第五章　大数定律和中心极限定理

关键词：

依概率收敛

切比雪夫不等式

大数定律

中心极限定理

§1 大数定律

随机变量序列依概率收敛的定义

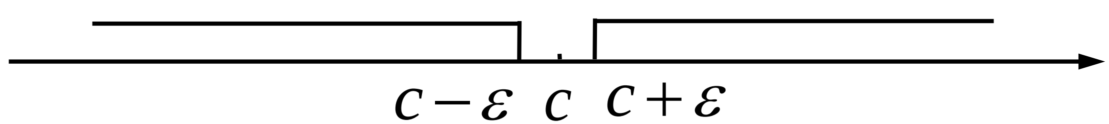
定义5.1.1： 设$Y_1,\ldots,Y_n,\ldots$(或用$\{Y_n, n\geq 1\}$记)为一个随机变量序列，c为常数量，若对于$\forall \varepsilon > 0$,均有：

$$\lim_{n\to+\infty} P\left\{\left|Y_n - c\right| \geq \varepsilon\right\} = 0$$

或

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

成立，则称随机变量序列 $\left\{Y_n, n\geq 1\right\}$ **依概率收敛于c**，记为：$Y_n \xrightarrow{P} c$, 当$n \to +\infty$.



例1：设$Y_n \sim N(0,\frac{1}{n}), n=1,2,\ldots,$则当$n \to +\infty$时，$Y_n \xrightarrow{P} 0$.

证明:对任意$\varepsilon > 0$,

$$P(|Y_n - 0| \geq \varepsilon)$$

$$= 1 - P(-\varepsilon \leq Y_n \leq \varepsilon)$$

$$= 1 - \Phi\left(\frac{\varepsilon - 0}{\sqrt{1/n}}\right) + \Phi\left(\frac{-\varepsilon - 0}{\sqrt{1/n}}\right)$$

$$= 2[1 - \Phi(\varepsilon\sqrt{n})] \xrightarrow{n \to +\infty} 0$$

$\therefore$ 当$n \to +\infty$时，$Y_n \xrightarrow{P} 0$

依概率收敛的性质：

若当$n\to\infty$时，$X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$，且函数$g(x,y)$在点$(a,b)$连续，则 $g(X_n,Y_n) \xrightarrow{P} g(a,b)$

例如：当$n\to\infty$时，

$$X_n + Y_n \xrightarrow{P} a+b,$$

$$X_n \times Y_n \xrightarrow{P} a\times b,$$

$$X_n / Y_n \xrightarrow{P} a/b \qquad (b\neq 0),$$

$$X_n e^{-Y_n} \xrightarrow{P} ae^{-b}.$$

在给出大数定律之前，先介绍两个重要不等式

马尔可夫不等式和切比雪夫不等式

定理5.1.1 (马尔可夫不等式):

设随机变量Y的k阶矩$E(Y^k)$存在$(k \geq 1)$,

则对于任意$\varepsilon > 0$,都有: $P\left\{|Y| \geq \varepsilon\right\} \leq \dfrac{E(|Y|^k)}{\varepsilon^k}$成立;

定理的等价形式为: $P\left\{|Y| < \varepsilon\right\} \geq 1 - \dfrac{E(|Y|^k)}{\varepsilon^k}$.

特别地，当Y为取非负值的随机变量时，

则有 $$P\left\{Y \geq \varepsilon\right\} \leq \frac{E(Y^k)}{\varepsilon^k}$$

证明$P\{|Y|\ge\varepsilon\}\le\dfrac{E(|Y|^k)}{\varepsilon^k}$, $k\ge 1$

证明: 对于任意$\varepsilon>0$，令$Z=\begin{cases}\varepsilon, & \text{当}|Y|\ge\varepsilon\text{时};\\ 0, & \text{当}|Y|<\varepsilon\text{时}.\end{cases}$

则$Z=\varepsilon\le|Y|$, $\quad\Rightarrow Z^k\le|Y|^k$,

$$\Rightarrow E(Z^k)\le E(|Y|^k).$$

Z	0	ε
P_k	$1-P(\|Y\|\ge\varepsilon)$	$P(\|Y\|\ge\varepsilon)$

$$E(Z^k)=\varepsilon^k P\{|Y|\ge\varepsilon\}\le E(|Y|^k)$$

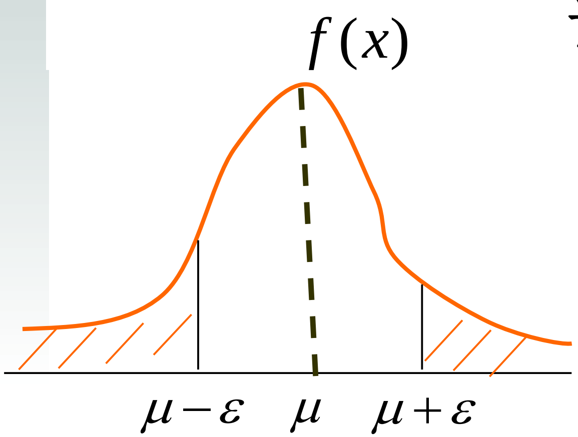
所以 $P\{|Y|\ge\varepsilon\}=\dfrac{E(Z^k)}{\varepsilon^k}\le\dfrac{E(|Y|^k)}{\varepsilon^k}$.

定理5.1.2 (切比雪夫不等式):

设随机变量X具有数学期望$E(X)=\mu$,方差$Var(X)=\sigma^2$, 则对于任意$\varepsilon>0$, 都有:

$$P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2};$$

定理的**等价形式**为: $P\{|X-\mu|<\varepsilon\}\geq 1-\dfrac{\sigma^2}{\varepsilon^2}.$



证明: 在定理5.1.1中取$Y=X-\mu$

且$k=2$ 就可把以下不等式:

$$P\{|Y|\geq\varepsilon\}\leq\frac{E(|Y|^k)}{\varepsilon^k}$$ 变为

$$P\{|X-\mu|\geq\varepsilon\}\leq\frac{E(|X-\mu|^2)}{\varepsilon^2}$$

切比雪夫不等式说明

对于任意分布的X，若记　$E(X)=\mu$,　$Var(X)=\sigma^2$,

则由切比雪夫不等式：

$$P(|X-\mu|<3\sigma)\ \ge 1-\frac{\sigma^2}{(3\sigma)^2}=0.8889$$

即对于任意分布的X落入区间（$\mu-3\sigma$，$\mu+3\sigma$）的概率均大于0.8889。

检验：当$X\sim N(\mu,\sigma^2)$时，则

$$P(|X-\mu|<3\sigma)=P(\mu-3\sigma<X<\mu+3\sigma)=2\Phi(3)-1=0.9974$$

可见 $0.9974>0.8889$，符合以上结论!

但要注意，虽然切比雪夫不等式可以对任意分布的随机变量落入其期望附近的对称区间概率进行估计，但只是粗略估计!

*例2：某天文机构想测量宇宙中两颗行星的距离，进行了n次独立的观测，测量值分别为X_i（光年）$, i=1,2,\cdots,n$.若$E(X_i)=\mu$(为两颗行星的真实距离, 未知)，$Var(X_i)=5.$

现取这n次观测的平均作为真实距离μ的估计.

(1) 若$n=100$,那么估计值与真实值之间的误差在± 0.5光年之内的概率至少有多大？

(2) 若要以不低于95%的把握控制估计值与真实值之间的误差在± 0.5光年之内，至少要观测多少次？

解:由于对$i=1,2,\cdots,n$，都有

$E(X_i)=\mu, Var(X_i)=5,$且X_i相互独立，故

$$E(\frac{1}{n}\sum_{i=1}^{n}X_i)=\mu,\quad Var(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}\sum_{i=1}^{n}Var(X_i)=\frac{5}{n}.$$

（1）当$n=100,$时，由切比雪夫不等式知

$$P\{|\frac{1}{100}\sum_{i=1}^{100}X_i-\mu|<0.5\}\geq 1-\frac{5/n}{0.5^2}=1-\frac{5/100}{0.5^2}=0.8;$$

(2)同样利用切比雪夫不等式，要使得

$$P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|<0.5\}\geq 1-\frac{5/n}{0.5^2}\geq 0.95,$$ n需满足$n\geq 400.$

例3：n重贝努里试验中，已知每次试验事件A出现的概率为0.75，试利用切比雪夫不等式,(1)若n=7500,估计A出现的频率在0.74至0.76之间的概率至少有多大；（2）估计n,使A出现的频率在0.74至0.76之间的概率不小于0.90。

解：设在n重贝努里试验中，事件A出现的次数为X，

则$X \sim B(n,0.75), E(X)=np=0.75n, Var(X)=npq=0.1875n,$

又A事件的频率为：$f_n(A)=\dfrac{X}{n}$

$$(1)\ n=7500,\ P\left\{0.74<\frac{X}{n}<0.76\right\}=P\left\{|X-0.75n|<0.01n\right\}$$

$$\geq 1-\frac{0.1875n}{(0.01n)^2} \quad =0.75$$

$$(2)\ P\left\{0.74<\frac{X}{n}<0.76\right\} =P\left\{|X-0.75n|<0.01n\right\}$$

$$\geq 1-\frac{0.1875n}{(0.01n)^2}=1-\frac{1875}{n}\geq 0.90 \qquad \Rightarrow n\geq 18750$$

大数定律

定义5.1.2： 设$Y_1,\dots,Y_n,\dots$为一个随机变量序列，

若存在常数序列$\{c_n,n\geq 1\}$，使得对$\forall\varepsilon>0$,均有：

$$\lim_{n\to+\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}Y_i-c_n\right|\geq\varepsilon\right\}=0,\text{或}\lim_{n\to+\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}Y_i-c_n\right|<\varepsilon\right\}=1$$

成立，即有当$n\to+\infty$， $\frac{1}{n}\sum_{i=1}^{n}Y_i\xrightarrow{P}c_n$

则称随机变量序列$\{Y_i,i\geq 1\}$服从(弱)大数定律.

说明：若令$Z_n=\frac{1}{n}\sum_{i=1}^{n}Y_i$，当$n\to\infty$, Z_n依概率收敛于c.

随机变量序列前n个变量的算术平均依概率收敛于c,

则这个随机变量序列服从大数定律。

定理5.1.3 (贝努里大数定律)

设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，则对$\forall \varepsilon > 0$,有：

$$\lim_{n\to+\infty} P\left\{\left|\frac{n_A}{n}-p\right|\geq\varepsilon\right\}=0 \qquad (n_A\text{为}n\text{个}0-1\text{分布变量之和})$$

证明：$\because n_A \sim B(n,p)$ $\quad E\left(\frac{n_A}{n}\right)=\frac{1}{n}E(n_A)=\frac{1}{n}\cdot np=p,$

$$Var\left(\frac{n_A}{n}\right)=\frac{1}{n^2}Var(n_A)=\frac{1}{n^2}\cdot npq=\frac{pq}{n}$$

由切比雪夫不等式：$P\left\{\left|\frac{n_A}{n}-p\right|\geq\varepsilon\right\}\leq\frac{pq}{n\varepsilon^2}\xrightarrow{n\to+\infty}0$

0-1分布$\{X_i \sim B(1,p), i\geq 1\}$的随机变量序列服从大数定律

贝努里大数定律的重要意义

贝努里大数定律揭示了在大量重复独立试验中事件出现频率的稳定性，正因为这种稳定性，概率的概念才有客观意义，贝努里大数定律还提供了通过试验来确定事件概率的方法，既然频率n_A/n与概率p有较大偏差的可能性很小，我们便可以通过做试验确定某事件发生的频率并把它作为相应的概率估计，这种方法就是第7章将要介绍的参数估计法，参数估计的重要理论基础之一就是大数定理。

例4：某种器件的寿命X(以小时计)具有以下的概率密度：

$$f(x)=\begin{cases}\dfrac{1000}{x^2}, x>1000\\ 0\quad ,其它\end{cases}$$，现有一大批此种器件（设各器件损坏与否相互独立），任取n只，Y_n表示n只中寿命大于3000小时的器件个数，则当$n\to\infty$时，$\dfrac{Y_n}{n}\xrightarrow{P}?$

解：设事件$A="X>3000"$

则$p=P(A)=P(X>3000)=\int_{3000}^{\infty}\dfrac{1000}{x^2}dx=\dfrac{1}{3}$

在n次试验中A发生的频率为$f_n(A)=\dfrac{Y_n}{n}$。

由大数定律：$f_n(A)\xrightarrow{P}P(A)$　　得：$\dfrac{Y_n}{n}\xrightarrow{P}\dfrac{1}{3}$.

定理5.1.4(辛钦大数定律)

设$\{X_i, i \geq 1\}$为独立同分布的随机变量序列，且其期望存在，记为μ，则对$\forall \varepsilon > 0$，有：

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_i - \mu\right| \geq \varepsilon\right\} = 0,$$

即随机变量序列$\{X_i, i \geq 1\}$服从大数定律，

也即，当$n \to +\infty$时，$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$

说明：当独立的随机变量服从相同分布时，就不要求随机变量的方差存在这一要求了。

推论5.1.1:

设$\{X_i, i \geq 1\}$为独立同分布的随机变量序列，若$h(x)$为连续函数, 且$E|h(X_1)| < +\infty$, 则对$\forall \varepsilon > 0$,

有: $$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} h(X_i) - E(h(X_1))\right| \geq \varepsilon\right\} = 0,$$

即随机变量$\{h(X_i), i \geq 1\}$也服从大数定律.

即 $$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow{P} E(h(X_1))$$

可由依概率收敛的性质推出。

若$X_1, X_2, \cdots, X_n$为独立同分布的，$h(x)$为连续函数，则$h(X_1), h(X_2), \cdots, h(X_n)$也为独立同分布的。

例5： 设随机变量$X_1,\cdots,X_n,\cdots,$相互独立同分布, $X_1 \sim U(-1,1)$. 则

(1)$\frac{1}{n}\sum_{k=1}^{n} X_k$, (2)$\frac{1}{n}\sum_{k=1}^{n}|X_k|$, (3)$\frac{1}{n}\sum_{k=1}^{n} X_k^2$分别依概率收敛吗？

如果依概率收敛，分别收敛于什么？

解：对照辛钦大数定律，$X_1,\cdots,X_n,$相互独立同分布，$E(X_1)$存在,

$|X_1|,\cdots,|X_n|,$相互独立同分布，$E(|X_1|)$存在,

$X_1^2,\cdots,X_n^2,$相互独立同分布，$E(X_1^2)$存在,

故它们各前n个算术平均：$\frac{1}{n}\sum_{k=1}^{n} X_k$, $\frac{1}{n}\sum_{k=1}^{n}|X_k|$, $\frac{1}{n}\sum_{k=1}^{n} X_k^2$均依概率收敛。

因为，$E(X_1)=0,$ $\Rightarrow \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} 0,$

同理，$E(|X_1|)=\int_{-1}^{1}|x|\frac{1}{2}dx=\frac{1}{2},$ $\Rightarrow \frac{1}{n}\sum_{k=1}^{n}|X_k| \xrightarrow{P} \frac{1}{2},$

$E(X_1^2)=\int_{-1}^{1}x^2\frac{1}{2}dx=\frac{1}{3},$ $\Rightarrow \frac{1}{n}\sum_{k=1}^{n} X_k^2 \xrightarrow{P} \frac{1}{3}$。

例6：设 $X \sim f(x)=\begin{cases}3x^2, 0<x<1\\ 0, \quad 其他\end{cases}$， 现对$X$独立观察$n$次，

观察值记为$X_1, X_2, \cdots, X_n$， $\forall\varepsilon>0$,如果这些观察值满足

$$\lim_{n\to\infty} P\left(\left|\frac{X_1^2+X_2^2+\cdots+X_n^2}{n}-a\right|<\varepsilon\right)=1,$$ 求a=？

解：由题意知，$X_1, X_2, \cdots, X_n$是具有独立同分布的随机变量，

所以，它们的连续函数$X_1^2, X_2^2, \cdots, X_n^2$也是独立同分布的。

$\dfrac{X_1^2+X_1^2+\cdots+X_n^2}{n}$是变量序列$X_1^2$，$X_1^2$,…，$X_n^2$的前$n$个算术平均，

故由辛钦定理得：算术平均依概率收敛于$E(X^2)$

$$\therefore\ a=E(X^2)=\int_0^1 x^2 3x^2 dx=\frac{3}{5}$$

例7：设 $X \sim f(x)=\begin{cases}3x^2, 0<x<1\\ 0, \quad 其他\end{cases}$，现对$X$独立观察$n$次，

观察值记为$X_1, X_2, \cdots, X_n$，问$\dfrac{X_1+X_2+\cdots+X_n}{X_1^2+X_2^2+\cdots+X_n^2}$依概率

收敛吗？若收敛，求其极限值。

解：由题意知，$X_1, X_2, \cdots, X_n$是具有独立同分布的随机变量,

所以，它们的连续函数$X_1^2, X_2^2, \cdots, X_n^2$也是独立同分布的。

$$\frac{X_1+X_2+\cdots+X_n}{X_1^2+X_1^2+\cdots+X_n^2}=\frac{(X_1+X_2+\cdots+X_n)/n}{(X_1^2+X_1^2+\cdots+X_n^2)/n}$$

$$\xrightarrow{P}\frac{E(X)}{E(X^2)}=\frac{\int_0^1 x3x^2dx}{\int_0^1 x^2 3x^2dx}=\frac{3/4}{3/5}=\frac{5}{4}$$

例8：设随机变量$X_1,\cdots,X_n,\cdots$,相互独立同分布，$X_1\sim U(0,1)$,则$\sqrt[n]{X_1X_2\dots X_n}$依概率收敛吗？如果依概率收敛，收敛于什么？

解：令$Y_n=\sqrt[n]{X_1\dots X_n}$,则$Z_n=\ln Y_n=\frac{1}{n}(\ln X_1+\dots+\ln X_n)$

而$\ln X_1,\cdots,\ln X_n,\cdots$是相互独立同分布的，并且

$$E(\ln X_1)=\int_{-\infty}^{\infty}\ln x\cdot f(x)dx=\int_0^1\ln xdx$$
$$=(x\ln x-x)_0^1=-1$$

由辛钦大数定律，$Z_n\xrightarrow{P}-1.$

所以Z_n的连续函数$Y_n=e^{Z_n}\xrightarrow{P}e^{-1}$

*例 设随机变量$X_1,\cdots,X_n,\cdots$,相互独立，且它们的分布律为

$$P\{X_i=0\}=1-\frac{1}{i},P\{X_i=\sqrt{i}\}=P\{X_i=-\sqrt{i}\}=\frac{1}{2i},i=1,2,\cdots.$$

试判断$\{X_i,i\geq 1\}$是否服从大数定律？

解：由于对任意的$i\geq 1$, 有 $E(X_i)=0+\sqrt{i}\cdot\frac{1}{2i}-\sqrt{i}\cdot\frac{1}{2i}=0,$

$$Var(X_i)=E(X_i^2)=0+(\sqrt{i})^2\cdot\frac{1}{2i}+(\sqrt{i})^2\cdot\frac{1}{2i}=1,$$

所以$\{X_i,i\geq 1\}$相互独立，方差相同，由推论知满足大数定律，且

$$\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}\frac{1}{n}\sum_{i=1}^{n}E(X_i)=0,\text{ 当}n\to+\infty.$$

*例 设随机变量$X_1,\cdots,X_n,\cdots$,相互独立，且它们的分布律为

$$P\{X_i=0\}=1-\frac{1}{i},P\{X_i=\sqrt{i}\}=P\{X_i=-\sqrt{i}\}=\frac{1}{2i},i=1,2,\cdots.$$

试判断$\{X_i,i\geq 1\}$是否服从大数定律？

用切比雪夫不等式解：$E(X_i)=0,\ Var(X_i)=E(X_i^2)=1$

$$E(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=0,\quad Var(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}\sum_{i=1}^{n}Var(X_i)=\frac{1}{n}$$

$$P(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-E(\frac{1}{n}\sum_{i=1}^{n}X_i)\right|\geq\varepsilon)\leq\frac{Var(\frac{1}{n}\sum_{i=1}^{n}X_i)}{\varepsilon^2}=\frac{1}{n\varepsilon^2}\xrightarrow{n\to+\infty}0$$

$\therefore$当$n\to+\infty$, $\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}\frac{1}{n}\sum_{i=1}^{n}E(X_i)=0.$

§2 中心极限定理

- 背景：

　　某些指标(随机变量)是由大量的相互独立因素的综合影响所形成的，而其中每个因素作用都很小，则这种指标(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。

定理5.2.1 (独立同分布的中心极限定理)

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

$E(X_i) = \mu,\ Var(X_i) = \sigma^2, i = 1, 2, \cdots, \forall x \in R$,有：

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

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

此定理表明，当n充分大时，Y_n近似服从$N(0,1)$.

思考题（n足够大）：

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

答：$N(\mu, \frac{\sigma^2}{n})$

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

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

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

推论5.2.1 （德莫弗–拉普拉斯定理）

设n_A为n重贝努里试验中A发生的次数，$P(A)=p(0<p<1)$,则

$$\forall x\in R，\text{有：}\lim_{n\to+\infty}P\left(\frac{n_A-np}{\sqrt{np(1-p)}}\le x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x),$$

即：$\boxed{\text{若}n\text{足够大}, n_A\sim B(n,p),\ \text{则}n_A\overset{\text{近似}}{\sim}N(np,npq)}$

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

证明：令$X_i=\begin{cases}1 & \text{第}i\text{次试验时}A\text{发生}\\ 0 & \text{第}i\text{次试验时}A\text{未发生}\end{cases}$

则 $X_1,X_2,\cdots,X_n,\cdots$相互独立同分布, $X_i\sim B(1,p)$.

由于$n_A=X_1+X_2+\cdots+X_n$,

符合独立同分布的中心极限定理, 所以$n_A\overset{\text{近似}}{\sim}N(np,npq)$

例1：在 (0,10)上任取48个实数，求这些数的和不超过260的概率。

解：设48个实数用$X_1, X_2, \cdots, X_{48}$表示，总和$X = X_1 + X_2 + \cdots + X_{48}$

则 $X_1, X_2, \cdots, X_{48}$是相互独立的，且都服从$U(0,10)$

$$E(X_1) = E(X_2) = \cdots = E(X_{48}) = \frac{0+10}{2} = 5$$

$$Var(X_1) = Var(X_2) = \cdots = Var(X_{48}) = \frac{(10-0)^2}{12}$$

$$E(X) = E(X_1 + X_2 + \cdots + X_{48}) = 48*5 = 240$$

$$Var(X) = Var(X_1 + X_2 + \cdots + X_{48}) = 48*\frac{100}{12} = 400$$

$\therefore$由中心极限定理，X近似服从 $N(240, 400)$

$$P(X \le 260) \approx \Phi\left(\frac{260-240}{20}\right) = \Phi(1) = 0.8413$$

例2：设某种电器元件的寿命服从均值为100小时的指数分布，现随机取得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：记16只电器元件的寿命分别为$X_1, X_2, \cdots, X_{16}$,

则16只电器元件的寿命总和为$X = \sum_{i=1}^{16} X_i$,

由题设$E\left(X_i\right)=100$, $Var\left(X_i\right)=100^2$

$$E(X)=E(\sum_{i=1}^{16} X_i)=\sum_{i=1}^{16} E(X_i)=16*100=1600$$

$$Var(X)=Var(\sum_{i=1}^{16} X_i)=\sum_{i=1}^{16} Var(X_i)=16*100^2=400^2$$

根据独立同分布的中心极限定理： $X \overset{近似}{\sim} N(1600, 400^2)$

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

$$=1-\Phi\left(0.8\right)=0.2119$$

例3：设随机变量$X_1,\cdots,X_{20}$,相互独立同分布，$X_1\sim U(-1,1)$。分别求

$$(1)\frac{1}{20}\sum_{k=1}^{20}X_k,\quad (2)\frac{1}{20}\sum_{k=1}^{20}|X_k|,\quad (3)\frac{1}{20}\sum_{k=1}^{20}X_k^2 \text{的近似分布。}$$

解：由中心极限定理,$\frac{1}{20}\sum_{k=1}^{20}X_k$，$\frac{1}{20}\sum_{k=1}^{20}|X_k|$，$\frac{1}{20}\sum_{k=1}^{20}X_k^2$均近似服从正态分布。

$$E(X_1)=0,\qquad D(X_1)=\frac{4}{12}=\frac{1}{3},\qquad \Rightarrow \frac{1}{20}\sum_{k=1}^{20}X_k \overset{\text{近似}}{\sim} N(0,\frac{1}{60}),$$

$$E(|X_1|)=\int_{-1}^{1}|x|\frac{1}{2}dx=\int_0^1 xdx=\frac{1}{2},\qquad E(X_1^2)=\int_{-1}^{1}x^2\frac{1}{2}dx=\int_0^1 x^2dx=\frac{1}{3},$$

$$E(X_1^4)=\int_{-1}^{1}x^4\frac{1}{2}dx=\int_0^1 x^4dx=\frac{1}{5},$$

$$Var(|X_1|)=E(X_1^2)-[E(|X_1|)]^2=\frac{1}{12},\qquad \Rightarrow \frac{1}{20}\sum_{k=1}^{20}|X_k| \overset{\text{近似}}{\sim} N(\frac{1}{2},\frac{1}{240}),$$

$$Var(X_1^2)=E(X_1^4)-[E(X_1^2)]^2=\frac{1}{5}-\frac{1}{9}=\frac{4}{45},\qquad \Rightarrow \frac{1}{20}\sum_{k=1}^{20}X_k^2 \overset{\text{近似}}{\sim} N(\frac{1}{3},\frac{1}{225})\text{。}$$

例4：设保险公司某项保险业务有5000人参加，投保人交16元，若符合赔付条件，保险公司付给投保人2000元。设赔付率为0.005，试求保险公司在这项保险业务中盈利2万到4万元之间的概率.

解：设X为一年内符合赔付条件的人数，则$X \sim B(5000, 0.005)$

$np = 25, npq \approx 25$，由中心极限定理得 $X \overset{\text{近似}}{\sim} N(25, 25)$

$$P(20000 < 5000*16 - 2000X < 40000) = P(20 < X < 30)$$

$$\approx \Phi\left(\frac{30-25}{\sqrt{25}}\right) - \Phi\left(\frac{20-25}{\sqrt{25}}\right) \qquad C_{5000}^{21} 0.005^{21} 0.995^{4979} + \cdots$$

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

例4:设保险公司某项保险业务有5000人参加，投保人交16元,若符合赔付条件，保险公司付给投保人2000元。设赔付率为0.005，试求保险公司在这项保险业务中盈利2万到4万元之间的概率.

另解:设X为公司在该业务中的总利润,

X_i为公司在第i人业务中的所获得的利润。

则：$X=\sum_{i=1}^{5000} X_i \overset{\text{近似}}{\sim} N(\mu,\sigma^2)$

X_i	16	-1984
P_k	0.995	0.005

$$E(X_i)=6, Var(X_i)=19900, E(X)=30000, Var(X)=9975^2$$

$$\therefore X \overset{\text{近似}}{\sim} N(30000, 9975^2)$$

$$P(20000 < X < 40000) \approx 2\Phi(1)-1=0.6826$$

例5：某种器件的寿命X(以小时计)具有以下的概率密度：

$f(x)=\begin{cases}\dfrac{1000}{x^2}, x>1000\\ 0\quad ，其它\end{cases}$，现有一大批此种器件（设各器件损坏与否相互独立），任取n只，Y_n表示n只中寿命大于3000小时的器件个数，则$P(Y_{50}\le 20)\approx ?$

可知：$Y_n \sim B(n,p)$, 其中 $p=P(X>3000)=1/3.$

由中心极限定理得：$Y_n \overset{\text{近似}}{\sim} N(np,npq)$

即$Y_{50} \overset{\text{近似}}{\sim} N(50/3,100/9)$

$$P(Y_{50}\le 20)\approx \Phi\left(\frac{20-np}{\sqrt{npq}}\right)=\Phi(1)=0.8413$$

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

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

2. 用泊松分布近似计算（泊松定理）

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

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

3. 用正态分布近似计算

$$np=8，\quad npq=400\times 0.02\times 0.98=7.84 \qquad \therefore\quad X\overset{\text{近似}}{\sim} N(8,7.84)$$

$$P(X\ge 2)=1-P(X\le 1)\approx 1-\Phi\left(\frac{1-8}{\sqrt{7.84}}\right)=\Phi\left(\frac{7}{2.8}\right)=0.9938$$

$$P(X\ge 2)=1-P(X<2)\approx 1-\Phi\left(\frac{2-8}{\sqrt{7.84}}\right)=\Phi\left(\frac{6}{2.8}\right)=0.9838$$

例7:某校1500名学生选修“概率论与数理统计”课程,共有10名教师主讲此课,假定每位学生可以随意地选择一位教师(即,选择任意一位教师的可能性均为1/10),而且学生之间的选择是相互独立的.问:每位教师的上课教室应该设有多少座位才能保证该班因没有座位而使学生离开的概率小于5%.

解：由于每位学生可以随意地选择一位老师，因此我们只需要考虑某个老师甲的上课教室的座位即可.

引入随机变量

$$X_i=\begin{cases}1, & \text{第}i\text{个学生选择教师甲}\\ 0, & \text{其它}\end{cases},\ i=1,2,\cdots,1500.$$

则各$X_1,X_2,\cdots,X_{1500}$是独立同分布的，$X_i\sim B(1,0.1)$

记教师甲的课堂学生数$Y=\sum_{i=1}^{1500}X_i$,则$Y\sim B(1500,0.1)$

由德莫弗--拉普拉斯中心极限定理，知

$Y\overset{\text{近似}}{\sim}N(1500*0.1\ ，\ 1500*0.1*0.9)$ ，即

$Y\overset{\text{近似}}{\sim}N(150\ ，135)$

设教室需要设a个座位，由题意知a需要满足

$$P\{Y > a\} = 1 - P\{Y \le a\} \approx 1 - \Phi\left(\frac{a-150}{\sqrt{135}}\right) < 5\%$$

即 $\Phi\left(\frac{a-150}{\sqrt{135}}\right) > 95\%$

查表得，$\Phi(1.645) = 95\%$, 故需 $\frac{a-150}{\sqrt{135}} > 1.645$,

解得 $a > 169.11.$

故每位老师的上课教室应该至少设170个座位才能保证因没有座位而使得学生离开的概率小于5%.

例8:在n重贝努里试验中，若已知每次试验事件A出现的概率为0.75，试利用中心极限定理,(1)若n=7500,估计A出现的频率在0.74至0.76之间的概率近似值；（2）估计n,使A出现的频率在0.74至0.76之间的概率不小于0.90。

解：设在n重贝努里试验中，事件A出现的次数为X，

$$\text{则} X \sim b(n, 0.75), E(X) = np = 0.75n, Var(X) = npq = 0.1875n,$$

$$(1)\ n = 7500,\ P\left\{0.74 < \frac{X}{n} < 0.76\right\} \approx \Phi\left(\frac{0.76n - 0.75n}{\sqrt{0.1875n}}\right) - \Phi$$

$$\approx 2\Phi\left(\frac{0.04\sqrt{n}}{\sqrt{3}}\right) - 1 = 2\Phi(2) - 1 = 0.9544$$

切比雪夫

不等式估计

$P(.) \geq 0.75$。

$$(2)\ P\left\{0.74 < \frac{X}{n} < 0.76\right\} \approx \Phi\left(\frac{0.76n - 0.75n}{\sqrt{0.1875n}}\right) - \Phi\left(\frac{0.74n - 0.75n}{\sqrt{0.1875n}}\right)$$

$$\approx 2\Phi\left(\frac{0.04\sqrt{n}}{\sqrt{3}}\right) - 1 \geq 0.9, \quad \Rightarrow \Phi\left(\frac{0.04\sqrt{n}}{\sqrt{3}}\right) \geq 0.95,$$

$$\frac{0.04\sqrt{n}}{\sqrt{3}} \geq 1.645, \quad n \geq (25 \times 1.645)^2 \times 3 = 5074$$

切比雪夫

不等式估计

$n \geq 18750$。

复习思考题 5

1. 依概率收敛与高等数学的收敛含义有何区别?

2. 马尔可夫不等式与切比雪夫不等式分别适用于哪些随机变量?

3. 大数定律与中心极限定理的联系与区别。

4. 什么情况下适合用中心极限定理?