第二章　条件概率与统计独立性

第4节　二项分布与泊松分布

李 娟

Email: juanlisdu@163.com

山东大学(威海)　　数学与统计学院

授课要求： 1. 掌握二项分布的性质及计算；

2. 掌握Poisson定理、Poisson分布的定义及性质；

3. 了解Poisson过程。

一、二项分布的性质及计算：

1. 二项分布的计算：

$$b(k;n,p)=C_n^k p^k(1-p)^{n-k},\ k=0,1,\cdots,n. \qquad (2.4.1)$$

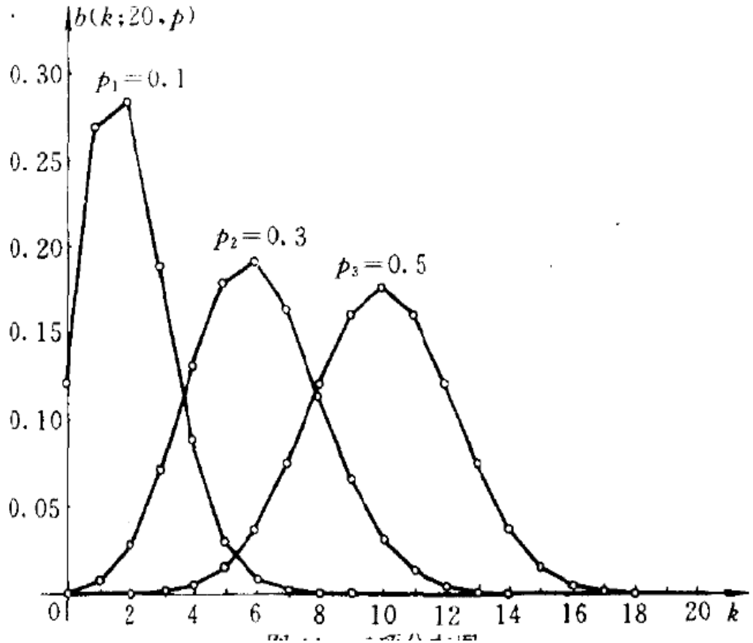
性质：

①对于固定的n及p，当k增加时，$b(k;n,p)$先随之增加并达到某最大值，然后再下降。

一、二项分布的性质及计算

②对于固定的p，随着n的增大，$b(k;n,p)$的图形趋于对称（$0<p<1$）。

③当概率p越与$\frac{1}{2}$接近时，分布越接近对称。



一、二项分布的性质及计算

注1: 二项分布的计算：二项分布表（$p \le 0.5$）

$$b(k;n,p) = b(n-k;n,1-p). \tag{2.4.2}$$

注2: 从图中可以看出，对于固定的n及p，当k增加时，$b(k;n,p)$先随之增加并达到某最大值，然后再下降。

例1. 一大批电子管中有10%已损坏，若我们从这批电子管中随机地选取20个来组成一个线路，问这线路能正常工作（即所选取的20个电子管全部是好的）的概率是多少？(0.1216)

分析：$b(20;20,0.9) = 0.9^{20}$。查表$b(0;20,0.1) = 0.01216$.

一、二项分布的性质及计算

p

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疾病有效，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？

解：假如血清毫无价值，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$b(0;20,0.3)+b(1;20,0.3)=0.0076.$

这个概率如此之小，因此我们不能认为血清毫无价值。 □

这里的做法是：先依照我们关心的问题提出一个假设，然后用实验得出的数据，利用概率论方法，计算某个事件在假设成立下的概率，最后根据这概率的大小来决定是接受还是拒绝原来的假设，这是数理统计中有名的**统计假设检验法**。

一、二项分布的性质及计算

2. 二项分布的性质：

①对于固定的n及p，当k增加时，$b(k;n,p)$ 先随k 增加而增大，达到某一最大值后又逐渐下降。

②对于固定的p，随着n的增大，$b(k;n,p)$的图形趋于对称（$0<p<1$）。

证明：对$0<p<1$，$\frac{b(k;n,p)}{b(k-1;n,p)}=\frac{(n-k+1)p}{kq}=1+\frac{(n+1)p-k}{kq}\cdots$

注: 使$b(k;n,p)$ 取最大值的项$b(m;n,p)$称为$b(k;n,p)$的中心项；而m称为最可能成功次数。$m=[(n+1)p]$。若$(n+1)p$ 是整数，则$m-1$ 也是最可能成功次数。

$K=(n+1)p$ 相等

$K<(n+1)p$ $b(k)>b(k-1)$ $K>(n+1)p$ $b(k)<b(k-1)$

一、二项分布的性质及计算

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。(0.17635)

解：$n=500,\ p=0.01,\ (n+1)p=5.01,\ [(n+1)p]=5.$
$b(5;500,0.01)=C_{500}^{5}(0.01)^5(0.99)^{495}=0.17635.$ □

注: 应该注意，若$0<p<1$，则当n值相当大时，即便是最可能成功次数m发生的概率也相当小，对于其他的k，则$b(k;n,p)$自然更小了，以后将会看到最可能成功次数m发生的概率接近于

$$(2\pi npq)^{-\frac{1}{2}}$$

（当n相当大时）。因此当$n\to\infty$时，这概率趋于0。

3. 产品抽样验收与（n, c）方案：

由于生产过程总有种种无法完全控制的因素，因此工艺规范也允许加工的尺寸有一定的误差，或允许产品中含有少量废品，这事实上是承认生产过程的随机性。

在产品质量管理中，全面检验一般是不可能的，因此采用抽样检查的方法。

抽样检查若用于生产过程中，则成为在线生产过程质量管理的一部分，此外就是用于产品的验收。

如果每个产品要么是好品要么是废品，那么这时关心的是废品数或废品率，这是计算抽样验收中最简单的情况。

3. 产品抽样验收与（n,c）方案：

①对质量的要求大体上可以归结为：存在p_0及p_1满足$0 < p_0 \le p_1 < 1$，当废品率$p \le p_0$时，接收这批产品；而当$p \ge p_1$时，拒绝这批产品。

②验收方案是：抽n件产品进行检验，当废品数$\le c$时，接收该批产品；否则拒绝。这个方案称为（n,c）方案。

③由于抽样的随机性，任何验收方案都可能犯两类错误：其一，拒收一批合格品；其二，接收一批不合格品。前者为生产者风险；后者为消费者风险。当然希望减小这两类风险，即降低犯两类错误的概率，这也为比较两种不同验收方案的优劣提供了客观的标准。

3. 产品抽样验收与（n, c）方案：

为刻画验收方案的性能，一般引进$L(p)$，它表示当废品率为p时，接收该产品的概率。若以p为横坐标，$L(p)$为纵坐标作图，则所得的曲线称为抽检特性曲线（operating characteristic curve），简称OC曲线。

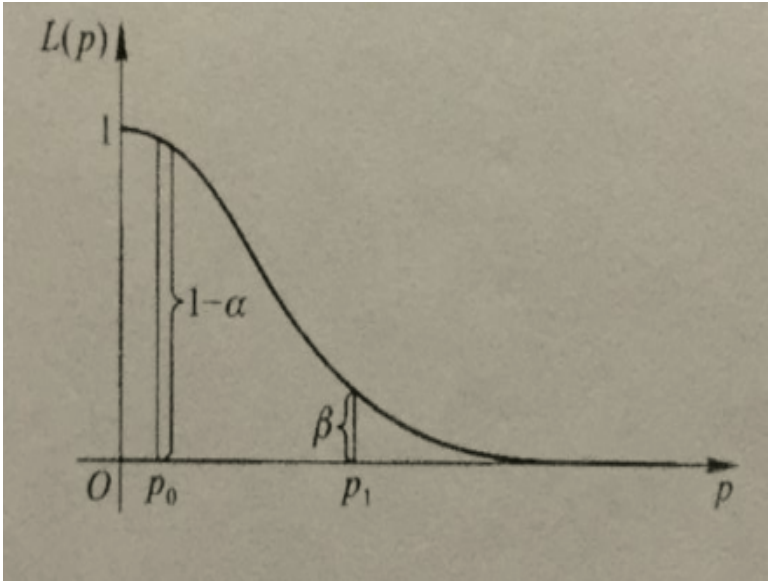


图 2.4.2 OC 曲线

3. 产品抽样验收与（n,c）方案：

④对（n,c）方案而言，若抽样是放回的，则利用二项分布容易得到

$$L(p)=\sum_{k=0}^{c}C_n^k p^k(1-p)^{(n-k)}. \tag{2.4.4}$$

⑤因此，问题归结为找n及c，使得

$$\begin{cases} L(p)\geq 1-\alpha,\ \text{当}p\leq p_0\text{时}; \\ L(p)\leq \beta,\ \text{当}p>p_1\text{时}; \end{cases} \tag{2.4.5}$$

这里α，β是两个不大的正数，按需要给定。

理想的验收方案要求$\alpha=\beta=0$，这是无法实现的，但可作为比较的基准。

4. 应用实例：

我们举一些应用的例子，说明二项分布的重要性，同时也提出一些问题。

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10000个这类人参加人寿保险，试求在未来一年中这些保险者里面，

（1）有40个人死亡的概率；（2）死亡人数不超过70个的概率。

直接计算这些数值相当困难，要有更好的计算方法。

例5（机票超售）：某航线历史资料表明：订座旅客有5%不来登机，问一架200座飞机应出售多少座位？

例6（车间用电）：某车间有200台车床，由于经常需要检修、测量、调换刀具、变换位置等种种原因，每台只有60%的时间在开动用电，若每台开动时耗电1千瓦，问应供给这个车间多少电力才能保证正常生产？（$1-\frac{0.5}{8\times 60}\approx 0.999.\ r=141.$）

例7（分子运动）：甲、乙两容器，容量各为1升，每个各含有2.7×10^{22}个气体分子，现将两容器接触，经过相当长的时间后（即这时每个分子落在两容器中的概率各为$\frac{1}{2}$），求两容器中分子数之差超过分子总数的100亿分之一(即10^{-10})的概率。

一、二项分布

从上面几个例子中可以看到，计算二项分布的数值时，由于试验次数n经常很大，因此实际计算$b(k;n,p)$及$\sum\limits_{k} b(k;n,p)$都很困难，有时甚至不可能。例如上面例7的计算过程中需算的项数有5.4×10^{12}之多，逐项计算是不可能的。在这种情况下，寻找更有效的计算法是必要的，即便是近似公式也好。这可以利用概率论中的极限定理来实现，关于极限定理的讨论将在第五章进行。

二、二项分布的泊松逼近

在很多应用问题中，我们常常遇到这样的Bernoulli试验，其中，相对地说，n大，p小，而乘积$\lambda = np$大小适中。对这种情况，泊松（Poisson，1781-1840）找到了一个便于使用的近似公式，下面我们来推导它。

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：

二、二项分布的泊松逼近

注1：$p(k;\lambda)=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,2,\ldots$ (2.4.10)

称为**泊松分布**（Poisson distribution），λ称为它的参数。

特别地，$\sum\limits_{k=0}^{\infty}p(k;\lambda)=\sum\limits_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}=e^{\lambda}e^{-\lambda}=1.$ (2.4.11)

泊松分布是概率论中很重要的一种分布。

注2：在应用中，当p相当小（一般当$p\leq 0.1$）时，我们用下面近似公式：

$$b(k;n,p)\approx\frac{(np)^k}{k!}e^{-np}=p(k;np). \tag{2.4.12}$$

例如，在例3中，要计算$b(5;500,0.01)(=0.17635)$时，这时$n=500$，相当大，而$p=0.01$相当小，但$np=5$，正好适中，所以很适合用Poisson逼近，查表得到：$p(5;5)=0.175467$，与0.17635十分接近。

注3：二项分布概率的计算($n < 50$)，有表可查；也有Poisson分布表可查。

例8：假如生三胞胎的概率为10^{-4}，求在100000次生育中，分别有$0, 1, 2$次生三胞胎的概率。

解：这可以看作Bernoulli试验；$n = 100000, p = 0.0001$, 所求的概率直接计算为：

$b(0; 100000, 0.0001) = 0.000045378$; $b(1; 100000, 0.0001) = 0.00045382$;

$b(2; 100000, 0.0001) = 0.0022693$. □

这时也可用Poisson逼近，$\lambda = np = 10$, 而

$$p(0; 10) = 0.000045;\ p(1; 10) = 0.000454;\ p(2; 10) = 0.002270.$$

可见近似程度很令人满意。

二、二项分布的泊松逼近

图2.4.3给出了泊松分布逼近二项分布的一个图示，吻合度甚好。

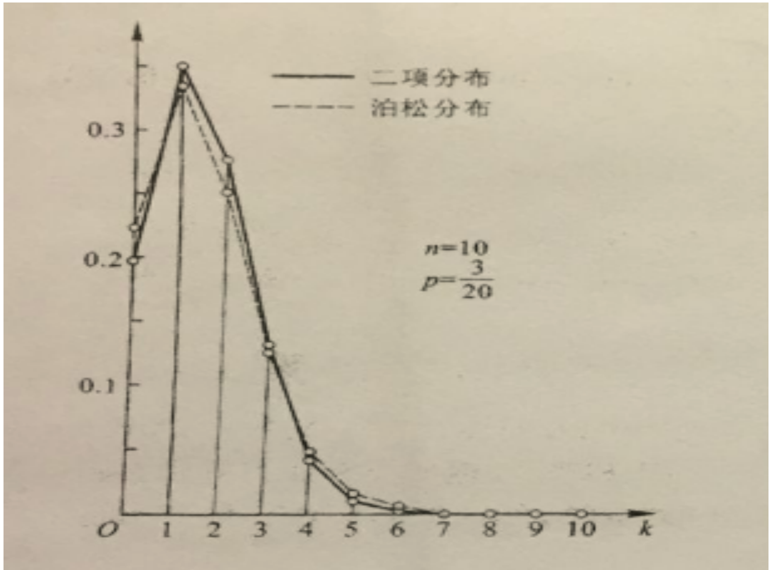


图 2.4.3 二项分布与泊松分布

三、泊松分布

1、直观背景

①一是社会生活，对服务的各种要求，诸如：某电话交换台中来到的呼叫数，网站访问数，公共汽车站来到的乘客数等都近似地服从泊松分布。因此在运筹学及管理科学中泊松分布占有突出的地位；
②另一领域是物理科学，放射性分裂落到某区域的质点数，热电子的发散，显微镜下落在某区域中的血球或微生物的数目等等都服从泊松分布。

……对泊松分布的深入研究（特别是通过对Poisson过程的研究）已发现它具有许多特殊的性质和作用，打个不很恰当的譬如，似乎泊松分布是构造随机现象的“基本粒子”之一。

三、泊松分布

图2.4.4是对于不同λ值的泊松分布图。为了计算泊松分布的数值，有许多专门的表格可供查用。本书附录一中也附有这样的表。例8的数值可从该表查到。

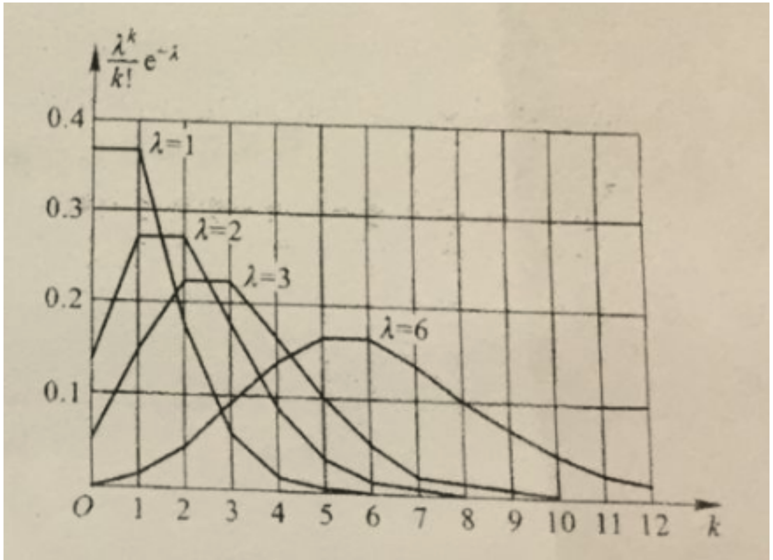


图 2.4.4 泊松分布图

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，结果列于下表中，其相应的频率与$\lambda = 0.87$的泊松分布符合得很好。

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

三、泊松分布

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$N = 2608$次，下表给出了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，理论值为$N_p(k; 3.870)$，理论值与实验值很近似。

三、泊松分布

表 2.4.3 Rutherford 实验理论值与实验值对照表①

k	N_k	$N_p(k;\ 3.870)$
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	2608	2608.000

三、泊松分布

在说明了泊松分布的常见性之后，我们转入介绍产生泊松分布的机制。经过研究，已经弄清了服从泊松分布的条件。为了便于理解，我们将结合电话呼叫流来叙述这个重要结果。

我们先证明一个以后屡次要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$x \geq 0$, $y \geq 0$）成立：

$$f(x)f(y) = f(x+y), \qquad (2.4.13)$$

则$f(x) = a^x$ (2.4.14)，其中$a \geq 0$，是某一常数。

三、泊松分布

2、泊松过程：以ξ_t表示某交换装置在时间区间$[0,t)$ 上来到的电话呼叫数，若它满足下述条件，则称它为一个Poisson过程(Poisson流)：

(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有关而与时间起点t_0无关。若以$P_k(t)$记在长度为t的时间区间中来到k个呼叫的概率，当然有 $\sum\limits_{k=0}^{\infty} P_k(t)=1, \qquad (2.4.15)$ 对任何$t>0$成立。

过程的平稳性表示了它的概率规律不随时间的推移而改变。

(ii)独立增量性（无后效性）：在$[t_0,t_0+t)$内来到k个呼叫这一事件与时刻t_0以前发生的事件独立。换言之，在对时刻t_0以前的事件发生情况所作的任何假定之下，计算出来的在$[t_0,t_0+t)$内发生k个呼叫的条件概率都等于同一事件的无条件概率。独立增量性表明在互不相交的时间区间内过程进行的相互独立性。

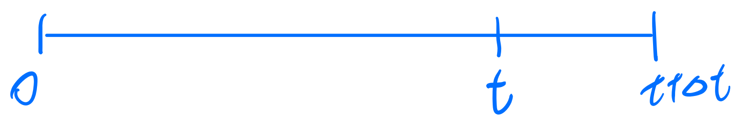
三、泊松分布

(iii)普通性：在充分小的时间间隔中，最多来到一个呼叫，即，若记：

$$\psi(t) = \sum_{k=2}^{\infty} P_k(t) = 1 - P_0(t) - P_1(t), \qquad (2.4.16)$$

应有$\psi(t) = o(t)$，i.e.，$\lim\limits_{t\to 0} \frac{\psi(t)}{t} = 0.$ (2.4.17)

普通性表明，在同一时间瞬间来两个或两个以上呼叫实际上是不可能的。



$$P_k(t+\Delta t) = \sum_{i=0}^{k} P_i(t) P_{k-i}(\Delta t) = P_k(t) P_0(\Delta t) + P_{k-1}(t) P_1(\Delta t) + \boxed{\quad} \to 0$$

三、泊松分布

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

对于泊松过程，成立下面的定理。

定理

泊松过程ξ_t必服从以λt为参数的泊松分布，其中$\lambda > 0$为常数，i.e.,

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

证明：

$P_k(t+\Delta t) - P_k(t) = P_k(t)[e^{-\lambda \Delta t} - 1] + P_{k-1}(t) P_1(\Delta t)$

$P_k'(t) = -\lambda P_k(t) + \frac{P_{k-1}(t) P_1(\Delta t)}{\Delta t}$

$P_1(\Delta t) \le 1 - P_0(\Delta t)$

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

$$= \lambda\left(P_{k-1}(t) - P_k(t)\right)$$

由 $P_0 = e^{-\lambda t}$

$$P_1' = \lambda\left(e^{-\lambda t} - P_1\right)$$

$$\Rightarrow P_1 = \lambda t e^{-\lambda t}$$

$\vdots$

$$P_k = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$