概率论与数理统计

第五章: 大数定律与中心极限定理

李立颖 (lily@sustech.edu.cn)

南方科技大学数学系

2023 秋季

问题背景

“概率”的产生

随机试验 $\Rightarrow$ 统计数据 $\Rightarrow$ 统计规律性 $\Rightarrow$ 频率稳定性 $\Rightarrow$ 概率

频率稳定性

设 n 次独立重复试验中事件 A 发生的概率为 n_A，则当 $n \to \infty$ 时，有

$$\xi_n = \frac{n_A}{n} \to p = \mathsf{P}(A)$$

大数定律的客观背景

大量随机现象中的平均结果的稳定性

- 怎样定义极限 $\lim\limits_{n\to\infty} \frac{n_A}{n} = p$?
- “频率稳定性”的严格数学描述是什么?

实例

"抛硬币" 实验

将一枚硬币连续抛 n 次, 记 A 为正面朝上的事件, n_A 为 n 次试验中 A 发生的次数. 则 A 发生的频率为 $\xi_n = \dfrac{n_A}{n}$, $n = 1, 2, \ldots$. 可见:

- $\{\xi_n\}_{n=1}^{\infty}$ 是随机变量列
- $\{\xi_n = \xi_n(\omega)\}_{n=1}^{\infty}$ 是定义在样本空间 Ω 上的函数列.

复习: 函数列的收敛

设函数 $f(x)$, $f_n(x)$, $n = 1, 2, \ldots$ 在区间 (a, b) 上有定义, $f_n(x)$ 点点收敛于 $f(x)$ 是指, 对任意 $x \in (a, b)$,

$$\lim_{n \to \infty} f_n(x) = f(x)$$

问题

作为样本空间上的函数列 $\{\xi_n(\omega)\}_{n=1}^{\infty}$, 是否有

$$\lim_{n \to \infty} \xi_n(\omega) = p, \quad \forall \omega \in \Omega?$$

依概率收敛

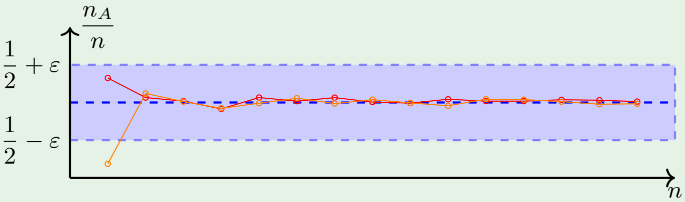
设 $\xi, \xi_1, \xi_2, \ldots$ 是一列随机变量. 若对于任意 $\varepsilon > 0$, 有

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

则称 $\{\xi_n\}$ **依概率收敛于** ξ, 记为 $\xi_n \xrightarrow{\mathsf{P}} \xi$.

如何理解 “依概率收敛”?

- $\lim\limits_{n\to\infty} \mathsf{P}(|\xi_n - \xi| \geq \varepsilon) = 0 \Leftrightarrow \lim\limits_{n\to\infty} \mathsf{P}(|\xi_n - \xi| < \varepsilon) = 1.$
- $\xi_n \xrightarrow{\mathsf{P}} \xi$ 的直观意义: 随着 n 的增大, 绝对误差 $|\xi_n - \xi|$ 较大的可能性越来越小.
- 对抛硬币实验: $\xi_n = \dfrac{n_A}{n} \xrightarrow{\mathsf{P}} \dfrac{1}{2}$.



伯努利大数定律

设 n_A 是 n 次独立重复试验中事件 A 发生的次数，且 $\mathsf{P}(A)=p$. 则 $\forall\varepsilon>0$，有

$$\lim_{n\to\infty}\mathsf{P}\left(\left|\frac{n_A}{n}-p\right|\geq\varepsilon\right)=0.$$

伯努利大数定律的意义

在概率的统计定义中，事件 A 发生的频率 $\frac{n_A}{n}$ “稳定于”事件 A 在一次试验中发生的概率 p，即“频率稳定性”：

当 n 足够大时，**概率** p 可被**频率** $\frac{n_A}{n}$ 近似代替.

证明方法

令 $n_A=X_1+\cdots+X_n$，其中 X_k 独立且服从参数为 p 的两点分布. 则 $\mathbb{E}X_1=p$ 且

$$\lim_{n\to\infty}\mathsf{P}\left(\left|\frac{n_A}{n}-p\right|\geq\varepsilon\right)=0\quad\Leftrightarrow\quad\lim_{n\to\infty}\mathsf{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mathbb{E}X_1\right|\geq\varepsilon\right)=0.$$

我们将证明更一般的切比雪夫大数律.

切比雪夫大数律

设 $\{X_n\}$ 为相互独立的随机变量列，且具有相同的数学期望 μ 和方差 σ^2. 则 $\forall \varepsilon > 0$, 有

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

切比雪夫不等式

设 ξ 的期望和方差存在，则 $\mathsf{P}\Big(|\xi - \mathbb{E}\xi| \geq \varepsilon\Big) \leq \dfrac{\mathrm{Var}(\xi)}{\varepsilon^2}$.

对 $\xi = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ 应用 Chebyshev 不等式，我们有

$$\mathbb{E}\xi = \mathbb{E}X_i = \mu, \quad \mathrm{Var}(\xi) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\sigma^2}{n}.$$

$$\text{故 } \forall \varepsilon > 0: \quad \mathsf{P}(|\frac{1}{n}\sum_{i=1}^{n} X_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0, \quad n \to \infty.$$

切比雪夫大数律的意义

具有相同数学期望和方差的独立随机变量序列的算术平均值依概率收敛于数学期望. 因此, 当 n 足够大时, 算术平均值几乎是一常数.

数学期望可被**算术均值**代替.

Kolmogorov 强大数律 (1933)

设 $\{X_i\}$ 为一列独立同分布随机变量，且期望 $\mu = \mathbb{E}X_i$ 存在. 令 $\xi_n = \frac{1}{n}\sum_{i=1}^{n} X_i$，则

$$\mathsf{P}\left(\lim_{n\to\infty} |\xi_n - \mu| = 0\right) = 1, \quad \text{或等价地,} \ \forall\varepsilon > 0, \mathsf{P}(\limsup_{n\to\infty} |\xi_n - \mu| \geq \varepsilon) = 0.$$

Grundbegriffe Der Wahrscheinlichkeitsrechnung, Kolmogorov 1933

mit $n \to \infty$ gegen Null konvergiert. Wenn alle $E(s_n)$ bestimmt sind und man

$$d_n = \mathsf{E}(s_n)$$

wählen kann, so ist die Stabilität *normal*.

Sind alle s_n gleichmäßig beschränkt, so folgt aus

$$(1) \qquad \mathsf{P}\{|s_n - d_n| \geqq \varepsilon\} \to 0 \qquad n \to +\infty$$

die Relation

$$|\mathsf{E}(s_n) - d_n| \to 0 \qquad n \to +\infty$$

und folglich

$$(2) \qquad \mathsf{P}\{|s_n - \mathsf{E}(s_n)| \geqq \varepsilon\} \to 0. \qquad n \to +\infty$$

Die Stabilität einer beschränkten stabilen Folge ist also notwendig normal.

Es sei

$$\mathsf{E}(s_n - \mathsf{E}(s_n))^2 = \sigma^2(s_n) = \sigma_n^2.$$

Nach der TCHEBYCHEFFschen Ungleichung ist

$$\mathsf{P}\{|s_n - E(s_n)| \geqq \varepsilon\} \leqq \frac{\sigma_n^2}{\varepsilon^2}.$$

Folglich ist die MARKOFF*sche Bedingung*

$$(3) \qquad \sigma_n^2 \to 0 \qquad n \to +\infty$$

für die normale Stabilität hinreichend.

Sind $s_n - E(s_n)$ gleichmäßig beschränkt:

$$|s_n - E(s_n)| \leqq M,$$

so ist nach der Ungleichung (9) aus § 3, viertes Kapitel

$$\mathsf{P}\{|s_n - E(s_n)| \geqq \varepsilon\} \geqq \frac{\sigma_n^2 - \varepsilon^2}{M^2}.$$

Folglich ist in diesem Falle die MARKOFFsche Bedingung (3) für die Stabilität von s_n auch notwendig.

Ist

$$s_n = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

und die Größen x_n paarweise unkorreliert, so hat man

$$\sigma_n^2 = \frac{1}{n^2}\{\sigma^2(x_1) + \sigma^2(x_2) + \cdots + \sigma^2(x_n)\}.$$

Folglich ist in diesem Falle für die normale Stabilität der Mittelwerte s_n die Bedingung

$$(4) \qquad \sigma_n^2 = \sigma^2(x_1) + \sigma^2(x_2) + \cdots + \sigma^2(x_n) = o(n^2)$$

hinreichend (*Satz von* TCHEBYCHEFF). Insbesondere ist die Bedingung (4) erfüllt, wenn alle Größen x_n gleichmäßig beschränkt sind.

Man kann diesen Satz auf den Fall der schwach korrelierten Größen x_n verallgemeinern: Setzt man voraus, daß der Korrelationskoeffizient r_{mn} * von x_m und x_n der Ungleichung

$$r_{mn} \leqq c(|n - m|)$$

genügt und daß

$$C_n = \sum_{k=0}^{k=n-1} c(k)$$

ist, so ist für die normale Stabilität der Mittelwerte s die Bedingung

$$(5) \qquad C_n \sigma_n^2 = o(n^2)$$

hinreichend[1].

Im Falle *unabhängiger* Summanden x_n kann man auch eine notwendige und hinreichende Bedingung für die Stabilität der Mittelwerte s_n geben. Für jedes x_n existiert eine Konstante m_n (Mediane von x_n), die den folgenden Ungleichungen genügt:

$$\mathsf{P}(x_n < m_n) \leqq \tfrac{1}{2},$$
$$\mathsf{P}(x_n > m_n) \leqq \tfrac{1}{2}.$$

Wir setzen nun

$$x_{nk} = x_k, \ \text{wenn} \ |x_k - m_k| \leqq n \ \text{ist},$$
$$x_{nk} = 0 \ \text{im entgegengesetzten Falle},$$
$$s_n^* = \frac{x_{n1} + x_{n2} + \cdots + x_{nn}}{n}.$$

Dann sind die Relationen

$$(6) \qquad \sum_{k=1}^{k=n} \mathsf{P}\{|x_k - m_k| > n\} = \sum_{k=1}^{k=n} \mathsf{P}(x_{nk} \neq x_k) \to 0, \qquad n \to +\infty$$

$$(7) \qquad \sigma^2(s_n^*) = \sum_{k=1}^{k=n} \sigma^2(x_{nk}) = o(n^2)$$

für die Stabilität der Größen s_n *notwendig und hinreichend*[2].

Man kann dabei die Konstanten d_n gleich den Größen $E(s_n^*)$ nehmen, so daß im Falle

$$\mathsf{E}(s_n^*) - \mathsf{E}(s_n) \to 0 \qquad n \to +\infty$$

(und nur in diesem Falle) die Stabilität *normal* ist.

Man erhält eine weitere Verallgemeinerung des TCHEBYCHEFFschen Satzes, wenn man voraussetzt, daß s_n irgendwie von den Ausgängen irgendwelcher n Versuche

$$\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_n$$

---

* Es ist offenbar immer $r_{nn} = 1$.

[1] Vgl. A. KHINTCHINE: Sur la loi forte des grandes nombres, C. R. de l'acad. sci. Paris, Bd. 186 (1928) S. 285.

[2] Vgl. A. KOLMOGOROFF: Über die Summen durch den Zufall bestimmter unabhängiger Größen. Math. Ann. Bd. 99 (1928) S. 309–319 — Berichtigungen ebenda Bd. 102 (1929) S. 484–488, Satz VIII (und der entsprechende Zusatz S. 318).

大数定律的意义

- 给出了“频率稳定性”的严格数学解释.
- 提供了通过试验来确定事件概率的方法.
- 是数理统计中参数估计的重要理论依据之一.
- 是 Monte Carlo 方法的主要数学理论基础.

几乎必然收敛与依概率收敛

对比

- X_n 依概率收敛到 0 是指 $\forall \varepsilon > 0: \lim_{n\to\infty} \mathsf{P}(|X_n| \geq \varepsilon) = 0.$
- X_n 几乎必然收敛到 0 是指 $\mathsf{P}\left(\left\{\omega : \lim_{n\to\infty} X_n(\omega) = 0\right\}\right) = 1.$

依概率收敛但不几乎必然收敛的例子

一个人被知乎网友永久地被关在一间 10 平米的小屋子里，各种生活必需品一应俱全，但只有一台电脑为伴. 坏消息是是电脑的硬盘容量只有 2 GB，好消息是电脑的容量会按如下的规律随机增长：在第 1 个小时会增长 1GB，在随后的 2 个小时中某一个随机小时会增长 1 GB，然后在随后的 3 个小时中某一个随机小时会增长 1GB，...，在第 $k(k-1)/2+1$ 至第 $k(k+1)/2$ 个小时中某随机小时会增长 1GB，依此类推.
记 X_n 为第 n 个小时硬盘容量的增长 (以 GB 为单位). 则

$$\mathsf{P}(|X_n| \geq \varepsilon) \lesssim \frac{c}{\sqrt{n}} \to 0 \Rightarrow X_n \xrightarrow{\mathsf{P}} 0.$$

但 $\lim_{n\to\infty} X_n \neq 0$，因为在无穷序列 $X_1, X_2, \ldots$ 中，对于任意大的 N，存在 $n > N$ 使得 $X_n = 1$.

为什么叫“大数定律”而不叫“大数定理”?

在概率论发展初期，由于概率的数学定义尚未明确，所以缺乏理解概率收敛的理论基础，故把频率“趋于”概率视为经大量试验而得到的结果，就像物理学的定律一样.
在概率论的公理化体系建立以后，大数定律可在理论上进行严格的证明而成为意义明确的定理，故现有有些教材上称为“大数定理”.

大数律的实际应用: Monte–Carlo 方法

Monte–Carlo 方法 或称为**计算机随机模拟方法、计算机仿真方法** 是科学与工程的一种重要工具.

Monte–Carlo 方法的原理主要基于大数定律.

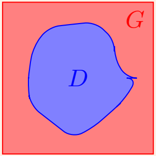
例

设计算机屏幕上有一矩形区域 G (不妨设 G 的面积为 1). 现用鼠标在 G 的内部任画一封闭曲线 L, 求 L 围成的内部图形 D 的面积 $|D|$.

解

用计算机产生一串相互独立、均服从 G 上均匀分布的随机变量 $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$. 记事件 $A = \{$产生的随机点落入 D 中$\}$, $n_A = (X_1, Y_1), \ldots, (X_n, Y_n)$ 中落入 D 的个数. 由伯努利大数定律有

$$\frac{n_A}{n} \xrightarrow{\mathsf{P}} \mathsf{P}(A) = \frac{|D|}{|G|}$$



因此当 n 充分大时, D 的面积 $|D| \approx \dfrac{n_A}{n}$.

1. 设随机变量 X 和 Y 数学期望都是 2, 方差分别为 1 和 4, 而相关系数为 0.5. 请根据切比雪夫不等式给出 $\mathsf{P}(|X-Y| \geq 6)$ 的上限.
2. 将一枚骰子重复掷 n 次, 则当 $n \to \infty$, 请写出 n 次掷出点数的算术平均值依概率收敛的极限.

引言

中心极限定理客观背景

在实际问题中，常常需要考虑许多随机因素所产生的总影响.
观察表示，如果一个量是由大量相互独立的随机因素的影响所造成，而每一个别因素在总影响中所起的作用不大，则这种量一般都服从或近似服从正态分布.

数学模型

设 $X_1, X_2, ..., X_n$ 是大量相互独立的随机变量，令 $Z_n = X_1 + X_2 + \cdots + X_n$，则 $Z \approx \mathcal{N}(\mu, \sigma^2)$. 或进行归一化:

$$\frac{Z_n - \mathbb{E}Z_n}{\mathrm{Var}(Z_n)} = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}X_i}{\sqrt{\sum_{i=1}^n \mathrm{Var}(X_i)}} \approx \mathcal{N}(0,1).$$

反例

取 $X_n = 0$, $n = 2, 3, \ldots$. 则

$$Z_n = \frac{X_1 - \mu_1}{\sigma_1}, \quad n = 1, 2, \ldots$$

除非 X_1 服从正态分布，否则 Z_n 不可能以正态分布为极限.

中心极限定理 (central limit theorem)

定义

令 $\{X_n\}$ 为独立随机变量列且 $\mathbb{E}X_n = \mu_n$, $\mathrm{Var}(X_n) = \sigma_n^2$. 若 $Z_n = X_1 + \cdots + X_n$ 的分布函数 $F_n(x)$ 对任意 x 满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} \mathsf{P}\left[\frac{\sum_{k=1}^n X_k - \sum_{k=1} \mu_k}{\sum_{k=1}^n \sigma_k^2} \le x\right] = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = \Phi(x),$$

则称 $\{X_n\}$ **服从中心极限定理**, 记作 $Z_n \Rightarrow \mathcal{N}(0,1)$.

独立同分布的中心极限定理

设 $\{X_n\}$ 为独立同分布的随机变量列, 其数学期望和方差分别为

$$\mathbb{E}X_k = \mu, \quad \mathrm{Var}(X_k) = \sigma^2 > 0, \qquad k = 1, 2, \ldots$$

则 $\{X_n\}$ 服从中心极限定理, 即

$$Z_n = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \Rightarrow \mathcal{N}(0,1).$$

De Moivre–Laplace 中心极限定理

设 $\{\eta_n\}$ 为服从参数为 n, p, $p \in (0,1)$ 的二项分布随机函数列，则对任意的 x，有

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

证明

因二项分布产生于 n 重伯努利试验，故 η_n 可分解为

$$\eta_n = X_1 + X_2 + \cdots + X_n,$$

其中 X_k, $k = 1, 2, \ldots$ 为独立同分布的 0-1 分布，且 $\mathbb{E}X_k = p$, $\mathrm{Var}(X_k) = p(1-p)$. 由独立同分布的中心极限定理，有

$$\lim_{n\to\infty} \mathsf{P}\left[\frac{\eta_n - np}{\sqrt{np(1-p)}} \le x\right] = \lim_{n\to\infty} \mathsf{P}\left[\frac{\sum_{i=1}^{n} - np}{\sqrt{np(1-p)}} \le x\right] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

De Moivre – Laplace 中心极限定理的应用

对于二项分布 $X \sim \text{Bin}(n,p)$，有

$$\frac{\eta_n - np}{\sqrt{np(1-p)}} \approx \mathcal{N}(0,1),$$

即当 n 很大时，

$$\eta_n \approx \mathcal{N}(np, np(1-p)).$$

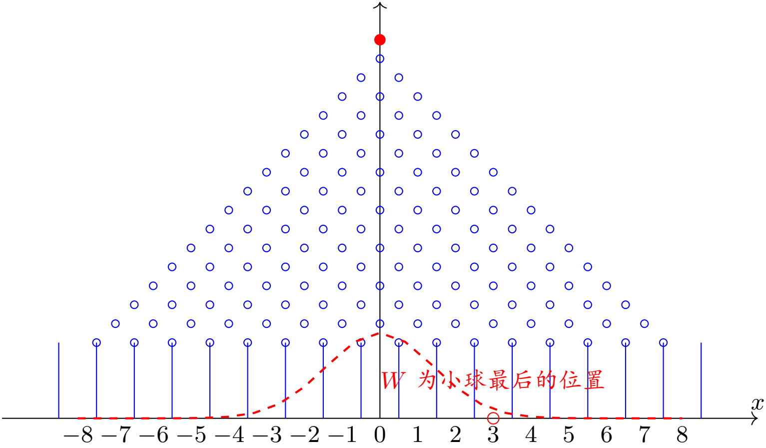
对比：泊松分布

当 n 很大，p 很小时，$np = \lambda$，有

$$\eta_n = \text{Bin}(n,p) \sim \pi(\lambda).$$

区别是什么？

高尔顿钉板试验



共 16 层钉板，X 为小球向右的次数，则

$$X \sim \mathrm{Bin}(16, 0.5), \quad W = \big(X - (16 - X)\big)/2 = X - 8 \approx \mathcal{N}(0, 4).$$