云南省普通高等学校“十二五”规划教材

概率论与数理统计 第二版

主编 戴琳 吴刘仓

高等教育出版社

[redacted]省普通高等学校“十二五”规划教材

概率论与数理统计

第二版

主　编　戴　琳　吴刘仓

编　者　戴　琳　付英姿　何维刚

　　　　秦叔明　吴刘仓　詹金龙

（按汉语拼音排序）

高等教育出版社·北京

内容提要

本书是在2009年第一版的基础之上修订而成的，一共九章，内容包括事件与概率、一维随机变量及其分布、多维随机变量及其分布、随机变量的数字特征、数理统计的基本概念、参数估计、假设检验、方差分析与正交试验、回归分析，各章配备了适量的精选习题，书末附有相应的习题答案以及常用的分布表、分位数表、临界值表、正交表。

本书可作为高等学校理工科各专业概率论与数理统计课程的教材，也可供科技工作者和工程技术人员参考。

图书在版编目(CIP)数据

概率论与数理统计 / 戴琳,吴刘仓主编. -- 2版. -- 北京：高等教育出版社，2017.8
ISBN 978-7-04-047819-8

Ⅰ.①概… Ⅱ.①戴… ②吴… Ⅲ.①概率论-高等学校-教材 ②数理统计-高等学校-教材 Ⅳ.①O21

中国版本图书馆CIP数据核字(2017)第119000号

策划编辑 高 丛　责任编辑 高 丛　封面设计 张 志　版式设计 马敬茹
插图绘制 杜晓丹　责任校对 胡美萍　责任印制 毛斯璐

出版发行 高等教育出版社
社　　址 北京市西城区德外大街4号
邮政编码 100120
印　　刷 三河市华骏印务包装有限公司
开　　本 787mm×960mm 1/16
印　　张 22.75
字　　数 410千字
购书热线 010-58581118
咨询电话 400-810-0598
网　　址 http://www.hep.edu.cn
　　　　 http://www.hep.com.cn
网上订购 http://www.hepmall.com.cn
　　　　 http://www.hepmall.com
　　　　 http://www.hepmall.cn
版　　次 2009年9月第1版
　　　　 2017年8月第2版
印　　次 2017年8月第1次印刷
定　　价 40.00元

物 料 号 47819-00

第二版前言

本书是在2009年第一版的基础之上,根据教育部高等学校大学数学教学指导委员会2014年颁布的“工科类本科数学基础课程教学基本要求”,结合八年来教学实践中不断积累的经验,重新修订而成的。本次修订着重概率统计的思想、原理、方法,加强针对性、实用性,兼顾系统性、严密性,力求所配例题与习题具有典型性、启发性,以期便于教学和自学。

本次修订主要进行了以下变动:

1. 调整了部分内容,增补了少量内容,主旨在于优化结构或突出重点。

第三章中,随机变量的独立性被分离出来,单立一节,以凸显其重要性。

第五章中,抽样分布一节被分成两节,其中三大分布与分位数另立一节;增补了比较实用的频率直方图,并与经验分布函数合并成一节。

第七章中,单个与两个正态总体的参数假设检验分别各立一节;分布拟合检验一节更名为非参数假设检验,其中增补了便于应用的符号检验。

原书第十章的内容被分拆到各章,以便读者循序渐进地了解 MATLAB 在概率统计中的应用。

删减了各章的思考题,不少章节添加了例题,对重点内容进行了总结,对前一版中的错漏和不当之处也作了订正,等等。

2. 适度增加了各章的习题量,并分成 A、B 两组,其中 A 组为基础练习,建议读者全部完成,有助于充分理解、牢固掌握所学知识;B 组为扩展练习,读者可以根据自己的能力和需求选择完成,有益于进一步拓展思路。

3. 斟酌了本书的叙述,使文字更加简炼、语言更加流畅、表述更加准确、层次更加分明、衔接更加自然。

修订的具体分工如下:戴琳、付英姿修改第一、二、五、六章,秦叔明修改第三、四章,吴刘仓修改第七、八、九章,何维刚负责绘制书中插图;统稿工作由戴琳、吴刘仓教授完成,最终由詹金龙教授审定。

本书第一版于2014年被列为云南省普通高等学校“十二五”规划教材,云南省教育厅为本次修订提供了专项资助,昆明理工大学教务处和高等教育出版

社也给予了大力支持,在此表示衷心感谢!

经本次修订,书中难免仍有疏漏和不当之处,恳请读者不吝指正。

编　者

2017年8月

第一版前言

本书根据最新的“工科类本科数学基础课程教学基本要求”编写而成,适用于高等学校理工类专业。本书内容包括:随机事件与概率、一维随机变量及其分布、二维随机变量及其分布、随机变量的数字特征、数理统计的基本概念、参数估计、假设检验、方差分析与正交试验设计、回归分析、MATLAB 在概率统计中的应用等。附录中包含了本书所需查找的各种分布表及各节习题的简明解答。

本书遵循高等教育的特点并结合理工类专业学生的实际情况,坚持“以基础为出发点、以应用为目的、以创新为导向”的编写原则,具有以下特点:

1. 对概率论与数理统计基本要求规定的基础内容部分着力下工夫,突出概率统计的思想、方法及应用。对基本概念的叙述尽量深入浅出,对基本定理的证明力求简明易懂,对基本方法的介绍力求充实到位,对例题及习题的选择力求典型性和富有启发及应用性,每章后面的思考题有利于拓展学生对基本概念、基本理论及基本方法的全面深入理解。

2. 在内容组织上,根据学生的认识发展规律和心理特点精心安排,便于学生自学。

3. 更新教学内容,增加了正交试验设计的内容,并编写了 MATLAB 在概率统计中的应用等,使教材更加直观化和现代化。

4. 正确处理好理论与实践的结合及应用,在教材中重视概率统计的应用,特别是加入了正交试验设计、MATLAB 在概率统计中的应用等内容,加强了整本教材的实用性。

本书第一、二章由陈秀华编写,第三、四章由秦叔明编写,第五、六章由戴琳编写,第七章由徐润林编写,第八、九章由吴刘仓编写,第十章由李建飞编写。初稿完成后由戴琳、陈秀华、秦叔明对全书进行了系统的统稿、定稿和加工工作。本书的全部插图由徐润林完成。

在本书的编写过程中得到了昆明理工大学理学院和学校教务处的大力支持;数学系教师张卫锋、张志坚及张江红等参加了教材的录入、校稿等工作;高等教育出版社为本书的出版做了大量工作,在此表示衷心的感谢。

由于编审人员水平有限，本书不足之处在所难免，诚恳希望广大教师及读者批评指正。

编　者

2009 年 6 月

目　　录

第一章　事件与概率

自然界和人类社会中的现象是多种多样的,大体上可分为两类:随机现象与必然现象.

必然现象:在一定的条件下必然发生的现象,具有因果关系,又称**确定性现象**. 例如,月球绕地球旋转,同性电荷相斥,水在 4℃时密度最大,等等.

随机现象:在一定的条件下可能发生、也可能不发生的现象,具有不确定性,又称**偶然现象**. 例如,掷一硬币,可能出现正面(带币值的面),也可能出现反面,事先无法确定;播下的种子,可能发芽,也可能不发芽,事先同样无法确定.

从个别观察或试验的结果来看,偶然现象是无规律可循的,但在大量独立重复试验中,却呈现出确定的规律性. 例如:

(1) 在相同条件下重复抛掷同一硬币,出现正面的次数大约有一半,如果硬币均匀的话.

(2) 个别婴儿在降生之前不能预言是男是女,但观察大量新生儿的性别后,可以得出男女大约各占一半的结论.

(3) 打靶时,无论怎样控制射击条件,每一次射击之前不能预测弹着点的准确位置,射击次数不多时,弹着点的分布是无规律的,但随着射击次数不断增加,弹着点的分布就逐渐显示出规律性.枪法越好,弹着点就越集中于靶心附近.

大量独立重复试验中,随机现象所呈现的规律性叫做**统计规律性**. 概率论与数理统计就是研究和揭示随机现象统计规律性的一门数学学科.

概率论与数理统计大约起源于 17 世纪中叶法国社会中盛行的赌博游戏,经过 300 多年,概率论与数理统计取得了长足的进展,已成为数学学科中最活跃的分支之一. 它不仅有自身的理论体系,而且日益广泛地应用到科学技术的许多方面,甚至深入人文学科.

§1　事件及其运算

事件是概率论中最基本的概念,概率论研究的中心问题就是事件发生可能性的定量描述,而概率则是用来度量这种可能性大小的一个数值. 下面先来介绍事件的概念.

一、随机试验与事件

为了研究随机现象,需要进行大量观察或试验. 若试验 E 具有以下特性,则 E 称为**随机试验**:

(1) 在相同的条件下可以重复进行;

(2) 事先可以明确试验的所有可能结果;

(3) 每次试验之前,不能预言究竟哪个结果会发生.

随机试验的每个可能结果称为**样本点**,全体样本点组成的集合称为**样本空间**,记作 Ω.

例 1 观察新生儿的性别,样本点有 2 个,样本空间

$$\Omega = \{e_1, e_2\},$$

其中 e_1、e_2 分别表示男性、女性.

例 2 掷一骰子,观察出现的点数,样本点有 6 个,样本空间

$$\Omega = \{1,2,3,4,5,6\}.$$

例 3 观察某电话交换机(站)在一小时内接到的呼唤次数,样本点有无限多个,样本空间

$$\Omega = \{0,1,2,\cdots\}.$$

例 4 在一批灯泡中任取一只,观察其寿命,用 t 表示灯泡寿命,则样本点是一非负实数. 由于不能确定寿命上界,任一非负实数都可以认为是一个可能结果,所以样本空间

$$\Omega = \{t \mid t \geqslant 0\}.$$

例 5 将长为 1 的直尺折为 3 段,用 x,y,z 记各段的长度,则样本点是空间中的点 (x,y,z),样本空间

$$\Omega = \{(x,y,z) \mid x > 0, y > 0, z > 0, x + y + z = 1\}.$$

从上面的例子可以看出,样本空间可以是有限集合(例 1,例 2),也可以是无限集合(例 3,例 4),甚至可以是一个区域(例 5).

考察随机试验时,我们感兴趣的是具备特定条件的样本点出现与否. 例如,若规定某种灯泡的寿命 $t<1000$ 为次品,关心的则是 $t\geqslant 1000$,满足这一条件的样本点构成样本空间的子集 $\{t \mid t\geqslant 1000\}$.

一般地,样本空间 Ω 的子集称为**随机事件**,简称**事件**,用大写字母 A,B,C 等表示.

规定:每次试验中,事件 A 发生当且仅当 A 中的一个样本点出现.

特别地,一个样本点构成的单点集称为**基本事件**. 在随机试验中,必然发生的事件称为**必然事件**,样本空间 Ω 作为自身的子集,每次试验必定发生,即为必

然事件.在随机试验中,一定不会发生的事件称为**不可能事件**,而空集$\varnothing$不含任何样本点,作为样本空间的子集,每次试验都不发生,即为不可能事件.

二、事件的关系及运算

由一个随机试验的样本空间,可以产生若干事件,有的复杂,有的简单. 为了用简单事件表示复杂事件,引入事件的关系及运算.

设试验的样本空间为Ω,事件$A,B,A_1,A_2,\cdots,A_n$均为Ω的子集.

(1) 子事件

若事件A发生必然导致事件B发生,则A称为B的**子事件**,记作$A\subset B$.

特别地,$A\subset B$与$B\subset A$同时成立时,称A与B**相等**,记作$A=B$.

(2) 和事件

由事件A与B中至少有一个发生构成的事件,称为A与B的**和事件**,记作$A\cup B$.

(3) 差事件

由事件A发生而事件B不发生构成的事件,称为A与B的**差事件**,记作$A-B$.

如高考体检中,事件A为身高合格,事件B为血压合格,则$A-B$表示身高合格,但血压不合格.

(4) 积事件

由事件A与B同时发生构成的事件,称为A与B的**积事件**,记作$A\cap B$或AB.

(5) 互斥与互逆事件

若事件A与B不能同时发生,则称A与B**互斥**或**互不相容**,记作$AB=\varnothing$.

对任一随机试验,其基本事件是互斥的,即一次试验中不可能有两个基本事件同时出现.

A与B互斥时,$A\cup B$也写成$A+B$.

进而,如果A,B满足条件

(i) $AB=\varnothing$;

(ii) $A\cup B=\Omega$,

那么A,B称为**互逆**或**对立事件**.

A的对立事件记作$\bar{A}$,表示A不发生. 事实上,$\bar{A}=\Omega-A$,且$A-B=A\bar{B}$.

互逆事件必然是互斥的,但互斥事件未必互逆.

(6) 完备事件组

若$A_1,A_2,\cdots,A_n$满足$A_1\cup A_2\cup\cdots\cup A_n=\Omega$,且两两互斥,则$A_1,A_2,\cdots,A_n$称为**完备事件组**.

其实,完备事件组就是样本空间 Ω 的一个分割.

所有基本事件构成一个完备组;对任何事件 A,A 与 $\bar{A}$ 同样构成一个完备组.

若矩形区域表示样本空间 Ω,圆形区域表示事件 A、B,则事件的关系及运算可用所谓文氏图(图 1-1)直观表示如下:

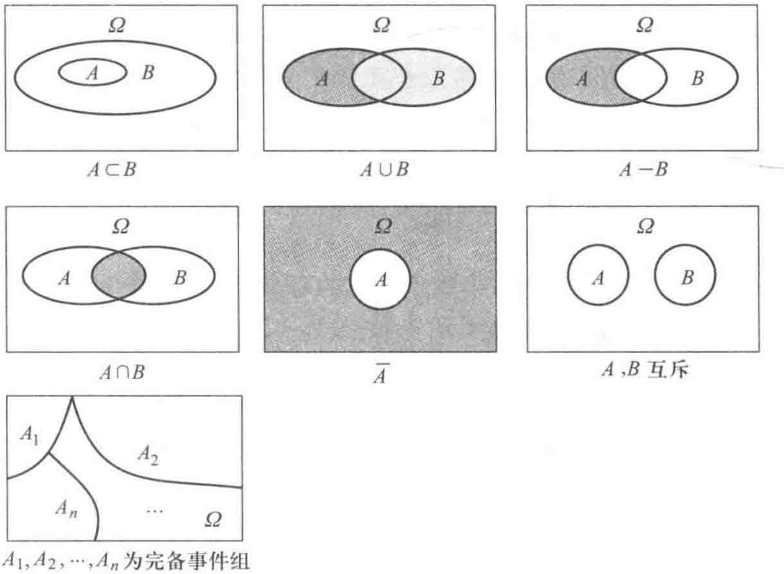

图 1-1

重要的是,要学会用符号表示事件;反之,对于用符号表示的事件,需清楚其具体含义. 简言之,就是要能正确无误地“互译”出来.

例 6　检验某种圆柱形产品,要求直径和长度都合格才算合格. 以事件 A 记直径合格,事件 B 记长度合格,则

$A\cup B$:直径或长度合格,不能确定产品是否合格;

$A-B$:直径合格但长度不合格,产品不合格;

AB:直径和长度都合格,产品合格;

$\bar{A}\bar{B}$:直径和长度都不合格,产品不合格　;

$\bar{A}\cup\bar{B}$:直径和长度至少有一个不合格,产品不合格.

例 7　设随机试验 E 为同时掷两枚硬币,观察出现正、反面的情形. 记基本事件

$$e_1:(正,正),\quad e_2:(正,反),\quad e_3:(反,正),\quad e_4:(反,反).$$

随机事件

$$A=\{e_1\},\quad B=\{e_2,e_3\},\quad C=\{e_1,e_2,e_3\},$$

试说明 A,B,C 的含义,并指出它们之间的关系.

解　$A=\{e_1\}$:两枚都正面朝上,

$B=\{e_2,e_3\}$:只有一枚正面朝上,

$C=\{e_1,e_2,e_3\}$:至少有一枚正面朝上.

易见 $A\cup B=C,AC=A,AB=\varnothing$.

例 8　射击目标 3 次,设事件 A_i 为第 i 次击中($i=1,2,3$),试用 A_1,A_2,A_3 表示下列随机事件 A,B,C,D,并指出它们之间的关系.

(1) A:3 次都击中;(2) B:击中 2 次;(3) C:击中 1 次;(4) D:3 次都未击中.

解　(1) $A=A_1A_2A_3$.

(2) $B=A_1A_2\overline{A}_3+A_1\overline{A}_2A_3+\overline{A}_1A_2A_3$.

(3) $C=A_1\overline{A}_2\overline{A}_3+\overline{A}_1A_2\overline{A}_3+\overline{A}_1\overline{A}_2A_3$.

(4) $D=\overline{A}_1\overline{A}_2\overline{A}_3$.

显然 A,B,C,D 两两互斥,且它们的和构成必然事件,即 $A+B+C+D=\Omega$.

本例中,若第 i 次射击记作 E_i,则 E_i 的样本空间 $\Omega_i=\{A_i,\overline{A}_i\}$,且本例的随机试验 E 由 $E_i(i=1,2,3)$ 复合而成,E_1,E_2,E_3 各做一次,E 才算完成,故 E 称为**复合随机试验**,其样本空间

$$\Omega=\{A_1A_2A_3,A_1A_2\overline{A}_3,A_1\overline{A}_2A_3,\overline{A}_1A_2A_3,A_1\overline{A}_2\overline{A}_3,\overline{A}_1A_2\overline{A}_3,\overline{A}_1\overline{A}_2A_3,\overline{A}_1\overline{A}_2\overline{A}_3\}.$$

事件与集合具有完全相同的运算规律,在此仅简略介绍下列常用结论:

(1) **结合律**:$(A\cup B)\cup C=A\cup(B\cup C),(AB)C=A(BC)$;

(2) **交换律**:$A\cup B=B\cup A,AB=BA$;

(3) **分配律**:$A\cup(BC)=(A\cup B)(A\cup C),A(B\cup C)=(AB)\cup(AC)$;

(4) **对偶律**:$\overline{A\cup B}=\overline{A}\overline{B},\overline{AB}=\overline{A}\cup\overline{B}$,

其中对偶律又称**德摩根**(De Morgan)公式.

§2　事件的概率

在一次试验中,事件可能发生,也可能不发生,而人们常常需要了解某些事件发生的可能性大小. 为此,首先引入描述事件发生频繁程度的数——频率,进而引出在一次试验中刻画事件发生可能性大小的量——概率.

一、概率的统计定义

定义 1　设 A 为随机试验 E 的事件,将 E 独立重复 n 次,称 A 发生的次数

n_A 为 A 在 n 次试验中发生的**频数**,称比值$\frac{n_A}{n}$为 A 在 n 次试验中发生的**频率**,记作$f_n(A)$,即

$$f_n(A)=\frac{n_A}{n}.$$

频率具有下列基本性质:

(1) $f_n(\varnothing)=0, f_n(\Omega)=1$;

(2) 对任何事件 A,都有 $0\leqslant f_n(A)\leqslant 1$;

(3) 若事件 A 与 B 互斥,则有 $f_n(A+B)=f_n(A)+f_n(B)$.

频率描述的是事件发生的频繁程度,在一定程度上反映了事件发生的可能性大小,但它不是一个确定的数值,当试验重复进行时,它也不断变化. 然而,大量试验表明频率还具有稳定性,历史上曾做过许多验证频率稳定性的试验,比如下表所示的掷硬币试验.

试验者	掷硬币次数 n	出现正面次数	出现正面的频率 f_n
德摩根	2 048	1 061	0.518 1
蒲　丰	4 040	2 048	0.506 9
皮尔逊	12 000	6 019	0.501 6
皮尔逊	24 000	12 012	0.500 5

从表中可以看出,出现正面的频率总在 0.5 附近波动,随着试验次数的增加,逐渐接近于 0.5.

综上所述,随着试验次数的不断增加,事件 A 的频率逐渐稳定于某常数 p 附近,这个性质称为**频率的稳定性**.这个稳定值 p 刻画了事件 A 发生的可能性大小,由此引出概率的统计定义.

定义 2　设 A 为随机试验 E 的事件,在 E 独立重复发生过程中,A 发生之频率趋于某一稳定值,该稳定值称为事件 A 的**概率**,记作 $P(A)$.

概率的统计定义既肯定了事件概率的存在性,又阐明了概率与频率之间的关系,即频率是概率的近似值. 因此,在实际问题中,当概率难以得到时,可以用频率来替代.

下面介绍两种简单而特殊的概率:古典概率与几何概率.

二、概率的古典定义

定义 3 设试验的样本空间 Ω 中样本点只有有限多个(比如 n 个),且每个样本点出现的可能性都相同(称为**等可能性**). 若事件 A 所含样本点个数为 k,则 A 的概率规定为

$$P(A)=\frac{k}{n},$$

并称之为**古典概率**,而相应的概率模型称为**古典概型**.

用定义中的公式计算概率时,通常需用排列或组合计算样本点总数 n 及事件 A 所含的样本点个数 k.

例 1(抽签原理) 袋中有 a 个白球、b 个红球,依次取一球,取后不放回,且在一次取球时每球被取到的可能性相同,求第 m 次抽到白球的概率.

解法一 将球全部编号并依次取出放入 $a+b$ 个格子,每个格子一球,样本点总数 $n=(a+b)!$. 以 A 记第 m 次抽到白球,则 A 所含的样本点数 $k=\mathrm{P}_a^1(a+b-1)!$,其概率为

$$P(A)=\frac{\mathrm{P}_a^1(a+b-1)!}{(a+b)!}=\frac{a}{a+b}.$$

解法二 将球全部编号并依次取出放入 m 个格子,同样每个格子一球,样本点总数 $n=\mathrm{P}_{a+b}^m$. 而有利于 A 发生的样本点数 $k=\mathrm{P}_a^1\mathrm{P}_{a+b-1}^{m-1}$,故所求概率

$$P(A)=\frac{\mathrm{P}_a^1\mathrm{P}_{a+b-1}^{m-1}}{\mathrm{P}_{a+b}^m}=\frac{a}{a+b}.$$

本例说明,抽签与先后次序无关.

例 2 某人有 5 把钥匙,其中 2 把为房门钥匙,但忘记了究竟是哪两把,只有逐一试开,求试开 3 次能打开房门的概率.

解法一 从 5 把钥匙中每次取 1 把,无放回地逐一试开 3 次,总的取法为 $n=\mathrm{C}_5^1\times\mathrm{C}_4^1\times\mathrm{C}_3^1=60$;3 次都未抽到房门钥匙的方法有 $\mathrm{C}_3^1\times\mathrm{C}_2^1\times\mathrm{C}_1^1=6$ 种,从而 3 次至少取到 1 把房门钥匙的方法有 $k=60-6=54$ 种. 用 A 表示房门能打开,则有

$$P(A)=\frac{k}{n}=\frac{54}{60}=\frac{9}{10}=0.9.$$

解法二 从 5 把钥匙中任取 3 把,共有 $n=\mathrm{C}_5^3$ 种取法,取到 1 或 2 把房门钥匙的方法有 $k=\mathrm{C}_2^1\times\mathrm{C}_3^2+\mathrm{C}_2^2\times\mathrm{C}_3^1$ 种,事件 A 发生的概率

$$P(A)=\frac{\mathrm{C}_2^1\times\mathrm{C}_3^2+\mathrm{C}_2^2\times\mathrm{C}_3^1}{\mathrm{C}_5^3}=\frac{9}{10}=0.9.$$