第二章　条件概率与统计独立性

§2.1 条件概率,全概率公式与贝叶斯公式

一. 条件概率

条件概率的直观定义:

某个事件发生的可能性大小经常会受到另一相关事件发生与否的影响.若在事件B已发生的条件下，事件A发生的概率为p,则称p为在已知B发生的条件下A发生的条件概率，记为P(A|B).

问题的提出:

1) 10个人摸彩，有3张中彩.

问：第1个人中彩的概率为多少?

第2个人中彩的概率为多少?

2) 10个人摸彩，有3张中彩.

问：已知第1个人没摸中，

第2个人中彩的概率为多少?

例 2.1 假定生男生女是等可能.若已知某一个家庭有俩孩子,求这个家庭有一个男孩、一个女孩的概率；若已知这个家庭至少有一个女孩，求这家有一个男孩、一个女孩的概率。

解：设A表示“这个家庭有一个男孩、一个女孩”；

B表示“这个家庭至少有一个女孩”。

于是，所求概率分别 P(A),P(A|B).

由题意知样本空间和事件分别可表示为

$$\Omega=\{(男,男),(男,女),(女,男),(女,女)\}$$

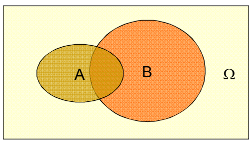
$$A=\{(男,女),(女,男)\}$$

$$B=\{(男,女),(女,男),(女,女)\}$$

所以有 $P(A)=\dfrac{1}{2}, P(A|B)=\dfrac{2}{3}$

注意,求解例2.1中$P(A|B)$过程可进行如下转换:

$$P(A|B)=\frac{2}{3}=\frac{2/4}{3/4}=\frac{P(AB)}{P(B)}$$



➤ 条件概率的概念

定义：设$(\Omega,\mathcal{F},P)$是一个概率空间，$B\in\mathcal{F}$,且P(B)>0,则对任意$A\in\mathcal{F}$,记

$$P(A\mid B)=\frac{P(AB)}{P(B)},$$

并称P(A|B)为已知事件B发生的条件下,事件A发生的条件概率(conditional probability).

以后,若出现条件概率P(A|B)时，都假定P(B)>0.

例1 体检发现，某地区自然人群中，每10万人内平均有40人患原发性肝癌，有34人甲胎球蛋白含量高，有32人患原发性肝癌又出现甲胎球蛋白含量高。现从这一地区随机抽查一人，发现其甲胎球蛋白含量高，求其患原发性肝癌的概率有多大？若在这个人群中，已知一人患原发性肝癌，求该人甲胎球蛋白含量高的概率？

解：设A表示“所抽人患原发性肝癌”

B表示“所抽人甲胎球蛋白含量高”

于是，所求概率分别为$P(A|B), P(B|A)$

由题设知

$$P(A)=0.0004, P(B)=0.00034, P(AB)=0.00032$$

$$P(A|B)=\frac{P(AB)}{P(B)}=0.9412,\ P(B|A)=\frac{P(AB)}{P(A)}=0.8$$

■ 条件概率是概率

条件概率P(A|B)满足概率的三条公理，以及其他一切性质，例如

(1) $P(A|B)\ge 0$

(2) $P(\Omega|B)=1$

(3) $P(\sum_{i=1}^{\infty} A_i|B)=\sum_{i=1}^{\infty} P(A_i|B)$

(4) $P(\bar{A}|B)=1-P(A|B)$

(5) $P(A\cup B|C)=P(A|C)+P(B|C)-P(AB|C)$

注 意 点

$P(\Omega|B) = 1$;　　$P(B|\Omega) \neq 1$;

$P(A|\Omega) = P(A)$;　　$P(A|A) = 1$.

$P(A|\bar{B}) \neq P(\bar{A}|B)$　　$P(A|\bar{B}) \neq 1-P(A|B)$

一般总有P(A|B)≥P(AB)成立，但P(A|B)与P(A)不可比.

条件概率的三大公式

- 乘法公式;
- 全概率公式;
- 贝叶斯公式.

二. 乘法公式

(1)若 P(B)>0，则P(AB)=P(B)P(A|B)；

若 P(A)>0，则P(AB)=P(A)P(B|A).

(2) 若 $P(A_1A_2\cdots A_{n-1})>0$，则

$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)\cdots P(A_n|A_1A_2\cdots A_{n-1})$

注：n个事件的概率乘法公式并不只有上面这种形式。事实上，对于n个事件，这样形式的公式一定有n！个.

■ 乘法公式的应用：

乘法公式主要用于求几个事件同时发生的概率.

例. 一批零件共有100个，其中有10个不合格品。从中一个一个不返回取出，求第三次才取出不合格品的概率.

解：记 A_i=“第i次取出的是不合格品” B_i=“第i次取出的是合格品”，目的求 $P(B_1B_2A_3)$.

用乘法公式

$P(B_1B_2A_3)=P(B_1)P(B_2|B_1)\ P(A_3|B_1B_2)$

$$=\frac{90}{100}\times\frac{89}{99}\times\frac{10}{98}.$$

例2. 波利亚(Polya)坛子模型

坛子中有b只黑球及r只红球，随机取出一只，把原球放回，并加进与抽出球同色的球c只，再摸第二次，这样下去共摸了n次，问前面的n_1次出现黑球，后面的$n_2=n-n_1$次出现红球的概率是多少？

解：A_k表示第k次摸出黑球这一事件，需要计算交事件的概率 $P(A_1A_2\cdots A_{n_1}\bar{A}_{n_1+1}\cdots\bar{A}_n)$. 则

$$P(A_1)=\frac{b}{b+r},\qquad P(A_2\mid A_1)=\frac{b+c}{b+r+c},$$

$$P(A_3\mid A_1A_2)=\frac{b+2c}{b+r+2c},\ \cdots,$$

$$P(A_{n_1}\mid A_1A_2\cdots A_{n_1-1})=\frac{b+(n_1-1)c}{b+r+(n_1-1)c},$$

$$P(\bar{A}_{n_1+1}\mid A_1A_2\cdots A_{n_1})=\frac{r}{b+r+n_1c},$$

$$P(\bar{A}_{n_1+2}\mid A_1A_2\cdots A_{n_1}\bar{A}_{n_1+1})=\frac{r+c}{b+r+(n_1+1)c},\quad \cdots,$$

$$P(\bar{A}_n\mid A_1A_2\cdots A_{n_1}\bar{A}_{n_1+1}\cdots\bar{A}_{n-1})=\frac{r+(n_2-1)c}{b+r+(n-1)c}.$$

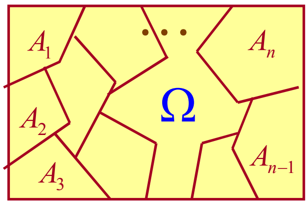
因此，
$$P(A_1A_2\cdots A_{n_1}\bar{A}_{n_1+1}\cdots\bar{A}_n)=\frac{b}{b+r}\cdot\frac{b+c}{b+r+c}\cdot\frac{b+2c}{b+r+2c}\cdots\frac{b+(n_1-1)c}{b+r+(n_1-1)c}\cdot\frac{r}{b+r+n_1c}\cdot\frac{r+c}{b+r+(n_1+1)c}\cdots\frac{r+(n_2-1)c}{b+r+(n-1)c}.$$

所求概率为
$$p=\frac{\prod_{k=0}^{n_1-1}(b+kc)\times\prod_{k=0}^{n_2-1}(r+kc)}{\prod_{k=0}^{n-1}(b+r+kc)}$$

这个模型曾被波利亚用来作为描述传染病的数学模型。这是一个很一般的模型，特别取 $c=0$，则是有放回摸球，$c=-1$ 则是不放回摸球。

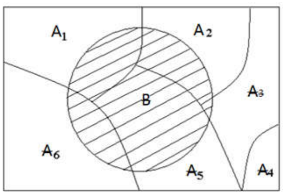
三. 全概率公式

若 $A_1,A_2,\cdots$ 两两互不相容，且 $\Omega=\bigcup_{i=1}^{\infty}A_i$，则称 $A_1,A_2,\cdots$ 为Ω的一个分割(划分)，亦称完备事件组。



◆ 全概率公式：若事件 A_1, A_2 ,… 是样本空间Ω的一组分割，且 $P(A_i)>0$，则 $B=\sum_{i=1}^{\infty}BA_i$,

$$P(B)=\sum_{i=1}^{\infty}P(BA_i)=\sum_{i=1}^{\infty}P(A_i)P(B\mid A_i).$$



Remark:

- 全概率公式用于求复杂事件的概率. 公式中的B是较复杂事件，A_i是引起B发生的各原因、情况或途径。
- 使用全概率公式关键在于寻找另一组事件来“分割”样本空间.
- 全空间可以由有限个A_i来分割，即A_1, A_2 , …, A_n
  $$P(B)=\sum_{i=1}^{n}P(A_iB)=\sum_{i=1}^{n}P(A_i)P(B\mid A_i)$$
- 全概率公式最简单的形式：
  $$P(B)=P(A)P(B\mid A)+P(\bar{A})P(B\mid\bar{A})$$

［例3］ 雨伞掉了. 落在图书馆中的概率为50%,这种情况下找回的概率为0.80;落在教室里的概率为30%,这种情况下找回的概率为0.60;落在商场的概率为20%,这种情况下找回的概率为0.05,求找回雨伞的概率.

［解］ 以 B 表示找回雨伞,而以 A_1,A_2,A_3 分别记雨伞落在图书馆,教室和商场,显然 A_1,A_2,A_3 满足

$$B=\sum_{i=1}^{3}A_iB$$

而且 $P(A_1)=0.5, P(A_2)=0.3, P(A_3)=0.2, P(B|A_1)=0.8, P(B|A_2)=0.6, P(B|A_3)=0.05$,因此

$$P(B)=\sum_{i=1}^{3}P(A_i)P(B|A_i)$$
$$=0.5\times0.8+0.3\times0.6+0.2\times0.05=0.59$$

例 设播种用一批小麦种子中混有一等、二等、三等、四等四个等级的种子，分别各占**95.5**%、**2**%、**1.5**%、**1**%，用一等、二等、三等、四等种子长出的穗含**50**颗以上麦粒的概率分别为**0.5**、**0.15**、**0.10**、**0.05**，求这批种子所结的穗含有**50**颗以上麦粒的概率。

解：设从这批种子中任选一颗是一等，二等，三等，四等种子的事件分别是 A_1,A_2,A_3,A_4，则它们构成完备事件组，又设B表示这批种子所结的穗含有**50**粒以上麦粒这一事件，于是，由题设条件有

$$P(A_1)=95.5\% \quad P(A_2)=2\% \quad P(A_3)=1.5\% \quad P(A_4)=1\%$$
$$P(B|A_1)=0.5 \quad P(B|A_2)=0.15$$
$$P(B|A_3)=0.10 \quad P(B|A_4)=0.05$$

则由全概率公式：

$$P(B)=\sum_{i=1}^{4}P(A_i)P(B|A_i)=0.4825$$

例 在你外出度假时，你请朋友给您的花浇水。如果没浇水的话，它死去的概率为0.8，如果浇水的话，它死去的概率为0.15。设朋友给花浇水的概率为0.9。求当您度假回来时，花还活着的概率是多少？若当您度假回来时，如果花死了，那么您的朋友忘记浇水的概率是多少？

解：设 **A**---朋友给花浇水，**B**--- 花活着. 由已知得

$$P(A)=0.9 \quad P(\bar{A})=0.1 \quad P(B|A)=0.85 \quad P(B|\bar{A})=0.2$$

则

$$P(B)=P(A)P(B|A)+P(\bar{A})P(B|\bar{A})=0.785$$

$$P(\bar{A}|\bar{B})=\frac{P(\bar{A})P(\bar{B}|\bar{A})}{P(\bar{A})P(\bar{B}|\bar{A})+P(A)P(\bar{B}|A)}$$
$$=\frac{0.1\times0.8}{0.1\times0.8+0.9\times0.15}\doteq0.3721$$

四. 贝叶斯公式

- 乘法公式是求“几个事件同时发生”的概率;
- 全概率公式是求“结果”的概率;
- 贝叶斯公式是已知“结果” ，求“原因”的概率.

◆ 贝叶斯(Bayes)公式：若事件 $A_1, A_2, \cdots$ 是样本空间Ω的一组分割，且 P(B)>0，则

用乘法公式

$$P(A_i|B)=\frac{P(A_iB)}{P(B)}=\frac{P(A_i)P(B|A_i)}{P(B)}$$
$$=\frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{\infty}P(A_j)P(B|A_j)}$$

用全概率公式

Bayes公式是英国统计学家Bayes于1763年首先提出的,是先验概率与后验概率转化工具。

经过多年的发展和完善，Bayes公式以及由此发展起来的一整套理论与方法，已经形成为概率统计中的贝叶斯统计。

Bayes公式的意义:

先验概率

当不知道某信息(事件B)时，我们对各事件A_1, A_2, …发生的可能性大小的认识为:$P(A_1)$, $P(A_2)$, …。

当知道某信息(事件B)已经发生时，我们对各事件A_1, A_2, … 发生的可能性大小要重新认识：$P(A_1|B)$, $P(A_2|B)$, …。

后验概率

例4 假定用血清甲胎球蛋白诊断肝癌：P(阳性|患者)= 0.95，P(阴性|健康者)=0.90；已知自然人群中 P(患者)=0.0004。现随机抽查一人，血清甲胎球蛋白诊断结果为阳性，求其患肝癌的概率有多大？

解：记A:诊断结果阳性，C:的确患有肝癌，则

$$P(C)=0.0004,\ P(A|C)=0.95,\ P(\bar{A}|\bar{C})=0.90$$

$$P(C\mid A)=\frac{P(C)P(A|C)}{P(C)P(A|C)+P(\bar{C})P(A|\bar{C})}$$

$$=\frac{0.0004\times 0.95}{0.0004\times 0.95+0.9996\times 0.1}=0.0038$$

Why?

过去的经验或知识 $P(C)$=0.0004 (先验概率) + 新信息 A ⇨ 修正过去认识 $P(C|A)$=0.0038 (后验概率)

后验概率小关键原因在于先验概率（人群中感染比例）非常小。

如果我们的检查对象是一个肝癌可疑人群, 比如甲胎球蛋白量高者, 其先验概率提高为例1中的**0.9412**, 则

$$P(C\mid A)=\frac{P(C)P(A|C)}{P(C)P(A|C)+P(\bar{C})P(A|\bar{C})}$$

$$=\frac{0.9412\times 0.95}{0.9412\times 0.95+0.0588\times 0.1}=0.9935$$

例5 若发报机以 0.7 和 0.3 的概率发出信号 0 和 1（譬如分别用低电平与高电平表示），由于随机干扰的影响，当发出信号 0 时，接收机不一定收到 0，而是以概率 0.8 和 0.2 收到信号 0 和 1；同样地，当发报机发出信号 1 时，接收机以概率 0.9 和 0.1 收到信号 1 和 0. 其关系如图 9 所示.

求:“当接收机收到信号 0 时，发报机是发出信号 0 的概率”.

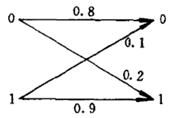


图 9

解：把发报机发出信号 0 记为事件 A_0，发出信号 1 记为事件 A_1，接收机接到信号 0 记为事件 B，我们要求的是 $P(A_0|B)$.

由于 $P(A_0)=0.7, P(A_1)=0.3, P(B|A_0)=0.8, P(B|A_1)=0.1$，用贝叶斯公式，

$$P(A_0|B)=\frac{P(A_0)P(B|A_0)}{P(A_0)P(B|A_0)+P(A_1)P(B|A_1)}$$

$$=\frac{0.7\times 0.8}{0.7\times 0.8+0.3\times 0.1}=\frac{0.56}{0.59}=0.949$$

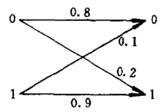


图 9

§ 2.2 事件独立性

一. 两个事件的独立性

二. 多个事件的独立性

三. 事件独立性与概率的计算

四. 试验的独立性

一. 两个事件的独立性

直观说法：对于两事件，若其中任何一个事件的发生不影响另一个事件发生的概率，则这两事件是独立的.

$$\Leftrightarrow P(A|B)=P(A)$$

$$\Leftrightarrow P(AB)/P(B)=P(A)$$

$$\Leftrightarrow P(AB)=P(A)P(B)$$

定义2.2.1 若两事件A与B满足

P(AB)=P(A)P(B)

则称A与B是统计独立的，简称A与B独立.

推论1 若事件A,B独立,且P(B)>0,则 P(A|B)=P(A).

推论2 若事件A与B独立，则A与$\bar{B}$独立、$\bar{A}$与B独立，$\bar{A}$与$\bar{B}$独立.

Remark:

① 必然事件Ω及不可能事件∅与任何事件都独立.

② “相互独立”与“互不相容”是无关的两个概念.

例1：有a只黑球，b只白球．每次随机从中取出一球，取后放回．求：

1. 在已知第一次摸得黑球的条件下，第二次摸出黑球的概率。
2. 第二次摸出黑球的概率。

解: 记 A={第一次取出黑球},B={ 第二次取出黑球 }. 则

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{a^2/(a+b)^2}{a/(a+b)}=\frac{a}{a+b}$$

$$P(B)=P(AB)+P(\bar{A}B)=\frac{a^2}{(a+b)^2}+\frac{ab}{(a+b)^2}=\frac{a}{a+b}$$

例2：有a只黑球，b只白球．每次随机从中取出一球，取后不放回．求：

1. 在已知第一次摸得黑球的条件下，第二次摸出黑球的概率。
2. 第二次摸出黑球的概率。

解: 记 A={第一次取出黑球},B={ 第二次取出黑球 }. 则

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{a(a-1)/[(a+b)(a+b-1)]}{a/(a+b)}=\frac{a-1}{a+b-1}$$

$$P(B)=P(AB)+P(\bar{A}B)=\frac{a(a-1)}{(a+b)(a+b-1)}+\frac{ab}{(a+b)(a+b-1)}$$

$$=\frac{a}{a+b}$$

二. 多个事件的独立性

定义2.2.2：对于三个事件A,B,C，若下列四个等式同时成立，则称它们相互独立。

$$\left.\begin{aligned}P(AB)&=P(A)P(B)\\P(BC)&=P(B)P(C)\\P(AC)&=P(A)P(C)\end{aligned}\right\}\quad(2.2.3)$$

$$P(ABC)=P(A)P(B)P(C)\quad(2.2.4)$$

若只满足前面三式，称A,B,C两两独立。

问题1：两两独立 ⇔ 相互独立?

［例 3］(伯恩斯坦反例)　一个均匀的正四面体,其第一面染成红色,第二面染成白色,第三面染成黑色,而第四面同时染上红,白,黑三种颜色. 现在以 A,B,C 分别记投一次四面体出现红,白,黑颜色朝下的事件,则由于在四面体中有两面有红色,因此

$$P(A)=\frac{1}{2}$$

同理 $P(B)=P(C)=\frac{1}{2}$,容易算出

$$P(AB)=P(BC)=P(AC)=\frac{1}{4}$$

所以(2.2.3)成立,即 A,B,C 两两独立,但是

$$P(ABC)=\frac{1}{4}\neq\frac{1}{8}=P(A)P(B)P(C)$$

因此(2.2.4)不成立,从而 A,B,C 不相互独立.

问题2：(2.2.4) ⇒ (2.2.3) ？

［例 4］　若有一个均匀正八面体,其第 1,2,3,4 面染红色,第 1,2,3,5 面染白色,第 1,6,7,8 面染上黑色,现在以 A,B,C 分别表示投一次正八面体出现红,白,黑的事件,则

$$P(A)=P(B)=P(C)=\frac{4}{8}=\frac{1}{2}$$

$$P(ABC)=\frac{1}{8}=P(A)P(B)P(C)$$

但是

$$P(AB)=\frac{3}{8}\neq\frac{1}{4}=P(A)P(B)$$

定义 2.2.3 对 n 个事件 $A_1, A_2, \cdots, A_n$，若对于所有可能的组合 $1 \leqslant i<j<k<\cdots \leqslant n$ 成立着

$$\left.\begin{aligned} &P(A_iA_j)=P(A_i)P(A_j)\\ &P(A_iA_jA_k)=P(A_i)P(A_j)P(A_k)\\ &\cdots\cdots\\ &P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n)\end{aligned}\right\} \qquad (2.2.5)$$

则称 $A_1, A_2, \cdots, A_n$ **相互独立**.

这里第一行有 $\binom{n}{2}$ 个式子，第二行有 $\binom{n}{3}$ 个式子等等，因此共应满足

$$\binom{n}{2}+\binom{n}{3}+\cdots+\binom{n}{n}=2^n-n-1$$

个等式. 由三个事件的场合可看出同时满足这些关系式是必须的.

推论: 设n个事件$A_1, A_2, \cdots, A_n$是相互独立的，则其中任意m个事件 $A_{i_1}, A_{i_2}, \cdots, A_{i_m}$也是相互独立的，其中，$1 \le m \le n$，而 $i_1, i_2 \ldots i_m$是1,2, …,n的一个选排列.

注：1. 对于对立事件也成立

2. 称无穷多个事件相互独立，如果其中任意有限多个事件都相互独立。

三. 事件独立性与概率的计算

[相互独立事件至少发生其一的概率的计算]

若n个事件A_1， A_2， …， A_n是相互独立的, 则由于

$$\overline{A_1 \cup A_2 \cup \cdots \cup A_n} = \overline{A_1}\,\overline{A_2}\cdots\overline{A_n},$$

因此，

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = 1 - P(\overline{A_1}\,\overline{A_2}\cdots\overline{A_n})$$
$$= 1 - P(\overline{A_1})P(\overline{A_2})\cdots P(\overline{A_n})$$

例5 假若每个人的血清中含有肝炎病毒的概率为0.4%，混合100个人的血清，求此血清中含有肝炎病毒的概率。

解：$P(A_1 \cup A_2 \cup \cdots \cup A_{100})$

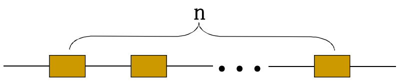
$= 1 - P(\overline{A_1}\,\overline{A_2}\cdots\overline{A_{100}}) = 1 - P(\overline{A_1})P(\overline{A_2})\cdots P(\overline{A_n})$

$= 1 - (1-0.004)^{100} = 1 - 0.996^{100} \approx 0.33$

[在可靠性理论中的应用]

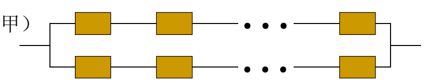
对于一个元件，它能正常工作的概率p，称为它的可靠性。元件组成系统，系统正常工作的概率称为该系统的可靠性。

例6：如果构成系统的每个元件的可靠性均为r, 0<r<1. 且各元件能否正常工作是相互独立的，试求下列系统的可靠性：



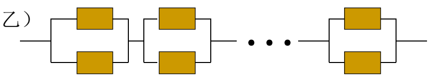
r^n

甲）



$r^n(2-r^n)$

乙）



$r^n(2-r)^n$

$P(G_1) = P(G_2) = r^n,$

$R_甲 = 1 - P(\overline{G_1}\,\overline{G_2}) = 1-(1-r^n)^2 = r^n(2-r^n),$

$or\ \ R_甲 = P(G_1 \cup G_2) = P(G_1)+P(G_2)-P(G_1G_2) = 2r^n - r^{2n},$

$R_乙 = (2r-r^2)(2r-r^2)\cdots(2r-r^2) = (2r-r^2)^n = r^n(2-r)^n.$

由于当 $n \ge 2$时，总有 $(2-r)^n > 2-r^n$

所以，$R_甲 < R_乙$，即乙系统比甲系统可靠。

四. 试验的独立性

所谓试验相互独立, 就是其中一试验所得到的结果, 对其它各试验取得其可能结果的概率都没有影响。

若试验E_1的任一结果与试验E_2的任一结果都是相互独立的事件，则称这两个试验相互独立，或称独立试验.

设试验E_1的样本空间是 $\Omega_1=\{\omega^{(1)}\}$，试验$E_2$的样本空间是$\Omega_2=\{\omega^{(2)}\}$, … E_n的样本空间是 $\Omega_n=\{\omega^{(n)}\}$, 为了描述这n次试验，应构造复合试验E，它表示依次进行试验$E_1, E_2, \cdots E_n$, 其样本点为

$$\omega=\{\omega^{(1)},\ \omega^{(2)},\ \cdots,\ \omega^{(n)}\}$$

这样的样本空间记作

$$\Omega=\Omega_1\times\Omega_2\times\cdots\times\Omega_n$$

例8. 若试验E_1是掷一枚硬币, Ω_1={正, 反}, 试验E_2是从装有红白黑三球的袋子中摸出一球, Ω_2={红, 白, 黑}, 则复合试验E表示先掷一枚硬币再摸一球，它相应的样本空间 $\Omega=\Omega_1\times\Omega_2$ 由下列6个样本点构成: (正, 红)，(正, 白)，(正, 黑)，(反, 红)，(反, 白)，(反, 黑)。

“<u>与第k次试验有关的事件</u>”：这种事件发生与否仅与第k次试验的结果有关。因此判断某一样本点是否属于这个事件，只需察看它的第k个分量。

定义2.2.4 以$\mathcal{A}_k$记为与第k次实验有关的事件全体。若对于任意的

$$A^{(1)}\in\mathcal{A}_1,\ A^{(2)}\in\mathcal{A}_2,\ \ldots,\ A^{(k)}\in\mathcal{A}_n$$

均成立

$$P(A^{(1)}A^{(2)}\cdots A^{(n)})=P(A^{(1)})P(A^{(2)})\cdots P(A^{(n)})$$

则称试验$E_1, E_2, \ldots, E_n$是**相互独立的**.

思考: 例8中, 试验E_1与试验E_2是否相互独立?
答：由古典概率的计算，可验证其独立.

是否还可以构造其它相互独立的试验?

例如:

1. n次有放回摸球所构成的n个试验是相互独立的;
2. n次不放回摸球所构成的n个试验不独立。

重复独立试验:

研究“在同样条件下重复试验”的数学模型，其满足：

1. $\Omega_1=\Omega_2=\cdots=\Omega_n$ ；
2. 有关事件的概率保持不变；
3. 各次试验是相互独立的。

例如：投n个硬币或进行n次有放回摸球。

§ 2.3 伯努利试验与直线上的随机游动

一 伯努利概型

二 伯努利概型中的一些分布

三 直线上的随机游动

四 推广的伯努利试验和多项分布

一. 伯努利(Bernoulli)概型

在许多问题中，人们往往关心实验中某一事件A是否发生。例如:

1. 在产品质量抽样检测中是否抽到次品；
2. 在掷硬币试验中是否出现正面；
3. 在股票市场中股票是涨还是跌等。

在这类问题中,我们可把事件域取为

$$\mathcal{F}=\{\varnothing, A, \bar{A}, \Omega\},$$

并称试验出现事件A为“成功”,反之称为“失败”.这种只有两个结果的试验为伯努利(Bernoulli)试验.

具体地，如果随机试验E只有两个结果：A和$\bar{A}$,其中,P(A)=p，P($\bar{A}$)=q，(p+q=1,p>0,q>0)，则称E为伯努利试验.

- n重伯努利试验: 独立重复的进行n次伯努利试验,记作E^n,其特点是:

① 每次试验只有两个可能结果，A和$\bar{A}$;

② A在每次试验中出现的概率p保持不变;

③ 各次试验相互独立;

④ 共进行了n次试验.

n 重伯努利试验的样本点$w=(w_1,w_2,\ldots,w_n)$,

w_i =A 或 $\bar{A}$, 表示第i次试验A是否发生。

n重独立重复的伯努里试验共有2^n个样本点.

样本点$(A_1,A_2,\ldots,A_{n-1},\bar{A}_n)$,可简记做$A_1A_2\ldots A_{n-1}\bar{A}_n$,其出现的概率为:

$$P(A_1A_2\ldots A_{n-1}\bar{A}_n)=pp\ldots pq=p^{n-1}q.$$

伯努利试验是一种非常重要的概率模型，它是“在同样条件下进行重复试验”的一种数学模型，特别在讨论某事件出现的频率时常用这种模型。

在历史上，伯努利概型是概率论中最早研究的模型之一，也是得到最多研究的模型之一，在理论上具有重要的意义。

另一方面,它有着广泛的应用，在我们这门课程中,一些较为深入的结果也是结合伯努利概型进行讨论的。

◆ 独立重复伯努利试验中的三个重要问题：

(1) n 次试验中A 恰好发生 k 次的概率是多少？

(2) 到第 k 次试验 A才首次发生的概率是多少？

(3) 一直不停试验，A最终发生的概率是多少？

二. 伯努利概型中的一些分布

1. 伯努利分布

- 只进行一次伯努利试验,
- 概率: $P(A)=p$, $P(\bar{A})=q$, $(p+q=1, p>0, q>0)$,
- 这种概率分布称为伯努利分布(或两点分布).
- 伯努利概型中最简单的情形.

2.二项(binomial)分布

在n重Bernoulli试验中,事件A恰好发生k次的概率为

$$P_n(k)=C_n^k p^k q^{n-k}, (q=1-p)$$
$$k=0, 1, 2, \cdots, n.$$

分析: 在n次试验中，考虑某k次A发生，另外$(n-k)$次A不发生，例如前k次A发生，后$(n-k)$次A不发生的概率为:

$P(A_1A_2\cdots A_k\bar{A}_{k+1}\bar{A}_{k+2}\cdots\bar{A}_n)=p^kq^{n-k}$,

而A发生k次总共有C_n^k种选择方法，即得.

为方便, 记n重伯努利试验中A出现k次的概率为$b(k; n, p)$, 即

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

称形如上式的概率分布为二项分布, 且有

$$\sum_{k=0}^{n} b(k;n,p)=\sum_{k=0}^{n} C_n^k p^k q^{n-k}=(p+q)^n=1.$$

例1 设一批产品中有a件是次品, b件是正品.现有放回地从中抽取了n件产品. 求: 事件A={n件产品中恰有k件次品}的概率 (k=0, 1, 2, …, n).

解: 视为n重伯努利试验, $p=a/(a+b)$, 从而有:

$$b(k;n,p)=C_n^k p^k q^{n-k}=C_n^k\left(\frac{a}{a+b}\right)^k\left(\frac{b}{a+b}\right)^{n-k}.$$

3. 几何(Geometric)分布

考虑伯努利试验中首次成功出现在第k次的概率, 有

$$P(\bar{A}_1\bar{A}_2\ldots\bar{A}_{k-1}A_k)=q^{k-1}p,$$

为方便，记$g(k;p)=q^{k-1}p, k=1,2,\ldots$

$g(k;p)$为几何级数的一般项，故称为几何分布。

另外，我们有：

$$\sum_{k=1}^{\infty} g(k;p)=\sum_{k=1}^{\infty} q^{k-1}p=p\frac{1}{1-q}=1.$$

注：伯努利试验中首次成功之前失败次数为l次的概率为

$$g^*(l;p)=q^l p, l=0,1,2,\ldots$$

例3. 一个人要开门, 共有n把钥匙, 其中仅有一把钥匙开门, 每次试开后放回，则这人在第s次试开时才首次成功的概率是多少?

分析：n重伯努利试验，$p=1/n$,

第s次首次成功的概率：

$g(s;1/n)=[(n-1)/n]^{s-1}\cdot 1/n.$

4. 帕斯卡(Pascal)分布

- 记C_k={第r次成功发生在第k次},
- 记$f(k;r,p)=P(C_k)$,
- C_k={前k-1次成功r-1次,且第k次成功}, 则

$$P(C_k)=C_{k-1}^{r-1}p^{r-1}q^{k-r}\cdot p=C_{k-1}^{r-1}p^r q^{k-r}$$
$$f(k;r,p)=C_{k-1}^{r-1}p^r q^{k-r},k=r,r+1,\cdots$$

称$f(k;r,p)$为帕斯卡分布。当r =1时，为几何分布。

令$l=k-r$, 即第r次成功之前失败的次数，则有

$$f(k;r,p)=\binom{k-1}{r-1}p^r q^{k-r}=\binom{r+l-1}{l}p^r q^l=\binom{-r}{l}(-1)^l p^r q^l,$$
$$\sum_{k=r}^{\infty}f(k;r,p)=\sum_{l=0}^{\infty}\binom{-r}{l}(-1)^l p^r q^l=p^r(1-q)^{-r}=1.$$

注:伯努利试验中第r次成功之前失败次数为l次的概率为:

$$f^*(l;r,p)=\binom{l+r-1}{l}p^r q^l=\binom{-r}{l}(-1)^l p^r q^l,l=0,1,2,\cdots$$

◆ 分赌注问题: 帕斯卡分布的起源

甲、乙两赌徒按某种方式下注赌博，先胜t局者将赢得全部赌注，但进行到甲胜r局、乙胜s局($r<t$, $s<t$)时，因故不得不中止,试问如何分配这些赌注才公平合理？

分配建议：

1. 用r : s来分配
2. 用甲乙最终取胜的概率$P_甲$：$P_乙$来分配.

分析：

- 甲若想获胜，需要再胜n=t-r局;
- 乙若想获胜，需要再胜m=t-s局;
- 对每一局，记A={甲胜}, P(A)=p，P(A^c)=q=1-p;
- 甲若想获胜，当甲再胜n局时，乙再胜的局数k<m局，即A的第n次出现发生在第n+k次(k<m)试验，则

$$P_甲=\sum_{k=0}^{m-1}C_{n+k-1}^{n-1}p^n q^k$$

或者当乙再胜m局时，甲再胜的局数k>=n局，即A^c的第m次出现发生在第m+k次(k≥n)试验，则

$$P_甲=\sum_{k=n}^{\infty}C_{m+k-1}^{k}p^k q^m$$

易证: 再赌n+m-1局可以决定胜负

甲若想获胜,必须在n+m-1局中至少胜n次，根据二项分布:

$$P_甲=\sum_{k=n}^{n+m-1}C_{n+m-1}^{k}p^k q^{n-m-1-k}$$

可以证明上述三个答案是一致的。

下面是与帕斯卡分布有关的另一个有名例子.

［例 4］（巴拿赫火柴盒问题）数学家的左、右衣袋中各放有一盒装有 N 根火柴的火柴盒，每次抽烟时任取一盒用一根，求发现一盒用光时，另一盒有 r 根的概率.

［解］ 看作 $p=\frac{1}{2}$ 的伯努利试验. 要左边空而右边剩 r 根，应该是左边摸过 $N+1$ 次（前 N 次用去 N 根火柴，最后一次发觉火柴盒是空的），而右边摸过 $N-r$ 次，这事件的概率为

$$f\left(2N-r+1;N+1,\frac{1}{2}\right)=\binom{2N-r}{N}\left(\frac{1}{2}\right)^{2N-r+1}$$

对于右边先空的情况可同样考虑，因此所求的概率为

$$u_r=2\cdot f\left(2N-r+1;N+1,\frac{1}{2}\right)=\binom{2N-r}{N}2^{-2N+r}$$

◆ 小概率事件必然发生

假定随机事件A在一次试验中发生概率是 p（无论多小、只需 $p>0$），如果不停地独立重复进行试验，那么 A 最终必然要发生。

证明. 记 $D_k=\{A$在第k次试验时发生，$k\geq 1\}$，由于是独立试验，这些D_k相互独立，因此它们的对立事件 $\{A$在第 k 次试验时没有发生$\}$ 也相互独立。我们只需要证明：

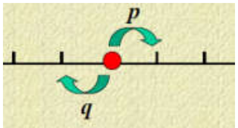
$$P\left(\bigcup_{k=1}^{\infty} D_k\right)=1$$

$$P\left(\bigcup_{k=1}^{\infty} D_k\right)=1-P\left(\bigcap_{k=1}^{\infty} \overline{D_k}\right)=1-\prod_{k=1}^{\infty}(1-p)$$

因为A每次试验时发生的概率 $p>0$，即 $1-p<1$，所以当不停独立重复试验时无论每次发生的概率有多小，这个随机事件最终必然会发生。

三. 直线上的随机游动

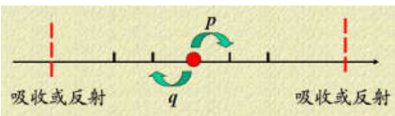
考虑x轴上的一个质点，在时刻$t=0$时，它处于初始位置 a (a是整数)，以后每隔单位时间，分别以概率p及概率$q=1-p$向正的或负的方向移动一个单位。用这种方式描述的质点的运动称为随机游动。



当p=q=1/2时，随机游动称为对称的，这时质点向左或向右移动的可能性相等。

若质点可以在整个数轴的整数点上游动，则称这种随机游动为无限制随机游动。

若在d点设有一个吸收壁，质点一到达这点即被吸收而不再游动，因而整个游动就结束，这种随机游动称为在d点有吸收壁的随机游动。



在随机游动模型中，我们所关心的是质点在时刻$t=n$ 时的位置。

1. 无限制随机游动：有无穷赌本的赌徒在n局后的输赢

假定质点在时刻0从原点出发，以S_n记它在时刻n的位置。

为了使质点在时刻 $t=n$ 时位于k(k可以是负整数，$-n\leq k\leq n$)，必须且只须在前n次游动中向右移动的次数比向左移动的次数多k 次。

若以 x 记它在n 次游动中向右移动的次数，y 记向左移动的次数，则

$$\begin{cases} x+y=n \\ x-y=k \end{cases} \Rightarrow x=\frac{n+k}{2},\ y=\frac{n-k}{2}.$$

因为 x是整数，所以 k 必须与 n 具有相同的奇偶性。

事件$\{S_n=k\}$ (赌徒在n局后赢\$$k$) 发生相当于要求在前$n$ 次游动中有$x=(n+k)/2$次向右，$y=(n-k)/2$ 次向左，利用二项分布可得

$$P\{S_n=k\}=\binom{n}{\frac{n+k}{2}}q^{\frac{n-k}{2}}p^{\frac{n+k}{2}}.$$

当k与n奇偶性相反时，$P\{S_n=k\}=0$.

◆ 赌徒在n局后赢的概率:

(1) n为偶数($n=2m$)时,

$$\sum_{k=1}^{n}P\{S_n=k\}=\sum_{i=1}^{m}P\{S_n=2i\}=\sum_{i=1}^{m}\binom{2m}{m+i}q^{m-i}p^{m+i}=\sum_{l=m+1}^{2m}b(l;2m,p)$$

(2) 当n为奇数($n=2m+1$)时,

$$\sum_{k=1}^{n}P\{S_n=k\}=\sum_{i=0}^{m}P\{S_n=2i+1\}$$

$$=\sum_{i=0}^{m}\binom{2m+1}{m+i+1}q^{m-i}p^{m+i+1}=\sum_{l=m+1}^{2m+1}b(l;2m+1,p)$$

$$\therefore\sum_{k=1}^{n}P\{S_n=k\}=\sum_{l=[\frac{n}{2}]+1}^{n}b(l;n,p).$$

特别地，当$p=q=0.5$时，$\sum_{k=1}^{n}P\{S_n=k\}=0.5$.

2. 两端带有吸收壁的随机游动:有穷赌本赌徒的输赢

$x=0$ 与 $x=a+b$ 为两端吸收点的随机游动

记 $q_0(n)=P\{$质点从$x=n$出发被0点吸收$\}$,
$p_{a+b}(n)=P\{$质点从$x=n$出发被$a+b$点吸收$\}$

q-递归关系: $q_0(n)=pq_0(n+1)+qq_0(n-1)$
$1\leq n\leq a+b-1$, 初值 $q_0(0)=1$, $q_0(a+b)=0$

p-递归关系: $p_{a+b}(n)=pp_{a+b}(n+1)+qp_{a+b}(n-1)$
$1\leq n\leq a+b-1$, 初值$p_{a+b}(0)=0$, $p_{a+b}(a+b)=1$

齐次二阶差分方程

$af(n+2)+bf(n+1)+cf(n)=0$, 两个初值

$\Rightarrow$ 解出$ax^2+bx+c=0$ 的两个根 x_1,x_2

(1) 当$x_1=x_2$, 则$f(n)=(c_1+c_2n)x_1^n$

(2) 当$x_1\neq x_2$, 则$f(n)=c_1x_1^n+c_2x_2^n$

根据差分方程初值解出待定系数c_1、c_2

方程 $px^2-x+q=0$ 的根显然为 $x=\dfrac{1\pm|p-q|}{2p}$,

(1). 如果 $p=q=0.5$, 重根 $x_1=x_2=1$,
递归关系的通解为 $f(n)=(c_1+c_2n)$

q-递归初值 $f(0)=1, f(a+b)=0$ 解出
$c_1=1, c_2=-1/(a+b)$ 即 $q_0(n)=\dfrac{a+b-n}{a+b}$;

p-递归初值 $f(0)=0, f(a+b)=1$ 解出
$c_1=0, c_2=1/(a+b)$ 即 $p_{a+b}(n)=\dfrac{n}{a+b}$;

甲有赌本a,乙有赌本b, 公平赌博中甲输光概率为$b/(a+b)$, 甲赢光乙的概率$a/(a+b)$; 必居其一。

(2). 如果 $p\neq q$, $x_1=1$, $x_2=q/p$, 记 $r=q/p$;
递归关系的通解为 $f(n)=(c_1+c_2r^n)$

q-递归初值 $f(0)=1, f(a+b)=0$ $\begin{cases}c_1+c_2=1\\ c_1+c_2r^{a+b}=0\end{cases}$

$c_1=\dfrac{-r^{a+b}}{1-r^{a+b}}, c_2=\dfrac{1}{1-r^{a+b}}$; $q_0(n)=\dfrac{r^n-r^{a+b}}{1-r^{a+b}}$

p-递归初值 $f(0)=0, f(a+b)=1$ $\begin{cases}c_1+c_2=0\\ c_1+c_2r^{a+b}=1\end{cases}$

$c_1=\dfrac{1}{1-r^{a+b}}, c_2=\dfrac{-1}{1-r^{a+b}}$; $p_{a+b}(n)=\dfrac{1-r^n}{1-r^{a+b}}$

甲有赌本a,乙有赌本b, 不公平赌博中甲输光概率为$q_0(a)$, 甲赢光乙的概率$p_{a+b}(a)$; 必居其一。

四. 推广的伯努利试验与多项分布

二项分布可以容易地推广到 n 次重复独立试验且每次试验可能有若干个结果的情形。

把每次试验的可能结果记为 $A_1, A_2, \cdots A_r$ ，而

$P(A_i) = p_i,\ i = 1, 2, \cdots, r,$ 且 $p_1 + p_2 + \cdots p_r = 1,\ p_i \ge 0$

当 $r = 2$ 时，我们得到伯努利试验。

不难导出，在 n 次试验中 A_1 出现 k_1 次，A_2 出现 k_2 次，A_r 出现 k_r 次的概率为

$$\frac{n!}{k_1!k_2!\cdots k_r!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r} \qquad (5)$$

这里 $k_i \ge 0$ ，且 $k_1 + k_2 + \cdots + k_r = n$ 。

公式（5）称为多项分布，它是二项分布的推广，二项分布中的很多结果都能平行地推广到多项分布的场合。

例5 平面上的随机游动

一质点从平面上某点出发，等可能的向上、下、左及右方向移动，每次移动的距离为1，求经过2n次移动后回到出发点的概率。

解：这可以归结为上述推广的伯努利试验的问题。

分别以事件 A_1, A_2, A_3, A_4 表示质点向上，下，左，右移动一格，则 $p_1 = p_2 = p_3 = p_4 = 1/4$

若要在2n次移动后回到原来的出发点，则向左移动的次数与向右移动的次数应该相等，向上移动的次数与向下移动的次数也应该相等。

而总移动次数为2n，故所求的概率为

$$P = \sum_{k+m=n} \frac{(2n)!}{(k!)^2(m!)^2}\left(\frac{1}{4}\right)^{2n} = \sum_{k=0}^{n} \frac{(2n)!}{(k!)^2[(n-k)!]^2}\left(\frac{1}{4}\right)^{2n}$$

$$= \left(\frac{1}{4}\right)^{2n} \frac{(2n)!}{(n!)^2} \sum_{k=0}^{n} \left[\frac{n!}{k!(n-k)!}\right]^2$$

$$= \left(\frac{1}{4}\right)^{2n} \binom{2n}{n} \sum_{k=0}^{n} \binom{n}{k}^2 = \left(\frac{1}{4}\right)^{2n} \binom{2n}{n}^2 .$$

第四节 二项分布与泊松分布

1. 二项分布
2. 二项分布的泊松逼近
3. 泊松分布

一、二项分布

在 n 重Bernoulli试验中，事件A恰好发生 k 次的概率为

$$b(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \cdots, n.$$

称形如上式的概率分布为二项分布.

◆二项分布的计算：

Excel函数：BINOMDIST(k, n, p, 0/1).

BINOMDIST(k, n, p, 0) = $C_n^k p^k (1-p)^{n-k}$;

BINOMDIST(k, n, p, 1) = $\sum_{i=0}^{k} C_n^i p^i (1-p)^{n-i}$.

表 2.4.1 二项分布数值表

k	$b(k;20,p)$			k	$b(k;20,p)$		
	$p_1=0.1$	$p_2=0.3$	$p_3=0.5$		$p_1=0.1$	$p_2=0.3$	$p_3=0.5$
0	0.121 6	0.000 8	—	11	—	0.012 0	0.160 2
1	0.270 2	0.006 8	—	12	—	0.003 9	0.120 1
2	0.285 2	0.027 8	0.000 2	13	—	0.001 0	0.073 9
3	0.190 1	0.071 6	0.001 1	14	—	0.000 2	0.037 0
4	0.089 8	0.130 4	0.004 6	15	—	—	0.014 8
5	0.031 9	0.178 9	0.014 8	16	—	—	0.004 6
6	0.008 9	0.191 6	0.037 0	17	—	—	0.001 1
7	0.002 0	0.164 3	0.073 9	18	—	—	0.000 2
8	0.000 4	0.114 4	0.120 1	19	—	—	—
9	0.000 1	0.065 4	0.160 2	20	—	—	—
10	—	0.030 8	0.176 2				

◆ $b(k;n,p) = b(n-k; n, 1-p)$.

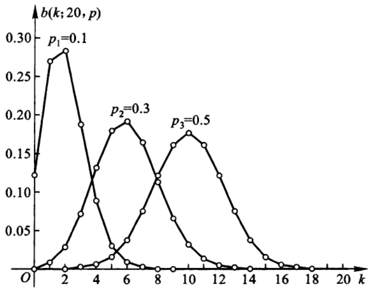


图 2.4.1　二项分布图

［例 2］（血清的试验）设在家畜中感染某种疾病的概率是 30%，新发现了一种血清可能对预防此病有效，为此对 20 只健康的动物注射这种血清. 若注射后只有一只动物受感染，我们应对此种血清的作用作何评价？

假如血清毫无价值，那么注射后的动物受感染的概率还是 30%，则这 20 只动物中有 k 只受感染的概率为 $b(k;20,0.3)$.

发生只有一只动物受感染或更好的情况（无动物受感染）的概率为

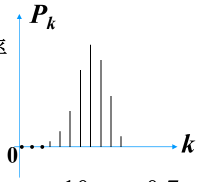
$$b(0;20,0.3)+b(1;20,0.3)=0.000\,8+0.006\,8=0.007\,6$$

这个概率如此之小，因此我们不能认为血清毫无价值.

如果注射后的 20 只动物中有 4 只受感染，我们是否相信此种血清有效？

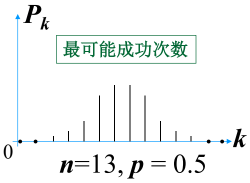
◆二项分布的性质：

对于固定 n 及 p，当 k 增加时，概率 $b(k;n,p)$ 先是随之增加直至达到最大值，随后单调减少.



n=10, p = 0.7

当 $(n+1)p$ 不为整数时，$b(k;n,p)$ 在 $k=[(n+1)p]$ 达到最大值；



n=13, p = 0.5

当 $(n+1)p=m$ 为整数时，$b(k;n,p)$ 在 $k=m$ 和 $k=m-1$ 处达到最大值.

▪ 证明：

由于对 $0<p<1$，

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

因此

当 $k<(n+1)p$ 时，$b(k;n,p)>b(k-1;n,p)$

当 $k=(n+1)p$ 时，$b(k;n,p)=b(k-1;n,p)$

当 $k>(n+1)p$ 时，$b(k;n,p)<b(k-1;n,p)$

因为 $(n+1)p$ 不一定是整数，而二项分布中的 k 只取整数值，所以存在整数 m，使得 $(n+1)p-1<m\leqslant(n+1)p$，而且当 k 从 0 变到 n 时，$b(k;n,p)$ 起先单调上升，当 $k=m$ 时达到极大值，后来又单调下降. 但若 $(n+1)p=m$，则这时 $b(m;n,p)=b(m-1;n,p)$ 同时达到极大值.

［例 3］　设某种疾病的发病率为 0.01，问在 500 人的社区中进行普查最可能的发病人数是多少？并求其相应的概率.

［解］　这是伯努利概型，发病人数服从二项分布. $n=500$，$p=0.01$，$(n+1)p=5.01$，$[(n+1)p]=5$. 所以最可能发病人数为 5. 相应概率为

$$b(5;500,0.01)=\binom{500}{5}(0.01)^5\cdot(0.99)^{495}=0.176\,35 \qquad (2.4.3)$$

［例 4］（人寿保险）　保险业是最早使用概率论的部门之一. 保险公司为了决定保险金数额，估算公司的利润和破产的风险，需要计算各种各样的概率. 下面是典型问题之一. 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 10 000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 个的概率.

[解] 作为初步近似，可以利用伯努利概型，$n=10\ 000$，$p=0.005$，设 μ 为未来一年中这些人里面死亡的人数，则所求的概率分别为

(1) $b(40;10\ 000,0.005)$

$$=\binom{10\ 000}{40}(0.005)^{40}(0.995)^{9\ 960} \qquad (2.4.6)$$

(2) $P\{\mu\leqslant 70\}=\sum_{k=0}^{70}b(k;10\ 000,0.005)$

$$=\sum_{k=0}^{70}\binom{10\ 000}{k}(0.005)^{k}(0.995)^{10\ 000-k} \qquad (2.4.7)$$

直接计算这些数值相当困难，要有更好的计算方法.

二、二项分布的泊松(Poisson)逼近

◆ 泊松分布：概率分布

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

称为泊松分布，λ>0 称为它的参数.

定理 2.4.1(泊松) 在独立试验中，以 p_n 代表事件 A 在试验中出现的概率，它与试验总数 n 有关，如果 $np_n\to\lambda$，则当 $n\to\infty$ 时，

$$b(k;n,p_n)\to\frac{\lambda^k}{k!}e^{-\lambda}$$

[证明] 记 $\lambda_n=np_n$，则

$$b(k;n,p_n)=\binom{n}{k}p_n^k(1-p_n)^{n-k}$$

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

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

由于对固定的 k 有

$$\lim_{n\to\infty}\lambda_n^k=\lambda^k,\quad \lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^{n-k}=e^{-\lambda}$$

及

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

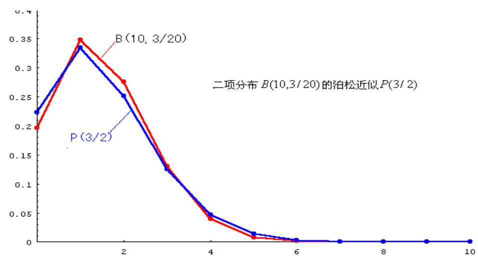
因此

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

定理证毕.

泊松定理表明：泊松分布是二项分布的极限分布.

当 n 很大，p 很小时，二项分布就可近似看成是参数 $\lambda=np$ 的泊松分布，即：$b(k;n,p)\approx\frac{(np)^k}{k!}e^{-np}$.



用二项分布的泊松近似计算前面的例3：

$$b(5;500,0.01)=\binom{500}{5}(0.01)^5\cdot(0.99)^{495}=0.176\ 35$$

$$b(5;500,0.01)\approx\frac{5^5}{5!}e^{-5}=0.17547.$$

例4.(1) $b(40;10000,0.005)\approx\frac{50^{40}}{40!}e^{-50}=0.0215.$

◆ 泊松分布的计算：

Excel函数：POISSON(k,λ,0/1).

POISSON(k,λ,0)$=\frac{\lambda^k}{k!}e^{-\lambda}$;

POISSON(k,λ,1)$=\sum_{i=0}^{k}\frac{\lambda^i}{i!}e^{-\lambda}$.

三、泊松分布

近几十年来，泊松分布日益显示其重要性，成为概率论中最重要的几个分布之一。泊松分布在管理科学、运筹学以及自然科学的某些问题中都占有重要的地位。

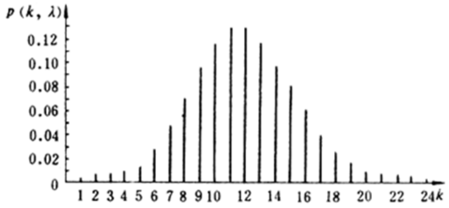
①在大量重复试验中稀有事件出现的次数近似服从泊松分布，如意外事故，非常见病，大的自然灾害.

②排队问题：在一段时间内窗口等待服务的顾客数，电话交换台的呼叫数，网站访问数，公共汽车站来到的乘客数.

③放射源衰变产生的粒子数.

图形特点:

$$p(k;\lambda)=\frac{\lambda^{k}}{k!}e^{-\lambda},k=0,1,2,\cdots.$$



泊松分布$P(\lambda)$ (λ=12)

下面提供两个有关的统计资料作为例子.

[例9] 对上海市某公共汽车站的客流进行调查,统计了某天上午10:30至11:47左右每隔20秒钟来到的乘客批数(每批可能有数人同时来到),共得230个记录,分别计算了来到0批,1批,2批,3批,4批及4批以上乘客的时间区间的频数,结果列于表2.4.2中,其相应的频率与$\lambda=0.87$的泊松分布符合得很好.

表 2.4.2 公共汽车客流统计

来到批数 i	0	1	2	3	≥4	总 共
频数 n_i	100	81	34	9	6	230
频率 $f_i=\frac{n_i}{n}$	0.43	0.35	0.15	0.04	0.03	
$p_i=\frac{\lambda^i}{i!}e^{-\lambda}$	0.42	0.36	0.16	0.05	0.01	

[例10] 放射性物质放射出的α质点数是服从泊松分布的有名例子. 1910年Rutherford等人的著名实验揭露了这个事实.

在这个实验中,观察了长为7.5秒的时间间隔里到达某指定区域的质点数,共观察了$N=2\ 608$次,表2.4.3给出观察值与理论值的对照,N_k表示在N次观察中发生"在7.5秒内落到指定区域的质点数为k"的观察次数,理论值是$Np(k;3.870)$,理论值与实验值很近似.

表 2.4.3 Rutherford 实验理论值与实验值对照表①

k	N_k	$Np(k;3.870)$
0	57	54.399
1	203	210.523
2	383	407.361
3	525	525.496
4	532	508.418
5	408	393.515
6	273	253.817
7	139	140.325
8	45	67.882
9	27	29.189
$k\geqslant 10$	16	17.075
总 计	2 608	2 608.000

◆ 产生泊松分布的机制分析

我们先证明一个以后要屡次用到的数学分析结论.

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

$$f(x)f(y)=f(x+y) \tag{2.4.13}$$

则

$$f(x)=a^{x} \tag{2.4.14}$$

其中$a\geqslant 0$,是某一常数.

[证明] 由(2.4.13)知对任意x,

$$f(x)=\left[f\left(\frac{x}{2}\right)\right]^{2}\geqslant 0$$

因此$f(x)$非负. 反复使用(2.4.13),对任意正整数n及实数x有

$$f(nx)=[f(x)]^{n}$$

在上式中取$x=\frac{1}{n}$,得

$$f(1)=\left[f\left(\frac{1}{n}\right)\right]^{n}$$

记$a=f(1)\geqslant 0$,则

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

因此,对任意正整数m及n,成立

$$f\left(\frac{m}{n}\right)=\left[f\left(\frac{1}{n}\right)\right]^{m}=a^{\frac{m}{n}}$$

这样,我们已证得(2.4.14)对一切有理数成立,再利用连续性或单调性可以证明对无理数也成立,从而证明了引理.

◆ 泊松过程：考虑来到某交换装置的电话呼叫数，假定它具有下面三个性质：

（i）平稳性：在$[t_0, t_0+t)$中来到的呼叫数只与时间间隔长度 t 有关而与时间的起点t_0无关。

（ii）独立增量性(无后效性)：在$[t_0, t_0+t)$中来到k次呼叫与时刻t_0以前发生的事件独立。

(iii) 普通性：在充分小的时间间隔内，最多来到一个呼叫。

若以 $N(t)$ 记在长度为t的时间区间中来到的呼叫数，而 $P_k(t)=P(N(t)=k)$，则显然有：$\sum_{k=0}^{\infty} P_k(t)=1$.

若记$\varphi(t)=\sum_{k=2}^{\infty} P_k(t)=1-P_0(t)-P_1(t)$，则普通性意味着：

$\varphi(t)=o(t)$，即$\lim\limits_{t\to\infty}\dfrac{\varphi(t)}{t}=0$.

➢ 平稳性表示了它的概率规律不随时间的推移而改变.

➢ 独立增量性表明互不相交的时间区间内过程进行的相互独立性。

➢ 普通性表明，在同一时间瞬间来到两个或两个以上呼叫实际上是不可能的。

一般地，称满足上述三个条件的随机过程$\{N(t), t\ge 0\}$为泊松过程，

且可以证明(见下)：$P_k(t)=P[N(t)=k]=\dfrac{(\lambda t)^k}{k!}e^{-\lambda t}, k=0,1,2,\cdots$.

下面我们来求 $P_k(t)$.

对$\Delta t>0$，考虑 $[0, t+\Delta t)$中发生k次事件概率$P_k(t+\Delta t)$，由独立增量性及全概率公式

$$P_k(t+\Delta t)=P_k(t)P_0(\Delta t)+P_{k-1}(t)P_1(\Delta t)+\cdots+P_0(t)P_k(\Delta t), (6)$$

（$k\ge 0$，对 $n\ge 1$，假定 $P_{-n}(t)=0$.）

特别地，$P_0(t+\Delta t)=P_0(t)P_0(\Delta t)$.

$P_0(t)$表示在长度为t 的时间间隔中没有来呼叫的概率，因此它关于t 单调下降，由前面的引理知：

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

其中，$0\le a\le 1$，但$a=0$(or 1)不在我们的考虑之列。所以，应有：$0<a<1$.

从而存在$\lambda(=-\ln a)>0$，使

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

因此，当 $\Delta t\to 0$ 时，我们有

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

$$P_1(\Delta t)=1-P_0(\Delta t)-P[N(\Delta t)\ge 2]=\lambda\Delta t+o(\Delta t)$$

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

故由(6)式得：

$$P_k(t+\Delta t)=P_k(t)\cdot(1-\lambda\Delta t)+P_{k-1}(t)\cdot\lambda\Delta t+o(\Delta t)$$

$$P_k(t+\Delta t)=P_k(t)-P_k(t)\cdot\lambda\Delta t+P_{k-1}(t)\cdot\lambda\Delta t+o(\Delta t)$$

因此，

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

令$\Delta t\to 0$，得　$P_k'(t)=\lambda[P_{k-1}(t)-P_k(t)],\quad k\ge 1$

$$[e^{\lambda t}P_k(t)]'=\lambda e^{\lambda t}P_{k-1}(t),\quad k\ge 1$$

由于已知 $P_0(t)=e^{-\lambda t}$，故有 $[e^{\lambda t}P_1(t)]'=\lambda e^{\lambda t}P_0(t)=\lambda$，$P_1(t)=\lambda te^{-\lambda t}$，　这样下去，可解得一切$P_k(t)$.

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

这正是参数为λt的泊松分布.　证毕.

例2.4.2　假定服务器在长度 t 分钟的时间内被攻击的次数近似服从 $P(2t)$，问 3 分钟内至少被攻击一次与 5 分钟内至少被两次攻击哪一个更可能出现？

解．3 分钟内被攻击次数服从参数6的Poisson分布，因此 $1-p(0;6)\approx 0.997521$；
5 分钟内被攻击次数服从参数10的Poisson分布，因此 $1-p(0;10)-p(1;10)\approx 0.999501$；
即更可能出现 5 分钟里至少被两次攻击。