第 5 章　大数定律与中心极限定律

安徽财经大学

统计与应用数学学院

目录

在前面几章中，我们介绍了一维和多维随机变量的统计特性，在实践和理论上还需要研究随机变量序列的统计特性.

关于随机变量序列的研究最基本的就是如下的两类问题:

- 一类是研究大量随机现象的平均结果在什么条件下具有稳定性的问题. 关于这类问题的一系列定理统称为**大数定律**.
- 另一类是研究大量随机因素的总效应在什么条件下近似服从正态分布的问题. 关于此类问题的一系列定理统称为**中心极限定律**.

本章将首先介绍在解决这两类问题时的一个有用工具——**特征函数**，然后再分别研究这两类问题.

定义 (5.1.1 特征函数)

设 X 为实随机变量，称

$$\varphi(t) = E(e^{itX}), \tag{1}$$

为 X 的特征函数 (Characteristic function)，这里 t 是任意实数.

由于 $|e^{itX}| \leqslant 1$，因此对任意 X，对一切 $t \in (-\infty, +\infty)$，$E(e^{itX})$ 总存在，即 (1) 式都有意义，也就是说任意随机变量的特征函数总存在.

特别地，若 X 为离散型，$p_k = P(X = x_k)$，$k = 1, 2, \cdots$，则其特征函数为

$$\varphi(t) = \sum_{k=1}^{+\infty} p_k, \quad -\infty < t < +\infty,$$

若 X 是连续型，其密度 $f(x)$，则

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)\mathrm{d}x,$$

它就是函数 $f(x)$ 的傅里叶变换.

下面不加证明的给出特征函数的一些重要性质，其中 $\varphi_X(t)$ 表示 X 的特征函数，其他类似.

性质 (5.1.1)

$|\varphi(t)| \leqslant \varphi(0) = 1.$

性质 (5.1.2)

$\varphi(-t) = \overline{\varphi(t)}.$

性质 (5.1.3)

若 $Y = aX + b$，其中 a, b 是常数，则有

$$\varphi_Y(t) = e^{ibt}\varphi_X(at).$$

性质 (5.1.4)

设 X, Y 独立，则有

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t).$$

即独立随机变量和的特征函数为特征函数的乘积.

性质 (5.1.5)

若 $E(X^l)$ 存在，则 X 的特征函数 $\varphi(t)$ 存在 l 阶导数，且对于任意的 $1 \leqslant k \leqslant l$，有

$$\varphi^{(k)}(0) = i^k E(X^k),$$

也就是说在上述条件下，X 的 k 阶矩可以通过特征函数在 $t=0$ 处的 k 阶导数表示出来.

$$E(X^k) = i^{-k} \cdot \varphi^{(k)}(0).$$

例 (5.1.1 常见分布的特征函数)

(1) 单点分布: $P(X=c)=1$ 的特征函数 $\varphi(t)=e^{itc}$.

(2) $0-1$ 分布: $P(X=1)=p$, $P(X=0)=q$, $p+q=1$ 的特征函数为

$$\varphi(t)=pe^{it}+q.$$

(3) 泊松分布: $P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,2,\cdots$ 的特征函数为

$$\varphi(t)=\sum_{k=0}^{+\infty}e^{it}\frac{\lambda^k}{k!}e^{-\lambda}=e^{\lambda(e^{it}-1)}.$$

(4) 均匀分布 $U(a,b)$ 的特征函数为

$$\varphi(t)=\int_{-\infty}^{+\infty}e^{itx}f(x)\mathrm{d}x=\int_a^b e^{itx}\frac{1}{b-a}\mathrm{d}x=\frac{e^{ibt}-e^{iat}}{it(b-a)}.$$

例 (5.1.1 常见分布的特征函数)

(5) 标准正态分布 $N(0,1)$ 的特征函数为

$$\varphi(t) = e^{-\frac{1}{2}t^2}.$$

(6) 指数分布 $Exp(\lambda)$ 的特征函数为

$$\varphi(t) = \left(1 - \frac{it}{\lambda}\right)^{-1}.$$

(7) 二项分布 $b(n,p)$ 的特征函数为

$$\varphi(t) = (pe^{it} + q)^n.$$

(8) 正态分布 $N(\mu,\sigma^2)$ 的特征函数为

$$\varphi(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}.$$

例 (5.1.1)

(6) 指数分布 $Exp(\lambda)$ 的特征函数为

$$\varphi(t) = \left(1 - \frac{it}{\lambda}\right)^{-1}.$$

证明.

(6)

$$\begin{aligned}
\varphi(t) &= \int_{-\infty}^{+\infty} e^{itx} f(x)\mathrm{d}x = \int_{0}^{+\infty} e^{itx}\lambda e^{-\lambda}\mathrm{d}x \\
&= \lambda\left\{\int_0^{+\infty}\cos(tx)e^{-\lambda}\mathrm{d}x + i\int_0^{+\infty}\sin(tx)e^{-\lambda}\mathrm{d}x\right\} \\
&= \lambda\left\{\frac{\lambda}{\lambda^2+t^2} + i\frac{t}{\lambda^2+t^2}\right\} = \left(1-\frac{it}{\lambda}\right)^{-1}.
\end{aligned}$$

□

例 (5.1.1)

(7) 二项分布 $b(n,p)$ 的特征函数为

$$\varphi(t) = (pe^{it} + q)^n.$$

证明.

(7) 设 $Y \sim b(n,p)$, 则由前一章知: $Y = \sum_{i=1}^{n} X_i$, , 其中 X_i 相互独立, 且都服从 $0-1$ 分布 $b(1,p)$. 由特征函数性质 5.1.4 知

$$\varphi_Y(t) = \prod_{i=1}^{n} \varphi_{X_i}(t).$$

由本例 (2) 可知: $\varphi_{X_i}(t) = pe^{it} + q$, 于是有

$$\varphi_Y(t) = (pe^{it} + q)^n.$$

$\square$

例 (5.1.1)

(8) 正态分布 $N(\mu, \sigma^2)$ 的特征函数为

$$\varphi(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}.$$

证明.

(8) 设 $Y \sim N(0,1)$. 则由本例 (5) 式知

$$\varphi(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}.$$

注意到 $Y = aX + \mu$, 利用性质 5.1.3 即可得证. □

例 (5.1.2)

试用特征函数的方法求正态分布 $N(\mu,\sigma^2)$ 的期望与方差.

证明.

因为正态分布 $N(\mu,\sigma^2)$ 的特征函数及其一、二阶导数为

$$\varphi(t)=e^{i\mu t-\frac{1}{2}\sigma^2t^2};$$

$$\varphi'(t)=(i\mu-\sigma^2t)e^{i\mu t-\frac{1}{2}\sigma^2t^2};\ \varphi'(0)=i\mu;$$

$$\varphi''(t)=(i\mu-\sigma^2t)e^{i\mu t-\frac{1}{2}\sigma^2t^2}-\sigma^2e^{i\mu t-\frac{1}{2}\sigma^2t^2};\ \varphi''(0)=(i\mu)^2-\sigma^2.$$

于是由 (5.1.8) 知

$$E(X)=\frac{\varphi'(0)}{i}=\mu,\ E(X^2)=\frac{(i\mu)^2-\sigma^2}{i^2}=\mu^2+\sigma^2.$$

从而有 $D(X)=E(X^2)-E^2(X)=\sigma^2$. □

定理 (5.1.1 唯一性定理)

随机变量的分布由其特征函数唯一决定.

证明需要较多的数学知识, 在此从略.

特征函数的定义及唯一性定理表明: 分布函数与特征函数是一一对应的.

在概率论中, 独立随机变量和的问题占有 "中心" 地位, 利用卷积公式去处理独立随机变量和的问题相当复杂, 而引入特征函数可以很方便地应用函数乘法得到独立随机变量和的特征函数.

例 (5.1.3)

试用特征函数的方法证明二项分布的可加性: 设 $X \sim b(n_1, p)$, $Y \sim b(n_2, p)$, 且 X, Y 独立, 则

$$X + Y \sim b(n_1 + n_2, p).$$

证明.

因为 X, Y 的特征函数分别为

$$\varphi_X(t) = (pe^{it} + q)^{n_1}, \ \varphi_Y(t) = (pe^{it} + q)^{n_2}.$$

又 X, Y 独立, 所以有

$$\varphi_{X+Y}(t) = (pe^{it} + q)^{n_1} \cdot (pe^{it} + q)^{n_2} = (pe^{it} + q)^{n_1+n_2},$$

这正是 $b(n_1 + n_2, p)$ 的特征函数, 故由唯一性定理知

$$X + Y \sim b(n_1 + n_2, p).$$

1. 特征函数
2. 大数定律
   - 契比雪夫不等式
   - 大数定律
3. 随机变量序列的两种收敛性
4. 中心极限定理

概率论是研究随机现象的统计规律的科学，而这种统计规律性往往需要在相同的条件下进行大量的重复试验才能体现出来.
在第一章中我们已经指出，人们经过长期实践认识到，虽然个别随机事件在某次试验中可能发生也可能不发生，但是在大量重复试验中却呈现明显的规律性，即随着试验次数的增大，一个随机事件发生的频率在某一固定值附近摆动. 这就是所谓的频率具有稳定性，本章通过大数定律从理论上给予说明.
在引入大数定律之前，我们先证一个重要的不等式——契比雪夫 (Chebyshev) 不等式.

定理 (5.2.1 契比雪夫不等式)

设随机变量 X 存在有限方差 $D(X)$, 则有对任意 $\varepsilon > 0$, 有

$$P\{|X - E(X)| \geqslant \varepsilon\} \leqslant \frac{D(X)}{\varepsilon^2}. \tag{2}$$

定理 (5.2.1 契比雪夫不等式)

设随机变量 X 存在有限方差 $D(X)$, 则有对任意 $\varepsilon > 0$, 有

$$P\{|X - E(X)| \geqslant \varepsilon\} \leqslant \frac{D(X)}{\varepsilon^2}. \tag{2}$$

证明.

以连续性随机变量为例证明该不等式. 设 X 的概率密度为 $f(x)$, 则有

$$\begin{aligned} P\{|X - E(X)| \geqslant \varepsilon\} &= \int\limits_{|x-E(X)|\geqslant\varepsilon} f(x)\mathrm{d}x \\ &\leqslant \int\limits_{|x-E(X)|\geqslant\varepsilon} \frac{|x - E(X)|^2}{\varepsilon^2} f(x)\mathrm{d}x \\ &\leqslant \frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty} [x - E(X)]^2 f(x)\mathrm{d}x = \frac{D(X)}{\varepsilon^2}. \end{aligned}$$

契比雪夫不等式的含义

定理 (5.2.1 契比雪夫不等式)

设随机变量 X 存在有限方差 $D(X)$, 则有对任意 $\varepsilon > 0$, 有

$$P\{|X - E(X)| \geqslant \varepsilon\} \leqslant \frac{D(X)}{\varepsilon^2}. \tag{3}$$

契比雪夫不等式也可表示成

$$P\{|X - E(X)| < \varepsilon\} \geqslant 1 - \frac{D(X)}{\varepsilon^2}. \tag{4}$$

事件 “$|X - E(X)| \geqslant \varepsilon$” 称为大偏差, 其概率 $P(|X - E(X)| \geqslant \varepsilon)$称为大偏差发生的概率.

显然契比雪夫不等式给出了大偏差发生的概率上界, 这个上界与方差成正比, 方差愈大上界也愈大.

契比雪夫不等式的含义

定理 (契比雪夫不等式的另一种形式)

设随机变量 X 的期望为 μ, 存在有限方差 σ^2, $k>0$, 则

$$P\{|X-\mu| \geqslant k\sigma\} \leqslant \frac{1}{k^2}. \tag{5}$$

特别地,

$$P\{|X-\mu| \geqslant 2\sigma\} \leqslant \frac{1}{4},$$

$$P\{|X-\mu| \geqslant 3\sigma\} \leqslant \frac{1}{9},$$

$$P\{|X-\mu| \geqslant 4\sigma\} \leqslant \frac{1}{16}.$$

例 (5.2.1)

已知某厂产品一周的产量是均值为 50 的随机变量, 若已知周产量的方差为 25, 求这一周产量在区间 $(40, 60)$ 内的可能性至少有多大?

例 (5.2.1)

已知某厂产品一周的产量是均值为 50 的随机变量，若已知周产量的方差为 25，求这一周产量在区间 $(40, 60)$ 内的可能性至少有多大?

解

用 X 表示周产量，则 $E(X) = 50$，$D(X) = 25$. 由契比雪夫不等式可得

$$P\{|X - 50| \geqslant 10\} \leqslant \frac{D(X)}{10^2} = \frac{1}{4} = 0.25,$$

$$P\{|X - 50| < 10\} \geqslant 1 - P\{|X - 50| \geqslant 10\} = 1 - 0.25 = 0.75,$$

可见这一周产量在区间 $(40, 60)$ 内的可能性至少为 0.75.

例 (5.2.2)

设电站供电网有 10000 盏电灯，夜晚每一盏灯开灯的概率都是 0.7，而假定开、关时间彼此独立，试用契比雪夫不等式估计夜晚同时开着的灯数在 6800 与 7200 之间的概率.

例 (5.2.2)

设电站供电网有 10000 盏电灯，夜晚每一盏灯开灯的概率都是 0.7，而假定开、关时间彼此独立，试用契比雪夫不等式估计夜晚同时开着的灯数在 6800 与 7200 之间的概率.

解

设 X 表示在夜晚同时开着的灯的数目，则 $X \sim b(10000, 0.7)$，故

$$E(X) = np = 10000 \times 0.7 = 7000,$$

$$D(X) = npq = 10000 \times 0.7 \times 0.3 = 2100,$$

$$P\{6800 < X < 7200\} = P\{|X - 7000| < 200\} \geqslant 1 - \frac{2100}{200^2} \approx 0.95.$$

可见，虽然有 10000 盏灯，但是只要有供应 7200 盏灯的电力就能够以相当大的概率保证够用. 事实上，契比雪夫不等式的估计只说明概率大于 0.95，后面我们将具体求出这个概率约为 0.99999. 可见契比雪夫不等式在理论上具有重大意义，但估计的精确度不高.

1. 特征函数
2. 大数定律
   - 契比雪夫不等式
   - 大数定律
3. 随机变量序列的两种收敛性
4. 中心极限定理

依概率收敛

定义 (依概率收敛)

设 $\{X_n\}$ 为一随机变量序列，X 为一随机变量，如果对于任意的 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\{|X_n - X| < \varepsilon\} = 1,$$

则称 $\{X_n\}$ 依概率收敛于 X，记作 $X_n \xrightarrow{P} X$.

长期以来，人们发现随机变量序列的算术平均值具有某种稳定性. 习惯上，称研究这种稳定性规律的定理为大数定律 (Law of Large Numbers). 它主要解决的问题是: 在什么样的条件下，一个随机变量序列的算术平均值收敛到所希望的期望值. 本节只考虑几个常见的大数定律.
契比雪夫不等式作为一个理论工具，在大数定律证明中，可使证明非常简洁.

定理 (5.2.2 契比雪夫 (Chebyshev) 大数定律)

设 $X_1, X_2, \cdots$ 是两两不相关的随机变量序列，各有数学期望 $E(X_1)$, $E(X_2), \cdots$ 及方差 $D(X_1), D(X_2), \cdots$, 并且对于所有 $i = 1, 2, \cdots$ 都有 $D(X_i) < C$, (C 是与 i 无关的常数), 则 $X_1, X_2, \cdots$ 服从大数定律，简记为 $X_n \sim LLN$. 即对任给 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right\} = 1. \tag{6}$$

证明.

因 $X_1, X_2, \cdots$ 两两不相关, 所以

$$D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)<\frac{1}{n^2}\cdot nC=\frac{C}{n},$$

又因

$$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i),$$

由 (4) 式, 对于任意 $\varepsilon>0$, 有

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)\right|<\varepsilon\right\}\geqslant 1-\frac{C}{n\varepsilon^2},$$

$\square$

证明.

但是任何事件的概率都不超过 1，即

$$1-\frac{C}{n\varepsilon^2}\leqslant P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)\right|<\varepsilon\right\}\leqslant 1,$$

因此

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)\right|<\varepsilon\right\}=1.$$

□

契比雪夫大数定律说明：在定理的条件下，当 n 充分大时，n 个两两不相关随机变量的平均数构成的随机变量的离散程度是很小的. 这意味，经过算术平均以后得到的随机变量 $\frac{1}{n}\sum_{i=1}^{n}X_i$ 将比较密地聚集在它的数学期望 $\frac{1}{n}\sum_{i=1}^{n}E(X_i)$ 的附近，它与数学期望之差依概率收敛到 0.

定理 (5.2.3 伯努利 (Bernoulli) 大数定律)

设 μ_A 是 n 次独立重复试验中事件 A 发生的次数. p 是事件 A 在每次试验中发生的概率, 则对于任意正数 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left\{\left|\frac{\mu_A}{n} - p\right| < \varepsilon\right\} = 1, \tag{7}$$

其等价形式为

$$\lim_{n\to\infty} P\left\{\left|\frac{\mu_A}{n} - p\right| \geqslant \varepsilon\right\} = 0. \tag{8}$$

证明.

引入随机变量

$$X_k = \begin{cases} 0, & \text{若在第 } k \text{ 次试验中 } A \text{ 不发生;} \\ 1, & \text{若在第 } k \text{ 次试验中 } A \text{ 发生.} \end{cases} \quad k = 1, 2, \cdots, n.$$

显然

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

由于 X_k 只依赖于第 k 次试验, 而各次试验是独立的. 于是 $X_1, X_2, \cdots, X_n$ 是相互独立的; 又由于 $X_k \sim b(1, p)$, 故有

$$E(X_k) = p, D(X_k) = p(1-p), \ k = 1, 2, \cdots, n.$$

□

证明.

由契比雪夫大数定律有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right\}$$

$$= \lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| < \varepsilon\right\} = 1,$$

即

$$\lim_{n\to\infty} P\left\{\left|\frac{\mu_A}{n} - p\right| < \varepsilon\right\} = 1.$$

□

伯努利大数定律的意义

- 伯努利大数定律说明: 随着试验次数 n 的增大, 事件 A 发生的频率 $\frac{\mu_A}{n}$ 与事件 A 发生的概率 p 的偏差 $\left|\frac{\mu_A}{n}-p\right|$ 大于预先给定的精度 ε 的可能性愈来愈小, 小到可以忽略不计, 这就是频率稳定于概率的含义.
- 因此, 本定律从理论上证明了大量重复独立试验中事件 A 发生的频率具有稳定性, 正因为这种稳定性, 概率的概念才有实际意义.
- 伯努利大数定律还提供了通过试验来确定事件的概率的方法, 即既然频率 $\frac{\mu_A}{n}$ 与概率 p 有较大偏差的可能性很小, 于是我们就可以通过做试验确定某事件发生的频率, 并把它作为相应概率的估计.
- 因此, 在实际应用中, 如果试验的次数很大时, 就可以用事件发生的频率代替事件发生的概率.

注意到在契比雪夫大数定律的证明中，只要有

$$\frac{1}{n^2}D(\sum_{i=1}^{n}X_i)\to 0 \tag{9}$$

则序列就服从大数定律. 这个条件 (9) 被称为马尔可夫条件.

定理 (5.2.4 马尔可夫大数定律)

对随机变量序列 $X_1, X_2, \cdots$，若马尔可夫条件 $\frac{1}{n^2}D(\sum\limits_{i=1}^{n}X_i)\to 0$ 成立，则序列 $X_1, X_2, \cdots$ 服从大数定律，即对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)\right|<\varepsilon\right\}=1.$$

马尔可夫大数定律的重要性在于：对随机变量序列 $X_1, X_2, \cdots$，已经没有任何同分布、独立性、不相关性的假定. 契比雪夫大数定律显然可以由马尔可夫大数定律推出.

例 (5.2.3)

设两两独立的随机变量序列 $\{X_n\}$ 的分布列分别为:

$$P(X_n = \sqrt{\ln n}) = \frac{1}{2},\ P(X_n = -\sqrt{\ln n}) = \frac{1}{2},\ n = 1, 2, \cdots .$$

试证明随机变量序列满足大数定律.

例 (5.2.3)

设两两独立的随机变量序列 $\{X_n\}$ 的分布列分别为:

$$P(X_n = \sqrt{\ln n}) = \frac{1}{2},\ P(X_n = -\sqrt{\ln n}) = \frac{1}{2},\ n = 1, 2, \cdots .$$

试证明随机变量序列满足大数定律.

证明.

容易得到: $E(X_n) = 0,\ D(X_n) = \ln n$. 又

$$\frac{1}{n^2} D(\sum_{i=1}^{n} X_i) = \frac{n \ln n}{n^2} = \frac{\ln n}{n} \to 0,\ n \to \infty.$$

满足马尔可夫条件, 所以随机变量序列 $\{X_n\}$ 服从大数定律. □

我们知道，一个随机变量的方差存在，则其数学期望一定存在；反之不成立. 上述的几个大数定律要求随机变量 X_k $(k = 1, 2, \cdots, n)$ 的方差存在. 但在随机变量服从同一分布的场合，并不需要这一要求，以下辛钦大数定律表明了这一点.

定理 (5.2.5 辛钦 (Khinchin) 大数定律)

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从同一分布，且具有数学期望 $E(X_k) = \mu$ $(k = 1, 2, \cdots)$，则序列 $X_n \sim LLN$. 即对任给 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \varepsilon\right\} = 1. \tag{10}$$

显然，伯努利大数定律是辛钦大数定律的特殊情况，辛钦大数定律在实际中应用很广泛.

这一定律使算术平均值的法则有了理论根据. 如要测定某一物理量 a, 在不变的条件下重复测量 n 次, 得观测值 $x_1, x_2, \cdots, x_n$, 求得实测值的算术平均值 $\frac{1}{n}\sum_{i=1}^{n} x_i$, 根据此定理, 当 n 足够大时, 取 $\frac{1}{n}\sum_{i=1}^{n} x_i$ 作为 a 的近似值, 可以认为所发生的误差是很小的, 所以实用上往往用某物体的某一指标值的一系列实测值的算术平均值来作为该指标值的近似值.
事实上, 用观测值的平均去作为随机变量的均值在实际生活中是常用的方法. 比如, 用观察到的某地区 50000 个人的平均寿命作为该地区的人均寿命的近似值是合适的, 这样做法的依据就是辛钦大数定律.

例 (5.2.4 蒙特卡洛方法计算定积分 (平均值法))

近似计算定积分

$$J = \int_0^1 f(x)\mathrm{d}x.$$

解

设随机变量 X 服从 $(0,1)$ 的均匀分布，则 $Y=f(X)$ 的数学期望为

$$E(f(X))=\int_0^1 f(x)\mathrm{d}x=J,$$

所以估计积分 J 的值就是估计 $f(X)$ 的期望值. 由辛钦大数定律，可以用 $f(X)$ 的观测值的平均值去估计 $f(X)$ 的期望值. 可按如下步骤来完成:

(1) 先用计算机产生 n 个 $(0,1)$ 上的均匀分布的随机数: $x_i,\ i=1,\cdots,n$.

(2) 计算每个 x_i 对应的 $f(x_i)$.

(3) 最后可得 J 的估计值为

$$J=\int_0^1 f(x)\mathrm{d}x\approx\frac{1}{n}\sum_{i=1}^{n}f(x_i).$$

思考: 如何用平均值法近似计算定积分 $J=\int_a^b f(x)\mathrm{d}x$?

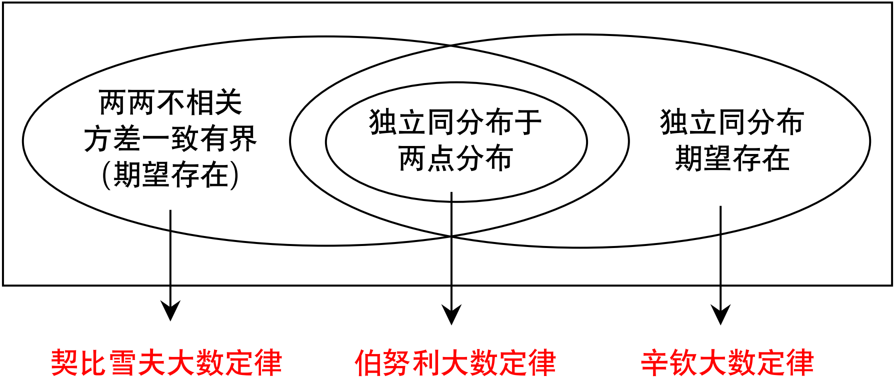

图: 三个大数定律的条件关系

1. 特征函数
2. 大数定律
3. 随机变量序列的两种收敛性
   - 依概率收敛
   - 依分布收敛
4. 中心极限定理

首先由随机变量序列 $\{X_n\}$ 服从大数定律启发我们引进如下定义

定义 (5.3.1 依概率收敛)

设 $\{X_n\}$ 为一随机变量序列，X 为一随机变量，如果对于任意的 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\{|X_n - X| < \varepsilon\} = 1,$$

则称 $\{X_n\}$ 依概率收敛于 X，记作 $X_n \xrightarrow{P} X$.

有了依概率收敛的定义，随机变量序列 $\{X_n\}$ 服从大数定律可以表述为

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i) \xrightarrow{P} 0, \ (n \to \infty).$$

例 (5.3.1)

设 $\{X_n\}$ 为一独立同分布的随机变量序列, 期望 $E(X_i) = a$, 方差存在. 试证明:

$$\frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i \xrightarrow{P} a, \ (n \to \infty).$$

例 (5.3.1)

设 $\{X_n\}$ 为一独立同分布的随机变量序列, 期望 $E(X_i) = a$, 方差存在. 试证明:

$$\frac{2}{n(n+1)} \sum_{i=1}^{n} iX_i \xrightarrow{P} a, \quad (n \to \infty).$$

证明.

记 $D(X_i) = \sigma^2$, $Y_n = \dfrac{2}{n(n+1)} \sum\limits_{i=1}^{n} iX_i$, 则

$$E(Y_n) = \frac{2}{n(n+1)} \sum_{i=1}^{n} ia = a, \quad D(Y_n) = \frac{4}{n^2(n+1)^2} \sum_{i=1}^{n} i^2\sigma^2 < \frac{4}{3n}\sigma^2,$$

$\square$

例 (5.3.1)

设 $\{X_n\}$ 为一独立同分布的随机变量序列，期望 $E(X_i)=a$，方差存在. 试证明:

$$\frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i \xrightarrow{P} a,\ (n\to\infty).$$

证明.

对于任意的 $\varepsilon>0$，由契比雪夫不等式有

$$P(|Y_n-a|\geqslant\varepsilon)\leqslant\frac{1}{\varepsilon^2}D(Y_n)\leqslant\frac{1}{\varepsilon^2}\frac{4\sigma^2}{3n}\to 0\ (n\to\infty),$$

即 $P(|Y_n-a|<\varepsilon)\to 1,\ (n\to\infty)$，所以

$$\frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i \xrightarrow{P} a,\ (n\to\infty).$$

□

下面给出依概率收敛的一个重要性质，其证明参见参考文献 [7].

定理 (5.3.1)

设 $\{X_n\}$, $\{Y_n\}$ 是两随机变量序列，a, b 是常数，且

$$X_n \xrightarrow{P} a, \quad Y_n \xrightarrow{P} b,$$

则

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

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

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

我们知道分布函数全面描述了随机变量的统计规律性，因此讨论一个随机变量序引的分布函数列 $\{F_n(x)\}$ 收敛到一个极限分布函数 $F(x)$ 是有实际意义的. 下面给出随机变量序列依分布收敛的定义.

定义 (5.3.2 依分布收敛)

设一个随机变量序列 $\{X_n\}$ 的分布函数列为 $\{F_n(x)\}$，$F(x)$ 为随机变量 X 的分布函数，若在 $F(x)$ 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 $F_n(x) \xrightarrow{W} F(x)$. 也称随机变量序列 $\{X_n\}$ 依分布收敛于 X，记作 $X_n \xrightarrow{L} X$.

下面给出依概率收敛与依分布收敛之间的一个重要关系，其证明参见参考文献 [7].

定理 (5.3.2)

设 $\{X_n\}$ 是随机变量序列，c 为常数，则有

(1) $X_n \xrightarrow{P} X \Longrightarrow X_n \xrightarrow{L} X$.

(2) $X_n \xrightarrow{P} c \Longleftrightarrow X_n \xrightarrow{L} c$.

利用定义来证明依分布收敛有时是相当麻烦的，这时可以利用特征函数这样有用工具. 下面给出一个判断依分布收敛的一个重要结论，证明从略.

定理 (5.3.3)

分布函数列 $\{F_n(x)\}$ 弱收敛于分布函数 $F(x)$ 的充分必要条件是 $\{F_n(x)\}$ 的特征函数列 $\{\varphi_n(t)\}$ 收敛于 $F(x)$ 的特征函数 $\varphi(t)$.

例 (5.3.2 证明辛钦大数定律)

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立, 服从同一分布, 且具有数学期望 $E(X_k) = a\ (k = 1, 2, \cdots)$, 证明

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

例 (5.3.2 证明辛钦大数定律)

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从同一分布，且具有数学期望 $E(X_k) = a\ (k = 1, 2, \cdots)$，证明

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

证明.

注意到 a 为常数，所以只需证明

$$Y = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{L} a.$$

由定理 5.3.3 知只需证明 $\varphi_{Y_n}(t) \to e^{iat}$. □

证明.

因为 X_i 独立同分布，所以具有相同的特征函数，记为 $\varphi(t)$. 由特征函数的性质知 $\varphi'(0) = iE(X_i) = ia$，从而其在 $t = 0$ 的展开式为

$$\varphi(t) = \varphi(0) + \varphi'(0)t + o(t) = 1 + iat + o(t).$$

由 X_i 之间的独立性知 Y_n 的特征函数为

$$\varphi_{Y_n}(t) = \left[\varphi\left(\frac{t}{n}\right)\right]^n = \left[1 + ia\frac{t}{n} + o\left(\frac{1}{n}\right)\right]^n \to e^{iat}.$$

定律得证. □

1. 特征函数
2. 大数定律
3. 随机变量序列的两种收敛性
4. 中心极限定理

在实践中常常需要考虑大量随机因素所产生的总效应. 例如误差是人们经常遇到且感兴趣的随机变量, 大量的研究表明, 误差的产生是由大量微小的相互独立的随机因素叠加而成的. 比如一位操作者在机床加工机械轴, 使其直径符合规定要求, 但加工后的机械轴与规定要求总有一定的误差, 这是因为在加工时受到一些随机因素的影响:

(1) 在机床方面有机床振动与转速的影响.

(2) 在刀具方面有装配与磨损的影响.

(3) 在材料方面有钢材的成分和产地的影响.

(4) 在操作者方面有注意力集中程度、当天的情绪的影响.

(5) 在测量方面有误差、测量技术的影响.

(6) 在环境方面有车间的温度、湿度、照明、工作电压的影响.

(7) 在具体场合下还可以列出许多其他的影响.

由于这些因素很多，每个因素对加工精度的影响都是微小的，每个因素的出现都是人们无法控制的、随机的、时有时无、时大时小、时正时负. 这些因素的综合影响最后使每个加工轴的直径产生误差，若将这个误差记为 Y_n，那么 Y_n 是随机变量，且可以将 Y_n 看成是许多微小的随机波动 $X_1, X_2, \cdots, X_n$ 之和，即

$$Y_n = X_1 + X_2 + \cdots + X_n.$$

人们关心的是当 $n \to \infty$ 时，Y_n 的极限分布是什么？下面介绍的中心极限定理 (Central Limit Theorem) 将告诉我们：在一定的条件下，Y_n 的极限分布是正态分布.

定理 (5.4.1 独立同分布的中心极限定理)

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立, 服从同一分布, 且具有数学期望 $E(X_k) = \mu$ 和方差 $D(X_k) = \sigma^2 > 0$, $(k = 1, 2, \cdots)$. 记

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}},$$

则对于任意实数 y 满足

$$\lim_{n\to\infty} P\{Y_n \leqslant y\} = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t,$$

即: $\{X_n\}$ 服从中心极限定理, 简记为 $X_n \sim CLT$.

从上述定理的结论可知, 当 n 充分大时, 近似地有 $Y_n \sim N(0, 1)$, 或者说, 当 n 充分大时, 近似地有

$$\sum_{k=1}^{n} X_k \sim N(n\mu, n\sigma^2), \text{ 或 } \frac{1}{n}\sum_{k=1}^{n} X_k \sim N(\mu, \frac{\sigma^2}{n}).$$

例 (5.4.1)

电压加法器上收到 48 个信号电压: $V_1, V_2, \cdots, V_{48}$, 它们相互独立, 且服从 $(0, 10)$ 的均匀分布 (单位: 伏), 试求加法器上总电压不超过 260 伏的近似概率.

例 (5.4.1)

电压加法器上收到 48 个信号电压：$V_1, V_2, \cdots, V_{48}$，它们相互独立，且服从 $(0, 10)$ 的均匀分布 (单位：伏)，试求加法器上总电压不超过 260 伏的近似概率.

解

$$E(V_k) = \frac{a+b}{2} = 5,\ D(V_k) = \frac{(b-a)^2}{12} = \frac{100}{12},\ k = 1, 2, \cdots, 48.$$

故由独立同分布的中心极限定理可得

$$P\left(\sum_{k=1}^{48} V_k \leqslant 260\right) = P\left(\frac{\sum\limits_{k=1}^{48} V_k - 48 \times 5}{\sqrt{48 \times 100/12}} \leqslant \frac{260 - 48 \times 5}{\sqrt{48 \times 100/12}}\right)$$

$$\approx \Phi(1) = 0.8413.$$

下面介绍另一个中心极限定理.

定理 (5.4.2 棣莫佛 – 拉普拉斯 (De Moivre-Laplace) 中心极限定理)

设随机变量 X 服从参数为 n, p $(0 < p < 1)$ 的二项分布, 则对于任意的 x, 有

$$\lim_{n\to\infty} P\left\{\frac{X-np}{\sqrt{np(1-p)}} \leqslant x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t.$$

正态分布是二项分布的极限分布

- 棣莫佛一拉普拉斯中心极限定理是概率论历史上第一个中心极限定理，它是专门针对二项分布的，因此也称为“二项分布的正态近似”.
- 前面第 2 章泊松定理给出了“二项分布的泊松近似”. 两者相比，一般在 p 较小时用泊松近似较好，而在 $np > 5$ 和 $n(1-p) > 5$ 时，用正态分布近似较好.
- 在利用该定理近似计算二项分布概率时，由于二项分布是离散分布而正态分布是连续分布，所以适当修正可以提高精度.
- 一般地，若 $k_1 < k_2$ 均为整数，可以先做如下修正，再用正态分布近似.

$$P(k_1 \leqslant X \leqslant k_2) = P(k_1 - 0.5 < X < k_2 + 0.5).$$

例 (5.4.2)

某厂有 25 台同类机器，每台机器发生故障的概率都是 0.4，假设每台机器工作是相互独立的，试计算发生故障的机器台数在 $5-15$ 之间的概率.

例 (5.4.2)

某厂有 25 台同类机器，每台机器发生故障的概率都是 0.4，假设每台机器工作是相互独立的，试计算发生故障的机器台数在 5 − 15 之间的概率.

解

设 X 表示故障发生的台数，由题意知$X \sim b(25, 0.4)$. 所求概率为

$$P(5 \leqslant X \leqslant 15),$$

由中心极限定理知可用正态分布来近似计算:

(1) 不做修正直接近似计算:

$$\begin{aligned} P(a \leqslant X \leqslant b) &\approx \Phi\left(\frac{b - n \times p}{\sqrt{npq}}\right) - \Phi\left(\frac{a - n \times p}{\sqrt{npq}}\right) \\ &= \Phi\left(\frac{15 - 25 \times 0.4}{\sqrt{25 \times 0.4 \times 0.6}}\right) - \Phi\left(\frac{5 - 25 \times 0.4}{\sqrt{25 \times 0.4 \times 0.6}}\right) \\ &= 2\Phi(2.041) - 1 = 0.9588. \end{aligned}$$

例 (5.4.2)

某厂有 25 台同类机器，每台机器发生故障的概率都是 0.4，假设每台机器工作是相互独立的，试计算发生故障的机器台数在 $5-15$ 之间的概率.

解

(2) 适当修正后近似计算:

$$
\begin{aligned}
P(5 \leqslant X \leqslant 15) &= P(4.5 < X < 15.5) \\
&\approx \Phi\left(\frac{15.5-25\times 0.4}{\sqrt{25\times 0.4\times 0.6}}\right)-\Phi\left(\frac{4.5-25\times 0.4}{\sqrt{25\times 0.4\times 0.6}}\right) \\
&= 2\Phi(2.245)-1 = 0.9754.
\end{aligned}
$$

事实上，利用二项分布的分布列可以求出的精确值为 0.9780，显然未经修正的近似值误差要大.

本章小结

(1) 特征函数.

(2) 契比雪夫不等式、大数定律.

(3) 中心极限定理.

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立, 且服从同一分布, 具有数学期望和方差 $E(X_k) = \mu$, $D(X_k) = \sigma^2 > 0$ $(k = 1, 2, \cdots)$. 则随机变量 $Y_n = \dfrac{\sum\limits_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$ 的分布函数 $F_n(x)$ 对任意实数 x 满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{\frac{\sum\limits_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \leqslant x\right\} = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \,\mathrm{d}t.$$

- 在实际应用时，当 $n \geqslant 30$ 时近似的有:

$$Y_n = \frac{\sum\limits_{k=1}^{n} X_k - n\mu}{\sqrt{n\sigma^2}} \sim N(0,1), \quad \sum_{k=1}^{n} X_k \sim N(n\mu, n\sigma^2),$$

用于计算一些概率的近似值.

- 在理论上可以说明正态分布的普遍性; 在数理统计中有重要的作用.

(4) * 随机变量序列的两种收敛性:

- 依概率收敛、依分布收敛.
- 大数定律是依概率收敛的问题; 中心极限定理是依分布收敛的问题.
- 两种收敛性之间的关系.