人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE

目录

重点难点

掌握：二项分布、Poisson分布、负二项分布的基本概念与适用条件

熟悉：二项分布、Poisson分布的参数区间估计与假设检验

了解：负二项分布的参数估计与应用

第一节

二项分布

在生物医学领域，服从二项分布的试验较为常见。如用某种药物治疗某种非传染性疾病，其疗效分为有效与无效；在动物的急性毒性实验中，观测动物的死亡与存活；接触某种病毒性疾病的传播媒介后，出现感染与非感染等。对于抽样而言，若从阳性率（如患病率）为 π 的总体中，有放回地随机抽取个体数为 n 的样本，则出现阳性数为 X 的概率分布即呈二项分布。若是无放回地随机抽样，当抽取的个体数 n 远小于总体的个体数 N（如 $n<\dfrac{N}{10}$）时，也可近似当作二项分布处理。

例6-1 某种医学技能测试的通过率为0.80。今有10名学生参加测试，试分别计算这10名学生中有6人、7人和8人获得通过的概率。

本例 $n=10$ ，$\pi=0.80$ ，$X=6,7,8$ 。按公式（6-1）计算相应的概率为

$$P(6)=\frac{10!}{6!(10-6)!}0.80^{6}(1-0.80)^{10-6}=0.08808$$

$$P(7)=\frac{10!}{7!(10-7)!}0.80^{7}(1-0.80)^{10-7}=0.20133$$

$$P(8)=\frac{10!}{8!(10-8)!}0.80^{8}(1-0.80)^{10-8}=0.30199$$

一、二项分布的适用条件和性质

（一）二项分布的适用条件

（二）二项分布的性质

（一）二项分布的适用条件

1. 每次试验只会发生两种对立的可能结果之一，即分别发生两种结果的概率之和恒等于1。

2. 每次试验产生某种结果（如“阳性”）的概率 π 固定不变。

3. 重复试验是相互独立的，即任何一次试验结果的出现不会影响其他试验结果出现的概率。

在上面的例6-1中，对这10名学生的测试，可看作10次独立的重复试验，通过与否为二分类结果，且测试的通过率（π=0.80）是恒定的。这样，10人中测试通过的人数 $X \sim B(10,0.80)$。

（二）二项分布的性质

1. 二项分布的均数与标准差

2.二项分布的图形

1.二项分布的均数与标准差

在 n 次独立重复试验中，出现“阳性”次数 X 的总体均数为 $\mu = n\pi$ （6-2）

X 的总体方差为 $\sigma^2 = n\pi(1-\pi)$ （6-3）

X 的总体标准差为 $\sigma = \sqrt{n\pi(1-\pi)}$ （6-4）

若以率表示，则样本阳性率 $p(p = 0/n, 1/n, 2/n, \ldots, n/n)$ 也服从公式（6-1）的二项分布，其总体均数为 $\mu_p = \pi$ （6-5）

p 的总体方差为 $\sigma_p^2 = \dfrac{\pi(1-\pi)}{n}$ （6-6）

p 的总体标准差为 $\sigma_p = \sqrt{\dfrac{\pi(1-\pi)}{n}}$ （6-7）

1.二项分布的均数与标准差

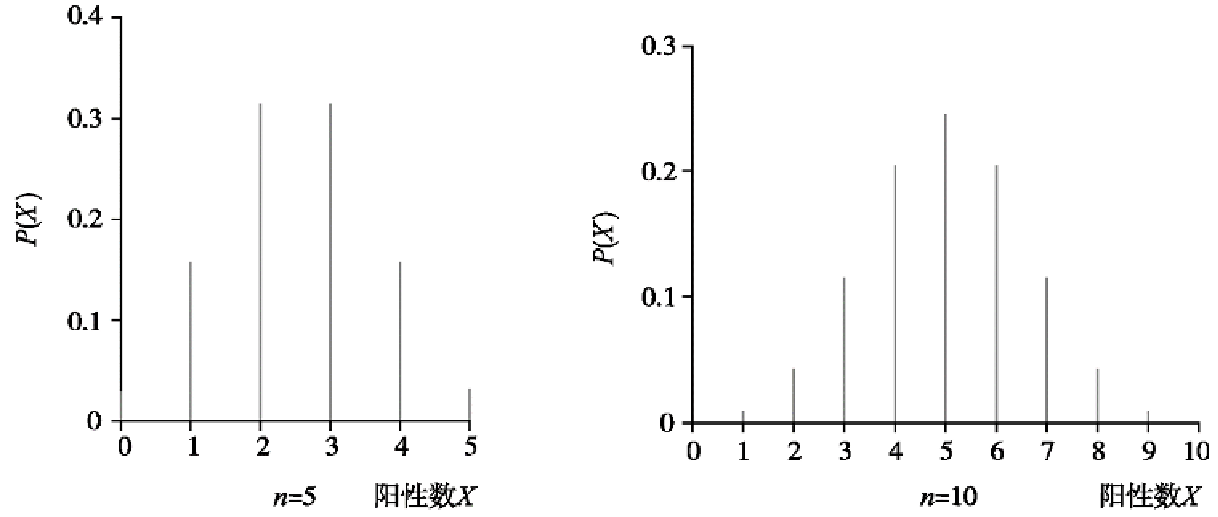
样本率的标准差也称为率的标准误，可用来描述样本率的抽样误差，率的标准误越小，则率的抽样误差就越小。

在一般情形下，总体率 π 往往并不知道。此时若用样本资料计算样本率 $p = X/n$ 作为 π 的估计值，则 σ_p 的估计为

$$S_p = \sqrt{\frac{p(1-p)}{n}} \qquad (6\text{-}8)$$

2.二项分布的图形

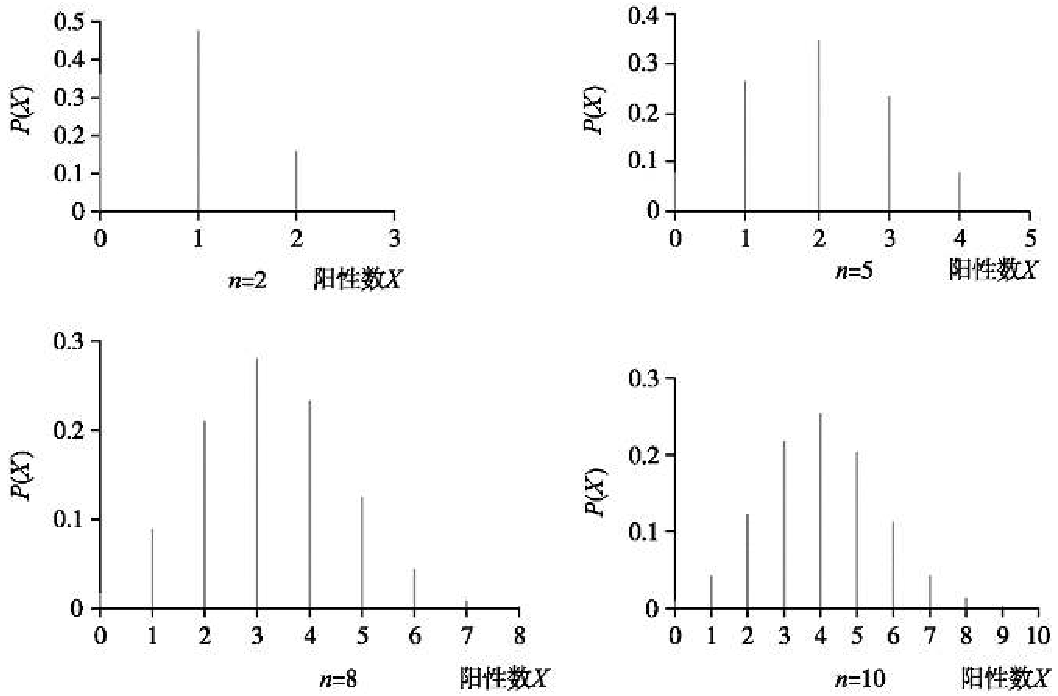
当 π=0.5 ，二项分布图形是对称的；



π=0.5时，不同n值下的二项分布图

2.二项分布的图形

当 $\pi \neq 0.5$，图形是偏态的；随着 n 增大，图形趋于对称。当$n \to \infty$时，只要π 不太靠近0或1，二项分布则近似正态分布。



π=0.4时，不同n值下的二项分布图

二、二项分布的应用

（一）总体率的区间估计

（二）样本率与总体率的比较

（三）两样本率的比较

（四）非遗传性疾病的家族集聚性

（五）群检验

（一）总体率的区间估计

1. 查表法

2. 正态近似法

1.查表法

对于$n \leq 50$的小样本资料，直接查附表6百分率的可信区间表，即可得到其总体率的 $1-\alpha$ 可信区间。

1.查表法

例6-2 在对13名输卵管结扎的育龄妇女经壶腹部-壶腹部吻合术后，观察其受孕情况，发现有6人受孕，据此资料估计该吻合术妇女受孕率的95%可信区间。

本例$n=13$，$X=6$。查附表6，α取0.05时，在$n=13$（横行）与$X=6$（纵列）的交叉处数值为19～75，即该吻合术妇女受孕率的95%可信区间为（19%，75%）。

在附表6百分率的可信区间表中，通常只列出$X \le n/2$的部分。当$X > n/2$时，可先按"阴性"数$n-X$查得总体阴性率的$1-\alpha$可信区间$Q_L \sim Q_U$，再用下面的公式转换成所需的阳性率的$1-\alpha$可信区间$P_L \sim P_U$。

$$P_L = 1 - Q_U,\ P_U = 1 - Q_L$$

2.正态近似法

当 n 较大，π 或（$1-\pi$）不接近0，也不接近1时，二项分布B（n，π）近似正态分布 $N[n\pi,n\pi(1-\pi)]$，而对应的样本率 p 也近似正态分布 $N(\pi,\sigma_p^2)$。为此，当 n 较大、p和$1-p$ 均不太小，如 np 和 n（$1-p$）均大于5时，可利用样本率 p 的分布近似正态分布来估计总体率的 $1-\alpha$ 可信区间。计算公式为

$$(p-u_{\alpha/2}S_p,\ p+u_{\alpha/2}S_p) \qquad (6\text{-}10)$$

式中 $\alpha=0.05$ 时，$u_{0.05/2}=1.96$；$\alpha=0.01$ 时，$u_{0.01/2}=2.58$。

2.正态近似法

例6-3 在一项光动力疗法治疗伴有完全梗阻或不完全梗阻的原发性晚期食管癌的单臂临床试验中，采用治疗28天基于目标肿瘤病灶缩小≥50%为判定标准的客观缓解作为主要有效性评价指标。试验共入组受试者100例，有45例患者治疗28天获得客观缓解，试据此估计该光动力疗法客观缓解率的95%可信区间。

2.正态近似法

本例 n=100 $p=45/100=0.45$，。代入公式（6-8）

$$S_p=\sqrt{\frac{0.45(1-0.45)}{100}}=0.0497$$

代入公式（6-10）

$$0.45-1.96\times0.0497=0.3526$$

$$0.45+1.96\times0.0497=0.5474$$

即该光动力疗法客观缓解率的95%可信区间为（35.26%，54.74%）。

（二）样本率与总体率的比较

1. 直接法

2. 正态近似法

1.直接法

在诸如疗效评价中，利用二项分布直接计算相关概率来推断样本所在的总体率与已知总体率有无差别。比较时，经常遇到单侧检验，即“优”或“劣”的问题。那么，在总体阳性率为 π 的 n 次独立重复试验中，一般有下面两种情形的概率计算。

（1）若是回答“差”或“低”的问题，则需计算出现“阳性”次数至多为 k 次的概率，即

$$P(X \le k)=\sum_{X=0}^{k} P(X)=\sum_{X=0}^{k} \frac{n!}{X!(n-X)!}\pi^{X}(1-\pi)^{n-X} \qquad （6\text{-}11）$$

（2）若是回答“优”或“高”的问题，则需计算出现“阳性”次数至少为 k 次的概率，即

$$P(X \ge k)=\sum_{X=k}^{n} P(X)=\sum_{X=k}^{n} \frac{n!}{X!(n-X)!}\pi^{X}(1-\pi)^{n-X}$$

（6-1
2）

$$P(X \le k)+P(X \ge k)=1+P(k)$$

1.直接法

对于双侧检验而言，由于要回答的是“有无差别”，即备择假设$H_1:\pi \neq \pi_0$是否成立，因此所要计算的双侧检验概率P值应为实际样本（记“阳性”次数为k次）出现的概率与更背离无效假设的事件（记“阳性”次数为i次$i \neq k$）出现的概率之和，即$P = P(X=k) + \sum_i P(X=i)$，其中$i$满足$P(X=i) \leq P(X=k)$。

1.直接法

例6-4　已知输卵管结扎的育龄妇女实施壶腹部–壶腹部吻合术后的受孕率为0.55。今对10名输卵管结扎了的育龄妇女实施峡部–峡部吻合术，结果有9人受孕。问实施峡部–峡部吻合术妇女的受孕率是否高于壶腹部–壶腹部吻合术的受孕率？

1.直接法

显然，这是单侧检验的问题，属上述第（2）种情况，记峡部-峡部吻合术后的受孕率为 π ，其假设检验为

$$H_0: \pi = 0.55$$

$$H_1: \pi > 0.55$$

$$\alpha = 0.05$$

对这10名实施峡部-峡部吻合术的妇女，按0.55的受孕率，若出现至少9人受孕的概率大于0.05，则不拒绝H_0；否则，可视为小概率事件，拒绝H_0，接受H_1。

1.直接法

本例 $n=10$ $\pi=0.55$， $k=9$ ， 。按公式（6-12）有

$$P=P(X\geq 9)=\sum_{X=9}^{10}P(X)=\sum_{X=9}^{10}\frac{10!}{X!(10-X)!}0.55^{X}(1-0.55)^{10-X}=0.023$$

$0.01<P<0.05$ $\alpha=0.05$ ，按 水准，拒绝H_0，接受H_1，即认为实施峡部-峡部吻合术妇女的受孕率要高于壶腹部-壶腹部吻合术的受孕率。

2.正态近似法

当 n 较大、p 和 $1\text{-}p$ 均不太小，如 np 和 n（$1\text{-}p$）均大于5时，利用样本率的分布近似正态分布的原理，可作样本所在的总体率 π 与已知总体率 π_0 的比较。检验统计量 u 值的计算公式为

$$u=\frac{p-\pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \tag{6-13}$$

（三）两样本率的比较

两样本率的比较目的在于对相应的两总体率进行统计推断。设两样本率分别为 p_1 和 p_2，当 n_1 与 n_2 均较大，且 p_1、$1-p_1$ 与 p_2、$1-p_2$ 均不太小，如 n_1p_1、$n_1(1-p_1)$ 与 n_2p_2、$n_2(1-p_2)$ 均大于5时，可利用样本率的分布近似正态分布以及独立的两个正态变量之差也服从正态分布的性质，采用正态近似法对两总体率作统计推断。检验统计量 u 的计算公式为

$$u=\frac{p_1-p_2}{S_{p_1-p_2}} \qquad (6\text{-}14)$$

其中

$$S_{p_1-p_2}=\sqrt{\frac{X_1+X_2}{n_1+n_2}\left(1-\frac{X_1+X_2}{n_1+n_2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \qquad (6\text{-}15)$$

（三）两样本率的比较

例6-7 为研究某职业人群颈椎病发病的性别差异，今随机抽查了该职业人群男性120人和女性110人，发现男性中有36人患有颈椎病，女性中有22人患有颈椎病。试作统计推断。

令该职业人群颈椎病的患病率男性为 π_1 ，女性为 π_2 ，其检验假设为

H_0: $\pi_1 = \pi_2$

$H_1: \pi_1 \neq \pi_2$

α=0.05

（三）两样本率的比较

本例 $n_1 = 120$ ，$X_1 = 36$ ，$p_1 = X_1 / n_1 = 36/120 = 0.30$

$n_2 = 110$ ，$X_2 = 22$ ，$p_2 = X_2 / n_2 = 22/110 = 0.20$

按公式（6-15）有

$$S_{p_1 - p_2} = \sqrt{\frac{36+22}{120+110}\left(1 - \frac{36+22}{120+110}\right)\left(\frac{1}{120} + \frac{1}{110}\right)} = 0.0573$$

按公式（6-14）有

$$u = \frac{0.30 - 0.20}{0.0573} = 1.745$$

查 u 界值表得 $0.05 < P < 0.10$ 。按 $\alpha = 0.05$ 水准，不拒绝 H_0 ，尚不能认为该职业人群颈椎病的发病有性别差异。

（四）非遗传性疾病的家族集聚性

非遗传性疾病的家族集聚性（clustering in families），系指该疾病的发生在家族成员间是否有传染性，如果没有传染性，则家族成员间的患病是独立的，该疾病无家族集聚性。否则，家族成员间的患病是非独立的，该疾病存在家族集聚性。

当这种非遗传性疾病无家族集聚性时，以家族为样本，在 n 个成员中，出现 X 个成员患病的概率分布可认为服从二项分布；否则，便不服从二项分布。

（四）非遗传性疾病的家族集聚性

例6-8 某研究者为研究某种非遗传性疾病的家族集聚性，对一社区82户3口人的家庭进行了该种疾病患病情况调查，所得数据资料见表中的第（1）、（2）栏。试分析其家族集聚性。

患病数据资料与二项分布拟合优度的检验

X （1）	实际户数A （2）	概率 P（X）（3）	理论户数 (4)＝82×(3)	$T-A$ （5）	$(T-A)^2$ （6）	$(T-A)^2/T$ （7）
0	26	0.13265	10.8774	-15.1226	228.6936	21.0247
1	10	0.38235	31.3525	21.3525	455.9273	14.5420
2	28	0.36735	30.1229	2.1229	4.5069	0.1496
3	18	0.11765	9.6472	-8.3528	69.7690	7.2320
合计	82	—	82.0000	—	—	42.9483

（四）非遗传性疾病的家族集聚性

如果该社区的此种疾病不存在家族集聚性，则以每户3口人的家庭为样本，在3个家庭成员中，出现 $X(=0, 1, 2, 3)$ 个成员患病的概率分布可认为服从二项分布。假设检验为

H_0: 该疾病的发生无家族集聚性

H_1: 该疾病的发生有家族集聚性

$\alpha=0.10$

（四）非遗传性疾病的家族集聚性

本例调查的总人数为 $N=82\times3=246$（人），

其中患病人数为 $D=0\times26+1\times10+2\times28+3\times18=120$（人）。以这246人的患病率作为总体患病率的估计值，即 $\pi=D/N=120/246=0.49$。

在 $n=3$、$\pi=0.49$ 时，利用二项分布，求得 $X=0,1,2,3$ 的概率 P（X），并以此得到相应的理论户数。对理论户数与实际户数进行拟合优度（goodness of fit）χ^2 检验。此时，自由度为 $\nu=$ 组数 $-2=4-2=2$。计算结果列于表中的第（3）至（7）栏。

（四）非遗传性疾病的家族集聚性

以 $\nu=2$ 、$\chi^2=42.95$ 查附表8的 χ^2 界值表得 $P<0.005$ 。按 $\alpha=0.10$ 水准，拒绝H_0，接受H_1，即此种疾病存在家族集聚性。