课程介绍

- 了解概率统计含义
- 了解事件的计算
- 概率的定义和性质
- 掌握期望、方差意义
- 掌握正态分布的相关概念

1. 伯努利分布（0--1分布）

又叫做0-1分布，指一次随机试验，结果只有两种。也就是一个随机变量的取值只有0和1。 记为: 0-1分布 或B(1,p)，其中 p 表示一次伯努利实验中结果为正或为1的概率。

硬币正面朝上的概率

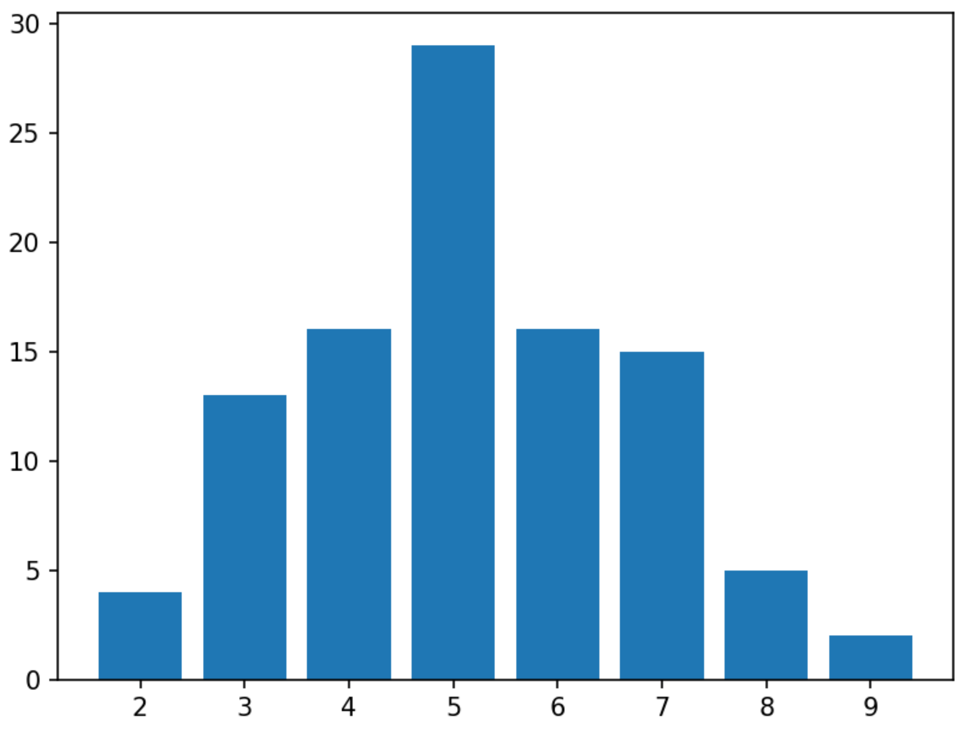
抛硬币只有2种结果，每种结果的概率性相等，而且互不干扰。正面朝上定义为 1、背面朝上定义为 0，来计算结果的概率。

假设我们抛同一枚均匀硬币。把抛10次硬币这样一件事记为第1组实验。

同样的实验我们做100次，得到一下数据

```
7,7,9,4,4,4,5,2,5,7,5,....
{7: 15, 9: 2, 4: 16, 5: 29, 2: 4, 8: 5, 6: 16, 3: 13}
```

正面次数2出现4次, 正面次数3出现了13次。我们将结果绘制成柱状图如下所示



代码实现:

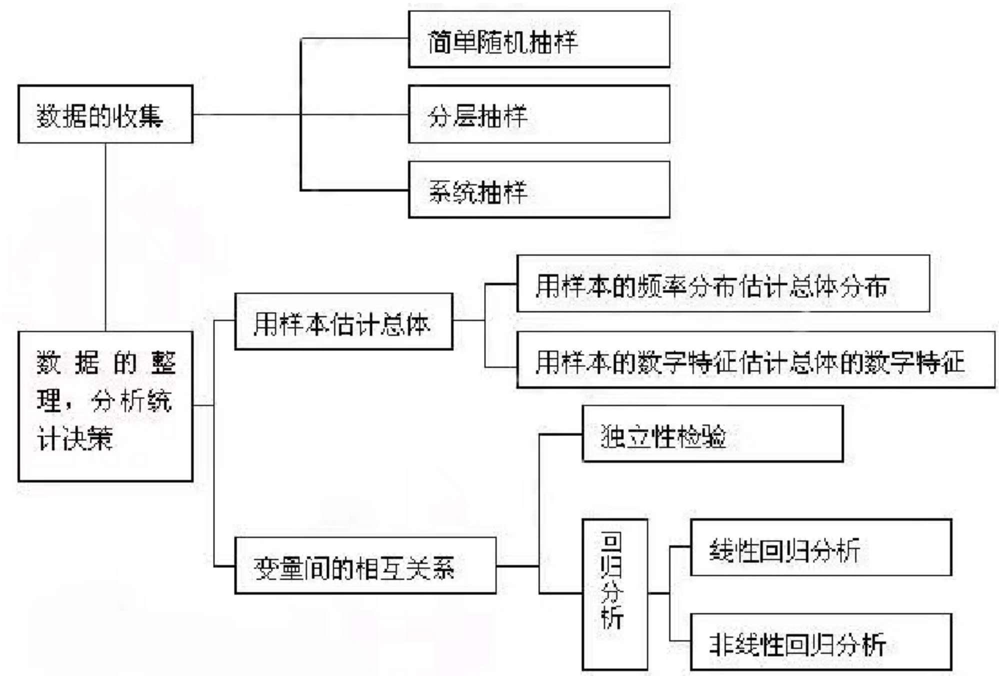
```python
import random
def coin_head_sum(n):
    total = 0
    for _ in range(0, n):  # 执行n次
        if random.random() >= 0.5:
            total += 1
    return total

n = 100
for i in range(100):
    print(coin_head_sum(n) / n)
```

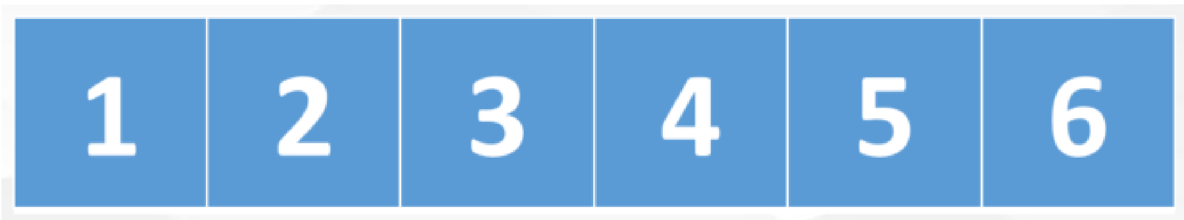
```
    coin_head_sum(n) / n 的目的是计算每次抛硬币实验中正面出现的概率。这是因为
coin_head_sum(n) 返回的是在一次实验中正面出现的次数，而将其除以总的抛硬币次数 n 就得到了正面
出现的概率。
```

2. 概率统计

概率：是反映随机事件出现的可能性大小。主要研究随机变量、随机事件、随机过程。
统计：收集、处理、分析、解释数据并从数据中得出结论的科学。



3. 古典型概率



随机变量　i = 1, 2, 3, 4, 5, 6（离散型随机变量）

古典型概率: 实验的所有结果只有有限个，且每个结果发生的可能性相同，其概率计算公式：

概率P(E)（古典概型）= 事件包含的结果数 / 事件可能的结果总数

P(i = 1) = 1 / 6
P(i = 2) = 1 / 6
P(i >= 3) = ?
Ω（样本空间） = 1 + 2 + 3 + 4 + 5 + 6（所有可能结果组成的集合）

总结:

如果一个事件的发生满足：

（1）事件的结果有限

（2）事件的每个结果出现的概率一样

这样的试验便是古典试验。

```
P = 概率
E = 事件
Ω（Omega） =  样本空间、必然事件、全集
Φ（Phi） =  空集、不可能事件
Σ（sigma） =  总和
```

频率定义

在做大量重复试验时，随着试验次数的增加，一个事件出现的频率，总在一个固定数的附近摆动，具有一定的稳定性。可以把这个固定数定义为该事件的概率，这就是概率的频率定义。

4. 概率具有以下7个不同的性质

- 性质1：P(Φ)=0；
- 性质2：（有限可加性）当n个事件A1,...,An两两互不相容时：
  P(A1∪ ... ∪An) = P(A1) + ... + P(An)；
- 性质3：对于任意一个事件A：P(A) = 1 - P(!A)；
- 性质4：当事件A,B满足A包含于B时：P(B - A) = P(B) - P(A)，P(A) ≤ P(B)；
- 性质5：对于任意一个事件A，P(A) ≤ 1；
- 性质6：对任意两个事件A和B，P(B - A) = P(B) - P(A∩B)；
- 性质7：（加法公式）对任意两个事件A和B，P(A∪B) = P(A) + P(B) - P(A∩B)。

5. 事件

在一个随机试验中，每一可能出现的结果是一个基本事件，全体基本事件的集合称为样本空间。

连续扔两次骰子，用a，b表示第一次和第二次出现的点数，a和b可以取值1、2、3、4、5、6，每一点（a，b）表示一个基本事件，因而基本空间包含？个元素。

点数之和为4，它由哪几个基本事件组成，P = ?

```
点数之和为4，由基本事件 (1,3),(3,1),(2,2) 组成
P(A)=P((1,3)∪(3,1)∪(2,2))=P(1,3)+P(3,1)+P(2,2)
```

点数之和为1，P(不可能事件) = 0

点数之和小于40，P = 1

样本空间总数: 36

6. 随机事件和概率

排列组合公式:

从m个人中挑选n个人进行排列的可能数

$$A_m^n = \frac{m!}{(m-n)!}$$

从m个人中挑选n个人进行组合的可能数

$$C_m^n = \frac{m!}{n!(m-n)!}$$

7. 事件计算

设A、B为随机事件，若同时发生的概率等于各自发生的概率的乘积，则A、B相互独立。

P(AB) = P(A)P(B)

例如: 抛一枚硬币正面向上出现的概率P(A)， 陨石撞击地球发生的概率P(B)互不影响

同时发生的概率等于各自发生概率的乘积

P(同时发生的概率) = P(A)P(B)

例如: 小强 和 小红 各自打靶射击，小强射中的概率是1/3, 小红射中的概率是 1/4。

同时击中 的概率:
P(同时击中 的概率) = 1/3 * 1/4 = 1 / 12

7.1 加法原理

某事件有N类方式完成，第一类方式可由m种方法完成，第二类方式可由n种方法来完成，则这件事可由 m + n 种方法来完成。

条件：每一种方法都是独立、完整、互斥的。

加法公式：P(A+B) = P(A) + P(B)

7.2 乘法原理

某件事由两个步骤来完成，第一个步骤可由m种方法来完成，第二个步骤可由n种方法来完成，则这件事可由 m * n 种方法来完成。

乘法公式： P(AB) = P(A|B)P(B) 或 P(AB) = P(B|A)P(A)

P(A|B)——在B条件下A 的概率. 即事件A 在另外一个事件B已经发生条件下的发生概率。

P(AB)——事件A、B同时发生的概率,即联合概率.

例: 两个仓库分别运送水果到销售点，到达目的地后从这两个仓库物品中随机抽查100件，发现合格品为80件和90件

	仓库1(B1)	仓库2(B2)	
合格品(A)	80	90	170
次品	20	10	30
	100	100	200

现在从这200件产品中随机挑选一件，发现它来自仓库1，请问该产品是正品的概率是多少?

求:
(1) P(A|B1) 从上述可得 已知来自仓库1， 该产品为正品的概率为80/100

条件概率：事件A在事件B1已经发生条件下的发生概率

P(A|B1) = P(AB1)/P(B1)

(2) 200件商品中，假如既来自仓库1，同时又是正品的概率
联合概率：既满足 A 条件，又满足 B 条件的概率
记为 P(A∩B1)、P (A,B1) 、P(AB1)

P(AB1) = P(A|B1)*P(B1) = 80/200

(3) 200件商品中，正品的概率

P(A)=?

8. 抽样

放回抽样：逐个抽取样本时，每次被抽到的样本放回总体中后，再进行下次抽取的抽样方法

不放回抽样：逐个抽取个体时，每次被抽到的个体不放回总体中参加下一次抽取的方法

例: 有5个球， 3个红球，2个黄球，从中取2个球(不放回)，求这2个球都是红球的概率

总事件次数:

- 总事件的次数 C(5, 2) = 10

取红球事件的次数:

- 取红球事件的次数 C(3, 2) = 3

取红球事件的次数/总事件次数 3/10= 0.3

抽样情况:

```
['红1,红2', '红1,红3', '红1,黄1', '红1,黄2', '红2,红3', '红2,黄1', '红2,黄2', '红3,黄1', '红3,黄2', '黄1,黄2']

['红1,红2', '红1,红3', '红2,红3']
```

结果:

```
0.3
```

例: 有5个球， 3个红球，2个黄球，从中取2个球(放回抽样)，求这2个球都是红球的概率

抽样情况:

```
15
['红1,红1', '红1,红2', '红1,红3', '红1,黄1', '红1,黄2', '红2,红2', '红2,红3', '红2,黄1', '红2,黄2', '红3,红3', '红3,黄1', '红3,黄2', '黄1,黄1', '黄1,黄2', '黄2,黄2']
6
['红1,红1', '红1,红2', '红1,红3', '红2,红2', '红2,红3', '红3,红3']
```

结果:

```
0.4
```

9. 条件概率、联合概率、边缘概率

条件概率：事件A在事件B已经发生条件下的发生概率
记为 P(A|B)

联合概率：既满足 A 条件，又满足 B 条件的概率
记为 P(A∩B)、P (A,B) 、P(AB)

边缘概率：在多元概率分布中只考虑单个概率
记为 P(X=a)或P(Y=b)

条件概率:

在P(Y=b) 已经发生的前提下，P(X=a)发生的概率
P(X=a,|Y=b) = P(X=a,Y=b) / P(Y=b)

例: 在16张扑克牌中随机抽取一张，设X为颜色(红色或者黑色), Y为牌型(数字或者字母)

已知抽到的是红牌，请问抽到的是数字牌的概率是?

J	Q	K	J
Q	K	1	2
K	1	2	3
3	4	5	6

	Y=字母	Y=数字
X=红	6/16	3/16
X=黑	1/16	6/16

边缘概率：P(X=红) = P(Y=字母，X=红) + P(Y=数，X=红) = 9/16

条件概率：P(Y=数字|X=红) = P(Y=数，X=红) / P(X=红) = 3/9 = 1/3

联合概率: P(Y=数，X=红) = P(Y=数字|X=红) * P(X=红) = (1/3) * (9/16) = 3/16

10. 全概率定理

将对复杂事件A的概率求解问题转化为在不同情况下发生的简单事件的概率的求和问题。

$$P(B)=\sum_{j=1}^{n}P(A_j)P(B|A_j)$$

全概率公式： $P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \ldots + P(A_n)P(B|A_n)$

贝叶斯公式：

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)} = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \ldots + P(A_n)P(B|A_n)}$$

场景描述：

你正在开发一个图像分类系统，该系统需要识别不同种类的动物。你的训练数据包括各种动物的图像，但由于图像质量和环境变化，每个类别的图像可能有不同的清晰度和背景噪声。

数据集：

假设你的数据集包含三个类别的动物：狗（Dog）、猫（Cat）、鸟（Bird）。

图像清晰度的分布：

- 狗的图像：
  - 清晰图像：P(Clear /Dog)=0.9（90% 的狗图像是清晰的）
  - 模糊图像：*P*(Blurry|Dog)=0.1（10% 的狗图像是模糊的）
- 猫的图像：
  - 清晰图像：P(Clear|Cat)=0.8（80% 的猫图像是清晰的）
  - 模糊图像：*P*(*Blurry*|*Cat*)=0.2（20% 的猫图像是模糊的）
- 鸟的图像：
  - 清晰图像：P(Clear|Bird)=0.7（70% 的鸟图像是清晰的）

    - 模糊图像：P(Blurry | Bird)=0.3（30% 的鸟图像是模糊的）
- **先验概率：**
    - 狗的先验概率：P(Dog)=0.4（40% 的图像是狗的图像）
    - 猫的先验概率：$P(Cat)$=0.3（30% 的图像是猫的图像）
    - 鸟的先验概率：P(Bird)=0.3（30% 的图像是鸟的图像）

计算后验概率：

假设有一张新的测试图像，该图像模糊。我们想要计算在这个条件下，图像属于每个类别的后验概率。

$P(Dog|Blurry) = \frac{P(Blurry|Dog)P(Dog)}{P(Blurry)}$

$P(Blurry) = P(Blurry|Dog)P(Dog) + P(Blurry|Cat)P(Cat) + P(Blurry|Bird)P(Bird)$

其中，$P(Blurry|Dog)$ 是在给定图像是狗的条件下，该图像模糊的概率。通过类似的方式，我们计算了 $P(Blurry|Cat)$ 和 $P(Blurry|Bird)$。

然后，我们可以计算 $P(Dog|Blurry)$、$P(Cat|Blurry)$ 和 $P(Bird|Blurry)$。选择具有最高后验概率的类别作为预测结果。

这个案例演示了如何利用全概率定理结合图像清晰度的信息和先验概率来计算后验概率，以更全面地理解模型在模糊条件下的分类情况。

11. 期望、方差、标准差(掌握)

11.1 期望

数学期望(mathematic expectation)（均值，亦简称期望）是试验中每次可能结果的概率乘以其结果的总和。
$E(X)$ 表示随机变量 X 的期望值

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

例如: : 下图是小毛射击的环数分布律

X	10	9	8	7
P	0.4	0.3	0.2	0.1

E(X) = 10 * 0.4 + 9 * 0.3 + 8 * 0.2 + 7 * 0.1 = 9

表示最有可能射中的环数是9

例如: 下图是分别是小毛和大毛射中的环数分布律

大毛:

X1	10	9	8
P	0.1	0.8	0.1

E(X1) = 10 * 0.1 + 9 * 0.8 + 8 * 0.1 = 9

小毛:

X2	10	9	8
P	0.3	0.4	0.3

E(X2) = 10 * 0.3 + 9 * 0.4 + 8 * 0.3 = 9

通过期望我们能看出大毛和小毛的水平吗？很显然不能, 引入方差。

11.2 方差

方差用来计算每一个变量（观察值）与总体均数之间的差异

$$D(x) = \sum_{k=1}^{\infty}[x_k - E(X)]^2 p_k \quad 或 \quad D(x) = \int_{-\infty}^{\infty}[x_k - E(X)]^2 f(x)dx$$

变换后可由下式来计算:

$$D(X) = E(X^2) - [E(X)]^2$$

大毛:

X1	10	9	8
P	0.1	0.8	0.1

E(X1) = 10 * 0.1 + 9 * 0.8 + 8 * 0.1 = 9

D(X1) = (10-9)^2*0.1 + (9-9)^2 * 0.8 + (8-9)^2 * 0.1 = 0.2

小毛:

X2	10	9	8
P	0.3	0.4	0.3

E(X2) = 10 * 0.3 + 9 * 0.4 + 8 * 0.3 = 9

D(X2) = (10-9)^2*0.3 + (9-9)^2 * 0.4 + (8-9)^2 * 0.3 = 0.6

通过上述我们发现:
D(X1) < D(X2) 我们就说大毛的设计水平更稳定

> D(X) 表示随机变量 X的方差，方差越小代表数据点相对于其均值的分散程度越小，数据更加集中，波动性更小。而方差越大则表示数据点相对于均值的分散程度越大，数据更加分散，波动性更大。

11.3 标准差

标准差（均方差）：标准差是方差的算术平方根。标准差能反映一个数据集的离散程度。

期望描述随机变量平均取值的大小

方差描述随机变量对于数学期望的偏离程度。

> 标准差的数值量纲（单位）与原始数据的数值量纲一致。让我们通过一个具体的例子来解释这个概念：
> 假设你在研究一组学生的考试成绩，这些成绩以百分制为单位。你计算了这组学生的考试成绩的标准差，并发现标准差为10分。
> 在这种情况下，标准差的数值单位与原始数据的数值单位（百分制成绩）是一致的，都是分（分数）。这是因为标准差的计算涉及将每个数据点与均值的差值进行平方，最后再开平方根，这个过程保留了原始数据的单位。
>
> 如果你在研究另一组数据，比如体重数据，单位是千克。在计算体重数据的标准差时，如果标准差的计算结果是2千克，那么标准差的单位就是千克。

12. 概率分布(了解)

概率分布：表述随机变量取值的概率规律。

- **随机变量的概率取值范围[0,1]：** 这意味着任何随机变量取某个特定值的概率都在0和1之间。概率不能为负，也不能超过1。
- **所有取值概率的和必须为1：** 对于一个随机变量的所有可能取值，它们的概率之和必须等于1。这表示随机变量取任何可能值的概率分配是完备的。
- **离散分布：** 当随机变量的取值是有限或可数的时候，我们称其为离散随机变量。概率分布描述了这种情况下每个可能取值的概率。

  > 意味着随机变量只能取一系列分离的、可以数清的数值，而不是连续的范围

- **连续概率分布：** 当随机变量的取值是连续的，即在一个区间内可以取任意值，我们称其为连续随机变量。概率分布在这种情况下通常通过概率密度函数来表示。

12.1 伯努利分布（两点分布或者0-1分布）

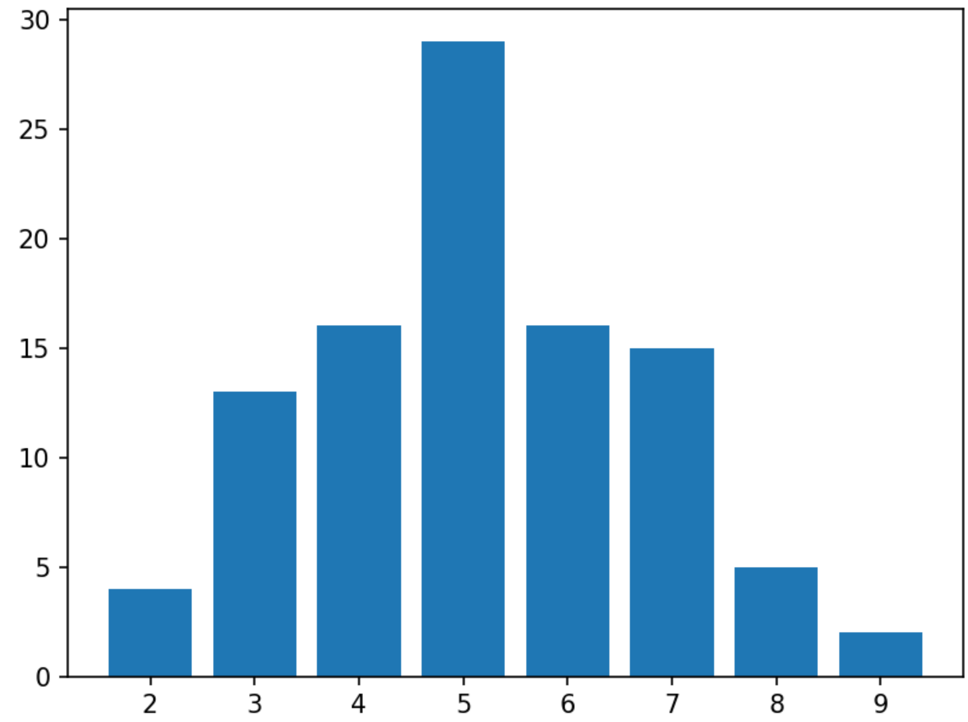
- 成功：1，记其成功概率为p
- 失败：0，记其失败概率为q = 1 − p

$$P(x) = p^x(1-p)^{1-x} = \begin{cases} p & \text{if } x=1 \\ q & \text{if } x=0 \end{cases}$$

n次实验，实验结果是互斥的，

案例：抛硬币实验

在抛硬币的过程中，我们进行了100组实验，称为: 100次10重伯努利实验。



12.2. 正态分布(高斯分布)密度函数(掌握曲线形状)

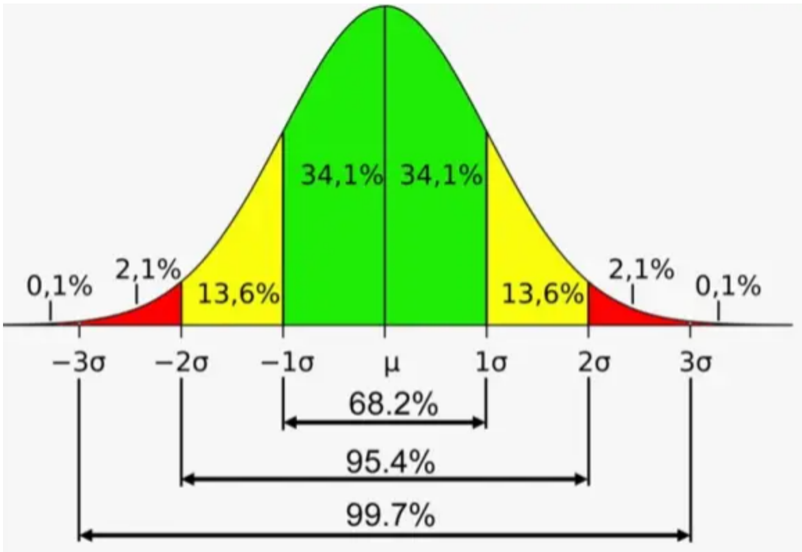
正态分布：正态曲线呈钟型，两头低，中间高，左右对称曲线呈钟形。

N(μ, σ^2)

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

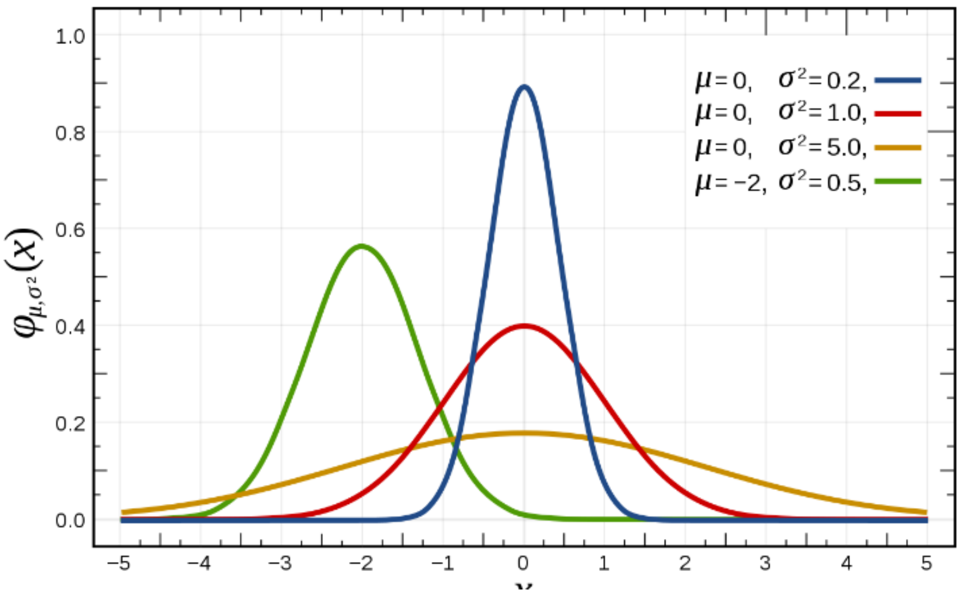
- 正态分布：若随机变量X服从一个数学期望为μ、方差为σ^2的正态分布，记为N(μ，σ^2)。μ决定了其位置（中心线），其标准差σ决定了分布的幅度（胖瘦）。
- 标准正态分布：当μ = 0，σ = 1时的正态分布是标准正态分布

http://vr.1zhidian.cn/bean/index.html



μ期望

σ标准差(标准差是方差的算术平方根)



代码:

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

# 定义随机变量
mean = 0  # 平均值
sd = 1  # 标准差
x = np.arange(-10, 10, 0.1)
# 正态分布概率密度函数(PDF)
y = stats.norm.pdf(x, mean, sd)
plt.plot(x, y)
plt.title(f"mu:{mean} -- sd:{sd}")
plt.grid()
plt.show()
```