

大数定律与中心极限定理

第 8 讲 · 概率论与数理统计

复旦大学经济学院 ECON130001

上一讲我们做了什么

把整个分布压成几个数

期望 EX 、方差 $D(X) = E(X^2) - (EX)^2$ 、协方差 $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ 、相关系数 $\rho = \text{Cov}/(\sigma_X \sigma_Y)$ 。

两条最要紧的性质

$E(X + Y) = E(X) + E(Y)$ **不需要独立**； $D(X + Y) = D(X) + D(Y)$ **需要不相关**，否则要加 2Cov 。

★ 留下的问题

E 与 D 描述的还是一个随机变量。如果把很多个独立同分布的随机变量**平均**起来，会怎样？

绪论里布丰投针的频率会稳定到概率上、第 6 讲两个均匀分布相加变梯形再加会变钟形——**这两件事其实是同一件事**。

为什么需要这一讲

场景 1913 年 8 月 18 日，蒙特卡洛赌场。轮盘的小球连续落在**黑色**——一次、两次……一直到**第 26 次**。

冲突 赌客疯狂加注押红。理由听起来很有学问：**「大数定律说红黑各占一半。黑连开了 26 次，红欠得太多了，该补回来。」**

那一晚赌场赚走了数百万法郎。而这些人引用的，正是今天要讲的这条定理。

悬念 **大数定律确实说频率会趋于 $1/2$ 。可它并不意味着「落后的一方会追上来」——轮盘没有记忆。那它到底说了什么？「趋于」这两个字，究竟是什么意思？**

来源

1913 年 8 月 18 日蒙特卡洛赌场轮盘连开 26 次黑色，是「赌徒谬误」又称 *Monte Carlo fallacy* 的名称由来。单零轮盘上连续 26 次同色的概率约为 $(18/37)^{26} \approx 1.5 \times 10^{-8}$ （红黑任一色则约 $2 \times$ ，约 6840 万分之一）。en.wikipedia.org/wiki/Gambler's_fallacy | 访问日期 2026-08-02。「赚走数百万法郎」为该事件的通行记述，非精确账目。

本讲学习目标

- ① 用**切比雪夫不等式**把「偏离期望多远」变成可估计的量
- ② 说清**大数定律**到底断言了什么、没有断言什么
- ③ 用**中心极限定理**算出「偏离多远」的精确近似，而不只是一个上界

本讲回答概率论的两个终极问题

问题一：凭什么相信反复抛硬币，经验频率会趋向理论概率 0.5？ → **大数定律**

问题二：为什么从身高、测量误差到金融收益率，处处都是正态分布？ → **中心极限定理**

先看一组数：多抛几次会怎样

例

抛 10 次硬币，恰好 5 次正面的概率是多少？正面比率落在 0.4 到 0.6 之间呢？抛 100 次呢？

抛 10 次

$$P(X = 5) = C_{10}^5 \left(\frac{1}{2}\right)^{10} = 0.2461$$

$$P(0.4 \leq \frac{X}{10} \leq 0.6) = \sum_{i=4}^6 C_{10}^i \left(\frac{1}{2}\right)^{10} = 0.6563$$

抛 100 次

$$P(X = 50) = C_{100}^{50} \left(\frac{1}{2}\right)^{100} = 0.0796$$

$$P(0.4 \leq \frac{X}{100} \leq 0.6) = \sum_{i=40}^{60} C_{100}^i \left(\frac{1}{2}\right)^{100} = 0.9648$$

★ 两个数走了相反的方向

「恰好一半」的概率下降了 ($0.2461 \rightarrow 0.0796$)，而「比率接近一半」的概率上升了 ($0.6563 \rightarrow 0.9648$)。

大数定律说的是后者，不是前者。这正是开场那些赌客弄反的地方。

问题的提法

设定

- $X_1, X_2, \dots$ 独立同分布, 都有期望 μ
- 直观上, n 很大时 $\bar{X}_n = \frac{1}{n} \sum X_i$ 应该接近 μ
- **但 $\bar{X}_n$ 的分布本身可能很复杂**, 不好直接算

大数定律要做的, 就是把这个直观变成定理。

而为了证它, 我们需要一个工具: **只知道期望和方差, 就能估计「偏离多远」的概率** ——即使完全不知道分布长什么样。

切比雪夫不等式

定理

设 X 有期望 EX 与方差 DX ，则对任意 $\varepsilon > 0$

$$P(|X - EX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2} \iff P(|X - EX| < \varepsilon) \geq 1 - \frac{DX}{\varepsilon^2}$$

证明思路

以连续型为例： $DX = \int (x - EX)^2 f(x) dx \geq \int_{|x - EX| \geq \varepsilon} (x - EX)^2 f(x) dx \geq \varepsilon^2 \int_{|x - EX| \geq \varepsilon} f(x) dx = \varepsilon^2 P(|X - EX| \geq \varepsilon)$ 。

两次放缩：先扔掉一部分非负积分，再把被积函数换成它在该区域上的下界。完整证明见教师笔记。

它的价值与它的代价

价值：不需要知道分布，只要有期望和方差就能给出保证。这在实际中极为难得。

代价：非常松。证明里放缩了两次，所以给出的只是个粗糙上界——本讲后半会看到它有多松。

切比雪夫不等式的两道用法

正着用：估计概率

10000 盏灯，每盏开的概率 0.7，彼此独立。求同时开着的灯数在 6800 至 7200 之间的概率。

$X \sim B(10000, 0.7)$, $EX = 7000$ 、 $DX = 10000 \times 0.7 \times 0.3 = 2100$ ：

$$P(|X - 7000| < 200) \geq 1 - \frac{2100}{200^2} \approx 0.95$$

反着用：估计方差

已知 $EX = 10$ 、 $P(X \leq 7) = 0.2$ 、 $P(X \geq 13) = 0.3$ 。证明 $DX \geq \frac{9}{2}$ 。

取 $\varepsilon = 3$ 。左边 $P(|X - 10| \geq 3) \geq 0.2 + 0.3 = 0.5$ ，而不等式给出 $P(|X - 10| \geq 3) \leq \frac{DX}{9}$ ，故

$$\frac{DX}{9} \geq 0.5 \implies DX \geq \frac{9}{2}$$

同一个不等式，两个方向都能用：知道方差估概率，或者知道概率反推方差的下界。第二种用法在考试里出现得不少。

样本均值的期望与方差

核心结论（本讲反复要用）

设 $X_1, \dots, X_n$ 来自期望 μ 、方差 σ^2 的总体， $\bar{X}_n = \frac{1}{n} \sum X_i$ ，则

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu, \quad D(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n D(X_i) = \frac{\sigma^2}{n}$$

注意 D 里那个 $\frac{1}{n^2}$ ——由第 7 讲 $D(aX) = a^2 D(X)$ ，且各项独立才能相加。

★ 这两个式子是整个统计学的地基

期望不变，方差缩小 n 倍。 标准差缩小 $\sqrt{n}$ 倍——想把精度提高一倍，样本量要变成**四倍**。

代入切比雪夫：

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

右边趋于 0——大数定律已经证完了。

例：需要多大的样本

例

某总体期望未知，但已知标准差不超过 2。要保证样本均值与期望的距离小于 1 的概率至少 0.99，样本量至少多大？

解

由 $D(\bar{X}_n) = \frac{\sigma^2}{n} \leq \frac{4}{n}$ 与切比雪夫：

$$P(|\bar{X}_n - \mu| < 1) \geq 1 - \frac{D(\bar{X}_n)}{1^2} \geq 1 - \frac{4}{n}$$

要求 $1 - \frac{4}{n} \geq 0.99$ ，即 $\frac{4}{n} \leq 0.01$ ：

$$n \geq 400$$

注意分母是 n 不是 n^2 —— $D(\bar{X}_n) = \sigma^2/n$ 。这一步写错是最常见的失误。

另外：400 是**切比雪夫给出的保守答案**。本讲后半用中心极限定理重算，需要的样本量会小得多。

「趋于」是什么意思：依概率收敛

定义

若存在常数 a ，使得对任意 $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|X_n - a| < \varepsilon) = 1$$

则称随机变量序列 $\{X_n\}$ **依概率收敛**于 a 。

★ 这就是开场那个问题的答案

依概率收敛说的是：**「偏离超过 ε 」这件事的概率趋于 0。**

它**没有**说：某一次的结果会被「修正」；也**没有**说：黑色多了红色就会补上。**轮盘没有记忆。**

频率之所以趋于 $1/2$ ，不是因为后面的结果去抵消前面的，而是因为**后面的次数越来越多，把前面那 26 次稀释掉了。**

三个版本的大数定律

切比雪夫定理（最一般）

$X_1, X_2, \dots$ 相互独立，方差**一致有界**（存在 l 使所有 $DX_i < l$ ），则

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

不要求同分布，只要求独立 + 方差有界。

伯努利大数律

n 次独立试验中事件 A 发生 X 次，则频率依概率收敛于概率：

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{X}{n} - p \right| < \varepsilon \right\} = 1$$

这就是绪论那个承诺：布丰投针的频率为什么会稳定到 π 相关的那个数上。

辛钦大数律

$X_1, X_2, \dots$ 独立**同分布**， $EX_i = a$ ，则

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

只要期望存在，不要求方差存在——比切比雪夫定理的条件弱。

两道例题

例：部件长度

10 个部分独立同分布，各长期期望 2 mm、标准差 0.01 mm。总长 (20 ± 0.1) mm 为合格，求合格概率。

$Y = \sum_{i=1}^{10} X_i$, $EY = 20$ 、 $DY = 10 \times 0.0001 = 0.001$:

$$P(|Y - 20| < 0.1) \geq 1 - \frac{0.001}{0.01} = 0.9$$

例：计算器舍入误差

每个加数舍入到最近整数，误差独立且服从 $U(-0.5, 0.5)$ 。

(1) 1500 个数相加，误差总和绝对值超过 15 的概率？

$EY = 0$, $DY = 1500 \times \frac{1}{12} = 125$:

$$P(|Y| \geq 15) \leq \frac{125}{225} = 55.6\%$$

(2) 最多几个数相加，能让 $P(|Y| < 10) \geq 0.9$?

需 $1 - \frac{DY}{100} \geq 0.9$ 即 $DY \leq 10$, 故 $\frac{n}{12} \leq 10$, $n \leq 120$ 。

应用：保险公司凭什么稳赚

问题

一份保费 500 元、保额 10 万元的意外险。对**单个**投保人，公司要么净赚 500，要么巨亏 99 500——这是极度不确定的赌局。

大数定律的保证

- 设出险概率 p ，第 i 张保单的收益 X_i ，则 $E[X_i] = 500(1 - p) - 99500p$ 。**精算师的工作就是确保 p 足够小，使 $E[X_i] > 0$**
- 投保人数 n 足够大时，实际平均收益 $\frac{1}{n} \sum X_i$ 以极高概率收敛于 $E[X_i]$

结论：只要定价合理，大数定律就保证了公司整体经营的稳定。它把个体的巨大不确定性，转化成了总体的确定性。

但请注意前提：**各保单要相互独立**。一场地震让整个区域同时出险时，这个前提就塌了——第 3 讲结构化产品那一课的保险版本。

应用：机器学习与蒙特卡洛

让机器学习成为可能

- 训练目标是 최소화 损失的**期望** $E[L]$ ，但真实数据的分布我们永远不知道， $E[L]$ 算不出来
- 于是改算 n 个样本上的**平均损失** $\frac{1}{n} \sum L_i$
- **大数定律保证**： $n \rightarrow \infty$ 时两者一致

这是数据驱动方法的理论支点。没有大数定律，「大数据」就没有意义。

蒙特卡洛模拟

- 怎么算复杂积分 $I = E[g(X)]$ （衍生品定价、物理模拟里到处都是）？
- 从分布中抽 N 个样本，用 $\frac{1}{N} \sum g(x_i)$ 近似
- **大数定律保证**这个近似可靠

绪论的布丰投针、第 2 讲 AlphaGo 的蒙特卡洛树搜索，**用的都是这一条。**

接下来

大数定律告诉我们：样本均值会趋近期望。

但它只给了一个粗糙的上界——切比雪夫算出 55.6%，实际只有 18%。我们无法比较精确地估计「偏离到底有多大」。

中心极限定理给出的不是上界，是分布本身。

中心极限定理 (Liapunov 定理)

定理

设 $X_1, X_2, \dots$ 相互独立, $E(X_i) = \mu_i$ 、 $D(X_i) = \sigma_i^2$ 。记 $Z_n = X_1 + \dots + X_n$, 将其标准化:

$$\frac{Z_n - E(Z_n)}{\sqrt{D(Z_n)}} = \frac{\sum_i X_i - \sum_i \mu_i}{\sqrt{\sum_i \sigma_i^2}}$$

若**每个 X_i 对总和的影响都不大**, 则

$$\lim_{n \rightarrow \infty} P \left\{ \frac{Z_n - E(Z_n)}{\sqrt{D(Z_n)}} \leq x \right\} = \Phi_0(x)$$

★ 读懂这句话的分量

不管 X_i 服从什么分布——只要独立、方差有限、没有哪一项特别大——它们的**和标准化后总是趋于标准正态**。

这就是第 5 讲那个问题的答案: **正态分布无处不在, 不是因为世界本来正态, 而是因为「大量微小独立因素之和」必然趋近正态。**

回到舍入误差：CLT 比切比雪夫准多少

同一道题，换个工具

1500 个数相加，误差独立且 $\sim U(-0.5, 0.5)$, $EY = 0$ 、 $DY = 125$ 。求 $P(|Y| \geq 15)$ 。

切比雪夫（上界）

$$P(|Y| \geq 15) \leq \frac{125}{225} = 55.6\%$$

中心极限定理（近似值）

$$\frac{15}{\sqrt{125}} = 1.34$$

$$P(|Y| \geq 15) \approx 1 - [\Phi_0(1.34) - \Phi_0(-1.34)] = 0.1802$$

55.6% 对 18%——切比雪夫松了三倍。

第 (2) 问同样：切比雪夫给 $n \leq 120$ ，CLT 给 $n \leq 441$ （由 $2\Phi_0(10/\sqrt{DY}) - 1 \geq 0.9$ 得 $DY \leq 36.7$, $n \leq 12 \times 36.7$ ）。**能多算三倍多的数。**

代价：切比雪夫是严格成立的不等式，CLT 是近似。要精度就用 CLT，要保证就用切比雪夫。

样本均值的近似分布

CLT 的样本均值版本（最常用的形式）

$X_1, X_2, \dots$ 独立同分布, $EX_i = \mu$ 、 $DX_i = \sigma^2$, 则

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x \right\} = \Phi_0(x)$$

即 n 充分大时

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \dot{\sim} N(0, 1), \quad \bar{X}_n \dot{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

分母那个 $\sigma/\sqrt{n}$ 正是 slide 11 算出的 $\sqrt{D(\bar{X}_n)}$, 它有个专门的名字: **标准误** (standard error)。

这个式子是第 9–12 讲全部内容的出发点。 从下一讲起, 我们做的每一件事都从它开始。

两道标准应用

例：要多少个器件

寿命方差 $\sigma^2 = 400$ ，用 $\bar{X}$ 估计 μ 。为使 $P(|\bar{X} - \mu| < 1) \geq 0.95$ ， n 至少多大？

$$2\Phi_0\left(\frac{\sqrt{n}}{\sigma}\right) - 1 \geq 0.95 \implies \frac{\sqrt{n}}{20} \geq 1.96$$
$$n \geq (1.96 \times 20)^2 \approx 1537$$

1.96 这个数要记住——95% 的双侧分位点，第 10 讲的置信区间天天用。

例：木柱

80% 的木柱长度不小于 3 m。随机选 100 根，求至少 30 根短于 3 m 的概率。

$X_i = 1$ 表示短于 3 m， $EX_i = 0.2$ 、 $DX_i = 0.16$ 。

$Y = \sum X_i$ ， $EY = 20$ 、 $DY = 16$ ：

$$P(Y \geq 30) \approx 1 - \Phi_0\left(\frac{30 - 20}{4}\right) = 1 - \Phi_0(2.5) = 0.0062$$

二项分布的精确值是 0.0112——近似值偏小了近一半。尾部是 CLT 近似最差的地方。

棣莫弗—拉普拉斯定理

定理（CLT 用在二项分布上的特例）

若 $X \sim B(n, p)$, 则 $n \rightarrow \infty$ 时

$$P(a < X < b) \approx \Phi_0\left(\frac{b - np}{\sqrt{np(1 - p)}}\right) - \Phi_0\left(\frac{a - np}{\sqrt{np(1 - p)}}\right)$$

现在二项分布有两个近似了

近似	适用条件	出处
泊松 $\lambda = np$	n 大、 p 小 、 np 适中	第 4 讲
正态 $N(np, np(1 - p))$	n 大、 p 不太靠近 0 或 1	本讲

经验判据： $np > 5$ 且 $n(1 - p) > 5$ 时用正态近似。 p **极小**就用泊松。

综合例题：家长会

例

每个学生有 0、1、2 名家长参会，概率分别 0.05、0.8、0.15。全校 400 名学生，相互独立。求 (1) 参会家长数超过 450 的概率；(2) 恰有 1 名家长参会的学生数不多于 340 的概率。

(1) 用 Liapunov

$EX_i = 1.1, EX_i^2 = 1.4, DX_i = 1.4 - 1.21 = 0.19$
。 $X = \sum_{i=1}^{400} X_i, EX = 440, DX = 76$ ：

$$\begin{aligned} P(X > 450) &= 1 - \Phi_0\left(\frac{450 - 440}{\sqrt{76}}\right) \\ &= 1 - \Phi_0(1.147) = 0.1257 \end{aligned}$$

(2) 用拉普拉斯

$Y = \text{恰 1 名家长的学生数} \sim B(400, 0.8), EY = 320$
、 $DY = 64$ ：

$$\begin{aligned} P(Y \leq 340) &= \Phi_0\left(\frac{340 - 320}{8}\right) \\ &= \Phi_0(2.5) = 0.9938 \end{aligned}$$

同一道题的两问用了两个定理：(1) 的 X_i 取三个值不是伯努利，只能用一般的 Liapunov；(2) 的 Y 是标准二项，用拉普拉斯更直接。**先看清随机变量是什么。**

应用：零件寿命与电梯载重

零件寿命

100 个元件，每个寿命期望 1000 小时、标准差 100，**分布未知**。求设备总寿命超过 98000 小时的概率。

不需要知道分布！ $S \sim N(100000, 1000^2)$ ：

$$P(S > 98000) = P(Z > -2) \approx 0.9772$$

电梯载重

限重 800 kg。成年男性体重期望 70 kg、标准差 10 kg，10 人同乘，求超重概率。

$S \sim N(700, 1000)$ ，标准差 ≈ 31.6 ：

$$P(S > 800) = P(Z > 3.16) \approx 0.0008$$

但要留个心眼： $n = 10$ **不算大**，而且这里问的是**尾部**概率——正是 CLT 近似最不准的地方（对照木柱那题）。工程上会留更大余量。

应用：A/B 测试与置信区间

A/B 测试

测试新版网页 B 的转化率是否优于旧版 A。

- 样本量足够大时， $\hat{p}_A$ 与 $\hat{p}_B$ 都近似正态
- **更要紧的是：它们的差值也近似正态**（正态之和仍正态，第 6 讲）
- 于是能用 z 检验判断：观测到的差异是真实效果，还是随机波动

置信区间从哪来

调查机构凭 1000 个样本，怎么得出「比例为 $55\% \pm 3\%$ 」？

- 样本比例 $\hat{p}$ 是大量独立变量的均值，CLT 保证它近似正态
- 正因为有这个正态性，才能构造出那个 $\pm 3\%$

没有中心极限定理，绝大多数统计推断都失去理论基础。

这两件事分别是**第 12 讲假设检验**与**第 10 讲区间估计**的内容。本讲讲完，后半学期的地基就打好了。

本讲总结

① 回到蒙特卡洛的那一夜

赌客们没有算错，他们**把定理用反了方向**。

大数定律说的是「**比率**趋于 $1/2$ 」，不是「**次数**会被补平」。看 slide 6 那组数：抛得越多，「恰好一半」的概率下降，「比率接近一半」的概率上升——这是两件相反的事。

频率之所以收敛，是因为后来的次数把前面那 26 次稀释了，而不是因为后来的结果去抵消它。轮盘没有记忆。

本讲总结（续）

② 核心结论

- **切比雪夫不等式** $P(|X - EX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2}$ ——不需要分布，但很松
- **样本均值**: $E(\bar{X}_n) = \mu$ 、 $D(\bar{X}_n) = \frac{\sigma^2}{n}$ ——期望不变，标准差按 $\sqrt{n}$ 缩小
- **大数定律**: $\bar{X}_n$ 依概率收敛于 μ
- **中心极限定理**: $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ ——不管原分布是什么

③ 易错提醒

- $D(\bar{X}_n) = \sigma^2/n$, **分母是 n 不是 n^2**
- 大数定律说的是**比率**收敛，不是次数会被补平
- 切比雪夫给**上界**（严格），CLT 给**近似**（准但不保证）
- **CLT 在尾部不准**（木柱题 0.0062 vs 精确 0.0112）， n 小时更要当心

④ 下一讲

本讲那个式子里 μ 与 σ 是**已知**的——我们从模型推数据。可现实完全相反：**手里只有数据， μ 和 σ 恰恰是要求的東西**。从下一讲起方向反过来：**由数据推模型**，这就是绪论说的统计学。

下一讲：样本及抽样分布。