

§ 1. 大数定律

§ 1 大数定律

在实践中，不仅事件发生的频率具有稳定性，还有大量测量值的算术平均值也具有稳定性。

定义1:

设 Y_1, \dots, Y_n, \dots 是随机变量序列， a 是一个常数；
若对任意 $\varepsilon > 0$ ，有： $\lim_{n \rightarrow \infty} P\{|Y_n - a| < \varepsilon\} = 1$
则称 Y_1, \dots, Y_n, \dots 依概率收敛于 a ，记为 $Y_n \xrightarrow{P} a$ 。

定理1: 若 $X_n \xrightarrow{P} a$ ， $Y_n \xrightarrow{P} b$ ， $g(x, y)$ 在点 (a, b) 连续，
则： $g(X_n, Y_n) \xrightarrow{P} g(a, b)$ 。



定理 2 (切比晓夫大数定律)

设随机变量 X_1, \dots, X_n, \dots 相互独立, 且具有相同的数学期望及方差, $EX_k = \mu$, $DX_k = \sigma^2$, $k = 1, 2, \dots$, 令 $Y_n = \frac{1}{n} \sum_{k=1}^n X_k$, 则: 对任意的 $\varepsilon > 0$, 有:

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

$$\text{或 } \lim_{n \rightarrow \infty} P\{|\frac{1}{n} \sum_{k=1}^n X_k - \mu| \geq \varepsilon\} = 0$$



证:
$$E\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n EX_k = \frac{1}{n} \sum_{k=1}^n \mu = \mu$$

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

由切比晓夫不等式得:
$$P\left\{\left|\frac{1}{n} \sum_{k=1}^n X_k - \mu\right| < \varepsilon\right\} \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$$

当 $n \rightarrow \infty$ 时,
$$P\left\{\left|\frac{1}{n} \sum_{k=1}^n X_k - \mu\right| < \varepsilon\right\} = 1 \text{。}$$



定理 3 (贝努里大数定律)

设 n_A 是 n 次独立重复试验中事件 A 发生的次数,
 p 是事件 A 发生的概率,
则: 对任意的 $\varepsilon > 0$, 有

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

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

则 $n_A = \sum_{k=1}^n X_k$, 且 X_1, \dots, X_n 相互独立同服从于(0-1)分布

故 $EX_k = p, \quad DX_k = p(1-p), \quad k = 1, 2, \dots, n, \dots$

由定理2有 $\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - p\right| < \varepsilon\right\} = 1,$

即 $\lim_{n \rightarrow \infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1$ 。 此定理说明了频率的稳定性。

定理 4（辛钦大数定律）

设 X_1, \dots, X_n, \dots 相互独立同分布，且具有数学期望 $EX_k = \mu$, $k = 1, 2, \dots, n, \dots$,

则：对任意的 $\varepsilon > 0$ ，有

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

注： 贝努里大数定律是辛钦大数定律的特殊情况。



§ 2. 中心极限定理

定义:

设 X_1, \dots, X_n, \dots 是独立的随机变量序列,

EX_k, DX_k 存在, 令: $Z_n = (\sum_{k=1}^n X_k - \sum_{k=1}^n EX_k) / \sqrt{\sum_{k=1}^n DX_k}$,

若对任意 $x \in R_1$, 有 $\lim_{n \rightarrow \infty} P\{Z_n \leq x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$ 。

则称 $\{X_n\}$ 服从中心极限定理。



定理1 (独立同分布的中心极限定理)

设 X_1, \dots, X_n, \dots 是独立同分布的随机变量序列, 且 $EX_k = \mu$, $DX_k = \sigma^2 \neq 0, (k = 1, 2, \dots)$
则 $\{X_n\}$ 服从中心极限定理, 即:

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



定理2 (李雅普诺夫定理)(Liapunov定理) § 2 中心极限定理

设 X_1, \dots, X_n, \dots 相互独立, 且 $EX_k = \mu_k$, $DX_k = \sigma_k^2 \neq 0$,

($k = 1, 2, \dots$), 设 $B_n^2 = \sum_{k=1}^n \sigma_k^2$, 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 $\{X_n\}$ 服从中心极限定理, 即:

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



定理3 (德莫佛-拉普拉斯定理) (De Moivre--Laplace)

设随机变量 $\eta_n (n = 1, 2, \dots)$ 服从参数为 $n, p (0 < p < 1)$ 的二项分布, 即 $\eta_n \sim B(n, p)$.

则对于任意 x , 恒有:

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

证: $\eta_n = \sum_{k=1}^n X_k,$

其中 X_1, \dots, X_n 相互独立且都服从于 (0-1) 分布。

$$EX_k = p, \quad DX_k = pq.$$

由定理1有结论成立。

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

推论:

设随机变量 $\eta_n (n = 1, 2, \dots)$ 服从参数为 n, p ($0 < p < 1$) 的二项分布, 即 $\eta_n \sim B(n, p)$. 当 n 充分大时有:

$$\begin{aligned} P\{a < \eta_n \leq b\} &= P\left\{\frac{a - np}{\sqrt{npq}} < \frac{\eta_n - np}{\sqrt{npq}} \leq \frac{b - np}{\sqrt{npq}}\right\} \\ &\approx \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right) \end{aligned}$$

说明: 这个公式给出了 n 较大时二项分布的概率计算方法。

例1

某车间有200台车床，它们独立地工作着，开工率为0.6,开工时耗电各为1千瓦，问供电所至少要供给这个车间多少电力才能以99.9%的概率保证这个车间不会因供电不足而影响生产。

解：记某时在工作着的车床数为 X , 则 $X \sim B(200, 0.6)$.

设至少要供给这个车间 r 千瓦电才能以99.9%的概率保证这个车间不会因供电不足而影响生产。由题意有：

$$P\{X \leq r\} = \sum_{k=0}^r C_{200}^k (0.6)^k (0.4)^{200-k}$$

$$\approx \Phi\left(\frac{r - 200 \times 0.6}{\sqrt{200 \times 0.6 \times 0.4}}\right) - \Phi\left(\frac{-200 \times 0.6}{\sqrt{200 \times 0.6 \times 0.4}}\right)$$



$$= \Phi\left(\frac{r-120}{\sqrt{48}}\right) - \Phi(-17.32) \approx \Phi\left(\frac{r-120}{\sqrt{48}}\right) \geq 0.999,$$

查表得

$$\frac{r-120}{\sqrt{48}} \geq 3.1 \quad \text{所以} \quad r \geq 141.$$

即供给141千瓦电就能以99.9%的概率保证这个车间不会因供电不足而影响生产。



用频率估计概率时误差的估计：

由上面的定理知

$$\begin{aligned} P\left\{\left|\frac{\eta_n}{n} - p\right| < \varepsilon\right\} &= P\left\{\left|\frac{\eta_n - np}{n}\right| < \varepsilon\right\} = \\ &= P\left\{-\varepsilon\sqrt{\frac{n}{pq}} < \frac{\eta_n - np}{\sqrt{npq}} < \varepsilon\sqrt{\frac{n}{pq}}\right\} \\ &\approx \Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) - \Phi\left(-\varepsilon\sqrt{\frac{n}{pq}}\right) = 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) - 1 \end{aligned}$$

用这个关系式可解决许多计算问题。



例2.

现有一批种子，其中良种占1/6。今任取6000粒，问能以0.99的概率保证在这6000粒种子中良种所占的比例与1/6的差不超过多少？相应的良种粒数在哪个范围内？

解：

设良种数为 X ，则 $X \sim B(n, p)$ ，其中 $n = 6000$ ， $p = 1/6$ 。

设不超过的界限为 α ，则应有：

$$P\left\{\left|\frac{X}{6000} - \frac{1}{6}\right| \leq \alpha\right\} = 0.99$$

由德莫佛-拉普拉斯定理

$$P\left\{\left|\frac{X}{6000} - \frac{1}{6}\right| \leq \alpha\right\}$$

$$n = 6000, p = 1/6.$$

$$\lim_{n \rightarrow \infty} P\left\{\frac{\eta_n - np}{\sqrt{npq}} \leq x\right\} = \Phi(x)$$

$$= P\left\{\left|\frac{X - 6000 \times 1/6}{\sqrt{6000 \times 1/6 \times 5/6}}\right| \leq \frac{6000 \alpha}{\sqrt{6000 \times 1/6 \times 5/6}}\right\}$$

$$\approx 2\Phi\left[\frac{6000 \alpha}{\sqrt{6000 \times 1/6 \times 5/6}}\right] - 1$$

故近似地有

$$2\Phi\left[\frac{6000 \alpha}{\sqrt{6000 \times 1/6 \times 5/6}}\right] - 1 = 0.99,$$



即
$$\Phi\left[\frac{6000\alpha}{\sqrt{6000 \times 1/6 \times 5/6}}\right] = 0.995,$$

查表得
$$\frac{6000\alpha}{\sqrt{6000 \times 1/6 \times 5/6}} = 2.58,$$

解得
$$\alpha = 0.0124.$$

良种粒数X的范围为

$$\left| \frac{X}{6000} - \frac{1}{6} \right| \leq \alpha$$

$$(1/6 - 0.0124) \times 6000 \leq X \leq (1/6 + 0.0124) \times 6000,$$

即
$$925 \leq X \leq 1075.$$

思考题:

假设一批种子的良种率为 $\frac{1}{6}$ ，从中任意选出600粒，试用切比晓夫（Chebyshev）不等式和中心极限定理分别估计：这600粒种子中良种所占比例与 $\frac{1}{6}$ 差的绝对值不超过0.02的概率。

$$EX = 600 \times \frac{1}{6}, \quad DX = 600 \times \frac{1}{6} \times \frac{5}{6}.$$

由切比晓夫不等式有

$$\begin{aligned} P\left\{\left|\frac{X}{600} - \frac{1}{6}\right| \leq 0.02\right\} &= P\left\{\left|\frac{X - 100}{600}\right| \leq 0.02\right\} \\ &= P\{|X - 100| \leq 12\} \geq 1 - \frac{DX}{12^2} = 1 - \frac{600 \times \frac{1}{6} \times \frac{5}{6}}{144} = 0.4213 \end{aligned}$$

例3

设一个系统由100个相互独立起作用的部件组成，每个部件的损坏率为0.1。为了使整个系统正常工作，至少必须有85个部件正常工作，求整个系统正常工作的概率。

解： 设 X 是损坏的部件数，则 $X \sim B(100, 0.1)$ 。则整个系统能正常工作当且仅当 $X \leq 15$ 。

由德莫佛-拉普拉斯定理有

$$\begin{aligned} P\{X \leq 15\} &= P\left\{ \frac{X - 100 \times 0.1}{\sqrt{100 \times 0.1 \times 0.9}} \leq \frac{15 - 100 \times 0.1}{\sqrt{100 \times 0.1 \times 0.9}} \right\} \\ &\approx \Phi\left(\frac{15 - 100 \times 0.1}{\sqrt{100 \times 0.1 \times 0.9}} \right) = \Phi\left(\frac{5}{3} \right) = 0.952. \end{aligned}$$



例4某单位有200台电话分机，每台分机有5%的时间要使用外线通话。假定每台分机是否使用外线是相互独立的，问该单位总机要安装多少条外线，才能以90%以上的概率保证分机用外线时不等待？

解：设有X部分机同时使用外线，则有 $X \sim B(n, p)$ ，其中 $n = 200, p = 0.05, np = 10, \sqrt{np(1-p)} = 3.08$ 。

设有N条外线。由题意有 $P\{X \leq N\} \geq 0.9$

由德莫佛-拉普拉斯定理有

$$P\{X \leq N\} = P\left\{\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{N - np}{\sqrt{np(1-p)}}\right\} \approx \Phi\left(\frac{N - np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{N - 10}{3.08}\right).$$

查表得 $\Phi(1.28) = 0.90$.故 N 应满足条件 $\frac{N - 10}{3.08} \geq 1.28$,

即 $N \geq 13.94$. 取 $N = 14$, 即至少要安装14条外线。

例5 一加法器同时收到20个噪声电压 $V_k (k = 1, 2, \dots, 20)$, 设它们是互相独立的随机变量, 且都在区间(0,10)上服从均匀分布, 记

$$V = \sum_{k=1}^{20} V_k$$

求 $P\{V > 105\}$ 近似值。

解: $EV_k = 5, DV_k = \frac{10^2}{12}, (k = 1, 2, \dots, 20)$, 由定理 1 知:

$$\begin{aligned} P\{V > 105\} &= P\left\{ \frac{V - 20 \times 5}{\sqrt{10^2 / 12} \times \sqrt{20}} > \frac{105 - 20 \times 5}{\sqrt{10^2 / 12} \times \sqrt{20}} \right\} \\ &= P\left\{ \frac{V - 100}{(10 / \sqrt{12}) \times \sqrt{20}} > 0.387 \right\} = 1 - P\left\{ \frac{V - 100}{(10 / \sqrt{12}) \times \sqrt{20}} \leq 0.387 \right\} \end{aligned}$$

$$\approx 1 - \Phi(0.387) = 0.348$$



- 1 引进了大数定律的概念，要了解大数定律的意义和内容，理解贝努里、辛钦大数定律，了解契比雪夫大数定律。
- 2 阐述了中心极限定理的含义及其客观背景，要掌握独立同分布的中心极限定理和德莫佛-拉普拉斯定理，会利用中心极限定理解决一般实际应用问题。

作业：1;3;5;7;9;11

实验：5.1节 例2 利用随机数计算积分

