Statistics and Induction
(Long and Loosely Organised Version)

Palash Sarkar

Indian Statistical Institute, Kolkata

June 29, 2017

"Statistics is the universal tool of inductive inference, research in natural and social sciences, and technological applications.

Statistics, therefore, must always have purpose, either in the pursuit of knowledge or in the promotion of human welfare."

– Prasanta Chandra Mahalanobis
(2nd December, 1956)

What is knowledge?

What is knowledge?

"Extreme precision of definition is often not worthwhile, and sometimes it is not possible – in fact mostly it is not possible"

– Richard P. Feynman, The Uncertainty of Science (1963)

Means of Valid Knowledge

- Perception
- *Inductive inference*
- Deductive inference
- Analogy and comparison
- Testimony (of authority)

Means of Valid Knowledge

- Perception
- *Inductive inference*
- Deductive inference
- Analogy and comparison
- Testimony (of authority)

> *"All that passes for knowledge can be arranged in a hierarchy of degrees of certainty, with arithmetic and the facts of perception at the top"*
>
> – Bertrand Russell, Philosophy For Laymen (1950)

The Method of Induction

Examples of Inductive Inference

The sun rises in the East.

All human beings are mortal.

Where there is smoke there is fire.

Examples of Inductive Inference

The sun rises in the East.

All human beings are mortal.

Where there is smoke there is fire.

If you do not carry an umbrella then it rains (?)

Women have fewer teeth than men (?)

Examples of Inductive Inference

The sun rises in the East.

All human beings are mortal.

Where there is smoke there is fire.

If you do not carry an umbrella then it rains (?)

Women have fewer teeth than men (?)

Enumerative induction or universal inference

Inference from particular inferences/observations:

"$a_1, a_2, \ldots, a_n$ are all Fs that are also G"

to a general principle: "All Fs are G."

Other Inductions

All observed rubies have been red.
Therefore, the next ruby to be observed will also be red.

Induction with a general premise and particular conclusion.

Other Inductions

All observed rubies have been red.
Therefore, the next ruby to be observed will also be red.

Induction with a general premise and particular conclusion.

Mercury is spherical, Venus is spherical, Earth is spherical, ...
Therefore, the next yet to be discovered planet is also spherical.

Singular predictive inference.

- Induction with particular premises to a particular conclusion.
- The middle step of generalisation is dispensed with.

Hume on Induction

"instances of which we have had no experience resemble those of which we have had experience"

– David Hume, Treatise of Human Nature (1738)

- Suggesting uniformity of nature.
- Logically false:
  'All observed *F*s have also been *G*s'
  and
  '*a* is an *F*'
  do not imply
  '*a* is a *G*'.

Mill's Methods of Induction

- Method of agreement.
- Method of difference.
- Joint method of agreement and difference.
- Method of residues.
- Method of concomittant variations.

Abductive Reasoning

Introduced by Peirce.

- Inferring cause from the observed effect.

Observation: The grass is wet in the morning.
Inference: It rained during the night.

- Inference to the best explanation (IBE).
- Use of Occam's razor to choose from a variety of possible causes.
- Abductive reasoning is the logic of pragmatism (pragmaticism).

Taxonomy of Inductive Inference by Carnap

Rudolf Carnap (1891-1970)

- *Direct inference:* infers the relative frequency of a trait in a sample from its known relative frequency in the population.
- *Predictive inference:* inference from one sample to another not overlapping the first;
  special case: singular predictive inference, where the second sample is a singleton.
- *Inference by analogy:* inference from traits of one individual to those of another based on the traits they share.
- *Inverse inference:* infers something about a population from the premises about a sample.
- *Universal inference:* infers a hypothesis of universal form based on a sample.

Features of Inductive Inferences

Ampliative (Peirce).

- "amplify" and generalise our experience;
- broaden and deepen our empirical knowledge.

Contingent.

- Conclusions of an inductive inference do not follow necessarily from the premises.

Non-monotonic.

- Adding true premises to sound induction may make it unsound. So, inductive inference should take into account all available data.

Non-preservation of truth.

- True premises may lead to false conclusions.

Occam's Razor

William of Ockham (c. 1287-1347): 'Among competing hypotheses, the one with the fewest assumptions should be used.'

> *"Nature operates in the shortest way possible."*
>
> – Aristotle

> *"When you hear hoofbeats, think of horses not zebras"*
>
> – Theodore Woodward (1940s)

Occam's Razor

William of Ockham (c. 1287-1347): 'Among competing hypotheses, the one with the fewest assumptions should be used.'

> *"Nature operates in the shortest way possible."*
>
> – Aristotle

> *"When you hear hoofbeats, think of horses not zebras"*
>
> – Theodore Woodward (1940s)

> *"But it is just this characteristic of simplicity in the laws of nature hitherto discovered which it would be fallacious to generalize, for it is obvious that simplicity has been a part cause of their discovery, and can, therefore, give no ground for the supposition that other undiscovered laws are equally simple."*
>
> – Bertrand Russell, On Scientific Method in Philosophy (1914)

Pragmatic Motivation

Among competing hypotheses/methods choose the one which has proved most useful in the past.

- Something has proved useful in the past is used to justify that it will be useful in the future too.
- This justification is based on induction.
- The principle is used for determining one among several *methods*. A second order induction.

The Problem of Induction

How to distinguish reliable from unreliable inductions?

- An inductive inference method cannot be justified using deductive logic.
- Justifying an inductive inference method from past experience would amount to *petitio principii*.

The Problem of Induction

How to distinguish reliable from unreliable inductions?

- An inductive inference method cannot be justified using deductive logic.
- Justifying an inductive inference method from past experience would amount to *petitio principii*.

Consequences:

- Procedural/epistemological: there is no *method* which can distinguish good from bad inductions.
- Fundamental/metaphysical: there is no *objective difference* between reliable and unreliable inductions.

Induction in Indian Philosophy

Schools which admit inductive inference as a valid means of knowledge:

- Nyaya (Udayana, Gangesa)
  (and the other orthodox schools: Samkhya, Yoga, Vaisheshika, Purva Mimansa, Vedanta (Uttar Mimansa)).
- Buddhism (Dignaga, Dharmakirti), Jainism.

The Carvaka school rejected induction.

- The works of Carvaka and his followers are not available.
- Their arguments are known from the summary of these arguments made by the Nyaya philosophers.

Jayarasi Bhatta rejects perception, induction and testimony.

- It is not clear whether he was a Carvaka.

Carvaka Criticisms of Induction

Problem of infinite regress or petitio principii.

- Predates the Humean critique of induction.
- Presented in more metaphysical terms.

Presence or absence of adjuncts (hidden issues).

- Example: From observing a few brittle earthenware one infers ”All earthenware is brittle”;
  this ignores the way the earthenware has been baked.
- One can never be sure that all adjuncts have been eliminated.

Multiple observations do not provide more support.

- If a single observation does not provide support, then neither does multiple observations.
- Metaphysical: If the ‘content’ is not present in a single observation, then it is not present in multiple observations.
- Generalisations supported by multiple observations could be false.

Induction and Expectation: The Carvaka View

After seeing smoke one searches for fire. Is this not inference?

- It is necessary to go beyond what is perceived and form opinions about the past and *expectations* about the future.
- A search for unperceived fire after seeing smoke is based on expectation.
- It is both unnecessary and unjustified to claim that there is inferred knowledge of fire.
- Expectation is a doubt one side of which is stronger than the other; if both sides were equally matched, expectation would not lead to action.

Induction and Expectation: The Carvaka View

Suppose search for fire leads to a fire. So, what was expected is now perceived. Being perceived, there is knowledge of fire and the acceptance of inference as a source of knowledge is necessary.

- No. The success of action prompted by expectation does not turn expectation into knowledge.
- Such success generates confidence in expectations and make them appear as knowledge.
- Appearing as knowledge is all that is required to initiate action.

Modern Notions and the Carvaka View

Anticipation of some modern notions:

- The frequentist view of probability is anticipated when mention is made that each success generates confidence in expectations.
- The quantification of uncertainty is anticipated when mention is made of two sides of an expectation.
- Pragmatism is anticipated when mention is made that expectation initiates action.

Inductive Inference: A Paradoxical Position

‘This much is certain: nothing is certain.’

Inductive Inference: A Paradoxical Position

'This much is certain: nothing is certain.'

"Dogmatism and skepticism are both, in a sense, absolute philosophies; one is certain of knowing, the other of not knowing. What philosophy should dissipate is certainty, whether of knowledge or ignorance."

– Bertrand Russell

Induction and Scientific Inference

"Aristotle, in spite of his reputation, is full of absurdities. He says that children should be conceived in the Winter, when the wind is in the North, and that if people marry too young the children will be female. He tells us that the blood of females is blacker then that of males; that the pig is the only animal liable to measles; that an elephant suffering from insomnia should have its shoulders rubbed with salt, olive-oil, and warm water; that women have fewer teeth than men, and so on. Nevertheless, he is considered by the great majority of philosophers a paragon of wisdom."

– Bertrand Russell, An Outline of Intellectual Rubbish (1937, 1943).

Empiricism in Science

Novum Organum (1620), Francis Bacon.

Suppose the intention is to discover the nature of heat.
Observation: draw up a list of hot bodies;
draw up a list of cold bodies.
draw up a list of bodies of varying degrees of heat.

Empiricism in Science

Novum Organum (1620), Francis Bacon.

Suppose the intention is to discover the nature of heat.
Observation: draw up a list of hot bodies;
draw up a list of cold bodies.
draw up a list of bodies of varying degrees of heat.
Inference:

- The lists would show characteristics always present in hot bodies, absent in cold bodies and present in varying degrees in bodies of different degrees of heat.
- By this method one arrives at a general law of the lowest degree of generality.
- From such laws one arrives at laws of the second degree of generality, and so on.

Newtonian Postulates

The Regulae (rules for philosophising) in Book III of the Principia Mathematica.

- No more causes of natural things should be admitted than are both true and sufficient to explain their phenomena.
- Therefore, the causes assigned to natural effects of the same kind must be, so far as possible, the same.
- Those qualities of bodies that cannot be intended and remitted and that belong to all bodies on which experiments can be made should be taken as qualities of all bodies universally.
- In experimental philosophy, propositions gathered from phenomena by induction should be considered either exactly or very nearly true notwithstanding any contrary hypotheses, until yet other phenomena make such propositions either more exact or liable to exceptions.

Newtonian Induction

Newton's famous phrase:

"hypotheses non fingo"
("I frame no hypotheses")

- Inductivism: the scientist was not to invent systems but infer explanations from observations.

Auguste Comte (1798–1857)

Positivism.

- Only authentic knowledge is that which is based on sense experience and positive verification.
- Different forms of the positive method: astronomy–observation, physics–experimentation, biology–comparison.

Theory-observation circularity.

- Observation involves both perception and cognition.
- Depends on our conception of the world; that conception may influence what is perceived or considered worthy of perception.
- Most scientific observation must be done within a theoretical context in order to be useful.
- Empirical observation and acceptability of a hypothesis.
  - The observation is framed in terms of a theory that also contains the hypothesis it is meant to verify or falsify.
  - So, the observation can only arbitrate between the hypotheses within the context of the underlying theory.

Karl Popper (1902-1994)

> *"[A] theory of induction is superfluous. It has no function in a logic of science."*
>
> – Karl Popper, The Logic of Scientific Discovery (1959)

Problem of induction: two formulations.

- Using emperical evidence to establish the truth of a theory.
- Using emperical evidence to choose between competing theories.

Both problems are insoluable:

- Scientific theories have infinite scope and no finite evidence can adjudicate among them.

Popper's Falsifiability

- A counter example can falsify a theory. This is a virtue in a scientific theory.
- Highly falsifiable theories are in general more informative.
- Though theories cannot be supported, then can be corroborated. A better corroborated theory is one that has been subjected to more and more rigorous tests without having been falsified.

Falsifiability is similar to the fallibilism introduced earlier by Peirce.

> *"No amount of experimentation can ever prove me right; a single experiment can prove me wrong."*
>
> – Albert Einstein

Bertrand Russell (1872-1970)

> *"[W]e must either accept the inductive principle on the ground of its intrinsic evidence, or forgo all justification of our expectations about the future."*
>
> – Bertrand Russell, The Problems of Philosophy (1912).

Human Knowledge: Its scope and limits (1948)

- Introduced five (rather complex) postulates of scientific inference. Essentially inductive inferences.
- Suggested that the assuming the validity of the postulates is sufficient for validity of scientific inference.
- The postulates may not be in their simplest form; no claim to their necessity is made.

Russell's Postulates of Scientific Inference

Quasi-permanence: 'Given any event *A*, it happens very frequently that, at a neighbouring time, there is at some neighbouring place an event similar to *A*.'

Separable causal lines: 'It is frequently possible to form a series of events such that from one or two members of the series something can be inferred as to all the other members.'

Spatio-temporal continuity in causal lines: This postulates says that 'when there is a causal connection between two events that are not contiguous, there must be intermediate links in the causal chain such that each is contiguous to the next.'

Russell's Postulates of Scientific Inference

Structural postulate: 'When a number of structurally similar complex events are ranged about a centre in regions not widely separated, it is usually the case that all belong to causal lines having their origin in an even of the same structure at the centre.'

Analogy: 'Given two classes of events *A* and *B*, and given that, whenever both *A* and *B* can be observed, there is reason to believe that *A* causes *B*, then if, in a given case, *A* is observed, but there is no way of observing whether *B* occurs or not, it is probable that *B* occurs; and similarly if *B* is observed, but the presence or absence of *A* cannot be observed.'

Richard Phillips Feynman (1918-1988)

"Philosophy of science is about as useful to scientists as ornithology is to birds."

"The great difficulty is in trying to imagine something that you have never seen, that is consistent in every detail with what has already been seen, and that is different from what has been thought of; furthermore, it must be definite and not a vague proposition."

The Uncertainty of Science (Feynman 1963)

Method of Science:

- Observation is the ultimate and final judge of the truth of an idea.
- Principle of science: “The exception proves that the rule is wrong.”
- Observing an exception spurs attempts at understanding and trying to find other exceptions and new rules.

Discipline of Observation: thoroughness, interpretation, objectivity.

Formation of Rules with the goal of prediction:

- The more specific a rule is:
  the more powerful it is (in the sense of prediction),
  the more liable it is to exceptions,
  the more interesting and valuable it is to check.

Uncertainty of Science

Feynman on scientific knowledge:

- Body of statements of varying degrees of certainty.
- Some are most unsure; some are nearly sure; but none is absolutely certain.

All scientific knowledge is uncertain!

Uncertainty of Science

Feynman on scientific knowledge:

- Body of statements of varying degrees of certainty.
- Some are most unsure; some are nearly sure; but none is absolutely certain.

All scientific knowledge is uncertain!

"Although this may seem a paradox, all exact science is dominated by the idea of approximation. When a man tells you that he knows the exact truth about anything, you are safe in infering that he is an inexact man. Every careful measurement in science is always given with the probable error ... every observer admits that he is likely wrong, and knows about how much wrong he is likely to be."

– Bertrand Russell, The Scientific Outlook (1931)

Doubt and Ignorance

Feynman calls for a satisfactory ‘philosophy of ignorance’.

- Doubt is not to be feared; it is to be welcomed.
- Feynman demands the freedom to doubt for future generations.

Doubt and Ignorance

Feynman calls for a satisfactory 'philosophy of ignorance'.

- Doubt is not to be feared; it is to be welcomed.
- Feynman demands the freedom to doubt for future generations.

> *"William James used to preach the 'will to believe.' For my part, I should wish to preach the 'will to doubt.' ... What is wanted is not the will to believe, but the wish to find out, which is the exact opposite."*
>
> – Bertrand Russell, Free Thought and Official Propaganda (1922)

Inferring Knowledge with Uncertainty

Where there is smoke there is 80% chance of fire.

- Quantification of uncertainty.
- The content of the statement pertains to uncertainty.
- A definite/certain statement about uncertainty.
- Inductive inference: ampliative, non-monotonic, contingent.
- Does not solve the problem of induction.

Calculus of Probability

- Notion of sample space by von Mises.
- Rigorous axiomatic treatment based on measure theory by Kolmogorov.

We shall no more attempt to explain the "true meaning" of probability than the modern physicist dwells on the "real meaning" of mass and energy or the geometer discusses the nature of a point.

– William Feller

Probability as Relative Frequency

Infinite sequence of causally independent identical repetitions of an experiment.

Probability of an event A is the limiting value of f_n/n where f_n is the number of times A occurs in the first n trials.

- Von Mises makes this more precise in terms of a *collective* and the stipulation that the limit should also hold for any sub-sequence that can be derived using a place selection rule.
- Later developments by Reichenbach, Fisher, Russell and others.

Problem of instantiation:

- How to interpret the probability of a successful surgery?

Probability as a Belief: Impersonal (Keynes-Jeffreys)

Probability of an uncertain proposition A can only be expressed in relation to another proposition H which represents the body of knowledge. This is written as $P(A|H)$.

$P(A|H)$ may be interpreted as the degree of belief that *any rational person* who is in possession of H will have about A.

- Keynes: probabilities of different propositions are not necessarily comparable.
- Jeffreys: with respect to the same knowledge, the probabilities of any two propositions can be compared.

Probability as a Belief: Personal (Ramsey-deFinnetti)

- Probability is specific to a particular person at a particular time.
- In assigning probability, one would draw upon one's current stock of knowledge (conscious or sub-conscious).
- Not necessary to explicitly mention the body of knowledge to which the probability relates.
- Coherence/consistency is required (Dutch book issue).

Statistical Notions and Induction

Sufficient Statistics

A *test statistic* is a function of the data.

Sufficient for an unknown parameter if “no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter.” (Fisher, 1922)

Ensures that all available information in the sample about the parameter is accounted for.

- Using less information can provide a wrong (or less accurate) inference.
- Relates roughly to the non-monotonic principle of inductive inference.

Maximum Likelihood Estimate (MLE)

Likelihood function: A function of the parameter which gives the probability of obtaining the sample given the value of the parameter.

MLE: The value of the parameter which maximises the likelihood function.

- The justification for MLE is based on induction to the best explanation.
- Given the data, the MLE of the parameter is the best explanation of the setting.

Null Hypothesis Testing

H_0: the null hypothesis;

p-value: given H_0, the probability that the test statistic equals the observation or more extreme;

α: level of signficance.

H_0 is rejected at α level of significance if the *p*-value is less than α.

Null Hypothesis Testing and Induction

- Ampliative: infer something about a hypothesis from the data.
- Non-monotonic: a hypothesis which was not priorly rejected can become rejected with the availability of additional data.
- Choice of α is based on prior experience which again involves an induction.

> *"[I]t is a fallacy, ..., to conclude from a test of significance that the null hypothesis is thereby established; at most it may be said to be confirmed or strengthened."*

– Ronald Fisher, Statistical Methods and Scientific Induction (1955)

Null versus Alternative Hypotheses

Testing for H_0 versus H_1.

- Type-1 error: rejecting H_0 when it is true;
- Type-2 error: accepting H_0 when it is false.

Aspects of inductive inference:

- Ampliative, non-monotonic and contingent.

Fisher (1955) calls such tests "acceptance procedures."

- These are different from level of significance based null hypothesis testing.
- Different from "the work of scientific discovery by physical or biological experimentation."

Model Selection

- Determine a set of possible models.
  - Inductive justification.
- Choice amongst the models.
  - Find the model that gives the 'best' prediction.
  - Find the 'true' model.
  - Find a distribution over the models.

"All models are wrong but some are useful"

– George Box (1978)

- Usefulness is a pragmatic justification which is determined by induction.

Some Model Selection Criteria

https://en.wikipedia.org/wiki/Model_selection

- Akaike information criterion
- Bayes factor
- Bayesian information criterion
- Cross-validation
- Deviance information criterion
- False discovery rate
- Focused information criterion
- Likelihood-ratio test
- Mallows's C_p
- Minimum description length (Algorithmic information theory)
- Minimum message length (Algorithmic information theory)
- Structural Risk Minimization
- Stepwise regression

AIC and BIC

Statistical model M; *# parameters k*; *data x*; *sample size n*.

$$\begin{aligned} \text{AIC} &= \ln \hat{L} - k; \\ \text{BIC} &= \ln \hat{L} - k \ln n/2. \end{aligned}$$

where $\hat{L} = P(x|\hat{\theta}, M)$ and $\hat{\theta}$ maximises the likelihood function.

- Simplicity: penalises complex models.
- MLE is justified from inference to the best explanation (IBE).
- AIC obtained by minimising Kullback-Leibler divergence; choice of KL divergence is based on induction.
- BIC obtained by maximising the posterior distribution of a model given the data (IBE/abduction).

Inductive Bias in Learning Algorithms

Learning algorithms:

- Training using examples of relation between input and output.
- Prediction of correct output on new input.

Problem of induction:

- For the new input, the output value can be arbitrary.
- Inductive bias: additional assumptions required to solve the problem.

Common Inductive Biases

https://en.wikipedia.org/wiki/Inductive_bias

Maximum conditional independence:

- In a Bayesian framework, try to maximize conditional independence.
- Used in the Naive Bayes classifier.

Minimum cross-validation error:

- To choose among several hypotheses, select the one with the lowest cross-validation error.

Maximum margin: distinct classes tend to be separated by thick slabs.

- Try to maximize the width of the boundary.
- Used in support vector machines. The assumption is that distinct classes tend to be separated by wide boundaries.

Common Inductive Biases

Minimum description length:

- Attempt to minimise the length of the description of the hypothesis.
- A form of Occam's razor.

Minimum features: Basis for feature selection algorithms.

- Delete features unless there is evidence that it is useful.
- Pragmatism.

Nearest neighbours:

- Assumption: most cases in a small neighbourhood in the feature space belong to the same class.
- Given a case for which the class is unknown, guess that it belongs to the same class as the majority in its immediate neighborhood.
- Used in the *k*-nearest neighbors algorithm.

Transduction

Formulated by Vapnik in the 1990s.

Given a set of points of which some are labelled and rest are unlabelled, to perform a labelling of all the points.

- Avoids the middle step of first inferring classes and then assigning unlabelled points to these classes.
- Infers from particular premises directly to conclusions.

Induction and Heuristics

Heuristic Method

A method to solve a problem which is not necessarily guaranteed to result in an optimal solution.

- No guarantee on the worst/average case error or, on the run time.
- Based on a rudimentary/inadequate understanding of the problem.
- Applied to problems for which methods with sufficiently good solution guarantees are not known.

Justification:

- Obtained through *trial and error*.
- Pragmatism: works in practice.

Meta-heuristics: Heuristics principles which apply to many problems.

- Justification is again inductive.

Heuristics and Integrity

Feynman on *scientific integrity*.

- Report both the positive and negative results.
- If known, details which could cast doubt should be provided.

> *"We have the duty of formulating, of summarising, and of communicating our conclusions, in intelligible form, in recognition of the right of* other *free minds to utilize them in making* their own *decisions.*
>
> – Ronald Fisher, Statistical Methods and Scientific Induction (1955)

Simplistic heuristics are "cargo-cult" inductive inferences.

An Enigmatic Inference

Mahalanobis:

> *"Statistics is the universal tool of inductive inference ...*
> *Statistics,* **therefore***, must always have purpose ..."*

Is this an inductive inference (clearly it is not deductive)?

- Force of the argument: a universal tool must have purpose.
- Is there support for such an induction?

Is this a moral (or belief based) inference?

- Then 'should' should have been used instead of 'must'.

References

- Hanne Andersen and Brian Hepburn. Scientific Method, The Stanford Encyclopedia of Philosophy (Spring 2016 Edition), Edward N. Zalta (ed.).
- Prasanta S. Bandyopadhyay and Malcolm R. Forster (eds.). Philosophy of Statistics. Volume 7 of the Handbook of Philosophy of Science, Elsevier.
- Kisor K. Chakrabarti. Classical Indian Philosophy of Induction: The Nyaya Viewpoint. Lexington Books (2010).
- Shoutir K. Chatterjee. Statistical Thought: A Perspective and History. OUP Oxford (2003).
- Jan-Willem Romeijn. Philosophy of Statistics. The Stanford Encyclopedia of Philosophy (Spring 2016 Edition), Edward N. Zalta (ed.).

References



John Vickers. The Problem of Induction, The Stanford Encyclopedia of Philosophy (Spring 2016 Edition), Edward N. Zalta (ed.).

Thank you for your kind attention!