# 目录

文件名

# 第一部分 “离散型数据”(PMF) 的分布

# 1 伯努利分布

---

## 2 二项分布 (binomial distribution) : B(试验次数n, 每项试验成功的概率p)

**二项, 代表“有两个结果”. 比如, 一个为“成功”, 另一个为“失败”.**

- 如: 投硬币10次(而不是只做一次实验)，让X 代表“正面向上的次数”． 那么X 就是一个服从“二项分布”的随机变量 — 每投一次硬币只有两种结果: 要么是”正面朝上”, 要么是”反面朝上”.
- 你的教授给来了一个惊喜的突击测验，考试是10个判断题. 你对某一道题的猜测, 就属于“伯努利事件 a Binomial Event” (因为它只有两种选择, ”对”或”错”). 而整个测验(连续做n次伯努利事件), 是属于一个“二项事件” the entire quiz is a Binomial Event.

所以**本质上，“二项事件”是一系列相同的“伯努利事件”.**

我们用字母“B” 来表示二项分布，即: $\boxed{B(\text{试验次数}n, \text{每项试验成功的概率}p)}$.

例

如: 我们将 $X \sim B(10, 0.6)$ 读作: 变量X 遵循10次试验中, 每项试验成功的可能性为0.6的二项分布.
Variable “X” follows a Binomial distribution with 10 trials /and a likelihood of success of 0.6 /on each individual trial.

**二项分布表示, 在特定的次数内, 能达到我们“期望结果”的可能性.** the graph of the binomial distribution /represents(v.) the likelihood of /attaining(v.) our desired outcome /a specific number of times.

某事件A发生的概率是P, 我们在做了n次试验后, 得到"该事件A 发生了k次", 则:

$$P(X=k) = \underbrace{C_n^k}_{\text{总}n\text{中取}k} \cdot \underbrace{P^k}_{k\text{次成功}} \cdot \underbrace{(1-P)^{n-k}}_{n-k\text{次失败}} \quad k = 0, 1, 2, ..., n.$$

$$\text{记作:} X \sim B\left( \underbrace{n}_{\text{一共做了}n\text{次实验}}, \underbrace{p}_{\text{事件}A\text{发生的概率}} \right)$$

例

某药物, 临床有效率为0.95. 今有10人服用, 问"至少有8人能治愈"的概率是多少? (即做10次实验, 8次成功)
代入"二项分布"公式:

$$P(X \geq 8) = \underbrace{P(X=8)} + \underbrace{P(X=9)} + \underbrace{P(X=10)}$$

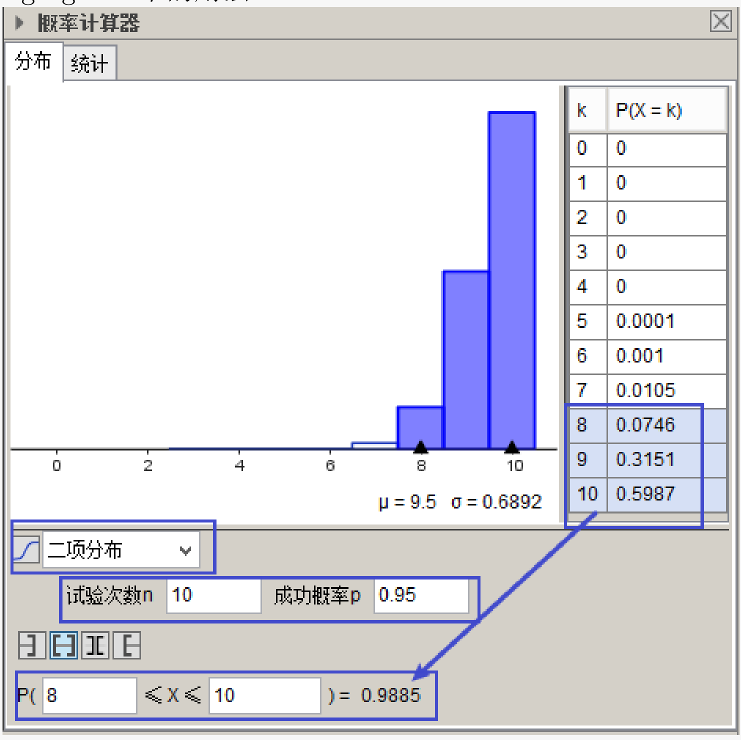
$$= \underbrace{C_{10}^8 \cdot 0.95^8 \cdot (1-0.95)^2}_{P(X=8)} + \underbrace{C_{10}^9 \cdot 0.95^9 \cdot (1-0.95)^1}_{P(X=9)} + \underbrace{C_{10}^{10} \cdot 0.95^{10} \cdot (1-0.95)^0}_{P(X=10)}$$

$$= 0.0746348 + 0.315125 + 0.598737$$

$$= 0.988496$$

所以, 10人中有8人以上被治愈的概率, 为0.98.

- geogebra 中的用法:




上表中, P(X=8)=0.0746 的意思, 就是 (对于单次实验是0.95的成功率的事件,) "做10次实验, 里面会成功8次" 的概率=7.46%.
同理, P(X=9) 的意思, 是"做10次实验, 里面会成功9次" 的概率.

- mathematica 中的用法:

某特效药的临床有效率为0.95，今有10人服用，问至少有8人治愈的概率是多少？即:
n=10次实验,
p=单次成功概率是0.95

```
In[9]:= resDist = BinomialDistribution[10, 0.95]
                  二项分布

Out[9]= BinomialDistribution[10, 0.95]

        1 - CDF[resDist, 7]
            累积分布函数

Out[15]= 0.988496
```

← 黄色部分, 只是0-7次成功的累积函数值, 题目问的是至少8次以上, 即 >8次成功 的累积概率值. 所以概率上就 = 1- (0到7次的累积概率值)

**即：我们只要知道单次的成功概率，就能计算n次成功中，会成功k次的概率.**

例

某报警器, 在发生危险时, 成功报警的概率是0.8. 问: 要将报警成功率提高到99%, 至少要安装多少台才行?

我们令:
- n: 表示总共安装的台数.
- X: 表示成功报警的台数.

则, 安装的总n台中, 只要至少有一台能报警 (即 $P(X \geq 1)$), 就成功了.

本例即: $X \sim B$(一共做n次实验, 单次实验的成功概率0.8)

$$\text{即:}\quad \underbrace{P(X \geq 1)}_{\text{至少1台报警 的概率}} \geq 0.99$$

$$\underbrace{1 - P(X = 0)}_{\text{全都没报警 的概率}} \geq 0.99$$

$$1 - \underbrace{C_n^0 \cdot \underbrace{0.8^0}_{0\text{台报警 的概率}} \cdot \underbrace{0.2^n}_{n\text{台没报警 的概率}}}_{\text{总}n\text{台里面,只有0台报警}} \geq 0.99$$

$$1 - 0.2^n \geq 0.99$$

$$1 - 0.99 \geq 0.2^n$$

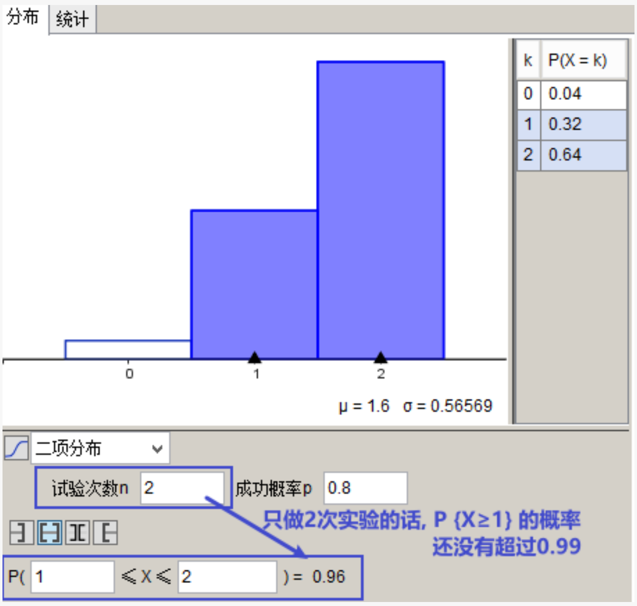
$$\ln 0.01 \geq \ln 0.2^n \leftarrow \text{两边取} \ln$$

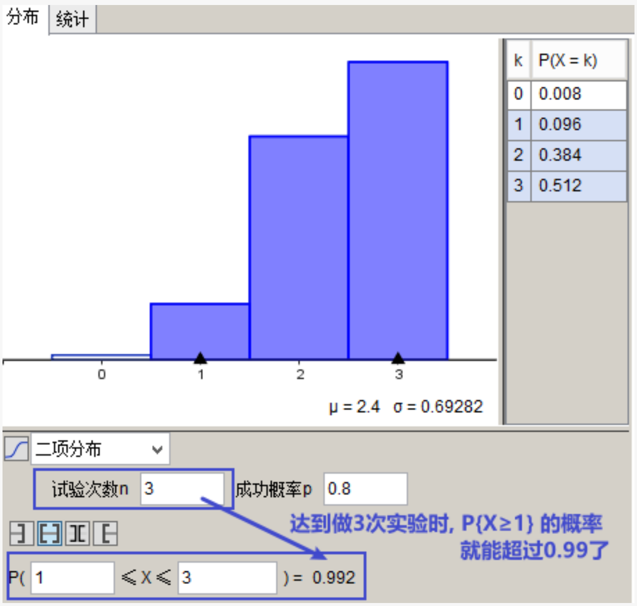
$$\ln 0.01 \geq n \ln 0.2$$

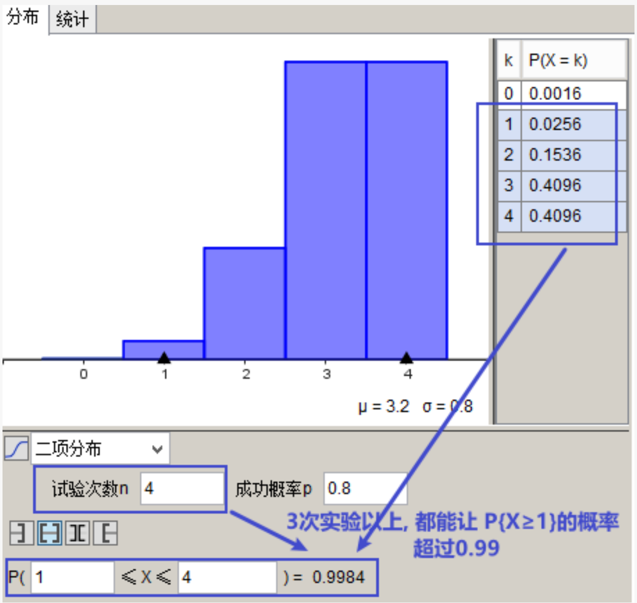
$$n \leq \frac{\ln 0.01}{\ln 0.2}$$

$$n \leq \frac{-4.60517}{-1.60944}$$

$$n \geq 2.86135$$


分布
统计
k P(X = k)
0 0.04
1 0.32
2 0.64
μ = 1.6 σ = 0.56569
二项分布
试验次数n 2
成功概率p 0.8
只做2次实验的话, P {X≥1} 的概率
还没有超过0.99
P( 1 ≤X≤ 2 ) = 0.96


分布
统计
k P(X = k)
0 0.008
1 0.096
2 0.384
3 0.512
μ = 2.4 σ = 0.69282
二项分布
试验次数n 3
成功概率p 0.8
达到做3次实验时, P{X≥1} 的概率
就能超过0.99了
P( 1 ≤X≤ 3 ) = 0.992

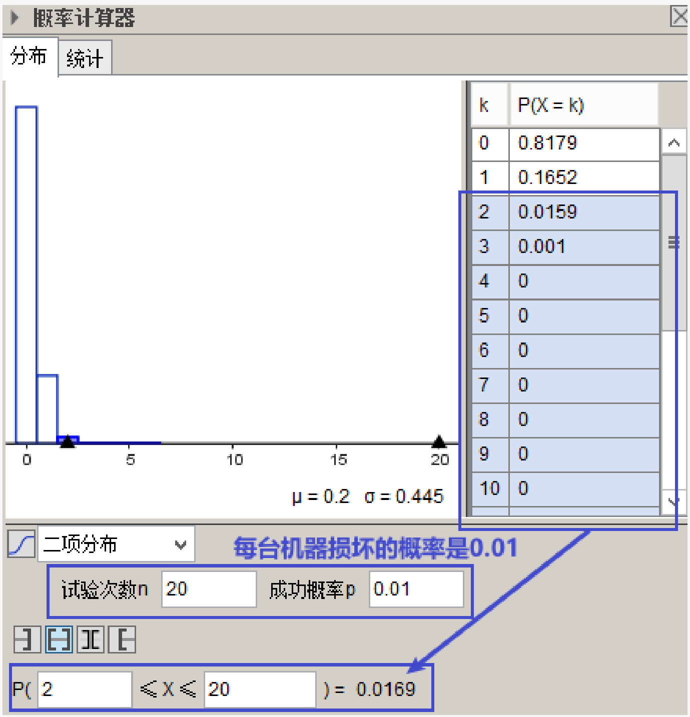

分布
统计
k P(X = k)
0 0.0016
1 0.0256
2 0.1536
3 0.4096
4 0.4096
μ = 3.2 σ = 0.8
二项分布
试验次数n 4
成功概率p 0.8
3次实验以上, 都能让 P{X≥1}的概率
超过0.99
P( 1 ≤X≤ 4 ) = 0.9984

例

每台机器(机床), 会坏的概率是0.01. (即一台机器只有两种结果: 要么处在“正常工作”的状态, 要么处在“损坏”的状态.) 问:

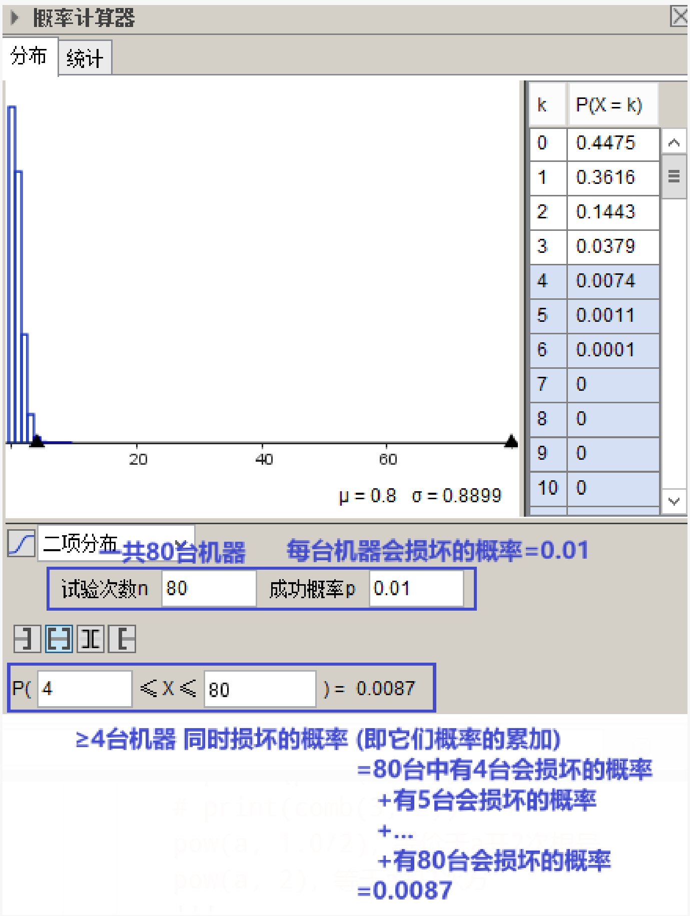
→ 若1个人(维修工)要看护20台机器. 他无法及时维修的概率是多少? 那么对1个人来说, 什么叫做“他无法及时维修”? 就是同时有≥ 2台机器处在“损坏”状态.
我们令:
- n: 代表总机器数. 本处 n=20.
- 随机变量X: 表示“机器处在损坏状态”的台数. 即本处要求的就是 $P(X \geq 2)$ 的概率.

$$P(X \geq 2) = 1 - P(X < 2)$$
$$= 1 - \left[\underbrace{P(X=0)}_{\text{坏0台的概率}} + \underbrace{P(X=1)}_{\text{坏1台的概率}}\right]$$
$$= 1 - \left[\left(\underbrace{C_{20}^{0}}_{\text{0台坏了,即全没坏}} \cdot \underbrace{0.01}_{\text{每台坏的概率}}{}^{0} \cdot \underbrace{0.99}_{\text{每台没坏的概率}}{}^{20-0}\right) + \left(\underbrace{C_{20}^{1}}_{\text{20台中坏了1台}} \cdot 0.01^{1} \cdot 0.99^{20-1}\right)\right]$$
$$= 0.0169$$




→ 若3个人看护80台机器, 问他们无法及时维修的概率? 那就是说, 同时有≥ 4台机器处在“损坏”状态. 即我们要求的是 $P(X \geq 4)$ 的概率.

$$P(X \geq 4) = 1 - P(X < 4)$$
$$= 1 - \left[\underbrace{P(X=0)}_{\text{坏0台的概率}} + \underbrace{P(X=1)}_{\text{坏1台的概率}} + \underbrace{P(X=2)}_{\text{坏2台的概率}} + \underbrace{P(X=3)}_{\text{坏3台的概率}}\right]$$
$$= 1 - \left[\left(C_{80}^{0} \cdot 0.01^{0} \cdot 0.99^{80}\right) + \left(C_{80}^{1} \cdot 0.01^{1} \cdot 0.99^{80-1}\right) + ... + \left(C_{80}^{3} \cdot 0.01^{3} \cdot 0.99^{80-3}\right)\right]$$
$$= 0.0087$$




| | |
|---|---|
| "伯努利分布"(投1次硬币)的"期望值" E(Bernoulli event) | **就表明我们对单个试验的预期结果.** the expected value of the Bernoulli distribution /suggests(v.) which outcome we expect for a single trial. |
| "二项分布"(投n次硬币)的"期望值" E(Binomial Event) | **是我们期望获得特定结果的次数.** the expected value of the Binomial distribution /would suggest(v.) the number of times we expect to get a specific outcome. |

---

# 3 0-1分布 (0-1 distribution) :

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

"0-1分布", 只是"二项分布"的一种特例而已.

0-1分布, 即随机变量X, 只能取 1 和 0. 即:只有两种试验结果. 而且试验只做一次.
- 当 X=1 时, 其概率 P=P
- 当 X=0 时, 其概率 P=1-P

| X= | 1 | 0 |
|---|---|---|
| P= | P | 1-P |

"0-1分布"的公式是:

$\boxed{P\{X=k\}=P^k\cdot(1-P)^{1-k}\quad k=0,1}$ ← 这个公式, 其实是"二项分布"的一种特例.

即: $\begin{cases} 当k=0时,\ P^k\cdot(1-P)^{1-k}=P^0\cdot(1-P)^{1-0}=1-P \\ 当k=1时,\ P^k\cdot(1-P)^{1-k}=P^1\cdot\underbrace{(1-P)^{1-1}}_{1}=P \end{cases}$

例

有一批产品, 次品率是10%. 从中抽1个来检验(即只做一次实验). 我们令:

随机变量$X=\begin{cases} 1 & :表示抽到“合格品” \\ 0 & :表示抽到“次品” \end{cases}$

所以: $\begin{cases} P\{X=0\}=0.1 & \leftarrow 次品率 \\ P\{X=1\}=1-0.1=0.9 \end{cases}$

---

# 4 几何分布

---

# 5 泊松分布

---

# 第二部分 “连续型数据”(PDF) 的分布

# 6 正态分布

---

# 7 指数分布