UCL

The Role of Model Assumptions in Statistics

Christian Hennig

15 June 2017

1 "The Death of Homeopathy"

A standard random effects meta-analysis

From Shang et al. "Are the clinical effects of homoeopathy placebo effects? Comparative study of placebo-controlled trials of homoeopathy and allopathy" (The Lancet, 2005)

Famous study,
prompted Lancet-editorial "The Death of Homeopathy".

$$L_i = \theta + \eta_i + e_i, \quad i = 1, \ldots, n,$$
$$\eta_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2) \text{ independent.}$$

$$L_i = \theta + \eta_i + e_i, \quad i = 1, \ldots, n,$$
$$\eta_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2) \text{ independent.}$$

$n = 8$. For trial i : p_i emp. prob. that "placebo works", q_i emp. prob. that "homeopathy works", $L_i = \log \frac{p_i/(1-p_i)}{q_i/(1-q_i)}$,

$$L_i = \theta + \eta_i + e_i, \quad i = 1, \ldots, n,$$
$$\eta_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2) \text{ independent.}$$

$n = 8$. For trial i : p_i emp. prob. that "placebo works", q_i emp. prob. that "homeopathy works", $L_i = \log \frac{p_i/(1-p_i)}{q_i/(1-q_i)}$, L_i negative $\Rightarrow$ homeopathy better than placebo

$$L_i = \theta + \eta_i + e_i, \quad i = 1, \ldots, n,$$
$$\eta_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2) \text{ independent.}$$

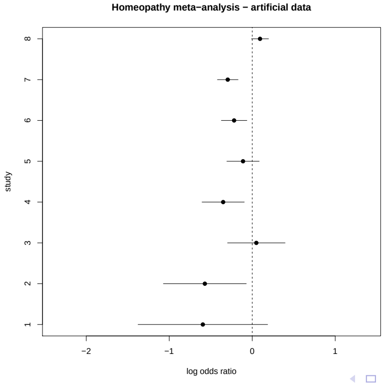
$n = 8$. For trial i : p_i emp. prob. that "placebo works",
q_i emp. prob. that "homeopathy works", $L_i = \log \frac{p_i/(1-p_i)}{q_i/(1-q_i)}$,
L_i negative $\Rightarrow$ homeopathy better than placebo
Aim: evidence for/against $\theta = 0$,
interpreted as "homeopathy equals placebo"?

$$L_i = \theta + \eta_i + e_i, \quad i = 1, \ldots, n,$$
$$\eta_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2) \text{ independent.}$$

$n = 8$. For trial i : p_i emp. prob. that "placebo works",
q_i emp. prob. that "homeopathy works", $L_i = \log \frac{p_i/(1-p_i)}{q_i/(1-q_i)}$,
L_i negative $\Rightarrow$ homeopathy better than placebo
Aim: evidence for/against $\theta = 0$,
interpreted as "homeopathy equals placebo"?

η_i trial specific effect, e_i within trial variation.
σ_i standard error of L_i (small if n_i large),

$$L_i = \theta + \eta_i + e_i, \quad i = 1, \ldots, n,$$
$$\eta_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2) \text{ independent.}$$



$$L_i = \theta + \eta_i + e_i, \quad i = 1, \ldots, n,$$
$$\eta_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2).$$

Shang et al. find
overall estimate $\hat{\theta} = -0.13$,
95%-confidence interval $[-0.43, 0.17]$ for θ,
conclude (because 0 is in CI)
"no evidence for homeopathy better than placebo."

$$L_i = \theta + \eta_i + e_i, \quad i = 1, \ldots, n,$$
$$\eta_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2).$$

Model assumptions:

1. Independence of trials, η_i and e_i
2. Additive model for L_i.
3. Normal distribution for η_i.
4. Normal distribution for e_i.

$$L_i = \theta + \eta_i + e_i, \quad i = 1, \ldots, n,$$
$$\eta_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2).$$

How to check whether these are "true"?
...or at least "approximately true",
"appropriate", "helpful" etc.

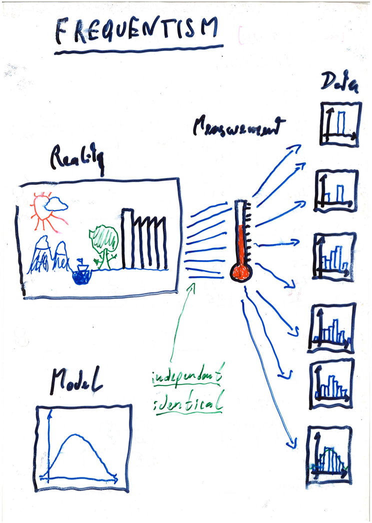
2. Frequentist model assumptions

(more general, “aleatory”, incl. propensities etc.)

2.1 What do the model assumptions mean?

“We think of the situation as . . . ”

- Potentially infinite repetition (of experimental conditions)
- $P(A)$: relative frequency limit of occurrence of A
  (e.g., normal distribution is defined by $P(A)$ $\forall A$.)


FREQUENTISM
Data
Measurement
Reality
Model
independent
identical

2. Frequentist model assumptions

(more general, "aleatory", incl. propensities etc.)

2.1 What do the model assumptions mean?

"We think of the situation as ..."

- Potentially infinite repetition (of experimental conditions)
- $P(A)$: relative frequency limit of occurrence of A
  (e.g., normal distribution is defined by $P(A)\ \forall A$.)

This is obviously an idealisation -
what constitutes a "repetition"?

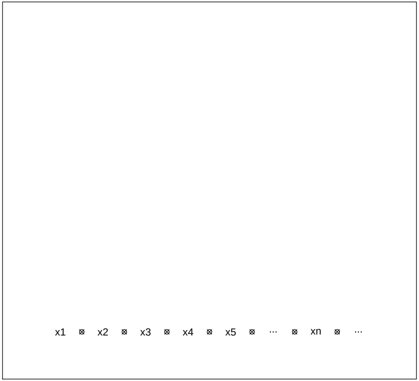
"Whatever can be distinguished cannot be identical."
(B. de Finetti)

2.2 Independent repetitions (“i.i.d.”)

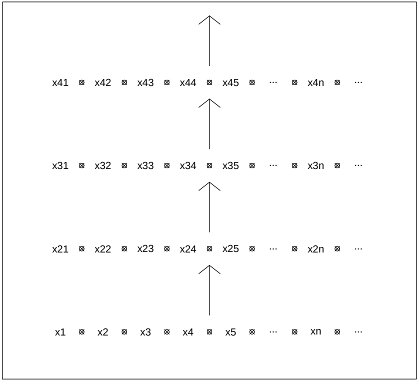
Frequentism relies on “repetitions of experiments”,
e.g. results from different patients in clinical study.

Possible to *define* i.i.d. in terms of probabilities
(in order to get results about distribution of L_i
from independent patients)
but this cannot justify defining probabilities in terms of
i.i.d. repetitions.

In order to define “i.i.d.” sequences, i.i.d. repetitions and defining repetitions are required on different levels.



In order to define "i.i.d." sequences, i.i.d. repetitions and defining repetitions are required on different levels.



In practice, there’s only one level of repetition.

The effective sample size for
(in-)dependence assumptions is usually 1.

But independent repetition *of some kind*
is always required
in order to learn from data.

In practice, there's only one level of repetition.

The effective sample size for
(in-)dependence assumptions is usually 1.

But independent repetition *of some kind*
is always required
in order to learn from data.

Independent repetition is constructed by conscious decision
to ignore potential dependencies and differences.

2.3 Can frequentist model assumptions be checked?

Goodness-of-fit/misspecification tests

If something modelled as very unlikely happens,
the model is interpreted to be *falsified*.

2.3 Can frequentist model assumptions be checked?

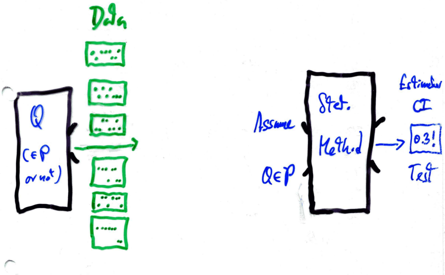
Goodness-of-fit/misspecification tests

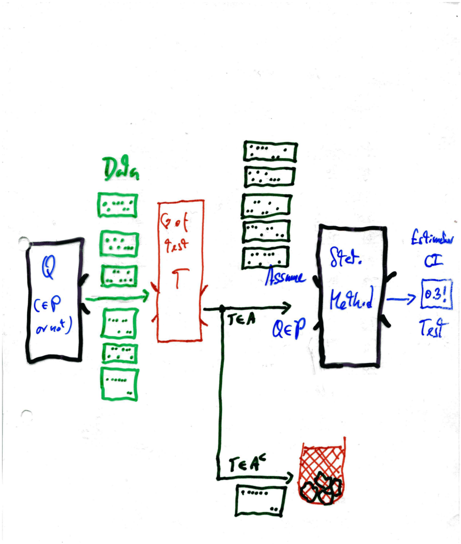
If something modelled as very unlikely happens,
the model is interpreted to be *falsified*.

For example, could “falsify” independence
from data where patients examined on same day
have similar results.

2.4 The goodness-of-fit paradox (H, 2007)

Checking the model assumptions violates them automatically
because the *possibility* of unlikely events
is constitutive part of the models.


Data
Q
(∈P
or not)
Stat.
Method
Assume
Q∈P
Estimator
CI
0.3!
Test

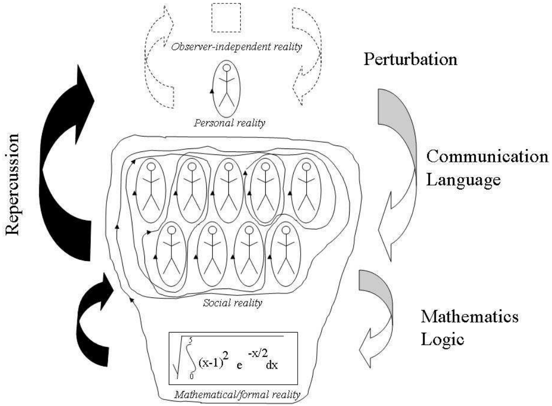

Data
GoF test
T
Q
(∈P
or not)
Assume
T∈A
Q∈P
Stat.
Method
Estimator
CI
0.3!
Test
T∈A^c

3. Mathematical models and reality

General mathematical modelling (broad and naive): mapping reality to mathematical objects.

Identification of items of perceived reality
with mathematical objects
and interpretation of results of mathematical operations
in terms of the items.

A constructivist view (H 2010, Foundations of Science)



Implications for mathematical modelling

- Science: establishing agreement in open exchange.

Implications for mathematical modelling

- Science: establishing agreement in open exchange.
- Mathematics is about creating a system that makes absolute agreement possible
  (but only *within mathematics*).

Implications for mathematical modelling

- Science: establishing agreement in open exchange.
- Mathematics is about creating a system that makes absolute agreement possible
(but only *within mathematics*).
- Mathematical modelling is not about how things are, but about how we think and communicate about them.
(Models communicate and change views of reality.)

Implications for mathematical modelling

- Science: establishing agreement in open exchange.
- Mathematics is about creating a system that makes absolute agreement possible
(but only *within mathematics*).
- Mathematical modelling is not about how things are, but about how we think and communicate about them.
(Models communicate and change views of reality.)
- It cannot be formally analysed how formal models are related to informal reality.

4. "Frequentism-as-model"

$$L_i = \theta + \eta_i + e_i, \quad i = 1, \ldots, n,$$
$$\eta_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2).$$

Not: is this (approximately) true?
But: can we learn what we want to learn
about our subject matter (homeopathy)
if we decide to look at the data "through" this model?

What does this view imply?

Is model consistent with our (agreed) perception of data?

(Frequentist) models are useful to. . .

- communicate researcher's perception of situation,
- inspire methodology,
- check quality of methodology
  in situations with known (made up) truth.
  (Tukey, Davies)

This is helped by. . .
Model-based mathematical theory:

- How well does method work under ideal conditions?
- How much variation to expect?
- Assume model, derive optimal method.
- Other properties of method (equivariance etc.)

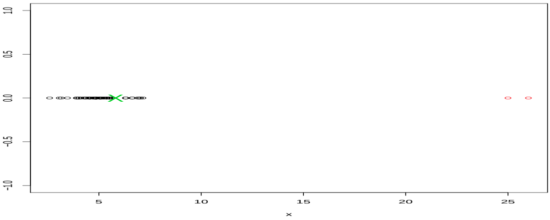
This is helped by...
Model-based mathematical theory:

- How well does method work under ideal conditions?
- How much variation to expect?
- Assume model, derive optimal method.
- Other properties of method (equivariance etc.)

Also informs about
effects of model deviations, e.g., data outliers.

Some violations of assumptions harm interpretation of results

Reasonable model checking:
Outliers not expected under $\mathcal{N}$,
affect $\mathcal{N}$-based statistics.



Some violations are *not* problematic
(e.g., g.o.f.-testing, often).

Some violations are *not* problematic
(e.g., g.o.f.-testing, often).

Re-formulation of "model checking":
Find out whether data could lead
statistical method (derived from model) astray.

Aspects of homeopathy meta-analysis

$$L_i = \theta + \eta_i + e_i, \quad i = 1, \dots, n,$$
$$\eta_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2) \text{ independent.}$$

"Trial effect" is treated as result from i.i.d. repetition,
$\Rightarrow$ "effective sample size" is $n = 8$, not $\sum n_i > 5,000$,
bad power, i.e.,
non-significance can easily happen under $\theta < 0$.

Aspects of homeopathy meta-analysis

$$L_i = \theta + \eta_i + e_i, \quad i = 1, \ldots, n,$$
$$\eta_i \sim \mathcal{N}(0, \sigma_0^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2) \text{ independent.}$$

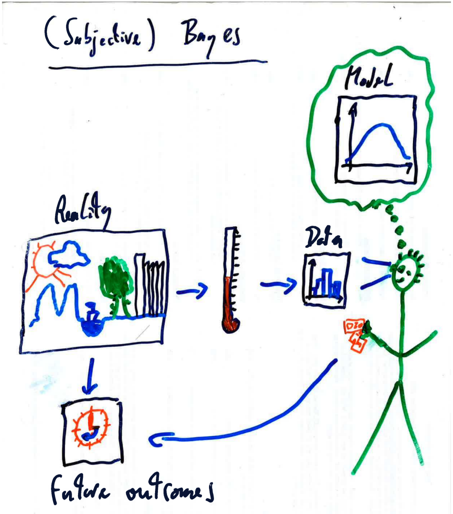
Furthermore, trials apply homeopathy differently;
only 2 trials treat "classical homeopathy".

Modelling all as i.i.d repetitions implies that
how homeopathy is applied is not "difference of interest".
$\Rightarrow$ "Classical homeopathy" paradigm
is not tested by this study.

5. Beyond classical frequentism

5.1 Bayesian philosophy and methodology

- Prior distribution on parameters θ
  allows posterior probability statements $P(\theta \in B) =?$
- Epistemological rather than aleatory probability
  (although in practice often ignored)


(Subjective) Bayes
Model
Reality
Data
Future outcomes

5. Beyond classical frequentism

5.1 Bayesian philosophy and methodology

- Prior distribution on parameters θ
  allows posterior probability statements $P(\theta \in B) =?$
- Epistemological rather than aleatory probability
  (although in practice often ignored)
- Requires very similar idealisation;
  ultimately same modeling issues as frequentism.

5.2 Nonparametrics/machine learning

- Often claim “no assumptions”,

5.2 Nonparametrics/machine learning

- Often claim "no assumptions",
- big datasets allow very flexible fitting, but. . .

5.2 Nonparametrics/machine learning

- Often claim "no assumptions",
- big datasets allow very flexible fitting, but. . .
- generalisation theory still assumes frequentism/i.i.d., (can be seriously harmed by bad data quality),

5.2 Nonparametrics/machine learning

- Often claim “no assumptions”,
- big datasets allow very flexible fitting, but. . .
- generalisation theory still assumes frequentism/i.i.d., (can be seriously harmed by bad data quality),
- implicit structural assumptions (e.g., class shapes, similarity).

6. Conclusions

- Decision whether a model is appropriate/helpful is quite different from checking whether it's true (the latter can't be done, not even approximately)

6. Conclusions

- Decision whether a model is appropriate/helpful is quite different from checking whether it's true (the latter can't be done, not even approximately)
- Major role of model assumptions: understanding methods under idealised conditions

6. Conclusions

- Decision whether a model is appropriate/helpful is quite different from checking whether it’s true (the latter can’t be done, not even approximately)
- Major role of model assumptions: understanding methods under idealised conditions
- Rather than “model check”: Will data lead method astray?