第五章 大数定律及中心极限定理

本章所介绍的内容，是概率论中理论性较强的一部分，难度较大。但教学大纲对这部分内容要求较弱，只要求大家能用切比雪夫不等式，大数定律与中心极限定理来估计或近似计算某些事件的概率即可，并了解正态分布在近似计算中的应用。

定　理

设 RV X 有 $E(X)=\mu$，$D(X)=\sigma^2$，则 $\forall \varepsilon>0$，都有

$$P\{|X-\mu|\ge\varepsilon\}\le\frac{\sigma^2}{\varepsilon^2} \text{ 或 } P\{|X-\mu|<\varepsilon\}\ge 1-\frac{\sigma^2}{\varepsilon^2}$$

这一不等称为切比雪夫不等式(**Chebyshev Inequality**).

重要不等式

证：这里仅就连续型RV的情况来证明

设RV X 的PDF为$f(x)$

$$\because |x-\mu| \geq \varepsilon \Rightarrow |x-\mu|^2 \geq \varepsilon^2 \Rightarrow \frac{|x-\mu|^2}{\varepsilon^2} \geq 1$$

$$\therefore P\{|X-\mu| \geq \varepsilon\} = \int_{|x-\mu| \geq \varepsilon} f(x)\mathrm{d}x$$

$$\leq \int_{|x-\mu| \geq \varepsilon} \frac{|x-\mu|^2}{\varepsilon^2} f(x)\mathrm{d}x$$

$$\leq \int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\varepsilon^2} f(x)\mathrm{d}x = \frac{\sigma^2}{\varepsilon^2}$$ 得证。

$$P\{|X-\mu|<\varepsilon\}\geq 1-\frac{\sigma^2}{\varepsilon^2}$$

说 明

① 在RV *X*的分布未知的情况下，如果知道$E(X)$和$D(X)$，那么切比雪夫不等式就给出了事件$\{|X-\mu|<\varepsilon\}$概率(也即：RV *X*取值落在$(\mu\pm\varepsilon)$区间内的概率)下限的一种估计方法。

当$\varepsilon=3\sigma$时，$P\{|X-\mu|<3\sigma\}\geq 1-\frac{1}{3^2}=0.8889$

但若已知$X\sim N(\mu,\sigma^2)$，$P\{|X-\mu|<3\sigma\}=0.9974$。

$$P\left\{\left|X-\mu\right|<\varepsilon\right\}\ge 1-\frac{\sigma^2}{\varepsilon^2}$$

说 明

② 从切比雪夫不等式可以看出，当$D(X)=\sigma^2$越小，事件 $\left\{\left|X-\mu\right|<\varepsilon\right\}$ 的概率越大，这表明方差就是刻画RV的取值与其数学期望偏离程度的一个度量。

③ 切比雪夫不等式作为一个理论工具，有着广泛的应用，例如：证明大数定律。

§1 大数定律 The law of large numbers

一、客观背景

大量随机试验
- 事件发生的频率稳定于某常数

  如：抛一枚硬币，观察出现正面的情况，当试验次数很多时，硬币出现正面的频率接近1/2。

- 测量值的算术平均值具有稳定性

  如：测量一个长度*a*,每次测量结果不见得就等于*a*，但测量次数很多时，测量结果的算术平均值接近于*a*几乎是必然的。

二、大数定律的概念

概率论中用来阐明大量随机现象平均结果的稳定性的一系列定理，称为大数定理。

本章将介绍三个大数定理:

（1）切比雪夫大数定律

（2）伯努利大数定律

（3）辛钦大数定律

它们之间既有区别，也有联系。

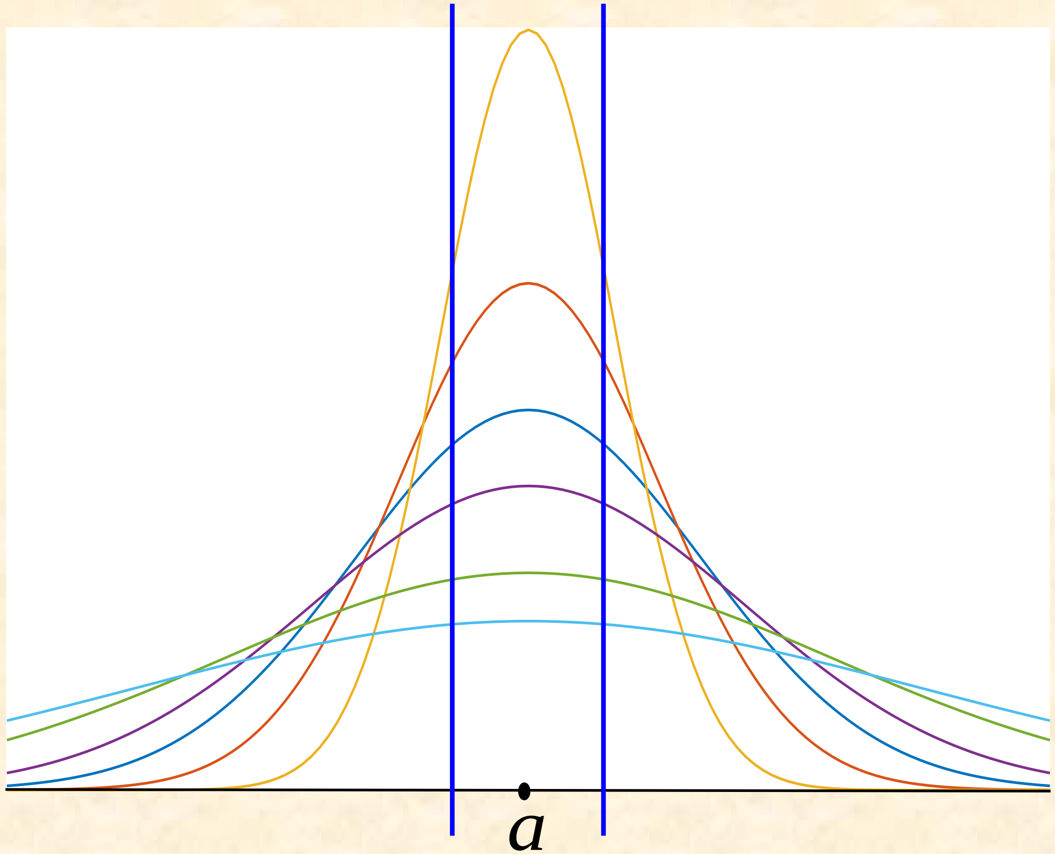
三、依概率收敛

定义 设$Y_1, Y_2, \cdots, Y_n, \cdots$ 是一个随机变量序列，a是常数，$\forall \varepsilon>0$，有

$$\lim_{n\to\infty} P\left\{\left|Y_n - a\right| < \varepsilon\right\} = 1$$

则称序列$Y_1, Y_2, \cdots, Y_n, \cdots$依概率收敛于$a$ (convergence in probability)，

记为 $$Y_n \xrightarrow{P} a$$



意义 依概率收敛表示Y_n与a的绝对误差小于任意正数ε的概率将随n增大而越来越大，直至趋于1；

也就是说，当$n\to\infty$时，Y_n取a的概率接近1。

性质 设$X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$，函数$g(x,y)$在点$(a,b)$连续，

则$g(X_n, Y_n) \xrightarrow{P} g(a,b)$。

一般地，对m元函数$g(x_1,\cdots,x_m)$，上述结论亦成立。

注意 $X_n \xrightarrow{P} a$，意味着$\forall \varepsilon > 0$，当n充分大时，事件$\{|X_n - a| < \varepsilon\}$的概率很大，接近于1；并不排除事件$\{|X_n - a| \geq \varepsilon\}$的发生，只是说他发生的可能性很小。

依概率收敛比高数中普通意义下的收敛弱些，它具有某种不确定性。

应用 在讨论未知参数的估计量是否具有一致性/相合性时，常会用到依概率收敛的性质。

四、大数定律

定理一（切比雪夫大数定理）弱大数定理

设**RV**序列$X_1,\cdots,X_n,\cdots$相互独立，有相同的期望和方差：$E\left(X_i\right)=\mu, D(X_i)=\sigma^2,(i=1,2,\cdots)$，前$n$个**RV**的算术平均值$\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$，则$\forall\varepsilon>0$，有

$$\lim_{n\to\infty}P\left\{\left|\overline{X}-\mu\right|<\varepsilon\right\}=1$$

也即 $\overline{X}\xrightarrow{P}\mu$。

要证 $\lim\limits_{n\to\infty} P\left\{\left|\bar{X}-\mu\right|<\varepsilon\right\}=1$

$$P\left\{\left|X-\mu\right|<\varepsilon\right\}\geq 1-\frac{\sigma^2}{\varepsilon^2}$$

证: **由切雪夫不等式知** $\mathrm{P}\left\{\left|\bar{X}-E\left(\bar{X}\right)\right|<\varepsilon\right\}\geq 1-\frac{D\left(\bar{X}\right)}{\varepsilon^2}$

$$E\left(\bar{X}\right)=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E\left(X_i\right)=\frac{1}{n}\cdot n\mu=\mu$$

$$D\left(\bar{X}\right)=D\left(\frac{1}{n}\sum_{i=1}^{\infty}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}D\left(X_i\right)=\frac{1}{n^2}\cdot n\sigma^2=\frac{\sigma^2}{n}$$

代入不等式有 $P\left\{\left|\bar{X}-\mu\right|<\varepsilon\right\}\geq 1-\frac{\sigma^2/n}{\varepsilon^2}$

$\therefore$ 当$n\to\infty$时, 有 $\lim\limits_{n\to\infty} P\left\{\left|\bar{X}-\mu\right|<\varepsilon\right\}=1$, 即$\bar{X}\xrightarrow{P}\mu$, 得证。

定理一：$\lim_{n\to\infty} P\left\{\left|\bar{X}-\mu\right|<\varepsilon\right\}=1$

说　明

① 当$n\to\infty$时，随机事件$\left\{\left|\bar{X}-\mu\right|<\varepsilon\right\}$的概率趋于1。即$\forall\varepsilon>0$，当$n$充分大时，不等式$\left|\bar{X}-\mu\right|<\varepsilon$成立的概率很大。

② 定理以数学形式证明了**RV** $X_1, X_2, \ldots, X_n$的算术平均$\bar{X}=\frac{1}{n}\sum_{i=1}^{n} X_i$接近$E(X_i)=\mu$。这种接近说明$\bar{X}$具有稳定性.

通俗地说，n个**RV**的算术平均$\bar{X}$,当n无限增大时将几乎变为一个常数。

切比雪夫定理（不要求）

设**RV**序列$X_1,\cdots,X_n,\cdots$相互独立, 各有数学期望$E(X_1),\cdots,E(X_n)$和方差$D(X_1),\cdots,D(X_n)$, 且$D(X_i)<C$ (C是常数), $i=1,\cdots,n$, 则$\forall\varepsilon>0$, 前n个**RV**的算术平均值

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

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

也即 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}\frac{1}{n}\sum_{i=1}^{n}E(X_i)$。

定理二（伯努利大数定理）

设n_A是n次独立重复试验中事件A发生的次数，A在每次试验中发生的概率为p，则$\forall \varepsilon > 0$，有

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

也即 $\dfrac{n_A}{n} \xrightarrow{P} p$。

要证 $\dfrac{n_A}{n} \xrightarrow{P} p$。

证： 关键n_A/n能否表示为RV序列的算术平均？

$$令X_i=\begin{cases}0, & 在第i次试验中A不发生\\ 1, & 在第i次试验中A发生\end{cases}，\quad i=1,2,\cdots,n$$

则$X_i \sim b(1,p)$，且$X_1,\ldots,X_n$相互独立，

$$故n_A=\sum_{i=1}^{n}X_i,\quad E\left(X_i\right)=p,\quad D\left(X_i\right)=p\left(1-p\right), i=1,2,\cdots,n$$

利用定理一，$X_1,\cdots,X_n$独立，且有相同E和D, 可知

$$\frac{n_A}{n} \xrightarrow{P} p \qquad 得证。$$

说　明

① 伯努利大数定理从理论上证明了频率具有稳定性。也就是说：当试验次数n充分大时，事件A发生的频率$\frac{n_A}{n}$与概率p有较大偏差的可能性很小。

② 提供了通过试验来确定事件概率的方法：

$$\frac{n_A}{n} \approx p = P(A),$$（用频率近似代替概率）。

举 例

① 在装有8只白球与2只红球的盒子中任取一只球记录记录颜色后再放回去，当抽到白球时记为0，抽到红球时记为1，不停地重复下去，就得到一组有0和1构成的数字，比如：010000000100001100010001110000

000101000100000000000000000000

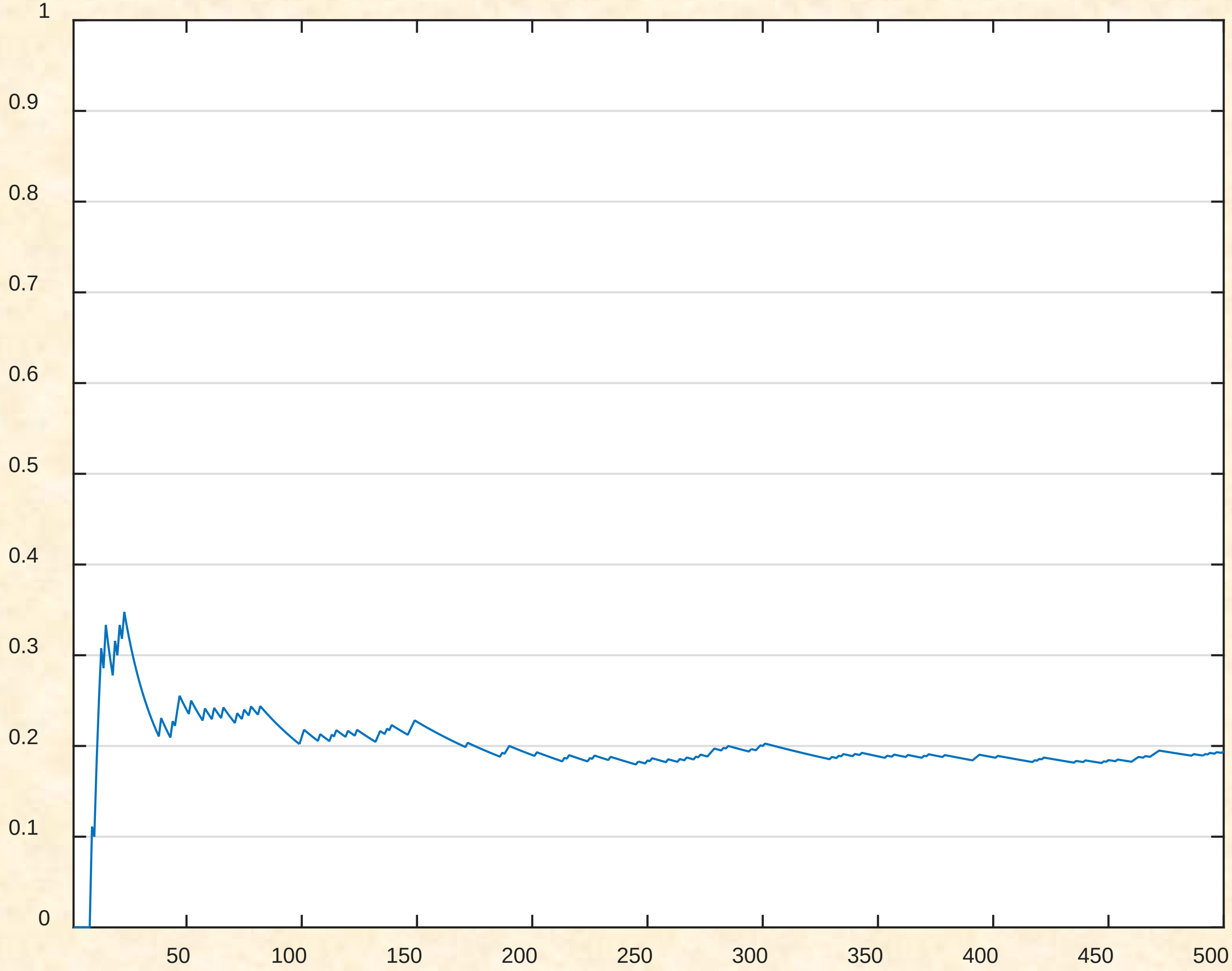
规律？数字串杂乱无章

$$\frac{n_{\text{红色}}}{n}=\frac{11}{60}\approx 0.1833$$

$$n\to\infty,\ \frac{n_{\text{红色}}}{n}=P\{\text{取得红色的球}\}=\frac{2}{10}$$

附：Matlab程序

```
R = binornd(1,0.2,1,5000);%产生二项分布随机数
P = mean(R)
```


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
50
100
150
200
250
300
350
400
450
500

定理三（辛钦大数定理）

设**RV**序列$X_1,\cdots,X_n,\cdots$相互独立, 服从同一分布, 且有数学期望$E(X_i)=\mu\ (i=1,2,\cdots)$, 则序列

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}\mu$$

说　明

① 辛钦大数定理说明了在相同条件下大量重复测量值的平均结果具有稳定性。因此，当n充分大时，可以用测量结果$X_i(1 \le i \le n)$的算术平均值$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$近似代替被观测值的真值$\mu$。

如：要估算某地区的平均亩产量，可以选择收割某些有代表性地块n块，计算其平均亩产量，则当n较大时，可以用它作为整个地区平均亩产量的一个估计。

说 明

② 伯努利大数定理 ？ 辛钦大数定理

伯努利大数定理是辛钦大数定理的特殊情况：

辛钦大数定理中**RV** X_i分布为$(0-1)$分布时的情况

③ 切比雪夫大数定理与辛钦大数定理相比：

$$\text{辛钦大数定理}\begin{cases}\text{多一条“同分布”要求}\\ \text{少一条“方差存在”要求}\end{cases}$$

本节小结

- 本节介绍了大数定律，以严格的数学形式表达了随机现象最根本的性质之一，平均结果的稳定性。

§2 中心极限定理 Central limit theorem

一、客观背景

在实际问题中，许多RV是由相互独立随机因素的综合影响所形成的。

例如：炮弹射击的落点与目标的偏差，就受到许多随机因素（如果瞄准，空气阻力，炮弹或炮身结构等）综合影响的，每个随机因素对弹着点所起的作用都很小，那么弹着点服从怎样分布?

如果一个RV是由大量相互独立的随机因素的综合影响所造成，而每一个因素对这种综合影响所起的作用不大，则这种RV一般都服从or近似服从正态分布。

二、中心极限定理定义

概率论中有关论证独立RV的和的极限分布是正态分布的一系列定理称为中心极限定理。

由于无穷个RV之和可能趋于∞，故我们不研究n个RV之和本身，而考虑它的标准化RV，以此来讨论其极限分布。

三、中心极限定理

定理一（独立同分布的中心极限定理）

设**RV**序列$X_1,\cdots,X_n,\cdots$相互独立且服从同一分布，具有期望和方差：$E(X_i)=\mu, D(X_i)=\sigma^2>0,$

$(i=1,2,\cdots)$, 则**RV**之和$\sum_{i=1}^{n}X_i$的标准化变量

$$Y_n=\frac{\sum_{i=1}^{n}X_i-E\left(\sum_{i=1}^{n}X_i\right)}{\sqrt{D\left(\sum_{i=1}^{n}X_i\right)}}=\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\,\sigma}$$

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

Y_n的分布函数$F_n(x)$对任意x满足

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

证略。

说 明

① 定理给出了**相互独立同分布的*n*个RV之和的近似分布**：当***n***充分大时，

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} \overset{近似}{\sim} N(0,1) \quad 或 \quad \sum_{i=1}^{n} X_i \overset{近似}{\sim} N\left(n\mu, n\sigma^2\right)$$

② 定理也给出了相互独立同分布的**n**个**RV**算术平均的近似分布，

$$\text{由于 } Y_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}}$$

因此，当**n**充分大时有

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \overset{\text{近似}}{\sim} N(0,1) \quad \text{或} \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

③ 虽然一般情况下，很难求出**RV**之和的确切分布，但当**n**很大时，可以求出其它近似分布。

定理二（李雅普诺夫定理）

设**RV**序列$X_1,\cdots,X_n,\cdots$相互独立,且具有期望和方差

$E(X_i)=\mu_i$,$D(X_i)=\sigma_i^2>0,(i=1,2,\cdots)$, 记$B_n^2=\sum_{i=1}^{n}\sigma_i^2$,

若存在$\delta>0$, 使得当$n\to\infty$时,

$$\frac{1}{B_n^{2+\delta}}\sum_{i=1}^{n}E\left\{\left|x_i-\mu_i\right|^{2+\delta}\right\}\to 0$$

则**RV**之和$\sum_{i=1}^{n}X_i$的标准化变量

$$Z_n = \frac{\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)}{\sqrt{D\left(\sum_{i=1}^{n} X_i\right)}} = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{B_n}$$

Z_n的分布函数$F_n(x)$对任意x满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\{Z_n \le x\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = \Phi(x)$$

说 明

① 在上述定理条件下，当**n**充分大时

$$\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N\left(\sum_{i=1}^{n} \mu_i, B_n^2\right) \quad \text{或} \quad \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{B_n} \overset{\text{近似}}{\sim} N(0,1)$$

② **RV**无论服从什么分布，只要满足条件，则**RV**之和 $\sum_{i=1}^{n} X_i$ ，当**n**很大时，就近似服从正态分布，这就是为什么正态分布在概率论中所占重要地位的一个基本原因。

独立同分布的中心极限定理的特殊情况

定理三（棣莫弗-拉普拉斯定理）

设**RV** $\eta_n(n=1,2,\cdots)\sim b(n,p),(0<p<1)$，则$\forall x\in R$，有

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

说明　定理表明：当**n**充分大时

$$\eta_n \overset{\text{近似}}{\sim} N\left(np,np(1-p)\right)$$

即　<u>二项分布的极限分布是正态分布</u>

若$\eta_n(n=1,2,\cdots)\sim b(n,p)$，要证$\eta_n \overset{\text{近似}}{\sim} N\left(np, np(1-p)\right)$

证：由n重伯努利试验背景易知

$$\eta_n=\sum_{i=1}^{n}X_i,\quad \text{其中} X_i\sim b(1,p),\ i=1,\cdots,n$$

$$\text{且知} E\left(X_i\right)=p,\ \ D\left(X_i\right)=p\left(1-p\right), i=1,2,\cdots,n$$

利用独立同分布的中心极限定理得

$$\lim_{n\to\infty}\mathrm{P}\left\{\frac{\sum_{i=1}^{n}X_i-E\left(\sum_{i=1}^{n}X_i\right)}{\sqrt{D\left(\sum_{i=1}^{n}X_i\right)}}\le x\right\}=\lim_{x\to\infty}\mathrm{P}\left\{\frac{\eta_n-np}{\sqrt{np(1-p)}}\le x\right\}=\Phi\left(x\right)\ \text{得证。}$$

本节小结

- 本节介绍了中心极限定理，它不仅提供了**计算独立RV之和的近似概率的简单方法**，而且有助于解释为什么很多自然群体的经验频率呈现出钟形(正态)曲线这一事实。因此，正态分布在概率统计中有广泛的应用。

例1 已知某产品的次品率为0.04，现有这样一批产品100件，求这批产品中不少于4件次品的概率？

解:设$X=\{100件产品中的次品数\}$,则$X \sim b(100, 0.04)$

需求 $P\{4 \le X \le 100\}$ ？

分别采用二项分布、泊松近似、正态近似来计算。

(1) 二项分布 **$X \sim b(100, 0.04)$**

$$P\{4 \le X \le 100\} = 1 - P\{X < 4\}$$

$$= 1 - \sum_{k=0}^{3} C_n^k p^k (1-p)^{100-k} \approx 0.5705$$

(1) 二项分布 $X \sim b(100, 0.04)$， $P\{4 \le X \le 100\} \approx 0.5705$

(2) 泊松近似 $X \overset{\text{近似}}{\sim} \pi(\lambda)$， $\lambda = np = 100 \times 0.04 = 4$

$$P\{4 \le X \le 100\} = 1 - P\{X < 4\}$$

$$= 1 - \sum_{k=0}^{3} 4^k e^{-4} / k! \approx 0.5669$$

(3) 正态近似 $X \overset{\text{近似}}{\sim} N(np, np(1-p))$

$$np = 100 \times 0.04 = 4,\ np(1-p) = 4 \times (1 - 0.04) = 3.84$$

$$P\{4 \le X \le 100\} = \Phi\left(\frac{100-4}{\sqrt{3.84}}\right) - \Phi\left(\frac{4-4}{\sqrt{3.84}}\right)$$

$$= \Phi(48.98) - \Phi(0) \approx 1 - 0.5 = 0.5$$

说明　二项分布概率的计算

设RV $X \sim b(n, p)$，则

① 当n不太大时，直接利用**二项分布**计算:

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

② 当$n \geq 20$，$p \leq 0.05$时，利用**泊松定理**近似计算:

$$P\{X=k\}=\lambda^k e^{-\lambda}/k!, \quad k=0,1,2,\cdots, \lambda=np.$$

③ 当n较大而p不太较小时: $p<(1-p)$且$np \geq 5$时，或 $p>(1-p)$且$n(1-p) \geq 5$ 时,利用**正态分布**近似计算:

$$P\{a<X<b\}=\Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right)$$

例2 投掷一枚质地均匀的硬币，需要多少次才能保证出现正面的频率在**0.4～0.6**之间的概率不少于**0.9**？

采用两种方法：切比雪夫不等式、中心极限定理

解：设需要投n次，

$X=\{n$次投掷中硬币出现正面的次数$\}$, 则$X \sim b(n, 0.5)$

要保证 $$P\left\{0.4 \le \frac{X}{n} \le 0.6\right\} \ge 0.9$$

也即 $$P\{0.4n \le X \le 0.6n\} \ge 0.9$$

由于 $E(X)=np=0.5n,\ D(X)=np(1-p)=0.25n$

$$E(X)=0.5n,\ D(X)=0.25n$$

(1) 切比雪夫不等式

$$\mathrm{P}\left\{\left|X-E(X)\right|<\varepsilon\right\}\geq 1-\frac{D(X)}{\varepsilon^2}$$

即 $\mathrm{P}\left\{0.5n-\varepsilon\leq X\leq 0.5n+\varepsilon\right\}\geq 1-\dfrac{0.25n}{\varepsilon^2}$

要求 $\mathrm{P}\left\{0.4n\leq X\leq 0.6n\right\}\geq 0.9$

令 $\varepsilon=0.1n$, **上式即为**

$$\mathrm{P}\left\{0.4n\leq X\leq 0.6n\right\}\geq 1-\frac{0.25n}{(0.1n)^2}=1-\frac{1}{0.4n}\overset{令}{\geq}0.9$$

解得 $n\geq 250$

$$E(X)=np=0.5n,\quad D(X)=np(1-p)=0.25n$$

(2) **中心极限定理** 易知 $X \overset{\text{近似}}{\sim} N(np, np(1-p))$

则 $P\{0.4n \le X \le 0.6n\} = \Phi\left(\frac{0.6n-0.5n}{\sqrt{0.25n}}\right) - \Phi\left(\frac{0.4n-0.5n}{\sqrt{0.25n}}\right)$

$$=\Phi\left(0.2\sqrt{n}\right)-\Phi\left(-0.2\sqrt{n}\right)$$

$$=2\Phi\left(0.2\sqrt{n}\right)-1 \overset{\text{令}}{\ge} 0.9$$

$\Rightarrow \Phi\left(0.2\sqrt{n}\right) \ge 0.95$ 查表得 $0.2\sqrt{n} \ge 1.645$

解得 $n \ge 68$

本章小结

- 本章介绍了切比雪夫不等式，大数定律，中心极限定理。要求大家能够利用它们来估计or近似计算某些事件的概率即可:

  ① 用契比雪夫不等式估计概率问题。

  ② RV之和的概率问题

  ③ RV算术平均的概率问题

  ④ 二项分布的极限分布问题

作业

Pages 126, 127：

第1, 3, 4, 8, 13题

练习2：设X的PDF为 $f(x)=\begin{cases}\dfrac{x^m}{m!}e^{-x}, & x>0\\ 0, & else\end{cases}$

试证：$P\{0<X<2(m+1)\}\geq\dfrac{m}{m+1}$。

提示：$P\{0<X<2(m+1)\}$

$$=P\{-(m+1)<X-(m+1)<(m+1)\}$$

$$=P\{|X-(m+1)|<(m+1)\}$$

$E(X)=?\quad D(X)=?$

$$f(x)=\begin{cases}\dfrac{x^m}{m!}e^{-x}, x>0\\ 0, \qquad else\end{cases}$$

$$E(X)=\int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x=\int_0^{+\infty} x\frac{x^m}{m!}e^{-x}\,\mathrm{d}x$$

$$=\frac{1}{m!}\int_0^{+\infty} x^{m+1}e^{-x}\,\mathrm{d}x=\frac{1}{m!}\Gamma(m+2)=(m+1)$$

$$E(X^2)=\int_{-\infty}^{+\infty} x^2 f(x)\,\mathrm{d}x=\int_0^{+\infty} x^2\frac{x^m}{m!}e^{-x}\,\mathrm{d}x=\frac{1}{m!}\int_0^{+\infty} x^{m+2}e^{-x}\,\mathrm{d}x$$

$$=\frac{1}{m!}\Gamma(m+3)=(m+2)(m+1)$$

$$D(X)=E(X^2)-E^2(X)=(m+1)$$

切比雪夫不等式 $P\{|X-E(X)|<\varepsilon\}\geq 1-\dfrac{D(X)}{\varepsilon^2}$

令$\varepsilon=m+1$, 即 $P\{|X-(m+1)|<(m+1)\}\geq 1-\dfrac{(m+1)}{(m+1)^2}=\dfrac{m}{m+1}$ 得证

附录 PPT P18 Matlab程序

```
%%绘制频率稳定性曲线
clear all
N=0;
nframes=500;
for k=1:nframes
    N=N+1;
    x=1:1:N;
r(k)=binornd(1,0.2);%产生二项分布随机数
y(k)=mean(r); %求平均值
plot(x,y);
axis([1 nframes  0  1]);
set(gca,'Ygrid','on')% 打开 Y 坐标轴网格线
m(k)=getframe(gcf);
end
movie2avi(m,'Frequency');%转存为 AVI 格式
```