第5章 大数定律与中心极限定理

大数定律与中心极限定理通称极限理论;
概率论中比较深刻的理论结果;
数理统计学的理论基础。

第四章 大数定律与中心极限定理

概述

大数定律和中心极限定理就是使用极限方法研究大量随机现象统计规律性.

阐明大量重复试验的平均结果具有稳定性的一系列定律都称为大数定律.

论证随机变量（试验结果）之和渐进服从某一分布的定理称为中心极限定理.

大数定律与中心极限定理

要解决的问题

答复

1. 为何能以频率估计概率？

2. 为何能以样本均值估计总体期望？

大数定律

大数定律：　　描述一系列<u>随机变量的和的平均结果</u>的稳定性

3. 为何正态分布有极其重要的地位？

4. 大样本统计推断的理论基础？

中心极限定理

中心极限定理：描述一系列<u>随机变量的和的概率分布</u>的稳定到正态分布

一.依概率收敛

设$\{X_n\}$为随机变量序列，X为随机变量，若任给$\varepsilon>0$, 使得

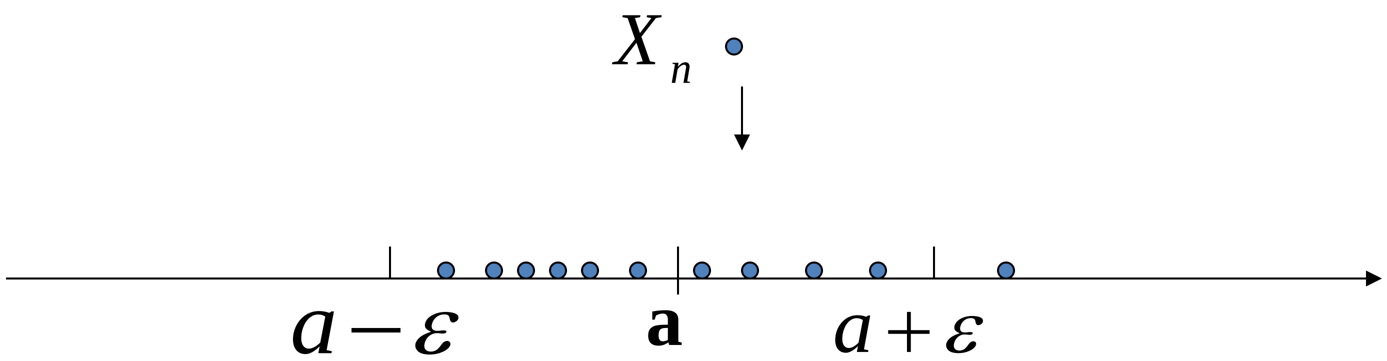
$$\lim_{n\to\infty} P\{|X_n - X| < \varepsilon\} = 1$$

则称$\{X_n\}$**依概率收敛**于X. 可记为

$$X_n \xrightarrow{P} X.$$

如 $X_n \xrightarrow{P} a$ 意思是:当 $n\to\infty$ 时,X_n落在

$(a-\varepsilon, a+\varepsilon)$ 内的概率越来越大. $\forall n_0, n>n_0$



而 $X_n \to a$ 意思是: $\forall\varepsilon>0, \exists n_0$,当 $n>n_0$

$$|X_n - a| < \varepsilon$$

一、依概率收敛

- 定义：设 $X_1, X_2, \cdots, X_n, \cdots$ 是一列随机变量，如果对于任意的 $\varepsilon > 0$，恒有

$$\lim_{n\to\infty} P\{|X_n - X| > \varepsilon\} = 0$$

或

$$\lim_{n\to\infty} P\{|X_n - X| \le \varepsilon\} = 1$$

则称$\{X_n\}$依概率收敛到X，记做$X_n \xrightarrow{P} X$

或 $P - \lim\limits_{n\to\infty} X_n = X$

二、大数定律

大数定律— 概率论中有关阐明大量随机现象平均结果的稳定性的一系列定理。

迄今为止,人们已发现很多大数定律(laws of large numbers)所谓大数定律，简单地说，就是大量数目的随机变量所呈现出的规律，这种规律一般用随机变量序列的某种收敛性来刻画。

二、大数定律

- 不是经验规律，而是严格证明了的定理;
- 成立条件：试验次数足够多；
- 频率的稳定性：“频率稳定到概率”；
  - “概率”是无法测量的；
  - 只有“频率”是可以被测量的；
  - 大数定律告诉我们，用“频率”去推测“概率”是合理的

二、大数定律

伯努利大数定律

历史上，J. Bernoulli第一个研究了大数定律，在其1713年出版的《猜测术》中，他建立了本定理，

这是全部大数定律中的第一个。

Jacob Bernoulli
(1654-1705)

二、大数定律

算术平均

伯努利大数定律

设 $\mu_n \sim B(n,p)$, 则对任意 $\varepsilon > 0$,有

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

上述极限记作 $\frac{\mu_n}{n} \xrightarrow{P} p$ 或 $P\text{-}\lim\limits_{n\to 0}\left(\frac{\mu_n}{n}\right) = p$,读作 $\frac{\mu_n}{n}$ 依概率收敛到 p.

数学期望

频率的稳定性

在独立重复试验中,事件 A 发生的频率 $\frac{\mu_n}{n}$ 会随着试验次数 n 的增加逐渐稳定到 A 发生的概率 $p = P(A)$.

二、大数定律

伯努利大数定律

设 $\mu_n \sim B(n,p)$, 则对任意$\varepsilon > 0$,有

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

在一定条件下，当试验次数增加时，事件发生的频率接近于事件发生的概率

该定律为用频率估计概率提供了理论依据

问题：将伯努利大数定律的条件放宽，是否可以得到更一般的定律？

切比雪夫大数定律

期望和方差存在

不相关

方差有界

设随机变量列 $X_1, X_2, \cdots, X_n, L$ 两两不相关,数学期望EX_i 和方差 DX_i 均存在, 且 DX_i 有界,即存在 C使得 $DX_i \le C,\ i = 1, 2, L\ , n.$ 则对于任意$\varepsilon > 0$, 有

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

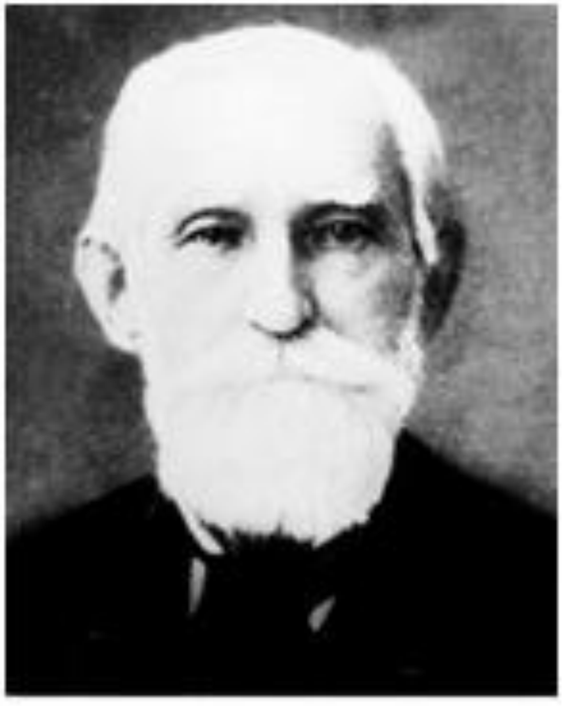
算术平均

数学期望

这个结果1866年由俄国数学家切比雪夫证得。

它是关于大数定律的一个相当普遍的结论，许多大数定律的古典结果是它的特例。

其证明方法（先构造一个不等式）也很有创造性，在此基础上发展起来的一系列不等式是研究各种极限定理的有力工具。



Pafnuty Lvovich Chebychef
(1821-1894)

切比雪夫大数定律的证明

----------切比雪夫不等式的回顾

如果随机变量X的数学期望 EX和方差 DX存在，则对于任一正数ε，都有 $P\left(|X-EX|\geq\varepsilon\right)\leq\dfrac{DX}{\varepsilon^2}$.

$$P\left\{|X-EX|\leq\varepsilon\right\}\geq 1-\frac{DX}{\varepsilon^2}$$

切比雪夫不等式将期望和方差联系起来

切比雪夫大数定律的证明

记$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}EX_i\right|>\varepsilon\right)=P\left(\left|\bar{X}-E\bar{X}\right|>\varepsilon\right)$$

$$\le\frac{1}{\varepsilon^2}D\left(\bar{X}\right)=\frac{1}{n^2\varepsilon^2}D\left(\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n}DX_i\le\frac{C}{n\varepsilon^2}$$

两边取极限,即得

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

●切比雪夫大数定律表明，当n很大时，

X_1，X_2，…，X_n的算术平均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

的取值，集中在其数学期望 $E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$ 附近。

这使我们关于算术平均值的法则有了理论上的依据。

如我们要测量某段距离，在相同条件下重复进行n次，得n个测量值 $X_1, X_2, \cdots, X_n$，它们可以看成是n个相互独立的随机变量具有相同的分布、相同的数学期望 μ 和方差 σ^2，

如我们要测量某段距离，在相同条件下重复进行*n*次，得*n*个测量值 $X_1, X_2, \cdots, X_n$ 它们可以看成是*n*个相互独立的随机变量具有相同的分布、相同的数学期望 μ 和方差 σ^2，

由大数定律知，只要*n*充分大，则以接近于1的概率保证

$$\mu \approx \frac{1}{n}\sum_{i=1}^{n} X_i$$

这便是在*n*较大情况下反映出的客观规律,故称为“大数”定律

切比雪夫大数定律推论

有相同的概率分布，并且互相独立，那么这些随机变量是独立同分布

设随 $\cdots$ $.X_n$是一列独立同分布的随机变量，$\cdots$ 差均存在，记 $E\xi_i=\mu$，则对任意 $\cdots$

算术平均

独立同分布 $\cdots$
前提是它们的 $\cdots$

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

辛钦大数定律

设随机变量列 $X_1,X_2,\cdots,X_n,\cdots$ 独立同分布且数学期望 μ存在,则对于任意$\varepsilon>0$, 有

$$\lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|>\varepsilon\right)=0.$$

数

切比雪夫大数定律推论

设随机变量X_1, X_2，…X_n是一列独立同分布的随机变量，其数学期望和方差均存在，记 $E X_i = \mu$，则对任意的$\varepsilon > 0$，有

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

在一定条件下，当样本的个数增加时，样本的算术平均接近于总体的平均值

该定律为统计推断中依据样本平均数估计总体平均数提供了理论依据

辛钦证明了在独立同分布的场合大数定律成立并不要求方差存在.

辛钦大数定律

设随机变量列 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布且数学期望 μ 存在,则对于任意 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right) = 0.$$

Xinchin(1894-1959)

三大定律的区别----附加条件不同

- **切比雪夫大数定律**：
- （1）不相关（2）期望和方差均存在（3）方差有上界

- **切比雪夫大数定律推论**：
- （1）独立同分布
- （2）期望和方差均存在

- **辛钦大数定律**：
- （1）独立同分布
- （2）期望存在

- **伯努利大数定律**：
- （1）独立同0-1分布
- （2）期望方差均相同

例1 判断下列说法的对错, 并简述理由:

(1) 设随机变量列 $X_1, X_2, \cdots, X_n, \mathrm{L}$ 独立且具有相同的密度函数,则序列 $X_1, X_2, \cdots, X_n, \mathrm{L}$ 满足辛钦大数定律.

(2) 设随机变量列 $X_1, X_2, \cdots, X_n, \mathrm{L}$ 独立同服从参数为λ的泊松分布,则 $X_1, 2X_2, \cdots, nX_n \mathrm{L}$ 满足切比雪夫大数定律.

解 (1) 错. 虽然 $X_1, X_2, \cdots, X_n, \mathrm{L}$ 独立且有相同的密度,但不能保证它们的数学期望存在. (例如, 柯西分布.)

(2) 错. 虽然 $X_1, 2X_2, \cdots, nX_n, \mathrm{L}$ 两两不相关且存在期望和方差, 但是它们的方差 $D(nX_n) = n\lambda$不存在上界.

大数定律的意义

当 n 足够大时, 算术平均值几乎是一常数.

数学期望 可被 算术均值 近似代替

三、　中心极限定理

一、客观背景

在实际问题中许多随机变量是由**相互独立随机因素的综合（或和）影响**所形成。

例如：炮弹射击的落点与目标的偏差，就受着许多随机因素（如瞄准，空气阻力，炮弹或炮身结构等）综合影响的。每个随机因素的对弹着点（随机变量和）所起的作用都是很小的。

那么弹着点服从怎样分布呢？

自从高斯发现测量误差服从正态分布之后，人们通过大量的观察和研究发现，**正态分布在自然界中极为常见**。

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。**

二、中心极限定理

引例：
在自然界与生产中，一些现象受到许多相互独立的随机因素的影响，如果每个因素所产生的影响都很微小时，总的影响是每一个微小影响的加和。例如：随机误差

依大数定律，**总的影响的数学期望**可以由每一个独立因素的算数平均估计得到；

总的影响的分布情况?

中心极限定理论证了这个总的影响是服从正态分布的

论证随机变量（试验结果）之和渐进服从某一分布的定理称为中心极限定理.

由于无穷个随机变量之和可能趋于∞，故我们不研究n个随机变量之和本身而考虑它的标准化的随机变量.

即考虑随机变量$X_k\ (k=1,\cdots n)$的和$\sum_{k=1}^{n} X_k$

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

讨论Y_n的极限分布是否为标准正态分布

二、中心极限定理

为了使问题简单，这里讨论的是中心极限定理中比较特殊的情形---独立同分布的随机变量的和。

林德伯格—列维中心极限定理（独立同分布的中心极限定理）

随机变量和的中心化

$\cdots, X_n, \cdots$ 独立同分布,且该分布的数学期望 μ 和方差 $\sigma^2 > 0$ 存在, 则对于任意 x, 有

$$\lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \le x\right\} = \Phi_0(x)$$

无论随机变量序列服从何
它们渐进正态分布，这说
态分布中心地位

独立同分布的中心极限定理：

独立、同分布、同期望、同方差的（几个）随机变量之和[的标准化变量]（当n充分大时）近似服从[标准]正态分布。

林德伯格—列维中心极限定理

$$\lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \le x\right\} = \Phi_0(x)$$

该定理说明独立同分布随机变量和的标准化变量渐进服从标准正态分布，通常记做

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{\mathrm{n}}\sigma} \overset{a}{\sim} N(0,1)$$

当n充分大时，独立同分布随机变量和渐进服从正态分布

$$\sum_{i=1}^{n} X_i \overset{a}{\sim} N(\mathrm{n}\mu, n\sigma^2) \qquad \frac{1}{\mathrm{n}}\sum_{i=1}^{n} X_i \overset{a}{\sim} N(\mu, \frac{1}{\mathrm{n}}\sigma^2)$$

注

2、独立同分布中心极限定理的另一种形式可写为

$$\overline{X} \overset{近似地}{\sim} N(\mu,\sigma^2/n) \quad 或 \quad \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \overset{近似地}{\sim} N(0,1)$$

其中$\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$

林德伯格—列维中心极限定理有什么用？？？？？

3、在一般情况下，我们很难求出 $\sum_{k=1}^{n}X_k$ 的分布函数。但当n很大时，可用正态分布来近似求解。

应用一：　利用中心极限定理产生随机数

中心极限定理：

$x_1, x_2, \cdots, x_n$ 服从均值为μ，方差为σ^2 的某一分布，令

$$\xi = \frac{\bar{x}_n - \mu}{\sigma}\sqrt{n}$$

则当 n 充分大时，ξ 渐近地服从标准正态分布N（0，1）

注意，这个定理没有指出随机变量x 是服从什么分布的。

可以利用[0，1]均匀分布这个定理来产生标准正态分布的随机数。

神奇

现在我们产生n个[0，1]均匀分布随机数，期望为1/2，方差为1/12

1）按均匀分布随机生成 $r_1, r_2, \cdots, r_n$

2）由中心极限定理， 我们有：

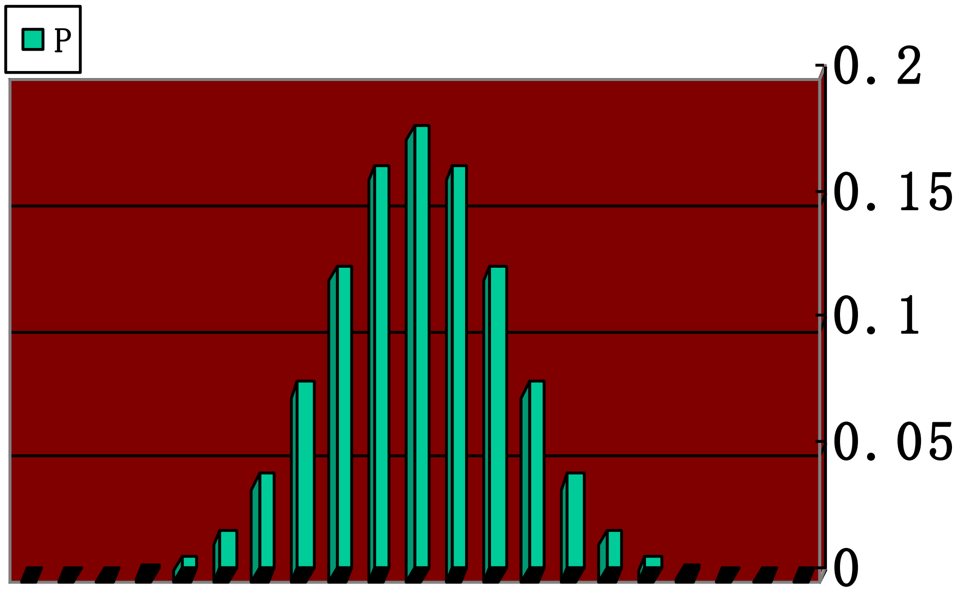
$$u = \sqrt{12n}\left(\frac{1}{n}\sum_{i=1}^{n} r_i - \frac{1}{2}\right)$$

3）为编程方便，我们特别选 n = 12，则

$$u = \sum_{i=1}^{12} r_i - 6$$

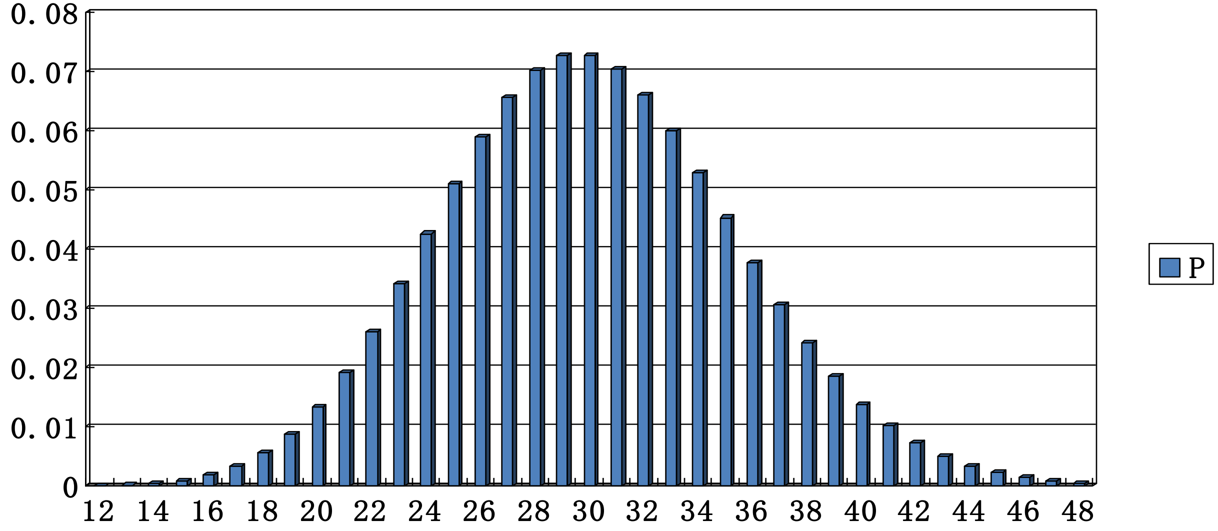
4）得到标准正态分布随机数

(1)二项分布的随机变量可看作许多相互独立的**0-1**分布的随机变量之和，下面是当ξ~**B(20,0.5)**时，ξ的概率分布图



(2)泊松分布相当于二项分布中p很小n很大的分布，当参数λ=np很大时也相当于n特别大，这个时候普阿松（泊松）分布也近似服从正态分布

下面是λ=30时的泊松概率分布图.



大数定律和中心极限定理的比较

大数定律：当n趋向于无穷大，独立同分布的随机变量的算术平均值依概率收敛到期望值，即对于任意的 $\varepsilon > 0$

$$P\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\} \to 0$$

对固定的 $\varepsilon > 0$ ，$P\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\}$有多大？

$$P\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\} = 1 - P\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \le \varepsilon\}$$

$$= 1 - P\{\left|\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}}\right| \le \frac{\varepsilon}{\sigma/\sqrt{n}}\}$$

由于 $\frac{1}{n}\sum_{i=1}^{n} X_i \overset{a}{\sim} N(\mu, \frac{1}{n}\sigma^2)$

$$= 1 - [2\Phi_0(\frac{\varepsilon}{\sigma/\sqrt{n}}) - 1] = 2[1 - \Phi_0(\frac{\varepsilon}{\sigma/\sqrt{n}})]$$

$$P\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\} = 2[1 - \Phi_0(\frac{\varepsilon}{\sigma/\sqrt{n}})]$$

说明1：当n充分大时，$\Phi_0(\frac{\varepsilon}{\sigma/\sqrt{n}}) \to 1$

从而 $P\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\} \to 0$，**即大数定律！**

说明2：当n充分大时，可以得到

$\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\} = \{\left|\sum_{i=1}^{n} X_i - n\mu\right| > n\varepsilon\}$ 概率近似值

即当n充分大时，可以得到随机变量和的**近似分布**，**即中心极限定理！！**

所以，中心极限定理比大数定律揭示的现象更深刻

例： 一盒同型号的螺丝钉共有100个，已知该型号的螺丝钉的质量是一个随机变量，期望值是100g，标准差是10g,求一盒螺丝钉的质量超过10.2kg的概率

解： 设 ξ_i是第 i个螺丝钉的质量（i=1,2,···,100）,且他们之间相互独立同分布，于是一盒螺丝钉的质量

为 $\xi = \sum_{i=1}^{100} \xi_i$，且由 $E\xi_i = 100, \sqrt{D\xi_i} = 10$

知 $E\xi = 100 \times E\xi_i = 10000, \sqrt{D\xi} = \sqrt{D\xi_i} = 100$

由中心极限定理有

$$p = P\{\xi > 10200\} = P\{\frac{\xi - 10000}{100} > \frac{10200 - 10000}{100}\}$$

$$= P\{\frac{\xi - 10000}{100} > 2\} = 1 - P\{\frac{\xi - 10000}{100} \le 2\}$$

$$= 1 - \Phi_0(2) = 0.02275$$

二项分布的正态近似

定理5.2.2　棣莫弗—拉普拉斯中心极限定理

设μ_n 为服从二项分布 $b(n, p)$ 的随机变量，则当 n 充分大时，有

$$\lim_{n\to\infty} P\left\{\frac{\mu_n - np}{\sqrt{npq}} \le y\right\} = \Phi(y)$$

是林德贝格—勒维中心极限定理的特例.

林德伯格—勒维中心极限定理推论------
棣莫弗—拉普拉斯中心极限定理

设 $\mu_n \sim B(n,p)$,则对于任意实数 x, 有

$$\lim_{n\to\infty} P\left(\frac{\mu_n - np}{\sqrt{npq}} \le x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt = \Phi_0(x)$$

结论：正态分布是二项分布的极限分布

当n充分大时，二项分布可以用正态分布来近似

即

$$\frac{\mu_n - np}{\sqrt{npq}} \overset{a}{\sim} N(0,1)$$

当n充分大时，二项分布可以用正态分布来近似
即

$$\frac{\mu_n - np}{\sqrt{npq}} \overset{a}{\sim} N(0,1)$$

$$P\{a \le \mu_n < b\} = P\{\frac{a-np}{\sqrt{npq}} \le \frac{\mu_n - np}{\sqrt{npq}} < \frac{b-np}{\sqrt{npq}}\}$$

$$\approx \Phi_0(\frac{b-np}{\sqrt{npq}}) - \Phi_0(\frac{a-np}{\sqrt{npq}})$$

查表即可得其具体值。

思考：用二项分布是否可以直接计算？为何一般不直接计算？

中心极限定理的应用

- Lindeberg-Levy中心极限定理应用

对于独立同分布随机变量序列$\{X_n\}$,不管他们服从什么分布，只要存在有限数学期望和方差，当n充分大时，就有

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

所以，$\sum_{i=1}^{n} X_i$的有关概率问题可利用正态分布求解。

- De Moivre-Laplace中心极限定理应用

对于随机变量$X \sim B(n, p)$，总有$X \overset{n\to+\infty}{\sim} N(np, npq)$，因此，当$n$充分大时，二项分布的概率问题可利用正态分布解决。一般在实际中$n \geq 50, 0.1 < p < 0.9$应用效果较理想。

三大定律的区别----附加条件不同

- **切比雪夫大数定律**：
- （1）不相关（2）期望和方差均存在（3）方差有上界

- **切比雪夫大数定律推论**：
- （1）独立同分布
- （2）期望和方差均存在

- **辛钦大数定律**：
- （1）独立同分布
- （2）期望存在

- **伯努利大数定律**：
- （1）独立同0-1分布
- （2）期望方差均相同

- **了解：**
  - 切比雪夫大数定律
  - 伯努利大数定律
  - 辛钦大数定律

  - 林德伯格-列维中心极限定理（**独立同分布随机变量的中心极限定理**）
  - 棣莫弗-拉普拉斯中心极限定理（二项分布以正态分布为极限）

- **重点: 棣莫弗—拉普拉斯中心极限定理**

大数定律和中心极限定理的比较

大数定律：当n趋向于无穷大，独立同分布的随机变量的算术平均值依概率收敛到期望值，即对于任意的 $\varepsilon > 0$

$$P\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\} \to 0$$

对固定的 $\varepsilon > 0$ ，$P\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\}$有多大？

$$P\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\} = 1 - P\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \le \varepsilon\}$$

$$= 1 - P\{\left|\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}}\right| \le \frac{\varepsilon}{\sigma/\sqrt{n}}\}$$

由于 $\frac{1}{n}\sum_{i=1}^{n} X_i \overset{a}{\sim} N(\mu, \frac{1}{n}\sigma^2)$

$$= 1 - [2\Phi_0(\frac{\varepsilon}{\sigma/\sqrt{n}}) - 1] = 2[1 - \Phi_0(\frac{\varepsilon}{\sigma/\sqrt{n}})]$$

注意点(2)

中心极限定理的应用有三大类：

i) 已知 n 和 x，求概率(求随机变量之和的概率Sn)；

ii) 已知 n 和概率，求x；

iii) 已知 x 和概率，求 n
(已知随机变量之和的概率Sn，求随机变量的个数)

一、给定 n 和 x，求概率

例 100个独立工作(工作的概率为0.9)的部件组成一个系统，求系统中至少有85个部件工作的概率.

解：用 $X_i=1$表示第i个部件正常工作, 反之记为$X_i=0$. 又记$Y=X_1+X_2+\ldots+X_{100}$，则 $E(Y)=90$，$\mathrm{Var}(Y)=9$.

由此得:

$$P\{Y \geq 85\} \approx 1-\Phi\left(\frac{85-0.5-90}{\sqrt{9}}\right)=0.966.$$

注 意 点(1)

二项分布是<u>离散分布</u>，而正态分布是<u>连续分布</u>，所以用正态分布作为二项分布的近似时，可作如下<u>修正</u>：

$$P\left(k_1 \le \mu_n \le k_2\right) = P\left(k_1 - 0.5 < \mu_n < k_2 + 0.5\right)$$

$$\approx \Phi\left(\frac{k_2 + 0.5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - 0.5 - np}{\sqrt{npq}}\right)$$

二、给定 n 和概率，求 x

例　有200台独立工作(工作的概率为0.7)的机床，每台机床工作时需15kw电力. 问共需多少电力, 才可有95%的可能性保证正常生产?

解：用　$X_i=1$表示第i台机床正常工作, 反之记为$X_i=0$. 又记$Y=X_1+X_2+\ldots+X_{200}$，则 $E(Y)=140$，$\mathrm{Var}(Y)=42$. 设供电量为y, 则从

$$P\{15Y \le y\} \approx \Phi\left(\frac{y/15+0.5-140}{\sqrt{42}}\right) \ge 0.95$$

中解得　$y \ge 2252$.

三、给定 x 和概率，求 n

例　用调查对象中的收看比例 k/n 作为某电视节目的收视率 p 的估计。要有 90% 的把握，使k/n与p的差异不大于0.05，问至少要调查多少对象?

解：用 Y_n表示n 个调查对象中收看此节目的人数，则 Y_n服从 $b(n, p)$ 分布，k 为Y_n的实际取值。根据题意

$$P\left(\left|Y_n / n - p\right| < 0.05\right) \approx 2\Phi\left(0.05\sqrt{n / p(1-p)}\right) - 1 \geq 0.90$$

从中解得 $0.05\sqrt{n / p(1-p)} \geq 1.645$

又由 $p(1-p) \leq 0.25$ 可解得 $n \geq 270.6$ 　 $n = 271$

例1 一加法器同时收到20个噪声电压，$V_k\ (k=1,2,\cdots,20)$ 设它们是互相独立的随机变量，且都在区间(0,10)上服从均匀分布，记

$$V=\sum_{k=1}^{20}V_k \qquad \text{求 } P\{V>105\}\text{的近似值。}$$

解： $EV_k=5,\ DV_k=10^2/12,\ (k=1,2,\cdots,20),$

则 $EV=20\times5=100,\qquad DV=20\times10^2/12$

$$P\{V>105\}=P\left\{\frac{V-20\times5}{\sqrt{20}\times\sqrt{10^2/12}}>\frac{105-20\times5}{\sqrt{20}\times\sqrt{10^2/12}}\right\}$$

$$=P\left\{\frac{V-100}{\sqrt{20}\times(10/\sqrt{12})}>0.387\right\}\approx1-\Phi(0.387)=0.348$$

例2 车间有200台车床，它们独立地工作着，开工率为0.6,开工时耗电各为1千瓦，问

(1)开工的车床数在100到120台之间的概率多大?

(2)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正常生产。

解：记某时刻工作着的车床数为 X, 则 X ~B(200,0.6).

$$EX = 120 \qquad DX = 48$$

由德莫佛-拉普拉斯中心极限定理, X 近似服从正态分布 $N(120, 48)$,

$$P\{a < \eta_n \le b\}$$
$$\approx \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right)$$

(1) $\quad P(100 \le X \le 120)$

$$\approx \Phi\left(\frac{120-120}{\sqrt{48}}\right) - \Phi\left(\frac{100-120}{\sqrt{48}}\right)$$

$$= \Phi(0) - \Phi(-2.887)$$

$$= 0.5 - 1 + 0.9981 = 0.4981$$

(2) 设至少要供给这个车间 r 千瓦电才能以99.9%的概率保证这个车间正常生产。由题意有

$$P\{X \le r\} \ge 0.999$$

而 $P\{X \le r\} \approx \Phi(\frac{r - 200 \times 0.6}{\sqrt{200 \times 0.6 \times 0.4}})$

$$\approx \Phi(\frac{r - 120}{\sqrt{48}}) \ge 0.999,$$

查表得 $\frac{r - 120}{\sqrt{48}} \ge 3.1$,所以 $r \ge 141.48$

即供给142千瓦电就能以99.9%的概率保证这个车间正常生产。

例3 每个学生来参加家长会的人数是一个随机变量,设其分布如下表.该校共400名学生,设各学生参加家长会的家长数相互独立,且服从同一分布. 求:

(1) 参加会议的家长数X超过450的概率.

(2) 有1名家长参加会议的学生数不多于340的概率。

X_k	0	1	2
p_k	0.05	0.8	0.15

解:
$$EX_k = 0\times0.05 + 1\times0.8 + 2\times0.15 = 1.1,$$
$$EX_k^2 = 0\times0.05 + 1\times0.8 + 4\times0.15 = 1.4,$$
$$DX_k = EX_k^2 - EX_k = 1.4 - 1.1^2 = 0.19$$

(1) 而 $X=\sum_{k=1}^{400}X_k \ ,k=1,2,\cdots,400$

又 $EX=\sum_{k=1}^{400}EX_k=400\times 1.1=440$

$$DX=\sum_{k=1}^{400}DX_k=400\times 0.19=76$$

由独立同分布的中心极限定理, X近似服$N(440,76)$

$$P(X>450)=1-P(X\le 450)$$

$$=1-P\left(\frac{X-440}{\sqrt{76}}\le\frac{450-440}{\sqrt{76}}\right)$$

$$=1-P\left(\frac{X-440}{\sqrt{76}}\le 1.147\right)\approx 1-\Phi(1.147)=0.1257$$

(2) 记Y为有一名家长来参加会议的学生数,则

$$Y \sim B(400, 0.8)$$

由德莫佛-拉普拉斯中心极限定理,

Y近似服从正态分布 $N(320, 64)$

$$P(Y \leq 340) = F(340) \approx \Phi\left(\frac{340-320}{8}\right)$$

$$= \Phi(2.5) = 0.9938$$

例2 设有一大批种子,其中良种占1/6. 试估计在任取的6000粒种子中, 良种比例与1/6相比上下不超过1%的概率.

解 记 X 为 6000 粒种子中的良种数,则 $X \sim B(n, p)$, $n = 6000,\ p = 1/6$, 因而 $EX = 1000,\ DX = 5000/6$. 所求概率为

$$P\left(\left|\frac{X}{6000} - \frac{1}{6}\right| < 0.01\right) = P\left(\left|X - 1000\right| < 60\right)$$

$$= P\left(\left|\frac{X - 1000}{\sqrt{5000/6}}\right| < \frac{60}{\sqrt{5000/6}}\right) \approx 2\Phi_0\left(\frac{60}{\sqrt{5000/6}}\right) - 1$$

$$\approx 2\Phi_0(2.078) - 1 \approx 0.9624$$

注：用二项分布 0.9590, 用泊松分布 0.9379.

例： 设某仪器有n个电子元件组成，每个电子元件的寿命服从[0,1000]上（单位为h）的均匀分布，当有20%的元件烧坏时，仪器便报废，求为使该仪器的寿命超过100h的概率不低于0.95，n至少为多大?

解： 设 X 表示仪器的寿命，X_i 表示第i个电子元器件的寿命，记 $A_i = \{X_i \geq 100\}$

ξ 表示事件 $A_i(i = 1,2,\ldots,n)$ 发生的个数;

由 $P(A_i) = 0.9$，知 $\xi \sim B(n,0.9)$，显然有

$$P\{X > 100\} = P\{\frac{\xi}{n} > 0.8\} = P\{\xi > 0.8n\}$$

$$= P\{\frac{\xi - 0.9n}{\sqrt{n \times 0.9 \times 0.1}} > \frac{0.8 - 0.9n}{\sqrt{n \times 0.9 \times 0.1}}\} \approx 1 - \Phi_0(\frac{-\sqrt{n}}{3})$$

$$= \Phi_0(\frac{\sqrt{n}}{3}) \geq 0.95$$ 从而 $\frac{\sqrt{n}}{3} \geq 1.64$，$n \geq 25$

三大定律的区别----附加条件不同

切比雪夫大数定律：

- （1）不相关（2）期望和方差均存在（3）方差有上界

切比雪夫大数定律推论：

- （1）独立同分布
- （2）期望和方差均存在

辛钦大数定律：

- （1）独立同分布
- （2）期望存在

伯努利大数定律：

- （1）独立同0-1分布
- （2）期望方差均相同

- **了解：**
  - 切比雪夫大数定律
  - 伯努利大数定律
  - 辛钦大数定律

  - 林德伯格-列维中心极限定理（**独立同分布随机变量的中心极限定理**）
  - 棣莫弗-拉普拉斯中心极限定理（二项分布以正态分布为极限）

- 重点: **棣莫弗—拉普拉斯中心极限定理**

二、中心极限定理

为了使问题简单，这里讨论的是中心极限定理中比较特殊的情形---独立同分布的随机变量的和。

林德伯格—列维中心极限定理（独立同分布的中心极限定理）

设随机变量 $X_1, X_2, \cdots, X_n, \ldots$ 独立同分布,且该分布的数学期望μ 和方差 $\sigma^2 > 0$ 存在, 则对于任意 x, 有

$$\lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \le x\right\} = \Phi_0(x)$$

独立同分布的中心极限定理:

独立、同分布、同期望、同方差的（几个）随机变量之和[的标准化变量]（当n充分大时）近似服从[标准]正态分布。

注意点(2)

中心极限定理的应用有三大类:

i) 已知 n 和 x，求概率(求随机变量之和的概率Sn)；

ii) 已知 n 和概率，求x；

iii) 已知 x 和概率，求 n
(已知随机变量之和的概率Sn，求随机变量的个数)

例3 检查员逐个地检查某种产品, 每检查一个产品需要用10秒钟．但有的产品需重复检查一次，再用去10秒钟. 假设产品需要重复检查的概率为 0.5, 求检验员在 8 小时内检查的产品多于1900个的概率.

解 检验员在 8 小时内检查的产品多于1900个, 等价于说检查1900个产品所用的时间小于8 小时. 设 T 为检查1900 个产品所用的时间(单位秒), 则所求概率为

$$P\{T < 3600 \times 8\}.$$

设 T_k 为检查第 k 个产品所用的时间 (单位: 秒), $k = 1, \ldots, 1900$, 则 $T = T_1 + T_2 + \cdots + T_{1900}$.

依题意知 $T_1, \cdots, T_{1900}$ 独立且同服从如下分布:

T_k	10	20
P	0.5	0.5

所以 $T = T_1 + T_2 + \cdots + T_{1900}$ 近似服从 $N(\mu, \sigma^2)$, 其中

$$\mu = E(T_1) + E(T_2) + \cdots + E(T_{1900}) = 28500,$$

$$\sigma^2 = D(T_1) + D(T_2) + \cdots + D(T_{1900}) = 47500,$$

所以 $P\{T < 3600 \times 8\} \approx \Phi\left(\dfrac{28800 - 28500}{\sqrt{47500}}\right) \approx 0.9162$

例3 检查员逐个地检查某种产品, 每检查一个产品需要用10秒钟．但有的产品需重复检查一次，再用去10秒钟. 假设产品需要重复检查的概率为 0.5, 求检验员在 8 小时内检查的产品多于1900个的概率.

解 检验员在 8 小时内检查的产品多于1900个, 等价于说, 检查1900个产品所用的时间T 小于8小时.

设 X 表示1900个产品中需要重复检验的个数, 则根据题意可知, $X \sim B(1900, 0.5)$, $T = 10 \times 1900 + 10X$, 且所求概率为 $p = P\{T < 3600 \times 8\}$.

$$p = P\{X < 980\} = P\left\{\frac{X - EX}{\sqrt{DX}} < \frac{980 - 950}{\sqrt{475}}\right\}$$

$$\approx \Phi\left(\frac{30}{\sqrt{475}}\right) \approx 0.9162$$

例 甲乙两电影院在竞争1000名观众，假设每位观众在选择时随机的，且彼此相互独立，问甲至少应设多少个座位，才能使观众因无座位而离去的概率小于1%？

设X表示来甲电影院的人数，甲至少设N个座位。

则$X \sim b(1000, 0.5)$, $E(X)=500$, $D(X)=250$,

由中心极限定理 $\dfrac{X-500}{\sqrt{250}} \overset{\text{近似}}{\sim} N(0,1)$

故 $P\{X>N\}$

$$=P\left\{\frac{X-500}{\sqrt{250}}>\frac{N-500}{\sqrt{250}}\right\}=1-\Phi\left(\frac{N-500}{\sqrt{250}}\right)$$

于是 $P\{X>N\}=P\left\{\frac{X-500}{\sqrt{250}}>\frac{N-500}{\sqrt{250}}\right\}$

$$=1-\Phi\left(\frac{N-500}{\sqrt{250}}\right)<1\%$$

即 $\Phi\left(\frac{N-500}{\sqrt{250}}\right)>99\%$

因 $\Phi(2.327)=0.99$ 所以 $\frac{N-500}{\sqrt{250}}>2.327,$

解得 $N>536.79$ 即 $N=537$

四、小结

中心极限定理

独立同分布中心极限定理

$$\begin{cases} E(X_k)=\mu, D(X_k)=\sigma^2 \\ \Rightarrow \sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(n\mu, n\sigma^2) \end{cases}$$

棣莫弗－拉普拉斯中心极限定理

$$\begin{cases} \eta_n \sim b(n,p) \\ \Rightarrow \eta_n \overset{\text{近似地}}{\sim} N(np, np(1-p)) \end{cases}$$

- **了解：**
  - 切比雪夫大数定律
  - 伯努利大数定律
  - 辛钦大数定律

  - 列维—林德伯格中心极限定理（**独立同分布随机变量的中心极限定理**）
  - 棣莫弗-拉普拉斯中心极限定理（**二项分布以正态分布为极限**）

- **重点: 棣莫弗—拉普拉斯中心极限定理**

作业 126（154）页 3,5, 7

补充题： 甲乙两电影院在竞争1000名观众，假设每位观众在选择时随机的，且彼此相互独立，问甲至少应设多少个座位，才能使观众因无座位而离去的概率小于1%？

补充题4：供电站供应某小区1000户居民用电，各户用电相互独立。已知每户日用电量（单位：千万时）服从区间[0,20]上的均匀分布，

（1）求这1000户居民每日用电量超过10100千瓦时的概率；

（2）若以0.95的概率保证居民供电需求，问每天至少向该小区供电多少千瓦？（结果只保留整数）