概率论与数理统计
实用案例分析

金明　编著

中国统计出版社
China Statistics Press

概率论与数理统计
实用案例分析

金明　编著

中国统计出版社
China Statistics Press

图书在版编目(CIP)数据

概率论与数理统计实用案例分析 / 金明编著. —— 北京：中国统计出版社，2014.8
ISBN 978－7－5037－7081－4

Ⅰ. ①概… Ⅱ. ①金… Ⅲ. ①概率论－高等学校－教学参考资料②数理统计－高等学校－教学参考资料 Ⅳ. ①O21

中国版本图书馆 CIP 数据核字(2014)第 088802 号

概率论与数理统计实用案例分析

作　　者/金　明
责任编辑/徐　颖
封面设计/李雪燕
出版发行/中国统计出版社
通信地址/北京市丰台区西三环南路甲 6 号　邮政编码/100073
电　　话/邮购(010)63376909　书店(010)68783171
网　　址/http://csp.stats.gov.cn
印　　刷/河北天普润印刷厂
经　　销/新华书店
开　　本/787×1092mm　1/16
字　　数/330 千字
印　　张/14.5
版　　别/2014 年 8 月第 1 版
版　　次/2014 年 8 月第 1 次印刷
定　　价/29.00 元

前　言

20世纪以来,由于物理学、生物学、工程技术、农业技术和军事技术发展的推动,概率论飞速发展,理论课题不断扩大与深入,应用范围大大拓宽。在最近几十年中,概率论的方法被引入各个工程技术学科和社会学科。目前,概率论在近代物理、自动控制、地震预报和气象预报、工厂产品质量控制、农业试验和公用事业等方面都得到了重要应用。有越来越多的概率论方法被引入到经济、金融和管理科学,概率论成为有力工具。

现在,概率论已发展成为一门与实际紧密相连的理论严谨的数学科学。它内容丰富,结论深刻,有别开生面的研究课题,有自己独特的概念和方法,已经成为了近代数学一个有特色的分支。

数理统计是伴随着概率论的发展而发展起来的一个数学分支,研究如何有效地收集、整理和分析受随机因素影响的数据,并对所考虑的问题作出推断或预测,为采取某种决策和行动提供依据或建议。当前,数理统计的应用范围愈来愈广泛,已渗透到许多科学领域,应用到国民经济各个部门,成为科学研究不可缺少的工具。

笔者从事概率论与数理统计的教学和科研工作有十余年,和国内同行们进行交流探讨,得出的观点几乎是一致的:这门课程的教学如果完全采用传统的"理论一大堆,公式一大堆"的教学模式,对普通本科学校的学生学习会有较大的难度,多数学生对这门课的感受是抽象、枯燥、难以理解。教师强行"喂",学生被迫"吃";教师"教"得辛苦,学生"学"得也辛苦。

纵观现有文献,概率统计教材内容大同小异,偏理论轻应用,且目前国内没有一本适合普通高等院校学生学习概率统计案例的正规出版物。与此同时,由于概率统计应用的广泛性,越来越多的学子对于概率统计的学习有较高的期待,这也是我写这本书的动力。本书是我在为本科生、研究生讲授课程的讲义的基础上写成的。笔者认为:对于概率统计这门有着自身特色的课程,好的案例讲解是教学

成功的一半。为了让学生更好地理解随机世界的特点，本书从对学生要求的实际出发，以各种实际问题为背景，精选了108个案例。这些案例多数是笔者在多年的教学实践和科研中精心构思、设计的，部分是参考国内外同行的图书、学术论文及部分网站内容(笔者在其中都作了标注，如果有遗漏，请见谅)。通过生动、有趣的案例教学，引出概率统计主要分支的基本概念、基本模型和基本方法，并且侧重各种方法及其应用，让学生深刻理解知识点，掌握概率统计方法在经济学、管理学及其他学科中的重要应用，学习起来也更有激情和针对性。书中的一些案例对概率统计中"全概率公式""贝叶斯公式""概率密度函数""极大似然法"等比较抽象而又非常重要的理论进行了新的阐述和说明，弥补了传统教材的不足。少数标注*的案例可供学有余力的学生参考。

本书顺应教学改革的潮流和CDIO理念，将概率统计理论体系与应用紧密结合，突出案例教学和统计软件操作的重要地位。本书还包括了相关内容的背景介绍、统计人文知识及少量习题等内容。这些素材与正文内容互相补充，为学生提供了丰富的学习材料，有利于培养学生的创新意识和独立思考的能力；在内容上充分考虑到实用性、科学性、先进性和前沿性；让学生在学习的过程中真正体会到学有所用，也有利于学生自主学习。

本书取材广泛，理论联系实际，可作为本、专科生及研究生学习概率论与数理统计、统计学原理、统计建模、高等数理统计等统计系列课程参考用书，也可供有关人员学习和自学参考。

由于编写这样的案例集是首次尝试，经验不足，水平有限，书中难免存在疏漏和不足，恳请同行批评指正。

金　明

2014年6月

目 录

第一部分 概率指标

第二部分 概率分布模型

第三部分 参数估计

第四部分　假设检验

第五部分　统计指标

第六部分 随机实验

第七部分 其他

第一部分

概率指标

案例1

概率论的起源:赌徒案例

我什么也不害怕,也不害怕丢钱,但我害怕不确定性。——索罗斯

【引言】概率统计只说“可能性”是实际世界的真实体现,因为真实世界充满了不确定性,有多少人会享受和陶醉于其未来每一时刻全部已经确定了的世界呢?从某种意义来说:**生活中唯一确定的事情就是其不确定性。**正是不确定性使得生活充满了挑战、魅力和迷人的色彩。

概率问题的历史可以追溯到遥远的过去,很早以前,人们就用抽签、抓阄的方法解决彼此间的争端,这可能是概率最早的应用。文艺复兴时期,在欧洲两个特殊行业的发展促进了概率论这门学科的诞生:**保险和赌博。**

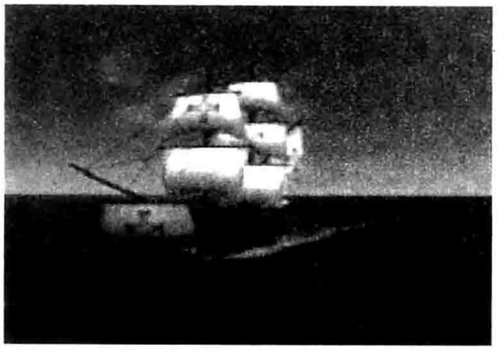

图1 海上保险

图2 赌博

当时的保险业已在欧洲蓬勃发展起来,不过,当时的保险业非常不成熟,只是一种完全靠估计形势而出现的赌博性事业,保险公司要承担很大的不确定性风险(为什么?),保险业的发展渴望能有指导保险的计算工具的出现。

这一技术难题因15世纪末赌博现象的大量出现而逐步得到解决。当时的主要赌博形式有玩纸牌、掷骰子、转铜币等。参加赌博的人,特别是那些专门从事以赢利为生的职业赌徒,鏖战赌场,天长日久就逐渐悟出了一个道理:在少数几次赌博中无法预料到输赢的结果,如果多次进行下去,就可能有所预料,这并不是完全的碰巧。这无意中就给学者们提供了一个比较简单而又非常典型的概率研究模型。

【案例】1654年,法国赌徒梅勒遇到了一个难解的问题:梅勒和他的一个朋友每人出30个金币,两人谁先赢满3局谁就得到全部赌注,梅勒每次赢的概率是1/2。在游戏进行了一会儿后,梅勒赢了2局,他的朋友赢了1局。这时候,梅勒由于一个紧急事情必须离开,游戏不得不停止。他们该如何分配赌桌上的60个金币的赌注呢?试说明理由。

【分析】梅勒的朋友认为,既然他接下来赢的机会是梅勒的一半,那么他该拿到梅勒所得的一半,即他拿20个金币,梅勒拿40个金币。然而梅勒争执道:再掷一次骰子,即使他输了,游戏是平局,他最少也能得到全部赌注的一半——30个金币;但如果他赢了,

并可拿走全部的60个金币。在下一次掷骰子之前,他实际上已经拥有了30个金币,他还有50%的机会赢得另外30个金币,所以,他应分得45个金币。

赌本究竟如何分配才合理呢?

后来梅勒把这个问题告诉了当时法国著名的数学家帕斯卡,这居然也难住了帕斯卡,因为当时并没有相关知识来解决此类问题,而且两人说的似乎都有道理。帕斯卡又写信告诉了另一个著名的数学家费马,于是在这两位伟大的法国数学家之间开始了具有划时代意义的通信,在通信中,他们最终正确地解决了这个问题。

【解决方案】 可以设想:如果继续赌下去,梅勒(设为甲)和他朋友(设为乙)最终获胜的机会如何呢?他们俩至多再赌2局即可分出胜负,这2局有4种可能结果:甲甲、甲乙、乙甲、乙乙。前3种情况都是甲最后取胜,只有最后一种情况才是乙取胜,所以赌注应按3∶1的比例分配,即甲得45个金币,乙15个。虽然梅勒的计算方式不一样,但他的分配方法是对的。

【结论】 三年后,也就是1657年,荷兰著名的天文、物理兼数学家惠更斯把这一问题置于更复杂的情形下,试图总结出更一般的规律,结果写成了《论掷骰子游戏中的计算》一书,这就是最早的概率论著作。正是他们把这一类问题提高到了理论的高度,并总结出了其中的一般规律。同时,他们的研究还吸引了许多学者,由此把赌博的数理讨论推向了一个新的台阶,逐渐建立起一些重要概念及运算法则,从而使这类研究从对机会性游戏的分析发展上升为一个新的数学分支。

由赌徒的问题引起,一直到20世纪初,概率论逐渐演变成一门**严谨的科学**。

【点评】 赌博和保险两个行业促进了概率论的诞生,也揭示了概率论这门学科并非空中楼阁,而是来源于**人类的实践**,"应用"是它的精髓。

【思考题】 利用概率的思想,解决下列问题:

1. 在一副扑克牌中任意抽出一张牌,这张牌是大王的可能性大还是红桃的可能性大?
2. 小明任意买一张电影票,座位号是2的倍数与座位号是5的倍数的可能性哪个大?
3. 在一个班任意找2名学生,他们是同年出生的和同月出生的哪一种可能性较大?
4. 能否用6个球设计一个摸球游戏,使得摸到黄球的可能性比摸到红球的可能性大?
5. 把赌徒案例进行推广:赌金是100法币,先赢 m 局的获胜,甲每次获胜的概率是1/2,已知甲、乙分别赢 k、t 局($k,t<m$),则赌金100法币如何分配?

【统计名人】

克里斯蒂安·惠更斯(Christiaan Huygens,1629~1695)荷兰物理学家、天文学家、数学家。在概率论学科上的贡献是:他的《机遇的规律》把帕斯卡和费马讨论的赌博问题从数理逻辑上推向了一个新的台阶,建立起一些重要概念及运算法则,是概率论的开创者之一。

案例2

概率的应用:警察断案

我是一个精炼的赌博者。——Girolano Cardano

【引言】长期以来,由于在中学数学教育中对概率统计内容的忽视,人们认为数学严谨,逻辑推理严格,故深受“数学只能研究确定性现象”这种思维的影响(即 1+1=2 确定性的思维模式),这种确定性的思维模式使得人们对于随机现象方面的理解很不习惯,或对于概率的理解仅仅停留在“生儿生女”和“买彩票中 500 万”等简单的现象上。

其实我们生活的世界就是一个随机的世界,无论自然界或社会学领域,我们面临的绝大多数现象都是随机现象,它们构成了一个丰富多彩的世界。无论我们学习什么专业,如经济学、管理学、统计学、通信、气象学、计算机科学、电子工程、法律、人文、历史等,我们处理的绝大多数问题也都是随机问题。所以,我们要研究“随机性的规律”。

图 1　天气

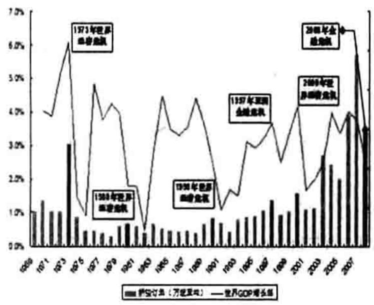

图 2　经济危机

图 3　人的一生

现在,概率统计内容的学习将进入一个全面普及的阶段。我们逐渐认识到统计学可以研究一些偶然现象后面的必然规律性。

【案例】深夜,成都市成华区一辆出租车被牵涉进一起交通事故,该市有两家出租车公司:红色出租车公司和蓝色出租车公司,其中蓝色出租车公司和红色出租车公司分别占整个城市出租车的 85%和 15%,据现场目击证人说,事故现场的出租车是红色的,对证人的辨别能力作了测试,测得他辨认颜色的正确率为 80%,于是警察就认定红色出租车具有较大的肇事嫌疑。请问警察的认定对红色出租车公平吗? 试说明理由。

【分析】该案例考查**概率**这个指标在实际生活中的功能:推断。

【解决方案】根据案例中相关数据分析,可得到下页表。

从表中可以看出,当证人说出租车是红色时,且经确认是红色的概率为

$$P_1=120/290=0.41$$

而是蓝色的概率为

$$P_2=170/290=0.59$$

图 4　断案需要概率?

表 1　调查数据分析表

		证人所说的颜色(正确率 80%)		
		蓝色	红色	合计
真实颜色	蓝色(85%)	680	170	850
	红色(15%)	30	120	150
	合计	710	290	1000

【结论】在这种情况下，以证人的证词作为推断的依据，对红色出租车显然是不公平的，做出错误结论的原因是对证人的辨认率的独立认识的基础做出的。

【点评】什么是“概率思维”? 概率是关于“可能性”的数学，关于“机会”的数学。概率思维是与确定性思维是相对应的一个概念。我们生活在一个并不“总是”和“一定”的世界里，这个世界更多地充满了“也许”和“可能”。当社会变得越来越复杂时，我们就更需要所谓的“概率思维”。

该案例告诉我们：在现实生活中一些人很相信自己的“直觉”，直觉有时候也没有错，但直觉往往是不可靠的，不严谨的。

【思考题】

1. 苗族风俗：苗族姑娘把 6 根草握在手掌中，只露出出其头与尾。然后请她的男友把 6 根头两两连接，6 根尾也两两连接。姑娘放手后，如果六根草恰好连成一个环的话，苗族姑娘就嫁给小伙。计算姑娘嫁给小伙的概率。

2. 在 17 世纪，意大利专业赌徒们都认为：三颗骰子掷出的点数之和为 9 的概率与为 10 的概率是相等的，真的吗?

【统计名人】

卡尔达诺(Girolamo Cardano，1501～1576 年)，意大利文艺复兴时期百科全书式的学者，他的职业是数学家、概率计算的先行者、医生、赌徒。

卡尔达诺死后发表的《论赌博游戏》一书被认为是第一部概率论著作，大约完成于 1526 年，这一著作是根据他多年的赌博经验和数学理论写成的，探讨了有关机遇的数学理论。弗洛伦斯 · 戴维把该著作称作是“理论数学家和赌徒的结合体”。

他对现代概率论有开创之功，为帕斯卡(1623～1662)和费尔马(1601～1665)全面展开这个课题的研究揭开了序幕。

案例3

概率的应用:洛杉矶抢劫案

统计学是人类 20 世纪取得突出成就的学科之一,统计学的普遍存在以及在开拓新知识领域方面的应用已远远超过 20 世纪内任何技术或科学发明。——C. R. Rao

【引言】 历史上最早、最著名的涉及争议的案例要数 1964 年夏天发生在美国洛杉矶的一起抢劫案。

【案例】 一天中午,在美国洛杉矶,一位老妇人从杂货店买了东西推着小车回家,途经一条小巷时,突然被一位冲过来的年轻女子推倒,等老妇人醒过神来,发现自己身上的钱包已被偷走,女贼也跑了很远。虽然老妇人没有看清罪犯是什么样子,可小巷周围的不少住户都曾与这位女子擦肩而过,并且看到她在街头跳上一辆车逃离现场。

【分析】 后来警方根据目击者描述的犯罪者特征,几天后在附近逮捕了一对夫妻(Malcolm Collins 和 Janet Collins. Malcolm)。可是在法庭上,目击者中并没有人能够清晰地指认出罪犯,因此检方很难将 2 人定罪。于是检察官们想出了一个“新颖的办法”,他们把目击证人说出的几条主要特征列了出来,并且根据洛杉矶地区的数据估算了这些特征会出现的概率:

黄色的汽车:1/10

嘴上面有短胡子的男性:1/4

络腮胡子的黑人:1/10

马尾辫女孩:1/10

金发女孩:1/3

汽车中有肤色不同的夫妻俩:1/1000

图 1 洛杉矶犯罪率

检察官找来一位“数学专业人士”,计算了在整个洛杉矶地区符合上述各条特征的夫妇存在的概率,这位“数学专业人士”认为最后的概率应该是 6 个概率值乘到一起,结果就是 1/12000000。检察官据此告知评审团,如此小的概率很难发生,附近地区很难再找到另外一对 6 项特征全部符合的夫妇,所以这对嫌疑人一定是罪犯。陪审团最终采纳了检方的意见,判定这对夫妇抢劫罪成立。

【解决方案】 可是后来加州高等法院驳回了这个判决,他们认为检方使用的概率作为证据的方式是错误的。

首先,概率乘法公式 $P(ABC)=P(A)P(B)P(C)$ 一定要在 A、B、C 都是独立事件的时候才成立,可是目击者提供的那些特征并不相互独立,比如留八字胡的男性和留络腮胡的男性这两项,“男性”这个信息是重叠的,而喜欢留胡子的人往往两个位置都会留胡子,两个特征高度关联,同时发生的概率远远大于两个数字相乘。马尾辫女孩和短发女孩也是同样的道理。这样的话,正确的概率可能会是 1/12000000 的很多倍,并没有那

么小。

退一步说，假定概率真的是 1/12000000，以案发附近地区有 400 万人算，至少有 2 位夫妇符合目击者全部特征的概率是$(1-11999999/12000000)\times4000000$，超过 30%，也就是说，仅仅根据 1/12000000 的概率就判定这对夫妇是罪犯没有道理。

【点评】概率是对随机现象规律的研究。而随机现象是普遍的，因此，我们只有正确地理解概率，才能对日常生活中的随机现象做出科学的解释，指导我们的实践活动。

要注意联系生活，学会科学地澄清日常生活中一些错误认识。

【思考题】《甄嬛传》中甄嬛在被贬时和果郡王生了双胞胎，后因人告密，多疑的雍正皇上要滴血认亲，在皇帝和皇后（甄嬛死对头）及众多妃嫔的监督下，滴血过程很严谨，却依然没有发现甄嬛给皇帝戴"绿帽子"的事实（可见至高无上的皇帝真是可怜人）。甄嬛继续深得皇帝宠爱，并在端妃、沈贵人、崔槿汐等几位好姐妹的帮助下，斗倒了华妃、安嫔、皇后，害死了皇上，最终坐上了皇太后的宝座。

是滴血认亲检测方法有问题？还是其他问题？在当时的技术条件下从概率的角度应如何设计检测流程来提高滴血认亲的可信度？

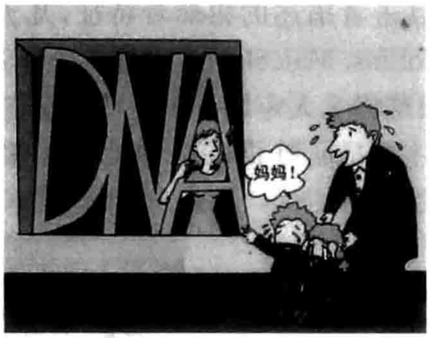

图 2　现代人的 DNA 鉴定

【统计趣谈】

DNA 鉴定一直被视为鉴定亲子关系的确凿证据，只要两份 DNA 样品相互匹配，即可确认二者之间的血缘关系。在一些 DNA 专家看来，DNA 鉴定的误差只有十亿分之一。

但是也有反例，《大自然探索》就有一个报导：安德鲁·克莱基是美国华盛顿的一位名人发型师，由于美国前第一夫人劳拉·布什的惠顾，他的事业一度如日中天。一位女士找上门来告诉他说"你有个孩子"时，克莱基认为不可能，因为在孩子出生几年前他就已经停止和这位女士交往了。他相信 DNA 鉴定不会说谎，所以毫不犹豫地向 LabCorp（美国最大的亲子鉴定中心之一）提交了细胞样本，但是鉴定报告显示：他是这名孩子亲生父亲的可能性高达 99.99%。

克莱基通过关系找到了联邦调查局 DNA 分析部，在经过了两个星期的分析后，联邦调查局 DNA 分析部终于给出了新的鉴定报告，指出克莱基并非那名孩子的父亲。法院终审裁决克莱基胜诉。