西安电子科技大学
XIDIAN UNIVERSITY

第5章 大数定律及中心极限定理

概率论与数理统计课程组

CHAPTER 5

大数定律及中心极限定理

由概率论通向数理统计的理论基础

CHAPTER 5

大数定律及中心极限定理

5.1 Chebyshev（切比雪夫）不等式

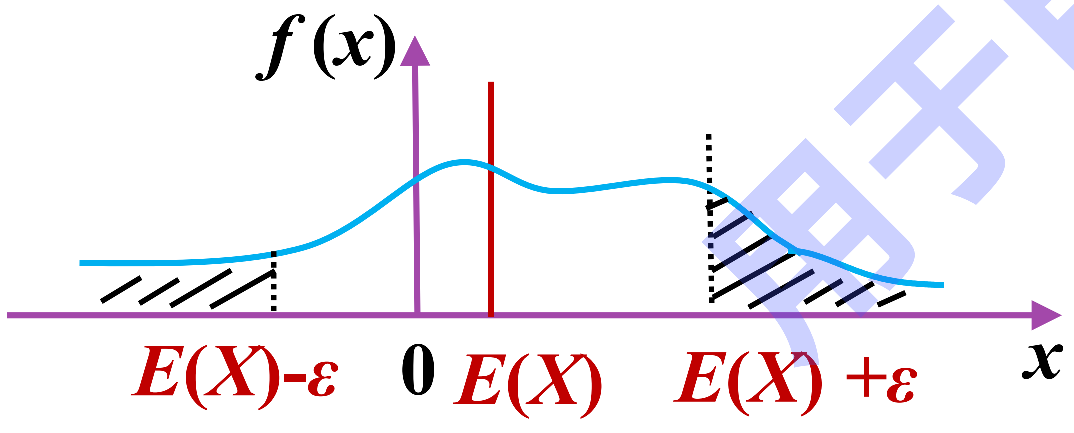
Chebyshev（切比雪夫）不等式

定理

设随机变量X的数学期望$E(X)$、方差$D(X)$存在，则对于任意$\varepsilon>0$，不等式

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

成立，称为**切比雪夫(*Chebyshev*)不等式**



证明

仅对连续型随机变量给予证明

设X的概率密度为$f(x)$，则有

$$P\left(|X-E(X)|\geq\varepsilon\right)=\int_{|x-E(X)|\geq\varepsilon}f(x)dx\leq\int_{|x-E(X)|\geq\varepsilon}\overbrace{\boxed{\frac{|x-E(X)|^2}{\varepsilon^2}}}^{\geq 1}f(x)dx$$

$$\leq\frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}[x-E(X)]^2f(x)dx=\frac{D(X)}{\varepsilon^2}$$

等价形式为

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

Chebyshev（切比雪夫）不等式

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

$$P\left(\left|X-E(X)\right|<\varepsilon\right)\geq1-\frac{D(X)}{\varepsilon^2}$$

说明：

1. 注意两式中≤、≥准确写法，逆事件和不等式变形。
2. 在分布未知，而期望$E(X)$和方差$D(X)$已知的情况下，可以估计概率$P(|X-E(X)|\geq\varepsilon)$或$P(|X-E(X)|<\varepsilon)$的界限。
3. 可用于证明方差性质5。
4. 可用于证明大数定理。

例 已知正常男性成年人的血液中，平均每毫升含白细胞数是 7300，标准差是700，试估计每毫升男性成年人血液中含白细胞数在 5200 至 9400 之间的概率。

解 设每毫升男性成年人血液中所含白细胞数为X，则 $E(X)=7300$，$D(X)=700^2$，从而

$$P(5200 < X < 9400) = P(5200-7300 < X-7300 < 9400-7300)$$

$$= P(-2100 < X-7300 < 2100) = P(|X-7300| < 2100)$$

由*Chebyshev*不等式得

$$P(5200 < X < 9400) = P(|X-7300| < 2100) = P(|X-E(X)| < 2100) \geq 1-\frac{D(X)}{2100^2} = 1-\frac{700^2}{2100^2} = \frac{8}{9}$$

即每毫升男性成年人血液中所含白细胞数在5200至9400之间的概率不小于 $\frac{8}{9}$

例 在n重*Bernoulli*试验中，若已知每次试验事件A出现的概率为0.75，试利用*Chebyshev*不等式估计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 *Bernoulli*试验中事件A出现次数为X，则有$X\sim B(n,0.75)$

$$E(X)=np=0.75n,\quad D(X)=npq=0.1875n$$

$$f_n(A)=\frac{X}{n}$$

$$P\left(0.74<\frac{X}{n}<0.76\right)=P\left(|X-0.75n|<0.01n\right)\geq 1-\frac{0.1875n}{(0.01n)^2}=1-\frac{1875}{n}\geq 0.90$$

$$\Rightarrow n\geq 18750$$

*Chebyshev*不等式可在只知$E(X)$和$D(X)$前提下（分布未知）估计随机变量X落在均值左右ε界限内的概率

本节回顾

Chebyshev（切比雪夫）不等式

设随机变量X的数学期望$E(X)$、方差$D(X)$存在，则对于任意$\varepsilon>0$，以下不等式成立

$$P(|X-E(X)|\geq\varepsilon)\leq\frac{D(X)}{\varepsilon^2}\qquad P(|X-E(X)|<\varepsilon)\geq1-\frac{D(X)}{\varepsilon^2}$$

1. 注意两式中≤、≥准确写法，逆事件和不等式变形。
2. 在分布未知，而期望$E(X)$和方差$D(X)$已知的情况下，可以估计概率$P(|X-E(X)|\geq\varepsilon)$或$P(|X-E(X)|<\varepsilon)$的界限。

CHAPTER 5

大数定律及中心极限定理

5.2 大数定律

大数定律是叙述随机变量序列的前若干项的算数平均值在某种条件下收敛于这些项均值的算数平均值

定义1

设$X_1, X_2, \ldots$是随机变量序列，如果存在数列$a_1, a_2, \ldots$使得对任意的$\varepsilon>0$，有

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

则称随机变量序列$\{X_i\}$服从大数定律。

定义2

设$X_1, X_2, \ldots$是随机变量序列，X是随机变量，若对任意的$\varepsilon>0$，有

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$$

则称随机变量序列$\{X_i\}$依概率收敛于X，记为 $X_n \xrightarrow{P} X, n \to \infty$

Chebyshev（切比雪夫）大数定律

设$X_1, X_2, \ldots$ 是**相互独立**的随机变量序列，具有相同的期望和方差，数学期望$E(X_k)=\mu$，方差$D(X_k)=\sigma^2(k=1, 2, \ldots)$，作前$n$个变量的算数平均 $\frac{1}{n}\sum_{k=1}^{n} X_k$

则对于任意$\varepsilon>0$，有 $\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k-\mu\right|\ge\varepsilon\right)=0$ 或 $\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k-\mu\right|<\varepsilon\right)=1$

该表述形式是*Chebyshev*不等式的等价形式

也可记为依概率收敛于μ的形式，即 $\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu, \quad n\to\infty$

证明

在方差$D(X_k)=\sigma^2$，$k=1, 2, \dots$存在的情况下证明

$$E\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{1}{n}\sum_{k=1}^{n}E(X_k)=\frac{1}{n}\cdot n\mu=\mu$$

由独立性 $$D\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}D(X_k)=\frac{1}{n^2}\cdot n\sigma^2=\frac{\sigma^2}{n}$$

由*Chebyshev*不等式

$$0\le P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|\ge\varepsilon\right)=P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-E\left[\frac{1}{n}\sum_{k=1}^{n}X_k\right]\right|\ge\varepsilon\right)$$

$$\le\frac{D\left[\frac{1}{n}\sum_{k=1}^{n}X_k\right]}{\varepsilon^2}=\frac{\sigma^2/n}{\varepsilon^2}\to 0 \qquad n\to\infty$$

由此可证 $\lim\limits_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|\ge\varepsilon\right)=0$ ■

说明：$\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}$ 是一个随机事件

而当$n\to\infty$时事件的概率趋近于1

对于相互独立且具有相同均值μ和相同方差σ^2的随机变量$X_1, X_2, \dots$，当n很大时它们算数平均 $\frac{1}{n}\sum_{k=1}^{n}X_k$ 很可能接近于μ

注意：不能简单理解为 $E(\frac{1}{n}\sum_{k=1}^{n}X_k)$ 接近为μ

要讨论的是一个随机变量，不是随机变量的期望值

*Chebyshev*大数定律（更一般的形式）

设$X_1, X_2, \ldots$ 是相互独立的随机变量序列，具有相同的期望，即数学期望$E(X_k)=\mu$(k =1, 2, …)，若存在常数C>0，使得$D(X_k)\leq C$ (k =1, 2, …)

则对于任意$\varepsilon>0$，有 $$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| \geq \varepsilon\right) = 0$$

其依概率收敛于μ的形式为 $$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu, \quad n\to\infty$$

Markov（马尔可夫）大数定律

虽各项期望方差不同，但仍可用*Chebyshev*不等式证明

设$X_1, X_2, \ldots$ 是随机变量序列，且 $\lim_{n\to\infty}\frac{1}{n^2} D\left[\sum_{k=1}^{n} X_k\right] = 0$

则对于任意$\varepsilon>0$，有 $$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| \geq \varepsilon\right) = 0$$ 即 $$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \frac{1}{n}\sum_{k=1}^{n} E(X_k),\ n\to\infty$$

例 设$X_1, X_2, \dots$， X_n是相互独立的随机变量序列，且X_i (i=1, 2, ...)有如下分布律，

$$P(X_i=-ia)=\frac{1}{2i^2}, P(X_i=0)=1-\frac{1}{i^2}, P(X_i=ia)=\frac{1}{2i^2}$$

问$X_1, X_2, \dots$， X_n是否满足*Chebyshev*大数定律。

解 由题意知， $X_1, X_2, \dots$， X_n相互独立

$$E(X_i)=-ia\times\frac{1}{2i^2}+0\times(1-\frac{1}{i^2})+ia\times\frac{1}{2i^2}=0, i=1,2,\cdots$$

$$E(X_i^2)=(-ia)^2\times\frac{1}{2i^2}+0^2\times(1-\frac{1}{i^2})+(ia)^2\times\frac{1}{2i^2}=a^2, i=1,2,\cdots$$

$$D(X_i)=E(X_i^2)-\left[E(X_i)\right]^2=a^2, i=1,2,\cdots$$

故$X_1, X_2, \dots$， X_n具有相同的数学期望的相同的方差，且方差有界，所以$X_1, X_2, \dots$， X_n满足*Chebyshev*大数定律

以上大数定律说明在方差满足一定的条件下，随机变量序列服从大数定律。但在方差不存在时，随机变量序列是否服从大数定律呢? 由 *Khintchine* 大数定律回答这个问题。

Khintchine（辛钦）大数定律

设$X_1, X_2, \ldots$ 是相互独立且服从同一分布的随机变量序列，具有有限的数学期望，即数学期望 $E(X_k)=\mu(k=1, 2, \ldots)$，

则对于任意$\varepsilon>0$，有
$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k-\mu\right|\geq\varepsilon\right)=0$$

其依概率收敛于μ的形式为
$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu, \qquad n\to\infty$$

注意：“同分布”未必要求每一项的分布参数一模一样

上述大数定律的成立条件不同，但是均表明了当n充分大时，随机变量序列的前n项的算数平均会收敛到这些项期望的算术平均值

*Bernoulli*大数定律是对第1章中提出的“频率稳定性”，给出理论上的论证

Bernoulli（伯努利）大数定律

设n_A是n次独立重复试验（n重*Bernoulli*试验）事件A发生的次数，p是事件A在每次试验中发生的概率，即$P(A)=p$，则对任意的$\varepsilon>0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n}-p\right|\ge\varepsilon\right)=0 \quad 即 \quad \frac{n_A}{n}\xrightarrow{P} p,\quad n\to\infty$$

证明 $n_A \sim B(n,p)$ $\quad E\left(\frac{n_A}{n}\right)=\frac{1}{n}E(n_A)=\frac{1}{n}\cdot np=p \quad D\left(\frac{n_A}{n}\right)=\frac{1}{n^2}D(n_A)=\frac{1}{n^2}\cdot npq=\frac{p(1-p)}{n}$

由*Chebyshev*不等式，对任意$\varepsilon>0$ $\quad P\left(\left|\frac{n_A}{n}-p\right|\ge\varepsilon\right)\le\frac{p(1-p)}{n\varepsilon^2} \quad \lim_{n\to\infty} P\left(\left|\frac{n_A}{n}-p\right|\ge\varepsilon\right)=0$

也可直接用*Chebyshev*大数定律，$X_1, X_2, \ldots$ 是相互独立的随机变量，$E(X_i)=p,\ D(X_i)=p(1-p)$

$$n_A=\sum_{k=1}^{n}X_k \qquad \frac{n_A}{n}=\frac{1}{n}\sum_{k=1}^{n}X_k \qquad \lim_{n\to\infty} P\left(\left|\frac{n_A}{n}-p\right|\ge\varepsilon\right)=\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-p\right|\ge\varepsilon\right)=0 \quad \blacksquare$$

说明：

1. *Bernoulli*大数定律建立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。

2. *Bernoulli*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$f_n(A)$与概率p有较大偏差的可能性很小，便可通过多次试验确定某事件发生的频率并把它作为相应的概率估计。

3. 大数定律是参数估计的理论依据（采用大量的样本观测值估计分布中的未知参数是可行的）

本节回顾

$$\frac{1}{n}\sum_{k=1}^{n}X_k \xrightarrow{P} \mu, \quad n\to\infty$$

- **Chebyshev大数定律**

设$X_1, X_2, \ldots$是相互独立的随机变量序列，具有相同的期望，即数学期望$E(X_k)=\mu$(k =1, 2, …)，若存在常数C>0，使得$D(X_k)\le C$ (k =1, 2, …)

- **Markov（马尔可夫）大数定律**

设$X_1, X_2, \ldots$是随机变量序列，且 $\lim\limits_{n\to\infty}\frac{1}{n^2}D\left[\sum_{k=1}^{n}X_k\right]=0$

- **Khintchine（辛钦）大数定律**

设$X_1, X_2, \ldots$是相互独立且服从同一分布的随机变量序列，具有有限的数学期望，即数学期望$E(X_k)=\mu$(k =1, 2, …)，

- **Bernoulli（伯努利）大数定律**

设n_A是n次独立重复试验（n重Bernoulli试验）事件A发生的次数，p是事件A在每次试验中发生的概率，即$P(A)=p$，则对任意的ε>0，有 $\frac{n_A}{n}\xrightarrow{P}p, \quad n\to\infty$

CHAPTER 5

大数定律及中心极限定理

5.3 中心极限定理

有许多随机变量，它们是由大量的相互独立的随机变量的综合影响所形成的，而其中每个个别因素对总的影响的作用都很小，这种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。

中心极限定理

本章介绍三个常用的中心极限定理

独立同分布中心极限定理

又称为Lindeberg-Levy（林德伯格–勒维）中心极限定理

设$X_1, X_2, \ldots, X_n, \ldots$是**相互独立且同分布**的随机变量序列，且具有相同的数学期望$E(X_k)=\mu$，与相同的方差$D(X_k)=\sigma^2>0\ (k=1, 2, \ldots)$，

则随机变量之和 $\sum_{k=1}^{n} X_k$ 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$F(x)$满足，对任意的x

$$\lim_{n\to\infty} F_n(x)=\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}\le x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}=\Phi(x)$$

数理统计中抽样理论的基础

该定理在实际中有着广泛的应用。当n充分大时，独立同分布且具有相同均值μ，方差$\sigma^2>0$的随机变量$X_1, X_2, \dots X_n$之和$\sum_{k=1}^{n} X_k$近似服从以它的均值为均值，它的方差为方差的正态分布，即

当n充分大时，有 $\sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(n\mu, n\sigma^2)$ 或者 $\dfrac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{\text{近似地}}{\sim} N(0,1)$

因此对于任意的实数x及$a<b$，有

$$P\left(\sum_{k=1}^{n} X_k \le x\right) \approx \Phi\left(\frac{x-n\mu}{\sqrt{n}\sigma}\right) \qquad P\left(a<\sum_{k=1}^{n} X_k \le b\right) \approx \Phi\left(\frac{b-n\mu}{\sqrt{n}\sigma}\right)-\Phi\left(\frac{a-n\mu}{\sqrt{n}\sigma}\right)$$

记 $\overline{X}=\frac{1}{n}\sum_{k=1}^{n} X_k$ 则上述结果变为

当n充分大时，有 $\overline{X} \overset{\text{近似地}}{\sim} N(\mu, \sigma^2/n)$ 或者 $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \overset{\text{近似地}}{\sim} N(0,1)$

例 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 设第 i 个元件的寿命为X_i，有 $E(X_i)=100$，$D(X_i)=100^2$

16个元件的寿命总和为X，$X=\sum_{i=1}^{16} X_i$

根据独立同分布中心极限定理：

$$Y=\frac{\sum_{i=1}^{16} X_i-16\times 100}{4\times 100}=\frac{X-1600}{400} \text{ 近似服从} N(0,1)$$

$$P(X>1920)=1-P(X\le 1920)\approx 1-\Phi\left(\frac{1920-1600}{400}\right)=1-\Phi(0.8)=0.2119$$

不能认为 $X=16X_1$ 将问题转为$X_1>1920/16$

例 测量一个物理量a共n次，每次测量的误差记为X_i (i=1, 2, …, n)，假设在适当选取单位的条件下X_i (i=1, 2, …, n)在区间(−0.5，0.5)上服从均匀分布，取 n次测量结果的算术平均值作为 a 的估计值。试求：

(i)它与真值a之差的绝对值小于预先指定的正数ε的概率；

(ii)若要估计值与真值a之差的绝对值小于0.1的概率不小于0.95，至少需要进行多少次测量?

解 (i) 设X表示n次测量结果的算数平均值，由于各次测量结果为 $a+X_i$ (i=1, 2, …, n)，故

$$X = a + \frac{1}{n}\sum_{i=1}^{n} X_i$$

又X_i (i=1, 2, …, n)在区间(−0.5，0.5)上服从均匀分布 $E(X_i)=0,\ D(X_i)=\dfrac{1}{12}$

根据独立同分布中心极限定理： $\sum_{i=1}^{n} X_i$ 近似服从$N(0, n/12)$ $\dfrac{\sum_{i=1}^{n} X_i - 0}{\sqrt{n/12}}$ 近似服从$N(0, 1)$

$$P(|X-a|<\varepsilon)=P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right|<\varepsilon\right)=P\left(\left|\sum_{i=1}^{n}X_i\right|<n\varepsilon\right)\approx\Phi\left(\frac{n\varepsilon}{\sqrt{n/12}}\right)-\Phi\left(\frac{-n\varepsilon}{\sqrt{n/12}}\right)$$

$$=2\Phi\left(\sqrt{12n}\varepsilon\right)-1$$

(ii) 取$\varepsilon=0.1$，可知$2\Phi\left(0.01\sqrt{12n}\right)-1\geq 0.95$

$$2\Phi\left(0.1\sqrt{12n}\right)-1\geq\frac{0.95+1}{2}=0.975$$

查表得 $0.1\sqrt{12n}\geq 1.96$

从而 $n\geq\dfrac{1.96^2}{12\times 0.1^2}\approx 32.01$

例 检查员逐个地检查某种产品，每次需花 10 秒钟检查一个产品，但也可能有的产品需要重复检查一次再用去 10 秒钟。假定每个产品需要重复检查的概率为 1 / 2 ，求在8 小时内检查员检查的产品不少于 1900 个的概率。

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

解 设X_i (i=1, 2, …, 1900) 表示检查第i个产品所需要花费的时间，则$X_1, X_2, \dots, X_{1900}$相互独立同分布

$\sum_{i=1}^{1900} X_i$ 表示检查1900个产品所需花费的总时间，由题设知

$$X_i=\begin{cases}10, & \text{第}i\text{个产品没有重复检查}\\ 20, & \text{第}i\text{个产品重复检查}\end{cases}\qquad P(X_i=10)=P(X_i=20)=\frac{1}{2},\ i=1,2,\dots,1900$$

$$E(X_i)=15,\ D(X_i)=E(X_i^2)-\left[E(X_i)\right]^2=25$$

根据独立同分布中心极限定理： $\sum_{i=1}^{1900} X_i$ 近似服从$N(1900\times5, 1900\times25)$

故所求的概率为 $P\left(\sum_{i=1}^{1900} X_i\le 8\times3600\right)\approx\Phi\left(\frac{28800-28500}{\sqrt{47500}}\right)=\Phi\left(\frac{6}{\sqrt{19}}\right)=0.9162$

Lyapunov（李雅普诺夫）中心极限定理

设随机变量$X_1, X_2, \ldots, X_n, \ldots$相互独立，其具有如下数学期望和方差：

$E(X_k)=\mu_k, D(X_k)=\sigma_k^2>0\ (k=1, 2, \ldots)$，记 $B_n^2=\sum_{k=1}^{n}\sigma_k^2$

若存在正数δ使得当$n\to\infty$时

则 的标准化变量 的分布函数$F(x)$满足，对任意的x有

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P\left(\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}\mu_k}{B_n}\le x\right)=\Phi(x)$$

很多问题随机变量Y可表示为大量独立随机变量（不一定是同分布，也不一定正态）的和，则Y近似正态分布

De Moivre-Laplace（棣莫弗-拉普拉斯）中心极限定理

独立同分布中心极限定理的特殊情况

设随机变量n_A表示n重*Bernoulli*试验中事件A发生的次数，p是A在一次试验发生的概率，$P(A)=p$，即n_A服从参数n, p $(0<p<1)$的二项分布，则对于任意的x，随机变量

的分布函数$F(x)$满足

即，若$X\sim B(n, p)$，则当n充分大时

证明思路：二项分布可分解为若干个独立同分布的0-1分布之和，即二项分布X有其中X_k $(k=1, 2, \ldots)$服从0-1分布，$X_k\sim B(1, p)$。然后再由独立同分布中心极限定理推出

说明：正态分布是二项分布的极限分布

例 某车间有200台车床，由于各种原因每台车床只有60%的时间在开动，每台车床开动期间耗电量为1单位，问至少供给此车间多少电量才能以不少于99.9%的概率保证此车间不因供电不足而影响生产。

解 设不影响生产需要开动的车床数为n，X表示200台车床中开动的车床数，则$X \sim B(200, 0.6)$

由*De Moivre-Laplace*中心极限定理，X近似地服从正态分布$N\,(120，48)$，所以n取决于如下条件:

查表得

例 对于一个学生而言，来参加家长会的家长人数是一个随机变量，设一个学生无家长、有 1 名家长与有2名家长来参加会议的概率分别为0.05， 0.80与0.15，若学校共有400名学生，设各学生参加会议的家长人数相互独立，且服从同一分布。

(i)求参加会议的家长人数X超过450的概率;

(ii)求有1名家长来参加会议的学生人数不多于 340的概率($\Phi(1.147)=0.8749$，$\Phi(2.5)=0.9938$)。

解 (i) 设X_i (i=1, 2, …, 400)表示第i个学生来参加会议的家长人数，X_i 分布律为

X_i	0	1	2
P	0.05	0.80	0.15

由独立同分布中心极限定理， 近似地服从正态分布$N\,(400\times 1.1，400\times 0.19)$，

所求概率为:

例 对于一个学生而言，来参加家长会的家长人数是一个随机变量，设一个学生无家长、有 1 名家长与有2名家长来参加会议的概率分别为0.05， 0.80与0.15，若学校共有400名学生，设各学生参加会议的家长人数相互独立，且服从同一分布。

(i)求参加会议的家长人数X超过450的概率;

(ii)求有1名家长来参加会议的学生人数不多于 340的概率($\Phi(1.147)=0.8749$，$\Phi(2.5)=0.9938$)。

解 (ii) 设Y表示有1名家长参加会议的学生人数，则$Y\sim B(400, 0.80)$

由*De Moivre-Laplace*中心极限定理，Y近似地服从正态分布$N(320，64)$，所求概率为:

例 某单位有200部电话,每部电话机大约有5%的时间要使用外线通话。若每部电话机是否使用外线是相互独立的，问该单位总机至少需要安装多少条外线，才能以90%以上的概率保证每部电话机需要外线时不需要等待($\Phi(1.29)=0.90$)。

解 设X表示同一时刻200部电话机中需要使用外线的部数，则$X\sim B(200, 0.05)$

由*De Moivre-Laplace*中心极限定理， X近似地服从正态分布$N(10，9.5)$，所以单位装的外线条数m取决于如下条件:

查表得

故该单位总机至少需要安装14 条外线才能以90%以上的概率保证每部电话机需要外线时不需要等待

例 一台总机共有300台分机，总机拥有 20条外线，假设每台分机向总机要外线的概率为 5%，试求：

(i) 每台分机向总机要外线时，能及时得到满足的概率；

(ii) 同时向总机要外线的最可能台数的概率($\Phi(1.32)=0.9066$，$\Phi(0.13)=0.5517$)。

解 (i) 设X 表示同时向总机要外线的分机台数，则 $X \sim B(300, 0.05)$

从而 $E(X)=300\times0.05=15$，$D(X)=300\times0.05\times0.95=14.25$

由*De Moivre-Laplace*中心极限定理，X近似地服从正态分布$N(15, 14.25)$，所求概率为

例 一台总机共有300台分机，总机拥有 20条外线，假设每台分机向总机要外线的概率为 5%，试求：

(i) 每台分机向总机要外线时，能及时得到满足的概率；

(ii) 同时向总机要外线的最可能台数的概率(Φ (1.32)=0.9066，Φ(0.13)=0.5517)。

解 (ii) 题设的问题可视为n重*Bernoulli*试验，则最可能的台数k为，

二项分布最大项的确定可用第k项与第k-1项的比值

所以最可能台数为[(300+1) ×0.05]=[15.05]=15，其概率为

[ ]表示向下取整

离散型二项分布的数值范围（0, 1， …, n）与连续型正态分布（-∞, ∞）不同，近似时应注意

例 某保险公司的重疾保险有1万人参加，每人每年交200元，若被保险人在该年内罹患条款中的疾病，公司付给受益人1万元。设患疾病率为0.017，试求保险公司在一年内这项保险亏本的概率。

解 设X为一年中投保人的患病数，则$X \sim B(10000, 0.017)$

由*De Moivre-Laplace*中心极限定理，亏本概率为

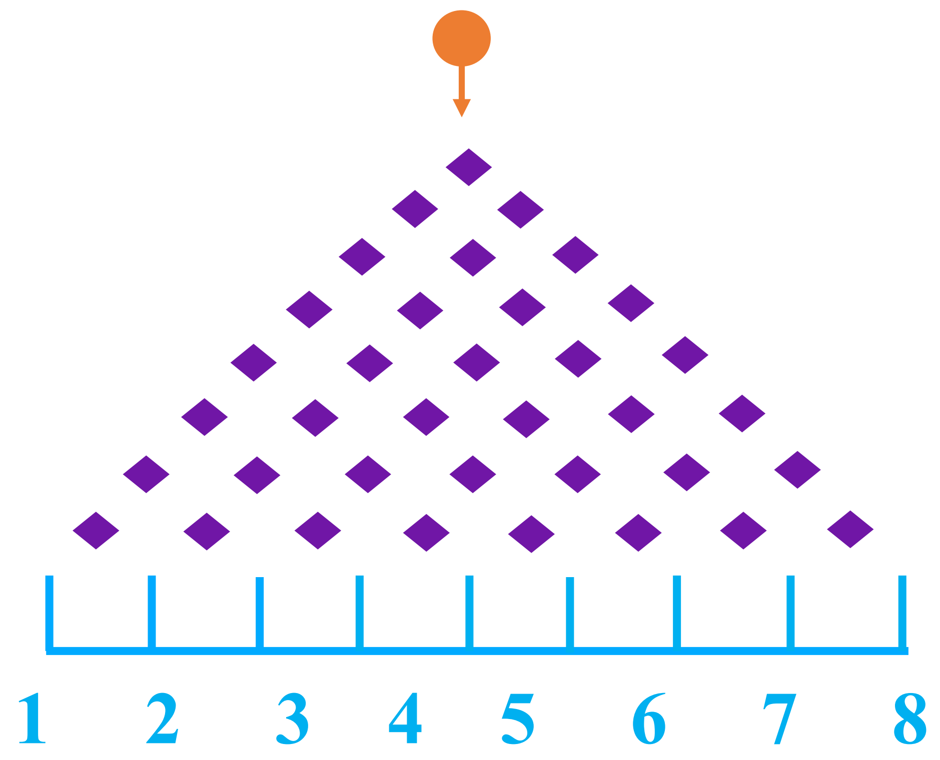
思考：求保险公司至少盈利10万元的概率

实例：伽尔顿板

下面是等间隔的竖槽，中间是规则排列的横杆，上面容器中装有大量颗粒。大量颗粒从上向下泻下来，与各级横杆碰撞后落入槽中。模拟颗粒运动轨迹，统计各槽中的颗粒数及占总粒子数比例。

假设一个粒子落下来都会与每层中的一个横杆发生碰撞，碰撞后向左和向右运动的概率相等。

共n个竖槽，每个粒子落在第m个槽中的概率为



说明

- ✓ 大量试验的平均值就趋于理论值。
- ✓ 每次试验的结果不同，但是每次试验值都在二项分布的理论值附近波动（涨落）。

二项分布在n很大时可用正态分布和泊松分布（p很小）来近似

泊松分布$np_n = \lambda$

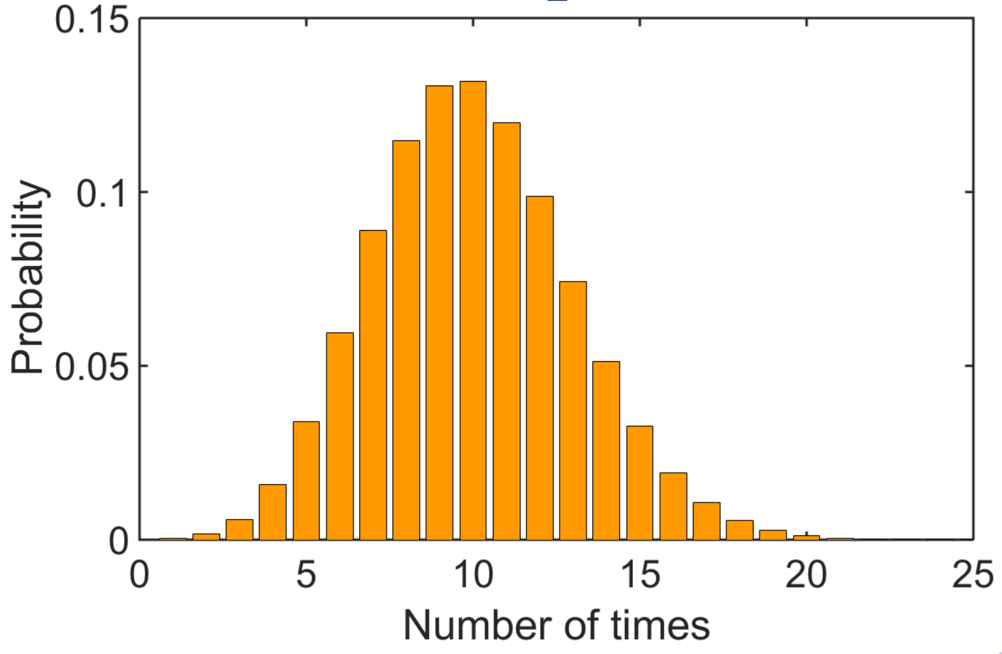
正态分布 $N(np, np(1-p))$

p值不够小！

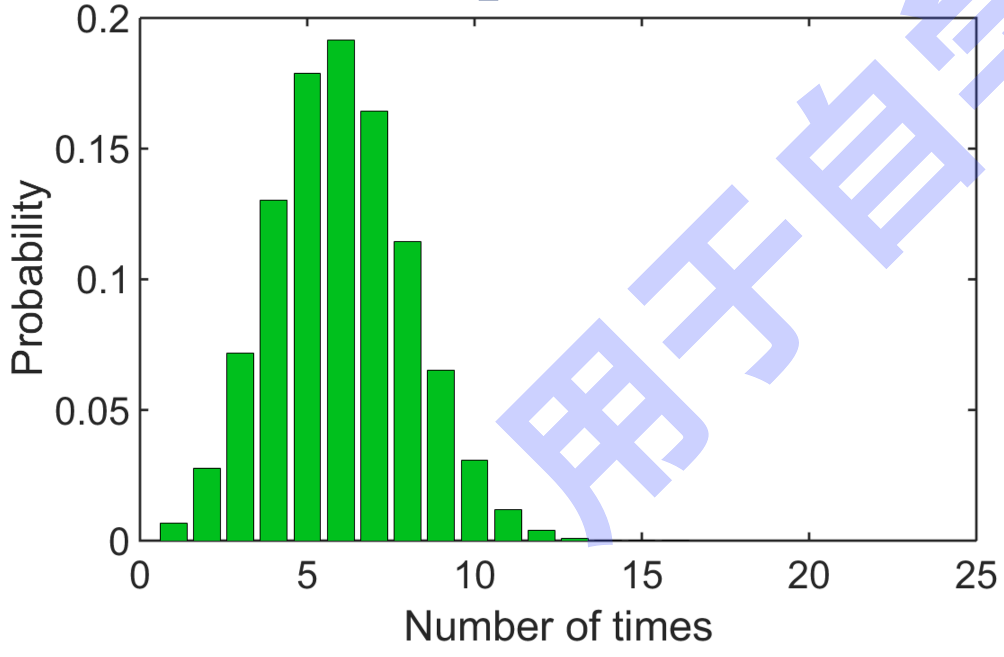
一般地$n \geq 20, p \leq 0.05$

二项分布

n=100, p=0.1

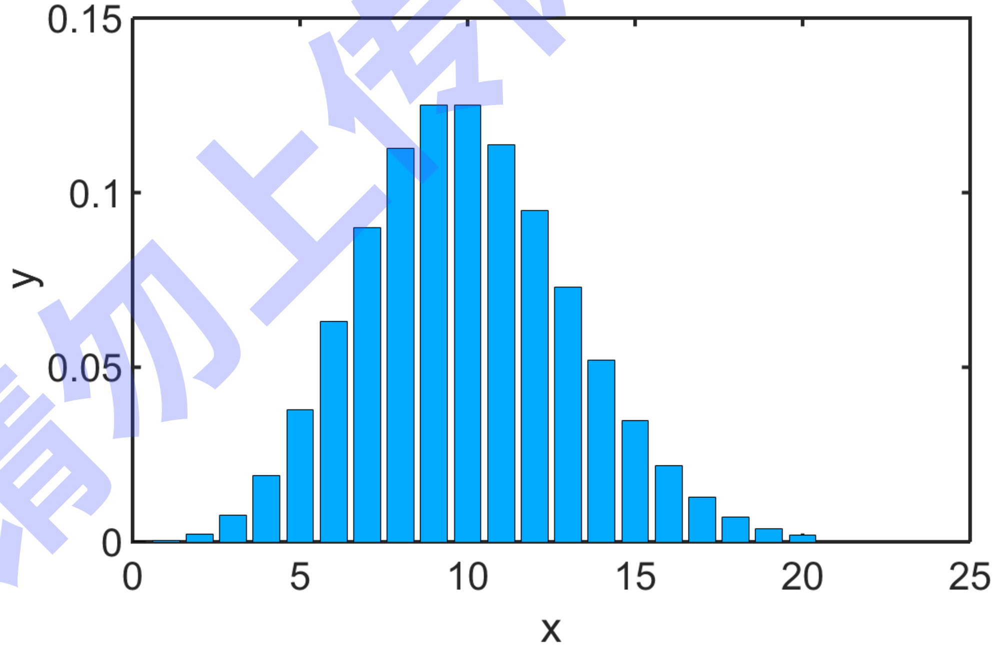


n=20, p=0.3

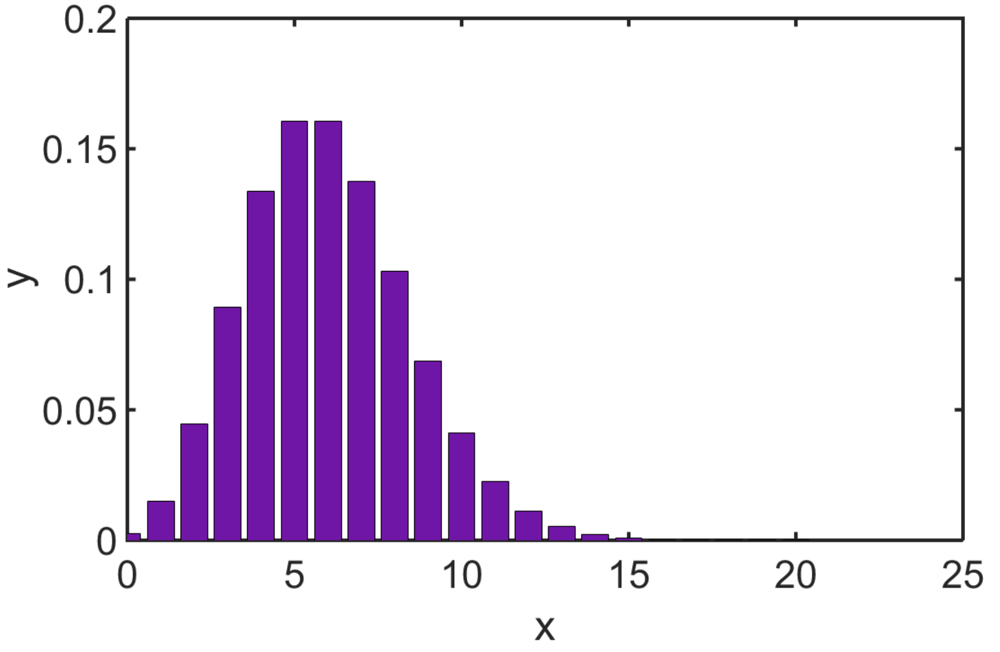


泊松分布

λ=10

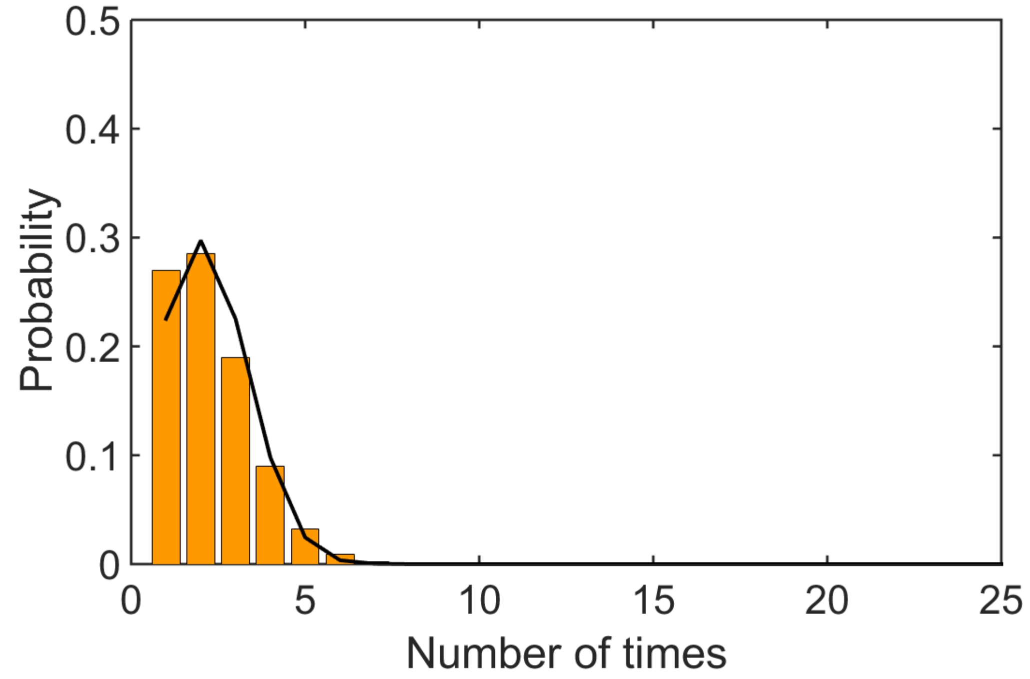


λ=6

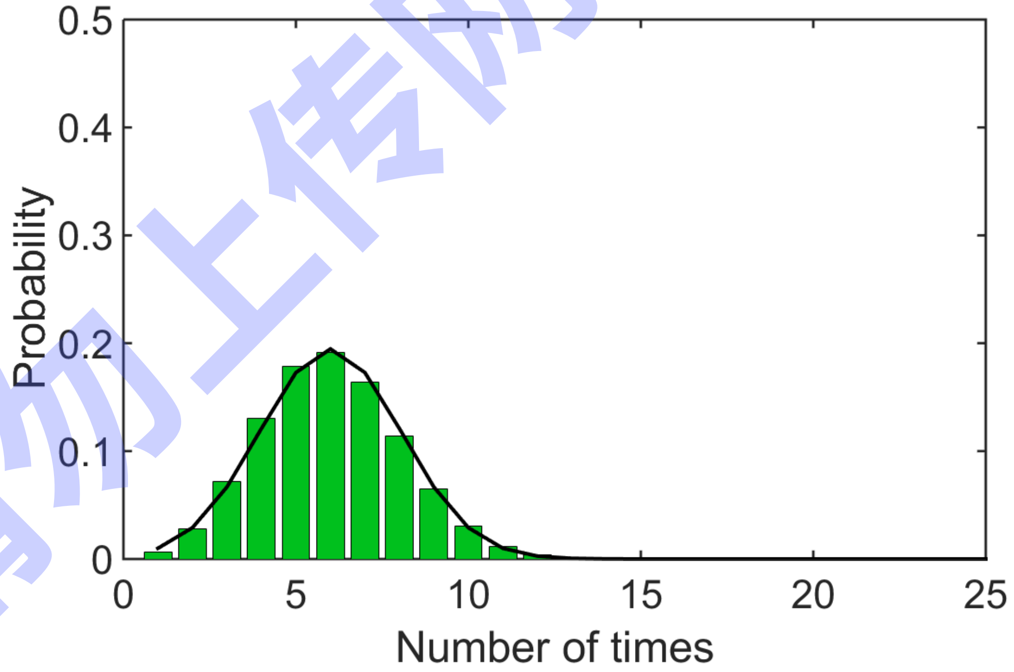


二项分布与正态分布

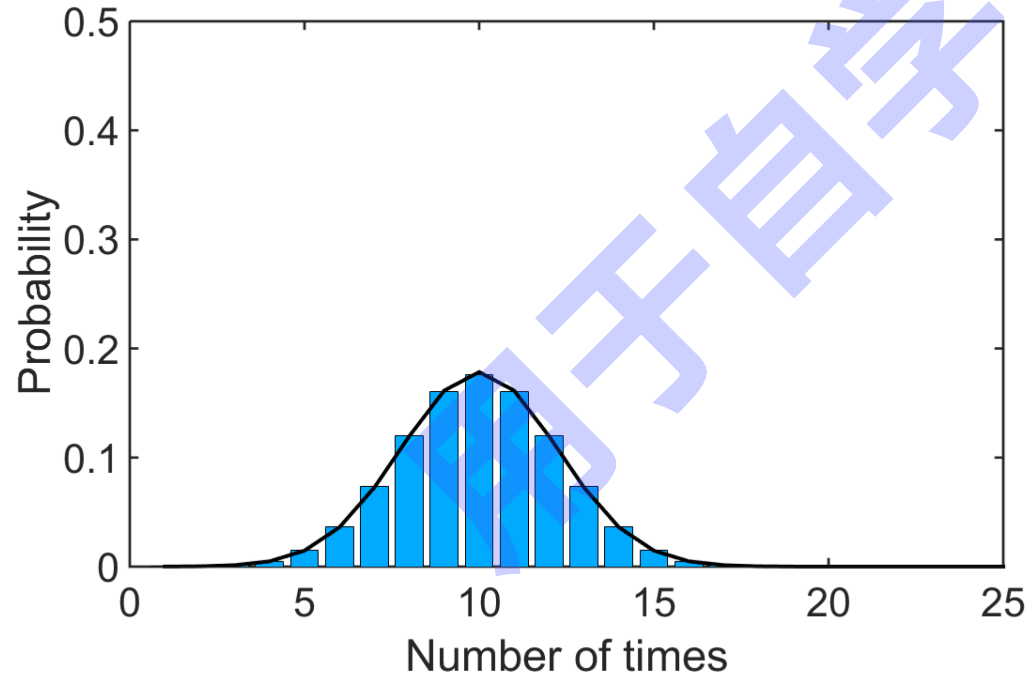
***n*=20, *p*=0.1**



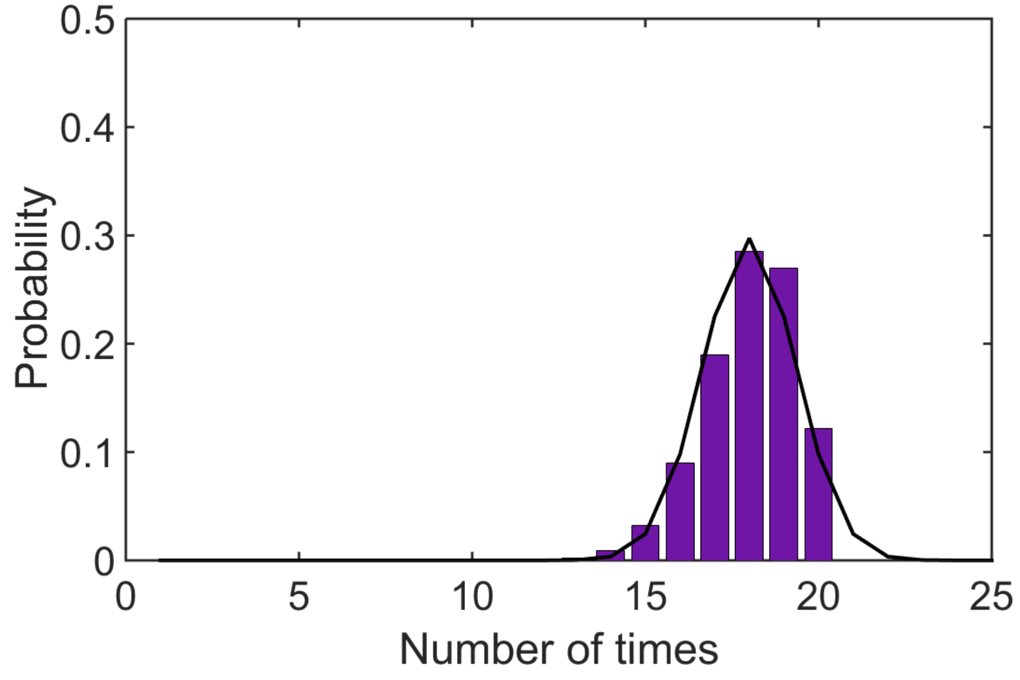
***n*=20, *p*=0.3**



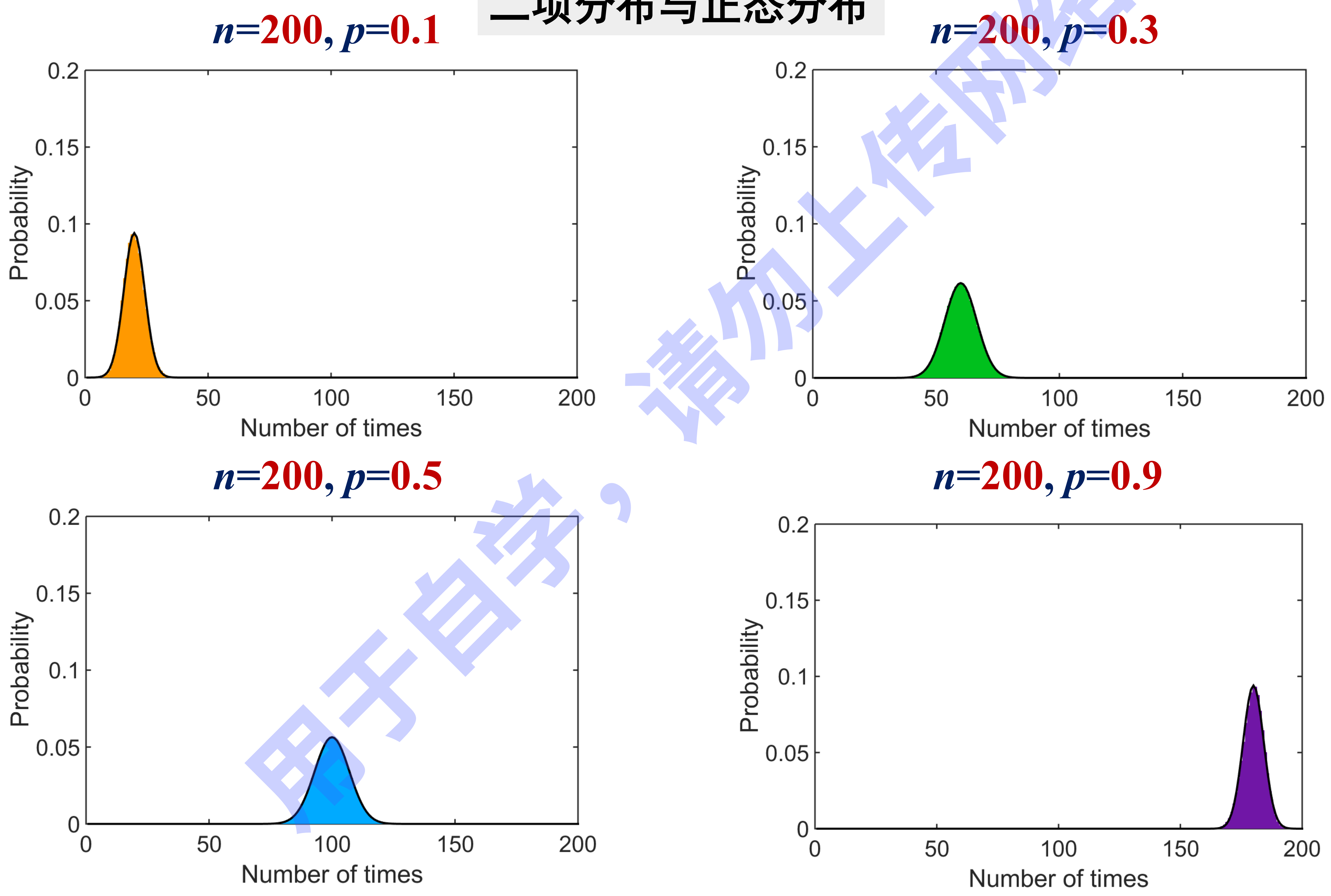
***n*=20, *p*=0.5**



***n*=20, *p*=0.9**



二项分布与正态分布


n=200, p=0.1
n=200, p=0.3
n=200, p=0.5
n=200, p=0.9
Probability
Number of times
0
0.05
0.1
0.15
0.2
0
50
100
150
200

例 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X为机器故障台数，$X \sim B(400, 0.02)$，三种方法求解

(i) 二项分布

(ii) 泊松分布近似

(iii) 正态分布近似

本节回顾

- 独立同分布中心极限定理

当n充分大时，有　　　　或者

- Lyapunov（李雅普诺夫）中心极限定理

- De Moivre-Laplace（棣莫弗-拉普拉斯）中心极限定理

若$X \sim B(n, p)$，则当n充分大时