Statistical Science
2016, Vol. 31, No. 4, 521–540
DOI: 10.1214/16-STS565

Filtering and Tracking Survival Propensity (Reconsidering the Foundations of Reliability)[1]

Nozer D. Singpurwalla

Abstract. The work described here was motivated by the need to address a long standing problem in engineering, namely, the tracking of *reliability growth*. An archetypal scenario is the performance of software as it evolves over time. Computer scientists are challenged by the task of when to release software. The same is also true for complex engineered systems like aircraft, automobiles and ballistic missiles. Tracking problems also arise in actuarial science, biostatistics, cancer research and mathematical finance.

A natural approach for addressing such problems is via the control theory methods of filtering, smoothing and prediction. But to invoke such methods, one needs a proper philosophical foundation, and this has been lacking. The first three sections of this paper endeavour to fill this gap. A consequence is the point of view proposed here, namely, that reliability *not* be interpreted as a probability. Rather, reliability should be conceptualized as a dynamically evolving propensity in the sense of Pierce and Popper. Whereas propensity is to be taken as an undefined primitive, it manifests as a chance (or frequency) in the sense of de Finetti. The idea of looking at reliability as a propensity also appears in the philosophical writings of Kolmogorov. Furthermore, *survivability* which quantifies ones uncertainty about a propensity should be the metric of performance that needs to be tracked. The first part of this paper is thus a proposal for a paradigm shift in the manner in which one conceptualizes reliability, and by extension, survival analysis. This message is also germane to other areas of applied probability and statistics, like queueing, inventory and time series analysis.

The second part of this paper is technical. Its purpose is to show how the philosophical material of the first part can be incorporated into a framework that leads to a methodological package. To do so, we focus on the problem which motivated the first part, and develop a mathematical model for describing the evolution of an item's propensity to survive. The item could be a component, a system or a biological entity. The proposed model is based on the notion of competing risks. Models like this also appear in biostatistics under the label of *cure models*. Whereas the competing risks scenario is instructive, it is not the only way to describe the phenomenon of growth; its use

Nozer D. Singpurwalla is Chair Professor, Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong, China (e-mail: nsingpur@um.cityu.edu.hk).

[1]Discussed in 10.1214/16-STS583, 10.1214/16-STS577, 10.1214/16-STS585 and 10.1214/16-STS575; rejoinder at 10.1214/16-STS593.

here is illustrative. All the same, one of its virtues is that it paves the path towards a contribution to the state of the art of filtering by considering the case of censored observations. Even though censoring is the hallmark of survival analysis, it could also arise in time series analysis and control theory, making the development here of a broader and more general appeal.

Key words and phrases: Chance, competing risks, cure models, exchangeability, filtering censored observations, frailty, propensity, proportional hazards, time series analysis.

0. INTRODUCTION: MOTIVATION AND BACKGROUND

The performance of many an engineered system changes over time due to improvements in design, manufacture and fault elimination. This is also true of biological systems which gain immunity after inception, and subsequent to this, experience ageing; also, patients undergoing medical therapies. Tracking the trustworthy performance and survivability of such systems is of concern in the biological as well as the engineering sciences. Problems like these go under the label of *reliability growth* about which much has been written, both in the engineering and the statistics outlets. It turns out that a satisfactory resolution of the matter of assessing reliability growth raises philosophical and technical issues whose effects boil down to a reconsideration of the foundations of reliability and survival analysis. In what follows, the term reliability includes survival analysis. The purpose of this paper is to articulate on the above matter, and to propose a change in the manner in which one conceptualizes reliability and addresses the problems that it spawns.

By way of background, as it is currently understood, *reliability theory* pertains to an assessment of the *probability* of an item completing its prescribed function within a specified period of time, under specified conditions. This theory is the basis of reliability engineering, survival analysis, actuarial science and *duration analysis* in finance.

1. FOUNDATIONS OF RELIABILITY THEORY

It may be claimed that the science of reliability begat the field of statistics, in a manner akin to how gambling begat probability. The first treatise on statistics appeared in 1662, in Graunt's "London Bills of Mortality". This was a life-table whose appearance signaled the official registration of births, deaths and marriages. Per Karl Pearson, Graunt was the father of statistics, and its modern day evolution as "big data".

Whereas much has been written on the mathematical and methodological aspects of reliability, it appears that little has been documented about its foundational underpinnings; some preliminary (but unsatisfactory) attempts are in Singpurwalla (1988, 2002). This is understandable because reliability has been defined as a probability, and there is plenty written on the meaning of probability in both the mathematical and the philosophical literatures. The caveat here is that there are several interpretations of probability. These are: a *relative frequency*, a *personalistic* 2-*sided bet*, a *propensity* and to Kolmogorov (1969), who takes refuge in the principle of "ignotum per ignotius"—an *undefined primitive*. Thus per Kolmogorov, were reliability to be interpreted as a probability—which he did—then reliability too would be an undefined primitive. Which of the above interpretations is germane to performance assessment, wherein one often deals with a one of a kind item? This is more a matter of philosophy and mathematics than of engineering or biostatistics. However, implicit in the current work in reliability is an adherence to either the relative frequency or the personalistic interpretation. The former is embodied in the comprehensive treatise by Meeker and Escobar (1998), the latter in the recent monograph by Hamada et al. (2008). Can these two differently interpretive viewpoints be aligned so that they constructively reinforce each other? The mathematically elegant and classic books by Barlow and Proschan (1965, 1975) predominantly focus on the structure of probability distributions taking the notion of probability as a given.

Whereas Popper's propensity interpretation of probability has been discarded by philosophers of science, like Humphreys (1985), and correctly so, the valuable role that this notion can play in performance assessment remains to be explored. In particular, one can lean on de Finetti's theorem (cf. de Finetti, 1937) on exchangeable sequences, and exchangeability in general, to make propensity work in *tandem* with personal probability, and produce a workable methodological package. It seems that such a linkage can provide a meaningful foundation for a theory of reliability and bring some conceptual clarity to the diverse approaches to life data analysis. Furthermore, the linkage can also bring about a rapprochement between the relative frequency and the personalistic interpretations.

The linking of propensity and personal probability is both appealing and necessary. Appealing because propensity connotes a causal relationship between what is observed and the underlying physical circumstances which produce the observables. Useful, because the linkage provides a formal platform for invoking the techniques of filtering, smoothing and predictions which are a pathway to tracking survivability, be it static or dynamic. However, to formalize this linkage we need to think of reliability as a propensity, and not as a probability, as is currently done, both in the frequentist and the personalistic paradigms. Indeed, in order to formally justify using a proper (personalistic) Bayesian approach in the contexts of reliability and survival analysis, it seems *necessary* to bring into the foray the notion of propensity or something equivalent. Not doing so raises the dilemma of placing a 2-sided bet on a 2-sided bet, making the Bayesian analysis a circular development.

2. PROPENSITY, FREQUENCY AND DE FINETTI'S THEOREM

Because the relative frequency theory of probability, upon which sample theoretic statistical inference is based, pertains to observable mass phenomena and repetitive events, it is unable to address issues pertaining to one of a kind events (cf. von Mises, 1941). Other difficulties with this theory, like indefinite repetitions under almost similar—but not exact—conditions, are outlined by de Groot (1988), and by Gillies (2000). By contrast, the notion of propensity is germane to one of a kind events, and this is its most attractive feature.

The basic idea behind propensities appeared in the early 1900s in the writings of the American philosopher Charles Pierce (cf. Miller, 1975). Its strongest proponent, however, was the British philosopher Karl Popper (1959). The notion of propensities is grounded in the objectivity of physical sciences in the sense that the propensity of an object for exhibiting a particular outcome is a physical attribute of the object and its environmental conditions of use. Thus, the notion of propensity, be it for a component, a system or a biological entity, encompasses a consideration of all the key qualities of the object, as well as the manner in which it is used. To quote Popper (1957), propensity is "an *unobservable dispositional property of the physical world*, which includes its measurable attributes, as well as the experimental conditions...". Consequently, a propensity is like the Newtonian notions of mass, volume and specific gravity. Stated differently, the propensity of an object is its tendency to yield an outcome of a certain kind, or to yield a long run frequency of such outcomes. To Popper (1957), propensities are the purported causes of relative frequencies. This means that we may regard relative frequency as the strength of a propensity, with the value of 1 implying a certainty and the value of 0 implying impossibility. This interpretation is important; it connects an observed frequency with an objective but unobservable dispositional property of the physical world.

The notion of propensity was thought useful in quantum mechanics wherein interest centers on isolated events, like the position of an atom at a particular time. Indeed, Popper's motivation for proposing his theory of propensities, was a dissatisfaction with the Bohr–Heisenberg subjectivist interpretation of quantum theory (cf. Popper, 1957). Popper wanted an objective interpretation of quantum theory which accommodates the probability of isolated events. Because propensities are likened to notions like mass, force and gravity, Good (1950) has called propensity a *physical probability*; also see Good and Card (1971) wherein the term *demiurge* is used in relationship to the notion of propensity.

2.1 Propensities Are Not Probabilities

Despite Popper's insistence on interpreting a propensity as a probability, and Good's labeling it a physical probability, a propensity is *not* a probability! Rather, a propensity implies a nondeterministic *causal* relationship between an event and its generating conditions. Thus, unlike conditional probabilities which are symmetric in their arguments, in the sense that both $P(A|B)$ and $P(B|A)$ are meaningful to consider, conditional propensities cannot be symmetric; see Humphreys (1985). In von Mises' frequency theory, probability is a property of a collective, whereas Popper's propensity is a property of the generating conditions of a sequence, where by the term generating conditions, we include a description of the key qualities of the object generating the sequence. Because a conditional probability is a definition in Kolmogorov's theory, and a corporate state of mind in the personalistic theory, conditional probabilities do not encapsulate causality, which propensities do. The inability of probability to encapsulate causality has prompted the philosopher Humphreys (1985) to "reject the current theory of probability as the correct theory of chance". Indeed, Fetzer (1981) in his probabilistic causal calculus has proposed a set of axioms for propensities that are unlike the Kolmogorov calculus. Finally, whereas the strength of a propensity can take the value 0 and 1, the degree of a personal probability, as reflected via a 2-sided bet, must be *between* 0 and 1.

Perhaps the closest that the notion of propensity comes to probability is the view expressed by Kolmogorov (1969), page 123, that probability is an undefined primitive (governed by a system of axioms) which "connects the useful life of a lamp with the materials and the technological condition of its manufacture"; also see Gillies (2000), and Kendall (1949), who states that "Any theory of probability which does not take probability as a primitive idea must, in some form or other, introduce an equivalent primitive before it can be applied". The quote by Kolmogorov is the basis of our claim that in equating reliability to a probability Kolmogorov would have interpreted reliability as an undefined primitive, characterized by the qualities of an object.

2.2 De Finetti's Theorem: Linking Propensity and Probability

Whereas propensity cannot be an interpretation of probability, there is a connection between the two which is made explicit by de Finetti's theorem on *exchangeable sequences*. To appreciate the spirit of the theorem, consider n tosses of a metaphorical coin conducted under similar conditions. Suppose that the *n future* outcomes of these tosses are *judged* exchangeable, in the sense that one's *personal* probability of any sequence of r heads and s tails, with $n = r + s$, is the same as any other sequence of r heads and s tails. For example, independent Bernoulli trails with a constant parameter θ is an exchangeable sequence. However, not all exchangeable sequences are independent, implying that the judgment of independence is stronger than that of exchangeability. Exchangeability connotes a sense of similarity, or positive dependence, and without exchangeability, it is not possible to justify inductive statistical inference.

THEOREM 2.1. *For every exchangeable probability assignment that can be extended to a probability assignment on an infinite sequence of zeros and ones, there corresponds a unique probability distribution function* F, *concentrated on* $[0, 1]$, *such that for all* n, *and* $0 \le r \le n$,

$$P(X_1 = 1, \ldots, X_r = 1, X_{r+1} = 0, X_n = 0) = \int_0^1 \theta^r (1-\theta)^{n-r} F(d\theta),$$

where $\lim_{n\to\infty} \sum X_i/n$ *exists, and is equal to* θ.

Here, $F(d\theta) = f(\theta)d\theta$, and $F(\theta)$ is the probability distribution function of θ. The left-hand side of the theorem is a personal probability, which can be operationalized via a 2-sided bet. The right-hand side of Theorem 2.1, represents a mixture of Bernoulli probabilities, with $F(d\theta)$ as the mixing distribution.

There are two aspects of the theorem that warrant discussion. These have to do with θ and $F(\theta)$. Clearly, θ being the limit of a relative frequency cannot be a personal probability. Rather θ is to be seen as a property of the world which is a manifestation of the propensity of any X_i to take the value 1. Alternatively put, it is the propensity of each X_i, $i = 1, 2, \ldots$, to take the value 1, which spawns θ, making θ a reflection of the underlying propensity. De Finetti refers to θ as a *chance*, a term also used by Pierce, though not in the same sense as de Finetti. This means that de Finetti's theorem can be seen as one which links personal probability and the strength of the underlying propensity. The term $F(\theta)$ is a personal probability which encapsulates one's uncertainty about θ, which (because of the caveat that $n \to \infty$) is unobservable. All the personal probabilities mentioned here are those of the individual(s) making the judgment of exchangeability.

A final, albeit important, point of note is that the term $\theta^\tau(1-\theta)^{n-\tau}$ in de Finetti's theorem is the Bernoulli (chance) distribution. Thus, more generally, one may also interpret probability models with *unknown* parameters as the reflection of an underlying propensity, and de Finetti's theorem as one which links an objective probability distribution with a subjective personal probability. This viewpoint is the basis of the material of Section 4.

2.3 Generalized Versions of de Finetti's Theorem

Whereas Theorem 2.1 above serves the purpose of justifying the point of view that we espouse here, there are more general and stronger versions of the theorem which further assert our viewpoint. These stronger versions come about because Theorem 2.1 pertains to Bernoulli events and is germane under the metaphysical concept of infinite sequences. In the context of occurrence data that is characteristic of the assurance sciences and biostatistics, one encounters sequences other than the Bernoulli; furthermore, one also needs to confront finiteness. Diaconis and Freedman in a series of highly influential papers addressed these issues culminating in their 1987 paper. Here, using arguments of invariance and symmetry, they produce de Finetti style theorems for mixtures of distributions other than the Bernoulli, namely, the geometric, Poisson, exponential and the Gaussian; also see Smith (1981). In addition to the above, Diaconis and Freedman (1987) provide error bounds on the resulting probabilities when the scenario of infinite sequences is replaced by that of finite sequences. The net effect of these extensions and generalizations is that a relationship between a subjective personal probability and the strength of an objective dispositional propensity holds under a wider set of circumstances.

In a striking, but less known paper whose theme is more in tune with lifetime models used in connection recurrent event data, Barlow and Mendel (1992), using the principle of indifference derive mixtures of the gamma and the Weibull distributions. In the former case, it is the sums of lifetimes that are endowed with an indifference; in the latter case, indifference is invoked on the sums of certain utility functions—maintenance costs in the case of Barlow and Mendel (1992). A summarization of the above contributions, plus other related material is in Spizzichino (2001), and in Chapter 3 of Singpurwalla (2006).

3. RE-INTERPRETING RELIABILITY: METRIC OF SURVIVABILITY

In the context of reliability, the θ of Theorem 2.1 can be seen as encapsulating the strength (or degree) of an item's propensity to survive. We take this measure of strength to be the *definition* of reliability. Thus, as was claimed in Lindley and Singpurwalla (2002), *reliability is a chance not a probability*.[2] 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. Because personal probabilities can be operationalized via 2-sided bets, it is an item's survivability that should be the measure of performance. Because survivability is a manifestation of reliability, the latter becomes a stepping stone to the former. This change in mode of conceptualizing reliability constitutes the paradigm shift mentioned before.

As stated in Section 2.3, analogues of de Finetti's theorem for continuous quantities, like lifetimes, have been developed; also see, for example, Hewitt and Savage (1955). These bring into play the commonly used failure models like the exponential, the gamma and the Weibull, as chance distributions. Such chance distributions are, in principle, unobservable, but serve the useful purpose of making predictive assessments of an item's survival.

Thus, if T denotes an item's lifetime, then the item's survivability assuming an exponential (λ) chance distribution will be given as: $P(T > t) = \int_0^\infty e^{-\lambda t} F(d\lambda)$. Here, $e^{-\lambda t}$ is the strength of the item's propensity to survive to t, and $F(d\lambda)$ encapsulates one's uncertainty about λ. Here again, the degree of propensity can be expressed in terms of a probability model with unknown parameters.

3.1 Some General Remarks on the Paradigm Change

There are two attractive features of the suggested paradigm change. The first is conceptual clarity, which helps justify the use of Bayesian methods in reliability and survival analysis. The second is a foundation for invoking filtering and control theory methods for tracking reliability growth. In the subsequent sections of this paper, an argument will be made that all problems of reliability are essentially a matter of tracking survivability, so that assessing reliability growth is quintessential to the biomedical and the engineering sciences.

Because we have equated the notion of reliability to that of an objective chance, a question may arise about the need to introduce propensities in the discussion. There are several reasons for doing so. One is inclusiveness, in the sense that the notion of propensities embraces the diverse but similar views of Pierce, Popper and Kolmogorov about the cause of observed frequencies, and causal relationships are the hallmark of science. The second reason is a facility to think in terms of one of a kind entities; frequencies are generally associated with collectives. Finally, in the context of filtering and control, propensities can serve as the state of nature, and an evolving propensity becomes a rationale for "system equation" of dynamic models, such as the Kalman filter model.

[2]This paradigm-changing insight is solely attributed to the late Dennis Lindley.

4. THE PHENOMENON OF RELIABILITY GROWTH

Many complex systems experience, over time, a change in their ability to survive any specified (mission) time. For example, biological systems first attain increasing levels of immunity, either naturally or by immunizations, and then experience deterioration due to ageing. This is also true of engineered systems and computer software which undergo design changes and fault elimination, which enhance their ability to survive. Similarly, businesses and organizations gain maturity due to improved practices and procedures. Thus, from a Popperian point of view, the key qualities of an entity, and the conditions of its use change over time, leading to a change in an item's propensity to survive. This phenomenon of a time-evolving propensity is what should be interpreted as *reliability growth*. However, since propensities are unobservable, and nor are the chances they spawn, tracking reliability growth is not possible. By contrast, survivability can be tracked. Thus, the commonly discussed procedures for assessing reliability growth boil down to the matter of tracking survivability. The change in focus from tracking reliability to monitoring survivability is not a mere change of verbiage; the underlying fundamentals are philosophically different.

The perspective mentioned above provides a foundation for casting the topic of performance assessment in the general framework of filtering and control. Specifically, an item's unobserved reliability (i.e., the strength of its survival propensity) is viewed as its *state* of *nature*, and its survivability as its *observable* and predictable quantity. But to implement this framework, we need to endow an evolutionary feature to the item's reliability. This we do via arguments which parallel a dynamic competing risks framework; see Section 5. In the case of static systems, that is, systems which do not experience time indexed changes, one may invoke the *steady model* of a propensity (cf. Meinhold and Singpurwalla, 1983).

To summarise, the essential import of this section is the viewpoint that filtering reliability and tracking survivability is a generic approach for performance assessment and the treatment of occurrence data in the biological, the engineering and the actuarial sciences. Differences in approaches are because of differences in the models used to characterize the evolving strength of propensities. Section 5.1 pertains to one such model. However, this is one among several such possibilities. The choice of this model is not coincidental. It was during the process of commenting on a model like this that the author encountered the foundational issues discussed in this paper; see Singpurwalla (2010).

5. A COMPETING RISKS MODEL FOR EVOLVING PROPENSITIES

We develop here a model for describing an evolving strength of propensity (to survive) of a complex system. Such a model will form a basis for the *state equation* of a dynamic model. In control theory, such models are specified using arguments based on the science of the problem. However, the strategy used here is suggested by the term $\theta^r(1-\theta)^{n-r}$ of Theorem 2.1, or the $e^{-\lambda t}$ of Section 3, wherein the strength of propensity is a probability model with unknown parameters. With the above in mind, our model for the strength of propensity is based on arguments that are grounded in the calculus of probability, and is conditioned on unknown parameters. Doing this has led to a comment—personal conversation with James Berger—that using probabilistic arguments to encapsulate propensity seems circular because doing so makes the two isomorphic. Our response is that probability and propensity are different notions, and that our approach is legislated by de Finetti's theorem. Furthermore, there is nothing in the underpinnings of the notion of propensity which dictates how its strength should be specified. Also, there exists a scenario wherein probability based arguments have been used for specifying nonprobabilistic entities. For example, the likelihood function, which is *not* a probability, is often specified by reversing the arguments in a probability model. We are invoking here a constitutional principle of English Law, namely that "everything that is not forbidden is allowed", a theme spectacularly exploited by the physicist Murray Gell-Mann in his discovery of the quark particle.

5.1 An Omnibus Model for the Strength of Evolving Propensities

We start by re-emphasizing that the set-up described below is very generic and universally applicable, and especially so in the biomedical and engineering sciences. Furthermore, that even though the arguments used below are based on the calculus of probability, the entity of interest is a mathematical model for an item's strength of propensity and the evolution of this propensity over time.

Suppose that a system is prone to failure due to an unknown number, say M, of faults or defects. The system could be a biological entity, a piece of software (cf. Martin, 2012), an algorithm or an engineered unit like an airplane. We regard these faults as the causes of failure in the sense that whenever the system fails, one or more of the M faults are revealed. To achieve reliability growth, one attempts to eliminate these faults, and with the hope of not introducing new faults; however, this does not always happen. The dynamics of the proposed model is governed by a parameter whose distribution changes over time, to reflect the possibilities of eliminating or not the observed fault and/or introducing new faults. We assume that M is Poisson with parameter $\mu > 0$. This assumption makes provision for a fault-free system, but doing so causes the model to have a positive mass at infinity.

Suppose that fault i ($i = 1, \dots, M$) spawns a discrete random quantity T_i taking the value $t_i = 1, 2, \dots$. We take T_i to be the number of trials (or runs) of the system until it encounters a failure which reveals i as the failure causing fault. Before testing, the nature of the faults is unknown. This means that we are a priori unable to a priori discern any systematic pattern to the possibly infinite number of T_i's. It therefore makes sense to assume that the T_i's are *infinitely exchangeable*, and we construct an exchangeable sequence via mixtures of independent and identically distributed random variables. To do so, we assume that each T_i has a geometric distribution with parameter P_i, where P_i is the strength of the propensity of fault i to cause a failure in any particular run of the system. Consequently,

$$P(T_i = t_i | P_i = p_i) = (1 - p_i)^{t_i - 1} p_i, \quad i = 1, 2, \dots.$$

To generate an exchangeable sequence of the T_i's, we assume that the P_i's are random draws from a beta distribution with specified parameters α and β. Then it is easily seen that the joint probability mass of $(T_1, \dots, T_M)$ at $(t_1, \dots, t_m)$ is

$$P(T_1, \dots, T_M; \alpha, \beta) = \prod_{i=1}^{M} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + 1)\Gamma(\beta + t_i - 1)}{\Gamma(\alpha + \beta + t_i)}.$$

This is the product of independent *Beta-Geometric distributions* (cf. Dubey, 1966), implying that exchangeability here is due to independence of the T_i's. To achieve exchangeability without independence, we need to either generate a dependent collection of P_i's, $i = 1, \dots, M$ from a multivariate beta distribution, or we may set $P_1 = P_2 = \cdots = P_M = P$, and generate a single P from a beta distribution. Doing the former requires prior knowledge about a relationship between the P_i's. The latter can be justified if and only if, for every M, the joint distribution of $(T_1, \dots, T_M)$, given $S_M = \sum_1^M T_i$, is a *Bose–Einstein distribution*; that is,

$$P(T_1 = t_1, \dots, T_M = t_M | S_M) = \frac{1}{k},$$

where k is the total number of M-tuples whose sum is S_M. This means that the joint distribution of $(T_1, \dots, T_M)$ given S_M is judged to be equiprobable over all nonnegative M-tuples of integers whose sum is S_M (Diaconis and Freedman, 1987, Theorem 4.7, page 414). Whereas symmetry judgements like this are difficult to conceptualize, the act of generating a dependent sequence of T_i's via the mechanism of sampling via a single P seems attractive. Because dependence among T_i's poses challenges namely a theory of extreme values, we continue with the assumption that the P_i's are random draws from a common beta distribution, making the T_i's an independent sequence.

With the above in place, we next assume that the T_i's compete with each other to cause a failure of the system. Thus, if T denotes the number of runs (or demands) of the system until its first observed failure, then a competing risks model for T, conditioned on $M = m \geq 1$, will be a consequence of $T = \min(T_1, \dots, T_M)$. Thus,

$$\begin{aligned} P(T > t | M = m \geq 1; \alpha, \beta) &= \prod_{i=1}^{m} P(T_i > t | M = m \geq 1; \alpha, \beta) \\ &= \prod_{i=1}^{m} \sum_{s=t+1}^{\infty} P(T_i = s | M = m \geq 1; \alpha, \beta) \\ &= \left[\sum_{s=t+1}^{\infty} \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + 1)\Gamma(\beta + s - 1)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + s)} \right]^m. \end{aligned}$$

Unconditioning on $M \geq 0$ via its Poisson (μ) distribution, we have, for $t = 0, 1, 2, \dots$,

$$\begin{aligned} P(T > t | \mu; \alpha, \beta) &= P(T > t | M = 0, \mu; \alpha, \beta) P(M = 0) \\ &\quad + \sum_{m=1}^{\infty} P(T > t | M = m \geq 1, \mu; \alpha, \beta) P(M \geq 1) \\ &= 1 \cdot e^{-\mu} \\ &\quad + \sum_{m=1}^{\infty} \left[\sum_{s=t+1}^{\infty} \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + 1)\Gamma(\beta + s - 1)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + s)} \right]^m \\ &\quad \cdot \frac{e^{-\mu} \mu^m}{m!}. \end{aligned}$$

Conditional on μ, and with α and β specified, the above expression is the reliability (or the strength of propensity to survive) of the system, when it is born.

As pointed out by a referee, and correctly so, these parameters can be interpreted as being spawned by *epistemic* considerations.

Simplification results of $\alpha = \beta = 1$, so that the p_i's are random draws from a uniform distribution on $(0, 1)$. Now

$$\begin{aligned} P(T > t|\mu) &= e^{-\mu} + \sum_{m=1}^{\infty} \left[\sum_{s=t+1}^{\infty} \frac{1}{s(s+1)} \right]^m \frac{e^{-\mu}\mu^m}{m!} \\ &= e^{-\mu} + \sum_{m=1}^{\infty} \left(\frac{1}{t+1} \right)^m \frac{e^{-\mu}\mu^m}{m!} \qquad (5.1) \\ &= \exp\left(-\frac{\mu t}{t+1} \right) \stackrel{\text{def}}{=} \bar{F}_\mu(t), \end{aligned}$$

is the system's *nascent reliability* (or *strength of propensity*) *function*.

5.2 Exploring the Nascent Reliability Function

A plot of $\bar{F}_\mu(t)$ versus t, for $\mu = 1, 3$ and 5, is shown in Figure 1.

A noteworthy feature of $\bar{F}_\mu(t)$ is that it takes a value greater than 0 when $t \to \infty$. This implies that the system has a nonzero strength of propensity for indefinite survival. This property is atypical of the complementary distribution functions used in the literature, wherein $\bar{F}_\mu(t)$ must tend to zero as $t \to \infty$, or else the underlying distribution is improper. Thus, $\bar{F}_\mu(t)$ cannot be interpreted as a complementary probability distribution function; rather, as stated before, it is to be seen as a strength of propensity to survive until t.

Being able to accommodate the feature of indefinite survival is attractive, because indefinite survival can be

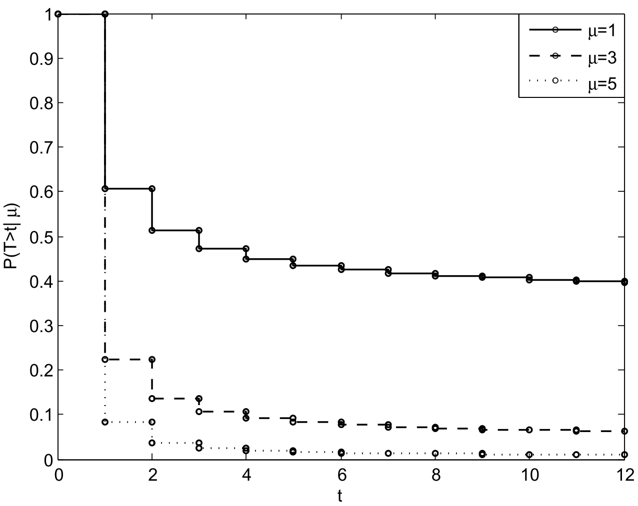

FIG. 1. *The nascent reliability function.*

realistic. In our case, it is a consequence of the Poisson distribution having mass $e^{-\mu}$ at $M = 0$. With $M = 0$, we are making provision for a fault free system. Since

$$\begin{aligned} &P(T = t|\mu, \cdot) \\ &\quad = P(T > t-1|\mu, \cdot) - P(T > t|\mu, \cdot) \qquad (5.2) \\ &\quad = e^{-\mu}(e^{\mu/t} - e^{\mu/(t+1)}), \end{aligned}$$

the right-hand side of the above expression is the strength of an item's propensity to fail at t. Note that this expression is undefined at $t = 0$, and it tends to zero as $t \to \infty$. However, since $P(T > t|\mu, \cdot) > e^{-\mu t}$, when $t \to \infty$, the model reflects a strength of propensity to survive, of at least $e^{-\mu}$, at $t = +\infty$.

As a parallel to the *model failure rate* (Singpurwalla, 2006, page 66), it is tempting to define the *model propensity rate* of $\bar{F}_\mu(t)$, at $T = t$, for $t = 1, 2, \ldots$, as

$$P(T = t|T > t-1, \mu; \cdot) = \frac{P(T = t|\mu; \cdot)}{P(T > t-1|\mu; \cdot)},$$

where

$$P(T = t|\mu; \cdot) = P(T > t-1|\mu; \cdot) - P(T > t|\mu; \cdot).$$

Thus,

$$P(T = t|T > t-1, \mu; \cdot) = 1 - \frac{P(T > t|\mu; \cdot)}{P(T > t-1|\mu; \cdot)},$$

from which it follows that the model propensity rate of $\bar{F}_\mu(t)$, at $T = t$, is of the form:

$$P(T = t|T > t-1, \mu; \cdot) = 1 - e^{-\mu/t(t+1)}.$$

This rate is undefined at $t = 0$; it takes the value $1 - e^{-\mu/2}$ at $t = 1$, and decreases to 0 as $t \to \infty$, suggesting a growth in an item's propensity to survive over time.

But the above feature is contrary to intuition, because in the absence of an external intervention, one expects a propensity to be constant over (calendar) time. A contradiction such as this arises because unlike conditional probabilities (which are a corporate state of mind) conditional propensities do not carry a meaningful interpretive import. The notion of a model propensity rate must therefore be viewed as being mainly technical. It is presented here to underscore the fact that propensities are not probabilities.

5.3 The System Survivability Function

Section 5.1 pertained to the system's reliability, interpreted as its strength of propensity to survive. A model for this is $\bar{F}_\mu(t)$; it is based on the premise that $\alpha = \beta = 1$, and that μ is known. A prior for μ

should be based on one's subjective opinion about the number of possible faults in the system, μ being the parameter of the Poisson distribution of M. For purpose of discussion, suppose that μ is endowed with an exponential distribution with a scale parameter $\lambda > 0$.

Were $\bar{F}_\mu(t)$ to be a proper survival function (it is not, because it has a mass $e^{-\mu}$ at $t = +\infty$), then the system's survivability would be given by averaging out $\bar{F}_\mu(t)$ with respect to the exponential (λ) distribution of μ. This means that to obtain the system's survivability, the probability calculus requires that we normalize $\bar{F}_\mu(t)$ to make it a proper survival function, and then interpret it as a probability. This is achieved by considering lifetimes that do not survive indefinitely, namely, by considering

$$P(T > t|T < +\infty, \mu) = \frac{P(t < T < +\infty|\mu)}{P(T < +\infty|\mu)} = \frac{F(+\infty|\mu) - F(t|\mu)}{F(+\infty|\mu)},$$

where $F(t|\mu) = P(T \le t|\mu) = 1 - P(T > t|\mu) = 1 - e^{-\mu(t/(t+1))}$. Thus,

$$P(T > t|T < +\infty, \mu) = \frac{e^{-\mu\tau} - e^{-\mu}}{1 - e^{-\mu}}, \tag{5.3}$$

where $\tau = t/(t+1)$.

Verify that as $t \uparrow +\infty$, $P(T > t|T < +\infty, \mu) \downarrow 0$, and that at $t = 0$, $P(T > t|T < +\infty, \mu) = 1$, making $P(T > t|T < +\infty, \mu)$ a proper probability survival function.

We may now average out $P(T > t|T < +\infty, \mu)$ over the exponential (λ) distribution of μ, to obtain $P(T > t|T < +\infty; \lambda)$, the system's survivability. That is, obtain

$$P(T > t|T < +\infty; \lambda) = \int_0^\infty \frac{e^{-\mu\tau} - e^{-\mu}}{1 - e^{-\mu}} \lambda e^{-\lambda\mu}\, d\mu.$$

The above expression is cumbersome to evaluate. However, for μ large (i.e., for M large—a system with many faults) we may consider the approximation

$$P(T > t|T < +\infty; \lambda) \approx \int_0^\infty e^{-\mu\tau} \lambda e^{-\lambda\mu}\, d\mu = \frac{\lambda}{\lambda + \tau}, \tag{5.4}$$

for $\tau = 0, 1/2, 2/3, \ldots, 1$. Since $\frac{\lambda}{\lambda+\tau} \to \frac{\lambda}{\lambda+1}$, as $t \uparrow \infty$, the above approximation results in a survival function that is defective. To correct for this defect, we subtract $\frac{\lambda}{\lambda+1}$ and normalize (details in Appendix A) to obtain the system's survivability function as

$$P(T > t|T < +\infty; \lambda) = \frac{\lambda(1-\tau)}{\lambda + \tau}.$$

Since $\lambda(1-\tau)/(\lambda+\tau) = 1$ when $t = 0$, and it $\downarrow 0$ as $t \uparrow \infty$, the survivability function is proper.

To summarize, under the assumption that μ is large (i.e., λ is small), the system survivability function is approximated as

$$P(T > t|T < +\infty; \lambda) = \frac{\lambda(1-\tau)}{\lambda + \tau}, \quad \tau = 0, \frac{1}{2}, \frac{2}{3}, \ldots, 1. \tag{5.5}$$

5.3.1 *The conditional reliability and survivability functions.* The matter of filtering reliability and tracking survivability (see Section 6) entails an assessment of quantities like $P(T > t+k|T > t, T < +\infty, \mu)$ and $P(T > t+k|T > t, T < +\infty; \lambda)$. Using the fact that $P(T > t|T < +\infty, \mu)$ is given by equation (5.3), it is easy to verify that

$$P(T > t+k|T > t, T < +\infty, \mu) = \frac{e^{-\mu\tau^*} - e^{-\mu}}{e^{-\mu\tau} - e^{-\mu}}, \tag{5.6}$$

where $\tau^* = (t+k)/(t+k+1)$.

Using L' Hospital's rule, it can be verified that for any value of k, $P(T > t+k|T > t, T < +\infty, \mu)$ equals to one, as $t \to \infty$, and this is intuitively satisfying.

Averaging the right-hand side of equation (5.6) with respect to $\Pi(\mu|T > t, T < +\infty; \lambda)$—the posterior distribution of μ—gives us the conditional survivability function

$$P(T > t+k|T > t, T < +\infty; \lambda) = \int_0^\infty \frac{e^{-\mu\tau^*} - e^{-\mu}}{e^{-\mu\tau} - e^{-\mu}} \Pi(\mu|T > t, T < +\infty; \lambda)\, d\mu.$$

Here again, this expression is cumbersome to evaluate, but for large μ, it can be approximated as

$$P(T > t+k|T > t, T < +\infty; \lambda) \approx \int_0^\infty \frac{e^{-\mu\tau^*}}{e^{-\mu\tau}} \Pi(\mu|T > t, T < +\infty; \lambda),$$

where by Bayes' law

$$\Pi(\mu|T > t, T < +\infty; \lambda) \propto P(T > t|T < +\infty, \mu)\Pi(\mu; \lambda).$$

The middle term of the above expression is given by the right-hand side of equation (5.3), and the last term is the prior for μ, namely, $\lambda e^{-\lambda\mu}$. For μ large, the middle term can be approximated as $e^{-\mu\tau}$, and so the posterior of μ is approximated as $\Pi(\mu|T > t, T < +\infty; \lambda) = c^* \cdot e^{-\mu\tau} \cdot \lambda e^{-\lambda\mu} = c^*\lambda e^{-\mu(\lambda+\tau)}$, where c^*

is the constant of proportionality; its value turns out to be $c^* = (\lambda + \tau)/\lambda$. With the above in place, the posterior distribution of μ is

$$\Pi(\mu | T > t, T < +\infty; \lambda) \approx (\lambda + \tau)e^{-(\lambda+\tau)\mu}, \quad \tau = 0, \frac{1}{2}, \frac{2}{3}, \ldots, 1. \tag{5.7}$$

Note that as $t \to \infty$, this posterior converges to a *fixed* distribution, namely, an exponential distribution with scale parameter $(\lambda + 1)$. The implication of this result is that when μ is large, there is a limit to what can be learned about μ, irrespective of how many successful runs of the system are observed. Interestingly, when there is no restriction placed on μ, the posterior distribution of μ is proportional to $[(e^{-\mu\tau} - e^{-\mu})/(1 - e^{-\mu})]\lambda e^{-\lambda\mu}$, and this approaches zero as t goes to infinity.

Plugging the exponential $(\lambda + \tau)$ distribution in the expression for the conditional survivability function given before, we have

$$\begin{aligned} P(T > t + k | T > t, T < +\infty; \lambda) &\approx \int_0^\infty e^{-\mu(\tau^* - \tau)}(\lambda + \tau)e^{-(\lambda+\tau)\mu}\, d\mu \\ &= \frac{\lambda + \tau}{\lambda + \tau^*}. \end{aligned} \tag{5.8}$$

Analogous to equation (5.6), the result of equation (5.8) is intuitively satisfying, because as $t \to \infty$, $\frac{\lambda+\tau}{\lambda+\tau^*} \to 1$; similarly, for $k = 0$. Furthermore, when $t = 0$,

$$\frac{\lambda + \tau}{\lambda + \tau^*} = \frac{\lambda(k + 1)}{\lambda(k + 1) + k},$$

which is what we would also obtain via equation (5.4), when τ is taken to be $k/(k + 1)$. Note that $\lambda(k + 1)/[\lambda(k + 1) + k]$ is not the result we will get using equation (5.5) with τ taken as $k/(k + 1)$. This is because the right-hand side of equation (5.6) is tantamount to taking the ratio of the improper distribution of $P(T > t | T < +\infty; \lambda)$—namely, equation (5.4).

5.3.2 *Failure rate of the survivability function.* Analogous to the material of Section 5.2, we consider here the *predictive failure rate* (Singpurwalla, 2006, page 66) of the survivability function at $T = \tau$, namely,

$$\begin{aligned} P(T = t | T > t - 1, T < +\infty; \lambda) &= 1 - \frac{P(T > t | T < +\infty; \lambda)}{P(T > t - 1 | T < +\infty; \lambda)}, \end{aligned}$$

where

$$P(T > t | T < +\infty; \lambda) = \lambda(1 - \tau)/(\lambda + \tau);$$

see equation (5.5). It now follows that the predictive failure rate at $T = t$ is

$$1 - \frac{(t - 1)(\lambda + 1)}{t(\lambda + 1) + \lambda} \cdot \frac{\lambda}{t(\lambda + 1) + \lambda}.$$

Verify that the failure rate function decreases in t, starting from 1 at $t = 1$, and decreasing to zero as $t \to \infty$. This is to be expected because having not observed a failure over several trials, increases our strength of belief that the number of faults in the system is small, and also that the P which spawns the T_i's is small.

5.4 Relationship to Cure and Frailty Models of Biostatistics

The scenario leading to equation (5.1) has a parallel in the "cure models" of biostatistics (cf. Chen, Ibrahim and Sinha, 1999). Conceptually, consider an individual possessing an unknown number, say M, of carcinogenic cells, each of which can spawn a continuous latent random variable T_i, $i = 1, \ldots, M$. Each T_i is the time for cell i to produce a detectable cancer. The T_i's are (perhaps unrealistically), assumed independent and identically distributed with distribution function $F(T)$, where $F(0) \equiv 0$. If T is the time for the cancer to relapse, then $T = \min(T_1, \ldots, T_m)$, and

$$P(T > t | M) = \bigl[1 - F(t)\bigr]^M.$$

If M is assumed Poisson (μ), then

$$\bar{G}(t; \mu) = P(T > t; \mu) = \exp\bigl(-\mu F(t)\bigr).$$

As suggested by a referee, the above formula can also be reproduced via alternate arguments involving additive hazard rates.

As $t \to \infty$, $F(t) \to 1$, and so $\bar{G}(t; \mu) \to e^{-\mu}$, implying a nonzero probability mass at $t = \infty$. In biostatistical contexts, $t = \infty$ is interpreted as an individual being cured of the disease, and thus $\bar{G}(t; \mu)$ is known as a *cure model*. The quantity $e^{-\mu}$ is known (cf. Tsodikov, Ibrahim and Yakovlev, 2003) as the *cure fraction*.

It is common to assume that $F(t)$ is a Weibull distribution with unknown parameters. However, if $F(t) = t/(t + 1)$, a location-shifted Pareto distribution with scale 1, then

$$\bar{G}(t; \mu) = \exp\bigl(-\mu t/(t + 1)\bigr), \quad t > 0,$$

which parallels our equation (5.1). Thus, in the spirit of this paper, a cure model of biostatistics would be interpreted as the strength of propensity to survive of an individual undergoing medical therapy.

Despite the above parallel, there are differences, two minor and one substantive. Regarding the minor differences, the first pertains to the fact that the T in our case is discrete because it represents the number of successful trials an entity experiences, whereas the T of cure models represents survival times. The second difference is brought about by the feature that the Beta-Geometric distribution of Section 5.1 has a constructive development, whereas the $F(t)$ in cure models is assumed, and then statistically validated. Inference in cure models pertains to the estimation of μ and the unknown parameters of $F(t)$. For the former, it is common to assume a *link function* of the form $\mu = \exp(\boldsymbol{x}'\boldsymbol{\beta})$, where $\boldsymbol{x}$ is a $(k \times 1)$ vector of covariates, and $\boldsymbol{\beta}$ a vector of unknown parameters. This heavy infrastructure of parameters is typical of work on cure models. It makes the task of inference challenging, a task admirably well addressed in the biostatical literature (cf. Section 7 of Tsodikov, Ibrahim and Yakovlev, 2003).

The key point of departure between the material of this paper, and that on cure models pertains to dynamics. The μ in cure models is assumed unknown, fixed, and dependent only on the values of the covariates at the time T is assessed. This means that the cure fraction $e^{-\mu}$ remains static throughout the course of the therapy. Because the effect of the therapy would be to eliminate or transform cancerous cells, it is meaningful to make μ a function of the time of assessment. This would be akin to how, in the model of this paper, μ is assumed to change because of fault elimination. Thus, from a synergestic viewpoint, our proposed enhancement of cure models would be to endow a dynamic to μ. A way to do this would be via a *system equation* of a filtering mechanism. A strategy for doing so is the topic of Section 6. For now, it suffices to say that the nature of the system equation should encapsulate the perceived character (aggressive or benign) of the cure therapy. Connection with the commonly considered cure models, and a possible approach for enhancing them, can be seen as an added motivation for considering the model proposed here.

5.4.1 *The incorporation of key qualities and frailty models.* A referee of this paper inquired about its possible relationships with *frailty models* of biostatistics, whereas a second referee expressed concerns via the claim that any concept of propensity should entail a description of an item's key qualities. Interestingly, and fortuitously, the two matters are linked. This is because expressions for the strength of propensity can also be produced via the proportional hazards model of Cox (1972).

Specifically, recall, that up until now we have discussed two directions via which a model for the strength of propensity can be developed. One is by invoking the judgment of indifference and symmetry of lifetimes; Sections 2.2 and 2.3 embody this approach. The second direction is via a constructive development based on probability modeling. The material of Section 5.1 and the cure model of this section are examples. There is a third approach popular in biostatistics (cf. Clayton, 1991), and in accelerated life testing (cf. Escobar and Meeker, 2006). The virtue of this third approach is its facility to incorporate biological covariates and the physical qualities of the entity of interest into the analysis.

This third approach focusses on the failure rate of an item's survival function, instead of the lifetimes themselves, as the first two approaches do. Specifically, it parameterizes the failure rate function in terms of an item's biometric covariates, or its physical qualities, depending on the situation at hand. Thus, in the context of this paper, if $h(t)$ denotes an item's model propensity rate (see Section 5.2), then $h(t)$ is parameterized as

$$h(t|\boldsymbol{\beta}; \lambda_0(t), z) = \lambda_0(t) \exp(\boldsymbol{\beta}' z), \tag{5.9}$$

where $h_0(t)$ is some baseline propensity rate, z a vector of covariates (or an item's physical quantities), and $\boldsymbol{\beta}$ a vector of unknown parameters. Implicit to equation (5.9) is the assumption that $\lambda_0(t)$ and z are specified, the latter by direct measurement, so that the only unknown is the vector $\boldsymbol{\beta}$. Conditional on $\boldsymbol{\beta}$, an analogue of equation (5.1) would be

$$P(T > t|\boldsymbol{\beta}; \cdot) = \exp\left(-\int_0^t h(t|\boldsymbol{\beta}; \cdot)\, dt\right),$$

and unconditioning on $\boldsymbol{\beta}$ with respect to its prior or posterior distribution will yield an item's survivability function—an analogue of equation (5.5). The relationship of equation (5.9) generalizes to the case wherein z is a function of time, and even when z is an unknown function of time, so that $\boldsymbol{Z}(t)$ is a stochastic process (cf. Singpurwalla, 1995).

Another kind of generalization of the proportional hazards model leads to frailty models (see Clayton and Cuzick, 1985). Such models allow for positive association between survival times, an association similar to that of exchangeability. To see why, recall that an exchangeable sequence can be generated via mixtures of independent identically distributed random variables in the spirit of Theorem 2.1, and the related theorems of Hewitt and Savage (1955). A frailty model generates a dependent sequence of lifetimes by considering mixtures of failure rates, in the sense described below.

Suppose that the relationship of equation (5.9) is extended to include an additional unknown and unobservable parameter ξ which is endowed with a prior distribution; ξ is called a *frailty*. Now

$$h(t|\boldsymbol{\beta}, \xi; \lambda_0(t), z) = \lambda_0(t)\xi \exp(\boldsymbol{\beta}' z). \tag{5.10}$$

The motivation behind the above relationship is that it is often the case that disease occurrence happens in clusters within families, either because of shared environmental exposures or genetic dispositions. The frailty parameter ξ is therefore shared by all members within a family. Furthermore, individual members have different frailties, and those who are the most frail will experience adverse effects earlier than the others. Consequently, it makes sense to endow a probability distribution to ξ, and such distributions serve a purpose similar in spirit to the mixing distributions of de Finetti's theorem and its generalizations. A consequence is that a hierarchical construction is induced on an item's model propensity rate. Alternatively put, frailty models entail mixtures of failure rates, whereas exchangeability entails mixtures of lifetime distributions. The net effect of both is a collection of interdependent lifetimes.

In frailty models, a commonly used distribution for ξ is the gamma distribution; see equation (4.2) of Clayton (1991). Aalen (1988) discusses other mixing distributions for ξ and invokes these in the context of analyzing two data sets of interest to him.

In the context of reliability, frailty models have appeared in connection with models for system reliability (cf. Lindley and Singpurwalla, 1986) and in matters pertaining to *burn-in testing*; see Block and Savits (1997). Whereas a connection between frailty modeling and burn-in testing remains to be articulated, it does seems to be there all the same. This is because the purpose of burn-in is to eliminate components that are the most frail in order to obtain a batch of items that have more robust lifetimes.

6. FILTERING RELIABILITY AND TRACKING SURVIVABILITY

By filtering, we mean the updating of one's assessment of uncertainty about an unknown state of nature in the light of data and/or the physical dynamics of a system. Filtering entails two relationships, an *observation equation*, and a *state* (or *system*) *equation*. The former describes a connection between what is observed and the state of nature. The latter—irrespective of the basis on which it is specified—is tantamount to a prior distribution on the state of nature. This distribution changes over time, either due to the added knowledge gleaned from observations, or due to the physical evolution of the system. The basic idea underlying the above conceptualization dates back to Laplace, who claimed that the Bernoulli parameter P was the cause of an observed binary random variable, and that one's knowledge about P changes as more and more binary observations are obtained. With filtering, the added dimension is to make provision for the P itself to change over time due to the physics of the scenario.

In the context of our set-up for reliability growth, the unknown state of nature is μ, where μ is embedded in equation (5.1). The dynamics of μ are the result of the fault removal/addition process, and is encapsulated in the prior for μ. In any scenario pertaining to filtering based on lifetimes, one encounters two types of observations, censored or uncensored. That is, at any trial $t > 0$, one either observes a success, wherein $\{T > t\}$, or a failure wherein $\{T = t\}$. Each case spawns its own observation equation and this may one strike as being anathema to the Bayesian paradigm, because the model that is used for inference depends on the observed data. However, this is not an issue, because it is the prior on μ that should not depend on the observed data.

In Section 6.1, we discuss the case of filtering under $\{T > t\}$, and in Section 6.2, the case of $\{T = t\}$. Section 6.3 describes the more realistic case of filtering under the mixed case of both complete and censored observations.

6.1 Filtering under Censorship

Suppose that a system under consideration is observed to have survived $t > 0$ demands. Then tracking its survivability entails assessing, at time t, the unconditional probability $P(T > t + k; T > t, T < +\infty, \lambda)$, for any specified $k = 0, 1, 2, \ldots$. Contrast this probability to the conditional probability $P(T > t + k|T > t, T < +\infty; \lambda)$, conditioned on the event $\{T > t\}$. An assessment of this conditional probability was discussed in Section 5.3.1. In the filtering scenario, the event $\{T > t\}$ has actually been observed.

To proceed formally, we adopt the *irrealis* (see Singpurwalla, 2016) mood and start by considering the conditional probability $P(T > t + k|T > t, T <$

$+\infty; \lambda)$, and follow the development of Section 5.3.1 to write

$$P(T > t+k | T > t, T < +\infty; \lambda) = \int_0^\infty \frac{e^{-\mu\tau^*} - e^{-\mu}}{e^{-\mu\tau} - e^{-\mu}} \Pi(\mu | T > t, T < +\infty; \lambda)\, d\mu,$$

where the middle term on the right-hand side can be interpreted as the *observation equation*

$$P(T > t+k | T > t, T < +\infty, \mu) = \frac{e^{-\mu\tau^*} - e^{-\mu}}{e^{-\mu\tau} - e^{-\mu}},$$

and the last term, namely,

$$\Pi(\mu | T > t, T < +\infty; \lambda) \propto P(T > t | T < +\infty, \mu)\Pi(\mu; \lambda),$$

as the *system equation*.

The above is straightforward, save for the caveat that since $\{T > t\}$ is actually observed, $P(T > t | T < +\infty, \mu)$ cannot be a probability. Rather, it is the *likelihood* of μ with $\{T > t\}$ known, written $\mathcal{L}(\mu; T > t, T < +\infty)$. Thus, the system equation gets re-written as

$$\Pi(\mu; T > t, T < +\infty, \lambda) \propto \mathcal{L}(\mu; T > t, T < +\infty)\Pi(\mu; \lambda), \tag{6.1}$$

and the expression for tracking survivability gets written as $P(T > t+k; T > t, T < +\infty, \lambda)$.

Were one to subscribe to the philosophical *principle of conditionalization* (cf. Howson and Urbach, 2006, page 81), commonly invoked in Bayesian inference, then the likelihood of μ is obtained by reversing the arguments in the probability model for $\{T > t\}$, namely equation (5.3). When such is the case

$$\mathcal{L}(\mu; T > t, T < +\infty) = \frac{e^{-\mu\tau} - e^{-\mu}}{1 - e^{-\mu}} \approx e^{-\mu\tau},$$

for μ large, and $\tau = t/(t+1)$. Since τ takes the values $\frac{1}{2}, \frac{2}{3}, \ldots, 1$, the likelihood of μ is enveloped between $e^{-\mu}$ and $e^{-\mu/2}$; see Figure 2.

Under this choice of the likelihood function, the posterior distribution of μ, with $\{T > t\}$ observed, is an exponential distribution with a scale parameter $(\lambda + \tau)$. This means that under censorship, a change in μ is purely due to a change of opinion brought about by added information. It is not due to the dynamics of the system—a matter that arises when the observation is a complete one, namely, $\{T = t\}$; see Section 6.2.

As a consequence of the above posterior distribution, the system's survivability function (having assumed that μ is large), can be approximated as

$$P(T > t+k; T > t, T < +\infty, \lambda) \approx \frac{\lambda + \tau}{\lambda + \tau^*}, \tag{6.2}$$

where $\tau = \frac{t}{t+1}$ and $\tau^* = \frac{\lambda+k}{t+k+1}$.

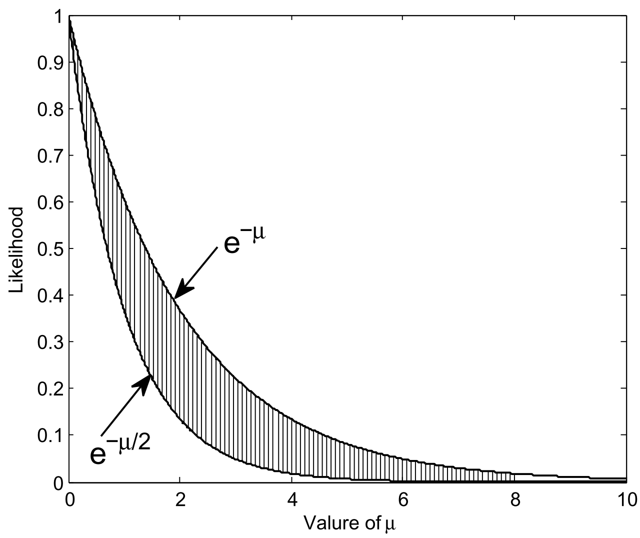

FIG. 2. *The likelihood envelope for any fixed τ.*

The development above parallels that of Section 5.3.1, the only difference being that the $T > t$ is no longer a conditioning event, and that the conditional probability $P(T > t | T < +\infty, \mu)$ of Section 5.3.1 is now interpreted as a likelihood. However, there is a philosophical import of the result of equation (6.2). It has to do with the claim that no amount of observation can ever prove a law (e.g., Laplace's rule of succession), and the dilemma this posed to the likes of Jeffreys (see Singpurwalla and Wilson, 2004). Since $P(T > t+k; T > t, T < +\infty, \lambda)$ tends to 1 as $t \to \infty$, irrespective of whether $k < 0$ or $k > 0$, we have here a scenario wherein an infinite number of successful tests do prove a law, namely, the law that the system is *infallible*.

6.2 Filtering under Observed Failures

Suppose that the system under test experiences a failure at $T = t$, for some $t > 0$. Associated with the event $\{T = t\}$, one will also observe n_t number of faults, where $n_t \geq 1$; in what follows we do not account for the nature and the type of faults. Once a system failure is encountered, the failure-causing faults are attempted to be eliminated. However, the fault elimination process may also result in the introduction of new faults. It is the process of fault elimination and introduction which gives the system its dynamics, because the net effect of these actions is to cause μ to change. The dynamics of μ are encapsulated via a hierarchically constructed prior for μ; see Section 6.2.1. This prior distribution will go to determine the filtering mechanism's observation equation. Under the above set-up, one's assessment of the uncertainty about μ will be updated based on the consequence of observing the event $\{T = t\}$, and also on the value of t itself. A large t signals a small μ and vice-versa. Alternatively put, a failure causes μ to change because of the dynamics of fault removal, and the exact time of failure changes one's opinion of μ.

Following the line of reasoning adopted in Section 6.1, tracking the system's survivability having observed a failure at say $T^* = t$, and n_t the number of faults, and having attempted to eliminate the faults, we inquire as to what $T > t + k$ is. That is, we need to assess $P(T > t+k; T^* = t, n_t, \cdot)$. To do so, we adopt the irrealis mood by conditioning on $T^* = t$, and extending the conversation to μ (see Lindley, 1990), obtain $P(T > t+k|T^* = t; n_t, \cdot)$ as

$$\int_0^\infty P(T > t+k|T^* = t, \mu) \cdot \Pi(\mu|T^* = t; n_t, \cdot)\, d\mu. \tag{6.3}$$

Using equation (5.3), and assuming that μ is large and t is moderate, the first term in the above expression can be approximated as $e^{-\mu\hat{\tau}}$, where $\hat{\tau} = \frac{k+1}{t(t+k+1)}$; see Appendix B for details. Thus, equation (6.3) simplifies as

$$P(T > t+k|T^* = t; n_t, \cdot) \approx \int_0^\infty e^{-\mu\hat{\tau}}\, \Pi(\mu|T^* = t; n_t, \cdot)\, d\mu. \tag{6.4}$$

The last term in the above expression is the posterior distribution μ, and the challenge is to obtain this distribution incorporating the nuances of the fault elimination/introduction process. Recall that this posterior distribution is to be interpreted as the system equation of the filtering and tracking mechanism.

6.2.1 *A hierarchically constructed prior on μ.* We start with the assumption that conditional on some $\delta > 0$, and with λ specified, μ has an exponential distribution with a scale parameter $\delta\lambda$. That is,

$$\Pi(\mu|\delta; \lambda) = \delta\lambda \exp(-\delta\lambda\mu), \quad \mu > 0.$$

We next assume that δ itself has an exponential distribution with parameter ω, where ω is also specified. That is,

$$\Pi(\delta; \omega) = \omega \exp(-\omega\delta), \quad \delta > 0.$$

The rationale for these choices is the following. When $\delta > 1$, μ decreases (stochastically); this accommodates the feature that the failure-causing faults are eliminated. When $\delta < 1$, μ increases (stochastically), and this incorporates the feature that new faults could be introduced into the system. When $\delta = 1$, μ does not experience a physically caused change; this happens when $T > t$ is observed. The behavior of δ therefore captures the dynamics of the fault elimination/introduction process. As a consequence of the above, the prior on μ is a *location-scale Pareto* distribution with λ and ω as specified parameters; that is,

$$\Pi(\mu; \lambda, \omega) = \frac{\lambda\omega}{(\lambda\mu + \omega)^2}, \quad \mu > 0. \tag{6.5}$$

This prior distribution is heavy tailed, and has no moments. Its virtue is a type of conjugacy which simplifies the process of obtaining a closed form posterior distribution of μ; see Section 6.2.2 below.

6.2.2 *Subjective likelihood of δ and induced posterior of μ.* Since the parameter δ influences μ, and since δ controls the dynamics of the process, it makes sense to specify a likelihood of δ under the observed $(T = t)$ and n_t, and to then obtain its posterior distribution. This posterior distribution can be used to induce the posterior distribution of μ. Such a schemata obviates the need for specifying a likelihood of μ—which can be a cumbersome exercise—and then obtaining a posterior distribution of μ. In what follows, we describe the mechanics of the above process.

Likelihood functions are conventionally specified by leaning on the principle of conditionalization, which boils down to reversing the arguments in a probability model. However, likelihood functions can also be specified via subjective arguments (cf. Singpurwalla, 2007), and this is what is done in the case of δ. Specifically, the likelihood of δ in its posterior distribution

$$\Pi(\delta; T = t, n_t, \omega) \propto \mathcal{L}(\delta; T = t, n_t)\Pi(\delta; \omega),$$

is subjectively specified as

$$\mathcal{L}(\delta; T = t, n_t) = \exp(-\delta f(t) g(n_t)),$$

where $f(t)$ is a decreasing function of t and $g(n_t)$ an increasing function of n_t. As an example, $f(t) = 1/t$ and $g(n_t) = n_t$. This likelihood is an exponentially decreasing function of δ, with the rate of decrease determined by $J = f(t)g(n_t)$. The rationale behind this choice is:

(i) When t is large, J is small, and thus δ is large signaling a growth in reliability, because $E(\mu|\delta; \lambda) = 1/\lambda\delta$.

(ii) When n_t is large, J is large, and this encapsulates the feature that in the process of eliminating a large number of the n_t faults, a large number of new faults may also be introduced in the system.

(iii) When both n_t and t are large, their effects tend to cancel out, suggesting that at the later stages of testing, there is a greater tendency to eliminate faults without introducing new ones.

Other forms of the likelihood are of course possible. With the likelihood given above, the posterior distribution of δ is an exponential with a scale parameter $(\omega + n_t/t)$. That is,

$$\Pi(\delta; T=t, n_t, \omega) = \left(\omega + \frac{n_t}{t}\right)\exp\left(-\delta\left(\omega + \frac{n_t}{t}\right)\right).$$

From the hierarchical construction which resulted in the location-scale Pareto as a prior distribution of μ—namely, equation (6.5)—the posterior distribution of μ with $T = t$ and n_t specified, is also a location-scale Pareto; specifically,

$$\begin{aligned}(6.6)\qquad &\Pi(\mu; T=t, n_t, \lambda, \omega)\\ &\quad = \frac{\lambda(\omega + n_t/t)}{(\lambda\mu + \omega + n_t/t)^2}, \qquad \mu > 0,\end{aligned}$$

and as mentioned before, this is the system equation of the filtering algorithm. As pointed out by a referee, the relationship between equations (6.5) and (6.6) suggest a connection between the proposed construction of the prior, and the theory of *invariant conditional distributions* due to Bather (1965).

6.2.3 *Assessing the survivability function.* The penultimate step of this section is assessing the system's survivability, namely, assessing the predictive distribution of T, having observed $T = t$, and n_t. This would entail invoking the result of equation (6.6) in the context of equation (6.4), suitably re-written to account for the fact that $(T = t)$ has actually been observed. That is we need to assess $P(T > t + k; T = t, n_t, \lambda, \omega)$, which for large μ can be approximated [see equation (6.4)] as

$$\begin{aligned}&P(T > t+k; T=t, n_t, \lambda, \omega)\\ &\quad \approx \int_0^\infty e^{-\mu[\frac{k+1}{t(t+k+1)}]}\Pi(\mu; T=t, n_t, \lambda, \omega)\,d\mu.\end{aligned}$$

If we set $\tilde{\omega} = (\omega + n_t/t)$, then with $\hat{\tau} = \frac{k+1}{t(t+k+1)}$, the above relationship becomes

$$\begin{aligned}&\int_0^\infty e^{-\mu\hat{\tau}}\frac{\lambda\tilde{\omega}}{(\lambda\mu + \tilde{\omega})^2}\,d\mu\\ &\quad = 1 + e^{\tilde{\tau}\tilde{\omega}/\lambda}\left(\frac{\tilde{\tau}\tilde{\omega}}{\lambda}\right)E_i\left(-\frac{\tilde{\tau}\tilde{\omega}}{\lambda}\right),\end{aligned}$$

where $E_i(z) = -\int_{-z}^{\infty}\frac{e^{-u}}{u}\,du$ is the *exponential integral function*; see Appendix C for details.

To obtain a closed form expression for the survivability function, one possibility is to approximate the expression for the posterior distribution of μ, namely, $\lambda\tilde{\omega}/(\lambda\mu + \tilde{\omega})^2$ by an exponential function $\theta e^{-\mu\theta}$, where $\theta = \lambda/\tilde{\omega}$ and $\lambda \ll \tilde{\omega}$. The rationale behind this latter choice is that when $\mu = 0$, $\lambda\tilde{\omega}/(\lambda\mu + \tilde{\omega})^2 = \lambda/\tilde{\omega}$, whereas with $\mu = 0$, $\theta e^{-\mu\theta} = \theta$. Thus to match the two functions at $\mu = 0$, we set $\theta = \lambda/\tilde{\omega}$. The requirement that $\lambda \ll \tilde{\omega}$ is motivated by the feature that the location-scale Pareto distribution has an infinite first moment, which is tantamount to the requirement that $1/\theta$—the mean of the exponential distribution—be infinite. This can happen when $\lambda \ll \tilde{\omega}$. Because small values of λ and large values of $\tilde{\omega}$ correspond to large values of μ, the proposed approximation is appropriate when the prior assumption that the system has many faults is germane. The quality of the approximation is discussed in Appendix D.

As a consequence of the proposed approximation, the system's predictive distribution is given as $P(T > t + k; T = t, n_t, \lambda, \omega) \approx \int_0^\infty e^{-\mu\hat{\tau}}\theta e^{-\theta\mu}\,d\mu$, from which it follows that for $\lambda \ll \tilde{\omega} = (\omega + n_t/t)$,

$$(6.7)\quad P(T > t+k; T=t, n_t, \lambda, \omega) = \frac{\theta}{\theta + \frac{k+1}{t(t+k+1)}},$$

where $\theta = \lambda/\tilde{\omega}$. Verify that when t gets large, the above probability becomes closer to one, as is to be expected. Furthermore, when $k = 0$, this probability is, for any $t < +\infty$, less than one, and this makes sense. Equation (6.7) prescribes the survivability of a system that has encountered a failure at t with n_t observed faults, which have been attempted to be eliminated.

6.3 Filtering Under Censored and Complete Observations

In this section, we propose an algorithm for addressing the more realistic scenario of filtering when censored and complete observations get intermingled. That is, an observed failure at t is, after suitable attempts at fault elimination, followed by a survival at $(t + 1)$, or a survival at t is followed by a failure at $(t + 1)$. The matter is less than straightforward because the posterior distributions of μ under censored and under complete observations are members of different families, exponential per Section 6.1, and a location-scale Pareto, per Section 6.2. An observed survival adds to the parameter λ whereas an observed failure adds to the parameter ω. The lack of a common family poses a difficulty when transiting from an observed survival at trial t, to an observed failure at trial $(t + 1)$. A way out of this difficulty is to make the schemata of

Sections 6.1 and 6.2 work in concert via an algorithm described below. The algorithm entails an approximation, the basis of which is the material of Appendix D which pertains to a strategy for replacing a location-scale Pareto with a suitably matched exponential. In the case of observed failures, the location-scale Pareto provides a prescription for updating the underlying parameters, whereas the approximating exponential provides a mechanism for transiting from a prior to a posterior in closed form.

In the interest of clarity, it is desirable to index the parameters λ and ω by their associated trial numbers, so that with λ_0 and ω_0 denoting their starting values, λ_{t-1} denotes the value of λ prior to an observation at t, and λ_t the value of λ posterior to an observation at t. In principle, the posterior value of a parameter at t, becomes its prior value at $(t+1)$, similarly, with the parameter ω.

6.3.1 *The filtering algorithm.* For purposes of discussions, focus on trial t. Prior to an observation at t, we have at hand λ_{t-1} and ω_{t-1}. Were a survival to be observed at t, that is, $T > t$, then the posterior distribution of μ_t would be an exponential (λ_t), where $\lambda_t = \lambda_{t-1} + \frac{t}{t+1}$. By contrast, were $T = t$ to be observed, then the posterior of μ_t would be a location-scale Pareto with parameters $\lambda_t = \lambda_{t-1}$ and $\omega_t = \omega_{t-1} + n_t/t$, where n_t is the observed number of faults at t.

Now suppose that at trial $(t+1)$ we observe a survival, that is, $T > (t+1)$. Then the posterior distribution of μ_{t+1} will depend on what was observed at trial t. If it was a survival, then the posterior of μ_{t+1} will be an exponential with parameter $\lambda_{t+1} = \lambda_t + \frac{t+1}{t+2}$. If, however, $T = t$, the posterior of μ_{t+1} will be the approximating exponential λ_{t+1} where $\lambda_{t+1} = \frac{\lambda_t}{\omega_t} + \frac{t+1}{t+2}$.

By contrast, suppose that the trial at $(t+1)$ results in an observed failure, that is, $T = (t+1)$. Then, the posterior distribution of μ_{t+1}, should $T = t$, will be a location-scale Pareto with $\lambda_{t+1} = \lambda_t$ and $\omega_{t+1} = \omega_t + \frac{n_{t+1}}{t+1}$, where n_{t+1} is the number of faults associated with the observed failure at $(t+1)$. However, were $T > t$ to be observed at trial t, then the posterior distribution of μ_{t+1} would also be a location-scale Pareto but with $\lambda_{t+1} = \lambda_t$ and $\omega_{t+1} = \omega_{t-1} + \frac{n_{t+1}}{t+1}$; we use here ω_{t-1} and not ω_t, because the observed survival at t does not update the parameter ω.

Appendix E provides a graphic of the proposed algorithm.

6.3.2 *Discussion of the algorithm.* The proposed algorithm is atypical of filtering, because of the switch in the prior of μ from an exponential to a Pareto, depending on whether a survival of a failure is observed. This could, from a Bayesian point of view, be objectionable. Priors are supposed to be chosen before observing any data.

However, such an objection can be overruled on two grounds:

(i) The switched priors pertain to two different observation equations: one for censored observations, and the other for complete observations.

(ii) The switched priors carry information about μ from its immediately preceding posterior, and the strategy of when to switch is declared in advance, contingent on what will be observed.

Indeed, the filtering procedure proposed here is unique, because it entails two types of observation and system equations. These are necessitated by the intermingled occurrence of censored and complete observations. Such observations are characteristic of survival time data. To the best of our knowledge, there appears to be a dearth of literature on filtering under censorship. Thus, it is our view that the approach proposed here is a *prototype* for filtering, prediction and control under complete and censored observations.

7. SUMMARY AND CONCLUSIONS

In essence, this paper comprises of two parts, a conceptual part and a methodological part. In principle, each part may be appreciated independently of the other. The conceptual part is spawned by the methodological part. The latter is motivated by the need to address a class of problems that occur in the actuarial, the biological and the engineering sciences. To address such problems, a conceptual framework is needed, and the philosophical part of the paper pertains to developing the framework. But the process of doing so raises issues that question the foundational underpinnings of reliability and of survival analysis. These issues stem from the meaning of probability, the methods of quantifying uncertainty and their relevance in the contexts of the life sciences. Embedded within these issues are the long ignored notions of propensity and chance. Also overlooked, or perhaps unrecognized, are some parallels in the perspectives of Pierce, Popper, de Finetti and Kolmogorov on matters of uncertainty and causality. The viewpoint arrived upon in this paper is that the notions of chance and propensity need to work in concert with that of personal probability to produce the framework we need. A consequence is that reliability should be interpreted as a strength of propensity—and not as a probability—the strength of propensity being quantified via a mathematical model with unknown parameters. Often this model is a well-known failure model like the exponential and the Weibull, though this need not be so. One's uncertainty about the strength of propensity, encapsulated as a personal probability, is known as survivability. Since survivability can be operationalized via a 2-sided bet, it is the survivability of an item that should be of relevance in practice. There are broader implications to this paradigm shift because it goes beyond the life sciences and sits at the doorstep of applied probability itself. In particular, topics such as inventory, queueing, Kalman filtering and time series analysis need to be revisited, namely their foundational underpinnings.

From a methodological point of view, this paper makes the claim that an item's performance characteristics change over time due to the acquisition of immunity, clinical therapies and defect elimination, and must therefore be time indexed. This means that the lifetime characteristics of an item should be analyzed using the techniques of filtering and tracking. An illustration of how this can be done, albeit with a plethora of approximations, occupies the bulk of the latter part of this paper. The filtering algorithm presented here is by no means a final word. It is merely an indication of possibilities without resorting to numerical methods or heavy computing via Monte Carlo. More important, 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.

APPENDIX A: NORMALIZING THE APPROXIMATE SYSTEM SURVIVABILITY FUNCTION

Subtracting $\lambda/(\lambda+1)$ from $\lambda/(\lambda+\tau)$ and normalizing via a constant C, we have

$$P(T>t|T<+\infty;\lambda)=C\left(\frac{\lambda}{\lambda+\tau}-\frac{\lambda}{\lambda+1}\right)=\frac{\lambda}{\lambda+1}\frac{(1-\tau)}{\lambda+\tau}C,$$

which for $t\uparrow\infty$ (i.e., $\tau=1$) must be 0 for any value of C, which it is. However, when $t=0$ (i.e., $\tau=0$),

$$\frac{\lambda}{\lambda+1}\frac{(1-\tau)}{\lambda+\tau}C=1,$$

only when $C=(\lambda+1)$. Thus, the normalized system survivability function is $P(T>t|T<+\infty;\lambda)=\frac{\lambda(1-\tau)}{\lambda+\tau}$.

APPENDIX B: ASSESSING CONDITIONAL PROBABILITY UNDER FAILURE

We need to assess

$$P(T>t+k|T=t,\mu,T<+\infty)=\frac{P(T>t+k|\mu,T<+\infty)}{P(T=t|\mu,T<+\infty)},$$

where [see equation (5.3)],

$$P(T>t+k|\mu,T<+\infty)=\frac{e^{-\mu(\frac{t+k}{t+k+1})}-e^{-\mu}}{1-e^{-\mu}},$$

and

$$P(T=t|\mu,T<+\infty)=P(T>t-1|\mu,T<+\infty)-P(T>t|\mu,T<+\infty).$$

Thus, the required probability, when μ is large, is approximately $e^{-\mu(\frac{t+k}{t+k+1})}/[e^{-\mu(\frac{t-1}{t})}-e^{-\mu(\frac{t}{t+1})}]$. Dividing the numerator and denominator by $e^{-\mu(\frac{t-1}{t})}$, we have

$$P(T>t+k|T=t,\mu,T<+\infty)\approx\frac{e^{-\mu[\frac{k+1}{t(t+k+1)}]}}{1-e^{-\mu(\frac{t}{t+1})}}\approx e^{-\mu[\frac{k+1}{t(t+k+1)}]},$$

if $\mu\gg t$.

APPENDIX C

To evaluate the quantity $\int_0^\infty e^{-\mu\hat{\tau}}\frac{\lambda\tilde{\omega}}{(\lambda\mu+\tilde{\omega})^2}\,d\mu$, where $\hat{\tau}=(k+1)/[t(t+k+1)]$ and $\tilde{\omega}=(\omega+n_t/t)$, we

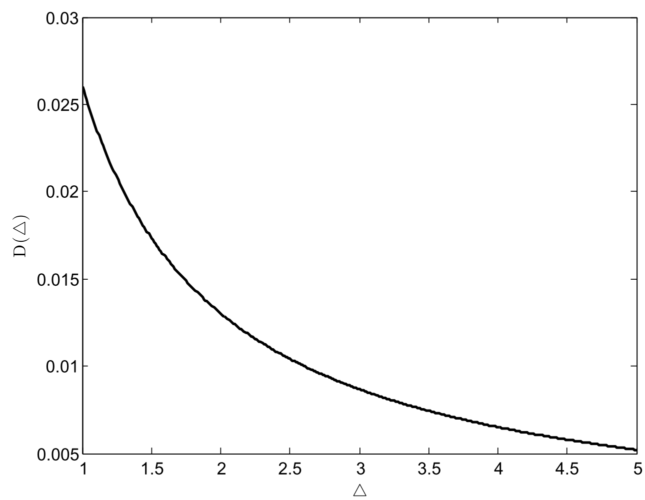

FIG. D.1. *The plot of $D(\triangle)$ as a function of $\triangle$.*

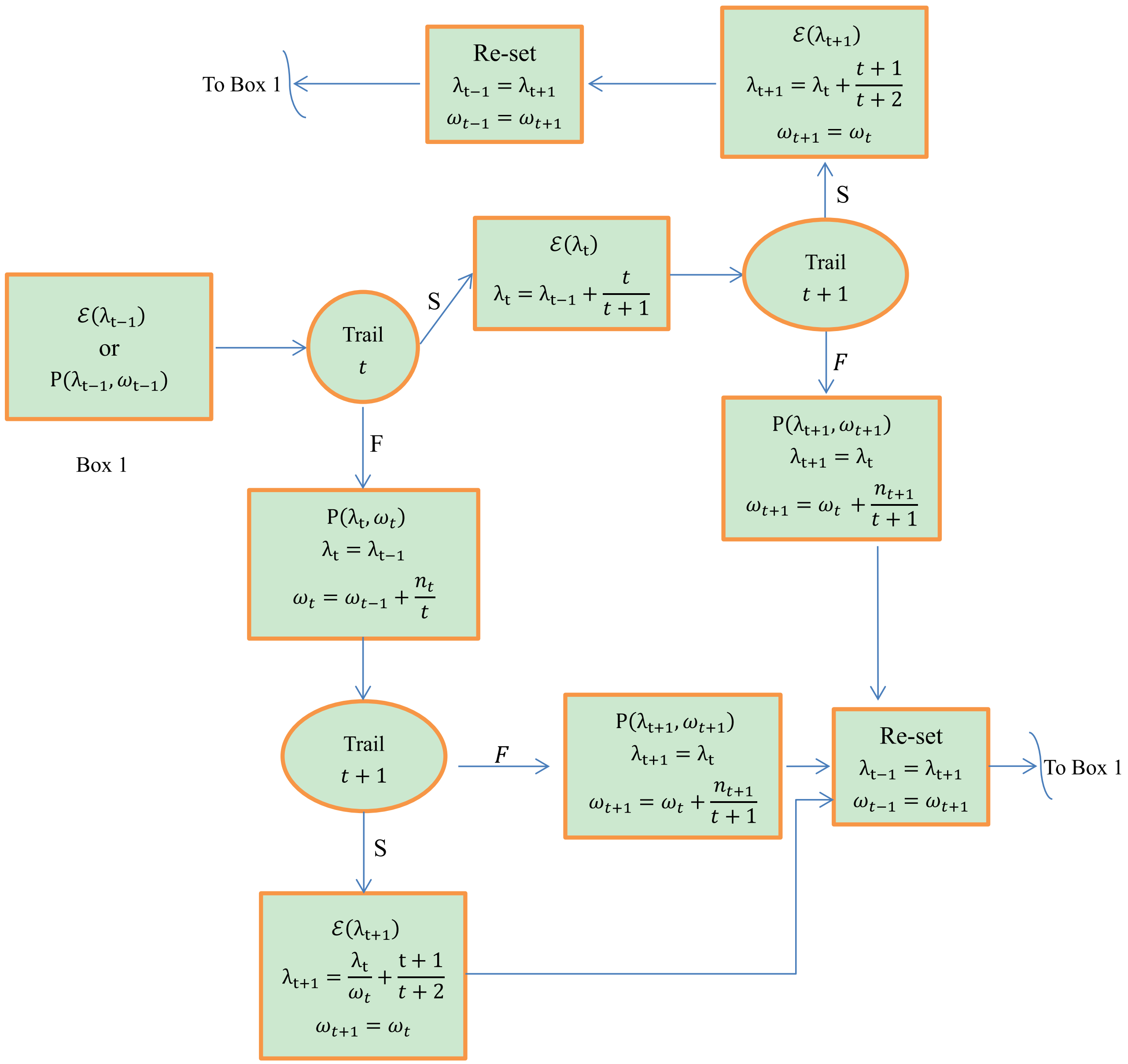

FIG. E.1. *Schemata of the filtering algorithm.*

integrate by parts to obtain

$$\int_0^\infty e^{-\mu\hat{\tau}} \frac{\lambda\tilde{\omega}}{(\lambda\mu+\tilde{\omega})^2}\,d\mu = 1 - \int_0^\infty e^{-\mu\hat{\tau}} \frac{\hat{\tau}\omega}{\lambda\mu+\omega}\,d\mu$$
$$= 1 - \frac{\hat{\tau}\omega}{\lambda} e^{\hat{\tau}\omega/\lambda} \int_{\hat{\tau}\omega/\lambda}^\infty \frac{e^{-u}}{u}\,du,$$

from which it follows that

$$\int_0^\infty e^{-\mu\hat{\tau}} \frac{\lambda\tilde{\omega}}{(\lambda\mu+\tilde{\omega})^2}\,d\mu = 1 + \frac{\hat{\tau}\omega}{\lambda} e^{\hat{\tau}\omega/\lambda} E_i\left(-\frac{\tilde{\tau}\tilde{\omega}}{\lambda}\right),$$

where $E_i(z) = -\int_{-z}^\infty \frac{e^{-u}}{u}\,du$.

APPENDIX D: APPROXIMATING A LOCATION-SCALE PARETO BY AN EXPONENTIAL

We wish to approximate the function $\frac{\lambda\tilde{\omega}}{(\lambda\mu+\tilde{\omega})^2}$ by $\theta e^{-\mu\theta}$, where $\theta = \lambda/\tilde{\omega}$. If we set $\triangle = 1/\theta = \tilde{\omega}/\lambda$, then $\frac{\lambda\tilde{\omega}}{(\lambda\mu+\tilde{\omega})^2} = \frac{\triangle}{(\mu+\triangle)^2} = h(\mu)$, say, and $\theta e^{-\mu\theta} = \frac{1}{\triangle}e^{-\mu/\triangle} = \ell(\mu)$, say.

Define $D(\triangle) = \int_0^\infty |h(\mu) - \ell(\mu)|^2\,d\mu$. Then a plot of $D(\triangle)$ as a function $\triangle$—shown in Figure D.1—reveals the feature that as $\triangle$ increases $D(\triangle)$ decreases, and that when $\triangle > 3$, $D(\triangle) < 0.001$. Thus, the larger the ratio $\tilde{\omega}/\lambda$, the better the approximation.

APPENDIX E: A GRAPHIC OF THE FILTERING ALGORITHM FOR μ

In Figure E.1, $S(F)$ denotes survival (failure), where survival means that $T > t$ and failure means that $T = t$. Furthermore, $\mathcal{E}(\lambda)$ denotes an exponential distribution with scale parameter (λ) and $P(\lambda, \omega)$ denotes a location-scale Pareto distribution with parameters λ and ω. The state of nature μ, encapsulated via a posterior distribution and its parameters, is encased in the rectangular boxes, whereas the circles denote random nodes at which either an S or an F is observed; associated with an F is also n_t, the number of observed faults.

ACKNOWLEDGMENTS

Research supported in part by a grant from the City University of Hong Kong (Project Nos. 9380068 and 9042083, and by the U.S. Office of Naval Research - Global, Grant N62909-15-1-2026). Aspects of the work described here have benefited from input by Owen Martin to whom the author is greatly obliged. J. Sethuraman's comments on the cure model being an improper distribution have provided rigor to the material, and several corrections by Michael Edesess have provided editorial clarity to this version of the paper. Lai Boya and Mimi Zhang helped put the material in a better format. Finally, an offhand comment by Jim Berger has provided a clearer pathway to the author's thinking. To all I am deeply obliged. The very substantive comments by the two referees of this paper have expanded its scope.

REFERENCES

AALEN, O. O. (1988). Heterogenity in survival analysis. *Stat. Med.* **7** 1121–1137.

BARLOW, R. E. and MENDEL, M. B. (1992). De Finetti-type representations for life distributions. *J. Amer. Statist. Assoc.* **87** 1116–1122. MR1209569

BARLOW, R. E. and PROSCHAN, F. (1965). *Mathematical Theory of Reliability*. Wiley, New York-London-Sydney. With Contributions by Larry C. Hunter. MR0195566

BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing*: *Probability Models*. Holt, Rinehart and Winston, Inc., New York. MR0438625

BATHER, J. A. (1965). Invariant conditional distributions. *Ann. Math. Stat.* **36** 829–846. MR0179840

BLOCK, H. W. and SAVITS, T. H. (1997). Burn-in. *Statist. Sci.* **12** 1–13.

CHEN, M.-H., IBRAHIM, J. G. and SINHA, D. (1999). A new Bayesian model for survival data with a surviving fraction. *J. Amer. Statist. Assoc.* **94** 909–919. MR1723307

CLAYTON, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* **47** 467–485.

CLAYTON, D. and CUZICK, J. (1985). Multivariate generalizations of the proportional hazards model. *J. Roy. Statist. Soc. Ser. A* **148** 82–117. MR0806480

COX, D. R. (1972). Regression models and life-tables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **34** 187–220. With discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox. MR0341758

DE FINETTI, B. (1937). La prévision : Ses lois logiques, ses sources subjectives. *Ann. Inst. Henri Poincaré* **7** 1–68. MR1508036

DE GROOT, M. H. (1988). A Bayesian view of assessing uncertainty and comparing expert opinion. *J. Statist. Plann. Inference* **20** 295–306. MR0976182

DIACONIS, P. and FREEDMAN, D. (1987). A dozen de Finetti-style results in search of a theory. *Ann. Inst. Henri Poincaré Probab. Stat.* **23** 397–423. MR0898502

DUBEY, S. D. (1966). Compound Pascal distributions. *Ann. Inst. Statist. Math.* **18** 357–365. MR0208712

ESCOBAR, L. A. and MEEKER, W. Q. (2006). A review of accelerated test models. *Statist. Sci.* **21** 552–577. MR2380715

FETZER, J. H. (1981). *Scientific Knowledge*: *Causation*, *Explanation*, *and Corroboration*. *Boston Stud. Philos. Sci.* **69**. D. Reidel, Dordrecht.

GILLIES, D. (2000). Varieties of propensity. *British J. Philos. Sci.* **51** 807–835. MR1813908

GOOD, I. J. (1950). *Probability and the Weighing of Evidence*. Charles Griffin & Co., Ltd., London. MR0041366

GOOD, I. J. and CARD, W. I. (1971). The diagnostic process with special reference to errors. *Methods Inf. Med.* **10** 176–188.

HAMADA, M. S., WILSON, A. G., REESE, C. S. and MARTZ, H. F. (2008). *Bayesian Reliability*. Springer, New York. MR2724437

HEWITT, E. and SAVAGE, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80** 470–501. MR0076206

HOWSON, C. and URBACH, P. (2006). *Scientific Reasoning*: *The Bayesian Approach*. Open Court, Chicago, IL.

HUMPHREYS, P. (1985). Why propensities cannot be probabilities. *Philos. Rev.* **94** 557–570.

KENDALL, M. G. (1949). On the reconciliation of theories of probability. *Biometrika* **36** 101–116. MR0035407

KOLMOGOROV, A. N. (1969). The theory of probability. In *Mathematics*: *Its Content*, *Methods and Meaning*, Vol. 2, Part 3. (A. D. Aleksandrov, A. N. Kolmogorov and M. A. Lavrentav, eds.). MIT Press, Cambridge, MA.

LINDLEY, D. V. (1990). The 1988 Wald memorial lectures: The present position in Bayesian statistics. *Statist. Sci.* **5** 44–89. With comments and a rejoinder by the author. MR1054857

LINDLEY, D. V. and SINGPURWALLA, N. D. (1986). Multivariate distributions for the life lengths of components of a system sharing a common environment. *J. Appl. Probab.* **23** 418–431. MR0839996

LINDLEY, D. V. and SINGPURWALLA, N. D. (2002). On exchangeable, causal and cascading failures. *Statist. Sci.* **17** 209–219. MR1939340

MARTIN, O. S. (2012). *A Dynamic Competing Risk Model for Filtering Reliability and Tracking Survivability*. ProQuest LLC, Ann Arbor, MI. Ph.D. thesis, The George Washington University. MR2995920

MEEKER, W. Q. and ESCOBAR, L. A. (1998). *Statistical Methods for Reliability Data*. Wiley, New York.

MEINHOLD, R. and SINGPURWALLA, N. D. (1983). Understanding the Kalman filter. *Amer. Statist.* **37** 123–127.

MILLER, R. W. (1975). Propensity: Popper or peirce? *British J. Philos. Sci.* 123–132.

POPPER, K. R. (1957). The propensity interpretation of the calculus of probability, and the quantum theory. In *Observation and Interpretation in the Philosophy of Physics* (S. Korner, ed.). Dover, Mineola, NY.

POPPER, K. R. (1959). The propensity interpretation of probability. *British J. Philos. Sci.* **10** 25–42.

SINGPURWALLA, N. D. (1988). Foundational issues in reliability and risk analysis. *SIAM Rev.* **30** 264–282. MR0941113

SINGPURWALLA, N. D. (1995). Survival in dynamic environments. *Statist. Sci.* **10** 86–113.

SINGPURWALLA, N. D. (2002). Some cracks in the empire of chance (flaws in the foundations of reliability). *Int. Stat. Rev.* **70** 53–65.

SINGPURWALLA, N. D. (2006). *Reliability and Risk*: *A Bayesian Perspective*. Wiley, Chichester. MR2265917

SINGPURWALLA, N. D. (2007). Betting on residual life: The caveats of conditioning. *Statist. Probab. Lett.* **77** 1354–1361. MR2392806

SINGPURWALLA, N. D. (2010). Comment: "Reliability growth management metrics and statistical methods for discrete-use systems" [MR2789242]. *Technometrics* **52** 396–397. MR2779425

SINGPURWALLA, N. D. (2016). The indicative and irrealis moods in Bayesian inference. Under review.

SINGPURWALLA, N. D. and WILSON, P. (2004). When can finite testing ensure infinite trustworthiness? *J. Iran. Stat. Soc.* (*JIRSS*) **3** 1–37.

SMITH, A. F. M. (1981). On random sequences with centred spherical symmetry. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **43** 208–209. MR0626767

SPIZZICHINO, F. (2001). *Subjective Probability Models for Lifetimes*. *Monographs on Statistics and Applied Probability* **91**. Chapman & Hall/CRC, Boca Raton, FL. MR1980207

TSODIKOV, A. D., IBRAHIM, J. G. and YAKOVLEV, A. Y. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *J. Amer. Statist. Assoc.* **98** 1063–1078. MR2055496

VON MISES, R. (1941). On the foundations of probability and statistics. *Ann. Math. Stat.* **12** 191–205. MR0004386