
条件概率与事件独立性

第 3 讲 · 概率论与数理统计

复旦大学经济学院 ECON130001

上一讲我们做了什么

两种概率模型

古典概型 $P(A) = \frac{\#A}{\#\Omega}$ (结果有限且等可能) ; **几何概型** $P(A) = \frac{m(A)}{m(\Omega)}$ (同一个式子, 计数换成测度) 。

两种模型共用的那条假设

两种模型都建立在**等可能**上, 而「对什么等可能」由建模的人指定, 不由公理决定。贝特朗悖论: 指定的对象不同, 同一道题能给出 $\frac{1}{3}$ 、 $\frac{1}{4}$ 、 $\frac{1}{2}$ 三个答案。

★ 留下的问题

到现在为止, 算概率都是一**次算完**的: 定下样本空间, 数一次或量一次, 答案就出来了, 手上的信息自始至终没有变过。可现实里信息是**陆续到来**的——三门问题里, 主持人一开门, 那扇门的概率就变了。 **已经知道了一部分事实, 剩下的概率该怎么改?**

为什么需要这一讲

背景 一种病，人群中每一万人有一个。检测方法「可靠性 90%」：真患病的人 90% 会测出阳性，没病的人有 10% 会误报。你去做了检测，**结果是阳性。**

转折 你多半会想：「90% 准，那我大概有九成可能真的病了。」

这不是外行的想法。1978 年 Casscells 等人在《新英格兰医学杂志》上，拿一道结构完全相同的题去问**哈佛医学院的教师、住院医师和学生**：发病率千分之一、误报率 5%，测出阳性后患病概率是多少？**多数人答 95%。**正确答案约 2%。

问题 我们这道题的正确答案是 **0.09%——不到千分之一。** 直觉的答案和它差了一千倍。
同样几个数字，为什么算出来天差地别？问题出在哪一步？

来源

Casscells W., Schoenberger A., Graboys T. B. (1978). *Interpretation by Physicians of Clinical Laboratory Results*. *New England Journal of Medicine*, 299(18), 999–1001. 题设为发病率 1/1000、假阳性率 5%，多数受访者答 95%，正确答案约 2%。 pubmed.ncbi.nlm.nih.gov/692627 | 访问日期 2026-08-01。

本页正文所用的「发病率万分之一、检出率 90%、误报率 10%」是本课程教材的题设，与该研究的数字不同，结论方向一致。

本讲学习目标

- ① 会用**条件概率**把「已知一部分事实」写成样本空间的收缩
- ② 掌握由它推出的三个公式——**乘法法则**、**全概率公式**、**贝叶斯定理**——并说清各自用在什么场合
- ③ 说清**独立**与**互斥**的区别，并知道独立性不成立时，按独立去算会错多少

经济金融

- 市场出了一个利好消息，这家公司股价上涨的概率变成多少？
- 客户出现逾期记录之后，他将来违约的概率该调整到多少？

人工智能

- 收到一封带「免费」「中奖」字样的邮件，它是垃圾邮件的概率有多大？
- 传感器测到一个移动的物体，判断它是「行人」的概率该更新为多少？

四个问题问的是同一件事：**拿到新信息之后，原来那个概率该改成多少**。条件概率就是做这件事的工具，而贝叶斯定理是AI 做推理与学习的核心算法之一。

先看一个概率怎么被信息改变

例

彩票从 1–30 中选 6 个数字。你买了一注 {1, 7, 15, 21, 26, 30}。

1. 中奖的概率是多少？
2. 如果你**已经知道 15 是中奖号码之一**，中奖概率变成多少？

解

$$(1) \quad \frac{1}{C_{30}^6} = \frac{1}{593775} \approx 1.7 \times 10^{-6}$$
$$(2) \quad \frac{1}{C_{29}^5} = \frac{1}{118755} \approx 8.4 \times 10^{-6}$$

知道一个号码之后，中奖概率**涨到 5 倍**。

发生了什么

概率本身没变，**变的是我们站在哪个样本空间上看它**。第 (2) 问里， Ω 已经从「全部 C_{30}^6 注」缩成了「含 15 的那些注」。

条件概率的定义

定义

在事件 B 发生的情况下事件 A 发生的概率，称为 A 在给定 B 下的 **条件概率**，记为 $P(A | B)$ 。

相对地， $P(A)$ 称为**无条件概率**。本课程只考虑 $P(B) > 0$ 的情形。

为什么它重要

概率论在数理统计中的一个主要用途，就是**当观察到某事件发生时，更新概率**。统计推断的整个过程——从数据反推模型——本质上都是在算条件概率。

读法上的一个提醒

$P(A | B)$ 读作「在 B 已发生的条件下 A 的概率」。竖线**不是**除号，也**没有**先后顺序的含义—— B 可以是后发生的事（比如已知结果反推原因）。

公式是怎么来的：样本空间的收缩

从古典概型推出条件概率公式

在古典概型下 $P(A) = \frac{\#A}{\#\Omega}$, $P(B) = \frac{\#B}{\#\Omega}$ 。

已知 B 发生，样本空间就**缩成了** B 。此时 A 也发生，意味着结果落在 $A \cap B$ 里：

$$P(A | B) = \frac{\#(A \cap B)}{\#B} = \frac{\#(A \cap B)/\#\Omega}{\#B/\#\Omega} = \frac{P(AB)}{P(B)}$$

于是有**条件概率公式**

$$P(A | B) = \frac{P(AB)}{P(B)}$$

关键在分母

无条件概率的分母是 Ω ，条件概率的分母是 B 。「除以 $P(B)$ 」这个动作，就是「把 B 当成新的全集重新归一化」。

条件概率公式：两道例题

例

回到彩票题，用公式重算第 (2) 问：

$$P(\text{中奖} \mid 15 \text{ 中}) = \frac{1/C_{30}^6}{C_{29}^5/C_{30}^6} = \frac{1}{C_{29}^5}$$

与刚才「直接在缩小的空间里数」结果一致。

例

同时掷两个骰子，已知**两数之和是奇数**，求和小于 8 的概率。

$B = \text{和为奇数}$, $P(B) = \frac{1}{2}$; $A \cap B = \text{和为 } 3, 5, 7$, 共 $2 + 4 + 6 = 12$ 种, $P(AB) = \frac{12}{36} = \frac{1}{3}$ 。

$$P(A \mid B) = \frac{1/3}{1/2} = \frac{2}{3}$$

例题：和为 7 先于和为 8

例

反复掷两个骰子，观察点数之和。求**和为 7 先于和为 8 出现**的概率。

解：把无穷次试验压成一次

大部分回合既不是 7 也不是 8，它们**什么也没决定**，可以整体忽略。只需看「第一次出现 7 或 8 时，它是 7 吗」。

记 A = 和为 7， B = 和为 7 或 8。和为 7 有 6 种，和为 8 有 5 种：

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{6/36}{11/36} = \frac{6}{11}$$

这道题的价值

题面是一个**无穷次**的试验，条件概率把它化成了一次。做法是**只在「已经分出 7 或 8」的那部分样本空间里比较**，其余回合与结果无关，可以整体略去。

乘法法则

概率的乘法法则

把条件概率公式移项即得

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

推广到 n 个事件（链式法则）：

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 A_2) \cdots$$

例

盒中有 r 个红球、 b 个蓝球，**依次不放回**取两个球。求第一个红、第二个蓝的概率。

$$P(AB) = P(A)P(B | A) = \frac{r}{r+b} \cdot \frac{b}{r+b-1}$$

什么时候用它

有些试验里**条件概率比无条件概率容易算**（比如「已经拿走一个红球之后」）。这时就反过来，用条件概率去拼出无条件概率。

例题：连续四次买牙膏

例

某人每次买牙膏时，有 $1/3$ 的概率买与上次相同的品牌， $2/3$ 的概率换品牌。第一次等可能地选 A 或 B。求**前两次买 A、后两次买 B** 的概率。

解：用链式法则逐步展开

记 A_k = 第 k 次买 A, B_k = 第 k 次买 B:

$$\begin{aligned} P(A_1 A_2 B_3 B_4) &= P(A_1) P(A_2 \mid A_1) P(B_3 \mid A_1 A_2) P(B_4 \mid A_1 A_2 B_3) \\ &= \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} = \frac{1}{27} \end{aligned}$$

每一个因子只依赖**上一次买了什么**，所以四重条件都能立刻写出来。

例题：抽两张牌

例

一张蓝牌、四张红牌（记为 A、B、C、D），随机取两张。

- (1) 已知 **A 被选中**，求两张都是红牌的概率；
- (2) 已知**至少一张是红牌**，求两张都是红牌的概率。

解

$$(1) \frac{C_3^1/C_5^2}{C_4^1/C_5^2} = \frac{3}{4} = 0.75 \quad (2) \frac{C_4^2/C_5^2}{1} = \frac{6}{10} = 0.6$$

两问值得对照着看：「A 被选中」比「至少一张红牌」是**更强的信息**——只有一张蓝牌，取两张必然至少有一张红牌，所以 (2) 的条件其实必然发生， $P = 1$ 。条件越强，条件概率与无条件概率差得越多。第 27 页会专门辨析这件事。

全概率公式：把复杂事件拆开算

例

两盒球：第一盒 60 红 40 蓝，第二盒 10 红 20 蓝。先随机取一盒，再从中随机取一球。求取到红球的概率。

$$P(B) = P(B | A_1)P(A_1) + P(B | A_2)P(A_2) = \frac{60}{100} \cdot \frac{1}{2} + \frac{10}{30} \cdot \frac{1}{2} = \frac{7}{15}$$

全概率定理

若 $A_1, A_2, \dots$ 构成**完备事件组**且都有正概率，则对任一事件 B

$$P(B) = \sum_i P(B | A_i)P(A_i)$$

证明思路

由完备事件组， $B = \bigcup_i (B \cap A_i)$ 且各项两两互斥；用可列可加性拆开，再对每一项用乘法法则 $P(BA_i) = P(B | A_i)P(A_i)$ 。这正是第 1 讲命题 5「按一个事件分情形」的概率版本。完整证明见教师笔记。

全概率公式的两道例题

例：硬币

盒里有 3 枚两面都是正的硬币、4 枚两面都是反的、2 枚正常硬币。随机取一枚抛一次，求得到正面的概率。

按取到哪种硬币分三类：

$$P(A) = 1 \cdot \frac{3}{9} + 0 \cdot \frac{4}{9} + \frac{1}{2} \cdot \frac{2}{9} = \frac{4}{9}$$

例：换牙膏（续）

规则同前，但第一次买 A 的概率为 $1/4$ 、买 B 为 $3/4$ 。求**第二次买 B** 的概率。

按第一次买了什么分两类：

$$\begin{aligned} P(B_2) &= P(B_2 | A_1)P(A_1) + P(B_2 | B_1)P(B_1) \\ &= \frac{2}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{3}{4} = \frac{5}{12} \end{aligned}$$

用全概率公式的三步

三步：① 找一个完备事件组（「按什么分类」）② 算每一类的权重 $P(A_i)$ ③ 算每一类里的条件概率 $P(B | A_i)$ ，然后加权平均。难点永远在第 ① 步——分对了类，后面是机械的。

例题：按第一轮得分分成 50 类

例

一个游戏，第一轮得分 X 等可能地取 1 至 50。之后反复玩，直到出现分数 $Y \geq X$ 为止。求 $Y = 50$ 的概率。

解：按第一轮得分分类

记 $B_i = \{X = i\}$ ，它们构成完备事件组， $P(B_i) = \frac{1}{50}$ 。

已知 $X = i$ ，后续只在 $\{i, i + 1, \dots, 50\}$ 这 $51 - i$ 个分数里等可能地首次命中，故

$$P(Y = 50 \mid B_i) = \frac{1}{51 - i}$$

$$P(Y = 50) = \sum_{i=1}^{50} \frac{1}{50} \cdot \frac{1}{51 - i} \approx 0.090$$

$P(Y = 50 \mid B_i)$ 这一步用的是上一节的做法：把「反复试到成功」只放在**相关的那些结果**里比较，与第 9 页的「7 先于 8」是同一个思路。

条件概率也是概率

运算性质

固定 B ，把 $P(\cdot | B)$ 看成一个函数，它**满足第 1 讲**的三条公理，因此第 1 讲的全部性质原样成立：

$$P(\bar{A} | B) = 1 - P(A | B)$$

$$P(A \cup C | B) = P(A | B) + P(C | B) - P(A \cap C | B)$$

证明思路

逐条验证三条公理：非负性显然； $P(\Omega | B) = P(\Omega B)/P(B) = 1$ ；可列可加性由 $(\bigcup_j A_j) \cap B = \bigcup_j (A_j \cap B)$ 与原可列可加性得到。完整证明见教师笔记。

不必另立一套性质

条件概率不需要新的运算法则。**但竖线右边必须始终是同一个 B** 。 $P(A | B) + P(A | \bar{B})$ 一般不等于 1——那是两个不同的样本空间上的数。

小结：三个工具的关系

- **条件概率** $P(A | B) = \frac{P(AB)}{P(B)}$ ——把已知的事实变成样本空间的收缩
- **乘法法则** $P(AB) = P(A | B)P(B)$ ——同一个式子反过来用，由条件概率拼出联合概率
- **全概率公式** $P(B) = \sum_i P(B | A_i)P(A_i)$ ——分类、加权、求和

三者是同一个式子的三种用法。把它们合起来，就得到下面的贝叶斯定理。

回到开场那道题

例

某病发病率**万分之一**。检测方法：真患病者 90% 检出阳性，未患病者 10% 误报。某人检测为阳性，求其实际患病的概率。

现在能说清楚在求什么了

检测之前，此人患病的概率是 0.0001——这是**先验概率**。

现在有了「检测阳性」这条新信息，要把它更新成 $P(\text{患病} | \text{阳性})$ ——这是**后验概率**。

注意方向：题目给的是 $P(\text{阳性} | \text{患病}) = 0.9$ ，要求的是 $P(\text{患病} | \text{阳性})$ 。**两者不是一回事。**

贝叶斯定理

定理

若 $A_1, A_2, \dots$ 构成完备事件组且都有正概率，则对任一 $P(B) > 0$ 的事件 B

$$P(A_m | B) = \frac{P(A_m)P(B | A_m)}{\sum_i P(A_i)P(B | A_i)}$$

证明思路

分子用乘法法则 $P(A_mB) = P(A_m)P(B | A_m)$ ，分母用全概率公式展开 $P(B)$ ，两者相除即条件概率的定义。**贝叶斯定理不是新东西，是前面三个工具的组合。** 完整证明见教师笔记。

怎么读这个式子

$$\underbrace{P(A_m | B)}_{\text{后验}} = \frac{\overbrace{P(B | A_m)}^{\text{似然}} \times \overbrace{P(A_m)}^{\text{先验}}}{\underbrace{P(B)}_{\text{归一化}}}$$

只看新证据不够——它要乘上「这件事本来有多常见」，再除以证据本身的概率。

解：0.09%

解

记 B_1 = 患病、 B_2 = 未患病、 A = 检测阳性：

$$\begin{aligned} P(B_1 | A) &= \frac{P(A | B_1)P(B_1)}{P(A | B_1)P(B_1) + P(A | B_2)P(B_2)} \\ &= \frac{0.9 \times 0.0001}{0.9 \times 0.0001 + 0.1 \times 0.9999} = \frac{0.00009}{0.10008} \approx 0.0009 \end{aligned}$$

为什么直觉会错一千倍

关键在**分母里的第二项**。设想 100 万人：真患病的只有 **100** 人，其中 90 人测出阳性；而健康的 999 900 人里，有 **99 990** 人被误报。

阳性的人共约 10 万，真患病的只有 90 个。**误报的绝对数量压倒了真报**——不是因为检测不准，是因为**健康的人实在太多了**。

交互：信贷违约的贝叶斯更新

怎么用

三个滑块：违约率（先验）、查全率、误报率。右边一万个方块是「自然频率」的图形版。

当堂要做的实验

把**违约率**从 2% 拖到 20%，其余不动。后验从 15% 变成 82%——**查全率与误报率一个都没动，只换了先验。**



交互演示：拖动先验与两类错误率，观察后验概率

https://fin-tech.fun/courseware/probability-statistics-undergrad/toolkit/bayes_credit.html

应用：AI 信用风险评估

场景

- **先验**：历史数据显示任一申请人未来违约的概率 $P(\text{违约}) = 0.01$
- **证据**：模型发现该申请人「近 3 个月有信用卡逾期记录」
- 违约客户中有此特征的比例 $P(\text{逾期} | \text{违约}) = 0.6$
- 未违约客户中有此特征的比例 $P(\text{逾期} | \text{不违约}) = 0.05$

后验

$$P(\text{违约} | \text{逾期}) = \frac{0.6 \times 0.01}{0.6 \times 0.01 + 0.05 \times 0.99} = \frac{0.006}{0.0555} \approx 10.8\%$$

风险判断从 **1% 更新到 10.8%**，涨了十倍。

10.8% 的另一面

10.8% 意味着**被标记的人里，近九成其实不会违约**。直接据此拒贷会错杀大量好客户。「**风险上升十倍**」与「**这个人会违约**」是两回事。

例题：次品来自哪台机器

例

三台机器 M_1, M_2, M_3 分别生产 20%、30%、50% 的产品，次品率分别为 1%、2%、3%。随机抽到一件次品，求它是 M_2 生产的概率。

解

$$P(B_2 | A) = \frac{0.02 \times 0.30}{0.01 \times 0.20 + 0.02 \times 0.30 + 0.03 \times 0.50} = \frac{0.006}{0.023} \approx 0.26$$

先验 $P(B_2) = 0.30$ ，后验 $P(B_2 | A) = 0.26$ ——**反而降了**。

为什么会降

把三台机器各自贡献的次品算出来： M_1 是 $0.20 \times 0.01 = 0.002$ ， M_2 是 $0.30 \times 0.02 = 0.006$ ， M_3 是 $0.50 \times 0.03 = 0.015$ ，合计 0.023。

于是在**全部次品**里， M_3 占 65%、 M_2 占 26%、 M_1 占 9%；而在**全部产品**里，三者是 50%、30%、20%。

M_2 在次品里的比重（26%）低于它在产品里的比重（30%），后验就比先验小。 后验与先验的大小关系，看的就是这两个比重哪个高。

独立性：什么时候信息不起作用

定义

若事件 A 发生的可能性**不受** B **发生与否的影响**，即

$$P(A \mid B) = P(A)$$

则称 A 对 B **独立**。

若 n 个事件 $A_1, \dots, A_n$ 中**任一事件发生的可能性都不受其余一个或几个事件发生与否的影响**，则称 $A_1, \dots, A_n$ **相互独立**。

等价刻画与性质

- A, B 独立 $\iff P(AB) = P(A)P(B)$ (对称, 故「独立」不分主客)
- 若 A, B 独立, 则 A 与 $\bar{B}$ 、 $\bar{A}$ 与 B 、 $\bar{A}$ 与 $\bar{B}$ 都独立
- 若 $A_1, \dots, A_n$ 相互独立, 则 $P(\bigcap_i A_i) = \prod_i P(A_i)$, $P(\bigcup_i A_i) = 1 - \prod_i P(\bar{A}_i)$

★ 独立 $\neq$ 互斥

互斥是集合关系 ($A \cap B = \emptyset$, 一个发生另一个必不发生, 极强的关联); **独立**是概率关系 ($P(AB) = P(A)P(B)$, 一个发生与否不改变另一个)。两个正概率事件若互斥, 则一定**不**独立。

独立性的两个例子

例

掷一枚骰子， $A =$ 掷得偶数， $B =$ 掷得 $\{1, 2, 3, 4\}$ 。 A 与 B 独立吗？

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{2}{3}, \quad AB = \{2, 4\}:$$

$$P(AB) = \frac{1}{3} = \frac{1}{2} \times \frac{2}{3} = P(A)P(B)$$

独立。——虽然两个事件明显有重叠，看上去毫不「无关」。

例

抛硬币两次。 $A =$ 第一次正面， $B =$ 第二次正面， $C =$ 两次结果相同。 考察三者的独立性。

三个事件**两两独立**（各自验证 $P = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2}$ ），但 $P(ABC) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$ 。

两两独立不等于相互独立。

左例：独立与否要靠验证 $P(AB) = P(A)P(B)$ 来判定，**看不出来**。 右例：多个事件的独立性，必须**每一个子集**都验证乘法式，不能只两两验。

独立重复试验与贝努里定理

例：同一台机器的三问

机器次品率 p ，各件是否次品相互独立。

- 抽 6 件，**恰有 2 件**次品： $C_6^2 p^2 (1-p)^4$
- 抽 6 件，**至少 1 件**次品： $1 - (1-p)^6$
- 抽 n 件，**恰有 k 件**次品： $C_n^k p^k (1-p)^{n-k}$

第二问用了对立事件：**「至少一个」先算反面**，与第 1 讲的做法相同。

贝努里定理

n 次重复试验相互独立、每次事件 A 的概率均为 p ，称为 n **重贝努里试验**。 A 恰好发生 k 次的概率

$$p_n(k) = C_n^k p^k (1-p)^{n-k}$$

C_n^k 又称**二项式系数**，因为 $(a+b)^n = \sum_{k=0}^n C_n^k a^{n-k} b^k$ 。

再看一道：一直抽到发现**第 5 件**次品为止，恰好抽了 n 件的概率是 $C_{n-1}^4 p^5 (1-p)^{n-5}$ ——前 $n-1$ 件里恰好 4 件次品，第 n 件必须是次品。

$p_n(k)$ **这个式子请记住**。第 4 讲它会有个名字：**二项分布**。

辨析：两条听起来差不多的信息

设定

某家庭有 5 个孩子，每个孩子蓝眼睛的概率 $1/4$ ，相互独立。问「至少有三个孩子是蓝眼睛」的概率，在以下两种已知条件下分别是多少？

(a) 已知至少一个是蓝眼睛

「至少一个」的概率是 $1 - (3/4)^5 = \frac{781}{1024}$ ：

$$\begin{aligned} P(B_{\geq 3} \mid B_{\geq 1}) &= \frac{106/1024}{781/1024} \\ &= \frac{106}{781} \approx 0.136 \end{aligned}$$

(b) 已知最小的那个是蓝眼睛

最小的已确定，只需剩下 4 个里再出现 ≥ 2 个：

$$\begin{aligned} &\sum_{k=2}^4 C_4^k \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{4-k} \\ &= \frac{67}{256} \approx 0.262 \end{aligned}$$

为什么差了近一倍

因为「**最小的那个是蓝眼睛**」是强得多的信息。(a) 只排除了「五个都不是蓝眼睛」这一种情形，排掉的概率是 $(3/4)^5 \approx 24\%$ ；(b) 排除了「最小的那个不是蓝眼睛」的全部情形，排掉的概率是 $3/4$ 。**条件排除掉的可能性越多，剩下的样本空间越小，条件概率的值就越大。**

公式小结

名称	公式	什么时候用
条件概率	$P(A B) = \frac{P(AB)}{P(B)}$	已知一部分事实，重算概率
乘法法则	$P(AB) = P(A B)P(B)$	条件概率好算，联合概率难算
全概率公式	$P(B) = \sum_i P(B A_i)P(A_i)$	由「原因」算「结果」
贝叶斯定理	$P(A_m B) = \frac{P(A_m)P(B A_m)}{\sum_i P(A_i)P(B A_i)}$	由「结果」反推「原因」
独立性	$P(A B) = P(A) \iff P(AB) = P(A)P(B)$	判断信息是否起作用
贝努里定理	$p_n(k) = C_n^k p^k (1 - p)^{n-k}$	n 次独立重复，恰好成功 k 次

独立性假设：一把双刃剑

假设不成立时

- 现实中大量事件是**相互关联**的
- 错误地假设独立、然后套用 $P(A_1 \cdots A_n) = \prod P(A_i)$, 结论可能与事实差**若干个数量级**
- 下一页就是一个真实的例子

明知不成立也照用：朴素贝叶斯

- **朴素贝叶斯分类器**做文本分类时, 「天真地」假设一封邮件里每个词的出现相互独立
- 即 $P(\text{「免费」且「发票」} \mid \text{垃圾}) = P(\text{「免费」} \mid \text{垃圾}) \times P(\text{「发票」} \mid \text{垃圾})$
- 这个假设明显不符合语言事实, **但分类效果相当好, 而且计算量小得多**

两边的区别在于代价：朴素贝叶斯即使判错一封邮件, 成本也有限; 下一页那个例子里, 假设错了会让风险被低估十几个数量级。 **能不能借独立性简化, 要看假设失效时会付出什么。**

案例： *People v. Collins*

案情

1964 年洛杉矶。一位老妇人推着购物车走过小巷时被人推倒，钱包被抢。 警方依目击者描述的特征，几天后逮捕了一对夫妇。

检方请来一位数学教师，为目击者提到的六项特征各指定一个出现频率， 然后把它们**相乘**：

黄色汽车	1/10	留小胡子的男子	1/4
扎马尾的女子	1/10	金发女子	1/3
蓄络腮胡的黑人男子	1/10	车内为跨种族夫妇	1/1000

$$\frac{1}{10} \times \frac{1}{4} \times \frac{1}{10} \times \frac{1}{10} \times \frac{1}{3} \times \frac{1}{1000} = \frac{1}{12\,000\,000}$$

检方据此主张：符合全部特征的夫妇千万里挑一，被告因此被定罪。

问题一：这六项特征独立吗

相乘的前提是相互独立，而这六项显然不是：留小胡子与蓄络腮胡不独立，扎马尾与金发也不独立。 **把相关的特征当独立相乘，会把「存在这样一对夫妇」的概率算得过小。**

案例：结构化产品的独立性假设

场景

把 100 笔住房抵押贷款打包成一个结构化产品。单笔违约概率 $p = 5\%$ 。产品的优先级份额只在**违约笔数超过 30** 时才会受损。这份额有多安全？

假设各笔违约相互独立

用贝努里定理：

$$P(\text{至少 30 笔违约}) = \sum_{k=30}^{100} C_{100}^k (0.05)^k (0.95)^{100-k} \\ \approx 8.6 \times 10^{-16}$$

按这个数，这份额看起来几乎无风险。

但如果它们高度相关

房价是**共同因素**：一个区域的房价一起跌，违约就会一起发生。极端情形（完全相关）下，要么几乎没人违约，要么大家一起违约：

$$P(\text{至少 30 笔违约}) \approx p = 5\%$$

同一批贷款，两个假设下的风险差了约 14 个数量级。

2007–2008 年，大量结构化产品正是在「违约相关性很低」这一假设下拿到最高评级的。房价同步下跌后假设失效——**错的不是乘法法则，是它的前提。**

来源与说明

上述两个概率为本课程的**教学化演算**： $n = 100$ 、 $p = 0.05$ 的二项分布尾概率，与完全相关情形下的对照，均可用本讲公式直接验证，不代表任何具体产品的真实参数。2007–2008 年结构化产品评级失效与违约相关性假设的关系，属金融危机研究的公认结论，此处仅作定性引用。

用本讲的工具搭一个垃圾邮件分类器

第一步：用贝叶斯定理反推

想知道 $P(\text{垃圾} \mid \text{内容})$ ，直接算很难；但反过来统计「垃圾邮件里出现某些词的频率」很容易。于是

$$\underbrace{P(\text{垃圾} \mid \text{内容})}_{\text{后验}} = \frac{\overbrace{P(\text{内容} \mid \text{垃圾})}^{\text{似然}} \times \overbrace{P(\text{垃圾})}^{\text{先验}}}{P(\text{内容})}$$

第二步：那个「朴素」的独立性假设

「内容」含很多词， $P(\text{「免费」且「发票」} \mid \text{垃圾})$ 依然难算。于是假设**给定类别后各词相互独立**：

$$P(\text{「免费」且「发票」} \mid \text{垃圾}) = P(\text{「免费」} \mid \text{垃圾}) \times P(\text{「发票」} \mid \text{垃圾})$$

复杂的联合概率被拆成一串乘积。

第三步：决策

分别算出「是垃圾」与「是正常」的后验，哪个大就归哪一类。 **整个分类器用到的工具，本讲全部讲过了：条件概率、乘法法则、全概率公式、贝叶斯定理、独立性。**

本讲总结

① 回到开场那个问题

阳性结果下真患病的概率是 **0.09%**，不是 90%。直觉之所以错一千倍，是因为它把 $P(\text{阳性} | \text{患病})$ 当成了 $P(\text{患病} | \text{阳性})$ ——方向反了。

贝叶斯定理给出的正是这两个方向之间的换算关系：由 $P(\text{阳性} | \text{患病})$ 求 $P(\text{患病} | \text{阳性})$ ，必须乘上先验 $P(\text{患病})$ 。哈佛医学院的受访者不是不会算术，是漏掉了这一项。

② 核心结论

- 条件概率就是**样本空间的收缩**： $P(A | B) = P(AB)/P(B)$
- 由因算果用**全概率公式**，由果推因用**贝叶斯定理**
- 独立 $\iff P(AB) = P(A)P(B)$ ； n 重贝努里试验 $p_n(k) = C_n^k p^k (1-p)^{n-k}$

本讲总结（续）

③ 易错提醒

- $P(A | B) \neq P(B | A)$ 。本讲一切错误的源头
- **独立** $\neq$ **互斥**。互斥是集合关系（强关联），独立是概率关系（无影响）
- **两两独立** $\neq$ **相互独立**。必须每个子集都验证
- $P(A | B) + P(A | \bar{B})$ 一般**不等于 1**——两个不同的样本空间

④ 下一讲

注意本讲最后那个式子： $p_n(k) = C_n^k p^k (1 - p)^{n-k}$ 。它把「恰好成功 k 次」的概率写成了 k 的一个函数—— k 取遍 $0, 1, \dots, n$ ，就得到一整张概率表。

与其每次都问「某个事件的概率是多少」，不如**一次性描述一个数量的全部可能取值及其概率**。下一讲：把这个「数量」正式定义为随机变量。