§ 2.2 离散型随机变量及分布律

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discrete随机变量**

描述X 的概率特性常用**概率分布律 probability function/probability mass function**

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

或

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

分布律：也就是一个二维的表格，分别是随机变量和这个随机变量所对应的事件的概率

只有离散型的随机变量才会有分布律，连续型的随机变量取值太多了，没办法画出来

重要

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

用性质可以判断是否为分布律

离散型随机变量的分布函数distribution function/cumulative distribution function

重要

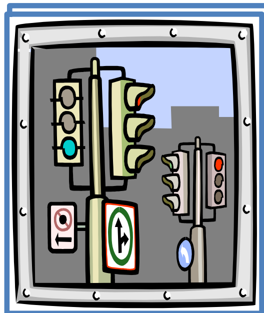
用分布函数计算 X 落在$(a,b]$里的概率：

$$P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a)$$

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

$$p_k = f(x_k) = P(X = x_k) = F(x_k) - F(x_{k-1}) \quad x_{k-1} < x_k$$

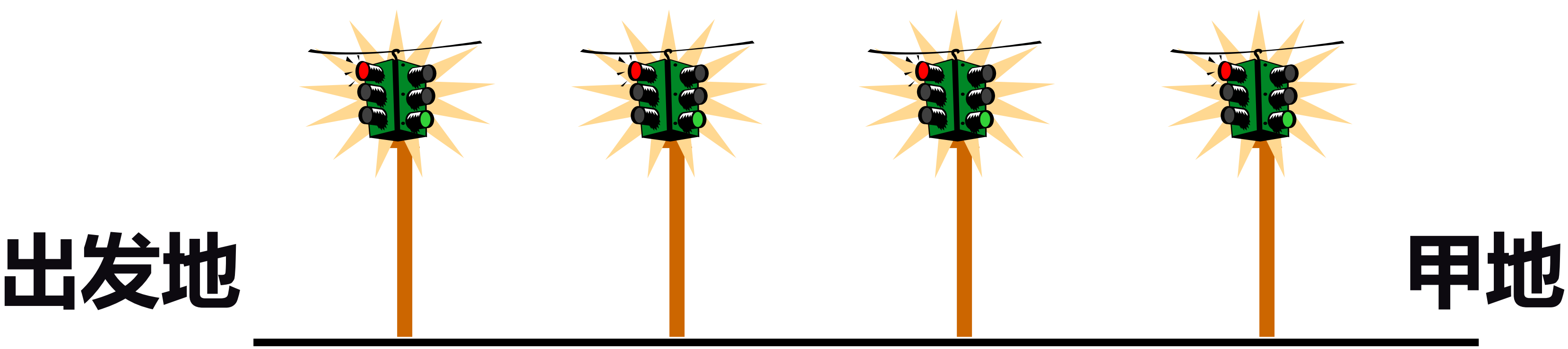
$F(x)$ 是分段阶梯函数, 在 X 的可能取值 x_k 处发生间断, 间断点为第一类跳跃间断点, 在间断点处有跃度 p_k.



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 的概率分布律与 $p=0.4$ 时的分布函数

伯努利？

解



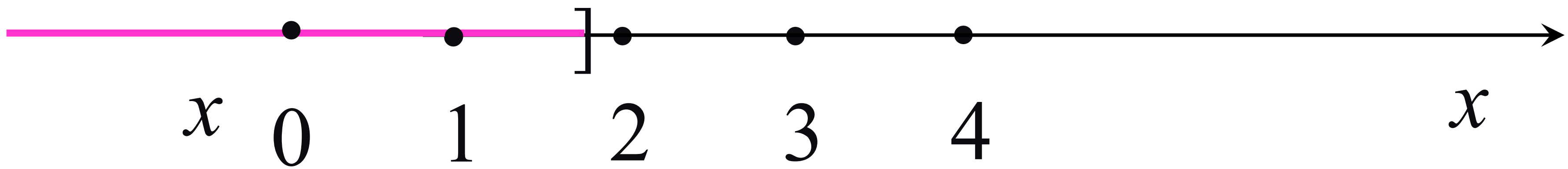
$$P(X=k)=p^k(1-p),\ k=0,1,2,3$$

k=4?

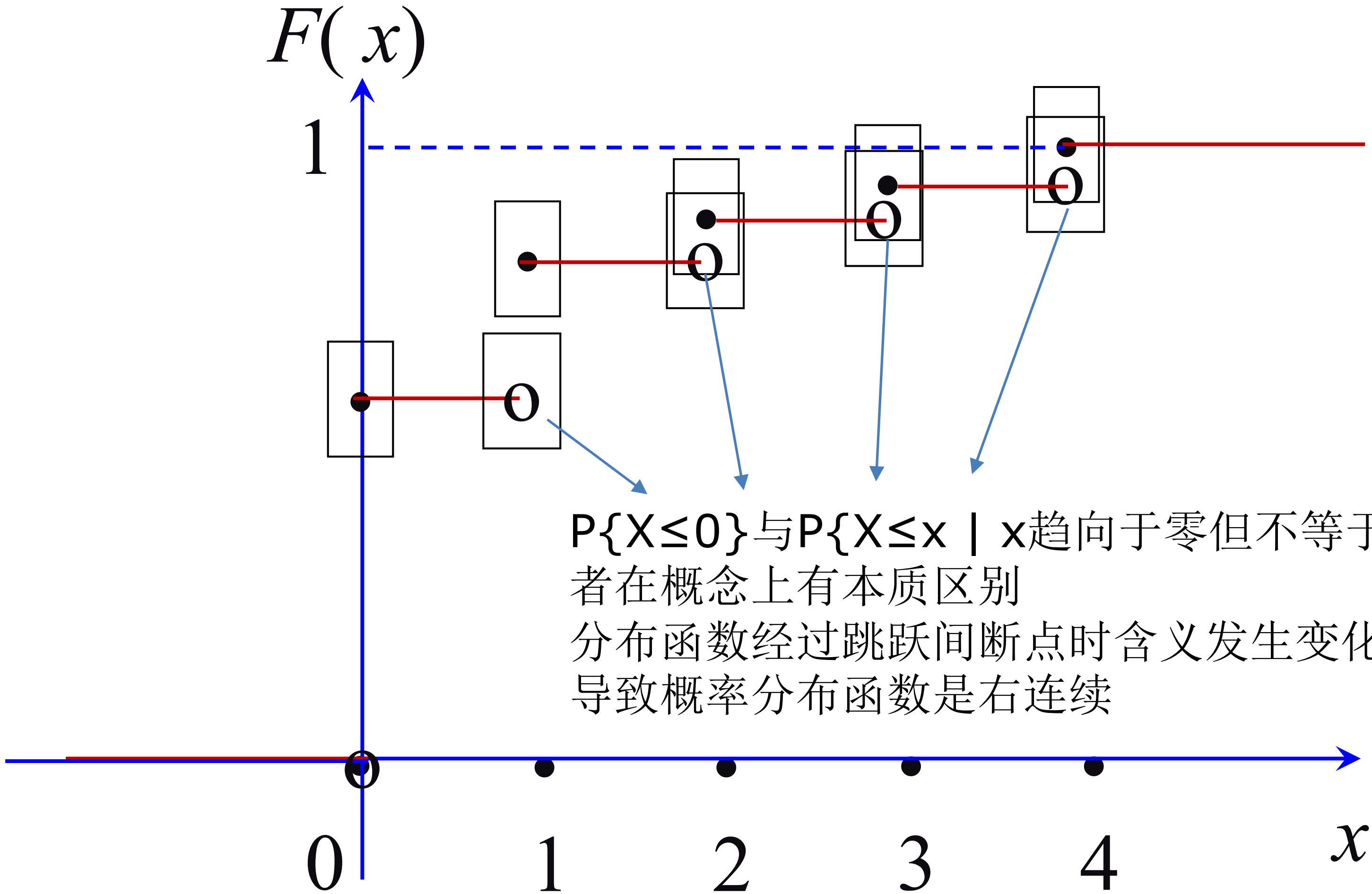
$$P(X=4)=p^4$$

$p = 0.4$ 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
P{X≤0}与P{X≤x | x趋向于零但不等于零} 二者在概念上有本质区别
分布函数经过跳跃间断点时含义发生变化的特点导致概率分布函数是右连续

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

$$P(1\le X\le 3)=P(X\le 3)-P(X<1)$$

$$=F(3)-F(1-0)=0.9744-0.6$$

此式应理解为极限 $\lim\limits_{x\to 1^-}F(x)$

例3. 设随机变量X的概率函数为:

$$f(X=k)=P(X=k)=a\frac{\lambda^k}{k!},k=0,1,2,\ldots,\lambda>0$$

试确定常数a.

解: 依据概率函数的性质:

$$\begin{cases} P(X=k)\geq 0, \\ \sum_k P(X=k)=1 \end{cases}$$

欲使上述函数为概率函数

应有 $a\geq 0$

$$\boxed{\sum_{k=0}^{\infty} a\frac{\lambda^k}{k!}=ae^{\lambda}=1}$$

从中解得 $a=e^{-\lambda}$

这里用到了常见的幂级数展开式

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

Summary

重要

1. $F(x)$、p_k定义和相互关系

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

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

2. 性质

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

四. 常见离散型随机变量的分布

1.超几何分布

例 设有 N 件产品，其中有 M 件次品，现从中任取 n 件，用 X 表示其中的次品数，求其分布律。

超几何公式

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

$$l=\min(M,n)$$

超几何分布

2.几何分布Geometric Distribution

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,… ,

A_k = {第k发命中}，k =1, 2, …,

$$P(X=1)=P(A_1)=p,$$

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

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

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

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

若随机变量*X*的概率分布如上式，则称*X*具有几何分布.

不难验证:

$$\sum_{k=1}^{\infty}(1-p)^{k-1}\cdot p=1$$

3. 两点分布(0 – 1 分布) Bernoulli Distribution

X	0	1
P_k	$1-p$	p

$0 < p < 1$

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

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.

4. 二项分布Binomial Distribution

重要

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

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

$$\text{or } f(x)=C_n^x p^x(1-p)^{n-x},\quad x=0,1,\cdots,n$$

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 $X \sim B(8, 1/3)$

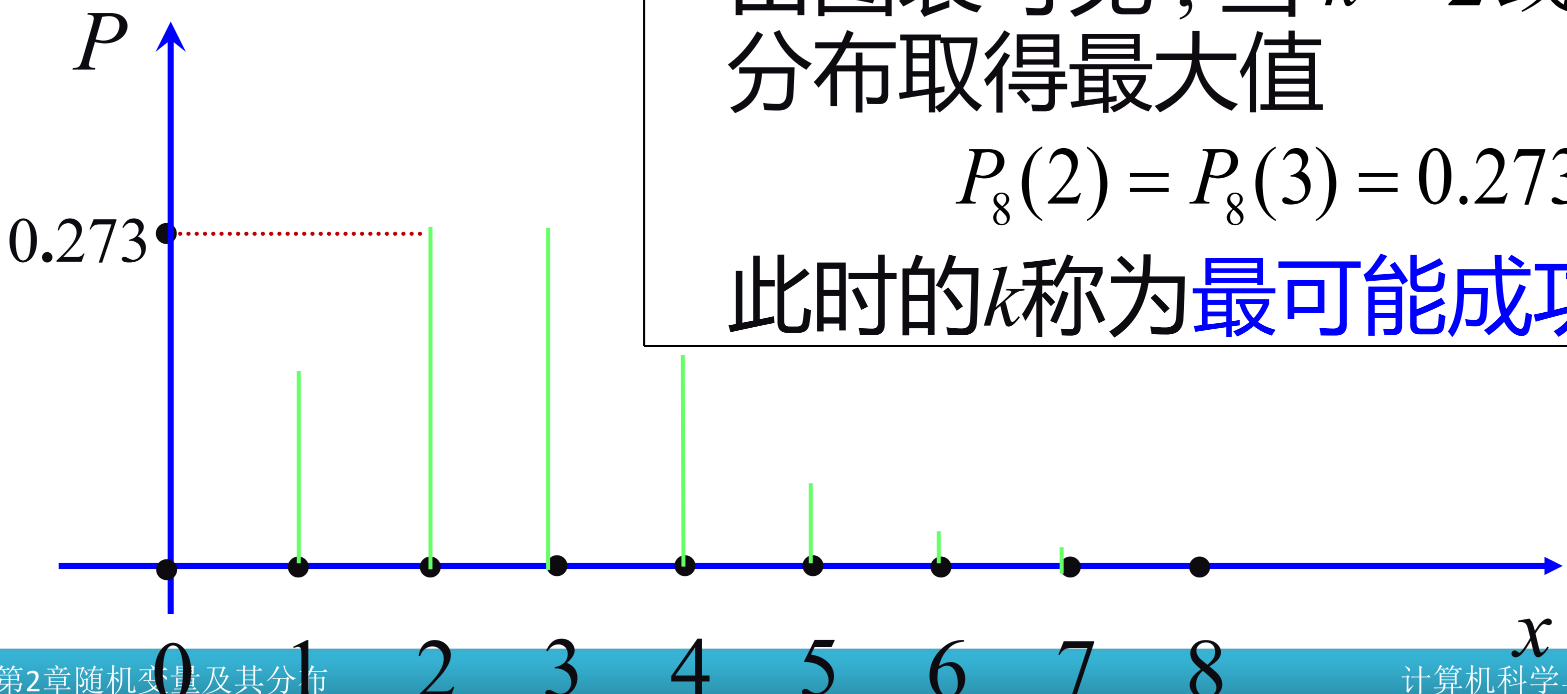
$$P_8(k) = P(X = k) = C_8^k (1/3)^k (1 - 1/3)^{8-k}, \quad k = 0, 1, \cdots, 8$$

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

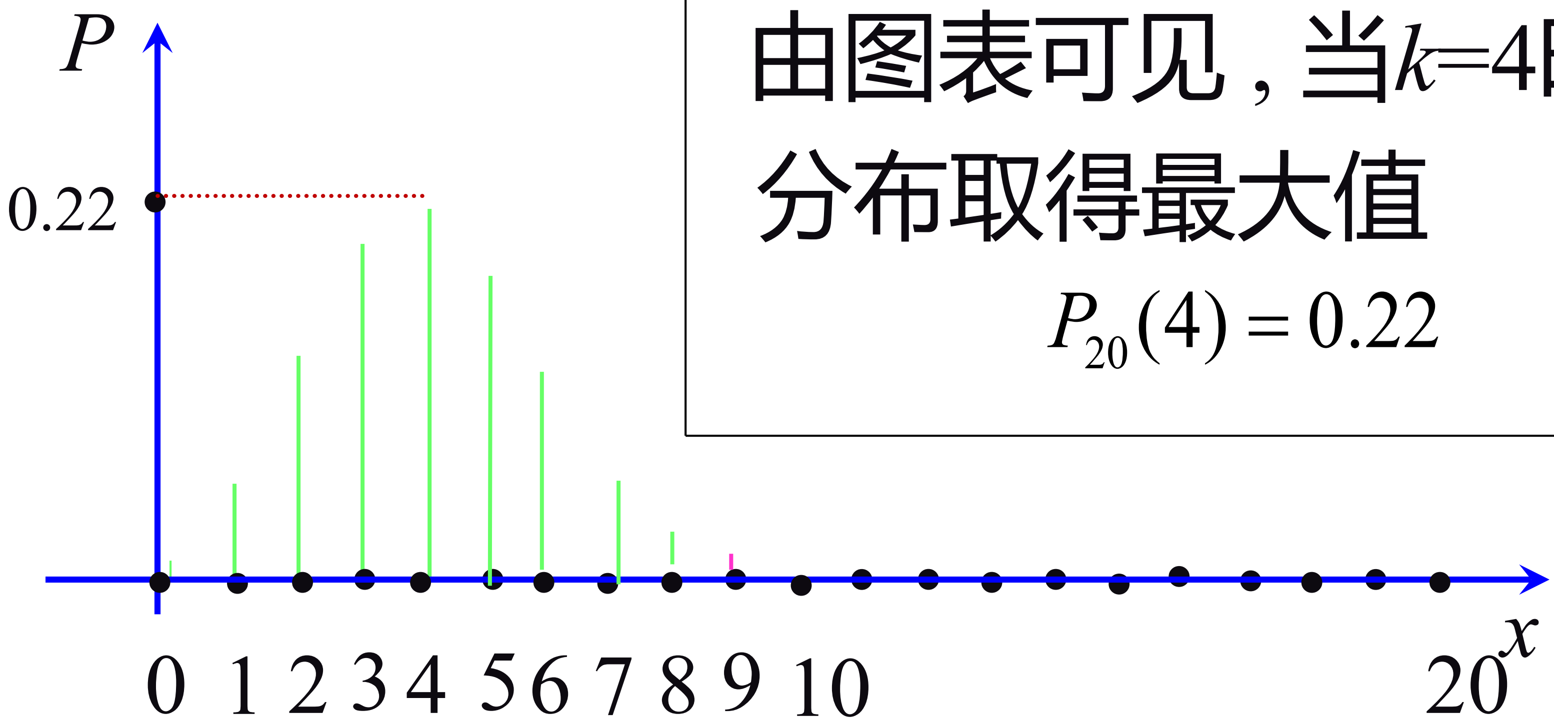
$$P_8(2) = P_8(3) = 0.273$$

此时的k称为**最可能成功次数**



设 $X \sim B(20, 0.2)$

0	1	2	3	4	5	6	7	8	9	10	$11 \sim 20$
.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

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

二项分布中最可能出现次数

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

$$\left.\begin{aligned} \frac{p_{k-1}}{p_k} &= \frac{(1-p)k}{p(n-k-1)} \le 1 \\ \frac{p_k}{p_{k+1}} &= \frac{(1-p)(k+1)}{p(n-k)} \ge 1 \end{aligned}\right\} \Rightarrow (n+1)p - 1 \le k \le (n+1)p$$

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

$\lfloor x \rfloor$ 向下取整 $\lceil x \rceil$ 向上取整

$[x]$ 取整，根据上下文约定，一般是向下取整

例 独立射击400次, 命中率为0.01,

求 (1) 最可能命中次数及相应的概率;

(2) 命中次数不少于3次的概率.

令X表示命中次数,则 $X \sim B(400,0.01)$

(1) $k=[(n+1)p]=[(400+1)0.01]=4$

$$P(X=4)=C_{400}^{4}(0.01)^{4}(0.99)^{396} \approx 0.1954$$

(2) $P(X \geq 3)=\sum_{k=3}^{400} P(X=k)$

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

$=1-\sum_{k=0}^{2} C_{400}^{k}(0.01)^{k}(0.99)^{400-k}$

问题 如何计算

$P(X \geq 300)$?

泊松近似

二项分布性质

$$X_1 \sim B(n_1, p) \quad X_2 \sim B(n_2, p)$$

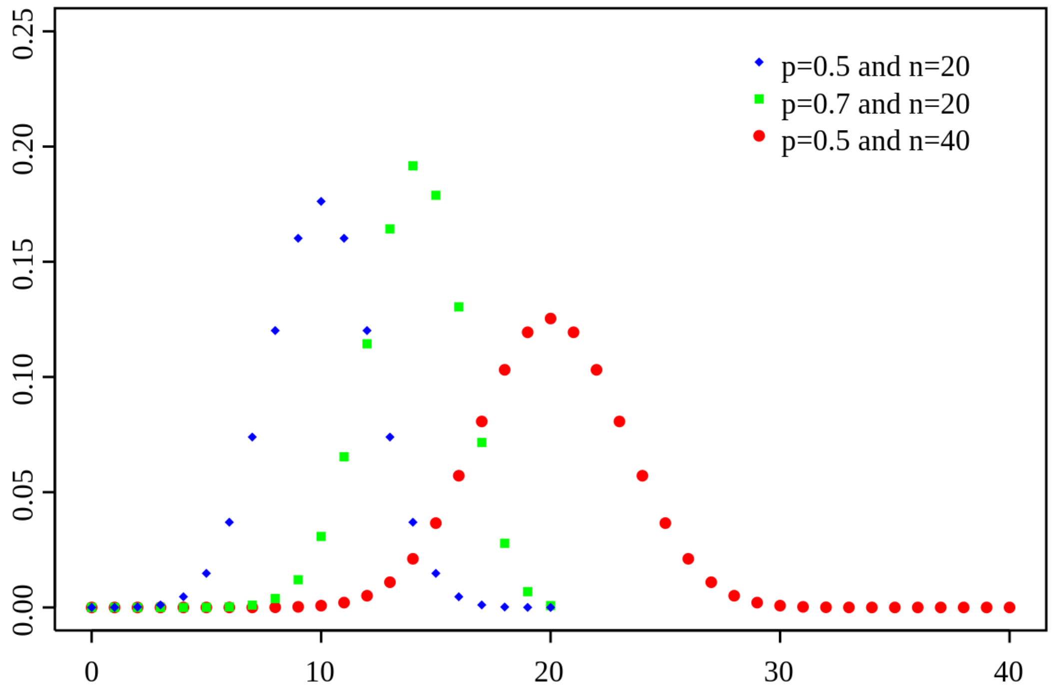
then

$$X_1 + X_2 \sim B(n_1 + n_2, p)$$

二项分布

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

- 注意比较 n, p 不同时图像的特点



5. 泊松分布Poisson Distribution

重要

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 的泊松(Poisson)分布. 记作 $X\sim\pi(\lambda)$ 或 $P(\lambda)$

应用场合

适合于描述单位时间内随机事件发生次数的概率分布。在某个时段内：

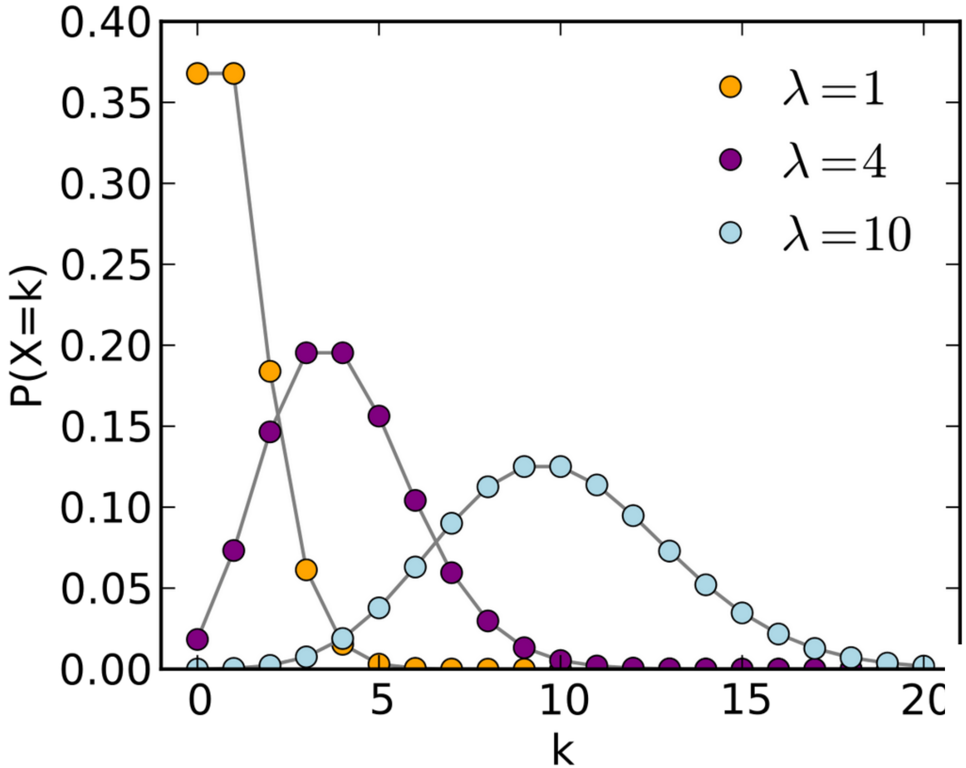
市级医院急诊病人数.

某地区拨错号的电话呼唤次数.

某地区发生的交通事故的次数.

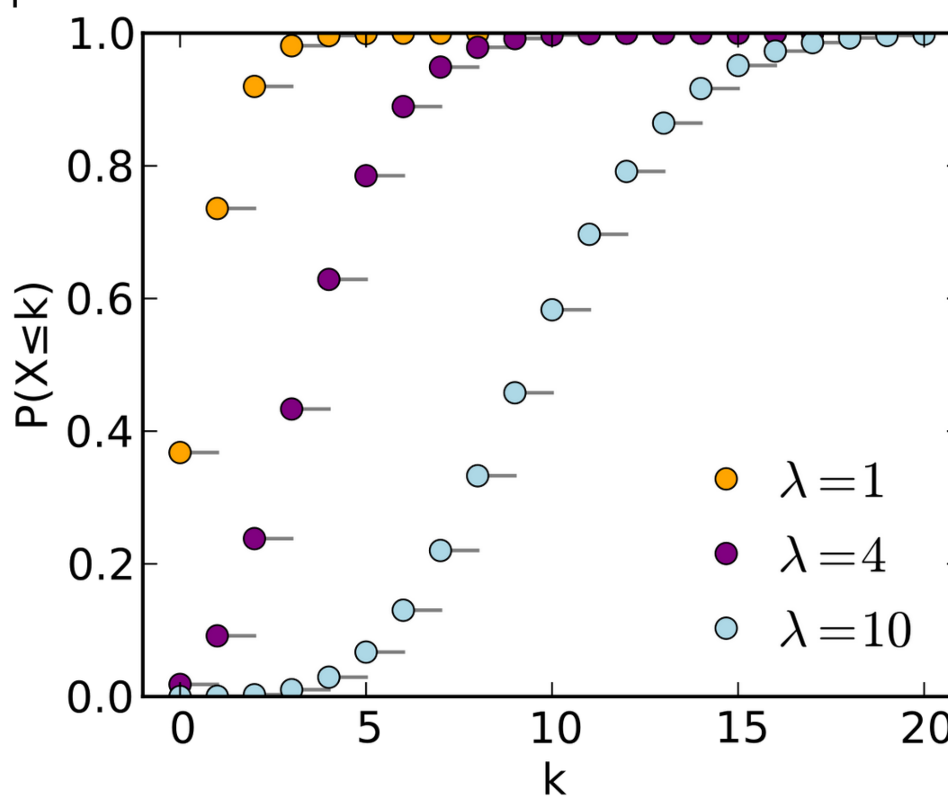
一本书一页中的印刷错误数.

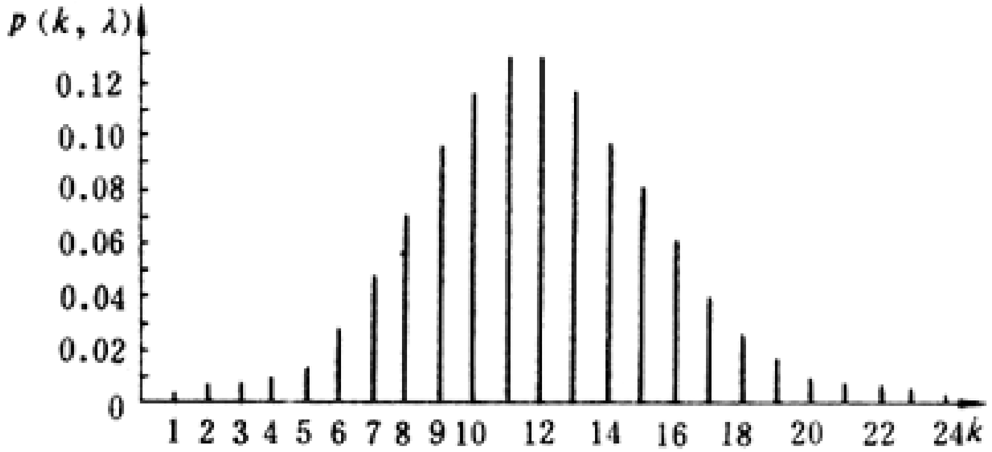
泊松分布的图形特点



Probability Distribution

Cumulative distribution function





泊松分布$P(\lambda)$ ($\lambda=12$)

泊松分布中**最可能出现次数**

当$\lambda=$ 整数时,在λ与$\lambda-1$处的概率取得最大值

当$\lambda\neq$ 整数时, 在 $[\lambda]$处的概率取得最大值

泊松分布性质

$$X_1 \sim \pi(\lambda_1) \quad X_2 \sim \pi(\lambda_2)$$

then

$$X_1 + X_2 \sim \pi(\lambda_1 + \lambda_2)$$

例 一家商店由过去的销售记录知道，某种商品每月的销售数（随机变量）服从参数 $\lambda=5$ 的泊松分布，为了以95%以上的把握保证不脱销（销售数<进货数），问商店在月底至少应进该种商品多少件？

设该商品每月的销售数为X，月底应进m件商品

$$P(X\leq m)>0.95 \qquad P(X>m)\leq 0.05$$

$$\sum_{k=m+1}^{\infty}\frac{e^{-5}5^k}{k!}\leq 0.05$$

查泊松分布表得 $\displaystyle\sum_{k=10}^{\infty}\frac{e^{-5}5^k}{k!}\approx 0.032,\quad \sum_{k=9}^{\infty}\frac{e^{-5}5^k}{k!}\approx 0.068$

$m+1=10,$ $\boxed{m=9\text{件}}$

二项分布的泊松近似

当试验次数n很大时，计算二项概率变得很麻烦，必须寻求近似方法.

历史上，泊松分布是作为二项分布的近似，于1837年由法国数学家泊松引入的.

我们先来介绍二项分布的泊松近似，后面我们将介绍二项分布的正态近似.

Possion定理

重要

设 $np=\lambda>0$，则对固定的 k

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

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

结论 二项分布的极限分布是 Poisson 分布

若 $X\sim B(n,p)$，则当 n 较大，p 较小，则

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

$n>10,\ p<0.1$时近似效果较好

利用Poisson定理再求前例

(2) 命中次数不少于3次的概率.

令X表示命中次数,则 $X \sim B(400, 0.01)$

$$\lambda = np = 4$$

泊松近似

查附表3泊松分布表

$$(2)\ P(X \geq 3) = \sum_{k=3}^{400} P(X = k) \approx \sum_{k=3}^{\infty} \frac{4^k}{k!} e^{-4} = 0.7619$$

同理,(1) $P(X = 4) = P(X \geq 4) - P(X \geq 5)$

$$\approx 0.1954$$

例 保险公司里有2500人参加某种事故保险，每人每年付120元保险费，在一年中一个人发生此种事故的概率为0.002（随机变量），发生事故时家人可向保险公司领得20000元. 问：

(1) 对该项保险保险公司亏本（赔偿>收入）的概率有多大？

(2) 该项保险的利润不少于10万元（收入-赔偿）的概率有多大？

(1)总收入：$120 \times 2,500 = 300,000$ 元，设有 x 人出现意外，则要支付赔偿金 $20,000x$ 元，只要 $20\,000x > 300\,000$，即 $x > 15$ 人，保险公司亏本。

令X表示出事故人数,则 $X \sim B(2500, 0.002)$

亏本

$$\lambda = np = 5$$

泊松近似

$$(1)\ P(X > 15) \approx \sum_{k=16}^{2500} \frac{5^k}{k!} e^{-5} = 6.9 \times 10^{-5}$$

几乎不亏本

(2) 获利不少于 10 万，即 $300{,}000 - 20{,}000x \geq 100{,}000$，$x \leq 10$人, 故:

$$P(X \leq 10) = 1 - P(X \geq 11) \approx 1 - \sum_{k=11}^{2500} \frac{5^k}{k!} e^{-5} = 0.9863$$

利润不少于10万

可能性极大

思考 设同类型设备100台，每台工作相互独立，每台设备发生故障的概率都是 0.01（随机变量）. 一台设备发生故障可由一个人维修. 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（同时故障数>工人数）的概率小于0.01?

解 设需要配备 n 个维修工人,

设 X 为同时发生故障的设备台数,

则 $X \sim B(100, 0.01)$

泊松近似

$$P(X > n) = \sum_{k=n+1}^{100} C_{100}^{k}(0.01)^{k}(0.99)^{100-k} < 0.01$$

$$\lambda = 100 \times 0.01 = 1$$

$$P(X > n) \approx \sum_{k=n+1}^{\infty} e^{-1} \frac{1^k}{k!} < 0.01$$

查附表3得 $n+1=5$

$$n = 4$$

至少要配备4名维修工人

Summary

重要

1、超几何分布
2、几何分布
3、两点分布
4、二项分布
5、泊松分布