Time and chance happeneth to them all: Mutation, selection and recombination

Steven N. Evans

Department of Mathematics & Department of Statistics
University of California at Berkeley

October, 2011

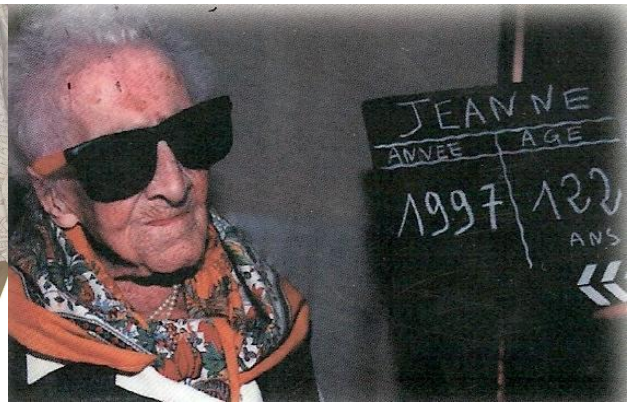
I returned, and saw under the sun, that the race is not to the swift, nor the battle to the strong, neither yet bread to the wise, nor yet riches to men of understanding, nor yet favour to men of skill; but time and chance happeneth to them all. Ecclesiastes 9:11

Collaborators

- David Steinsaltz
  Statistics
  Oxford
- Kenneth W. Wachter
  Demography
  U.C. Berkeley

A mutation-selection model for general genotypes with recombination.
To appear in *Memoirs of the American Mathematical Society*.
Available at arXiv:q-bio.PE/0609046

Multicellular organisms mature, age and die



Why do organisms age?

- Things fall apart.
- BUT, organisms can make repairs.
- There are physical constraints on repair (cf. modern toasters - modularity).
- Repairs can introduce "bugs" (cf. software, my attempts at plumbing).
- Reproduction is the ultimate repair – despite things falling apart, life has continued to exist for billions of years.

"You fix it by buying a new one."

Human mortality rates INCREASE with chronological age after adolescence

$$\lim_{\Delta \downarrow 0} \mathbb{P}\{\text{age at death} \in [t, t+\Delta] \,|\, \text{live to age } t\}$$

increases with t after adolescence

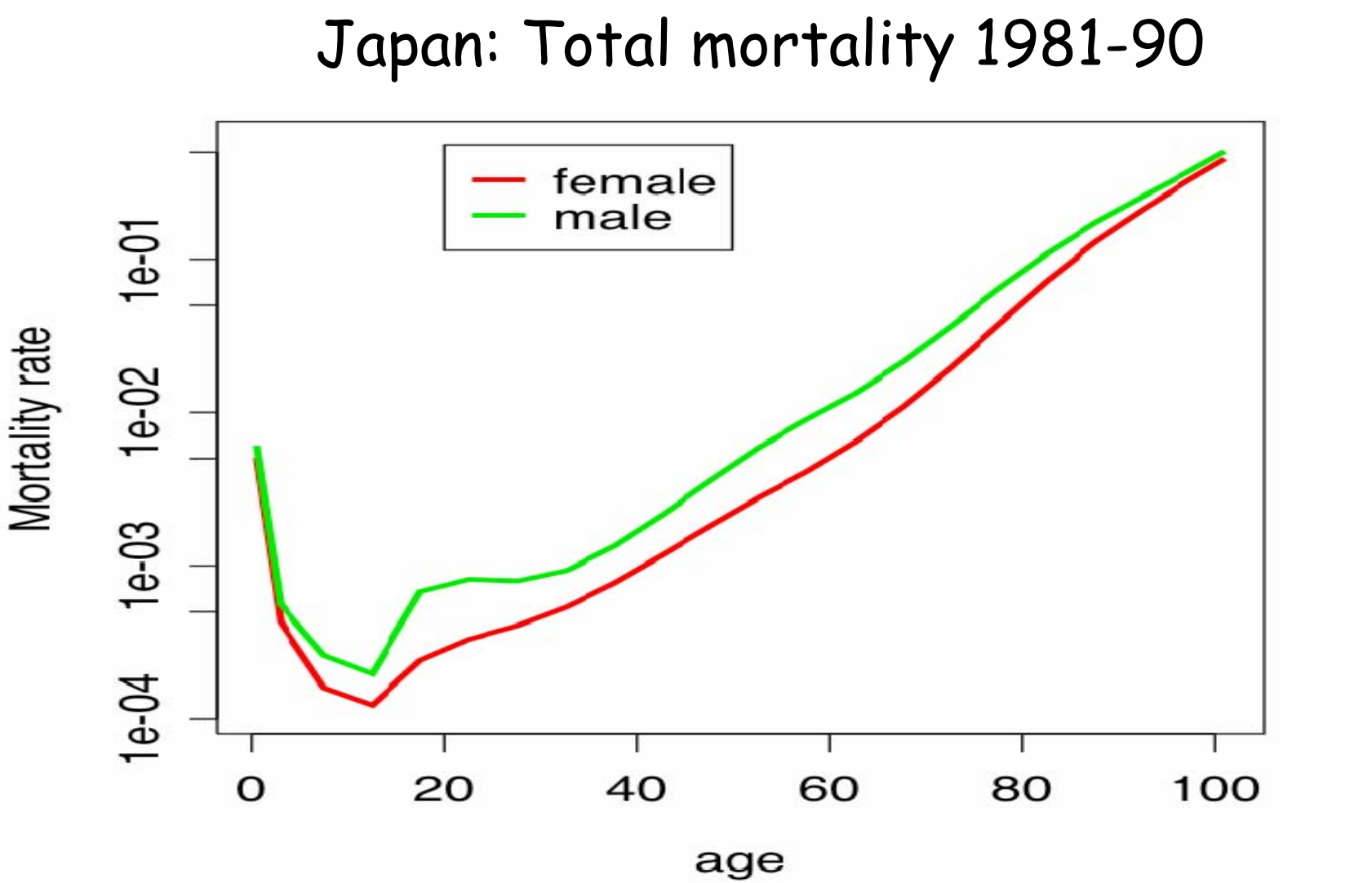
Mortality for many organisms is an exponential function of age

BENJAMIN GOMPERTZ

We observe that in those tables the numbers of living in each yearly increase of age are from 25 to 45 nearly, in geometrical progression.

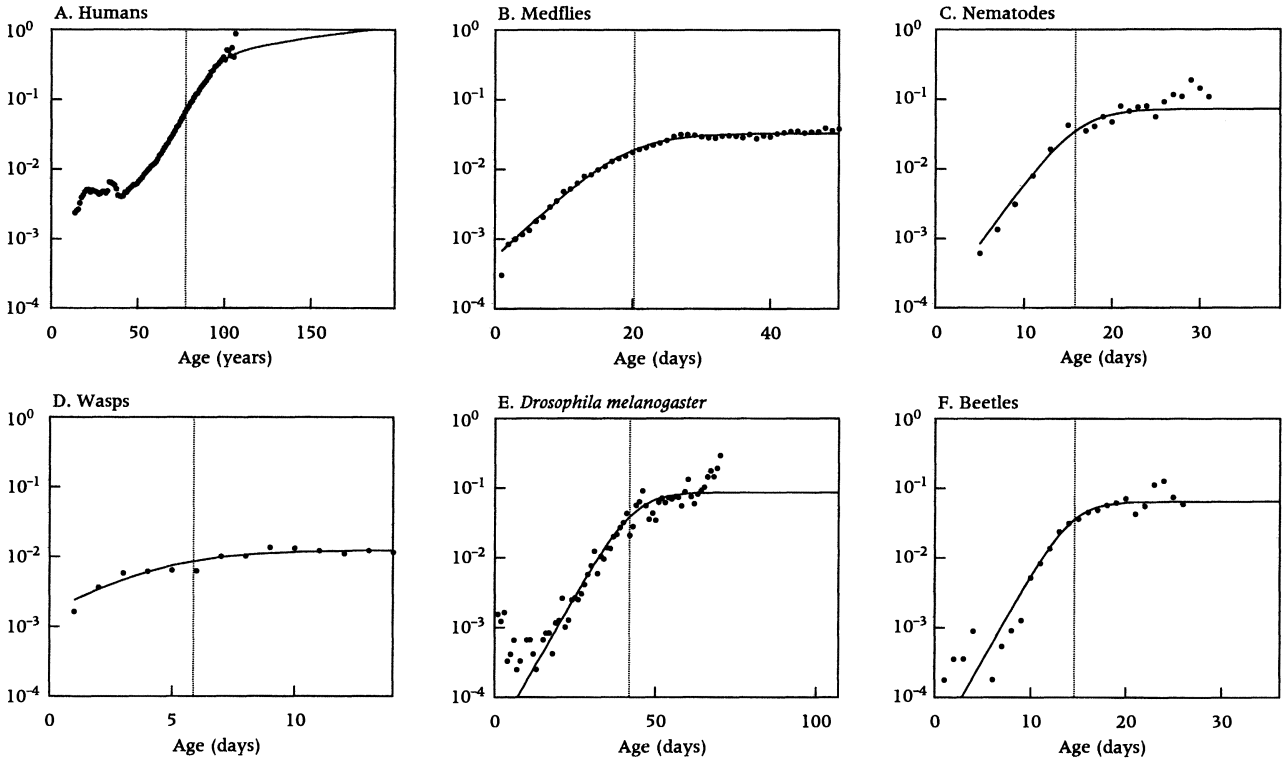
Gompertz 1825

An example


Japan: Total mortality 1981-90
female
male
Mortality rate
1e-01
1e-02
1e-03
1e-04
0
20
40
60
80
100
age

More examples

FIGURE 2 Age-specific death rates for six species

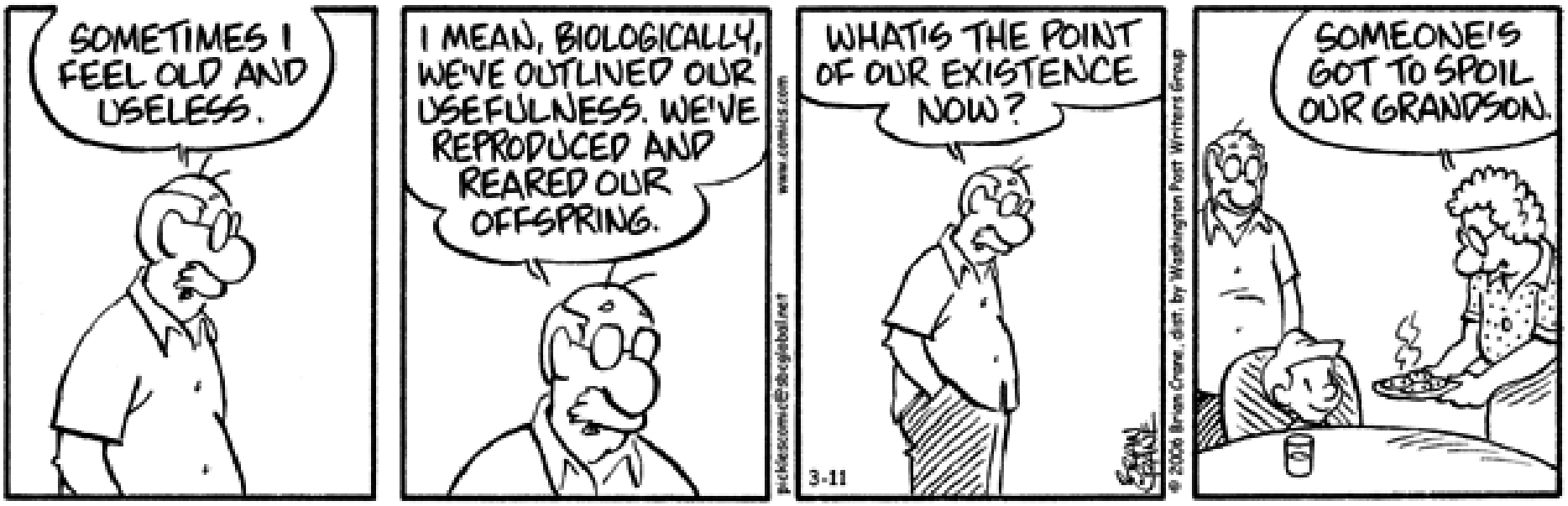


NOTE: Dots indicate values directly calculated from data, and solid lines indicate estimates using frailty models. In each panel, the vertical dotted line indicates the estimated modal life span, and the vertical axis indicates the number of deaths per individual-OHM of exposure. (OHM is one-hundredth of the modal life span.)
SOURCES: See Figure 1.

Evolutionary explanations of senescence and mortality

Biologists have proposed the following informal model.

- There are large numbers of mildly deleterious mutations that meander towards extinction in the population due to natural selection but are constantly reintroduced.
- The adverse effects of these mutations are mainly felt later in life.
- Natural selection will not oppose mutations with negative effects that occur after the individual has been able to reproduce.



A challenge

CAN WE TURN THESE IDEAS INTO MATHEMATICS?

Biological assumptions

- the population is infinite,
- the genome may consist of infinitely many loci (a locus is a site where a mutation can occur),
- each individual has two parents,
- mating is random,
- an individual's genotype is a random mosaic of the genotypes of its parents produced by recombination,
- an individual has one copy of each gene,
- starting from an ancestral wild type, mutations only accumulate,
- fitness is calculated for individuals rather than for mating pairs,
- genotypes with additional mutant alleles are less fit,
- recombination acts on a faster time scale than mutation or selection.

Describing genotypes

- Let $\mathcal{M} :=$ the collection of loci of interest.
- Take $\mathcal{M}$ to be an arbitrary complete, separable metric space.
- An individual's genotype is the set of loci at which mutant alleles are present.
- So, a genotype is an element of the space $\mathcal{G}$ of integer–valued finite Borel measures on $\mathcal{M}$.
- The genotype $\sum_i \delta_{m_i}$, where δ_m is the unit point mass at the locus $m \in \mathcal{M}$, has mutations away from the ancestral wild type at loci $m_1, m_2, \ldots$.
- The wild genotype is the null measure.

Describing population structure

- The genetic composition of the population at some time is completely described by a probability measure P on the space of genotypes $\mathcal{G}$.
- For a subset $G \subseteq \mathcal{G}$, $P(G)$ is the proportion of individuals that have genotypes belonging to G.

Describing mutation

- New mutations from the ancestral type appear in a subset A of the locus space $\mathcal{M}$ at rate $\nu(A)$, where ν is a finite measure on $\mathcal{M}$.
- Write X^ν for a Poisson random measure on $\mathcal{M}$ with intensity measure ν.
- Mutation in one generation transforms the probability measure P to the probability measure $\mathfrak{M}P$, where

$$\begin{aligned}\mathfrak{M}P[F] &= \int_{\mathcal{G}} F(g)\,\mathfrak{M}P(dg)\\ &:= \int_{\mathcal{G}} \mathbb{E}[F(g+X^\nu)]\,P(dg).\end{aligned}$$

  – individuals get an extra Poisson load of mutations.

- Note: If P is the distribution of a Poisson random measure, then so is the probability measure $\mathfrak{M}P$.

Describing fitness

- A genotype $g \in \mathcal{G}$ has an associated selective cost $S(g)$.
- The difference in the rate of sub-population growth between
  the sub-population of individuals with genotype g''
  and
  the sub-population of individuals with genotype g'
  is $S(g') - S(g'')$.
- Genotypes with more accumulated mutations are less fit, so

$$S(g+h) \geq S(h), \quad g, h \in \mathcal{G}.$$

- Normalize so that $S(0) = 0$ (only differences in costs matter).

Example of a demographic selective cost

- There is a constant background hazard λ.
- An mutation at locus $m \in \mathcal{M}$ contributes an increment $\theta(m, x)$ to the cumulative hazard at age x.
- The probability an individual with genotype $g \in \mathcal{G}$ lives beyond age x is

$$\ell_x(g) := \exp\left(-\lambda x - \int_{\mathcal{M}} \theta(m, x)\, g(dm)\right).$$

- At age x an individual has offspring at rate $f(x)$ – fertility.
- For the sub-population with genotype g, the relative size of the next generation is $\int_0^\infty f(x)\, \ell_x(g)\, dx$.
- The selective cost of genotype g is thus

$$S(g) = \int_0^\infty f(x)\, \exp(-\lambda x) \left[1 - \exp\left(-\int_{\mathcal{M}} \theta(m, x)\, g(dm)\right)\right] dx$$

(normalizing so that $S(0) = 0$).

Describing selection

- Selection in one generation transforms the probability measure P to the probability measure $\mathfrak{S}P$, where

$$\begin{aligned}\mathfrak{S}P[F] &= \int_{\mathcal{G}} F(g)\,\mathfrak{S}P(dg)\\ &:= \frac{\int_{\mathcal{G}} e^{-S(g)} F(g)\,P(dg)}{\int_{\mathcal{G}} e^{-S(g)}\,P(dg)}\\ &= \frac{P[e^{-S}F]}{P[e^{-S}]}\end{aligned}$$

  – "tilting" with a Radon-Nikodym derivative.

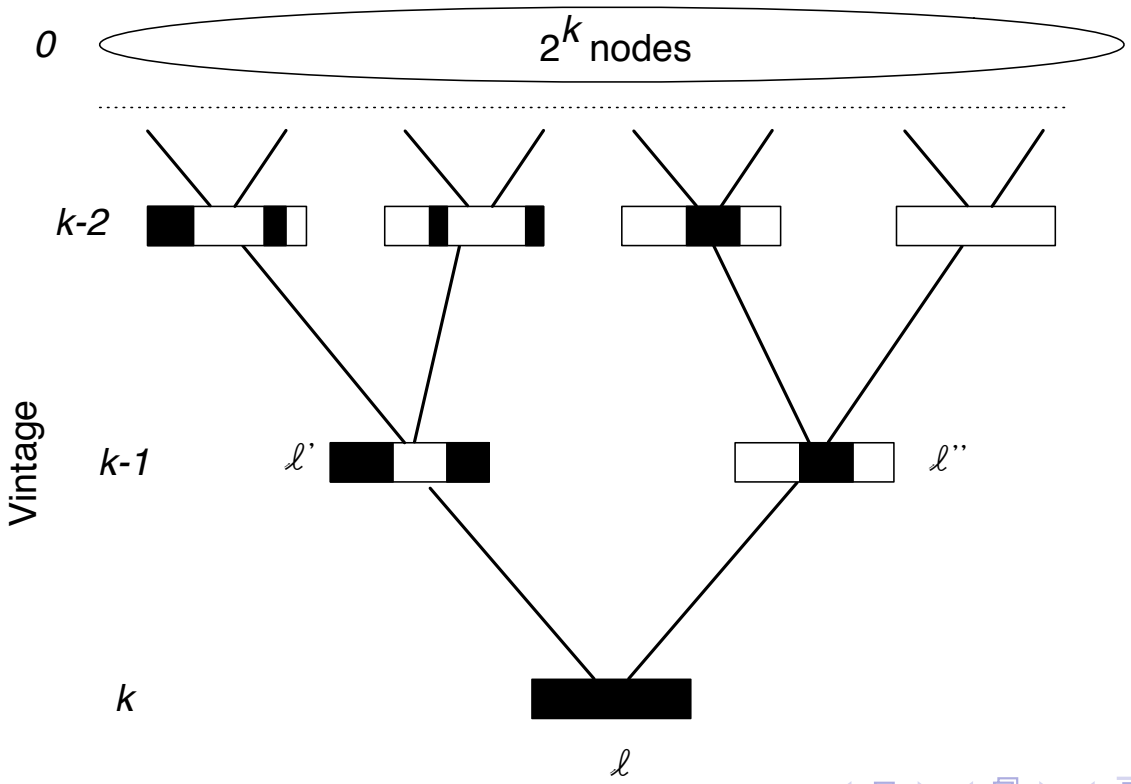
- Note: If P is the distribution of a Poisson random measure, then $\mathfrak{S}P$ will not be Poisson unless $S(g+h) = S(g) + S(h)$ – non-additive selection introduces linkage.

Describing recombination

- Recombination takes two genotypes $g', g'' \in \mathcal{G}$ and replaces the genotype g' by the genotype g defined by

$$g(A) := g'(A \cap R) + g''(A \cap R^c),$$

where the random set $R \subseteq \mathcal{M}$ is chosen according to a probability measure $\mathcal{R}$ on the set $\mathcal{B}(\mathcal{M})$ of Borel subsets of $\mathcal{M}$.



Describing recombination – continued

- Recombination in one generation transforms a probability measure P to $\mathfrak{R}P$, where

$$\mathfrak{R}P[F] := \int_{\mathcal{B}(\mathcal{M})} \int_{\mathcal{G}} \int_{\mathcal{G}} F(g'(\cdot \cap R) + g''(\cdot \cap R^c))\, P(dg')\, P(dg'')\, \mathcal{R}(dR).$$

- Note: Under weak assumptions on P and $\mathcal{R}$, the limit $\lim_{k\to\infty} \mathfrak{R}^k P$ is the distribution of a Poisson random measure with the same intensity measure as P – recombination reduces linkage.

Notation reminder

- $\mathcal{M}$:= space of loci (places where mutations can occur),
- $\mathcal{G}$:= space of genotypes (finite integer-valued measures on $\mathcal{M}$),
- a population is a probability measure on $\mathcal{G}$,
- ν := mutation intensity measure (a finite measure on $\mathcal{M}$),
- S := selective cost (an increasing function from $\mathcal{G}$ to $\mathbb{R}_+$),
- $\mathfrak{M}$:= mutation operator (transforms probability measures on $\mathcal{G}$),
- $\mathfrak{S}$:= selection operator (transforms probability measures on $\mathcal{G}$),
- $\mathfrak{R}$:= recombination operator (transforms probability measures on $\mathcal{G}$).

Combining mutation, selection and recombination

- If the population in generation 0 is described by the probability measure Q_0, then the population in generation k is described by $(\mathfrak{RMS})^k Q_0$ – it is usually intractable to determine this probability measure explicitly.
- Recall our assumption that recombination acts on a faster time scale than both mutation and selection.
- For $n \in \mathbb{N}$, define $\mathfrak{M}_n$ and $\mathfrak{S}_n$ like $\mathfrak{M}$ and $\mathfrak{S}$, but with the mutation intensity measure ν replaced by ν/n and the selective cost S replaced by S/n
- Note that

$$\lim_{n\to\infty} n\big(\mathfrak{M}_n P[F] - P[F]\big) = \int_{\mathcal{G}} \left(\int_{\mathcal{M}} F(g+\delta_m) - F(g)\, \nu(dm) \right) P(dg)$$

and

$$\lim_{n\to\infty} n\big(\mathfrak{S}_n P[F] - P[F]\big) = P[S]\, P[F] - P[S \cdot F].$$

Taking the limit: discrete generations $\longrightarrow$ continuous time

Theorem 1

Suppose that Q_0 is the distribution of a random measure on $\mathcal{M}$ with intensity measure ρ_0. Under mild assumptions, the probability measure $(\mathfrak{RM}_n\mathfrak{S}_n)^{\lfloor nt \rfloor}Q_0$, $t > 0$, converges to a probability measure P_t that is the distribution of a Poisson random measure with intensity measure ρ_t, where $(\rho_t)_{t\geq 0}$ is the unique solution of the non-linear, measure-valued ODE

$$\frac{d}{dt}\rho_t(dm) = \nu(dm) - \mathbb{E}\left[S(X^{\rho_t} + \delta_m) - S(X^{\rho_t})\right]\rho_t(dm).$$

- Recall for a finite measure π on $\mathcal{M}$ that X^π is a Poisson random measure on $\mathcal{M}$ with intensity measure π.
- Equip probability measures on $\mathcal{G}$ with a Wasserstein metric.
- Showing convergence is technically very demanding – selection drives the population away from Poisson while recombination drives it towards Poisson.

Equilibria

- Write $\mathcal{H}^+$ for the space of finite measures on $\mathcal{M}$.
- Define $F : \mathcal{M} \times \mathcal{H}^+ \to \mathbb{R}_+$ by

$$F_\pi(m) := \mathbb{E}\big[S(X^\pi + \delta_m) - S(X^\pi)\big], \quad m \in \mathcal{M},\ \pi \in \mathcal{H}^+$$

$=$ expected marginal cost of an additional mutation at m.

- Recall that the intensity measures $(\rho_t)_{t \geq 0}$ evolve according to the measure-valued dynamical system

$$\frac{d}{dt}\rho_t(dm) = \nu(dm) - F_{\rho_t}(m) \cdot \rho_t(dm).$$

- An equilibrium is a measure $\rho_* \in \mathcal{H}^+$ such that $\nu - F_{\rho_*} \cdot \rho_* = 0$.
- That is, ρ_* is absolutely continuous with respect to ν with Radon-Nikodym derivative satisfying

$$F_{\rho_*}(m)\frac{d\rho_*}{d\nu}(m) = 1 \quad \text{for } \nu\text{-a.e. } m \in \mathcal{M}.$$

Equilibria may or may not exist

- Suppose that

$$S(g) := 1 - \exp\left(-\int_{\mathcal{M}} \sigma(m)\, g(dm)\right)$$

for some $\sigma : \mathcal{M} \to \mathbb{R}_+$.

- Can show an equilibrium exists if and only if

$$\int_{\mathcal{M}} \frac{1}{1 - \exp(-\sigma(m))}\, \nu(dm) < \infty$$

and $\nu(\mathcal{M}) \leq e^{-1}$.

- If an equilibrium exists, it is of the form

$$\rho_*(dm) = \frac{\exp(c)}{1 - \exp(-\sigma(m))}\, \nu(dm),$$

where $ce^{-c} = \nu(\mathcal{M})$.

Equilibria for small mutation rates

- Recall that $F_\pi(m) := \mathbb{E}[S(X^\pi + \delta_m) - S(X^\pi)]$ for $\pi \in \mathcal{H}^+$ and ρ_* is an equilibrium if $F_{\rho_*} \cdot \rho_* = \nu$.
- If $S(g+h) = S(g) + S(h)$ for all $g, h \in \mathcal{G}$ (additivity), then $F_\pi(m) = S(\delta_m)$ and

$$\rho_*(dm) := S(\delta_m)^{-1} \nu(dm)$$

  is an equilibrium provided $\rho_* \in \mathcal{H}^+$.
- If $\pi \approx 0$, then $F_\pi(m) \approx S(\delta_m)$ – perhaps equilibria exist if the mutation intensity measure ν is small.

Theorem 2

Suppose $\inf\{S(\delta_m) : m \in \mathcal{M}\} > 0$. *If* $\epsilon > 0$ *is sufficiently small, then there exists* $\rho_*^{(\epsilon)} \in \mathcal{H}^+$ *satisfying*

$$F_{\rho_*^{(\epsilon)}} \cdot \rho_*^{(\epsilon)} = \epsilon \nu.$$

- Note: $\epsilon \mapsto \rho_*^{(\epsilon)}$ satisfies a non-linear, measure-valued ODE with $\rho_*^{(0)} = 0$ by an "implicit function theorem".

Concave selective costs, monotonicity and comparison

The selective cost S is concave if

$$S(g+h+k) - S(g+h) \leq S(g+k) - S(g) \quad g,h,k \in \mathcal{G};$$

that is, the marginal cost of an additional mutation decreases as more mutations are added.
Note: The selective cost in the demographic example is always concave.

Theorem 3 (monotonicity)

Suppose that the selective cost S is concave. If $\dot{\rho}_0 \geq 0$ (respectively, ≤ 0), then $\rho_s \leq \rho_t$ (resp. $\rho_s \geq \rho_t$) for all $0 \leq s \leq t < \infty$.

Theorem 4 (comparison)

Suppose that the selective cost functions S' and S'' are concave and $S'(g+\delta_m) - S'(g) \geq S''(g+\delta_m) - S''(g)$ for all $g \in \mathcal{G}$ and $m \in \mathcal{M}$. Let $(\rho')_{t\geq 0}$ and $(\rho'')_{t\geq 0}$ be corresponding families of intensity measures with $\rho'_0 \leq \rho''_0$. Then, $\rho'_t \leq \rho''_t$ for all $t \geq 0$.

Concave selective costs and minimal equilibria

- Suppose that the selective cost S is concave and $\rho_0 = 0$.
- Then, $\dot{\rho}_0 = \nu \geq 0$ and $\rho_s \leq \rho_t$ for $s \leq t$.
- Therefore, either

$$\lim_{t \to \infty} \rho_t(\mathcal{M}) = \infty$$

or

$$\lim_{t \to \infty} \rho_t = \rho_* \in \mathcal{H}^+ \text{ exists.}$$

In the latter case, ρ_* is an equilibrium.

- The latter case occurs if and only if there is some equilibrium ρ_{**}, in which case $\rho_* \leq \rho_{**}$ – if any equilibria exist, then there is a well-defined minimal equilibrium.

Returning to the demographic example

Recall the "demographic" selective cost

$$S(g) = \int_0^\infty f(x)\, \exp(-\lambda x) \left[1 - \exp\left(-\int_{\mathcal{M}} \theta(m,x)\, g(dm)\right)\right] dx.$$

The measure ρ_* is an equilibrium if

$$\left[\int_0^\infty \left(1 - e^{-\theta(m',x)}\right) f(x) \exp(-\lambda x) \times \exp\left(-\int_{\mathcal{M}} \left(1 - e^{-\theta(m'',x)}\right) \rho_*(dm'')\right) dx\right] \rho_*(dm') = \nu(dm').$$

Demographic example with localized hazard increments

- Suppose that $\mathcal{M} = \mathbb{R}_+$ and

$$\theta(m,x) = \begin{cases} 0, & x < m, \\ \eta(m), & x \geq m \end{cases}$$

  for some function $\eta : \mathbb{R}_+ \to \mathbb{R}_+$.
- That is:
  - Each mutation is identified with a specific age at which it has an effect.
  - A mutation having an effect at age m has an effect of magnitude $\eta(m)$.
- In this case, if $\nu(dm) = q(m)\,dm$, then ρ_* is an equilibrium if $\rho_*(dm) = p_*(m)\,dm$ with

$$\left[\int_m^\infty f(x)\exp(-\lambda x)\exp\left(-\int_x^\infty \left(1-e^{-\eta(n)}\right)p_*(n)\,dn\right)dx\right] \times \left(1-e^{-\eta(m)}\right)p_*(m) = q(m).$$

- This leads to a second order non-linear ODE for p_*.

Validating the demographic model

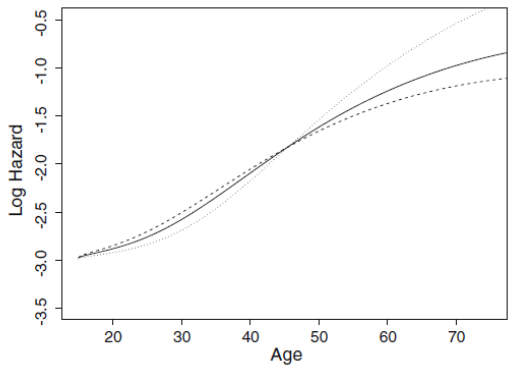
- Just looking at a population we only observe the fertility $f(x)$ and

  proportion of the population that lives to age x

$$= \exp(-\lambda x) \exp\left(- \int_{\mathcal{M}} \left(1 - e^{-\theta(m'',x)} \right) \rho_*(dm'') \right).$$

- We need breeding experiments and genomics to determine $\mathcal{M}$, ν, λ, and θ.
- Simple choices of $\mathcal{M}$, ν and θ can produce Gompertz-like mortality.

Fig. 3 Logarithm of predicted hazard for three cases showing early upward bend, straight middle Gompertzian stretch, and late downward bend. The cases all have $\lambda = \phi = 1/20$ and $\eta \equiv 0.100$. The *solid curve* has $\nu_{tot} = 0.150$, and $\xi = 6.0$; the *dashed curve* has $\nu_{tot} = 0.170$, and $\xi = 5.5$; the *dotted curve* has $\nu_{tot} = 0.120$, and $\xi = 7.0$



- What, if anything, does this mean?
- We would like an analogue of the central limit theorem stating that Gompertz behavior appears whenever suitable general, qualitative assumptions hold.

Conclusion

THE CHALLENGE CONTINUES . . .