§2.2　离散型随机变量及其分布

一　离散随机变量的分布列

设$X(\omega)$是概率空间$(\Omega,\mathbb{F},\mathrm{P})$上的随机变量．若集合

$$\{X(\omega):\omega\in\Omega\}=\{x_1,x_2,\cdots,x_n,\cdots\}$$

是有限集或可列集，则称是离散型随机变量．由于

$$\{\omega:X(\omega)\le x\}=\sum_{x_i\le x}\{\omega:X(\omega)=x_i\}$$

故只要给出了事件$\{\omega:X(\omega)=x_i\}$的概率$p(x_i)$，就可算出事件$\{\omega:X(\omega)\le x\}$的概率. 称

X	x_1	x_2	$\cdots\cdots$	x_n	$\cdots\cdots$
P	p_1	p_2	$\cdots\cdots$	p_n	$\cdots\cdots$

为离散型随机变量X的概率分布列——直观. 其中

（1） $p_k \geq 0$, 非负性；

（2） $\sum_k p_k = 1$. 正则性.

事实上： $\{\omega : X(\omega) = x_1\}, \{\omega : X(\omega) = x_2\}, \cdots\cdots$

都是随机事件，它们构成了对 Ω 的一个分割.

离散型随机变量的分布函数$F(x)$的特点：$F(x)$是一个跳跃函数，它在x_k处的跳跃度是p_k.

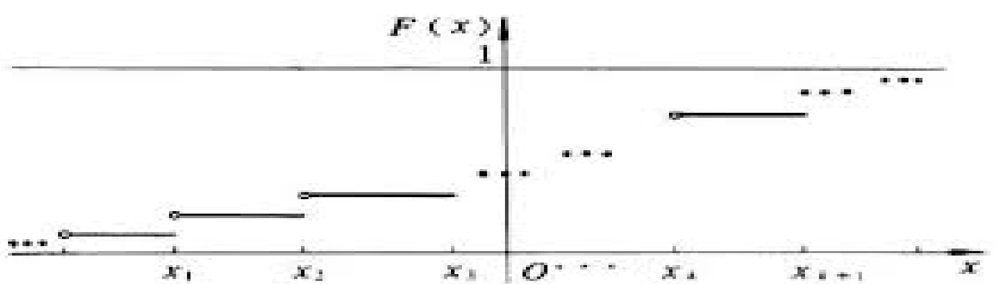


图 2 离散型随机变量的分布函数

对于离散型随机变量:

概率分布表 $\longleftrightarrow$ 分布函数

$\longrightarrow$ 由于 $\{\omega : X(\omega) \le x\} = \bigcup\limits_{k: x_k \le x} \{\omega : X(\omega) = x_k\}$

所以 $F(x) = P\{\omega : X(\omega) \le x\}$

$$= \sum_{k: x_k \le x} P\{\omega : X(\omega) = x_k\} = \sum_{k: x_k \le x} p_k$$

$\longleftarrow$ $P(X = x_k) = F(x_k) - F(x_k - 0)$

例1 设随机变量X具分布律如下表

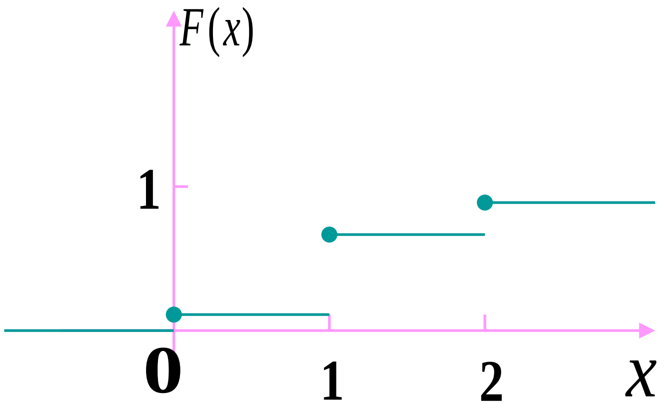
X	0	1	2
P	0.1	0.6	0.3

试求出X的分布函数。

解：

$$F(x)=P\{X\le x\}$$

$$=\begin{cases}0 & ,x<0\\ 0.1, & 0\le x<1\\ 0.7, & 1\le x<2\\ 1 & ,x\ge 2\end{cases}$$



求离散随机变量的分布列:

(1)确定随机变量的所有可能取值;

(2)计算每个取值点的概率.

求离散随机变量的分布函数:

(1)$F(x)$是递增的阶梯函数;

(2)其间断点均为右连续的;

(3)其间断点即为X的可能取值点;

(4)其间断点的跳跃高度是对应的概率值.

二、常见离散型随机变量的概率分布

1. 退化分布

若随机变量X取常数值C的概率为1，即

$$P(X=C)=1$$

则称X服从退化分布.

2. 两点分布

设随机变量X只可能取0与1两个值,它的分布律为

X	0	1
P	1-p	p

则称X服从两点分布或$(0-1)$分布.

两点分布是最简单的一种分布，任何一个只有两种可能结果的随机现象，如新生婴儿是男还是女、明天是否下雨、种籽是否发芽等，都属于两点分布.

3 二项分布

在n重贝努利试验中，若事件A出现的次数记为X，则随机变量X可能取值为$0, 1, 2, \cdots, n$. 相应的概率分布列为

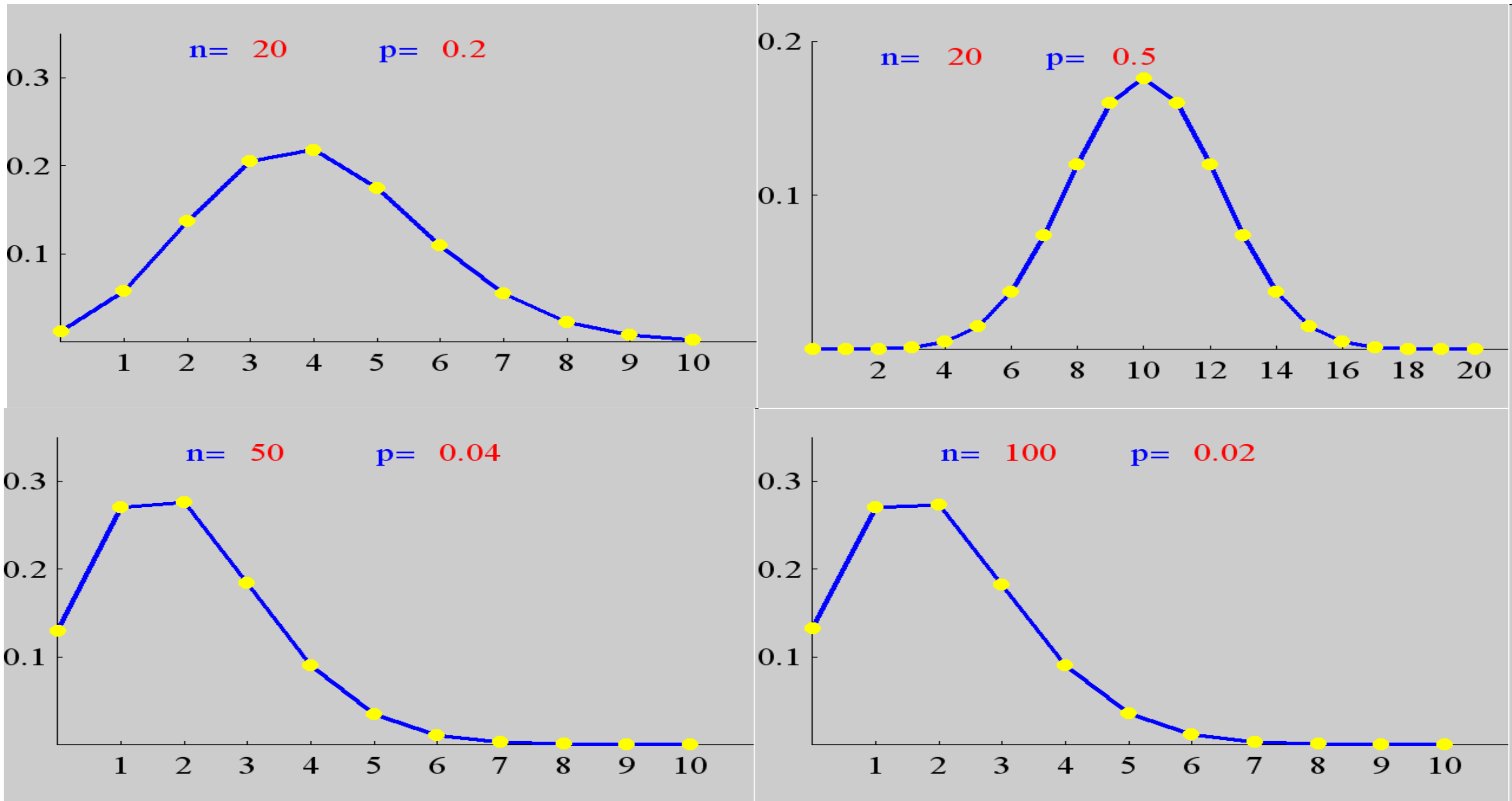
$$P(X=k)=C_n^k p^k q^{n-k}, \quad 0 \le k \le n.$$

称为参数为n, p的二项分布，记作 $X \sim B(n, p)$.

其中 $p=P(A), \ q=1-p$.

注：当$n=1$时，$X \sim B(1, p)$，即是两点分布.

二项分布的图形



例2 按规定，某种型号的电子元件的使用寿命超过2000小时为一等品. 已知一大批该种产品的一等品率为0.2，现从中随机地抽取20件，问20件产品中有k件一等品的概率是多少？

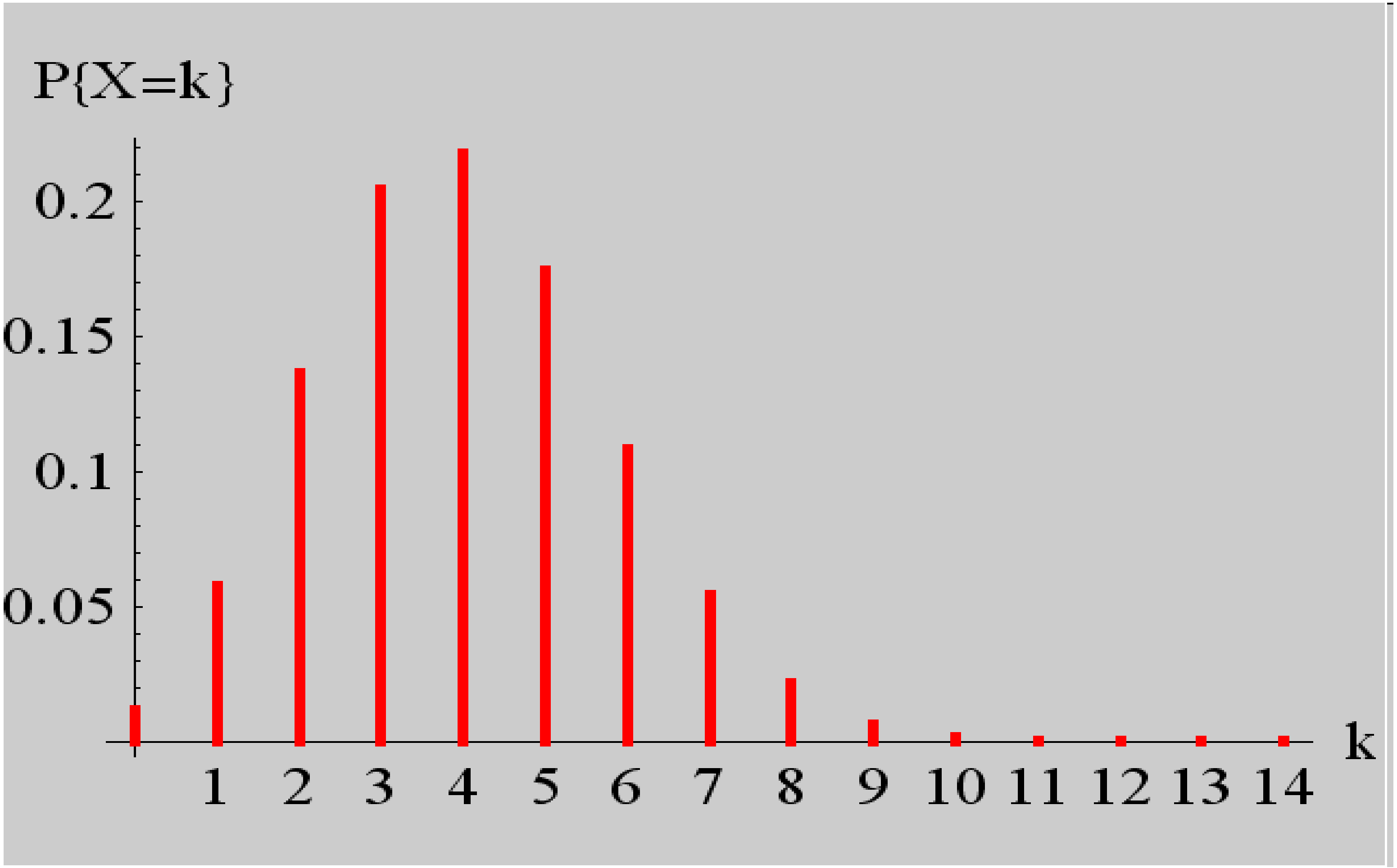
解：设X为20件产品中一级品的件数，则

$$X \sim B(20,\ 0.2).$$

$$P\{X=k\}=C_{20}^{k}(0.2)^{k}(0.8)^{20-k},\quad k=0,1,\cdots,20.$$

$P\{X=0\}=0.012$	$P\{X=4\}=0.218$	$P\{X=8\}=0.022$
$P\{X=1\}=0.058$	$P\{X=5\}=0.175$	$P\{X=9\}=0.007$
$P\{X=2\}=0.137$	$P\{X=6\}=0.109$	$P\{X=10\}=0.002$
$P\{X=3\}=0.205$	$P\{X=7\}=0.055$	
$P\{X=k\}<0.001$，当$k\geq 11$时		

概率分布图示



二项分布中最可能出现次数

若 $P(X=k)=\max\left\{P(X=j):0\le j\le n\right\}$

则称 k 为**最可能出现次数.**

当 $(n+1)p\neq$ 整数时, 在 $k=[(n+1)p]$ 处的概率取得最大值.

当 $(n+1)p=$ 整数时, 在 $k=(n+1)p$ 与 $(n+1)p-1$ 处的概率取得最大值.

例3: 某人进行射击练习，假设他每次射击的命中率为0.02，现独立射击400次，试求命中目标的概率.

解：设命中目标的次数为X，则$X \sim B(400, 0.02)$

$$P\{X=k\}=C_{400}^{k}(0,02)^{k}(0.98)^{400-k},$$

$$k=0,1,2,\cdots,400.$$

于是，所求概率为

$$P\left\{X \geq 1\right\}=1-P\left\{X=0\right\}=1-0.98^{400} \approx 0.9997.$$

注：随着试验次数的增加，小概率事件发生的可能性也将增加，即小概率事件不再小概率.

若在400次的射击中, 没有一次击中目标, 根据实际推断原理, 我们有理由怀疑原假设(命中率为0.02).

二项分布的泊松近似

先观察一个随机试验：从集合$\Omega=\{1,2,\cdots,n\}$随机的取一个数，记A表示事件"取出的数不超过2". 则

$$P(A)=\frac{2}{n}\triangleq p_n \quad (n\geq 2),$$

今将该试验独立重复n次，得到n重伯努利试验.

记X_n表示事件A在n重伯努利试验中发生的次数, 则

$$X_n \sim B(n,p_n),\ \ P(X_n=k)=C_n^k p_n^k(1-p_n)^{n-k}.$$

当n较大时, 上述概率的计算是相当复杂的.

上述二项分布的特点是：n较大，p_n较小，而且有

$$\lim_{n\to+\infty} np_n = 2 > 0.$$

法国数学家泊松（*Poisson*）研究了这类二项分布概率的极限问题，得到了其概率的一种近似计算方法.

定理1（泊松定理） 设$X_n \sim B(n, p_n)$，若当$n \to \infty$时，$np_n \to \lambda(\lambda > 0$常数）. 则对固定的$k$，$k = 0,1,2,\cdots$，有

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

证明： 记$np_n = \lambda_n$，则

$$C_n^k p_n^k (1-p_n)^{n-k}$$

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

$$= \frac{\lambda_n^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda_n}{n}\right)^{n-k}$$

任意固定的k，有 $\lim\limits_{n\to\infty}\lambda_n^{\ k} = \lambda^k$

又$\lim\limits_{n\to\infty}(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{k-1}{n}) = 1,$

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

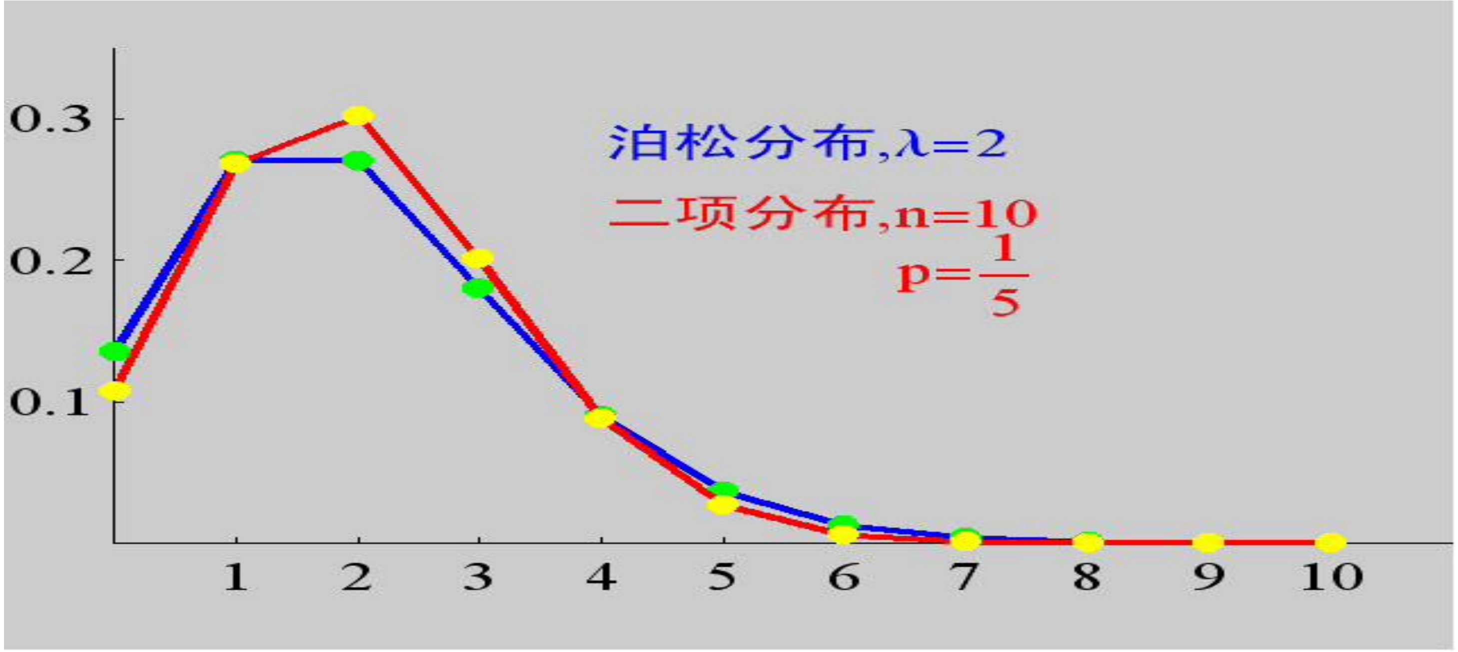
所以 $\lim\limits_{n\to\infty}P\{X_n = k\} = \frac{\lambda^k}{k!}e^{-\lambda}.$

故当n 较大，p较小，np大小适中时：

$$C_n^k p_n^k (1-p_n)^{n-k} \approx \frac{(np)^k}{k!}e^{-np}.$$

泊松定理表明

$$\text{二项分布} \xrightarrow{n\text{很大}, p\text{ 很小}} \text{泊松分布}$$



4 泊松分布

若随机变量X的分布列为

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}, \quad k=0,1,2,\cdots$$

则称X服从参数为λ的泊松分布. 记为$X\sim P(\lambda)$.

背景：泊松分布主要用于估计某事件在特定时间或空间中发生的次数，如：

①社会服务问题：电话交换台中的呼叫数、公共汽车的乘客数；

②物理学：放射性分裂落在某区间的质点数；

③………

【泊松过程】考虑某交换装置的电话呼叫数，假设它满足下面三个性质

(1) 平稳性　在时间$[t_0, t_0+t)$内来到的呼叫数只与时间长度t有关而与时间起点t_0无关. 以$P_k(t)$表示长度为t的时间内来到k个呼叫的概率. 有

$$\sum_{k=0}^{\infty} P_k(t) = 1$$

对任意$t > 0$成立.

平稳性表示过程的概率规律不随时间的推移而改变.

(2) 独立增量性(无后效性)　在$[t_0, t_0+t)$时间段内来k个呼叫这一事件与时刻t_0以前的事件独立，此表明在不相交的区间内发生的事件是相互独立的.

(3) 普通性　在充分小的时间间隔内，最多来到一个呼叫．若记

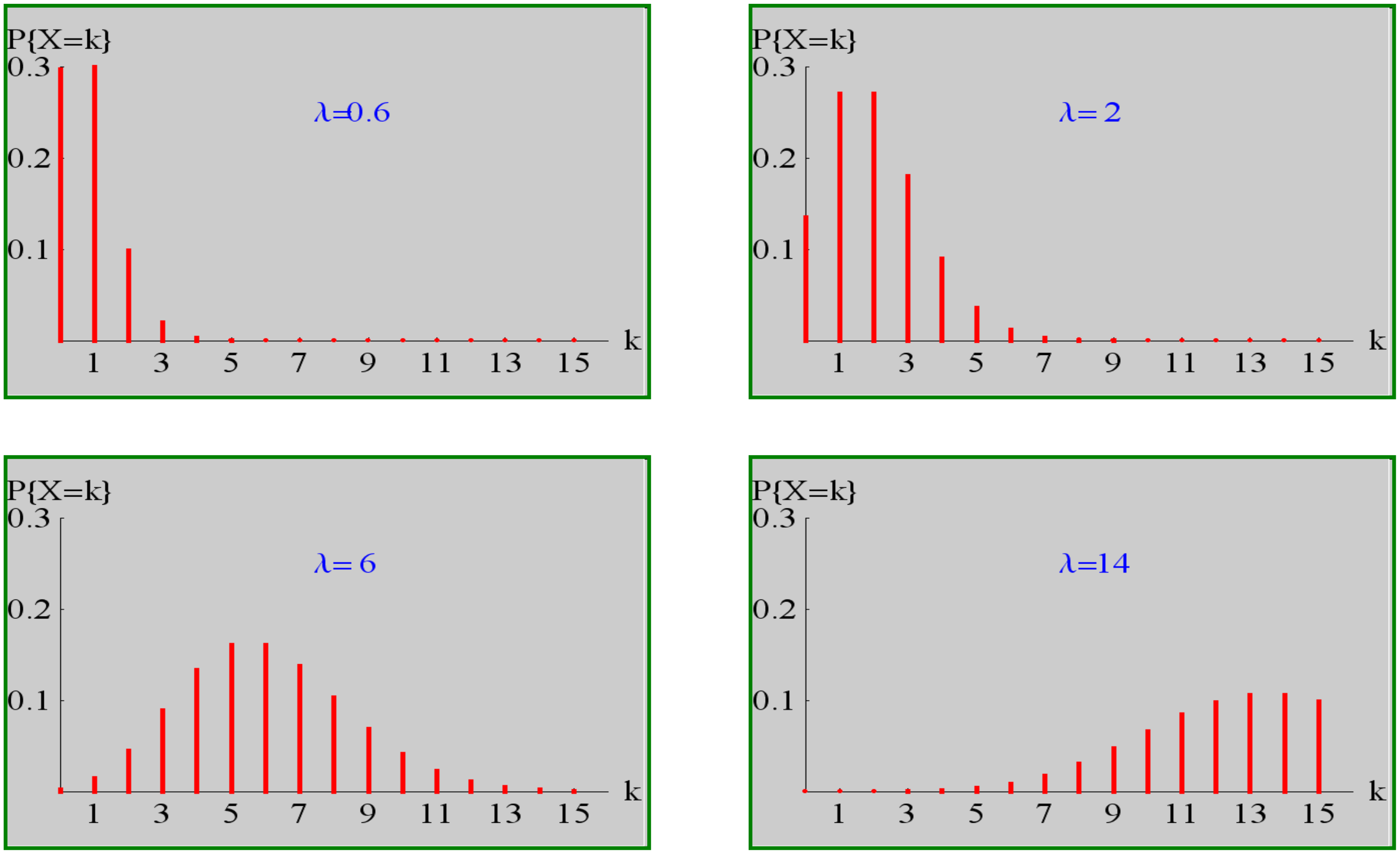
$$\varphi(t)=\sum_{k=2}^{\infty}P_k(t)=1-P_0(t)-P_1(t)$$

应有$\varphi(t)=o(t)$.　此表明在同一时间瞬间来两个或两个以上呼叫是不可能的.

【泊松分布的推导】 利用全概率公式，一阶微分方程

$$P_k(t)=\frac{(at)^k}{k!}e^{-at},\qquad k=0,1,2,\cdots$$

泊松分布的图形



P{X=k}
0.3
0.2
0.1
λ=0.6
k
1 3 5 7 9 11 13 15
P{X=k}
0.3
0.2
0.1
λ= 2
k
1 3 5 7 9 11 13 15
P{X=k}
0.3
0.2
0.1
λ= 6
k
1 3 5 7 9 11 13 15
P{X=k}
0.3
0.2
0.1
λ=14
k
1 3 5 7 9 11 13 15

例4：假设书的某一页上印刷错误的个数服从参数为0.5的泊松分布，求在这一页上至少有一处印刷错误的概率.

解：设X表示一页书上印刷错误的个数，则

$$X \sim P(0.5).$$

因此

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - \mathrm{e}^{-1/2} \approx 0.393.$$

5、几何分布

若随机变量X的分布律为

$$P\{X=k\}=q^{k-1}p,$$

其中$q=1-p,\quad k=1,2,\cdots.$

则称X服从几何分布。记为$X\sim G(p)$.

注：在伯努利试验中，事件A第1次发生时的试验次数X服从几何分布.

背景：放回抽样。

例5 设箱中有N个白球与M个黑球，每次随机取一个球，直到取出黑球为止．如果每取出一个球后立即放回，再取出一个球，试求下列概率：

(1) 正好需要取n次；?(2)?至少需要取k次.

解： 令X表示取到黑球所需的次数，则

$$P\{X=n\}=\left(\frac{N}{N+M}\right)^{n-1}\frac{M}{N+M}=\frac{MN^{n-1}}{(N+M)^n},$$

$$P\{X\ge k\}=\sum_{n=k}^{\infty}\left(\frac{N}{N+M}\right)^{n-1}\frac{M}{N+M}=\left(\frac{N}{N+M}\right)^{k-1}.$$

6. 超几何分布

若随机变量X的分布律为

$$P\{X=k\}=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n},$$

$$k=0,1,2,\cdots,\min\{n,M\}.$$

则称X服从超几何分布. 记为 $X \sim H(M,N,n)$.

背景：产品检验和药物试验等实际问题.