中心极限定理

问题的引出

- 自从高斯指出测量误差服从正态分布之后，人们发现，正态分布在自然界中极为常见。

高斯

- 大量实验观察也表明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。

研究的问题：

研究独立随机变量之和所特有的规律性问题。当 n 无限增大时，在什么条件下这个和的极限分布会是正态的呢？

在概率论中，习惯把和的分布收敛于正态分布这一类定理都叫做中心极限定理。故：

在一定条件下，大量的随机变量之和的概率分布以正态分布为极限的定理称为中心极限定理。

- 对于何种条件下服从正态分布这一个重要问题，在长达两个世纪内一直成为概率论研究的中心问题。最终，数学家们建立了一系列定理，解决了这一问题:

  - 具有有限数学期望和方差的一列独立同分布的随机变量的和经过标准化后是以标准正态分布为极限的，这就是独立同分布的中心极限定理，或称为林德贝尔格—勒维中心极限定理。

  - 当同分布为二项分布时就得出该定理的特例，即为：棣莫弗—拉普拉斯定理，它也是二项分布的正态近似。

- 对“由大量微小的、独立的随机因素”（不要求同分布）引起并累积成的变量，当随机因素个数趋于无穷时以正态分布为极限。这就是李雅普诺夫中心极限定理。

比如：一台机床已经调试良好，操作正常。但由于机床的微小震动、工具的微小变形、原材料质量上的微小差异、工作操作上的微小偏差等等数不清的随机因素，它们每一个因素在总的影响中所起的作用都是微小的。

而综合起来在产品质量上就形成一定的误差，这误差近似服从正态分布。

一. 独立同分布的中心极限定理

（林德贝尔格—勒维(Levy－Lindberg)定理）

定理1. 设随机变量 $X_1, X_2 \cdots X_n \cdots$ 相互独立且服从同一分布，其数学期望与方差：$E(X_k)=\mu$ $D(X_k)=\sigma^2$， $(k=1,2\cdots)$ 则随机变量之和 $\sum_{k=1}^{n} X_k$ 的标准化变量：

$$Y_n=\frac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}}=\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n\sigma^2}}=\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\,\sigma}$$

的分布函数 $F_n(x)$ 对于任意 x 满足:

近似地服从 N(0,1)

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{\boxed{\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\,\sigma}} \le x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

注: ➢ 定理1 表明，当 n 充分大时，n 个具有期望和方差的独立同分布的随机变量之和的标准化变量近似服从标准正态分布。

➢ 虽然在一般情况下，很难求出 $X_1+X_2+\dots+X_n$ 的分布的确切形式，但当 n 很大时，可以求出其近似分布。

▲ 定理1 表达了正态分布在概率论中的特殊地位：

尽管 $X_1, X_2 \cdots X_n \cdots$ 分布是任意的，但只要 n 充分大后，其样本平均值 $\frac{1}{n}\sum_{k=1}^{n} X_k$ 的分布却是近似服从正态分布的：

$$\because \quad Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\,\sigma} = \frac{\frac{1}{n}\sum_{k=1}^{n} X_k - \mu}{\frac{1}{n}\sqrt{n}\,\sigma} = \frac{\frac{1}{n}\sum_{k=1}^{n} X_k - E(X)}{\sqrt{\frac{1}{n}}\sqrt{D(X)}}$$

即 $\frac{1}{n}\sum_{k=1}^{n} X_k$ 服从正态分布 $N[E(X), (\sqrt{\frac{1}{n}}\sqrt{D(X)})^2]$

或 $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

这一结果是数理统计中大样本统计推断的基础

二. 李雅普诺夫定理

(Lyapunov 中心极限定理)

定理2. 设随机变量 $X_1, X_2 \cdots X_n \cdots$ 相互独立，它们具有数学期望和方差为：

$$E(X_k) = \mu_k\ ,\ \ D(X_k) = \sigma_k^2 \neq 0\ ,\ \ k = 1, 2 \cdots$$

记 $B_n^2 = \sum_{k=1}^{n} \sigma_k^2$, 若存在正数 δ, 使得当 $n \to \infty$

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E\left\{ \left| X_k - \mu_k \right|^{2+\delta} \right\} \to 0\ ,$$

则随机变量之和 $\sum_{k=1}^{n} X_k$ 的标准化变量：

$$Z_n = \frac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}} = \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n}$$

的分布函数 $F_n(x)$ 对于任意 x 满足：

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{ \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n} \le x \right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

注: ▲ 定理2表明, 当 n 充分大时，随机变量:

$$Z_n = \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n}$$ 近似服从标准正态分布。

即, $\sum_{k=1}^{n} X_k = B_n Z_n + \sum_{k=1}^{n} \mu_k$ 近似服从正态分布

$$N \sim (\sum_{k=1}^{n} \mu_k,\ B_n^2)$$

▲ 由此, 定理2再次表达了正态分布在概率论中的特殊地位: 无论各个随机变量 X_k 服从什么分布，只要满足定理2的条件，那么当n 充分大时，它们的和就近似服从正态分布。

三. 棣莫弗---拉普拉斯定理

（De Moivere—laplace 中心极限定理）

定理3. 设随机变量 $\eta_1,\eta_2\cdots\eta_n\cdots$ 相互独立，η_n服从参数为 $n,\ p\ (0<p<1)$ 的二项分布，则对任意 x 恒有:

$$\eta_1,\eta_2\cdots\eta_n\cdots=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt$$

证明: η_n 可分解为 n 个相互独立，服从同一 (0-1) 分布的 $X_1, X_2 \cdots X_n$ 之和。即： $\eta_n = \sum_{k=1}^{n} X_k$

其中 X_k $(k = 1, 2 \cdots n)$ 的分布律为：

$$P(X_k = i) = p^i (1-p)^{1-i} \qquad i = 0, 1$$

$$\because \quad E(X_k) = p, \ D(X_k) = p(1-p) \quad (k = 1, 2 \cdots n)$$

$\therefore$ 由定理1得：

$$\lim_{n\to\infty} P\left\{ \frac{\eta_n - np}{\sqrt{np(1-p)}} \le x \right\} = \lim_{n\to\infty} P\left\{ \frac{\sum_{k=1}^{n} X_k - np}{\sqrt{np(1-p)}} \le x \right\}$$

$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

注：▲ 定理3表明，正态分布是二项分布的极限分布，当 n 充分大时可以用正态分布来计算二项分布的概率。

▲ 在第二章中已介绍当 $n\to\infty$ 时，二项分布以泊松分布为极限分布；而在本章中二项分布又以正态分布为极限分布。这两者的区别是：

注：▲ 定理3表明，正态分布是二项分布的极限分布，当 n 充分大时可以用正态分布来计算二项分布的概率。

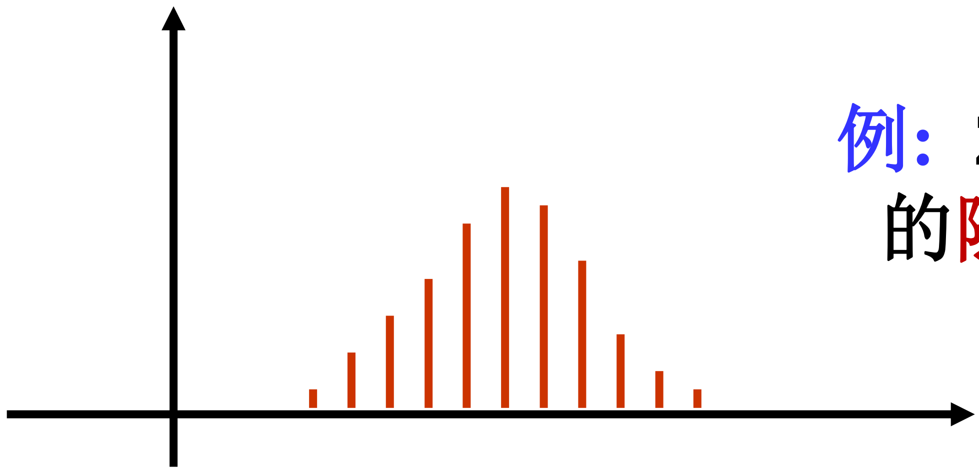
▲ 在第二章中已介绍当 $n \to \infty$ 时，二项分布以泊松分布为极限分布；而在本章中二项分布又以正态分布为极限分布。这两者的区别是：

在泊松定理中要求 $np \to \lambda$ (λ为常数)

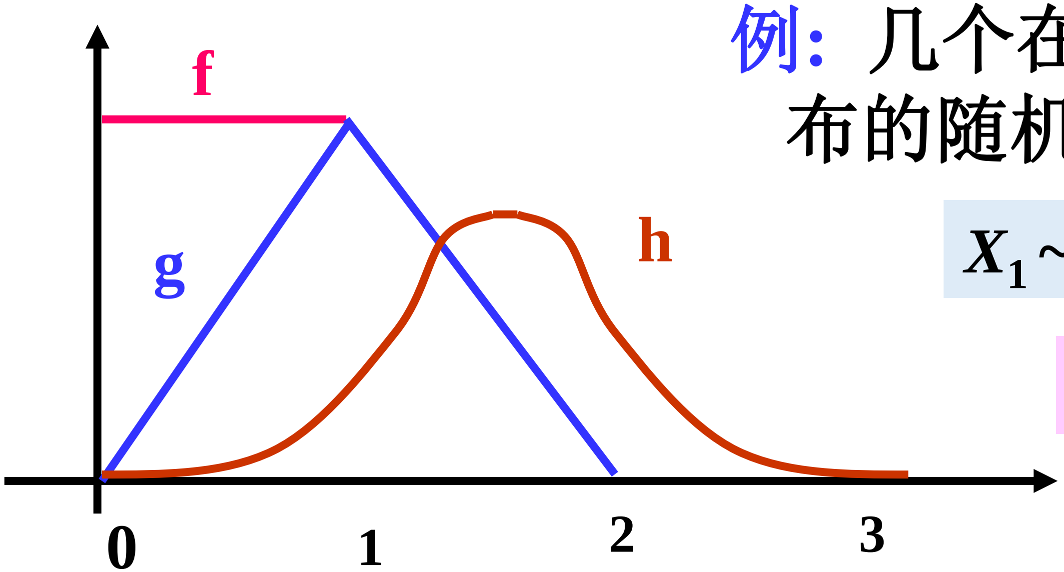
在中心极限定理中要求 $np \to \infty$

▲ 所以在实际计算中，如果 n 很大但 np或 nq 不大 (即 p 很小或 q =1-p 很小)，那么应该用泊松定理去近似；如果 n，np 或 nq 都较大，那么应该用中心极限定理去近似。

▲ 中心极限定理的直观图示



例：20个服从（0-1）分布的随机变量的和的分布



例：几个在(0, 1)上服从均匀分布的随机变量的和的分布。

$X_1 \sim f(x)$

$X_1 + X_2 \sim g(x)$

$X_1 + X_2 + X_3 \sim h(x)$

例1. 抽样检查产品质量时，如果发现**次品多于10个**则认为这批产品不能接受。

求： 应该检查多少个产品，可使次品率为 10% 的一批次品不能接受的概率达到 0. 9?

例1. 抽样检查产品质量时，如果发现次品多于10个则认为这批产品不能接受。

求： 应该检查多少个产品，可使次品率为 10% 的一批次品不能接受的概率达到 0. 9?

解： 设应检查产品个数为 n，其中次品数为 X，则

$$X \sim B(n, 0.1),\quad np = 0.1n,\ \sqrt{npq} = 0.3\sqrt{n}$$

现要求 n，使得：$P(10 \le X \le n) = 0.9$

$$\because\quad P(10 \le X \le n)$$

由定理3

近似服从N(0, 1)

$$= P\left(\frac{10-0.1n}{0.3\sqrt{n}} \le \frac{X-0.1n}{0.3\sqrt{n}} \le \frac{n-0.1n}{0.3\sqrt{n}}\right)$$

$$= P\left(\frac{10-0.1n}{0.3\sqrt{n}} \le \frac{X-0.1n}{0.3\sqrt{n}} \le 3\sqrt{n}\right)$$

由3σ准则，$\Phi(3\sqrt{n})$为 1

$$\approx \Phi(3\sqrt{n}) - \Phi\left(\frac{10-0.1n}{0.3\sqrt{n}}\right) \approx 1 - \Phi\left(\frac{10-0.1n}{0.3\sqrt{n}}\right)$$

$\therefore$ 要 $P(10 \le X \le n) = 0.9$

只要： $1-\Phi\left(\dfrac{10-0.1n}{0.3\sqrt{n}}\right) = 0.9$

即要： $\Phi\left(\dfrac{10-0.1n}{0.3\sqrt{n}}\right) = 0.1$

此时由于：$\Phi\left(\dfrac{10-0.1n}{0.3\sqrt{n}}\right) < 0.5$

必定有： $\dfrac{10-0.1n}{0.3\sqrt{n}}<0$

所以要： $\Phi\left(\dfrac{10-0.1n}{0.3\sqrt{n}}\right)=0.1$

因为 $\Phi(-x)=1-\Phi(x)$

只要：$1-\Phi\left(\dfrac{0.1n-10}{0.3\sqrt{n}}\right)=0.1$

即 $\Phi\left(\dfrac{0.1n-10}{0.3\sqrt{n}}\right)=0.9$ 查表得： $\Phi(1.28)\approx 0.9$

故 $\dfrac{0.1n-10}{0.3\sqrt{n}}=1.28 \Rightarrow n\approx 146$

结论：应检查 146 个产品时，可使这批产品不被接受的概率为0. 9

例 2. 计算机进行加法计算时，把每个加数取为最接近它的整数来计算。设所有的取数误差是相互独立的随机变量，并且都在区间[-0.5, 0.5] 上服从均匀分布。

求: (1) 现有1200个数相加，误差总和的绝对值小于10的概率。

(2) 应有多少个数相加时可使误差总和的绝对值小于10 的概率大于0. 9。

解: 设 $X_1, X_2, \cdots X_n$ 为各个加数的取数误差

则这是一列独立同分布的随机变量，其所有加数的误差总和为: $\sum_{k=1}^{n} X_k$

独立同分布中心极限定理

$\because\ X_k\quad (k=1,2,\cdots n)$ 在 $[-0.5,\ 0.5]$ 服从均匀分布

$$\therefore\ E(X_k)=\frac{-0.5+0.5}{2}=0,\ D(X_k)=\frac{[0.5-(-0.5)]^2}{12}=\frac{1}{12}$$

从而: $$E(\sum_{k=1}^{n}X_k)=\sum_{k=1}^{n}E(X_k)=0,$$

$$\sqrt{D(\sum_{k=1}^{n}X_k)}=\sqrt{\sum_{k=1}^{n}D(X_k)}=\sqrt{\frac{n}{12}}$$

(1). 这里: $$\sqrt{\sum_{k=1}^{n}D(X_k)}=\sqrt{\frac{1200}{12}}=10$$

$$\because\quad P\left(\left|\sum_{k=1}^{1200}X_k\right|<10\right)\quad =P\left(-10<\sum_{k=1}^{1200}X_k<10\right)$$

由定理1

$$= P\left(\frac{-10-0}{10} < \frac{\sum_{k=1}^{n} X_k - 0}{10} < \frac{10-0}{10} \right)$$

近似服从N (0, 1)

$$= P\left(-1 < \frac{\sum_{k=1}^{n} X_k}{10} < 1 \right) \approx \Phi(1) - \Phi(-1)$$

$$= 2\Phi(1) - 1 \quad = 2 \times 0.8413 - 1 = 0.6826$$

(2). $\because \; P\left(\left| \sum_{k=1}^{n} X_k \right| < 10 \right)$

$$= P\left(\frac{-10-0}{\sqrt{n/12}} < \frac{\sum_{k=1}^{n} X_k - 0}{\sqrt{n/12}} < \frac{10-0}{\sqrt{n/12}} \right)$$

$$= P\left(-20\sqrt{\frac{3}{n}} < \frac{\sum_{k=1}^{n} X_k}{\sqrt{\frac{n}{2}}} < 20\sqrt{\frac{3}{n}}\right) = 2\Phi\left(20\sqrt{\frac{3}{n}}\right) - 1$$

所以要 $P\left(\left|\sum_{k=1}^{n} X_k\right| < 10\right) > 0.9,$ 只要:

$$2\Phi\left(20\sqrt{\frac{3}{n}}\right) - 1 > 0.9 \implies \Phi\left(20\sqrt{\frac{3}{n}}\right) > 0.95$$

查表得: $\Phi(1.65) = 0.9505 > 0.95$

$\therefore \quad 20\sqrt{\frac{3}{n}} = 1.65$ 解得: $n \approx 441$

结论: 441 个数相加时可使误差总和的绝对值小于10 的概率大于0. 9。

例3. 在人寿保险公司里，有16000名同一年龄的人参加人寿保险。一年里这些人的死亡率为0.1%；参加保险的人在一年的第一天交付保险费3元，死亡时家属可以从保险公司领取2000元。

求: (1). 保险公司因开展这项业务获利不少于10000元的概率。

(2). 保险公司因开展这项业务亏本的概率。

解: 由题意，死亡人数 $X \sim B(16000,\ 0.001)$

这里，$np = 16000 \times 0.001 = 16$

$$\sqrt{npq} \approx \sqrt{16} = 4$$

棣莫弗-拉普拉斯中心极限定理

保险公司一年内这项保险收入是：$3\times 16000=48000$元

(1). 获利不少于10000元，即赔偿不大于 38000(元)，

即一 年内至多有 $\dfrac{38000}{2000}=19$（人）死亡

所以：
$$P(X\le 19)=P\left(\frac{X-16}{4}\le\frac{19-16}{4}\right)$$
$$\approx\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0.75}e^{-\frac{t^2}{2}}dt$$
$$=\Phi(0.75)=0.7734$$

即该公司获利不少于 10000(元)的概率为 0.7734。

(2). 公司亏本即赔款大于48000元，即一年内有多于 $\frac{48000}{2000}=24$（人）死亡

$$\therefore P(X>24)=1-P(X\le 24)$$

$$=1-P\left(\frac{X-16}{4}\le\frac{24-16}{4}\right)$$

$$=1-P\left(\frac{X-16}{4}\le 2\right)\approx 1-\Phi(2)$$

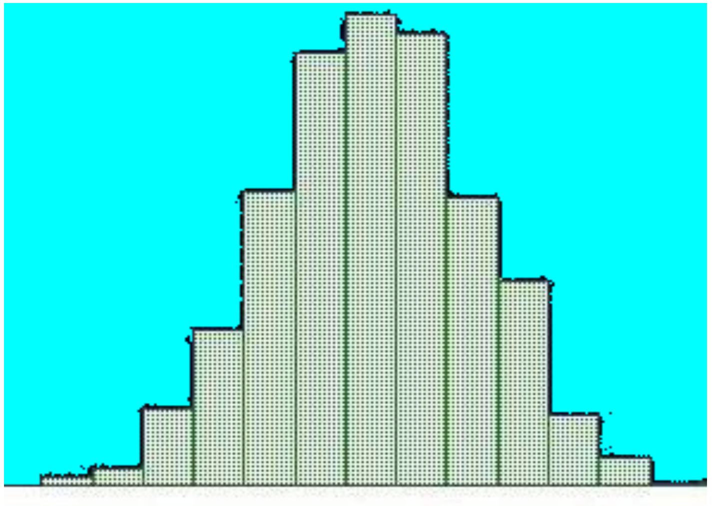
$$=1-0.9772=0.0228$$

即该公司亏本的概率为 **0.0228**

总结

中心极限定理是概率论中最著名的结果之一：

- 它不仅提供了计算独立随机变量之和的近似概率的简单方法；
- 而且有助于解释为什么很多自然群体的经验频率呈现出正态曲线这一值得注意的事实。



第五章作业（教材第五版）：

P128：2

P129：6、9、10、11

P130: 13

注：作业不得抄袭；写上姓名、班级、学号和页码（如1/5），同六、七章作业一起提交至教学云平台，标明题目属于第五章/第六章/第七章。