LLN(大数定律)和CLT(中心极限定理):

随机变量序列$\{X_n\}$, $n= 1,2,\ldots$;

随机变量序列$\{X^2_n\}$, $n= 1,2,\ldots$;

随机变量序列$\{X_n - E(X_n)\}$, $n= 1,2,\ldots$

1、(Chebyshev)不等式

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

2、依概率收敛: $X_n \xrightarrow{P} X$: $\lim\limits_{n\to\infty} P\{|X_n - X| < \varepsilon\} = 1$

3、依分布收敛: $X_n \xrightarrow{L} X$: $\lim\limits_{n\to\infty} F_n(x) = F(x)$

$\overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$

4、大数定律: $\lim\limits_{n\to\infty} P\{|\frac{1}{n}\sum\limits_{i=1}^{n} X_i - \frac{1}{n}\sum\limits_{i=1}^{n} E(X_i)| < \varepsilon\} = 1$

i.e., $\overline{X}_n - E(\overline{X}_n) \xrightarrow{p} 0$

$E(X_n)$都存在

1) 切比雪夫大数定律 ⇒ 2) 独立同分布大数定律

$D(X_i)<C$, $i=1,2,3,\ldots\ldots$

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

3) 贝努里(Bernulli)大数定律

$$X_i = \begin{cases} 1, & \text{第}i\text{次}A\text{发生;} \\ 0, & \text{第}i\text{次}A\text{不发生.} \end{cases}$$

令$\sum\limits_{i=1}^{n} X_i = m$

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

重要结论：小概率事件原理

大数定律是由“频率依概率收敛于概率值”引申而来的.

大数定律的应用:

实际工作中, 以大量测量值的平均值作为精确值的估计值 → 以独立同分布大数定律 为理论依据

频率的稳定性：试验次数充分多，则A发生的频率 $\frac{m}{n}$ 趋于A发生的概率p → 以贝努里大数定律为理论依据

小概率事件原理: 概率很小的事件, 在一次试验中几乎不可能发生, 实际中看作不可能事件→ 以贝努里大数定律 为理论依据

三倍标准差原理(3σ原理)：若 $X\sim N(\mu,\sigma^2)$，则当X取值在$(\mu-3\sigma,\mu+3\sigma)$之外时，看作小概率事件 →

$$P\{|X-\mu|>3\sigma\}=0.0026$$

大数定律是由“频率依概率收敛于概率值”引申而来的.

大数定律：（“频率依概率收敛于概率值”引申而来）

只是对随机变量序列前**n**项算术平均值与它的期望值之差依概率收敛于**0**，给出了概率趋势定性的说明：

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

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n}X_i$$

$$\overline{X}_n - E(\overline{X}_n) \xrightarrow{p} 0$$

切比雪夫不等式只能对概率作出很粗略的估计。

$$P\{|\frac{1}{n}\sum_{i=1}^{n}X_i - \frac{1}{n}\sum_{i=1}^{n}E(X_i)|<\varepsilon\} \ge 1 - \frac{D(\sum_{i=1}^{n}X_i)}{\varepsilon^2}$$

需要较精确的概率值 $P\{\underset{-\varepsilon}{a} < \frac{1}{n}\sum_{i=1}^{n}X_i - \frac{1}{n}\sum_{i=1}^{n}E(X_i) < \underset{\varepsilon}{b}\}$

中心极限定理给出概率近似计算公式(定量的说明)

重要结论:

1. 随机变量X的$E(X),D(X)$存在, 且$D(X)>0$,则$X^*=\dfrac{X-E(X)}{\sqrt{D(X)}}$

$$E(X^*)=0,\ D(X^*)=1$$

2. 设随机变量$X_1,X_2,\ldots,X_n$相互独立, 且$E(X_i)=a,D(X_i)=b,$

则:$E(\bar{X})=E\left(\dfrac{1}{n}(X_1+X_2+\ldots+X_n)\right)=a,D(\bar{X})=\dfrac{1}{n}b.$

3. 设随机变量$X_1,X_2,\ldots,X_n$,相互独立且同分布,均值为a和方差为b,

则:$E(\bar{X})=E\left(\dfrac{1}{n}(X_1+X_2+\ldots+X_n)\right)=a,D(\bar{X})=\dfrac{1}{n}b$

三. 常 见 的 中 心 极 限 定 理(两个)

1.中心极限定理: 前n项和$\sum_{k=1}^{n} X_k$的标准化随机变量序列

以下$X_1, \dots, X_k, \dots$相互独立, 且$E(X_k)$和$D(X_k)$都存在

$$Z_n=\frac{\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}E(X_i)}{\sqrt{\sum_{i=1}^{n}D(X_i)}},\ \text{从而}\ \lim_{n\to\infty}F_{Z_n}(x)=\lim_{n\to\infty}P\{\frac{\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}E(X_i)}{\sqrt{\sum_{i=1}^{n}D(X_i)}}\le x\}=\Phi(x)$$

•独立同分布中心极限定理:

X_k有相同分布,

$E(X_k)=\mu,$

$D(X_k)=\sigma^2, k=1,2\dots$

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

n充分大

$$\Rightarrow \sum_{k=1}^{n}X_k\sim N\left(n\mu,n\sigma^2\right)\qquad \overline{X}_n=\frac{\sum_{k=1}^{n}X_k}{n}\sim N\left(\mu,\frac{1}{n}\sigma^2\right)$$

"随机变量和的极限分布是正态分布"

独立同分布中心极限定理的应用

1）求随机变量之和 $S_n = \sum_{i=1}^{n} X_i$ 取值的概率.

产 品 检 验

2）已知 $S_n = \sum_{i=1}^{n} X_i$ 取值的概率，反求n.

产 品 测 重(略)

3）近似计算与用频率估计概率的有关问题.

重复试验次数估计

注: 切比雪夫不等式只能对概率作出很粗略的估计.

例1: 检验员逐个地检查某种产品, 每次花10 秒检查一个, 但也可能有的产品需要重复检查一次再用去10秒, 假设每个产品需要重复检查的概率为0. 5, 求在8小时内检查员检查的产品多于1900个的概率.

如何用独立同分布中心极限定理估计概率?

方法: 首先将复杂的随机变量分解成独立同分布的随机变量之和;
然后把变量的和看作正态分布计算出期望和方差;
最后当作正态分布求在某区间上的概率.

分析: 问题等价于求检查员检查1900个产品所花的总时间 X 不超过8h的概率 $P\{X \le 8*3600\}$.

例1: 检验员逐个地检查某种产品, 每次花10 秒检查一个, 但也可能有的产品需要重复检查一次再用去10秒, 假设每个产品需要重复检查的概率为0. 5, 求在8小时内检查员检查的产品多于1900个的概率.

解: 设检查第i个产品所花时间为X_i $(i = 1,2,3,\ldots,1900)$

则检查1900个产品所花的总时间为: $X = \sum_{i=1}^{1900} X_i$

$$X_i = \begin{cases} 10 & \text{第 } i \text{ 个产品没有重复检查;} \\ 20 & \text{第 } i \text{ 个产品需要重复检查一次;} \end{cases}$$

同时 $P(X_i = 10) = P(X_i = 20) = 0.5$, $(i = 1,2,3,\ldots,1900)$

即X_i相互独立都服从同一分布.

例1: 检验员逐个地检查某种产品, 每次花10 秒检查一个, 但也可能有的产品需要重复检查一次再用去10秒,假设每个产品需要重复检查的概率为0. 5, 求在8小时内检查员检查的产品多于1900个的概率.

$$X_i = \begin{cases} 10 & \text{第 } i \text{ 个产品没有重复检查;} \\ 20 & \text{第 } i \text{ 个产品需要重复检查一次;} \end{cases} \quad X = \sum_{i=1}^{1900} X_i$$

$$E(X_i) = 10 \times 0.5 + 20 \times 0.5 = 15$$

$$D(X_i) = E(X_i^2) - [E(X_i)]^2 = 25$$

$$i = 1,2,\ldots,1900$$

由独立同分布中心极限定理，当n充分大时,

$$X = \sum_{i=1}^{1900} X_i \overset{\cdot}{\sim} N(1900 \times 15 , 1900 \times 25)$$

近似成立.

方法：首先将复杂的随机变量分解成独立同分布的随机变量之和；

然后把变量的和看作正态分布计算出期望和方差；

最后当作正态分布求在某区间上的概率.

由独立同分布中心极限定理, 当n充分大时,

$$X=\sum_{i=1}^{1900}X_i \overset{\cdot}{\sim} N(1900\times 15, 1900\times 25)$$ 近似成立.

所求概率为: $P(X\le 8\times 3600)$

$$\approx \Phi\left(\frac{8*3600-1900*15}{\sqrt{1900*25}}\right)$$

$$=\Phi\left(\frac{6}{\sqrt{19}}\right)\approx 0.9162$$

例2: 将一枚均匀硬币连续抛 n 次, 试用中心极定理来估计n, 使下式成立. $P\{| f_n(A) - P(A) | < 0.01\} \geq 0.99$

其中 A ={ 出现正面 }.

解: $P(A)$=1/2, 令 $X_i = \begin{cases} 1, & 第i次出现正面; \\ 0, & 其它. \end{cases} \quad i = 1, 2, \cdots n$

则随机变量序列$\{X_i\}, i = 1,2,\ldots$是相互独立且同分布

的, 且有 $$f_n(A) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$E(X_i) = \frac{1}{2}, \quad D(X_i) = \frac{1}{4}, \quad i = 1,2,\cdots$$

所以随机变量序列$\{X_i\}$, 满足独立同分布中心极限定理.

例2: 将一枚均匀硬币连续抛 n 次, 试用中心极定理来估计n, 使下式成立. $P\{|f_n(A)-P(A)|<0.01\}\geq 0.99$

其中 $A=\{$ 出现正面 $\}$.

$$f_n(A)=\frac{1}{n}\sum_{i=1}^{n}X_i \quad E(\sum_{i=1}^{n}X_i)=\frac{n}{2} \quad D(\sum_{i=1}^{n}X_i)=\frac{1}{4}n$$

解: $P\{|f_n(A)-P(A)|<0.01\}$

$$=P\{-0.01<\frac{\sum_{i=1}^{n}X_i}{n}-\frac{1}{2}<0.01\}=P\{-0.01n+\frac{n}{2}<\sum_{i=1}^{n}X_i<\frac{n}{2}+0.01n\}$$

$$=P\{-\frac{0.01n}{1/2\sqrt{n}}<\frac{\sum_{i=1}^{n}X_i-\frac{n}{2}}{\underline{1/2\sqrt{n}}}<\frac{0.01n}{1/2\sqrt{n}}\}$$

$\dfrac{\sum_{i=1}^{n}X_i-\frac{n}{2}}{1/2\sqrt{n}}$ 可由 $N(0,1)$来近似.

例2: 将一枚均匀硬币连续抛 n 次, 试用中心极定理来估计n, 使下式成立. $P\{|f_n(A)-P(A)|<0.01\}\geq 0.99$

其中 $A=\{$ 出现正面 $\}$.

随机变量序列$\{X_i\}$, 满足独立同分布中心极限定理:

$$0.99\leq P\{-\frac{0.01n}{1/2\sqrt{n}}<\frac{\sum_{i=1}^{n}X_i-\frac{n}{2}}{1/2\sqrt{n}}<\frac{0.01n}{1/2\sqrt{n}}\}$$

$$\approx 2\Phi(0.02\sqrt{n})-1$$

$Y\sim N(0,1)$:

$P\{-a<Y<a\}=2\Phi(a)-1$

$$\Rightarrow \Phi(0.02\sqrt{n})\geq 0.995$$

因$\Phi(x)$单调不降 $\Rightarrow 0.02\sqrt{n}\geq 2.58$

解得 $n\geq 16{,}641$ (次).

注: 由切比雪夫不等式估计需至少250,000次.

(两个中心极限定理)以下 X_k ,…相互独立, 且$E(X_k)$和$D(X_k)$ 都存在.

1. 中心极限定理: 前n项和 $\sum_{k=1}^{n} X_k$ 的标准化随机变量序列

$$Z_n=\frac{\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}E(X_i)}{\sqrt{\sum_{i=1}^{n}D(X_i)}},\text{从而 } \lim_{n\to\infty}F_{Z_n}(x)=\lim_{n\to\infty}P\left\{\frac{\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}E(X_i)}{\sqrt{\sum_{i=1}^{n}D(X_i)}}\le x\right\}=\Phi(x)$$

X_k有相同分布,
$E(X_k)=\mu,$
$D(X_k)=\sigma^2, k=1,2\ldots$

独立同分布中心极限定理:

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

2. 棣莫佛(De Moivre)-拉普拉斯中心极限定理:

$$Y_n=\sum_{i=1}^{n}X_i,$$

$$X_i\sim B(1,p)$$

$$Y_n\sim B(n,p), n=1,2,\ldots$$

$$n\mu=np,\quad n\sigma^2=np(1-p)$$

$$\lim_{n\to\infty}P\left\{\frac{Y_n-np}{\sqrt{np(1-p)}}\le x\right\}=\Phi(x)$$

中心极限定理：“随机变量和的极限分布是正态分布”

$$Z_n=\frac{\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}E(X_i)}{\sqrt{\sum_{i=1}^{n}D(X_i)}},\quad \lim_{n\to\infty}F_{Z_n}(x)=\lim_{n\to\infty}P\{\frac{\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}E(X_i)}{\sqrt{\sum_{i=1}^{n}D(X_i)}}\le x\}=\Phi(x)$$

独立同分布中心极限定理

棣莫佛-拉普拉斯中心极限定理

$Y_n \sim B(n,p), n=1,2,\ldots$

$$\lim_{n\to\infty}P\{\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\le x\}=\Phi(x)\qquad \lim_{n\to\infty}P\{\frac{Y_n-np}{\sqrt{np(1-p)}}\le x\}=\Phi(x)$$

当n足够大时，$\sum_{k=1}^{n}X_k\sim N\left(n\mu,n\sigma^2\right)$近似成立.

$Y_n\sim N(np,np(1-p))$近似成立.

$$\overline{X}_n=\frac{\sum_{k=1}^{n}X_k}{n}\sim N\left(\mu,\frac{1}{n}\sigma^2\right)$$

在n充分大情况下二项分布概率的近似计算公式.

若$X \sim B(n, p)$, 对于足够大的n, 有

$$P\{m_1 < X \le m_2\} = \sum_{m_1+1}^{m_2} C_n^k p^k (1-p)^{n-k}$$

$$= P\left\{\frac{m_1 - np}{\sqrt{np(1-p)}} < \frac{X - np}{\sqrt{np(1-p)}} \le \frac{m_2 - np}{\sqrt{np(1-p)}}\right\}$$

$$\approx \Phi\left(\frac{m_2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{np(1-p)}}\right)$$

注: 与泊松逼近定理区别.

棣莫佛—拉普拉斯中心极限定理的应用

1) 计算在*n*充分大情况下二项分布概率的近似值;

航船的稳定性

2) 已知在*n*充分大情况下二项分布在某范围内取值的概率, 求该范围.

方法： 这实际是1)的反问题.

参见教材例题5.3.3

例3：一船舶在某海区航行，已知每遭受一次波浪的冲击，纵摇角大于3°的概率为 $p = 1/3$，若船舶遭受了90000次波浪冲击，问其中有 29500 ~ 30500 次纵摇角大于3°的概率是多少？

解：假定船舶遭受波浪的各次冲击是独立的.

记X为90000次冲击下纵摇角大于3°的次数, 故有

$$X \sim B(90000, \frac{1}{3}), \qquad \underline{n = 90000,\ p = \frac{1}{3}}$$

所求事件的概率为：$P(29500 \le X \le 30500)$

$$= \sum_{29500}^{30500} C_n^k p^k (1-p)^{n-k}$$

由拉普拉斯中心极限定理：$X \sim B(90000, \frac{1}{3}), n = 90000, p = \frac{1}{3}$

$$P(29500 \le X \le 30500)$$

$$= P\left\{\frac{29500 - np}{\sqrt{np(1-p)}} \le \frac{X - np}{\sqrt{np(1-p)}} \le \frac{30500 - np}{\sqrt{np(1-p)}}\right\}$$

$$\approx \Phi\left\{\frac{30500 - np}{\sqrt{np(1-p)}}\right\} - \Phi\left\{\frac{29500 - np}{\sqrt{np(1-p)}}\right\}$$

$$= \Phi\left\{\frac{5\sqrt{2}}{2}\right\} - \Phi\left\{-\frac{5\sqrt{2}}{2}\right\}$$

$$= 2\Phi\left\{\frac{5\sqrt{2}}{2}\right\} - 1 = 0.995$$

n充分大，

由Laplace的CLT

$$\frac{X - np}{\sqrt{np(1-p)}}$$

近似服从$N(0,1)$

例4(机票超售): 航空公司电脑订座系统的普遍采用给旅客和公司都带来极大的方便，但是也对管理工作提出更高的要求, 例如一架200座的飞机应出售多少座位?

每个订座旅客当作一次试验，则他或到时不登机，记为**A**，或到时登机, 记为 $\bar{A}$, **P(A)**按过去统计资料取为**5%**。

主要的难点在于能否把旅客是否登机看作是独立的，显然对于购买团体票的旅客作此假定是不合适的，此外，大型的交通堵塞等偶然事件也会使这个假定偏离，不过在一般场合作此假定还是合适的. 全体订座旅客数n作为试验总数, 这便构成伯努利概型.

假定超售m个座位, 则共售出200+m个座位,这时要求登机的旅客数X服从二项分布B(200+m, 0.95), 发生拒登机的可能性：$P=P\{X>200\}$

N	201	202	203	204	205	206	207
P	0.000	0.002	0.007	0.015	0.032	0.062	0.109

例4(机票超售): 航空公司电脑订座系统的普遍采用给旅客和公司都带来极大的方便，但是也对管理工作提出更高的要求, 例如一架200座的飞机应出售多少座位?

每个订座旅客当作一次试验，则他或到时不登机，记为A，或到时登机, 记为 $\overline{A}$, P(A)按过去统计资料取为5%。

假定各个旅客登机是独立的, 归为伯努利概型, 并把问题的要求简化与明确,即对不同的超售额*m*计算发生麻烦(即有旅客被拒登机)的概率(没有标准答案的).

假定超售*m*个座位, 则共售出200+*m*个座位, 这时要求登机的旅客数X服从二项分布B(200+*m*, 0.95), 发生拒登机的可能性:

$$P=P\{X>200\}=\sum_{k=201}^{200+m} C_{200+m}^{k}\,0.95^{k}(0.05)^{200+m-k}$$

$$\approx \Phi\left(\frac{(200+m)-(200+m)p}{\sqrt{(200+m)p(1-p)}}\right)-\Phi\left(\frac{201-(200+m)p}{\sqrt{(200+m)p(1-p)}}\right)$$

实际问题的最终解决大都要求应用数学家与实际部门反复磋商.

例、设某种工艺需要某合格的产品100个，该产品的合格率为96%,问需要采购多少个这种产品，才能有95%以上的把握保证合格品数够用？

解：设采购n件产品，其中合格品数为X，

要保证合格品数够用，X必须满足 $X \geq 100$

因此问题归结为求最小的正整数n，使得

$$P(X \geq 100) \geq 0.95$$

由于$X \sim B(n,\ 0.96)$,

由棣莫弗—拉普拉斯中心极限定理可得

$X \sim N(np,\ np(1-p))$ 即 $N(0.96n,\ 0.96 \times 0.04n)$

从而有 $P(X \geq 100) = 1 - P(X < 100)$

$$\approx 1 - \Phi\left(\frac{100 - 0.96n}{\sqrt{0.96 \times 0.04n}}\right) \geq 0.95$$

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

$$= \Phi\left(\frac{0.96n - 100}{\sqrt{0.96 \times 0.04n}}\right) \geq 0.95$$

查表得：$\dfrac{0.96n - 100}{\sqrt{0.96 \times 0.04n}} \geq 1.64$

化简得：$n - 0.2858\sqrt{n} - 104.1667 \geq 0$

求解得：$\sqrt{n} \geq 10.3501$ 或 $\sqrt{n} \leq -20.1286$(舍弃)

也即 $n \geq 107.1432$

故 至少需要采购108个产品才能满足要求