第五章　概率与概率分布

李德山

四川师范大学商学院

2022 年 3 月 21 日

Contents

问题的提出

- 日常生活中，你遇到过哪些现象符合概率分布？
- 概率分布学了对我有啥用？
- 概率分布就很好，为什么还要提出概率密度的概念？

1. 随机事件及其概率
2. 离散型随机变量及其分布
3. 连续型随机变量的概率分布

随机事件及其概率

- 实验：在相同条件下，对事物或现象所进行的观察。例如：掷一枚骰子，观察其出现的点数
- 试验的特点
  - 可以在相同的条件下重复进行
  - 每次试验的可能结果可能不止一个，但试验的所有可能结果在试验之前是确切知道的
  - 在试验结束之前，不能确定该次试验的确切结果

随机事件及其概率

- 实验：在相同条件下，对事物或现象所进行的观察。例如：掷一枚骰子，观察其出现的点数
- 试验的特点
  - 可以在相同的条件下重复进行
  - 每次试验的可能结果可能不止一个，但试验的所有可能结果在试验之前是确切知道的
  - 在试验结束之前，不能确定该次试验的确切结果

随机事件及其概率

- 实验：在相同条件下，对事物或现象所进行的观察。例如：掷一枚骰子，观察其出现的点数
- 试验的特点
  - 可以在相同的条件下重复进行
  - 每次试验的可能结果可能不止一个，但试验的所有可能结果在试验之前是确切知道的
  - 在试验结束之前，不能确定该次试验的确切结果

随机事件及其概率

- 事件 (event)：随机试验的每一个可能结果 (任何样本点集合)。例如：掷一枚骰子出现的点数为 3
- 随机事件 (random event)：每次试验可能出现也可能不出现的事件。例如：掷一枚骰子可能出现的点数
- 必然事件 (certain event)：每次试验一定出现的事件，用 Ω 表示。例如：掷一枚骰子出现的点数小于 7
- 不可能事件 (impossible event)：每次试验一定不出现的事件，用 Φ 表示。例如：掷一枚骰子出现的点数大于 6

随机事件及其概率

- 基本事件 (elementary event): 一个不可能再分的随机事件。例如：掷一枚骰子出现的点数
- 样本空间 (sample space)：一个试验中所有基本事件的集合，用 Ω 表示。例如：在掷枚骰子的试验中，$\Omega = \{1, 2, 3, 4, 5, 6\}$

随机事件及其概率

- 事件 A 的概率是对事件 A 在试验中出现的可能性大小的一种度量
- 表示事件 A 出现可能性大小的数值
- 事件 A 的概率表示为 $P(A)$
- 概率的定义有：古典定义、统计定义和主观概率定义

随机事件及其概率

- 概率的古典定义：如果某一随机试验的结果有限，而且各个结果在每次试验中出现的可能性相同，则事件 A 发生的概率为该事件所包含的基本事件个数 m 与样本空间中所包含的基本事件个数 n 的比值，记为

$$P(A) = \frac{\text{事件 A 所包含的基本事件个数}}{\text{样本空间所包含的基本事件个数}} = \frac{m}{n}$$

随机事件及其概率

- 概率的统计定义：在相同条件下进行 n 次随机试验，事件 A 出现 m 次，则比值 m/n 称为事件 A 发生的频率。随着 n 的增大，该频率围绕某一常数 P 上下摆动，且波动的幅度逐渐减小，取向于稳定，这个频率的稳定值即为事件 A 的概率，记为

$$P(A) = \frac{m}{n} = p$$

随机事件及其概率

- 主观概率的统计定义：对一些无法重复的试验，确定其结果的概率只能根据以往的经验人为确定
- 概率是一个决策者对某事件是否发生，根据个人掌握的信息对该事件发生可能性的判断
- 例如，我认为 2023 年的中国股市是...

离散型随机变量及其分布

- 随机变量：一次试验的结果的数值性描述。
- 一般用 X、Y、Z 来表示
- 根据取值情况的不同分为离散型随机变量和连续型随机变量

离散型随机变量及其分布

- 离散型随机变量

试验	随机变量	可能的取值
抽查100个产品	取到次品的个数	0,1,2, …,100
一家餐馆营业一天	顾客数	0,1,2, …
电脑公司一个月的销售	销售量	0,1, 2,…
销售一辆汽车	顾客性别	男性为0,女性为1

- 连续型随机变量

试验	随机变量	可能的取值
抽查一批电子元件	使用寿命(小时)	$X\geq 0$
新建一座住宅楼	半年后工程完成的百分比	$0\leq X\leq 100$
测量一个产品的长度	测量误差(cm)	$X\geq 0$

离散型随机变量及其分布

- 离散型随机变量的概率分布
- 通常用下面的表格来表示

$X=x_i$	x_1， x_2， …， x_n
$P(X=x_i)=p_i$	p_1， p_2， …， p_n

- $P(X=x_i)=p_i$ 称为离散型随机变量的概率函数

$$p_i \geq 0$$

$$\sum_{i=1}^{n} p_i = 1$$

0—1 分布

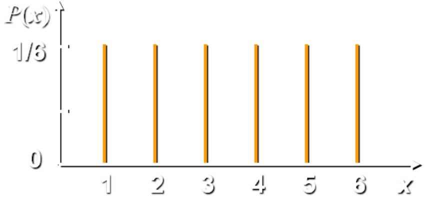
- 一个离散型随机变量 X 只取两个可能的值
- 【例】已知一批产品的次品率为 p = 0.05，合格率为 q=1-p=1-0.05=0.95。并指定废品用 1 表示，合格品用 0 表示。则任取一件为废品或合格品这一离散型随机变量，其概率分布为

$X=x_i$	1	0
$P(X=x_i)=p_i$	0.05	0.95

均匀分布

- 一个离散型随机变量取各个值的概率相同
- 【例】投掷一枚骰子，出现的点数是个离散型随机变量，其概率分布为

$X=x_i$	1	2	3	4	5	6
$P(X=x_i)=p_i$	1/6	1/6	1/6	1/6	1/6	1/6



离散型随机变量的期望值 expected value

- 在离散型随机变量 X 的一切可能取值的完备组中，各可能取值 x_i 与其取相对应的概率 p_i 乘积之和
- 描述离散型随机变量取值的集中程度

$$E(X)=\sum_{i=1}^{n} x_i p_i$$

$$E(X)=\sum_{i=1}^{\infty} x_i p_i$$

离散型随机变量的方差 variance

- 随机变量 X 的每一个取值与期望值的离差平方和的数学期望，记为 $D(X)$
- 描述离散型随机变量取值的分散程度

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

$$D(X) = \sum_{i=1}^{\infty} [x_i - E(X)]^2 \cdot p_i$$

离散型随机变量的方差

- 【例】投掷一枚骰子，出现的点数是个离散型随机变量，其概率分布为如下。计算数学期望和方差。

$X=x_i$	1	2	3	4	5	6
$P(X=x_i)=p_i$	1/6	1/6	1/6	1/6	1/6	1/6

- 解：期望和方差分别为：

$$E(X)=\sum_{i=1}^{6}x_ip_i=1\times\frac{1}{6}+\cdots+6\times\frac{1}{6}=3.5$$

$$D(X)=\sum_{i=1}^{6}[x_i-E(X)]^2\cdot p_i$$
$$=(1-3.5)^2\times\frac{1}{6}+\cdots+(6-3.5)^2\times\frac{1}{6}=2.9167$$

离散型随机变量的方差

- 【例】投掷一枚骰子，出现的点数是个离散型随机变量，其概率分布为如下。计算数学期望和方差。

$X=x_i$	1	2	3	4	5	6
$P(X=x_i)=p_i$	1/6	1/6	1/6	1/6	1/6	1/6

- 解：期望和方差分别为：

$$E(X)=\sum_{i=1}^{6}x_ip_i=1\times\frac{1}{6}+\cdots+6\times\frac{1}{6}=3.5$$

$$D(X)=\sum_{i=1}^{6}[x_i-E(X)]^2\cdot p_i$$

$$=(1-3.5)^2\times\frac{1}{6}+\cdots+(6-3.5)^2\times\frac{1}{6}=2.9167$$

二项分布 Binomial distribution

- 二项分布与贝努里试验有关
- 贝努里试验具有如下属性
  - 试验包含了 n 个相同的试验
  - 每次试验只有两个可能的结果，即“成功”和“失败”
  - 出现“成功”的概率 p 对每次试验结果是相同的；“失败”的概率 q 也相同，且 $p+q=1$
  - 试验是相互独立的
  - 试验“成功”或“失败”可以计数

二项分布

- 进行 n 次重复试验，出现“成功”的次数的概率分布称为二项分布
- 设 X 为 n 次重复试验中事件 A 出现的次数，X 取 x 的概率为

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

$$C_n^x=\frac{n!}{x!(n-x)!}$$

二项分布

- 显然，对于 $P\{X=x\} \geq 0 x=1,2,\cdots,n$，有

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

- 当 $n=1$ 时，二项分布化简为

$$P\{X=x\} = p^x q^{1-x} 其中 x=0,1$$

二项分布

- 二项分布的数学期望为

$$E(X) = np$$

- 方差为

$$D(X)=npq$$

二项分布

- 【例】已知 100 件产品中有 5 件次品，现从中任取一件，有放回地抽取 3 次。求在所抽取的 3 件产品中恰好有 2 件次品的概率
- 解：设 X 为所抽取的 3 件产品中的次品数，则 $X \sim B(3, 0.05)$，根据二项分布公式有

$$P\{X = 2\} = C_3^2 (0.05)^2 (0.95)^{3-2} = 0.007125$$

二项分布

- 【例】已知 100 件产品中有 5 件次品，现从中任取一件，有放回地抽取 3 次。求在所抽取的 3 件产品中恰好有 2 件次品的概率
- 解：设 X 为所抽取的 3 件产品中的次品数，则 $X \sim B(3, 0.05)$，根据二项分布公式有

$$P\{X = 2\} = C_3^2(0.05)^2(0.95)^{3-2} = 0.007125$$

泊松分布 Poisson distribution

- 用于描述在一指定时间范围内或在一定的长度、面积、体积之内每一事件出现次数的分布
- 一个城市在一个月内发生的交通事故次数
- 消费者协会一个星期内收到的消费者投诉次数

泊松分布

- 泊松概率分布函数

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

- λ 给定的时间间隔、长度、面积、体积内“成功”的平均数
- $e=2.71828$
- x 给定的时间间隔、长度、面积、体积内“成功”的次数

泊松分布

- 泊松分布的数学期望为

$$E(X) = \lambda$$

- 方差为

$$D(X)=\lambda$$

泊松分布

- 【例】假定某企业的职工中在周一请假的人数 X 服从泊松分布，且设周一请事假的平均人数为 2.5 人。求（1）X 的均值及标准差（2）在给定的某周一正好请事假是 5 人的概率
- 解：

$$E(X) = \lambda = 2.5$$

$$\sqrt{D(X)} = \sqrt{2.5} = 1.581$$

$$P\{X = 5\} = \frac{(2.5)^5 e^{-2.5}}{5!} = 0.067$$

泊松分布

- 【例】假定某企业的职工中在周一请假的人数 X 服从泊松分布，且设周一请事假的平均人数为 2.5 人。求（1）X 的均值及标准差（2）在给定的某周一正好请事假是 5 人的概率
- 解：

$$E(X) = \lambda = 2.5$$

$$\sqrt{D(X)} = \sqrt{2.5} = 1.581$$

$$P\{X = 5\} = \frac{(2.5)^5 \mathrm{e}^{-2.5}}{5!} = 0.067$$

泊松分布

- 当试验的次数 n 很大，成功的概率 p 很小时，可用泊松分布来近似地计算二项分布的概率，即

$$C_n^x p^x q^{n-x} \approx \frac{\lambda \mathrm{e}^{-\lambda}}{x!}$$

- 实际应用中，当 $P \leq 0.25$，$n > 20$，$np \leq 5$ 时，近似效果良好

1. 随机事件及其概率
2. 离散型随机变量及其分布
3. 连续型随机变量的概率分布

连续型随机变量的概率分布

- 连续型随机变量可以取某一区间或整个实数轴上的任意一个值
- 它取任何一个特定的值的概率都等于 0
- 不能列出每一个值及其相应的概率
- 通常研究它取某一区间值的概率
- 用数学函数的形式和分布函数的形式来描述

连续型随机变量的概率分布

- 设 X 为一连续型随机变量，x 为任意实数，X 的概率密度函数记为 $f(x)$，它满足条件

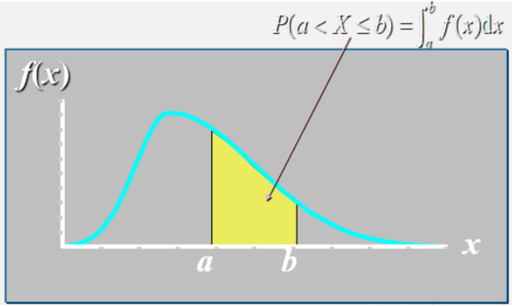
$$f(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$$

- $f(x)$ 不是概率

概率密度函数

- 在平面直角坐标系中画出 $f(x)$ 的图形，则对于任何实数 $x_1 < x_2$，$P(x_1 < X \leq x_2)$ 是该曲线下从 x_1 到 x_2 的面积



分布函数

- 连续型随机变量的概率也可以用分布函数 $F(x)$ 来表示
- 分布函数定义为

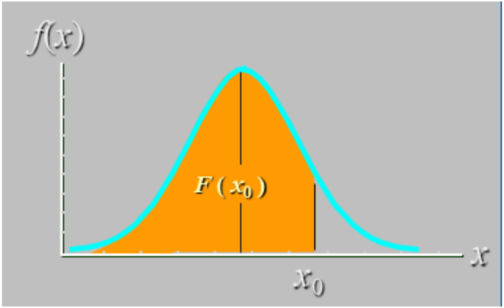
$$F(x)=P(X\leq x)=\int_{-\infty}^{x}f(t)\mathrm{d}t(-\infty<x<+\infty)$$

- 根据分布函数,$P(a<X<b)$ 可以写成

$$P(a<X<b)=\int_{a}^{b}f(x)\mathrm{d}x=F(b)-F(a)$$

连续型随机变量的概率分布

- $f(x)$ 密度函数曲线下的面积等于 1
- 分布函数是曲线下小于 x_0 的面积



连续型随机变量的期望与方差

- 连续型随机变量的数学期望为

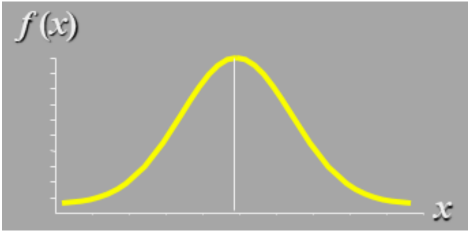
$$E(X) = \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x = \mu$$

- 方差为

$$D(X) = \int_{-\infty}^{+\infty} [x - E(X)] f(x) \mathrm{d}x = \sigma^2$$

正态分布 normal distribution

- 描述连续型随机变量的最重要的分布
- 可用于近似离散型随机变量的分布。例如：二项分布
- 经典统计推断的基础



正态分布

- 正态分布的概率密度函数

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

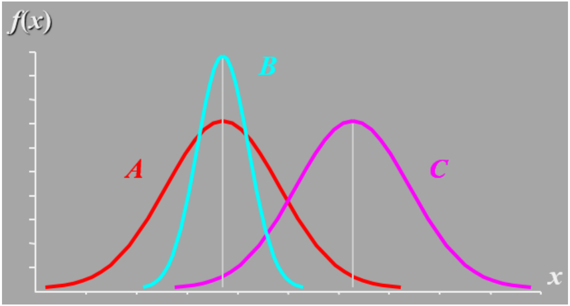
- $f(x)$ = 随机变量 X 的频数
  σ^2= 总体方差
- π=3.14159; e = 2.71828
  x = 随机变量的取值 $(-\infty < x < +\infty)$
  μ= 总体均值

正态分布

- 概率密度函数在 x 的上方，即 $f(x) > 0$
- 正态曲线的最高点在均值 μ，它也是分布的中位数和众数
- 正态分布是一个分布族，每一特定正态分布通过均值 μ 和标准差 σ 来区分。μ 决定了图形的中心位置,σ 决定曲线的平缓程度，即宽度
- 曲线 $f(x)$ 相对于均值 μ 对称，尾端向两个方向无限延伸，且理论上永远不会与横轴相交
- 正态曲线下的总面积等于 1
- 随机变量的概率由曲线下的面积给出

正态分布

- 均值 μ 和标准差 σ 对正态分布曲线的影响
- 正态分布的高度（平均值）与坡度（标准差）



正态分布

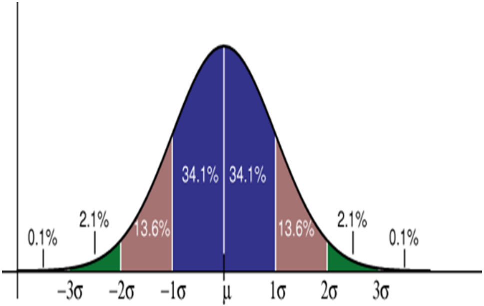
- 正态分布是正常分布，自然分布

万有不齐天地事

一无可寄古今情

正态分布

- 正态分布是一种世界观



正态分布

- 人的绝大多数属性的分布是正态分布
  - 身高、体重、颜值、脑力、体力，等等
  - 勤奋、灵活、创意、远见、格局，等等
  - 发现发扬正态分布对自己有利的属性
  - 躲避弥补正态分布对自己不利的属性

标准正态分布

- 一般的正态分布取决于均值 μ 和标准差 σ
- 计算概率时，每一个正态分布都需要有自己的正态概率分布表，这种表格是无穷多的
- 若能将一般的正态分布转化为标准正态分布，计算概率时只需要查一张表

标准正态分布

- 任何一个一般的正态分布，可通过下面的线性变换转化为标准正态分布

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

- 标准正态分布的概率密度函数

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x}{2}^2} \quad , \qquad -\infty < x < +\infty$$

- 标准正态分布的分布函数

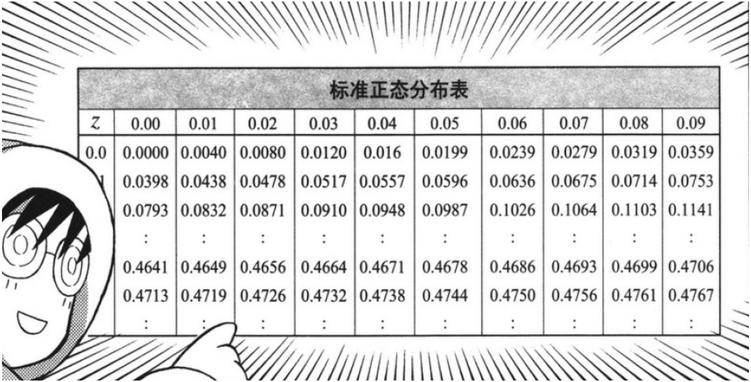
$$\Phi(x) = \int_{-\infty}^{x} \phi(t)\mathrm{d}t = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t$$

标准正态分布表

- 将一个一般的转换为标准正态分布
- 计算概率时，查标准正态概率分布表
- 对于负的 x ，可由 $\Phi(-x) = 1 - \Phi(x)$ 得到
- 对于标准正态分布，即 $X \sim N(0, 1)$，有 $P(a \leq X \leq b) = \Phi(b) - \Phi(a)$；$P(|X| \leq a) = 2\Phi(a) - 1$
- 对于一般正态分布，即 $X \sim N(\mu, \sigma)$，有

$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

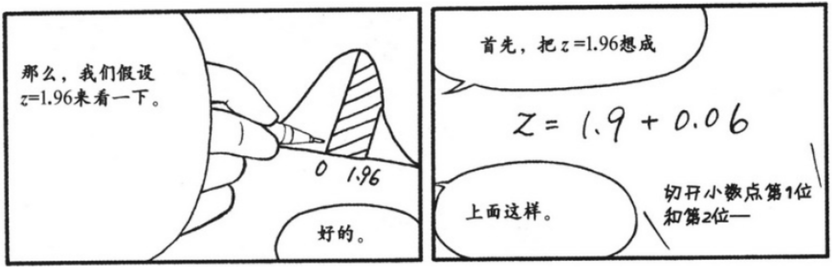
标准正态分布表



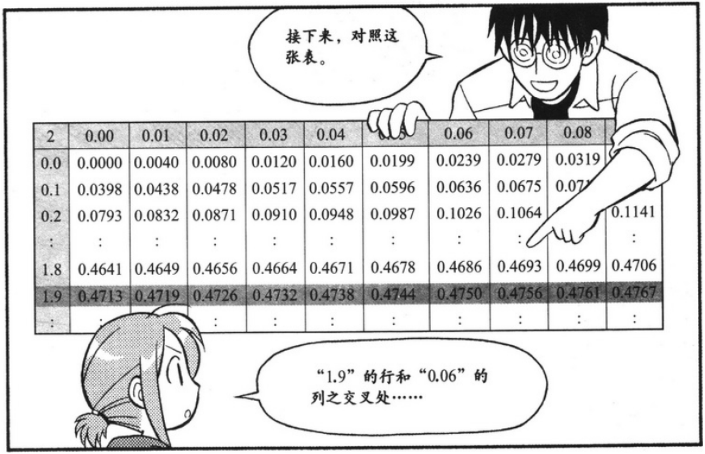
标准正态分布表

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.016	0.0199	0.0239	0.0279	0.0319	0.0359
1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

标准正态分布表


那么，我们假设
z=1.96来看一下。
0
1.96
好的。
首先，把z=1.96想成
Z = 1.9 + 0.06
上面这样。
切开小数点第1位
和第2位—

标准正态分布表


接下来，对照这张表。
2 0.00 0.01 0.02 0.03 0.04 0.06 0.07 0.08
0.0 0.0000 0.0040 0.0080 0.0120 0.0160 0.0199 0.0239 0.0279 0.0319
0.1 0.0398 0.0438 0.0478 0.0517 0.0557 0.0596 0.0636 0.0675
0.2 0.0793 0.0832 0.0871 0.0910 0.0948 0.0987 0.1026 0.1064 0.1141
1.8 0.4641 0.4649 0.4656 0.4664 0.4671 0.4678 0.4686 0.4693 0.4699 0.4706
1.9 0.4713 0.4719 0.4726 0.4732 0.4738 0.4744 0.4750 0.4756 0.4761 0.4767
“1.9”的行和“0.06”的列之交叉处……

标准正态分布表


是"0.4750"！
0.4750
0
1.96
是的！
这就是z=1.96时的面积。
啊！
差点忘了告诉你，所有的标准正态分布之机率密度函数的图形和横轴所围成的面积都是1哦！
面积=1
哦！

正态分布的前世今生

- 内容参见《统计学 A》-outline.nb

补充内容

- 内容参见 Bilibili《计量经济学》–概率论与统计相关知识

参考资料

- 贾俊平.《统计学》(第八版)[M]. 北京:中国人民大学出版社,2021。
- 向蓉美等. 统计学(第 2 版)[M]. 北京:机械工业出版社,2017。

Q&A

THANK YOU