1 大数定理和中心极限定理

1.1 大数定理

在抛一次硬币试验中，设随机变量X为正面出现的次数，则X的可能取值为0,1.

$$P(X=1)=\frac{1}{2}.$$

我们怎么知道概率是二分之一的呢？

假设抛了n次硬币，每次观察正面出现的次数, 记随机变量X_i为第i次抛硬币正面出现的次数. 则n次试验中正面一共出现$X_1+X_2+\cdots+X_n$次，频率为

$$p_n=\frac{X_1+X_2+\cdots+X_n}{n}=\overline{X}_n$$

当n很大时, **'频率趋向概率'**

$$\text{当n很大时，}\quad \overline{X}_n\text{趋向于}X_i\text{数学期望}$$

定理1.1 (大数定理)**.** $X_1,X_2,\cdots,X_n$ **独立同分布**，数学期望为μ, 方差σ^2, 算数平均值$\overline{X}_n=\sum_{i=1}^n X_i/n$, 则$\forall\epsilon>0$

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

注1.1. 上述定理说明，当 n 充分大时，$\overline{X}_n$ 与 μ 的偏离总可以充分小。或者说当 n 充分大时，我们有大概率说$\overline{X}_n$ 很靠近 μ。例如上面抛硬币，当 n 充分大时，我们有大概率说$\overline{X}_n$ 很靠近 $\frac{1}{2}$. 但不排除你抛了*1000*次，都是反面(只是这种可能性很小). 如果试验或者抽样是随机的，则这样的可能性会随n增大而减小.

证明需要**切比雪夫不等式**

$$P\{|X-E(X)|\le\epsilon\}\ge 1-\frac{\sigma^2}{\epsilon^2}$$

定义1.1 (依概率收敛)**.** $Y_1,Y_2,\cdots,Y_n$ 是一个随机变量序列, a为常数。若$\forall\epsilon>0$

$$\lim_{n\to\infty}P(|Y_n-a|<\epsilon)=1$$

则称Y_n依概率收敛到a,记为

$$Y_n\xrightarrow{P}a$$

定理1.2 (伯努利大数定理)**.** 设n_A是n次独立重复试验中事件A发生的次数，p是事件A在每次试验中发生的概率，则对于任意正数$\epsilon>0$ ，有

$$\lim_{n\to\infty}P(|\frac{n_A}{n}-p|<\epsilon)=1$$

定理1.3 (辛钦大数定理)**.**

$X_1,X_2,\cdots,X_n$ **独立同分布**，数学期望为μ, 算数平均值$\overline{X}_n=\sum_{i=1}^n X_i/n$, 则$\forall\epsilon>0$

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

辛钦大数定理最大的区别是**不需要方差**.(某些分布有期望但方差趋向无穷大)

辛钦大数定律为寻找随机变量的期望值提供了一条实际可行的途径. 要估计某地区的平均亩产量，要收割某些有代表性块，例如n 块地. 计算其平均亩产量，则当n较大时，可用它作为整个地区平均亩产量的一个估计.

Q.在一个罐子中,装有10个编号为0-9的同样的球，从罐中有放回地抽取若干次，每次抽一个，并记下号码. 设

$$X_k=\begin{cases}1, & \text{第k次取到号码6}\\ 0 & \text{其他}\end{cases}$$

问$\{X_k\}$ 是否能应用大数定理？

1.2 中心极限定理

Q. 已知

$$f_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

求$f_{100000}(1)$的近似值.

显然, 这个近似值就是 e. 概率论上有类似的结论:

定理1.4 (中心极限定理).

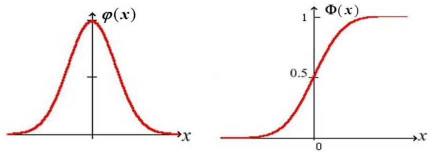
$X_1, X_2, \cdots, X_n$ 独立同分布，数学期望为μ, 方差σ^2, 则$\forall\ x$

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^n X_n - n\mu}{\sqrt{n}\sigma} \le x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt$$

$\sum_{i=1}^n X_n$ 期望为$n\mu$, 方差$n\sigma^2$, 故

$$\frac{\sum_{i=1}^n X_n - n\mu}{\sqrt{n}\sigma}$$

就是$\sum_{i=1}^n X_n$ 的标准化, 它恰好对应于标准的正态分布$N(0,1)$。**我们虽然不知道$X_1 + X_2 + \cdots + X_n$的分布是什么，但当n充分大时，可以用正态分布去近似。**



事实上, 世界上最早的中心极限定理是棣莫佛－拉普拉斯定理(二项分布), 棣莫佛讨论了$p = \frac{1}{2}$的情况，而拉普拉斯将它推广到一般的p.

二项分布$X \sim B(n,p)$可以看作n个独立试验，即引入随机变量$X_1, X_2, \cdots, X_n$, 其中

$$X_i = \begin{cases} 1, & \text{第i次试验A发生} \\ 0, & \text{第i次试验A不发生} \end{cases}$$

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

X_i 为0-1分布，其均值和方差为

$$E(X_i) = p,\ D(X_i) = E(X_i^2) - [E(X_i)]^2 = p - p^2 = p(1-p)$$

定理1.5 (棣莫佛－拉普拉斯定理(De Laplace)).

设随机变量$X_n (n = 1, 2, \cdots)$服从参数(n,p) 的二项分布，则

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

算术平均的n为1

一加法器同时收到20个噪声电压$V_k (k = 1,2,\cdots n)$, 设它们是相互独立的随机变量，且都在区间(0, 10)上服从均匀分布.记$V = \sum_{k=1}^{n} V_k$，求$P\{V > 105\}$的近似值.

解 易知$E(V_k) = 5, D(V_k) = 100/12 \quad (k = 1,2,\cdots 20)$.

$$V = \sum_{k=1}^{20} V_k \overset{\text{近似地}}{\sim} N(20\times 5, \frac{100}{12}\times 20)$$

于是
$$P\{V > 105\} = p\left\{\frac{V - 20\times 5}{(\sqrt{100/12})\sqrt{20}} > \frac{105 - 20\times 5}{(\sqrt{100/12})\sqrt{20}}\right\}$$

$$= p\left\{\frac{V - 20\times 5}{(\sqrt{100/12})\sqrt{20}} > 0.387\right\} = 1 - p\left\{\frac{V - 20\times 5}{(\sqrt{100/12})\sqrt{20}} \le 0.387\right\}$$

$$\approx 1 - \Phi(0.387) = 0.348$$

Q.(供电问题)某车间有200台车床,在生产期间由于需要检修、调换刀具、变换位置及调换工件等常需停车. 设开工率为0.6, 设每台车床的工作是独立的，且在开工时需电力1千瓦.
问应供应多少瓦电力就能以99.9%的概率保证该车间不会因供电不足而影响生产?

解：用X表示在某时刻工作着的车床数，二项分布$X \sim B(200, 0.6)$，设设需N台车床工作，求

$$P(X \le N) > 0.999 \quad \text{最小的N}$$

由中心极限定理

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^n X_n - np}{\sqrt{np(1-p)}} \le x\right) = \Phi(x)$$

$$\Phi(\frac{N - 120}{\sqrt{48}}) > 0.999 = \Phi(3.1) \Longrightarrow N \ge 141.5$$

Q. 家长会的家长人数为随机变量X_i(独立同分布), 无家长，一个家长，两个家长的概率分别为0.05，0.8，0.15。若有400学生，求来的家长超过450的概率(这样你好安排教室)。

$$P\{X > 450\} = P\left\{\frac{X - 400\times 0.8}{\sqrt{400}\sqrt{0.19}} > \frac{450 - 400\times 0.8}{\sqrt{400}\sqrt{0.19}}\right\}$$

$$= 1 - P\left\{\frac{X - 400\times 0.8}{\sqrt{400}\sqrt{0.19}} \le 1.147\right\}$$

$$\approx 1 - \Phi(1.147) = 0.1257$$

$$\frac{X - nE(x)}{\sqrt{n}\sqrt{D(x)}}$$