第 四 章

大数定律与中心极限定理

注记

- 本章特征函数部分不作要求；
- 应用特征函数方法证明大数律或中心极限定理的部分不作要求。

- **What should I do if I can't calculate a probability or expectation exactly?**

Don't panic.

- **Simulate it:** Using Monte Carlo simulation
- **Bound it:** Using inequalities
- **Approximate it:** Using limit theorems

第二节：RV 的两种收敛性

概率论中有多种收敛性。

最常用的两种收敛性：

(1) 依概率收敛（Convergence in Probability)

用于大数定律；

(2) 按分布收敛（Convergence in Distribution)

用于中心极限定理。

一、问题的提出

伯努利试验：只有**两个**可能结果的随机试验称为**伯努利试验**.

$$P(A)=p, P(\overline{A})=1-p.$$

n重伯努利试验：重复进行n次独立伯努利试验。

四个约定：

- 每次试验出现**两个**可能结果之一A或非A。
- A在每次试验中出现的**概率保持不变**。
- 各次试验相互**独立**。
- 共进行n次试验。

一、问题的提出

n重伯努利试验中，事件A发生的次数 μ_n 是RV，服从二项分布

$$P(\mu_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0,1,2,\ldots,n.$$

$$\Rightarrow E(\mu_n) = np, \quad D(\mu_n) = npq.$$

当n很大时，μ_n一般也很大 $(E(\mu_n), D(\mu_n)$均很大）。

所以直接研究μ_n不恰当。

考虑频率$\dfrac{\mu_n}{n}$：$E\left(\dfrac{\mu_n}{n}\right) = p, \quad D\left(\dfrac{\mu_n}{n}\right) = \dfrac{pq}{n}.$

均值不变，方差趋于**0**.

频率$\dfrac{\mu_n}{n}$是**RV**，如何定义其极限？

一、问题的提出

在N次重复试验中，事件A发生的频率 $f_n := \frac{\mu_n}{n}$.

虽然随机事件在某次试验中可能发生或不发生，但在大量的试验中却呈现出明显的规律性 --- 频率的稳定性。

例：抛硬币

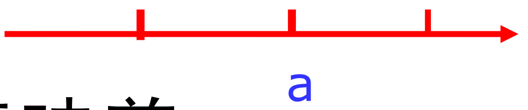
实验者	抛硬币次数	出现正面次数	频率	
Buffon	4040	2048	0.5069	
Pearson	12000	6019	0.5016	
Pearson	24000	12012	0.5005	

问题：如何刻画“频率的稳定性”？

频率的极限？

回顾：传统极限定义

$(a-\varepsilon,\ a-\varepsilon)$



若$\{x_n\}$为确定性序列,$\lim\limits_{n\to\infty} x_n = a$意味着：

$\forall \varepsilon > 0, \exists n_0,$使当$n \ge n_0$时，必定有$|x_n - a| < \varepsilon.$

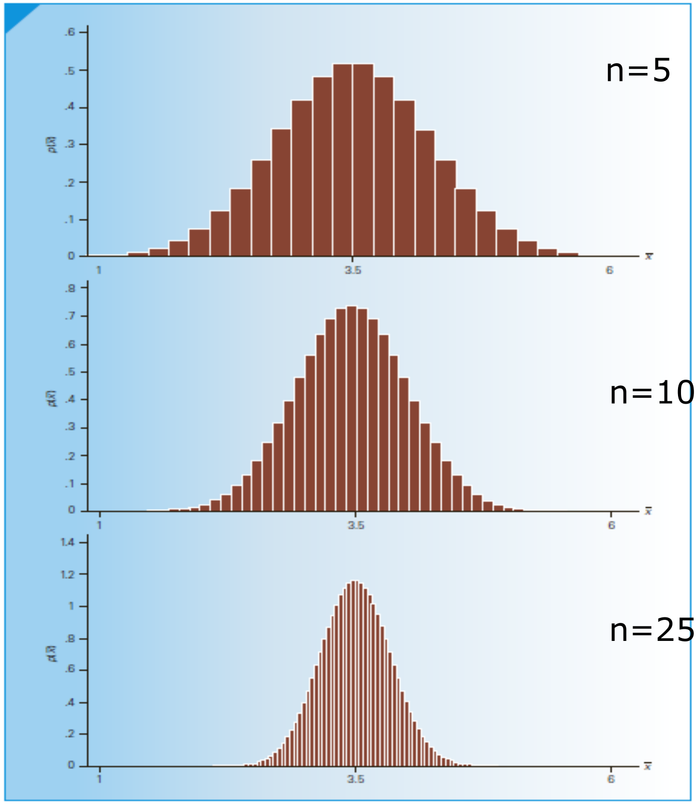
(而绝不会有$|x_n - a| \ge \varepsilon$)

这种收敛性是否适合随机变量的收敛性要求？

频率的极限？

在n重Bernoulli试验中，A发生的概率为 p，μ_n为A发生的次数。**频率** $f_n = \mu_n/n$. 考虑 **RV** 序列$\{f_n\}$.

频率fn（作为RV）的分布 ⟶



注意:

- 不论 n 多大，都有可能 $f_n = 0$

  (这一事件的概率为 $(1-p)^n$)

  导致 f_n 与p的总可能有较大差距。

- 但，f_n与p**有差距较大的概率很小**，即有较小差距的概率很大。

频率的极限？

Bernoulli 的提法（Law of Large Numbers, LLN）

$\{f_n\}$为频率序列，p为事件A发生的概率。

$$\forall \varepsilon > 0,\ \lim_{n\to\infty} P(|f_n - p| \geq \varepsilon) = 0,$$

或 $$\lim_{n\to\infty} P(|f_n - p| < \varepsilon) = 1.$$

称$\{f_n\}$依概率收敛于p, 记为 $f_n \xrightarrow{P} p.$

理解：

$\forall \varepsilon > 0$，当n充分大时，不等式$|f_n - p| < \varepsilon$成立的概率很大。但不论n多大，并不排除$|X_n - p| \geq \varepsilon$的发生，只是发生的概率很小。

频率的极限？

Borel的提法（Strong Law of Large Numbers）*

$\{f_n\}$为频率序列，p为事件A发生的概率:

$$P\left(\lim_{n\to\infty} f_n = p\right) = 1.$$

称$\{f_n\}$以概率**1**收敛于p, 记为 $f_n \xrightarrow{a.s.} p$.

频率的极限？

De Moivre的提法（Centre Limit Theorem）

$\{f_n\}$为频率序列，考虑标准化的随机变量

$$X_n = \frac{f_n - p}{\sqrt{pq/n}} = \frac{\mu_n - np}{\sqrt{npq}}.$$

研究X_n的分布函数的极限行为:

$$\lim_{n\to\infty} P(X_n < x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

X_n的分布函数的极限分布是正态分布**N(0,1).**

De Moivre 证明了$p = 1/2$的情况。

Laplace 推广到$0 < p < 1$的情况。

二、依概率收敛

定义： 若$\{X_n\}$为RV序列，a为一常数，若

$$\forall \varepsilon > 0,\ \lim_{n\to\infty} P(|X_n - a| \geq \varepsilon) = 0,$$

或
$$\lim_{n\to\infty} P(|X_n - a| < \varepsilon) = 1.$$

则称$\{X_n\}$依概率收敛于a,记为 $X_n \xrightarrow{P} a.$

理解： $\forall \varepsilon > 0$，当n充分大时，不等式$|X_n - a| < \varepsilon$成立的概率很大。但不论n多大，并不排除$|X_n - a| \geq \varepsilon$的发生，只是发生的概率很小。

注：$X_n \xrightarrow{P} X \iff X_n - X \xrightarrow{P} 0.$

定理：

设$X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$ (a,b为常数)，又函数$g(x,y)$

在(a,b)点连续，则$g(X_n,Y_n) \xrightarrow{P} g(a,b)$.

证：由$g(x,y)$在(a,b)点的连续性，有 $\forall \varepsilon > 0, \exists \delta > 0$,

使当$|x-a|<\delta$及$|y-b|<\delta$时，有$|g(x,y)-g(a,b)|<\varepsilon$.

则随机事件

$$\{|g(X_n,Y_n)-g(a,b)|<\varepsilon\} \supset \{|X_n-a|<\delta\} \cap \{|Y_n-b|<\delta\}.$$

对立事件

$$\{|g(X_n,Y_n)-g(a,b)|\geq\varepsilon\} \subset \{|X_n-a|\geq\delta\} \cup \{|Y_n-b|\geq\delta\}.$$

$\Rightarrow$

$$P\{|g(X_n,Y_n)-g(a,b)|\geq\varepsilon\} \leq P\{|X_n-a|\geq\delta\} + P\{|Y_n-b|\geq\delta\}.$$

右边两项在$n\to\infty$时，趋于0（因$X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$），

故右边在$n\to\infty$时趋于0。

推论：

设 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$ (a,b为常数),

则 $$X_n \pm Y_n \xrightarrow{P} a \pm b;$$

$$X_n \cdot Y_n \xrightarrow{P} a \cdot b;$$

$$\frac{X_n}{Y_n} \xrightarrow{P} \frac{a}{b} \quad (b \neq 0).$$

例：

设$X_1, X_2,\ldots$独立同分布，均服从$U(0,1)$.

定义$Y_n=\min\{X_1,\ldots,X_n\}$.

直观：Y_n RV 序列，可能依概率收敛于0.

$$\forall \varepsilon>0,\ P(|Y_n-0|\geq \varepsilon)=P(\min\{X_1,\ldots,X_n\}\geq \varepsilon)$$
$$=P(X_1\geq \varepsilon,\ldots,X_n\geq \varepsilon)$$
$$=P(X_1\geq \varepsilon)\cdots P(X_n\geq \varepsilon)$$
$$=(1-\varepsilon)^n\to 0\,(n\to\infty).$$

确定性极限

所以，Y_n依概率收敛于0.

例：

设Y_n依概率收敛于某常数a，问：$E[Y_n]$是否收敛于a？

设

Y_n	0	n^2
P	$1-1/n$	$1/n$

则$\forall \varepsilon > 0, \lim\limits_{n\to\infty} P(|Y_n - 0| > \varepsilon) = \lim\limits_{n\to\infty} \frac{1}{n} = 0,$

所以，$Y_n \xrightarrow{P} 0.$

但，$E[Y_n] = n^2 \cdot \frac{1}{n} = n \to \infty.$

三、弱收敛、按分布收敛

想法：RV都有分布函数，研究对应的分布函数序列。

定义： 设**RV** X的分布函数为$F(x)$, X_n的分布函数为$F_n(x)$.

若对$F(x)$的任一连续点x, 都有

$$\lim_{n\to\infty} F_n(x) = F(x),$$

则称 $\{F_n(x)\}$弱收敛于$F(x)$,

记为 $F_n(x) \xrightarrow{\mathrm{W}} F(x)$.

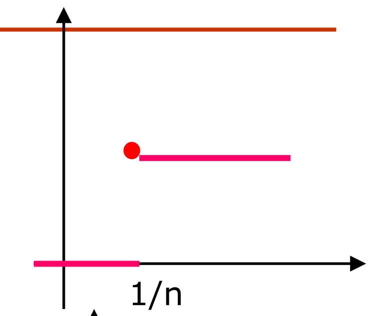
也称对应的**RV**序列$\{X_n\}$按分布收敛于X,

记为 $X_n \xrightarrow{\mathrm{L}} X$.

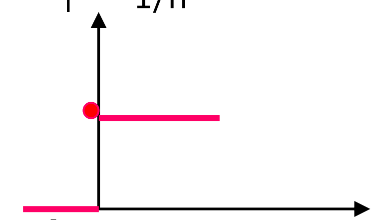
注记：“弱收敛”比“点点收敛”更弱。见下例

例：

设$\{X_n\}$为RV序列：$P(X_n = 1/n) = 1$，$n = 1,2,\ldots$



则其分布函数：$F_n(x) = \begin{cases} 1, x \geq 1/n, \\ 0, x < 1/n. \end{cases}$



设X为RV：$P(X = 0) = 1$，

其分布函数：$F(x) = \begin{cases} 1, x \geq 0, \\ 0, x < 0. \end{cases}$ $F(x)$在$x \neq 0$均连续。

易知：在$F(x)$的连续点，即对$x \neq 0$，均有$\lim\limits_{n \to \infty} F_n(x) = F(x)$；

所以，$F_n(x) \xrightarrow{W} F(x)$；或$X_n \xrightarrow{L} X$.

X<0时，Fn(x)=F(x)=0.

X>0时，F(x)=1,
Fn(x)=1（对 n>1/x）

注：对$x = 0$，有$\lim\limits_{n \to \infty} F_n(0) = 0 \neq F(0) = 1$，

即$F_n(x)$不是点点收敛到$F(x)$.

弱收敛与按分布收敛的关系

定理： 设$X_n \xrightarrow{P} X$，则$X_n \xrightarrow{L} X$.

(但反过来不成立)

如：定义

X	-1	1
P	$1/2$	$1/2$

定义$X_n = -X$,则X_n与X同分布，具有相同的分布函数，

所以，$X_n \xrightarrow{L} X$.

但对$0<\varepsilon<2, P(|X_n - X| \geq \varepsilon) = P(2|X| \geq \varepsilon) = P(2 \geq \varepsilon) = 1.$

即，X_n不依概率收敛到X。

定理： 若c为常数，则$X_n \xrightarrow{P} c \Leftrightarrow X_n \xrightarrow{L} c.$

第三节：大数定律

一、Markov不等式和Chebyshev不等式 - bounds on tail probabilities

定理：（Markov不等式）

设X为取非负值的**RV**,则对于任何常数$a>0$,

有 $$P(X \ge a) \le \frac{E(X)}{a}.$$ （只依赖于期望，只是一个界）

证：对于$a>0$, 令$\mathbf{I} = \mathbf{I}_{\{X \ge a\}} = \begin{cases} 1, X \ge a \\ 0, \text{其它。} \end{cases}$ (*)

由于$X \ge 0$,有 $\mathbf{I} \le \dfrac{X}{a}$.

两边取期望:$E(\mathbf{I}) \le \dfrac{E(X)}{a}$. 由(*),

$$E(\mathbf{I}) = P(X \ge a) \Rightarrow P(X \ge a) \le \frac{E(X)}{a}.$$

期望：求和、积分 ...

Let X be the income. Take a =2EX. Then

$$P(X \ge 2EX) \le 1/2.$$

It is impossible for more than half the population to make at least twice the average income.

定理：（Chebyshev不等式）

设RV X有数学期望$E(X)=\mu$,有方差$D(X)=\sigma^2$.

则$\forall\ \varepsilon>0$,有$P(|X-E(X)|\geq\varepsilon|)\leq\dfrac{\sigma^2}{\varepsilon^2}$.（方差小，有大偏差的概率小）

证1: $P(|X-\mu|\geq\varepsilon)=\displaystyle\int_{|x-\mu|\geq\varepsilon} f(x)dx\leq\int_{|x-\mu|\geq\varepsilon}\frac{(x-\mu)^2}{\varepsilon^2}f(x)dx$ (放大被积函数)

已证

$$\leq\int_{-\infty}^{+\infty}\frac{(x-\mu)^2}{\varepsilon^2}f(x)dx \text{ (放大积分区域)} =\frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx=\frac{\sigma^2}{\varepsilon^2}.$$

证2: 由于$(X-\mu)^2$为非负RV,利用**Markov**不等式得

$$P\{(X-E(X))^2\geq\varepsilon^2\}\leq\frac{E[(X-E(X))^2]}{\varepsilon^2},$$

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

If we are $\varepsilon=k\sigma$ away from the mean μ, then the total prob in the tails is less than or equal to 1/k^2.

注记

- Markov不等式和 Chebyshev不等式对RV的分布没有限制（知道分布的信息很少）。
  - 只要 E(X)和 D(x)有限，Chebyshev不等式就可用；
  - 只要 E(X) 有限，RV本身非负，Markov不等式就可用。
- Chebyshev不等式所给出的估计一般比Markov不等式所给出的估计更好（用到的信息更多）。
- 两个不等式所给出的估计比较保守。但它们在没有添加其它条件下不能改进（它们的界可以被某些分布达到）。
- 更好的估计可以在添加附加条件后得到。

Bounds on Normal tail probability

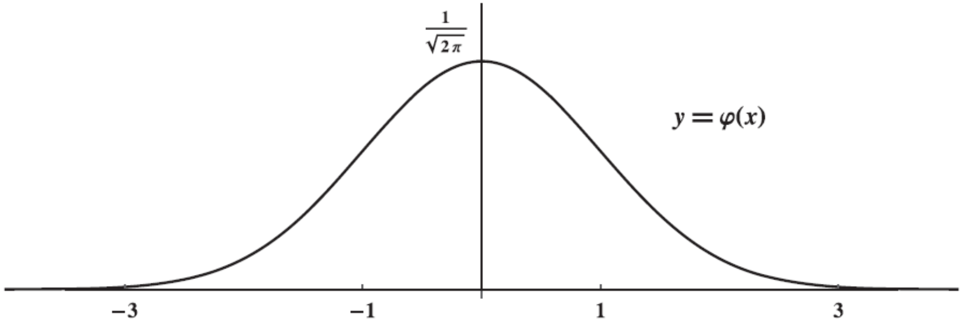
- Markov: we found that $E|Z| = \sqrt{2/\pi}$. Then

$$P(|Z| > 3) \leq \frac{E|Z|}{3} = \frac{1}{3} \cdot \sqrt{\frac{2}{\pi}} \approx 0.27.$$

- Chebyshev:

$$P(|Z| > 3) \leq \frac{1}{9} \approx 0.11.$$

What is the exact value of P(|Z|>3)? P = 0.0027.



例：期中测试成绩为RV，均值80.

（1） 估计概率 $P(X \geq 90)$

（2） 若$D(X)=25$，$P(X\geq 90)$的估计有何变化？

若X是正态分布呢？

利用Markov不等式，

$$P(X\geq 90)\leq E(X)/90 = 8/9.$$

利用Chebyshev不等式，

$$
\begin{aligned}
P(X\geq 90) &= P(X-E(X)\geq 90-\overset{80}{E(X)}) \\
&\leq P(|X-E(X)|\geq 10) \\
&\leq D(X)/100 = 1/4.
\end{aligned}
$$

二、大数定律 (Law of Large Numbers, LLN)

1. 定理（Bernoulli 大数律）

若 X_n 为 n 重**Bernoulli**试验中事件A发生的频率，p为每次试验中A发生的概率，则X_n依概率收敛于 p：

$$X_n \xrightarrow{P} p,\quad \text{即}$$

$$\forall \varepsilon > 0,\ \lim_{n\to\infty} P(|X_n - p| \geq \varepsilon) = 0,$$

或 $$\lim_{n\to\infty} P(|X_n - p| < \varepsilon) = 1.$$

要证X_n依概率收敛于p: $X_n \xrightarrow{P} p$.

证：记Y_n为n重**Bernoulli**试验中事件**A**发生的次数，

则$Y_n \sim B(n, p)$, $E(Y_n) = np$, $D(Y_n) = np(1-p)$.

因 $X_n = Y_n / n \Rightarrow E(X_n) = p$,

$D(X_n) = D(Y_n)/n^2 = p(1-p)/n$.

$\forall \varepsilon > 0$, 由**Chebyshev**不等式

$$0 \le P(|X_n - p| \ge \varepsilon) \le \frac{D(X_n)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}$$

$$\Rightarrow \lim_{n\to\infty} P(|X_n - p| \ge \varepsilon) = 0.$$

当$\varepsilon = 0.01$, $p = 0.5$,

$P(|X_n - 0.5| \ge 0.01) \le \frac{10^4}{4n}$

However, Bernoulli大数律发表在1713年（在Chebyshev之前），如何用Chebyshev不等式？

定义：（序列{Xn}服从大数律）

若$\{X_n\}$为**RV**序列，称$\{X_n\}$服从大数律，如果

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i) \xrightarrow{P} 0,\text{即}$$

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

或

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

2. 定理：（Chebyshev大数律）

设$\{X_n\}$为相互独立(或两两不相关)RV序列，

各X_i方差均存在，且有公共上界$D(X_i)\le C, i=1,2,\cdots$

则$\{X_n\}$服从大数律：

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i) \xrightarrow{P} 0,\text{即}$$

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

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

证明：

记 $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$， 则$E(Y_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$，

$$D(Y_n) = \frac{1}{n^2} D\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) \text{（独立或两两不相关）}$$

$$\le \frac{1}{n^2}\cdot nC = \frac{C}{n}.$$

$\forall \varepsilon > 0$，由Chebyshev不等式

$$0 \le P(|Y_n - E(Y_n)| \ge \varepsilon) \le \frac{D(Y_n)}{\varepsilon^2} \le \frac{C}{n\varepsilon^2}.$$

$$\Rightarrow \lim_{n\to\infty} P(|Y_n - E(Y_n)| \ge \varepsilon) = 0.$$

注记(两个特例)

(1) 当$\{X_n\}$为相互独立(或两两不相关)**RV**序列，

且各X_i具有相同的数学期望μ和方差时，

$\{X_n\}$服从大数律：$\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \xrightarrow{P} 0$，即

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

或

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

(2) 当$X_1, X_2, \ldots,$相互独立(或两两不相关)，同分布（独立同分布），

且方差有限时，则$\{X_n\}$服从大数律。 一样的形式

3. 定理：（Markov大数律）

设$\{X_n\}$为**RV**序列，$E(X_i), D(X_i)$存在，

若$\frac{1}{n^2} D\left(\sum_{i=1}^{n} X_i\right) \to 0 \ (n \to \infty)$ (**Markov**条件),

则$\{X_n\}$服从大数律：

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i) \xrightarrow{P} 0, \text{即}$$

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

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

注记：没有独立性或不相关或方差一致有界的假设。

证明： 记$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$,

则$E(Y_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$

$D(Y_n) = \frac{1}{n^2} D\left(\sum_{i=1}^{n} X_i\right) \to 0$ (**Markov条件**)

$\forall \varepsilon > 0$,由**Chebyshev不等式**

$$0 \le P(|Y_n - E(Y_n)| \ge \varepsilon) \le \frac{D(Y_n)}{\varepsilon^2}$$

$$\Rightarrow \lim_{n\to\infty} P(|Y_n - E(Y_n)| \ge \varepsilon) = 0.$$

例： $\{X_n\}$ 为同分布，方差存在的RV序列，且 X_n 仅与 X_{n-1} 和 X_{n+1} 相关，与其它 X_k 不相关，问 $\{X_n\}$ 是否服从大数律？

解： 考虑：$D\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} D(X_i)+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbf{Cov}(X_i,X_j)$

$$=\sum_{i=1}^{n} D(X_i)+2\sum_{i=1}^{n-1}\mathbf{Cov}(X_i,X_{i+1}).\ (*)$$

设 $D(X_i)=\sigma^2$，则

$|\mathbf{Cov}(X_i,X_{i+1})|\le\sqrt{D(X_i)}\sqrt{D(X_{i+1})}=\sigma^2.$ (**Cauchy**不等式)

由(*)有 $\frac{1}{n^2}D\left(\sum_{i=1}^{n} X_i\right)\le\frac{1}{n^2}\left[n\sigma^2+2(n-1)\sigma^2\right]\to 0\ (n\to\infty)$

即**Markov**条件成立，从而 $\{X_n\}$ 服从大数律：

$$\frac{1}{n}\sum_{i=1}^{n} X_i-\frac{1}{n}\sum_{i=1}^{n} E(X_i)\xrightarrow{P}0.$$

4. 定理：（辛钦大数律）

设RV $X_{1,} X_{2,}\ldots$独立同分布，有数学期望μ

(对方差不作要求)，则$\{X_n\}$服从大数律：

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \xrightarrow{P} 0,\text{即}$$

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

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

注记：方差不一定存在，不能用Chebyshev不等式。

必须用别的方法证明！→ 特征函数方法

Summary

辛钦

$\Downarrow$

Markov $\Rightarrow$ Chebyshev $\Rightarrow$ Bernoulli

(1) **Bernoulli大数律大数定律是Chebyshev大数定律的特例。**

(2) **Chebyshev大数定律是Markov大数定律的特例。**

(3) **Bernoulli大数定律是辛钦大数定律的特例。**

(4) **辛钦大数定律不是的Chebyshev大数定律的推广**

✓ **辛钦大数律要求同分布，对方差没有要求；**

✓ **Chebyshev大数律不要求同分布，但对方差有要求。**

(5) **大数律中所假设的条件是充分条件。**

条件不满足的随机变量序列也可能服从大数律。

三、大数律的应用：Monte Carlo方法

- **Computing probability $P(A)$ by simulation**

- **Simulate a trial:** Model the random experiment using random numbers on the computer.
- **Determine success:** Based on the outcome of the trial, determine whether or not the event A occurs. If yes, call that a “success”.
- **Replication:** Repeat the two steps above many times. The proportion of successful trials is the estimate of $P(A)$.

Computing probability by simulation

令 $X_i = \begin{cases} 1，事件A发生， \\ 0，否则。 \end{cases}$

则$X_1,\ldots,X_n$独立同分布，

$E(X_i) = P(X_i = 1) = P(A).$

由LLN,$\bar{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} P(A).$ 频率趋于概率

例：设X ~ Gamma(3, 5)，计算P(1<X<10)

- MC estimate

R code

```
x <- rgamma(n=100000, shape=3, scale=5)  #generate the random sample
mean( 1<x & x<10)
```

Output: 14.99879

- True value

We can compare the Monte Carlo estimate with their true value.

R Code:

```
integrate(function(x) dgamma(x, shape=3, scale=5), lower=1, upper=10)$value
```

Output: 15.0000000

Monte Carlo方法计算定积分

计算积分$\int_a^b f(x)dx,\ \ 0\le f(x)\le c, f(x)$复杂。

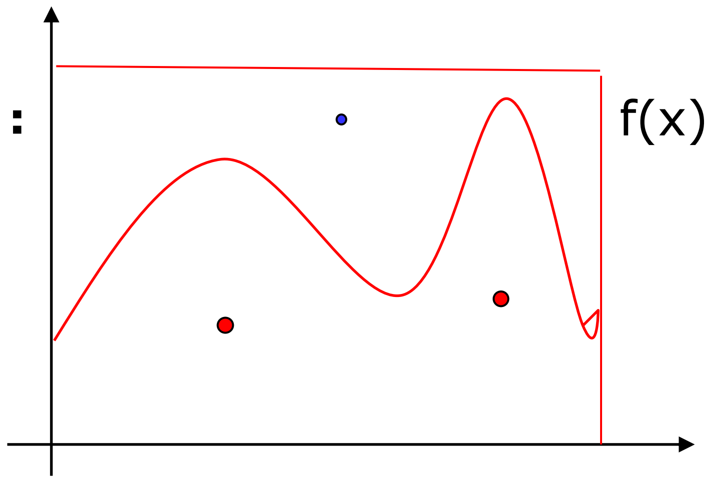
考虑矩形$[a,b]\times[0,c]$.

所求积分就是曲线下方的面积。方法：

(1) 往矩形$[a,b]\times[0,c]$中随机投n个点；

(2) 计算落于曲线下方的点的比例$\overline{p}$;

(3) $\int_a^b f(x)dx \approx \overline{p}\times(b-a)c.$ （频率）



Note：点落于曲线下方的概率$p=\int_a^b f(x)dx/[(b-a)c].$ 几何概型

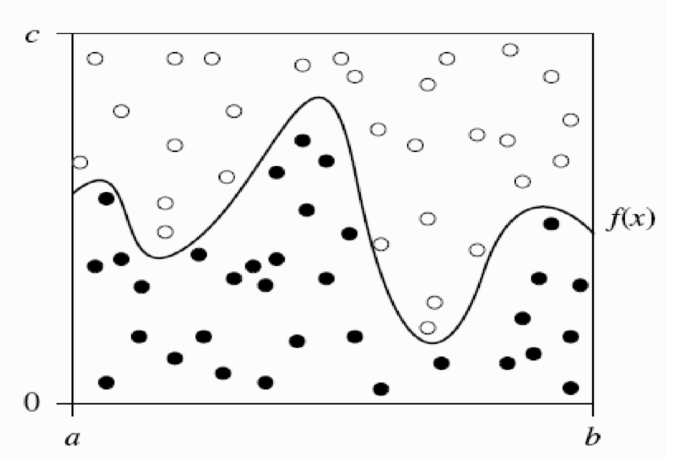
由**LLN**，$\overline{p}\xrightarrow{\mathbf{P}} p=\int_a^b f(x)dx/[(b-a)c].$ 频率趋于概率

Acceptance/Rejection

令 $X_i = \begin{cases} 1, & \text{随机点落于曲线下方,} \\ 0, & \text{否则。} \end{cases}$

则$X_1,\ldots,X_n$独立同分布，

$$E(X_i) = P(X_i = 1) = \int_a^b f(x)dx/[(b-a)c].$$



几何概型

由LLN，$\overline{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \int_a^b f(x)dx/[(b-a)c].$

频率趋于概率

$$\Rightarrow \int_a^b f(x)dx \approx \overline{p}\,[(b-a)c].$$

Monte Carlo方法一般原理

- Let μ be the quantity to be approximated. Assume that we can design a RV X so that

$$E(X) = \mu.$$

- If we now generate a sequence of i.i.d. RV $X_1, X_2, X_3, \ldots$ with the same distribution as X, then the Law of Large Numbers tells us

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} \mu \quad (\text{as } n \to \infty)$$

- Take n large enough.

例：一维积分

- 计算积分 $$I = \int_0^{\infty} e^{\sin(\ln(x)) - x^2}\, dx.$$

因为 $I = \int_0^{\infty} e^{\sin(\ln(x)) - x^2} e^x e^{-x}\, dx = E\left[e^{\sin(\ln(X)) - X^2 + X}\right]$, $X \sim \mathrm{Exp}(1)$.

令 Xi 为相互独立 Exp(1) RV, 且 $Y_i = e^{\sin(\ln(X_i)) - X_i^2 + X_i}$.

用 $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ 估计积分。

```
set.seed(123456)
N <- 1000000
x <- rexp(N, 1)
integral <- sum(exp(sin(log(x))- x^2 + x))/N
integral

## [1] 0.6542623
```

例：一维积分

一般：

$$\int_a^b g(x)\,dx = \int_a^b \frac{g(x)}{f(x)} f(x)\,dx = E\left[\frac{g(X)}{f(X)}\right],$$

其中，X的密度函数为f(x)，满足$\int_a^b f(x)dx = 1$，如f(x)=$\frac{1}{b-a} I_{\{a<x<b\}}$.

产生 X 的 i.i.d. 样本 X_1，…，X_N。令

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n}\frac{g(X_i)}{f(X_i)},\ X_i \sim f(x).$$

如：

$$\int_a^b g(x)dx = (b-a)\int_a^b g(x)\frac{1}{b-a}dx \approx \frac{(b-a)}{n}\sum_{i=1}^{n} g(X_i),\ X_i \sim U(a,b).$$

例：高维积分

计算积分：$I=\int_0^1\cdots\int_0^1 g(x_1,\ldots,x_s)dx_1\ldots dx_s=\int_{[0,1]^s} g(x)dx.$ (g连续)

Step 1. 随机化：作**RV** $X\sim U[0,1]^s$，则其密度为$f(x)=1, x\in[0,1]^s$.

令$Y=g(X)$，则$E(Y)=E[g(X)]=\int_{[0,1]^s} g(x)dx=I.$ LOTUS

方差$D(Y)=E(Y^2)-I^2=\int_{[0,1]^s} g^2(x)dx-I^2$.(必存在)

Step 2. 构造**RV**序列$X_1, X_2,\ldots,$独立同分布，均服从$U[0,1]^s$;

对应地，有**RV**序列$Y_1=g(X_1), Y_2=g(X_2),\ldots,$

序列$\{Y_i\}$独立同分布, 方差存在，数学期望均为I.

Step 3. 对**RV**序列$\{Y_i\},\ldots,$应用**Chebyshev**大数定律：

$$\overline{\mathbf{Y}}:=\frac{1}{n}\sum_{i=1}^n Y_i=\frac{1}{n}\sum_{i=1}^n g(X_i)\xrightarrow{P} I.$$

例：高维积分

最后，$I \approx \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} g(X_i).$

实际计算中，可用“伪随机数”代人进行计算

注意到：$E(\overline{Y}) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = I,$

$$D(\overline{Y}) = D\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(Y_i) = \frac{D(Y)}{n},$$

均方误差：

$$\mathbf{RMSE} = \sqrt{E[\overline{Y} - I]^2} = \sqrt{D(\overline{Y})} = \sqrt{\frac{D(Y)}{n}}.$$

每增加一位精度，需要增加100倍的计算量。

注：该误差与维数s无关！— 打破“维数的灾难”！！

Monte Carlo在科学和工程计算中具有广泛应用。

但是，收敛速度较慢。

第四节：中心极限定理
(Central Limit Theorem, CLT)

- **大数定律**：研究RV的平均 $\frac{1}{n}\sum_{i=1}^{n} X_i$ 的稳定性或渐近性。
- **中心极限定理**：研究独立RV 的和 $\frac{1}{n}\sum_{i=1}^{n} X_i$ 的极限分布。

实际问题中，某RV是众多独立的随机因素的综合叠加影响所形成。独立RV的和的精确分布可尝试用“卷积公式”，

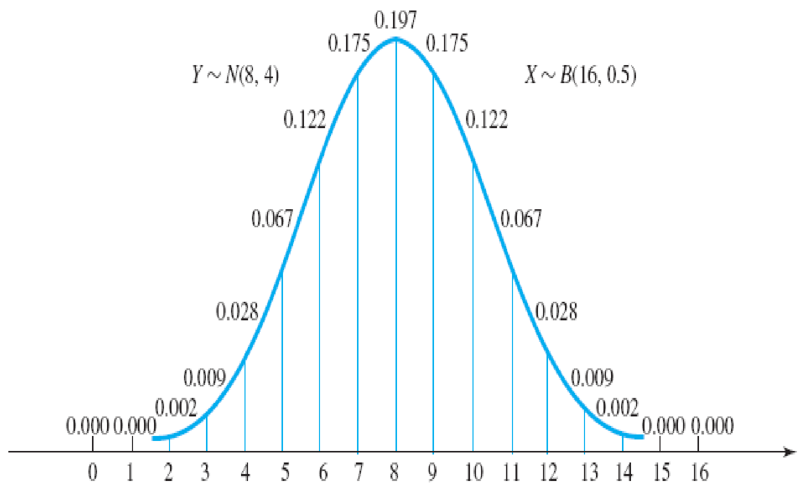
某些情况下可用分布的可加性，如Poission分布、正态分布。

在正态场合：$\frac{1}{n}\sum_{i=1}^{n} X_i$ 精确服从正态分布，均值方差易得。

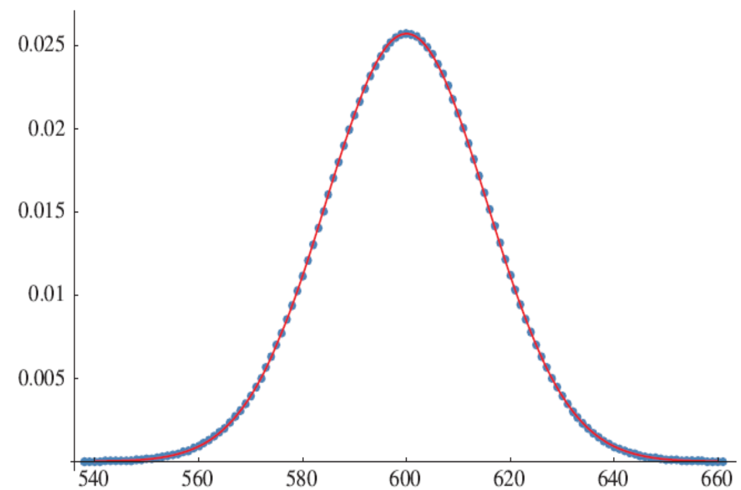
但一般情况下，无此规律，比较复杂，如均匀分布的和。

Normal Approximation to Binomial Dist?

X~B(16, 0.5), Y~N(8,4)



X~B(1000,0.6), Y~N(600,240)



Even though RV *X* has a *discrete* distribution and RV *Y* has a *continuous* distribution, the shapes of their respective PMF and PDF are quite similar.

注意到：二项分布RV **X = X1 + ··· + Xn** ，Xi 为iid的两点分布。

一、问题的提法

设$S_n=\sum_{k=1}^{n}X_k$，$\{X_k\}$独立同分布，$E(X_k)=\mu, D(X_k)=\sigma^2.$

研究S_n本身？ $D(S_n)=n\sigma^2\to\infty$， S_n“太大”。

研究$\frac{S_n}{n}$？ $D(\frac{S_n}{n})=\frac{1}{n^2}n\sigma^2\to 0$， $\frac{S_n}{n}$“太小”。

对S_n进行标准化： $S_n^*=\frac{S_n-E(S_n)}{\sqrt{D(S_n)}}=\frac{S_n-n\mu}{\sqrt{n}\,\sigma}$,

此时，$E(S_n^*)=0, D(S_n^*)=1.$

问题：S_n^*的极限分布是什么？

二、Lindeberg-Levy中心极限定理

定理：

设$\{X_k\}$是独立同分布RV序列，$E(X_k)=\mu$,

$$D(X_k)=\sigma^2>0.\quad 记\quad S_n^*=\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\,\sigma},$$

标准化

则 $S_n^* \xrightarrow{L} Z$, $Z\sim N(0,1)$, 即

$$\lim_{n\to\infty}P(S_n^*\le x)=\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt,\ \forall x\in R.$$

关于中心极限定理

- Loosely speaking, CLT theorem asserts that the sum of a large number of independent RVs (not necessarily normal) has a distribution that is approximately normal.

- “几乎没有一种理论能够像中心极限定理那样神奇，那样贴切地体现宇宙次序。如果古希腊人知道这个规律的话，就一定会把它神化。它在混乱中保持着平衡，情况越复杂、混乱，它的主导作用就越完善。它是最卓越、最不可思议的规律”。

注记：

CLT研究独立RV和的极限分布（没有对 X_k 的具体分布作出限制，可以任意，只要均值、方差有限），极限分布即为正态：

$$S_n^* = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\,\sigma} = \frac{\frac{1}{n}\sum_{k=1}^{n} X_k - \mu}{\sigma/\sqrt{n}} \overset{\text{近似}}{\sim} N(0,1).$$

$$\Rightarrow \quad \frac{1}{n}\sum_{k=1}^{n} X_k \overset{\text{近似}}{\sim} N(\mu, \sigma^2/n),$$

$$\sum_{k=1}^{n} X_k \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2).$$

这就是正态分布广泛存在的原因。

不应用CLT，

$\frac{1}{n}\sum_{i=1}^{n} X_i$有均值$\mu$及方差$\sigma^2/n$。

CLT告诉我们额外的信息：

$\frac{1}{n}\sum_{i=1}^{n} X_i$近似服从正态分布。

Remarks

- CLT is one of the most important theorems in the whole of probability theory and in math.
- It may explain why many naturally occurring phenomena are observed to have distributions similar to normal distribution, because they may be considered to be composed of the aggregate of many smaller random events.
- CLT provides a convenient way of approximating the probability values of an average of a set of iid RVs (through the exact distribution of this average will in general be complicated).

Remarks

- CLT leaves open how large sample size *n* needs to be for the normal approximation to be valid. The answer depends on population distribution.
- For instance, if underlying population distribution is normal, then the sample mean X will also be normal regardless of the sample size.
- There is no general answer how large n should be.

例：正态随机数的产生

只要$\{X_k\}$独立同分布，$E(X_k)=\mu, D(X_k)=\sigma^2>0,$

则 $$S_n^*=\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\,\sigma}\overset{\text{近似}}{\sim} N(0,1)$$ ($L-L$中心极限定理)

选$X_k\sim U(0,1)$,则$\mu=1/2$， $\sigma^2=1/12$， 选$n=12$,

则 $$S_n^*=\sum_{k=1}^{12}X_k-6\overset{\text{近似}}{\sim} N(0,1).$$

1* 产生12个$U(0,1)$的随机数$x_1,x_2,\ldots,x_{12}$;

2* 令$y=\sum_{k=1}^{12}X_k-6$,则y可认为是服从$N(0,1)$的一个随机数；

3* 重复以上步骤，可得多个$N(0,1)$随机数；

4* 作变换$w=\mu+\sigma y$,可得一般正态$N(\mu,\sigma^2)$的随机数。

例：误差分析

设$A=\sum_{i=1}^{n}X_i$.计算机在进行加法运算时对每个加数X_i取k位小数,

即每个加数的误差$\varepsilon_i \sim U(-0.5\times10^{-k}, 0.5\times10^{-k})$，且相互独立。

(1) 求总误差$S=\sum_{i=1}^{n}\varepsilon_i$的近似分布。

(2) 问：以0.99的概率可使总误差的绝对值不超过多少？

解：（1）因为$\mu := E(\varepsilon_i)=0$, $\sigma^2=D(\varepsilon_i)=10^{-2k}/12$,

根据$L-L$中心极限定理,有

$$S_n^* = \frac{\sum_{i=1}^{n}\varepsilon_i - n\mu}{\sqrt{n}\,\sigma} = \frac{\sum_{i=1}^{n}\varepsilon_i}{\sqrt{n10^{-2k}/12}} \overset{\text{近似}}{\sim} N(0,1).$$

$$\Rightarrow \sum_{i=1}^{n}\varepsilon_i \overset{\text{近似}}{\sim} N(0, n10^{-2k}/12).$$

例：误差分析

(2) 需要确定$z>0$, 使总误差S满足 $P(|\mathbf{S}|\le \mathbf{z})=\mathbf{0.99}$.

因为$S_n^* = \dfrac{\sum_{i=1}^{n}\varepsilon_i}{\sqrt{n10^{-2k}/12}} = \dfrac{\mathbf{S}}{\sqrt{n10^{-2k}/12}} \overset{\text{近似}}{\sim} N(0,1)$，所以，

$$P(|\mathbf{S}|\le \mathbf{z}) = P\left(\frac{|\mathbf{S}|}{\sqrt{n10^{-2k}/12}} \le \frac{\mathbf{z}}{\sqrt{n10^{-2k}/12}}\right) \approx 2\Phi\left(\frac{\mathbf{z}}{\sqrt{n10^{-2k}/12}}\right) - 1 = 0.99$$

$$\Phi\left(\frac{\mathbf{z}}{\sqrt{n10^{-2k}/12}}\right) = 0.995 \overset{\text{查表}}{\Rightarrow} \frac{\mathbf{z}}{\sqrt{n10^{-2k}/12}} = 2.575$$

$$\Rightarrow z = 2.575\times\sqrt{n10^{-2k}/12} = 0.743\times 10^{-k}\sqrt{n}.$$

即，以概率**0.99**保证总误差不超过误差 $0.743\times 10^{-k}\sqrt{n}$.

比较:

确定性的误差界 $|\mathbf{S}| \le \mathbf{0.5}\times 10^{-k}\times n.$

如：100个数相加，k=1时
(每个加数的误差不超过0.05)，
概率误差界为：0.743。
确定性误差界为 5.

三、De Moivre-Laplace中心极限定理

（二项分布的正态逼近）

设 RV μ_n为n重Bernoulli试验中事件A发生的次数，

即$\mu_n \sim B(n,p),(0<p<1)$，记 $S_n^* = \dfrac{\mu_n - np}{\sqrt{npq}}, (q=1-p)$

标准化

则 $S_n^* \xrightarrow{L} Z, Z \sim N(0,1)$,

即 $$\lim_{n\to\infty} P(S_n^* \le x) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt,\ \forall x \in R.$$

$\{X_k\}$独立同分布，$E(X_k) = \mu$, $D(X_k) = \sigma^2 > 0$.

记 $S_n^* = \dfrac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\,\sigma}$，则 $S_n^* \xrightarrow{L} Z$.

证：引入 $X_k = \begin{cases} 1, & 第k次试验时A发生， \\ 0, & 否则。\end{cases}$

则$\{X_k\}$独立同分布，且 $E(X_k) = p, D(X_k) = pq.$

注意到 $\mu_n = \sum_{i=1}^{n} X_i$，由$L-L$中心极限定理即得。

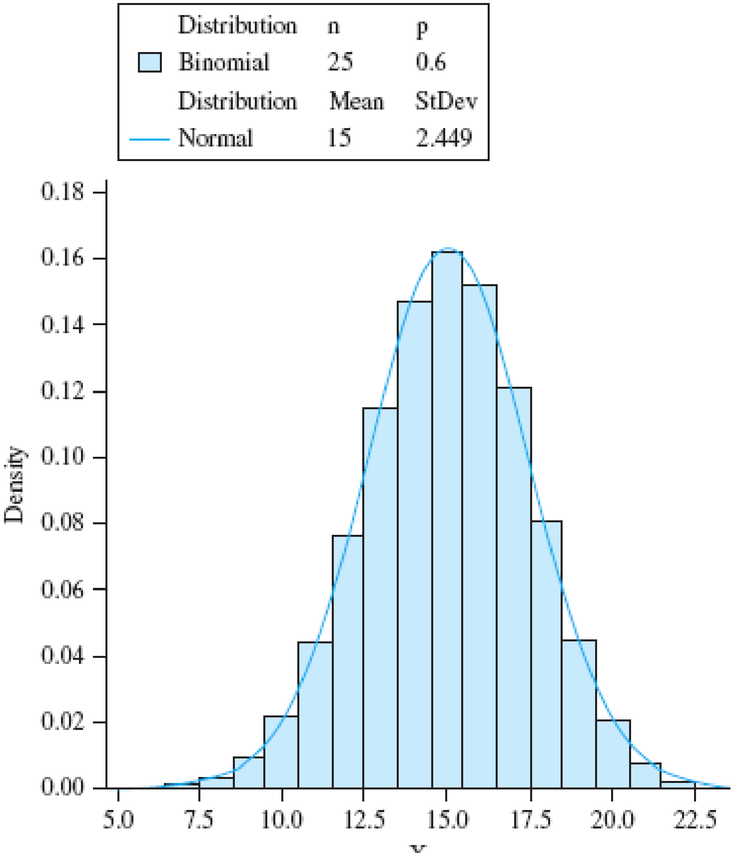
二项分布的正态逼近

$$S_n^* = \frac{\mu_n - np}{\sqrt{npq}},$$

标准化

$$S_n^* \xrightarrow{L} Z$$

$\Rightarrow \mu_n$近似服从N(np, npq)



注记：大数律与中心极限定理的关系

设 $\{X_k\}$是独立同分布RV序列，$E(X_k)=\mu, D(X_k)=\sigma^2>0.$

则由大数律：$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|\le\varepsilon\right)\to 1.$

但括号中事件概率有多大，没有回答。

中心极限定理：$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|\le\varepsilon\right)=P\left(\left|\frac{\frac{1}{n}\sum_{k=1}^{n}X_k-\mu}{\sigma/\sqrt{n}}\right|\le\varepsilon\frac{\sqrt{n}}{\sigma}\right)$

$$\approx\Phi\left(\varepsilon\frac{\sqrt{n}}{\sigma}\right)-\Phi\left(-\varepsilon\frac{\sqrt{n}}{\sigma}\right)=2\Phi\left(\varepsilon\frac{\sqrt{n}}{\sigma}\right)-1.$$

中心极限定理比大数律更精确。

是n的单调函数，n大时，接近1

LLN and CLT

- CLT and LLN require that mean and variance of X_i be finite.
- Cauchy distribution has no mean or variance, so it obeys neither the LLN nor the CLT.

- It can be shown that sample mean of n Cauchys is still Cauchy, no matter how large n gets. So its sample mean never approaches Normal distribution, contrary to CLT.
- There is also no true mean for $\overline{X}_n$ to converge to, so LLN does not apply either.

四、中心极限定理的应用

Recall: Lindeberg-Levy中心极限定理

设 $\{X_k\}$是独立同分布RV序列，$E(X_k)=\mu$,

$$D(X_k)=\sigma^2>0. \quad 记 \quad S_n^*=\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\,\sigma},$$

则 $S_n^* \xrightarrow{L} Z$, $Z\sim N(0,1)$, 即

$$\lim_{n\to\infty}P(S_n^*\le x)=\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt,\ \forall x\in R.$$

Recall: De Moivre-Laplace中心极限定理

（二项分布的正态逼近）

设 RV μ_n为n重Bernoulli试验中事件A发生的次数，

即 $\mu_n \sim B(n, p), (0 < p < 1)$, 记 $S_n^* = \dfrac{\mu_n - np}{\sqrt{npq}}, (q = 1 - p)$

则 $S_n^* \xrightarrow{L} Z, Z \sim N(0,1)$,

即 $\lim\limits_{n \to \infty} P(S_n^* \le x) = \Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-t^2/2} dt, \ \forall x \in R.$

四、中心极限定理的应用

1. 原理 设$\mu_n \sim B(n,p),(0<p<1)$,

则$P(m_1 \le \mu_n \le m_2)=\sum_{k=m_1}^{m_2} C_n^k p^k (1-p)^{n-k}$(直接用二项分布)

用正态逼近:

$$\frac{\mu_n - np}{\sqrt{npq}} \overset{\text{近似}}{\sim} N(0,1).$$ **(De Moivre – Laplace中心极限定理)**

$$\Rightarrow P(m_1 \le \mu_n \le m_2) = P\left(\frac{m_1 - np}{\sqrt{npq}} \le \frac{\mu_n - np}{\sqrt{npq}} \le \frac{m_2 - np}{\sqrt{npq}}\right)$$

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

2. 连续性修正（Continuity Correction）

为什么需要修正？

二项分布是离散分布，而正态分布是连续分布，所以用正态分布作为二项分布的近似时，应作修正。

设$\mu_n \sim B(n, p), (0 < p < 1)$，

$$P(m_1 \le \mu_n \le m_2) \approx \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right).$$

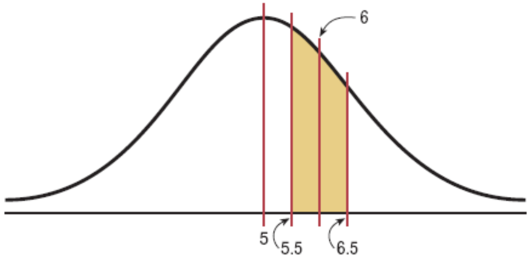
不修正

显然，当$m_1 = m_2$时，误差较大

(右边为0，左边不为0)。

如何修正？

(把 $\mu_n = k$ 当成 $k-0.5 \le \mu_n \le k+0.5$)



$$P(m_1 \le \mu_n \le m_2) = P(m_1 - 0.5 \le \mu_n \le m_2 + 0.5)$$

二项分布RV，往外扩展半个单位

$$= P\left(\frac{m_1 - 0.5 - np}{\sqrt{npq}} \le \frac{\mu_n - np}{\sqrt{npq}} \le \frac{m_2 + 0.5 - np}{\sqrt{npq}}\right)$$

$$\approx \Phi\left(\frac{m_2 + 0.5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - 0.5 - np}{\sqrt{npq}}\right).$$

结果更精确(使用修正法，可以计算μ_n 在单点的概率。)

注记： $P(m_1 < \mu_n < m_2)$ $\qquad$ μ_n取整数值

$$= P(m_1 + 1 \le \mu_n \le m_2 - 1)$$

先化为两边带等号

例：

设每颗炮弹命中目标的概率为0.01，求500发炮弹中命中 5 发的概率.

解：设X表示命中目标的次数，则$X \sim B(500, 0.01)$.

(1) 直接用二项分布计算:

$$P(X=5)=C_{500}^{5}\times 0.01^{5}\times 0.99^{495}=0.17635.$$

dbinom(5, 500, 0.01)=0.176351

(2) 用正态逼近：

$$P(X=5)=P(4.5<X<5.5)$$

修正

$$\approx \Phi\left(\frac{5.5-5}{\sqrt{4.95}}\right)-\Phi\left(\frac{4.5-0.5}{\sqrt{4.95}}\right)=0.1742.$$

例：修正 or 不修正

pbinom(15, 25, 0.4) - pbinom(4, 25, 0.4) = 0.977

设$\mu_n \sim B(25,0.4)$,求$P(5 \le \mu_n \le 15)$

(1)精确值：$P(5 \le \mu_n \le 15) = 0.97$ 70 直接用二项分布计算)

(2)不修正：$P(5 \le \mu_n \le 15) \approx \Phi\left(\dfrac{15-10}{\sqrt{6}}\right) - \Phi\left(\dfrac{5-10}{\sqrt{6}}\right)$

$$= 2\Phi(2.041) = 0.9588.$$

(3)修正:

$$P(5 \le \mu_n \le 15) = P(5-0.5 \le \mu_n \le 15+0.5)$$

$$\approx \Phi\left(\frac{15+0.5-10}{\sqrt{6}}\right) - \Phi\left(\frac{5-0.5-10}{\sqrt{6}}\right)$$

$$= 2\Phi(2.245-1) = 0.9754.$$ 结果更精确

例：设某课程注册的学生数服从Poisson分布，均值为100，学校决定如果注册学生数大于等于120，则分两个班授课。求分两个班授课的概率。

解：方法1. 设X为注册学生数，则$X \sim P(100)$,

所求概率为：$\sum_{i=120}^{\infty} \frac{100^i}{i!} e^{-100}$.

= 1-ppois(119, 100) = 0.0282

方法2. 注意到Poisson分布具有可加性，可把X写成：

$$X = X_1 + X_2 + \cdots + X_{100}, \quad X_i \sim P(1)\text{且相互独立。}$$

则$P(X \geq 120) = P(120 \leq X < \infty) = P(119.5 \leq X < \infty)$

修正

$$= P\left(\frac{119.5 - 100}{\sqrt{100}} \leq \frac{X - 100}{\sqrt{100}} < \infty \right)$$

中心极限定理

$$\approx 1 - \Phi(1.95) = 0.0256.$$

3. 应用中的各种情况

设 $\mu_n \sim B(n,p)\ (0<p<1)$, 记 $S_n^* = \dfrac{\mu_n - np}{\sqrt{npq}}$,

有 $P\left(S_n^* \le y\right) \approx \Phi(y) = \beta.$ (*)

近似等式中涉及三个数: $n, y, \beta.$

已知其中任何二个数，可以求第三个数。

(1) 已知 n, y, 求 β; (直接用(*))

(2) 已知 n, β, 求 y; (由(*)反解出 y)

(3) 已知 y, β, 求 n; (由(*)反解出 n)

例： 设$Y_n \sim B(100,0.9)$, 求$P(Y_n \geq 85)$.

1-pbinom(84, 100, 84)=0.9601

解：$E(Y_n)=90; D(Y_n)=9;$

可对 $P(Y_n < 85)$= P(-∞ < $Y_n \leq 84$)
= P(-∞ < $Y_n \leq 84.5$)

(1)不修正：

$$P(Y_n \geq 85)=1-P(Y_n<85)$$

$$=1-P\left(\frac{Y_n-np}{\sqrt{npq}}<\frac{85-np}{\sqrt{npq}}\right)\approx 1-\Phi\left(\frac{85-90}{3}\right)=\Phi\left(\frac{5}{3}\right)=0.9525.$$

(2)修正：

$$P(Y_n \geq 85)=P(85 \leq Y_n <+\infty)=P(85-0.5 \leq Y_n <+\infty)$$

带等号的地方往外拓展

$$=P\left(\frac{85-0.5-np}{\sqrt{npq}} \leq \frac{Y_n-np}{\sqrt{npq}}<+\infty\right)$$

$$\approx 1-\Phi\left(\frac{85-0.5-90}{3}\right)=\Phi(5.5/3)=0.9664.$$

例：

某项保险业务，保费60元/年。若一年内有事故，赔金2万，事故概率$p=0.005$.有$n=5000$人投保。问保险公司在此项业务的收益在20万~40万的概率。

解：令η表示发生事故总数，则$\eta \sim B(5000,0.005)$.

保险公司收益：$0.016\times 5000-2\eta=80-2\eta$.

$P(20\le 80-2\eta\le 40)=P(20\le \eta\le 30)$ = pbinom(30, 5000, 0.005) - pbinom(19, 5000, 0.005)=0.7310

$$=P\left(\frac{20-np}{\sqrt{npq}}\le\frac{\eta-np}{\sqrt{npq}}\le\frac{30-np}{\sqrt{npq}}\right)\approx\Phi(1)-\Phi(-1)=2\Phi(1)-1=0.6826.$$

*修正：

$P(20\le\eta\le 30)=P(20-0.5\le\eta\le 30+0.5)$ **带等号的地方往外拓展**

$$=P\left(\frac{20-0.5-np}{\sqrt{npq}}\le\frac{\eta-np}{\sqrt{npq}}\le\frac{30+0.5-np}{\sqrt{npq}}\right)\approx 2\Phi(1.10)-1=0.7286.$$

第2类问题：

设 $\mu_n \sim B(n, p)\ (0 < p < 1)$, 记 $S_n^* = \dfrac{\mu_n - np}{\sqrt{npq}}$,

有 $P\left(S_n^* \le y\right) \approx \Phi(y) = \beta. \qquad (*)$

近似等式中涉及三个数: $n, y, \beta.$

(2) 已知 n, β, 求 y; (由(*)反解出 y)

例： 有$n=200$台机床，一小时内，每台机床有70%时间在工作。工作时每台机床耗能$15\mathbf{KW\cdot h}$。各机床工作相互独立。问：至少供电多少，才能有95%的可能性保证正常运转。

解：令Y表示同时工作的机床台数，则$Y\sim B(200,0.7)$.

$E(Y)=np=140;D(Y)=npq=42.$

Y台机床同时工作时的耗能：$15Y(\mathbf{KW\cdot h})$.

设y为供电量，由题$P(15Y\le y)\ge 0.95$.

$\Rightarrow P(Y\le y/15)\ge 0.95 \quad \Rightarrow P(Y\le y/15+0.5)\ge 0.95$ (修正)　　(*)

$$\text{左边}=P\left(\frac{Y-np}{\sqrt{npq}}\le\frac{y/15+0.5-np}{\sqrt{npq}}\right)\approx\Phi\left(\frac{y/15+0.5-140}{\sqrt{42}}\right)$$

由（*），$\Phi\left(\dfrac{y/15+0.5-140}{\sqrt{42}}\right)\ge 0.95.$

查表：$\dfrac{y/15+0.5-140}{\sqrt{42}}\ge 1.645\Rightarrow y\ge 2252.$

注：不修正时：$y\ge 2260.$

第3类问题：

设 $\mu_n \sim B(n, p)\ (0 < p < 1)$，记 $S_n^* = \dfrac{\mu_n - np}{\sqrt{npq}}$，

有 $P\left(S_n^* \le y\right) \approx \Phi(y) = \beta.$ $\qquad (*)$

(3) 已知 y, β，求 n；(由(*)反解出n)

例： 调查电视收视率p，用收看的频率$\overline{p}$估计p. 为保证90%的把握保证$|\overline{p}-p|\le 0.05$, 要调查多少对象？

解：设调查n个对象，记$X_i=\begin{cases}1, & \text{第}i\text{个调查对象收看}\\ 0, & \text{否则。}\end{cases}$

收看总人数：$Y=\sum_{i=1}^{n} X_i \Rightarrow Y \sim B(n,p).\ \ \overline{p}=Y/n,$

由题：$P(|\overline{p}-p|\le 0.05)\ge 0.90$，即$P\left(\left|Y/n-p\right|\le 0.05\right)\ge 0.90$

也即：$P\left(\left|Y-np\right|\le 0.05n\right)\ \ge\ 0.90\ (*)$

$$\text{左边}=P\left(\left|\frac{Y-np}{\sqrt{npq}}\right|\le\frac{0.05n}{\sqrt{npq}}\right)\approx 2\Phi\left(\frac{0.05n}{\sqrt{npq}}\right)-1,$$

此处没有修正

$$\text{由}(*),\ 2\Phi\left(\frac{0.05n}{\sqrt{npq}}\right)-1\ge 0.90\Rightarrow\Phi\left(\frac{0.05n}{\sqrt{npq}}\right)\ge 0.95$$

$$\Rightarrow\frac{0.05n}{\sqrt{npq}}\ge 1.645\Rightarrow n\ge(1.645/0.05)^2\,\underline{pq}\qquad (pq\le 1/4)$$

对所有p成立

$\Rightarrow n\ge 270.6.$ 调查271个对象。

例： 抛均匀硬币，需抛掷多少次，才能保证出现正面的频率在0.4至0.6的概率不小于90%.

解：设抛掷n次，记Y_n为出现正面次数$\Rightarrow Y_n \sim B(n,0.5)$.

由题：$P(0.4 \le Y_n / n \le 0.6) \ \ge \ 0.90, \qquad (*)$

左边$= P(0.4n \le Y_n \le 0.6n)$

$$= P\left(\frac{0.4n - 0.5n}{\sqrt{0.5^2 n}} \le \frac{Y_n - 0.5n}{\sqrt{0.5^2 n}} \le \frac{0.6n - 0.5n}{\sqrt{0.5^2 n}} \right)$$

没有修正

$$= P\left(-0.2\sqrt{n} \le \frac{Y_n - 0.5n}{\sqrt{0.5^2 n}} \le 0.2\sqrt{n} \right) \approx 2\Phi\left(0.2\sqrt{n}\right) - 1,$$

由$(*)$,$2\Phi\left(0.2\sqrt{n}\right) - 1 \ge 0.90 \Rightarrow \Phi\left(0.2\sqrt{n}\right) \ge 0.95$

$$\Rightarrow 0.2\sqrt{n} \ge 1.645 \Rightarrow n \ge 68.$$

本题可尝试用切比雪夫不等式求解

五、独立不同分布下的中心极限定理*

了解

问题：设$\{X_k\}$独立,不同分布，$E(X_k)=\mu_k, D(X_k)=\sigma_k^2$.

记$S_n=\sum_{k=1}^{n}X_k$，则$E(S_n)=\sum_{k=1}^{n}\mu_k$，$D(S_n)=\sum_{k=1}^{n}\sigma_k^2=: B_n^2$.

对S_n进行标准化：

$$S_n^*=\frac{S_n-E(S_n)}{\sqrt{D(S_n)}}=\sum_{k=1}^{n}\frac{X_k-\mu_k}{B_n},\quad (*)$$

此时，$E(S_n^*)=0, D(S_n^*)=1$.

问题：S_n^*的极限分布是什么？

Idea：希望(*)式加项中各项“均匀地小”(对总和的分布没有决定性影响)

即：$\forall\tau>0, P\left(\max_{1\le k\le n}\left|\frac{X_k-\mu_k}{B_n}\right|>\tau\right)\to 0.$

Lindeberg 中心极限定理

设 $\{X_k\}$是独立RV序列, $E(X_k)=\mu_k, D(X_k)=\sigma_k^2. X_k$密度为$f_k(x)$.

记 $S_n=\sum_{k=1}^{n} X_k,\ B_n=\sqrt{D(S_n)},\ S_n^*=\dfrac{\sum_{k=1}^{n} X_k - E(S_n)}{\sqrt{D(S_n)}}=\sum_{k=1}^{n}\dfrac{X_k-\mu_k}{B_n},$

若 $\dfrac{1}{\tau^2 B_n^2}\sum_{i=1}^{n}\int_{|x-\mu_i|>\tau B_n}(x-\mu_i)^2 f_i(x)dx \to 0$ (**Lindeberg条件**)

则 $S_n^* \xrightarrow{L} Z, Z \sim N(0,1),$

即 $\lim\limits_{n\to\infty} P(S_n^* \le x)=\Phi(x)=\dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt,\ \forall x\in R.$

Liapunov 中心极限定理

设 $\{X_k\}$是独立RV序列，$E(X_k)=\mu_k, D(X_k)=\sigma_k^2.$

记 $S_n=\sum_{k=1}^{n} X_k,\ B_n=\sqrt{D(S_n)},\ S_n^*=\sum_{k=1}^{n}\frac{X_k-\mu_k}{B_n},$

若存在$\delta>0,$满足

$$\lim_{n\to+\infty}\frac{1}{B_n^{2+\delta}}\sum_{i=1}^{n} E[|X_i-\mu_i|^{2+\delta}]=0\quad \text{(Liapunov条件)}$$

则 $S_n^*\xrightarrow{L} Z,\ Z\sim N(0,1),$

即 $\lim_{n\to\infty} P(S_n^*\le x)=\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt,\ \forall x\in R.$

(该条件比Lindeberg条件更好验证)

例：考卷有99道题，某学生答对第 k 道题的概率为1−k/100. 该生解答各题相互独立。答对60道题(含)算考试通过。问该生通过考试的概率？

解： 设 $X_k=\begin{cases}1,\text{若学生答对第}k\text{题;}\\0,\text{若学生答错第}k\text{题。}\end{cases}$

则$\{X_k\}$相互独立，服从不同的两点分布：

X_k	0	1
P	k/100	1 - k/100

记S $=\sum_{k=1}^{99} X_k$,要求$P(S\geq 60)$.

可以验证$\{X_k\}$满足**Lyapunov**条件,从而可用中心极限定理。

注意到：$E(S)=E(\sum_{k=1}^{99} X_k)=49.5, D(S)=D(\sum_{k=1}^{99} X_k)=16.665,$

$$\Rightarrow P(S\geq 60)=P\left(\frac{S-49.5}{\sqrt{16.665}}\geq\frac{60-49.5}{\sqrt{16.665}}\right)\approx 1-\Phi\left(2.5735\right)=0.005.$$

验证Liapnov条件

$\{X_k\}$相互独立，服从不同的两点分布：

X_k	0	1
P	**k/100**	**1-k/100**

设想从X_{100}起的RV与X_{99}同分布，且独立。

以下验证$\delta=1$的Lyapunov条件成立。

$$B_n^2=\sum_{k=1}^{n} D(X_k)=\sum_{k=1}^{n} p_k(1-p_k)\to\infty \quad (p_k=1\text{-}k/100)$$

$$E(|X_k-p_k|^3)=(1-p_k)^3p_k+p_k^3(1-p_k)$$

$$=p_k(1-p_k)[(1-p_k)^2+p_k^2]\le p_k(1-p_k).$$

$$\frac{1}{B_n^3}\sum_{k=1}^{n} E(|X_k-p_k|^3)=\frac{1}{B_n}\cdot\frac{1}{B_n^2}\sum_{k=1}^{n} E(|X_k-p_k|^3)\le\frac{1}{B_n}\to 0.$$

满足Lyapunov条件,从而可用中心极限定理。

Summary 极限定理

	Weak LLN	Strong LLN	CLT
Bernoulli 试验序列	Bernoulli	Borel	De Moivre- Laplace
独立同分布	辛钦	Kolmogorov	Lindeberg-Levy
(独立)不同分布	切比雪夫 Markov	Kolmogorov	Lindeberg- Feller Lyapunov

Strong LLN 不要求

The End of Chapter 4