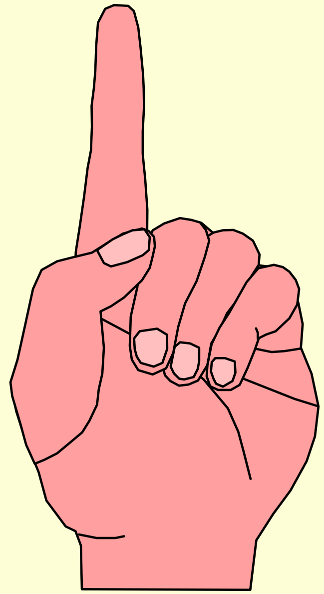
第五章 大数定律与中心极限定理

- 大数定律
- 中心极限定理



§1 大数定律

一.依概率收敛

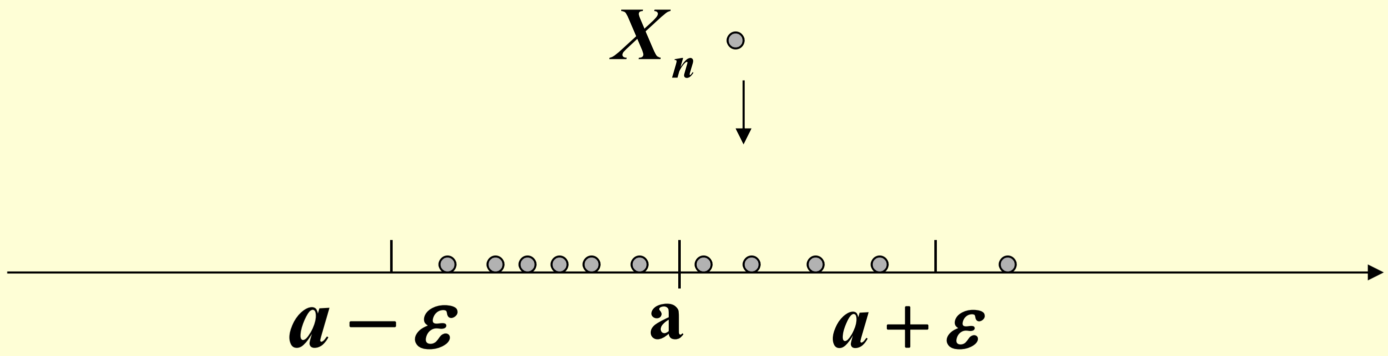
设$\{X_n\}$为随机变量序列，X为随机变量，若任给$\varepsilon>0$, 使得

$$\lim_{n\to\infty} P\{|X_n - X| < \varepsilon\} = 1$$

则称$\{X_n\}$依概率收敛于X. 可记为

$$X_n \xrightarrow{P} X.$$

例如：$X_n \xrightarrow{P} a$ 意思是：当 $n \to \infty$ 时，X_n 落在 $(a-\varepsilon, a+\varepsilon)$ 内的概率越来越大。$\forall n_0, n > n_0$



而 $X_n \to a$ 意思是：$\forall \varepsilon > 0, \exists n_0$，当 $n > n_0$

$$|X_n - a| < \varepsilon$$

二.几个常用的大数定律

1.切比雪夫大数定律

设$\{X_k, k=1,2,\ldots\}$为独立的随机变量序列，且有相同的数学期望μ，及方差$\sigma^2>0$，则

$$Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu$$

即若任给ε>0, 使得

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

证明:由切比雪夫不等式

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

这里

$$E(Y_n) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \mu$$

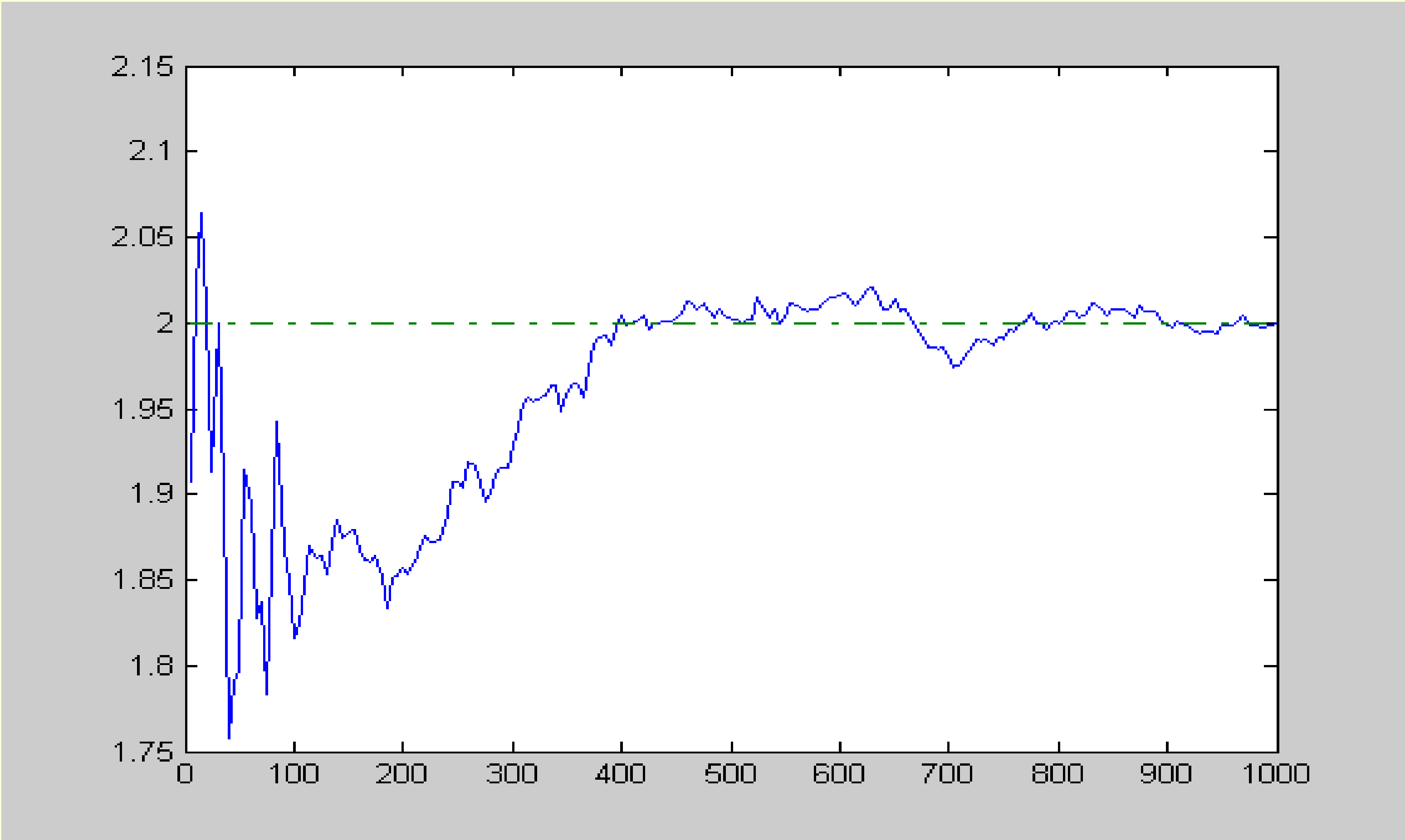
$$D(Y_n) = \frac{1}{n^2}\sum_{k=1}^{n} D(X_k) = \frac{\sigma^2}{n}$$

故

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

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

1000个[0，4]均匀分布随机数前n项算术平均值的变化趋势



2.伯努利大数定律

设进行n次独立重复试验，每次试验中事件A发生的概率为p，记f_n为n次试验中事件A发生的频率，则

$$f_n \xrightarrow{p} p \qquad n \to \infty$$

证明:设

$$X_i = \begin{cases} 1 & \text{第i次试验事件A发生} \\ 0 & \text{第i次试验事件A不发生} \end{cases}$$

则

$$E(X_i) = p, D(X_i) = p(1-p)$$

由切比雪夫大数定理

$$f_n = \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{P} p$$

3. 辛钦大数定律

若{X_k,k=1,2,...}为独立同分布随机变量序列, $EX_k=\mu<\infty$, k=1, 2, … 则

$$Y_n=\frac{1}{n}\sum_{k=1}^{n}X_k \xrightarrow{P} \mu$$

推论:若{X_i,i=1,2,...}为独立同分布随机变量序列, $E(X_1^k)=<\infty$, 则

$$\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} E(X_1^k)$$

§2 中心极限定理

一.依分布收敛

设{X_n}为随机变量序列，X为随机变量，其对应的分布函数分别为$F_n(x)$, F(x). 若在F(x)的连续点，有

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

则称{X_n}依分布收敛于X. 可记为

$$X_n \xrightarrow{w} X.$$

现令 $Y_n=\sum_{k=1}^{n} X_k$, 若 Y_n 的标准化 $Y_n^* \xrightarrow{w} \xi \sim N(0,1)$,

则称 $\{X_n\}$ 满足中心极限定理.

二.几个常用的中心极限定理

1.独立同分布中心极限定理

设$\{X_n\}$为独立同分布随机变量序列，若$EX_k=\mu<\infty$，$DX_k=\sigma^2>0$，$k=1, 2, \ldots$, 则$\{X_n\}$满足中心极限定理。

根据上述定理，当n充分大时

$$P\{\sum_{i=1}^{n} X_i \le x\} \approx \Phi(\frac{x-n\mu}{\sqrt{n}\sigma})$$

例1. 将一颗骰子连掷100次，则点数之和不少于300的概率是多少？

解:设 X_k 为第k 次掷出的点数,k=1,2,…,100,则

$X_1,\ldots,X_{100}$ 独立同分布.

$$E(X_1)=\frac{7}{2},D(X_1)=\frac{1}{6}\sum_{k=1}^{6}k^2-\frac{49}{4}=\frac{35}{12}$$

由中心极限定理

$$P\{\sum_{i=1}^{100}X_i\ge 300\}\approx 1-\Phi\left(\frac{300-100\times\frac{7}{2}}{10\sqrt{\frac{35}{12}}}\right)=1-\Phi(-2.93)$$

$$=0.9983$$

2.棣莫弗-拉普拉斯中心极限定理(De Moivre-Laplace)

设随机变量η_n(n=1, 2, ...)服从参数为n, p(0<p<1)的二项分布，则

$$\frac{\eta_n - np}{\sqrt{npq}} \xrightarrow{w} \xi \sim N(0,1).$$

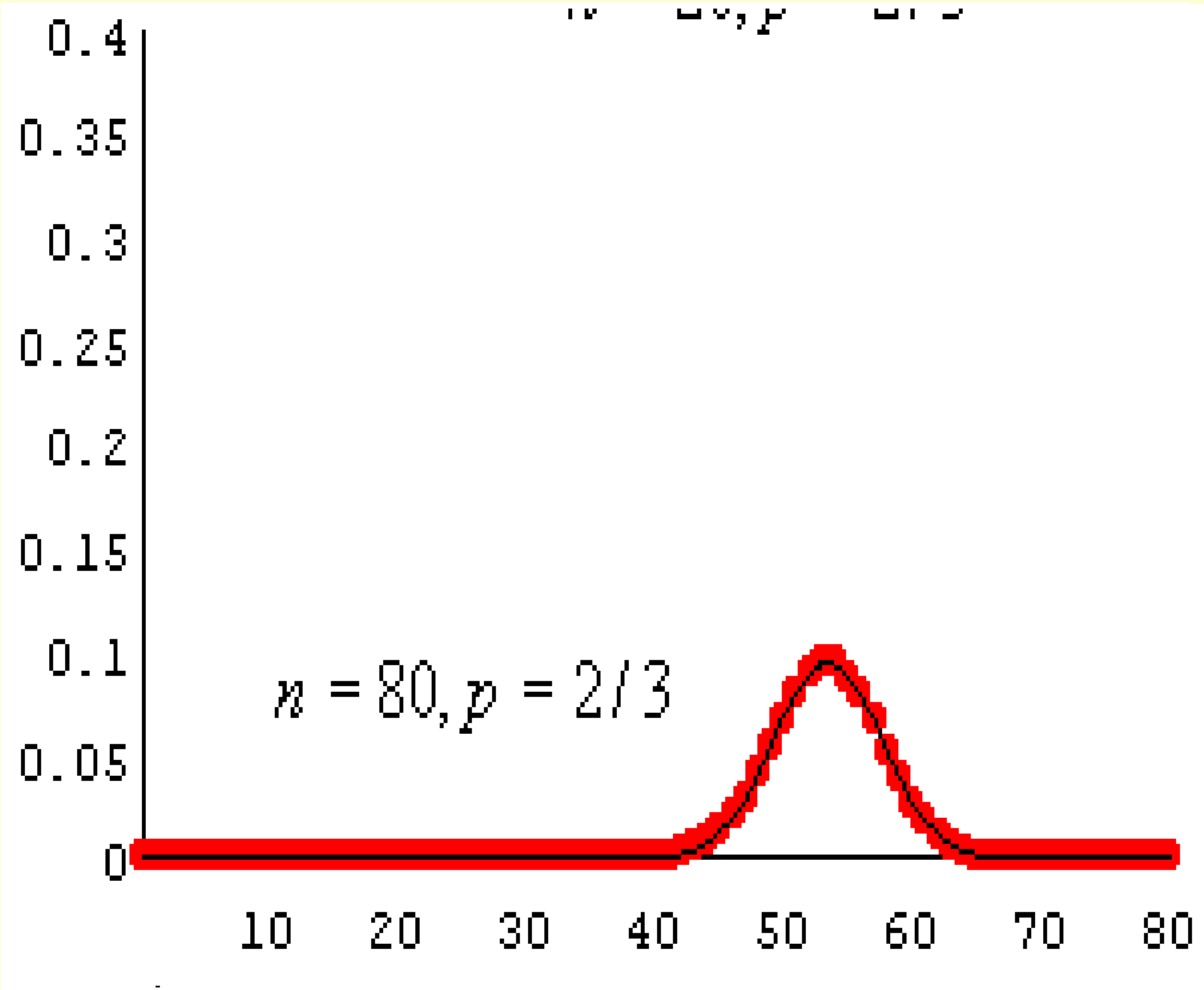
证明:设

$$X_i = \begin{cases} 1 & \text{第i次试验事件A发生} \\ 0 & \text{第i次试验事件A不发生} \end{cases}$$

则

$$E(X_i) = p, D(X_i) = p(1-p), \eta_n = \sum_{i=1}^{n} X_i$$

由中心极限定理,结论得证



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

例2 在一家保险公司里有10000个人参加寿命保险，每人每年付12元保险费。在一年内一个人死亡的概率为0.6%，死亡时其家属可向保险公司领得1000元，问：

(1)保险公司亏本的概率有多大？

(2)其他条件不变，为使保险公司一年的利润不少于60000元的概率不小于90%，赔偿金至多可设为多少？

解： 设X表示一年内死亡的人数，则$X \sim B(n, p)$,

其中

$$n = 10000,\ p = 0.6\%,$$

设Y表示保险公司一年的利润，

$$Y = 10000 \times 12 - 1000X$$

于是由中心极限定理

$$(1) P\{Y<0\} = P\{10000 \times 12 - 1000X < 0\}$$

$$= 1 - P\{X \leq 120\}$$

$$\approx 1 - \Phi(7.75) = 0;$$

（2）设赔偿金为a元，则令

$$P\{Y \geq 60000\} \geq 0.9$$

P{Y≥60000}=P{10000×12-a**X ≥ 60000}**

=P{X≤60000/a}≥0.9;

由中心极限定理,上式等价于

$$\Phi\left(\frac{\frac{60000}{a} - 10000 \times 0.006}{\sqrt{10000 \times 0.006 \times 0.994}}\right) \geq 0.9$$

$$\Rightarrow a \leq 3017$$

依概率收敛

↓

切比雪夫大数定律

↓

伯努利大数定律

↓

辛钦大数定律

依分布收敛

↓

独立同分布中心极限定理

↓

德莫弗-拉普拉斯中心极限定理

EX1 **某螺丝钉厂废品率为0.01，问一盒中应装多少个螺丝钉才能使得盒中合格品数目不少于100个的概率不少于0.95？**

解：设应装n个螺丝钉，其中有X个合格品

则X～B(n,0.99)；令$P\{X \geq 100\} \geq 0.95$

$\because n > 100$,由中心极限定理，

$$P\{X \geq 100\} \approx 1 - \Phi\left(\frac{100 - 0.99n}{\sqrt{0.99 \times 0.01n}}\right) \geq 0.95$$

$$\Rightarrow \frac{0.99n - 100}{\sqrt{0.99 \times 0.01n}} \geq 1.645 \quad \Rightarrow n \geq 103$$

EX2 检验员逐个检查某种产品，每查一件花**10**秒时间，有的产品可能要复查一次而再花**10**秒时间. 假定每一件产品需复查的概率为**1/2**，求在**8**小时内检验员能够至少检查**1900**件的概率.

解法一：设X_i为检查第i件产品所花时间,则

$$X_i=\begin{cases}10 & \text{此件不需复查}\\ 20 & \text{此件需复查}\end{cases}\Rightarrow E(X_i)=15,D(X_i)=25$$

于是,检查1900件所花时间为$\sum_{i=1}^{1900}X_i$,则在**8**小时内检验员能够至少检查**1900**件的概率为

$$P\{\sum_{i=1}^{1900}X_i<8\times3600\}\approx\Phi\left(\frac{28800-1900\times15}{\sqrt{1900\times25}}\right)\approx0.916$$

解法二：设X为1900件产品中需复查的件数,Y为检查1900件产品所花时间,则 $X \sim B(1900, \frac{1}{2})$.

$$Y = 1900 \times 10 + 10X$$

在8小时内检验员能够至少检查1900件的概率为

$$P\{Y < 8 \times 3600\} = P\{19000 + 10X < 28800\}$$

$$= P\{X < 980\}$$

$$\approx \Phi\left(\frac{980 - 1900 \times 0.5}{\sqrt{1900 \times 0.25}}\right) \approx 0.916$$