概率论 **Spring 2024**

第 1 节: 概率空间简介

Lecturer: 尹一通 *Scribes:* 张桄玮

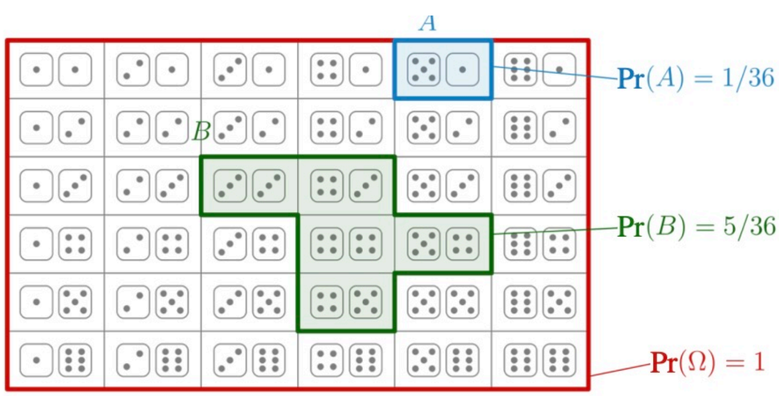
§1 概率空间

1.1. 定义的由来. 随机试验是概率论的基本概念, 试验的结果虽说事先不能准确地预言, 但具有如下三个特性:

1. 可以在相同的条件下重复进行;
2. 每次试验的结果不止一个, 但预先知道试验的所有可能的结果;
3. 每次试验前不能确定哪个结果会出现.

这就使得我们把一次试验中所有可能出现的结果 (possible outcomes)的集合放在一起考虑. 构成**样本空间 (sample space)**, 一般用 Ω 表示. 对于样本空间的每一个元素 $\omega \in \Omega$, 称为**样本 (sample)**(或**基本事件 (elementary event)**). 下文所讨论的**事件 (event)** 一般认为是 Ω 的一个子集.

例子 1.1. 下图展示了投掷两枚骰子的所有可能结果. 如果我们认为事件 $A :=$ 第一枚是 6, 第二枚是 1, 可以看到这是样本空间的一个子集. 事件 B 和 C 表示什么? 我们为什么会认为 "一小格" 代表的概率是 1/36?



实际上, 我们不假思索地认为一小格就是 1/36, 是因为我们认为这 36 种情况是等可能发生的. 因此将某一个小格子 (事件)映射成了一个实数. 假设这些小格子不是等可能出现的, 我们当然可以把这些小格子映成互不相同的实数. 但是这样的映射 $p : \Omega \to [0, 1]$ 要满足:

- $\sum_{\omega \in \Omega} p(\omega) = 1$. 我们希望所有的事件发生的概率加起来等于 1.

有了这样对基本元素的映射, 我们就可以定义一个事件 $A \subseteq \Omega$ 的概率, 定义作 $\Pr(A) := \sum_{\omega \in A} p(\omega)$, 称为**事件的概率 (Probability of event)**. 这也引出下一个问题: 到底什么样的集合簇, 才可以被称为事件构成的集合?

对于一个有限 (或者可数) 的样本空间 Ω 而言, 我们可以写出样本空间上所有可能的事件: 2^{Ω}. 但是, 这个要求能不能放松些? 通过我们以前的知识, 我们发现:

- $\emptyset$ 和 Ω 一定要在事件的集合里面. 它们是基础.
- 如果 A 在事件的集合中, 其补集 A^c 也要在. 不然我们就没有办法很方便地对一个事件取反.
- 如果有可数多个集合 $A_1, A_2, \ldots$ 在事件集合里面, 那么 $\bigcup_i A_i$ 和 $\bigcap_i A_i$ 也要在事件的集合里面. 这是为了我们方便地做并集和交集运算.

可以看到, 我们这里定义这样的规则 (2), (3), 完全是基于我们希望我们定义出来的这簇集合的元素在交、并、补这些运算下面封闭. 也就是事件集合中拿出任意可数多个元素进行上述集合的运算, 得到的结果一定还在这定义的事件集合里面. 这样我们定义的概率映射 p 才有意义.

于是经过上述探讨, 得到下面的概率空间的定义:

定义 1.1. 设 Ω 是一个样本空间 (或任意一个集合), Σ 是 Ω 的某些子集组成的集合族. 如果满足:

(1) $\Omega \in \Sigma$;

(2) 若 $A \in \Sigma$ ， 则 $A^c = \Omega \backslash A \in \Sigma$;

(3) 若 $A_n \in \Sigma, n = 1, 2, \cdots$, 则 $\bigcup_{n=1}^{\infty} A_n \in \Sigma$;

则称 Σ 为 σ **代数 (σ-algebra)**. (Ω, Σ) 称为**可测空间 (measurable space)**, Σ 中的元素称为**事件 (events)**.

这里没有定义交集的原因是，交集可以通过并集取反得到. 这定义是一个最精简的版本.

例子 1.2. 如果研究掷一次硬币的结果, 可用如下 Ω 表示样本空间:

$$\Omega = \{H, T\}$$

如果研究掷两次次硬币的结果，可用如下 Ω 表示样本空间:

$$\Omega = \{\{H, H\}, \{H, T\}, \{T, H\}, \{T, T\}\}$$

如果研究掷一次六面骰子出现 2 或 3 的概率, 可用如下可测空间严格描述:

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

如果研究掷一次六面骰子出现 2 和 3 的概率, 可用如下可测空间严格描述

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

定义 1.2. 设 (Ω, Σ) 是可测空间, Pr 是一个从事件 Σ 的集合到实数集合的映射 ($\mathrm{Pr}: \Sigma \to [0, 1]$), 满足

(1) (归一化) $\mathrm{Pr}(\Omega) = 1$;

(2) (σ-可加性) 对两两互不相容事件 $A_1, A_2, \cdots$, (即当 $i \neq j$ 时, $A_i \cap A_j = \emptyset$) 有

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

则称 Pr 是 (Ω, Σ) 上的**概率测度 (probability measure)**, 三元组 (Ω, Σ, P) 称为概率空间, Σ 中的元素称为事件, $P(A)$ 称为事件 A 的概率.

一经建立上述的公理体系 (定义), 我们任何常见的对于概率的想象都可以被验证.

例子 1.3. 下面的内容是我们在中学 "不假思索" 就用的. 我们现在诉诸公理, 证明他们.

- $\Pr(A^c) = 1 - \Pr(A)$. 这是因为考虑 A 和它的补 A^c 没有交集, 且组成了全空间, 根据 $\Pr(\Omega) = 1$ 以及 σ-可加性可知.
- $\Pr(\varnothing) = 0$. 考虑反证法说明: 因为 $\Pr(A) > 0 \Longrightarrow A \neq \varnothing$ 这样的方法被称为**概率方法 (the probabilistic method)**
- $\Pr(A \backslash B) = \Pr(A) - \Pr(A \cap B)$ 可以借鉴问题 1 的思路.
- $A \subseteq B \Longrightarrow \Pr(A) \leq \Pr(B)$. 直观来看, 由于 $\Pr$ 的值域是 $[0, 1]$, 点越多, 概率至少不应该变小.
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.这是容斥原理的一个简单的形式.

1.2. 几个重要的不等式. 下面我们介绍几个重要的不等式和一些重要的思想.

a) 并的上界 (union bound)

命题 1.1. 对于事件 $A_1, A_2, \cdots, A_n$,

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

这可以通过上述例子中 $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$ 得到证明, 并且可以知道等号取得的条件是 A_i 两两没有交集.

b) 容斥原理 (principle of inclusion-exclusion, PIE) 继续推广 $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$, 当有三个集合的时候, 要算它们的并集就可以转换为它们的交集. 我们接下来先说明一般的容斥原理.

若 S 是一个集合, 记 $|S|$ 表示集合 S 中元素的个数, 对于两个集合而言, 有 $|A \cup B| = |A| + |B| - |A \cap B|$. 如果是三个集合, 可以写作

$$\begin{aligned} |A \cup B \cup C| = & |A| + |B| + |C| \\ & -|A \cap B| - |A \cap C| - |B \cap C| \\ & +|A \cap B \cap C| \end{aligned}$$

其特点如下:

1. 这是一个把求集合并集化为交集之组合的变换.
2. 如果我们把参与并集那些集合 (如 A, B, C) 当做一个大集合 K, 其会生成 K 的所有子集, 这每个子集的元素 (也是集合) 交起来, 便得到了构成等式右端的基本组件.
3. 如果这个子集的大小为奇数, 前面的符号为正; 反之就是负.

根据我们上面的观察, 实际上我们就可以得到一般的容斥原理:

定理 1.2 (容斥原理)**.** 假设是 $A_1, A_2, ..., A_n$ 是 n 个集合, 那么有

$$\begin{aligned}\left|\bigcup_{i=1}^{n} A_i\right| &= \sum_{i=1}^{n} |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \cdots \\ &= \sum_{\substack{s\subseteq\{1,2,\ldots,n\}\\ S\neq\varnothing}} (-1)^{|S|-1} \left|\bigcap_{i\in S} A_i\right|\end{aligned}$$

我们来证明这个定理. 最无脑 (但不一定总是可以这样做) 的方法当然是可以使用数学归纳法了. 但是我们今天换一种方法.

Proof. 我们的目标是计数的时候每个元素都被算过, 且只被算过了一次. 假设 $A_1, ..., A_n \subset X$, 对于每一个 A_i, 我们定义一个指示函数. 这个指示函数写作

$$f_i(x) := \begin{cases} 1 & x \in A_i \\ 0 & x \notin A_i \end{cases}$$

我们考虑

$$F(x) = \prod_{1\leq i\leq n} (1 - f_i(x)).$$

这个函数是集合 $\cup_{1\leq i\leq n} A_i$ 的补的指示函数. 也就是说, $F(x) = 1 \iff x$ 不在任何一个 A_i 集合中. 于是,

$$\sum_{x\in X} F(x) = |X \setminus \cup_{1\leq i\leq n} A_i|.$$

现在我们用另一种视角看 $F(x)$. 通过把对应的指示函数展开, 可以得到 2^n 个元素, 就是

$$F(x) = \prod_{1\leq i\leq n} (1 - f_i(x)) = \sum_{I\subset[n]} (-1)^{|I|} \prod_{i\in I} f_i(x).$$

实际上, $\prod_{i\in I} f_i(x)$ 就是 $\cap_{i\in I} A_i$ 的特征函数, 因此得到

$$\sum_{x\in X} F(x) = \sum_{I\subset[n]} (-1)^{|I|} \sum_{x\in X} \prod_{i\in I} f_i(x) = \sum_{I\subset[n]} (-1)^{|I|} |\cap_{i\in I} A_i|$$

这样一来, 我们就得到了对应的大小信息

$$|X \setminus \cup_{1\leq i\leq n} A_i| = |X| - |\cup_{1\leq i\leq n} A_i| = \sum_{I\subset[n]} (-1)^{|I|} |\cap_{i\in I} A_i|$$

于是我们就证明了这个定理.

□

实际上, 事件只是特殊的集合, 而概率函数 (Pr) 和个数函数 ($|\cdot|$) 有类似的性质, 因此固然有

对于事件 $A_1, A_2, \cdots, A_n \in \Sigma$, 有

$$\begin{aligned}\operatorname{Pr}\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} \operatorname{Pr}(A_i) - \sum_{i<j} \operatorname{Pr}(A_i \cap A_j) + \sum_{i<j<k} \operatorname{Pr}(A_i \cap A_j \cap A_k) - \cdots \\ &= \sum_{\substack{S \subseteq 11,2,\ldots,n\} \\ S \neq \varnothing}} (-1)^{|S|-1} \operatorname{Pr}\left(\bigcap_{i \in S} A_i\right)\end{aligned}$$

例子 1.4 (错排问题)**.** 取数列 $a = [1, 2, 3, ..., n]$, 将它随机打乱. 请问 $a_i \neq i$ 的情况有多少?

对计数任务做合理的划分. 有时候从反面考虑会有些帮助. 我们考虑计算 "存在 $a_i = i$ 的情况的个数". 定义 A_i 表示 $a_i = i$ 的可能性. 并且我们发现 $\operatorname{Pr}\left(\bigcap_{i \in S} A_i\right) = \frac{(n-|S|)!}{n!}$ 这个式子比较好算, 我们的任务成功了一大半. 然后应用容斥原理的表达:

$$\operatorname{Pr}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} \sum_{S \in \left(\begin{array}{c}\{1,2,\ldots,n\} \\ k\end{array}\right)} (-1)^{k-1} \operatorname{Pr}\left(\bigcap_{i \in S} A_i\right) = \sum_{k=1}^{n} \left(\begin{array}{c} n \\ k \end{array}\right) (-1)^{k-1} \frac{(n-k)!}{n!}$$

由于我们的问题是回答的反面, 自然用 1 减, 得到

$$\operatorname{Pr}\left(\bigcap_{i=1}^{n} A_i^c\right) = 1 - \operatorname{Pr}\left(\bigcup_{i=1}^{n} A_i\right) = 1 + \sum_{k=1}^{n} \frac{(-1)^k}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

当 n 很大的时候, 它接近 $1/e$.

1.3. 概率测度的连续性. 就像我们在 $\mathbb{R}$, $\mathbb{R}^2$ 上面定义了距离的度量, 然后由此导出了邻域的定义. 一旦定义了邻域, 就可以定义映射极限的概念.

定理 1.3 (概率测度是连续的)**.** 如果 $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ 是一系列事件, 我们定义这一簇集合的极限为 A, 表示

$$A := \bigcup_{i=1}^{\infty} A_i = \lim_{i \to \infty} A_i.$$

那么有 $\operatorname{Pr}(A) = \lim_{i \to \infty} \operatorname{Pr}(A_i)$

这条定理说明了: 只要搞明白这一簇集合最终的极限是什么, 我们就能够得到与之对应的 $\operatorname{Pr}(A)$ 是多少. 不必再去考虑每一个集合 A_i 单独做映射得到 $\operatorname{Pr}(A_i)$ 并求极限.

Proof. 关键是使用不相交集合的可加性. 如果我们把最终的极限 A 写作若干个不交和 $A = A_1 \uplus (A_2 \backslash A_1) \uplus (A_3 \backslash A_2) \uplus \ldots$, 那么

$$\begin{aligned}\operatorname{Pr}(A) &= \operatorname{Pr}(A_1) + \sum_{i=1}^{\infty} \operatorname{Pr}(A_{i+1} \backslash A_i) \\ &= \operatorname{Pr}(A_1) + \lim_{n \to \infty} \sum_{i=1}^{n-1} [\operatorname{Pr}(A_{i+1}) - \operatorname{Pr}(A_i)] \\ &= \lim_{n \to \infty} \operatorname{Pr}(A_n)\end{aligned}$$

□

同样的, 如果我们有不断变小的一系列集合, 也有同样的定理, 即: 如果 B 是事件集合 $B_1, B_2, ...$, 满足 $B_1 \supseteq B_2 \supseteq B_3 \supseteq \ldots$ 的集合的极限 (定义作 $B = \bigcap_{i=1}^{\infty} B_i = \lim_{i\to\infty} B_i$), 那么有 $\Pr(B) = \lim_{i\to\infty} \Pr(B_i)$. 实际上我们可以考虑上述各个事件的补, 也就是考虑 $B_1^c \subseteq B_2^c \subseteq B_3^c \subseteq \ldots$ – 就转化到了上述定理的内容了.

有了上述的铺垫, 我们可以看一看**零事件 (null event)**.

定义 1.3 (零事件)**.** 如果一个事件 $A \in \Sigma$, 满足 $\Pr(A) = 0$, 那么就称它为零事件.

注意零事件和不可能事件的区别. 零事件不一定是不可能事件. 即: 概率为 0 不一定不会发生.

与之对应, 概率为 1 也不一定会发生. 像概率为 1 事件我们叫做几乎必然事件.

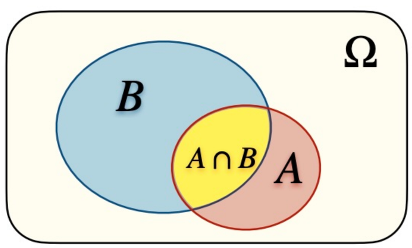
定义 1.4 (**几乎必然 (almost surely, a.s.)**)**.** 如果一个事件 $A \in \Sigma$, 满足 $\Pr(A) = 1$, 那么就称它为几乎必然事件.

§2 条件概率

考虑 "在事件 B 发生的条件下, A 发生的概率是 p". 在中学学过条件概率公式

$$p = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{\Pr(A \cap B)}{\Pr(B)}$$

这是因为考虑直观的意义, 可以把 $A \cup B$ 和 B 在全空间的意义下分别表示.



用集合的术语来说, 这就相当于在原有的集合上面做了一个限制. 我们现在为它下一个定义, 看一看概率空间的基础上, 如何定义条件概率.

2.1. 基本定义.

定义 2.1. 如果 A 是一个事件, B 是一个 $\Pr(B) > 0$ 的事件, 那么 A 在 B 发生的条件下发生的**条件概率 (conditional probability)** 被定义作

$$\Pr(A \mid B) := \frac{\Pr(A \cap B)}{\Pr(B)}$$

这样的定义是合理的. 只要我们

- 把样本空间看做 B;
- $\Sigma^B := \{A \cap B : A \in \Sigma\}$ 满足 σ-代数的条件;
- $\Pr(\cdot|B)$ 满足概率的公理.

那么, 由 $(B, \Sigma^B, \Pr)$ 也构成一个概率空间.

例子 2.1. 假设你有一个以概率为 p 出现正面 (H) 的硬币, 但是不知道 p 是多少. 你应该如何用这个硬币生成均匀的硬币投掷?

答案: 只要我们不断地投掷, 当连续两次投掷结果是 HT 的情况下, 就等效地认为扔出了正面; 连续两次投掷出现 TH, 就认为扔出了反面.

这是因为

$$\Pr(\mathrm{HT} \mid \{\mathrm{HT}, \mathrm{TH}\}) = \Pr(\mathrm{TH} \mid \{\mathrm{HT}, \mathrm{TH}\}) = \frac{p(1-p)}{2p(1-p)} = \frac{1}{2}$$

例子 2.2. 已知两个孩子, 至少一个是女孩. 那么两个孩子都是女孩的概率是多少?(假设男孩和女孩出现等概率).

假设在空间有 $\Omega = \{\mathrm{BB}, \mathrm{BG}, \mathrm{GB}, \mathrm{GG}\}$, 每种情况发生的可能性相同. 那么

$$\begin{aligned}\Pr(\{\mathrm{GG}\} \mid \{\mathrm{BG}, \mathrm{GB}, \mathrm{GG}\}) &= \frac{\Pr(\{\mathrm{GG}\})}{\Pr(\{\mathrm{BG}, \mathrm{GB}, \mathrm{GG}\})} \\ &= \frac{1/4}{3/4} = \frac{1}{3}\end{aligned}$$

2.2. 条件概率的性质.

a) 链式法则 链式法则 (Chain rule) 描述了我们为什么像计数的时候应用乘法原理那样对 “分步决策的” 两个概率乘起来. 其表述如下:

定理 2.1 (链式法则)**.** 假设 $A_1, A_2, ..., A_n$ 有正概率, 那么

$$\Pr(A_1 \cap A_2 \cap A_3 \cap ... \cap A_n) = \Pr(A_1)\Pr(A_2|A_1)\Pr(A_3|A_1 \cap A_2) \cdots \Pr(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

也就是

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \Pr\left(A_i \mid \bigcap_{j<i} A_j\right)$$

Proof. 首先回到定义上面去. 可以使用裂项的方法证明.

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \frac{\Pr\left(\bigcap_{i=1}^{n} A_i\right)}{\Pr\left(\bigcap_{i=1}^{n-1} A_i\right)} \cdot \frac{\Pr\left(\bigcap_{i=1}^{n-1} A_i\right)}{\Pr\left(\bigcap_{i=1}^{n-2} A_i\right)} \cdots \frac{\Pr(A_1 \cap A_2)}{\Pr(A_1)} \cdot \Pr(A_1)$$

□

例子 2.3. 把 n 个球一个一个地随机扔进 m 个盒子里面, 请问每个球都扔进了一个空盒子的概率是多少?

答案: 考虑分步进行. 很容易求 “对于编号 j 的球 $(j < i)$ 都扔进了一个空桶的情况下, 编号 i 的球扔进了一个空桶的概率” 为 $1 - \frac{i-1}{m}$. 那么根据上述的定理:

概率就是

$$\prod_{i=1}^{n} \left(1 - \frac{i-1}{m}\right)$$

2.3. 全概率公式. 全概率公式说的事情是对于一个复杂的样本空间, 与其在其上面一整个上面求概率, 不如先把样本空间做一个划分, 在划分的小片上面计算概率, 然后把它们加起来.

定理 2.2 (全概率公式)**.** 令事件 $B_1, B_2, ..., B_n$ 是 Ω 的一个分割, 其中对任何的 i, 都有 $\Pr(B_i) > 0$ 那么,

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \cap B_i) = \sum_{i=1}^{n} \Pr(A \mid B_i) \Pr(B_i)$$

Proof. 由于 $A \cap B_1, A \cap B_2, \ldots, A \cap B_n$ 是互不相交的, 且 $A = \bigcup_{i=1}^{n} (A \cap B_i)$. 这就表明 $\Pr(A) = \sum_{i=1}^{n} \Pr(A \cap B_i)$.

这同样表明了 $\Pr(A \cap B_i) = \Pr(A \mid B_i) \Pr(B_i)$. □

例子 2.4 (三门问题)**.** 假设你参加一个游戏节目, 有三扇门. 其中一扇门后面有大奖, 另外两扇门后面是什么都没有. 你选择了一扇门 (比如第一个门), 主持人知道每扇门后面是什么. 然后, 他打开了另外一扇门 (比如第三个门), 里面是什么都没有. 然后他问你:"你要换成第二号门吗?" 你应该换门吗?

答案: 应该换. 定义事件 A 为我们最后得到了大奖; B 为一开始我们选了含有车的那个门. 那么

$$\Pr(A) = \begin{cases} \Pr(B) = 1/3 & \text{不换门} \\ \Pr(A \mid B)\Pr(B) + \Pr(A \mid B^c)\Pr(B^c) = 0 + 1 \times 2/3 = 2/3 & \text{换门} \end{cases}$$

例子 2.5 (赌徒破产问题)**.** 一赌徒扔一枚均匀硬币. 如果他得到正面, 会得到 1 元; 如果得到反面, 就会输掉 1 元. 他从 k 元开始游戏. 在他有的总钱数超过 k 的时候 (赢了) 或者小于 0 的时候 (破产, 输了) 停止游戏. 问他最后赢的概率是多少.

答案: 定义事件 $A :=$ 赌徒破产输了; $B :=$ 第一次掷硬币的时候得到了正面.

以及让 $\Pr_k$ 表示赌徒以 k 元开始游戏. 由于输掉一元之后可以重新看做以 $k-1$ 元开始, 有如下的递归式:

$$\Pr_k(A) = \frac{1}{2}\Pr_k(A \mid B) + \frac{1}{2}\Pr_k(A \mid B^c) = \frac{1}{2}\Pr_{k+1}(A) + \frac{1}{2}\Pr_{k-1}(A)$$

解这个递归式, 就有

$$\Pr_k(A) = \begin{cases} \frac{1}{2}(\Pr_{k+1}(A) + \Pr_{k-1}(A)) = 1 - \frac{k}{n} & \text{if } 0 < k < n \\ 1 & \text{if } k = 0 \\ 0 & \text{if } k = n \end{cases}$$

也就是赌徒最后一定会破产.

下面来看全概率公式的一重要变形. 我们会成它为 Bayes 定理. 这项定理可以帮助把条件概率的条件 "反过来". 有时候也会称它为 "逆概率公式".

定理 2.3 (**Bayes 定理 (Bayes Theorem)**)**.** 对于事件 A, B 满足 $\Pr(A), \Pr(B) > 0$, 有

$$\Pr(B \mid A) = \frac{\Pr(B)\Pr(A \mid B)}{\Pr(A)}$$

如果 $B_1, B_2, ..., B_n$ 是 Ω 的一个划分, $\forall i, \Pr(B_i) > 0, \Pr(A) > 0$, 那么

$$\Pr\left(B_i \mid A\right) = \frac{\Pr\left(B_i\right)\Pr\left(A \mid B_i\right)}{\Pr(A)} = \frac{\Pr\left(B_i\right)\Pr\left(A \mid B_i\right)}{\Pr\left(A \mid B_1\right)\Pr\left(B_1\right) + \cdots + \Pr\left(A \mid B_n\right)\Pr\left(B_n\right)}$$

这定律可以用全概率公式和链式法则想办法把 A, B 互换顺序就可以推出.

例子 2.6 (假阳性问题)**.** 一个函件的疾病以 0.001 的概率发生. 但是有 5% 的测验误差. 也就是说, 有病的人 $\begin{cases}\text{测出有病 (+)95\%} \\ \text{测出没病 (-)5\%}\end{cases}$, 没病的人 $\begin{cases}\text{测出没病 (-)95\%} \\ \text{测出有病 (+)5\%}\end{cases}$.

现在问, 如果一个人测出有病的条件下, 他得病的概率是多少?

答案:

$$\begin{aligned}\Pr(\text{有病} \mid +) &= \frac{\Pr(\text{有病})\Pr(+ \mid \text{有病})}{\Pr(+)} = \frac{\Pr(\text{有病})\Pr(+ \mid \text{有病})}{\Pr(+ \mid \text{有病})\Pr(\text{有病}) + \Pr(+ \mid \neg\text{有病})\Pr(\neg\text{有病})} \\ &= \frac{0.001 \times 95\%}{95\% \times 0.001 + 5\% \times 0.999} \approx 1.87\%\end{aligned}$$

实际上, 在不同的条件下, 我们看到的内容会很不一样. Simonson 发现了这样的一个 “悖论”:

例子 2.7 (Simonson 悖论)**.** 如下是两个药品的临床试验数据:

	女		男	
	药品 I	药品 II	药品 I	药品 II
成功	200	10	19	1000
失败	1800	190	1	1000

请问哪一个药品更有效?

1. 药品 I 更有效, 因为总体来看 I 的成功率高于 II.
2. 药品 II 更有效, 对于女性和男性分别而言, II 的成功率都要大于 I.

用概率的语言来讲, 就是对于事件 A, B 以及 Ω 的一个划分 $C_1, C_2, ..., C_n$, 有可能出现:

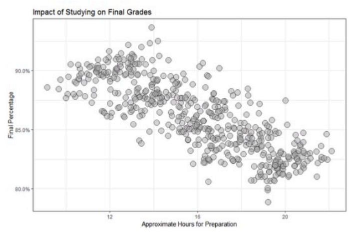
- 每一个 C_i, B 的发生对 A 有积极影响, 即

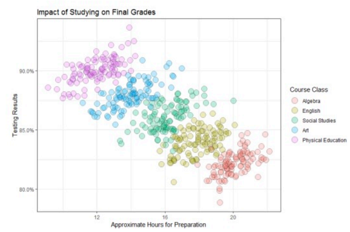
$$\Pr\left(A \mid B \cap C_i\right) > \Pr\left(A \mid B^c \cap C_i\right), \forall i$$

- 但是, 总体来看, B 的发生对 A 的发生有消极影响. 即

$$\Pr(A \mid B) < \Pr\left(A \mid B^c\right)$$

一个例子是, 总体来看, 下图表示了学习时长越长, 成绩越低; 但是一门课一门课来看, 总体学习时长越长, 分数越高.





§3 事件的独立性

3.1. 基本定义. 有时候, 有些事件的发生会影响另一个事件发生的概率. 也就是从 $\Pr(A)$ 变为 $\Pr(A|B)$. 如果我们发现 $\Pr(A|B) = \Pr(A)$, 我们就可以称这两个事件独立. 但是, 这样的定义可能会不能处理 $\Pr(B) = 0$ 的情形, 于是更换一个更加方便的定义:

定义 3.1 (两个事件的独立性)**.** 如果两个事件满足

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

我们就称两个事件**独立 (independent)**.

关于独立, 有如下的几个常用的性质.
1° 如果 $\Pr(B) > 0, \Pr(A \mid B) = \Pr(A) \Longleftrightarrow \Pr(A \cap B) = \Pr(A)\Pr(B)$;
2° $\Pr(A \cap B) = \Pr(A)\Pr(B) \Longleftrightarrow \Pr\left(A \cap B^c\right) = \Pr(A)\Pr\left(B^c\right)$
受刚才定义的启发, 我们定义一簇事件 $\{A_i \mid i \in I\}$ 互相独立的条件.

定义 3.2. 一簇事件 $I = \{A_i \mid i \in I\}$ 成为 (相互) 独立的, 当且仅当对于任何 I 的有限子集 $J \subset I$, 满足

$$\Pr\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \Pr\left(A_i\right).$$

如果我们想要说一个事件和其他的一簇事件独立, 则有如下的定义.

定义 3.3. 事件 A 与另一簇事件 $\{B_1, B_2, ..., B_k\}$ 互相独立, 当且仅当对任意的下标集合 $S \subset \{1, 2, 3, ..., k\}$ 满足

$$\Pr\left(B \mid \bigcap_{i \in S} B_i\right) = \Pr(B).$$

按独立性的定义来判定, 有时候我们与生俱来的直觉可能会起到反作用. 例如, 如下的几种可能的情况都是可能的:

1. $A_1, A_2, \ldots, A_n$ 相互独立, $B_1, B_2, \ldots, B_n$ 也相互独立. 但是 $\forall 1 \le i \le n, A_i$ 和 B_i 不相互独立.
2. $\forall 1 \le i \le n, A_i$ 和 B_i 相互独立; 但是 $\forall 1 \le i < j \le n$, A_i 和 A_j, B_i 和 B_j 都不互相独立.

但是我们可以试图把这些事件的依赖关系画成图. 有连边的两个顶点说明他们有某种相关性. 比如我们有事件 $B_1, B_2, ..., B_n$, 如果构造一个有向图 $D = (\{1, 2, 3, ..., n\}, E)$ QUEST: 无向图还是有向图?

如果 B_i 与所有 $(i,j)\notin E$ 的 B_j 这一簇事件 ($\{B_j:(i,j)\notin E\}$) 独立, 我们就称这个图为依赖关系的依赖图. 对于互相独立的一组事件, 它的依赖图就是孤零零的一些点, 并没有连边.

依赖关系图并不是唯一的. 我们可以构造出不同的依赖图.

与互相独立相对, 我们也可以定义一簇事件中两两独立.

定义 3.4 (**两两独立 (pairwise independent)**)**.** 对于一簇事件 $\{A_i \mid i\in I\}$, 我们称它为两两独立, 当且仅当对于不同的下标 $i,j\in I$, 有

$$\Pr\left(A_i\cap A_j\right)=\Pr\left(A_i\right)\Pr\left(A_j\right)$$

实际上, 互相独立是一个很强的定义. 即: 互相独立的一簇事件一定是两两独立的; 但是两两独立的事件不一定是互相独立的.

最后我们来定义条件概率意义下的独立.

定义 3.5 (条件独立)**.** 两个事件 A,B 在 C 事件下独立 ($\Pr(C)>0$), 意味着

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

这同样也没有偏离我们对独立的刻画: 如果 $\Pr(B\cap C)>0$, 那么 $\Pr(A\cap B\mid C)=\Pr(A\mid C)\Pr(B\mid C)\Longleftrightarrow\Pr(A\mid B\cap C)=\Pr(A\mid C)$

3.2. 由 Cartesian product 生成的概率空间. 在离散数学中学习过两个集合的 Cartesian product: 如果有两个集合 A,B, 那么定义 $A\times B=\{(i,j):i\in A,j\in B\}$. 通常会从两次或者多次独立实验序列中构造一个概率空间. 例如考虑离散的概率空间 $\left(\Omega_1,p_1\right),\left(\Omega_2,p_2\right),\ldots,\left(\Omega_n,p_n\right)$, 那么这个大的概率空间 (Ω,p) 可以这样构造:

- 样本空间 $\Omega=\Omega_1\times\Omega_2\times\cdots\times\Omega_n$
- $\forall\omega=\left(\omega_1,\ldots,\omega_n\right)\in\Omega$, 对应的 $p(\omega)=p_1\left(\omega_1\right)\cdots p_n\left(\omega_n\right)$.

对于一般的概率空间 $\left(\Omega_1,\Sigma_1,\Pr_1\right),\ldots,\left(\Omega_n,\Sigma_n,\Pr_n\right)$, 在这些概率空间上面做 Cartesian product, 需要满足:

- Σ 是 (唯一) 最小的包含 $\Sigma_1\times\cdots\times\Sigma_n$ 的 σ-代数;
- $\Pr$ 扩展为 $\forall A=\left(A_1,\ldots,A_n\right)\in\Sigma_1\times\cdots\times\Sigma_n,\Pr(A)=\Pr\left(A_1\right)\cdots\Pr\left(A_n\right)$.

注记

容斥原理的证明来自`https://math.mit.edu/~fox/MAT307-lecture04.pdf`, 像这样使用多项式的性质证明组合结构的内容会在《组合数学》课程上面频繁用到. 蒋炎岩老师也曾经面向中学生提到过容斥原理. 详见视频`https://www.bilibili.com/video/BV1G3411h7f5`.

一个事件和另一簇事件的独立性采用的定义来自`https://cse.buffalo.edu/~hungngo/classes/2011/Spring-694/lectures/lm.pdf`. 这份定义随后说明了依赖图, 并且由它导出了一个比较有趣的定理. 但是上面的定义和 Slides 定义不太一致. 我也不知道应该用哪个更妥当一点.

本节需要特别注意的是: 概率空间以及上面的不等式; 条件概率及其的性质; 事件的独立性的定义. 还要尤其体会在解决问题的时候用到的递归思想.