数学、线性代数、概率论

《线性代数及其应用》(美+第三版)（中文版）.pdf.pdf
46.50MB

2021李永乐王式安武忠祥复习全书 基础版（数学一二三）.pdf
163.60MB

数学课地址点我

整体概念

概率统计关注的是随机变量及其概率分布，以及如何通过观测数据来推断这些分布。可是，在解决很多问题的时候，我们不仅要关心单个变量之间的关系，还要进一步研究多个变量之间的关系，最典型的例子就是基于多个特征的信息检索和机器学习。在信息检索中，我们需要考虑多个关键词特征对最终相关性的影响，而在机器学习中，无论是监督式还是非监督式学习，我们都需要考虑多个特征对模型拟合的影响。在研究多个变量之间关系的时候，线性代数成为了解决这类问题的有力工具。

泰勒公式

https://www.matongxue.com/madocs/7.html
泰勒公式是用多项式函数去逼近光滑函数

$$f(x)=\frac{f(a)}{0!}+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\ldots+\frac{f^{(n)}(a)}{n!}(x-a)^n+R_n(x)$$

初等行变换有哪些?

- 把某一行的倍数加到另一行上
- 把两行对换
- 把某一行的所有元素乘以同一个非0数

矩阵经过初等行变换变为另一个矩阵，称他们是行等价的。

线性代数常见定义

1. 逆矩阵：设A是n阶矩阵，如果存在n阶矩阵B使得AB=BA=E成立，则称A是可逆矩阵，B是A的逆矩阵
2. 等价矩阵：矩阵A经过有限次初等变换变成矩阵B，则称A和B等价
3. 等价向量组：两个向量组I和II可以相互表出，称I和II为等价向量组
4. 正交矩阵：设A为n阶矩阵，若AA^T=A^TA=E，则称A为正交矩阵
5. 相似矩阵：

定义 设 $\boldsymbol{A},\boldsymbol{B}$ 都是 n 阶矩阵，若存在可逆矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{B}$，则称 $\boldsymbol{B}$ 是 $\boldsymbol{A}$ 的**相似矩阵**，或 $\boldsymbol{A}$ 相似于 $\boldsymbol{B}$，记成 $\boldsymbol{A}\sim\boldsymbol{B}$.

若 $\boldsymbol{A}\sim\boldsymbol{\Lambda}$，其中 $\boldsymbol{\Lambda}$ 是对角阵，则称 $\boldsymbol{A}$ 可相似对角化. $\boldsymbol{\Lambda}$ 是 $\boldsymbol{A}$ 的**相似标准形**.

两个矩阵相似的必要条件

$$\boldsymbol{A}\sim\boldsymbol{B}\begin{cases}\Rightarrow (1)\ \text{特征多项式相同，即}\ |\lambda\boldsymbol{E}-\boldsymbol{A}|=|\lambda\boldsymbol{E}-\boldsymbol{B}|;\\ \Rightarrow (2)\ r(\boldsymbol{A})=r(\boldsymbol{B});\\ \Rightarrow (3)\ \boldsymbol{A},\boldsymbol{B}\ \text{有相同的特征值};\\ \Rightarrow (4)\ |\boldsymbol{A}|=|\boldsymbol{B}|=\prod_{i=1}^{n}\lambda_i;\\ \Rightarrow (5)\ \sum_{i=1}^{n}a_{ii}=\sum_{i=1}^{n}b_{ii}=\sum_{i=1}^{n}\lambda_i.\end{cases}$$

n阶方阵A可对角化的充分必要条件是A有n个线性无关的特征向量。

6. 合同矩阵：

定义（合同） 设 $\boldsymbol{A},\boldsymbol{B}$ 是两个 n 阶方阵，若存在可逆阵 $\boldsymbol{C}$，使得 $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}=\boldsymbol{B}$，则称 $\boldsymbol{A}$ 合同于 $\boldsymbol{B}$，记成 $\boldsymbol{A}\simeq\boldsymbol{B}$.

7. 正定矩阵

即A的全部顺序主子式大于0或A的全部特征值大于0

定义（正定） 若对于任意的非零向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，恒有

$$f(x_1,x_2,\cdots,x_n)=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}>0$$

则称二次型 f 为**正定二次型**，对应矩阵为**正定矩阵**.

什么是线性相关，什么是线性无关？

(平凡解就是0解。)

定义 $\mathbb{R}^n$ 中一组向量 $\{v_1, \cdots, v_p\}$ 称为线性无关的，若向量方程

$$x_1 v_1 + x_2 v_2 + \cdots + x_p v_p = 0$$

仅有平凡解. 向量组（集）$\{v_1, \cdots, v_p\}$ 称为线性相关的，若存在不全为零的权 $c_1, \cdots, c_p$，使

$$c_1 v_1 + c_2 v_2 + \cdots + c_p v_p = 0 \tag{2}$$

什么是向量空间?

假设 V 是 Fn 的非零子集，如果对任意的向量 x、向量 y∈V，都有 (x+y)∈V，我们称为 V 对向量的加法封闭；对任意的标量 k∈V，向量 x∈V，都有 kx 属于 V，我们称 V 对标量与向量的乘法封闭。如果 V 满足向量的加法和乘法封闭性，我们就称 V 是 F 上的向量空间。

The basis of a vector space is a set of linearly independent vectors that span the full space

向量空间的一组基是张成该空间的一个线性无关向量集

$$(AB)^T = B^T A^T. \quad (AB)^{-1} = B^{-1} A^{-1} \quad (A^T)^{-1} = (A^{-1})^T.$$

什么是矩阵的秩?

矩阵的秩=存在的最高阶不为零的子式的阶数=行向量组的秩=列向量组的秩=向量组的极大线性无关组的向量个数

子式：

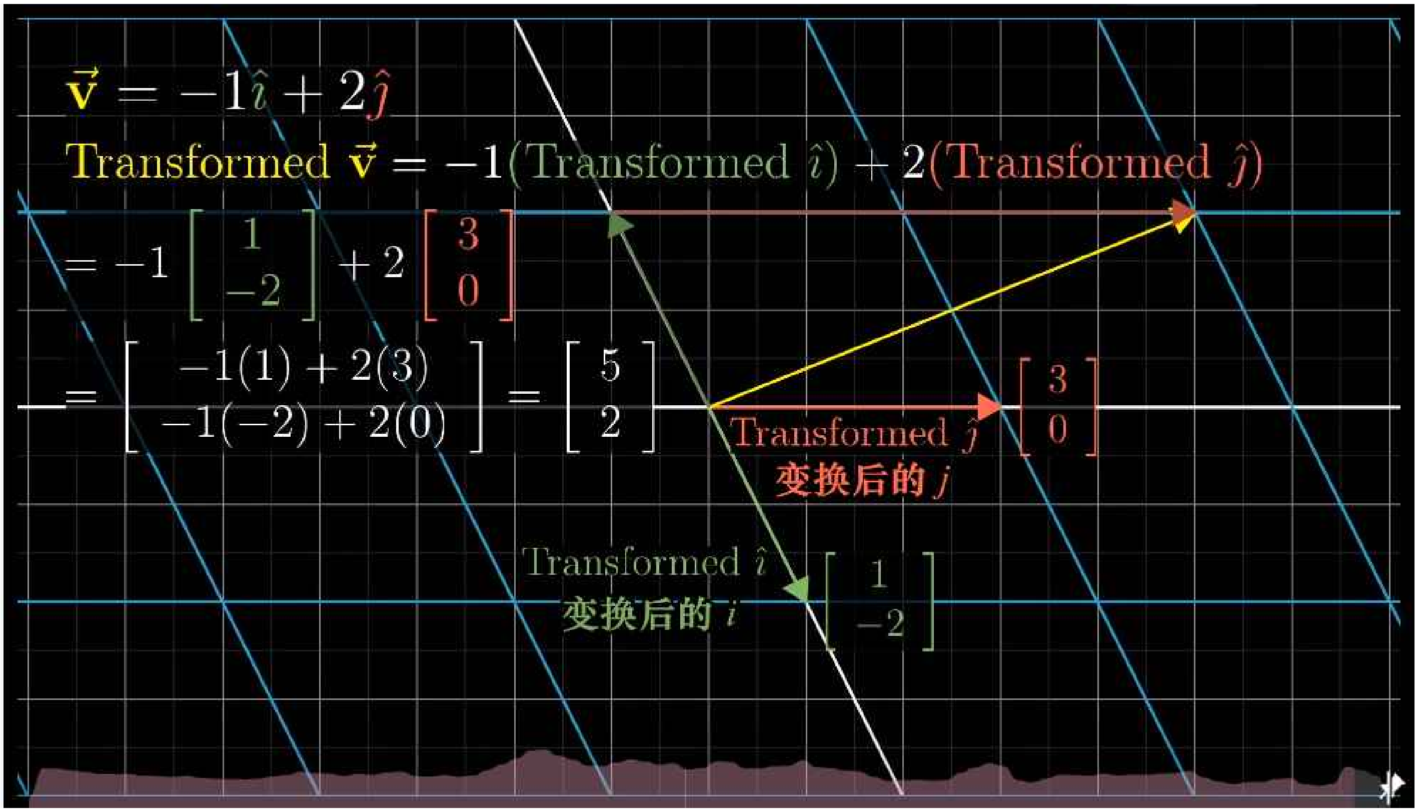
定义 在 $m \times n$ 矩阵 A 中，任取 k 行与 k 列($k \leqslant m, k \leqslant n$)，位于这些行与列的交叉点上的 k^2 个元素按其在原来矩阵 A 中的次序可构成一个 k 阶行列式，称其为矩阵 A 的一个 k 阶子式.

初等变换矩阵的秩不变。

如何理解矩阵乘向量？（视频P4）

记矩阵A=[**a1**,**a2**,**a3**]，列向量线性无关。列向量b=[b1,b2,b3]，那么A*b可以视为，在**a1 a2 a3**这三个向量张成的空间内的一个向量。其每个轴的长度分别为 b1 b2 b3，即结果

c=A***b**=b1***a1**+b2***a2**+b3***a3**



进而，记基本的三维空间由xyz轴构成，三个基为**e1 e2 e3**，有

c1=[**e1**,**e2**,**e3**]***b**=b1***e1**+b2***e2**+b3***e3**

e1 e2 e3表示了**c1**，将空间扭曲下，三个基变为**a1 a2 a3**，**c1**变为**c2**，但是他们的权值（b1 b2 b3）不变，我们有：

c2=[**a1**,**a2**,**a3**]***b**=b1***a1**+b2***a2**+b3***a3**

总结：矩阵乘向量，其实就是根据矩阵列向量想象出由基本的三维空间扭曲后的空间，因为矩阵列向量就是扭曲空间的多个基，再根据权值延长或缩减基，加在一起就得到结果。

特别地，若**a1 a2 a3**线性相关，他们扭曲后的空间就被压缩成一条线，结果**c**必然在这条线上。

特征值和特征向量

定义 设 A 是 n 阶矩阵，如果存在一个数 λ 及非零的 n 维列向量 $\boldsymbol{\alpha}$，使得

$$A\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha}$$

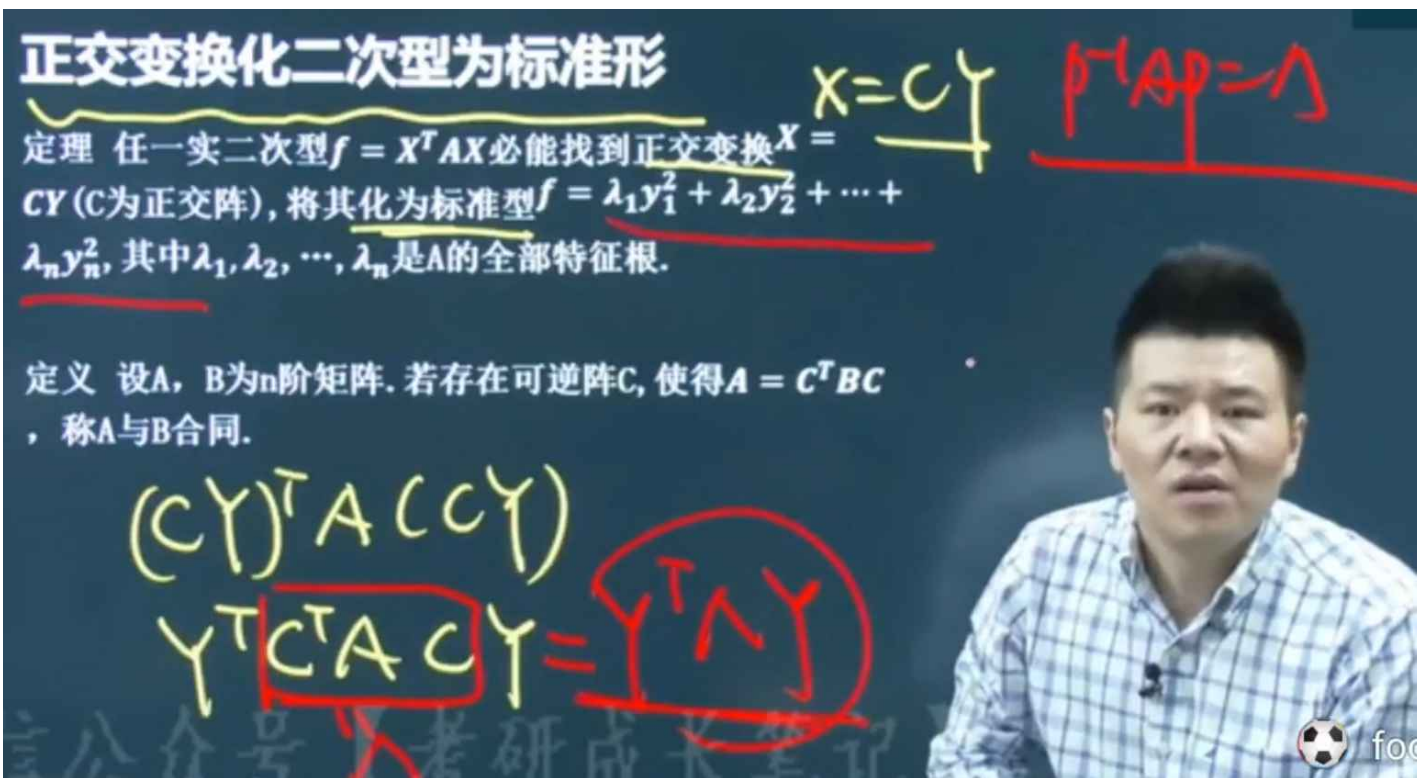
成立，则称 λ 是矩阵 A 的一个**特征值**，称非零向量 $\boldsymbol{\alpha}$ 是矩阵 A 属于特征值 λ 的一个**特征向量**.

求特征向量：先通过|λE-A|=0，解出不同的特征值，再通过Aα=λα求出不同的特征向量α

如何把自然界里物体的属性，转换为能够用数字表达的特征？由于特征有很多维，因此我们可以使用向量来表示某个物体的特征。其中，向量的每个元素就代表一维特征，而元素的值代表了相应特征的值，我们称这类向量为特征向量（Feature Vector）。需要注意的是，这个特征向量和矩阵的特征向量（Eigenvector）是两码事。那么矩阵的特征向量是什么意思呢？矩阵的几何意义是坐标的变换。如果一个矩阵存在特征向量和特征值，那么这个矩阵的特征向量就表示了它在空间中最主要的运动方向。

从定义出发，Ax=cx：A为矩阵，c为特征值，x为特征向量。矩阵A乘以x表示，对向量x进行一次转换（旋转或拉伸）（是一种线性转换），而该转换的效果为常数c乘以向量x（即只进行拉伸）。我们通常求特征值和特征向量即为求出该矩阵能使哪些向量（当然是特征向量）只发生拉伸，使其发生拉伸的程度如何（特征值大小）。这样做的意义在于，看清一个矩阵在那些方面能产生最大的效果（power），并根据所产生的每个特征向量（一般研究特征值最大的那几个）进行分类讨论与研究。

正交变换是什么？



正交变换就是将列向量X化为正交矩阵C与列向量Y的乘积，即X=CY（我没说反吧）

正交变换能借助一个正交矩阵C（需要我们自己找出这个正交矩阵）将二次型化为标准型，也就是将一个普通的矩阵化为对角矩阵。为什么可以呢？看上图，把C找出来后，令X=CY，代入。这里注意下可逆矩阵能够使得矩阵A相似对角化，那么正交矩阵也可以。

朴素贝叶斯？

（“朴素”的原因是假设各个特征是相互独立的。）

$$P(c|o)=P(c|f_i,f_j)=P(c|f_i)\times P(c|f_j)$$
$$=\frac{P(f_i|c)\times P(c)}{P(f_i)}\times\frac{P(f_j|c)\times P(c)}{P(f_j)}$$

其中，c 表示一个分类（class），f 表示属性对应的数据字段（field）。假设我们有一个新的水果，它的形状是圆形，口感是甜的，那么根据朴素贝叶斯，它属于苹果、甜橙和西瓜的概率分别是多少呢？我们先来计算一下，它属于苹果的概率有多大

$$\begin{aligned}P(apple|o)&=P(apple|shape-2,taste-2)\\&=P(apple|shape-2)\times P(apple|taste-2)\\&=\frac{P(shape-2|apple)\times P(apple)}{P(shape-2)}\times\frac{P(taste-2|apple)\times P(apple)}{P(taste-2)}\\&=\frac{0.33\times0.30}{0.30}\times\frac{0.01\times0.30}{0.50}\\&=0.33\times0.006\\&=0.00198\end{aligned}$$

再解释下贝叶斯公式，一般情况下，我们知道某事物有特征A、B，想知道它属于某一类的概率（后验概率）是多少：

`P（属于某类|具有特征A、B）=P(属于某类|具有特征A） * P（属于某类|具有特征B)`

而

`P(属于某类|具有特征A)=P（属于某类且具有特征A）/P（所有事物中特征A出现的概率）=`

`P（具有特征A|属于某类）*P(在特征A中，某类出现的概率)/P(所有事物中特征A出现的概率)`

解释下右边出现的三个概率：

- P（具有特征A|属于某类）：在样本中，该类事物具有 特征A的频率，如样本的所以苹果中，味道是甜的频率。--似然概率(似然函数)
- P(在特征A中，某类出现的概率)：在所有水果样本观测值中，味道是甜的有西瓜、葡萄等，其中苹果出现的频率--先验概率
- P(所有事物中特征A出现的概率)：在所有水果中，味道是甜的样本出现的频率。--先验概率

> 统计学中，似然函数是一种关于统计模型参数的函数。给定输出x时，关于参数θ的似然函数L(θ|x)（在数值上）等于给定参数θ后变量X的概率：L(θ|x)=P(X=x|θ)。
> 概率描述了已知参数时的随机变量的输出结果；似然则用来描述已知随机变量输出结果时，未知参数的可能取值。例如，对于“一枚正反对称的硬币上抛十次”这种事件，我们可以问硬币落地时十次都是正面向上的“概率”是多少；而对于“一枚硬币上抛十次”，我们则可以问，这枚硬币正反面对称的“似然”程度是多少。

常见离散随机变量的分布

伯努利试验是在同样的条件下重复地、相互独立地进行的一种随机试验，其特点是该随机试验只有两种可能结果：发生或者不发生。我们假设该项试验独立重复地进行了n次，那么就称这一系列重复独立的随机试验为n重伯努利试验

定义 如果随机变量 X 有分布律

X	0	1
P	$1-p$	p

$0<p<1$，则称 X 服从参数为 p 的 $0-1$ 分布，或称 X 具有 **$0-1$ 分布**.

定义 如果随机变量 X 有分布律

$$P\{X=k\}=C_n^k p^k q^{n-k},\ k=0,1,2,\cdots,n$$

其中 $0<p<1$，$q=1-p$，则称 X 服从参数为 n,p 的**二项分布**，记作 $X\sim B(n,p)$.

在 n 重伯努利试验中，若每次试验成功率为 $p(0<p<1)$，则在 n 次独立重复试验中成功的总次数 X 服从二项分布.

当 $n=1$ 时，不难验证二项分布就退化成 $0-1$ 分布. 所以 $0-1$ 分布也可以记为 $B(1,p)$.

定义 如果随机变量 X 的分布律为

$$P\{X=k\}=pq^{k-1},\ k=1,2,\cdots$$

其中 $0<p<1$，$q=1-p$，则称 X 服从参数为 p 的几何分布，或称 X 具有**几何分布**.

【注】 在独立地重复做一系列伯努利试验中，若每次试验成功率为 $p(0<p<1)$，则在第 k 次试验时才首次试验成功的概率服从几何分布.

定义 如果随机变量 X 的分布律为

$$P\{X=k\}=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n},\ k=l_1,\cdots,l_2$$

其中 $l_1=\max(0,n-N+M)$，$l_2=\min(M,n)$. 则称随机变量 X 服从参数为 n,N,M 的**超几何分布**.

如果 N 件产品中含有 M 件次品，从中任意一次取出 n 件(或从中一件接一件不放回地取 n 件)，令 $X=$ 抽取的 n 件产品中的次品件数，则 X 服从参数为 n,N,M 的超几何分布.

如果 N 件产品中含有 M 件次品，从中一件接一件有放回地取 n 次(即每次取出记录后就放回，再取下一个)，则 X 服从 $B\left(n,\dfrac{M}{N}\right)$.

定义 如果随机变量 X 的分布律为 $P\{X=k\}=\dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$，$k=0,1,2,\cdots$，其中 $\lambda>0$ 为常数，则称随机变量 X 服从参数为 λ 的**泊松分布**，记为 $X\sim P(\lambda)$.

【注】 在一段时间内电话总机接到的呼叫次数、候车的旅客数、保险索赔的次数等都服从泊松分布.

常见连续随机变量的分布

定义 如果连续型随机变量 X 的概率密度为

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

则称 X 在区间 $[a,b]$ 上服从**均匀分布**，记作 $X\sim U[a,b]$.

定义 如果连续型随机变量 X 的概率密度为

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

则称 X 服从参数为 λ 的**指数分布**，记作 $X\sim E(\lambda)$.

定义 如果随机变量 X 的概率密度为

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

其中 μ,σ 为常数且 $\sigma>0$，则称 X 服从参数为 μ,σ 的**正态分布**，记作 $X\sim N(\mu,\sigma^2)$.

当 $\mu=0,\sigma^2=1$ 时，即 $X\sim N(0,1)$，称 X 服从**标准正态分布**，此时用 $\varphi(x)$ 表示 X 的概率密度，即 $$\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}},-\infty<x<+\infty$$

常用随机变量的期望和方差

常用随机变量的数学期望和方差

(1) 0－1 分布

$E(X) = p, \quad D(X) = p(1-p).$

(2) 二项分布, $X \sim B(n,p)$

$E(X) = np, \quad D(X) = np(1-p).$

(3) 泊松分布, $X \sim P(\lambda)$

$E(X) = \lambda, \quad D(X) = \lambda.$

(4) 几何分布, $P\{X = k\} = p(1-p)^{k-1}, k = 1,2,\cdots, 0 < p < 1.$

$$E(X) = \frac{1}{p}, \quad D(X) = \frac{1-p}{p^2}.$$

(5) 均匀分布, $X \sim U(a,b)$

$$E(X) = \frac{a+b}{2}, \quad D(X) = \frac{(b-a)^2}{12}.$$

(6) 指数分布, $X \sim E(\lambda)$

$$E(X) = \frac{1}{\lambda}, \quad D(X) = \frac{1}{\lambda^2}.$$

(7) 正态分布, $X \sim N(\mu,\sigma^2)$

$E(X) = \mu, \quad D(X) = \sigma^2.$

大数定律

大数定律告诉我们，随着数据样本的增加，频率变得稳定，且趋向于概率。

马尔科夫不等式：

马尔科夫不等式

如果X是只取非负值的随机变量，则对于任意$a>0$:

$$P\{X \geq a\} \leq \frac{EX}{a}$$

切比雪夫不等式：

切比雪夫不等式

如果 X 是具有期望 μ 和方差 σ^2 的随机变量，则对于任意 $\varepsilon > 0$

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

大数定律：

定义 设 $X_1, X_2, \cdots, X_n, \cdots$ 是一个随机变量序列，A 是一个常数，如果对任意 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P\{|X_n - A| < \varepsilon\} = 1$$

则称随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ **依概率收敛于常数** A，记作 $X_n \xrightarrow{P} A$.

切比雪夫大数定律 设 $X_1, X_2, \cdots, X_n, \cdots$ 为两两不相关的随机变量序列，存在常数 C，使 $D(X_i) \leqslant C \quad (i = 1, 2, \cdots)$，则对任意 $\varepsilon > 0$，有

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

伯努利大数定律 设随机变量 $X_n \sim B(n, p)$，$n = 1, 2, \cdots$，则对于任意 $\varepsilon > 0$，有

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

辛钦大数定律 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，具有数学期望 $E(X_i) = \mu$，$i = 1, 2, \cdots$，则对任意 $\varepsilon > 0$ 有

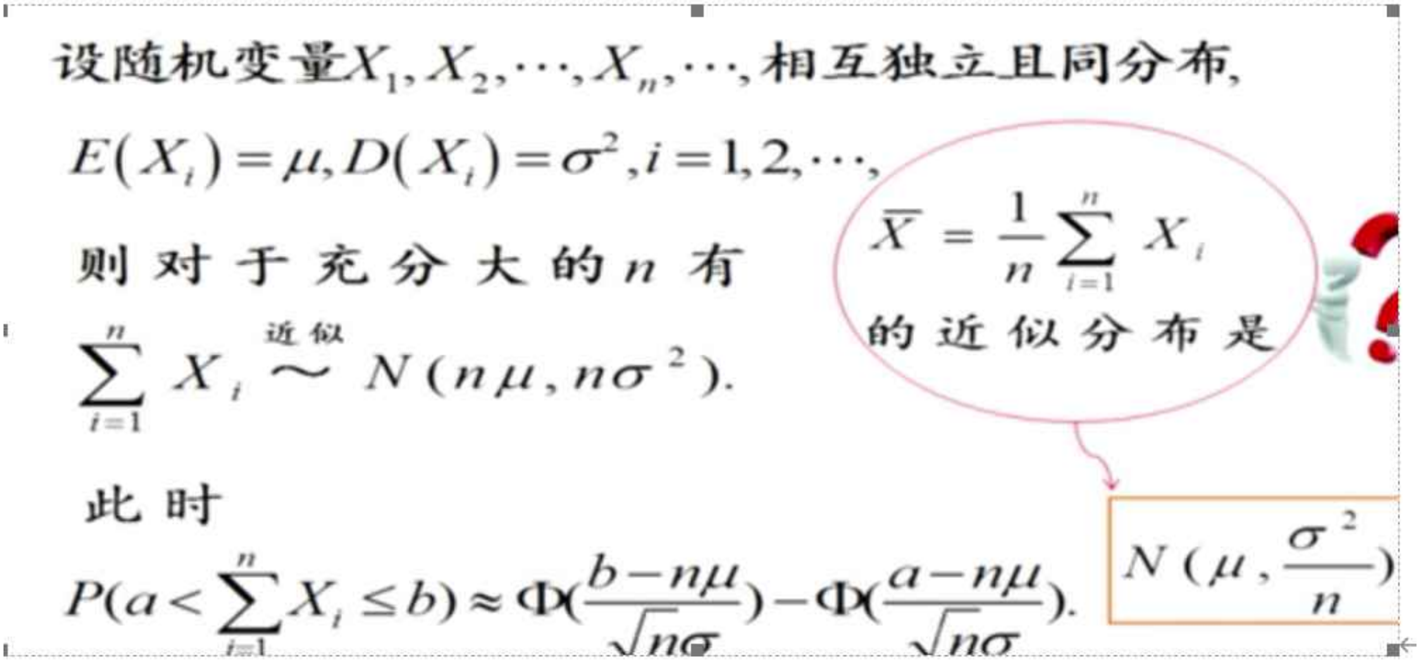
$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \varepsilon\right\} = 1$$

我的理解：观测值趋于已知常数、频率趋于概率。就像摇骰子，统计无数次，点数为6的频率等于点数为6这件事发生的概率。

中心极限定律

指出了大量随机变量近似服从正态分布的条件。

在自然界与生产中，一些现象受到许多相互独立的随机因素的影响，如果每个因素所产生的影响都很微小时，**总的影响**可以看作是服从正态分布的



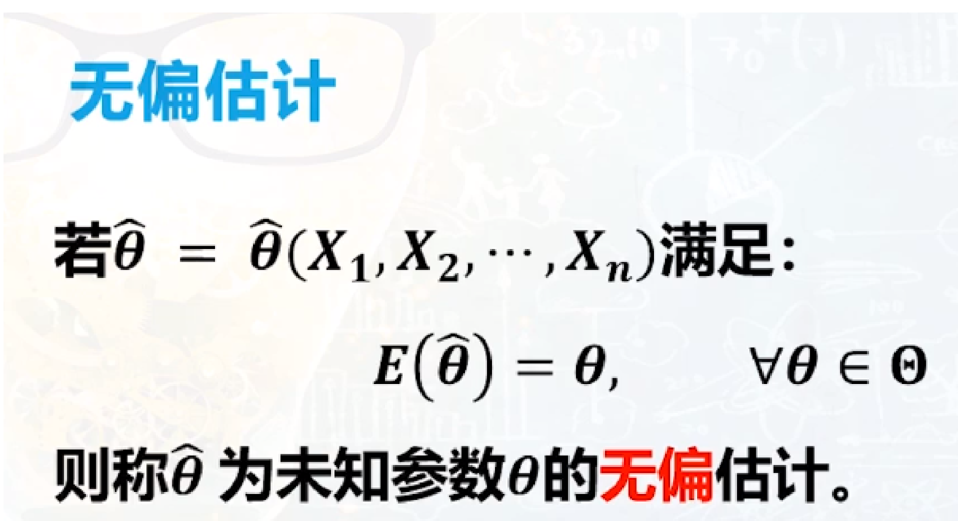
点估计

定义：设总体X的分布函数的形式已知，但它的一个或多个参数未知，借助于总体X的一个样本来估计总体未知参数的问题称为参数的点估计问题。

参数估计分为点估计和区间估计。

点估计有：

- 无偏估计



注意，大写的X1表示统计量，是随机的；小写的x1表示观测值，是确定的。参数θ是确定的，对它的估计量称为S，如果S的均值是确定的一个数——θ，那么就称S是θ的无偏估计。

例：若样本$X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2)$

试证：

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

为未知参数σ^2的**无偏**估计。

- 有效估计

如果两个估计量都是无偏估计，他们中方差最小的称为有效估计，即有效性。

- 一致性/相合性

当样本量越来越多时，估计量和真实值之间的差距是不是越来越小?

相合性

相合性

设$\theta \in \Theta$为未知参数，

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$$

为θ的一个估计，若当$n \to +\infty$时，

对于任意的$\varepsilon > 0$均有

$$P\left(\left|\hat{\theta} - \theta\right| \geq \varepsilon\right) \to 0,$$

则称$\hat{\theta}$是θ的相合估计。

例题：

例：若样本$X_1, X_2, \cdots, X_n \overset{\text{iid}}{\sim} f(x; \theta)$

其中$\theta = E(X_1)$，且$Var(X_1) < +\infty$试证：$\overline{X}$是θ的相合估计。

证：由切比雪夫不等式，对于任意$\varepsilon > 0$有：

$$P\left(\left|\hat{\theta} - \theta\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} Var(\hat{\theta})$$

$$= \frac{1}{\varepsilon^2} \frac{Var(X_1)}{n} \to 0, \qquad n \to +\infty.$$

- 矩估计

例. 设总体X服从$[0, \theta]$上的均匀分布，其密度函数为

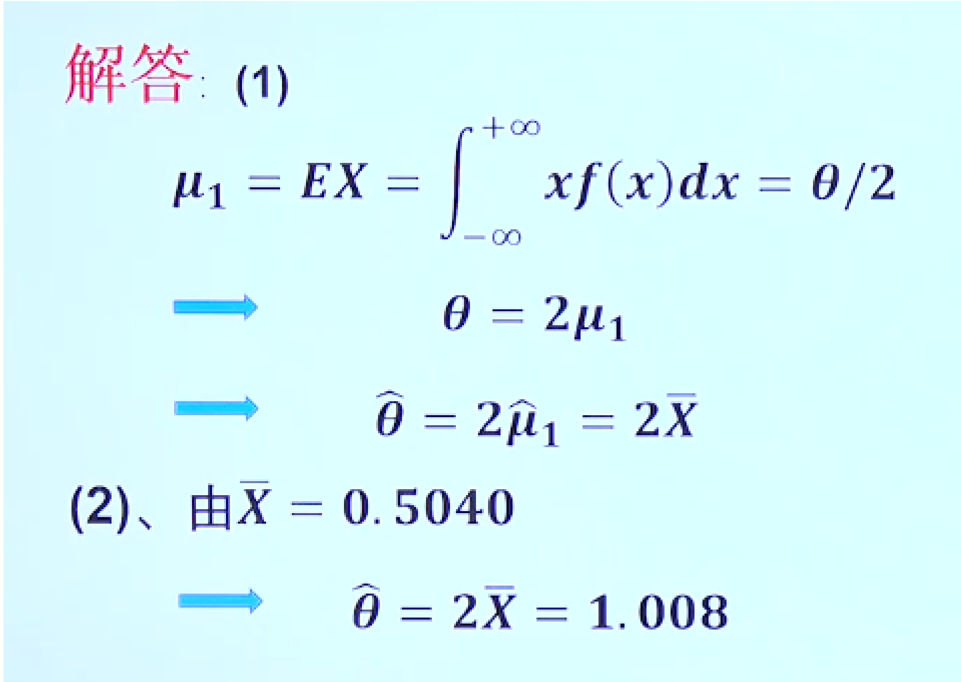
$$f(x) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq \theta \\ 0, & 其他 \end{cases}$$

(1)、求未知参数θ的矩估计量;
(2)、若样本观察值为

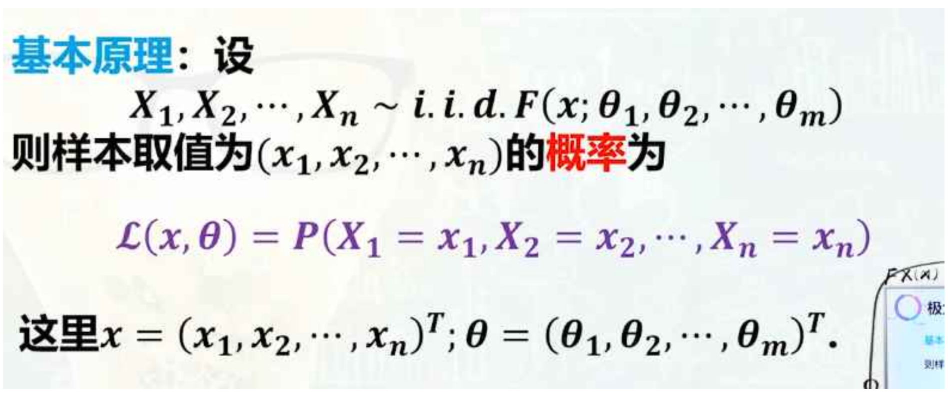
$$0.3, 0.7, 0.37, 0.61, 0.54$$

求θ的矩估计值。

先求出总体均值EX，因为密度函数中包含参数θ，所以这个均值也一定包含参数θ。
那么θ就成为了总体均值的函数，将总体均值用一阶原点矩（即样本均值，所有观测值的平均值，是个确定的值）替换掉即可。
本题用的是一阶矩，用高阶矩来做也行，但是一般用最低阶矩。



- 极大似然估计



X1、X2等独立同分布。设样本取到值分别为x1，x2等的概率为L（x，θ），称为极大似然函数。求解θ取何值时L最大，就是进行极大似然估计。若F是离散分布，L为概率连乘；若为连续分布，L与概率连乘正相关。所以将极大似然函数表示为每个具体点的密度函数连乘，对于固定的x来说，它是θ的函数。

例题：

【例 3】 设某种元件的使用寿命 X 的概率密度为

$$f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\leqslant 0\end{cases}\text{，其中 }\lambda>0\text{ 为未知参数}$$

又设 $x_1,x_2,\cdots,x_n$ 是 X 的一组样本观测值，求参数 λ 的最大似然估计值.

【解】 似然函数为

$$L(\lambda)=\prod_{i=1}^{n}f(x_i;\lambda)=\begin{cases}\lambda^n e^{-\lambda\sum\limits_{i=1}^{n}x_i}, & \begin{matrix}x_1\\x_2\\\vdots\\x_n\end{matrix}>0\\ 0, & \text{其他}\end{cases}.$$

当 $x_1,x_2,\cdots,x_n>0$ 时，$L(\lambda)>0$，取对数得

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

解得 $\hat{\lambda}=\frac{1}{x}$.

首先要知道，似然函数是关于观测值x1、x2……xn以及未知参数θ的函数，但是变量只有θ，它的值等于值为x1、x2、xn的概率密度连乘。似然函数的思路与概率分布的函数相反，前者是不知道参数，由观测值估计参数，后者是由参数计算概率。似然函数值越大，表明在θ=θ的情况下，取得这些观测值的概率越大，所以我们要求哪个θ能使得似然函数取到最大值，这个值就是θ的最大似然估计值。

如何求sin0.1?

- 已知cos(Π/3)=0.5，又sin2x=根号下（1-cos^2（2x））,cos2x=1-2sin^2(x)，可以根据sin2x求得sinx。依此迭代求得sin0.1
- 麦克劳林展开式来求
- 画直角三角形

蒙特卡洛求不规则图形面积

将不规则图形放在一个更大的规则图形上，随机撒点，计算落入不规则图形的概率。这个概率乘以规则图形的面积就是不规则图形的面积。

两事件相关、独立、对立、互斥?

- 相关是指，相关系数不等于0，不相关是指相关系数等于0

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

其中cov(X,Y)=E(XY)-E(X)*E(Y)=cov(Y,X)

- 独立是指，A事件的发生并不影响B事件的发生，即P(AB)=P(A)*P(B)，独立一定不相关
- 对立是指，A事件和B事件必发生一个，要么发生A，要么发生B
- 互斥是指，A事件和B事件不可能同时发生

全概率公式

全概率公式为概率论中的重要公式，它将对一复杂事件A的概率求解问题转化为了在不同情况下发生的简单事件的概率的求和问题。
内容：如果事件B1、B2、B3…Bn 构成一个完备事件组，即它们两两互不相容，其和为全集；并且P（Bi)大于0，则对任一事件A有
P(A)=P(A|B1)P(B1) + P(A|B2)P(B2) + ... + P(A|Bn)P(Bn)。

贝叶斯公式

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

卷积公式?

https://www.zhihu.com/question/22298352

教科书上一般定义函数 f, g 的卷积 $f * g(n)$ 如下：

连续形式：

$$(f * g)(n) = \int_{-\infty}^{\infty} f(\tau)g(n-\tau)d\tau$$

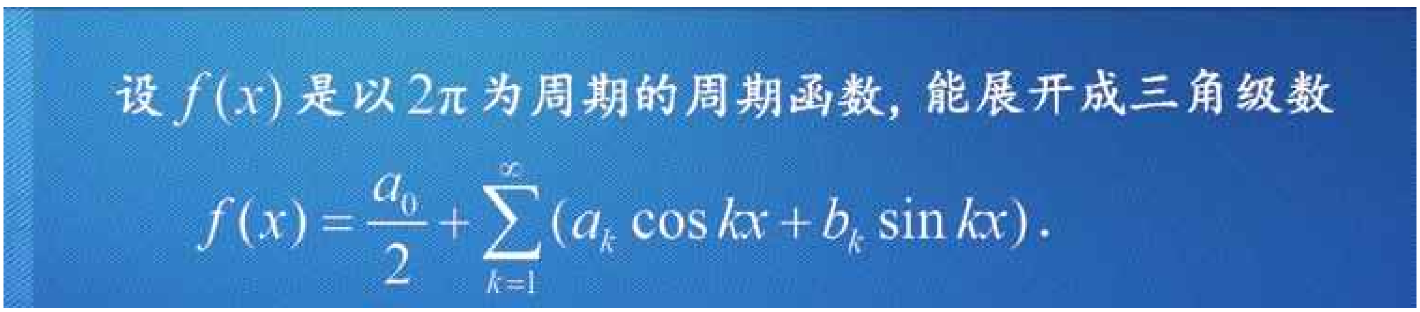
离散形式：

$$(f * g)(n) = \sum_{\tau=-\infty}^{\infty} f(\tau)g(n-\tau)$$

在概率论中，已知随机变量X和Y的概率密度，求Z=X+Y的概率密度时，需要用到卷积。
对卷积这个名词的理解：**所谓两个函数的卷积，本质上就是先将一个函数翻转，然后进行滑动叠加。**
~~在连续情况下，叠加指的是对两个函数的乘积求积分，在离散情况下就是加权求和，为简单起见就统一称为叠加。~~
~~整体看来是这么个过程：~~
~~翻转——>滑动——>叠加——>滑动——>叠加——>滑动——>叠加.....~~

信号中的卷积：
卷积法的原理是根据线性定常电路的性质(齐次性、叠加性、时不变性、积分性等),借助电路的单位冲激响应h(t),求解系统响应的工具,系统的激励一般都可以表示为冲击函数和激励的函数的卷积
图像里面的卷积，输入一幅图像f(x,y)，经过特定设计的卷积核g(x,y)进行卷积处理以后，输出图像将会得到模糊，边缘强化等各种效果。

傅里叶级数



将一个周期的信号分解成无限多分开的（离散的）正弦波。
- 傅里叶变换可以将一个时间域信号变换到频率域，可以用于修复图片上的污点。
- 可以大大减少运算复杂度。因为正弦余弦函数有奇偶性和对称性，微分积分也简单
- 用在计算机网络，说明在物理介质上发送模拟信号或者数字信号 在频域空间本质是一样

怎样在球面上均匀排列许多点?

各点均满足球面方程（用球坐标系），求解让各点之间距离的最小。这样定义出来的问题叫 Tammes problem，是密铺问题（Packing problems）的一个特例

逻辑题

10框苹果，有9框每个半斤重，有一筐每个四两，如何通过只称重一次，找出那个不一样的
答：第一个框取一个、第二个框取两个、依次类推，第10个框取十个，一起称重。如果全都是半斤重，结果应该为55*0.5=22.5，但实际上如果第i个框是4两的，就少了i*0.1斤，也就是说实际称得的重量加上i*0.1=22.5。由此可以解出来i等于多少，就是那个4斤的框。

复习打卡

- [x] ~~我复习并且记忆了一遍~~
- [x] ~~我复习并且记忆了一遍~~
- [x] ~~我复习并且记忆了一遍~~
- [x] ~~我复习并且记忆了一遍~~
- [x] ~~我复习并且记忆了一遍~~
- [x] ~~我复习并且记忆了一遍~~
- [x] ~~我复习并且记忆了一遍~~
- [x] ~~我复习并且记忆了一遍~~
- [x] ~~我复习并且记忆了一遍~~
- [x] ~~我复习并且记忆了一遍~~
- [x] ~~我复习并且记忆了一遍~~