Technometrics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/utch20

The Equation of Knowledge: From Bayes' Rule to a Unified Philosophy of Science

Stan Lipovetsky

approaches demonstrate that it is possible to obtain a much more precise picture of terrorism activity and forecast it.

Each chapter suggests mathematical definitions, glossary, and additional reading sources. Besides those, the book supplies with bibliography of 153 most recent works and multiple links to the internet sites. The book presents an incredibly fascinating research, and can be interesting and useful not only to specialists but to general public for understanding and making informed judgments on terrorism and its debunking with help of statistical data analysis and prediction to prevent future attacks.

Stan Lipovetsky
Minneapolis, MN

Check for updates

The Equation of Knowledge: From Bayes' Rule to a Unified Philosophy of Science, by Lê Nguyên Hoang. Boca Raton, FL: Chapman and Hall/CRC Press, Taylor & Francis Group, 2020, xxi+438 pp., $64.95 (hbk), ISBN: 978-0-367-42815-0.

The monograph is devoted to the Bayesian statistical data analysis in the inductive reasoning to infer scientific knowledge. In contrast to the classic logic with true or false dichotomy in deductive thinking, the Bayesian-based epistemology operating with estimation of continuous probability values is more adequate to complex problems with which human mind deals in the contemporary sciences from physics and biology to social and economics disciplines. The Bayesian philosophy in approach to various scientific and practical problems (the author calls it Bayesianism) has been becoming a compelling approach in the modern framework of reasoning. The famous Bayes' rule serves for updating the knowledge credence with new empirical data which can support or contradict the achieved theories, and Bayes' formula can actually be called the equation of knowledge. Besides the pure Bayesianism, the rise of artificial intelligence through machine learning and massive data available nowadays produces the new abilities of numerous state-of-art algorithms and leads to combining computer science and probability theory into a wide variety of approaches called pragmatic Bayesianism.

The monograph consists of 22 chapters arranged in 4 sections. Section I of *Pure Bayesianism* starts with Chapter 1 *On A Transformative Journey*, where the author shares his enthusiastic discoveries on helpfulness of Bayes' law in his teaching and life experience, claims that Bayesianism is a universal philosophy of knowledge, discusses the subjective frequency as a model-dependent probability, and describes the structure of the book. Chapter 2 *Bayes' Theorem*, presents several illustrations of conditional probability estimations. One example is a known puzzle: a man has two kids, one is a boy, what is the probability that the other is a boy too? Suppressing an intuitive impulse to a solution $1/4$ or $1/2$, denote a boy as b, and a girl as g, and from all the variants of two kids bb, bg, gb, and gg, exclude the last one, then the bb probability equals 1/3. The next are the famous Monty Hall problem and its correct solution proposed by Marilyn vos Savant, the trial case of Sally Clark and the legal arguments, the reliability of a test asserting Ebola disease evaluated for the posterior conditional probability in the Bayes' formula expressed via the prior and likelihood terms. Chapter 3 of *Logically Speaking…* considers Aristotle's syllogism in logical deduction and the inference rules decomposed to the steps of universal instantiation and modus ponens, with other properties of modus tollens and contrapositive, and propositions expressed via quantifiers and predicates. Axioms of Peano for natural numbers and of Zermelo–Fraenkel (ZF) and its extension with the axiom of choice (ZFC) are noted in the context of Gödel incompleteness theorem, which asserts that there exist formulae that axioms cannot prove or disprove, including ZF and ZFC. Classical logicians, aka Platonists, typically interpret Gödel's theorem as a deficiency of axioms, so there are true theorems without proof. For intuitionists, aka constructivists, Gödel's theorem asserts that in any theory there are sentences for which no proof in favor or against it can be constructed. Besides Gödel's theorem, the intuitionists reject all nonconstructive proofs that Platonists proved, for instance, the Banach–Tarski paradox, the existence of bases of vector spaces, or the uniqueness of algebraic closures. The Bayesian deductive logic is neither classical nor intuitionist, it is not limited to 0 and 1 values, so it exceeds the true-false dichotomy, and allows a generalization to operating with different degrees of certainty. Thus, the binary language of classical logic is limited because it ignores the extent of confirmation or the magnitude of rejection. The Bayesian approach can incorporate different theories' predictions into the weighted estimates by the ensembling or bagging techniques known in machine learning. Chapter 4 of *Let's Generalize!* recaps David Hume and Karl Popper works in epistemology, Karl Pearson, Egon Pearson Jerzy Neyman, and Ronald Fisher on frequentist understanding of probability, hypotheses testing, problems with p-values for big data, and p-hacking for scientifically publishable results. The equation of knowledge, which is the Bayes formula, is suggested for checking the how adequate is a Theory to the Data:

$$P(\text{Theory}|\text{Data}) = \frac{P(\text{Theory}|\text{Data})}{P(\text{Data}|\text{Theory})P(\text{Theory}) + \Sigma_{\text{Alter}}P(\text{Data}|\text{Alter})P(\text{Alter})},$$

where Alter denotes all the alternative theories. From the prior $P(\text{Theory})$, with the likelihood $P(\text{Data}|\text{Theory})$ and partition function $P(\text{Data})$ which equals the expression in the denominator, the posterior probability $P(\text{Theory}|\text{Data})$ of the Theory supported by Data yields. In the cumulative process of integration of New collected data to refine the credence of the Theory, the Bayesian inference or update is presented as follows:

$$P(T|\text{News and } D) = \frac{P(\text{News}|T \text{ and } D)P(T|D)}{P(\text{News}|D)},$$

where T and D denote the Theory and Data, respectively. When the new data News have been obtained independently from the past data D, the Bayesian inference reduces to

$$P(T|\text{News and } D) = \frac{P(\text{News}|T)P(T|D)}{P(\text{News}|T)P(T|D) + \Sigma_{\text{Alter}}P(\text{News}|A)P(A|D)}.$$

The prior $P(T|D)$ and the alternatives $P(A|D)$ are the probabilities updated subject to the past data, where A denotes the Alter theories. The last formula defines the pragmatic Bayesian formulation where the updated prior $P(T|D)$ coincides with the posterior probability of the original Bayes formula (the left-hand side of the first formula), so the fundamental priors $P(T)$ and $P(A)$ are substituted by the probabilities updated with the new knowledge. "Today's prior is yesterday's posterior,"—quoting D. Lindley. Chapter 5 of *All Hail Prejudices*, describes power of Bayesian thinking in solving several puzzle or paradox problems. For example, the famous Conjunction fallacy originated by A. Tversky and D. Kahneman is formulated as follows: *Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.* Which is more probable: (1) Linda is a bank teller; (2) Linda is a bank teller and is active in the feminist movement. The majority of respondents chose option 2, although probability of two events is always lesser than of only one of them, so the option 1 is the correct one. Prejudices and non-informative priors are discussed in cases of pseudo-sciences and para-normal phenomena, experiments on checking the theory of relativity and genetically modified organisms, erroneous prejudices and moral questions. Chapter 6 of *The Bayesian Prophets*, describes origins of probability developing by Pascal, Fermat, de Moivre, and by Presbyterian minister Thomas Bayes with the problem on inverse probability, his followers Richard Price and Laplace, and recalls many historic anecdotes. In spite of the frequentists fight with Bayesianism in 20th century, the Bayesian approaches were successfully used by Turing in computations to crack Enigma code in WWII, supported by theoretical developments by Kolmogorov, Savage, Lindley, and others, applied by Bailey in insurance, Cornfield in study of tobacco harmful effects, Tukey and Silver in prediction of presidential elections, if to mention just a few in the infinite list of modern applications. This list continues with Bayesian implementations via the Monte Carlo technique developed by physicists working on the Manhattan Project, especially Markov Chain Monte Carlo (MCMC), with the software such as Bayesian inference Using Gibbs Sampling (BUGS), in modern deep learning and machine learning methods. Chapter 7 of *Solomonoff's Demon* contributes to the works of Ray Solomonoff who originated the algorithmic information and probability theory of inductive inference by combining Turing's theory of computation with Bayesian framework, that was actually a base for the artificial intelligence. Expressing the Church–Turing theories of computation and Kolmogorov complexity measure via the Bayesian presentation, Solomonoff introduced a new complexity theory of computation given in terms of predictive algorithmic probability. For an observed bit string $d_1, d_2, \ldots, d_n$, the probability that the next bit is 1, averaged across all *Theories* equals:

$$P(d_{n+1} = 1|d_1, \ldots, d_n) = \frac{\Sigma_T P(d_1, \ldots, d_n, 1|T)P(T)}{\Sigma_T P(d_1, \ldots, d_n|T)P(T)}$$

which is the formula of Solomonoff's induction. Concepts of completeness and computability in Solomonoff's complexity theory are discussed.

Section II of *Applied Bayesianism* reveals the hidden presence of Bayesian principals in numerous phenomena. It starts with Chapter 8 *Can You Keep A Secret?* on cryptography and encoding systems of various kind, for instance, the homomorphic encryption. Randomization in survey sampling for keeping privacy of respondents is discussed, for example, a percent of marihuana-smoking soldiers was estimated using the Bayesian rule, tuned by data pseudonymization and the differential privacy technique. Chapter 9 of *Game, Set and Math* considers examples of the assignment problem and prisoner dilemma problem of the antagonistic, cooperative and bargaining game theory, Nash equilibrium and its extension to the Bayes-Nash mixed equilibrium given by Harsanyi in the purification theorem, Bayesian mechanism design and Myerson's auction defined by the revenue-equivalence theorem. Chapter 10 of *Will Darwin Select Bayes?* describes evolution of the competing species and its modeling via the Lotka–Volterra differential equations which solution can be presented by the structure similar to the Bayes' rule. Many other processes, for example, the evolution of scientific theories, can be also expressed as Bayesian inference in disguise:

$$P_T(t+1) = \frac{\text{fitness}(T,t)P_T(t)}{\text{fitness}(T,t)P_T(t) + \Sigma_{A \neq T}\text{fitness}(A,t)P_A(t)},$$

where the probability $P_T(t+1)$ of credibility for the theory T in the step $t+1$ is defined via its prior value $P_T(t)$ and the characteristic of its fitness (birth minus death for extinguished species) among the other Alternatives A. The predictive power of traders' market shares is given in a similar presentation of an implicit Bayesian formula, and many other examples are discussed as well. Chapter 11 of *Exponentially Counterintuitive* shares thoughts on exponential growth in physics, technologies, computation, describes the Benford's law, multiplicative weights update used in Adaboost algorithm which also corresponds to an approximation of the Bayes' rule (an analogue of the last formula above). Chapter 12 of *Ockham Cuts to the Chase* considers the famous concept of this philosopher—"The multitude must not be advanced without necessity," also known as Ockham's razor. This principle is useful in applied statistical modeling and machine learning to prevent the overfitting of approximation by using redundant parameters. The cross-validation criterion is suggested for finding the right amount of fitting by the training set, together with Tibshirani's (note—there is a typo in this name in the book) regularization to penalize some measure of complexity in multiple regressions and the robust optimization. It is shown that a regularization is the consequence of the Bayes' rule presented in the logarithmic scale, and Ockham's razor can be seen as a mathematical theorem of the Bayesian paradigm. Chapter 13 of *Facts Are Misleading*, considers various examples of Simpson's paradox, correlation is not causality notion, BLM movement and accusations not supported by statistical data, and explanation for these problems via the confounding variables. Regression to the mean and Stein's paradox with identification of conciseness variables are discussed in relation to the principles of Bayesianism. Endogenous stratification, randomization methods, controlled trials, and poetic naturalism epistemological position are discussed as well.

Section III of *Pragmatic Bayesianism* starts with Chapter 14 of *Quick And Not Too Dirty*, which describes amazing features of distribution of the prime numbers, Gauss theorem on the distance between them and Riemann zeta function, and concludes that the problem in finding the new ones consists in the needed computing power which exceeds what physics allows us to calculate. Linearization helps to approximate the nonlinear complexities and to make calculations for huge possible number of combinations. Turing's imitation game and test on learning machines are described, with the modern solutions in such algorithms as linear and logistic regressions, decision trees and forest, support vector machines and Markov fields, neural and Bayesian networks. The *pragmatic Bayesian* prefers methods of realistic computation times, focusing on the most credible variants or theories which are most useful among all the options. Fourier transform, in its fast and sparse versions runs in a sublinear time even on big data. Sublinear estimates correspond to that Daniel Kahneman called System 1 of spontaneous thinking in contrast to System 2 of logical contemplation. To make the exact calculation to be feasible, the pragmatic Bayesian uses fast heuristics to approximate Bayes' rule by different innovations. Those include the following approaches: (1) using parameterized models and multiplicative weights update algorithms; (2) calculating only a high-credence model with implemented maximum a posteriori (MAP), gradient descent, expectation-maximization (EM) and generative adversarial networks (GANs); (3) ignoring the partition function in the denominator of Bayes' formula, with MCMC or contrastive divergence algorithms for still possible predictions; (4) modification of definition of conditional probability with introduction of pseudo-probabilities, and using simulated computationally bounded traders as probabilistic belief on the truth, proposed by Machine Intelligence Research Institute (MIRI); (5) taking a limited set of probability laws, in such methods as Gaussian mixture models, variational Bayesian methods, and expectation-propagation technique. Chapter 15 of *Wish Me Luck*, deals with measures of uncertainty, beginning with unpredicted by all polls Donald Trump election in 2016, then describing quantum mechanics probabilistic problems, chaos theory, unpredictable deterministic automata, and Boltzmann's laws of thermodynamics and statistical physics. Shannon's information theory with concepts of bits, entropy, optimal compression, channel capacity or redundancy, Kullback–Leibler (KL) divergence, proper scoring rules, and Wasserstein metric, useful for probabilistic predictions are discussed. The GANs, or generative adversarial networks introduced by Goodfellow and his collaborators, use the scoring weights based on KL divergence, and this approach is nowadays on the rise in deep learning and convolutional neural networks. Chapter 16 of *Down Memory Lane*, provides with contemplations on big data, the so-called secretary problem of decision making on the next step of 37% (that is $1/e$) trying samples, efficiency of algorithms for data processing, Kalman filter and hidden Markov models, long-short-term and false memory, recurrent neural networks and auto-decoder in relation to the Bayes' rule. Chapter 17 of *Let's Sleep on It*, focuses on the Bayesian networks and Markov fields, describing the latent Dirichlet allocation (LDA) which is a typical example of a Bayes network, and a hierarchical LDA adapted to big data. Monte Carlo simulations, stochastic gradient descent (SGD), pseudo-random numbers, and importance sampling for LDA via the Bayes' rule, GANs models, Lenz and Ising model, the Boltzmann machine, Metropolis-Hasting and Gibbs sampling schemes are discussed. MCMC in relation to the cognitive biases described by Tversky and Kahneman in terms of the availability bias, priming and anchor effects, and loss aversion. The algorithm of contrastive divergence for maximization of a posteriori probability for a hidden-variable, useful in the LDA and Boltzmann machine, is presented via the derivative by the parameter of the logarithm of Bayes' formula:

$$\partial_\theta \ln P(\theta|D) = \partial_\theta \ln P(\theta) + \partial_\theta \ln p(D|\theta) - E_{x|\theta}[\partial_\theta \ln p(x|\theta)],$$

where the expectation E corresponds to the total in the partitioning function. It is interesting to note that according to F. Crick and G. Mitchison, human dreams correspond such a formula of the maximized a posteriori estimation by sampling the effect of a change in brain parameters on alterations to observed data. Chapter 18, *The Unreasonable Effectiveness of Abstraction*, continues with problems of deep learning algorithmic properties measured in terms of the Solomonoff sophistication and Bennett logical depth, feature and representation learning, convolution neural networks, world vector representation and exponential expressivity, complexity and Kolmogorov sophistication. The aspects of depth, concision, and modularity of mathematics in the main formulae of mathematical and quantum physics are also presented. Chapter 19, *The Bayesian Brain*, presents some ideas on the Bayesian developments in cognitive sciences, for instance, in perception of illusions and motion, learning to talk and to count. Large paradigms of thoughts are operated by the hierarchical Bayesian approach, and the human mind performs like a sophisticated optimizer in complex Bayesian evaluations.

Section IV of *Beyond Bayesianism* is devoted to its impact on moral philosophy. Chapter 20 of *It's All Fiction*, presents the allegory of Plato's cave and the movie Matrix to discuss what is real, how to define and to prove it, if our perceptions are actually electrical signals interpreted by human brain. Multiple questions are raised, such as does life exist, does money exist, is teleology a scientific dead end, is the Church-Turing thesis versus reality correct if the entire universe can be simulated in the machine, is antirealism aka fictionalism useful? Such hard questions are legitimate, especially after the works by K. Friston who developed a model in which a Markov blanket separates the brain from the outside world but the former can reconstruct the latter by the variational Bayesian inference. Chapter 21 of *Exploring the Origins of Beliefs*, presents some funny summing of infinite divergent series and the author's personal life stories, psychological and statistical experiments, superstitions and Darwinian evolution of ideologies, again in the context of Bayesian ideology. Finally, Chapter 22 of *Beyond Bayesianism* concludes that Bayesian way of scientific thinking does not consider of what ought to be, so it is not a prescriptive moral philosophy. Individual intuitive morals are adjusted to groups' norms in a process of reinforcement learning. However, even if members of a society have coherent priorities, the famous Condorcet paradox and Arrow's impossibility theorem show that there is no way to deduce a public preference from individual preferences, that easily can lead to solutions like dictatorship. Group negotiations and collegial decisions are often made in a consensus within a small number of influential individuals. The voting systems nowadays have poor mathematical properties and are prone to fraudulence and manipulations, and some new approaches of majority judgment and randomized Condorcet voting are mentioned. Bayesian consequentialism and utilitarianism which lead to a desirable consequence and more happiness are discussed, with the precautionary principle for Bayesian decision making, and an amazing conclusion that to be a good Bayesian is a moral duty.

Each chapter is opened with a fascinating epigraph quoting famous persons, and is completed by the most recent references. There are multiple illustrations, and the Bayes' formulae are many times presented via various funny symbols of emoji kind. The book is addressed to a wide audience of students, professionals, and actually any reader interested to be better acquainted with modern ideas in various sciences and philosophy of science, and their Bayesian statistical description and interpretation. For some other books and articles on the related to the monograph topics see Lipovetsky (2012, 2017, 2019a, 2019b, 2020), Lipovetsky and Conklin (2021), and Mandel (2020).

Stan Lipovetsky
Minneapolis, MN

References

Lipovetsky, S. (2012), "Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach With Mathematica™ Support, by Phil Gregory," *Technometrics*, 54, 443. [143]

——— (2017), "Risk Assessment and Decision Analysis With Bayesian Networks, by Norman Fenton and Martin Nail," *Technometrics*, 59, 129–130. [143]

——— (2019a), "Express Analysis for Prioritization: Best-Worst Scaling Alteration to System 1," *Journal of Management Analytics*, 7, 12–27. [143]

——— (2019b), "Strategic Economic Decision-Making: Using Bayesian Belief Networks to Solve Complex Problems, by J. Grover," *Technometrics*, 61, 422. [143]

——— (2020), "Let the Evidence Speak—Using Bayesian Thinking, by Alan Jessop," *Technometrics*, 62, 137–138. [143]

Lipovetsky, S., and Conklin, M. (2021), "Bayesian Sensitivity-Specificity and ROC Analysis Finding Key Drivers, *Journal of Modern Applied Statistical Methods*, 19 (forthcoming). [143]

Mandel, I. (2020), "The Science of Statistics in Victimhood Culture: Dignity and Dishonor," *Model Assisted Statistics and Applications*, 15, 181–195. [143]

Understanding Elections Through Statistics: Polling, Prediction, and Testing, by Ole J. Forsberg. Boca Raton, FL: Chapman and Hall/CRC, Taylor & Francis Group, 2021, 225 pp., 54 b/w illustrations, $180.00, ISBN: 978-0-367-89537-2.

The monograph belongs to the *Statistics in the Social and Behavioral Sciences Series*, and it is devoted to the statistical description, modeling, and testing of the election process in its various aspects. As the author notes in the Preface, the elections "tend to be heavily influenced, both indirectly and in the form of the media, social and not—and directly in the form of the governments that hold the elections, count the ballots, and report the results," and it is easy to agree with this observation in the dramatic events of 2020 voting and election in the United States. The book consists of eight chapters combined into two parts dealing with modeling elections and testing election results on their fairness. R statistical language and packages are used in the estimations.

Part I is titled *Estimating Electoral Support*. It starts with Chapter 1 called *Polling 101* describes methods for evaluating proportions of the voting population views. On the example of the 2014 referendum on Scottish independence from England when over 100 inconsistent polls were taken, the problem of obtaining reliable information is considered. The simple random sampling (SRS) polling technique is considered when a randomly taken people from the population are contacted with a question do they support a proposition or not. SRS can be described by the binomial distribution permitting the unbiased estimation of the sample proportion and the mean square error (MSE). More estimators are given as well, including Forsberg estimator for MSE, likely sample proportions in the confidence intervals, Bayesian credible intervals obtained with priors following the beta distribution, Agresti–Coull estimator and the corresponding intervals, and SRS without replacement, and expected value and variance in the hypergeometric distribution. Mathematical derivations and illustrations in R simulations are presented. Chapter 2 of *Polling 399* continues with the stratified sampling methods which use the estimates weighted by demography (age, gender, race, etc.) and yield the reduced MSE values. An example of the polls for the 2016 US presidential election is considered, the formulas for the weighted characteristics are derived, and various numerical experiments are simulated in R. Chapter 3 of *Combining Polls* focuses on improving accuracy by the combining results of different polls, on the example of South Korea president election in 2012. Weighted averaging, averaging in time, more sophisticated schemes with different distributions, and ordinary linear regression (OLS) for predicting trends are considered. Chapter 4 of *In-Depth Analysis: Brexit 2016* employs the described approaches to the example of data for the referendum on the single question: "Should the United Kingdom remain a member of the European Union or leave the European Union?" and the voters supported the Brexit position by the vote of 51.89%–48.11%. The data were studied in time, telephone versus online, by regions, by stratified sampling and demographics.

Part II is titled *Testing Election Results*, is devoted to checking the election and the electoral system for fairness (note: there is a typo in the page of this part's title coinciding with the title of the first part). Chapter 5 of *Digit Tests* deals with the methods for detecting unfairness or fraud by the vote counts in each electoral division. The well-known Benford law (earlier stated by Newcomb in 1881, then applied by Benford in 1938 to fraud identification in data) is described and discussed in detail, with additional estimation of the mean value and standard error, and with innovative generalizations on the underlying distributions suggested by the author. Application of this test is easy: to count