The Role of Model Assumptions in Statistics

Christian Hennig

October 19, 2011

1. Motivation

"Statistical methods have certain model assumptions, and before applying the methods, it should be checked whether the assumptions hold."

1. Motivation

"Statistical methods have certain model assumptions, and before applying the methods, it should be checked whether the assumptions hold."

"We don't believe that the models are really true - all models are wrong but some are useful."

1. Motivation

“Statistical methods have certain model assumptions, and before applying the methods, it should be checked whether the assumptions hold.”

“We don’t believe that the models are really true - all models are wrong but some are useful.”

So what is this business of checking model assumptions really about?

1. Motivation

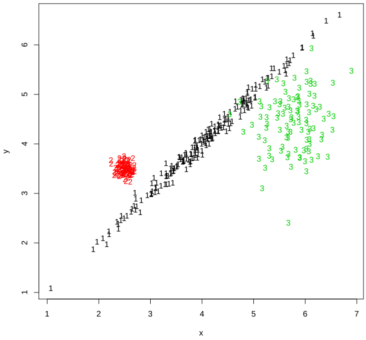
"Statistical methods have certain model assumptions, and before applying the methods, it should be checked whether the assumptions hold."

"We don't believe that the models are really true - all models are wrong but some are useful."

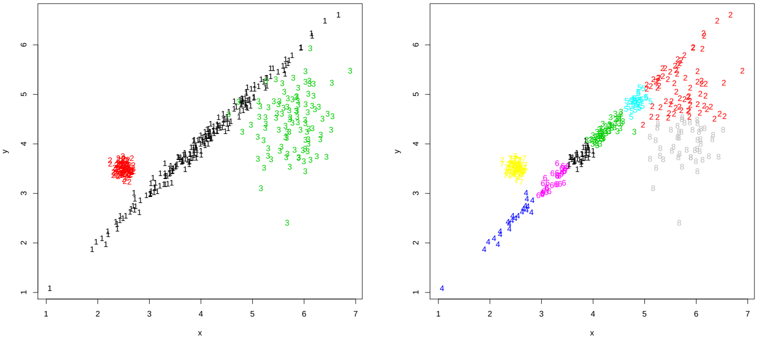
So what is this business of checking model assumptions really about?

Check whether model assumptions hold *approximately*?
Be robust? Be Bayesian?

$$f(\mathbf{x}) = \sum_{i=1}^{k} \pi_i \varphi_{\mathbf{a}_i,\Sigma_i}(\mathbf{x}).$$



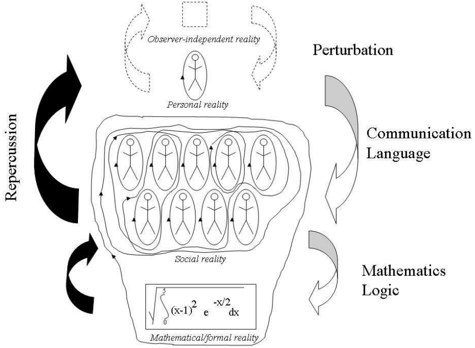
$f(\mathbf{x}) = \sum_{i=1}^{k} \pi_i \varphi_{\mathbf{a}_i,\Sigma_i}(\mathbf{x})$ - Σ_i diagonal?



Overview

1. Introduction
2. Mathematical models and reality
3. What frequentist model assumptions mean
4. The Bayesian assumptions
5. The Bayes-frequentist controversy
6. Conclusion

2. Mathematical models and reality
(H., 2010, Foundations of Science)



- 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.
- It cannot be formally analysed how formal models are related to 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.
- It cannot be formally analysed how formal models are related to informal reality.
- Mathematical modelling is not about how things are, but about how we think about them.

- 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.
- It cannot be formally analysed how formal models are related to informal reality.
- Mathematical modelling is not about how things are, but about how we think about them.
- Pragmatist attitude: what do we get out of 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.
- It cannot be formally analysed how formal models are related to informal reality.
- Mathematical modelling is not about how things are, but about how we think about them.
- Pragmatist attitude: what do we get out of it?
- Science is about establishing agreement in open exchange.
  Mathematics is about creating a system that makes absolute agreement possible.
  (But that's *within mathematics*.)

Antony Gormley - Allotment

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?
- 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?
- 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?

3.1 What do frequentist model assumptions mean?

"We think of the situation as . . . "

- Potentially infinite repetition (of experimental conditions)
- $P(A)$: relative frequency limit of occurrence of A

3.1 What do frequentist model assumptions mean?

“We think of the situation as . . . ”

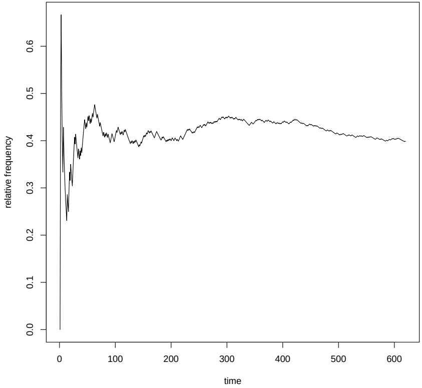
- Potentially infinite repetition (of experimental conditions)
- $P(A)$: relative frequency limit of occurrence of A

“Infinite repetition” is obviously an idealisation.

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

A “repetition” is something where
the observer assesses the differences as irrelevant.

Myth 1: it is an “empirical fact” that relative frequencies converge.

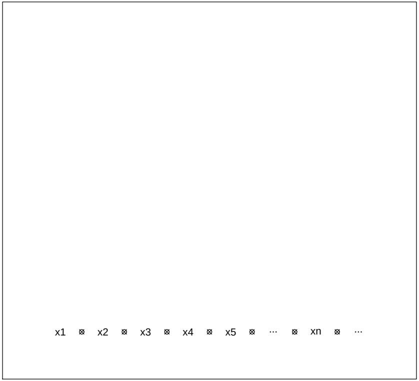


Myth 2: the laws of large numbers confirm frequentist probability concept.

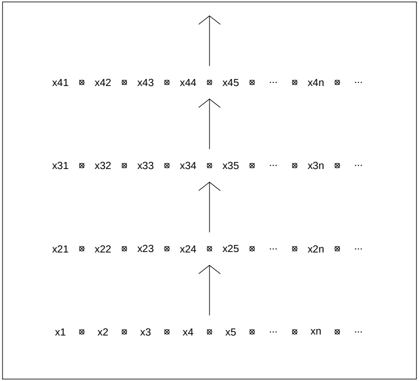
Note: If repetitions *to define* probabilities are modelled as i.i.d., this leads to circularity (as von Mises knew).

“i.i.d.” is defined in terms of probabilities.
Laws of large numbers *assume* frequentism.

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.

This can (only?) be resolved by accepting that
probabilities cannot be *defined* in terms of reality,
and allowing reality only to *falsify* models
by use of (goodness-of-fit) tests.
(*Gillies's propensity interpretation of probability*)

“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."

Dissimilarity measure d is required.
Davies (1995, 2008) distinguishes metrics inducing "weak" and "strong" topologies.
Weak topologies: enable Glivenko-Cantelli theorem:
"$d(P, Q)$ large" $\Leftrightarrow$ P, Q can be distinguished by data.

“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.”

Dissimilarity measure d is required.
Davies (1995, 2008) distinguishes metrics inducing “weak” and “strong” topologies.
Weak topologies: enable Glivenko-Cantelli theorem:
“$d(P, Q)$ large” $\Leftrightarrow$ P, Q can be distinguished by data.
(Kolmorogov, Prohorov, Kuiper metric,
but *not* Kullback-Leibler, total variation, L^p.)

According to weak topologies:
Distributions *can* approximate data, *but*. . .

According to weak topologies:
Distributions *can* approximate data, *but*. . .
In every neighborhood of any P there are distributions. . .

- . . . without Lebesgue-density,
- . . . without existing mean and variance.

According to weak topologies:
Distributions *can* approximate data, *but*...
In every neighborhood of any P there are distributions...

- ...without Lebesgue-density,
- ...without existing mean and variance.

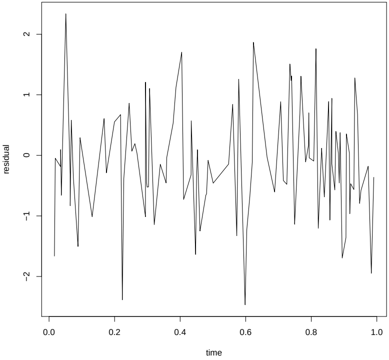
Glivenko-Cantelli still assumes i.i.d.

3.2 Can and should the model assumptions be checked?
3.2.1 Observability, identifiability

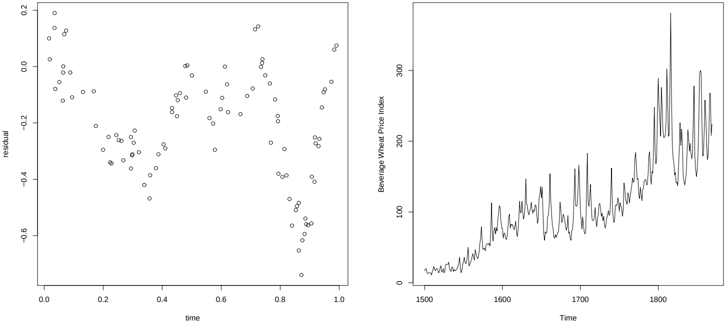
Probability model $\mathcal{P}$ can be distinguished from $\mathcal{Q}$

- by repeating observations
- based on events A with $\exists P \in \mathcal{P}: \ P(A) > Q(A) + \epsilon \ \forall Q \in \mathcal{Q}$.

Approximate independence cannot be distinguished from arbitrary dependence patterns.



Arbitrary dependence cannot be distinguished from arbitrary non-identity.



Assuming repetition (i.i.d.) components in models is required to generate effective sample sizes > 1.

Such assumptions cannot be justified by observation, because they are *required* in order to make meaningful observations about probabilities.

Assuming repetition (i.i.d.) components in models is required to generate effective sample sizes > 1.

Such assumptions cannot be justified by observation, because they are *required* in order to make meaningful observations about probabilities.

"i.i.d." is, in reality, usually justified by arguing that
"we don't know how dependence could work" in given situation.

In fact, "i.i.d.", at some level, is a necessity of statistical thinking.
To have it justified by observation is hopeless.

Existence and value of moments are unobservable…

$$(1-\epsilon)\mathcal{N}(\mu,\sigma^2)+\epsilon G$$

…because depend on extreme events
that may occur with so low probability
that they just don't happen for any real n.

Existence and value of moments are unobservable...

$$(1-\epsilon)\mathcal{N}(\mu,\sigma^2)+\epsilon G$$

...because depend on extreme events
that may occur with so low probability
that they just don't happen for any real n.

In reality we may be rescued by finite value ranges,
so (Gaussian) model-based inference may work
because the model is violated.

3.2.2 The goodness-of-fit paradox

Want to apply statistical method S assuming $\mathcal{P}$.

3.2.2 The goodness-of-fit paradox

Want to apply statistical method S assuming $\mathcal{P}$.
Check model assumptions first by test T with
$Q \in \mathcal{P} \Rightarrow Q(T \in A^c) = \alpha > 0$ small.

3.2.2 The goodness-of-fit paradox

Want to apply statistical method S assuming $\mathcal{P}$.
Check model assumptions first by test T with
$Q \in \mathcal{P} \Rightarrow Q(T \in A^c) = \alpha > 0$ small.
Apply S to data with $T \in A$, generated by $R = Q(\bullet|A)$.

3.2.2 The goodness-of-fit paradox

Want to apply statistical method S assuming $\mathcal{P}$.
Check model assumptions first by test T with
$Q \in \mathcal{P} \Rightarrow Q(T \in A^c) = \alpha > 0$ small.
Apply S to data with $T \in A$, generated by $R = Q(\bullet|A)$.
$R(A^c) = 0 \Rightarrow R \notin \mathcal{P}$.

3.2.2 The goodness-of-fit paradox

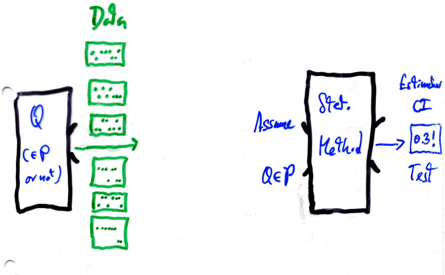
Want to apply statistical method S assuming $\mathcal{P}$.
Check model assumptions first by test T with
$Q \in \mathcal{P} \Rightarrow Q(T \in A^c) = \alpha > 0$ small.
Apply S to data with $T \in A$, generated by $R = Q(\bullet|A)$.
$R(A^c) = 0 \Rightarrow R \notin \mathcal{P}$.

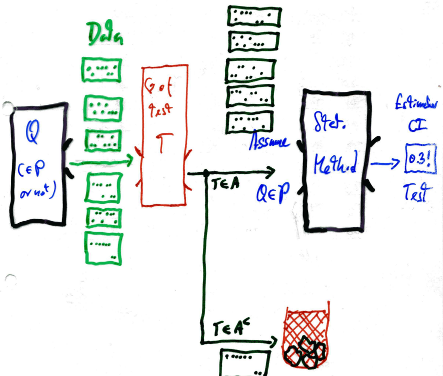
Checking the model assumptions violates them automatically (by introducing dependence so that T cannot be failed).

3.2.2 The goodness-of-fit paradox

Want to apply statistical method S assuming $\mathcal{P}$.
Check model assumptions first by test T with
$Q \in \mathcal{P} \Rightarrow Q(T \in A^c) = \alpha > 0$ small.
Apply S to data with $T \in A$, generated by $R = Q(\bullet|A)$.
$R(A^c) = 0 \Rightarrow R \notin \mathcal{P}$.

Checking the model assumptions violates them automatically
(by introducing dependence so that T cannot be failed).
Holds for informal (visual) checks as well.


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


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

3.2.3 Summary

- Some crucial assumptions are essentially unobservable.
- Without assuming i.i.d. on some level we won’t get anywhere.
- Model assumptions are violated by checking them.

3.2.3 Summary

- Some crucial assumptions are essentially unobservable.
- Without assuming i.i.d. on some level we won’t get anywhere.
- Model assumptions are violated by checking them.

…but I claim that it still sometimes makes sense to do it.

3.3 "Subjective frequentism", a new perspective

(Frequentist) models are still useful to...

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

3.3 “Subjective frequentism”, a new perspective

(Frequentist) models are still useful to. . .

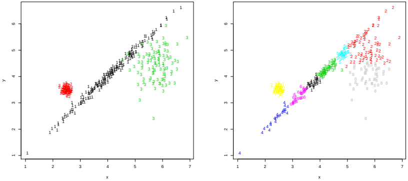
- communicate researcher’s perception of situation,
- inspire methodology,
- check quality of methodology
  in situations with known (made up) truth.
  (Tukey, Davies)

But major motivation of methodology should come from *desired interpretation of results* (subjective).

Social stratification clustering:

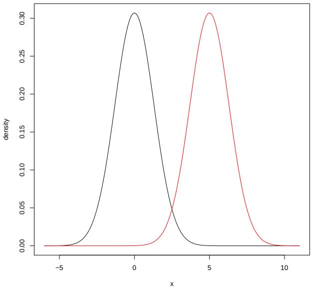
Interpret clusters in terms of marginal distributions, should not model within-cluster dependence.

$f(\mathbf{x}) = \sum_{i=1}^{k} \pi_i \varphi_{\mathbf{a}_i,\Sigma_i}(\mathbf{x})$ - Σ_i diagonal!

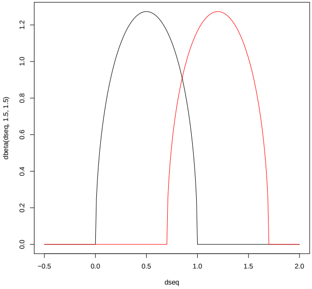


Comparing two samples

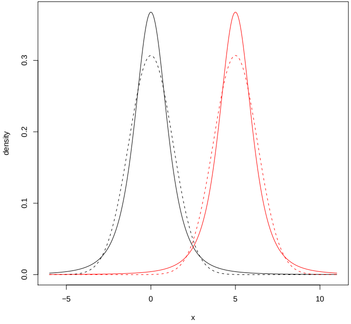
Comparing means is optimal to compare two samples modelled as Gaussian with equal variances.



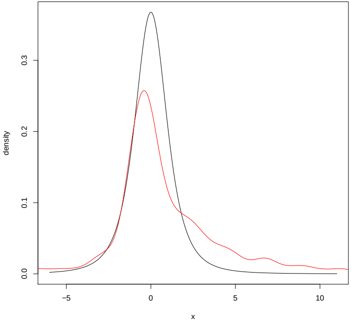
For some clearly un-Gaussian distributions,
comparing means loses “optimality”,
but may still be very good.



With heavier tails, means become much worse, even if “approximation of Gaussian” isn’t bad.



May need to compare distributions between which
there is no clear stochastic order.
Model-based theory won't tell how to aggregate features.



May want to use mean if summing up values across observations is meaningful (money).

May want to use mean if summing up values across observations is meaningful (money).

May not want to use the median,
because it disregards so much information,
even though it may be efficient and robust
- may want "compromise statistic" with nonzero
but much lower than 50% breakdown point.

May want to use mean if summing up values across observations is meaningful (money).

May not want to use the median,
because it disregards so much information,
even though it may be efficient and robust
- may want “compromise statistic” with nonzero
but much lower than 50% breakdown point.

May want to treat differences asymmetrically.

May want to use mean if summing up values across observations is meaningful (money).

May not want to use the median,
because it disregards so much information,
even though it may be efficient and robust
- may want “compromise statistic” with nonzero
but much lower than 50% breakdown point.

May want to treat differences asymmetrically.

Want or not want rank sums depending on whether quantitative differences between values are relevant.

Model assumptions

. . . can be given interpretative meaning;
it is informative how well methods do under any “interpretatively appropriate” model,

but should we be concerned about their “truth”?

Need to check whether the data
may lead methodology astray.

Outliers affect mean and variance.
Systematic dependence affects confidence levels.
Heteroscedasticity implies that
shouldn’t treat observations as equally informative.

Need to check whether the data may lead methodology astray.

Outliers affect mean and variance.
Systematic dependence affects confidence levels.
Heteroscedasticity implies that shouldn't treat observations as equally informative.
Some violations are *not* problematic.

Need to check whether the data may lead methodology astray.

Outliers affect mean and variance.
Systematic dependence affects confidence levels.
Heteroscedasticity implies that shouldn't treat observations as equally informative.
Some violations are *not* problematic.

"Model checks" actually don't check the model, but whether method delivers desired interpretation.

4. What the Bayesians assume

4.1 Bayesian interpretations of probability

Subjective Bayes Probabilities measure individual's strength of belief in future outcomes, formalised as betting rates.

Objective Bayes Same, but believe that objectively rational strengths of belief can be found (if at all well defined).

Hidden frequentist Use Bayesian methodology but present as if there were true underlying distributions.

Bayesian assumptions imply...

- that strength of belief can be properly quantified,
- that this is done by betting rates (de Finetti),
- that coherence is always desirable,
- a neither fully normative
  nor fully descriptive account (subj.),
- principles for objective priors (obj.).

Bayesian assumptions imply. . .

- that strength of belief can be properly quantified,
- that this is done by betting rates (de Finetti),
- that coherence is always desirable,
- a neither fully normative
  nor fully descriptive account (subj.),
- principles for objective priors (obj.).

Much of this cannot be checked by observation either.

Subjective Bayesians acknowledge subjectivity,
but locate it only in prior choice.

4.2 Bayesian idealisation: coherence and exchangeability

Coherence

Coherence commits Bayesians to not learning anything from the data that was not expected in advance.

Bayesians these days do model checks,
may be prepared to violate coherence
by posterior adjustment (Box, Dawid).

But this undermines their motivation
of the probability calculus.

Exchangeability

Exchangeability (on some level) has a similar role as i.i.d. for frequentists.
Bayesians need it to get their methods going.

Exchangeability

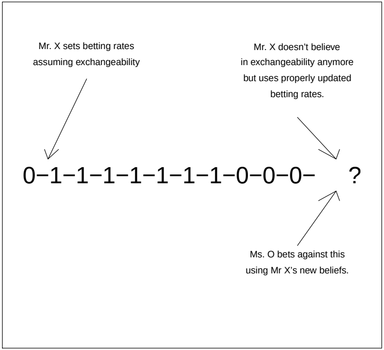
Exchangeability (on some level) has a similar role as i.i.d. for frequentists.
Bayesians need it to get their methods going.

Exchangeability implies that
$P\{1\}$ in the next go doesn't depend on whether you observe
0,0,1,0,1,1,1,0,0,1,0,1,1,0,1,1,0,0,1,0,1,0,0,1 or

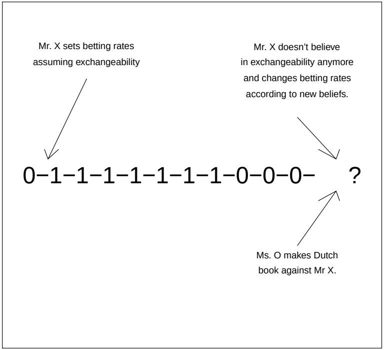
0,0,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0.

This seems counterintuitive.

Incoherence may be good. Scenario 1



Scenario 2 is on average better for Mr. X.



But we can still be Bayesian,
if we accept that this is an idealisation, too,
and do not insist in “truth” of these assumptions,
but rather ask whether they help us
to get the desired interpretation.

From my point of view,
the major philosophical problems
with Bayesian statistics
are the same as with frequentism.

From my point of view,
the major philosophical problems
with Bayesian statistics
are the same as with frequentism.

General problems of mathematical modelling,

From my point of view,
the major philosophical problems
with Bayesian statistics
are the same as with frequentism.

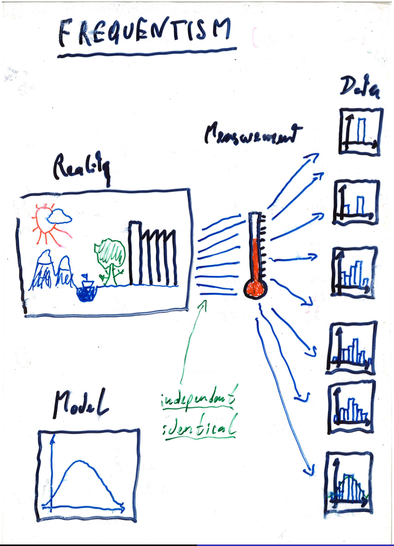
General problems of mathematical modelling,

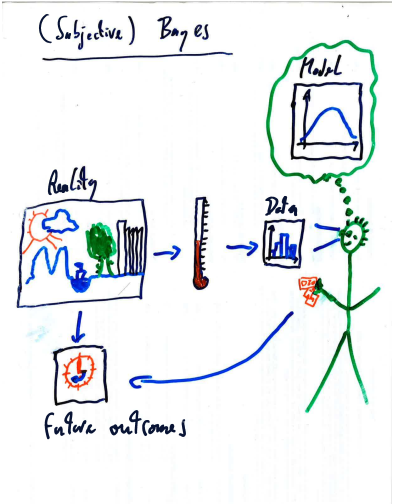
an essential problem with probability models:
they model what could have happened
apart from what happened.

5. The Bayes-frequentist controversy

. . . loses its bitterness if it is . . .

- recognised that they model essentially different “realities”,


FREQUENTISM
Data
Measurement
Reality
Model
independent
identical


(Subjective) Bayes
Model
Reality
Data
Future outcomes

5. The Bayes-frequentist controversy

. . . loses its bitterness if it is . . .

- recognised that they model essentially different "realities",
- accepted that both suffer from essential problems of mathematical modelling and should not be interpreted as reflecting any "real truth" properly,
- accepted that subjective choices and idealisations are inevitable.

Choice of approach:
not “which one is correct”, but
“which approach is implied by our decisions,
and what does the approach imply?”,
“How do we want to interpret results?”,
“How do we want to think about our topic?”

Frequentist vs. Bayes:

- Do we model "world outside" or "rational betting"?
- Do we have background information nicely to be modelled as prior?
- Do we want "$P(A$ given data)" to the price that we have to decide $P(A)$ in advance?
- How can we get scientific agreement about prior, model, decision rule?

..., and then many more decisions are required.

6. Conclusions

- Acknowledge what has to be decided subjectively (hopefully well informed),
- acknowledge basic problems and limits of mathematical modelling,
- acknowledge need for assumptions that cannot be checked,
- motivate chosen method and (model) checks from desired interpretation.

Practical implications

- Present and discuss statistics in a different way.
- Different research questions become interesting:
  - How well do methods work under "interpretatively equal" models?
  - Connect methodology to potentially desired interpretations:
  - Design of loss functions, distance measures, test statistics.
  - How well does mixture clustering do according to distance-based criteria?
  - Define "true number of clusters" by list of desired criteria etc.