INFORMAL INFERENCE – APPROACHES TOWARDS STATISTICAL INFERENCE

Manfred Borovcnik
Department of Statistics, University of Klagenfurt
manfred.borovcnik@aau.at

The development of methods suitable to tackle the problem of inductive logic – how to justify arguments that generalise findings from data – has been signified by great controversies in the foundations and – later – also in statistics education. There have been several attempts to reconcile the various approaches or to simplify statistical inference: EDA, Non-parametric statistics, and the Bootstrap. EDA focuses on a strong connection between data and context, non parametrics reduces the complexity of the model, and Bootstrap rests solely on the data. Informal inference subsumes two different areas of didactic endeavour: teaching strategies to simplify the full complexity of inference by analogies, simulations, or visualisations on the one hand, and reduce the complexity of inference by a novel approach of Bootstrap and re-randomisation. The considerations about statistical inference will remain important in the era of Big Data. In this paper, the various approaches are compared for their merits and drawbacks.

INTRODUCTION

Controversies in the foundation of stochastics

Historically, the development of methods suitable to tackle the problem of inductive logic – how to conclude rationally from empirically data to hypotheses – has been signified by great controversies (see, Stegmüller, 1973; Hacking, 1975; Borovcnik, 1984; Gigerenzer 1993). Barnett (1982) tried to reconcile the various approaches. The controversy does not restrict itself to the dispute between R. A. Fisher with his pure significance test and Neyman and Pearson with their test policy of uniformly most powerful tests (see Barnett, 1982, or Borovcnik & Kapadia, 2015); it extends also to inference beyond the paradigm of frequentist probability, which is the main interpretation of Kolmogorov's axiomatic justification of probability. The controversy on statistical inference was fuelled by different conceptions of probability: the frequentist and the subjectivist theories of probability, which lead to different theories for statistical inference and caused a fierce controversy on the foundations of stochastics beginning from de Finetti's (1937) axiomatic approach toward probability based on (personal) preference patterns. After the translation of de Finetti's theory into English, the controversy raged in the 1950's to 1970's with no satisfying solution over it. Nowadays, at least in the practice of statistics, Bayesian (going back to de Finetti's perception of probability) and classical methods for statistical inference (based on a frequentist probability conception) are used interchangeably in the way of whatever seems to provide better models for a specific situation. That means, in practice, the decision between the two approaches is purely pragmatic while in theory, the controversy is unresolved and the community has agreed to let it unresolved. Berger (1985) with his "dry" exposition of the mathematical background paved the way to this "approach".

Controversies in statistics education on statistical inference

In 1997, a fierce discussion emerged in the *American Statistician* on statistical inference from a didactical point of view (Albert, 1997; Berry, 1997; Moore, 1997). It is obvious that from a didactical perspective, the basic concept of probability cannot be treated as a pure mathematical concept that leads to models that are sometimes better and sometime fit less well as is expressed by the famous statement of G. E. P. Box "All models are wrong but some of them are useful" (Box & Draper, 1987). What remains from all the debate is the multifaceted character of statistical inference and that it is difficult to conceptualise, understand, and teach (Carranza & Kuzniak, 2008; Borovcnik & Kapadia, 2013).

Informal ways towards statistical inference

In statistics education, there have been distinct directions to deal with the complexity: one goes back to Vancsó (2009), who develops a parallel approach to classical and Bayesian methods for statistical inference in order to make it easier for learners to understand each position and the related

In: S Budgett (Ed.), *Decision Making Based on Data*
Proceedings of the Satellite conference of the International Association for Statistical Education (IASE),
August 2019, Kuala Lumpur, Malaysia. iase-web.org/Conference_Proceedings.php

solution for the inference better. Vancsó (2013) continues this endeavour and Vancsó (2018) enriches the approach with visualisation.

The other direction to deal with the complexity of the statistical inference is to begin with a reduced situation of inference and make the developed method plausible by informal ways to "motivate" and explain the procedure (by analogies, or by visualisation, or by simulation, or by teaching methods that combine these didactical strategies). Early suggestions are Borovcnik (1996; 2006a; 2006b), refined ways to deal with statistical inference from an informal point of view are Batanero and Borovcnik (2016) or Borovcnik (2019). The approaches resume several attempts of the discipline of statistics to simplify the situation in statistical inference, especially for the reason of providing methods that have to rely on fewer assumptions so that the "wrong model" gets less severe implications on the final decisions, which are based on the result of applying the method of inference on the problem under scrutiny. The non-parametric approach of Noether (1967) is to name here as well as the ever-growing field of resampling and Bootstrap since the seminal publication of Efron and Tibshirani (1993).

Exploratory data analysis

A side-step of the historical development is marked by the Exploratory Data Analysis (Tukey, 1977). This has been perfectly received by the research community in the sense of a hypothesis generating method but is less tractable for hypothesis testing as is required in statistical inference. That has to do with the novel approach of an interactive modeller who adapts the model step-by-step by interpreting intermediate results from the analysis by the modeller's knowledge of the context of the investigated problem. By the end, the insight into the result from the knowledge on the context forms the justification of the results. Yet, this feature attributes too much responsibility to the modeller and the results have the touch of subjectivity due to the personal decisions in the modelling phase, which may cause problems in the usual case when the modeller and those who have to follow the solution of the model are distinct. As in this case, subjective acts of the modeller are "forced" upon others who – as a normal reaction to it – would then reject the modelling and the result as relevant for them.

"Informal inference"

Noether's non-parametric approach has gained more attention by the huge computer facilities, which make it easy to simulate from a large number of combinatorial possible cases, which is now called re-randomisation. Computer facilities have also reinforced the implementation of Bootstrap methods. The methods are developed, for example, in Edgington (1995), Efron (2000), or Manly (2007). A quick orientation on Bootstrap and re-randomisation is offered by didactical analyses such as Engel (2008; 2010), or Borovcnik (2019).

In statistics education, a school of "informal inference" has been established that simplifies statistical inference to the reduced situation either in re-randomisation or in Bootstrap. This development was initiated by the milestone paper of Cobb (2007) who criticised the appropriateness of the full complexity of statistical inference in the light of the new developments of the discipline of resampling. Early attempts to investigate the viability of such a reduction go back to Garfield and Ben-Zvi (2008), or Rossman (2008). Meanwhile, the approach has been extended to cover the curriculum of statistical inference (in the reduced form) over the secondary level and the introduction into statistical inference at universities (Stohl Lee, Angotti, & Tarr, 2010; delMas, 2017; Ben-Zvi, Makar, & Garfield, 2018).

Bayesian decision theory

The Bayesian approach towards decision theory (using a more general, discrete decision situation rather than the parametric model with a family of continuous distributions such as that of the normal distribution) has attracted less attention though the 1997 discussion has been fiercely in favour of Bayesian methods but as Moore (1997) stated in this discussion, the required prior distributions on the parameter of the model distribution multiplies the mathematical complexities beyond feasible teaching solutions at least at the secondary level. Yet, see Vancsó's (2008) approach who uses software for the required complex calculations and substitutes mathematical formalism by a visual approach towards the prior and posterior distributions (Vancsó, 2018); that means he uses a qualitative perception of distribution highly connected to the visual display of the distribution and illustrates the feasibility of the approach by a qualitative comparison of prior distributions (which “model” any knowledge prior to the data, which are to be analysed) with posterior distributions (which summarise the information prior to the data AND the information from the data). Stangl (2017) advocates using the Bayesian paradigm for teaching and illustrates her demand by examples that can be used also for secondary-level education.

The future of statistical inference

In the era of Big Data, methods for learning from data (Hastie, Tibshirani, & Friedman, 2009) by extracting patterns by sophisticated data-reduction algorithms, seemingly bypass inference. Yet, methods of statistical inference will even increase in relevance to identify features in data that deviate from representativeness so that the bias in them may be removed. Furthermore, empirical research will continue to investigate experiments that are specifically designed to produce data that can be analysed by methods of statistical inference in order to partially answer the raised research questions.

INFORMAL AND “INFORMAL” STATISTICAL INFERENCE

We will denote the specific approach to statistical inference that simplifies the complexity of the inferential situation to the reduced situation, which is hosted by resampling methods (which comprise the re-randomisation and the Bootstrap approach) as “informal inference”.

Didactic target of “informal” and informal inference

Didactic target of this “informal inference” is to leave out the full complexity of statistical inference once and for all in-line with Cobb’s (2007) statement that the historic approach of statistical inference is outdated with the new facilities by computer power, which makes any probability distribution superfluous as every statistical inference can now be done by resampling from the existing data so that no further assumption has to be granted except that the data originates from a random sample. Informal inference (without quotation marks) is reserved here for informal ways to illustrate and to make accessible the full complexity of statistical inference. That means that – temporarily – the situation could be restricted to the same situation as is used in “informal inference”; yet, this is characterised by a conscious step of simplification of the full situation from the outset.

For the methods of both approaches, see the vastly growing literature on “informal inference” (e.g., Ben-Zvi, Makar, & Garfield, 2018) or Borovcnik (2019). The effectivity of teaching statistical inference by “informal inference” has been the scope of studies such as Zieffler, Garfield, delMas, and Reading (2008), Makar and Rubin (2009; 2014), or Pfannkuch and Wild (2012); delMas (2017) also contributes to the results in favour of “informal inference”.

Critique on the “informal inference” approach

Critique on the approach – if seen as a universal solution for the whole statistical inference and not as a pathway towards it (as an intermediate state towards the full complexity) – has been expressed by Biehler (2014) who debates the basic problem of an “informal approach” as informal to which approach towards statistical inference as there are many (see the controversy in the foundations above and see Barnett, 1982). From the discipline of statistics, it has to be stated that – against all predictions that traditional methods of statistical inference are outdated (Efron & Tibshirani, 1993; Cobb, 2007) – classical statistical methods are still widely used, much more than resampling methods. Furthermore, the solutions of resampling do not provide the solutions that are promised: Howell (n. d.) and Lunneborg (2000) show that Bootstrap intervals are different from classical confidence intervals and there is no guarantee that they “converge” to them; they have other boundaries and other coverage properties. Re-randomisation provides no substitute for the power of statistical tests (the complement of the type-II error) as there is no way to formulate and embed an alternative hypothesis in the method. Borovcnik (2013; 2017; 2019), Kapadia and Borovcnik (2015) summarise the flaws of an “informal inference” approach.

CONCLUSION

Borovcnik (2019, p. 14) states

> “‘Informal inference’ is not inference in a mathematical sense. It is NOT an informal approach to what the discipline of statistics calls inference. It presents a rather restricted approach to making inferences

with no obvious links how to proceed from there to formal inference (unless one omits traditional statistical inference)."

- With "informal inference" it is impossible to address key issues of statistical inference (type-II error neither with re-randomisation nor with Bootstrap as it is not possible to formulate alternative hypotheses within the same framework.
- "Informal inference" reduces all statistical activities to the data; no hypotheses are any longer involved. By the sole method of simulation (resampling from the data, which is under investigation), probability is reduced to a pure frequentist concept leaving all Bayesian methods and views outside the reach of considerations.
- The approach fails with small probabilities as small probabilities (as they are usually connected to investigations of risks) have no equivalent in the given data. Either they are not represented at all or small probabilities are overestimated by the natural fluctuation of data (see Borovcnik, 2016).

Borovcnik (2019, pp. 15) summarises issues to re-consider for an "informal inference" approach. For reasons of space here, only three points are repeated:

> "'Informal Inference' is very convincing but leads to a restricted methodology that is a strict subset of statistical inference. […]
> How to continue the curriculum within such a setting? There is no path from resampling to decision theory, which is much closer to many problems of everyday concern […]. There is no connection from resampling to Bayes methods [...].
> Conceptual understanding differs from easier access and solving of tasks. Furthermore, modelling is absorbed in simulation. This may result in data as facts while models represent a hypothetical way of thinking. […]"

We suggest using resampling (Bootstrap and re-randomisation) as a transient stage to statistical inference and focus on ways of simplifying the full complexity of statistical inference. Examples that fit to such an informal approach are elaborated in detail in Batanero and Borovcnik (2016). Applets that support the approach are available from Borovcnik (n. d.). The wider field of decisions under uncertainty (Borovcnik, 2020a; 2020b) that is connected to statistical inference from a Bayesian viewpoint may serve as a final argument against narrowing the perspective in teaching.

REFERENCES

Albert, J. (1997). Teaching Bayes' rule: a data-oriented approach. *The American Statistician*, 51(3), 247–253.

Barnett, V. (1982). *Comparative statistical inference (2nd ed.)*. New York: Wiley.

Batanero, C. & Borovcnik, M. (2016). *Statistics and probability in high school*. Rotterdam: Sense Publishers.

Ben-Zvi, D., Makar, K., & Garfield, J. (Eds.) (2018). *International handbook of research in statistics education*. Cham, Switzerland: Springer International Publishing.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis (2nd ed.)*. New York: Springer.

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician*, 51(3), 241–246.

Biehler, R. (2014). On the delicate relation between informal statistical inference and formal statistical inference. In K. Makar (Ed.), *Sustainability in Statistics Education. Proceedings of the Ninth International Conference on Teaching Statistics* (2 pp.). The Hague, International Statistical Institute. Online: iase-web.org/Conference_Proceedings.php.

Borovcnik, M. (1984). *Was bedeuten statistische Aussagen* [What do statistical statements signify?] Vienna: Hölder-Pichler-Tempsky.

Borovcnik, M. (1996). Trends und Perspektiven in der Stochastik-Didaktik [Trends and perspectives in the didactics of stochastics]. In G. Kadunz, H. Kautschitsch, G. Ossimitz, E. Schneider (Eds.): *Trends und Perspektiven*. Wien: Hölder-Pichler-Tempsky, 39–60.

Borovcnik, M. (2006a). Daten – Zufall – Resampling [Data – randomness – resampling]. In J. Meyer (Ed.), *Anregungen zum Stochastikunterricht Band 3. Tagungsband 2004/5 des Arbeitskreises "Stochastik in der Schule"* (p. 143–158). Berlin: Franzbecker. Online:

www.researchgate.net/publication/304496776_Data_-_Chance_-_Resampling_Daten_-_Zufall_-_Resampling.
Borovcnik, M. (2006b). *On outliers, statistical risks, and statistical inference*. Paper presented at CERME VI. Larnaka. Online: www.researchgate.net/publication/304120534_On_Outliers_Statistical_Risks_and_a_Resampling_Approach_towards_Statistical_Inference.
Borovcnik, M. (2013). A comparative educational study of statistical inference. In *Proceedings of the 59th World Statistics Congress* (10 pp.). Voorburg: International Statistical Institute (ISI). Online: http://2013.isiproceedings.org/index.php?r3=Borovc&r4=Authors&r5=All&Search=Go&r1=NR.
Borovcnik, M. (2015). Risk and decision making: The "logic" of probability. *The Mathematics Enthusiast, 12*(1, 2 & 3), 113–139. Online: scholarworks.umt.edu/tme/vol12/iss1/.
Borovcnik, M. (2016). Probabilistic thinking and probability literacy in the context of risk. *Educação Matemática Pesquisa, 18*(3), 1491–1516. Online: revistas.pucsp.br/emp/issue/view/1658.
Borovcnik, M. (2017). Informal inference – Some thoughts to reconsider. In *Proceedings of the 61st World Statistics Congress* (6 pp.). The Hague: International Statistical Institute (ISI).
Borovcnik, M. (2019). Informal and "Informal" Inference. In J. M. Contreras, M. M. Gea, M. M. López-Martín, & E. Molina (Eds.), *Actas del Congreso Internacional Virtual de Educación Estadística* (17 pp.). Granada: Universidad de Granada. Online: www.ugr.es/~fqm126/civeest.html.
Borovcnik, M. (2020a). Risk and decision making – Fundamental aspects. In S. Lehrman (Ed.), *Encyclopedia of mathematics education* (15 pp.). New York: Springer.
Borovcnik, M. (2020b). Risk and decision making – Psychological and educational aspects. In S. Lehrman (Ed.), *Encyclopedia of mathematics education* (17 pp.). New York: Springer.
Borovcnik, M. (n.d.). *Spreadsheets in Statistics Education*. Online: wwwg.uni-klu.ac.at/stochastik.schule/Boro/index_inhalt.
Borovcnik, M., & Kapadia, R. (2013). A historical and philosophical perspective on probability. In E. J. Chernoff, & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives.* New York: Springer.
Borovcnik, M. & R. Kapadia (2015). A comparative study of statistical inference from an educational point of view. In *Proceedings of the 60th World Statistics Congress* (6 pp.). Voorburg: International Statistical Institute (ISI). Online: www.isi2015.ibge.gov.br/components/com_users/views/registration/tmpl/media/uploadedFiles/paper/2961/201/OR-D01-P4-S.pdf.
Box, G. E. P. & Draper, N. R. (1987): *Empirical model-building and response surfaces*. New York: Wiley.
Carranza, P. & Kuzniak, A. (2008). Duality of probability and statistics teaching in French education. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education.* Monterrey: ICMI and IASE. iase-web.org/Conference_Proceedings.php?p=Joint_ICMI-IASE_Study_2008.
Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum. *Technology Innovations in Statistics Education, 1*(1). Online: escholarship.org/uc/item/6hb3k0nz#page-1.
delMas, R. (2017). A 21st century approach towards statistical inference – Evaluating the effects of teaching randomization methods on students' conceptual understanding. In *Proceedings of the 61st World Statistics Congress* (6 pp.). The Hague: International Statistical Institute (ISI).
Efron, B. (2000). The bootstrap and modern statistics. *Journal of the American Statistical Association, 95*(452), 1293–1296.
Efron, B. & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
Engel, J. (2008). Bootstrap – oder die Kunst, sich selbst aus dem Sumpf zu ziehen [Bootstrap – or the art of dragging oneself out of the swamps]. *Mathematische Semesterberichte, 55*, 113–130.
Engel, J. (2010). On teaching Bootstrap confidence intervals. In C. Reading (Ed.), *Data and context: towards an evidence-based society.* Voorburg: International Statistical Institute. Online: iase-web.org/Conference_Proceedings.php?p=ICOTS_8_2010.
Edgington, E. S. (1995). *Randomization tests (3rd ed.)*. New York: Marcel-Dekker.
Finetti, B. de (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales Institut Henri Poincaré*, 7, 1–68. Foresight: Its logical laws, its subjective sources. In S. Kotz, & N.L. Johnson

(1992), *Breakthroughs in statistics. Volume I. Foundations and Basic Theory* (pp. 134–174). New York, Berlin: Springer.
Gigerenzer, G. (1993). The Superego, the Ego, and the Id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handboook for data analysis in the behavioral sciences: Methodological issues* (pp. 311–339). Hillsdale, NJ: Lawrence Erlbaum.
Hacking, I. (1975). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference (2nd ed.)*. Cambridge, MA: Cambridge University Press.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning. Data mining, inference, and prediction (2nd ed.)*. New York, NY: Springer.
Howell, D. (n.d.). Resampling statistics: randomization and the Bootstrap. *Statistical homepage David C. Howell*. Online: www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html.
Lunneborg, C. E. (2000). *Data analysis by resampling: concepts and applications.* Belmont, CA: Duxbury Press.
Makar, K. & Rubin, R. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105.
Online: iase-web.org/Publications.php?p=SERJ_issues.
Makar, K. & Rubin, R. (2014). Informal statistical inference revisited. In K. Makar (Ed.), *Proceedings of the Ninth International Conference on Teaching Statistics* (6 pp.). The Hague, International Statistical Institute. Online: iase-web.org/Conference_Proceedings.php.
Manly, B. (2007). *Randomization, Bootstrap and Monte Carlo Methods in Biology (3rd ed.)*. Boca Raton, FL: Chapman and Hall.
Moore, D. S. (1997). Bayes for beginners? Some reasons to hesitate. *The American Statistician*, 51(3), 254–261.
Noether, G. (1967). *Elements of noparametric statistics*. New York: Wiley.
Pfannkuch, M., & Wild, C. (2012). Laying foundations for statistical inference. *Proceedings of ICME 12*. Seoul. Online: www.icme12.org/upload/submission/1883_F.pdf.
Rossman, A. J. (2008). Reasoning about informal statistical inference: One statistician's view. Statistics Education Research Journal, 7(2), 5–19.
Online: iase-web.org/documents/SERJ/SERJ7(2)_Rossman.pdf.
Stangl, D. (2017). Urging a paradigm change: Why and how to train introductory statistics students in Bayesian thinking? In *Proceedings of the 61st World Statistics Congress* (6 pp.). The Hague: International Statistical Institute (ISI).
Stegmüller, W. (1973). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, vol.4, first part: Personelle Wahrscheinlichkeit und Rationale Entscheidung, second part: Personelle und statistische Wahrscheinlichkeit.* Berlin-New York: Springer.
Stohl Lee, H., Angotti, R. L., & Tarr, J. E. (2010). Making comparisons between observed data and expected outcomes: students' informal hypothesis testing with probability simulation tools. *Statistics Education Research Journal, 9*(1), 68–96.
Online: iase-web.org/Publications.php?p=SERJ_issues.
Tukey, J. W. (1977). *Exploratory data analysis*. Reading: Addison Wesley.
Vancsó, Ö. (2009). Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal of Mathematics Education (IEJME), 4*(3), 291-322.
Online: www.iejme.com/article/parallel-discussion-of-classical-and-bayesian-ways-as-an-introduction-to-statistical-inference .
Vancsó, Ö. (2013). Teaching statistical inference from multiple perspectives integrating diverging schools of inference. In *Proceedings of 59th ISI World Statistics Congress* (pp. 1108–1113). The Hague: ISI. Online: 2013.isiproceedings.org/Files/IPS112-P3-S.pdf .
Vancsó, Ö. (2018). How visualisation using software helps understanding classical and Bayesian statistics. *Invited paper in session 6C. ICOTS 10, Kyoto, 8-13 July, 2018*. Online: www.researchgate.net/profile/Oedoen_Vancso.
Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 40–58. Online: iase-web.org/Publications.php?p=SERJ_issues.