Inductive vs. Deductive Statistical Inference

Konstantin Genin
Kevin T. Kelly

February 28, 2018

Abstract

The distinction between deductive (infallible, monotonic) and inductive (fallible, non-monotonic) inference is fundamental in the philosophy of science. However, virtually all scientific inference is statistical, which falls on the inductive side of the traditional distinction. We propose that deduction should be *nearly* infallible and monotonic, up to an arbitrarily small, a priori bound on chance of error. A challenge to that revision is that deduction, so conceived, has a structure entirely distinct from ideal, infallible deduction, blocking useful analogies from the logical to the statistical domain. We respond by tracing the logical insights of traditional philosophy of science to the underlying **information topology** over possible worlds, which corresponds to deductive verifiability. Then we isolate the unique information topology over probabilistic worlds that corresponds to statistical verifiability. That topology provides a structural bridge between statistics and logical insights in the philosophy science.

1 Traditional Taxonomy of Inference

Nothing could be more familiar or more obvious in the philosophy of science than the elementary distinction between inductive and deductive inference. For example:

> Any . . . inference in science belongs to one of two kinds: either it yields certainty in the sense that the conclusion is necessarily true, provided that the premises are true, or it does not. The first kind is . . . deductive inference . . . . The second kind will . . . be called 'inductive inference' [Carnap, 1952, p. 3].

Carnap identified deductive inference in science with logically valid deduction. Logically valid deductions have two crucial properties. They are **truth preserving** or **infallible**—the conclusion must be true if the premises are, and they are **monotonic**, in the sense that adding further premises does not invalidate them. The suggested taxonomy is:

inference:	deductive:	truth preserving, monotonic
	inductive:	everything else.

On that view, deduction appears to apply to such problems as calculation, refutation of a universal law, verification of an existential law, deciding between two universal laws, deriving predictions from a law, and determining which laws are logically compatible with the given evidence. But one must rely on induction to infer a universal law from the observations, to choose among an infinite range of universal laws, etc. Thus, the distinction between induction and deduction provides a revealing, high-level, epistemic distinction among scientific activities.

2 Trivialization by Probable Error

The trouble is that the preceding taxonomy doesn't really work. Statistical laws don't necessarily imply anything about what will happen. The predictions hold, at best, with high probability. Hempel therefore placed all such inferences on the side of induction, including prediction and explanation from a statistical theory:

> Explanatory arguments which ... account for a phenomenon by reference to statistical laws are not of the strictly deductive type. An account of this type will be called an ... inductive explanation [Hempel, 1965, p. 302].

Alas, even deterministic laws must be tested with concrete measurements, all of which are subject to error. Error is usually attributed to random causes that result in a probability distribution of possible measurement values. One can reduce the probable error of measurement by improvements in instrumentation, by diligence in its employment, and by making repeated measurements and averaging them, but one can never eliminate it entirely. So testing laws should fall, after all, on the side of induction.

Deriving predictions from a universal law is also fraught with probable error. Laws typically have free parameters that must be estimated from samples. One can tighten the predictions by basing the estimates on larger samples, but the random error never goes away entirely. So deriving predictions from a law should, again, fall on the side of induction.

Real-life calculations are also subject to probable error. Mathematicians can be distracted or make mistakes. Every concrete computation is subject to error due to faulty programming. More fundamentally, every physical computational process is subject to interference by cosmic rays. As a result, "data decay" is a real issue in computing that becomes a significant risk for computations in deep space probes [Niranjan and Frenzel, 1996]. One can deal with the problem by redundancy and repeating the computations, but it never goes away entirely.

To summarize; infallible, monotonic reasoning is very desirable, and there are detailed logical and computational theories of it. But no concrete scientific inference rises to that standard. So all scientific reasoning must be inductive. Therefore, the distinction between induction and deduction ultimately provides no useful distinctions concerning the various inferential tasks in science.

3 A Response

One could leave it at that, and place the distinction between induction and deduction on the shelf of failed concepts in the philosophy of science. A more engaging response is that the tasks usually identified as deductive in science are *nearly* deductive, but for a small chance of error that can often be estimated and controlled, whereas the tasks that are considered inductive come with no bounds on probable error whatsoever [Gelman and Shalizi, 2013, Ionides et al., 2017]. In particular, an inferential procedure is **truth preserving in chance** iff it has the property that the chance of drawing a false conclusion is low, given that the premises are true. It is **monotonic in chance** iff the chance of producing a conclusion is guaranteed not to drop by much, given further information.

Those slightly weakened versions of truth preservation and monotonicity can serve as useful guides for understanding what is at stake in the various inferential tasks faced in concrete scientific inferences. The revised taxonomy is:

inference:	deductive:	truth preserving and monotonic *in chance*
	inductive:	everything else.

On that slightly revised taxonomy, deduction does apply to such problems as calculation, refutation of a universal law, verification of an existential law, deciding between two universal laws, deriving predictions from a law, and determining which laws are logically compatible with the given evidence. And one really does need to resort to induction to infer a universal law from the observations, to choose among an infinite range of universal laws, etc. Thus, the revised distinction between induction and deduction really does provides a revealing, high-level, epistemic distinction among scientific activities.

4 An Objection

One potential objection is that statistical deduction in the sense proposed is nothing like traditional, logical deduction, so that traditional insights have essentially nothing to do with what is going on in statistical inference. That would be a compelling objection, if true. Our main object is to show that it isn't—that the structural insights from the traditional view, based on propositional information, really do transfer, in a very direct and precise way, to statistical problems and to statistical inference procedures. In short, we propose a firm, mathematical bridge between traditional philosophy of science, based on logic and propositional information, and statistical inference, based on random samples. Given such a bridge, the epistemological and statistical traffic can flow back and forth with relative ease.

The relevant, common structure is **topological**. That may sound a bit jarring to the philosophical ear, which is habituated to associate deduction and induction with logic and probability, but it is very apt in the classical setting in which propositional information bears logical connections with scientific hypotheses. In that setting, the space of possible, propositional information states one might encounter generates an **information topology** over epistemically possible worlds, in which topological "closeness" is a matter of informational distinguishability. The deductively verifiable propositions are just the open sets in that topology. One can also characterize refutability, decidability, learnability in the limit, empirical simplicity, and Ockham's razor in terms of the information topology.

Since the information topology is based on propositional information states and their logical connections with hypotheses, there appears to be no corresponding notion in statistics, in which propositions about samples are logically independent from the hypotheses of interest. Our strategy is to sidestep that difficulty by identifying the unique topology on possible statistical worlds that characterizes verifiability *in chance*. Then refutablity in chance, decidability in chance, learnability in the limit in chance, empirical simplicity of statistical hypotheses, and a statistical version of Ockham's razor, all using the same definitions as in the classical setting that assumes deductive relations between hypotheses and propositional information.

5 Information Topology

Let W be a set of **possible worlds** of interest. The sense of possibility is broadly epistemic and pragmatic—they are the worlds one takes seriously as possibilities of error, and one does not take it as settled, a priori, which of them is actual. A proposition is identified, for present purposes, with the set of possible worlds in which it is true. Hypotheses and information are both assumed, for now, to be propositional. An **information state** is the conjunction of all information available in a given, possible situation. Let $\mathcal{I}$ be the set of all possible information states corresponding to all worlds and all situations, and let $\mathcal{I}(w)$ denote the set of all information states in $\mathcal{I}$ that are true in w. Consider the propositional structure $\mathfrak{I} = (W, \mathcal{I})$. Say that $\mathfrak{I}$ is an **information**

space iff the following, three assumptions are met.

B.1. The information states in $\mathcal{I}$ cover W.

B.2. Any two information states in $\mathcal{I}$ that are true in world w are entailed by an information state that is true in w.

B.3. There are only countably many distinct information states in $\mathcal{I}$.

Assumption B.1 is a conceptual truth. In every situation in a world, one is in some information state—at worst the uninformative information state W. Assumption B.2 follows from an unstated but intended **plenitude** assumption that information *accumulates*. For if one is in true information state E, and there is some true information state F that does not entail E, then the only way to receive further information that entails F is to enter a true information state G that entails both E and F. Assumption B.3 reflects the Turing-like observation that only a discrete range of possible situations are distinguishable by us, and what is not distinguishable is not information. By definition, an information space is just a countable **topological basis**. That directly connects topology to propositional information.

Let H be the proposition expressed by some hypothesis. The underlying interpretation of $\mathfrak{I}$, is that if E is an information state true in w, then diligent inquiry will turn up information F entailing E eventually. For such an inquirer, one can say that hypothesis H **will be verified** in w iff there exists information state E in $\mathcal{I}(w)$ such that E entails H—i.e., E is a subset of H. In topology, w is then said to be an **interior point** of H. The **interior** of H, denoted by $\mathsf{int}\, H$, is defined to be the set of all interior points of H. Therefore, $\mathsf{int}\, H$ is the proposition *that* H will be verified, so int can be viewed as the modal operator "it will be verified that".

Hypothesis H is **verifiable** iff H will be verified in every world in which H is true; i.e., if H entails $\mathsf{int}\, H$. That is one of the many equivalent definitions for saying that H is a topologically **open set** relative to the topological basis $\mathfrak{I}$. The open sets are just the verifiable propositions. Let $\mathcal{V}$ denote the collection of all open propositions relative to $\mathfrak{I}$. Then it is a basic topological fact that $\mathfrak{V} = (W, \mathcal{V})$ is a topological space.

Verifiability can also be understood methodologically. A propositional inference **method** M is a propositional operator that takes information states to conclusions. Method M is **infallible** iff there is no world that presents information in which the method draws a false conclusion. Method M is a **verifier** of H iff M infers only H or W, is infallible in every world, and converges to the conclusion H in each world in which H is true. Then H is methodologically verifiable iff H is informationally verifiable (open). The verifier M simply waits for verifying information and then concludes that H.[1] **Closed** sets are defined to be the complements of open sets, so they are the

[1] The strategy for M is also monotonic, since it never drops H after having inferred it.

refutable propositions. **Clopen** sets are both open and closed, so they correspond to decidable propositions. As in the preceding paragraph, one can define methodological refutability and decidability, which are equivalent to purely informational refutability and decidability.

The dual operator to interior is the closure operator $\mathsf{cl}\, H$. It says that not-H will never be verified, which is just to say that H will never be refuted. Of course, H is never refuted if H is true. What is bad is if H is never refuted even though H is false. Then a Popperian might say that H poses the **problem of metaphysics** in w. The proposition that H poses the problem of metaphysics is called the **frontier** of H in topology:

$$\mathsf{frnt}\, H \;=\; \mathsf{cl}\, H \cap \neg H.$$

Again, following Popper, hypothesis H poses the **problem of induction** in w iff H is true, but will never be verified. So the proposition that H poses the problem of induction is just $\mathsf{frnt}\neg H$. If H poses neither the problem of metaphysics nor the problem of induction, then H will be decided one way or the other; otherwise not. The proposition that H will never be decided is called the **boundary** of H, definable by:

$$\mathsf{bdry}\, H \;=\; \mathsf{frnt}\, H \cup \mathsf{frnt}\, \neg H.$$

The frontier of H and the frontier of $\neg H$ partition the boundary into the problem of metaphysics and the problem of induction, respectively—the two fundamental problems with which Popper begins *The Logic of Scientific Discovery* .

Popper thought the problem of induction is insuperable for universal theories, and topology concurs. His response to the problem of metaphysics was to favor more falsifiable hypotheses. It is a standard topological fact that H is closed (refutable) iff $\mathsf{frnt}\, H$ is empty—i.e., iff there is no possible problem of metaphysics for H. Thus, Popper's philosophy makes perfect topological sense.

Popper did make one big mistake. Popper proposed that A is **as simple** as B iff the **potential falsifiers** of B are included among those for A. For us, a potential falsifier of A is just an information state disjoint from A. Taking the contrapositive, the information states compatible with A are included among the information states compatible with B. But then B is never falsified if A is true. So Popper's simplicity order (less than means simpler) is topological:

$$A \preceq B \;\text{ iff }\; A \subseteq \mathsf{cl}\, B,$$

which says that A entails that B will never be refuted. Alas, $\mathsf{cl}\, A$ includes not only worlds in which A is false and never refuted, but also all worlds in which A is not refuted because A is *true*. But worlds in which A is true do *not* make A metaphysical, else the tautology W would be maximally metaphysical and complex, which it clearly is not.

The evident remedy is to replace $\mathsf{cl}\,B$, which says that B will never be refuted (possibly because B is true) with $\mathsf{frnt}\,B$, which says that B poses the problem of metaphysics.

$$A \preceq B \;\text{ iff }\; A \subseteq \mathsf{frnt}\,B.$$

The revised definition has many virtues. First, every refutable (closed) proposition is maximally simple. Second, W is trivially closed, so it is also maximally simple. Third, the definition is notationally invariant (grue-proof). Fourth, the definition can be glossed as saying that the problem of induction arises **from** A to B, since however A is true, all resulting information is compatible with the possibility that B is true *and* A is false. That explains directly the relevance of simplicity to inductive reasoning.

Popper was right that favoring refutable hypotheses at each stage avoids the problem of metaphysics. He was also right that, given that one faces the problem of metaphysics, one *could* end up stuck with a metaphysical falsehood forever. But it hardly follows that the *best* truth-finding strategy is to favor refutable hypotheses. Science has repeatedly dealt with violations of Ockham's razor by giving up on complex hypotheses whose complex effects have taken *too long* to appear—e.g., failure of the classically predicted ether drift to appear prior to 1905. Since one can recover from error either way, favoring the simpler hypothesis over the complex one sounds like a mere case of robbing Peter to pay Paul. So what *mandates* favoring the simpler theory, so far as finding the truth is concerned? Call that **Popper's gap**.

Here is an answer, based on the revised definition of simplicity. Suppose that, contrary to Ockham's razor, one were to favor B more complex than A when A is still compatible with available information. Then there is a relevant possibility w in which A is true and B is false. On pain of not converging to the truth at all, the scientist must eventually reverse opinion from B to A in w. But since the scientist can only see the available information, there is some information state E true in w on which the scientist reverses opinion back to A. *By the definition of simplicity*, information E does not rule out A. So there is some world in which A is true such that, on pain of not converging to the truth at all, the scientist reverses opinion *back to* B. So the Ockham violator's reward for favoring complex B over simple A is a cycle of opinion from B to A to B in a world in which the original conclusion B was *true*. It is not, after all, a matter of robbing Peter to pay Paul. Prematurely favoring Peter robs Peter! So given that one converges to the truth at all, avoidance of cycles of opinion (a kind of monotonicity requirement) mandates Ockham's razor.

Most scientific hypotheses are neither verifiable nor refutable, but have a less familiar topological property of great methodological importance. Consider the hypothesis H, which says:

$$Y \;=\; \alpha X^2 + \beta X.$$

Suppose that the truth is:

$$Y \;=\; \beta X.$$

Hypothesis H is not verifiable, because finitely many inexact observations along a parabola are compatible with a cubic function with a very small cubic term. Hypothesis H is not refutable, because the truth might be linear, in which case inexact measurements would never rule out arbitrarily flat parabolas. But H does have this important property: however H is true, one receives, eventually, information ruling out all simpler laws, after which H would be refuted if H were false. In general, say that H is **verifutable** iff H entails that H will *become* refutable in light of further information. In topology, H is said to be **locally closed**. Local closure is the characteristic epistemological property of concrete, scientific hypotheses and models.

Scientific paradigms or research programs are supposedly not even verifutable—they must be articulated with auxiliary assumptions of increasing complexity to make them verifutable. That familiar idea motivates the concept of a **limiting open** proposition, which is a countable union (disjunction) of locally closed propositions that may be viewed as its possible, concrete articulations.

Limiting open propositions are closely connected to the concepts of epistemic access, empirical underdetermination, and learnability. Say that H is **decidable in the limit** if there exists a method that, in every possible world, converges to H if H is true and to $\neg H$ if $\neg H$ is true. It is a basic result, proved independently by a number of authors in philosophy and informatics (de Brecht and Yamamoto [2009], Genin and Kelly [2015], and Baltag et al. [2015]), that H is decidable in the limit iff H and $\neg H$ are both countable unions of locally closed sets—i.e., iff H and $\neg H$ are both research programs! That implies that the catch-all hypothesis "neither H nor $\neg H$" is off the table, either by presupposition or by assumption, and explains why science typically focuses on competitions between two salient research programs. If the catch-all hypothesis is taken seriously, one can still **verify** research program H **in the limit**, in the sense that M converges to an articulation of H iff H is true. Otherwise, M may cycle forever through alternative articulations of H. The converse is also true—verification in the limit is demonstrably possible only for research programs.

6 Transition to Statistics

Having illustrated the topological structure behind traditional issues in the philosophy of science, we turn to the problem of transferring that entire story to statistics, by identifying the unique topology on probabilistic worlds that characterizes statistical verifiability. Then, although there is no such thing as propositional information in statistics, it is nonetheless just *as if* there were.

In statistical inference, one starts with a set Ω of **sample points**. Let $\mathcal{B}$ be a topological basis on Ω and let $\mathcal{F}$ be the sigma field that results from closing $\mathcal{B}$ under countable union and complementation. The set W of possible statistical worlds is some collection of probability measures over measurable space $(\Omega, \mathcal{F})$. Evidence consists of a random sample of size n from Ω, so there is typically no logical connection between the sample and the world from which it is sampled—anything goes.

A **statistical inference method** M draws propositional conclusions from samples of arbitrary size, rather than from propositional information.[2]

Let $\alpha > 0$ be a given, small tolerance for chance of error. Statistical inference method M **statistically verifies** H at error bound $\alpha > 0$ iff in each world w:

1. M produces only conclusion H or conclusion W;
2. M converges in chance to conclusion H if H is true;
3. M concludes W with chance $> 1 - \alpha$ if H is false.

The bound α is said to be the **significance level** of M. Say that H is **statistically verifiable** iff there exists an α verifier for H, for every $\alpha > 0$.[3]

Method $M$is a **limiting** statistical α verification method for H iff in each world w:

1. M converges in probability to a subset H' of H, if H is true;
2. The chance that M infers H' converges to zero, for each subset H' of H, if H is false.

Here is the main result.

Proposition 1 (Genin and Kelly [2017])**.** *Under a weak and natural condition on W,*[4] *there exists a unique topology on W such that:*

1. *H is statistically verifiable iff H is open;*
2. *H is verifiable in the limit iff H is a countable disjunction of locally closed sets.*

[2]There is one further regularity condition. For each answer produced by the method, there is the **zone** of possible samples over which the answer is produced. In all real applications of statistics, the probability that a sample hits exactly on the boundary between two such zones is zero, so we require M to have that property.

[3]Nothing in the following development goes wrong if one also requires α to go monotonically to 0 with sample size.

[4]The "weak and natural condition" is that the underlying topology on which probability measures are defined has a basis for which every probability measure assigns probability zero to the boundary of every basis element. Call that the **feasible basis** condition. For example, if the sample space is the real line with open intervals as the basis, then every continuous distribution assigns zero probability the endpoints of a given open interval. If the sample space is discrete, then there is already a clopen basis, so every collection of measures satisfies the feasible basis condition.

The final question concerns the identity of the unique information topology of statistical inference. If it is so important, what *is* it? It turns out to be the most familiar and widely employed topology on probability, called the **weak topology**. Every topology is uniquely definable in terms of its closure operation. For the weak topology, world p is in $\mathsf{cl}\, H$ iff there exists a sequence of statistical worlds $(p_n : n \in \mathbb{N})$ in H such that for all statistical methods M:

$$\mathsf{lim}_{n\to\infty} p_n(M \neq H) \;\to\; p(M \neq H).$$

Thus, the weak topology is just the topology that tracks the problem of induction for arbitrary, feasible statistical methods.

With that bridge in place, the topological insights of traditional philosophy of science flow with only minor resistance into the statistical domain. For example, H is statistically verifiable in the limit iff H is a countable union of propositions that are locally closed in the weak topology; H is statistically refutable in the limit iff $\neg H$ is limiting open; and H is statistically decidable in the limit iff H and $\neg H$ are both countable unions of locally closed propositions.

Another illustration is Ockham's razor. Define empirical simplicity just as before, but in terms of the weak topology on probability measures. It is still the case that simpler hypotheses pose the (statistical) problem of induction with respect to more complex ones, so the underlying motivation is exactly the same. It remains to define a statistical version of Ockham's razor. It doesn't work to speak of simplest hypotheses given current information, since all hypotheses are. But one can still define Ockham's α-razor as the principle that the method's chance of producing an answer more complex than the *true* answer is bounded by α, in all possible worlds—a principle that happens to be equivalent to the usual version of Ockham's razor in the case of propositional information. An α-cycle of opinions is performed (in chance) if the chance of producing answer H drops by more than α, the chance of producing some alternative hypothesis G rises by more than α thereafter, and then the chance of producing H again rises by more than α. Then:

Proposition 2. *If you violate Ockham's α-razor, then:*

1. *either you fail to converge to the truth in chance at all,*
2. *or nature can force you into an avoidable α-cycle of opinions in a world that satisfies the hypothesis on which the violation occurred.*

7 Conclusion

The trouble with the traditional distinction between inductive and deductive inference is that it fails to track crucial epistemological analogies between propositional and statistical reasoning. Indeed, it has become scientifically irrelevant—one doesn't even see the terms in statistics and machine learning. We have proposed what we take to be a more useful way to talk. On that proposal, the distinction between statistical inductive and deductive inference is ubiquitous and revealing. Rejecting a point null hypothesis is deduction; accepting it is induction. Confidence intervals are deduction. Non-experimental causal orientation of linear causes is induction. Estimating correlations is deduction. Inferring independence is induction. Estimating the parameters of a model is deduction. Inferring a model is induction. Experimental causal orientation is deduction. Actual prediction is deduction. Counterfactual prediction is induction. These are fundamental epistemic facts that should be familiar to everyone employing the techniques.

Regarding future work, we would like the proposed analogy between propositional and statistical deduction to extend to monotonicity. Monotonicity is a recognized feature of deductive logic. It is also attracting attention in statistics, since the probable (true) conclusion of a non-monotonic method may be less probable as sample size increases [Chernick and Liu, 2012]. As defined above, a statistical verification method need not converge to H monotonically in chance. Our current results suggest that monotonicity in chance up to a slop factor α is achievable.

References

Alexandru Baltag, Nina Gierasimczuk, and Sonja Smets. On the solvability of inductive problems: a study in epistemic topology. In *Proceedings of the Fifteenth Conference on Theoretical Aspects of Rationality and Knowledge*, 2015. doi: 10.4204/EPTCS.215.7.

Rudolf Carnap. *The Continuum of Inductive Methods.* The University of Chicago Press, 1952.

M. Chernick and C. Liu. The saw-toothed behavior of power vs. sample size and software solutions. *The American Statistician*, 2012.

Matthew de Brecht and Akihiro Yamamoto. Interpreting learners as realizers for Σ^0_2-measurable functions. *(Manuscript)*, 2009.

Andrew Gelman and Cosma Rohilla Shalizi. Philosophy and the practice of bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66(1):8–38, 2013.

Konstantin Genin and Kevin T. Kelly. Theory choice, theory change, and inductive truth-conduciveness. In *Proceedings of the Fifteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, 2015.

Konstantin Genin and Kevin T. Kelly. The topology of statistical verifiability. In *Proceedings of the Sixteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, 2017.

C. G. Hempel. *Aspects of Scientific Explanation.* Free Press, 1965.

Edward L Ionides, Alexander Giessing, Yaacov Ritov, and Scott E Page. Response to the ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 71(1):88–89, 2017.

Shailesh Niranjan and James F Frenzel. A comparison of fault-tolerant state machine architectures for space-borne electronics. *IEEE Transactions on Reliability*, 45(1): 109–113, 1996.

Karl R. Popper. *The Logic of Scientific Discovery.* London: Hutchinson, 1959.