第二章 随机变量及其分布

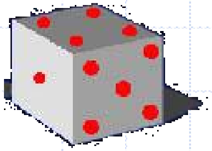
§1 随机变量

- 随机变量的概念
- 随机变量与普通函数的区别

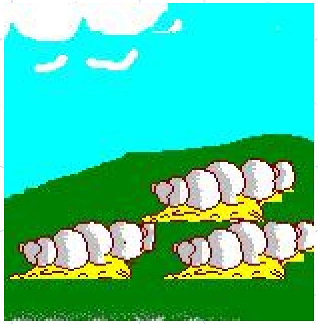
随机变量概念的产生

- 有些试验结果本身与数值有关。

例如，掷一颗骰子上面出现的点数；

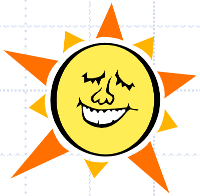


每天从火车站下车的人数；



昆虫的产卵数；

四月份青岛的最高温度；

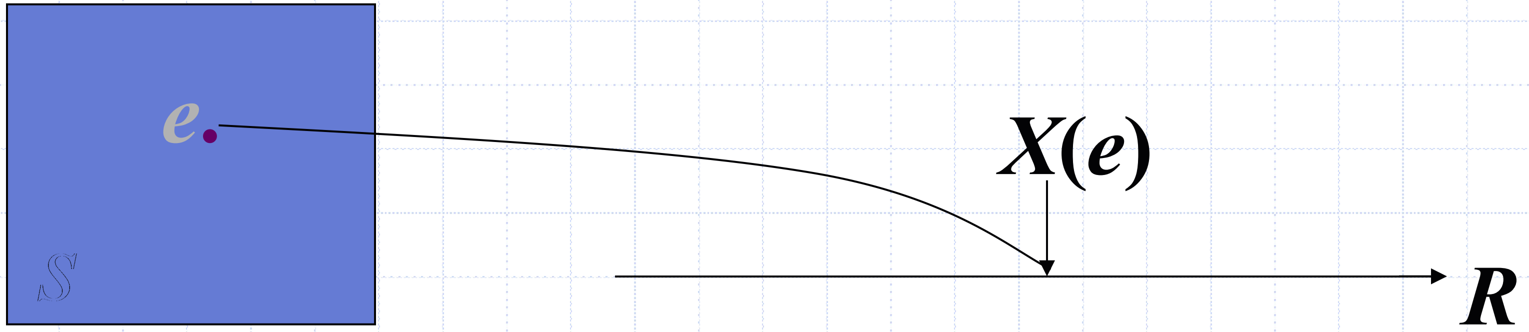


…

- 有些试验的试验结果本身与数值无关，可根据研究需要设置随机变量.

在数学上理解为定义了一种实值单值函数.

将随机试验的每一个结果,即S的每一个元素 $e \xleftrightarrow{\text{对应}}$ 实数x



例 1　袋中有3只黑球，2只白球，从中任意取出3只球．我们将3只黑球分别记作1，2，3号，2只白球分别记作4，5号，则该试验的样本空间为

$$S=\left\{\begin{matrix}(1,&2,&3) & (1,&2,&4) & (1,&2,&5)\\(1,&3,&4) & (1,&3,&5) & (1,&4,&5)\\(2,&3,&4) & (2,&3,&5) & (2,&4,&5)\\(3,&4,&5) & & & & & & & \end{matrix}\right\}$$

考察取出的3只球中的黑球的个数。

记取出的黑球数为X，则 X 的可能取值为1, 2, 3.

X 是一个变量，X 取什么值依赖于试验结果，即 X 的取值带有随机性，所以称 X为随机变量.

X 的取值情况可由下表给出：

样本点	黑球数 X	样本点	黑球数 X
$(1,\ 2,\ 3)$	3	$(1,\ 4,\ 5)$	1
$(1,\ 2,\ 4)$	2	$(2,\ 3,\ 4)$	2
$(1,\ 2,\ 5)$	2	$(2,\ 3,\ 5)$	2
$(1,\ 3,\ 4)$	2	$(2,\ 4,\ 5)$	1
$(1,\ 3,\ 5)$	2	$(3,\ 4,\ 5)$	1

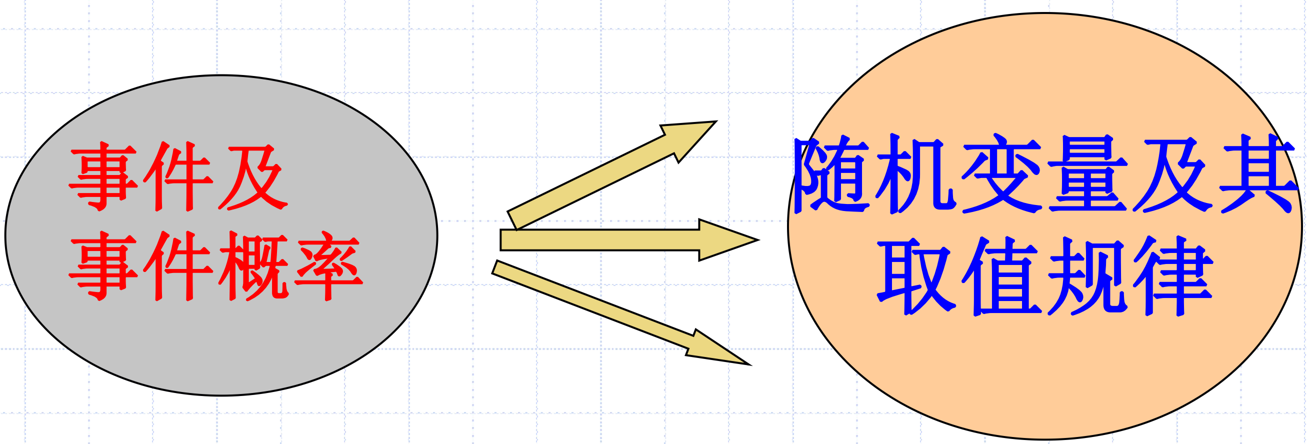
随机试验的每一个结果都对应着变量 X 的一个确定的取值，因此变量 X 是样本空间S上的函数：

$$X = X(e) \qquad (e \in S)$$

1 随机变量的概念

（1）定义 设随机试验的样本空间为 $S=\{e\}$，称定义在 S 上的实值单值函数 $X=X(e)$ 为随机变量。

通常随机变量用大写的英文字母 X、Y、Z、$\cdots$ 或希腊字母 ξ、η、ς、$\cdots$ 等来表示.



（2）用随机变量的关系式表示随机事件

$$\{e: X(e)=2\}=\{X=2\}$$

表示“取出2个黑球”这一事件；

$\{X \geq 2\}$ 表示“至少取出2个黑球”这一事件，等等.

例2 上午 8:00～9:00 在某路口观察，令：

Y：该时间间隔内通过的汽车数.

则 Y 就是一个随机变量. 它的取值为 0，1，….

$$\{50 < Y \le 100\}$$

表示“通过的汽车数大于 50 辆但不超过 100 辆”这一随机事件.

注意 Y 的取值是可列无穷个！

例 3 观察某电子元件的寿命（单位：小时），令

Z：该电子元件的寿命.

则Z 就是一个随机变量．它的取值为所有非负实数.

$$\{Z > 1000\}$$

表示“该电子元件的寿命大于 1000小时”这一随机事件.

注意 Z 的取值是不可列无穷个！

例 4 掷一枚硬币，令：

$$X=\begin{cases}1 & \text{掷硬币出现正面；}\\ 0 & \text{掷硬币出现反面．}\end{cases}$$

则X是一个随机变量.

说 明：

在同一个样本空间上可以定义不同的随机变量.

(3)随机变量的取值有一定的概率

随机变量的取值随试验的结果而定,而试验的各个结果出现有一定的概率,

——因而随机变量的取值有一定的概率.

例如 例1 中，随机事件 $\{X = 2\}$ 对应于事件

$$A = \{(1,2,4),(1,2,5),(1,3,4),(1,3,5),(2,3,4),(2,3,5)\}$$

当且仅当事件A发生时有随机事件$\{X = 2\}$.

A 发生的概率 $P(A) = P\{X = 2\} = \dfrac{6}{10}$

2 随机变量与普通函数有本质区别

(1) 随机变量的取值随试验的结果而定, 在试验之前只知它可能取值的范围，但不能预知它将取什么值。

—普通函数的取值是完全确定的。

(2) 试验结果的出现具有一定的概率, 因而随机变量的取值具有一定的概率。

—普通函数的取值没有概率要求。

(3) 随机变量是定义在样本空间 S上的函数, S 中的元素不一定是实数。

—而普通函数只是定义在实数轴上.

§2离散型随机变量

- 离散型随机变量的分布律与性质
- 一些常用的离散型随机变量

一、离散型随机变量的分布律与性质

1)离散型随机变量的定义

如果随机变量 X 的取值是有限个或可列无穷个，则称 X 为离散型随机变量.

2)离散型随机变量的分布律

设离散型随机变量 X 的所有可能取值为

$$x_1,\quad x_2,\quad \cdots,\quad x_k,\quad \cdots$$

并设 $P\{X=x_k\}=p_k \qquad (k=1,\ 2,\ \cdots)$

则称上式或

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

为离散型随机变量 X 的分布律.

3) 离散型随机变量分布律的性质:

(1)非负性:对任意的自然数 k，有 $p_k \geq 0$;

(2)归范性: $\sum_k p_k = 1.$

重要性质,有重要应用

注: 性质(2)常用来决定分布律中的常数,困难之处在于无穷级数的求和.

要求: 会求离散型随机变量的分布律.

弄清两点: a)所有可能的取值; b) 取值的概率.

例1 如果随机变量X的分布律为

$$P\{X=k\}=c\frac{\lambda^k}{k!}\quad(k=1,2,\cdots).$$

（其中$\lambda>0$为常数）试确定未知常数c.

解： 由分布律的性质有$\sum_{k=1}^{\infty}c\frac{\lambda^k}{k!}=c\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}=1,$

而 $\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}-1=e^{\lambda}-1$

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

所以 $c=\dfrac{1}{e^{\lambda}-1}.$

例 2 从1～10这10个数字中随机取出5个数字，令 X：取出的5个数字中的最大值．试求X的分布律．

解： X 的可能取值为5，6，7，

$$P\{X=k\}=\frac{C_{k-1}^{4}}{C_{10}^{5}} \qquad k=5,\ 6,\ \cdots,\ 10.$$

具体写出，即可得 X 的分布律：

X	5	6	7	8	9	10
P	$\frac{1}{252}$	$\frac{5}{252}$	$\frac{15}{252}$	$\frac{35}{252}$	$\frac{70}{252}$	$\frac{126}{252}$

例 3 将 1 枚硬币掷 3 次，令

X：出现的正面次数与反面次数之差.

试求：（1）X 的分布律；(2) $P\{0.5 \le X < 3\}$.

解: X 的可能取值为 -3，-1，1，3.

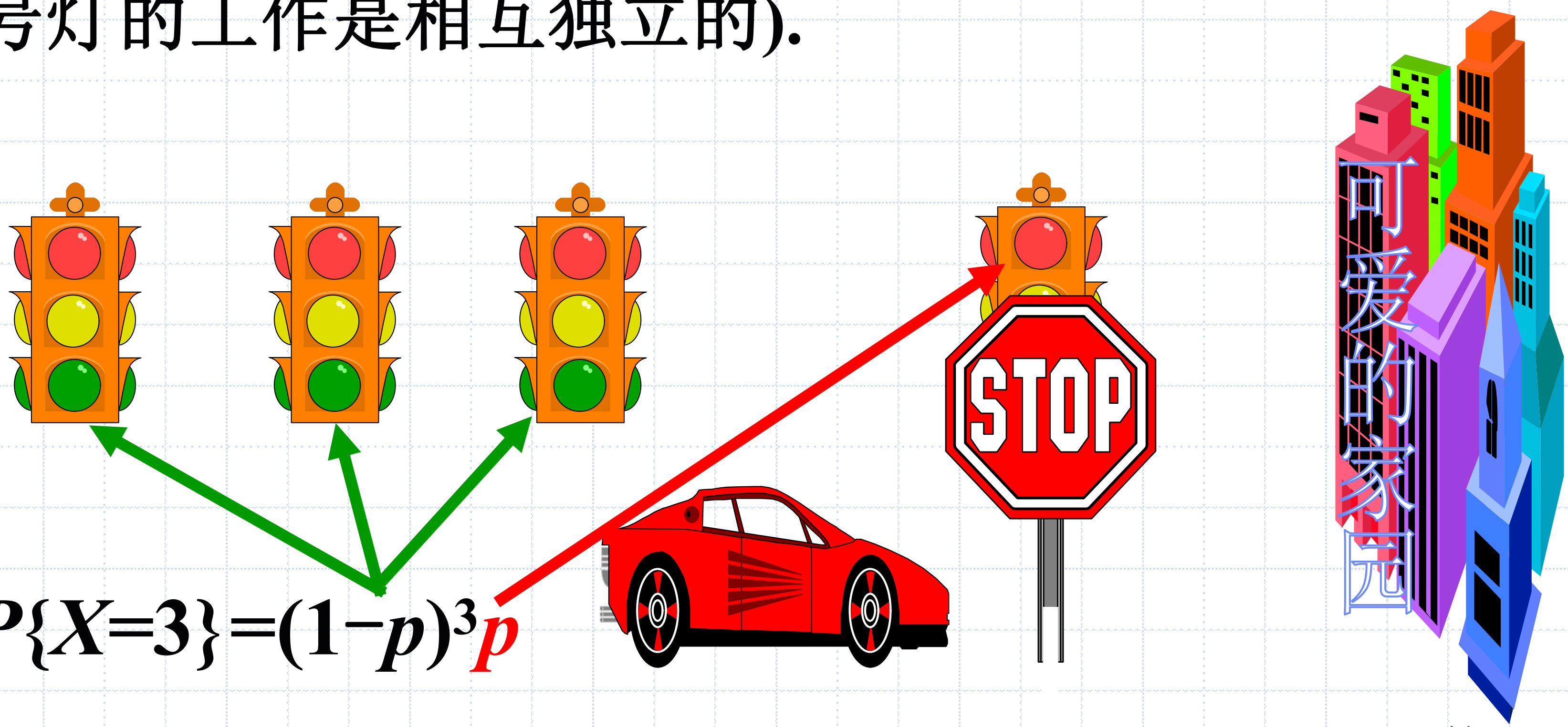
并且分布律为

X	-3	-1	1	3
P_k	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$P\{0.5 \le X < 3\} = P\{X = 1\} = \frac{3}{8}.$$

例 4

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，求 X 的分布律. (信号灯的工作是相互独立的).



$$P\{X=3\}=(1-p)^3p$$

例 4(续)

解： 以 p 表示每盏信号灯禁止汽车通过的概率，则 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 $P\{X=k\}=(1-p)^k p$，$k=0,1,2,3$

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

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、一些常用的离散型随机变量

1) （0–1）分布

如果随机变量 X 的分布律为

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

或

X	0	1
P	$1-p$	p

则称随机变量 X 服从参数为 p 的（0–1）分布或两点分布.

记作 $X\sim b(1,\ p)$ （其中 $0\le p\le 1$ 为参数）

（0-1）分布的概率背景

设试验E只有两个可能结果：A 及 $\overline{A}$，则称E为伯努利（*Bernoulli*）试验。

进行一次*Bernoulli*试验，A是随机事件。设：

$$P(A)=p\,(0<p<1),\qquad P(\overline{A})=1-p=q$$

设X表示这次*Bernoulli*试验中事件A发生的次数.

或者设

$$X=\begin{cases}1 & 若事件A发生;\\ 0 & 若事件A不发生.\end{cases}$$

则 $X\sim b(1,p)$

2）二 项 分 布

如果随机变量 X 的分布律为

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

则称随机变量 X 服从参数为 $(n,\ p)$ 的二项分布，

记作 $X \sim b(n,\ p)$

（其中 n 为自然数，$0<p<1$ 为参数）

分布律的验证

(1) 由于 $0 < P < 1$ 以及 n 为自然数，可知

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

(2) 又由二项式定理，可知

$$\sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = [p + (1-p)]^n = 1$$

所以

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

是分布律.

说 明

显然，当 n=1 时

$$X \sim b(1, \quad p)$$

此时，X 服从(0−1)分布.

这说明，(0−1)分布是二项分布的一个特例.

二项分布的概率背景

定义：如果随机试验E只有两个可能结果：A（成功）及$\overline{A}$（失败），则称 E 为*Bernoulli* 试验.

设 $P(A)=p$ ， $P(\overline{A})=1-p=q$

将E独立重复地进行n 次，则称这一串重复地独立试验为n 重 ***Bernoulli***（伯努利）试验。

重复：指在每次试验中$P(A)=p$ 保持不变；

独立：指各次试验的结果互不影响。

令 X 表示这 n 重 ***Bernoulli*** 试验中事件A发生的次数.

则 $X \sim b(n, \ p)$.

说明：

设 $A_i=\{$第 i 次试验 A 出现$\}$,则

$$\{X=k\}=A_1\cdots A_k\overline{A}_{k+1}\cdots\overline{A}_n\cup\overline{A}_1A_2\cdots A_{k+1}\overline{A}_{k+2}\cdots\overline{A}_n\cup$$

$$\cdots\cup\overline{A}_1\overline{A}_2\cdots\overline{A}_{n-k}A_{n-k+1}\cdots A_n$$

在 n 次试验中，指定 k 次出现 A(成功)，其余 $n-k$ 次出现 $\overline{A}$(失败)，这种指定的方法共有 C_n^k 种.

所以

$$P\{X=k\}=C_n^k p^k q^{n-k} \qquad (q=1-p)$$

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

例5 一大批产品的次品率为0.1，现从中取出15件. 试求下列事件的概率：

B={ 取出的15件产品中恰有2件次品 }

C={ 取出的15件产品中至少有2件次品 }

解: 由于从一大批产品中取15件产品，故可看作是一15重*Bernoulli*试验.

$$A=\{取出一件产品为次品\},\quad 则\quad P(A)=0.1.$$

所以，$P(B)=C_{15}^{2}\times 0.1^{2}\times 0.9^{13}$

$$P(C)=1-P(\overline{C})$$

$$=1-C_{15}^{0}\times 0.1^{0}\times 0.9^{15}-C_{15}^{1}\times 0.1\times 0.9^{14}$$

例 6 一张考卷上有5道选择题，每道题列出4个可能答案，其中只有一个答案是正确的．某学生靠猜测至少能答对4道题以上的概率是多少？

解： 每答一道题相当于做一次*Bernoulli*试验，

$A=\{答对一道题\}$，则 $P(A)=\dfrac{1}{4}$

则答5道题相当于做5重Bernoulli试验．

设X表示该学生靠猜测能答 对的题数，

则 $X\sim b\left(5,\ \dfrac{1}{4}\right)$

所以

$$P\{\text{至少能答对4道题}\}=P\{X\geq 4\}$$

$$=P\{X=4\}+P\{X=5\}$$

$$=C_5^4\left(\frac{1}{4}\right)^4\cdot\frac{3}{4}+\left(\frac{1}{4}\right)^5$$

$$=\frac{1}{64}$$

例 7 某人进行射击，设每次射击的命中率为0.02，对同一目标独立射击400次，试求至少击中两次的概率？

解： 一次射击是一次试验，有射中和射不中两种结果。对目标进行400次射击相当于做400重*Bernoulli* 试验．令：

X 表示400次射击中击中目标的次数．

则由题意 $X \sim b(400,\ 0.02).$

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

所求概率为

$$P\{X\geq 2\}=1-P\{X=0\}-P\{X=1\}=0.9972$$

注意:

(1) 虽然每次射击的命中率不高，但射击400次而至少击中两次是几乎可以肯定的。

(2)若射手在400次射击中击中目标的次数竟不到两次，由 $P\{X<2\}\approx 0.003$ 很小，由实际推断原理，将“怀疑每次射击的命中率为0. 02”，即命中率达不到0. 02。

3) 泊松(*Poisson*)分布

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

$$P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!}, k=0,1,2,\cdots, \text{其中}\lambda>0\text{是常数}$$

则称随机变量 X 服从参数为 λ 的 泊松分布，记

$$X \sim \pi(\lambda)$$

注意:

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

泊松（*Poisson* ）分布的应用

- 泊松分布是概率论中重要的分布之一.
- 自然界及工程技术中的许多随机指标都服从泊松分布.
- 例如，可以证明，

电话总机在某一时间间隔内收到的呼叫次数，

放射物在某一时间间隔内发射的粒子数，

容器在某一时间间隔内产生的细菌数，

某一时间间隔内来到某服务台要求服务的人数，等等，在一定条件下，都是服从泊松分布的.

例 8

设一个人在一年内的感冒次数服从参数 $\lambda=5$ 的 *Poisson* 分布，现有一种预防感冒的药，它对30%的人来讲，可将上述参数 λ 降为 $\lambda=1$（疗效显著）；对另45%的人来讲，可将参数 λ 降为 $\lambda=4$（疗效一般）；而对其余25%的人来讲，则是无效的．现某人服用此药一年，在这一年中，他得了3次感冒，试求此药对他“疗效显著”的概率．

解：设 B={ 此人在一年中得3次感冒 }

$A_1=\{该药疗效显著\}$, $A_2=\{该药疗效一般\}$,

$A_3=\{该药无效\}$, 则由Bayes公式，得

$$P(A_1|B)=\frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1)+P(A_2)P(B|A_2)+P(A_3)P(B|A_3)}$$

$$=\frac{0.30\times\frac{1^3}{3!}e^{-1}}{0.30\times\frac{1^3}{3!}e^{-1}+0.45\times\frac{4^3}{3!}e^{-4}+0.25\times\frac{5^3}{3!}e^{-5}}$$

$$=0.1301$$

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

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

证明： $C_n^k p_n^k \left(1-p_n\right)^{n-k}$

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

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

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

对于固定的 k，有 $\lim\limits_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n=e^{-\lambda},\lim\limits_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}=1$

所以，

$$\lim_{n\to\infty}C_n^k p_n^k(1-p_n)^{n-k}$$

$$=\frac{\lambda^k}{k!}\cdot\lim_{n\to\infty}\left\{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right\}\cdot\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n\cdot\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}$$

$$=\frac{\lambda^k}{k!}\cdot e^{-\lambda}$$

泊松（Poisson）定理的应用

由 泊松 定理，可用泊松分布来近似二项分布。

若随机变量 $X \sim b(n, \ p)$,

$\boxed{\text{则当}n\text{比较大，}p\text{比较小时，}}$

令： $\lambda = np$

则有 $P\{X = k\} = C_n^k p^k (1-p)^{n-k}$

$$\approx \frac{\lambda^k}{k!} e^{-\lambda}$$

例 9 保险公司售出某种寿险（一年）保单2500份.每单交保费100元，当被保人一年内死亡时，家属可从保险公司获得2万元的赔偿.若此类被保人一年内死亡的概率为0.001，求

（1）保险公司亏本的概率;

（2）保险公司获利不少于10万元的概率.

解：设此类被保人一年内死亡的人数为 X，

则 $X \sim b\,(2500，0.001)$.

（1）$P\{\text{保险公司亏本}\} = P\{25-2X<0\}$

$$=1-P\{X\le 12\}=1-\sum_{k=0}^{12}C_{2500}^{k}\cdot(0.001)^{k}\cdot(0.999)^{2500-k}$$

$$\approx 1-\sum_{k=0}^{12}\frac{2.5^{k}}{k!}e^{-2.5}=0.000002.$$

（2）$P\{\text{保险公司获利不少于10万元}\}=P\{25-2X\ge 10\}$

$$=P\{X\le 7\}\approx\sum_{k=0}^{7}\frac{2.5^{k}}{k!}e^{-2.5}=0.995753.$$

本节小结：

1）离散型随机变量的分布律及其性质；

2）两点 (0–1)分布、二项分布、泊松分布；

要求：

1）掌握分布律的性质；

2）熟练运用（0–1）分布、二项分布、泊松分布几个分布模型解决实际问题，特别是二项分布。