第五章：大数定律及中心极限定理习题课

第五章作业提交截止时间:
2018年11月15日

一、重点与难点

二、主要内容

三、典型例题

一、重点及难点

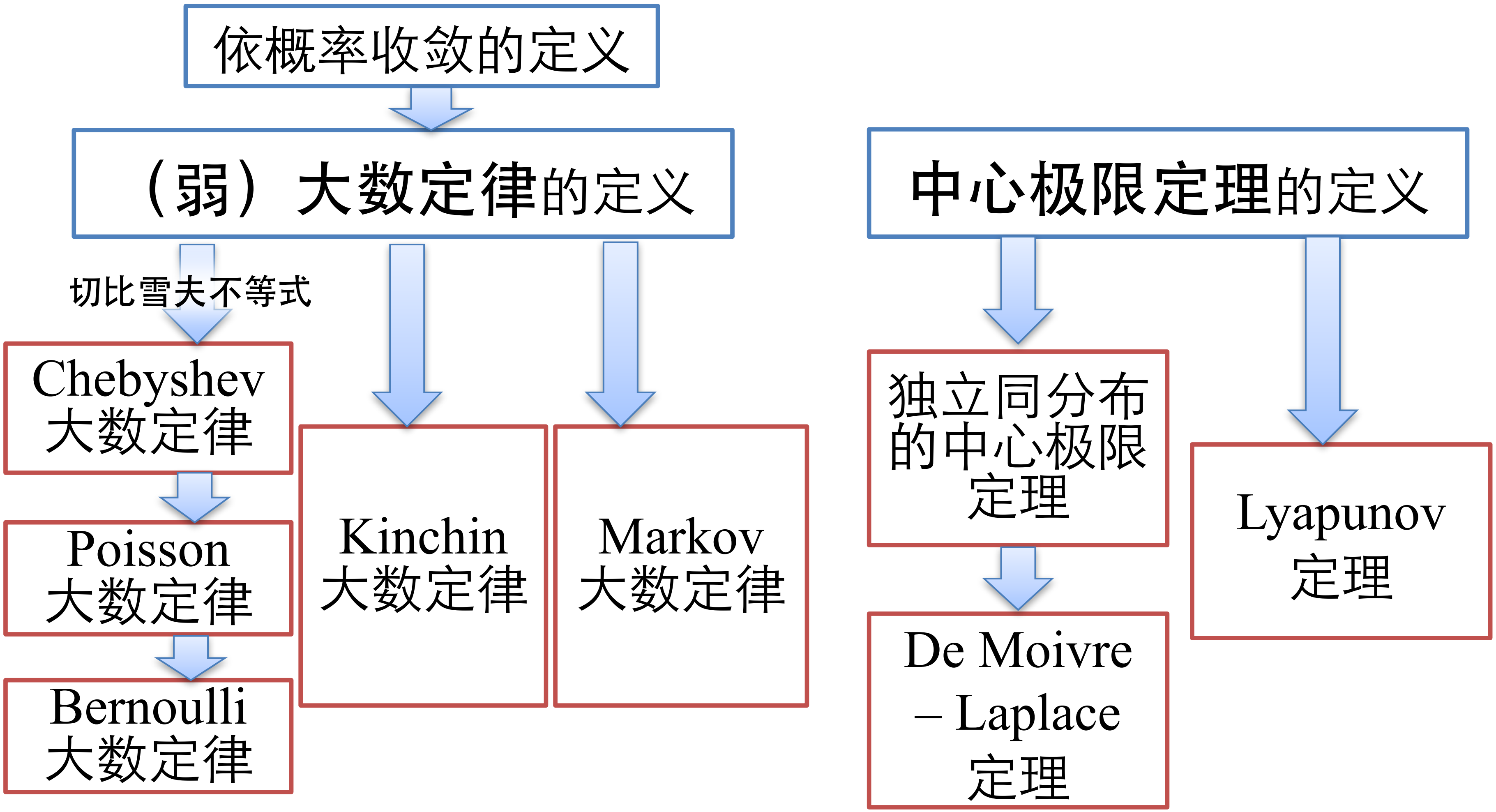
1.重点

- 利用切比雪夫不等式和中心极限定理估计和近似计算概率。
- 了解正态分布在近似计算中的作用。

2.难点

- 证明随机变量服从大数定律。

二、主要内容


依概率收敛的定义
（弱）大数定律的定义
中心极限定理的定义
切比雪夫不等式
Chebyshev 大数定律
Poisson 大数定律
Bernoulli 大数定律
Kinchin 大数定律
Markov 大数定律
独立同分布的中心极限定理
De Moivre – Laplace 定理
Lyapunov 定理

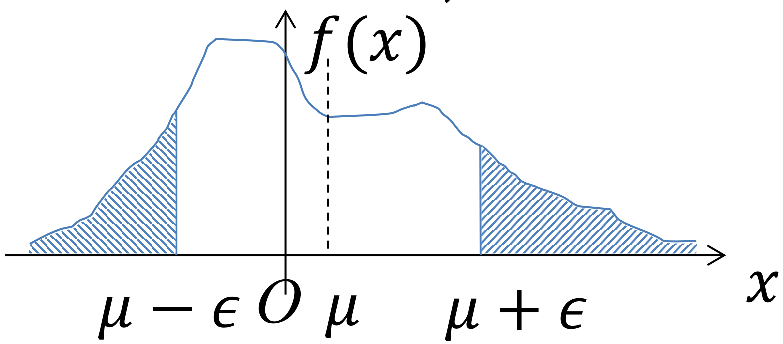
切比雪夫（Chebyshev）不等式

❑ 回顾：切比雪夫（Chebyshev）不等式

设随机变量X的**期望和方差都存在**，设期望$E(X)=\mu$，方差$D(X)=\sigma^2$，则对于任意正数ϵ，有

$$P\{|X-\mu|\geq\epsilon\}\leq\frac{\sigma^2}{\epsilon^2}$$

$$\left(\text{等价地：}P\{|X-\mu|<\epsilon\}\geq 1-\frac{\sigma^2}{\epsilon^2}\right)$$



依概率收敛

❑ 依概率收敛的定义

设$Y_1, Y_2, \dots, Y_n, \dots$是一个随机变量序列， a是一个常数，若对任意正数ε, 有

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

则称序列$\{Y_n\}$依概率收敛于a, 记为

$$Y_n \xrightarrow{P} a$$

（弱）大数定律

❑ (弱)大数定律的定义

设$\{X_n, n = 1,2, \dots\}$是一个随机变量序列，而且对每个n，**$E(X_n)$存在**，如果对于任意给定的$\epsilon > 0$，有

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

则称随机变量序列$\{X_n\}$服从（弱）大数定律.

等价形式：

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| > \epsilon\right\} = 0$$

即$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \frac{1}{n}\sum_{k=1}^{n} E(X_k)$

什么时候（弱）大数定律成立

❑ 常见的大数定律

如果随机变量序列$\{X_n\}$满足以下条件之一：

(1) $X_1, X_2, \ldots, X_n, \ldots$ **两两独立**, 且 $E(X_i)$, $D(X_i)$ **存在且有界**；**Chebyshev大数定律**

(2) $X_1, X_2, \ldots, X_n, \ldots$ 服从**0-1分布**, 且两两**独立**（或**独立同分布**）；**Poisson大数定律(或Bernoulli大数定律)**

(3) $X_1, X_2, \ldots, X_n, \ldots$ **独立同分布**, $E(X_i)$**存在**; **Khinchin大数定律**

(4) $E(X_i)$**存在**, $D(X_i)$满足 $\lim_{n\to\infty} D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = 0$；**Markov大数定律**

那么，随机变量序列$\{X_n\}$服从大数定律。

切比雪夫(Chebyshev)大数定律

设随机变量序列$\{X_n\}$满足如下条件：

(1) $X_1, X_2, \dots, X_n, \dots$**两两相互独立**；

(2) 对每一个$n(n=1,2,\dots)$，$\boldsymbol{D(X_n)}$**存在**；

(3) 数列$\boldsymbol{\{D(X_n)\}}$**有界**，即存在常数c，使得对于任意的$n(n=1,2,\dots)$，有$D(X_n)\le c$，则对任意$\epsilon>0$，有

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

即随机变量序列$\{X_n\}$ 服从大数定律

切比雪夫(Chebyshev)大数定律的特殊形式

设随机变量$X_1, X_2, \ldots, X_n, \ldots$是**独立同分布**的随机变量，且$E(X_k) = \mu, D(X_k) = \sigma^2 \ (k = 1,2, \ldots)$，做前n个随机变量的算术平均

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

则对任意$\epsilon > 0$，有

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

泊松(Poisson)大数定律

如果在**独立**试验序列中，事件A在第n次试验中出现概率为p_n，设n_A是前n次试验中事件A出现的次数，则对任意$\epsilon > 0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - \frac{p_1 + p_2 + \cdots + p_n}{n}\right| < \epsilon\right\} = 1$$

Bernoulli大数定律

设f_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$\epsilon > 0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{f_A}{n} - p\right| < \epsilon\right\} = 1$$

或

$$\lim_{n\to\infty} P\left\{\left|\frac{f_A}{n} - p\right| > \epsilon\right\} = 0$$

Khinchin大数定律

设随机变量$X_1, X_2, \dots, X_n, \dots$是**独立同分布**的随机变量序列，且**具有数学期望$\mathbf{E(X_k) = \mu}$**，$(k = 1,2,\dots)$，做前n个变量的算术平均

$$\frac{1}{n}\sum_{k=1}^{n} X_k$$

则对任意$\epsilon > 0$，有

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

注意：与前面Chebyshev大数定律的特殊情况相比，这里没有要求$D(X_n)$存在

Markov大数定律

设随机变量$\{X_n\}$是随机变量序列，如果

$$\lim_{n\to\infty} D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = 0 \text{ (称为Markov条件)},$$

则对任意$\epsilon > 0$，有

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

不要求随机变量的独立性，因此提供了一种研究非独立随机变量序列服从大数定律的方法。

大数定律的关系

大数定律的一般形式

对随机变量序列$\{X_n\}$ $(n=1,2,\ldots)$，$E(X_n)$存在，

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

不要求独立同分布

Khinchin大数定律

$X_1,X_2,\ldots,X_n,\ldots$ 独立同分布，

$E(X_k)=\mu,\ (k=1,2,\ldots)$

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

Markov大数定律

随机变量序列$\{X_n\}$，$E(X_n)$存在，且

$$\lim_{n\to\infty} D\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=0\ (\text{Markov条件})$$

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

Bernoulli试验

Bernoulli试验

Bernoulli大数定律

n次独立重复试验发生f_A次，p是每次试验中发生的概率

$$\frac{f_A}{n} \xrightarrow{P} p$$

不要求独立性，但对方差有约束

Poisson大数定律

n 次独立试验发生 f_A 次，第k次发生概率为p_k

$$\frac{n_A}{n} \xrightarrow{P} \frac{1}{n}\sum_{k=1}^{n}p_k$$

不要求方差的存在性

0-1分布

0-1分布

Chebyshev大数定律

$X_1,X_2,\ldots,X_n,\ldots$两两相互独立

$E(X_n)$、$D(X_k)$存在$(k=1,2,\ldots)$

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

独立同分布

独立同分布下的Chebyshev大数定律

$E(X_k)=\mu, D(X_k)$存在

$(k=1,2,\ldots)$

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

中心极限定理

定义：设$X_1, X_2, \dots, X_n, \dots$是**独立**的随机变量序列，若$\boldsymbol{E(X_k)}$, $\boldsymbol{D(X_k)\ (k=1,2,\dots)}$**都存在**，令

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

若对任意$x \in R$，有，

$$\lim_{n\to\infty} P\{Y_n \le x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

则称该随机变量序列服从**中心极限定理**。

等价地，若$\{X_n\}$服从中心极限定理，则当n足够大时, Y_n近似服从标准正态分布。

独立同分布的中心极限定理(Lindburg-Levy定理)

设随机变量$X_1, X_2, \dots, X_n, \dots$**独立同分布**，且$\boldsymbol{E(X_k) = \mu, D(X_k) = \sigma^2 > 0, (k = 1, 2, \dots)}$，则随机变量之和$\sum_{k=1}^{n} X_k$的标准化变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - E\left(\sum_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n} X_k\right)}} = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

的分布函数$F_n(x)$对于任意的x满足

$$F_n(x) = \lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \le x\right\}$$

$$= \int_{-\infty}^{x} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt = \Phi(x)$$

李雅普诺夫(Lyapunov)定理

设随机变量$X_1, X_2, \dots, X_n, \dots$**相互独立**，且$\boldsymbol{E(X_k) = \mu_k, D(X_k) = \sigma_k^2 > 0}$, $(k = 1,2, \dots)$，记$B_n^2 = \sum_{k=1}^{n} \sigma_k^2$，

若存在$\delta > 0$，使得当$n \to \infty$时

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E\{|X_k - \mu_k|^{2+\delta}\} \to 0$$

则随机变量之和$\sum_{k=1}^{n} X_k$的标准化变量

$$Z_n = \frac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}} = \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n}$$

的分布函数$F_n(x)$对于任意的x满足

$$\lim_{n \to \infty} F_n(x) = \lim_{n \to \infty} P\{\frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n} \le x\}$$

$$= \int_{-\infty}^{x} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt = \Phi(x)$$

棣莫佛－拉普拉斯(De Moivre－Laplace)定理

设随机变量$\eta_n(n=1,2,\ldots)$服从参数为$n,p\ (0<p<1)$的**二项分布**，则对于任意x，有

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

De Moivre-Laplace定理可以认为是**独立同分布的中心极限定理**的一个**特例**： η_n可以分解成***n*个相互独立、服从同一（0-1）分布**的随机变量$X_1, X_2, \ldots, X_n$**之和**。

二项分布的逼近

对于一个参数为n, p的二项分布，当$n \to \infty$时：

（1）若$np = \lambda$，则该二项分布逼近参数为λ的泊松分布；泊松定理

（2）若$np \to \infty$，则该二项分布逼近正态分布。

De Moivre – Laplace定理

在实际计算中：
如果n很大，但是p很小或$q = 1 - p$很小，应该用泊松定理去近似二项分布；
如果n, np和nq都较大，应该用中心极限定理去近似。

三个中心极限定理

Lyapunov定理

$\{X_n\}$ 独立，且$E(X_k)=\mu_k$, $D(X_k)=\sigma_k^2>0$, $B_n^2=\sum_{k=1}^n\sigma_k^2$，

若存在$\delta>0$，使得当$n\to\infty$时

$$\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}E\{|X_k-\mu_k|^{2+\delta}\}\to 0$$

$$\lim_{n\to\infty}Z_n=\frac{\sum_{k=1}^n X_k-\sum_{k=1}^n\mu_k}{B_n}\sim N(0,1)$$

不要求同分布，但要求方差满足收敛性

独立同分布的中心极限定理

$\{X_n\}$独立同分布，且$E(X_k)=\mu$, $D(X_k)=\sigma^2>0$, $(k=1,2,\ldots)$，

$$\lim_{n\to\infty}Y_n=\frac{\sum_{k=1}^n X_k-n\mu}{\sqrt{n}\sigma}\sim N(0,1)$$

二项分布逼近

De Moivre – Laplace定理

$\eta_n\sim b(n,p)$ $(0<p<1)$，则对于任意x，有

$$\lim_{n\to\infty}\frac{\eta_n-np}{\sqrt{np(1-p)}}\sim N(0,1)$$

三、典型例题

例1. 设独立同分布随机变量序列$\{X_n\}$均服从$[0,a]$上的均匀分布，令

$$Y_n = \max(X_1, X_2, \dots, X_n),$$

试证明：$Y_n \xrightarrow{P} a.$

【思路】要证$Y_n \xrightarrow{P} a$，即证对任给的$\epsilon > 0$，有

$$\lim_{n\to\infty} P\{|Y_n - a| > \epsilon\} = 0,$$

因此，只需要先求出概率$P\{|Y_n - a| > \epsilon\}$即可。

证： 由于 $\{X_n\}$均服从$[0,a]$上的均匀分布，因此，对任意n，当$x>a$时，有

$$P\{\mathrm{Y_n} > x\} = 0.$$

对任意的$0<\epsilon<a$，有

$$P\{|\mathrm{Y_n} - a| > \epsilon\}$$
$$= P\{Y_n < a - \epsilon\} + P\{Y_N > a + \epsilon\}$$
$$= P\{Y_n < a - \epsilon\}$$
$$= P\left\{\max(X_1, X_2, \ldots, X_n) < a - \epsilon\right\}$$
$$= P\{X_1 < a - \epsilon, X_2 < a - \epsilon, \ldots, X_n < a - \epsilon\}$$
$$= P\{X_1 < a - \epsilon\}P\{X_2 < a - \epsilon\}\ldots P\{X_n < a - \epsilon\}$$
$$= \left(\frac{a-\epsilon}{a}\right)^n$$

从而$\lim\limits_{n\to\infty} P\{|Y_n - a| > \epsilon\} = 0$，即$Y_n \xrightarrow{P} a$。

例2. 设$X_1, X_2, \ldots, X_n$是独立同分布的随机变量序列，$E(X_i)=\mu$，$D(X_i)=\sigma^2$均存在，证明

$$Y_n = \frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i \xrightarrow{P} \mu.$$

【思路】：由依概率收敛的定义，

即证$\forall\varepsilon>0$，有$\lim\limits_{n\to\infty} P\{|Y_n-\mu|<\varepsilon\}=1$;

即 $\forall\varepsilon>0$，有$\lim\limits_{n\to\infty} P\{|Y_n-\mu|\geq\varepsilon\}=0$。

考虑到$E(Y_n)=\mu$，$D(Y_n)$存在，可由切比雪夫不等式证出。

证： 因为 $E(Y_n) = \mathrm{E}\left\{\frac{2}{\mathrm{n(n+1)}}\sum_{i=1}^{n} iX_i\right\}$

$$= \frac{2}{\mathrm{n(n+1)}}\sum_{i=1}^{n} iE(X_i)$$

$$= \frac{2\mu}{\mathrm{n(n+1)}}\sum_{i=1}^{n} i \ = \mu$$

$$D(Y_n) = \frac{4}{\mathrm{n^2(n+1)^2}}\sum_{i=1}^{n} i^2 D(X_i) = \frac{4\sigma^2}{\mathrm{n^2(n+1)^2}}\sum_{i=1}^{n} i^2$$

$$= \frac{4n(n+1)(2n+1)\sigma^2}{6n^2(n+1)^2} = \frac{2(2n+1)\sigma^2}{3n(n+1)}$$

随机变量Y_n的数学期望$E(Y_n)=\mu$，且方差$D(Y_n)$存在，则对于任意正数$\varepsilon>0$，由切比雪夫不等式得：

$$P\{|Y_n-\mu|\geq\varepsilon\}\leq\frac{D(Y_n)}{\varepsilon^2}=\frac{2\sigma^2(2n+1)}{2n(n+1)\varepsilon^2}$$

$$\lim_{n\to\infty}P\{|Y_n-\mu|\geq\varepsilon\}\leq\lim_{n\to\infty}\frac{2\sigma^2(2n+1)}{2n(n+1)\varepsilon^2}=0$$

即$Y_n\xrightarrow{P}\mu$.

例3. 设X的概率密度为$\mathrm{f}(x)=\frac{x^m}{m!}e^{-x}, x\geq 0$,m为正整数，试证

$$P\{0<X<2(m+1)\}\geq\frac{m}{m+1}$$

【思路】：直接计算$\int_0^{2(m+1)}\frac{x^m}{m!}e^{-x}dx$，过程繁琐，不等式可考虑用切比雪夫不等式证明。
另，考虑到当m为正整数时，Γ函数的形式$\Gamma(\mathrm{m})=\int_0^{\infty}x^{m-1}e^{-x}dx=(m-1)!$，可简化X的期望与方差的计算。

证明： $E(X)=\int_0^{+\infty} x\cdot\frac{x^m}{m!}e^{-x}dx$

$$=\frac{1}{m!}\int_0^{+\infty} x^{(m+2)-1}\,e^{-x}dx=\frac{1}{m!}\Gamma(m+2)$$

$$=\frac{(m+1)!}{m!}=m+1$$

$$D(X)=E(X^2)-[E(X)]^2$$

$$=\int_0^{+\infty} x^2\cdot\frac{x^m}{m!}e^{-x}dx-(m+1)^2$$

$$=\frac{1}{m!}\Gamma(m+3)-(m+1)^2=m+1$$

X的期望和方差均存在，则对任意正数ε，由切比雪夫不等式得

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

取$\varepsilon = \mathrm{m}+1$，则

$$P\{|X-(m+1))|\leq m+1\}\geq 1-\frac{m+1}{(m+1)^2}$$

$$P\{0\leq X\leq 2(m+1)\}\geq 1-\frac{m+1}{(m+1)^2}=\frac{m}{m+1}$$

得证。

例4. 设$\boldsymbol{X_1, X_2, \ldots, X_n}$ $\boldsymbol{i.i.d,}$ $\mathbf{E}(\boldsymbol{X_k}) = \boldsymbol{\mu}, \boldsymbol{D}(\boldsymbol{X_k}) = \mathbf{8}, \boldsymbol{k} = \mathbf{1, 2, \cdots}$，令$\boldsymbol{\bar{X}} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{k=1}^{n} \boldsymbol{X_k}$，利用切比雪夫不等式估计$\boldsymbol{P(|\bar{X} - \mu| < 4)}$。

解：

$$\mathrm{E}(\bar{X}) = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \mu,$$

$$\mathrm{D}(\bar{X}) = D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n} D(X_k) = \frac{8}{n},$$

由切比雪夫不等式，

$$P(|\bar{X} - \mu| < 4) = P(|\bar{X} - \mathrm{E}(\bar{X})| < 4) \geq 1 - \frac{\frac{8}{n}}{4^2}$$

$$= 1 - \frac{1}{2n}$$

例5. 掷六颗骰子，利用切比雪夫不等式估计六颗骰子出现点数和在15~27之间的概率。

解： 令随机变量X_i表示第i颗骰子出现的点数$(i=1,2,3,4,5,6)$，则X_i独立同分布，有

$$\mathrm{E}(X_i)=\frac{1}{6}(1+2+3+4+5+6)=\frac{7}{2},$$

$$\mathrm{E}\left(\mathrm{X}_i^2\right)=\frac{1}{6}\left(1^2+2^2+3^2+4^2+5^2+6^2\right)=\frac{91}{6},$$

$$\mathrm{D}(X_i)=\mathrm{E}\left({X_i}^2\right)-[\mathrm{E}(X_i)]^2=\frac{91}{6}-\left(\frac{7}{2}\right)^2=\frac{35}{12}.$$

令$X=\sum_{i=1}^{6}X_i$表示六颗骰子的点数和,有

$$\mathrm{E}(X)=\mathrm{E}(\sum_{i=1}^{6}X_i)=\sum_{i=1}^{6}E(X_i)=\sum_{i=1}^{6}\frac{7}{2}=6\times\frac{7}{2}=21,$$

$$\mathrm{D}(X)=\mathrm{D}(\sum_{i=1}^{6}X_i)=\sum_{i=1}^{6}D(X_i)=6\times\frac{35}{12}=\frac{35}{2},$$

根据切比雪夫不等式有:

$$\mathrm{P}(15<X<27)=\mathrm{P}(|\mathrm{X}-21|<6)$$
$$\geq 1-\frac{\frac{35}{2}}{6^2}=\frac{37}{72}$$

例6. 100个独立工作(工作的概率为0.9)的部件组成一个系统，求系统中至少有85个部件工作的概率.

【思路】中心极限定理已知n和$\sum X$的取值范围，求概率p.

解： 令$X_i=1$表示第i个部件正常工作，反之记为$X_i=0$, $E(X_i)=0.9$，$D(X_i)=0.09$。记$Y=X_1+X_2+\cdots+X_{100}$，因为X_i独立同分布，则随机变量序列$\{X_n\}$服从中心极限定理，即近似地

$$\frac{Y-100E(X_i)}{\sqrt{100D(X_i)}}\sim N(0,1)$$

由此得：$P\{Y\geq 85\}=1-P\{Y\leq 85\}$

$$=1-\emptyset\left(\frac{85-90}{\sqrt{9}}\right)=0.9525$$

例7. 有200台独立工作(工作的概率为0.7)的机床，每台机床工作时需15kw电力. 问共需多少电力，才可有95%的可能性保证供电充足?

【思路】中心极限定理已知n和概率p，求$\sum X$的取值范围.

解：用$X_i=1$表示第i台机床正常工作，反之记为$X_i=0$.

则 $E(X_i)=0.7$，$D(X_i)=0.21$. 记$Y=X_1+X_2+\ldots+X_{200}$，因为

X_i独立同分布，由中心极限定理有$\frac{Y-200E(X_i)}{\sqrt{200D(X_i)}}\sim N(0,1)$.

设供电量为y，供电充足即为$15Y\le y$，则

$$P\{15Y\le y\}\approx \emptyset\left(\frac{y/15-140}{\sqrt{42}}\right)\ge 0.95$$

解得： $y\ge 2260.4$

例8. 用调查对象中的收看比例k/n作为某电视节目的收视率p的估计。要有90%的把握，使k/n与p的差异不大于0.05，问至少要调查多少对象？

【思路】中心极限定理已知Y和概率p，求n.

解：用$X_i=1$表示第i个调查对象观看某电视节目，否则$X_i=0$，$EX_i=p$，$DX_i=p(1-p)$。$Y_n=X_1+\cdots+X_n$表示前n个调查对象中观看节目的人数，因为X_i相互独立，所以$\frac{Y_n-np}{\sqrt{np(1-p)}}\sim N(0,1)$。

根据题意：$\mathrm{P}\left\{\left|\frac{Y_n}{n}-p\right|<0.05\right\}=\mathrm{P}\left\{\left|\frac{Y_n-np}{\sqrt{np(1-p)}}\right|<0.05\sqrt{\frac{n}{p(1-p)}}\right\}\approx 2\emptyset\left(0.05\sqrt{\frac{n}{p(1-p)}}\right)-1\geq 0.90$

解得：$\mathrm{n}\geq 272.25$，即：$\mathrm{n}=273$

例9. 设每颗炮弹命中目标的概率为0. 01，求500发炮弹中命中 5 发的概率.

解：设 X 表示命中的炮弹数，则 $X \sim b(500, 0.01)$

(1) $\mathrm{P}(\mathrm{X}=5)=C_{500}^{5}\times 0.01^{5}\times 0.99^{495}=0.17635$

(2) 应用正态逼近，近似地 $\dfrac{X-500\times 0.01}{\sqrt{500\times 0.01\times 0.99}} \sim N(0,1)$

$$\mathrm{P}(X=5)=P(4.5<X<5.5)$$

$$\approx \emptyset\left(\frac{5.5-5}{\sqrt{4.95}}\right)-\emptyset\left(\frac{4.5-5}{\sqrt{4.95}}\right)=0.1742$$

“标准化”和“正态近似”，n越大所得的近似值越精确。

例10. 一生产线生产的产品成箱包装，每箱的重量是随机的。假设每箱平均重50千克，标准差为5千克。若用最大载重量为5吨的汽车承运，试利用中心极限定理说明每辆车最多可以装多少箱，才能保证不超载的概率大于0.977。

【思路】独立同分布中心极限定理已知概率，求n

解：设$X_i(i=1,2,\ldots,n)$是装运第i箱的重量，n是所求箱数，则载重量$Y_n=X_1+X_2+\cdots+X_n$是独立同分布的随机变量$X_i(i=1,2,\ldots,n)$之和。

$$\mu=EX_i=50,\qquad \sigma^2=DX_i=25\ (i=1,2,..,n).$$

根据独立同分布的中心极限定理知$\frac{Y_n-50n}{5\sqrt{n}}$的极限分布是标准正态分布。

$$P\{不超载\}=P\{0<Y_n\le 5000\}$$

$$=P\{\frac{0-50n}{5\sqrt{n}}<\frac{Y_n-50n}{5\sqrt{n}}<\frac{5000-50n}{5\sqrt{n}}\}$$

$$\approx\Phi\left\{\frac{1000-10n}{\sqrt{n}}\right\}-\Phi\{-10\sqrt{n}\}>0.977$$

解得$\frac{1000-10n}{\sqrt{n}}>2$，即$n<98.0199$，故最多可装98箱。

例11. 某保险公司的老年人寿保险有1万人参加, 每人每年交200元. 若老人在该年内死亡, 公司付给家属1万元. 设老年人死亡率为0. 017, 试求保险公司在一年内的这项保险中亏本的概率.

解：设X为一年中投保老人的死亡数，则X服从二项分布$B(n,p)$, 其中$n=10000$，$p=0.017$。

由棣莫佛—拉普拉斯定理知保险公司亏本的概率：

$$P\{10000X>10000\times 200\}=P\{X>200\}$$

$$=P\left\{\frac{X-np}{\sqrt{np(1-p)}}>\frac{200-np}{\sqrt{np(1-p)}}\right\}=P\{\frac{X-np}{\sqrt{np(1-p)}}>2.321\}$$

$$\approx 1-\Phi\{2.321\}\approx 0.01。$$

例12　假设一条自动生产线生产的产品的合格率为0.8，试用中心极限定理计算，要使一批产品的合格率在76%与84%之间的概率不小于90%，问这批产品至少要生产多少件？

【思路】：根据De Moivre-Laplace定理求二项式分布$B(n,p)$的参数n。

解：设该批产品至少需要生产n件，才能使其合格率在76%与84%之间的概率不小于90%。

再设随机变量$X \sim B(n,p)$为这批产品中的合格品数目。因此，n需满足下面的不等式：

$$P\left\{0.76 \le \frac{X}{n} \le 0.84\right\} \ge 0.90$$

由De Moivre-Laplace定理知, 近似地

$$\frac{X-np}{\sqrt{np(1-p)}}\sim N(0,1)$$

则$P\left\{0.76 \le \frac{X}{n} \le 0.84\right\}$

$$= P\left\{\frac{0.76n-0.8n}{\sqrt{0.16n}} \le \frac{X-0.8n}{\sqrt{0.16n}} \le \frac{0.84n-0.8n}{\sqrt{0.16n}}\right\}$$

$$= P\left\{-0.1\sqrt{n} \le \frac{X-0.8n}{\sqrt{0.16n}} \le 0.1\sqrt{n}\right\}$$

$$\approx \Phi(0.1\sqrt{n}) - \Phi(-0.1\sqrt{n}) = 2\Phi(0.1\sqrt{n}) - 1$$

因此, 要使$P\left\{0.76 \le \frac{X}{n} \le 0.84\right\} \ge 0.90$, 只需

$$2\Phi(0.1\sqrt{n}) - 1 \ge 0.90,$$

即$\Phi(0.1\sqrt{n}) \ge 0.95.$

查表得$0.1\sqrt{n} \ge 1.65$，故 $n \ge 272.25.$

因此，由中心极限定理估计，可知这批产品至少要生产273件，才能使其合格率在76%与84%之间的概率不小于90%。

例13. 假设某大学中报名选修统计课的学生人数服从参数为100的泊松分布。负责开课的教授决定，如果学生的报名人数不少于120人，就分成两个班授课，如果少于120人，就集中在一个班讲授。试问：教授将讲授两个班的概率是多少？

【思路】设报名选修统计课的学生人数为X，故所求概率为

$$P\{X \geq 120\} = 1 - \sum_{k=120}^{\infty} \frac{100^k e^{-100}}{k!}$$

上式计算量惊人，考虑应用中心极限定理求近似值。根据泊松分布的可加性，将随机变量X（服从参数100的泊松分布）看成是100个相互独立的服从参数为1的泊松分布的随机变量之和。

泊松分布的可加性

$X \sim P(\lambda_1), Y \sim P(\lambda_2)$，若$X$与$Y$独立，则$X+Y \sim P(\lambda_1+\lambda_2)$

证明：$P(X+Y=n)=\sum_{k=0}^{n} P(X=k, Y=n-k)=\sum_{k=0}^{n} P(X=k)P(Y=n-k)$

$$=\sum_{k=0}^{n} \frac{\lambda_1^k}{k!} e^{-\lambda_1} \cdot \frac{\lambda_2^{n-k}}{(n-k)!} e^{-\lambda_2}=\frac{1}{n!} e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k}$$

$$=\frac{1}{n!} e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} C_n^k \lambda_1^k \lambda_2^{n-k}=\frac{(\lambda_1+\lambda_2)^n}{n!} e^{-(\lambda_1+\lambda_2)}$$

$\therefore \quad X+Y \sim P(\lambda_1+\lambda_2)$

解： 设报名选修统计课的学生人数为$X \sim \mathcal{P}(100)$，设随机变量$X_1, X_2, \dots, X_{100}$相互独立且$X_i \sim \mathcal{P}(1), i = 1,2,\dots,100$。由泊松分布的可加性知，

$$X_1 + X_2 + \cdots + X_{100} \sim \mathcal{P}(100)。$$

又因为$E(X_i) = 1, D(X_i) = 1, i = 1,2,\dots,100$, 由独立同分布的中心极限定理可得

$$P\{X \geq 120\} = P\{X_1 + X_2 + \cdots + X_{100} \geq 120\}$$

$$= P\left\{\frac{\sum_{i=1}^{100} X_i - 100}{\sqrt{100}} \geq \frac{120-100}{\sqrt{100}}\right\}$$

$$\approx 1 - \Phi\left(\frac{120-100}{\sqrt{100}}\right) = 1 - \Phi(2)$$

$$= 0.0228$$

作业题《概率论及其应用》Chpt. 10, 7:

设$\{X_k\}$是相互独立的具有同分布的随机变量序列，且有均值μ和有限的方差。令$S_n = X_1 + X_2 + \cdots + X_n$.试证：大数定律对随机变量序列$\{S_n\}$不成立。如果$na_n \to 0$，则$\{a_n S_n\}$服从大数定律。

思路：（1）弱大数定律的定义要求:如果$\{S_n\}$满足大数定律，那么对于每个n，$E(S_n)$存在，并且对于任意给定的$\epsilon > 0$，有

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

考虑用反证法证明。如果$\{S_n\}$服从大数定律，则可以推出$\lim\limits_{n\to\infty} D\left(\frac{1}{n}\sum_{i=1}^{n} S_n\right) = 0$；但是事实上$\lim\limits_{n\to\infty} D\left(\frac{1}{n}\sum_{i=1}^{n} S_n\right) = \infty$。因此，得出假设不成立。

证明：（1）假设$\{S_n\}$满足大数定律，则$\overline{S_n}=\frac{1}{n}\sum_{i=1}^{n}S_i$依概率收敛到$ES=\frac{1}{n}\sum_{i=1}^{n}E(S_i)$，即对任意$\varepsilon>0$，都有

$$\lim_{n\to\infty}P\{|\overline{S_n}-ES|<\varepsilon\}=1。$$

又$\lim_{n\to\infty}D(\overline{S_n})=\lim_{n\to\infty}E\left[\left(\overline{S_n}-E(\overline{S_n})\right)^2\right]=\lim_{n\to\infty}\int\left(\overline{S_n}-E(S_n)\right)^2p(\overline{S_n})d\overline{S_n}$

则对任意$\varepsilon>0$，都有

$$\lim_{n\to\infty}D(\overline{S_n})=\lim_{n\to\infty}\{\int_{|\overline{S_n}-E(\overline{S_n})|<\varepsilon}\left(\overline{S_n}-E(\overline{S_n})\right)^2p(\overline{S_n})d\overline{S_n}+\int_{|\overline{S_n}-E(\overline{S_n})|\geq\varepsilon}\left(\overline{S_n}-E(\overline{S_n})\right)^2p(\overline{S_n})d\overline{S_n}\}$$

$$\leq\varepsilon^2\underbrace{\lim_{n\to\infty}P\{|\overline{S_n}-ES|<\varepsilon\}}_{1}+\underbrace{\lim_{n\to\infty}P\{|\overline{S_n}-ES|\geq\varepsilon\}}_{0}\sup_{|\overline{S_n}-ES|\geq\varepsilon}(\overline{S_n}-ES)^2$$

$$=\varepsilon^2\cdot1+0=\varepsilon^2$$

取$\varepsilon\to0$，即可得$\lim_{n\to\infty}D(\overline{S_n})\to0$。

又因为

$$\lim_{n\to\infty} \mathrm{D}(\overline{\mathrm{S}_n}) = \lim_{n\to\infty} D(\frac{1}{n}\sum_{i=1}^{n} S_i) = \lim_{n\to\infty} \frac{1}{n^2} D(S_1 + S_2 + \cdots + S_n)$$

$$= \lim_{n\to\infty} \frac{1}{n^2} D[X_1 + (X_1+X_2) + \cdots + (X_1 + X_2 \cdots + X_n)]$$

$$= \lim_{n\to\infty} \frac{1}{n^2} D[nX_1 + (n-1)X_2 + (n-2)X_3 \cdots + X_n]$$

$$= \lim_{n\to\infty} \frac{1}{n^2} [n^2\, D(X_1) + (n-1)^2 D(X_2) + \ \cdots + D(X_n)]$$

由题意知$\{X_i\}$有有界的方差，不妨设方差为 $0 < D(X_2) = \sigma^2 < \infty$，则

$$\lim_{n\to\infty} \mathrm{D}(\overline{\mathrm{S}_n}) = \lim_{n\to\infty} \frac{\sigma^2}{n^2}(1 + 4 + \cdots + n^2) = \lim_{n\to\infty} \frac{\sigma^2}{n^2} \frac{n(n+1)(2n+1)}{6}$$

$$= \lim_{n\to\infty} \frac{\sigma^2 n}{6} = \infty$$

与 $\lim\limits_{n\to\infty} \mathrm{D}(\overline{\mathrm{S}_n}) \to 0$ 矛盾。

因此，假设不成立，即$\{S_n\}$不满足大数定律。

思路：（2）先证明随机变量$\overline{Y_n}=\frac{1}{n}\sum_{i=1}^{n}a_iS_i$的期望$E(\overline{Y_n})$和方差$D(\overline{Y_n})$存在；再根据切比雪夫不等式$P\{|\overline{Y_n}-E(\overline{Y_n})|<\varepsilon\}\geq 1-\frac{D(\overline{Y_n})}{\varepsilon^2}$可得证。

证明： 因为$na_n\to 0$，所以a_n有界。

不防设存在M，使得$a_n<M$，n=1,2,…

令$\overline{Y_n}=\frac{1}{n}\sum_{i=1}^{n}a_iS_i$，则

$$E(\overline{Y_n})=E\left(\frac{1}{n}\sum_{i=1}^{n}a_iS_i\right)=\frac{1}{n}\sum_{i=1}^{n}a_iE(S_i)$$

$$=\frac{1}{n}\sum_{i=1}^{n}a_i\mu<\frac{nM\mu}{n}=M\mu$$

所以$\overline{Y_n}$的期望存在。

$$D(\overline{Y_n}) = D\left(\frac{1}{n}\sum_{i=1}^{n} a_i S_i\right) = \frac{1}{n^2} D\left(\sum_{i=1}^{n} a_i S_i\right)$$

$$= \frac{1}{n^2} D[a_1X_1 + a_2(X_1 + X_2) + \cdots a_n(X_1 + X_2 + \cdots + X_n)]$$

$$= \frac{1}{n^2} D[(a_1+a_2 + \cdots a_n)X_1 + (a_2 + \cdots a_n)X_2 + \cdots a_nX_n]$$

$$= \frac{1}{n^2} [(a_1+a_2 + \cdots a_n)^2\sigma^2 + (a_2 + \cdots a_n)^2\sigma^2 + \cdots a_n^2\sigma^2]$$

$$= \underbrace{\left[\left(\frac{a_1 + a_2 + \cdots + a_n}{n}\right)^2 + \left(\frac{a_2 + \cdots + a_n}{n}\right)^2 + \cdots + \left(\frac{a_n}{n}\right)^2\right]}_{\to 0}\sigma^2$$

当n → ∞时， $D(\overline{Y_n}) \to 0$

所以 $D(\overline{Y_n})$有界，Y_n的方差存在。

已知$\lim_{n\to\infty} na_n = 0$，以下证明当$n \to +\infty$时，

$$\left(\frac{a_1 + a_2 + \cdots + a_n}{n}\right)^2 + \left(\frac{a_2 + \cdots + a_n}{n}\right)^2 + \cdots + \left(\frac{a_n}{n}\right)^2 \to 0 \quad \text{式(1)}$$

因为$\lim\limits_{n\to\infty} na_n = 0$，所以$na_n$有界，不妨设$\forall n, na_n < A$。

令$f_k = (a_k + a_{k+1} + \cdots + a_n)^2 = \sum_{k\le \mathrm{i}, j\le \mathrm{n}} a_i a_j$

$$= \sum_{k\le \mathrm{i}, j\le \mathrm{n}} \frac{(ia_i)(ja_j)}{ij} < A^2 \sum_{k\le \mathrm{i}, j\le \mathrm{n}} \frac{1}{ij}$$

$$\sim A^2(\ln n - \ln(k-1))^2$$

式(1)左边为

$$\sum_{k=1}^{n} \frac{f_k}{n^2} < \frac{A^2}{n^2}\left((\ln n)^2 + \sum_{k=2}^{n} (\ln n - \ln(k-1))^2\right)$$

$$\sim \mathrm{A}^2 \left(\frac{(\ln n)^2}{n} - \frac{2(n-1)\ln n \ln(n-1)}{n^2} + \frac{2(n-1)\ln n}{n^2}\right) +$$

$$A^2\left(\frac{(n-1)(\ln(n-1))^2}{n^2} - \frac{2(n-1)\ln(n-1)}{n^2} - \frac{2(n-1)}{n^2}\right) \to 0, \text{ 当} n \to +\infty \text{时}.$$

即式(1)成立。

因为$\overline{Y_n}$的期望和方差存在，所以对任意正数$\varepsilon > 0$,
由切比雪夫不等式得

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

即$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} a_i S_i - \frac{1}{n}\sum_{i=1}^{n} E(a_i S_i)\right| < \varepsilon\right\} \geq 1 - \frac{D(Y_n)}{\varepsilon^2}$。

取$n \to \infty$, 则有

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

$$\geq 1 - \frac{\lim\limits_{n\to\infty} D(Y_n)}{\varepsilon^2} = 1$$

所以$\{a_n S_n\}$满足大数定律。