大数定律与中心极限定理

1、常用的随机变量序列的收敛性有两种：**依概率收敛**（用于大数定律）**和按分布收敛**（用于中心极限定理）

（一）依概率收敛： 设$\{X_n\}$为一随机变量序列，X为一随机变量，如果对任意的$\varepsilon > 0$，有$P(|X_n - X| \geq \varepsilon) \to 0\ (n \to \infty)$，则称序列$\{X_n\}$依概率收敛于$X$，记作$X_n \xrightarrow{P} X$

依概率收敛的含义是：X_n对X的绝对偏差不小于任一给定量的可能性将随着n增大而越来越小。或者说，绝对偏差$|X_n - X|$小于任意给定量的可能性将随着n增大而越来越接近于1，特别的，当X为退化分布时，即等价于$P(X = c) = 1$，则称序列$\{X_n\}$依概率收敛于c，即$X_n \xrightarrow{P} X$

定理1：设$\{X_n\}, \{Y_n\}$是两个随机变量序列，a, b是两个常数。如果$X_n \xrightarrow{P} a$，$Y_n \xrightarrow{P} b$，则有：

（1）$X_n \pm Y_n \xrightarrow{P} a \pm b$

（2）$X_n \times Y_n \xrightarrow{P} a \times b$

（3）$X_n \div Y_n \xrightarrow{P} a \div b (b \neq 0)$

由定理可以看出，随机变量序列在概率意义上的极限在四则运算下依然成立

（二）按分布收敛（弱收敛）

设随机变量$X, X_1, X_2, \ldots$的分布函数分别为$F(X), F_1(X), F_2(X), \ldots$若对$F(X)$的任意连续点$x$，都有$\lim_{n\to\infty} F_n(x) = F(x)$，则称$\{F_n(X)\}$弱收敛于$F(X)$，记作$F_n(X) \xrightarrow{W} F(X)$，也称$\{X_n\}$按分布收敛于$X$，记作$X_n \xrightarrow{L} X$

若$F(X)$是直线上的连续函数，则弱收敛就是点点收敛

定理 2：$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{L} X$

定理 3：若c为常数，则$X_n \xrightarrow{P} c$的充分必要条件是$X_n \xrightarrow{L} c$

2、伯努利大数定律

设s_n是n重伯努利实验中事件A发生的次数，每次试验中事件A发生的概率为p，则对任意的$\varepsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = 1$$

伯努利大数定律说明：随着n的增大，事件A发生的频率$\frac{S_n}{n}$与其概率p的偏差$|\frac{S_n}{n} - p|$大于预先给定的精度ε的可能性越来越小。

例题：抛一枚硬币出现正面的概率$p = 0.5$. 若把这枚硬币连抛 10 次，则因为n较小，发生偏差的可能性有时会大一些，有时会小一些. 若把这枚硬币连抛 10 万次，由切比雪夫不等式知：正面出现的频率与 0.5 的偏差大于预先给定的精度(若取精度$\varepsilon = 0.01$)的可能性

$$p\left(\left|\frac{S_n}{n} - 0.5\right| > 0.01\right) \le \frac{0.5 * 0.5}{0.01^2 * n} = \frac{10^4}{4n}$$

大偏差发生的可能性小于$1/_{40} = 2.5\%$ 当$n = 10^6$时，大偏差发生的可能性小于$1/_{400} = 0.25\%$ 可见试验次数越多，大偏差发生的可能性越小。

3、常用的几个大数定律

（一）大数定理的一般形式

伯努利大数定律讨论的是一个相互独立同分布的随机变量序列$\{X_n\}$，其共同分布为二点分布，即记 $X_i = \begin{cases} 1，第\ \mathrm{i}\ 次实验中事件\ \mathrm{A}\ 发生 \\ 0，第\ \mathrm{i}\ 次实验中事件\ \mathrm{A}\ 不发生 \end{cases}$ $i = 1,2,\dots,n\dots$ 则$\{X_n\}$是独立的二点分布随机变量序列，先考察该序列的前n个随机变量之和$S_n = \sum_{i=1}^{n} X_i$，则$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$　$p = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}E(X_i)$，那么伯努利大数定律的结论为：对任意的$\varepsilon > 0$，有$\lim_{n\to\infty}(\left|\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} E(x_i)\right| < \varepsilon) = 1$

定义：设有一随机变量序列$\{X_n\}$，假设它具有形式$\lim\limits_{n\to\infty}(\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\frac{1}{n}\sum_{i=1}^{n}E(x_i)\right|<\varepsilon)=1$的性质，则称该随机变量序列$\{X_n\}$服从大数定律

（二）切比雪夫大数定律

设$\{X_n\}$为一列两两不相关的随机变量序列，若每个X_i的方差存在，且有共同的上界，即$Var(X_i)\le c, i=1,2,\ldots$,则$\{X_n\}$服从大数定律，即对任意的$\varepsilon>0$，$\lim\limits_{n\to\infty}(\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\frac{1}{n}\sum_{i=1}^{n}E(x_i)\right|<\varepsilon)=1$

（三）马尔可夫大数定律

定理：对随机变量序列$\{X_n\}$，若$\frac{1}{n^2}Var(\sum_{i=1}^{n}X_i)\to 0$ 成立，则$\{X_n\}$服从大数定律，即对任意的$\varepsilon>0$ $\lim\limits_{n\to\infty}(\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\frac{1}{n}\sum_{i=1}^{n}E(x_i)\right|<\varepsilon)=1$式成立

（四）辛钦大数定律（弱大数定律）

设$\{X_n\}$为一独立同分布的随机变量序列，若X_i的数学期望存在，则$\{X_n\}$服从大数定律，即对任意的$\varepsilon>0$，$\lim\limits_{n\to\infty}(\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\frac{1}{n}\sum_{i=1}^{n}E(x_i)\right|<\varepsilon)=1$式成立

注：（1）伯努利大数定律是切比雪夫大数定律的特例

（2）切比雪夫大数定律是马尔科夫大数定律的特例

（3）伯努利大数定律是辛钦大数定律的特例

4、中心极限定理

设$\{X_n\}$为一独立同分布的随机变量序列，记其和为$Y_n=\sum_{i=1}^{n}X_i$需要研究讨论独立随机变量和的极限分布

（1）独立同分布下的中心极限定理

定理（林德伯格−莱维中心极限定理）：设$\{X_n\}$为一独立同分布的随机变量序列，数学期望为μ，方差为$\sigma^2>0$存在，若记

$$Y_n{}^* = \frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \sim N(0,1)$$

则对任意实数y有

$$\lim_{n\to\infty} P(Y_n{}^* \le y) = \Phi(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{t^2}{2}}\,dt$$

例题：每袋味精的净重为随机变量，平均重量为 100 克，标准差为 10 克，一箱内装 200 袋味精，求一箱味精的净重大于 20500 克的概率？

解：设箱中第i袋味精的净重为X_i，则X_i独立同分布，且$E(X_i)=100$，有中心极限定理得，所求概率为$P(\sum_{i=1}^{200}X_i>20500)\approx 1-\Phi\left(\frac{20500-200\times100}{\sqrt{200\times100}}\right)=1-\Phi(3.54)=0.0002$，故一箱味精的净重大于 20500 克的概率为 0.0002

定理（拉普拉斯中心极限定理）设n重伯努利试验中，事件 A 在每次试验中出现的概率p（$0<p<1$），记s_n为n次试验中事件 A 出现的次数，且记$Y_n{}^*=\frac{s_n-np}{\sqrt{npq}}$（减去均值，除以标准差）则对任意实数$y$，有$\lim_{n\to\infty}P(Y_n{}^*\le y)=\Phi(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}e^{-\frac{t^2}{2}}dt$，注：二项分布是离散分布，而正态分布是连续分布，所以用正态分布作为二项分布的近似时，可做如下修正：

$$P(k_1\le\mu_n\le k_2)=P(k_1-0.5\le\mu_n\le k_2+0.5)$$
$$\approx\Phi\left(\frac{k_2+0.5-np}{\sqrt{npq}}\right)-\Phi\left(\frac{k_1-0.5-np}{\sqrt{npq}}\right)$$

（加减符号要注意）

（2）中心极限定理的应用有三大类

若记$\beta=\Phi(y)$，有中心极限定理可知$P(Y_n{}^*\le y)\approx\Phi(y)=\beta$，可以来解决三类计算问题：

（1）已知n和y，求概率β

100 个独立工作（工作的概率为 0.9）的部件组成一个系统，求系统中至少有 85 个部件工作的概率.

解：用$X_i=1$表示第i个部件正常工作，反之记为$X_i=0$，又记$Y=X_1+X_2+\cdots+X_{100}$，则$E(Y)=90, Var(Y)=9$，由此得：$P(Y\geq 85)\approx 1-\Phi\left(\frac{85-0.5-90}{\sqrt{9}}\right)=0.966$（85 是小于等于$y$的，所以对$y$做修正的时候是-0.5）

（2） 已知n和概率β，求y

有 200 台独立工作（工作的概率为 0.7）的机床，每台机床工作时需$15kw$电力，问共需多少电力，才可有 95%的可能性保证正常生产？

用$X_i=1$表示第i个机床正常工作，反之记为$X_i=0$，又记$Y=X_1+X_2+\cdots+X_{200}$，则$E(Y)=140, Var(Y)=42$，设供电量为y，则从

$P(15Y\leq y)\approx\Phi(\frac{\frac{y}{15}+0.5-140}{\sqrt{42}})\geq 0.95$中解得：$y\geq 2252$

（3） 已知y和概率β，求n

用调查对象中的收看比例k/n作为某电视节目的收视率p的估计，要有 90%的把握，使$\frac{k}{n}$与p的差异不大于 0.05，问至少要调查多少对象？

解：用Y_n表示n个调查对象中收看此节目的人数，则Y_n服从$b(n,p)$分布，k为Y_n的实际取值，根据题意

$$P\left(\left|\frac{Y_n}{n}-p\right|<0.05\right)\approx 2\Phi\left(0.05\sqrt{\frac{n}{p(1-p)}}\right)-1\geq 0.90$$

$$P\left(\left|\frac{Y_n}{n}-p\right|<0.05\right)=P(n(-0.05+p)<Y_n<n(0.05+p))$$

从中解得$0.05\sqrt{n/p(1-p)}\geq 1.645$，又由于$p(1-p)\leq 0.25$，可解得$n\geq 270.6$，所以$n=271$

（3）独立不同分布下的中心极限定理

林德伯格条件：$\lim\limits_{n\to\infty}\frac{1}{\tau^2 {B_n}^2}\sum_{i=1}^{n}\int_{|x-\mu_i|>\tau B_n}(x-\mu_i)^2 p_i(x)\,dx=0$

定理（林德伯格中心极限定理）设独立随机变量序列$\{X_n\}$满足林德伯格条件，则对任意的x有

$$\lim_{n\to\infty}P\left\{\frac{1}{B_n}\sum_{i=1}^{n}(X_i-\mu_i)\leq y\right\}=\Phi(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}\,dt$$

李雅普诺夫条件：$\lim\limits_{n\to\infty}\frac{1}{{B_n}^{2+\delta}}\sum_{i=1}^{n}E\left(|X_i-\mu_i|^{2+\delta}\right)=0$

定理（林德伯格中心极限定理）设独立随机变量序列$\{X_n\}$满足林德伯格条件，则对任意的x有

$$\lim_{n\to\infty} P\left\{\frac{1}{B_n}\sum_{i=1}^{n}(X_i-\mu_i)\le y\right\}=\Phi(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}\,dt$$