Probability, Mathematical Statistics and Applications of SPSS

概率论与数理统计及SPSS软件应用

孟 璀 吴培群◎编著

中国工信出版集团 人民邮电出版社 POSTS & TELECOM PRESS

Probability, Mathematical Statistics and Applications of SPSS

概率论与数理统计及SPSS软件应用

孟　璀　吴培群◎编著

人民邮电出版社
北京

图书在版编目（C I P）数据

概率论与数理统计及SPSS软件应用 / 孟璀，吴培群编著. -- 北京 : 人民邮电出版社，2017.11
ISBN 978-7-115-46967-0

Ⅰ. ①概… Ⅱ. ①孟… ②吴… Ⅲ. ①概率论②数理统计③统计分析－软件包 Ⅳ. ①O21②C819

中国版本图书馆CIP数据核字(2017)第233704号

内 容 提 要

本书内容包括概率论、数理统计、SPSS 软件应用三部分，共 9 章，每章附有习题。第 1~4 章是概率论，包括概率论的基本概念、随机变量及其分布和数字特征、几种特殊随机变量的分布、多维随机变量及其分布和数字特征。第 5~7 章是数理统计，包括参数估计、假设检验、方差分析和回归分析。第 8~9 章是 SPSS 软件应用，包括 SPSS 软件基本使用方法，用 SPSS 进行描述性统计分析，用 SPSS 进行均值比较、方差分析、相关分析和回归分析等统计推断，这部分是前两部分内容的直接应用。

本书可作为高等学校信息管理、经济管理类，以及其他非数学专业概率论与数理统计课程的教材，也可供各类管理和专业技术人员及有志于考研的学生参考。

◆ 编　　著　孟　璀　吴培群
责任编辑　邢建春
责任印制　彭志环

◆ 人民邮电出版社出版发行　　北京市丰台区成寿寺路 11 号
邮编　100164　　电子邮件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
大厂聚鑫印刷有限责任公司印刷

◆ 开本：700×1000　1/16
印张：10.5　　2017 年 11 月第 1 版
字数：206 千字　　2017 年 11 月河北第 1 次印刷

定价：69.00 元

读者服务热线：(010) 81055488　印装质量热线：(010) 81055316
反盗版热线：(010) 81055315

前　言

为适应信息化发展的需要和大数据时代的特点，概率论与数理统计已作为一门强调应用性、工具性的独立课程，面向各个非数学专业的本科生开设。在这些非数学专业中，有相当一部分专业对概率统计的要求与强调严密逻辑体系、较多定理证明的传统同名教材的内容相差较大。为适应新的教学需要，笔者在多年相关课程研究和教学实践的基础上，按照教育部最新颁布的概率论与数理统计教学基本要求，撰写了这本可用于该课程理论课和实验课的教学用书。

1. 本书特色

（1）兼容概率统计理论知识与常用统计软件操作技术

本书包括概率论与数理统计的基本内容，又有作为常用统计软件之一的 SPSS 的基本操作方法内容，并且内容前后呼应。

（2）降低理论性，强调统计应用，强化概率统计课程的工具价值

与传统教材相比，本书减少了概率论部分的内容，增加了数理统计内容；例题多数是笔者在长期教学实践中积累的与学生实际生活接近的问题；书中命题证明很少，对多数定理着重于分析其条件和结论，强调应用定理解决实际问题。另外，将概率统计的思想方法、理论知识，落实于软件实际应用技能，也是本书的主要特色之一。

（3）内容结构简单，便于读者学习使用

本书内容基本呈现为线性结构，SPSS 统计软件内容是数理统计的技术实现，而数理统计内容是 SPSS 统计技术的理论基础，书中概率论部分又是数理统计内容的理论基础。

2. 本书主要内容、适用对象及教学建议

本书内容分为三大部分：概率论、数理统计和 SPSS 统计软件。全书共 9 章，第 1~4 章是概率论内容，概率论的基本概念、随机变量及其分布和数字特征。为了强化概率论内容为数理统计和统计软件提供基础服务的特点，本书专设一章探

讨几种常用随机变量的分布。

第 5~7 章是数理统计内容，在传统的数理统计基本概念、参数估计和假设检验的基础上，增加了方差分析和线性回归的初步内容。

第 8~9 章介绍 SPSS 统计软件及其使用方法。其中，第 8 章包括 SPSS 基础和数据统计前的数据录入和整理、用 SPSS 做描述性统计分析；第 9 章探讨用 SPSS 做均值比较、方差分析、相关分析和线性回归的方法和操作技术。

本书可作为高等学校信息管理类、经济管理类、人文社科类及其他非数学类专业概率论与数理统计课程的教材，也可供各类专业技术人员及有志于考研的学生参考。运用本书进行“概率论与数理统计”等相关课程的教材时，建议为 58 课时左右，其中理论课 46 课时左右，上机实验课 12 课时左右；若理论课教学课时为 32 学时左右，建议对本书的第 4 章和第 7 章略讲；若无实验课时，建议将 SPSS 统计软件内容作为学生自学内容，教师在数理统计内容教学时对 SPSS 软件做适当演示。

本书第 1~7 章由孟瓘负责完成，第 8~9 章由吴培群负责完成，全书由孟瓘统稿。北京电子科技学院张师瑜同学帮助整理了部分书稿，在此表示感谢。

本书虽经作者反复修改，但难免有疏漏和错误之处，诚望读者批评指正。

作者

2017 年 8 月

目　录

第 1 章 概率论的基本概念

微积分研究函数的微分（或导数）、积分运算等，线性代数研究向量、线性空间、线性变换和有限维线性方程组等。那么概率论这门数学分支研究什么？

1.1 概率论的研究对象和概率论简史

（一）随机现象

现实世界中的现象可以分为两类，分别是确定性现象和随机现象。确定性现象是指在一定条件下必然发生的现象。例如，向空中抛一个物体，这个物体必然落向地面；水加热到 100℃必然沸腾等。随机现象是指在实验或观察前无法预知确切结果的现象。但人们经过长期实践并深入研究之后，发现这类现象在大量重复实验或观察下，它的结果却呈现出某种规律性。在生活中，我们能接触到的随机现象有很多。例如，连续重复多次掷一枚骰子，出现每个点数的频率大致都有六分之一；向同一目标射击，各次弹着点都不相同，但弹着点的分布有一定的规律等。这种在大量重复实验或观察中所呈现出的固有规律性，就是我们以后所说的统计规律性。

1. 随机现象的特点

（1）随机性

在一定的条件下进行的实验或观察，可能出现这样的结果，也可能出现那样的结果，即在实验或观察之前不能预知确切的结果。

（2）统计规律性

在大量重复进行同一个实验或观察的时候，其结果会呈现固有的规律性。

2. 概率论的研究对象

概率论是研究和揭示随机现象的统计规律性的一门数学学科。

（二）概率论的产生和发展

概率论最早是探索赌博（博弈）游戏开始的。有关赌博最早的一个数学问题出现在1494年意大利修士、数学家巴乔罗(Luca Pacciolo)的著作《算术，几何，比例和比值要义》中。

甲、乙两人相约赌若干局，谁先赢 s 局就获胜。现在甲赢 a 局($a<s$)，而乙赢 b 局($b<s$)时赌博中止了，问赌本应如何分?

巴乔罗的计算方法是按赌博中止时甲、乙已赢的局数分配赌本。例如，$s=3$，$a=2$，$b=1$，则按2:1分配。

热衷于占星术和掷骰子的代数学家卡丹(J.Cardan)和塔塔利亚(N.Tartanlia)指出巴乔罗的分法是错误的，认为巴乔罗的分法没有考虑甲乙双方取得最终胜利还需要赢的局数。但是他们两人也没有给出正确的解法。

17世纪中叶，法国流行用牌和骰子赌博的游戏，法国贵族德·梅理(De Méré)向法国著名数学家、物理学家帕斯卡(B.Pascal)再次提出同一问题。

这一貌似简单的问题难住了天才数学家帕斯卡，他思索了很久仍没有解决。于是，他与费马(P.Fermat)开始了关于这一问题的通信讨论。帕斯卡在1654年7月29日寄给费马的信中给出了这一问题的解。在这一问题的讨论中，产生了“概率”和“数学期望”等基本概念。帕斯卡的这封信被公认为是概率论的第一篇文献，是数学史上的一个里程碑。

在随后的200多年里，概率论不仅在理论上获得了一定发展，而且在人口统计、保险业、误差理论、天文学等自然科学和社会科学中得到了应用。

在这一时期，对概率论在理论和应用方面做出重要贡献的数学家有雅格布·伯努利(Jakob Bernoulli)、丹尼尔·伯努利(Daniel Bernoulli)、棣莫弗(De Moivre)、拉普拉斯(P.Lapace)、欧拉(L.Euler)、贝叶斯(T.Bayes)、蒲丰(G.Buffon)、高斯(F.Gauss)、泊松(S.Poisson)、布尼亚可夫斯基(V.Bunjakovski)、切比雪夫(Chebyshev)、马尔可夫(A.Markov)、李雅普诺夫(A.Lyapunov)等。

目前，概率论还成功应用到了包括语言、管理、心理、教育等绝大多数的人文和社会科学中，成为最有魅力的数学分支之一。

1.2 随机事件及其运算

（一）随机试验与随机事件

我们遇到过各种试验。这里，我们把试验作为一个含义广泛的术语。它包括

施以操作条件再观察的实验，也包括自然条件下的单纯观察。在概率论中，我们将对随机现象进行观察的实验称为随机试验。

随机试验具有以下特点：

（1）可以在相同的条件下重复进行；

（2）试验的可能结果不止一个，并且在试验前能预先知道全部可能结果；

（3）在每次试验前不能预先知道哪个结果会出现。

对于随机试验，尽管在每次试验之前不能预知试验的结果，但试验的所有可能结果组成的集合是已知的。我们将随机试验 E 的所有可能结果组成的集合称为 E 的样本空间，记为 Ω 或 S。样本空间的元素，即 E 的每个结果，称为样本点，记为 e。下面举两个例子。

随机掷三枚硬币，观察其出现数字面(H)还是国徽面(T)。

S={HHT, HHH, HTH, HTT, THH, THT, TTH, TTT}

明年夏天的最高气温 t。

$$S=\{ t \mid t\geqslant 0\}$$

注意 样本空间的元素是由试验目的所决定的。如下。

（1）观察正反面出现的情况，$S_1=\{\text{HHH, HHT},\cdots\}$

（2）观察正面（数字面）出现的次数，$S_1=\{0,1,2,3\}$

在实际中，当进行随机试验时，人们常常关心满足某种条件的那些样本点所组成的集合。一般，我们称试验 E 的样本空间 S 的子集为 E 的随机事件，简称事件，一般记为 A, B, C 等。例如，抛两个骰子，骰子可分辨，观察其出现的点数之和。

$$S=\{(1,1),(1,2),(1,3),\cdots,(6,1),\cdots,(6,6)\}$$

设 A 表示事件“点数之和为 7”，即

$$A=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

在每次试验中，当且仅当子集的一个样本点出现时，称这一事件发生。

特别地，由一个样本点组成的单点集，称为基本事件。

样本空间 S 包含所有的样本点，它是自身的子集，在每次试验中它总是发生的，S 成为必然事件。所以，必然事件一般也记为 Ω 或 S。空集 $\varnothing$ 不包含任何样本点，它也是样本空间的子集，它在每次试验中都不发生，$\varnothing$ 称为不可能事件。

（二）事件的关系及其运算

事件是一个集合，因而事件间的关系与事件的运算自然按照集合论中集合之间的关系和集合运算处理。下面给出这些关系和运算在概率论中的提法，并根据“事件发生”的含义，给出它们在概率论中的含义。

设试验 E 的样本空间为 S，而 A,B 是 S 的子集。

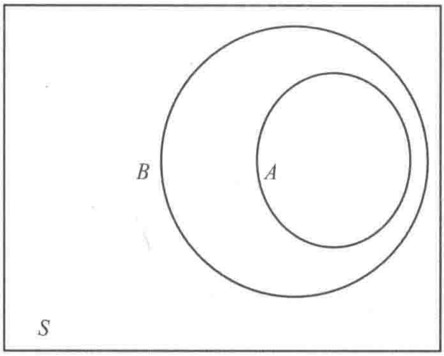
1. 包含、相等关系

（1）包含

若 $A \subset B$，则称事件 B 包含事件 A。如图 1-1 所示。即

$$\forall x \in A \Rightarrow x \in B$$

这指的是事件 A 发生必导致事件 B 发生。



$A \subset B$

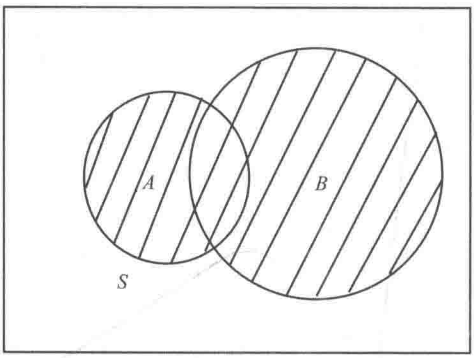
图 1-1　包含

（2）事件相等

若 $A \subset B$ 且 $B \subset A$，即 $A = B$，则称事件 A 与事件 B 相等。

2. 事件的和

事件 $A \cup B = \{x \mid x \in A 或 x \in B\}$，称为事件 A 与事件 B 的和事件。如图 1-2 所示。当且仅当 A,B 中至少有一个发生时，事件 $A \cup B$ 发生。

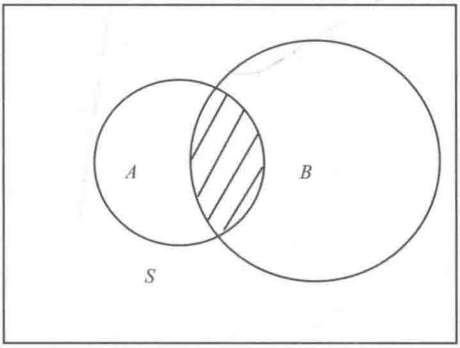


$A \cup B$

图 1-2　事件的和

3. 事件的积

事件 $A \cap B = \{x \mid x \in A且x \in B\}$，称为事件 A 与事件 B 的积事件。如图 1-3 所示。当且仅当 A、B 同时发生时，事件 $A \cap B$ 发生。$A \cap B$ 通常简记为 AB。

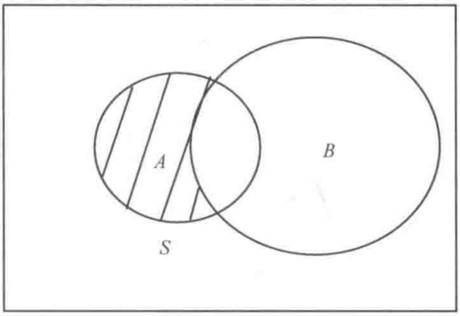


$A \cap B$

图 1-3　事件的积

4. 事件的差

事件 $A - B = \{x \mid x \in A且x \notin B\}$，称为事件 A 与事件 B 的差事件。如图 1-4 所示。当且仅当 A 发生、B 不发生时，事件 $A - B$ 发生。



$A-B$

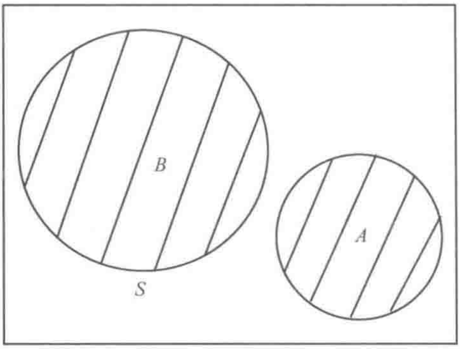
图 1-4　事件的差

5. 互斥事件（互不相容）

若 $A \cap B = \varnothing$，则称事件 A 与 B 是互不相容的，或互斥的。如图 1-5 所示。这指的是事件 A 与事件 B 不能同时发生。基本事件是两两互不相容的。

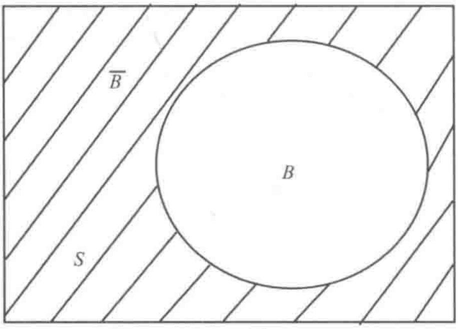
6. 对立事件（逆事件）

若 $A \cup B = S$ 且 $A \cap B = \varnothing$，则称事件 A 与事件 B 互为逆事件，又称事件 A 与事件 B 互为对立事件。如图 1-6 所示。这指的是对每次试验而言，事件 A、B 中必有一个发生，且仅有一个发生。A 的对立事件记为 $\overline{A}=S-A$。



$A \cap B = \varnothing$

图 1-5　互斥事件



$B \cup \overline{B} = S$

图 1-6　对立事件

7. 事件的运算律

在进行事件运算时，经常要用到下述定律。

设 A, B, C 为事件，则有

交换律：$A \cup B = B \cup A; A \cap B = B \cap A$

结合律：$A \cup (B \cup C) = (A \cup B) \cup C$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

分配律：$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

德 • 摩根律：$\overline{A \cup B} = \overline{A} \cap \overline{B}$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

推广：$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}; \overline{\bigcap_{k=1}^{n} A_k} = \bigcup_{k=1}^{n} \overline{A_k}$

注　事件的一些关系式如下。

① 设 $A \subset B$，则 $\overline{A} \supset \overline{B}$，$AB = A$，$A \cup B = B$，$A - B = \varnothing$

② $A - B = A\overline{B} = A - AB$

③ $A = AB \cup A\overline{B}$；

$A \subset (A \cup B)$，$B \subset (A \cup B)$；

$AB \subset B$；

$A \cup A = A$，$AA = A$

例　设 A,B,C 表示三个事件，试表示下列事件

（1）A 发生，B 与 C 不发生

（2）A 与 B 发生，C 不发生

（3）A, B 与 C 都发生

（4）A, B 与 C 至少有一个发生

（5）A, B 与 C 全不发生

（6）A, B 与 C 至少有两个发生

解　（1）$A\overline{B}\overline{C}$；（2）$AB\overline{C}$；（3）$ABC$；（4）$A \cup B \cup C$；（5）$\overline{ABC}$；（6）$AB \cup BC \cup AC$。

1.3　随机事件的概率

（一）概率的统计定义

如何定量表示事件在一次试验中发生的可能性大小？我们首先引入频率，它描述了事件发生的频繁程度，进而引出定量表示事件在一次试验中发生的可能性大小的数——概率。

1. 频率定义

设 A 为 E 中某一事件，在相同条件下进行了 n 次独立重复试验，事件 A 发生的次数记为 n_A，则比值

$$f_n(A) = n_A / n$$

称为 A 的频率。

2. 频率性质

（1）$0 \leqslant f_n(A) \leqslant 1$

（2）$f_n(S) = 1$

（3）若 $A_1,A_2,\cdots,A_k$ 两两互不相容（$\forall i\neq j, A_iA_j=\varnothing$），则

$$f_n\left(A_1\cup A_2\cup\cdots\cup A_k\right)=f_n\left(A_1\right)+f_n\left(A_2\right)+\cdots+f_n\left(A_k\right)$$

下面观察试验次数增大时频率的变化。

历史上著名的统计学家蒲丰和皮尔逊曾进行过大量掷硬币的试验，所得结果如表 1-1 所示。

表 1-1　掷硬币试验结果

试验者	次数	正面的次数	正面的频率
蒲丰	4040	2048	0.5069
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005

结论　当 n 较小时，频率呈偶然性，波动性很大；随着 n 的增加，波动幅度减小，最后集中在某一个数附近。

这种观察出来的性质称为频率稳定性，也就是通常所说的统计规律性，频率的稳定值即概率，事件 A 的概率记为 $P(A)$。

由频率的性质可以得到概率具有以下性质。

（1）非负性：$P(A)\geqslant 0$；

（2）规范性：$P(S)=1$；

（3）有限可加性：

若 $A_1,A_2,\cdots,A_n$ 是两两互不相容的事件，则 $P\left(A_1\cup A_2\cup\cdots\cup A_n\right)=\sum\limits_{i=1}^{n}P\left(A_i\right)$。

（二）概率的古典定义

当试验具有以下两个特点：

（1）试验的样本空间只包含有限个元素；

（2）试验中每个基本事件发生的可能性相同。

则称这种试验为等可能概型。它在概率论发展的初期曾是主要的研究对象，所以也称为古典概型。

下面讨论等可能概型中事件概率的计算公式。

设试验的样本空间为 $S=\{e_1,e_2,\cdots,e_n\}$。由于在试验中每个基本事件发生的可能性相同，即有

$$P\left(\{e_1\}\right)=P\left(\{e_2\}\right)=\cdots=P\left(\{e_n\}\right)$$

又由于基本事件是两两互不相容的。于是

$$
\begin{aligned}
1 = P(S) &= P(\{e_1\} \cup \{e_2\} \cup \cdots \cup \{e_n\}) \\
&= P(\{e_1\}) + P(\{e_2\}) + \cdots + P(\{e_n\}) \\
&= nP(\{e_i\}),
\end{aligned}
$$

$$P(\{e_i\}) = \frac{1}{n}, i = 1,2,\cdots,n$$

若事件 A 包含 k 个基本事件，即 $A = \{e_{i_1}, e_{i_2}, \cdots, e_{i_k}\} = \{e_{i_1}\} \cup \{e_{i_2}\} \cup \cdots \cup \{e_{i_k}\}$，这里 $i_1, i_2, \cdots, i_k$ 是 $1,2,\cdots,n$ 中 k 个不同的数。则有

$$P(A) = \sum_{j=1}^{k} P(\{e_{i_j}\}) = \frac{k}{n} = \frac{A\text{包含的基本事件数}}{S\text{中基本事件的总数}}$$

上式就是等可能概型中事件 A 的概率的计算公式。

1. 古典概率的计算方法

（1）直接用定义式计算

构造 A 和 S 的样本点，当样本空间 S 的元素较少时，先一一列出 S 和 A 中的元素，直接利用下式求解。

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

（2）用排列组合方法计算

当 A 和 S 的样本点个数较复杂时，往往要用到计数法原理和排列组合公式计算。

① 加法原理：完成一项工作有 m 类方法，第 i 类方法有 n_i 种（i=1,2,$\cdots$,m），则完成这项工作共有 $n_1 + n_2 + \cdots + n_m$ 种方法。

② 乘法原理：完成一项工作有 m 个步骤，第 i 步有 n_i 种方法（i=1,2,$\cdots$,m），则完成该项工作共有 $n_1 n_2 \cdots n_m$ 种方法。

③ 排列数：从 n 个元素中取出 r 个元素，按一定顺序排成一列，称为从 n 个元素中取出 r 个元素的排列（n，r 均为整数）。有以下两种情况。

第一种情况，无放回选取。从 n 个不同元素中无放回地取出 m 个 $(m \leqslant n)$ 进行排列，共有

$$P_n^m = n(n-1)\cdots(n-m+1) = \frac{n!}{(n-m)!}$$

种方法。当 m=n 时

$$P_n^n = n!$$

这叫作全排列。

第二种情况，有放回选取。从 n 个不同元素中有放回地抽取 r 个，依次排成