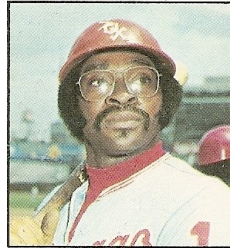PML seminar
UvA 2010

Observations and objectivity in statistics

Jan-Willem Romeijn

Faculty of Philosophy
University of Groningen

Observing theory

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

We see patches of colour, a man with funny 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 set of hypotheses, each representing a probability function over sample space.

Such dependence is often characterised as a violation of the likelihood principle: the impact of the data is not completely captured by the likelihoods for the data actually obtained.

1.1 Stopping rules

One famed controversy over the role of the sampling plan concerns ordinary null hypothesis testing. Say that two ethologists study the sleeps of a fish, to test the hypothesis that it sinks overnight with probability $\frac{2}{3}$.

One of them decides to record 5 sleeps, the other allows herself to stop recording if the weather merits a trip to the beach.

Stopping rules

As it happens, weather is awful and they both record 5 sleeps, in which the fish sinks only once. Because they have different rules for collecting data, the results of their analyses will differ.

After checking the weather reports, the sun-loving researcher can reject the null hypothesis. The more disciplined researcher cannot.

Stopping rules
Via the stopping rules, the intentions of the researcher influence the statistical analysis, introducing an element of subjectivity. But this dependence is not always obviously wrong.

- Arguably, the researchers are testing different hypotheses: the sun-loving researcher brings in the chance on sunny weather. It is correct that this is reflected in the different results.

- Stopping rules that depend on the potential results of the analysis seem dodgy: they allow for a so-called persistent experimenter. Perhaps we must restrict rules to be independent of the results.

The only common ground in the debate seems to be that the issue is important. For how long can we justify not treating a control group with a drug that shows to be highly effective? Can we stop running time-consuming and expensive tests if the results are overwhelmingly clear?

Persistent experimentation
Mayo (2004) claims that if we leave Bayesians—who allow for optional stopping—enough time, they can reject any hypothesis with probability 1. But we must be careful in positioning and interpreting this claim.

- It is more precise to say that given enough time, Bayesians will at some intermediate stage have arrived at an arbitrarily low probability for the true hypothesis.

- This is not in conflict with the convergence results according to which, given enough time, the probability of the true hypothesis tends to 1.

This means that even with optional stopping, Bayesian inference does not lead us astray.

1.2 Not exactly wysiwyg

To get a grip on the stopping rule controversy, consider an example gleaned from Hacking (1965). Two pear orchards producing pears of three different colours. We sample one pear from a truck load.

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

According to Neyman-Pearson hypothesis testing, we can rule out with 5% significance that the truck came from Alma's orchard if the sampled pear turns out to be green.

Not exactly wysiwyg
But now consider that we compare the orchard of Ben to the one of Anna.

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

We cannot rule out that the truck is from Anna's orchard on the basis of a green pear. Now if the truck actually comes from Alma, we falsely reject this hypothesis because, in the words of Jeffreys (1931), "it fails to predict an outcome that does not occur".

1.3   Violating the likelihood principle

The common denominator for the above cases is that we violate the likelihood principle: the result depend not just on the likelihoods for the actual observation, but also on what we 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 theoretical framework in which the observations are received.

Or violating the principle of total evidence?

We can choose to maintain the likelihood principle at the cost of violating the principle of total evidence. We simply reorganise sample space so that it follows the test statistic.

Hypothesis \ Data	Red or Green	Yellow
Alma	0.05	0.95

Hypothesis \ Data	Red	Green or Yellow
Anna	0.05	0.95

In doing so, we redefine what it is that we are observing. We are then explicitly tailoring the content of observation to fit the theoretical framework.

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.

- Also when refraining from the use of priors, 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 Testing causal priors

Say that we compare two Bernoulli models for two binary variables and hence four elementary possibilities. The models only differ in the prior probability assignments over the hypotheses.
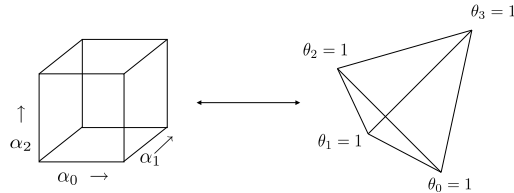
The different priors are both uniform, but over different parameterisations. One of these is associated with a causal relation between the variables, the other is a $3D$ simplex.

Testing causal priors

While the likelihoods of the hypotheses in the two models are identical, the model predictions will differ. We can derive an analytic expression for the likelihood ratio:

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

We find the following interesting points:

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

Testing causal priors

We can 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.

But in the context of the present paper, the following 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 likeli-hoods of the hypotheses under consideration, is illustrated by Forster (2008).
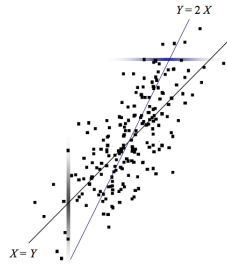
The very same scatterplot can be generated by an exogenous variable X and a dependent variable Y, or by the converse roles for X and Y. Swapping these roles leads to a different regression line.

Regression analysis

Following Borsboom, Wicherts, and Romeijn (2008), with some algebra and the substitution

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

we find the following relations between the two regressions,

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

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

Regression analysis

We can perhaps see what causes this seeming mismatch by considering 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,

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

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 $\frac{1}{\sqrt{2}}$.

Regression analysis

Forster employs the foregoing in the construction of alledged violations of the likelihood principle. But I think the cases show us something else.

- The examples reveal that the decision to view a variable as exogenous may have its impact on the conclusions we draw from the estimation in a model.

- Alternatively, they show that the probability function determining a statistical model may be described in different ways, associated with different theoretical content.

The only common ground in the debate seems to be that the issue is important. For how long can we justify not treating a control group with a drug that shows to be highly effective? Can we stop running time-consuming and expensive tests if the results are overwhelmingly clear?

2.3   The content of observation

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

- The choice of a prior expresses additional theoretical knowledge concerning the variables, and thereby influences how the observations affect our theory.

- The decision to view variables as exogenous determines how the observations are decomposed into structural component and noise.

So, it is also regulated by the theoretical starting points of the statistical analyses.

3   The use of theory-ladenness

We might consider all of this bad news. Even in the seemingly straightfor-
ward setting of statistics, we fail to isolate a neutral notion of observational
content.
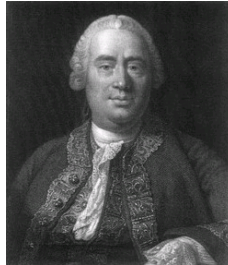
This is yet another version of Hume's problem of induction, and tied up with
the failure of the Carnapian programme in inductive logic.

The Kantian response

I propose to view the same fact from another angle: it is because of the theoretical content we endow the observations with, that we can conclude anything interesting from the observations.

By carefully choosing what it is we see, 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.