概率论与数理统计

第二章　随机变量及其分布

§2　离散型随机变量及其分布律

◆离散型随机变量的分布律

◆常见离散型随机变量的概率分布

一、离散型随机变量的分布律

设离散型随机变量 X所有可能取的值为

$$x_k\,(k=1,2,\cdots),$$

X取各个可能值的概率，即事件$\{X=x_k\}$的概率,为

$$P(X=x_k)=p_k \quad k=1,2,\cdots.$$

称此为离散型随机变量 X 的分布律.

说明：由概率的定义,p_k满足如下两个条件：

1° $p_k \geq 0$，$k=1,2,\cdots;$

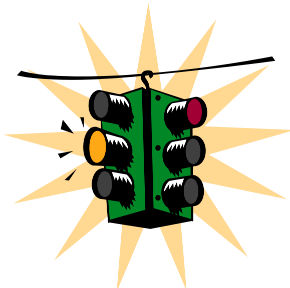
2° $\sum_{k=1}^{\infty} p_k = 1.$

分布律也可以用表格的形式来表示:

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

表格直观地表示了随机 变量X 取各个值的概率的规律. X取各个值各占一些概率 ，这些概率合起来是1. 可以想象成：概率1以一定的规律分布在各个可能值上.

例1 设一汽车在开往目的地的道路上需经过4组信号灯，每组信号灯以1/2的概率允许或禁止汽车 通过. 以X表示汽车首次停下时 ,它已通过的信号灯组数, 假设各组信号灯的工作是相互独立的, 求X分布律 .



解 以p表示每组信号灯禁止汽 车通过的概率 .

易知X的分布律为

X	0	1	2	3	4
p_k	p	$(1-p)p$	$(1-p)^2 p$	$(1-p)^3 p$	$(1-p)^4$

或写成

$$P\{X=k\}=(1-p)^k p, \quad k=0,1,2,3,$$

$$P\{X=4\}=(1-p)^4.$$

以$p=1/2$代入得

X	0	1	2	3	4
p_k	0.5	0.25	0.125	0.0625	0.0625

二、常见离散型随机变量的概率分布

(一) (0—1)分布

设随机变量 X 只可能取 0与1两个值，其分布是

$$P\{X=k\}=p^k(1-p)^{1-k},\ k=0,1\quad (0<p<1),$$

则称X服从以p为参数的$(0-1)$分布或两点分布．

(0—1)分布的分布律也可写成

X	0	1
p_k	$1-p$	p

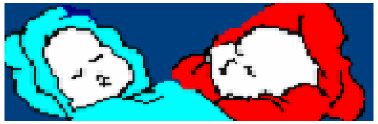
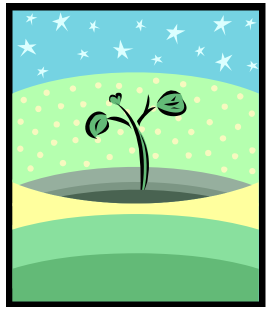
实例　“抛硬币”试验,观察正、反两面情况.

$$X = X(e) = \begin{cases} 0, & \text{当} e = \text{正面}, \\ 1, & \text{当} e = \text{反面}. \end{cases}$$

随机变量X服从(0—1)分布.

其分布律为

X	0	1
p_k	$\frac{1}{2}$	$\frac{1}{2}$

对于一个随机试验，如果它的样本空间只包含两个元素，即$S=\{e_1,e_2\}$，我们总能在 S上定义一个服从(0—1)分布的随机变量

$$X=X(e)=\begin{cases}0, & \text{当}e=e_1, \\ 1, & \text{当}e=e_2.\end{cases}$$

来描述这个随机试验的结果.

(二) 伯努利试验、二项分布

伯努利资料

设试验E只有两个可能结果 : A及$\overline{A}$,则称E为**伯努利(Bernoulli)实验.** 设$p(A)=p(0<p<1)$, 此时$P(\overline{A})=1-p$.将E独立重复进行 n次 ,则称这一串重复独立的试验为 n重伯努利试验 .

n重伯努利试验是一种非常重要的数学模型，它有广泛的应用, 是研究最多的模型之一.

实例1 抛一枚硬币观察得到正面或反面. 若将硬币抛 n 次,就是n重伯努利试验.

实例2 抛一颗骰子n次,观察是否 “出现 1 点”, 就是 n重伯努利试验.

二项概率公式

若 X 表示 n 重伯努利试验中事件 A 发生的次数 , 则 X 所有可能取的值为

$$0,\quad 1,\quad 2,\quad \cdots,\quad n.$$

当 $X = k\ (0 \le k \le n)$ 时，即 A 在 n 次试验中发生 k 次.

$$\underbrace{A\,A\cdots A}_{k\text{次}}\ \underbrace{\overline{A}\,\overline{A}\cdots\overline{A}}_{n-k\text{ 次}},$$

$$\underbrace{A\,A\cdots A}_{k-1\text{ 次}}\ \overline{A}\ \ A\ \underbrace{\overline{A}\,\overline{A}\cdots\overline{A}}_{n-k-1\text{ 次}}\qquad \cdots\cdots$$

得 A 在 n 次试验中发生 k 次的方式共有 $\binom{n}{k}$ 种，

且两两互不相容.

因此 A在 n 次试验中发生 k 次的概率为

$$\binom{n}{k} p^k (1-p)^{n-k} \xrightarrow{\text{记 } q=1-p} \binom{n}{k} p^k q^{n-k}$$

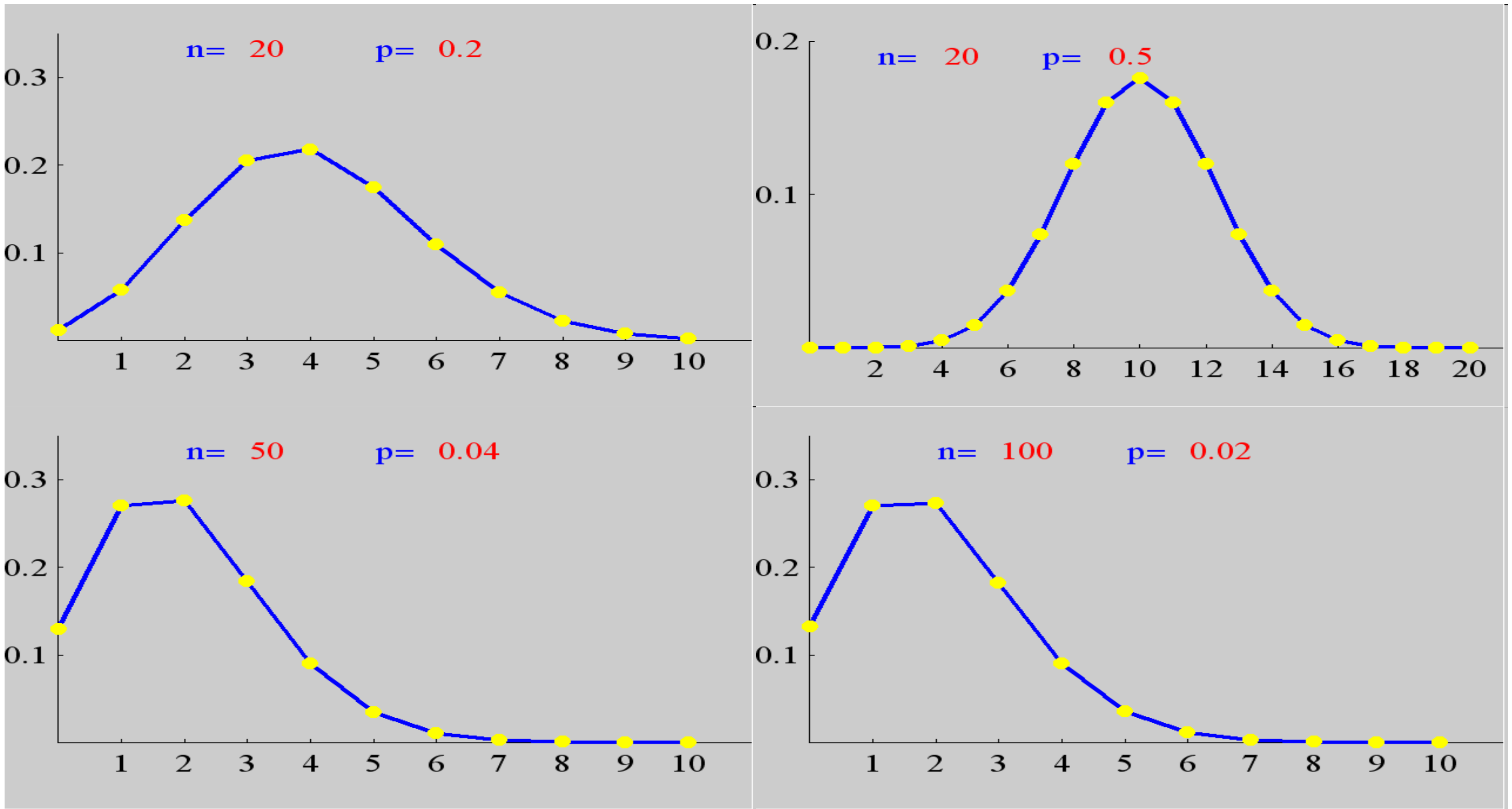
得 X 的分布律为

X	0	1	$\cdots$	k	$\cdots$	n
p_k	q^n	$\binom{n}{1} pq^{n-1}$	$\cdots$	$\binom{n}{k} p^k q^{n-k}$	$\cdots$	p^n

称这样的分布为二项分布.记为 $X \sim b(n,p)$.

$$\text{二项分布} \xrightarrow{n=1} \text{两点分布}$$

二项分布的图形



n= 20 p= 0.2
0.3
0.2
0.1
1 2 3 4 5 6 7 8 9 10
n= 20 p= 0.5
0.2
0.1
2 4 6 8 10 12 14 16 18 20
n= 50 p= 0.04
0.3
0.2
0.1
1 2 3 4 5 6 7 8 9 10
n= 100 p= 0.02
0.3
0.2
0.1
1 2 3 4 5 6 7 8 9 10

例如　在相同条件下相互独立地进行 5 次射击,每次射击时击中目标的概率为 0.6 ,则击中目标的次数 X 服从 b (5,0.6) 的二项分布.

X	0	1	2	3	4	5
p_k	$(0.4)^5$	$\binom{5}{1}0.6\cdot 0.4^4$	$\binom{5}{2}0.6^2\cdot 0.4^3$	$\binom{5}{3}0.6^3\cdot 0.4^2$	$\binom{5}{4}0.6^4\cdot 0.4$	0.6^5

例2　按规定，某种型号的电子元件的使用寿命超过1500小时的为一等品．已知某一大批产品的一等品率为0.2，现在从中随机抽查20只．问20只元件中恰有k只$(k=0,1,2,\cdots,20)$为一级品的概率是多少　？

解　这是不放回抽样．但由于这批元件的总数很大，且抽查元件的数量相对于元件的总数来说又很小，因而此抽样可近似当作放回抽样来处理，这样做有一些误差，但误差不大．我们把检查一只会元件看它是否为一等品看成是一次试验，检查20只元件相

当于做20重伯努利试验. 以 X 记 20 只元件中一级品的只数, 那么, X是一个随机变量 , 且有

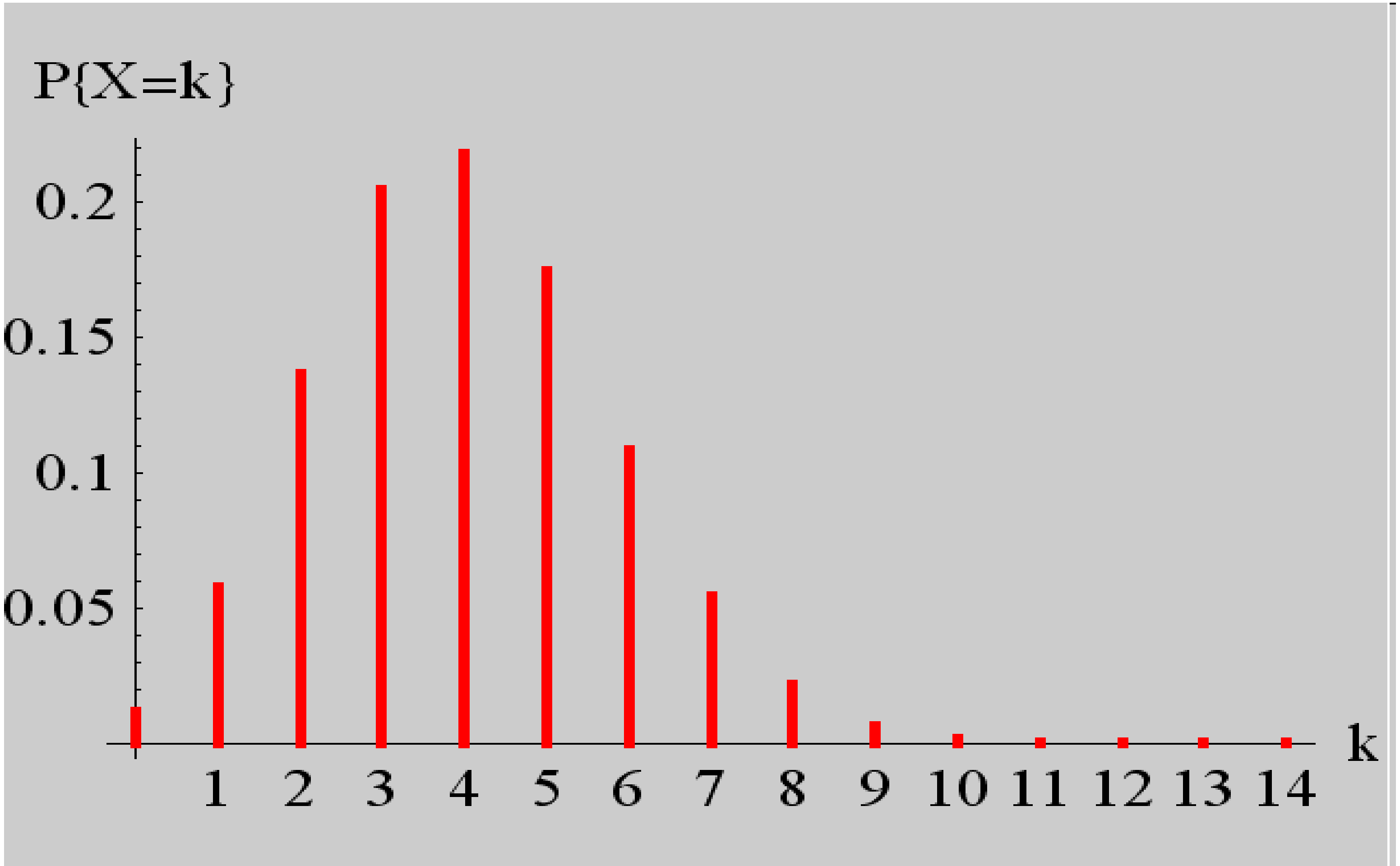
$X \sim b(20, 0.2)$. 所求概率为

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

将计算结果列表如下:

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

图示概率分布



例3 某人进行射击, 假设每次射击的命中率为0.02, 独立射击400次, 试求至少击中两次的概率.

解 将一次射击看成是一次试验. 设击中的次数为 X, 则 $X \sim b(400, 0.02)$. X 的分布律为

$$P\{X=k\}=\binom{400}{k}(0.02)^k(0.98)^{400-k},\ k=0,1,\cdots,400.$$

于是所求概率为

$$P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\}$$

$$= 1 - (0.98)^{400} - 400(0.02)(0.98)^{399} = 0.9972 .$$

结果的实际意义:

1. 决不能轻视小概率事件.

2.由实际推断原理, 我们怀疑"每次射击命中率为0.02"这一假设, 认为该射手射击的命中率不到0.02

例4　设有80台同类型设备，各台工作是相互独立的，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人维护，每人负责20台；其二是由3人共共同维护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的大小．

解　按第一种方法，以X记“第1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 20台中发生故障不能及时　维修”，

则知80台中发生故障而不能及时维修的概率为

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) \geq P(A_1) = P\{X \geq 2\}.$$

而$X \sim b(20,0.01)$，故有

$$P\{X \geq 2\} = 1 - \sum_{k=0}^{1} P\{X = k\}$$

$$= 1 - \sum_{k=0}^{1} \binom{20}{k} (0.01)^k (0.99)^{20-k}$$

$$= 0.0169.$$

按第二种方法, 以Y记80台中同一时刻发生故障的台数, 此时, $Y \sim b(80,0.01)$, 故80台中发生故障而不能及时维修的概率为

$$P\{Y \geq 4\}=1-\sum_{k=0}^{3}\binom{80}{k}(0.01)^{k}(0.99)^{80-k}=0.0087 .$$

我们发现, 在后一种情况尽管任务重了(每人平均维护约27台), 但工作效率不仅没有降低, 反而提高了.

(三)　泊松分布

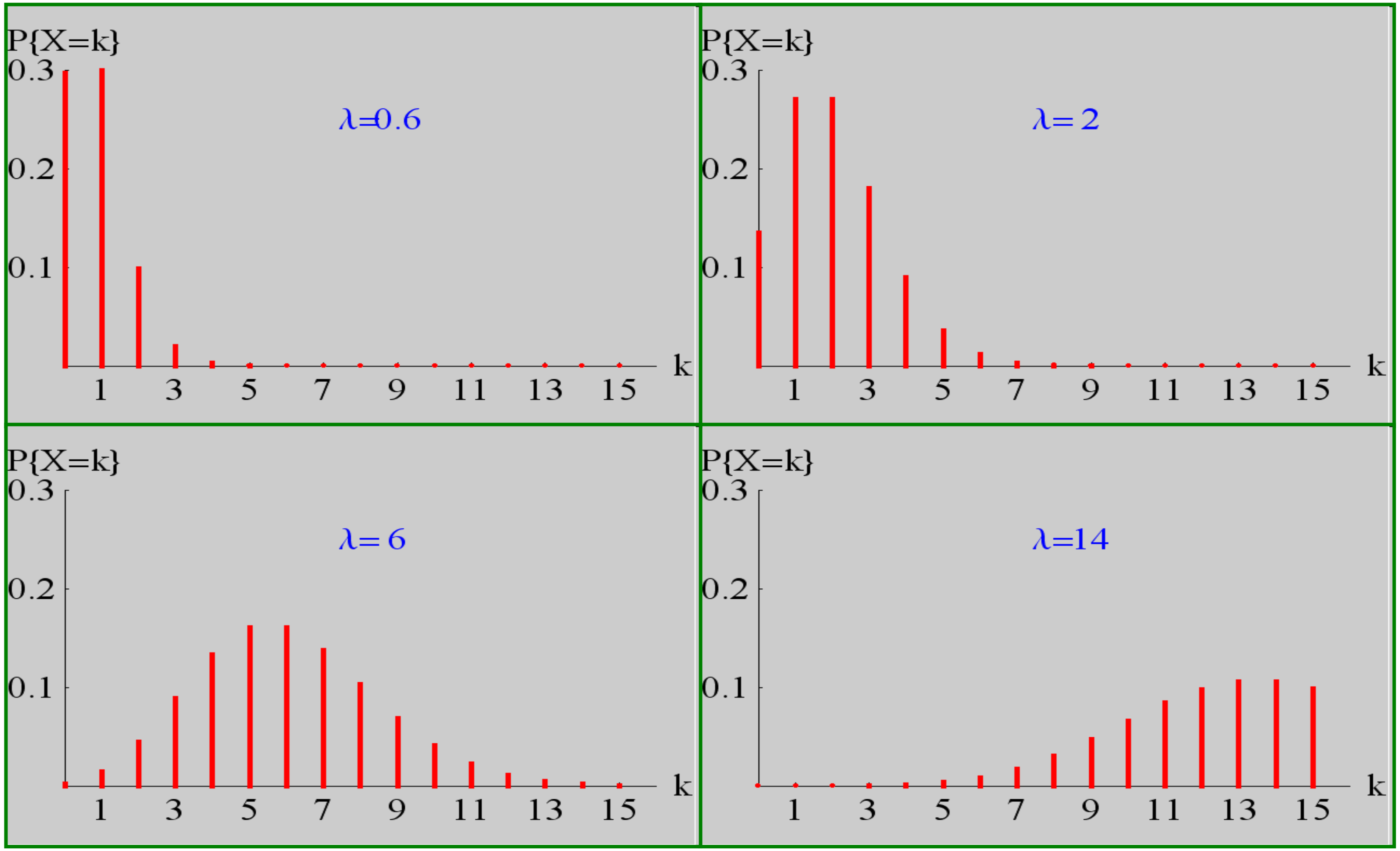
泊松资料

设随机变量 X 所有可能取的值为 $0,1,2,\cdots$，而取各个值的概率为

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

其中$\lambda>0$是常数，则称X服从参数为λ的泊松分布，记为$X\sim\pi(\lambda)$.

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

泊松分布的背景及应用

二十世纪初卢瑟福和盖克两位科学家在观察与分析放射性物质放出的粒子个数的情况时, 他们做了**2608**次观察(每次时间为**7.5**秒), 发现放射性物质在规定的一段时间内, 其放射的粒子数**X**服从泊松分布.

地震

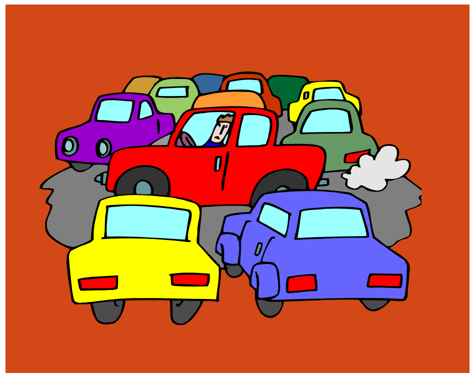
火山爆发

特大洪水

商场接待的顾客数

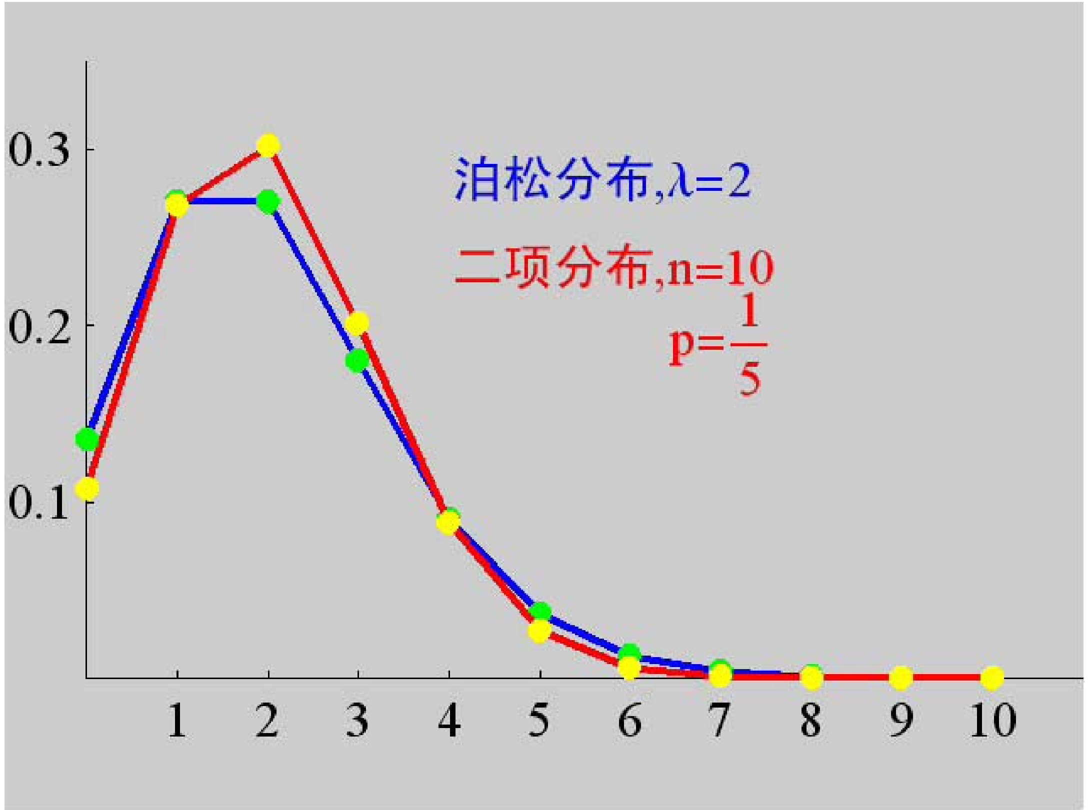
电话呼唤次数

交通事故次数



上面我们提到

二项分布 $\xrightarrow{np \to \lambda (n \to +\infty)}$ 泊松分布



泊松定理 设$\lambda > 0$是一个常数，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

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

证 由$p_n=\dfrac{\lambda}{n}$，有

$$\binom{n}{k}p_n^k(1-p_n)^{n-k}$$

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

$$\binom{n}{k} p_n^k (1-p_n)^{n-k}$$

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

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

对于固定的 k，当$n\to\infty$时

$$1\cdot\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\to 1,$$

$$\left(1-\frac{\lambda}{n}\right)^{n}\to \mathrm{e}^{-\lambda}, \quad \left(1-\frac{\lambda}{n}\right)^{-k}\to 1.$$

故有

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

定理的条件 $np_n=\lambda$(常数)意味着当 n 很大时 p_n 必定很小, 因此, 上述定理表明当 n 很大, p 很小 ($np=\lambda$)时有以下近似式

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!} \quad (其中\lambda = np).$$

也就是说以 n, p 为参数的二项分布的概 率值可以由参数为 $\lambda = np$ 的泊松分布的概率值近 似． 上式也能用来作二项分布概率的近似计算.

例5　计算机硬件公司制造某种特殊型号的微型芯片，次品率达0.1%，各芯片成为次品相互独立．求在1000只产品中至少有2只次品的概率．以X记产品中的次品数，$X \sim b(1000, 0.001)$.

解　所求概率为

$$P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\}$$

$$= 1 - (0.999)^{1000} - \binom{1000}{1}(0.999)^{999}(0.001)$$

$$\approx 1 - 0.3676954 - 0.3680635 \approx 0.2642411$$

利用近似计算得： $\lambda=1000\times0.001=1,$

$$P\{X\geq2\}=1-P\{X=0\}-P\{X=1\}$$

$$=1-\mathrm{e}^{-1}-\mathrm{e}^{-1}\approx0.2642411$$

显然利用近似计算来得方便.

一般，当$n\geq20$，$p\leq0.05$时采用近似计算：

用$\dfrac{\lambda^k e^{-\lambda}}{k!}$作为$\dbinom{n}{k}p^k(1-p)^{n-k}$的近似值颇佳．

(四) 几何分布

若随机变量 X 的分布律为

X	1	2	$\cdots$	k	$\cdots$
p_k	p	qp	$\cdots$	$q^{k-1}p$	$\cdots$

$, p+q=1,$

则称 X 服从几何分布.

实例　设某批产品的次品率为 p,对该批产品做有放回的抽样检查，直到第一次抽到一只次品为止 (在此之前抽到的全是正品), 那么所抽到的产品数 X 是一个随机变量 , 求X 的分布律.

解　X 所取的可能值是 $1, 2, 3, \cdots.$

设 A_i 表示“抽到的第 i 个产品是正品”,

$$P\{X=k\}=P(A_1A_2\cdots A_{k-1}\overline{A_k})$$

$$=P(A_1)\cdot P(A_2)\cdot\ \cdots\ \cdot P(A_{k-1})\cdot P(\overline{A_k})$$

$$=\underbrace{(1-p)(1-p)\cdot\ \cdots\ \cdot(1-p)}_{(k-1)}\cdot p=q^{k-1}p.$$

所以 X 服从几何分布.　　$(k=1,2,\cdots)$

说明　几何分布可作为描述某个试验“首次成功”的概率模型.

小结 1. 离散型随机变量的分布

$$\begin{cases} \text{两点分布} \\ \text{二项分布} \\ \text{泊松分布} \\ \text{几何分布} \end{cases}$$

两点分布

$\uparrow\; n = 1$

二项分布

$\downarrow\; n > 10, p < 0.1$

泊松分布

以 $n, p\ (np = \lambda)$ 为参数的二项分布，当 $n \to \infty$ 时趋于以 λ 为参数的泊松分布，即

$$P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{(np)^k}{k!} e^{-np},$$

$$(k = 0, 1, 2, \cdots, n).$$

作业：课后习题 2、5、7、9

伯努利资料

Jacob Bernoulli

Born: 27 Dec 1654 in Basel, Switzerland
Died: 16 Aug 1705 in Basel, Switzerland

泊松资料

Siméon Poisson

Born: 21 June 1781 in Pithiviers, France
Died: 25 April 1840 in Sceaux (near Paris), France