高等学校数学辅导教材之 3

概率论与数理统计辅导讲义

编著 龚兆仁 王雪标

国家行政学院出版社

高等学校数学辅导教材之③

概率论与数理统计
辅导讲义

编著　龚兆仁　王雪标

国家行政学院出版社

图书在版编目（CIP）数据

概率论与数理统计辅导讲义/龚兆仁，王雪标编著. －北京：国家行政学院出版社，2001.8
ISBN 7-80140-182-4

Ⅰ.概…　Ⅱ.①龚…　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　Ⅳ.021

中国版本图书馆 CIP 数据核字（2001）第 053901 号

概率论与数理统计辅导讲义

龚兆仁　王雪标　编著

*

国家行政学院出版社出版发行
北京市海淀区长春桥路 6 号
邮政编码：100089
发行部电话：68920615，68929949
新华书店经销
北京市朝阳印刷厂印刷

*

850×1168　1/32 开本　15.25 印张　490 千字
2002 年 10 月北京第 1 版　2002 年 10 月第 1 次印刷
ISBN 7-80140-182-4/O·17　　定价：18.00 元

前　言

概率论与数理统计是研究随机现象统计规律的一门数学学科.概率论是对随机现象统计规律演绎的研究,即从理论上是对随机现象进行研究,而数理统计则是对随机现象统计规律归纳的研究,即从随机现象的简单随机样本对总体进行统计推断的研究.

概率论是建立在随机事件这个概念基础上的.理解、掌握概率论的基本概念及其实际背景、正确分析给定随机试验中的随机事件(或随机变量)之间的关系,并选择合适的等价表示形式,是学好概率论的关键.数理统计则应着重于掌握其统计思想,它是所有不同的统计方法的依据.初等概率论(非数学专业)的问题主要是计算概率、数字特征及求随机变量 X 的分布(即是计算概率($P\{X \leqslant x\}$).其解题方法和技巧并不多,主要是通过对随机试验的分析,把所叙述的问题转换成概率论语言,即表示成事件、概率、条件概率、数字特征、随机变量的分布等,而后选择合适的概率模型及正确的计算公式,最后经过较为简单的数学运算就可以解决问题了.由于概率统计有非常强的实际背景与直观意义,这有利于我们理解和想象,有利于我们学好、掌握好这门课程.

全书共分九章.前八章均设有**知识结构网络图**、**基本内容诠释与重要结论归纳**、**重点**、**难点提示**,希望能帮助同学把握住该章的核心,掌握基本概念的涵义及重要公式、定理的应用,通过**典型例题分析**,或是澄清基本概念与基本运算,或是指出同学常犯的错误,或是介绍概率论与数理统计中常用思路与方法,并且许多题目给出一题多解,通过以上这些希望能帮助同学开阔思路,活跃思维,举一反三,触类旁通.同学做各章设置的**练习题**可达到巩固、理解、提高的目的.同学在做练习题时,一定要独立思考,动手做题,实在有困难再看提示和参考答案.第九章为总复习,通过**典型的综合例题分析**,希望帮助同学提高综合分析问题、解决问题以及逻辑

思维的能力;通过精编的**水平测试题**,帮助同学检查前阶段复习之效果,查漏补缺.

编写本书时,依据大学本科现行教材及教学大纲的基本要求,参考了**北京大学、清华大学、同济大学、浙江大学、四川大学、西安交通大学、上海交通大学、东北财经大学**等高等院校的教材,并结合硕士研究生入学考试《数学考试大纲》的要求进行编写.

由于编者水平所限,疏漏错误难免,欢迎广大同仁、读者批评指正.

编者

2002 年 10 月

目　　录

第一章 随机事件与概率

本章知识结构网络图

- 随机现象—随机试验—随机事件
  - 表述方式
    - 语言叙述
    - 点集：基本事件 ω；样本空间 $\Omega=\{\omega\}$；事件 $A\subset\Omega$，必然事件，不可能事件
    - 关系
      - 包含、相等、互不相容、对立、完备事件组
      - 相互独立
        - 定义：描述性定义与严格定义
        - 性质：① 相互独立必两两独立；② 改若干个为对立事件仍相互独立；③ 不含相同事件的事件组相互独立.
    - 运算
      - 形式：加(并)，交(积)，减，逆；
      - 法则：交换、结合、分配、对偶、吸收，分解
  - 概率
    - 定义
      - 描述性定义：度量事件发生可能性大小的非负数值；
      - 统计定义：频率 $f_A=\mu A/n$ 的稳定值；
      - 古典定义：$P(A)=n(A)/n(\Omega)$；
      - 几何定义：$P(A)=\angle(A)/\angle(\Omega)$；
      - 公理化定义：非负、规范、可列可加的事件函数.
    - 性质
      - 非负、有界、单调不减、连续、有限可加、半可加
      - 加法公式：$P(A\cup B)=P(A)+P(B)-P(AB)$
      - 减法公式：$P(A-B)=P(A\bar{B})=P(A)-P(AB)$
      - 求逆公式：$P(\bar{A})=1-P(A)$
    - 条件概率
      - 定义
      - 三公式
        - 乘法公式 $P(AB)=P(A)P(B\mid A)(P(A)>0)$
        - 全概公式 $P(B)=\sum\limits_i P(A_i)P(B\mid A_i)$
        - Bayes 公式 $$P(A_i\mid B)=\frac{P(A_i)P(B\mid A_i)}{\sum\limits_i P(A_i)P(B\mid A_i)}$$
    - 概型
      - 等可能概型
        - 古典概型
        - 几何概型
      - 独立试验序列概型(贝努里概型、二项分布概型)

        n 次独立重复试验，事件 A 发生 k 次的概率为 $C_n^k p^k(1-p)^{n-k}$

基本内容诠释与重要结论归纳

一、随机试验与随机事件

1. 随机试验

称一个试验为随机试验,如果

(1) 试验可以在相同条件下重复进行;

(2) 试验所有可能结果是明确可知道的,并且不止一个;

(3) 每次试验会出现哪一个结果事先不能确定.

我们是通过研究随机试验来研究随机现象的,为方便起见,将随机试验简称为试验,并用字母 E 或 $E_1,E_2,\cdots$ 表示.

2. 随机事件,基本事件与样本空间

随机事件 在一次试验中可能出现,也可能不出现的结果称为随机事件,简称为事件,并用大写字母 A,B,C 等表示;

基本事件 在一次试验中,每一个可能出现的最简单、最基本的结果称为基本事件,在每次试验中能且只能发生该试验的一个基本事件.基本事件有时也称为样本点,常用 ω 表示.

基本事件空间(或样本空间) 一个试验所有基本事件(样本点)的全体组成的集合,称为基本事件空间(样本空间),通常用 Ω 表示,即 $\Omega=\{\omega\}$.

必然事件 每次试验一定发生的事件称为必然事件,记为 Ω.

不可能事件 每次试验一定不发生的事件,称为不可能事件,记为 $\varnothing$.

随机事件与集合 从集合论的观点来看,如果把样本空间 $\Omega=\{\omega\}$ 作为全集(基本集),那么基本事件 ω 就是 Ω 中的无素(记为 $\omega\in\Omega$).随机事件 A 总是由若干个具有某些特性的基本事件所组成,因此事件 A 就为 Ω 的子集,即 $A\subset\Omega$.事件 A 发生等价于构成 A 的基本事件中有一个发生.不可能事件 $\varnothing$ 是不包含 Ω 中任何一个元素的空集,在任何一次试验中,必然要出现 Ω 中的某一基本事件 ω,即 Ω 必然会发生,因此 Ω 为必然事件.为方便起见,我们把不可能事件,必然事件作为随机事件两个极端情况,并约定 $\varnothing\subset A\subset\Omega$,这样就将事件与集合等同起来,它们之间的关系和运算也是一致的,这对现代概率论的研究是极为方便、必要的.

3. 随机事件间关系与运算

研究随机事件的规律，总是通过对较简单的事件规律的研究去掌握复杂事件的规律.为此需要研究事件之间的关系与运算.

1. **包含**:$A \subset B$

称事件B包含事件A(记为$A \subset B$)，如果事件A发生必然导致事件B发生，即A为B的子集:若$\omega \in A$，必有$\omega \in B$.

2. **相等**:$A = B$

称事件A与B相等(记为$A = B$)，如果$A \subset B$且$B \subset A$，即A与B是由相同的元素构成:$\omega \in A \Leftrightarrow \omega \in B$.

3. **互不相容(互斥)**:$AB = \varnothing$

称事件A与B互不相容(互斥)(记为$AB = \varnothing$)，如果A与B不能同时发生，即$\{\omega : \omega \in A$ 且 $\omega \in B\} = \varnothing$.

称可列个(或有限个)事件$A_1, A_2, \cdots, A_n, \cdots$是互不相容的，如果$A_iA_j = \varnothing (i \neq j, i, j \geqslant 1)$，即两两互不相容.

4. **和(并)**:$A \cup B$, $\bigcup_{i=1}^{n} A_i$, $\bigcup_{i=1}^{\infty} A_i$

称"事件A与B至少有一个发生"的事件为事件A与B的和(并)，记为$A \cup B$，即$A \cup B = \{\omega : \omega \in A$ 或 $\omega \in B\}$.

称"有限个事件$A_1, A_2, \cdots, A_n$至少有一个发生"的事件为事件$A_1, A_2, \cdots A_n$的和，记为$\bigcup_{i=1}^{n} A_i$，即$\bigcup_{i=1}^{n} A_i = \{\omega$:存在$i (1 \leqslant i \leqslant n)$，使$\omega \in A_i\}$.

类似地可定义可列个事件的和$\bigcup_{i=1}^{\infty} A_i = \{\omega$:存在$i (i \geqslant 1)$，使$\omega \in A_i\}$，即$\bigcup_{i=1}^{\infty} A_i$是可列个事件$A_1, A_2, \cdots$中至少有一个发生的事件.

5. **积(交)**:$A \cap B$(简记为AB), $\bigcap_{i=1}^{n} A_i$, $\bigcap_{i=1}^{\infty} A_i$

称"事件A与B同时发生"的事件为事件A与B的积(交)，记为$A \cap B$或AB，即$AB = \{\omega : \omega \in A$ 且 $\omega \in B\}$.

称"有限个事件$A_1, A_2, \cdots, A_n$同时发生"的事件为事件$A_1, A_2, \cdots, A_n$的积，记为$\bigcap_{i=1}^{n} A_i$，即$\bigcap_{i=1}^{n} A_i = \{\omega$:对一切$i (1 \leqslant i \leqslant n)$有，$\omega \in A_i\}$.

类似地可定义$\bigcap_{i=1}^{\infty} A_i$为"可列个事件$A_1, A_2, \cdots A_n \cdots$同时发生"的事件.

6. **差**:$A - B$

称"事件A发生而B不发生"的事件为事件A与B的差，记为$A - B$，即

$A-B=\{\omega:\omega\in A$ 但 $\omega\overline{\in}B\}$.由定义知 $A-B=A-AB=A\bar{B}$.

7. 逆(对立):$\bar{A}$

称"事件 A 不发生"的事件为事件 A 的逆事件,或对立事件,记为 $\bar{A}$,即 $\bar{A}=\{\omega:\omega\overline{\in}A\}=\Omega-A$.由定义知 $\bar{A}=B\Leftrightarrow AB=\varnothing$ 且 $A\cup B=\Omega$.

8. 完备事件组:$\{A_n,n\geqslant1\}$

称可列个(或有限个)事件 $A_1,A_2,\cdots,A_n\cdots$ 构成一个完备事件组,如果 $\bigcup\limits_i A_i=\Omega$,且 $A_iA_j=\varnothing$(一切 $i\neq j$).有时也称 $\{A_n,n\geqslant1\}$ 为 Ω 的一个不交的分割.显然,基本事件 $A_i=\{\omega_i\}$ 全体构成一个完备事件组,反之不然.

4. 事件运算法则

事件运算与集合运算具有相同的运算法则:

吸收律:("并"取大,"交"取小):若 $A\subset B$,则 $A\cup B=B,AB=A$.

交换律:$A\cup B=B\cup A,\quad AB=BA$.

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

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

对偶律(De · Morgan 法则):

$$\overline{A\cup B}=\bar{A}\cap\bar{B},\qquad \overline{A\cap B}=\bar{A}\cup\bar{B};$$

$$\overline{\bigcup_i A_i}=\bigcap_i\bar{A}_i,\qquad \overline{\bigcap_i A_i}=\bigcup_i\bar{A}_i\quad(i\geqslant1).$$

分解律:

(1) 直和分解式 —— 将事件的并分解为互不相容事件的并.

$$A_1\cup A_2=A_1\cup(A_2-A_1)=A_1\cup A_2\bar{A}_1=A_1\bar{A}_2\cup A_2$$
$$=A_1\bar{A}_2\cup A_1A_2\cup\bar{A}_1A_2,$$
$$A_1\cup A_2\cup A_3=A_1\cup(A_2-A_1)\cup(A_3-A_1-A_2)$$
$$=A_1\cup A_2\bar{A}_1\cup A_3\bar{A}_1\bar{A}_2,$$

$\bigcup\limits_{n=1}^{\infty}A_n=\sum\limits_{n=1}^{\infty}B_n$(互不相容事件的并常用 $\sum\limits_i B_i$ 表示,有时也用 $\bigcup\limits_i B_i$).

其中 $A_0=\varnothing,B_n=A_n-\bigcup\limits_{k=0}^{n-1}A_k=A_n\bigcap\limits_{k=0}^{n-1}\bar{A}_k\subset A_n(n\geqslant1),B_iB_j=\varnothing(i\neq j)$.

(2) 全集分解式 —— 事件对完备事件组的分解.

$$A=AB+A\bar{B}.$$

若 $\bigcup\limits_i B_i=\Omega,B_iB_j=\varnothing(i\neq j)$,则 $A=\sum\limits_i AB_i$.

全集分解式给出了构造事件 A 与 B 联系的一个公式,我们可以借助于

A、B 之间的关系，分析、讨论事件 A，这在计算概率、求随机变量分布时，常常被用到.

【评注】 ① 事件的关系与运算非常重要，这不仅在讨论各事件间关系时经常用到，而且在概率计算，求随机变量的分布中，也经常需要将一些事件用另一些事件运算关系来表示，或作等价变换，或进行某种必要的变形，或考虑包含关系.要学会用概率论的语言来解释这些关系及运算，并会用这些运算关系来表示事件.

② 在分析和理解事件间的关系时，常常借助于图示法即文(Venn)图，既直观又简便.

③ 在事件运算法则中，吸收律、对偶律及分解律，在事件的化简运算及概率计算中常常起着极为重要的作用，应掌握并会应用.

④ 减法运算常化为积考虑，即 $A-B=A\bar{B}$；$A-B-C=A\bar{B}\,\bar{C}=A(\overline{B\cup C})$. 此外，应注意等价关系：$A\bar{B}=\varnothing \Leftrightarrow A\subset B \Leftrightarrow \bar{A}\supset\bar{B} \Leftrightarrow AB=A \Leftrightarrow A\cup B=B$.

二、随机事件的概率

对于一个试验，我们不仅关心它可能出现哪些结果，而更为重要的是要知道这些结果即随机事件发生的可能性大小.随机事件 A 发生可能性大小的度量(非负值)，称为 A 发生的概率，记作 $\boldsymbol{P}(A)$.

1. 概率的统计定义

在不变的一组条件下，独立重复地进行了几次试验，事件 A 发生的次数 μ_A 称为事件 A 发生的频数，比值 $f_n(A)\triangleq\frac{\mu_A}{n}$ 称为 A 发生的频率.当试验次数 n 增大时，频率 $f_n(A)$ 呈现出某种稳定性，即它在某一常数 p 附近摆动，随着 n 的增大，摆动的幅度越来越小，我们称这个客观存在的频率稳定值 p 为事件 A 的概率，即 $\boldsymbol{P}(A)=p$.

【评注】 ① "频率稳定性"的确切含义，大数定律将作进一步阐述；

② 在很多情况下，无法确定出 p 值，通常多是用频率值 $f_n(A)=\frac{\mu_n}{n}$ 近似地估计概率 $\boldsymbol{P}(A)$.例如，某种疾病的死亡率，种子的发芽率等等.

2. 古典概型与概率的古典定义

称随机试验的概率模型为**古典概型**，如果(1) 试验的基本事件总数是有

限个的;(2) 试验的每个基本事件发生的可能性都一样.

如果古典概型随机试验的基本事件总数为 n,事件 A 包含 k 个基本事件,则 A 发生的概率定义为

$$P(A) \triangleq \frac{k}{n} = \frac{A\text{ 中所含基本事件个数}}{\text{试验的基本事件总数}} \xlongequal{\text{记}} \frac{n(A)}{n(\Omega)}.$$

由此式计算的概率称为**古典概率**.

3. 几何概型与概率的几何定义

称随机试验的概率模型为几何概型,如果(1) 试验的样本空间 Ω 是一个可度量的几何区域(这个区域可以是一维、二维,甚至是 n 维的);(2) 试验的每个基本事件发生的可能性都一样,即样本点落入 Ω 的某一个可度量的子集 A 的可能性大小与 A 的几何测度成正比,而与 A 的位置及形状无关.

在几何概型随机试验中,A 为 Ω 的一个可度量的子集,则随机事件"点落入区域 A"的概率定义为

$$P(\mathrm{A}) \triangleq \frac{\mathrm{A}\text{ 的几何测度}}{\Omega\text{ 的几何测度}} = \frac{\angle(\mathrm{A})}{\angle(\Omega)}.$$

其中 $\angle(\mathrm{A})$ 表示区域 A 的几何测度,可以是长度、面积或体积.由上式计算的概率称为几何概率.

4. 概率的公理化定义

设试验 E 的样本空间为 Ω,对于 E 的每一个事件 A 都赋于一个确定的实数 $P(A)$,如果事件函数 $P(\cdot)$ 满足:

(1) 非负性:$P(A) \geqslant 0$;

(2) 规范性:$P(\Omega) = 1$;

(3) 可列可加性:对任意可列个互不相容事件 $A_1, A_2, \cdots, A_n, \cdots$ 有

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

则称 $P(\cdot)$ 为概率,$P(A)$ 为事件 A 的概率.

三、概率的性质

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

2. 有限可加性:设 $A_1, A_2, \cdots A_n$ 是两两互不相容的事件,则

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

3. $P(\overline{A}) = 1 - P(A)$;

4. **减法公式**:若 $A \subset B$,则 $\boldsymbol{P}(B-A)=\boldsymbol{P}(B)-\boldsymbol{P}(A)$,且 $\boldsymbol{P}(A) \leqslant \boldsymbol{P}(B)$;一般的减法公式为 $\boldsymbol{P}(B-A)=\boldsymbol{P}(B)-\boldsymbol{P}(AB)$.

5. **加法公式**:$\boldsymbol{P}(A \cup B)=\boldsymbol{P}(A)+\boldsymbol{P}(B)-\boldsymbol{P}(AB)$.一般地,对于任意 n 个事件 $A_1,A_2,\cdots,A_n$,有

$$\boldsymbol{P}\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} \boldsymbol{P}(A_i)-\sum_{1 \leqslant i<j \leqslant n} \boldsymbol{P}(A_iA_j)+\sum_{1 \leqslant i<j<k \leqslant n} \boldsymbol{P}(A_iA_jA_k)-\cdots\cdots +(-1)^{n-1} \boldsymbol{P}(A_1,A_2,\cdots A_n).$$

例如,$\boldsymbol{P}(A \cup B \cup C)=\boldsymbol{P}(A)+\boldsymbol{P}(B)+\boldsymbol{P}(C)-\boldsymbol{P}(AB)-\boldsymbol{P}(AC)-\boldsymbol{P}(BC)+\boldsymbol{P}(ABC)$,

由此可知, $\boldsymbol{P}(A \cup B) \leqslant \boldsymbol{P}(A)+\boldsymbol{P}(B)$.

由归纳法可证得 $\boldsymbol{P}\left(\bigcup_{n=1}^{n} A_n\right) \leqslant \sum_{i=1}^{n} \boldsymbol{P}(A_i)$.

6. **半可加性**:$\boldsymbol{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leqslant \sum_{n=1}^{\infty} \boldsymbol{P}(A_n)$.

【证明】 由直和分解式 $\bigcup_{n=1}^{\infty} A_n=\sum_{n=1}^{\infty} B_n$,其中 $B_n=A_n-\bigcup_{k=0}^{n-1} A_k \subset A_n$,$A_0=\varnothing$,$B_iB_j=\varnothing(i \neq j)$,故 $\boldsymbol{P}(B_n) \leqslant \boldsymbol{P}(A_n)$,且

$$\boldsymbol{P}\left(\bigcup_{n=1}^{\infty}\right)=\boldsymbol{P}\left(\sum_{n=1}^{\infty} B_n\right)=\sum_{n=1}^{\infty} \boldsymbol{P}(B_n) \leqslant \sum_{n=1}^{\infty} \boldsymbol{P}(A_n).$$

7. **连续性定理**

下连续性:如果事件序列 $A_1,A_2,\cdots,A_n\cdots$,满足 $A_1 \subset A_2 \subset A_3 \subset \cdots$,记 $A=\bigcup_{n=1}^{\infty} A_n$,则 $\lim\limits_{n \to \infty} \boldsymbol{P}(A_n)=\boldsymbol{P}\left(\bigcup_{n=1}^{\infty} A_n\right)=\boldsymbol{P}(A)$.

上连续性:如果事件序列 $A_1,A_2,\cdots,A_n,\cdots$ 满足,$A_1 \supset A_2 \supset A_3 \supset \cdots$,记 $A=\bigcap_{n=1}^{\infty} A_n$,则 $\lim\limits_{n \to \infty} \boldsymbol{P}(A_n)=\boldsymbol{P}\left(\bigcap_{n=1}^{\infty} A_n\right)=\boldsymbol{P}(A)$.

【证明】 如果 $A_n \subset A_{n+1}$,则由直和分解式 $A=\bigcup_{n=1}^{\infty} A_n=\sum_{n=1}^{\infty} B_n$,其中 $B_n=A_n-\bigcup_{k=0}^{n-1} A_k=A_n-A_{n-1}$,$A_0=\varnothing$,$B_iB_j=\varnothing(i \neq j)$,$\bigcup_{k=1}^{n} B_k=A_n$,根据可列可加性,得下连续性

$$\boldsymbol{P}(A)=\boldsymbol{P}\left(\bigcup_{n=1}^{\infty} A_n\right)=\boldsymbol{P}\left(\bigcup_{n=1}^{\infty} B_n\right)=\sum_{n=1}^{\infty} \boldsymbol{P}(B_n)=\lim_{n \to \infty} \sum_{k=1}^{n} \boldsymbol{P}(B_k)$$
$$=\lim_{n \to \infty} \boldsymbol{P}\left(\bigcup_{k=1}^{n} B_k\right)=\lim_{n \to \infty} \boldsymbol{P}(A_n).$$

如果 $A_1 \supset A_2 \supset A_3 \supset \cdots$，则 $\bar{A}_1 \subset \bar{A}_2 \subset \bar{A}_3 \subset \cdots$，由下连续性，得

$$P\left(\bigcup_{n=1}^{\infty} \bar{A}_n\right) = \lim_{n\to\infty} P(\bar{A}_n) = 1 - \lim_{n\to\infty} P(A_n).$$

故 $$\lim_{n\to\infty} P(A_n) = 1 - P\left(\bigcup_{n=1}^{\infty} \bar{A}_n\right) = P\left(\overline{\bigcup_{n=1}^{\infty} \bar{A}_n}\right) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = P(A).$$

四、条件概率及与其有关的公式：乘法公式、全概公式、贝叶斯(Bayes)公式

1. 条件概率

设 A、B 为任意两个事件，若 $P(A) > 0$，我们称在已知事件 A 发生的条件下，事件 B 发生的概率为条件概率，记为 $P(B \mid A)$，并定义

$$P(B \mid A) \triangleq \frac{P(AB)}{P(A)}.$$

容易验证，条件概率 $P(\cdot \mid A)$ 满足概率公理化定义中的三个条件，因此条件概率也是一种概率，概率的一切性质和重要结果都适用于条件概率. 例如：$P(\bar{B} \mid A) = 1 - P(B \mid A)$，

$$P(B_1 \cup B_2 \mid A) = P(B_1 \mid A) + P(B_2 \mid A) - P(B_1 B_2 \mid A).$$

2. 乘法公式

如果 $P(A) > 0$，则 $P(AB) = P(A)P(B \mid A)$.

对于 $n(n \geqslant 2)$ 个事件 $A_1, A_2, \cdots, A_n$，如果 $P(A_1 A_2 \cdots A_{n-1}) > 0$，则

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 A_2)\cdots P(A_n \mid A_1 A_2 \cdots A_{n-1}).$$

3. 全概公式

如果事件组 $A_1, A_2, \cdots, A_n$ 是一个完备事件组，即 $\bigcup_{i=1}^{n} A_i = \Omega$，$A_i A_j = \varnothing (i \neq j)$，且 $P(A_i) > 0 (1 \leqslant i \leqslant n)$，则对任一事件 B，有

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

4. 贝叶斯(Beyes)公式(逆概公式)

如果事件组 $A_1, A_2, \cdots, A_n$ 是一个完备事件组，且 $P(A_i) > 0 (1 \leqslant i \leqslant n)$，则对任一事件 B，只要 $P(B) > 0$，有

$$P(A_i \mid B) = \frac{P(A_i) P(B \mid A_i)}{\sum_{k=1}^{n} P(A_i) P(B \mid A_i)}, 1 \leqslant i \leqslant n.$$

【评注】 ① 乘法公式与条件概率公式在 $P(A)>0$ 时，实际上是一个公式. 在解决实际问题时，要区分 $P(AB)$ 与 $P(B\mid A)$ 的不同. $P(B\mid A)$ 是表示在 A 已经发生的条件下，B 发生的可能性，此时样本空间已从 Ω 缩减为 A 了. 因此计算 $P(B\mid A)$ 时，可以在 Ω 中用定义的公式进行计算，也可以在 A 中考虑 B 发生的可能性大小. 而 $P(AB)$ 则表示在样本空间 Ω 中 A 与 B 同时发生的可能性. 例如，袋中有 8 个球，其中白球 3 个，黑球 5 个. 随意从袋中先后取出两球(取后不放回)，记 $A_i=$"第 i 次取到的球是白球"$(i=1,2)$. 那么，"第二次取到白球"的事件可以表示为 $A_2=A_1A_2\cup\overline{A}_1A_2$；"二次都取到白球"的事件为 A_1A_2；"第二次才取到白球"的事件为 $\overline{A}_1A$；"二次中仅有一次取到白球"的事件为 $A_1\overline{A}_2+\overline{A}_1A_2$；"二次抽取中取到了白球"的事件为 $A_1\cup A_2$；"已知第一次取到了白球，那么第二次也取到白球"的事件是条件事件，表示为"$A_2\mid A_1$". 只要题目中有前提条件："在 A 发生的条件下"或"已知 A 发生"等等，均要考虑条件事件.

计算同时发生的概率可以用乘法公式，也可以在样本空间 Ω 中考虑同时发生的可能性. 例如：上例中

$$P(\overline{A}_1A_2)=P(\overline{A}_1)P(A_2\mid\overline{A}_1)=\frac{5}{8}\cdot\frac{3}{7}=\frac{15}{56},$$

或 $$P(\overline{A}_1A_2)=\frac{C_5^1\cdot C_3^1}{8\cdot 7}=\frac{15}{56}.$$

而在计算条件概率时，则应注意样本空间的不同. 例如：上例中

$$P(A_2\mid A_1)=\frac{P(A_1A_2)}{P(A_1)}=\frac{P(A_1A_2)}{P(A_1A_2)+P(A_1\overline{A}_2)}$$
$$=\frac{3\cdot 2/8\cdot 7}{3\cdot 2/8\cdot 8+3\cdot 5/8\cdot 7}=\frac{2}{7}.$$

此时应用条件概率的定义公式计算，是在样本空间 Ω 中考虑的. 如果在缩减的样本空间中计算条件概率，那么在 A_1 已发生的条件下，意味着袋中少一个白球，因此再从袋中任取一球是白球的概率 $P(A_2\mid A_1)=\dfrac{2}{7}$. 如果仍以先后取出的两球作为基本事件，那么在 A_1 已发生的条件下，缩减的样本空间中的基本事件是那些含有白球的基本事件全体，其总数为 $C_3^1C_7^1=21$(由于取球有先后之分，因而基本事件要考虑顺序). 而有利的基本事件则是两个全是白球的结果，其总数为 $C_3^1C_2^1=6$，故 $P(A_2\mid A_1)=\dfrac{6}{21}=\dfrac{2}{7}$.

② 全概公式是用于计算某个"结果"B 发生的可能性大小. 而该"结果"