Science and belief: A plea for epistemic statistics

Richard D. Morey
Department of Behavioural and Social Sciences
University of Groningen

Jan-Willem Romeijn
Department of Philosophy
University of Groningen

Abstract

We identify a problem in the foundations of statistics that has received little attention among scientists: the results of most statistical methods cannot warrant scientists to believe, disbelieve, or indeed take any epistemic attitude, to the hypotheses at issue. We discuss the historical and conceptual roots of this problem, present a perspective on the its resolution, and argue for the importance of belief-driven statistical methods that better serve science.

For a long time, the main controversy in statistics has been the opposition of classical and Bayesian methods. But these woeful times seem to be coming to an end: many statistical practitioners accept Bayesian tools as a welcome addition, and seem at ease with a pragmatic attitude to the choice of method. This might lead us to believe that problems in the foundations of statistics have finally been resolved. But we believe that a far more critical philosophical distinction has been missed: belief versus decision. Although accounting for rational belief is critical in science, researchers would be surprised to learn that

Send correspondence to r.d.morey@rug.nl, or Grote Kruisstraat 2/1, 9712NS Groningen, The Netherlands

current statistical theory does not support it. This presents the foundations of statistics with problems that rival the older controversy in importance and scope.

One of the core features of science — perhaps the core feature — is its significance for what we believe. It is beyond contention that our beliefs should somehow be constrained by scientifically established empirical facts. Of course, such matters as what exactly counts as an empirical fact and how beliefs are supposed to be constrained by them are subject to debate. But at any rate, once the empirical facts are established scientists are not completely free to adapt their beliefs in whatever way they want. If we say that the facts provide evidence for one hypothesis over another, we are at the very least saying that these facts support believing in the one hypothesis more than believing in the other, that they bring us closer to accepting the hypothesis, or the like.

The epistemic authority of science has to be exercised by means of an independently-motivated theory that relates beliefs and facts in a particular way. The stock example of such a theory is deductive logic. If a the conjunction of a theory and its auxiliaries predicts a particular experimental outcome, and that outcome does not occur, then one is compelled by logic to believe that the theory is false. In such simple cases, logic provides the link between empirical fact and belief. In more realistic cases, deductive logic alone will not suffice. But even so, it remains essential to scientists' epistemic authority that beliefs be constrained by the facts according to some justifiable method.

The notion that beliefs must be constrained by empirical fact in a regulated way will appear especially obvious to anyone involved in statistical data analysis: statistics is directly concerned with the link between the facts, codified in data, and hypotheses. Scientific writing is rife with epistemic language: results may be convincing, a hypothesis may be credible or supported, another hypothesis may be doubtful. The prevalence of epistemic language in reporting statistical results makes it all the more surprising that for most of the past century, statistics has been overwhelmingly dominated by methods that, by design, have no interface with belief. In an embarrassment for science, dominant classical methods place no constraints on belief at all, in spite of the common belief that they do. Moreover, the increasingly popular Bayesian methods — often touted as the epistemically-oriented alternative to classical methods — do not unequivocally relate to beliefs either. In foundational work concerning Bayesianism, beliefs are treated as a dependent category in a wider decision-theoretic framework. And in practical applications, Bayesian methods are often not interpreted in terms of belief at all. Thus, methods of statistical inference mostly lack a central feature required for scientific inference.

In what follows we elaborate this shortcoming of both classical and Bayesian statistics. Following this, we outline a resolution of the problems, in both statistical methodology and practice.

Classical statistics

In the first half of the 20th century, the statistical methods that we now call classical were taking shape. Both Fisher's work on likelihood and Neyman's work on hypothesis testing represented shifts from older statistical methods, then known as inverse probability and presently called Bayesian methods. Many statisticians had grown suspicious of inverse probability and its requirement of prior distributions. While Fisher and Neyman shared this suspicion about priors, they disagreed vociferously about what should replace them. At the heart of their debate was the question of what kinds of questions statistics should answer.

Fisher, on this point agreeing with the inverse probability school, held that statistics should be concerned with rational belief. When a researcher obtains data from an experiment, she wants to know what to believe about the relative plausibility of different possible conclusions from those data. Fisher developed the theory of likelihood, a concept intended to measure support, and later his theory of fiducial probability, a notion measuring trust: the term "fiducial" deriving from the Latin term *fiducia* which means "trust" or "belief". That statistics should be about rational belief was self-evident to Fisher through his applied work as a scientist, and he criticized statisticians who thought otherwise as not understanding science (**Fisher:1935**; **Fisher:1955**).

Neyman was perhaps the foremost statistician opposing the older way of thinking. In line with the empiricist and behaviorist philosophy popular in his day, Neyman introduced the idea of "inductive behavior" (**Neyman:1957**). Under this statistical philosophy, researchers adopt rules for behavior that have certain optimal properties, in the long run. We might, for instance, adopt a rule that guarantees that if a drug has no effect, we will only act as if it does in 5% of all cases, thereby putting an upper bound on so-called Type I errors. An optimal rule for inductive behavior then minimizes the frequency of so-called Type II errors: that is, of failures to declare an effect if it does exist. Importantly, this is solely a rule for behavior: it determines that we will act as though an effect obtains if the observations meet particular criteria, and not otherwise. Such rules have nothing to do with reasonable beliefs about particular hypotheses. In his 1957 paper describing inductive behavior, Neyman even rejects the idea that statistical hypothesis testing is a form of reasoning at all.

Neyman and Pearson, in fact, "won" the debate over the purpose of statistics. Their action-oriented approach to statistical hypothesis testing is the dominant one taught to students and used by researchers. In the theory of classical statistics, procedures thus supply advice for actions, but no constraints on belief.

Bayesian statistics

It may be thought that this particular view on statistical methods is typical of classical statistics, and that Bayesian statistics is directed towards constraining beliefs. To some extent this is correct, because Bayesian statistics provides conceptual tools for expressing relations between beliefs and data. While some Bayesian methodologies sit relatively close to epistemology (e.g. the subjective Bayesian approach of **Goldstein:2006** and others), modern Bayesians may or may not have an epistemological orientation. Gelman and Shalizi's pragmatic Bayesianism (**Gelman:Shalizi:2013**) explicitly eschews the interpretation of the Bayesian prior in terms of belief as unnecessary, while objective Bayesianism (e.g. **Berger:2006**) can result in procedures that are impossible to interpret as expressing belief. Objective Bayesians may view their procedures as being a part of a larger Bayesian framework, some of which may be belief-based (**Berger:2006**; **Bernardo:1997**); our point here is that just because a procedure is "Bayesian" does not guarantee an epistemological orientation. Objective Bayesianism is one of the most popular forms of Bayesianism perhaps precisely because it does not require the explicit interpretation of the Bayesian prior in terms of belief, thus avoiding the critique that it is "subjective."

On the foundations side of Bayesianism, the logical nature of statistical inference is emphasized, linking statistics to an epistemic perspective (**Hacking:1965**; **Good:1983**; **Lindley:2000**; **Kadane:2011**). However, Bayesian statistics grew out of a foundational programme in which the very same empiricist and behaviorist ideas were dominant as those that influenced Neyman. As a result, Bayesian statistics is not uniformly geared towards beliefs either. Frank Ramsey and Bruno de Finetti put on firm conceptual footing the idea that one might represent uncertain opinion by a probability function (**Ramsey:1931**; **deFinetti:1937**). They defined uncertain opinion in terms of a willingness to act, where the acts were represented as bets on possible results. In Ramsey's famous example, if I run to catch the train, I effectively buy a bet that is likely to pay out the reward of catching the train, at the cost of the discomfort involved in running. The general idea behind this is rooted in a form of empiricism: what matters in statistical inference are only empirical consequences, namely the decisions and actions of deliberating agents. The development of an action-oriented philosophy of statistics culminated in the simultaneous axiomatization of beliefs and decisions by Leonard Savage (**Savage:1972**). In this axiomatization beliefs became a subordinate category in the theory of rational choice. The locus of rationality is therefore not belief per se, but rather decision or action. In fact, without a decision to be made whose eventual consequences can be measured as a cost or benefit, it is not clear whether rational belief has any meaning in these axiomatizations.

More recently philosophers have provided justifications for representing beliefs by probability that do not rely on a behaviorist reduction of epistemic attitudes(**Joyce:1998**; **Leitgeb:Pettigrew:2010a**; **Leitgeb:Pettigrew:2010b**). Rather the probabilistic nature of beliefs is derived from the fact that they must accurately represent the world. We believe these developments to be a step in the right direction, because the aim of science is at least partly to offer accurate representations, and not merely representations that promote good decisions. We hope that epistemic concerns are more fully embraced by the developers and users of Bayesian methods. Without an epistemic foundation, scientists' justification of their use of Bayesian procedures on the grounds that they are "rational" may be hollow.

Summing up, we contend that Bayesian statistics falls short of providing unequivocal epistemological constraints. In early foundational work on Bayesian statistics, interpretations of probability ultimately rest on the decisions and actions that they motivate. And while some recent and philosophically-oriented approaches do emphasize epistemic aspects, Bayesian statistics has since developed into a practical toolkit that often shows little affiliation with epistemological concerns.

Epistemological statistics

Both classical and Bayesian statistics are currently insufficiently equipped to serve scientists who look for constraints on belief. This state of affairs is most immediate for the classical statistical methods that presently enjoy most popularity, but we argued above that Bayesian methods fare less well on this count than is often considered. However, the scientific spirit that Fisher identified with — the need for statistics to address rational belief — remains. For instance, researchers often use a rejection of the null hypothesis to argue that it is reasonable to believe that the null hypothesis is false, although this is incorrect usage of the method of null hypothesis testing. Because statistical methods do not provide epistemic advice, researchers end up constraining their beliefs by convention, custom, or case-based methodology, and not by an independently-motivated theory on how data relate to beliefs in hypotheses.

One might argue that this situation is acceptable. Scientists might talk about their beliefs informally over coffee, but, following an influential philosophical view on science (**Duhem:1954**; **vanFraassen:1980**), one could deny that the main goal of science is rational belief. Instead we might maintain that its core business is to provide predictive tools and guidelines for action, or merely to accord with empirical fact irrespectively of the truth or falsity of its theoretical claims. Along these lines, we should continue to keep statistical methods and epistemological advice separate, and improve the training of scientists so that they avoid making epistemological claims. Talk of convincing results, evidence, support, credibility, plausibility — all common language with epistemic meaning — should all be struck from scientists' writing. Scientists would thus be robbed of a crucial component of their communicative resources. They would not have recourse to formal epistemic concepts, so they must either regard epistemic talk as merely a manner of speaking, or make epistemic claims unregulated by theory and informed only by custom.

To our mind, however, such a situation is unnecessary and ultimately not conducive to good science. We prefer a different approach: to restore epistemology as a core focus of statistics. In the remainder of this paper, we sketch what epistemological statistics might look like.

Inferentialism in statistics

There are quite a few theoretical and philosophical perspectives on statistics that do have an explicit epistemological focus. None of these approaches to statistics has made it to the statistical mainstream, but some enjoy cult status among philosophers of statistics and more reflectively oriented statisticians. We will briefly discuss some of them, to indicate the various ways in which statistics can be made to fit epistemological goals.

In classical statistics, arguably the oldest attempt to orient on the epistemic is Fisher's fiducial probability. Fisher developed the method of maximum likelihood and then claimed that under certain conditions it allows us to derive constraints on rational belief, in the form of fiducial probability, from data alone (**Dawid:Stone:1982**; **Seidenfeld:1992**). However, fiducial probability is highly controversial and hardly used today, though **Jeffreys:1935** pointed out that the best argument for using likelihood in such a manner is actually Bayesian, and **Dempster:1968** also shows its intimate connection to Bayesian inference. In the 1960s Henry Kyburg (**Kyburg:1961**) developed a defeasible logic to match classical statistics, the theory of so-called evidential probability. The logic focuses in particular on the problem of collating statistical results such as estimations and confidence intervals from different studies, thus providing a guideline for belief as an add-on to classical statistics itself.

There have also been attempts to reinterpret classical statistics as pertaining to the epistemic realm. Deborah Mayo developed the ideas of Neyman and Pearson on stringent testing (**Mayo:1996**) into a theory on error avoidance, learning, and reliable knowledge. Notably, those concepts concern beliefs rather than action. Many others have argued for an epistemological reading of classical statistics by emphasizing that it helps us to determine support and evidential strength, both of which are clearly epistemic in nature. More precisely, **Jeffreys:1935**; **Hacking:1965** and **Edwards:1972** proposed that the likelihoods of hypotheses express the relative strength of evidence for those hypotheses. More recently this so-called likelihoodism is defended by **Royall:1997** and **Sober:2008** All of these present more or less systematic attempts to interpret classical statistics as an epistemological project. **Dawid:etal:2011** discuss evidence from a broad perspective, featuring several approaches to evidence that bear on statistics (**Dawid:etal:2011a**; **Gardner-Medwin:2011**).

On the Bayesian side, several subjectivists (**Howson:2001**; **Kadane:2011**) have developed the logical approach of de Finetti, thereby revealing a sensitivity for epistemological considerations. Moreover, in recent years philosophy has seen active research in so-called probabilistic epistemology (**Jeffrey:1992**; **Howson:Urbach:2006**; **Hajek:Hartmann:2010**). This budding branch of epistemology provides conceptual analyses of beliefs in probabilistic terms and discusses many issues that are central to statistics: the transition from probabilistic to full belief, the interpretation of probability assignments as epistemic or physical, the nature of evidence, and so on. It is naturally related to the systematic study of support relations between data and hypotheses, as carried out in inductive logic and confirmation theory. We believe that these philosophical disciplines have a fruitful common ground with epistemically-oriented statistics.

This is not the place to argue in detail for one of the many perspectives mentioned above. Some focus on evidence, others on opinion, some are permissive and others restrictive; the choice for a perspective depends on too many considerations. But we submit that in any of these perspectives, statistics is best understood as a theory of sound reasoning, comparable to logic. We therefore urge for *inferentialism* about statistics: the statistical model and the data act as premises, the statistical method as an inference rule, and the evaluations, such as predictions and rejections, as conclusions. The constraints on belief thereby take a conditional form. If we believe model and data, then our further beliefs must be constrained in such-and-such a way. Thus conceived, statistics will at the very least provide consistent constraints on epistemic attitudes, but one may hope that it will constrain our beliefs more, perhaps even, in some cases, to a uniquely most rational belief given model, initial beliefs, and data.

This is of course not to say that a "most rational" belief will be the end of analysis: new possibilities may present themselves which can then be evaluated after choosing a refined model, and possibly a new prior. As we argued elsewhere (**Morey:etal:2013**), an epistemic approach to statistics need not suffer from the closed-mindedness that Bayesian statistics is sometimes criticized for (**Gelman:Shalizi:2013**), certainly not if statistics is given the conditional form alluded to above. It is a challenge for any epistemic account of statistics to balance the constraints from rationality with the flexibility required for open-minded science.

The future of data analysis

To readers familiar with the foundational debate on classical and Bayesian statistics, our call for methods addressing belief may appear to be another pro-Bayesian salvo, since Bayesians are by and large more receptive to epistemic aspects of statistical inference. But our contribution here is not geared towards the classical versus Bayes debate. We call for a more comprehensive consideration of epistemology in statistics, be it by classical or Bayesian means. We believe that the debate on methodology will be more fruitful if it addresses these concerns, which are by and large orthogonal to the traditionally prominent classical vs. Bayesian dividing lines.

Our proposal has implications down the chain from the practice of statistical inference by researchers, through the theoretical development of statisticians, to the foundational work of philosophers. At present, the relationship between researchers and statisticians is decidedly lopsided. Statistical consulting often involves researchers asking statisticians what the "right" procedure is for their data, or is focused on model building. Researchers, however, should take the initiative in demanding statistics whose meaning conforms to their epistemic goals, and should ask statisticians with whom they consult to describe how statistical methodologies interface with the claims the researcher wants to make on the basis of her data. To support this in the future, statistical training needs a reorientation: an understanding of the philosophy underlying statistics must be taken as seriously as the mathematical acumen needed for statistics.

Further consequences concern the relationship between philosophers and statisticians. Statistics, we argue, is unfortunately positioned as applied mathematics, and is much better viewed as a form of applied philosophy carried out by mathematical means. The tools of the statistician are indeed mathematical but the questions concern relations between data and hypothesis, and rightfully fall within the domain of epistemology. More tight cooperation is therefore needed between philosophers and statisticians to investigate how statistical tools serve epistemic goals. As a start, we think that statistical training should involve a basic introduction to confirmation theory, both logical and probabilistic, and that philosophers of science should come to view confirmation theory in constant conjunction to statistical practice.

A reorientation of the field of statistics may sound radical, but we believe that the alternative leaves us in an even stranger place. Under the status quo, we have two options with regard to talking about beliefs concerning statistical hypotheses: discouraging scientific talk about beliefs, on the basis that it is unprincipled; and allowing all such talk, on the basis that the methods don't constrain them. To purge the language of beliefs from discussion of statistical results would be nearly impossible. Scientists, after all, consider a major goal of science to be accumulation of knowledge, and it is nonsensical to talk about a scientist's knowledge without talking about belief. Scientists will continue to talk about beliefs, and nothing a philosopher or statistician can say will stop them. Likewise, scientists will continue to be guided by the particular discipline-relative ways in which data constrain beliefs. If the constraints are not independently justified, they give rise to a potentially damaging variety of epistemic standards across science. This has, in fact, already occurred: **Wagenmakers:etal:2008** for example, discusses the disagreement across researchers about such a fundamental matter as whether a larger sample size corresponds to more evidence or less for a given p value. Consider also the fact that when interpreted in epistemic terms, the common null hypothesis significance test yields an invalid inference **Cohen:1994**; **Pollard:Richardson:1987** The understandable difficulty researchers have in coming to grips with the epistemic implications of their statistical procedures makes this matter a pressing issue.

Conclusion

We wish to be clear about what we are saying and what not: we are not claiming that all previous statements about evidence without a firm grounding in epistemology are fatally flawed. Epistemology, scientific methodology, and statistics are all dynamic fields that improve over time: our call for a refocus in statistics is forwarding-looking, towards improved methods that build on older attempts that may have fallen short. Nor are we calling for the formalization of all aspects of scientific reasoning; there will always remain an aspect to science that relies on intuition and creativity. But increasing the prominence of epistemological considerations in statistics is both an important and an achievable goal. Just as the 20th century was the century of action-oriented statistics, the 21st century should be dedicated to realizing Fisher's goal: a statistics that can help achieve the epistemological objectives of science.