第二节　中心极限定理

一、林德伯格-莱维中心极限定理
(独立同分布中心极限定理)

二、棣莫弗-拉普拉斯中心极限定理
(二项分布的正态近似)

三、李雅谱诺夫中心极限定理
(独立不同分布下的中心极限定理)

Chapter 5 Law of large numbers and Limit theorem

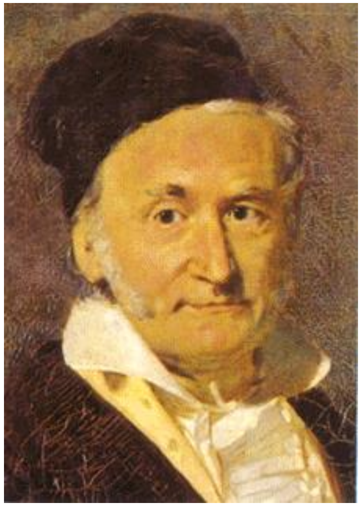
Law of Large Numbers (LLN)	Chebyshev's LLN
	Bernoullis's LLN
	Khinchin's LLN

Central Limit Theorem (CLT)	Lindberger-Levy CLT
	De Moirve-Laplace CLT
	Liapunov CLT

Introduction

- In probability theory, it has been known that normal distribution occupies the first important position, and many random variables obey normal distribution.

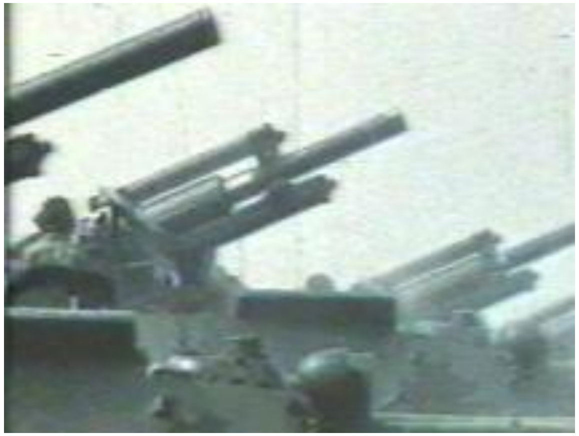
- Since Gauss pointed out that the measurement error obeys the normal distribution, it has been found that the normal distribution is very common in nature.



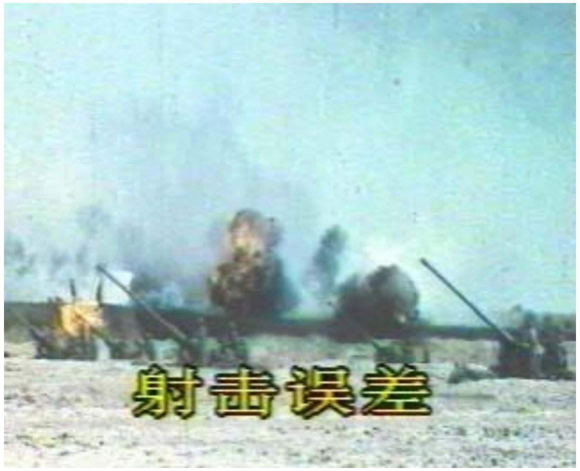
Gauss

中心极限定理的背景

在实际问题中，常常需要考虑许多随机因素所产生总影响.



如：炮弹射击的落点与目标的偏差，就受着许多随机因素的影响，如瞄准误差、空气阻力所产生的误差、炮身结构所引起的误差等。每个随机因素对弹着点（随机变量和）所起作用都是很小的，则弹着点服从怎样分布？



在概率论中习惯于把随机变量和的分布收敛于正态分布这一类定理叫作中心极限定理.

Introduction

A large number of experimental observations indicate that:

If a **variable** is caused by the joint influence of a large number of independent and random factors, **and each individual factor plays little role** in the total influence, then this variable generally **obeys** or **approximately obeys** the normal distribution.

Is it based on experience, or **is it based on theory?**

Central Limit Theorems (CLT)

- The sum of a series of independent identically distributed random variables with finite variance, that it takes the standard normal distribution as limit after standardization.

  Lindberger-Levy CLT

  When the "same distribution" is a binomial distribution, a special case of the theorem is obtained.

  De Moirve-Laplace CLT

- For the variables caused and accumulated by a large number of small independent random factors (do not seek the same distribution), when the number of random factors tends to infinity, the normal distribution is the limit.

  Liapunov CLT

中心极限定理

大数定律讨论的是在什么条件下，随机变量的算数平均依概率收敛到其均值的问题；中心极限定理所讨论的内容是在什么条件下，相互独立的随机变量的和的分布函数会收敛于正态分布。

设$\{X_n\}$为独立随机变量序列，且$E(X_n)$和$D(X_n)$存在，$n = 1,2,\dots$.设

$$Y_n = \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} E(X_k)}{\sqrt{\sum_{k=1}^{n} D(X_k)}}$$

若对任意的$x \in (-\infty, +\infty)$,

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

即Y_n的分布函数的极限是标准正态分布，则称$\{X_n\}$服从中心极限定理.

5.2.1 林德伯格-莱维中心极限定理（独立同分布中心极限定理）

定理5.2.1　设$\{X_n\}$为独立同分布的随机变量序列，且$E(X_n)=\mu$和$D(X_n)=\sigma^2>0$，$n=1,2,\ldots$.则对任意x,随机变量序列

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

的分布函数$F_n(x)$当$n\to+\infty$时趋近标准正态分布函数，即

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P\left(\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}\le x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\phi(x)$$

说明：定理5.2.1只假设$X_1,X_2,\ldots$独立同分布且方差存在，而不必关注他的分布是什么，只要n充分大，就可以用正态分布去逼近随机变量和的分布，所以有广泛的应用。

5.2.2 棣莫弗-拉普拉斯中心极限定理（二项分布的正态近似）

定理5.2.2 设n_A为n重伯努利试验中事件A发生的次数，$p(0<p<1)$是事件A在每次试验中发生的概率，即$n_A \sim B(n,p)$,则对于任意的实数x,

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

证明 设$X_i=\begin{cases}1, & 在第i次试验中事件A发生,\\ 0, & 在第i次试验中事件A不发生\end{cases}$ $i=1,2,\dots,n.$

则有$n_A=X_1+X_2+\cdots+X_n$,

显然随机变量$X_1,X_2,\cdots,X_n,\cdots$相互独立，且都服从参数为$p$的0-1分布，

$$E(X_i)=p \qquad D(X_i)=p(1-p) \qquad i=1,2,\cdots,n$$

则由林德伯格-莱维中心极限定理

$$\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}\le x\right)=\lim_{n\to\infty} P\left(\frac{n_A-np}{\sqrt{np(1-p)}}\le x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\phi(x)$$

5.2.2 棣莫弗-拉普拉斯中心极限定理（二项分布的正态近似）

注意

1) 棣莫佛－拉普拉斯定理是中心极限定理的特例，但它是概率论发展史上第一个中心极限定理，专门针对二项分布，因此称为“二项分布的正态近似”.

2） 德莫佛—拉普拉斯定理表明:二项分布以正态分布为极限，即当 n 充分大时

$$n_A \quad \text{近似服从正态分布} \quad N(np, np(1-p))$$

5.2.3 李雅谱诺夫中心极限定理（独立不同分布的中心极限定理）

定理5.2.3 设$\{X_n\}$为独立随机变量序列，具有有限的数学期望和方差：

$$E(X_k) = \mu_k,\ D(X_k) = \sigma_k^2 > 0，\mathrm{k} = 1,2,\ldots.$$

记 $B_n^2 = \sum\limits_{k=1}^{n} \sigma_k^2$ 并设

$$Y_n = \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} E(X_k)}{\sqrt{\sum_{k=1}^{n} D(X_k)}} = \frac{\sum_{k=1}^{n}(X_k - \mu_k)}{B_n}$$

若存在正常数δ，使得 $\dfrac{1}{B_n^{2+\delta}} \sum\limits_{k=1}^{n} E\left|X_k - \mu_k\right|^{2+\delta} \to 0, n \to \infty$ 李雅谱诺夫条件

则对任意的$x\in(-\infty,+\infty)$,

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

5.2.3 李雅谱诺夫中心极限定理（独立不同分布的中心极限定理）

定理5.2.3 表明：在李雅普诺夫条件下，当n充分大时，

$$Y_n = \frac{\sum_{k=1}^{n}(X_k-\mu_k)}{B_n} \dot{\sim} N(0,1)$$

即 Y_n 近似服从标准正态分布 $N(0,1)$

所以当n充分大时，

$$\sum_{k=1}^{n} X_k = B_n Y_n + \sum_{k=1}^{n} \mu_k$$ 近似服从正态分布 $N(\sum_{k=1}^{n}\mu_k, B_n^2)$

无论各随机变量$X_1, X_2, \ldots$服从什么分布，只要满足定理的条件，当n充分大时，它们的和$\sum_{k=1}^{n} X_k$就近似地服从正态分布。这就是正态随机变量在概率论中占有重要地位的一个基本原因。

小结

1.在实际问题中，若某随机变量可以看作是有相互独立的大量随机变量综合作用的结果，每一个因素在总的影响中的作用都很微小，则**综合作用的结果服从正态分布**.

Examples

Question: If more than 10 defective products are found during sampling inspection of product quality, this batch of products will be considered unacceptable.

Find: how many products should be inspected to make the probability of a batch of products with a defect rate of 10% unacceptable reach 0.9?

Solution: let n be the number of sampling products, X be the number of defective products. Then :

$$X \sim B(n, 0.1), \quad np = 0.1n, \quad \sqrt{npq} = 0.3\sqrt{n}$$

$$P(10 \le X \le n) = 0.9$$

$$\because \quad P(10 \le X \le n)$$

$$= P\left(\frac{10 - 0.1n}{0.3\sqrt{n}} \le \frac{X - 0.1n}{0.3\sqrt{n}} \le \frac{n - 0.1n}{0.3\sqrt{n}}\right)$$

Examples

$$= P\left(\frac{10-0.1n}{0.3\sqrt{n}} \le \frac{X-0.1n}{0.3\sqrt{n}} \le 3\sqrt{n}\right)$$

"3σ rule"

$\Phi(3\sqrt{n}) = 1$

$$\approx \Phi(3\sqrt{n}) - \Phi\left(\frac{10-0.1n}{0.3\sqrt{n}}\right) = 1 - \Phi\left(\frac{10-0.1n}{0.3\sqrt{n}}\right)$$

$$\therefore \quad P(10 \le X \le n) = 1 - \Phi\left(\frac{10-0.1n}{0.3\sqrt{n}}\right) = 0.9$$

$$\frac{0.1n-10}{0.3\sqrt{n}} = 1.28 \quad \Rightarrow \quad n \approx 146$$

Examples

Question: When calculating addition, the computer takes each addend as the integer closest to it. Let all retrieval errors be independent random variables and obey uniform distribution in the interval [- 0.5,0.5].

Find: (1) The probability that the absolute value of the total error is less than 10 when the existing 1200 numbers are added;

(2) How many numbers should be added to make the absolute value of the error sum less than 10 and the probability greater than 0.9

Examples

Solution: Let X_1, X_2,... X_n be the retrieval error of each addend, then this is a column of independent and identically distributed random variables.

The error sum of all addends is : $\sum_{k=1}^{n} X_k$

$\because$ X_k $(k = 1, 2, \cdots n)$ obeys uniform distribution in $[-0.5, 0.5]$,

$$\therefore \quad E(X_k) = \frac{-0.5 + 0.5}{2} = 0, \quad D(X_k) = \frac{[0.5 - (-0.5)]^2}{12} = \frac{1}{12}$$

So, $\sum_{k=1}^{n} E(X_k) = 0, \quad \sqrt{\sum_{k=1}^{n} D(X_k)} = \sqrt{\frac{n}{12}}$

Examples

Find: (1) The probability that the absolute value of the total error is less than 10 when the existing 1200 numbers are added;

(1) $\because n = 1200, \quad \therefore \sqrt{\sum_{k=1}^{n} D(X_k)} = \sqrt{\frac{1200}{12}} = 10$

Thus, $P\left(\left| \sum_{k=1}^{1200} X_k \right| < 10 \right) = P\left(-10 < \sum_{k=1}^{1200} X_k < 10 \right)$

$$= P\left(\frac{-10-0}{10} < \frac{\sum_{k=1}^{n} X_k - 0}{10} < \frac{10-0}{10} \right)$$

$$= P\left(-1 < \frac{\sum_{k=1}^{n} X_k}{10} < 1 \right) \approx \Phi(1) - \Phi(-1)$$

$$= 2\Phi(1) - 1$$

$$= 2 \times 0.8453 - 1 = 0.6826$$

Examples

(2) How many numbers should be added to make the probability of the absolute value of sum error less than 10 is greater than 0.9?

$$P\left(\left|\sum_{k=1}^{n} X_k\right| < 10\right) > 0.9$$

Examples

$$\because \; P\left(\left|\sum_{k=1}^{n} X_k\right| < 10\right) = P\left(\frac{-10-0}{\sqrt{n/12}} < \frac{\sum_{k=1}^{n} X_k - 0}{\sqrt{n/12}} < \frac{10-0}{\sqrt{n/12}}\right)$$

$$= P\left(-20\sqrt{\frac{3}{n}} < \frac{\sum_{k=1}^{n} X_k}{\sqrt{\frac{n}{2}}} < 20\sqrt{\frac{3}{n}}\right) = 2\Phi\left(20\sqrt{\frac{3}{n}}\right) - 1$$

$$\Rightarrow \quad P\left(\left|\sum_{k=1}^{n} X_k\right| < 10\right) > = 2\Phi\left(20\sqrt{\frac{3}{n}}\right) - 1 > 0.9$$

$$\therefore \quad 20\sqrt{\frac{3}{n}} = 1.65 \qquad \Rightarrow \qquad n \approx 441$$