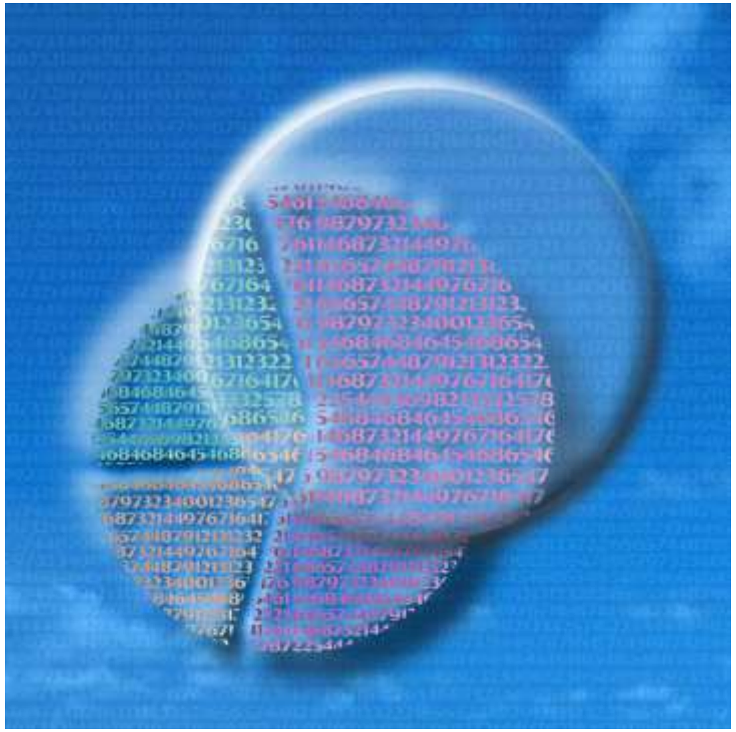


Lesson 3
Probabilities

Andrea Saltelli

Introduction to Statistics
March 30, 2023 UPF-BSM

BARCELONA SCHOOL OF MANAGEMENT

CNR ISTC ISTITUTO DI SCIENZE E TECNOLOGIE DELLA COGNIZIONE

UNIVERSITAS BERGENSIS

Fair games: is it fair to play a 7 against a 3?

Photo: Creator: Bob Adelman | Credit: © Bob Adelman/Corbis. Dices https://creazilla.com/

Fair games: is it fair to play a 7 against a 3?

How can a 7 come about?						
6	5	4	3	2	1	
1	2	3	4	5	6	

Fair games: is it fair to play a 7 against a 3?

How can a 3 come about?		
2	1	
1	2	

How can a 7 come about?						
6	5	4	3	2	1	
1	2	3	4	5	6	

How can a 3 come about?		
2	1	
1	2	

Bad luck: strike and rain

Photo: Toronto Public Library Archive. Drawing: Shutterstock

Strikes on average once every 100 days

Rains on average once every 10 days

What is the probability of rain and strike in any given day?

Strikes on average once every 100 days

Rains on average once every 10 days

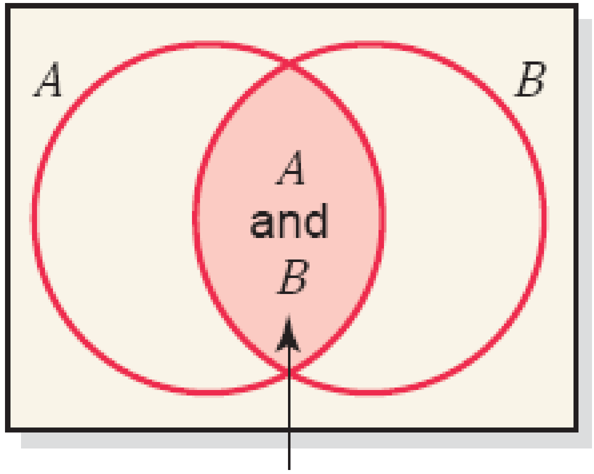
Definition

Joint Probability The probability of the intersection of two events is called their *joint probability*. It is written as

$$P(A \text{ and } B)$$

P(A)=P(strike)=1/100
B(B)=P(rain)=1/10

A and B are independent



Intersection of *A* and *B*

P(A and B)=P(A)P(B)=
(1/10)/(1/100)=1/1000

Discerning tea drinker

From Mary Evans picture library
https://www.maryevans.com/search.php?prv=preview&job=5382258&itm=2&pic=10029395&row=1

Claims she can distinguish a cup of tea where milk has been added before pouring the tea or after

Eight cups of tea are given to her where four times the milk is added before, and four after. She can tell the difference every time.

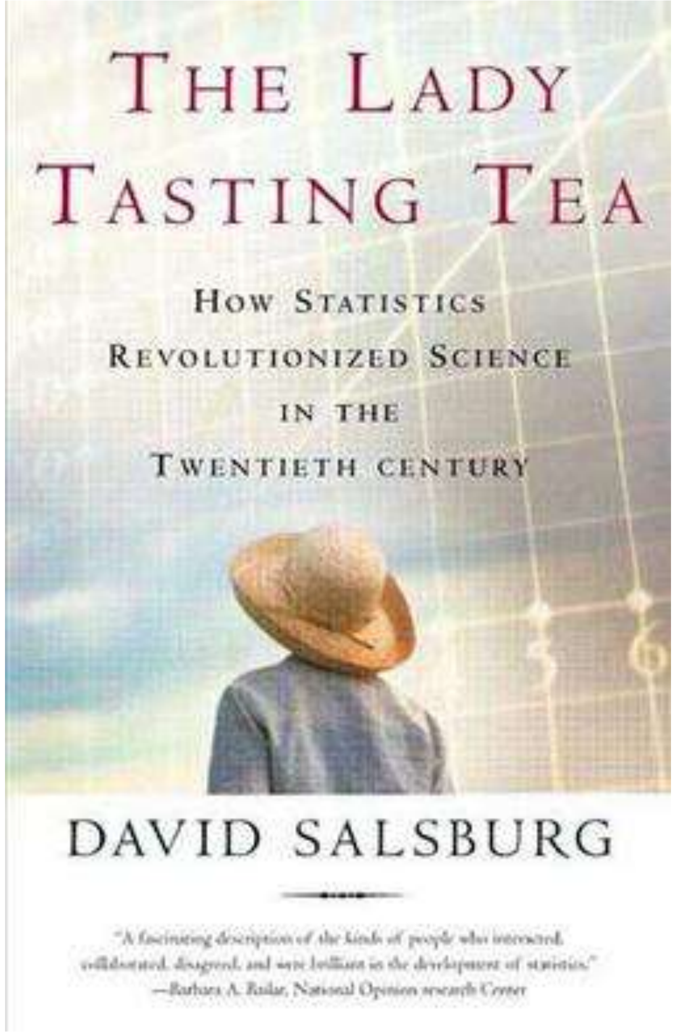
Luck?

https://www.twinkl.es/resource/teacup-display-cut-outs-t-m-3640

The problem is hence to pick the four ‘right’ cups / colors out of four …

Solving this demands some combinatorial calculus, but the solution is one chance in seventy of getting this by luck

Discerning tea drinker



This was in fact a real experiment run with statistician Ronald Fisher and reported in 1935

Some definitions

Definition

Experiment, Outcomes, and Sample Space An *experiment* is a process that, when performed, results in one and only one of many observations. These observations are called the *outcomes* of the experiment. The collection of all outcomes for an experiment is called a *sample space.*

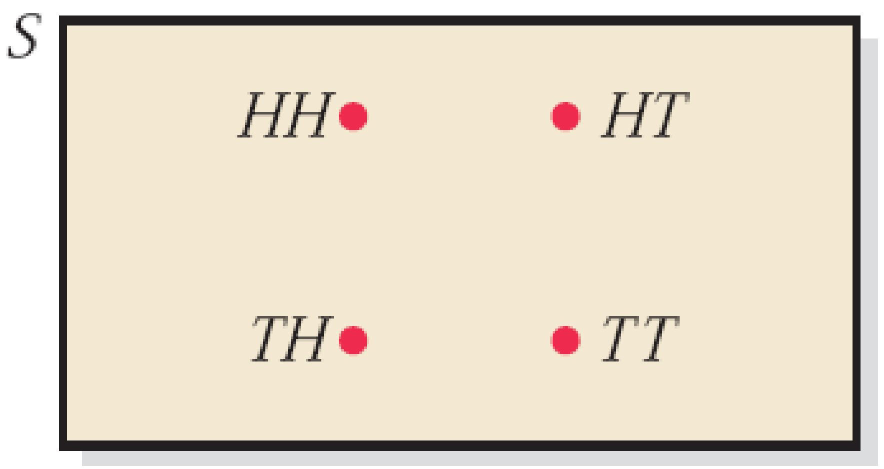
Source: Mann, Prem S. 2010. Introductory Statistics. Wiley.

iStock by Getty images

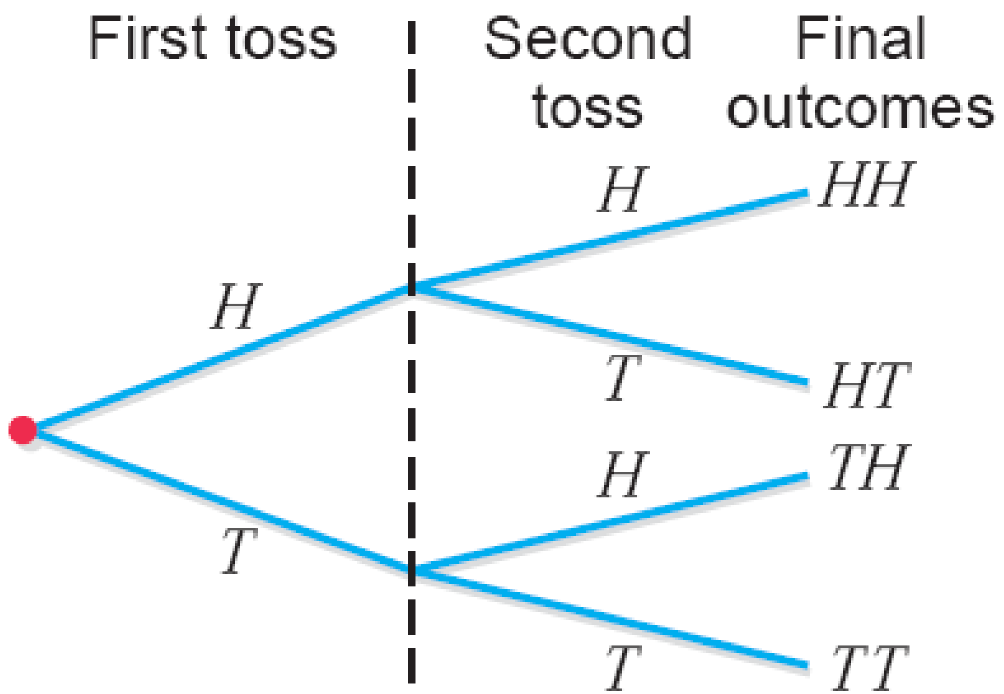
Experiment = tossing a coin
Outcomes = H,T
Sample space = {H,T}

Experiment = rolling a dice
Outcomes = 1,2,3,4,5,6
Sample space = {1,2,3,4,5,6}

Venn diagram representing the sample space when experiment = tossing two coins



Tree diagram for the same experiment



The probability is a number between 0 and 1

$P(Head)=P(Tail)=1/2$

$P(Head)+P(Tail)=1$

The sum of the probabilities of all possible outcomes is one

Definition

Law of Large Numbers If an experiment is repeated again and again, the probability of an event obtained from the relative frequency approaches the actual or theoretical probability.

Source: Mann, Prem S. 2010. Introductory Statistics. Wiley.

The theoretical probability of a head in a toss of a coin is $\frac{1}{2}$

True or false?

Definition

Law of Large Numbers If an experiment is repeated again and again, the probability of an event obtained from the relative frequency approaches the actual or theoretical probability.

Source: Mann, Prem S. 2010. Introductory Statistics. Wiley.

The theoretical probability of a head in a toss of a fair coin is $\frac{1}{2}$

↑

assumption

Definition

Law of Large Numbers If an experiment is repeated again and again, the probability of an event obtained from the relative frequency approaches the actual or theoretical probability.

Law formulated (published) posthumously in 1713

The same of the Euler number e

Jacob Bernoulli
1655-1705

Definition

Law of Large Numbers If an experiment is repeated again and again, the probability of an event obtained from the relative frequency approaches the actual or theoretical probability.

John Edmund Kerrich: tossing a coin 10,000 times ➔ 5067 heads (1946, before large computers)

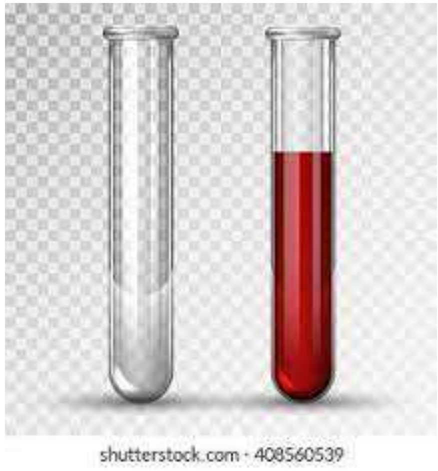
Source: https://en.wikipedia.org/wiki/John_Edmund_Kerrich

Thomas Bayes and his theorem

Thomas Bayes, 1701-1761

Two ways of looking at probabilities

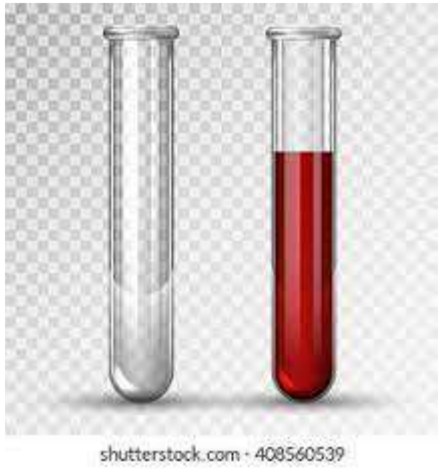
Bayesian, reason to believe, degree of conviction, ‘subjective’, inductive…



Frequentist, the limit of a sequence of equiprobable trials; also called ‘objective’, statistical…

Strong beliefs (even wars) in both positions

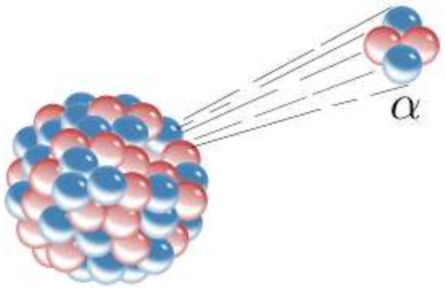
Bayesian, XVIII century, a comeback in the XX



Frequentist, XIX century, mainstream e.g. in teaching

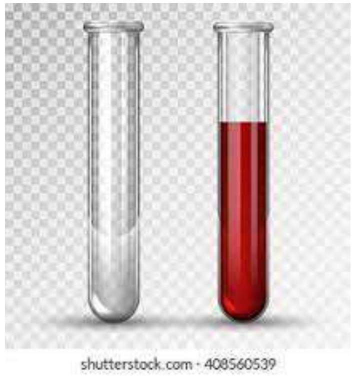
Other distinguish not two but three ways of defining probabilities

Equal possibilities based on physical symmetry (coins dices)



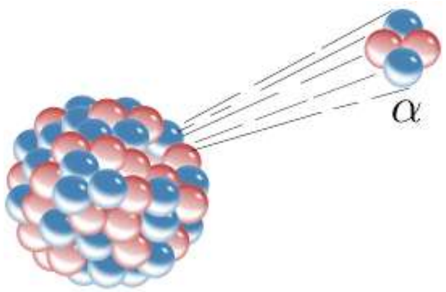
Observed frequencies (decay)

Degree of conviction
Experiments

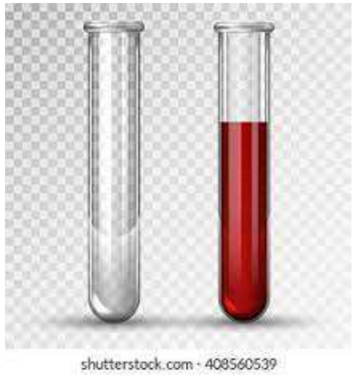


Equal possibilities

Frequencies



Degrees of belief

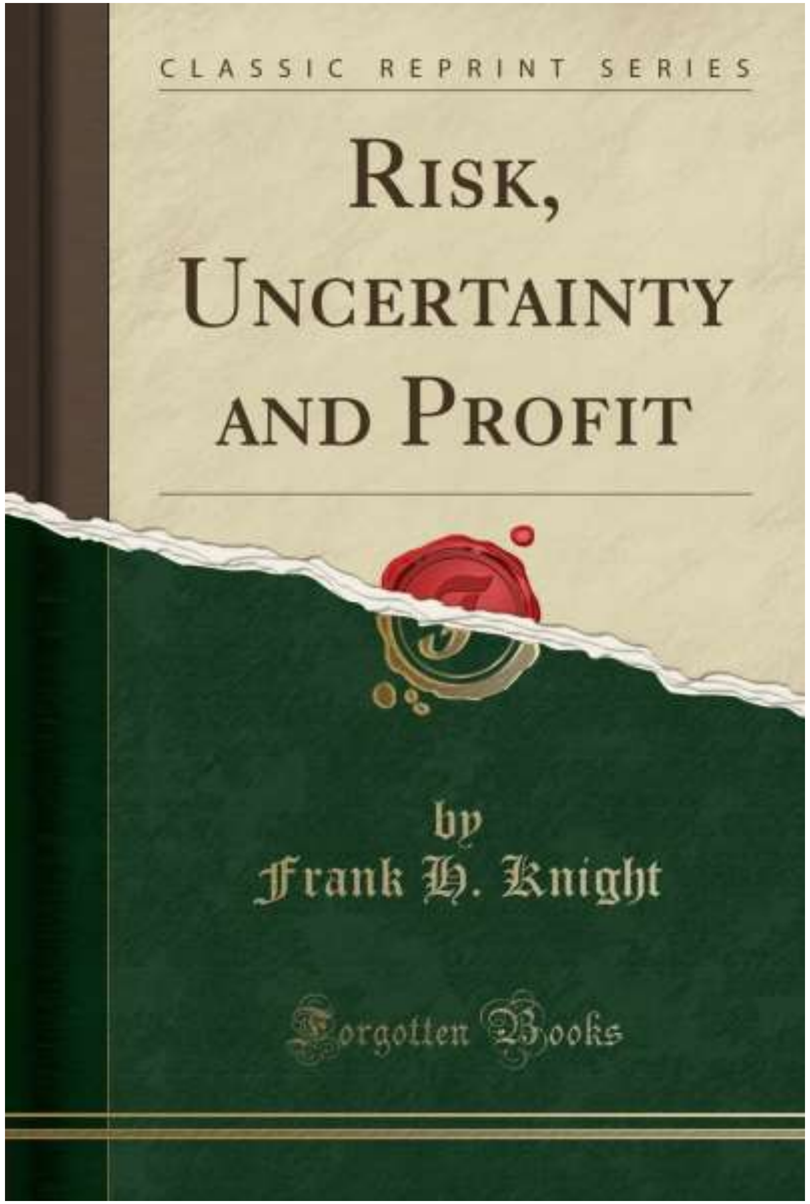


Are these separate domains?

Frank Knight (1921) distinguished risk from uncertainty

Risk = know outcomes & probabilities

Uncertainty = unsure about the probabilities



Frank H. Knight
1885–1972

Quote:

"We live in a world of contradiction and paradox, a fact of which perhaps the most fundamental illustration is this: that the existence of a problem of knowledge depends on the future being different from the past, while the possibility of the solution of the problem depends on the future being like the past."

Probability of A given B: $P(A|B)$

Probability of B given A: $P(B|A)$

Example:

A=I have disease COVID-19
B=My clinical test for disease COVID-19 is positive

Example:

A=I have disease X
B=My clinical test for disease X is positive

In general $P(A|B)$ is neither 1 nor $P(A)$,
nor in general is $P(B|A)$ 1 or $P(B)$ because …

The fact that the test is positive does not guarantee that I have the disease, nor the fact that I have the disease guarantees that the test will be positive (more soon)

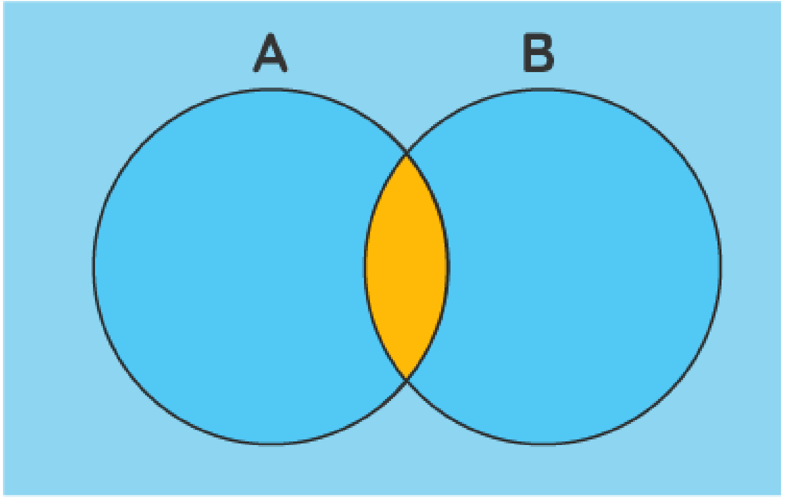
Source: https://www.123rf.com/photo_30337574_sick-boy-lying-in-bed.html

$P(A|B)=P(B|A)=P(A)P(B)$ if A and B are independent (example strike and rain)

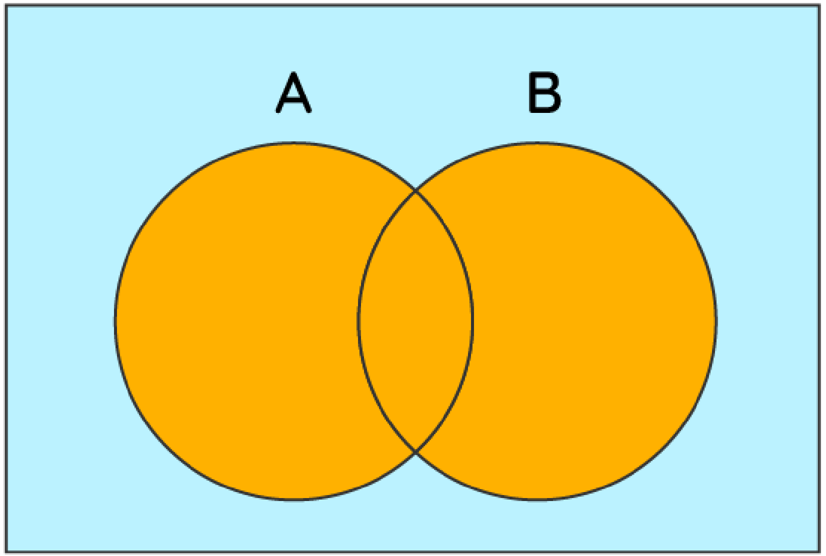
Probability of A given B: $P(A|B)$

Probability of B given A: $P(B|A)$

Probability and A **and** B being simultaneously true $P(A \cap B)$



Probability and A **or** B being true P(AUB)



In general $P(A|B) \neq P(B|A)$ but

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

A classic exercise in screening

You test positive for AIDS (one test only). Time for despair?

Only one 1 in 100,000 has AIDS in your population

The test has a 5% false positive rate

Already one can say: in a population of say 100,000 one person will have AIDS and 5,000 (5% of 100,000) will test positive

➔ Don't despair (yet)

A classic exercise in screening

You test positive for AIDS (one test only). Time for despair?

Only one 1 in 100,000 has AIDS in your population

The test has a 5% false positive rate

Already one can say: in a population of say 100,000 one person will have AIDS and 5,000 (5% of 100,000) will test positive

➔ Don't despair (yet), why?

In a population of say 100,000 one person will have AIDS and 5,000 (5% of 100,000) will test positive; so your chance of having AIDS is
1/5000 = 0.0002 = 0.02%

… but we can use Bayes instead

$$P(AIDS|TestPositive)P(TestPositive)$$
$$= P(TestPositive|AIDS)P(AIDS)$$

$$P(AIDS|TestPositive) =$$
$$= \frac{P(TestPositive|AIDS)P(AIDS)}{P(TestPositive)}$$

The values to plug in

P(Test Positivo)=0.05 [approximation]
P(AIDS)=0.00001 [prevalence]
P(Test Positivo|AIDS)= 1 [assumption of no false negative]

$$P(AIDS|TestPositive) = \frac{1*0.00001}{0.05} = 0.0002 = 0.02\%$$

as before

The power of Bayesian statistics is foremost in saying something about unknown causes given known events

Theory Hi is under discussion and an experiment is performed that gives the outcome E

What is **now** the probability of Hi?

If the experiment gives the outcome E, how does the probability of Hi change?

- If I know the probability of Hi before the experiment P(Hi), known as ‘the prior’
- And I also know how probably event E would be if Hi were true, then

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{P(E)}$$

If there are alternative theories that would result in outcome E, H1,H2,⋯

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{P(E|H_1) + P(E|H_2 \dots)}$$

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{P(E|H_1) + P(E|H_2 \dots)}$$

"the probability of the existence of each cause is equal to the probability of the event given that cause, divided by the sum of all the probabilities of the event given each of these causes"

(Laplace, 1774)

Pierre Simon Laplace
1749-1827

Inferring the causes from observed events is "the way of the historian, the policemen, and a doctor, who suggest a diagnosis based on symptoms" (Desrosières 1993)

Arthur Conan Doyle was a medical doctor …

Source: https://www.pngwing.com/en/free-png-nsxet

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{P(E|H_1) + P(E|H_2 \dots)}$$

The probability of theory Hi given outcome E also depends upon the probability of E for all possible theories

$$P(H_i|E) = \frac{P(E|H_i)}{P(E)} P(H_i)$$

posterior

likelihood

prior

The P-test saga

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Review

CrossMark click for updates

Cite this article: Colquhoun D. 2014 An investigation of the false discovery rate and the misinterpretation of *p*-values. *R. Soc. open sci.* **1**: 140216.
http://dx.doi.org/10.1098/rsos.140216

An investigation of the false discovery rate and the misinterpretation of *p*-values

David Colquhoun

Department of Neuroscience, Physiology and Pharmacology, University College London, Gower Street, London WC1 6BT, UK

P values by way of an example

- Two groups, one with a placebo, one with the treatment
- Random allocation to groups
- The difference *d* between the means of the two groups is tested (is it different from zero?)
- $p=0.05$ implies that if there were no effect the probability of observing a value equal to *d* or higher would be *5%*

“At first sight, it might be thought that this procedure would guarantee that you would make a fool of yourself only once in every 20 times that you do a test”

Colquhoun D. 2014 An investigation of the false discovery rate and the misinterpretation of p-values. R. Soc. Open sci. 1: 140216. http://dx.doi.org/10.1098/rsos.140216

Another exercise in screening (Colquhoun 2014)

You test positive for mild cognitive impairment (MCI) (one test only). Time to retire?

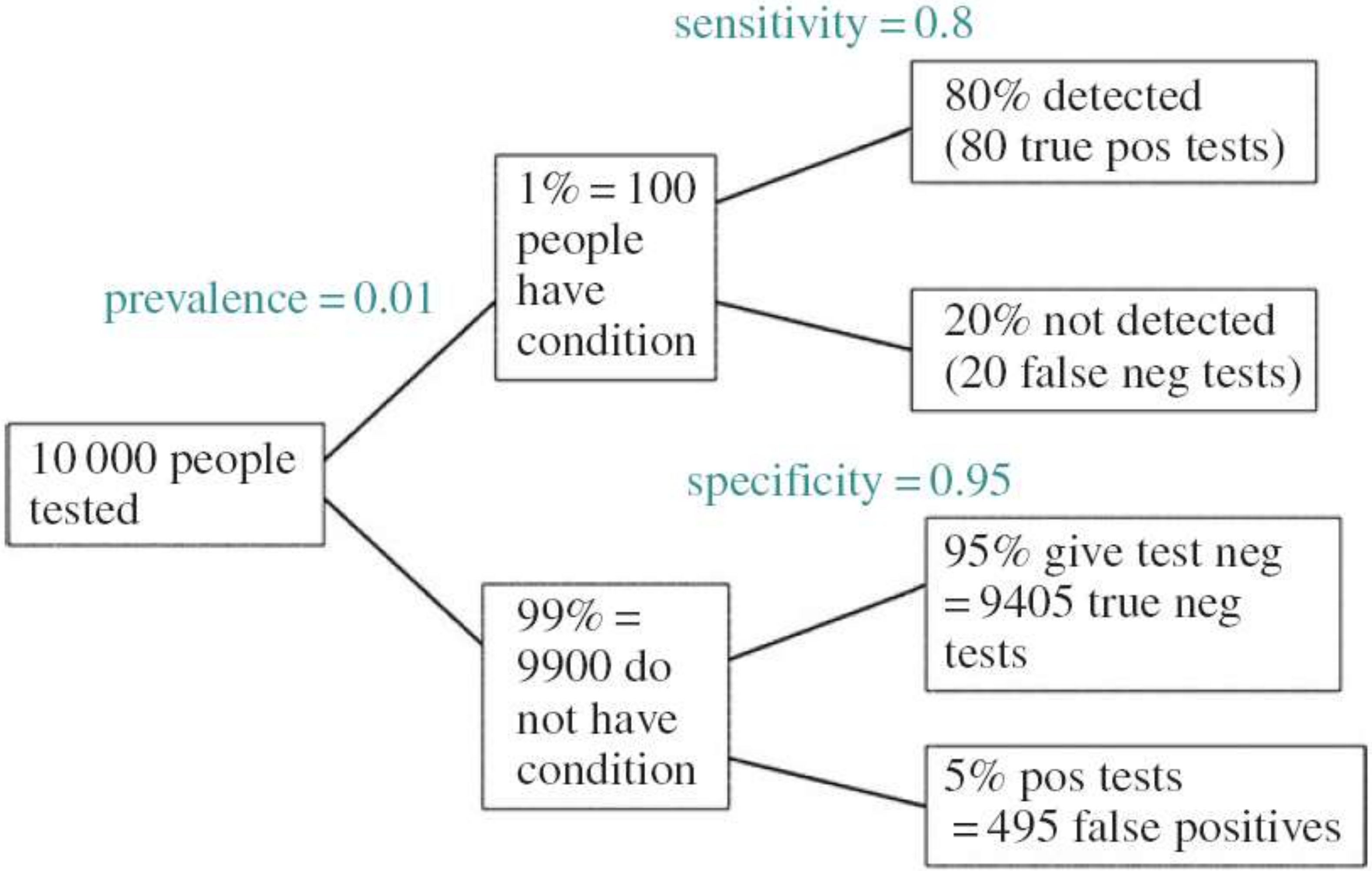
MCI prevalence in the population 1%, i.e. in a sample of 10,000 then 100 have MCI and 9,900 don't

The test has a 5% false positive rate; of the 9,900 who don't have MCI 495 test (false) positive and the remaining 9,405 (true) negative

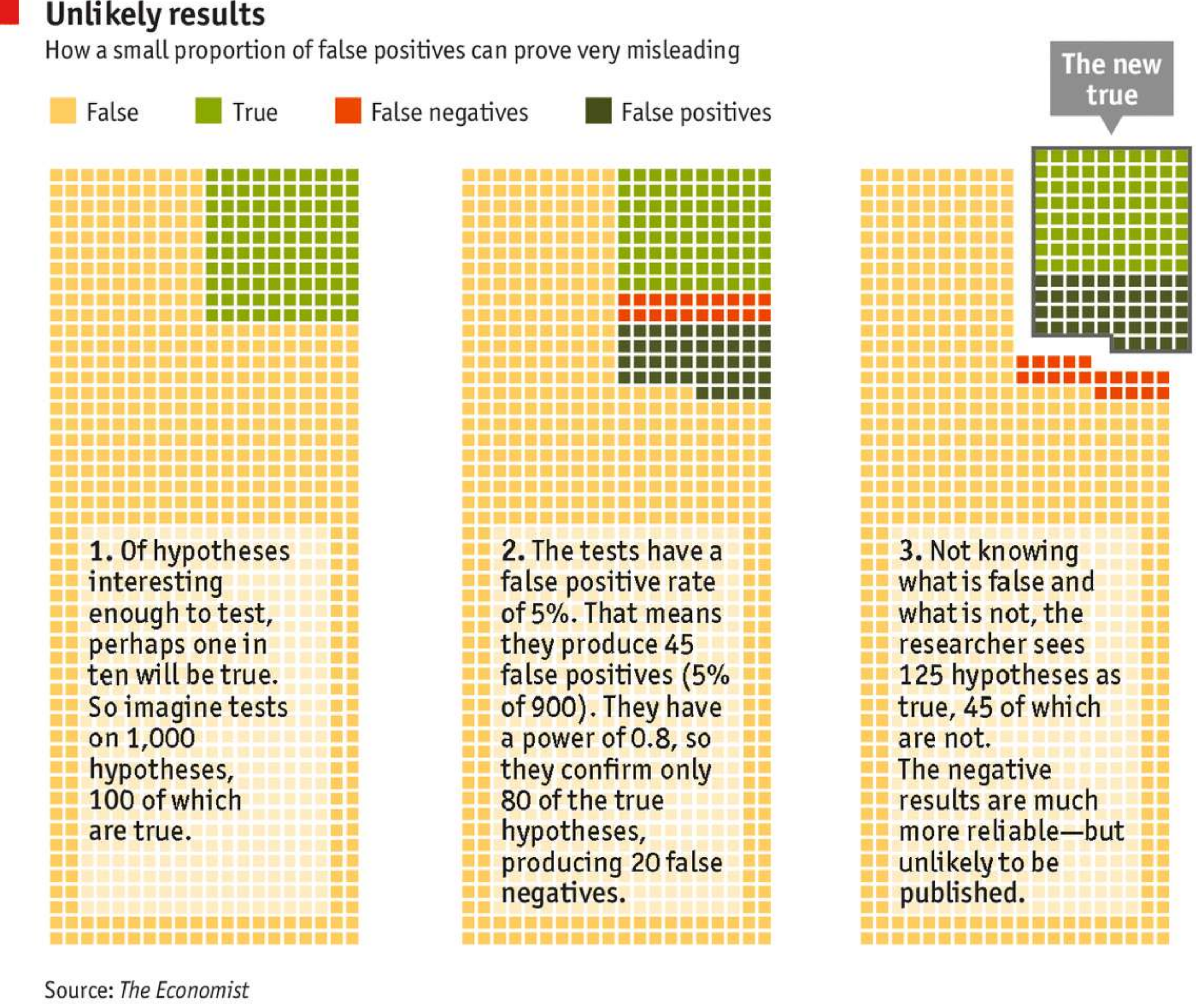
The test does not pick all the 100 MCI but only 80; there will be 20 false negative. So we see 80+ 495=575 positive of which only 80 (a 14%) are true and the remaining 495 (a 86%) false

➔ It does not make sense to screen the population for MCI!

The number 86% = 495/(495+ 80) is our false discovery rate



Colquhoun D. 2014 An investigation of the false discovery rate and the misinterpretation of p-values. R. Soc. Open sci. 1: 140216. http://dx.doi.org/10.1098/rsos.140216



The false discovery rate is the black area divided by the (green + black) one

“20% of the faculty teaching statistics in psychology, 39% of the professors and lecturers, and 66% of the students” don’t understand what the P-test is about

Gigerenzer, G., 2018, Statistical Rituals: The Replication Delusion and How We Got There, Advances in Methods and Practices in Psychological Science, 1–21

[Main source]


Introductory
Statistics
SEVENTH EDITION
MANN
7
PREM S. MANN

The End

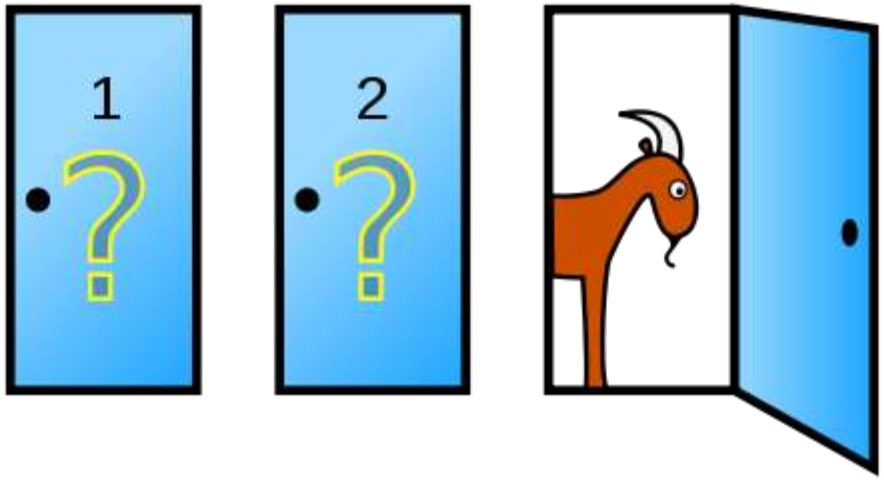
https://mstdn.social/@AndreaSaltelli/

Practicum

Luca's problem; the interrupted game. Wins the first who gets at 5 victories. The game is halted on a 3 to 5 situation. How to share?

Luca Pacioli 1447-1517
Source: Wikipedia Commons

Posed by Pacioli, solved by Pascal



Monty Hall problem

Stay or change door?