第5.1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

概率论与数理统计是研究随机现象统计规律性的学科．随机现象的规律性只有在相同的条件下进行大量重复试验时才会呈现出来．也就是说，要从随机现象中去寻求必然的法则，应该研究大量随机现象.

研究大量的随机现象，常常采用极限形式，由此导致对极限定理进行研究．极限定理的内容很广泛，其中最重要的有两种：

大数定律 与 **中心极限定理**

下面我们先介绍大数定律

切比雪夫不等式

定理　设随机变量 X 具有数学期望 $E(X)=\mu$, 方差 $D(X)=\sigma^2$,则对于任意正数 ε, 不等式

$$P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}$$

成立.

切比雪夫不等式

证明　取连续型随机变量的情况来证明.

设 X 的概率密度为 $f(x)$,则有

$$P\{|X-\mu| \geq \varepsilon\} = \int_{|x-\mu|\geq\varepsilon} f(x)\,\mathrm{d}x$$

$$\leq \int_{|x-\mu|\geq\varepsilon} \frac{|x-\mu|^2}{\varepsilon^2} f(x)\,\mathrm{d}x$$

$$\leq \frac{1}{\varepsilon^2}\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,\mathrm{d}x = \frac{1}{\varepsilon^2}\sigma^2.$$

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

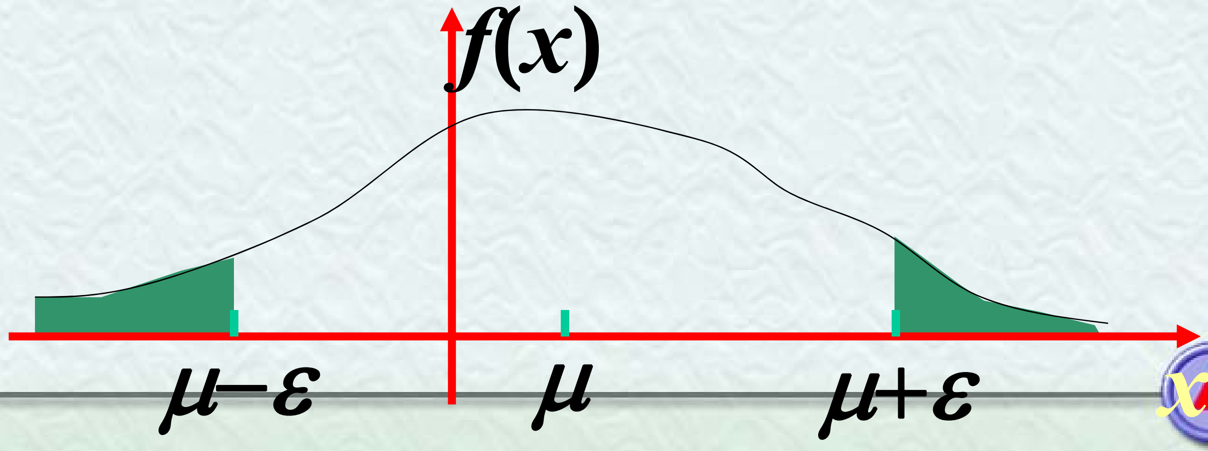
$$P\{|X-\mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2} \Leftrightarrow P\{|X-\mu| < \varepsilon\} \geq 1-\frac{\sigma^2}{\varepsilon^2}.$$

此不等式也可写为:

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

由切比雪夫不等式可以看出，若 σ^2 越小，则事件{|X-E(X)|< ε }的概率越大，即随机变量X 集中在期望附近的可能性越大.

这个不等式给出了, 在随机变量X的分布未知的情况下事件$\{|X-\mu|<\varepsilon\}$的概率的下限估计.



例1　已知正常男性成人血液中，每一毫升白细胞数平均是7300，均方差是700．利用切比雪夫不等式估计每毫升白细胞数在5200~9400之间的概率．

例2　某地区有10000盏电灯，每晚每盏电灯开灯概率为0.7，假定灯开关相互独立，试用切比雪夫不等式估计每晚开着的灯数载6800到7200之间的概率。

大数定律的客观背景

大量的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

……

依概率收敛：

设 $X_1,\ldots,X_n,\ldots$是随机变量序列，A 是一个常数；若对任意 $\varepsilon>0$ ，有：$\lim\limits_{n\to\infty}P\{|X_n-A|<\varepsilon\}=1$ ，则称 $X_1,\ldots,X_n\cdots$ 依概率收敛于 A，记为$X_n\xrightarrow{P}A$ 。

注：依概率收敛表示当n无限增大时，对任意事件$\{|X_n-A|<\varepsilon\}$的概率无限接近于1.

请注意：

$\{X_n\}$依概率收敛于A，意味着对任意给定的$\varepsilon > 0$，当n充分大时，事件$|X_n - A| < \varepsilon$的概率很大，接近于1；并不排除事件$|X_n - A| \geq \varepsilon$的发生，而只是说他发生的可能性很小.

依概率收敛比高等数学中的普通意义下的收敛弱些，它具有某种不确定性.

定理5.1（切比雪夫大数定理）

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 是相互独立的随机变量序列，l是正常数，且$E(X_i) = \mu_i < +\infty$, $D(X_i) \le l < +\infty, (i = 1, 2, \ldots)$,则对于任意正数$\varepsilon$有

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

即$\lim\limits_{n\to\infty} P\{|\bar{X} - \frac{1}{n}\sum\limits_{i=1}^{n}\mu_i| \ge \varepsilon\}$

$$= \lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| \ge \varepsilon\right\} = 0.$$

$\{|\overline{X}-\frac{1}{n}\sum_{i=1}^{n}E(X_i)|\geq\varepsilon\}$是一个随机事件,等式表明,当$n\to\infty$时这个事件的概率趋于0,

即对于任意正数ε,当n充分大时,不等式$|\overline{X}-\frac{1}{n}\sum_{i=1}^{n}E(X_i)|\geq\varepsilon$成立的概率很小.

关于定理5.1的说明:

当 n 很大时,随机变量 $X_1, X_2, \cdots, X_n$ 的算术平均 $\frac{1}{n}\sum_{i=1}^{n} X_i$ 接近于它们的数学期望 的算术平均值 $\frac{1}{n}\sum_{i=1}^{n} E(X_i)$.

（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5.1的另一种叙述:

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

l是正常数，且$E(X_i) = \mu_i < +\infty$,

$D(X_i) \le l < +\infty, (i = 1,2,\ldots)$,

序列 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 依概率收敛于 $\frac{1}{n}\sum_{i=1}^{n} E(X_i)$,

即 $\overline{X} \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} E(X_i)$.

定理5.2（辛钦定理）

辛钦资料

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

则对于任意正数 ε，有 $\lim\limits_{n\to\infty} P\left\{\left|\frac{1}{n}\sum\limits_{k=1}^{n} X_k - \mu\right| \geq \varepsilon\right\} = 0.$

即 $\lim\limits_{n\to\infty} P\{|\overline{X}-\mu|<\varepsilon\}$

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

说明

1、定理中$\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|<\varepsilon\}$是指一个随机事件，当$n\to\infty$时，这个事件的概率趋于1.

2、定理以数学形式证明了随机变量$X_1,\cdots X_n$的算术平均$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$接近数学期望E（$X_k$）$=\mu$（$k=1,2\cdots,n$），这种接近说明其具有的稳定性.

这种稳定性的含义说明算术平均值是依概率收敛的意义下逼近某一常数.

问题：

设n_A是n重贝努里试验中事件A发生的次数，p是事件A发生的概率， $\frac{n_A}{n}$ **是事件A发生的频率.**

事件发生的频率能否代替事件的概率，频率是否具有稳定性呢？

定理5.3（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

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

证明 引入随机变量

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

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 $E(X_k) = p,\quad D(X_k) = p(1-p) \le \frac{1}{4},\quad k = 1,2,\cdots$

根据定理**4.1**有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - p\right| \ge \varepsilon\right\} = 0,$$

即 $$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - p\right| \ge \varepsilon\right\} = 0.$$ [证毕]

关于贝努利定理的说明:

贝努利定理表明事件发生的频率 $\frac{\mu_A}{n}$ 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.

小概率事件的实际不可能原理（小概率原理）：

在实际中概率很小的随机事件在个别试验中几乎不可能发生。

因此常忽略小概率事件发生的可能性。

但这个“小”是相对的，要视实际问题的要求和性质而定。

三、典型例题

例1 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，

具有如下分布律：

X_n	$-na$	0	na
P	$\frac{1}{2n^2}$	$1-\frac{1}{n^2}$	$\frac{1}{2n^2}$

问是否满足切比雪夫定理？

解 独立性依题意可知，检验是否具有数学期望？

$$E(X_n) = -\mathbf{na} \cdot \frac{1}{2n^2} + 0 \cdot \left(1 - \frac{1}{n^2}\right) + \mathbf{na} \cdot \frac{1}{2n^2} = 0,$$

说明每一个随机变量都有数学期望,

检验是否具有有限方差?

$\because$

X_n^2	$(na)^2$	0	$(na)^2$
P	$\dfrac{1}{2n^2}$	$1-\dfrac{1}{n^2}$	$\dfrac{1}{2n^2}$

$$\therefore E(X_n^2)=2(na)^2\cdot\frac{1}{2n^2}=a^2$$

$$\therefore D(X_n)=E(X_n^2)-[E(X_n)]^2=a^2,$$

说明离散型随机变量有有限方差,

故满足切比雪夫定理的条件.

例2 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，且 $E(X_k)=0$, $D(X_k)=\sigma^2, k=1,2,\cdots$, 证明对任意正数 ε 有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k^2-\sigma^2\right|<\varepsilon\right\}=1.$$

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

所以 $X_1^2, X_2^2, \cdots, X_n^2, \cdots$ 也是相互独立的，

由 $E(X_k)=0$, 得 $E(X_k^2)=D(X_k)+[E(X_k)]^2=\sigma^2$,

由**辛钦定理**知

对于任意正数 ε, 有 $\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k^2-\sigma^2\right|<\varepsilon\right\}=1.$

例3 设$X_1, X_2, \cdots, X_n$是独立同分布的随机变 量序列，

$E(X_i)=\mu, D(X_i)=\sigma^2$均存在，证明

$$Y_n = \frac{2}{n(n+1)}\sum_{i=1}^{2} iX_i$$

依概率收敛到μ.

证明　因为 $\mathbf{E(Y_n) = E\{\frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i\}}$

$$\mathbf{= \frac{2}{n(n+1)}\sum_{i=1}^{n} iE(X_i)}$$

$$= \frac{2\mu}{n(n+1)} \sum_{i=1}^{n} i = \mu$$

$$D(Y_n) = \frac{4}{n^2(n+1)^2} \sum_{i=1}^{n} i^2 D(X_i) = \frac{4\sigma^2}{n^2(n+1)^2} \sum_{i=1}^{n} i^2$$

$$= \frac{4n(n+1)(2n+1)\sigma^2}{6n^2(n+1)^2} = \frac{2(2n+1)\sigma^2}{3n(n+1)}$$

从而对任意给定的$\varepsilon > 0$，由切比谢夫不等式得

$$0 \le P\{|Y_n - \mu| \ge \varepsilon\} \le \frac{D(Y_n)}{\varepsilon^2}$$

$$= \frac{2(2n+1)\sigma^2}{3n(n+1)\varepsilon^2} \to 0 \qquad (n \to \infty)$$

因此$Y_n \xrightarrow{P} \mu$.

四、小结

三个大数定理 $\begin{cases} \text{切比雪夫大数定理} \\ \text{辛钦定理} \\ \text{伯努利大数定理} \end{cases}$

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.