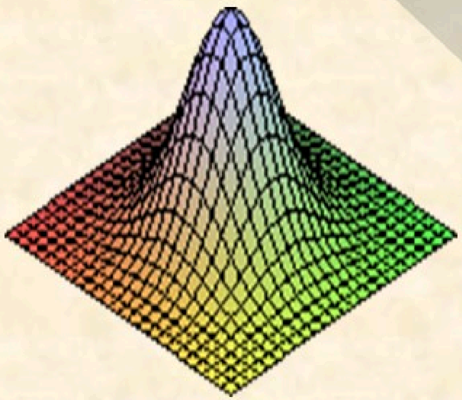
概率论与数理统计

制作人：叶鹰　吴娟

主讲人：吴 娟

wujuan@hust.edu.cn



第五章　随机变量的极限定理

一、问题的提出

1）X_1，X_2，…，X_n为多次“测量”，则

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{} \mu$$ （真值）合理吗?

2）$X_1 + X_2 + \cdots + X_n \sim$?

$$X_1 + X_2 + \cdots + X_n \sim \ ?\ (n\to\infty)$$

二、切比雪夫чебышев不等式

设**R.V.*X*** 有 $E(X)=\mu$，$D(X)=\sigma^2$，则对任何 $\varepsilon>0$ 有

$$P(|X-\mu|\ge\varepsilon)\le\frac{\sigma^2}{\varepsilon^2}$$

或

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

证　仅对 C.R.V.X 证明，设 $f(x)$ 为 X 的密度函数，则

$$P(|X-\mu|\ge\varepsilon)=\int_{|x-\mu|\ge\varepsilon} f(x)\mathrm{d}x \le \int_{|x-\mu|\ge\varepsilon}\frac{(x-\mu)^2}{\varepsilon^2}f(x)\mathrm{d}x$$

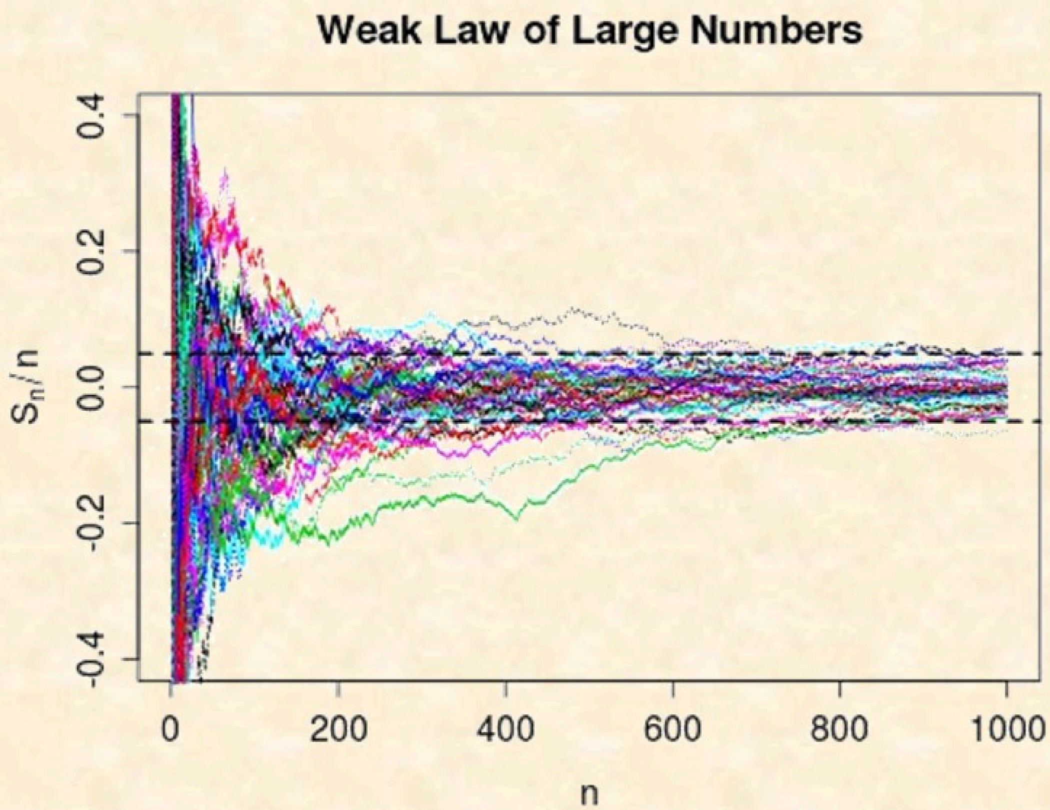
$$\le\frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)\mathrm{d}x=\frac{\sigma^2}{\varepsilon^2}$$

三、大数定律（0-1律）

记 $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ 若 $\forall \varepsilon > 0$

$$\lim_{n\to\infty} P(\left|\overline{X}_n - E(\overline{X}_n)\right| < \varepsilon) = 1$$

则称$X_1, X_2, \dots, X_n, \dots$服从大数定律.


Weak Law of Large Numbers
0.4
0.2
0.0
-0.2
-0.4
S_n/n
0
200
400
600
800
1000
n

四、中心极限定理

1. 独立同分布的中心极限定理

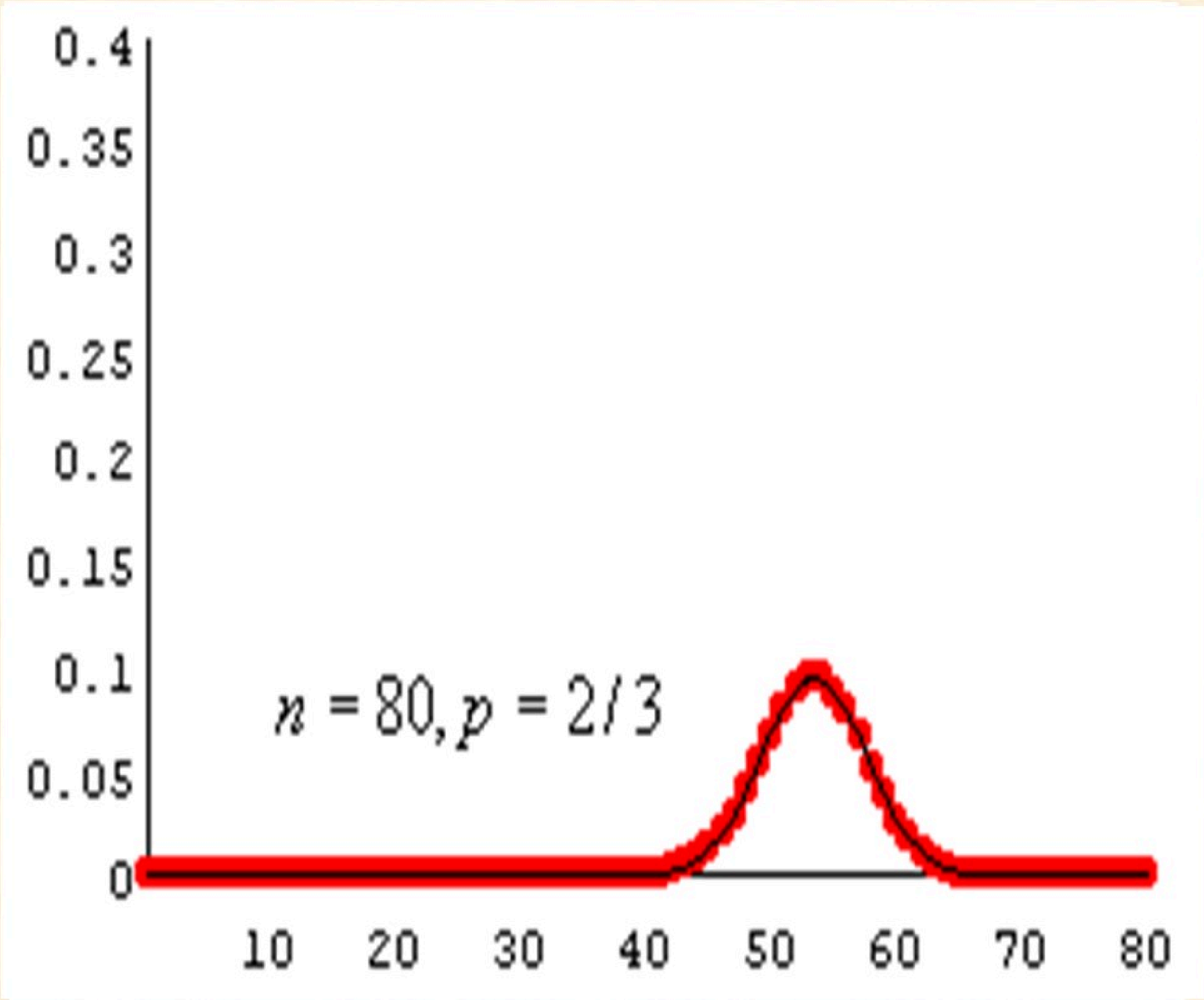
设$\{X_n, n=1,2,\ldots\}$是独立同分布的随机变量序列，且$E(X_i)=\mu$，$D(X_i)=\sigma^2$，$i=1,2,\ldots$，设标准化随机变量

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

的分布函数为$F_n(x)$，则

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

称 Y_n 依分布收敛于$X\sim N(0,1)$.


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
10
20
30
40
50
60
70
80
$n = 80, p = 2/3$

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，$Y_n = \sum_{i=1}^{n} X_i$

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1 保险公司根据调查，假设每人在购买保单后一年内意外死亡的概率为0.003. 现有10000人参保，每人每年保费120元，若投保人一年内意外死亡，保险公司需赔付20000元，问一年内保险公司亏本的概率有多大?

解 设X表示一年内死亡的人数，则 $X\sim B(n,p)$,

$$n=10000，p=0.003.$$

设Y表示保险公司一年的利润， $Y=10000\times120-20000X$

于是由中心极限定理

$$P\{Y<0\}=P\{10000\times120-20000X<0\}$$
$$=1-P\{X\le60\}$$
$$\approx1-\Phi(5.49)=0$$

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 $DX_i=4$，$i=1,2,\ldots,n$，现天文学家用 $\bar{X}_n = 1/n\sum X_i$ 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： $E\bar{X}_n = D,\ D\bar{X}_n = \dfrac{4}{n}$ 当 n 充分大时

$$\sum_{i=1}^{n} X_i \sim N(nD, 4n),\quad \bar{X}_n \sim N\left(D, \frac{4}{n}\right),\quad \frac{\bar{X}_n - D}{\sqrt{4/n}} \sim N(0,1)$$

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

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

故这位天文学家至少要作出347次独立的观测.

第六章　数理统计的基本概念

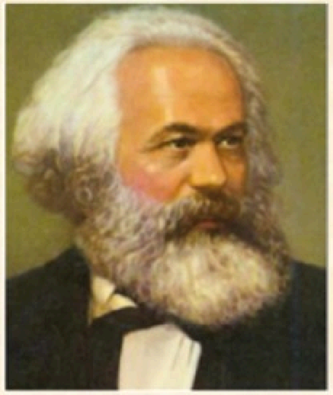
什么是统计学

统计学是收集和分析数据的科学与艺术.

——不列颠百科全书

历史上著名的统计学家

- Jacob Bernoulli (伯努利) (1654-1705)
- Edmond Halley (哈雷) (1656-1742)
- De Moivre (棣美佛) (1667-1754)
- Thomas Bayes (贝叶斯) (1702-1761)
- Leonhard Euler (欧拉) (1707-1783)
- Pierre Simon Laplace (拉普拉斯) (1749-1827)
- Adrien Marie Legendre (勒让德) (1752-1833)
- Thomas Robert Malthus (马尔萨斯) (1766-1834)
- Friedrich Gauss (高斯) (1777-1855)
- Johann Gregor Mendel (孟德尔) (1822-1884)
- Karl Pearson (皮尔森) (1857-1936)
- Ronald Aylmer Fisher (费歇) (1890-1962)
- Jerzy Neyman (内曼)(1894-1981)
- Egon Sharpe Pearson (皮尔森) (1895-1980)
- William Feller (费勒)(1906-1970)



Karl ← Carl

Karl Pearson
(1857年3月27-1936年4月27)

1879年，他以一等荣誉中第三名的
优异成绩毕业于
剑桥大学数学系

Karl Heinrich Marx
(1818年5月5日-1883年3月14日)

Sir RA Fisher
(1890. 2. 17-1962. 7. 29)

- B. A. in Math, Cambridge University, 1912
- Rothamsted Experimental Station, 1919-1933
- Professor of eugenics at University College London, 1933-1943
- Balfour Chair of Genetics at Cambridge, 1943

- RA Fisher(1925). *Statistical Methods for Research Workers*, Edinburgh, Oliver and Boyd
- RA Fisher(1930). *The Genetical Theory of Natural Selection*. The Clarendon Press

许宝騄(PL HSU)

MS Bertlett, H Cramer, HSU
(1946, Chapel HIll)

- 1910年9月1日，生于北京
- 1928-1930年，燕京大学化学系，1930年未入清华大学数学系
- 1936年赴英国伦敦University of College学习，1938年获哲学博士学位，1940年获科学博士学位
- 1940年回国，任北京大学教授
- 1945-1947年，访问美国
- 1970年12月18日，去世
- 1945年，在重庆参加了地下革命组织“民主革命同盟”。

统计学三大内容

- 抽样分布(Sampling Distribution)
- 估计(Estimation)
- 假设检验(Hypothesis Testing)

问题：一个成年男子甲的身高为175cm,一个成年女子乙的身高为170cm，请问是女的高还是男的高?

某制衣公司自1986至1990年期间，在我国不同地区随机测量了15200个人的身高数据，其均值与标准差如下：

成年男子(5115人)		成年女子(5507)	
均值	标准差	均值	标准差
167.48	6.09	156.58	5.47

基于以上数据，我们假设成年男子与成年女子身高的分布为：

$$X \sim N(167.48, 6.09^2), \quad Y \sim N(156.58, 5.47^2)$$

由此分布，我们有：

$$P\{X \geq 175\} = 0.108, \quad P\{Y \geq 170\} = 0.007.$$

标准化 $\frac{175-167.48}{6.09} = 1.23$ $\frac{170-156.58}{5.47} = 2.45$

- 在统计学中，数字的等与不等以及大与小并不是确定无疑的。它既是绝对的，又是相对的。
- 如何衡量大小，则要从概率分布角度。