第十七章. 不等式与大数定律.

17.1 不等式.

(讲得乱七八糟看不懂)

17.2 马尔可夫不等式

马尔可夫不等式: 设 X 是一个均值有限的非负随机变量, 均值为 $\mathbb{E}[X]$(这意味着 $\text{Prob}(X < 0) = 0$). 那么, 对于任意的正数 a, 有

$$\text{Prob}(X \geqslant a) \leqslant \frac{\mathbb{E}[X]}{a}.$$

有些作者会把 $\mathbb{E}[X]$ 写成 μ_X. 另一个等价公式是

$$\text{Prob}(X < a) \geqslant 1 - \frac{\mathbb{E}[X]}{a}.$$

仅在 $a \geqslant E(x)$ 时不等式有意义.

当 $a \geqslant E(x)$ 时. $\text{Prob}(x \geqslant a) \leq 1$,

马尔可夫不等式旨在量化小的程度.

证明：

$$\mathrm{Prob}(X \geq a) = \int_a^{+\infty} f_X(x)\,dx$$

$$\leq \int_a^{+\infty} \frac{x}{a} f_X(x)\,dx$$

$$= \frac{1}{a} \int_a^{+\infty} x f_X(x)\,dx$$

$$\leq \frac{1}{a} \cdot \int_0^{+\infty} x f_X(x)\,dx$$

$$= \frac{E(X)}{a}$$

得证.

例. 美国家庭平均收入 60,000 \$. 问出现收入大于 120,000 \$ 的家庭概率是？ 1,000,000 \$ 呢？

$$\mathrm{Pr}(X \geq 120{,}000) \leq \frac{60{,}000}{120{,}000} = 0.5$$

$$\mathrm{Pr}(X \geq 1{,}000{,}000) \leq 0.06.$$

设 X 是一个非负的随机变量，且均值 $\mathbb{E}[X]$ 是有限的. X 的取值不小于 l 乘以均值的概率最多等于 $1/l$:

$$\operatorname{Prob}(X \geqslant l\mathbb{E}[X]) \leqslant \frac{1}{l}.$$

马尔可夫不等式仅要求：

① 随机变量非负.

② $a \geqslant \mu_X$

条件极弱，适用范围极广，但能得到的信息相应很弱.

17.3 切比雪夫不等式.

定理 17.3.1(切比雪夫不等式) 设 X 是一个随机变量，它的均值 μ_X 和方差 σ_X^2 都是有限的. 那么，对于任意的 $k>0$，有

$$\operatorname{Prob}(|X-\mu_X| \geqslant k\sigma_X) \leqslant \frac{1}{k^2}.$$

有些作者会把 μ_X 写成 $\mathbb{E}[X]$. 这意味着，随机变量与均值的距离至少为 k 个标准差的概率不超过 $1/k^2$. 另一个有用的等价公式为

$$\operatorname{Prob}(|X-\mu_X| < k\sigma_X) \geqslant 1-\frac{1}{k^2}.$$

马尔可夫不等式仅输入均值，而切比雪夫不等式还需输入标准差 σ_x.

两个不等式都是通过衡量相对于均值的偏置量测算概率，但马…依靠数量级，切…依靠标准差。

直观理解："随机变量 X 偏出均值 μ k 倍标准差的距离"的概率小于等于 $\frac{1}{k^2}$.

要求 $k>1$，否则无法得到有效信息

逼近效果：

到为

设 $X \geq c\cdot\mu_x$ （这一点同马…式）

$$|X-\mu_x| \geq |c\cdot\mu_x-\mu_x|$$

$$\geq (c-1)\cdot\mu_x.$$

$$\Rightarrow \quad (l-1)\mu_x = k \cdot \sigma_x$$

$$k = \frac{(l-1)\mu_x}{\sigma_x}$$

代入原式，X至少为k倍μ_x的概率

$$\geq \frac{\sigma_x^2}{(l-1)^2\mu_x^2}$$

σ_x和μ_x都是常量，但对比马式，l较大时代数式衰减得更快，得到的范围也更加准确。

证明一：

令 $W = (X-\mu_x)^2$

$$E(W) = \int_{w=0}^{+\infty} w\, f_W(w)\, dw$$

$$= \int_{x=-\infty}^{+\infty} (x-\mu_x)^2 \cdot f_X(x)\, dx$$

$$= \sigma_x^2$$

现在对 w 使用马尔可夫不等式：

$$\text{Prob}(w \geq a) \leq \frac{E(w)}{a}$$

$$\text{Prob}((x-\mu_x)^2 \geq a) \leq \frac{E(w)}{a}$$

$$\Rightarrow \text{Prob}(|x-\mu_x| \geq \sqrt{a}) \leq \frac{E(w)}{a}$$

令 $\sqrt{a} = k\sigma_x$，代入，有

$$\text{Prob}(|x-\mu_x| \geq k\sigma_x) \leq \frac{1}{k^2}$$

得证

证明二：（直接证明法）

$$\text{Prob}(|X-\mu_x| \geq k\sigma_x)$$

$$= \int_{|x-\mu_x| \geq k\sigma_x} 1 \cdot f_x(x)\,dx$$

$$\leq \int_{|x-\mu_x| \geq k\sigma_x} \frac{(X-\mu_x)^2}{k^2\sigma_x^2} f_x(x)\,dx$$

$$\leq \frac{1}{k^2\sigma^2_X}\int_{-\infty}^{+\infty}(x-\mu_X)^2 f_X(x)\,dx$$

$$= \frac{1}{k^2}$$

得证

17.3.3 指数分布与均匀分布的例子.

$X \sim Unif(0,1)$, $Y \sim Exp(1)$

$$\begin{cases} E(X) = \frac{1}{2}, & \sigma^2_X = \frac{1}{12} \\ E(Y) = 1, & \sigma^2_Y = 1 \end{cases}$$

考察 $X \geq 0.95$ 和 $Y \geq 4$ 的概率.

马: $c_1 = \frac{0.5}{0.95} \doteq 0.55$

$Pr_1(X \geq 0.95) \leq 0.55$.

$c_2 = \frac{1}{4} = 0.25$

$$Pr_1(Y \geq 4) \leq 0.25.$$

切：把想求的概率写成 $|X-\mu_x|<k\sigma_x$ 进行计算.

X：$10.95-0.51 = k\cdot\frac{1}{\sqrt{12}}$

$$k \doteq 1.56.$$

$$Pr_1 \leq \frac{1}{k^2} \doteq 0.41.$$

Y：$14-11 = k\cdot 1$

$$k = 3$$

$$Pr_2 \leq \frac{1}{9} \doteq 0.11$$

对于衰减得过快的正态分布、切比雪夫的上切就比实际值高得过头了。因为切比雪夫也是一个弱条件不等式

17-4 布尔不等式与邦弗伦尼不等式

布尔不等式： 我们有

$$\mathrm{Prob}\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} \mathrm{Prob}(A_i).$$

对于可数多个 A_i, 这个结果仍然成立.

记 $S_k = \sum_{1 \le a_1 < a_2 < \cdots < a_k \le n} \mathrm{Prob}(a_1 \cap a_2 \cap a_3 \cdots \cap a_k)$

→ 有 $\binom{n}{k}$ 个数.

邦弗伦尼不等式： 对于上述 S_k 以及正整数 l 和 m, 我们有

$$\sum_{k=1}^{2l} (-1)^{k-1} S_k \leqslant \mathrm{Prob}\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{k=1}^{2m-1} (-1)^{k-1} S_k.$$

例. 四个人拿牌. 至少一个人拿到全同花的概率是多少?

解: 记 A_i 表示第 i 个人拿到同花.

$$\mathrm{Prob}(A_i) = \frac{\binom{4}{1}\binom{13}{13}}{\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}}$$

$$\text{Prob}(A_i \cap A_j) = \frac{4 \times 3 \times \binom{13}{13}\binom{13}{13}}{\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}}$$

$$S_1 = \sum_{i=1}^{4} \text{Prob}(A_i) = 4 \cdot \text{Prob}(A_i)$$

$$S_2 = \sum_{1 \leq i < j \leq 4} \text{Prob}(A_i \cap A_j) = 6\, \text{Prob}(A_i \cap A_j)$$

$$S_1 - S_2 \leqslant P(\text{至少1人拿同花}) \leqslant S_1$$

$$0.000\,000\,000\,025\,184\,2 \leqslant \text{Prob}\left(\bigcup_{i=1}^{4} A_i\right) \leqslant 0.000\,000\,000\,025\,196\,4.$$

17.5 收敛类型

17.5.1 依分布收敛

依分布收敛 (或弱收敛)：设 $X, X_1, X_2, \cdots$ 都是随机变量, 它们的累积分布函数分别是 $F, F_1, F_2, \cdots$. 设 $\mathcal{C}$ 是由 F 的连续点构成的实数集. 如果 $\lim_{n\to\infty} F_n(x) = F(x)$ 对所有的 $x \in \mathcal{C}$ 均成立, 那么随机变量序列 $X_1, X_2, \cdots$ **依分布收敛**(或**弱收敛**) 于随机变量 X. 换句话说, 如果 F 在 x 处是连续的, 那么累积分布函数列在 x 处的极限就等于 $F(x)$. 这一结论通常被记作 $X_n \xrightarrow{d} X$ 或者 $X_n \xrightarrow{D} X$. 如果知道随机变量 X 的类型, 我们有时会把 X 替换成它服从的分布. 因此, 用 $X_n \xrightarrow{d} N(0,1)$ 表示收敛于一个服从标准正态分布的随机变量, 用 $X_n \xrightarrow{d} \mathrm{Exp}(2)$ 表示收敛于一个服从参数为 2 的指数分布的随机变量.

服从均匀分布的随机变量 X, 有

$$f_X(x) = \begin{cases} \frac{1}{n} & x \in \{0, \frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-1}{n}\} \\ 0 & \text{其它} \end{cases}$$

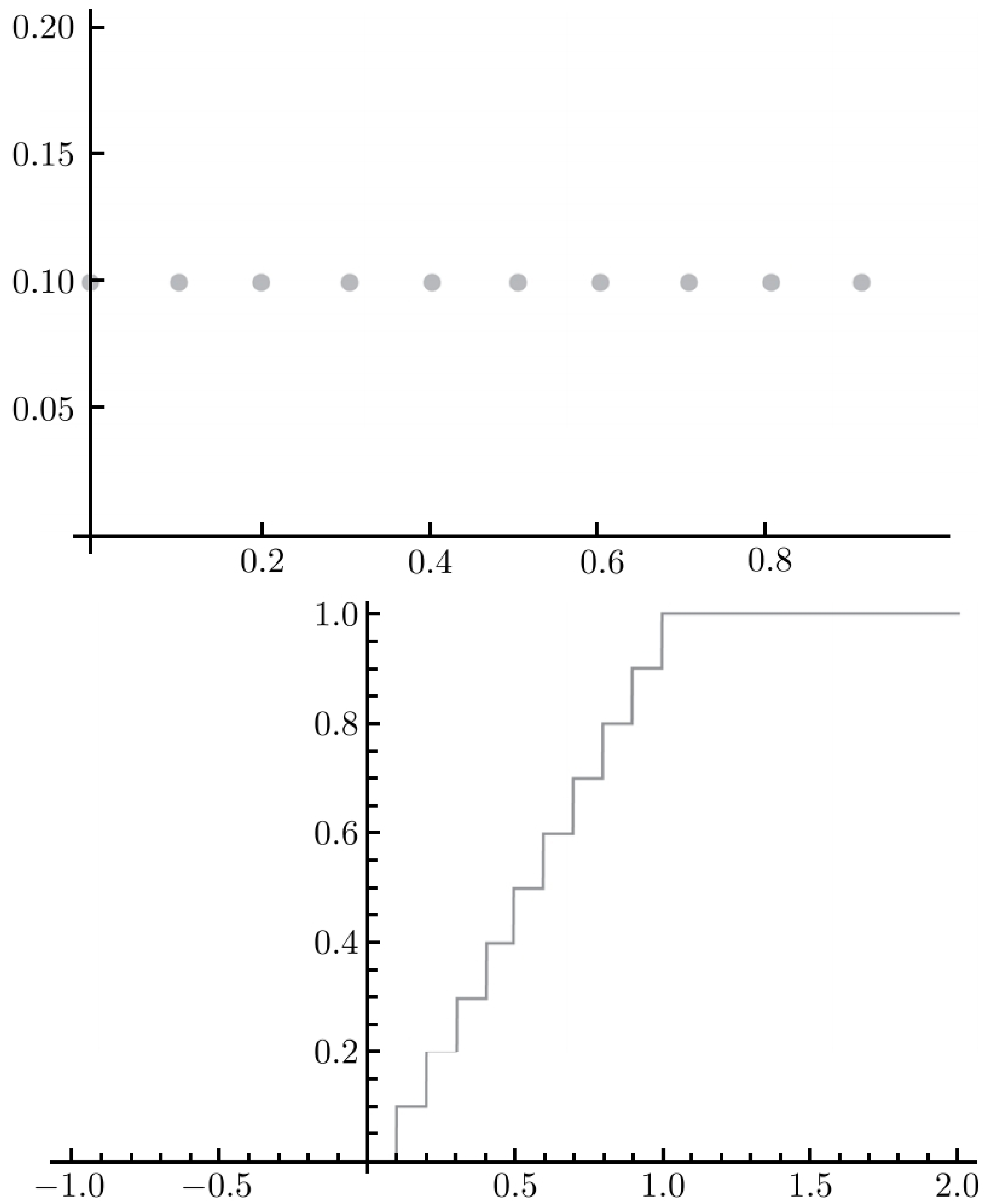

图 17-1　当 $n = 10$ 时, 离散均匀分布的 PDF(上图) 和 CDF(下图)

对于 pdf，当 $n \to \infty$ 时，无论是否取一些特殊的值，$f_X(x)$ 都 $\to 0$；

cdf 则截然不同，观察到

$$F_n(x) = \sum_{\substack{0 \le k \le n-1 \\ \frac{k}{n} \le x}} \frac{1}{n} = \frac{\lfloor nx \rfloor}{n}$$

当 $n \to \infty$ 时，$F_n(x) \to x$，这正是服从 $[0,1]$ 上均匀分布的累积分布函数

$$\Rightarrow \quad X_n \xrightarrow{d} \mathrm{Unif}(0,1)$$

17.5.2 依概率收敛

依概率收敛：设 $X, X_1, X_2, \cdots$ 都是随机变量. 如果对于任意的 $\epsilon > 0$, 都有

$$\lim_{n \to \infty} \mathrm{Prob}(|X_n - X| \geqslant \epsilon) = 0,$$

那么说序列 $\{X_n\}_{n=1}^{\infty}$ 依概率收敛于 X, 并记作 $X_n \xrightarrow{p} X$ 或者 $X_n \xrightarrow{P} X$.

翻译：$n \to \infty$ 时，X_n 与 X 的距离大于任何正数均为不可能事件

$\Longleftrightarrow$ $|X_n - X|$ 要"足够小"

例、$X_n \sim N(1701, \frac{1}{n^2})$，断言：

$X_n \xrightarrow{P} 1701$，证明之。

证：

$$\lim_{n \to \infty} Prob(|X_n - 1701| \geq \varepsilon) = 0.$$

由上式左边联想到切比雪夫不等式

$$Prob(|X_n - 1701| \geq k\sigma) = \frac{1}{k^2}$$

令 $\varepsilon = k\sigma$，有

$$Prob(|X_n - 1701| \geq \varepsilon) = \frac{\sigma^2}{\varepsilon^2}$$

$$= \frac{1}{n^2 \varepsilon^2}$$

$$\lim_{n\to\infty} \cdots = 0.$$

得证

17.5.3 几乎必然收敛与必然收敛.

几乎必然收敛：设 $(\Omega, \mathcal{F}, \text{Prob})$ 是一个概率空间，设 $X, X_1, X_2, \cdots$ 都是随机变量. 如果

$$\text{Prob}(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1,$$

那么说 $\{X_n\}_{n=1}^{\infty}$ **几乎必然**(或者**几乎处处**, 又或者**以概率 1** 收敛于 X), 并记作 $X_n \xrightarrow{a.s.} X$.

必然收敛：设 $(\Omega, \mathcal{F}, \text{Prob})$ 是一个概率空间，设 $X, X_1, X_2, \cdots$ 都是随机变量. 如果对于所有的 $\omega \in \Omega$, 均有

$$\lim_{n\to\infty} X_n(\omega) = X(\omega),$$

那么说 $\{X_n\}_{n=1}^{\infty}$ 必然收敛于 X.

17.6 弱大数定律与强大数定律

弱大数定律：设 $X_1, X_2, \cdots$ 是独立同分布的随机变量，且均值为 μ，并设 $\bar{X}_n = \frac{X_1+\cdots+X_n}{n}$，那么 $\bar{X}_n \xrightarrow{p} \mu$(即 $\bar{X}_n$ 依概率收敛于 μ).

证明：

$$E(\overline{X_n}) = E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)$$

$$= \frac{1}{n}\cdot[E(X_1)+E(X_2)+\cdots+E(X_n)]$$

$$= \frac{1}{n}\cdot n\mu \quad (\text{同分布})$$

$$= \mu.$$

$$Var(\overline{X_n}) = Var\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)$$

$$= \frac{1}{n^2} Var(X_1+X_2+\cdots+X_n)$$

$$= \frac{\sigma_x^2}{n}$$

当 $n \to \infty$，有 $Var(\overline{X_n}) \to 0$.

故收敛，且收敛到 μ. 这并不是严格的证明

再严格证明一次 $X_n \xrightarrow{P} \mu$：

$$\text{Prob}(|X_n - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$\forall \varepsilon > 0$，令 $\varepsilon = k\sigma$

有 $k = \frac{\varepsilon}{\sigma}$. 则

$$\text{Prob}(|X_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

$$= \frac{\sigma_x^2}{n\varepsilon^2}$$

$$\lim_{n\to\infty} \text{Prob}(|X_n - \mu| \geq \varepsilon) = 0.$$

得证.

强大数定律：设 $X_1, X_2, \cdots$ 是独立同分布的随机变量，且均值为 μ，并设 $\bar{X}_n = \frac{X_1+\cdots+X_n}{n}$，那么 $\bar{X}_n \xrightarrow{a.s.} \mu$.