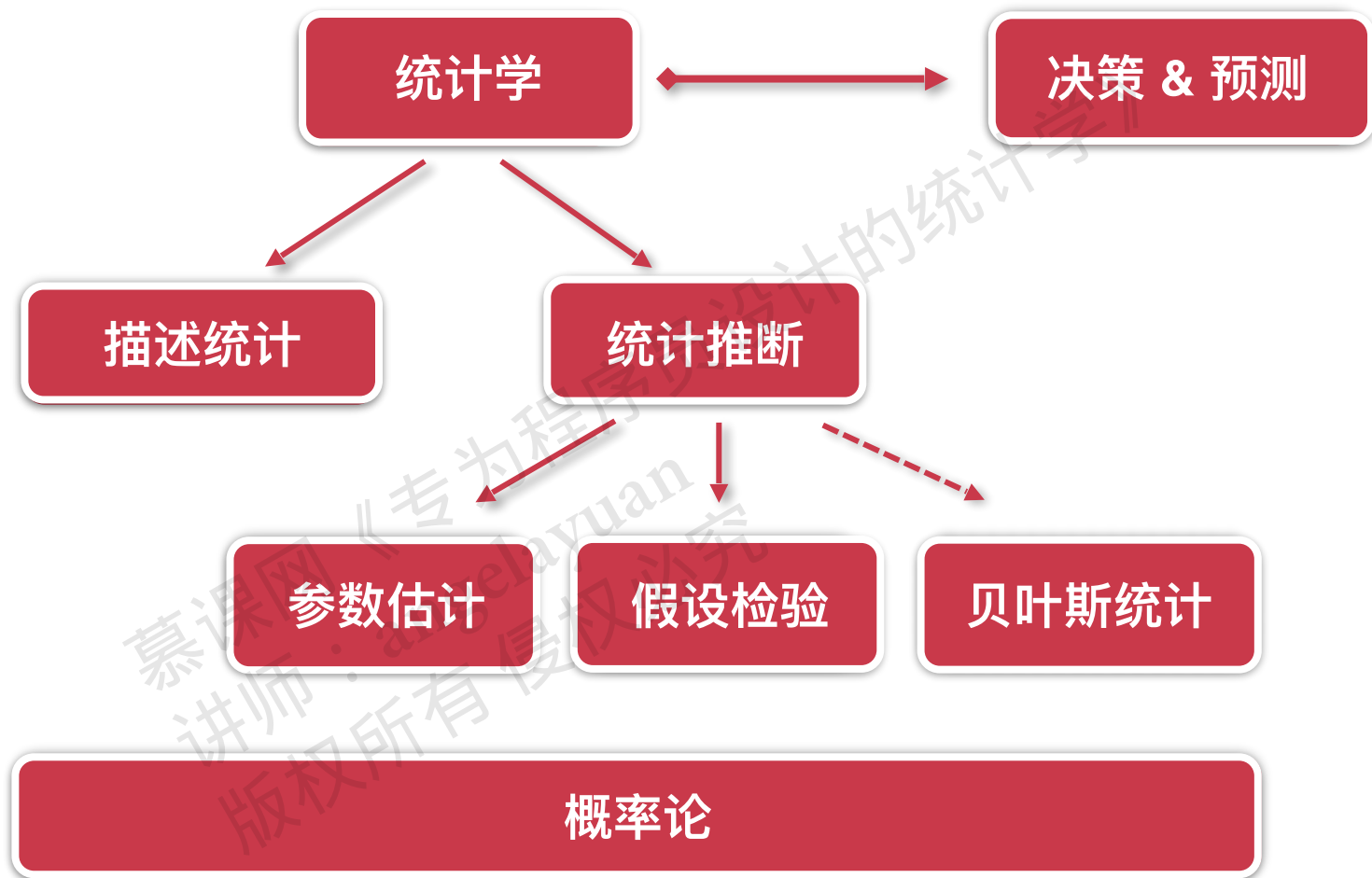


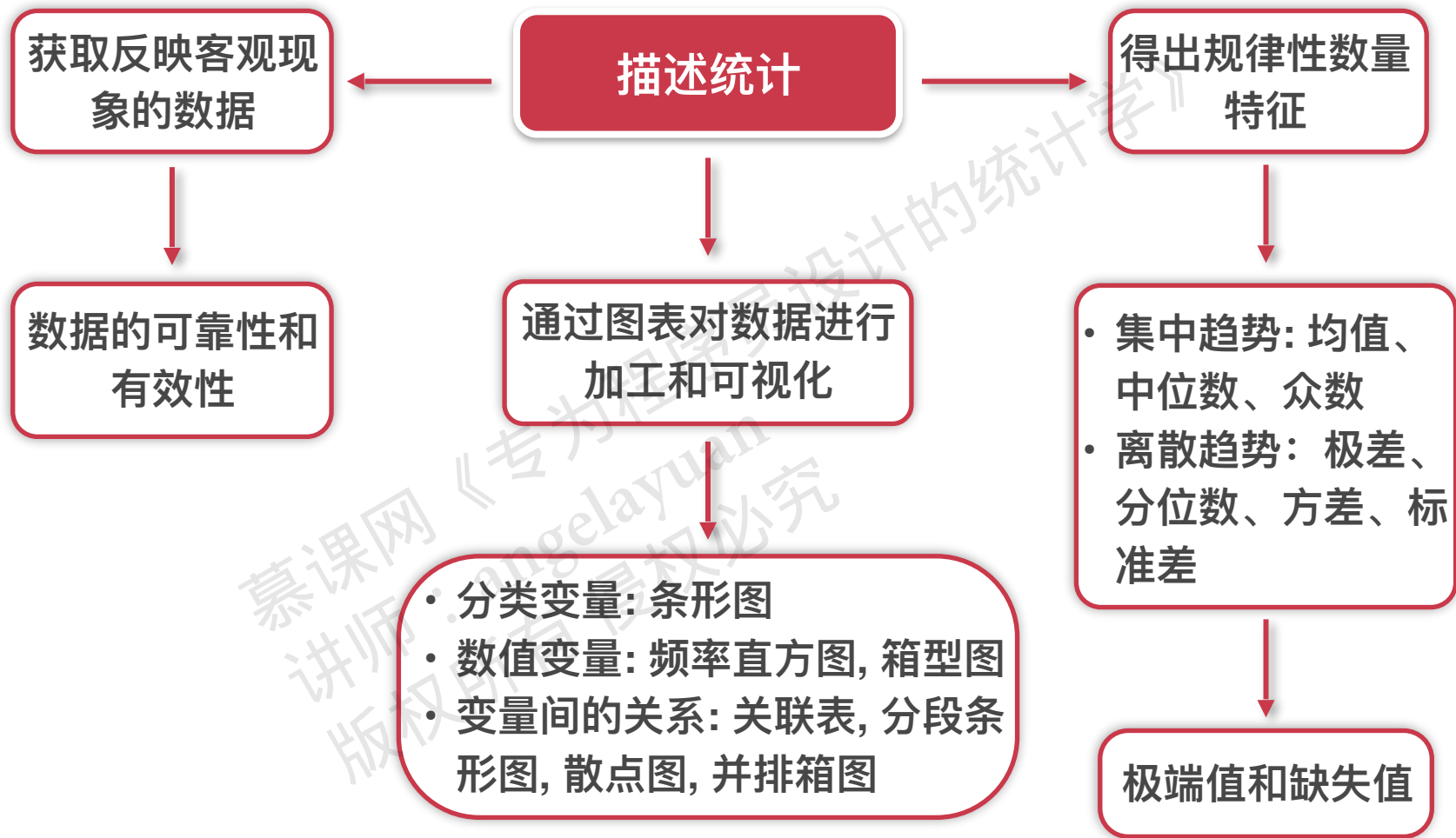
## 课程回顾和展望

慕课网《专为程序员设计的统计学》  
讲师：angelayuan  
版权所有 侵权必究



# 认识数据

尺度	举例	逻辑与数学运算	类别
名目	性别、颜色	$=, \neq$	定性/(无序)分类变量
次序	教育程度、评价	$=, \neq, >, <$	定性/(有序)分类变量
等距	温度、年份、时间	$=, \neq, >, <, +, -$	定量/数值变量 离散型
等比	身高、体重、年龄	$=, \neq, >, <, +, -, \times, \div$	定量/数值变量 连续型



## 必须了解的概率论知识

随机现象

在个别**试验**中结果具有不确定性

在大量重复试验中结果呈现出**固有规律性**  
(统计规律性)

- 可在相同条件下**重复**进行
- 能够事先明确**所有可能结果**
- 试验前**不确定**哪个结果会出现

样本空间

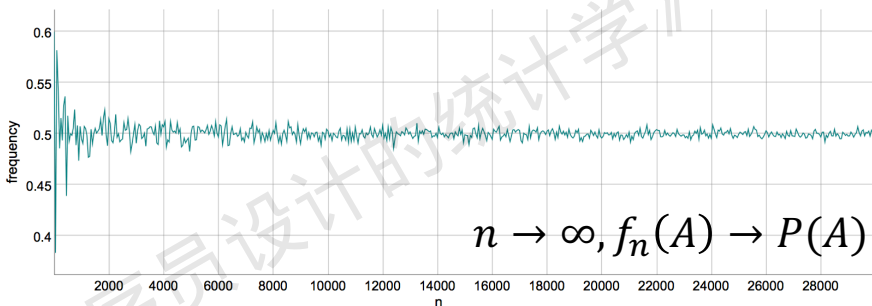
子集

(随机)事件

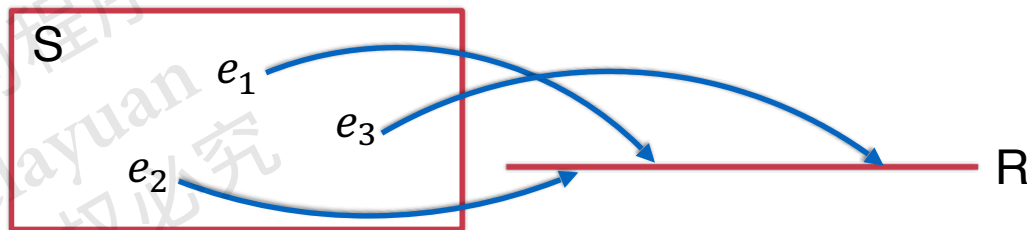


# 必须了解的概率论知识

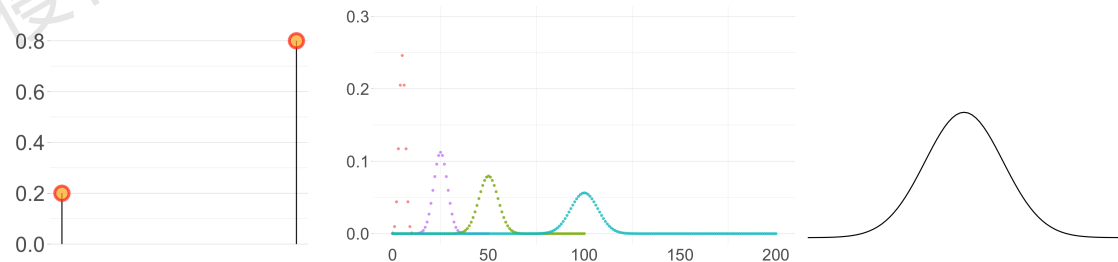
- 概率
- 频率vs概率
- 小数/大数定律



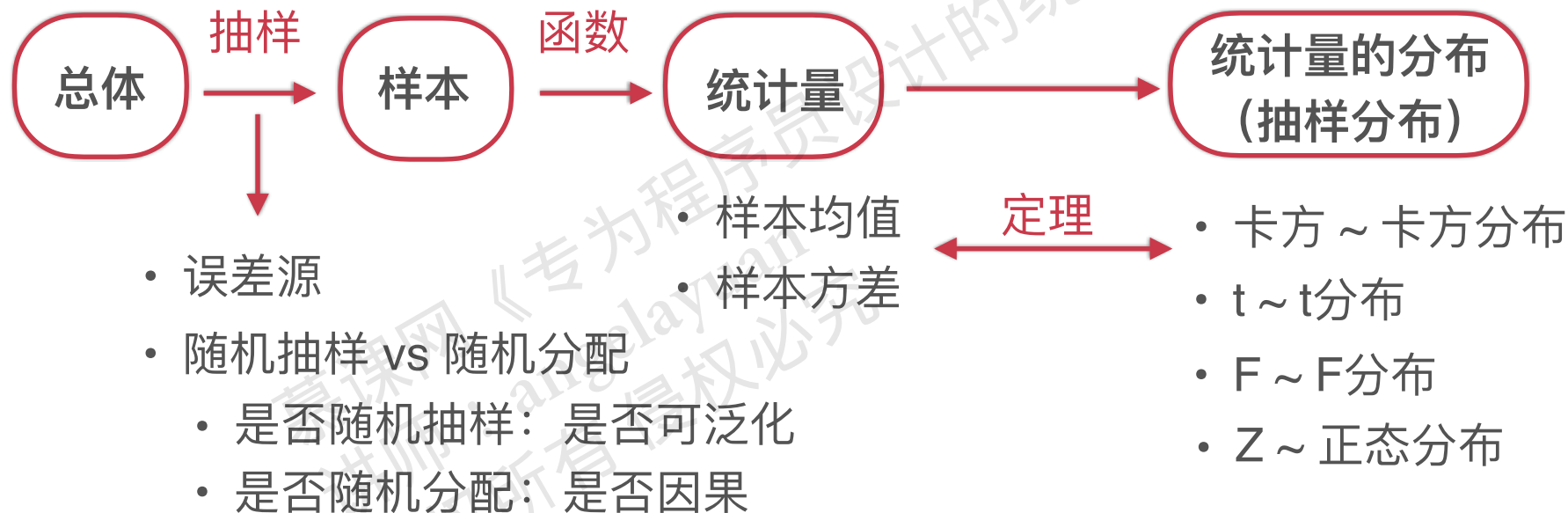
- 把数据/变量和概率论中的概念建立联系



- 概率分布
  - 两点分布, 二项分布
  - 正态分布



# 样本和抽样分布



# 参数估计

点估计

无偏性、有效性、相合性  
样本均值、样本方差

区间估计

一个正态  
总体

- 方差已知, 求均值
- 方差未知, 求均值
- 均值未知, 求方差

单侧置信  
区间

两个正态  
总体

- 两个方差已知, 求均值差
- 两个方差未知, 求均值差
- 两个均值未知, 求方差比



# 假设检验

```
graph LR; A[假设检验] --> B[什么是假设检验<br/>• 频率论 vs 贝叶斯]; A --> C[正态总体均值的假设检验]; A --> D[正态总体方差的假设检验]; A --> E[决策错误, 统计功效, 统计显著性与<br/>实际显著性, 各种检验的前提条件]; C --> F[• z检验<br/>• t检验]; D --> G[• 卡方检验<br/>• F检验]
```

- 什么是假设检验
- 频率论 vs 贝叶斯

## 正态总体均值的假设检验

- z检验
- t检验

## 正态总体方差的假设检验

- 卡方检验
- F检验

决策错误, 统计功效, 统计显著性与实际显著性, 各种检验的前提条件

# 方差分析

```
graph TD; A[方差分析] --> B[方差的分配]; A --> C[单因素方差分析]; A --> D[双因素方差分析]; C --> E[多重比较]; D --> E;
```

A flowchart illustrating the structure of ANOVA. At the top is a red box labeled '方差分析' (ANOVA). Three arrows point down from it to '方差的分配' (Allocation of Variance), '单因素方差分析' (One-way ANOVA), and '双因素方差分析' (Two-way ANOVA). From '单因素方差分析', an arrow points down to '多重比较' (Multiple Comparisons). From '双因素方差分析', an arrow points down and to the left to '多重比较'. To the left of the '多重比较' box, there is a horizontal line. Above the line is the text '自变量引起的变化 (可以解释的变化)' and below the line is '其他因素引起的变化 (无法解释的变化)'.

方差的分配

单因素方差分析

双因素方差分析

自变量引起的变化  
(可以解释的变化)

其他因素引起的变化  
(无法解释的变化)

多重比较

# 线性回归

- 协方差
- 相关及假设检验

## 一元线性回归

- 回归方程
- 最小二乘法
- 偏导数
- 系数的含义
- 前提条件
- 评价指标
- 假设检验
- 编程实现

## 多元线性回归

- 回归方程
- 最小二乘法
- 偏导数
- 系数的含义
- 前提条件
- 共线性
- 评价指标
- 假设检验

- 回归统一各种检验
- 统计学vs机器学习

## 非参数方法

```
graph LR; A[非参数方法] --> B[非参数方法的特点, 优缺点  
什么时候选择非参数方法]; A --> C["Wilcoxon signed-rank test  
Wilcoxon rank-sum test  
Spearman correlation test"]; A --> D[Bootstrap  
Permutation]; C --> E[从回归的角度认识非参数方法];
```

非参数方法的特点, 优缺点  
什么时候选择非参数方法

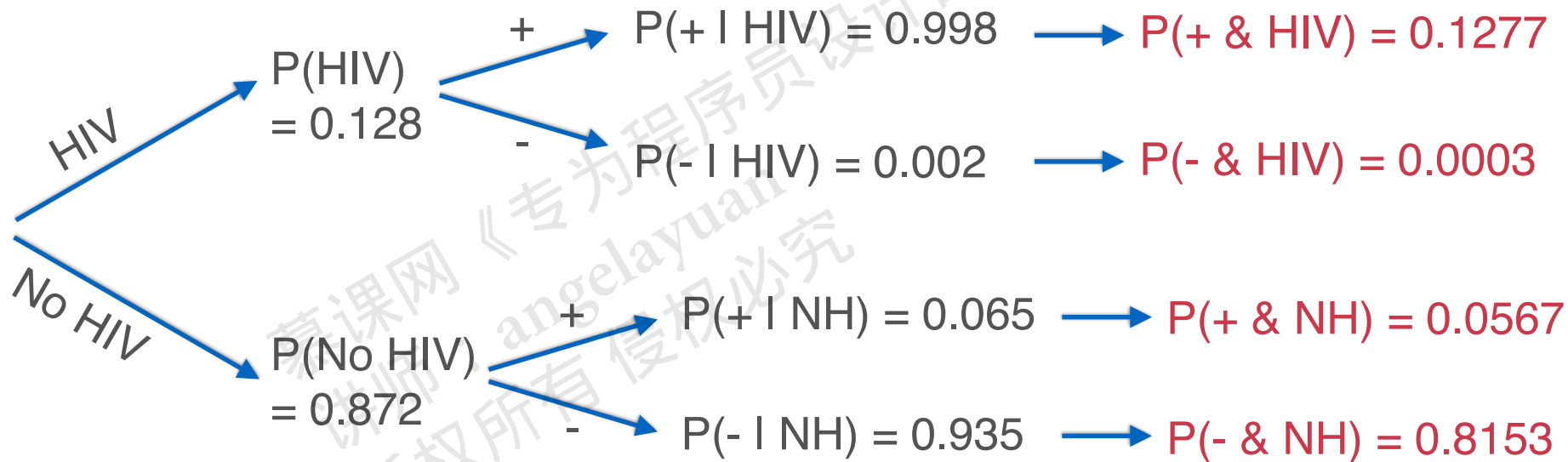
Wilcoxon signed-rank test  
Wilcoxon rank-sum test  
Spearman correlation test

Bootstrap  
Permutation

从回归的角度认识非参数方法

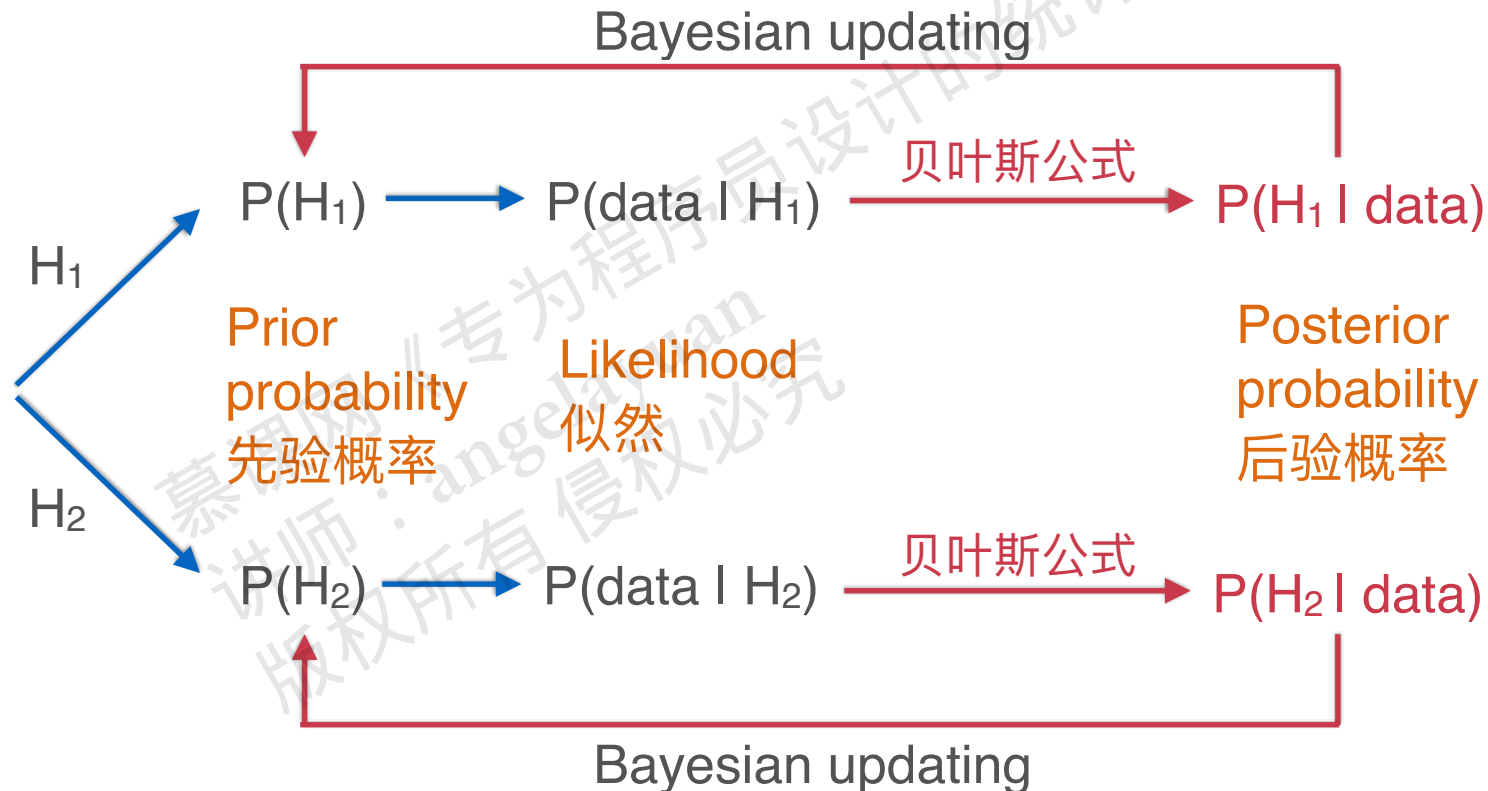
## 贝叶斯统计

条件概率, 乘法定理, 全概率公式, 贝叶斯公式; 概率树



## 贝叶斯统计

- 贝叶斯推断 (疾病诊断; 参数估计)
- 置信区间 (credible interval)



**希望这个课程让你真正学懂了统计学 :)**

慕课网《专为程序员设计的统计学》  
讲师：angelayuan  
版权所有 侵权必究