第四节　二项分布与泊松分布

一　二项分布的性质及其计算

在n重伯努利试验中事件A出现k次的概率为

$$b(k;n,p)=C_n^k p^k q^{n-k},\quad k=0,1,2,\cdots,n$$

此概率分布称为二项分布.

1. $b(k;n,p)$的计算：

对部分n,p可直接查$b(k;n,p)$的数值. $b(k;n,p)=b(n-k;n,1-p)$. 只须给出$p\leq 0.5$的数值即可;

2. 二项分布的最可能次数

$$\frac{b(k;n,p)}{b(k-1;n,p)}=1+\frac{(n+1)p-k}{kq}$$

若$(n+1)p$不是整数． 不妨设$m=\left[(n+1)p\right]$. 由上式得

$$b(0;n,p)<b(1;n,p)<\cdots<b(m;n,p)$$
$$>b(m+1;n,p)>\cdots>b(n;n,p).$$

从而$b(m;n,p)$为最大概率.称m为二项分布的最可能次数.

若$(n+1)p$是整数． 不妨设$m=(n+1)p$. 由上式得

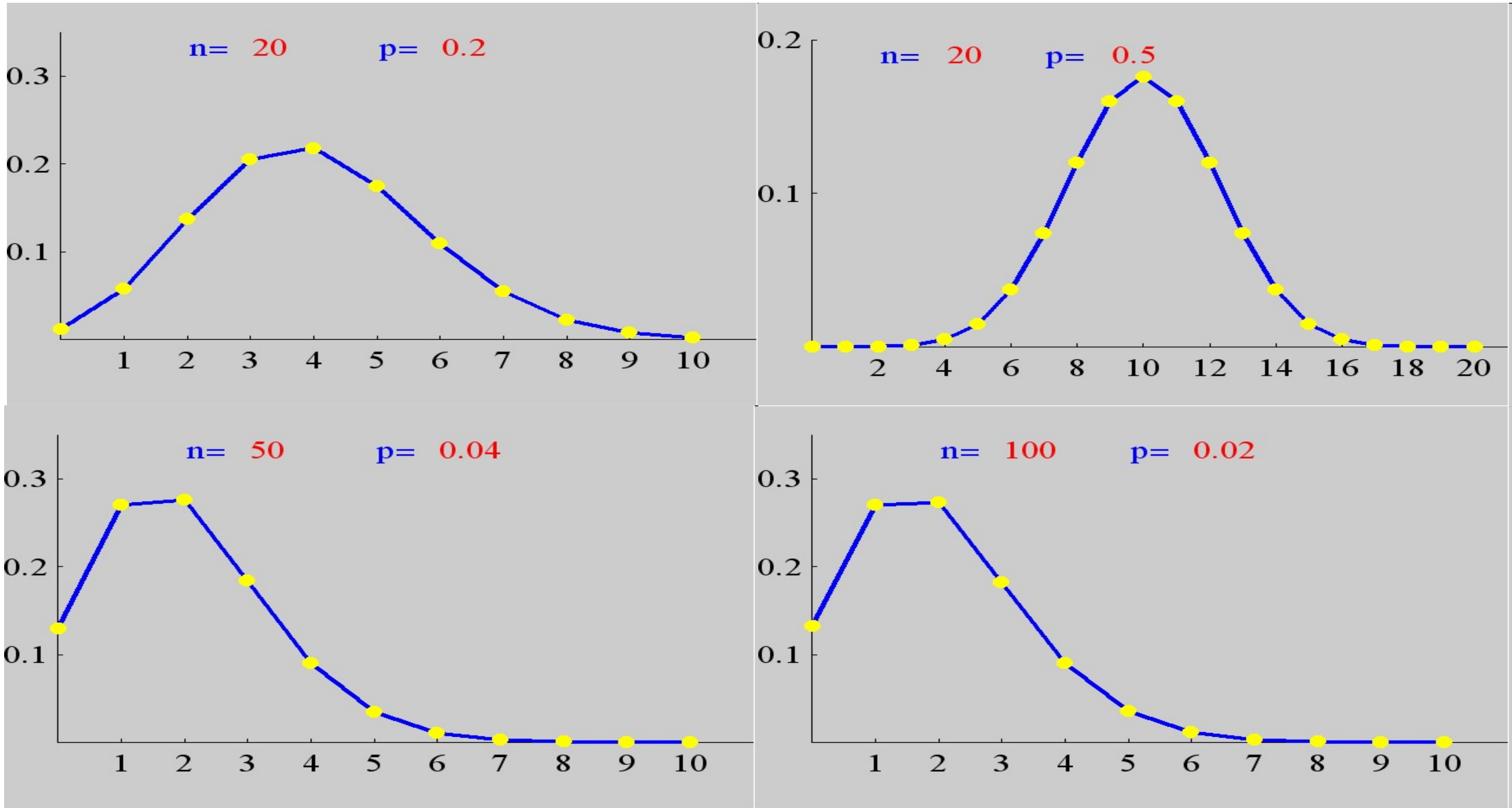
$$b(0;n,p)<b(1;n,p)<\cdots<b(m-1;n,p)=b(m;n,p)$$
$$>b(m+1;n,p)>\cdots>b(n;n,p).$$

从而$b(m-1;n,p)=b(m;n,p)$为最大概率.

二项分布的最可能次数$m-1$和m.

注：当试验次数较大时，$b(m;n,p)\sim(2\pi npq)^{-0.5}$.
其绝对数值较小（趋于零）.

二项分布的图形



例1 按规定，某种型号的电子元件的使用寿命超过**20000**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？

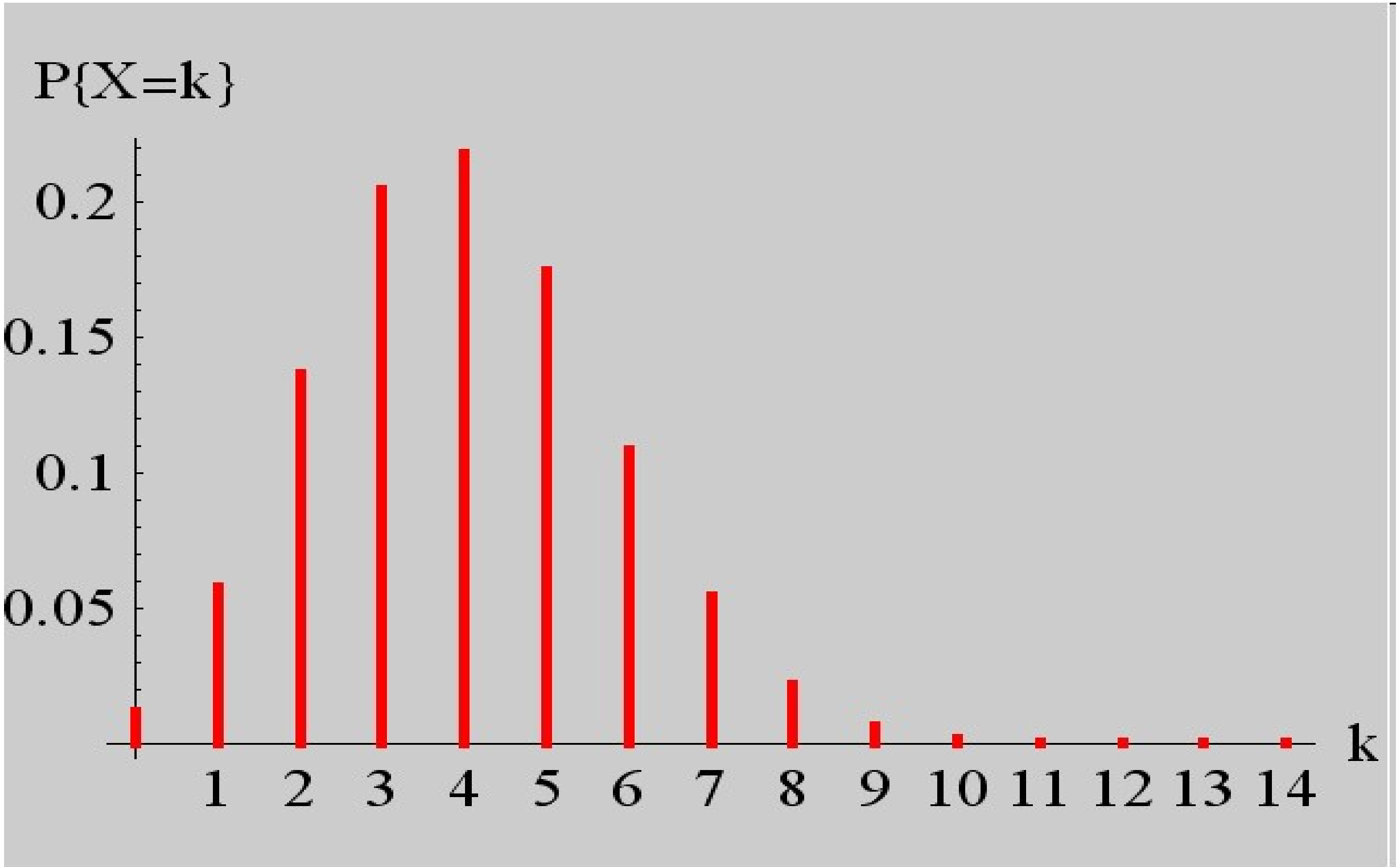
解：设X为**20**件产品中一级品的件数，则

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

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

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

概率分布图示



二　二项分布的泊松逼近

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

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

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.

例题　书本P100-101 (人寿保险、机票超售、车间用电、分子运动)——n 较大时，计算复杂.

例题（人寿保险）根据生命表知道某年龄段保险者里，一年中每个人死亡的概率为0.005，现有10000个这类人参加人寿保险，试求在未来一年中在这些保险者里面，(1)有**40**个人死亡的概率；(2)死亡人数不超过70人的概率.

解：利用伯努利概型来近似，

$$n = 10000, \quad p = 0.005.$$

(1) $b(40;10000,0.005) = C_{10000}^{40} \cdot 0.005^{40} \cdot 0.995^{9960};$

(2) $P(\mu \le 70) = \sum_{k=0}^{70} b(k;10000,0.005)$

$$= \sum_{k=0}^{70} C_{10000}^{k} \cdot 0.005^{k} \cdot 0.995^{10000-k}$$

例题 （机票超售）某航线历史资料表明：订单旅客有5%不来登机，问一架200座飞机应出售多少张机票？

解： 设超售m张机票. 利用伯努利概型来近似，

$$n = 200 + m, \quad p = 0.95.$$

记登机的旅客数为μ，则$\mu > 200$时，有人无法登机.

$$P(\mu > 200) = \sum_{k=201}^{200+m} b(k; 200+m, 0.95)$$

例题 书本（车间用电、分子运动）

——n 较大时，计算复杂.（略）

二　二项分布的泊松逼近

定理2.4.1（泊松定理）　在独立试验中，以p_n代表事件A在试验中出现的概率，它与试验次数n有关，如果 $\lim\limits_{n\to+\infty} np_n = \lambda$，则有

$$\lim_{n\to+\infty} b(k;n,p) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

证明：记$np_n = \lambda_n$，则

$$b(k;n,p) = C_n^k p_n^k (1-p_n)^{n-k}$$

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

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

$$\longrightarrow\frac{\lambda^k}{k!}e^{-\lambda}.$$

故当n 较大，p较小，np大小适中时：

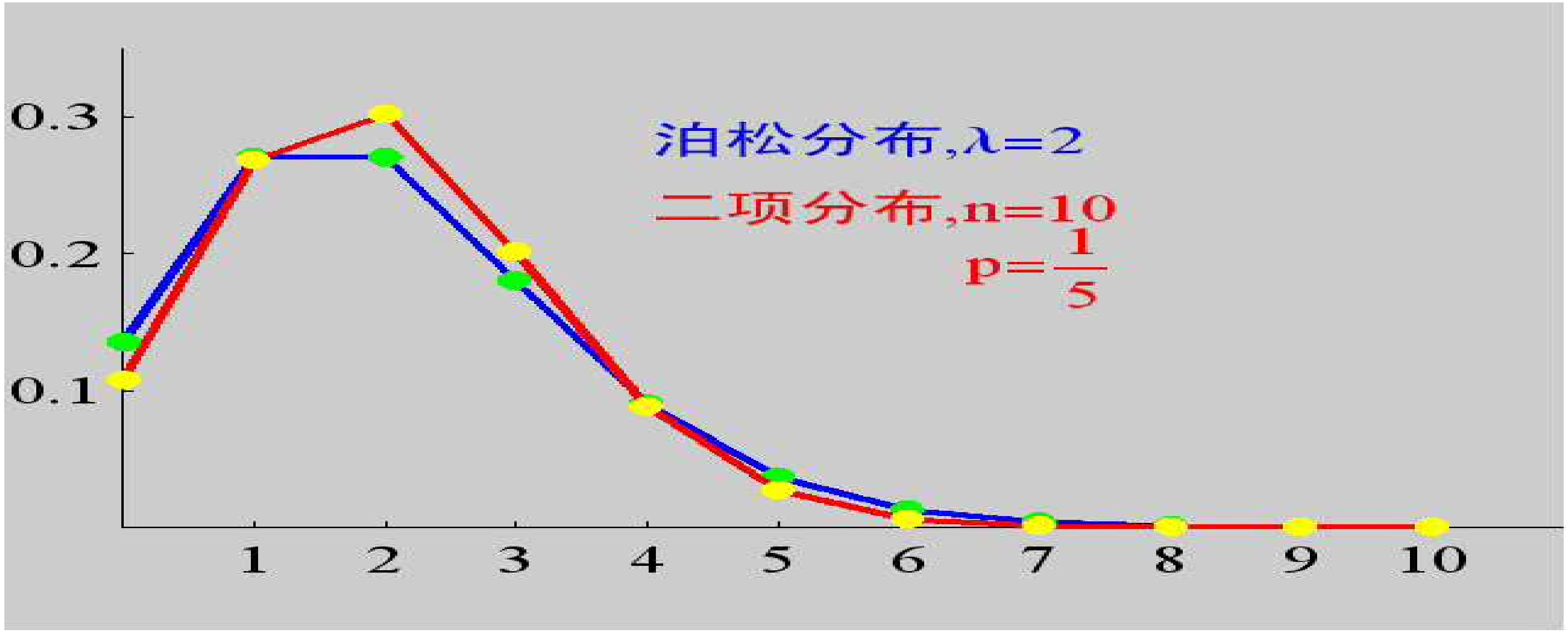
$$b(k;n,p)\approx\frac{(np)^k}{k!}e^{-np}.$$

泊松定理表明

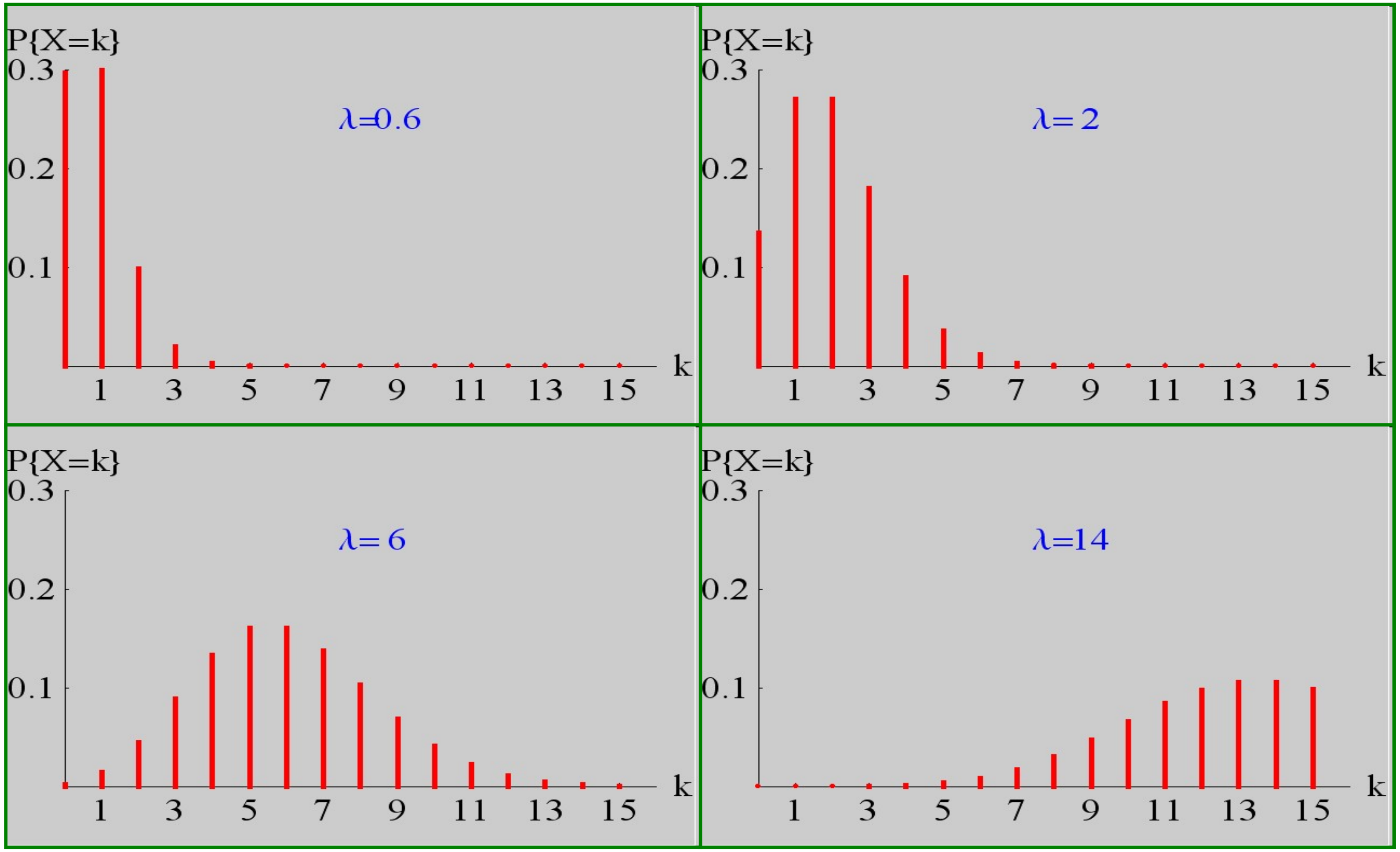
$$\text{二项分布}\xrightarrow{n\text{很大},p\text{ 很小}}\text{泊松分布}$$

泊松定理表明

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



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

三　泊松分布

引理 2.4.1　若$f(x)$是连续函数(或单调函数)，且对一切x, y（或一切$x \geq 0, y \geq 0$）成立

$$f(x)f(y) = f(x+y),$$

则

$$f(x) = a^x.$$

其中$a \geq 0$，是某一常数.

证明：由题设知对任意 x,

$$f(x) = f(\frac{x}{2})^2 \geq 0$$

因此$f(x)$非负. 又$f(1) = f(\frac{1}{n} + \cdots + \frac{1}{n}) = f(\frac{1}{n})^n$

记$a = f(1) \geq 0$，则

$$f(\frac{1}{n}) = a^{\frac{1}{n}}$$

因此，对一切正整数m及n，成立

$$f(\frac{m}{n}) = a^{\frac{m}{n}}.$$

这样, 我们得到对一切有理数成立$f(x) = a^x$,利用连续性得对无理数也成立.

单调函数时(不妨设单调上升):

设x_0是任一无理数, 寻找一单调上升趋向于x_0的有理数列$\{x_n\}_1^\infty$和另一单调下降趋于x_0的有理数列$\{y_n\}_1^\infty$.

由$x_n < x_0 < y_n$知： $f(x_n) \le f(x_0) \le f(y_n)$. 两边取极限得 $a^{x_0} \le f(x_0) \le a^{x_0}$，故 $f(x_0) = a^{x_0}$.

【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】

对$\Delta t>0$，$t>0$. 考虑$[0, t+\Delta t)$时间内来到k个呼叫

的概率$P_k(t+\Delta t)$. 由独立增量性及全概率公式

$$P_k(t+\Delta t)=P_k(t)P_0(\Delta t)+P_{k-1}(t)P_1(\Delta t)+\cdots+P_0(t)P_k(\Delta t)$$

特别地

$$P_0(t+\Delta t)=P_0(t)P_0(\Delta t)$$

由于$P_0(t)$是t 的单调下降函数. 由引理得

$$P_0(t)=a^t.$$

其中$a\geq 0$，因$P_0(t)$是概率，故$a\leq 1$. 所以$0\leq a\leq 1$.

若$a=0$，则$P_0(t)\equiv 0$. 从而$\sum_{k=1}^{\infty}P_k(t)\equiv 1$. 这说明无论时间间隔$t$ 多小都要来呼叫，这种情况我们不考虑.

若$a=1$，则$P_0(t)\equiv 1$．从而$\sum_{k=1}^{\infty}P_k(t)\equiv 0$．这说明是无论时间间隔 t 多大都不来呼叫，这种情况我们也不考虑．所以$0<a<1$，从而存在$\lambda>0$，使得

$$P_0(t)=e^{-\lambda t}$$

因此当$\Delta t\to 0$时，有

$$P_0(\Delta t)=e^{-\lambda\Delta t}=1-\lambda\Delta t+o(\Delta t)$$

$$P_1(\Delta t)=1-P_0(\Delta t)-\varphi(\Delta t)=\lambda\Delta t+o(\Delta t)$$

由于

$$P_k(t+\Delta t)=P_k(t)P_0(\Delta t)+P_{k-1}(t)P_1(\Delta t)$$

$$+\sum_{l=2}^{k}P_{k-l}(t)P_l(\Delta t).$$

$$\sum_{l=2}^{k} P_{k-l}(t)P_l(\Delta t) \le \sum_{l=2}^{k} P_l(\Delta t) \le \sum_{l=2}^{\infty} P_l(\Delta t) = o(\Delta t)$$

所以

$$\begin{aligned} P_k(t+\Delta t) &= P_k(t)P_0(\Delta t) + P_{k-1}(t)P_1(\Delta t) + o(\Delta t) \\ &= P_k(t)(1-\lambda\Delta t) + P_{k-1}(t)\cdot\lambda\Delta t + o(\Delta t) \end{aligned}$$

移项整理得

$$\frac{P_k(t+\Delta t) - P_k(t)}{\Delta t} = \lambda\left[P_{k-1}(t) - P_k(t)\right] + o(1)$$

令$\Delta t \to 0$，得

$$P_k'(t) = \lambda\left[P_{k-1}(t) - P_k(t)\right]$$

由$P_0(t) = e^{-\lambda t}$，可解得

$$P_1(t) = \lambda t e^{-\lambda t},$$

一直做下去就可得到

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

在随机过程中，这种方法将被推广.

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……