第四章　极限定理

Tianxiao Pang

Zhejiang University

September 12, 2021

内容

内容

内容

在第一章我们曾经指出：在大量重复试验中，随机事件出现的频率在某个常数附近摆动，这就是所谓的“频率稳定性”．但对于这点，我们至今尚未给予理论上的说明．

Bernoulli首先给出了“频率稳定性”的理论证明，由此建立了概率论的第一个极限定理—-大数定律．

德莫佛和拉普拉斯提出将观察的误差看作大量独立微小误差的累加，证明了观察误差的分布一定渐近正态，由此得到中心极限定理．

随后，出现了许多各种收敛意义下的极限定理．本章主要介绍大数定律和中心极限定理等有关内容．

依分布收敛与中心极限定理

分布函数可以刻画随机变量的全部概率性质. 我们首先来研究分布函数的极限行为.

依分布收敛与中心极限定理

分布函数可以刻画随机变量的全部概率性质. 我们首先来研究分布函数的极限行为.

4.1.1 分布函数弱收敛

例1 令

$$F_n(x) = \begin{cases} 0, & x < 1/n, \\ 1, & x \geq 1/n, \end{cases}$$

这是一个退化分布, 它可以解释为一个单位质量全部集中在$x = 1/n$这一点的分布. 当$n \to \infty$时, 我们自然认为$F_n(x)$应该收敛于一个单位质量全部集中在$x = 0$这一点的分布, 即

$$F(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$$

依分布收敛与中心极限定理

分布函数可以刻画随机变量的全部概率性质. 我们首先来研究分布函数的极限行为.

4.1.1 分布函数弱收敛

例1 令

$$F_n(x) = \begin{cases} 0, & x < 1/n, \\ 1, & x \geq 1/n, \end{cases}$$

这是一个退化分布, 它可以解释为一个单位质量全部集中在$x = 1/n$这一点的分布. 当$n \to \infty$时, 我们自然认为$F_n(x)$应该收敛于一个单位质量全部集中在$x = 0$这一点的分布, 即

$$F(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$$

但是, $F_n(0) \equiv 0$, 而$F(0) = 1$, 显然, $F_n(0) \nrightarrow F(0)$.

因此看来要求分布函数列在所有的点都收敛到极限分布函数是太严格了. 这个例子中不收敛的点是极限分布函数$F(x)$的不连续点.

因此看来要求分布函数列在所有的点都收敛到极限分布函数是太严格了. 这个例子中不收敛的点是极限分布函数$F(x)$的不连续点.

定义（分布函数弱收敛）

设F是一分布函数, $\{F_n, n \geq 1\}$是一列分布函数, 如果对F的每个连续点$x \in \mathbb{R}$, 都有$F_n(x) \to F(x)$ $(n \to \infty)$, 则称F_n弱收敛(weak convergence)于F, 记作$F_n \xrightarrow{w} F$.

注: 分布函数逐点收敛的极限函数未必是分布函数. 例如,

$$F_n(x) = \begin{cases} 0, & x < n, \\ 1, & x \geq n. \end{cases}$$

该分布函数列处处收敛于

注: 分布函数逐点收敛的极限函数未必是分布函数. 例如,

$$F_n(x) = \begin{cases} 0, & x < n, \\ 1, & x \geq n. \end{cases}$$

该分布函数列处处收敛于0, 但极限函数$G(x) \equiv 0$不是分布函数.

注: 分布函数逐点收敛的极限函数未必是分布函数. 例如,

$$F_n(x) = \begin{cases} 0, & x < n, \\ 1, & x \geq n. \end{cases}$$

该分布函数列处处收敛于0, 但极限函数$G(x) \equiv 0$不是分布函数.

定义（依分布收敛）

设ξ为一随机变量，$\{\xi_n, n \geq 1\}$是一列随机变量，如果ξ_n的分布函数弱收敛于ξ的分布函数，则称ξ_n依分布收敛(convergence in distribution）于ξ，记作$\xi_n \xrightarrow{d} \xi$.

定理（Helly第一定理）

设$\{F_n, n \geq 1\}$是一列分布函数, 那么存在一个单调不减右连续的函数F(不一定是分布函数), $0 \leq F(x) \leq 1$, $x \in \mathbb{R}$, 和一列子列$\{F_{n_k}, k \geq 1\}$, 使得对F的每个连续点x, $F_{n_k}(x) \to F(x)$ $(k \to \infty)$.

证明: 略.

定理（Helly第二定理）

设F是一分布函数，$\{F_n, n \geq 1\}$是一列分布函数，$F_n \xrightarrow{w} F$. 如果$g(x)$是$\mathbb{R}$上的有界连续函数，则

$$\int_{-\infty}^{\infty} g(x)\mathrm{d}F_n(x) \to \int_{-\infty}^{\infty} g(x)\mathrm{d}F(x).$$

设F, F_n是单调不减右连续函数(不一定是分布函数)，并且对F的任一连续点x有$F_n(x) \to F(x)$. 如果$a < b$是F的连续点，$g(x)$是$[a, b]$上的连续函数，则

$$\int_a^b g(x)\mathrm{d}F_n(x) \to \int_a^b g(x)\mathrm{d}F(x).$$

证明：略.

若在Helly第二定理中取$g(x) = e^{\mathrm{i}tx}$, 则可得到下面的Lévy连续性定理:

若在Helly第二定理中取$g(x)=e^{itx}$, 则可得到下面的Lévy连续性定理:

定理 (Lévy连续性定理/正极限定理)

设F是一分布函数, $\{F_n, n\geq 1\}$是一列分布函数. 如果$F_n \xrightarrow{w} F$. 则相应的特征函数$\{f_n(t), n\geq 1\}$收敛于F的特征函数$f(t)$, 且在t 的任一有限区间内收敛是一致的.

若在Helly第二定理中取$g(x) = e^{itx}$, 则可得到下面的Lévy连续性定理:

定理（Lévy连续性定理/正极限定理）

设F是一分布函数, $\{F_n, n \geq 1\}$是一列分布函数. 如果$F_n \xrightarrow{w} F$. 则相应的特征函数$\{f_n(t), n \geq 1\}$收敛于F的特征函数$f(t)$, 且在t 的任一有限区间内收敛是一致的.

证明: 在Helly第二定理中取$g_t(x) = e^{itx}$, $x \in \mathbb{R}$. 显然$g_t(x)$是$\mathbb{R}$上的有界连续函数, 所以收敛性得证. 至于在任一有限区间上的一致收敛性, 证明方法可参考Helly第二定理的证明, 或参考《概率论》(第二版, 苏淳著).

在第三章, 我们已经知道特征函数和分布函数可以相互唯一确定. 同样, Lévy连续性定理的逆命题也成立.

在第三章, 我们已经知道特征函数和分布函数可以相互唯一确定. 同样, Lévy连续性定理的逆命题也成立.

定理 (逆极限定理)

设$\{f_n(t), n \geq 1\}$是分布函数$\{F_n(x), n \geq 1\}$的特征函数, 如果对每一t, $f_n(t) \to f(t)$, 且$f(t)$在$t = 0$处连续, 则$f(t)$一定是某个分布函数F的特征函数, 且$F_n \xrightarrow{w} F$.

证明: 略.

设随机变量序列$\{\xi_n, n \geq 1\}$的特征函数为$\{f_n(t), n \geq 1\}$, ξ的特征函数为$f(t)$. 则由正极限定理和逆极限定理知

$$\xi_n \xrightarrow{d} \xi \iff f_n(t) \to f(t), \ \forall t \in \mathbb{R}.$$

例2 用特征函数方法证明二项分布的Poisson逼近定理.

例2 用特征函数方法证明二项分布的Poisson逼近定理.

证明: 设$\xi_n \sim B(n, p_n)$, 且$\lim_{n\to\infty} np_n = \lambda$. 它的特征函数为

$$f_n(t) = (p_n e^{\mathrm{i}t} + q_n)^n, \quad q_n = 1 - p_n.$$

例2 用特征函数方法证明二项分布的Poisson逼近定理.

证明: 设$\xi_n \sim B(n, p_n)$, 且$\lim_{n\to\infty} np_n = \lambda$. 它的特征函数为

$$f_n(t) = (p_n e^{\mathrm{i}t} + q_n)^n, \quad q_n = 1 - p_n.$$

因此

$$\lim_{n\to\infty} f_n(t) = \lim_{n\to\infty} \left(1 + \frac{np_n(e^{\mathrm{i}t} - 1)}{n}\right)^n = e^{\lambda(e^{\mathrm{i}t}-1)},$$

这正是Poisson分布的特征函数.

例2 用特征函数方法证明二项分布的Poisson逼近定理.

证明: 设$\xi_n \sim B(n, p_n)$, 且$\lim_{n\to\infty} np_n = \lambda$. 它的特征函数为

$$f_n(t) = (p_n e^{\mathrm{i}t} + q_n)^n, \quad q_n = 1 - p_n.$$

因此

$$\lim_{n\to\infty} f_n(t) = \lim_{n\to\infty} \left(1 + \frac{np_n(e^{\mathrm{i}t} - 1)}{n}\right)^n = e^{\lambda(e^{\mathrm{i}t} - 1)},$$

这正是Poisson分布的特征函数. 所以由逆极限定理, ξ_n依分布收敛于$\xi \sim P(\lambda)$.

例3 设$\xi_n \sim P(\lambda_n)$，且$\lambda_n \to \infty$. 证明

$$\frac{\xi_n - \lambda_n}{\sqrt{\lambda_n}} \xrightarrow{d} N(0,1).$$

例3 设$\xi_n \sim P(\lambda_n)$, 且$\lambda_n \to \infty$. 证明

$$\frac{\xi_n - \lambda_n}{\sqrt{\lambda_n}} \xrightarrow{d} N(0,1).$$

证明: 因为ξ_n的特征函数为$e^{\lambda_n(e^{\mathrm{i}t}-1)}$, 所以$\frac{\xi_n-\lambda_n}{\sqrt{\lambda_n}}$的特征函数为

$$\begin{aligned} f_n(t) &= \exp\left\{\lambda_n(e^{\mathrm{i}t/\sqrt{\lambda_n}} - 1)\right\} \cdot \exp\{-\mathrm{i}t\sqrt{\lambda_n}\} \\ &= \exp\left\{\lambda_n\left(\frac{\mathrm{i}t}{\sqrt{\lambda_n}} - \frac{t^2}{2\lambda_n} + o(\frac{1}{\lambda_n})\right)\right\} \cdot \exp\{-\mathrm{i}t\sqrt{\lambda_n}\} \\ &= \exp\left\{-\frac{t^2}{2} + o(1)\right\} \to e^{-t^2/2}, \end{aligned}$$

后者正是$N(0,1)$分布的特征函数. 所以由逆极限定理, 可知结论成立.

例3 设$\xi_n \sim P(\lambda_n)$，且$\lambda_n \to \infty$. 证明

$$\frac{\xi_n - \lambda_n}{\sqrt{\lambda_n}} \xrightarrow{d} N(0,1).$$

证明: 因为ξ_n的特征函数为$e^{\lambda_n(e^{\mathrm{i}t}-1)}$，所以$\frac{\xi_n-\lambda_n}{\sqrt{\lambda_n}}$的特征函数为

$$\begin{aligned} f_n(t) &= \exp\left\{\lambda_n(e^{\mathrm{i}t/\sqrt{\lambda_n}} - 1)\right\} \cdot \exp\{-\mathrm{i}t\sqrt{\lambda_n}\} \\ &= \exp\left\{\lambda_n\left(\frac{\mathrm{i}t}{\sqrt{\lambda_n}} - \frac{t^2}{2\lambda_n} + o(\frac{1}{\lambda_n})\right)\right\} \cdot \exp\{-\mathrm{i}t\sqrt{\lambda_n}\} \\ &= \exp\left\{-\frac{t^2}{2} + o(1)\right\} \to e^{-t^2/2}, \end{aligned}$$

后者正是$N(0,1)$分布的特征函数. 所以由逆极限定理，可知结论成立.

4.1.2 性质

除lévy连续性定理外, 分布函数弱收敛还有下列性质:

4.1.2 性质

除lévy连续性定理外，分布函数弱收敛还有下列性质：

性质1：设$\{F_n, n \geq 1\}$是一列分布函数，如果$F_n \xrightarrow{w} F$，F是一连续的分布函数，则F_n在$\mathbb{R}$上一致收敛于$F(x)$.

4.1.2 性质

除lévy连续性定理外，分布函数弱收敛还有下列性质：

性质1：设$\{F_n, n \geq 1\}$是一列分布函数，如果$F_n \xrightarrow{w} F$，F是一连续的分布函数，则F_n在$\mathbb{R}$上一致收敛于$F(x)$.

*证明：令

$$x_0 = -\infty < x_1 = -M < x_2 < \cdots < x_{K-1} = M < x_K = \infty.$$

若$x \in [x_k, x_{k+1}]$，可知

$$\begin{aligned} F_n(x) - F(x) \leq & F_n(x_{k+1}) - F(x_k) \\ \leq & |F_n(x_{k+1}) - F(x_{k+1})| + F(x_{k+1}) - F(x_k), \end{aligned}$$

以及

$$\begin{aligned} F(x)-F_n(x) \le & F(x_{k+1})-F_n(x_k) \\ \le & |F_n(x_k)-F(x_k)|+F(x_{k+1})-F(x_k). \end{aligned}$$

结合以上两个不等式可知

$$\begin{aligned} & \sup_{x\in\mathbb{R}} |F_n(x)-F(x)| \\ \le & \max_{0\le k\le K} |F_n(x_k)-F(x_k)| + \max_{0\le k\le K-1} [F(x_{k+1})-F(x_k)]. \end{aligned}$$

以及

$$\begin{aligned} F(x) - F_n(x) \leq & F(x_{k+1}) - F_n(x_k) \\ \leq & |F_n(x_k) - F(x_k)| + F(x_{k+1}) - F(x_k). \end{aligned}$$

结合以上两个不等式可知

$$\begin{aligned} & \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \\ \leq & \max_{0\leq k\leq K} |F_n(x_k) - F(x_k)| + \max_{0\leq k\leq K-1} [F(x_{k+1}) - F(x_k)]. \end{aligned}$$

因为F是连续函数, 所以对任意的$\varepsilon > 0$, 可以找到$\{x_0, \cdots, x_K\}$使得

$$\begin{aligned} & |F(x_0) - F(x_1)| = F(-M) < \varepsilon, \\ & |F(x_K) - F(x_{K-1})| = 1 - F(M) < \varepsilon, \\ & |F(x_{k+1}) - F(x_k)| < \varepsilon, \ \forall\, k = 1, \cdots, K-2. \end{aligned}$$

因此可知

$$\sup_{x\in\mathbb{R}}|F_n(x)-F(x)|\le \max_{0\le k\le K}|F_n(x_k)-F(x_k)|+\varepsilon.$$

因此可知

$$\sup_{x\in\mathbb{R}}|F_n(x)-F(x)|\le\max_{0\le k\le K}|F_n(x_k)-F(x_k)|+\varepsilon.$$

又因为对任意的x_k，存在一个N_k，使得当$n\ge N_k$时，

$$|F_n(x_k)-F(x_k)|<\varepsilon.$$

因此当$n>\max\{N_k,k=0,\cdots,K\}$时，

$$\sup_{x\in\mathbb{R}}|F_n(x)-F(x)|\le 2\varepsilon.$$

性质2: 设ξ是一随机变量, $\{\xi_n, n \geq 1\}$是一列随机变量, $g(x)$是$\mathbb{R}$上的连续函数. 如果$\xi_n \xrightarrow{d} \xi$, 则$g(\xi_n) \xrightarrow{d} g(\xi)$.

性质2: 设ξ是一随机变量, $\{\xi_n, n \geq 1\}$是一列随机变量, $g(x)$是$\mathbb{R}$上的连续函数. 如果$\xi_n \xrightarrow{d} \xi$, 则$g(\xi_n) \xrightarrow{d} g(\xi)$.

证明: 假设ξ和$\{\xi_n, n \geq 1\}$的分布函数为别为F和$\{F_n, n \geq 1\}$. 若$\xi_n \xrightarrow{d} \xi$, 则$F_n \xrightarrow{w} F$. 由Helly第二定理知

$$\int_{-\infty}^{\infty} e^{\mathrm{i}tg(x)} \mathrm{d}F_n(x) \to \int_{-\infty}^{\infty} e^{\mathrm{i}tg(x)} \mathrm{d}F(x).$$

再由逆极限定理可知$g(\xi_n)$的分布函数弱收敛于$g(\xi)$的分布函数, 即$g(\xi_n) \xrightarrow{d} g(\xi)$.

性质3: 设$\{a_n, n \geq 1\}$和$\{b_n, n \geq 1\}$是两列常数, F是一分布函数. $\{F_n, n \geq 1\}$是一列分布函数. 如果$a_n \to a$, $b_n \to b$, $F_n \xrightarrow{w} F$, 则

$$F_n(a_n x + b_n) \to F(ax + b),$$

其中x使得$ax + b$是F的连续点.

性质3: 设$\{a_n, n\geq 1\}$和$\{b_n, n\geq 1\}$是两列常数, F是一分布函数. $\{F_n, n\geq 1\}$是一列分布函数. 如果$a_n\to a$, $b_n\to b$, $F_n\xrightarrow{w} F$, 则

$$F_n(a_n x+b_n)\to F(ax+b),$$

其中x使得$ax+b$是F的连续点.

证明: 令$\varepsilon>0$使得F在$ax+b\pm\varepsilon$处连续(因为F的连续点在$\mathbb{R}$上是稠密的). 显然$a_nx+b_n\to ax+b$, 因此对充分大的n有

$$ax+b-\varepsilon\leq a_nx+b_n\leq ax+b+\varepsilon.$$

因此

$$F_n(ax+b-\varepsilon)\leq F_n(a_nx+b_n)\leq F_n(ax+b+\varepsilon).$$

由于$F_n \xrightarrow{w} F$, 则

$$\begin{aligned}F(ax+b-\varepsilon) &\le \liminf_{n\to\infty} F_n(a_n x + b_n)\\ &\le \limsup_{n\to\infty} F_n(a_n x + b_n)\\ &\le F(ax+b+\varepsilon).\end{aligned}$$

令$\varepsilon \to 0$, 由于F在$ax+b$处连续, 所以结论得证.

推论

若$\xi_n \xrightarrow{d} \xi$, $a_n \to a$, $b_n \to b$, 则$a_n\xi_n + b_n \xrightarrow{d} a\xi + b$, $(a_n, a \neq 0)$.

推论

若$\xi_n \xrightarrow{d} \xi$, $a_n \to a$, $b_n \to b$, 则$a_n\xi_n + b_n \xrightarrow{d} a\xi + b$, $(a_n, a \neq 0)$.

证明: 不妨假设$a > 0$($a < 0$情形类似可证). 由于$a_n \to a$, 所以当n充分大时, $a_n > 0$. 此时, 对于任意的使得$\frac{x-b}{a}$为F的连续点的实数x, 根据性质3, 有

$$\mathsf{P}(a_n\xi_n + b_n \le x) = F_n\left(\frac{x - b_n}{a_n}\right) \to F\left(\frac{x - b}{a}\right) = \mathsf{P}(a\xi + b \le x).$$

所以$a_n\xi_n + b_n \xrightarrow{d} a\xi + b$.

4.1.3 中心极限定理(Central Limit Theorem/CLT)

4.1.3 中心极限定理(Central Limit Theorem/CLT)

考虑一个n重Bernoulli试验, 每次试验中成功的概率为p $(0 < p < 1)$, 记S_n为n重Bernoulli试验中成功的总次数. 显然

$$\mathsf{P}(S_n = k) = b(k; n, p).$$

人们往往对下列的概率大小感兴趣:

$$\mathsf{P}(\alpha \le S_n \le \beta) = \sum_{\alpha \le k \le \beta} b(k; n, p),$$

其中α和β为非负整数. 直接计算上述的和式十分困难.

定理 (德莫佛-拉普拉斯定理)

设$\Phi(x)$为标准正态随机变量的分布函数. 则对$-\infty < x < \infty$, 有

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{npq}} \le x\right) = \Phi(x),$$

其中$q = 1 - p$.

定理 (德莫佛-拉普拉斯定理)

设$\Phi(x)$为标准正态随机变量的分布函数. 则对$-\infty < x < \infty$, 有

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{npq}} \leq x\right) = \Phi(x),$$

其中$q = 1 - p$.

事实上, $\frac{S_n-np}{\sqrt{npq}}$是$S_n$的标准化随机变量. 德莫佛-拉普拉斯定理告诉我们, 当n充分大时, 可把$\frac{S_n-np}{\sqrt{npq}}$近似看成$N(0,1)$.

定理的应用:

定理的应用:

$$
\begin{aligned}
\mathrm{P}(\alpha \le S_n \le \beta) =& \mathrm{P}\left(\frac{\alpha - np}{\sqrt{npq}} \le \frac{S_n - np}{\sqrt{npq}} \le \frac{\beta - np}{\sqrt{npq}}\right) \\
\approx& \Phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha - np}{\sqrt{npq}}\right). \qquad (4.1.1)
\end{aligned}
$$

定理的应用:

$$\begin{aligned}\mathsf{P}(\alpha \le S_n \le \beta) =& \mathsf{P}\left(\frac{\alpha - np}{\sqrt{npq}} \le \frac{S_n - np}{\sqrt{npq}} \le \frac{\beta - np}{\sqrt{npq}}\right) \\ \approx& \Phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha - np}{\sqrt{npq}}\right). \end{aligned} \tag{4.1.1}$$

注1: 德莫佛-拉普拉斯定理与第二章中的二项分布的泊松逼近定理并无矛盾. 在泊松逼近定理中, 要求$np_n \to \lambda$, 而德莫佛-拉普拉斯定理中要求p是常数. 在实际应用中, 当n很大时,

- 若p大小适中, 用正态分布逼近二项分布;
- 若p接近0或者1, 且np较小或较大, 此时二项分布的图像偏斜度太大, 宜用泊松分布去逼近二项分布.

注2: 实际计算中, 若n不是很大, 则通常把(4.1.1)修正为

$$\Phi\left(\frac{\beta+0.5-np}{\sqrt{npq}}\right)-\Phi\left(\frac{\alpha-0.5-np}{\sqrt{npq}}\right).$$

理由如下:

注2: 实际计算中, 若n不是很大, 则通常把(4.1.1)修正为

$$\Phi\left(\frac{\beta+0.5-np}{\sqrt{npq}}\right)-\Phi\left(\frac{\alpha-0.5-np}{\sqrt{npq}}\right).$$

理由如下:

由于α和β是非负整数, 所以

$$\mathsf{P}(\alpha\le S_n\le\beta)\approx\Phi\left(\frac{\beta-np}{\sqrt{npq}}\right)-\Phi\left(\frac{\alpha-np}{\sqrt{npq}}\right)$$

以及

$$\begin{aligned}\mathsf{P}(\alpha\le S_n\le\beta)=&\mathsf{P}(\alpha-1<S_n<\beta+1)\\ \approx&\Phi\left(\frac{\beta+1-np}{\sqrt{npq}}\right)-\Phi\left(\frac{\alpha-1-np}{\sqrt{npq}}\right).\end{aligned}$$

因此加减0.5是合理的.

例4 抛一枚质量均匀的硬币，需要抛多少次才能保证出现正面的频率在$(0.4, 0.6]$之间的概率不小于90%?

例4 抛一枚质量均匀的硬币，需要抛多少次才能保证出现正面的频率在$(0.4, 0.6]$之间的概率不小于90%?

解：令n为抛掷次数，S_n为出现正面的次数．则$S_n \sim B(n, 0.5)$．由题意，我们需要求n使得

$$\mathsf{P}(0.4 < S_n/n \leq 0.6) \geq 0.9.$$

例4 抛一枚质量均匀的硬币，需要抛多少次才能保证出现正面的频率在$(0.4, 0.6]$之间的概率不小于90%?

解：令n为抛掷次数，S_n为出现正面的次数．则$S_n \sim B(n, 0.5)$．由题意，我们需要求n使得

$$\mathsf{P}(0.4 < S_n/n \le 0.6) \ge 0.9.$$

利用德莫佛-拉普拉斯定理，

$$\begin{aligned}
&\mathsf{P}(0.4 < S_n/n \le 0.6)\\
=&\mathsf{P}\left(\frac{0.4n - 0.5n}{\sqrt{n/4}} < \frac{S_n - 0.5n}{\sqrt{n/4}} \le \frac{0.6n - 0.5n}{\sqrt{n/4}}\right)\\
\approx&\Phi(0.2\sqrt{n}) - \Phi(-0.2\sqrt{n})\\
=&2\Phi(0.2\sqrt{n}) - 1.
\end{aligned}$$

现在问题转化为求n使得

$$2\Phi(0.2\sqrt{n}) - 1 \geq 0.9.$$

易知当$n \geq 69$时, 满足要求.

定义 (中心极限定理)

设$\{\xi_n, n \geq 1\}$是一列随机变量. 如果存在常数列$B_n > 0$以及A_n, 使得

$$\frac{\sum_{k=1}^{n} \xi_k - A_n}{B_n} \xrightarrow{d} N(0,1),$$

则称$\{\xi_n, n \geq 1\}$满足中心极限定理(Central Limit Theorem).

通常, 我们称A_n为中心化因子, 称B_n为正则化因子, 称$\sum_{k=1}^{n} \xi_k$为部分和(partial sums).

定理（Lindeberg-Lévy CLT）

设$\{\xi_n, n \geq 1\}$是一列独立同分布的(independent and identically distributed, 以后记成i.i.d.)随机变量序列. 记

$$S_n = \sum_{k=1}^{n} \xi_k, \ E\xi_1 = a, \ Var\xi_1 = \sigma^2 (0 < \sigma^2 < \infty),$$

则CLT成立, 即

$$\frac{S_n - na}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1).$$

证明: 我们利用特征函数来证明此定理. 记$f(t)$与$f_n(t)$分别为$\xi_1 - a$与$\frac{S_n - na}{\sqrt{n}\sigma}$的特征函数.

证明: 我们利用特征函数来证明此定理. 记$f(t)$与$f_n(t)$分别为ξ_1-a与$\frac{S_n-na}{\sqrt{n}\sigma}$的特征函数. 因为序列$\{\xi_n, n\geq 1\}$是i.i.d.的, 所以有

$$f_n(t)=[f(\frac{t}{\sqrt{n}\sigma})]^n.$$

又因为E$\xi_1=a$, Var$\xi_1=\sigma^2$. 所以$f(t)$有二阶连续导数且有下列的Taylor展开式:

$$f(x)=f(0)+f^{'}(0)x+\frac{1}{2}f^{''}(0)x^2+o(x^2),\ x\to 0.$$

因为

$$f^{'}(0) = \mathrm{i} \cdot \mathsf{E}(\xi_1 - a) = 0, \;\; f^{''}(0) = \mathrm{i}^2 \cdot \mathsf{E}(\xi_1 - a)^2 = -\sigma^2.$$

所以

$$\forall t \in \mathbb{R}, \;\; f(\frac{t}{\sqrt{n}\sigma}) = 1 - \frac{t^2}{2n} + o(\frac{1}{n}), \;\; n \to \infty,$$

从而

$$\forall t \in \mathbb{R}, \;\; f_n(t) = [f(\frac{t}{\sqrt{n}\sigma})]^n \to e^{-t^2/2}.$$

由逆极限定理可知Lindeberg-Lévy CLT成立.

例5 当辐射的强度超过每小时0.5毫伦琴(mr)时，辐射会对人的健康造成伤害. 设一台彩电工作时的平均辐射强度是0.036mr/h，方差是0.0081，则家庭中一台彩电的辐射一般不会对人造成健康伤害. 但是彩电销售店同时有多台彩电工作时，辐射可能对人造成健康伤害. 现在有16台彩电同时工作，问这16台彩电的辐射量可以对人造成健康伤害的概率.

例5 当辐射的强度超过每小时0.5毫伦琴(mr)时，辐射会对人的健康造成伤害. 设一台彩电工作时的平均辐射强度是0.036mr/h，方差是0.0081，则家庭中一台彩电的辐射一般不会对人造成健康伤害. 但是彩电销售店同时有多台彩电工作时，辐射可能对人造成健康伤害. 现在有16台彩电同时工作，问这16台彩电的辐射量可以对人造成健康伤害的概率.

解：用ξ_k表示第k台彩电的辐射量(mr/h)，则E$\xi_k = 0.036$，Var$\xi_k =$ 0.0081，并且$S_{16} = \sum_{k=1}^{16} \xi_k$是$n = 16$台彩电的辐射量. 题目要求P$(S_{16} > 0.5)$. 认为$\{\xi_k, k \geq 1\}$是i.i.d.的，由Lindeberg-Lévy CLT知：

$$\frac{S_{16} - 16 \times 0.036}{\sqrt{16} \times 0.09} \overset{\text{近似}}{\sim} N(0,1).$$

所以

$$
\begin{aligned}
\mathsf{P}(S_{16} > 0.5) =& \mathsf{P}\left(\frac{S_{16} - 16 \times 0.036}{\sqrt{16} \times 0.09} > \frac{0.5 - 16 \times 0.036}{\sqrt{16} \times 0.09}\right) \\
=& \mathsf{P}\left(\frac{S_{16} - 16 \times 0.036}{\sqrt{16} \times 0.09} > -0.211\right) \\
\approx& \Phi(0.211) = 0.58.
\end{aligned}
$$

Lindeberg-Lévy CLT要求各ξ_k同分布，事实上这一条件可被放宽. Lindeberg证明了在和式

$$\frac{\sum_{k=1}^n \xi_k - \mathsf{E}(\sum_{k=1}^n \xi_k)}{\sqrt{\sum_{k=1}^n \mathsf{Var}\xi_k}}$$

中，只要被加项

$$\frac{\xi_k - \mathsf{E}\xi_k}{\sqrt{\sum_{k=1}^n \mathsf{Var}\xi_k}}$$

“均匀地小”，则CLT仍然成立.

定理（Lindeberg-Feller CLT）

设$\{\xi_k, k \geq 1\}$是独立随机变量序列, 则

$$\lim_{n\to\infty} \max_{1\leq k\leq n} \frac{Var\xi_k}{\sum_{k=1}^n Var\xi_k} = 0 \quad \text{(Feller 条件)}$$

和

$$\frac{\sum_{k=1}^n \xi_k - E(\sum_{k=1}^n \xi_k)}{\sqrt{\sum_{k=1}^n Var\xi_k}} \xrightarrow{d} N(0,1)$$

成立的充要条件是Lindeberg条件被满足:

$$\frac{1}{\sum_{k=1}^n Var\xi_k} \sum_{k=1}^n \int_{\frac{|x-E\xi_k|}{\sqrt{\sum_{k=1}^n Var\xi_k}} \geq \tau} (x - E\xi_k)^2 \mathrm{d}F_k(x) \to 0, \quad \forall \tau > 0.$$

定理（Lyapunov CLT)

若对独立随机变量序列$\{\xi_k, k \geq 1\}$存在常数$\delta > 0$, 使得当$n \to \infty$时有

$$\frac{1}{(\sum_{k=1}^{n} Var\xi_k)^{1+\delta/2}} \sum_{k=1}^{n} E|\xi_k - E\xi_k|^{2+\delta} \to 0, \tag{4.1.2}$$

则CLT成立.

证明: 易知(4.1.2)可推得Lindeberg条件成立. 所以由Lindeberg-Feller CLT知Lyapunov CLT成立.

例6 设$\{\xi_k, k \geq 1\}$是相互独立的随机变量序列, ξ_k的分布列是

$$\mathsf{P}(\xi_k = \pm k) = 1/2.$$

例6 设$\{\xi_k, k \geq 1\}$是相互独立的随机变量序列, ξ_k的分布列是

$$\mathsf{P}(\xi_k = \pm k) = 1/2.$$

易知

$$\mathsf{E}\xi_k = 0,\ \mathsf{Var}\xi_k = k^2,\ \ \mathsf{E}|\xi_k|^3 = k^3.$$

因此

$$\sum_{k=1}^{n} \mathsf{E}|\xi_k - \mathsf{E}\xi_k|^3 \Big/ \Big(\sum_{k=1}^{n} \mathsf{Var}\xi_k\Big)^{3/2} \to 0,$$

例6 设$\{\xi_k, k \geq 1\}$是相互独立的随机变量序列, ξ_k的分布列是

$$\mathsf{P}(\xi_k = \pm k) = 1/2.$$

易知

$$\mathsf{E}\xi_k = 0,\ \mathsf{Var}\xi_k = k^2,\ \ \mathsf{E}|\xi_k|^3 = k^3.$$

因此

$$\sum_{k=1}^{n} \mathsf{E}|\xi_k - \mathsf{E}\xi_k|^3 \Big/ \Big(\sum_{k=1}^{n} \mathsf{Var}\xi_k\Big)^{3/2} \to 0,$$

所以由Lyapunov CLT知$\{\xi_k, k \geq 1\}$服从CLT.

*我们把CLT推广到多元情形:

*我们把CLT推广到多元情形:

定理 (多元中心极限定理)

若p维随机向量$\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n, \cdots$相互独立, 具有相同的分布, 其数学期望为$\boldsymbol{a}$, 协方差矩阵为$\boldsymbol{\Sigma}$, 则

$$\boldsymbol{\eta}_n = \frac{\sum_{k=1}^{n}(\boldsymbol{\xi}_k - \boldsymbol{a})}{\sqrt{n}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

*我们把CLT推广到多元情形:

定理 (多元中心极限定理)

若p维随机向量$\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n, \cdots$相互独立, 具有相同的分布, 其数学期望为$\boldsymbol{a}$, 协方差矩阵为$\boldsymbol{\Sigma}$, 则

$$\boldsymbol{\eta}_n = \frac{\sum_{k=1}^{n}(\boldsymbol{\xi}_k - \boldsymbol{a})}{\sqrt{n}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

证明: 对p维列向量$\boldsymbol{\lambda}$, 构造

$$\zeta_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \boldsymbol{\lambda}'(\boldsymbol{\xi}_k - \boldsymbol{a}) = \boldsymbol{\lambda}'\boldsymbol{\eta}_n.$$

*我们把CLT推广到多元情形:

定理 (多元中心极限定理)

若p维随机向量$\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n, \cdots$相互独立, 具有相同的分布, 其数学期望为$\boldsymbol{a}$, 协方差矩阵为$\boldsymbol{\Sigma}$, 则

$$\boldsymbol{\eta}_n = \frac{\sum_{k=1}^{n}(\boldsymbol{\xi}_k - \boldsymbol{a})}{\sqrt{n}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

证明: 对p维列向量$\boldsymbol{\lambda}$, 构造

$$\zeta_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \boldsymbol{\lambda}'(\boldsymbol{\xi}_k - \boldsymbol{a}) = \boldsymbol{\lambda}'\boldsymbol{\eta}_n.$$

由于

$$\mathsf{E}\zeta_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \boldsymbol{\lambda}'(\mathsf{E}\boldsymbol{\xi}_k - \boldsymbol{a}) = 0,$$

$$\begin{aligned}
\mathrm{Var}\zeta_n =& \mathsf{E}\zeta_n^2 = \frac{1}{n}\mathsf{E}\left[\sum_{k=1}^{n}\boldsymbol{\lambda}'(\boldsymbol{\xi}_k-\boldsymbol{a})\cdot\sum_{k=1}^{n}(\boldsymbol{\xi}_k-\boldsymbol{a})'\boldsymbol{\lambda}\right] \\
=& \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\left[\boldsymbol{\lambda}'(\boldsymbol{\xi}_k-\boldsymbol{a})(\boldsymbol{\xi}_k-\boldsymbol{a})'\boldsymbol{\lambda}\right] \\
=& \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda} = \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}.
\end{aligned}$$

$$\begin{aligned}\mathrm{Var}\zeta_n =& \mathrm{E}\zeta_n^2 = \frac{1}{n}\mathrm{E}\left[\sum_{k=1}^n \boldsymbol{\lambda}'(\boldsymbol{\xi}_k - \boldsymbol{a}) \cdot \sum_{k=1}^n (\boldsymbol{\xi}_k - \boldsymbol{a})'\boldsymbol{\lambda}\right]\\ =& \frac{1}{n}\sum_{k=1}^n \mathrm{E}\left[\boldsymbol{\lambda}'(\boldsymbol{\xi}_k - \boldsymbol{a})(\boldsymbol{\xi}_k - \boldsymbol{a})'\boldsymbol{\lambda}\right]\\ =& \frac{1}{n}\sum_{k=1}^n \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda} = \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}.\end{aligned}$$

因此ζ_n是均值为0, 方差为$\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}$的一维随机变量. 由Lindeberg-Lévy CLT知$\{\zeta_n, n \geq 1\}$的分布函数弱收敛于$N(0, \lambda'\boldsymbol{\Sigma}\lambda)$的分布函数. 因此, 若以$f_n(t)$记$\zeta_n$的特征函数,则由正极限定理知

$$f_n(t) \to f(t) = \exp\{-\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}t^2/2\}. \tag{4.1.3}$$

$$\begin{aligned}\mathrm{Var}\zeta_n =& \mathrm{E}\zeta_n^2 = \frac{1}{n}\mathrm{E}\left[\sum_{k=1}^{n}\boldsymbol{\lambda}'(\boldsymbol{\xi}_k-\boldsymbol{a})\cdot\sum_{k=1}^{n}(\boldsymbol{\xi}_k-\boldsymbol{a})'\boldsymbol{\lambda}\right]\\ =&\frac{1}{n}\sum_{k=1}^{n}\mathrm{E}\left[\boldsymbol{\lambda}'(\boldsymbol{\xi}_k-\boldsymbol{a})(\boldsymbol{\xi}_k-\boldsymbol{a})'\boldsymbol{\lambda}\right]\\ =&\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}=\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}.\end{aligned}$$

因此ζ_n是均值为0, 方差为$\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}$的一维随机变量. 由Lindeberg-Lévy CLT知$\{\zeta_n, n\geq 1\}$的分布函数弱收敛于$N(0,\lambda'\boldsymbol{\Sigma}\lambda)$的分布函数. 因此, 若以$f_n(t)$记$\zeta_n$的特征函数,则由正极限定理知

$$f_n(t)\to f(t)=\exp\{-\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}t^2/2\}. \tag{4.1.3}$$

而

$$f_n(t)=\mathrm{E}e^{\mathrm{i}t\zeta_n}=\mathrm{E}\exp\{\mathrm{i}t\boldsymbol{\lambda}'\boldsymbol{\eta}_n\},$$

因此

$$f_n(1) = \mathsf{E}e^{\mathrm{i}\zeta_n} = \mathsf{E}\exp\{\mathrm{i}\boldsymbol{\lambda}'\boldsymbol{\eta}_n\}.$$

它作为$\boldsymbol{\lambda}$的函数，是$\boldsymbol{\eta}_n$的特征函数. 在(4.1.3)中令$t=1$得

$$f_n(1) \to \exp\{-\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}/2\}.$$

这正是p元正态分布$N(\mathbf{0}, \boldsymbol{\Sigma})$的特征函数. 再由逆极限定理知定理得证.

依概率收敛与弱大数定律

4.2.1 依概率收敛

依概率收敛与弱大数定律

4.2.1 依概率收敛

分布函数可完整描述随机变量取值的分布规律, 但是分布函数的收敛性并不能反映随机变量序列取值之间的接近程度. 举例如下:

依概率收敛与弱大数定律

4.2.1 依概率收敛

分布函数可完整描述随机变量取值的分布规律, 但是分布函数的收敛性并不能反映随机变量序列取值之间的接近程度. 举例如下:

向区间$[0,1]$上随机等可能投点, 样本点ω表示落点的位置, 定义

$$\xi(\omega)=\begin{cases}1, & \omega\in[0,0.5],\\ 0, & \omega\in(0.5,1],\end{cases}\qquad \eta(\omega)=\begin{cases}0, & \omega\in[0,0.5],\\ 1, & \omega\in(0.5,1].\end{cases}$$

依概率收敛与弱大数定律

4.2.1 依概率收敛

分布函数可完整描述随机变量取值的分布规律, 但是分布函数的收敛性并不能反映随机变量序列取值之间的接近程度. 举例如下:

向区间$[0,1]$上随机等可能投点, 样本点ω表示落点的位置, 定义

$$\xi(\omega)=\begin{cases}1, & \omega\in[0,0.5],\\ 0, & \omega\in(0.5,1],\end{cases}\qquad \eta(\omega)=\begin{cases}0, & \omega\in[0,0.5],\\ 1, & \omega\in(0.5,1].\end{cases}$$

则ξ与η具有相同的分布函数

$$F(x)=$$

依概率收敛与弱大数定律

4.2.1 依概率收敛

分布函数可完整描述随机变量取值的分布规律, 但是分布函数的收敛性并不能反映随机变量序列取值之间的接近程度. 举例如下:

向区间$[0,1]$上随机等可能投点, 样本点ω表示落点的位置, 定义

$$\xi(\omega)=\begin{cases}1, & \omega\in[0,0.5],\\ 0, & \omega\in(0.5,1],\end{cases}\qquad \eta(\omega)=\begin{cases}0, & \omega\in[0,0.5],\\ 1, & \omega\in(0.5,1].\end{cases}$$

则ξ与η具有相同的分布函数

$$F(x)=\begin{cases}0, & x<0,\\ 1/2, & 0\le x<1,\\ 1, & x\ge 1.\end{cases}$$

即两个不同的随机变量可以有相同的分布函数.

依概率收敛与弱大数定律

4.2.1 依概率收敛

分布函数可完整描述随机变量取值的分布规律, 但是分布函数的收敛性并不能反映随机变量序列取值之间的接近程度. 举例如下:

向区间$[0,1]$上随机等可能投点, 样本点ω表示落点的位置, 定义

$$\xi(\omega)=\begin{cases}1, & \omega\in[0,0.5],\\ 0, & \omega\in(0.5,1],\end{cases}\qquad \eta(\omega)=\begin{cases}0, & \omega\in[0,0.5],\\ 1, & \omega\in(0.5,1].\end{cases}$$

则ξ与η具有相同的分布函数

$$F(x)=\begin{cases}0, & x<0,\\ 1/2, & 0\le x<1,\\ 1, & x\ge 1.\end{cases}$$

即两个不同的随机变量可以有相同的分布函数. 若定义$\xi_n=\xi$, $n\ge 1$, 则$\xi_n\xrightarrow{d}\eta$, 但$|\xi_n-\eta|=|\xi-\eta|\equiv 1$.

因此，我们需要引入另外的收敛性来度量随机变量取值的接近程度:

因此，我们需要引入另外的收敛性来度量随机变量取值的接近程度:

定义 (依概率收敛)

设ξ和$\{\xi_n, n \geq 1\}$是定义在同一概率空间上的随机变量. 如果对任意的$\varepsilon > 0$,

$$\lim_{n\to\infty} P(|\xi_n - \xi| \geq \varepsilon) = 0,$$

或

$$\lim_{n\to\infty} P(|\xi_n - \xi| < \varepsilon) = 1,$$

则称$\{\xi_n, n \geq 1\}$依概率收敛(convergence in probability)于ξ, 记作$\xi_n \xrightarrow{P} \xi$.

因此，我们需要引入另外的收敛性来度量随机变量取值的接近程度:

定义 (依概率收敛)

设ξ和$\{\xi_n, n \geq 1\}$是定义在同一概率空间上的随机变量. 如果对任意的$\varepsilon > 0$,

$$\lim_{n\to\infty} P(|\xi_n - \xi| \geq \varepsilon) = 0,$$

或

$$\lim_{n\to\infty} P(|\xi_n - \xi| < \varepsilon) = 1,$$

则称$\{\xi_n, n \geq 1\}$依概率收敛(convergence in probability)于ξ, 记作$\xi_n \xrightarrow{P} \xi$.

注1: $\xi_n \xrightarrow{P} \xi$可直观地理解为: 除去极小的可能性, 只要$n$充分大, ξ和ξ_n的取值可任意接近.

注2: 显然, 依分布收敛不能推出依概率收敛.

关于依分布收敛和依概率收敛的关系, 我们有下列的定理.

关于依分布收敛和依概率收敛的关系，我们有下列的定理.

定理

设ξ和$\{\xi_n, n \geq 1\}$是定义在同一概率空间上的随机变量.

- 如果$\xi_n \xrightarrow{P} \xi$, 则$\xi_n \xrightarrow{d} \xi$.
- 如果$\xi_n \xrightarrow{d} c$, c是常数, 则$\xi_n \xrightarrow{P} c$.

关于依分布收敛和依概率收敛的关系，我们有下列的定理.

定理

设ξ和$\{\xi_n, n \geq 1\}$是定义在同一概率空间上的随机变量.

- 如果$\xi_n \xrightarrow{P} \xi$, 则$\xi_n \xrightarrow{d} \xi$.
- 如果$\xi_n \xrightarrow{d} c$, c是常数, 则$\xi_n \xrightarrow{P} c$.

证明: 第一部分的证明. 设F和F_n分别是ξ和ξ_n的分布函数, x为F的连续点.

关于依分布收敛和依概率收敛的关系，我们有下列的定理.

定理

设ξ和$\{\xi_n, n\geq 1\}$是定义在同一概率空间上的随机变量.

- 如果$\xi_n \xrightarrow{P} \xi$, 则$\xi_n \xrightarrow{d} \xi$.
- 如果$\xi_n \xrightarrow{d} c$, c是常数, 则$\xi_n \xrightarrow{P} c$.

证明: 第一部分的证明. 设F和F_n分别是ξ和ξ_n的分布函数, x为F的连续点. 任意给定$\varepsilon > 0$, 有

$$\begin{aligned}\{\xi \leq x-\varepsilon\} &= \{\xi \leq x-\varepsilon, \xi_n \leq x\} \cup \{\xi \leq x-\varepsilon, \xi_n > x\} \\ &\subset \{\xi_n \leq x\} \cup \{\xi_n - \xi \geq \varepsilon\},\end{aligned}$$

所以

$$F(x-\varepsilon) \leq F_n(x) + \mathsf{P}(\xi_n - \xi \geq \varepsilon).$$

因为$\xi_n \xrightarrow{P} \xi$, 所以

$$\mathsf{P}(\xi_n - \xi \geq \varepsilon) \leq \mathsf{P}(|\xi_n - \xi| \geq \varepsilon) \to 0,$$

从而

$$F(x-\varepsilon) \leq \liminf_{n\to\infty} F_n(x). \tag{4.2.1}$$

类似地,

$$\begin{aligned}\{\xi_n \leq x\} &= \{\xi_n \leq x, \xi \leq x+\varepsilon\} \cup \{\xi_n \leq x, \xi > x+\varepsilon\} \\ &\subset \{\xi \leq x+\varepsilon\} \cup \{\xi - \xi_n \geq \varepsilon\},\end{aligned}$$

从而

$$F_n(x) \leq F(x+\varepsilon) + \mathsf{P}(\xi - \xi_n \geq \varepsilon).$$

因此

$$\limsup_{n\to\infty} F_n(x) \leq F(x+\varepsilon). \tag{4.2.2}$$

因为$\xi_n \xrightarrow{P} \xi$, 所以

$$\mathsf{P}(\xi_n - \xi \geq \varepsilon) \leq \mathsf{P}(|\xi_n - \xi| \geq \varepsilon) \to 0,$$

从而

$$F(x-\varepsilon) \leq \liminf_{n\to\infty} F_n(x). \tag{4.2.1}$$

类似地,

$$\begin{aligned} \{\xi_n \leq x\} &= \{\xi_n \leq x, \xi \leq x+\varepsilon\} \cup \{\xi_n \leq x, \xi > x+\varepsilon\} \\ &\subset \{\xi \leq x+\varepsilon\} \cup \{\xi - \xi_n \geq \varepsilon\}, \end{aligned}$$

从而

$$F_n(x) \leq F(x+\varepsilon) + \mathsf{P}(\xi - \xi_n \geq \varepsilon).$$

因此

$$\limsup_{n\to\infty} F_n(x) \leq F(x+\varepsilon). \tag{4.2.2}$$

结合(4.2.1)和(4.2.2)可知对任意的$\varepsilon > 0$,

$$F(x-\varepsilon) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x+\varepsilon).$$

由于x是F的连续点, 所以令$\varepsilon \to 0$得

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

即$\xi_n \xrightarrow{d} \xi$.

由于x是F的连续点，所以令$\varepsilon \to 0$得

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

即$\xi_n \xrightarrow{d} \xi$.

第二部分的证明. 如果$\xi_n \xrightarrow{d} c$, 则当$x \neq c$时

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0, & x < c, \\ 1, & x > c. \end{cases}$$

因此对任意的$\varepsilon > 0$有

$$\begin{aligned} \mathsf{P}(|\xi_n - c| \geq \varepsilon) =& \mathsf{P}(\xi_n \geq c+\varepsilon) + \mathsf{P}(\xi_n \leq c-\varepsilon) \\ =& 1 - \mathsf{P}(\xi_n < c+\varepsilon) + \mathsf{P}(\xi_n \leq c-\varepsilon) \\ =& 1 - F_n(c+\varepsilon-0) + F_n(c-\varepsilon) \to 0. \end{aligned}$$

引理（Slutsky引理）

如果$\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{d} c$, 则$\xi_n + \eta_n \xrightarrow{d} \xi + c$.

引理 (Slutsky引理)

如果$\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{d} c$, 则$\xi_n + \eta_n \xrightarrow{d} \xi + c$.

*证明: 不妨假设$c = 0$(否则考虑$\eta_n' = \eta_n - c$). 记ξ的分布函数为F, $\zeta_n = \xi_n + \eta_n$. 任取F的连续点x, 对任意给定的$\varepsilon > 0$, 取$0 < \varepsilon_1 < \varepsilon$使得$x \pm \varepsilon_1$为$F$的连续点. 易见

$$\mathsf{P}(\zeta_n \le x) \le \mathsf{P}(\xi_n \le x + \varepsilon_1) + \mathsf{P}(|\eta_n| \ge \varepsilon_1),$$
$$\mathsf{P}(\xi_n \le x - \varepsilon_1) \le \mathsf{P}(\zeta_n \le x) + \mathsf{P}(|\eta_n| \ge \varepsilon_1).$$

由上述两个式子得

$$\begin{aligned} & \mathsf{P}(\xi_n \le x - \varepsilon_1) - \mathsf{P}(|\eta_n| \ge \varepsilon_1) \\ \le & \mathsf{P}(\zeta_n \le x) \\ \le & \mathsf{P}(\xi_n \le x + \varepsilon_1) + \mathsf{P}(|\eta_n| \ge \varepsilon_1). \end{aligned}$$

令$n \to \infty$得

$$F(x-\varepsilon_1) \le \liminf_{n\to\infty} \mathsf{P}(\zeta_n \le x) \le \limsup_{n\to\infty} \mathsf{P}(\zeta_n \le x) \le F(x+\varepsilon_1).$$

再令$\varepsilon_1 \searrow 0$, 得

$$\lim_{n\to\infty} \mathsf{P}(\zeta_n \le x) = F(x).$$

令$n \to \infty$得

$$F(x-\varepsilon_1) \le \liminf_{n\to\infty} \mathsf{P}(\zeta_n \le x) \le \limsup_{n\to\infty} \mathsf{P}(\zeta_n \le x) \le F(x+\varepsilon_1).$$

再令$\varepsilon_1 \searrow 0$, 得

$$\lim_{n\to\infty} \mathsf{P}(\zeta_n \le x) = F(x).$$

注: 在Slutsky引理中, 对ξ_n与η_n的独立性没有要求, 所以这是一个非常有用的引理. 此外, 还可以得到$\xi_n\eta_n \xrightarrow{d} c\xi$这一有用的结论.

***注**: 如果ξ_n与η_n独立, 且$\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{d} \eta$. 则$\xi_n + \eta_n \xrightarrow{d} \xi + \eta$. 但如果独立性不满足, 则在一般情况下无法得到$\xi_n + \eta_n \xrightarrow{d} \xi + \eta$.

例1 设$\{\xi_n, n \geq 1\}$独立同分布, 都为$[0, a]$上的均匀分布, $\eta_n = \max_{1 \leq k \leq n} \xi_k$. 求证$\eta_n \xrightarrow{P} a$.

例1 设$\{\xi_n, n \geq 1\}$独立同分布, 都为$[0,a]$上的均匀分布, $\eta_n = \max_{1\leq k\leq n} \xi_k$. 求证$\eta_n \xrightarrow{P} a$.

证明: 只需证$\eta_n \xrightarrow{d} a$.

例1　设$\{\xi_n, n \geq 1\}$独立同分布, 都为$[0,a]$上的均匀分布, $\eta_n = \max_{1\leq k\leq n}\xi_k$. 求证$\eta_n \xrightarrow{P} a$.

证明: 只需证$\eta_n \xrightarrow{d} a$. 易知ξ_k的分布函数为

$$F(x) = \begin{cases} 0, & x < 0, \\ x/a, & 0 \leq x < a, \\ 1, & x \geq a. \end{cases}$$

例1 设$\{\xi_n, n \geq 1\}$独立同分布, 都为$[0, a]$上的均匀分布, $\eta_n = \max_{1\leq k\leq n} \xi_k$. 求证$\eta_n \xrightarrow{P} a$.

证明: 只需证$\eta_n \xrightarrow{d} a$. 易知ξ_k的分布函数为

$$F(x) = \begin{cases} 0, & x < 0, \\ x/a, & 0 \leq x < a, \\ 1, & x \geq a. \end{cases}$$

所以η_n的分布函数$G_n(x)$为

$$G_n(x) = [F(x)]^n = \begin{cases} 0, & x < 0, \\ (x/a)^n, & 0 \leq x < a, \\ 1, & x \geq a. \end{cases} \to \begin{cases} 0, & x < a, \\ 1, & x \geq a. \end{cases}$$

后者正是在a点退化的分布函数. 证毕.

关于依概率收敛, 我们有如下的重要结论.

关于依概率收敛, 我们有如下的重要结论.

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在同一概率空间上的随机变量序列. 则有

- 若$\xi_n \xrightarrow{P} \xi$, $\xi_n \xrightarrow{P} \eta$, 则$P(\xi = \eta) = 1$.
- 若$\xi_n \xrightarrow{P} \xi$, g是$(-\infty, \infty)$上的连续函数, 则$g(\xi_n) \xrightarrow{P} g(\xi)$.

关于依概率收敛, 我们有如下的重要结论.

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在同一概率空间上的随机变量序列. 则有

- 若$\xi_n \xrightarrow{P} \xi$, $\xi_n \xrightarrow{P} \eta$, 则$P(\xi = \eta) = 1$.
- 若$\xi_n \xrightarrow{P} \xi$, g是$(-\infty, \infty)$上的连续函数, 则$g(\xi_n) \xrightarrow{P} g(\xi)$.

第一部分的证明: 对任意的$\varepsilon > 0$, 有

$$\{|\xi - \eta| \geq \varepsilon\} \subset \{|\xi_n - \xi| \geq \varepsilon/2\} \cup \{|\xi_n - \eta| \geq \varepsilon/2\}.$$

因此

$$\mathsf{P}(|\xi - \eta| \geq \varepsilon) \leq \mathsf{P}(|\xi_n - \xi| \geq \varepsilon/2) + \mathsf{P}(|\xi_n - \eta| \geq \varepsilon/2).$$

由于$\xi_n \xrightarrow{P} \xi$且$\eta_n \xrightarrow{P} \eta$, 在上式中令$n \to \infty$得$\mathsf{P}(|\xi - \eta| \geq \varepsilon) = 0$.

进一步地, 可知

$$\begin{aligned}\mathsf{P}(|\xi-\eta|>0) =&\mathsf{P}(\cup_{n=1}^{\infty}\{|\xi-\eta|\geq 1/n\}) \\ \leq& \sum_{n=1}^{\infty}\mathsf{P}(|\xi-\eta|\geq 1/n)=0.\end{aligned}$$

所以$\mathsf{P}(\xi=\eta)=1$.

进一步地, 可知

$$\begin{aligned}\mathsf{P}(|\xi-\eta|>0) =&\mathsf{P}(\cup_{n=1}^{\infty}\{|\xi-\eta|\geq 1/n\})\\ &\leq \sum_{n=1}^{\infty}\mathsf{P}(|\xi-\eta|\geq 1/n)=0.\end{aligned}$$

所以$\mathsf{P}(\xi=\eta)=1$.

第二部分的证明: 对任意给定的$\varepsilon^{'}>0$, 存在$M>0$使得

$$\mathsf{P}(|\xi|\geq M)\leq \mathsf{P}(|\xi|\geq M/2)<\varepsilon^{'}/4. \tag{4.2.3}$$

由于$\xi_n\xrightarrow{P}\xi$, 故存在$N_1\geq 1$使得当$n\geq N_1$时,

$$\mathsf{P}(|\xi_n-\xi|\geq M/2)<\varepsilon^{'}/4.$$

因此当$n\geq N_1$时,

$$\mathsf{P}(|\xi_n|\geq M)\leq \mathsf{P}(|\xi_n-\xi|\geq M/2)+\mathsf{P}(|\xi|\geq M/2)<\frac{\varepsilon^{'}}{2}. \tag{4.2.4}$$

又因为$g(x)$在$(-\infty,\infty)$上连续, 所以它在$[-M,M]$上一致连续. 因此可知: 当$|x|<M$, $|y|<M$时, 对给定的$\varepsilon>0$, 存在$\delta>0$使得当$|x-y|<\delta$ 时, $|g(x)-g(y)|<\varepsilon$. 这样,

$$\begin{aligned}
& \mathsf{P}(|g(\xi_n)-g(\xi)|\geq\varepsilon) \\
\leq\ & \mathsf{P}(|g(\xi_n)-g(\xi)|\geq\varepsilon,|\xi_n|<M,|\xi|<M) \\
& \quad +\mathsf{P}(|\xi_n|\geq M)+\mathsf{P}(|\xi|\geq M) \\
\leq\ & \mathsf{P}(|\xi_n-\xi|\geq\delta,|\xi_n|<M,|\xi|<M) \\
& \quad +\mathsf{P}(|\xi_n|\geq M)+\mathsf{P}(|\xi|\geq M) \\
\leq\ & \mathsf{P}(|\xi_n-\xi|\geq\delta)+\mathsf{P}(|\xi_n|\geq M)+\mathsf{P}(|\xi|\geq M).
\end{aligned} \tag{4.2.5}$$

对上式的δ, 存在$N_2\geq 1$, 使得当$n\geq N_2$时,

$$\mathsf{P}(|\xi_n-\xi|\geq\delta)<\frac{\varepsilon'}{4}. \tag{4.2.6}$$

结合(4.2.3)-(4.2.6)可知: 当$n \geq \max\{N_1, N_2\}$时,

$$\mathsf{P}(|g(\xi_n) - g(\xi)| \geq \varepsilon) < \frac{\varepsilon^{'}}{4} + \frac{\varepsilon^{'}}{4} + \frac{\varepsilon^{'}}{2} = \varepsilon^{'},$$

所以$g(\xi_n) \xrightarrow{P} g(\xi)$.

定理

- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则$\xi_n \pm \eta_n \xrightarrow{P} \xi \pm \eta$.
- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则$\xi_n \eta_n \xrightarrow{P} \xi\eta$.
- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} c$, c为一常数, 假设η_n与c都不为零, 则$\xi_n/\eta_n \xrightarrow{P} \xi/c$.

定理

- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则$\xi_n \pm \eta_n \xrightarrow{P} \xi \pm \eta$.
- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则$\xi_n \eta_n \xrightarrow{P} \xi\eta$.
- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} c$, c为一常数, 假设η_n与c都不为零, 则则$\xi_n/\eta_n \xrightarrow{P} \xi/c$.

证明: 这里, 我们只给出第三个结论的证明. 由第二个结论, 我们只需证明$1/\eta_n \xrightarrow{P} 1/c$.

定理

- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则$\xi_n \pm \eta_n \xrightarrow{P} \xi \pm \eta$.
- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则$\xi_n\eta_n \xrightarrow{P} \xi\eta$.
- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} c$, c为一常数, 假设η_n与c都不为零, 则 则$\xi_n/\eta_n \xrightarrow{P} \xi/c$.

证明: 这里, 我们只给出第三个结论的证明. 由第二个结论, 我们只需证明$1/\eta_n \xrightarrow{P} 1/c$. 对任意的$\varepsilon > 0$, 记

$$\delta = \min\left\{\frac{1}{2}c^2\varepsilon, \frac{1}{2}|c|\right\}.$$

定理

- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则$\xi_n \pm \eta_n \xrightarrow{P} \xi \pm \eta$.
- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则$\xi_n \eta_n \xrightarrow{P} \xi\eta$.
- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} c$, c为一常数, 假设η_n与c都不为零, 则则$\xi_n/\eta_n \xrightarrow{P} \xi/c$.

证明: 这里, 我们只给出第三个结论的证明. 由第二个结论, 我们只需证明$1/\eta_n \xrightarrow{P} 1/c$. 对任意的$\varepsilon > 0$, 记

$$\delta = \min\left\{\frac{1}{2}c^2\varepsilon, \frac{1}{2}|c|\right\}.$$

若$|\eta_n - c| < \delta$, 则$|\eta_n| > |c| - \delta \geq \frac{1}{2}|c|$, 因此

$$\left|\frac{1}{\eta_n} - \frac{1}{c}\right| = \frac{|\eta_n - c|}{|c\eta_n|} < \frac{\frac{1}{2}c^2\varepsilon}{\frac{1}{2}c^2} = \varepsilon.$$

所以

$$\mathsf{P}\left(\left|\frac{1}{\eta_n}-\frac{1}{c}\right|\geq\varepsilon\right)\leq\mathsf{P}(|\eta_n-c|\geq\delta)\to 0.$$

所以

$$\mathsf{P}\left(\left|\frac{1}{\eta_n}-\frac{1}{c}\right|\geq\varepsilon\right)\leq\mathsf{P}(|\eta_n-c|\geq\delta)\to 0.$$

注: 在此定理中, 我们无需假设ξ_n和η_n独立.

定理（Markov不等式）

设ξ是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量，$g(x)$是$[0, \infty)$上非负单调不减函数. 则对任意的$x > 0$,

$$P(|\xi| \geq x) \leq \frac{Eg(|\xi|)}{g(x)}.$$

定理（Markov不等式）

设ξ是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量，$g(x)$是$[0,\infty)$上非负单调不减函数. 则对任意的$x>0$,

$$P(|\xi| \geq x) \leq \frac{Eg(|\xi|)}{g(x)}.$$

证明：当$\mathrm{E}g(|\xi|)=\infty$，结论显然成立. 假设$\mathrm{E}g(|\xi|)<\infty$，$\xi$的分布函数为$F(x)$. 由$g(x)$的单调不减性知：当$|y|>x$时，$g(|y|)\geq g(x)$. 因此

$$\begin{aligned}\mathrm{P}(|\xi| \geq x) &= \int_{|y|\geq x} \mathrm{d}F(y) \leq \int_{|y|\geq x} \frac{g(|y|)}{g(x)} \mathrm{d}F(y) \\ &\leq \frac{1}{g(x)} \int_{-\infty}^{\infty} g(|y|)\mathrm{d}F(y) \\ &= \frac{\mathrm{E}g(|\xi|)}{g(x)}.\end{aligned}$$

定理

$\xi_n \xrightarrow{P} \xi$当且仅当$E\frac{|\xi_n-\xi|^2}{1+|\xi_n-\xi|^2} \to 0.$

定理

$\xi_n \xrightarrow{P} \xi$当且仅当$E\frac{|\xi_n-\xi|^2}{1+|\xi_n-\xi|^2} \to 0.$

证明: 充分性. 由于$f(x) = \frac{x^2}{1+x^2}$在$[0,\infty)$上单调不减, 所以对任意的$\varepsilon > 0$, 由Markov不等式知

$$\mathsf{P}(|\xi_n - \xi| \geq \varepsilon) \leq \frac{1+\varepsilon^2}{\varepsilon^2}\mathsf{E}\frac{|\xi_n - \xi|^2}{1+|\xi_n-\xi|^2} \to 0,$$

所以$\xi_n \xrightarrow{P} \xi$.

必要性. 设$F_n(x)$为$\xi_n-\xi$的分布函数. 对任意的$\varepsilon>0$,

$$\begin{aligned}\mathsf{E}\frac{|\xi_n-\xi|^2}{1+|\xi_n-\xi|^2} &= \int_{-\infty}^{\infty}\frac{x^2}{1+x^2}\mathrm{d}F_n(x)\\ &= \int_{|x|<\varepsilon}\frac{x^2}{1+x^2}\mathrm{d}F_n(x)+\int_{|x|\geq\varepsilon}\frac{x^2}{1+x^2}\mathrm{d}F_n(x)\\ &\leq \frac{\varepsilon^2}{1+\varepsilon^2}+\int_{|x|\geq\varepsilon}\mathrm{d}F_n(x)\\ &= \frac{\varepsilon^2}{1+\varepsilon^2}+\mathsf{P}(|\xi_n-\xi|\geq\varepsilon).\end{aligned}$$

因为$\xi_n\xrightarrow{P}\xi$, 所以先令$n\to\infty$, 再令$\varepsilon\to 0$, 得

$$\mathsf{E}\frac{|\xi_n-\xi|^2}{1+|\xi_n-\xi|^2}\to 0.$$

4.2.2 弱大数定律

4.2.2 弱大数定律

考虑随机试验E中的事件A, 假设其发生的概率为$p\ (0 < p < 1)$. 独立重复做n次试验, 得到一个n重Bernoulli试验. 令

$$\xi_k = \begin{cases} 1, & A\text{在第}k\text{次试验中出现}, \\ 0, & A\text{在第}k\text{次试验中不出现}. \end{cases}$$

4.2.2 弱大数定律

考虑随机试验E中的事件A, 假设其发生的概率为p $(0 < p < 1)$. 独立重复做n次试验, 得到一个n重Bernoulli试验. 令

$$\xi_k = \begin{cases} 1, & A\text{在第}k\text{次试验中出现}, \\ 0, & A\text{在第}k\text{次试验中不出现}. \end{cases}$$

则$\mathsf{P}(\xi_k = 1) = p$, $S_n = \sum_{k=1}^{n} \xi_k$是$n$重Bernoulli试验中$A$发生的总次数. 在第一章中, 我们曾提到一个"事实": 事件A发生的频率S_n/n在某种意义下稳定到其概率p. 这一事实最早由Bernoulli发现: 对任意给定的$\varepsilon > 0$, 当n充分大以后, $\mathsf{P}(|S_n/n - p| \geq \varepsilon)$可以任意小.

在以定理形式给出Bernoulli的发现之前, 我们先介绍一个概念——-弱大数定律.

在以定理形式给出Bernoulli的发现之前，我们先介绍一个概念——-弱大数定律.

定义（弱大数定律/WLLN）

设$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量序列，如果存在常数列$\{a_n, n \geq 1\}$和$\{b_n, n \geq 1\}$使得

$$\frac{1}{a_n}\sum_{k=1}^{n}\xi_k - b_n \xrightarrow{P} 0.$$

则称$\{\xi_n, n \geq 1\}$服从弱大数定律(weak law of large numbers)，简称$\{\xi_n, n \geq 1\}$服从大数定律.

定理（Bernoulli大数定律）

设$\{\xi_n, n \geq 1\}$是一列独立同分布的随机变量序列，$P(\xi_k = 1) = p$，$P(\xi_k = 0) = 1 - p$，$0 < p < 1$，记$S_n = \sum_{k=1}^{n} \xi_k$. 则

$$\frac{S_n}{n} \xrightarrow{P} p.$$

定理（Bernoulli大数定律）

设$\{\xi_n, n \geq 1\}$是一列独立同分布的随机变量序列，$P(\xi_k = 1) = p$，$P(\xi_k = 0) = 1 - p$，$0 < p < 1$，记$S_n = \sum_{k=1}^n \xi_k$. 则

$$\frac{S_n}{n} \xrightarrow{P} p.$$

Bernoulli大数定律是下面的Chebyshev大数定律的特例.

定理（Chebyshev大数定律）

设$\{\xi_n, n \geq 1\}$是一列独立随机变量序列，$E\xi_k = \mu_k$，$Var\xi_k = \sigma_k^2$. 如果

$$\frac{1}{n^2}\sum_{k=1}^{n}\sigma_k^2 \to 0,$$

则$\{\xi_n, n \geq 1\}$服从大数定律，即

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}\mu_k \xrightarrow{P} 0.$$

定理（Chebyshev大数定律）

设$\{\xi_n, n\geq 1\}$是一列独立随机变量序列，$E\xi_k=\mu_k$，$Var\xi_k=\sigma_k^2$. 如果

$$\frac{1}{n^2}\sum_{k=1}^{n}\sigma_k^2\to 0,$$

则$\{\xi_n, n\geq 1\}$服从大数定律，即

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}\mu_k\xrightarrow{P}0.$$

注：若把条件$\frac{1}{n^2}\sum_{k=1}^{n}\sigma_k^2\to 0$改成$\frac{1}{n^2}\mathrm{Var}\left(\sum_{k=1}^{n}\xi_k\right)\to 0$(此时不需要假设$\{\xi_n, n\geq 1\}$是独立随机变量序列)，则定理的结论仍然成立，称为Markov大数定律.

证明: 因为

$$\mathsf{E}\Big(\frac{1}{n}\sum_{k=1}^{n}\xi_k\Big)=\frac{1}{n}\sum_{k=1}^{n}\mu_k,\quad \mathsf{Var}\Big(\frac{1}{n}\sum_{k=1}^{n}\xi_k\Big)=\frac{1}{n^2}\sum_{k=1}^{n}\sigma_k^2,$$

所以由Chebyshev不等式知: 对任意的$\varepsilon>0$,

$$\begin{aligned}\mathsf{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}\mu_k\Big|\geq\varepsilon\Big)\leq&\frac{1}{\varepsilon^2}\mathsf{Var}\Big(\frac{1}{n}\sum_{k=1}^{n}\xi_k\Big)\\=&\frac{1}{\varepsilon^2}\cdot\frac{1}{n^2}\sum_{k=1}^{n}\sigma_k^2\to 0,\end{aligned}$$

所以结论成立.

若$\{\xi_n, n \geq 1\}$不仅独立，而且同分布，则可把Chebyshev大数定律改进到下面的Khinchine大数定律.

若$\{\xi_n, n \geq 1\}$不仅独立，而且同分布，则可把Chebyshev大数定律改进到下面的Khinchine大数定律.

定理（Khinchine大数定律）

设$\{\xi_n, n \geq 1\}$是一列独立同分布随机变量序列，$E|\xi_1| < \infty$. 记$E\xi_1 = \mu$, $S_n = \sum_{k=1}^n \xi_k$, 则$\{\xi_n, n \geq 1\}$服从大数定律，即

$$\frac{S_n}{n} \xrightarrow{P} \mu.$$

证明: 分别令$f(t)$和$f_n(t)$为ξ_1与S_n/n的特征函数. 因为$\{\xi_n, n \geq 1\}$是独立同分布随机变量序列, 所以

$$f_n(t) = \mathsf{E}e^{\mathrm{i}tS_n/n} = \prod_{k=1}^{n} \mathsf{E}e^{\mathrm{i}t\xi_k/n} = [f(t/n)]^n.$$

证明: 分别令$f(t)$和$f_n(t)$为ξ_1与S_n/n的特征函数. 因为$\{\xi_n, n \geq 1\}$是独立同分布随机变量序列, 所以

$$f_n(t) = \mathsf{E}e^{\mathrm{i}tS_n/n} = \prod_{k=1}^{n} \mathsf{E}e^{\mathrm{i}t\xi_k/n} = [f(t/n)]^n.$$

另外, 由于$\mathsf{E}\xi_1 = \mu$, 所以$f^{'}(0) = \mathrm{i} \cdot \mathsf{E}\xi_1 = \mathrm{i}\mu$, 从而

$$f(x) = 1 + f^{'}(0)x + o(x) = 1 + \mathrm{i}\mu x + o(x), \ \ x \to 0.$$

因此对任意的$t \in \mathbb{R}$, 有

$$f_n(t) = [1 + \mathrm{i}\mu t/n + o(1/n)]^n \to e^{\mathrm{i}\mu t}, \ \ n \to \infty.$$

后者正是恒等于μ的退化随机变量的特征函数, 所以由逆极限定理知$\frac{S_n}{n} \xrightarrow{d} \mu$, 即$\frac{S_n}{n} \xrightarrow{P} \mu$.

例2 设$\{\xi_k, k \geq 1\}$有分布列P$(\xi_k = \pm k^s) = 1/2$, $s < 1/2$为正常数, 且$\{\xi_k, k \geq 1\}$相互独立. 试证$\{\xi_k, k \geq 1\}$服从大数定律.

例2 设$\{\xi_k, k \geq 1\}$有分布列P$(\xi_k = \pm k^s) = 1/2$, $s < 1/2$为正常数, 且$\{\xi_k, k \geq 1\}$相互独立. 试证$\{\xi_k, k \geq 1\}$服从大数定律.

证明: 因为$\{\xi_k, k \geq 1\}$独立但不一定同分布, 所以我们用Chebyshev大数定律来证明.

例2 设$\{\xi_k, k \geq 1\}$有分布列$\mathsf{P}(\xi_k = \pm k^s) = 1/2$, $s < 1/2$为正常数, 且$\{\xi_k, k \geq 1\}$相互独立. 试证$\{\xi_k, k \geq 1\}$服从大数定律.

证明: 因为$\{\xi_k, k \geq 1\}$独立但不一定同分布, 所以我们用Chebyshev大数定律来证明. 因为

$$\mathsf{E}\xi_k = 0, \mathsf{Var}\xi_k = k^{2s},$$

所以当$0 < s < 1/2$时,

$$\frac{1}{n^2}\sum_{k=1}^{n}\mathsf{Var}\xi_k = \frac{1}{n^2}\sum_{k=1}^{n}k^{2s} < \frac{1}{n^2}\sum_{k=1}^{n}n^{2s} = n^{2s-1} \to 0,$$

所以$\{\xi_k, k \geq 1\}$服从大数定律.

例3(矩估计的相合性) 假定总体ξ的一阶矩(即数学期望)m_1存在(隐含E$|\xi| < \infty$)但未知, 通常的做法是对ξ进行n次独立重复观测, 得到样本$\xi_1, \cdots, \xi_n$, 并以它们的算术平均$A_1 = \frac{1}{n}\sum_{j=1}^{n}\xi_j$作为$m_1$的估计量, 这种做法的依据就是Khinchine大数定律:

$$A_1 \xrightarrow{P} m_1.$$

例3(矩估计的相合性) 假定总体ξ的一阶矩(即数学期望)m_1存在(隐含$\mathsf{E}|\xi| < \infty$)但未知，通常的做法是对ξ进行n次独立重复观测，得到样本$\xi_1, \cdots, \xi_n$，并以它们的算术平均$A_1 = \frac{1}{n}\sum_{j=1}^{n}\xi_j$作为$m_1$的估计量，这种做法的依据就是Khinchine大数定律:

$$A_1 \xrightarrow{P} m_1.$$

更为重要的是，根据Khinchine大数定律，若总体的k阶矩$m_k = \mathsf{E}\xi^k$存在(隐含$\mathsf{E}|\xi|^k < \infty$)，这时样本的$k$阶矩$A_k = \frac{1}{n}\sum_{j=1}^{n}\xi_j^k$ 可作为m_k的估计量，依据是

$$A_k \xrightarrow{P} m_k.$$

例4(用Monte Carlo方法计算定积分) 用概率论方法数值计算定积分$J = \int_a^b g(x)\mathrm{d}x$.

例4(用Monte Carlo方法计算定积分) 用概率论方法数值计算定积分$J=\int_a^b g(x)\mathrm{d}x$.

解: 任取一列独立同分布服从$U[a,b]$的随机变量$\{\xi_k, k\geq 1\}$, 则$\{g(\xi_k), k\geq 1\}$也是一列独立同分布的随机变量, 且

$$\mathsf{E}[g(\xi_k)] =$$

例4(用Monte Carlo方法计算定积分) 用概率论方法数值计算定积分$J=\int_a^b g(x)\mathrm{d}x$.

解: 任取一列独立同分布服从$U[a,b]$的随机变量$\{\xi_k, k\geq 1\}$, 则$\{g(\xi_k),\, k\geq 1\}$也是一列独立同分布的随机变量, 且

$$\mathsf{E}[g(\xi_k)]=\frac{1}{b-a}\int_a^b g(x)\mathrm{d}x=\frac{J}{b-a}.$$

例4(用Monte Carlo方法计算定积分) 用概率论方法数值计算定积分$J=\int_a^b g(x)\mathrm{d}x$.

解: 任取一列独立同分布服从$U[a,b]$的随机变量$\{\xi_k,k\geq 1\}$, 则$\{g(\xi_k),\, k\geq 1\}$也是一列独立同分布的随机变量, 且

$$\mathrm{E}[g(\xi_k)]=\frac{1}{b-a}\int_a^b g(x)\mathrm{d}x=\frac{J}{b-a}.$$

因此只要能数值计算出$\mathrm{E}[g(\xi_k)]$, 就能得到

$$J=(b-a)\cdot\mathrm{E}[g(\xi_k)] \tag{4.2.7}$$

的值.

例4(用Monte Carlo方法计算定积分) 用概率论方法数值计算定积分$J=\int_a^b g(x)\mathrm{d}x$.

解: 任取一列独立同分布服从$U[a,b]$的随机变量$\{\xi_k, k\geq 1\}$, 则$\{g(\xi_k), k\geq 1\}$也是一列独立同分布的随机变量, 且

$$\mathsf{E}[g(\xi_k)]=\frac{1}{b-a}\int_a^b g(x)\mathrm{d}x=\frac{J}{b-a}.$$

因此只要能数值计算出$\mathsf{E}[g(\xi_k)]$, 就能得到

$$J=(b-a)\cdot\mathsf{E}[g(\xi_k)] \tag{4.2.7}$$

的值. 为求$\mathsf{E}[g(\xi_k)]$, 自然想到Khinchine大数定律.

因为

$$\frac{g(\xi_1)+\cdots+g(\xi_n)}{n} \xrightarrow{P} \mathsf{E}[g(\xi_k)], \tag{4.2.8}$$

所以我们可先在计算机上产生服从$U[a,b]$的随机数

$$\{\xi_k, k=1,\cdots,n\},$$

从而得到随机数

$$\{g(\xi_k), k=1,\cdots,n\}.$$

再根据(4.2.8)和(4.2.7)估算J的大小.

最后，我们给出随机变量序列的另一种收敛性概念.

最后, 我们给出随机变量序列的另一种收敛性概念.

定义 (r-阶平均收敛)

设ξ和$\{\xi_n, n \geq 1\}$是定义在同一概率空间$(\Omega, \mathscr{F}, P)$上的随机变量序列, $E|\xi|^r < \infty$, $E|\xi_n|^r < \infty$, $0 < r < \infty$. 如果

$$E|\xi_n - \xi|^r \to 0,$$

则称$\{\xi_n, n \geq 1\}$ r-阶平均收敛(convergence in the mean of order r)于ξ, 记作$\xi_n \xrightarrow{L_r} \xi$.

注: $r = 2$时, 称为均方收敛.

依概率收敛与r-阶平均收敛的关系:

- r-阶平均收敛 $\Rightarrow$ 依概率收敛;
- 依概率收敛 $\nRightarrow$ r-阶平均收敛.

依概率收敛与r-阶平均收敛的关系:

- r-阶平均收敛 $\Rightarrow$ 依概率收敛;
- 依概率收敛 $\nRightarrow$ r-阶平均收敛.

证明: 关于第一部分. 若$\xi_n \xrightarrow{L_r} \xi$, 则由Markov不等式知: 对任意的$\varepsilon > 0$,

$$\mathsf{P}(|\xi_n - \xi| \geq \varepsilon) \leq \frac{1}{\varepsilon^r}\mathsf{E}|\xi_n - \xi|^r \to 0.$$

所以有$\xi_n \xrightarrow{P} \xi$.

关于第二部分. 举例说明: 定义

$$\mathsf{P}(\xi_n = n) = \frac{1}{\log(n+3)}, \mathsf{P}(\xi_n = 0) = 1 - \frac{1}{\log(n+3)}, n = 1, 2, \cdots .$$

关于第二部分. 举例说明: 定义

$$\mathsf{P}(\xi_n = n) = \frac{1}{\log(n+3)}, \mathsf{P}(\xi_n = 0) = 1 - \frac{1}{\log(n+3)}, n = 1, 2, \cdots .$$

则$\mathsf{P}(|\xi_n - 0| \geq \varepsilon) = \mathsf{P}(\xi_n = n) = 1/\log(n+3) \to 0$. 所以$\xi_n \xrightarrow{P} 0$. 但对任何的$0 < r < \infty$,

$$\mathsf{E}|\xi_n - 0|^r = \mathsf{E}|\xi_n|^r =$$

关于第二部分. 举例说明: 定义

$$\mathsf{P}(\xi_n = n) = \frac{1}{\log(n+3)}, \mathsf{P}(\xi_n = 0) = 1 - \frac{1}{\log(n+3)}, n = 1, 2, \cdots .$$

则$\mathsf{P}(|\xi_n - 0| \geq \varepsilon) = \mathsf{P}(\xi_n = n) = 1/\log(n+3) \to 0$. 所以$\xi_n \xrightarrow{P} 0$.
但对任何的$0 < r < \infty$,

$$\mathsf{E}|\xi_n - 0|^r = \mathsf{E}|\xi_n|^r = \frac{n^r}{\log(n+3)} \to \infty.$$

所以$\xi_n \overset{L_r}{\nrightarrow} 0$.

以概率1收敛与强大数定律

以概率1收敛与强大数定律

4.3.1 以概率1收敛

以概率1收敛与强大数定律

4.3.1 以概率1收敛

我们已经知道随机变量是定义在概率空间$(\Omega, \mathscr{F}, \mathsf{P})$上取值为实数的函数, 对于固定的$\omega \in \Omega$, $\{\xi_n(\omega), n \in \mathbb{N}\}$就是实数列, 因此我们可以讨论$\xi_n(\omega)$的收敛性.

以概率1收敛与强大数定律

4.3.1 以概率1收敛

我们已经知道随机变量是定义在概率空间$(\Omega, \mathscr{F}, \mathsf{P})$上取值为实数的函数，对于固定的$\omega \in \Omega$，$\{\xi_n(\omega), n \in \mathbb{N}\}$就是实数列，因此我们可以讨论$\xi_n(\omega)$的收敛性.

但由于随机变量取值的随机性，我们常常不可能期望随机变量序列在所有样本点处都存在极限. 我们关心的问题是：极限是否在一个概率为1的点集上存在.

定义（以概率1收敛）

设ξ和$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量序列.

- 如果存在$\Omega_0 \in \mathscr{F}$, $P(\Omega_0) = 0$, 且对任意的$\omega \in \Omega \setminus \Omega_0$, 都有$\xi_n(\omega) \to \xi(\omega)$, 则称$\xi_n$以概率1收敛(converges with probability one)或几乎必然收敛(converges almost surely)于ξ, 记作$\xi_n \to \xi \;\; a.s.$
- 如果存在$\Omega_0 \in \mathscr{F}$, $P(\Omega_0) = 0$, 且对任意的$\omega \in \Omega \setminus \Omega_0$, 数列$\{\xi_n(\omega), n \geq 1\}$是柯西基本列, 即$\xi_n(\omega) - \xi_m(\omega) \to 0 \; (n > m \to \infty)$, 则称$\xi_n$以概率1是柯西基本列.

定义（以概率1收敛）

设ξ和$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量序列.

- 如果存在$\Omega_0 \in \mathscr{F}$, $P(\Omega_0) = 0$, 且对任意的$\omega \in \Omega \backslash \Omega_0$, 都有$\xi_n(\omega) \to \xi(\omega)$, 则称$\xi_n$以概率1收敛(converges with probability one)或几乎必然收敛(converges almost surely)于ξ, 记作$\xi_n \to \xi \ \ a.s.$
- 如果存在$\Omega_0 \in \mathscr{F}$, $P(\Omega_0) = 0$, 且对任意的$\omega \in \Omega \backslash \Omega_0$, 数列$\{\xi_n(\omega), n \geq 1\}$是柯西基本列, 即$\xi_n(\omega) - \xi_m(\omega) \to 0 \ (n > m \to \infty)$, 则称$\xi_n$以概率1是柯西基本列.

注: 以概率1收敛/a.s.收敛意味着最多除去一个零概率点集外, ξ_n逐点收敛于ξ. ξ_n以概率1收敛当且仅当ξ_n以概率1是柯西基本列.

以概率1收敛的判别准则:

以概率1收敛的判别准则:

定理

设ξ和$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量序列.
(1) $\xi_n \to \xi \ \ a.s.$当且仅当对任意的$\varepsilon > 0$,

$$\lim_{n\to\infty} P(\cup_{k\geq n}\{|\xi_k - \xi| \geq \varepsilon\}) = 0,$$

或者等价地

$$\lim_{n\to\infty} P(\sup_{k\geq n} |\xi_k - \xi| \geq \varepsilon) = 0.$$

(2) $\{\xi_n, n \geq 1\}$以概率1是柯西基本列当且仅当对任意的$\varepsilon > 0$,

$$\lim_{n\to\infty} P(\cup_{k\geq 0}\{|\xi_{n+k} - \xi_k| \geq \varepsilon\}) = 0,$$

或者等价地

$$\lim_{n\to\infty} P(\sup_{k\geq 0} |\xi_{n+k} - \xi_k| \geq \varepsilon) = 0.$$

证明: 第一部分. 记

$$A_n^\varepsilon = \{|\xi_n - \xi| \geq \varepsilon\}, \quad A^\varepsilon = \cap_{n=1}^\infty \cup_{k\geq n} A_k^\varepsilon.$$

则可知

$$\{\xi_n \nrightarrow \xi\} = \cup_{m=1}^\infty A^{1/m}.$$

由概率的连续性定理得

$$\mathsf{P}(A^\varepsilon) = \lim_{n\to\infty} \mathsf{P}(\cup_{k\geq n} A_k^\varepsilon).$$

则下列关系式成立:

$$\begin{aligned}
0 = \mathsf{P}(\xi_n \nrightarrow \xi) \quad & \Leftrightarrow \quad \mathsf{P}(\cup_{m=1}^\infty A^{1/m}) = 0 \\
& \Leftrightarrow \quad \mathsf{P}(A^{1/m}) = 0, \quad \forall m \geq 1 \\
& \Leftrightarrow \quad \mathsf{P}(\cup_{k\geq n} A_k^{1/m}) \to 0, \quad \forall m \geq 1 \\
& \Leftrightarrow \quad \mathsf{P}(\cup_{k\geq n}\{|\xi_k - \xi| \geq 1/m\}) \to 0, \quad \forall m \geq 1 \\
& \Leftrightarrow \quad \mathsf{P}(\cup_{k\geq n}\{|\xi_k - \xi| \geq \varepsilon\}) \to 0, \quad \forall \varepsilon > 0.
\end{aligned}$$

又

$$\mathsf{P}(\sup_{k\geq n}|\xi_k-\xi|\geq\varepsilon)=\mathsf{P}(\cup_{k\geq n}\{|\xi_k-\xi|\geq\varepsilon\}),$$

第一部分得证.

又

$$\mathsf{P}(\sup_{k\geq n}|\xi_k-\xi|\geq\varepsilon)=\mathsf{P}(\cup_{k\geq n}\{|\xi_k-\xi|\geq\varepsilon\}),$$

第一部分得证.

第二部分的结论类似可证, 只需令

$$B_{n,k}^{\varepsilon}=\{|\xi_{n+k}-\xi_n|\geq\varepsilon\},\ \ B^{\varepsilon}=\cap_{m=1}^{\infty}\cup_{n\geq m}\cup_{k\geq 1}B_{n,k}^{\varepsilon}$$

并注意到

$$\{\xi_n\text{不是柯西基本列}\}=\cup_{\varepsilon>0}B^{\varepsilon}.$$

推论

如果对任意的$\varepsilon > 0$, $\sum_{n=1}^{\infty} P(|\xi_n - \xi| \geq \varepsilon) < \infty$, 则$\xi_n \to \xi$ *a.s.*

推论

如果对任意的$\varepsilon > 0$, $\sum_{n=1}^{\infty} P(|\xi_n - \xi| \geq \varepsilon) < \infty$, 则$\xi_n \to \xi$ *a.s.*

证明: 因为

$$\mathsf{P}(\cup_{k\geq n}\{|\xi_k - \xi| \geq \varepsilon\}) \leq \sum_{k=n}^{\infty} \mathsf{P}(|\xi_k - \xi| \geq \varepsilon) \to 0,$$

所以$\xi_n \to \xi$ a.s.

注: 若$\sum_{n=1}^{\infty} \mathsf{P}(|\xi_n - \xi| \geq \varepsilon) < \infty$, 则称$\xi_n$完全收敛于$\xi$, 参见Hsu, P. L. and Robbins, H. (1947), Complete Convergence and the Law of Large Numbers, PNAS 33(2): 25-31.

许宝騄(1910—1970年)，字闲若，祖籍浙江杭州，1910年9月1日生于北京，因此小名京生

1936年许宝騄考取了赴英留学, 派往伦敦大学学院, 在统计系学习数理统计, 攻读博士学位, 师从著名统计学家奈曼(Neyman).

许宝騄在中国开创了概率论、数理统计的教学与研究工作. 在奈曼—皮尔逊理论、参数估计理论、多元分析、极限理论等方面取得卓越成就, 是多元统计分析学科的开拓者之一, 其研究成果推动了概率论与数理统计的发展.

他被公认为在数理统计和概率论方面第一个具有国际声望的中国数学家. 《中国大百科全书:数学》称赞说: "许宝騄是中国早期从事数理统计学和概率论研究并达到世界先进水平的一位杰出学者."

许宝騄曾执教加州大学伯克利分校、哥伦比亚大学、西南联合大学、北京大学, 与郝泰林(Hotelling)一起创办了北卡罗来纳大学统计系.

许宝騄的像片悬挂在斯坦福大学统计系的走廊上, 与世界著名的统计学家并列.

许宝騄是二十世纪和华罗庚、陈省身齐名的中国数学家, 中央研究院第一届当选的5名数学所院士之一(另外四位首届数学院士是姜立夫、陈省身、华罗庚、苏步青).

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$的随机变量序列. 若$\xi_n \to \xi$ a.s., 则$\xi_n \xrightarrow{P} \xi$. 反之不成立.

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$的随机变量序列. 若$\xi_n \to \xi$ a.s., 则$\xi_n \xrightarrow{P} \xi$. 反之不成立.

证明: 若$\xi_n \to \xi$ a.s., 则

$$\lim_{n\to\infty} \mathsf{P}(\sup_{k\geq n} |\xi_k - \xi| \geq \varepsilon) = 0.$$

所以

$$\lim_{n\to\infty} \mathsf{P}(|\xi_n - \xi| \geq \varepsilon) \leq \lim_{n\to\infty} \mathsf{P}(\sup_{k\geq n} |\xi_k - \xi| \geq \varepsilon) = 0,$$

即$\xi_n \xrightarrow{P} \xi$.

*取$(\Omega, \mathscr{F}, \mathsf{P})$为$(0,1)$上的几何概率空间, 定义$\xi = \xi(\omega) \equiv 0$, 令

$$\xi_n = I\left\{\frac{n-2^m}{2^m} < \omega < \frac{n+1-2^m}{2^m}\right\}, \ 2^m \le n < 2^{m+1}, \forall m \in \mathbb{N}.$$

则对任意的$r > 0$均有

$$\mathsf{E}|\xi_n - \xi|^r = \frac{1}{2^m} \to 0.$$

所以$\xi_n \xrightarrow{L_r} \xi$, 自然地有$\xi_n \xrightarrow{P} \xi$. 另一方面, 只要$\omega$不是有理数, 那么就有无穷多个$n$, 使得$\xi_n(\omega) = 1$, 所以

$$\mathsf{P}(\omega : \xi_n(\omega) \nrightarrow \xi(\omega)) = 1,$$

即$\xi_n \nrightarrow \xi$ a.s.

关于随机变量序列的依分布收敛, 依概率收敛, r-阶平均收敛, a.s.收敛, 我们有如下的结论:

关于随机变量序列的依分布收敛, 依概率收敛, r-阶平均收敛, a.s.收敛, 我们有如下的结论:

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$的随机变量序列.

- r-阶平均收敛与a.s.收敛互不蕴含;

关于随机变量序列的依分布收敛，依概率收敛，r-阶平均收敛，a.s.收敛，我们有如下的结论：

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$的随机变量序列.

- r-阶平均收敛与a.s.收敛互不蕴含；
- r-阶平均收敛与a.s.收敛都蕴含依概率收敛，但依概率收敛不蕴含r-阶平均收敛与a.s.收敛；

关于随机变量序列的依分布收敛, 依概率收敛, r-阶平均收敛, a.s.收敛, 我们有如下的结论:

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$的随机变量序列.

- r-阶平均收敛与a.s.收敛互不蕴含;
- r-阶平均收敛与a.s.收敛都蕴含依概率收敛, 但依概率收敛不蕴含r-阶平均收敛与a.s.收敛;
- 依概率收敛蕴含依分布收敛, 但依分布收敛不蕴含依概率收敛;

关于随机变量序列的依分布收敛, 依概率收敛, r-阶平均收敛, a.s.收敛, 我们有如下的结论:

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$的随机变量序列.

- r-阶平均收敛与a.s.收敛互不蕴含;
- r-阶平均收敛与a.s.收敛都蕴含依概率收敛, 但依概率收敛不蕴含r-阶平均收敛与a.s.收敛;
- 依概率收敛蕴含依分布收敛, 但依分布收敛不蕴含依概率收敛;
- 对于退化的随机变量c, 有$\xi_n \xrightarrow{d} c \Leftrightarrow \xi_n \xrightarrow{P} c$.

4.3.2 强大数定律

4.3.2 强大数定律

我们以前讨论的大数定律只要求依概率收敛, 若把收敛性要求提高为以概率1收敛, 则得到的大数定律称为强大数定律.

4.3.2 强大数定律

我们以前讨论的大数定律只要求依概率收敛, 若把收敛性要求提高为以概率1收敛, 则得到的大数定律称为强大数定律.

定义 (强大数定律/SLLN)

设$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量序列, 如果存在常数列$\{a_n, n \geq 1\}$和$\{b_n, n \geq 1\}$使得

$$\frac{1}{a_n}\sum_{k=1}^{n}\xi_k - b_n \to 0 \;\; a.s.,$$

则称$\{\xi_n, n \geq 1\}$服从强大数定律(strong law of large numbers).

第一个强大数定律是Borel在1909年针对Bernoulli试验场合给出的. 它是Bernoulli大数定律的加强版.

第一个强大数定律是Borel在1909年针对Bernoulli试验场合给出的. 它是Bernoulli大数定律的加强版.

定理 (Borel强大数定律)

设$\{\xi_k, k \geq 1\}$是一列独立同分布的随机变量序列, $P(\xi_k = 1) = p$, $P(\xi_k = 0) = 1 - p$, $0 < p < 1$, 记$S_n = \sum_{k=1}^n \xi_k$. 则

$$\frac{S_n}{n} \to p, \quad a.s.$$

第一个强大数定律是Borel在1909年针对Bernoulli试验场合给出的. 它是Bernoulli大数定律的加强版.

定理（Borel强大数定律）

设$\{\xi_k, k \geq 1\}$是一列独立同分布的随机变量序列, $P(\xi_k = 1) = p$, $P(\xi_k = 0) = 1 - p$, $0 < p < 1$, 记$S_n = \sum_{k=1}^n \xi_k$. 则

$$\frac{S_n}{n} \to p, \quad a.s.$$

证明: 只需证明对任意的$\varepsilon > 0$, 有

$$\mathrm{P}(\cup_{k\geq n}\{|S_k/k - p| \geq \varepsilon\}) \to 0.$$

因为

$$\mathrm{P}(\cup_{k\geq n}\{|S_k/k - p| \geq \varepsilon\}) \leq \sum_{k=n}^{\infty} \mathrm{P}(|S_k/k - p| \geq \varepsilon),$$

所以只需证明

$$\sum_{n=1}^{\infty} \mathsf{P}(|S_n/n - p| \geq \varepsilon) < \infty.$$

根据Markov不等式,

$$\mathsf{P}(|S_n/n - p| \geq \varepsilon) \leq \frac{1}{\varepsilon^4} \mathsf{E} \left| \frac{S_n}{n} - p \right|^4 .$$

这样, 问题就转化为计算S_n/n的四阶中心矩. 易知

$$\begin{aligned} & \mathsf{E} \left| \frac{S_n}{n} - p \right|^4 \\ = & \frac{1}{n^4} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \mathsf{E}[(\xi_i - p)(\xi_j - p)(\xi_k - p)(\xi_l - p)] \end{aligned}$$

上面的和式中只有$\mathsf{E}(\xi_i - p)^4$以及$\mathsf{E}[(\xi_i - p)^2(\xi_j - p)^2]$的项才不等于0, 显然

$$\mathsf{E}(\xi_i - p)^4 = pq(p^3 + q^3), \tag{4.3.1}$$

$$\mathsf{E}[(\xi_i - p)^2(\xi_j - p)^2] = p^2q^2 \ \ (i \neq j). \tag{4.3.2}$$

(4.3.1)形式的项有n项, (4.3.2)形式的项有$\binom{4}{2}\binom{n}{2} = 3n(n-1)$项, 因此

$$\mathsf{E}\left|\frac{S_n}{n} - p\right|^4 = \frac{pq}{n^4}[n(p^3 + q^3) + 3pq(n^2 - n)] < \frac{1}{4n^2}.$$

于是

$$\sum_{n=1}^{\infty} \mathsf{P}(|S_n/n - p| \geq \varepsilon) \leq \frac{1}{4\varepsilon^4} \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

结论得证.

Kolmogorov在1930年将Borel强大数定律进行了推广，使之适用于一般随机变量.

Kolmogorov在1930年将Borel强大数定律进行了推广，使之适用于一般随机变量.

定理（Kolmogorov强大数定律）

设$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的独立同分布随机变量序列. 记$S_n = \sum_{k=1}^n \xi_k$, 则

$$\frac{S_n}{n} \to \mu \ \ a.s.$$

成立的充要条件是$E|\xi_1| < \infty$且$\mu = E\xi_1$.

Kolmogorov在1930年将Borel强大数定律进行了推广，使之适用于一般随机变量.

定理（Kolmogorov强大数定律）

设$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的独立同分布随机变量序列. 记$S_n = \sum_{k=1}^{n} \xi_k$, 则

$$\frac{S_n}{n} \to \mu \ \ a.s.$$

成立的充要条件是$E|\xi_1| < \infty$且$\mu = E\xi_1$.

在Borel强大数定律中，ξ_n是有界随机变量，证明过程中需计算ξ_n的四阶矩. 而在Kolmogorov强大数定律，我们只假定ξ_n有一阶矩，更高阶矩可能不存在. 所以无法机械地模仿Borel强大数定律的证明过程去证明Kolmogorov强大数定律的充分性部分.

注: 以下内容供有兴趣的同学阅读, 目的是证明Kolmogorov强大数定律.

注: 以下内容供有兴趣的同学阅读, 目的是证明Kolmogorov强大数定律.

记

$$\limsup_{n\to\infty} A_n = \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_n, \quad \liminf_{n\to\infty} A_n = \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} A_n.$$

称$\limsup_{n\to\infty} A_n$为事件序列$\{A_n\}$的上限事件, 它表示A_n发生无穷多次, 因为$\omega \in \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_n$当且仅当$\omega$属于无穷多个$A_n$. 类似地称$\liminf_{n\to\infty} A_n$为事件序列$\{A_n\}$的下限事件, 它表示$A_n$至多只有有限个不发生, 因为$\omega \in \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} A_n$当且仅当存在一个$N$, 使$\omega \in \cap_{n=N}^{\infty}$, 因此若$\omega$发生, 则$A_N, A_{N+1}, \cdots$同时发生, 这时至多只有前面$N-1$个事件$A_1, \cdots, A_{N-1}$可能不发生.

注: 以下内容供有兴趣的同学阅读, 目的是证明Kolmogorov强大数定律.

记

$$\limsup_{n\to\infty} A_n = \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_n, \quad \liminf_{n\to\infty} A_n = \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} A_n.$$

称$\limsup_{n\to\infty} A_n$为事件序列$\{A_n\}$的上限事件, 它表示A_n发生无穷多次, 因为$\omega \in \cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_n$当且仅当$\omega$属于无穷多个$A_n$. 类似地称$\liminf_{n\to\infty} A_n$为事件序列$\{A_n\}$的下限事件, 它表示$A_n$至多只有有限个不发生, 因为$\omega \in \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} A_n$当且仅当存在一个$N$, 使$\omega \in \cap_{n=N}^{\infty}$, 因此若$\omega$发生, 则$A_N, A_{N+1}, \cdots$同时发生, 这时至多只有前面$N-1$个事件$A_1, \cdots, A_{N-1}$可能不发生.

容易看出$\xi_n \to \xi$ a.s.等价于对任意的$\varepsilon > 0$,

$$\mathsf{P}(\cap_{k=1}^{\infty} \cup_{n=k}^{\infty} \{|\xi_n - \xi| \geq \varepsilon\}) = 0,$$

即$\mathsf{P}(\limsup_{n\to\infty} A_n) = 0$.

下面的Borel-Cantelli引理在概率论中有着广泛的应用.

引理（Borel-Cantelli Lemma）

- 若事件序列$\{A_n\}$满足$\sum_{n=1}^{\infty} P(A_n) < \infty$, 则
$$P(\limsup_{n\to\infty} A_n) = 0, \quad P(\liminf_{n\to\infty} \bar{A}_n) = 1.$$
- 若$\{A_n\}$是相互独立的事件序列, 则$\sum_{n=1}^{\infty} P(A_n) = \infty$成立的充要条件是
$$P(\limsup_{n\to\infty} A_n) = 1 \text{ 或 } P(\liminf_{n\to\infty} \bar{A}_n) = 0.$$

证明: 第一部分. 注意到$\cup_{n=k}^{\infty} A_n$关于k是单调递减序列, 所以

$$\mathsf{P}(\limsup_{n\to\infty} A_n) = \lim_{k\to\infty} \mathsf{P}(\cup_{n=k}^{\infty} A_n) \le \lim_{k\to\infty} \sum_{n=k}^{\infty} \mathsf{P}(A_n) = 0.$$

证明：第一部分．注意到$\cup_{n=k}^{\infty} A_n$关于k是单调递减序列，所以

$$\mathsf{P}(\limsup_{n\to\infty} A_n) = \lim_{k\to\infty} \mathsf{P}(\cup_{n=k}^{\infty} A_n) \le \lim_{k\to\infty} \sum_{n=k}^{\infty} \mathsf{P}(A_n) = 0.$$

第二部分．先证必要性：注意到$\{A_n\}$的独立性，有

$$\begin{aligned}
\mathsf{P}(\liminf_{n\to\infty} \bar{A}_n) &= \mathsf{P}(\cup_{k=1}^{\infty} \cap_{n=k}^{\infty} \bar{A}_n) \\
&\le \sum_{k=1}^{\infty} \mathsf{P}(\cap_{n=k}^{\infty} \bar{A}_n) \\
&= \sum_{k=1}^{\infty} \prod_{n=k}^{\infty} \mathsf{P}(\bar{A}_n) \\
&= \sum_{k=1}^{\infty} \prod_{n=k}^{\infty} [1 - \mathsf{P}(A_n)].
\end{aligned}$$

由于$1+x\le e^x$对任意的$x\in\mathbb{R}$成立，所以由$\sum_{n=1}^{\infty}\mathsf{P}(A_n)=\infty$知

$$\prod_{n=k}^{\infty}[1-\mathsf{P}(A_n)]\le\lim_{N\to\infty}\exp\left\{-\sum_{n=k}^{N}\mathsf{P}(A_n)\right\}=0.$$

所以$\mathsf{P}(\liminf_{n\to\infty}\bar{A}_n)=0.$

由于$1+x\le e^x$对任意的$x\in\mathbb{R}$成立，所以由$\sum_{n=1}^{\infty}\mathsf{P}(A_n)=\infty$知

$$\prod_{n=k}^{\infty}[1-\mathsf{P}(A_n)]\le\lim_{N\to\infty}\exp\left\{-\sum_{n=k}^{N}\mathsf{P}(A_n)\right\}=0.$$

所以$\mathsf{P}(\liminf_{n\to\infty}\bar{A}_n)=0$.

再证充分性. 若$\mathsf{P}(\limsup_{n\to\infty}A_n)=1$. 假定$\sum_{n=1}^{\infty}\mathsf{P}(A_n)$ $<\infty$, 则由第一部分的结论知$\mathsf{P}(\limsup_{n\to\infty}A_n)=0$, 产生矛盾. 因$\mathsf{P}(A_n)\ge 0$, 故只能是$\sum_{n=1}^{\infty}\mathsf{P}(A_n)=\infty$.

引理

设ξ是一随机变量, 则下列三条等价:

(1) $E|\xi| < \infty$,

(2) $\sum_{n=0}^{\infty} P(|\xi| \geq n) < \infty$,

(3) $\sum_{n=1}^{\infty} nP(n-1 \leq |\xi| < n) < \infty$.

引理

设ξ是一随机变量，则下列三条等价:

(1) $\mathsf{E}|\xi| < \infty$,

(2) $\sum_{n=0}^{\infty} P(|\xi| \geq n) < \infty$,

(3) $\sum_{n=1}^{\infty} nP(n-1 \leq |\xi| < n) < \infty$.

证明: (2)和(3)等价是因为

$$\begin{aligned}\sum_{n=0}^{\infty} \mathsf{P}(|\xi| \geq n) &= \sum_{n=0}^{\infty} \sum_{k=n}^{\infty} \mathsf{P}(k \leq |\xi| < k+1) \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{k} \mathsf{P}(k \leq |\xi| < k+1) \\ &= \sum_{k=0}^{\infty} (k+1)\mathsf{P}(k \leq |\xi| < k+1).\end{aligned}$$

下证(1)和(3)等价. 因为

$$\begin{aligned}\mathsf{E}|\xi| &= \sum_{n=1}^{\infty} \mathsf{E}|\xi| I\{n-1 \le |\xi| < n\} \\ &\le \sum_{n=1}^{\infty} n \mathsf{E} I\{n-1 \le |\xi| < n\} \\ &= \sum_{n=1}^{\infty} n \mathsf{P}(n-1 \le |\xi| < n),\end{aligned}$$

因此(3)推出(1); 反过来,

$$\begin{aligned}\mathsf{E}|\xi| &\ge \sum_{n=1}^{\infty} (n-1) \mathsf{E} I\{n-1 \le |\xi| < n\} \\ &= \sum_{n=1}^{\infty} n \mathsf{P}(n-1 \le |\xi| < n) - 1,\end{aligned}$$

故(1)推出(3).

引理 (Hájek-Rényi Inequality)

若$\{\xi_i, i \geq 1\}$是独立随机变量序列, $Var\xi = \sigma_i^2 < \infty$, $i = 1, 2, \cdots$. 而$\{C_n, n \geq 1\}$是一列正的非增常数序列, 则对任意正整数m, n $(m < n)$及$\varepsilon > 0$有

$$P\left(\max_{m \leq j \leq n} C_j \left|\sum_{i=1}^{j}(\xi_i - E\xi_i)\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\left(C_m^2 \sum_{j=1}^{m} \sigma_j^2 + \sum_{j=m+1}^{n} C_j^2 \sigma_j^2\right).$$

证明: 略.

在Hájek-Rényi 不等式中令$m = 1$, $C_j = 1$, 则得到Kolmogorov不等式.

在Hájek-Rényi 不等式中令$m=1$, $C_j=1$, 则得到Kolmogorov不等式.

引理（Kolmogorov Inequality）

设$\xi_1,\cdots,\xi_n$是独立随机变量序列, 方差存在, 则对任意的$\varepsilon>0$, 成立

$$P\left(\max_{1\le j\le n}\left|\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\ge\varepsilon\right)\le\frac{1}{\varepsilon^2}\sum_{i=1}^{n}Var\xi_i.$$

定理

设$\{\xi_n, n \geq 1\}$ 是独立随机变量序列，且$\sum_{n=1}^{\infty} \frac{Var\xi_n}{n^2} < \infty$，则

$$\frac{1}{n}\sum_{n=1}^{\infty}(\xi_n - E\xi_n) \to 0 \ \ a.s.$$

定理

设$\{\xi_n, n \geq 1\}$ 是独立随机变量序列，且$\sum_{n=1}^{\infty} \frac{Var\xi_n}{n^2} < \infty$，则

$$\frac{1}{n}\sum_{n=1}^{\infty}(\xi_n - E\xi_n) \to 0 \ \ a.s.$$

证明：由概率的连续性以及Hájek-Rényi不等式知：对任意的$\varepsilon > 0$,

$$\begin{aligned} &\mathsf{P}\left(\sup_{j\geq m}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i - \mathsf{E}\xi_i)\right| \geq \varepsilon\right) \\ =&\lim_{n\to\infty}\mathsf{P}\left(\sup_{m\leq j\leq n}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i - \mathsf{E}\xi_i)\right| \geq \varepsilon\right) \\ \leq&\frac{1}{\varepsilon^2}\left(\frac{1}{m^2}\sum_{i=1}^{m}\mathsf{Var}\xi_i + \sum_{i=m+1}^{\infty}\frac{\mathsf{Var}\xi_i}{i^2}\right). \end{aligned}$$

因为$\sum_{n=1}^{\infty}\frac{\mathsf{Var}\xi_i}{n^2}<\infty$, 所以

$$\lim_{m\to\infty}\mathsf{P}\left(\sup_{j\geq m}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-\mathsf{E}\xi_i)\right|\geq\varepsilon\right)=0.$$

由此可得$\frac{1}{n}\sum_{n=1}^{\infty}(\xi_n-\mathsf{E}\xi_n)\to 0$ a.s.

Kolmogorov强大数律的证明: 必要性. 若$S_n/n \to a$ a.s., 这里a是某一常数. 则

$$\frac{\xi_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \cdot \frac{S_{n-1}}{n-1} \to 0 \;\; a.s.$$

所以

$$\mathsf{P}(\limsup_{n\to\infty}\{|\xi_n| \geq n\}) = 0.$$

由Borel-Cantelli引理的第二部分结论知

$$\sum_{n=1}^{\infty} \mathsf{P}(|\xi_n| \geq n) < \infty.$$

这等价于$\mathsf{E}|\xi_k| < \infty$. 这时显然有$a = \mathsf{E}\xi_k$(否则, 用反证法可得到矛盾的结论).

充分性. 用截尾方法(truncation method), 令

$$\xi_n^* = \xi_n I\{|\xi_n| \le n\}.$$

容易看出

$$\mathsf{Var}\xi_n^* \le \mathsf{E}\xi_n^{*2} = \int_{-n}^{n} x^2 \mathrm{d}F(x) \le \sum_{k=1}^{n} k^2 \mathsf{P}(k-1 \le |\xi_n| < k),$$

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{\mathsf{Var}\xi_n^*}{n^2} &\le \sum_{n=1}^{\infty}\sum_{k=1}^{n} \frac{k^2}{n^2} \mathsf{P}(k-1 \le |\xi_n| < k) \\ &= \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} \mathsf{P}(k-1 \le |\xi_n| < k) \\ &= \sum_{k=1}^{\infty} k^2 \mathsf{P}(k-1 \le |\xi_n| < k) \sum_{n=k}^{\infty} \frac{1}{n^2}.\end{aligned}$$

由于

$$\sum_{n=k}^{\infty}\frac{1}{n^2} < \frac{1}{k^2} + \sum_{n=k+1}^{\infty}\frac{1}{n(n-1)} = \frac{1}{k^2} + \frac{1}{k} \le \frac{2}{k},$$

故

$$\sum_{n=1}^{\infty}\frac{\mathsf{Var}\xi_n^*}{n^2} < 2\sum_{k=1}^{\infty} k\mathsf{P}(k-1 \le |\xi_n| < k) = 2\mathsf{E}|\xi_n| < \infty.$$

因此由前面的强大数定律得

$$\frac{1}{n}\sum_{n=1}^{\infty}(\xi_n^* - \mathsf{E}\xi_n^*) \to 0 \;\; a.s.$$

显然

$$\mathsf{E}\xi_n^* = \int_{-n}^{n} x\mathsf{d}F(x) \to \mathsf{E}\xi_n = \mu,$$

所以

$$\frac{1}{n}\sum_{k=1}^{n}(\mathsf{E}\xi_k^* - \mu) \to 0.$$

由于

$$\left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mu)\right| \le \left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k - \xi_k^*)\right| + \left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k^* - \mathsf{E}\xi_k^*)\right| + \left|\frac{1}{n}\sum_{k=1}^{n}(\mathsf{E}\xi_k^* - \mu)\right|,$$

所以为了完成整个证明, 我们只需证明

$$\frac{1}{n}\sum_{k=1}^{n}(\xi_k - \xi_k^*) \to 0 \ \ a.s. \tag{4.3.3}$$

然而

$$\sum_{k=1}^{\infty} \mathsf{P}(\xi_k \neq \xi_k^*) = \sum_{k=1}^{\infty} \mathsf{P}(|\xi_k| \geq k) = \sum_{k=1}^{\infty} \mathsf{P}(|\xi_1| \geq k) \leq \mathsf{E}|\xi_1| < \infty,$$

由Borel-Cantelli引理知, $\xi_k(\omega) \neq \xi_k^*(\omega)$无穷多次发生的概率为0. 因此(4.3.3)成立. 证毕.