西北工業大學

NORTHWESTERN POLYTECHNICAL UNIVERSITY

概率论与数理统计

徐 爽

西北工业大学理学院概率统计教研室

第二节 中心极限定理

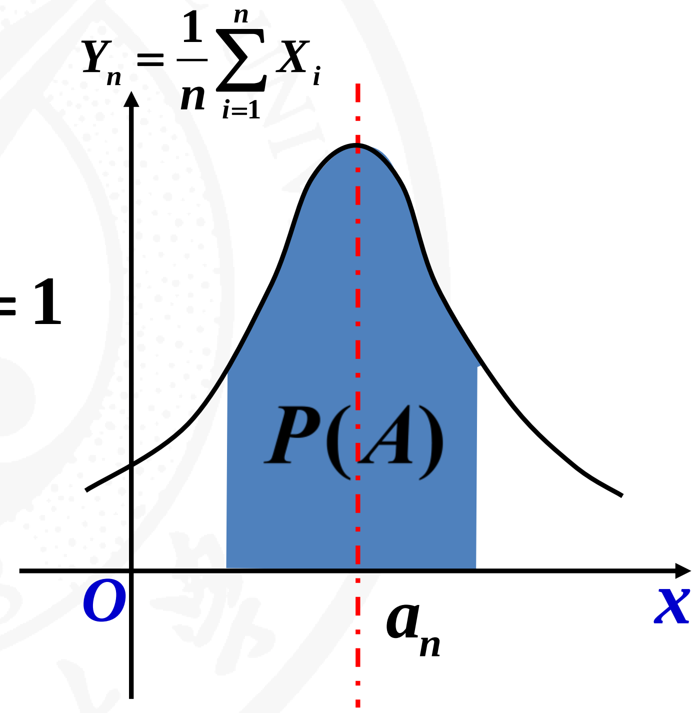
- 一、问题的提出
- 二、中心极限定理

一、问题的提出

大数定律： 满足一定条件的随机变量序列的

算数平均值依概率收敛。

$$\lim_{n\to\infty} P\left\{\underbrace{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon}_{A}\right\} = 1$$



$$\boxed{Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} E(Y_n)}$$

若$X_i \sim N(2,3)$，　　则$\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i \xrightarrow{p} 2$

X_1,	X_2,	X_3,	X_4,	X_5,	X_6,	X_7,	X_8,	X_9,	X_{10},	$\overline{X}$
4.5211	-4.4151	10.7240	0.9385	2.0687	3.5602	1.1187	-1.9960	-2.0851	1.4143	2.1200
-0.6641	-0.5188	4.4757	-0.4708	1.2140	1.9399	-0.5438	-4.9896	3.3651	1.3472	1.6740
2.3003	6.0638	6.1369	-2.7312	-3.2506	1.8957	-1.3604	-2.3473	-0.5461	1.0907	1.8494
0.3664	-1.2165	-1.1745	3.5239	1.1430	-0.3945	9.5780	3.0005	0.9953	2.0691	1.6199
2.9106	4.8829	0.5942	2.8460	-0.4941	5.0561	6.9665	3.1741	3.6584	2.1539	2.0553
0.1990	2.3721	1.1826	2.1004	-0.9376	1.6003	2.9226	3.3550	5.1173	4.4782	1.6360
3.4699	6.3101	5.2953	-2.0010	-1.4692	-0.1436	-1.7714	1.6091	-1.3529	6.5809	1.9810
4.2181	-3.8827	1.1664	5.3825	0.3993	6.0542	-0.5964	2.5511	5.7820	3.4007	1.8734
7.1357	1.4069	4.1046	3.0505	-4.0079	1.3257	1.4704	0.5715	3.9804	1.3709	2.1550
1.4176	-1.6235	-4.1554	1.1028	4.8927	0.2329	4.3742	4.5861	1.7964	3.8756	1.9405

但我们无法得知其收敛的速度，即：　给定n，$P\left\{\left|\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i-2\right|<\varepsilon\right\}=?$

$$Y_n=\frac{1}{n}\sum_{i=1}^{n} X_i \sim F(x)?$$

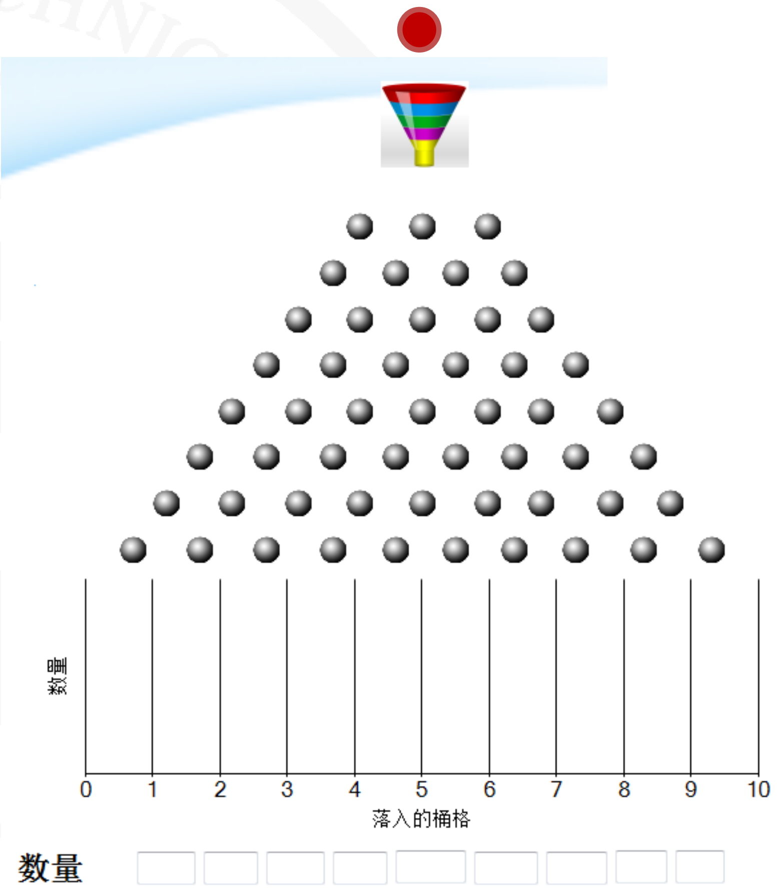
引例 高尔顿(Galton)板实验

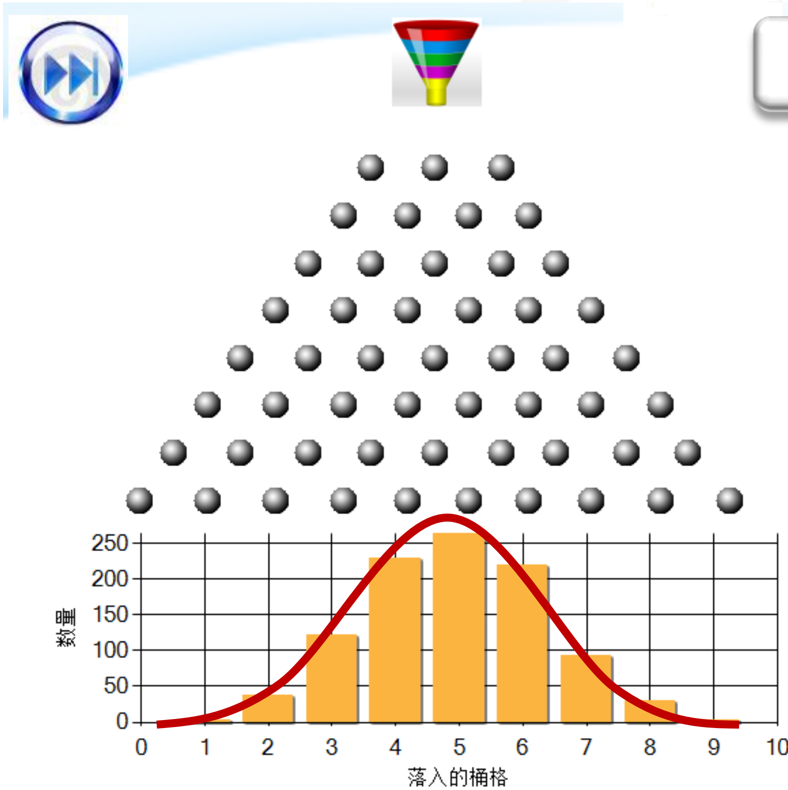
*n*层钉子

$P_{左}=P_{右}=1/2$

每个格子中小球的个数

统计特性





数量	1	38	122	230	263	219	94	30	3

分析

令$X_k = \begin{cases} 1 & 第k层向左 \\ -1 & 第k层向右 \end{cases}$

则$X_k \sim B(1,\dfrac{1}{2}),\ \ k=1,2,\cdots n,$

且$X_1, X_2, \cdots, X_n$相互独立。

小球下落到最底层的水平位置

$$Y_n = \sum_{k=1}^{n} X_k \sim B(n,\frac{1}{2})$$

$$\sim N(\mu,\sigma^2)$$

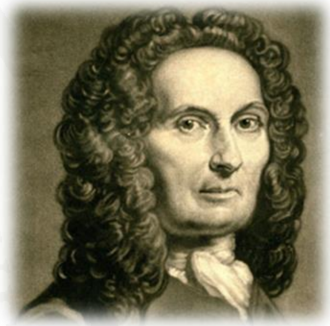
二、中心极限定理

定理4.7　棣莫佛-拉普拉斯(*De Movire - Laplace*)中心极限定理

设随机变量Y_n服从二项分布$B(n, p)$,则对任意x，

有

$$\lim_{n\to\infty} P\left\{\boxed{\frac{Y_n - np}{\sqrt{np(1-p)}} \le x}\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t = \Phi(x)$$



A. De Moivre (1667-1754)

表 明　若$Y_n \sim B(n,p)$,

$$E(Y_n) = np$$
$$D(Y_n) = np(1-p)$$

则有$Y_n^* = \dfrac{Y_n - np}{\sqrt{np(1-p)}} \sim AN(0,1)$

P.-S. Laplace (1749-1827)

近似服从标准正态分布

正态分布是二项分布的极限分布

$np \geq 5 \,\&\, n(1-p) \geq 5$

$n \to \infty$

正态分布

$N(\mu, \sigma^2)$

二项分布

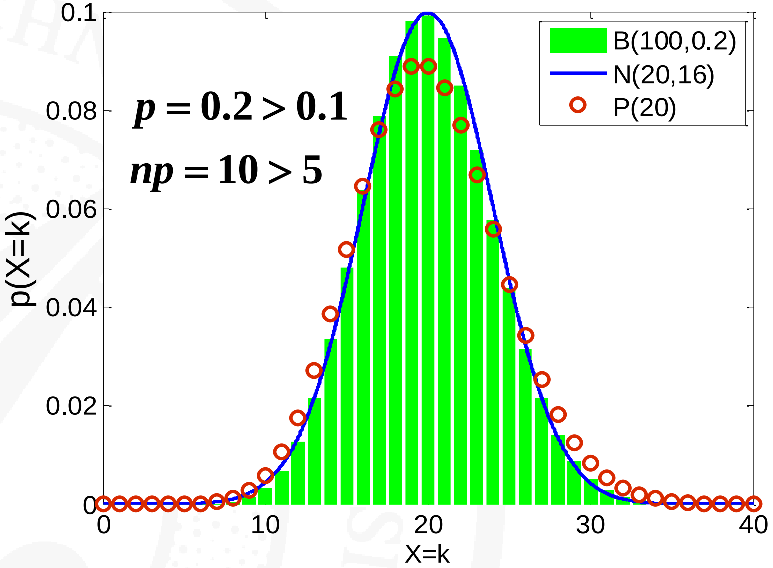
$B(n, p)$

$p \leq 0.1, np \to \lambda < 5$

$n \to \infty$

泊松分布

$P(\lambda)$



$$X \sim B(100, 0.2)$$

$$P(X \leq b) = \sum_{k=0}^{b} C_{100}^{k} 0.2^{k} 0.8^{100-k}$$

$$\approx \Phi\left(\frac{b-20}{\sqrt{16}}\right)$$

中心极限定理

De Moivre-Laplace中心极限定理

$$Y_n \sim B(n,p) \Longrightarrow \frac{Y_n - np}{\sqrt{np(1-p)}} \sim AN(0,1)$$

$$Y_n = \sum_{i=1}^{n} X_i \qquad \frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \sim AN(0,1)$$

其中$X_i \sim B(1,p)$

特殊 $X_i \sim B(1,p)$ ⇒ 一般$X_i \sim F(x)$?

定理4.6 林德伯格-列维(*Lindeberg-Levy*)中心极限定理

设随机变量$X_1, X_2, \ldots, X_n$相互独立，服从同一分布，

且存在

$$E(X_i)=\mu,\quad D(X_i)=\sigma^2\neq 0\quad (i=1, 2, \ldots, n)$$

则对任意x,有

$$\lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}\leq x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=\Phi(x)$$

（证明略）

表明 $E(X_i)=\mu, D(X_i)=\sigma^2$ 独立同分布

$$\Rightarrow E(\sum_{i=1}^{n} X_i)=n\mu, D(\sum_{i=1}^{n} X_i)=n\sigma^2$$

大量相互<u>独立同分布</u>的随机变量的和

标准化后近似服从标

X_i之间没有相互依赖关系

对总和的影响在概率意义下均衡

$$Y_n^*=\frac{\sum_{i=1}^{n} X_i-n\mu}{\sqrt{n}\sigma}\sim AN(0,1)$$

$$\sum_{i=1}^{n} X_i \sim AN\left(n\mu, n\sigma^2\right)$$

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n} X_i \sim AN\left(\mu, \frac{\sigma^2}{n}\right)$$

例2　一船舶在某海区航行, 已知每遭受一次海浪的冲击, 纵摇角大于 3° 的概率为1/3, 若船舶遭受了90000次波浪冲击, 问其中有29500～30500次纵摇角大于 3° 的概率是多少?

解　设A:纵摇角大于 3°，则P(*A*)=1/3

假设海浪冲击相互独立。

X：90000次波浪冲击中纵摇角大于 3° 的次数

则 $X \sim B(90000, 1/3)$. 分布律为

$$P\{X=k\}=\binom{90000}{k}\left(\frac{1}{3}\right)^{k}\left(\frac{2}{3}\right)^{90000-k}, k=1,\cdots,90000.$$

所求概率为 $P\{29500 < X \leq 30500\}$

$$= \sum_{k=29501}^{30500} \binom{90000}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{90000-k}$$

直接计算很麻烦，利用**切比雪夫不等式**

$$\because n = 90000, \quad p = \frac{1}{3},$$

$P\{29500 < X \leq 30500\}$

$$= P\{29500 - np < X - np \leq 30500 - np\}$$

$$= P\left\{\left|X - 30000\right| \leq 500\right\} \geq 1 - \frac{npq}{500^2} \approx 0.92$$

只能估计范围，利用**棣莫佛-拉普拉斯定理**

$$P\{29500 < X \le 30500\}$$

$$= P\left\{\frac{29500 - np}{\sqrt{np(1-p)}} < \frac{X - np}{\sqrt{np(1-p)}} \le \frac{30500 - np}{\sqrt{np(1-p)}}\right\}$$

$$= \Phi\left(\frac{30500 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{29500 - np}{\sqrt{np(1-p)}}\right)$$

$$\approx \Phi\left(\frac{5\sqrt{2}}{2}\right) - \Phi\left(-\frac{5\sqrt{2}}{2}\right)$$

$$= 2\Phi(3.53) - 1 \quad = 0.9995.$$

例3 设随机变量$X_1, X_2,\ldots, X_n$相互独立，且 X_i 在区间[–1, 1] 上服从均匀分布(i=1, 2,…, n)，试证当 n充分大时，随机变量 $Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ 近似服从正态分布并指出其分布参数.

证 记 $Y_i = X_i^2, \quad (i = 1,2,\cdots,n)$

因为$X_1, X_2,\ldots, X_n$相互独立，所以$Y_1, Y_2,\ldots,Y_n$相互独立，根据**林德伯格–列维中心极限定理**

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \sim AN\left(E(Y_i), \frac{D(Y_i)}{n}\right)$$

$X_i \sim U[-1, 1]$ $\qquad E(X_i) = 0, D(X_i) = \dfrac{1}{3}$

$$E(Y_i) = E(X_i^2) = D(X_i) + E^2(X_i) = \frac{1}{3}$$

$$D(Y_i) = E(Y_i^2) - [E(Y_i)]^2 = E(X_i^4) - [E(Y_i)]^2$$

因为 $E(X_i^4) = \int_{-1}^{1} x_i^4 \cdot \dfrac{1}{2} \mathrm{d}x_i = \dfrac{1}{5},$

所以 $D(Y_i) = \dfrac{1}{5} - \left(\dfrac{1}{3}\right)^2 = \dfrac{4}{45},$

$$\therefore Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \sim N\left(\frac{1}{3}, \frac{4}{45n}\right).$$

例5 设总体X服从参数为2的指数分布，$X_1, X_1, \cdots X_n$为来自总体X的简单随机样本，则当$n \to \infty$时，随机变量序列$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$依概率收敛于 ___$\frac{1}{2}.$___。

例6 设随机变量$X_1, X_1, \cdots X_n$相互独立，均服从$B(1,0.2)$，则由中心极限定理知，当$n \to \infty$时，随机变量$U = \sum_{i=1}^{n} X_i$近似服从 ___$N(0.2n, 0.16n)$___ 分布.（写出分布参数）.

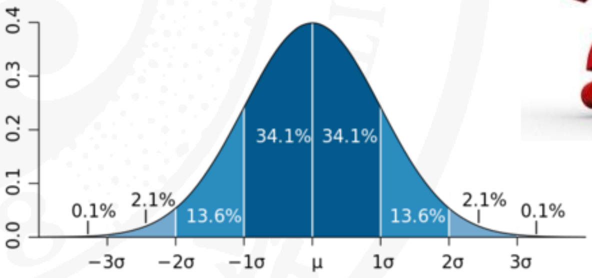
课后思考

中心极限定理

独立同分布

独立不同分布

$$\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} \text{正态分布}$$



内容小结

$E(X_i)=\mu \quad D(X_i)=\sigma^2$

中心极限定理

林德贝格－列维中心极限定理

独立同分布情形

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \sim AN(0,1)$$

李雅普诺夫定理

独立不同分布情形

$$Y_n^* = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{B_n} \sim AN(0,1)$$

棣莫佛－拉普拉斯定理

二项分布的正态近似

$$Y_n = \sum_{k=1}^{n} X_k \sim B(n,p)$$

$$\frac{Y_n - np}{\sqrt{np(1-p)}} \sim AN(0,1)$$

4-2 中心极限定理

Thank You !