7 大数定律及中心极限定理

研究 $\frac{1}{n}\sum_{i=1}^{n} X_i$ 和 $\frac{1}{n}\sum_{i=1}^{n} E[X_i]$ 的关系

7.1 大数定律

给定n个随机变量$X_1, X_2, \cdots, X_n$, 我们考虑n个随机变量的平均值

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

设$Y_1, Y_2, \cdots Y_n$是随机变量序列, Y是另一随机变量
若 $\lim_{n\to\infty} E[|Y_n - Y|] \to 0$, 称$Y_i$依概率收敛于$Y$.

大量随机变量的算术平均值具有稳定性问题.

定义7.1 (依概率收敛). 设$Y_1, Y_2, \cdots, Y_n, \cdots$是随机变量序列, a是一常数, 如果对任意$\epsilon > 0$, 有

$$\lim_{n\to\infty} \Pr\{|Y_n - a| < \epsilon\} = 1 \text{或} \lim_{n\to\infty} \Pr\{|Y_n - a| > \epsilon\} = 0$$

成立, 则称随机变量序列$Y_1, Y_2, \cdots, Y_n, \cdots$依概率收敛于$a$, 记$Y_n \xrightarrow{P} a$.

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$X_n + Y_n \xrightarrow{P} a + b$, $X_n Y_n \xrightarrow{P} ab$.

定理7.1 (大数定律). 设$X_1, X_2, \cdots, X_n, \cdots$是随机变量序列, 如果有

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

则称$\{X_n\}$服从大数定律.

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

定理7.2 (马尔可夫大数定律). 如果随机变量序列$X_1, X_2, \cdots, X_n, \cdots$满足 (大数定律的充分条件)

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

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$X_1, X_2, \cdots, X_n, \cdots$之间相互独立、或同分布, 其证明直接通过Chebyshev不等式有

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

作用在于，将被加从变量之和的方差转变为各个变量各自的方差.

定理7.3 (切比雪夫大数定律). 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$Var(X_n) \le c$, 则$\{X_n\}$服从大数定律.

独立，且方差有界

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

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

协方差为0即可

定理7.4 (辛钦大数定律). 设$X_1, X_2, \cdots, X_n, \cdots$为独立同分布随机变量序列, 且期望存在$E[X_i] = \mu$, 则$\{X_n\}$服从大数定律.

证明超出了本书范围, 一般书中只证明方差存在σ^2.

和之前的大数定律不太一样.

定理7.5 (Bernoulli大数定律). 设X_n为n重*Bernoulli*试验中事件A发生的次数, 记$p = \Pr(A)$, 即$X_n \sim B(n,p)$. 对任意$\epsilon > 0$有$X_n/n \xrightarrow{P} p$, 即

$$\lim_{n\to\infty} \Pr\left[\left|\frac{X_n}{n} - p\right| \ge \epsilon\right] = 0. \quad \Leftrightarrow \quad \frac{1}{n}X_n \xrightarrow{P} p$$

例7.1. 设$X_1, X_2, \cdots, X_n, \cdots$是独立的随机变量序列, 且满足$\Pr\{X_n = n^{1/4}\} = \Pr\{X_n = -n^{1/4}\} = 1/2$. 求证: $\{X_n\}$服从大数定律.

Proof. 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

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

当$n \to 0$时上式趋于零. □

下面补充一个Chebysheve不等式的应用例子: 本题与大数定律无关.

例7.2. 设随机变量X和Y满足$E(X) = -2$, $E(Y) = 2$, $Var(X) = 1$, $Var(Y) = 4$, $\rho_{XY} = -1/2$. 利用*Chebyshev*不等式估计$\Pr(|X+Y| \ge 6)$的上界.

解. 根据期望的线性关系有$E[X+Y] = 0$, 根据相关系数的定义有

$$\rho_{XY} = \frac{E[(X - E(X))(Y - E(Y))]}{\sqrt{Var(X)}\sqrt{Var(Y)}} = -\frac{1}{2}.$$

由此可得$Var(X+Y) = Var(X) + Var(Y) + 2E_{XY}[X - E(X)][Y - E(Y)] = 3$. 根据Chebysheve不等式有$\Pr\{|X+Y| \ge 6\} \le \mathrm{Var}(X+Y)/36 = 1/12$. □

大数定律小结

- 伯努力大数定律: 对二项分布$X_n \sim \mathrm{B}(n,p)$, 有$\Pr[|X_n/n - p| < \epsilon] \to 1$ $(n \to \infty)$;
- 切比雪夫大数定律: 对独立随机变量序列$\{X_i\}$满足$\mathrm{Var}(X_i) \le c$, 有$\Pr[|\sum_{i=1}^{n}(X_i - E[X_i])/n| < \epsilon] \to 1$ $(n \to \infty)$;

- 辛钦大数定律: 对独立同分布随机变量序列$\{X_i\}$, 如果期望存在, 则有$\Pr[|\sum_{i=1}^n (X_i - E[X_i])|/n < \epsilon] \to 1\ (n \to \infty)$;
- 马尔可夫大数定律: 如果随机变量序列$\{X_i\}$满足$Var(\sum_{i=1}^n X_n)/n \to 0$, 则有$\Pr[|\sum_{i=1}^n (X_i - E[X_i])|/n < \epsilon] \to 1\ (n \to \infty)$.

7.2 中心极限定理

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

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

的极限分布是否为是正态分布.

首先介绍一个定义: 依分布收敛.

定义7.2 (依分布收敛). 设$Y_1, Y_2, \cdots, Y_n, \cdots$是随机变量序列, Y是一随机变量, 设分布函数$F_{Y_n}(y) = \Pr(Y_n \le y)$和$F_Y(y) = \Pr(Y \le y)$, 如果

$$\lim_{n\to\infty} \Pr[Y_n \le y] = \Pr[Y \le y], \quad i.e., \quad \lim_{n\to\infty} F_{Y_n}(y) = F_Y(y),$$

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

定理7.6 (林德贝格-勒维中心极限定理, 又称“独立同分布中心极限定理”). 设$X_1, X_2, \cdots, X_n$是独立同分布*(i.i.d)*随机变量, $E(X_1) = \mu$, $Var(X_1) = \sigma^2$, 则

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

Y_n是n个随机变量的标准化, 其极限分布为标准正态分布, 上述中心极限定理可表示为

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

当n足够大, 近似有$Y_n \sim \mathcal{N}(0,1)$, 上述中心极限定理的变形公式:

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

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

例7.3. 假设一个接收器同时接收到 20 个信号电压V_k $(k \in [20])$, 它们独立且均服从$U(0,10)$, 求电压和大于105的概率.

解. 由题意可知$V_1, V_2, \ldots, V_{20}$独立同分布于均匀分布$U(0,10)$, 有$E(V_k) = 5$和$\mathrm{Var}(V_k) = 100/12 = 25/3$. 设$V = \sum_{k=1}^{20} V_k$, 则

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

根据中心极限定理近似有

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

根据$\mathcal{N}(0,1)$的分布函数$\Phi(x)$, 我们得到

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

查表完成证明. □

例7.4. 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, 才能以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($i\in[n]$), 易有$E(X_i)=50$, $\mathrm{Var}(X_i)=25$. 设总重量$X=\sum_{i=1}^{n}X_i$, 则有$E(X)=50n$, $\mathrm{Var}(X)=25n$. 由中心极限定理近似有

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

根据$\mathcal{N}(0,1)$的分布函数$\Phi(x)$, 我们得到

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

有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$n=98$. □

推论7.1 (棣莫弗-拉普拉斯中心极限定理)**.** 设$X_n\sim B(n,p)$, 则

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

伯努利分布可被看成为 $X_i\in\{0,1\}$ 进行 n 次采样的结果, 因此实际上蕴含着一个求和号.

棣莫弗-拉普拉斯中心极限定理表明: 若随机变量$X_n\sim B(n,p)$且n非常大时, 有$X_n\overset{近似}{\sim}\mathcal{N}(np,np(1-p))$, 从而有如下近似估计:

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

针对上式, 可以考虑利用棣莫弗-拉普拉斯中心极限的三种情况:

- 已知n和$\Pr[X_n\leq y]$, 求y;
- 已知n和y, 求$\Pr[X_n\leq y]$;
- 已知y 和$\Pr[X_n\leq y]$, 求n.

下面看三个例子:

例7.5. 车间有200台独立工作的车床, 每台工作概率为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. 设工作车床数为X, 则有$X \sim B(200, 0.6)$. 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

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

所以$\frac{y-120}{\sqrt{48}} \ge 3.1$, 即$y \ge 141$. □

例7.6. 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$X \sim B(100, 0.1)$, 则有$E(X) = 10$, $\mathrm{Var}(X) = 9$. 根据棣莫弗-拉普拉斯中心定理近似有$X \sim \mathcal{N}(10, 9)$, 所以

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

□

例7.7. 在一次电视节目调查中, 假设调查了n个人, 其中k个人观看了电视节目, 因此收看比例k/n作为某电视节目收视率p的估计, 要以90%的把握有$|k/n - p| \le 0.05$成立, 问需要调查多少对象?

解. 设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$X_n \sim \mathcal{N}(0, 1)$, 进一步有

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

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

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

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

定理7.7 (李雅普诺夫中心极限定理, 又称“独立不同分布中心极限定理”). 设$\{X_n\}$为独立随机变量序列, 其期望$E[X_n] = \mu_n$, 方差$Var(X_n) = \sigma_n^2 > 0$. 设$B_n^2 = \sum_{k=1}^n \sigma_k^2$, 若存在$\delta > 0$, 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

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

中心极限定理小结:

- 独立同分布中心极限定理: 若$E[X_k] = \mu$, $\mathrm{Var}(X_k) = \sigma^2$, 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).