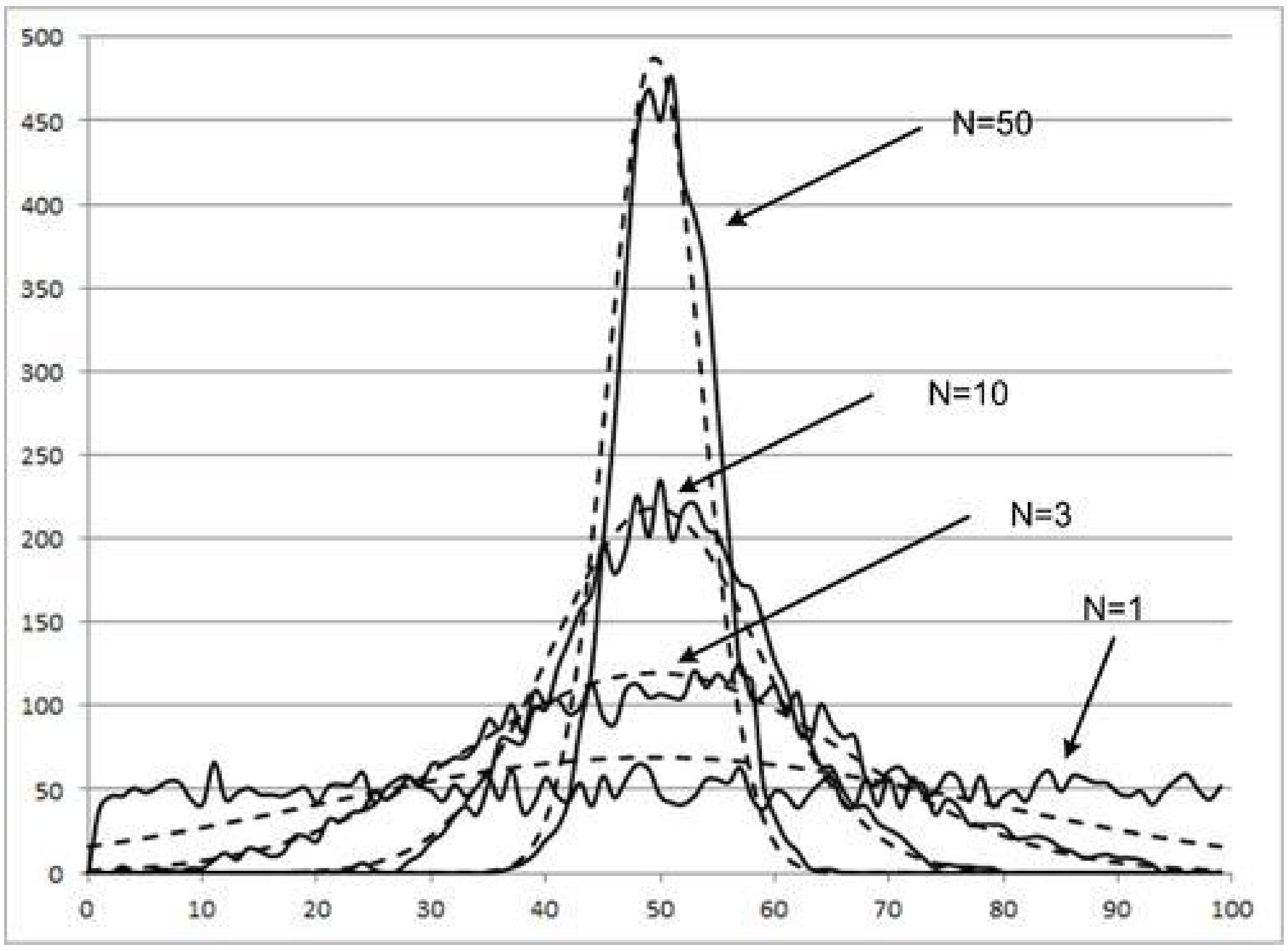

概率论——大数定律与中心极限定理

tetradecane
上海交通大学 工学硕士在读

学习阶段：大学数学。

前置知识：微积分、随机变量、数学期望、方差。

1. 切比雪夫不等式

切比雪夫不等式可以对随机变量偏离期望值的概率做出估计，这是大数定律的推理基础。

以下介绍一个对切比雪夫不等式的直观证明。

1.1 示性函数

对于随机事件A，我们引入一个**示性函数** $I_A=\begin{cases}1, & A\text{发生}\\0, & A\text{不发生}\end{cases}$ ，即一次试验中，若A发生了，则 I 的值为1，否则为0.

现在思考一个问题：这个函数的自变量是什么？

我们知道，随机事件在做一次试验后有一个确定的观察结果，称这个观察结果为**样本点** ω ，所有可能的样本点的集合称为**样本空间** $\Omega=\{\omega\}$. 称 Ω 的一个子集 A 为随机事件。

例如，掷一个六面骰子，记得到数字k的样本点为 ω_k ，则 $\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\}$ ，随机事件“得到的数字为偶数”为 $A=\{\omega_2,\omega_4,\omega_6\}$.

由此可知，示性函数是关于样本点的函数，即

$$I_A(\omega)=\begin{cases}1, & \omega\in A\\0, & \omega\notin A\end{cases}$$ （试验后）

在试验之前，我们能获得哪个样本点也未知的，因此样本点也是个随机事件，记为 ξ ，相应地示性函数可以记为

$$I_A = \begin{cases} 1, & \xi \in A \\ 0, & \xi \notin A \end{cases} \quad \text{（试验前）}$$

在试验之前， I 值也是未知的，因此 I 是个二值随机变量。这样，我们就建立了随机事件A和随机变量 I 之间的一一对应关系。

对 I 求数学期望可得

$$\mathbb{E}I_A = 1 \times P(\xi \in A) + 0 \times P(\xi \notin A) = P(\xi \in A)$$

$P(\xi \in A)$ 是什么？是样本点落在A里面的概率，也就是A事件发生的概率 $P(A)$ ，由此我们就得到了示性函数很重要的性质：其期望值正是对应的随机事件的概率，即

$$\mathbb{E}I_A = P(A)$$

1.2 马尔可夫不等式

对于非负的随机变量 X 和定值 a ，考虑随机事件 $A = \{X \geq a\}$ ，我们可以画出示性函数 I_A 关于观察值x的图像，如图1所示：

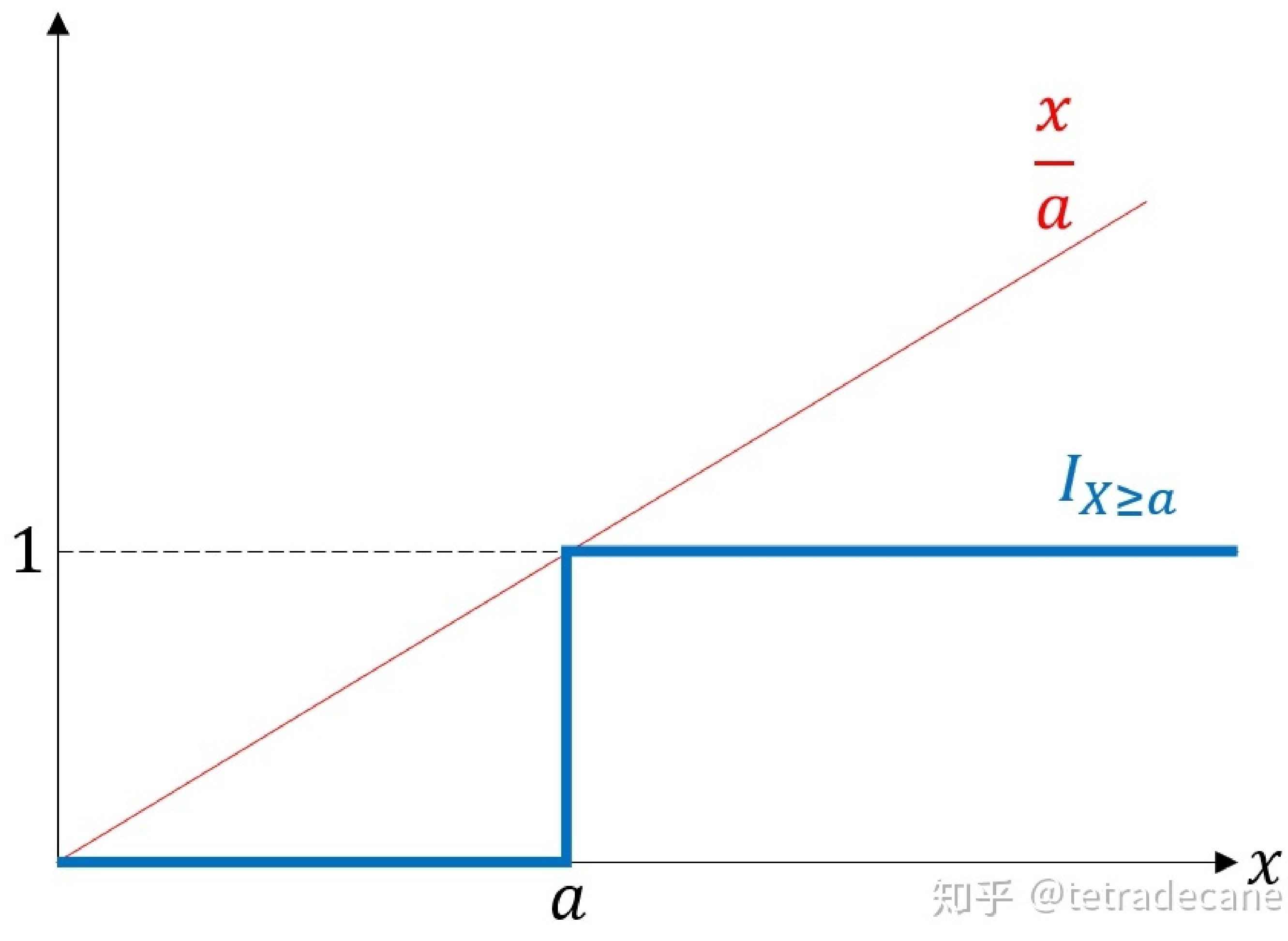

图1

容易发现 $I_{X \geq a}(x) \leq \dfrac{x}{a}$ 恒成立。把x换为随机变量X，再对该式取数学期望得

$$\mathbb{E}I_{X \geq a} = P(X \geq a) \leq \frac{\mathbb{E}X}{a}$$

称该不等式为**马尔可夫Markov不等式。**

从理解上来说，如果非负随机变量X的期望存在，则X超过某个定值a的概率不超过 $\dfrac{\mathbb{E}X}{a}$. 举个简单的例子：如果我们知道所有人收入的平均数a，那么随机抽一个人收入超过10a的概率不超过10%.

1.3 切比雪夫不等式

对于随机变量 X ，记 $\mu=\mathbb{E}X$ ，考虑随机事件 $A=\{|X-\mu|\geq a\}$ ，其示性函数的图像如图2所示：

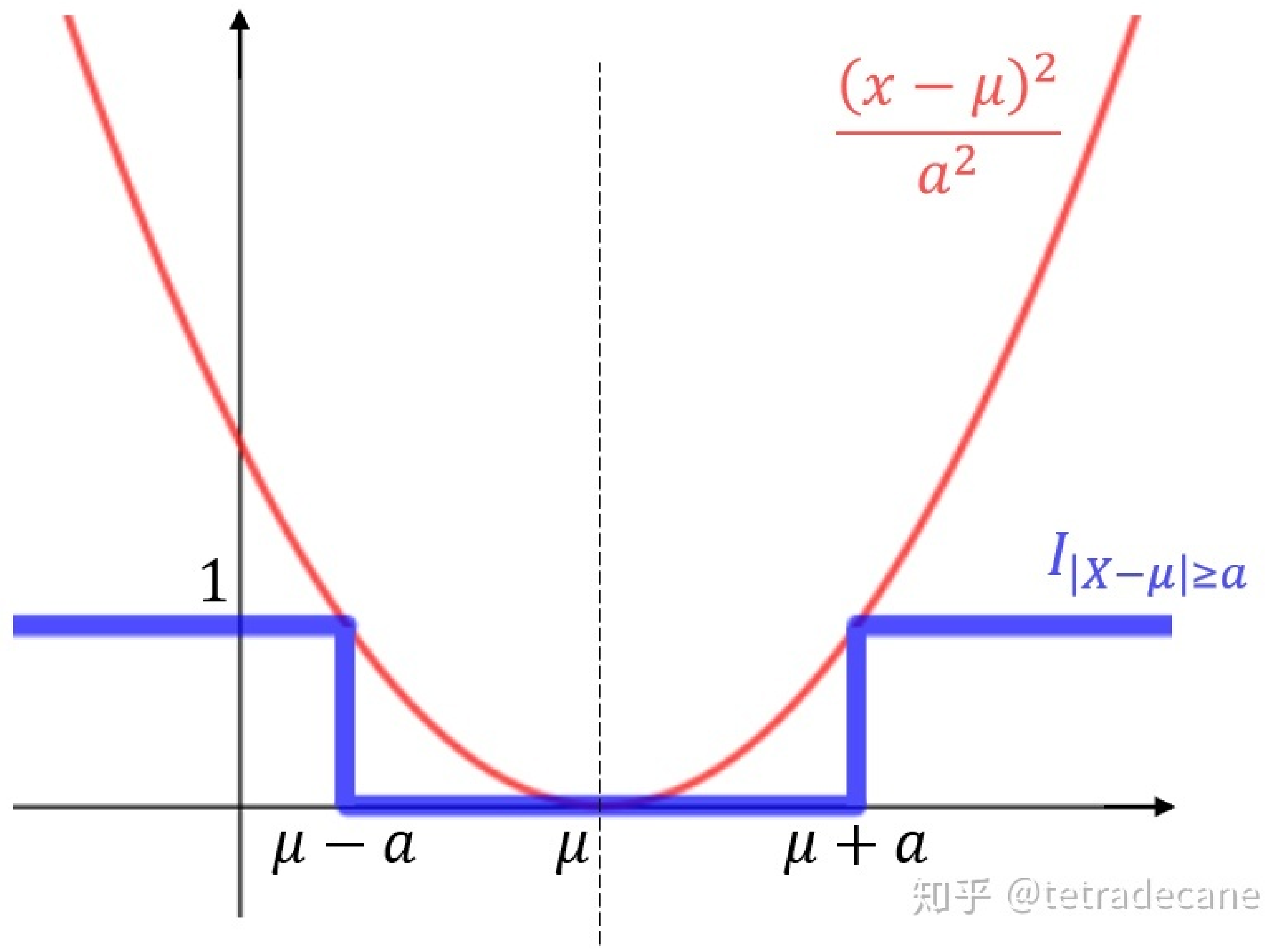

图2

易知 $I_{|X-\mu|\geq a}\leq\dfrac{(x-\mu)^2}{a^2}$ 恒成立。将该式的x换成X并取数学期望得

$$\mathbb{E}I_{|X-\mu|\geq a}=P(|X-\mu|\geq a)\leq\frac{\mathbb{D}X}{a^2}$$

称上面这个不等式为**切比雪夫Chebyshev不等式。**

从理解上来说，如果随机变量X的期望和方差存在，则X和期望值的距离大于a的概率不超过 $\dfrac{\mathbb{D}X}{a^2}$. 给定的范围越大（a越大），或X的方差越小，则偏离的概率越小，这和直觉是相符的。

同样地，切比雪夫不等式对概率的估计也比较粗糙。

以下再给出一个书本上常见的切比雪夫不等式的证明：

记 $p(x)$ 为随机变量 X 的概率密度函数，则

$$P(|X-\mu|\geq a)=\left(\int_{-\infty}^{\mu-a}+\int_{\mu+a}^{+\infty}\right)p(x)dx$$

上式求的是图3中阴影部分的面积。

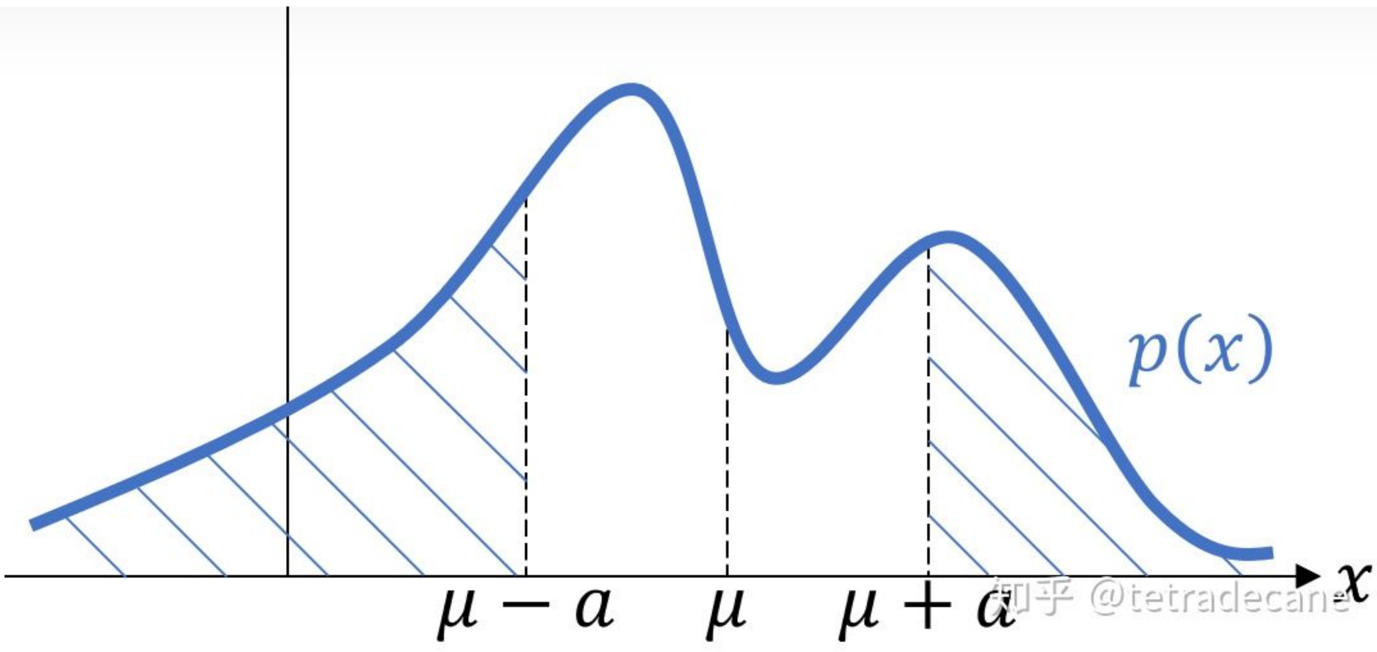

图3

显然，在积分范围内恒有 $\frac{(x-\mu)^2}{a^2} \geq 1$ ，故

$$P(|X-\mu| \geq a) \leq \left(\int_{-\infty}^{\mu-a} + \int_{\mu+a}^{+\infty}\right) \frac{(x-\mu)^2}{a^2} p(x)dx$$

被积函数是非负的，x轴上一部分的积分必然不大于整个x轴上的积分，故

$$P(|X-\mu| \geq a) \leq \int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{a^2} p(x)dx$$

$$= \frac{1}{a^2}\mathbb{E}(X-\mu)^2 = \frac{\mathbb{D}X}{a^2}$$

证毕。

2. 大数定律

对于一系列随机变量 $\{X_n\}$ ，设每个随机变量都有期望。由于随机变量之和 $\sum_{i=1}^{n} X_i$ 很有可能发散到无穷大 ，我们转而考虑随机变量的均值 $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ 和其期望 $\mathbb{E}\left(\overline{X}_n\right)$ 之间的距离。若 $\{X_n\}$ 满足一定条件，当n足够大时，这个距离会以非常大的概率接近0，这就是大数定律的主要思想。

定义：

任取 $\varepsilon > 0$ ，若恒有 $\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mathbb{E}\overline{X}_n\right| < \varepsilon\right) = 1$ ，称 $\{X_n\}$ **服从（弱）大数定律**，称 $\overline{X}_n$ **依概率收敛于** $\mathbb{E}\overline{X}_n$ ，记作

$$\overline{X}_n \xrightarrow{P} \mathbb{E}\overline{X}_n$$

每个“大数定律”其实都是定理，需要证明，只是大家习惯叫他定律罢了。

这里只讨论弱大数定律，并且把弱大数定律简称为大数定律。

2.1 马尔可夫大数定律

任取 $\varepsilon > 0$ ，由切比雪夫不等式知

$$= 1 - \frac{1}{\varepsilon^2 n^2}\mathbb{D}\left(\sum_{i=1}^{n} X_i\right)$$

由此得到**马尔可夫大数定律**：

如果 $\lim\limits_{n\to\infty} \frac{1}{n^2}\mathbb{D}\left(\sum_{i=1}^{n} X_i\right) = 0$ ，则 $\{X_n\}$ 服从大数定律。

2.2 切比雪夫大数定律

在马尔可夫大数定律的基础上，如果 $\{X_n\}$ 两两不相关，则方差可以拆开：

$$\frac{1}{n^2}\mathbb{D}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{D}X_i$$

如果 $\mathbb{D}X_i$ 有共同的上界c，则

$$\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{D}X_i \le \frac{nc}{n^2} = \frac{c}{n}$$

$$P\left(\left|\overline{X}_n - \mathbb{E}\overline{X}_n\right| < \varepsilon\right) \ge 1 - \frac{c}{\varepsilon^2 n}$$

由此得到**切比雪夫大数定律**：

如果 $\{X_n\}$ 两两不相关，且方差有共同的上界，则 $\{X_n\}$ 服从大数定律。

2.3 独立同分布大数定律

在切比雪夫大数定律的基础上，进一步限制 $\{X_n\}$ 独立同分布，立刻得到**独立同分布大数定律**：

如果 $\{X_n\}$ 独立同分布且方差有界，则 $\{X_n\}$ 服从大数定律，即

$$\overline{X}_n \xrightarrow{P} \mathbb{E}\overline{X}_n = \mathbb{E}X$$

2.4 伯努利大数定律

根据经验，在做了大量独立重复实验后，某随机事件A发生的频率与概率往往会十分接近，这正是大数定律在发挥作用。

记第k次试验中A的示性函数为 $I_{A,k}$ ，则所有n次试验中A发生的频数 是 $\sum\limits_{i=1}^{n} I_{A,k}$ ，频率是 $\frac{1}{n}\sum\limits_{i=1}^{n} I_{A,k}$ ，易知

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} I_{A,k}\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}I_{A,k} = \frac{nP(A)}{n} = P(A)$$

又知这n个 $I_{A,k}$ 独立同分布且方差有界，由独立同分布大数定律知 $\{I_{A,k}\}$ 服从大数定律，这就是**伯努利Bernoulli大数定律**：

记 n_A 为n次伯努利实验中事件A发生的次数，记p为事件A发生的概率，则

伯努利大数定律是最早被发现的大数定律，因为这是生活中最容易发现的规律。

2.5 辛钦大数定律

以上2.1至2.4的大数定律都对 $\{X_n\}$ 的方差有所约束，而接下来的**辛钦Khinchin大数定律**可以完全不考虑方差：

如果 $\{X_n\}$ 独立同分布且具有有限的数学期望 $\mathbb{E}X$ ，则 $\{X_n\}$ 服从大数定律。

这个定理的证明较复杂，此处不予证明。

3. 中心极限定理

大数定律研究的是一系列随机变量 $\{X_n\}$ 的均值 $\overline{X}_n=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$ 是否会依概率收敛于其期望 $\mathbb{E}\overline{X}_n$ 这个数值，而中心极限定理进一步研究 $\overline{X}_n$ 服从什么分布。若 $\{X_n\}$ 满足一定的条件，当n足够大时， $\overline{X}_n$ 近似服从正态分布，这就是中心极限定理的主要思想，这也体现了正态分布的重要性与普遍性。

3.1 林德贝格-勒维/独立同分布中心极限定理

如果 $\{X_n\}$ 独立同分布，且 $\mathbb{E}X=\mu,\quad \mathbb{D}X=\sigma^2>0$ ，则n足够大时 $\overline{X}_n$ 近似服从正态分布 $N\left(\mu,\dfrac{\sigma^2}{n}\right)$ ，即

$$\lim_{n\to\infty}P\left(\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}<a\right)=\Phi(a)=\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$$

上述定理就是**林德贝格-勒维Lindeberg-Levy中心极限定理**，又称**独立同分布中心极限定理**。

这个定理的证明也比较复杂，此处不予证明。

这个定理是容易理解、记忆的。首先记住 $\{X_n\}$ 的均值 $\overline{X}_n$ 近似服从正态分布，接下来只需要解出这个正态分布的期望和方差。期望有

$$\mathbb{E}\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}X_i=\frac{n\mu}{n}=\mu$$

方差有

$$\mathbb{D}\overline{X}_n=\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{D}X_i=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}$$

那么 $\overline{X}_n$ 近似服从的正态分布就是 $N\left(\mu,\dfrac{\sigma^2}{n}\right)$ ，归一化后的随机变量 $\dfrac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}$ 近似服从标准正态分布 $N(0,1)$.

3.2 棣莫弗-拉普拉斯/二项分布中心极限定理

棣莫弗-拉普拉斯De Moivre-Laplace中心极限定理是独立同分布中心极限定理的特殊情况，它是最先被发现的中心极限定理。

$\mathbb{E}\xi_n = np, \quad \mathbb{D}\xi_n = np(1-p)$ ，当n足够大时 ξ_n 近似服从正态分布 $N(np, np(1-p))$ ，即

$$\lim_{n\to\infty} P\left(\frac{\xi_n - np}{\sqrt{np(1-p)}} < a\right) = \Phi(a)$$

该定理表明：当试验次数n足够大时，二项分布近似于正态分布。

*3.3 独立不同分布下的中心极限定理

长度、重量、时间等等实际测量量一般符合正态分布，因为它们受各种微小的随机因素的扰动。这些随机因素的独立性是很普遍的，但很难说它们一定同分布。

实际上，一系列独立不同分布的随机变量也可能满足中心极限定理，只是这些不同分布的随机变量要有所限制。以下给出两个独立不同分布下的中心极限定理，不予证明，仅供欣赏：

林德伯格中心极限定理

设 $\{X_n\}$ 是一系列相互独立的连续随机变量，它们具有有限的期望 $\mathbb{E}X_i = \mu_i$ 和方差 $\mathbb{D}X_i = \sigma_i^2$ ，记 $Y_n = \sum_{i=1}^n X_i, \quad \mathbb{D}Y_n = \sum_{i=1}^n \sigma_i^2 = B_n^2$ ，记 X_i 的密度函数是 $p_i(x)$ ，若

$$\forall \tau > 0 : \lim_{n\to\infty} \frac{1}{\tau^2 B_n^2} \sum_{i=1}^n \int_{|x-\mu_i|>\tau B_n} (x-\mu_i)^2 p_i(x) dx = 0$$

则

$$\lim_{n\to\infty} P\left(\frac{1}{B_n}\sum_{i=1}^n (X_i - \mu_i) < a\right) = \Phi(a)$$

林德伯格中心极限定理对 $\{X_n\}$ 的约束基本上是最弱的，也就是最强的中心极限定理。然而该定理的条件较难运用与验证，以下的定理是它的特例：

李雅普诺夫Lyapunov中心极限定理

设 $\{X_n\}$ 是一系列相互独立的随机变量，，若

$$\exists \delta > 0 : \lim_{n\to\infty} \frac{1}{B_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}\left(|X_i - \mu_i|^{2+\delta}\right) = 0$$

则

$$\lim_{n\to\infty} P\left(\frac{1}{B_n}\sum_{i=1}^n (X_i - \mu_i) < a\right) = \Phi(a)$$

李雅普诺夫中心极限定理的条件在很多情况下是满足的，因此适用性也很广。

编辑于 2021-07-15 01:32