第四章 几种重要的分布

4.1 二项分布

在第一章介绍过独立试验概型

作n次相互独立的试验，每次试验事件A出现的概率为p，不出现的概率为$q=1-p$，事件A出现的次数ξ为一离散型随机变量，则我们已经知道

$$P\{\xi=k\}=C_n^k p^k q^{n-k}$$

如假设第i次试验时事件A发生的次数为随机变量ξ_i，则ξ_i服从0-1分布，

$P\{\xi_i=1\}=p$, $P\{\xi_i=0\}=q=1-p$, $(i=1,2,\ldots,n)$

因此有$\xi=\xi_1+\xi_2+\ldots+\xi_n$

定义 4.1 如果随机变量ξ有概率函数

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

其中$0<p<1$, $q=1-p$, 则称ξ服从参数为n,p的二项分布. 简记作$\xi \sim B(n,p)$.

在这里$P\{\xi=k\}$的值恰好是二项式$(q+px)^n$展开式中第$k+1$项x^k的系数.

如果$\xi \sim B(n,p)$, 则ξ可看作是由n个取1概率为p的相互独立的0-1分布的随机变量$\xi_i, i=1,2,\ldots,n$的和,

$$\xi=\xi_1+\xi_2+\ldots+\xi_n$$

ξ的分布函数为

$$F(x)=\sum_{k\le x}C_n^k p^k q^{n-k}$$

事件A至多出现m次的概率是

$$P\{0\le\xi\le m\}=\sum_{k=0}^{m}C_n^k p^k q^{n-k}$$

事件A出现次数不小于l不大于m的概率是

$$P\{0\le\xi\le m\}=\sum_{k=l}^{m}C_n^k p^k q^{n-k}$$

例1 某工厂每天用水量保持正常的概率为3/4, 求最近6天内用水量正常的天数的分布.

解 设最近6天内用水量保持正常的天数为ξ, 则ξ~B(6,0.75), 因此

$$P\{\xi=0\}=\left(\frac{1}{4}\right)^6=0.0002$$

$$P\{\xi=1\}=C_6^1\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^5=0.0044$$

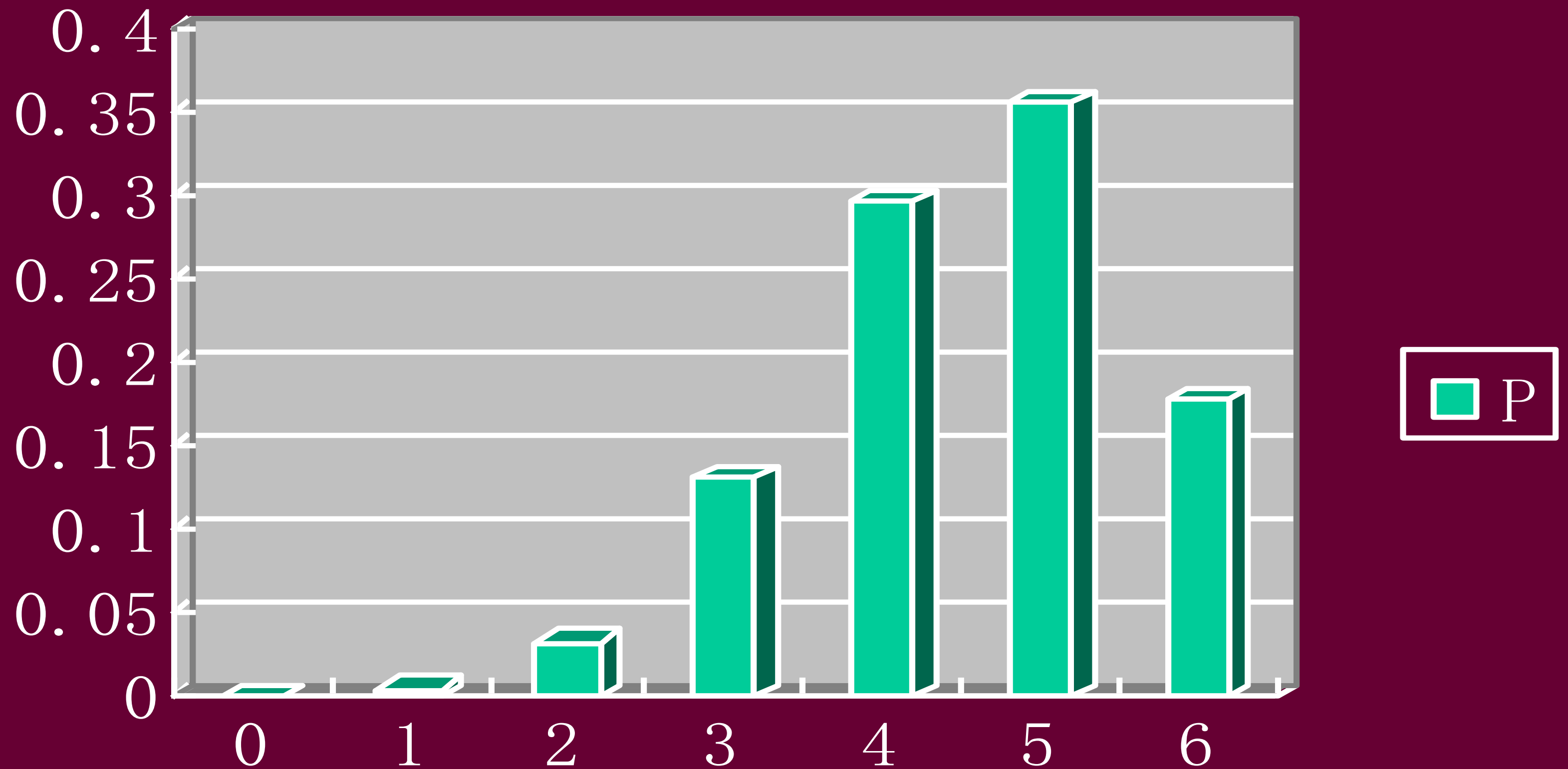
$$\cdots$$

$$P\{\xi=6\}=\left(\frac{3}{4}\right)^4=0.178$$

其分布表如下表所示

ξ	0	1	2	3	4	5	6
P	0.0002	0.0044	0.033	0.1318	0.2966	0.356	0.178

分布图:



例3 一批产品的废品率p=0.03, 进行20次重复抽样(每次抽一个, 观察后放回去再抽下一个), 求出现废品的频率为0.1的概率.

解 令ξ表示20次重复抽取中废品出现的次数, $\xi \sim B(20, 0.03)$

$$P\left(\frac{\xi}{20}=0.1\right)=P(\xi=2)$$

$$=C_{20}^{2}\times 0.03^{2}\times 0.97^{18}=0.0988$$

二项分布的期望和方差

如ξ~$B(n,p)$, 则ξ可看作n个相互独立的0-1分布的随机变量$\xi_1,\xi_2,\ldots,\xi_n$之和,

$\xi=\xi_1+\xi_2+\ldots+\xi_n$

而且我们知道0-1分布的期望为p, 方差为pq, 因此易得

$E\xi=E\xi_1+E\xi_2+\ldots+E\xi_n=np$

$D\xi=D\xi_1+D\xi_2+\ldots+D\xi_n=npq$

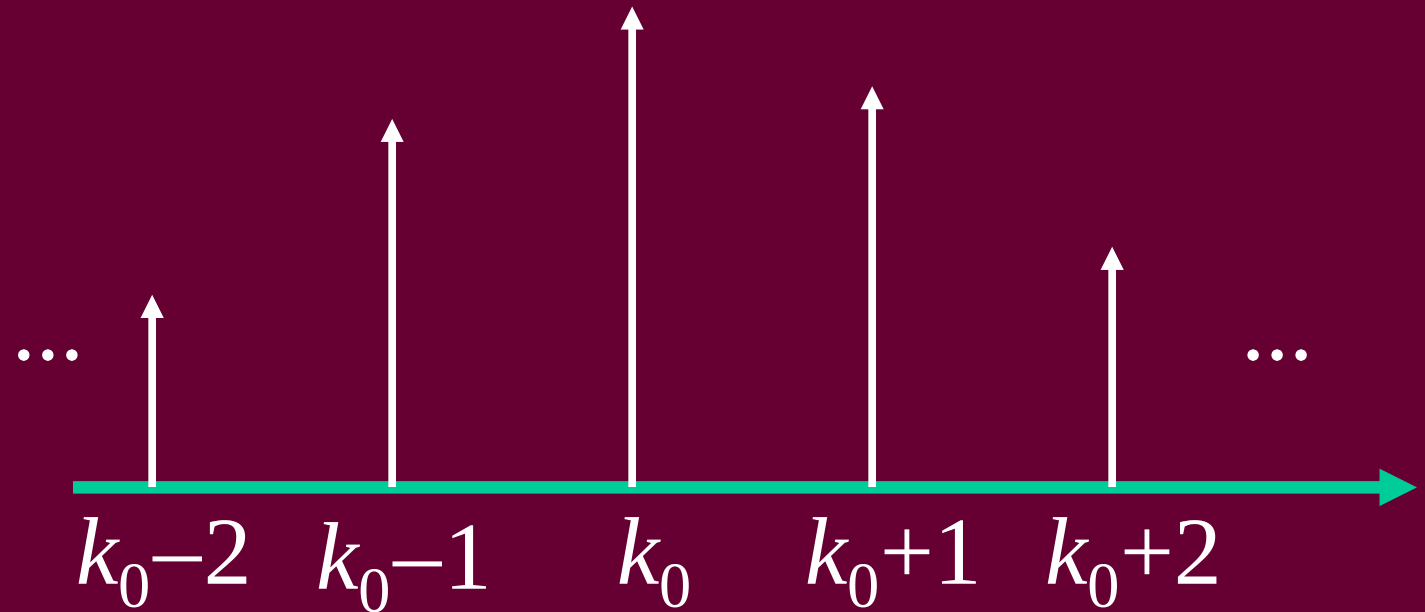
即

$$E\xi=np,\ D\xi=npq,\ \sigma_\xi=\sqrt{npq}$$

当然，也可直接计算二项分布的数学期望和方差

二项分布的最可能值

使概率$P\{\xi=k\}$取最大值的k记作k_0, 称k_0为二项分布的最可能值, 如图示意



由上图可知$P(\xi=k_0)\geq P(\xi=k_0+1)$

且$P(\xi=k_0)\geq P(\xi=k_0-1)$

因此得

因 $$\frac{C_n^k}{C_n^{k-1}} = \frac{n!}{k!(n-k)!} \cdot \frac{(k-1)!(n-k+1)!}{n!} = \frac{n-k+1}{k}$$

由 $$\frac{C_n^{k_0} p^{k_0} q^{n-k_0}}{C_n^{k_0-1} p^{k_0-1} q^{n-k_0+1}} = \frac{(n-k_0+1)p}{k_0 q} \geq 1$$

$(n+1)p \geq k_0$, 得$np + p \geq k_0$

再由 $$\frac{C_n^{k_0+1} p^{k_0+1} q^{n-k_0-1}}{C_n^{k_0} p^{k_0} q^{n-k_0}} = \frac{(n-k_0)p}{(k_0+1)q} \leq 1$$

$np + p - 1 \leq k_0$,即$np + p - 1 \leq k_0 \leq np + p$

分析$np+p-1\le k_0\le np+p$

知道$np+p$比$np+p-1$大了1, 因此挤在这两个数中间的整数有1个还是2个取决于$np+p$是否正好是整数. 如果正好是整数, 则无论是$np+p$还是$np+p-1$都满足上面的不等式, 这个时候就有两个最可能值$np+p-1$和$np+p$.

如果$np+p$不是整数, 则k_0取被$np+p-1$和$np+p$夹在中间的整数才能够满足上面的不等式. 因此可以看作是不大于$np+p$的最大整数, 记作$[np+p]$

例4 某批产品有80%的一等品, 对它们进行重复抽样检验, 共取出4个样品, 求其中一等品数ξ的最可能值k_0, 并用贝努里公式验证.

解 $\xi\sim B(4, 0.8)$,

因$np+p=4\times0.8+0.8=4$是整数, 所以$k_0=4$和$k_0=3$时$P\{\xi=k\}$为最大, 即3和4为最可能值.

ξ	0	1	2	3	4
P	0.0016	0.0256	0.1536	0.4096	0.4096