

概率论与数理统计

刘维湘

深圳大学医学部
生物医学工程学院

Table Of Contents

目录

课程背景

应用情况

课程要求

大学我们都学了哪些数学

教材与参考书

考核与考试

课程内容

概率论：从模型出发

数理统计：从数据出发

个人体会：数学课程的内容安排与学习

一般的数学知识组织框架

Outline

课程背景

应用情况

课程要求

大学我们都学了哪些数学

教材与参考书

考核与考试

课程内容

概率论：从模型出发

数理统计：从数据出发

个人体会：数学课程的内容安排与学习

一般的数学知识组织框架

应用情况：考研用 vs 挣学分 vs 想学（感兴趣）

国内情况

- ▶ 数据统计与分析：北京时间 2015 年 11 月 25 日消息，今天下午国家体育总局正式发布《2014 年国民体质监测公报》，而本次公报指标体系由身体形态、机能和素质三大类共计 20 项指标组成，全面解读近五年来中国国民体质变化情况，值得一提的是，一些数据绝对出乎很多人的预料，国民体质达标率达到 89.6%，公报也体现了 2014 年中国人体质状况。
- ▶ 发现数据中的规律：根据身体形态等指标系数，中国女人更加健康。
- ▶ 利用其中的规律：分别针对男人，女人，小孩，老人，等不同层次需求设计产品

应用情况：都用

国外情况

- ▶ 数据统计与分析：如老美喜欢拿中国人权问题说事
- ▶ 发现数据中的规律：
- ▶ 利用其中的规律：

课程的使用实例

- ▶ 数据统计与分析：被富有 \Rightarrow 不要被数据忽悠

课程的使用实例

- ▶ 数据统计与分析：被富有 \Rightarrow 不要被数据忽悠
- ▶ 发现数据中的规律：啤酒与尿布的故事 \Leftarrow 数据挖掘 \Rightarrow 大数据时代

课程的使用实例

- ▶ 数据统计与分析：被富有 \Rightarrow 不要被数据忽悠
- ▶ 发现数据中的规律：啤酒与尿布的故事 \Leftarrow 数据挖掘 \Rightarrow 大数据时代
- ▶ 利用其中的规律：推荐系统

课程的使用实例

- ▶ 数据统计与分析：被富有 \Rightarrow 不要被数据忽悠
- ▶ 发现数据中的规律：啤酒与尿布的故事 \Leftarrow 数据挖掘 \Rightarrow 大数据时代
- ▶ 利用其中的规律：推荐系统
- ▶ 人机大战的技术支持：Alpha GO \Leftarrow 人工智能，机器学习，深度学习，模式识别 \rightarrow

课程的使用实例

- ▶ 数据统计与分析：被富有 \Rightarrow 不要被数据忽悠
- ▶ 发现数据中的规律：啤酒与尿布的故事 \Leftarrow 数据挖掘 \Rightarrow 大数据时代
- ▶ 利用其中的规律：推荐系统
- ▶ 人机大战的技术支持：Alpha GO \Leftarrow 人工智能，机器学习，深度学习，模式识别 \Rightarrow
- ▶ 无处不在的 AI：图像，语音，文本，…世界各国的竞赛通道

课程的使用实例

- ▶ 数据统计与分析：被富有 \Rightarrow 不要被数据忽悠
- ▶ 发现数据中的规律：啤酒与尿布的故事 \Leftarrow 数据挖掘 \Rightarrow 大数据时代
- ▶ 利用其中的规律：推荐系统
- ▶ 人机大战的技术支持：Alpha GO \Leftarrow 人工智能，机器学习，深度学习，模式识别 \Rightarrow
- ▶ 无处不在的 AI：图像，语音，文本，…世界各国的竞赛通道
- ▶ …保险业，彩票设计，金融业 …

课程的使用实例

- ▶ 数据统计与分析：被富有 => 不要被数据忽悠
- ▶ 发现数据中的规律：啤酒与尿布的故事 <= 数据挖掘 => 大数据时代
- ▶ 利用其中的规律：推荐系统
- ▶ 人机大战的技术支持：Alpha GO <+ 人工智能，机器学习，深度学习，模式识别 +>
- ▶ 无处不在的 AI：图像，语音，文本，…世界各国的竞赛通道
- ▶ …保险业，彩票设计，金融业 …
- ▶ **BME: AI4Health**

课程的使用实例

- ▶ 数据统计与分析：被富有 \Rightarrow 不要被数据忽悠
- ▶ 发现数据中的规律：啤酒与尿布的故事 \Leftarrow 数据挖掘 \Rightarrow 大数据时代
- ▶ 利用其中的规律：推荐系统
- ▶ 人机大战的技术支持：Alpha GO \Leftarrow 人工智能，机器学习，深度学习，模式识别 \Rightarrow
- ▶ 无处不在的 AI：图像，语音，文本，…世界各国的竞赛通道
- ▶ …保险业，彩票设计，金融业 …
- ▶ BME: AI4Health
- ▶ …(请三位同学讲解发生在自己身边的与统计有关的事情、故事或事故)

真实的故事

问题：1943 年，美国空军向美国统计学家亚伯拉罕·沃尔德 (Abraham Wald) 寻求加固飞机的建议。飞机只能搭载有限重量的护甲，怎么办？

真实的故事

问题：1943 年，美国空军向美国统计学家亚伯拉罕·沃尔德 (Abraham Wald) 寻求加固飞机的建议。飞机只能搭载有限重量的护甲，怎么办？

建议 1：当时已有的建议是将护甲装在机翼、机身中央和尾部。为何选择这些部位？因为完成任务返航的那些轰炸机在这些部位布满了弹孔。

真实的故事

问题：1943 年，美国空军向美国统计学家亚伯拉罕·沃尔德 (Abraham Wald) 寻求加固飞机的建议。飞机只能搭载有限重量的护甲，怎么办？

建议 1：当时已有的建议是将护甲装在机翼、机身中央和尾部。为何选择这些部位？因为完成任务返航的那些轰炸机在这些部位布满了弹孔。

建议 2：沃尔德解释称，这样做是错误的。空军之所以见到完成任务返航的那些轰炸机在这些部位布满弹孔，是因为当飞机被击中机翼、尾部或机身中央时，它仍然能返航。沃尔德当时问道，那些被击中其他部位的飞机哪里去了？它们始终没有返航。瓦尔德的建议是，将护甲装在返航的飞机毫无损伤的那些部位。

讨论：空战中，专家根据返航飞机翼部和尾部等位置中弹最多的特点，判断出，如果要强化飞机装甲，应该强化翼部和尾部。但如果仅仅根据返航的飞机上的弹孔分布，来研究该加强飞机的哪部分时，就忽略掉了“飞机被击落”这个筛选的过程。也就是说，能返航的飞机都是躲过防空炮火筛选的、没有受过致命伤的。他们多中弹于翼部和尾部，而发动机、油箱和驾驶员仓位完好，这恰恰说明翼部和尾部不是致命伤，不应该加强；而油箱和驾驶员仓位受伤的飞机，基本没有活着回来的，因此发动机、油箱和驾驶员仓位才是真正致命的地方，这些部位受伤的飞机却因为被击落而被筛出研究样本，从而变成了“不会说话的死人”，或“沉默的数据”。

[Survival Bias, 生存者偏差]

应用情况：数据中的噪音，我怎么能听见!?

Noisy data

- 现象：淘宝数据显示，男人们比女人更趋向于浏览和购买女人用品？[来自一个数据分析大牛的分享，忘了链接了，谁知道了请告诉我，谢谢先！]

应用情况：数据中的噪音，我怎么能听见!?

Noisy data

- ▶ 现象：淘宝数据显示，男人们比女人更趋向于浏览和购买女人用品？[来自一个数据分析大牛的分享，忘了链接了，谁知道了请告诉我，谢谢先！]
- ▶ 原因？你想到了吗

应用情况：数据中的噪音，我怎么能听见!?

Noisy data

- ▶ 现象：淘宝数据显示，男人们比女人更趋向于浏览和购买女人用品？[来自一个数据分析大牛的分享，忘了链接了，谁知道了请告诉我，谢谢先！]
- ▶ 原因？你想到了吗
- ▶ 原因之一：因为更多的女人用其丈夫或男友的账户进行淘宝购物活动！

应用情况：数据中的规律，我们怎么去发现？

大学生吃早餐和成绩的关系

► 现象：大学生作息时间与大学生学习成绩！]

应用情况：数据中的规律，我们怎么去发现？

大学生吃早餐和成绩的关系

- ▶ 现象：大学生作息时间与大学生学习成绩！]
- ▶ 规律？你想到了吗
- ▶ 规律之一：发现按时早餐的同学成绩都不错！

Outline

课程背景

应用情况

课程要求

大学我们都学了哪些数学

教材与参考书

考核与考试

课程内容

概率论：从模型出发

数理统计：从数据出发

个人体会：数学课程的内容安排与学习

一般的数学知识组织框架

大学我们都学了哪些数学：我的天，全还给老师啦

- ▶ 集合论
- ▶ 线性代数
- ▶ 高等数学
- ▶ … (请三位同学讲解自己对集合论, 线性代数或高等数学的学习体会)

大学我们都学了哪些数学：我的天，全还给老师啦

- ▶ 集合论
- ▶ 线性代数
- ▶ 高等数学
- ▶ …(请三位同学讲解自己对集合论, 线性代数或高等数学的学习体会)
- ▶ 会用一门编程语言 (BASIC, C/C++, Java, Matlab, Python, R, Julia, …) 来解决数学问题?

Outline

课程背景

应用情况

课程要求

大学我们都学了哪些数学

教材与参考书

考核与考试

课程内容

概率论：从模型出发

数理统计：从数据出发

个人体会：数学课程的内容安排与学习

一般的数学知识组织框架

教材与参考书：别贪多，搞定一本足也

- ▶ 盛骤等. 概率论与数理统计. 浙大第四版.
- ▶ 附带的习题集。先自己做，不要抄答案。

教材与参考书：别贪多，搞定一本足也

- ▶ 盛骤等. 概率论与数理统计. 浙大第四版.
- ▶ 附带的习题集。先自己做，不要抄答案。
- ▶ 思考

Outline

课程背景

应用情况

课程要求

大学我们都学了哪些数学

教材与参考书

考核与考试

课程内容

概率论：从模型出发

数理统计：从数据出发

个人体会：数学课程的内容安排与学习

一般的数学知识组织框架

考核与考试：临时要抱佛脚，平时更要努力

- ▶ 平时成绩, 如课堂测验与课后作业
- ▶ 考试成绩

考核与考试：临时要抱佛脚，平时更要努力

- ▶ 平时成绩, 如课堂测验与课后作业
- ▶ 考试成绩
- ▶ 课堂出勤

考核与考试：临时要抱佛脚，平时更要努力

- ▶ 平时成绩, 如课堂测验与课后作业
- ▶ 考试成绩
- ▶ 课堂出勤
- ▶ 课堂提问

考核与考试：临时要抱佛脚，平时更要努力

- ▶ 平时成绩，如课堂测验与课后作业
- ▶ 考试成绩
- ▶ 课堂出勤
- ▶ 课堂提问
- ▶ 老师对你的印象...

考核与考试：临时要抱佛脚，平时更要努力

- ▶ 平时成绩，如课堂测验与课后作业
- ▶ 考试成绩
- ▶ 课堂出勤
- ▶ 课堂提问
- ▶ 老师对你的印象...
- ▶ 加入我的课题组... 约吗? email: wxliu@szu.edu.cn

Outline

课程背景

应用情况

课程要求

大学我们都学了哪些数学

教材与参考书

考核与考试

课程内容

概率论：从模型出发

数理统计：从数据出发

个人体会：数学课程的内容安排与学习

一般的数学知识组织框架

概率论：从模型出发：石头剪刀布游戏

上帝真的在玩骰子吗

- ▶ 概率
- ▶ 概率的计算

概率论：从模型出发：石头剪刀布游戏

上帝真的在玩骰子吗

- ▶ 概率
- ▶ 概率的计算
- ▶ 随机变量

Outline

课程背景

应用情况

课程要求

大学我们都学了哪些数学

教材与参考书

考核与考试

课程内容

概率论：从模型出发

数理统计：从数据出发

个人体会：数学课程的内容安排与学习

一般的数学知识组织框架

数理统计：从数据出发 ： 我们每天都在产生数据

世界即数据

- ▶ 样本与分布
- ▶ 参数估计

数理统计：从数据出发 ： 我们每天都在产生数据

世界即数据

- ▶ 样本与分布
- ▶ 参数估计
- ▶ 点估计

数理统计：从数据出发：我们每天都在产生数据

世界即数据

- ▶ 样本与分布
- ▶ 参数估计
- ▶ 点估计
- ▶ 矩估计与最大似然估计 (MLE)，估计的无偏性 [机器学习的很多方法都用 MLE!]
- ▶ 区间估计 (*)

数理统计：从数据出发：我们每天都在产生数据

世界即数据

- ▶ 样本与分布
- ▶ 参数估计
- ▶ 点估计
- ▶ 矩估计与最大似然估计 (MLE)，估计的无偏性 [机器学习的很多方法都用 MLE!]
- ▶ 区间估计 (*)
- ▶ 假设检验 (*)

Outline

课程背景

应用情况

课程要求

大学我们都学了哪些数学

教材与参考书

考核与考试

课程内容

概率论：从模型出发

数理统计：从数据出发

个人体会：数学课程的内容安排与学习

一般的数学知识组织框架

我们想学习什么内容

- ▶ 发现已知数学不能解决的问题和现象
- ▶ 定义，公理
- ▶ 性质，定理，推论
- ▶ 应用来解决问题
- ▶ 碰到目前的方法解决不了的问题
- ▶ 需要新的定义和方法 [新假设、新概念和新方法]

历史上有好几首咏蛙的诗。下面这首据说为毛泽东所作。

- ▶ 咏蛙
- ▶ 毛泽东
- ▶ 独坐池塘如虎踞，绿荫树下养精神。
- ▶ 春来我不先开口，哪个虫儿敢作声？