§4.4 大数定律及中心极限定理

概率论与数理统计是研究随机现象统计规律性的学科，随机现象的规律性只有在相同的条件下进行大量重复试验时才会呈现出来.

研究大量的随机现象，常常采用极限形式，由此导致对极限定理进行研究. 极限定理的内容很广泛，其中最重要的有两种——大数定律与中心极限定理.

大数定律的客观背景

大量的随机现象中平均结果的稳定性

大量抛掷硬币正面出现频率

生产过程中的废品率

字母使用频率

…… ……

一　问题的提出

从第一章我们知道：虽然随机事件在某次试验中是否发生带有偶然性，但在大量的重复试验中却呈现明显的规律性——频率的稳定性. 如果以μ_n记n次贝努利试验中事件A出现的次数，则μ_n/n是事件A出现的频率，频率的稳定性就是指当n增大时μ_n/n与$P(A)$(某个固定的常数)越来越靠近. 这说明了频率所具有的极限性质, 这类极限问题, 我们还没有给出理论上的的刻画. 历史上，雅·贝努利第一个对这种极限进行了研究.

第一节　依概率收敛与大数定律

一　问题的提出

从第一章我们知道：虽然随机事件在某次试验中是否发生带有偶然性，但在大量的重复试验中却呈现明显的规律性——频率的稳定性. 如果以μ_n记n次贝努利试验中事件A出现的次数，则μ_n/n是事件A出现的频率，频率的稳定性就是指当n增大时μ_n/n与$P(A)$(某个固定的常数)越来越靠近. 这说明了频率所具有的极限性质, 这类极限问题, 我们还没有给出理论上的的刻画. 历史上，雅·贝努利第一个对这种极限进行了研究.

下面先讨论频率μ_n/n的性质：

μ_n是随机变量，它服从二项分布$B(n,p)$.

$$P(\mu_n = k) = C_n^k p^k (1-p)^{n-k}，\ k = 0,1,2,\cdots,n.$$

因而μ_n/n也是随机变量，且

$$P(\mu_n/n = k/n) = C_n^k p^k (1-p)^{n-k}，\ k = 0,1,2,\cdots,n.$$

由此可见：从理论上讲，频率μ_n/n全部可能的取值为

$$0,\frac{1}{n},\cdots,\frac{n-1}{n},1.$$

仅从取值本身看，没有什么极限性质，但从上面我们知道,μ_n/n取这些值的概率有很大的不同．因为

$$E(\mu_n/n) = p, \quad D(\mu_n/n) = pq/n,$$

如何描述 $\frac{\mu_n}{n} \longrightarrow p$

数列极限的定义

数　　列　　$x_n \longrightarrow a$

数列极限的描述：

$\forall \varepsilon > 0$，$\exists$正整数N，当$n > N$时，总有

$$|x_n - a| < \varepsilon$$

$\frac{\mu_n}{n}$与x_n的区别：

（1） $\{x_n\}_1^{\infty}$是数列：给定n，x_n是确定的实数。

$\left\{\frac{\mu_n}{n}\right\}_1^{\infty}$是随机变量序列：给定$n$， $\frac{\mu_n}{n}$不是确定的实数，

取值为$0,\frac{1}{n},\cdots,\frac{n-1}{n},1$

（2） $\forall\varepsilon>0$， 存在N， 当 $n>N$时， $|x_n-a|<\varepsilon$一定成立.

$\forall\varepsilon>0$， $\left|\frac{\mu_1}{1}-p\right|<\varepsilon,\cdots\cdots,\left|\frac{\mu_n}{n}-p\right|<\varepsilon,\cdots\cdots$是事件序列：

无论N多么大， 当$n>N$时，事件$\left|\frac{\mu_n}{n}-p\right|<\varepsilon$ 都可能不发生。

我们可以研究数列$P\left(\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right)$的性质：

$$P\left(\left|\frac{\mu_1}{1}-p\right|<\varepsilon\right),P\left(\left|\frac{\mu_2}{2}-p\right|<\varepsilon\right),\cdots\cdots,P\left(\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right),\cdots$$

一种提法是：当n足够大时，频率μ_n/n与概率p有较大偏差的可能性（概率）很小．用数学语言来讲，就是要证明：对于任意$\varepsilon>0$

$$\lim_{n\to+\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|\geq\varepsilon\right\}=0,$$

或者与它等价的式子成立，即

$$\lim_{n\to+\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}=1.$$

下面先讨论频率μ_n/n的性质：

μ_n是随机变量，它服从二项分布$B(n,p)$.

$$P(\mu_n=k)=C_n^k p^k(1-p)^{n-k}，\ k=0,1,2,\cdots,n.$$

因而μ_n/n也是随机变量，且

$$P(\mu_n/n=k/n)=C_n^k p^k(1-p)^{n-k}，\ k=0,1,2,\cdots,n.$$

由此可见：从理论上讲，频率μ_n/n全部可能的取值为

$$0,\frac{1}{n},\cdots,\frac{n-1}{n},1.$$

仅从取值本身看，没有什么极限性质，但从上面我们知道，μ_n/n取这些值的概率有很大的不同．因为

$$E(\mu_n/n)=p,\quad D(\mu_n/n)=pq/n,$$

所以当$n \to +\infty$时，频率的数学期望保持不变，而方差趋于0. 我们知道方差为零的随机变量是常数，于是我们自然预期频率将趋于常数p（即A发生的概率），那么如何描述随机变量的极限？

一种提法是：当n足够大时，频率μ_n/n与概率p有较大偏差的可能性（概率）很小. 用数学语言来讲，就是要证明：对于任意$\varepsilon > 0$

$$\lim_{n \to +\infty} P\left\{\left|\frac{\mu_n}{n} - p\right| \geq \varepsilon\right\} = 0,$$

或者与它等价的式子成立，即

$$\lim_{n \to +\infty} P\left\{\left|\frac{\mu_n}{n} - p\right| < \varepsilon\right\} = 1.$$

$$\mu_n \sim B(n,p):$$

随机变量序列： $\mu_1, \mu_2, \cdots\cdots, \mu_n, \cdots\cdots$

随机变量序列： $\frac{\mu_1}{1}, \frac{\mu_2}{2}, \cdots\cdots, \frac{\mu_n}{n}, \cdots\cdots$

$\forall \varepsilon > 0$

事件序列： $\left|\frac{\mu_1}{1} - p\right| < \varepsilon, \left|\frac{\mu_2}{2} - p\right| < \varepsilon, \cdots\cdots, \left|\frac{\mu_n}{n} - p\right| < \varepsilon, \cdots\cdots$

数列： $P(\left|\frac{\mu_1}{1} - p\right| < \varepsilon), P(\left|\frac{\mu_2}{2} - p\right| < \varepsilon), \cdots\cdots, P(\left|\frac{\mu_n}{n} - p\right| < \varepsilon), \cdots$

性质：（大数定律）

$$\lim_{n \to +\infty} P(\left|\frac{\mu_n}{n} - p\right| < \varepsilon) = 1,$$

或者与它等价的式子成立，即

$$\lim_{n\to+\infty} P\left(\left|\frac{\mu_n}{n}-p\right|\geq\varepsilon\right)=0.$$

这类极限定理——大数定律.

为了研究μ_n的极限性质，可以讨论它的分布函数$P\{\mu_n \le x\}$的变化情况. 因$E\mu_n = np, D\mu_n = npq$，所以当$n \to +\infty$时，对于固定的$x$而言，它将趋于0, 因而没什么意义. 同样，研究$\mu_n/n$的分布函数$P(\mu_n/n \le x)$的变化情况，对于固定的$x$而言，若$x < p$,

$$P\left\{\frac{\mu_n}{n} \le x\right\} = P\left\{\frac{\mu_n}{n} - p \le x - p\right\} \le P\left\{\left|\frac{\mu_n}{n} - p\right| \ge p - x\right\}$$

它将趋于0，若$x > p$,

$$P\left\{\frac{\mu_n}{n} \le x\right\} = P\left\{\frac{\mu_n}{n} - p \le x - p\right\} \ge P\left\{\left|\frac{\mu_n}{n} - p\right| \le x - p\right\}$$

它将趋于1，同样没什么意义. 为此我们通常研究"标准化"的随机变量$Y_n = \dfrac{\mu_n - np}{\sqrt{npq}}$的分布函数$P\{Y_n \le x\}$的极限性质. 由$Y_n$的分布函数不难得到$\mu_n$的分布函数.

已经证明上述分布函数的极限分布是$N(0,1)$，即

$$\lim_{n\to+\infty} P\{Y_n \le x\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x).$$

这是另一类极限定理 — —中心极限定理.

$$Y_n = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} EX_i}{\sqrt{\sum_{i=1}^{n} DX_i}}$$

我们的目的是寻找使

$$\lim_{n\to+\infty} P\{Y_n \le x\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x).$$

成立的条件. 上式成立称$\{X_n\}_1^{\infty}$服从中心极限定理.

这类极限定理的研究也顺便解决了n 较大场合时的二项分布的计算问题——后面研究该类问题.

$\mu_n \sim B(n,p)$:

随机变量序列：$\mu_1, \mu_2, \cdots\cdots, \mu_n, \cdots\cdots$

随机变量序列：$\dfrac{\mu_1 - p}{\sqrt{pq}}, \dfrac{\mu_2 - 2p}{\sqrt{2pq}}, \cdots\cdots, \dfrac{\mu_n - np}{\sqrt{npq}}, \cdots\cdots$

$\forall x > 0$

事件序列：$\dfrac{\mu_1 - p}{\sqrt{pq}} \le x, \dfrac{\mu_2 - 2p}{\sqrt{2pq}} \le x, \cdots\cdots, \dfrac{\mu_n - np}{\sqrt{npq}} \le x, \cdots\cdots$

数列：$P(\dfrac{\mu_1 - p}{\sqrt{pq}} \le x), P(\dfrac{\mu_2 - 2p}{\sqrt{2pq}} \le x), \cdots\cdots, P(\dfrac{\mu_n - np}{\sqrt{npq}} \le x), \cdots$

性质：（中心极限定理）

$$\lim_{n \to +\infty} P(\frac{\mu_n - np}{\sqrt{npq}} \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{dt} = \Phi(x)$$

关系：中心极限定理$\Rightarrow$大数定律

$$P\left(\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right)=P\left(\left|\frac{\mu_n-np}{\sqrt{npq}}\right|<\varepsilon\sqrt{\frac{n}{pq}}\right)$$

$$\approx\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)-\Phi\left(-\varepsilon\sqrt{\frac{n}{pq}}\right)$$

$$=2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)-1\to 1$$

二　切比雪夫(chebyshev)不等式

定理1（Chebyshev不等式） 设随机变量X的方差$D(X)$存在，则对于任意实数$\varepsilon > 0$，有

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2},$$

或
$$P(|X - E(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}.$$

证明 仅证连续型随机变量的情形，

设X的概率度为$f(x)$，则对于任意实数$\varepsilon > 0$, 有

$$P(|X - E(X)| \geq \varepsilon) = \int_{|x - E(X)| \geq \varepsilon} f(x)dx$$

$$\leq \int_{|x - E(X)| \geq \varepsilon} \frac{|x - E(X)|^2}{\varepsilon^2} f(x)dx$$

性质1 $DX \geq 0$，且

$$DX = 0 \Leftrightarrow P\{X = EX\} = 1.$$

特别地 常数的方差为零.

利用 $\{X \neq E(X)\} = \bigcup_{n=1}^{\infty} \left\{ |X - E(X)| \geq \frac{1}{n} \right\}$

及 $P(|X - E(X)| \geq \varepsilon) \leq \dfrac{D(X)}{\varepsilon^2}$

证明

$$\leq \frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}|x-E(X)|^2 f(x)dx=\frac{D(X)}{\varepsilon^2}.$$

从而

$$P(|X-E(X)|<\varepsilon)=1-P(|X-E(X)|\geq\varepsilon)$$

$$\geq 1-\frac{D(X)}{\varepsilon^2}.$$

切比雪夫不等式表明：

当误差ε取定后，DX越小，概率$P\{|X-EX|\geq\varepsilon\}$就越小. 即随机变量$X$落入$EX$的$\varepsilon$邻域之外的可能性很小, 也即落入$EX$的$\varepsilon$邻域内可能性很大. 由此说明$X$的取值比较集中，也即离散程度较小，这进一步说明了方差能描述随机变量取值的离散程度.

三　大数定律

定义 设X_1，$X_2,\cdots$，$X_n,\cdots$是随机变量序列，EX_i $(i=1,2,\cdots)$存在，若对于任意实数$\varepsilon>0$，有

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

或

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}EX_i\right|\geq\varepsilon\right\}=0.$$

则称随机变量序列X_1，$X_2,\cdots$，$X_n,\cdots$服从大数定律.

定理2（切比雪夫大数定律）

设随机变量序列$X_1, X_2, \cdots, X_n, \cdots$两两不相关，且它们具有相同的数学期望, 方差存在且具有公共的上界，即

$$E(X_k) = \mu, \quad D(X_k) \le C, \quad k = 1, 2, \cdots$$

则对于任意实数$\varepsilon > 0$，有

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

证明: (利用切比雪夫不等式）

$$E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \frac{1}{n} n\mu = \mu,$$

$$D\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}D(X_k)\le\frac{1}{n^2}nC=\frac{C}{n}.$$

根据切比雪夫不等式得

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}\ge 1-\frac{1}{\varepsilon^2}D\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)\ge 1-\frac{C}{n\varepsilon^2},$$

$$1-\frac{C}{n\varepsilon^2}\le P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}\le 1.$$

令n趋于无穷大，上式取极限即得证.

注：将定理的条件改为：若随机变量序列$\{X_k\}_1^{+\infty}$满足

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}X_k\right)\to 0,$$

结论仍然成立——马尔科夫大数定律.

定理3 伯努利（Bernoulli）大数定律　设μ_n是n次独立重复试验中事件A发生的次数, p是每次试验中A发生的概率,则对于任意实数$\varepsilon > 0$, 有

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

或

$$\lim_{n\to\infty} P\left(\left|\frac{\mu_n}{n} - p\right| < \varepsilon\right) = 1.$$

注：(1)频率的稳定性是概率定义的客观基础, 伯努利大数定理以严密的数学形式论证了频率的稳定性的确切含义.

(2)定理说明：当试验次数很大时，可以用事件A发生的频率作为概率p的估计——参数估计法.

（1）伯努利大数定律是本定理的特例——练习题.

（2）定理中的方差相同，改为“方差有公共的上界”，结论仍然成立——切比雪夫大数定律.

四　中心极限定理

研究独立随机变量之和所特有的规律性问题.

当n无限增大时，随机变量之和的分布是什么？

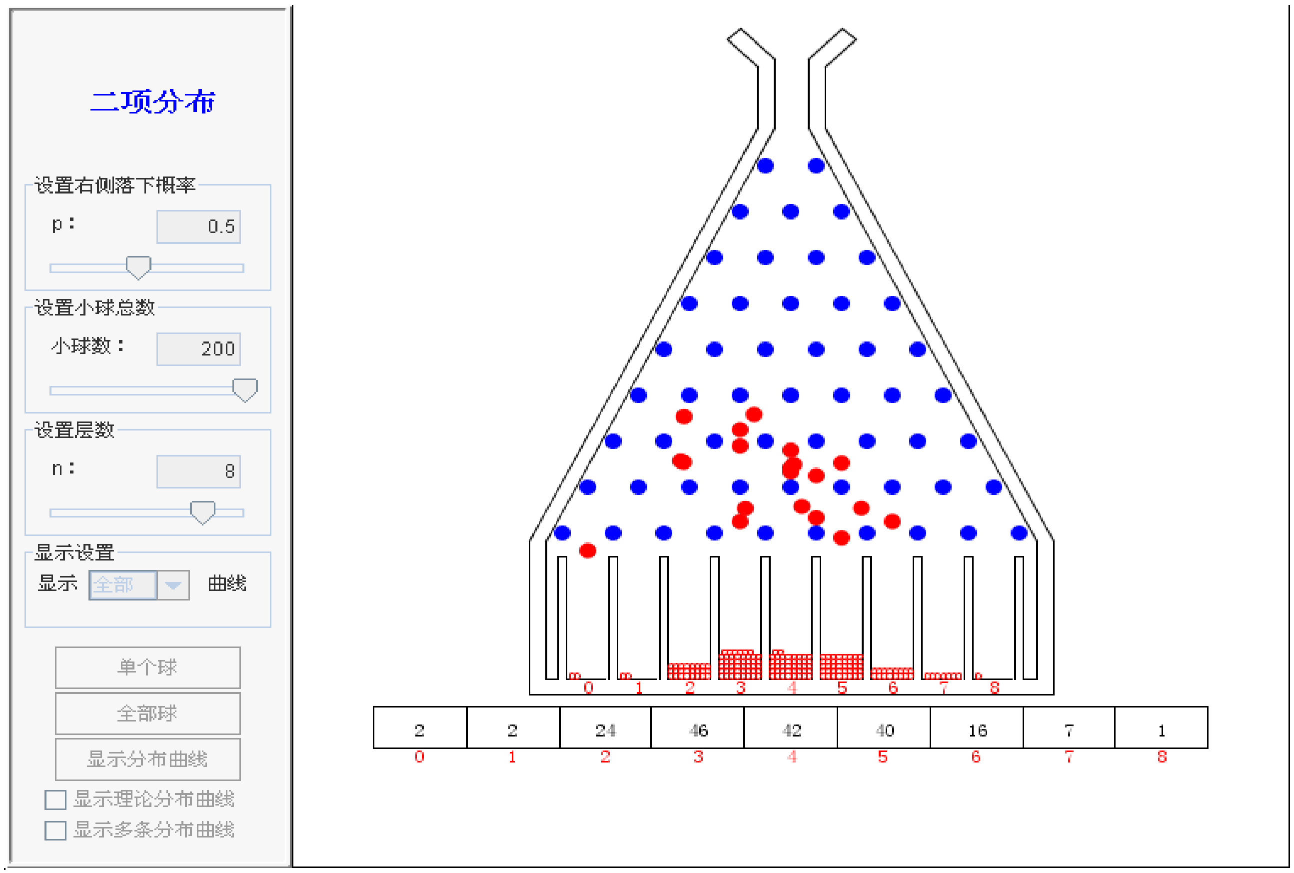
先看一个随机试验.

四　中心极限定理

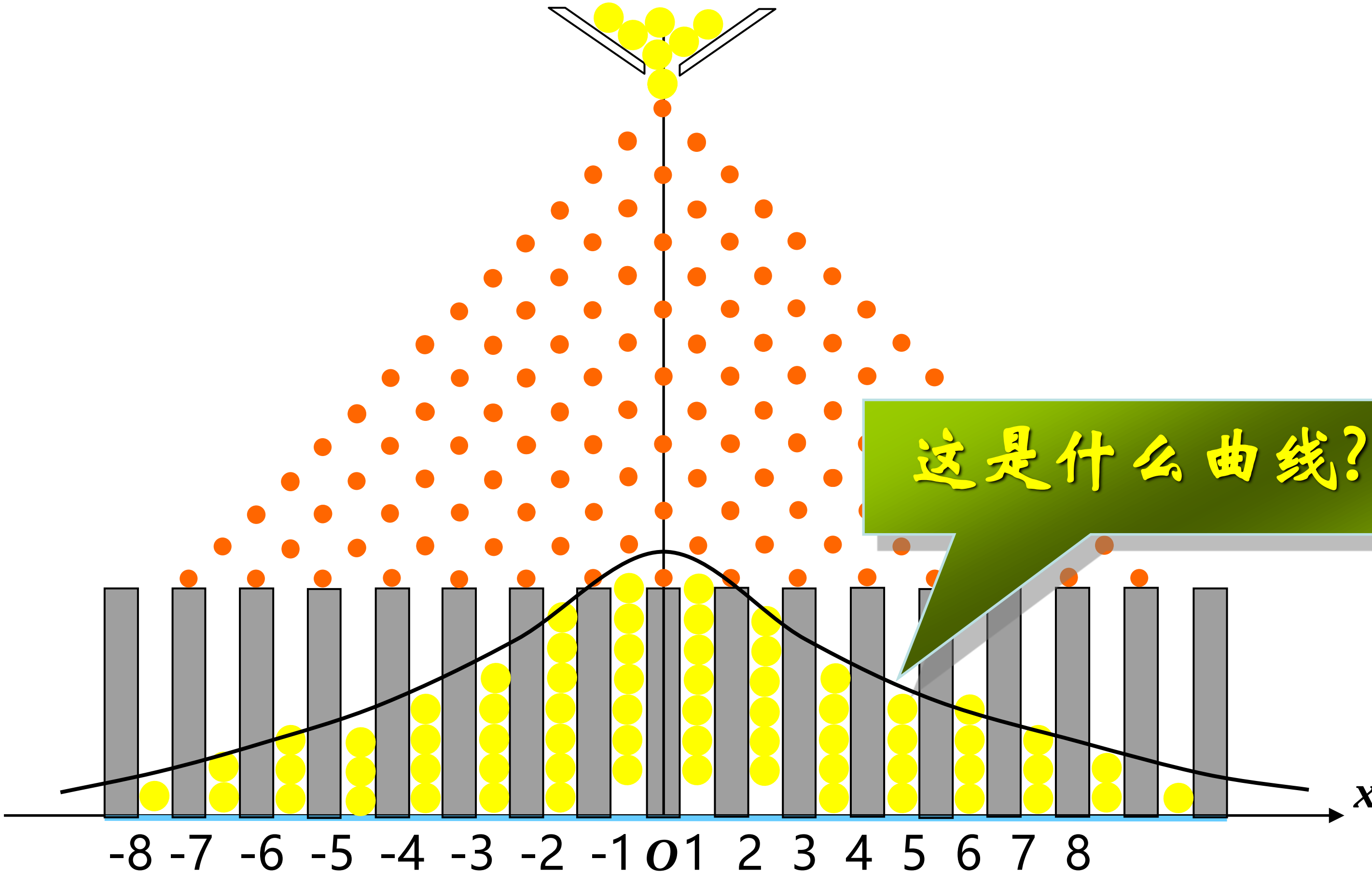
研究独立随机变量之和所特有的规律性问题.
当n无限增大时，随机变量之和的分布是什么？

先看一个随机试验.

高尔顿（Galton）钉板试验：



高尔顿（Galton）钉板试验：



我们作初步解释如下

令X_k表示某一个小球在第k次碰了钉子后向左或向右落下这一随机现象相联系的随机变量（$X_k=1$表示向右落下，$X_k=-1$表示向左落下），由题意，X_k的分布列可设为下列形式：

$$P(X_k=1)=0.5，P(X_k=-1)=0.5.\quad k=1,2,3,\cdots$$

令

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

其中$X_1,X_2,\cdots,X_n$相互独立，而Y_n则表示这个小球第n次碰钉后的位置. 试验表明Y_n近似地服从正态分布.

上述例子表明，需要研究相互独立随机变量和的极限分布是正态分布的问题.

概率论中关于论证“大量独立随机变量的和的极限分布是正态分布”的一系列定理统称为中心极限定理.

本节我们叙述其中的两个最简单，也是最重要的定理, 并举例说明它们在实际问题中的应用.

定理4 (林德伯格——莱维)中心极限定理

设随机变量序列$X_1, X_2, \cdots, X_n, \cdots$相互独立，且服从同一分布，

$$E(X_k) = \mu,\ D(X_k) = \sigma^2 > 0,\ k = 1, 2, \cdots$$

则随机变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sigma\sqrt{n}}$$

的分布函数$F_n(x)$满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P(Y_n \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t = \Phi(x).$$

这个定理通常称为“独立同分布中心极限定理”. 它的证明超出了本书的范围, 我们略去, 这里仅对定理的含义做一些说明. 注意到等式右端就是标准正态分布的分布函数$\Phi(x)$. 定理表明:

$$
n\text{充分大时}\quad\begin{cases} Y_n = \dfrac{1}{\sigma\sqrt{n}}\left[\displaystyle\sum_{k=1}^{n} X_k - n\mu\right] \overset{\text{近似}}{\sim} N(0,1), \\ \displaystyle\sum_{k=1}^{n} X_k \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2). \end{cases}
$$

定理5 （德莫佛—拉普拉斯）中心极限定理

设$Y_n \sim B(n,p)$, $0<p<1$, $n=1,2,\cdots$. 则对任一实数x，有

$$\lim_{n\to\infty} P\left(\frac{Y_n - np}{\sqrt{np(1-p)}} \le x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \mathrm{d}t.$$

即 $\dfrac{Y_n - np}{\sqrt{np(1-p)}} \overset{\text{近似}}{\sim} N(0,1)$. $\quad Y_n \overset{\text{近似}}{\sim} N(np, npq)$.

注: (1)定理5表明:二项分布以正态分布为极限分布.

(2)定理5是定理4的特例.

例1 售报员在报摊上卖报，已知每个过路人在报摊上买报的概率为1 / 3. 令Y是出售了100份报时过路人的数目，求$P(280 \le Y \le 320)$.

解：令X_i为售出了第 $i-1$ 份报纸后到售出第 i 份报纸时的过路人数，$i = 1, 2, \cdots, 100$. 则

$$P(X_i = k) = p(1-p)^{k-1}\Big|_{p=1/3}, \quad k = 1, 2, \cdots$$

$$E(X_i) = (1/p)\Big|_{p=1/3} = 3, \ D(X_i) = (q/p^2)\Big|_{p=1/3} = 6,$$

$$Y = \sum_{k=1}^{100} X_k.$$

由于$X_1, X_2, \cdots, X_{100}$相互独立，根据林德贝格－列维中心极限定理.

$$Y=\sum_{k=1}^{100}X_k \overset{\text{近似}}{\sim} N(300,600).$$

$$P(280\le Y\le 320)\approx \Phi(\frac{320-300}{\sqrt{600}})-\Phi(\frac{280-300}{\sqrt{600}})$$

$$=2\Phi(\frac{20}{\sqrt{600}})-1\approx 2\Phi(0.8165)-1\approx 0.5878.$$

例2　某车间有200台车床,每台独立工作,设开工率为0.6. 开工时每台耗电量为1千瓦. 问供电所至少要供给这个车间多少电力, 才能以99.9% 的概率保证这个车间不会因供电不足而影响生产？

解：设至少要供给这个车间a千瓦的电力，X为开工的车床数，则$X \sim B(200,\ 0.6)$.

由德莫佛 — 拉普拉斯中心极限定理, 有

$X \sim N(120,\ 48)$(近似)

要使 $P(0 \le X \le a) = 99.9\%$.

$$P(0 \le X \le a) \approx \Phi(\frac{a-120}{\sqrt{48}}) - \Phi(\frac{0-120}{\sqrt{48}})$$

$$\approx \Phi(\frac{a-120}{\sqrt{48}}).$$

反查标准正态函数分布表，得 $\Phi(3.09) = 99.9\%$.

$$\frac{a-120}{\sqrt{48}} = 3.09 \quad \Longrightarrow \quad a \approx 141(\text{千瓦}).$$

例3 蒲丰试验中投掷硬币4040次. 出现正面2048次，试计算当重复蒲丰试验时，正面出现的频率与概率之差的偏离程度不大于蒲丰试验中所发生的偏差的概率.

解：蒲丰实验的频率与概率的偏差为

$$\varepsilon = \left| \frac{2048}{4040} - \frac{1}{2} \right| = 0.00693$$

所求概率为

$$P\left\{ \left| \frac{\mu_{4040}}{4040} - \frac{1}{2} \right| < 0.00693 \right\} = 2\Phi(0.00693\sqrt{\frac{4040}{0.5^2}}) - 1$$

$$= 2\Phi(0.8810) - 1 = 2 \times 0.8109 - 1 = 0.622.$$