西安電子科技大學
XIDIAN UNIVERSITY

第2章 随机变量及其分布

概率论与数理统计课程组

CHAPTER 2

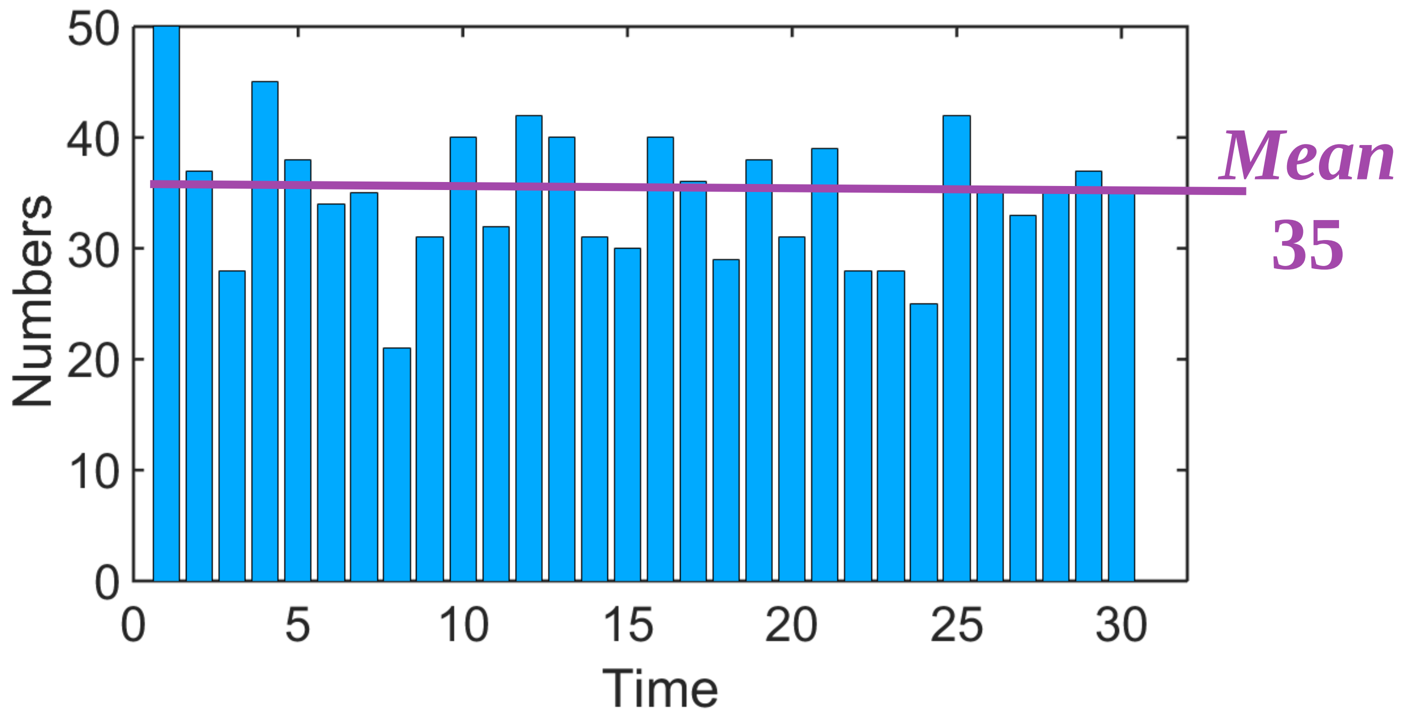
随机变量及其分布

2.3 离散型随机变量及其分布律

自然现象和社会生活中很多随机现象服从某些特定规律

如下例：卖早点的大叔应准备多少包子合适呢？

取一段时间的销量平均值固然是一种方法



定义

有些随机变量，它全部可能取到的值是**有限个或可列无限个**，称为离散型随机变量

掌握离散型随机变量X的统计规律，必须且只需知道X的所有可能取值及取每个值的概率

离散型随机变量的分布律

设X的所有可能取值为x_k $(k = 1, 2, 3, \ldots)$，每个可能取值的概率即事件$\{X=x_k\}$的概率，记为

(1) 解析式法 $P(X=x_k)=P(e: X(e)=x_k)=p_k$ $(k = 1, 2, 3, \ldots)$ 称为**分布律**

p_k 满足两个条件：

1° $p_k \geq 0 \ (k = 1, 2, 3, \dots)$

2° $\sum_{k=1}^{\infty} p_k = 1$ 概率1以一定规律分布在各可能值上

即 $1 = P(\Omega) = P\left\{\bigcup_{i=1}^{\infty}(X = x_i)\right\} = \sum_{i=1}^{\infty} P(X = x_i) = \sum_{i=1}^{\infty} p_i$

分布律可以用**列表法**表示

X	x_1	x_2	$\dots$	x_i	$\dots$
P	p_1	p_2	$\dots$	p_i	$\dots$

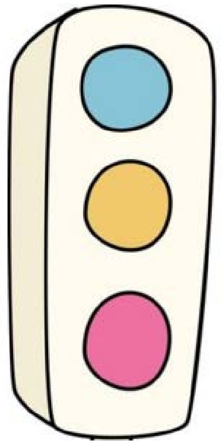
或矩阵法表示

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix}$$

步骤

1. 写出所有可能取值，即写出样本点
2. 写出相应的概率，即写出每个样本点出现的概率

例 某人骑自行车从学校到火车站，一路上要经过3个独立的交通灯，设各灯工作独立，且设各灯为红灯的概率为p，$0<p<1$，以X表示首次停车时所通过的交通灯数，求X的分布律。



解 设A_i={第i个灯为红灯}，则$P(A_i)=p$， $i=1, 2, 3$

且A_1, A_2, A_3相互独立

$$P(X=0)=P(A_1)=p\ ;\quad P(X=1)=P(\bar{A}_1A_2)=(1-p)p\ ;$$

$$P(X=2)=P(\bar{A}_1\bar{A}_2A_3)=(1-p)^2p\ ;\quad P(X=3)=P(\bar{A}_1\bar{A}_2\bar{A}_3)=(1-p)^3\ ;$$

分布律

X	0	1	2	3
P	p	$p(1-p)$	$(1-p)^2p$	$(1-p)^3$

上述事件是Ω的一个划分，即完备事件组

例 袋中有 5 只球，其中 3 只红球、 2 只白球。从中任取 2 只，设 X 表示其中的红球数，试求 X 的分布律。

解 X 可能的取值为 0 、 1 、 2，且

$$P(X=0)=\frac{C_2^2}{C_5^2}=\frac{1}{10}$$

$$P(X=1)=\frac{C_3^1C_2^1}{C_5^2}=\frac{3}{5}$$

$$P(X=2)=\frac{C_3^2}{C_5^2}=\frac{3}{10}$$

亦即 X 的分布律为

X	0	1	2
P	$\frac{1}{10}$	$\frac{3}{5}$	$\frac{3}{10}$

几种重要的离散型随机变量

1. 0-1分布

设随机变量X只可能取0与1两个值，分布律是

$P(X=k)=p^k(1-p)^{1-k}\ (k=0, 1)$　　$0<p<1$　　称X服从以 p 为参数的 0-1分布

分布律也可记为

X	0	1
P	$1-p$	p

随机试验的样本空间只包含两个元素，即$\Omega=\{\omega_1, \omega_2\}$

总可以定义随机变量 $X=X(\omega)=\begin{cases} 0 & 当\ \omega=\omega_1 \\ 1 & 当\ \omega=\omega_2 \end{cases}$

例如，硬币正反面、性别、合格与否、通过与否

2. 二项分布

如果试验E只有两个可能结果，A及$\bar{A}$，称E为**伯努利(*Bernoulli*)试验**

设$P(A)=p$，$0<p<1$，则 $P(\bar{A})=1-p$

将E**独立重复**进行n次，则为n重**伯努利试验**

注意

- **独立：** 每一次试验结果互相不影响
- **重复：** 相同条件多次进行，每次$P(A)$相同

例如，抛多次硬币（正反面）、骰子（奇数偶数）、有放回摸牌（红色黑色）等；

不放回抽样不是n重伯努利试验，**即使每次$P(A)$相同，但不独立**

伯努利

Jacob Bernoulli

Born: 27 Dec 1654 in Basel, Switzerland

Died: 16 Aug 1705 in Basel, Switzerland

设A在n重伯努利试验中发生X (X = 0, 1, 2, …, n) 次，每次事件相互独立，记 q=1-p，则

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

$$1=(p+q)^n=\sum_{k=0}^{n} C_n^k p^k q^{n-k}=\sum_{k=0}^{n} P(X=k)$$

$C_n^k p^k q^{n-k}$是二项式$(p+q)^n$展开式中出现 p^k 的一项，

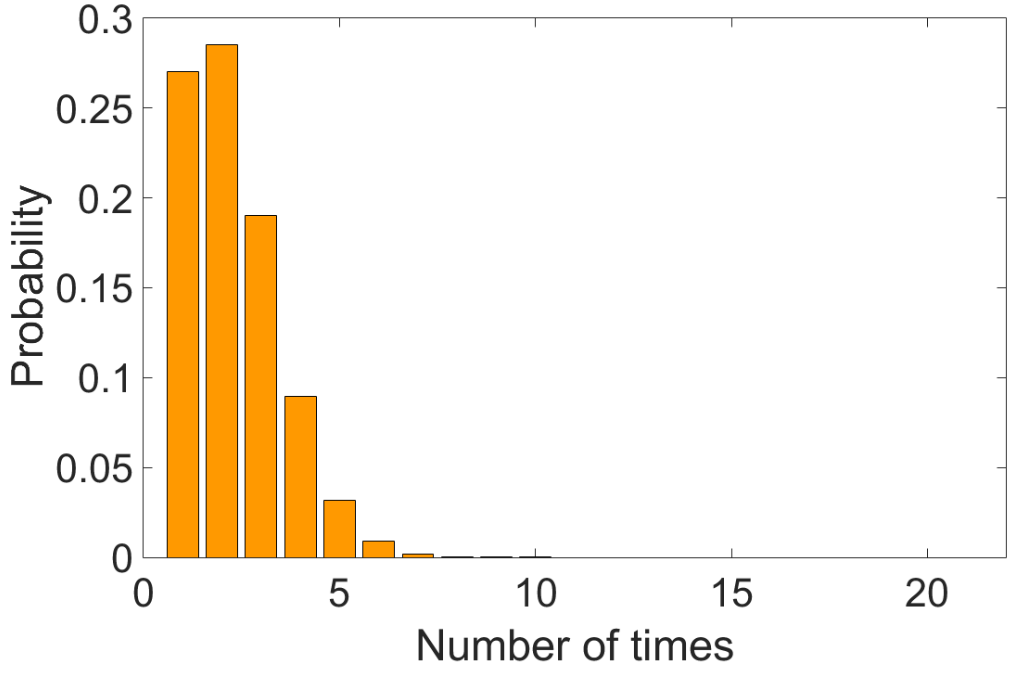
因此称X服从参数为p的二项分布，$X \sim B(n, p)$

特别地，当n=1时，二项分布退化为参数 p 的0-1分布

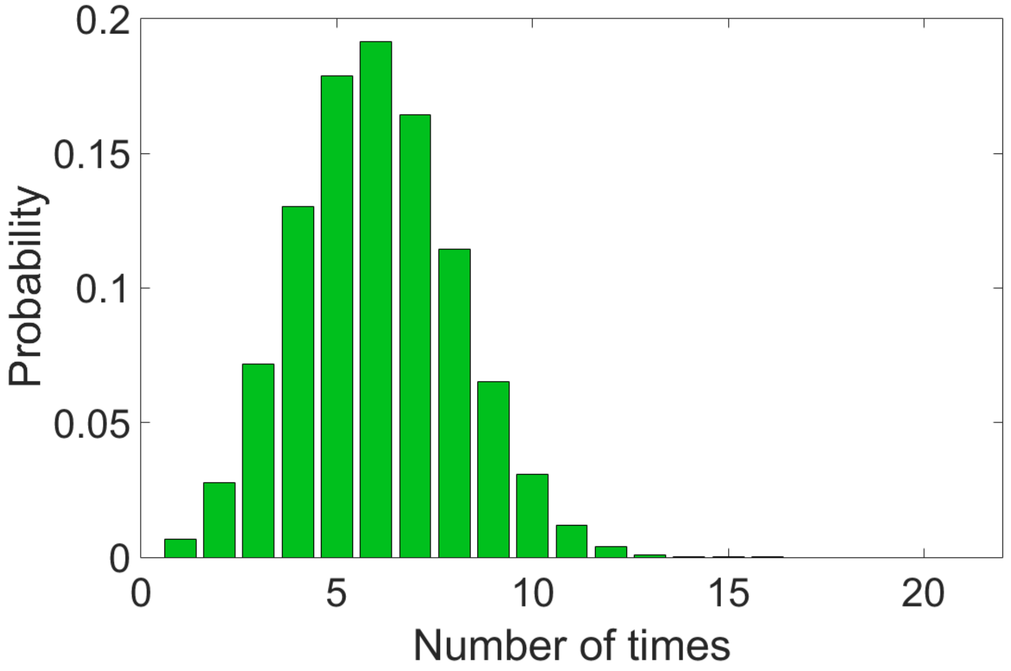
注意：二项分布是数轴上不是对称图形（除了p = 0.5 时）

概率计算值

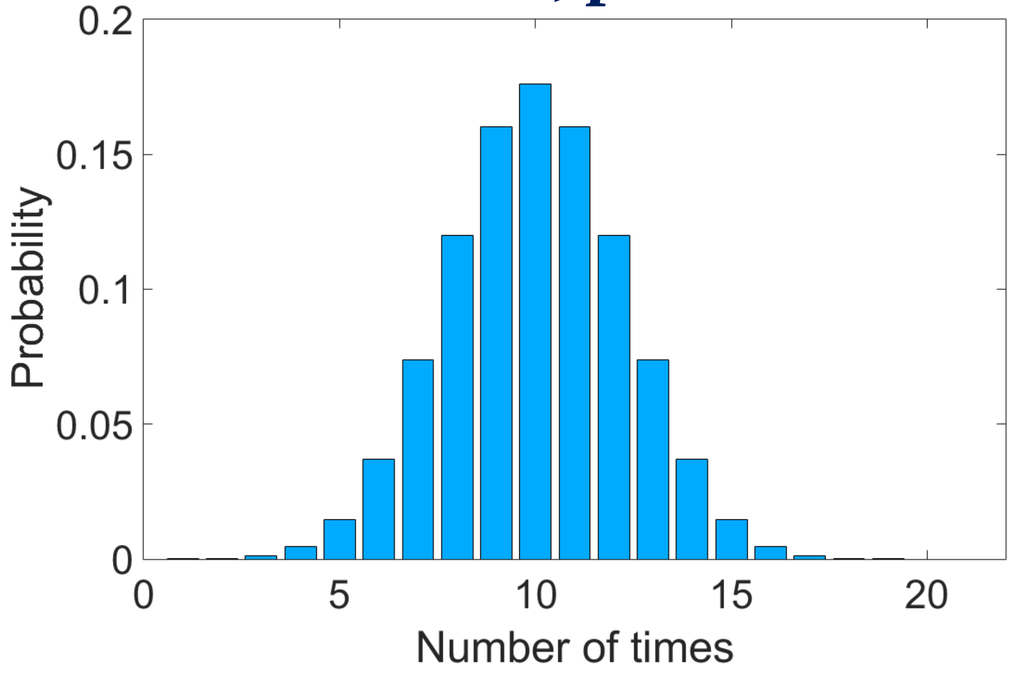
n=20, p=0.1



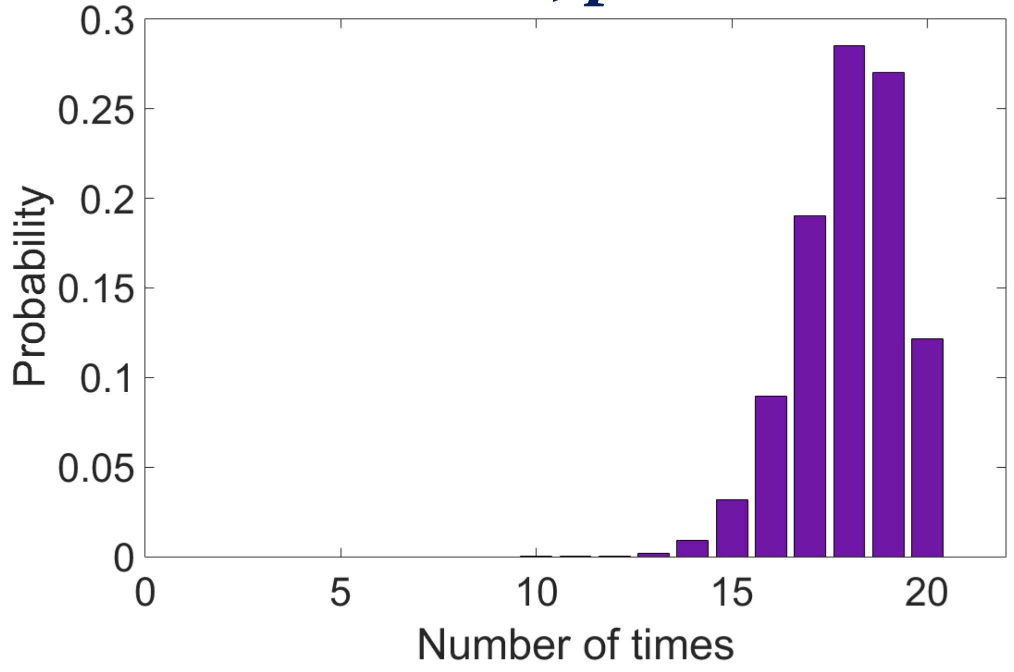
n=20, p=0.3



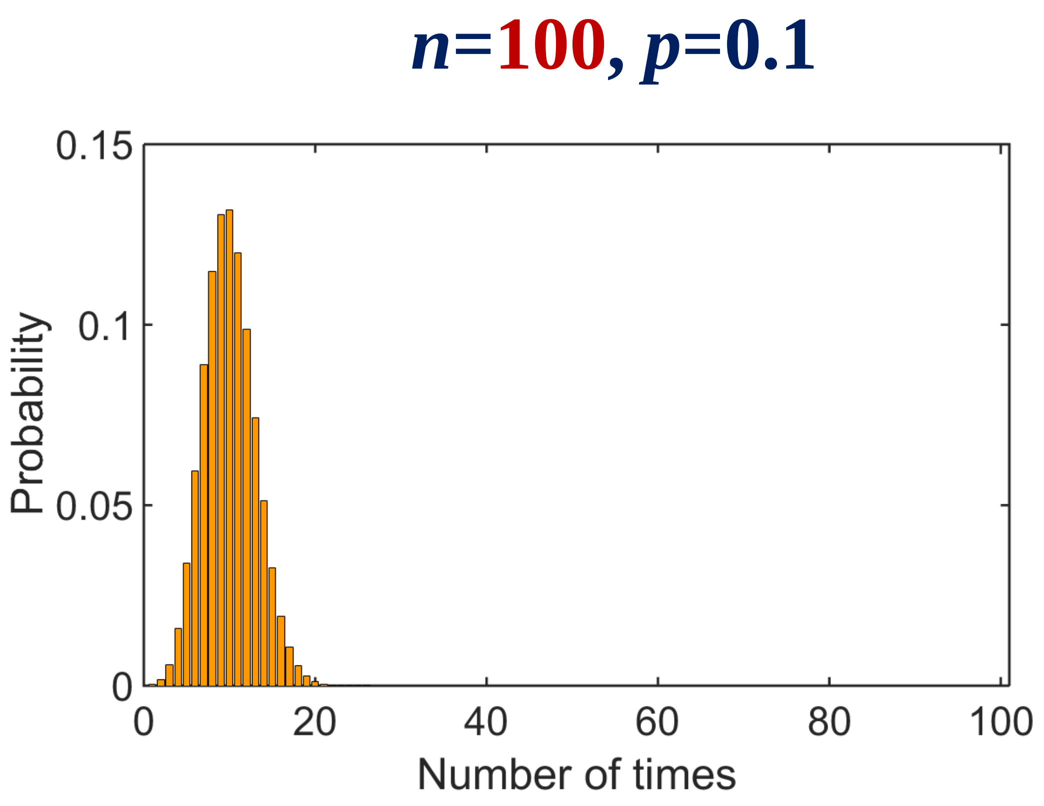
n=20, p=0.5

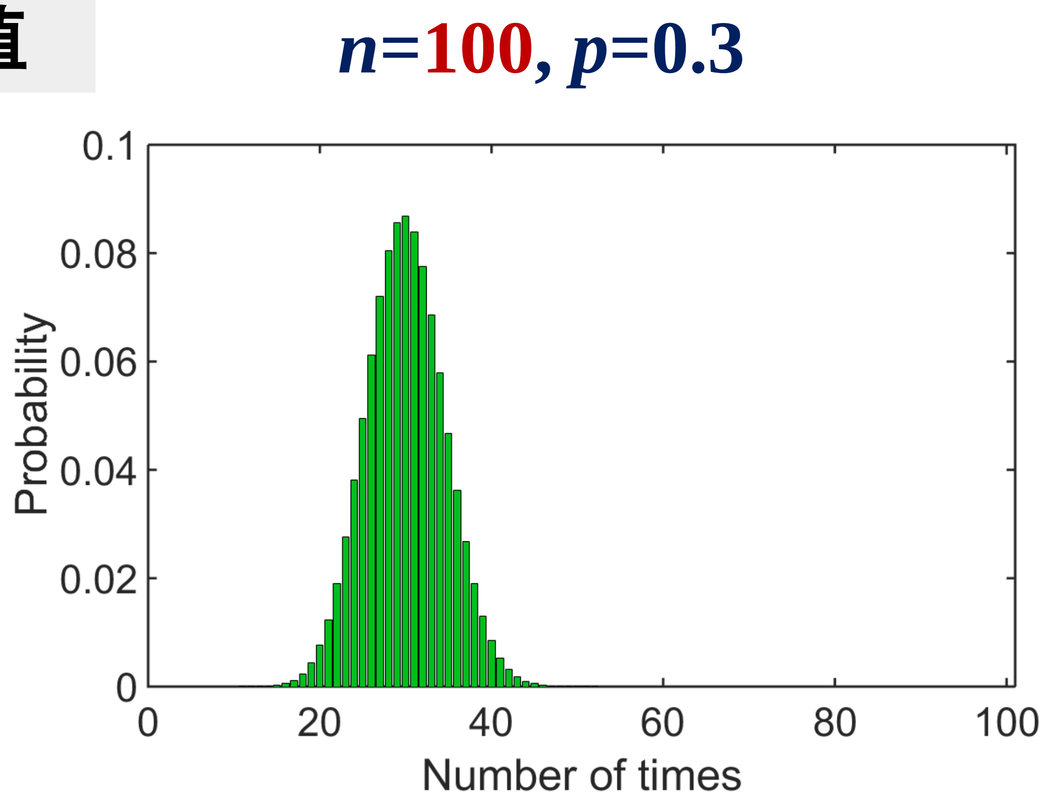


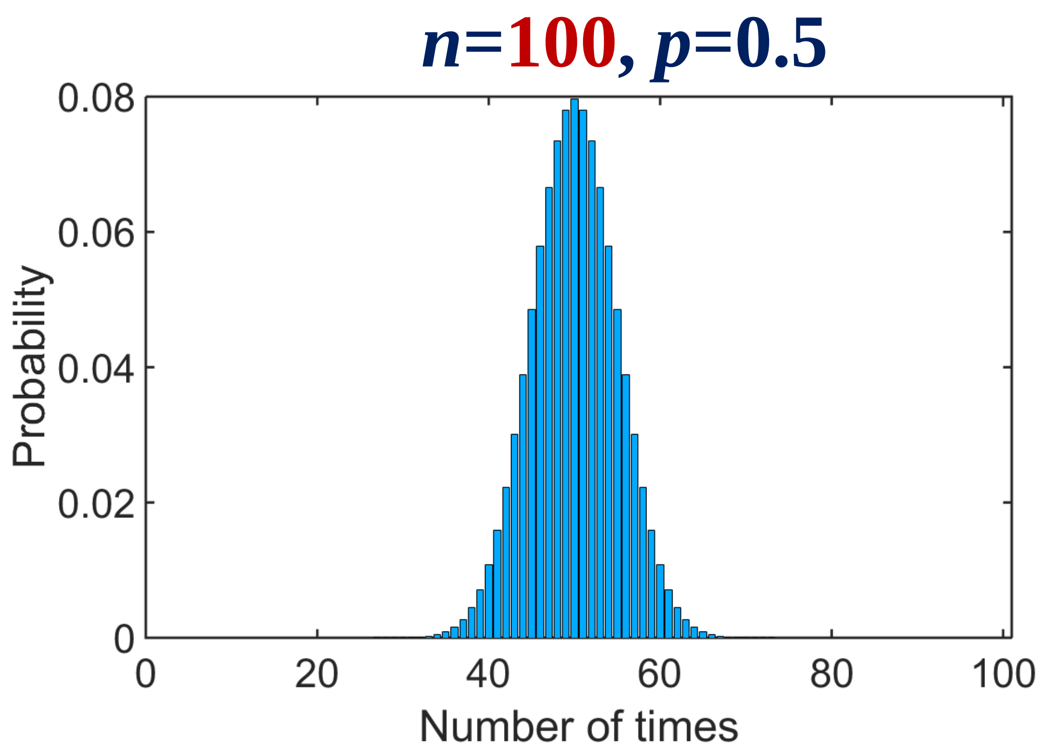
n=20, p=0.9

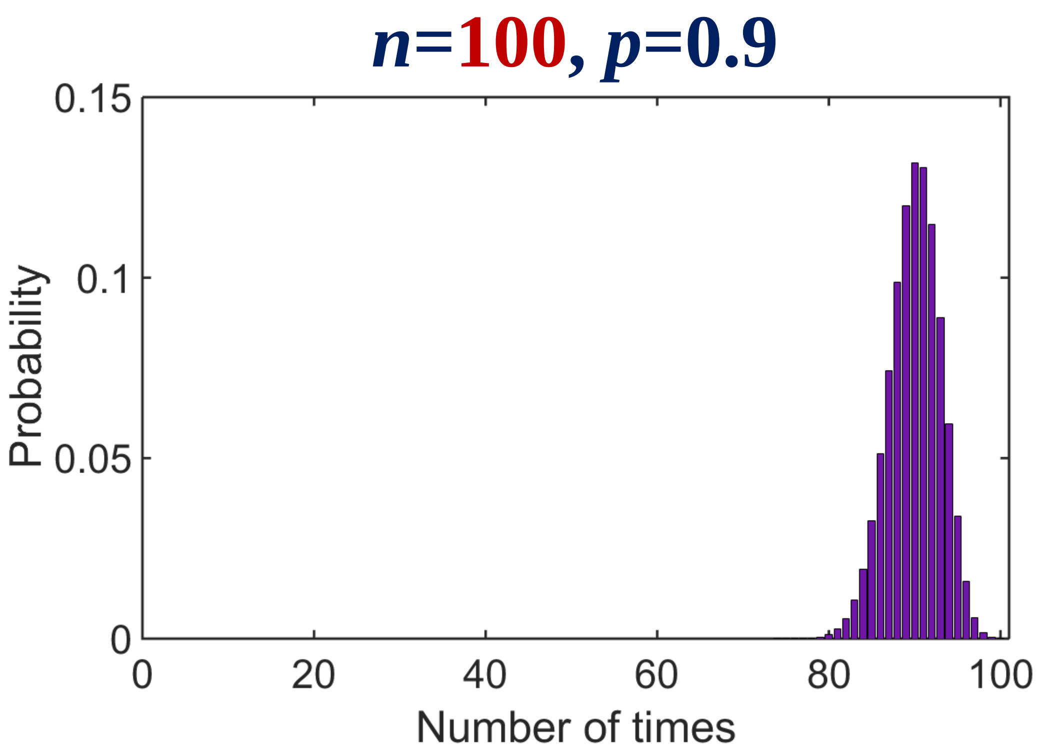


概率计算值


n=100, p=0.1
0.15
0.1
0.05
0
Probability
0
20
40
60
80
100
Number of times


n=100, p=0.3
0.1
0.08
0.06
0.04
0.02
0
Probability
0
20
40
60
80
100
Number of times


n=100, p=0.5
0.08
0.06
0.04
0.02
0
Probability
0
20
40
60
80
100
Number of times


n=100, p=0.9
0.15
0.1
0.05
0
Probability
0
20
40
60
80
100
Number of times

例 设有80台同类型设备，各台工作相互独立，发生故障的概率都是0.01，且一台设备的故障能由一个人处理。考虑两种配备维修工人的方法，其一是由4个人维护，每人负责20台；其二是由3个人共同维护80台。

试比较这两种方法在设备发生故障时不能及时维修的概率大小。

解 按照第一种方法　以 X 记“第一人维护的20台中同一时刻发生故障的台数”，以A_i (i=1, 2, 3, 4) 表示事件“第 i 人维护的20台中发生故障不能及时维修”

则80台中发生故障不能及时维修的概率为 $P(A_1\cup A_2\cup A_3\cup A_4)\ge P\left(A_1\right)=P(X\ge 2)$

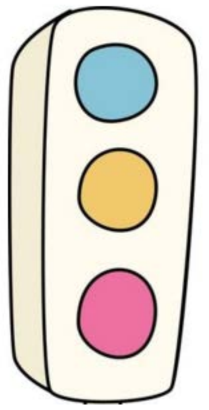
$X\sim B(20,0.01)$　$P(X\ge 2)=1-\sum_{k=0}^{1}P(X=k)=1-\sum_{k=0}^{1}C_{20}^{k}(0.01)^{k}(0.99)^{20-k}=0.0169$

$$P(A_1\cup A_2\cup A_3\cup A_4)\ge 0.0169$$

按照第二种方法　以 Y 记“80台中同一时刻发生故障的台数”　$Y\sim B(80,0.01)$

则80台中发生故障不能及时维修的概率为 $P(Y\ge 4)=1-\sum_{k=0}^{3}C_{80}^{k}(0.01)^{k}(0.99)^{80-k}=0.0087$

例 某人骑自行车从学校到火车站，一路上要经过3个独立的交通灯，设各灯工作独立，且设各灯为红灯的概率为p，$0<p<1$，以Y表示路上遇到红灯的次数。



求 (i) Y 的分布律； (ii) 求恰好遇2次红灯的次数。

解 这是三重伯努利试验 $Y \sim B(3, p)$

(i) $P(Y=k)=C_3^k p^k (1-p)^{3-k}, k=0, 1, 2, 3$

(ii) $P(Y=2)=C_3^2 p^2 (1-p)$

例 某人独立射击n次，设每次命中率为p，$0<p<1$，设命中X 次，(i) 求X的分布律；(ii) 求至少有一次命中的概率。

解 这是n重伯努利试验 $X \sim B(n, p)$

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

$\lim\limits_{n\to\infty} P(X\geq 1)=1$

(ii) $P(X\geq 1)=1-P(X=0)=1-(1-p)^n$

小概率事件在大量重复试验必然发生

例 有一大批产品，其验收方案如下：先作第一次检验，从中任取10件，经检验无次品，则接受这批产品，次品数大于2，则拒收；否则作第二次检验，从中任取5件，仅当5件中无次品便接受这批产品，设产品的次品率为p。

求这批产品能被接受的概率$L(p)$。

解 设X为第一次抽得的次品数，Y为第二次抽得的次品数；

则$X \sim B(10, p)$，$Y \sim B(5, p)$，且$\{X=i\}$与$\{Y=j\}$独立

A={接受该批产品}

$$L(p) = P(X=0)P(A|X=0) + P(1 \le X \le 2)P(A|1 \le X \le 2) + P(X>2)P(A|X>2)$$

$$P(A|1 \le X \le 2) = P(Y=0|1 \le X \le 2) = P(Y=0)$$

$$= (1-p)^{10} + [10p(1-p)^9 + 45p^2(1-p)^8](1-p)^5$$

3. 泊松(*Poisson*)分布

若随机变量X的概率分布律为 $P(X=k)=\dfrac{\lambda^k e^{-\lambda}}{k!}, k=0,1,2,\cdots, \lambda>0$

称X服从参数为λ ($\lambda>0$) 的泊松分布，记 $X\sim P(\lambda)$

可证 $\sum_{k=0}^{\infty}P(X=k)=\sum_{k=0}^{\infty}\dfrac{\lambda^k e^{-\lambda}}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\dfrac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1$

适合描述单位时间或空间内随机事件发生的次数

某夜值班台收到的呼叫、某路口一段时间交通事故次数、
一小时内到车站的乘客、某山区一年自然灾害的次数、
单位面积放射性物质发射的粒子、显微镜下某区域的白血球等

泊松

Siméon

Poisson

Born: 21 June 1781 in Pithiviers, France

Died: 25 April 1840 in Sceaux (near Paris), France

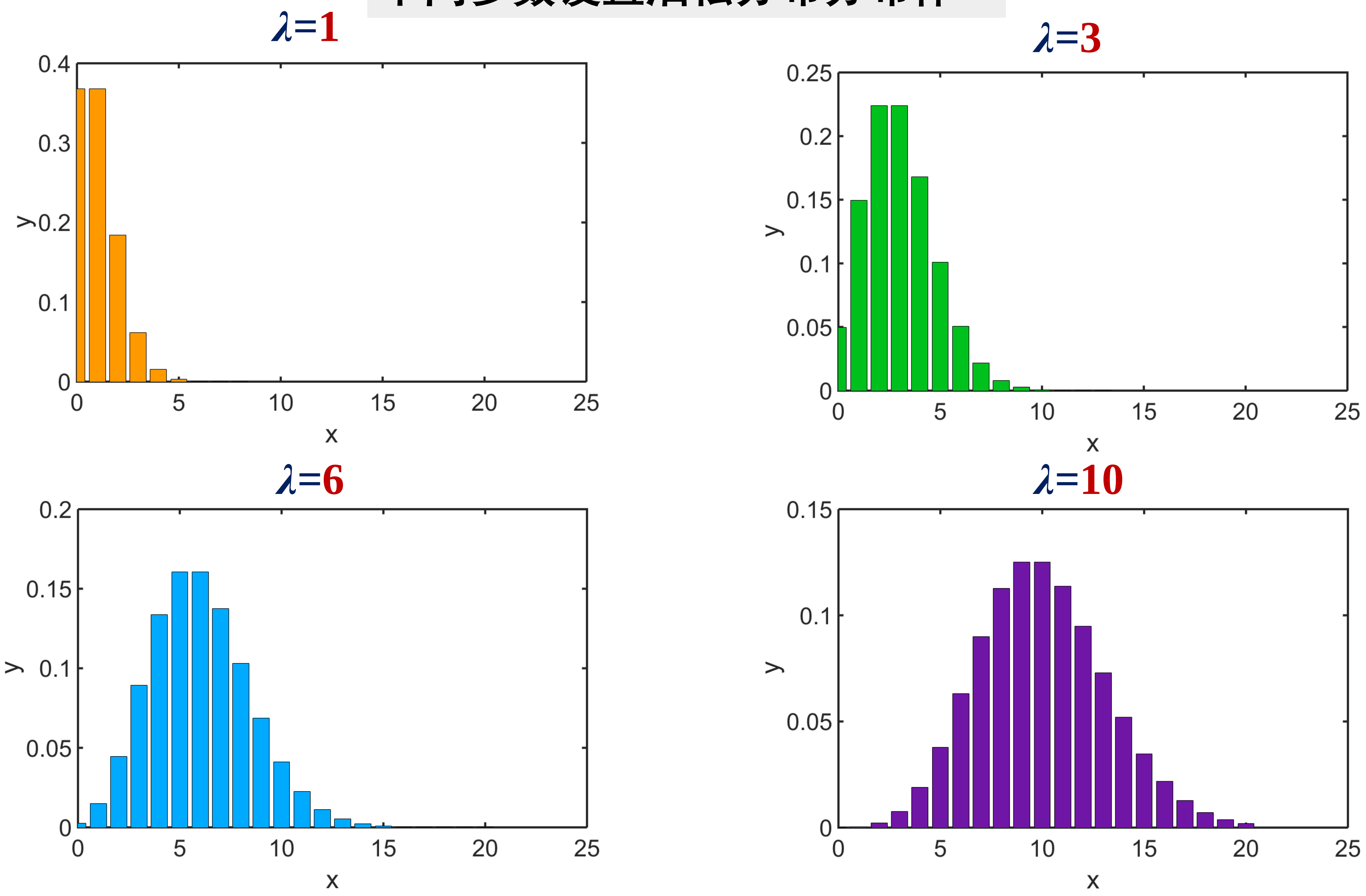
泊松分布表

$$P(X \leqslant x) = \sum_{k=0}^{x} \frac{\lambda^k e^{-\lambda}}{k!}$$

x	λ = 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.9048	0.8187	0.7408	0.6730	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865
4		1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977
5				1.0000	1.0000	1.0000	0.9999	0.9998	0.9997
6							1.0000	1.0000	1.0000

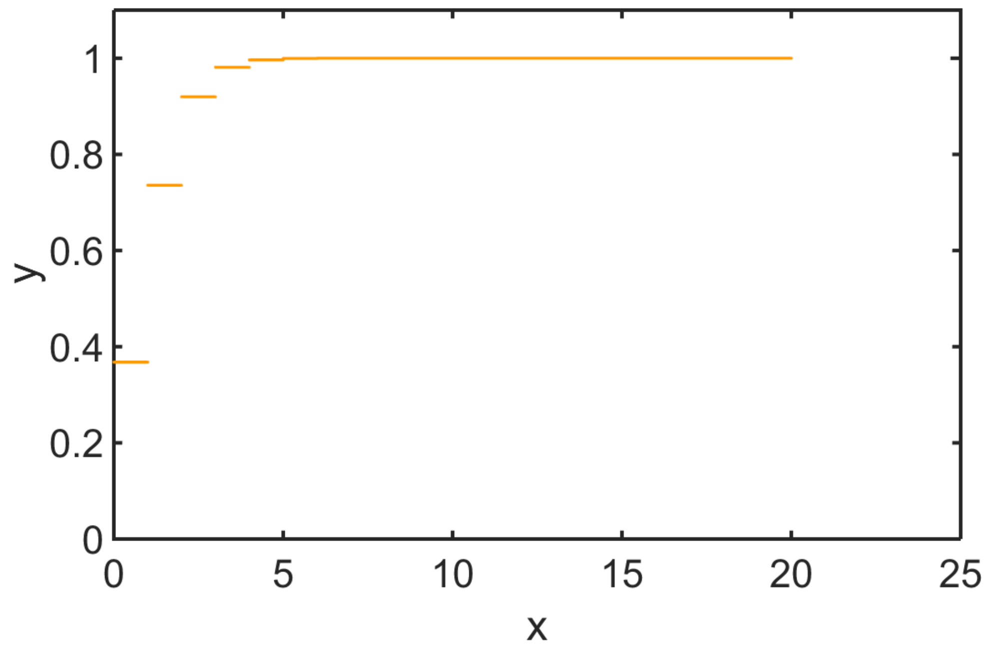
x	λ = 1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0	0.3679	0.2231	0.1353	0.0821	0.0498	0.0302	0.0183	0.0111	0.0067
1	0.7358	0.5578	0.4060	0.2873	0.1991	0.1359	0.0916	0.0611	0.0404
2	0.9197	0.8088	0.6767	0.5438	0.4232	0.3208	0.2381	0.1736	0.1247
3	0.9810	0.9344	0.8571	0.7576	0.6472	0.5366	0.4335	0.3423	0.2650
4	0.9963	0.9814	0.9473	0.8912	0.8153	0.7254	0.6288	0.5321	0.4405
5	0.9994	0.9955	0.9834	0.9580	0.9161	0.8576	0.7851	0.7029	0.6160
6	0.9999	0.9991	0.9955	0.9858	0.9665	0.9347	0.8893	0.8311	0.7622
7	1.0000	0.9998	0.9989	0.9958	0.9881	0.9733	0.9489	0.9134	0.8666
8		1.0000	0.9998	0.9989	0.9962	0.9901	0.9786	0.9597	0.9319
9			1.0000	0.9997	0.9989	0.9967	0.9919	0.9829	0.9682
10				0.9999	0.9997	0.9990	0.9972	0.9933	0.9863
11				1.0000	0.9999	0.9997	0.9991	0.9976	0.9945
12					1.0000	0.9999	0.9997	0.9992	0.9980

不同参数设置泊松分布分布律

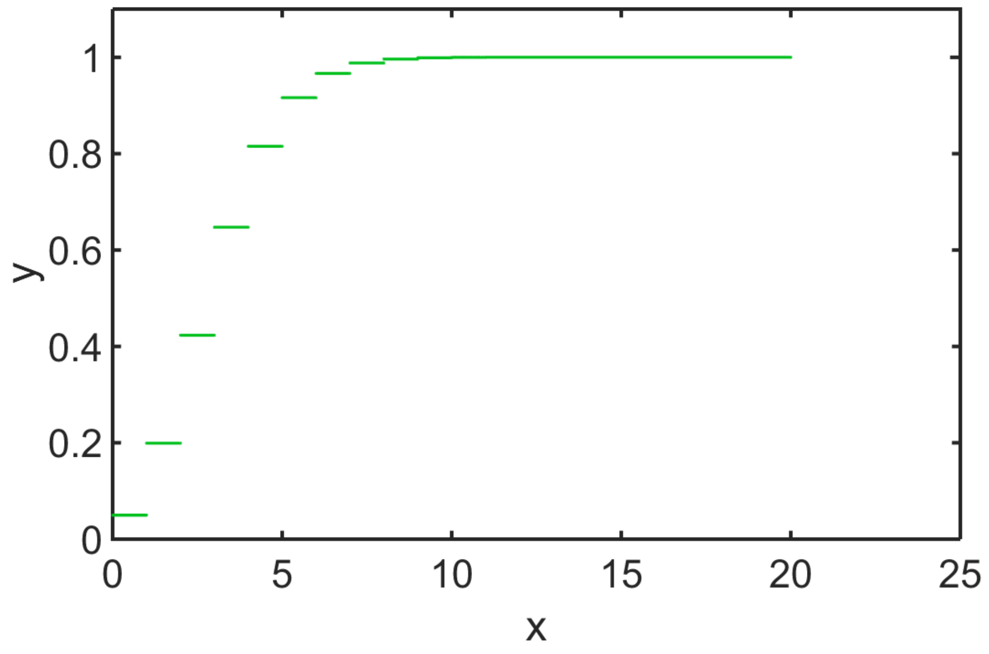


不同参数设置泊松分布的分布函数

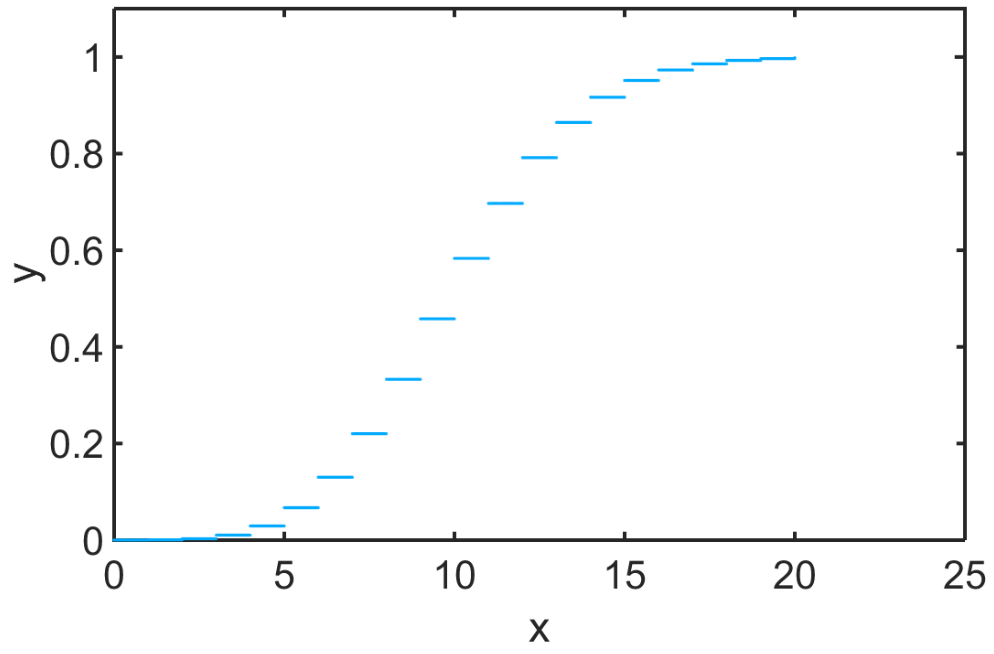
$\lambda=1$



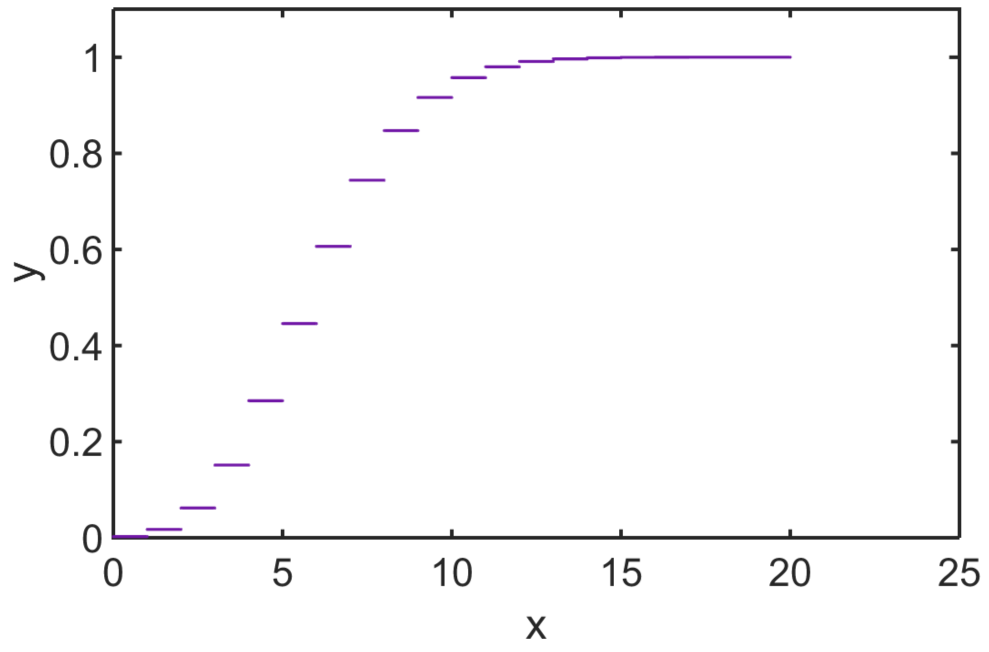
$\lambda=3$



$\lambda=6$



$\lambda=10$



例 某电话机交换台每分钟收到用户的呼叫次数 X 服从参数为 4 的 Poisson 分布。

试求: (1)某一分钟恰有 8 次呼叫的概率; (2)某一分钟的呼叫次数大于 3 的概率。

解 由于 $X \sim P(4)$，因此 X 的分布律为

$$P(X=k)=\frac{e^{-4}4^{k}}{k!},\ k=0,1,2,\cdots$$

(1)所求的概率为 $P(X=8)=\dfrac{e^{-4}4^{8}}{8!}=0.0298$

(2)所求的概率为 $P(X>3)=1-P(X=0)-P(X=1)-P(X=2)-P(X=3)=1-\sum\limits_{k=0}^{3}\dfrac{e^{-4}4^{k}}{k!}=0.5665$

例 设某汽车停靠站候车人数 $X \sim P(\lambda),\ \lambda = 4.5$

(i) 求至少有两人候车的概率；

(ii) 已知至少有两人候车，求恰有两人候车的概率。

解 $P(X = k) = \dfrac{e^{-4.5}4.5^k}{k!},\ k = 0, 1, 2, \cdots$

(i) $P(X \ge 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-4.5}(1 + 4.5) = 0.9389$

(ii) $P(X = 2 / X \ge 2) = \dfrac{P(X = 2)}{P(X \ge 2)} = 0.1198$

泊松分布满足如下性质：

1° 每个时间段（或区域）发生事件的概率相同

2° 每个时间段（或区域）之间彼此之间互不影响，相互独立

泊松(*Poisson*)定理

设$\lambda>0$是一常数，n是任意正整数，设$np_n=\lambda$，则对于任一固定的非负整数k，有

$$\lim_{n\to\infty} C_n^k p_n^{\ k}(1-p_n)^{n-k}=\frac{\lambda^k e^{-\lambda}}{k!}$$

二项分布与泊松分布之间的数值关系

- 二项分布 $X\sim B(n,p)$ $\quad P(X=k)=C_n^k p^k(1-p)^{n-k},\ k=0,1,\cdots,n$
- 泊松分布 $X\sim P(\lambda)$ $\quad P(X=k)=\dfrac{\lambda^k e^{-\lambda}}{k!},\ k=0,1,2,\cdots$

证 $np_n=\lambda$

$$C_n^k p_n^{\ k}(1-p_n)^{n-k}=\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$=\frac{\lambda^k}{k!}\left[1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right]\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$$

对任意固定k, 当$n\to\infty$时

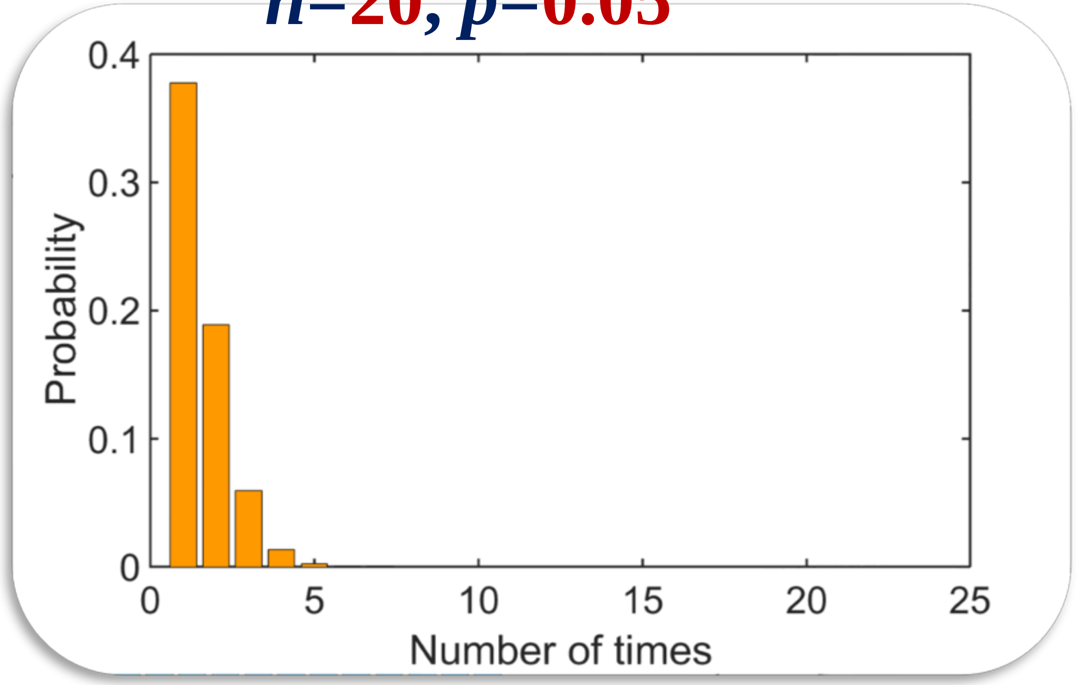
$$1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\to 1 \qquad \left(1-\frac{\lambda}{n}\right)^n\to e^{-\lambda} \qquad \left(1-\frac{\lambda}{n}\right)^{-k}\to 1$$

$$\therefore\quad \lim_{n\to\infty} C_n^k p_n^{\ k}(1-p_n)^{n-k}=\frac{\lambda^k e^{-\lambda}}{k!}$$

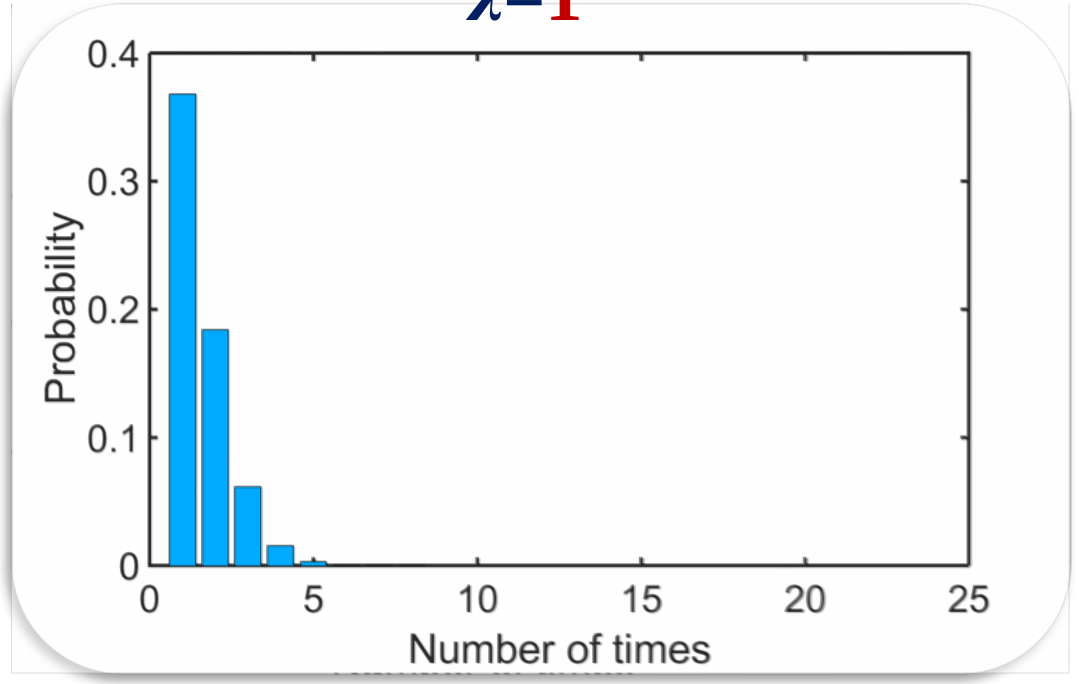
$$e=\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x$$

二项分布 **分布律比较** **泊松分布**

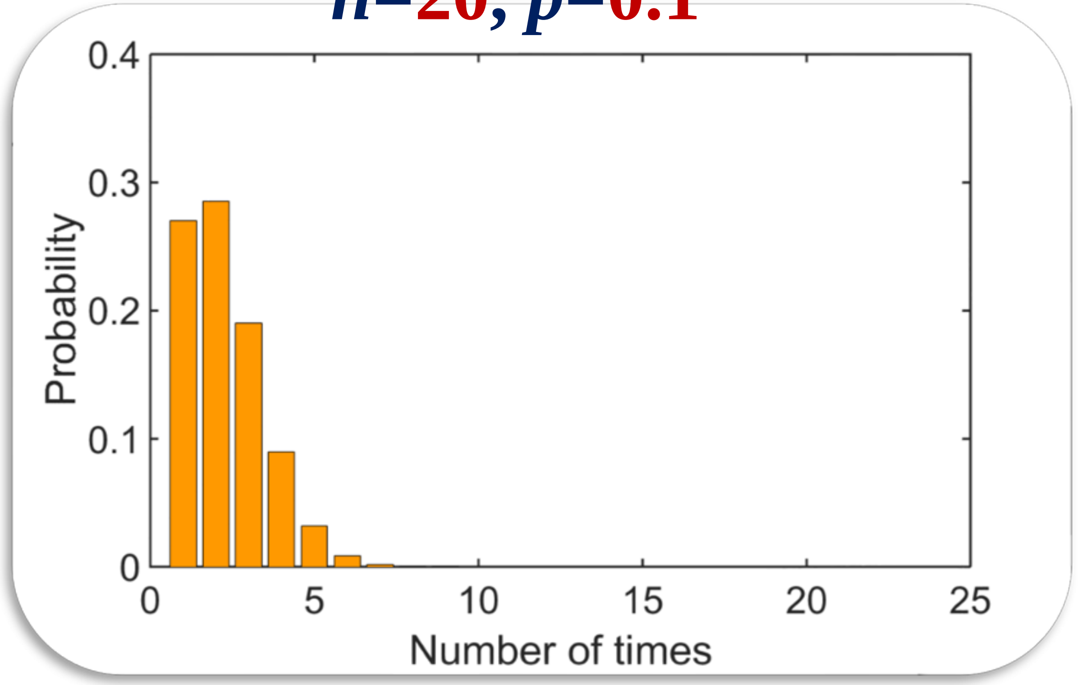
n=20, p=0.05



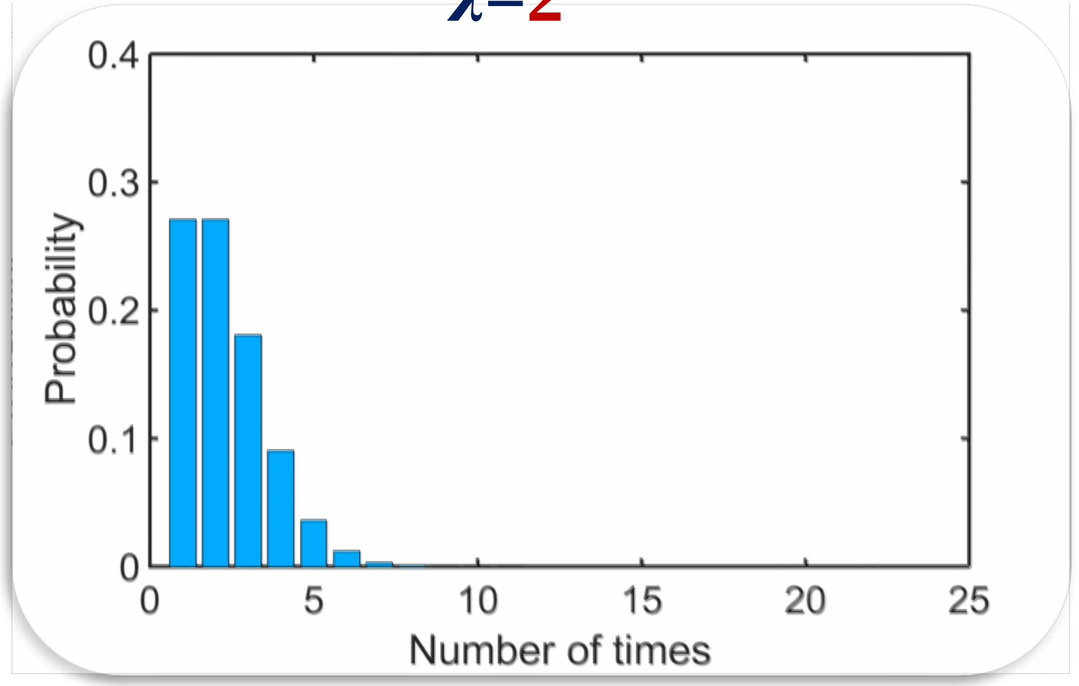
λ=1



n=20, p=0.1

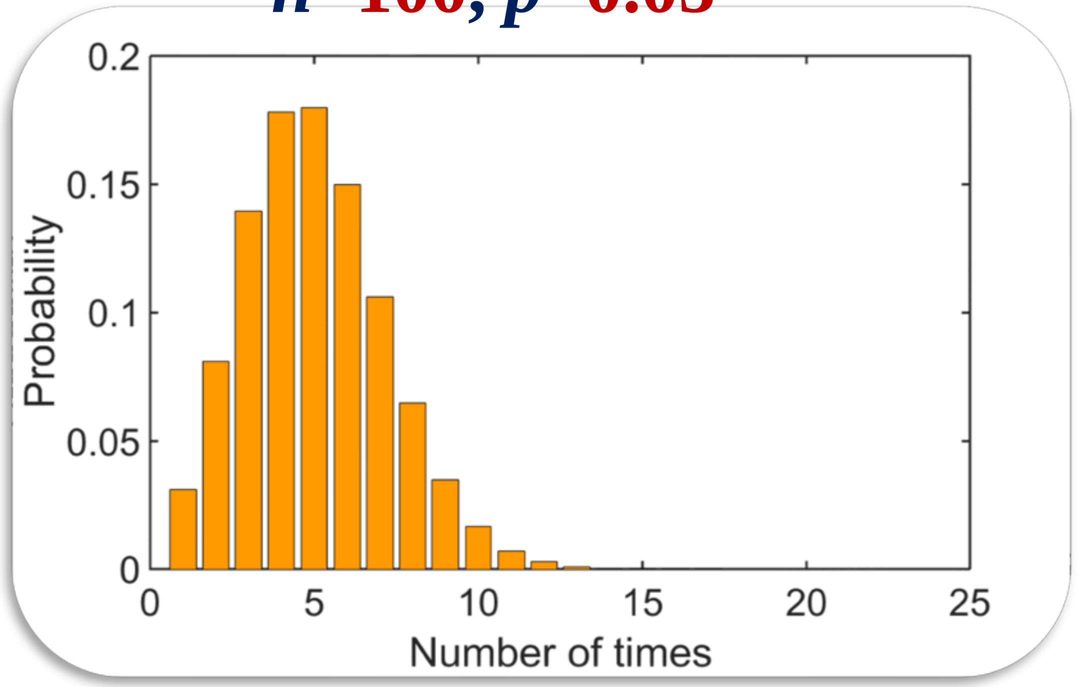


λ=2

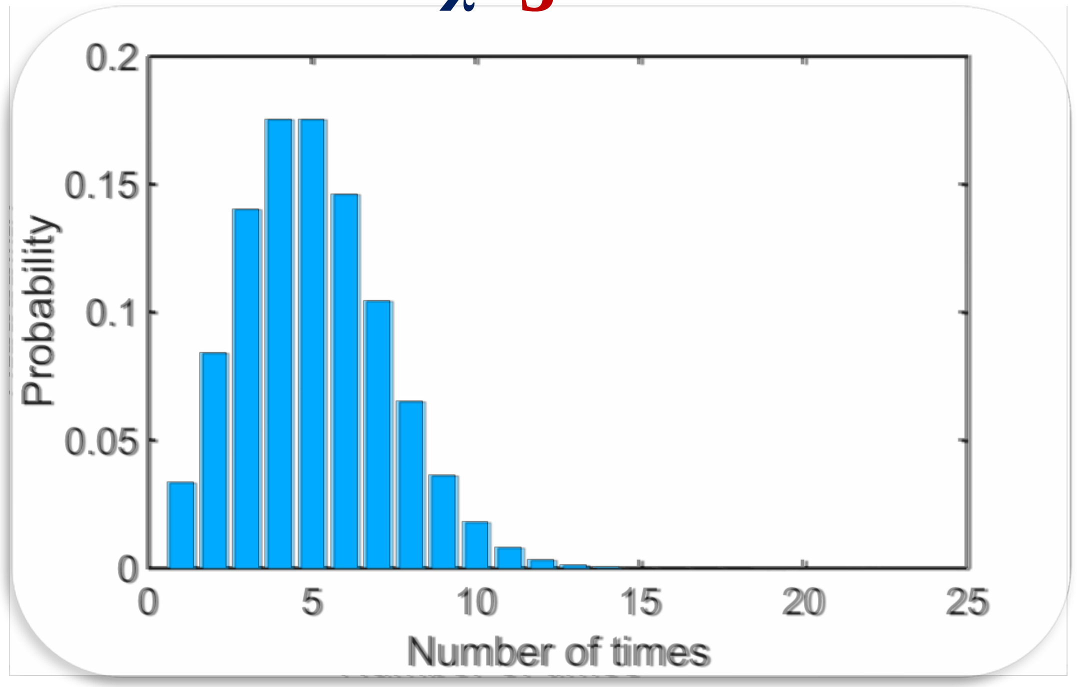


二项分布 　　分布律比较 　　泊松分布

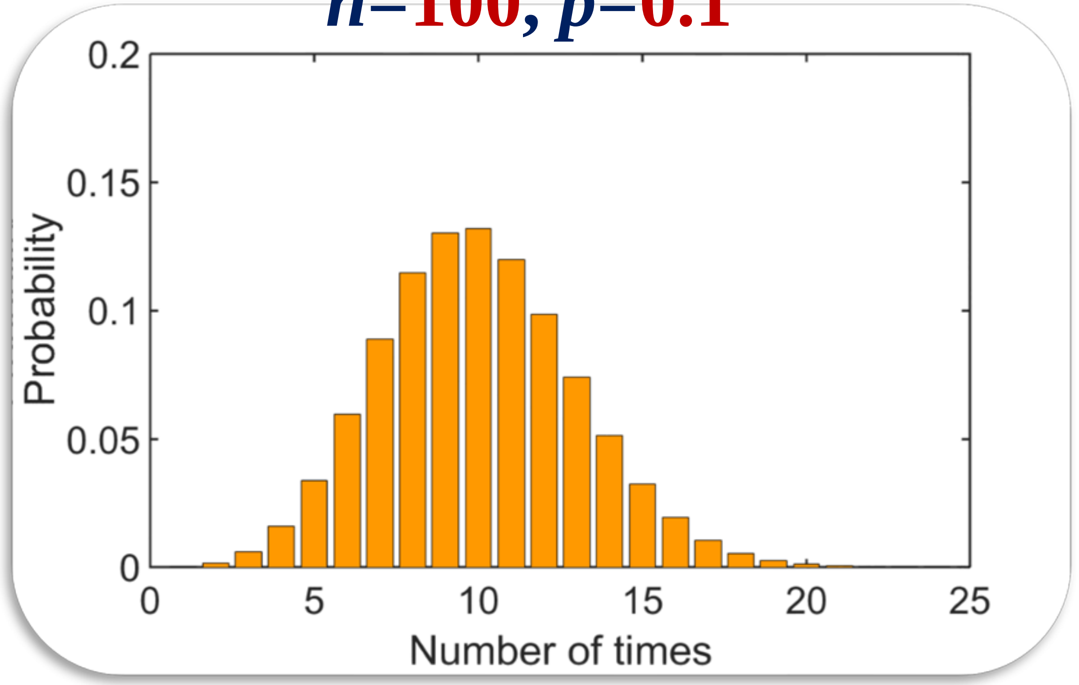
n=100, p=0.05



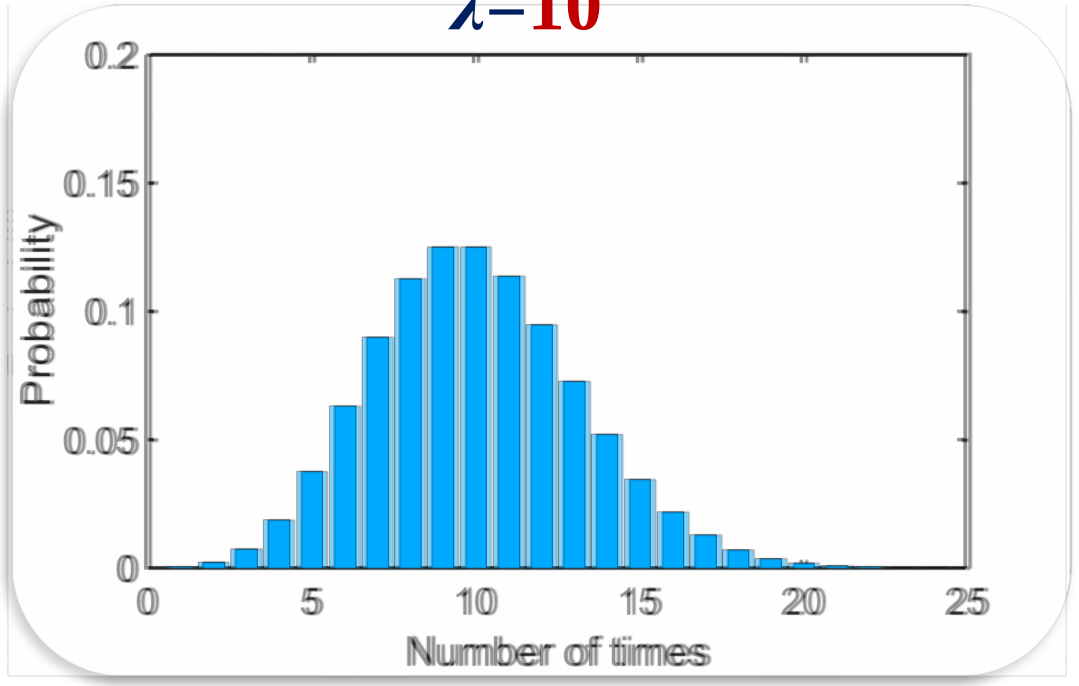
λ=5



n=100, p=0.1



λ=10



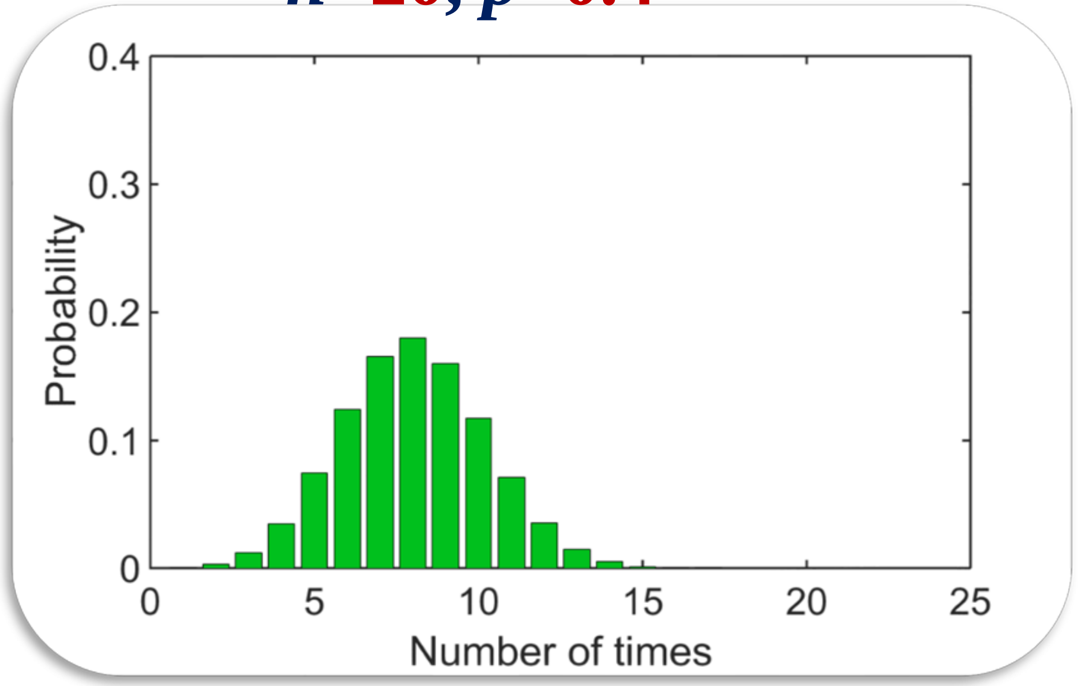
分布律比较

二项分布

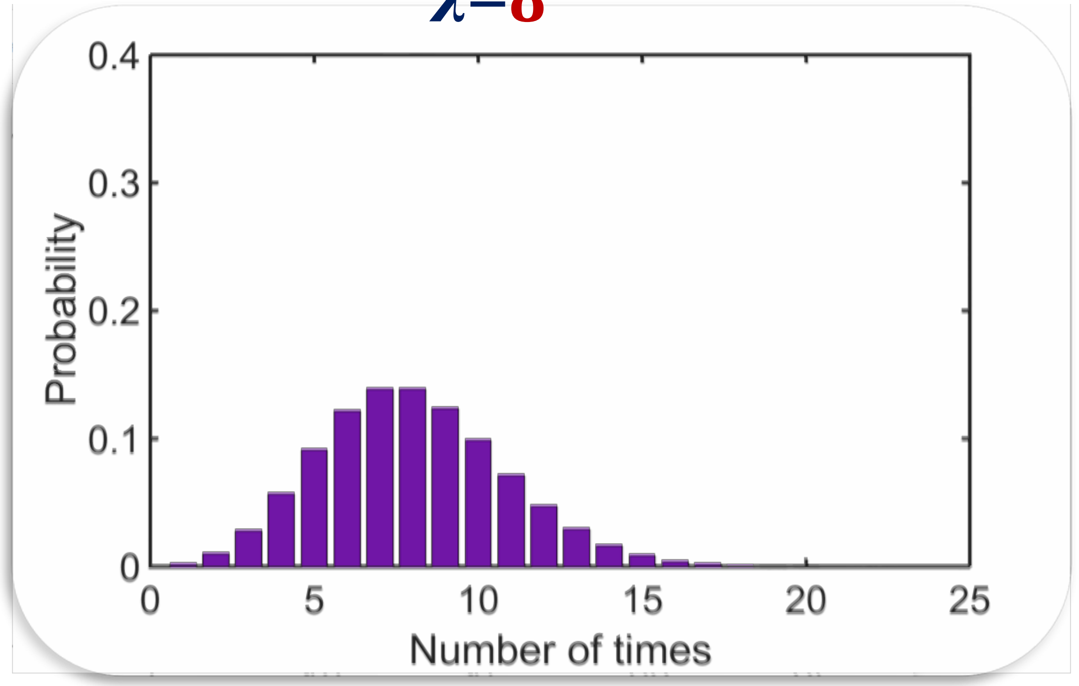
泊松分布

***p* 值不够小！**

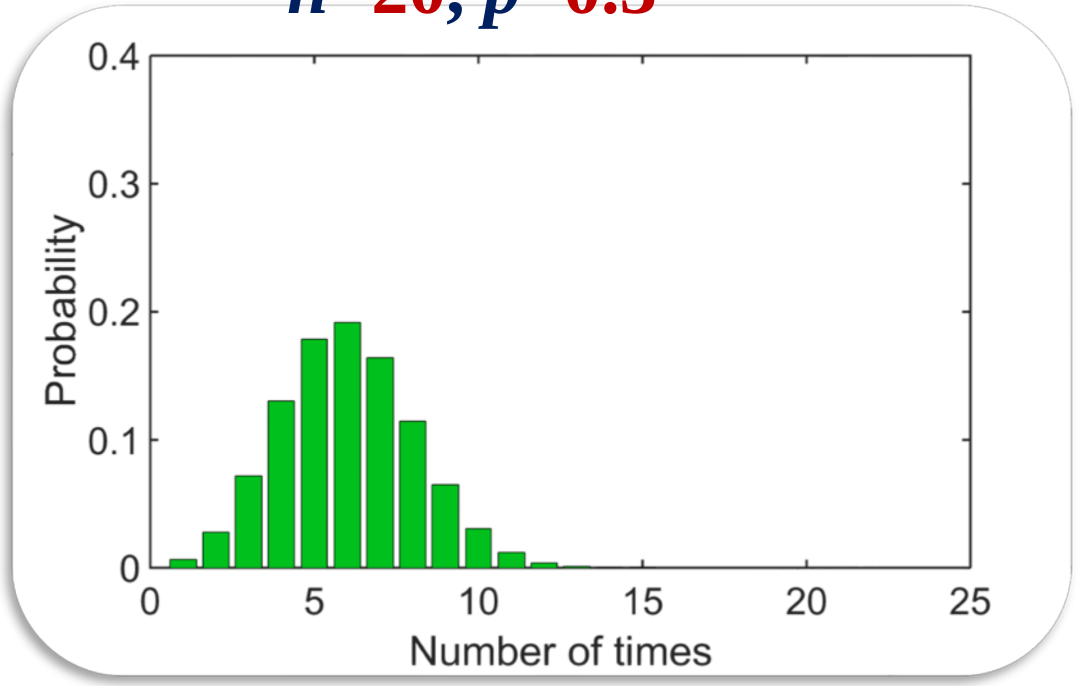
n=20, p=0.4



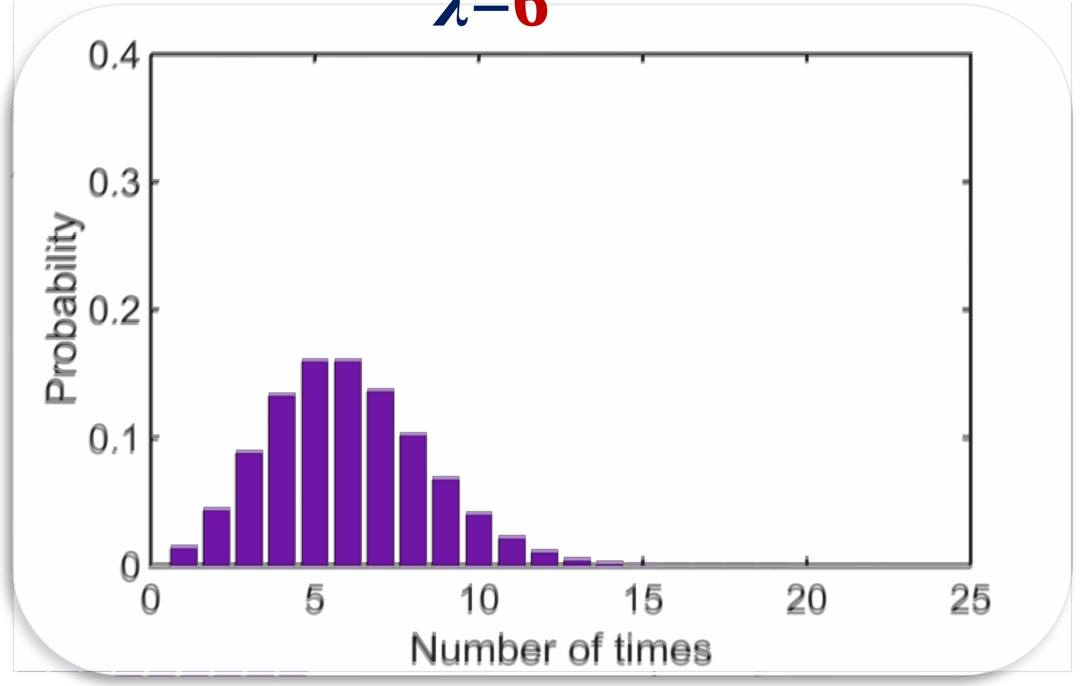
λ=8



n=20, p=0.3

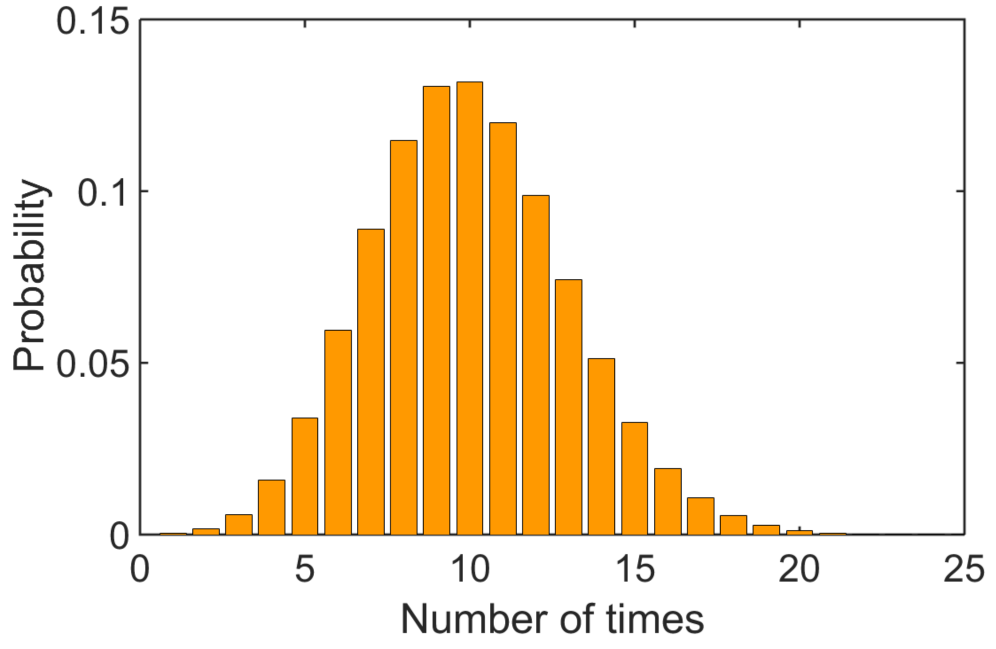


λ=6

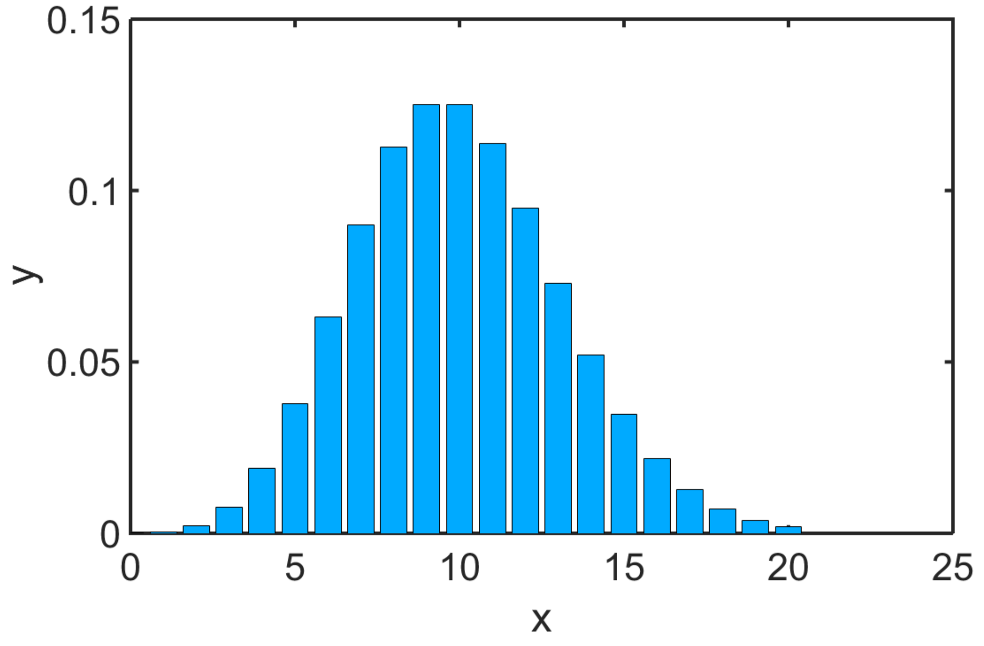


二项分布　　**分布律比较**　　**泊松分布**　　***p* 值不够小！**

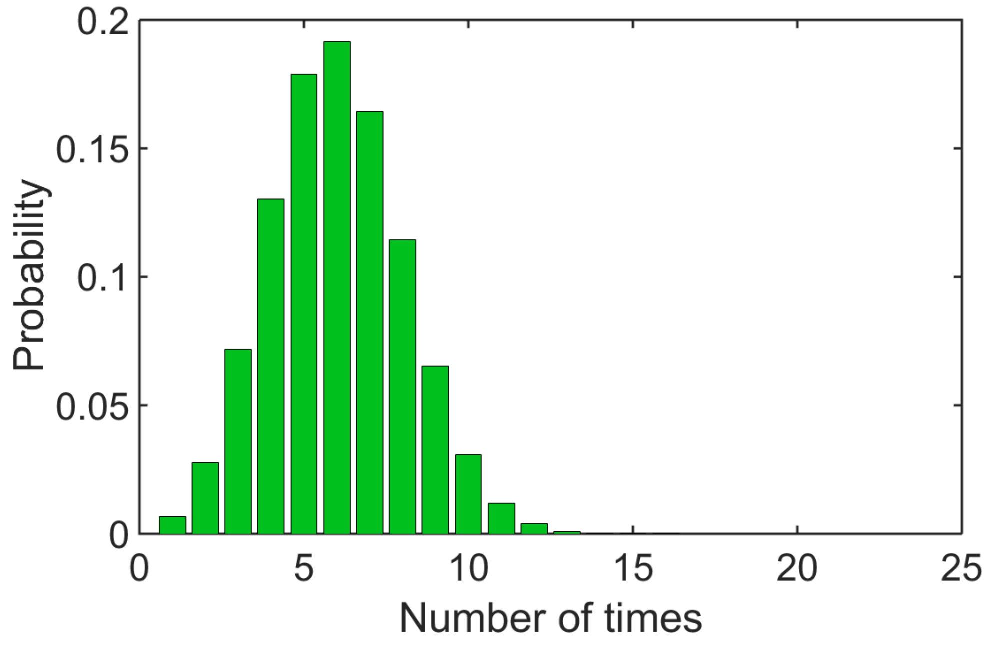
n=100, p=0.1



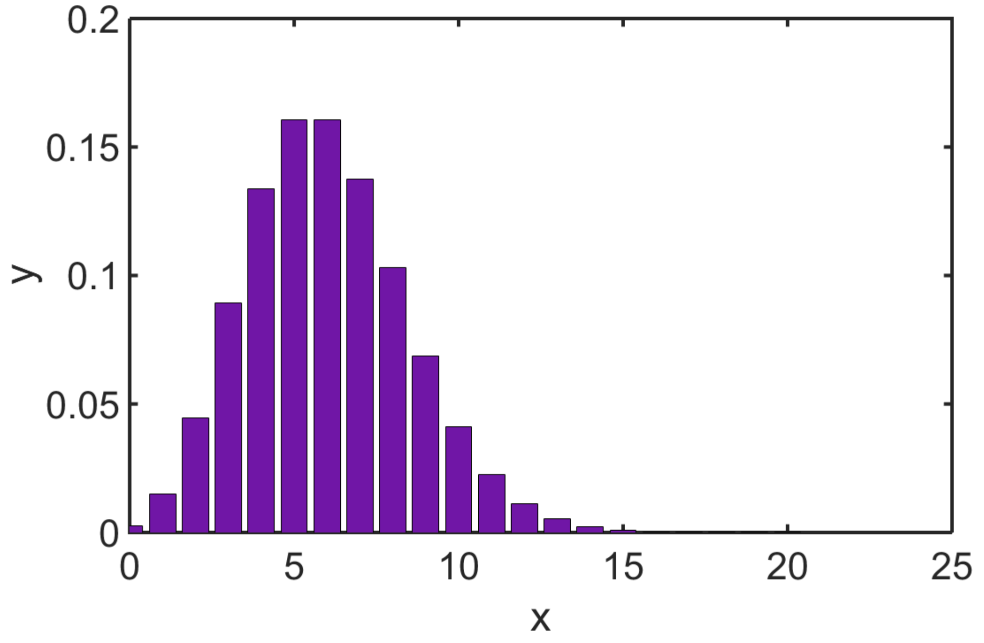
λ=10



n=20, p=0.3



λ=6



Poisson定理

二项分布在n很大，p很小时($np=\lambda$) 可以用泊松分布来近似

当 $n>10$，$p<0.1$ 时，

$$C_n^k p^k (1-p)^{n-k} \approx \frac{e^{-\lambda}\lambda^k}{k!}$$

在 $n>20$，$p<0.05$ 时近似效果颇佳

二项分布的实验次数n 趋近∞ 时，可得到正态分布曲线（第5章证明）

二项分布可以用泊松分布（离散）和正态分布（连续）来近似，但使用条件不同

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，$X \sim B(400, 0.02)$，两种方法求解

(i) 二项分布

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$$

$$= 1 - 0.98^{400} - 400 \times 0.02 \times 0.98^{399} = 0.9972$$

(ii) 泊松分布近似

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

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

4.超几何分布

引例 袋中有 a 只白球，b 只红球. 从袋中任取 n 只球，求取到k （$\le \min(n,a)$ ）只白球的概率。

解 从 $a+b$ 只球中任取 n 只，样本点总数为 C_{a+b}^{n}

取到 k 只白球的有利场合数为 $C_a^k \cdot C_b^{n-k}$

故所求概率为 $p_k = \dfrac{C_a^k \cdot C_b^{n-k}}{C_{a+b}^{n}}$

定义 设有 N 件产品，其中有 M 件次品，从中任取 n 件，则取出的次品数 X 的分布律为

$$P(X=k)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k=0,1,2,\dots,\min\{M,n\}$$

则称 X 服从参数为 N、M、n 的**超几何分布**。

定理 设 X 服从参数为 N、M、n 的超几何分布，对于固定的n，当 $N \to \infty$ 时，$\dfrac{M}{N} \to p$，则

$$\lim_{N\to\infty}\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} = C_n^k p^k (1-p)^{n-k}$$

事实上，超几何分布的背景是不放回抽样，而二项分布的背景是有放回抽样，当N很大时，不放回抽样近似于放回抽样。

例 **有一批种子共计10000颗，发芽率为99%，现从中随机抽取200颗，求至多1颗不发芽的概率。**

解 **不发芽数量：*M*=10000×1%=100**

发芽数量：10000-100=9900

***n*=200**

***X*“取出种子中不发芽的数量”**

$$P(X\le 1)=P(X=0)+P(X=1)$$

$$=\frac{C_{100}^{0}C_{9900}^{200}}{C_{10000}^{200}}+\frac{C_{100}^{1}C_{9900}^{199}}{C_{10000}^{200}}$$

$$P(X\le 1)=P(X=0)+P(X=1)$$

$$=C_{200}^{0}0.01^{0}0.99^{200}+C_{200}^{1}0.01^{1}0.99^{19}$$

***n*=200, *p*=0.01, *np*=2, $X\sim P(2)$**

$P(X\le 1)=P(X=0)+P(X=1)$**=0.406**

5.几何分布

引例 某射手连续向一目标射击，直到命中为止，已知他每发命中的概率是p，求所需射击发数X的分布律.

解 显然，X 可能取的值是1,2,…， 为了计算 $P\{X=k\}$， $k=1,2,\dots$,

设 $A_k=\{$第k发命中$\}$，$k=1,2,\dots$,

于是 $P(X=1)=P(A_1)=p$ $P(X=2)=P(\bar{A}_1A_2)=(1-p)p$

$P(X=3)=P(\bar{A}_1\bar{A}_2A_3)=(1-p)^2p$ ……

可见 $P(X=k)=(1-p)^{k-1}\cdot p$ $k=1,2,\cdots$ 就是求所需射击发数 X 的分布律。

定义 若随机变量 X 的分布律为 $P(X=k)=(1-p)^{k-1}\cdot p$, $k=1,2,\cdots$ $0<p<1$

则称 X 服从参数为 p 的几何分布

离散型随机变量的分布函数

$P(X=x_k)=p_k \ (k = 1, 2, 3, \ldots)$

$$F(x)=P(X\leq x) = \sum_{x_k \leq x} P(X = x_k) = \sum_{x_k \leq x} p_k$$

$F(x)$在$x=x_k$处有跳跃，跳跃值为$p_k = P(X = X_k)$

如已知分布函数，也可求分布律：X的可能取值为$F(x)$的间断点，x_k，从而X的分布律为

$$p_k = P(X = X_k) = F(X_k + 0) - F(X_k - 0) = F(X_k) - F(X_k - 0)$$

例 设离散型随机变量X的分布函数为$F(x)=\begin{cases}0, x<-1\\ \frac{1}{4}, -1\le x<2\\ \frac{3}{4}, 2\le x<3\\ 1, x\ge 3\end{cases}$

试求随机变量X的分布律。

解 由于分布函数$F(x)$的分界点为 -1，2，3，因此随机变量X可能的取值为 -1， 2， 3

$$P(X=-1)=F(-1)-F(-1-0)=\frac{1}{4}$$

$$P(X=2)=F(2)-F(2-0)=\frac{1}{2}$$

$$P(X=3)=F(3)-F(3-0)=\frac{1}{4}$$

X	-1	2	3
P	1/4	1/2	1/4

本节回顾

离散型随机变量的分布律

设X的所有可能取值为x_k $(k = 1, 2, 3, \ldots)$，每个可能取值的概率即事件$\{X=x_k\}$的概率，记为

$$P(X=x_k)=P(e:\ X(e)=x_k)=p_k\ (k = 1, 2, 3, \ldots)$$

X	x_1	x_2	…	x_i	…
P	p_1	p_2	…	p_i	…

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix}$$

几种典型的离散型随机变量

0-1分布、二项分布、泊松分布、几何分布、超几何分布