

第 4 章

条件概率、贝叶斯公式与独立性

教材节选 · 发布版本 2026-07-18

- 学习目标
- 本章的目标不是只会把数字代入公式，而是能回答三件事：
- 1. 已经知道的条件是谁？观察范围因此缩小到哪里？
 - 2. 结果可能由哪些互斥路径产生？每条路径的基准率是多少？
 - 3. 换一个基准率或删除一个前提后，原来的方法是否仍成立？

建议阅读顺序

页码	主题	阅读时自问
2	事件、交集与补集	我是否能说清同一总体里的四种对象？
3	条件概率	竖线右边改变了哪个分母？
4	全概率	观察结果由哪些互斥路径组成？
5	贝叶斯公式	反向条件为什么还要乘基准率？
6	10000 件产品例题	15.4% 的分子和分母分别是谁？
7	独立与互斥	不能同时发生是否等于互不影响？
8	练习与自检	我能否在新情境中先判断再计算？

材料说明

教材中的产品、设备与数据均为教学用虚构设定。请把例题当作概率模型，不对应任何真实厂商或产品质量结论。

4.1 先统一事件语言

概率问题中的字母不是公式装饰，而是把自然语言拆成可检查的事件。若事件定义含糊，后面的条件方向、交集与分母都会跟着出错。

一、事件、交集与补集

记号	在质检例中的含义	检查方法
D	随机抽到的产品确有缺陷	对象状态，只能是缺陷或合格
非 D	产品合格	D 的补事件
+	设备发出疑似缺陷报警	设备观察结果
-	设备不报警	+ 的补事件
Dn+	产品有缺陷且设备报警	两件事同时发生，即真报警
非 Dn+	产品合格但设备报警	误报警

二、把总体拆成四格

真实状态 / 设备结果	报警 +	不报警 -	合计
缺陷 D	Dn+	Dn-	D
合格 非 D	非 Dn+	非 Dn-	非 D
合计	+	-	全部产品

先记住这件事

同一个“报警”结果包含两类对象：真报警 Dn+ 与误报 非 Dn+。当题目问“报警后真实有缺陷”时，分母必须覆盖这两类报警对象。

三、最小前置自检

- “产品有缺陷”和“设备报警”能同时发生，因此不是互斥事件。
- D 与非 D 互斥且完备：每件产品必须且只能落在其中一类。
- + 与 - 也互斥且完备。
- 要把不同百分数相加，必须确认它们已经换到同一个总体或同一个分母。

4.2 条件概率：先缩小范围，再数目标

条件概率回答的是：在某件事已经发生的范围里，另一件事所占的比例。竖线右边不是补充说明，而是新的观察范围。

定义

$P(A|B) = P(A \cap B) / P(B)$ ，要求 $P(B) > 0$ 。

竖线阅读法

读法：已知 B 已经发生，问 A 的概率。

操作：先把全部对象缩小到 B，再在 B 里面数同时属于 A 的对象。

分母：全部 B。分子：A 与 B 同时发生。

方向对照

问题	概率符号	分母是谁	回答什么
缺陷品被设备报警的比例	$P(+ D)$	全部缺陷品	检出率
设备报警后确有缺陷的比例	$P(D +)$	全部报警品	后验概率
合格品被设备报警的比例	$P(+ \text{非 } D)$	全部合格品	误报率
未报警产品中合格品的比例	$P(\text{非 } D -)$	全部未报警品	阴性后的判断

常见误区

$P(+|D)$ 与 $P(D|+)$ 不是同一句话换语序。交换竖线两边会同时改变分母和问题。除非有额外条件，两者通常不相等。

小例：选课

某班 50 人，40 人选修 Python，其中 16 人同时选修统计学。已知一名学生选修 Python，他也选统计学的概率是 $16/40=40\%$ 。这里的分母不是全班 50 人，因为条件已经把范围缩小到 40 名 Python 学生。

自问

- 看到“在……中”“已知……”“……以后”时，我能否先圈出条件群体？
- 如果把竖线两侧交换，我能否说出分母发生了什么变化？

4.3 乘法与全概率：结果来自哪些路径

从条件概率定义移项，可以得到乘法公式：

$$P(A \cap B) = P(A|B)P(B)$$

在质检题中，设备报警可能来自两条路径：产品有缺陷且被检出，或产品合格但被误报。两条路径互斥，并覆盖全部报警。

报警的全概率

$$P(+)=P(+|D)P(D)+P(+|\text{非 } D)P(\text{非 } D)$$

路径	起始比例	路径内报警比例	在全部产品中的占比
D -> +	P(D)	P(+ D)	P(+ D)P(D)
非 D -> +	P(非 D)	P(+ 非 D)	P(+ 非 D)P(非 D)

为什么不能直接把两个条件概率相加

‘P(+|D)’以缺陷品为分母，‘P(+|非 D)’以合格品为分母。只有先乘各自基准率，把两条路径都换成“在全部产品中的占比”，才能相加。

分母警报

检查法：如果两个百分数后面的“每 100 个谁”不同，就先不要相加。先写出各自分母，再换到同一总体。

完备路径的要求

- 路径彼此不能重叠，否则会重复计算。
- 路径合起来必须覆盖要计算的全部结果，否则会漏项。
- 每条路径的条件概率必须与该路径的起始群体对应。

反例

若把报警只写成“缺陷且报警”，就漏掉了合格品误报；若把“缺陷且报警”与“所有报警”相加，又会把真报警重复计算。

4.4 贝叶斯公式：看到结果后反推原因

贝叶斯公式用于把原因到结果的条件概率，结合原因的基准率，转换为看到结果后原因成立的概率。

一般形式

$$P(A|B) = P(B|A)P(A) / P(B)$$

若 A 与非 A 构成完备划分：

$$P(A|B) = P(B|A)P(A) / [P(B|A)P(A) + P(B|\text{非 } A)P(\text{非 } A)]$$

质检形式

$$P(D|+) = P(+|D)P(D) / [P(+|D)P(D) + P(+|\text{非 } D)P(\text{非 } D)]$$

量	本题含义	是否可省略
$P(D)$	产品缺陷率，即基准率	不可省略
$P(+ D)$	缺陷品被检出的比例	不可省略
$P(+ \text{非 } D)$	合格品被误报的比例	不可省略
$P(D +)$	报警后真实有缺陷的比例	这是要求的结果

为什么基准率重要

假设设备参数完全不变。如果缺陷极少，合格品数量会远大于缺陷品；一个不高的误报率乘上巨大合格品基数，也可能产生很多误报。反之，缺陷率升高时，真报警在全部报警中的比例会上升。

信息不足时要停下

缺少 $P(D)$ 时，通常不能只凭检出率和误报率求出 $P(D|+)$ 。不要默认基准率是 50%，也不要把检出率直接当成后验概率。

两个不同问题

- 设备能力问题：真实缺陷品中，设备能抓到多少？看 $P(+|D)$ 。
- 报警解释问题：所有报警品中，真实缺陷有多少？看 $P(D|+)$ 。

4.5 例题：为什么答案不是 90%

某工厂一批产品中有 1% 的产品存在缺陷。设备对缺陷品的检出率为 90%，对合格品的误报率为 5%。随机抽取一件产品，设备报警。求产品真实有缺陷的概率。

第一步：选一个便于计数的总体

取 10000 件产品。用 100 件会出现 0.9 件，不便于直观计数。

第二步：按基准率拆总体

- 缺陷品：10000×1%=100 件。
- 合格品：10000×99%=9900 件。

第三步：在各自群体内使用检出率和误报率

- 缺陷品报警：100×90%=90 件；不报警 10 件。
- 合格品误报：9900×5%=495 件；不报警 9405 件。

真实状态 / 设备结果	报警 +	不报警 -	合计
缺陷 D	90	10	100
合格 非 D	495	9405	9900
合计	585	9415	10000

第四步：在全部报警品中数真实缺陷

$$P(D|+) = 90 / (90+495) = 90/585 \approx 15.38\%$$

解释结果

15.38% 不表示设备只能抓住 15.38% 的缺陷品。设备仍抓住了 100 件缺陷品中的 90 件。后验较低，是因为 9900 件合格品产生了 495 次误报。

单变量检查

只改变一项	真报警	误报	报警后真缺陷
缺陷率改为 10%	900	450	900/1350=66.67%
误报率降为 2%	90	198	90/288=31.25%

4.6 独立与互斥：两个不同维度

“互斥”和“独立”在日常语言里都容易被说成“互不相关”，但它们回答的问题不同。

关系	回答的问题	判断式	关键词
互斥	能不能同时发生？	$P(A \cap B) = 0$	共同部分为空
独立	知道一个发生会不会改变另一个概率？	$P(A B) = P(A)$ 或 $P(A \cap B) = P(A)P(B)$	概率不被改变

例 1：正概率互斥事件通常不独立

一次掷骰子， A =点数为 1， B =点数为 6。 A 与 B 互斥。原本 $P(B)=1/6$ ；若已知 A 发生，则 $P(B|A)=0$ 。条件改变了 B 的概率，所以不独立。

例 2：独立事件可以同时发生

掷两枚公平硬币， A =第一枚正面， B =第二枚正面。 $P(A)=1/2$ ， $P(A|B)=1/2$ ，因此独立。两枚都为正面时 A 、 B 同时发生，所以它们不互斥。

例 3：可同时发生也可能不独立

一次掷骰子， A =偶数， B =点数大于 3。它们能同时发生（4、6），因此不互斥；但 $P(A)=1/2$ ， $P(A|B)=2/3$ ，因此也不独立。

两步判断法

先问“能否同时发生”判断互斥；再问“条件概率是否改变”判断独立。不要只凭‘看起来互不影响’下结论。

边界提醒

- 条件概率定义要求条件事件概率大于 0。
- 当 A 、 B 中有零概率事件时，独立与条件概率的表述需要按课程定义谨慎处理。
- 本章练习默认有限离散情形，后续连续型情形会使用密度与条件分布。

4.7 章末练习与自检

闭卷规则

先独立作答，再核对教材与教师反馈。不要一边做题一边看例题 4-5 的人数表。

A. 条件方向

1. 垃圾邮件过滤器能拦下 96% 的垃圾邮件。用概率符号表示这句话，并写出分母群体。
2. 已知一封邮件被系统拦截，求它实际是垃圾邮件的概率。写出符号，并说明它与第 1 题是否相同。

B. 人数表与基准率

3. 5000 笔交易中有 2% 是欺诈交易。模型能识别 85% 的欺诈交易，并把 3% 的正常交易误报为欺诈。列四格人数表，求报警后为欺诈的概率。
4. 保持模型参数不变，只把欺诈率从 2% 改为 8%。先预测报警后为欺诈的概率会上升、下降还是不变，再计算。
5. 若第 3 题不提供欺诈率，只给识别率和误报率，能否求出唯一后验概率？说明理由。

C. 独立与互斥

6. 一次抽牌中，A=抽到红桃，B=抽到黑色。判断是否互斥、是否独立，并写出判断依据。
7. 有放回地连续抽两次牌，C=第一次红桃，D=第二次红桃。判断是否互斥、是否独立，并说明“有放回”为什么重要。

D. 解释

8. 不超过 80 字解释：为什么一个检出率很高的系统，报警后真的命中比例仍可能不高？

自检清单

- 我先写清已知条件和目标事件，再写符号。
- 我能指出每个百分数的分母，不直接相加不同分母的比例。
- 我会检查基准率是否出现并进入计算。
- 信息不足时，我能停来说明缺什么。
- 我用概率关系判断独立与互斥，不只写日常语言。

本节选结束。课程结论以本文件版本 2026-07-18 为准；若后续勘误与本文件冲突，以课程平台最新正式勘误为准。