Basics of Probs and Stats

1. 随机事件与概率

梁雪吟

地球科学与资源学院
liangxueyin@email.cugb.edu.cn

March 17, 2023

Outline

1.1 概率论基础

样本空间和随机事件

- 随机现象：事先不可预知结果的现象（天气、掷硬币）
- 随机试验：相同条件下随机观察、重复测量（可重复的随机现象）
- 样本点：随机试验的结果，用 ω 表示
- 样本空间：所有样本点组成的集合，用 Ω 表示
- 随机事件：样本空间 Ω 的一个子集

事件和集合的语言

- $A \subset B$ or $B \supset A$： A 发生 B 也发生
- $A + B$ or $A \cup B$： A、B 至少发生一个
- AB or $A \cap B$： A、B 同时发生
- $A - B$ or A/B： A 发生，B 不发生。$A - B = A \cup \bar{B}$
- $\bar{A}$: A 的对立事件（逆事件）表示 A 不发生。
- $A \cap B = \varnothing$: AB 不相容，不会同时发生。

Tips:

通过 Venn 图理解记忆

事件和集合的语言

Question 1.1

写出连续扔两枚硬币的样本空间

Question 1.2

利用事件运算表示下列各事件：

- ABC 不多于一个出现
- AB 至少有一个出现，C 不出现

事件和集合的语言

Question 1.1

写出连续扔两枚硬币的样本空间:
假设 1 表示正面，0 表示反面。
$\Omega = \{(0,0),(0,1),(1,0),(1,1)\}$

Question 1.2

利用事件运算表示下列各事件:

- ABC 不多于一个出现:
  $A\bar{B}\bar{C} + \bar{A}\bar{B}C + \bar{A}B\bar{C} + \bar{A}\bar{B}\bar{C}$
- AB 至少有一个出现，C 不出现:
  $(A+B)\bar{C}$

事件的运算法则

- **分配律**： $A\cap(B\cup C)=(AB)\cup(AC)$

  $A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$
- **交换律**： $A\cap B=B\cap A \quad A\cup B=B\cup A$
- **结合律**： $A\cap(B\cap C)=(A\cap B)\cap C \quad A\cup(B\cup C)=(A\cup B)\cup C$
- **对偶律 (De-Morgan 公式)**： $\overline{A\cap B}=\bar{A}\cup\bar{B} \quad \overline{A\cup B}=\bar{A}\cap\bar{B}$

1.2 概率的定义及其确定方法

公理

- **非负性**: **对于任意的事件** A, $P(A) > 0$;
- **正则性**: **对于必然事件** Ω, $P(\Omega) = 1$;
- **可列可加性**: **对于可列个互不相容的事件**, $A_1, A_2, ..., A_n, ...$, **有** $P(\sum_i A_i) = \sum_i P(A_i)$.

概率的性质

- $P(A) < 1$
- $P(\varnothing) = 0$
- (有限可加性)：对于有限个互不相容的事件，$A_1, A_2, ..., A_n$ 有 $P(\sum\limits_{i=1}^{n} A_i) = \sum\limits_{i=1}^{n} P(A_i)$;
- 若 $B \subset A$，则 P(A − B) = P (A) − P (B);
- 加法公式：P(A + B) = P(A) + P(B) − P(AB);

加法原理和乘法原理

- 分类相加
  1. 做成某件事有 n 种途径可选择，每种途径有 m_n 种完成方法。
  2. 从 A 地前往 B 地有三种交通工具：飞机，火车，汽车可选择，一天类分别的班次为：3、5、8，所以旅行者到达 B 地可选择 $8+5+3=16$ 种班次。
- 分步相乘（一件事一步不能做完）
  1. 做成某件事需要通过 n 个步骤完成，每种步骤有 m_n 种完成方法。
  2. 从 A 地前往 B 有 3 条路，B 地前往 C 地有 2 条路，从 A 地前往 C 地（必须经过 B 地）有 $3\times2=6$ 条路线可选择。
- P12 例 3

Question 2.1

1. 假设学号由八位数字组成，但第一位不能是 0 或 1. 一共可以有多少个不同的号码呢?
2. 小明在这些号码中随便选择一个，选中认识的同学的号码的概率有多大 (假定每位同学都有且仅有一个号码，其中小明认识十位同学)?

Question 2.1

1. 假设学号由八位数字组成，但第一位不能是 0 或 1. 一共可以有多少个不同的号码呢?
2. 小明在这些号码中随便选择一个，选中认识的同学的号码的概率有多大 (假定每位同学都有且仅有一个号码，其中小明认识十位同学)?

Solution

1. 第一位有 8 种选择，其余位有 10 种选择，所以有 8×10^7 个不同学号
2. 样本空间 $\Omega = \{(a,b,c,d,e,f,g,h) | a \neq 0,1\}, |\Omega| = 8 \times 10^7$
选中认识的同学的号码记为事件 A，则 $|A| = 10$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{10}{8 \times 10^7} = \frac{1}{8 \times 10^6}$$

排列与组合

1.1 排列 Arrange

从 n 个不同对象中顺序选出 k 个排成一列的方法，称为 n 取 k 排列，记作 $A_n^k = \frac{n!}{(n-k)!}$

1.2 组合 Combine

从 n 个不同对象中不计顺序选出 k 个为一组的方法，称为 n 取 k 组合，记作 $C_n^k = \frac{n!}{k!(n-k)!}$

排列与组合——捆绑法

有 3 名小学生，4 名初中生，5 名高中生. 现举行升国旗仪式，要求所有学生排成一列，相同年级站在一起，请问有多少种排队方式?

排列与组合——捆绑法

有 3 名小学生，4 名初中生，5 名高中生. 现举行升国旗仪式，要求所有学生排成一列，相同年级站在一起，请问有多少种排队方式?

Solution

分类讨论：1. 将不同类别的学生分别看作一个整体排一次 A_3^3
2. 对同一类别的学生进行内部排序 $A_3^3A_4^4A_5^5$
共有 $A_3^3A_3^3A_4^4A_5^5$ 种排序方法

排列与组合——二项式公式

组合系数为二项式展开系数

$$(a+b)^n = \sum_{i=0}^{n} C_n^i a^i b^{n-i} \tag{1}$$

算两次

Question 2.3

有 m 个女生，n 个男生，从中选出 k 人，有多少种选法？

解：$C_{m+n}^{k}=\sum_{i=0}^{k}C_{m}^{i}C_{n}^{k-i}$

男女一起选 = 男生女生分别选.

分割

n 个相异的物件分成 k 堆，各堆物件数分别为 $r_1 ... r_k$ 的分法分别是：

$$\frac{n!}{r_1!r_2!...r_k!}$$

证明

先从 n 个中取出 r_1 个作为第 1 堆，取法有 C(n, r_1) 种；在剩余的 $n - r_1$ 中取出 r_2 个作为第 2 堆，取法有 C($n - r_1$,r_2);⋯⋯以此类推，得到全部不同的分法为：

$$\prod_k C^{r_k}_{n-\sum_{j=1}^{k-1} r_j} \tag{2}$$

分割

n 双相异的鞋共 2n 只，随机的分成 n 堆，每堆 2 只，问："各堆自成一双鞋"这个事件 E 发生的概率？

分割

n 双相异的鞋共 2n 只，随机的分成 n 堆，每堆 2 只，问："各堆自成一双鞋"这个事件 E 发生的概率？(根据上述公式)

解答

2n 只鞋随机的分成 n 堆，每堆 2 只的分法为：$N=\dfrac{(2n)!}{2^n}$ 种

有利于事件 E 的分法为：把每一双鞋捆绑为一个整体看为一个物件，然后把这 n 个相异的物件分为 n 堆，每堆一件，分法有：$M=n!$ 种

于是，$P(E)=\dfrac{M}{N}=\dfrac{n!}{\dfrac{(2n)!}{2^n}}$

插空法

如果有 n 个男孩，m 个女孩，要求将这 n+m 个孩子不是排在一个圆周上，且女孩彼此不相邻，问此事件 E 发生的概率是多少？(m≤n)

插空法

如果有 n 个男孩，m 个女孩，要求将这 n+m 个孩子不是排在一个圆周上，且女孩彼此不相邻，问此事件 E 发生的概率是多少？$(m \leq n)$

解

因为排成一个圆周，头和尾没有分别，固定一个人把其他人展开成直线共有 $(N-1)!$ 种排法；

所以，先把孩子不分性别一起排一共有：$N=(m+n-1)!$ 种排法；

若先排男生，有：$(n-1)!$ 种排法；

圆桌排完男生后，男生之间有 n 个空位；

在其中选出 m 个空位排女生，有：C_n^m 种选法；

将女生安排到空位，共有：$m!$ 种排法。

所以，$M=(n-1)!C_n^m m!$

$$P(E)=\frac{M}{N}=\frac{(n-1)!C_n^m m!}{(m+n-1)!}$$

隔板法

将 n 个球放入 k 个不同的盒子，要使每个盒子都有球，请问有多少种不同的方法？若不限制每个盒子必须有球，则又有多少种方法？

隔板法

将 n 个球放入 k 个不同的盒子，要使每个盒子都有球，请问有多少种不同的方法？若不限制每个盒子必须有球，则又有多少种方法？

解

将 n 个球排成一排，在空隙处插入隔板。注意，只需要插入 $k-1$ 块板。对应的方式为 C_{n-1}^{k-1}

延伸

实际上，这可看作方程 $x_1+x_2+...+x_k=n$ 求正整数解的问题（将 n 拆分成 n 个 1，问题就转化成将 n 个 1 分成 k 份的组合数）。
若不限制每个盒子必须有球，则化为该方程求非负整数解的问题。同理，记 $y_i=x_i+1$，则相当于讨论 $y_1+y_2+\cdots+y_k=n+k$ 的正整数解，又回到了上面的情况。

四种概型

- 古典概型：基本事件的个数有限，且每个事件可能性相同
- 几何概型：均匀投点，面积可测，画图
- 超几何概型：无放回取球
- 二项（伯努利）概型：有放回取球

古典概型

1.3 一口袋装有外形相同的小球 15 个，其中红球 8 个，白球 7 个，从中无放回抽取 4 个，求下列概率：

(1) 恰有 3 个红的，1 个白的

(2) 至少有 1 个红的

(3) 至多有 1 个红的

(4) 颜色相同

古典概型

1.3 一口袋装有外形相同的小球 15 个，其中红球 8 个，白球 7 个，从中无放回抽取 4 个，求下列概率：

(1) 恰有 3 个红的，1 个白的

(2) 至少有 1 个红的

(3) 至多有 1 个红的

(4) 颜色相同

若改为有放回抽取呢（二项概型）

几何概型

从区间 (0,1) 中随机取两个数，求两数之和小于 1.2 的概率。

几何概型

从区间 (0,1) 中随机取两个数，求两数之和小于 1.2 的概率。

Solution

解　由已知，此题符合几何概型的特征：无限等可能结果.

设 x,y 是从区间 $(0,1)$ 内随机取出的两个数，则样本空间为

$$\Omega=\{(x,y)\mid 0<x<1,0<y<1\},$$

所求事件为 $A=\{(x,y)\mid x+y<1.2\}$（如图 1-2）.

由几何概型的计算公式，有

$$P(A)=\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}=\frac{1-\frac{1}{2}\times 0.8^2}{1}=0.68.$$

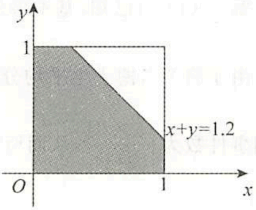

图 1-2

Figure 1: 几何概型

1.3 条件概率、乘法公式与事件的独立性

条件概率

- 已知 A 发生的条件下，发生 B 的概率：$P(B|A)=\dfrac{P(AB)}{P(A)}$，若 $P(A)\neq 0$
- 条件概率满足概率的三条公理
- $P(\overline{B}|A)=1-P(B|A)$，若 $P(A)\neq 0$
- 条件概率反映事件之间的相互影响。
- 若 $\mathbb{F}$ 中的 $A_1, A_2, ..., A_n, ...$ 互不相容，则：$P(\cup_{n=1}^{\infty}A_n|B)=\Sigma_{n=1}^{\infty}P(A_n|B)$

乘法公式

- $P(AB) = P(B|A)P(A)$;
- $P(A_1A_2 \cdot \cdot \cdot A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)...P(A_n|A_1...A_{n-1})$

乘法公式——官员受贿问题

某官员第一次收回没被查处的概率是 $q_1 = 98/100$，第一次没被查处后，第二次受贿没被查处的概率是 $q_2 = 96/98 = 0.9796...$ 前 j-1 次没被查处后，第 j 次受贿不被查处的概率是 $q_j = (100-2j)(100-2(j-1))$，求他受贿 n 次不被查处的概率 P_n。

乘法公式——官员受贿问题

某官员第一次收回没被查处的概率是 $q_1 = 98/100$，第一次没被查处后，第二次受贿没被查处的概率是 $q_2 = 96/98 = 0.9796...$ 前 j-1 次没被查处后，第 j 次受贿不被查处的概率是 $q_j = (100-2j)(100-2(j-1))$，求他受贿 n 次不被查处的概率 P_n。

Question 3.2

用 A_j 表示该官员 j 次受贿没被查处，则 $A_1A_2,..,A_n$ 表示受贿 n 次都没有被查处
所以，

$$
\begin{aligned}
P_n &= P(A_1A_2,..,A_n) = P(A_1)(A_2|A_1)...P(A_n|A_1...A_{n-1}) && (3)\\
&= q_1q_2...q_n && (4)\\
&= \frac{100-2n}{100} && (5)\\
&= 1 - \frac{n}{50} && (6)
\end{aligned}
$$

独立性 independent

- Define： $P(AB) = P(A)P(B)$ **则事件** A， B **相互独立**。
- **不可能事件、必然事件与任何事件独立**
- **当** $P(B) > 0$， $0 < P(A) < 1$， A、B **独立当且仅当** $P(B|A) = P(B|\overline{A}) = P(B)$， **即** B **事件发生概率不受** A **事件影响**。
- A **与** B， A **和** $\overline{B}$, $\overline{A}$ **和** B, $\overline{A}$ **和** $\overline{B}$, **只要一个独立**，**则另三个也独立**。
- **独立的性质**： **若** A、B、C、D **相互独立**，**则** $A\overline{B}$ **和** $C\overline{D}$ **一定相互独立**，**但** $A\overline{B}$ **和** $\overline{BCD}$ **不一定相互独立**。**（注意分组不能重合）**

独立与互不相容

- 独立与不相容的区别：分别是事件和概率的概念。

  $AB = \varnothing$ 与概率无关。

- $P(A) > 0, P(B) > 0$ 时，A 与 B 相互独立 $\Rightarrow$ AB 必相容。

  $P(AB) = P(A)P(B) > 0, P(\varnothing) = 0$

- 对立 $\Rightarrow$ 互不相容，但 A 互不相容 $\nRightarrow$ 对立。

Example

事件 A 与 B 对立, 即 B=$\overline{A}$, 是指在一次试验中，A 与 B 必有一个发生，且至多只有一个发生；但 A 与 B 互不相容是指 $AB = \varnothing$，即 A 与 B 可以同时不发生。例如，在一次考试中，及格与不及格总有一个发生，它们对立且不相容；70 分与 80 分互不相容，但是不对立.

Question

判断下列说法是否正确，并说明理由

(1) 若 A,B 不相容，则有 $P(AB) = P(A)P(B)$。

(2) 若 P(A),P(B) 均大于 0，A,B 独立，则 A,B 相容。

Question

判断下列说法是否正确，并说明理由

(1) **若** A,B **不相容，则有** $P(AB) = P(A)P(B)$。

(2) **若** P(A),P(B) **均大于** 0，A,B **独立，则** A,B **相容。**

解答

显然 (1) **错** (2) **对**.

Question

甲、乙、丙三人进行投篮练习，每人一次，如果他们的命中率分别为 0.8，0.7，0.6，计算下列事件的概率: (1) 只有一人投中; (2) 最多有一人投中; (3) 最少有一人投中

Question

甲、乙、丙三人进行投篮练习，每人一次，如果他们的命中率分别为 0.8，0.7，0.6，计算下列事件的概率: (1) 只有一人投中; (2) 最多有一人投中; (3) 最少有一人投中

解答

设事件 A、B、C 分别表示 “甲投中”、“乙投中”、“丙投中”，显然 A、B、C 相互独立. 设 A_i 表示 “三人中有 i 人投中”，i = 0, 1, 2, 3, 依题意

$$
\begin{aligned}
P(A_0) &= P(\bar{A}\bar{B}\bar{C}) = P(\bar{A})P(\bar{B})P(\bar{B}) \\
&= 0.2 \times 0.3 \times 0.4 = 0.024 \\
P(A_3) &= P(ABC) = P(A)P(B)P(C) \\
&= 0.8 \times 0.7 \times 0.6 = 0.336 \\
P(A_2) &= P(AB\bar{C}) + P(A\bar{B}C) + P(\bar{A}BC) \\
&= 0.8 \times 0.7 \times 0.4 + 0.8 \times 0.3 \times 0.6 + 0.2 \times 0.7 \times 0.6 = 0.452
\end{aligned}
$$

Question 接上

解答

解:

(1)

$P(A_1) = 1 - P(A_0) - P(A_2) - P(A_3) = 1 - 0.024 - 0.452 - 0.336 = 0.188$

(2) $P(A_0 + A_1) = P(A_0) + P(A_1) = 0.024 + 0.188 = 0.212$

(3) $P(\overline{A_0}) = 1 - P(A_0) = 1 - 0.024 = 0.976$

1.4 全概率公式和贝叶斯公式

全概率公式

全概率公式

若 $A_1, A_2, ..., A_n$ 构成一个完备事件组，且 $P(A_1) \geq 0, \forall i$, 则对任意事件 B，有 $P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$

- 完备事件组（前提，必须声明出来）:
  - 互不相容
  - 和为必然事件
  - 构成样本空间的一个划分
  - 每一个试验中，完备事件组中有且仅有一个事件发生
- 已知原因求结果
- $P(B)$ 后发生的事 (结果事件)，$P(A_i)$ 先发生的事 (原因事件)，$\sum_{i=1}^{n} P(A_i)P(B|A_i)$ 把所有原因下导致结果发生的可能性求和。

全概率公式

你参加一个王者荣耀单挑比赛，其中 50% 是黄金选手，你赢他们的概率为 0.3；25% 为白银选手，你赢他们的概率是 0.4；剩下的是英勇黄铜选手，你赢他们的概率是 0.618。从他们中间随机选择一名选手与你比赛你的胜算有多大？

全概率公式

你参加一个王者荣耀单挑比赛，其中 50% 是黄金选手，你赢他们的概率为 0.3；25% 为白银选手，你赢他们的概率是 0.4；剩下的是英勇黄铜选手，你赢他们的概率是 0.618。从他们中间随机选择一名选手与你比赛你的胜算有多大？

解答

选手段位为黄金 (A_1)、白银 (A_2)、英勇黄铜 (A_3) 构成一个完备事件组，且 $P(A_3) = 1 - P(A_1) - P(A_2) = 1 - 0.5 - 0.25 = 0.25$ 故可以使用全概率公式.

记事件 B 为你获胜，由全概率公式可以得到：

$$\begin{aligned} P(B) &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) \\ &= 0.5 \times 0.3 + 0.25 \times 0.4 + 0.25 \times 0.618 = 0.4045 \end{aligned}$$

贝叶斯公式

贝叶斯公式

设 $A_1, A_2, ..., A_n$ 构成一个完备事件组，且 $P(A_1) \geq 0, \forall i$, 则对任意事件 B，若有 $P(B) > 0$，有 $P(A_m|B) = \frac{P(A_m)P(B|A_m)}{P(B)} = \frac{P(A_m)P(B|A_m)}{\sum\limits_{i=1}^{n} P(A_i)P(B|A_i)}$

- 已知原因求结果：已经观察到了结果事件 B，那么导致 B 发生的是哪个原因？
- 理解：此种原因下发生 B 的概率除以所有原因导致 B 发生的原因。

贝叶斯公式

你参加一个王者荣耀单挑比赛，其中 50% 是黄金选手，你赢他们的概率为 0.3；25% 为白银选手，你赢他们的概率是 0.4；剩下的是英勇黄铜选手，你赢他们的概率是 0.618。现在已知你赢了，求你遇到黄金选手的概率是多大？

贝叶斯公式

你参加一个王者荣耀单挑比赛，其中 50% 是黄金选手，你赢他们的概率为 0.3；25% 为白银选手，你赢他们的概率是 0.4；剩下的是英勇黄铜选手，你赢他们的概率是 0.618。现在已知你赢了，求你遇到黄金选手的概率是多大？

解答

记选手段位为黄金 (A_1)、白银 (A_2)、英勇黄铜 (A_3) 构成一个完备事件组，$P(A_3) = 1 - P(A_1) - P(A_2) = 1 - 0.5 - 0.25 = 0.25$ 故可以使用贝叶斯公式.

记事件 B 为你获胜，由贝叶斯公式可以得到：

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(B)} = \frac{P(A_1)P(B|A_1)}{\sum_{i=1}^{3} P(A_i)P(B|A_i)}$$

$$= \frac{0.5 \times 0.3}{0.4045} \approx 0.3708$$

贝叶斯公式

（假阳性问题）设对于新冠肺炎的检出率为 0.95，如果被检人有这种病，其检查结果为阳性的概率为 0.95；如果该人没有这种病，其检查结果为阴性的概率为 0.95。现在假定某市市民患有这种病的概率为 0.001，并从这个总体中随机抽取一个人进行检测，检查结果为阳性。现在问这个人患这种病的概率有多大？

贝叶斯公式

（假阳性问题）设对于新冠肺炎的检出率为 0.95，如果被检人有这种病，其检查结果为阳性的概率为 0.95；如果该人没有这种病，其检查结果为阴性的概率为 0.95。现在假定某市市民患有这种病的概率为 0.001，并从这个总体中随机抽取一个人进行检测，检查结果为阳性。现在问这个人患这种病的概率有多大？

解答

记事件 A 为此人患有这种疾病，事件 B 为经检验这个人为阳性，利用贝叶斯公式可得：

$$
\begin{aligned}
P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A)+P(\bar{A})P(B|\bar{A})} \\
&= \frac{0.001\times 0.95}{0.001\times 0.95+0.999\times 0.05} \approx 0.0187
\end{aligned}
$$

Thank You
March 17, 2023