Monte Hall Problem: In the game show Let's Make a Deal, there are 3 doors — behind one door is a car and behind the other two are goats. You (the contestant) are asked to pick a door. After you choose, Monte Hall, who knows where the car is, opens one of the other doors to reveal a goat, and he offers you the choice to switch from your original choice to the other closed door. Should you switch or stay with your choice?

Quick History

Bayes 1760 – A village had some people dying (data) and they needed to know what caused it. Possible causes were bad air, bad food, or diseases.

Switched from deterministic causality to probabilistic

$P(\text{effect} \mid \text{cause}) = \text{easy}$ $P(\overset{\text{unknown}}{\text{cause}} \mid \overset{\text{data}}{\text{effect}}) = \text{hard}$

$$P(U \mid D) = \frac{P(U \text{ and } D)}{P(D)} \rightarrow P(U \text{ and } D) = P(D)\,P(U \mid D)$$

$$P(D \mid U) = \frac{P(D \text{ and } U)}{P(U)} \rightarrow P(D \text{ and } U) = P(U)\,P(D \mid U)$$

Therefore $P(U)\,P(D \mid U) = P(D)\,P(U \mid D)$

$$P(U \mid D) = \frac{P(D)(U \mid D)}{P(U)}$$

$$P(\text{unknown} \mid \text{data}) = \frac{P(\text{unknown}) \cdot P(\text{data} \mid \text{unknown})}{P(\text{data})}$$

Bayes Theorem for T/F propositions

The denominator is often hard to compute.

$$P(U \mid D) = \frac{P(U)\,P(D \mid U)}{P(D)}$$

Suppose T/F

$$P(\text{not } U \mid D) = \frac{P(\text{not } U)\,P(D \mid \text{not } U)}{P(D)}$$

$$\frac{P(U \mid D)}{P(\text{not } U \mid D)} = \left[\frac{P(U)}{P(\text{not } U)}\right] \cdot \left[\frac{P(D \mid U)}{P(D \mid \text{not } U)}\right]$$

$P(H) = p \qquad P(T) = 1 - p = P(\text{not } H)$

$\dfrac{P(H)}{P(\text{not } H)} = \dfrac{p}{1-p}$ = odds ratio in favor of H

= odds ratio against T

Odds ratio against H = $\dfrac{1-p}{p}$

a priori — data — a posterior → time

before info — (new) info arrives — after info

$$\left[\frac{P(U \mid D)}{P(\text{not } U \mid D)}\right] = \left[\frac{P(D)}{P(\text{not } U)}\right] \cdot \left[\frac{P(D \mid U)}{P(D \mid \text{not } U)}\right]$$

(posterior odds ratio given D) = (prior odds ratio in favor of U) $\cdot$ (Bayes' Factor)

also called likelihood ratio

$$O = \frac{p}{1-p} \longleftrightarrow p = \frac{O}{1+O}$$

Monte Hall Problem

$i, j, k = 1, 2, 3$

$Y_i = \{$you initially choose door $i\}$

$M_j = \{$Monte Hall then opens door j with a goat$\}$

$C_k = \{$car actually behind door $k\}$

You pick door 1 ($Y1$) and Monte opens door 2 ($M2$)
We want to compare $P(C1 \mid M2, Y1)$ with $P(C3 \mid M2, Y1)$
and

This is like ELISA

	Monte	ELISA
unknown :	location of car	true HIV status
data :	Monte showing goat behind door 2	What ELISA said

We want P(unknown | data) but we got P(data | unknown)
so use Bayes' Theorem to reverse the order of conditioning

$$\underbrace{\frac{P(C_1 \mid M_2, Y_1)}{P(C_3 \mid M_2, Y_1)}}_{\text{posterior odds}} = \underbrace{\left[\frac{P(C_1)}{P(C_3)}\right]}_{\text{prior odds}} \cdot \underbrace{\left[\frac{P(M_2, Y_1 \mid C_1)}{P(M_2, Y_1 \mid C_3)}\right]}_{\text{Bayes factor}}$$

By the rules $P(C_1) = P(C_3) = \frac{1}{3}$ so the prior odds are $\frac{P(C_1)}{P(C_3)} = \frac{1/3}{1/3} = 1$

To evaluate probability like $P(M_2, Y_1 \mid C_1)$, use the general form of product rule for AND

$$\frac{P(M_2, Y_1 \mid C_1)}{P(M_2, Y_2 \mid C_3)} = \frac{\overset{1}{\cancel{P(Y_1 \mid C_1)}} \cdot P(M_2 \mid Y_1, C_1)}{\cancel{P(Y_1 \mid C_3)} \cdot P(M_2 \mid Y_1, C_3)}$$

Y_i and C_j are independent so $P(Y_1 \mid C_1) = P(Y_1) = \frac{1}{3}$ and
$P(Y_1 \mid C_3) = P(Y_1) = \frac{1}{3}$

So $$\frac{P(C_1 \mid M_2, Y_1)}{P(C_3 \mid M_2, Y_1)} = \frac{P(M_2 \mid Y_1, C_1)}{P(M_2 \mid Y_1, C_3)} = \frac{1/2}{1} = \frac{1}{2}$$

Which means that the odds are behind door 1 are 2:1 against that, therefore you should switch.

After M_2 (given Y_i, C_j), the posterior odds in favor of the car behind door 3 are 2:1, so $P(C_3 \mid M_2, Y_i) = \frac{2}{3}$ and you should switch.

Your chance of guessing the car right to begin with is $\frac{1}{3}$ so $\frac{1}{3}$ of the time by him showing you where the goat is he hasn't given you any new information but 2/3s of the time he has shown you where the car is. Therefore you should switch because he's shown you that the car is behind door 3 $\frac{2}{3}$s of the time.

Cromwell's Rule Case Study: Prove the following two facts: for any D such that $P(D) > 0$

(a) If $P(A) = 0$ then $P(A|D) = 0$
(b) If $P(A) = 1$ then $P(A|D) = 1$

A: proposition whose truth value is known to you

D: some data relevant to A

What are the implications of Cromwell's Rule for the use of Bayes' Theorem as a formal model for learning from data? Explain Briefly

Anything you put prior probability 0 or 1 on has to have posterior probability 0 or 1, no matter how the data set comes out; this destroys the possibility of learning from data

$P(A|D) = \dfrac{P(A \text{ and } D)}{P(D)}$ but if $P(A) = 0$ then $P(A \cap D) = 0$

$P(A|D) = \dfrac{P(A \text{ and } D)}{P(D)}$ $P(A = 1)$ so $(A \cap D) = D$ and $P(D)/P(D) = 1$

Case Study: The Rasmussen Report - "The Reactor Safety Study" - problem: Estimate prob of a catastrophic accident at a nuclear power plant... No such event occurred before (1975)

"Solution": Use expert judgement to break down into a collection of simpler events connected together with AND/OR

Result: Their estimate of the prob. was extremely small: 10^{-12} and yet 4 years later; 3 mile island accident

Right calculation:

$P(*) = P\left(\begin{smallmatrix}\text{thing 1}\\ \text{breaks}\end{smallmatrix}\right) \cdot P\left(\begin{smallmatrix}\text{alarm 1}\\ \text{fails}\end{smallmatrix} \middle| \begin{smallmatrix}\text{thing 1}\\ \text{breaks}\end{smallmatrix}\right) \cdot P\left(\begin{smallmatrix}\text{thing 2}\\ \text{breaks}\end{smallmatrix} \middle| \begin{smallmatrix}\text{thing 1 breaks,}\\ \text{alarm 1 fails}\end{smallmatrix}\right) \ldots$

What they did was assume everything was independent!

$$P(*) = \underbrace{P\begin{pmatrix}\text{thing 1}\\ \text{breaks}\end{pmatrix}}_{\text{small}} \cdot \underbrace{P\begin{pmatrix}\text{alarm 1}\\ \text{fails}\end{pmatrix}}_{\text{small}} \cdot \underbrace{P\begin{pmatrix}\text{thing 2}\\ \text{breaks}\end{pmatrix}}_{\text{small}} \cdots = \text{tiny}$$

Multiplying a bunch of small probabilities close to zero

The actual conditional probabilities are fairly close to 1.

Consequences that follow Kolmogorov's Axioms

① $P(\emptyset) = 0$

② $P(A^c) = 1 - P(A)$

③ If $A \subset B$ then $P(A) \leq P(B)$

④ For all events A, $0 \leq P(A) \leq 1$

⑤ For all events A, B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

⑥ For any events $A_1, A_2, \ldots, A_n$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i) \quad \text{and}$$

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq 1 - \sum_{i=1}^{n} P(A_i^c)$$

T-S Disease in more detail

NNNNN	0: # of T-S babies = Y
TNNNN	
NTNNN	
NNTNN	
NNNTN	
NNNNT	1
TTNNN	
TNTNN	
TNNTN	
TNNNT	
NTTNN	
NTNTN	
NTNNT	
NNTTN	
NNTNT	
NNNTT	2
⋮	⋮
TTTTT	5

$P(Y=0) = P(\text{exactly } 0 \text{ T-S babies})$

$= P\begin{pmatrix}\text{1st baby} \\ \text{not T-S}\end{pmatrix} \,\&\, \begin{matrix}\text{2nd baby} \\ \text{not T-S}\end{matrix} \,\&\cdots\&\, \begin{matrix}\text{5th baby} \\ \text{not T-S}\end{matrix}\Bigg)$

Independence

$= P\begin{pmatrix}\text{1st baby} \\ \text{not T-S}\end{pmatrix} \cdot P\begin{pmatrix}\text{2nd baby} \\ \text{not T-S}\end{pmatrix} \cdots P\begin{pmatrix}\text{5th baby} \\ \text{not T-S}\end{pmatrix}$

Identical distribution

$= \left[1 - P\begin{pmatrix}\text{1st baby} \\ \text{T-S}\end{pmatrix}\right] \cdots \left[1 - P\begin{pmatrix}\text{5th baby} \\ \text{T-S}\end{pmatrix}\right]$

$= 1 p^0 (1-p)^5 = (1-p)^5$ where $p = \frac{1}{4}$

$= 24\%$

A similar line of reasoning gives $P(\mathbb{Y}=5)=P(TTTTT)$
$= 1p^5(1-p)^0 = p^5$

What about $P(\mathbb{Y}=1)$? The outcomes with 1 T-S baby all have 1 T and 4 Ns, so each one has probability $p^1(1-p)^4$, and there are 5 of them so $P(\mathbb{Y}=1)=5p^1(1-p)^4$.

By similar reasoning $P(\mathbb{Y}=2)=10p^2(1-p)^3$

The outcomes with $(\mathbb{Y}=3)$ are mirror images of those with $(\mathbb{Y}=2)$: $\left\{\begin{matrix} TTNNN \\ NNTTT \end{matrix}\right\}$

So there must also be 10 elements of S with $(\mathbb{Y}=3)$ and
$P(\mathbb{Y}=3)=10p^3(1-p)^2$

$(Y=4)$ is a mirror image of $(Y=1)$ so

$P(Y=4) = 5p^4(1-p)^1$

Soon we'll call Y a random variable and y to stand for a possible value of Y.

# of T-S babies (y)	$P(Y=y)$	With $p=\frac{1}{4}$
0	$1p^0(1-p)^5$	0.2373
1	$5p^1(1-p)^4$	0.3955 : Highest chance
2	$10p^2(1-p)^3$	0.2637
3	$10p^3(1-p)^2$	0.0879
4	$5p^4(1-p)^1$	0.0146
5	$1p^5(1-p)^0$	0.0010
	Sum: 1	1.0000

It looks like $P(Y=y) = \, ? \, p^y(1-p)^{n-y}$ where $n=5$

The multiplier ? comes from Pascal's triangle

Ex: You have an ordinary deck of $n=52$ playing cards. How many possible poker hands of $k=5$ cards can you draw at random w/out replacement from the deck?

It's like filling in 5 slots: $\underset{52}{\underline{8}}$ $\underset{51}{\underline{\quad}}$ $\underset{50}{\underline{\quad}}$ $\underset{49}{\underline{\quad}}$ $\underset{48}{\underline{\quad}}$: # of ways that slot can be filled

8 of diamonds

The total # of ways you can do this is $52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 = n(n-1)\cdots(n-k+1) = 311{,}875{,}200$ ways. This is the # of permutations of 52 things taken 5 at a time.

Def: The # of permutations of n distinct things taken k at a time is written as

$$P_{n,k} = n(n-1)\dots(n-k+1)$$

How many possible orderings of a 52-card deck are there? Now there are 52 slots so the total # must be

$$52 \cdot 51 \cdot \dots \cdot 1 = n(n-1)\dots 1 = n! \quad (n \text{ factorial})$$

$$= 8.1 \cdot 10^{67}$$

Any time you have to solve calculations with calculus, it's perfectly okay to use Wolfram Alpha online to do it for you.