第五章：大数定律和中心极限定理

问题引入 I

考虑：一个接一个的检查产品的合格情况。

记X_i为第i个产品的不合格数，$X_i \sim B(1,p)$。

记$S_n = \sum_{i=1}^n X_i$，以及n次检查的不合格率$v_n = \frac{S_n}{n}$，考虑下面两个问题

1. 是否存在实数p，使得$\lim_{n\to+\infty} v_n = p$(弱大数定律)
2. 是否存在某个随机变量Y，使得$\lim_{n\to\infty} S_n = Y$？(依分布收敛(中心极限定理))

弱大数定律

若对任意的$\epsilon > 0$，

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

则称$X_1, X_2, \cdots$服从**弱大数定律**。

n重伯努利试验 I

对于n重伯努利试验，弱大数定律阐述的是下面的事实，即

概率是频率的极限

$$S_n = \sum_{i=1}^{n} X_i, \quad X_i = \begin{cases} 1, & \text{成功} \\ 0, & \text{失败} \end{cases}$$

对$\forall \epsilon > 0$

$$\lim_{n\to\infty} P\left(\omega : \left|\frac{S_n}{n} - p\right| > \epsilon\right) = 0.$$

几个常用的弱大数定律

1. 伯努利大数定律
2. 切比雪夫大数定律
3. 辛钦大数定律
4. 马尔可夫大数定律

各类型的弱大数定律之间只有适用条件的不同，表达公式均为(1)

切比雪夫不等式 I

切比雪夫不等式

设随机变量X的方差存在，则对任意的$\epsilon > 0$，有

$$\boldsymbol{P(|X - E[X]| \geq \epsilon) \leq \frac{Var[X]}{\epsilon^2}}$$

切比雪夫不等式 II

证明：对于连续型随机变量

$$P(|X - E[X]| \geq \epsilon) = \int_{|x-E[X]|\geq\epsilon} f_X(x)\,\mathrm{d}x \tag{2}$$

$$\leq \int_{|x-E[X]|\geq\epsilon} \frac{(x - E[X])^2}{\epsilon^2} f_X(x)\,\mathrm{d}x \tag{3}$$

$$\leq \int_{-\infty}^{\infty} \frac{(x - E[X])^2}{\epsilon^2} f_X(x)\,\mathrm{d}x \tag{4}$$

$$= \frac{Var[X]}{\epsilon^2} \tag{5}$$

切比雪夫不等式 III

对于离散型随机变量

$$
\begin{aligned}
P(|X - E[X]| \geq \epsilon) &= \sum_{|x_i - E[X]| \geq \epsilon} p_i \\
&\leq \sum_{|x_i - E[X]| \geq \epsilon} \frac{(x_i - E[X])^2}{\epsilon^2} p_i \\
&\leq \sum_i \frac{(x_i - E[X])^2}{\epsilon^2} p_i \\
&= \frac{Var[X]}{\epsilon^2}
\end{aligned}
$$

切比雪夫不等式的等价表述

1. 随机变量的波动幅度是由标准差控制的

$$P(|X - E[X]| \geq k\sqrt{Var[X]}) \leq \frac{1}{k^2}$$

2.

$$P(|X - E[X]| \leq \epsilon) \geq 1 - \frac{Var[X]}{\epsilon^2}$$

练习

1. 设随机变量X的数学期望$E[X]=\mu$，方差$Var[X]=\sigma^2$，那么由切比雪夫不等式估计$P(|X-\mu|\geq 3\sigma)$
2. 设X与Y为随机变量，且满足$E[X]=-2$，$E[Y]=2$，$Var[X]=1$，$Var[Y]=9$，且相关系数$r(X,Y)=-0.5$，请根据切比雪夫不等式确定满足不等式$P(|X+Y|\geq 6)\leq c$的最小正数

伯努利大数定律

由切比雪夫不等式可以直接推出

伯努利大数定律

记S_n为n重**伯努利试验**中的成功次数，p为一次试验成功的概率，那么

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n}-p\right|\geq\epsilon\right)=0$$

伯努利大数定律使用条件是:$X_1, X_2, \cdots$为独立同分布的两点分布。

设$X_1, X_2, \cdots, X_n$为**独立**的随机变量序列，$\boldsymbol{E[X_i] = \mu}$，$\boldsymbol{Var[X_i] \leq C}$，$i = 1, 2, \cdots,$，则对任意的$\epsilon > 0$，有

$$\boldsymbol{\lim_{n \to \infty} P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) = 0}$$

证明 记随机变量$Y_n = \frac{X_1+X_2+\cdots+X_n}{n}$，那么由期望的线性性

$$E[Y_n] = \mu$$

切比雪夫弱大数定律 II

有$X_1, X_2, \cdots$的独立性，有

$$Var[Y_n] = \frac{\sum_{i=1}^{n} Var[X_i]}{n^2} \leq \frac{C}{n}$$

由切比雪夫不等式

$$0 \leq \lim_{n\to\infty} P(|Y_n - \mu| \geq \epsilon) \leq \lim_{n\to\infty} \frac{C}{n\epsilon} = 0$$

切比雪夫弱大数定律的使用条件是

1. 随机变量序列$X_1, X_2, \cdots$相互独立,
2. X_i有相同的均值（并不一定同分布），
3. X_i的方差有公共的上界。

辛钦弱大数定律 I

当随机变量X_i的方差不一定存在时，要使用下面的辛钦弱大数定律

辛钦弱大数定律

设$X_1, X_2, \cdots$为**独立同分布**的随机变量序列，且具有**相同**的数学期望μ，则对于任意的$\epsilon > 0$有

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) = 0$$

辛钦弱大数定律 II

辛钦弱大数定律的使用条件是

1. 随机变量序列$X_1, X_2, \cdots$相互独立同分布,
2. X_i的均值存在

马尔可夫大数定律

对于任意的随机变量序列，若有

$$\lim_{n\to\infty}\frac{1}{n^2}Var[\sum_{i=1}^{n}X_i]=0$$

成立，那么$\{X_n\}$服从弱大数定律。

马尔可夫弱大数定律是适用最广泛的弱大数定律，因为没有对随机变量序列独立性，同分布性，或者均值方差有界性的要求。

例子

1. 设X_k为独立的随机变量序列，且

$$P(X_k=\pm 2^k)=\frac{1}{2^{2k+1}},\quad P(X_k=0)=1-\frac{1}{2^{2k}},\ k=1,2,\cdots$$

证明$\{X_k\}$服从弱大数定律。

2. 设$\{X_n\}$为独立的随机变量序列，且

$$P(X_n=1)=p_n,\quad P(X_n=0)=1-p_n,\ n=1,2,\cdots$$

证明$\{X_n\}$服从弱大数定律。

3. 设$\{X_n\}$ 是独立的随机变量序列，且假设

$$P(X_n=\sqrt{\ln n})=P(X_n=-\sqrt{\ln n})=0.5,\ n=1,2,\cdots$$

请问$\{X_n\}$是否服从大数定律。

中心极限定理

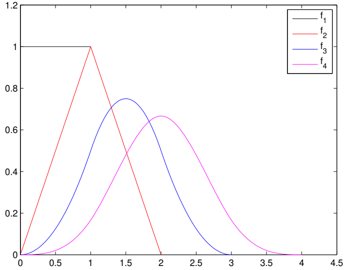
$X_1, X_2, \cdots, X_n$独立同分布，问

$$Y_n = X_1 + X_2 + \cdots + X_n$$

的分布函数是什么？当$n \to \infty$，这个极限分布是什么？

中心极限定理引入 I

设$\{X_n\}$为独立同分布的随机变量序列，且$X_i \sim U(0,1)$，记$Y_n = \sum_{i=1}^{n} X_i$，设Y_1, Y_2, Y_3, Y_4的分布密度函数分别为f_1, f_2, f_3, f_4，有



中心极限定理引入 II

随着n的增加，f_n越来越接近正态分布的密度函数。但$\sum_{i=1}^{n} X_i$的均值与方差都随着n变大而趋于无穷大。

问题提法

当$n \to \infty$，$E[Y_n] \to \infty$，$Var[Y_n] \to \infty$，即Y_n不收敛到一个有限的分布。我们对Y_n作标准化

$$\boldsymbol{Y_n^* = \frac{Y_n - E[Y_n]}{\sqrt{Var[Y_n]}}}$$

$E[Y_n^*] = 0$，$Var[Y_n^*] = 1$，此时Y_n^*就有可能用标准正态分布去代替，即

$$\boldsymbol{Y_n^* \to Z \sim N(0,1)}$$

依分布收敛

对于随机变量Z的分布函数$F_Z(a)$的任意连续点$a \in \mathbb{R}$，都有

$$\lim_{n\to\infty} F_{Z_n}(a) = F_Z(a)$$

则称随机变量序列$\{Z_n\}$依分布收敛到Z，记作$Z_n \xrightarrow{d} Z$。

中心极限定理 I

定义5.2.1

设$X_1, X_2, \cdots, X_n$为**独立的**随机变量序列，具有有限的数学期望和方差，记$Y_n = \sum_{i=1}^{n} X_i$，若有

$$Z_n = \frac{Y_n - E[Y_n]}{\sqrt{Var[Y_n]}} \xrightarrow{d} Z,\ Z \sim N(0,1) \tag{6}$$

则称$X_1, X_2, \cdots, X_n$服从中心极限定理。

L-L定理

下面的定理被称为**林德伯格－莱维**中心极限定理，它是最常使用的中心极限定理。

定理5.2.1

设$X_1, X_2, \cdots$,为**独立同分布**的随机变量序列，具有**有限并相同的**的数学期望μ和方差σ^2，那么$X_1, X_2, \cdots$,服从中心极限定理，

$$\lim_{n\to\infty} P\left(\frac{1}{\sigma\sqrt{n}}[\sum_{k=1}^{n} X_k - n\mu] \leq x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{s^2}{2}}\,\mathrm{d}s$$

课堂练习

例题5.2.1已知$X_1, X_2, \cdots$独立同分布，且$E[X_i]=1$，$Var[X_i]=4$，求$P(X_1+X_2+\cdots+X_{100}\leq 125)$

习题5.7 一仪器同时收到50个信号X_k，$k=1,2,\cdots,50$，设$X_1, X_2, \cdots, X_{50}$独立同分布，且$X_i \sim U[0,9]$，求$P(\sum_{k=1}^{50} X_k > 250)$的近似值。

棣莫弗－拉普拉斯定理

设$X_1, X_2, \cdots,$为**独立同分布**的随机变量序列，且都服从$\boldsymbol{B(1,p)}$，那么$X_1, X_2, \cdots,$服从中心极限定理，

$$\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n} X_k - np}{\sqrt{np(1-p)}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{s^2}{2}}\, \mathrm{d}s$$

这个定理与泊松定理都是对二项分布的近似，当$p \ll 1$的时候，用泊松分布近似较好，当p并不十分小的时候，用正态分布较好。