

# 第七讲大数定律 (详细讲稿)

龚鹤扬

2025 年 5 月 7 日

## 幻灯片 1: 标题页

- (展示标题页)
- 好的, 今天我们探讨的主题是” 第七讲大数定律”。我是龚鹤扬。

## 幻灯片 2: 目录

- (展示目录页)
- 在正式开始之前, 我们先看一下本讲的主要内容安排:
  1. **引言**: 首先, 我们会简单介绍一下大数定律的重要性, 以及一个非常关键的点——它的局限性, 也就是说, 它并非对所有情况都成立。
  2. **切比雪夫不等式**: 这是推导某些大数定律的重要工具, 我们会先学习这个不等式。
  3. **弱大数定律 (WLLN)**: 接着, 我们会详细讨论弱大数定律, 包括它的不同形式, 比如切比雪夫弱大数定律和伯努利弱大数定律, 还会引入非常重要的辛钦弱大数定律。
  4. **强大数定律 (SLLN)**: 之后, 我们会学习比弱大数定律更强的收敛形式——强大数定律, 主要是柯尔莫哥洛夫强大数定律。
  5. **WLLN 与 SLLN 的区别与联系**: 这两种大数定律有什么不同? 又有什么联系? 我们会通过对比, 特别是引入正态分布和柯西分布的例子来加深理解。
  6. **大数定律的应用**: 了解了理论之后, 我们会看看大数定律在实际中是如何应用的, 以及应用时需要注意什么。
  7. **总结与展望**: 最后, 我们会对本讲内容进行总结, 并展望一下接下来可能涉及的相关概念, 比如中心极限定理。

## 幻灯片 3: 引言: 大数定律的重要性与局限

- (展示引言页)
- 那么, 我们正式开始第一部分: 引言。

- 大家可以看到，大数定律是概率论中一组非常重要的极限定理。它的核心思想其实非常直观：当进行大量重复试验的时候，一个事件发生的”频率”会越来越接近它真实的”概率”；同样地，样本的均值也会趋向于总体的”期望”。
- 这一定律是我们理解从理论概率到实际统计推断的关键。很多我们习以为常的统计方法，比如用样本均值去估计总体均值，其背后都是大数定律在提供理论保障。
- **但是，这里有一个非常重要的提示，也是我们本讲会反复强调的一点：**大数定律的结论并不是无条件成立的！它依赖于随机变量序列满足特定的”条件”，比如期望存在、方差有限等等。
- 如果这些条件不满足，大数定律的结论可能就不成立了。为了让大家更深刻地理解这一点，我们本讲会特别介绍一个著名的反例——”柯西分布”，通过它来探讨这些条件的必要性，以及当条件不满足时会发生什么。所以，请大家在学习过程中特别留意这些”条件”。

## 幻灯片 4: 7.1 切比雪夫不等式

- (展示切比雪夫不等式页)
- 接下来，我们学习”7.1 切比雪夫不等式”。这个不等式本身非常有用，也是我们后续证明弱大数定律的一个关键工具。
- 切比雪夫不等式的作用是什么呢？它给出了一个随机变量偏离其期望值的概率的”上界”。也就是说，它告诉我们，一个随机变量的值落在其期望值 一定范围之外的概率，最大不会超过多少。
- 这个不等式的厉害之处在于，它不依赖于随机变量的具体分布形式——无论是正态分布、均匀分布还是其他什么奇奇怪怪的分布——只要这个随机变量的”期望”和”方差”存在并且有限，这个不等式就成立。
- 我们来看一下定理 7.1，也就是切比雪夫不等式的具体形式。
- 假设随机变量  $X$  的期望是  $E(X) = \mu$ ，方差是  $Var(X) = \sigma^2$ ，并且这里要求方差  $\sigma^2$  是一个大于 0 的有限值。那么对于任何一个正数  $\epsilon > 0$ ，我们有：

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

- 这就是说， $X$  取值与它的均值 的偏差的绝对值大于或等于  $\epsilon$  的概率，不会超过方差  $\sigma^2$  除以  $\epsilon^2$ 。
- 这个不等式也可以等价地写成：

$$P(|X - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2}$$

- 也就是说， $X$  落在以  $\mu$  为中心，宽度为  $2\epsilon$  的区间  $(\mu - \epsilon, \mu + \epsilon)$  内的概率，至少是  $1 - \frac{\sigma^2}{\epsilon^2}$ 。

- 那么，这个不等式有什么意义和特点呢？
  - 首先是它的”普适性”：正如我们前面提到的，只要期望和方差存在，它对任何分布都成立。这一点非常强大。
  - 其次是它的”实用性”：当随机变量的分布未知，我们又想对它偏离期望的概率进行估计时，切比雪夫不等式就能提供一个（尽管可能比较粗略的）估计。
  - 最后，也要认识到它的”局限性”：通常情况下，这个不等式给出的界是比较宽松的，也就是说，实际的概率可能远小于这个上界。它给的是一个”最坏情况”的保证。所以，如果对精度要求很高，切比雪夫不等式可能不够用，但它在理论推导中非常重要。

## 幻灯片 5: 7.2 弱大数定律 (WLLN)

- (展示弱大数定律 - 切比雪夫 WLLN 页)
- 了解了切比雪夫不等式之后，我们来看本讲的第一个核心内容：”7.2 弱大数定律”，英文是 Weak Law of Large Numbers，简称 WLLN。
- 弱大数定律的核心思想是：在特定条件下，当试验次数  $n$  非常大时，样本均值会”依概率收敛”于总体的期望。这里的”依概率收敛”是一个很重要的数学概念，我们马上会看到它的定义。
- 我们首先介绍的是”定理 7.2：切比雪夫弱大数定律”。
- 这个定理的条件是：我们有一列随机变量  $X_1, X_2, \dots, X_n, \dots$ 。它们需要满足三个条件：
  - (i) ”相互独立”；
  - (ii) 具有相同的期望  $E(X_i) = \mu$ ；
  - (iii) 具有相同的”有限方差”  $Var(X_i) = \sigma^2 < +\infty$ 。
- 如果这些条件都满足，我们令样本均值为  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ 。那么，对于任意给定的正数  $\epsilon > 0$ ，我们有：

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$$

- 这个式子是什么意思呢？它表示，当样本量  $n$  趋向于无穷大时，样本均值  $\bar{X}_n$  与总体期望  $\mu$  之间的偏差的绝对值大于或等于任意小的正数  $\epsilon$  的概率，趋向于 0。
- 或者等价地说：

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$$

- 这就是说，当  $n$  趋于无穷大时，样本均值  $\bar{X}_n$  落在总体期望  $\mu$  的任意小邻域  $(\mu - \epsilon, \mu + \epsilon)$  内的概率趋向于 1。
- 这种收敛方式，我们就称之为样本均值  $\bar{X}_n$  ”依概率收敛”于  $\mu$ ，记作  $\bar{X}_n \xrightarrow{P} \mu$ 。

- 下面有个小字注释：这个定理的证明依赖于方差存在且有限，并且可以直接通过我们前面学的切比雪夫不等式来证明。这也是为什么我们先讲切比雪夫不等式的原因。

## 幻灯片 6: 伯努利弱大数定律

- (展示伯努利弱大数定律页)
- 接下来，我们看一个弱大数定律的非常经典且重要的特例——“伯努利弱大数定律”，也就是定理 7.3。
- 这个定律描述的是我们非常熟悉的“频率的稳定性”。
- 它的背景是这样的：我们进行  $n$  次独立重复的伯努利试验。在每次试验中，事件  $A$  发生的概率都是  $p$ 。我们用  $n_A$  来表示在这  $n$  次试验中事件  $A$  实际发生的次数。
- 那么，事件  $A$  发生的“频率”就是  $f_n = \frac{n_A}{n}$ 。
- 伯努利弱大数定律告诉我们，对于任意给定的正数  $\epsilon > 0$ ，有：

$$\lim_{n \rightarrow \infty} P(|f_n - p| \geq \epsilon) = 0$$

- 这个式子表明，当试验次数  $n$  趋于无穷大时，事件  $A$  发生的频率  $f_n$  与其真实概率  $p$  之间的偏差的绝对值大于或等于任意小的正数  $\epsilon$  的概率，趋向于 0。
- 换句话说，当试验次数  $n$  非常大时，事件的频率  $f_n$  会以很高的概率无限接近其真实的概率  $p$ 。这就是我们常说的“频率稳定性”的数学表述。
- 下面小字注释也指出了：伯努利大数定律其实是切比雪夫弱大数定律的一个重要特例。为什么呢？因为每次伯努利试验可以看作一个随机变量（比如，事件  $A$  发生记为 1，不发生记为 0），这些随机变量独立同分布，有期望  $p$ ，有方差  $p(1-p)$ （这是有限的）。所以它满足切比雪夫 WLLN 的条件。这个定律为我们用频率来估计概率提供了坚实的理论依据。

## 幻灯片 7: 期望存在的重要性：引入柯西分布

- (展示引入柯西分布页)
- 前面我们学习了切比雪夫弱大数定律，它要求随机变量序列具有相同的有限方差。现在，我们来思考一个问题：这个“方差有限”的条件是不是必需的？或者说，有没有条件更弱一些的大数定律呢？
- 答案是有的。许多其他版本的大数定律，比如我们稍后会讲到的辛钦弱大数定律和柯尔莫哥洛夫强大数定律，它们对随机变量的条件有所放宽，不再要求方差有限，但它们通常会保留一个更弱但依然非常关键的条件，那就是：“期望的存在性”。

- 这就引出了一个自然的问题：是不是所有随机变量的期望都存在呢？或者说，是否存在期望不存在的随机变量？

- 答案是：是的，确实存在期望不存在的随机变量。而”柯西分布”就是这样一个非常典型的例子。柯西分布在大数定律乃至整个概率统计领域都是一个非常重要的反例，它可以帮助我们理解很多理论的边界和条件的重要性。

- 我们来看一下柯西分布。幻灯片左侧给出了它更普适的概率密度函数 (PDF)：

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma \left[ 1 + \left( \frac{x-x_0}{\gamma} \right)^2 \right]}$$

- 这里的  $x_0$  是位置参数，它可以是任何实数，决定了分布的峰值在哪里，它也是这个分布的中位数和众数。 $\gamma$  (gamma) 是尺度参数，必须大于 0，它决定了分布的”胖瘦”程度，也就是峰的宽度，它等于半峰全宽 (Half-Width at Half-Maximum)。

- 当我们取  $x_0 = 0$  和  $\gamma = 1$  时，就得到了”标准柯西分布”，也就是  $f(x) = \frac{1}{\pi(1+x^2)}$ ，这是我们之前可能接触得比较多的一种形式。

- 这个分布有什么特点呢？

- 首先，它的形状是类似钟形的，对称于位置参数  $x_0$ 。
- 请大家同时看一下幻灯片右侧的图像。这张图对比了标准正态分布（蓝色）和标准柯西分布（红色）的概率密度函数。可以非常清楚地看到，柯西分布的尾部要比正态分布”厚得多”，我们称之为”重尾分布”。这意味着它出现极端值的概率相对更高。
- 回到左侧文字，关键在于第二点：”期望  $E(X)$  不存在”。对于任何参数  $x_0$  和  $\gamma$  的柯西分布，如果我们去计算它的期望  $\int_{-\infty}^{\infty} xf(x; x_0, \gamma)dx$ ，我们会发现这个积分是发散的，更准确地说绝对是绝对期望  $\int_{-\infty}^{\infty} |x|f(x; x_0, \gamma)dx$  发散。因此，柯西分布的期望不存在。
- 既然期望都不存在，那么依赖于期望的”方差  $Var(X)$ ”自然也就不存在了。

- (点按，出现”这意味着什么？”)

- 那么，柯西分布的这些特点，特别是期望不存在，对我们讨论大数定律意味着什么呢？

- 首先，柯西分布不满足绝大多数大数定律版本所要求的”期望存在”或”方差有限”的条件。
- 其次，我们将会在后面看到，对于一系列独立同分布 (i.i.d.) 的柯西随机变量，它们的样本均值并”不会”像我们通常期望的那样收敛到任何常数。这是一个非常反直觉但又极其重要的结论。

- 所以，柯西分布的引入，就是为了提醒我们，大数定律的条件是不能随意忽略的。

## 幻灯片 8: 可视化: 正态 vs. 柯西样本均值路径

- (展示可视化: 正态 vs. 柯西样本均值路径页)

- 在理论上理解了柯西分布的特性，并直观对比了它的 PDF 之后，我们来看一个更动态的可视化对比。这张幻灯片展示了正态分布和柯西分布的样本均值，是如何随着样本量  $n$  的变化而变化的轨迹。
- 请看左边的图：这是标准正态分布  $N(0, 1)$  的情况。我们可以看到，随着样本量  $n$ （横轴）的增加，样本均值  $\bar{X}_n$ （纵轴）很快就稳定地收敛到了总体的期望  $\mu = 0$ 。这完全符合我们大数定律的预期，表现出良好的收敛性。
- 现在请看右边的图：这是标准柯西分布  $C(0, 1)$  的情况。大家可以看到，它的样本均值  $\bar{X}_n$  表现出持续且剧烈的波动，完全没有要收敛到某一个特定值的趋势，即便样本量  $n$  已经很大了也是如此。图中的那些“尖峰”就代表了柯西分布的极端值（离群点）对样本均值的巨大影响。
- 这种极端值持续“污染”样本均值的现象，使得柯西分布的样本均值无法稳定下来，也就不会收敛。这与左边正态分布的平稳收敛形成了鲜明对比。这个可视化再次强调了期望（以及有时是方差）存在对于大数定律的重要性。

## 幻灯片 9: 7.2 弱大数定律 (续): 辛钦弱大数定律

- (展示辛钦弱大数定律页)
- 了解了柯西分布这个特殊的例子，以及它样本均值的奇特行为后，我们继续讨论弱大数定律。前面讲的切比雪夫弱大数定律要求方差有限，现在我们介绍一个条件更宽松的弱大数定律，这就是“定理 7.4: 辛钦弱大数定律”，也叫 Khinchin's WLLN。
- 辛钦弱大数定律对随机变量序列的方差没有要求，它仅仅要求“期望存在”并且序列是“独立同分布的”。
- 具体来说，定理叙述如下：
- 设  $X_1, X_2, \dots, X_n, \dots$  是一列“独立同分布 (i.i.d.)”的随机变量序列。并且，它们共同的期望  $E(X_i) = \mu$  “存在”（也就是说， $\mu$  是一个有限的常数，即  $|\mu| < \infty$ ）。
- 令样本均值为  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ 。
- 那么，对于任意  $\epsilon > 0$ ，结论与切比雪夫 WLLN 相同：

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$$

- 即样本均值  $\bar{X}_n$  依然“依概率收敛”于总体期望  $\mu$  ( $\bar{X}_n \xrightarrow{P} \mu$ )。
- 讲到这里，我们自然要回顾一下前面提到的柯西分布。
  - 如果我们有一列独立同分布的柯西随机变量，它们是 i.i.d. 序列，这没问题。但是，关键在于它们的期望“不存在”。

- 因此，即使是 i.i.d. 的柯西随机变量序列，它们也”不满足”辛钦弱大数定律所要求的”期望存在”这一核心条件。
- 那么结论是什么呢？结论就是，对于这样的序列，它们的样本均值  $\bar{X}_n$  ”不会”依概率收敛到任何常数  $\mu$ 。这与辛钦弱大数定律的结论是相反的，也再次印证了期望存在这个条件的重要性。
- 所以，辛钦弱大数定律虽然放宽了对独立同分布序列的方差要求，但”期望存在”仍然是其成立的基石。

## 幻灯片 10: 7.3 强大数定律 (SLLN)

- (展示强大数定律 - 柯尔莫哥洛夫 SLLN 页)
- 前面我们讨论了弱大数定律 (WLLN)，它描述的是样本均值的”依概率收敛”。现在，我们要学习一种更强的收敛形式，它对应的是”7.3 强大数定律”，英文是 Strong Law of Large Numbers，简称 SLLN。
- 强大数定律 (SLLN) 表明，在特定条件下，样本均值会”几乎必然收敛”（或者也称为以概率 1 收敛）于总体期望。这是一种比依概率收敛更强的收敛模式。与辛钦弱大数定律类似，它通常也要求期望存在。
- 我们主要介绍的是”定理 7.5: 柯尔莫哥洛夫强大数定律”。
- 这个定理的条件与辛钦弱大数定律非常相似：
- 设  $X_1, X_2, \dots, X_n, \dots$  是一列”独立同分布 (i.i.d.)”的随机变量序列，并且它们共同的期望  $E(X_i) = \mu$  ”存在”（即  $|\mu| < \infty$ ）。
- 令样本均值为  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ 。
- 那么，柯尔莫哥洛夫强大数定律的结论是：

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1$$

- 这个式子是什么意思呢？它表示，样本均值序列  $\bar{X}_n$  当  $n$  趋向于无穷大时的”极限”等于总体期望  $\mu$ ——这件事情发生的”概率”是 1。
- 这种收敛方式，我们称之为样本均值  $\bar{X}_n$  ”几乎处处收敛”或”几乎必然收敛”(Almost Surely Converge) 于  $\mu$ ，记作  $\bar{X}_n \xrightarrow{a.s.} \mu$ 。
- 同样地，我们再次回顾一下柯西分布。
  - 对于 i.i.d. 的柯西随机变量序列，它们的期望是”不存在”的。
  - 因此，它们也”不满足”柯尔莫哥洛夫强大数定律所要求的”期望存在”这一条件。

- 结论就是：对于 i.i.d. 的柯西随机变量序列，它们的样本均值  $\bar{X}_n$  ”不会”几乎必然收敛到任何常数。
- 下面小字注释强调：柯尔莫哥洛夫强大数定律的条件（i.i.d. 且期望存在）与我们前面学的辛钦弱大数定律的条件是相同的，但是它的结论（几乎必然收敛）要比弱大数定律的结论（依概率收敛）更强。我们稍后会更详细地讨论这两种收敛的区别。

## 幻灯片 11: WLLN 与 SLLN 的区别与联系 (1): 核心概念与正态分布

- (展示 WLLN 与 SLLN 区别联系 - 第一部分页)
- 好了，我们已经学习了弱大数定律 (WLLN) 和强大数定律 (SLLN)。现在，我们来更深入地探讨一下它们之间的区别与联系。这一页我们先关注核心概念，并以正态分布为例。
- 首先，核心区别在于收敛的类型：
  - **弱大数定律 (WLLN)** 描述的是”依概率收敛”，我们用  $\xrightarrow{P}$  表示。它的数学形式是  $\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$ 。这本质上是说，当  $n$  足够大时，样本均值  $\bar{X}_n$  落在总体期望  $\mu$  的一个小邻域内的”概率”会趋于 1。它关注的是概率的极限。
  - **强大数定律 (SLLN)** 描述的是”几乎必然收敛”，我们用  $\xrightarrow{a.s.}$  表示。它的数学形式是  $P(\lim_{n \rightarrow \infty} \bar{X}_n = \mu) = 1$ 。这说的是，对于几乎所有的样本序列（除去一个概率为 0 的例外集合），样本均值序列  $\bar{X}_n$  本身的”极限值”就是  $\mu$ 。它关注的是样本路径的极限行为。
- 接着，我们看一下它们的条件与对比：
  - 首先，从收敛强度来看，SLLN（几乎必然收敛）要比 WLLN（依概率收敛）”更强”。也就是说，如果一个序列几乎必然收敛，那么它一定依概率收敛。反之则不一定成立。所以， $\xrightarrow{a.s.} \implies \xrightarrow{P}$ 。
  - 其次，我们看到，无论是辛钦 WLLN 还是柯尔莫哥洛夫 SLLN，它们最广泛应用的 i.i.d. 版本都要求”期望存在”（即  $|\mu| < \infty$ ）。这是它们共同的核心条件。
  - 现在，我们来看一个具体的例子：**正态分布**，比如  $N(\mu, \sigma^2)$ 。
    - 对于正态分布，它的期望是  $\mu$ ，这是存在的；它的方差是  $\sigma^2$ ，也是有限的。所以它满足大数定律的条件。
    - 那么结论是什么呢？对于一系列 i.i.d. 的正态随机变量，它们的样本均值  $\bar{X}_n$  既”依概率收敛”于其期望  $\mu$ （根据辛钦 WLLN），也”几乎必然收敛”于其期望  $\mu$ （根据柯尔莫哥洛夫 SLLN）。
- 正态分布是一个很好的例子，它清楚地展现了当大数定律的条件得到满足时，样本均值会如何稳定地趋向总体均值。下一页，我们将对比柯西分布的情况。



## 幻灯片 12: WLLN 与 SLLN 的区别与联系 (2): 柯西分布反例

- (展示 WLLN 与 SLLN 区别联系 - 第二部分, 柯西分布页)
- 这一页我们继续讨论 WLLN 与 SLLN 的区别和联系, 重点来看之前引入的”柯西分布”这个反例, 看看当大数定律的期望存在条件不满足时会发生什么。这个讨论对应的是我们 Beamer 中, 在样本均值路径可视化之后的那张纯文本解释柯西分布行为的幻灯片。
- 首先, 我们再次回顾一下柯西分布的关键特性: 它的”期望不存在”。这是核心。因为期望不存在, 所以它不满足我们前面讨论的辛钦 WLLN 和柯尔莫哥洛夫 SLLN 关于期望存在的前提条件。
- 那么, 对于 i.i.d. 的柯西随机变量序列, 它们的样本均值  $\bar{X}_n$  会怎么样呢? (结合上一张可视化幻灯片的直观印象)
  - 结论是: 这个样本均值  $\bar{X}_n$  ”不”会依概率收敛于任何常数。
  - 并且, 它也”不”会几乎必然收敛于任何常数。
  - 更令人惊讶的是, 这里有一个”惊人事实”: 如果  $X_1, \dots, X_n$  是一列 i.i.d. 服从标准柯西分布的随机变量, 那么它们的样本均值  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  的分布”仍然是标准柯西分布”! 这个结论与样本量  $n$  是无关的!
- 这是什么意思呢? 通常我们期望, 随着样本量  $n$  的增大, 样本均值的分布会越来越集中在总体均值附近 (比如正态分布那样)。但是对于柯西分布, 无论你取多少样本, 它们的平均值的分布形态和弥散程度和单个柯西分布完全一样, 根本不会向某个点”收缩”。
- 正如幻灯片下方斜体字所说: 柯西分布的这种行为与我们前面看到的、满足大数定律条件的正态分布形成了极其鲜明的对比。它非常有力地突出了”期望存在性”对于大数定律成立是何等重要。如果这个基本条件缺失, 大数定律的结论可能就完全不适用了。
- 理解柯西分布这个反例, 对于我们深刻把握大数定律的本质和适用边界非常有帮助。

## 幻灯片 13: 7.4 大数定律的应用

- (展示大数定律的应用页)
- 学习了这么多理论之后, 我们来看看”7.4 大数定律的应用”。大数定律作为概率论与数理统计之间的重要理论基石, 其应用非常广泛。
- 但是, 在讨论具体应用之前, 我们必须再次强调一个”前提”: 所有这些应用, 无论是直接的还是间接的, 都隐式或显式地假设了相关的随机变量序列满足大数定律的条件, 特别是我们反复强调的”期望存在”。如果条件不满足, 那么基于大数定律的推论和应用可能会失效。
- 好了, 我们来看一些主要的应用领域:

### 1. 理论基础与解释现象:

- 首先是”频率的稳定性”。伯努利大数定律为我们用频率来近似概率提供了坚实的理论依据。比如我们抛硬币，抛的次数越多，正面朝上的频率就越接近 0.5。
- 它也解释了为什么在大量重复试验中，许多随机事件的平均结果会趋向于一个稳定的值。

## 2. 统计推断：这是大数定律非常重要的应用领域。

- ”参数估计”：一个核心应用是，样本均值是总体期望的”一致估计量”。这意味着当样本量足够大时，样本均值会非常接近真实的总体期望。这是矩估计法等参数估计方法的基础。但是，这里要特别注意，幻灯片上用红色标出了：”但只有当总体期望存在时才成立！”如果总体期望不存在（比如柯西分布），那么样本均值就不再是总体某个中心位置参数（比如对称中心）的一个好的一致估计量了。
- ”蒙特卡洛方法”：这是一种非常强大的数值计算方法。它通过进行大量的随机抽样，来估计一些难以直接计算的积分、期望等。比如用随机投点法计算圆周率，或者在金融中模拟股票价格路径等，都依赖于”样本均值逼近期望”这个性质。

## 3. 风险管理与保险：保险公司如何厘定保费？它们会利用大数定律来预测在大量保单中可能发生的平均赔付金额，从而确保保费的合理性和公司的经营稳定。这里也隐含了平均损失是存在且可预测的。

## 4. 质量控制与抽样检验：在工业生产中，我们不可能对每一件产品都进行检测。通常是进行抽样检查，然后根据样本的平均质量（比如合格率）来推断整批产品的质量水平。这背后就是样本均值的收敛性在起作用。

## 5. 物理与工程：在进行物理或工程测量时，通常会存在测量误差。为什么我们提倡多次测量取平均？因为根据大数定律，多次测量的平均值通常比单次测量更接近真实值。

- 但是，这里也要特别小心！幻灯片下方的斜体字给出了一个警示：如果测量误差恰好服从柯西分布（这种分布尾部很重，容易出现大的离群值），那么简单地增加测量次数并计算平均值是”没有帮助的”！因为我们已经知道，柯西分布的样本均值不会收敛到真实值，它的分布形状甚至都不会改变！这种情况下就需要更稳健的统计方法了。
- 所以，最后再次强调幻灯片底部的黑体字：”大数定律的条件（尤其是期望存在）是其应用有效性的关键。对于像柯西分布这样期望不存在的情况，基于样本均值的大数定律应用会失效。”我们在实际应用中，一定要对这一点有清醒的认识。

## 幻灯片 14：总结与展望（小结部分）

- (展示总结与展望页 - 小结部分)
- 好的，我们这讲的主要内容就差不多了。现在我们来对本讲进行一个”小结”。
  1. 首先，我们学习了”切比雪夫不等式”。它提供了一个概率的通用上界，但需要随机变量的方差存在且有限。
  2. 其次，我们详细讨论了”大数定律”的核心思想：样本均值在特定条件下会收敛于总体期望。

- 其中，“弱大数定律 (WLLN)”描述的是“依概率收敛”（用  $\xrightarrow{P}$  表示）。我们接触到的辛钦 WLLN 要求随机变量序列是独立同分布 (i.i.d.) 并且期望存在。
  - 而“强大数定律 (SLLN)”描述的是“几乎必然收敛”（用  $\xrightarrow{a.s.}$  表示）。我们学习的柯尔莫哥洛夫 SLLN 也要求 i.i.d. 且期望存在。SLLN 的结论比 WLLN 更强。
3. 第三点，也是我们反复强调的：“条件的重要性”。大数定律的结论是依赖于随机变量满足特定条件的，其中最核心的一个条件就是“期望的存在性”。
  4. 第四点，我们通过一个关键的“反例：柯西分布”来理解条件的重要性。
    - 柯西分布的期望是“不存在”的。
    - 因此，对于 i.i.d. 的柯西变量序列，它们的样本均值并“不服从”大数定律，也就是说，它既不会依概率收敛，也不会几乎必然收敛到任何常数。
    - 特别地，柯西分布的样本均值“仍服从”柯西分布，这个特性非常鲜明地突出了当期望不存在时可能发生的特殊行为。
  5. 最后，我们要认识到，大数定律是连接概率论与统计实践的“核心桥梁”，深刻理解它的内容、条件以及局限性，对于我们正确地学习和应用后续的统计方法至关重要。

## 幻灯片 15: 总结与展望 (展望部分)

- (展示总结与展望页 - 展望部分，保持在同一张幻灯片，内容在 alertblock 中)
- 讲完了小结，我们再对未来做一个简短的“展望”。
  - 首先，大数定律实际上描述的是随机变量序列“平均”行为的极限。
  - 那么，除了平均行为，我们可能还会关心这些随机变量的和或者均值的“分布”会趋向于什么形态。这就引出了概率论中另一类核心的极限定理——“中心极限定理 (CLT)”。中心极限定理告诉我们，在相当广泛的条件下（通常也要求期望和方差存在），大量独立随机变量的和或均值的分布会趋向于正态分布。这是非常非常重要的一个结论，我们将在后续的课程中详细学习。
  - 可以说，大数定律和中心极限定理共同构成了现代统计推断的两大理论基石。它们为我们从样本数据中推断总体信息提供了强大的数学工具。
- 希望通过本讲的学习，大家对大数定律有了更深入的理解，并能体会到其在概率统计中的核心地位。

## 幻灯片 16: 谢谢聆听

- (展示致谢页)
- 好的，以上就是我们第七讲“大数定律”的全部内容。
- 非常感谢大家的聆听！

- 现在是问题与讨论时间，大家如果有什么疑问或者想要交流的想法，都可以提出来。