西安交通大学

第五章 大数定律与中心极限定理

人工智能学院
周三平
Email：spzhou@xjtu.edu.cn

本章主要内容

5.1 随机变量的收敛性与切比雪夫不等式

5.2 大数定律

5.3 中心极限定理

§5.1 随机变量的收敛性与切比雪夫不等式

5.1.1 依概率收敛

如 果 对 任 意 $\varepsilon>0$， 有 下 式 成 立：

$$\lim_{n\to\infty} P\{|\xi_n(\omega)-\xi(\omega)|\geq\varepsilon\}=0$$

则称 $\{\xi_n(\omega)\}$ 依概率收敛于 $\xi(\omega)$ ，并记为

$$\xi_n(\omega)\xrightarrow{P}\xi(\omega)$$

依概率收敛表明：随机变量 ξ_n 对 ξ 的绝对偏差不小于一个给定正数的概率随着 n 的增大而越来越趋向于零。

§5.1 随机变量的收敛性与切比雪夫不等式

5.1.1 依概率收敛

依概率收敛于常数的随机变量有一条重要的性质：

若 $(p)\lim\limits_{n\to\infty} X_n = a, (p)\lim\limits_{n\to\infty} Y_n = b$ **，又** $g(x,y)$ **在点**(a,b)**处连续，则**

$$(p)\lim_{n\to\infty} g(X_n, Y_n) = g(a,b)$$

特别地：

$$(p)\lim_{n\to\infty} X_n \pm Y_n = a \pm b, (p)\lim_{n\to\infty} X_n Y_n = ab, (p)\lim_{n\to\infty} \frac{X_n}{Y_n} = \frac{a}{b}(Y_n \neq 0, b \neq 0)$$

§5.1 随机变量的收敛性与切比雪夫不等式

案例1 设随机变量 X_n 服从柯西分布 $C\left(0,\frac{1}{n}\right)$，其概率密度为：

$$f_n(x)=\frac{n}{\pi\left(1+n^2x^2\right)},-\infty<x<+\infty$$

试证：随机变量序列$\{X_n\}$依概率收敛于0，即 $(p)\lim\limits_{n\to\infty}X_n=0$ 。

证明： 由依概率收敛的定义，对任意 $\varepsilon>0$ ，

$$P\{|X_n-0|<\varepsilon\}=P\{|X_n|<\varepsilon\}=\int_{-\varepsilon}^{+\varepsilon}\frac{n}{\pi\left(1+n^2x^2\right)}dx=\frac{2}{\pi}\arctan n\varepsilon$$

则：$P\{|X_n-0|<\varepsilon\}=\frac{2}{\pi}\arctan n\varepsilon=1$ 因此 X_n 依概率收敛于0。

§5.1 随机变量的收敛性与切比雪夫不等式

5.1.2 依分布收敛

设随机变量$\xi_n(\omega),\xi(\omega)$的分布函数分别为$F_n(x)$及$F(x)$

如果 $F_n(x)\xrightarrow{W}F(x)$，则称 $\{\xi_n(\omega)\}$ 依分布收敛于

$\xi(\omega)$，记为 $\xi_n(\omega)\xrightarrow{L}\xi(\omega)$ 或 $\xi_n(\omega)\xrightarrow{d}\xi(\omega)$。

§5.1 随机变量的收敛性与切比雪夫不等式

定理1 $\xi_n(\omega) \xrightarrow{P} \xi(\omega) \Rightarrow \xi_n(\omega) \xrightarrow{L} \xi(\omega).$

证明： 因为，对 $x' < x$，有

$$\{\xi < x'\} = \{\xi_n < x, \xi < x'\} + \{\xi_n \geq x, \xi < x'\}$$
$$\subset \{\xi_n < x\} + \{\xi_n \geq x, \xi < x'\}$$

所以, $F(x') \leq F_n(x) + P\{\xi_n \geq x, \xi < x'\}$

因为 $\{\xi_n\}$ 依概率收敛于 ξ ，则

$$P\{\xi_n \geq x, \xi < x'\} \leq P\{|\xi_n - \xi| \geq x - x'\} \to 0$$

因而有 $F(x') \leq \varliminf_{n\to\infty} F_n(x)$

§5.1 随机变量的收敛性与切比雪夫不等式

同理，对 $x'' > x,$

$$\{\xi \ge x''\} = \{\xi_n < x, \xi \ge x''\} + \{\xi_n \ge x, \xi \ge x''\}$$

$$\subset \{\xi_n < x, \xi \ge x''\} + \{\xi_n \ge x\}$$

类似可得： $\overline{\lim_{n\to\infty}} F_n(x) \le F(x'')$

所以对 $x' < x < x''$，有 $F(x') \le \underline{\lim_{x\to\infty}} F_n(x) \le \overline{\lim_{n\to\infty}} F_n(x) \le F(x'')$

如果 x 是 $F(x)$ 的连续点，则令 x', x'' 趋于 x 可得

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

证毕.

§5.1 随机变量的收敛性与切比雪夫不等式

定理1逆命题不成立.

案例1 若样本空间 $\Omega=\{\omega_1,\omega_2\},P(\omega_1)=P(\omega_2)=\dfrac{1}{2}$，定义随机变量 $\xi(\omega)$ 如下： $\xi(\omega_1)=-1,\xi(\omega_2)=1$，则 $\xi(\omega)$ 的分布列为

$$\begin{pmatrix} -1 & 1 \\ 1/2 & 1/2 \end{pmatrix} \qquad (1)$$

若对一切 n，令 $\xi_n(\omega)=-\xi(\omega)$，显然 $\xi_n(\omega)$ 的分布列也是(1) 因此，$\xi_n(\omega)\xrightarrow{L}\xi(\omega)$

§5.1 随机变量的收敛性与切比雪夫不等式

但是, 对任意的$0<\varepsilon<2$，因 $P\{|\xi_n(\omega)-\xi(\omega)|>\varepsilon\}=P(\Omega)=1$

因此, $\{\xi_n(\omega)\}$不依概率收敛于$\xi(\omega)$。但是在特殊场合却有下面的结果:

定理2 设C是常数，则 $\xi_n(\omega)\xrightarrow{P}C \Leftrightarrow \xi_n(\omega)\xrightarrow{L}C.$

§5.1 随机变量的收敛性与切比雪夫不等式

定理2 设C是常数，则 $\xi_n(\omega) \xrightarrow{P} C \Leftrightarrow \xi_n(\omega) \xrightarrow{L} C.$

证明： 由前面的定理可知，只须证明由依分布收敛于常数可推出依概率收敛于常数。事实上，对任意ε >0,

$$P\{|\xi_n - C| \ge \varepsilon\} = P\{\xi_n \ge C+\varepsilon\} + P\{\xi_n \le C-\varepsilon\}$$

$$= 1 - F_n(C+\varepsilon) + F_n(C-\varepsilon+) \xrightarrow{n\to\infty} 1 - F(C+\varepsilon) + F(C-\varepsilon+)$$

$$= 1 - 1 + 0 = 0.$$

$$F(x) = \begin{cases} 0, & x < C \\ 1, & x \ge C \end{cases}$$

证毕.

§5.1 随机变量的收敛性与切比雪夫不等式

拓展1：r阶矩收敛

设对随机变量 ξ_n 及 ξ 有 $E|\xi_n|^r<\infty, E|\xi|^r<\infty$ ，其中 $r>0$ 为常数，如果 $\lim\limits_{n\to\infty} E|\xi_n-\xi|^r=0$，则称 $\{\xi_n\}$ r 阶(矩)收敛于ξ，并记为 $\xi_n \xrightarrow{r} \xi$

在r阶收敛中，最重要的是r=2的情况，称为均方收敛。

§5.1 随机变量的收敛性与切比雪夫不等式

拓展2：几乎处处收敛

如果 $P\{\lim\limits_{n\to\infty}\xi_n(\omega)=\xi(\omega)\}=1$ ，则称$\{\xi_n(\omega)\}$ **以概率1收敛**于 $\xi(\omega)$ ，又称 $\{\xi_n(\omega)\}$ **几乎处处收敛**于 $\xi(\omega)$ ，记为

$$\xi_n(\omega)\xrightarrow{a.s.}\xi(\omega).$$

§5.1 随机变量的收敛性与切比雪夫不等式

r 阶收敛与依概率收敛的关系

定理3 $\xi_n \xrightarrow{r} \xi \Rightarrow \xi_n \xrightarrow{P} \xi.$

证明:

先证对于任意$\varepsilon > 0$，成立 $P\{|\xi_n - \xi| \geq \varepsilon\} \leq \dfrac{E|\xi_n - \xi|^r}{\varepsilon^r}$

这个不等式是Chebyshev不等式的推广，称作**Markov不等式**。

$$P\{|\xi - C| \geq \varepsilon\} \leq \frac{E|\xi - C|^r}{\varepsilon^r}$$

§5.1 随机变量的收敛性与切比雪夫不等式

事实上，若以 $F(x)$ 记 $\xi_n-\xi$ 的分布函数，则有

$$P\{|\xi_n-\xi|\geq\varepsilon\}=\int_{|x|\geq\varepsilon}\mathrm{d}F(x)\leq\int_{|x|\geq\varepsilon}\frac{|x|^r}{\varepsilon^r}\mathrm{d}F(x)$$

$$\leq\frac{1}{\varepsilon^{\mathrm{r}}}\int_{|x|\geq\varepsilon}|x|^r\,\mathrm{d}F(x)\leq\frac{E|\xi_n-\xi|^r}{\varepsilon^{\mathrm{r}}}\xrightarrow{n\to\infty}0$$

定理3逆命题不成立。 下例说明：由依概率收敛或几乎处处收敛不能推出以r阶收敛。

§5.1 随机变量的收敛性与切比雪夫不等式

定理4 $\xi_n(\omega) \xrightarrow{a.s} \xi(\omega) \Rightarrow \xi_n(\omega) \xrightarrow{P} \xi(\omega).$

反例：依概率收敛不能导致几乎处处收敛

$\forall\, k\in\mathrm{N}$，把 $(0,1]$ k 等分，定义 k 个随机变量：

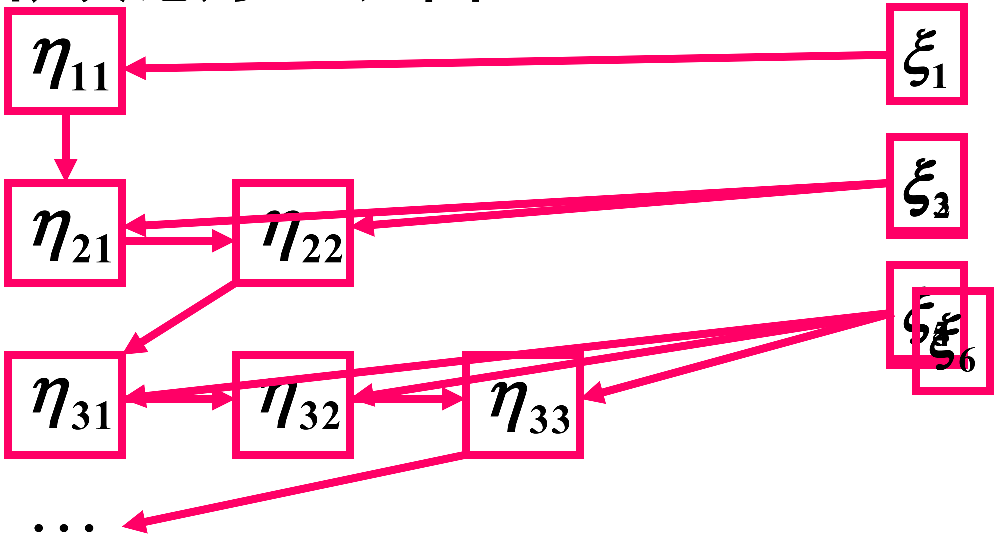
$$\eta_{ki}(\omega)=\begin{cases}1, & \omega\in\left(\dfrac{i-1}{k},\dfrac{i}{k}\right], \quad i=1,2,\cdots,k.\\ 0, & \text{otherwise}\end{cases}$$

取 P 为勒贝格测度，则

$$\forall\,\varepsilon>0,\ P(|\eta_{ki}(\omega)|\geq\varepsilon)=\frac{1}{k},\quad i=1,2,\cdots,k.\quad (*)$$

§5.1 随机变量的收敛性与切比雪夫不等式

将 η_{ki} 依次记为 ξ_n 如图:



即 $\xi_n(\omega)=\eta_{ki}(\omega),\quad n=i+\dfrac{k(k-1)}{2}.$

§5.1 随机变量的收敛性与切比雪夫不等式

由(*)式，$\forall\ \varepsilon > 0$ ，$\lim\limits_{n\to\infty} P(|\xi_n(\omega)| \geq \varepsilon) = \lim\limits_{k\to\infty} P(|\eta_{ki}(\omega)| \geq \varepsilon) = 0.$

从而，$\xi_n(\omega) \xrightarrow{P} 0$，同时，$\forall\ \omega \in (0,1]$，总有无数个 $\eta_{ki} = 1$，以及无数个 $\eta_{ki} = 0$，所以 ξ_n 处处不收敛。

上述 ξ_n 满足 $E|\xi_n|^r = E|\eta_{ki}|^r = \dfrac{1}{k} \to 0, as\ n \to \infty$，即 $\xi_n(\omega) \xrightarrow{r} 0$，可见，矩收敛也不能蕴涵几乎处处收敛。

§5.1 随机变量的收敛性与切比雪夫不等式

收敛性小结

依分布收敛 $\xi_n(\omega) \xrightarrow{d} \xi(\omega)$

依概率收敛 $\xi_n(\omega) \xrightarrow{P} \xi(\omega)$

r阶矩收敛 $\xi_n \xrightarrow{r} \xi$

几乎处处收敛 $\xi_n(\omega) \xrightarrow{a.s.} \xi(\omega).$

§5.1 随机变量的收敛性与切比雪夫不等式

依分布收敛

依概率收敛

几乎处处收敛

r阶矩收敛

§5.1 随机变量的收敛性与切比雪夫不等式

5.1.3 切比雪夫不等式

定理 设随机变量X的数学期望$E(X)=\mu$，方差$D(X)=\sigma^2$，则对任意的正数ε，有 $P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}$ 或 $P\{|X-\mu|<\varepsilon\}\geq 1-\frac{\sigma^2}{\varepsilon^2}$

证明（X为连续型）设X的概率密度为$f(x)$，则

$$P\{|X-\mu|\geq\varepsilon\}=\int_{|x-\mu|\geq\varepsilon} f(x)dx \leq \int_{|x-\mu|\geq\varepsilon}\frac{(x-\mu)^2}{\varepsilon^2}f(x)dx$$

$$\leq\frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx=\frac{D(X)}{\varepsilon^2}=\frac{\sigma^2}{\varepsilon^2}$$ 证毕.

§5.1 随机变量的收敛性与切比雪夫不等式

切比雪夫不等式的意义

□　这个不等式给出了在随机变量X的分布未知的情况下事件|x−μ|<ε 的概率的一种估计方法。例如

$$P\{|X-\mu|<3\sigma\}\geq 1-\frac{1}{9}=0.8889$$

□　切比雪夫不等式从另一角度体现了方差D(X)的意义。从切比雪夫不等式可以看出，随机变量X的方差越小，则X的取值越集中在其中心E(X)的附近。方差越小，X取值越集中在区间(E(X)−ε，E(X)+ε)之内。

§5.1 随机变量的收敛性与切比雪夫不等式

案例2 设有一大批种子,其中良种占1/6.试估计在任选的 6000 粒种子中,良种所占比例与1/6比较上下小于1% 的概率。

解:设X表示 6000 粒种子中的良种数, 则

$$X \sim B(6000, 1/6) \quad E(X) = 1000, D(X) = \frac{5000}{6}$$

$$P\left\{\left|\frac{X}{6000} - \frac{1}{6}\right| < 0.01\right\} = P\{|X - 1000| < 60\} \geq 1 - \frac{5000 - 6}{60^2}$$

$$= \frac{83}{108} = 0.7685$$

§5.1 随机变量的收敛性与切比雪夫不等式

案例3 若某班某次考试的平均成绩是75，方差为10，试估计及格率至少是多少？

解： 设随机变量X表示学生的成绩，由题意知E(x)=75，D(x)=10,由切比雪夫不等式知

$$P\{60 \le X \le 100\} = P\{-15 \le X - 75 \le 25\} \ge P\{|X - 75| < 15\}$$

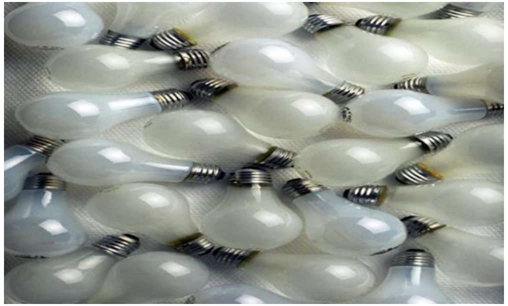
$$\ge 1 - \frac{10}{225} = \frac{215}{225} = 0.956$$

因此估计及格率至少95.6%

§5.2 大数定律

大数定律的客观背景

大量抛掷硬币
正面出现频率



生产过程中
的废品率

字母使用频率

（1）频率稳定性

（2）大量测量结果算术平均值的稳定性。

§5.2 大数定律

抛硬币试验 将一枚硬币连续抛 n 次,记A={ 正面朝上次数为n_A 次} 则 A 发生的频率为 $f_n(A)=\frac{n_A}{n}$

概率论历史上几个有名的"抛硬币"试验:

实验者	n	n_A	ξ_n
蒲丰(18世纪)	4048	2048	0.5069
德 · 摩根(19世纪)	2048	1061	0.5181
皮尔逊(19世纪)	12000	6019	0.5016
皮尔逊(19世纪)	24000	12012	0.5005
罗曼诺夫斯基(20世纪)	80640	39699	0.4923

可见 $f_n(A)=\frac{n_A}{n}\to P(A)=0.5 \quad (n\to\infty)$

§5.2 大数定律

思考 在抛硬币试验中会不会出现下列情形？

情形(1) 每次试验全出现正面 $f_n(A)=\dfrac{n_A}{n}=1$

情形(2) 每次试验全出现反面 $f_n(A)=\dfrac{n_A}{n}=0$

在理论上完全可能出现，但出现的概率非常小！

即当 $n\to\infty$ 时，概率越来越小、趋于零。

§5.2 大数定律

Representing Multimodal Behaviors With Mean Location for Pedestrian Trajectory Prediction

Publisher: IEEE

Cite This | PDF

Liushuai Shi ; Le Wang ; Chengjiang Long ; Sanping Zhou ; Wei Tang ; Nanning Zheng ; Gang Hua   All Authors

Abstract

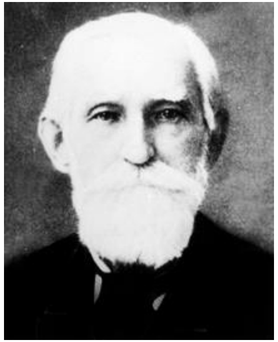
Abstract:
Representing multimodal behaviors is a critical challenge for pedestrian trajectory prediction. Previous methods commonly represent this multimodality with multiple latent variables repeatedly sampled from a latent space, encountering difficulties in interpretable trajectory prediction. Moreover, the latent space is usually built by encoding global interaction into future trajectory, which inevitably introduces superfluous interactions and thus leads to performance reduction. To tackle these issues, we propose a novel Interpretable Multimodality Predictor (IMP) for pedestrian trajectory prediction, whose core is to represent a specific mode by its mean location. We model the distribution of mean location as a Gaussian Mixture Model (GMM) conditioned on sparse spatio-temporal features, and sample multiple mean locations from the decoupled components of GMM to encourage multimodality. Our IMP brings four-fold benefits: 1) Interpretable prediction to provide semantics about the motion behavior of a specific mode; 2) Friendly visualization to present multimodal behaviors; 3) Well theoretical feasibility to estimate the distribution of

§5.2 大数定律

5.2.1 切比雪夫大数定律

定理(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，它们分别存在数学期望与方差 $E(X_k), D(X_k)$, 若存在常数 C，使得 $D(X_k) \le C, k = 1, 2, \cdots$ 则对任意的$\varepsilon > 0$, 有

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



切比雪夫

§5.2 大数定律

证明 由于$X_1, X_2, \cdots, X_n, \cdots$ 相互独立. 于是令$Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k$

$$D\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] = \frac{1}{n^2}\sum_{k=1}^{n} D(X_k) \le \frac{C}{n}$$

由切比雪夫不等式可得

$$1 \ge P\{|Y_n - E(Y_n)| < \varepsilon\} \ge 1 - \frac{D(Y_n)}{\varepsilon^2} \ge 1 - \frac{C}{\varepsilon^2 n},$$

在上式中令$n \to \infty$，即

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

证毕.

§5.2 大数定律

推论1 (独立同分布下的大数定律)

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，且具有相同的数学期望和方差 $E(X_k)=\mu, D(X_k)=\sigma^2 (k=1,2,\cdots)$

作前n个随机变量的算术平均 $\bar{X}=\frac{1}{n}\sum_{k=1}^{n} X_k$，则对任意正数$\varepsilon$有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k-\mu\right|<\varepsilon\right\}=1.$$

§5.2 大数定律

解释： 推论使我们关于算术平均值的法则有了理论上的依据。如我们要测量某段距离，在相同条件下重复进行n次，得n个测量值，它们可以看成是n个相互独立的随机变量,具有相同的分布、相同的数学期望μ和方差σ^2，由推论知，只要n充分大，则以接近于1的概率保证

$$\boxed{\mu \approx \frac{1}{n}\sum_{i=1}^{n} X_i}$$

这便是在n较大情况下反映出的客观规律，故称为“大数”定律。

§5.2 大数定律

推论2（泊松大数定律）

如果在一个独立试验序列中，事件A在第k次试验中出现的概率等于p_k，以μ_n记在前n次试验中事件A出现的次数，则对任意$\varepsilon>0$，都有

$$\lim_{n\to\infty} P\left\{\left|\frac{\mu_n}{n}-\frac{p_1+p_2+\cdots+p_n}{n}\right|<\varepsilon\right\}=1$$

§5.2 大数定律

证明： 定义ξ_k为第k次试验中事件A 出现的次数，则

$$E\xi_k = p_k, \quad D\xi_k = p_k(1-p_k) \le \frac{1}{4}, \ \mu_n = \sum_{k=1}^{n} \xi_k,$$

再利用切比雪夫大数定律立刻推出结论。

明显，当$p_k = p$，泊松大数定律即为伯努利大数定律

§5.2 大数定律

5.2.2 伯努利大数定律

设n_A是n次重复独立试验中A发生的次数，p是事件A在每次试验中发生的概率,则对任意$\varepsilon > 0$,有

$$\lim_{n\to\infty} \mathrm{P}\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1 \quad \text{或} \quad \lim_{n\to\infty} \mathrm{P}\left\{\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right\} = 0$$

证明： 因为 $n_A \sim b(n, p)$, 有 $n_A = X_1 + X_2 + \cdots + X_n$

因而$E(X_k)=p$, $k=1,2,\ldots$, 由切比雪夫大数定理的推论,

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - p\right| < \varepsilon\right\} = 1 \quad \text{即} \quad \lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1$$

$$\therefore \frac{n_A}{n} \xrightarrow{P} P(A), (n \to \infty)$$

证毕.

§5.2 大数定律

[注]

1. 伯努利大数定理以严格的数学形式表达了频率的稳定性.
2. 伯努利大数定律提供了通过试验来确定事件概率的方法.

在实际应用中，当试验次数很大时，往往用事件发生的频率来代替事件的概率。

§5.2 大数定律

伯努利大数定律表明事件发生的频率 $\frac{n_A}{n}$ 依概率收敛于事件的概率 p ，它以严格的数学形式表达了频率的稳定性。

故而当 n 很大时，事件发生的频率与概率有较大偏差的可能性很小。根据实际推断原理，当试验次数很大时，便可以用事件发生的频率来代替事件的概率。

§5.2 大数定律

5.2.3 辛钦大数定律

定理（辛钦大数定律） 设 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是相互独立的随机变量序列，它们服从相同的分布，且具有有限的数学期望 a，则对任意的 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - a\right| < \varepsilon\right\} = 1 \quad 即 \quad \frac{1}{n}\sum_{i=1}^{n}\xi_i \xrightarrow{P} a.$$

§5.2 大数定律

证明：由于 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 具有相同分布，故有相同的特征函数，设为 $f(t)$，因为数学期望存在，故 $f(t)$ 可展开成：

$$f(t)=f(0)+f'(0)t+o(t)=1+iat+o(t)$$

而 $\frac{1}{n}\sum_{i=1}^{n}\xi_i$ 的特征函数为

$$\left[f\left(\frac{t}{n}\right)\right]^n=\left[1+ia\frac{t}{n}+o\left(\frac{t}{n}\right)\right]^n$$

对于固定的 t，$\left[f\left(\frac{t}{n}\right)\right]^n\to e^{iat}\quad(n\to\infty)$

§5.2 大数定律

极限函数 e^{iat} 是连续函数，它是退化分布 $I_a(x)$ 所对应的特征函数。由逆极限定理，有

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

最后由依分布收敛和依概率收敛的关系定理知：

$\frac{1}{n}\sum_{i=1}^{n}\xi_i$ 依概率收敛于常数 a ，从而证明了定理。

§5.2 大数定律

案例4 设$\{X_n\}$是独立同分布的随机变量序列，且每个X_n都服从参数为$\lambda>0$的泊松分布， 试问当$n\to\infty$时，$Y_n=\frac{1}{n}\sum_{k=1}^{n}X_k^2$依概率收敛于何值？

解： 由于 $X_1^2, X_2^2, \cdots, X_n^2, \cdots$ 是独立同分布的随机变量，且

$E\left(X_k^2\right)=\left[E\left(X_k\right)\right]^2+D\left(X_k\right)=\lambda^2+\lambda$ 对于$\left\{X_n^2, n=1,2,\cdots\right\}$应用辛钦大数定律可知，当$n\to\infty$ 时，$Y_n=\frac{1}{n}\sum_{k=1}^{n}X_k^2$ 依概率收敛于：$E\left(Y_n\right)=E\left(X_k^2\right)=\lambda^2+\lambda$ 。

§5.2 大数定律

小结：大数定律的意义

(1)Khintchin大数定律

这一定理表明：同一量X在相同条件下观测n次，当观测次数n充分大时，“观测值的算术平均值接近期望值”是一个大概率事件，即下式以大概率成立：

$$\frac{1}{n}\sum_{i=1}^{n} X_i \overset{n\text{充分大}}{\approx} E(X)$$

为寻找随机变量的期望值提供了一条实际可行的途径

§5.2 大数定律

(2)Bernoulli大数定律

这一定律表明：在相同条件下重复同一随机试验n次，当试验次数n充分大时，" 事件A发生的频率接近其概率" 是一个大概率事件，即下式以大概率成立：

$$f_A \overset{n\text{充分大}}{\approx} P(A)$$

寻找随机事件概率提供了一条实际可行的途径

§5.2 大数定律

大数定律的应用

案例5（用蒙特卡洛方法计算定积分）为计算积分$J=\int_a^b g(x)\mathrm{d}x$可以通过下面的概率方法实现：

任取一列相互独立的，都具有$[a,b]$中均匀分布的随机变量$\{\xi_i\}$，则$\{g(\xi_i)\}$也是一列相互独立同分布的随机变量，且

$$Eg(\xi_i)=\frac{1}{b-a}\int_a^b g(x)\mathrm{d}x=\frac{J}{b-a}$$

§5.2 大数定律

既然 $J=(b-a)\cdot Eg(\xi_i)$，因此只要能求得 $Eg(\xi_i)$，便能得到J 的数值。为求$Eg(\xi_i)$，使用大数定律，因为

$$\frac{g(\xi_1)+g(\xi_2)+\cdots+g(\xi_n)}{n}\xrightarrow{P}Eg(\xi_i)$$

只要能生成随机变量序列$\{g(\xi_i)\}$ 就能对前面的积分进行数值计算。而生成$\{g(\xi_i)\}$ 的关键是生成相互独立同分布的$\{\xi_i\}$，这里的ξ_i 服从$[a,b]$ 上的均匀分布。

§5.2 大数定律

现在已经可以把上述想法变成现实。这就是在电子计算机上产生服从均匀分布 $[a,b]$ 的随机数 $\{\xi_i\}$ 。

这种通过概率论的想法构造模型从而实现数值计算的方法，随着电子计算机的发展， 已形成一种新的计算方法——概率计算方法，亦称蒙特卡洛(Monte Carlo)方法。它在原子物理、公用事业理论中发挥了不少作用，这个方法的理论根据之一就是大数定律。

§5.2 大数定律

至于计算积分，蒙特卡洛方法的实用场合是计算重积分

$$I = \int_K g(P)\mathrm{d}P$$

其中P是m 维空间中的点。

§5.3 中心极限定理

中心极限定理的客观背景

自然界许多随机指标均服从或近似服正态分布

例 子弹和炮弹的弹着点

测量误差

一个班级的课程考试成绩

人的身高和体重

一个城市的日平均耗电量

农作物的产量

海浪的高度

……

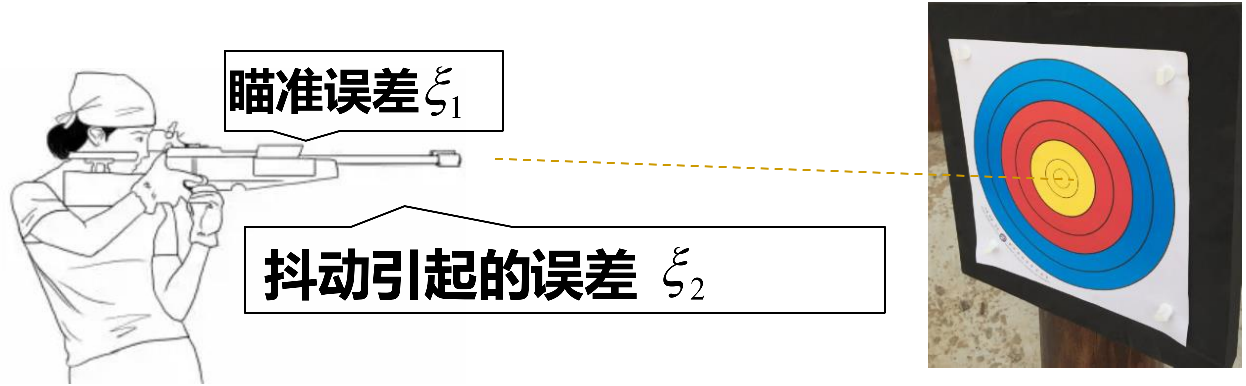
问题： 产生这一现象的原因是什么？

§5.3 中心极限定理

同时，在实际问题中，常需考虑许多随机因素所产生总影响。例如：炮弹射击的落点与目标的偏差就受着许多随机因素的影响。如空气阻力所产生的误差，瞄准时的误差，炮弹或炮身结构所引起的误差等等。对我们来说重要的是这些随机因素的总影响。

§5.3 中心极限定理

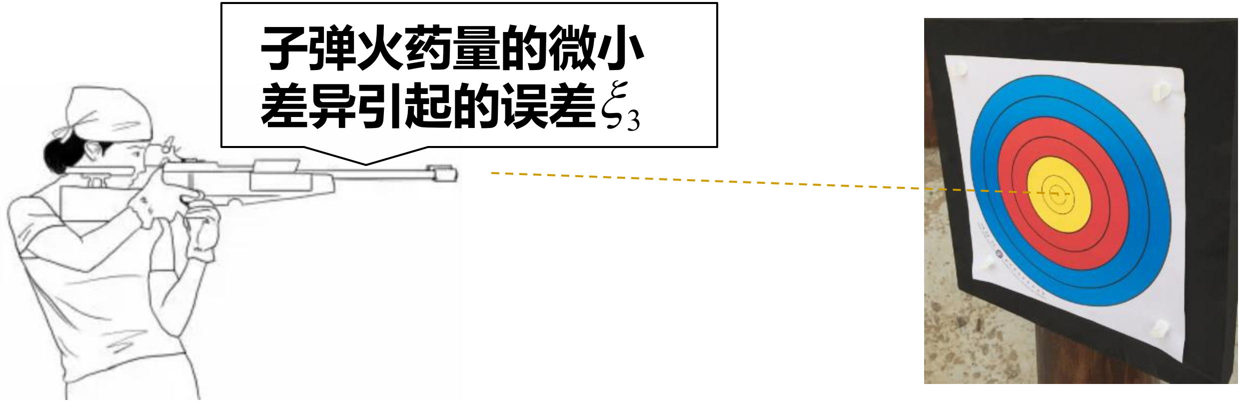
例: 步枪射击时, 子弹落点的横向偏差 η服从正态分布



$$\eta = \xi_1 + \xi_2$$

§5.3 中心极限定理

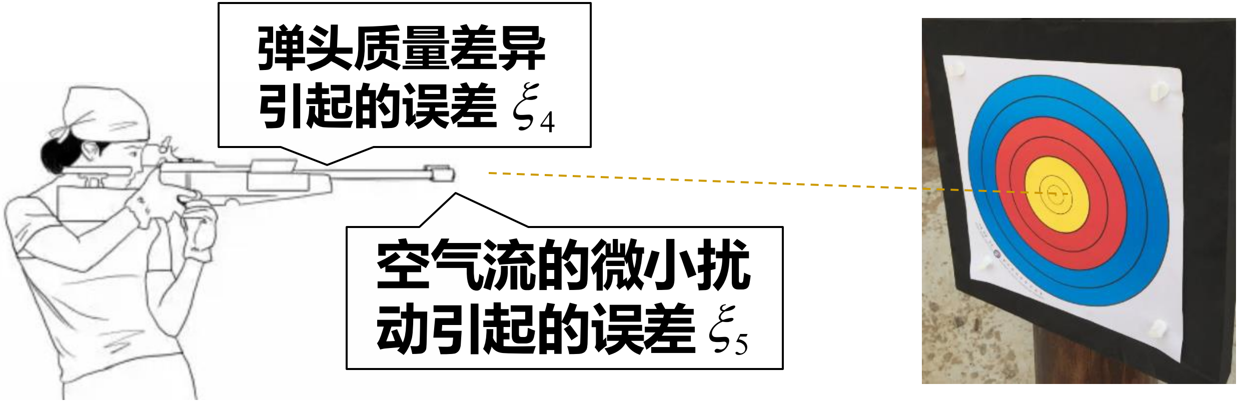
例: 步枪射击时, 子弹落点的横向偏差η服从正态分布



$$\eta = \xi_1 + \xi_2 + \xi_3$$

§5.3 中心极限定理

例: 步枪射击时, 子弹落点的横向偏差η服从正态分布



$$\eta = \xi_1 + \xi_2 + \xi_3 + \xi_4 + \xi_5$$

§5.3 中心极限定理

例: 步枪射击时, 子弹落点的横向偏差η服从正态分布

弹头外形上细小差别而导致空气阻力不同引起的误差ξ_6

独立随机变量之和

$$\eta = \xi_1 + \xi_2 + \xi_3 + \xi_4 + \xi_5 + \xi_6 + \cdots + \xi_n$$

问题: 当 $n \to \infty$ 时, 在什么情况下, $\eta_n = \sum_{i=1}^{n} \xi_i$ 的极限分布是正态分布?

§5.3 中心极限定理

5.3.1 独立同分布的中心极限定理

设随机变量$X_1,X_2,\ldots,X_n,\ldots$ 相互独立,服从同一分布,且

$E(X_k)=\mu$, $D(X_k)=\sigma^2\neq 0$ $(k=1,2,\ldots)$, 则 $Y_n=\dfrac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}$

的分布函数$F_n(x)$满足:对任意实数x, 有

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P\left\{\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\leq x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x).$$

定理表明, 当n充分大时, Y_n近似服从标准正态分布.

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\overset{\text{近似}}{\sim}N\left(\mu,\frac{\sigma^2}{n}\right)$$

§5.3 中心极限定理

表明：$\dfrac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \overset{a}{\sim} N(0,1)$，$n \to +\infty$。由正态分布的性质，

有 $\sum\limits_{i=1}^{n} X_i \overset{a}{\sim} N(n\mu, n\sigma^2)$。这就是说：当$n$充分大时，只要

$X_1, X_2, \cdots, X_n$独立同分布，无论他们服从什么分布，

一定有$\sum\limits_{i=1}^{n} X_i \overset{a}{\sim} N(n\mu, n\sigma^2)$

即：一个由许多独立同分布随机变量作用形成的随机变量，其概率分布一定是近似正态分布。

§5.3 中心极限定理

证明： 设$\mathbf{x}_n-\mu$的特征函数为$f(t)$，则由于

$$\eta_n=\frac{\sum\limits_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}=\frac{\sum\limits_{i=1}^{n}(X_i-\mu)}{\sqrt{n}\sigma}$$,则其特征函数为

$$f_n(t)=\left(f\left(\frac{t}{\sqrt{n}\sigma}\right)\right)^n$$,又由于 $\frac{f'(0)}{i}=E(X_n-\mu)=0$,

知$f'(0)=0-f''(0)+(f'(0))^2=D(X_n-\mu)=\sigma^2$

§5.3 中心极限定理

而$f(t)$Taylor展开为$f(t)=f(0)+f'(0)t+\dfrac{f''(0)}{2!}t^2$

$+o(t^2)=1-\dfrac{\sigma^2}{2}t^2+o(t^2)$所以，$\left[f\left(\dfrac{t}{\sigma\sqrt{t}}\right)\right]^n=\left[1-\dfrac{1}{2n}t^2+o\left(\dfrac{t^2}{n}\right)\right]^n\to e^{-t^2/2}$

由于$e^{-t^2/2}$是连续函数，它对应的分布函数为N(0,1)，因此由逆极限定理知

$$\lim_{n\to\infty}P\{\eta_n<x\}=\Phi(x)$$

证毕.

§5.3 中心极限定理

案例5 一盒同型号螺丝钉共100个,已知该型号的螺丝钉的重量是一个随机变量，期望值是100g，标准差是10g，求一盒螺丝钉的重量超过10.2kg的概率。

§5.3 中心极限定理

解 设X_i 为第 i个螺丝钉的重量， i=1, 2, ···, 100，则X_i相互独立同分布. 于是，一盒螺丝钉的重量为 $X=\sum_{i=1}^{100}X_i$

且 $\mu=E(X_i)=100,\sigma=\sqrt{D(X_i)}=10,i=1,2,\dots,100.$

由中心极限定理，

$$P\{\sum_{i=1}^{100}X_i>10200\}=P\left\{\frac{\sum_{i=1}^{100}X_i-100\mu}{\sigma\sqrt{100}}>\frac{10200-100\mu}{\sigma\sqrt{100}}\right\}$$

$$=P\left\{\frac{\sum_{i=1}^{100}X_i-10000}{100}>2\right\}\approx 1-\Phi(2)=1-0.97725=0.02275$$

§5.3 中心极限定理

案例6 根据以往经验，某种电器元件的寿命服从均值为100小时的指数分布，现随机抽取36只，设它们的寿命是相互独立的，求这36只元件寿命总和超过4000小时的概率

§5.3 中心极限定理

解 假设 X_i 表示第i只元件的寿命，i=1,2,3,...36。由题意知 $E(X_i)=100, D(X_i)=100^2$. 所求概率为：

$$P\{\sum_{i=1}^{36} X_i > 4000\} = P\left\{\frac{\sum_{i=1}^{36} X_i - 100\times 36}{\sqrt{36}\times 100} > \frac{4000-3600}{600}\right\}$$

$$\approx 1-\Phi(\frac{2}{3}) = 1-0.7486 = 0.2514$$

§5.3 中心极限定理

5.3.2 棣莫弗–拉普拉斯定理中心极限定理

设随机变量$\eta_n(n=1,2,\ldots)$服从参数为n，$p(0<p<1)$的二项分布，则对任意x，恒有

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

§5.3 中心极限定理

证明： η_n可以看成n个相互独立的服从同一(0–1)分布的随机变量$X_1,\dots,X_n$之和，即 $\eta_n = X_1 + X_2 + \cdots + X_n$

$$E(X_i) = p, D(X_i) = p(1-p), i = 1,2,\cdots,n$$

由定理1知，

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

此定理表明，**正态分布是二项分布的极限分布**，所以当n充分大时，我们可以用正态分布近似二项分布.

§5.3 中心极限定理

案例7 某车间有200台车床独立工作,设每台车床的开工率为0.6,开工时耗电1千瓦，问供电所至少要供多少电才能以不小于99.9%的概率保证该车间不会因供电不足而影响生产?

§5.3 中心极限定理

解 记X为200台车床中工作着的车床台数，则$X\sim b(200, 0.6)$.

按题意，要求最小的 k，使P$\{X\leq k\}\geq 0.999$

由定理2 $P\{X\leq k\}=P\left\{\frac{X\text{-}200\times 0.6}{\sqrt{200\times 0.6\times 0.4}}\leq\frac{k-200\times 0.6}{\sqrt{200\times 0.6\times 0.4}}\right\}$

$$=P\left\{\frac{X\text{-}120}{\sqrt{48}}\leq\frac{k-120}{\sqrt{48}}\right\}\approx\Phi\left(\frac{k-120}{\sqrt{48}}\right)\geq 0.999$$

$$\frac{k-120}{\sqrt{48}}\geq 3.1 \quad k\geq 141.48,$$

至少供电142千瓦,才能保证车间以不小于99.9%的概率正常工作。

§5.3 中心极限定理

案例8 售报员在报摊上卖报,已知每个过路人在报摊上买报的概率为$1/3$. 令 X 是出售了100份报时过路人的数目,求$P\{280 < X \leq 320\}$.

§5.3 中心极限定理

解： 令X_i表示售出了第$i-1$份报纸后到售出第i份报纸时的过路人数, $i=1,2,\cdots,100$, 则 $P\{X_i=k\}=p(1-p)^{k-1}\big|_{p=1/3}$,

$$E(X_i)=\frac{1}{p}\bigg|_{p=1/3}=3,\quad D(X_i)=\frac{1-p}{p^2}\bigg|_{p=1/3}=6$$

因为$X_1,X_2,\cdots,X_{100}$相互独立, $X=\sum_{k=1}^{100}X_k$, $E(X)=300$,

$D(X)=600$ 所以

$$P\{280<X\le 320\}\approx\Phi\left(\frac{320-300}{\sqrt{600}}\right)-\Phi\left(\frac{280-300}{\sqrt{600}}\right)=2\Phi\left(\frac{20}{\sqrt{600}}\right)-1=0.5858$$

§5.3 中心极限定理

小结： 中心极限定理的意义

（1）Lindeberg-Levy中心极限定理

对于独立同分布随机变量序列 $\{X_n\}$,不管他们服从什么分布，只要存在有限数学期望和方差，当 n充分大时，就有

$$\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2)$$

所以，$\sum_{i=1}^{n} X_i$的有关概率问题可利用正态分布求解。

§5.3 中心极限定理

（2） De Moivre-Laplace中心极限定理

对于随机变量$X \sim B(n,p)$，总有$X \overset{n\to+\infty}{\sim} N(np,npq)$，因此，当$n$充分大时，二项分布的概率问题可利用正态分布解决。

§5.3 中心极限定理

中心极限定理的应用

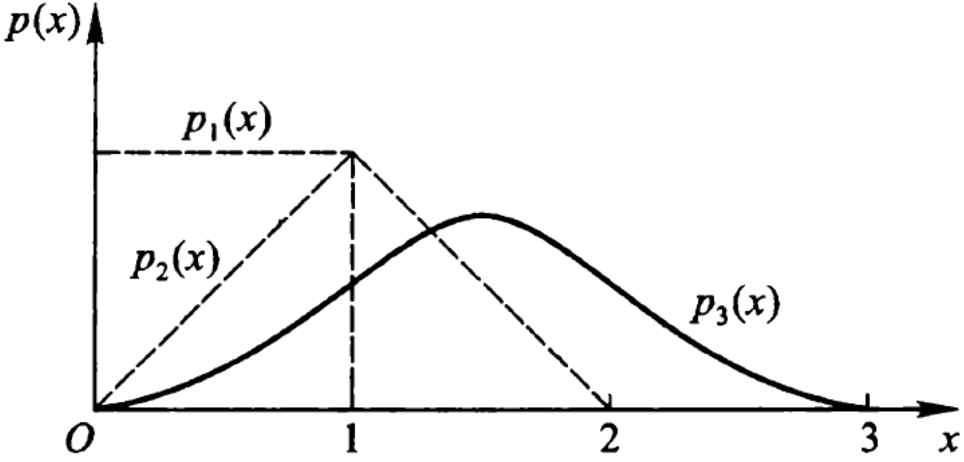
Lindeberg-Levy（林德伯格–列维）中心极限定理有着广泛应用。在实际工作中，只要 n 足够大，便可以把独立同分布的随机变量之和当作是正态变量。此做法在数理统计中用得很普遍。

例（正态随机数的产生）

设 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是相互独立、均服从[0, 1]均匀分布的随机变量，这时Lindeberg-Levy中心极限定理的条件得到满足，故 $\xi_1 + \xi_2 + \cdots + \xi_n + \cdots$ 渐进于正态变量。

§5.3 中心极限定理

一般n 取不太大的值就可满足实际要求，在蒙特卡洛方法中，一般取n=12。



均匀分布卷积

§5.3 中心极限定理

◆在二项分布计算中的应用

由积分极限定理, 当p不太接近于0或1,而n又不太小时,对二项分布的近似计算有下面的公式:

$$P\{k_1 \le \mu_n \le k_2\} = P\left\{\frac{k_1 - np}{\sqrt{npq}} \le \frac{\mu_n - np}{\sqrt{npq}} \le \frac{k_2 - np}{\sqrt{npq}}\right\}$$

$$\approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right)$$

§5.3 中心极限定理

实际计算中，往往用下面的修正公式计算效果更好。

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

§5.3 中心极限定理

案例9（近似数定点运算的误差分析）

数值计算时，任何数x都只能用一定位数的有限小数y来近似，这就产生了一个误差$\xi = x - y$。

在下面的讨论中，我们假定参加运算的数都用十进制定点表示，每个数都用四舍五入的方法取到小数点后五位，这时相应的舍入误差可以看作是 $[-0.5\times10^{-5}, 0.5\times10^{-5})$ 上的均匀分布。

§5.3 中心极限定理

现在如果要求 n个数$x_i(i=1,2,\mathrm{L}\ ,n)$ 的和S，在数值计算中就只能求出相应的有限位小数 $y_i(i=1,\mathrm{L}\ ,n)$ 的和T，并用T作为S的近似值。

下面计算这样做造成的误差$\eta=S-T$是多少？因为

$$S=\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}(y_i+\xi_i)=\sum_{i=1}^{n}y_i+\sum_{i=1}^{n}\xi_i \quad \text{故} \quad \eta=\sum_{i=1}^{n}\xi_i$$

§5.3 中心极限定理

一种传统的估计方法是这样的：由于 $|\xi_i| \le 0.5\times10^{-5}$

所以 $|\eta| \le \sum_{i=1}^{n} |\xi_i| \le n\times0.5\times10^{-5}$ 以 $n=10000$ 为例，

所得的误差估计为 $|\eta| \le 0.05$

这种估计方法显然太保守，下用概率论方法估计。这时直接求 $\eta=\sum_{i=1}^{n}\xi_i$ 的分布不容易，但当n较大时用极限定理作为工具，则能使问题很快得到解决。因为

$$\mu = E\xi_i = 0, \sigma = \sqrt{D\xi_i} = \sqrt{\frac{(1\times10^{-5})^2}{12}} = \frac{0.5\times10^{-5}}{\sqrt{3}}.$$

§5.3 中心极限定理

如果假定舍入误差 ξ_i 是相互独立的，n又较大，那么根据Lindeberg-Levy中心极限定理知：

$$P\left\{\left|\sum_{i=1}^{n}\xi_i\right|<k\sqrt{n}\sigma\right\}\approx\frac{1}{\sqrt{2\pi}}\int_{-k}^{k}e^{-t^2/2}\mathrm{d}t$$

取 $k=3$ 时上式右边为0.997，因此我们能以 99.7%的概率断言：

$$|\eta|<3\times100\times\frac{0.5\times10^{-5}}{\sqrt{3}}=0.866\times10^{-3}.$$

这仅仅是传统估计法中误差上限的60分之一。

谢谢大家！