MARKOV MORTALITY MODELS: IMPLICATIONS OF QUASISTATIONARITY AND VARYING INITIAL DISTRIBUTIONS

DAVID STEINSALTZ AND STEVEN N. EVANS*

ABSTRACT. This paper explains some implications of markov-process theory for models of mortality. We show that an important qualitative feature which has been found in certain models — the convergence to a "mortality plateau" — is, in fact, a generic consequence of the convergence to a "quasistationary distribution", which has been explored extensively in the mathematical literature. This serves not merely to free these results from specifics of the models, but also to offer a new explanation of the convergence to constant mortality. At the same time that we show that the late behavior — convergence to a finite asymptote — of these models is almost logically immutable, we also show that the early behavior of the mortality rates can be more flexible than has been generally acknowledged. We point, in particular, that an appropriate choice of initial conditions enables one popular model to approximate any reasonable hazard-rate data. This illustrates how precarious it can be to read a model's vindication from the consilience with a favored hazard-rate function, such as the Gompertz exponential.

1. INTRODUCTION

1.1. **Outline of the problem.** A host of probabilistic models have been pressed into service to explain the randomness of aging and mortality. While many are ingenious and even enlightening, they are in a certain respect scattershot efforts. Quite commonly, some demographic principles or biological theory indicates a general class of mathematically related models. The analysis, however, is typically confined to a few selected examples, distinguished, as often as not, for being especially tractable. Even when the methods are competent to apprehend any particular model we might choose, how ought we to identify the ones of interest? Efforts to draw broad lessons from the behavior of a model are stymied when we cannot be sure how widespread the behavior is, even among trivial variants of the model. At the same time, when we attempt to validate a class of models empirically, our judgment of success must depend on the ease with which the parameters might equally have been tuned to match any arbitrary alternative data.

Date: June 10, 2003.

* Supported in part by NSF Grant DMS-00-71468 and by the Miller Foundation.

More sophisticated mathematical theory can help to redress these defects, by working with the models in their broader context. The first problem is overspecificity of particular models. Here it may be possible to show that the behavior under consideration is indeed common to the entire class of models, without ever considering any particular example. Not only will this provide assurance that the match of theory and phenomenon was not an artifact of the arbitrary choice of representative model, but the more general derivation will likely also be more illuminating than the bare computation that may suffice in the special case. Solving the general case typically forces us to come to grips with the fundamental issues. The second problem is overgenerality of classes of models. Here we may hope to characterize the widest class of patterns which could possibly be matched by suitable parameter choices. From this we can judge whether our success with matching actual data truly validates model, or whether it was simply unavoidable, given the loose definition of the model. In some cases, the class is defined by parameters explicitly built into the model. In other cases, the class is merely an implicit nimbus around a single model, defined by alternatives to arbitrary or indifferent choices which may have been fixed purely for technical convenience. For this reason, it will be appropriate to seek a mathematically natural class of models which will be equivalent from the perspective of the biodemographic justification of the model.

Both of these problems have dogged the study of markov mortality models. By a "markov mortality model" we will mean a process which is "killed" at a random stopping time, according to the behavior of a markov process, which itself is typically unobserved. In principle, requiring the hidden process to be markov is no constraint at all, since the current state may be made arbitrarily complex, so as to include whatever memory factors influence the future. In practice, though, markov models are generally assumed to have simple state spaces, representing a "vitality" often in one dimension.

Several authors (including [GG91, Section 6.4], [GG01], [WF01], [Bai00]) have proposed markov mortality models, and pointed out that they produce plateaus in late-life mortality. On a very general level, this could advance the discussion of mortality plateaus, by offering an explanation essentially distinct from the conventional dyad of physical heterogeneity of the population, as against temporal heterogeneity of the aging process. ("Mortality correlation", described by J. Vaupel *et al.* [Vau98], and modelled by B. Charlesworth [Cha01], is an evolutionary explanation of underlying causes, not a functional explanation, and so exists on a different plane from this pair.) What J. Weitz and H. Fraser, in particular, point out explicitly, is that no special pleading is required to generate mortality plateaus.

A very simple model — in this case, a Brownian motion with constant drift, killed at the origin — does generate these features. While this "constant-drift model" is not profoundly realistic, what is essential for their argument is merely the absence of *a priori* reference to mortality at late ages. This important point deserves to be rescued from the specificity of their methods.

The trouble begins with the "inverse Gaussian" computation, by which Weitz and Fraser derive the asymptotics of the mortality rates in their model. This approach resists generalization even to slight variants, as do the computations in several of the other mentioned papers. (An exception, of sorts, is the paper by W. Bains [Bai00], which presents only numerical computations. These could, in principle, be extended to a broad class of related models, though only one by one.) The mortality plateau in each of these analyses is *sui generis*, absent any unifying principle. And yet, as we explain in section 3, when viewed in its proper mathematical context, each model mortality plateau is an example of a generic phenomenon, the convergence to quasistationary distributions. The decelerating mortality rates, we see, are not the consequences of any special properties of these models; rather, a markov process would need to be fairly pathological to escape such behavior.

The benefit here, we wish to emphasize, is not merely in the generality of the results. Where before we had merely an observation — this model, and that one, produce mortality plateaus — we now have something more closely resembling an explanation. In the biodemographic context, this might be called an *evolving heterogeneity* theory of mortality tapering. As a markov process progresses, the distribution of its state is being shaped by two forces: random motion, which tends to spread the mass out and shift it in certain preassigned directions; and deaths, which lop off mass at each point, at a fixed rate. In most cases, there are certain distributions of probability mass whose shapes are stable, so that their levels sink proportionately at every location. Intuitively, this often means that mass is concentrated in locations with lower mortality rates, so that the flow out will balance the slow death. What is more, it can be shown in many cases that no matter where the process starts, if we wait long enough the distribution of those individuals who survive will approach a certain one of these quasistationary distributions. The mortality rate, of course, will also approach the mortality rate averaged over this distribution. In other words, the mortality rate stops increasing, not because we have selected out an exceptional subset of the population, but because the condition

This model has been applied to demography, with different goals, by J. Anderson [And00]. It earlier appeared as a lifetime model in the engineering literature [CF77].

of the survivors is reflective of their being survivors, even though they started out the same as everyone else. (A suggestion along these lines may be found on page 210 of [FK00].) A population piled up in the low-vitality domain is unstable. What is more, when the state space has regions of differing mortality, the surviving population will be found concentrated in the more salubrious realms.

To make a banal analogy, compare this problem to the "mortality" of automobiles, such as has been described by J. Vaupel [Vau97]. It is natural to expect that a 2000 automobile will be in better condition than a 1990 automobile (barring a decline in the craft of automobile engineering), and consequently less likely to break down. This will continue for some time into the past. But will the 1970 automobile be more or less prone to breaking down than an old-timer from 1930? There are very few 1930s automobiles still in service, but those that remain seem likely to be in especially good condition. This is not because the survivors necessarily were unusually sturdy to begin with, but rather, that their survival reflects a special life course.

Some parallels might be drawn to the notion of "induced demographic schedules", proposed by Vaupel *et al.* [Vau98] as one explanation for mortality deceleration. This refers to the ability of some organisms to switch among two or more distinct life histories, often with wide disparities in the typical lifespans. This is similar to our notion, in that the organisms may begin life identical, but become heterogeneous in the course of their lives, perhaps because of environmental influences. An important difference is that we require no specialized life-history adaptations. The heterogeneity that develops, and that induces the mortality plateaus, is in the vitality itself, not in an extraneous genetic switch. (This is not intended as an argument against induced demographic schedules, which are indisputably a real phenomenon, and which may contribute to mortality deceleration; we wish merely to point up the differences to our proposal.)

1.2. **Outline of the paper.** In sections 2.2 and 3.1 we offer an account of some general theory of quasistationary distributions, for finite and general state spaces respectively. Much of this is hoary in the mathematical literature, although we offer a few improvements. While these results are not specifically applied in the sequel, they are central to our program, in that they reveal the convergence to quasistationary distributions, and the convergence to asymptotic killing rates, as truly generic features of markov models. The key idea, which is best understood in this abstract setting, is the connection between asymptotic killing rates for markov processes, and their invariant functions. It must be admitted that the general theory is not

as complete as one might wish. Neither the convergence to quasistationary distributions nor to asymptotic mortality rates can be guaranteed in perfect generality, even when the state space is compact. What we have instead is a long-term average killing rate, called the *decay parameter*, and its link to sub- and superinvariant functions, weaker forms of quasistationary distributions.

More complete results can be given when the markov processes are diffusions, particularly in one dimension, and this theory is discussed in sections 3.2 and 3.3. Numerous examples, both discrete-space and diffusions, including several which have already been intensively studied in the aging literature, are then analyzed in section 4. In particular, we see that in the constant-drift model, the limit mortality is $b^2/2\sigma^2$, where b is the drift and σ the diffusion constant. (This was computed directly from the inverse Gaussian density by R. Chhikara and J. Folks [CF77].) The vitality converges to a gamma distribution, with shape parameter 2 and exponential rate b/σ.

In section 5 we move from the behavior as time goes to infinity, to the early behavior of some models. Here we mean to address the second problem mentioned above, the unexplored flexibility of some models. Researchers often highlight the hazard rates that appear from their models, seeing in them passable reflections of real data. We point out, in passing, in section 5.1, how some have allowed themselves to be misled by the appearance of the Gompertz curve into accepting an erroneous computation. Turning then to the example of one-dimensional diffusions, particularly those with constant drift, we see that even when the computation is correct it may be rash to infer much from mere similarity of theoretical and empirical curves. In most cases, the model presented is only one of a family of possibilities that are all equally plausible, at least superficially. Viewing the class as a whole, it may be possible to tune parameters to match almost any data we might happen to be given. To illustrate this, we characterize all hazard rates which can be obtained from the constant-drift model by varying the initial distributions. There is nothing in the nature of the model which gives us a clear direction about the appropriate starting state. And yet, we see that the indefiniteness of the initial distribution translates into enormous latitude to shape the mortality rates in such models. While this does not negate the value of such models, it suggests that different and more careful work would be needed to identify appropriate models, and appropriate versions of models.

2. Finite state spaces

2.1. **An introductory example.** Consider a model in which an organism has two possible states: healthy and sick. Healthy individuals may stay healthy or get sick; sick individuals may stay sick or get healthy, but they may also die. For definiteness, let us say that the rate of healthy individuals getting sick is α, the rate of sick individuals getting well is β, and their rate of dying is δ. A markov process is represented by a matrix Q with entries q_{ij}, being the rate at which a process currently in state i makes jumps to the state j. The sum $\sum_j q_{ij}$ of the transition rates in any given row is 0. The unkilled process described above has matrix $\left(\begin{smallmatrix} -\alpha & \alpha \\ \beta & -\beta \end{smallmatrix}\right)$, which means that it converges to a stationary state (corresponding to the left eigenvector with eigenvalue 0) with a proportion $\beta/(\alpha+\beta)$ healthy and $\alpha/(\alpha+\beta)$ sick.

Naively, one might expect that the limiting mortality rate would be $\delta \cdot \alpha/(\alpha+\beta)$. In fact, though, it will be smaller. Since being sick raises the likelihood of dying, survival implies a greater chance of being healthy. An individual who has survived a very long time is more likely to be healthy than an average individual who has not been exposed to mortality. To compute this, we take the infinitesimal matrix of the killed markov process. We can either define a third "cemetery" state ∂, such that transitions to ∂ are equivalent to dying. Then we get the 3×3 transition matrix

$$Q = \begin{pmatrix} -\alpha & \alpha & 0 \\ \beta & -\beta-\delta & \delta \\ 0 & 0 & 0 \end{pmatrix}. \tag{1}$$

Equivalently, we can just drop the last row and column, and work with the *sub-markov* transition matrix

$$Q = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta-\delta \end{pmatrix}. \tag{2}$$

This has two eigenvalues,

$$r_{1,2} = -\frac{\alpha+\beta+\delta}{2} \pm \sqrt{\frac{(\alpha+\beta+\delta)^2}{4} - \alpha\delta}.$$

Both eigenvalues are real and negative. The larger one, r_1, will be the limiting mortality rate. If, for example, we take $\alpha = \delta = 1$, and $\beta = 2$, we see that the naive mortality asymptote is 1/3, while the correct value is about .268 It is easy to see that the naive asymptote is always too large.

To make this more tangible, we know that if the starting distribution is (p_h, p_s), then the distribution at time t will be $(p_h, p_s)\exp(tQ)$. To write down the explicit formula would be more messy than illuminating; instead, we pursue the numerical

example. By diagonalizing Q, we have

$$\exp(tQ) = \begin{pmatrix} .789e^{-.268t} + .211e^{-3.73t} & .289e^{-.268t} - .289e^{-3.73t} \\ .577e^{-.268t} - .577e^{-3.73t} & .211e^{-.268t} + .789e^{-3.73t} \end{pmatrix}.$$

Thus

$$\begin{aligned} &e^{.268t}\big(P^t\{\text{healthy}\}, P^t\{\text{sick}\}\big) \\ &\quad = \big(.789p_h + .577p_s, .289p_h + .211p_s\big) + e^{-3.46t}\big(.211p_h - .577p_s, -.289p_h + .789p_s\big). \end{aligned}$$

For a healthy individual, we start with $p_h = 1$ and $p_s = 0$:

$$\mathrm{P}\big\{\text{survive until time } t\big\} = P^t\{\text{healthy}\} + P^t\{\text{sick}\} = 1.078e^{-.268t} - 0.078e^{-3.73t}.$$

The conditional probability of survival is then

$$\begin{aligned} &\mathrm{P}\big\{\text{survive until time } t+s \,\big|\, \text{survive until time } t\big\} \\ &\qquad = \frac{e^{-.268s} - 0.072e^{-3.46t-3.73s}}{1 - 0.072e^{-3.46t}} \xrightarrow{t\to\infty} e^{-.268s}. \end{aligned}$$

Note, as well, that conditioned on survival until t, as $t \to \infty$ the probability of being healthy converges to

$$\frac{.789p_h + .577p_s}{1.078p_h + .788p_s} = .732,$$

independent of the starting state. This distribution of healthy and sick — which is the same as the top eigenvector — is what we call the quasistationary distribution.

2.2. **General theory for finite state spaces.** As in the above example, we may represent death as entrance into the cemetery state ∂. A markov process with a single absorbing state will be referred to as a *killed markov process*. We will follow the alternative route of dropping the cemetery state, and allowing the particle simply to disappear at death. This turns the markov process (where the sum of the probabilities among the possible states is always 1) into a submarkov process (where the sum of the probabilities is decreasing with time).

Suppose we have a submarkov process, whose state space is a finite set $\mathcal{X}$. We write the infinitesimal rate of transition from state i to state j as q_{ij}. The rate of leaving state i is q_{ii}, so it is equal to $-k_i - \sum_{j\neq i} q_{ij}$, where k_i is the rate of dying at state i. We say that a state y is *accessible* from the state x if there is a sequence of states $x = i_0, i_1, \ldots, i_k = y$ such that $q_{i_m i_{m+1}} > 0$ for every m. Stated simply, this means that if the process starts at x, it has a nonzero probability of arriving eventually at y. The process is said to be *irreducible* if every state is accessible from every other.

Let $\mathcal{A}$ be the set of states which are accessible from all other states. Suppose that $\mathcal{A}$ is nonempty. Let T be the time of "death", and let $T_{\mathcal{A}}$ be the first time the

process arrives in $\mathcal{A}$. Conditioned on $T > t$, the probability of $T_{\mathcal{A}} > t$ (that is, the probability that the process has never been in $\mathcal{A}$ up to time t) goes exponentially to 0. Note that $\mathcal{A}$ is an absorbing set. Thus, conditioned on survival up to time t, the process becomes concentrated on $\mathcal{A}$.

Proposition 1. *Let λ be the eigenvalue of $-Q$ with the smallest real part. Then λ is purely real and nonnegative, and is a simple eigenvalue. The unique left eigenvector π with $\sum \pi_i = 1$, and right eigenvector v with $\sum \pi_i v_i = 1$, corresponding to λ, are nonnegative, and all components in $\mathcal{A}$ are positive. For any states i and j,*

$$\lim_{t\to\infty} \mathrm{P}_i\big\{X_t = j \,\big|\, T > t\big\} = \pi_j$$

and

$$\lim_{t\to\infty} \mathrm{P}\big\{T > t + s \,\big|\, T > t\big\} = e^{-\lambda s}.$$

(The i in the subscript represents the starting state.) In addition, for any states i and j,

$$\lim_{t\to\infty} e^{\lambda t}\, \mathrm{P}\big\{X_t = i \,\big|\, X_0 = j\big\} = v_j \pi_i.$$

Proof. The positivity results are simply the Perron-Frobenius Theorem [HJ85], applied to e^{tQ} for arbitrary positive t. The rest is elementary linear algebra. The theory of such "Perron-Frobenius eigenvalues" is discussed at length in [Sen73]. □

Note that when the set A is empty, the state space may be decomposed into subsets which are mutually inaccessible. The above result may then be applied to the subsets separately.

A version of this theory for finite state spaces and discrete time was worked out in [DS65], and extended to some cases of infinite discrete state spaces in [SVJ66]. Infinite state spaces are not so easy to treat in a comprehensive framework, but some general results for the special case of birth-death processes may be found in [Doo91].

3. Quasistationary distributions for general state spaces

A general fact about killed markov processes, under fairly general conditions, is that their rates of killing, averaged over a long time, converge to a finite rate, called the *decay parameter*. The major results in section 3.1 are all due to R. Tweedie and several coauthors. In the greatest generality, the existence of a decay parameter λ is a fairly weak statement. It states, essentially, that $P^t(x, A)$ (the probability that the process started from x will be in the set A at time t) declines, on an exponential scale, at a rate on the order of $e^{-\lambda t}$. The decay parameter is defined by the existence of so-called subinvariant measures and functions, a generalization

of eigenmeasures and eigenfunctions. The exact statement is given in Proposition 3.

We will say that the decay parameter is the *asymptotic killing rate* for the process — a much stronger statement, and one that is closer to our concerns — if the killing time for the process started at any point x satisfies

$$\lim_{t\to\infty} \mathrm{P}_x\big\{T > t+s \,\big|\, T > t\big\} = e^{-\lambda s}. \tag{3}$$

Clearly, if for each $x \in \mathcal{X}$ there is a positive constant c_x such that

$$\lim_{t\to\infty} e^{\lambda t}\,\mathrm{P}_x\big\{T > t\big\} = c_x,$$

then λ is the asymptotic killing rate. As we explain below, a sufficient condition for the decay parameter to be the asymptotic killing rate is that the process have a property called λ-*positivity*. A characterization of λ-positive processes in terms of subinvariant measures and functions, due to Tweedie and P. Tuominen, is given here as Proposition 4.

Even when the process is not λ-positive, so that the rate of decay is not precisely exponential, we may still have an asymptotic killing rate if the conditional probability converges to a quasistationary distribution. That is,

$$\lim_{t\to\infty} \mathrm{P}_x\big\{X_t \in A \,\big|\, T > t\big\} = \mu(A),$$

where μ satisfies

$$\int_{\mathcal{X}} P^t(x,A)d\mu(x) \le e^{-\lambda t}\mu(A).$$

This follows from

$$\begin{aligned}\lim_{t\to\infty} \mathrm{P}_x\big\{T > t+s \,\big|\, T > t\big\} &= \lim_{t\to\infty} \frac{\int_{\mathcal{X}} \mathrm{P}^t(x,dy)P^t(y,\mathcal{X})}{\int_{\mathcal{X}} \mathrm{P}^t(x,dy)} \\ &= \int_{\mathcal{X}} P^s(y,\mathcal{X})\mu(dy) \\ &\le e^{-\lambda s}.\end{aligned}$$

It is this criterion that we will use in section 3.2 for diffusions on an unbounded interval. Here

$$P^t(x,A) = \mathrm{P}\big\{X_t \in A \,\big|\, X_0 = x\big\}.$$

While these theorems provide a useful overview of the convergence properties of submarkov processes, they are limited in two ways. On the one hand, the conditions are difficult to check in general. In particular, finding subinvariant measures and functions (and, which is essential, knowing that we have found all of them) is rarely possible in infinite state spaces, except when the state-space is the real line, where we have the machinery of ordinary differential equations to hand. On the other

hand, the conditions are too strict. When the process is not λ-positive, the general theory tells us very little. Many of the models that would interest us — in particular, diffusions which are not confined to a compact set — are not λ-positive. They may, nonetheless, have asymptotic killing rates. The reason is straightforward: when the process is λ-positive, the decay is exactly exponential. But in processes such as the diffusion models we examine in sections 4.1 and 4.3, the probability of survival up to time t falls as $\frac{1}{t}e^{-\lambda t}$. (In technical terms, this happens because the generator has a continuous spectrum, rather than the discrete spectrum which is are typical in diffusions on compact intervals.) This rules out an application of positivity methods, but has no effect on the results that we are concerned with, which refer only to relative probabilities: the probability of being in a given set conditioned on having survived for a very long time. In section 3.2 we derive a general result for the long-term behavior of one-dimensional diffusion processes. In section 3.3 we explain why killed Brownian motions in general dimensions are indeed λ-positive, as long as they are confined to a compact set.

3.1. **General theory.** Here we present the broadest current understanding of the asymptotic killing rate of submarkov processes. For those unfamiliar with the terminology of abstract markov processes, useful introductions may be found in the books [KS88] and [RW00a], among many others.

For any real number ℓ, the σ-finite measure μ on $\mathcal{X}$ is said to be *ℓ-subinvariant* for P^t if for all t and all measurable A with $\mu(A)$ finite,

$$\int_{\mathcal{X}} P^t(x, A)d\mu(x) \leq e^{-\ell t}\mu(A); \tag{4}$$

if the inequality (4) is an equality, then we call μ *ℓ-invariant*. If f is a function from $\mathcal{X}$ to $\mathbb{R}^+$, it is said to be *ℓ-subinvariant* for P^t if for all t

$$\int_{\mathcal{X}} P^t(x, dy)f(y) \leq e^{-\ell t}f(x), \tag{5}$$

except for at most an exceptional set of points x with ϕ-measure 0. *Superinvariant* functions and measures are defined by reversing these inequalities. A function or measure is *ℓ-invariant* if it is both ℓ-subinvariant and ℓ-superinvariant. Except for some technical details, which we will not discuss here, an ℓ-invariant function (or measure) is the same as an eigenfunction (or eigenmeasure) of the generator (or its adjoint) with eigenvalue $-\ell$. In the case of diffusions, the generator is a differential operator, allowing the problem to be attacked with the methods of differential equations. This case will be discussed in sections 3.2 and 3.3.

There is a powerful, though almost trivial, relationship between invariant measures and decay rates.

Proposition 2. *Let μ be the starting distribution for the submarkov process X_t. If there exists a λ-subinvariant measure ν such that $C := \sup_{x\in\mathcal{X}} d\mu/d\nu(x)$ is finite, then*

$$\mathrm{P}\big\{\textit{survive until time } t\big\} \leq Ce^{-\lambda t}, \tag{6}$$

which gives a lower bound on the decay parameter. If there exists a λ-superinvariant measure ν such that $c := \inf_{x\in\mathcal{X}} d\mu/d\nu(x) > 0$, then

$$\mathrm{P}\big\{\textit{survive until time } t\big\} \geq ce^{-\lambda t}, \tag{7}$$

which gives an upper bound on the decay parameter. If there exists a λ-invariant measure ν such that $\infty > C \geq c > 0$, then

$$Ce^{-\lambda t} \geq \mathrm{P}\big\{\textit{survive until time } t\big\} \geq ce^{-\lambda t}, \tag{8}$$

which implies that the decay parameter exists, and is equal to λ.

Proof. We prove statement (6). The other proofs are essentially the same. The probability of survival is given by

$$\int_{\mathcal{X}} P^t(x,\mathcal{X})\mu(dx) \leq C\int_{\mathcal{X}} P^t(x,\mathcal{X})\nu(dx) \leq Ce^{-\lambda t}\nu(\mathcal{X}) = Ce^{-\lambda t}.$$

□

This result is useful, in that it allows us to derive bounds on the decay rate in particular cases. The upper bound on the decay parameter holds, loosely speaking, if the starting condition is sufficiently diffuse (not concentrated at points, in particular) and if it avoids overemphasizing the boundary (when the invariant measure is forced to 0 there). The lower bound holds, more or less, when the starting measure is sufficiently spread out: putting positive weight whereever the invariant measure does. Of course, we may start with a point mass and let the process run for a short time, until the distribution has spread out, and then treat what results as the new starting distribution.

Intuitively, one would expect that if the process is sufficiently smooth, and the state space is compact, then it should be possible to compare the top invariant measure with the result of running the process for a fixed time, and so apply equation (8). For diffusions on a bounded domain in $\mathbb{R}^n$ with Hölder differentiable coefficients, this is shown in [GQZ88].

It would help to know, more generally, when such sub- and super-invariant measures exist. In addition, we might like to know whether the average decay rate holds uniformly over all starting measures. The most important results along these lines are due to R. Tweedie and various coauthors. We summarize some of this

work in the following Theorem, extracted from more extensive results in Theorems 2, 3, and 6 of [TT79], and Proposition 3.1 of [Twe74].

Theorem 3 (Tuominen and Tweedie)**.** *Let X_t be a ϕ-irreducible, submarkov process, such that $P^t(x, A)$ is continuous in t for each x and measurable A. Let λ be the supremum of those ℓ such that there is an ℓ-subinvariant measure for P^t. Then there exists a λ-subinvariant function f and a λ-subinvariant measure π. If A is a measurable set with $0 < \pi(A) < \infty$, then*

$$\int_0^\infty e^{st} P^t(x, A)dt < \infty \text{ if } s < \lambda, \text{ and} \qquad \int_0^\infty e^{st} P^t(x, A)dt = \infty \text{ if } s > \lambda. \tag{9}$$

If, in addition, $\inf_{x\in A} f(x) > 0$*,*

$$\lim_{t\to\infty} \frac{1}{t} \log P^t(x, A) = -\lambda. \tag{10}$$

These results hold for all starting points x, except perhaps for an exceptional set with ϕ-measure 0.

We say that the process X_t is *λ-transient* if

$$\int_0^\infty e^{\lambda t} P^t(x, A)dt$$

is finite for all $x \in \mathcal{X}$ and all A with $\phi(A) > 0$. Otherwise, X_t is said to be *λ-recurrent*. Theorem 3 of [TT79] tells us that when X_t is recurrent, there is a unique λ-subinvariant measure π and a unique λ-subinvariant function f, and these are λ-invariant. If $\int f d\pi$ is finite, then we say that P^t is *positive λ-recurrent* (or *λ-positive*).

Conversely, we have (combining Theorems 4 and 5 of [TT79] with Proposition 4.4 of [Twe74])

Proposition 4. *Let X_t be a ϕ-irreducible submarkov process such that $P^t(x, A)$ is a continuous function of t for every x and A. If for some ℓ there exists an ℓ-invariant function f and a ℓ-invariant measure π, such that $\int f d\pi$ is finite, then $\ell = \lambda$ and the process is λ-positive.*

For a λ-positive process,

$$\lim_{t\to\infty} e^{\lambda t} P^t(x, A) = \frac{f(x)\pi(A)}{\int f d\pi} \tag{11}$$

for π-almost every x, and every measurable A such that $\inf_{x\in A} f(x) > 0$*. The asymptotic killing rate is λ.*

3.2. **One-dimensional diffusions with killing.** A strong markov process with continuous sample paths on some subset of $\mathbb{R}^d$ is called a *diffusion*. Most reasonable one-dimensional diffusions that appear in practice may be defined, for state space $\mathcal{X} = (r_1, r_2)$, where r_1 may be $-\infty$ and r_2 may be $+\infty$, by three functions: the *diffusion rate* $\sigma : \mathcal{X} \to [0, \infty)$, the *drift* $b : \mathcal{X} \to \mathbb{R}$, and the *killing rate* $\kappa : \mathcal{X} \to [0, \infty)$. General terminology and theory of diffusions may be found in any number of books, including [KS88], [RY90], and [RW00a]. Those unfamiliar with the theory of boundaries for one-dimensional diffusions can find two very different presentations in [Fel52] and [RW00b, section V.51]. We will assume the drift to be continuously differentiable, the diffusion rate to be continuously twice differentiable, and the killing rate to be continuous. Proofs of all the results in this section may be found in [SE03].

The infinitesimal generator of the unkilled process is (see, for example, section VII.2 of [RY90])

$$\mathcal{L}\phi(x) := \lim_{t \downarrow 0} \frac{E_x[\phi(X_t)] - \phi(x)}{t} = \frac{1}{2}\sigma^2(x)\phi''(x) + b(x)\phi'(x), \tag{12}$$

and the adjoint operator is

$$\mathcal{L}^*\phi = \frac{1}{2}\big(\sigma^2(x)\phi(x)\big)'' - \big(b(x)\phi(x)\big)'. \tag{13}$$

The generator describes the behavior of the diffusion in the interior of $\mathcal{X}$. To describe the complete behavior of the diffusion we need to add boundary conditions (as discussed below), to determine whether it is reflected or killed at the endpoints. By the Feynman-Kac formula (Proposition VIII.3.10 of [RY90]), the infinitesimal generator of the process with killing is $\mathcal{L} - \kappa$.

We may simplify the problem somewhat by assuming that σ is identically 1. There is no loss of generality since (following [Fel52]) we can replace X_t by $Y_t = F(X_t)$, where $F(x) := \int_{x_0}^{x} du/\sigma(u)$, where c is an arbitrary point in (r_1, r_2). The killing rate for Y_t becomes $\kappa(F^{-1}(Y_t))$, while the drift may be computed by Itô's formula to be

$$\frac{b(F^{-1}(Y_t))}{\sigma(F^{-1}(Y_t))} - \sigma'(F^{-1}(Y_t)). \tag{14}$$

Since σ is positive on the interior of the interval, this transformation is finite. From now on, we will always assume, unless otherwise stated, that the diffusion rates of our one-dimensional diffusions are identically 1.

Fix any point x_0 in the interior of the interval (r_1, r_2), and define

$$B(x) := \int_{x_0}^{x} 2b(z)dz.$$

We define the function $\phi_\lambda : \mathcal{X} \to \mathbb{R}$ to be the unique nonzero solution to the initial value problem

$$\begin{aligned} &\frac{1}{2}\frac{d^2}{dx^2}\phi_\lambda - \frac{d}{dx}\big(b \cdot \phi_\lambda\big) - \kappa\phi_\lambda = -\lambda\phi_\lambda \\ &\qquad \text{with } (1-p_i)e^{-B(r_i)}\phi_\lambda(r_i) = p_i\big(\phi'(r_i) - b(r_i)\phi_\lambda(r_i)\big), \end{aligned} \tag{15}$$

if it exists. Here p_i represents the degree of reflection at the boundary point r_i. It is implicitly 0 at an inaccessible boundary.

Theorem 5. *Suppose that r_1 and r_2 are both regular boundaries. Let $\underline{\lambda} = \min\{\lambda : \phi_\lambda \text{ exists}\}$. Suppose*

$$\phi_{\underline{\lambda}} \text{ is nonnegative.} \tag{16}$$

If both boundaries are reflecting and $\kappa = 0$ almost everywhere then $\underline{\lambda} = 0$; otherwise, $\underline{\lambda}$ is positive. For any measurable subset $A \subset (r_1, r_2)$, and any starting point x,

$$\lim_{t\to\infty} e^{\underline{\lambda}t} P_x\big\{X_t \in A\big\} = \frac{e^{-B(x)}\phi_{\underline{\lambda}}(x) \cdot \int_A \phi_{\underline{\lambda}}(y)dy}{\int_{r_1}^{r_2} e^{-B(y)}\phi^2_{\underline{\lambda}}(y)dy}. \tag{17}$$

The asymptotic killing rate is $\underline{\lambda}$.

If b is finite at both boundaries then condition (16) *is automatically satisfied.*

Note that the orthogonality of eigenfunctions implies that $\underline{\lambda}$ is the only eigenvalue for which the boundary value problem (15) has a nonnegative solution.

Theorem 6. *Suppose that r_1 is a regular boundary, the drift b is continuous at r_1 and $r_2 = \infty$ a natural or entrance boundary. (An equivalent result holds if r_2 is the regular boundary and r_1 the natural boundary.) Suppose, too, that*

$$\int_{r_1}^{\infty} e^{B(z)}dz < \infty, \tag{18}$$

and

$$\liminf_{z\to\infty} z^{-2}\big(b(z)^2 + b'(z) + 2\kappa(z)\big) > -\infty. \tag{19}$$

Then $\underline{\lambda}$ is finite, and is equal to the decay parameter for the process X_t. For any measurable subset $A \subset [r_1, \infty)$,

$$\lim_{t\to\infty} \mathrm{P}\big\{X_t \in A \,\big|\, T > t\big\} = \frac{\int_A \phi_{\underline{\lambda}}(z)dz}{\int_{r_1}^{\infty} \phi_{\underline{\lambda}}(z)dz}, \tag{20}$$

and $\int_{r_1}^{\infty} \phi_{\underline{\lambda}}(z)dz$ is finite. The asymptotic killing rate is $\underline{\lambda}$.

Note that the condition (19) is fairly weak: it is automatically satisfied if the drift $b(z)$ does not drop suddenly for large z. There is little cost in imposing this condition, since it constrains only the local fluctuations of the drift. The condition

(18), on the other hand, is a global constraint on the behavior of $b(z)$ as $z \to \infty$. Essentially, this requires the drift to be sufficiently negative to keep the mass from drifting off to ∞. This is clearly too restrictive, since in some models (such as the one discussed in section 4.3) the process is prevented from drifting off to ∞ not by a countervailing drift, but by an increasing killing rate, which forces the conditioned process to lurk near the origin. For such models, we have the following result:

Theorem 7. *The conclusions of Theorem 6 remain valid when the condition* (18) *is replaced by*

$$\liminf_{z\to\infty} \kappa(z) > \underline{\lambda}. \tag{21}$$

3.3. **Killed Brownian motion on a compact set.** Let $\mathcal{X}$ be a bounded open subset of $\mathbb{R}^n$, with a smooth boundary. Let $\kappa : \mathcal{X} \to \mathbb{R}^+$ be a measurable function, and let $\mathcal{L}$ be the generator of a diffusion, whose diffusion rate is differentiable and whose drift is continuous on a neighborhood of $\mathcal{X}$. The diffusion rate for a process in $\mathbb{R}^n$ is a differentiable function on $\mathcal{X}$, whose values are $n \times n$ matrices. The drift is a continuous function that takes values in $\mathbb{R}^n$. For any distribution μ on $\mathcal{X}$ we may define a process X_t which starts at μ, and continues as a diffusion with generator $\mathcal{L}$ until the time T, which is the minimum of the killing time defined by κ and the first time when the process hits the boundary of $\mathcal{X}$.

In general, we have the following theorem due to M. Donsker and S. Varadhan [DV76, Theorem 2.2]:

Theorem 8 (Donsker and Varadhan)**.**

$$\lim_{t\to\infty} \sup_{x\in\mathcal{X}} \log \mathrm{P}_x\{T > t\} = \sup_{\mu} \inf_{\phi} \int_{\mathcal{X}} \frac{\mathcal{L}\phi(z) - \kappa(z)\phi(z)}{\phi(z)} d\mu(z), \tag{22}$$

where μ ranges over all probability measures on the closure of $\mathcal{X}$, and ϕ ranges over the functions in the domain of $\mathcal{L}$ which are bounded away from 0.

R. Pinsky [Pin85] has cited this theorem in a stronger form, with the supremum over $x \in \mathcal{X}$ on the left-hand side removed (so that the result applies to the decay parameter from any starting point). It is not clear to us how the stronger form follows from the weaker, in general.

If the base diffusion is Brownian motion — that is, with no drift, and with σ everywhere equal to the identity matrix — we may apply the same arguments as for Theorem 5 to show

Theorem 9. *Let X_t be Brownian motion on the bounded domain $\mathcal{X}$ in $\mathbb{R}^n$, killed at rate κ. The process X_t is λ-positive. Let $\underline{\lambda}$ be the maximum λ such that there is*

a nontrivial $\phi_\lambda : \mathrm{Int}(\mathcal{X}) \to \mathbb{R}^+$ *satisfying*

$$\frac{1}{2}\Delta\phi_\lambda - \kappa\phi_\lambda = -\lambda\phi_\lambda, \tag{23}$$

which converges to 0 on the boundary. (Here Δ *is the Laplacian operator.) Then* $\underline{\lambda}$ *is the decay parameter for* X_t*, and the asymptotic killing rate is* $\underline{\lambda}$*. In addition, for any measurable set* $A \subset \mathcal{X}$*,*

$$\lim_{t\to\infty} \mathrm{P}_\mu\big\{X_t \in A \,\big|\, T > t\big\} = \frac{\int_A \phi_{\underline{\lambda}}(z)d^n z}{\int_{\mathcal{X}} \phi_{\underline{\lambda}}(z)d^n z} \tag{24}$$

where T *is the killing time, and*

$$\lim_{t\to\infty} \mathrm{P}_x\big\{T > t+s \,\big|\, T > t\big\} = e^{-\underline{\lambda}s}. \tag{25}$$

Furthermore,

$$\lim_{t\to\infty} e^{\underline{\lambda}t} P_x\big\{X_t \in A\big\} = \frac{\int_A \phi_{\underline{\lambda}}(z)d^n z \cdot \int_{\mathcal{X}} \phi(z)d\mu(z)}{\int_{\mathcal{X}} \phi_{\underline{\lambda}}(z)^2 d^n z}.$$

More generally, as pointed out in the appendix to [Pin85], the same methods apply if $\sigma^{-1}b$ is a gradient function.

4. Examples

4.1. **Constant drift and killing at 0.** In the case when b is a negative constant, $\kappa \equiv 0$ on $\mathbb{R}^+$, and 0 is an absorbing boundary (the model, we recall, considered in [WF01], [And00], [CF77]), Theorem 6 becomes particularly easy to apply. For $\lambda \neq b^2/2$ we get the solution

$$\phi_\lambda(x) = \frac{1}{2\sqrt{b^2-2\lambda}}\left[\exp\left\{\left(b+\sqrt{b^2-2\lambda}\right)x\right\} - \exp\left\{\left(b-\sqrt{b^2-2\lambda}\right)x\right\}\right],$$

which changes sign for $\lambda > b^2/2$. Thus the limit mortality rate is $b^2/2$, and the density of the fitnesses of the survivors converges to

$$\phi_{b^2/2}(x) = xe^{bx},$$

which is a gamma distribution with exponential rate $|b|$ and shape parameter 2. If σ is not 1, we simply need to replace b by b/σ. (Remember that σ is constant here.)

4.2. **General killing: a discrete-space example.** The well-known "cascading failure" model was introduced by H. Le Bras [Bra76], and further studied by L. Gavrilov and N. Gavrilova [GG91]. This represents senescence as a discrete variable, with motion only in the increasing direction by steps of size 1. The rate of jumps from state x to $x+1$ is a constant λ times x, and the process is killed at rate μx when it is in state x. One diffusion analogue of this would have a Brownian component of intensity σx when the process is at x, upward drift of magnitude bx, and killing rate $\kappa(x) = \mu x$. We will discuss this variant in section 4.3.

The concept of quasistationary distribution is still relevant to this model, even though it progresses always in the same direction, hence is not irreducible. Conditioned on survival, the process will indeed converge to a fixed quasistationary distribution, from any starting state. The difference from the standard setting is, that the distribution will depend on the starting state. We can analyze the model as follows: Assume first that X_0 is 1. Let $p = \mu/(\mu + \lambda)$ and $\alpha = \mu + \lambda$. If the particle is at state x, the probability that it dies before moving on to $x + 1$ is p. Thus, the final senescence state X_ω of the particle has geometric distribution with ratio $q = 1 - p$. Conditioned on $X_\omega = k$, the time of death T may be represented as $T = \tau_1 + \tau_2 + \cdots + \tau_k$, where τ_i is exponentially distributed with parameter $i\alpha$, and all are independent. We have then

$$\mathrm{P}\big\{T > t \,\big|\, X_\omega = k\big\} = 1 - \big\{1 - e^{-\alpha t}\big\}^k.$$

(To see this, observe that the sum of exponential variables with parameters $1, \ldots, k$ has the same distribution as the maximum of k exponential variables with parameter 1.) This yields

$$\mathrm{P}\big\{T > t\big\} = 1 - \sum_{k=1}^{\infty} pq^{k-1}\left(1 - e^{-\alpha t}\right)^k = \frac{1}{q + pe^{\alpha t}} = \frac{\mu + \lambda}{\lambda + \mu e^{(\mu+\lambda)t}}.$$

The hazard rate is then

$$\frac{\mu(\mu + \lambda)e^{(\mu+\lambda)t}}{\lambda + \mu e^{(\mu+\lambda)t}}.$$

The rate is approximately exponential early on, and converges to $\mu + \lambda$ as $t \to \infty$. (That the rate begins as an exponential is hardly surprising. If we remove all randomness from the motion, and simply have it remain at site x for a deterministic time $1/\lambda x$, the position at time t will be approximately a constant times $e^{\lambda t}$. The hazard rate at time t will be μ times the position.)

Suppose, now, we know that an individual has survived to time t, where t is very large. What is the distribution of the individual's senescence state? It is important to notice that this is not the same as the distribution of X_t if killing were eliminated. Conditioned on survival to time t, X_t will be smaller than it would be without killing, since survival is more likely lower down. Let T_k be the time when X_t leaves state k (or ∞ if it never reaches k). Then

$$\begin{aligned}\mathrm{P}\big\{X_t = k \,\big|\, T > t\big\} &= \frac{\mathrm{P}\big\{T_{k-1} < t < T_k \text{ and } X_\omega \geq k\big\}}{\mathrm{P}\big\{T > t\big\}} \\ &= \left(\left(1 - e^{-\alpha t}\right)^{k-1} - \left(1 - e^{-\alpha t}\right)^k\right) q^{k-1}\left(q + pe^{\alpha t}\right).\end{aligned}$$

The distribution of X_t conditioned on survival is geometric with parameter $q(1-e^{-\alpha t})$. As t goes to infinity, this settles into a quasistationary distribution which is geometric with parameter q: coincidentally, the same as the distribution of the state at death, X_ω.

For a general starting distribution, the limit distribution is simply the same as for $X_0 = 1$, conditioned on being no smaller than X_0. This is

$$\mathrm{P}\big\{X_t = k \,\big|\, T > t\big\} = \mathrm{E}\big[pq^{k-X_0}\mathbf{1}_{X_0\le k}\big] = pq^k \sum_{j=1}^{k} q^{-j}\,\mathrm{P}\big\{X_0 = j\big\}. \tag{26}$$

4.3. **A continuous version of Le Bras' process.** The defining feature of Le Bras' model is that the rate of the random motion, as well as the killing rate, increase linearly with the state. We keep these features, but allow the process to move up and down, in order to arrive at a continuous process (which must be a diffusion).

We define a process on $[1,\infty)$ by the stochastic differential equation

$$dX_t = \sigma X_t dW_t + bX_t dt, \tag{27}$$

where σ is a positive constant and b is a constant larger than $\sigma^2/2$. The process starts at $X_0 = 1$, is killed at the rate kX_t, and is reflected when it hits 1. By Ito's formula we see that X_t is geometric Brownian motion, and can be written as

$$X_t = \exp\left\{\sigma(W_t + b't)\right\},$$

where $b' = \frac{b}{\sigma} - \frac{\sigma}{2}$. Equivalently, then, we could consider the Brownian motion with drift:

$$Y_t = W_t + b't,$$

killed at a rate $ke^{\sigma y}$ and reflected at 0. This will have the same mortality distribution as X_t. This makes it trivial to see that if σk is small, then for intermediate times t, such that

$$\frac{b'}{\sigma k} \gg e^{b't}$$

we will have $Y_t \approx b't/\sigma$, so the killing rate will be about $k\exp\{b't\}$.

In the notation of section 3.2 we have $B(x) = b'x$, where b' is positive. The left boundary 0 is regular, while the right boundary ∞ is natural. Since $\kappa(x) \to \infty$ as $x \to \infty$, we can apply Theorem 7, as long as $\underline{\lambda}$ is finite. This guarantees the convergence to a quasistationary distribution.

The details may be found in section 3 of [SE03]. The result is that the process conditioned on survival far a long time converges to the quasistationary distribution

$$\phi(x) = \frac{x^{\frac{b}{\sigma^2}-\frac{3}{2}} K_{i\tilde{y}}\left(\frac{\sqrt{8kx}}{\sigma}\right)}{\int_0^\infty x^{\frac{b}{\sigma^2}-\frac{3}{2}} K_{i\tilde{y}}\left(\frac{\sqrt{8kx}}{\sigma}\right) dx},$$

where K is the modified Bessel function, and $\tilde{y}$ is the smallest y such that $K'_{iy}(\sqrt{8k}/\sigma) = 0$. This function ϕ behaves asymptotically as

$$x^{\frac{b}{\sigma^2}-2} e^{-\sqrt{8kx}/\sigma}$$

as $x \to \infty$. The asymptotic killing rate is

$$\frac{\sigma^2}{8}\left[\left(\frac{2b}{\sigma^2}-1\right)^2 + \tilde{y}^2\right].$$

4.4. **Multidimensional fitness.** An interesting variant of the model proposed by Anderson and Weitz-Fraser, would view fitness as having several — perhaps many — components $(X_1(t), \ldots, X_n(t))$, all carrying out independent Brownian motions. We think of 0 as representing the optimum, and we represent total senescence by a continuous function $f(x_1, \ldots, x_n)$ which is taken to be increasing in $|x_i|$ for each i; we assume, as well, that $\{(x_1, \ldots, x_n) : f(x_1, \ldots, x_n) \le K\}$ is bounded, for each K. The process is killed at a rate $\kappa(f(x_1, \ldots, x_n))$ when it is at $(x_1, \ldots, x_n)$, but there is also a maximum senescence K, such that the process is killed as soon as its senescence reaches K. Since the state space of the process is compact, we may infer from Theorem 9 that the killing rate and the distribution of the fitness conditioned on survival converges to a limit given by the solution to the eigenvalue problem (23).

In general, explicit solutions will be difficult (though numeric solutions could be computed with standard algorithms). One case which we can solve in closed form, though, is that where $f(x_1, \ldots, x_n) = x_1^2 + \cdots + x_n^2$, with $n \ge 3$, and the killing occurs only when f reaches a fixed magnitude K^2. What makes this case so straightforward is the fact that the magnitude of the n-dimensional Brownian motion is itself a well-explored one-dimensional Markov process, known as the Bessel process of order ν, where $\nu = n/2 - 1$. The generator is

$$\mathcal{L}\phi(x) = \frac{1}{2}\phi''(x) + \left(\nu + \frac{1}{2}\right)x^{-1}\phi'(x). \tag{28}$$

We may either apply Theorem 6, with the inaccessible (entrance) boundary at 0 and the regular boundary at 1

The adjoint eigenvalue problem is

$$\frac{1}{2}\phi''(x) + \left(\nu + \frac{1}{2}\right)\left[x^{-1}\phi(x)\right]' + \lambda\phi(x) = 0,$$

for $x \in (0, K)$, with $\phi(K) = 0$ and $\phi'(0) = 0$. It has the general solution

$$x^{\nu+1}\left(c_1 J_\nu(\sqrt{2\lambda}x) + c_2 Y_\nu(\sqrt{2\lambda}x)\right), \tag{29}$$

in terms of the Bessel functions J_ν and Y_ν. Since $\nu > 0$, and $J_\nu(z) \sim (\frac{1}{2}z)^\nu/\Gamma(\nu+1)$ (by 9.1.7 of [AS65]), we see that $x^{\nu+1}J_\nu(\sqrt{2\lambda}x)$ has derivative 0 at $x = 0$. On the other hand, by 9.1.9 of [AS65] we see that $x^{\nu+1}Y_\nu(\sqrt{2\lambda}x)$ behaves asymptotically like a nonzero constant times x near 0, so it has a nonzero derivative at 0. Consequently, our solution to the boundary-value problem must have the form

$$\phi(x) = cx^{\nu+1}J_\nu(\sqrt{2\lambda}x).$$

The other boundary value is $\phi(K) = 0$. This means that $\sqrt{2\lambda}K$ is a zero of J_ν. The function J_ν will be positive on $(0, K)$ only if it is the smallest positive zero, conventionally denoted $j_{\nu,1}$. (In general, the i-th positive zero is denoted $j_{\nu,i}$. Thus, the limiting rate of mortality for this process is

$$\underline{\lambda} = \frac{j_{\nu,1}^2}{2K^2},$$

and the senescence states of those individuals who survive until time t converges to

$$cx^{\nu+1}J_\nu\big(j_{\nu,1}x/K\big) \tag{30}$$

as $t \to \infty$.

In fact, the eigenvalue expansion may be continued to give the exact distribution of the time when the Bessel process of order ν hits K. This is given by formula 6.2.0.2 of [BS96]:

$$\mathrm{P}_x\big\{T \in dt\big\} = \sum_{i=1}^{\infty} \frac{j_{\nu,i}x^{-\nu}J_\nu(j_{\nu,i}x/K)K^{\nu-2}}{J_{\nu+1}(j_{\nu,i})} e^{-j_{\nu,i}^2 t/2K^2} dt. \tag{31}$$

When x (the starting point) is 0, this simplifies slightly to

$$\mathrm{P}_0\big\{T \in dt\big\} = \sum_{i=1}^{\infty} \frac{(j_{\nu,i})^{\nu+1}}{\Gamma(\nu+1)2^\nu K^2 J_{\nu+1}(j_{\nu,i})} e^{-j_{\nu,i}^2 t/2K^2} dt. \tag{32}$$

4.5. **The series-parallel model.** Another discrete-space model that has been advanced by Gavrilov and Gavrilova in [GG91] and [GG01] represents the organism as an assemblage of independent "elements" in each of k independent "organ systems". A system fails when all of its components has failed, but the organism dies as soon as any one of its organs fails. The failure times of the components are independent exponential random variables with expectation $1/\lambda$. (The particular assumptions that Gavrilov and Gavrilova impose on the starting condition we defer to section 5.1.)

We represent this model as a Markov process with state space k-tuples of nonnegative integers. If the process is in state $(i_1, i_2, \ldots, i_r, \ldots, i_k)$, with all the i_j

positive, the rate of transition to state $(i_1, i_2, \ldots, i_r - 1, \ldots, i_k)$ is $i_r\lambda$, for each $1 \le r \le k$.

Let T be the time of death, and let T_i be the time when component i first reaches 0. Then $T = \min T_i$. Each T_i is the sum of exponential waiting times with rates $\lambda, 2\lambda, \ldots, x_i\lambda$, where x_i is the initial number of elements in system i. T_i has the same distribution as the maximum of x_i independent exponential waiting times with rate λ, so

$$\mathrm{P}\{T_i > t\} = 1 - \left(1 - e^{-\lambda t}\right)^{x_i}.$$

This gives us the distribution

$$\mathrm{P}\{T > t\} = \prod_{i=1}^{k} \left(1 - \left(1 - e^{-\lambda t}\right)^{x_i}\right).$$

The hazard rate is then given by the logarithmic derivative with respect to t:

$$h(t) = \sum_{i=1}^{k} \frac{\lambda e^{-\lambda t} x_i (1 - e^{-\lambda t})^{x_i - 1}}{1 - \left(1 - e^{-\lambda t}\right)^{x_i}} \tag{33}$$

$$= kx \frac{\lambda e^{-\lambda t}(1 - e^{-\lambda t})^{x-1}}{1 - \left(1 - e^{-\lambda t}\right)^{x}} \text{ when all } x_i \text{ have the same value } x. \tag{34}$$

(It is perhaps worth noting that, although great emphasis is placed on the two-stage structure of this model — multiple serial systems, each comprised of multiple parallel components — the serial systems have no effect on the hazard rate, but to multiply it by a constant.)

It follows from (34) that

$$\lim_{t \to \infty} h(t) = k\lambda. \tag{35}$$

This may also be understood in terms of a quasistationary distribution. Conditioned on a given system surviving up to time t, as t becomes very large, the number of components surviving in the system converges to 1 with high probability. Since the systems are independent, we see that the distribution of the total state converges to one wholly concentrated at $(1, 1, \ldots, 1)$, the state in which all systems have one surviving component. From that state, clearly, the hazard rate is $k\lambda$.

5. Initial distributions

When a mathematical model succeeds in reproducing target features of the empirical data, it is natural to suppose that the model is correct in some significant sense. But mortality models are generally too vague, and leave too much latitude to arbitrary manipulation of parameters, to assign much significance to the resulting mortality distributions. Moreover, researchers typically aim to match a particular

formula only in a part of the life course, and this allows for arbitrary definition of "small t" or "large t". In one case, the Brownian motion with drift, we show how a large class of mortality-time distributions, including all mixtures of gamma distributions with certain bounds on their parameters, could be the exact outcome of this model, if the initial distribution were chosen appropriately. In addition, any distribution at all could appear for small t, to any desired degree of accuracy, if we allow the drift b to be increased *ad libitum*.

5.1. **Series-parallel model redux.** As we intend to suggest that the freedom to choose an initial distribution (or other arbitrary parameters of a model) could create the impression of a model magically matching the empirical facts, it will perhaps be useful to consider an example in which wishful thinking, combined with loose application of initial distributions, have contrived a Gompertz curve out of whole cloth. Here the computations which yield the desired hazard rates are wrong, and would presumably quickly have been recognized as such if the result were not otherwise so pleasing.

We return to the series-parallel model, which we described in section 4.5, and begin by explaining how its inventors derived Gompertz hazard rates in [GG91] and [GG01]. For t which are small in comparison with $1/\lambda$ — that is, times when a typical component should still be functioning — we have the approximation to first order in λt

$$h(t) \approx \sum_{i=1}^{k} x_i \lambda^{x_i} t^{x_i - 1} \tag{36}$$

$$= kx\lambda^x t^{x-1} \text{ when all } x_i \text{ have the same value } x. \tag{37}$$

It would be tempting to meld (37) and (35), and so to imagine a hazard rate which is initially Weibull — a power of t, that is — until it gradually flattens out at a mortality plateau. The temptation becomes even greater when Gavrilov and Gavrilova in [GG01] claim that by mixing these Weibull distributions with a Poisson starting state — that is, letting the number of components initially functioning in each system be independent Poisson random variables with common expectation μ — the early hazard rate looks like

$$h(t) \approx k \sum_{x=1}^{\infty} \frac{e^{-\mu}\mu^x}{x!} x\lambda^x t^{x-1} = k\lambda\mu e^{\mu(\lambda t - 1)}, \tag{38}$$

which is the coveted Gompertz hazard rate.

Unfortunately, this depends on two significant errors. The approximation which gives the Weibull hazard rate, is accurate only very close to $t = 0$, while the asymptote is only relevant for large t. For intermediate values of t, the hazard rate

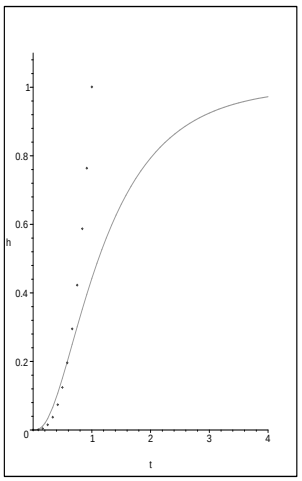

FIGURE 1. Weibull hazard rates (equation 36) with $\lambda = 1$, $k = 1$, $x = 4$ (solid curve), and the exact hazard rate for the series-parallel model (equation 33) with the same parameters (dots).

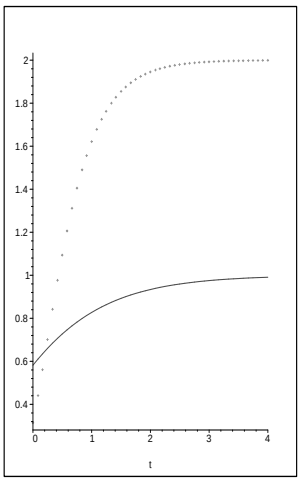

FIGURE 2. Hazard rate (equation 40) for the series-parallel process with $k = 1$, and $\lambda = 1$ (solid) or $\lambda = 2$ (dots), started with the initial number of components having a Poisson distribution with expectation $\mu = 1$ (solid) or $\mu = 3$ (dots), conditioned on at least one working component.

for this model is no more similar to the Weibull or Gompertz approximation than many another curve might be. Once we have started with the assumption that $e^{\lambda t}$ is indistinguishable from $1 + \lambda t$, we cannot hope to make fine distinctions among the various powers and exponentials of t. As an example, we plot in figure 1 the correct hazard rate for $\lambda = 1$, $k = 1$, $x = 4$, together with the curve $4t^3$ (the Weibull hazard rate for the same parameters). These curves are treated as interchangeable in the Gavrilovs's analysis, though they are clearly very different.

A more serious problem is the computation which underlies (38). A hazard rate is $h(t) = -F'(t)/F(t)$, where $F(t) = \mathrm{P}\{T > t\}$, and T is the failure time. If $F(t)$ depends on a random variable X (in this case, the number of initially functioning components in a system), with $\mathrm{P}\{T > t|X = x\} = F_x(t)$ and $\mathrm{P}\{X = x\} = p_x$, then $F(t) = \sum_x p_x F_x(t)$. We have then the hazard rate

$$h(t) = -\frac{\sum_x p_x F'_x(t)}{\sum_x p_x F_x(t)}.$$

On the other hand, the Gompertz hazard rate results from averaging the individual hazard rates

$$-\sum_x p_k \frac{F_k'(t)}{F_k(t)}.$$

If we start from the approximation (36), and mix with the Poisson distribution, we get

$$\frac{k}{1-e^{-\mu}} \sum_{x=1}^{\infty} \frac{e^{-\mu}\mu^x}{x!} x\lambda(\lambda t)^{x-1} = \frac{k\lambda\mu}{e^{\mu}-1} e^{\mu\lambda t}. \tag{39}$$

(The factor $(1-e^{-\mu})^{-1}$ comes from the fact that we condition on there being at least one working component at the beginning.)

As it happens, the exact hazard rate, when computed correctly, has a simple form. The distribution function is

$$\left[\sum_{x=1}^{\infty} \frac{e^{-\mu}\mu^x}{x!(1-e^{-\mu})} \left(1-\left(1-e^{-\lambda t}\right)^x\right)\right]^k = \left(\frac{1-e^{-\mu e^{-\lambda t}}}{1-e^{-\mu}}\right)^k.$$

Thus, the hazard rate becomes

$$\frac{k\mu\lambda e^{-\lambda t}}{e^{\mu e^{-\lambda t}}-1}. \tag{40}$$

This is very different from an exponentially increasing function. Examples are shown in figure 2. In particular, for μ smaller than about 1.8, the second derivative at $t=0$, given by the expression

$$\frac{\lambda^2\left(e^{2\,\mu}-2\,e^{\mu}+1-3\,\mu\,e^{2\,\mu}+3\,\mu\,e^{\mu}+\mu^2e^{2\,\mu}+\mu^2e^{\mu}\right)}{\left(e^{\mu}-1\right)^3}$$

is negative. Thus, the hazard rate is already concave at $t=0$ for these values of μ.

This is not to say that no version of this model can generate anything remotely like a Gompertz hazard rate. There are several parameters, offering ample scope for cherry-picking. In particular, if we make λ very small, say .005, this will have the effect of stretching the mortality curve out in time, creating the appearance of a fairly constant linear growth. This makes the hazard rate extremely small, so we can blow it up by choosing a very large value of k, such as 20000. We still have a very compressed hazard rate: it seems exponential for the first 75 years or so, but barely completes a single doubling in that time. This may be remedied by choosing μ somewhat larger: say 15. The result may be seen on the left of figure 3.

It might be argued that the model ought to be applied in exactly this range of parameters. However, as we have already pointed out, it is a frail defense for a model, to say that it can be compelled to approximate a line by skillful manipulation

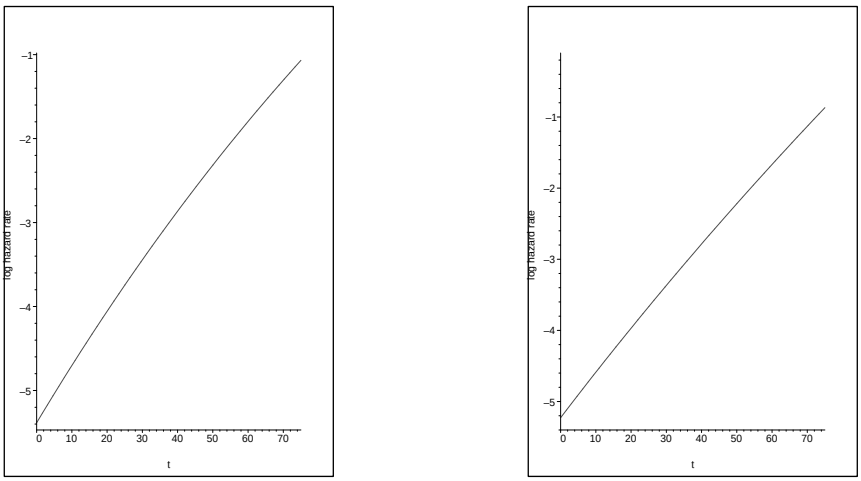

FIGURE 3. Left: Natural logarithm of the hazard rate (equation 40) for the series-parallel process with $k = 20000$, $\lambda = .005$, and $\mu = 15$, conditioned on at least one working component. Right: Natural logarithm of the Weibull hazard rate given in equation (41).

of multiple parameters. The same may be achieved with the Weibull model. For instance, if the hazard rate were given by

$$(.0025t + .77)^{20}, \tag{41}$$

we get the log hazard shown on the right of figure 3. Merely matching the coarse features of the hazard rate is well known to be only a weak success of a model. What was promised here — erroneously — was an analytic derivation of the Gompertz curve.

5.2. **Brownian motion with drift.** After introducing their model for arbitrary intitial distributions of fitness, Weitz and Fraser carry out their computations only for "homogeneous" populations; that is, with X_0 a deterministic constant. Anderson constrains his model similarly. In fact, though, there is no obvious justification for the starting condition to be deterministic. It turns out that the choice of starting distribution is enough to realize almost any hitting-time distribution, up to an error which can be made as small as you like, though at the expense of making the limiting mortality rate larger. The results in this section are all proved in section 6

To fix our conventions, we give the gamma density as

$$\gamma_{r,\beta}(x) = \Gamma(r)^{-1}\beta^r x^{r-1}e^{-\beta x},$$

where the shape parameter $r \geq -1$ and the scale parameter β is any positive number. We remind the reader that a gamma random variable has expectation r/β and variance r/β^2. For $r = 1$, this is the exponential distribution, while large values of r converge to the normal distribution.

If μ and μ^* are probability distributions on $\mathbb{R}$, we define the distance between them to be

$$d(\mu,\mu^*) = \inf_{\substack{X\sim\mu \\ X^*\sim\mu^*}} \mathrm{E}\big|X - X^*\big| = \int_0^\infty \big|F(t) - F^*(t)\big|dt, \tag{42}$$

where F and F^* are the corresponding distribution functions. (The infimum is taken over X and X^* with distribution μ and μ^* respectively.) This means (see [SW86]) that μ and μ^* are close if $\int f(x)\mu(dx)$ and $\int f(x)\mu^*(dx)$ are close, whenever f is a function which does not vary too rapidly (for example, $|f'| \leq 1$).

If μ is a probability distribution on $[0,\infty)$, σ a positive constant, and b a function, we denote by $\nu_{\sigma,b}(\mu)$ the distribution of the time when a Brownian motion started in the distribution μ, with drift b and diffusion constant σ, first hits 0. We will also write $\nu_{\sigma,b}(g)$ where g is a probability density. When σ is dropped from the notation, it is implicitly assumed to be 1. We will call a distribution ν on $\mathbb{R}^+$ *b-attainable* if there is a distribution μ such that $\nu = \nu_b(\mu)$. It is *ϵ-approximately b-attainable* if there is a distribution μ such that $d(\nu, \nu_b(\mu)) < \epsilon$.

Note that the time of killing at 0 is unchanged by a linear rescaling of the space. This implies that $\nu_{\sigma,b}(\mu) = \nu_{b/\sigma}(\mu)$.

Our first result tells us that any target mortality-time distribution ν is ϵ-approximately b/σ-attainable when σ is a sufficiently small rescaling factor.

Theorem 10. *Suppose that ∞ is a natural boundary for the process. Assume, in addition, that b is Lipschitz, bounded away from zero, and sublinear; that is, there are positive constants α_1 and α_2 such that*

$$\alpha_1 \leq -b(x) \leq \alpha_2(1+x), \tag{43}$$

and a constant L such that

$$\big|b(x) - b(y)\big| \leq L|x-y| \qquad \text{for all } x, y \in \mathbb{R}^+. \tag{44}$$

Then for any probability distribution ν on $\mathbb{R}^+$ such that

$$\int_0^\infty e^{\alpha_2 z}\sqrt{z}\nu(dz) < \infty, \tag{45}$$

and any positive ϵ, for every σ sufficiently small there exists a distribution μ on $\mathbb{R}^+$ such that $d(\nu, \nu_{b/\sigma}(\mu)) \leq \epsilon$.

In particular, any potential exit-time distribution may be approximated from some initial distribution, as long as the constant drift is made sufficiently large. The rest of the results are specific to the case of constant drift.

We remind the reader of a few facts about Laplace transforms. If μ is a probability distribution on $\mathbb{R}^+$, its Laplace transform is the function

$$f(s) = \int_0^\infty e^{-sz}\mu(dz).$$

have the same Laplace transform.

Theorem 11. *Let b be a negative constant, and let ν be a probability distribution on $\mathbb{R}^+$, with Laplace transform f. Then ν is b-representable if and only if*

$$\phi(s) := f\left(\frac{s^2}{2} - bs\right)$$

is the Laplace transform of some distribution μ. This is equivalent to

$$\sum_{j=0}^{\lfloor n/2 \rfloor} (-1)^j \binom{n-j}{j} 4^{-j} \left(1 - \frac{b^2 - 2r}{(-b+s)^2}\right)^j \xi_{n-j}\left(\frac{s^2}{2} - bs\right) > 0 \qquad \forall s \in \mathbb{R}^+,\ n \in \mathbb{Z}^+, \tag{46}$$

where

$$\xi_k(s) := (-1)^k \frac{(s+r)^k}{k!} f^{(k)}(s) > 0.$$

In that case, μ is the desired starting distribution.

This allows us to show, by direct computation, that a given distribution is not b-attainable. For example, if we consider a logistic Gompertz distribution with $\theta = r = a = 1$, since the asymptotic hazard rate is 1, it could, in principle, be b-attainable for any $b \leq -\sqrt{2}$. But it is straightforward to compute that (46) is negative for $b = -\sqrt{2}$, $s = 0$, $n = 3$. If we let $b = -2$, we get a negative sum for $s = 0$ when $n = 19$.

On the other hand, Theorem 11 also allows us to show that a large class of distributions is b-attainable.

Corollary 12. *Let b be a fixed negative number. Any convex combination of gamma distributions with scale parameter $\beta \leq b^2/2$ and arbitrary shape parameters, is b-attainable.*

In the constant-drift setting, we can give a simple bound on the error of approximating a given distribution.

Corollary 13. *Let μ be any distribution on $\mathbb{R}^+$. For any fixed negative b there exists a probability distribution μ^* on $\mathbb{R}^+$ such that*

$$\int_0^\infty x\nu_b(\mu^*)(dx) = \int_0^\infty x\mu(dx) \tag{47}$$

(that is, the two distributions define stopping times with the same expectation) and

$$d\big(\mu, \nu_b(\mu^*)\big) \le |b|^{-1}\sqrt{2}\int_0^\infty x^{1/2}\mu(dx). \tag{48}$$

As long as the drift is reasonably well behaved — in particular, if the drift is constant — one could go even further, to show that the densities and the hazard rates may also be made to converge to target functions. The details are not very enlightening, though, so we leave them out. In addition, it may reasonably be criticized, that this convergence as $b \to \infty$ is somewhat trivial, since it essentially just means that as b grows large, the process behaves more and more like a deterministic process. The target mortality is programmed into the distribution which slides down the axis, with minimal stochastic perturbation. This criticism does not affect Corollary 12, which tells us that a large class of killing-time distributions may be attained with fixed drift.

6. Proofs of results from section 5.2

6.1. **Proof of Theorem 10.** Let

$$P(x) = \int_0^x \frac{dy}{b(y)}\,,$$

μ^* the push-forward of ν by P; that is, for any $A \subset \mathbb{R}^+$, $\mu^*(A) = \nu\big(P^{-1}(A)\big)$. Let W_t be a single Brownian motion, and define $X_t^{(\sigma)}$ to be the strong solution to the SDE

$$dX_t = \sigma dW_t + b(X_t)dt$$

started in the distribution μ^*. The assumption that ∞ is a natural boundary guarantees the existence of a strong solution[KS88, Theorem 2.9].

Let $\tau^{(\sigma)}$ be the first time when $X_t^{(\sigma)}$ hits 0. Note that for $\sigma > 0$, the distribution of $\tau^{(\sigma)}$ is the same as the one we have denoted $\nu_{b/\sigma}(\mu^*)$. The only randomness in $X^{(0)}$ is the starting point, and it is easy to see that $\tau^{(0)}$ has the distribution ν. We need only to show that

$$\lim_{\sigma\to 0} d\big(\tau^{(\sigma)}, \tau^{(0)}\big) = 0.$$

where the distance between the random variables is understood to be the distance between their distributions. But for this it will suffice to show that

$$\lim_{\sigma\to 0} \mathrm{E}\big|\tau^{(\sigma)} - \tau^{(0)}\big| = 0. \tag{49}$$

Fix $\sigma > 0$, and let

$$Z_t = X_t^{(\sigma)} - X_t^{(0)}.$$

Also, let Z_t^* be a strong solution to

$$dZ_t^* = \sigma dW_t + \alpha_2\big(\big|Z_t^*\big| + 1\big)dt,$$

with initial condition $Z_0^* = Z_0$. Then Z^* and Z have the same diffusion term, and (by the Lipschitz condition) the drift of Z^* is greater. Thus, by the comparison theorem [KS88, Proposition 2.18] we see that $Z_t^* \geq Z_t$ for all $0 \leq t \leq T$ almost surely. By Gronwall's Lemma [DZ93, Lemma E.6], it follows that

$$\zeta := \sup\big\{Z_t : 0 \leq t \leq \tau^{(0)}\big\} \leq \sigma e^{\alpha_2 \tau^{(0)}} \cdot \sup_{0\leq t\leq \tau^{(0)}} \big|W_t\big|. \tag{50}$$

Since $X^{(\sigma)}$ is a diffusion with negative drift at least α_1, and since $X^{(\sigma)}_{\tau^{(0)}} \leq \zeta$, it must be that

$$\mathrm{E}\big[\big(\tau^{(\sigma)} - \tau^{(0)}\big)^+ \,\big|\, \zeta\big] \leq \frac{\zeta}{\alpha_1}.$$

(That is, a diffusion with drift $-\alpha$ has an expected time of arrival at 0 no more than α^{-1} times its starting point.) Thus the expected overhang of $\tau^{(\sigma)}$ over $\tau^{(0)}$ is no more than α_1^{-1} times its distance from 0 when $X^{(0)}$ hits. Similarly, if $\tau^{(\sigma)} \leq \tau^{(0)}$, it must be that $X^{(0)}_{\tau^{(\sigma)}} \leq \zeta$. The difference between the hitting times is bounded by

$$\mathrm{E}\big|\tau^{(\sigma)} - \tau^{(0)}\big| \leq \frac{2\,\mathrm{E}\,\zeta}{\alpha_1}.$$

By (50), then, and using the reflection principle (see, for example, [KS88]) to see that

$$\mathrm{E}\big[\sup_{0\leq t\leq s} |W_t|\big] = 2\,\mathrm{E}\big[|W_s|\big] = 4\sqrt{s}/\sqrt{2\pi},$$

we see that

$$\mathrm{E}\big|\tau^{(\sigma)} - \tau^{(0)}\big| \leq \frac{4\sigma}{\alpha_1}\,\mathrm{E}\Big[\sqrt{\tau^{(0)}}e^{\alpha_2\tau^{(0)}}\Big].$$

By assumption (45) this is finite, and goes to 0 with σ.

6.2. **Proof of Theorem 11.** Suppose μ is a starting distribution such that $\nu_b(\mu) = \nu$, and call its Laplace transform g. For positive x, let τ_x be the time when a Brownian motion with drift $b < 0$, started at x, first hits 0. Define for positive α,

$$\zeta_\alpha(x) := \mathrm{E}\big[e^{-\alpha\tau_x}\big].$$

By the Feynman-Kac formula [RS90, 8.10b], ζ satisfies

$$\frac{1}{2}\zeta_\alpha'' + b\zeta_\alpha' = \alpha\zeta_\alpha, \tag{51}$$

with boundary conditions $\zeta_\alpha(0) = 1$ and $\zeta_\alpha(\infty) = 0$. This yields

$$\zeta_\alpha(x) = \exp\big\{(-b - \sqrt{b^2 + 2\alpha})x\big\}.$$

Since the hitting time from a start in distribution μ has distribution ν, it must be that

$$f(\alpha) = \int_0^\infty \zeta_\alpha(x)\mu(dx) = g\big(b + \sqrt{b^2 + 2\alpha}\big).$$

Thus for s positive,

$$\phi(s) = f\left(\frac{s^2}{2} - bs\right) = g(s),$$

which is the Laplace transform of μ.

Now suppose that ϕ is the Laplace transform of a distribution μ. By the above computation, if μ is the initial distribution, and we let $\psi(\alpha)$ be the Laplace transform of the time of first hitting 0, we have

$$\psi(\alpha) = \phi\big(b + \sqrt{b^2 + 2\alpha}\big) = f(\alpha).$$

Since the hitting time has the same Laplace transform as ν, they are the same distribution.

By Theorem XIII.1.4. of [Fel71], ϕ is the Laplace transform of some probability distribution if and only if $\lim_{z\to 0}\phi(z) = 1$ and *phi* is totally monotone (meaning that $(-1)^n\phi^{(n)}(s)$ is positive for all n and s; here $\phi^{(n)}$ is the n-th derivative of ϕ.) It is straightforward to show by induction that

$$\phi^{(n)}(s) = \sum_{k=\lceil n/2\rceil}^{n} \frac{n!}{2^{n-k}(2k-n)!(n-k)!}(s-b)^{2k-n} f^{(k)}\left(\frac{s^2}{2} - bs\right).$$

A change of variables $j = n - k$ shows then that the sum in (46) is precisely

$$(-1)^n(n!)^{-1}\left(\frac{s^2}{2} - bs + r\right)^n (-b+s)^{-n}\phi^{(n)}(s),$$

whose positivity for all s and n is the definition of ϕ being totally monotone, which is equivalent to ϕ being the Laplace transform of some distribution.

6.3. **Proof of Corollary 12.** We begin by noting that any convex combination of b-attainable distributions is also b-attainable, simply by starting from the corresponding convex combination of the starting distributions. Thus, we need only show that all gamma distributions with exponential rate $\beta \le b^2/2$ are b-attainable. (It is also worth noting that any convolution of b-attainable distributions is also b-attainable.)

The gamma distribution with parameters (r, β) has Laplace transform

$$f(s) = \beta^r\,(\beta + s)^{-r}$$

Thus

$$f\left(\frac{s^2}{2} - bs\right) = \beta^r \left(\beta + \frac{s^2}{2} - bs\right)^{-r}$$
$$= (2\beta)^r \left(s - b + \sqrt{b^2 - 2\beta}\right)^{-r} \left(s - b - \sqrt{b^2 - 2\beta}\right)^{-r}$$

For $\beta \le b^2/2$, the roots are real, so each of the two factors is the Laplace transform of a gamma distribution. The product of two Laplace transforms is itself a Laplace transform (of the convolution of the two distributions), which completes the proof.

6.4. **Proof of Corollary 13.** Let $\beta = \sqrt{2|b|}$. Define the kernel

$$K_\beta(x, y) = \gamma_{\beta x, x}(y), \tag{52}$$

and let

$$g(y) := \int_0^\infty K_\beta(x, y)\mu(dx). \tag{53}$$

Observe first that g is a probability density, since it is nonnegative and

$$\int_0^\infty g(y)dy = \int_0^\infty \int_0^\infty K_\beta(x, y)\mu(dx)dy = 1$$

by changing the order of integration. Similarly, the expectation of μ^* is

$$\int_0^\infty yg(y)dy = \int_0^\infty \int_0^\infty yK_\beta(x, y)\mu(dx)dy$$
$$= \int_0^\infty \left(\int_0^\infty yK_\beta(x, y)dy\right)\mu(dx)$$
$$= \int_0^\infty x\mu(dx)$$
$$= \|\mu\|_1.$$

Since g is a mixture of gamma densities with scale $\beta = b^2/2$, we know from Corollary 12 that a distribution μ^* exists such that $\nu_b(\mu^*)$ has density g. It only remains then to show that this distribution does indeed have the right distance from μ.

Let X be a random variable with distribution μ. Conditioned on X, let Y be a random variable with a gamma distribution, with parameters $(\beta X, X)$. Observe that, conditioned on X, the random variable $(X-Y)$ has expectation 0 and standard deviation $\sqrt{x/\beta}$. Then Y has density $g(y)$ and

$$\mathrm{E}\big[|X - Y|\big] = \mathrm{E}\big[\mathrm{E}[|X - Y| \;\big|\; X]\big]$$
$$\le \mathrm{E}\big[\sqrt{X/\beta}\big].$$

Acknowledgement

The authors would like to thank Persi Diaconis, Hunter Fraser, Jim Pitman, and Ken Wachter, who all proferred helpful comments, advice, and especially guidance to the literature in the various fields touched on by this paper.

References

[And00] James J. Anderson. A vitality-based model relating stressors and environmental properties to organism survival. *Ecological monographs*, 70(3):445–70, 2000.

[AS65] Milton Abramowitz and Irene Stegun. *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*. Dover, New York, 1965.

[Bai00] William Bains. Statistical mechanic prediction of non-gompertzian ageing in extremely aged populations. *Mechanisms of Ageing and Development*, 112:89–97, 2000.

[Bra76] H. Le Bras. Lois de mortalité et age limite. *Population*, 33(3):655–91, May-June 1976.

[BS96] Andrei N. Borodin and Paavo Salminen. *Handbook of Brownian Motion — Facts and Formulae*. Birkhäuser, Basel, 1996.

[CF77] R. S. Chhikara and J. L. Folks. The inverse gaussian distribution as a lifetime model. *Technometrics*, 19(4):461—8, November 1977.

[Cha01] Brian Charlesworth. Patterns of age-specific means and genetic variances of mortality rates predicted by the mutation-accumulation theory of ageing. *Journal of Theoretical Biology*, 210(1):47–65, 2001.

[Doo91] Erik A. Van Doorn. Quasistationary distributions and convergence to quasi-stationarity of birth-death processes. *Advances in Applied Probability*, 23:683–700, 1991.

[DS65] J. N. Darroch and E. Seneta. On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *Journal of Applied Probability*, 2:88–100, 1965.

[DV76] M. D. Donsker and S. R. S. Varadhan. Second order elliptic differential operators. *Communications on Pure and Applied Mathematics*, 29:595–621, 1976.

[DZ93] A. Dembo and O. Zeitouni. *Large Deviation Techniques and Applications*. Jones and Bartlett, Boston, 1993.

[Fel52] William Feller. The parabolic differential equations and the associated semi-groups of transformations. *Annals of Mathematics*, 55:468–519, 1952.

[Fel71] William Feller. *An Introduction to Probability and its Applications*, volume 2. John Wiley & Sons, New York, 1971.

[FK00] Caleb E. Finch and Thomas B. L. Kirkwood. *Chance, Development, and Aging*. Oxford University Press, New York, 2000.

[GG91] Leonid A. Gavrilov and Natalia S. Gavrilova. *The biology of lifespan: a quantitative approach*. Harwood Academic Publishers, Chur, Switzerland, 1991.

[GG01] Leonid A. Gavrilov and Natalia S. Gavrilova. The reliability theory of aging and longevity. *Journal of Theoretical Biology*, 213:527–45, 2001.

[GQZ88] Guanglu Gong, Minping Qian, and Zhongxin Zhao. Killed diffusions and their conditioning. *Probability Theory and Related Fields*, 80:151–67, 1988.

[HJ85] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1985.

[KS88] Ioannis Karatzas and Steven Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, Heidelberg, Berlin, 1988.

[Pin85] Ross G. Pinsky. On the convergence of diffusion processes conditioned to remain in a bounded region for large time to limiting positive recurrent diffusion processes. *The Annals of Probability*, 1985.

[RS90] Luigi M. Ricciardi and Shunsuke Sato. Diffusion processes and first-passage-time problems. In *Lectures in applied mathematics and informatics*, pages 206–285. Manchester Univ. Press, Manchester, 1990.

[RW00a] L. C. G. Rogers and David Williams. *Diffusions, markov processes and martingales*, volume 1. Cambridge University Press, Cambridge, 2nd edition, 2000.

[RW00b] L. C. G. Rogers and David Williams. *Diffusions, markov processes and martingales: Itô calculus*, volume 2. Cambridge University Press, Cambridge, 2nd edition, 2000.

[RY90] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*. Springer-Verlag, New York, Heidelberg, Berlin, first edition, 1990.

[SE03] David Steinsaltz and Steven N. Evans. Quasistationary distributions for one-dimensional diffusions with killing. Technical Report 637, U. C. Berkeley, Department of Statistics, March 2003.

[Sen73] E. Seneta. *Non-Negative Matrices*. Wiley, New York, 1973.

[SVJ66] E. Seneta and D. Vere-Jones. On quasi-stationary distributions in discrete time for Markov chains with a denumerable infinity of states. *Journal of Applied Probability*, 3:403–34, 1966.

[SW86] Galen Shorack and Jon Wellner. *Empirical processes with applications to statistics*. Wiley, New York, 1986.

[TT79] Pekka Tuominen and Richard L. Tweedie. Exponential decay and ergodicity of general Markov processes and their discrete skeletons. *Adv. in Appl. Probab.*, 11(4):784–803, 1979.

[Twe74] Richard Tweedie. r-theory for Markov chains on a general state space I: solidarity properties and r-recurrent chains. *The Annals of Probability*, 2:840–64, 1974.

[Vau97] James W. Vaupel. Trajectories of mortality at advanced ages. In Kenneth W. Wachter and Caleb E. Finch, editors, *Between Zeus and the Salmon: The Biodemography of Longevity*, chapter 2, pages 17–37. National Academies Press, Washington, D.C., 1997.

[Vau98] James W. Vaupel *et al.* Biodemographic trajectories of longevity. *Science*, 280(5365):855–60, May 1998.

[WF01] Joshua Weitz and Hunter Fraser. Explaining mortality rate plateaus. *Proc. Natl. Acad. Sci. USA*, 98(26):15383–6, December 18 2001.

David Steinsaltz, Department of Demography, University of California, 2232 Piedmont Ave., Berkeley, CA 94720-2120, Steven N. Evans, Department of Statistics #3860, 367 Evans Hall, University of California, Berkeley, CA 94720-3860

E-mail address: dstein@demog.berkeley.edu//evans@demog.berkeley.edu