# A GENERALIZED MODEL OF MUTATION-SELECTION BALANCE WITH APPLICATIONS TO AGING

DAVID STEINSALTZ+, STEVEN N. EVANS, AND KENNETH W. WACHTER

Abstract. A probability model is presented for the dynamics of mutation-selection balance in a haploid infinite-population infinite-sites setting sufficiently general to cover mutation-driven changes in full age-specific demographic schedules. The model accommodates epistatic as well as additive selective costs. Closed form characterizations are obtained for solutions in finite time, along with proofs of convergence to stationary distributions and a proof of the uniqueness of solutions in a restricted case. Examples are given of applications to the biodemography of aging.

## 1. Introduction

Arguments from the mathematical genetics of mutation-selection balance figure broadly in evolutionary theories of senescence. Available formal models, however, do not cover cases brought to the fore by recent progress in biodemography [1]. In this paper, we present a rigorous general model encompassing these cases, prove results concerning existence, uniqueness, and convergence, obtain closed-form representations for solutions to the model, and give examples of its application to questions in the demography of aging.

The whole mathematical theory of natural selection may be divided into three parts: positive mutations, neutral mutations, and deleterious mutations. Positive mutations may be thought to add up to an optimal adaptation, at least under some conditions, and they are generally studied in that context by demographers. Neutral mutations have their primary effects in alleles which drift randomly to fixation. Deleterious mutations, the focal subject for theories of aging and for this paper, are expected never to achieve fixation in populations, except, through founder effects, in very small populations. Their influence in large populations derives from their persistent reintroduction and slow meander to extinction.

Sir Peter Medawar [2], in 1952, descried an explanation for senescence in the accumulation of deleterious alleles with age-specific effects, given the declining force of natural selection with adult age. W. D. Hamilton [3] presented expressions for this declining age-specific force, helping others quantify the resulting balance

+ To whom correspondence should be addressed at Department of Demography, University of California, 2232 Piedmont Avenue, Berkeley, CA. 94720-2120; email dstein@demog.berkeley.edu.

between mutation and selection. B. Charlesworth [4] analyzed the dynamics of age-specific selection. His work guides the thinking of many experimentalists.

At stake are the cumulative effects of numerous mildly deleterious mutations showing up at some large collection of loci. In our setting, the genotypes determine full age-specific schedules of mortality and fertility, and the effects of a mutation have to be represented as a perturbation of a whole function of age. A rigorous treatment demands that mutations correspond to points in abstract spaces, such as function spaces. Relationships between our work and the large literature on mutation and selection reviewed by Bürger [5] are discussed in Section 8.

Up to now, researchers have relied on linear approximations to cost functions and restricted their representations of the age-specific effects of mutations to stylized patterns like step-functions. Intriguing results have been obtained. Some are discussed in Section 7. The linear analysis, however, can be deceptive, and the stylized patterns are remote from realistic portrayals of gene action. Cases chosen for analytic tractability give a misleading picture of the full range of possibilities.

Our model is an infinite-population, multiple-sites or infinite-sites model in continuous time. The dynamical equation is a fairly standard one, but the space of mathematical objects to which it applies is novel. Our model allows a highly flexible specification of pleiotropic gene action. It is especially suited to demographic applications with mutant alleles affecting age-specific schedules. The model is a haploid infinite-population model with no recombination. A parallel model with free recombination, introduced in Section 9, will be developed in a future paper.

Our contribution is to allow large numbers of interacting genes to make small contributions to a continuum of linked traits. Traditional analyses which recognize individual alleles (thus admitting, in principle, arbitrary configurations of pleiotropy) are amenable only to small numbers of loci; quantitative genetics, which reduces the contributions of individual genes to a continuum, reduces the complexity of pleiotropy to covariance matrices.

Although multi-locus models without recombination like our own can be formally imbedded in single-locus models, this imbedding will not generally yield useful results. When a multilocus model is translated into the single-locus framework, it brings along an extra structure of transition rates, whose complexity grows exponentially with the number of loci. When the number of loci is large or, as in our model, effectively infinite, this extra structure overwhelms the single-locus infrastructure. In our function-space setting, the formal embedding itself also poses difficulties. As a consequence, results for single-locus models are mainly helpful as analogies.

Unlike most models of which we are aware, our model comfortably accommodates epistasis. (A very different approach to epistasis, in the two-allele setting, may be found in [6].) The selective cost of a mutant allele can depend on the configuration of other mutant alleles present in a genome. This property is critical to the study of senescence, even without special assumptions about interactions among genes, because the fitness costs of cumulative demographic changes are not linear.

We are able to obtain closed-form representations of the entire time path of solutions to our dynamical equation (Theorem 3.1). Our results are not restricted, like much previous work, to limiting states and equilibrium distributions. We give proofs of convergence over time (Theorem 4.1), and set machinery into place to compute rates of convergence and to cope with changing fitness conditions as well. In Section 5, we present some results about the asymptotic behavior of solutions. Theorem 5.1 gives sufficient conditions for the numbers of certain classes of mutant alleles to increase without limit, generalizing the well-known "error threshold" (cf., [7].) In Section 6 we derive the Poisson limit for the non-epistatic case, as well as proving uniqueness of the solution. In the general epistatic case we do not yet have a proof of uniqueness. In Section 7 we discuss some implications of our results for the theory of longevity. In Section 8 we review earlier work on related problems.

## 2. The model

We consider an infinite population subject to mutation and selection. There is a complete, separable metric space $\mathcal{M}$ of potential mutations, on which is defined a boundedly finite Borel measure $\nu$. (In other words, $\nu$ assigns finite mass to bounded sets; together with the assumptions on $\mathcal{M}$, this condition implies that $\nu$ is $\sigma$-finite.) We refer to this measure as the "mutation rate"; for any set $B$, the quantity $\nu(B)$ represents the rate at which there spontaneously arises a mutant allele from $B$. Our picture is one in which new mutant alleles are steadily arising, each one tagged by a corresponding point of $\mathcal{M}$. For convenience, we identify the tag with a description of the effects that the mutant allele produces: for instance, a function on the non-negative real line $\mathbb{R}^+$ giving the increases in mortality attributed to the action of that allele at each age.

The space of "genotypes" $\mathcal{G}$ is identified with the integer–valued boundedly finite Borel measures on $\mathcal{M}$, with a topology to be described shortly. An element of $\mathcal{G}$ has the form $\sum \delta_{m_i}$, where the $m_i \in \mathcal{M}$ are not necessarily distinct and the number of $m_i$ in any bounded subset of $\mathcal{M}$ is finite. The notation $\delta_x$ stands in general for a unit mass at the point $x$ in the space to which $x$ belongs. Each genotype represents a set of mutant alleles that an individual may carry. The "null genotype" has wild-type alleles at every locus and carries none of these mutant alleles.

The state of the population at time $t$ is denoted $P_t$, which is a Borel probability measure on the measures in $\mathcal{G}$. Thus $P_t$ is the distribution of a random measure [8, 9]. The evolution of the population is presumed to be so slow that it can be represented as occurring in continuous time, without reference to discrete generations.

To each genotype $g$ we assign a"selection cost" $S(g)$; $S$ is a continuous function from $\mathcal{G}$ to $\mathbb{R}^+$. (Including negative costs would be feasible for the finite-time solutions, at the expense of slightly more complicated statements for theorems.) In applications, costs will typically be decrements to growth rates, in effect measuring fitness on a logarithmic scale.

We normalize costs so that $S$ vanishes on the null genotype, and vanishes for no other $g$. On $\mathcal{M}$ we write $S(m)$ for the cost of the singleton $g = \delta_m$. When $S$ is linear, so that $S(g + \delta_m) - S(g)$ is independent of $g$, the model is additive, or *nonepistatic.*

Any measure $P$ on $\mathcal{G}$, like $P_t$, may be determined by the expectation values it assigns to a suitably rich collection of functions $F$ from $\mathcal{G}$ to $\mathbb{R}$, as specified below. For brevity, we write $PF$ or $P(F)$ for the expectation value $PF = \int_{\mathcal{G}} F(g)dP(g)$ of any measurable function from $\mathcal{G}$ to $\mathbb{R}$ such that $\int |F(g)|dP(g) < \infty$. Since genotypes are measures, we can also write $gf = g(f) = \int f(m)dg(m) = \sum f(m_i)$ when $f : \mathcal{M} \to \mathbb{R}$, and $g = \sum \delta_{m_i}$.

Our dynamic equation for $P_t$ is

$$(1) \qquad \frac{d}{dt}P_tF = P_t\left(\int \big[F(\cdot + \delta_m) - F(\cdot)\big]d\nu(m)\right) - P_t(FS) + (P_tF)(P_tS)$$

The meaning of the equation is readily described when $F$ is the indicator function of a set $G$ of genotypes. The first term inside the integral measures the rate at which the population is flowing into the states in $G$ out of all sorts of other states because of the addition of a mutant $m$ that lines up just right to enter $G$. The second term inside the integral measures the rate at which population flows out of $G$ because of new mutations. The remaining two terms measure the effect of selection. The proportional rate of change in mass of the population in $G$ equals the difference between the average fitness cost of genotypes in $G$ and the average fitness cost of the whole population.

Measuring fitness relative to the changing average fitness of the whole population keeps total mass constant and lets the measure $P_t$ represent the probability of finding a randomly selected individual in a given state, modeling population distribution rather than population size. While our equation may be novel, it is strongly analogous to standard mutation-selection dynamics on quantitative traits, such as those given in equation V.2.11 of [5].

When mutations are identical (so that $\mathcal{M}$ comprises only a single point) we have the "mutation-counting model" going back to Kimura and Maruyama [10], whose history will be described in Section 8. A genotype is specified by a natural number, the number of mutant alleles present in it, and (1) becomes

$$\frac{dP_t(n)}{dt} = \nu P_t(n-1) - \nu P_t(n) - P_t(n)\left(S(n) - \sum_m S(m)P_t(m)\right) \tag{2}$$

In the non-epistatic case, where $S$ is additive, the mutation-counting model (or its discrete-time counterpart) has a Poisson distribution with parameter $\nu/S(1)$ as its stationary distribution.

For general $\mathcal{G}$, the counterpart of the Poisson distribution is a Poisson random measure. For the non-epistatic case, Theorem 6.1 establishes conditions for uniqueness and convergence to a stationary distribution given by a Poisson random measure with intensity $(1/S(m))d\nu(m)$, the measure on $\mathcal{M}$ whose Radon-Nikodym derivative with respect to $\nu$ is $1/S$. (A Poisson random measure assigns a Poisson-distributed random integer mass to each measurable set, the mean of the mass assigned to a set is the intensity measure of the set, and the random masses of disjoint sets are independent random variables.) The general theory [9, Chapter 7] takes care of technical details. Even in the non-epistatic case, only rather special starting states lead to the Poisson limit. In the epistatic case, covered by Theorem 3.1, asymptotic distributions, when they exist, may not be Poisson.

We need a suitable notion of weak convergence for boundedly finite random measures. Following Appendix A.2 of [9], we equip the space $\mathcal{G}$ with the metrizable $\hat{w}$-topology under which a sequence of measures $g_1, g_2, \ldots \in \mathcal{G}$ converges to a measure $g \in \mathcal{G}$ if and only if $\lim_n g_n(f) = g(f)$ for each bounded continuous function $f : \mathcal{M} \to \mathbb{R}$ that is supported on a bounded set. A sequence of probability measures $Q_1, Q_2, \ldots$ on $\mathcal{G}$ (that is, a sequence of distributions of boundedly finite random measures on $\mathcal{M}$) converges weakly to the probability measure $Q$ on $\mathcal{G}$ with respect to the $\hat{w}$-topology if and only if $\lim_n Q_n(F) = Q(F)$ for every bounded $\hat{w}$-continuous funtion $F$ on $\mathcal{G}$. This turns out to be equivalent to the requirement that $\lim_n Q_n(F) = Q(F)$ for all $F$ of the form $F(g) = e^{-g(f)}$ for some continuous boundedly supported $f : \mathcal{M} \to \mathbb{R}^+$. This class $\mathcal{F}$ is the sufficiently rich class of functions required for our expectation values to determine our measures: Equality of expectations for $F$ in $\mathcal{F}$ implies equality of expectations for all bounded Borel-measurable $F$ [9, Section 6.4].

We must bear in mind that ordinary theorems guaranteeing existence and uniqueness of solutions to differential equations do not extend to our abstract setting. The derivatives on the left-hand side of (1) might not exist, and, in the presence of unbounded $S$, we might have infinity minus infinity on the right-hand side. Our proofs are constructive, so the meaningfulness of the equation will follow from the properties of the proferred solutions. We prove a reasonable version of uniqueness in the non-epistatic case. In the general epistatic case, we have not yet ruled out multiple alternative meaningful solutions; there could be a complicated mathematical question lurking here.

## 3. Existence of solutions

We express the solution in terms of a certain random meaure on $\mathcal{M} \times \mathbb{R}^+$. We let $\Pi$ denote the Poisson random measure on this space with intensity measure $\nu \otimes$ Lebesgue. Define a time-homogeneous $\mathcal{G}$-valued Markov process $(X_t)_{t\geq 0}$ by

$$X_t := X_0 + \int_{\mathcal{M}\times[0,t]} \delta_m \, d\Pi(m,u), \tag{3}$$

where $X_0$ is a random measure with distribution $P_0$, independent of $\Pi$. Each realization of $X_t$ may be pictured as a discrete set of points, possibly with duplication; as time passes, new points accrete. The cost function $S$ could be allowed to depend on time, but we keep time-independent notation here.

**Theorem 3.1.** *Suppose that there is a positive $T$ such that*

$$\mathbb{E} \exp\left(-\int_0^t S(X_u)du\right) S(X_t) < \infty \tag{4}$$

*for all $t \in [0,T)$. Then the equations (1) have a solution on $[0,T)$, given by*

$$P_t F = \frac{\mathbb{E}\left[\exp\left(-\int_0^t S(X_u)\,du\right) F(X_t)\right]}{\mathbb{E}\left[\exp\left(-\int_0^t S(X_u)\,du\right)\right]}. \tag{5}$$

*Proof.* Define a linear operator on the continuous functions on the genotype space by

$$AF = \int \left[F(\cdot + \delta_m) - F(\cdot)\right] d\nu(m) - S(\cdot)F(\cdot). \tag{6}$$

Given an integrable function $\sigma(t)$, put $\tilde{P}_t = \exp\left(-\int_0^t \sigma(u)\,du\right) P_t$. If we can arrange for $\sigma(t)$ to equal the average selective cost $P_t S$, then, thanks to the chain rule, the derivative of $\tilde{P}_t F$ must equal $\tilde{P}_t AF$.

The operator $A$ may be unbounded if $\nu$ has infinite total mass, but it is well-defined on the class $\mathcal{F}$ and is the generator of a sub-Markovian semigroup $(\Gamma_t)_{t\geq 0}$.

By the Feynman-Kac formula [11, Section III.19], $\Gamma_t$ may be described as

$$\Gamma_t F(g) = \mathbb{E}\left[\exp\left(-\int_0^t S(g + X_u - X_0)\,du\right) F(g + X_t - X_0)\right]. \tag{7}$$

Now, the semigroup $(\Gamma_t)_{t\geq 0}$ solves the forward equation $\frac{d}{dt}\Gamma_t F = \Gamma_t(AF)$ and it follows that $\tilde{P}_t F = \tilde{P}_0 \Gamma_t F$, which equals the numerator of (5). By the condition on $T$, $\tilde{P}_t S$ is finite on $[0, T)$, equalling the derivative of $\tilde{P}_t I$, so that we may put $P_t F = \tilde{P}_t F / \tilde{P}_t I$ and achieve $\sigma(t) = P_t S < \infty$ on [0, T). □

## 4. Representations

Although our solution (5) may look abstract, as long as $\nu(\mathcal{M})$ is finite $P_t F$ can be expressed as a series expansion whose terms can be evaluated by multiple integration. We now derive this expansion, which makes direct calculations feasible in applications. When $\nu(\mathcal{M})$ is finite, we order the points put down by $\Pi$ according to their arrival times $\tau(1), \tau(2)\dots$ and write $Y_n := X_{\tau(n)}$. Let $J_n$ be the indicator function of genotypes with exactly $n$ (possibly overlapping) points: $J_n(g) = 1$ if $g(\mathcal{M}) = n$, and 0 otherwise. Our series expansion for $P_t F$ will take the form $\sum_n P_t J_n F$. We write $x \wedge y$ for the lesser of any two quantities $x$ and $y$. Renewal theory calculations turn (5) into a handy formula for probabilities of $n$-point genotypes:

**Theorem 4.1.** *Suppose $\nu(\mathcal{M}) < \infty$ and $P_0$ puts unit mass at the null state 0, with $S(0) = 0$. Then the solution (5) may be written as $P_t F = \tilde{P}_t F / \tilde{P}_t \mathbf{1}$, with*

$$\tilde{P}_t J_n F = \nu(\mathcal{M})^n e^{-\nu(\mathcal{M})t}\,\mathbb{E}\left[\big(S(Y_1)\dots S(Y_n)\big)^{-1} H_{t,n} F(Y_n)\right]. \tag{8}$$

*Here $H_{t,n}$ is a conditional probability defined in terms of independent unit-rate exponential variables $Z_1, Z_2, \dots$ by the formula*

$$H_{t,n} = \mathbb{P}\left\{\sum Z_j / S(Y_j) < t \,|\, Y_1, \dots Y_n\right\} \tag{9}$$

*If $\sum \nu(\mathcal{M})^n \mathbb{E}[((S(Y_1)\dots S(Y_n))^{-1}]$ is finite, $P_t$ converges in distribution as t goes to infinity. If the sum is infinite, $P_t J_n$ goes to zero for all $n$.*

*Proof.* Consider the numerator of (5) with $J_n F$ in place of $F$. The integral inside the exponential is the sum of terms $S(Y_j)(\tau(j+1) \wedge t - \tau(j))$ for $j$ from 1 to $n$. The factor $J_n$ restricts the domain to the event $\{X_t(\mathcal{M}) = n\}$, an event with probability $e^{-\nu t}\nu^n t^n / n!$, where we write $\nu$ for $\nu(\mathcal{M})$. Conditional on this event, the $Y$'s are independent of the $\tau$'s, and the $\tau$'s are distributed like the order statistics of a sample of $n$ uniform random variables $u_1 \dots u_n$ on $[0, t]$ which may occur in any of $n!$ orderings. Put $u_{n+1} = t$.

To obtain the expectation over the $\tau$'s, we evaluate the integral

$$n!t^{-n}\int_0^t\int_{u_1}^t\cdots\int_{u_{n-1}}^t \exp\left\{-\sum_{i=1}^n S(Y_i)(u_{i+1}-u_i)\right\}du_n\cdots du_2du_1, \tag{10}$$

The change of variables $z_i = u_{i+1}-u_i$ transforms this integral into the product $n!/(t^nS(Y_1)\dots S(Y_n))$ times

$$\int\cdots\int\left(S(Y_1)e^{-S(Y_1)z_1}\right)\dots\left(S(Y_n)e^{-S(Y_n)z_n}\right)dz_1dz_2\dots dz_n, \tag{11}$$

The integrations range over all non-negative $z_1\dots z_n$ such that $z_1+\dots+z_n<t$, yielding the exponential probability expression $H_{t,n}$. Closed-form formulas for $H$ are given in [12, Ch. 1, 13.12]. The probability $e^{-\nu}\nu^nt^n/n!$ times $n!/(t^nS(Y_1)\dots S(Y_n))$ times $H_{t,n}$ gives (8).

We bound $\tilde{P}_tJ_nS$ by $\nu\tilde{P}_tJ_{n-1}$, noting that $H_{t,n}\le H_{t,n-1}$. Summing over $n$, we find $\tilde{P}_tS\le\nu\tilde{P}_t\mathbf{1}\le\nu$, verifying the supremum condition for all finite $T$. The factors of $e^{-\nu t}$ in the numerator and denominator of $P_tJ_nF$ cancel. The conditional probability $H_{t,n}$, is monotone increasing in $t$ toward a limit of 1 for each choice of $n$ and $Y_1,\dots,Y_n$. Hence the limit claim follows by monotone convergence. □

In demographic applications we are typically interested in counting the average number of mutant alleles of a given type that a randomly chosen individual would bear. For $B$ a measurable subset of $\mathcal{M}$, write $R_t(B)$ for the expected number of mutations from $B$ at time $t$; that is, $R_t(B)=\int_{\mathcal{G}}g(B)dP_t(g)$. For special starting states, we can obtain a closed-form density for $R_t$.

**Theorem 4.2.** *Suppose the starting distribution $P_0$ is a Poisson measure with intensity $\pi_0$. Then the measure $R_t$ has the form $\zeta_t(m)d\nu(m)+\eta_t(m)d\pi_0(m)$ where*

$$\begin{aligned}\zeta_t(m)&=\frac{\mathbb{E}\left[\exp\left(-\int_0^t S(X_u)du\right)\int_0^t\exp\left(-\int_\tau^t[S(X_u+\delta_m)-S(X_u)]du\right)d\tau\right]}{\mathbb{E}[\exp\left(-\int_0^t S(X_u)du\right)]}\\ \eta_t(m)&=\frac{\mathbb{E}\left[\exp\left(-\int_0^t S(X_u+\delta_m)du\right)\right]}{\mathbb{E}\left[\exp\left(-\int_0^t S(X_u)du\right)\right]}.\end{aligned} \tag{12}$$

*Proof.* When the initial distribution is Poisson, the entire process $X_t$, including $X_0$, is defined from a Poisson random measure $\xi=\Pi+(X_0,\delta_0)$ on the product space $\mathcal{M}\times\mathbb{R}^+$ with intensity measure $H=\nu\otimes\text{Lebesgue}+\pi_0\otimes\delta_0$. The local Palm distribution for the Poisson random measure $\xi$ at $(m,\tau)$ in $\mathcal{M}\times\mathbb{R}^+$ is the distribution of $\xi$ itself augmented by an atom at $(m,\tau)$ [9, Example 12.1(b)]. For any non-negative bounded Borel-measurable function $G(m,\tau,\xi)$ the Palm integral formula [9, Proposition 12.1.IV] makes

$$\mathbb{E}\int G(m,\tau,\xi)d\xi(m,\tau)=\int\mathbb{E}\,G(m,\tau,\xi+\delta_{(m,\tau)})dH(m,\tau) \tag{13}$$

The integrals are taken over $\mathcal{M} \times \mathbb{R}^+$, and $\mathbb{E}$ operates on $\xi$. Fix $t$ and $B \subset \mathcal{M}$ and choose the function $G$ to be

$$G(m,\tau,\xi) = \exp\left(-\int_0^t S(X_u)\,du\right)\mathbf{1}_{\{m\in B\}}\mathbf{1}_{\{\tau\le t\}}. \tag{14}$$

Bear in mind that $X_u$ is a function of $\xi$. With this $G$, plugging into Equation (5), $\tilde{P}_t X_t(B)$ is given by the left-hand side of (13). On the right-hand side, the extra atom at $(m,\tau)$ changes the argument of the exponential function inside (14) into $-\int_0^\tau S(X_u)du - \int_\tau^t S(X_u+\delta_m)du$. The first term in $H$, which is $\nu\otimes$Lebesgue, calls for integration over $\tau$ and gives the contribution in the numerator of $\zeta$ in (12) with respect to $\nu$. The second term in $H$ puts $\tau$ equal to zero and gives the contribution in the numerator of $\eta$ with respect to $\pi_0$. The denominator in $\zeta$ and $\eta$ is a constant independent of the set $B$. It converts $\tilde{P}_t$ to $P_t$. The indicator function in $G$ arranges that the measure $R_t(B)$ is obtained by integrating over $B$, so $\zeta$ and $\eta$ are indeed Radon-Nikodym derivatives for $R_t$ as claimed. □

When the process $P_t$ starts from the null genotype we set $\pi_0 = 0$. Equations (12) allow us to compare the influences of different cost functions:

**Corollary 4.3.** *Assume that the conditions of Theorems 3.1 and 4.2 are satisfied, and suppose that $S$ is sub-additive; that is, $S(g+g') \le S(g)+S(g')$. Define the corresponding additive cost function $\bar{S}(g) := \int S(\delta_m)dg(m)$. Let $P_t$ and $\bar{P}_t$ be corresponding genotype distributions produced by (5). Then $P_tF \ge \bar{P}_tF$ for any linear $F$ of the form $F(g) = g(f)$, where $f$ is nonnegative, measurable, and has bounded support.*

*Proof.* The sublinearity of $S$ and the linearity of $\bar{S}$ imply

$$\begin{aligned}\zeta_t(m) &\ge \frac{\mathbb{E}\big[\exp\big(-\int_0^t S(X_u)du\big)\int\exp\big(-(t-\tau)S(\delta_m)\big)d\tau\big]}{\mathbb{E}\exp\big(-\int_0^t S(X_u)du\big)}\\ &= \frac{1-e^{-S(m)t}}{S(m)} = \bar{\zeta}_t(m).\end{aligned} \tag{15}$$

Similarly $\eta_t(m) \ge e^{-S(m)t}\bar{\eta}_t(m)$. The result follows from the special case of the Palm integral formula known as Campbell's Theorem [9, (6.4.11)]. □

## 5. Asymptotic Behavior

In contrast to the additive case, genotypes subject to subadditive cost functions may tend to explode. In age-structured models, the total effect of mutant alleles acting after some given age is limited, regardless of how many of them may accumulate. If the rate at which some class of mutant alleles is generated exceeds any countervailing selection-cost increment which they may incur, the number of mutant alleles in that class may be expected to grow without limit.

**Theorem 5.1.** *Assume the conditions of Theorem 3.1 are satisfied. Let $B \subset \mathcal{M}$ be a subset with finite $\nu$-mass. Suppose $0 \le S(g+g^*) - S(g) \le s$ for all $g$ and for all those $g^*$ with masses only at points in $B$, that is, with $g^*(B) = g^*(\mathcal{M})$. Let $J_n^*$ be the indicator function of the set of genotypes with $g(B) = n$. Then $s < \nu(B)$ implies that $P_t J_n^*$ goes to zero for every $n$ as $t$ goes to infinity.*

*Proof.* We write our Poisson process $X_t$ as $X_t^* + X_t^r$, where $X_t^*$ is the restriction of $X_t$ to $B$ and $X_t^r$ is the remainder. These components are independent of each other. Let $U := \inf\{u : X_t(B) > 0\}$ be the arrival time of the first point in $B$, an exponential random variable with mean $1/\nu(B)$. We have

$$0 \le \int_0^t S(X_u^r + X_u^*) - S(X_u^r)du \le s(t-U) \wedge 0. \tag{16}$$

To bound $P_t J_n^*$ , we write the numerator of (5), $\tilde{P}_t J_n^*$, as the expectation of a product of three factors, $\exp\big(-\int_0^t S(X_u^r)du\big)$, $\exp\big(-\int_0^t S(X_u) - S(X_u^r)du\big)$ and $J_n^*(X_t^*)$. The second factor is bounded above by 1 and the third factor is independent of the first. The denominator of (5), $\tilde{P}_t I$, has the same first factor, the same second factor, bounded below by $\exp(-s(t-U)\wedge 0)$, and a third factor identically equal to 1. Using independence, we may cancel the expectations of the first factors in numerator and denominator, so that $P_t J_n^*$ is less than or equal to the quotient of $\mathbb{E}J_n^*(X_t^*)$ and $\mathbb{E}\exp(-s(t-U)\wedge 0)$. Writing $\nu^*$ for $\nu(B)$, this quotient equals $(\nu^* t)^n/n!$ divided by $(\nu^* e^{(\nu^*-s)t} - s)/(\nu^* - s)$. The quotient goes to 0 as $t \to \infty$ for every $n$. □

## 6. Non-epistatic cost functions

In the non-epistatic case, when the cost function $S$ is additive, a proof of uniqueness and an eminently computable formula can be obtained which lead to conditions for convergence as $t$ goes to infinity:

**Theorem 6.1.** *Suppose that $S$ is an additive (nonepistatic) cost function such that the expectation value $\nu(S \wedge 1)$ is finite and suppose that $P_0$ is an initial probability measure such that $P_0 S$ is finite. Then the equations (1) have a unique solution on $[0,\infty)$. A random measure chosen according to $P_t$ may be represented as the sum of two independent random measures. The first component is a Poisson random measure with intensity $(1/S(m))(1-e^{-S(m)t})d\nu(m)$. The second is the initial measure $P_0$, tilted by the weighting $e^{-tgS}$. That is, the second component $Q_t$ satisfies*

*$Q_tF = \tilde{Q}_tF/\tilde{Q}_t1$ with*

$$\tilde{Q}_tF = \int e^{-S(g)t}F(g)dP_0(g) \tag{17}$$

*If $\nu$ is finite, this solution is identical with that given in Theorem 3.1.*

*Proof.* Linearity of $S$ allows us to transform Equation (1) into a first-order linear partial differential equation. Suppose we are given an integrable non-negative function $\sigma(t)$ which serves as a candidate for $P_tS$. Let $z$ be a positive real number and let $f$ on $\mathcal{G}$ be a bounded nonnegative function with bounded support. We take our test functions $F$ now to be of the combined form $F(g) = e^{-gf-zS(g)}$. We write $h(t,z)$ for the real function which will turn out to be $\log(P_tF)$ satisfying given boundary conditions

$$h(0,z) = \eta(z) = \log P_0F. \tag{18}$$

Thanks to the form of $F$ and the linearity of $S$, the expression $\int(F(g+\delta_m) - F(g))d\nu(m)$ from (1) now equals $F(g)\zeta(z)$, where

$$\zeta(z) := \nu(F(\delta_m) - 1). \tag{19}$$

Since $f$ is non-negative, $\left|\zeta(z)\right|$ is bounded by $\nu(f) + (1+z)\nu(S \wedge 1)$. The first term is finite because $f$ is bounded with bounded support. The second term is finite by assumption. If $\exp(h)$ is to satisfy (1), we need $h$ to satisfy the following partial differential equation of the McKendrick type familiar to demographers:

$$\frac{\partial h(t,z)}{\partial t} - \frac{\partial h(t,z)}{\partial z} = \zeta(z) + \sigma(t). \tag{20}$$

We have shown that the right-hand side is well-defined for all non-negative $z$ and $t$.

We solve (20) uniquely for $h$ by exploiting the method of characteristic curves to tranform it into a system of ordinary differential equations. The characteristic curve passing through the point $(t,z)$ is the line $\tau \mapsto (\tau, t+z-\tau)$ [13, Section 3.2]. Defining $\tilde{h}(\tau) := h(\tau, t+z-\tau)$, we get $\tilde{h}'(\tau) = \sigma(\tau) + \zeta(t+z-\tau)$ for $0 \le \tau \le t+z$. Integrating this equation from 0 to $t$ gives

$$h(t,z) = \eta(t+z) + \int_0^t \sigma(\tau)d\tau + \int_z^{t+z} \zeta(r)dr \tag{21}$$

The final term in $\zeta$ is equal to

$$\nu\left[-t + \left(e^{-f(m)-zs(m)} - e^{-f(m)-(z+t)s(m)}\right)/s(m)\right]. \tag{22}$$

We now set $P_tF = \exp(h(t,z))$. Additivity of $S$ makes the derivative of $P_tF$ equal the right-hand side of (1) plus $(\sigma(t) - P_tS)(P_tF)$. Also, $-P_tS$ is the partial derivative of $h$ with respect to $z$ at $z = 0$ and $f \equiv 0$, which is the sum of $\nu(1 - e^{tS})$ and $P_0 S e^{-tS} / P_0 e^{-tS}$. Setting $-\sigma(t)$ equal to this sum is therefore the unique choice which makes $P_t$ satisfy (1). Writing out $h$ and setting $z = 0$, we recognize the Laplace functional of the convolution of probability measures specified in the theorem. □

The first piece of $P_t$ clearly converges to a Poisson random measure as long as $\nu/S$ is boundedly finite. But that is not the complete story of asymptotic behavior. In general, the influence of $P_0$ in $Q_t$ may persist. In the limit, however, we may apply Varadhan's Lemma [14, Theorem III.13] to show that $Q_t$ becomes concentrated on the set of genotypes of minimum selective cost.

**Corollary 6.2.** *Suppose $P_0$ and $S$ satisfy the conditions of Theorem 6.1, that the support* $\operatorname{supp} P_0$ *is compact, and that $S$ is continuous. Let $\sigma = \inf\{S(g) : g \in \operatorname{supp} P_0\}$. If $\mathcal{O}$ is any open neighborhood of $\{g \in \operatorname{supp} P_0 : S(g) = \sigma\}$, then $\lim_{t\to\infty} Q_t(\mathcal{O}) = 1$. In particular, if $P_0\{0\} > 0$, the tilted measure $Q_t$ converges to $\delta_0$.*

## 7. Applications to the theory of longevity

We outline a few of the many applications to the biodemography of longevity. We take the space of potential mutations $\mathcal{M}$ to be $C[0, \infty)$, the continuous real-valued functions on $\mathbb{R}^+$, supplied with any of the usual metrics corresponding to uniform convergence on bounded intervals [15, Section 1.44]. A mutation measure $\nu$ on this space is the distribution of a stochastic process. We base our selective cost function $S(g)$ on Equation 4.9 of Charlesworth [4, p. 140], taking into account more recent discussion [16, p. 930]. The cost function is defined in terms of the age-specific survival function $l_x(g)$ and the age-specific fertility rate $f_x(g)$ specific to each genotype, along with a conversion factor $T$, representing a baseline value for the length of a generation, and a rate $r_0$ representing a population-wide baseline intrinsic rate of natural increase usually set to zero in applications.

$$S(g) = (1 - \int e^{-r_0 x} l_x(g) f_x(g) dx)(1/T) \tag{23}$$

7.1. **Gompertz hazards.** Charlesworth [17] has suggested a possible origin for Gompertzian (exponentially increasing) hazard rates through a process of mutation-selection balance which fits into our generalized model. Members of a species are taken to be subject to a common high background age-independent hazard rate $\lambda$ plus age-dependent contributions from mutations. Each mutant allele $m$ may be represented as a continuous function $m(x)$ of age added onto the hazard function for an individual. Charlesworth's elementary models assume constant fertility at all ages above an age $b$ of sexual maturity, forgoing any *a priori* upper age cutoff.

The selective cost $S(g)$ for a genotype $g$ from (23) takes the following form when time is measured in generations rather than years:

$$S(g) = 1 - \int_b^\infty \lambda \exp\Big(-\lambda x + \lambda b - \int \int_0^x m(a) da\, dg(m)\Big) dx \tag{24}$$

This cost function $S$ is a non-additive epistatic cost function. Following established practice, Charlesworth substitutes the additive cost function

$$\hat{S}(g) = \int \Big(\int_0^\infty (e^{-\lambda(x-b)} \wedge 1) m(x) dx\Big)\ dg(m). \tag{25}$$

This function $\hat{S}$ is an additive approximation to $S$.

We first show that under the same premises as [17] our model confirms the same conclusions. With additive costs as in (25), Theorem 6.1 and Corollary 6.2 give us sufficient conditions for the distribution of genotypes to converge to the Poisson random measure with intensity $\nu/\hat{S}$. It suffices that the starting state put positive weight on the null state and have compact support and that $\nu$ and $\nu/S$ be boundedly finite. Then the average of the hazard rates over genotypes will converge to $\lambda + \int_{\mathcal{M}} (m(x)/\hat{S}(m)) d\nu(m)$, equivalent to [17, Eq.4a].

It is worth mentioning that this expression for the average of the hazard rates is not the equilibrium aggregate hazard rate for the whole population, because the heterogeneity mediated by the Poisson distribution implies attrition of higher-risk genotypes with advancing age. The Poisson expression for the additive genetic variance and covariance also require modification for age-specific attrition.

Charlesworth focuses on translation families of mutations, which we may write as $m_y(x) = m_0(x-y)$ with effects only after an age of onset $y$. With $d\nu(m_y) = \nu_0 dy$ on some $[b', \infty)$, he displays choices for $m_0$ which make the average of the hazard rates into an exact Gompertz-Makeham function $\lambda + \nu_0 \exp(-\lambda(x-b))$ on the support of $\nu$, and others which approximate Gompertz-Makeham shapes for large $x$. (These shapes do not include heterogeneity corrections.)

We now observe that our generalized model predicts different qualitative behavior when the additive approximation of (24) by (25) is not guaranteed to hold. The additive theory predicts that a "wall of death" with an infinite equilibrium mean hazard rate appears at, but not before, the age at which reproduction comes to an end [17, p. 60]. Theorem 5.1 implies a more dramatic breakdown. The mean hazard function can actually reach infinity at ages at which fertility is still strictly positive, if the full epistatic cost function $S$ in (24) is kept in place of the additive approximation (25). The same is also true, if the bounded cost function $S$ is replaced by an unbounded cost function defined, as in Equation 4.12 of [4, p. 141], to equal the decrement to the intrinsic rate of natural increase resulting from the mutations contributing to each genotype. Contrary to additive theory, the "wall of death" is not tied to the end of reproduction but involves a fine-tuned balance between mutation rates and tapering costs.

7.2. **Gaussian process mutations.** We now apply our model to move beyond stylized cases and investigate a wider range of possible specifications for the age-specific effects of mutations. The cases considered in 7.1, in which a constant background mortality imprints a Gompertzian pattern onto increments to the hazard function, share the property that every mutant is deleterious at every affected age. Is this property essential to the imprinting, or can the age-specific force of selection readily produce the same kind of outcomes with mutants that mix positive and negative effects?

Our framework allows quite general pleiotropic specifications. A natural starting point is the case of Gaussian processes. The fitness cost for this brief discussion will be the additive approximation (25). Suppose that the mutation process generates mutations proportionately to a positive-real-parameter Gaussian process with expectation $a(x)$ and covariance function $c(x, x')$, conditioned on fitness cost bounded away from 0. That is, if we look at the pattern of age effects in a randomly chosen mutation, it looks like a realization of this Gaussian process, subject to $\hat{S}(m) > s > 0$ for some $s$. The overall rate of mutation is a constant $\nu_0$. For rigorous treatment, we also need to condition on events which keep the resulting hazard functions non-negative and insure the validity of the additive approximation, but here we shall assume that the choices of parameters keep misbehavior rare enough that it can be neglected.

The average over genotypes of the hazard function is given by

$$h(x) = \lambda + \int_{\mathcal{M}} \frac{m(x) d\nu(m)}{\int_0^\infty \lambda e^{-\lambda z} dz \int_0^{z+b} m(y) dy}. \tag{26}$$

The denominator $\hat{S}(m)$ is obtained from (25) by integration by parts. Since the numerator and denominator of (26), linear functionals of $m$, are both Gaussian random variables, we can describe their joint distribution simply by computing their covariance. The conditional expectation of $m(x)/\hat{S}(m)$ conditional on $\hat{X}(m)$, is obtained via linear regression.

When $a(x) \equiv 0$, the conditional expectation turns out to be independent of $\hat{S}(m)$, making the integral in (26) independent of the bound $s$ (except for a small change in the proportionality constant) and equal to $\nu_0 \, \mathbb{E}\, m(x)\hat{S}(m)/ \operatorname{Var}\big(S(m)\big)$.

For processes with zero mean, then, we may treat $s$ as 0. As one example, take the mutations to be realizations of Brownian motion, so that $c(x, x') = x \wedge x'$ and $a(x) = 0$. The increment to the mean hazard rate then has the form $c_1 - c_2 e^{-\lambda x}$ for $x \geq b$.

Can any Gaussian mutation process with zero mean generate approximations to Gompertz-Makeham hazard functions? The covariance kernel must satisfy $\left|c(x, y)\right| \leq c(x, x)/2 + c(y, y)/2$. With $a(x) \equiv 0$, the incremental mortality is bounded above by $c_1 + c_2 c(x, x)$ for constants $c_1$ and $c_2$. The mortality thus cannot be exponentially increasing over a long range of ages, unless this exponential increase is built into the mutation process itself.

## 8. Historical Background

We discuss in this section the relationship of our model to the existing corpus of work on related topics. It was J.B.S. Haldane [18] who articulated the concept of mutation-selection balance as early as 1937. Crow and Kimura [19], Ewens [20], and Kingman [21] give the foundations of the subject. Bürger [22] and [5] covers the present state of the art. These authors put only limited emphasis on age structure; Charlesworth [4] and [23] propounds the age-specific side.

Infinite population models in which fitness is a function of the number but not the identity of mutant genes go back to Kimura and Maruyama [10]. They state discrete-time and continuous-time dynamic equations for special cases which readily suggest the general "mutation-counting model" (2). They obtain some closed-form equilibrium distributions. Conditions for convergence to stationary states follow from a theorem of Kingman [24], generalizing theorems of Moran [25, 26]. Bürger [5, pp. 298-308] traces the subsequent history. Markov-chain versions with stepwise mutations of identical deleterious effect have Poisson stationary distributions (see e.g. Haigh [27] or Durrett [28, p. 137]). The Poisson limit is implicit in estimations of equilibrium genetic variance [17].

Mutation-counting models in the tradition of Kimura and Maruyama are more tractable than the general case considered here because they are a kind of multi-locus model that can be subsumed under the theory of single-locus models. The count of mutant alleles at different sites can be likened to the integer label on a countable set of alleles at a single site, subject to constraints on the non-zero interallelic mutation rates. Models defined by various alternative sets of constraints have been studied in some detail.

The most famous of these single-locus models is Sir John Kingman's "House of Cards" (HC) described in [29] and [21]. Kingman's infinite-population discrete-time model posits a single gene with potentially infinitely many alleles. Alleles mutate to new alleles at a constant rate; each new allele has a random fitness, given by a probability distribution on $[0, 1]$. The state of the system is given by a distribution of fitnesses on $[0, 1]$, and the dynamics are governed by a standard evolution equation. Kingman [29] gives the original proof that the distribution of fitnesses for the HC process converges to a limiting distribution. This model has many descendants, including the "HC-approximation" [30] for stabilizing selection around an intermediate optimum.

Our model differs from HC and its counterparts in four main ways. Mutant alleles in HC have no properties other than fitness. Mutant alleles in our model are tagged by an effect represented by a point in a general metric space whose specification determines the fitness through the impact on demographic rates. In HC there is only a single locus. In our model we are concerned with the heterogeneity of whole sets of mutant alleles across a large number of loci within the population. Because HC includes only a single locus, it offers no possibility for interactions between the fitnesses of different alleles. Our model is open to general epistasis. Finally, HC is well-suited to the use of Markovian methods and sample-path analysis, whereas our proofs require non-Markovian machinery.

A highly versatile general formulation of single-locus models has been developed by Reinhard Bürger [5, chapter IV.2]. His "general mutation-selection model" warrants close comparison with our own. Like us, Bürger draws mutants from a general space, according to an arbitrary distribution. Bürger requires his space to be locally compact, whereas we allow any complete, separable metric space so as to include mutants identified with continuous functions on $[0, \infty)$.

The substantive difference between Bürger's model and our own is in their contrasting views of the genome. Bürger focuses on a single locus, with (perhaps) infinitely many potential alleles. Each individual's genotype is characterized by a single quantity and the population is characterized by a distribution on the mutation space. We take a more synoptic view, watching the (perhaps) infinitely many alleles pop up at (perhaps) infinitely many loci, thereby opening up the representation of population heterogeneity. In our model, the only mutation process is the conversion of an undifferentiated wild type to a random mutant allele, so we need not introduce transition rates between alleles. However, our flexible treatment of heterogeneity means that even the description of the state of the system has to be more abstract than is customary in population genetics.

In a different setting, Del Moral and Miclo [31] present results which parallel our Theorem 3.1. Their conditions are more general in some respects and more restrictive in others. While their concerns are remote from biology, they use the terminology of "mutation generators" and "adaptation" in their descriptions. They prove that the differential equation model which we analyze can be derived as an infinite-population limit of finite nondeterministic Moran models for interacting particles. A new book [32] expands this line of investigation.

We accompany Del Moral and Miclo, on a road that diverges from the Markov modeling, branching processes, branching diffusions, and superprocesses which are so important to stochastic population theory [33]. Pioneering work in these areas by [34, 35, 36] has been followed by extensive results on particle processes and measure-valued diffusions with selection, including [37, 38, 39, 40, 41, 42, 43, 44]. Such Markov processes, even when they involve selection, are essentially linear. Lineages rise or fall at their own rates, according to their fitnesses, independent of the outside population. By contrast, the mutation-selection paradigm on which we focus has to be nonlinear, since every lineage has negative fitness. The models are saved from trivial degeneracy by a renormalization, conditioning the process on long-term survival. This ingredient introduces a quadratic nonlinearity into the evolution equation, inasmuch as the entire population contributes to the selective pressure on each individual, bringing non-Markovian arguments to the fore.

## 9. Prospects

The generalized model for mutation-selection balance presented here can be applied widely to settings where age structure matters. Because the model allows mutations with a mixture of positive and negative effects, it gives scope to some blending of ideas about mutation accumulation with ideas about antagonistic pleiotropy. It offers a handle on responses to changing fitness conditions through the finite-time solutions, along with machinery for treating epistatic cost functions.

The Palm formula in Section 4 facilitates the construction of an alternative version of our model which, in contrast to (1), allows for free recombination (FR). In line with [10], [45], and [46], we postulate conditions on the relative rates of recombination, selection, and mutation which lead, in the continuous-time limit, to a process in which $P_t$ is always a Poisson random measure on $\mathcal{G}$ with some intensity measure $\rho_t$. Our results in Section 4, derived in the absence of recombination, give us the form of an equation for $\rho_t$ in this generalized free-recombination model.

Differentiating (12) at $t = 0$ leads to a representation for $\rho_t$ of the form

$$\rho_t = \rho_0 + \nu t - \int_0^t D\rho_\tau d\tau \tag{27}$$

Here $D\rho$ is a measure whose density with respect to $\rho$ at $m$ is $\mathbb{E}[S(X^\rho + \delta_m) - S(X^\rho)]$, and $X^\rho$ is the Poisson random measure with intensity $\rho$. Rigorous development of this alternative is reserved for a future paper.

## Acknowledgements

D.R.S. was supported by Grant K12-AG00981 from the National Institute on Aging, S.N.E. was supported in part by Grants DMS-0071468 and DMS-0405778 from the National Science Foundation and by the Miller Foundation, and K.W.W. was supported by Grant P01-008454 from the National Institute on Aging. The authors thank Montgomery Slatkin and Nick Barton for helpful discussions.

## References

[1] K. W. Wachter, Hazard curves and life span prospects, in: J. R. Carey, S. Tuljapurkar (Eds.), Life Span: Evolutionary, Ecological, and demographic perspectives, Vol. 29 (Supplement) of Population and Development Review, The Population Council, New York, 2003, pp. 270–291.

[2] P. Medawar, An unsolved problem in biology: An inaugural lecture delivered at University College, London, 6 December, 1951, H. K. Lewis and Co., London, 1952.

[3] W. D. Hamilton, The moulding of senescence by natural selection, Journal of Theoretical Biology 12 (1966) 12–45.

[4] B. Charlesworth, Evolution in age-structured populations, Cambridge University Press, Cambridge, 1994.

[5] R. Bürger, The mathematical theory of selection, recombination, and mutation, John Wiley, Chichester, New York, 2000.

[6] L. A. Zhivotovsky, K. V. Pylkov, A general multilocus two-allele model for epistatic selection, Journal of Genetics 77 (2/3) (1998) 115–21.

[7] G. P. Wagner, P. Krall, What is the difference between models of error thresholds and muller's ratchet, Journal of Mathematical Biology 32 (1) (1993) 33–44.

[8] O. Kallenberg, Foundations of modern probability, 2nd Edition, Probability and its Applications (New York), Springer-Verlag, New York, 2002.

[9] D. J. Daley, D. Vere-Jones, An introduction to the theory of point processes, Springer Series in Statistics, Springer-Verlag, New York, 1988.

[10] M. Kimura, T. Maruyama, The mutational load with epistatic gene interaction, Genetics 54 (1966) 1337–51.

[11] L. C. G. Rogers, D. Williams, Diffusions, Markov Processes and Martingales, 2nd Edition, Vol. 1, Cambridge University Press, Cambridge, 2000.

[12] W. Feller, An Introduction to Probability and its Applications, Vol. 2, John Wiley & Sons, New York, 1971.

[13] L. C. Evans, Partial differential equations, Vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 1998.

[14] F. den Hollander, Large Deviations, American Mathematical Society, Providence, RI, 2000.

[15] W. Rudin, Functional Analysis, 2nd Edition, McGraw-Hill, New York, 1991.

[16] B. Charlesworth, Fisher, Medawar, Hamilton, and the evolution of aging, Genetics 156 (2000) 927–931.

[17] B. 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) (2001) 47–65.

[18] J. B. S. Haldane, The effect of variation on fitness, American Naturalist 71 (1937) 337–49.

[19] J. F. Crow, M. Kimura, An Introduction to Population Genetics Theory, Harper and Row, New York, 1970.

[20] W. J. Ewens, Mathematical Population Genetics, Springer-Verlag, New York, Heidelberg, Berlin, 1979.

[21] J. F. C. Kingman, Mathematics of Genetic Diversity, Society for Industrial and Applied Mathematics, 1980.

[22] R. Bürger, Mathematical properties of mutation-selection models, Genetica 102/103 (1998) 279–98.

[23] B. Charlesworth, Mutation-selection balance and the evolutionary advantage of sex and recombination, Genetical Research 55 (2000) 199–221.

[24] J. F. C. Kingman, On the properties of bilinear models for the balance between genetic mutation and selection, Mathematical Proceedings of the Cambridge Philosophical Society 81 (1977) 443–53.

[25] P. A. P. Moran, Global stability of genetic systems governed by mutation and selection, Math. Proc. Cambridge Philos. Soc. 80 (2) (1976) 331–336.

[26] P. A. P. Moran, Global stability of genetic systems governed by mutation and selection. II, Math. Proc. Cambridge Philos. Soc. 81 (3) (1977) 435–441.

[27] J. Haigh, The accumulation of deleterious genes in a population – muller's ratchet, Theoretical Population Biology 14 (1978) 251–67.

[28] R. Durrett, Probability Models for DNA Sequence Evolution, Probability and its Applications (New York), Springer-Verlag, New York, 2002.

[29] J. F. C. Kingman, A simple model for the balance between mutation and selection, Journal of Applied Probability 15 (1978) 1–12.

[30] M. Turelli, Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle, Theoretical Population Biology 25 (1984) 138–93.

[31] P. Del Moral, L. Miclo, A Moran particle system approximation of Feynman-Kac formulae, Stochastic Process. Appl. 86 (2) (2000) 193–216.

[32] P. del Moral, Feynman-Kac formulae. Genealogical and interacting particle approximations, Springer-Verlag, Heidelberg, 2004.

[33] S. Ethier, T. Kurtz, Markov Processes: Characterization and Convergence, John Wiley & Sons, 1986.

[34] S. Sawyer, Branching diffusion processes in population genetics, Advances in Appl. Probability 8 (4) (1976) 659–689.

[35] R. C. Griffiths, On the distribution of allele frequencies in a diffusion model, Theoret. Population Biol. 15 (1) (1979) 140–158.

[36] R. C. Griffiths, Allele frequencies with genic selection, J. Math. Biol. 17 (1) (1983) 1–10.

[37] P. Donnelly, T. G. Kurtz, Genealogical processes for Fleming-Viot models with selection and recombination, Ann. Appl. Probab. 9 (4) (1999) 1091–1148.

[38] S. N. Ethier, T. G. Kurtz, The infinitely-many-alleles model with selection as a measure-valued diffusion, in: Stochastic methods in biology (Nagoya, 1985), Vol. 70 of Lecture Notes in Biomath., Springer, Berlin, 1987, pp. 72–86.

[39] S. N. Ethier, T. G. Kurtz, Fleming-Viot processes in population genetics, SIAM J. Control Optim. 31 (2) (1993) 345–386.

[40] D. A. Dawson, S. Feng, Large deviations for the Fleming-Viot process with neutral mutation and selection, Stochastic Process. Appl. 77 (2) (1998) 207–232.

[41] D. A. Dawson, S. Feng, Large deviations for the Fleming-Viot process with neutral mutation and selection. II, Stochastic Process. Appl. 92 (1) (2001) 131–162.

[42] H.-C. Kang, S. M. Krone, C. Neuhauser, Stepping-stone models with extinction and recolonization, Ann. Appl. Probab. 5 (4) (1995) 1025–1060.

[43] E. Perkins, Dawson-Watanabe superprocesses and measure-valued diffusions, in: Lectures on probability theory and statistics (Saint-Flour, 1999), Vol. 1781 of Lecture Notes in Math., Springer, Berlin, 2002, pp. 125–324.

[44] D. A. Dawson, Measure-valued Markov processes, in: École d'Été de Probabilités de Saint-Flour XXI—1991, Vol. 1541 of Lecture Notes in Math., Springer, Berlin, 1993, pp. 1–260.

[45] P. G. Higgs, Error thresholds and stationary mutant distributions in multi-locus diploid genetics models, Genetical Research 63 (1994) 63–78.

[46] K. J. Dawson, The dynamics of infinitesmially rare alleles, applied to the evolution of mutaiton rates and the expression of deleterious mutations, Theoretical Population Biology 55 (1999) 1–22.