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

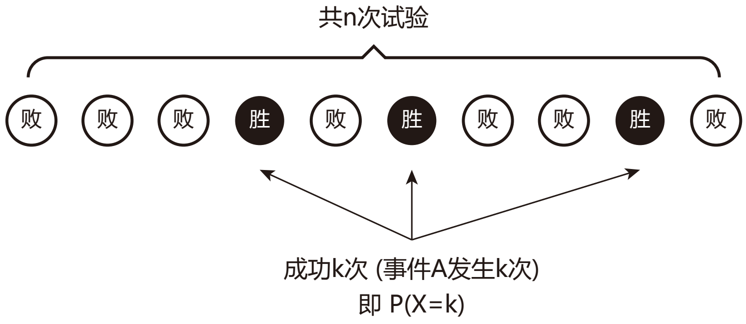
如: 我们将 $X \sim B(10, 0.6)$ 读作: 变量X 遵循10次试验中, 每项试验成功的可能性为0.6的二项分布.
Variable “X” follows a Binomial distribution with 10 trials /and a likelihood of success of 0.6 /on each individual trial.

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:

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

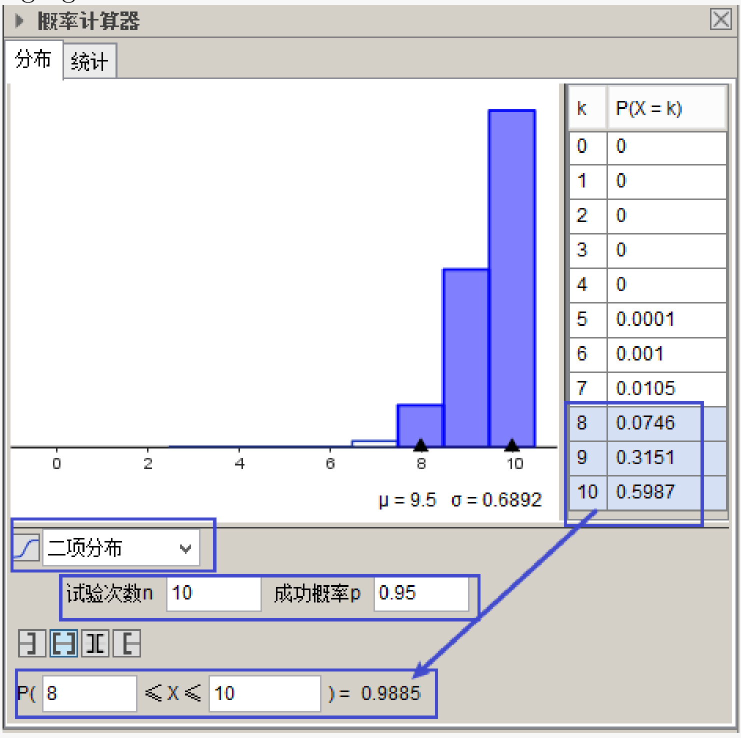
$$=\underbrace{C_{10}^8\cdot 0.95^8\cdot(1-0.95)^2}_{P(X=8)}+\underbrace{C_{10}^9\cdot 0.95^9\cdot(1-0.95)^1}_{P(X=9)}+\underbrace{C_{10}^{10}\cdot 0.95^{10}\cdot(1-0.95)^0}_{P(X=10)}$$

$$=0.0746348+0.315125+0.598737$$

$$=0.988496$$

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]  ← 黄色部分, 只是0-7次成功
            累积分布函数        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.
Out[15]= 0.988496              所以概率上就 = 1- (0到7次的累积概率值)
```

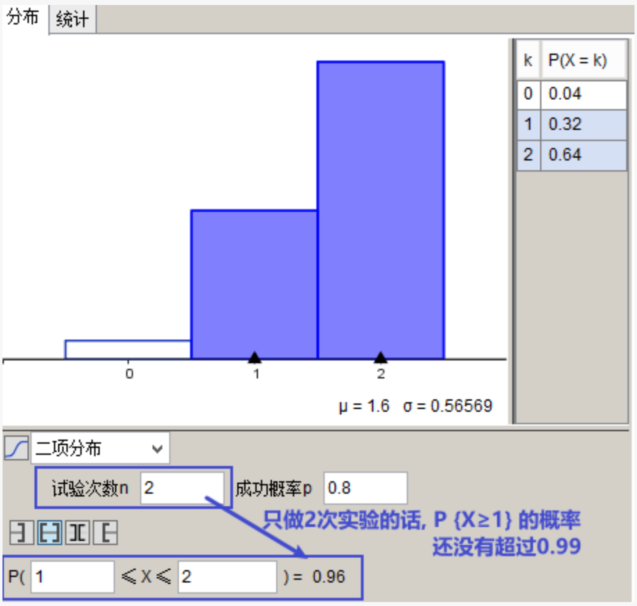
即：我们只要知道单次的成功概率，就能计算n次成功中，会成功k次的概率.

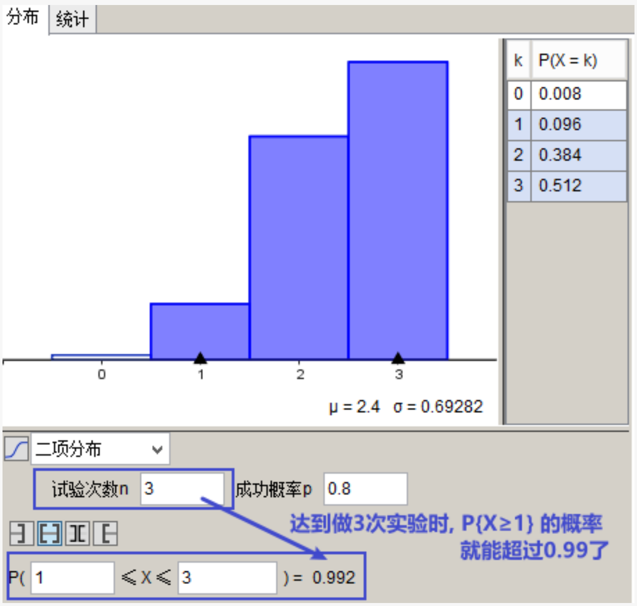
例

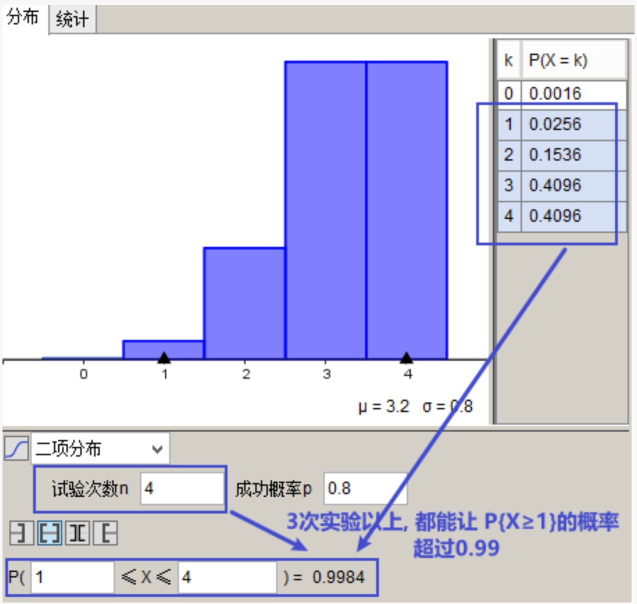
某报警器, 在发生危险时, 成功报警的概率是0.8. 问: 要将报警成功率提高到99%, 至少要安装多少台才行?
我们令:
- n: 表示总共安装的台数.
- X: 表示成功报警的台数.
则, 安装的总n台中, 只要至少有一台能报警 (即 $P(X \geq 1)$), 就成功了.

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

$$\begin{aligned}
&\text{即:}\quad \underbrace{P\left(X \geq 1\right)}_{\text{至少1台报警 的概率}} \geq 0.99 \\
&\underbrace{1 - P\left(X = 0\right)}_{\text{全都没报警 的概率}} \geq 0.99 \\
&1 - \underbrace{C_n^0 \cdot \underbrace{0.8^0}_{0\text{台报警 的概率}} \cdot \underbrace{0.2^n}_{n\text{台没报警 的概率}}}_{\text{总}n\text{台里面,只有0台报警}} \geq 0.99 \\
&1 - 0.2^n \geq 0.99 \\
&1 - 0.99 \geq 0.2^n \\
&\ln 0.01 \geq \ln 0.2^n \quad \leftarrow \text{两边取} \ln \\
&\ln 0.01 \geq n \ln 0.2 \\
&n \leq \frac{\ln 0.01}{\ln 0.2} \\
&n \leq \frac{-4.60517}{-1.60944} \\
&n \geq 2.86135
\end{aligned}$$


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

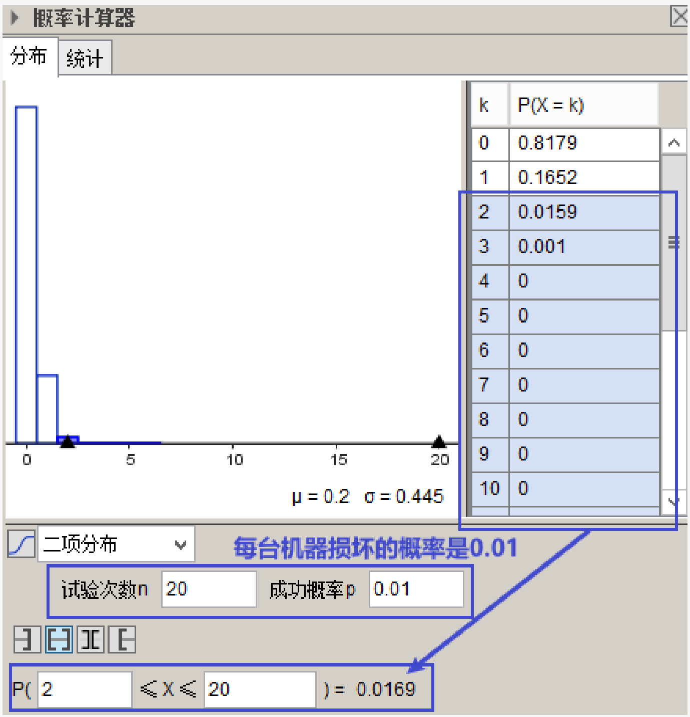

分布
统计
k
P(X = k)
0 0.0016
1 0.0256
2 0.1536
3 0.4096
4 0.4096
μ = 3.2 σ = 0.8
二项分布
试验次数n 4
成功概率p 0.8
3次实验以上, 都能让 P{X≥1}的概率
超过0.99
P(1 ≤X≤ 4) = 0.9984

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:

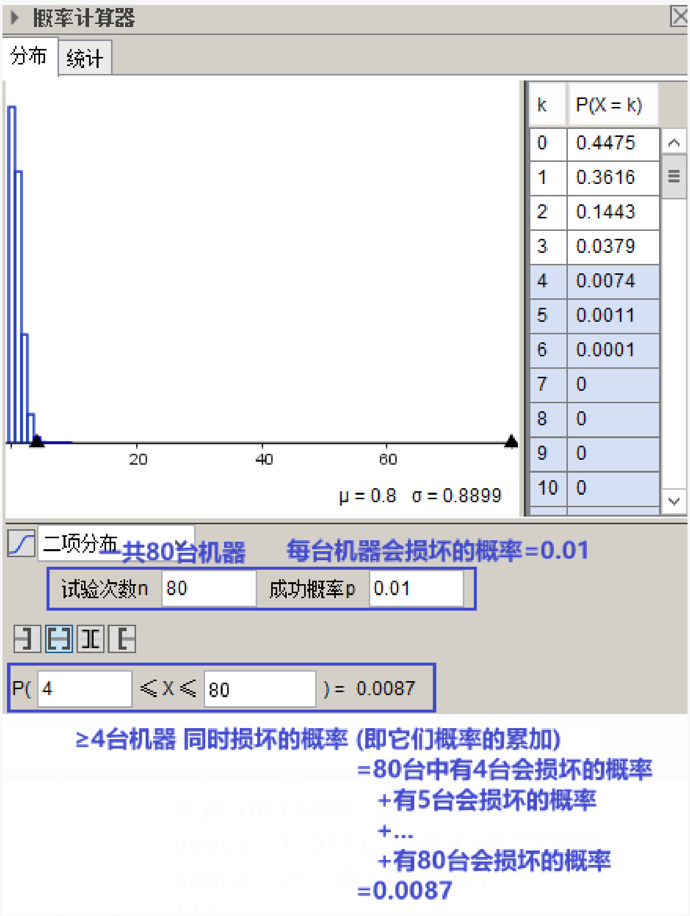
→ 若1个人(维修工)要看护20台机器. 他无法及时维修的概率是多少? 那么对1个人来说, 什么叫做“他无法及时维修”? 就是同时有≥ 2台机器处在“损坏”状态.
我们令:
- n: 代表总机器数. 本处 n=20.
- 随机变量X: 表示“机器处在损坏状态”的台数. 即本处要求的就是 $P(X \geq 2)$ 的概率.

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



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

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

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

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

---

4 几何分布 (Geometric distribution) : $P(X=k)=(1-P)^{k-1}\cdot P$

某事件A, 发生的概率是P, 即$P(A)=P$. 我们把试验重复做很多遍, 使得该事件A, 在第k次试验时首次发生了. 即前面的 k-1 次试验中, 都没发生事件A. 则:

$$\boxed{P(X=k)=\underbrace{(1-P)^{k-1}}_{\text{在前}k-1\text{次试验中,事件}A\text{没发生的概率}}\cdot\underbrace{P}_{(\text{在第}k\text{次试验时})\text{事件}A\text{发生的概率}},\quad k=1,2,\ldots}$$

上面的整体, 就是: 在n次伯努利试验中, “试验k次后, 才得到第一次成功”的机率. 即: “前k-1次皆失败, 第k次才成功”的概率.

上面这个就是“几何分布”的公式. 记作$X\sim G(P)$.

所以, 只要看到“首次发生”这个关键词, 我们就要想到使用“几何分布”来做.

例

射击, 命中率是0.6.

则我们令随机变量X表示“直到首次命中时, 所射击的次数” (即第一次成功时, 是第几次射击).

就有: $P(X=k)=(1-P)^{k-1}\cdot P=(1-0.6)^{k-1}\cdot 0.6,\ k=1,2,3\ldots$

第1次就成功的概率 就是:

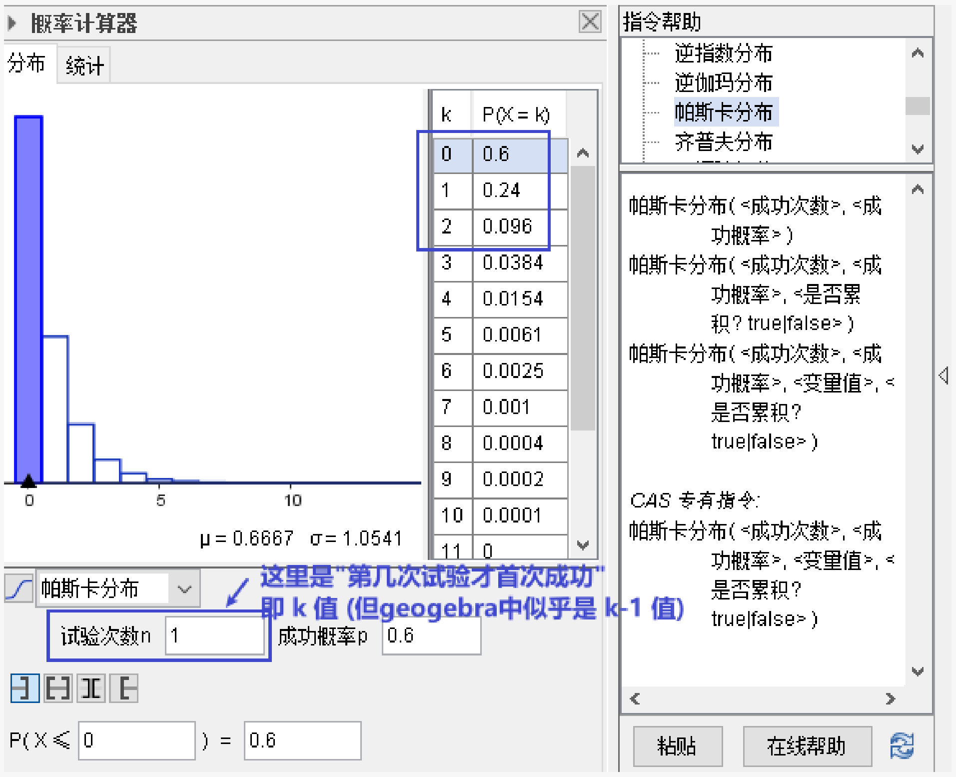
$$P\left(X=\underset{k}{\underset{\smile}{1}}\right)=(1-P)^{1-1}\cdot P=(1-0.6)^{1-1}\cdot 0.6=0.6$$

第2次才首次成功的概率 就是:

$$P(X=2)=(1-P)^{2-1}\cdot P=(1-0.6)^{2-1}\cdot 0.6=0.24$$

第3次才首次成功的概率 就是:

$$P(X=3)=(1-P)^{3-1}\cdot P=(1-0.6)^{3-1}\cdot 0.6=0.096$$



几何分布 Geometric distribution 是"离散型数据"的概率分布.
"几何分布"是"帕斯卡分布"当 r=1 时的特例.
(帕斯卡分布 Pascal distribution 是: 进行多次重复、独立的伯努利试验, 直到出现r次某事件成功为止. 即: 随机变量X表示所需的试验次数. 用 P(X=k)来表示"帕斯卡分布". 即: $P\left(X=k\right)=C_{k-1}^{r-1}\cdot P^{r}\cdot\left(1-P\right)^{k-r},\quad k=r,r+1,...)$

---

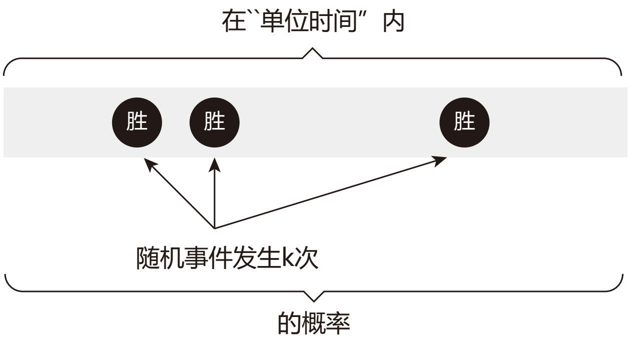
5 泊松分布 (poisson distribution) : $P\left(事X=想\right)=\frac{均^{想}\cdot e^{-均}}{想!}$

"泊松分布"研究的是 : 在"单位时间(或空间)"内, "随机事件发生任意次数"的概率.

即, "泊松分布"是为了解决这样的问题的: 单位时间内, 随机事件发生的次数, 即: 一件事发生的概率P是已知的, 但它的发生与否是随机的. 我们想要求它(即该随机事件)发生k次(或$\geq k$次, $\leq k$次 等问题) 的概率.

当一个随机事件, 以固定的"平均瞬时速率λ" (或称"密度") 随机且独立地出现时, 那么这个事件在"单位时间(面积或体积)"内出现的次数或个数, 就近似地服从"泊松分布".

泊松分布的参数λ：是单位时间(或单位面积)内，随机事件的平均发生次数.



“泊松分布” 的期望和方差, 均为λ.

$$P\{X = \text{你希望“随机事件}A\text{”发生的次数}k\}$$
$$= \frac{(\text{单位时间内,}A\text{的平均发生次数}\lambda)^{\text{你希望}A\text{发生的次数}k}}{(\text{你希望}A\text{发生的次数}k)!} e^{-\lambda}, \quad k = 0, 1, 2, ..., n$$
$$\text{即} = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$$

上面的公式中:
- λ：是单位时间内, 随机事件A的**平均发生次数**.
- k：是**你希望的, 想要的** 随机事件A 发生的次数.

所以, 泊松分布的“概率函数”就是: $\boxed{P\,(X = \text{你想要发生的次数}) = \frac{\text{均}^{\text{想}} \cdot e^{-\text{均}}}{\text{想}!}}$

记作：$\boxed{X \sim P(\lambda)}$ ← 即: 我们用 $Po(\lambda)$ 来表示“泊松分布”. 比如, 我们将$Y \sim Po(4)$ 读作: “变量Y” 遵循 “λ等于4” 的泊松分布.

例

“50年一遇”的大雨, 结果三年中下了两场, 这是怎么回事?
其实“50年一遇”是个数学语言, 它是指: “长期来看”, 这样的大暴雨是平均50年发生一次. 注意关键词“长期”. 长期是多长? 在数学中, 是指“很长很长”的时间段.
所以对“长期”的理解不到位, 就是概率问题的结果“反直觉”的原因.

平均50年发生一次, 可以是: 前4年, 每年都发生一次; 之后的196年一次都没发生. 200除以4，还是50年一次.

所以, 我们更想知道的是: 在任何一段具体的、有限的时间内, 比如5年之内，发生1次大暴雨的概率是多少? 发生2次大暴雨的概率是多少?

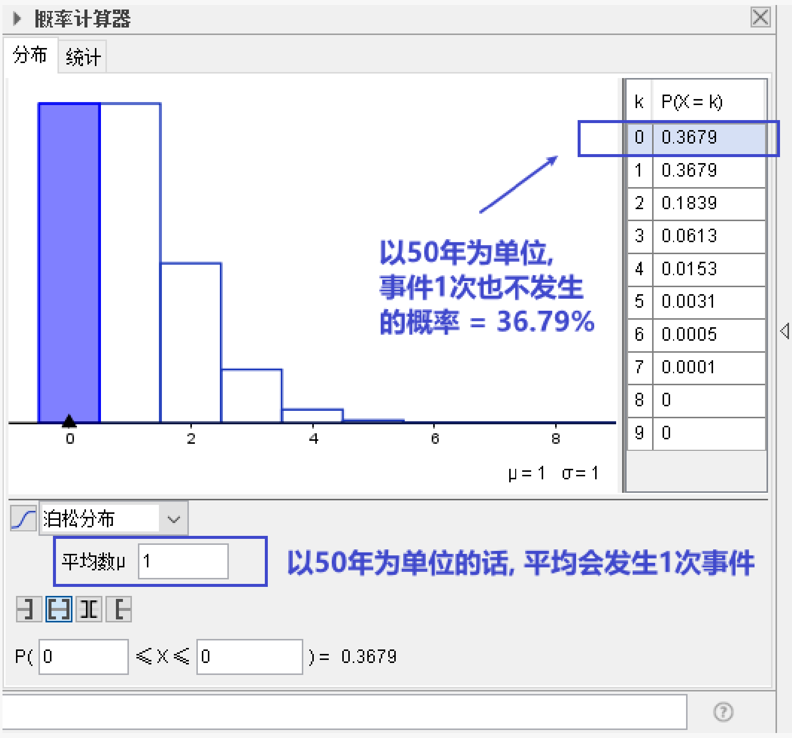
即: 当我们知道了一个随机事件A发生的概率，也知道A发生的概率符合“正态分布”之后, 那么在某一段时间或者空间间隔内,这个随机事件“发生的次数”的概率分布, 是怎样的呢? 这个问题, 就能用“泊松分布”来解决.

泊松分布的公式是: $\boxed{P(X = \text{你希望发生k次}) = \frac{\lambda^k}{k!} e^{-\lambda}}$

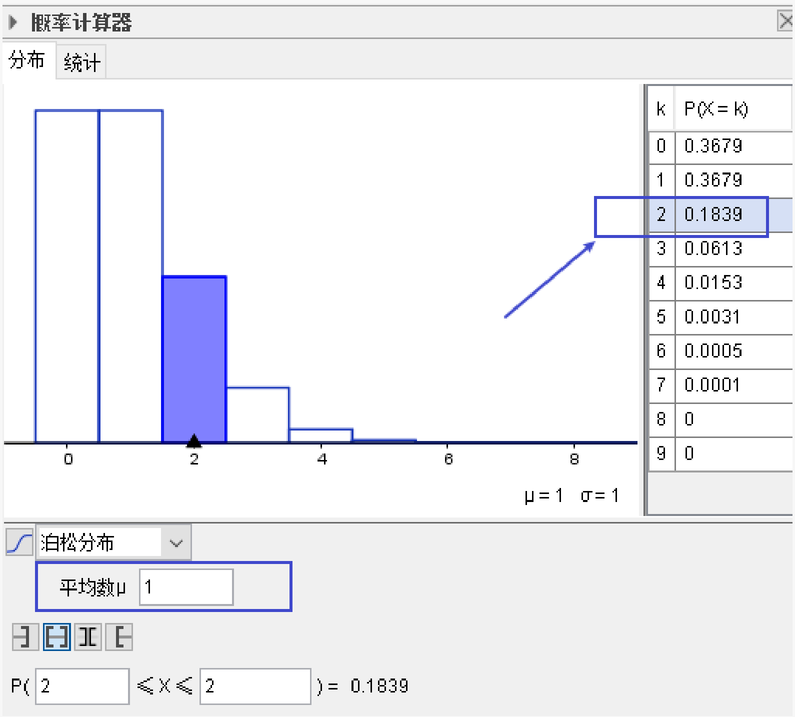
其中,
→ **k：为你希望“随机事件”发生的次数.**
→ λ：**为单位时间内, 随机事件的平均发生次数.** 比如, 50年一遇的大雨：

- 如果以50年为“单位时间”的话, 发生次数就是: 1次. (进一步, 我们可以算出即 : 每年发生$\frac{1}{50}$ 次).
- 如果以100年为单位的话, 发生次数就是 : 100年× 每年$\frac{1}{50}$次 = 2次
- 如果以5年为单位的话, 发生次数就是 : 5年× 每年$\frac{1}{50}$次 = 1/10次.

那么套用“泊松分布公式”, 来算一下, 50年中, 一次上面的大雨也不发生的概率: 即k=0次 :



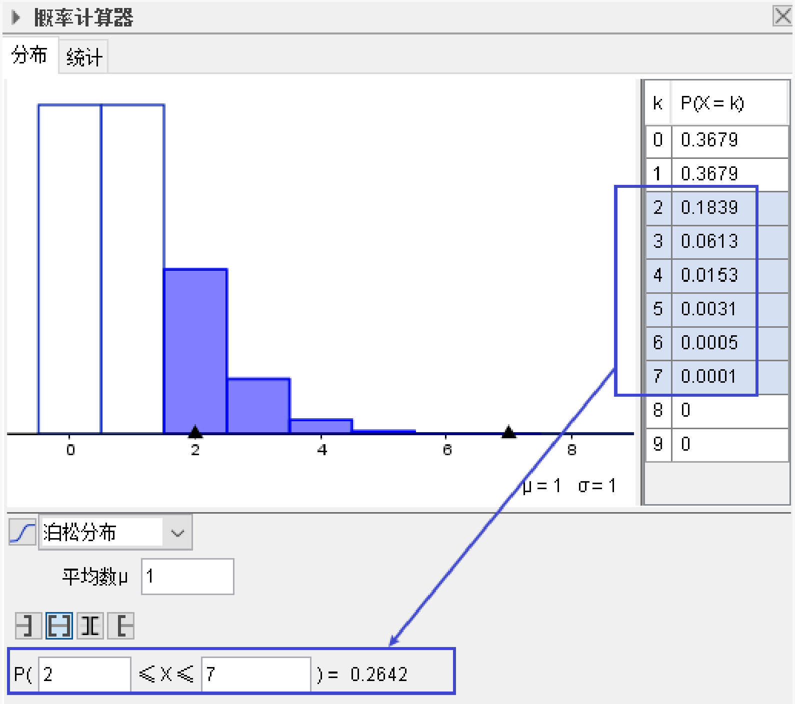
再算一下 K=2, 就是接下来的“50年为单位”的话, 其中发生2次大暴雨的概率, 答案是18%. (下图)



上图的右表中, 表示的就是 : 50年一遇的大雨. 你就以50年为“单位时间段” (即平均会

发生一次这种大雨, 即 μ 或 λ =1), 在这50年中, 你遇到0次, 1次, 2次, ... 这种大雨的真实概率, 是多少?

50年中, 发生2次和2次以上的概率是: 用1 减去发生0次和发生1次的概率. = 1 - (0.3679 × 2) = 26%, 说明这并不是很小的概率事件.

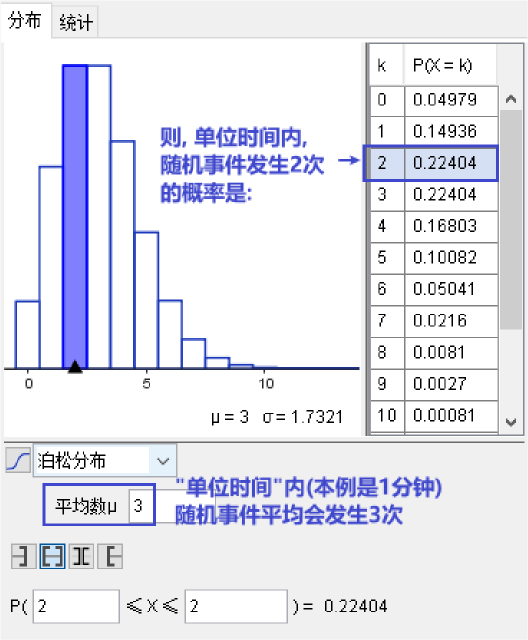


例

某收费站, 平均每分钟通过的车辆为3辆. 问: 1分钟内, 恰有2辆车经过的概率, 是多少? 即:

- 随机事件A : 收费站有车经过.
- λ (单位时间内, 随机事件平均发生的次数. 一般用 λ 或 mean 来表示.) : 本例, 单位时间就是"每分钟", 随机事件A 发生3次.
- k (你希望随机事件发生的次数. 一般用 k 或 x 来表示.) : 本例, 就是 2. (收费站有车经过, 发生2次)

$$P\{X=k\}=\frac{\lambda^k\cdot e^{-\lambda}}{k!}$$
$$P\{X=想\}=\frac{均^{想}\cdot e^{-均}}{想!}=\frac{3^2e^{-3}}{2!}=0.224042$$



例

你创建了一个关于概率的在线课程. 通常，你的学生每天问你大约4个问题, 但昨天他们问了7个. 你想知道昨天这件事, 事实上发生的可能性算多大?

即：

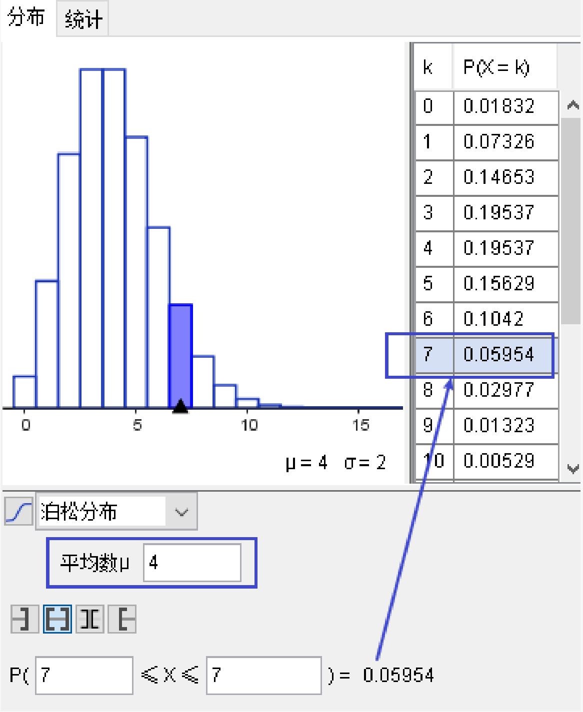
- 随机事件A：是学生提问.
- λ：表示在单位时间内, 随机事件发生的平均次数. 本例就是 =4 (单位时间1天里面, 平均上, 学生提问会发生4次).
- k：你感兴趣的"随机事件发生次数". 本例就是 k=7.

即:

$$P\{X = 想\} = \frac{均^{想} \cdot e^{-均}}{想!}$$

$$P\{X = 7\} = \frac{4^7 \cdot e^{-4}}{7!} = 0.0595404$$

因此, 你收到7个问题的几率, 只有6%.



例

某航空公司, 发生事故(即随机事件A) 的平均值为: 每月 0.05次.

问:

→ 1年内, 发生“0事故”的概率是?

随机事件, 平均每月发生0.05次, 这里的“单位时间”是以“月”为时间段的. 而问题问的是“1年内”, 是以“年”为“单位时间段”的. 所以我们要统一两者的“单位时间”段. 把“月”换算成“年”来做.

即: 事故的“月概率”是 0.05次, 则事故的“年概率”= 0.05 × 12 =0.6.

$$P\{X = \text{想}\} = \frac{\text{均}^{\text{想}} \cdot e^{-\text{均}}}{\text{想}!}$$

$$P\{X = \underbrace{0}_{\text{随机事件在单位时间内,发生0次}}\} = \frac{\underbrace{0.6}_{\text{事故的年概率}}{}^{0} \cdot e^{-0.6}}{0!} = 0.548812$$

→ 1年内, 发生了“1次事故”的概率是?

$$P\{X = 1\} = \frac{0.6^{1} \cdot e^{-0.6}}{1!} = 0.329287$$

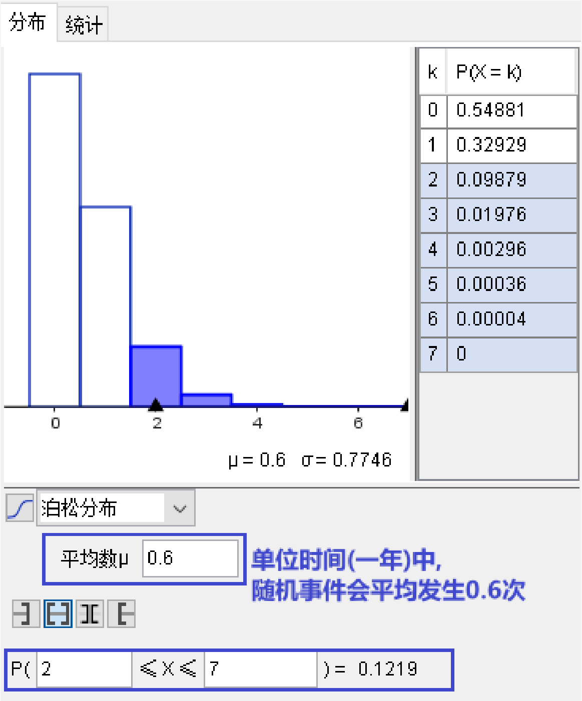
→ 1年内, 发生事故 ≥ 1次的概率是?

$$= P\{X = 2\} + P\{X = 3\} + ...$$

$$= 1 - [P\{X = 0\} + P\{X = 1\}]$$

把随机事件发生0次和1次的情况, 扣除掉后, 剩下的就是超过1次的所有情况了.

$$= 1 - \frac{0.6^{0} \cdot e^{-0.6}}{0!} - \frac{0.6^{1} \cdot e^{-0.6}}{1!} = 0.121901$$



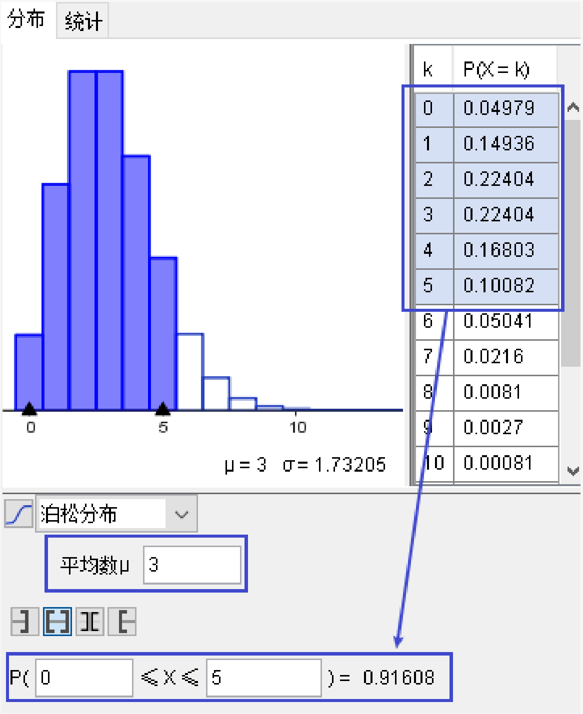
例

某客服工作, 每分钟收到客户来电的次数, 满足 $X \sim P\left(\underbrace{3}_{\text{即}\lambda}\right)$

问: 你“每分钟收到来电不超过5次”的概率.

即:
- 随机事件A : 你收到客户来电.
- λ : 表示在单位时间(本例是1分钟)内, 随机事件发生的平均次数. 本例就是 =3.
- k : 你感兴趣的“随机事件发生次数”. 本例就是 $k \leq 5$.

$$P\{X = 想\} = \frac{均^{想} \cdot e^{-均}}{想!}$$
$$P\{X = k \leq 5\} = \sum_{k=0}^{5} [P\{X = k\}]$$
$$= \frac{3^0}{0!}e^{-3} + \frac{3^1}{1!}e^{-3} + ... + \frac{3^5}{5!}e^{-3} = 0.916$$



5.1 泊松分布的意义 — 为我们开启了“统计推断”的大门

连续2年下大暴雨, 这个现象是否正常? 这个问题的困难在哪儿呢? — 数据太少. 我们没有1000年的降雨资料. 即便有，在长期、无限面前也是个渣渣, 还是太少.

同样，物理学家要研究放射性物质的半衰期, 可绝大多数物质, 衰变期极长, 长到我们没法直接测量. 连一个完整的衰变周期都观测不到, 那怎么办呢? 用“泊松分布”解决.
找一堆铋209原子，统计一下在几个确定的时间间隔中，这堆原子中有多少个发生了衰变? 只要这个数字服从“泊松分布”, 反过来就证明铋209原子的衰变, 也服从“正态分布”. 就可以用“正态分布”来直接计算.

在这些问题的解决中, 统计数据, 和概率论的“概率分布 f(x)”, 就被连在了一起.
而在“泊松分布”之前, 概率和统计是两个不同的学科. “概率”研究“未发生”的随机事件; “统计”描述“已发生”的现实. 那会儿只有描述统计, 没有推断统计. 泊松分布开启了“推断统计”的大门, 第一次把概率和统计连接在一起.

5.2 泊松分布, 其实就是“二项分布”的一种特殊情况. 当二项分布中的 n $\to \infty$; p$\to$ 0 时, 我们就能用“泊松分布”, 来近似该“二项分布”.

当“二项分布”的n很大, 而p很小时, 我们就适合用 “泊松分布”, 来作为“二项分布”的近似. 其中 λ 为np.
通常当$n \geq 20$, $p \leq 0.05$时，就可以用“泊松公式”近似的计算.

即: 当二项分布中的 n $\to \infty$; p$\to$ 0 时, 我们就能用“泊松分布”, 来近似该“二项分布”.
二项分布的“期望值”, 是 $E(X) = np = \lambda$, 所以也就是泊松分布中, $\lambda = np$.

例

某保险公司统计, 其单位时间 (1年)内, 随机事件(每位投保人发生意外死亡)的平均发生概率是 0.002.
现从投保者中抽出1000人 (即单位时间(1年)内, 这1000人里面, 会平均死亡 : 1000人$\times$ 0.002的概率/人 = 2人).
问: 下一年度, 会有1人死亡(而获理赔)的概率?

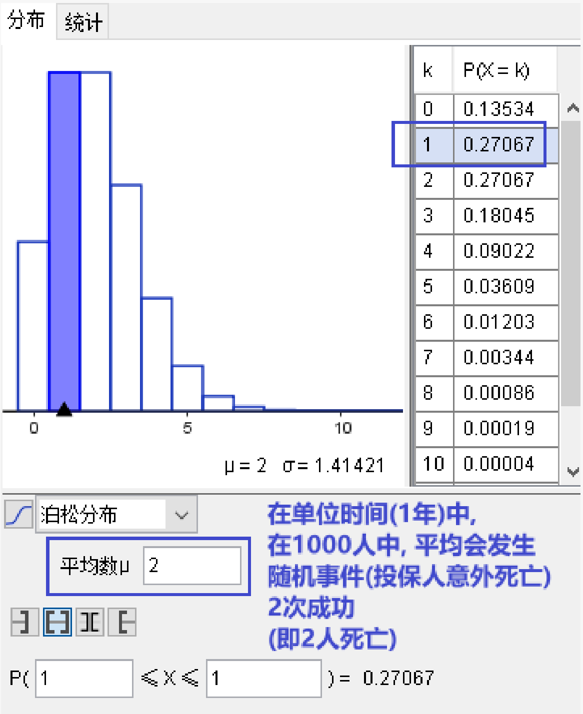
这是一个二项分布 (用来描述 “n次试验中, 事件A恰好发生k次”的概率. 即 $P\{X = k\} = C_n^k p^k (1-p)^{n-k}$). 本例中, n=1000, p=0.002. 即:

$$P\{X = \underbrace{1}_{成功发生死亡事件\ 1人}\} = C_{1000}^1 \cdot 0.002^1 \cdot (1-0.002)^{1000-1} = 0.27067$$

但, 由于 n很大, p很小, $np = 1000 \times 0.002 = 2 = \lambda$, np的值适中, 我们就能用“泊松分布”, 来近似“二项分布”.
即:

$$P\{X = 想\} = \frac{均^{想} \cdot e^{-均}}{想!}$$

$$P\{X = 1\} = \frac{\underbrace{2}_{均值是2人死亡}{}^{1} \cdot e^{-2}}{\underbrace{1}_{你想发生1人死亡}!} = 0.270671$$



例

某病(非传染病), 发病率是 $\frac{1}{1000}$, 某地区有5000人, 问至少2人得病的概率?

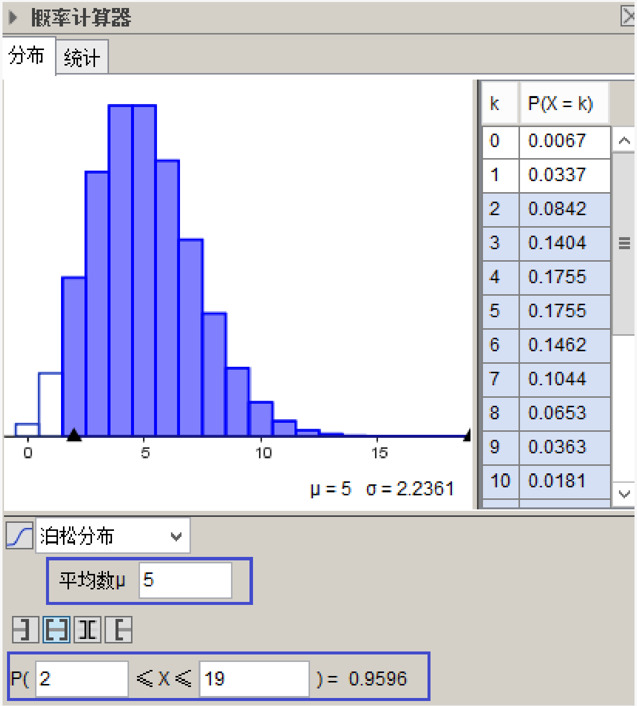
即:

- 随机事件A : 有人得病.
- λ : 表示在单位空间(本例是某地区)内, 随机事件发生的平均次数. 本例就是 = 5000人× 1/1000 = 5人.
- k : 你感兴趣的"随机事件发生次数". 本例就是≥2人(得病).

$$P\{X = 想\} = \frac{均^{想} \cdot e^{-均}}{想!}$$

$$P\{X = k \geq 2\} = 1 - [P(X = 0) + P(X = 1)]$$

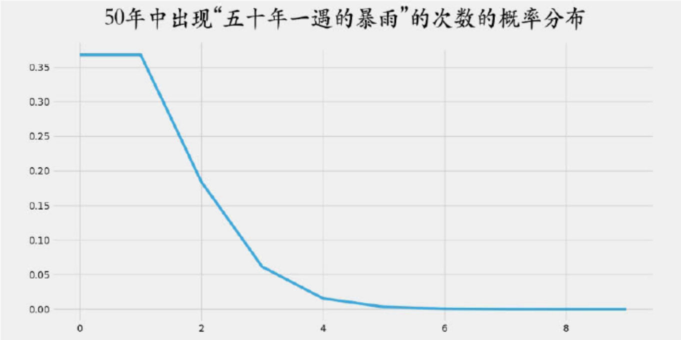
$$= 1 - \frac{5^0}{0!}e^{-5} - \frac{5^1}{1!}e^{-5} = 0.959572$$

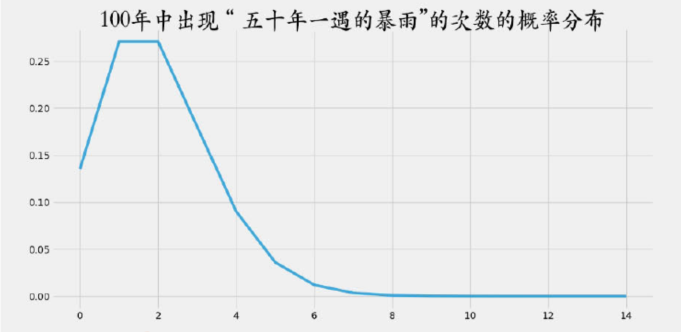


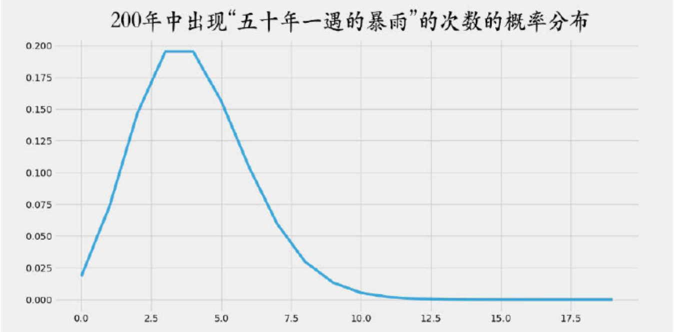
k	P(X = k)
0	0.0067
1	0.0337
2	0.0842
3	0.1404
4	0.1755
5	0.1755
6	0.1462
7	0.1044
8	0.0653
9	0.0363
10	0.0181

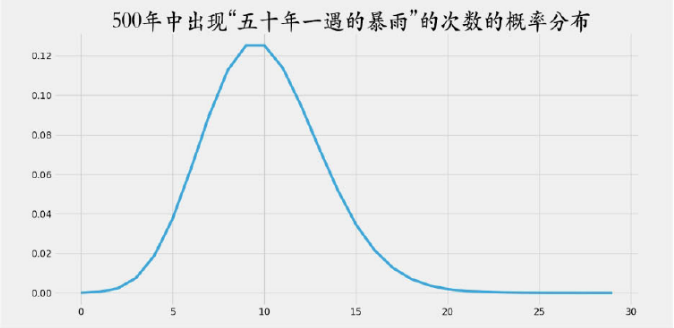
5.3 性质 : 单位时间段越长, "泊松分布"会向"正态分布"看齐

随着我们把"时间单位"拉长, 我们会发现: "泊松分布"的曲线越来越像"正态分布".









5.4 性质：前后两次事件的“发生时间间隔”，无“记忆性”

泊松分布中, 事件对两次发生的时间间隔, 是无”记忆性”的.

即：后一次事件不会记得“距离它前一次发生, 时间隔了多久”. 换言之, 事件之间是相互“独立”的关系.

正因此, 就一定存在一些“短间隔”和“长间隔”，而很难有“一长一短、一长一短”这样有规律的出现. 否则就不叫”无记忆”了.

第二部分 “连续型数据”(PDF) 的分布

6 正态分布

---

7 指数分布