“Your Honor, this was not a coincidence!”

On the (ab)use of statistics in the case against Lucia de B.

Peter Grünwald
CWI, Amsterdam

The Case of Lucia de Berk

- In 2004, the court of appeals convicted Lucia de Berk, a nurse from the Juliana Children’s Hospital in the Hague, to life imprisonment for 7 murders and 3 murder attempts.
- De B. has been in prison for more than 5 years now, but has never confessed.
- The case is now under review of the “Committee Posthumus-II”

2001 (september 16th) Juliana hospital notifies the police

- –There were many more “incidents” during her shifts than during those of other nurses in her ward
- –Statistician calculates that the probability that something like this would happen by pure coincidence, is less than **1 in 342 million**
- –Trial gets (more than) substantial media attention

2003 Court convicts L. of 4 murders, 3 attempts

- –Statistics is crucial part of the evidence
- –In appeal, experts for defense R. Meester (probabilist) and M. van Lambalgen (logician) claim calculation is flawed

2004 Court of appeals convicts L. of 7 murders,3 attempts

- –Court says that this time, “statistics in the form of probability calculations has not been used”
- –In fact, **the court’s report has flawed statistics written all over it**

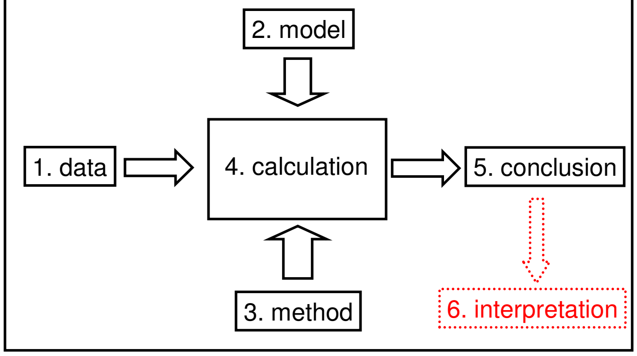
2006 Prof. **Ton Derksen** writes a book about the case

- –Case goes to “Posthumus-II committee”
- –**Richard Gill** and myself write letter to the committee supporting Derksen’s statistical analysis

Menu

1. The use of statistics
   - What the statistician did (evidence in first verdict)
   - What went wrong (everything)
   - The role of statistics in the final verdict
2. Can we do better?

The Statistician's Work



1. data
2. model
3. method
4. calculation
5. conclusion
6. interpretation

1. The Data - I

Juliana Hospital MCU-1, Oct 1 2000 – Sep 9 2001	no incident	incident	total
Nr of Shifts with L present	134	8	142
Nr of Shifts with L not present	887	**0**	887
Total Number of Shifts	1021	8	1029

Data from the Juliana Hospital Medium Care Unit-1, where suspicion first arose. "Incident" is sudden death or reanimation with no clear explanation

1. The Data - II

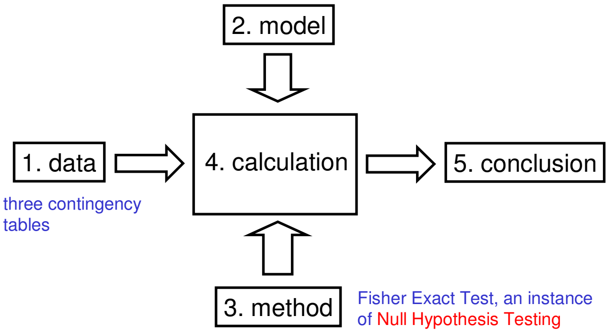
RKZ unit 42, 6/8 – 26/11 1997	no incident	incident	total
Nr of Shifts with L present	53	5	58
Nr of Shifts with L not present	272	9	281
Total Number of Shifts	325	14	339

Data from the RKZ (Red Cross) Hospital, unit 42

RKZ unit 43, 6/8 – 26/11 1997	no incident	incident	total
Nr of Shifts with L present	0	1	1
Nr of Shifts with L not present	361	4	365
Total Number of Shifts	361	5	366

Data from the RKZ (Red Cross) Hospital, unit 43

Homogeneity: For each nurse, the probability that an incident occurs during his/her shift is the same $p \in [0, 1]$



2. model
1. data
4. calculation
5. conclusion
3. method

three contingency tables

Fisher Exact Test, an instance of Null Hypothesis Testing

2. & 3. The Model and Method

- Statistician tested null hypothesis

  H0: “Lucia has same incident probability as other nurses”

  against the alternative

  H1: “Lucia has higher incident probability”

  using a standard test with a significance level of 1 in 10000.

2. & 3. The Model and Method

- Statistician tested null hypothesis

  H0: “Lucia has same incident probability as other nurses”

  against the alternative

  H1: “Lucia has higher incident probability”

  using a standard test with a significance level of 1 in 10000.
- i.e. he chooses some test statistic (a function of the data) T such that, as t increases, $\text{Pr}_{H_0}(T \geq t)$ goes to 0.
- If the actually observed data t_{obs} is so extreme that

  $$\text{p-value} := \text{Pr}_{H_0}(T \geq t_{\text{obs}}) \leq \frac{1}{10000}$$

  then one “rejects” the null hypothesis.

The Method: Fisher’s Exact Test

- Statistician used Fisher’s Exact test
  - a “conditional” test with test statistic

    $T = \#\text{incidents with Lucia present}$
  - For convenience, define

    $$\begin{array}{ll} \text{Pr}^*(T \geq \underset{\overset{\uparrow}{8}}{t}) := \text{Pr}_{H_0}(T \geq t \mid & \#\text{shifts with Lucia}, \longleftarrow 142 \\ & \#\text{shifts total}, \longleftarrow 1029 \\ & \#\text{incidents}) \longleftarrow 8 \end{array}$$

The Method: Fisher’s Exact Test

- Statistician used Fisher’s Exact test
  - a “conditional” test with test statistic

    $T = \#\text{incidents with Lucia present}$
  - For convenience, define

    $$\begin{array}{ll} \text{Pr}^*(T \geq \underset{\overset{\uparrow}{8}}{t}) := \text{Pr}_{H_0}(T \geq t \mid & \#\text{shifts with Lucia}, \longleftarrow 142 \\ & \#\text{shifts total}, \longleftarrow 1029 \\ & \#\text{incidents}) \longleftarrow 8 \end{array}$$

  - We reject “Lucia is like the others” if $\text{Pr}^*(T \geq 8) < 1/10000$
  - We find

    $\text{Pr}^*(T \geq 8) = \text{Pr}^*(T = 8) = .00000011057 \approx \text{1 in 9 million}$

Fisher’s Exact Test: Interpretation

- View nurse’s shifts as balls in an urn
  - Pr^* follows a hypergeometric distribution
- There are 1029 balls (shifts). 8 are black (incidents), the rest white
- We draw 142 balls without replacement from the urn (Lucia’s shifts)
- It turns out that all 8 black balls are among these 142
- $\text{Pr}^*(T \geq 8) \approx$ 1 in 9 million is the probability that this happens.

$$\text{Pr}^*(T = t) = \frac{\binom{\#\text{shifts Lucia}}{t}\binom{\#\text{shifts others}}{\text{total}\#\text{incidents}-t}}{\binom{\#\text{shifts total}}{\text{total}\#\text{incidents}}}$$

4. The Calculation

- Applying Fisher’s test to the Juliana data (first table) gives a *p*-value of 1 in 9 million

$$p := \text{Pr}^*(T \geq 8) \approx \frac{1}{9 \cdot 10^6}$$

- Classical Problem: Same data that was used to suggest hypothesis was also used to test it
- Statistician recognizes this and performs a post-hoc correction, by considering the H0-probability that some nurse in Lucia’s ward experienced a pattern as extreme as Lucia’s

$$p' := \text{Pr}^*(\text{there exists } j \in \{1, \ldots, 27\} \text{ s.t. } T_j \geq t_j) \approx 27\,\text{Pr}(T \geq 8) \approx \frac{3}{10^6}$$

4. The Calculation

- Another complication:There are three tables, and hence three *p*-values. How to combine these?
  - Using the fact that the data are independent, statistician combines them into one *p*-value by multiplying:

$$p_{\text{new}} :=$$

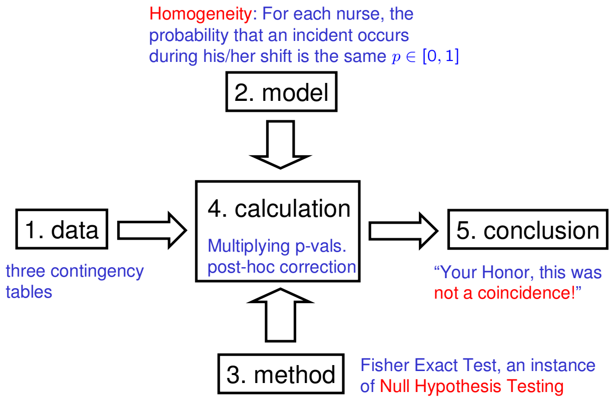
$$\text{Pr}^*\Big(\text{exists } j \in \{1, \ldots, 27\} \text{ s.t. } T_j^{(1)} \geq t_j^{(1)} \;\&\; T^{(2)} \geq t^{(2)} \;\&\; T^{(3)} \geq t^{(3)}\Big)$$

$$= p'_1 \cdot p_2 \cdot p_3 = \text{1 in 342 million}$$

$$p'_1 := 27\text{Pr}^*(T^{(1)} \geq 8)$$

$$p_2 := \text{Pr}^*(T^{(2)} \geq 5)$$

$$p_3 := \text{Pr}^*(T^{(3)} \geq 1)$$



Homogeneity: For each nurse, the probability that an incident occurs during his/her shift is the same $p \in [0, 1]$

2. model

1. data

three contingency tables

4. calculation

Multiplying p-vals.
post-hoc correction

5. conclusion

“Your Honor, this was not a coincidence!”

3. method

Fisher Exact Test, an instance of Null Hypothesis Testing

5. The Conclusion

- Statistician chose a significance level of 1 in 10000
- He observed a *p*-value of 1 in 342 million
- Therefore he rejects the null hypothesis
  "Lucia has same incidence probability as the others"
- Statistician explicitly mentions the *p*-value 1 in 342 million, and translates "rejection of null" into
  "your honor, this was not a coincidence!"

The Conclusion - II

- Statistician does add a very explicit warning that this does *not* imply that Lucia is a murderer!
- He explicitly lists five alternative explanations:
  1. Lucia prefers to work together with another nurse. That nurse is really causing the incidents
  2. Lucia often does the night shift, during which more incidents happen
  3. Lucia is, quite simply, a bad nurse
  4. Lucia prefers to take on the most ill patients
  5. Somebody hates Lucia and tries to discredit her

What Went Wrong
(Everything!)

1. The Data

- Derksen has uncovered evidence that data were gathered in a strongly biased manner
  1. Selection bias in choice of hospitals/wards
  2. Suspect-driven search
  3. Normative and fluctuating definition of "incident"
  4. Additional "epidemiological" data that suggest Lucia is innocent, was ignored

1. Hospital Selection Bias

- Possible bias in choice of hospital
  - Juliana Hospital: MCU-1 (table 1) and MCU-2 were adjacent, connected by a swing door
  - Lucia also worked in RKZ (table 2 and 3) and two other hospitals
  - Prosecutor tried to get L. convicted for some cases in MCU-2 and the two other hospitals as well
  - Yet no tables from these hospitals have been used…

2. Suspect-Driven Search

- In RKZ and the two other hospitals, explicit evidence that the search was suspect-driven
  - More thorough search for incidents when she was present than for incidents when she wasn't

  “We were asked to make a list of incidents that happened during or shortly after Lucia's shifts”
- In JKZ, an attempt was made to be “objective”, but
  - There was no record of reanimations. Doctors and nurses were asked whether they remembered such “incidents”. Everybody knew why they were being asked…

3. Definition of “incident”

- “incident” was first defined as:

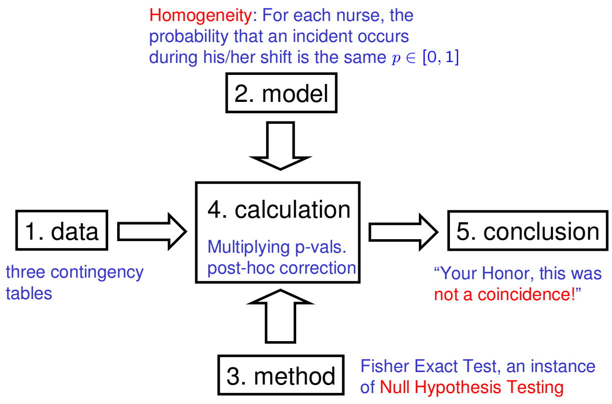
  a patient suddenly dies or needs reanimation
- Later the court changes this to

  a patient suddenly dies with no clear explanation, or reanimation is suspicious, i.e. without clear explanation
- This means that some sudden deaths and reanimations were not listed in the tables, because they were in no way suspicious
- All the people who have to report ‘suspicious incidents' know that they are asked because Lucia may be a serial killer

There is a considerable risk that “incident is suspicious” effectively becomes synonymous to “Lucia is present” (Van Zwet)

4. Highly Relevant Additional Data

- The statistician and the court ignored the following data that were available from the start:
  - From 1996-1998 (before Lucia worked there), there were seven deaths in her ward.
  - From 1999-2001 (when Lucia worked there), there were six deaths
- Less people die when there's a serial killer around!
- Also percentage of deaths in Lucia's ward compared to total number of deaths in hospital was lower than average while Lucia worked there (12.8% vs 16.6%)



Homogeneity: For each nurse, the probability that an incident occurs during his/her shift is the same $p \in [0,1]$

2. model

1. data

three contingency tables

4. calculation

Multiplying p-vals.
post-hoc correction

5. conclusion

“Your Honor, this was not a coincidence!”

3. method

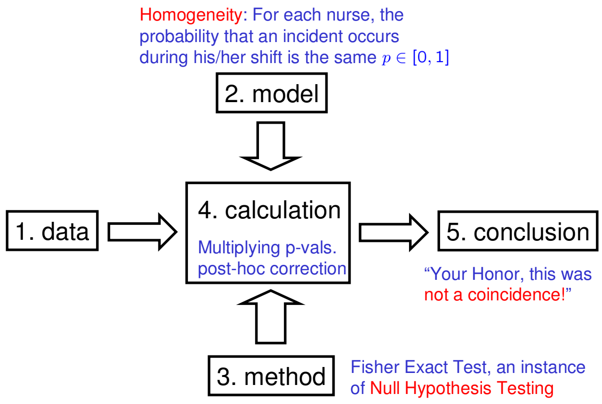
Fisher Exact Test, an instance of Null Hypothesis Testing

The Calculation

- Statistician combines three independent tests by *multiplying* the three *p*-values
- This is a mistake!
  - In this way, if you worked in 20 hospitals with a similar harmless incident pattern in each of them (say, a *p*-value of .5) you still end up with final *p*-value $(0.5)^{20} < 1/10^6$
  - if you change hospital a lot, you automatically become a suspect
  - Something close to .5 would be more reasonable!
- Various alternative, correct methods exist. A standard method such as Fisher’s gives a *p*-value that is a factor 300 larger

The Method (and main issue)

- Even when combining *p*-values in a correct manner, final *p*-value (slightly) is smaller than 1 in 10000…so may we conclude “no coincidence” after all?
- **NO:**
  Neyman-Pearson style Null Hypothesis Testing cannot be used if the same data is used both for suggesting and testing a hypothesis. The results are essentially meaningless and there is *no* way a post-hoc correction can correct for this!
  - This will be explained in detail in final part of talk



Homogeneity: For each nurse, the probability that an incident occurs during his/her shift is the same $p \in [0,1]$

2. model

1. data

4. calculation

Multiplying p-vals.
post-hoc correction

5. conclusion

“Your Honor, this was not a coincidence!”

3. method

Fisher Exact Test, an instance of Null Hypothesis Testing

The Model

- The assumption that there is no variation between ordinary nurses is wrong (but defensible – see later on).
- Following Lucy and Aitken (2002), A. de Vos and R. Gill propose the following model:
  - The nr of incidents witnessed by a nurse, is Poisson distributed with some parameter μ
  - For each nurse, μ is drawn independently from some distribution
  - Allows for **innocent heterogeneity** (e.g. clusters of shifts in time, caused by different vacation patterns, and so on)

The Model

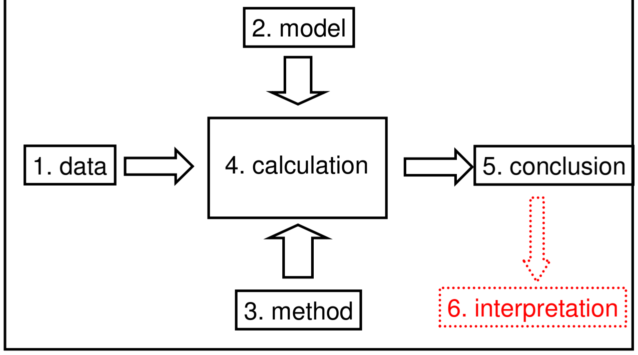
- The assumption that there is no variation between ordinary nurses is wrong (but defensible – see later on).
- Following Lucy and Aitken (2002), A. de Vos and R. Gill propose the following model:
  - The nr of incidents witnessed by a nurse, is Poisson distributed with some parameter μ
  - For each nurse, μ is drawn independently from some distribution
  - Allows for **innocent heterogeneity** (e.g. clusters of shifts in time, caused by different vacation patterns, and so on)
- With this model, the *p*-value **increases dramatically**
  - combined with a correct combination method for the three tests and presumably correct data, it becomes as large as 1 in 9

5. The Conclusion

Here we are counterfactually assuming that calculations were correct in the first place

5. The Conclusion

- Statistician might have warned that the conclusion is extremely sensitive to the data being 100% correct
- Statistician might have pointed out that conclusion “this is not a coincidence” depends on the chosen model
  - he does this to some extent, though
- Van Zwet, grand old man of Dutch statistics:
  - In statistical consulting, it is bad practice to just write down your conclusions and say “the rest is up to you”
  - Customer usually doesn’t realize that statistical conclusions are model-dependent
  - At the CWI, consultants used to have a **veto** on the further (re-) formulation of their conclusions by the customer



6. The Interpretation - I

- Report accompanying verdict in court of appeals:

11.13 “Not a single reasonable explanation has been found of the fact that the suspect was involved in so many deaths and life threatening incidents in such a short period”

  - The report goes on to discuss several alternative explanations that might be given. *These are exactly the same alternative explanations as those given by statistician.* (The statistician included them as mere examples, the court views them as an exhaustive list…)
  - The court dismisses each of these, and then (implicitly) concludes that Lucia’s incident pattern is a strong indication that Lucia caused the incidents

The Interpretation - II

- One of the “murders” for which Lucia has been convicted concerned the death, in 1997, of a 73-year old woman suffering from terminal cancer
- In 2004, 6 medical experts testify regarding her death
  - 5 say it was natural
  - 1 (the one who in 1997 had given the ‘natural death certificate’) says: “at the time I thought it was a natural death, but, given all these other cases reported by the media, I now think it was unnatural”
- The court follows the single dissenting expert who has implicitly used statistics!

What Went Wrong - Summary

1. Wrong data
2. Wrong model
3. Wrong method
4. Wrong calculation
5. Wrongly worded conclusion
6. Wrong interpretation of conclusion

I must add that the statistician was only involved in a few of the mistakes, and that nobody has any doubts concerning his integrity. Indeed (in contrast to medical experts) he is willing to publicly discuss all these issues, which is highly laudable

Menu

1. The use of statistics
    - What the statistician did (evidence in first verdict)
    - What went wrong
    - The role of statistics in the final verdict
2. Can we do better?

   I will only speak about “the method”, assuming that the data are correct

Small probability events happen!

- Report based on intuition that, when we observe something with “incredibly small probability”, this is a strong indication that something funny is going on
- Yet incredibly improbable things happen all the time
    - I met a good friend from high school in a coffee house in Marrakech
    - Sally Clark passes away just when I go to the UK to talk about a similar case
- The reason is, quite simply, that very many things can happen. If all these things are equally likely, they must all have very small probability. So whatever actually happens, will have very small probability.

- The fact that

  “something with incredibly small probability happened”

  is totally insufficient to conclude

  “this is most probably not a coincidence”
- To gain evidence that warrants such a conclusion, we need more. Two ways to get that:
    1. Use a (Neyman-Pearson) hypothesis test
    2. Incorporate prior probabilities and use Bayes’ rule

Neyman-Pearson Testing

- The idea is to identify, *before seeing the data*, a definite event with probability smaller than 1/10000
    - If that event happens, you reject the null hypothesis
    - If you have already seen the data before you decided on your event, this only works if you do an additional experiment to gain additional data

Example:

Somebody calls the newspaper and says that he bought a die that has magically landed 6 the last 10 times he threw it. Even if he is telling the truth, this doesn’t strongly indicate that the die is loaded

But if somebody **predicts** that he will throw 10 sixes in a row, and then this indeed happens, this does give a strong indication that the die is loaded

Neyman-Pearson Guarantee

- NP Testing has been designed such that, if it is performed repeatedly (and **correctly**!), then the following guarantee holds:
- on average, at most 1 in 10000 times that we do a NP test, we say “null hypothesis rejected” even though null hypothesis is true

$$\Pr_{H_0}(\text{I say ”reject”}) \leq \frac{1}{10000}$$

$$\Pr_{H_0}(\text{p-value} \leq \frac{1}{10000}) \leq \frac{1}{10000}$$

Neyman-Pearson Guarantee

- In Lucia’s case, statistician effectively promises that, if his method is used repeatedly, then at most 1 in 10000 times one would say “not a coincidence”, whereas in truth, it was just a coincidence
  - *But what does **‘used repeatedly’** mean here?*
- Do we say ‘coincidence/no coincidence’
  - Each time a nurse at Juliana hospital has a suspicious incident pattern?
  - Each time a nurse in the Netherlands/Europe/the world has a suspicious incident pattern?
  - Each time *a court case involving a statistical test is held?*

Neyman-Pearson Guarantee

- There is no way that the statistician can live up to his promise of “being right most of the time”
  - the post-hoc correction factor he has to apply depends on unknowable aspects of the problem
- If we say ‘coincidence/no coincidence’ each time a court case involving a statistical test is held, then,
  - in order to properly “correct” for the reuse of data, we would need to know the exact circumstances that induce a court case involving a statistical test
  - In any case, the correction factor that was actually used is many orders of magnitude too small

Separate Data

- Neyman-Pearson tests cannot be used directly in L.’s case, since there is no “correction” that gives the 1 in 10000 guarantee. For this, we need separate data
- In L.’s case, hypothesis was suggested only by the Juliana data (first table). Thus, we could have used the second and third table as an independent data set for testing (proposed by Gill et al.)
- essentially a sound approach, even though some grave problems remain

NP approach is unconditional

- Even in a population in which everybody is innocent and all patterns are coincidental, a NP approach will sometimes lead to the conclusion “not a coincidence”, suggesting “guilt”.
  We **guarantee** that
  $$\text{"Pr"}(\text{it is coincidence } \textbf{and} \text{ we say it's not}) < \frac{1}{10000}$$
  but we are more **interested** in
  $$\text{"Pr"}(\text{it is coincidence} \mid \text{we say it's not}) < \ldots$$
- I think in many cases the second “probability” is unobtainable, and then the first is still useful in court
  - Significance level should be put much higher though!
  - Several highly intelligent people disagree with me

NP vs. Bayes

- NP *if used correctly*
  - + Gives a (calibration-type) **performance guarantee**
  - +/- Is **unconditional**
  - \- Does not confront main issue (guilty vs innocent)
  - \- Can only be used if **independent data** can be obtained (in an ethical manner). We throw away potentially useful data….
- Bayes
  - + Is conditional
  - + Confronts main issue
  - \- Can only be used with a reasonable degree of intersubjectivity if it has already been established that crimes have been committed

Bayesian Approach

- Automatically compensates for selection bias by taking prior probabilities into account in mathematically sound way (Bayes rule)
  - Reuse of data is not a problem
- Confronts the real problem of interest.
  In a (simplistic) application, we would test
  H0: Lucia has same incident probability as other nurses
  against the more interesting alternative
  H1’: Lucia is murderous
- We calculate conditional probability
  $$\Pr(H_1' \mid D) = \Pr(\text{Lucia murderous} \mid \text{data})$$

Bayes rule

$\pi = \text{prior}$

$$\frac{\Pr(H_0 \mid D)}{\Pr(H_1' \mid D)} = \frac{\Pr(D \mid H_0) \cdot \pi(H_0)}{\Pr(D \mid H_1') \cdot \pi(H_1')}$$

$$\geq \frac{\Pr(D \mid \theta_L = \theta_{NL}) \cdot \pi(H_0)}{\Pr(D \mid \hat{\theta}_L(D), \hat{\theta}_{NL}(D)) \cdot \pi(H_1')} \approx \frac{\Pr(D \mid \hat{\theta}(D)) \cdot \pi(H_0)}{\Pr(D \mid \hat{\theta}_L(D), \hat{\theta}_{NL}(D)) \cdot \pi(H_1')}$$

$$= \frac{\left(\frac{8}{1029}\right)^8 \left(\frac{1021}{1029}\right)^{1021}}{\left(\frac{887}{887}\right)^{887} \left(\frac{8}{142}\right)^8 \left(\frac{134}{142}\right)^{134}} \cdot \frac{\pi(H_0)}{\pi(H_1')} \approx \frac{1}{9 \cdot 10^6} \frac{\pi(H_0)}{\pi(H_1')}$$

Likelihood ratio of H0 to H1’ is 1 in 9 million. Thus, for example, if 1 in 100001 nurses is murderous $(\pi(H_0) = 10^5 \pi(H_1))$, then posterior odds are 1 in 90. Not beyond a reasonable doubt!

A Problem

- What facts about Lucia should one take into account when determining the prior probability that she is murderous?
  - Lucia is a typical European nurse
  - Lucia is a typical Dutch nurse
  - Lucia is a typical female Dutch nurse
  - Lucia is a typical (?) female Dutch nurse who used to be a prostitute
  - Lucia is a typical female Dutch nurse who used to be a prostitute, faked her high school certificates, has an alcoholic father and has the ambition to write a whodunnit

Serial Killer Profiles

- It is unclear what to condition on when determining the prior
  - Similar to the NP testing post-hoc problem: it is unclear what is the relevant subpopulation
- This may seem a moot point
  - a reasonable person might say: every reasonable person will put the prior probability very small
  - But what if the prosecution finds a psychologist who testifies that Lucia's personality and history exactly fit the serial killer profile (this almost happened…)
  - i.e., the prior probability that she's a serial killer is substantially higher

Robust Bayesian Approach

- The judge may believe the psychologist
- The defense may feel cheated
- It seems safer to let both the defense and the prosecution produce experts who both suggest a prior, $\pi_{\text{def}}(H'_1)$ and $\pi_{\text{pro}}(H'_1)$ respectively. If the judge thinks both experts are reasonable people, she should consider a prior interval $\Pi(H'_1) = [\pi_{\text{def}}(H'_1), \pi_{\text{pro}}(H'_1)]$
- Given the data, she then ends up with a posterior interval $[\Pr_{\text{def}}(H'_1 \mid D), \Pr_{\text{pro}}(H'_1 \mid D)]$

Robust Bayesian Approach

- Judge may of course end up simply with interval $[0, 1]$
- This may not seem helpful, but at least
  - it's safe and nonarbitrary
  - It may be helpful after all in "reasoning towards innocence" (there is reasonable doubt of guilt)
- I think it's often the right thing to do; therefore I think a NP approach on separate data can be helpful as well

Counterarguments

1. “Every rational decision maker must be Bayesian”
2. “Expert witnesses shouldn’t talk about priors, only about likelihood ratios. The prior is the judge’s jurisdiction”

- I don’t agree. A judge has to come up with a posterior probability/decision that most informed people will be able to understand, and also find “reasonable”
- We need “intersubjective acceptance” of the judge’s priors

Bayes, again

- Bayes is method of choice, in, i.e. use of DNA-related evidence in court. Do I think this is all wrong?
- NO: in contrast to the Lucia case,

it is usually clear that a crime has been committed

Bayes, again

- Bayes is method of choice, in, i.e. use of DNA-related evidence in court. Do I think this is all wrong?
- NO: in contrast to the Lucia case,

it is usually clear that a crime has been committed

  - Now Bayes seems much less problematic
  - In a remote village with 10000 inhabitants, an old lady was stabbed to death. It may be reasonable to state that the prior probability that inhabitant X is the murderer, is 1 in 10000: (somebody in the village **must** have done it!)
  - Van Zwet even proposes that statistics should only be used in court if it is 100% sure that a crime has been committed

Conclusions – 1 (of 2)

- Bayesian thought experiments should always be performed
  - “what happens if the prior/population rates were this and this…”
  - The court’s report features negligence of prior probabilities all over the place
- This leads to a “robust Bayesian” approach
- Nonrobust Bayesian reasoning can be quite arbitrary **unless it is clear that a crime has been committed**
- NP approach on additional data may be helpful
  - despite numerous problems

Final Remark: Role of Statisticians

- Everybody except Van Zwet and Derksen (but including me) took the data for granted!
- "Statisticians don't speak with one voice"
  - While they completely disagree on the details, almost all statisticians do strongly agree that there was much less evidence against L. than reported by the court's statistician
  - This crucial point got completely lost in the debate
- Neither judges nor public nor Bayesian statisticians nor frequentist statisticians seem to understand that probabilistic statements are **meaningless** if, even in idealized circumstances, they do not allow you to make predictions that improve on random guessing

Thank you for your attention!

More information in English can be found on
Richard Gill's homepage www.math.leidenuniv.nl/~gill