Lecture 10

大数定律与中心极限定理

Probability and Statistics
Beihang University

10.1　大数定律与中心极限定理简介

大数定律是关于随机序列 $\{X_n\}$ 部分和 $S_n = \sum_{i=1}^{n} X_i$ 的渐进性质的结论. 记 $\overline{X}_n = S_n/n$ 为 X_n 的平均值序列或样本平均序列. 假设 $\{X_n\}$ 独立同分布, 具有期望值 μ. 那么, 大数定律给出了 $\overline{X}_n$ 到 μ 的收敛性 . 根据收敛性的不同类别,

$$\overline{X}_n \xrightarrow{a.e} \mu \text{或者} \overline{X}_n \xrightarrow{P} \mu$$

大数定律被分别称为强大数定律和弱大数定律. 在近代数学中, 该定理也出现不同形式的推广. 本课将介绍经典意义下的强大数定律及其常用形式. 并证明一个简单情形的大数定律.

中心极限定理是统计学的重要基础, 它指出在方差有限的条件下, 独立同分布序列的部分和近似地具有正太分布.

大数定律给出了 $\left|\overline{X}_n - \mu\right|$ 的收敛性, 而中心极限定理然回答了收敛速度的问题：当 n 多大时, $\overline{X}_n$ 才能达到我们所希望的接近程度?

10.2 大数定律

随机序列收敛性

定义 10.2.1 $X, \{X_n\}$ 为随机变量, 考虑集合 $\{X_n\}$ 的收敛集合

$$A = \left\{\omega : \lim_n X_n(\omega) = X(\omega)\right\}$$

我们称 $\{X_n\}$ 以概率 1 收敛到 X, 如果成立

$$P(A) = 1$$

记作 $P(\lim_n X_n = X) = 1$ 或者 $X_n \xrightarrow{a.e} X$.

对于给定的样本 ω, $\{X_n(\omega)\}$ 不收敛到 $X(\omega)$ 可以等价地表达为: 存在 $\varepsilon > 0$, 对任意 $k \geqslant 1$, 总存在 $n \geqslant k$ 使得

$$|X_n(\omega) - X(\omega)| \geqslant \varepsilon$$

用集合将这一叙述表达出来就是：

$$A^c = \left[\omega : \lim_n X_n = X\right]^c = \bigcup_{\epsilon>0} \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} [\omega : |X_n - X| \geqslant \varepsilon]$$

如果 $P(\lim_n X_n = X) = 1$, 那么 $P(A^c) = 0$, 并且由上式可知, 对 $\forall \varepsilon > 0$ 都有

$$P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} [\omega : |X_n - X| \geqslant \varepsilon]\right) = 0$$

根据概率的上连续性, 这就是

$$\lim_n P\left(\bigcup_{n=k}^{\infty} [\omega : |X_n - X| \geqslant \varepsilon]\right) = P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} [\omega : |X_n - X| \geqslant \varepsilon]\right) = 0$$

因此

$$P(\omega : |X_n - X| \geqslant \varepsilon) \leqslant P\left(\bigcup_{n=k}^{\infty} [\omega : |X_n - X| \geqslant \varepsilon]\right) \to 0$$

即对 $\forall\varepsilon>0$ 都有

$$\lim_n P\left(\omega:|X_n-X|\geqslant\varepsilon\right)=0. \tag{10.1}$$

基于这一极限形式我们引入：

定义 10.2.2 如果对 $\forall\varepsilon>0$, 都有 (*Equation* 10.1) 成立, 则称 $\{X_n\}$ 依概率收敛到 X, 并记为 $X_n\xrightarrow{P}X$

前面的推导以及定义总结起来就是

定理 10.2.1 $X,\{X_n\}$ 为随机变量. 如果 $X_n\xrightarrow{a.e}X$, 那么 $X_n\xrightarrow{P}X$.

下面给出两个将用到的性质,

定理 10.2.2 $X,Y,\{X_n\},\{Y_n\}$ 为随机变量

(1) $X_n\xrightarrow{a.e}X$, $Y_n\xrightarrow{a.e}Y$, 那么 $X_n\pm Y_n\xrightarrow{a.e}X\pm Y$

(2) $X_n\xrightarrow{P}X$, $Y_n\xrightarrow{P}Y$, 那么 $X_n\pm Y_n\xrightarrow{P}X\pm Y$

(1) $X_n \xrightarrow{a.e} X$, 即存在集合 $A_X, P(A_X) = 0$ 使得 $X_n(\omega) \to X(\omega)$, $\forall \omega \in A_X^c$. 同样存在集合 A_Y, $P(A_Y) = 0$ 使得 $Y_n(\omega) \to Y(\omega)$, $\forall \omega \in A_Y^c$. 因此 $X_n(\omega) \pm Y_n(\omega) \to X(\omega) \pm Y(\omega)$, $\forall \omega \in A_X^c \cap A_Y^c$, 并且

$$P\left(\left(A_X^c \cap A_Y^c\right)^c\right) = P\left(A_X \cup A_Y\right) \leqslant P\left(A_X\right) + P\left(A_Y\right) = 0$$

因此 $X_n \pm Y_n \xrightarrow{a.e} X \pm Y$.

(2) 只需注意到 $\forall \varepsilon > 0$,

$$\begin{aligned} P\left(\left|(X_n \pm Y_n) - (X \pm Y)\right| \geqslant \varepsilon\right) \leqslant\ & P\left(\left|X_n \pm X\right| \geqslant \varepsilon/2\right) + \\ & P\left(\left|Y_n \pm Y\right| \geqslant \varepsilon/2\right) \to 0 \end{aligned}$$

强大数定律

定理 10.2.3 (Kolmogorov) $\{X_n\}$ 为独立同分布随机序列, $\mu = E(X_1)$ 存在. 那么

$$\frac{S_n}{n} \xrightarrow{a.e.} \mu$$

注记 10.2.1 该定理并不对 $\{X_n\}$ 的二阶矩进行假设, 然而它们必须是可积的, 即期望存在. 由于 $\{X_n\}$ 同分布, 它们有公共的期望值 $\mu \in \mathbb{R}$. 根据定理 (10.2.1), 该定理的一个推论是

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

关于 S_n/n 依概率收敛到 μ 的结论通常被称为弱大数定律.

下面证明简单情形的弱大数定律, $\{X_n\}$ 的独立性可放宽为两两不相关.

定理 10.2.4 $\{X_n\}$ 随机序列, $\forall i, j$, X_i 与 X_j 不相关, 存在常数 K 使得 $E\left(X_i^2\right) \leqslant K, \forall i$. $S_n = \sum_{i=1}^n X_i$. 那么 $\frac{S_n - E(S_n)}{n} \xrightarrow{P} 0$.

- 注意由 Schwarz 不等式,

$$(E(X_i))^2 \leqslant E\left(X_i^2\right), \forall i$$

因此 $|E(X_i)| < \infty$, 并且

$$0 \leqslant Var(X_i) = E\left(X_i^2\right) - (E(X_i))^2 \leqslant E\left(X_i^2\right) \leqslant K.$$

由于 $\forall i, j$, X_i 与 X_j 不相关,

$$E\left[(S_n - E(S_n))^2\right] = E\left[\left(\sum_{i=1}^n (X_i - E(X_i))\right)^2\right]$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} E\left[\left(X_i - E\left(X_i\right)\right)^2\right] \\
&\leqslant nK
\end{aligned}
$$

从而

$$E\left[\left(\frac{S_n - E\left(S_n\right)}{n}\right)^2\right] \leqslant \frac{K}{n} \to 0 \ \left(n \to \infty\right)$$

同时, 由 Markov 不等式, $\forall \varepsilon > 0$,

$$P\left(\left|\frac{S_n - E\left(S_n\right)}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{1}{\varepsilon^2} E\left[\left(\frac{S_n - E\left(S_n\right)}{n}\right)^2\right]$$

综合上面两个不等式便有

$$\frac{S_n - E\left(S_n\right)}{n} \xrightarrow{P} 0$$

例题 10.2.1 (Bernoulli 序列) 假设 $\{X_n\}$ 为独立同分布的 Bernoulli 序列, $P(X_1 = 1) = p$, $P(X_1 = 0) = 1 - p$. $S_n = \sum_{i=1}^{n} X_i$ 为 Bernoulli 试验成功的次数. S_n/n 为 n 次试验中成功试验所占的比例. $E(S_n) = np$. 那么由大数定律 $(S_n - np)/n \xrightarrow{a.e.} 0$, 即有 $S_n/n \xrightarrow{a.e.} p$. 这实际上为试验成功的比例即为近似的成功概率这一直觉提供了依据.

例题 10.2.2 (Monte Carlo) 假设 $f(x)$ 是 $[0,1]$ 上的可积函数, 为了估计其积分数值, 取 $[0,1]$ 上均匀分布的独立随机序列 U_n, 根据大数定律则有

$$\frac{1}{n}\sum_{k=1}^{n} f(U_k) \xrightarrow{a.e.} E(f(U_1)) = \int_0^1 f(x)\,dx.$$

例题 10.2.3 (估计未知分布) 假设 $\{X_n\}$ 是从某分布 $F(x)$ 中随机抽取得到的观测序列, $F(x)$ 为未知, 希望对它进行估计. 给定 x, 令 $Y_n(\omega) = 1_{\{X_n \leqslant x\}}(\omega)$, 那么 Y_n 是独立同分布的有界随机序列,

因此由大数定律

$$F_n(x) \triangleq \frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow{a.e.} E(Y_1) = P(X_1 \leqslant x) = F(x).$$

例题 10.2.4 某水源中含有多种重金属, 为了检测其中铅的含量占到总重金属含量的百分比, 工作人员将铅的含量百分比记为随机变量 R, 于是对铅含量百分比的一个估计就是期望 $E(R)$. 由于 R 的分布是未知的. 工作人员抽取 n 个水样, 并测量得到其中铅的含量百分比 $R_1,...R_n$. 工作人员计算了 $\overline{R}_n$, 并把它作 $E(R)$ 的估计值. 试解释其合理性, 并说明工作人员应该至少抽取多少个水样, 才能使得样本均值与 $E(R)$ 的误差大于 0.005 的概率被控制在 0.02 之内.

■ 由题设 R_i 独立同分布, $|R_i| \leqslant 1, \forall i$. 因此 $E(|R_i|) \leqslant 1$, $Var(R_1) \leqslant 1/4$ (为什么?). 根据大数定律, $\overline{R}_n \xrightarrow{a.e.} E(R)$, 因此工作人员的做法是合理的. 为了估计所需的最少水样数目 n, 我们

借助于 Chebyshev 不等式, 取 $\varepsilon = 0.005$, 并注意 $E\left(\overline{R}_n\right) = E(R)$,

$$
\begin{aligned}
P\left(\left|\overline{R}_n - E(R)\right| \geqslant \varepsilon\right) &\leqslant \frac{E\left(\left(\overline{R}_n - E(R)\right)^2\right)}{\varepsilon^2} \\
&= \frac{E\left(\left(\overline{R}_n - E\left(\overline{R}_n\right)\right)^2\right)}{\varepsilon^2} \\
&= \frac{Var\left(R_1\right)}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2}
\end{aligned}
$$

可见, 只要 $\frac{1}{4n\varepsilon^2} \leqslant 0.02$ 就能满足所需要误差概率, 即

$$
n \geqslant \frac{1}{4 \times 0.02 \times (0.005)^2} = 500,000.
$$

10.3 中心极限定理

几个记号：用 $\Phi(x)$ 表示 $\mathcal{N}(0,1)$ 的分布函数：

$$\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy.$$

若 $\{Y_n\}$ 为任意随机序列, 方差存在且不为零, 若令 $Y_n^* = \frac{Y_n - E(Y_n)}{\sqrt{Var(Y_n)}}$, 那么 $E(Y_n^*) = 0$, $Var(Y_n^*) = 1$, 这时我们称 Y_n^* 为 Y_n 的正规化 .

定理 10.3.1 $\{X_n\}$ 为独立同分布随机序列, $\mu = E(X_1)$ 存在, $\sigma^2 = Var(X_1)$ 存在. $\sigma > 0$. $S_n = \sum_{i=1}^{n} X_i$. 那么, $\forall x$

$$F_{S_n^*}(x) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leqslant x\right) \to \Phi(x).$$

我们称 S_n^* **依分布收敛**到 $\mathcal{N}(0,1)$.

中心极限定理也常常写作

$$P\left(\frac{\overline{X}_n-\mu}{\sigma/n^{1/2}}\leqslant x\right)\to\Phi\left(x\right)$$

这一定理指出, 当 n 充分大, 近似地 $n^{1/2}\left(\overline{X}_n-\mu\right)/\sigma\sim\mathcal{N}\left(0,1\right)$.

例题 10.3.1 (Bernoulli 序列) 假设 $\{X_n\}$ 为独立同分布的 Bernoulli 序列, $P\left(X_1=1\right)=p$, $P\left(X_1=0\right)=1-p$. 那么 $S_n=\sum_{i=1}^{n}X_i$ 是参数为 n 和 p 的二项分布. $E\left(S_n\right)=np$, $Var\left(S_n\right)=np\left(1-p\right)$. 那么由中心极限定理

$$P\left(S_n\leqslant x\right)=P\left(\frac{S_n-n\mu}{\sigma\sqrt{n}}\leqslant\frac{x-n\mu}{\sigma\sqrt{n}}\right)\approx\Phi\left(\frac{x-n\mu}{\sigma\sqrt{n}}\right),\forall x.$$

例题 10.3.2 (Poisson 序列) 假设 $\{X_n\}$ 为独立同分布的随机序列, 每一个 X_i 都服从参数为 λ 的 Poisson 分布. 那么 $E\left(X_i\right)=Var\left(X_i\right)=\lambda$. 那么 $n^{1/2}\left(\overline{X}_n-\lambda\right)/\lambda^{1/2}$ 近似地服从 $\mathcal{N}\left(0,1\right)$. 因此

当 $c>0$, 我们有

$$\begin{aligned} P\left(\left|\overline{X}_n-\lambda\right|<c\right) &= P\left(\left|n^{1/2}\left(\overline{X}_n-\lambda\right)/\lambda^{1/2}\right|<cn^{1/2}\lambda^{-1/2}\right) \\ &\approx \Phi\left(cn^{1/2}\lambda^{-1/2}\right)-\Phi\left(-cn^{1/2}\lambda^{-1/2}\right) \\ &= 2\Phi\left(cn^{1/2}\lambda^{-1/2}\right)-1. \end{aligned}$$

例题 10.3.3 再回到水源污染检测的例题 (10.2.4), 试运用中心极限定理说明工作人员应该至少抽取多少个水样, 才能使得样本均值与 $E(R)$ 的误差大于 0.005 的概率被控制在 0.02 之内.

■ 在例题 (10.2.4) 中我们已经看到 $\{R_i\}$ 独立同分布, $E(|R_i|)\leqslant 1$, $Var(R_i)\leqslant 1/4$. 记 $\mu=E(R_1)$, $\sigma=(Var(R_1))^{1/2}$. 根据中心极限定理, 取 $\varepsilon=0.005$,

$$\begin{aligned} P\left(\left|\overline{R}_n-\mu\right|<\varepsilon\right) &\approx 2\Phi\left(\varepsilon n^{1/2}\sigma^{-1}\right)-1 \\ &\geqslant 2\Phi\left(2\varepsilon n^{1/2}\right)-1 \end{aligned}$$

可见, 只要 $2\Phi\left(2\varepsilon n^{1/2}\right)-1 \geqslant 0.98$ 就能满足所需要误差概率, 即

$$
\begin{aligned}
n &\geqslant \frac{1}{4\varepsilon^2}\left[\Phi^{-1}\left(\frac{1+0.98}{2}\right)\right]^2 \\
&\approx \frac{1}{4\times(0.005)^2}\left[\Phi^{-1}(0.99)\right]^2 \approx 54,103.
\end{aligned}
$$

这比我们之前估计的所需水样数目 $500,000$ 要小很多.