浙江大学
ZHEJIANG UNIVERSITY

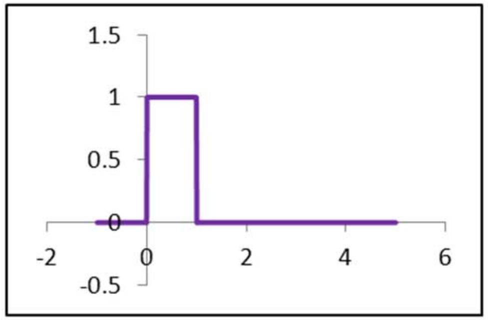
第37讲 中心极限定理

问题的提出：

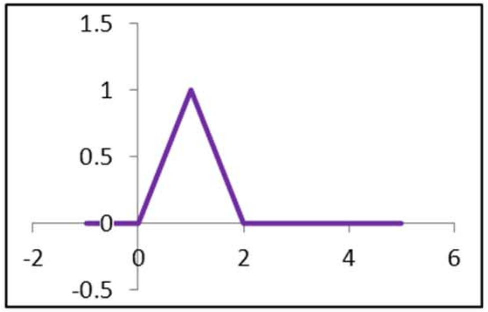
有许多随机变量，它们是由大量的相互独立的随机变量的综合影响所形成的，而其中每个个别的因素作用都很小，这种随机变量往往服从或近似服从正态分布，或者说它的极限分布是正态分布，中心极限定理正是从数学上论证了这一现象，它在长达两个世纪的时期内曾是概率论研究的中心课题.

n个独立同分布的均匀分布随机变量的和：

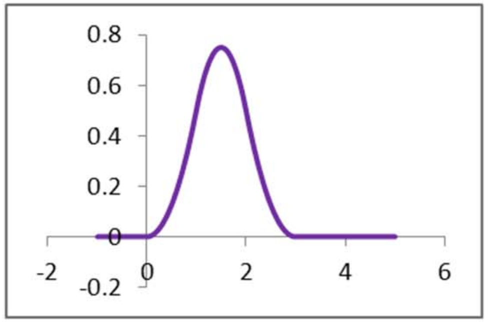
$$X_1$$



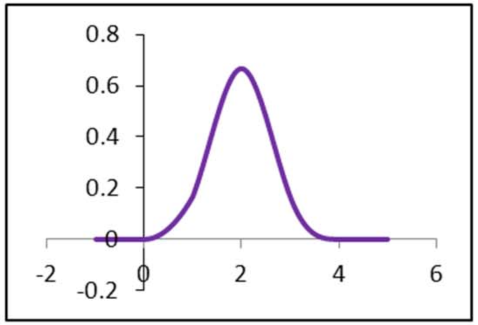
$$X_1+X_2$$



$$X_1+X_2+X_3$$



$$X_1+X_2+X_3+X_4$$



定理 (独立同分布的中心极限定理(CLT)):

设随机变量$X_1, X_2, \cdots, X_n, \cdots$,相互独立且同分布,

$$E(X_i) = \mu, D(X_i) = \sigma^2, i = 1, 2, \cdots,$$

则对于充分大的n,有

$$\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2).$$

此时

$$P(a < \sum_{i=1}^{n} X_i \le b) \approx \Phi\left(\frac{b - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{a - n\mu}{\sqrt{n}\sigma}\right).$$

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 的近似分布是

$$\boxed{N(\mu, \frac{\sigma^2}{n})}$$

注意到:当随机变量$X_1, X_2, \cdots, X_n, \cdots$,相互独立且同分布,

$$E(X_i) = \mu,\ D(X_i) = \sigma^2,\ i = 1, 2, \cdots,$$

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = n\mu.$$

$$D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} D(X_i) = n\sigma^2.$$

据CLT,有 $\sum_{i=1}^{n} X_i \overset{近似}{\sim} N(n\mu, n\sigma^2).$

故CLT仅仅是分布类型上的一种近似.("万物归一")

定理 (德莫弗-拉普拉斯中心极限定理):

记n_A为n重贝努里试验中事件A发生的次数,

并记事件A在每次试验中发生的概率为p $(0<p<1)$.

则对于充分大的n有 $n_A \overset{\text{近似}}{\sim} N(np, np(1-p))$.

即, 对于二项分布$B(n, p)$,当n充分大时,

可用正态分布来近似.

中心极限定理的作用通过实验10有一个直观的了解。

例1： 某宴会上提供一瓶6升(l)的大瓶法国红酒,假定与会者每次所倒红酒的重量服从同一分布，期望值为100毫升(ml)，标准差为32毫升.若每次所倒红酒都是相互独立的，试问：倒了55次后该瓶红酒仍有剩余的概率约为多少？

解： 设X_i为第i次所倒的红酒的重量(ml), 则相互独立,

分布相同,且$E(X_i)=100, D(X_i)=32^2, i=1,2,\cdots,55.$

$$P\{\text{倒了55次后该瓶红酒仍有剩余}\}=P\{\sum_{i=1}^{55}X_i<6000\}$$

根据独立同分布的CLT,可知

$$\sum_{i=1}^{55}X_i \overset{\text{近似}}{\sim} N(55\times100, 55\times32^2).$$

$$P\{\sum_{i=1}^{55}X_i<6000\}\approx\Phi\left(\frac{6000-55\times100}{32\sqrt{55}}\right)=\Phi(2.11)=0.9826.$$

例2: 一颗骰子投掷了11520次，出现□1点□的次数超过了2160次，由此可否推断该骰子是不均匀的呢?

解: 记n_A为骰子投掷了11520次出现"**1**点"的次数.

假设骰子是均匀的，则

$$n_A \sim B(11520, 1/6).$$

$$\text{那么} P\{n_A > 2160\} = \sum_{k=2161}^{11520} C_{11520}^{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{11520-k} = \dots$$

根据德莫弗$-$拉普拉斯CLT,可知

$$n_A \overset{\text{近似}}{\sim} N(np, np(1-p)).$$

即 $n_A \overset{\text{近似}}{\sim} N(1920, 1600).$

那么 $P\{n_A > 2160\} \approx 1-\Phi\left(\dfrac{2160-1920}{\sqrt{1600}}\right) = 1-\Phi(6) = 0.$

根据实际推断原理, 可以推断该骰子是不均匀的.