重庆市市级优质课程配套教材

应用数理统计

钟 波 刘琼荪 刘朝林
荣腾中 黄光辉 编著

科 学 出 版 社

重庆市市级优质课程配套教材

应用数理统计

钟 波 刘琼荪 刘朝林 荣腾中 黄光辉 编著

科 学 出 版 社
北 京

内 容 简 介

本书是重庆市市级优质课程配套教材，是重庆市研究生教育教学改革重大项目成果，全书共 7 章，分别是概率论基础及应用、数理统计基础、参数估计、假设检验、回归分析、方差分析与正交设计、多元统计分析. 本书每章均配有应用案例、章节总结、应用分享和习题，便于教师教学和学生自学.

本书可以作为高等院校工科专业研究生的数理统计教材，也可以作为相关专业研究生的参考资料.

图书在版编目（CIP）数据

应用数理统计/钟波等编著. —北京：科学出版社，2017.9
重庆市市级优质课程配套教材
ISBN 978-7-03-054214-4

Ⅰ. ①应… Ⅱ. ①钟… Ⅲ. ①数理统计-研究生-教材 Ⅳ. ①O212

中国版本图书馆 CIP 数据核字(2017)第 201810 号

责任编辑：任俊红 孙翠勤 / 责任校对：桂伟利
责任印制：霍 兵 / 封面设计：华路天然工作室

科 学 出 版 社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com
北京市密东印刷有限公司 印刷
科学出版社发行 各地新华书店经销
*
2017 年 9 月第 一 版 开本：787×1092 1/16
2017 年 9 月第一次印刷 印张：22
字数：522 000

定价：49.00 元

（如有印装质量问题，我社负责调换）

前　言

在科学研究和实际生产的各个领域中，需要对大量的数据进行处理和分析，尤其是随着科技的快速发展，数据在规模和类型上都发生了巨大变化，亟待工程技术人员等实际工作者对海量数据、大数据进行分析与挖掘.

数理统计是一门研究随机现象及其统计规律的数学学科，它以概率论为基础，研究如何获取随机现象的观测数据、如何根据观测数据推断随机现象的统计规律. 数理统计不仅为人们认识随机现象提供了有效的方法，而且为人们分析和挖掘数据集合的结构、特征、数据间的关系等提供了丰富的思想和技术.

重庆大学研究生数理统计课程是重庆市市级优质课程，有着丰富的教学成果. 本书结合作者多年积累的教学实践和教材编写经验，参考国内外相关优秀教材，以工科专业硕士研究生的培养目标为指导，在重庆市研究生教育教学改革重大研究项目——“加强专业硕士数理统计课程建设，提高学生应用能力”的支持下编写而成. 编写宗旨是：突出数理统计的应用性和实践性，重视统计思想的理解和统计方法在实际中的灵活应用，把现代科技与经济管理中的统计应用成果引入教材，提高读者的学习兴趣.

本书有如下特色.

(1) 内容浅显易懂. 结合实例，详细、透彻地介绍了数理统计的基本理论与方法，深入浅出地阐述了统计方法蕴含的统计思想与原理.

(2) 强调应用性，注意与统计软件结合. 书中引例、例题、应用案例、习题大量地选择了来源于工程管理、医学卫生、生物、信息、社会与经济等各个学科领域的实际问题. 各章的“应用案例”小节从实际问题出发，以真实数据为例，结合 Excel 或 R 等统计软件，呈现了“问题描述→统计模型建立→统计方法选择→软件求解（模拟）→结果分析”这种解决实际问题的全过程. 每章的“应用分享”小节呈现了统计在现代社会、科技以及工程管理等领域中的实际应用.各章习题中的 B 组题主要用于统计应用能力的训练.

(3) 强调实践性，注重学以致用. 各章设置了“结合你的专业背景研讨问题”的练习题，即要求读者结合自己的专业领域提出统计问题，收集数据，并解决问题. 引导学生勇于实践，用已学到的知识解决自己身边的问题.

(4) 引入随机模拟方法，以增强统计应用与研究能力. 随机模拟是进行计算机仿真、统计实践的一种重要而又方便、实用的现代统计方法. 此方法的引入，可以帮助读者更好地理解数理统计的基本理论，更好地用统计实验方法研讨一些较困难的问题.

(5) 注重研究性思维和自学能力的培养. 在习题中，编制了讨论题和查阅资料自学新知识的练习题. 这部分习题是对教材内容进一步的思考或扩展，以及新知识、统计应用新概念的学习.

本书主要面向高等院校工科专业硕士研究生，所需的预备知识是微积分和线性代数的基础知识.

本书共七章，具体分工是：第 1 章由荣腾中编写，第 2 章由刘朝林编写，第 3、4 章由钟波

编写, 第 5、6 章由刘琼荪编写, 第 7 章由黄光辉编写. 全书校对工作由潘明勇完成.

书中许多例题、习题是从大量参考文献激发的思想中产生的, 在此对参考文献的作者深表感谢!

本书的编写得到了学校各级领导和数理统计课程组教师的积极支持; 同时, 也得到了数学与统计学院部分研究生和本科生的热诚帮助, 在此对他们表示真诚的感谢!

由于编者水平有限, 文中难免不当之处, 恳请广大读者不吝赐教.

编　者

2017 年 5 月 30 日

于重庆大学数学与统计学院

目　　录

第 1 章　概率论基础及应用

数学被人们尊崇为自然科学的皇后. 长期以来, 在传统的数学王国里, 结论都是确定的. 5 个原始人均分 15 只猎物, 每人可分 3 只; 古巴比伦王国的土地分配, 无论是三角形还是平行四边形, 只关心土地面积是否相同; 直角三角形三边必定满足“勾股定理”; 祖冲之精确地计算了圆的周长与直径的比例常数π; 曲线长度可以用微积分准确计算. 然而, 现实生活并非如理想的数学模型一般, 常存在着随机因素, 使得用纯粹的数学来度量不够精确. 要更深刻地认识客观世界, 必须对这些随机因素进行探究. 在随机领域里, 结果是不确定的、随机的. 比如, 掷两颗骰子点数之和等于多少? 结果有很多, 一次抛掷两颗骰子出现哪一个结果具有随机性. 但如果能发现随机现象(random phenomenon)的客观规律性, 即在大量的观测、试验中随机现象表现出的统计规律, 就可以利用这些规律为生产生活服务.

随机数学(random mathematics)是研究随机现象及其统计规律性的一门数学学科, 它包含概率论(probability theory)、数理统计(mathematical statistics)、随机过程(stochastic processes)、试验设计(test design)、抽样调查(sample survey)、随机分析(stochastic analysis)等许多分支. 其中, 概率论与数理统计是随机数学的基础部分.

概率论给出了描述随机现象及其统计规律性的数学方法和模型, 从理论上研究随机事件发生的概率及性质; 数理统计运用概率论的理论与方法, 通过研究如何有效地收集、整理和分析随机现象的观测数据, 实现对随机现象及其统计规律性的认识.

概率论起源于 17 世纪中叶. 当时, 在误差分析、人口统计等范畴中有大量的随机数据资料需要整理和研究, 从而孕育出一门数学分支专门用于研究随机现象的规律性. 同时, 由于欧洲贵族赌博的兴起, 在利益的驱动下, 越来越多的精英投入到“可能性”的研究上来. 渐渐地, 人们不再满足于诸如“掷硬币赌正反面”这样简单的赌博方式, 希望增加赌局复杂度和趣味性, 把“可能性”进行“加密”, 让有概率运算能力的人靠技术拥有更多的赌博优势. 比如, 庄家掷骰子一次, 抛掷前下注三个点; 或是庄家掷骰子三次, 抛掷前下注一个点. 这两种赌博方式, 可能性是不是一样的呢? 应用古典概率计算方法, 可推断第一种方式对赌徒有利.

使概率论成为数学的一个分支, 并做出重要贡献的是瑞士数学家雅各布 · 伯努利(Jakob Bernoulli), 他于 1689 年建立了概率论中的第一个极限定理——“伯努利大数定律”. 法国数学家拉普拉斯(Laplace)于 1812 年出版了《概率的分析理论》, 首先明确地给出了概率的古典定义. 后来, 经过高斯(Gauss)和泊松(Poisson)等数学家的努力, 概率论在数学中的地位基本确立. 到了 20 世纪 30 年代, 数学家柯尔莫哥洛夫(Kolmogorov)提出了概率的公理化定义, 为概率论的发展做出了杰出贡献, 使概率论成为一门严谨的数学分支.

概率论作为理论严谨、应用广泛的数学学科日益受到人们的重视, 并随着社会和科学技术的发展而发展. 21 世纪以来, 概率论的理论与方法已渗透到各个学科和领域. 例如:

◆应用概率论的分析方法对通信数据、电子商务数据、金融数据等建立概率模型, 推断随机规律、进行科学预测.

◆建立在概率论基础上的随机化优化算法、概率结构化过程是计算机科学的基本工具,

其应用涉及组合优化、机器学习、通信网络以及安全协议等诸多领域. 越来越高级、越来越复杂的概率技术已用于更加广泛和更富有挑战性的问题.

◆基于概率论的可靠性数学, 广泛应用于航空航天、电力系统、核能系统、通信系统、软件系统、桥梁系统等的工程可靠性研究.

◆以概率论为基础的随机复杂网络理论, 在基因动态调控模型、干细胞等重要生物体的演化、计算机病毒在互联网或邮件网络中的传播、黑客对计算机网站攻击的演化, 以及禽流感、艾滋病等恶性传染病在人群构成的复杂网络中的传播、信用风险与非法资金在金融机构形成的复杂网络中的传播与扩散等方面, 有着深入的应用.

◆概率论是保险精算学科中保险模型、破产理论、分红理论、风险分析以及决策与风险控制理论的基础, 是保险公司对其风险进行定量分析和预测, 并根据这些结果进行管理与控制风险的基础理论.

诸如此类的例子, 举不胜举. 总之, 科学研究已经将随机性视为建模和分析中的基本组成部分.

概率论是数理统计的理论基础, 为了便于更好地学习数理统计知识, 本章对概率论的基础知识及其应用进行了概述.

1.1　随机事件及概率

1.1.1　随机事件

在一定条件下可能发生也可能不发生的现象称为随机现象或随机事件, 简称为事件(random event), 常用大写字母 A, B, C 等或带下标的 $A_i, B_j(i=1, 2, \cdots; j=1, 2, \cdots)$等表示. 在一定条件下必然发生的现象称为必然事件, 用Ω表示. 在一定条件下不可能发生的现象称为不可能事件, 用$\varnothing$表示. 在概率论中, 随机事件一定是某个随机试验 E 中的随机现象. 为了运用“集合”工具研究随机事件, 需要引入“样本点”和“样本空间”的概念. 将随机试验 E 中的基本事件称为样本点, 一般用小写字母ω, e 等或数字表示, 由所有样本点组成的集合称为E的样本空间, 记为Ω. 随机试验E中的任何一个随机事件A, 都有一个样本空间Ω的子集与之对应, 该子集具有这样的性质: 随机事件 A 发生当且仅当子集中的某一样本点发生. 因此, 任何一个随机事件都可以用样本空间的一个子集表示. 这样, 随机事件的问题可以转化为“集合”的问题.

运用集合论的知识, 可以方便地表示子事件 $A\subset B$、两事件相等 $A=B$、和事件 $A\cup B$、积事件 AB(或 $A\cap B$)、差事件 $A-B$、互斥事件 $AB=\varnothing$、逆事件 $\overline{A}$、完备事件组 $A_1, A_2, \cdots, A_n$ $\left(\text{两两互斥且}\bigcup\limits_{i=1}^{n}A_i=\Omega\right)$等事件的关系与运算, 并且得到下列事件的运算规律:

(1) 交换律: $A\cup B=B\cup A, AB=BA$.

(2) 结合律: $(A\cup B)\cup C=A\cup(B\cup C)$, $(AB)C=A(BC)$.

(3) 分配律: $(A\cup B)\cap C=AC\cup BC$, $(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$.

(4) 德摩根(De Morgan)定理(对偶律): $\overline{\bigcup\limits_{i=1}^{n}A_i}=\bigcap\limits_{i=1}^{n}\overline{A}_i$, $\overline{\bigcap\limits_{i=1}^{n}A_i}=\bigcup\limits_{i=1}^{n}\overline{A}_i$.

(5) 差化积: $A-B=A\overline{B}$.

(6) 吸收律: 如果 $A\subset B$, 则 $A\cup B=B, AB=A$.

1.1.2　概率的定义及性质

直观来讲, 概率(probability)是对事件在一次随机试验中发生可能性大小的一种数量度量, 介于 0 与 1 之间, 概率越接近数值 1, 则表明事件发生的可能性越大. 在一次随机试验中, 事件 A 的发生具有偶然性, 但在多次重复试验中, 事件 A 的发生却呈现出一定的规律性, 称这种规律性为统计规律, 即随着重复试验次数 n 的增加, 事件 A 发生的频率 $f_n(A)=\dfrac{r}{n}$ 会稳定在某一常数附近, 该常数称为频率的稳定值, 也就是事件 A 的概率 $P(A)$.

例 1.1.1　在足球比赛中, 罚点球是一个扣人心弦的场面. 若用 A 表示"罚点球射中球门"的事件, 那么事件 A 的概率是多少呢? 这可以用通过重复试验所得的数据资料计算出的频率估计概率 $P(A)$. 曾经有人对 1930～1988 年世界各地 53274 场重大足球比赛作了统计, 在判罚的 15382 个点球中, 有 11172 个射中球门, 其频率为 $11172/15382\approx 0.726$, 这就是概率 $P(A)$ 的一个估计值.

理论上 $P(A)$ 是存在的, 但在现实世界里人们无法将一个试验无限次地重复下去. 因此, 探寻事件的概率, 还不能完全依赖于做大量的重复试验, 需要进行理论上的研究、推断和假设. 比如, 古典概率、几何概率、概率的公理化定义.

1. 古典概率

假设随机试验 E 的样本空间 Ω 满足: ①Ω 是有限集合; ②Ω 中基本事件发生的可能性相等, 则随机试验 E 的任意事件 A 的概率:

$$P(A)=\frac{r}{n} \tag{1.1.1}$$

称由上述公式定义的概率为古典概率. 其中: n 表示 Ω 中基本事件的总数, r 表示事件 A 包含的基本事件数.

古典概率具有以下三个基本性质:

(1) 非负性: 对任意事件 A, 有 $0\leqslant P(A)\leqslant 1$;

(2) 规范性: $P(\Omega)=1$;

(3) 有限可加性: 设事件 $A_1, A_2, \cdots, A_m$ 为两两互斥的事件, 则 $P\left(\bigcup\limits_{i=1}^{m}A_i\right)=\sum\limits_{i=1}^{m}P(A_i)$.

2. 几何概率

当样本空间 Ω 是某个可度量的区域, 且基本事件在区域中是均匀等可能的. 则由区域 Ω 中的某个子区域 A 构成的随机事件的概率为

$$P(A)=\frac{A\text{的几何测度}}{\Omega\text{的几何测度}} \tag{1.1.2}$$

称由上述公式定义的概率为几何概率. 其中, 当区域 Ω 为一维、二维或三维区域时, 对应的几何测度分别为长度、面积及体积.

几何概率具有下列性质:

(1) 非负性: 对任意事件 A, 有 $0 \leqslant P(A) \leqslant 1$;

(2) 规范性: $P(\Omega)=1$;

(3) 可列可加性: 设事件 $A_1, A_2, \cdots, A_n, \cdots$ 为两两互斥的可列个事件, 则 $P\left(\bigcup_{i=1}^{+\infty} A_i\right)=\sum_{i=1}^{+\infty} P(A_i)$.

3. 概率的公理化定义

概率的公理化定义, 以一些不加证明而承认的前提为概率公理, 这些公理规定了概率理论的一些基本关系和所满足的条件, 为复杂随机事件的概率分析以及深入的理论推演奠定了基础.

1933 年, 苏联数学家柯尔莫哥洛夫提出了下列概率的公理化定义.

定义 1.1.1　设Ω是随机试验E的样本空间, F是Ω中事件组成的集合, 称为事件域, P是定义在F上的实值集合函数(测度), 即$P: F \mapsto [0,1]$. 如果P满足:

公理 1 (非负性)　$P(A) \geqslant 0$;

公理 2 (规范性)　$P(\Omega)=1$;

公理 3 (可列可加性)　$P\left(\bigcup_{i=1}^{+\infty} A_i\right)=\sum_{i=1}^{+\infty} P(A_i)$, 其中事件 $A_1, A_2, \cdots, A_n, \cdots$ 两两互斥.

则称 $P(A)$ 为事件 A 的概率.

在概率论体系里, Ω, F 和概率 P 三元素就构成了概率的全部, 称 (Ω, F, P) 为概率空间(probability space).

可以验证, 古典概率和几何概率皆满足上述概率的公理化定义.

利用概率的公理化定义可以导出概率的一系列性质.

性质 1.1.1　(1) 不可能事件的概率为零, 即 $P(\varnothing)=0$;

(2) 有限可加性: 设 $A_1, A_2, \cdots, A_n$ 为两两互斥的 n 个事件, 则

$$P\left(\sum_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i) \tag{1.1.3}$$

(3) 对任意一个事件 A, $P(\bar{A})=1-P(A)$;

(4) 减法公式: $P(A-B)=P(A)-P(AB)$;

(5) 单调性: 若 $A \subset B$, 则 $P(A) \leqslant P(B)$;

(6) 有界性: 对任意一个事件 A, $0 \leqslant P(A) \leqslant 1$;

(7) 加法公式: $P(A \cup B)=P(A)+P(B)-P(AB)$, 更一般的加法公式为

$$P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i)-\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_i A_j)+\cdots+(-1)^{n-1} P(A_1 A_2 \cdots A_n) \tag{1.1.4}$$

4. 条件概率

条件概率 $P(B \mid A)$ 描述的是在事件 A 已经发生的情况下事件 B 发生的概率, 可以理解为是对 $P(B)$ 的修正或在掌握了新信息 A 时对事件 B 发生概率的新认识. 当 $P(A)>0$ 时, 有下面的条件概率公式:

$$P(B \mid A)=\frac{P(AB)}{P(A)} \tag{1.1.5}$$

条件概率有如下性质.

性质 1.1.2 (1) 非负性: 对任意事件 B, 有 $P(B \mid A) \geqslant 0$;

(2) 规范性: $P(\Omega \mid A)=1$;

(3) 可列可加性: 若 $B_1, B_2, \cdots, B_n, \cdots$ 是两两互斥的随机事件, 则

$$P\left(\sum_{i=1}^{+\infty} B_i \mid A\right)=\sum_{i=1}^{+\infty} P(B_i \mid A)$$

(4) 对任意事件 B, 有 $P(\bar{B} \mid A)=1-P(B \mid A)$;

(5) 加法公式: $P(B_1 \cup B_2 \mid A)=P(B_1 \mid A)+P(B_2 \mid A)-P(B_1 B_2 \mid A)$.

将(1.1.5)式变形, 得到如下的乘法公式.

性质 1.1.3 (乘法公式) 设 A 与 B 是两事件, 且 $P(A)>0$, $P(B)>0$, 则

$$P(AB)=P(B)P(A \mid B) \tag{1.1.6}$$

$$P(AB)=P(A)P(B \mid A) \tag{1.1.7}$$

由乘法公式, 可导出下列两个著名的公式, 即全概率公式和贝叶斯(Bayes)公式.

性质 1.1.4 (全概率公式与贝叶斯公式) 设 $A_1, A_2, \cdots, A_n$ 是两两互斥的事件组, $P(A_i)>0$, $i=1,2,\cdots,n$, 且 $B \subset \sum_{i=1}^{n} A_i$, 则

$$P(B)=\sum_{i=1}^{n} P(B \mid A_i) P(A_i) \tag{1.1.8}$$

$$P(A_i \mid B)=\frac{P(A_i B)}{P(B)}=\frac{P(A_i) P(B \mid A_i)}{\sum_{j=1}^{n} P(A_j) P(B \mid A_j)}, \quad i=1,2,\cdots,n \tag{1.1.9}$$

称(1.1.8)式为全概率公式, (1.1.9)式为贝叶斯公式.

条件 $B \subset \sum_{i=1}^{n} A_i$ 表示: $A_1, A_2, \cdots, A_n$ 是与事件 B 有关的全部可能的、两两互斥的事件. 因此, B 发生时一定是 $A_1 B, A_2 B, \cdots, A_n B$ 中某个事件发生, 事件 B 的概率 $P(B)$ 就是全部事件 $A_1 B, A_2 B, \cdots, A_n B$ 的概率和. 全概率公式可以用图 1.1.1 所示的网络图形表示.

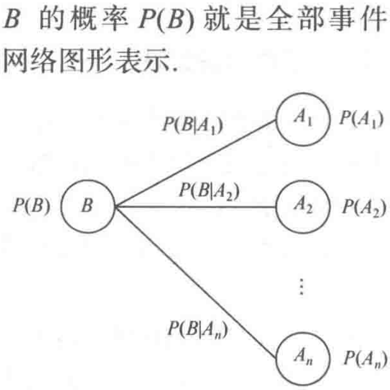

图 1.1.1　全概率公式的示意图

在图 1.1.1 中, 每条线的终端分别对应事件 $A_1, A_2, \cdots, A_n$ 及其概率, 表示与事件 B 有关的全部事件, 每条线上有一个条件概率 $P(B \mid A_i), i=1,2,\cdots,n$, 表示在 $A_i (i=1,2,\cdots,n)$ 发生的条件下, B 发生的可能性. 于是, 事件 B 发生的概率 $P(B)$ 就是各条线上的条件概率与对应终端概率乘积的全部和. 在实际应用中, 常常将事件 $A_1, A_2, \cdots, A_n$ 理解为引起事件 B 发生的各种"原因", 事件 B 理解为这些"原因"导致的"结果". 从这个角度讲, 全概率公式表示: "结果 B"发生的概率由各个"原因"发生的概率 $P(A_i) (i=1,2,\cdots,n)$ 和"原因"作为条件时　"结果 B"发生的条件概率

$P(B \mid A_i), i=1,2,\cdots,n$ 确定，是一种“由因推果”的概率计算．而贝叶斯公式计算的是“结果 B”已经发生的条件下各个“原因”发生的概率，即条件概率，是一种“由果推因”的概率计算．

在实际应用中，常把概率 $P(A_1), P(A_2), \cdots, P(A_n)$ 称为先验概率，一般可以通过经验获得；而 $P(A_i \mid B)$ 称为后验概率，是在知道 B 发生的信息下对事件 $A_1, A_2, \cdots, A_n$ 发生可能性大小的重新认识，也是利用信息 B 对先验概率 $P(A_i)$ 的一种修正．所以，贝叶斯公式也称为后验概率计算公式．

贝叶斯公式由英国数学家贝叶斯(Thomas Bayes)在 18 世纪提出．贝叶斯公式实现了先验信息到后验信息的转化，使人们对随机事件发生的概率有了更加客观的认识方法．贝叶斯公式在统计推断领域里有着重要的影响，以贝叶斯公式为基础形成的贝叶斯统计(Beyes statistics)是当今应用非常广泛的一个统计分支，尤其在数据挖掘、机器学习、机器翻译、模式识别、故障诊断等领域已得到深入的应用．

例 1.1.2　某个人欲向银行贷款，银行需要考察他的诚信度．如果此人首次贷款，那么银行会考察他的现状，如他的财产情况、工作状况等，利用这些信息可以估计他的诚信度，即按时还款可能性．如果此人曾经有一次贷款没有按期还款，那么他的诚信度肯定会下降．会下降多少呢？下面借助贝叶斯公式定量地估计他的诚信度下降的程度．

设事件 A 表示某人的“诚信”，且诚信度为 80%，即 $P(A)=0.8$．事件 B 表示此人“曾经有一次贷款没有按期还款”．一般地，在诚信度较高的情况下，曾经有一次贷款没有按期还款的可能性较小，而诚信度低时，曾经有一次贷款为按期还款的可能性较大．不妨设为

$$P(B \mid A)=0.1, \quad P(B \mid \bar{A})=0.7$$

当事件 B 已经发生了，那么，重新审视此人的诚信，就是对他诚信度的后验认识．由贝叶斯公式计算得

$$\begin{aligned} P(A \mid B) &= \frac{P(A)P(B \mid A)}{P(A)P(B \mid A)+P(\bar{A})P(B \mid \bar{A})} \\ &= \frac{0.8 \times 0.1}{0.8 \times 0.1+0.2 \times 0.7} \approx 0.3636 \end{aligned}$$

计算结果表明，一旦他曾经有一次贷款没有按期还款，那么他的诚信度将从 80%大幅度下降到 36.36%．

1.1.3　事件的独立性

由于事件的发生会对概率产生影响，故在一般情况下，$P(A \mid B) \neq P(A)$．但当事件 B 对事件 A 的概率没有影响时，即 $P(A \mid B)=P(A)$，此时称事件 A 与 B 相互独立．事件之间的独立关系常用积事件的概率来定义．

定义 1.1.2　设 A，B 是概率空间 (Ω, F, P) 的事件，若

$$P(AB)=P(A)P(B) \tag{1.1.10}$$

则称事件 A 与 B 相互独立(independent)．

设 $A_1, A_2, \cdots, A_n$ 为 n 个随机事件，如果对其中任意 $s(2 \leqslant s \leqslant n)$ 个事件 $A_{k_1}, A_{k_2}, \cdots, A_{k_s}$，均有

$$P(A_{k_1}A_{k_2}\cdots A_{k_s})=P(A_{k_1})P(A_{k_2})\cdots P(A_{k_s}) \tag{1.1.11}$$

则称事件 $A_1, A_2, \cdots, A_n$ 相互独立．

性质 1.1.5 (事件独立的性质)　(1) 若 $P(A)>0, P(B)>0$，则事件 A与B 相互独立的充要条件是

$$P(A|B)=P(A) \quad 或 \quad P(B|A)=P(B)$$

(2) 事件 A与B，A与$\overline{B}$，$\overline{A}$与B，$\overline{A}$与$\overline{B}$ 之间的独立性同时成立.

(3) 必然事件Ω与任意事件 A 相互独立, 不可能事件$\varnothing$与任意事件 A 相互独立.

(4) 若事件 $A_1, A_2, \cdots, A_n$ 相互独立, 则将其中任一部分事件改为对立事件, 所得事件组仍为相互独立的事件.

(5) 若事件 $A_1, A_2, \cdots, A_n$ 相互独立, 将 $A_1, A_2, \cdots, A_n$ 任意分成多组, 并对组内的事件施行"并、交、差、补" 运算后所得事件之间也相互独立.

有了事件的独立性概念就可以定义 n 重伯努利试验, 并得到重要的二项概率公式.

定义 1.1.3　设 E 为随机试验, 若每次试验要么事件 A 发生, 要么事件 $\overline{A}$ 发生, 则称该试验为伯努利试验. 将伯努利试验独立重复 n 次, 则称这 n 次试验为 n 重伯努利试验.

性质 1.1.6 (二项概率(binomial probability)公式)　在 n 重伯努利试验中, 设事件 A 每次发生的概率为 p，记 n 次试验中事件 A 发生 k 次的概率为 $P_n(k)$，则

$$P_n(k)=\mathrm{C}_n^k p^k (1-p)^{n-k}, \quad k=0,1,\cdots,n \tag{1.1.12}$$

例 1.1.3　某厂需要到外地采购 12 只集成电路装配仪表. 已知所需集成电路的不合格率为 0.1, 问要采购多少只, 才能以 99%的概率保证完成仪表装配.

解　要完成仪表装配, 即要求采购的集成电路中合格品不少于 12 只. 仅采购 12 只时, 其全部都合格的概率仅为 $0.9^{12} \approx 0.2824$. 要保证装配, 当然购买得越多越好了, 但由于采购成本的原因, 不可能买得太多.

设 X 表示购买的 n 只集成电路中合格品的个数, 则由二项概率公式(1.1.12)

$$P\{X \geqslant 12\}=\sum_{k=12}^{n} \mathrm{C}_n^k 0.9^k 0.1^{n-k} \geqslant 0.99$$

由于上式不能用解析的方法求出, 从而应用离散的方法进行推算.

表 1.1.1　采购数量与保证装配概率对应表

采购数 n	保证装配概率	采购数 n	保证装配概率
12	0.2824	17	0.9953
13	0.6213	18	0.9988
14	0.8416	19	0.9997
15	0.9444	20	0.9999
16	0.9830		

由表 1.1.1 可知, 采购 17 只时, 可以达到要求.

1.2　一维随机变量

随机变量(random variable)及其分布概念的引入, 为概率论和数理统计方法的研究提供了有效工具.

1.2.1　一维随机变量的分布

设 X 为随机变量, 称

$$F(x)=P\{X\leqslant x\},\quad x\in\mathbf{R} \tag{1.2.1}$$

为 X 的分布函数.

分布函数 $F(x)$ 蕴含了随机变量 X 所有取值及其分布情况.

分布函数具有下列基本性质.

性质 1.2.1　(1) 非负性: $0\leqslant F(x)\leqslant 1,\ x\in\mathbf{R}$;

(2) 单调不减性: 当 $x_1<x_2$ 时, 有 $F(x_1)\leqslant F(x_2)$;

(3) 规范性: $F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0,\ F(+\infty)=\lim\limits_{x\to+\infty}F(x)=1$;

(4) 右连续性: $F(x+0)=\lim\limits_{t\to x+0}F(t)=F(x), x\in\mathbf{R}$;

(5) 函数 $F(x)$ 在点 x 连续的充分必要条件是 $P\{X=x\}=0$.

随机变量分为离散型随机变量(discrete random variable)、连续型随机变量(continuous random variable)、既不离散也不连续型随机变量. 离散型随机变量、连续型随机变量是常见的两类随机变量. 如果随机变量 X 取有限多个或可列多个不同的值, 则称 X 为离散型随机变量, 设 X 所有可能的取值为 $a_1,a_2,\cdots,a_i,\cdots$, 则称 $p_i=P\{X=a_i\}\ (i=1,2,\cdots)$ 为 X 的分布律或概率分布. 分布律具有如下两个基本性质:

(1) 非负性: $p_i\geqslant 0,\ i=1,2,\cdots$;

(2) 正则性: $\sum\limits_{i=1}^{+\infty}p_i=1$.

连续型随机变量常用于刻画高度、长度、质量、时间、面积、体积等指标. 设 $F(x)$ 为随机变量 X 的分布函数, 如果 $F(x)$ 可表示为非负函数 $f(x)$ 的积分, 即

$$F(x)=\int_{-\infty}^{x}f(t)\mathrm{d}t,\quad x\in\mathbf{R} \tag{1.2.2}$$

则称 X 为连续型随机变量, $f(x)$ 为 X 的密度函数(density function), X 服从 $f(x)$, 记为 $X\sim f(x)$.

连续型随机变量 X 的密度函数 $f(x)$ 与分布函数 $F(x)$ 有下列性质.

性质 1.2.2　(1) 非负性: $f(x)\geqslant 0,\quad x\in\mathbf{R}$;

(2) 规范性: $\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$;

(3) $P\{a<X\leqslant b\}=\int_a^b f(x)\mathrm{d}x$, 其中 a,b 均为常数, 且 $a<b$;

(4) $F(x)$ 是连续函数;

(5) 在 $F(x)$ 的可微点 x 处, 有 $F'(x)=f(x)$.

1.2.2　常见一维随机变量的分布

1. 0-1 两点分布

如果随机变量 X 的分布律为

$$P\{X=k\}=p^k(1-p)^{1-k},\quad k=0,1 \tag{1.2.3}$$

其中 $0<p<1$，则称 X 服从 0-1 两点分布，记为 $X\sim B(1,p)$.

显然，伯努利试验结果可以用服从 0-1 两点分布的随机变量描述.

2. 二项分布

如果随机变量 X 的分布律为

$$P\{X=k\}=\mathrm{C}_n^k p^k(1-p)^{n-k},\quad k=0,1,2,\cdots,n \tag{1.2.4}$$

其中 $0<p<1$，则称 X 服从参数为 n 与 p 的二项分布(binomial distribution)，记为 $X\sim B(n,p)$.

二项分布描述的是 n 重伯努利试验中事件 A 出现次数的分布. 二项分布是离散型随机变量分布中重要的分布之一，有着广泛的应用. 例如，保险公司对企业或建设项目做风险担保时，必须对各种可能事件发生的概率进行估计. 做人寿和意外伤亡保险时需要计算各种各样伤亡、疾病甚至自然灾害的概率，这些概率计算需要用到二项分布. 又如，在质量管理中，不合格产品数控制图和不合格率控制图的绘制，一些抽样检验方案的制定等都要用到二项分布.

例 1.2.1　有 9 台设备，间歇地使用电力. 设在任一时刻每台设备都以同样的概率 0.3 需要一个单位的电力，且各设备是否需要电力相互独立. 求有 3 台设备同时需要供应一个单位电力的概率.

解　设 X 表示 9 台设备任一时刻同时需要供应一个单位电力的设备数，则 $X\sim B(9,0.3)$. 于是，3 台设备同时需要供应一个单位电力的概率为

$$P\{X=3\}=\mathrm{C}_9^3 0.3^3(1-0.3)^6\approx 0.2668$$

3. 泊松分布

如果随机变量 X 的分布律为

$$P\{X=k\}=\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!},\quad k=0,1,2,\cdots \tag{1.2.5}$$

其中 $\lambda>0$，则称 X 服从参数为 λ 的泊松分布(Poisson distribution)，记为 $X\sim P(\lambda)$.

泊松分布在确定一个事件在特定时间或空间间隔中发生次数的随机分布方面是非常有用的. 例如，某段时间内来到某售票口买票的人数、电话交换台在给定时间内收到用户的呼叫次数、交通枢纽在某高峰期的客流量和车流量、保险公司在一定时期内被索赔的次数、某天上午某大学上课迟到的总人数、某部手机一天内收到的短信数等，都可用服从泊松分布的随机变量来描述.

例 1.2.2　某商店根据过去的销售记录知道某种商品每月的需求量(单位：件)服从参数为 9 的泊松分布. 为了以 95% 及以上的概率保证商品不脱销，问商店一个月应准备多少件这种商品？

解　设商店一个月需要准备 a 件商品，商品每月的需求量为 X 件. 根据题意，$X\sim P(9)$. 商品要不脱销，必须 $X\leqslant a$. 于是，参数 a 应满足

$$P\{X\leqslant a\}\geqslant 0.95$$

即

$$\sum_{k=0}^{a}P\{X=k\}=\sum_{k=0}^{a}\frac{9^k}{k!}\mathrm{e}^{-9}\geqslant 0.95$$