

Some practical exercises

Giulia Cereda

Dipartimento di Statistica, Informatica, Applicazioni “G. Parenti”

Università di Firenze

giulia.cereda@unifi.it

- On June 18, 1964, Juanita Brooks was robbed. Someone pushed her to the ground and took her purse. She caught a glimpse of a young blonde woman running away. When questioned by the police, Brooks reported that the young woman wore her hair in a ponytail and had escaped in a yellow car driven by a dark-skinned man with a beard and moustache.
- A few days later, Janet and Malcolm Collins were arrested and prosecuted after a police officer happened to pass their house and noticed that their appearance matched the description given by Juanita Brooks.
- In court, neither the victim nor the other witness was able to confirm that the defendants were the people who had committed the robbery

The Collins case

-
- In court, the prosecution's case was entrusted to a mathematics instructor. He wrote the following table on a blackboard:

Characteristic	Probability in the population
Yellow car	1/10
Man with a moustache	1/4
Woman with a ponytail	1/10
Blonde woman	1/3
Man with a beard	1/10
Interracial couple	1/1000

- The instructor considered that the probability of observing simultaneously two people in the population with all six characteristics could be obtained by multiplying the probabilities of each characteristic using the product rule.
- In this way, a very small probability was obtained

$$\frac{1}{12M} = 0.00000000833,$$

for any couple to possess all the characteristics, and it was concluded that the couple was guilty.

Some problems:

- Are the probabilities in the table accurate?
- Is the witness 100% reliable?

But most importantly:

- Is it correct to consider the characteristics independent? If 10% of men have a beard and 25% have a moustache, it does not necessarily follow that 2.5% of men have both a beard and a moustache.
- Even assuming that the figure $1/12M$ is accurate, it corresponds to the probability that a couple randomly selected from the population has those characteristics, not to the probability that the two defendants are innocent.

- This probability, which we can express as $\Pr(E \mid H_d)$, was instead used as the probability of innocence, $\Pr(H_d \mid E)$.

This is the prosecutor's fallacy.

Un modo per risolvere il Caso Collins

Assuming that the figure $1/12M$ is accurate:

- H_p : the Collinses are guilty
- H_d : another couple is guilty
- E : match between the characteristics of the Collinses and those of the perpetrators
- N : number of couples in the city

$$\frac{\Pr(H_p | E)}{\Pr(H_d | E)} = \frac{\Pr(E | H_p) \Pr(H_p)}{\Pr(E | H_d) \Pr(H_d)} = \frac{1}{1/12M} \frac{1/N}{(N-1)/N} = \begin{cases} 4 & \text{if } N = 3 \text{ million,} \\ 2 & \text{if } N = 6 \text{ million,} \\ 1 & \text{if } N = 12 \text{ million.} \end{cases}$$

These correspond to

$$\Pr(H_p | E) = \begin{cases} 4/5 & \text{if } N = 3 \text{ million,} \\ 2/3 & \text{if } N = 6 \text{ million,} \\ 1/2 & \text{if } N = 12 \text{ million.} \end{cases}$$

- Sally Clark was an English solicitor who, in November 1999, became the victim of a miscarriage of justice when she was found guilty of murdering her two infant sons.
- Clark's first son died in December 1996, a few weeks after his birth, and her second son died in similar circumstances in January 1998.
- One month later, Clark was arrested and tried for both deaths. The defence argued that the children had died from sudden infant death syndrome (SIDS).

- The prosecution relied on statistical evidence presented by the paediatrician Sir Roy Meadow, who testified that the probability of two children from an affluent family suffering from SIDS was 1 in 73 million.
- He had obtained this figure from his estimate that the probability of a single SIDS death under similar circumstances was 1 in 8,500.
- The Royal Statistical Society subsequently issued a statement arguing that there was no statistical basis for Meadow's claim and expressing concern about the "misuse of statistics in the courts."
- Why?

- Ray Hill, Professor in the Department of Mathematics at the University of Salford, conducted a detailed analysis of the CESDI data and concluded that these deaths were certainly not independent.
- According to his estimates, the siblings of children who died from SIDS have a probability between 10 and 22 times higher than average of dying in the same way.
- Using the figure of 1 in 1,303 for the probability of a first SIDS death, the probability of a second SIDS death in the same family is between 1 in 60 and 1 in 130.
- There are insufficient data to be more precise or to take family factors into account, but it seems reasonable to use an approximate figure of 1 in 100.
- Multiplying $1/1,303$ by $1/100$ gives an estimated incidence of double SIDS deaths of approximately 1 in 130,000. Since 650,000 children are born each year in England and Wales, we may expect approximately five families to suffer a second tragic loss. This is supported by the FSID, which reports becoming aware of one or two such cases each year.

- **Prosecutor's fallacy.** Even using the more accurate figure of 1 in 130,000, this is not the probability that Sally Clark was innocent. It is the probability that a randomly selected family would lose two children to SIDS.

We can formulate two hypotheses:

- H_p : the two children were killed by their mother;
- H_d : the two children were not killed by their mother;
- E : the two children died.

$$\frac{\Pr(H_p | E)}{\Pr(H_d | E)} = \frac{\Pr(E | H_p) \Pr(H_p)}{\Pr(E | H_d) \Pr(H_d)}.$$

We assume that, if their mother did not kill them, the only possible cause of death was SIDS.

- $\Pr(H_p)$ depends on the available case information; suppose that $\Pr(H_p) = 1/(10 \text{ million})$;
- $\Pr(H_d) = 1 - \Pr(H_p)$;
- $\Pr(E | H_p) = 1$;
- $\Pr(E | H_d) = 1/130,000$.

This gives posterior odds of 0.13, corresponding to a probability of guilt of

$$\Pr(H_p | E) = 0.11.$$

Clearly, the result depends strongly on the prior probability.

$\Pr(H_p)$	$\Pr(H_p E)$
10^{-6}	0.115
10^{-5}	0.565
10^{-4}	0.928

Table 1: Different posterior probabilities obtained from different prior probabilities

The tragic conclusion of the story is that Sally Clark spent three years in prison before being released following evidence suggesting that the two children may have died from an infection. In the meantime, however, the experience had traumatised her so deeply that she died from alcohol abuse a few years later, at the age of only 42.

Il caso O.J. Simpson

On 12 June 1994, Nicole Brown Simpson and Ronald Goldman were found murdered outside Brown Simpson's home on Bundy Drive in Los Angeles. O. J. Simpson, Brown Simpson's former husband, was charged with both murders.

The prosecution presented blood and DNA evidence collected from several locations, including the Bundy crime scene, Simpson's Ford Bronco and his Rockingham estate. Different stains were analysed using conventional blood testing. Some findings were compatible with Simpson, some with the victims, and some were interpreted as mixtures involving more than one contributor.

The defence challenged the collection, storage and laboratory handling of the samples. It raised the possibilities of contamination, sample handling error and deliberate planting. Simpson was acquitted in the criminal trial on 3 October 1995.

The blood characteristics were compatible with Simpson and incompatible with both victims. Approximately 0.43% of the relevant population shared these characteristics, corresponding to about 1 person in 200.

On cross-examination, it was noted that approximately 40,000 people in Los Angeles County might share these blood characteristics.

Consider the following conclusion:

"The blood matches Simpson and only one person in 200 has these characteristics. Therefore, there is only a 1 in 200 probability that the blood came from someone else, and a 199 in 200 probability that it came from Simpson."

Is it ok?

Simpson had a documented history of violence against Nicole Brown Simpson. An argument associated with the defence stated that only approximately 1 in 2,500 men who abuse their partners subsequently murder them.

The statistic was used to suggest that the history of abuse provided little evidence concerning the identity of the murderer.

1. What is wrong with this reasoning?
2. Express the 1-in-2,500 statistic as a conditional probability. Define each event clearly.

Notation

G : the batterer is guilty,

$\bar{G}$: the batterer is not guilty,

B : the victim was battered,

M : the victim was murdered.

The defence emphasized

$$P(G \mid B) = \frac{1}{2500} = 0.0004.$$

Good argued that the murder is already known. The relevant question is therefore

$$P(G \mid B, M).$$

Why the denominator changes

Probability

$$P(G \mid B)$$

$$[0.4\text{em}] P(G \mid B, M)$$

Reference population

All battered women

Battered women who were murdered

A rare outcome among all battered women may account for a large share of the much smaller group of battered women who were murdered.

Good's point

The probability of murder given battering does not answer the probability of guilt given battering and murder.

Good uses $P(M \mid G, B) = 1$ and estimates

$$P(M \mid \bar{G}, B) = P(M \mid \bar{G}) = \frac{1}{20,000} = 0.00005.$$

Bayes' theorem gives

$$P(G \mid B, M) = \frac{P(G \mid B)}{P(G \mid B) + P(\bar{G} \mid B)P(M \mid \bar{G})}.$$

Using $P(G \mid B) = 1/2500$ and $P(\bar{G} \mid B) = 2499/2500$,

$$P(G \mid B, M) = \frac{0.0004}{0.0004 + (0.9996)(0.00005)} \approx 0.889.$$

Conclusion

Once both battering and murder are known, Good's calculation raises the probability from $1/2500$ to about 0.9.

The calculation supports the relevance of prior battering. It does not establish guilt by itself. Its result depends on the assumed rates and must be combined with the other evidence in the case.

References

I. J. Good, "When Batterer Turns Murderer," *Nature* 375 (1995), 541.
doi:10.1038/375541b0

I. J. Good, "When Batterer Becomes Murderer," *Nature* 381 (1996), 481.
doi:10.1038/381481a0