§ 2.2 离散型随机变量及其概率分布

离散随机变量及分布律

定义 若随机变量 X 的可能取值是有限个或可列个，则称 X 为离散型随机变量

描述 X 的概率特性常用概率分布或分布律

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

或

X	x_1	x_2	$\cdots$	x_k	$\cdots$
P	p_1	p_2	$\cdots$	p_k	$\cdots$

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

分布律的性质

- $p_k \geq 0,\ k = 1,2,\cdots$ ———— 非负性
- $\sum_{k=1}^{\infty} p_k = 1$ ———— 归一性

离散随机变量及分布函数

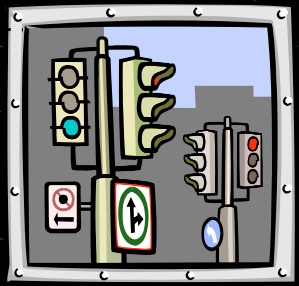
$$F(x)=P(X\le x)=P\Big(\bigcup_{x_k\le x}(X=x_k)\Big)$$

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

$$p_k=P(X=x_k)=F(x_k)-F(x_{k-1})$$

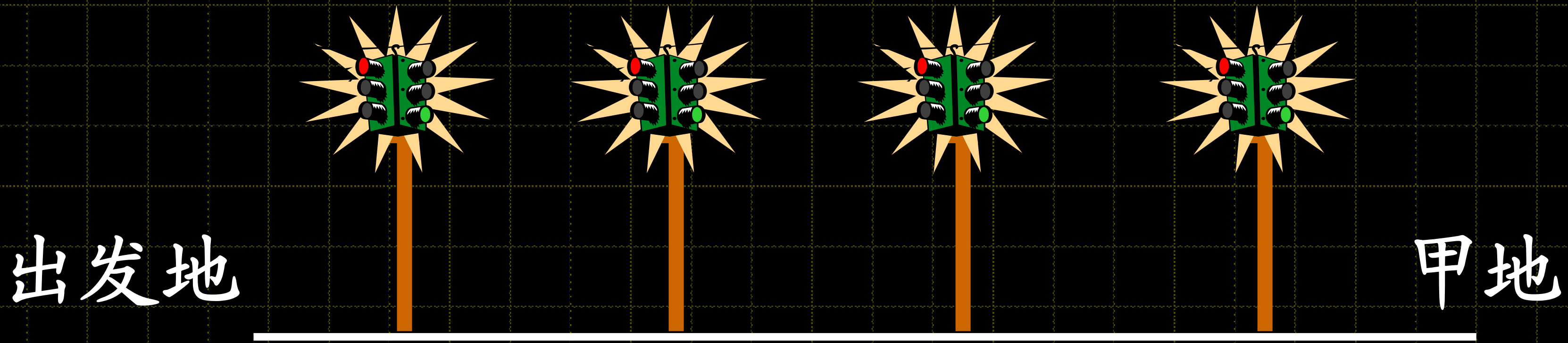
其中 $x_{k-1}<x_k$.

$F(x)$ 是分段阶梯函数, 在 X 的可能取值 x_k 处发生间断, 间断点为第一类跳跃间断点,在间断点处有跃度 p_k.



例1 设汽车在开往甲地途中需经过 4 盏信号灯, 每盏信号灯独立地以概率 p 允许汽车通过.令 X 表示首次停下时已通过的信号灯盏数, 求 X 的概率分布与 $p=0.4$ 时的分布函数.

解

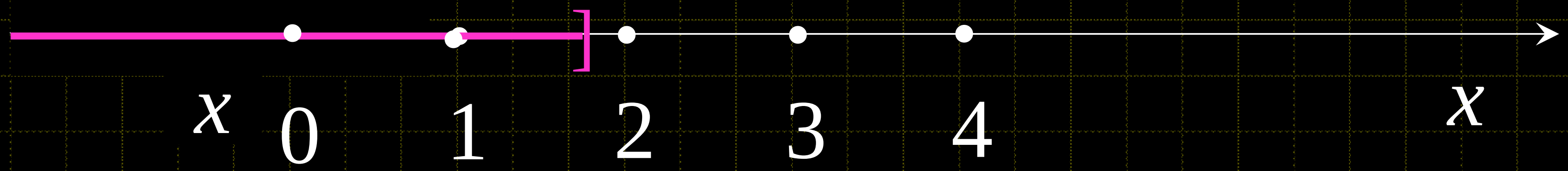


$$P(X=k)=p^k(1-p),\ k=0,1,2,3$$

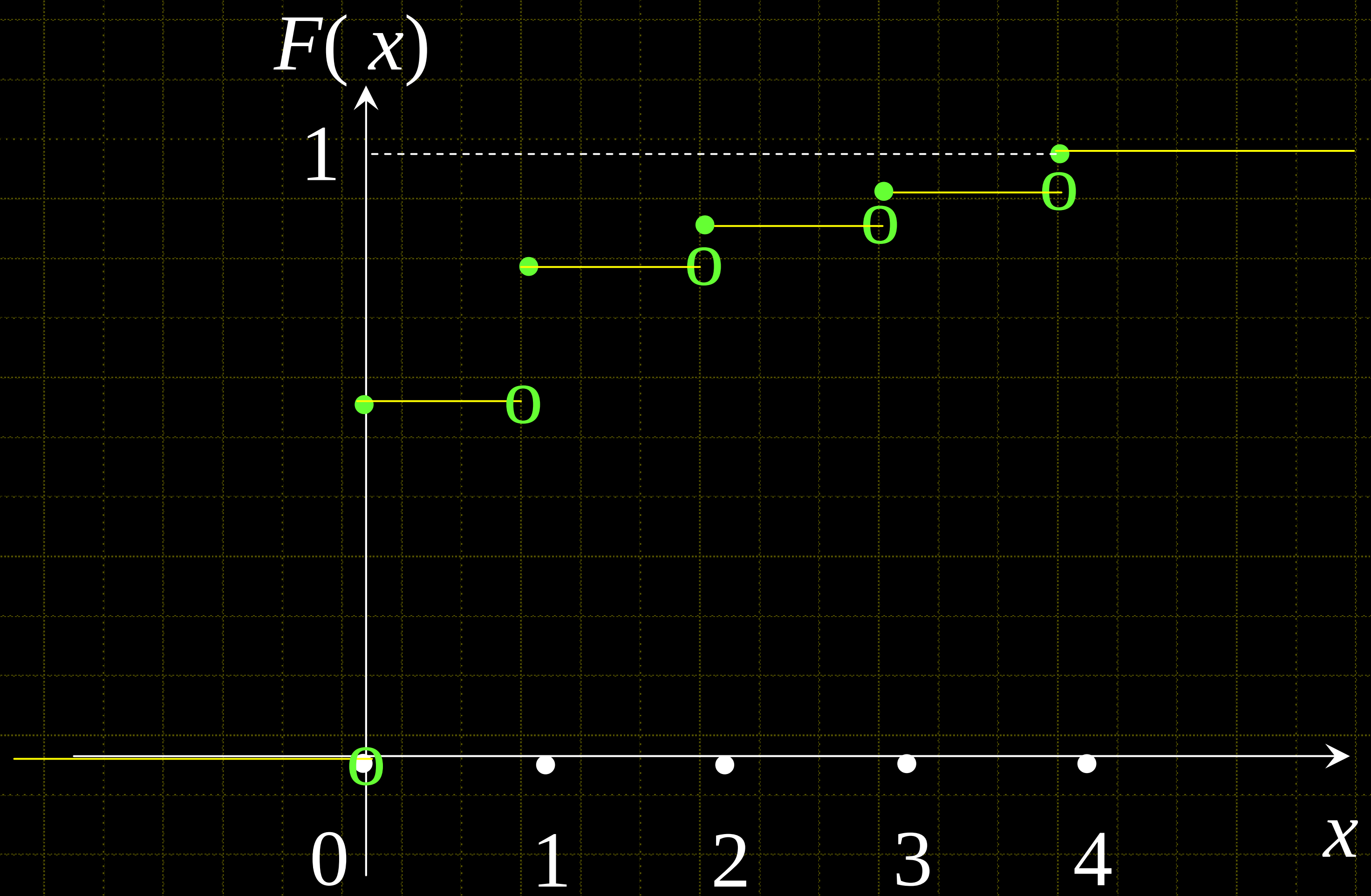
$$P(X=4)=p^4,$$

$p = 0.4$ 代入

k	0	1	2	3	4
p_k	0.6	0.24	0.096	0.0384	0.0256



$$F(x) = P(X \le x) = \begin{cases} 0, & x < 0 \\ 0.6, & 0 \le x < 1 \\ 0.6 + 0.24 = 0.84, & 1 \le x < 2 \\ 0.84 + 0.096 = 0.936, & 2 \le x < 3 \\ 0.936 + 0.0384 = 0.9744, & 3 \le x < 4 \\ 1 & x \ge 4 \end{cases}$$


F(x)
1
0
1
2
3
4
x

用分布律或分布函数来计算事件的概率

例2 在上例中，分别用分布律与分布函数计算 $P(1\le X\le 3)$.

解 $P(1\le X\le 3)=P(X=1)+P(X=2)+P(X=3)$

$$=0.6(0.4+0.4^2+0.4^3)=0.3744$$

或

$$P(1\le X\le 3)=F(3)-F(1-0)=0.9744-0.6$$

此式应理解为极限 $\lim\limits_{x\to 1^-}F(x)$

例3 一门大炮对目标进行轰击,假定此目标必须被击中 r 次才能被摧毁. 若每次击中目标的概率为 p $(0 < p < 1)$, 且各次轰击相互独立,一次次地轰击直到摧毁目标为止.求所需轰击次数 X 的概率分布.

解 $P(X = k) = P($前 $k-1$次击中 $r-1$次, 第 k 次击中目标$)$

$$= C_{k-1}^{r-1} p^{r-1} (1-p)^{k-r} \cdot p$$

$$= C_{k-1}^{r-1} p^{r} (1-p)^{k-r} \qquad k = r, r+1, \cdots$$

帕斯卡分布

注 $$\sum_{k=r}^{\infty} C_{k-1}^{r-1} p^r (1-p)^{k-r} = 1$$

利用幂级数在收敛域内可逐项求导的性质

当 $|x|<1$ $$\sum_{k=1}^{\infty} x^{k-1} = \frac{1}{1-x}$$

$$\sum_{k=2}^{\infty} (k-1) x^{k-2} = \frac{1}{(1-x)^2}$$

$$\sum_{k=3}^{\infty} (k-1)(k-2) x^{k-3} = \frac{2}{(1-x)^3}$$

$$\Rightarrow \sum_{k=3}^{\infty} C_{k-1}^{2} x^{k-3} = \frac{1}{(1-x)^3}$$

归纳地

$$\sum_{k=r}^{\infty} C_{k-1}^{r-1} x^{k-r} = \frac{1}{(1-x)^r}$$

令 $x = 1 - p$

$$\Longrightarrow \sum_{k=r}^{\infty} C_{k-1}^{r-1} (1-p)^{k-r} = \frac{1}{(1-(1-p))^r} = \frac{1}{p^r}$$

$$\Longrightarrow \sum_{k=r}^{\infty} C_{k-1}^{r-1} p^r (1-p)^{k-r} = 1$$

作业 P82 习题二

2 4

5 6

常见离散r.v.的分布

(1) $0-1$分布

$X=x_k$	1	0
P_k	p	$1-p$

$0<p<1$

或 $P(X=k)=p^k(1-p)^{1-k}$, $k=0,1$

应用场合

凡试验只有两个结果, 常用$0-1$分布描述, 如产品是否合格、人口性别统计、系统是否正常、电力消耗是否超标等等.

(2) 二项分布

n 重Bernoulli 试验中, X 是事件A 在 n 次试验中发生的次数 , $P(A) = p$,若

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

则称 X 服从参数为n, p 的二项分布，记作

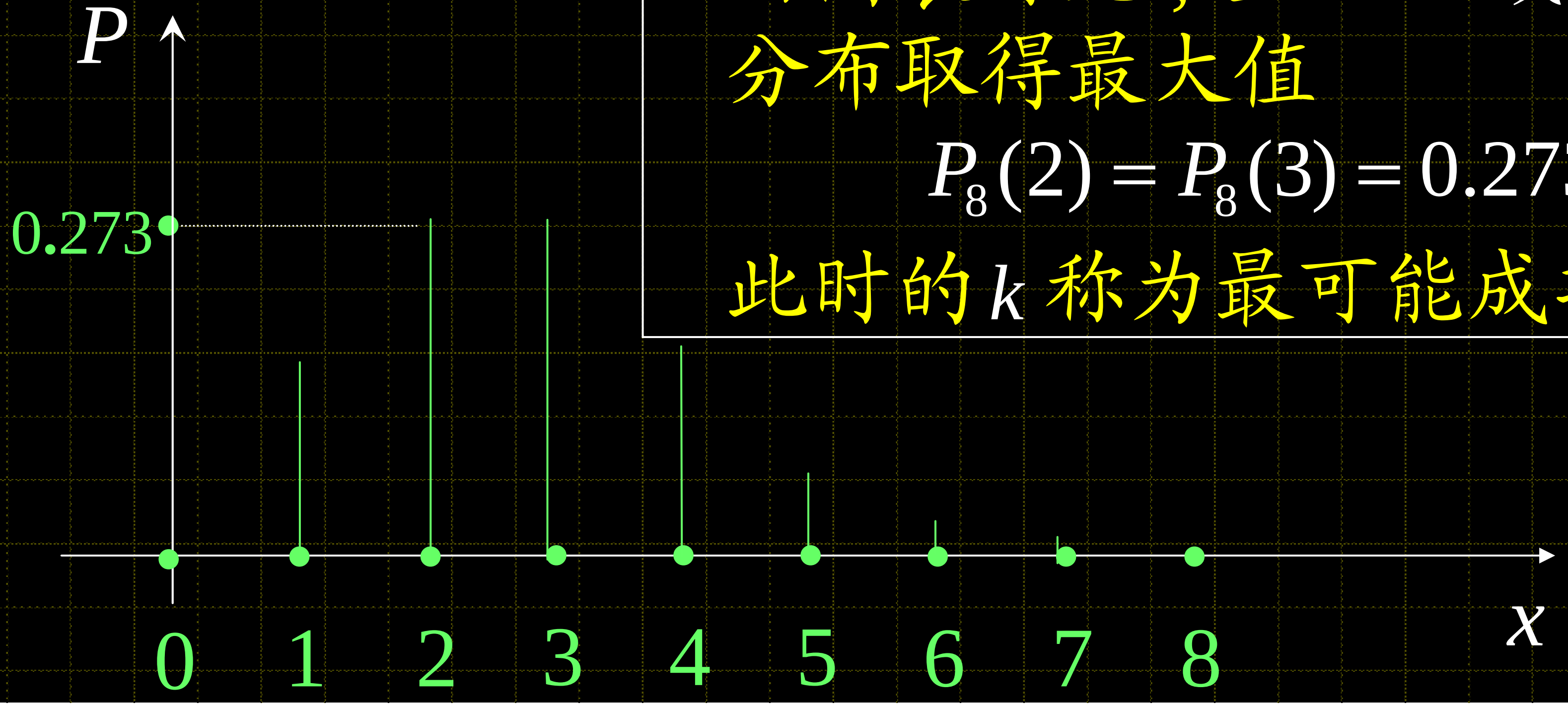
$$X \sim B(n, p)$$

0–1 分布是 $n = 1$ 的二项分布

二项分布的取值情况 设 $X \sim B(8, \frac{1}{3})$

$$P_8(k) = P(X = k) = C_8^k \left(\tfrac{1}{3}\right)^k \left(1 - \tfrac{1}{3}\right)^{8-k}, \quad k = 0,1,\cdots,8$$

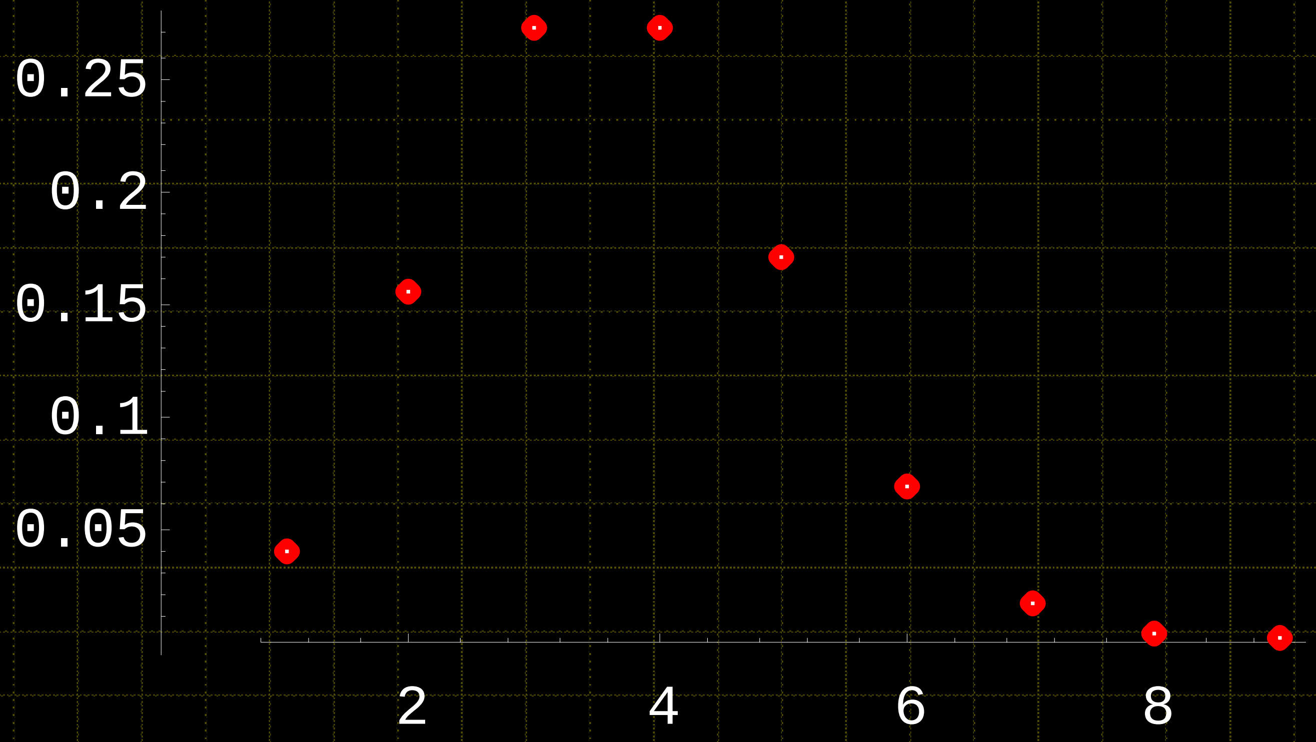
0	1	2	3	4	5	6	7	8
.039	.156	.273	.273	.179	.068	.017	.0024	.0000



由图表可见，当 $k = 2$ 或 3 时，分布取得最大值

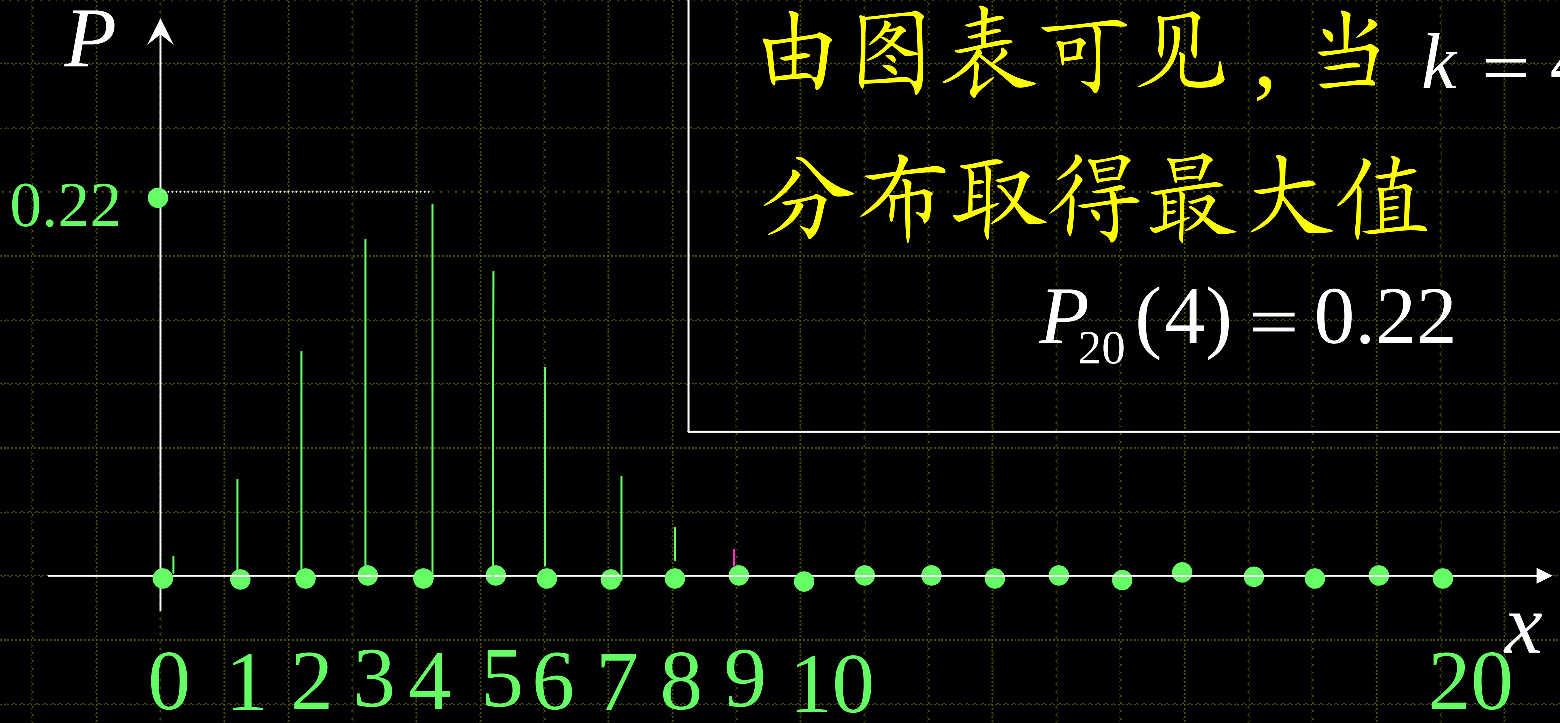
$$P_8(2) = P_8(3) = 0.273$$

此时的 k 称为最可能成功次数


0.25
0.2
0.15
0.1
0.05
2
4
6
8

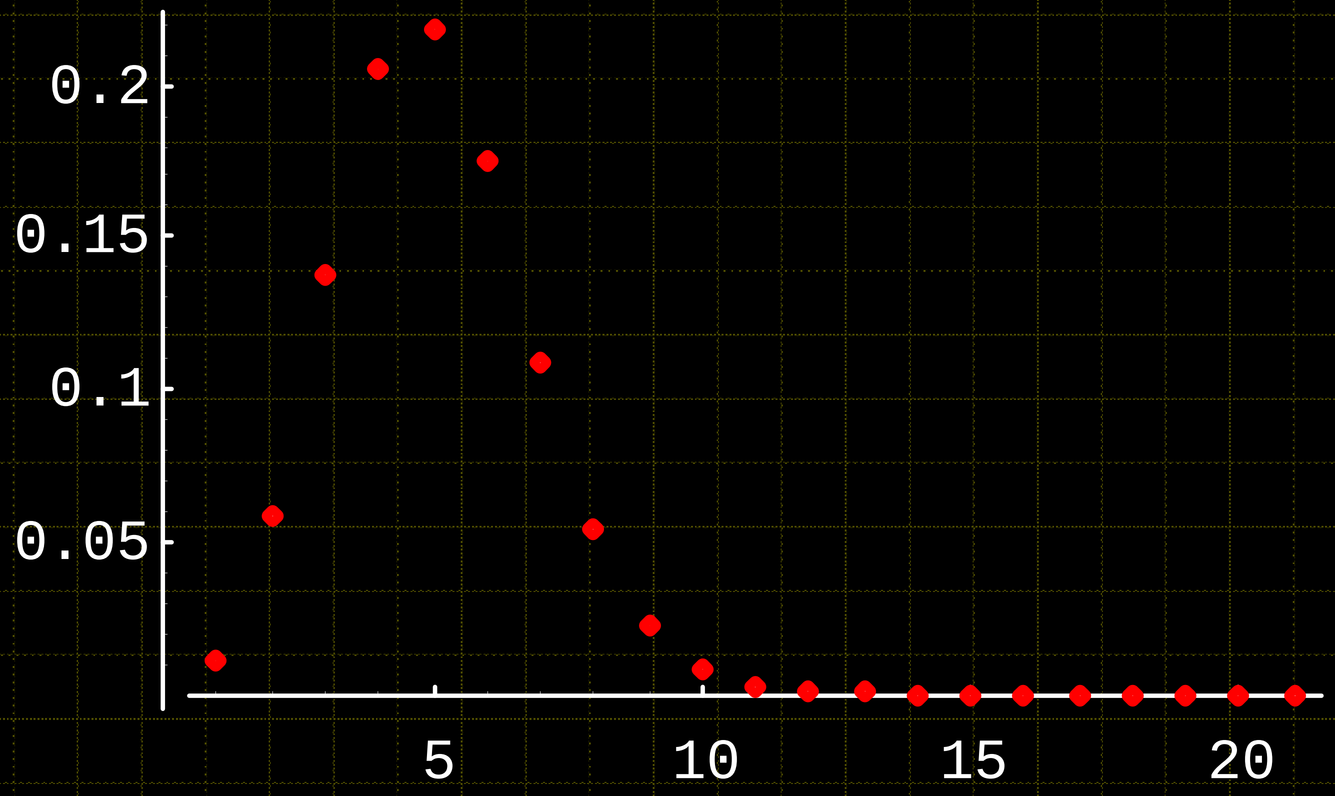
设 $X \sim B(20,0.2)$

0	1	2	3	4	5	6	7	8	9	10	11 ~ 20
.01	.06	.14	.21	.22	.18	.11	.06	.02	.01	.002	< .001



由图表可见，当 $k=4$ 时，分布取得最大值

$$P_{20}(4) = 0.22$$


0.2
0.15
0.1
0.05
5
10
15
20

二项分布中最可能出现次数的定义与推导

若 $P(X=k)\geq P(X=j)$, $j=X$ 可取的一切值

则称k 为最可能出现的次数

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

$$\left.\begin{aligned}\frac{p_{k-1}}{p_k}&=\frac{(1-p)k}{p(n-k-1)}\leq 1\\[2ex]\frac{p_k}{p_{k+1}}&=\frac{(1-p)(k+1)}{p(n-k)}\geq 1\end{aligned}\right\}$$

$\Rightarrow$ $(n+1)p-1\leq k\leq(n+1)p$

- 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

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

- 对固定的 n、p, $P(X=k)$ 的取值呈不对称分布

  固定 p, 随着 n 的增大，其取值的分布趋于对称

例4 独立射击5000次, 命中率为0.001, 求 (1) 最可能命中次数及相应的概率;

(2) 命中次数不少于1 次的概率.

解 (1) $k = [(n+1)p]$

$$= [(5000+1)0.001] = 5$$

$$P_{5000}(5) = C_{5000}^{5}(0.001)^5(0.999)^{4995}$$

$$\approx 0.1756$$

(2) 令X表示命中次数,则 $X \sim B(5000,0.001)$

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

$$= 1 - C_{5000}^{0}(0.001)^{0}(0.999)^{5000}$$

$$= 0.9934.$$

本例启示 小概率事件虽不易发生，但重复次数多了，就成大概率事件.

由此可见日常生活中“提高警惕,防火防盗”的重要性.

由于时间无限,自然界发生地震、海啸、空难、泥石流等都是必然的,早晚的事,不用奇怪,不用惊慌.

同样,人生中发生车祸、失恋、患绝症、考试不及格、炒股大亏损等都是正常现象,大可不必怨天尤人,更不要想不开而跳物理楼(交大闵行校区最高楼)自杀.

问题 如何计算 $P(X \geq 2500)$？

Possion定理

设 $np_n = \lambda > 0$，则对固定的 k

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

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

Poisson定理说明若$X \sim B(n, p)$，则当n较大，p较小，而 $np = \lambda$适中，则可以用近似公式

$$C_n^k p^k (1-p)^{n-k} \approx e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0,1,2,\cdots$$

证 记 $np_n = \lambda_n$

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

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

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

$$\to e^{-\lambda}\frac{\lambda^k}{k!} \qquad k = 1,2,\cdots$$

类似地, 从装有 a 个白球, b 个红球的袋中不放回地任取 n 个球, 其中恰有k 个白球的概率为 $C_a^k C_b^{n-k} / C_{a+b}^n$

当 $a+b \to \infty, \dfrac{a}{a+b} \to p$ 时,

对每个 n 有 $\dfrac{C_a^k C_b^{n-k}}{C_{a+b}^n} \to C_{a+b}^k p^k (1-p)^{n-k}$

结 论

超几何分布的极限分布是二项分布

二项分布的极限分布是 Poisson 分布

利用Poisson定理再求例4 (2)

解 令X表示命中次数, 则

$$X \sim B(5000, 0.001)$$

令 $\lambda = np = 5$

$$P(X \geq 1) \approx 1 - e^{-5} = 0.9933.$$

此结果也可直接查 P.378 附表2 泊松分布表得到，它与用二项分布算得的结果0.9934仅相差万分之一.

例5 某厂产品不合格率为0.03, 现将产品装箱, 若要以不小于 90%的概率保证每箱中至少有 100 个合格品, 则每箱至少应装多少个产品?

解 设每箱至少应装100 + n 个, 每箱的不合格品个数为X, 则$X \sim B(100+n, 0.03)$

由题意 $$P(X \le n) = \sum_{k=0}^{n} P_{100+n}(k) \ge 0.9$$

$(100+n)0.03=3+0.03n \approx 3$ 取 $\lambda = 3$

应用Poisson定理

$$\sum_{k=0}^{n} P_{100+n}(k) \approx \sum_{k=0}^{n} \frac{3^k}{k!} e^{-3} = 1 - \sum_{k=n+1}^{\infty} \frac{3^k}{k!} e^{-3} \geq 0.9$$

$$\Longrightarrow \sum_{k=n+1}^{\infty} \frac{3^k}{k!} e^{-3} \leq 0.1$$ 查Poisson分布表, $\lambda = 3$

得 $n + 1 = 6$, $n = 5$

故每箱至少应装105个产品, 才能符合要求.

在实际计算中,当 $n \geq 20, p \leq 0.05$时, 可用上述公式近似计算; 而当 $n \geq 100,\ np \leq 10$ 时, 精度更好

k	按二项分布 $n=10$ $p=0.1$	$n=20$ $p=0.05$	$n=40$ $p=0.025$	$n=100$ $p=0.01$	按Possion公式 $\lambda=np=1$
0	0.349	0.358	0.369	0.366	0.368
1	0.305	0.377	0.372	0.370	0.368
2	0.194	0.189	0.186	0.185	0.184
3	0.057	0.060	0.060	0.061	0.061
4	0.011	0.013	0.014	0.015	0.015

在Poisson 定理中，

$$e^{-\lambda}\frac{\lambda^k}{k!} > 0$$

$$\sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}$$

$$= e^{-\lambda}\left(1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\cdots\right)$$

$$= e^{-\lambda}\cdot e^{\lambda} = 1$$

由此产生了一种离散型随机变量的概率分布
— Poisson 分布

(3) Poisson 分布

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

其中 $\lambda>0$ 是常数，则称 X 服从参数为 λ 的Poisson 分布. 记作 $X\sim\pi(\lambda)$ 或 $P(\lambda)$

应用场合

在某个时段内：

① 大卖场的顾客数;
② 市级医院急诊病人数;
③ 某地区拨错号的电话呼唤次数;
④ 某地区发生的交通事故的次数.
⑤ 放射性物质发出的 α 粒子数;
⑥ 一匹布上的疵点个数;
⑦ 一个容器中的细菌数;
⑧ 一本书一页中的印刷错误数;

…………

都可以看作是源源不断出现的随机质点流，若它们满足一定的条件，则称为 Poisson 流，在长为 t 的时间内出现的质点数 $X_t \sim P(\lambda t)$

例6 设一只昆虫所生虫卵数为随机变量 X，已知 $X \sim P(\lambda)$，且每个虫卵发育成幼虫的概率为 p.

设各个虫卵是否能发育成幼虫是相互独立的.

求一昆虫所生的虫卵发育成幼虫数 Y 的概率分布.

解 昆虫 $\longrightarrow$ X 个虫卵 $\longrightarrow$ Y 个幼虫

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

$$P(Y=m \mid X=k)=C_k^m p^m(1-p)^{k-m},\ m=0,1,2,\cdots,k$$

$$(Y=m)\subset\bigcup_{k=m}^{\infty}(X=k),\quad m=0,1,2,\cdots$$

$$(X=k)\cap(X=l)=\varnothing,\quad k\neq l$$

由全概率公式

$$P(Y=m)=\sum_{k=m}^{\infty}P(X=k)P(Y=m|X=k)$$

$$=\sum_{k=m}^{\infty}e^{-\lambda}\frac{\lambda^k}{k!}C_k^m p^m(1-p)^{k-m}$$

$$=e^{-\lambda}\frac{(\lambda p)^m}{m!}\sum_{k=m}^{\infty}\frac{\lambda^{k-m}}{(k-m)!}(1-p)^{k-m}$$

$$\overset{令 k-m=s}{=} e^{-\lambda}\frac{(\lambda p)^m}{m!}\sum_{s=0}^{\infty}\frac{\lambda^s}{s!}(1-p)^s$$

$$=e^{-\lambda}\frac{(\lambda p)^m}{m!}e^{\lambda(1-p)}=e^{-\lambda p}\frac{(\lambda p)^m}{m!},\quad m=0,1,2,\cdots$$

故 $Y \sim P(\lambda p)$

作业 P82 习题二

8 (1)
12
14
15

第5周

问 题

自动生产线调整以后出现废品的概率为 p，当生产过程中出现废品时立即重新进行调整，求在两次调整之间的合格产品数的分布.

第五周

问题

已知运载火箭在飞行中进入其仪器舱的宇宙粒子数服从参数为 2 的泊松分布．而进入仪器舱的粒子随机落到仪器重要部位的概率为 0.1, 求落到仪器重要部位的粒子数的概率分布．

帕斯卡

Blaise Pascal

1623-1662

法国数学家

物理学家

思想家

帕斯卡简介

帕斯卡四岁丧母，在父亲精心培养下, 16岁时发现帕斯卡六边形定理,写成《圆锥曲线论》, 由此定理导出400余条推论，这是古希腊阿波罗尼奥斯以来圆锥曲线论的最大进步.

1642年发明世界上第一台机械加法计算机——帕斯卡计算器.

1647年他发现了流体静力学的帕斯卡原理.

1654年研究二项系数性质,写出《论算术三角形》一文, 还深入讨论不可分原理, 这实际上相当于已知道

$$\int_0^a x^n dx = \frac{a^{n+1}}{n+1}$$

他应用此方法解决了摆线问题.

1658年完成了《摆线论》，这给G.W.莱布尼茨以很大启发，促使了微积分的建立.

三十岁时他曾研究过赌博问题，对早期概率论的发展颇有影响.

在离散型随机变量的分布中有个以帕斯卡名字命名的分布，它应用于重复独立试验中，事件发生 r 次的场

合. 而有名的几何分布正是其 $r=1$ 时的特例.

帕斯卡还写过不少文学著作.

1654年他进入修道院,献身于哲学和宗教.

附例 设同类型设备90台，每台工作相互独立，每台设备发生故障的概率都是 0.01. 在通常情况下，一台设备发生故障可由一个人独立维修，每人同时也只能维修一台设备.

(1) 问至少要配备多少维修工人，才能保证当设备发生故障时不能及时维修的概率小于0.01?

(2) 问3个人共同负责90台还是3个人各自独立负责30台设备发生故障不能及时维修的概率低?

自学（详解见教材 P.61例6）

解 (1) 设需要配备 N 个维修工人,设 X 为90 台设备中发生故障的台数，则 $X \sim B(90, 0.01)$

$$P(X > N) = \sum_{k=N+1}^{90} C_{90}^{k}(0.01)^{k}(0.99)^{N-k}$$

令 $\lambda = 90 \times 0.01 = 0.9$

则 $P(X > N) \approx \sum_{k=N+1}^{90} e^{-0.9}\frac{0.9^{k}}{k!}$

$$= \sum_{k=N+1}^{\infty} e^{-0.9}\frac{0.9^{k}}{k!} - \sum_{k=91}^{\infty} e^{-0.9}\frac{0.9^{k}}{k!}$$

$$\approx \sum_{k=N+1}^{\infty} e^{-0.9}\frac{0.9^{k}}{k!} < 0.01$$

查附表2得 $\underline{N = 4}$

(2)三个人共同负责90台设备发生故障不能及时维修的概率为

$$P(X>3)\approx\sum_{k=4}^{90}e^{-0.9}\frac{0.9^k}{k!}$$

$$=\sum_{k=4}^{\infty}e^{-0.9}\frac{0.9^k}{k!}-\sum_{k=91}^{\infty}e^{-0.9}\frac{0.9^k}{k!}$$

$$\approx\sum_{k=4}^{\infty}e^{-0.9}\frac{0.9^k}{k!}$$

$$=0.013459$$

设30台设备中发生故障的台数为 $Y \sim B(30, 0.01)$

设每个人独立负责30台设备，第 i 个人负责的30台设备发生故障不能及时维修为事件 A_i

则 $$P(A_i) = P(Y \geq 2) \approx \sum_{k=2}^{\infty} e^{-0.3} \frac{0.3^k}{k!}$$

$$= 0.0369 \qquad i = 1,2,3$$

三个人各独立负责30台设备发生故障不能及时维修为事件 $A_1 \cup A_2 \cup A_3$

$$P(A_1 \cup A_2 \cup A_3) = 1 - \prod_{i=1}^{3} P(\overline{A_i})$$

$$= 1 - (1 - 0.0369)^3 \approx 0.1067 > 0.013459$$

故 三个人共同负责90 台设备比各自负责好!