数据科学基础知识

第一部分:概率论

ISEP第二年
2023-2024

基于 Nathalie Colin 和 Jean-Claude Guillerot 教授的课程

概率论

第一届会议（2023 年 9 月 29 日）：

第 1 章:事件和概率的概念

第 2 章:条件概率 &
独立

第一章:简介 主要定义

简介 概率的三种
定义 ·频率概率 ·基于结果数量的定义 ·
基于公理的定义集合和
事件 ·集合代数提示 ·公理推论

介绍

概率：≠确定性

概率:事件发生的确定程度

历史上……

从八世纪和十三世纪开始,阿拉伯数学家就开始研究密码学
首次使用排列和组合来列出所有可能带或不带元音的阿拉伯语单词。

现代概率数学理论源于十六世纪杰罗拉莫·卡尔达诺、十七世纪皮埃尔·德·费马和布莱斯·帕斯卡对机会游戏的分析尝试。

最初,概率论主要考虑离散事件,其
方法主要是组合方法。后来分析考虑
迫使连续变量的结合。

如今,它存在于许多科学领域,如经济学、物理学（统计学）、遗传学、通信学、计算机科学等。

概率的三种定义

- Ø频率概率
- Ø基于结果数量的定义
- Ø基于公理的定义

词汇

我们对产生单一结果的随机实验感兴趣

有限数量的可能结果中的结果表示为： ω1, ω2,

...，ωn。我们将 Ω 表示为所有可能结果的集合（样本空间）：

$$\Omega=\{\omega 1,\omega 2,\ldots,\omega n\}$$

我们称事件为 Ω 的子集 A

例子：

- 实验 :滚动六面骰子
- 可能的结果： Ω = {1,2,3,4,5,6}
- 事件： A = {结果为偶数}

我们将不可能事件表示为空集 ∅ 和确定事件

将是设定的Ω。

定义1:频率概率

结果ωk的相对频率fk是

结果为ωk的实验与实验总数 n: !

= ” !
,,

事件 A 的概率,记为 P(A),是其发生的极限

许多试验中的相对频率。

$$P(A) = \lim_{n \to \infty} \left(\frac{n_A}{n} \right)$$

其中： nA： 实验结果是事件 A 的次数

nA/n :事件 A 的相对频率

该定义的优点/缺点:

- 满足直观的概率概念

- 它通过重复试验来衡量事件发生的概率

- 表达式包含限制。

定义 2:基于结果数量的定义

$$(\ ) = \frac{!}{\quad} = \frac{\text{事件 A 的结果数}}{\text{样本空间中的结果总数}}$$

样本空间中的所有结果必须具有相同的可能性。

示例:实验 -> 掷骰子

事件 -> A = {1}

可能结果的数量 = 6

事件 A = 1 的结果数P(A) = 1/6

该定义的优点/缺点

-非常简单的定义

- 例如,当结果集无穷大时,计数可能是不可能的!

定义3:基于公理的定义

给定一个实验,事件 A 的概率（表示为 P(A))是 0 到 1 之间的实数,满足:

公理 1 :P(A) 为正或为零。

公理 2 :如果 A 是特定事件:P(A) = 1

公理 3 :可加性

如果 A 和 B 不相交,即 $A \cap B = \varnothing$,

$$P(A \cup B) = P(A) + P(B)$$

该属性可以扩展到无限的不相交事件集 Ai

≠ (! ∩ ”) = ∅ (!#$, !) = !#$ ( ) !

设置代数！ => 在概率论中非常有用

集合代数⇔概率论的对应关系

- Ø套装⇔活动
- Ø集合的元素⇔实验的任何可能结果
- Ø空集∅⇔不可能事件
- Ø Ω :宇宙集或幂集⇔样本空间（所有可能的集合）结果）

提醒:设置代数

工会	在
十字路口	∩
幂集	在
空集	哦
A 的补集	
交换律	∪ = ∪ 且 ∩ = ∩
结合律	(∪) ∪ = ∪ (∪) 且 (∩) ∩ = ∩ (∩)
分配律	∩ (∪) = (∩) ∪ (∩) 且 ∪ (∩) = (∪) ∩ (∪)
补码定律 ∪ = Ω ∩	= ∅, ∪ Ω =, Ω, ∩ Ω = A, ∪ ∅ = A, ∩ ∅ = ∅
对合律	$\overline{\quad}$ () =
幂等定律	∪ = 和 ∩ =
摩根定律	$\overline{(\cup)}$ = ∩ 0 且 $\overline{(\cap)}$ = ∩ 0

4个公理推论：

推论1： $P(\varnothing) = 0$

推论2： $P(A) = 1 - P(\overline{A}) \leq 1$

推论 3： $(\quad \cup =) \quad (\quad) + (\quad) - (\cap)$

推论 4：

如果 ⊂ 则 $(\quad) = (\quad) + (\cap \geq (\,) \;)$

⊂ 表示 A 是 B 的子集

第 2 章:条件概率与独立

- 定义和解释
- 独立
- 全概率定律
- 贝叶斯定理

条件概率

条件概率的定义：

设A和M为两个事件，并假设M已发生（先验信息），则P(M) ≠ 0。给定M 时A的条件概率，表示为P(A|M)：

$$(\,|\,) = \frac{(\;\cap\;)}{()}$$

P(A|M) 是概率吗?让我们看看它是否验证了 3 个公理：

Ø正数或零。

Ø某个事件的条件概率P(Ω/M)等于1。

Ø两个不相交事件并集的条件概率
(A∩B∩M = ∅) 等于它们的条件概率之和，也就是说：

$$(\;\cup\; | \;) = (\,|\,) + (\,|\,)$$

按相对频率解释

- ( ) :事件A发生的次数
- ( ) :事件M发生的次数
- (∩ :) 事件 A 和 M 发生的次数

  同时地。
- : 试验总数

$$(\ ) = \frac{-(\ )}{-} ; \quad (\ ) = \frac{-/(\ )}{-} ; \quad (\ \cap =) \ \frac{-\cap/(\ )}{-}$$

根据条件概率的定义：

$$(\ | \ ) = \frac{\frac{(\ \cap \ )}{}}{\underline{(\ )}} = \frac{(\ \cap \ )}{(\ )}$$

P(A)计算
关于n(M)而不是
n,试验总数

独立活动

两个事件 A 和 B 是独立的当且仅当：

$(\quad \cap =)) (\quad (\quad)$

两个事件 A 和 B 不相交当且仅当： $\cap = (\varnothing) = 0$

$(\quad)$

~~备注:如果~~ A 和 B 独立 :P(A/B) = P(A)

按相对频率解释：

$(\quad) = \frac{(\quad)}{\quad}; \quad (\quad | \quad) = \frac{(\quad \cap \quad)}{(\quad)}$ 这意味着……

如果 A 和 B 独立

$\frac{“\# \,)}{”} = \frac{“(\# \cap \% \,)}{” \,(\% \,)}$

如果事件A和B是独立的,则事件A的相对频率为

无论我们考虑试验总数还是仅考虑试验次数,结果都是一样的

事件 B 已发生。

例子

考虑一副 52 张牌（13 张牌分为四种花色：梅花 (♣)、方块 (♦)、红心 (♥) 和黑桃 (♠)。每种花色有 1 张 K、1 张 Q 和 1 张 Jack 牌）。下面两个事件A和B独立吗？

A = {我们画一个皇后} B = {我们画一颗心}

如果 A 和 B 是独立的，则 B 发生的事实根本不会改变 A 的概率！

示例:解决方案!

考虑一副 52 张牌。下面两个事件A和B独立吗?

A = {我们画一个皇后} B = {我们画一颗心}

根据概率的第二个定义:

$$P(A) = \frac{4}{52} \quad P(B) = \frac{13}{52} = \frac{1}{4} \quad \Rightarrow \quad P(A).P(B) = \frac{1}{52}$$

另外A∩B = {我们画一个红心皇后},即:

$$\frac{1}{52}$$

如果 A 和 B 是独立的,则 B 发生的事实根本不会改变 A 的概率!

全概率定律

考虑一组N成对不相交事件,表示为M1,⋯, MN
其并集是整个样本空间 ,即：

%

C „ =哦 和 ∩ ” & = ∅ , ≠

“#$ 相互排斥和详尽的

那么对于任意事件A：

1 1
( ) =- (∩ !) =- (| !)(!)

!0$!0$

例子：

实验:从 4500 枚硬币中抽取一枚硬币,其中 800 枚为假币,3700 枚为真币。硬币被放入 4 个盒子中:

盒子	全部的	公平的	伪造的
B1	2000	1 600	400
B2	500	300	200
B3	1000	900	100
B4	1000	900	100

问题:抽到假硬币的概率是多少?

$$P(F) = \sum_{i=1}^{4} P(F \mid \mathrm{B_i}).P(B_i)$$

$$P(F) = \frac{1}{4}(0{,}2 + 0{,}4 + 0{,}1 + 0{,}1) = 0{,}2$$

贝叶斯定理

给定一组N成对不相交事件,表示为

M1,···, MN其并集是整个样本空间

先验概率 P(Mi) ∀ i ∈ {1, . 。 。 , N},然后给定事件 A,其

中 P(A) > 0,假设 A 已发生， Mk的后验概率为：

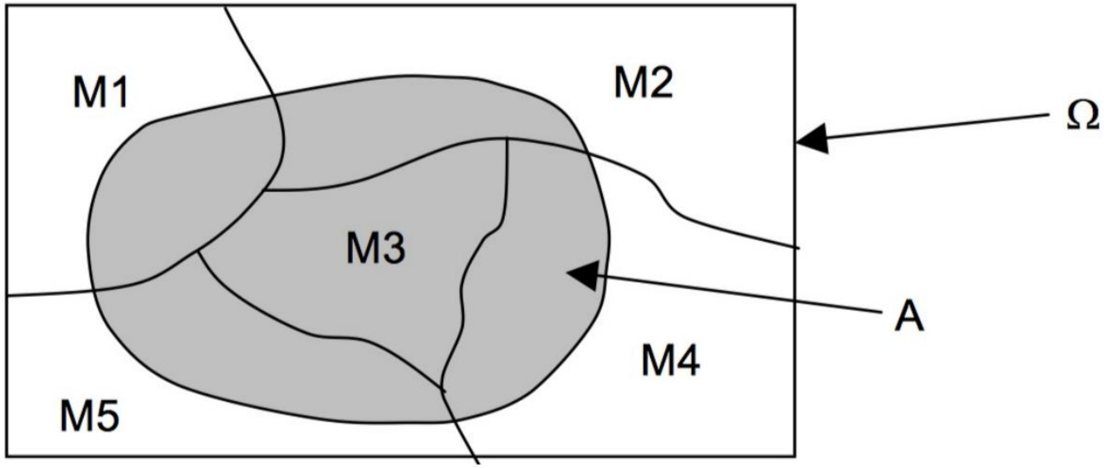
(2| =) (| 2) (2)

1

Σ!0$ (| !)(!)

总概率定律

贝叶斯定
理
方案



例子：

实验:从 4500 枚硬币中抽取一枚硬币,其中 800 枚为假币，
3700 枚为真币。硬币被放入
4 盒。我们抽了一枚假硬币。

问题： B2盒中出现假币的概率是多少?

$$P(B_2 \mid \mathrm{F}) = \frac{P(F|B_2)P(B_2)}{P(F)} = \frac{0{,}4 \,.\, 0{,}25}{0{,}2} = 0{,}5$$

备注： P(B2) = 0,25（先验概率）

P(B2/F) = 0,5（后验概率）

附录:计数方法

排列和组合

Type	Formulas	Explanation of Variables	Example
Permutation with repetition (Use permutation formulas *when order matters* in the problem.)	n^r	Where n is the number of things to choose from, and you choose r of them.	A lock has a 5 digit code. Each digit is chosen from 0-9, and a digit can be repeated. How many different codes can you have? n = 10, r = 5 $10^5 = 100{,}000$ codes
Permutation without repetition (Use permutation formulas *when order matters* in the problem.)	$\frac{n!}{(n-r)!}$	Where n is the number of things to choose from, and you choose r of them. Sometimes you can see the following notation for the same concept: $P(n,r) = {}^nP_r = {}_nP_r = \frac{n!}{(n-r)!}$	How many ways can you order 3 out of 16 different pool balls? n = 16, r = 3 $\frac{16!}{(16-3)!} = 3{,}360\ ways$
Combination with repetition (Use combination formulas *when order doesn't matter* in the problem.)	$\frac{(n+r-1)!}{r!\,(n-1)!}$	Where n is the number of things to choose from, and you choose r of them.	If there are 5 flavors of ice cream and you can have 3 scoops of ice cream, how many combinations can you have? You can repeat flavors. n = 5, r = 3 $\frac{(5+3-1)!}{3!\,(5-1)!} = 35\ combinations$
Combination without repetition (Use combination formulas *when order doesn't matter* in the problem.)	$\frac{n!}{r!\,(n-r)!}$	Where n is the number of things to choose from, and you choose r of them. Sometimes you can see the following notation for the same concept: $C(n,r) = {}^nC_r = {}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$	The state lottery chooses 6 different numbers between 1 and 50 to determine the winning numbers. How many combinations are possible? n = 50, r = 6 $\frac{50!}{6!\,(50-6)!} = 15{,}890{,}700\ combinations$