大学数学系列丛书

概率论与数理统计学习辅导及R语言解析

桂文豪　王立春　编著

清华大学出版社
http://www.tup.com.cn
北京交通大学出版社
http://www.bjtup.com.cn

大学数学系列丛书

概率论与数理统计学习辅导及R语言解析

桂文豪　王立春　编著

清华大学出版社
北京交通大学出版社
·北京·

内容简介

本书是根据作者多年的教学经验编写而成的，与北京交通大学概率统计课程组编写的《概率论与数理统计》相配套的学习辅导用书。

本书通过系统的知识点归纳及详尽的解答分析来帮助读者进一步提高概率论与数理统计的基本理论和实际应用，并且引进强大的统计软件R，从程序的角度对习题进行解析，强调统计理论知识和统计软件工具的有效结合，进一步加深统计的实践应用。

本书可作为理工科本科生，考研学生的学习辅导用书，也可供工程技术人员、科技工作者参考。

图书在版编目（CIP）数据

概率论与数理统计学习辅导及R语言解析/桂文豪，王立春编著．—北京：北京交通大学出版社：清华大学出版社，2017.2

ISBN 978－7－5121－3164－4

Ⅰ.①概…　Ⅱ.①桂…　②王…　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　③程序语言－程序设计－高等学校－教学参考资料　Ⅳ.①O21　②TP312

中国版本图书馆CIP数据核字（2017）第024202号

概率论与数理统计学习辅导及R语言解析

GAILÜLUN YU SHULI TONGJI XUEXI FUDAO JI R YUYAN JIEXI

责任编辑：谭文芳　　助理编辑：龙嫚嫚

出版发行：清华大学出版社　邮编：100084　电话：010－62776969　http://www.tup.com.cn

　　　　　北京交通大学出版社　邮编：100044　电话：010－51686414　http://www.bjtup.com.cn

印 刷 者：北京时代华都印刷有限公司

经　　销：全国新华书店

开　　本：185 mm×230 mm　印张：24.25　字数：544千字

版　　次：2017年2月第1版　2017年2月第1次印刷

书　　号：ISBN 978－7－5121－3164－4/O·160

印　　数：1～1 000册　定价：42.00元

本书如有质量问题，请向北京交通大学出版社质监组反映。

投诉电话：010－51686043，51686008；传真：010－62225406；E-mail：press@bjtu.edu.cn。

前　言

概率论与数理统计是高等院校一门重要的基础课程，它的理论和方法广泛应用于后续课程及实际问题中. 本书力求主线清晰，重点突出，体系完整. 在知识体系方面，注重归纳共性和总结规律，增加章节之间的联系，使学生对知识的掌握更加系统，提高解决实际问题的技能. 在习题讲解中，不是简单地给出答案，而是注重分析过程，并对题目涉及的知识点进行认真点评，启发和引导学生深入思考，促进学生更好地掌握解题的方法和技巧. 另外，编者会注重对习题的进一步开发和挖掘，对某些题目进行多角度解析，增加概率统计和微积分、几何代数、经济学、工程学等其他学科之间的横向联系，丰富学生学习的内容，提高学生解决实际问题的本领，扩大课程的影响力.

概率统计方法与计算机相结合是本书的一大特色，也是统计学科发展的需要. 本书首次引进优秀的统计软件 R，利用 R 的强大统计功能，对概率统计相关理论知识进行详细的解析. 概率统计与 R 相结合，不仅有助于学生更有效地利用 R 的超强功能来处理概率统计中的问题，而且有助于学生锻炼理论联系实际解决具体问题的能力，避免那种仅仅会考试或偏重理论学习的现象发生.

本书吸取了国内外优秀教材和学习辅导书的优点，取材新颖、特色鲜明、富有启发性，便于教学与自学. 本书是编者在教学改革中一种新的探索，可能会有不妥或不完善的地方，望广大读者给予批评指正，欢迎提出宝贵意见和建议.

编　者

2017 年 1 月

目　　录

第1章　概率与随机事件

§1.1　知识点归纳

1.1.1　随机现象和随机试验

自然界中的现象和人们在社会实践中发生的现象一般分为两类：一类是在一定条件下必然发生的现象，称为确定性现象．另一类是在一定条件下可能出现也可能不出现的现象，称为随机现象．随机现象在一次观察中出现什么结果具有偶然性，但在大量试验或观察中，这种结果的出现具有一定的统计规律性，概率论就是研究随机现象这种本质规律的一门数学学科．

随机现象是通过随机试验来研究的．在概率论中，把具有以下3个特征的试验称为随机试验：

（1）可以在相同的条件下重复地进行；

（2）每次试验的可能结果不止一个，并且能事先明确试验的所有可能结果；

（3）进行一次试验之前不能确定哪一个结果会出现．

1.1.2　样本空间与事件

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 S 或 Ω．样本空间的元素，即试验的每一个结果，称为样本点．试验不同，对应的样本空间也不同．同一试验，若试验目的不同，则对应的样本空间也不同．建立样本空间，事实上就是建立随机现象的数学模型．

随机试验 E 的样本空间 Ω 的子集称为 E 的随机事件，简称事件，以 A，B，C 来表示事件．由一个样本点组成的单点集称为基本事件．样本空间本身称为必然事件．空集称为不可能事件．

1.1.3　事件的关系和运算

随机事件间的关系可按照集合之间的关系来解决.

(1) **包含**　若事件 A 出现，必然导致 B 出现，记作 $A\subset B$，如图 1-1 所示.

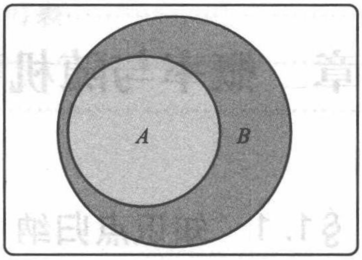

图 1-1

(2) **等价**　若事件 A 包含事件 B，而且事件 B 包含事件 A，则称事件 A 与事件 B 等价，记作 $A=B$.

(3) **交（积）事件**　事件 A 和事件 B 同时发生，记作 $A\cap B$ 或 AB，如图 1-2 所示. 可将交事件推广到有限个或可列个事件的情形. $\bigcap\limits_{i=1}^{n} A_i$ 表示 n 个事件 A_1，…，A_n 的交，$\bigcap\limits_{i=1}^{+\infty} A_i$ 表示可列个事件 A_1，A_2，…的交.

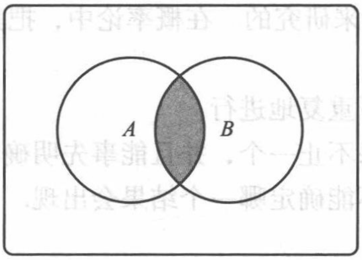

图 1-2

(4) **并（和）事件**　事件 A 和事件 B 至少有一个发生，记作 $A\cup B$，如图 1-3 所示. 可将并事件推广到有限个或可列个事件的情形. $\bigcup\limits_{i=1}^{n} A_i$ 表示 n 个事件 A_1，…，A_n 的并，$\bigcup\limits_{i=1}^{+\infty} A_i$ 表示可列个事件 A_1，A_2，…的并.

(5) **差事件**　事件 A 发生但是事件 B 不发生，称为事件 A 和事件 B 的差，记作 $A-B$.

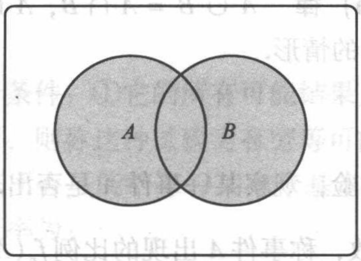

图 1－3

(6) **互不相容（互斥）事件** 事件 A 和事件 B 不能同时发生，记作 $A \cap B = \varnothing$，如图 1－4 所示.

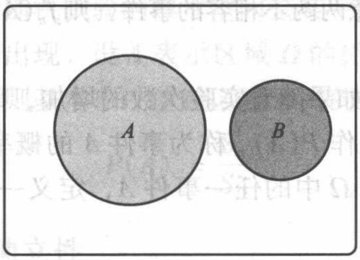

图 1－4

(7) **对立事件或逆事件** 设 A 表示“事件 A 出现”，则“事件 A 不出现”称为事件 A 的对立事件或逆事件，记作 $\bar{A}$，如图 1－5 所示.

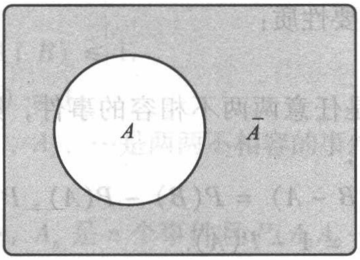

图 1－5

同样的，随机事件间的运算也可以按照集合之间的运算规律来解决.

(1) **交换律** $A \cup B = B \cup A, A \cap B = B \cap A.$

(2) **结合律** $A \cup (B \cup C) = (A \cup B) \cup C, A \cap (B \cap C) = (A \cap B) \cap C.$

(3) **分配律** $A \cup (B \cap C) = (A \cup B) \cap (A \cup C), A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$

(4) **德摩根 (De Morgan) 律** $\overline{A \cup B} = \bar{A} \cap \bar{B}$, $\overline{A \cap B} = \bar{A} \cup \bar{B}$. De Morgan 律可以推广到有限个或可列个事件的情形.

1.1.4 事件的概率

在相同条件下重复 n 次试验，观察某一事件 A 是否出现，称 n 次试验中事件 A 出现的次数 n_A 为事件 A 出现的频数，称事件 A 出现的比例 $f_n(A) = \dfrac{n_A}{n}$ 为事件 A 出现的频率. 事件 A 发生的频率 $f_n(A)$ 具有下列性质：

(1) $0 \leqslant f_n(A) \leqslant 1$;

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

(3) 设 A_1, …, A_m 是任意两两不相容的事件，则 $f_n(A_1 \cup A_2 \cup \cdots \cup A_m) = f_n(A_1) + f_n(A_2) + \cdots + f_n(A_m)$.

对于给定的随机事件 A，如果随着实验次数的增加，事件 A 发生的频率 $f_n(A)$ 稳定在某个常数上，把这个常数记作 $P(A)$，称为事件 A 的概率，简称为 A 的概率. 概率的公理化定义为，对于样本空间 Ω 中的任一事件 A，定义一个实单值集合函数 $P(A)$，满足下列 3 个条件：

(1) 非负性，即 $0 \leqslant P(A) \leqslant 1$;

(2) 正规性，即 $P(\Omega) = 1$;

(3) 可列可加性，设 A_1, A_2, …是任意两两不相容的事件，则 $P(A_1 \cup A_2 \cdots) = P(A_1) + P(A_2) + \cdots$.

此外，概率还具有一些重要性质：

(1) $P(\varnothing) = 0$;

(2) 设 A_1, A_2, …, A_n 是任意两两不相容的事件，则 $P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$;

(3) 设事件 $A \subset B$，则 $P(B - A) = P(B) - P(A)$, $P(A) \leqslant P(B)$;

(4) 对任一事件 A, $P(\bar{A}) = 1 - P(A)$.

(5) 对任意两个事件 A 和 B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. 该性质可以推广到多个事件情形，

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

(6) 对于任意的事件 A_1, A_2, …, 有 $P(A_1 \cup A_2 \cdots) \leqslant P(A_1) + P(A_2) + \cdots$.

1.1.5 古典概型

若随机试验满足下述两个条件：①它的所有可能结果只有有限多个基本事件；②每个基本事件出现的可能性相同，则称这种试验为有穷等可能随机试验或古典概型.

设试验E是古典概型，其所有可能结果Ω由n个基本事件组成，事件A由k个基本事件组成. 则定义事件A的概率为：

$$P(A)=\frac{k}{n}=\frac{A\text{所含基本事件个数}}{\Omega\text{所含基本事件总数}}$$

排列组合是计算古典概型中随机事件的概率的重要工具.

把等可能推广到无限个基本事件场合，人们引入了几何概型. 由此形成了确定概率的另一方法——几何方法. 设随机试验的样本空间是某一区域Ω，基本事件是区域中的一个点，并且在区域中等可能出现. 设A表示区域Ω的任意子区域，并设S_A、S_Ω分别表示区域A、区域Ω的度量，则基本事件落在区域A内的概率为：

$$P(A)=\frac{S_A}{S_\Omega}$$

1.1.6 条件概率、事件的独立性

设A，B为两个事件，且$P(B)>0$，称$P(A\mid B)=\dfrac{P(AB)}{P(B)}$为事件$B$发生的条件下事件$A$发生的条件概率. 条件概率同样满足概率定义中的非负性、正规性和可列可加性，即若$P(B)>0$，则满足：

(1) 非负性，即$0\leqslant P(A\mid B)\leqslant 1$.

(2) 正规性，即$P(\Omega\mid B)=1$.

(3) 可列可加性，即设A_1，A_2，…是两两不相容的事件，则$P(A_1\cup A_2\cup\cdots\mid B)=P(A_1\mid B)+P(A_2\mid B)+\cdots$.

乘法定理 设A_1，A_2，…，A_n是n个事件且$P(A_1A_2\cdots A_n)>0$，则有：

$$P(A_1\cap A_2\cap\cdots\cap A_n)=P(A_1)P(A_2\mid A_1)\cdots P(A_n\mid A_1A_2\cdots A_{n-1})$$

样本空间的划分 设$\{B_k:1\leqslant k\leqslant n\}$是一组事件且满足：

$$\bigcup_{k=1}^{n}B_k=\Omega,\ B_jB_k=\varnothing,\ j\neq k$$

则称$\{B_k:1\leqslant k\leqslant n\}$是样本空间的一个划分.

全概率公式 设$\{B_k:1\leqslant k\leqslant n\}$是样本空间的一个划分，且$P(B_k)>0$，$1\leqslant k\leqslant n$，则对任一事件$A$有：

$$P(A) = \sum_{k=1}^{n} P(A \mid B_k)P(B_k)$$

贝叶斯（Bayes）公式 设 $\{B_k: 1 \leqslant k \leqslant n\}$ 是样本空间的一个划分，且 $P(B_k) > 0, 1 \leqslant k \leqslant n, P(A) > 0$，则

$$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{\sum_{j=1}^{n} P(A \mid B_j)P(B_j)}, 1 \leqslant k \leqslant n$$

对事件 A，B，如果有

$$P(AB) = P(A)P(B)$$

则称事件 A 与事件 B 相互独立，简称事件 A，B 独立．若事件 A，B 独立，且 $P(B) > 0$，则 $P(A \mid B) = P(A)$．若事件 A，B 独立，则 $\bar{A}$ 与 B 独立，A 与 $\bar{B}$ 独立，$\bar{A}$ 与 $\bar{B}$ 独立．

对事件 A，B，C，如果以下 4 个等式均成立，

$$P(AB) = P(A)P(B)$$
$$P(AC) = P(A)P(C)$$
$$P(BC) = P(B)P(C)$$
$$P(ABC) = P(A)P(B)P(C)$$

则称事件 A，B，C 相互独立．若只有前 3 个等式成立，则称 A，B，C 两两独立．两两独立和相互独立是不同的概念．

§1.2 例题讲解

例 1 已知 A，B，C 是三个事件，且 $P(AB) = 0$，$P(AC) = P(BC) = \frac{1}{6}$，$P(A) = P(B) = P(C) = \frac{1}{4}$，求 A，B，C 全不发生的概率．

解：

$$\begin{aligned} P(\bar{A}\bar{B}\bar{C}) &= 1 - P(A \cup B \cup C) \\ &= 1 - [P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)] \\ &= \frac{7}{12} - P(ABC) \end{aligned}$$

因为 $ABC \subset AB, 0 \leqslant P(ABC) \leqslant P(AB) = 0$，所以 $P(ABC) = 0, P(\bar{A}\bar{B}\bar{C}) = \frac{7}{12}$．

注意： 一种错误的解法是，$P(AB) = 0 \Rightarrow AB = \varnothing \Rightarrow ABC = \varnothing \Rightarrow P(ABC) = 0$. 概率是 0 的事件是

有可能发生的. 由 $P(AB)=0$ 得不到 $AB=\varnothing$.

例2 一口袋装有6个球，其中4个白球、2个红球. 从袋中取球两次，每次随机地取一个. 考虑两种取球方式：

放回抽样 第一次取一个球，观察其颜色后放回袋中，搅匀后再取一个球；

不放回抽样 第一次取一个球不放回袋中，第二次从剩余的球中再取一个球.

分别就上述两种方式求：

(1) 取到的两个都是白球的概率；

(2) 取到的两个球中至少有一个是白球的概率.

解： 从袋中取两个球，每一种取法就是一个基本事件.

设 $A=$ "取到的两个都是白球"，$B=$ "取到的两个球中至少有一个是白球".

放回抽样：样本点总数是 $n=6^2$ 是排列数.

$$P(A)=\frac{4^2}{6^2}\approx 0.444,\ P(B)=1-P(\bar{B})=1-\frac{2^2}{6^2}\approx 0.889$$

注意： $P(B)\neq\frac{P_4^1P_6^1}{6^2}=\frac{P_4^1P_4^1+P_4^1P_2^1}{6^2}$，这种算法只考虑了第一次是白球，漏算了第一次是红球，第二次是白球的情形.

R程序和输出：

```
>urn< -c(1,1,1,1,0,0)
>sims< -10000
>success< -NULL
>for(i in 1:sims){
+draw< -sample(urn,2,replace=T)
+success[i]< -sum(draw)= =2
+}
>mean(success)
[1]0.4437
>
>success2< -NULL
>for(i in 1:sims){
+draw< -sample(urn,2,replace=T)
+success2[i]< -sum(draw)>0
+}
```

```
>mean(success2)
[1]0.8867
```

不放回抽样：

$$P(A) = \frac{C_4^2}{C_6^2} = 0.4,\ P(B) = 1 - P(\bar{B}) = 1 - \frac{C_2^2}{C_6^2} \approx 0.933$$

B 的概率也可以如下计算，

$$P(B) = \frac{C_4^2 + C_4^1 C_2^1}{C_6^2} \approx 0.933$$

R 程序和输出：

```
>urn<-c(1,1,1,1,0,0)
>sims<-10000
>success<-NULL
>for(i in 1:sims){
+draw<-sample(urn,2,replace=F)
+success[i]<-sum(draw)==2
+}
>mean(success)
[1]0.3951
>
>success2<-NULL
>for(i in 1:sims){
+draw<-sample(urn,2,replace=F)
+success2[i]<-sum(draw)>0
+}
>mean(success2)
[1]0.9305
```

例 3 将 n 个球随机地放入 $N(N \geqslant n)$ 个盒子中去，求每个盒子至多有一个球的概率（设盒子的容量不限）.

解： n 只球看成是 n 个不同的球，N 个盒子也看成是 N 个不同的盒子，设每个球都以等可能性放在各个盒子中，所以每个球都有 N 种不同放法，n 个球总共有 N^n 种放法.

N^n 是从 N 个不同元素中取 n 个元素允许重复的排列，每一种放法即为一种排列，并且是一基本事件. 每个盒子中至多有一个球的概率为

$$p = \frac{N \times (N-1) \times \cdots \times [N-(n-1)]}{N^n} = \frac{\mathrm{P}_N^n}{N^n}$$

至少有两个球放在同一盒子中的概率为

$$q = 1 - p = 1 - \frac{\mathrm{P}_N^n}{N^n}$$

注意：该数学模型可用于许多实际问题，如生日问题、住房问题、车站下车问题等. $n(n \leqslant 365)$ 个人在365天的生日，可看成是 n 个球放入365个盒子中. 随机取 $n(n \leqslant 365)$ 人他们的生日各不相同的概率为

$$P(A) = \frac{365 \times 364 \times \cdots \times (365-n+1)}{365^n}$$

因而，n 个人中至少有两人生日相同的概率为

$$1 - P(A) = 1 - \frac{365 \times 364 \times \cdots \times (365-n+1)}{365^n} = 1 - \frac{\mathrm{P}_{365}^n}{365^n}$$

可以计算"在一个有64人的班级里，至少有两人生日相同"的概率为99.7%.

R 程序和输出：

```
>n< -64
>1 -prod((365:(365 -n +1))/365)
[1]0.9971905
>
>
>sim< -10000
>x< -numeric(sim)
>for(i in 1:sim)
+{a< -sample(1:365,n,replace =T)
+x[i]< -n -length(unique(a))
+}
>1 -mean(x = =0)
[1]0.9963
```

例4 在1~2 000的整数中随机地取一个数，问取到的整数既不能被6整除，又不