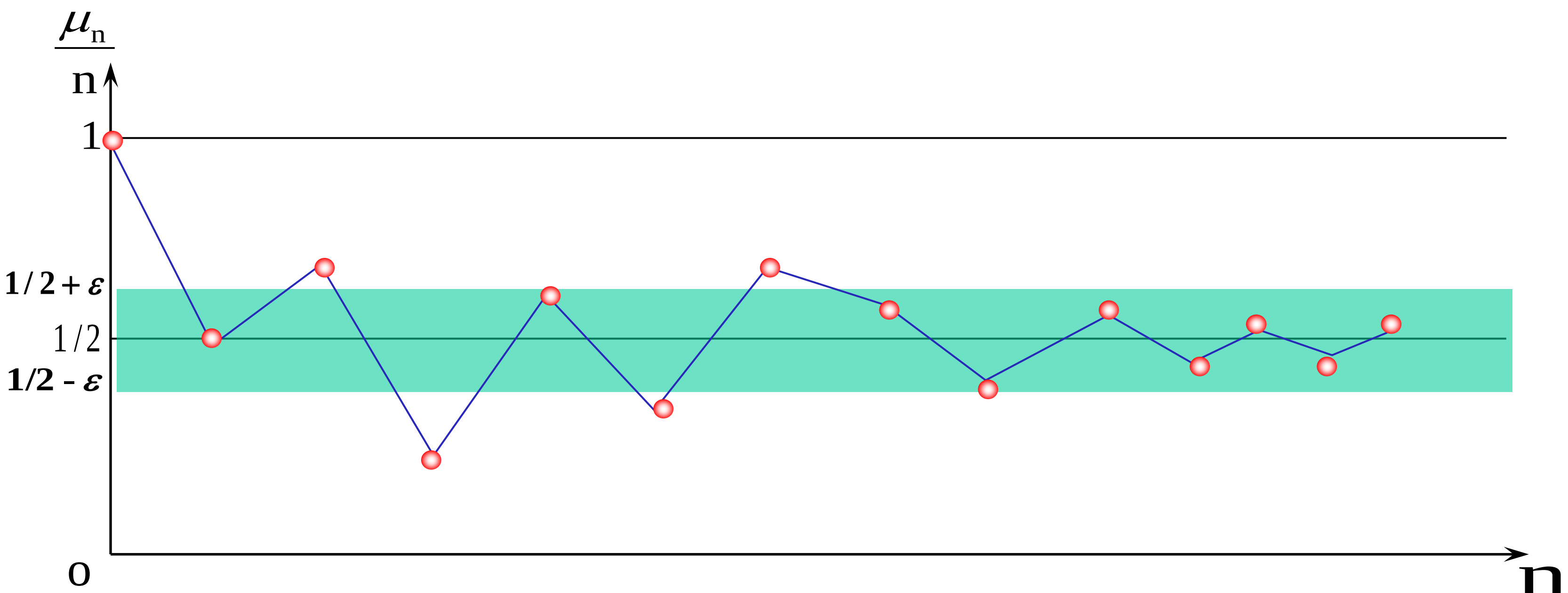
频率的稳定性

试验者	抛掷次数 n	正面次数 μ_n	频率 $\frac{\mu_n}{n}$
德·摩根	2048	1061	0.5181
浦丰	4040	2048	0.5069
费勒	10000	4979	0.4979
皮尔逊	12000	6019	0.5016
杰万斯	20480	10379	0.5068
皮尔逊	24000	12012	0.5005
罗曼洛夫斯基	80640	39699	0.4923

频率在某常数附近波动。

频率的稳定性

频率在概率值附近波动？并逐渐稳定于概率值？



$$\cancel{\lim_{n\to\infty}\frac{\mu_n}{n}=\frac{1}{2}} \Leftrightarrow \forall\varepsilon>0,\exists N(\varepsilon)>0, n\geq N(\varepsilon), \left|\frac{\mu_n}{n}-\frac{1}{2}\right|\leq\varepsilon.$$

不成立

$$P\left\{\left|\frac{\mu_n}{n}-\frac{1}{2}\right|\geq\frac{1}{2}\right\}=P\left\{\frac{\mu_n}{n}=0\right\}+P\left\{\frac{\mu_n}{n}=1\right\}=\frac{1}{2^{n-1}}\to 0, n\to\infty.$$

$$P\left\{\left|\frac{\mu_n}{n}-\frac{1}{2}\right|\geq\varepsilon\right\}\to 0, n\to\infty.$$ ？

伯努利大数定律

Bernouli's Law of Large Numbers

数学与统计学院 刘显明

Bernouli大数定律

设μ_n为n重贝努利试验中A发生的次数，

$P(A) = p$，则对任何$\varepsilon>0$，有

$$\lim_{n\to\infty} P\{|\frac{\mu_n}{n} - p| \geq \varepsilon\} = 0.$$

Jacob Bernoulli

(1647-1705)

证明解析：对任何$\varepsilon>0$

$$0 \leq P\{|\frac{\mu_n}{n} - p| \geq \varepsilon\} \leq a(n) \to 0, n \to \infty.$$

Bernouli大数定律

$$0 \le P\{|\frac{\mu_n}{n} - p| \ge \varepsilon\} \le a(n) \to 0, n \to \infty.$$

如何构造 $a(n)$?

Chebyshev不等式

$$\forall \varepsilon > 0, P\{|X - EX| \ge \varepsilon\} \le \frac{DX}{\varepsilon^2}.$$

$$E(\frac{\mu_n}{n}) = p \qquad D(\frac{\mu_n}{n}) \to 0, n \to \infty.$$

$$P\{|\frac{\mu_n}{n} - E(\frac{\mu_n}{n})| \ge \varepsilon\} \le \frac{D(\frac{\mu_n}{n})}{\varepsilon^2} \to 0, n \to \infty.$$

Bernouli大数定律

设μ_n为n重贝努利试验中 A发生的次数，　$P(A)=p$,

则对任何$\varepsilon>0$，有 $\lim\limits_{n\to\infty} P\{|\frac{\mu_n}{n}-p|\geq\varepsilon\}=0.$

证明：$\mu_n \sim B(n,p), E\mu_n=np, D\mu_n=np(1-p).$

$$E(\frac{\mu_n}{n})=p, D(\frac{\mu_n}{n})=\frac{p(1-p)}{n}.$$

由Chebyshev不等式得对任何$\varepsilon>0$，

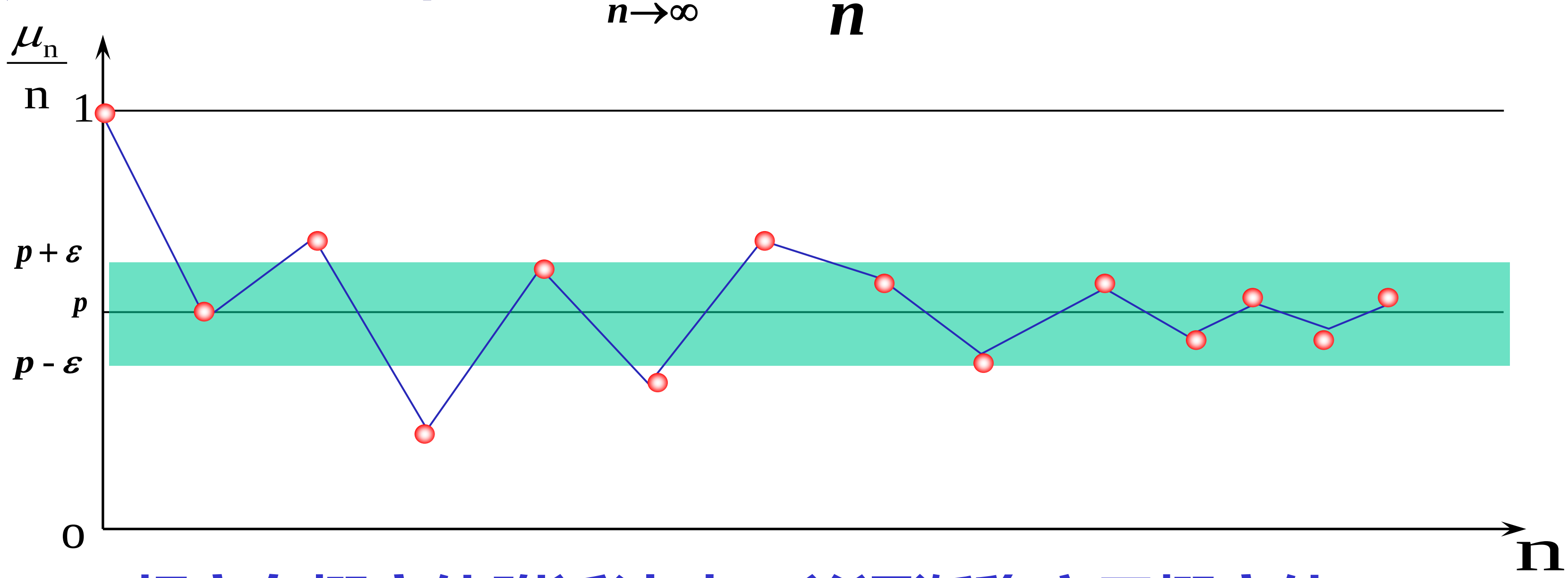
$$0\leq P\{|\frac{\mu_n}{n}-p|\geq\varepsilon\}\leq\frac{D(\frac{\mu_n}{n})}{\varepsilon^2}=\frac{p(1-p)}{n\varepsilon^2}\to 0, n\to\infty.$$

由夹挤原理得：$\lim\limits_{n\to\infty} P\{|\frac{\mu_n}{n}-p|\geq\varepsilon\}=0.$

Bernouli大数定律

设μ_n为n重贝努利试验中 A 发生的次数， $P(A)=p$,

则对任何$\varepsilon>0$，有 $\lim\limits_{n\to\infty} P\{|\frac{\mu_n}{n}-p|\geq\varepsilon\}=0$



频率在概率值附近波动，并逐渐稳定于概率值。

$\lim\limits_{n\to\infty} P\{|\frac{\mu_n}{n}-p|<\varepsilon\}=1$ $\Rightarrow$ 当n足够大，$\frac{\mu_n}{n}\approx p$

总结

$$\forall\varepsilon>0,\lim_{n\to\infty}P\{|\frac{\mu_n}{n}-p|\geq\varepsilon\}=0.$$

- **准确刻画了“频率稳定性”。**

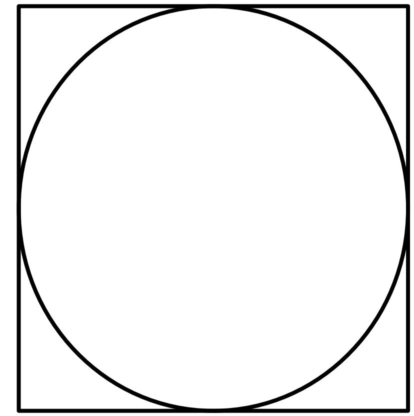
  是数理统计中参数估计的重要理论依据之一。

- **提供了试验（频率）近似事件概率的方法。**

  是Monte Carlo 方法主要的理论依据。

Monte Carlo 方法模拟圆周率

向单位正方形及内切圆均匀投点n次。
记录落在圆内点的个数为I。



$$\lim_{n\to\infty} P\{|\frac{I}{n} - \frac{\pi}{4}| \geq \varepsilon\} = 0.$$ 当n足够大 $$\frac{4I}{n} \approx \pi.$$

投点总数n	圆内点数I	Pi的近似值	相对误差*
1000	811	3.24400000	3.2597%
8000	6292	3.14600000	0.1403%
100000	78723	3.14892000	0.2332%
1000000	785009	3.14003600	-0.0495%
10000000	7854335	3.14173400	0.0045%
500000000	392689724	3.14151779	-0.0024%
2000000000	1570795484	3.14159097	-0.0001%

独立同分布大数定律

•记 $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ •若 $\forall \varepsilon > 0$ $\lim_{n\to\infty} P(|\overline{X}_n - E(\overline{X}_n)| < \varepsilon) = 1$

•则称 $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n, \ldots$服从大数定律。

设$\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n, \ldots$相互独立，且$E(\boldsymbol{X}_n)=\mu, D(\boldsymbol{X}_n)=\sigma^2$存在，则

$$E(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu \text{，} \quad D(\overline{X}_n) = \frac{1}{n^2} D(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{\sigma^2}{n}$$

$$1 \geq P(|\overline{X}_n - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2/n}{\varepsilon^2} \xrightarrow[n\to\infty]{} 1$$

几个著名的大数定律

•名　称	条　　件	结　　论
•马尔科夫	$\lim\limits_{n\to\infty}\frac{1}{n^2}D(\sum\limits_{i=1}^{n}X_i)=0$	$\lim\limits_{n\to\infty}P(\lvert\bar{X}_n-E(\bar{X}_n)\rvert<\varepsilon)=1$
•切比雪夫	•$Cov(X_i, X_j)=0$，i≠j，且$D(X_n)<C$（有界）	$\lim\limits_{n\to\infty}P(\lvert\bar{X}_n-E(\bar{X}_n)\rvert<\varepsilon)=1$
•伯 努 利	•$\mu_n\sim B(n,p)$	$\lim\limits_{n\to\infty}P(\left\lvert\frac{\mu_n}{n}-p\right\rvert<\varepsilon)=1$
•辛　钦	•X_1，$X_2,\ldots,X_n,\ldots$ 独立同分布，且 $E(X_n)=\mu$ (有限)	$\lim\limits_{n\to\infty}P(\lvert\bar{X}_n-\mu\rvert<\varepsilon)=1$

•$\{\bar{X}_n\}$依概率收敛于 μ

三、中心极限定理

1、独立同分布的中心极限定理

设$\{X_n, n=1,2,\ldots\}$是独立同分布的随机变量序列，且

$E(X_i)=\mu$，$D(X_i)=\sigma^2$，$i=1,2,\ldots$，设标准化随机变量

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

•的分布函数为$F_n(x)$，则

$$\lim_{n\to\infty}F_n(x)=\Phi(x)$$

•称 $\boldsymbol{Y}_n$ 依分布收敛于 $\boldsymbol{X}\sim N(0,1)$.

2、德莫佛·拉普拉斯（De Moivre-Laplace）中心极限定理

•设 $Y_n \sim B(n, p)$，$n=1,2,\ldots$，则

$$\lim_{n\to\infty} P\left(\frac{Y_n - np}{\sqrt{np(1-p)}} \le x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \qquad -\infty < x < +\infty$$

•**证明**：•取 $X_i \sim B(1, p)$，$i=1,2,\ldots$，相互独立，则 $Y_n = \sum_{i=1}^{n} X_i$

•而 $E(X_i)=p$，$D(X_i)=p(1-p)$，由独立同分布的中心极限定理

•即得。

•**应用**： •$Y_n \sim B(n, p)$，当 n 充分大时

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

例0： 对一个学生而言，来参加家长会的家长人数是一个随机变量。设一个学生无家长、1名家长、2名家长来参加会议的概率分别是0.05、0.8、0.15. 若学校共有400名学生，设各学生参加会议的家长数相互独立，且服从同一分布。（1）求家长数X超过450的概率。（2）求恰有一名家长参加会议的学生数Y不多于340的概率。

解 （1）$P(\mathrm{X}>450)=P(\frac{X-400\times 1.1}{\sqrt{400\times 0.19}}>\frac{450-400\times 1.1}{\sqrt{400\times 0.19}})$

$$\approx 1\text{-}\Phi(1.147)=0.1357.$$

（2）$P(\mathrm{Y}\le 340)=P(\frac{\mathrm{Y}-400\times 0.8}{\sqrt{400\times 0.8\times 0.2}}>\frac{340-400\times 0.8}{\sqrt{400\times 0.8\times 0.2}})$

$$\approx\Phi(2.5)=0.9938.$$

例1 已知一批同型号的电子元件，次品率为1/6，试以99%的把握断定：从这批电子元件中任取6000只，其中次品所占的比例与1/6之差的绝对值不超过多少？这时6000电子元件中，次品数又落在一个什么范围内？

解 记X为6000只电子元件中的次品数，则

$X \sim B(6000, 1/6)$，要求 ε 使

$$0.99 = P\left(\left|\frac{X}{6000} - \frac{1}{6}\right| \le \varepsilon\right) = P\left(\left|\frac{X - 6000 \times \frac{1}{6}}{\sqrt{6000 \times \frac{1}{6} \times \frac{5}{6}}}\right| \le \frac{\varepsilon 6000}{\sqrt{6000 \times \frac{1}{6} \times \frac{5}{6}}}\right)$$

$$\approx 2\Phi\left(\varepsilon \frac{6}{\sqrt{5}} \sqrt{6000}\right) - 1 \Rightarrow \Phi\left(\varepsilon \times 60\sqrt{\frac{60}{5}}\right) = \frac{1.99}{2} = 0.995$$

$$\Rightarrow \varepsilon = 2.576 \times \frac{1}{60\sqrt{12}} = 0.01239$$

$$\left|\frac{X}{6000} - \frac{1}{6}\right| < 0.01239$$

$$926 < X < 1074$$

例2 设有某天文学家试图观测某星球与他所在天文台的距离D，他计划作出 n 次独立的观测$X_1, X_2, \ldots, X_n$ (单位：光年)，设这 n 次独立的观测的期望 EX_i=D，方差DX_i=4，i =1,2,…,n，现天文学家用 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ 作为D的估计，为使对D的估计的精度在±0.25光年之间的概率大于0.98，问这位天文学家至少要作出多少次独立的观测？

解　当n充分大时　$\dfrac{\bar{X}-D}{\sqrt{4/n}} \sim N(0,1)$

$$0.98 < P(\left|\bar{X}-D\right| \le 0.25) = P\left(\left|\frac{\bar{X}-D}{2/\sqrt{n}}\right| \le \frac{0.25}{2}\sqrt{n}\right) \approx 2\Phi\left(\frac{0.25}{2}\sqrt{n}\right)-1$$

$$\Phi\left(\frac{0.25}{2}\sqrt{n}\right) > \frac{1.98}{2} = 0.99 \Rightarrow 0.125\sqrt{n} > 2.3264 \quad \Rightarrow n > 346.376765$$

故这位天文学家至少要作出347次独立的观测。