

Demographic window to aging in the wild: constructing life tables and estimating survival functions from marked individuals of unknown age

Hans-Georg Müller,¹ Jane-Ling Wang,²
James R. Carey,^{2,3} Edward P. Caswell-Chen,⁴
Carl Chen,⁴ Nikos Papadopoulos² and Fang Yao⁵

¹Department of Statistics, ²Department of Entomology, and
³Center for the Economics and Demography of Aging, University

of California, Berkeley, CA 94720, USA

⁴Department of Nematology, University of California, Davis,
One Shields Ave., Davis, CA 95616, USA

⁵University of Thessaly, Laboratory of Entomology and Applied
Zoology, Phytokou St. N.m. 384 46 N. Ionia Magnesia, Greece

⁶Department of Statistics, Colorado State University, Fort Collins,
CO 82523, USA

Summary

We address the problem of establishing a survival schedule for wild populations. A demographic key identity is established, leading to a method whereby age-specific survival and mortality can be deduced from a marked cohort life table established for individuals that are randomly sampled at unknown age and marked, with subsequent recording of time-to-death. This identity permits the construction of life tables from data where the birth date of subjects is unknown. An analogous key identity is established for the continuous case in which the survival schedule of the wild population is related to the density of the survival distribution in the marked cohort. These identities are explored for both life tables and continuous lifetime data. For the continuous case, they are implemented with statistical methods using non-parametric density estimation methods to obtain flexible estimates for the unknown survival distribution of the wild population. The analytical model provided here serves as a starting point to develop more complex models for residual demography, i.e. models for estimating survival of wild populations in which age-at-entry is unknown and using remaining information in randomly encountered individuals. This is a first step towards a broad new concept of 'expressed demographic information content of marked or captured individuals'.

Key words: capture; demographic identity; density estimation; information content; life table; non-parametric estimation; remaining lifetime; residual demography; survival function.

Introduction

The life table is one of the most important tools in demographic and gerontological research because it is used to characterize the mortality and survival properties of cohorts and to quantify the actuarial rate of aging. The historical application of classical life table methods in aging science has been largely restricted to the use of mortality data from either humans or experimental animals maintained in the laboratory, or to life tables based on capture–recapture methods to assess aging in wild populations (Udevitz & Ballachey, 1998). In both applications, it is mandatory that age-at-entry is known. This has limited the use of life tables because in the analysis of field populations one often encounters and marks individuals of unknown age. However, capture–recapture and other current field methods generally require capturing and marking of young individuals, or alternatively of individuals of known age, for monitoring throughout their lives until they die.

The predominance of capture–recapture methods has had a limiting effect on the use of flexible non-parametric statistical methods that make minimal assumptions on survival schemes, and have the desirable property that they do not presume statistical parametric survival models. Because in non-parametric modelling one does not specify the functional form of hazard or survival functions, these methods require the exact recording of lifetimes and therefore are not applicable to usual capture–recapture designs, which correspond to usually coarsely graded life tables (Lebreton *et al.*, 1992; Williams *et al.*, 2002).

Because of the importance of the life table in aging research and the growing interest in understanding aging in the wild (Austad, 1993; Congdon *et al.*, 1994; Finch, 2001; Reznick *et al.*, 2001; Tatar & Yin, 2001), the case of life table analysis with unknown age at entry and the analogous situation for continuous lifetimes is clearly of great interest. We describe a life table identity that, by making certain key assumptions, enables us to estimate the age-specific life table rates from data based on the mark, release and monitoring of randomly captured individuals of unknown age from the time of their entry into the study (i.e. marking) to their death. We also discuss an identity for the continuous case in which marked animals are continuously monitored until their death. Continuous monitoring, when feasible, enables the continuous version of the analysis, which provides us with substantially more detailed information

Correspondence

Dr James R. Carey, Department of Entomology, University of California, Davis,
One Shields Ave., Davis, CA 95616, USA. Tel.: +1 530 7526217; fax: +1 530
7521537; e-mail: jrcarey@ucdavis.edu

Accepted for publication 17 March 2004

about the behaviour of survival functions and hazard rates (force of mortality). Our approach can also be used in conjunction with life tables that are obtained from capture–recapture experiments in those situations in which age-at-entry is unknown. Current designs of capture–recapture experiments, however, are not amenable to continuous lifetime analysis with the preferred flexible non-parametric methods, allowing for the construction of hazard rate estimates.

Consider a population that is assumed to be stable, stationary and closed. Individuals are randomly captured at an unknown age and marked, and their time-to-death recorded. The question we address is this: Can the information on time-to-death for this randomly captured marked subgroup provide the necessary information to construct a life table for the population at large? We will demonstrate that the answer to this question is yes because of a life table identity that reveals a mathematical relationship between the distribution of deaths in the marked cohort and the age structure of the original population. Individuals in the captured and marked sample are assumed to have remaining lifetimes similar to those in the wild. This model may be particularly adequate for some human populations.

The problem of constructing a survival schedule from incomplete data has been studied in anthropology (Müller et al., 2002) and has applications to human populations such as the !Kung and the Ache for which only incomplete demographic data are available (Howell, 1979; Hill & Hurtado, 1996; Hawkes et al., 1998; Jones et al., 2002). An anthropologist may encounter a group of people whose ages are unknown but whose remaining lifetime can be recorded. The key identity, on which the reconstruction of the survival schedule that we propose is based, asserts that for such situations a life table for the population can be obtained, under certain assumptions. Application of the key identity then establishes a new way to construct life tables and estimate survival functions.

We derive this key identity for both discrete life tables and situations that are modelled by continuous survival times. In the continuous case, this identity is a consequence of a close relationship between the density of the remaining lifetimes in a cohort of randomly sampled subjects and the survival schedule of the population from which the subjects were sampled. We provide statistical implementations of this identity by applying suitably adapted non-parametric density estimation methods. The proposed model is developed for a stable, stationary and closed population but possesses sufficient flexibility to allow for modifications of these assumptions.

The concept and techniques that we describe in this paper will help to advance understanding of senescence in the wild in several areas that were outlined by Gaillard et al. (1994). First, refinements of this concept have the potential to improve the reliability of survival data because, unlike the approach used in virtually all long-term field studies in which only newborn are marked and their survival monitored throughout their lives, this approach estimates survival using information from individuals first marked at any age. Therefore, for many species it may be possible to mark many more individuals than are available from

only a single (newborn) age group and therefore to increase sample size. Second, this approach introduces new biological concepts for measuring senescence in the wild that differ from Nesse's (1988) intensity of selection, Finch's (1990) mortality rate doubling time, Promislow's (1991) log slope mortality and Abrams' (1993) fitness cost of senescence. Additionally, our method is a useful addition to capture–recapture studies with unknown age-at-entry.

The method we outline in this paper focuses on the information content of wild-caught, living individuals and will ultimately not only include information on survival that can be used to estimate actuarial aging as in conventional approaches, but also information on fertility, behaviour, mating and other life history categories that can be used to shed new light on senescence in the wild. The methods we introduce will provide new techniques for expanding the taxonomic horizons of senescence studies in the wild beyond mammals, to include other vertebrates such as birds, reptiles, fishes and amphibians as well as invertebrates ranging from nematodes to insects. These methods will be especially important for studying aging of invertebrates in the wild, such as nematodes, that cannot be marked and released into the wild for later recapture.

A key demographic identity

The data on remaining lifetime after capture and marking, obtained from the marked sample, are assembled in a 'marked sample life table'. Assuming that the process of capture and marking does not alter an individual's remaining lifespan, the corresponding 'marked sample' and 'wild' life tables are compared for a hypothetical situation in Table 1.

That it is possible to obtain the survival schedule in the wild, as summarized by the wild life table, from the marked sample life table, is due to a basic relationship between these two life tables. Assuming that the population is stable, stationary and closed, i.e. is neither increasing nor decreasing, and without immigration or emigration, the number of subjects of age x is $c_x = l_x / \sum l_y = c_0 l_x$ (Caswell, 2001; see Table 1 for definitions). The death rates in the marked sample life table at age x' are by definition $d_{x'}^* = l_{x'}^* - l_{x'+1}^*$. These death rates are generated by subjects that enter the marked sample life table at various (unknown) ages, survive to 'marked age' (i.e. age counted in days after capture and marking) x and do not survive to 'marked age' $x' + 1$. For all subjects that enter the marked sample cohort at age z , the contribution to $d_{x'}^*$ is therefore

$$c_z \frac{l_{z+1}}{l_z} \cdots \frac{l_{z+x'}}{l_{z+x'-1}} \left(1 - \frac{l_{z+x'+1}}{l_{z+x'}} \right) = c_z \left(\frac{l_{z+x'}}{l_z} - \frac{l_{z+x'+1}}{l_z} \right) \\ = c_0 (l_{z+x'} - l_{z+x'+1}),$$

where l_z refers to the survival function or survival schedule of the wild population at age z .

The contributions of subjects entering the marked sample life table at various ages are additive. Therefore, adding the contributions over all ages of entry z :

Table 1 Illustration of the relationship between hypothetical ‘wild’ and ‘marked sample’ life tables in the stationary case (J. R. Carey, unpublished data). The ‘wild’ cohort consists of N_x individuals at each age x with corresponding schedules of survival I_x and age structure $c_x = I_x/\sum_j I_j$, with life table in the leftmost subtable. The ‘marked sample’ cohort consists initially of 20 ‘marked’ individuals with the same age structure as the ‘wild’ cohort, all simultaneously entering the marked sample cohort at the age of capture and marking $x^* = 0$. Remaining lifetimes are recorded for the marked sample, N_{x^*} is the number of animals that remain alive at age x^* after marking, and I_{x^*} is the survival schedule of the marked sample cohort, with death rates $d_{x^*} = I_{x^*+1} - I_{x^*}$, as listed in the rightmost subtable. The survival schedules given separately for age cohorts $x = 0, x = 1, x = 2$ and $x = 3$ in dependency on marked sample cohort age x^* , are listed in the corresponding columns of the subtable in the middle. In this hypothetical example, the initial marked sample cohort at marked sample cohort age $x^* = 0$ has an age structure identical to c_x (row in bold type in the middle subtable is identical to the bold type c_x column of the leftmost subtable). The key identity is revealed by the equality of the bold type columns c_x and d_{x^*} in the leftmost and rightmost subtables. This key relationship allows us to deduce the wild survival schedule from the marked sample survival schedule

Wild cohort				Age distribution in marked sample cohort				Marked sample			
x	N_x	I_x	c_x	$x = 0$	$x = 1$	$x = 2$	$x = 3$	x^*	N_{x^*}	I_{x^*}	d_{x^*}
0	40	1.000	0.40	0.40	0.30	0.25	0.05	0	20	1.00	0.40
1	30	0.750	0.30	0.30	0.25	0.05		1	12	0.60	0.30
2	25	0.625	0.25	0.25	0.05			2	6	0.30	0.25
3	5	0.125	0.05	0.05				3	1	0.05	0.05
4	0	0.000	0.00	0.00				4	0	0.00	0.00
	100	2.5								1.95	1.00

$$d_{x^*}^* = \sum_z c_0 (I_{z+x^*} - I_{z+x^*+1}) = c(0) I_{x^*} = c_{x^*},$$

and this relationship implies that the columns c_x indicating the age distribution in the wild life table and $d_{x^*}^*$ indicating the distribution of deaths in the marked sample life table are identical. We can see from Table 1 that this is indeed the case for the hypothetical case considered there. As $I_x = c_x/c_0$, this relationship between the two life tables leads to

$$I_x = d_{x^*}^*/d_0^*$$

thus enabling the reconstruction of the survival schedule I_x in the wild life table from the survival schedule I_x^* of the marked sample life table.

Statistical estimates implementing this probabilistic relationship can be easily found, for example by plugging in empirical observed frequencies for I_x^* and $d_{x^*}^*$, thus replacing expected population values as they appear in Table 1 with their corresponding sample estimates. Based on the binomial distribution of these observed frequencies, we can derive large-sample confidence intervals for the resulting estimates of I_x . Details and formulas are provided in the Appendix.

Continuous lifetimes

These considerations can be extended to the case in which age-at-death is considered to be measured on a continuous scale and the smooth nature of the underlying survival distributions can be discerned. The power of analysing hazard functions from continuous lifetimes has been illustrated in Müller & Wang (1994). Information loss and recovery of features related to smoothness and derivatives such as hazard rates from aggregated survival data as encountered in life tables are well known (Müller et al., 1997; Wang et al., 1998). For these reasons, it is therefore clearly preferable to work with continuous lifetime

data rather than life tables whenever feasible. Therein lies one of the promises of the proposed methodology – the continuous case is supported without the need to specify a parametric model for the survival distribution as is usually required. The downside of parametric modelling is lack of flexibility because these models are tied to the correctness of the assumed parametric model, and such an assumption cannot be easily verified. The continuous model can be implemented whenever the marked cohorts can be continuously monitored.

In the following, we discuss the continuous version of the key identity. This identity enables us to estimate hazard rates and other continuous features of survival distributions by means of flexible non-parametric curve estimation methods. Denoting by X the age-at-death (lifetime) for an individual in the wild, by $\bar{F}(x) = P(X > x)$ the survival function in the wild, where x is a continuous age variable and P denotes probability, we find for the density of the age-distribution in the wild $c(x) =$

$$\bar{F}(x)/\int_0^\infty \bar{F}(x)dx, \text{ and consequently}$$

$$\bar{F}'(x) = \frac{c(x)}{c(0)}.$$

The unknown age A at the time of capture and marking and the unknown age-at-death X are related with the known remaining lifetime X^* of an individual by $X^* = X - A$. Denote the densities of the distributions of X, X^* by f_X, f_{X^*} and consider the conditional density $f_X(\cdot | X \geq x)$ of lifetime conditional on the event that the individual survives to age x . Then one obtains for the density $f_{X^*}(a)$ of X^* , evaluated at the age-at-death a ,

$$\begin{aligned} f_{X^*}(a) &= \int_0^\infty c(x)f_X(x+a | X \geq x)dx = \int_0^\infty \frac{\bar{F}(x)}{\int_0^\infty \bar{F}(x)dx} \frac{f_X(x+a)}{\bar{F}(x)} dx \\ &= \frac{\bar{F}(a)}{\int_0^\infty \bar{F}(a)da} = c(a) \end{aligned}$$

This relationship implies the key identity for the continuous case,

$$\hat{F}(x) = \frac{\hat{f}_{X^*}(x)}{\hat{f}_{X^*}(0)},$$

providing the relationship between the marked cohort mortality and survival in the wild. This type of relationship has been noted previously in the literature on renewal processes (Doob, 1948; Feller, 1968; Winter, 1989). Statistical estimation and inference based on this continuous version of the key identity is discussed in the next section.

Estimating the survival schedule of the wild population

Given a sample of continuous lifetimes X_1^*, \dots, X_n^* that are observed in the marked sample cohort and measured in terms of relative age counted from the time of marking, we may substitute non-parametric kernel density estimators (compare, for example, Müller, 1997) for $f_{X^*}(z)$, given by

$$\hat{f}_{X^*}(z) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{z - X_i^*}{h}\right).$$

Here $h = h(n)$ is a sequence of bandwidths and K is a kernel function. Specific kernel functions are listed in the Appendix.

We implement a flexible non-parametric smoothing approach that does not depend on an assumption that the survival schedule is likelihood based, which Bayesian methods would require when dealing with continuous lifetimes. Given the enormous plasticity of mortality schedules in biological populations, these methods are very limited in their applicability whereas non-parametric methods do not make any assumptions on the underlying survival distributions except for some basic smoothness. In return, a bandwidth or smoothing parameter h in the above kernel density estimator needs to be specified to control the trade-off between variance and bias of the resulting non-parametric estimates. Methods for data-adaptive specification of bandwidths and also for efficient numerical implementations of the above estimator are described in Müller (1997).

We then obtain asymptotically consistent estimates of the survival function of the wild population,

$$\hat{F}(x) = \frac{\hat{f}_{X^*}(x)}{\hat{f}_{X^*}(0)}.$$

The implementation of this estimate is less straightforward than it may seem. One difficulty is that the estimates $\hat{f}_{X^*}(0)$ that appear in the denominator are density estimates at a boundary point of the support of the data and therefore are subject to higher variability than density estimates in the interior of the support (Müller & Wang, 1994). We replace the kernel K in the definition of the kernel density estimator above by a boundary kernel K_0 when estimating the density of X^* at the boundary point $x = 0$ (see end of Appendix). A second difficulty is that the above estimate is not necessarily a survival function, which

by definition is monotone declining from 1 to 0. This can be ensured by adding a monotonization step through the pool adjacent violators algorithm (PAVA; Robertson et al., 1988).

Using analogous kernel density estimators $\hat{f}'_{X^*}(x)$ for the derivative of f_{X^*} , we may obtain estimates for the density f of the survival schedule of the wild cohort, $\hat{f}(x) = -\hat{f}'_{X^*}(x)/\hat{F}_{X^*}(0)$. Analogously, estimates for the hazard rate $h(x) = f(x)/\hat{F}(x)$ are obtained as $\hat{h}(x) = -\hat{f}'_{X^*}(x)/\hat{f}_{X^*}(x)$. To obtain the density derivative estimates that appear in these formulas we replace the kernel K in the kernel density estimator above by a derivative kernel K_1 (often chosen as $K_1 = K^{(1)}$, see Appendix) and the scaling factor $1/(nh)$ by $1/(nh^2)$. The construction of confidence intervals and thus inference for these non-parametric estimates can be obtained through asymptotic methods. The asymptotic arguments, corresponding variance estimates and resulting formulas for confidence intervals are summarized in the Appendix.

To assess the age at capture for a subject for which an additional lifetime x was observed in the marked cohort life table, we may use the conditional density $f_{A|x^*}(a|x) = f_X(x+a)/\hat{F}(x)$ to infer the conditional expectation

$$E(A|X^* = x) = \frac{1}{\hat{F}(x)} \int_x^\infty (z-x)f_X(z)dz.$$

Plugging the above estimates into the right-hand side of this equation then leads to consistent estimates of conditional mean age at capture. Monotonized density estimates similar to those above were proposed by Watelet & Winter (1991) in a reliability setting.

Illustration

We illustrate the reconstruction of the survival schedule of the wild population from the observations made on the marked sample in a simulation study. The underlying survival schedule of the wild population is modelled as the survival function of a real cohort. The starting point is a cohort consisting of 1000 female Mediterranean fruit flies, *Ceratitis capitata*, commonly known as the medfly, whose survival has been described and analysed in Carey et al. (1998).

Using acceptance-rejection sampling based on the graph of the survival function for these 1000 flies, we randomly sample N flies (with replacement) to create one simulated marked sample. Each of the flies selected for the marked sample has a random age, following the age distribution of the flies in the entire 'wild population', and also an associated remaining lifetime that is recorded as 'marked lifespan'. Kernel density estimation as described above is implemented by local linear smoothing after an initial prebinning step (see Müller, 1997) and combined with the PAVA method.

The resulting survival function estimates, along with the target survival function for six generated marked sample cohorts of sizes $N = 1000$ and $N = 50$, can be seen in Fig. 1. We find that the method of reconstructing the survival schedule of the wild population works very well for the larger sample and

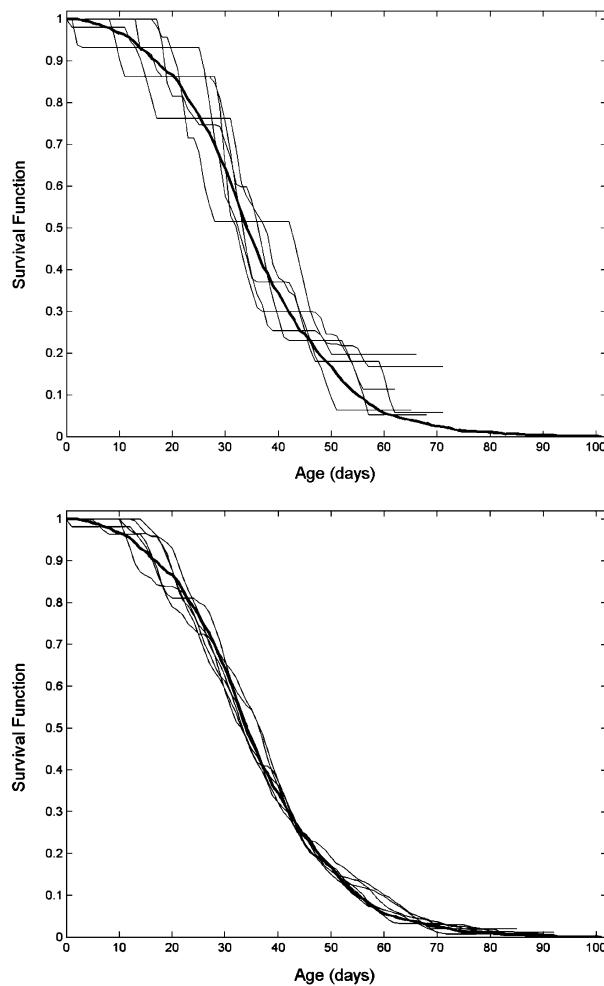


Fig. 1 Reconstruction of survival function of wild population from six simulated marked sample cohorts of sizes $N = 50$ (upper panel) and $N = 1000$ (lower panel). This reconstruction is based on a key demographic identity and corresponding non-parametric estimation methods as described in the text. The solid curve is the target survival function that corresponds to the observed survival schedule of a cohort of medflies.

reasonably well for the smaller sample. The infant survival estimates show a higher degree of variability than the survival estimates for the mid-age period because not very many early deaths will be recorded in the marked cohort.

Discussion: window on aging in the wild and a generalization

In this paper we demonstrated that age-specific life tables can be constructed from mortality data derived from randomly captured individuals of unknown age in stable, stationary and closed populations. The importance of our model is that it provides a starting point to develop more complex models whose purpose is to estimate the life table properties of populations based on more realistic assumptions (non-stable, non-stationary populations). However, we believe that the significance of the

general approach extends beyond the life table and applies to the concept of expressed information content of marked (or captured) individuals. For the current case the expressed information is the remaining post-capture lifespan of marked individuals that is used to estimate the life table of the population at large.

The idea of expressed information content generalizes if it is assumed that: (1) the experiences of individuals early in life influence the expression and pattern of their life history traits (mortality, reproduction, behaviour) later in life; and (2) these patterns expressed in later life can be traced to early life experience. The concept of extracting knowledge of both an individual's age and its early life experience to gain insights into the demographic and gerontological characteristics of the field population can then be used as the conceptual foundation for a new sampling concept for understanding aging in the wild. Examples of the types of information that can be extracted from wild-caught (or marked) individuals at the individual level include remaining lifespan, age-specific reproduction (relative to time of capture), details of reproduction including birth interval, clutch size, post-reproductive period, overall patterns of individual reproduction, total reproduction and time from capture to first egg, timing and magnitude of peak reproduction (Carey *et al.*, 1998; Müller *et al.*, 2001), mating status and frequency of mating, behavioural measures such as supine behaviour (Papadopoulos *et al.*, 2002) or calling (males, see Papadopoulos *et al.*, 2004), mating, oviposition and overall activity, and physiological measures such as metabolic rate.

We believe that this new concept for extracting information about aging in the wild is important for several reasons. First, life course analysis will both encourage and require a deep understanding of the interdependencies of various components of an individual's life course, including reproduction, behaviour and death. In particular, the approach will require an understanding of the relationship between reproduction at young ages and mortality risk at older ages, the age patterns of reproduction that are unique to different stages in the adult life course, and the linkages between different behavioural patterns and death. Second, the approach will encourage a greater integration of laboratory and field studies. Specifically, the method will require the creation of reference 'libraries' consisting of the life history patterns of individuals maintained under different conditions in the laboratory. These 'libraries' will be used for comparing the observed life history patterns (birth and death) of wild-caught flies maintained in the laboratory. Third, the results of studies using the methods we propose to develop will shed new light on both aging and aging structure of wild populations. This includes aging data on populations of invertebrate species such as *C. elegans* that are difficult to study under natural conditions in the wild but that are extraordinarily important model organisms in aging science (Reznick, 1993; Gershon & Gershon, 2002). The combination of laboratory and field studies will provide the means for testing various theories about aging in the wild and also for testing models used in both forecasting and back-casting.

Acknowledgments

This research was supported by NIH grant P01-AG08761 and NSF grant DMS-02-04869. We thank J. Cardenas for technical assistance, L. Harshman, L. Partridge and A. Yashin for discussion, and J. Vaupel and K. Wachter for comments on a previous draft.

References

- Abrams PA (1993) Does increased mortality favor the evolution of more rapid senescence? *Evolution* **47**, 877–887.
- Austad SN (1993) Retarded senescence in an insular population of Virginia opossums (*Didelphis virginiana*). *J. Zool.* **229**, 695–708.
- Carey JR, Liedo P, Müller HG, Wang JL, Chiou JM (1998) Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large cohort of Mediterranean fruit fly females. *J. Gerontol. – Biol. Sci.* **53**, 245–251.
- Caswell H (2001) *Matrix Population Models: Construction, Analysis, and Interpretation*, 2nd edn. Sunderland, MA: Sinauer Associates.
- Congdon JD, Dunham AE, VanLobenSels RC (1994) Demographics of common snapping turtles (*Chelydra serpentina*): implications for conservation and management of long-lived organisms. *Am. Zool.* **34**, 397–408.
- Doob JL (1948) Renewal theory from the point of view of the theory of probability. *Trans. Am. Mathemat. Soc.* **63**, 429.
- Feller W (1968) *Introduction to Probability Theory*, Vol. I. New York: Wiley.
- Finch CE (1990) *Longevity, Senescence, and the Genome*. Chicago: The University of Chicago Press.
- Finch CE (2001) History and prospects: symposium on organisms with slow aging. *Exp. Gerontol.* **36**, 593–597.
- Gaillard JM, Allaine D, Pontier D, Yoccoz NG, Promislow DEL (1994) Senescence in natural populations of mammals: a reanalysis. *Evolution* **48**, 509–516.
- Gershon H, Gershon D (2002) *Caenorhabditis elegans* – a paradigm for aging research: advantages and limitations. *Mechanisms Ageing Dev.* **123**, 261–274.
- Hawkes K, O'Connell JF, Jones NGB, Alvarez H, Charnov EL (1998) Grandmothering, menopause, and the evolution of human life histories. *Proc. Natl Acad. Sci. USA* **95**, 1336–1339.
- Hill K, Hurtado AM (1996) *Ache Life History*. New York: A. de Gruyter.
- Howell N (1979) *Demography of the Dobe !Kung*. New York: Academic Press.
- Jones NGB, Hawkes K, O'Connell JF (2002) Antiquity of postreproductive life: are there modern impacts on hunter-gatherer postreproductive lifespans? *Am. J. Hum. Biol.* **14**, 184–205.
- Lebreton JD, Burnham KP, Clobert J, Anderson DR (1992) Modelling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecol. Monogr.* **62**, 67–118.
- Müller HG (1997) Density estimation. In *Encyclopedia of Statistical Science* (eds Kotz S, Read CB, Banks DL). New York: Wiley, pp. 185–200.
- Müller HG, Carey JR, Wu D, Liedo P, Vaupel JW (2001) Reproductive potential predicts longevity of female Mediterranean fruit flies. *Proc. Royal Soc. B* **268**, 445–450.
- Müller HG, Love B, Hoppa R. (2002) A semiparametric method for estimating paleodemographic profiles from age indicator data. *Am. J. Phys. Anthropol.* **116**, 1–14.
- Müller HG, Wang JL (1994) Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50**, 61–76.
- Müller HG, Wang JL, Capra WB (1997) From lifetables to hazard rates: the transformation approach. *Biometrika* **84**, 881–892.
- Nesse RM (1988) Life table tests of evolutionary theories of senescence. *Exp. Gerontol.* **23**, 445–453.
- Papadopoulos NT, Carey JR, Katsoyannos BI, Kouloussis NA, Müller HG, Liu X (2002) Supine behaviour predicts time to death in Mediterranean fruit flies (*Ceratitis capitata*). *Proc. Royal Soc. B* **269**, 1633–1637.
- Papadopoulos NT, Katsoyannos BI, Kouloussis NA, Carey JR, Müller HG, Zhang Y (2004) High sexual calling rates predict extended lifespan in male Mediterranean fruit flies. *Oecologia* **138**, 127–134.
- Promislow DEL (1991) Senescence in natural populations of mammals: a comparative study. *Evolution* **45**, 1869–1887.
- Reznick D (1993) New model systems for studying the evolutionary biology of aging: crustacea. *Genetica* **91**, 79–88.
- Reznick D, Buckwalter G, Gro. J, Elder D (2001) The evolution of senescence in natural populations of guppies (*Poecilia reticulata*): a comparative approach. *Exp. Gerontol.* **36**, 791–812.
- Robertson T, Wright FT, Dykstra RL (1988) *Order Restricted Statistical Inference*. New York: Wiley.
- Tatar M, Yin CM (2001) Slow aging during insect reproductive diapause: why butterflies, grasshoppers and flies are like worms. *Exp. Gerontol.* **36**, 723–738.
- Udevitz MS, Ballachey BE (1998) Estimating survival rates with age-structure data. *J. Wildlife Manage.* **62**, 779–792.
- Wang JL, Müller HG, Capra WB (1998) Analysis of oldest-old mortality: lifetables revisited. *Ann. Statistics* **26**, 126–163.
- Watelet L, Winter BB (1991) Non-parametric estimation of nonincreasing densities and use of data from renewal processes. *Comms. Statistics – Theory Meth.* **20**, 2073–2094.
- Williams BK, Nichols JD, Conroy MJ (2002) *Analysis and Management of Animal Populations*. London: Academic Press.
- Winter BB (1989) Joint simulation and forward recurrence times in a renewal process. *J. Appl. Probability* **26**, 404–407.

Appendix: asymptotic confidence intervals and variances

Based on the estimation of the survival schedule of the wild population, one can derive asymptotic confidence intervals for important characteristics of the survival schedule of the wild population. This includes confidence intervals and associated inference for the survival function $\hat{F}(x)$ for discrete and continuous lifetimes, and the density $f_{X^*}(x)$ and hazard rate $h_{X^*}(x)$ for continuous lifetimes. Another option is to employ a suitable bootstrap.

We first investigate confidence intervals for the survival function $\hat{F}(x)$ for discrete lifetimes, i.e. the survival schedule at age x , given by $I_x = d_x^*/d_0^*$, where x is an arbitrary non-negative integer. Let \hat{d}_0^* and \hat{d}_x^* denote the estimates obtained by plugging in empirical observed frequencies for d_0^* and d_x^* . Let $W_n(x)$ denote the number of deaths in $(x, x+1]$. It is easily seen that $W_n(x) \sim \mathbf{B}(n, d_x^*)$, for $x = 0, 1, \dots$, and $\hat{d}_x^*/\hat{d}_0^* = W_n(x)/W_n(0)$, where $\mathbf{B}(n, d_x^*)$ denotes the binomial distribution with n trials and probability of success d_x^* , and n is the total number of subjects. Then from the central limit theorem, one can obtain the asymptotic joint distribution of the multinomial random variable $(W_n(x), W_n(0))^T$, which is $\mathbf{N}_2((d_x^*, d_0^*)^T, \Sigma)$, where \mathbf{N}_2 denotes the bivariate normal distribution, and Σ is a 2×2 matrix with $(\Sigma)_{11} = d_x^*(1 - d_x^*)$, $(\Sigma)_{22} = d_0^*(1 - d_0^*)$ and with $(\Sigma)_{12} = (\Sigma)_{21} = -d_x^*d_0^*$. Applying the delta method leads to the asymptotic normal distribution of $\hat{F}(x)$,

$$(I_x - I_x)/\sqrt{n} \xrightarrow{D} \mathbf{N}\left(0, \frac{d_x^*(1 - d_x^*)}{d_0^{*2}} + \frac{d_x^{*2}(1 - d_0^*)}{d_0^{*3}} - \frac{d_x^{*2}}{d_0^{*2}}\right).$$

Then the $100(1 - \alpha)\%$ confidence interval of I_x is obtained by substituting the empirical estimates of I_x and applying Slutsky's theorem:

$$\hat{I}_x \pm \Phi(1 - \alpha/2) \sqrt{\hat{d}_x^*(1 - \hat{d}_x^*)/\hat{d}_0^{*2} + \hat{d}_x^{*2}(1 - \hat{d}_0^*)/\hat{d}_0^{*3} - 2\hat{d}_x^{*2}/\hat{d}_0^{*2}}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable.

For the case of continuous lifetimes, the survival function is estimated by $\hat{F}(x) = \hat{f}_{X^*}(x)/\hat{f}_{X^*}(0)$. Assume that a kernel K supported on $[-1, 1]$ is used for $\hat{f}_{X^*}(x)$ and the boundary kernel K_0 supported on $[-1, 0]$ for $\hat{f}_{X^*}(0)$. The bandwidth h for the kernel density estimates $\hat{f}_{X^*}(x)$ and $\hat{f}_{X^*}(0)$ is assumed to satisfy $h \rightarrow 0$ and $nh \rightarrow \infty$, as $n \rightarrow \infty$. For any fixed $x > 0$, when n is sufficiently large, one has $h < x - h$, i.e. no X_i^* 's are included in both $[0, h]$ and $[x - h, x + h]$, whence the estimates $\hat{f}_{X^*}(x)$ and $\hat{f}_{X^*}(0)$ are asymptotically independent. From standard results for kernel density estimation (see Müller, 1997, for references), one can easily obtain the asymptotic joint distribution of $[\hat{f}_{X^*}(x), \hat{f}_{X^*}(0)]$ as follows,

$$\sqrt{nh}\{\hat{f}_{X^*}(x) - E[\hat{f}_{X^*}(x)], \hat{f}_{X^*}(0) - E[\hat{f}_{X^*}(0)]\} \xrightarrow{D} \mathbf{N}_2\left(0, \begin{pmatrix} f_{X^*}(x) \| K \|^2 & 0 \\ 0 & f_{X^*}(0) \| K_0 \|^2 \end{pmatrix}\right),$$

which is bivariate normal with mean vector 0. Here $\| K \|^2 = \int K^2(u)du$ and $\| K_0 \|^2 = \int K_0^2(u)du$. Because the bias $E(\hat{f}_{X^*}(x)) - f_{X^*}(x) = O(h^2)$ for both $x = 0$ and $x > 0$, we can ignore biases for small values of h . Assuming this is the case and applying the delta method, we obtain the asymptotic normal approximation to the distribution of $\hat{F}(x) = \hat{f}_{X^*}(x)/\hat{f}_{X^*}(0)$:

$$\hat{F}(x) - F(x) \approx N\left(0, \frac{1}{nh} \left[\frac{f_{X^*}(x) \| K \|^2}{f_{X^*}(0)} + \frac{f_{X^*}^2(x) \| K_0 \|^2}{f_{X^*}^3(0)} \right]\right).$$

Then the $100(1 - \alpha)\%$ confidence interval for $\hat{F}(x)$ is obtained by substituting the consistent kernel estimates $\hat{f}_{X^*}(x)$ and $\hat{f}_{X^*}(0)$ for $f_{X^*}(x)$ and $f_{X^*}(0)$ in the formula, applying Slutsky's theorem, i.e. the $100(1 - \alpha)\%$ confidence interval for $\hat{F}(x)$ is

$$\hat{F}(x) \pm \Phi(1 - \alpha/2) \sqrt{[\hat{f}_{X^*}(x) \| K \|^2/\hat{f}_{X^*}^2(0) + \hat{f}_{X^*}^2(x) \| K_0 \|^2/\hat{f}_{X^*}^3(0)]/(nh)}$$

To construct the confidence interval for the density estimate $\hat{f}(x) = \hat{f}'_{X^*}(x)/\hat{f}_{X^*}(0)$, we note that the derivative estimate $\hat{f}'_{X^*}(x)$ has slower convergence rate than $\hat{f}_{X^*}(0)$. Slutsky's theorem implies that $\hat{f}'_{X^*}(x)/\hat{f}_{X^*}(0)$ is asymptotically equivalent to $f'_{X^*}(x)/f_{X^*}(0)$. From the asymptotic distribution of the kernel estimator for the derivative $\hat{f}'_{X^*}(x)$, and ignoring the bias terms as argued earlier, one has

$$\hat{f}'_{X^*}(x) - f'_{X^*}(x) \approx N(0, f_{X^*}(x) \| K_1 \|^2/(nh^3)),$$

where K_1 is the kernel function used in $\hat{f}'_{X^*}(x)$. Thus the asymptotic distribution of the density estimate $\hat{f}'_{X^*}(x)$ is approximately $\mathbf{N}(0, f_{X^*}(x) \| K_1 \|^2/[nh^3 f_{X^*}^2(0)])$, and the $100(1 - \alpha)\%$ confidence intervals can be obtained by substituting the kernel estimates for $\hat{f}'_{X^*}(x)$ and $f_{X^*}(0)$, whence one obtains the intervals

$$\hat{f}'(x) \pm \Phi(1 - \alpha/2) \sqrt{\hat{f}_{X^*}(x) \| K_1 \|^2}/[(nh^3) \hat{f}_{X^*}^2(0)].$$

Similarly, the $100(1 - \alpha)\%$ confidence interval for the hazard rate $h(x)$, estimated by $\hat{h}(x) = \hat{f}'_{X^*}(x)/\hat{f}_{X^*}(x)$, is obtained by

$$\hat{h}(x) \pm \Phi(1 - \alpha/2) \sqrt{\hat{f}_{X^*}(x) \| K_1 \|^2}/[(nh^3) \hat{f}_{X^*}^2(x)].$$

We note that common choices for kernels for interior, boundary and derivative estimation K , K_0 and K_1 are $K(x) = 0.75(1 - x^2)$ on $[-1, 1]$, $K_0(x) = 12(x + 1)(x + 1/2)$ on $[-1, 0]$ and $K_1(x) = -(3/2)x$ on $[-1, 1]$.