第二节　中心极限定理

一、问题的引入

二、基本定理

三、典型例题

四、小结

一、问题的引入

实例: 考察射击命中点与靶心距离的偏差.

这种偏差是大量微小的偶然因素造成的微小误差的总和, 这些因素包括: 瞄准误差、测量误差、子弹制造过程方面 (如外形、重量等) 的误差以及射击时武器的振动、气象因素(如风速、风向、能见度、温度等) 的作用, 所有这些不同因素所引起的微小误差是相互独立的, 并且它们中每一个对总和产生的影响不大.

问题: 某个随机变量是由大量相互独立且均匀小的随机变量相加而成的, 研究其概率分布情况.

中心极限定理的客观背景

在实际问题中，常常需要考虑许多随机因素所产生总影响.

例如：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.

如瞄准时的误差,

空气阻力所产生的误差,

测量误差、子弹制造过程方面 (如外形、重量等) 的误差

射击时武器的振动、气象因素(如风速、风向、能见度、温度等)

对我们来说重要的是这些随机因素的总影响.

自从高斯指出测量误差服从正态分布之后，人们发现，正态分布在自然界中极为常见.

观察表明，如果一个量是由大量相互独立的随机因素的影响所造成，而每一个别因素在总影响中所起的作用不大. 则这种量一般都服从或近似服从正态分布.

考虑 $Z_n = \dfrac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}}$ 的分布函数的极限.

标准化

可以证明，满足一定的条件，上述极限分布是标准正态分布.这就是下面要介绍的

中心极限定理

在概率论中，习惯于把和的分布收敛于正态分布这一类定理都叫做中心极限定理.

我们只讨论几种简单情形.

二、基本定理

定理四（独立同分布的中心极限定理）

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立, 服从同一分布, 且具有数学期望和方差：$E(X_k) = \mu$, $D(X_k) = \sigma^2 > 0 \ (k = 1,2,\cdots)$, 则随机变量之和的

标准化变量 $$Y_n = \frac{\sum_{k=1}^{n} X_k - E\left(\sum_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n} X_k\right)}} = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\,\sigma}$$

的分布函数 $F_n(x)$ 对于任意 x 满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{ \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\,\sigma} \le x \right\}$$

$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t = \Phi(x).$$

定理四表明:

当 $n \to \infty$, 随机变量序列 Y_n 的分布函数收敛于标准正态分布的分布函数.

推论1： 设相互独立的随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 服从同一分布，已知均值为 μ ，方差为 $\sigma^2 > 0$ 但分布函数未知，当 n 充分大时，$X = \sum_{k=1}^{n} X_k$ 近似服从正态分布 $N(n\mu, (\sigma\sqrt{n})^2)$

推论1表明:

无论各个随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 服从什么分布，只要满足定理的条件，那么它们的和 $\sum_{k=1}^{n} X_k$ 当n很大的时候，近似地服从正态分布·

(如实例中射击偏差服从正态分布)

推论2 设相互独立的随机变量 $X_1, X_2, \cdots, X_n, \cdots$

服从同一分布，已知均值为 μ ，方差为 $\sigma^2 > 0$

但分布函数未知，当n 充分大时，$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$

近似服从正态分布 $N(\mu, (\frac{\sigma}{\sqrt{n}})^2).$

虽然在一般情况下，我们很难求出 $X_1+X_2+\ldots+X_n$ 的分布的确切形式，但当 n 很大时，可以求出近似分布.

定理五(德莫佛—拉普拉斯定理)

设随机变量 $\eta_n\ (n=1,2,\cdots)$ 服从参数为 n, p $(0<p<1)$的二项分布, 则对于任意 x, 恒有

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

证明 $\eta_n = \sum_{k=1}^{n} X_k,$

其中 $X_1, X_2, \cdots, X_n$ 是相互独立的、服从同一 (0—1) 分布的随机变量, 分布律为

$$P\{X_k = i\} = p^i (1-p)^{1-i}, \qquad i = 0, 1.$$

$$\because E(X_k) = p, \quad D(X_k) = p(1-p) \quad (k = 1, 2, \cdots, n),$$

根据定理四得

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

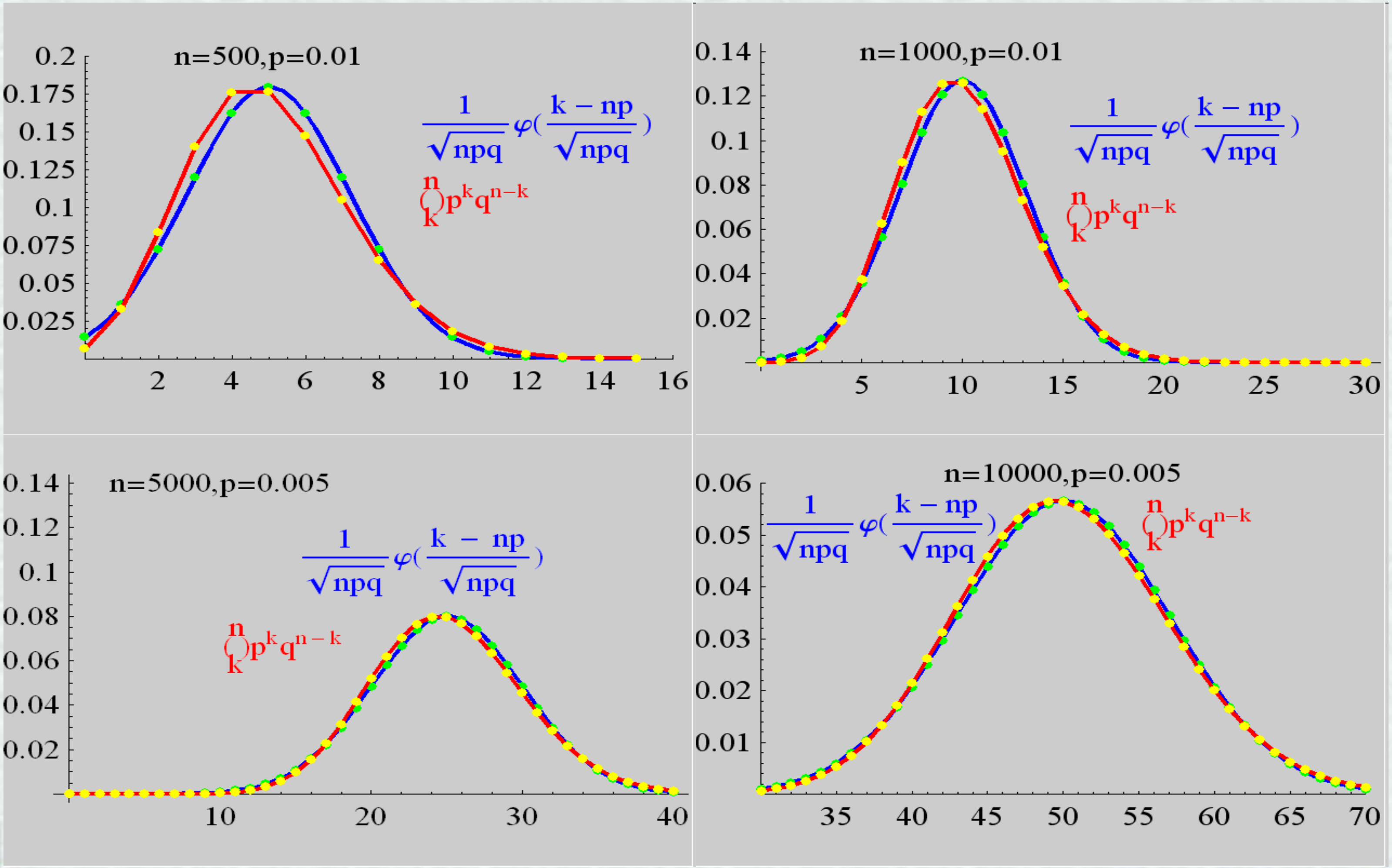
$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t = \Phi(x).$$

$$P(a < n_A \le b) \approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right)$$

定理五表明:

正态分布是二项分布的极限分布, 当**n**充分大时, 可以利用该定理来计算二项分布的概率.

下面的图形表明:正态分布是二项分布的逼近.



三、典型例题

例1：设 $X_1, X_2, \ldots, X_n$ **为独立同分布的随机变量序列，且均服从参数为** λ **的指数分布，记** $\Phi(x)$ **为标准正态分布函数，则（C ）.**

$$A.\ \lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - n\lambda}{\lambda\sqrt{n}} \le x\right\} = \Phi(x)$$

$$C.\ \lim_{n\to\infty} P\left\{\frac{\lambda\sum_{i=1}^{n} X_i - n\lambda}{\sqrt{n}} \le x\right\} = \Phi(x)$$

$$B.\ \lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - n\lambda}{\sqrt{n\lambda}} \le x\right\} = \Phi(x)$$

$$D.\ \lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - \lambda}{\sqrt{n\lambda}} \le x\right\} = \Phi(x)$$

解：独立、同分布，期望、方差存在，可见条件齐了。这里考察结论

$$E(X)=\theta$$

$$D(X)=\theta^2$$

$$\theta=\frac{1}{\lambda}$$

$$\sum_{i=1}^{n} X_i \text{近似} N\left(\frac{n}{\lambda},\frac{n}{\lambda^2}\right)$$

标准化

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

例2 计算器在进行加法时，将每个加数舍入最靠近它的整数。设所有舍入误差 $X_1, X_2, \ldots, X_n$ 是独立的，且在 [-0.5 , 0.5] 上服从均匀分布。1) 若将1500个数相加，问误差总和的绝对值超过15的概率是多少？ 2） 最多可以有几个数相加才能使得误差总和的绝对值小于10的概率不小于0.90?

解

解 $\because E(X_k)=0,\quad D(X_k)=\dfrac{1}{12}\,(k=1,2,\cdots,1500).$

$$E(X)=\int_{-\infty}^{+\infty}xp(x)\mathrm{d}x=\int_a^b\frac{1}{b-a}x\mathrm{d}x=\frac{1}{2}(a+b).$$

$$DX=EX^2-(EX)^2=\frac{(b-a)^2}{12}$$

由定理四, 随机变量 $Z=\dfrac{\sum\limits_{k=1}^{1500}X_k-1500\times 0}{\sqrt{1500\times\dfrac{1}{12}}}$ 近似服从正态分布 $N\,(0,1)$,

$$P\left\{\left|\sum_{k=1}^{1500} X_k\right| > 15\right\} = P\left\{\left|\frac{\sum_{k=1}^{1500} X_k}{\sqrt{125}}\right| > \frac{15}{\sqrt{125}}\right\} \approx 2 - 2\Phi\left(\frac{15}{\sqrt{125}}\right)$$

$$= 2\left[1 - \Phi\left(1.3416\right)\right] = 0.1796$$

$$P\left\{\left|\sum_{k=1}^{n} X_k\right| < 10\right\} = P\left\{\left|\frac{\sum_{k=1}^{n} X_k - 0}{\sqrt{n/12}}\right| < \frac{10}{\sqrt{n/12}}\right\} \geq 0.9$$

即 $$2\Phi\left(\frac{10\sqrt{12}}{\sqrt{n}}\right) - 1 \geq 0.9$$

$$\Phi\left(\frac{10\sqrt{12}}{\sqrt{n}}\right) \geq 0.95$$

例3　某保险公司的老年人寿保险有1万人参加,每人每年交200元. 若老人在该年内死亡,公司付给家属1万元. 设老年人死亡率为0.017,试求保险公司在一年内的这项保险中亏本的概率.

解　设 X 为一年中投保老人的死亡数,

则 $X \sim B(n, p)$,

其中 $n = 10000$,　　$p = 0.017$,

由**德莫佛—拉普拉斯定理**知,

$$P(a < Y < b) \approx \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

保险公司亏本的概率

$$P\{10000X > 10000\times 200\} = P\{X > 200\}$$

$$= P\left\{\frac{X-np}{\sqrt{np(1-p)}} > \frac{200-np}{\sqrt{np(1-p)}}\right\}$$

$$= P\left\{\frac{X-np}{\sqrt{np(1-p)}} > 2.321\right\}$$

$$\approx 1-\Phi(2.321) \approx 0.01 .$$

例4 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立, 且 X_i 在区间 $(-1, 1)$ 上服从均匀分布 $(i = 1, 2, \cdots, n)$, 试证当 n 充分大时, 随机变量 $Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ 近似服从正态分布, 并指出其分布参数.

证 记 $Y_i = X_i^2$, $(i = 1,2,\cdots,n)$

$$E(Y_i) = E(X_i^2) = D(X_i) = \frac{1}{3},$$

$$D(Y_i) = E(Y_i^2) - [E(Y_i)]^2 = E(X_i^4) - [E(Y_i)]^2.$$

$$f(x)=\begin{cases}0, & 其它\\ \dfrac{1}{b-a}, & x\in[a,b]\end{cases}$$

因为 $E(X_i^4)=\int_{-1}^{1}x_i^4\cdot\frac{1}{2}\mathrm{d}x_i=\frac{1}{5}$,

所以 $D(Y_i)=\frac{1}{5}-\left(\frac{1}{3}\right)^2=\frac{4}{45}$,

因为 $X_1,X_2,\cdots,X_n$ 相互独立,

所以 $Y_1,Y_2,\cdots,Y_n$ 相互独立.

根据独立同分布的中心极限定理,

$$n \cdot Z_n = \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i$$

近似服从正态分布 $N\left(\frac{n}{3}, \frac{4n}{45}\right)$,

故 Z 近似地服从正态分布 $N\left(\frac{1}{3}, \frac{4}{45n}\right)$.

例5 根据以往经验，某种电器元件的寿命服从均值为100小时的指数分布. 现随机地取16只，设它们的寿命是相互独立的. 求这16只元件的寿命的总和大于1920小时的概率.

解：设第i只元件的寿命为X_i, $i=1,2, \ldots,16$

由题给条件知，诸X_i独立，

$E(X_i)=100$, $D(X_i)=10000$

16只元件的寿命的总和为 $Y=\sum_{k=1}^{16} X_k$

依题意，所求为$P(Y>1920)$

$E(X_i)=100, D(X_i)=10000$ $\quad Y=\sum_{k=1}^{16} X_k$

依题意，所求为$P(Y>1920)$

由于$E(Y)=1600$, $\quad D(Y)=160000$

由中心极限定理, $\dfrac{Y-1600}{400}$近似$N(0,1)$

$$P(Y>1920)=1-P(Y\le 1920)\approx 1-\Phi\left(\frac{1920-1600}{400}\right)$$

$$=1-\Phi(0.8)=1-0.7881=0.2119$$

例6：设某工厂有400台同类机器，各台机器发生故障的概率都是0.02，各台机器工作是相互独立的，试求机器出故障的台数不小于2的概率。

解：设机器出故障的台数为X,则$X \sim b(400, 0.02)$，分别用三种方法计算：

1. 用二项分布计算

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - 0.98^{400} - 400 \times 0.02 \times 0.98^{399} = 0.9972$$

2. 用泊松分布近似计算

$$\lambda = np = 400 \times 0.02 = 8 \ ,$$

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) \approx 1 - 0.000335 - 0.002684 \approx 0.9969.$$

3. 用正态分布近似计算

$$\sqrt{npq} = \sqrt{400 \times 0.02 \times 0.98} = 2.8$$

$$P(X \geq 2) = 1 - P(X \leq 1) \approx 1 - \Phi\left(\frac{1 - np}{\sqrt{npq}}\right)$$

$$= \Phi\left(\frac{7}{2.8}\right) = 0.9938$$

例7　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.

(1) 设 $X_k = \begin{cases} 1 & \text{第}k\text{次取到号码}0 \\ 0 & \text{否则} \end{cases}$，$k$=1,2, …

问对序列$\{X_k\}$,能否应用大数定律?

解: $X_k \sim \begin{Bmatrix} 1 & 0 \\ 0.1 & 0.9 \end{Bmatrix}$, 　$E(X_k)$=0.1, 　k=1,2, …

诸X_k 独立同分布，且期望存在，故能使用大数定律.

解：$X_k \sim \begin{Bmatrix} 1 & 0 \\ 0.1 & 0.9 \end{Bmatrix}$, $E(X_k)=0.1$, $k=1,2,\ldots$

诸X_k 独立同分布，且期望存在，故能使用大数定律.

即对任意的$\varepsilon>0$,

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^{n} X_k - 0.1| < \varepsilon\} = 1$$

(2) 至少应取球多少次才能使“0”出现的频率在0.09-0.11之间的概率至少是0.95?

解：设应取球n次，0出现频率为 $\frac{1}{n}\sum_{k=1}^{n}X_k$

$$E\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=0.1,\quad D\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{0.09}{n}$$

由中心极限定理

$$\frac{\sum_{k=1}^{n}X_k-0.1n}{0.3\sqrt{n}}=\frac{\frac{1}{n}\sum_{k=1}^{n}X_k-0.1}{0.3/\sqrt{n}}$$ 近似$N(0,1)$

$$\frac{\frac{1}{n}\sum_{k=1}^{n} X_k - 0.1}{0.3/\sqrt{n}} \text{ 近似} N(0,1)$$

$$P\{0.09 \le \frac{1}{n}\sum_{k=1}^{n} X_k \le 0.11\}$$

$$= P\{|\frac{1}{n}\sum_{k=1}^{n} X_k - 0.1| \le 0.01\}$$

$$= P\{|\frac{\frac{1}{n}\sum_{k=1}^{n} X_k - 0.1}{0.3/\sqrt{n}}| \le \frac{\sqrt{n}}{30}\} \approx 2\Phi(\frac{\sqrt{n}}{30}) - 1$$

欲使 $$2\Phi(\frac{\sqrt{n}}{30})-1\geq 0.95$$

即 $$\Phi(\frac{\sqrt{n}}{30})\geq 0.975$$

查表得 $$\frac{\sqrt{n}}{30}\geq 1.96$$

从中解得 $$n\geq 3458$$

即至少应取球3458次才能使“0”出现的频率在0.09-0.11之间的概率至少是0.95.

(3) 用中心极限定理计算在100次抽取中，数码“0”出现次数在7和13之间的概率.

解：在100次抽取中，数码“0”出现次数为$\sum_{k=1}^{100} X_k$

由中心极限定理，

$$\frac{\sum_{k=1}^{100} X_k - \sum_{k=1}^{100} E(X_k)}{\sqrt{\sum_{k=1}^{100} D(X_k)}}$$

近似$N(0,1)$

$$E(X_k)=0.1,\ D(X_k)=0.09$$

即 $\dfrac{\sum_{k=1}^{100} X_k - 10}{3}$ 近似$N(0,1)$

$$\frac{\sum_{k=1}^{100} X_k - 10}{3}$$ 近似$N(0,1)$

$$P(7 \le \sum_{k=1}^{100} X_k \le 13) = P(-1 \le \frac{\sum_{k=1}^{100} X_k - 10}{3} \le 1)$$

$$\approx \Phi(1) - \Phi(-1)$$

$$= 2\Phi(1) - 1 = 0.6826$$

即在100次抽取中，数码“0”出现次数在7和13之间的概率为0.6826.

四、小结

三个中心极限定理
- 独立同分布的中心极限定理
- 德莫佛－拉普拉斯定理

中心极限定理表明, 在相当一般的条件下, 当独立随机变量的个数增加时，其和的分布趋于正态分布.

第三章内容总框图

数字特征的引入

矩、协方差矩阵

协方差
相关系数

几种重要分布的期望和方差

数学期望
E(X)

方差
$D(X)= E(X-E(X))^2$

离散型
连续型

随机变量函数的数学期望

离散型
连续型

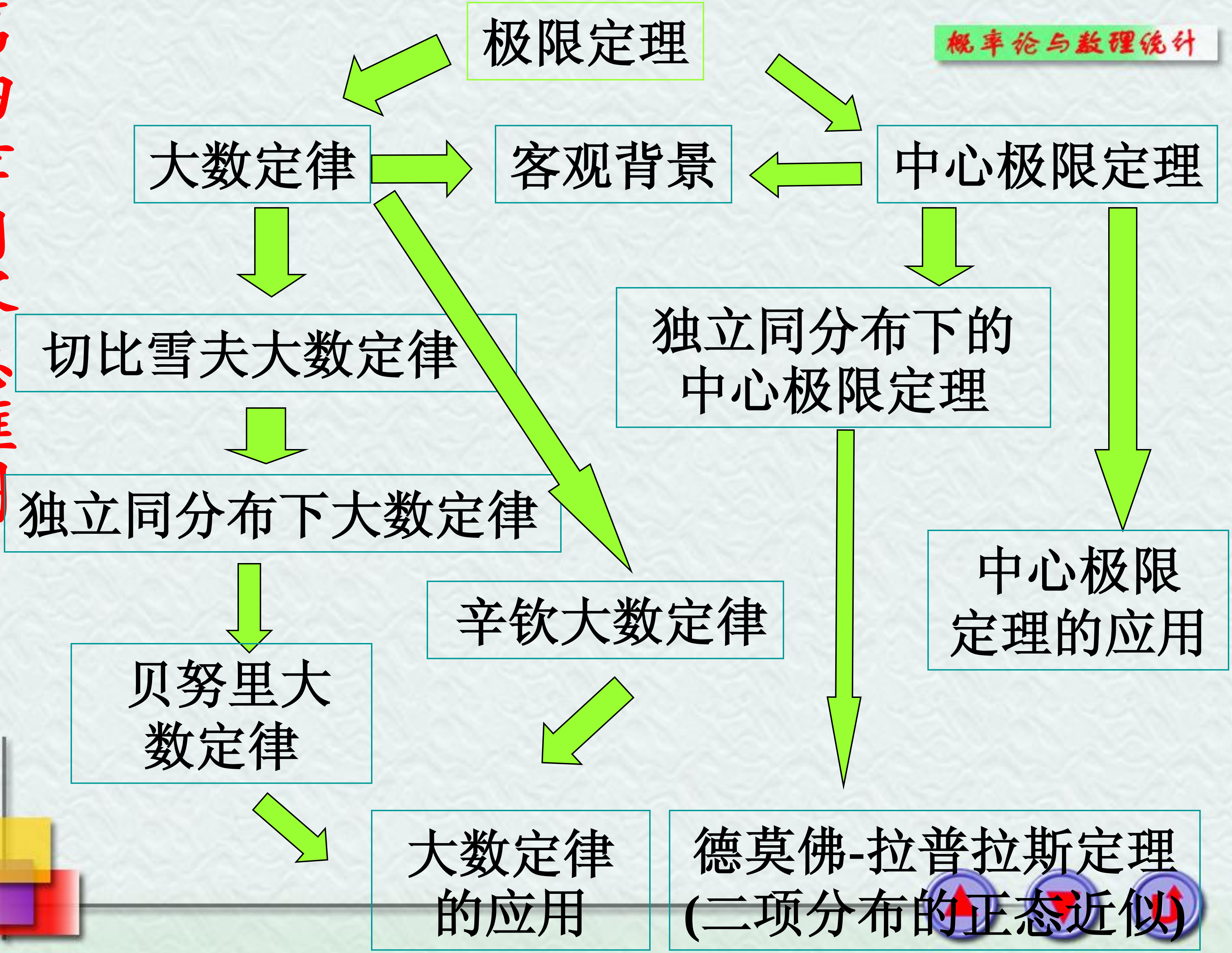
期望性质

多维随机变量函数的数学期望

方差性质

性质的应用

第四章内容总框图


极限定理
大数定律
客观背景
中心极限定理
切比雪夫大数定律
独立同分布下的中心极限定理
独立同分布下大数定律
辛钦大数定律
中心极限定理的应用
贝努里大数定律
大数定律的应用
德莫佛-拉普拉斯定理(二项分布的正态近似)