第一节　大数定律

一、问题的引入

二、基本定理

三、典型例题

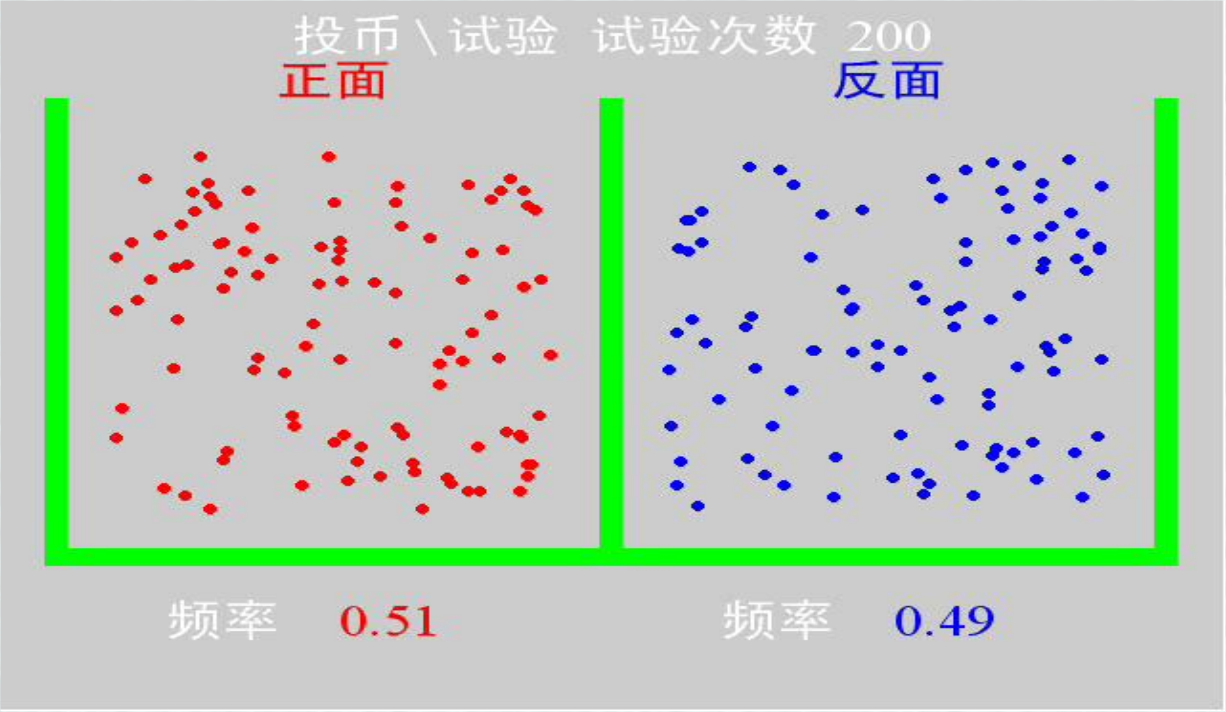
四、小结

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.

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.



启示:从实践中人们发现大量测量值的算术平均值有稳定性.

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**

1. 契比雪夫不等式

设随机变量 X 的均值 $E(X)$ 及方差 $D(X)$ 都存在，则对于任意给定的 $\varepsilon > 0$，有不等式

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{D(X)}{\varepsilon^2}$$

或

$$P\{|X - E(X)| < \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon^2}$$

证明 （仅对连续性的随机变量进行证明）

设 $f(x)$ 为 X 的密度函数，记 $E(X)=\mu,$ $D(X)=\sigma^2.$ 则

$$P\{|X-E(X)|\geq\varepsilon\}$$

$$=\int_{|x-\mu|\geq\varepsilon} f(x)dx \leq \int_{|x-\mu|\geq\varepsilon} \frac{(x-\mu)^2}{\varepsilon^2} f(x)dx$$

$$\leq \frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx \leq \frac{1}{\varepsilon^2}\times\sigma^2=\frac{D(X)}{\varepsilon^2}$$

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

由此可体会方差的概率意义：
它刻划了随机变量取值的离散程度.

契贝雪夫不等式的用途：

（1）证明大数定律；（2）估计事件的概率。

契贝雪夫不等式只利用随机变量的数学期望及方差就可对的概率分布进行估计。

$$P\{|X-E(X)|\geq\varepsilon\}\leq\frac{D(X)}{\varepsilon^2}$$

从契贝雪夫不等式还可以看出，对于给定的$\varepsilon>0$，当方差越小时，事件$\{|X-E(X)|\geq\varepsilon\}$发生的概率也越小，即$X$的取值越集中在$E(X)$附近．这进一步说明方差确实是一个描述随机变量与其期望值离散程度的一个量.

当$D(X)$已知时，契贝雪夫不等式给出了X与$E(X)$的偏差小于ε的概率的估计值.

$$P\{|X-E(X)|<\varepsilon\}\geq 1-\frac{D(X)}{\varepsilon^2}$$

例1 设随机变量 X的方差是2，则根据切比雪夫不等式估计

$P\left\{\left|X-E(X)\right|\geq 2\right\}\leq$ ____________

解

$$P\left\{\left|X-E(X)\right|\geq 2\right\}\leq \frac{DX}{\varepsilon^2}=\frac{2}{2^2}=\frac{1}{2}$$

例2:　设随机变量X的数学期望为11，方差为9，则根据切比雪夫不等式估计

$$P\left\{2 < X < 20\right\} \geq \underline{\qquad\qquad\qquad}$$

解

$$P\left\{2 < X < 20\right\}$$

$$= P\left\{2-11 < X - 11 < 20-11\right\}$$

$$= P\left\{-9 < X - 11 < 9\right\}$$

$$= P\left\{\left|X - 11\right| < 9\right\} \geq 1 - \frac{DX}{\varepsilon^2} = 1 - \frac{9}{9^2} = \frac{8}{9}$$

例3:　设随机变量 X 和 Y 的数学期望分别为 -2 和 2，方差分别为 1 和 4，而 X 与 Y 的相关系数，则根据切比雪夫不等式估计 $P\{|X+Y|<6\}\geq$ ____________

解:

$$E(X+Y)=E(X)+E(Y)=0$$

$$P\left\{\left|X+Y-E(X+Y)\right|<6\right\}$$

$$=P\left\{\left|X+Y)\right|<6\right\}\geq 1-\frac{D\left(X+Y\right)}{\varepsilon^{2}}$$

$$D(X+Y)=D(X)+D(Y)+2Cov(X,Y)$$

$$=D(X)+D(Y)+2\rho_{XY}\sqrt{DX}\sqrt{DY}$$

$$=1+4+2\times\left(-0.5\right)\times 1\times 2$$

$$=3$$

$$\therefore P\left\{\left|X+Y)\right|<6\right\}\geq 1-\frac{D\left(X+Y\right)}{\varepsilon^{2}}=1-\frac{3}{6^{2}}=\frac{33}{36}=\frac{11}{12}$$

例4: 设 $E(X)=\mu$ ，$D(X)=\sigma^2$，由切比雪夫不等式可得

$$P\{|X-\mu|<2\sigma\}\geq 1-\frac{\sigma^2}{4\sigma^2}=0.75$$

$$P\{|X-\mu|<3\sigma\}\geq 1-\frac{\sigma^2}{9\sigma^2}=0.89$$

如果 $X\sim N\left(\mu,\sigma^2\right)$ **那么**

$$P\{|X-\mu|<2\sigma\}=P\left\{\left|\frac{X-\mu}{\sigma}\right|<2\right\}=\Phi(2)-\Phi(-2)=0.955$$

$$P\{|X-\mu|<3\sigma\}=P\left\{\left|\frac{X-\mu}{\sigma}\right|<3\right\}=\Phi(3)-\Phi(-3)=0.997$$

例5：n重贝努里试验中，已知每次试验事件A出现的概率为0.75，试利用契比雪夫不等式,(1)若n=7500,估计A出现的频率在0.74至0.76之间的概率至少有多大；（2）估计n,使A出现的频率在0.74至0.76之间的概率不小于0.90

解：设在n重贝努里试验中，事件A出现的次数为X，

则$X \sim B(n,0.75), E(X)=np=0.75n, D(X)=npq=0.1875n,$

又A事件的频率为：$f_n(A)=\dfrac{X}{n}$

$$(1)\ n=7500,\ P\left\{0.74<\frac{X}{n}<0.76\right\}=P\{|X-0.75n|<0.01n\}$$

$$\geq 1-\frac{0.1875n}{(0.01n)^2} \quad =0.75$$

化成切比雪夫不等式形式

$$P\left\{0.74<\frac{X}{n}<0.76\right\}=P\left\{0.74n<\frac{X}{n}\cdot n<0.76n\right\}$$

$$=P\{0.74n-0.75n<X-0.75n<0.76n-0.75n\}$$

$$=P\{-0.01n<X-0.75n<0.01n\}$$

$$=P\{|X-0.75n|<0.01n\}$$

$$\geq 1-\frac{0.1875n}{(0.01n)^2}\quad =0.75$$

（2）估计n,使A出现的频率在0.74至0.76之间的概率不小于0.90。

解：设在n重贝努里试验中，事件A出现的次数为X,

则$X \sim B(n,0.75), E(X)=np=0.75n, D(X)=npq=0.1875n,$

又A事件的频率为：$f_n(A)=\dfrac{X}{n}$

$$(2)\ P\left\{0.74<\frac{X}{n}<0.76\right\}=P\{|X-0.75n|<0.01n\}$$

$$\geq 1-\frac{0.1875n}{(0.01n)^2}=1-\frac{1875}{n}\geq 0.90 \quad \Rightarrow n\geq 18750$$

2. 经典大数定律

定理一　契比雪夫大数定理

设相互独立的随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 分别具有均值 $E(X_1), E(X_2), \cdots, E(X_n), \cdots$ 及方差 $D(X_2), \cdots, D(X_n), \cdots$ 且若存在常数C，使 $D(X_k) \le C,$ 则对于任意给定的 $\varepsilon > 0$ ，有

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

证明 由于$X_1, X_2, \cdots, X_n, \cdots$相互独立，那么对于任意的$X_1, X_2, \cdots, X_n$相互独立. 于是

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

令$Y_n=\dfrac{1}{n}\sum_{k=1}^{n}X_k$ 由契比雪夫不等式可得

$$1\ge P\{|Y_n-E(Y_n)|<\varepsilon\}\ge 1-\frac{D(Y_n)}{\varepsilon^2}\ge 1-\frac{C}{\varepsilon^2 n},$$

在上式中令 $n\to\infty$，即

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

推论

且

$D($

的

表达式的意义

$\{|\overline{X}-\mu|<\varepsilon\}$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 1，即对于任意正数 ε，当n充分大时，不等式$|\overline{X}-\mu|<\varepsilon$成立的概率很大.

数ε有

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

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 $n\to\infty$, 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 $X_1, X_2, \cdots, X_n$ 的算术平均 $\frac{1}{n}\sum_{k=1}^{n} X_k$ 接近于数学期望

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

(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.

切比雪夫大数定律表明，独立随机变量序列$\{X$ 则

$\frac{1}{n}\sum_{i=1}^{n} X_i$ 与 小的

切比雪夫大数定律给出了
平均值稳定性的科学描述

概率接近于1.

即当n充分大时，$\frac{1}{n}\sum_{i=1}^{n} X_i$差不多不再是随机的了，取值接近于其数学期望的概率接近于1.

一般地，称概率接近于1 的事件为大概率事件，而称概率接近于0 的事件为小概率事件. 在一次试验中大概率事件几乎肯定要发生，而小概率事件几乎不可能发生，这一规律称之为实际推断原理.

定义4.1 设 $Y_1, Y_2, \cdots, Y_n, \cdots$ 是一随机变量序列， a 为一常数，对于任意给定的 $\varepsilon > 0$ ，有

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

则称随机变量序列 $Y_1, Y_2, \cdots, Y_n, \cdots$ 依概率收敛于 a .

记为 $Y_n \xrightarrow{P} a.$

定理一的另一种叙述:

设随机变量 X_1, X

且具有相同的数学期望

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

依概率收敛于 μ, 即 $\overline{X}$

设$Y_1, Y_2, \cdots, Y_n$是一个随机变量序列, a是一个常数, 若对于任意正数ε有 $\lim\limits_{n\to\infty} P\{|Y_n - a| < \varepsilon\} = 1$,

则称序列$Y_1, Y_2, \cdots, Y_n$依概率收敛于a, 记为

$$Y_n \xrightarrow{P} a$$

定义4.2 设 $X_1, X_2, \cdots, X_n, \cdots$ 是一随机变量序列，$Y_n = \dfrac{1}{n}\sum_{k=1}^{n} X_k$，若存在常数列 a_n 使对于任意给定的 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P\left\{\left|Y_n - a_n\right| < \varepsilon\right\} = 1$$

则称随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 服从大数定律.

依概率收敛序列的性质:

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

又设函数 $g(x,y)$ 在点 (a,b) 连续,

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

若: $g(X_n,Y_n)=cX_n+dY_n$，c，d 为常数，则

$$cX_n+dY_n \xrightarrow{P} ca+db$$

若: $g(X_n,Y_n)=X_nY_n$，则 $X_nY_n \xrightarrow{p} ab$

若: $g(X_n,Y_n)=\dfrac{X_n}{Y_n}, Y_n \neq 0$，则 $\dfrac{X_n}{Y_n} \xrightarrow{p} \dfrac{a}{b}(b\neq 0)$

定理二（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

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

证明　　引入随机变量

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

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

根据定理一有

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

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

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.

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，这种方法就是第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。

定理三（辛钦定理）

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

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特殊情况.

辛钦大数定律为寻找随机变量的期望值提供了一条实际可行的途径.


例如要估计某地区的平均亩产量，要收割某些有代表性的地块，例如n 块．计算其平均亩产量，则当n 较大时，可用它作为整个地区平均亩产量的一个估计．

这一讲我们介绍了大数定律

大数定律以严格的数学形式表达了随机现象最根本的性质之一：

平均结果的稳定性

它是随机现象统计规律的具体表现.

大数定律在理论和实际中都有广泛的应用.

例6 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	$\dfrac{1}{2n^2}$	$1-\dfrac{1}{n^2}$	$\dfrac{1}{2n^2}$

问是否满足契比雪夫定理?

解 独立性依题意可知, 检验是否具有数学期望?

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

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,

故满足契比雪夫定理的条件.

例7 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 $\lim\limits_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k^2-\sigma^2\right|<\varepsilon\right\}=1.$

例8：设 $X \sim f(x)=\begin{cases}3x^2 & ,\quad 0<x<1\\ 0 & ,\quad 其它\end{cases}$

现对X独立观察 n 次，观察值记为 $X_1,\ldots,X_n$ ，$\forall\varepsilon>0$ **，**

如果这些观察值满足 $\lim\limits_{n\to\infty}P\left(\left|\dfrac{X_1^2+X_2^2+\ldots+X_n^2}{n}-a\right|<\varepsilon\right)=1$

求a？

解：由题意知，$X_1,X_2,\cdots,X_n$是具有独立同分布的随机变量

所以，它们的连续函数$X_1^2,X_2^2,\cdots,X_n^2$也是独立同分布的。

$\dfrac{X_1^2+X_1^2+\cdots+X_n^2}{n}$是变量序列$X_1^2$，$X_1^2,\cdots$，$X_n^2$的前$n$个算术平均，

故由定理三（辛钦定理）得：算术平均依概率收敛于$E(X^2)$

$$\therefore\ a=E(X^2)=\int_0^1 x^2 3x^2dx=\frac{3}{5}$$

四、小结

三个大数定理
- 契比雪夫定理的特殊情况
- 伯努利大数定理
- 辛钦定理

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.