Advances in Education 教育进展, 2024, 14(7), 498-503
Published Online July 2024 in Hans. https://www.hanspub.org/journal/ae
https://doi.org/10.12677/ae.2024.1471192

Hans汉斯

基于数学实验的贝叶斯公式重点概念解析

刘　倩*，李　伟

西安电子科技大学数学与统计学院，陕西　西安

收稿日期：2024年6月7日；录用日期：2024年7月10日；发布日期：2024年7月17日

摘　要

本文对贝叶斯公式的教学方案进行了设计，通过对两个经典实例进行统计软件编程辅助课堂教学，这种互动式的示例实验易于让学生深刻理解贝叶斯公式中所包含的先验概率、后验概率以及公式的重要意义，从而突破以往理论教学的缺陷。

关键词

数学实验，贝叶斯公式，先验概率，后验概率，教学设计

An Analysis of Key Concepts of Bayes Formula Based on Mathematical Experiments

Qian Liu*, Wei Li

School of Mathematics and Statistics, Xidian University, Xi'an Shaanxi

Received: Jun. 7th, 2024; accepted: Jul. 10th, 2024; published: Jul. 17th, 2024

Abstract

This paper designs a teaching scheme for Bayes formula, by using statistical software programming to assist classroom teaching for two classic examples. This interactive example experiment is easy to help students deeply understand the prior probability, posterior probability, and the significance of the formula in Bayes formula, thus overcoming the shortcomings of theoretical teaching in the past.

*通讯作者。

文章引用: 刘倩, 李伟. 基于数学实验的贝叶斯公式重点概念解析[J]. 教育进展, 2024, 14(7): 498-503.
DOI: 10.12677/ae.2024.1471192

Keywords

Mathematical Experiments, Bayes Formula, Prior Probability, Posterior Probability, Teaching Design

1. 引言

概率统计以概率论为基础，利用模型和统计推断方法对收集的数据进行分析和处理。课程中包含了大量抽象复杂概念，以往传统课堂教学所采用的教学方法、教学内容和手段过于注重理论知识的传授，而忽视学生解决实际问题能力的培养以及课程蕴含思政方面的思考。我校概率统计课程组以“突破创新思维，强化实践能力”为指导思想，在课堂教学过程中注重实践教学环节，并强化统计软件的应用，这些手段在一定程度上可以弥补理论教学和习题课教学的缺陷。主讲教师通过动态、可视化、可操作的互动实验示例辅助教学，强化基本概念，解决实际问题。

贝叶斯公式是概率统计教学过程中的一个重点、难点问题。案例式、启发式教学，以应用性为导向并融入课程思政的教学设计[1]-[3]被广泛使用。本文尝试通过对经典案例进行统计软件编程，重点剖析先验概率、后验概率两个重要概念，深化应用，让学生形成对公式的直观性理解，这种基于数学实验的方法在教学中取得了良好的效果。

2. 教学目标

在浙江省永嘉中学的校园里面，该校 93 届校友赠送了一块纪念石。石头一面显示了一个数学公式，另一面刻着这样一句话“人工智能的奠基理论之一——贝叶斯定理”。看来，永嘉中学是一所有文化，懂教育，爱数学的中学。那它为什么要设立这样一个纪念石？这个看似简单的公式对当代中学生的人生又具有什么样启示呢？我们希望通过这节课的学习使学生既获得知识上的收获，又有人生体验的收获。

2.1. 本节专业知识教学目标

掌握贝叶斯公式的数学基础和原理；掌握贝叶斯公式的应用领域，能够正确运用贝叶斯公式解决实际问题；培养学生的逻辑思维和推理能力，提高解决问题的能力。

2.2. 本节思政教学目标

贝叶斯公式蕴含执果寻因，探求真理的辩证思想，同时体现了信息更新的重要性，人们普遍都有保守主义情结，即便出现了新信息，主观上不愿意根据新信息来及时更新、调整先验概率。贝叶斯公式告诉我们应当根据新情况及时更新已有认知。另一方面，课程通过案例教学，引导学生铸就诚信品质，树立正确的人生观，培养学生的核心价值观。

3. 教学设计

首先以定理的形式给出贝叶斯公式严格的数学描述，其次给出两个典型案例的教学设计过程。

3.1. 定理

定理[4]假设 $B_1, B_2, \cdots, B_n$ 是两两互不相容的随机事件，且 $P(B_i)>0$，($i = 1, 2, \cdots, n$)。若对于任意事件 A，有 $A \subset B_1 \cup B_2 \cup \cdots \cup B_n$，且 $P(A)>0$，则

$$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n}P(B_j)P(A|B_j)},\ i=1,2,\cdots,n \tag{1}$$

称(1)式为贝叶斯公式。

根据全概率公式和条件概率，易于完成贝叶斯定理的证明。

3.2. 典型案例的教学设计

案例 1：贝叶斯定理的发明者托马斯·贝叶斯提出了一个很有意思的假设：“如果一个袋子中共有 10 个球，分别是黑球和白球，但是我们不知道它们之间的比例是怎么样的，现在仅通过摸出的球的颜色，是否能判断出袋子里面黑白球的比例？

上述问题与学生高中时期所接受的概率问题“一个袋子里面有 10 个球，其中 4 个黑球，6 个白球，如果你随机抓取一个球，那么是黑球的概率是多少？”有所区别，毫无疑问，该问题的答案是 0.4。

这个问题非常简单，因为我们事先知道了袋子里面黑球和白球的比例，所以很容易算出摸出某种颜色小球的概率，但是在某些复杂情况下，我们无法得知“比例”。此时就引出了贝叶斯提出的问题。这种问题称为逆概率问题，现实生活中，大部分都是像上面的“逆概率”问题。怎么解决这个问题？

授课教师继续提问：假设有两个各装有 100 个小球的箱子，甲箱子中有 70 个红球，30 个绿球；乙箱子中有 30 个红球，70 个绿球。假设我们随机选择其中一个箱子，从中拿出一个球观察颜色，记录小球颜色后再放回原箱中。如此重复 12 次，记录得到 8 次红球，4 次绿球。问题是，你认为被选择的箱子是甲箱子的概率有多大？

课前授课教师先对学生进行实际调查，让学生估计这个概率。结果表明，大部分人都低估了选择的是甲箱子的概率。为什么呢？再次请同学们基于贝叶斯公式编写 Python 程序，最终发现根据贝叶斯公式，正确答案是 96.7%。

详细 Python 代码如下所示：

```
defbayesFunc(pIsBox1, pBox1, pBox2):
return (pIsBox1 * pBox1)/((pIsBox1 * pBox1) + (1 - pIsBox1) * pBox2)
defredGreenBallProblem():
pIsBox1 = 0.5
# consider 8 red ball
fori in range(1, 9):
pIsBox1 = bayesFunc(pIsBox1, 0.7, 0.3)
print " After red %d > in  甲  box: %f" % (i, pIsBox1)
# consider 4 green ball
fori in range(1, 5):
pIsBox1 = bayesFunc(pIsBox1, 0.3, 0.7)
print " After green %d > in  甲  box: %f" % (i, pIsBox1)
redGreenBallProblem()
```

运行结果如下：

After red 1 > in 甲 box: 0.700000

After red 2 > in 甲 box: 0.844828

After red 3 > in 甲 box: 0.927027

After red 4 > in 甲 box: 0.967365

After red 5 > in 甲 box: 0.985748

After red 6 > in 甲 box: 0.993842

After red 7 > in 甲 box: 0.997351

After red 8 > in 甲 box: 0.998863

After green 1 > in 甲 box: 0.997351

After green 2 > in 甲 box: 0.993842

After green 3 > in 甲 box: 0.985748

After green 4 > in 甲 box: 0.967365

在这个代码中，A 代表从甲盒中取球，R 代表取到红球。当取出第一个红球的时候，$P(A|R)=P(R|A)*P(A)/P(R)$。分别展开，在甲盒子中取出红球的概率 $P(R|A)=70/100=0.7$，选择甲盒子的概率 $P(A)=0.5$，取出红球的概率(从两个盒子中)，根据全概率公式等于从甲盒子中取出红球的概率加上从乙盒子中取出红球的概率 $P(R)=0.7*0.5+0.3*0.5=0.5$，所以 $P(A|R)=0.7*0.5/0.5=0.7$。

请注意，第一个红球的出现导致了我们对两个盒子的概率猜测发生了修正。现在甲盒子被选取的概率是 0.7，乙盒子被选取的概率自然就是 0.3。当第二个红球出现的时候，$P(A)=0.7$，$P(R)=0.7*0.7+0.3*0.3=0.58$，$P(A|R)=0.7*0.7/0.58=0.845$，于是第二次迭代之后，甲盒子的概率称为了 0.845。

在这个调查问题里面，8 次红球与 4 次绿球出现的顺序并不重要，因为红球的出现总是增加选择甲箱子的概率，而绿球的出现总是减少选择甲箱子的概率。不妨将红球出现的情况以及绿球出现的情况摆在一起，从程序结果来看，结果确实如此。

人们普遍具有的保守主义情结，用这个例子解释就是，新信息是 R 事件不断发生，人们本应该根据这个信息去更新 A 事件发生的概率，但人们却更愿意固守之前估计的 A 事件发生的概率。这个例子告诉我们应当根据新情况更新先验概率 $P(A)$。

案例 2：应用某种检测方法进行疾病诊断，设随机事件 A 表示“试验反应呈阳性”，随机事件 C 表示“被诊断者确实患有该疾病”，已知条件概率 $P(A|C)=0.95$，$P(\bar{A}|C)=0.05$，$P(A|\bar{C})=0.05$，$P(\bar{A}|\bar{C})=0.95$。在自然人群中对该种疾病进行普查，且 $P(C)=0.005$，求 $P(C|A)=?$

根据题设，该诊断方法具有如下的检验效果：条件概率 $P(A|C)$ 表示患者对试验反应为阳性的概率，0.95 表明由于检查方法的不完善，患者也未必一定会出现阳性结果，那么由条件概率的加法公式得到 $P(\bar{A}|C)=0.05$；类似的，健康人也可能检测为阳性，条件概率 $P(A|\bar{C})$ 为 0.05，那么对立事件的概率 $P(\bar{A}|\bar{C})$ 为 0.95。若对人群进行普查，假设某地区患这种疾病的人占 0.005，即 $P(C)=0.005$，如果随机抽查了一个人，其试验反应为阳性，那么此人真正患病的概率？

随机事件 C 和 $\bar{C}$ 都会导致结果事件 A 的发生，探求导致 A 发生的一种潜在原因发生的概率，这正是一个执果寻因的过程，而且涉及因果关系的转换，是一个典型的贝叶斯公式的应用。

课堂上，授课教师先请同学们凭直觉判断此人真正患病的可能性的大小，再根据贝叶斯公式给 Python 程序，进行推断。具体程序如下：

```
defbayes(prioProb, liklihood, i)
temp=prioProb*liklihood
total=temp.sum()
return temp[i-1]/total
importnumpy as np
fromsympy import Rational as R
prioProb=np.array([R(1,200),R(199,200)])
likelihood=np.array([R(19,20), R(1, 20)])
p=bayes(prioProb,likelihood, 1)
print('P(C|A)=%s'%p)
```

运行结果如下：

$P(C|A)=19/218=0.087$

面对真阳性率高达 95%的疾病检测结果，此人“有病”的概率到底多大呢？结果发现，平均一千名具有阳性反应的人群中，真正患病的人数量很少，大约只有 87 人。这究竟是什么原因造成的呢？

授课教师可以提示同学们通过改变先验概率 $P(C)$ 的数值，观察后验概率 $P(C|A)$ 如何变化？比如，当先验概率 $P(C)=0.5$ 时，后验概率 $P(C|A)$ 将达到 0.96。事实上，从贝叶斯公式本身可以得到解释：尽管健康人呈现阳性反应的概率为 0.05，但是由于该疾病的发病率仅为 0.005，实在太小，从而导致检测结果为假阳性的部分相对较大，最终造成 $P(C|A)$ 值较小。

当在不同的健康和患病比例实验条件下，以及健康和患病状态下不同的检测阴阳性条件概率下，患者中的假阳性与假阴性会有不同。当固定条件概率的时候，健康人群比例增大，假阳性占比会增大。当固定健康和患病的先验分布的时候，当健康状态下阳性的条件概率增大时，假阳性占比也会增大。

这个典型案例告诉我们，对于发病率很低的疾病，进行普查没有实际意义，尤其是检查的准确性也不是很高的情况。通常，医生先辅助其他简单易行的手段进行分析判断，当他高度怀疑某个对象时，才会建议应用该种疾病检测方法，此时疾病发病率(先验概率)已经显著地增加了。当医生再进行一次检测，一旦呈现阳性结果，那么应用贝叶斯公式，该对象患病的概率将上升至 0.73；再做一次检测，还是呈现阳性，此时医生就可以确诊了，患者患病的概率高达 0.985。

4. 重点概念

1) 贝叶斯公式是从先验概率到后验概率的转化公式

获取先验假设(先验概率)是贝叶斯分析中的一个关键步骤，它基于实验或观察之前对某个问题的理解和信念。先验假设可以通过多种方式获得，具体方法可由问题的性质、可用信息的数量和质量，以及分析者的专业判断决定。常见的方法包括历史数据、专家知识、文献调研、均匀或无信息先验、实验或调查等方法。

回顾前面的引例，先验概率就是对过去已经掌握生产情况的反映，对试验将要出现的结果提供一定的信息，当试验结果 A 出现之后，利用这个新的信息修正先验概率，应用贝叶斯公式得到的条件概率正是后验概率。比如，学生们熟悉的伊索寓言中《狼来了》的故事，用贝叶斯公式可以给出这个故事的概率论解释：诚信重要的原因就在于人们会根据与你交往过程中发生的各种事件修正对你的看法，不断用事件的后验概率代替它的先验概率。这种修正过程会反复地进行，本质正是贝叶斯公式，量化了从先验概率到后验概率的转化过程。

2) 贝叶斯公式又称为逆概率公式

从形式上看，全概率公式是求一个复杂事件发生的概率，而贝叶斯公式是求一个事件发生的条件下，另一个事件的条件概率。从思想上看，全概率公式是将一个复杂的事件分解为若干个简单的子事件，然后利用子事件发生的概率和条件概率来求出复杂事件发生的概率。贝叶斯公式是利用已知的结果，反推出原因的可能性，然后利用原因发生的概率和条件概率来更新对原因发生的概率的估计。基于这种思想，贝叶斯公式有很多有趣的逆问题应用[5]-[8]。

这里，我们给出几个贝叶斯公式的应用领域。地震定位：通过从多个地震站接收到的地震波到达时间反推地震的震源位置。医学成像：从体外测得的信号反推体内组织的结构或功能状态。天气预报：通过观测数据(如温度、湿度、风速等)来估计天气模型的参数，进而预测未来的天气状况。机器学习：在监督学习任务中，通过从输入(特征)和输出(标签)的数据集中学习模型的参数，以预测新输入数据的输出。

可以说，凡是需要做出概率预测的地方都能见到贝叶斯公式的影子。

5. 结束语

贝叶斯的方法在贝叶斯去世后的很长一段时间内并未受到重视，直到 19 世纪末、20 世纪初，随着统计学和概率论的发展，贝叶斯定理被广泛用于科学研究、经济学、医学、工程学等领域。20 世纪中叶，随着计算技术的进步，贝叶斯方法开始得到更加广泛的应用。它的灵活性和在处理不确定性信息方面的能力，使其成为许多领域不可或缺的工具。

贝叶斯公式让我们可以在面对不确定性和有限的信息时，进行合理的推断和决策。通过不断地收集数据并更新我们的信念，我们能够逐渐接近真相。在人生中，我们会面临许多不确定的情况，应该保持对新知识、新经验的开放态度，学会适应和应对不确定性，这些都是提高生活质量的关键。至此，我们基于数学实验这个有利工具揭开了贝叶斯公式的神秘面纱。

本文，我们基于数学实验不仅讲授了贝叶斯公式的原理，更是通过这种互动式的实验让学生明白实验中各步骤、参数和理论概念间的联系。这种将教学实验与大数据、动画、图表等现代信息化技术相结合的教学手段值得推广。广大授课教师可以采用 Python、Matlab、R、SAS、SPSS 等多种统计软件进行虚拟仿真交互性实验，比如说明频率稳定性的抛硬币实验、利用蒙特卡洛方法估计积分的实验、进行骰子实验掌握带权平均和期望的关系的实验、通过改变正态分布均值观察似然函数变化，理解掌握极大似然估计的思想和计算的实验等等。无论是验证性还是探索性实验，这些教学内容都极大地激发了学生的学习兴趣，将学生的主观能动性发挥到极致。

基金项目

2024 年西安电子科技大学教育教学改革研究项目(B2316)。

参考文献

[1] 韦来生, 张伟平. 贝叶斯分析[M]. 合肥: 中国科学技术大学出版社, 2013: 5-6.

[2] 张倩, 杨文国. 基于案例教学法的贝叶斯公式教学设计与实践[J]. 科技风, 2021(31): 91-93.

[3] 刘倩. 由古典概型引入贝叶斯公式的一种教学设计[J]. 高师理科学刊, 2016, 36(6): 56-60.

[4] 魏宗舒. 概率论与数理统计教程[M]. 北京: 高等教育出版社, 2008: 40-42.

[5] 杨静, 陈冬等. 贝叶斯公式的几个应用[J]. 大学数学, 2011, 27(2): 166-169.

[6] 王丽. 浅析贝叶斯公式及其在概率推理中的应用[J]. 科技创新导报, 2010(24): 136.

[7] 周丽华. 市场预测中的贝叶斯公式应用[J]. 商场现代化, 2006(34): 55-56.

[8] 廖杰. 贝叶斯公式在河流水质综合评价中的应用[J]. 四川师范大学学报(自然科学版), 2007, 30(4): 519-522.