On Statistics and Induction

Palash Sarkar

Indian Statistical Institute, Kolkata

Indian Statistical Institute Retired Employees Association
Silver Jubilee Celebrations
29th June 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 will also be 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

- *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* (in the sense of deductive logic) 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?

Is there at all any basic difference between reliable and 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* (begging the question).

The Problem of Induction

How to distinguish reliable from unreliable inductions?

Is there at all any basic difference between reliable and 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* (begging the question).

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.

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.

- 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.
- *Falsifiability* is anticipated when it is mentioned that positive verification is not sufficient for inferring knowledge.

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

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 (Mathematics) 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

Setting: 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 inference to the best explanation (IBE) or abductive inference.
- 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?

References

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

Thank you for your kind attention!