概率论与数理统计
第五章 概率极限定理

金玲飞

复旦大学软件学院
Email: lfjin@fudan.edu.cn

2019.12.3

5.1 大数定律

- 大数定律：大量重复试验所呈现的统计规律。“频率的稳定性”，"随机变量算数均值的稳定性"。
- 大数定律最早可追溯到公元1500年左右。
- 1713，Bernoulli正式提出并证明了最初的大数定律。
- 直到1930年，现在概率论奠基人Kolgomorov才真正证明了最后的强大数定律。

几种收敛性

定义 (5.1.1 依概率收敛)

设$\{X_n : n = 1, 2, \cdots\}$为随机变量序列，$X$为随机变量。若对任意$\epsilon > 0$，有

$$\lim_{n\to\infty} P(|X_n| > \epsilon) = 0$$

则称X_n依概率收敛于*0*,记为$X_n \xrightarrow{P} 0$。
若$X_n - X \xrightarrow{P} 0$，则称$X_n$依概率收敛于$X$，记为$X_n \xrightarrow{P} X$。

几种收敛性

定义 (5.1.1 依概率收敛)

设$\{X_n : n = 1, 2, \cdots\}$为随机变量序列，$X$为随机变量。若对任意$\epsilon > 0$，有

$$\lim_{n\to\infty} P(|X_n| > \epsilon) = 0$$

则称X_n依概率收敛于*0*,记为$X_n \xrightarrow{P} 0$。
若$X_n - X \xrightarrow{P} 0$，则称$X_n$依概率收敛于$X$，记为$X_n \xrightarrow{P} X$。

若$X_n \xrightarrow{P} X$，$g(x)$为连续函数，则$g(X_n) \xrightarrow{P} g(X)$。

定义 (5.1.2 按分布收敛)

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$F(x)$的任一连续点$x$，都有

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

则称$\{F_n(x)\}$弱收敛于$F(x)$，记为$F_n(x) \xrightarrow{W} F(x)$。也称 $\{X_n\}$按分布收敛于X，记为$X_n \xrightarrow{L} X$。

定义 (5.1.3 几乎必然收敛 Almost sure convergence)

设X，$X_n(n = 1, 2, \cdots)$均为$(\Omega, \mathscr{F}, P)$上的随机变量。若

$$P(\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1$$

则称X_n几乎必然收敛于X，记为$X_n \xrightarrow{a.s.} X$，
或$X_n \to X\ a.s.$。

设$X, X_n (n = 1, 2, \cdots)$均为$(\Omega, \mathscr{F}, P)$上的随机变量。则

$$X_n \stackrel{a.s.}{\to} X \Rightarrow X_n \stackrel{P}{\to} X \Rightarrow X_n \stackrel{L}{\to} X$$

定义 (5.1.4)

- 若$X_1, \ldots, X_n$相互独立，则称$\{X_n : n = 1, 2, \cdots\}$为独立随机变量序列
- 若$X_1, \ldots, X_n$相互独立且具有相同的分布函数，则称$\{X_n : n = 1, 2, \cdots\}$为独立同分布随机变量序列

切比雪夫大数定律

定理 (5.1.1 Chebyshev law of large numbers)

设$\{X_n : n = 1, 2, \cdots\}$为独立随机变量序列。若存在常数$C$，使得$D(X_n) \leq C, n = 1, 2, \cdots$，则对任意$\epsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| > \epsilon\right) = 0 \tag{1}$$

即

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - E(X_k)) \xrightarrow{P} 0$$

- 若随机变量序列$\{X_n : n = 1, 2, \cdots\}$满足(1)式，则称它服从大数定律。

- 若随机变量序列$\{X_n : n = 1, 2, \cdots\}$满足(1)式，则称它服从大数定律。
- 在什么条件下，$\{X_n : n = 1, 2, \cdots\}$服从大数定律？

推论 (5.1.2)

如果$\{X_n : n = 1, 2, \cdots\}$是独立同分布的，期望$E(X_n) = \mu$，方差$D(X_n) = \sigma^2$，则有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| > \epsilon\right) = 0$$

即

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu$$

推论 (5.1.2)

如果$\{X_n : n = 1, 2, \cdots\}$是独立同分布的，期望$E(X_n) = \mu$，方差$D(X_n) = \sigma^2$，则有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| > \epsilon\right) = 0$$

即

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu$$

- 算术均值的稳定性：当试验次数$n \to \infty$，平均值$\frac{1}{n}\sum_{k=1}^{n} X_k$依概率收敛于期望。

例子 (5.1.1)

设$\{X_n : n = 1, 2, \cdots\}$是相互独立的随机变量序列。若

$$P(X_n = -3^n) = P(X_n = 3^n) = 3^{-(2n+2)}$$
$$P(X_n = 0) = 1 - \frac{2}{3^{2n+2}}$$

问此随机变量序列是否服从大数定律?

例子 (5.1.2)

设$\{X_n : n = 1, 2, \cdots\}$是独立同分布的随机变量序列，$E(X_n^4) < \infty$。若令$E(X_n) = \mu, D(X_n) = \sigma^2$。考察

$$Y_n = (X_n - \mu)^2, n = 1, 2, \ldots,$$

则$\{Y_n : n = 1, 2, \cdots\}$ 是否服从大数定律。

伯努利大数定律

最早提出的大数定律是1713年伯努利大数定律。

定理 (5.1.3 Bernoulli law of large numbers)

设n_A为n重伯努利试验中事件A发生的次数，又A在每次试验中发生的概率为$p(0 < p < 1)$，则对任意$\epsilon > 0$，有

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

伯努利大数定律

最早提出的大数定律是1713年伯努利大数定律。

定理 (5.1.3 Bernoulli law of large numbers)

设n_A为n重伯努利试验中事件A发生的次数，又A在每次试验中发生的概率为$p(0 < p < 1)$，则对任意$\epsilon > 0$，有

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

- 抛一枚硬币10^4次，$n = 10^4$。设$\epsilon = 0.01$,

$$P\left(\left|\frac{n_A}{n} - 0.5\right| > 0.01\right) \leq \frac{10^4}{4n}$$

辛钦大数定律

可否去掉方差存在的条件？

定理 (5.1.4 辛钦大数定律)

设$\{X_n : n = 1, 2, \cdots\}$是独立同分布随机变量序列，$E(X_n) = \mu$。则对任意的$\epsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| > \epsilon\right) = 0$$

即

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

辛钦大数定律

可否去掉方差存在的条件？

定理 (5.1.4 辛钦大数定律)

设$\{X_n : n = 1, 2, \cdots\}$是独立同分布随机变量序列，$E(X_n) = \mu$。则对任意的$\epsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| > \epsilon\right) = 0$$

即

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

辛钦大数定律提供了求期望$E(X)$的近似值的方法。

Strong law of large numbers

定理 (5.1.5 科尔莫戈罗夫强大数定律)

设$\{X_n : n = 1, 2, \cdots\}$是独立同分布的随机变量序列，$E(X_n) = \mu$，则

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s} \mu$$

定理 (5.1.6 博雷尔强大数定律)

设n_A为n重伯努利试验中事件A发生的次数，又A在每次试验中发生的概率为$p(0 < p < 1)$，则

$$\frac{n_A}{n} \xrightarrow{a.s.} p$$

例子 (5.1.3 蒙特卡罗方法计算积分)

设$g(x)$在区间$[0, 1]$上绝对可积，试计算积分$\int_0^1 g(x)dx$。

例子 (5.1.3 蒙特卡罗方法计算积分)

设$g(x)$在区间$[0,1]$上绝对可积，试计算积分$\int_0^1 g(x)dx$。

- 设X 是服从[0,1]上的均匀分布。
- $Y=g(X)$的数学期望$E(Y)=\int_0^1 g(x)dx=J$。
- 根据辛钦大数定律，可以用$g(X)$的观察值的平均值去估计它的期望。

例子 (5.1.3 蒙特卡罗方法计算积分)

设$g(x)$在区间$[0,1]$上绝对可积，试计算积分$\int_0^1 g(x)dx$。

- 设X 是服从[0,1]上的均匀分布。
- $Y = g(X)$的数学期望$E(Y) = \int_0^1 g(x)dx = J$。
- 根据辛钦大数定律，可以用$g(X)$的观察值的平均值去估计它的期望。

  先用计算机产生n个均匀分布的随机数x_i，再计算$g(x_i)$。

5.2 中心极限定理

中心极限定理

设$\{X_n : n = 1, 2, \cdots\}$是独立随机变量序列，当$n$很大时，研究

$$Y_n = X_1 + X_2 + \ldots + X_n$$

的分布？

中心极限定理

设$\{X_n : n = 1, 2, \cdots\}$是独立随机变量序列，当$n$很大时，研究

$$Y_n = X_1 + X_2 + \ldots + X_n$$

的分布？

$$Y_n^* = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}}$$

- 在某种条件下，使得随机变量序列的和的极限分布是正态分布的结果，统称为中心极限定理。

定理 (5.2.1 莱维-林德伯格中心极限定理)

设$\{X_n : n = 1, 2, \cdots\}$是独立同分布随机变量序列，$E(X_n) = \mu$，$D(X_n) = \sigma^2 > 0$。则随机变量

$$\frac{1}{\sqrt{n}\sigma}\left(\sum_{k=1}^{n} X_k - n\mu\right)$$

的分布函数$F_n(X)$收敛到标准正态分布函数$\Phi(x)$，即对任意实数x，成立

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}\sigma}\left(\sum_{k=1}^{n} X_k - n\mu\right) \le x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

定义 (5.2.1)

当随机变量序列的极限分布是正态分布时，我们也称该随机变量序列是渐近正态分布的。

定义 (5.2.1)

当随机变量序列的极限分布是正态分布时，我们也称该随机变量序列是渐近正态分布的。

例子 (5.2.1 误差估计)

计算机在进行数字计算时，遵从四舍五入的原则。为简单计，现对小事实数点后面第一位进行舍入运算，则误差X可以认为服从均匀分布$U[-0.5, 0.5]$。若在一项计算中进行了 *100* 次数字计算，求平均误差落在区间$\left[-\frac{\sqrt{3}}{20}, \frac{\sqrt{3}}{20}\right]$上的概率。

例子 (5.2.2 正态随机数的产生)

在随机模拟（蒙特卡洛方法）中经常需要产生正态分布$N(\mu, \sigma^2)$的随机数，一般的统计软件都没有产生正态随机数的功能。那如何产生的呢？

例子 (5.2.2 正态随机数的产生)

在随机模拟（蒙特卡洛方法）中经常需要产生正态分布$N(\mu, \sigma^2)$的随机数，一般的统计软件都没有产生正态随机数的功能。那如何产生的呢？

其中一种方法：用中心极限定理通过$(0, 1)$上的均匀分布的随机数来产生正态分布$N(\mu, \sigma^2)$ 的随机数。

拉普拉斯极限定理

二项分布的正态近似

定理 (5.2.2 棣莫弗-拉普拉斯极限定理)

设n_A为n重伯努利试验中事件A出现的次数，又A在每次试验中发生的概率为$p(0 < p < 1)$。则对任意实数x，有

$$\lim_{n\to\infty} P\left(\frac{n_A - np}{\sqrt{np(1-p)}} \leq x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

推论 (5.2.3)

上面定理表明，二项分布可以用正态分布来近似。即对$a \le b$，有

$$
\begin{aligned}
& P(a \le n_A \le b) \\
& = P\left(\frac{a - np}{\sqrt{np(1-p)}} \le \frac{n_A - np}{\sqrt{np(1-p)}} \le \frac{b - np}{\sqrt{np(1-p)}}\right) \\
& \approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right)
\end{aligned}
$$

我们做如下修正来提高精度

- 由于n_A只取整数，故对整数a, b和任意$0 \le \epsilon_1, \epsilon_2 < 1$，有

$$P(a \le n_A \le b) = P(a - \epsilon_1 \le n_A \le b + \epsilon_2) \approx \Phi\left(\frac{b + \epsilon_2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \epsilon_1 - np}{\sqrt{np(1-p)}}\right)$$

我们做如下修正来提高精度

- 由于n_A只取整数，故对整数a, b和任意$0 \le \epsilon_1, \epsilon_2 < 1$，有

$$P(a \le n_A \le b) = P(a - \epsilon_1 \le n_A \le b + \epsilon_2)$$
$$\approx \Phi\left(\frac{b + \epsilon_2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \epsilon_1 - np}{\sqrt{np(1-p)}}\right)$$

- 进一步分析表明，选择$\epsilon_1 = \epsilon_2 = 0.5$得到的修正公式效果较好，

$$P(a \le n_A \le b) \approx \Phi\left(\frac{b + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

- 对二项分布的计算，用修正的正态分布还可以近似求得

$$P(n_A = b) = P(b - 0.5 < n_A < b + 0.5)$$

比如$n_A \sim b(25, 0.4)$，则$P(n_A = 10) = 0.1612$，而用正态近似估计可得$P(n_A = 10) = 0.1629$。

利用中心极限定理可以解决三种计算问题：

$$P(Y_n^* \leq y) \approx \Phi(y) = \beta$$

1. 已知n, y，求β；
2. 已知n, β，求y；
3. 已知β, y，求n；

例子 (5.2.3)

一复杂系统由*100*个相互独立工作的部件组成，每个部件正常工作的概率为*0.9.*已知整个系统中至少要有*85*个部件正常工作，系统才能正常工作。求该系统正常工作的概率。

例子 (5.2.3)

一复杂系统由*100*个相互独立工作的部件组成，每个部件正常工作的概率为*0.9*.已知整个系统中至少要有*85*个部件正常工作，系统才能正常工作。求该系统正常工作的概率。

例子 (5.2.4)

某单位有*260*部电话，每部电话约有*4%*的时间要使用外线通话。设每部电话是否使用外线是相互独立的，问该单位总机至少需要安装多少条外线，才能以*95%*以上的概率保证每部电话需要使用外线时可以打通？

例子 (5.2.5)

某厂生产的螺丝钉的不合格品率为*0.01*.问一盒中应至少装多少只螺丝钉才能保证其中含有*100*只合格品的概率不小于*0.95*?

例子 (5.2.6)

电视台做某节目A收视率的调查，在每天节目A播出时随机地向当地居民打电话，问是否在看电视，如在，再问是否在看节目A。设回答在看电视的居民数位n，为保证以95%的概率使调查误差不超过10%，问n至少应取多大？

- 当n很大时，$\frac{1}{\sqrt{n}\sigma}\left(\sum_{k=1}^{n} X_k - n\mu\right)$ 的分布函数$F_n(x)$收敛于标准正态分布$\Phi(x)$。那么$F_n(x)$收敛于$\Phi(x)$ 的速度怎样？

- 当n很大时，$\frac{1}{\sqrt{n}\sigma}\left(\sum_{k=1}^{n} X_k - n\mu\right)$ 的分布函数$F_n(x)$收敛于标准正态分布$\Phi(x)$。那么$F_n(x)$收敛于$\Phi(x)$ 的速度怎样？

定理 (Berry-Esseen Theorem)

设$\{X_n : n = 1, 2, \cdots\}$是独立同分布随机变量序列，$E(X_n) = \mu$，$D(X_n) = \sigma^2 > 0$，$E(|X_1 - \mu|^3) = \rho < \infty$。则对任意的实数$x$和任意$n \geq 1$，

$$\left|P\left(\frac{1}{\sqrt{n}\sigma}\left(\sum_{k=1}^{n} X_k - n\mu\right) \leq x\right) - \Phi(x)\right| \leq \frac{3\rho}{\sigma^3\sqrt{n}}$$

- 独立同分布随机变量序列的中心极限定理

  （1）林德贝格—勒维中心极限定理

  （2）棣莫弗—拉普拉斯中心极限定理

- 独立不同分布随机变量序列的中心极限定理

  （1）林德贝格中心极限定理

  （2）李雅普诺夫中心极限定理

中心极限定理

- 若$\{X_n\}$服从中心极限定理，则$\{X_n \pm a_n\}$也服从中心极限定理，其中$\{a_n\}$为常数列。

中心极限定理

- 若$\{X_n\}$服从中心极限定理，则$\{X_n \pm a_n\}$也服从中心极限定理，其中$\{a_n\}$为常数列。

例子

用概率的方法证明

$$\lim_{n\to\infty}(1+n+\frac{n^2}{2!}+\cdots+\frac{n^n}{n!})e^{-n}=\frac{1}{2}$$

中心极限定理

- 若$\{X_n\}$服从中心极限定理，则$\{X_n \pm a_n\}$也服从中心极限定理，其中$\{a_n\}$为常数列。

例子

用概率的方法证明

$$\lim_{n\to\infty}(1+n+\frac{n^2}{2!}+\cdots+\frac{n^n}{n!})e^{-n}=\frac{1}{2}$$

设$\{X_n\}$是独立同分布的的随机变量序列，服从$P(1)$。$E(X_n)=D(X_n)=1$。根据可加性，$Y_n=\sum_{i=1}^{n}X_i\sim P(n)$。

中心极限定理

- 若$\{X_n\}$服从中心极限定理，则$\{X_n \pm a_n\}$也服从中心极限定理，其中$\{a_n\}$为常数列。

例子

用概率的方法证明

$$\lim_{n\to\infty}(1+n+\frac{n^2}{2!}+\cdots+\frac{n^n}{n!})e^{-n}=\frac{1}{2}$$

设$\{X_n\}$是独立同分布的的随机变量序列，服从$P(1)$。$E(X_n)=D(X_n)=1$。根据可加性，$Y_n=\sum_{i=1}^{n}X_i\sim P(n)$。根据林德伯格-莱维中心极限定理

$$P(\frac{Y_n-n}{\sqrt{n}}\leq 0)=P(Y_n\leq n)$$