Frailty, Profile Likelihood and (Medfly) Mortality

Roger Koenker

University of Illinois, Urbana-Champaign

Tbilisi: 8 September 2015

Joint work with Jiaying Gu (U. of Toronto)

Carey et al (1992) Medfly Experiment

- 1,203,646 medflies survival times recorded in days
- 167 aluminum mesh cages of roughly 7200 flies each
- Adults were given a diet of sugar and water *ad libitum*
- Sex determined at time of death
- Pupae were sorted into one of five size classes

Medfly Findings

Three biologically surprising conclusions emerged from this study:

Medfly Findings

Three biologically surprising conclusions emerged from this study:

- Mortality rates actually declined at the oldest observed ages, contradicting the view that aging is an inevitable, monotone process of senescence.

Medfly Findings

Three biologically surprising conclusions emerged from this study:

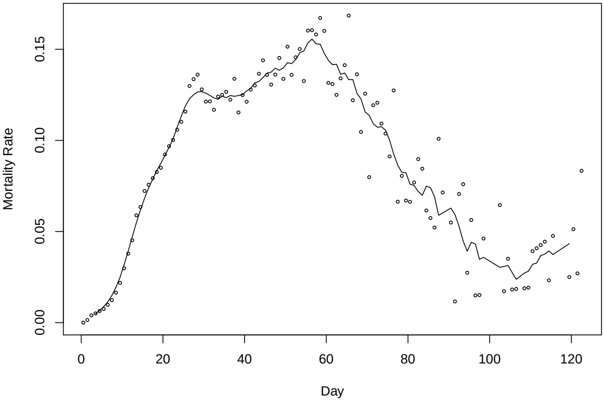
- Mortality rates actually declined at the oldest observed ages, contradicting the view that aging is an inevitable, monotone process of senescence.
- The right tail of the survival distribution was, at least by human standards, remarkably long.

Medfly Findings

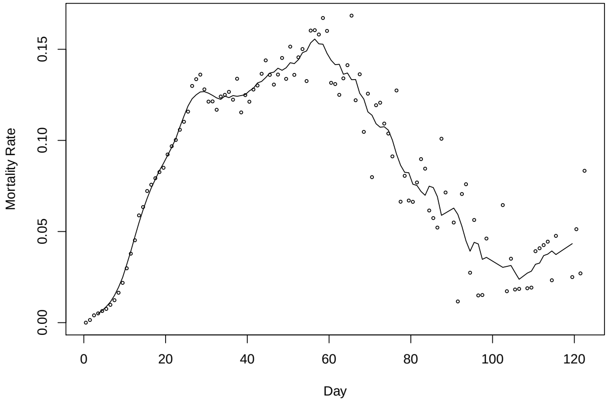
Three biologically surprising conclusions emerged from this study:

- Mortality rates actually declined at the oldest observed ages, contradicting the view that aging is an inevitable, monotone process of senescence.
- The right tail of the survival distribution was, at least by human standards, remarkably long.
- The experiment provided strong evidence for a crossover in gender specific mortality rates.

Raw Daily Medfly Mortality Rates and MA(7) Smooth



Raw Daily Medfly Mortality Rates and MA(7) Smooth



"The observed decline in mortality offered no consolation to the 99.8 percent of the flies that were already dead by age 60, but to the remaining 0.02 percent, more than 2000 less frail ones, it offered some hope of a prolonged retirement. The oldest flies in the experiment expired on day 172."

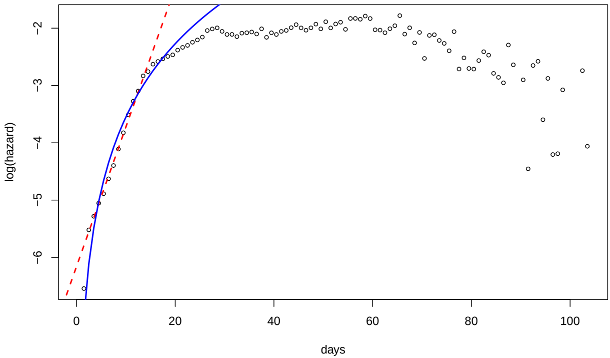
Mixture Models

There are several approaches to modeling this surprising tail behavior. An early suggestion, by Vaupel and Carey (1998) was to consider mixtures of the classical Weibull or Gompertz survival distributions:

$$g(x) = \int \varphi(x, \theta) dF(\theta),$$

where φ is either the Weibull or Gompertz density, and F is an unknown mixing density. In both cases we consider scale mixtures, and we therefore need to determine an appropriate shape parameter for each distribution.

Estimating the Shape Parameter



Estimated Baseline Gompertz and Weibull Hazard Models: Linear (Gompertz) and log linear (Weibull) fits to the initial $k = 15$ observations of daily log mortality rates. This is somewhat analogous to Hill estimation of the Pareto exponent.

Kiefer-Wolfowitz Nonparametric MLE for Mixture Models

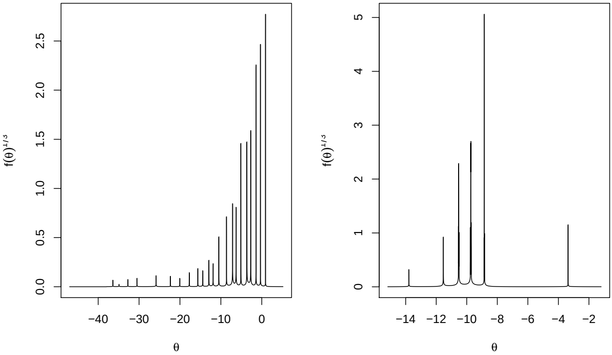
Given iid observations, $x_1, \cdots, x_n$ from g, we wish to solve

$$\max_{F \in \mathcal{F}} \{ \sum_{i=1}^{n} \log(g(x_i)) \mid g(x) = \int \varphi(x, \theta) dF(\theta) \}$$

where $\mathcal{F}$ is the (convex) set of distribution functions.

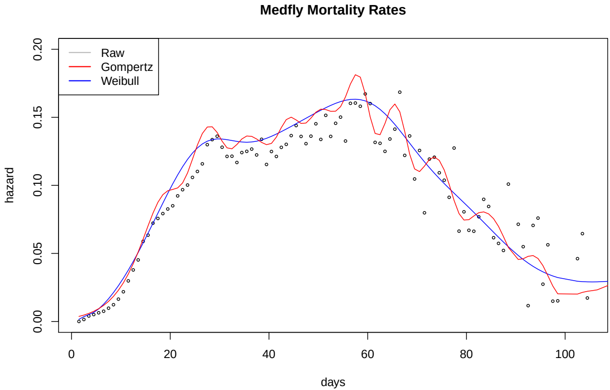
- Solutions are discrete distributions with fewer than n points of support,
- Laird (1978) proposed solving with the EM algorithm,
- Recent interior point algorithms for additively separable convex programs enable much more efficient and accurate solutions, K and Mizera (2014).

Kiefer Wolfowitz Estimated Mixing Distributions



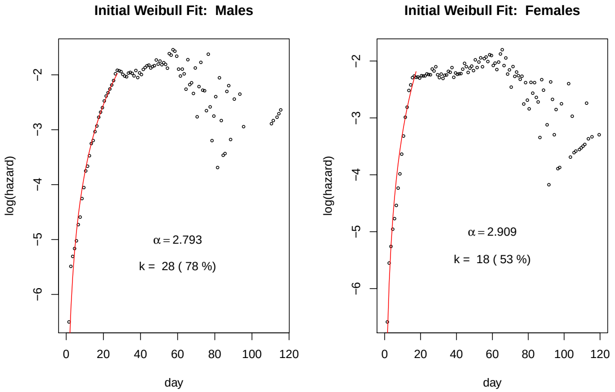
Estimated Mixing Distributions for the Gompertz (left) and Weibull (right) Models

Estimated (Mixture) Hazard Functions



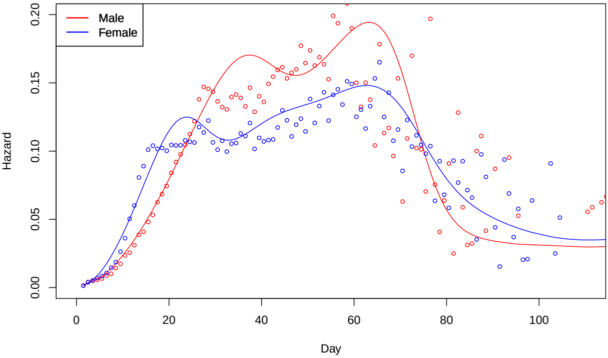
Hazard Functions for the Estimated Gompertz and Weibull Models

Gender Specific Mixture Models



Gender Specific Baseline Weibull Models: Weighted least squares fitting of the initial 20 daily mortality rates. The percentage of the sample population dead by day 20 is given in parentheses. The estimated shape parameter of the baseline Weibull model is α.

Gender Specific Estimated Hazard Rates



Gender Specific Hazard Functions for the Weibull Mixture Model: Raw daily mortality rates are plotted in black for males and grey for females, superimposed are the estimated hazard functions for the Weibull mixture models.

Gender Crossover

Several interesting features:

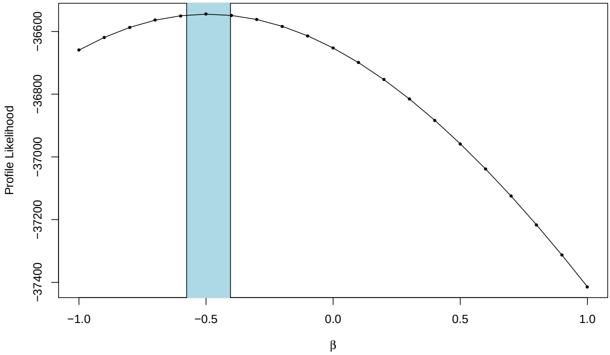
- Until about age 20 female mortality is higher than for males.
- But after age 20 female mortality is substantially below that of males.
- Hazard crossover implies survival function crossover at about age 36.
- Reverses human pattern in which males are more frail than females.
- The second hazard crossover at 75 shouldn't be taken very seriously since it is quite imprecise.

The Cage Density Controversy

A controversial aspect of the original experiment was the effect of cage density on mortality; critics argued that high density would make flies unhappy, and lead to earlier mortality.

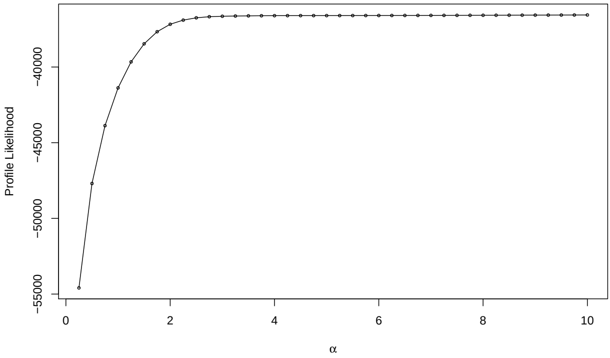
- Let the baseline Weibull scale take the form $\theta_0 \exp(d_i \beta)$ where d_i is the initial cage density,
- Evaluate the profile KW mixture likelihood on a equally spaced grid on $[-1, 1]$,
- This yields point estimate of $\beta = -0.5$ implying higher cage density shifts the survival distribution to the right, thus prolonging lives, and contradicting the critics.
- The classical Wilks $2 \log \lambda \rightsquigarrow \chi_1^2$ interval for β is quite precise.

Profile Likelihood for Cage Density Effect



Profile Likelihood for the Initial Cage Density Effect in the Weibull Mixture Model

Profile Likelihood for the Weibull Shape Parameter



Sometimes profile likelihood is just plain ugly! Ishwaran (1999) has shown that in Weibull models with shape parameter α_0, one can find Weibull mixture models arbitrarily close (in Hellinger distance) for any $\alpha > \alpha_0$.

Five Medfly Life Lessons

Males are tough ...	but only until 40.
Bigger is better ...	but only before 18.
Small is beautiful...	after 25.
Crowds are good ...	especially of guys.
Life gets safer ...	but only after 60.

Five Medfly Statistical Lessons

- Nonparametric mixture models can be easily estimated by the Kiefer-Wolfowitz MLE
- Finite dimensional mixture models are considerably more difficult to estimate
- Profile likelihood can be used to estimate semiparametric extended models
- Sometimes semiparametric inference with profile likelihood is feasible, van der Vaart(1996)
- Sometimes profile likelihood behaves badly.

Some References

CAREY, J., P. LIEDO, D. OROZCO, AND J. VAUPEL (1992): "Slowing of mortality rates at older ages in large Medfly cohorts," *Science*, 258, 457–61.

KIEFER, J., AND J. WOLFOWITZ (1956): "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *The Annals of Mathematical Statistics*, 27, 887–906.

KOENKER, R. (2014): "REBayes: An R package for empirical Bayes methods," Available from CRAN `http://www.r-project.org`.

KOENKER, R., AND J. GU (2014): "Frailty, Profile Likelihood and Medfly Mortality," in *Contemporary Developments in Statistical Theory: A Festschrift for Hira Lal Koul*, ed. by S. Lahiri, A. Schick, A. SenGupta, and T. Sriram, pp. 227–238. Springer.

KOENKER, R., AND I. MIZERA (2014): "Convex Optimization, Shape Constraints, Compound Decisions and Empirical Bayes Rules," *Journal of the American Statistical Association*, 109, 674–685.

LAIRD, N. (1978): "Nonparametric Maximum Likelihood Estimation of a Mixing Distribution," *Journal of the American Statistical Association*, 73, 805–811.

VAN DER VAART, A. (1996): "Efficient Maximum Likelihood Estimation in Semiparametric Mixture Models," *The Annals of Statistics*, 24, 862–878.

VAUPEL, J., AND J. CAREY (1998): "Compositional interpretations of Medfly mortality," *Science*, 260, 1666–1667.