LEGAL PROBABILISM

MARCELLO DI BELLO – ASU

HANDOUT #3 – SEPTEMBER 5, 2024

1. NON-COINCIDENCE ARGUMENTS

Rex v. Smith, 11 Cr. App. R. 229 (1915)

> "George Joseph Smith was accused of drowning Bessie Mundy in the small bathtub of their quarters in a boarding house. Mundy had left all her property to Smith in a will executed after their 'marriage' (Smith was already married). The trial court allowed the prosecution to prove the deaths of two other women who had gone through marriage ceremonies with Smith and to argue that the circumstances surrounding their deaths in their bathtubs were remarkably similar . . . The Court of Criminal Appeal affirmed the resulting conviction on the ground that the evidence in connection with Mundy's death alone made out a prima facie case" (p. 63).
>
> Fienberg and Kaye (1991), 'Legal and Statistical Aspects of Some Mysterious Clusters,' *Journal of the Royal Statistical Society*, 154(1), pp. 61-74.

Schema of the argument

> OCCURRENCE. Event E occurred when defendant was present.
>
> SIMILARITY. Other prior events $E', E'', \ldots$, all similar to E, also occurred when defendant was present.
>
> STATISTICAL FREQUENCY. Assuming everything happened by chance, the sequence of similar events $E', E'', \ldots$ culminating in E, all occurring in the presence of defendant, is statistically very unlikely.
>
> NON-COINCIDENCE. Event E could not have occurred as a result of chance alone. It was—most likely—the result of the defendant's purposeful conduct.

$\boxed{\text{QUESTIONS:}}$ Is this a good argument?

2. LUCIA DE BERK

Suspicion and arrest On September 4, 2001, at Juliana's children Hospital in the Hague, a baby died unexpectedly. The nurse in charge when the baby died was Lucia de Berk, a licensed pediatric nurse. That day another nurse reported to the supervisor that Lucia had been present at too many resuscitations. On December 13, 2001, Lucia was arrested and charged with multiple murders.

The numbers The numbers do look striking:

Shifts at Juliana's hosp.	without incidents	with incident (death or resuscitation)
without Lucia	887	0
with Lucia	134	8

Calculations At trial, Dr. Elffers, a law professor with a degree in statistics, calculated the probability that all incidents occurred during Lucia's shift, given the total number of incidents and total number of shifts. With the data in the above table and with other data regarding Lucia's shifts in two other hospitals, Elffers estimated this probability to be 1 in 342 million.[1]

$\boxed{\text{QUESTIONS:}}$ What probability is 1 in 342 million intended to be? Can we trust this number?

The court "The court is of the opinion that the probabilistic calculations given by Dr H. Elffers ... entail that it must be considered extremely improbable that the suspect experienced all incidents mentioned in the indictment coincidentally. These calculations consequently show that it is highly probable that there is a connection between the presence of the suspect and the occurrence of an incident (p.241)." Cited in Meester et al (2006), 'On the (ab)use of Statistics in the Legal Case Against the Nurse Lucia de B.' *Law, Probability and Risk*, 5(3-4), pp. 233–250.

$\boxed{\text{QUESTION:}}$ Is the Court correct or did it commit a fallacy?

3. OTHER CASES

Daniela Poggiani (IT) Again the numbers are striking:

Nurse	Same Zone	Opposite Zone	Total Deaths	Hours on Duty	Same Zone Mortality Rate	Opposite Zone Mortality Rate
N.1	68	58	126	3686	0.54	0.46
N.2	51	68	119	3545	0.43	0.57
N.3	64	60	124	3554	0.52	0.48
N.4	70	53	123	3535	0.57	0.43
N.5	64	41	105	3625	0.61	0.39
N.6	43	65	108	3532	0.40	0.60
DP	139	52	191	3577	0.73	0.27
N.8	60	44	104	3710	0.58	0.42
N.9	66	53	119	3741	0.55	0.45

It was an eventful series of trials from 2018 to 2023 which resulted in a definite acquittal.[2]

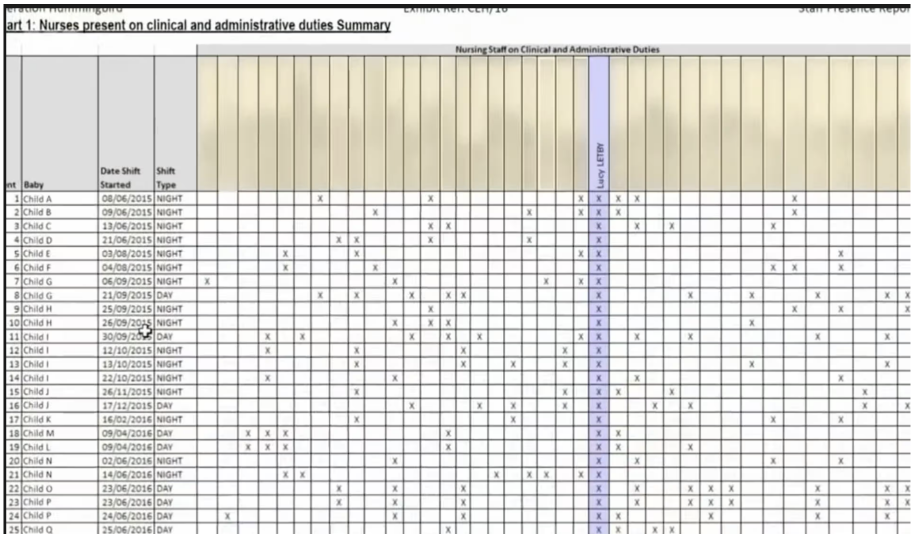
Lucy Letby (UK) Here is a case of a recently convicted nurse who was preset at all deaths!

[1] He used the formula for the *hypergeometric distribution*, where n is the total number of shifts, r is the number of Lucia's shifts, x is the number of incidents during Lucia's shifts, k is the total number of incidents, and p is the probability of one incident occurring, as follows:

$$\frac{\binom{r}{x}p^x(1-p)^{r-x}\binom{n-r}{k-x}p^{k-x}(1-p)^{n-r-k+x}}{\binom{n}{k}p^k(1-p)^{n-k}}$$

The numerator is the probability that Lucia witnessed an x number of incidents, during a total number of r shifts, and that the other nurses witnessed a $k-x$ number of incidents over a total of $n-r$ shifts. The denominator is the probability that a k number of incidents happened. So, the above formula describes the conditional probability that Lucia witnessed the number of deaths she witnessed, given how many deaths occurred overall.

[2] March 2016, Daniela Poggiani was sentenced by the Ravenna Court of Assizes to life imprisonment. July 7, 2017, the Court of Appeal of Bologna acquitted her. In 2018 the Supreme Court reversed and ordered a new trial. In 2019, she was acquitted again. In 2020 the Supreme Court ordered yet another trial. In two subsequent appeal trials, she was acquitted, but the Supreme Court reserved. October 2021: the Court of Appeal of Bologna acquitted Daniela yet again of murder charges. Finally in 2023 the Supreme Court upheld the acquittal.



4. A DISTINCTION

- *identification question*: who did it?
- *question about how it happened*: was it intentional or an accident?

5. SALLY CLARK

What happened Sally Clark's sons, a few months old, both died when their mother was present, the first in 1996 and the second in 1998. Sally was arrested and charged with murdering them. At trial, pediatrician Roy Meadow testified that the probability of two consecutive cot deaths, in a family similar to the Clarks, was 1 in 73 million. Theses statistics were interpreted to show that the possibility that the two deaths occurred accidentally or because of natural causes was minuscule. Sally was convicted to life imprisonment in 1999 and the conviction was upheld on appeal in 2000. In a second appeal, in 2003, she was acquitted because of new medical evidence.

QUESTIONS: How was the 1 in 73 million statistic calculated? Should we question it? What conditional probability does it represent?

Informal reasoning Suppose—although this was not the case in the Clark case—that reliable statistics show that the chance that two sons, in the same family, die of natural causes one after the other is 1 in 100 million. We might be impressed by such a low probability and reason as follows:

> If the chance that the sons died of natural causes is so low (1 in 100 million), the chance that the mother killed them must be overwhelmingly high.

QUESTION: Does this reasoning make sense? Is it correct?

6. BAYES' THEOREM IN SALLY CLARK

The set up Bayes' theorem offers a principled way to bring clarity to the question and weigh the evidence. Recall the formulation in terms of odds ratio, easier to use in calculations:

$$\frac{P(H|E)}{P(H'|E)} = \frac{P(E|H)}{P(E|H')} \times \frac{P(H)}{P(H')}.$$

Note that H' does not have to be the negation of H. We have:

- The two competing hypothesis H and H' to compare are that the sons died of natural causes—call it *natural*—and that Sally intentionally killed them—call it *kill twice*.
- The evidence is that both sons died.
- We need to assess the value of the posterior ratio $\frac{P(\textit{kill twice}|\textit{two deaths})}{P(\textit{natural}|\textit{two deaths})}$
- Putting it all together, our set up is this:

$$\frac{P(\textit{kill twice}|\textit{two deaths})}{P(\textit{natural}|\textit{two deaths})} = \frac{P(\textit{two deaths}|\textit{kill twice})}{P(\textit{two deaths}|\textit{natural})} \times \frac{P(\textit{kill twice})}{P(\textit{natural})}.$$

QUESTION: Could we have set up things differently, say by comparing different hypotheses?

Plug in the numbers

- The likelihood ratio equals one. If the babies died of natural causes or because they were killed, they would be found dead with equal probability. So $\frac{P(\textit{two deaths}|\textit{kill twice})}{P(\textit{two deaths}|\textit{natural})} = 1$.
- What makes a difference are the prior odds $\frac{P(\textit{kill twice})}{P(\textit{natural})}$. We assumed the probability that the sons died of natural causes is 1 in 100 million, so $P(\textit{natural})$ =1 in 100 million.
- What is the prior probability that a mother kills both her sons, $P(\textit{kill twice})$? If in a mid-size country like the UK 1 million babies are born every year of whom 100 are murdered by their mothers, he chance that a mother kills one baby in a year is 1 in 10,000. Assuming independence, the chance that a mother kills *two* sons equals 1 in 100,000,000 (why?).
- So prior odds $\frac{P(\textit{kill twice})}{P(\textit{natural})}$ equal 1.

Posterior probability of guilt

- Since both prior odds and likelihood ratio equal one, the posterior odds $\frac{P(\textit{kill twice}|\textit{two deaths})}{P(\textit{natural}|\textit{two deaths})}$ must equal one, a value clearly insufficient for a conviction. (Why?)
- To have greater posterior odds, the prior probability of *natural* must be much lower than the prior probability of *kill*. With a fixed likelihood ratio of one, we have:

$P(\textit{kill twice})$	$P(\textit{natural})$	Posterior Odds
1 in 100 million	1 in 100 million	1
1 in 100 million	1 in 1 billion	10
1 in 100 million	1 in 10 billion	100

QUESTION: What is the difference between posterior odds and posterior probability?

EXERCISE: Repeat calculation using a different set up, as follows:

$$\frac{P(\textit{kill once}|\textit{two deaths})}{P(\textit{natural}|\textit{two deaths})} = \frac{P(\textit{two deaths}|\textit{kill once})}{P(\textit{two deaths}|\textit{natural})} \times \frac{P(\textit{kill once})}{P(\textit{natural})}.$$

The hypothesis *kill once* should be understood as: Sally Clark killed one of her two children and the other died of natural causes.