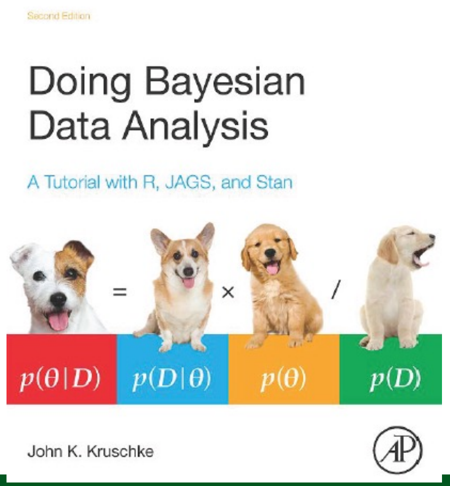
统计决策与贝叶斯分析

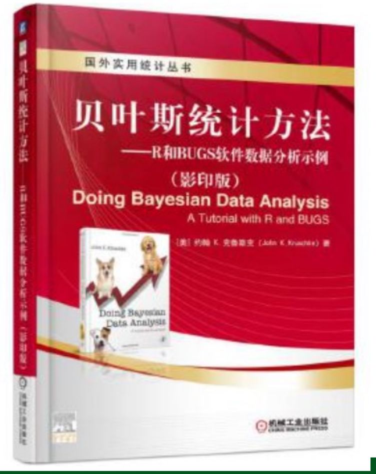
中山大学人工智能学院
毛旭东

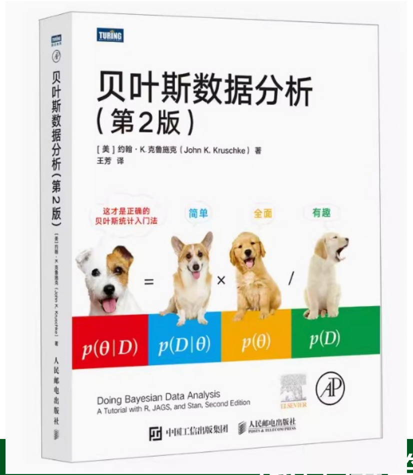
Email：maoxd3@mail.sysu.edu.cn

课程教材信息

- 《Doing Bayesian Data Analysis》，第二版，作者 John K. Kruschke。
  - 课程群有电子版可下载。
- “The puppy book”
- 2024年1月刚出版翻译版。







参考书

- **《统计反思》**，作者理查德麦克尔里思，机械工业出版社。
- **《贝叶斯统计方法（影印版）》**，作者约翰 K. 克鲁斯克，机械工业出版社。
- **《贝叶斯数据分析》**，作者吴喜之，中国人民大学出版社。
- **《贝叶斯数据分析（英文导读版 原书第3版）》**，作者安德鲁格尔曼等，机械工业出版社。

成绩考核和学时

- 总学时：36学时
- 成绩评定：
  - 平时成绩占40%，其中作业占30%，出勤情况占10%。
  - 期末考试占60%。

- 关于编程
  - 课堂会讲授用PyMC库的python代码
  - 平时作业会有编程题，来实践贝叶斯分析
  - 期末考试没有编程题

教学团队

- 教师：毛旭东
  - maoxd3@mail.sysu.edu.cn

- 助教：周斐毓、董晓宇、于云聪

- 课程QQ群：扫码右图。
  - 课程信息在群内发布。
  - 密码: bayes2024

为什么学习统计决策与贝叶斯分析?

- “机器学习”时间轴：

Year	Event type	Caption	Event
1763	Discovery	The Underpinnings of Bayes' Theorem	Thomas Bayes's work *An Essay towards solving a Problem in the Doctrine of Chances* is published two years after his death, having been amended and edited by a friend of Bayes, Richard Price.[8] The essay presents work which underpins Bayes theorem.
1805	Discovery	Least Square	Adrien-Marie Legendre describes the "méthode des moindres carrés", known in English as the least squares method.[9] The least squares method is used widely in data fitting.
1812		Bayes' Theorem	Pierre-Simon Laplace publishes *Théorie Analytique des Probabilités*, in which he expands upon the work of Bayes and defines what is now known as Bayes' Theorem.[10]
1913	Discovery	Markov Chains	Andrey Markov first describes techniques he used to analyse a poem. The techniques later become known as Markov chains.[11]
1943	Discovery	Artificial Neuron	Warren McCulloch and Walter Pitts develop a mathematical model that imitates the functioning of a biological neuron, the artificial neuron which is considered to be the first neural model invented.[12]

为什么学习统计决策与贝叶斯分析?

“*Machine learning is statistics + linear algebra.*”

- 机器学习从统计学发展而来。

什么是统计（Statistics）？

- 概率论是从模型映射到数据的过程，比如：
  - 给定一个正常的骰子（模型），得到1点（数据）的概率是$\frac{1}{6}$；
  - 给定一个高斯分布$X\sim N(\mu,\sigma^2)$（模型），x取值（数据）在$[u-\sigma,u+\sigma]$的概率约为68.27%。

- 统计是一个相反的过程，**从“数据”估计“模型参数”**。
  - 比如：给定一个骰子的多次结果，估计这个骰子每个点数的概率。

- 统计在早期被称为逆概率论（inverse probability theory）。

什么是贝叶斯分析？ （Bayesian Analysis）

- 贝叶斯分析是统计学的一种分析方法，一种基于贝叶斯法则的方法。

- 统计学主要有两个学派：

1. 贝叶斯学派
2. 频率学派

课程大纲

1. 贝叶斯分析简介和基本概念
2. 概率论回顾
3. 极大似然估计与贝叶斯估计
4. 贝叶斯推理方法：准确数学分析
5. 贝叶斯推理方法：MCMC近似
6. 广义线性模型：回归
7. 广义线性模型：分类
8. 层级模型

- 作业的预计时间点：
  1. 第3章讲完
  2. 第5章讲完
  3. 第7章讲完
- 每次作业包含3-4题简答题、1-2题编程题。

课程目标

- 掌握贝叶斯分析的理论基础；
- 掌握常用的贝叶斯分析方法，包括网格近似方法、准确数学分析方法、以及MCMC近似方法；
- 掌握使用贝叶斯方法分析广义线性模型，包括线性回归、逻辑回归等；
- 掌握极大似然估计，了解极大后验估计，以及他们与贝叶斯估计的区别；
- 了解层级模型；
- 了解运用PyMC等Python概率编程库来解决实际问题。

第1章 贝叶斯分析简介和基本概念

贝叶斯推理---引例

- 有一天早上，我们在路上走，发现**人行道是湿的**。。。
- 可能的原因包括：
  - 刚下过雨
  - 周围的草坪刚浇过水
  - 地下水喷出来
  - 污水管破了
  - 行人的饮料洒了
  - 。。。
- 对于以上可能的原因，我们根据以前的知识，都有一个***先验可信度(prior credibility)***，比如我们觉得“刚下过雨”是最有可能的。

2.1 贝叶斯推理是重新分配可信度的过程

- 我们接着行走。。。
- 如果发现路边的树和汽车也是湿的，
  - 我们重新分配可信度（credibility），更确信“刚下过雨”。
- 如果发现仅仅只是一小块地是湿的，并且旁边有个空的饮料杯，
  - 我们重新分配可信度（credibility），相信是“饮料洒了”。

- 这种根据观察（数据）重新分配可信度的过程，就是贝叶斯推理的核心思想。

引例2：福尔摩斯的推理

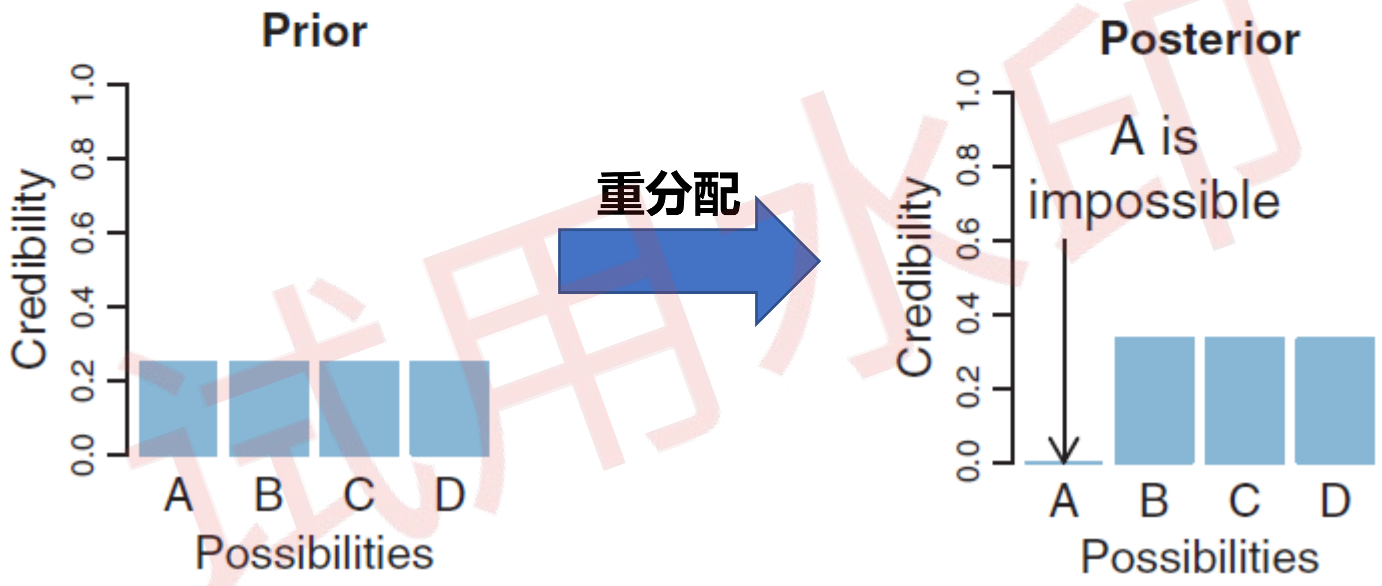
- “How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?”

  ---(Doyle, 1890, chap. 6)

- 福尔摩斯通过观察，收集证据，来排除一些不可能的情况。

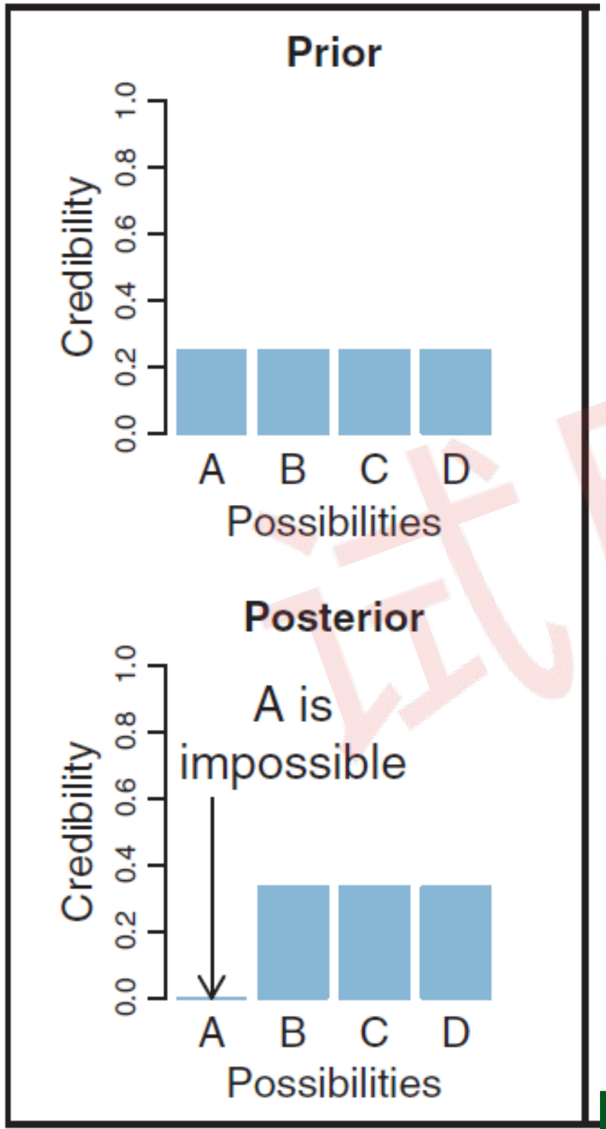
福尔摩斯的推理

- 福尔摩斯通过观察，收集证据，来排除一些不可能的情况。



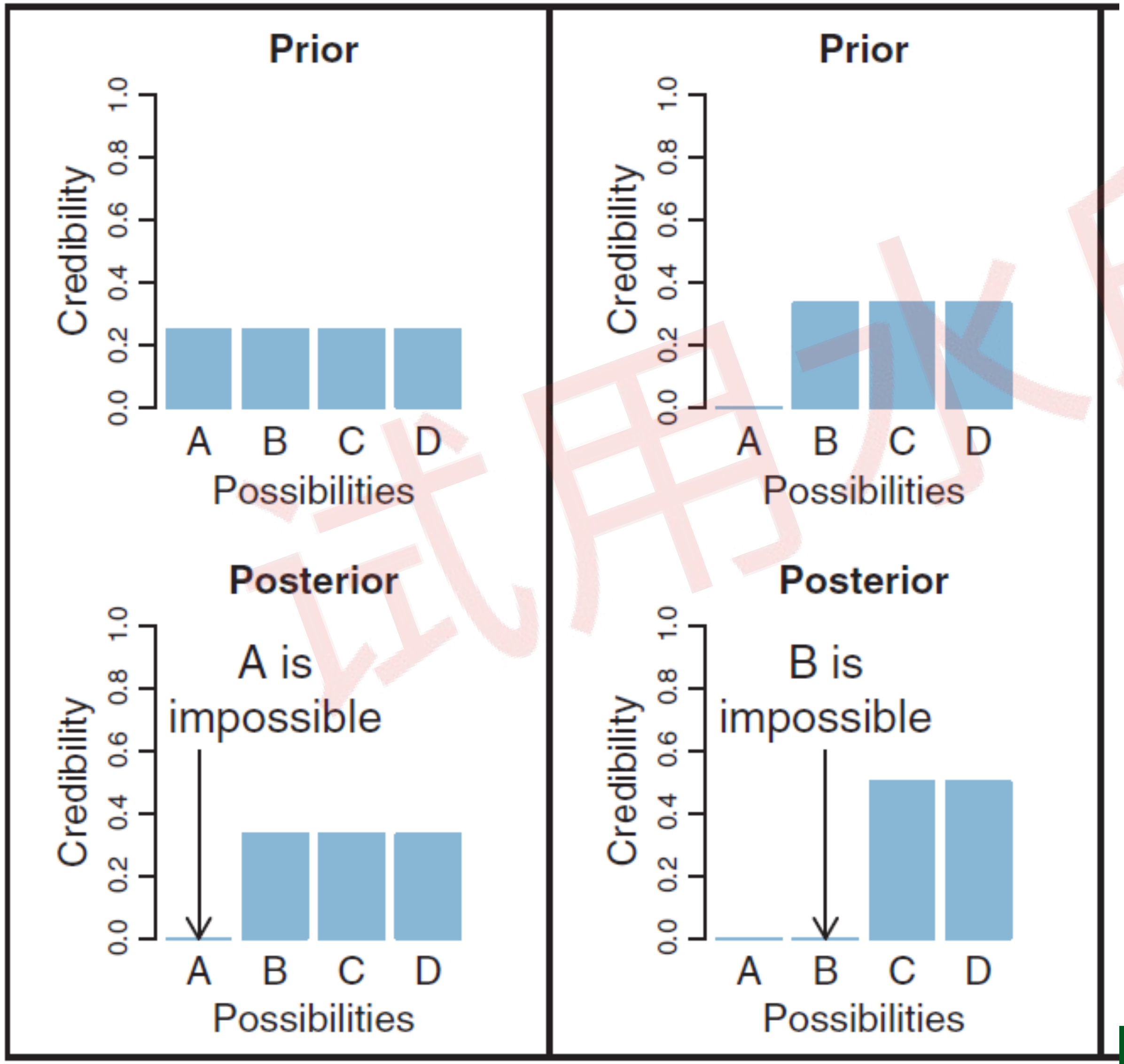
- 左图称为***先验概率分布（prior distribution）***。
- 重分配后的右图称为***后验概率分布（posterior distribution）***。

福尔摩斯的推理

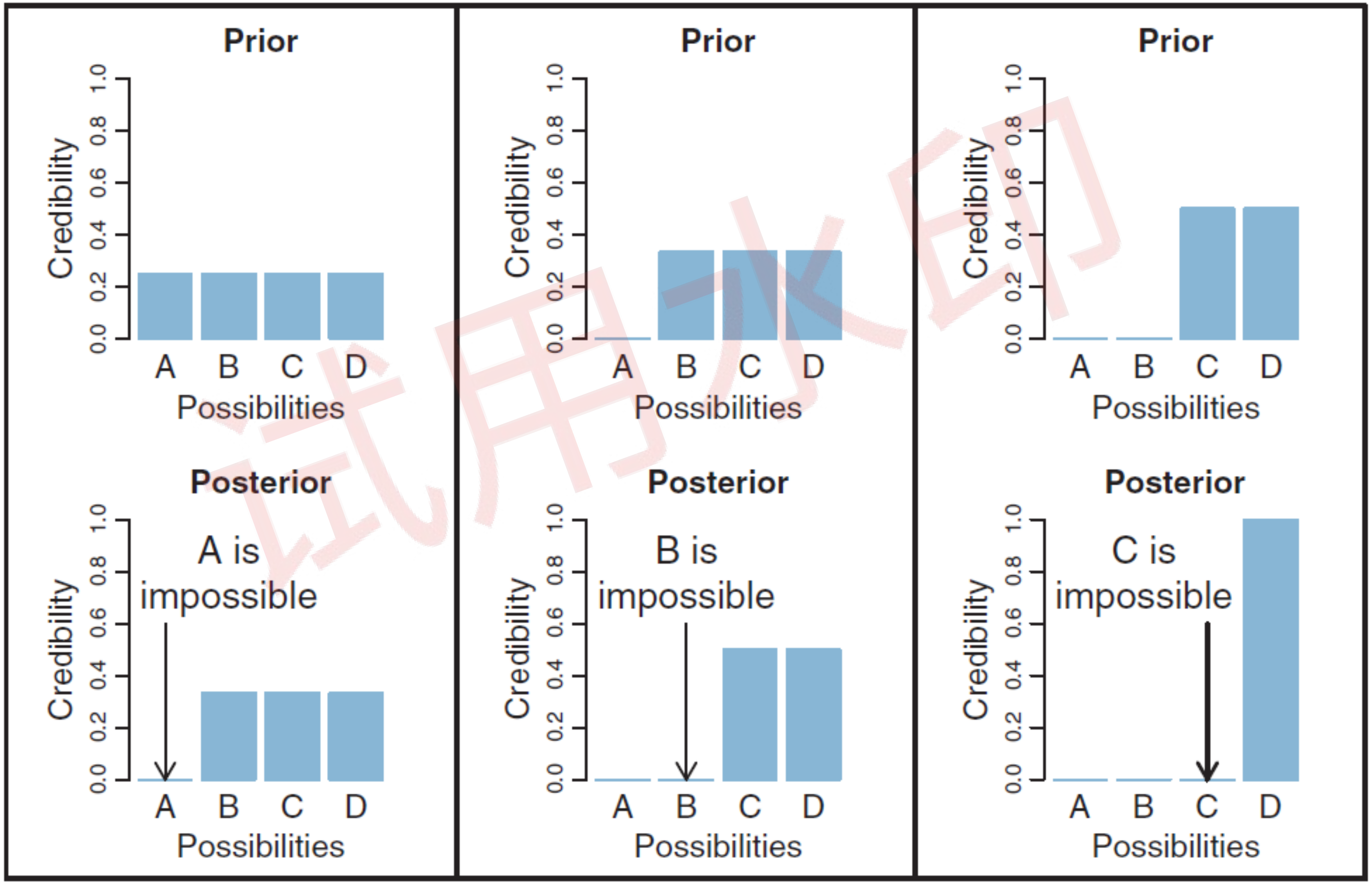


Prior
Credibility
0.0
0.2
0.4
0.6
0.8
1.0
A
B
C
D
Possibilities
Posterior
Credibility
A is
impossible
0.0
0.2
0.4
0.6
0.8
1.0
A
B
C
D
Possibilities

福尔摩斯的推理


Prior
Credibility
0.0 0.2 0.4 0.6 0.8 1.0
A B C D
Possibilities
Prior
Credibility
0.0 0.2 0.4 0.6 0.8 1.0
A B C D
Possibilities
Posterior
Credibility
0.0 0.2 0.4 0.6 0.8 1.0
A is impossible
A B C D
Possibilities
Posterior
Credibility
0.0 0.2 0.4 0.6 0.8 1.0
B is impossible
A B C D
Possibilities

福尔摩斯的推理



Prior
Credibility
0.0 0.2 0.4 0.6 0.8 1.0
A B C D
Possibilities
Prior
Credibility
0.0 0.2 0.4 0.6 0.8 1.0
A B C D
Possibilities
Prior
Credibility
0.0 0.2 0.4 0.6 0.8 1.0
A B C D
Possibilities
Posterior
Credibility
0.0 0.2 0.4 0.6 0.8 1.0
A is impossible
A B C D
Possibilities
Posterior
Credibility
0.0 0.2 0.4 0.6 0.8 1.0
B is impossible
A B C D
Possibilities
Posterior
Credibility
0.0 0.2 0.4 0.6 0.8 1.0
C is impossible
A B C D
Possibilities

2.1.1 数据是有噪声的

- 福尔摩斯发现了一个脚印，然后推测出了鞋子的大小。
- 然而，在实际中，测量脚印大小并不是准确的，只能推测出鞋子大小的范围。

数据的测量是充满随机性的

- 比如，我们要测试一种新药能否降低血压，
  - 有一组是实验组，吃新药；
  - 另一组是控制组，吃安慰剂。
- 我们每隔一段时间测试一次血压，测出的血压有很多不确定因素：
  - 人的血压受运动、压力、吃的食物等的影响；
  - 血压测试方法本身并不完全准确，有一定的浮动值；
  - 不同人的血压不同。
- 最终，每一组中的数据，波动都是很大的。两组之间还会有很大的重叠。

推理是有不确定性的

- 我们拿到的数据，都是有一定的“噪声”的。

- 贝叶斯分析是一种从噪声数据中推理可能性的方法。

例子：推理是有不确定性的

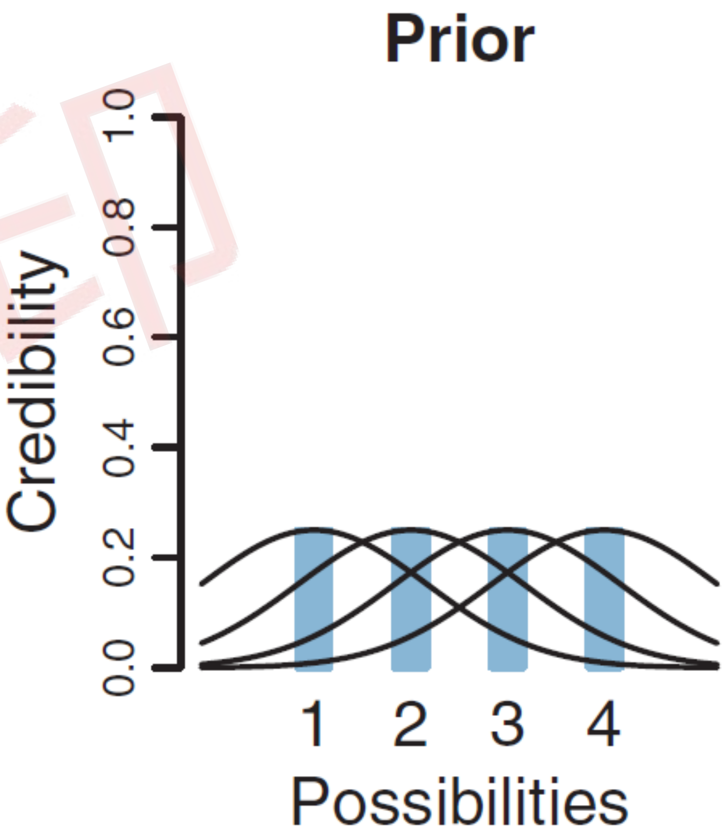
- 假设有一个制造弹力球的工厂，生产4种大小的球，分别是1，2，3，4。然而，制造过程中会有误差，比如生产3的球，实际可能是1.8或者4.2，但是均值还是3。

- 假设工厂生产了3个某一种大小的球。我们拿到的3个球的大小分别是1.77，2.23，2.70。

- 根据拿到的3个球的大小，我们能推理出工厂生产的是大小是2的球吗？大小是1,3或者4，有可能吗？

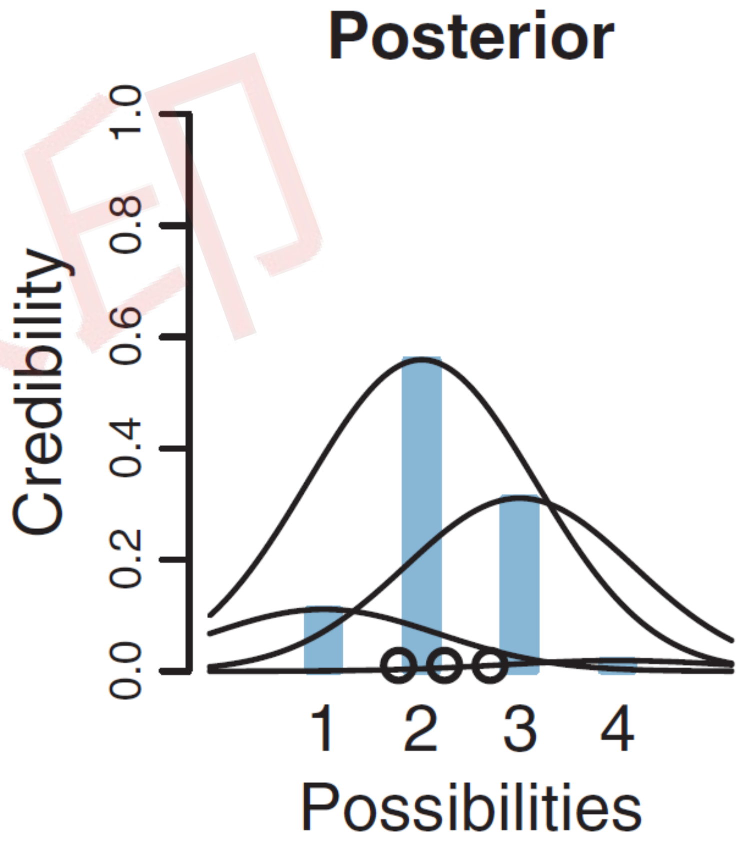
先验概率分布

- 假设工厂生产不同大小的概率是一样的，都是0.25。
  - P(大小=1)=0.25
  - P(大小=2)=0.25
  - P(大小=3)=0.25
  - P(大小=4)=0.25
- 由于生产会有误差，假设生产的球大小是以1,2,3,4为中心的“钟形“概率分布。



后验概率分布

- 根据得到的数据1.77，2.23，2.7，来重分配不同大小的概率值。
- 根据贝叶斯推理，最后可得：
  - P(大小=2|D)=0.56
  - P(大小=3|D)=0.31
  - P(大小=1|D)=0.11
  - P(大小=4|D)=0.02



正态分布（Normal Distribution）

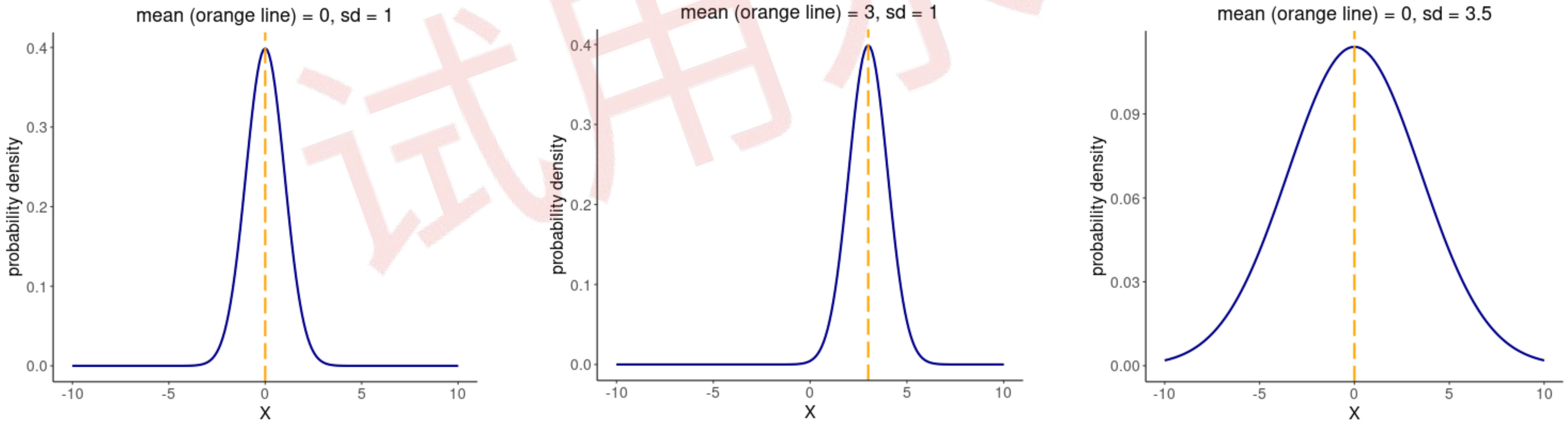
- 概率密度函数为：

$$p(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

- 参数μ是均值，被称为位置参数。
- 参数σ是标准差，被称为尺度参数。

正态分布

- 形状是钟形（bell-shaped）。
- 均值被称为***位置参数*** *(location parameter)*。
- 标准差被称为***尺度参数*** *(scale parameter)*。

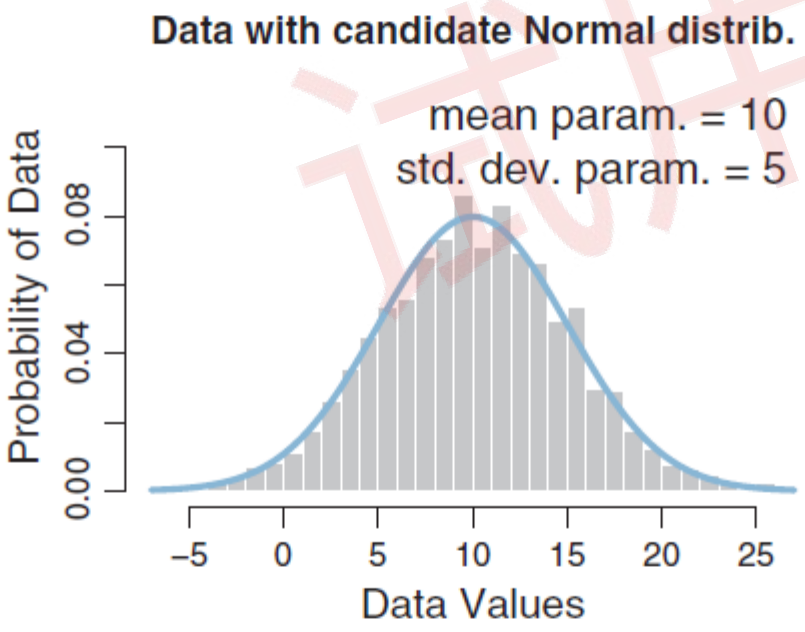


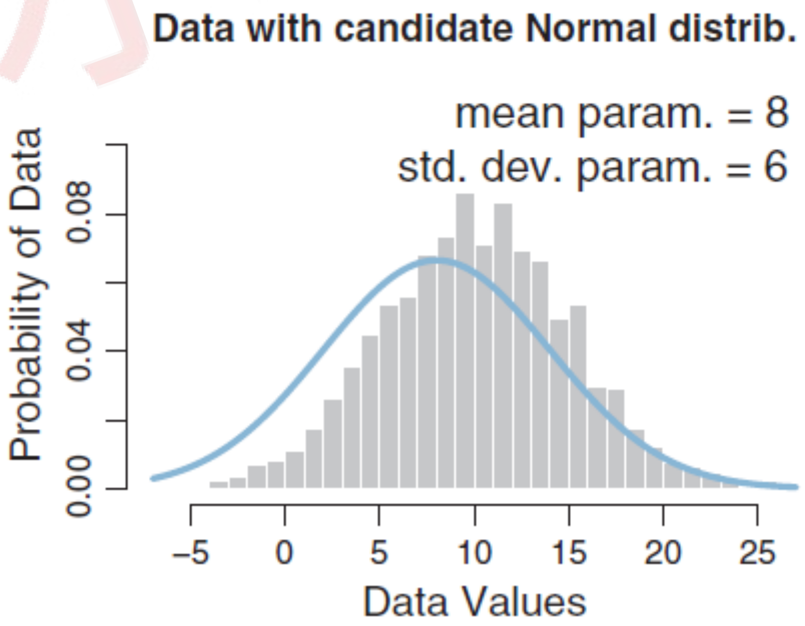
2.2 模型和参数

- 数据分析的第一步：找一个能够描述数据的***模型***，模型由数学公式来表达。
- 模型的公式中，往往有一些可变的***参数（parameters）***，这些参数可以用来控制模型的形式，来描述不同的数据。
- 比如正态分布（normal distribution），又被称为高斯分布（Gaussian distribution），有2个参数：均值（mean）和标准差（standard deviation）。

模型选择

- 下图为用2个不同参数的正态分布来描述数据。
- 贝叶斯推理要做的：给定数据，对于每一个参数，计算不同值的可能性。

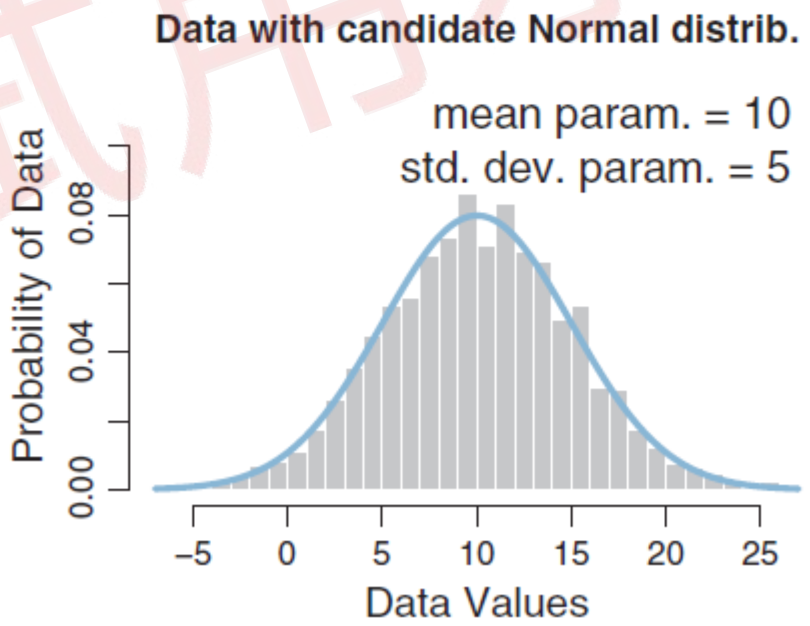




模型选择

- 给定一些数据，一个优秀的描述该数据的模型，需具备两个条件：

1. 足够强的描述能力，也就是“看起来和数据很像”。
2. 有意义的参数。对于参数不仅需要和数据非常的贴合，还需要有意义能够进行解释



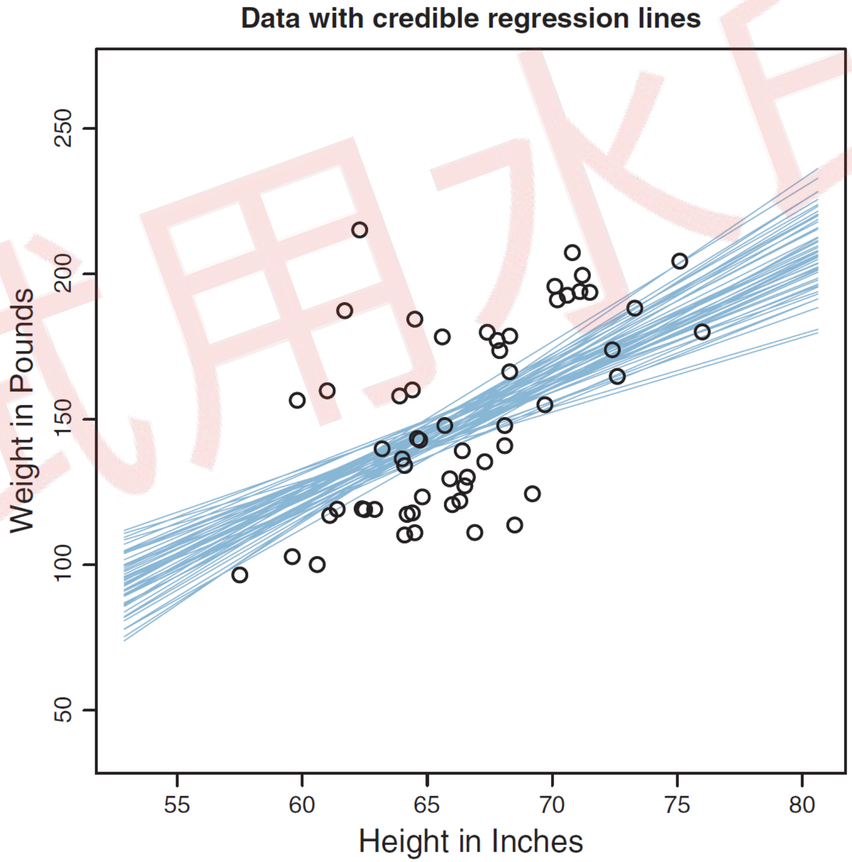
2.3 贝叶斯分析的步骤

中山大學 SUN YAT-SEN UNIVERSITY

1. 确定和问题相关的数据。
2. 确定适合数据的模型和相应的参数。
3. 给要估计的参数指定一个先验概率分布。
4. 根据数据，使用贝叶斯推理来重分配参数的概率分布，得到参数的后验概率分布。
5. 检验后验概率分布能够准确地描述数据。如果不行，考虑换一个模型。

例子：身高预测体重

1.确定和问题相关的数据。



例子：身高预测体重

2. 确定适合数据的模型和相应的参数。

- 从数据看，体重和身高似乎成正比，假设线性关系：

$$\hat{y} = \beta_1 x + \beta_0$$

- 加上随机变化：

$$y \sim \text{normal}(\hat{y}, \sigma)$$

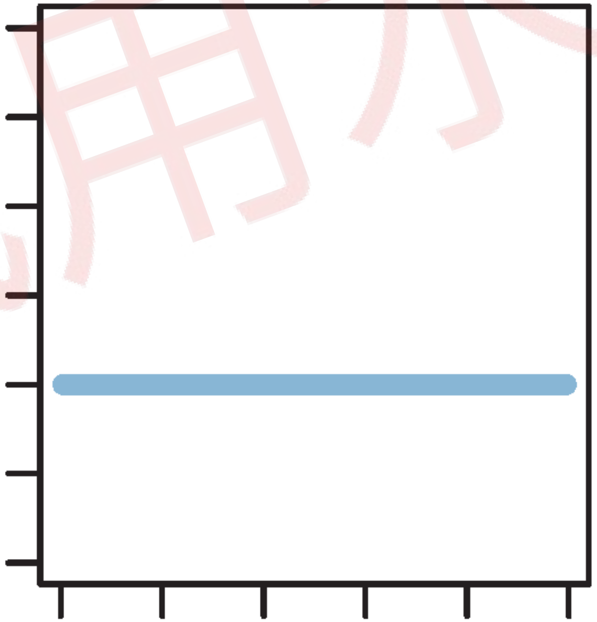
- 模型共有三个参数：

1. β_1
2. β_0
3. σ

例子：身高预测体重

3. 给要估计的参数指定一个先验概率分布。

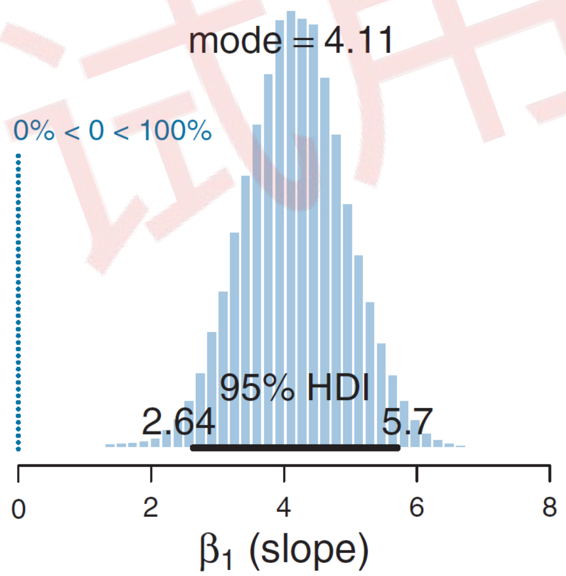
- 假设我们对三个参数都没有先验知识，我们采用“不明确的先验”，也就是所有取值都是一样的概率。



例子：身高预测体重

4. 根据数据，使用贝叶斯推理来重分配参数的概率分布，得到参数的后验概率分布。

- 这里，我们先略过推理过程。最终得到如下图的后验概率分布：

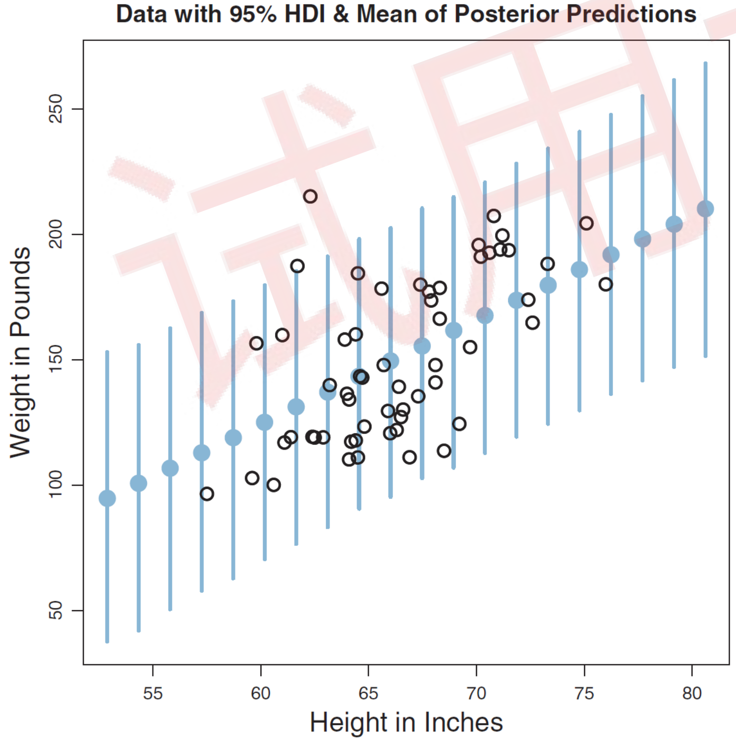


- 众数（mode）是4.11。
- 95%最高密度区间（Highest Density Interval，HDI)是[2.64, 5.7]。

例子：身高预测体重

5. 检验后验概率分布能够准确地描述数据。如果不行，考虑换一个模型。

- 有很多方法，其中一种是画出模型和实际数据。



- 蓝色点表示预测的均值。
- 蓝色线条代表95% HDI。

总结

- 统计
- 先验概率分布
- 后验概率分布 根据数据来重新分配参数的概率
- 贝叶斯分析
- 贝叶斯分析的5个步骤
  1. 数据
  2. 选择合适的模型和参数
  3. 设置先验
  4. 根据数据更新后验
  5. 判断描述数据的能力