Statistical Science
2016, Vol. 31, No. 4, 545–548
DOI: 10.1214/16-STS577
Main article DOI: 10.1214/16-STS565

How About Wearing Two Hats, First Popper's and then de Finetti's?

Elja Arjas

I am grateful to Nozer Singpurwalla for having brought up so clearly and openly these puzzling, and partly confusing, foundational issues in reliability theory, and to the Editors for the opportunity to take part in this discussion.

My comments are concerned mainly with the first part of the paper. There an important distinction is made between the concepts of *reliability* and *survivability*, by linking the former to the, in some sense, *physical* or *objective propensity* interpretation of the probability concept advocated by Popper, and the latter to the *personalistic* or *subjective* probability concept of de Finetti.

There is a clear need for both types of perceptions: While the personalistic Bayesian point of view offers a systematic approach for statistical inference from data, well anchored in probability calculus, it does not make direct reference to the "true" states of the considered physical objects or systems. Such states, or changes in them, such as the repair of a defective part in a mechanical device, are of intrinsic importance in nearly all problems relating to reliability and risk assessment.

The existence of a certain gap between physical reality and a corresponding statistical modelling framework of reliability problems, even when based on the more traditional frequentist interpretation of probabilities, has been noted already much earlier. For example, thirty years ago Bo Bergman wrote in his review paper (Bergman, 1985): "However, some care has to be taken when this (repair) model is used; we have to distinguish between *statistical minimum repair*, for which the above interpretation (the equality between two failure rates) is taken as the definition, and *physical minimum repair*, in which case the failed unit is restored to the exact physical condition as it was just before the failure. These two kinds of minimum repair are not necessarily the same!"

Singpurwalla not only makes a distinction between the concepts of reliability and survivability; he also suggests that there would be a single conceptual framework which contains, and combines, objective physical entities and statistical tools, the latter based on de Finetti's personalistic approach to probability. Adopting this framework, he says, would entail a change in the current paradigm of reliability theory. This is not a modest claim.

I believe it is useful to first consider this possibility from a wider perspective, which is not restricted to reliability problems. To continue with another quotation, Philip Dawid has written (Dawid, 2004) on the relationship between the physical reality and our theories on it as follows: "I regard it as of vital importance to distinguish, carefully and constantly, between two very different universes, which I will term "intellectual" and "physical". Any kind of scientific, mathematical or logical theory is a purely intellectual construct. It will typically involve a variety of symbols and concepts, together with rules for manipulating them. The physical universe, on the other hand, just does its own thing, entirely ignorant of, and careless of, any of our intellectual theories. It manifests itself to us by means of observations".

This, I think, is a very fitting description of the situation which we face in statistics in general. In reliability problems, for example, nuclear power plants, cars or computer codes when in use, "just do their own thing", ignorant of our intellectual constructs or theories, whether they be based on Popperian propensities, on de Finetti's epistemic probabilities, or something else. The key link between the two universes, as stated above in the last sentence, is in being able to make observations on objects and processes belonging to the physical universe, and then transporting these observations into the intellectual one as data. Data consisting of registered values of observables in the physical world can be smuggled, through a back door, into the intellectual world, and then treated there in a probabilistic inferential framework as fixed values. What

Elja Arjas is Professor Emeritus, Department of Mathematics and Statistics, P.O. Box 68 (Gustaf Hällströmin katu 2b), FI-00014 University of Helsinki, Finland, and Part Time Researcher, Oslo Centre for Biostatistics and Epidemiology (OCBE), University of Oslo, Norway (e-mail: elja.arjas@helsinki.fi).

comes out as a result, through the front door, are explicit probabilistic *predictions* for future observables.

Of particular practical interest are predictive probabilities of the form $P(X_{n+1} = x_{n+1}, \dots, X_{n+m} = x_{n+m} | X_1 = x_1, \dots, X_n = x_n)$, $n, m \geq 1$, where $\{X_1 = x_1, \dots, X_n = x_n\}$ are the data observed up to the *present* time n, and $\{X_{n+1}, \dots, X_{n+m}\}$ are future observable responses of interest. In the context of survival problems, Singpurwalla uses the term *tracking survivability* for the inductive process of monitoring such conditional probabilities as n takes on larger values, and the observer is thereby learning from increasing amounts of data. This process is a version of the more general predictive statistical inference, the approach originated by Laplace and de Finetti and propounded, among others, by Geisser (1993). I should think that predictive inference has particular relevance in reliability applications: the main interest is in what is going to happen to the considered system or device in the future, not whether we can reject some hypothesis concerning an unobservable model parameter. At a later point in time we are then in a position where we can see how well a prediction we made matches with what the physical universe actually happened to produce.

There are systematic ways in the literature for specifying predictive probabilities in a closed form, without making recourse to standard parametric models with real valued parameters; examples can be found in Berliner and Hill (1988), Pitman (1996) and Spizzichino (2001). I am not aware of that such approaches would have been widely applied in reliability, however.

In reliability problems, one faces the additional challenge that, even if Popper's idea of the existence of an objective propensity concept in the physical sense were accepted and it were used as a model parameter, there is no way in which it, or its strength, could be measured on individual units. Unlike in the case of entities such as mass, length, etc., there is no instrument which would provide us with an operational definition of this concept. With such an operationalisation lacking, Singpurwalla's solution for establishing a link to observables, in the case of infinitely exchangeable sequences, is to use de Finetti's representation theorem (Theorem 1). Considering the prototype example of sequential Bernoulli trials, the theorem implies that probabilities $P(X_1 = x_1, \dots, X_n = x_n)$, $n \geq 1$, can be expressed as integrals, with respect to some distribution $F(d\theta)$, of the function $\theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}$. Thus, different coherent people, who would all agree to consider the variables $\{X_i; 1 \leq i \leq n\}$ exchangeable with respect to their personal probabilities, would end up using the same integrand $\theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}$ in the integral representation of their $P(X_1 = x_1, \dots, X_n = x_n)$. Moreover, they would expect the empirical relative frequencies $\frac{1}{n}\sum_1^n X_i$ of "success" to converge almost surely to a limit; this limit could then be taken as the definition of parameter θ, also called *chance*. In this sense, by assuming infinite exchangeability, one can indeed behave *as if* the model parameter θ existed. It is a matter of semantics, or perhaps taste, whether these mathematical results would justify calling θ *objective*.

But Singpurwalla wants more: he wants to establish another link between the two universes, and now in the opposite direction, from intellectual to physical. He writes: "The left-hand side of the theorem is a personal probability which we shall refer to as an item's survivability under its key quantities and the conditions which characterize its propensity. This makes reliability an objective, albeit unobservable, physical quantity, whereas its survivability is a subjective predictive entity". Jaynes [(1990), page 22] coined the term *mind projection fallacy* for such attempts to transport entities belonging to the intellectual world to the physical world. In my view, too, and apparently in contrast to Singpurwalla's, θ remains firmly in the intellectual universe.

Here, however, as a thought experiment, it may be useful to study what consequences, and possible advantages, *would* follow if we, as statisticians, behaved *as if* such a mind projection were possible. In fact, such behaviour, be it conscious or not, appears to be consistent with practices routinely followed in both the frequentist and the Bayesian camps of statistical inference when specifying models. This is also necessary, because de Finetti's theorem, even in the case of exchangeable sequences, is only a mathematical existence result, not a practical recipe for constructing a prior $F(d\theta)$.

The standard practice in setting up a statistical model is to move from parameters to observables, and do this by specifying a likelihood. This is only possible if the parameters in question are given some meaningful interpretation relating it to the physical world, at least in the mind of the person in question. The Bernoulli likelihood for a binary outcome is again the prototype of such reasoning, and then the parameter itself is given the probability interpretation $\theta = P_\theta(X_i = 1)$. Adopting Popper's idea of θ as a physical propensity parameter solves then the problem of such interpretation.

In standard Bayesian modelling, θ is considered as a random variable, which means simply that its "true" value is unknown to the observer, and a (prior) distribution F for θ is understood to be an expression of his or her uncertainty about this value. Consideration of θ as a random variable implies that Bernoulli likelihood, for a Bayesian, becomes the conditional probability $P(X_i = 1 \mid \theta)$. But at this point the Bayesian, too, usually claims to have some understanding of what θ would represent in the physical world in which he lives; otherwise quantifying uncertainty would not make sense. This line of thought has been followed in Bayesian approaches to reliability theory for a long time. For example, Natvig and Eide (1987) write: "Assume that we have described our present uncertainty on the reliabilities of the components, at a fixed point of time, by the moments up till order m of their marginal distributions".

The same recipe can be applied more generally, for defining more involved multivariate or hierarchical statistical models. Consideration of such multi-level structures is indeed necessary for all system reliability considerations. This process uses sequential conditioning of, and on, additional variables being introduced into the model, and then applying the chain multiplication rule for deriving joint distributions, assuming conditional independence where it seems appropriate. Similar ideas are followed when making use of the popular framework of graphical models, or the theory of stochastic processes in modelling of developments in time; importantly, both these approaches allow for explicit consideration of causal dependences.

However, such straightforward inclusion of new variables into the models on several different levels of hierarchy happens at a certain cost: it becomes increasingly difficult to keep faith in the idea that the new variables, if their values cannot be measured, would have an objective existence, let alone correspond to entities in the physical universe. A relatively simple example of this are the *frailty* parameters ξ considered in Section 5.4.1 of the paper. Lack of objective existence does not mean that introducing such variables into statistical modelling would necessarily be a bad idea, however. Shared frailties have turned out to be a convenient way of introducing dependence between recurrent failure times of the same repairable unit, or between the life lengths of genetically closely related individuals. A Google search of *frailty models in survival analysis* gave 157.000 hits, showing the enormous popularity of such ideas in practice.

In the Popper–de Finetti controversy, although Singpurwalla explicitly states that "propensity is not a probability", he nevertheless uses exclusively the rules of logic and probability calculus in his technical treatment of reliability problems, based on the adopted dictum "everything that is not forbidden is allowed". This, in my view, is indeed necessary for any meaningful systematic treatment for such problems. But it makes the models technically inseparable from what would be arrived at by applying the principles and practices of hierarchical Bayesian modelling. A variant of the well-known *duck test* would therefore give the conclusion that, what Singpurwalla is using, and in spite of his claiming the opposite, is probability.

Somewhat optimistically, Singpurwalla suggests that "the notions of chance and propensity need to work in concert with that of personal probability to produce the framework we need". If both Popper's and de Finetti's tunes were played together in a concert, I would expect to hear more dissonances than harmony. But perhaps Popper's tune could be played first, and then de Finetti's. Likewise, as statisticians, we could follow Popper in our attempts to set up a reasonably descriptive statistical model in the considered real world context, and then de Finetti, in performing posterior predictive inference based on that model and the acquired data. Or, to put it differently, we could first fit on Popper's hat, not caring much about de Finetti, and then switch to de Finetti's hat, forgetting Popper.

There is, admittedly, an element of *the end justifies the means* mentality in this suggestion of changing tunes or hats. But the added flexibility gained from applying different interpretations of probability is useful, as it plays an important role in our cognitive ability to create in our mind reasonable representations of physical world entities. Concrete reliability problems often involve complex physical structures and processes, and of causal dependences between them. This requires elaborate statistical modelling, in which considerations based on exchangeability alone are not sufficient. Using probability calculus in the modelling *as if* Popper's concept of physical probability existed, jointly with epistemic and frequentist interpretations, may then be the only practical recipe which works.

Lack of space does not allow me to comment here on the technical details of the material in Section 6 of Singpurwalla's paper. He ends the paper by writing: "… the algorithm presented here has the makings of a prototype approach for filtering and control in the presence of complete and partial observations, a topic on which there appears to be a dearth of literature". I have not followed closely the research literature in reliability during the past twenty years, so I cannot say whether this is indeed true. But I believe there are papers which are methodologically closely related. I would like to use this opportunity to provide references to two older ones of my own, with co-authors, Arjas, Haara and Norros (1992) and Arjas and Holmberg (1995).

REFERENCES

ARJAS, E., HAARA, P. and NORROS, I. (1992). Filtering the histories of a partially observed marked point process. *Stochastic Process. Appl.* **40** 225–250. MR1158025

ARJAS, E. and HOLMBERG, J. (1995). Marked point process framework for living probabilistic safety assessment and risk follow-up. *Reliab. Eng. Syst. Saf.* **49** 59–73.

BERGMAN, B. (1985). On reliability theory and its applications. *Scand. J. Stat.* **12** 1–41. MR0804223

BERLINER, L. M. and HILL, B. M. (1988). Bayesian nonparametric survival analysis. *J. Amer. Statist. Assoc.* **83** 772–784. MR0963805

DAWID, A. P. (2004). Probability, causality and the empirical world: A Bayes–de Finetti–Popper–Borel synthesis. *Statist. Sci.* **19** 44–57. MR2082146

GEISSER, S. (1993). *Predictive Inference. Monographs on Statistics and Applied Probability* **55**. Chapman & Hall, New York. MR1252174

JAYNES, E. T. (1990). Maximum entropy and Bayesian methods. In *Proceedings of the Ninth Annual Workshop Held at Dartmouth College, Hanover, New Hampshire, August* 14–18, 1989. *Fundamental Theories of Physics* **39** (Paul F. Fougère, ed.). Kluwer Academic, Dordrecht. MR1093675

NATVIG, B. and EIDE, H. (1987). Bayesian estimation of system reliability. *Scand. J. Stat.* **14** 319–327. MR0943292

PITMAN, J. (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **30** 245–267. IMS, Hayward, CA. MR1481784

SPIZZICHINO, F. (2001). *Subjective Probability Models for Lifetimes*. Chapman & Hall/CRC, Boca Raton, FL. MR1980207