Markov models of aging: Theory and practice

David Steinsaltz[1,*], Gurjinder Mohan[1], Martin Kolb[2]

Abstract

We review and structure some of the mathematical and statistical models that have been developed over the past half century to grapple with theoretical and experimental questions about the stochastic development of aging over the life course. We suggest that the mathematical models are in large part addressing the problem of partitioning the randomness in aging: How does aging vary between individuals, and within an individual over the lifecourse? How much of the variation is inherently related to some qualities of the individual, and how much is entirely random? How much of the randomness is cumulative, and how much is merely short-term flutter?

We propose that recent lines of statistical inquiry in survival analysis could usefully grapple with these questions, all the more so if they were more explicitly linked to the relevant mathematical and biological models of aging. To this end, we describe points of contact among the various lines of mathematical and statistical research. We suggest some directions for future work, including the exploration of information-theoretic measures for evaluating components of stochastic models as the basis for analyzing experiments and anchoring theoretical discussions of aging.

1. Introduction: Stochasticity and aging

A famous definition Arking (2006) identifies aging with *CUPID* developments in an organism — changes that are cumulative, universal, progressive, inherent, and deleterious. And yet, one of the key markers of aging, advancing mortality, is essentially random: for most species we may speak only of a universally increasing risk of death — *force of mortality* to the demographers, *mortality rate* to the biologists, *hazard rate* to the statisticians. This creates peculiar challenges for mathematical and statistical modeling, since this risk is not observable or measurable in any single individual. It is an artifact of the ensemble.

*Corresponding author

Email addresses: steinsal@stats.ox.ac.uk (David Steinsaltz), mohan@stats.ox.ac.uk (Gurjinder Mohan), kolb@math.lmu.de (Martin Kolb)

[1]Dept. of Statistics, University of Oxford, Oxford OX1 3TG, United Kingdom

[2]Dept. of Statistics, University of Warwick, Coventry, United Kingdom

An important topic of investigation in recent decades has been the scope of randomness in aging. There are three fundamentally distinct entry points in the life course for stochasticity: The beginning, the middle, and the end. At the beginning of life it appears as inherent variability in the quality of organisms. (This is the main topic of Finch and Kirkwood's provocative Finch and Kirkwood (2002).) At the end of life it is the annihilating *coup de grace*, whether an illness or a predator, that may certainly be facilitated by progressive weakness, but which might have come earlier or later, and not infrequently carry off an organism in its prime. And in between there are the "thousand natural shocks that flesh is heir to."

Suppose we maintain the operationally useful identification of aging with increasing risk of death. The stochasticity at the end, of the timing of mortality, is not in doubt. Stochasticity at the beginning and during the life course are difficult to identify. For a given species we face the questions: To what extent would it be possible, in principle, to sharpen our estimate of an individual's ultimate fate, based on measurements and observations of characteristics present in the newborn? To what extent are the shocks of life both cumulative and rare, in the sense that they come in large increments that are few enough to yield identifiably random trajectories?

In this paper we

1. Describe and structure some of the disparate approaches which have been presented, in mathematics and statistics, for examining the stochasticity of the life course, particularly as regards aging and mortality; and
2. Propose how future research could meld these approaches to give clear answers to the substantive questions of how to account for stochasticity in aging.

1.1. The task of mathematical models

An important distinction between the literature we summarize and the directions we outline in section 1.2 and section 4 for current and future research in aging is in the role assigned to the longitudinal stochastic model. Most of the progress in this area in recent years has come from the domain of survival analysis, where the goal is fundamentally to measure the effect of covariate measures on hazard rates. When a stochastic model is included, it is commonly a kind of large nuisance parameter, whose properties are only abstractly defined, primarily to buttress the parameters estimation against the bias that arises when covariate measurements vary over time. A different perspective, and new tools, are needed to apply survival analysis to examine the nature of unobserved senescence processes.

Mathematical models have played an important part in our understanding of biological aging for at least the past 50 years: at least since Strehler and Mildvan (1960). This celebrated but flawed work is an early version of what might be called a *mechanistic deterministic* model. *Mechanistic* because the main object of investigation is the internal mechanism of aging, as opposed to the process by which such mechanisms arise or are maintained. *Deterministic*

because the mechanism explicitly portrays senescence as a loss of function that progresses at the same rate in all individuals, as measured by calendar age.

In this paper we will be focused on mechanistic *stochastic* models of aging: Models which place the variability of aging (and not merely of the consequences of aging, mortality in particular) at their core.

Why do we need such models? Mathematical models serve, first of all, to sharpen our formulation of theoretical descriptions of the aging process. Their characteristics may either be compared to qualitative observations, or inserted as a module into an evolutionary or other teleological model. For such applications the premium is on analytical tractability rather than precise matching to observable or measurable features of the organism.

Second, mathematical models are an important component of statistical modeling. When analyzing aging experiments we rely, whether explicitly or implicitly, on underlying mathematical models that fit the measurements and other data into a comprehensible and testable framework. For this purpose mathematical elegance is less important than an appropriate choice of variables, sufficiently linked to measurable quantities and consequential for the hypotheses under consideration. Mathematical tractability or even comprehensibility recedes in significance in modern computational methods in statistics, but it remains crucial when we aim to interpret the results, to compare and link them to results of other experiments, and to integrate them with broader theoretical understanding.

1.2. Classifying models

In Figure 1 we propose a structure for some of the features typically represented in statistical models of aging. The ovals represent observable factors, while the rectangles denote characteristics of a hidden layer: Factors which are not directly observable, but which have most of the causal power in the model. These may be of theoretical interest, or may simplify the description of the observed factors. All real models limit themselves to representing explicitly only some subset of these factors.

Arrows in the graph represent causal connections. Green ovals represent measurable features of the organism, while blue rectangles are properties that are posited either because they are theoretically posited to exist and be significant, or because they are practically useful components of a statistical model. While we have assigned suggestive names to these factors, the distinctions are meant to be structural. The observable features are:

1. **Mortality**: The time (and perhaps cause) of death. Observable except when censored.
2. **Age**: Time since birth (or whatever counts as birth in the given organism). Not causally affected by anything else.
3. **Signs of aging**: The visible effect of age. By definition, while this is subject to individual variation, these variations are not directly linked to the health effects of aging, nor do they affect the progress of senescence. They may be influenced by senescence. The classic examples are gray hair and wrinkles in humans.

4. **Physiological measures**: A catch-all for all visible effects of aging which do give evidence of health status. However, as *measures* they are only indicators of the progress of senescence and health, but are not direct causes of mortality or deterioration. Examples are grip strength and verbal fluency. Behavioral changes, such as slower walking or change in clothing preference, may be included in this category, but to the extent that these changes do not influence the future course of senescence.

The unobserved features may be modeled in many different ways; here we categorize them by the distinction between features that fixed, fluctuating, and cumulatively random. Their fundamental role is represented by their causal connections to observable features. Of course, in some cases the correlation between an unobserved causal feature and an observed measurement or behavior will be so tight that the distinction is merely formal.

1. **Fixed frailty**: These are genetic, epigenetic, and perhaps other inherent characteristics of an individual that affect their mortality risk throughout their lives.
2. **Individual characteristics**: These include fixed characteristics of an individual that influence the expression of senescence, but do not necessarily determine mortality: Individual optimum blood pressure, family predisposition to gray hair, exposure to testosterone *in utero*, and so on.
3. **Health and Senescence**: The distinction here is somewhat arbitrary. Health is defined as that which, if it does not kill us, leaves us no weaker than before. Acute pneumonia, for instance, or an automobile accident. Of course, many illnesses have both an acute phase and chronic damage. It is simply a matter of definition that we designate those cumulative health consequences as "Senescence". It is for this formal reason that a causal arrow points from senescence to health — aging makes itself felt in increased risks for all manner of nonfatal illnesses — and from health to mortality, but not from health back to senescence.

We note that this graph is restricted to the individual level. A more complete model of senescence would include social and environmental influences, which in some species are likely to be of paramount significance.

The general approach we describe here is generically described as "joint modeling": In each model there is a mathematical description of some subset of the unobserved (blue square) aging, and some portion of the observed (green oval) aging, and a link between them. A complete version would be impossibly complex, so every model that has been substantively analyzed simplifies the causal graph somewhat.

The theoretical demography and biology of aging literature has described a substantial number of mathematical models that represent formalized versions of different biological conceptions of the aging process. Each of these could serve as the infrastructure for a joint model, when linked to a process for generating measurable quantities. The task is to link these in such a way that the hidden model effectively represents and organizes information about observable

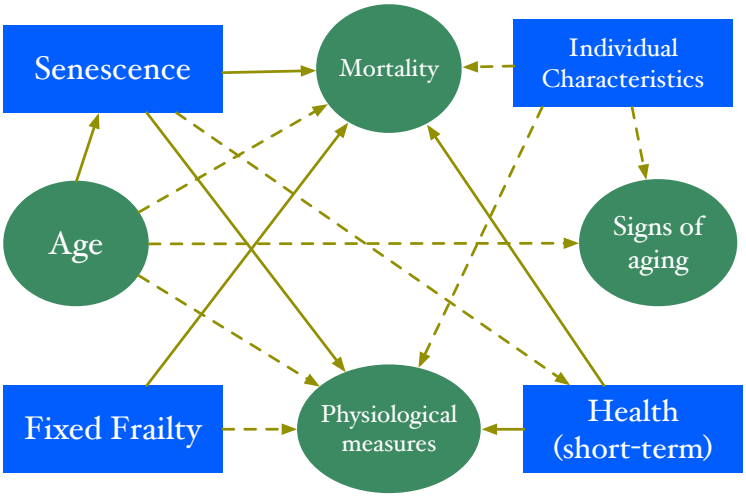

Figure 1: Causal graph for a statistical model of aging. Green ovals represent measurable features of the organism, while blue rectangles represent unobserved components of the aging process which are presumed by the model. solid arrows show causal effects that are essential to the nature of the phenomena being described; if these features are present in a model then it is almost unavoidable that there should be a direct causal link between them. Dashed arrows are more optional causal links.

phenomena, particularly trends in mortality. As we will discuss in section 4.1, according to one way of measuring the statistical utility of a joint model, the potential value of the senescence factor in some classic models is surprisingly small.

2. Markov models: Mathematical theory

Underlying the *joint modeling* of observed and unobserved aging must be a mathematical model of the aging process. While some of the most celebrated aging models, such as Strehler and Mildvan (1960), represent aging itself as deterministic (Strehler and Mildvan limited stochasticity to the "challenges" that convert the deterministic decline in "vitality" into increased probability of death), we focus here on stochastic models, which largely means Markov process models. In such a model there is an internal state — often termed a "senescence" or "vitality" state — that evolves randomly, but with no internal structure or memory of the past trajectory.

We describe here briefly some of the prominent types of mathematical models. A good review article that goes into more depth on a variety of models is Yashin et al. (2000).

2.1. Crude vitality models

There is a significant body of work proposing simple structures for vitality or senescence, without any detailed underlying story. Fixed frailty and individual characteristics typically play no role in these models, no do the sporadic random effects that we have denoted as "health". An influential example of this style is the famous "cascading failures" model of Le Bras (1976): Here senescence is a positive integer state $(1, 2, 3, \dots)$, and an individual in state k moves up to state $k + 1$ at a rate proportional to k, and also has mortality rate increasing proportionately to k. We discuss this example at greater length in section 4.1.

Typically attempts to link these kinds of models to data have focused on mortality rates. Le Bras's original paper purported to show that Gompertz-like mortality would arise, but Gavrilov and Gavrilova (1991) later pointed out that this model ultimately yields mortality plateaus rather than exponentially increasing mortality. Mortality plateaus were also the object of comparison in Weitz and Fraser (2001), and were comprehensively analyzed in Steinsaltz and Evans (2004). Aalen and Gjessing (2001, 2003) have provided perhaps the most general applications of this type of model to generating mortality distributions, primarily for biomedical applications. The consequences of vitality as a purely unobserved process driving mortality have been perhaps most thoroughly developed by T. Li and J. Anderson Li and Anderson (2009).

Yashin et al. (1994) first made explicit the inescapable weakness of attempts to infer from mortality data back to the process of aging: Models with fundamentally different longitudinal structures produce the same mortality rates. If we observe only the binary variable dead or alive, it is a challenge even to infer the mortality rates. Interpreting the underlying process driving the mortality rates is impossible. Yashin and Manton (1997) brought the study of joint models of covariates with mortality into the ambit of aging studies. More recent developments, particularly of Yashin and his collaborators, will be discussed in section 3.1. In recent years the linking of vitality models to individual longitudinal data has become a major focus of statistical work in clinical trials and in gerontology.

2.2. Reliability-type models

The link between the wearing out of mechanical devices and the senescent decline of organisms has been treated as the intuitively obvious foil to more sophisticated theories at least since Weismann (1892) in the 19th century. In the 1970s and 1980s a number of theorists proposed specific mathematical models intended to formalize this intuition. Early pioneers of reliability models in Eastern Europe were Gavrilov and Gavrilova (1991, 2001), Koltover (1982, 1997), and Doubal (1982); their counterparts in the west include Rosen (1978) and Witten (1985). All presented elementary stochastic models that aimed to produce age-specific increasing mortality from non-aging components. To the extent that these models have been linked to data, this was done in a rudimentary way, typically by comparing broad features of predicted and observed mortality curves. More recently some researchers, in particular Pletcher and Neuhauser

(2000) and Laird and Sherratt (2009) have embedded reliability models of aging within the framework of mathematical models of evolution.

More recently, statistically sophisticated versions of engineering reliability models have started to be applied to problems close to the concerns of aging research, under the name "degradation models". These are described in section 3.3.2.

2.3. Damage-accumulation models

One of the most influential modern theories of aging, "disposable soma" Kirkwood (1977), takes a slightly different approach to the accumulation of damage by blending the long tradition of metabolic theories of aging (going back to Pearl (1928)) with Orgel's error catastrophe Orgel (1963). It is an optimization theory, of the sort that suggests that the apparent inefficiency of senescence is an illusion of not appropriately weighing the alternatives, or not considering the tradeoffs that would be required to incorporate perfect repair.

This theory has directly or indirectly inspired a long line of what might be called "metabolic budget" models. Instead of trying to model the components and structures whose performance degrades over time — something that Kirkwood himself essayed in two impressive papers Kowald and Kirkwood (1994, 1996) — damage as such is quantified as it accumulates in cells and organisms, is repaired or not, and is removed or diluted through reproduction. In broadest terms these models link together an organism's growth, reproduction, damage accumulation and damage repair by some kind of budgeting constraint, and then maximizes fitness by dynamical programming or some other method. This is a large literature, but a few of the signal contributions are Abrams and Ludwig (1995); Cichón (1997); Mangel (2001); Chu and Lee (2006); Drenos and Kirkwood (2005). The strength of this work lies in the ability to illustrate the different paths that a fitness-maximizer could take in subverting survival to selectively relevant goals. The weakness is that the tradeoffs which are at the core of the theory are represented purely abstractly, by functional relations that have little justification other than mathematical or computational convenience.

One novel line of research restores the inherent stochasticity of damage accumulation. The mathematical work here, particularly Johnson and Mangel (2006); Evans and Steinsaltz (2007); Watve et al. (2006), has been inspired, at least in part, by a new generation of aging experiments — Webb and Blaser (2002); Stewart and Taddei (2005); Lindner et al. (2008) — in protozoans, as well as by the theoretical and science-historical ideas of Bell (1988). In some sense these models reverse the burden of proof — rather than "explaining" senescence they seek to explain reproduction as a tool for disposing of damage that has accumulated over a lifetime. The link between reproduction and senescence is productively problematized when modeling the senescence of organisms that have no clear individual beginning, hence only at best a partial cleansing of accumulated damage through reproduction.

We will not describe in detail the statistical side of these models, but a modest body of statistical theory and methodology has been built up in very recent time — *cf.* Bansaye et al. (2011); Delmas and Marsalle (2010); de Saporta et al.

(2011) — inspired by the requirements of the experiments. These permit statistical tests of the hypothesis that organisms are dividing their damage unequally, as well as hypotheses that damage allocation is linked to future measured vitality.

2.4. Fixed-frailty models

There is a class of models that simply attempts to capture the variability in individual trajectories through a fixed variability in a basic individual characteristic, such as individual hazard rate. This idea was notably applied to the analysis of mortality plateaus in Vaupel et al. (1979) and Vaupel and Carey (1993). In Yashin et al. (1994) it was pointed out how it is essentially impossible to distinguish fixed-frailty from Markov models simply by observing mortality rates. This observation is an important motivation for joint modeling in aging studies.

Many of the models and techniques used in this work come originally from reliability theory in engineering. A review of mathematical and statistical theory of mixture models, including some applications to survival models, may be found in McLachlan and Peel (2000). The asymptotic behavior of fixed-frailty models, particularly as regards their long-term hazard rates, has been described in Finkelstein and Esaulova (2006) and Steinsaltz and Wachter (2006).

3. Markov models: Statistical applications

Consider a survival experiment. The data from such an experiment may include survival times (possibly censored), some measurements of permanent characteristics for each individual (including birthdate), and some longitudinal measurements of time-varying characteristics of each individual.

Such data provide a natural testbed for stochastic theories of aging. Different theories make different predictions about the relative strength of fixed and time-varying deterministic factors, and fixed and time-varying stochastic factors, in determining the future life course. Much of the published literature is centered on the remaining lifetime as the principle observable, reflecting in part the constraints of earlier experiments and studies. We will retain this perspective in much of the discussion to follow, but the reader should keep in mind that there is no reason, in principle, why survival needs to be at the center of aging studies, once longitudinal data are being collected.

3.1. Joint longitudinal survival

With the increasing number of long-term studies in recent decades has come a flowering of research on the statistical problem that arise when combining longitudinal measures with survival data. We give only a brief summary here, directed at the particular concerns of this paper. For further statistical background we recommend consulting some of the many excellent reviews of the joint longitudinal survival modeling literature to date, including Troxel (2002);

Tsiatis and Davidian (2004); Yu et al. (2004); Ibrahim et al. (2005); Verbeke et al. (2010); Sousa (2011).

Methods for analyzing survival data jointly with longitudinal observations may be sorted into three categories:

1. The simplest approach separates the longitudinal and survival analyses into a 2-stage analysis where in the 1st stage, a longitudinal model, such as a linear mixed effects model, is estimated for the covariates ignoring survival information; at the 2nd stage the fitted values from this model, are plugged into a standard survival model, such as Cox proportional hazards model, as time-dependent covariates Tsiatis et al. (1995); Bycott and Taylor (1998); Dafni and Tsiatis (1998).
2. A more demanding approach bases estimation and inference on a joint likelihood for the longitudinal covariates and survival analysis, within the framework of classical statistical models. Typical joint likelihood constructions are based on a linear mixed effects model for the subject-specific covariate trajectories and a Cox proportional hazards model. In the language of section 1.2, these tend to be based on age, individual characteristics, fixed frailty, and health. That is, time-varying components are deterministic, while random effects are either fixed for all time (sometimes in the form of an individual rate of change) or short-term.
3. A number of recent studies have attempted to model more directly the cumulative random component of senescence, allowing an unobserved senescence process to drive mortality, but also some physiological measures and signs of aging. As we have discussed, there is a wealth of existing stochastic senescence models — generally Markov models — that could be used as the hidden component of such a joint model, but efforts to exploit this stockpile, and explore the statistical utility of these theoretical models have so far been sporadic.

The first approach ignores survival information when modeling the longitudinal process, and consequently underestimates the uncertainty in the estimated values of survival parameters. Numerous authors highlight the resulting inefficiency, and endorse the second approach, including Faucett and Thomas (1996); Wulfsohn and Tsiatis (1997); Yu et al. (2004); Sweeting and Thompson (2011).

The second and third approaches are versions of joint modeling. We describe here some of the essential features of current joint modeling methods. A joint likelihood approach is expected to provide more precise estimates of the relationship between the longitudinal process and time to event, as well as being more efficient in its use of the data, cf. Wulfsohn and Tsiatis (1997); Yu et al. (2004). The third approach is distinguished by its reliance on a stochastic process model, which could allow it to better capture (and measure) the stochasticity in the longitudinal component. The main drawback is that this expands the scope for arbitrary choice to the component of the model that is inherently least subject to verification by direct observation, thus complicating responsible efforts at model selection and verification.

Let $Z_i(t)$ denote the latent true covariate value for a process at time t for the ith subject and $X_i(t)$ denote the corresponding longitudinal measurement available for the process; let $\bar{Z}_i(t) = \{Z_i(u), u \leq t\}$ denote the 'true' history of the process up till time t and $\bar{X}_i(t)$ the corresponding observed longitudinal measurements up till time t; let T_i denote the observed survival time for the ith subject and Y_i any other time-independent covariates for the ith subject.

Tsiatis et al. (1995) propose a joint model based on a linear mixed effects model:

$$X_i(t) = Z_i(t) + \epsilon(t), \tag{1}$$
$$Z_i(t) = \theta_{0i} + \theta_{1i}t, \tag{2}$$

where $\epsilon(t)$ denotes the measurement error with $\mathbb{E}(\epsilon(t)) = 0$, $\mathrm{Var}(\epsilon(t)) = \sigma^2$ and $\mathrm{Cov}(\epsilon(t_1), \epsilon(t_2)) = 0$ for $t_1 \neq t_2$. The random effects $\boldsymbol{\theta}_i = (\theta_{0i}, \theta_{1i})$ given Y_i are bivariate normal with fixed parameters.

The random intercepts and slopes for each subject can then be combined with the hazard of failure within a larger 'metamodel', using the Cox proportional hazards framework, as in Wulfsohn and Tsiatis (1997) and Yu et al. (2004):

$$\text{hazard rate} = \lambda(t \mid \boldsymbol{\theta}_i, \bar{Z}_i(t), Y_i) = \lambda_0(t)\exp[\gamma(\theta_{0i} + \theta_{1i}t) + \beta Z_i], \tag{3}$$

where γ and β are regression coefficients for the time-dependent and time-independent covariates.

Several authors, particularly Wang and Taylor (2001), but also Taylor et al. (1994) and LaValley and DeGruttola (1996), expand (2) by including an integrated Ornstein-Uhlenbeck (IOU) stochastic process. This is a continuous version of an autoregressive process, so the effect is to expand the range of stochasticity to include a process of medium-term memory:

$$X_i(t_{ij}) = Z_i(t_{ij}) + \epsilon_{ij}, \quad \epsilon_{ij} \approx N(0, \sigma_e^2), \tag{4}$$
$$Z_i(t) = a_i + bt + \beta X_i(t) + W_i(t), \tag{5}$$

where t_{ij} refer to the time points j at which values were measured for subject i; $a_i \approx N(\mu_a, \sigma_a^2)$ and $W_i(t)$ denotes the IOU process. Henderson et al. (2000) and Xu and Zeger (2001) consider $W_i(t)$ to be a stationary gaussian process, which enables the trend to vary with time and allows a within-subject autocorrelation structure which can be thought of as biological fluctuations about a smooth trend, as formalized by Tsiatis and Davidian (2004).

The accelerated failure-time (AFT) model — *cf.* Cox and Oakes (1984) — offers a viable alternative to the Cox proportional hazards model when the proportionality assumption breaks down. Wei (1992) and Cox (1997) in particular highlight how modelling covariates directly on the survival time can provide a more intuitive understanding of the relationship between a longitudinal process and the survival time. An AFT model formalizes the notion of each individual having his or her own senescence clock, with mortality, physiologic changes, and

signs of aging all being driven by this hidden internal clock, hence only indirectly by calendar time. In Hsieh et al. (2005) a joint longitudinal-AFT model is proposed, which models the longitudinal process by a linear mixed effects measurement error model, as in the examples described above. Swindell (2009) has also recommended the AFT as a useful statistical framework for aging research. The statistical techniques for accelerated failure are not yet as well developed or as easy to use as for proportional hazards.

3.2. Physiological measures with error

In many of the earliest versions of joint modeling of survival with longitudinal covariates the cumulative factor that drives mortality rates — is not an abstract "senescence" quality, but a concrete physiological feature. It is not observable because of the limitations of measurement. An important early example is CD4 count in HIV studies Wulfsohn and Tsiatis (1997), where the crucial factor is subject to measurement error and short-term fluctuations ("short-term health" in the language of section 1.2).

Longitudinal data such as the physiological measurements can be fitted in a Cox proportional hazards model as time-dependent covariates Cox and Oakes (1984) but as Sweeting and Thompson (2011) note, this is not recommended. Typically, the longitudinal measurements are incomplete and prone to measurement error and simply using the raw measurements as recorded, can lead to biased estimation of model parameters, as discussed by Prentice (1982); Hughes (1993); Raboud et al. (1993); Hu et al. (1998).

Lange et al. (1992) and Hoover et al. (1992) draw further attention to the high variability within subjects when making biological and physiological measurements. In response, longitudinal models are typically used when fitting covariates measured over time, with error. Carroll (2006) provides a comprehensive review of the models currently available.

3.3. Applications to aging

In recent times, joint longitudinal survival models have been extended to include and build upon on the earlier aging models, some of which we have outlined in section 2. In principle, almost any mathematical model may serve as the engine of a joint model. It is important to be aware of the theoretical implications and constraints imposed by the choice of longitudinal model, and to consider these as seriously as questions of statistical tractability.

3.3.1. Modeling signs of aging

A number of existing statistical methodologies have been applied recently to the problem of inferring underlying senescence from longitudinal observations of signs of aging — that is, in the terminology of section 1.2, measurements of properties that do not themselves participate in the causal processes of senescence, but are presumed to be influenced by senescence.

Pavlov (2010) builds upon the joint modeling ideas of Wang and Taylor (2001),Taylor et al. (1994), and LaValley and DeGruttola (1996) by applying a

Kalman filtering approach of the sort that was introduced by Fahrmeir (1994) for discrete-time survival data. Using the hidden Markov model setting of a state space model he explores the dynamics of an unobserved aging process based on damage accumulation and other stochastic aging theories. An attempt to capture heterogeneity is made through the introduction of a frailty component within the joint longitudinal survival likelihood. Using fruit-fly lifetime behavioral data (from the studies described in Zou et al. (2011)), Pavlov (2010) postulates the hidden aging process to follow Brownian motion, an IOU, or other basic stochastic processes. This allows him to extract an aging signal of sorts from observed fly eating habits over age. Pavlov's approach highlights the flexibility of a state-space model — based on an unobserved process driving the system — to explore and model the aging process. The unobserved process can be considered as the latent aging process and models can be tailored to extract aging information from a wide variety of data. The direct conclusions of this work were limited by the very small number of individual flies included in the study.

Yashin et al. (2007), on the other hand, propose stochastic differential equations for capturing physiological changes in a joint longitudinal survival setting. They in particular advocate the use of stochastic differential equations for capturing allostasis and the decline in adaptive capacity associated with aging. But as Yashin et al. (2011) later remark, this fails to describe changes in 'health status accompanying physiological aging', unlike Pavlov (2010). Yashin et al. (2011) uses the notion of a health history — modeled using a finite state jumping stochastic process — and attempts to jointly analyze this with physiological measurements and survival, as in Yashin et al. (2007).

A different approach was introduced by Müller and Zhang (2005). Using functional principal component analysis (Rice and Silverman, 1991) and the concept of functional regression (Cardot et al., 2003) from functional data analysis (*cf.* Ramsay and Silverman (2005)) they construct a time-varying functional regression framework that predicts the remaining lifetime and lifetime distributions from longitudinal trajectories for individual subjects. They remark, as we have emphasized here, that in studies on aging the predicted remaining lifetime can provide a useful measure for senescence.

3.3.2. Degradation models

Aging is often defined as the accumulation of damage or wear and tear over time leading to death. Degradation models thus provide a useful alternative framework for analysing survival data by linking failure times with stochastic time-varying covariates. Singpurwalla (1995) provides key early expository work outlining useful stochastic processes for modeling degradation and stochastic 'wear and tear', which can be related to the biological 'damage accumulation' aging theory.

From the perspective of aging research, it is notable that Whitmore et al. (1998) introduced a failure time model based on a bivariate Wiener process, whereby one process represents an observed marker process, and the second represents a latent process. The failure time is defined by the latent process

crossing a threshold. Lee et al. (2000) extended this to model a joint marker process and latent health status; as with the hidden diffusion "senescence state" analyzed by Pavlov (see above), the notion of a latent health status is appealing, since it represents the 'unobservable' aging process.

Degradation models are a distinct scientific lineage from joint survival models, and there is limited cross-citation between them, but they are formally almost identical. The two approaches have tended to emphasize different concerns, reflecting their divergent origins: Degradation models grew out of reliability modeling in engineering, while joint survival modeling are one of the many adaptations of classical survival analysis in medicine to the problems of longitudinally measured covariates.

Building on the work of Whitmore et al. (1998) and Lee et al. (2000), Lee and Whitmore (2006) introduced 'threshold regression' for survival data by postulating the deterioration of a subject's health as following a stochastic process and thereby utilising first-hitting time model ideas as previously — see Aalen and Gjessing (2001) for a comprehensive review of first-hitting time models. Following the notation in Lee and Whitmore (2006), let $\{X(t), t \geq 0\}$ denote a Wiener process with mean μ and variance σ^2, with initial value $X(0) = x_0 > 0$. With $\mu < 0$ the process will tend to drift towards 0 and hence the first-hitting time for the zero level follows an inverse Gaussian distribution. Lee et al. (2010) extended these ideas to introduce a threshold survival regression model with time-varying covariates in a health status setting, and further advances by Li and Lee (2011) allow for more flexible representation of the varying coefficients.

Some more recent work on degradation models such as Bagdonavičius and Nikulin (2000) and Bagdonavičius and Nikulin (2004) incorporate longitudinal measures, producing a framework that differs only in nomenclature from the survival models. However, instead of the Gaussian-process models that dominate the joint-modeling literature, this work reflects its engineering influences by building upon a gamma process, which has substantially different properties. Most importantly, it only increases, so it represents a model of aging with "wear and tear" but no repair.

3.3.3. Discrete state-spaces and deficit models

Discrete-state models like the Le Bras mortality model offer some important simplifications relative to continuous-state models. They are particularly appealing for settings where the observations fall naturally into discrete ordered classes. For example, Mesbah et al. (2004) use a simple four-state model (good, neutral, bad, dead) to analyze joint observations of longitudinal quality of life reports and survival. This work is very much in the spirit that we are advocating here, since the goal is not merely to use the longitudinal observations to predict mortality, but also to understand the quality-of-life process itself.

A substantial body of work by A. Mitnitski and K. Rockwood, together with various collaborators — in particular Mitnitski et al. (2002) and Rockwood and Mitnitski (2007) — has promoted the use of simple counts of discretely observable physical deficits as a proxy for ageing. This stands currently as one of the most thorough attempts to date to measure the individual pace of biological

aging in humans, and to validate the measurement procedures through a range of statistical, epidemiological, and mathematical modeling methods. Interestingly, their preferred mathematical model (Mitnitski et al. (2006)) represents the frailty index itself as Markov, with no hidden layer. It is not clear from the published record what efforts were made to validate this feature of the model, or if any alternatives were considered.

4. Future work: Integrating the mathematical and statistical perspectives

We have described developments in theoretically modeling aging systems in terms of the progressive changes in measurable senescence states, and in linking measured longitudinal changes statistically to survival. There are still important gaps in our understanding, which need to be addressed before this work can effectively shape our view of biological aging. We discuss here two of these gaps, which fall in the space linking mathematical and statistical approaches: Evaluating the effectiveness of a senescence model, and parcelling out stochasticity automatically among different timescales.

4.1. Evaluating the potential significance of Markov models

Suppose we are faced with a choice between various models of senescence, one of which instantiates an individual inherent robustness, another which adds a random slope of vitality loss, a third with "health" variation, represented as stochastic short-term fluctuations in vitality. Considered as a statistical problem, given a collection of covariate measures (linked to the vitality) and outcomes we have a variety of tools for deciding which model best connects the covariates to the outcomes.

In the biostatistical practice of survival analysis it is common to focus on predictive accuracy: We score a model based on the average difference between the best prediction we can make of the lifetime from the model and the true lifetime. If the average is judged in the mean-square sense, we obtain the *Brier score* or *conditional mean square error of prediction* in Lawless and Yuan (2010). In the simplest setting, where we know the true model (so we are not using the observations to estimate the model) the Brier score is essentially the expected residual variance when the survival is conditioned on the covariates.

For purposes of selecting relevant factors for a theoretical model of aging it is reasonable to foreground not the potential prediction accuracy, but rather the opposite: How much does the distribution shift when we incorporate a new measurement into the model. For the remainder of this section we will continue to assume that the model is known.

4.1.1. Stable covariates

Suppose for the moment we are considering inherent individual properties, measuring stable covariates which are fixed throughout life, represented by an underlying characteristic Z or a measurement X. If π is the distribution of

T, and π_z is the conditional distribution, given that the covariates Z have value z, then the value of the information $Z = z$ is just $\delta(\pi, \pi_z)$ — the change in our evaluation of the distribution of T that comes from incorporating the information $\{Z = z\}$ — and the average value of Z is

$$\mathbb{E}_Z \left[\delta(\pi, \pi_Z)\right], \tag{6}$$

where the expectation $\mathbb{E}_Z$ gives us the mean distance between the unconditional distribution and the conditional distribution. For instance, if π and π' have means μ_π and $\mu_{\pi'}$, and the quadratic distance $\delta(\pi, \pi') = (\mu_\pi - \mu_{\pi'})^2$, then the distance in (6) turns out to be precisely the difference between the variance of π and the average residual variance conditioned on Z.

The same arguments apply with comparable force to other measures of efficacy of a covariate model. One appealing possibility for the distance is Kullback-Leibler divergence (not, formally speaking, a true distance). In this case, the value of Z for predicting T becomes the same simply the mutual information between T and Z:

$$I(T; Z) = \mathbb{E}\left[\int_0^\infty f(t) \log \frac{f(t)}{f_Z(t)} dt\right].$$

To take a simple example, suppose Z is the sex of an individual, with a fraction p of the population being male and $1 - p$ female, and that the survival times for male and female have different distributions with densities f_F and f_M, with means μ_M and μ_F respectively. The overall mean lifespan is then $\mu = p\mu_M + (1 - p)\mu_F$. Using the quadratic distance defined above, the "value" of knowing that an individual is female for predicting lifespan is $(p\mu_M - p\mu_F)^2$, and the value of knowing that an individual is male is $((1-p)\mu_M - (1-p)\mu_F)^2$. The *average* value of knowing an individual's sex weights these two values by the appropriate proportion in the population, yielding $p(1-p)(\mu_M - \mu_F)^2$. This has a natural interpretation: Sex is informative for lifespan when there is a large difference between male and female life expectancy, but also when there are equal numbers of male and female. In the extreme case, where the population is nearly all of one sex, observing an individual's sex tells us on average very little about survival time.

If we do not observe the individual's sex directly, but only a covariate X that is correlated with sex (and otherwise independent of survival time — for example, the individual's name or hair length) then the factor $p(1-p)$ is replaced by a smaller factor $p(1 - p) - \mathbb{E}[p(X)(1 - p(X))]$, where $p(x)$ is the proportion of males among those with $\{X = x\}$.

If we measure value by mutual information, we see that

$$I(T; Z) = \int_0^\infty f(t) \log f(t) dt - p \int_0^\infty f(t) \log f_M(t) dt - (1-p) \int_0^\infty f(t) \log f_F(t) dt.$$

(The first term on the right-hand side is the *entropy* of the survival time.) This may be understood as the reduction in the quantity of information (measured in bits, if the logarithm is base 2) required to specify the survival time, when

the sex is known as well. As before, if X is a covariate that is correlated with sex, but otherwise independent of survival time, we have $I(T;X) \leq I(T;Z)$, and

$$I(T;Z) - I(T;X) = I(X,T;Z) - I(X;Z).$$

That is, the extra information that knowing Z provides about T over and above what X provides is the same as the increased certainty about the sex acquired by knowing T and X, compared with just knowing X.

4.1.2. Longitudinally changing covariates

One key insight developed in particular by Schoop et al. (2008) and Commenges et al. (2012) is that dynamic models of survival times require measures of prognostic loss based on the conditional distribution of remaining time at the time of observation. Along these lines, in his analysis of longitudinal aging-related observations of fruit flies, Pavlov (2010) proposed evaluating joint models based on differences in the expected remaining lifetime at time t, conditioned on Z_s for all $s \leq t$.

For the rest of this section $(Z_t)_{t\geq 0}$ will be a stochastic process describing the state of an individual, and X_t is any function of $\{Z_s : 0 \leq s \leq t\}$ (that is, X_t is any score that could be based on information that could be observed in the process up to time t). We write T_t for a random variable whose distribution is that of t conditioned on survival to age t.

In the context of longitudinal measurements the single prediction problem for the distribution of T becomes a sequence of prediction problems for the distribution of remaining lifetime $T_t - t$ at each age t. We may choose either to consider a sequence of expected distances — listing (or plotting) for a range of values of t the strength of a certain measurement in shifting the conditional distribution of remaining lifetime; for instance, we may wish to compute the mutual information $I(T_t - t; X_t)$ for each t, conditioned on survival at least to age t. Alternatively, we may choose to integrate the different measurements into a single average score. Effectively what this means is that we choose an individual at random from the population, and ask for the average effect (as scored by Kullback-Leibler, or some other measure) on the conditional remaining lifetime distribution induced by knowing the value of that individual's current X_t.

One advantage of population averaging, if it is done purely on the basis of age, is that it offers a convenient way of calibrating the information. The random age A of a randomly sampled individual serves as an alternative observable that provides a baseline minimum power for predicting future lifespan. It is natural then to compare the average efficacy of a model variable X_A, computed on the basis of observations at age A, with the average efficacy of observing A alone. That is, we consider the information that X_A (together with A) provides about the remaining lifetime $T_A - A$, for a random individual sampled from the population whose (random) current age is observed to be A, compared with the information provided simply by knowing the age is A.

We are free to choose any sampling distribution we like for this averaging. One appealing choice is the stationary age distribution of the population under conditions of zero growth. This is the distribution whose density at age t is proportional to the probability of survival to age t. This is equivalent to looking at a single birth cohort and taking a uniform random sample of all the moments of life lived by that cohort, and then predicting each individual's remaining lifespan at that moment. We illustrate the calculation of mutual information averaged over the stationary age distribution in section 4.1.4 and in the appendix.

Commenges et al. (2012) offer a framework for applying information-theoretic measures of prognostic loss in survival models with longitudinal covariates. In the context of longitudinal covariates, the mutual information has some useful properties. For example, if the process $Z.$ driving senescence is a Markov process, then by the Data Processing Inequality (Cover and Thomas, 2006, section 2.8)

$$I(T_t - t; X_t) \leq I(T_t - t; Z_t).$$

That is, the most information about $T - t$ that could possibly be available by time t would be if it allowed a perfect determination of Z_t. To put it somewhat differently, to adopt a Markov model is to make a claim that there is a theoretical upper limit to the predictability of future survival, or indeed of any of the future trajectory of an individual. In particular, if the Markov chain is time-homogeneous — that is, if time and/or age have no influence on the process *independent* of the internal state — then knowing the age and current state of an individual provides no further information about the future beyond simply knowing the current state.

Other measures of the prognostic efficacy of hidden stochastic processes and longitudinal observations have their own strengths and weaknesses. Several of these have been developed and usefully compared in Schoop et al. (2011), both theoretically and by simulations. These include loss-function approaches like expected Brier score and conditional mean absolute deviation, and graphical approaches like time-dependent ROC curves and predictiveness curves. Their framework is more general than ours here, as it includes errors from parameter misestimation and model misspecification.

4.1.3. Graphical tools

A common tool used by epidemiologists in evaluating screening tests is the ROC curve, a graphical representation of the tradeoff between true positive and false positive results — see Fawcett (2006) for background. Suppose we are designing a test for predicting whether an individual will have a certain property — for instance, dead within 5 years. We have some candidate measure of risk X, and we want to predict "yes" for anyone whose X score is above a threshold s to be determined. As we move the screening threshold gradually down the number of positive results increases, catching an increasing fraction of the true positive cases, but also of the false positives simultaneously. The ROC curve plots the fraction of true positives captured at threshold s against the fraction of false positives. It starts at $(0, 0)$ (when s is so high that no one passes the test)

and moves up to $(1,1)$. A completely useless measurement will randomly sample true and false positives equally, and so will produce the upward sloping 45° line. A perfect predictor will go straight up from $(0,0)$ to $(0,1)$ (catching 100% of the true positives with no false positives mixed in), and then horizontally to $(1,1)$. In general we expect to find ROC curves somewhere in between, with superior predictors yielding greater separation from the diagonal.

If we think of an observable or non-observable feature of the model as predicting an individual's death within some fixed T units of time from the present, we have a family of ROC curves, depending on T and on the distribution used to sample the individuals from the population. Let T become very small we have what might be called the ROC curve for predicting imminent death, giving a snapshot of risk in the population as stratified by the predictor X. This is equivalent to plotting the distribution of hazard rate within the population. In other words, suppose we take a sample from the population — for example, all individuals of a given age, or (as we will do in our example in section 4.1.4) sampled from the stationary age distribution. Let $\mathcal{H}_x$ be the total current hazard rate of all individuals in the sample with $X \geq x$, and let F_x be the fraction of the sample with $X \geq x$. The ROC for imminent mortality plots the pairs $(F_x, \mathcal{H}_x)$ for all x; when X takes on only discrete values, the points are joined up linearly.

A slightly different formulation is the Lorenz curve, more commonly used in economics, which plots the fraction of true positives against the fraction of the total population. Of course, if the true positives are only a small fraction of the whole population, Lorenz curves and ROC curves are essentially the same. In the case of predicting imminent mortality the proportion of true positives in the population goes to 0, so the ROC curve described above is actually identical to the Lorenz curve.

4.1.4. Example: The Le Bras cascading failures model

Predictive measures have been developed and extensively applied in survival analysis, although the dynamic versions that are useful for longitudinal covariates are fairly new. What we are advocating here is applying these tools for interpreting the kinds of stochastic models that underly joint models of survival and longitudinal measures, and that also play a role in theoretical discussions of the biology and evolution of aging. When we propose a stochastic model of aging, in which mortality rates are driven by individual random accumulation of senescence (or random loss of vitality), it seems reasonable that we ought to ask how much this model actually differs from the naïve deterministic model, in which mortality is determined solely by age. How significant is the variability in remaining lifespan coded in the population's hidden stratification by senescence state?

For purposes of illustration we show how these calculations come out for the Le Bras cascading failures model, already mentioned in section 2.1. (For details, see the appendix.) In Figure 2 we show the mutual information of X_s and $T_s - s$ for different fixed values of s, for three different choices of the parameters (λ, μ): $(0.05, 0.1)$, $(0.1, 0.1)$, $(0.1, 0.05)$. For example, consider the model with

parameters $\lambda = \mu = 0.1$, and suppose we look at a random individual at age 4, with an eye toward predicting their lifespan. Being told the individual's value of X_4 reduces our uncertainty about the remaining lifespan by 0.06. On this scale the full uncertainty of T_4 — the entropy — is about 3.1. The entropy changes very little over age, and is very similar for for the other parameter choices under consideration here. Thus, knowing X_4 removes only a small fraction of the total uncertainty about remaining lifespan at age 4.

The amount of information obtained by observing the state increases as age advances and the states become more variable. The maximum amount of information is larger when λ is bigger, relative to μ. This makes sense, because larger λ implies a larger dispersion of states for a given age, meaning that knowing the state is relatively more informative. Nonetheless, the mutual information is never more than about 6% of the total entropy.

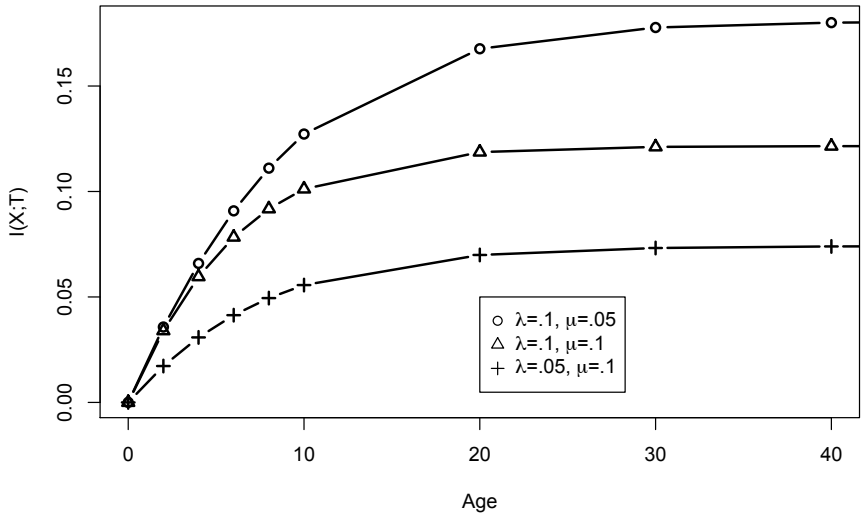

Figure 2: Information value of observing the underlying senescence state for predicting remaining lifespan, as a function of current age.

We note that the information $I(X_s; T - s)$ approaches its maximum value at ages s when nearly the entire population has already died. If we consider instead the mutual information between state and remaining lifespan for a randomly chosen individual in a random time of life, and compare this to the mutual information between age and remaining lifespan, we get the results in Table 1. We see that a very large ratio of λ/μ is needed — meaning a large change in senescence states over an ordinary lifespan — to give the unobserved state a significant information advantage over simple age. (Calculations may be found

in the appendix.)

Table 1: Mutual information of remaining lifespan with age (third column) and current senescence state (fourth column) for an individual chosen at random from the stable population in the Le Bras model of senescence, for several different choices of the parameters.

λ	μ	$I(A; T_A - A)$	$I(X_A; T_A - A)$
.01	.1	0.193	0.199
.05	.1	0.194	0.226
.1	.1	0.196	0.256
.1	.05	0.199	0.309
.1	.01	0.224	0.546

The basic lesson is that if a model like this did reflect the underlying truth of senescence, no possible biomarker could yield more information about future lifespan than observation of the hidden state X_t. Despite the fact that all survival is driven by X_t, for significant ranges of parameters observing X_t is (on average) of little use in predicting future survival.

As described in section 4.1.3, this principle may be illustrated graphically by plotting ROC curves for the population hazard rate, shown in Figure 3. Here we plot the fraction of total population hazard that can be accumulated for a given fraction of the population, using individual age (solid curves) or senescence state (dashed curves) as the basis of the stratification. We see that for the case $\lambda = \mu = 0.1$ current age is not very useful as a predictor of mortality risk; the prediction is improved somewhat, but not enormously, by observing the true senescence state. Age is a somewhat better predictor when λ is reduced to 0.05, and the information value of senescence state is also significantly increased.

4.2. Stochasticity on different timescales

We suggested in section 1.2 that stochastic models could best be understood as partitioning the randomness in aging and mortality across different timescales: short-term health events, long-term senescence, inherent robustness. As we have already discussed, each mathematical model implies a choice of one or more unobservable and observable properties, allowing us to explore the interaction and effects. What is missing is a set of tools that would enable us to assimilate diverse data into a generic model that apportions randomness across different timescales and measures their influence on survival.

In principle we could hope to encompass the diverse components in Figure 1 within a hidden diffusion model, taking a proportional hazards or other link between the hidden state and the hazard rates. The Wang-Taylor model described at the end of section 3.1 provides a useful starting point. This decomposes the vitality status of an individual into a common deterministic trend and individual intercept, plus a white-noise process (short-term random "health"

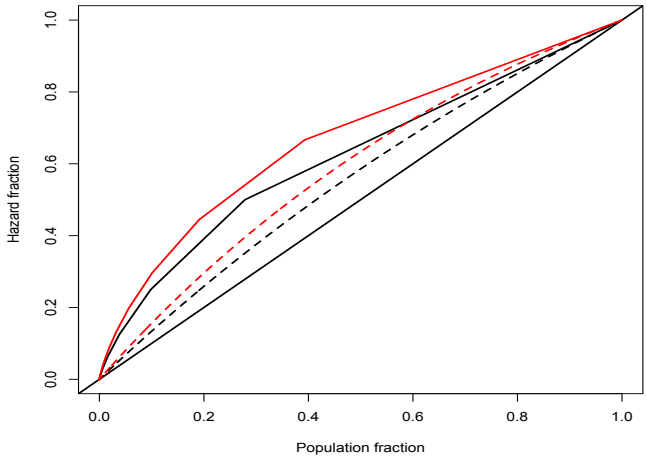

Figure 3: Receiver operator curves for current hazard rate in the Le Bras model for parameter values $(\lambda, \mu) = (0.1, 0.1)$ (black curves, lower down) or $(0.1, 0.05)$ (red curves,higher up). The solid curve represents prediction on the basis of observing the senescence state; the dashed curve represents prediction based on age. The 45° upward-sloping line, corresponds to a completely informationless predictor.

events, in the nomenclature of section 1.2), plus an integrated stochastic process ("senescence"). Alternatively, if we choose to model cumulative senescence as strictly increasing we could use the gamma process favored by Bagdonavičius and Nikulin (2000), and further developed in their more recent work on degradation models (*cf.* section 3.3.2). As far as we are aware it is an open question to what extent the distinction between a hidden process that is strictly increasing and one that rises and falls is an identifiable property, and to the extent that it is, how it might best be inferred from data.

There is no reason, in principle, why the model cannot be extended to include, in addition to the zero-memory and infinite-memory components, stochastic processes that revert to the mean on intermediate timescales. Significant progress has been made in recent years in the estimation of diffusions from low-frequency observations — see in particular the early review of estimating function techniques Sørensen (1997) and the introduction of spectral techniques Gobet et al. (2004). More recent work has been undertaken largely with a view to finance and econometric applications, and leave untouched many of the natural questions for applications to biological aging (although the work on modeling default risk in finance, such as Nakagawa (2001) and Duffie et al. (2009), is formally very close to the context of aging models).

An important goal would be to analyze separate time-scale components along the lines described in section 4.1, to evaluate the leverage that these components

of randomness can have over the observable components, above all the survival.

5. Conclusion

There has been a profusion of mathematical and statistical models in recent years that are intended to describe the senescence process and link it to measurable features of organisms. What we have described here is only a corner of this burgeoning literature. We have suggested that more attention will be required to linking the biological, mathematical, and statistical approaches to aging, before experiments will be able to provide clear answers to biological questions about aging. A more systematic exploitation of theoretical models of aging for application to longitudinal individual observations has become imperative.

We have proposed adapting some of the standard information theoretic and graphical tools used in model selection may be useful for defining the effectiveness and potential applicability of certain mathematical models of aging, and for putting bounds on their potential contribution to the analysis of experimental data.

References

Aalen, O. O., Gjessing, H. K., 2001. Understanding the shape of the hazard rate: A process point of view. Statistical Science 16 (1), 1–22.

Aalen, O. O., Gjessing, H. K., 2003. A look behind survival data: Underlying processes and quasi-stationarity. In: Lindqvist, B. H., Doksum, K. A. (Eds.), Mathematical and Statistical Methods in Reliability. World Scientific Publishing, Singapore, pp. 221–34.

Abrams, P. A., Ludwig, D., December 1995. Optimality theory, Gompertz' law, and the disposable soma theory of senescence. Evolution 49 (6), 1055–66.

Arking, R., 2006. The Biology of Aging: Observations and Principles, 3rd Edition. Oxford University Press.

Bagdonavičius, V., Nikulin, M. S., 2000. Estimation in degradation models with explanatory variables. Lifetime Data Analysis 7, 85–103.

Bagdonavičius, V., Nikulin, M. S., 2004. Semiparametric analysis of degradation and failure time data with covariates. In: Parametric and Semipara- metric Models with Applications to Reliability, Survival Analysis, and Quality of Life. Birkhäuser, Ch. 1, pp. 1–22.

Bansaye, V., Delmas, J.-F., Marsalle, L., Tran, V. C., 2011. Limit theorems for markov processes indexed by continuous time Galton-Watson trees. Annals of Applied Probability, 2263–2314.

Bell, G., 1988. Sex and death in protozoa : The history of an obsession. Cambridge University Press, Cambridge.

Bycott, P., Taylor, J., 1998. A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. Statistics in medicine 17 (18), 2061–2077.

Cardot, H., Ferraty, F., Mas, A., Sarda, P., 2003. Testing hypotheses in the functional linear model. Scandinavian Journal of Statistics 30 (1), 241–255.

Carroll, R., 2006. Measurement error in nonlinear models: a modern perspective. Vol. 105. London: CRC Press.

Chu, C. Y., Lee, R. D., March 2006. The co-evolution of intergenerational transfers and longevity: an optimal life history approach. Theoretical Population Biology 69 (2), 193–2001.

Cichón, M., 1997. Evolution of longevity through optimal resource allocation. Proceedings of the Royal Society of London, Series B: Biological Sciences 264, 1383–8.

Commenges, D., Liquet, B., Proust-Lima, C., 2012. Choice of prognostic estimators in joint models by estimating differences of expected conditional Kullback–Leibler risks. Biometrics, no–no.
URL http://dx.doi.org/10.1111/j.1541-0420.2012.01753.x

Cover, T. M., Thomas, J. A., 2006. Elements of Information Theory. John Wiley & Sons.

Cox, D. R., 1997. Some remarks on the analysis of survival data. Lecture Notes in Statistics, 1–10.

Cox, D. R., Oakes, D., 1984. Analysis of survival data. London: Chapman & Hall.

Dafni, U., Tsiatis, A., 1998. Evaluating surrogate markers of clinical outcome when measured with error. Biometrics, 1445–1462.

de Saporta, B., Gégout-Petit, A., Marsalle, L., December 16 2011. Asymmetry tests for bifurcating auto-regressive processes with missing data, preprint available arxiv.org/abs/1112.3745.

Delmas, J.-F., Marsalle, L., 2010. Detection of cellular aging in a Galton-Watson process. Stochastic Processes and their Applications 120 (12), 2495–2519.

Doubal, S., 1982. Theory of reliability, biological systems and aging. Mechanisms of Ageing and Development 18, 339–53.

Drenos, F., Kirkwood, T. B. L., 2005. Modelling the disposable soma theory of ageing. Mechanisms of Ageing and Development 126, 99–103.

Duffie, D., Eckner, A., Horel, G., Saita, L., October 2009. Frailty correlated default. Journal of Finance 64 (5), 2089–2123.

Evans, S. N., Steinsaltz, D., 2007. Damage segregation at fissioning may increase growth rates: A superprocess model. Theoretical Population Biology 71 (4), 473–90.

Fahrmeir, L., 1994. Dynamic modelling and penalized likelihood estimation for discrete time survival data. Biometrika 81 (2), 317–330.

Faucett, C. L., Thomas, D. C., 1996. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Statistics in Medicine 15 (15), 1663–1685.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognition Letters 27, 861–74.

Finch, C. E., Kirkwood, T. B. L., 2002. Chance, Development, and Aging. Oxford University Press.

Finkelstein, M., Esaulova, V., 2006. Asymptotic behavior of a general class of mixture failure rates. Advances in Applied Probability 38 (1), 244–62.

Gavrilov, L. A., Gavrilova, N. S., 1991. The Biology of Lifespan: A Quantitative Approach. Harwood Academic Publishers, Chur, Switzerland.

Gavrilov, L. A., Gavrilova, N. S., 2001. The reliability theory of aging and longevity. Journal of Theoretical Biology 213, 527–45.

Gobet, E., Hoffmann, M., Reiß, M., 2004. Nonparametric estimation of scalar diffusions based on low frequency data. Annals of Statistics 32 (5), 2223–53.

Henderson, R., Diggle, P., Dobson, A., 2000. Joint modelling of longitudinal measurements and event time data. Biostatistics 1 (4), 465.

Hoover, D., Graham, N., Chen, B., Taylor, J., Phair, J., Zhou, S., Muñoz, A., et al., 1992. Effect of CD4+ cell count measurement variability on staging HIV-1 infection. Journal of acquired immune deficiency syndromes 5 (8), 794.

Hsieh, F., Tseng, Y.-K., Wang, J.-L., 2005. Joint modelling of accelerated failure time and longitudinal data. Biometrika 92 (3), 587–603.

Hu, P., Tsiatis, A. A., Davidian, M., 1998. Estimating the parameters in the cox model when covariate variables are measured with error. Biometrics, 1407–1419.

Hughes, M. D., 1993. Regression dilution in the proportional hazards model. Biometrics, 1056–1066.

Ibrahim, J., Chen, M., Sinha, D., 2005. Bayesian survival analysis. New York: Springer.

Johnson, L. R., Mangel, M., 2006. Life histories and the evolution of aging in bacteria and other single-celled organisms. Mechanisms of Ageing and Development 127 (10), 786–93.

Kirkwood, T., November 24 1977. Evolution of ageing. Nature 270, 301–304.

Koltover, V. K., 1982. Reliability of enzyme systems and molecular mechanisms of ageing. Biophysics 27, 635–9.

Koltover, V. K., 1997. Reliability concept as a trend in biophysics of aging. J. Theor. Bio. 184, 157–63.

Kowald, A., Kirkwood, T. B., 1994. Towards a network theory of aging: A model combining the free radical theory and the protein error theory. Journal of Theoretical Biology.

Kowald, A., Kirkwood, T. B., 1996. A network theory of aging: The interactions of defective mitochondria, aberrant proteins, free radicals and scavengers in the ageing process. Mutation research.

Laird, R. A., Sherratt, T. N., May 2009. The evolution of senescence through decelerating selection for system reliability. Journal of Evolutionary Biology 22 (5), 974–82.

Lange, N., Carlin, B. P., Gelfand, A. E., 1992. Hierarchical bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers. Journal of the American Statistical Association, 615–626.

LaValley, M., DeGruttola, V., 1996. Models for empirical bayes estimators of longitudinal cd4 counts. Statistics in Medicine 15 (21), 2289–2305.

Lawless, J. F., Yuan, Y., 2010. Estimation of prediction error for survival models. Statistics in Medicine 29, 262–74.

Le Bras, H., May-June 1976. Lois de mortalité et age limite. Population 33 (3), 655–91.

Lee, M. L. T., DeGruttola, V., Schoenfeld, D., 2000. A model for markers and latent health status. Journal of the Royal Statistical Society. Series B, Statistical Methodology, 747–762.

Lee, M. L. T., Whitmore, G. A., 2006. Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary. Statistical Science, 501–513.

Lee, M. L. T., Whitmore, G. A., Rosner, B., 2010. Threshold regression for survival data with time-varying covariates. Statistics in medicine 29 (7-8), 896–905.

Li, J., Lee, M., 2011. Analysis of failure time using threshold regression with semi-parametric varying coefficients. Statistica Neerlandica.

Li, T., Anderson, J. J., 2009. The vitality model: A way to understand population survival and demographic heterogeneity. Theoretical Population Biology 76 (2), 118–31.

Lindner, A. B., Madden, R., Demarez, A., Stewart, E. J., Taddei, F., February 26 2008. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. PNAS 105 (8), 3076–81.

Mangel, M., 2001. Complex adaptive systems, aging and longevity. J. Theor. Bio. 213, 559–71.

McLachlan, G., Peel, D., 2000. Finite Mixture Models. John Wiley & Sons Inc.

Mesbah, M., Dupuy, J.-F., Heutte, N., Awad, L., 2004. Joint analysis of longitudinal quality of life and survival processes. In: Balakrishnan, N., Rao, C. R. (Eds.), Advances in Survival Analysis. Vol. 23 of Handbook of Statistics. Elsevier, Ch. 38, pp. 689–728.

Mitnitski, A. B., Bao, L., Rockwood, K., 2006. Going from bad to worse: A stochastic model of transitions in deficit accumulation, in relation to mortality. Mechanisms of Ageing and Development 127 (5), 490–3.

Mitnitski, A. B., Graham, J. E., Mogilner, A. J., Rockwood, K., 2002. Frailty, fitness and late-life mortality in relation to chronological and biological age. BMC Geriatrics 2 (1).

Müller, H., Zhang, Y., 2005. Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. Biometrics 61 (4), 1064–1075.

Nakagawa, H., 2001. A filtering model on default risk. Journal of Mathematical Sciences, The University of Tokyo 8, 107–42.

Orgel, L. E., 1963. The maintenance of the accuracy of protein synthesis and its relevance to aging. PNAS 49, 517–21.

Pavlov, A., 2010. A new approach in survival analysis with longitudinal covariates. Ph.D. thesis, Queen's University (Canada).
URL `http://qspace.library.queensu.ca/jspui/bitstream/1974/5585/1/Pavlov_Andrey_201004_PhD.pdf`

Pearl, R., 1928. The Rate of Living, Being an Account of Some Experimental Studies on the Biology of Life Duration. Alfred A. Knopf, New York.

Pletcher, S. D., Neuhauser, C., 2000. Biological aging — criteria for modeling and a new mechanistic model. International Journal of Modern Physics C 11 (3), 525–46.

Prentice, R. L., 1982. Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika 69 (2), 331–342.

Raboud, J., Reid, N., Coates, R., Farewell, V., 1993. Estimating risks of progressing to aids when covariates are measured with error. Journal of the Royal Statistical Society. Series A (Statistics in Society), 393–406.

Ramsay, J., Silverman, B., 2005. Functional Data Analysis. SV, New York.

Rice, J., Silverman, B., 1991. Estimating the mean and covariance structure nonparametrically when the data are curves. Journal of the Royal Statistical Society. Series B (Methodological), 233–243.

Rockwood, K., Mitnitski, A. B., 2007. Frailty in relation to the accumulation of deficits. Journal of Gerontology: Medical Sciences 62A (7), 722–7.

Rosen, R., 1978. Feedforwards and global system failure: A general mechanism for senescence. Journal of Theoretical Biology 74, 579–90.

Schoop, R., Graf, E., Schumacher, M., 2008. Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. Biometrics 64, 603–10.

Schoop, R., Schumacher, M., Graf, E., 2011. Measures of prediction error for survival data with longitudinal covariates. Biometrical Journal 53 (2), 275–93.

Singpurwalla, N., 1995. Survival in dynamic environments. Statistical Science 10 (1), 86–103.

Sørensen, M., 1997. Estimating functions for discretely observed diffusions: A review. Lecture Notes — Monograph Series. Institute of Mathematical Statistics.

Sousa, I., 2011. A review on joint modelling of longitudinal measurements and time-to-event. REVSTAT–Statistical Journal 9 (1), 57–81.

Steinsaltz, D., Evans, S. N., June 2004. Markov mortality models: Implications of quasistationarity and varying initial conditions. Theoretical Population Biology 65 (4), 319–37.

Steinsaltz, D., Wachter, K. W., 2006. Understanding mortality rate deceleration and heterogeneity. Mathematical Population Studies 13 (1), 19–37, accepted for *Mathematical Population Studies*, Sept. 2005.

Stewart, E., Taddei, F., 2005. Aging in *Esherichia coli*: signals in the noise. BioEssays 27 (9), 983.

Strehler, B., Mildvan, A., 1960. General theory of mortality and aging. Science 132 (3418), 14–21.

Sweeting, M., Thompson, S., 2011. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. Biometrical Journal.

Swindell, W. R., 2009. Accelerated failure time models provide a useful statistical framework for aging research. Experimental gerontology 44 (3), 190–200.

Taylor, J. M. G., Cumberland, W. G., Sy, J. P., 1994. A stochastic model for analysis of longitudinal AIDS data. Journal of the American Statistical Association, 727–736.

Troxel, A. B., 2002. Techniques for incorporating longitudinal measurements into analyses of survival data from clinical trials. Statistical methods in medical research 11 (3), 237–245.

Tsiatis, A., Davidian, M., 2004. Joint modeling of longitudinal and time-to-event data: an overview. Statistica Sinica 14 (3), 809–834.

Tsiatis, A. A., DeGruttola, V., Wulfsohn, M. S., March 1995. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and CD4 counts in patients with AIDS. Journal of the American Statistical Association 90 (429), 27–37.

Vaupel, J. W., Carey, J. R., June 11 1993. Compositional interpretations of medfly mortality. Science 260 (5114), 1666–7.

Vaupel, J. W., Manton, K. G., Stallard, E., 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16 (3), 439–54.

Verbeke, G., Molenberghs, G., Rizopoulos, D., 2010. Random effects models for longitudinal data. In: van Montfort, K., Oud, J. H., Satorra, A. (Eds.), Longitudinal Research with Latent Variables. Berlin: Springer-Verlag, pp. 37–96.

Wang, Y., Taylor, J., 2001. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. Journal of the American Statistical Association 96 (455), 895–905.

Watve, M., Parab, S., Jogdand, P., Keni, S., October 3 2006. Aging may be a conditional strategic choice and not an inevitable outcome for bacteria. Proceedings of the National Academy of Sciences 103 (40), 14831–5.

Webb, G. F., Blaser, M. J., February 26 2002. Dynamics of bacterial phenotype selection in a colonized host. Proceedings of the National Academy of Sciences 99 (5), 3135–40.

Wei, L., 1992. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. Statistics in Medicine 11 (14-15), 1871–1879.

Weismann, A., 1892. Ueber die Dauer des Lebens. In: Aufsätze über Vererbung und verwandte biologische Fragen. Verlag von Gustav Fischer, Jena, Ch. 1, pp. 1–72.
URL `http://www.zum.de/stueber/weismann/`

Weitz, J., Fraser, H., December 18 2001. Explaining mortality rate plateaus. Proceedings of the National Academy of Sciences, USA 98 (26), 15383–6.

Whitmore, G., Crowder, M., Lawless, J., 1998. Failure inference from a marker process based on a bivariate wiener model. Lifetime Data Analysis 4 (3), 229–251.

Witten, M., 1985. A return to time, cells, systems, and aging: III. Gompertzian models of biological aging and some possible roles for critical elements. Mechanisms of Ageing and Development 32, 141–77.

Wulfsohn, M. S., Tsiatis, A. A., 1997. A joint model for survival and longitudinal data measured with error. Biometrics, 330–339.

Xu, J., Zeger, S., 2001. Joint analysis of longitudinal data comprising repeated measures and times to events. Journal of the Royal Statistical Society: Series C (Applied Statistics) 50 (3), 375–387.

Yashin, A., Arbeev, K., Akushevich, I., Kulminski, A., Akushevich, L., Ukraintseva, S., 2007. Stochastic model for analysis of longitudinal data on aging and mortality. Mathematical biosciences 208 (2), 538–551.

Yashin, A. I., Akushevich, I. V., Konstantin, G. A., Kulminski, A., Ukraintseva, S. V., 2011. Joint analysis of health histories, physiological state, and survival. Mathematical Population Studies 18 (4), 207–233.

Yashin, A. I., Iachine, I. A., Begun, A. S., 2000. Mortality modeling: A review. Mathematical Population Studies 8 (4), 305–32.

Yashin, A. I., Manton, K. G., 1997. Effects of unobserved and partially observed covariate processes on system failure: a review of models and estimation strategies. Statistical Science 12 (1), 20–34.

Yashin, A. I., Vaupel, J. W., Iachine, I. A., 1994. A duality in aging: the equivalence of mortality models based on radically different concepts. Mechanisms of Ageing and Development.

Yu, M., Law, N. J., Taylor, J. M. G., Sandler, H. M., 2004. Joint longitudinal-survival-cure models and their application to prostate cancer. Statistica Sinica 14 (3), 835–862.

Zou, S., Liedo, P., Altamirano-Robles, L., Cruz-Enriquez, J., Morice, A., Ingram, D. K., Kaub, K., Papadopoulos, N., Carey, J. R., 04 2011. Recording lifetime behavior and movement in an invertebrate model. PLoS ONE 6 (4), e18151.
URL http://dx.doi.org/10.1371%2Fjournal.pone.0018151

Appendix A. Calculations for the Le Bras model

We show here (briefly) the calculations for the mutual information in the Le Bras model. We recall that the model is a continuous-time birth process $(X_t)_{t\geq 0}$ on states $\{1, 2, 3, \dots\} \cup \{\partial\}$, where $X_0 = 1$ and ∂ is an absorbing state

(representing death). The transition rates from state x to $x+1$ are $q_{x,x+1} = \lambda x$, and $q_{x,\partial} = \mu x$.

Let L be the last senescence state before death, and T the time of death. Then L is distributed geometrically with parameter $\mu/(\mu+\lambda)$, and setting $p_t := e^{-(\mu+\lambda)t}$ we have

$$\mathbb{P}\{L = i\} = \frac{\mu\lambda^{i-1}}{(\mu+\lambda)^i}, \tag{A.1}$$

$$\mathbb{P}\{T \le t \mid L = i\} = \left(1 - p_t\right)^i \text{ if } X_0 = 1, \text{ so} \tag{A.2}$$

$$\mathbb{P}\{T > t\} = \frac{\mu+\lambda}{\mu/p_t + \lambda}, \tag{A.3}$$

$$\mathbb{E}[T] = \int_0^\infty \mathbb{P}\{T > t\}dt = \lambda^{-1}\log\left(1 + \frac{\lambda}{\mu}\right), \tag{A.4}$$

$$\mathbb{P}\{X_t = x \mid T > t\} = \left(\frac{\lambda}{\lambda+\mu}\right)^x \left(\frac{\mu}{\lambda} + p_t\right)(1-p_t)^{x-1}. \tag{A.5}$$

The stationary population distribution — equivalent to sampling a random individual at a random point of the lifespan — is

$$f(t) = \frac{\mathbb{P}\{T > t\}}{\mathbb{E}[T]} = \frac{(\mu+\lambda)\lambda}{\log(1+\lambda/\mu)}\left(\mu e^{(\mu+\lambda)t} + \lambda\right)^{-1}. \tag{A.6}$$

Let (τ, X_τ, S) be the current age, current senescence state, and remaining lifespan ($S = T - \tau$) of an individual selected from the stationary distribution. In other words, we pick a random individual (according to the stationary distribution), and find that that individual has simultaneously a random age τ, and a random current senescence state X_τ whose distribution may be computed conditional on the value of τ.

Given that the age is $\tau = t$, we find X_t as a sample from the distribution conditioned on survival to age t given in (A.5). In particular,

$$\mathbb{P}\{X_\tau = x\} =: \pi(x) = \frac{1}{\log(1+\lambda/\mu)}\frac{1}{x}\left(\frac{\lambda}{\lambda+\mu}\right)^x. \tag{A.7}$$

Somewhat more complicated is the remaining lifespan of an individual known to be in state x. There are no simple formulas like (A.2) and (A.3). We may write the probability as a sum

$$\mathbb{P}\{T > t+s \mid X_s = x\} = \sum_{i=1}^\infty \frac{\mu\lambda^{i-1}}{(\mu+\lambda)^i}\sum_{j=0}^{i-1}\binom{x+i-1}{j} p_t^{x+i-1-j}(1-p_t)^j. \tag{A.8}$$

Differentiating yields a conditional density $g_x(t)$, which may be computed and integrated numerically. The mutual information $I(T-t; X_t)$ that we plot in Figure 2 may then be computed (numerically) as

$$\sum_{x=1}^\infty \mathbb{P}\{X_t = x \mid T > t\}\int_0^\infty g_1(s)\log\frac{g_1(s)}{g_x(s)}ds.$$

To compute the mutual information for a random individual we compute the survival distribution weighted by the stationary age distribution, which is

$$s_a(t) = \mathbb{P}\{T > t + a \mid T > a\} = \frac{\mu}{\lambda} \log\left(1 + \frac{\lambda}{\mu} p_t\right) \frac{\mu p_t + \lambda p_{a+t}}{\mu + \lambda p_{a+t}}. \tag{A.9}$$

Integrating this with respect to the density f yields the survival function of an individual whose age is sampled from the stationary age distribution:

$$G(t) = \frac{\log(1 + \lambda p_t/\mu)}{\log(1 + \lambda/\mu)} \frac{(\mu + \lambda)(\mu p_t + \lambda p_{a+t})}{1 + \lambda p_{a+t}/\mu}. \tag{A.10}$$

Then

$$I(A; T_A - A) = \int_0^\infty f(a) \int_0^\infty -G'(t) \log \frac{G'(t)}{s_a(t)} dt da;$$
$$I(X_A; T_A - A) = \sum_{x=1}^\infty \pi(x) \int_0^\infty -G'(t) \log \frac{G'(t)}{p_x(t)} dt.$$

We turn now to the ROC curves for imminent mortality (equivalent to Lorenz curves for hazard rate, as described in section 4.1.3). An individual sampled from the stationary distribution has states distributed according to (A.7), so the fraction in state $\geq x$ is $\mathcal{P}_x := \sum_{y \geq x} \pi(y)$. Since the hazard rate in state x is μx, the total hazard corresponding to individuals in states $\geq x$ is $\sum_{y \geq x} \mu y \pi(y)$, and the proportion of total hazard is $\mathcal{H}_x = (1 + \mu/\lambda)^{-(x-1)}$. The solid curves in Figure 3 are computed by plotting pairs $(\mathcal{P}_x, \mathcal{H}_x)$ for $x = 1, 2, 3, \ldots$ (and interpolating linearly). For stratification by age, we note that the deaths at age s in the stationary population — equivalent to the deaths at age s within a cohort — are proportional to the product of the density of individuals at age s, which is $f(s)$, and the hazard rate at age s, which is $G'(s)/G(s)$. Thus the proportion of total hazard in the stationary population distribution belonging to individuals above age t computed as

$$\mathcal{H}_t^* = \left(\int_t^\infty \frac{-G'(s)}{G(s)} f(s) ds\right) \left(\int_0^\infty \frac{-G'(s)}{G(s)} f(s) ds\right)^{-1}.$$

The dashed curves in Figure 3 come from plotting $\mathcal{H}_t^*$ against $\mathcal{P}_t^* = \int_t^\infty f(a) da$, the survival probability to age t of the stationary distribution, equivalent to the fraction of total cohort lifespan spent above age t.