第 8 章　大数定律及中心极限定理

8.1 大数定律

给定随机变量 $X_1, X_2, \cdots, X_n$, 这些随机变量的均值 (算术平均值) 为

$$\frac{1}{n}\sum_{i=1}^{n} X_i.$$

当 n 非常大时, 大数定律考虑随机变量的均值是否具有稳定性.

定义 8.1 (依概率收敛) 设 $X_1, X_2, \cdots, X_n, \cdots$ 是一随机变量序列, a 是一常数, 如果对任意 $\epsilon > 0$ 有

$$\lim_{n\to\infty} \Pr\{|X_n - a| < \epsilon\} = 1 \quad 或 \quad \lim_{n\to\infty} \Pr\{|X_n - a| > \epsilon\} = 0,$$

则称随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 依概率收敛于 a, 记 $X_n \xrightarrow{P} a$.

问题: 与数列极限的区别? 下面我们给出依概率的性质:

1) 若 $X_n \xrightarrow{P} a$ 且函数 $g\colon \mathbb{R} \to \mathbb{R}$ 在 $X = a$ 点连续, 则$g(X_n) \xrightarrow{P} g(a)$.
2) 若 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$, 函数 $g\colon \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ 在点 $(X, Y) = (a, b)$ 处连续, 则 $g(X_n, Y_n) \xrightarrow{P} g(a, b)$.

例如: 如果 $X_n \xrightarrow{P} a$ 和 $Y_n \xrightarrow{P} b$, 那么 $X_n + Y_n \xrightarrow{P} a + b$ 和 $X_nY_n \xrightarrow{P} ab$.

定理 8.1 (大数定律) 若随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 满足

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} E[X_i],$$

则称 $\{X_n\}$ 服从大数定律.

大数定理刻画了随机变量的均值 (算术平均值) 依概率收敛于期望的均值 (算术平均值). 下面介绍几种大数定律:

定理 8.2 (马尔可夫 Markov 大数定律) 如果随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 满足

$$\frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \to 0 \qquad n \to \infty,$$

则 $\{X_n\}$ 服从大数定理.

马尔可夫大数定律不要求随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立或同分布, 其证明直接通过 Chebyshev 不等式有

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X_i])\right| \geqslant \epsilon\right] \leqslant \frac{1}{n^2\epsilon^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \to 0 \qquad n \to \infty.$$

定理 8.3 (切比雪夫 Chebyshev 大数定律) 设随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立, 且存在常数 $c > 0$ 使得 $\mathrm{Var}(X_n) \leqslant c$, 则 $\{X_n\}$ 服从大数定律.

此处独立的随机变量可以修改为'不相关随机变量'. 证明直接通过切比雪夫不等式

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X_i])\right| \geqslant \epsilon\right] \leqslant \frac{1}{\epsilon^2 n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \leqslant \frac{c}{n\epsilon^2} \to 0 \qquad n \to \infty.$$

定理 8.4 (辛钦 Khintchine 大数定律) 设 $X_1, X_2, \cdots, X_n, \cdots$ 为独立同分布随机变量序列, 且每个随机变量的期望 $E[X_i] = \mu$ 存在, 则 $\{X_n\}$ 服从大数定律.

辛钦大数定律不要求方差一定存在, 其证明超出了本书范围.

定理 8.5 (Bernoulli大数定律) 设随机变量序列 $X_n \sim B(n,p)$ $(p > 0)$, 对任意 $\epsilon > 0$ 有

$$\lim_{n\to\infty} \Pr\left[\left|\frac{X_n}{n} - p\right| \geqslant \epsilon\right] = 0,$$

即 $X_n/n \xrightarrow{P} p$.

定理的证明依据二项分布的性质: 独立同分布随机变量 $Y_1, Y_2, \ldots, Y_n$ 满足 $Y_i \sim \mathrm{Ber}(p)$, 则

$$X_n = \sum_{i=1}^{n} Y_i \sim \mathrm{B}(n,p).$$

于是得到

$$\lim_{n\to\infty} \Pr\left[\left|\frac{X_n}{n} - p\right| \geqslant \epsilon\right] = \lim_{n\to\infty} \Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - E[Y_i]\right| \geqslant \epsilon\right] \leqslant \frac{1}{\epsilon^2 n^2}\mathrm{Var}(\sum_{i=1}^{n} Y_i) = \frac{p(1-p)}{\epsilon^2 n} \to 0.$$

如何判断随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 满足大数定律:

- 若随机变量独立同分布, 则利用辛钦大数定律查看期望是否存在;
- 对非独立同分布随机变量, 则利用 Markov 大数定律判断方差是否趋于零.

例 8.1 独立的随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 满足 $\Pr\{X_n = n^{1/4}\} = \Pr\{X_n = -n^{1/4}\} = 1/2$. 证明 $\{X_n\}$ 服从大数定律.

证明 根据题意可得 $E[X_i]=0$, 以及 $\mathrm{Var}(X_i)=E[X_i^2]=i^{1/2}$, 根据 Chebysheve 不等式和独立性有

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^n X_i\right|\geqslant\epsilon\right]\leqslant\frac{1}{n^2\epsilon^2}\mathrm{Var}(\sum_{i=1}^n X_i)=\frac{1}{n^2\epsilon^2}\sum_{i=1}^n Var(X_i)=\frac{1}{\epsilon^2}\frac{1}{n^2}\sum_{i=1}^n i^{1/2}\leqslant\frac{1}{\epsilon^2\sqrt{n}}$$

再根据

$$\sum_{i=1}^n i^{1/2}\leqslant\sum_{i=1}^n\int_i^{i+1}i^{1/2}dx\leqslant\sum_{i=1}^n\int_i^{i+1}x^{1/2}dx=\int_1^{n+1}x^{1/2}dx=2((n+1)^{3/2}-1)/3$$

由此可得当 $n\to+\infty$ 时有

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^n X_i\right|\geqslant\epsilon\right]\leqslant\frac{2((n+1)^{3/2}-1)/3}{\epsilon^2n^2}\to 0$$

大数定律小结:

- Markov 大数定律: 若随机变量序列 $\{X_i\}$ 满足 $\mathrm{Var}(\sum_{i=1}^n X_n)/n^2\to 0$, 则满足大数定律;
- Chebyshev 大数定律: 若独立随机变量序列 $\{X_i\}$ 满足 $\mathrm{Var}(X_i)\leqslant c$, 则满足大数定律;
- Khintchine 大数定律: 若独立同分布随机变量序列 $\{X_i\}$ 期望存在, 则满足大数定律;
- Bernoulli 大数定律: 对二项分布 $X_n\sim\mathrm{B}(n,p)$, 有 $X_n/n\xrightarrow{P}p$.

8.2 中心极限定理

对独立的随机变量序列 $X_1,X_2,\cdots,X_n,\cdots$, 我们考虑标准化后随机变量

$$Y_n=\frac{\sum_{i=1}^n X_i-\sum_{i=1}^n E(X_i)}{\sqrt{\mathrm{Var}(\sum_{i=1}^n X_i)}}$$

的极限分布是否为服从正态分布. 首先介绍依分布收敛.

定义 8.2 设随机变量 Y 的分布函数为 $F_Y(y)=\Pr(Y\leqslant y)$, 以及随机变量序列 $Y_1,Y_2,\cdots,Y_n,\cdots$ 的分布函数分别为 $F_{Y_n}(y)=\Pr(Y_n\leqslant y)$, 如果

$$\lim_{n\to\infty}\Pr[Y_n\leqslant y]=\Pr[Y\leqslant y],\quad \text{即}\quad \lim_{n\to\infty}F_{Y_n}(y)=F_Y(y),$$

则称随机变量序列 $Y_1,Y_2,\cdots,Y_n,\cdots$ 依分布收敛于 Y, 记 $Y_n\xrightarrow{d}Y$.

下面介绍独立同分布中心极限定理, 又被称为林德贝格-勒维 (Lindeberg-Lévy) 中心极限定理":

定理 8.6 设独立同分布的随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 的期望 $E(X_1) = \mu$ 和方差 $Var(X_1) = \sigma^2$, 则

$$Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

前面介绍标准正态分布的分布函数为 $\Phi(x)$, 则上述中心极限定理等价于

$$\lim_{n\to\infty} \Pr[Y_n \leqslant y] = \Phi(y).$$

随机变量 Y_n 是随机变量 $X_1, X_2, \ldots, X_n$ 的标准化, 其极限服从标准正态分布. 当 n 足够大时近似有 $Y_n \sim \mathcal{N}(0,1)$, 中心极限定理的变形公式为

$$\sum_{i=1}^n X_i \xrightarrow{d} \mathcal{N}(n\mu, n\sigma^2), \qquad \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{d} \mathcal{N}(\mu, \sigma^2/n).$$

大数定律给出了当 $n \to \infty$ 时随机变量平均值 $\frac{1}{n}\sum_{i=1}^n X_i$ 的趋势, 而中心极限定理给出了 $\frac{1}{n}\sum_{i=1}^n X_i$ 的具体分布.

例 8.2 设一电压接收器同时接收到 20 个独立同分布的信号电压 V_k $(k \in [20])$, 且 $V_k \sim U(0,10)$, 求电压和大于 105 的概率.

解 根据题意可知独立同分布的随机变量 $V_1, V_2, \ldots, V_{20}$ 服从均匀分布 $U(0,10)$, 于是有 $E(V_k) = 5$ 和 $\mathrm{Var}(V_k) = 100/12 = 25/3$. 设 $V = \sum_{k=1}^{20} V_k$, 则有

$$E(V) = 100 \qquad \mathrm{Var}(V) = 500/3.$$

根据中心极限定理近似有

$$\frac{V - E(V)}{\sqrt{\mathrm{Var}(V)}} = \frac{V - 100}{\sqrt{500/3}} \sim \mathcal{N}(0,1).$$

根据标准正态分布的分布函数 $\Phi(x)$ 有

$$\Pr(V \geqslant 105) = \Pr\left(\frac{V-100}{\sqrt{500/3}} \geqslant \frac{105-100}{\sqrt{500/3}}\right) = \Pr\left(\frac{V-100}{\sqrt{500/3}} \geqslant 0.387\right) = 1 - \Phi(0.387).$$

查表完成证明.

例 8.3 某产品装箱, 每箱重量是随机的, 假设其期望是 50 公斤, 标准差为 5 公斤. 若最大载重量为 5 吨, 问每车最多可装多少箱能以 0.997 以上的概率保证不超载?

解 假设最多可装 n 箱不超重, 用 X_i 表示第 i 箱重量 $(i \in [n])$, 有 $E(X_i) = 50$ 和 $\mathrm{Var}(X_i) = 25$. 设总重量 $X = \sum_{i=1}^n X_i$, 则有 $E(X) = 50n$ 和 $\mathrm{Var}(X) = 25n$. 由中心极限定理近似有

$$(X - 50n)/\sqrt{25n} \sim \mathcal{N}(0,1).$$

根据标准正态分布的分布函数 $\Phi(x)$ 有

$$\Pr(X \leqslant 5000) = \Pr\left(\frac{X-50n}{\sqrt{25n}} \leqslant \frac{5000-50n}{\sqrt{25n}}\right) = \Phi\left(\frac{5000-50n}{\sqrt{25n}}\right) > 0.977 = \Phi(2).$$

根据分布函数的单调性有

$$\frac{1000-10n}{\sqrt{n}} > 2 \Longrightarrow 1000n^2 - 2000n + 1000^2 > 4n.$$

求解可得 $n > 102.02$ 或 $n < 98.02$, 根据由题意可知 $n = 98$.

下面介绍另一个中心极限定理: 棣莫弗-拉普拉斯 (De Moivre-Laplace) 中心极限定理:

推论 8.1 设随机变量 $X_n \sim B(n,p)$, 则

$$Y_n = \frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0,1).$$

由此中心极限定理可知: 当 n 非常大时随机变量 $X_n \sim B(n,p)$ 满足 $X_n \overset{\text{近似}}{\sim} \mathcal{N}(np, np(1-p))$, 从而有如下近似估计:

$$\Pr[X_n \leqslant y] = \Pr\left[\frac{X_n - np}{\sqrt{np(1-p)}} \leqslant \frac{y-np}{\sqrt{np(1-p)}}\right] \approx \Phi\left(\frac{y-np}{\sqrt{np(1-p)}}\right).$$

针对上式, 可以考虑三种问题: i) 已知 n 和 $\Pr[X_n \leqslant y]$, 求 y; ii) 已知 n 和 y, 求 $\Pr[X_n \leqslant y]$; iii)已知 y 和 $\Pr[X_n \leqslant y]$, 求 n. 下面看三个例子:

例 8.4 车间有 200 台独立工作的车床, 每台工作的概率为 0.6, 工作时每台耗电 1 千瓦, 至少供电多少千瓦才能以 99.9% 的概率保证正常生产.

解 设工作的车床数为 X, 则 $X \sim B(200, 0.6)$. 设至少供电 y 千瓦. 根据棣莫弗-拉普拉斯中心定理近似有 $X \sim \mathcal{N}(120, 48)$, 进一步有

$$\Pr(X \leqslant y) \geqslant 0.999 \quad \Rightarrow \quad \Pr\left(\frac{X-120}{\sqrt{48}} \leqslant \frac{y-120}{\sqrt{48}}\right) \approx \Phi\left(\frac{y-120}{\sqrt{48}}\right) \geqslant 0.999 = \Phi(3.1).$$

所以有 $\frac{y-120}{\sqrt{48}} \geqslant 3.1$, 求解可得 $y \geqslant 141$.

例 8.5 系统由 100 个相互独立的部件组成, 每部件损坏率为 0.1, 至少 85 个部件正常工作系统才能运行, 求系统运行的概率.

解 设 X 是损坏的部件数, 则 $X \sim B(100, 0.1)$, 有 $E(X) = 10$ 和 $\mathrm{Var}(X) = 9$. 根据棣莫弗-拉普拉斯中心定理近似有 $X \sim \mathcal{N}(10, 9)$, 求系统运行的概率为

$$\Pr(X \leqslant 15) = \Pr\left(\frac{X-10}{\sqrt{9}} \leqslant \frac{15-10}{\sqrt{9}}\right) \approx \Phi(5/3).$$

例 8.6 一次电视节目调查中调查 n 人, 其中 k 人观看了电视节目, 因此收看比例 k/n 作为电视节目收视率 p 的估计, 要以 90% 的概率有 $|k/n - p| \leqslant 0.05$ 成立, 需要调查多少对象?

解 用 X_n 表示 n 个调查对象中收看节目的人数, 则有 $X_n \sim B(n, p)$. 根据棣莫弗-拉普拉斯中心定理近似有 $(X_n - np)/\sqrt{np(1-p)} \sim \mathcal{N}(0, 1)$, 进一步有

$$\begin{aligned}\Pr\left[\left|\frac{X_n}{n} - p\right| \leqslant 0.05\right] &= \Pr\left[\frac{|X_n - np|}{n} \leqslant 0.05\right] = \Pr\left[\frac{|X_n - np|}{\sqrt{np(1-p)}} \leqslant \frac{0.05\sqrt{n}}{\sqrt{p(1-p)}}\right] \\ &= \Phi\left(\frac{0.05\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(-\frac{0.05\sqrt{n}}{\sqrt{p(1-p)}}\right)\end{aligned}$$

对于标准正太分布函数有 $\Phi(-\alpha) = 1 - \Phi(\alpha)$ 以及 $p(1-p) \leqslant 1/4$, 于是有

$$\Pr\left[\left|\frac{X_n}{n} - p\right| \leqslant 0.05\right] = 2\Phi\left(\frac{0.05\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 > 2\Phi\left(\sqrt{n}/10\right) - 1 > 0.9.$$

所以 $\Phi\left(\sqrt{n}/10\right) \geqslant 0.95$, 查表解得 $n \geqslant 271$.

对独立不同分布的随机变量序列, 有李雅普诺夫 (Lyapunov) 中心极限定理:

定理 8.7 设独立随机变量 $X_1, X_2, \ldots, X_n, \ldots$ 的期望 $E[X_n] = \mu_n$ 和方差 $\mathrm{Var}(X_n) = \sigma_n^2 > 0$. 记 $B_n^2 = \sum_{k=1}^n \sigma_k^2$, 若存在 $\delta > 0$, 当 $n \to \infty$ 时有

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

成立, 则有

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

中心极限定理小结:

- 独立同分布中心极限定理: 若 $E[X_k] = \mu$ 和 $\mathrm{Var}(X_k) = \sigma^2$, 则 $\sum_{k=1}^n X_k \xrightarrow{d} \mathcal{N}(n\mu, n\sigma^2)$;
- 棣莫弗-拉普拉斯中心极限定理: 若 $X_k \sim B(k, p)$, 则 $X_k \xrightarrow{d} \mathcal{N}(np, np(1-p))$;
- 独立不同分布中心极限定理: 李雅普诺夫定理.

习题

8.1 设随机变量 X 的期望 $E[X]=\mu>0$, 方差为 σ^2, 证明对任意 $\epsilon>0$ 有

$$P(X-\mu\leqslant-\epsilon)\leqslant\frac{\sigma^2}{\sigma^2+\epsilon^2}.$$

8.2 设随机变量X和Y满足$E(X)=-2$, $E(Y)=2$, $\mathrm{Var}(X)=1$, $Var(Y)=4$, $\rho_{XY}=-1/2$. 利用Chebyshev不等式估计$\Pr(|X+Y|\geqslant 6)$的上界.

8.3 独立同分布随机变量 $X_1,X_2,\cdots,X_n$ 满足 $E[X_i]=\mu$ 和 $\mathrm{Var}(X_i)\leqslant v$. 证明对任意 $\epsilon>0$ 有

$$\left[\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|\geqslant\epsilon\right]\leqslant\frac{v}{n\epsilon^2}\ .$$

8.4 阐述什么是chernoff方法。

8.5 随机变量 $X_1,X_2,\cdots,X_n$ 相互独立且满足 $X_i\sim\mathrm{Ber}(p_i)$ $(p_i>0)$. 利用chernoff方法给出下列概率的上界

$$P\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-E[X_i])\geqslant\epsilon\right]\quad\text{和}\quad P\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-E[X_i])\leqslant-\epsilon\right]\ .$$

8.6 若独立同分布随机变量 $X_1,X_2,\cdots,X_n$ 满足 $X_i\in\{a,b\}$ $(b>a)$ 且 $P(X_i=a)=P(X_i=b)=1/2$. 求下列概率的上界

$$P\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\frac{a+b}{2}\right)\geqslant\epsilon\right]\quad\text{和}\quad\Pr\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\frac{a+b}{2}\right)\leqslant-\epsilon\right]\ .$$

8.7 随机变量 $X_1,X_2,\cdots,X_n$ 相互独立且满足 $X_i\sim\mathrm{Ber}(p_i)$ $(p_i>0)$. 证明对任意 $0<\epsilon<1$ 有不等式

$$P\left[\sum_{i=1}^{n}X_i\geqslant(1+\epsilon)\sum_{i=1}^{n}p_i\right]\leqslant e^{-\mu\epsilon^2/3}\ .$$

8.8 随机变量 $X\in[a,b]$ 且期望 $\mu=\mathbb{E}[x]$, 证明对任意 $t>0$ 有

$$\mathbb{E}\left[e^{tx}\right]\leqslant\exp\left(\mu t+t^2(b-a)^2/8\right)$$