

第五章

大数定律及中心极限定理

第一节 大数定律

一、问题的引入

二、基本定理

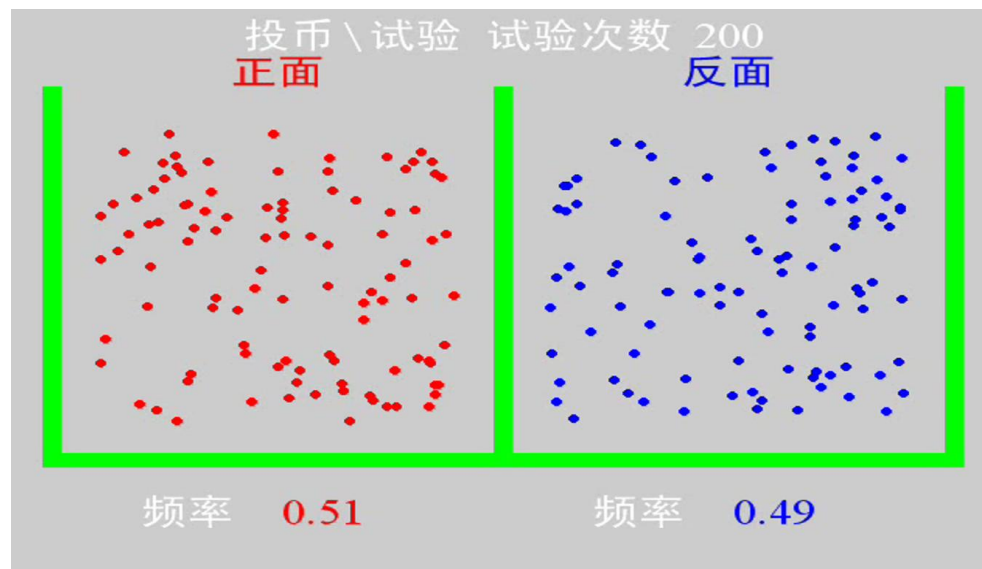
三、典型例题

四、小结

一、问题的引入

实例 频率的稳定性

随着试验次数的增加, 事件发生的频率逐渐稳定于某个常数.



频率的稳定性是概率定义的客观基础。

本节将对频率的稳定性做出理论的说明。

二、基本定理

定理一（切比雪夫定理的特殊情况）

设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立,
且具有相同的数学期望和方差: $E(X_k) = \mu$,
 $D(X_k) = \sigma^2$ ($k = 1, 2, \dots$), 作前 n 个随机变量

的算术平均 $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$, 则对于任意正

数 ε 有

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

二、基本定理

定理一（契比雪夫定理的特殊情况）

设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立,
且具有相同的数学期望和方差: $E(X_k) = \mu$,
 $D(X_k) = \sigma^2$ ($k = 1, 2, \dots$), 作前 n 个随机变量
的算术平均 $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$, 则对于任意正
数 ε 有

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

二、基本定理

定理

表达式的意义

$\{|\bar{X} - \mu| < \varepsilon\}$ 是一个随机事件, 等式表明, 当 $n \rightarrow \infty$ 时这个事件的概率趋于1, 即对于任意正数 ε , 当 n 充分大时, 不等式 $|\bar{X} - \mu| < \varepsilon$ 成立的概率很大.

且:

$D(\bar{X})$

的算

$\sum_{k=1}^n$

数 ε 有

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

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

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

由契比雪夫不等式可得

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

在上式中令 $n \rightarrow \infty$, 并注意到概率不能大于1, 则

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

关于定理一的说明:

当 n 很大时, 随机变量 X_1, X_2, \dots, X_n 的算术平

均 $\frac{1}{n} \sum_{k=1}^n X_k$ 接近于数学期望

$$E(X_1) = E(X_2) = \dots = E(X_k) = \mu,$$

(这个接近是概率意义下的接近)

即在定理条件下, n 个随机变量的算术平均, 当 n 无限增加时, 几乎变成一个常数.

定理一的另一种叙述:

设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立,
且具有相同的数学期望和方差: $E(X_k) = \mu$,

$D(X_k) = \sigma^2$ ($k = 1, 2, \dots$), 则序列 $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$

依概率收敛于 μ , 即 $\bar{X} \xrightarrow{P} \mu$.

定理一的另一种叙述:

设随机变量 X_1, X_2, \dots, X_n 且具有相同的数学期望

$$D(X_k) = \sigma^2 \quad (k = 1, 2, \dots)$$

依概率收敛于 μ , 即 \bar{X} .

设 Y_1, Y_2, \dots, Y_n 是一个随机变量序列, a 是一个常数, 若对于任意正数 ε 有 $\lim_{n \rightarrow \infty} P\{|Y_n - a| < \varepsilon\} = 1$, 则称序列 Y_1, Y_2, \dots, Y_n 依概率收敛于 a , 记为

$$Y_n \xrightarrow{P} a$$

依概率收敛序列的性质:

设 $X_n \xrightarrow{P} a, \quad Y_n \xrightarrow{P} b,$

又设函数 $g(x, y)$ 在点 (a, b) 连续,

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



定理二 弱大数定理（**辛钦大数定理**）

设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立,
服从同一分布, 且具有数学期望 $E(X_k) = \mu$
($k = 1, 2, \dots$),

则对于任意正数 ε , 有 $\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{k=1}^n X_k - \mu\right| < \varepsilon\right\} = 1$.

关于辛钦定理的说明:

- (1) 与定理一相比, 不要求方差存在;
- (2) 伯努利定理是辛钦定理的一个重要推论.

定理三（伯努利大数定理）

设 n_A 是 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, 则对于任意正数 $\varepsilon > 0$, 有

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

证明 引入随机变量

$$X_k = \begin{cases} 0, & \text{若在第 } k \text{ 次试验中 } A \text{ 不发生,} \\ 1, & \text{若在第 } k \text{ 次试验中 } A \text{ 发生, } k = 1, 2, \dots \end{cases}$$

显然 $n_A = X_1 + X_2 + \cdots + X_n$,

因为 $X_1, X_2, \cdots, X_n, \cdots$ 是相互独立的,

且 X_k 服从以 p 为参数的 $(0-1)$ 分布,

所以 $E(X_k) = p$, $D(X_k) = p(1-p)$, $k = 1, 2, \cdots$.

根据弱大数定理有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} (X_1 + X_2 + \cdots + X_n) - p \right| < \varepsilon \right\} = 1,$$

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

关于伯努利定理的说明:

伯努利定理表明事件发生的频率 $\frac{n_A}{n}$ 依概率收敛于事件的概率 p , 它以严格的数学形式表达了频率的稳定性.

故而当 n 很大时, 事件发生的频率与概率有较大偏差的可能性很小. 在实际应用中, 当试验次数很大时, 便可以用事件发生的频率来代替事件的概率.

四、小结

三个大数定理 { 契比雪夫定理的特殊情况
辛钦定理
伯努利大数定理

频率的稳定性是概率定义的客观基础，而伯努利大数定理以严密的数学形式论证了频率的稳定性.

第二节 中心极限定理

一、问题的引入

二、基本定理

三、典型例题

四、小结

一、问题的引入

实例：考察射击命中点与靶心距离的偏差.

这种偏差是大量微小的偶然因素造成的微小误差的总和, 这些因素包括: 瞄准误差、测量误差、子弹制造过程方面 (如外形、重量等) 的误差以及射击时武器的振动、气象因素(如风速、风向、能见度、温度等) 的作用, 所有这些不同因素所引起的微小误差是相互独立的, 并且它们中每一个对总和产生的影响不大.

问题：某个随机变量是由大量相互独立且均匀小的随机变量相加而成的, 研究其概率分布情况.

二、基本定理

定理一（独立同分布的中心极限定理）

设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立, 服从同一分布, 且具有数学期望和方差: $E(X_k) = \mu$, $D(X_k) = \sigma^2 > 0$ ($k = 1, 2, \dots$), 则随机变量之和的

标准化变量 $Y_n = \frac{\sum_{k=1}^n X_k - E\left(\sum_{k=1}^n X_k\right)}{\sqrt{D\left(\sum_{k=1}^n X_k\right)}} = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n} \sigma}$

的分布函数 $F_n(x)$ 对于任意 x 满足

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n} \sigma} \leq x \right\} \\ = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x).$$

定理一表明:

当 $n \rightarrow \infty$, 随机变量序列 Y_n 的分布函数收敛于标准正态分布的分布函数.

定理二(李雅普诺夫定理)

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

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

记
$$B_n^2 = \sum_{k=1}^n \sigma_k^2,$$

若存在正数 δ , 使得当 $n \rightarrow \infty$ 时,

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

则随机变量之和的标准化变量

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

的分布函数 $F_n(x)$ 对于任意 x 满足

$$\begin{aligned} \lim_{n \rightarrow \infty} F_n(x) &= \lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k}{B_n} \leq x \right\} \\ &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x). \end{aligned}$$

定理二表明:

无论各个随机变量 $X_1, X_2, \dots, X_n, \dots$ 服从什么分布, 只要满足定理的条件, 那么它们的和 $\sum_{k=1}^n X_k$ 当 n 很大时, 近似地服从正态分布.

(如实例中射击偏差服从正态分布)

下面介绍的定理三是定理一的特殊情况.

定理三(棣莫佛—拉普拉斯定理)

设随机变量 η_n ($n = 1, 2, \dots$) 服从参数为 n, p ($0 < p < 1$) 的二项分布, 则 对于任意 x , 恒有

$$\lim_{n \rightarrow \infty} P\left\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \leq x\right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x).$$

证明 根据第四章第二节例题可知 $\eta_n = \sum_{k=1}^n X_k$,

其中 X_1, X_2, \dots, X_n 是相互独立的、服从同一 (0—1) 分布的随机变量, 分布律为

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

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

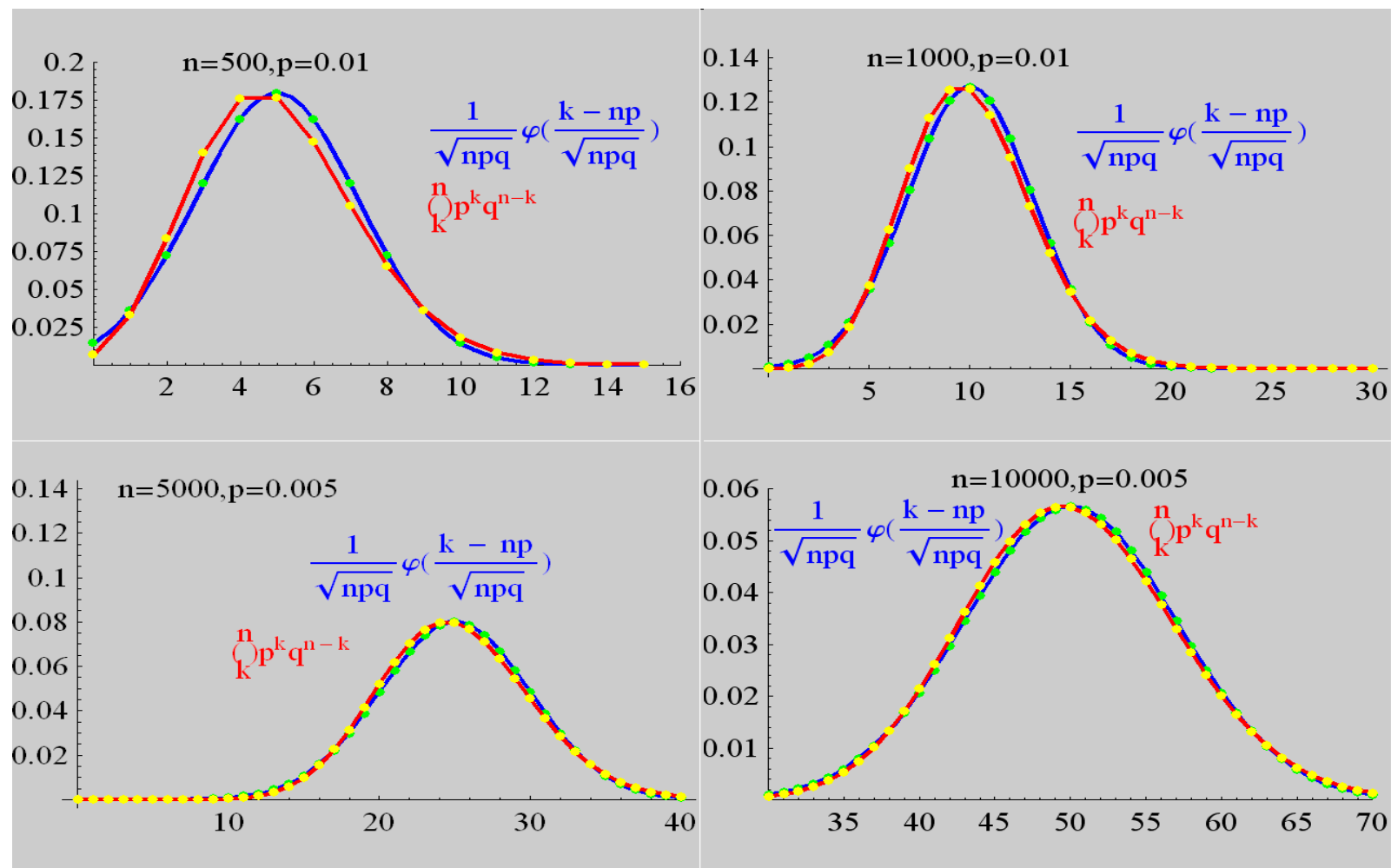
根据定理一得

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left\{ \frac{\eta_n - np}{\sqrt{np(1-p)}} \leq x \right\} &= \lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n X_k - np}{\sqrt{np(1-p)}} \leq x \right\} \\ &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x). \end{aligned}$$

定理三表明:

正态分布是二项分布的极限分布, 当 n 充分大时, 可以利用该定理来计算二项分布的概率.

下面的图形表明:正态分布是二项分布的逼近.



例1 一船舶在某海区航行, 已知每遭受一次海浪的冲击, 纵摇角大于 3° 的概率为 $1/3$, 若船舶遭受了 90 000 次波浪冲击, 问其中有 29 500 ~ 30 500 次纵摇角大于 3° 的概率是多少?

解 将船舶每遭受一次海浪

的冲击看作一次试验,

并假设各次试验是独立的,

在 90 000 次波浪冲击中纵摇角大于 3° 的次数为 X ,

则 X 是一个随机变量, 且 $X \sim b(90\,000, \frac{1}{3})$.



分布律为 $P\{X = k\} = \binom{90\,000}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{90\,000-k}$,

$k = 1, \dots, 90\,000.$

所求概率为

$$P\{29500 < X \leq 30500\} = \sum_{k=29501}^{30500} \binom{90\,000}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{90\,000-k}.$$

直接计算很麻烦，利用棣莫佛—拉普拉斯定理

$$P\{29500 < X \leq 30500\}$$

$$= P\left\{ \frac{29500 - np}{\sqrt{np(1-p)}} < \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{30500 - np}{\sqrt{np(1-p)}} \right\}$$

$$\approx \int_{\frac{29500-np}{\sqrt{np(1-p)}}}^{\frac{30500-np}{\sqrt{np(1-p)}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$= \Phi\left(\frac{30500-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{29500-np}{\sqrt{np(1-p)}}\right)$$

$$\because n = 90000, \quad p = \frac{1}{3},$$

$$\therefore P\{29500 < X \leq 30500\} \approx \Phi\left(\frac{5\sqrt{2}}{2}\right) - \Phi\left(-\frac{5\sqrt{2}}{2}\right)$$

$$= 0.9995.$$

例2 某保险公司的老年人寿保险有1万人参加,每人每年交**200**元. 若老人在该年内死亡,公司付给家属**1**万元. 设老年人死亡率为**0.017**,试求保险公司在一年内的这项保险中亏本的概率.

解 设 X 为一年中投保老人的死亡数,

则 $X \sim B(n, p)$,

其中 $n = 10000$, $p = 0.017$,

由**德莫佛—拉普拉斯定理**知,



保险公司亏本的概率

$$P\{10000X > 10000 \times 200\} = P\{X > 200\}$$

$$= P\left\{\frac{X - np}{\sqrt{np(1-p)}} > \frac{200 - np}{\sqrt{np(1-p)}}\right\}$$

$$= P\left\{\frac{X - np}{\sqrt{np(1-p)}} > 2.321\right\}$$

$$\approx 1 - \Phi(2.321) \approx 0.01 .$$

四、小结

三个中心极限定理 { 独立同分布的中心极限定理
李雅普诺夫定理
德莫佛—拉普拉斯定理

中心极限定理表明, 在相当一般的条件下,
当独立随机变量的个数增加时, 其和的分布趋于
正态分布.