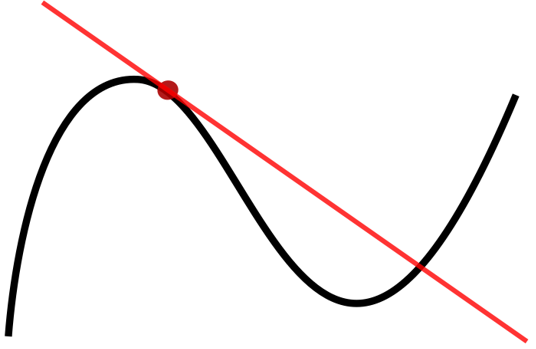
尚硅谷大模型技术之数学基础

（作者：尚硅谷研究院）

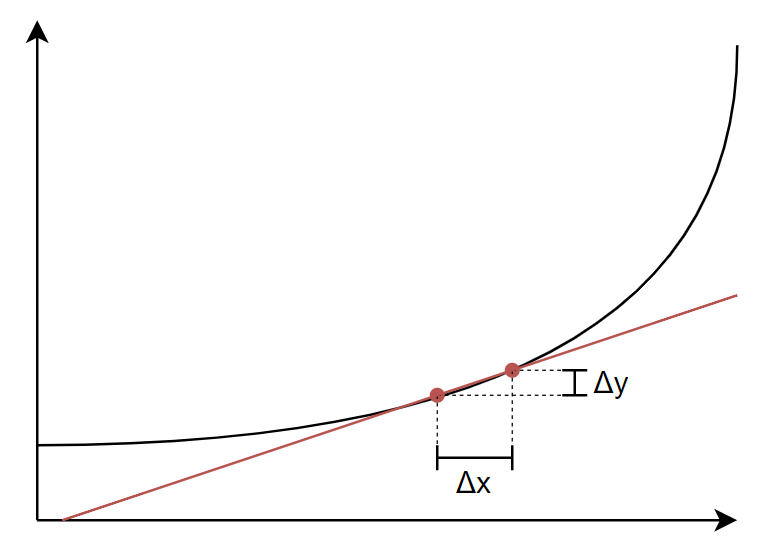
版本：V0.9.1

1. 高等数学
   1. 导数
      1. 导数的概念

导数（derivative）是微积分中的一个概念。函数在某一点的导数是指这个函数在这一点附近的变化率（即函数在这一点的切线斜率）。导数的本质是通过极限的概念对函数进行局部的线性逼近。



当函数的自变量在一点上产生一个增量时，函数输出值的增量与自变量增量的比值在趋于0时的极限如果存在，即为在处的导数，记作、或。



例如在运动学中，物体的位移对于时间的导数就是物体的瞬时速度：。

* + 1. 基本函数的导数

|  |  |  |
| --- | --- | --- |
| **说明** | **公式** | **例子** |
| **常数的导数** |  |  |
| **幂函数的导数** |  |  |
| **指数函数的导数** |  |  |
|  | — |
| **对数函数的导数** |  |  |
|  | — |
| **三角函数的导数** |  | — |
|  | — |
|  | — |
|  | — |

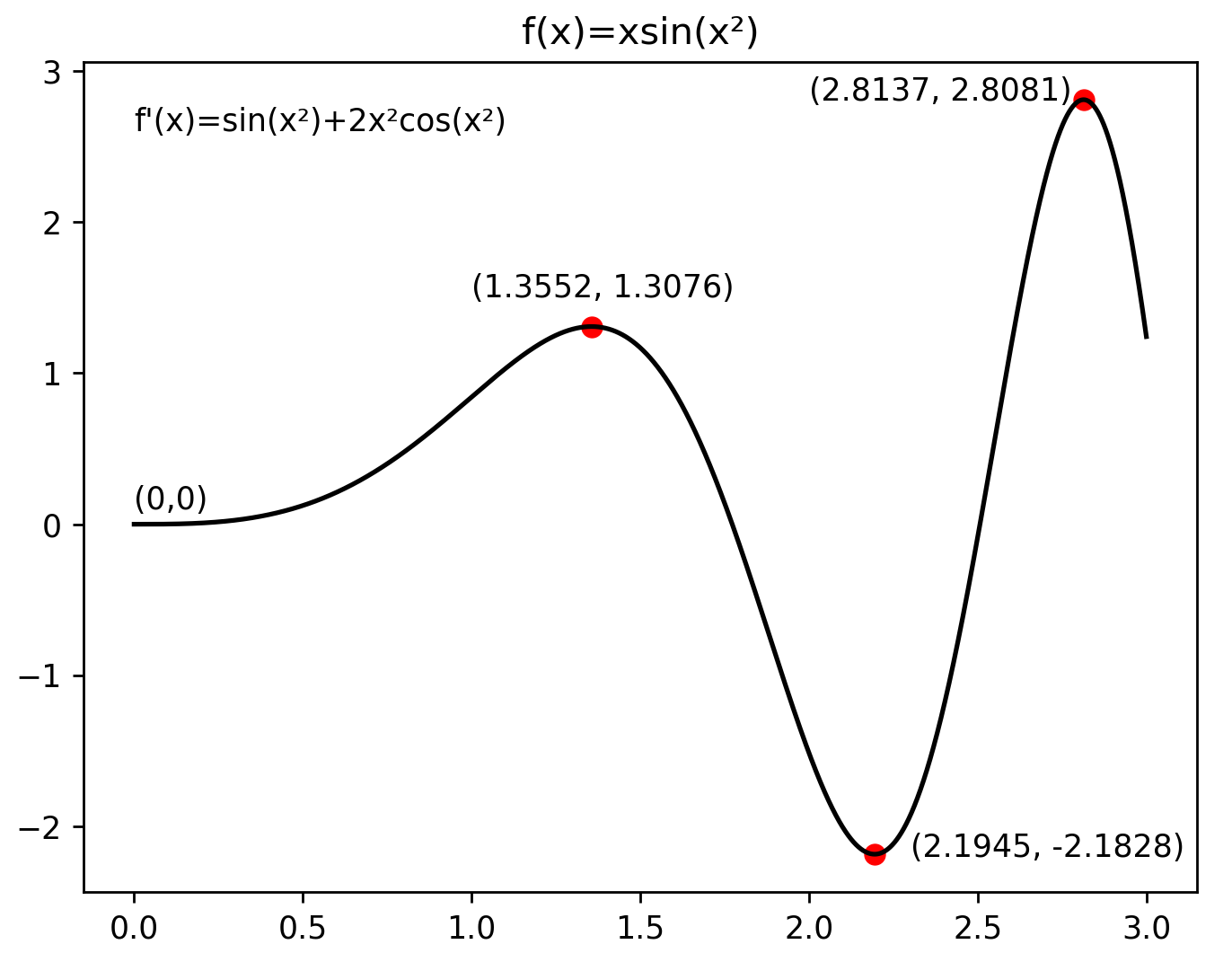
* + 1. 导数的求导法则

|  |  |
| --- | --- |
| **说明** | **公式** |
| **两函数之和求导** |  |
| **两函数之积求导** |  |
| **两函数之商求导** |  |
| **复合函数的导数** |  |

例如：求函数在处的导数。

* + 1. 利用导数求极值

导数等于零的点称为函数的驻点（或极值可疑点），在这类点上函数可能会取得极大值或极小值。进一步判断则需要知道导数在附近的符号。



* + 1. 二阶导数
       - 1. 二阶导数的概念

在微积分中，函数的二阶导数是函数导数的导数。粗略来说，某个量的二阶导数描述该量变化率变化的快慢。例如物体位置对时间的二阶导数是物体的瞬时加速度，即该物体速度随时间的变化率：。

函数的二阶导数通常记作、或。

* + - * 1. 二阶导数与函数凹凸的关系

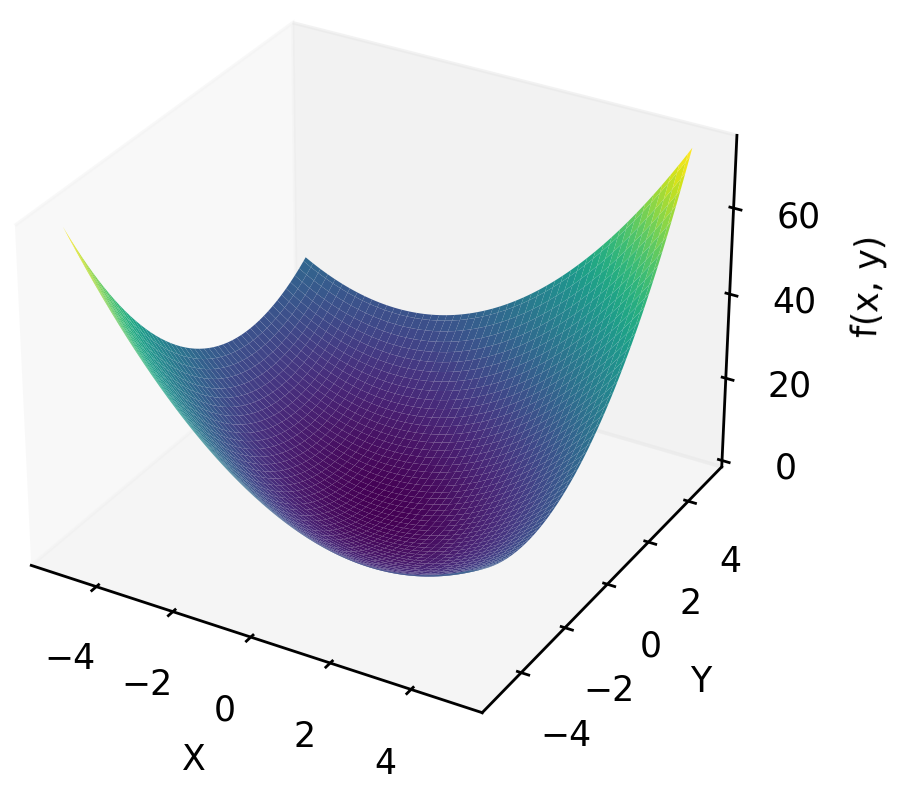
函数的二阶导数描述了函数图像的凹凸方向和程度。若二阶导数在某区间恒为正，则函数在该区间向上弯（也称下凸函数）。反之，若二阶导数在某区间恒为负，则函数在该区间向下弯（也称上凸函数）。

若函数的二阶导数在某点左右异号，则图像由向上弯转为向下弯，或反之。这种点称之为拐点。若二阶导数连续，则在该点处二阶导数为0。但反之二阶导数为0的点不一定是拐点。如，在处有，但在实数系上无拐点。

二阶导数与凹凸性的关系有助于判断函数的驻点是否为极大值点或极小值点：

* 若，，则在取得极大值。
* 若，，则在取得极小值。
* 若，，则该点可能是拐点，也可能是极大值点或极小值点。
  1. 偏导与梯度
     1. 偏导数

如果函数的自变量并非单个元素，而是多个元素，例如：



可将其中一个元素看作参数，此时可看作关于另一元素的函数。

在固定的情况下，可计算关于的导数：

这种导数称为偏导数，一般记作：

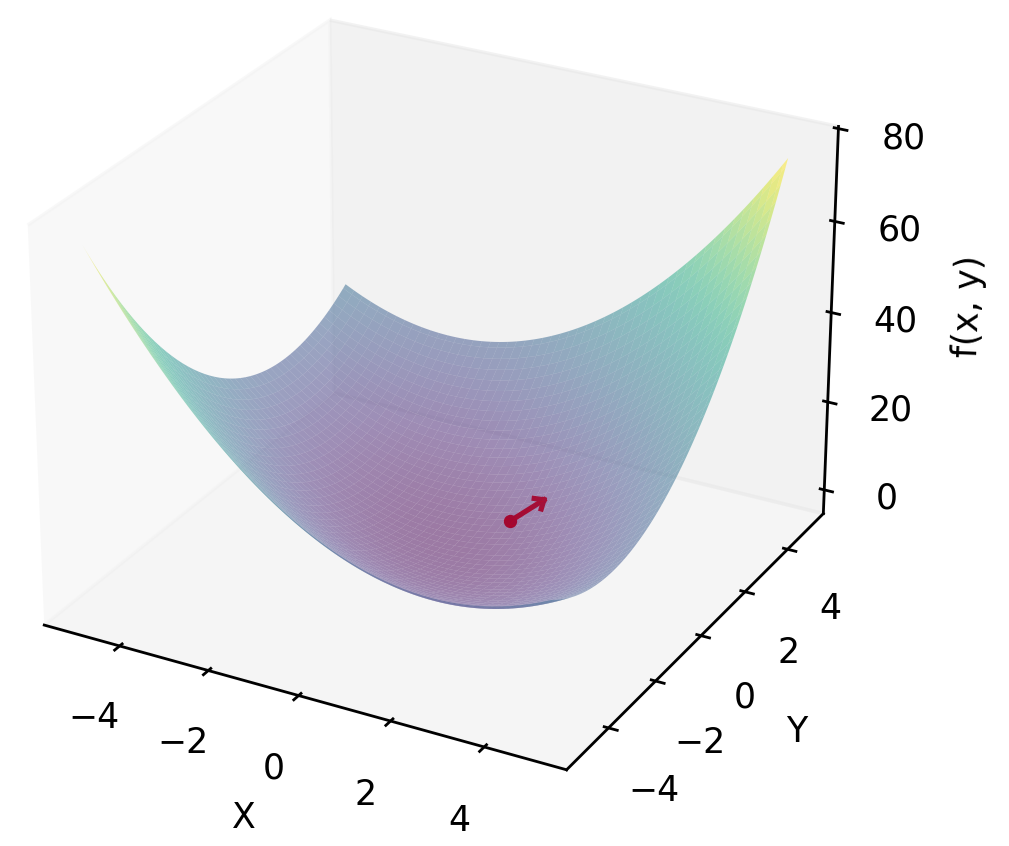
更一般地来说，一个多元函数在点处对的偏导数定义为：

* + 1. 梯度

多元函数关于每个变量都有偏导数，在点处，这些偏导数定义了一个向量。

这个向量称为在点的梯度。

例如：在处的梯度为。



1. 线性代数
   1. 标量与向量
      1. 标量与向量的概念
         * 1. 标量（scalar）

标量是一个单独的数，只有大小。

* + - * 1. 向量（vector）

向量由标量组成，有大小有方向。

行向量：

列向量：

* + 1. 向量运算
       - 1. 向量转置：列向量转置结果为行向量
         2. 向量相加：对应元素相加
         3. 向量与标量相乘：标量与向量每个元素相乘
         4. 向量内积：又称向量点乘，两向量对应元素乘积之和，结果为标量

两向量之间夹角表示为

* + 1. 向量范数

范数（norm）是具有“长度”概念的函数。

* + - * 1. L0范数（也称0范数）

例如：

* + - * 1. L1范数（也称和范数或1范数）

例如：

* + - * 1. L2范数（也称欧几里得范数或2范数）

例如：

* + - * 1. Lp范数
  1. 矩阵与张量
     1. 矩阵的概念

一个的矩阵（matrix）是一个有m行n列元素的矩形阵列。用表示所有实数矩阵的向量空间。

* + - * 1. 方阵：行数等于列数的矩阵
        2. 对角矩阵：主对角线以外元素全为0的方阵
        3. 单位矩阵：主对角线元素全为1的对角矩阵
    1. 矩阵乘法
       - 1. 矩阵乘法运算

两个矩阵的乘法仅当矩阵的列数和矩阵的行数相等时才能定义。如，，它们的乘积

例如：

特别地，矩阵与单位矩阵相乘等于矩阵本身：

或

例如：

* + - * 1. 矩阵乘法的性质

矩阵乘法满足结合律、左分配律和右分配律。不满足交换律即。

结合律：若，则

左分配律：若，则

右分配律：若，则

* + 1. 矩阵转置
       - 1. 矩阵转置运算

矩阵的转置是一个的矩阵，记为。其中的第个行向量是原矩阵的第个列向量；或者说，转置矩阵第行第列的元素是原矩阵第行第列的元素。

* + - * 1. 矩阵转置的性质
    1. 矩阵的逆

对于方阵，如果存在另一个方阵，使得成立，此时也同样成立。称为的逆矩阵。例如：

* + 1. 其他矩阵运算
       - 1. 矩阵的向量化

矩阵的向量化将矩阵的元素按列排列成一个的向量。

矩阵也可以转化为行向量，称为矩阵的行向量化。

例如：**。**

* + - * 1. 矩阵的直和

矩阵和矩阵的直和记作，它是一个的矩阵。

* + - * 1. 矩阵的内积

矩阵和矩阵的内积记作，它是两个矩阵对应元素乘积之和，是一个标量。

* + - * 1. 矩阵的Hadamard积

矩阵和矩阵的Hadamard积记作，它是两个矩阵对应元素的乘积，是一个的矩阵。

* + - * 1. 矩阵的Kronecker积

矩阵和矩阵的Kronecker积记作，它是矩阵中每个元素与矩阵的乘积，是一个的矩阵。

Kronecker积也称为直积或张量积。

* + 1. 张量

张量（tensor）可视为多维数组，是标量，1维向量和2维矩阵的n维推广。

例如：3维张量

* 1. 矩阵求导

矩阵求导的本质就是函数对变元的每个元素逐个求导，只是写成了向量、矩阵的形式。

为方便理解，首先对变元和函数作统一的符号规定：

为实向量变元。

为实矩阵变元。

为实标量函数，其变元为实向量。

为实标量函数，其变元为实矩阵。

为实向量函数，其变元为实向量。

为实向量函数，其变元为实矩阵。

为实矩阵函数，其变元为实向量。

为实矩阵函数，其变元为实矩阵。

* + 1. 常用求导结果
    2. Jacobian矩阵与梯度矩阵
       - 1. Jacobian矩阵

实标量函数在的偏导向量为的行向量：

当实标量函数的变元为实矩阵时，存在两种可能的定义：

和

其中与分别称为实标量函数关于实矩阵变元的Jacobian矩阵和行偏导矩阵。两者之间的关系为。

实矩阵函数，其实矩阵变元。为得到该函数的Jacobian矩阵，先通过列向量化，将的矩阵函数转换为的列向量：

然后，该列向量对列向量化的求偏导，得到的Jacobian矩阵：

其具体表达式为：

* + - * 1. 梯度矩阵

采用列向量形式定义的偏导算子称为列向量偏导算子，习惯称为梯度算子。

实标量函数的梯度向量为的列向量：

将矩阵变元列向量化，得到的梯度算子：

另外，可直接得到其梯度矩阵：

可以发现，实标量函数的梯度矩阵等于Jacobian矩阵的转置，。另外，实矩阵函数的梯度矩阵也是其Jacobian矩阵的转置，。

* + 1. 一阶实矩阵微分
       - 1. 矩阵的迹

方阵的主对角线元素之和为矩阵的迹（trace），记作：

标量的迹是其本身：

转置：

线性：

交换律：

* + - * 1. 矩阵微分的性质

矩阵微分用符号表示，定义为。

考虑标量函数的微分，有**：**

考虑矩阵乘积的微分矩阵，有**：**

上两例表明，实矩阵微分具有以下两个基本性质：

* 转置：矩阵转置的微分等于矩阵微分的转置，即。
* 线性：**。**

下面汇总矩阵微分常用的计算公式：

* + - * 1. 标量函数的矩阵微分

考虑标量函数的全微分：

令，则有：

这表明，标量函数的Jacobian矩阵与微分矩阵之间存在等价关系：

换言之，若函数的微分可以写作，则矩阵就是函数关于其变元向量的Jacobian矩阵。

考虑标量函数，其变元为的全微分：

利用行向量偏导与Jacobian矩阵的关系，上式可以改写为：

式中是标量函数的Jacobian矩阵。

利用向量化算子和迹函数之间的关系，令，，则有：

综上：

因此可以利用矩阵微分直接求标量函数的Jacobian矩阵和梯度矩阵。

二次型函数

标量函数可直接写成迹的形式：

根据可得：

矩阵的标量函数：迹

因为，所以有：

三矩阵乘积的迹函数

包含逆矩阵的迹函数

四矩阵乘积的迹函数

总结

标量函数总可以写为迹函数的形式：

无论出现在迹函数内的任何位置，总可以通过迹的交换律将其写到最右边：

对于，总可以通过迹函数的性质转换为**：**

1. 概率论
   1. 概率
      1. 概率的概念

概率是对事件发生的可能性的度量。通常将事件A的概率写作。

* + 1. 概率的计算

|  |  |
| --- | --- |
| **事件** | **概率** |
| **A** |  |
| **非A** |  |
| **A和B**  **（联合概率）** |  |
| **A或B** |  |
| **B的情况下A的概率**  **（条件概率）** |  |

例如：现有一个装有10个球的袋子，其中有6个红球和4个蓝球。从中随机抽取两个球。我们定义以下事件：

事件A：第一个抽到的是红球。

事件B：两个抽到的球都是红球。

* + - * 1. 计算联合概率

第一个球是红球的概率：

在第一个球是红球的情况下，两个球都是红球的概率：

联合概率：

* + - * 1. 计算条件概率

条件概率表示在已知两个球都是红球的情况下，第一个球是红球的概率。

两个球都是红球的概率：

在两个球都是红球的情况下，第一个球是红球的概率：

* 1. 似然函数
     1. 似然函数的概念

概率用于在已知一些参数的情况下，预测接下来在观测上所得到的结果。似然性则是用于在已知某些观测所得到的结果时，对有关事物之性质的参数进行估值。

似然函数是对参数的函数，其定义为在给定参数值的条件下，观察到某个特定数据的概率。换句话说，似然函数是一个关于参数的函数，而不是关于数据的函数。

如果我们有一个参数化的概率模型，其中是观测数据，是模型参数，似然函数定义为：

这里，表示在参数为的情况下，观察到数据的概率。

设有一组独立同分布的观测数据，并且这些数据服从某个分布（例如正态分布、二项分布等），比如服从参数为的某个分布，那么似然函数可以写作：

针对其中存在的乘法，可以使对数函数将其转化为加法：

* + 1. 极大似然估计

似然函数常用于极大似然估计。我们希望找到使似然函数最大化的参数。这意味着在给定观测数据的情况下，选择最可能生成这些数据的参数值。

例如，掷硬币3次，2次正面1次背面，能否依据此结果逆推出正面的概率；正面概率为0.5的概率为多少、正面概率为0.6的概率为多少；最有可能的正面概率是多少？

我们用代表硬币正面朝上的概率，用代表2次正面1次背面的结果

当正面概率为0.5时：

当正面概率为0.6时：

为了找出极大似然估计，对似然函数取对数并求导，使其等于0

解得，意味着当掷硬币3次，出现2次正面1次背面的结果时，硬币正面朝上的概率最有可能为。

* 1. 贝叶斯定理

贝叶斯定理（Bayes' Theorem）是概率论中的一个核心定理，用于描述在已有条件概率信息的基础上，如何更新或计算事件的概率。它以英国数学家托马斯·贝叶斯的名字命名。贝叶斯定理特别适合处理“逆向概率”问题，即从结果反推原因的概率。

贝叶斯定理建立在条件概率的基础上，假设有两事件，贝叶斯定理描述了在已知发生的情况下，发生的概率：

* ：后验概率，发生的情况下发生的概率。
* ：似然概率，发生的情况下发生的概率。
* ：先验概率，发生的概率。
* ：发生的概率，通常通过全概率公式计算。

在实际问题中通常不是直接给出，而是需要通过全概率公式计算。假设样本空间被一组互斥且完备的事件划分，则：

例如：某疾病发病率为1%，如果一个人有疾病，检测呈阳性的概率为95%；如果一个人没有疾病，检测呈阳性的概率为5%，现有一人检测结果呈阳性，问他真正患病的概率是多少？

检测呈阳性的人真正患病的概率为16.1%。

* 1. 概率分布
     1. 伯努利分布

伯努利分布（Bernoulli Distribution）又称两点分布或0-1分布，是一个离散概率分布，适用于只有两种可能结果的随机试验。

伯努利分布描述了一个随机变量，它只有两个可能的取值：

* ，发生概率为。
* ，发生概率为。

伯努利分布的概率质量函数可写为：

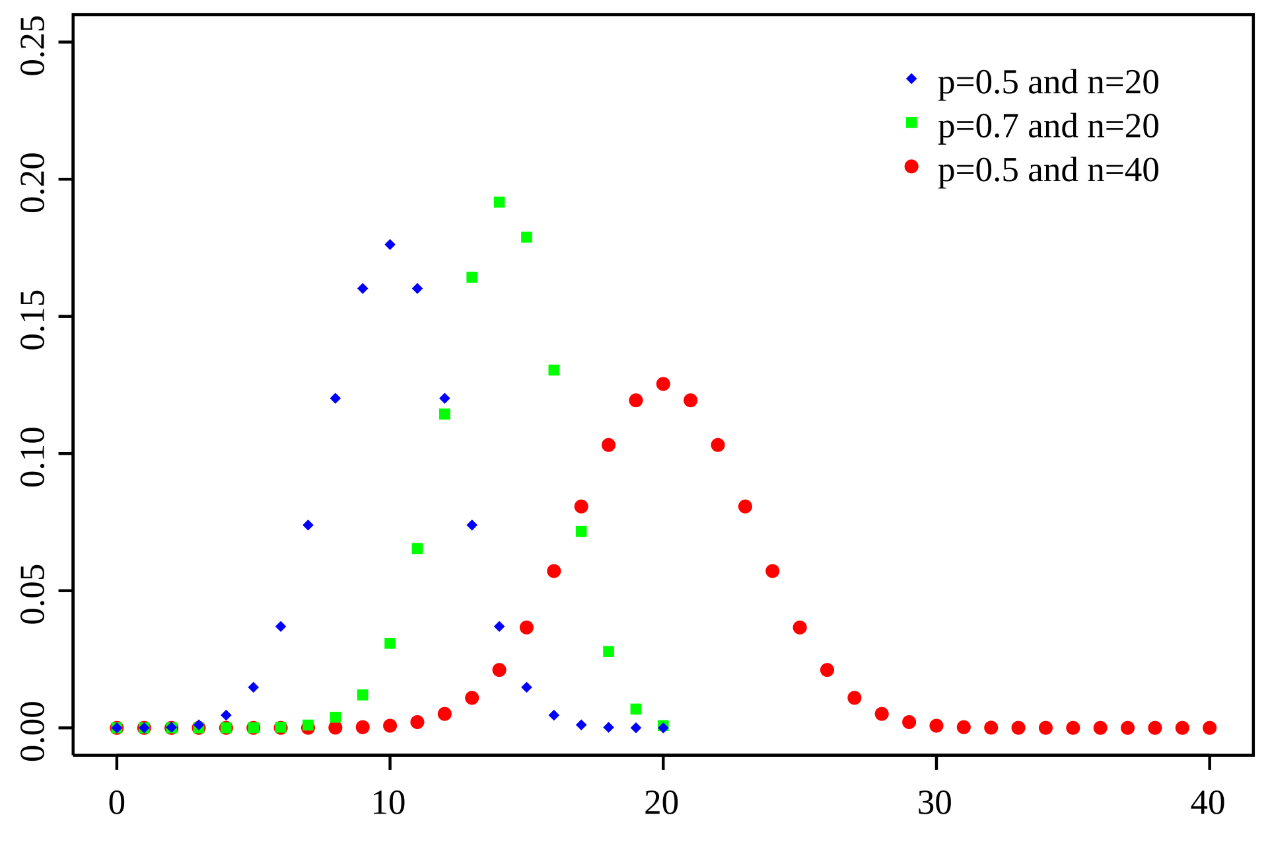
* + 1. 二项分布

二项分布（Binomial Distribution）是一种离散概率分布，它描述了在固定次数的独立伯努利试验（每次试验只有成功和失败两种结果）中，成功次数的概率。二项分布可以看作是伯努利分布的扩展，从单次试验推广到多次试验。

对于随机变量，有概率质量函数：

* 为试验的总次数
* 为成功的次数
* 为单次试验成功的概率
* 表示从𝑛次试验中选择𝑘次成功的组合数

二项分布概率质量函数图：



* + 1. 正态分布

正态分布（Normal Distribution）也称高斯分布，是常见的连续概率分布。正态分布在统计学上十分重要，经常用在自然和社会科学来代表一个不明的随机变量。

若随机变量服从一个平均数为、标准差为（）的正态分布，则记为，其概率密度函数

正态分布的期望可解释为位置参数，决定了分布的位置；其方差可解释为尺度参数，决定了分布的幅度。

中心极限定理指出，在特定条件下，一个具有有限均值和方差的随机变量的多个样本的平均值本身就是一个随机变量，其分布随着样本数量的增加而收敛于正态分布。因此，许多与独立过程总和有关的物理量，例如测量误差，通常可被近似为正态分布。

正态分布的概率密度函数曲线呈钟形，因此人们又经常称之为钟形曲线。我们通常所说的标准正态分布是位置参数，尺度参数的正态分布。

正态分布概率密度函数图：

