If all models are wrong, are the robust ones better?

The role of robust statistics in a constructivist philosophy of statistics

Christian Hennig

July 7, 2013

1. A constructivist philosophy of mathematical models

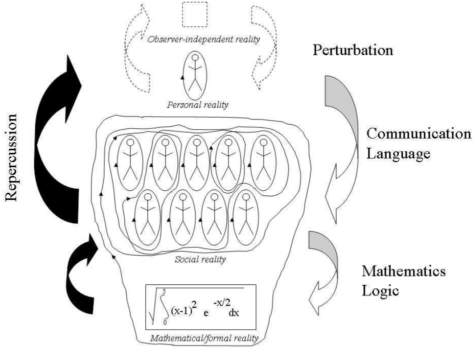
Most people know models are “wrong but sometimes useful”, but most philosophy of statistics ignores this (and many researchers do, too, when using them).

Missing: philosophy of mathematical modelling.

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

Note: in this sense, not only “$\mathcal{N}(\mu, \sigma^2)$”
is a mathematical model,
but also a breakdown point is a mathematical model of
what we think of as the “quality of robustness”.

Mathematics and reality
(H., 2010, Foundations of Science)



Mathematical modelling makes connections between mathematical social system and (agreed) perceptions.

Mathematical modelling makes connections between mathematical social system and (agreed) perceptions.

“Absolute agreement” within mathematics makes people believe that mathematical modelling can achieve agreement about reality.

But this relies on connection between formal models and informal reality, which is inaccessible to formal analysis.

- Mathematical models are items of social reality. They are a way of thinking and communicating about reality.

- Mathematical models are items of social reality. They are a way of thinking and communicating about reality.
- Models don't only refer to reality; also influence thoughts and communication, and change reality through it.

- Mathematical models are items of social reality. They are a way of thinking and communicating about reality.
- Models don't only refer to reality; also influence thoughts and communication, and change reality through it.
- Mathematical modelling is not about how things are,
  but about how we think about them.
  Models live in "world of mathematics"
  and cannot be "true" in informal reality.

- Mathematical models are items of social reality. They are a way of thinking and communicating about reality.
- Models don't only refer to reality; also influence thoughts and communication, and change reality through it.
- Mathematical modelling is not about how things are,
  but about how we think about them.
  Models live in "world of mathematics"
  and cannot be "true" in informal reality.
- Pragmatist attitude: what do we get out of it?

Relevant questions when discussing models

- What are the implications for modelling of not really believing in the “truth” of models?

Relevant questions when discussing models

- What are the implications for modelling of not really believing in the “truth” of models?
- How is perceived reality mapped on models? Do people agree about this reality?

Relevant questions when discussing models

- What are the implications for modelling of not really believing in the "truth" of models?
- How is perceived reality mapped on models? Do people agree about this reality?
- What aims can modelling help to achieve, and how are different modelling methods/strategies related to these aims?

Relevant questions when discussing models

- What are the implications for modelling of not really believing in the “truth” of models?
- How is perceived reality mapped on models? Do people agree about this reality?
- What aims can modelling help to achieve, and how are different modelling methods/strategies related to these aims?
- What are the implications for modelling of models influencing communication, thoughts, and reality?

2. Some thoughts on frequentism

Frequentism is an interpretation of probability;
connecting models to how we perceive reality.

2. Some thoughts on frequentism

Frequentism is an interpretation of probability;
connecting models to how we perceive reality.

“We think of the situation as . . . ”

- potentially infinite repetition (of experimental conditions),
- no systematic dependence between repetitions,
- $P(A)$: relative frequency limit of occurrence of A

"Infinite repetition" is obviously an idealisation.

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

Bayesians don't accept this but tend to ignore
the need of idealisation in their own use of models.

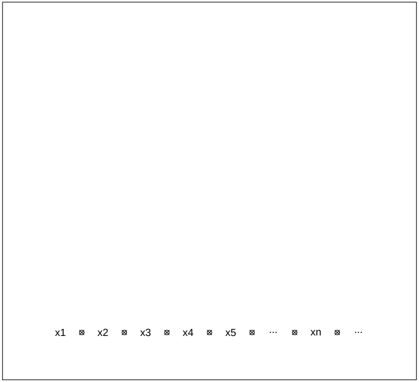
"Infinite repetition" is obviously an idealisation.

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

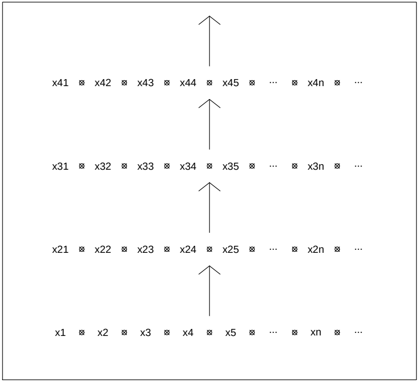
Bayesians don't accept this but tend to ignore
the need of idealisation in their own use of models.

"Infinite repetition" is usually modelled as i.i.d.,
but i.i.d. is defined *in terms of* probability models,
so cannot explain what such models mean.

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.



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

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

Independence is usually argued by
“we don’t see reasons for dependence”
(or “we don’t see regular dependence patterns”).

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

Independence is usually argued by
“we don’t see reasons for dependence”
(or “we don’t see regular dependence patterns”).

Repetition is constructed by selective ignorance.

"I.i.d"-type repetition assumptions are needed
(on some level; e.g. for "error terms")
not because they "hold" in reality,
but because only the idea of regular repetition
allows to learn from data.

"I.i.d"-type repetition assumptions are needed
(on some level; e.g. for "error terms")
not because they "hold" in reality,
but because only the idea of regular repetition
allows to learn from data.

(Similarly, Bayesians need exchangeability.)

“Approximately true?”

“The model assumptions hold approximately” - precise meaning could only be:
“We assume $Q \in \mathcal{P}$ and there is a true P so that $\min_{Q \in \mathcal{P}} d(P, Q)$ is small.”

“Approximately true?”

“The model assumptions hold approximately” - precise meaning could only be:
“We assume $Q \in \mathcal{P}$ and there is a true P so that $\min_{Q \in \mathcal{P}} d(P, Q)$ is small.”

“Approximately true”-arguments are *within* mathematical models;
not about models vs. informal reality.

Davies (1995, 2008):
distributions *can* approximate *data*,
depending on “data features” chosen by researcher.
(Data are formal, too!)

Davies (1995, 2008):
distributions *can* approximate *data*,
depending on “data features” chosen by researcher.
(Data are formal, too!)

However, with every data set,
huge classes of models are compatible.

So this doesn’t give a useful notion of
“approximately true models”.

3. Frequentism-as-model

Exploit frequentist modelling to

- communicate researcher's perception of situation,
- inspire methodology,
- understand implications of methods,
- approximate data,
- check quality of methods
  against known (made up) truth. (Tukey, Davies)

3. Frequentism-as-model

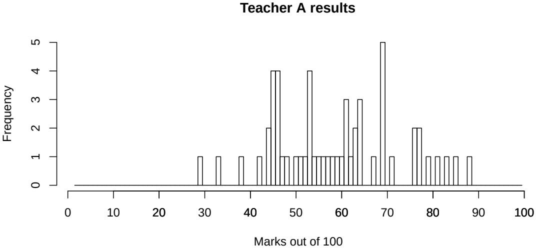
Exploit frequentist modelling to

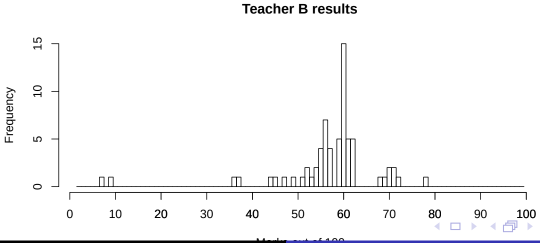
- communicate researcher's perception of situation,
- inspire methodology,
- understand implications of methods,
- approximate data,
- check quality of methods
  against known (made up) truth. (Tukey, Davies)

Keep in mind that
models are neither true nor approximately true;

consider other aspects of "usefulness".

Example


Teacher A results
Frequency
0 1 2 3 4 5
0 10 20 30 40 50 60 70 80 90 100
Marks out of 100


Teacher B results
Frequency
0 5 10 15
0 10 20 30 40 50 60 70 80 90 100

May use frequentist models
for test, confidence intervals (quantifying uncertainty).

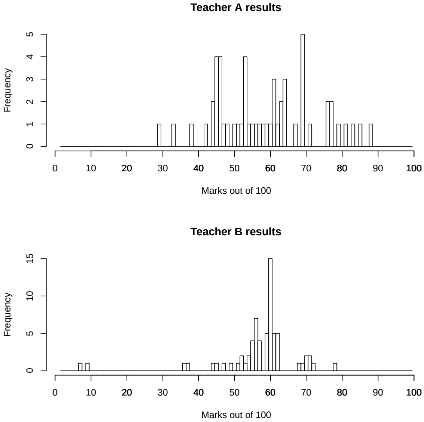
May use frequentist models
for test, confidence intervals (quantifying uncertainty).

Students from same year: clearly not independent
(but may not show dependence pattern),
"identical" only by ignoring background
(*iid is constructed by selective ignorance*).

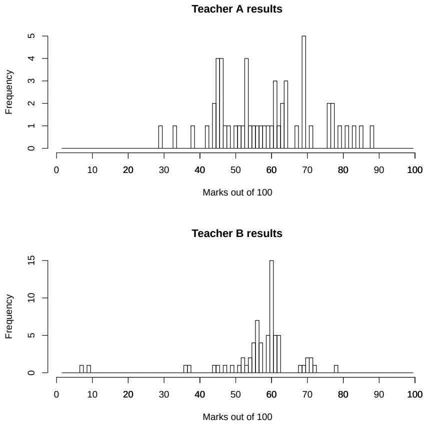
May use frequentist models
for test, confidence intervals (quantifying uncertainty).

Students from same year: clearly not independent
(but may not show dependence pattern),
“identical” only by ignoring background
(*iid is constructed by selective ignorance*).

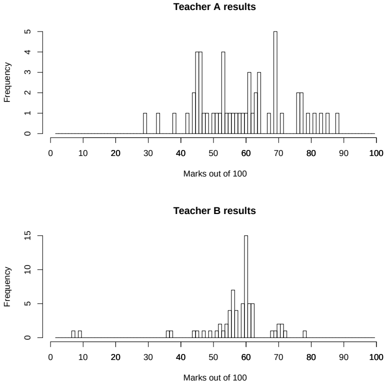
Mean vs. median vs. . . . by optimality in model?
Max breakdown?
(No ultimate criteria, but aspects of understanding.)



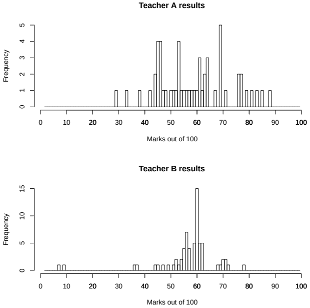
Means: 58.6 (A), 56.9 (B)
Medians: 58 (A), 59 (B).



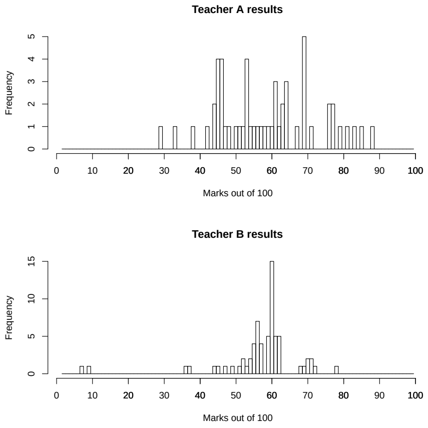
For median: lower outliers in B.



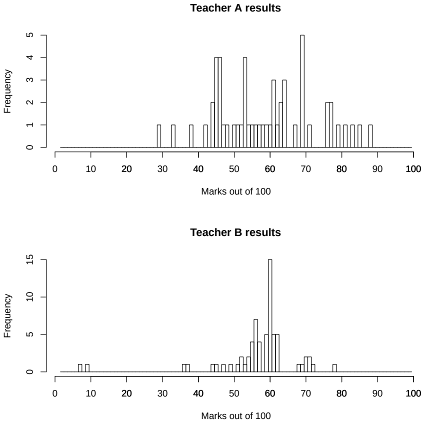
For mean: observations are not erroneous,
so *all* observations should contribute!?
High breakdown may imply ignoring relevant information.



Against mean: it's really not very relevant
how bad the clearly failed students exactly are.
(Would be different for upper outliers.)



Could look at pass ($\geq$ 50) rate (B clearly better)
and other meaningful statistics instead.



"Which aspect is of interest?"/meaning of data dominates "Which model is close to the truth."

4. Connections to robust statistics

What is better about $(1-\epsilon)\mathcal{N}(\mu,\sigma^2)+\epsilon G$ than about $\mathcal{N}(\mu,\sigma^2)$?

4. Connections to robust statistics

What is better about $(1-\epsilon)\mathcal{N}(\mu,\sigma^2)+\epsilon G$ than about $\mathcal{N}(\mu,\sigma^2)$?

Not that it's more "true"/"realistic".

4. Connections to robust statistics

What is better about $(1-\epsilon)\mathcal{N}(\mu, \sigma^2) + \epsilon G$
than about $\mathcal{N}(\mu, \sigma^2)$?

Not that it's more "true"/"realistic".

Better argument: If there are outliers,
mean optimal under $\mathcal{N}(\mu, \sigma^2)$ may not measure
what *we want it to measure*,
as opposed to, e.g., Huber-M or median, . . .

Better argument: If there are outliers,
mean optimal under $\mathcal{N}(\mu, \sigma^2)$ may not measure
what *we want it to measure*,
as opposed to, e.g., Huber-M or median. . .

unless we *want* outliers to carry full weight
(top students; insurance payouts,. . .);

Better argument: If there are outliers,
mean optimal under $\mathcal{N}(\mu, \sigma^2)$ may not measure
what *we want it to measure*,
as opposed to, e.g., Huber-M or median…

unless we *want* outliers to carry full weight
(top students; insurance payouts,…);
although theory then warns us about instability.

The great achievement of robust statistics:

Making the problem of deviation from model assumptions visible (and accessible) by modelling it.

The great achievement of robust statistics:

Making the problem of deviation
from model assumptions visible (and accessible)
by modelling it.

Though problem with deviation extends to *this* model.

Issues with robust statistics

- Influences researcher's perception (as all models): e.g., "good observations" vs. "outliers",

Issues with robust statistics

- Influences researcher's perception (as all models): e.g., "good observations" vs. "outliers",
- Do standard robustness measures really measure what's desirable?

Issues with robust statistics

- Influences researcher's perception (as all models): e.g., "good observations" vs. "outliers",
- Do standard robustness measures really measure what's desirable?
  Sometimes...
  May not want to throw as much information away as high breakdown does.

Issues with robust statistics

- Influences researcher's perception (as all models): e.g., "good observations" vs. "outliers",
- Do standard robustness measures really measure what's desirable? Sometimes...
  May not want to throw as much information away as high breakdown does.
- May tempt researchers into believing that the "violation of assumptions"-problem will go away.

Conclusion

Frequentism-as-model:
“What aspects of data are of interest”;
connecting methods to what we want to know
helped by statistical (robust) theory

Conclusion

Frequentism-as-model:
“What aspects of data are of interest”;
connecting methods to what we want to know
helped by statistical (robust) theory

vs.
decisions made by “the data alone, not the researcher”
or mathematical optimisation
(popular in science in general)

Robust statistics models
what happens when models are wrong.

That's a big step forward,
but also a paradox.