university of groningen

NWO
Netherlands Organisation for Scientific Research

Colloquium talk CMU
22 September 2011

Observations and objectivity in statistics

Jan-Willem Romeijn
University of Groningen

Research seminar CUNY
9 September 2011

Meaning shifts and Conditioning

Jan-Willem Romeijn
University of Groningen

To appear in *Studia Logica*

university of groningen

NWO
Netherlands Organisation for Scientific Research

M&E Group Toronto
3 October 2011

How to Frame Experimental Facts?

Jan-Willem Romeijn
University of Groningen

Work in progress, in part with Jon Williamson

Games and Decision Lunchtime meeting
12 October 2011

Learning Juror Competence:
a generalised Condorcet Jury Theorem

Jan-Willem Romeijn and David Atkinson
University of Groningen

Appeared this year in *Politics, Philosophy and Economics*

Kevin's seminar on simplicity
October 2011

Specificity, Accommodation and the Sub-family Problem

Jan-Willem Romeijn
University of Groningen

To appear in an edited volume on *Plurality in Statistics*

LPS Colloquium UC Irvine
21 October 2011

A new resolution of the Judy Benjamin problem

Igor Douven and Jan-Willem Romeijn
University of Groningen

To appear in *Mind*

Lunchtime Colloquium University of Pittsburgh
1 November 2011

Frequencies, Chances and Undefinable Sets

Jan-Willem Romeijn
University of Groningen

Work in progress for my research project on chance

Colloquium talk CMU
22 September 2011

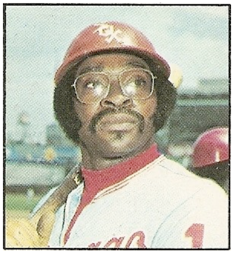
Observations and objectivity in statistics

Jan-Willem Romeijn
University of Groningen

Work in progress, in part for *SEP*

Observing theory

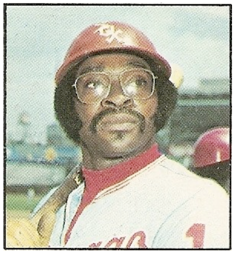
Observation is never independent of implicit, or explicit, interpretation.



We see patches of colour. . .

Observing theory

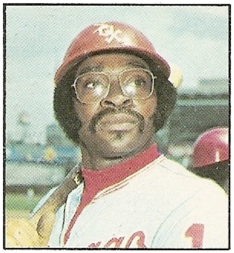
Observation is never independent of implicit, or explicit, interpretation.



We see patches of colour, a man with contraptions on his head. . .

Observing theory

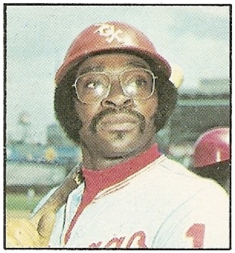
Observation is never independent of implicit, or explicit, interpretation.



We see patches of colour, a man with contraptions on his head, a baseball player with glasses. . .

Observing theory

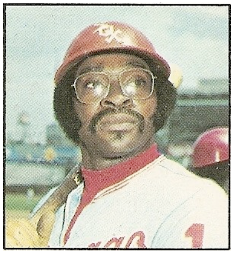
Observation is never independent of implicit, or explicit, interpretation.



We see patches of colour, a man with contraptions on his head, a baseball player with glasses, or Dick Allen, the White Sox homerun leader of 1974.

Observing theory

Observation is never independent of implicit, or explicit, interpretation.



We see patches of colour, a man with contraptions on his head, a baseball player with glasses, or Dick Allen, the White Sox homerun leader of 1974.

But what do we see *objectively*?

Contents

1 Observation in classical statistics

Classical statistical procedures are known to depend on more than just the observations. The results of the statistical analysis depend on. . .

- the *sampling plan*, consisting of rules and procedures for collecting the data, and

- the *statistical model*, or the set of distributions that is under consideration.

Such dependencies can be characterised as violations of the likelihood principle.

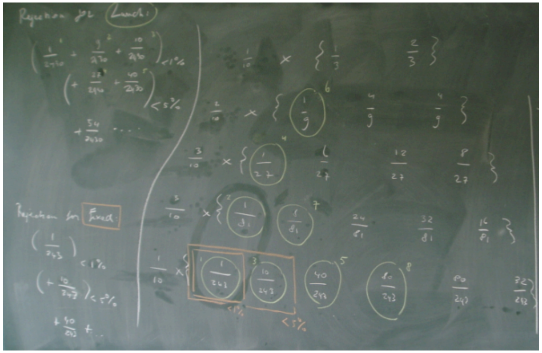
1.1 Optional stopping

Say that three ethologists study the sleeps of a fish, to test the hypothesis that it sinks overnight with probability $\frac{2}{3}$.

The first researcher records for one week, the second stops recording if the weather is good, and the third records until boredom sets in.

Conflicting analyses

As it happens, weather is awful and researchers 1 and 2 record 5 sleeps, in which the fish sinks only once. Researcher 3 is bored after two days.



After checking the weather reports, researcher 2 can reject the null hypothesis. Researcher 1 cannot.

Stopping rule controversy

The weather, and even the mood of the researcher, may influence the analysis. But this is not always wrong.

- The researchers are testing different hypotheses: the sun-loving one brings in the chance on sunny weather, or her assessment thereof.
- Stopping itself may be *informative* and thus can be absorbed into the likelihood: researcher 3 gets bored because sinking is seldom.

The issue is also controversial in more important settings: how long can we justify not treating a control group if the tested drug appears effective?

The persistent experimenter

Another often cited argument for involving stopping rules is that if we do not, we can at long last reject any true hypothesis. But this seems false.

$$P(H_0) = \sum_i P(H_0|D_i)P(D_i)$$

For a Bayesian the prior is a mixture of posteriors. Intuitively, whatever the shape of the sample space, these posteriors must even out.

Persistent worries

Nevertheless there may be reasons for worrying about persistent experimentation.

- Optional stopping may still be detrimental to the *ex ante* quality of a test procedure.
- It can trivially be exploited if the test compares the null with a composite alternative that has an improper prior over it.
- Finitely additive probability distributions may be *non-conglomerable*, i.e. violate the above property.

We conclude that stopping rules matter to what the observations tell us.

1.2 Neyman-Pearson testing

Consider an example inspired by Hacking: two pear orchards producing pears of three different colours. We sample one pear from a truck load.

Hypothesis \ Data	Red	Green	Yellow
Anna	0.00	0.05	0.95
Ben	0.40	0.30	0.30

If the sampled pear is green, the optimal test rules out that the truck came from Anna with 5% significance.

Not exactly wysiwyg

But now consider that we compare the orchard of Ben to the one of Hanna. A green pear cannot license the conclusion that the truck is from Hanna.

Hypothesis \ Data	Red	Green	Yellow
Hanna	0.05	0.05	0.90
Ben	0.40	0.30	0.30

Now, if the truck actually came from Anna, we falsely rejected this hypothesis because, in the words of Jeffreys, “it fails to predict an outcome that does not occur”.

1.3 Sample space dependence

In all the above cases we violate the likelihood principle: the results depend on what we did not, but could have observed.

- The statistical procedures are sensitive to what is deemed observable in a study or an experimental setting.
- Even if they agree on that, they depend on differences between hypotheses concerning events that are not observed.

In other words: what is conveyed by an observation hinges on the full framework in which the observations are received.

Violating total evidence

We can maintain the likelihood principle by violating the principle of total evidence: we reorganise sample space so that it follows the test statistic.

Hypothesis \ Data	Red or Green	Yellow
Anna	0.05	0.95

Hypothesis \ Data	Red	Green or Yellow
Hanna	0.05	0.95

In doing so, we redefine what it is that we are observing. We explicitly tailor the content of observation.

2 Observation in likelihoodist statistics

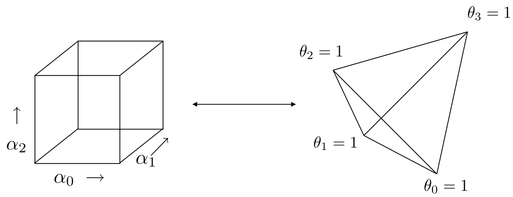
We might think that violations of the likelihood principle are to blame for the apparent theory-ladenness of observations in statistics. Not so.

- In Bayesian inference, the choice of a prior directly influences what conclusions we can draw from the observations.
- Simple regression analysis by maximum likelihood estimation depends on what we take to be exogenous variables.

In other words, the dependence on theoretical context also shows up if the likelihood principle is adhered to.

2.1 Bayesian model selection

We compare two Bernoulli models for two binary variables A and B. The models only differ in the prior probability assignments over the hypotheses.



The different priors are both uniform, but over different parameterisations. One is associated with a causal relation between the variables, $P(A) = \alpha_0$, $P(B|\bar{A}) = \alpha_1$ and $P(B|A) = \alpha_2$, the other is the $3D$ simplex.

Comparing priors

While the likelihoods of the hypotheses in the two models are identical, the model predictions will differ:

$$P_{\text{causal}}(\alpha_0) = 1\,, \qquad\qquad P_{\text{non-causal}}(\alpha_0) = \alpha_0(1-\alpha_0).$$

We can derive an analytic expression for the likelihood ratio:

$$BF = \frac{P(D_n|M_{\text{non-causal}})}{P(D_n|M_{\text{causal}})} = \frac{6(n_0+n_1)(n_2+n_3)}{(n+2)(n+3)}.$$

Different predictive behaviour

We find the following differences in predictive behaviour.

Number of observations n	Interval $\frac{n_0+n_1}{n}$ in which $BF > 1$
< 12	–
12	$\frac{1}{2}$
48	$\left[\frac{1}{4}, \frac{3}{4}\right]$
∞	$\left[\frac{1}{2} - \frac{1}{2\sqrt{3}}, \frac{1}{2} + \frac{1}{2\sqrt{3}}\right]$

Notice that for the first 12 observations, the prior based on the causal parameterization uniformly outperforms the prior over the simplex.

The impact of observations

We draw some tentative conclusions from this example on the influence of priors.

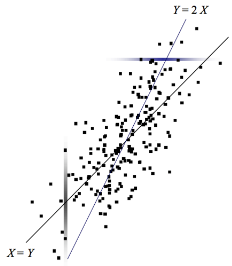
- We improve the short-term predictions by adopting a uniform prior over the parameters associated with a causal model.
- In exchange for this short-term advantage, the long-run predictions of the non-causal model are better.

In the context of the present paper, another conclusion is more relevant.

- The impact of the observations on a statistical model is partly determined by a theoretically motivated prior.

2.2 Regression analysis

A similar dependence on theoretical background, over and above the likelihoods, is illustrated by simple regression.



The same scatterplot can be generated by an exogenous X and a dependent Y, or by the converse roles for X and Y. Swapping these roles leads to a different regression line.

Different regression lines

With some algebra and the substitution

$$u = \frac{\sigma_X \lambda_X}{\epsilon_X},$$

we find the following relations between the two regressions:

$$\mu_Y = \lambda_X \mu_X \,, \qquad \sigma_Y = \epsilon_X \left(1 + \frac{2u^2}{1+u^2}\right)^{-\frac{1}{2}} \,,$$

$$\lambda_Y = \frac{u^2}{\lambda_X(1+u^2)} \,, \qquad \epsilon_Y = \frac{\sigma_X}{\sqrt{1+u^2}} \,.$$

A simple case

To see what underlies this seeming mismatch, consider a simple case with zero means, unit variance, and unit slope,

$$P(X,Y) \sim \exp\left[-\frac{1}{2}X^2\right]\exp\left[-\frac{1}{2}(Y-X)^2\right] = \exp\left[-\frac{1}{2}(2X^2+XY+Y^2)\right].$$

We can only write this as the product of a Gaussian over Y and Gaussians around some regression line by tweaking the parameters,

$$\left(2X^2+XY+Y^2\right) = \left(\frac{Y}{\sqrt{2}}\right)^2 + \left(\frac{X-\frac{1}{2}Y}{\frac{1}{\sqrt{2}}}\right)^2 .$$

So the standard deviation for Y is $\sqrt{2}$, the slope of the regression of X on Y is $\frac{1}{2}$, and the errors are $\frac{1}{\sqrt{2}}$.

No violation of likelihoodism

Forster (2008) constructs the foregoing as a violation of the likelihood principle. I think the cases show us something else.

- They reveal that the decision to view a variable as exogenous has an impact on the estimations in a model.
- In other words, they show that the same distribution may be described in different ways, associated with different theoretical content.

The case is notably similar to the one before. The difference seems to be in putting the improper prior over either X or Y.

2.3 The role of parameterisation

Much like classical statistics, likelihoodist and Bayesian statistics are regulated by the theoretical starting points of the analyses.

- The choice of a prior imports additional knowledge concerning the variables, thereby influencing how the observations impact on them.
- The decision to view variables as exogenous determines how the observations are decomposed into structural component and noise.

The cases illustrate that the import of the observations is not only regulated by the likelihoods of the hypotheses under consideration.

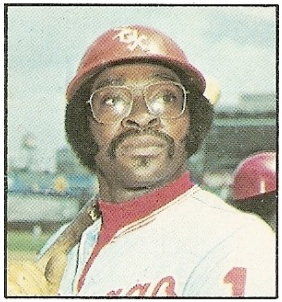
Observations and the prior

Parameterisation and prior have an independent impact on what the observations tell us.

This impact is sometimes qualitative and lasting, despite the fact that numerically the prior washes out for large data sets.

3 The use of theory-ladenness

I started by emphasizing that what we see is determined by the choice of a theoretical framework.



Whether we see Dick Allen or patches of colour depends on our starting points.

Theory-ladenness in statistics

The same phenomenon can be drawn out of a variety of statistical methods.

- In the case of optional stopping, the impact of observations depends on the framework of possible observations.
- Observations in Neyman-Pearson statistics have content only relative to the hypotheses under consideration.
- In Bayesian model selection, differences between the priors allow observations to tell otherwise identical models apart.
- In regression analysis, the choice of exogenous variable determines how the observations are decomposed into structure and noise.

Hume’s problem

We might consider all of this bad news. Once we isolate a neutral notion of observation, subjective starting points seem necessary for learning anything.

This is yet another version of Hume’s problem and Goodman’s new riddle of induction.

The Kantian response

I propose to view this from another angle: it is because of their theoretical content that we can conclude anything interesting from the observations.

Critik
der
reinen Vernunft
von
Immanuel Kant

RUDOLF CARNAP
Der logische Aufbau der Welt

By choosing our language well, we allow the observations to guide us to informative theory.

Thank you

The slides for this talk will be available at http://www.philos.rug.nl/∼romeyn.

For comments and questions, email j.w.romeijn@rug.nl.