第五章: 极限定理

5.1 大数定律

5.2 中心极限定理

5.1 大数定律

切比雪夫不等式

定理 1.1 (切比雪夫不等式)

设随机变量 X 有 $E(X) = \mu, Var(X) = \sigma^2$, 则对 $\forall \epsilon > 0$, 成立不等式

$$P(|X - \mu| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}.$$

证明: 设 X 是一个连续随机变量, 且有密度 $f(x)$. 则

$$\begin{aligned}
P(|X - \mu| \geq \varepsilon) &= \int_{\{x:|x-\mu|\geq\varepsilon\}} f(x)dx \\
&\leq \int_{\{x:|x-\mu|\geq\varepsilon\}} \frac{(x-\mu)^2}{\varepsilon^2} f(x)dx \\
&\leq \frac{1}{\varepsilon^2} \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \frac{Var(X)}{\varepsilon^2}.
\end{aligned}$$

$\square$

切比雪夫不等式

切比雪夫不等式也可以写为

$$P\{|X-\mu|<\varepsilon\} \geq 1-\frac{\sigma^2}{\varepsilon^2}.$$

切比雪夫不等式说明, **X 的方差越小, 则事件 $\{|X-\mu|<\varepsilon\}$ 发生的概率就越大.** 即 X 取的值越集中于它的期望 μ 附近.

- 在 X 的分布未知的情况下，估计概率值 $P\{|X-\mu|<\varepsilon\}$ 或 $P(|X-\mu|\geq\varepsilon)$.
- 例如, 令 $\varepsilon=3\sigma$,

$$P\{|X-\mu|<3\sigma\} \geq 1-\frac{\sigma^2}{9\sigma^2}=0.8889.$$

依概率收敛: 背景

该怎样用极限语言描述频率稳定性?

频率收敛于概率!

当 n 足够大, 频率 $\frac{n_A}{n}$ 与概率 p 有较大偏差的概率很小, 即对 $\forall \varepsilon > 0$, 有

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

称为**频率稳定性.**

依概率收敛: 定义

定义 1.1

设 $\{X_n\}$ 为一随机变量序列, X 为一随机变量, 如果对任意的 $\epsilon > 0$, 有

$$P(|X_n - X| \geq \epsilon) \to 0, \quad n \to \infty,$$

则称序列 $\{X_n\}$ 依概率收敛于 X, 记作 $X_n \xrightarrow{P} X$. 其等价形式为

$$P(|X_n - X| < \epsilon) \to 1, \quad n \to \infty.$$

特别地, 当 X 为退化分布时, 即 $P(X = c) = 1$, 则称序列 $\{X_n\}$ 依概率收敛于 c, 即

$$X_n \xrightarrow{P} c.$$

大数定律

定理 1.2

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立, 并且具有相同的期望和方差, $E(X_i) = \mu$, $Var(X_i) = \sigma^2$, $i = 1, 2, \cdots$. 作前 n 个随机变量的平均 $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, 则 $\forall \epsilon > 0$, 有

$$\lim_{n\to\infty} P\{|Y_n - \mu| < \varepsilon\} = 1.$$

证明: 计算 Y_n 的期望和方差

$$E(Y_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu, \; Var(Y_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}.$$

由切比雪夫不等式可知, $P\{|Y_n - \mu| < \varepsilon\} \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$. 因为概率不可能大于 1. 令 $n \to \infty$, $P\{|Y_n - \mu| < \varepsilon\} = 1$. $\square$

定理 1.3 (伯努利大数定律)

设在 n 次独立重复试验中事件 A 发生的次数为 n_A, p 是事件 A 在每次试验中发生的概率, 则对 $\forall\epsilon$, 有

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

证: 因为

$$E\left(\frac{n_A}{n}\right) = p, \quad Var\left(\frac{n_A}{n}\right) = \frac{p(1-p)}{n}.$$

根据切比雪夫不等式,

$$P\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right) \geq 1 - \frac{p(1-p)}{n\epsilon^2} \to 1, n \to \infty.$$

□

伯努利大数定律: 例子

例 1.1

抛一枚硬币出现正面的概率是 $p=0.5$. 若把这枚硬币连续抛 n 次, 取精度 $\epsilon=0.01$, 由切比雪夫不等式知:

$$P\left(\left|\frac{n_A}{n}-0.5\right|\geq 0.01\right)\leq\frac{0.5\times 0.5}{n0.01^2}=\frac{10^4}{4n}.$$

- 当 $n=10^5$ 时, 大偏差发生可能性小于 $1/40=2.5\%$.
- 当 $n=10^6$ 时, 大偏差发生可能性小于 $1/400=0.25\%$.

由此, 试验次数越多, 大偏差发生的概率越低.

5.2 中心极限定理

中心极限定理: 独立同分布的中心极限定理

定理 2.1 (林德伯格-莱维中心极限定理)

设随机变量 $X_1, X_2, \cdots, n, \cdots$ 相互独立, 服从同一分布, 且具有期望和方差: $E(X_i) = \mu, Var(X_i) = \sigma^2 > 0, i = 1, 2, \cdots,$ 则随机变量

$$Y_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

的分布函数 $F_n(x)$ 收敛到标准正态分布 $N(0,1)$ 的分布函数, 即对任意实数 x 满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\{Y_n \leq x\} = \Phi(x),$$

其中 $\Phi(x)$ 是标准正态分布 $N(0,1)$ 的分布函数.

中心极限定理: 独立同分布的中心极限定理

X_i 的分布是任意的, $\sum_{i=1}^n X_i$ 的分布难以确切求得, 这时只要 n 很大, 就能通过 $\Phi(x)$ 给出 $\sum_{i=1}^n X_i$ 的分布函数的近似值.

注意到 $E(\sum_{i=1}^n X_i) = n\mu$, $Var(\sum_{i=1}^n) = n\sigma^2$, Y_n 可写为

$$Y_n = \frac{\sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right)}{\sqrt{Var\left(\sum_{i=1}^n X_i\right)}},$$

从而 Y_n 的期望是 0, 方差是 1, Y_n 是 $\sum_{i=1}^n X_i$ 的标准化随机变量.

中心极限定理: 独立同分布的中心极限定理

例 2.1

设一批产品的强度的服从期望未 14, 方差为 4 的分布. 每箱中装有这种产品 100 件, 问

(1) 每箱产品的平均强度超过 14.5 的概率是多少?

(2) 每箱产品的平均强度超过期望 14 的概率是多少?

解: 样本量 $n = 100$. 设 X_i 是第 i 件产品的强度, $E(X_i) = 14$, $Var(X_i) = 4$, $i = 1, 2, \cdots, 100$. 记 $\bar{X} = \sum_{i=1}^{100} X_i/100$, 利用中心极限定理, 近似地有

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - 14}{0.2} \sim N(0, 1).$$

(1) 计算

$$P\{\bar{X} > 14.5\} = P\left\{\frac{\bar{X} - 14}{0.2} > \frac{14.5 - 14}{0.2}\right\} = P\left\{\frac{\bar{X} - 14}{0.2} > 2.5\right\}$$
$$\approx 1 - \Phi(2.5) = 0.0062.$$

可见, 100 件产品的平均强度超过 14.5 的概率非常之小.
(2) 计算

$$P\{\bar{X} > 14\} = P\left\{\frac{\bar{X} - 14}{0.2} > 0\right\} \approx \Phi(0) = 0.5.$$

可以说, 每箱产品的平均强度超过 14 的概率约为 50%. □

例 2.2

计算机进行数字计算时, 遵循四舍五入原则. 为简单计, 现在对小数点后吗第一位近下舍入运算, 则舍入误差可以认为服从 $[-0.5, 0.5]$ 上的均匀分布. 若独立进行了 100 次数字计算, 求这些计算的平均舍入误差落在 $[-\sqrt{3}/20, \sqrt{3}/20]$ 上的概率.

解: 样本容量 $n = 100$. 用 X_i 表示第 i 次运算中产生的舍入误差. $X_1, X_2, \cdots, X_{100}$ 相互独立, 服从 $[-0.5, 0.5]$ 上的均匀分布. 这时,

$$E(X_i) = 0, \quad Var(X_i) = 1/12,$$

$i = 1, 2, \cdots, 100.$

从而, 近似地有

$$Y_{100} = \frac{\sum_{i=1}^{100} X_i - 100}{\sqrt{100/12}} = \frac{\sqrt{3}}{5} \sum_{i=1}^{100} X_i \sim N(0,1).$$

于是, 平均误差 $\bar{X}$ 落在区间 $[-\sqrt{3}/20, \sqrt{3}/20]$ 上的概率为

$$\begin{aligned} P\left\{-\frac{\sqrt{3}}{20} \le \bar{X} \le \frac{\sqrt{3}}{20}\right\} &= P\left\{-\frac{\sqrt{3}}{20} \le \frac{1}{100}\sum_{i=1}^{100} X_i \le \frac{\sqrt{3}}{20}\right\} \\ &= P\left\{-3 \le \frac{\sqrt{3}}{5}\sum_{i=1}^{100} X_i \le 3\right\} \\ &\approx \Phi(3) - \Phi(-3) = 0.9973. \end{aligned}$$

□

中心极限定理: 棣莫弗-拉普拉斯中心极限定理

定理 2.2 (棣莫弗-拉普拉斯中心极限定理)

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立且服从参数为 p 的两点分布, 则对任意实数 x, 有

$$\lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \le x\right\} = \Phi(x).$$

由于 $\sum_{i=1}^{n} X_i \sim B(n,p)$, 有关二项分布的计算, 当 n 很大时, 可以通过 $\Phi(x)$ 来计算. 另外,

n 很大时, 二项分布可以用正态分布来近似!

例 2.3

某公司有 200 名员工参加一种资格证书考试, 按往年经验该考试的通过率为 0.8. 试计算这 200 名员工至少有 150 人考试通过的概率.

解: 令

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 人通过考试}, \\ 0, & \text{第 } i \text{ 人未通过考试}, \end{cases}$$

$i = 1, 2, \cdots, 200$. 且 $P\{X_i = 1\} = 0.8$, 计算

$$np = 200 \times 0.8 = 160,$$

$$np(1-p) = 200 \times 0.8 \times 0.2 = 32.$$

中心极限定理: 棣莫弗-拉普拉斯中心极限定理

显然, $\sum_{i=1}^{200} X_i$ 是考试通过人数, 因为 X_i 满足棣莫弗-拉普拉斯中心极限定理, 故近似地有

$$\frac{\sum_{i=1}^{200} X_i - 160}{\sqrt{32}} \sim N(0,1).$$

于是

$$\begin{aligned} P\left\{ \sum_{i=1}^{200} X_i \geq 150 \right\} &= P\left\{ \frac{\sum_{i=1}^{200} X_i - 160}{\sqrt{32}} \geq -1.77 \right\} \\ &\approx 1 - \Phi(-1.77) = \Phi(1.77) \\ &= 0.9616. \end{aligned}$$

即至少有 150 人通过考试的概率为 0.9616. □