Scandinavian Journal of Statistics — Theory and Applications

Scandinavian Journal of Statistics, Vol. 38: 185–194, 2011
doi: 10.1111/j.1467-9469.2011.00737.x


# On Future Directions in Statistical Methodologies – Some Speculations

ELJA ARJAS

*Department of Mathematics and Statistics, University of Helsinki and National Institute for Health and Welfare (THL)*

**ABSTRACT. We, as statisticians, are living in interesting times. New scientifically significant questions are waiting for our contributions, new data accumulate at a fast rate, and the rapid increase of computing power gives us unprecedented opportunities to meet these challenges. Yet, many members of our community are still turning the old wheel as if nothing dramatic had happened. There are ideas, methods and techniques which are commonly used but outdated and should be replaced by new ones. Can we expect to see, as has been suggested, a consolidation of statistical methodologies towards a new synthesis, or is perhaps an even wider separation and greater divergence the more likely scenario? In this talk these issues are discussed, and some conjectures and suggestions are made.**

*Key words:* Bayesian inference, frequentist inference, interpretation of probability, statistical paradigms

## 1. Big Question: how is statistics going to develop from its present status today?

The only recipe for making predictions we have available is to look at what has happened in the past, and then try to make informed guesses about the future. Generally speaking, such predictions should not involve serious personal biases. The proposals laid out below have a different character, however, and might be best described as something I would hope to be true.

A first observation concerning the past is obvious: the increase of computational power and speed, in terms of numerical and logical operations that can be carried out, has changed dramatically what actually is feasible to do – and thereby our discipline as a whole. Therefore, statistics has already changed in a fundamental manner in recent past, and no doubt continues to do so in the future. It has even been suggested that we will reach, in a not very distant future, a point of Technological Singularity. For example, Kurzweil (2005) writes as follows: '*Of the three primary revolutions underlying the Singularity, genetics, nanotechnology, and robotics, the most profound is robotics, which refers to the nonbiological intelligence that exceeds that of unenhanced humans. A more intelligent process will inherently outcompete one that is less intelligent, making intelligence the most powerful force in the universe.*'

Given the reasonable expectation that speed of computation continues to increase as rapidly as before, one may ask whether statistical methods, as we know them today, will still be needed in a more distant future. Could we not just simulate and then see what it leads to? This is a fair question, and as statisticians we must face competition from areas and backgrounds represented by computer scientists, engineers, physicists and others, for whom computer simulations are the natural mode of thinking and experimentation. If statistics is to remain a true and viable scientific discipline – and statisticians as experts representing it – it is important for us to ask what are the particular strengths that would make its survival and continued vitality possible.

In this regard, I believe that by far the most interesting development in the recent past has been the emergence of 'neo-Bayesian' statistical inference. This has not happened indepen-

This paper is based on the SJS Lecture given at the NORDSTAT Meeting in Voss Norway, June 2010.

dently from aspects relating to computation; indeed, combined with inferential ideas based on probability theory, possibilities to apply algorithmic computational methods for carrying out the numerical work have been an absolute necessity for the new developments we have witnessed.

## 2. Lindley's lectures in 1977 and their discussion

Given that this first 'SJS Lecture' is a part of the programme of the 23rd Nordic Conference on Mathematical Statistics, I think it is appropriate to look more closely into a particular event in this same series of conferences, 33 years earlier: Dennis Lindley's invited lectures at the 7th Nordic Conference on Mathematical Statistics, which was held in Ystad, Sweden, 6–10 June 1977, and later published as a paper in SJS (Lindley, 1978).

Lindley made a bold attempt to bring 'the good news' of Bayesian inference to the Nordic countries, as is evidenced by the abstract of his paper:

> *The paper reviews some of the developments in Bayesian statistics over the past seven years. The topics discussed include (i) de Finetti's work on the personalistic view and its implications for statistics (ii) operational techniques for Bayesian analyses (iii) marginalization paradoxes and proper, σ-additive distributions (iv) incoherence of non-Bayesian work (v) probability assessment by scoring-rules and by coherence (vi) the estimation of the dispersion matrix of a multivariate normal distribution and analogous multiple regression problems (vii) the problem of 'over-fitting' and the choice of numbers of variables in classificatory problems (viii) the fitting of polynomials (ix) the contrast between inference and decision-making (x) experimental design.*

The lectures were followed by a discussion, with contributions by Ole Barndorff-Nielsen, Gustav Elfving, Erik Harsaae, Daniel Thorburn, Anders Hald and Emil Spjøtvoll. These names are familiar at least to Scandinavian statisticians belonging to my generation, as teachers or as colleagues, and they have influenced in different ways how statistics has developed here as a scientific discipline during the times I have been able to follow closely myself, which is, starting from the early 1960s. I had as a student attended Elfving's introductory lecture courses on probability and mathematical statistics at the University of Helsinki, and they were the original source of inspiration that got me interested in this whole area.

But, as it turned out, the Vikings attending the Ystad conference were not to be convinced so easily, with Elfving and Thorburn being perhaps most receptive. The logical arguments presented by Lindley, in essence saying that in order to behave in a rational ('coherent') way a statistician has no other choice than become Bayesian, seem to have had rather little effect on the discussants. Erik Harsaae, in his commentary, asked: '*Why do Bayesians show so much missionary zeal?*'

I believe that most statisticians here in the North had been influenced in various ways, and not forgetting the important role which the famous monograph of Cramér (1946) had, by the writings of an earlier missionary, R. A. Fisher. Fisher (1922) had declared the use of Bayes' formula in statistical inference (then commonly referred to as 'inverse probability') to be a heresy: '*It is this last confusion, … which has led to the survival to the present day of the fundamental paradox of inverse probability, which like an impenetrable jungle arrests progress towards precision of statistical concepts.*'

Lindley's paper and its discussion take 26 journal pages, and it would be impossible to try to cover here all essential points of this exchange of ideas. But I would like to single out the following three:

- whether the Fisherian mode of statistical inference could be more appropriately thought of as representing 'objective' science than the Bayesian,

– the specification of the prior distribution and, in particular, whether there is a form of the prior which could be reasonably called 'non-informative', and
– *whether the uncertainties handled in statistical activity are all comparable, on a single scale, or whether these uncertainties fall in qualitatively distinct classes and are not, in general, comparable* (quotation from Barndorff-Nielsen's commentary).

Essentially these same arguments, with slight variations, come up today in most discussions in which these statistical paradigms are compared. Concerning the first issue, Efron (1986) states an empirically correct fact, still true in 2010, when he writes: '*The high ground of scientific objectivity has been seized by the frequentists.*' But be this as it may, in my view Draper (2006) settles the issue conclusively by: '*There is no need to apologize for the role of subjectivity in statistical analysis: ... all scientific activity that has an inferential character inescapably involves judgment (the data never really* "*speak for themselves*" *when you look closely at how scientific inferences are made), and indeed to be human is to make choices based on assumptions and judgments every waking moment.*'

The second issue, the choice of the prior, is closely related to the first and equally controversial, as the choice of the prior can be seen either as

– an advantage, by providing an opportunity to bring existing subject matter knowledge into the considered inferential problem, or
– an inherent weakness of the Bayesian approach, as the results of the data analysis (the posterior) will depend on how the prior is chosen.

In consequence, a large number of alternative approaches have been introduced to establish 'non-informative' or 'objective' prior distributions, with particular versions carrying names such as Jeffreys' prior, reference prior, or minimax prior. By comparison, the choice of the likelihood from some narrow family of parametric distributions is mostly considered to be a routine matter and rarely generates any concern, and the same holds for the specification of 'penalty' or 'cost' functions, which are used in classical inference to eliminate non-identifiability problems of parameter estimates in models of a high dimension.

The third question in the list, which is, how the concept of probability is understood and whether its use in statistical inference should be restricted to a pure frequency interpretation, is intrinsic and more interesting than the other two. Such a narrow interpretation is apparently an important reason why many people find the concept of the prior distribution to be problematic.

Let us again have a look at what Fisher (1922) had written: '*When we speak of the probability of a certain object fulfilling a certain condition, we imagine all such objects to be divided into two classes, according as they do or do not fulfil the condition. This is the only characteristic in them of which we take cognisance. For this reason probability is the most elementary of statistical concepts. It is a parameter which specifies a simple dichotomy in an infinite hypothetical population, and it represents neither more nor less than the frequency ratio which we imagine such a population to exhibit.*' And Fisher went on: '*... known as that of inverse probability ... This amounts to assuming that before A was observed, it was known that our universe had been selected at random for an infinite population in which X was true in one half, and Y true in the other half. Clearly such an assumption is entirely arbitrary, nor has any method been put forward by which such assumptions can be made even with consistent uniqueness.*'

In other words, Fisher first says that probabilities can only be interpreted as relative frequencies in hypothetical infinite populations, then he forces that interpretation to a context where it is plainly inappropriate, and finally ridicules what he just did. It would of course be a mistake not to acknowledge the fact that Fisher, in this same 1922 paper, actually lays the foundations of statistical science as it is commonly practiced today. Alternative, and more general, interpretations of probability certainly existed at Fisher's time as well as earlier, see, e.g. Hacking (1975) or von Plato (1998). Harold Jeffreys (1939), a contemporary of Fisher, summarized his views by '*The essence of the present theory is that no probability, direct, prior, or posterior, is simply a frequency*'. Even if Fisher had no respect for Bayes or Jeffreys, his 1922 paper witnesses some puzzlement about how the great genius Laplace could have thought of probability so differently than he himself, by viewing the world essentially as deterministic but then needing probability because we, as humans, do not have prefect information. According to Laplace '*Probability theory is nothing but common sense reduced to calculation*'.

Jeffreys' ideas on probability get an even more radical expression in de Finetti's famous exclamation '*Probability Does Not Exist*'. de Finetti, whose work (1974, 1975) forms the central doctrine behind Lindley's 1978 paper, is no second to Fisher in expressing sarcasm towards those with whom he disagrees: '*The numerous, different, opposed attempts to put forward particular points of view which, in the opinion of their supporters, would endow Probability Theory with a "nobler" status, or a more "scientific" character, or "firmer" philosophical or logical foundations, have only served to generate confusion and obscurity, and to provide well-known polemics and disagreements, even between supporters of essentially the same framework*.'

## 3. Emergence of neo-Bayesian statistics

But let us come back to the present time. Given the perfect wisdom provided by hindsight, and equipped with the Internet and Google, one can now see that there is ample frequentist evidence that Fisher's warnings have not been universally accepted. Here is a short list of 'hits' found by Google at the time this lecture was prepared:

| | |
|---|---|
| Bayes' theorem | 380 000 |
| Bayes' formula | 444 000 |
| Bayesian statistics | 1 120 000 |
| Bayesian inference | 615 000 |
| Prior distribution | 8 830 000 |
| Posterior distribution | 1 830 000 |

Curiously, many more people mentioned priors than posteriors. Perhaps they were unable to complete their Bayesian analysis! Moreover, in some later checks with Google somewhat different frequencies turned up. However, this shows convincingly enough that Lindley, after all, had a case. For a very thorough and informative account on these neo-Bayesian developments see Fienberg (2006).

Indeed, there has been a spectacular expansion in Bayesian statistics or Bayesian inference during the past 50 years or so. As already mentioned, the driving force behind it has been computational: in addition to advances in hardware technology, there have been major developments in algorithmic simulation methods enabling efficient numerical integration in high-dimensional spaces. An equally important factor has been the realization by many scientists that, because of its more liberal interpretation of probability as a measure of uncertainty, Bayesian statistics provides useful answers in substantive contexts where the frequency interpretation is awkward. This has led to an expansion in the application of such methods in a large number of research areas including

- medicine and health sciences
- genetics, perhaps particularly phylogenetics
- bioinformatics
- various areas of technology and engineering, e.g. signal and image analysis, pattern recognition, computer vision, and reliability and safety studies, where Bayesian methods have been recently replacing methods such as neural networks.

Emerging new areas, where analogous developments have started to gain popularity, include

- archaeology
- environmental sciences, including climate science
- estimation of natural resources, e.g. in fisheries science
- astronomy and astrophysics.

But even Lindley (1978) acknowledged that there were problems:

'*... This is important because it is clear that one of the reasons that the Bayesian argument is not as popular as it should be is the lack of workable procedures to lay alongside the formidable array of non-Bayesian techniques currently available: until a scientist can analyse a simple one-way table coherently as easily as he can incoherently, the new ideas are not going to be adopted.*'

We have now, in the 33 years that have passed from Lindley's lectures, come a long way towards having methods for carrying out such workable procedures in practice, even in the form of user-friendly software such as WinBUGS / OpenBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/) and JAGS (http://www-fis.iarc.fr/~martyn/software/jags/). Still, the idea that 'you can just feed in the data, pick up your favourite method for the analysis, and press the button' seems unlikely to ever be realized.

## 4. Some practical suggestions

Given the multitude of interesting new areas in which statistical methods are being applied, it seems to me that the ways in which many key statistical concepts are taught today need revision. Here is a list of some such issues:

- In courses of probability theory, presently important aspects relating to the interpretation of the concept of probability are either bypassed completely (as when probability is treated merely as an abstract 'measure'), restricted to (in textbook examples mostly trivial) cases of symmetry, or hidden behind circular limiting arguments via 'the law of large numbers'.
- Widely different meanings and interpretations for probability are used in practice outside of statistical theory side by side, with the colloquial meaning referring directly to a 'degree of belief' or 'plausibility'. These different modes of interpretation appear to correspond in a natural way to how the human mind works, and should be utilized more explicitly also in statistics, instead of always trying to enforce the needlessly narrow frequency interpretation.
- In most text books of introductory statistics, model parameters are characterized by the epithets 'fixed but unknown'. This conveys the message that 'fixed unknown things' could not be described in terms of probabilities. This conflicts with the colloquial meaning of the word 'probability' and, in my view, with common sense.
- The semantic confusions concerning the different interpretations of probability are only made bigger by use of the term 'random variable', giving the impression that it is necessarily something that 'varies' (supposedly over time, in repeated sampling, etc.) A better term, also suggested by Lindley, would be 'random quantity', which unfortunately has not caught on.
- Probabilities depend of the information that is available. This idea is intrinsic, and probabilities should be viewed primarily from this perspective rather than, for example, as 'propensities' of physical objects. In this sense, all probabilities are conditional, and they change when the information changes even when the underlying

physical state does not. (These ideas, and much else, are presented in the mammoth size but very readable and entertaining monograph of Jaynes (2003), largely building on the axiomatic approach to the foundations of inductive inference presented by Cox (1946, 1961). In particular, Jaynes recommends that, to explicate in language this understanding of the probability concept, one should speak of 'probability *for* (a random quantity)', instead of 'probability *of*'. The book of Jaynes deserves to be read, and studied critically, by every professional statistician. For some related comments, see the 'Epilogue' of section 7.

- Probabilities can provide useful descriptions of one's knowledge of the past (retrodiction), present (state estimation), or the future (prediction). But prediction is where using probability is hardest. This is because our limited ability to extend our imagination beyond circumstances about which we already know a great deal. Terms such as 'ignorance', 'black swans', and 'unknown unknowns', are an attempt to say something about situations in which that is not the case. Unfortunately, any hope of bringing them into the domain of explicit probability calculations seems unrealistic.
- A clear distinction, also emphasized by Jaynes (2003), should be made between physical, or even causal, dependence on one hand and logical dependence on the other. Not doing so leads to frequent misunderstandings and much confusion, because neither type of independence implies each other. Phrases such as 'iid' when used without any qualification are likely to confuse such issues further, and should rather be banned from use.
- Sequential conditioning (as in chain multiplication rule) and the concept of conditional independence need to be taught to everybody. Now the former is typically hidden in the form of the so-called 'definition of conditional probability', and the latter, a crucially important concept in statistics, does not appear in most standard text-books at all.
- A statistical model should be viewed as a means of providing an adequate description of the knowledge, which the statistician (believes he or she) has. One should be very cautious with terminology even indirectly conveying the idea that 'the data were generated by the model' or that 'the model is true' (cf. the famous quote of G.E.P. Box (1979): '*All models are wrong, but some are useful*'). Perhaps we should completely dispense with phrases such as 'true model' and 'true parameter value'? If we don't, then some people are inevitably misled into thinking that such things actually exist, in some absolute sense in Nature. (Ideas like this may easily sneak even into Bayesian model formulations, as in hierarchical mixture models, where the weights are often interpreted as 'probabilities of a particular sub-model being true'. Traps like this can be avoided only if the very idea of a 'true model' is rejected, and models are thought of as approximate expressions of our understanding of reality, formulated in mathematical language.)
- It is a consequence of the perception that probabilities are expressions of the available information, that larger data sets typically justify more complex model structures. With more information models can be updated or revised. Such ideas are put to test most directly when making predictions about future observables, as suggested, e.g. by Geisser (1971), Dawid (1984), and Gelman *et al.* (2003).
- As a general rule, the sensitivity of the results to a particular choice of a model specification may be assessed most conveniently by techniques such as cross-validation, or dividing the data into separate training and validation sets. (Again, this idea was already presented in Lindley's paper.) For a confirmatory analysis, as in clinical trials, it would be a good idea to compare the results obtained by 'optimistic' and 'pessimistic' priors concerning possible effect sizes (Spiegelhalter & Freedman, 1986).

- In my view, the efforts made towards establishing a universal concept of non-informative prior distribution in Bayesian inference have been disproportionately large compared to the practical benefits. (Fienberg, 2006 called these efforts ‘the quest for the holy grail’, which I find very fitting because, most likely, neither object being searched actually exists.) In the Bayesian paradigm the prior distribution is a part of the statistical model, and its specification depends always on the considered context and the information available. (But I am willing to admit that in hierarchical models involving several levels one will often encounter situations in which the (hyper)parameters do not have a meaningful operational definition, making the specification of their prior problematic. This seems to be an area where suitably ‘non-informative’ or ‘vague’ prior distributions could be naturally introduced.)
- Linear models have been the work horse of statistics for decades, and are still used, routinely and understandably, for example in so-called $p >> n$ problems. However, except in rare special circumstances where linear models provide genuinely accurate descriptions of the studied phenomenon, such rigid parametric functional forms of models should be used only as first crude approximations when the data are sparse. From a Bayesian perspective, they would correspond to specification of a singular prior in the space of plausible models.
- Largely based on a tradition arising from linear models, people tend to be overly concerned with the values of individual coefficients and parameters in their models, and with testing the statistical significance of some simple hypotheses related to them. Instead, they should be considering the overall shape of the functions appearing in their models, and the consequences which their estimates have to our understanding, and to predicting the values of future observables.
- Important aspects that should be accounted for in models include: Hierarchical structures (involving latent variables, which then help one make a distinction between the underlying physical reality and the measurement data), and explicit consideration of time (using formulations based on stochastic processes, an important area in probability theory which is typically completely missing from the curricula designed for training professional statisticians).
- Sometimes (as in fisheries science, and climate science) one should view a realistic assessment of uncertainty to be at least as important as striving to obtain a highly precise estimate, which in the end may turn out to be far off the mark. Combined with a sound risk-averse attitude, the pay-off from the former approach may be much bigger.
- Finally, how about also, in order to remove an unnecessary terminological barrier, getting rid of the word ‘Bayesian’ and replacing it – where some qualification still seems necessary – by ‘probabilistic’?

## 5. Future: will there be a re-conciliation or synthesis of the Bayesian and frequentist approaches?

To get an idea about what statistical practices are currently most popular, let us first check the Google ‘hits’ of some basic frequentist concepts:

| | |
|---|---|
| maximum likelihood | 8 930 000 |
| *p*-value | 96 700 000 |
| confidence interval | 5 380 000 |

In frequentist terms, clearly, Fisher is the winner here. However, I believe that the main reason behind the enormous popularity of *p*-values is that most non-statisticians who make use of them think that *p*-value = Pr(null hypothesis | data).

Those who use confidence intervals – apparently a much less popular choice, although recommended by many journal editors these days – similarly think that the 95 per cent confidence interval they have determined from their data 'contains the unknown parameter with probability 0.95'. Such interpretations are perfectly natural, and indeed, what most scientists would like to get based on their analysis of the data. But they are Bayesian concepts, cannot be obtained without specifying a full probability model (including the prior), and the reference to probability is then to the corresponding posterior. As remarked by Savage (1961), such ambitions can be described as 'a bold attempt to make the Bayesian omelet without breaking the Bayesian eggs'.

Every now and again there is an outcry about this, e.g. Ioannidis (2005). However, it seems that these warnings will have little effect on the common practices that are followed. The situation is akin to the common use of the English language today: 'Broken English' cannot be effectively controlled by Oxford, Webster's, or even BBC. A partial defence of this practice could be that a word, even when its original semantics is not understood but is misunderstood in the same way by sufficiently many, can be a useful means for communication.

The present situation, parallel existence and use of two statistical paradigms, is puzzling. But there are some understandable reasons for this which are not based on mere popular misunderstandings, or on different semantics relating to the key concepts. The following quotations from three well-known statisticians describe this situation in ways which correspond well to my own understanding:

*Why is so much of applied statistics carried out in a Fisherian mode? One big reason is the automatic nature of Fisher's theory … the working statistician can apply maximum likelihood in an automatic fashion, with little chance (in experienced hands) of going far wrong and considerable chance of providing a nearly optimal inference. In short, he does not have to think a lot about the specific situation in order to get on toward its solution. ... Bayesian theory requires a great deal of thought about the given situation to apply sensibly.* (Efron, 1986).

*It seems quite clear that both Bayesian and frequentist methodology are here to stay, and that we should not expect either to disappear in the future. ... Philosophical unification of the Bayesian and frequentist positions is not likely, nor desirable, since each illuminates a different aspect of statistical inference.* (Bayarri & Berger, 2004).

*The lack of an agreed inferential basis for statistics makes life 'interesting' for academic statisticians, but at the price of negative implications for the status of statistics in industry, science, and government. ... An assessment of strengths and weaknesses of the frequentist and Bayes systems of inference suggests that ... inferences under a particular model should be Bayesian, but model assessment can and should involve frequentist ideas.* (Little, 2006).

## 6. Summary

A brief summary of the above lengthy argumentation and speculations on future developments could be given as follows:

- The well known prediction (de Finetti, Lindley) according to which the statistical world will be Bayesian in 2020 is not going to be true, and it is unlikely to be true even in 2050!
- Frequentist methods are easier to use, and therefore suitable for exploratory data analysis and model assessment and calibration.
- Bayesian probability based methods are the preferred means for serious scientific inference, as well as a general tool for dealing with uncertainty.

- Predictive inference concerning future observables should be emphasized, thus largely replacing the apparatus of hypothesis testing relating to parametric distribution families and specific parameter values.
- As professional statisticians we should be involved in scientific work which is challenging, both with respect to methods and substance. To adequately meet this challenge, many current statistical practices need to be thrown over board.
- A major part of the activities of mathematical statisticians has been devoted to first defining a new estimator or a test, and then establishing its asymptotic optimality properties when assuming that the sample size goes to infinity. This line of research has provided work to academic statisticians for decades, but is unlikely to lead to major discoveries of practical importance in the future. The need to establish analytic asymptotic expressions is already diminished by the easy availability of reference distributions obtained by straightforward numerical simulations. The energies should therefore be re-directed, with an emphasis on serious modelling issues arising from applications, and on consequent problems relating to computation.
- Useful ideas from machine learning and artificial intelligence should be employed to a greater extent: let us join forces with computer scientists and engineers in areas which they know well.
- In the end, what counts is good science!

One of the obvious advantages – and risks – in making explicit predictions concerning future observables is that, after some time has passed, they can be verified as having been either right or wrong. Therefore, those of you in the audience listening to this lecture (or, reading this paper) who happen to be around still in 2020, or in 2030, try to make an assessment of how statistics has developed by then, and check how wrong I was! (I hope to be able to still join in. Just for curiosity, of course.)

## 7. Epilogue

A referee of this written contribution based on the lecture wanted me to clarify the terminology I used in section 4 above, and asked: "Do you mean that a physical constant, e.g. the speed of light, should be described as a random quantity rather than fixed and unknown? Of course, it is known to an accurate approximation, but not exactly as $\pi$." This is a good diagnostic question, because it shows how the referee and I are using different semantics for the word 'random quantity'. My answer is: I would view 'fixed unknown quantities' as a subset of 'random quantities'. (Whether it is a proper subset is a deep ontological question which I don't have any real competence to answer. Note, however, that if we were strict followers of Laplace, we would actually say that they are the same, by definition.) But to give a more specific answer to the question, I would treat both the speed of light and the value of $\pi$ as random quantities in the (hypothetical) situation in which I did not know their true values to enough decimals in a concrete problem I was trying to solve, I would consider this lack of information to be potentially important to the solution of that problem, and I had no source from which I could find such lacking decimals. First looking at $\pi$, I would be quite content with describing my uncertainty of the remaining unknown decimal expansion of $\pi$ with a uniform distribution on the unit interval. In reality, of course, very much is known about $\pi$ (see, e.g. http://en.wikipedia.org/wiki/Pi), including the fact that some 1.2 trillion decimals have been computed to date, which is very much more than we will ever need in a real problem. Still, I am not quite sure I share the referee's opinion that '$\pi$ is known exactly'. How about the next 1.2 trillion decimals? Interestingly, from the web page http://www.super-computing.org/pi-decimal_current.html we can see that no statistical test based on the decimals of $\pi$ that are already known can reject the simple hypothesis that treats them as independent draws (in the sense that none can be predicted on the basis of the decimals preceding it) from the multinomial distribution with probability 1/10 assigned to all digits from 0 to 9. This, then, is exactly in line with modeling an unknown infinite decimal expansion of $\pi$ as a draw from the uniform distribution on the unit interval, the choice I made above. Describing my uncertainty about the speed of light with a probability distribution is obviously more problematic, and coming up with a particular version of such a distribution, detached from an actual problem context, would be beside the main point.

## Acknowledgements

Material for this lecture was prepared while visiting the Department of Mathematics, University of Bristol, with Peter Green as my host. I am also grateful to Ulf Grenander for sharing with me some of his ideas on the future of statistics, in correspondence after the Nordstat Conference, and to Esa Läärä for having provided some good references. Finally, I am grateful to Juha Alho, Editor of SJS, for extending the original invitation to present this lecture.

## References

Bayarri, M. J. & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statist. Sci.* **19**, 58–80.
Cox, R. T. (1946). Probability, frequency, and reasonable expectation. *Am. J. Phys.* **14**, 1–13.
Cox, R. T. (1961). *The algebra of probable inference*. Johns Hopkins University Press, Baltimore, MD.
Cramér, H. (1946). *Mathematical methods of statistics*. Princeton University Press, Princeton, NJ.
Dawid, A. P. (1984). Present position and potential developments: some personal views. Statistical theory. The prequential approach (with Discussion). *J. Roy. Statist. Soc. Ser. A* **147**, 278–292.
Draper, D. (2006). Coherence and calibration: comments on subjectivity and "objectivity" in Bayesian analysis (comment on articles by Berger and by Goldstein). *Bayesian Anal.* **1**, 423–428.
Efron, B. (1986). Why isn't everyone a Bayesian? *Amer. Statist.* **40**, 1–5.
Fienberg, S. E. (2006). When did Bayesian inference become Bayesian. *Bayesian Anal.* **1**, 1–40.
de Finetti, B. (1974). *Theory of probability*, Vol. 1. Wiley, New York.
de Finetti, B. (1975). *Theory of probability*, Vol. 2. Wiley, New York.
Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. Lond. Ser. A* **222**, 309–368.
Geisser, S. (1971). The inferential use of predictive distributions. In *Foundations of statistical inference* (eds V. P. Godambe & D. A. Sprott), 456–469. Holt, Rinehart, and Winston, Toronto.
Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2003). *Bayesian data analysis*, 2nd edn. Chapman & Hall, Boca Raton, FL.
Hacking, I. (1975). *The emergence of probability*. Cambridge University Press, Cambridge.
Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med.* **2**, e124.
Jaynes, E. T. (2003). *Probability theory: the logic of science*. Cambridge University Press, Cambridge.
Jeffreys, H. (1939). *Theory of probability*. Oxford University Press, Oxford.
Kurzweil, R. (2005). *The singularity is near*. Penguin Press, London.
Lindley, D. V. (1978). The Bayesian approach (with discussion). *Scand. J. Statist.* **5**, 1–26.
Little, R. D. (2006). Calibrated Bayes: a Bayes frequentist roadmap. *Amer. Statist.* **60**, 1–11.
von Plato, J. (1998). *Creating modern probability*. Cambridge University Press, Cambridge.
Savage, L. J. (1961). The foundations of statistical inference reconsidered. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, Vol. 1, 575–586. University of California Press, Berkeley, California.
Spiegelhalter, D. J. & Freedman, L. S. (1986). A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. *Statist. Med.* **5**, 1–13.

*Received October 2010, in final form February 2011*

Elja Arjas, Department of Mathematics and Statistics, P.O. Box 68, FI-00014 University of Helsinki, Finland.
E-mail: elja.arjas@helsinki.fi