第四章　大数定律和中心极限定理

- §1 大数定律

■ 背景

本章的大数定律，对第一章中提出的“频率稳定性”，给出理论上的论证

上页 下页

设随机变量序列$X_1, X_2, X_3, \cdots$,若存在某常数μ,

使得$\forall \varepsilon > 0$,均有:$\lim\limits_{n\to+\infty} P\{|X_n - \mu| \geq \varepsilon\} = 0$,

则称随机变量序列$\{X_n\}$依概率收敛于常数μ,

记为:$X_n \xrightarrow{p} \mu$。

定理5.2(契比雪夫不等式的特殊情形)：

设随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，且具有相同的数学期望 μ 和相同的方差 σ^2，

作前 n 个随机变量的算术平均：$Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k$

则 $\forall \varepsilon > 0$，有：

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

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

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

由契比雪夫不等式得：$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right\} \geq 1 - \frac{\sigma^2/n}{\varepsilon^2}$

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

定理5.3 (贝努里大数定理)

设事件A在每次试验中发生的概率为p，记n_A为n次独立重复试验中A发生的次数,则$\forall \varepsilon > 0$,有: $\lim\limits_{n\to+\infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1$

证明：利用契比雪夫不等式，因$n_A \sim b(n, p)$,故:

$$E\left(\frac{n_A}{n}\right) = \frac{1}{n}E(n_A) = \frac{1}{n}\cdot np = p, \quad D\left(\frac{n_A}{n}\right) = \frac{1}{n^2}D(n_A) = \frac{1}{n^2}\cdot npq = \frac{pq}{n}$$

于是，$\forall \varepsilon > 0$,有$P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} \ge 1 - \frac{pq}{n\varepsilon^2}$

即得: $\lim\limits_{n\to+\infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1$

大数定律的重要意义：

贝努里大数定律建立了在大量重复独立试验中事件出现频率的稳定性，正因为这种稳定性，概率的概念才有客观意义，贝努里大数定律还提供了通过试验来确定事件概率的方法，既然频率n_A/n与概率p有较大偏差的可能性很小，我们便可以通过做试验确定某事件发生的频率并把它作为相应的概率估计，这种方法即是在第7章将要介绍的参数估计法，参数估计的重要理论基础之一就是大数定理。

上页 下页

§2 中心极限定理

- 背景：

　　有许多随机变量，它们是由大量的相互独立的随机变量的综合影响所形成的，而其中每个个别的因素作用都很小，这种随机变量往往服从或近似服从正态分布，或者说它的极限分布是正态分布，中心极限定理正是从数学上论证了这一现象，它在长达两个世纪的时期内曾是概率论研究的中心课题。

定理5.4（独立同分布的中心极限定理）

设随机变量$X_1, X_2, \cdots, X_n, \cdots$相互独立同分布，

$E(X_i)=\mu, D(X_i)=\sigma^2, i=1,2,\cdots$

则前n个变量的和的标准化变量为：$Y_n=\dfrac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}$

$\forall x\in R$,有：
$$\lim_{n\to+\infty}P(Y_n\le x)=\lim_{n\to+\infty}P\left(\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\le x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$

证明略。

此定理表明，当n充分大时，Y_n近似服从$N(0,1)$.

即：$\sum_{i=1}^{n}X_i$（近似）$\sim N(n\mu, n\sigma^2)$,

从而，$P(a<\sum_{i=1}^{n}X_i\le b)\approx\Phi(\dfrac{b-n\mu}{\sqrt{n}\sigma})-\Phi(\dfrac{a-n\mu}{\sqrt{n}\sigma})$.

思考题：

$\bar{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$的近似分布是什么？

答案：$N(\mu,\dfrac{\sigma^2}{n})$

上页 下页

定理5.5 (德莫佛--拉普拉斯定理)

设n_A为n次贝努里试验中A发生的次数，$P(A)=p(0<p<1)$,

则对任何区间$[a,b]$，有：$\lim\limits_{n\to+\infty} P\left(a<\dfrac{n_A-np}{\sqrt{np(1-p)}}\le b\right)=\int_a^b \dfrac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$,

$$P(a<n_A\le b) \approx \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right)$$

证明：令$X_i=\begin{cases}1 & 第i次试验时A发生\\ 0 & 第i次试验时A未发生\end{cases}$

则$X_1,X_2,\cdots,X_n,\cdots$相互独立同分布，$X_i\sim b(1,p)$.

由于$n_A=X_1+X_2+\cdots+X_n$,

由定理5.4, $\lim\limits_{n\to+\infty} P\left(a<\dfrac{n_A-np}{\sqrt{np(1-p)}}\le b\right)=\int_a^b \dfrac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$

即：n_A(近似)$\sim N(np,np(1-p))$.

上页 下页

设某种电器元件的寿命服从均值为100小时的指数分布，现随机取得16只，设它们的寿命是相互独立的,求这16只元件的寿命的总和大于1920小时的概率。

解：记16只电器元件的寿命分别为$X_1, X_2, \cdots, X_{16}$，

则16只电器元件的寿命总和为$X = \sum_{i=1}^{16} X_i$，

由题设$E(X_i) = 100, D(X_i) = 100^2$

根据独立同分布的中心极限定理:

$$Y = \frac{\sum_{i=1}^{16} X_i - 16\times 100}{4\times 100} = \frac{X - 1600}{400} \text{ 近似服从 } N(0,1)$$

$$P(X > 1920) = 1 - P(X \le 1920) \approx 1 - \Phi\left(\frac{1920-1600}{400}\right)$$

$$= 1 - \Phi(0.8) = 0.2119$$

上页 下页

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，公司付给受益人1万元。设老年人死亡率为0.017，试求保险公司在一年内这项保险亏本的概率。

解：设X为一年中投保老人的死亡数，则$X \sim b(n,p), n=10000, p=0.017$

由德莫佛--拉普拉斯中心极限定理，保险公司亏本的概率为：

$$P(10000X > 10000\times 200)$$

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

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

思考题：

求保险公司至少盈利10万元的概率。

答案：0. 937