The Birthday Paradox

Valentin Cornaciu

2022-12-02

The Problem

- The birthday paradox, also known as the birthday problem, states that in a random group of 23 people, there is about a 50 percent chance that two people have the same birthday.
- There are multiple reasons why this seems like a paradox. New sources will be added it to this ever growing presentation.

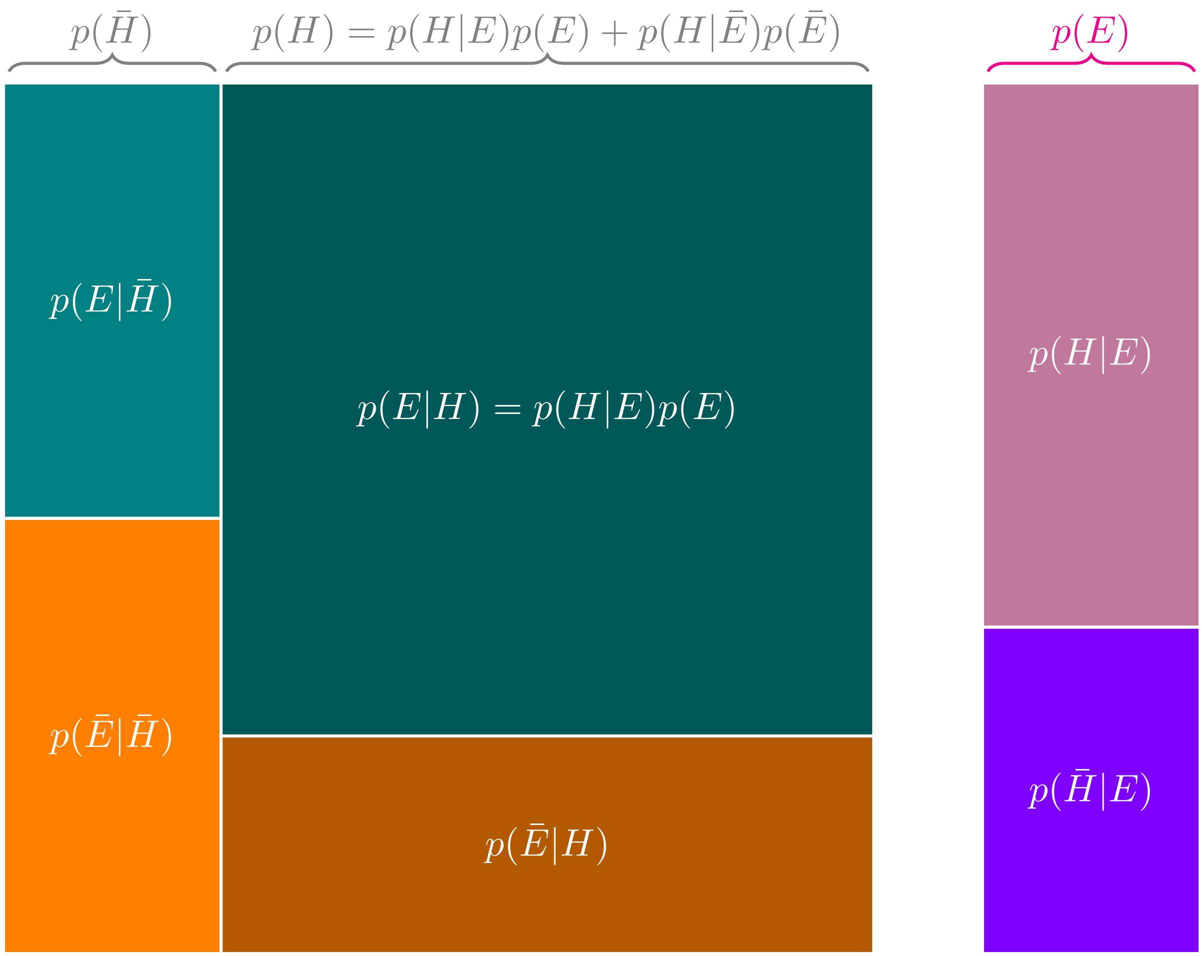
1. Scientific American proposes the same problem, creating another paradox, by stating:
   a) Every one of the 253 combinations (of two persons) has the same odds, $p = 0.99726027$, of not being a match, *they say*.
   b) If you calculate (364/365)^253, you'll find there's a 49.952 percent chance that all 253 comparisons contain no matches, *they say*.

Assuming a year of 365 days only

Remember Bayes

3 people only

- Why go for 23, before understanding the problem with only 3 persons?
- Define A_i the event where the i person will not share the same birthday with nobody and the negation by $\bar{A}_i$.
- We have 3 combinations (of two persons) A_1A_2, A_1A_3, A_2A_3.
- $P(A_1A_2)$ denotes the probability of both A_1 and A_2 being TRUE.
- If A_1A_2 and A_1A_3 both not being a match (TRUE), then A_2A_3 can be FALSE (share the same birthday).
- BUT: $P(A_1A_2A_3|G_3) = P(A_2A_3|G_3)$, so problem solved (one combination), why asking for p^3?


$p(\bar{H})$
$p(H) = p(H|E)p(E) + p(H|\bar{E})p(\bar{E})$
$p(E)$
$p(E|\bar{H})$
$p(E|H) = p(H|E)p(E)$
$p(H|E)$
$p(\bar{E}|\bar{H})$
$p(\bar{H}|E)$
$p(\bar{E}|H)$

Bayes settings

- We start by declaring the sample space, G_3, being a group of 3 persons, and here the negation $\bar{A}_i$ depends on G_3.
- The Problem is symmetric, so finding $P(A_3|G_3)$ will determine $P(A_2|G_3)$ as well.
- $P(\bar{A}_1\bar{A}_2\bar{A}_3|G_3) = 1/365^2$ and this can be generalized for any group
- $P(\bar{A}_1\bar{A}_2\bar{A}_3|G_3) = P(\bar{A}_1|\bar{A}_2\bar{A}_3G_3) \cdot P(\bar{A}_2\bar{A}_3|G_3) \Rightarrow P(\bar{A}_2\bar{A}_3|G_3) = 1/365$ (1)
- $P(A_1A_2A_3|G_3) = P(A_1|A_2A_3G_3) \cdot P(A_2A_3|G_3) = P(A_2A_3|G_3)$, by simple logic: if A_3 is TRUE and A_2 is TRUE, then the first person can't share the same birthday with nobody else in G_3, so the first probability is one. Also holds to be true for any $n > 2$:

$$P(A_1A_2...A_n|G_n) = P(A_1|A_2 \dots A_nG_n) \cdot P(A_2...A_n|G_n)$$

Bayes calculations

- (2): $P(\bar{A}_2|\bar{A}_3G_3) = 1/2$, by knowing that the third person shares the *birthday*, we have a 50% chance that will mach either of the remaining persons. This is so under the assumption that no more information is provided, so one must be indifferent (and its generalization, the principle of maximum entropy)!
- remember that probability represents a state of partial knowledge!
- (1)+(2): $P(\bar{A}_2\bar{A}_3|G_3) = P(\bar{A}_2|\bar{A}_3G_3) \cdot P(\bar{A}_3|G_3)$ so we get $P(\bar{A}_3|G_3) = 2/365$ or $P(A_3|G_3) = 363/365$
- Actually $P(A_2|A_3G_3) = 364/365$ so only by knowing that the third person is alone, will impose that the other persons will have *that claimed chance* of being alone (not matching the remaining person, or a G_2 problem)
- Finally, $P(A_2A_3|G_3) = P(A_2|A_3G_3)P(A_3|G_3) = \frac{363 \cdot 364}{365^2}$

Another solution

- Even though not recommended in general, Bayes being preferred, sampling all possible different birthdays of 23 persons can pave the way.

- The number of ordered arrangements of 23 days taken from 365 unlike days is $365!\,/(365-23)!$
- Not to confuse with the number of ways of selecting 23 different days! Not looking for $\binom{365}{23}$, because
- We consider the total number of cases 365^{23}, so we get $P(A_1A_2...A_{23}|G_{23}) = 364 \cdot 363 \cdots 343/365^{22} \approx 0.4927$

Mistakes explained

- $p = P(A_1 A_2)$ has different meanings for different groups of people, so claiming that $p = 364/365$ it is false in general, but TRUE in G_2
- Same mistakes can be found almost everywhere, the difference in calculations are luckily close (49.952% vs. 49.27%)
- Can not mix results from G_2 to G_n, for $n > 2$
- We might guess that the intention was to calculate the probabilities of having 2 different days as a pair and impose it to all pairs. But (A_2, A_3) or any other pair of events are not independent!
- Now imagine there were 366 persons in the group. It is clear that $P(A_1 A_2 ... A_{366} | G_{366}) = 0$, but $(364/365)^{66795} > 0$

Remarks

- Google returned the selected papers, no judgmental sampling was intended. The author has not been involved in a dispute with the editors of the mentioned papers
- The views expressed in this article are those of the author.
- This is an updated report using the latest information and tries to adapt the main idea using more information as they appear,
- actively pursuing error-correction by creating criticisms of both existing ideas and new proposals.
- To create an annotation, select any text and then select the Annotate button either in a private or public group.