

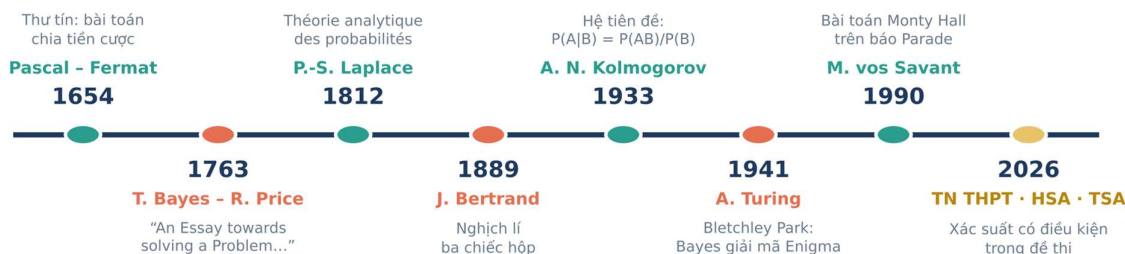
CHUYÊN ĐỀ BỒI DƯỠNG TƯ DUY ĐỊNH LƯỢNG



XÁC SUẤT CÓ ĐIỀU KIỆN CÔNG THỨC TOÀN PHẦN VÀ CÔNG THỨC BAYES

*Từ bài luận của Thomas Bayes (1763), nghịch lí Monty Hall (1990)
đến các câu hỏi Đúng/Sai trong đề tốt nghiệp THPT, HSA và TSA*

10 DẠNG TOÁN · 10 TÌNH HUỐNG THỰC TẾ · 30 BÀI LUYỆN TẬP CÓ LỜI GIẢI



Dành cho học sinh lớp 12 ôn thi tốt nghiệp THPT, Đánh giá năng lực (HSA – ĐHQG Hà Nội) và Đánh giá tư duy (TSA – ĐH Bách khoa Hà Nội)

MỤC LỤC

I. ĐẶT VẤN ĐỀ	4
1. Mười nghìn lá thư phản đối.....	4
2. Từ một bài luận được công bố sau khi tác giả qua đời	4
3. Vị trí trong chương trình và trong các kì thi	5
II. CƠ SỞ LÝ THUYẾT.....	6
1. Khi thông tin mới thu hẹp không gian mẫu	6
2. Tính chất.....	7
3. Công thức nhân và tính độc lập.....	7
4. Công thức xác suất toàn phần	8
5. Công thức Bayes	8
6. Ba công cụ trực quan.....	9
7. Mở rộng: dạng tỉ số của công thức Bayes	9
8. Quy trình bốn bước	10
III. CÁC DẠNG TOÁN VÀ VÍ DỤ MINH HOẠ.....	11
Dạng 1. Thu hẹp không gian mẫu – đếm trực tiếp.....	11
Dạng 2. Dữ kiện cho dưới dạng tỉ lệ – dùng định nghĩa	11
Dạng 3. Công thức nhân – dãy phép thử phụ thuộc.....	12
Dạng 4. Rút thăm lần lượt – thứ tự có ảnh hưởng không?.....	13
Dạng 5. Công thức xác suất toàn phần – phép thử nhiều giai đoạn.....	13
Dạng 6. Công thức Bayes trong xét nghiệm y khoa.....	15
Dạng 7. Bayes với nhiều giả thuyết – “ai là thủ phạm?”	18
Dạng 8. Cập nhật rồi dự báo – Bayes nối tiếp toàn phần.....	19
Dạng 9. Các nghịch lí kinh điển.....	20
Dạng 10. Độc lập, Bernoulli và điều kiện “có đúng k”.....	21
Tổng kết các dạng toán.....	22
IV. ỨNG DỤNG THỰC TẾ	23
V. BÀI TẬP LUYỆN TẬP THEO ĐỊNH DẠNG ĐỀ THI	28
A. Câu hỏi trắc nghiệm 4 lựa chọn.....	28
B. Câu hỏi trả lời ngắn	29
C. Câu hỏi Đúng/Sai.....	31
VI. ĐÁP ÁN VÀ HƯỚNG DẪN GIẢI	34
Bảng đáp án.....	34
Hướng dẫn giải	34
VII. KẾT LUẬN VÀ LỜI KHUYÊN KHI LÀM BÀI	37
VIII. TÀI LIỆU THAM KHẢO	37

I. ĐẶT VẤN ĐỀ

1. Mười nghìn lá thư phản đối

Tháng 9 năm 1990, trong chuyên mục “Ask Marilyn” của tạp chí *Parade*, Marilyn vos Savant trả lời câu hỏi của một độc giả: **Trong một trò chơi truyền hình có ba cánh cửa, sau một cửa là chiếc xe hơi, sau hai cửa còn lại là con dê. Bạn chọn một cửa. Người dẫn chương trình – người biết xe ở đâu – mở một trong hai cửa còn lại và cho thấy một con dê, rồi hỏi bạn có muốn đổi sang cửa chưa mở kia không. Nên đổi hay không?**

Câu trả lời của bà – *nên đổi, vì xác suất thắng tăng từ $1/3$ lên $2/3$* – khiến khoảng 10 000 độc giả viết thư phản đối, trong đó gần 1 000 người có học vị tiến sĩ, phần lớn khẳng định bà sai. Tương truyền ngay cả nhà toán học lừng danh Paul Erdős cũng chỉ chịu thuyết phục khi xem một mô phỏng trên máy tính. Câu đố thật ra đã được Steve Selvin nêu trên tạp chí *The American Statistician* từ năm 1975 dưới cái tên “bài toán Monty Hall”, và có một “tổ tiên” còn xa hơn: **ngịch lí ba chiếc hộp** của Joseph Bertrand trong cuốn *Calcul des probabilités* (1889).

Vì sao trực giác của cả những người được đào tạo bài bản lại hỏng ở đây? Bởi vì đây là một câu hỏi về **xác suất có điều kiện**: khi có thêm thông tin, xác suất phải được *cập nhật* – và cập nhật đúng cách thì không hề hiển nhiên.

2. Từ một bài luận được công bố sau khi tác giả qua đời

Mục sư người Anh Thomas Bayes (khoảng 1701 – 1761) để lại một bản thảo bàn về câu hỏi “ngược”: *biết một sự kiện đã xảy ra bao nhiêu lần, có thể nói gì về khả năng của nó?* Hai năm sau khi Bayes mất, bạn ông là Richard Price chỉnh lí và gửi đăng bài luận *An Essay towards solving a Problem in the Doctrine of Chances* trên tạp chí *Philosophical Transactions of the Royal Society* (tập 53, năm 1763, tr. 370 – 418). Price nhận xét rằng bài toán “ngược” này mới là thứ áp dụng trực tiếp được vào đời sống.

Độc lập với Bayes, Pierre-Simon Laplace phát biểu lại nguyên lí “xác suất của các nguyên nhân” năm 1774, rồi hệ thống hoá trong *Théorie analytique des probabilités* (1812). Đến năm 1933, A. N. Kolmogorov xây dựng lí thuyết xác suất trên hệ tiên đề, trong đó xác suất có điều kiện được **định nghĩa** bởi công thức quen thuộc $P(A | B) = \frac{P(AB)}{P(B)}$.

Trong Chiến tranh thế giới thứ hai, tại Bletchley Park, Alan Turing và cộng sự xây dựng phương pháp *Banburismus* – dùng xác suất có điều kiện để liên tục cập nhật niềm tin về các thiết lập của máy mã Enigma, rút ngắn đáng kể thời gian giải mã (theo S. B. McGrayne, *The Theory That Would Not Die*, 2011). Ngày nay, cùng một công thức đứng sau bộ lọc thư rác, chẩn đoán y khoa, hệ thống gợi ý và trí tuệ nhân tạo.

Nhìn lại bằng con mắt lịch sử

Pascal – Fermat (1654) đặt nền móng cho xác suất → Bayes – Price (1763) đặt ra “bài toán ngược” → Laplace (1774, 1812) tổng quát hoá → Bertrand (1889) chỉ ra các nghịch lí → Kolmogorov (1933) định nghĩa chặt chẽ → Turing (thập niên 1940) đưa vào ứng dụng. Một ý tưởng xuyên suốt: *thông tin mới không thay đổi thế giới, nhưng thay đổi điều ta biết về thế giới – và xác suất phải thay đổi theo.*

3. Vị trí trong chương trình và trong các kì thi

Theo Chương trình giáo dục phổ thông môn Toán 2018, xác suất có điều kiện, công thức xác suất toàn phần và công thức Bayes được học ở **lớp 12** (chẳng hạn trong bộ sách Kết nối tri thức với cuộc sống là Chương VI, gồm Bài 18 “Xác suất có điều kiện” và Bài 19 “Công thức xác suất toàn phần và công thức Bayes”). SGK chú trọng hai công cụ trực quan: **sơ đồ hình cây** và **bảng số liệu thống kê**.

Đề tham khảo tốt nghiệp THPT 2025 môn Toán gồm 12 câu trắc nghiệm, 4 câu Đúng/Sai (mỗi câu 4 ý) và 6 câu trả lời ngắn; các câu vận dụng tập trung vào bài toán thực tế, liên môn về hàm số, tích phân, **xác suất**, hình học không gian và tọa độ. Ở kì thi HSA, phần *Toán học và xử lí số liệu* (50 câu, 75 phút) có mạch Thống kê và xác suất; ở kì thi TSA, phần *Tư duy Toán học* (40 câu, 60 phút) ưa chuộng các tình huống y học, tài chính, kĩ thuật – đúng “sân nhà” của công thức Bayes.

Dạng **Đúng/Sai** đặc biệt hợp với chủ đề này: một bối cảnh (xét nghiệm, sản xuất, thi đấu...) kéo theo bốn ý hỏi lần lượt $P(A)$, $P(AB)$, $P(A | B)$ và $P(B | A)$ – thí sinh chỉ cần nhầm chiều điều kiện một lần là mất điểm cả câu.

Nguồn tư liệu và phạm vi kiến thức

Tài liệu tham khảo chính là bài giảng *Các bài toán cơ bản của xác suất có điều kiện* của PGS.TS Phan Huy Khải và ThS Nguyễn Thị Lanh (Hà Nội, 2024); nhiều ví dụ được lấy từ đó, có điều chỉnh cách trình bày và kiểm tra lại số liệu. Các bài toán kinh điển quốc tế được ghi rõ nguồn.

SGK Toán 12 trình bày công thức toàn phần và Bayes chủ yếu với hệ hai biến cố $B, \bar{B}$. Dạng tổng quát với n biến cố được đưa vào như phần **mở rộng** và được đánh dấu rõ. Các bài luyện tập ở mục V là **bài biên soạn theo định dạng** đề thi; mọi đáp số đã được kiểm tra bằng tính toán phân số chính xác.

II. CƠ SỞ LÝ THUYẾT

1. Khi thông tin mới thu hẹp không gian mẫu

Gieo hai con xúc xắc cân đối, không gian mẫu có 36 kết quả. Xét A : “tổng số chấm bằng 6” và B : “con thứ nhất ra số chấm lẻ”. Ta có $P(A) = \frac{5}{36}$. Nhưng nếu **biết** con thứ nhất ra số lẻ, những kết quả có con thứ nhất chẵn không còn khả năng xảy ra; không gian mẫu thu hẹp còn 18 kết quả, trong đó chỉ có 3 kết quả (1; 5), (3; 3), (5; 1) có tổng bằng 6.

		Con thứ hai →					
		1	2	3	4	5	6
Con thứ nhất	1	(1;1)	(1;2)	(1;3)	(1;4)	(1;5)	(1;6)
	2	(2;1)	(2;2)	(2;3)	(2;4)	(2;5)	(2;6)
	3	(3;1)	(3;2)	(3;3)	(3;4)	(3;5)	(3;6)
	4	(4;1)	(4;2)	(4;3)	(4;4)	(4;5)	(4;6)
	5	(5;1)	(5;2)	(5;3)	(5;4)	(5;5)	(5;6)
	6	(6;1)	(6;2)	(6;3)	(6;4)	(6;5)	(6;6)

Ω: 36 kết quả · A (tổng = 6): 5 kết quả

		Con thứ hai →					
		1	2	3	4	5	6
Con thứ nhất	1	(1;1)	(1;2)	(1;3)	(1;4)	(1;5)	(1;6)
	2	(2;1)	(2;2)	(2;3)	(2;4)	(2;5)	(2;6)
	3	(3;1)	(3;2)	(3;3)	(3;4)	(3;5)	(3;6)
	4	(4;1)	(4;2)	(4;3)	(4;4)	(4;5)	(4;6)
	5	(5;1)	(5;2)	(5;3)	(5;4)	(5;5)	(5;6)
	6	(6;1)	(6;2)	(6;3)	(6;4)	(6;5)	(6;6)

Biết B (con thứ nhất lẻ): còn 18 · AnB: 3

Hình 1. Thông tin “con thứ nhất lẻ” loại bỏ một nửa không gian mẫu

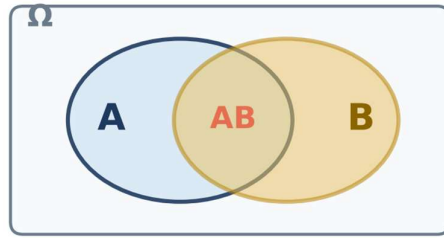
Xác suất của A khi biết B đã xảy ra là $\frac{3}{18} = \frac{1}{6}$ – lớn hơn $\frac{5}{36}$. Để ý $\frac{3}{18} = \frac{\frac{36}{18}}{\frac{36}{18}} = \frac{P(AB)}{P(B)}$. Từ đó có định nghĩa tổng quát.

Định nghĩa

Cho hai biến cố A và B với $P(B) > 0$. Xác suất của A với điều kiện B đã xảy ra, kí hiệu $P(A | B)$, là

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

Khi các kết quả đồng khả năng: $P(A | B) = \frac{n(AB)}{n(B)}$ – tức là “đếm lại trong không gian mẫu mới B ”.



Biết B đã xảy ra:

B trở thành “không gian mẫu mới”, A chỉ còn phần AB

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

Hình 2. Biết B đã xảy ra, B đóng vai trò không gian mẫu mới

2. Tính chất

Khi cố định B, hàm $A \mapsto P(A | B)$ là một xác suất “thật sự” trên không gian mẫu mới, nên mọi quy tắc quen thuộc vẫn đúng:

- $0 \leq P(A | B) \leq 1; P(B | B) = 1; P(\bar{A} | B) = 1 - P(A | B)$
- $P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B) - P(A_1 A_2 | B)$

Ba cặp hay bị nhầm

(1) $P(A | B)$ nói chung **khác** $P(B | A)$. Xác suất “dương tính khi có bệnh” khác hẳn xác suất “có bệnh khi dương tính” (Dạng 6).

(2) $P(A | B) + P(\bar{A} | B) = 1$ luôn đúng, nhưng $P(A | B) + P(A | \bar{B})$ nói chung **không** bằng 1 (Ví dụ 3).

(3) “A và B” ($P(AB)$) khác “A nếu B” ($P(A | B)$). Ví dụ 1 cho thấy hai cách hỏi chỉ khác một chữ mà đáp số chênh nhau gần 2,4 lần.

3. Công thức nhân và tính độc lập

Từ định nghĩa suy ra **công thức nhân xác suất**:

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

và tổng quát cho dãy biến cố xảy ra lần lượt:

$$P(A_1 A_2 \dots A_n) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 A_2) \dots P(A_n | A_1 \dots A_{n-1})$$

Hai biến cố A, B **độc lập** khi việc biết B xảy ra không làm thay đổi xác suất của A: $P(A | B) = P(A)$, tương đương $P(AB) = P(A) \cdot P(B)$. Khi đó các cặp $A, \bar{B}; \bar{A}, B; \bar{A}, \bar{B}$ cũng độc lập.

Độc lập từng đôi chưa chắc độc lập toàn phần (S. N. Bernstein)

Một khối tứ diện đều có ba mặt lần lượt sơn đỏ, xanh, vàng và mặt thứ tư sơn cả ba màu. Gieo khối tứ diện, gọi Đ, X, V là các biến cố “mặt đáy có màu đỏ / xanh / vàng”. Mỗi biến cố có xác suất $\frac{1}{2}$, mỗi cặp có xác suất $\frac{1}{4}$ (chỉ mặt bốn màu) – nên **độc lập từng đôi**. Nhưng $P(ĐXV) = \frac{1}{4} \neq \frac{1}{8}$ – ba biến cố **không độc lập toàn phần**.

4. Công thức xác suất toàn phần

Nhiều khi ta không tính trực tiếp được $P(A)$ nhưng biết xác suất của A trong từng trường hợp. Chia không gian mẫu theo B và $\bar{B}$: $A = AB \cup A\bar{B}$ với hai phần xung khắc, nên theo công thức nhân:

Công thức xác suất toàn phần (SGK Toán 12)

Cho hai biến cố A, B với $0 < P(B) < 1$. Khi đó

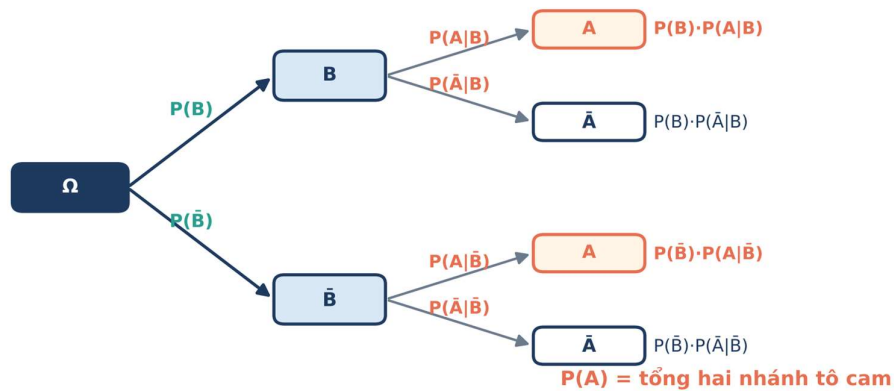
$$P(A) = P(B) \cdot P(A | B) + P(\bar{B}) \cdot P(A | \bar{B})$$

Mở rộng: hệ n biến cố

Nếu $B_1, \dots, B_n$ là một **hệ đầy đủ** (đôi một xung khắc, hợp là không gian mẫu, mỗi biến cố có xác suất dương) thì

$$P(A) = \sum_{i=1}^n P(B_i) \cdot P(A | B_i)$$

Dạng này không bắt buộc trong SGK nhưng rất tiện khi có 3 – 4 trường hợp (ba phân xưởng, bốn nhóm xạ thủ...). Học sinh hoàn toàn có thể làm bằng sơ đồ hình cây nhiều nhánh.



Hình 3. Sơ đồ hình cây: xác suất của một nhánh bằng tích các xác suất dọc nhánh

5. Công thức Bayes

Công thức toàn phần đi *xuôi* – từ nguyên nhân đến kết quả. Công thức Bayes đi *ngược*: biết kết quả A đã xảy ra, hỏi khả năng của nguyên nhân B . Theo định nghĩa và công thức nhân:

Công thức Bayes

$$P(B | A) = \frac{P(B) \cdot P(A | B)}{P(A)} = \frac{P(B) \cdot P(A | B)}{P(B) \cdot P(A | B) + P(\overline{B}) \cdot P(A | \overline{B})}$$

$P(B)$ gọi là xác suất **tiền nghiệm** (trước khi có thông tin), $P(B | A)$ là xác suất **hậu nghiệm** (sau khi biết A). Mở rộng cho hệ đầy đủ: $P(B_k | A) = \frac{P(B_k)P(A | B_k)}{\sum_i P(B_i)P(A | B_i)}$

Độc công thức trên cây: tử số là xác suất của *một* nhánh dẫn tới A , mẫu số là tổng *mọi* nhánh dẫn tới A . Câu khẩu quyết: **“Bayes = nhánh mình chia cho tất cả các nhánh cùng kết quả.”**

6. Ba công cụ trực quan

Công cụ	Khi nào dùng	Ưu điểm
Sơ đồ hình cây	Phép thử nhiều giai đoạn; dữ kiện cho dạng $P(\cdot \cdot)$	Thấy ngay công thức nhân và toàn phần
Bảng 2×2 (bảng tần số)	Số liệu khảo sát, thống kê; hai tiêu chí	Đổi xác suất thành “số người” – dễ hiểu, khó nhầm chiều
Hình vẽ diện tích / mảng chấm	Giải thích kết quả “phản trực giác”	Thấy được tỉ lệ nền (base rate) lần đầu

Nhà tâm lý học G. Gigerenzer (*Calculated Risks*, 2002) chỉ ra rằng khi dữ kiện được trình bày bằng **tần số tự nhiên** (“trong 1 000 người có 20 người bệnh...”) thay vì phần trăm, tỉ lệ người trả lời đúng bài toán Bayes tăng lên rõ rệt. Vì vậy trong tài liệu này, ta thường “đổi xác suất ra số người” trước khi tính.

7. Mở rộng: dạng tỉ số của công thức Bayes

Gọi **tỉ số khả năng** (odds) của B là $O(B) = \frac{P(B)}{P(\overline{B})}$. Chia công thức Bayes cho chính nó với B thay bằng $\overline{B}$ ta được:

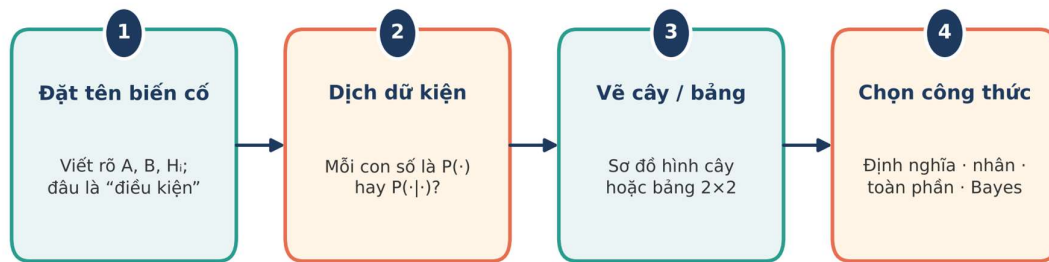
$$O(B | A) = O(B) \cdot \frac{P(A | B)}{P(A | \overline{B})}$$

Hệ số $L = \frac{P(A | B)}{P(A | \overline{B})}$ gọi là **hệ số khả dĩ** (likelihood ratio) của bằng chứng A . Câu khẩu quyết: **“tỉ số sau = tỉ số trước $\times$ hệ số khả dĩ”**. Khi có nhiều bằng chứng độc lập (khi đã biết B hay $\overline{B}$), chỉ việc nhân thêm các hệ số – rất tiện cho bài toán xét nghiệm nhiều lần (Ví dụ 13), bộ lọc thư rác (Tình huống 2). Đảo ngược: $P = \frac{O}{1 + O}$.

Ví dụ nhanh

Tỉ lệ mắc 2%: $O = \frac{2}{98} \approx 0,0204$. Xét nghiệm độ nhạy 95%, dương tính giả 10%: $L = 9,5$. Sau một lần dương tính: $O \approx 0,194$, $P \approx \frac{0,194}{1,194} \approx 0,162$ – khớp Ví dụ 12 mà không cần vẽ cây.

8. Quy trình bốn bước



Hình 4. Quy trình bốn bước giải bài toán xác suất có điều kiện

Dấu hiệu trong đề	Ý nghĩa	Kí hiệu
“biết rằng...”, “nếu...”, “trong số những...”, “khi đã...”	Phần sau là điều kiện	$P(A B)$
“...và...”, “đồng thời...”, “cả hai...”	Giao của hai biến cố	$P(AB)$
“tỉ lệ ... trong số ...”, “độ nhạy”, “tỉ lệ phế phẩm của máy I”	Xác suất có điều kiện cho sẵn	$P(A B_i)$
“chọn ngẫu nhiên..., thấy... Tính xác suất nó thuộc / do...”	Bài toán ngược	Bayes
“lấy lần lượt, không hoàn lại”, “chuyển từ hộp này sang hộp kia”	Phép thử nhiều giai đoạn phụ thuộc	Công thức nhân / toàn phần

III. CÁC DẠNG TOÁN VÀ VÍ DỤ MINH HOẠ

Dạng 1. Thu hẹp không gian mẫu – đếm trực tiếp

Ví dụ 1

Gieo ba con xúc xắc cân đối. a) Tính xác suất tổng số chấm bằng 8 **nếu** biết có ít nhất một con ra 1 chấm. b) Tính xác suất tổng số chấm bằng 8 **và** có ít nhất một con ra 1 chấm.

a) Gọi A : “tổng bằng 8”, B : “có ít nhất một con ra 1 chấm”. Biến cố đối $\bar{B}$ có $5^3 = 125$ kết quả nên $n(B) = 216 - 125 = 91$. Các bộ có tổng 8 chứa số 1: hoán vị của $(1; 1; 6) - 3$ bộ, $(1; 2; 5) - 6$ bộ, $(1; 3; 4) - 6$ bộ, tổng cộng 15. Vậy $P(A | B) = \frac{15}{91} \approx 0,165$.

b) $P(AB) = \frac{15}{216} = \frac{5}{72} \approx 0,069$ – nhỏ hơn gần 2,4 lần. Chữ “nếu” đã đổi mẫu số từ 216 thành 91.

Ví dụ 2 (bài toán hai đứa trẻ)

Một gia đình có hai con (xác suất sinh trai, gái như nhau). a) Biết gia đình có ít nhất một con trai, tính xác suất cả hai đều là trai. b) Biết con **đầu lòng** là trai, tính xác suất cả hai đều là trai.

Không gian mẫu $\{TT, TG, GT, GG\}$. a) Điều kiện còn $\{TT, TG, GT\}$ nên đáp số là $\frac{1}{3}$. b) Điều kiện còn $\{TT, TG\}$ nên đáp số là $\frac{1}{2}$. Hai câu hỏi tưởng như giống nhau nhưng thông tin khác nhau: “ít nhất một” không chỉ rõ đứa nào, còn “đầu lòng” xác định hẳn một đứa – không gian mẫu thu hẹp khác nhau.

Dạng 2. Dữ kiện cho dưới dạng tỉ lệ – dùng định nghĩa

Ví dụ 3

Một công ty dự thầu hai dự án. Xác suất thắng thầu dự án 1 là 0,4, dự án 2 là 0,5, cả hai dự án là 0,3. Tính xác suất công ty: a) chỉ thắng một dự án; b) thắng dự án 2, biết đã thắng dự án 1; c) thắng dự án 2, biết đã trượt dự án 1; d) thắng dự án 1, biết đã thắng dự án 2.

Gọi A, B là biến cố thắng thầu dự án 1, 2. Lập bảng xác suất 2×2 – mỗi ô điền được nhờ các tổng hàng, tổng cột:

	B	$\bar{B}$	Tổng
A	0,3	0,1	0,4
$\bar{A}$	0,2	0,4	0,6

	B	$\bar{B}$	Tổng
Tổng	0,5	0,5	1

a) $0,1 + 0,2 = 0,3$. b) $P(B | A) = \frac{0,3}{0,4} = 0,75$. c) $P(B | \bar{A}) = \frac{0,2}{0,6} = \frac{1}{3}$. d) $P(A | B) = \frac{0,3}{0,5} = 0,6$. Vì $0,4 \cdot 0,5 = 0,2 \neq 0,3$, hai biến cố **không độc lập** – thắng dự án này làm tăng cơ hội ở dự án kia. Lưu ý $P(B | A) + P(B | \bar{A}) = 0,75 + 0,33 \neq 1$.

Ví dụ 4

Xác suất một chuyến bay khởi hành đúng giờ là 0,83, đến đúng giờ là 0,82, vừa khởi hành vừa đến đúng giờ là 0,78. Tính xác suất chuyến bay: a) đến đúng giờ, biết đã khởi hành đúng giờ; b) đã khởi hành đúng giờ, biết đến đúng giờ; c) đến đúng giờ, biết khởi hành trễ.

a) $\frac{0,78}{0,83} \approx 0,94$. b) $\frac{0,78}{0,82} \approx 0,95$. c) $\frac{0,82 - 0,78}{1 - 0,83} = \frac{0,04}{0,17} \approx 0,24$. Câu c minh họa kỹ thuật hay dùng: $P(\bar{A}B) = P(A) - P(AB)$.

Dạng 3. Công thức nhân – dãy phép thử phụ thuộc

Ví dụ 5

Có 15 sinh viên, trong đó 5 sinh viên giỏi. Chia ngẫu nhiên 15 sinh viên thành 5 nhóm, mỗi nhóm 3 người. Tính xác suất mỗi nhóm đều có đúng một sinh viên giỏi.

Gọi A_i : “nhóm thứ i có đúng 1 sinh viên giỏi”. Lập lần lượt từng nhóm; mỗi lần chọn, số người còn lại giảm 3, số sinh viên giỏi giảm 1:

$$P = \frac{C_5^1 C_{10}^2}{C_{15}^3} \cdot \frac{C_4^1 C_8^2}{C_{12}^3} \cdot \frac{C_3^1 C_6^2}{C_9^3} \cdot \frac{C_2^1 C_4^2}{C_6^3} \cdot 1 = \frac{45}{91} \cdot \frac{28}{55} \cdot \frac{15}{28} \cdot \frac{3}{5} = \frac{81}{1001} \approx 0,081$$

Ví dụ 6

Cầu thủ A sút phạt đền lần đầu thành công với xác suất 0,8. Ở các lần sau, nếu lần trước ghi bàn thì xác suất ghi bàn là 0,9; nếu lần trước sút hỏng thì chỉ còn 0,4. Trong trận đấu, A sút 3 quả. Tính xác suất A ghi cả 3 bàn và xác suất A ghi đúng 2 bàn.

Cả 3 bàn: $0,8 \cdot 0,9 \cdot 0,9 = 0,648$. Đúng 2 bàn có ba kịch bản (G: ghi, H: hỏng): GGH = $0,8 \cdot 0,9 \cdot 0,1 = 0,072$; GHG = $0,8 \cdot 0,1 \cdot 0,4 = 0,032$; HGG = $0,2 \cdot 0,4 \cdot 0,9 = 0,072$. Tổng = 0,176. Đây chính là một xích Markov hai trạng thái – cầu nối sang chuyên đề *Quy hoạch động trong bài toán xác suất*.

Dạng 4. Rút thăm lần lượt – thứ tự có ảnh hưởng không?

Ví dụ 7

Trò chơi hái hoa có 10 phiếu, trong đó 5 phiếu có thưởng. Ba người lần lượt hái, mỗi người một phiếu (không hoàn lại). Tính xác suất mỗi người hái được phiếu thưởng.

Người thứ nhất: $\frac{5}{10} = \frac{1}{2}$. Người thứ hai, theo công thức toàn phần với hệ $A_1, \bar{A}_1$:

$$P(A_2) = \frac{1}{2} \cdot \frac{4}{9} + \frac{1}{2} \cdot \frac{5}{9} = \frac{1}{2}$$

Tương tự (chia bốn trường hợp theo kết quả của hai người đầu) ta được $P(A_3) = \frac{1}{2}$. **Mọi người đều có cơ hội như nhau, bất kể rút trước hay sau.**

Một bài thi thời sinh viên và người thầy

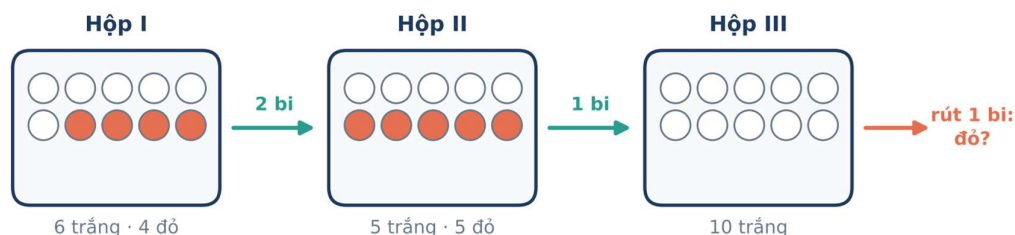
Tổng quát: trong n phiếu có m phiếu “có”, mỗi người rút lần lượt không hoàn lại đều có xác suất $\frac{m}{n}$ rút được phiếu “có”. Trong bài giảng của mình, PGS.TS Phan Huy Khải kể đây là bài thi kết thúc học phần Lí thuyết xác suất ông làm năm thứ hai tại khoa Toán, Trường Đại học Tổng hợp Hà Nội (khóa 1965 – 1969), do GS.TSKH Nguyễn Duy Tiến (1942 – 2020) giảng dạy – người đặt nền móng cho hệ đào tạo Cử nhân khoa học tài năng của ĐHQG Hà Nội.

Lập luận ngắn gọn: mọi cách xếp thứ tự các phiếu là đồng khả năng, nên phiếu ở vị trí thứ k “giống hệt” phiếu ở vị trí đầu tiên.

Dạng 5. Công thức xác suất toàn phần – phép thử nhiều giai đoạn

Ví dụ 8

Hộp I có 6 bi trắng, 4 bi đỏ; hộp II có 5 bi trắng, 5 bi đỏ; hộp III có 10 bi trắng. Lấy ngẫu nhiên 2 bi từ hộp I bỏ vào hộp II, trộn đều; sau đó lấy ngẫu nhiên 1 bi từ hộp II bỏ vào hộp III, trộn đều; cuối cùng lấy ngẫu nhiên 1 bi từ hộp III. Tính xác suất bi lấy ra cuối cùng màu đỏ.



Hình 5. Ba giai đoạn chuyển bi: mỗi giai đoạn chia trường hợp theo giai đoạn trước

Giai đoạn 1. Hai bi chuyển sang hộp II: 2 trắng ($\frac{C_6^2}{C_{10}^2} = \frac{1}{3}$), 2 đỏ ($\frac{2}{15}$), 1 trắng 1 đỏ ($\frac{8}{15}$). Hộp II khi đó có lần lượt 5, 7, 6 bi đỏ trên 12 bi.

Giai đoạn 2. Xác suất bi chuyển sang hộp III là đỏ:

$$P(C) = \frac{1}{3} \cdot \frac{5}{12} + \frac{2}{15} \cdot \frac{7}{12} + \frac{8}{15} \cdot \frac{6}{12} = \frac{29}{60}$$

Giai đoạn 3. Hộp III có 11 bi, chứa 1 bi đỏ nếu C xảy ra, không có bi đỏ nếu ngược lại: $P = \frac{29}{60} \cdot \frac{1}{11} = \frac{29}{660} \approx 0,044$.

Ví dụ 9

Hộp thứ nhất có 15 phiếu thi, hộp thứ hai có 9 phiếu; mỗi phiếu ghi một câu hỏi. Sinh viên A chỉ thuộc 10 câu ở hộp 1 và 8 câu ở hộp 2. a) Thầy rút ngẫu nhiên mỗi hộp 1 phiếu, rồi A rút ngẫu nhiên 1 trong 2 phiếu đó. b) Thầy rút 1 phiếu từ hộp 1 bỏ vào hộp 2, rồi A rút 1 phiếu từ hộp 2. c) Thầy rút 2 phiếu từ hộp 1 bỏ vào hộp 2, rồi A rút 2 phiếu từ hộp 2. Tính xác suất A trả lời được (câu c: cả hai câu).

a) $\frac{1}{2} \cdot \frac{10}{15} + \frac{1}{2} \cdot \frac{8}{9} = \frac{7}{9}$. b) $\frac{10}{15} \cdot \frac{9}{10} + \frac{5}{15} \cdot \frac{8}{10} = \frac{13}{15}$.

c) Hai phiếu chuyển sang: A thuộc cả hai ($\frac{3}{7}$), thuộc một ($\frac{10}{21}$), không thuộc ($\frac{2}{21}$). Hộp 2 khi đó có 11 phiếu, A thuộc lần lượt 10, 9, 8 phiếu:

$$P = \frac{3}{7} \cdot \frac{C_{10}^2}{C_{11}^2} + \frac{10}{21} \cdot \frac{C_9^2}{C_{11}^2} + \frac{2}{21} \cdot \frac{C_8^2}{C_{11}^2} = \frac{821}{1155} \approx 0,711$$

Ví dụ 10

Chuồng thứ nhất có 9 gà mái, 1 gà trống; chuồng thứ hai có 1 gà mái, 5 gà trống. Từ mỗi chuồng bắt ngẫu nhiên một con làm thịt, các con còn lại dồn vào chuồng thứ ba. Bắt ngẫu nhiên một con từ chuồng thứ ba. Tính xác suất bắt được gà trống.

Hai con bị bắt đi: cả hai trống ($\frac{1}{10} \cdot \frac{5}{6} = \frac{1}{12}$), cả hai mái ($\frac{9}{10} \cdot \frac{1}{6} = \frac{3}{20}$), một trống một mái ($1 - \frac{1}{12} - \frac{3}{20} = \frac{23}{30}$). Chuồng thứ ba có 14 con với số gà trống lần lượt 4, 6, 5:

$$P = \frac{1}{12} \cdot \frac{4}{14} + \frac{3}{20} \cdot \frac{6}{14} + \frac{23}{30} \cdot \frac{5}{14} = \frac{38}{105} \approx 0,362$$

Ví dụ 11 (bài toán trà sữa – theo bài giảng của PGS.TS Phan Huy Khải)

Hai anh em mỗi người một cốc trà sữa, mỗi cốc có 8 viên trân châu đen và 2 viên trắng. Anh chuyển ngẫu nhiên 1 viên sang cốc em; em trộn đều rồi chuyển lại ngẫu nhiên 1 viên

sang cốc anh. Sau đó anh hút ngẫu nhiên ở mỗi cốc 1 viên. Tính xác suất anh hút được 2 viên đen.

Hệ đầy đủ gồm bốn kịch bản theo màu viên đi (A : đen) và viên về (B : đen). Sau hai lần chuyển, số viên đen ở (cốc anh, cốc em) lần lượt là:

Kịch bản	Xác suất	Đen ở cốc anh / em	$P(\text{hai viên đen} \mid \text{kịch bản})$
đi đen – về đen	$\frac{8}{10} \cdot \frac{9}{11}$	8 / 8	$\frac{8}{10} \cdot \frac{8}{10}$
đi trắng – về đen	$\frac{2}{10} \cdot \frac{8}{11}$	9 / 7	$\frac{9}{10} \cdot \frac{7}{10}$
đi đen – về trắng	$\frac{8}{10} \cdot \frac{2}{11}$	7 / 9	$\frac{7}{10} \cdot \frac{9}{10}$
đi trắng – về trắng	$\frac{2}{10} \cdot \frac{3}{11}$	8 / 8	$\frac{8}{10} \cdot \frac{8}{10}$

$$P = \frac{72 \cdot 64 + 16 \cdot 63 + 16 \cdot 63 + 6 \cdot 64}{110 \cdot 100} = \frac{876}{1375} \approx 0,637$$

Bài toán đòi hỏi tổng hợp: công thức nhân (tính xác suất kịch bản), công thức toàn phần, và lập luận cẩn thận về thành phần mỗi cốc sau từng bước.

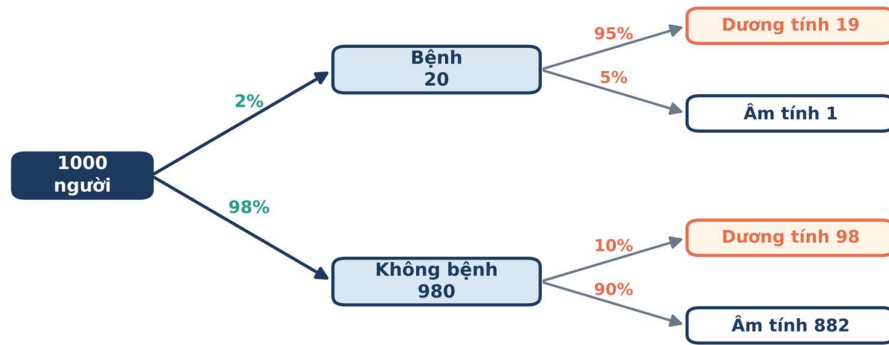
Dạng 6. Công thức Bayes trong xét nghiệm y khoa

Trong y học, với B : “có bệnh” và T^+ : “xét nghiệm dương tính”, người ta dùng các thuật ngữ:

Thuật ngữ	Kí hiệu	Ý nghĩa
Tỉ lệ mắc bệnh (tỉ lệ nền)	$P(B)$	Tiền nghiệm
Độ nhạy	$P(T^+ \mid B)$	Người bệnh được phát hiện
Độ đặc hiệu	$P(T^- \mid \overline{B})$	Người khỏe được kết luận âm tính
Tỉ lệ dương tính giả	$P(T^+ \mid \overline{B})$ $= 1 - \text{độ đặc hiệu}$	Người khỏe bị “báo động nhầm”
Giá trị dự đoán dương	$P(B \mid T^+)$	Điều bệnh nhân thật sự quan tâm

Ví dụ 12

Tỉ lệ mắc bệnh B ở một huyện miền núi là 2%. Xét nghiệm cho kết quả dương tính với 95% người bệnh và 10% người không bệnh. a) Tính xác suất một người được xét nghiệm có kết quả dương tính. b) Một người có kết quả dương tính; tính xác suất người đó mắc bệnh. c) Tính xác suất xét nghiệm cho kết luận đúng.



Trong 117 người dương tính, chỉ 19 người thật sự có bệnh: $19/117 \approx 16,2\%$

Hình 6. Đảo xác suất ra “số người” trên 1 000 người

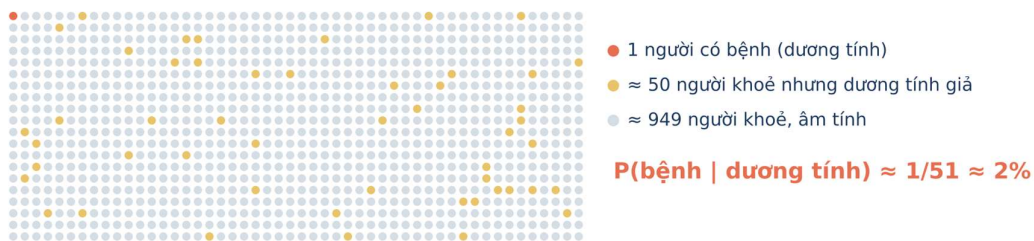
a) $P(T^+) = 0,02 \cdot 0,95 + 0,98 \cdot 0,1 = 0,117$. b) $P(B | T^+) = \frac{0,019}{0,117} = \frac{19}{117} \approx 0,162$. c) $0,02 \cdot 0,95 + 0,98 \cdot 0,9 = 0,901$.

Kết quả ở câu b gây ngạc nhiên: xét nghiệm “chính xác 90%” mà người dương tính chỉ có 16% khả năng mắc bệnh. Lí do nằm ở **tỉ lệ nền**: người khỏe đông gấp 49 lần người bệnh, nên dù chỉ 10% trong số họ bị báo nhầm, con số đó (98 người) vẫn lớn hơn 19 người bệnh thật.

Các bác sĩ cũng nhầm

Năm 1978, nhóm Casscells, Schoenberger và Graboys hỏi các bác sĩ, sinh viên y ở một bệnh viện thuộc Đại học Harvard: “Một bệnh có tỉ lệ mắc 1/1 000, xét nghiệm có tỉ lệ dương tính giả 5%. Một người dương tính – khả năng người đó mắc bệnh là bao nhiêu?” Câu trả lời phổ biến nhất là 95%. Năm 2014, Manrai và cộng sự (*JAMA Internal Medicine*) lặp lại đúng câu hỏi này và chỉ khoảng 23% người được hỏi trả lời đúng.

Đáp số (coi độ nhạy là 100%): $\frac{0,001}{0,001 + 0,999 \cdot 0,05} = \frac{20}{1019} \approx 2\%$.



Hình 7. 1 000 người: một chấm cam (người bệnh) giữa khoảng 50 chấm vàng (dương tính giả)

Ví dụ 13

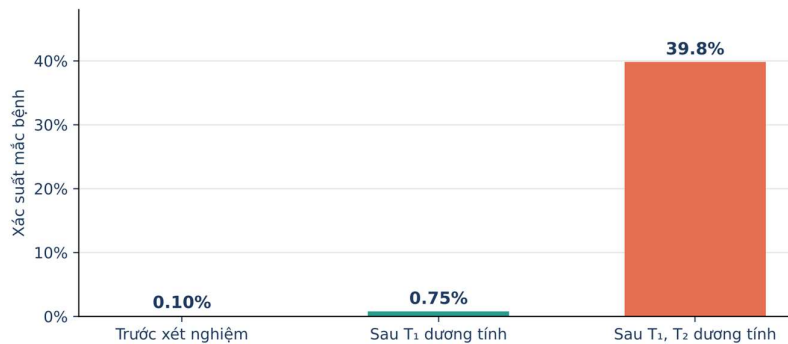
Một bệnh nhân bị nghi ung thư dạ dày với xác suất 0,001. Xét nghiệm T_1 có độ nhạy 90%, độ đặc hiệu 88%; xét nghiệm T_2 có độ nhạy 88%, độ đặc hiệu 99%. Hai xét nghiệm độc lập với nhau khi biết tình trạng bệnh. Bệnh nhân dương tính với cả hai. Tính xác suất người đó mắc bệnh. Thứ tự làm xét nghiệm có ảnh hưởng không?

Cách gọn nhất là dùng **dạng tỉ số (odds)** của công thức Bayes. Đặt $O = \frac{P(B)}{P(\bar{B})}$ (tỉ số mắc / không mắc). Mỗi kết quả dương tính nhân O với **hệ số khả dĩ** $L = \frac{\text{độ nhạy}}{1 - \text{độ đặc hiệu}}$:

$$L_1 = \frac{0,90}{0,12} = 7,5; \quad L_2 = \frac{0,88}{0,01} = 88; \quad O_{sau} = \frac{1}{999} \cdot 7,5 \cdot 88 = \frac{660}{999}$$

$$P(B | T_1^+ T_2^+) = \frac{O_{sau}}{1 + O_{sau}} = \frac{660}{1659} = \frac{220}{553} \approx 0,398$$

Phép nhân giao hoán nên **thứ tự xét nghiệm không ảnh hưởng** tới kết quả cuối. Hình dưới cho thấy “bước nhảy” của xác suất: $0,1\% \rightarrow 0,75\% \rightarrow 39,8\%$.



Hình 8. Mỗi lần có kết quả dương tính, xác suất mắc bệnh được cập nhật

Sai lầm thường gặp: nhầm độ đặc hiệu với tỉ lệ dương tính giả

Đề cho “độ đặc hiệu 80%” thì tỉ lệ dương tính giả là **20%**, không phải 10% hay 80%. Ví dụ: tỉ lệ mắc 20%, xét nghiệm có độ nhạy 90%, độ đặc hiệu 80%. Lời giải đúng: $P(T^+) = 0,2 \cdot 0,9 + 0,8 \cdot 0,2 = 0,34$ và $P(B | T^+) = \frac{0,18}{0,34} = \frac{9}{17} \approx 0,53$. Nếu lấy nhầm dương tính giả là 0,1 sẽ ra $\frac{9}{13} \approx 0,69$ – sai.

Tương tự, khi làm xét nghiệm thứ hai, phải dùng $P(T^+ | \bar{B}) = 1 - \text{độ đặc hiệu}$ của chính xét nghiệm đó cho nhánh “không bệnh”.

Ví dụ 14

Một người nghi mình mắc bệnh B, có thể chọn một trong hai xét nghiệm cùng chi phí: T_1 có độ nhạy 85%, độ đặc hiệu 90%; T_2 có độ nhạy 92%, độ đặc hiệu 82%. Tỉ lệ mắc bệnh trong dân số là 2%. a) Nên chọn xét nghiệm nào nếu tiêu chí là xác suất kết luận đúng? b) Người đó làm xét nghiệm đã chọn và dương tính; tính khả năng mắc bệnh. c) Nếu âm tính thì khả năng không mắc bệnh là bao nhiêu?

a) Xác suất kết luận đúng là trung bình có trọng số của độ nhạy và độ đặc hiệu: $P_1 = 0,002 \cdot 0,85 + 0,998 \cdot 0,90 = 0,8999$, $P_2 = 0,002 \cdot 0,92 + 0,998 \cdot 0,82 = 0,8202$. Chọn T_1 – khi bệnh hiếm, **độ đặc hiệu** quan trọng hơn độ nhạy.

b) $\frac{0,002 \cdot 0,85}{0,002 \cdot 0,85 + 0,998 \cdot 0,10} = \frac{17}{1015} \approx 1,7\%$. c) $\frac{0,998 \cdot 0,90}{0,998 \cdot 0,90 + 0,002 \cdot 0,15} = \frac{2994}{2995} \approx 99,97\%$. Với bệnh hiếm, kết quả âm tính gần như “loại trừ” bệnh, còn kết quả dương tính chỉ là tín hiệu cần xét nghiệm tiếp.

Dạng 7. Bayes với nhiều giả thuyết – “ai là thủ phạm?”

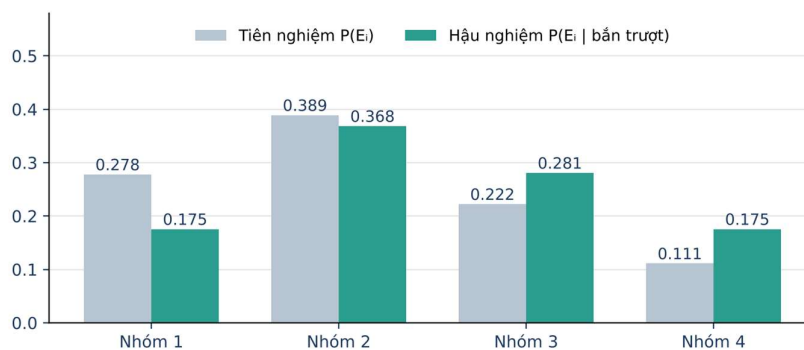
Ví dụ 15

Bốn nhóm xạ thủ lần lượt có 5, 7, 4, 2 người; xác suất bắn trúng của mỗi người trong từng nhóm là 0,8; 0,7; 0,6; 0,5. Chọn ngẫu nhiên một xạ thủ, người này bắn trượt. Xạ thủ đó có khả năng thuộc nhóm nào nhất?

Tiên nghiệm: $\frac{5}{18}, \frac{7}{18}, \frac{4}{18}, \frac{2}{18}$. Xác suất trượt theo nhóm: 0,2; 0,3; 0,4; 0,5. Toàn phần: $P(\text{trượt}) = \frac{5 \cdot 0,2 + 7 \cdot 0,3 + 4 \cdot 0,4 + 2 \cdot 0,5}{18} = \frac{19}{60}$. Bayes:

$$P(E_1 | \text{trượt}) = \frac{10}{57}, \quad P(E_2 | \text{trượt}) = \frac{21}{57}, \quad P(E_3 | \text{trượt}) = \frac{16}{57}, \\ P(E_4 | \text{trượt}) = \frac{10}{57}$$

Nhóm 2 có khả năng cao nhất. Để ý: sau khi biết “bắn trượt”, xác suất của nhóm 1 (bắn giỏi nhất) **giảm**, còn nhóm 3, 4 (bắn kém) **tăng** – thông tin mới đẩy niềm tin về phía các giả thuyết “giải thích tốt” dữ kiện hơn.



Hình 9. Tiên nghiệm và hậu nghiệm của bốn nhóm xạ thủ

Ví dụ 16

Một bệnh nhân nghi mắc một trong ba bệnh A, B, C với xác suất 0,3; 0,4; 0,3. Bốn bác sĩ khám độc lập, lần lượt chẩn đoán A, B, C, A. Mỗi bác sĩ chẩn đoán đúng với xác suất 0,6 và nhầm sang mỗi bệnh còn lại với xác suất 0,2. Đánh giá lại xác suất mắc từng bệnh.

Nếu bệnh là A: ba chẩn đoán (A, ?, ?, A) thì đúng ở bác sĩ 1 và 4, nhầm ở bác sĩ 2 và 3: $P(H | A) = 0,6 \cdot 0,2 \cdot 0,2 \cdot 0,6 = 0,0144$. Tương tự $P(H | B) = P(H | C) = 0,0048$. Khi đó $P(H) = 0,00768$ và hậu nghiệm: A: 0,5625; B: 0,25; C: 0,1875.

Ví dụ 17

Chuồng thứ nhất có 4 thỏ trắng, 3 thỏ đen; chuồng thứ hai có 5 thỏ trắng, 6 thỏ đen. Một con thỏ chạy từ chuồng một sang chuồng hai, sau đó một con chạy từ chuồng hai ra ngoài.
a) Tính xác suất con chạy ra ngoài là thỏ trắng. b) Biết con chạy ra ngoài là thỏ trắng, tính xác suất con chạy từ chuồng một sang là thỏ trắng.

$$\text{a) } \frac{4}{7} \cdot \frac{6}{12} + \frac{3}{7} \cdot \frac{5}{12} = \frac{39}{84} = \frac{13}{28}. \quad \text{b) } \frac{\frac{4}{7} \cdot \frac{6}{12}}{\frac{13}{28}} = \frac{8}{13} \approx 0,615 - \text{tăng so với tiên nghiệm } \frac{4}{7} \approx 0,571.$$

Dạng 8. Cập nhật rồi dự báo – Bayes nối tiếp toàn phần

Ví dụ 18

Lô hàng có 100 kiện: 60 kiện của xí nghiệp E (tỉ lệ phế phẩm 30%), 40 kiện của xí nghiệp F (tỉ lệ phế phẩm 10%). Mở ngẫu nhiên một kiện, lấy 1 sản phẩm thấy là phế phẩm; trả lại rồi lấy tiếp 1 sản phẩm từ cùng kiện. a) Tính xác suất sản phẩm thứ hai là phế phẩm. b) Biết sản phẩm thứ hai cũng là phế phẩm, tính xác suất kiện hàng của xí nghiệp E.

Bước 1 – cập nhật. $P(A) = 0,6 \cdot 0,3 + 0,4 \cdot 0,1 = 0,22$; $P(E | A) = \frac{0,18}{0,22} = \frac{9}{11}$,
 $P(F | A) = \frac{2}{11}$.

Bước 2 – dự báo. Dùng hậu nghiệm làm tiên nghiệm mới: $P = \frac{9}{11} \cdot 0,3 + \frac{2}{11} \cdot 0,1 = \frac{29}{110} \approx 0,264$ (lớn hơn 0,22 vì phế phẩm đầu tiên “tố cáo” kiện có nhiều khả năng của E). b) $\frac{\frac{27}{110}}{\frac{29}{110}} = \frac{27}{29} \approx 0,931$.

Ví dụ 19

Một hộp có 5 sản phẩm, hoàn toàn không rõ chất lượng (số sản phẩm tốt là 0, 1, ..., 5 với khả năng như nhau). Lấy ngẫu nhiên không hoàn lại 1 sản phẩm thì được sản phẩm tốt. Tính xác suất sản phẩm lấy lần thứ hai cũng tốt.

Gọi A_i : “hộp có i sản phẩm tốt”, $P(A_i) = \frac{1}{6}$. Lần 1 tốt: $P(B) = \frac{1}{6} \left(0 + \frac{1}{5} + \dots + 1 \right) = 0,5$, hậu nghiệm $P(A_i | B) = \frac{i}{15}$. Lần 2 tốt:

$$P(C) = \sum_{i=1}^5 \frac{i}{15} \cdot \frac{i-1}{4} = \frac{2}{3}$$

Ban đầu ta nghĩ mỗi sản phẩm “tốt hay xấu 50 – 50”; một sản phẩm tốt đã nâng dự báo cho sản phẩm tiếp theo lên $\frac{2}{3}$ – đây là tinh thần của “quy tắc kế tiếp” (rule of succession) mà Laplace đề xuất.

Ví dụ 20

Ba người thợ may cùng may một loại áo, xác suất mỗi áo đạt chất lượng cao lần lượt là 0,9; 0,9; 0,8. Một người (chưa biết là ai, ba khả năng như nhau) may 8 áo thì có 6 áo chất lượng cao. Tính xác suất người đó may thêm 8 áo nữa cũng có đúng 6 áo chất lượng cao.

Theo Bernoulli, $P(A | T_i) = C_8^6 p_i^6 (1 - p_i)^2$ bằng khoảng 0,1488; 0,1488; 0,2936. Hậu nghiệm: $P(T_1 | A) = P(T_2 | A) \approx 0,2517$, $P(T_3 | A) \approx 0,4966$ – người thợ thứ ba (tay nghề kém hơn) trở nên đáng ngờ nhất. Dự báo:

$$P(8 \text{ áo sau có đúng 6 áo tốt}) \approx 0,2517 \cdot 0,1488 \cdot 2 + 0,4966 \cdot 0,2936 \approx 0,221$$

Dạng 9. Các nghịch lí kinh điển

Ví dụ 21 (bài toán Monty Hall)

Ba cửa, một xe, hai dê. Người chơi chọn cửa 1; người dẫn chương trình (biết xe ở đâu) luôn mở một cửa có dê trong hai cửa còn lại – nếu cả hai đều có dê thì mở ngẫu nhiên – rồi mời đổi. Người chơi có nên đổi không?



Hình 10. Ba khả năng ban đầu như nhau; đổi cửa thắng ở hai trong ba khả năng

Gọi X_i : “xe sau cửa i ”, M_3 : “người dẫn mở cửa 3”. Ta có $P(M_3 | X_1) = \frac{1}{2}$, $P(M_3 | X_2) = 1$, $P(M_3 | X_3) = 0$. Theo Bayes:

$$P(X_2 | M_3) = \frac{\frac{1}{3} \cdot 1}{\frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0} = \frac{2}{3}$$

Vậy **nên đổi**. Mấu chốt là người dẫn *biết* và *bị ràng buộc* phải mở cửa có dê. Nếu người dẫn mở ngẫu nhiên (có thể lộ xe) mà tình cờ lộ dê, thì xác suất đổi cửa thắng chỉ là $\frac{1}{2}$ – cùng một quan sát nhưng **cơ chế sinh ra thông tin** khác nhau cho đáp số khác nhau.

Ví dụ 22 (nghịch lí ba chiếc hộp của Bertrand, 1889)

Có ba hộp giống nhau: hộp 1 có hai đồng vàng, hộp 2 có hai đồng bạc, hộp 3 có một đồng vàng, một đồng bạc. Chọn ngẫu nhiên một hộp, rút ngẫu nhiên một đồng thì được đồng vàng. Tính xác suất đồng còn lại trong hộp cũng là vàng.



Rút được đồng vàng: 3 đồng vàng đồng khả năng, 2 trong số đó nằm ở Hộp 1 $\Rightarrow P = 2/3$

Hình 11. Ba đồng vàng đồng khả năng được rút ra – hai trong số đó thuộc hộp 1

Trực giác nói $\frac{1}{2}$ (“chỉ còn hộp 1 hoặc hộp 3”). Nhưng $P(\text{vàng} | \text{hộp 1}) = 1$, $P(\text{vàng} | \text{hộp 3}) = \frac{1}{2}$, nên $P(\text{hộp 1} | \text{vàng}) = \frac{\frac{1}{3} \cdot 1}{\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{2}} = \frac{2}{3}$. Hộp 1 “dễ cho ra vàng” gấp đôi hộp 3, nên quan sát được vàng ủng hộ hộp 1 gấp đôi.

Dạng 10. Độc lập, Bernoulli và điều kiện “có đúng k”

Ví dụ 23

Ba xạ thủ bắn độc lập, mỗi người một viên, xác suất trúng lần lượt 0,7; 0,8; 0,9. Biết có đúng hai người bắn trúng, tính xác suất xạ thủ thứ nhất bắn trúng.

$P(\text{đúng 2}) = 0,7 \cdot 0,8 \cdot 0,1 + 0,7 \cdot 0,2 \cdot 0,9 + 0,3 \cdot 0,8 \cdot 0,9 = 0,056 + 0,126 + 0,216 = 0,398$. Trong đó hai kịch bản có người thứ nhất trúng: $0,056 + 0,126 = 0,182$. Vậy $P = \frac{0,182}{0,398} = \frac{91}{199} \approx 0,457$ – nhỏ hơn hẳn 0,7: biết “có một người trượt” làm tăng khả năng người trượt chính là người bắn kém nhất.

Ví dụ 24

Phân xưởng có 5 máy hoạt động độc lập, mỗi máy hỏng với xác suất 0,1. Biết có đúng 2 máy hỏng, tính xác suất máy số 1 không hỏng.

Bài giảng dùng công thức Bernoulli: $P = \frac{0,9 \cdot C_4^2 0,1^2 0,9^2}{C_5^2 0,1^2 0,9^3} = \frac{C_4^2}{C_5^2} = 0,6$. Nhận xét đẹp: khi các máy **như nhau**, biết có đúng 2 máy hỏng thì mọi cặp 2 máy đều đồng khả năng, nên đáp số $\frac{3}{5}$ **không phụ thuộc** vào 0,1 – xác suất có điều kiện đã “xóa” tham số.

Tổng kết các dạng toán

Dạng	Dấu hiệu nhận biết	Công cụ chính
1	Phép thử đồng khả năng, đề có “nếu / biết rằng”	Đếm lại trong không gian mẫu thu hẹp
2	Cho sẵn $P(A)$, $P(B)$, $P(AB)$ dạng tỉ lệ	Bảng xác suất 2×2 + định nghĩa
3	Lấy lần lượt, kết quả lần sau phụ thuộc lần trước	Công thức nhân
4	Nhiều người rút thăm không hoàn lại	Toàn phần hoặc lập luận đối xứng: m/n
5	Chuyển bi, chuyển vật giữa các hộp, chuồng, cốc	Hệ đầy đủ theo giai đoạn trước + toàn phần
6	Độ nhảy, độ đặc hiệu, dương tính	Cây tần số tự nhiên + Bayes
7	Nhiều nguồn, hỏi “thuộc nhóm nào / do ai”	Bayes với hệ n giả thuyết
8	Quan sát rồi dự báo lần tiếp theo	Bayes $\rightarrow$ hậu nghiệm $\rightarrow$ toàn phần
9	Monty Hall, ba chiếc hộp, đồng xu	Bayes, chú ý cơ chế sinh thông tin
10	“Biết có đúng k ...”	Bernoulli / độc lập + định nghĩa

IV. ỨNG DỤNG THỰC TẾ

Các tình huống dưới đây đến từ sản xuất, y tế, pháp luật, công nghệ và kinh doanh. Đề thi tốt nghiệp THPT, HSA và TSA rất ưa chuộng những bối cảnh như vậy vì chúng kiểm tra cả **khả năng đọc hiểu và mô hình hoá**, không chỉ kỹ năng tính toán.

Tình huống 1. Kiểm soát chất lượng

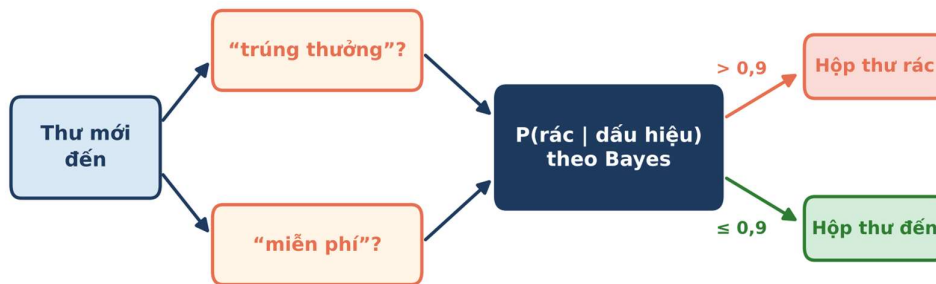
Một công ty sản xuất ba loại sản phẩm, chiếm 50%, 30%, 20% sản lượng, với tỉ lệ lỗi tương ứng 1%, 2%, 5% (số liệu minh hoạ trong bài giảng của PGS.TS Phan Huy Khải).

a) Chọn ngẫu nhiên một sản phẩm, tính xác suất bị lỗi. b) Một sản phẩm bị khách hàng trả lại vì lỗi; khả năng nó thuộc loại 3 là bao nhiêu?

a) $0,5 \cdot 0,01 + 0,3 \cdot 0,02 + 0,2 \cdot 0,05 = 0,005 + 0,006 + 0,010 = 0,021$. b) $\frac{0,010}{0,021} = \frac{10}{21} \approx 0,48$. Loại 3 chỉ chiếm 20% sản lượng nhưng gây ra gần một nửa số lỗi – người quản lí nên ưu tiên cải tiến dây chuyền loại 3.

Tình huống 2. Bộ lọc thư rác

Giả sử 40% thư gửi đến một hộp thư là thư rác. Cụm từ “trúng thưởng” có trong 50% thư rác nhưng chỉ 2% thư thường; cụm từ “miễn phí” có trong 30% thư rác và 5% thư thường (số liệu minh hoạ). Một thư mới chứa cả hai cụm từ. Bộ lọc nên xếp nó vào đâu?



Hình 12. Sơ đồ bộ lọc thư rác kiểu Bayes

Chỉ với “trúng thưởng”: $\frac{0,4 \cdot 0,5}{0,4 \cdot 0,5 + 0,6 \cdot 0,02} = \frac{50}{53} \approx 0,943$. Bộ lọc Bayes “ngây thơ” (naive Bayes) giả sử các từ xuất hiện độc lập khi đã biết loại thư, nên nhân thêm hệ số của “miễn phí”:

$$P(\text{rác} | \text{hai từ}) = \frac{0,4 \cdot 0,5 \cdot 0,3}{0,4 \cdot 0,5 \cdot 0,3 + 0,6 \cdot 0,02 \cdot 0,05} = \frac{100}{101} \approx 0,99$$

Mô hình đơn giản này, với hàng nghìn từ khoá và xác suất ước lượng từ dữ liệu thật, từng là nền tảng của nhiều bộ lọc thư rác đầu những năm 2000 – và vẫn là bài học nhập môn của học máy.

Tình huống 3. Ngụy biện của công tố viên – vụ án Sally Clark

Năm 1999, luật sư người Anh Sally Clark bị kết án vì hai con trai của bà lần lượt tử vong khi còn sơ sinh. Một nhân chứng chuyên gia cho rằng xác suất một gia đình như vậy có một trẻ đột tử là $\frac{1}{8\,543}$, nên xác suất hai trẻ cùng đột tử là $\left(\frac{1}{8\,543}\right)^2 \approx \frac{1}{73\,000\,000}$.

Lập luận này mắc hai lỗi mà Hội Thống kê Hoàng gia Anh đã chỉ ra trong một tuyên bố công khai năm 2001:

- **Giả định độc lập sai:** hai ca đột tử trong cùng một gia đình có thể cùng chịu yếu tố di truyền, môi trường – nên $P(ca\,2 \mid ca\,1)$ lớn hơn nhiều so với $\frac{1}{8\,543}$.
- **Ngụy biện công tố viên:** con số nhỏ đó (nếu đúng) là $P(bằng\,chứng \mid vô\,tội)$, trong khi điều toà cần là $P(vô\,tội \mid bằng\,chứng)$. Theo Bayes, phải so sánh với xác suất của giả thuyết thay thế (hai lần sát hại trẻ sơ sinh liên tiếp) – vốn cũng cực kì hiếm.

Bản án bị huỷ năm 2003. Vụ việc trở thành ví dụ kinh điển cho việc **nhầm chiều điều kiện** có thể gây hậu quả nghiêm trọng ra sao.

Tình huống 4. Nhân chứng và tỉ lệ nền (Tversky – Kahneman)

Một xe taxi gây tai nạn rồi bỏ chạy lúc đêm. Thành phố có hai hãng taxi: 85% xe màu xanh lá, 15% xe màu xanh dương. Nhân chứng khẳng định xe màu xanh dương. Kiểm tra trong điều kiện tương tự cho thấy nhân chứng nhận đúng màu 80% số lần. Khả năng chiếc xe thật sự màu xanh dương là bao nhiêu?

$$P(\text{dương} \mid \text{nói dương}) = \frac{0,15 \cdot 0,8}{0,15 \cdot 0,8 + 0,85 \cdot 0,2} = \frac{0,12}{0,29} \approx 0,41$$

Phần lớn người được hỏi trả lời 80% – bỏ qua tỉ lệ nền 15%. Bài toán do A. Tversky và D. Kahneman đưa ra (1972) là một trong những minh chứng nổi tiếng nhất cho “ngụy biện tỉ lệ nền”.

Tình huống 5. Truyền máu

Người nhóm máu AB nhận được máu của mọi nhóm; người nhóm O, A, B chỉ nhận được máu cùng nhóm hoặc nhóm O. Tỉ lệ các nhóm máu O, A, B, AB lần lượt là 33,7%; 37,5%; 20,9%; 7,9%. Chọn ngẫu nhiên một người cần máu. a) Có một người hiến máu ngẫu nhiên: tính xác suất truyền được. b) Có hai người hiến máu ngẫu nhiên.

a) $P = 0,337 \cdot 0,337 + 0,375 \cdot 0,712 + 0,209 \cdot 0,546 + 0,079 \cdot 1 \approx 0,574$.

b) Truyền không được khi **cả hai** người hiến đều không phù hợp – xác suất có điều kiện phải được **bình phương**:

$$P(\text{không được}) = 0,337 \cdot 0,663^2 + 0,375 \cdot 0,288^2 + 0,209 \cdot 0,454^2 \approx 0,222$$

Vậy $P(\text{truyền được}) \approx 0,778$. (Một lỗi dễ mắc là quên bình phương, dẫn tới đáp số sai 0,574 – giống hệt câu a.)

Tình huống 6. Bán vé vượt số ghế (overbooking)

Theo kinh nghiệm, 5% khách đặt vé sẽ huỷ chuyến. Hãng bán tối đa 52 vé cho chuyến bay 50 ghế; xác suất bán được 52 vé và 51 vé đều là 10%, còn lại bán không quá 50 vé. Tính xác suất mọi khách đến bay đều có ghế.

Thiếu ghế khi: bán 52 vé và có ít hơn 2 người huỷ, hoặc bán 51 vé và không ai huỷ. Theo công thức toàn phần kết hợp Bernoulli:

$$P(\text{thiếu ghế}) = 0,1 \cdot (0,95^{52} + 52 \cdot 0,95^{51} \cdot 0,05) + 0,1 \cdot 0,95^{51} \approx 0,033$$

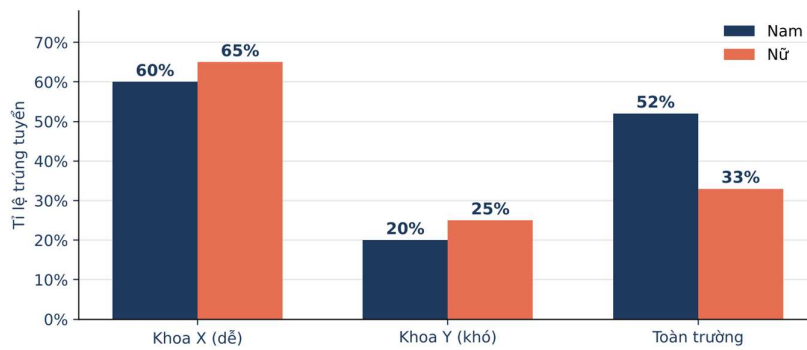
Vậy với xác suất khoảng 0,967 mọi hành khách đều có ghế. Các hãng hàng không cân nhắc con số này với lợi nhuận thu thêm từ ghế trống được bán.

Tình huống 7. Nghịch lý Simpson – tuyển sinh sau đại học ở Berkeley

Mùa thu năm 1973, Đại học California tại Berkeley nhận 8 442 hồ sơ nam (khoảng 44% trúng tuyển) và 4 321 hồ sơ nữ (khoảng 35% trúng tuyển). Nghiên cứu của P. Bickel, E. Hammel và J. O'Connell (*Science*, 1975) cho thấy khi xét **từng khoa**, không có bằng chứng phân biệt đối xử với nữ – ở nhiều khoa nữ còn có tỉ lệ đỗ cao hơn.

Ví dụ thu nhỏ (số liệu giả định) dưới đây cho thấy cơ chế: khoa X dễ, khoa Y khó; nam chủ yếu nộp khoa X (400 nam, 100 nữ), nữ chủ yếu nộp khoa Y (100 nam, 400 nữ).

	Khoa X: nộp / đỗ	Khoa Y: nộp / đỗ	Toàn trường
Nam	400 / 240 (60%)	100 / 20 (20%)	500 / 260 (52%)
Nữ	100 / 65 (65%)	400 / 100 (25%)	500 / 165 (33%)



Hình 13. Ở mỗi khoa nữ đông nhiều hơn, nhưng gộp chung thì nam đông nhiều hơn

Về xác suất: $P(\text{đỗ} | \text{nữ}) = \sum_k P(\text{khoa } k | \text{nữ}) \cdot P(\text{đỗ} | \text{nữ}, \text{khoa } k)$ – tỉ lệ chung là **trung bình có trọng số**, và trọng số $P(\text{khoa} | \text{giới tính})$ khác nhau giữa nam và nữ. Bài học: trước khi kết luận từ một tỉ lệ chung, hãy hỏi “điều kiện theo yếu tố nào?”.

Tình huống 8. Khảo sát thị trường

Phòng vấn 200 khách hàng về một sản phẩm mới: 34 người nói “sẽ mua”, 96 người nói “có thể mua”, 70 người nói “không mua”. Kinh nghiệm cho thấy tỉ lệ thực sự mua ở ba nhóm là 40%, 20%, 1% . a) Ước lượng thị phần tiềm năng. b) Trong số người thực sự mua, bao nhiêu phần trăm đã nói “sẽ mua”?

a) $0,17 \cdot 0,4 + 0,48 \cdot 0,2 + 0,35 \cdot 0,01 = 0,1675$. b) $\frac{0,068}{0,1675} \approx 0,406$ – hơn một nửa khách hàng thực sự đến từ nhóm “lưỡng lự”, nên chiến dịch tiếp thị nên nhắm vào nhóm này.

Tình huống 9. Bố trí phòng không – ra quyết định bằng xác suất toàn phần

Máy bay địch xuất hiện ở vị trí A với xác suất $\frac{2}{3}$, ở vị trí B với xác suất $\frac{1}{3}$. Có 4 khẩu pháo, mỗi khẩu bắn trúng với xác suất 0,7, độc lập. So sánh ba phương án: (3 tại A, 1 tại B), (2, 2), (1, 3) .

Phương án	Công thức	Xác suất hạ máy bay
3 – 1	$\frac{2}{3}(1 - 0,3^3) + \frac{1}{3} \cdot 0,7$	0,882
2 – 2	$\frac{2}{3}(1 - 0,3^2) + \frac{1}{3}(1 - 0,3^2)$	0,910
1 – 3	$\frac{2}{3} \cdot 0,7 + \frac{1}{3}(1 - 0,3^3)$	0,791

Phương án chia đều 2 – 2 tốt nhất dù vị trí A khả năng cao gấp đôi: khẩu pháo thứ ba ở A chỉ tăng thêm $0,09 - 0,027 = 0,063$ xác suất, trong khi khẩu thứ hai ở B tăng thêm 0,21. Đây là ý tưởng “lợi ích cận biên giảm dần” trong kinh tế học.

Tình huống 10. Alan Turing và cỗ máy Enigma

Không cần tính toán, chỉ cần một câu chuyện: mỗi bản tin bị chặn được, Turing và đồng nghiệp không hỏi “thiết lập nào *chắc chắn* đúng?” mà hỏi “bằng chứng này làm thiết lập nào *đáng tin hơn bao nhiêu?*” – tức là nhân tỉ số khả năng với hệ số khả dĩ, đúng như cách ta làm ở Ví dụ 13. Bằng chứng yếu, cộng dồn nhiều lần, trở thành bằng chứng mạnh.

V. BÀI TẬP LUYỆN TẬP THEO ĐỊNH DẠNG ĐỀ THI

Kí hiệu: **TN** trắc nghiệm 4 lựa chọn; **TLN** trả lời ngắn / điền đáp án; **Đ/S** Đúng/Sai (4 ý).
 Độ khó: ★☆☆ cơ bản, ★★☆☆ vận dụng, ★★★ vận dụng cao. Thời gian gợi ý: 2 phút cho câu TN, TLN; 5 phút cho câu Đ/S.

A. Câu hỏi trắc nghiệm 4 lựa chọn

Bài 1. **TN** ★☆☆ Gieo hai con xúc xắc cân đối. Biết tổng số chấm bằng 8, xác suất có một con ra 6 chấm là

- A. $\frac{2}{5}$ B. $\frac{1}{3}$ C. $\frac{1}{18}$ D. $\frac{5}{36}$

Bài 2. **TN** ★☆☆ Một gia đình có hai con, xác suất sinh trai hay gái như nhau. Biết con đầu lòng là con trai, xác suất cả hai đều là con trai là

- A. $\frac{1}{4}$ B. $\frac{1}{3}$ C. $\frac{1}{2}$ D. $\frac{3}{4}$

Bài 3. **TN** ★☆☆ Cho hai biến cố A, B với $P(A) = 0,4$, $P(B) = 0,5$, $P(A \cup B) = 0,7$. Giá trị của $P(A | \overline{B})$ bằng

- A. 0,2 B. 0,4 C. 0,5 D. 0,8

Bài 4. **TN** ★☆☆ Lớp 12A có 40 học sinh gồm 25 nữ và 15 nam; trong đó có 12 bạn nữ và 8 bạn nam đeo kính. Chọn ngẫu nhiên một bạn thì thấy bạn đó đeo kính. Xác suất bạn đó là nữ là

- A. 0,30 B. 0,48 C. 0,60 D. 0,50

Bài 5. **TN** ★☆☆ Một hộp có 5 bi đỏ và 3 bi xanh. Lấy lần lượt không hoàn lại 2 bi. Xác suất bi lấy ra ở lần thứ hai màu đỏ là

- A. $\frac{4}{7}$ B. $\frac{5}{8}$ C. $\frac{5}{14}$ D. $\frac{1}{2}$

Bài 6. **TN** ★☆☆ Một bệnh có tỉ lệ mắc 1%. Xét nghiệm có độ nhạy 90% và tỉ lệ dương tính giả 9%. Một người có kết quả dương tính. Xác suất người đó mắc bệnh xấp xỉ

- A. 90% B. 50% C. 9% D. 1%

Bài 7. TN ★★★ Phân xưởng I sản xuất 60%, phân xưởng II sản xuất 40% sản phẩm của nhà máy, tỉ lệ phế phẩm tương ứng là 2% và 5%. Chọn ngẫu nhiên một sản phẩm thấy là phế phẩm. Xác suất nó do phân xưởng II sản xuất là

- A. 0,375 B. 0,4 C. 0,625 D. 0,05

Bài 8. TN ★★★ Trò chơi Monty Hall với 4 cửa (1 xe, 3 dê). Người chơi chọn một cửa; người dẫn (biết xe ở đâu) mở một cửa có dê trong ba cửa còn lại. Người chơi đổi sang ngẫu nhiên một trong hai cửa chưa mở còn lại. Xác suất thắng xe là

- A. $\frac{1}{4}$ B. $\frac{3}{8}$ C. $\frac{1}{2}$ D. $\frac{3}{4}$

Bài 9. TN ★★★ Có ba tấm thẻ: một tấm hai mặt đỏ, một tấm hai mặt trắng, một tấm một mặt đỏ một mặt trắng. Rút ngẫu nhiên một tấm, đặt ngẫu nhiên lên bàn thì thấy mặt trên màu đỏ. Xác suất mặt dưới cũng màu đỏ là

- A. $\frac{1}{3}$ B. $\frac{1}{2}$ C. $\frac{2}{3}$ D. $\frac{3}{4}$

Bài 10. TN ★★★ Ba xạ thủ bắn độc lập vào một bia, xác suất trúng lần lượt 0,6; 0,7; 0,8. Biết có đúng một người bắn trúng. Xác suất người đó là xạ thủ thứ nhất là

- A. 0,6 B. $\frac{9}{47}$ C. 0,036 D. $\frac{1}{3}$

Bài 11. TN ★★★ Một học sinh thi hai môn. Xác suất đạt môn thứ nhất là 0,7; nếu đạt môn thứ nhất thì xác suất đạt môn thứ hai là 0,8, nếu không đạt môn thứ nhất thì xác suất đạt môn thứ hai là 0,4. Biết học sinh đạt đúng một môn. Xác suất môn đó là môn thứ hai là

- A. $\frac{6}{13}$ B. $\frac{7}{13}$ C. 0,12 D. 0,4

Bài 12. TN ★★★ Một hộp có 10 lá thăm, trong đó 3 lá trúng thưởng. Ba người lần lượt rút, mỗi người một lá, không hoàn lại. Xác suất người thứ ba rút được thăm trúng thưởng là

- A. $\frac{1}{10}$ B. $\frac{1}{8}$ C. $\frac{3}{10}$ D. $\frac{1}{3}$

B. Câu hỏi trả lời ngắn

Bài 13. TLN ★★★ Gieo ba con xúc xắc cân đối. Biết có ít nhất một con ra 6 chấm, tính xác suất tổng số chấm không nhỏ hơn 15 (làm tròn đến hàng phần trăm).

Bài 14. TLN ★☆☆ Hộp I có 4 bi đỏ, 6 bi trắng; hộp II có 5 bi đỏ, 5 bi trắng. Lấy ngẫu nhiên 1 bi từ hộp I bỏ vào hộp II, rồi lấy ngẫu nhiên 1 bi từ hộp II. Tính xác suất bi lấy ra màu đỏ (làm tròn đến hàng phần trăm).

Bài 15. TLN ★☆☆ 30% số thư đến một hộp thư là thư rác. Từ “khuyến mãi” xuất hiện trong 40% thư rác và 5% thư thường. Một thư chứa từ “khuyến mãi”; tính xác suất đó là thư rác (làm tròn đến hàng phần trăm).

Bài 16. TLN ★☆☆ Tỷ lệ mắc bệnh B là 2%. Một xét nghiệm có độ nhạy 95%, tỷ lệ dương tính giả 10%. Một người làm xét nghiệm này hai lần độc lập (khi đã biết tình trạng bệnh) và đều dương tính. Tính xác suất người đó mắc bệnh (làm tròn đến hàng phần trăm).

Bài 17. TLN ★☆☆ Lô 1 có 9 sản phẩm tốt, 1 sản phẩm xấu; lô 2 có 6 sản phẩm tốt, 4 sản phẩm xấu. Chọn ngẫu nhiên một lô, lấy ngẫu nhiên 1 sản phẩm thấy tốt, trả lại vào lô. Lấy tiếp ngẫu nhiên 1 sản phẩm từ cùng lô đó. Tính xác suất sản phẩm này xấu (làm tròn đến hàng phần trăm).

Bài 18. TLN ★☆☆ Một thành phố có 70% taxi màu xanh, 30% màu vàng. Một nhân chứng khẳng định chiếc taxi gây tai nạn màu vàng; nhân chứng nhận đúng màu 75% số lần. Tính xác suất chiếc taxi thật sự màu vàng (làm tròn đến hàng phần trăm).

Bài 19. TLN ★☆☆ Chia ngẫu nhiên 12 học sinh, trong đó có 4 học sinh giỏi, thành 4 nhóm, mỗi nhóm 3 người. Tính xác suất mỗi nhóm có đúng một học sinh giỏi (làm tròn đến hàng phần trăm).

Bài 20. TLN ★☆☆ Khảo sát 200 học sinh: 80 em dùng mạng xã hội trên 3 giờ mỗi ngày, trong đó 30 em có kết quả học tập yếu; 120 em còn lại có 18 em kết quả yếu. Chọn ngẫu nhiên một học sinh có kết quả yếu. Tính xác suất em đó dùng mạng xã hội trên 3 giờ mỗi ngày (làm tròn đến hàng phần trăm).

Bài 21. TLN ★☆☆ Cho $P(A) = 0,5$, $P(B | A) = 0,6$, $P(B | \bar{A}) = 0,2$. Tính $P(A | B)$.

Bài 22. TLN ★☆☆ Một cầu thủ sút phạt lần đầu thành công với xác suất 0,7. Ở các lần sau, nếu lần trước thành công thì xác suất thành công là 0,8, nếu lần trước hỏng thì là 0,5. Cầu thủ sút 3 quả. Tính xác suất ghi đúng 2 bàn (làm tròn đến hàng phần trăm).

C. Câu hỏi Đúng/Sai

Bài 23. **Đ/S** ★★★ Khảo sát 300 học sinh: 180 em ăn sáng đều đặn, trong đó 126 em có kết quả học tập khá, giỏi; 120 em không ăn sáng đều đặn, trong đó 54 em khá, giỏi. Chọn ngẫu nhiên một học sinh.

Mệnh đề	Đúng	Sai
a) Xác suất học sinh đó khá, giỏi là 0,6.	☐	☐
b) Xác suất học sinh đó khá, giỏi, biết em ăn sáng đều đặn, là 0,7.	☐	☐
c) Xác suất học sinh đó ăn sáng đều đặn, biết em khá, giỏi, là 0,6.	☐	☐
d) Hai biến cố “ăn sáng đều đặn” và “khá, giỏi” độc lập với nhau.	☐	☐

Bài 24. **Đ/S** ★★★ Tỷ lệ mắc bệnh B trong cộng đồng là 5%. Xét nghiệm X có độ nhạy 90% và độ đặc hiệu 95%.

Mệnh đề	Đúng	Sai
a) Xác suất một người ngẫu nhiên có kết quả dương tính là 0,0925.	☐	☐
b) Một người dương tính thì khả năng mắc bệnh lớn hơn 50%.	☐	☐
c) Một người âm tính thì khả năng không mắc bệnh lớn hơn 99%.	☐	☐
d) Nếu làm xét nghiệm X hai lần độc lập (khi đã biết tình trạng bệnh) và đều dương tính thì khả năng mắc bệnh lớn hơn 90%.	☐	☐

Bài 25. **Đ/S** ★★★ Máy A sản xuất 60%, máy B sản xuất 40% sản phẩm; tỉ lệ phế phẩm lần lượt là 3% và 2%. Chọn ngẫu nhiên một sản phẩm.

Mệnh đề	Đúng	Sai
a) Xác suất sản phẩm là phế phẩm bằng 0,026.	☐	☐
b) Biết sản phẩm là phế phẩm, xác suất nó do máy A sản xuất bằng $\frac{9}{13}$.	☐	☐
c) Biết sản phẩm là phế phẩm, xác suất nó do máy B sản xuất bằng 0,4.	☐	☐
d) Biết sản phẩm đạt chuẩn, xác suất nó do máy A sản xuất lớn hơn 0,6.	☐	☐

Bài 26. **Đ/S** ★★★ Trò chơi Monty Hall chuẩn: ba cửa, một xe; người chơi chọn cửa 1, người dẫn (biết xe ở đâu) luôn mở một cửa có dê trong hai cửa còn lại.

Mệnh đề	Đúng	Sai
a) Trước khi người dẫn mở cửa, xác suất xe ở sau cửa 1 là $\frac{1}{3}$.	☐	☐
b) Sau khi người dẫn mở một cửa có dê, xác suất xe ở sau cửa 1 tăng lên $\frac{1}{2}$.	☐	☐
c) Nếu đổi cửa, xác suất thắng là $\frac{2}{3}$.	☐	☐

Mệnh đề	Đúng	Sai
d) Nếu người dẫn <i>không biết</i> xe ở đâu, mở ngẫu nhiên một trong hai cửa còn lại và tình cờ thấy dê, thì đôi cửa thắng với xác suất $\frac{1}{2}$.	☐	☐

Bài 27. **Đ/S** ★☆☆ Một công ty dự thầu hai dự án. Xác suất thắng dự án 1 là 0,5; dự án 2 là 0,6; cả hai dự án là 0,35.

Mệnh đề	Đúng	Sai
a) Biết công ty thắng dự án 1, xác suất thắng dự án 2 là 0,7.	☐	☐
b) Xác suất công ty chỉ thắng đúng một dự án là 0,4.	☐	☐
c) Biết công ty trượt dự án 1, xác suất thắng dự án 2 là 0,3.	☐	☐
d) Việc thắng hai dự án là hai biến cố độc lập.	☐	☐

Bài 28. **Đ/S** ★☆☆ Một cuộc thi có 3 vòng. Vòng 1 lấy 80% thí sinh; vòng 2 lấy 75% số thí sinh qua vòng 1; vòng 3 lấy 90% số thí sinh qua vòng 2. Chọn ngẫu nhiên một thí sinh.

Mệnh đề	Đúng	Sai
a) Xác suất thí sinh vượt qua cả 3 vòng là 0,54.	☐	☐
b) Xác suất thí sinh bị loại ngay ở vòng 1 là 0,2.	☐	☐
c) Biết thí sinh bị loại, xác suất em bị loại ở vòng 2 xấp xỉ 0,43.	☐	☐
d) Biết thí sinh bị loại, khả năng em bị loại ở vòng 3 là cao nhất.	☐	☐

Bài 29. **Đ/S** ★☆☆ Ở một vùng, nếu hôm nay nắng thì ngày mai nắng với xác suất 0,8, mưa với xác suất 0,2; nếu hôm nay mưa thì ngày mai nắng hay mưa đều với xác suất 0,5. Hôm nay trời nắng.

Mệnh đề	Đúng	Sai
a) Xác suất ngày mai trời mưa là 0,2.	☐	☐
b) Xác suất ngày kia trời mưa là 0,26.	☐	☐
c) Biết ngày kia trời mưa, xác suất ngày mai trời mưa lớn hơn 0,5.	☐	☐
d) Xác suất ba ngày tới đều nắng là 0,64.	☐	☐

Bài 30. **Đ/S** ★☆☆ Hộp I có 3 bi đỏ, 2 bi xanh; hộp II có 1 bi đỏ, 4 bi xanh. Gieo một con xúc xắc cân đối: nếu ra mặt 6 thì lấy ngẫu nhiên 1 bi từ hộp I, ngược lại lấy ngẫu nhiên 1 bi từ hộp II.

Mệnh đề	Đúng	Sai
a) Xác suất lấy bi từ hộp I là $\frac{1}{6}$.	☐	☐

Mệnh đề	Đúng	Sai
b) Xác suất lấy được bi đỏ là $\frac{4}{15}$.	☐	☐
c) Biết lấy được bi đỏ, xác suất bi đỏ từ hộp I là $\frac{1}{2}$.	☐	☐
d) Biết lấy được bi xanh, xác suất bi đỏ từ hộp II lớn hơn 0,9.	☐	☐

VI. ĐÁP ÁN VÀ HƯỚNG DẪN GIẢI

Bảng đáp án

Bài	1	2	3	4	5	6	7	8	9	10	11	12
Đáp án	A	C	B	C	B	C	C	B	C	B	A	C

Bài	13	14	15	16	17	18	19	20	21	22
Đáp án	0,21	0,49	0,77	0,65	0,22	0,56	0,16	0,63	0,75	0,30

Bài	a)	b)	c)	d)
23	Đ	Đ	S	S
24	Đ	S	Đ	Đ
25	Đ	Đ	S	S
26	Đ	S	Đ	Đ
27	Đ	Đ	S	S
28	Đ	Đ	Đ	S
29	Đ	Đ	S	S
30	Đ	Đ	S	Đ

Hướng dẫn giải

Bài 1. Tổng bằng 8 có 5 kết quả (2; 6), (3; 5), (4; 4), (5; 3), (6; 2), trong đó 2 kết quả có số 6: $\frac{2}{5}$. (Phương án D là $P(AB)$ – nhầm “nếu” thành “và”).

Bài 2. Không gian mẫu thu hẹp $\{TT, TG\}$: $\frac{1}{2}$. Phương án B ($\frac{1}{3}$) ứng với điều kiện “có ít nhất một con trai”.

Bài 3. $P(AB) = 0,4 + 0,5 - 0,7 = 0,2$; $P(A | \overline{B}) = \frac{0,4 - 0,2}{1 - 0,5} = 0,4$. (Ở đây A, B độc lập nên $P(A | \overline{B}) = P(A)$.)

Bài 4. 20 bạn đeo kính, 12 bạn nữ: $\frac{12}{20} = 0,6$.

Bài 5. Dạng 4 – rút lần lượt công bằng: $\frac{5}{8}$. Kiểm tra: $\frac{5}{8} \cdot \frac{4}{7} + \frac{3}{8} \cdot \frac{5}{7} = \frac{35}{56} = \frac{5}{8}$.

Bài 6. $\frac{0,01 \cdot 0,9}{0,01 \cdot 0,9 + 0,99 \cdot 0,09} = \frac{10}{109} \approx 9\%$. Tỷ lệ nền 1% lần át độ chính xác của xét nghiệm.

Bài 7. $P(lỗi) = 0,012 + 0,02 = 0,032$; $\frac{0,02}{0,032} = 0,625$.

Bài 8. Cửa ban đầu có xe với xác suất $\frac{1}{4}$; khi đó đổi chắc chắn thua. Với xác suất $\frac{3}{4}$ xe ở một trong ba cửa kia; người dẫn loại một cửa dê, xe nằm ở một trong hai cửa còn lại và người chơi chọn trúng với xác suất $\frac{1}{2} \cdot \frac{3}{4} \cdot \frac{1}{2} = \frac{3}{8}$ (vẫn lớn hơn $\frac{1}{4}$ nếu giữ nguyên).

Bài 9. Có 3 mặt đỏ đồng khả năng nằm trên, 2 trong số đó thuộc tám hai mặt đỏ: $\frac{2}{3}$ (nghịch lí Bertrand dạng thẻ).

Bài 10. $P(\text{đúng 1}) = 0,6 \cdot 0,3 \cdot 0,2 + 0,4 \cdot 0,7 \cdot 0,2 + 0,4 \cdot 0,3 \cdot 0,8 = 0,036 + 0,056 + 0,096 = 0,188$; $\frac{0,036}{0,188} = \frac{9}{47}$.

Bài 11. Chỉ đạt môn 1: $0,7 \cdot 0,2 = 0,14$; chỉ đạt môn 2: $0,3 \cdot 0,4 = 0,12$. $\frac{0,12}{0,26} = \frac{6}{13}$.

Bài 12. Dạng 4: mọi người có cùng xác suất $\frac{3}{10}$.

Bài 13. Có $216 - 125 = 91$ kết quả có ít nhất một con 6. Các bộ có số 6 và tổng ≥ 15 : hoán vị của $(6; 6; 6) - 1$, $(6; 6; 5) - 3$, $(6; 6; 4) - 3$, $(6; 6; 3) - 3$, $(6; 5; 5) - 3$, $(6; 5; 4) - 6$; tổng 19. Đáp số $\frac{19}{91} \approx 0,21$.

Bài 14. $0,4 \cdot \frac{6}{11} + 0,6 \cdot \frac{5}{11} = \frac{27}{55} \approx 0,49$.

Bài 15. $\frac{0,3 \cdot 0,4}{0,3 \cdot 0,4 + 0,7 \cdot 0,05} = \frac{0,12}{0,155} \approx 0,77$.

Bài 16. Dạng tỉ số: $\frac{0,02}{0,98} \cdot \left(\frac{0,95}{0,1}\right)^2 = \frac{2}{98} \cdot 90,25 \approx 1,842$, nên $P \approx \frac{1,842}{2,842} \approx 0,65$. (Sau một lần dương tính chỉ khoảng 0,16 – Ví dụ 12.)

Bài 17. $P(\text{tốt}) = 0,5 \cdot 0,9 + 0,5 \cdot 0,6 = 0,75$; hậu nghiệm lô 1: $\frac{0,45}{0,75} = 0,6$, lô 2: 0,4. Lần sau xấu: $0,6 \cdot 0,1 + 0,4 \cdot 0,4 = 0,22$.

Bài 18. $\frac{0,3 \cdot 0,75}{0,3 \cdot 0,75 + 0,7 \cdot 0,25} = \frac{0,225}{0,4} = 0,5625 \approx 0,56$.

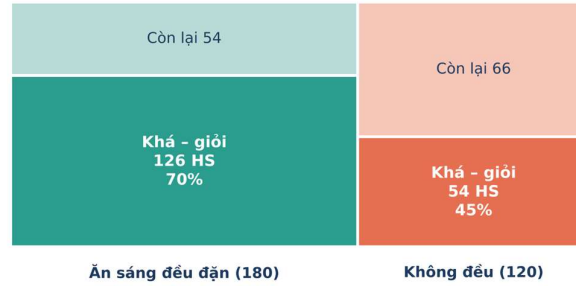
Bài 19. $\frac{C_4^1 C_8^2}{C_{12}^3} \cdot \frac{C_3^1 C_6^2}{C_9^3} \cdot \frac{C_2^1 C_4^2}{C_6^3} \cdot 1 = \frac{112}{220} \cdot \frac{45}{84} \cdot \frac{12}{20} = \frac{9}{55} \approx 0,16$.

Bài 20. 48 học sinh yếu, 30 em dùng trên 3 giờ: $\frac{30}{48} = 0,625 \approx 0,63$. (Chú ý chiều điều kiện: $P(\text{yếu} \mid \text{trên 3 giờ}) = \frac{30}{80} = 0,375$ là câu hỏi khác.)

Bài 21. $\frac{0,5 \cdot 0,6}{0,5 \cdot 0,6 + 0,5 \cdot 0,2} = \frac{0,3}{0,4} = 0,75$.

Bài 22. GGH: $0,7 \cdot 0,8 \cdot 0,2 = 0,112$; GHG: $0,7 \cdot 0,2 \cdot 0,5 = 0,07$; HGG: $0,3 \cdot 0,5 \cdot 0,8 = 0,12$. Tổng $0,302 \approx 0,30$.

Bài 23. a) $\frac{126+54}{300} = 0,6$ – Đ. b) $\frac{126}{180} = 0,7$ – Đ. c) $\frac{126}{180} = 0,7 \neq 0,6$ – S. d) $P(KG | \text{ăn sáng}) = 0,7 \neq 0,6 = P(KG)$ – không độc lập – S.



Hình 14. Biểu đồ diện tích cho Bài 23: chiều rộng là tỉ lệ nhóm, chiều cao là tỉ lệ khá – giỏi trong nhóm

Bài 24. a) $0,05 \cdot 0,9 + 0,95 \cdot 0,05 = 0,0925$ – Đ. b) $\frac{0,045}{0,0925} = \frac{18}{37} \approx 0,486 < 0,5$ – S. c) $\frac{0,95 \cdot 0,95}{1 - 0,0925} = \frac{361}{363} \approx 0,9945$ – Đ. d) Hệ số khả dĩ $\frac{0,9}{0,05} = 18$, tỉ số sau hai lần $\frac{5}{95} \cdot 18^2 \approx 17,05$, xác suất $\frac{324}{343} \approx 0,945$ – Đ.

Bài 25. a) $0,018 + 0,008 = 0,026$ – Đ. b) $\frac{0,018}{0,026} = \frac{9}{13}$ – Đ. c) $\frac{4}{13} \approx 0,31$ – S. d) $\frac{0,6 \cdot 0,97}{0,974} = \frac{291}{487} \approx 0,598 < 0,6$ – S.

Bài 26. a) Đ. b) Vẫn là $\frac{1}{3}$ – S. c) Đ (Ví dụ 21). d) Người dẫn không biết: $P(\text{lộ dề}) = \frac{2}{3}$; $P(\text{xe ở cửa 1 và lộ dề}) = \frac{1}{3}$, nên xe ở cửa 1 với xác suất $\frac{1}{2}$ và đổi cửa cũng thắng với xác suất $\frac{1}{2}$ – Đ.

Bài 27. a) $\frac{0,35}{0,5} = 0,7$ – Đ. b) $0,5 + 0,6 - 2 \cdot 0,35 = 0,4$ – Đ. c) $\frac{0,6 - 0,35}{0,5} = 0,5$ – S. d) $0,5 \cdot 0,6 = 0,3 \neq 0,35$ – S.

Bài 28. a) $0,8 \cdot 0,75 \cdot 0,9 = 0,54$ – Đ. b) Đ. c) Bị loại: 0,46; loại ở vòng 2: $0,8 \cdot 0,25 = 0,2$, $\frac{0,2}{0,46} = \frac{10}{23} \approx 0,43$ – Đ. d) Loại ở vòng 1: 0,2; vòng 2: 0,2; vòng 3: $0,8 \cdot 0,75 \cdot 0,1 = 0,06$ – thấp nhất – S.

Bài 29. a) Đ. b) $0,8 \cdot 0,2 + 0,2 \cdot 0,5 = 0,26$ – Đ. c) $\frac{0,2 \cdot 0,5}{0,26} = \frac{5}{13} \approx 0,385 < 0,5$ – S. d) $0,8^3 = 0,512$ – S.

Bài 30. a) Đ. b) $\frac{1}{6} \cdot \frac{3}{5} + \frac{5}{6} \cdot \frac{1}{5} = \frac{8}{30} = \frac{4}{15}$ – Đ.

c) $\frac{\frac{3}{30}}{\frac{8}{30}} = \frac{3}{8}$ – S. d) $\frac{\frac{5}{6} \cdot \frac{4}{5}}{\frac{11}{15}} = \frac{10}{11} \approx 0,909 > 0,9$ – Đ.

VII. KẾT LUẬN VÀ LỜI KHUYÊN KHI LÀM BÀI

- **Đặt tên biến cố trước, tính sau.** Viết ra giấy: A là gì, B là gì, đề cho $P(A | B)$ hay $P(B | A)$. Phần lớn lỗi sai không nằm ở phép tính mà ở việc dịch đề.
- **Nghe kĩ chữ “nếu”, “biết rằng”, “trong số”.** Chúng đánh dấu điều kiện. Chữ “và” thì không.
- **Đổi ra tần số tự nhiên khi bối cảnh y tế, sản xuất.** “Trong 1 000 người...” giúp thấy ngay tỉ lệ nền.
- **Độ đặc hiệu $\neq$ tỉ lệ dương tính giả.** Tỉ lệ dương tính giả bằng 1 trừ độ đặc hiệu.
- **Có hoàn lại hay không? Độc lập hay không?** Chỉ nhân các xác suất khi đã chắc chắn độc lập (vụ Sally Clark).
- **Hậu nghiệm hôm nay là tiên nghiệm ngày mai.** Thông tin đến tuần tự thì cập nhật tuần tự (Dạng 8, xét nghiệm hai lần).
- **Kiểm tra nhanh:** các hậu nghiệm của một hệ đầy đủ cộng lại bằng 1; $P(A | B)$ phải nằm giữa 0 và 1; tử số Bayes luôn là một số hạng của mẫu số.

Từ bài luận của một mục sư người Anh năm 1763, qua những lá thư phản đối gửi tạp chí *Parade* năm 1990, đến các câu hỏi Đúng/Sai trong đề thi hôm nay, bài học vẫn là một: *hãy để dữ kiện thay đổi niềm tin của mình – nhưng thay đổi đúng mức, không hơn không kém*. Đó là tư duy mà mọi công dân của thời đại dữ liệu đều cần.

VIII. TÀI LIỆU THAM KHẢO

Tài liệu trong nước

- [1] Phan Huy Khải, (2024). *Bài giảng: Các bài toán cơ bản của xác suất có điều kiện*.
- [2] Bộ Giáo dục và Đào tạo (2018). *Chương trình giáo dục phổ thông môn Toán*.
- [3] Hà Huy Khoái (Tổng chủ biên) và nhóm tác giả. *Toán 12, tập hai* (bộ sách Kết nối tri thức với cuộc sống), NXB Giáo dục Việt Nam – Chương VI, Bài 18, Bài 19.
- [4] Báo Đại biểu Nhân dân (2024). *Đề tham khảo môn Toán tốt nghiệp THPT 2025: Đảm bảo phân hoá rõ năng lực thí sinh*. daibieunhandan.vn.
- [5] Viện Đào tạo số và Khảo thí, ĐHQG Hà Nội (2026). *Dạng thức và đề cương chi tiết bài thi HSA*. hsa.edu.vn.
- [6] Đại học Bách khoa Hà Nội (2026). *Thông tin kì thi Đánh giá tư duy TSA 2026*. tsa.hust.edu.vn.

Giáo trình nước ngoài

- [7] Ross, S. M. *A First Course in Probability*, 10th ed. Pearson, 2019 (Chương 3: Conditional Probability and Independence).

- [8] Blitzstein, J. K. & Hwang, J. *Introduction to Probability*, 2nd ed. CRC Press, 2019 (Chương 2; bài toán Monty Hall, nghịch lý Simpson).
- [9] Grinstead, C. M. & Snell, J. L. *Introduction to Probability*. American Mathematical Society, 1997.
- [10] Feller, W. *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed. Wiley, 1968.
- [11] Gigerenzer, G. *Calculated Risks: How to Know When Numbers Deceive You*. Simon & Schuster, 2002.

Bài báo và tư liệu lịch sử

- [12] Bayes, T. & Price, R. (1763). *An Essay towards solving a Problem in the Doctrine of Chances*. Philosophical Transactions of the Royal Society of London, 53, 370 – 418.
- [13] Laplace, P.-S. (1812). *Théorie analytique des probabilités*. Paris.
- [14] Bertrand, J. (1889). *Calcul des probabilités*. Paris: Gauthier-Villars.
- [15] Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer.
- [16] Selvin, S. (1975). Letter to the editor (“A problem in probability”). *The American Statistician*, 29(1).
- [17] vos Savant, M. (1990). Ask Marilyn. *Parade Magazine*, 9/9/1990.
- [18] Casscells, W., Schoenberger, A. & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, 299, 999 – 1001.
- [19] Manrai, A. K. và cộng sự (2014). Medicine’s uncomfortable relationship with math. *JAMA Internal Medicine*, 174(6).
- [20] Bickel, P. J., Hammel, E. A. & O’Connell, J. W. (1975). Sex bias in graduate admissions: Data from Berkeley. *Science*, 187(4175), 398 – 404.
- [21] Tversky, A. & Kahneman, D. (1982). Evidential impact of base rates. Trong *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press.
- [22] Royal Statistical Society (2001, 2002). Statement và thư gửi Lord Chancellor về vụ án R v. Sally Clark.
- [23] McGrayne, S. B. (2011). *The Theory That Would Not Die*. Yale University Press.