XI'AN JIAOTONG UNIVERSITY

概率统计与随机过程

第四章：大数定律与中心极限定理

赵俊舟

junzhou.zhao@xjtu.edu.cn

2025 年 4 月 9 日

目录

目录

目录

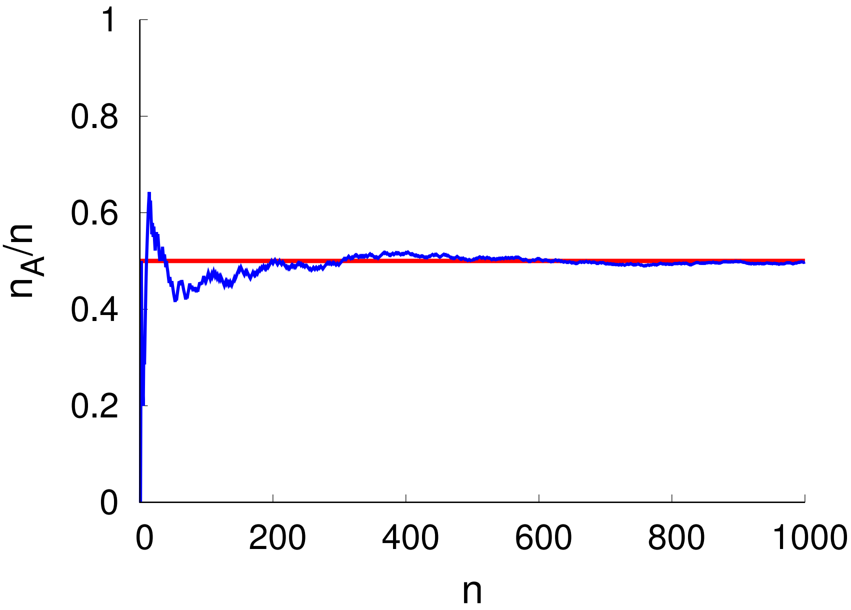
抛硬币试验

- **抛硬币试验**：连续抛一枚硬币 n 次，记 $A=\{正面朝上\}$，n_A 表示 n 次试验中事件 A 发生的次数，则事件 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	n_A/n
德·摩根	2048	1061	0.5181
蒲丰	4048	2048	0.5069
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005
罗曼诺夫斯基	80640	39699	0.4923

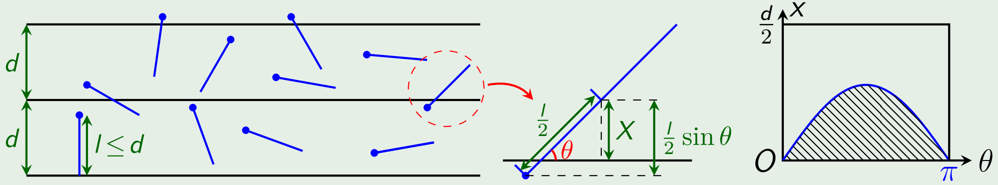
- 可见 $f_n(A)=\dfrac{n_A}{n}\to P(A)=0.5,\quad n\to\infty$

抛硬币试验中的统计规律性



蒲丰投针试验

例 (蒲丰投针试验，1777 年)



记 $A=\{$针与平行线相交$\}$，求 $P(A)$。

- 令针与平行线的夹角为 θ，针的中点与离它比较近的平行线的距离为 x，则 $\theta\sim U(0,\pi), X\sim U(0,d/2)$
- 针与平行线相交的条件为 $X<\frac{l}{2}\sin\theta$

$$P(A)=P(X<\frac{l}{2}\sin\theta)=\int_0^{\pi}\mathrm{d}\theta\int_0^{\frac{l}{2}\sin\theta}\frac{2}{\pi d}\,\mathrm{d}x=\frac{2l}{\pi d}$$

蒲丰投针试验

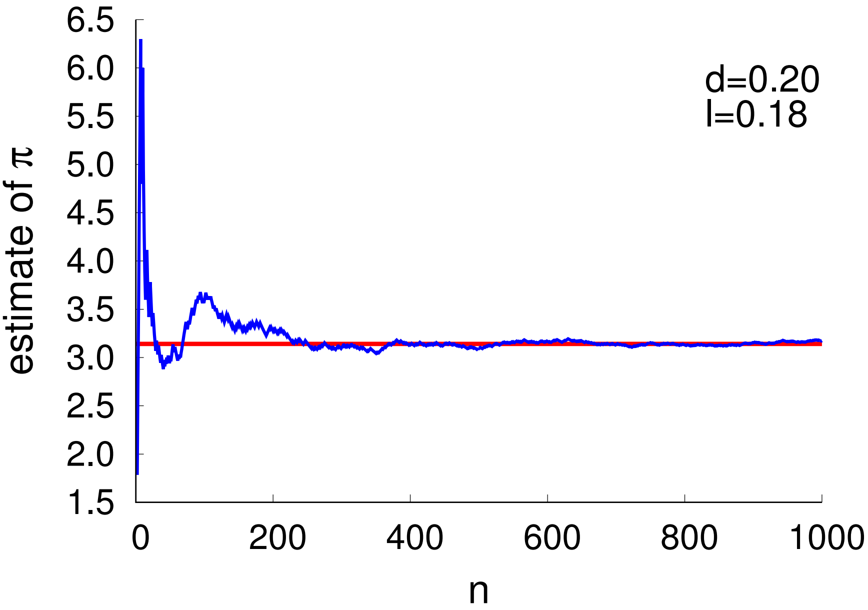
- 记投针总数为 n，针与平行线相交次数为 n_A，则

$$\frac{n_A}{n} \to \frac{2l}{\pi d} \quad \Rightarrow \quad \frac{2nl}{n_A d} \to \pi$$

- 19-20 世纪几个有名的“投针”试验

试验者	时间	针长 d	n	n_A	$\hat{\pi}$
Buffon	1777	0.5	2212	704	3.142
Wolf	1850	0.8	5000	2532	3.1596
Smith	1855	0.6	3204	1219	3.1554
De.morgan	1860	1	600	383	3.137
Fox	1884	0.75	1030	489	3.1595
Lazzerini	1901	0.83	3408	1801	3.14159
Reina	1925	0.54	2520	859	3.1683

蒲丰投针试验中的统计规律性



概率与事件发生频率的稳定值

随机试验 → 统计数据 → 统计规律性 → 频率稳定性 → **概率**

- 在大量重复试验中，事件发生的频率稳定于它发生的概率，即**频率稳定于概率**。
- **频率的稳定性**：设 n 次独立重复试验中事件 A 发生的次数为 n_A，则当 n 越来越大时，有

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

问题

1. 频率稳定于概率，但是“稳定”的严格数学定义是什么？
2. 如何定义随机变量的收敛性？

目录

切比雪夫不等式

定理 (切比雪夫不等式)

设随机变量 X 的数学期望 $\mathbb{E}(X)$ 和方差 $D(X)$ 都存在，则对任意常数 $\epsilon > 0$，有

$$P(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{D(X)}{\epsilon^2}$$

或

$$P(|X - \mathbb{E}(X)| < \epsilon) \geq 1 - \frac{D(X)}{\epsilon^2}$$

- 如果随机变量 X 的方差 $D(X)$ 越小，则随机变量 X 落入区间 $(\mathbb{E}(X) - \epsilon, \mathbb{E}(X) + \epsilon)$ 的概率就越大；
- 切比雪夫不等式表明，方差是反映随机变量取值集中在 $\mathbb{E}(X)$ 附近的程度的数量指标。

切比雪夫不等式的证明

引理 (Markov 不等式)

对随机变量 X，若 $\mathbb{E}\left(|X|^r\right) < \infty, r > 0$，则对任意 $\epsilon > 0$，有

$$P(|X| \geq \epsilon) \leq \frac{\mathbb{E}\left(|X|^r\right)}{\epsilon^r}$$

证明.

在 Markov 不等式中，以 $X - \mathbb{E}(X)$ 代替 X，并令 $r = 2$，则对任意正数 ϵ，有

$$P(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\mathbb{E}\left([X - \mathbb{E}(X)]^2\right)}{\epsilon^2} = \frac{D(X)}{\epsilon^2}$$

□

举例：切比雪夫不等式

例 (估计及格率的下界)

若某班考试平均成绩是 75 分，方差为 10，估计及格率下界。

- 设随机变量 X 表示学生成绩，则 $\mathbb{E}(X) = 75, D(X) = 10$。
- 于是

$$\begin{aligned} P(60 \le X \le 100) &= P(-15 \le X - 75 \le 25) \\ &\ge P(|X - 75| < 15) \ge 1 - \frac{10}{225} = 0.956 \end{aligned}$$

- 表明及格率至少为 95.6%。
- 若 $X \sim N(75, 10)$，则实际概率为

$$P(60 \le X \le 100) = P(\frac{-15}{\sqrt{10}} \le \frac{X-75}{\sqrt{10}} \le \frac{25}{\sqrt{10}}) \approx 1$$

切比雪夫不等式

- 切比雪夫不等式所得概率下界可能与实际概率相差较大。
- 切比雪夫不等式经常作为理论工具，用于证明某些定理。

目录

随机变量序列的收敛性

- 目标：研究随机变量序列 $\{X_n : n = 1, 2, \ldots\}$ 的收敛性。
- 高等数学中，级数 $\{x_n : n = 1, 2, \ldots\}$ 的收敛性可以定义为：
$$\lim_{n\to\infty} x_n = x$$
- 能否直接将级数的收敛性推广到随机变量序列的收敛性？

随机变量序列的收敛性

例

研究随机变量序列 $\{X_n : n = 1, 2, \ldots\}$。设 $X_n \sim N(0, 1)$，能否说 X_n 收敛到 $X \sim N(0, 1)$，即 $\lim_{n\to\infty} X_n = X$？

不能！因为对于连续型随机变量，$P(X_n = X) = 0, \forall n$。

例

研究随机变量序列 $\{X_n : n = 1, 2, \ldots\}$。设 $X_n \sim N(0, 1/n)$，当 n 很大时，X_n 集中到 0，能否说 X_n 收敛到 0，即 $\lim_{n\to\infty} X_n = 0$？

不能！因为对于连续型随机变量，$P(X_n = 0) = 0, \forall n$。

需要提出新的数学工具，能够适用于研究随机变量序列的收敛性！

依概率收敛

定义 (依概率收敛)

设 $\{X_n: n = 1, 2, \ldots\}$ 为随机变量序列，X 是另一个随机变量，如果对任意 $\epsilon > 0$ 有

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

或

$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$$

则称随机变量序列 $\{X_n\}$ 依概率收敛于 X，记作 $X_n \xrightarrow{p} X$。

特别地，若 X 为单点分布，即 $P(X = a) = 1$，则称随机变量序列 $\{X_n\}$ 依概率收敛于常数 a，记作 $X_n \xrightarrow{p} a$。

依概率收敛

定理

如果 $X_n \xrightarrow{p} a, Y_n \xrightarrow{p} b$，又函数 $g(x,y)$ 在点 (a,b) 处连续，则 $g(X_n, Y_n) \xrightarrow{p} g(a,b)$。

特例，如果 $X_n \xrightarrow{p} a, Y_n \xrightarrow{p} b$，则:

$$X_n \pm Y_n \xrightarrow{p} a \pm b$$

$$X_n Y_n \xrightarrow{p} ab$$

$$\frac{X_n}{Y_n} \xrightarrow{p} \frac{a}{b} \qquad (Y_n, b \neq 0)$$

依分布收敛

定义 (依分布收敛)

设 $\{X_n\}$ 为随机变量序列，其分布函数序列为 $\{F_n(x)\}$。X 是随机变量，其分布函数为 $F(x)$。若对 $F(x)$ 的每个连续点 x，有

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

则称**随机变量序列 $\{X_n\}$ 依分布收敛于 X**，记作 $X_n \xrightarrow{d} X$，或称**分布函数序列 $\{F_n(x)\}$ 弱收敛于 $F(x)$**，记作 $F_n(x) \xrightarrow{w} F(x)$。

- 依分布收敛是最宽松的收敛方式之一；
- 一个依概率收敛的随机变量序列必然也依分布收敛到同一个极限，即**依概率收敛强于依分布收敛**；
- 依分布收敛蕴含依概率收敛当且仅当依分布收敛的极限是一个常数。

举例：随机变量序列的收敛性

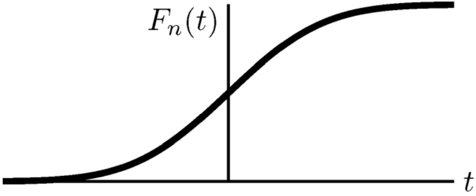
例 (随机变量序列的收敛性分析)

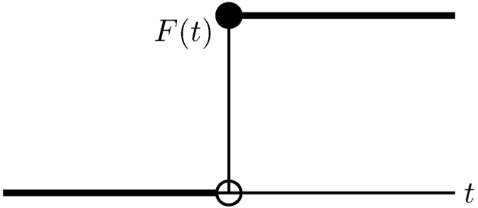
设随机变量 $X_n \sim N(0, 1/n)$，分析 $\{X_n\}$ 的收敛性。

- 对任意 $\epsilon > 0$，由切比雪夫不等式，有
$$P(|X_n| \geq \epsilon) \leq \frac{D(X_n)}{\epsilon^2} = \frac{1}{n\epsilon^2} \to 0$$
- 所以 $X_n \xrightarrow{p} 0$
- 注意到 $\sqrt{n}X_n \sim N(0,1)$，令 $Z \triangleq \sqrt{n}X_n$，则 $Z \sim N(0,1)$。
- 当 $t < 0$ 时，
$$F_n(t) = P(X_n < t) = P(\sqrt{n}X_n < \sqrt{n}t) = P(Z < \sqrt{n}t) \to 0$$
- 当 $t > 0$ 时，
$$F_n(t) = P(X_n < t) = P(Z < \sqrt{n}t) \to 1$$

举例：随机变量序列的收敛性

- 令 $F(t)$ 为单点分布 $P(X=0)=1$ 的分布函数，则对于任意 $t\neq 0$，有 $\lim_{n\to\infty}F_n(t)=F(t)$。





- 说明 $X_n\xrightarrow{d}0$
- 注意 $F_n(0)=1/2\neq F(0)=1$，所以在 $t=0$ 处不收敛，但不影响 X_n 的依分布收敛性，因为 $t=0$ 不是 F 的连续点。

举例：随机变量序列的收敛性

例 (依概率收敛性与期望)

设随机变量 X_n 的分布律为

x	0	n^2
p	$1-1/n$	$1/n$

分析 $\{X_n\}$ 的依概率收敛性及 X_n 的数学期望的极限。

- 因为对任意 $\epsilon > 0$，有
$$P(|X_n| < \epsilon) = P(X_n = 0) = 1 - \frac{1}{n} \to 1$$
- 所以 $X_n \xrightarrow{p} 0$。
- 但是 $\mathbb{E}(X_n) = n^2 \times 1/n = n \neq 0$，且 $\lim_{n\to\infty} \mathbb{E}(X_n) = \infty$。

目录

目录

伯努利大数定律

定理（伯努利大数定律）

设 $\eta_n \sim B(n,p)$，则对任意常数 $\epsilon > 0$，有

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

证明.

因为 $\mathbb{E}(\eta_n) = np$，$D(\eta_n) = np(1-p)$，由切比雪夫不等式，对任意 $\epsilon > 0$，有

$$P\left(\left|\frac{\eta_n}{n} - p\right| < \epsilon\right) = P(|\eta_n - np| < n\epsilon) \geq 1 - \frac{np(1-p)}{(n\epsilon)^2} = 1 - \frac{p(1-p)}{n\epsilon^2}$$

令 $n \to \infty$，得

$$1 \geq \lim_{n\to\infty} P\left(\left|\frac{\eta_n}{n} - p\right| < \epsilon\right) \geq \lim_{n\to\infty}(1 - \frac{p(1-p)}{n\epsilon^2}) = 1$$

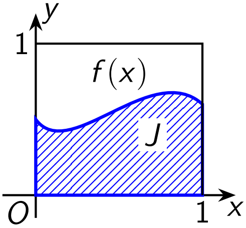
伯努利大数定律

- 伯努利大数定律是概率论历史上最早出现的大数定律，由瑞士数学家雅各布·伯努利在著作《猜度术》中首次提出。
- 伯努利大数定律描述了当试验次数 $n \to \infty$ 时，频率 η_n/n 的极限状态：事件发生的频率 η_n/n 依概率收敛于概率 p，即 $\dfrac{\eta_n}{n} \xrightarrow{P} p$，或者说频率依概率收敛于概率。
- 伯努利大数定律提供了用频率来确定概率的理论依据。
- 伯努利大数定律也是蒙特卡洛（Monte Carlo）方法的数学理论基础。

举例：伯努利大数定律

例 (蒙特卡洛方法计算定积分，也称为随机投点法)

设 $f(x) \in [0,1]$，求 $f(x)$ 在区间 $[0,1]$ 上的积分值 $\int_0^1 f(x)\,\mathrm{d}x$。



- 设 (X, Y) 服从正方形区域 $[0,1]\times[0,1]$ 上的均匀分布，即 $X \sim U(0,1), Y \sim U(0,1)$
- 记事件 $A = \{Y \le f(X)\}$，则事件 A 发生的概率为

$$p = P(Y \le f(X)\} = \int_0^1 \int_0^{f(x)} \mathrm{d}y\,\mathrm{d}x = \int_0^1 f(x)\,\mathrm{d}x \triangleq J$$

- 即定积分的值 J 就是事件 A 发生的概率。
- 由伯努利大数定律，可以用重复试验中事件 A 发生的频率作为 p 的估计值。

举例：伯努利大数定律

随机投点法

1. 用计算机产生 $[0,1]$ 上的均匀分布的 $2n$ 个随机数：x_i, y_i，$i = 1, \ldots, n$，例如 $n = 10^5$。
2. 对 n 对数据 $(x_i, y_i), i = 1, \ldots, n$，记录满足 $y_i \le f(x_i)$ 的次数，这就是事件 A 发生的频数 η_n，则事件 A 发生的频率 $\eta_n/n \approx J$。

例如计算 $\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^1 e^{-x^2/2}\,\mathrm{d}x$，其精确值和 $n = 10^4$，$n = 10^5$ 时的近似值如下

精确值	$n = 10^4$	$n = 10^5$
0.341344	0.340698	0.341355

举例：伯努利大数定律

- 对于一般区间 $[a, b]$ 上的定积分 $J' = \int_a^b g(x)\,\mathrm{d}x$，可以通过线性变换 $y = (x-a)/(b-a)$ 转化成 $[0, 1]$ 区间上的积分。
- 进一步，若 $c \le g(x) \le d$，可令
$$f(y) \triangleq \frac{g(a+(b-a)y)-c}{d-c}$$
则 $0 \le f(y) \le 1$。
- 此时有
$$J' = \int_a^b g(x)\,\mathrm{d}x = S_0 \int_0^1 f(y)\,\mathrm{d}y + c(b-a)$$
其中 $S_0 = (b-a)(d-c)$。
- 说明用蒙特卡罗方法计算定积分具有普遍意义。

大数定律的一般形式

- 伯努利大数定律讨论的是一个独立同分布随机变量序列 $\{X_n\}$，每个随机变量 X_i 都服从两点分布，即 $X_i \sim B(1,p)$。
- 考察该序列前 n 个随机变量之和 $\eta_n = \sum_{i=1}^{n} X_i$，其频率和频率的数学期望分别为

$$\frac{\eta_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i\right) = p$$

- 那么伯努利大数定律可表述为：对任意 $\epsilon > 0$，有

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

- 上式为大数定律的一般表达形式，不同大数定律的区别只是针对不同的随机变量序列。

目录

切比雪夫大数定律

定理 (切比雪夫大数定律)

设 $\{X_n\}$ 是两两不相关的随机变量序列，若每个 X_i 的方差存在，且有共同上界，即 $D(X_i) \leq c, i = 1, \dots, n$，则对任意 $\epsilon > 0$，都有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(X_i\right)\right| \geq \epsilon\right) = 0$$

或

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

注意：切比雪夫大数定律只要求 $\{X_n\}$ 两两不相关，并不要求它们独立同分布。

切比雪夫大数定律

证明.

- 令 $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$，则 $\mathbb{E}\left(Y_n\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i\right)$。
- 由于 $\{X_n\}$ 两两不相关，则有

$$D(Y_n) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) \leq \frac{c}{n}$$

- 对任意 $\epsilon > 0$，应用切比雪夫不等式

$$1 \geq P(|Y_n - \mathbb{E}\left(Y_n\right)| < \epsilon) \geq 1 - \frac{D(Y_n)}{\epsilon^2} \geq 1 - \frac{c}{n\epsilon^2}$$

- 令 $n \to \infty$，由夹逼法则得

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

切比雪夫大数定律

推论

设 $\{X_n\}$ 是相互独立的随机变量序列，且每个 X_i 有相同的期望 $\mathbb{E}(X_i)=\mu$ 和方差 $D(X_i)=\sigma^2$，则对任意 $\epsilon>0$，都有

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

- 当 n 很大时，事件 $\left|\frac{1}{n}\sum_{i=1}^{n} X_i-\mu\right|<\epsilon$ 发生的概率接近 1。
- 称概率接近于 1 的事件为大概率事件，称概率接近于 0 的事件为小概率事件。
- 实践发现，大概率事件在一次试验中几乎肯定要发生，而小概率事件在一次试验中几乎不可能发生，称为实际推断原理。

实际推断原理与数学期望的统计估计

- 设 X 是一个概率分布未知的随机变量，希望估计其数学期望 $\mathbb{E}(X)$。
- 对 X 进行一次 n 重观测试验，记 X_i 为第 i 次观测的结果，$i = 1, \ldots, n$，X_i 为随机变量且彼此相互独立。
- 当 n 很大时，$\left| \dfrac{1}{n} \sum\limits_{i=1}^{n} X_i - \mathbb{E}(X) \right| < \epsilon$ 为大概率事件。
- 因此对一次 n 重观测试验的结果 $x_1, \ldots, x_n$，几乎可以肯定有 $\left| \dfrac{1}{n} \sum\limits_{i=1}^{n} x_i - \mathbb{E}(X) \right| < \epsilon$，所以 $\mathbb{E}(X) \approx \dfrac{1}{n} \sum\limits_{i=1}^{n} x_i$。
- 以上方法称为数学期望的统计估计法。

弱大数定律

上述推论的一个特殊情况是随机变量序列独立同分布。

定理 (弱大数定律 The Weak Law of Large Numbers, WLLN)

设 $\{X_n\}$ 是独立同分布的随机变量序列，且存在数学期望 $\mathbb{E}(X_i) = \mu$ 与方差 $D(X_i) = \sigma^2, i = 1, 2, \ldots$，则对任意 $\epsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) = 0$$

或

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

目录

辛钦大数定律

- 注意：以上大数定律都要求方差存在（方差存在则数学期望必然存在，但反之不一定成立）。
- 辛钦大数定律去掉了方差存在的假设，仅要求数学期望存在。

定理（辛钦大数定律）

设 $\{X_n\}$ 是独立同分布的随机变量序列，若每个随机变量 X_i 的数学期望 $\mathbb{E}(X_i) = \mu$ 存在，则对任意 $\epsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) = 0$$

或

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

举例：辛钦大数定律

例

设 $\{X_n\}$ 是独立同分布的随机变量序列，其分布律为

$$P(X_k=(-1)^{k-1}k)=\frac{6}{\pi^2k^2}\quad k=1,2,\ldots$$

问 $\{X_n\}$ 是否依概率收敛于其数学期望？

由于级数

$$\sum_{k=1}^{\infty}|(-1)^{k-1}k|\frac{6}{\pi^2k^2}=\frac{6}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k}$$

不绝对收敛，故 X_k 的数学期望不存在，从而 $\{X_n\}$ 不依概率收敛于期望。

目录

目录

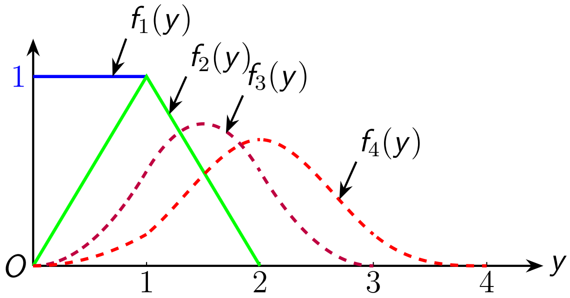
随机变量和的极限

- 大数定律讨论的是多个随机变量的平均 $\frac{1}{n}\sum_{i=1}^{n} X_i$ 的渐进性质，那么独立随机变量和 $\sum_{i=1}^{n} X_i$ 又有怎样的性质？
- 例如，机床加工工件时总有误差，导致误差的随机因素包括：
  - 机床振动与转速、刀具装配与磨损、钢材成分与产地、
  - 操作者注意力集中程度、当天的情绪、
  - 量具误差、测量技术、环境温度、湿度、照明、电压等。
- 误差由大量微小的相互独立的随机因素叠加而成，每个因素的出现无法控制、随机、时有时无、时大时小、时正时负。
- 记误差为 Y_n，那么 Y_n 是很多微小随机波动 $X_1, \ldots, X_n$ 之和
$$Y_n = X_1 + X_2 + \cdots + X_n,$$
人们关心当 $n \to \infty$ 时，Y_n 的分布是怎样的？

均匀分布和的极限

例 (独立同均匀分布随机变量和的极限)

设 $\{X_n\}$ 为独立同分布随机变量序列，且 $X_i \sim U(0,1)$。记 $Y_n = \sum_{i=1}^n X_i$，利用卷积公式可以求出 Y_n 的概率密度 $f_n(y)$。

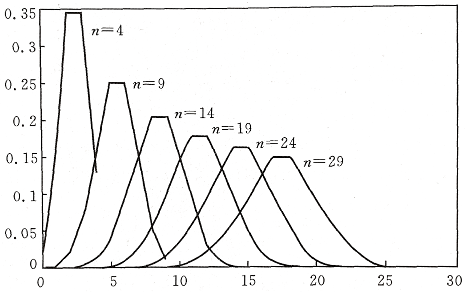


随着 n 的增大，$f_n(y)$ 愈来愈平滑，愈来愈接近正态分布的概率密度曲线。同时，曲线也在右移，方差变大。

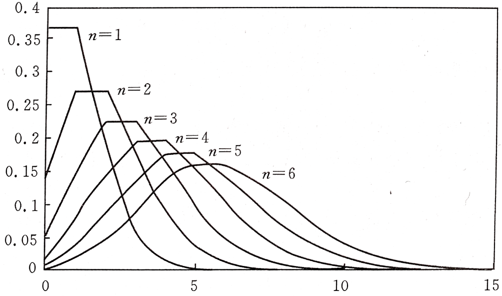
其他分布和的极限

- 当 X_i 服从其他分布时，$Y_n = \sum_{i=1}^{n} X_i$ 的分布也观察到类似的现象。

$X_i \sim B(1, 0.6)$



$X_i \sim P(1)$



随着 n 的增大，$f_n(y)$ 愈来愈接近正态分布的概率密度曲线的同时，曲线也在右移，方差变大。

目录

独立同分布随机变量和的正态近似

- 为避免当 $n \to \infty$ 时 Y_n 的均值和方差趋于无穷大，对 Y_n 进行标准化：$Y_n^* \triangleq \dfrac{Y_n - \mathbb{E}(Y_n)}{\sqrt{D(Y_n)}}$，使 $\mathbb{E}(Y_n^*) = 0, D(Y_n^*) = 1$。

定理 (林德贝格-勒维中心极限定理)

设 $\{X_n\}$ 是独立同分布的随机变量序列，且 $\mathbb{E}(X_n) = \mu$，$D(X_n) = \sigma^2 > 0$，则对任意 $y \in \mathbb{R}$，有

$$\lim_{n\to+\infty} P(Y_n^* \leq y) = \Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-t^2/2}\,\mathrm{d}t$$

- 无论随机变量序列 $\{X_n\}$ 服从什么分布，只要独立同分布且存在期望和方差，则 Y_n^* 依分布收敛于标准正态分布 $N(0,1)$。

举例：独立同分布随机变量和的正态近似

例 (元件寿命)

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，设它们的寿命是独立同分布的，求这 36 只元件的寿命总和超过 4000 小时的概率。

- 设 X_i 表示第 i 只元件的寿命，$i = 1, \ldots, 36$，则 $\mathbb{E}(X_i) = 100$，$D(X_i) = 100^2$。
- 这 36 只元件的总寿命为 $\sum_{i=1}^{36} X_i$，所求概率为

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

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

举例：独立同分布随机变量和的正态近似

例 (测量误差)

多次测量一个物理量，每次都产生一个随机误差且服从 $U(-0.5, 0.5)$。问：（1）100 次测量的算术平均值与真实值的绝对值小于 0.05 的概率是多少？(2) 需要进行多少次测量才能使测量的算术均值与其真值的差小于 0.05 的概率不小于 0.95？

- 设 a 为真值，X_i 表示第 i 次测量值，ϵ_i 表示第 i 次测量误差。
- 所以 $X_i = a + \epsilon_i$，$\epsilon_i \sim U(-0.5, 0.5)$，$\mathbb{E}(\epsilon_i) = 0$，$D(\epsilon_i) = 1/12$。

$$P\left(\left|\frac{1}{100}\sum_{i=1}^{100} X_i - a\right| < 0.05\right) = P\left(\left|\sum_{i=1}^{100} X_i - 100a\right| < 5\right)$$

$$= P\left(\left|\sum_{i=1}^{100}\epsilon_i\right| < 5\right) = P\left(\left|\frac{\sum_{i=1}^{100}\epsilon_i}{\sqrt{100/12}}\right| < \frac{5\sqrt{12}}{10}\right) = P\left(|Z| < \sqrt{3}\right)$$

举例：独立同分布随机变量和的正态近似

$$= \Phi(\sqrt{3}) - \Phi(-\sqrt{3}) = 2\Phi(\sqrt{3}) - 1 = 0.9164$$

- 设需要测量 n 次，测量的算术均值与真值差小于 0.05 的概率为

$$\begin{aligned} P\left(\frac{1}{n}\sum_{i=1}^{n} X_i - a < 0.05\right) &= P\left(\sum_{i=1}^{n} X_i - na < 0.05n\right) \\ &= P\left(\frac{\sum_{i=1}^{n}\epsilon_i}{\sqrt{n/12}} < \frac{0.05n}{\sqrt{n/12}}\right) \\ &\approx \Phi(0.1\sqrt{3n}) \end{aligned}$$

- 令 $\Phi(0.1\sqrt{3n}) \geq 0.95$，得 $n \geq 90.75$

二项分布的正态近似

定理 (德莫弗-拉普拉斯极限定理)

设 n 重伯努利试验中，事件 A 在每次试验中出现的概率为 $p \in (0,1)$。记 η_n 为 n 次试验中事件 A 出现的次数，且记

$$Y_n^* = \frac{\eta_n - np}{\sqrt{np(1-p)}}$$

则对任意实数 y, 有

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

- 二项分布的极限分布是正态分布，故也称为“二项分布的正态近似”。当 n 较大时，η_n 可用 $N(np, np(1-p))$ 来近似。
- 泊松定理同时给出了二项分布的泊松近似。一般 p 较小 np 适中时，泊松近似较好；在 $np > 5$ 时，正态近似较好。

举例：二项分布的正态近似

例 (性状遗传)

根据孟德尔遗传理论，红、黄两种番茄杂交的第二代结红果植株和黄果植株的比率为 $3:1$，现在种植番茄杂交种 400 株，试求黄果植株介于 $83 \sim 117$ 的概率。

- 令 X 表示结黄果的植株数，由于结红、黄果彼此独立，则 $X \sim B(400, 0.25)$，且 $\mathbb{E}(X) = 100$，$D(X) = 75$。
- 由德莫弗-拉普拉斯定理得

$$
\begin{aligned}
P(83 < X < 117) &= P(\frac{83-100}{\sqrt{75}} < \frac{X-100}{\sqrt{75}} < \frac{117-100}{75}) \\
&\approx \Phi(1.96) - \Phi(-1.96) \\
&= 2\Phi(1.96) - 1 \\
&= 0.95
\end{aligned}
$$

目录

中心极限定理的应用

例 (治愈率)

某药厂断言，该厂生产的某种药品对于医治一种疾病的治愈率为 0.8。医院任意抽查服用该药的 100 位病人，若其中多于 75 人治愈，就接收断言，否则拒绝断言。
(1) 若实际上该药治愈率为 0.8，接受该断言的概率是多少？
(2) 若实际上该药治愈率为 0.7，接受该断言的概率是多少？

用 X 表示 100 人中的治愈人数。
(1) $X \sim B(100, 0.8)$，则

$$
\begin{aligned}
P(X > 75) &= P\left(\frac{X - 100 \times 0.8}{\sqrt{100 \times 0.8 \times 0.2}} > \frac{75 - 80}{4}\right) \\
&\approx 1 - \Phi(-1.25) \\
&= \Phi(1.25) = 0.8944
\end{aligned}
$$

中心极限定理的应用

（2）$X \sim B(100, 0.7)$，则

$$
\begin{aligned}
P(X > 75) &= P\left(\frac{X - 100 \times 0.7}{\sqrt{100 \times 0.7 \times 0.3}} > \frac{75 - 70}{\sqrt{21}}\right) \\
&\approx 1 - \Phi(1.09) \\
&= 0.1379
\end{aligned}
$$

中心极限定理的应用

例 (保单售价)

某保险公司欲推出一项新业务，经分析，该业务每份保单的年赔付金额 $X \sim \exp(0.001)$。试建立每份保单的售价 Q 与参保人数 n 的关系，使得保险公司有 95% 的概率盈利。

用 X_i 表示第 i 个人的赔付金，$i = 1, \ldots, n$，则 $\{X_i\}$ 独立同分布。$\mathbb{E}(X_i) = 1000, D(X_i) = 1000^2$。盈利等价于 $\sum_{i=1}^n X_i \le nQ$。

$$
\begin{aligned}
P\left(\sum_{i=1}^n X_i \le nQ\right) &= P\left(\frac{\sum_{i=1}^n X_i - 1000n}{1000\sqrt{n}} \le \frac{n(Q-1000)}{1000\sqrt{n}}\right) \\
&\approx \Phi\left(\frac{n(Q-1000)}{1000\sqrt{n}}\right) = \Phi\left(\left(\frac{Q}{1000}-1\right)\sqrt{n}\right) \\
&= 0.95
\end{aligned}
$$

中心极限定理的应用

查表得 $(\frac{Q}{1000}-1)\sqrt{n}=1.65$，所以

$$Q=1000+\frac{1650}{\sqrt{n}}$$

中心极限定理的应用

例 (次品率)

某种产品的次品率为 5%，装箱时，
(1) 若每箱装 100 只，问至少有两件次品的概率？
(2) 若要以 99% 的把握保证每箱合格品不少于 100 只，问每箱至少应多装几只？

(1) 用 X 表示每箱次品数，则 $X \sim B(100, 0.05)$，于是

$$
\begin{aligned}
P(X \geq 2) &= P\left(\frac{X - 100 \times 0.05}{\sqrt{100 \times 0.05 \times 0.95}} \geq \frac{2-5}{\sqrt{4.75}}\right) \\
&\approx 1 - \Phi(-1.3764) \\
&= \Phi(1.3764) \\
&= 0.9162
\end{aligned}
$$

中心极限定理的应用

（2）设每箱应多装 k 只，则 $X \sim B(100+k, 0.05)$，希望 $P(X \le k) \ge 0.99$，于是

$$
\begin{aligned}
& P(X \le k) \\
= & P\left(\frac{X-(100+k)\times 0.05}{\sqrt{(100+k)\times 0.05\times 0.95}} \le \frac{k-(100+k)\times 0.05}{\sqrt{(100+k)\times 0.05\times 0.95}}\right) \\
\approx & \Phi\left(\frac{k-(100+k)\times 0.05}{\sqrt{(100+k)\times 0.05\times 0.95}}\right) \\
\ge & 0.99
\end{aligned}
$$

得 $k \ge 11$。

中心极限定理的应用

例 (电话交换机)

某单位电话交换机接有 500 部电话，在所有通话中有 96% 次通话是在各分机内进行的。假定每部分机是否需要打外线是相互独立的，问需要配备多少条外线才能以 95% 的概率保证每个分机要用外线时不必等候？

记 η 表示 500 台分机中同时打外线电话的分机台数，则 $\eta \sim B(500, 0.04)$。设需要外线 k 条使得 $P(\eta \le k) \ge 0.95$，即

$$P(\eta \le k) = P\left(\frac{\eta - 500 \times 0.04}{\sqrt{500 \times 0.04 \times 0.96}} \le \frac{k - 500 \times 0.04}{\sqrt{500 \times 0.04 \times 0.96}}\right)$$

$$\approx \Phi(\frac{k - 500 \times 0.04}{\sqrt{500 \times 0.04 \times 0.96}}) \ge 0.95$$

解得 $k \ge 27.23$，故需要 28 条。

小结

1. 随机变量序列的收敛性
2. 大数定律
3. 中心极限定理