第 1 章　概率论基本概念

🕮　1.1　随机事件&样本空间

❀　1.1.1　引言

随机试验满足 3 个条件：

1) 可以在相同的条件下重复进行；
2) 每次试验可能结果不止一个，且能事先明确所有可能的结果；
3) 每次试验之前不能确定出现哪一个结果。

随机试验发生的结果称为随机事件，一般用 A,B,C 等大写字母表示。

❀　1.1.2　样本空间

随机事件有两个事件较为特殊：在每次试验必然会发生的事件为必然事件(Ω)；必然不会发生的事件为不可能事件(Ø)。

在随机试验中可直接观察到的、最简单的不能再分解的结果称为基本事件。
由若干个基本事件组成的事件称为复合事件。
比如抛骰子结果为 1 的情况为基本事件；结果为偶数的情况为复合事件。

随机试验所有可能的结果组成的集合称为样本空间；样本空间的每一个元素称为样本点。样本点通常是基本事件，也可以是复合事件。

样本空间 Ω = {样本点 ω}
样本空间是必然事件；随机试验的任何一个随机事件都可以看出是样本空间的一个子集。

例 1：将一枚硬币抛两次，正面为 H，反面为 T
样本空间 Ω = {HH, HT, TH, TT}

例 2：连续向一目标射击直至命中
令 ω_i 表示前 i-1 次未能命中，第 i 次命中，i=1,2,3,...
样本空间 Ω = {ω_1, ω_2, ω_3,...}

❀　1.1.3　事件的关系与运算

1) 包含关系：A⊂B
事件 B 包含事件 A，事件 A 发生必然导致事件 B 发生。
特殊情况：如果 A⊂B 且 B⊂A，则 A=B，表示事件 A 与 B 相等。

2) 和事件(并)：A∪B，事件 A 与 B 至少有一个发生。
$A_1, A_2, ..., A_n$ 的和事件写为：

$$\bigcup_{k=1}^{n} A_k$$

3) 积事件(交)：A∩B 或 AB，事件 A 与 B 同时发生。
类似，$A_1, A_2, ..., A_n$ 的积事件写为：

$$\bigcap_{k=1}^{n} A_k$$

特殊情况：如果 A∩B = Ø，表示事件 A 与 B 不会同时发生，称事件 A 与 B 是互不相容的或互斥的。

$A_1, A_2, ..., A_n$ 是随机试验的一组事件，样本空间为 Ω。如果满足：

$$\begin{cases} A_1 \cup A_2 \cup \ldots \cup A_n = \Omega \\ A_i \cap A_j = \emptyset, i \neq j \end{cases}$$

则 $A_1, A_2, ..., A_n$ 是样本空间 Ω 的一个完备事件组。

4) 若 A∩B=Ø 且 A∪B=Ω (A 和 B 构成样本空间 Ω 的一个完备事件组)
称事件 A 与 B 互为逆事件或对立事件。每次试验，事件 A、B 必有一个发生，且仅有一个发生。
记为：$A = \overline{B}$

5) 差事件：A - B，事件 A 发生，B 不发生。

❁ 1.1.4 事件运算规律

1) 交换律：A∪B = B∪A；A∩B = B∩A
2) 结合律：(A∪B)∪C = A∪(B∪C)；(A∩B)∩C = A∩(B∩C)
3) 分配律：A∪(B∩C) = (A∪B)∩(A∪C)；A∩(B∪C) = (A∩B)∪(A∩C)
4) 德摩根律(对偶法则)
$\overline{A \cup B} = \overline{A} \cap \overline{B}$；$\overline{A \cap B} = \overline{A} \cup \overline{B}$

例 1：证明 A - B = A - AB
证：$A - B = A\overline{B} = A(\Omega - B) = A\Omega - AB = A - AB$

例 2：射击 3 次，令 A_i = {第 i 次命中了}, i=1,2,3；B_j = {在 3 次中恰好命中 j 次}, j=0,1,2,3，用 A_i 表示 B_0、B_1、B_3。
$B_0 = \overline{A_1 \cup A_2 \cup A_3} = \overline{A_1} \cap \overline{A_2} \cap \overline{A_3}$
$B_1 = (A_1 \cap \overline{A_2} \cap \overline{A_3}) \cup (\overline{A_1} \cap A_2 \cap \overline{A_3}) \cup (\overline{A_1} \cap \overline{A_2} \cap A_3)$
$B_3 = A_1 \cap A_2 \cap A_3$

🕮 1.2 概率的定义与性质

❁ 1.2.1 概率公理化定义

对样本空间 Ω 每一个事件 A，P(A)表示事件 A 的概率

1) $0 \le P(A) \le 1$ (非负性)
2) $P(\Omega) = 1$ (规范性/正则性)
3) 可加性：若 $A_1, A_2, ...$互不相容，则 $P(\cup A_i) = \Sigma P(A_i)$

❁ 1.2.2 概率的性质

1) $P(\emptyset) = 0$
2) 有限可加性，即上面的可加性也适用于有穷的情况：

$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$

3) 对任意事件 A：$P(A) + P(\overline{A}) = 1$

4) 若 $A \subset B$，则 $P(A) \leq P(B)$且 $P(B - A) = P(B) - P(A)$

5) 广义加法定理：对于任意两个事件 A 和 B，$P(A \cup B) = P(A) + P(B) - P(AB)$

🕮 1.3 古典概率模型

❀ 1.3.1 古典概型定义

若随机试验 E 满足：1) 样本空间有限多个元素；2) 每个基本事件发生的可能性相同，这种试验称为古典概型(等可能概型)。

若事件 A 包含 k (也称为有利场合数)个基本事件，n 为 Ω 包含的基本事件总数，则 P(A)=k/n

例：投掷两个骰子，设 A={和不小于 6}，B={点数相同}，求 P(A)、$P(\overline{A})$、$P(A \cup B)$、$P(A \cap B)$。

基本事件总数$n = 6 \times 6 = 36$

$$P(\overline{A}) = \frac{1+2+3+4}{36} = \frac{10}{36} \quad (11,12,21,13,22,31,14,23,32,41)$$

$$P(A) = 1 - \frac{10}{36} = \frac{26}{36}$$

$$P(B) = \frac{6}{36} \quad (11,22,33,44,55,66)$$

$$P(A \cup B) = \frac{26+2}{36} = \frac{28}{36} \quad (A \text{ 加上 } 11,22 \text{ 的情况})$$

$$P(A \cap B) = \frac{6-2}{36} = \frac{4}{36} \quad (B \text{减去 } 11,22 \text{ 的情况})$$

古典概率计算要统计基本事件的总数，通常要用到排列组合。

❀ 1.3.2 古典概型三个例子

例 1：盒子里有 10 个球，编号 1,2,...,10，从中任取 3 个球。事件 A：3 个球中第二大的数字是 4，求 P(A)。

总共取法：C_{10}^3

3 球最小数可以从 1,2,3 中选取：C_3^1

中间数只能选 4：C_1^1

最大数可以从 5,6,7,8,9,10 中选取：C_6^1

$$P(A) = \frac{C_3^1 \cdot C_1^1 \cdot C_6^1}{C_{10}^3} = \frac{3 \times 1 \times 6}{120} = \frac{3}{20}$$

例 2：袋子中有 a 个黑球，b 个白球，一个一个拿出来，第 k $(1 \leq k \leq a+b)$个球是黑球的概率？

1) 从排列角度考虑：

假设把球拿出来放到 a+b 个坑中，总共(a+b)!种放法；

第 k 个坑需要是黑球有 a 种取法，剩余 a+b-1 个坑随意放，有(a+b-1)!种；
有利场合数为 a×(a+b-1)!

$$P = \frac{a \times (a+b-1)!}{(a+b)!} = \frac{a}{a+b}$$

2) 从组合角度考虑：
从 a+b 个坑挑 a 个坑放黑球，剩余放白球：C_{a+b}^{a}
第 k 个坑放黑球，再挑 a-1 个坑放黑球，剩余放白球，有利场合数：C_{a+b-1}^{a-1}

$$P = \frac{C_{a+b-1}^{a-1}}{C_{a+b}^{a}} = \frac{a}{a+b}$$

这是一个抽签的模型，抽签结果和先后顺序无关，保证公平性。

例 3：将 n 个球随机放入 N $(N \geq n)$个坑中，求每个坑至多有一个球的概率(一个坑可以放多个球)。

样本空间：每次都有 N 个坑选择，总共放 n 次，总共有 N^n 个放法。
有利场合数：N 个坑选 n 个做全排列$C_N^n \cdot n!$

$$P = \frac{C_N^n \cdot n!}{N^n} = \frac{N!}{N^n(N-n)!} = \frac{A_N^n}{N^n}$$

类似问题，如假设一年 365 天，一个班 n $(n \leq 365)$个同学，求至少有两人生日相同的概率。

$$P = 1 - \frac{365 \times 364 \times \ldots \times (365-n+1)}{365^n}$$

```python
def cal(n, N):
    assert 1 <= n <= N
    p = 1
    for i in range(n):
        p *= (N-i)/N
    return 1-p

def main():
    arr = [20, 23, 30, 40, 50, 64]
    for i in arr:
        print('{}\t{:.4f}'.format(i, cal(i, 365)))

if __name__ == '__main__':
    main()
```

结果：

```
20	0.4114
23	0.5073
30	0.7063
40	0.8912
50	0.9704
64	0.9972
```

当一个班有 64 人时，至少有两人生日相同的概率已经很接近 100%。

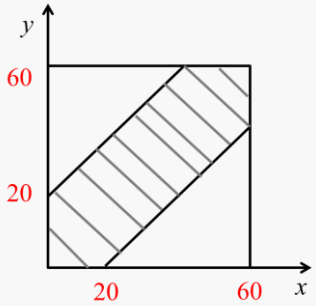
1.3.3 几何概型

古典概型的基本事件个数是有限的，对于无限的情况需要使用几何概型。
几何概型基本事件通常不可计数，只能通过一定的测度(如长度、面积、体积)的比值来表示。

例:(会面问题)两个人约定 5 点到 6 点在某地见面，先到者等待另一个人 20min，如果过时即可离开，求两人见面的概率。

x, y 表示两人到达时间，满足：$|x-y| \le 20$ 两人就能见面

其中 x, y 用相对时间表示都要满足：$0 \le x, y \le 60$



$$P = 1 - \frac{40 \times 40}{60 \times 60} = \frac{5}{9}$$

📖 1.4 条件概率

⚽ 1.4.1 条件概率定义与性质

例：一对夫妇有三个孩子。事件 A：三个都是女孩的概率；事件 B：已知有一个是女孩，剩余两个是女孩的概率。

$$P(A) = \frac{1}{2^3} = \frac{1}{8}$$

$$P(B) = \frac{1}{2^3 - 1} = \frac{1}{7}$$ (排除 3 男的一种情况)

设两个事件 A、B，$P(B) > 0$，事件 B 发生条件下事件 A 发生的条件概率记为：

$$P(A|B) = \frac{P(AB)}{P(B)}$$

试验的基本事件总数为 n，A 包含基本事件数为 a，B 为 b (b>0)，AB 为 k。
如果 B 发生，新的基本事件总数为 b，而 A 发生就是 AB 要发生：

$$P(A|B) = \frac{k}{b} = \frac{k/n}{b/n} = \frac{P(AB)}{P(B)}$$

例 1：盒子里有编号为 1 到 10 的 10 个小球，从中有放回地取两次，两次号码分别为 x 和 y。事件 A：x=4；事件 B：x+y=7。
求 P(A|B)和 P(B|A)。

$$P(A) = P(x = 4) = \frac{1}{10}$$

$$P(B) = P(x + y = 7) = \frac{6}{100}$$

$$P(AB) = P(BA) = P(x = 4, x + y = 7) = P(x = 4, y = 3) = \frac{1}{100}$$

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{1}{6}$$

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{1}{10}$$

例 2：证明$P(A|B)+P(\overline{A}|B)=1$

$$左边=\frac{P(AB)+P(\overline{A}B)}{P(B)}=\frac{P(AB\cup\overline{A}B)}{P(B)}=\frac{P(\Omega B)}{P(B)}=\frac{P(B)}{P(B)}=1$$

条件概率本质上仍然是古典概型，也具有普通概率的性质。

1) $0\le P(A|B)\le 1$
2) $P(\Omega|B)=1$ (Ω 是必然事件)
3) 若 $A_1, A_2, \ldots$互不相容，则：

$$P\left(\bigcup_{i=1}^{\infty}A_i|B\right)=\sum_{i=1}^{\infty}P(A_i|B)$$

1.4.2 乘法公式

将条件概率公式改写得到：

$$P(AB)=P(A|B)\times P(B)=P(B|A)\times P(A)$$

称为乘法公式。其中 P(B)或 P(A)需要>0。

三个事件 A、B、C，且 P(AB)>0，则有：

$$P(ABC)=P(C|AB)\times P(B|A)\times P(A)$$

例：对产品进行三种破坏性试验，产品没有通过第一种试验的概率是 0.3；通过第一种没有通过第二种的概率是 0.2；通过前两种没有通过第三种的概率是 0.1。事件 A：没有通过这三种试验。求 P(A)

设 A_i 为没有通过第 i 种试验 (i=1,2,3)
由题意可知：

$$P(A_1)=0.3$$
$$P(A_2|\overline{A_1})=0.2$$
$$P(A_3|\overline{A_1}\cap\overline{A_2})=0.1$$
$$P(A)=1-P(\overline{A})=1-P(\overline{A_1}\cap\overline{A_2}\cap\overline{A_3})=1-P(\overline{A_1})P(\overline{A_2}|\overline{A_1})P(\overline{A_3}|\overline{A_1}\cap\overline{A_2})$$
$$=1-0.7\times 0.8\times 0.9=0.496$$

1.4.3 全概率公式

例 1：有甲、乙两个箱子，甲中有编号 1~15 的红色卡片，乙中有编号 1~10 的白色卡片。求从甲、乙任选一个、再从中任选一张卡片，编号为偶数的概率。

设事件 A：抽到偶数；B_1：抽到红色；B_2：抽到白色
显然$B_1\cup B_2=\Omega, B_1\cup B_2=\emptyset$。$B_1$、$B_2$ 构成 Ω 的一个完备事件组。

$$P(A)=P(A\Omega)=P(AB_1\cup AB_2)=P(AB_1)+P(AB_2)$$
$$=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)=\frac{1}{2}\times\frac{7}{15}+\frac{1}{2}\times\frac{5}{10}=\frac{29}{60}$$

对于复杂事件 A，将样本空间 Ω 划分为多个不相容的事件 $B_1, B_2, ..., B_n$ (构成一个完备事件组，A 发生的原因或途径)，则：

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)$$

该公式就是全概率公式。

例 2：一批产品 50%来自甲厂、25%来自乙厂、25%来自丙厂；甲乙丙次品率分别为 2%、2%、4%，求所有产品中任取一个是次品的概率。

设 A 为抽到次品；B_1 为抽到产品来自甲厂；B_2 为抽到乙；B_3 为抽到丙

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)$$
$$= 2\% \times 50\% + 2\% \times 25\% + 4\% \times 25\% = 2.5\%$$

1.4.4 贝叶斯公式

根据条件概率和全概率公式可得：

$$P(B_i|A) = \frac{P(B_iA)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}$$

该公式为贝叶斯公式。

B_i 是诸多原因或途径，$P(B_i)$称为先验概率，反应各种原因或途径发生可能性大小，实际中通常是经验的总结或归纳。

$P(B_i|A)$称为后验概率，反应在 A 试验做完后，对各个概率的重新认识。也就是从结果来探讨不同原因的可能性大小。

例：用某个方法诊断肺癌，事件 A：被检验者患有肺癌，B：结果为阳性。已知 $P(B|A) = 0.95, P(\overline{B}|\overline{A}) = 0.95, P(A) = 0.005$，求 P(A|B)。

题意：人群中患肺癌概率是 0.005。患肺癌检出阳性、没肺癌检出阴性的概率都是 0.95，说明准确率比较高。

由贝叶斯公式得：

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\overline{A})P(B|\overline{A})} = \frac{0.005 \times 0.95}{0.005 \times 0.95 + 0.995 \times 0.05} = 0.087$$

结果检出阳性而确实患肺癌的概率只有 8.7%。

1.5 独立性

1.5.1 两个事件的独立性

A，B 是试验 E 的两个事件，一般而言 B 的发生对 A 有影响，即 $P(A|B) \neq P(A)$。如果满足 $P(A|B)=P(A)$，即：

$$P(AB) = P(A|B)P(B) = P(A)P(B)$$

称事件 A、B(相互)独立。

注：事件 A、B 相互独立和 A、B 互不相容不能同时成立。

推论：若 A、B 独立，则A 与$\overline{B}$，$\overline{A}$与 B，$\overline{A}$与$\overline{B}$ 也相互独立。

例：机器甲的次品率是 0.05，乙的次品率是 0.04.从甲乙各取一件，求：

1) 两件都是次品的概率 p1；
2) 至少一件次品的概率 p2；
3) 恰好一件次品的概率 p3.

设事件 A：抽到甲的是次品；B：抽到乙的是次品。

$$p1 = P(AB) = P(A)P(B) = 0.05 \times 0.04 = 0.002$$

$$p2 = P(A \cup B) = 1 - P(\overline{A} \cap \overline{B}) = 1 - P(\overline{A})P(\overline{B}) = 1 - 0.95 \times 0.96 = 0.088$$

$$p3 = P(A\overline{B}) + P(\overline{A}B) = P(A)P(\overline{B}) + P(\overline{A})P(B) = 0.05 \times 0.96 + 0.95 \times 0.04 = 0.086$$

1.5.2 多个事件的独立性

三个事件的 ABC 如果满足：

$$P(AB) = P(A)P(B)$$
$$P(BC) = P(B)P(C)$$
$$P(AC) = P(A)P(C)$$
$$P(ABC) = P(A)P(B)P(C)$$

则事件 ABC 相互独立。

对 n (n≥2)个事件 $A_1, A_2, ..., A_n$，如果对任意 2,3, ..., n 个事件的积事件概率都等于各事件概率之积，则成这 n 个事件相互独立。

例：掷两个骰子，事件 A：第 1 个骰子是奇数；B：第 2 个为奇数；C：两者之和为奇数。分析三个事件的独立性。

$$P(A) = P(B) = P(C) = \frac{1}{2}$$

$$P(AB) = P(BC) = P(AC) = \frac{1}{4} = P(A)P(B) = P(B)P(C) = P(A)P(C)$$

$$P(A)P(B)P(C) = \frac{1}{8}$$

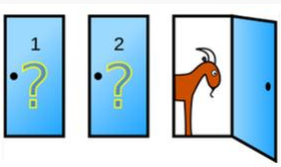
$$P(ABC) = 0 \neq P(A)P(B)P(C)$$

说明 A、B、C 两两独立，并不是相互独立。

1.6 蒙提霍尔三门问题

游戏规则：

1) 有三扇关闭的门，其中一扇后面有一辆汽车，另外两扇门后面各藏有一只山羊。选中后面有车的那扇门可赢得该汽车。
2) 当参赛者选定了一扇门，但未去开启它的时候，知道门后情形的节目主持人开启剩下两扇门的其中一扇，露出其中一只山羊。
3) 主持人其后会问参赛者要不要换另一扇仍然关上的门。



问题是：换另一扇门会否增加参赛者赢得汽车的机率？

该问题也称为蒙特霍尔悖论。因为很具迷惑性，违反大多人的直觉。

如果严格按照上述的条件，即主持人清楚地知道，自己打开的那扇门后是羊，那么换门赢得汽车的几率是 2/3；不换门赢得汽车的几率是 1/3。

```
from random import randint

def monty_hall(num, is_change):
    car = randint(1, 3)
    # 选中但是改变，或者没选中且不改变，则没有得到汽车
    if (car == num and is_change) or (car != num and not is_change):
        return False
    return True

def test(flag):
    # flag: 0 表示不改变,1 表示改变,其他表示随机
    n = 100000  # 运行次数
    win = 0
    for i in range(n):
        num = randint(1, 3)
        is_change = flag if flag in [0, 1] else randint(0, 1)
        win += 1 if monty_hall(num, is_change) else 0
    return win/n

def main():
    dct = {0: '不改变', 1: '改变', 2: '随机'}
    for k, v in dct.items():
        print('{}: {:.4f}'.format(v, test(k)))

if __name__ == '__main__':
    main()
```

结果：

```
不改变: 0.3341
改变: 0.6654
随机: 0.5034
```

使用简单的 Python 程序可以帮助理解较为迷惑性的条件概率问题。

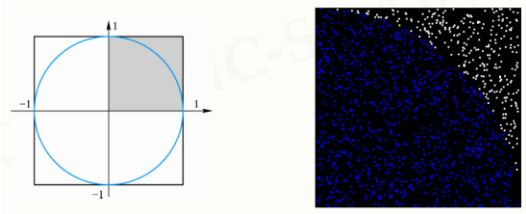
🕮 1.7 蒙特卡罗方法初步

也称统计模拟方法，使用(伪)随机数解决很多计算问题。

比如计算圆周率，在边长为 1 的正方形区域随机撒点 n 个，距离圆心≤1 的点个数为 m。

$$\frac{S_{扇形}}{S_{正方形}} = \frac{\pi}{4} = \frac{m}{n} \quad \rightarrow \quad \pi = \frac{4m}{n}$$

撒点个数越多，得到的结果越精确。



```
from random import random
from time import perf_counter

def cal_pi():  # 蒙特卡罗方法
```

```
    darts = 1000*1000  # 随机撒点个数
    hits = 0  # 记录落在扇形中的点数
    start = perf_counter()
    for i in range(darts):
        # 随机撒点
        x, y = random(), random()
        distance = (x**2+y**2)**0.5  # 该点到圆心距离
        if distance <= 1:
            hits += 1
    pi = hits/darts*4
    end = perf_counter()
    print('运行时间: {:.2f}s'.format(end-start))
    return pi

if __name__ == '__main__':
    print(cal_pi())
```

结果：

```
运行时间: 1.04s
3.143004
```

🕮 1.8 随机测试初步

某个应用需要一个计算多项式的程序，两个攻城狮实现了两个程序 F(x)和 G(x)，问如何验证 F(x)=G(x)?

假设它们最高阶(x 的最高次)为 d，随机算法首先从[1, 100d]均匀随机选择一个整数 t 进行测试。

如果 F(t)=G(t)，判断两个程序相同，但 F(x)≠G(x)，此时产生误判。

也就是 t 是 F(x)-G(x)=0 的根时，产生误判。

因为 F(x)-G(x)最高次不大于 d，F(x)-G(x)=0 最多 d 个根。

误判概率为：$\mathrm{P} = d/(100d) = 1/100$

如何降低误判概率？

① 增加数据范围，如从[1, 1000d]随机选择一个整数 t 进行测试，此时误判概率为 1/1000；

② 增加测试强度，每次从[1, 100d]随机选择一个整数 t，重复 k 次随机测试：

1)当 F(t)=G(t)，选择新的 t 进行下一轮测试；

2)当 F(t) ≠G(t)，结束测试。

因为每次测试都是独立事件，则 k 次测试的误判为 $1/100^k$(呈指数级降低)。

抽样方法分为有放回抽样和无放回抽样。之前讨论的是有放回抽样。

使用无放回抽样的误判概率：

$$\mathrm{P} \le \prod_{i=1}^{k} \frac{d-(i-1)}{100d-(i-1)} \le \frac{1}{100^k}$$

随机测试使用无放回抽样的误判概率比有放回抽样低，但有放回抽样的实现更简单和实用。

当 k>d 次随机测试，能确保误判概率为 0；随机测试的总复杂度为 O(d^2)(每一轮为 O(d))。但如果 d 特别大，这样做不合算。

第 2 章　随机变量及其分布

🕮 2.1　随机变量

随机变量取值随机会而定，是试验结果的函数。其反面是确定性变量。
随机试验样本空间为 Ω = {ω}，X=X(ω)是定义在样本空间 Ω 上的实值单值函数，称 X=X(ω)为随机变量。

示例：
① 掷一枚硬币：

$$X=\begin{cases}1, & \text{正面}\\ 0, & \text{反面}\end{cases}$$

② 检测 m 件产品，次品数 X = 0, 1, 2, ..., m
③ 人的身高：$H \in (-\infty,+\infty)$。近似处理方法，离散问题当作连续问题考虑。
④ 射击击中位置：以靶心为圆点，靶半径为 r：X=(x, y), x≤r, y≤r

🕮 2.2　离散型随机变量

如果随机变量的可能取到的值有限多个或可列无限多个，称为离散型随机变量。
离散型随机变量 X，其所有可能取值为 a_i，X 取到 a_i 的概率：

$P(X = a_i) = p_i, i = 1, 2, 3, ...$

如果满足：

1) $p_i \geq 0, i = 1,2,3 ...$
2) $\sum p_i = 1$

则称离散型随机变量 X 的分布列或分布律。使用表格的形式表示更加直观：

X	a_1	a_2	...	a_n	...
p_i	p_1	p_2	...	p_n	...

🕮 2.3　常用三种离散型随机分布

重复独立试验：每次试验相互独立且相应概率保持不变。
贝努里(Bernoulli)试验：只有两个结果的重复独立试验。

如$P(A) = p \ \ (0 < p < 1)$，则$P(\overline{A}) = 1 - p$

试验重复独立进行 n 次，称为 n 重贝努里试验。

① 两点分布
只进行 1 次贝努里试验：

X	0	1
p_i	1 - p	p

称 X 服从以 p 为参数的两点分布或(0-1)分布。
常见例子为抛硬币、性别统计、检查产品质量是否合格等。

② 二项分布
进行 n 重贝努里试验，事件 A 出现 k 次的概率。也就是 k 次试验 A 事件发生，n - k 次试验 A 没有发生。

比如前 k 次试验 A 发生，后 n - k 次 A 没有发生的概率为：$p^k(1-p)^{n-k}$

n 次试验选择 k 次表示 A 发生，总共取法是C_n^k种。
因为它们两两互不相容，所以 n 次试验 A 发生 k 次的概率为$C_n^k p^k (1-p)^{n-k}$
记$q = 1 - p$，有：

$$\mathrm{P(X = k)} = C_n^k p^k q^{n-k} \ , k = 0,1,2,\dots,n$$

X	0	1	…	k	…	n
p_k	$C_n^0 p^0 q^n$	$C_n^1 p^1 q^{n-1}$	…	$C_n^k p^k q^{n-k}$	…	$C_n^n p^n q^0$

$C_n^k p^k q^{n-k}$是二项式$(p+q)^n$的展开式的一项。
所以：

$$\sum_{k=0}^{n} p_k = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^n = 1$$

满足离散型随机分布要求。

由于是二项展开式的形式，故称为二项分布。
称随机变量 X 服从参数为 n, p 的二项分布，记为 X~$b(n, p)$;
每一项记为：$b(k; n, p) = C_n^k p^k q^{n-k}$
使 b(k; n, p)取最大值的项 b(m; n, p)称为中心项；m 为最可能成功的次数。

例 1：大批元件有 10%是次品，抽取 20 个，都是合格的概率。
不是放回抽样，但是因为总数很大，抽取的数目相对而言很少，所以可以近似认为是有放回抽样。相当于做了 n=20 重贝努里试验。
$\mathrm{p = b(20,20,0.9) = C_{20}^{20} 0.9^{20} = 0.1216}$

例 2：一次抽卡抽到 UR 的概率 0.01，500 次单抽，最可能抽到多少张 UR，并求相应的概率。
最可能抽到 5 张 UR，概率为：
$\mathrm{p = b(5,500,0.01) = C_{500}^{5} 0.01^5 0.99^{495} = 0.1764}$

③ 泊松(Poisson)分布
随机变量 X 所有可能取值为 0,1,2,...(非负整数)，各取值概率为：

$$\mathrm{P(X = k)} = \frac{\lambda^k e^{-\lambda}}{k!} \ , k = 0,1,2,\dots$$

其中 λ>0 是常数，称 X 服从参数为 λ 的泊松分布，记为X~π(λ)。

根据泰勒级数：

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

得：

$$\sum_{k=0}^{\infty} P(X = k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1$$，满足概率分布的要求

泊松分布可以作为二项分布的近似(逼近)。
泊松定理：λ>0 是常数，n 是任意正整数，设 np_n=λ，则对于任一固定非负整数 k，有：

$$\lim_{n\to\infty} C_n^k {p_n}^k (1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

也就是当 n 很大、p 很小时，以 n, p 为参数的二项分布可近似于以 λ=np 的泊松分布。

例：某个产品次品率为 0.1%，求 1000 个产品中至少有 2 个是次品的概率。

使用二项分布计算

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$$
$$= 1 - C_{1000}^0 0.999^{1000} - C_{1000}^1 0.001 \cdot 0.999^{999} \approx 0.264241087$$

使用泊松分布近似，λ=1

$$P(X \geq 2) = 1 - \frac{1^0 e^{-1}}{0!} - \frac{1^1 e^{-1}}{1!} = 1 - e^{-1} - e^{-1} \approx 0.264241118$$

当 n≥20、p≤0.05 时，使用泊松分布作为二项分布的近似，效果颇佳。

🕮 2.4 分赌本问题

1654 年，一个职业赌徒向法国数学家帕斯卡提出了一个令他苦恼已久的分赌本问题：
甲乙两赌徒赌技相同，各出赌注 100 法郎，每局无平局。约定谁先赢 3 局则得到全部 200 法郎的赌本。
当甲赢了 2 局，乙赢了 1 局时，因故中止赌博，问这 200 法郎怎么分才算公平？
赌徒的不同理解：
1) 谁都没赢，各得 100 法郎；2) 甲得 2/3，乙得 1/3。

帕斯卡与法国数学家费马就此问题展开了信件♂讨论。

假设比赛继续，如果乙要赢，需要赢下第 4、5 局，概率为 1/4；而甲要赢，只需赢第 4 局或输了第 4 局而赢下第 5 局，概率为 1/2+1/4=3/4。
所以甲赢的概率 3/4，分得 150 法郎；乙赢的概率 1/4，分得 50 法郎。

🕮 2.5 分布函数

对于非离散型(连续型)随机变量：
1) 其可能取值不能一一列举，因而不能使用分布律描述。
2) 连续型随机变量在任一点处的概率都是 0。
3) 更感兴趣的是随机变量在某个区间的概率，如取值落在区间$(x_1,x_2]$的概率 $P\{x_1<X\leq x_2\}$。

设 X 是一个随机变量，x 是任意实数，

$$F(x) = P\{X \leq x\}, -\infty < x < +\infty$$

函数 $F(x)$称为 X 的分布函数。

对于任意实数 $x_1,x_2(x_1<x_2)$，有：

$$P(x_1 < X \leq x_2) = P\{X \leq x_2\} - P\{X \leq x_1\} = F(x_2) - F(x_1)$$

分布函数完整地描述了随机变量的统计规律性。通过分布函数，可以使用数学分析(微积分)的方法研究随机变量。

分布函数 F(x)的性质：

1) F(x)单调不减；

$$\mathrm{P}(x_1 < \mathrm{X} \le x_2) = \mathrm{F}(x_2) - \mathrm{F}(x_1) \ge 0$$

2) $0 \le \mathrm{F}(\mathrm{x}) \le 1$，且

$$\mathrm{F}(-\infty) = \lim_{\mathrm{x}\to-\infty} \mathrm{F}(\mathrm{x}) = 0 \quad \text{(趋于不可能事件)}$$

$$\mathrm{F}(+\infty) = \lim_{\mathrm{x}\to+\infty} \mathrm{F}(\mathrm{x}) = 1 \quad \text{(趋于必然事件)}$$

3) F(x+0) = F(x)，即 F(x)是右连续的。

例 1：将一枚硬币连续抛两次，X 表示正面出现的次数，求 X 的分布函数 F(x)及 P{0<X≤1}和 P{1≤X≤2}。

X 的分布律为：

X	0	1	2
p	1/4	1/2	1/4

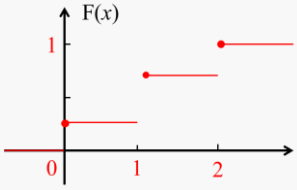
分布函数为：

$$\mathrm{F}(\mathrm{x}) = \begin{cases} 0, & \mathrm{x} < 0 \\ \frac{1}{4}, & 0 \le \mathrm{x} < 1 \\ \frac{3}{4}, & 1 \le \mathrm{x} < 2 \\ 1, & x \ge 2 \end{cases}$$

$$\mathrm{P}\{0 < \mathrm{X} \le 1\} = \mathrm{F}(1) - \mathrm{F}(0) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$$

$$\mathrm{P}\{1 \le \mathrm{X} \le 2\} = \mathrm{F}(2) - \mathrm{F}(1) + \mathrm{P}\{x = 1\} = 1 - \frac{3}{4} + \frac{1}{2} = \frac{3}{4}$$

F(x)是一个阶梯型的曲线：



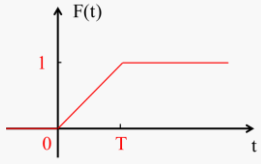
在 0,1,2 处有跳跃点。

例 2：某脉冲信号在时间区间(0, T)内随机出现，出现时刻记为 t。事件{t_1<t≤t_2}的概率为：

$$\mathrm{P}\{t_1 < \mathrm{t} \le t_2\} = \frac{1}{T}(t_2 - t_1), 0 < t_1 < t_2 < T$$

求 X(t)=t 的分布函数。

$$F(t)=\begin{cases}0, & t<0\\ \dfrac{t}{T}, & 0\le t<T\\ 1, & t\ge T\end{cases}$$



F(t)处处连续

📖 2.6 连续型随机变量及其概率密度函数

设随机变量 X 的分布函数为 F(x)，如果存在非负函数 $f(x)$，对任意实数 x 有：

$$F(x)=\int_{-\infty}^{x} f(t)dt$$

则称 X 为连续型随机变量，$f(x)$为 X 的概率密度函数。

概率密度函数 $f(x)$性质：

1) $f(x)\ge 0$

2) $\int_{-\infty}^{+\infty} f(x)dx=1$

3) 对于任意实数 $x_1, x_2(x_1\le x_2)$：

$$P\{x_1<X\le x_2\}=F(x_2)-F(x_1)=\int_{x_1}^{x_2} f(x)dx$$

4) 若 $f(x)$在点 x 处连续，则 $F'(x)=f(x)$

5) 对于连续型随机变量 X，取任意实数值 a 的概率都是 0

$$P\{X=a\}=\int_{a}^{a} f(x)dx=0$$

也就是若事件 A 是不可能事件，则 P(A)=0；反之若 P(A)=0，并不一定表示 A 是不可能事件。

约定：随机变量的概率分布，离散型→分布律；连续型→概率密度函数。

📖 2.7 常用的三种连续型随机分布

① 均匀分布

若连续型随机变量 X 的概率密度满足：

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b\\ 0, & \text{其他}\end{cases}$$

则称 X 在区间(a, b)上服从均匀分布，记为 X~U(a, b)

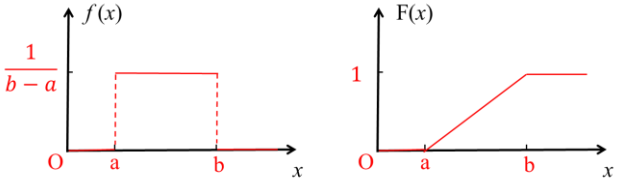
当$x<a$时，$F(x)=\int_{-\infty}^{x} 0dt=0$

当$a \le x < b$时，$\mathrm{F}(x) = \int_{-\infty}^{a} 0dt + \int_{a}^{x} \frac{1}{b-a} dt = \frac{x-a}{b-a}$

当$x \ge b$时，$\mathrm{F}(x) = \int_{-\infty}^{a} 0dt + \int_{a}^{b} \frac{1}{b-a} dt + \int_{b}^{x} 0dt = 1$

所以，X 的分布函数为：

$$\mathrm{F}(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b \end{cases}$$



② 指数分布

若连续型随机变量 X 的概率密度满足：

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{其他} \end{cases} \text{ 或 } f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & x > 0 \\ 0, & \text{其他} \end{cases}$$

其中 λ>0 (λθ=1)为常数，则称 X 服从参数 λ 的指数分布。

当 x<0 时，$\mathrm{F}(x) = \int_{-\infty}^{x} 0dt = 0$

当 x≥0 时，$\mathrm{F}(x) = \int_{-\infty}^{0} 0dt + \int_{0}^{x} \lambda e^{-\lambda t} dt = (-e^{-\lambda t})\Big|_{0}^{x} = e^{0} - e^{-\lambda x} = 1 - e^{-\lambda x}$

所以，X 的分布函数为：

$$\mathrm{F}(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

指数分布的一个有趣的性质：对任意 $s, t > 0$，有：

$$\mathrm{P}\{\mathrm{X} > \mathrm{s} + \mathrm{t} \mid \mathrm{X} > \mathrm{s}\} = \frac{P\{(X > s+t) \cap (X > s)\}}{P\{X > s\}} = \frac{P\{X > s+t\}}{P\{X > s\}}$$

$$= \frac{1 - P\{X \le s+t\}}{1 - P\{X \le s\}} = \frac{1 - F(s+t)}{1 - F(s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - F(t)$$

$$= P\{X > t\}$$

该性质称为无记忆性。比如 X 指某个产品的寿命，如果使用了 s 小时，在此条件下再使用 t 小时的概率和从头算能使用 t 小时的概率相等，也就是产品对已经使用过 s 小时没有记忆。所以，指数分布也称寿命分布。

③ **正态分布**

若连续型随机变量 X 的概率密度满足：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty$$

其中 μ, σ (σ>0)为常数，则称 X 服从参数为 μ, σ 的正态分布或高斯(Gauss)分布，记为 X~N(μ, σ^2)。

使用广义二重积分和极坐标变换成累次积分可得：$\int_{-\infty}^{+\infty}f(x)dx=1$

正态分布性质：

1) $f(x)$曲线关于 x=μ 对称。

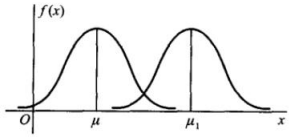
2) 当 x=μ 时，$f(x)$取得最大值：

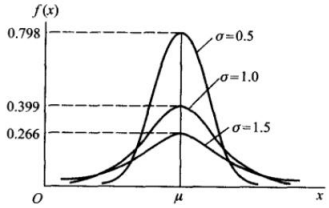
$$f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$$

x 离 μ 越远，$f(x)$越小。在$x=\mu\pm\sigma$处曲线有拐点；曲线以 x 轴为渐近线。

μ 决定$f(x)$的位置，而不影响形状，μ 称为位置参数。

σ 越小，图形变得越尖，X 落在 μ 附近的概率越大。





3) 当 μ=0, σ=1 时，随机变量 X 服从标准正态分布，X~N(0,1)。其概率密度和分布函数用 φ(x)、Φ(x)表示。

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$$

$$\Phi(x)+\Phi(-x)=1$$

人们已经编制了 Φ(x)的函数表，可供查找。

4) 对于一般 X~N(μ, σ^2)，可以做线性变换转化为标准正态分布。

$$\mathrm{Y}=\frac{X-\mu}{\sigma}\sim\mathrm{N}(0,1)$$

证明：

$$\mathrm{P}\{\mathrm{Y}\le x\}=\mathrm{P}\left(\frac{X-\mu}{\sigma}\le x\right)=\mathrm{P}\{\sigma x+\mu\}=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\sigma x+\mu}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$$

令$\frac{t-\mu}{\sigma}=s$，则 $\mathrm{ds}=\frac{dt}{\sigma}$

$\mathrm{P}\{\mathrm{Y}\le x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{s^2}{2}}ds=\Phi(x)$，因此 $\mathrm{Y}\sim\mathrm{N}(0,1)$

X 的分布函数 F(x)写为：

$$F(x)=P\{X\le x\}=P\left\{\frac{X-\mu}{\sigma}\le\frac{x-\mu}{\sigma}\right\}=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

5) 3σ 法则

$$P\{\mu-3\sigma<X<\mu+3\sigma\}=99.74\%$$

其值非常接近 1，所以认为取值落在$(\mu-3\sigma,\mu+3\sigma)$内几乎是肯定的。

例：将温度调节器放在有某种液体的容器内，调节器定在 d℃，液体温度 X～N(d, 0.5^2)。(1)若 d=90℃，求 X 小于 89℃的概率。(2)若要求液体的温度至少为 80℃的概率不低于 0.99，d 至少为多少？

1) 线性变换转为标准正态分布：

$$P\{X<89\}=P\left\{\frac{X-90}{0.5}<\frac{89-90}{0.5}\right\}=\Phi\left(\frac{89-90}{0.5}\right)=\Phi(-2)=1-\Phi(2)$$

查表的$\Phi(2.00)=0.9772$

所以$P\{X<89\}=1-0.9772=0.0228$

2) 由题意得 $P\{X\ge80\}\ge0.99$

$$P\{X\ge80\}=1-P\{X<80\}=1-\Phi\left(\frac{80-d}{0.5}\right)=\Phi\left(\frac{d-80}{0.5}\right)$$

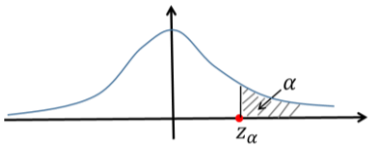
查表得$\Phi(2.33)=0.9901>0.99$

$$\frac{d-80}{0.5}>2.33\quad\rightarrow\quad d>81.165$$

设 X~N(0,1)，若z_α满足：

$$P\{X>z_\alpha\}=\alpha,0<\alpha<1$$

则称点z_α为标准正态分布的上 α 分位点。



由对称性可知：$z_{1-\alpha}=-z_\alpha$

生活中如男性成年人的身高、测量零件的误差、海洋波浪的高度等都服从正态分布。正态随机变量在概率论与数理统计理论和实际应用中都有着重要的作用。

🕮 2.8 庞加莱买面包问题

一个叫庞加莱的帅哥每天买 1kg 的面包，都回家称量并做记录；他发现一年平均质量是 0.95kg；于是他认为面包店缺斤少两，投诉了该面包店。

面包店的策略是：叮嘱店员，每次都给庞加莱最大的。

一年后，庞加莱又来投诉面包店还是缺斤少两，欺骗百姓，只不过每次都给自己大的面包。

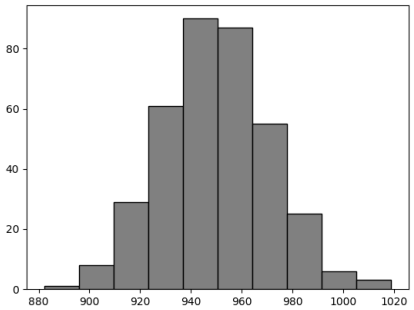
庞加莱是如何知道的？

```
import numpy
import scipy.stats as sta
import matplotlib.pyplot as plt
import random
```

```python
def buy_bread(f, pic_name):
    X = sta.norm(loc=950, scale=20)  # 平均值 950,标准差 20 的正态随机变量 X
    lst = []
    for i in range(365):
        x = X.rvs(size=100)  # 面包店一天生产 100 个面包
        lst.append(f(x))  # 根据函数 f 挑选一个
    print(f'平均值: {numpy.mean(lst)}')  # 一年平均值
    print(f'偏度: {sta.skew(lst)}')  # 偏度
    plt.hist(lst, color='grey', edgecolor='black')  # 直方图
    plt.savefig(f'{pic_name}.png', format='png')  # 存储为图片

def main():
    # 两个一起执行第 2 张图会有问题?
    buy_bread(f=random.choice, pic_name='first_year')
    # buy_bread(f=max, pic_name='second_year')

if __name__ == '__main__':
    main()
```

执行第 1 年结果：

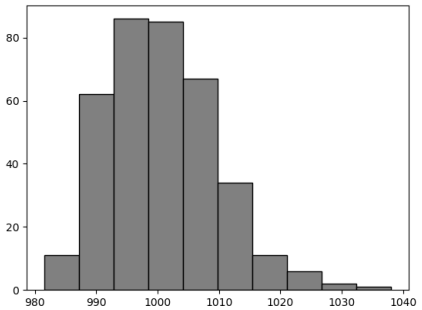
```
平均值: 949.5629805482257
偏度: 0.07426643059780352
```



偏度越小，数据分布越对称。

执行第 2 年结果：

```
平均值: 1000.7355805031303
偏度: 0.6751619291306432
```



虽然平均买到 1000，但是根据是否对称(数据正偏)，判断每次拿到都是大的。

🕮 2.9 随机变量的函数的分布

已知随机变量 X，Y=f(X)，求 Y 的分布。

比如测量一个圆的半径为 r，但是更关心圆的面积$S = \pi r^2$

例 1：已知离散型随机变量 X 的分布律：

X	-2	0	1	2
p	0.1	0.2	0.3	0.4

求$Y = X^2 + 1$的分布律。

Y	1	2	5
p	0.2	0.3	0.5

例 2：已知连续型随机变量 X 的概率密度：

$$f(x)=\begin{cases}2x, & 0<x<1\\ 0, & \text{other}\end{cases}$$

求Y = 3X + 1的概率密度。

1) 求 Y 的分布函数 G(y)

$$G(y)=P\{Y\le y\}=P\{3X+1\le y\}=P\left\{X\le\frac{y-1}{3}\right\}=\int_0^{\frac{y-1}{3}}2xdx=(\frac{y-1}{3})^2$$

$$0<x<1 \quad\rightarrow\quad 0<\frac{y-1}{3}<1 \quad\rightarrow\quad 1<y<4$$

$$\therefore G(y)=\begin{cases}(\frac{y-1}{3})^2, & 1<y<4\\ 0, & other\end{cases}$$

2) 求 Y 的概率密度 g(y)

$$g(y)=G'(y)=\begin{cases}\frac{2}{9}(x-1), & 1<y<4\\ 0, & other\end{cases}$$

设 X 的概率密度是$f(x)$，$a<x<b$，$y=g(x)$在(a,b)严格单调连续，且存在唯一的反函数$x=h(y)$, $\alpha<y<\beta$，且$h'(y)$连续，则 Y = g(X)也是连续型随机变量，其概率密度是：

$$f_Y(y)=f\big(h(y)\big)|h'(y)|,\alpha<y<\beta$$

例：设X~U$(-\frac{\pi}{2},\frac{\pi}{2})$，求 Y = tanX 的概率密度。

X 服从均匀分布，其概率密度为：

$$f(x)=\begin{cases}\frac{1}{\pi}, & -\frac{\pi}{2}<x<\frac{\pi}{2}\\ 0, & other\end{cases}$$

反函数$h(y)=\arctan y, y\in(-\infty,+\infty)$

$$f_Y(y)=\frac{1}{\pi}(arctany)'=\frac{1}{\pi}\cdot\frac{1}{1+y^2}, y\in(-\infty,+\infty)$$

这是柯西(Cauchy)分布

🕮 2.10 概率分布的 Python 实现

① 二项分布 X~$b(n,p)$

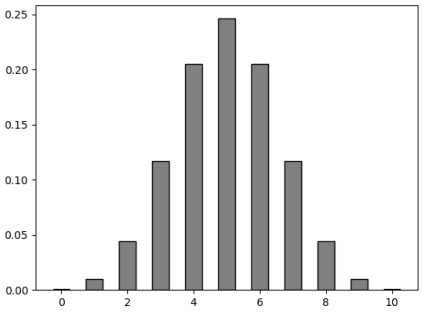
$$b(k;n,p)=C_n^k p^k q^{n-k}$$

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import binom

def b(n, p):  # 二项分布
```

```
    rv = binom(n, p)
    x = np.arange(n+1)
    y = rv.pmf(x)
    print(y)
    plt.bar(x, y, width=0.5, color='grey', edgecolor='black')
    plt.savefig('binom.png', format='png')

if __name__ == '__main__':
    b(10, 0.5)
```

结果：

```
[0.00097656 0.00976563 0.04394531 0.1171875  0.20507813 0.24609375
 0.20507813 0.1171875  0.04394531 0.00976563 0.00097656]
```



② 几何分布

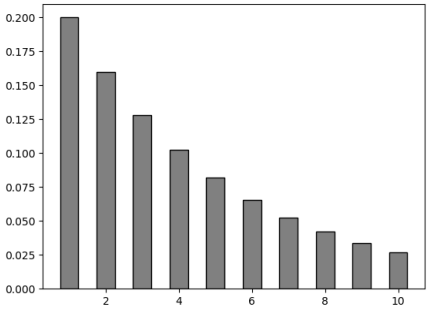
在 n 次贝努里试验中，试验 k 次才第一次成功的机率。也就是：前 k-1 次都失败但第 k 次成功的概率。

$$g(k;p) = (1-p)^{k-1}p$$

```
from scipy.stats import geom

def g(p):  # 几何分布
    N = 10  # 只画前 N 次
    rv = geom(p)
    x = np.arange(1, N+1)
    y = rv.pmf(x)
    print(y)
    plt.bar(x, y, width=0.5, color='grey', edgecolor='black')
    plt.savefig('geom.png', format='png')

if __name__ == '__main__':
    g(0.2)
```

结果：

```
[0.2        0.16       0.128      0.1024     0.08192    0.065536
 0.0524288  0.04194304 0.03355443 0.02684355]
```



③ 泊松分布X~π(λ)

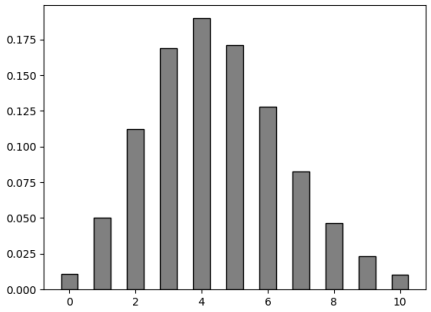
$$\mathrm{P(X=k)} = \frac{\lambda^k e^{-\lambda}}{k!} \quad ,k = 0,1,2,\ldots$$

```
from scipy.stats import poisson

def pi(lmd):  # 泊松分布
    N = 10  # 只画前 N 次
```

```
    rv = poisson(lmd)
    x = np.arange(N+1)
    y = rv.pmf(x)
    print(y)
    plt.bar(x, y, width=0.5, color='grey', edgecolor='black')
    plt.savefig('poisson.png', format='png')

if __name__ == '__main__':
    pi(4.5)
```

结果：

```
[0.011109   0.04999048 0.11247859 0.16871788 0.18980762 0.17082686
 0.12812014 0.08236295 0.04632916 0.02316458 0.01042406]
```

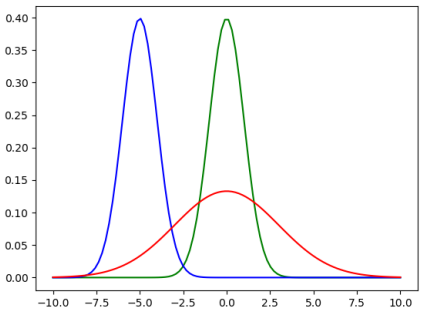


④ 正态分布 X~N(μ, σ^2)

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty$$

```
from scipy.stats import norm

def n():  # 正态分布
    x = np.linspace(-10, 10, 100)
    #loc 相当于 μ; scale 相当于 σ
    rv1 = norm(loc=0, scale=1)
    rv2 = norm(loc=-5, scale=1)
    rv3 = norm(loc=0, scale=3)
    # 3 个正态分布的叠加图
    plt.plot(x, rv1.pdf(x), color='green')
    plt.plot(x, rv2.pdf(x), color='blue')
    plt.plot(x, rv3.pdf(x), color='red')
    plt.savefig('norm.png', format='png')
```

结果：



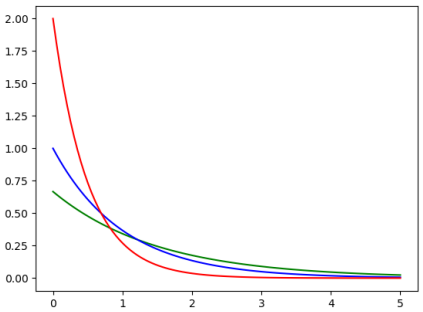
⑤ 指数分布

$$f(x)=\lambda e^{-\lambda x} \text{ 或 } f(x)=\frac{1}{\theta}e^{-x/\theta}$$

```
from scipy.stats import expon

def e():  # 指数分布
    x = np.linspace(0, 5, 100)
    # scale 相当于 1/λ 或 θ
    rv1 = expon(scale=1.5)
```

```
    rv2 = expon(scale=1)
    rv3 = expon(scale=0.5)
    # 3 个指数分布的叠加图
    plt.plot(x, rv1.pdf(x), color='green')
    plt.plot(x, rv2.pdf(x), color='blue')
    plt.plot(x, rv3.pdf(x), color='red')
    plt.savefig('expon.png', format='png')
```

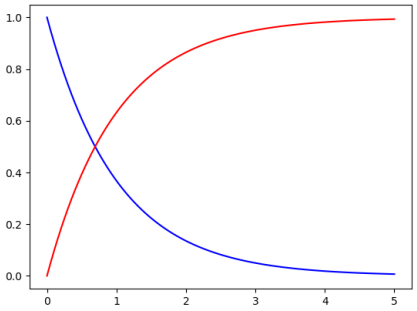
结果：



指数分布的分布函数：$\mathrm{F}(x)=1-e^{-\lambda x}$

```
def compare_pdf_cdf():
    x = np.linspace(0, 5, 100)
    rv = expon(scale=1)
    # pdf 为概率密度,cdf 为累计概率,就是分布函数
    plt.plot(x, rv.pdf(x), color='blue')
    plt.plot(x, rv.cdf(x), color='red')
    plt.savefig('指数分布概率密度和分布函数.png', format='png')
```

结果：



第 3 章　二维随机变量及其分布

🕮　3.1　二维随机变量

❀　3.1.1　二维随机变量的分布函数

设随机事件 E 的样本空间是 S={e}，X=X(e)和 Y=Y(e)是定义在 S 上的随机变量，向量(X, Y)叫做二维随机向量或二维随机变量。

对于二维随机变量(X, Y)的分布函数：

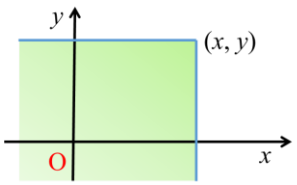
$$\mathrm{F}(x,y)=\mathrm{P}\{(\mathrm{X}\le x)\cap(\mathrm{Y}\le y)\}=\mathrm{P}\{\mathrm{X}\le x,\mathrm{Y}\le y\}$$

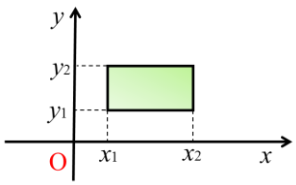
称为随机变量 X 和 Y 的联合分布函数。

如果将(X, Y)看成平面上的随机点的坐标，分布函数 F(x, y)的值就是随机点落在如图以点(x, y)为顶点是左下方无穷矩形区域内的概率。

由此可得随机点(X, Y)落在矩形区域$\{(x,y)|x_1<x\le x_2,y_1<y\le y_2\}$的概率是：

$$\mathrm{P}\{x_1<X\le x_2,y_1<Y\le y_2\}=\mathrm{F}(x_2,y_2)-\mathrm{F}(x_2,y_1)+\mathrm{F}(x_1,y_1)-\mathrm{F}(x_1,y_2)$$





联合分布函数 F(x, y)的性质：

1) F(x, y)是变量 x 和 y 的不减函数。如对固定的 y 当$x_2 > x_1$时$\mathrm{F}(x_2, \mathrm{y}) \geq \mathrm{F}(x_1, \mathrm{y})$

2) $0 \leq \mathrm{F}(x, y) \leq 1$；$\mathrm{F}(-\infty, y) = \mathrm{F}(x, -\infty) = \mathrm{F}(-\infty, -\infty) = 0$；$\mathrm{F}(+\infty, +\infty) = 1$

3) 右连续性：$\mathrm{F}(x + 0, y) = \mathrm{F}(x, y); \mathrm{F}(x, y + 0) = \mathrm{F}(x, y)$

4) $\mathrm{F}(x_2, y_2) - \mathrm{F}(x_2, y_1) + \mathrm{F}(x_1, y_1) - \mathrm{F}(x_1, y_2) \geq 0$

因为上式左边就是$\mathrm{P}\{x_1 < X \leq x_2, y_1 < Y \leq y_2\}$的值，由概率的非负性即得。

❀ 3.1.2 二维离散型和连续型随机变量

① 二维离散型

若(X, Y)取值有限多对或可列无限多对，则(X, Y)是离散型随机变量。

所有可能取值$(x_i, y_j), i, j = 1,2,3 \ldots$记为：

$$\mathrm{P}\{\mathrm{X} = x_i, \mathrm{Y} = y_j\} = p_{ij}, \qquad i, j = 1,2,3 \ldots$$

满足：

$$p_{ij} \geq 0, \qquad \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{ij} = 1$$

p_{ij} 称为二维离散型随机变量(X, Y)的分布律，或随机变量 X 和 Y 的联合分布律。

② 二维连续型

设(X, Y)的分布函数 F(x, y)，如果存在非负函数 $f(x, y)$使对任意 x,y 满足：

$$\mathrm{F}(x, y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(u, v)dudv$$

则(X, Y)是连续型二维随机变量，函数 f (x, y)称为二维随机变量(X, Y)的概率密度，或随机变量 X 和 Y 的联合概率密度。

概率密度$f(x, y)$的性质：

1) $f(x, y) \geq 0$

2) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)dxdy = F(+\infty, +\infty) = 1$

3) 点(X, Y)落在平面区域 D 的概率：

$$\mathrm{P}\{(\mathrm{X}, \mathrm{Y}) \in \mathrm{D}\} = \iint_D f(x, y)dxdy$$

4) 若$f(x, y)$在点(x, y)连续，则有：

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y)$$

二维随机变量的情况，也适用于 n ($n > 2$)维随机变量。

🕮 3.2 边缘分布

二维随机变量(X, Y)作为整体，具有分布函数$F(x,y)$，而 X 和 Y 各自也有分布函数，记为$F_X(x)$和$F_Y(y)$，称为二维随机变量(X, Y)关于 X 和 Y 的边缘分布函数。

$$F_X(x) = P\{X \le x\} = P\{X \le x, Y < +\infty\} = F(x, +\infty)$$

同理：

$$F_Y(y) = F(+\infty, y)$$

① 离散型

X 的分布律：

$$\mathrm{P}\{\mathrm{X} = x_i\} = p_{i1} + p_{i2} + \cdots + p_{i\infty} = \sum_{j=1}^{+\infty} p_{ij} \ , \quad i = 1,2,3 \ldots$$

记为$p_{i\cdot}$并称其为关于 X 的边缘分布律。

同理 Y 的边缘分布律：

$$p_{\cdot j} = \sum_{i=1}^{+\infty} p_{ij} \ , \quad j = 1,2,3 \ldots$$

② 连续型

设连续型随机变量(X, Y)的概率密度是$f(x,y)$，关于 X 的边缘分布函数：

$$F_X(x) = F(x, +\infty) = \int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} f(x,y)dy \right) dx$$

X 是连续型随机变量，其概率密度是分布函数求导所得：

$$f_X(x) = F_X'(x) = \int_{-\infty}^{+\infty} f(x,y)dy$$

同理，Y 的概率密度：

$$f_Y(y) = F_Y'(y) = \int_{-\infty}^{+\infty} f(x,y)dx$$

分别称$f_X(x)$和$f_Y(y)$是(X, Y)关于 X 和 Y 的边缘概率密度。

例：袋中有 2 个编号为 1 的球，3 个编号为 0 的球，有/无放回一个一个取出。X 为第 1 次取出的编号；Y 为第 2 次取出的编号。求 X 和 Y 的联合分布律和各自的边缘分布律。

1) 有放回

Y \ X	0	1	P{Y=y_j}
0	0.6×0.6	0.4×0.6	0.6
1	0.6×0.4	0.4×0.4	0.4
P{X=x_i}	0.6	0.4	1

2) 无放回

Y \ X	0	1	$P\{Y=y_j\}$
0	0.6×0.5	0.4×0.75	0.6
1	0.6×0.5	0.4×0.25	0.4
$P\{X=x_i\}$	0.6	0.4	1

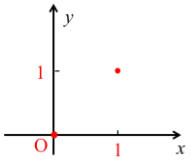
常常将边缘分布律写在联合分布律表格的边缘上，这就是边缘分布律的由来。可以看出两种情况对应边缘分布是一样的，但是联合分布不同。

所以，已知 X 和 Y 的联合分布可以确定边缘分布；但是已知 X 和 Y 的边缘分布，不能确定联合分布。

🕮 3.3 常见的三种二维分布

① 二维两点分布

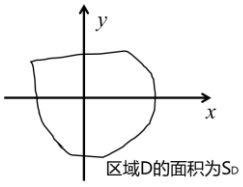
Y \ X	0	1	
0	1-p	0	1-p
1	0	p	p
	1-p	p	1



(X, Y)的联合分布函数 $F(x, y)$：

1) 当 x<0 或 y<0 时，$F(x, y) = 0$；
2) 当 0≤x<1,y≥0 或 0≤y<1,x≥0 时，$F(x, y) = 1-p$；
3) 当 x≥1,y≥1 时，$F(x, y) = 1$

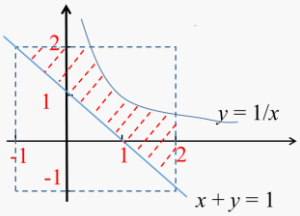
② 二维均匀分布



(X, Y)的联合概率密度 $f(x, y)$：

$$f(x,y)=\begin{cases}\dfrac{1}{S_D}, & (x,y)\in D\\ 0, & (x,y)\notin D\end{cases}$$

例：在[-1,2]任取两数 X 和 Y，求 X+Y>1 且 XY<1 的概率。



根据几何概型：

$$\mathrm{p}=\frac{S_D}{S_{矩形}}=\frac{3\times\frac{1}{2}+2\int_1^2\frac{1}{x}dx}{3\times3}=\frac{2}{9}\ln 2+\frac{1}{6}$$

③ 二维正态分布

概率密度为：

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\{\frac{-1}{2(1-\rho^2)}[\frac{(x-\mu_1)^2}{{\sigma_1}^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{{\sigma_2}^2}]\}$$

其中$\sigma_1,\sigma_2>0$，$-1<\rho<1$

称(X, Y)服从参数为$\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$的二维正态分布。

记为$(X,Y)\sim N(\mu_1,{\sigma_1}^2;\mu_2,{\sigma_2}^2;\rho)$

🕮 3.4 条件分布

① (X, Y)是二维离散型随机变量

分布律：$P\{X=x_i,Y=y_j\}=p_{ij},\ i,j=1,2,3\ldots$

对于固定的 j，若 $P\{Y=y_j\}>0$，由条件概率公式得：

$$P\{X=x_i|Y=y_j\}=\frac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\frac{p_{ij}}{p_{\cdot j}},\ i=1,2,3\ldots$$

称为在 $Y=y_j$ 条件下随机变量 X 的条件分布律。

② (X, Y)是二维连续型随机变量

对于固定的 y，若 $f_Y(y)>0$：

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$

称为在 $Y=y$ 条件下 X 的条件概率密度。

此条件下的条件分布函数为：

$$\int_{-\infty}^{x}f_{X|Y}(x|y)dx=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}dx$$

例：设(X, Y)的联合概率密度函数为：

$$f(x,y)=\begin{cases}\frac{1}{y}e^{-\frac{x}{y}}e^{-y}, & x,y>0\\ 0, & other\end{cases}$$

求 P{X>1|Y=y}。

在 Y=y>0 时，X 的条件概率密度是：

$$f(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{\frac{1}{y}e^{-\frac{x}{y}}e^{-y}}{\frac{1}{y}e^{-y}\int_0^{+\infty}e^{-\frac{x}{y}}dx}=\frac{e^{-\frac{x}{y}}}{\left(-ye^{-\frac{x}{y}}\right)|_0^{+\infty}}=\frac{e^{-\frac{x}{y}}}{y},\qquad x>0$$

所求条件概率是条件概率密度的积分：

$$P\{X>1|Y=y\}=\int_1^{+\infty}f(x|y)dx=\int_1^{+\infty}\frac{e^{-\frac{x}{y}}}{y}dx=\left(-e^{-\frac{x}{y}}\right)|_1^{+\infty}=e^{-\frac{1}{y}}$$

🕮 3.5 独立性

设二维随机变量(X, Y)的分布函数是 $F(x,y)$，边缘分布函数分别为 $F_X(x)$和 $F_Y(y)$，对于所有 x, y 有：$F(x,y)=F_X(x)F_Y(y)$

即：$P\{X\le x, Y\le y\}=P\{X\le x\}\cdot P\{Y\le y\}$

则称随机变量 X 和 Y 的相互独立的。

1) 当是离散型时有：

$$P\{X=x_i, Y=y_j\}=P\{X=x_i\}P\{Y=y_j\}$$

即：

$$p_{ij}=p_{i\cdot}\times p_{\cdot j}$$

2) 当是连续型时，概率密度有：

$$f(x,y)=f_X(x)\times f_Y(y)$$

例：设 X,Y 独立同分布，X 的概率密度是：

$$f(x)=\begin{cases}2x, & 0\le x\le 1\\ 0, & other\end{cases}$$

求 $P\{X+Y\le 1\}$

因为独立，所以 X,Y 的联合概率密度为：

$$f(x,y)=f(x)f(y)=\begin{cases}4xy, & 0\le x,y\le 1\\ 0, & other\end{cases}$$

根据二重积分的累次积分方法：

$$P\{X+Y\le 1\}=\iint_{x+y\le 1}f(x,y)dxdy=\int_0^1\int_0^{1-x}4xydxdy=\int_0^1 2x(1-x)^2dx=$$

$$=\left(\frac{2}{4}x^4-\frac{4}{3}x^3+x^2\right)|_0^1=\frac{1}{6}$$

独立性的性质：

1) X 与 Y 独立，则 f(X)和 g(Y)独立，f, g 是任意连续函数。

2) 常数 c 与任意随机变量独立。

3) 若联合概率密度 $f(x,y)$可分离成：

$$f(x,y)=g(x)h(y)$$

且 g(x)的非 0 区域与 y 无关，h(y)的非 0 区域与 x 无关，则 X, Y 独立。

🕮 3.6 两个随机变量的函数的分布

例 1：设 X, Y 独立，且服从泊松分布：$X\sim\pi(\lambda_1), Y\sim\pi(\lambda_2)$
求 X+Y 的分布律。

$$\begin{aligned}P\{X+Y=n\} &= P\{X=0,Y=n\}+P\{X=1,Y=n-1\}+\cdots+P\{X=n,Y=0\}\\ &=\sum_{k=0}^{n}P\{X=k,Y=n-k\}=\sum_{k=0}^{n}P\{X=k\}P\{Y=n-k\}\\ &=\sum_{k=0}^{n}\frac{{\lambda_1}^k}{k!}e^{-\lambda_1}\frac{{\lambda_2}^{n-k}}{(n-k)!}e^{-\lambda_2}=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!\,(n-k)!}{\lambda_1}^k{\lambda_2}^{n-k}\\ &=\frac{(\lambda_1+\lambda_2)^n}{n!}e^{-(\lambda_1+\lambda_2)}\end{aligned}$$

即：$X+Y\sim\pi(\lambda_1+\lambda_2)$

例 2：最大值最小值分布
设 X, Y 独立，分布函数分别为 $F_X(x)$和 $F_Y(y)$，求 $Z_1=\max(X,Y)$，$Z_2=\min(X,Y)$的分布函数。

$$\begin{aligned}F_{Z_1}(z) &= P\{\max(X,Y)\le z\}=P\{X\le z,Y\le z\}=P\{X\le z\}P\{Y\le z\}\\ &=F_X(z)F_Y(z)\\ F_{Z_2}(z) &= P\{\min(X,Y)\le z\}=1-P\{\min(X,Y)>z\}=1-P\{X>z,Y>z\}\\ &=1-P\{X>z\}P\{Y>z\}=1-(1-P\{X\le z\})(1-P\{Y\le z\})\\ &=1-(1-F_X(z))(1-F_Y(z))\end{aligned}$$

例 3：设(X, Y)的联合概率密度 $f(x,y)$，求 Z=X+Y 的分布。

Z 的分布函数为：

$$F_Z(z)=P\{X+Y\le z\}=\iint_{x+y\le z}f(x,y)dxdy=\int_{-\infty}^{+\infty}\int_{-\infty}^{z-y}f(x,y)dxdy$$

固定 z 与 y，令 $x=u-y$，则：

$$F_Z(z)=\int_{-\infty}^{+\infty}\int_{-\infty}^{z}f(u-y,y)dudy=\int_{-\infty}^{z}\left(\int_{-\infty}^{+\infty}f(u-y,y)dy\right)du$$

所以，Z 的概率密度函数为：

$$f_Z(z)=F_Z'(z)=\int_{-\infty}^{+\infty}f(z-y,y)dy=\int_{-\infty}^{+\infty}f(x,z-x)dx$$

也称为卷积公式。

同样方法也可以求出：如 Z=X-Y、X·Y、X/Y 等的分布。

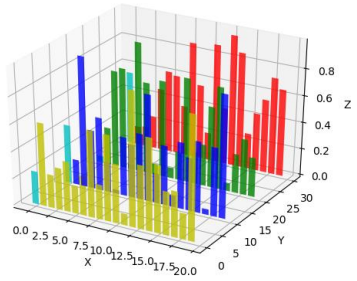
若 X, Y 独立，且服从正态分布：$X\sim N(\mu_1,{\sigma_1}^2), Y\sim N(\mu_2,{\sigma_2}^2)$
则$X+Y\sim N(\mu_1+\mu_2,{\sigma_1}^2+{\sigma_2}^2)$。
正态分布具有可加性：有限个相互独立的正态随机变量的线性组合仍然服从正态分布。

🕮 3.7 二维随机变量的 Python 实现

matplotlib 官网 mplot3d 的示例：

```
from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
import numpy as np

def bars3d_demo():
    """
    ========================================
    Create 2D bar graphs in different planes
    ========================================
    Demonstrates making a 3D plot which has 2D bar graphs projected onto
    planes y=0, y=1, etc.
    """
    fig = plt.figure()
    ax = fig.add_subplot(111, projection='3d')
    for c, z in zip(['r', 'g', 'b', 'y'], [30, 20, 10, 0]):
        xs = np.arange(20)
        ys = np.random.rand(20)
        # You can provide either a single color or an array. To demonstrate this,
        # the first bar of each set will be colored cyan.
        cs = [c] * len(xs)
        cs[0] = 'c'
        ax.bar(xs, ys, zs=z, zdir='y', color=cs, alpha=0.8)
    ax.set_xlabel('X')
    ax.set_ylabel('Y')
    ax.set_zlabel('Z')
    plt.savefig('bars3d_demo.png', format='png')

if __name__ == '__main__':
    bars3d_demo()
```

结果：



例：对一人群吸烟情况 X 和健康情况 Y 调查：

X=1 不吸烟；X=2 吸烟一般；X=3 吸烟严重

Y=-1 不健康；Y=0 一般；Y=1 健康

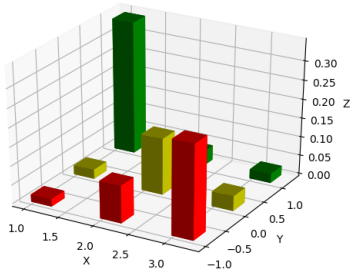
(X,Y)的联合分布律为：

Y \ X	1	2	3
-1	0.02	0.1	0.25
0	0.025	0.15	0.04
1	0.35	0.04	0.025

```
from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
import numpy as np

def smoke_demo():
    # 3D 柱状图
    fig = plt.figure()
    ax = fig.add_subplot(111, projection='3d')
    dx, dy = 0.3, 0.3  # 柱状底面是边长 0.3 的正方形
```

```
    # z 高度对应联合分布律的概率
    dz = [0.02, 0.025, 0.35, 0.1, 0.15, 0.04, 0.25, 0.04, 0.025]
    zpos = 0
    i = 0
    for xpos in range(1, 4):
        for c, ypos in zip(['r', 'y', 'g'], [-1, 0, 1]):
            ax.bar3d(xpos, ypos, zpos, dx, dy, dz[i], color=c)
            i += 1
    ax.set_xlabel('X')
    ax.set_ylabel('Y')
    ax.set_zlabel('Z')
    plt.savefig('smoke_demo.png', format='png')

if __name__ == '__main__':
    smoke_demo()
```

结果：



二维正态分布：bivariate_normal()函数

// The bivariate_normal function was deprecated in Matplotlib 2.2 and will be removed in 3.1

第 4 章　随机变量的数字特征

🕮　4.1　数学期望

❀　4.1.1　数学期望的定义

① 离散型

设离散型随机变量 X 的分布律是：

$$P\{X = x_k\} = p_k, \qquad k = 1,2,3,\dots$$

定义 E(X)为 X 的数学期望(简称期望，也称均值)：

$$E(X) = \sum_{k=1}^{\infty} x_k p_k$$

其中级数$\sum_{k=1}^{\infty} x_k p_k$需要绝对收敛。

设有级数$\sum u_n$：

1) 若$\sum |u_n|$收敛，则称$\sum u_n$绝对收敛，绝对收敛级数一定收敛；

2) 若$\sum u_n$收敛，而$\sum |u_n|$发散，则称$\sum u_n$条件收敛。

黎曼定理：绝对收敛的级数可以任意交换次序。

因为随机试验的结果具有任意次序性，所以任意次序的数学期望应是一样的。对于可列无限多的级数必须收敛。

② 连续型

设连续型随机变量 X 的概率密度是$f(x)$，数学期望 E(X)为：

$$\mathrm{E(X)} = \int_{-\infty}^{+\infty} xf(x)dx$$

其中广义积分$\int_{-\infty}^{+\infty} xf(x)dx$需要**绝对收敛**，也就是$\int_{-\infty}^{+\infty} |xf(x)|dx$需要收敛。

❀ 4.1.2 常见三种离散型随机变量的期望

① 两点分布

$\mathrm{E(X)} = 1 \times p + 0 \times (1-p) = p$

② 二项分布：X~B(n, p)

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k}$$

$$E(\mathrm{X}) = \sum_{k=0}^{n} kC_n^k p^k (1-p)^{n-k} = \sum_{k=1}^{n} k\frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$

$$= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\,(n-1-(k-1))!} p^{k-1}(1-p)^{n-k}$$

$$= np\sum_{i=0}^{n-1} \frac{(n-1)!}{i!\,(n-1-i)!} p^i (1-p)^{n-1-i} = np(p+1-p)^{n-1} = np$$

③ 泊松分布：X~π(λ)

$$\mathrm{P(X = k)} = \frac{\lambda^k e^{-\lambda}}{k!}\ , k = 0,1,2,\ldots$$

$$E(\mathrm{X}) = \sum_{k=0}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} = \lambda\sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

❀ 4.1.3 常见三种连续型随机变量的期望

① 均匀分布：X~U(*a*, *b*)

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0, & \text{其他} \end{cases}$$

$$\mathrm{E(X)} = \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\frac{1}{2}x^2\Big|_b^a = \frac{a+b}{2}$$

② 指数分布：X~EXP(λ)

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{其他} \end{cases}$$

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{0}^{+\infty}x\cdot\lambda e^{-\lambda x}dx=-\int_{0}^{+\infty}xde^{-\lambda x}$$

$$=-xe^{-\lambda x}\Big|_{0}^{+\infty}+\int_{0}^{+\infty}e^{-\lambda x}dx=0-\frac{1}{\lambda}e^{-\lambda x}\Big|_{0}^{+\infty}=\frac{1}{\lambda}$$

③ 正态分布：X~N(μ, σ^2)

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty$$

$$E(X)=\int_{-\infty}^{+\infty}x\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

令$t=\frac{x-\mu}{\sigma}$，则$x=\sigma t+\mu, dx=\sigma\cdot dt$

$$E(X)=\int_{-\infty}^{+\infty}(\sigma t+\mu)\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$

因为$\sigma te^{-\frac{t^2}{2}}$是奇函数，在$(-\infty,+\infty)$上积分为 0，所以：

$$E(X)=\frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt$$

根据广义二重积分和极坐标变换成累次积分，可得：

$$\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\sqrt{2\pi}$$

∴ $E(X)=\mu$

❀ 4.1.4 随机变量的函数的期望

设 X 分布已知，Y=g(X)，g 是连续函数，求 E(Y)。

1) 离散型

$$E(Y)=E\big(g(X)\big)=\sum_{k=1}^{\infty}g(x_k)p_k$$

2) 连续型

$$E(Y)=E\big(g(X)\big)=\int_{-\infty}^{+\infty}g(x)f(x)dx$$

也需要级数和广义积分绝对收敛。

数学期望的性质：

1) E(c) = c，c 是常数；

2) E(cX) = cE(X)，c 是常数；
3) 设有两个随机变量 X 和 Y，则 E(X±Y) = E(X) ± E(Y)，可以推广到有限个；
4) 若 X、Y 独立，则 E(X·Y)=E(X)·E(Y)，可以推广到有限个(独立)；

🕮 4.2 方差

❀ 4.2.1 方差的定义及其性质

研究随机变量与均值的偏离程度，可使用 E{|X-E(X)|}衡量。
但因为使用绝对值，计算不方便，通常取平方：E{(X-E(X))²}。

设随机变量 X，若 E{(X-E(X))²}存在，则称其为 X 的*方差*。
记为 D(X)或 Var(X)：

$$\mathrm{D(X)} = \mathrm{Var(X)} = \mathrm{E}\{(X - E(X))^2\}$$

引入σ(X) = $\sqrt{D(X)}$，称为*标准差*或*均方差*。

1) 离散型

$$\mathrm{D(X)} = \sum_{k=1}^{\infty}(x_k - E(X))^2 p_k$$

2) 连续型

$$\mathrm{D(X)} = \int_{-\infty}^{+\infty}\big(x - E(X)\big)^2 f(x)dx$$

方差的性质：
1) D(c) = 0，c 是常数；
2) $\mathrm{D(X)} = \mathrm{E}(X^2) - E(X)^2$

证明：$\mathrm{D(X)} = \mathrm{E}\left\{\big(X - E(X)\big)^2\right\} = \mathrm{E}\{X^2 - 2\mathrm{XE(X)} + E(X)^2\}$

{}里面的 E(X)是常数，可以提出来：

∴ $\mathrm{D(X)} = \mathrm{E}(X^2) - 2\mathrm{E(X)E(X)} + E(X)^2 = \mathrm{E}(X^2) - E(X)^2$

这是计算方差**最常用**的公式。
3) *a*,*b* 是常数，有：

$\mathrm{D}(a\mathrm{X} + b) = a^2 D(X)$

4) 两个随机变量 X、Y，X+Y 的方差：

$$\begin{aligned}\mathrm{D(X+Y)} &= \mathrm{E}\left\{\big((X+Y) - E(X+Y)\big)^2\right\} = \mathrm{E}\left\{\big((X - E(X)) + (Y - E(Y))\big)^2\right\}\\ &= E\left\{\big(X - E(X)\big)^2 + \big(Y - E(Y)\big)^2 + 2\big(X - E(X)\big)\big(Y - E(Y)\big)\right\}\\ &= D(X) + D(Y) + 2E\left\{\big(X - E(X)\big)\big(Y - E(Y)\big)\right\}\\ &= D(X) + D(Y) + 2(E(XY) - E(X)E(Y))\end{aligned}$$

当 X 与 Y 相互独立时，$\mathrm{D(X \pm Y)} = D(X) + D(Y)$

❁ 4.2.2 常见三种离散型随机变量的方差

① 两点分布

$$\mathrm{E(X)} = 1 \times \mathrm{p} + 0 \times (1 - \mathrm{p}) = \mathrm{p}$$
$$\mathrm{E}(X^2) = 1^2 \times \mathrm{p} + 0^2 \times (1 - \mathrm{p}) = \mathrm{p}$$
$$\mathrm{D(X)} = \mathrm{E}(X^2) - E(X)^2 = p - p^2 = p(1-p)$$

② 二项分布：X~B(n, p)

将二项分布看成 n 个独立的两点分布 $X_1, X_2, X_3 \ldots$：

$$\mathrm{D(X)} = \mathrm{D}\left(\sum_{k=1}^{n} X_i\right) = np(1-p)$$

③ 泊松分布：X~π(λ)

$$\mathrm{P(X = k)} = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0,1,2, \ldots$$

$$\mathrm{E}(X^2) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} (k-1+1) \frac{\lambda^k e^{-\lambda}}{(k-1)!}$$

$$= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2} e^{-\lambda}}{(k-2)!} + \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} = \lambda^2 + \lambda$$

$$\mathrm{D(X)} = \mathrm{E}(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

也就是泊松分布的期望和方差都是 λ。

❁ 4.2.3 常见三种连续型随机变量的方差

设 X 的期望 E(X)=μ，方差$\mathrm{D(X)} = \sigma^2 \neq 0$

令 $\mathrm{Y} = \frac{X-\mu}{\sigma}$，则：

$$\mathrm{E(Y)} = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma} E(X-\mu) = \frac{1}{\sigma}(\mu - \mu) = 0$$
$$\mathrm{D(Y)} = \mathrm{D}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} D(X) = 1$$

即 Y 的数学期望为 0，方差为 1，Y 称为 X 的标准化变量。

① 均匀分布：X~U(*a*, *b*)

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{其他} \end{cases}$$

$$\mathrm{E(X^2)} = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \frac{1}{3} x^3 \Big|_b^a = \frac{a^2 + b^2 + ab}{3}$$

$$\mathrm{D(X)} = \mathrm{E}(X^2) - E(X)^2 = \frac{a^2 + b^2 + ab}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

② 指数分布：X~EXP(λ)

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & \text{其他}\end{cases}$$

$$\mathrm{E}(\mathrm{X}^2)=\int_{-\infty}^{+\infty}x^2f(x)dx=\int_0^{+\infty}x^2\cdot\lambda e^{-\lambda x}dx=-\int_0^{+\infty}x^2de^{-\lambda x}$$

$$=-x^2e^{-\lambda x}\Big|_0^{+\infty}+\int_0^{+\infty}e^{-\lambda x}dx^2=2\int_0^{+\infty}xe^{-\lambda x}dx$$

$$=-\frac{2}{\lambda}\left(\int_0^{+\infty}xde^{-\lambda x}\right)=-\frac{2}{\lambda}xe^{-\lambda x}\Big|_0^{+\infty}+\frac{2}{\lambda}\int_0^{+\infty}e^{-\lambda x}dx$$

$$=-\frac{2}{\lambda^2}e^{-\lambda x}|_0^{+\infty}=\frac{2}{\lambda^2}$$

$$\mathrm{D}(\mathrm{X})=\mathrm{E}(X^2)-E(X)^2=\frac{2}{\lambda^2}-(\frac{1}{\lambda})^2=\frac{1}{\lambda^2}$$

③ 正态分布：X~N(μ, σ²)

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty$$

对于标准正态分布 X~N(0,1)：

$$\mathrm{E}(X^2)=\int_{-\infty}^{+\infty}x^2\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}-xde^{-\frac{x^2}{2}}$$

$$=\frac{1}{\sqrt{2\pi}}\left(-xe^{-\frac{x^2}{2}}\Big|_{-\infty}^{+\infty}-\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}d(-x)\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}dx=1$$

$$\mathrm{D}(\mathrm{X})=\mathrm{E}(X^2)-E(X)^2=1-0=1$$

对于一般的正态分布 Y，令：

$$\mathrm{X}=\frac{Y-\mu}{\sigma}\sim N(0,1)$$

$$\mathrm{D}(\mathrm{Y})=D(\sigma X+\mu)=\sigma^2D(X)=\sigma^2$$

所以正态分布的两个参数：μ代表数学期望，σ代表标准差，σ^2代表方差。

🕮 4.3 协方差&相关系数

二维随机变量(X, Y)，除了研究 X 和 Y 的数学期望和方差外，还需讨论描述 X 与 Y 之间相互关系的数字特征。

❀ 4.3.1 协方差

随机变量 X 与 Y 的协方差记为 Cov(X, Y)：

$$\mathrm{Cov}(\mathrm{X},\mathrm{Y})=\mathrm{E}\{[\mathrm{X}-\mathrm{E}(\mathrm{X})][\mathrm{Y}-\mathrm{E}(\mathrm{Y})]\}$$

P34：X±Y 的方差可改写为：

$$D(X \pm Y) = D(X) + D(Y) \pm 2Cov(X, Y)$$

协方差的性质：

1) Cov(X, Y) = Cov(Y, X)
2) Cov(X, X) = D(X)
3) Cov(X, Y) = E(XY) - E(X)E(Y)　// 计算协方差主要公式
4) 如果 X、Y 独立，则 Cov(X, Y) = 0；反之不一定。
5) $Cov(X_1+X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$
6) Cov(aX, bY) = abCov(X, Y)，其中 a、b 是常数

例：(X, Y)的联合概率密度函数为：

$$f(x,y) = \begin{cases} 2, & 0 \le x \le 1, 0 \le y \le x \\ 0, & other \end{cases}$$

求 Cov(X, Y)。

$$E(X) = \iint_D xf(x,y)dxdy = \int_0^1 \left(\int_0^x 2xdy \right) dx = \int_0^1 2x^2 dx = \frac{2}{3}$$

$$E(Y) = \iint_D yf(x,y)dxdy = \int_0^1 \left(\int_0^x 2ydy \right) dx = \int_0^1 x^2 dx = \frac{1}{3}$$

$$E(XY) = \iint_D xyf(x,y)dxdy = \int_0^1 \left(\int_0^x 2xydy \right) dx = \int_0^1 x^3 dx = \frac{1}{4}$$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{4} - \frac{2}{3} \times \frac{1}{3} = \frac{1}{36}$$

4.3.2 相关系数

随机变量 X 和 Y 的相关系数：

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)D(Y)}}$$

ρ_{XY} 没有单位(量纲)，消除了单位的影响。

XY 作标准化变换，令$X^* = \frac{X - E(X)}{\sqrt{D(X)}}, Y^* = \frac{Y - E(Y)}{\sqrt{D(Y)}}$，得到：

$$\rho_{XY} = Cov(X^*, Y^*)$$

也就是 ρ_{XY} 是标准尺度下的协方差。

相关系数的性质：

1) 若 X、Y 独立，则 $\rho_{XY} = 0$
2) $|\rho_{XY}| \le 1$

若$\rho_{XY} = 1$表示存在$a > 0, b \in R$，使得$Y = aX + b$　// 完全正相关

若$\rho_{XY}=-1$表示存在$a<0,b\in R$，使得$Y=aX+b$ // 完全负相关
若$\rho_{XY}=0$表示 X 与 Y 不相关
ρ_{XY}叫线性相关系数，反应 X 与 Y 直接的线性依赖程度。

例 1：设 $X\sim U\left(-\frac{1}{2},\frac{1}{2}\right),Y=\cos X$，求$\rho_{XY}$

$E(X)=0$

$$E(XY)=E(X\cos X)=\int_{-1/2}^{1/2}xcosx\cdot 1\cdot dx=0$$

$Cov(X,Y)=E(XY)-E(X)E(Y)=0$
$\rho_{XY}=0$，说明 X 与 Y 不相关，但它们之间有非线性关系。

例 2：抛硬币 n 次，正面出现次数 X，反面出现次数 Y。求ρ_{XY}。
$X+Y=n\rightarrow Y=-X+n$
$\therefore \rho_{XY}=-1$，表示完全的负相关。

例 3：(X, Y)服从平面上一单位圆内的均匀分布，求ρ_{XY}及 X 与 Y 的独立性。

$$f(x,y)=\begin{cases}\frac{1}{\pi}, & x^2+y^2\le 1\\ 0, & other\end{cases}$$

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dy=\frac{2}{\pi}\sqrt{1-x^2},-1\le x\le 1$$

同理：$f_Y(y)=\frac{2}{\pi}\sqrt{1-y^2},-1\le y\le 1$

$E(X)=\int_{-1}^{1}x\cdot\frac{2}{\pi}\sqrt{1-x^2}dx=0$, 奇函数在对称区间积分为 0

同理：$E(Y)=0$

$$E(XY)=\iint_D xy\frac{1}{\pi}dxdy=\frac{1}{\pi}\int_{-1}^{1}x\left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}ydy\right)dx=\frac{1}{\pi}\int_{-1}^{1}x\cdot 0dx=0$$

$Cov(X,Y)=E(XY)-E(X)E(Y)=0$
$\therefore \rho_{XY}=0$，代表 X 与 Y 不相关
因为$f(x,y)\neq f_X(x)\times f_Y(y)\therefore X,Y$不独立

对于二维正态分布$(X,Y)\sim N(\mu_1,{\sigma_1}^2;\mu_2,{\sigma_2}^2;\rho)$

$\rho_{XY}=\rho$

也就是二维正态分布的参数ρ就是 X 和 Y 的相关系数。因此二维正态分布完全可由 X、Y 各自的数学期望、方差和它们的相关系数确定。

🕮 4.4 矩&协方差矩阵

❀ 4.4.1 协方差矩阵

设 n 维随机变量$(X_1, X_2, \dots X_n)$，任意两个分量的协方差：

$$c_{ij} = Cov(X_i, X_j), \qquad i, j = 1,2,3 \dots n$$

都存在，则称矩阵：

$$\mathrm{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

是 n 维随机变量$(X_1, X_2, \dots X_n)$的协方差矩阵。

协方差矩阵的性质：

1) $c_{ii} = D(X_i), \mathrm{i} = 1,2,3, \dots \mathrm{n}$

2) $c_{ij} = c_{ji}$, 即：$\mathrm{C} = \mathrm{C}^{\mathrm{T}}$, C是对称矩阵

4.4.2 矩

随机变量 X，若：$\mathrm{E}(|X|^k) < +\infty$ (即存在)，则称：

$$\alpha_k = E(X^k), k = 1,2,3 \dots$$

是 X 的 k 阶原点矩，简称 k 阶矩。

若 E(X)存在，且$\mathrm{E}(|X - E(X)|^k) < +\infty$ (即存在)，则称：

$$\beta_k = E\{(X - E(X))^k\}, k = 1,2,3 \dots$$

是 X 的 k 阶中心矩。

数学期望 E(X)是 X 的一阶原点矩；方差 D(X)是 X 的二阶中心矩。
所以，矩是更具一般意义的数字特征。

4.5 数字特征的 Python 实现

① 二项分布

```
from scipy.stats import binom

rv = binom(10, 0.2)  # 二项分布
print(f'数学期望: {rv.mean()}')
print(f'方差: {rv.var()}')
for i in range(1, 5):
    print(f'{i}阶原点矩: {rv.moment(i)}')
print(rv.stats(moments='mvsk'))  # 获取统计数据
```

结果：

```
数学期望: 2.0
方差: 1.6
1阶原点矩: 2.0
2阶原点矩: 5.6
3阶原点矩: 18.560000000000002
4阶原点矩: 69.82400000000007
(array(2.), array(1.6), array(0.47434165), array(0.025))
```

② 泊松分布

```
from scipy.stats import poisson

rv = poisson(mu=5)  # 泊松分布 λ=5
print(f'数学期望: {rv.mean()}')
```

```
print(f'方差: {rv.var()}')
for i in range(1, 5):
    print(f'{i}阶原点矩: {rv.moment(i)}')
print(rv.stats(moments='mvsk'))  # 获取统计数据
```

结果：

```
数学期望: 5.0
方差: 5.0
1 阶原点矩: 5
2 阶原点矩: 30
3 阶原点矩: 205.0
4 阶原点矩: 1555.0
(array(5.), array(5.), array(0.4472136), array(0.2))
```

scipy.stats 内置离散型、连续型概率分布。只要更改分布，基本步骤类似。

③ 均匀分布

```
from scipy.stats import uniform

rv = uniform(loc=2, scale=6)  # 均匀分布 a=2,b=2+6=8
# ...
```

④ 指数分布

```
from scipy.stats import expon
import numpy as np
from scipy import integrate

rv = expon(scale=2)  # 指数分布 θ=1/λ=2
print(f'数学期望: {rv.mean()}')
print(f'方差: {rv.var()}')
for i in range(1, 5):
    print(f'{i}阶原点矩: {rv.moment(i)}')
print(rv.stats(moments='mvsk'))  # 获取统计数据
# ------------------------------------------------
# 使用积分求指数分布的数学期望和方差
F1 = lambda x: x/2*np.exp(-x/2)
F2 = lambda x: x**2/2*np.exp(-x/2)
# 积分结果是(积分,偏差)的元组
E1 = integrate.quad(F1, 0, np.inf)  # 0到正无穷积分
E2 = integrate.quad(F2, 0, np.inf)
print(f'数学期望=1阶原点矩={E1[0]}')
print(f'方差=2阶原点矩-1阶原点矩的平方={E2[0]-E1[0]**2}')
```

结果：

```
数学期望: 2.0
方差: 4.0
1 阶原点矩: 2.0
2 阶原点矩: 8.0
3 阶原点矩: 48.0
4 阶原点矩: 384.0
(array(2.), array(4.), array(2.), array(6.))
数学期望=1阶原点矩=1.9999999999999998
方差=2阶原点矩-1阶原点矩的平方=4.000000000000001
```

第 5 章　大数定律和中心极限定理

🕮　5.1　概率统计常用的两种收敛性

① 依概率收敛

设$X_1, X_2, \dots X_n \dots$是随机变量序列，X 是一随机变量或常数，对于任意$\varepsilon > 0$

$$\lim_{n\to\infty} P\{|X_n - X| < \varepsilon\} = 1 \quad or \quad \lim_{n\to\infty} P\{|X_n - X| \geq \varepsilon\} = 0$$

则称随机变量序列$X_1, X_2, \dots X_n \dots$依概率收敛于 X，记为：

$$X_n \xrightarrow{P} X$$

② 依分布收敛

设$X_1, X_2, \ldots X_n \ldots$是随机变量序列，$F_n(X)$是$X_n$的分布函数；

X 是一随机变量，F(X)是 X 的分布函数。

若在 F(X)的连续点 x 处都有：

$$\lim_{n\to\infty} F_n(x) = F(x)$$

则称随机变量序列$X_1, X_2, \ldots X_n \ldots$依分布收敛于 X，记为：

$$X_n \xrightarrow{L} X$$

这表明$\{X_n\}$以 X 的分布为极限分布。

而微积分的收敛性：依距离收敛

🕮 5.2 切比雪夫不等式

设随机变量 X 的 E(X)和 D(X)都存在，对于任意$\varepsilon > 0$，有：

$$\mathrm{P}\{|\mathrm{X} - \mathrm{E}(\mathrm{X})| \geq \varepsilon\} \leq \frac{D(X)}{\varepsilon^2} \quad or \quad \mathrm{P}\{|\mathrm{X} - \mathrm{E}(\mathrm{X})| < \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon^2}$$

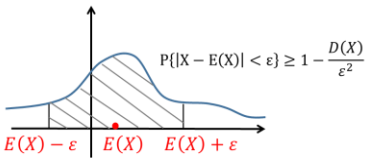
以连续型随机变量为例证明：

$$\mathrm{P}\{|\mathrm{X} - \mathrm{E}(\mathrm{X})| \geq \varepsilon\} = \int_{|\mathrm{X}-\mathrm{E}(\mathrm{X})|\geq\varepsilon} f(x)dx$$

因为$|\mathrm{X} - \mathrm{E}(\mathrm{X})| \geq \varepsilon$，所以$\frac{|\mathrm{X} - \mathrm{E}(\mathrm{X})|}{\varepsilon} \geq 1$

$$\text{原式} \leq \int_{|\mathrm{X}-\mathrm{E}(\mathrm{X})|\geq\varepsilon} \frac{\big(X - E(X)\big)^2}{\varepsilon^2} f(x)dx \leq \frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty} \big(X - E(X)\big)^2 f(x)dx = \frac{D(X)}{\varepsilon^2}$$

切比雪夫不等式用于未知估计分布概率，如在以 E(X)为中心的区间$(\mathrm{E}(\mathrm{X}) - \varepsilon, \mathrm{E}(\mathrm{X}) + \varepsilon)$发生的概率。但这个估计是比较粗糙的。



例 1：X 的分布未知，但$\mathrm{E}(\mathrm{X}) = \mu, \mathrm{D}(\mathrm{X}) = \sigma^2$。试估计：

$\mathrm{P}\{|\mathrm{X} - \mu| \geq 3\sigma\}$和$\mathrm{P}\{|\mathrm{X} - \mu| < 4\sigma\}$

$$\mathrm{P}\{|\mathrm{X} - \mu| \geq 3\sigma\} \leq \frac{D(X)}{(3\sigma)^2} = \frac{\sigma^2}{(3\sigma)^2} = \frac{1}{9}$$

$$\mathrm{P}\{|\mathrm{X} - \mu| < 4\sigma\} \geq 1 - \frac{\sigma^2}{(4\sigma)^2} = \frac{15}{16}$$

例 2：设$\{X_i\}$是独立同分布的随机变量序列，有相同的数学期望μ和方差σ^2。
证明：

$$\overline{\mathrm{X}} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

$$E(\overline{\mathrm{X}}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}n\mu = \mu$$

$$D(\overline{\mathrm{X}}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

代入切比雪夫不等式：

$$\mathrm{P}\{|\overline{\mathrm{X}} - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{n\varepsilon^2}$$

$$P\{|\overline{\mathrm{X}} - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{n\varepsilon^2}$$

当$\mathrm{n} \to \infty$时，$P\{|\overline{\mathrm{X}} - \mu| \geq \varepsilon\} \to 0$

此例的$\{X_i\}$称为服从弱大数定律。

🕮 5.3 大数定律

随机现象的统计规律：在相同条件下进行大量重复试验才能显现出来。
如抛一枚硬币，虽然不能准确预测每次的结果，但随着次数增加，正面出现的概率趋于 0.5。

设随机变量序列$\{X_i\}, \mathrm{i} = 1,2,3 \ldots, \mathrm{E}(X_i)$都存在，若对任意$\varepsilon > 0$，有：

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| < \varepsilon\} = 1$$

即：$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i) \xrightarrow{P} 0$
则称$\{X_i\}$服从大数定律。
这说明了算术平均值和频率(试验次数$\to \infty$)的稳定性。

① 切比雪夫大数定律
$\{X_i\}$是相互独立的随机变量序列，$\mathrm{E}(X_i), \mathrm{D}(X_i)$都存在，且$\{\mathrm{D}(X_i)\}$一致有界，则$\{X_i\}$服从大数定律。
一致有界：$\exists \mathrm{M} > 0, \forall i, \mathrm{D}(X_i) < \mathrm{M}$

② 独立同分布大数定律
$\{X_i\}$是独立同分布的随机变量序列，$\mathrm{E}(X_i) = \mu, \mathrm{D}(X_i) = \sigma^2$，则$\{X_i\}$服从大数定律。

② 贝努里大数定律
在 n 重贝努里试验中，A 事件出现 m 次，$p(\mathrm{A}) = p$，则对任意$\varepsilon > 0$

$$\lim_{n\to\infty} P\{|\frac{m}{n}-p|<\varepsilon\}=1$$

也就是 n 充分大时，频率m/n与概率p的偏差小于ε，这就是频率的稳定性。实际应用中，当试验次数很大时，常使用事件的频率代替事件的概率。

如果事件 A 的概率很小，则其发生的频率也很小。实际生活中，常常忽略概率很小的事件发生的可能性。

小概率事件原理：概率很小的事件在一次试验中几乎不会发生，认为它(在一次试验中)是不可能事件。

🕮 5.4 中心极限定理

设$\{X_i\}$是相互独立的随机变量序列，Z_n是$\{X_i\}$前 n 项和的标准化变量：

$$Z_n=\frac{\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}E(X_i)}{\sqrt{\sum_{i=1}^{n}D(X_i)}}$$

记Z_n的分布函数是$F_n(x)$。若：

$$\lim_{n\to\infty}F_n(x)=\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$

即：Z_n依分布收敛于标准正态变量

则称$\{X_i\}$服从中心极限定理。

现实中许多变量可以表示为诸多相互独立的随机变量的和，其中每个随机变量对总和的影响非常小，则这个总和近似服从正态分布。

① 独立同分布的中心极限定理

$\{X_i\}$是独立同分布的随机变量序列，$\mathrm{E}(X_i)=\mu,\mathrm{D}(X_i)=\sigma^2$，则：

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}\mathrm{P}\{\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\le x\}=\Phi(x)$$

② 棣莫弗-拉普拉斯中心极限定理

$\{Y_n\}$服从二项分布，$\sim\mathrm{B}(n,p),n=1,2,3\ldots$，则：

$$\lim_{n\to\infty}\mathrm{P}\left\{\frac{Y_n-np}{\sqrt{np(1-p)}}\le x\right\}=\Phi(x)$$

例：一批数量很大的产品，次品率为 0.02，从中抽取 10000 件，求次品数不超过 221 件的概率。

设 10000 件次品数是 Y，则Y~B(10000,0.02)

$\mathrm{E(Y)=np=200}$

$\mathrm{D(Y)=np(1-p)=196}$

$$\mathrm{P\{0\le Y\le 221\}}=\mathrm{P}\left\{\frac{0-E(Y)}{\sqrt{D(Y)}}\le\frac{Y-E(Y)}{\sqrt{D(Y)}}\le\frac{221-E(Y)}{\sqrt{D(Y)}}\right\}$$

$$= \mathrm{P}\left\{-\frac{200}{14} \leq \frac{Y-200}{14} \leq \frac{21}{14}\right\}$$
$$= \Phi(1.5) - \Phi(14.286) \approx \Phi(1.5) = 0.9332$$

中心极限定理奠定了正态分布的至高无上的地位。

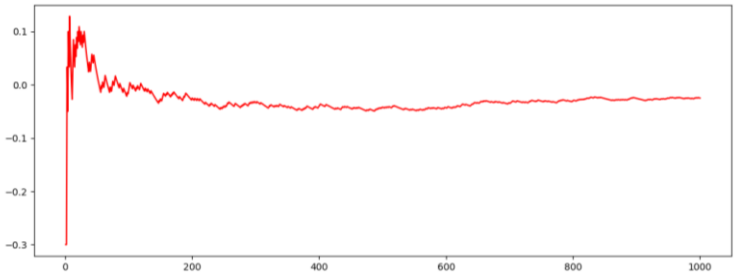
🕮 5.5 大数定律和中心极限定理的 Python 实现

① 贝努里大数定律

$$\lim_{n\to\infty} P\{|\frac{m}{n} - p| < \varepsilon\} = 1$$

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import bernoulli

n = 1000
x = np.arange(1, n+1)
p = 0.3
# 1000 次贝努里随机抽样,p=0.3
r = bernoulli.rvs(p, size=n)
y = []
m = 0
for i in range(n):
    if r[i] == 1:
        m += 1
    # y 表示 1000 次的频率-概率,当 x 足够大,y 趋于 0
    y.append(m/(i+1)-p)
plt.plot(x, y, color='red')
plt.show()
```

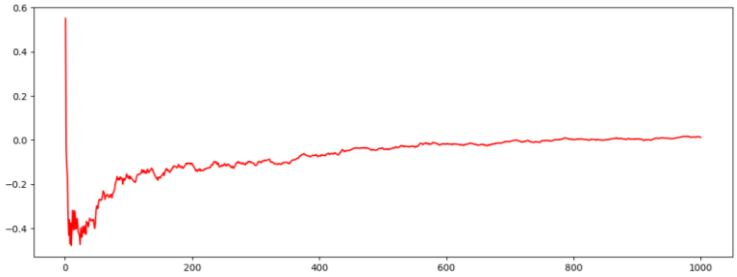
结果：



② 一般的大数定律

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import binom, poisson, norm

n = 1000
x = np.arange(1, n+1)
# 3 个分布期望都是 6
r1 = binom.rvs(10, 0.6, size=n)  # 二项分布抽样 n=10,p=0.6
r2 = poisson.rvs(mu=6, size=n)  # 泊松分布抽样 λ=6
r3 = norm.rvs(loc=6, size=n)  # 正态分布抽样 μ=6
y = []
rsum = 0
for i in range(n):
    rsum += r1[i]+r2[i]+r3[i]
    # y 表示 1000 次的频率-概率,当 x 足够大,y 趋于 0
    y.append(rsum/(i+1)/3-6)

plt.plot(x, y, color='red')
plt.show()
```

结果：

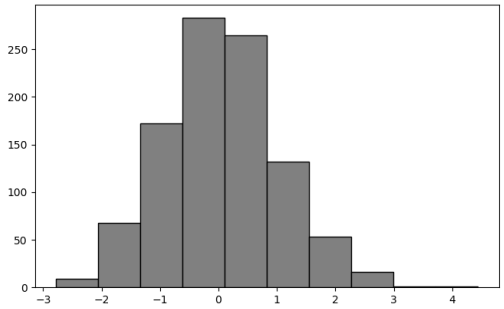


③ 独立同分布的中心极限定理：

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$$

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import expon

n = 100
y = []
for i in range(1000):  # 1000 次重复
    # 指数分布: λ=1,期望 μ=1,标准差 σ=1
    r = expon.rvs(scale=1, size=n)  # n 次抽样
    rsum = np.sum(r)  # n 次抽样求和
    # 标准化变换
    z = (rsum-n)/np.sqrt(n)
    y.append(z)
plt.hist(y, color='grey', edgecolor='black')
plt.show()
```

结果：

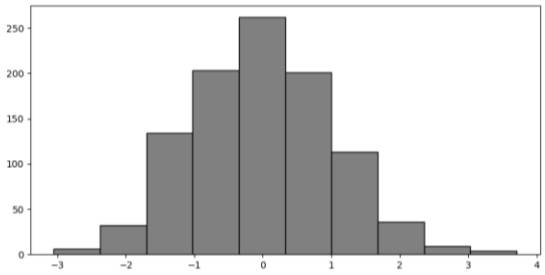


④ 棣莫弗-拉普拉斯中心极限定理：

$$\frac{X_n - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import binom

# 二项分布 B(n.p)
n, p = 100, 0.3
y = []
for i in range(1000):  # 1000 次重复
    r = binom.rvs(n, p)
    rsum = np.sum(r)  # n 次抽样求和
    # 标准化变换
    z = (rsum-n*p)/np.sqrt(n*p*(1-p))
    y.append(z)
plt.hist(y, color='grey', edgecolor='black')
plt.show()
```

结果：



第 6 章　抽样分布理论

🕮　6.1　随机样本&统计量

❀　6.1.1　数理统计引言

数理统计是数学的一个分支，主要研究如何收集和使用随机性数据。

1) 有效地收集数据：普查、抽样调查、安排试验
2) 有效地使用数据：分析和提取数据中的信息，对所研究问题作出统计推断(建立一个统计模型，给定某些准则)
3) 归纳性质：一般数学是演绎式推理；而统计是归纳推理。

由于归纳推理依赖与随机性数据，结果也具有不确定性。
统计就是提供归纳推理和计算不确定性程度的方法。

❀　6.1.2　数理统计基本概念

总体：由所研究问题的有关个体组成
按包含个体数目，分为：有限总体和无限总体
样本：从总体中抽取的一部分个体，不同抽样方法得到不同的样本

定义：
统计问题研究对象的全体称为总体，总体可以用一个随机变量及其概率分布描述。

样本的两重性：抽样前是随机变量；抽样后是具体的数。

简单随机样本：
1) 代表性：总体每个个体同等机会被抽入样本，即每个个体与总体分布相同
2) 独立性：样本每个个体取值不影响其他个体，即每个个体是相互独立的

此处的样本一般指简单随机样本，故简单随机样本在此处简称样本。

设总体为 X，分布函数为F(X), $(X_1, X_2, \dots, X_n)$是从总体中抽取的样本，则$(X_1, X_2, \dots X_n)$的联合分布为：

$$\mathrm{F}(X_1) \cdot \mathrm{F}(X_2) \cdots \mathrm{F}(X_n) = \prod_{i=1}^{n} \mathrm{F}(X_i)$$

若 X 的概率密度是$f(x)$，则$(X_1, X_2, \dots X_n)$的联合概率密度为：

$$f(x_1) \cdot f(x_2) \cdots f(x_n) = \prod_{i=1}^{n} f(x_i)$$

统计模型：指该问题所得样本的分布，如正态分布对应于正态模型
统计推断：
1) 样本的分布已知，对其所含未知参数的推断——参数统计推断。常见有参数估计、假设检验等。
2) 样本的分布未知——非参数统计推断。

❁ 6.1.3 统计量的概念

数理统计任务的要通过样本去推断总体，需要对样本的数据进行分析整理，构造出合适的量以解决总体的相关问题。

统计量是由样本所确定(计算)的量，是样本的函数，完全由样本决定。

注意：
1) 统计量只和样本有关，不能含有总体的未知参数
2) 由于样本具有两重性，因此统计量也具有两重性：既可以看成一个数，也可以看成随机变量。因此统计量有概率分布，是这统计推断的依据。
3) 统计量应最好的集中了样本的(与总体关联度最高)信息以解决总体相应问题。

❁ 6.1.4 常见的统计量

设 X 是总体，$(X_1, X_2, \dots, X_n)$是样本：
① 样本均值：

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

② 样本方差：

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$S_n{}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \text{ (样本原始方差)}$$

前者S^2也称为修正的方差，与"无偏性"有关。

③ 矩 (k 阶原点矩和 k 阶中心矩)

$$\alpha_k = \frac{1}{n}\sum_{i=1}^{n} X_i{}^k, k = 1,2,3 \dots$$

$$\beta_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k, k = 1,2,3 \dots$$

④ 将$X_1, X_2, \dots, X_n$按大小排列为：

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$$

则$(X_{(1)}, X_{(2)}, \dots, X_{(n)}$称为次序统计量。$X_{(1)}$为极小值；$X_{(n)}$为极大值。

🕮 6.2 三大抽样分布

❁ 6.2.1 卡方分布 (χ^2分布)

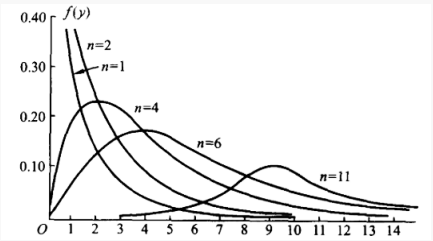
设$X_1, X_2, \dots, X_n$是来自总体N(0,1)的样本，称统计量：

$$Y = {X_1}^2 + {X_2}^2 + \cdots + {X_n}^2 = \sum_{i=1}^{n} {X_i}^2$$

服从自由度为 n 的χ^2分布，记为$Y \sim \chi^2(n)$。自由度是包含独立变量的个数。

χ^2分布的性质：

1) 概率密度函数$f(y)$比较复杂，图形如下：



图形只在第一象限，自由度 n 越大，密度曲线越对称。

2) 若$Y \sim \chi^2(n)$，有$\mathrm{E}(Y) = n,\ \mathrm{D}(\mathrm{Y}) = 2n$

3) 可加性：若$Y_1 \sim \chi^2(n_1), Y_2 \sim \chi^2(n_2)$，且相互独立，则$Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$

❁ 6.2.2 t 分布&分位数的定义

设$\mathrm{X} \sim \mathrm{N}(0,1), \mathrm{Y} \sim \chi^2(n)$，且相互独立，则称随机变量：

$$\mathrm{T} = \frac{X}{\sqrt{Y/\mathrm{n}}}$$

服从自由度为 n 的 t 分布，记为$\mathrm{T} \sim \mathrm{t}(n)$。

t 分布最早由戈塞(Gosset)发表，其笔名为 Student，故 t 分布又称学生分布。

t 分布的特点：

1) t 分布概率密度函数的特点：类似于N(0,1)

2) 若$\mathrm{T} \sim \mathrm{t}(n), n \geq 2$时，$\mathrm{E}(\mathrm{T}) = 0$；$n \geq 3$时，$\mathrm{D}(\mathrm{T}) = n/(n-2)$

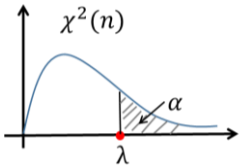
3) $n \to \infty$时，其极限分布为N(0,1)

分位点或分位数

一些分布的概率密度函数非常复杂，直接用其计算概率很麻烦。

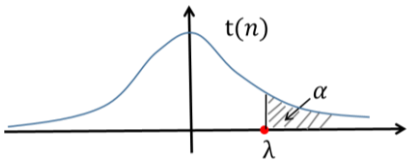
数学家将常见的值编成各种表(如正态分布表、卡方分布表等)，利用分位点的概念，通过查表就可得到结果。

如χ^2分布：



$\mathrm{P}(\mathrm{X} \geq \lambda) = \alpha$，记：$\lambda = \chi_\alpha^2(n)$为$\chi^2(n)$分布的上α分位点。

如 t 分布：



$P(X \geq \lambda) = \alpha$，记：$\lambda = t_\alpha(n)$为$t(n)$分布的上$\alpha$分位点。

⊛ 6.2.3 F 分布

设$X \sim \chi^2(m), Y \sim \chi^2(n)$，且相互独立，称随机变量：

$$F = \frac{X/m}{Y/n}$$

服从自由度为(m,n)的 F 分布，记为$F \sim F(m,n)$。

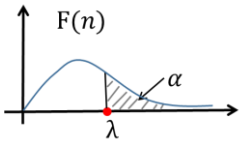
F 分布的性质：

1) 若$Z \sim F(m,n)$，则：

$$\frac{1}{Z} \sim F(n,m)$$

2) 若$T \sim t(n)$，则：$T^2 \sim F(1,n)$

3) F 分布的分位点：



$P(X \geq \lambda) = \alpha$，记：$\lambda = F_\alpha(m,n)$为$F(m,n)$分布的上$\alpha$分位点。

F 分布的上α分位点的性质：

$$F_{1-\alpha}(m,n) = \frac{1}{F_\alpha(n,m)}$$

🕮 6.3 抽样分布定理

正态总体样本均值和样本方差的分布。

设$X_1, X_2, \ldots, X_n$是来自正态总体$X \sim N(\mu, \sigma^2)$的样本，样本均值和样本方差为：

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

(1) $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\cdot(n\mu) = \mu$$

$$D(\overline{X}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\cdot(n\sigma^2) = \frac{\sigma^2}{n}$$

且$\overline{X}$服从正态分布，所以$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

如：测量某个元件长度时，多次测量取平均值能减少误差，就是基于这个原理(方差原来的 $1/n$)。

(2) $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

因为$X_i\sim N(\mu,\sigma^2)$，所以：

$$Y_i=\frac{X_i-\mu}{\sigma}\sim N(0,1)$$

$$\frac{(n-1)S^2}{\sigma^2}=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}=\sum_{i=1}^{n}(\frac{(X_i-\mu)-(\overline{X}-\mu)}{\sigma})^2=\sum_{i=1}^{n}(Y_i-\overline{Y})^2$$

$$=\sum_{i=1}^{n}({Y_i}^2+\overline{Y}^2-2Y_i\overline{Y})=\sum_{i=1}^{n}{Y_i}^2+n\overline{Y}^2-2\overline{Y}\sum_{i=1}^{n}Y_i$$

$$=\sum_{i=1}^{n}{Y_i}^2-n\overline{Y}^2$$

// 不会了...

(3) $\overline{X}$与S^2相互独立☆☆☆

(4) 由(1)(2)构造出：

$$T=\frac{\sqrt{n}(\overline{X}-\mu)}{S}\sim t(n-1)$$

由(1)$\overline{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$得出$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，和(2) $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

根据 t 分布定义得到：

$$T=\frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}}=\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$$

(5) $X_1,X_2,\dots,X_m$是来自正态总体$X\sim N(\mu_1,{\sigma_1}^2)$的样本；$Y_1,Y_2,\dots,Y_n$是来自正态总体$Y\sim N(\mu_2,{\sigma_2}^2)$的样本，两个样本相互独立。样本均值和方差为：$\overline{X},{S_X}^2,\overline{Y},{S_Y}^2$。

1° $F=\frac{{S_X}^2/{S_Y}^2}{{\sigma_1}^2/{\sigma_2}^2}\sim F(m-1,n-1)$，将 X、Y 分别代入(2)相除即得

2° 当${\sigma_1}^2={\sigma_2}^2$时：

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S^2}\sqrt{\frac{mn}{m+n}}\sim t(m+n-2)$$

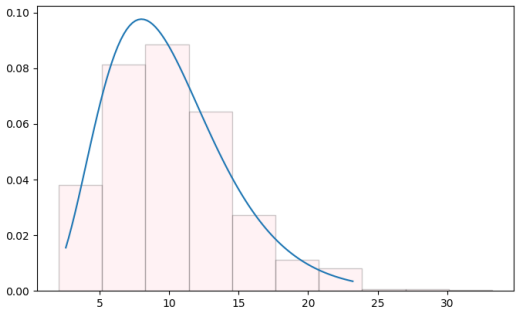
其中$(m+n-2)S^2=(m-1){S_X}^2+(n-1){S_Y}^2$

📖 6.4 抽样分布的 Python 实现

① 卡方分布

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import chi2, norm

fig, ax = plt.subplots(1, 1)
df = 10  # 自由度
x = np.linspace(chi2.ppf(0.01, df), chi2.ppf(0.99, df), 100)
# 画出 scipy 内置的卡方分布曲线,自由度 df=10
ax.plot(x, chi2.pdf(x, df))

# 根据卡方分布定义,模拟 1000 个样本近似,每次样本抽 10 个正态分布随机变量
N, n = 1000, 10
y = []
for i in range(N):
    chi2_ret = 0
    r = norm.rvs(size=n)
    for j in r:
        chi2_ret += j**2
    y.append(chi2_ret)

ax.hist(y, density=True, color='pink', alpha=0.2, edgecolor='black')
plt.show()
```

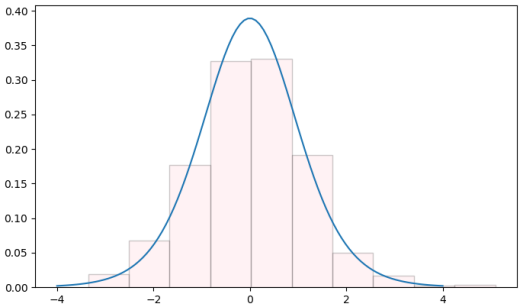
结果：



② t 分布

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import chi2, norm, t

def t_dis():
    fig, ax = plt.subplots(1, 1)
    df = 10  # 自由度
    x = np.linspace(-4, 4, 100)
    # scipy 内置的 t 分布曲线,自由度 df=10
    ax.plot(x, t.pdf(x, df))

    N = 1000
    y = []
    # 根据定义构造 t 分布
    for i in range(N):
        rx = norm.rvs()
        ry = chi2.rvs(df)
        ret = rx/np.sqrt(ry/df)
        y.append(ret)
    # 画出模拟的直方图和标准曲线的叠加图
    ax.hist(y, density=True, color='pink', alpha=0.2, edgecolor='black')
    plt.show()
```

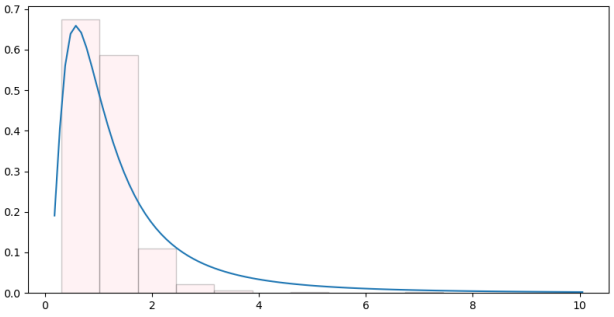
结果：



③ F 分布

```
def F_dis():
    fig, ax = plt.subplots(1, 1)
    dfm, dfn = 10, 5  # 自由度
    x = np.linspace(f.ppf(0.01, dfm, dfn), f.ppf(0.99, dfm, dfn), 100)
    # scipy 内置的 F 分布曲线~F(10,5)
    ax.plot(x, f.pdf(x, dfm, dfn))

    N = 1000
    y = []
    # 根据定义构造 F 分布
    for i in range(N):
        rx = chi2.rvs(dfm)
        ry = chi2.rvs(dfn)
        ret = np.sqrt((rx/dfm)/(ry/dfn))
        y.append(ret)
    # 画出模拟的直方图和标准曲线的叠加图
    ax.hist(y, density=True, color='pink', alpha=0.2, edgecolor='black')
    plt.show()
```

结果：

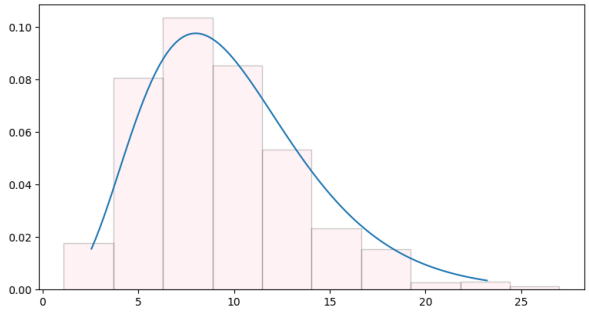


④ 样本方差抽样分布

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

```
def S2():
    fig, ax = plt.subplots(1, 1)
    df = 10
    x = np.linspace(chi2.ppf(0.01, df), chi2.ppf(0.99, df), 100)
    ax.plot(x, chi2.pdf(x, df))

    N = 1000
    y = []
    for i in range(N):
        # 每次样本抽取 n(即 df+1)个正态分布(μ=5,σ=2)随机变量
        sigma = 2
        n = df+1
        r = norm.rvs(loc=5, scale=sigma, size=n)
        # np.var(r)为求样本的方差
        ret = (n-1)*np.var(r)/(sigma**2)
        y.append(ret)
    ax.hist(y, density=True, color='pink', alpha=0.2, edgecolor='black')
    plt.show()
```

结果：

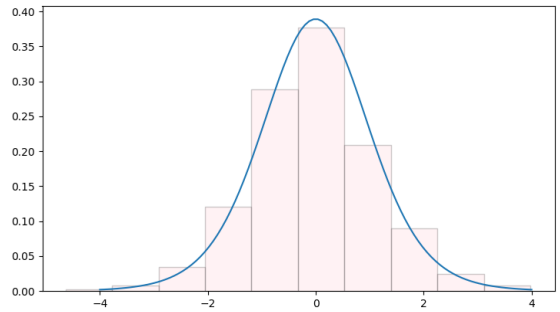


⑤ 抽样分布-t

$$\frac{\sqrt{n}(\overline{X}-\mu)}{S}\sim t(n-1)$$

```
def t1():
    fig, ax = plt.subplots(1, 1)
    df = 10
    x = np.linspace(-4, 4, 100)
    # scipy 内置的 t 分布曲线,自由度 df=10
    ax.plot(x, t.pdf(x, df))

    N = 1000
    y = []
    for i in range(N):
        # μ=5, σ=2 的正态分布,每次样本抽取 n 次
        miu, sigma, n = 5, 2, df+1
        r = norm.rvs(loc=miu, scale=sigma, size=n)
        ret = (np.mean(r)-miu)/np.sqrt(np.var(r)/n)
        y.append(ret)
    # 画出模拟的直方图和标准曲线的叠加图
    ax.hist(y, density=True, color='pink', alpha=0.2, edgecolor='black')
    plt.show()
```

结果：



第 7 章　描述统计

🕮　7.1　变量定义

为根据不同度量水平将变量分为：

1) 定类变量(Nominal Variable)，又称名义变量，通常用于代表不同的分类。如 1 代表男 0 代表女；C、Java、Python 定义为 0、1、2；

定类变量数学关系只有等于(=)和不等于(≠)。

2) 定序变量(Ordinal Variable)：采用数字表示顺序。每个分类有差别，还有等级之分。如优等品、合格品、次品，分别记为 2、1、0。

定序变量数学关系有=、≠、<、>等。注意：不一定是等距的。

3) 定距变量(Interval Variable)，也称间隔变量，描述事物类别或次序间距。

定距变量中，0 是强行规定的，不代表完全没有的意思，如温度 0℃不代表没有温度，只是代表水结冰的特定温度。

4) 定比变量(Ratio Variable)：在定距变量基础上，扩展可作为比率的基数而成。需要统一的单位，如 m、cm、kg、s 等。身高体重都是定比变量，和定距变量的一个根本区别是定比变量的 0 代表完全没有。

变量类型主要基于相应的变量数值可做的有意义数学运算区分的。
4 种变量的计量层次由低级到高级、由粗略到精确递进。高层次变量有低层次变量全部特性，反之没有。

变量类型	特点	关系和运算	举例
定类变量	无顺序分类	$=,\neq$	性别
定序变量	有顺序分类	$>,<$	严重级别
定距变量	等间隔但没有绝对零点	$+,-,\times,\div$	温度、标准成绩
定比变量	等间隔且有绝对零点	$x/y, x, y$ 是同类数据	身高、年龄、数量

🕮 7.2 统计图表

使用统计图表展示数据更加直观，令人信服。

❀ 7.2.1 分类型数据统计图表

分类型数据来自于定类或定序变量，主要用于事物的分类描述，计算每一类别的频数、频率、累积频率等。

① 频数分布表

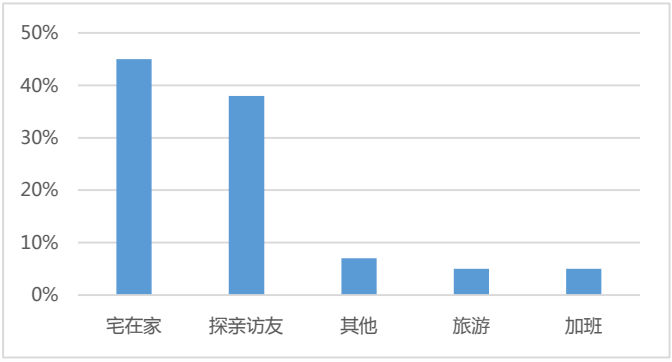
将每个类别及其频数以表格形式表现。频数是每个类别数据出现的次数。

如：黄金周过节方式的网络调查：

过节方式	频数	频率	百分比
宅在家	7853	0.45	45%
探亲访友	6632	0.38	38%
旅游	873	0.05	5%
加班	873	0.05	5%
其他	1221	0.07	7%
合计	17452	1	100%

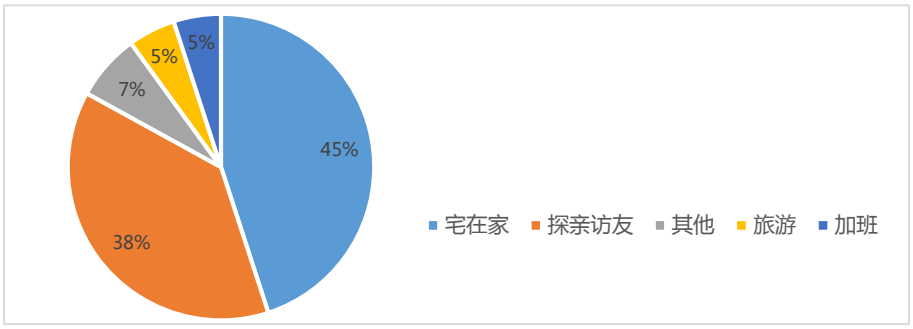
② 柱状图

等宽的垂直图条按类别分组显示，柱状高度描述统计量的值。用于各组之间的比较。水平显示的称为条状图。



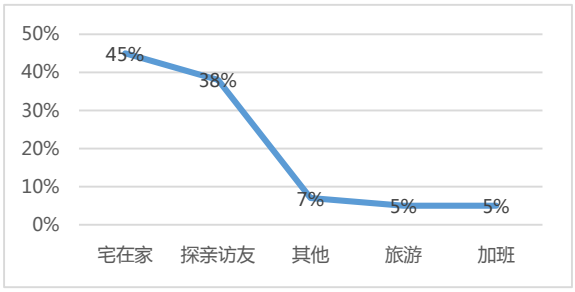
③ 饼图

主要描述频数、频率、百分比之间相对关系，研究结构性问题。
饼图具有较好的视觉效果，但表现的信息相对较少。



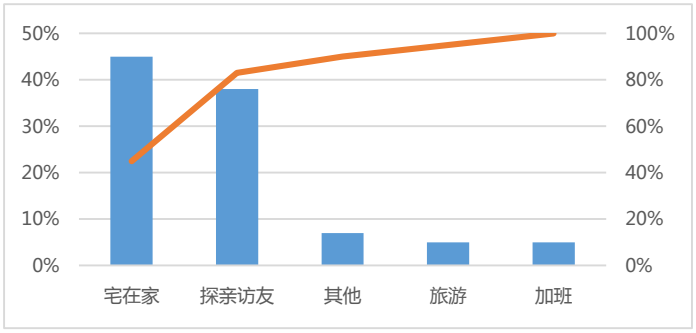
④ 折线图

用直线将数据点连接起来，显示数据的变化趋势。适用于显示相等时间间隔下数据的趋势。



⑤ 帕累托图

图中包含柱状图和折线图。柱状图按发生的频率降序排列，而折线图显示了累积频率。帕累托图来源于帕累托定律(即 28 定律)：80%的结果取决于 20%的原因。柱状图上方的折线是帕累托曲线。



❀ 7.2.2 数值型数据统计图表

数值型数据包括定距和定比数据。通常要进行数据分组。分组再计算各组的频数，就形成了频数分布表。

组数 k 可以根据 Sturges 提出的经验公式：

$$k = 1 + \log_2 n$$

确定组距的简单方法就是均分，即(max-min)/k。

① 直方图

矩形的宽度和高度表示频数分布。组距一般相等，也可不等。

直方图和柱状图类似，但直方图的横坐标是连续的数据，是刻度轴；而柱状图是分类轴。

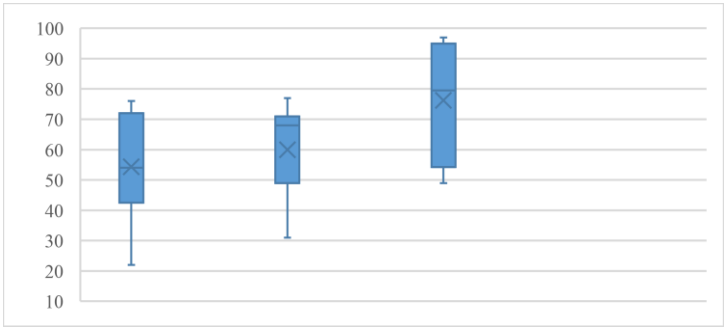
柱状图使用高度表示各类的频数，而直方图使用面积表示。

② 茎叶图

③ 盒状图

又称箱形图，用来表示数据集是位置和变异信息。

最上方的横线表示最大值；最下方的横表示最小值；方体中的三个横线从上到下为第三个四分位数(Q_3)、中位数(Q_2 或 M)，第一个四分位数(Q_1)。

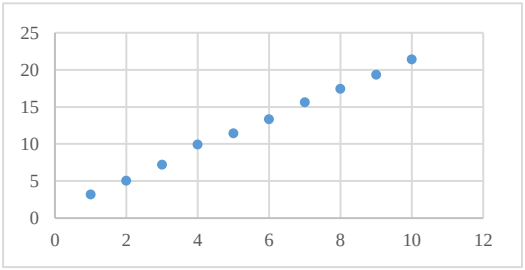


7.2.3 多变量数据统计图表

考虑来自两个(或多个)变量(X, Y)样本数据，

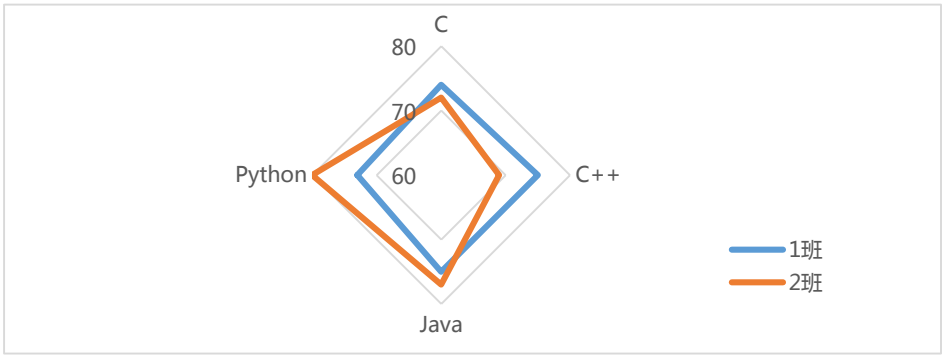
① 散点图

完全线性相关、线性相关、非线性相关、不相关。



② 雷达图

也称蜘蛛图，是显示多个变量的常用统计图形。



雷达图用于显示和比较不同批次的数据和不同变量的数值总和，或显示比较不同数据间的相似性。

③ 列联表

观测数据按两个或多个属性(定类或定序)分类时列出的频数表。

性别/等级	不及格	及格	中	良	优	合计
男生	4	13	8	4	3	32
女生	1	5	3	5	4	18
合计	5	18	11	9	7	50

🕮 7.3 数据汇总

❁ 7.3.1 集中趋势度量

集中趋势度量(Central Tendency)反映数据的平均水平或数据的中心值。

1) 众数：一组数据出现次数最多的那个数值。

2) 中位数：数据从小到大排序后，在中间位置的数值。

3) 四分位数：数据从小到大排序后四等分，第一、第二、第三四分位数为 Q_1、Q_2、Q_3；第二四分位数 Q2 就是中位数。

$$Q_i = \frac{i}{4}(n+1)\text{对应位置的数据}$$

如果求得位置不是整数，四舍六入，五取左右算术平均。

4) 算术平均值：也叫平均值或均值。

5) 加权平均值：每个数据和对应权重相乘，再求和。

6) 几何平均值：n 个数据相乘的 n 次方根。主要用于描述平均比率。

$$\overline{X}_G = \sqrt[n]{\prod_{i=1}^{n} X_i}$$

7) 调和平均值：每个数据倒数的平均的倒数。

$$\overline{X}_H = \frac{n}{\sum_{i=1}^{n}\frac{1}{X_i}}$$

❁ 7.3.2 离散趋势度量

离散(变异)趋势度量(Variation Tendency)：数据的离散程度。

① 全距，也叫极差，样本观察值的最大值和最小值的差。

② 内距：全称内四分位距，用来，描述中间 50%的样本观察值的离散程度，是第三四分位数 Q_3 与第一四分位数 Q_1 的差。

③ 方差与标准差

④ 变异系数：也称标准差率，是样本标准差的一种相对度量。比较多个数据离散趋势，如果单位不同或平均值差异较大，需要使用变异系数比较。

变异系数记为 CV，是样本标准差除以样本均值的百分比形式：

$$\text{CV} = \frac{S}{\overline{X}}100\%$$

❁ 7.3.3 形态度量(Shape Tendency)

① 偏度：描述数据的对称性，是三阶矩的一个形式。

偏度s^3为：

$$s^3 = \frac{B_3}{{B_2}^{1.5}} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^3}{(\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n})^{1.5}}$$

如偏度=0，说明数据是对称的，此时平均值=中位数；
偏度>0，称数据是正偏的，平均值>中位数；
偏度<0，称数据是负偏的，平均值<中位数；

实际应用，可以采用皮尔逊近似公式计算偏度：

$$s'^3 = \frac{3(\overline{X} - M_e)}{s}$$

其中M_e是中位数，s 是标准差。

② 峰度：描述数据的尖峰长度，即数据是否集中在均值附近。
根据尖峰程度分为 3 种形态：尖顶峰度、平顶峰度、标准峰度。
通常与正态分布的概率密度图进行比较，比正态分布更尖的为尖顶峰度；比正态分布更平的为平顶峰度；与正态分布差不多的就是标准峰度。

偏度s^4为：

$$s^4 = \frac{B_4}{{B_2}^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^4}{(\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n})^2}$$

标准正态分布的峰度为 3；$s^4 > 3$为尖顶；$s^4 < 3$为平顶。
$s^4 = 1.8$时数据分布接近矩形形态；$s^4 < 1.8$时，呈现两端翘起 U 字形态。

峰度近似计算公式：

$$s'^4 = 3 + \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

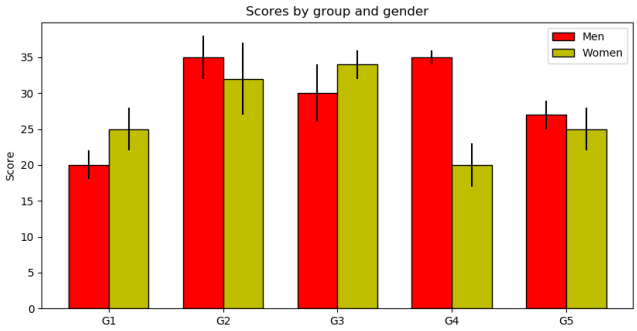
其中，Q_1, Q_3为 25%、75%的分位数；P_{10}, P_{90}为 10%、90%的分位数。

🕮 7.4 统计图表的 Python 实现

① 柱状图 bar

```
import matplotlib.pyplot as plt
import numpy as np

def bar_show():
    men_means = (20, 35, 30, 35, 27)
    men_std = (2, 3, 4, 1, 2)
    women_means = (25, 32, 34, 20, 25)
    woman_std = (3, 5, 2, 3, 3)
    ind = np.arange(5)
    width = 0.35
    fig, ax = plt.subplots()
    rects1 = ax.bar(ind, men_means, width, color='r',
                    edgecolor='black', yerr=men_std)
    rects2 = ax.bar(ind+width, women_means, width, color='y',
                    edgecolor='black', yerr=woman_std)
    ax.set_ylabel('Score')  # y 轴标签
    ax.set_title('Scores by group and gender')  # 标题
    ax.set_xticks(ind+width/2)  # x 轴的小锯齿
    ax.set_xticklabels(('G1', 'G2', 'G3', 'G4', 'G5'))  # x 轴每个记号的标签
    ax.legend((rects1[0], rects2[0]), ('Men', 'Women'))  # 显示图例
    plt.show()
```

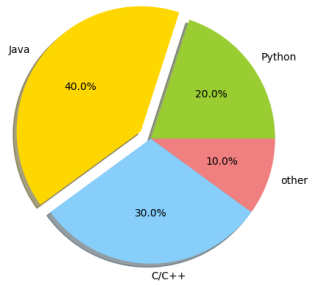
结果：



② 饼图 pie

```
def pie_show():
    labels = ['Python', 'Java', 'C/C++', 'other']
    sizes = [20, 40, 30, 10]
    colors = ['yellowgreen', 'gold', 'lightskyblue', 'lightcoral']
    explode = [0, 0.1, 0, 0]  # 饼图的凸出显示
    plt.pie(sizes, explode=explode, labels=labels,
            colors=colors, autopct='%1.1f%%', shadow=True)
    plt.axis('equal')
    plt.show()
```

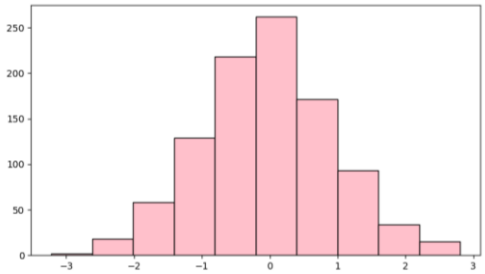
结果：



③ 直方图 hist

```
from scipy.stats import norm

def hist_show():
    X = norm(loc=0, scale=1)
    plt.hist(X.rvs(size=1000), color='pink', edgecolor='black')
    plt.show()
```

结果：



④ 折线图 plot

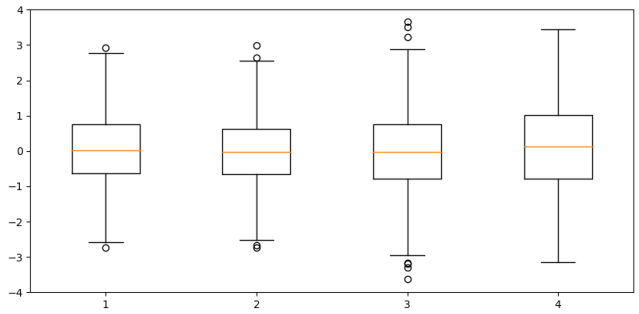
```
def plot_show():
    x = np.linspace(-10, 10, 100)
    y = np.random.normal(0, 1, size=(100,))
    plt.plot(x, y, color='red')
    plt.show()
```

⑤ 盒状图 boxplot

```
def boxplot_show():
    inc = 0.1
```

```
    e1 = np.random.normal(0, 1, size=(500,))
    e2 = np.random.normal(0, 1, size=(500,))
    e3 = np.random.normal(0, 1+inc, size=(500,))
    e4 = np.random.normal(0, 1+inc*2, size=(500,))
    plt.boxplot([e1, e2, e3, e4])
    plt.show()
```

结果：



第 8 章　参数估计

🕮 8.1　点估计

设总体X的分布函数为$F(x,\theta_1,\theta_2,\cdots\theta_k,)$，分布已确定，但有未知参数$\theta_1,\theta_2,\cdots\theta_k$，称为待估参数，$X_1,X_2,\ldots,X_n$是来自 X 的样本，相应样本值为：$x_1,x_2,\cdots,x_n$。
对未知参数θ_i，构造一个合适的统计量：$\hat{\theta}_i(X_1,X_2,\cdots,X_n)$
其对应值：$\hat{\theta}_i(x_1,x_2,\cdots,x_n)$作为$\theta_i$的估计，这种估计称为点估计。

点估计常用有：矩估计法和极大似然估计法。

❁ 8.1.1　矩估计 (皮尔逊提出)

将样本的矩作为总体矩的估计(假设取到的样本较好)。
总体矩：$a_k=E(X^k),b_k=E((X-E(X))^k)$

样本矩：$\alpha_k=\frac{1}{n}\sum_{i=1}^{n}{X_i}^k,\beta_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$

由大数定律可得：

$$\alpha_k\xrightarrow{P}a_k,\beta_k\xrightarrow{P}b_k$$

$\therefore\alpha_k=a_k,\beta_k=b_k$

注意：
1) 有几个未知参数就建立几个方程；
2) 能用低阶矩处理就不用高阶矩；
3) 矩估计不唯一。

例 1：总体期望和方差都存在，记为μ,σ^2，但其值未知，求μ,σ^2矩估计。

$\mu=\mathrm{E}(\mathrm{X})=a_1$
$\sigma^2=E\{[X-E(X)]^2\}=b_2$

$\hat{\mu}=\alpha_1=\overline{X}$

$\hat{\sigma}^2=\beta_2={S_n}^2$

例 2：设泊松分布π(λ)的λ未知，求λ的矩估计。

$\mu = \mathrm{E}(X) = a_1$

$\sigma^2 = E\{[X - E(X)]^2\} = b_2$

$\hat{\mu} = \hat{\lambda} = \alpha_1 = \overline{X}$

$\hat{\sigma}^2 = \hat{\lambda} = \beta_2 = {S_n}^2$

两个都可以，使用低阶矩，即：$\hat{\lambda} = \overline{X}$

例 3：设$X \sim U(a,b)$，求的a,b矩估计。

$$\mathrm{E}(X) = a_1 = \frac{a+b}{2}, D(X) = b_2 = \frac{(b-a)^2}{12}$$

$$\alpha_1 = \overline{X}, \beta_2 = {S_n}^2$$

$$\therefore \begin{cases} \dfrac{a+b}{2} = \overline{X} \\ \dfrac{(b-a)^2}{12} = {S_n}^2 \end{cases}$$

$$\Rightarrow \begin{cases} \hat{a} = \overline{X} - \sqrt{3}S_n \\ \hat{b} = \overline{X} + \sqrt{3}S_n \end{cases}$$

注意：矩估计简便易行，但估计的效果不好。

8.1.2 极大似然估计 (费歇尔提出)

应用最广泛的点估计方法，其基本思想：最大可能性原则。

总体分布已知，含有 1 个或几个未知参数$\theta_1, \theta_2, \cdots \theta_k$，$X_1, X_2, \ldots, X_n$是来自总体的样本，$x_1, x_2, \cdots, x_n$是样本的观测值。

根据最大可能性原则，相当于事件$\{X_1 = x_1, \cdots, X_n = x_n\}$是最可能发生的事件，认为有最大的概率。

样本联合概率密度为：

$$\prod_{i=1}^{n} f(x, \theta_1, \theta_2, \cdots \theta_k)$$

记为：$L(x_1, x_2, \cdots, x_n, \theta_1, \theta_2, \cdots \theta_k)$，称为似然函数。

L 取最大值的$\hat{\theta}_1, \hat{\theta}_2, \cdots \hat{\theta}_k$称为未知参数$\theta_1, \theta_2, \cdots \theta_k$的极大似然估计值；

统计量$\theta_i = \theta_i(X_1, X_2, \cdots, X_n)$是$\theta_i$的极大似然估计量。

例 1：$X_1, X_2, \ldots, X_n$是来自正态总体$N(\mu, \sigma^2)$，求μ, σ^2的极大似然估计。

设$x_1, x_2, \cdots, x_n$是观测值

$$\mathrm{L} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$$

取对数(不影响单调性)：

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\begin{cases}\dfrac{\partial \ln L}{\partial \mu} = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \\ \dfrac{\partial \ln L}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0\end{cases}$$

$$\therefore \begin{cases}\hat{\mu} = \dfrac{1}{n}\sum_{i=1}^{n}x_i = \overline{X} \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i-\overline{X})^2 = {S_n}^2\end{cases}$$

例 2：$X_1, X_2, \ldots, X_n$是来自泊松分布π(λ)，λ未知，求λ的极大似然估计。

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0,1,2,\ldots$$

$$L = \prod_{i=1}^{n}\left(\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}\right)$$

$$\ln L = \left(\sum_{i=1}^{n}x_i\right)\ln\lambda - n\lambda - \sum_{i=1}^{n}\ln(x_i!)$$

$$\frac{d\ln L}{d\lambda} = \frac{1}{\lambda}\sum_{i=1}^{n}x_i - n = 0$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{n}x_i}{n} = \overline{X}$$

$\frac{d^2\ln L}{d\lambda^2} = -\frac{1}{\lambda^2}\sum_{i=1}^{n}x_i < 0$，即$\lambda = \overline{X}$时，$\ln L$取最大值。

函数一阶导数为 0 的点，称为驻点(stationary point)，又称平稳点、稳定点或临界点(Critical Point)。多元函数的驻点是所有一阶偏导数都为 0 的点。

驻点不一定是极值点：如$y = x^3$在$x = 0$处导数为 0，是驻点，但没有极值；
极值点也不一定是驻点：如$y = |x|$在$x = 0$处不可导，但是极小值点。
二阶导数<0 的驻点为极大值点。

🕮 8.2 估计量的评价标准

① 无偏性

设$\hat{\theta} = \theta(X_1, X_2, \cdots, X_n)$是未知参数$\theta$的一个估计量。如果：

$$E(\hat{\theta}) = \theta$$

则称$\hat{\theta}$是θ的一个无偏估计量；
否则称为有偏估计量，此时称$E(\hat{\theta}-\theta)$是估计量$\hat{\theta}$的偏差，也叫$\hat{\theta}$的系统误差。

无偏估计就意味着：没有系统误差。

例：设总体的 k 阶原点矩$\alpha_k = E(X^k)$存在，证明样本的 k 阶原点矩$\alpha_k = \frac{1}{n}\sum_{i=1}^{n}{X_i}^k$

是a_k的无偏估计。

$$E(\alpha_k)=E\left(\frac{1}{n}\sum_{i=1}^{n}{X_i}^k\right)=\frac{1}{n}E\left(\sum_{i=1}^{n}{X_i}^k\right)$$

因为样本X_i的分布与总体X相同，所以：

$$E(\alpha_k)=\frac{1}{n}E\left(\sum_{i=1}^{n}X^k\right)=E(X^k)=a_k$$

思考题：设总体的方差存在，记为σ^2，则$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$是$\sigma^2$的无偏估计。

$$E(S^2)=E\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)$$

$$=\frac{1}{n-1}E\left(\sum_{i=1}^{n}\left({X_i}^2+\overline{X}^2-2X_i\overline{X}\right)\right)$$

$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}E\left({X_i}^2\right)+\sum_{i=1}^{n}E\left(\overline{X}^2\right)-2E(\sum_{i=1}^{n}X_i\overline{X})\right)$$

$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}E\left({X_i}^2\right)+nE\left(\overline{X}^2\right)-2E(n\overline{X}\cdot\overline{X})\right)$$

$$=\frac{1}{n-1}\left(nE(X^2)-nE\left(\overline{X}^2\right)\right)=\frac{1}{n-1}\left(n(\mu^2+\sigma^2)-n(\mu^2+\frac{\sigma^2}{n})\right)$$

$$=\frac{1}{n-1}(n\sigma^2-\sigma^2)=\sigma^2$$

② 有效性

设$\hat{\theta}_1(X_1,X_2,\cdots,X_n)$，$\hat{\theta}_2(X_1,X_2,\cdots,X_n)$是参数$\theta$的两个无偏估计。

如果$D(\hat{\theta}_1)\le D(\hat{\theta}_2)$，则称$\hat{\theta}_1$比$\hat{\theta}_2$更有效。$\hat{\theta}_1$的波动性小，稳定性好。

如果$\hat{\theta}_0$是θ的无偏估计，$\hat{\theta}$是θ的任何一个无偏估计，有$D(\hat{\theta}_0)\le D(\hat{\theta})$，则称$\hat{\theta}_0$是$\theta$的最小方差无偏估计。

③ 一致性(相合性)

估计量与样本的容量 n 有关，随着 n 的增加，估计量的值应该越来越趋向于被估计参数的真值。

当$n\to\infty$，$\hat{\theta}$依概率收敛于θ，则$\hat{\theta}$是θ的一致(相合)估计量。

一致性是对估计量的**根本要求**。

大数定律可以证明常见的矩估计量是一致的。

🕮 8.3 区间估计

❀ 8.3.1 区间估计引言

点估计只是未知参数的一个近似值，难以判断其误差大小。

区间估计给出一个范围，并在要求的可靠程度下保证此范围包含了未知参数。

设θ是总体分布的一个未知参数，$X_1, X_2, \cdots, X_n$为样本。
两个统计量$\hat{\theta}_1(X_1, X_2, \cdots, X_n)$和$\hat{\theta}_2(X_1, X_2, \cdots, X_n)$，其中$\hat{\theta}_1 \le \hat{\theta}_2$，以它们为端点的区间$[\hat{\theta}_1, \hat{\theta}_2]$，满足：
1) 可信度：此随机区间$[\hat{\theta}_1, \hat{\theta}_2]$以多大的概率包含了未知参数$\theta$；
2) 精度：区间长度$\hat{\theta}_2 - \hat{\theta}_1$尽可能小。

希望这两个要求能得到保证，但它们是互相矛盾的，不能同时满足：如一个实数一定在$(-\infty, +\infty)$区间上，但精度为 0。

Neyman 提出原则：先确定一个能接受的可靠度，在此前提下尽量提高精度。
定义：给定一个很小的数$\alpha > 0$，若对于参数θ的所有可能取值，都有：

$$P\{\hat{\theta}_1(X_1, X_2, \cdots, X_n) \le \theta \le \hat{\theta}_2(X_1, X_2, \cdots, X_n)\} = 1 - \alpha$$

则称随机区间$[\hat{\theta}_1, \hat{\theta}_2]$是$\theta$的置信度为$1-\alpha$的置信区间。

区间估计主要工作：给定置信度后，寻找估计精度尽可能高的置信区间。

⊛ 8.3.2 枢纽变量法

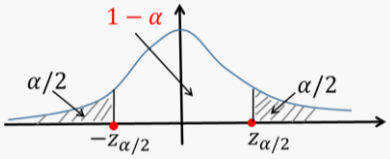
区间估计最常用的方法是枢纽变量法。

例：设$X_1, X_2, \ldots, X_n$来自正态总体$N(\mu, \sigma^2)$，σ^2为已知参数，求未知参数μ的置信度为$1-\alpha$的置信区间。

1) 选择样本均值$\overline{X}$来估计μ
2) 构造统计量(具有确定分布)：

$$Y = \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

即：$P\{-z_{\alpha/2} \le Y \le z_{\alpha/2}\} = 1 - \alpha$



3) 上式改写为：

$$P\{-z_{\alpha/2} \le \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} \le z_{\alpha/2}\} = P\{\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \le \mu \le \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\} = 1 - \alpha$$

$[\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}]$就是置区间信

根据例子总结枢轴变量法的步骤：
1) 找出一个未知参数θ良好的估计量；

2) 构造一个包含θ的函数 Y(枢轴变量)，Y 不能包含其他未知数，且其分布明确；
3) 将关于 Y 的不等式(范围)改写为关于θ的不等式；
4) 查表确定需要的分位点，即找出置信区间。

🕮 8.4 正态总体均值和方差的区间估计

❀ 8.4.1 单总体的常见四种区间估计

① 正态总体$N(\mu,\sigma^2)$，σ^2已知，μ的置信区间：

$$(\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2})$$

② 正态总体$N(\mu,\sigma^2)$，σ^2未知，μ的置信区间：

选取枢轴量：$T=\frac{\overline{X}-\mu}{\sqrt{S^2/n}}\sim t(n-1)$ // P50

$$P\{-t_{\alpha/2}(n-1)<\frac{\overline{X}-\mu}{\sqrt{S^2/n}}<t_{\alpha/2}(n-1)\}$$

$$=P\{\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)<\mu<\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\}=1-\alpha$$

所以μ的置信区间：

$$(\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1),\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1))$$

③ 正态总体$N(\mu,\sigma^2)$，μ已知，σ^2的置信区间：

$$(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)},\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)})$$

因为$\frac{X-\mu}{\sigma}\sim N(0,1)$，取枢轴量：

$$Q=\sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2\sim\chi^2(n)$$

④ 正态总体$N(\mu,\sigma^2)$，μ未知，σ^2的置信区间：

$$(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)})$$

选取枢轴量：$K=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ // P50

例：某工厂生产一批滚珠，直径 X 服从正态分布$N(\mu,\sigma^2)$，随机抽取 6 件，测得直径为：15.1、14.8、15.2、14.9、14.6、15.1。
(1) 若$\sigma^2=0.06$，求μ的置信区间；(2) 若σ^2未知，求μ的置信区间；(3) 求σ^2的置信区间。置信度均为 0.95。

(1)$\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$，即 $N(\mu,0.01)$

$\bar{x} = \frac{1}{6}\sum_{i=1}^{6} x_i = 14.95$

$\frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} = \frac{14.95 - \mu}{0.1} \sim N(0,1)$

查表得：$\Phi(1.96) = 0.975 \quad \therefore z_{0.025} = 1.96$

μ的置信区间为：$(14.95 \pm 1.96 \times 0.1) = (14.75, 15.15)$

(2) $s^2 = \frac{1}{5}\sum_{i=1}^{6} (x_i - \bar{x})^2 = 0.051 \therefore s = 0.226$

取 $T = \frac{\overline{X} - \mu}{\sqrt{S^2/n}} = \frac{14.95 - \mu}{0.226/\sqrt{6}} \sim t(5)$

查表得：$t_{0.025}(5) = 2.5706$

μ的置信区间为$\left(14.95 \pm \frac{2.5706 \times 0.226}{\sqrt{6}}\right) = (14.95 \pm 0.237) = (14.713, 15.187)$

(3) 选取枢轴量 $K = \frac{5S^2}{\sigma^2} \sim \chi^2(5)$

$P\{\chi^2_{0.975}(5) < \frac{5 \times 0.051}{\sigma^2} < \chi^2_{0.025}(5)\} = 0.95$

$0.831 < \frac{5 \times 0.051}{\sigma^2} < 12.833$

$0.0199 < \sigma^2 < 0.3069$

❁ 8.4.2 双总体的集中区间估计

$X_1, \cdots X_m$是来自$N(\mu_1, {\sigma_1}^2)$的样本；$Y_1, \cdots Y_n$是来自$N(\mu_2, {\sigma_2}^2)$的样本。

$\overline{X}, {S_1}^2, \overline{Y}, {S_2}^2$分别是两个样本的均值和方差，置信度为$1 - \alpha$

① ${\sigma_1}^2, {\sigma_2}^2$已知，$\mu_1 - \mu_2$的置信区间：

$\overline{X}, \overline{Y}$相互独立，所以：$\overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \frac{{\sigma_1}^2}{m} + \frac{{\sigma_2}^2}{n})$

$\therefore \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{{\sigma_1}^2}{m} + \frac{{\sigma_2}^2}{n}}} \sim N(0,1)$

置信区间为：$\left((\overline{X} - \overline{Y}) \pm z_{\alpha/2}\sqrt{\frac{{\sigma_1}^2}{m} + \frac{{\sigma_2}^2}{n}}\right)$

② ${\sigma_1}^2 = {\sigma_2}^2 = \sigma^2$，但$\sigma^2$未知，$\mu_1 - \mu_2$的置信区间：

$\frac{(m-1){S_1}^2}{\sigma^2} \sim \chi^2(m-1), \frac{(n-1){S_2}^2}{\sigma^2} \sim \chi^2(n-1)$

卡方分布的可加性：$\frac{(m-1){S_1}^2 + (n-1){S_2}^2}{\sigma^2} \sim \chi^2(m+n-2)$

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{1}{m}+\frac{1}{n}}\sigma}\sim N(0,1)$$

根据 t 分布定义，构造：

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{1}{m}+\frac{1}{n}}\sqrt{\frac{(m-1){S_1}^2+(n-1){S_2}^2}{m+n-2}}}\sim t(m+n-2)$$

③ ${\sigma_1}^2,{\sigma_2}^2$未知，$m,n>50$，可视为大样本：

$$\frac{{\sigma_1}^2}{m}+\frac{{\sigma_2}^2}{n}\approx\frac{{S_1}^2}{m}+\frac{{S_2}^2}{n}$$

④ μ_1,μ_2未知，${\sigma_1}^2/{\sigma_2}^2$的置信区间：

构造枢轴量：$\mathrm{F}=\frac{{S_1}^2/{S_2}^2}{{\sigma_1}^2/{\sigma_2}^2}\sim F(m-1,n-1)$

例：某工厂两条自动化流水线罐装番茄酱。分别从两条流水线抽取容量为 13 和 17 的两个相互独立的样本$X_1,\cdots X_{13}$；$Y_1,\cdots Y_{17}$

已知：$\overline{x}=10.6,\overline{y}=9.5,{s_1}^2=2.4,{s_2}^2=4.7$

假设都服从正态分布，均值分别为μ_1,μ_2。

(1) 若${\sigma_1}^2={\sigma_2}^2=\sigma^2$，求$\mu_1-\mu_2$的置信度为 0.95 的置信区间；

(2) 若不知道方差是否相同，求方差比的置信度为 0.95 的置信区间。

(1)枢轴量：$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{1}{m}+\frac{1}{n}}\sqrt{\frac{(m-1){S_1}^2+(n-1){S_2}^2}{m+n-2}}}\sim t(m+n-2)$

$$\frac{(10.6-9.5)-(\mu_1-\mu_2)}{\sqrt{\frac{1}{13}+\frac{1}{17}}\sqrt{\frac{12\times2.4+16\times4.7}{28}}}\sim t(28)$$

查表得：$t_{0.025}(28)=2.0484$

$\mu_1-\mu_2$的置信区间$(1.1\pm1.4545)=(-0.3545,2.5545)$

(2)枢轴量：$\mathrm{F}=\frac{{S_1}^2/{S_2}^2}{{\sigma_1}^2/{\sigma_2}^2}\sim F(m-1,n-1)$

$$\frac{2.4/4.7}{{\sigma_1}^2/{\sigma_2}^2}\sim F(12,16)$$

查表得：$F_{0.025}(12,16)=2.89$

$$F_{0.975}(12,16)=\frac{1}{F_{0.025}(16,12)}=\frac{1}{3.16}$$

${\sigma_1}^2/{\sigma_2}^2$置信区间为：

$$\left(\frac{2.4}{4.7\times2.89},\frac{2.4\times3.16}{4.7}\right)=(0.1767,1.614)$$

🕮 8.5 火车头数量估计问题

铁路上以 1 到 N 编号火车头。有一天看到一个编号 60 的火车头，估计铁路上有多少火车头(估计 N)？

① 极大似然估计：

似然函数$\mathrm{L}=1/\mathrm{N}$需要最大化，N 需要最小化，因为 N 不能小于看到的编号，即 60，所以 N 的极大似然估计为 60。

② 最小偏方差估计

设观察到的编号为i，估计量为$\mathrm{K}i$；偏方差为$(\mathrm{K}i-N)^2$；假设重复 N 次，即观测值从 1~N：$\sum_{i=1}^{N}(\mathrm{K}i-N)^2$

对 K 求一阶导数：

$$\sum_{i=1}^{N}[2i(\mathrm{K}i-N)]=0\rightarrow \mathrm{K}=\frac{3N}{2N+1}$$

估计值为：$\mathrm{K}i=\frac{3N}{2N+1}\times 60\approx 90$

③ 无偏估计

偏差$\mathrm{K}i-N$，假设重复 N 次，即观测值从 1~N：$\sum_{i=1}^{N}(\mathrm{K}i-N)$

令$\sum_{i=1}^{N}(\mathrm{K}i-N)=0\rightarrow K=\frac{2N}{N+1}$

估计值为：$\mathrm{K}i=\frac{2N}{N+1}\times 60\approx 120$

综上所述，没有一种估计方法具备所有优点。

实际应用根据需求选择合适的估计方法。

```python
def train():
    h1, h2, h3 = 0, 0, 0
    mse1, mse2, mse3 = 0, 0, 0
    me1, me2, me3 = 0, 0, 0

    CNT = 10000  # 总共测试次数
    for j in range(CNT):
        N = randint(1, 100)  # 火车头总数 1~100 随机
        i = randint(1, N)  # 随机看到一个火车头的编号
        # 3 种估计值
        e1 = i
        e2 = round(1.5*i)
        e3 = 2*i
        # 如果猜中,命中数+1
        if e1 == N:
            h1 += 1
        if e2 == N:
            h2 += 1
        if e3 == N:
            h3 += 1
        # 计算总偏方差和总偏差
        mse1 += (e1-N)**2
        mse2 += (e2-N)**2
        mse3 += (e3-N)**2
        me1 += e1-N
        me2 += e2-N
        me3 += e3-N
    print(f'命中率：{h1/CNT}, {h2/CNT}, {h3/CNT}')
    print(f'均偏方差：{mse1/CNT}, {mse2/CNT}, {mse3/CNT}')
    print(f'均偏差：{me1/CNT}, {me2/CNT}, {me3/CNT}')
```

结果：

```
命中率：0.0523, 0.0292, 0.0225
均偏方差：1145.7825, 855.2121, 1134.7286
均偏差：-25.4311, -12.5505, 0.2942
```

极大似然估计猜中次数最多；均偏方差第二种最小；均偏差第三种最小。

第 9 章　假设检验

🕮　9.1　假设检验的概念

❀　9.1.1　假设检验引言

① 总体分布已知，但有未知参数，可以对未知参数作出假设，利用样本的数据检验假设。如全体学生身高服从正态分布，但不知道平均身高。此时假设平均身高为 1.68m，抽取部分学生测量身高，检验 1.68m 的假设是否正确(可靠性)。

② 总体分布未知。如某个车站中午等车的人数，其分布未知。此时假设其服从泊松分布，抽取样本检验是否符合泊松分布。

❀　9.1.2　假设检验的基本概念

若对参数一无所知，可用参数估计的方法。
若对参数有所了解，但怀疑猜测，需要验证时，使用假设检验的方法。

假设检验是指施加于一个或多个总体的概率分布或参数的假设，假设不一定正确。为了验证，从总体中抽取样本，根据样本取值，按一定原则进行检验，最后接受或拒绝假设。

假设检验的内容：
1) 参数检验：总体均值、均值差的检验；总体方差、方差比的检验。
2) 非参数检验：分布拟合检验；符号检验；秩和检验。

假设检验的理论依据：小概率事件原理。

例 1：某产品次品率 p 不超过 4%才能出厂。现从 1 万件产品中抽取 12 件发现 3 件是次品，该批产品能否出厂?若抽查发现 1 件次品，能否出厂?

1) 假设$p \le 0.04$
$\mathrm{p1} = C_{12}^{3} p^{3}(1-p)^{9} \le 0.0097 < 0.01$
这是小概率事件，一般认为不会发生，但是竟然发生了，所以认为原假设不成立，也就是次品率大于 0.04，故不能出厂。
2) $\mathrm{p2} = C_{12}^{1} p^{1}(1-p)^{11} \approx 0.306 > 0.3$
这不是小概率事件，没理由拒绝原假设，认为合格，可以出厂。

注：此处使用的是概率意义下的反证法，因而不成立即拒绝原假设是有说服力的，接受原假设是没说服力的。很多时候把希望否定的假设作为原假设。

例 2：某厂生产螺钉，标准强度为 68/mm^2，实际强度$\mathrm{X}\sim\mathrm{N}(\mu, 3.6^2)$。若$\mathrm{E}(\mathrm{X}) = \mu = 68$，则认为符合要求，否则不符合。提出假设：
H_0: $\mu = 68$——称为原假设或零假设
原假设的对立面：H_1: $\mu \neq 68$——称为备择假设

而假设的检验任务是必须在原假设和备择假设之间作一选择。

现从其中抽取容量为 36 的样本，均值$\overline{x}=68.5$，问原假设是否正确?

若原假设正确：$\overline{\mathrm{X}}\sim N(68,\frac{3.6^2}{36})$

因而$\mathrm{E}(\overline{\mathrm{X}})=68$，即$\overline{\mathrm{X}}$偏离 68 不应该太远，故$|\frac{\overline{\mathrm{X}}-68}{3.6/6}|$取较大值是小概率事件。

因此可以确定一个常数 c 使得：

$$\mathrm{P}\{|\frac{\overline{\mathrm{X}}-68}{3.6/6}|>c\}=\alpha$$

取$\alpha=0.05$，$c=z_{0.025}=1.96$

$$|\frac{\overline{\mathrm{X}}-68}{3.6/6}|>1.96\rightarrow\overline{\mathrm{X}}>69.18\ \ or\ \ \overline{\mathrm{X}}<66.82$$

区间$(-\infty,66.824)\cup(69.18,+\infty)$为检验的拒绝域。
$(66.82,69.18)$为检验的接受域，没理由拒绝。
$\overline{x}=68.5$落在接受域，则接受原假设H_0: $\mu=68$

由上面**例 2** 可见，在给定α前提下，接受或拒绝原假设完全取决于样本值，因此所作检验可能导致两类错误的产生：

第一类错误：弃真错误(记为α)；
第二类错误：取伪错误(记为β)。

任何检验方法都不能完全排除犯错误的可能性。理想的检验方法应使犯两类错误的概率都很小。但在样本容量给定时，降低一个，往往使另一个增大。

假设检验的指导思想是：控制第一类错误的概率不超过α，若有必要，通过增大样本容量来减少β。

备择假设可以是单侧(单边)或双侧(双边)。如**例 2** 的备择假设，μ可能大于 68，也可能小于 68，这是双边备择假设，其对应假设检验称为双边假设检验。

如：H_0: $\mu\leq 68, H_1$: $\mu>68$形式的称为右边检验(此点右边区域全部拒绝)。
类似地如：H_0: $\mu\geq 68, H_1$: $\mu<68$形式的称为左边检验。
两者统称为单边检验。

❀ 9.1.3 假设检验的步骤

1) 根据实际问题建立H_0与H_1
2) 在H_0为真时，选择**合适统计量** V，由H_1确定拒绝域 R
双边检验(两边区间全部拒绝)、左边检验(左边区间全部拒绝)、右边检验
3) 计算并作出相应的判断

// 感觉和区间估计很类似啊...

🕮 9.2 单个正态总体的参数检验

❀ 9.2.1 关于均值 μ 的检验

① σ^2已知 (U 检验/Z 检验)

给定显著性水平α与样本值$(X_1,\cdots,X_n)$

设X~N(μ,σ^2),σ^2已知，需检验：

$$H_0:\mu=\mu_0;\quad H_1:\mu\neq\mu_0$$

构造统计量：

$$U=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$$

$$P\{\text{当}H_0\text{为真但拒绝}H_0\}=P_{H_0}\left(\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right|\geq z_{\alpha/2}\right)=\alpha$$

所以拒绝域为：$|U|\geq z_{\alpha/2}$

② σ^2未知 (t 检验)

使用样本标准差 S 代替σ，统计量采用：

$$T=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$$

拒绝域为：$|T|\geq t_{\alpha/2}(n-1)$

例：某厂小型马达在正常负载下平均消耗电流不超过 0.8A。测试 16 台马达，平均消耗电流 0.92A，标准差 0.32A。马达消耗电流服从正态分布，取显著水平$\alpha=0.05$，此样本能否否定厂方断言。

(1) 待检假设设为：$H_0:\mu\leq 0.8;\ H_1:\mu>0.8$

选取统计量：

$$T=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$$

因为是单边检验：$P\{T\geq t_\alpha(n-1)\}=\alpha$

拒绝域为：$T\geq t_\alpha(n-1)=t_{0.05}(15)=1.753$

$$T=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}=\frac{0.92-0.8}{0.32/\sqrt{16}}=1.5$$，落在拒绝域之外

所以接收原假设H_0，不能否定厂方断言。

(2) 待检假设设为：$H_0:\mu>0.8;\ H_1:\mu\leq 0.8$

选择和(1)相同的统计量。

因为是单边检验：$P\{T\leq -t_\alpha(n-1)\}=\alpha$

拒绝域为$T\leq -t_\alpha(n-1)=-t_{0.05}(15)=-1.753$

$T=1.5$，落在拒绝域之外。

所以接收原假设H_0，即否定厂方断言。

示例两种解法使用相同样本和统计量，不同的原假设，检验的结果也不同。

第一种假设是不轻易否定厂方结论；第二种假设是不轻易相信厂方结论。

引起检验结果不同的根本原因是样本容量不够大。
若样本容量足够大，不论把哪个假设作为原假设，结果应该是一样的，否则假设检验就无意义了。

假设检验是控制犯第一类错误的概率，使得拒绝原假设H_0的决策变得比较慎重，原假设H_0得到特别的保护。因而通常把有把握的、经验的结论作为原假设，或尽量使后果严重的错误成为第一类错误。

⊛ 9.2.2 关于方差σ^2的检验 (χ^2检验)

① μ已知

设$X\sim N(\mu,\sigma^2)$，μ已知，需检验：

$$H_0:\sigma^2=\sigma_0{}^2;\ H_1:\sigma^2\neq\sigma_0{}^2$$

构造统计量：

$$\chi^2=\sum_{i=1}^{n}(\frac{X_i-\sigma}{\sigma_0})^2\sim\chi^2(n)$$

拒绝域：$\chi^2\leq\chi^2_{1-\alpha/2}(n),\chi^2\geq\chi^2_{\alpha/2}(n)$

② μ未知

构造统计量：

$$\chi^2=\frac{(n-1)S}{\sigma_0{}^2}\sim\chi^2(n-1)$$

拒绝域：$\chi^2\leq\chi^2_{1-\alpha/2}(n-1),\chi^2\geq\chi^2_{\alpha/2}(n-1)$

🕮 9.3 双总体的假设检验

① 均值差$\mu_1-\mu_2$的检验

② 方差比σ_1^2/σ_2^2的检验 (F 检验)

例：检验机器 A 和 B 生产钢管内径的稳定程度，设它们生产钢管内径分别为 X 和 Y，服从正态分布$N(\mu_1,\sigma_1{}^2)$和$N(\mu_2,\sigma_2{}^2)$
现从 A 抽取 18 根，B 抽取 13 根，测得$s_1^2=0.34,s_2^2=0.29$

设：$H_0:\sigma_1{}^2=\sigma_2{}^2;\ H_1:\sigma_1{}^2\neq\sigma_2{}^2$

$$\frac{S_1^2}{S_2^2}\sim F(17,12)$$

查表得：$F_{0.05}(17,12)=2.59$

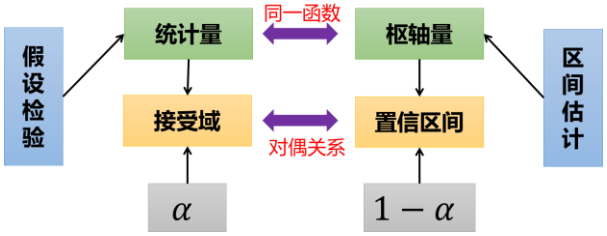
$$F_{0.95}(17,12)=\frac{1}{F_{0.05}(12,17)}=\frac{1}{2.38}=0.42$$

所以拒绝域为：$\frac{S_1^2}{S_2^2}>2.59$ or $\frac{S_1^2}{S_2^2}<0.42$

由抽样值计算得：$\frac{s_1^2}{s_2^2}=\frac{0.34}{0.29}=1.17$

落在拒绝域之外，所以接受原假设，认为 A、B 内径稳定程度相同。

🕮 9.4 区间估计和假设检验的联系



🕮 9.5 χ^2拟合检验

对同一样本，在不同总体分布的假设下，其结论可能不同，甚至矛盾。所以经常需要检验一个样本来自什么样的分布总体，这属于非参数假设检验。

χ^2拟合检验

英国著名统计学家 K. Pearson 于 1900 年提出。

① 分类数据的χ^2检验

根据某个指标，总体被分为 r 类，每类所占比例为：

$$p_i(i=1,2,\dots,r),\sum_{i=1}^{r}p_i=1$$

原假设：H_0:第 i 类占比例为p_i

每类有n_i个观察样本，样本总数为n

检验统计量为：

$$\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}$$

Pearson 给出的定理：在H_0为真时

$$\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}\sim\chi^2(r-1)$$

② 带参数的分类数据的检验

当$p_i(i=1,2,\dots,m)$未知，用样本求出极大似然估计$\hat{p}_i$，对应检验统计量为：

$$\chi^2=\sum_{i=1}^{r}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}\sim\chi^2(r-m-1)$$

③ 分布检验

$X_1,\cdots,X_n$是来自总体$F(x)$的样本，需要检验的原假设为：

$$H_0:F(x)=F_0(x)$$

其中$F_0(x)$称为理论分布。

1) 总体 X 为离散型随机变量

将若干个a_i值并为一类，使a_i被分为有限个类$B_1,\cdots,B_r$，并使样本的观察值x_i落在每个B_i内的个数n_i不小于 5。

记$\mathrm{P}\{\mathrm{X}\in B_i\}=p_i, i=1,\cdots,r$，则在$\mathrm{H}_0$成立时，$B_i$所占的比例是$p_i$
从而，检验问题与分类数据的检验问题是一样的。

1) 总体 X 为连续型随机变量
选适当的点$a_1<a_2<\cdots<a_{r-1}$把实数轴分为 r 个区间。
这 r 个区间相当于 r 个类。
$p_i=P\{a_{i-1}<X\le a_i\}=\mathrm{F}_0(a_i)-\mathrm{F}_0(a_{i-1})$

例：84 个成年男子身高数据

高度(cm)	154	155	156	157	159	160	161	162	163	164	165	166
人数	1	2	1	1	2	1	5	1	2	5	2	1
高度(cm)	167	168	169	170	171	172	173	174	175	176	177	178
人数	4	5	5	4	4	7	3	7	4	2	2	3
高度(cm)	179	180	181	182	185							
人数	1	1	3	3	2							

(1) 试检验成年男子身高服从正态分布
(2) 若公共汽车门高度按男子碰头机会不超过 1%设计，由以上数据设计车门高度至少为多少?

(1) $\hat{\mu}=\overline{X}=170, \hat{\sigma}^2=S^2=7.2^2$

$$\mathrm{F}_0(x)=\frac{1}{7.2\sqrt{2\pi}}\exp\{-\frac{1}{2}(\frac{x-170}{7.2})^2\}$$

$\mathrm{H}_0: \mathrm{F}(x)=\mathrm{F}_0(x)$
统计量$\chi^2=\sum_{i=1}^{r}(v_i^2/np_i)-n$
其中$r=8, p_i=\Phi(a_i)-\Phi(a_{i-1})$

	频数v_i	v_i^2	p_i	np_i	v_i^2/np_i
(, 158]	5	25	0.0475	3.99	6.2675
(158, 162]	8	64	0.0860	7.224	8.8594
(162,166]	10	100	0.1542	12.9528	7.7203
(166, 170]	15	225	0.2123	17.8332	12.6169
(170, 174]	18	324	0.2123	17.8332	18.1684
(174, 178]	15	225	0.1542	12.9528	17.3708
(178, 182]	8	64	0.0860	7.224	8.8594
(182,)	5	25	0.0475	3.99	6.2657

由数据计算统计量为：

$$\chi^2=\sum_{i=1}^{8}\left(\frac{v_i^2}{np_i}\right)-n=86.1266-84$$

$$=2.1266\le\chi^2_{0.05}(8-2-1)=\chi^2_{0.05}(5)=11.07$$

// 期望和方差是估计的，m=2，自由度为 r - m - 1 = 8 - 2 - 1 = 5
故接收原假设。
(2) 成年男子身高$\mathrm{X}\sim(170, 7.2^2)$
$\mathrm{P}\{\mathrm{X}\ge \mathrm{h}\}=1-\mathrm{P}\{\mathrm{X}<\mathrm{h}\}\le 0.01$

$$P\{X < h\} = \Phi\left(\frac{h-170}{7.2}\right) \geq 0.99 = \Phi(2.33)$$

$$h \geq 2.33 \times 7.2 + 170 = 186.776$$

第 10 章 方差分析