第二章 随机变量及其分布

§2.1 随机变量

§2.2 离散型随机变量的概率分布

§2.3 随机变量的分布函数

§2.4 连续型随机变量的概率密度

§2.5 随机变量的函数的分布

§2.1 随机变量

我们知道随机事件是由基本事件构成的，前面所给出的定义无论是基本事件还是随机事件都是用文字叙述给出，这有两个缺憾:一是非常繁琐，二是尽管事件可以看成子集（样本空间的子集）但是文字叙述却不符合数学的研究特点，因此为了更深入地研究随机现象，我们就需要将随机试验的结果数量化，也就是用某一变量取得各种不同的数值来描述随机试验的结果，这样就引进了<u>随机变量</u>的概念.

例1

考察抛硬币试验 $S=\{H, T\}$，其中$H=\{$出现正面$\}$, $T=\{$出现反面$\}$.

用数字 "1" 代表事件"出现正面",

用数字 "0" 代表事件"出现反面",

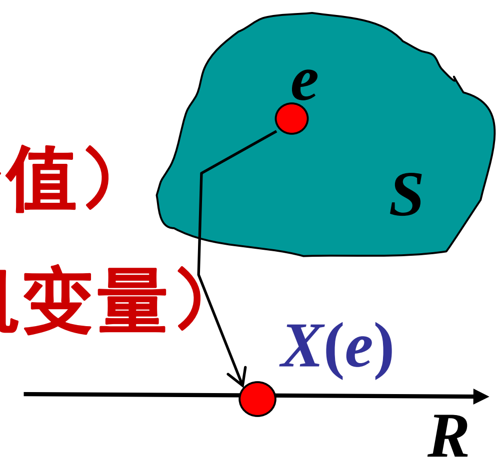
则构造随机变量X：$S\to\{0,1\}$，即$X(H)=1$, $X(T)=0$

此时，随机变量 X 随基本事件的变化而变化，当基本事件确定，对应值 X 也相应确定.

定义 设E是随机试验, 其样本空间是$S=\{e\}$, 如果对于每一个 $e\in S$，都有一个实数 $X(e)$ 与之对应，这样就得到一个定义在 S 上的单值实值函数 $X=X(e)$, 称为**随机变量**. 常用字母 X, Y, ξ, η 等表示随机变量.

随机变量是定义在样本空的实值集函数,它与普通的实函数有本质的区别. 一方面它的取值是随机的,而它取每一个可能值都有一定的概率；另一方面它的定义域是样本空间S, 而S不一定是实数集.

分类
- 离散型随机变量(取有限个或可数个值)
- 非离散型的随机变量（如连续型随机变量）



随机变量的概念在概率论与数理统计中既是基本的，又是非常重要的. 后面将会看到, 由于引入了随机变量, 高等数学的方法就可用来研究随机现象了.

例2 一袋中有6个球，分别标有 1, 2, 2, 2, 3, 3， 从袋中任取一个球，观察出现的数字.

解： 样本空间 $S=\{e_1, e_2, e_3\}$,

其中$e_i=\{$出现数字$i\}$，$i=1, 2, 3$

构造随机变量X：$S\to\{1, 2, 3\}$,

即 $X(e_1)=1$， $X(e_2)=2$， $X(e_3)=3$.

❖ **当试验的可能结果本身是用数量描述的，这时构造随机变量最容易.**

例3 对于一批灯泡, 设每一灯泡在某固定条件的耐用时间为X，则 $X:S\to[0,+\infty)$

随着取不同灯泡的试验结果不同，X 取不同的值. 取定灯泡，X 值才能确定，故X是随机变量.

例4 某射手每次射击打中目标的概率是$p\,(0<p<1)$, 现在他连续向一目标射击, 直到第一次击中目标为止,

则**射击次数**X是一个随机变量, X可以取到一切自然数.

✵ 定义了随机变量后, 随即事件就可以用随机变量的取值来描述.

小结

随机变量是定义在样本空间上的一种特殊的函数.

与普通的函数不同，随机变量的取值具有一定的概率规律.

§2.2离散型随机变量的概率分布

一、基本概念

❖**离散型随机变量：** 随机变量所有可能取值是有限个或可列个.

❖**离散型随机变量的分布律：**

定义 设离散型随机变量X所有可能取值为x_k, k=1,2,…
X取各个可能值的概率为 p_k， 即

$$P\{X=x_k\}=p_k, \quad k=1, 2, \ldots \qquad (*)$$

称(*)式为离散型随机变量 X 的**概率分布**或**分布律**.

❖**分布律常用表格表示：**

X	x_1	x_2	…	x_k	…
p_k	p_1	p_2	…	p_k	…

❖**分布律具有性质：**

1. $p_k \geq 0$, (k=1,2,...)，

2. $\sum_{k=1}^{\infty} p_k = 1$

例1 设一汽车在开往目的地的道路上需经过四盏信号灯,每盏信号灯以概率 p 禁止汽车通过. 以 X 表示汽车首次停下时, 它已通过的信号灯的盏数(设各信号灯的工作是相互独立的), 求X 的分布律.

解: X 的可能取值为 0, 1, 2, 3, 4， 故 X的分布律为

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

或写成

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

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

例2 设袋中有4个红球, 1个白球, 今从袋中随机抽取两次, 每次取一个, 设 X 表示所取得的白球数，试分两种情况：(1) 放回抽取; (2)不放回抽取，分别求出 X 的分布律.

解: (1) 放回

X	0	1	2
p_k	$\frac{4\cdot 4}{5\cdot 5}$	$\frac{4+4}{5\cdot 5}$	$\frac{1}{5\cdot 5}$

(2)不放回

X	0	1
p_k	$\frac{4\cdot 3}{5\cdot 4}$	$\frac{4+4}{5\cdot 4}$

离散型随机变量 X 的概率分布或分布律完全刻划了离散型随机变量的分布情况，已知 X 的概率分布，可以求得这个随机变量 X 所对应的样本空间中任何随机事件的概率.

作业

- **第二章习题2, 5, 7, 10, 12, 15**

二、几种常见的离散型随机变量

1. (0−1) 分布

如果随机变量X只能取 0, 1 两个值, 其分布律为

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

即

X	0	1
p_k	$1-p$	p

或

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

则称X服从**参数为 p** 的 **(0−1) 分布**或**两点分布**.

对于一次试验只有两种可能结果的概率分布都可用两点分布来描述. 如在射击中，只考虑“击中”与“不中”； 对产品质量进行检验，如果我们只关心“合格” 与“不合格, 则这类问题都可以归结为两点分布.

2. 二项分布

(1) **贝努利试验**：只有两个可能结果A及$\overline{A}$的试验.

将 E 独立地重复进行n次, 则称这一串重复的独立试验为**(n重)贝努利试验**.

(2) **二项分布**：设 X 表示 n 重贝努利试验中A事件发生的次数, $P(A)=p$, 则X是一个随机变量, X 的可能值为 0, 1, 2, …, n. X 的分布率为

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

则称X 服从参数为 n, p 的**二项分布**, 记作 $X\sim b(n,p)$.

注 (1) 显然 $P\{X=k\}\geq 0$，$k=0,1,2,\ldots,n$ ； $\sum_{k=0}^{n} C_n^k p^k q^{n-k}=1$

(2)当n=1时二项分布化为(0−1)分布

例3 设有20台机床, 独立地各加工一件齿轮, 若各机床加工的废品率都是0.2, 求20件齿轮产品中的废品数的分布律.

解 本题可看作是20次重复独立试验. 设X表示20件齿轮产品中的废品个数, 则$X \sim b(20, 0.2)$,

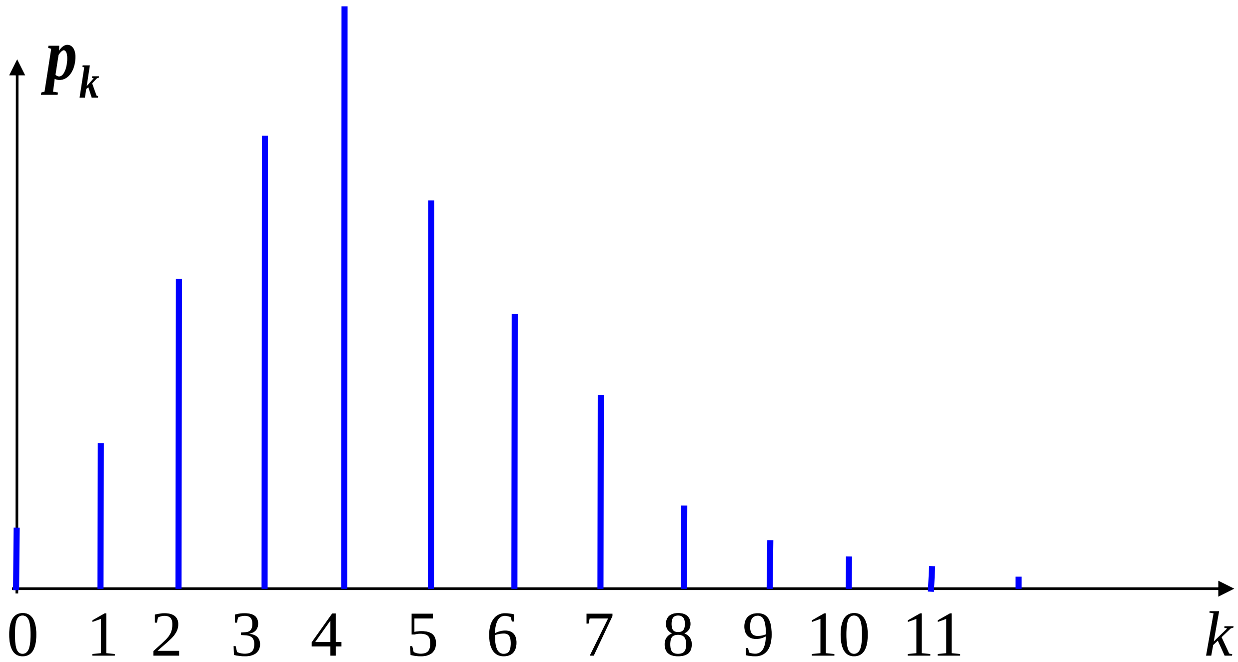
故

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

X	0	1	2	3	4	5	6
p	0.012	0.058	0.137	0.205	0.218	0.175	0.109

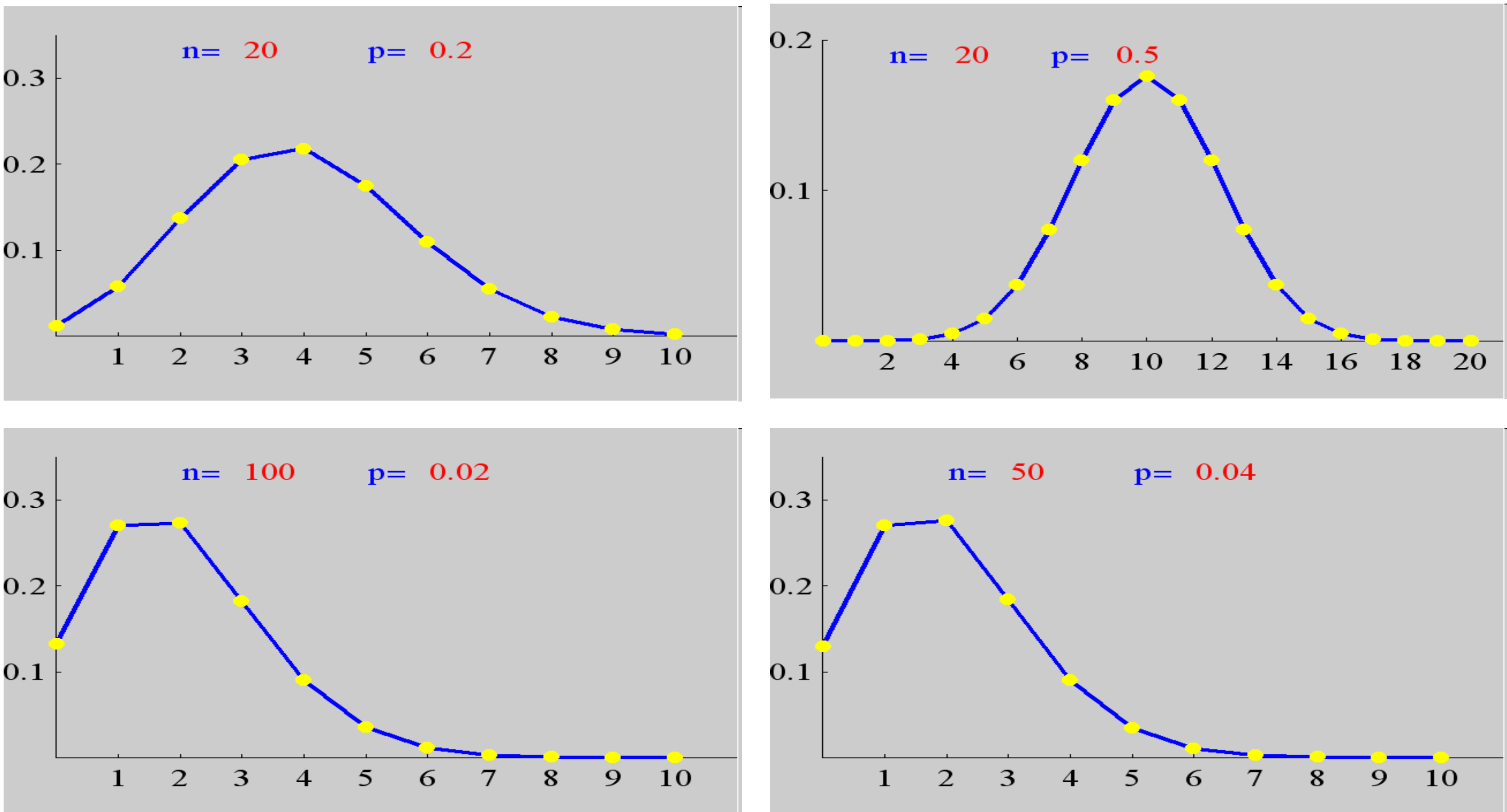
X	7	8	9	10	11	…	20
p	0.055	0.022	0.007	0.002	0.000	…	0.000

表中当$k \geq 11$时，$P\{X=k\}<0.001$. 为了对此结果有一个直观的了解，我们将表中数据用图形来表.



从上图中可看到, 概率$P\{X=k\}$先是随 k 的增加而单调上升, 当 k 增加到 4 时,$P\{X=k\}$取得最大值0.218, 然后$P\{X=k\}$再随着 k 的增加而单调下降. 一般来讲, 对于固定的n和p，二项分布 $b(n,p)$都具有这一性质.

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= 100 p= 0.02
0.3
0.2
0.1
1 2 3 4 5 6 7 8 9 10
n= 50 p= 0.04
0.3
0.2
0.1
1 2 3 4 5 6 7 8 9 10

练习

设X服从二项分布，其分布律为

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k}, k=0,1,2,\cdots,n$$

问当 k 取何值时P{X=k}为最大.

解: 设$k=N$ 时$P\{X=k\}$为最大,则有不等式

$$\begin{cases} \dfrac{P\{X=N\}}{P\{X=N-1\}} \geq 1 \\ \dfrac{P\{X=N\}}{P\{X=N+1\}} \geq 1 \end{cases}$$

解得 $(n+1)p-1 \leq N \leq (n+1)p$

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

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

例3 某人进行射击，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: 设X表示击中的次数，则$X \sim b(400, 0.02)$，

$$P\{X=k\}=C_{400}^{k}(0.02)^{k}(0.98)^{400-k}, k=0,1,2,\cdots,400$$

所求概率为

$$P\{X \geq 2\}=1-P\{X=0\}-P\{X=1\}$$
$$=1-(0.98)^{400}-400\times(0.02)\times(0.98)^{399}$$
$$\approx 0.9972$$

$P\{X<2\} = 0.0029$

这一结果的实际意义：其一，一事件尽管在一次试验中发生的概率很小，但只要试验独立进行次数很多，那么这一事件几乎是肯定的；其二，如果在400次射击中，击中目标的次数竟不到两次，据实际推断原理，我们将有理由怀疑“每次射击的命中率为0.02”。

泊松定理*：设λ>0是一常数, n是任意正整数,设np_n=λ, 则对于任一固定的非负整数 k，有

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

注 1° 显然此定理的条件 $np_n = \lambda$（常数）意味着当n很大时 p_n必定很小.

2° 当 $n \geq 20, p \leq 0.05$ 时可用近似式

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

利用此定理解例3

$$p \approx 1 - e^{-8} - 8e^{-8} \approx 0.997$$

证 由$p_n = \frac{\lambda}{n}$，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

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

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

故有

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

3. 泊松分布

若随机变量 X 的分布律为

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

其中λ>0是常数, 则称X服从参数为 λ 的泊松分布,

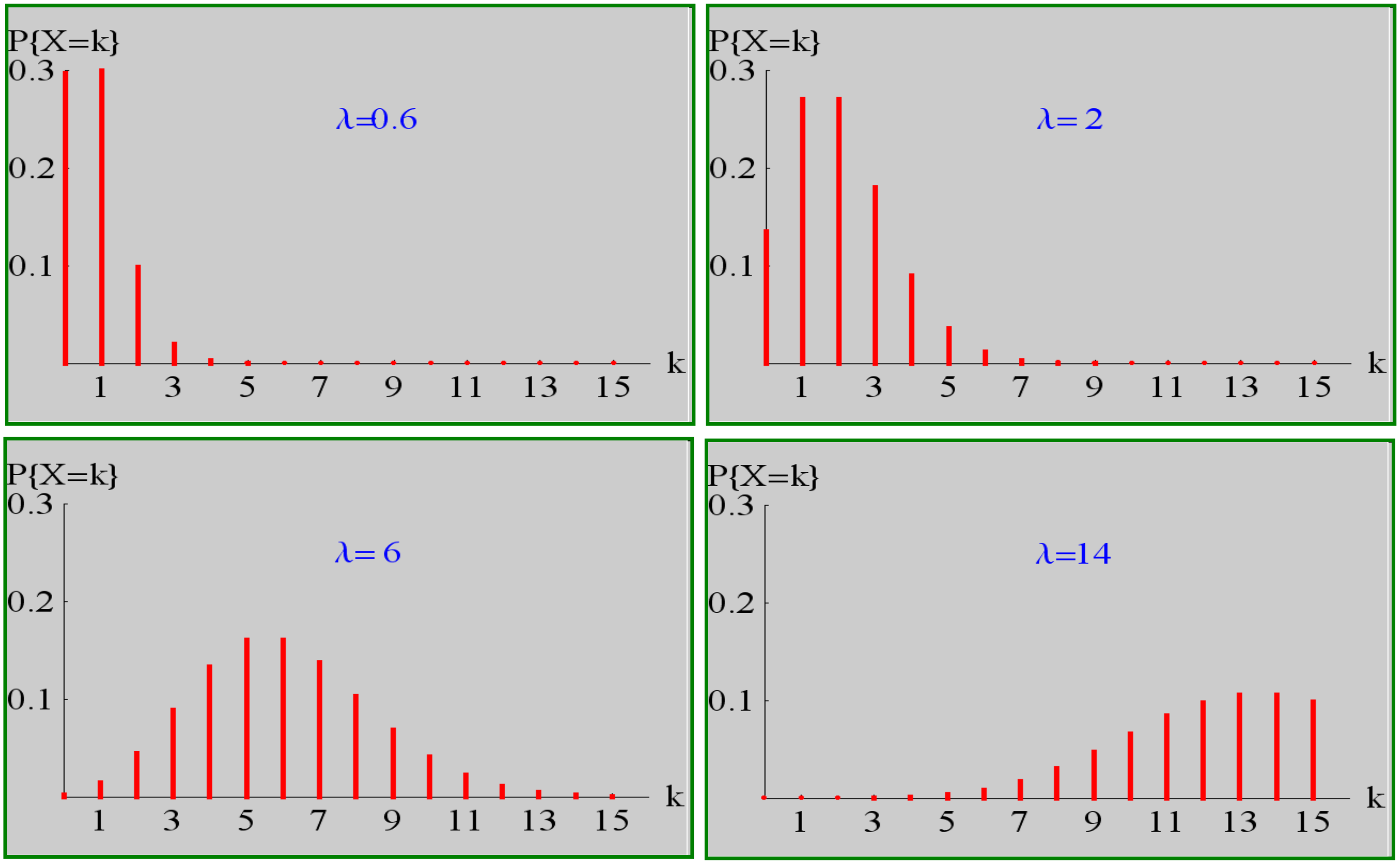
记作 $X \sim \pi(\lambda)$

可以验证:

1）$P\{X=k\}\geq 0,\ k=0,1,2,\cdots$

2）$\sum_{k=0}^{\infty}P\{X=k\}=\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1$

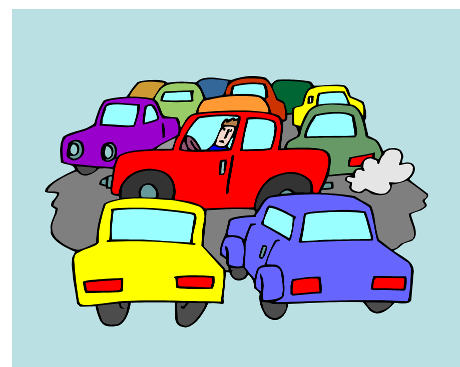
泊松分布的图形



泊松分布的背景及应用

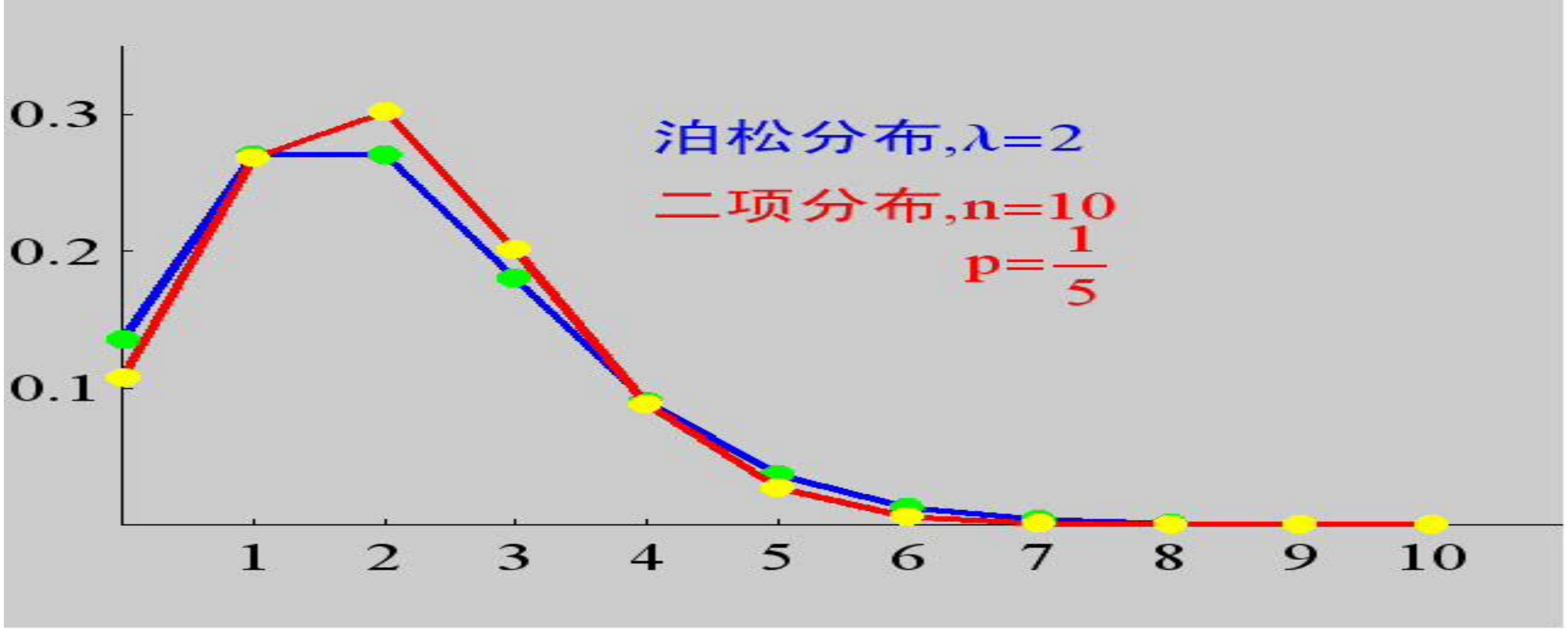
二十世纪初卢瑟福和盖克两位科学家在观察与分析放射性物质放出的粒子个数的情况时, 他们做了2608次观察(每次时间为7.5秒), 发现放射性物质在规定的一段时间内, 其放射的粒子数X服从泊松分布.

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



例4 已知某电话交换台每分钟接到的呼叫次数X服从参数 λ =4的泊松分布，求：

(1) 每分钟恰好接到3次呼唤的概率;

(2) 每分钟内接到呼唤的次数不超过4次的概率.

解 （1） $P\{X=3\}=\dfrac{4^3}{3!}e^{-4}\approx 0.1954$

（2） $P\{X\le 4\}=\sum\limits_{k=0}^{4}P\{X=k\}=\sum\limits_{k=0}^{4}\dfrac{4^k}{k!}e^{-4}$

$$=e^{-4}\left(1+4+\frac{4^2}{2!}+\frac{4^3}{3!}\right)\approx 0.371163$$

例5 设某商店每月销售某种商品的数量服从参数为5的泊松分布，问在月初要库存多少此种商品才能保证当月不脱销的概率为0.95以上.

解: 设X表示销售数量，n 为库存数量，

则 $P\{X \le n\} \ge 0.95$

由于 $X \sim \pi(5)$ $\therefore P\{X \le n\} = \sum_{k=0}^{n} \frac{5^k}{k!} e^{-5} > 0.95$

即 $\sum_{k=n+1}^{\infty} \frac{5^k}{k!} e^{-5} \le 0.05$ 查表得 $n=9$.

所以在月初要库存9 件此种商品才能保证当月不脱销的概率为0.95.

(4) 超几何分布

设有产品N件，其中次品D件，其余为正品，从中随机地抽取n件。记X为抽到的次品件数，求X的分布律。

抽到k件次品的概率为

$$P\{X=k\}=\frac{C_D^k C_{N-D}^{n-k}}{C_N^n}$$

$$\max(0, n-N+D) \le k \le \min(n, D)$$

称X服从超几何分布。

超几何分布在$N \to \infty$时的极限分布就是二项分布。

(5) 几何分布

设在投篮比赛中单次投中的概率为p，无限次地投篮，则首次投中时所需投篮的次数X服从参数为p的几何分布.即

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

容易验证，若在前 m 次投篮中未投中，那么,在此条件下，为了等到投中时刻所需要等待的时间也服从同一几何分布，该分布与 m 无关，这就是所谓的无记忆性.

$$P\{X=m+k \mid X>m\}=P\{X=k\}$$

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

$\downarrow$ $n \geq 20, p \leq 0.05$

泊松分布

2. 二项分布与$(0-1)$分布、泊松分布之间的关系.

二项分布是$(0-1)$分布的推广, 对于n次独立重复伯努利试验, 每次试验成功的概率为p, 设

$$X_i=\begin{cases}1, & \text{若第}\, i\, \text{次试验成功,}\\ 0, & \text{若第}\, i\, \text{次试验失败.}\end{cases}\qquad (i=1,2,\cdots,n)$$

它们都服从$(0-1)$分布并且相互独立, 那么

$$X=X_1+X_2+\cdots+X_n$$

服从二项分布, 参数为(n,p).

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

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

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

作业

- **第二章习题7, 10, 12, 15**