Statistical Science
2016, Vol. 31, No. 4, 549–551
DOI: 10.1214/16-STS585
Main article DOI: 10.1214/16-STS565

What Does “Propensity” Add?

Jane Hutton

Abstract. Singpurwalla addresses the important challenge of modelling a unique individual. He proposes “propensity” as an approach to describing the reliability or life time of “one of a kind”. My view is that mathematical modelling is only possible when we assume that nonunique features provide sufficient information for statistical prediction to be useful. As far as possible, we should test our assumptions. However, contrary to a popular perception of Hume, we always rely on some beliefs.

The stated focus of Professor Singpurwalla’s paper is “reliability growth”. He asserts as an important insight: “reliability is a chance not a probability”. I have tried to think through what this changes in my approach to deciding when I should change my cycle tyre, or estimate the effect of a drug on the time to a person’s next epileptic seizure. As yet, I have not found any benefit.

There are texts which treat “chance” as referring to a description of the physical world, in contrast to probability as a mathematical system. However, I am not sure that this distinction is sufficiently familiar for “reliability is a chance not a probability” to cause general excitement. Philosophers and mathematicians have long puzzled over the fit between the elegant and enjoyable abstractions of mathematics and our ability to design and make moon-rockets. Many of us come to understand probability and statistics though a combination of exploration and explanation. We explore definitions through increasingly complicated paper exercises, computer simulations, and applications to data. We provide explanations of our results to others. Our interpretations are developed and our definitions refined through applications and interactions with others. I understand Kolmogorov’s “undefined primitives” as axioms. Any system of knowledge requires assumptions, or axioms, or beliefs, not all of which can be questioned simultaneously. Some axioms can, and should, be questioned. Where possible, each axiom should be tested by changing it and working out the consequences of the changes.

Jane Hutton is Professor, Department of Statistics, The University of Warwick, Coventry, CV4 7AL, United Kingdom (e-mail: J.L.Hutton@warwick.ac.uk).

Singpurwalla includes “propensity” as one interpretation of probability, the most useful for assessing the performance of “one of a kind.” It is not clear to me that a different formulation is required for life times. Do we require different concepts to model the breaking strength or failure time of a beam from those used to model the weight of a beam? Consider the claim that the uniqueness of each patient implies that statistics has no relevance in medicine. I argue that in order to reach a diagnosis or a decision, a doctor has to select a subset of characteristics of the patient so as to use knowledge from a group with those characteristics, Hutton (1995). Astronomy often addresses unique events, and frequentist theory has been effectively used there. If the uniqueness of the event implies no regularities or similarities, we can say little or nothing in advance, and observation of the event will not contribute to knowledge. To create a mathematical model of the world, we have to think of a collection, the members of which are deemed to be indistinguishable for the purpose of the model. If we regard an individual as entirely unique, we must be silent. In order to think about trustworthy performance or life expectancy, we have to select some aspects on which to focus attention, some aspects which adequately map the unique unit into our mathematical symbols.

Confidence in the value of observation, abstraction and mathematical models requires a belief in regularities of a universe, a universe which is partially comprehensible to us. If the thought which haunted Darwin—there is no value in the ideas of a mind developed from a lower mind—were true, why would we try to find formal justifications or seek to assess causes? To assert that our mathematics can make some sense (and allow some control) of the world is to assert that we can observe the world, Clark (2016). We do believe that the sun will rise tomorrow. To argue that a particular approach to reliability is correct requires the assumption that there is truth that exists regardless of our preferences. What criteria do we use to choose between models and assumptions? Elegance, simplicity, confirmation through predictions, conformity with a grand theory? One of my criteria is the effectiveness of the approach in my daily work.

Singpurwalla wants to align frequentist and personal interpretations of probability. The link in his article is the use of infinite collections or sets. His approach depends on the concept of exchangeable sequences which can be extended to an infinite sequence. The logic of objecting to notional infinite repetitions in frequentist probability while relying on infinite sequences is obscure to me. Exchangeable sequences extended to infinity do not differ from relative frequencies in infinite repetitions in the critical aspect, which is that abstraction is required for mathematics to get traction on the real world.

Take a simple model for a life time: let T be a random variable which represents the time from fitting a tyre to my bicycle until the first puncture; assume the distribution function for T is $\exp(-\lambda t)$, with $E(T) = 1/\lambda$. I might regard λ as a fixed but unknown property, which might be a function of a set of covariates. perhaps including the history of the tyre. Or I might regard λ as a realisation for a particular tyre of a further random variable, Λ, which represents my beliefs about tyres, including their propensity to fail. A statistical idea of the convergence or alignment of the two models is that for sufficient information, the frequentist and Bayesian intervals for λ will become indistinguishable. What is sufficient information for one of a kind? This is the dilemma.

Reliability is defined by Singpurwalla as the measure of the strength of an item's propensity to survive, θ, where θ is the limit of an infinite sum of random variables. The strength of propensity to survive, "survivability" is added to "reliability" as a further remove from parameters. Reliability is a stepping stone to survivability. Singpurwalla's propensity is not a probability, nor a parameter, but implies a "nondeterministic causal relationship". Propensity might have the same role in Singpurwalla's physical world as latent factors have in psychology. Reliability is also said to be a metric, but there is no demonstration that it has properties of a metric. It is an objective chance, and objective physical quality. "Objective" refers to a physical world, ultimately described using probability theory in quantum mechanics. Survivability is personal probability, a subjective predictive entity, a measure of performance, a manifestation of reliability, simply a probability. There are multiple metaphors: "sums of lifetimes *endowed with an indifference*", "spawns", "invoke". I am drowning in equivalences, synonyms and metaphors. It is fun, and often educational to play with metaphors, but relying on metaphors to reach the truth is risky. That kind of bridge often breaks and throws us into an abyss when we try to cross to the truth.

Singpurwalla asserts that personalistic Bayesian methods are the proper approach to survival analysis, connecting propensity to reliability. I think he intends "propensity" to replace what I would call a "property" of a particular unit. Parameters provide a partial summary of a property: a subset of the unit's characteristics. As propensity "encompasses a consideration of all the *key qualities* of the object" (my emphasis), Singpurwalla does select a subset of characteristics. Choices of abstractions are made. So, does T in Section 3 denote the actual lifetime of *an* item, or is it a model which allows us to say something useful and general? What do these accumulated descriptions add to my understanding of the time to a person's next epileptic seizure, or to death? After a person's death, the observed lifetime is a manifestation of their reliability or a measure of performance. This does not provide me with any new insight. I await unenlightment.

Singpurwalla claims the appeal of propensity is that it connotes a causal relationship. This is his personal definition or understanding; to me, a "propensity" is similar to a "tendency", and is not a "cause". I can cause my back to hurt by lifting a heavy weight. I interpret the tendency of my wider family to have back pain associated with the HLA-B27 gene to mean that a higher proportion of those of us with this gene with have back pain than those without. We have a greater propensity to hurt our backs. Perhaps Singpurwalla intends "propensity" to include the entire history of all physical properties which result in—*cause*—an observed event. If so, I understand smoothing or filtering to mean disregarding some information about a unit so that it can be regarded as one of a group, exchangeable with other units. Then estimates or predictions of failure time can be made using observations on some units from that group. There is no mention of another common use of "propensity": "propensity scores" are used in adjusting estimates when data are missing. A causal story could be told about missing data.

As there is no definition of a conditional propensity, I cannot make sense of the claim that conditional propensities cannot be symmetric. I see no reason for probability theories to be theories of cause, as opposed to part of wider theories which we use to understand causes. I am surprised that there is no reference to recent discussions of causality and probability, Pearl (2009).

With regard to the methodological section of the article, we are offered models similar to those already in use. Tracking reliability growth could be useful: we might well wish to model residual life time, conditional on as much of the past as we can capture. I regard a hazard function as a mathematical model for an item's reliability over time, which conditions on the past. Singpurwalla's "model propensity rate" (Section 5.2) is similar to the hazard rate, but for discrete time, so can be written as a probability. I thought $P(\cdot)$ in the definition denoted probability, but the last paragraph of Section 5.2 tells me that it is a conditional propensity. In Section 5.1, $P(\cdot)$ denoted probability and P_i a parameter or "strength of propensity". The particular definition chosen yields a decreasing model propensity rate, which Singpurwalla claims is contrary to intuition. If there is a contradiction, why not track back through definitions and reasoning to check for an error?

In Section 5.4, we are told that cure models have "a heavy infrastructure"—a log-linear link between a parameter and covariates. In equation (5.9), the same equation is apparently unexceptional. Singpurwalla claims the parameter for cure fraction is assumed fixed and dependent only on the covariate values at time T. However, I have used a fairly simple model with different cure rates at different times for responses to anti-epileptic drugs, Rogers and Hutton (2013) and joint models for longitudinal and survival data allow the hazard rate to be updated over time, Rizopoulos (2012). This is similar to using filtering to model changes in P (here P is Singpurwalla's propensity) over time. Other more general stochastic process models with multiple states can also accommodate changing transition rates, cure fractions and frailties.

In Section 6, we read "the Bernoulli parameter P was the cause of an observed binary random variable, and that one's knowledge about P changes". Notice we have a cause of observed frequencies, no longer a single unique item. Imagine throwing a drawing pin from a paper cup, and setting $X = 1$ if it lands point up, $X = 0$ if it lands with the point and an edge of the base resting on the table. Model X_i, $i = 1, 2, \dots$ as Bernoulli random variables with $\mathrm{Prob}(X_i = 1) = \beta$. Those of us who have taught at the University of Newcastle, where this experiment is used in teaching undergraduates, had precise estimates of β: 95% intervals of (0.7320, 0.7324), but I do not think β *causes* the drawing pins to land point up.

The distinctions which Singpurwalla tries to draw have confused me. I hope he will be able to explain what "propensity" adds in the examples I give, as better understanding would improve my modest contributions.

REFERENCES

CLARK, S. R. L. (2016). Atheism considered as a Christian sect. *Philosophy* **90** 277–303.

HUTTON, J. L. (1995). Statistics is essential for professional ethics. *J. Appl. Philos.* **12** 253–261.

PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166

RIZOPOULOS, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data*. Chapman & Hall/CRC, Boca Raton, FL.

ROGERS, J. K. and HUTTON, J. L. (2013). Joint modelling of pre-randomisation event counts and multiple post-randomisation survival times with cure rates: Application to data for early epilepsy and single seizures. *J. Appl. Stat.* **40** 546–562. MR3047301